

THÈSE de DOCTORAT de L'UNIVERSITÉ du MAINE

Spécialité :
ACOUSTIQUE

Présentée par
M. Olivier HOUIX

pour obtenir le grade de
DOCTEUR de L'UNIVERSITÉ du MAINE

Sujet de la thèse

Catégorisation auditive des sources sonores

Thèse soutenue le 9 juillet 2003 devant le jury composé de :

Emmanuel Bigand (Rapporteur)
Georges Canévet (Rapporteur)
René Caussé (Directeur de thèse)
Antoine Chaigne (Examineur)
Joël Gilbert (Examineur)
Yves Guiard (Examineur)
Stephen McAdams (Directeur de thèse)

Résumé

Dans cette thèse, nous cherchons à définir les relations qui existent entre un événement sonore, contraint par ses caractéristiques physiques vibratoires, et la perception de celui-ci. La plupart des travaux sur l'audition se focalisent sur les dimensions perceptives des sons ou sur les processus de base qui traitent les signaux sonores, mais indépendamment des sources qui les produisent. Nous orientons l'étude vers la description des sources sonores, ici des barres de métal frappées, en termes de leurs propriétés audibles. Ces propriétés spécifient les caractéristiques physiques des sources telles que la géométrie, le matériau, le type d'interaction avec d'autres objets. Nous focalisons sur la perception de la géométrie. D'un point de vue théorique, notre approche est d'associer l'organisation catégorielle de nos connaissances et la perception des événements sonores. Nous l'abordons sous l'angle de l'acoustique écologique. Nous présentons aussi une synthèse sur les techniques d'analyses des données pour la modélisation et la représentation des proximités, ainsi que de nouvelles techniques de validation pour l'analyse des données de classification. Ces techniques nous permettent d'élaborer des modèles perceptifs et de mettre en lumière les propriétés physiques des sons qui les sous-tendent. Une première série d'expériences a mis en jeu la classification de sons de barres de métal excitées en différents points. Notre but était de savoir si les auditeurs arrivent à extraire des invariants perceptifs propres à chaque barre, spécifiant leurs propriétés géométriques. Les résultats suggèrent que les auditeurs utilisent les hauteurs spectrales proéminentes liées aux modes vibratoires d'une dimension géométrique privilégiée (largeur ou épaisseur de la barre) afin de comparer et classer ces sons. Il n'apparaît pas que les auditeurs utilisent l'information acoustique brute afin d'extraire des invariants perceptifs propres à une barre, bien que ces informations soient présentes et pourraient être utilisées à cette fin. Une deuxième série d'expériences a utilisé un paradigme d'appariement intermodal entre la représentation visuelle de paires de sources sonores et le son produit par celles-ci. Grâce à l'utilisation de sons de synthèse, nous avons cherché à tester quelles informations acoustiques liées aux modes de vibration permettent de spécifier la perception de la géométrie des barres. Les résultats montrent que les auditeurs mettent en place des stratégies perceptives afin d'associer un son avec sa représentation visuelle. Ils n'étaient pas l'hypothèse selon laquelle il existe une perception directe des propriétés géométriques de ces sources.

Summary

In this thesis, we sought to define the relations between a sound event, constrained by its physical vibratory characteristics, and its perception. Most work on hearing focuses on the perceptual dimensions of sound or on the basic mechanisms that process sound signals, but independently of the sources that produce them. We orient this work toward the description of sound sources, struck metal bars in our case, in terms of their audible properties. These properties specify the physical characteristics of the sources such as their geometry, material, and interactions with other objects. We focus on the perception of geometry. From a theoretical standpoint, our approach is to associate the categorical organization of knowledge with the perception of sound events, adopting the approach of ecological acoustics. The techniques of data analysis for modeling and representing proximities are summarized, and new techniques for the validation of classification data analysis are developed. These techniques allow us to elaborate perceptual models and to reveal the physical properties of sounds that give rise to them. A first series of experiments involves the classification of sounds of metal bars struck at different positions. The aim was to determine whether listeners can extract the perceptual invariants specific to each bar's geometric properties. The results suggest that listeners use prominent spectral pitches related to the vibration modes of a privileged geometric dimension (width or thickness of the bar) to compare and classify the sounds. They do not seem to use the raw acoustic information to extract the perceptual invariants that specify the bar, even though this information is present and could be used to this end. A second series of experiments used a cross-modal matching paradigm between the visual representation of pairs of sound sources and the sounds produced by them. Using synthesized sounds, we sought to determine the acoustic information related to vibration modes that specify the perception of bar geometry. The results show that listeners use perceptual strategies in order to associate sounds with their visual representation, but they do not support the hypothesis that a direct perception of the geometric properties of these sources exists.

Remerciements

Je voudrais tout d'abord remercier Stephen McAdams et René Caussé qui m'ont offert par ce sujet un champ d'ouvertures vers d'autres disciplines et en prenant le risque de me laisser m'y plonger. Je voudrais aussi les remercier pour tous les aspects humains qui ont permis cette entreprise.

Je remercie Emmanuel Bigand et Georges Canévet d'avoir accepté d'examiner et de discuter ce document.

Je remercie Joël Bensoam ainsi que Christophe Vergez pour leur contribution sur la modélisation par éléments finis ; Xavier Rodet et Goeffroy Peeters pour leur aide dans l'analyse des signaux.

Je remercie Hugues Vinet, directeur scientifique de l'Ircam, pour son accueil dans cette grande maison qu'est l'Ircam.

Je remercie tout particulièrement les personnes de l'équipe Perception et Cognition Musicales : Patrick, Daniel, Bennett, Sandrine, Anne, Julien, Alain, Jeremy, Mondher et Guillaume pour tous les "à côté". Je n'oublie pas pour autant les nombreuses personnes de l'Ircam.

Un grand merci à David, pour son amitié, son soutien et sa présence dans les derniers instants, à Marie France pour sa relecture et sa sympathie, à mes amis, à Johanna.

Je remercie avec tendresse les membres de ma famille, en m'excusant de les avoir souvent abandonnés pour rejoindre les sous-sols de l'Ircam.

Pour l'inspiration, merci à High Tone pour leurs basses profondes et envoûtantes et à Philippe Moroux, pour ses photos^[a] insérées dans ce manuscrit.

A Typhaine

^[a]<http://utopia.knoware.nl/users/philippe/>

Table des matières

Remerciements	iii
1 Introduction	1
1.1 Problématique générale	1
1.2 Démarche générale	3
1.3 Plan de l'étude	4
2 Ecognition	5
2.1 Introduction	6
2.2 De la perception à la cognition	7
2.2.1 Introduction	7
2.2.2 Vers une convergence	8
2.2.3 Quelle approche pour l'étude de la perception?	9
2.2.4 Ecologie de la perception	9
2.3 L'Homme et son environnement sonore	10
2.3.1 Introduction	10
2.3.2 Niveaux de description d'un d'événement sonore	13
2.3.3 Etudes des événements sonores	15
2.4 Approche psychophysique : le timbre	16
2.4.1 Introduction	16
2.4.2 Perception du timbre musical	16
2.4.3 Limites de cette approche	19
2.5 Acoustique écologique	19
2.5.1 Introduction	19
2.5.2 Spécification des événements	20
2.5.3 Propriétés physiques potentielles	21
2.5.4 Propriétés audibles des sources sonores	24
2.6 La reconnaissance des sources sonores	31
2.6.1 Introduction	31
2.6.2 Deux approches de la perception	32
2.6.3 Etapes de la reconnaissance	35

2.6.4	Discussion du modèle	37
2.7	Catégories et événements sonores	40
2.7.1	Introduction	40
2.7.2	Définition de la catégorisation	40
2.7.3	Catégorisation et mémoire	41
2.7.4	But de la catégorisation	41
2.7.5	Théories des catégories	47
2.7.6	Discussion	57
2.8	Similarité	58
2.8.1	Introduction	58
2.8.2	Dissociation entre similarité et catégorisation	58
2.8.3	Flexibilité de la similarité	59
2.8.4	Similarité base de la catégorisation	61
2.9	Discussion	65
3	Représentation des proximités	67
3.1	Introduction	67
3.1.1	Notion de proximité	68
3.1.2	Modèles	68
3.1.3	Synopsis d'une analyse	69
3.2	Les proximités	69
3.2.1	Introduction	69
3.2.2	Jugements de proximité	70
3.2.3	Classification	71
3.2.4	Remarque	74
3.3	Différents modèles de la proximité	74
3.3.1	Introduction	74
3.3.2	Approche géométrique	76
3.3.3	Modèle de contraste	81
3.3.4	Discussion sur les modèles	83
3.4	Représentation des proximités	84
3.4.1	Introduction	84
3.4.2	Analyses multidimensionnelles	85
3.4.3	Représentations arborées	87
3.4.4	Arbre additif	92
3.4.5	Arbre hiérarchique	94
3.5	Comparaison des partitions	105
3.5.1	Indice de Rand	105
3.5.2	Partition optimale : une nouvelle approche	107
3.6	Discussion	111

4	Analyse des signaux et modélisation mécanique	116
4.1	Introduction	116
4.2	Analyses acoustiques	117
4.2.1	Méthode paramétrique	117
4.2.2	Analyse en Modèles de Résonance : modRes	120
4.2.3	Comparaison des deux méthodes	122
4.3	Synthèse par FOF	123
4.3.1	Introduction	123
4.3.2	Formes d'Onde Formantiques	124
4.4	Méthode des éléments finis	125
4.4.1	Mouvements des barres	126
4.4.2	Principe général des éléments finis	126
4.4.3	Etapas de la méthode	127
4.5	Discussion	131
5	Classification des sources sonores	132
5.1	Introduction	132
5.1.1	Cadre d'étude	132
5.1.2	Hypothèses	133
5.2	Méthode	134
5.2.1	Présentation des stimuli	134
5.2.2	Enregistrements	137
5.2.3	Protocole expérimental	137
5.2.4	Matériel	138
5.2.5	Participants	139
5.3	Analyse des classifications	139
5.3.1	Analyse des données brutes	139
5.3.2	Analyse des différences interindividuelles	140
5.3.3	Analyse des groupements collectifs	143
5.3.4	Hypothèses	156
5.4	Modèle basé sur les analyses acoustiques	157
5.5	Expérience d'appariement	159
5.5.1	Protocole expérimental	160
5.5.2	Transformation des données	162
5.5.3	Modèle de contraste	165
5.6	Discussion	174
6	Représentation des sources sonores	177
6.1	Introduction	177
6.2	Représentation d'une source sonore	178
6.2.1	Rappel des travaux antérieurs	178

6.2.2	Hypothèses	179
6.3	Construction des stimuli	179
6.3.1	Sons originaux	180
6.3.2	Analyse et synthèse	180
6.4	Expériences	187
6.4.1	Protocole expérimental	187
6.4.2	Déroulement	188
6.4.3	Remarques	189
6.4.4	Matériel	189
6.4.5	Participants	190
6.5	Analyses des données	190
6.5.1	Analyse des données brutes	190
6.5.2	Transformation des données	192
6.5.3	Analyse des dissemblances	193
6.5.4	Analyses multidimensionnelles	198
6.5.5	Corrélat acoustiques potentiels	204
6.5.6	Analyses des corrélations	206
6.6	Discussion	213
7	Conclusions	216
7.1	Introduction	216
7.2	Synthèse des résultats	217
7.2.1	Catégorisation auditives	217
7.2.2	Représentation mentale des sources	219
7.3	Réflexions	220
A	Mécanique	222
A.1	Théorie des barres	223
A.1.1	Vibrations longitudinales	223
A.1.2	Vibrations de torsion	224
A.1.3	Vibrations de flexion	225
A.2	Résolution des équations de propagation	227
A.2.1	Vibrations longitudinales	227
A.2.2	Vibrations de torsion	228
A.2.3	Vibrations de flexion	228
B	Consignes	229
B.1	Classification : « non orientée »	229
B.2	Classification : « orientée »	230
B.3	Classification : « contrainte »	230
B.4	Appariement	231

B.5	Expérience « représentation »	233
C	Expériences de classification	235
C.1	Concordance entre individus	236
C.1.1	Classification non-orientée	236
C.1.2	Classification orientée	237
C.1.3	Classification contrainte	238
C.2	Arbres additifs	239
C.2.1	Classification non-orientée	239
C.2.2	Classification orientée	240
C.2.3	Classification contrainte	241
C.3	Analyses M.D.S	242
C.3.1	Classification non-orientée	242
C.3.2	Classification orientée	243
C.3.3	Classification contrainte	244
C.4	Analyses acoustiques	245
C.5	Modèle appariement	246
C.5.1	Modèle de contraste	246
D	Représentation des sources sonores	249
D.1	modRes : exemple de fichier SDIF	250
D.2	Code CASTEM pour une barre	251

Chapitre 1

Introduction



1.1 Problématique générale

Mark Kac [Kac66] a posé un jour la question : « pouvez-vous entendre la forme d'un tambour ? ». Cette question peut paraître saugrenue mais elle révèle une activité auditive que nous pratiquons tous les jours. En effet, immergés dans une ambiance sonore, nous associons constamment les sons à leur source et essayons de les décrire.

La figure 1.1 montre un cadre d'écoute général. Lorsque nous écoutons un son, nous pouvons écouter aussi bien le son, indépendamment de sa source, que la source elle-même pour qu'elle nous informe sur ses propriétés. Nous voyons, ici, que nous pouvons décrire les sons sur différents niveaux : mécanique, acoustique et perceptif. D'un point de vue psychophysique, nous essayons de faire le lien entre ces différents niveaux.

La reconnaissance des événements sonores est généralement associée à la perception des sons instrumentaux et plus spécifiquement abordée par l'étude de la perception du timbre [Gre77] [Wes79] [Kru89] [MWD⁺95]. Ces études psychoacoustiques ont permis de dégager des dimensions perceptives du timbre qui peuvent être impliquées dans la reconnaissance des événements

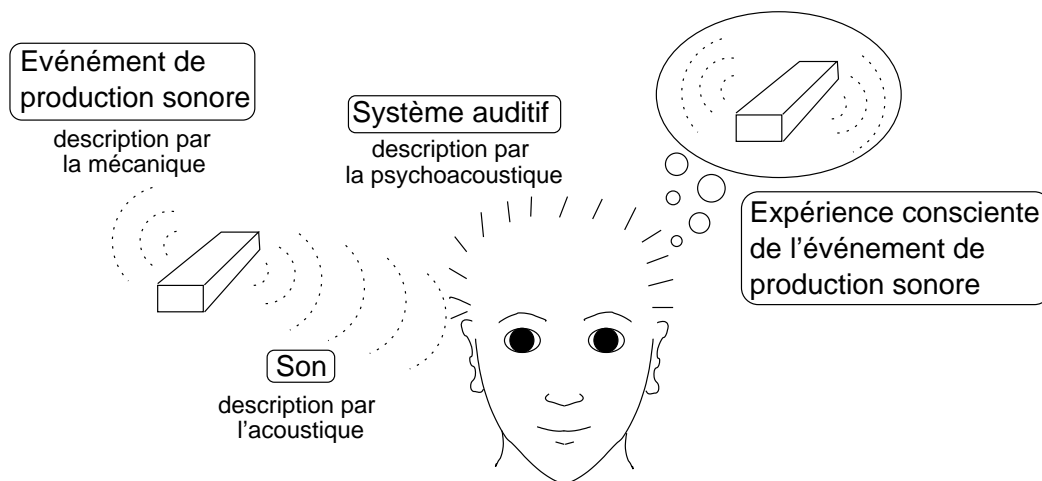


Figure 1.1 - Processus allant de l'événement de la production sonore à l'expérience de cet événement par l'auditeur. Premièrement, un événement externe (décrit par la mécanique) produit une vibration dans le milieu environnant (décrit par l'acoustique). Cette vibration est traitée par le système auditif (décrit par la psychoacoustique) et produit une expérience de la source que nous devons étudier. Figure tirée de [Gav88].

sonores. Mais, comme d'autres travaux qui se focalisent sur les propriétés perceptives telles que la sonie, la hauteur, ..., la perception des événements sonores est étudiée dans ces cas, indépendamment des propriétés mécaniques des sources qui les produisent.

Parallèlement à cette démarche, l'acoustique écologique [Gav93] [CWTss] a pour but de définir les relations perceptives qui existent entre un événement sonore, contraint par ses caractéristiques physiques vibratoires, et la perception de celui-ci. Cette approche permet de définir les propriétés audibles des sources sonores nous permettant de décrire, à travers la perception, leurs propriétés physiques telles que la géométrie, le matériau, le type d'interactions entre objets. Ces propriétés perceptives, orientées vers la description de la source, peuvent être, elles aussi, une source potentielle d'informations pour la reconnaissance des événements sonores.

Ces deux approches, l'une, basée sur l'étude de la perception du timbre et l'autre sur les propriétés audibles des sources sonores, tentent chacune de proposer une base théorique pour la reconnaissance des événements sonores. Mais ces deux approches tendent à s'opposer quant à la nature des processus mis en jeu. Dans le cas d'une approche associée à la théorie du traitement de l'information, la reconnaissance de la source sonore est effectuée par des traitements perceptifs successifs qui permettent l'élaboration des propriétés pertinentes, telles que le timbre et l'activation en mémoire de nos connais-

sances, par une représentation interne de l'événement. L'approche écologique, quant à elle, postule que l'information provenant de la source sonore est riche et structurée et que nos systèmes perceptifs n'ont qu'à saisir cette information pour la reconnaissance des événements sonores.

Nous voyons ici l'importance de la relation et des interactions que nous entretenons avec notre environnement sonore, à travers ces deux théories. La théorie du traitement de l'information met l'emphase sur nos connaissances de notre environnement qui permettent de mettre en relation nos expériences passées avec nos expériences présentes, par la catégorisation. La théorie écologique part du principe que nous sommes « accordés » à notre environnement, c'est l'interaction avec celui-ci qui nous rend sensibles à la structure du monde externe. Pour l'étude de la perception des sources sonores, nous voyons ici l'importance de prendre en compte ces deux approches théoriques, pour une approche complète des processus perceptifs.

1.2 Démarche générale

Pour aborder la perception des propriétés des sources sonores, nos études ont porté sur une famille d'objets simples : des barres de métal frappées. Nous étudions l'acuité de l'audition à juger la perception de la section géométrique de barres de métal.

Nos travaux font suite à ceux de Lakatos et coll. [LMC97] sur la représentation des sources sonores. Cette étude a montré que les auditeurs se représentent la section géométrique de barres frappées (métal ou bois) en utilisant des informations acoustiques vraisemblablement basées sur le rapport épaisseur/largeur de la section géométrique des barres. Ces informations sont liées aux propriétés vibratoires des barres.

Nous avons étendu les résultats de ces expériences en étudiant la perception de ces sources, afin de tester si la représentation auditive de la géométrie des barres peut être une source d'information potentielle pour la reconnaissance de ces objets. Pour cela, nous étudions la catégorisation auditive de ces événements sonores. Nous testons aussi les résultats des expériences de Lakatos et coll., afin de mettre en lumière l'information physique réellement impliquée dans la représentation auditive de ces sources sonores.

De plus, cette thèse a pour but de confronter, d'un point de vue théorique, l'approche du traitement de l'information et l'approche écologique, qui présentent deux cadres théoriques de la perception des sources sonores. D'un point de vue méthodologique, nous présentons une synthèse critique des méthodes d'analyse de données expérimentales, notamment les représentations arborées des données de proximité et proposons des techniques de validation

de ces analyses. Ces méthodes d'analyse ont été associées à des modèles de classement permettant l'étude des processus impliqués dans la catégorisation auditive.

1.3 Plan de l'étude

Nous introduisons dans le chapitre 2, la relation entre l'organisation catégorielle de nos connaissances et la perception des événements sonores. La spécificité de notre démarche est d'aborder sous l'angle de l'acoustique écologique, le lien entre les descriptions perceptive et physique des sources sonores.

Les techniques d'analyse des données expérimentales introduites dans le chapitre 3, nous permettent d'élaborer des modèles perceptifs et de mettre en lumière les propriétés physiques des sons qui sous-tendent ces processus.

Dans le chapitre 4, nous présentons des outils de traitement de signal, propres à la caractérisation des sons de type impulsif, tels que les barres de métal frappées par un marteau. Ces analyses ainsi que l'étude vibratoire des barres par éléments finis, nous ont permis la resynthèse de sons modifiés de barres de métal, utilisés dans la seconde série d'expériences.

Deux séries d'expériences sont présentées dans les chapitres 5 et 6. Une première série d'expériences met en jeu la classification de sons de barres de métal excitées en différents points. Notre but est de connaître si les auditeurs arrivent à extraire des invariances perceptives propres à chaque barre, spécifiant leurs propriétés géométriques. Une deuxième série d'expériences reprend un paradigme d'appariement intermodal entre la représentation visuelle de la source sonore (la section d'une barre de métal) et le son produit par celle-ci. Grâce à l'utilisation de sons de synthèse, nous cherchons à tester quelles informations acoustiques permettent de spécifier la perception de la géométrie des barres. Enfin, nous concluons sur les résultats de ces expériences en les mettant en perspective avec notre approche théorique.

Dans les annexes nous introduisons la théorie mécanique des mouvements vibratoires des barres afin d'initier le lecteur non familiarisé à ce domaine. Les consignes des expériences ainsi que les détails de l'analyse des résultats expérimentaux, sont également présentés en annexe.

Chapitre 2

Ecognition



« La laideur est de la beauté au repos : quand il parlait, sa voix était enrouée et sourde, elle avait encore quelques stries acides qui étaient comme des craquelures, des gerçures, et songeant à la beauté de sa voix quand il chantait, j'examinai cette voix parlée avec plus d'attention. Je fis cette découverte : c'était l'enrouement énervant, qui forcé par le chant, se transformait en une teinte velouté, si douce, et les craquelures devenaient les notes les plus claires. C'est quelque chose comme si, en filant d'une pelote au repos, ces notes se fussent épurées. Un physicien expliquera très bien ce phénomène, moi je reste troublé devant lui qui m'indiquait que la beauté est la projection de la laideur et qu'en « développant » certaines monstruosité, on obtient les plus purs ornements. »

Jean Genet, *Miracle de la rose*

2.1 Introduction

Il va de soi que nous interagissons avec l'environnement dans lequel nous évoluons. Cette interaction est vitale car elle permet une connaissance de notre environnement, le développement des stratégies d'action et des conduites adaptatives. Cette interaction est complexe car malgré notre déplacement et celui des objets, nous continuons à percevoir un monde stable et organisé [Gar87]. Cette complexité se prolonge aussi dans nos expériences introspectives lorsqu'en l'absence de sources sonores nous arrivons à créer notre propre « musique interne ».

L'interaction privilégiée avec le monde externe passe par la perception. La conceptualisation de cette interaction reflète la vision de notre rapport au monde. La perception apparaît comme une double entité et ce caractère dual nourrit les débats philosophiques. Elle permet un accès à la réalité mais en même temps elle est une expérience sensible. « La perception se fait là-bas dans le monde et elle se fait en nous. » [Bar94]. Berkeley [Ber70] note que lorsque nous entendons passer une voiture dans la rue, il n'y a que le son qui est perçu immédiatement, mais l'expérience que nous avons acquise de la connexion d'un son semblable avec une voiture nous fait dire que nous entendons une voiture. Le son est entendu, mais la voiture est suggérée par association.

Nous sommes ainsi amenés à nous questionner sur cette association : comment s'effectue-t-elle, c'est-à-dire pourquoi telle sensation va-t-elle réactiver tel autre groupe de sensations (ou de concepts) habituellement associées ? À l'évidence, il faut que quelque chose, dans la sensation, oriente et limite le recours au passé. Ce qui revient à dire que la sensation doit déjà être saisie comme sensation de cet objet [Bar94].

Ce constat peut apparaître comme une impasse, mais il laisse entrevoir les problématiques auxquelles les théories de la perception sont confrontées. Les informations nécessaires à la perception existent-elles dans le monde extérieur ou avons-nous déjà une certaine connaissance et certaines attentes lors de l'acte de perception [Gar87] ? Ces questions nous amènent à resituer la perception et la cognition dans un cadre plus « écologique » des interactions « humain-environnement » et ainsi de définir les traitements mis en jeu, d'où le néologisme du titre de ce chapitre.

Dans la continuité de cette problématique, Bregman [Bre90b] aborde l'audition sous deux aspects qui nous semblent particulièrement intéressants lorsqu'ils sont présentés conjointement : un aspect perceptif, un autre « écolo-

gique »^[a]. Le premier aspect tente de comprendre comment nos systèmes auditifs construisent une « image » du monde autour de nous à travers leur sensibilité au son. La question « écologique » aborde la façon dont notre environnement crée et forme le son autour de nous. Une approche complète de la perception doit relier ces deux aspects ; elle permet d'étendre le champ de la recherche en psychophysique au-delà des stimuli « simples » et de les replacer dans le contexte environnemental.

Comme l'a fait A. Bregman, nous soulèverons la question :

« Which physical property of the sound gives rise to the perceptual quality that we experience ? » [Bre90b].

Cette question permet de joindre deux approches qui semblent parfois inconciliables : une approche spécifiant les propriétés mécaniques des événements sonores et une approche de la psychologie expérimentale qui spécifie les paramètres acoustiques pertinents dans la perception et la reconnaissance de ces événements. C'est dans cette perspective que nous nous placerons lors de l'étude des propriétés audibles des sources sonores et de leur reconnaissance.

2.2 De la perception à la cognition

2.2.1 Introduction

L'interaction avec notre environnement passe par une saisie de la structure de celui-ci. Les organes de nos sens sont l'interface privilégiée avec le monde extérieur. Ils permettent le passage de l'extériorité à l'intériorité, à une expérience du sensible [Bar94].

La perception est un processus de transformation de l'information qu'acquière les organes sensoriels en des expériences subjectives des événements sonores [EK00]. La perception utilise les informations sensorielles (sensations) pour donner naissance aux représentations internes du monde qui nous entoure. Elle sert de point de départ à la connaissance de l'environnement dans lequel nous vivons et à des processus plus élaborés regroupés sous le terme de cognition.

Comme définition de la cognition, nous donnerons celle de Neisser [Nei76], qui la présente de manière synthétique et qui s'inscrit pleinement dans notre problématique de recherche :

« Cognition is the activity of knowing: the acquisition, organization, and use of knowledge » p. 1 [Nei76].

^[a] Cette appellation est mise entre guillemets car elle est généralement associée à l'approche de la théorie écologique de la psychologie prônée par J.J. Gibson.

Cette définition nous permet d'appréhender les problèmes que soulève l'élaboration de nos connaissances. Un premier aspect est lié au type de support de nos connaissances et l'organisation de celles-ci. Nous nous focaliserons avant tout sur leur organisation par le biais de l'étude de la perception des objets et de leur représentation en mémoire. La perception nous permet en effet d'acquérir des connaissances sur le monde et d'étudier les processus cognitifs qui permettent d'élaborer une organisation structurée des connaissances qui reflète notre environnement.

2.2.2 Vers une convergence

Généralement, il existe une forte divergence conceptuelle entre la perception et la cognition [Bar99]. Les systèmes perceptifs saisissent l'information dans l'environnement et la passent dans des systèmes séparés qui supportent différentes tâches cognitives telles que le langage, la mémoire ou la pensée. Cette dichotomie est souvent abordée comme une indépendance ou une modularité de ces systèmes dans le cerveau dont les fonctions et les spécificités sont propres à chacun.

La divergence la plus forte est liée au caractère amodal (qui n'est pas propre à une modalité sensorielle spécifique : audition, vision, etc.) des représentations servant de supports à ces fonctions cognitives, comparativement aux états perceptifs liés à leur modalité sensorielle propre. La notion de représentation qui est au cœur des débats reste toujours délicate à définir car elle véhicule des concepts très différents suivant la discipline – philosophie, psychologie, neuroscience Nous donnerons une définition proche du champ de la psychologie :

« La représentation mentale peut être vue comme une entité interne, le correspondant cognitif individuel des réalités externes expérimentées par un sujet » [HKK⁺98].

Malgré cet aspect polysémique de la représentation, il existe pourtant une forte convergence pour affirmer que le fonctionnement du système cognitif humain repose sur la mise en œuvre et la manipulation de représentations mentales.

Aujourd'hui, la dichotomie entre perception et cognition a tendance à s'estomper. Par exemple, Barsalou [Bar99] (ou [GB98] pour une introduction) propose une nouvelle théorie qui ancre la cognition dans les états perceptifs en rejetant le passage à des représentations internes amodales. Il faut remarquer que le débat reste de nos jours encore ouvert. Devant cette difficulté à proposer une approche unifiée de la perception et de la cognition, différentes

théories se sont développées. Nous les introduisons pour présenter notre cadre d'étude.

2.2.3 Quelle approche pour l'étude de la perception ?

Il est réducteur de présenter la perception sous l'aspect purement associationniste. En effet si nous réduisons la perception à l'addition de sensations, cette théorie ne permet pas d'expliquer les propriétés perceptives émergentes ou holistiques d'événements sonores ou d'objets visuels. Cette impasse a été abondamment illustrée par la théorie de la Gestalt [Gui79], l'exemple classique étant l'image représentant soit des visages soit un vase suivant notre état cognitif. Il convient donc de parler d'une théorie qui intègre nos connaissances sur le monde et nos croyances au processus de perception : pour Helmholtz [vH89] la perception était fondée sur des processus d'inférence inconsciente.

Dans une perspective cognitive, la perception a une fonction d'interprétation des données sensorielles et suppose une activité de traitement de l'information [HKK⁺98]. Cette fonction est conçue comme mobilisant deux formes de traitement : ascendant ou « dirigé par les données » (« bottom-up » en anglais) et descendant, dirigé par les concepts, connaissances ou représentations (« top-down » en anglais). Deux processus rentrent donc en jeu : un processus qui exploite les informations provenant de notre environnement et un autre qui intègre nos connaissances sur cet environnement. Nous voyons ici l'importance de resituer la perception dans le cadre de cette interaction.

2.2.4 Ecologie de la perception

Etudier la perception demande d'intégrer cette problématique dans le cadre d'une interaction « Homme-environnement », c'est-à-dire d'intégrer la dimension de validité écologique. Généralement, une des critiques apportées à la psychologie expérimentale est de ne pas toujours permettre une généralisation de ses résultats au « vrai » monde [EK00]. Dit autrement, cette approche est parfois loin de l'expérience que nous avons de tous les jours [Nei76]. Néanmoins, cette validité écologique ne doit pas avoir comme conséquence l'étude unique des phénomènes naturels dans toute leur complexité non contrôlée. Elle doit plutôt orienter les recherches sur la structure des situations naturelles qui dirigent les phénomènes comportementaux à l'intérieur et à l'extérieur du laboratoire [And91].

Cette critique peut se retrouver dans la réémergence de la théorie écologique initiée par J.J. Gibson [Gib66] que nous reprendrons ultérieurement, §2.6.2.1, p. 32. Cette théorie est dite écologique car elle tente de mettre l'accent sur la fonction primaire de la perception, c'est-à-dire de faciliter les

interactions entre les individus et l'environnement.

Pour illustrer ces perspectives nous reprendrons Neisser qui voyait déjà, dans les années soixante-dix, la nécessité de s'approcher d'une étude de la cognition plus ancrée dans la réalité de la vie quotidienne :

« ..., cognitive psychologists must make a greater effort to understand cognition as it occurs in the ordinary environment and in the context of natural purposeful activity. This would not mean an end to laboratory experiments, but a commitment to the study of variables that are ecologically important rather than those that are easily manageable. Second, it will be necessary to pay more attention to the details of the real world in which perceivers and thinkers live, and the fine structure of information which that world makes available to them. » p. 7 [Nei76].

Il nous semble important d'inscrire nos travaux dans ces perspectives « écologiques » car elles ouvrent une voie pour appréhender les processus perceptifs qui peuvent être impliqués dans une écoute de tous les jours.

2.3 L'Homme et son environnement sonore

2.3.1 Introduction

Lorsque nous sommes plongés dans un environnement sonore, différentes attitudes s'offrent à nous. En écoutant un morceau de musique, nous pouvons écouter la musique en tant que telle, mais il nous arrive de nous focaliser sur un instrument de musique et d'aller plus loin en tentant de reconnaître les caractéristiques de la source. Au contraire dans le métro, quand nous tendons l'oreille pour écouter la rame du train qui s'approche, il nous arrive de transformer ce rythme en une mélodie.

Par ces exemples simples, nous voyons que la notion d'écoute semble prendre différentes formes. Si nous mettons l'emphase sur le « message » véhiculé par le son, le son peut être décrit comme un indice qui renseigne sur les sources sonores et les événements situés dans notre environnement (écoute naturelle [Sch77]). Mais le son peut être aussi décrit comme un « message » lié à un code comme dans le cadre par exemple de la musique ou du langage (écoute culturelle [Sch77]).

Dans le cadre de ces travaux, nous abordons « l'écoute » de manière plus spécifique. Nous nous intéressons à la relation entre la perception des sons et la représentation des propriétés des sources sonores qui ont produit ces sons. C'est-à-dire qu'écoutons nous dans un son ? : nous focalisons nous sur ses propriétés acoustiques intrinsèques permettant de le décrire indépendamment de la source sonore ou au contraire sur les propriétés de la source sonore ?

2.3.1.1 Ecoute « musicale » et « quotidienne »

Etudions ces deux processus qui nous intéressent plus particulièrement, l'écoute musicale et l'écoute quotidienne [Gav93]. Nous pouvons écouter un son en termes d'attributs acoustiques propres (écoute musicale) ou selon l'événement qui l'a produit (écoute quotidienne). L'écoute quotidienne fait référence au processus de focalisation sur les propriétés de la source sonore contrairement à l'écoute musicale pour laquelle nous jugeons les qualités du son indépendamment de la source qui l'a produit. Nous décrivons donc un événement sonore sur deux niveaux :

- z Les attributs perceptifs de l'onde acoustique émise par la source sonore, nous permettant de décrire le son selon les propriétés telles que la hauteur, la sonie, la durée subjective, le timbre ...
- z Les propriétés audibles de la source nous permettant de décrire l'événement sonore du point de vue écologique : sa forme, son mode d'excitation, son matériau ...

Pierre Schaeffer a abordé cette problématique dans un cadre théorique pour la recherche musicale. Dans le « Traité des objets musicaux », [Sch77], il aborde la notion d'objet sonore, d'écoute et de typologie sonore. Bien que sa recherche s'insère dans le cadre d'une réflexion plus large sur la musique concrète, elle aborde cette question de l'écoute.

Dans le cadre de ses travaux, il a développé quatre secteurs d'écoute impliqués dans la perception :

- (1) Ecouter** ⊢ Nous prêtons l'oreille à quelqu'un ou à quelque chose. Par l'intermédiaire du son, nous visons la source, l'événement. Le son est traité comme indice de la source, de cet événement. Nous nous focalisons sur la causalité de la source.
- (2) Ouïr** ⊢ C'est le niveau le plus élémentaire de la perception. C'est une perception brute et passive, nous ne cherchons ni à écouter, ni à comprendre.
- (3) Entendre** ⊢ Nous manifestons une intention d'écouter, nous sélectionnons dans la perception brute ce qui nous intéresse afin d'opérer une qualification de ce que nous entendons. C'est « l'écoute réduite ».
- (4) Comprendre** ⊢ Nous saisissons un sens, des valeurs. Le son est traité comme un signe qui renvoie à un sens en fonction d'un langage, d'un code (p. ex. le langage ou la musique).

Ces secteurs correspondent à différentes écoutes qui sont naturellement présentes dans notre perception et qui se complètent. Il nous semble intéressant de montrer que nous retrouvons les deux notions abordées précédem-

ment, l'écoute « musicale » et « quotidienne » respectivement sous les termes d'« entendre » et d'« écouter ».

2.3.1.2 Une dichotomie ?

Existe-t-il deux traitements séparés ? Notre but n'est pas de poser une dichotomie entre ces deux processus. Nous pouvons voir en cela un passage, un continuum, de l'écoute des qualités intrinsèques aux sons, amenant parfois à un jugement « esthétique » par exemple lors de l'écoute d'une pièce musicale, à une reconnaissance des propriétés de la source sonore qui a produit cette expérience perceptive, voire la reconnaissance ou l'identification de celle-ci.

Nous spécifions cette dichotomie, dans le but d'orienter nos recherches vers la conceptualisation d'attributs audibles de la source sonore [Gav93] en lien avec les attributs de « surface » tels que le timbre, la hauteur et la sonie. Nous mettons ainsi l'emphase sur la source qui produit le son.

2.3.1.3 Qu'écoutons nous ?

Pour répondre à cette question, différentes études ont exploré la perception des événements sonores par une description de l'expérience auditive des auditeurs. Par exemple Gaver, [Gav88] [Gav93] par la description de l'écoute d'événements, a proposé une organisation hiérarchique des attributs de sources caractéristiques. Elle est organisée suivant le type de matériaux qui interagissent (produisant des sons d'objets vibrants, sons aérodynamiques ou des sons liquides) et suivant la façon dont ils sont mis en excitation.

Cette étude fait suite à celle de Van Derveer [VD79] cité dans [Gav88] montrant que les auditeurs ont une capacité à identifier et surtout à décrire un grand nombre de sources sonores enregistrées (voir aussi [BH87]). C'est uniquement dans le cas où les auditeurs ne reconnaissent pas la source qu'ils décrivent les sons sur des aspects purement sensoriels sans faire référence à la source. De plus les confusions semblent se porter sur les événements dont les profils temporels sont similaires. De même, Ballas & Howard [BH87] en utilisant divers sons courts, ont montré que la reconnaissance des sources sonores est liée au nombre d'événements différents pouvant être à la source de production du son. Une mesure d'entropie permet alors de quantifier l'ambiguïté causale d'un son, c'est-à-dire que la confiance dans l'identification d'un son diminue lorsque le nombre de causes, pour un son, augmente.

A. Faure [Fau00], dans une étude sur la verbalisation du timbre, a montré que dans plus de la moitié des cas les auditeurs décrivent non pas le son lui-même mais son mode de production – un instrument, une action, un matériau, un mécanisme – dans le cadre d'une verbalisation libre. De plus,

les gens ont une forte tendance à décrire la forme temporelle du son et son déroulement, qui semblent très importants dans la perception des événements sonores complexes.

Nous voyons ici une certaine faculté à décrire les propriétés des sources sonores. Mais ces études ne permettent pas de décrire avec précision les propriétés des sons qui sous-tendent la perception des événements sonores. Par contre l'utilisation de verbalisation dans le cadre de la perception d'événements sonores semble particulièrement intéressante afin de comprendre les propriétés acoustiques sous-jacentes au processus de perception, elles seront un guide précieux lors de l'analyse de nos résultats.

2.3.2 Niveaux de description d'un événement sonore

Les deux façons d'écouter que nous venons d'introduire montrent qu'il existe différents niveaux de description de ce qu'est un événement sonore. Dans le cadre de la vision, Shepard [She81] a spécifié la notion d'objet qui possède une double entité physique et perceptive, voire cognitive. Une entité physique dont les attributs peuvent être mesurables directement, une entité perceptive, voire cognitive qui a un aspect conceptuel qui découle de la catégorisation [Tij01]. Contrairement à certains auteurs comme Kubovy & Van Valkenburg [KVV01] qui parlent de la même façon d'« objet sonore », nous préférons parler d'événement sonore pour le défaire de sous-entendus visuels.

Comme le montre la figure 2.1, nous étudions l'événement sonore à différents niveaux :

Source sonore P Nous définissons l'événement au niveau de sa production sonore, le stimulus distal, c'est-à-dire au niveau du corps vibrant qui peut être décrit suivant sa forme, son matériau, la façon dont il est excité ...

Onde sonore P Lorsque le corps sonore est mis en vibration, l'énergie acoustique, le stimulus proximal, transporte aussi bien les caractéristiques de la source que celle du milieu dans lequel l'onde se propage. Cette information acoustique est saisie par nos récepteurs sensoriels.

Expérience interne P L'événement sonore est perçu suivant ses qualités perceptives ou bien il peut faire référence à une représentation interne de l'événement.

Naturellement, ces différents niveaux de description et leurs interactions sont étudiés par des disciplines spécifiques comme le montre la figure 2.1.

Par exemple, une barre de métal (la source sonore) qui est excitée en un point peut être étudiée d'un point de vue de la mécanique, afin de connaître ses propriétés physiques et vibratoires. Les propriétés de l'onde sonore émise

par la barre peuvent être décrites par l'acoustique. Par l'étude du signal, nous allons connaître ses propriétés telles que l'amplitude du signal, le contenu fréquentiel... La psychophysique, quant à elle, nous permet de relier les propriétés perceptives du son perçu (expérience interne) aux propriétés acoustiques de ce son. La psychomécanique essaie d'étudier quelles peuvent être les propriétés mécaniques de la source qui permettent de spécifier les propriétés perçues de l'événement sonore telles que sa forme, son matériau, son type d'excitation. La reconnaissance de l'événement sonore, son identification, sa relation perceptive et cognitive aux autres événements..., vont être étudiées par la psychologie cognitive en intégrant les résultats de la psychophysique et de la psychomécanique. Dans le cadre de cette thèse nous aborderons ces différents niveaux afin d'étudier les propriétés audibles de sources simples telles que des barres de métal qui vont nous permettre une reconnaissance de ces événements.

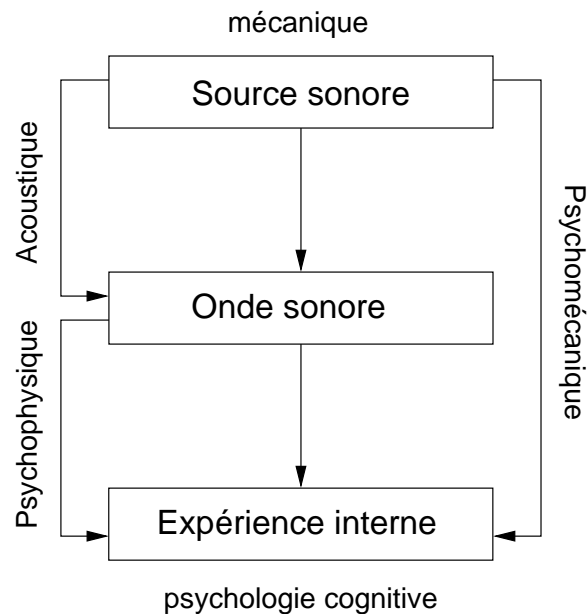


Figure 2.1 - Les différents niveaux d'analyse associés à la perception des événements sonores sont représentés ici. Le premier niveau est la source sonore qui représente le corps producteur de son qui peut être étudié du point de vue de la mécanique. Le second niveau est l'onde sonore correspondant à l'onde acoustique dans le milieu de propagation. Le troisième niveau fait référence à l'expérience interne, c'est-à-dire aux propriétés perceptives ou la représentation interne de notre événement sonore. Le lien entre la source sonore et l'onde sonore est étudié par l'acoustique. Le lien entre l'onde sonore et l'expérience interne de l'événement sonore peut être étudié par le biais de la psychophysique. Le lien entre la source sonore et l'expérience interne de l'événement sonore est étudié par la psychomécanique mais aussi par la psychologie cognitive notamment par l'étude de la catégorisation.

2.3.3 Etudes des événements sonores

Suivant les différentes propriétés des événements sonores que nous venons d'évoquer, propriétés intrinsèques aux sons ou propriétés audibles spécifiant la source sonore, différentes approches se sont développées.

Historiquement, l'approche de la psychophysique s'est focalisée sur les relations entre paramètres physiques et attributs perceptifs tels que la sonie, la hauteur ... Un développement plus récent a permis d'étudier les propriétés perceptives qui sous-tendent la perception des sources sonores par l'étude du timbre musical. Cette démarche associée à la théorie du traitement de l'information étudie séparément les propriétés des sons, indépendamment de la source. Elle permet de définir les propriétés perceptives potentielles engagées dans le processus de perception des sources sonores.

Une approche moins « traditionnelle » de la perception, la théorie écologique, s'est construite en se focalisant, entre autres, sur la structure des sources et sur la perception des invariants spécifiant la source [Gib66]. Cette théorie privilégie le complexe « humain-environnement » plutôt que l'étude de la perception centrée uniquement sur l'organisme percevant.

Il faut remarquer que cette approche écologique tend à influencer les recherches en sciences cognitives plus particulièrement dans des domaines tels que l'action située, l'ergonomie et dans le domaine qui nous intéresse : la perception des sources sonores.

Sous l'impulsion de Gaver [Gav88] [Gav93], le courant de l'« acoustique écologique » s'est développé dans la continuité de cette théorie. Cette approche écologique tente de décrire les attributs audibles de la source liés à ses caractéristiques physiques : sa géométrie, son matériau, son excitation... Ainsi la perception des sources sonores est étudiée non plus en termes de propriétés isolées mais selon les propriétés écologiquement valides. Le but de cette approche est de développer une théorie de l'écoute quotidienne telle que Gaver l'a proposée [Gav88].

Il faut remarquer que l'acoustique écologique actuelle s'oriente vers une approche intermédiaire faisant converger ces deux approches (psychophysique et approche écologique). De nouveaux axes de recherche émergent tels que la psychomécanique [Rou99] qui étudie la relation entre les propriétés mécaniques de la source sonore et la perception de celle-ci. Nous nous placerons dans ce cadre afin d'étudier la perception de la forme géométrique d'objets simples tels que des barres de métal.

Dans la suite du texte, nous commencerons par introduire l'approche psychophysique de l'étude du timbre, puis l'acoustique écologique, afin de montrer ce qu'elle apporte à l'étude des propriétés audibles des sources sonores. Ces propriétés audibles vont être une source potentielle d'informations

pour la reconnaissance des événements sonores que nous aborderons ensuite. La reconnaissance des événements sonores nous permettra de faire une comparaison critique de la théorie écologique (inspiratrice de l'acoustique écologique) et de l'approche théorique plus répandue, associée au « traitement de l'information ».

2.4 Approche psychophysique : le timbre

2.4.1 Introduction

L'étude du timbre s'est développée, afin de comprendre son rôle potentiel dans le processus de reconnaissance des sources sonores comme la voix, les instruments de musique et par extension les sons non-musicaux. La recherche sur le timbre s'est affinée depuis les années 1960, parallèlement aux perfectionnements des outils d'analyse et de synthèse du son et des outils statistiques.

A l'origine, le travail sur la perception du timbre a été principalement effectué sur des sons instrumentaux. Cette direction s'explique tout d'abord par la connaissance des sons harmoniques (voir la connaissance des Pythagoriciens sur l'harmonie et le travail de référence de Helmholtz sur la consonance). De par cette connaissance des sons harmoniques et de leur analyse acoustique, l'étude de la perception du timbre musical a permis d'obtenir un modèle perceptif simplifié, dans le but de l'étendre aux autres sources sonores.

2.4.2 Perception du timbre musical

Le timbre est en premier lieu un attribut qui a été décrit par la négative. Ainsi l'American National Standards Institute (ANSI) donne comme définition :

« Timbre is that attribute of auditory perception in terms of which a subject can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. » [ASA60]

Cette définition apporte peu d'informations sur les propriétés physiques sous-jacentes à la perception du timbre. C'est pourquoi elle est accompagnée de la note suivante :

« Timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus. »

Même si cet ajout tend à décrire le timbre suivant des paramètres acoustiques, mais qui restent encore à préciser, il convient de noter que le timbre n'est nullement décrit du point de vue perceptif. Les expériences que nous allons présenter ont étudié le timbre afin de spécifier ses attributs perceptifs et leurs corrélats acoustiques. Nous ne présenterons pas ici une revue historique sur le timbre (voir [Don97] et [McA94b]).

2.4.2.1 Etudes multidimensionnelles

■ **Méthodologie** Le caractère multidimensionnel du timbre ne permet pas de le mesurer sur des échelles simples comme « grave à aigu » ou « faible à fort ».

Pour ce type d'études, la proximité entre les sons est d'abord estimée par des jugements perceptifs de similarité/dissembance effectués par des auditeurs sur des paires de sons. Les analyses multidimensionnelles (voir §3.4.2, p. 85 pour une présentation de la méthode) sont pertinentes pour étudier les relations perceptives entre des stimuli et connaître les attributs sous-jacents.

Ces analyses statistiques tentent de placer les sons dans un espace psychologique multidimensionnel représentant au mieux la distance perceptive entre ceux-ci. Lors des analyses, des critères statistiques permettent d'estimer le nombre de dimensions appropriées. L'intérêt de ce paradigme est de ne poser aucun a priori, ni sur le nombre, ni sur la nature des dimensions perceptives qui sous-tendent les jugements perceptifs.

■ **Etudes de référence** Les études sur le timbre, telles que [Plo70] et [Gre77], ont montré que le timbre est un attribut perceptif multidimensionnel pouvant se décomposer sur deux ou trois dimensions.

J.M. Grey [Gre77] dans une étude de référence, a pu établir un espace de timbre tridimensionnel pour un ensemble de 16 sons synthétisés d'instruments à cordes et à vent. Sa démarche a été d'associer qualitativement ces dimensions perceptives à des corrélats acoustiques.

La première dimension est liée à la distribution de l'énergie dans le spectre. Elle peut être décrite comme « la brillance » du son, terme introduit par D. Wessel [Wes79]. La deuxième dimension est associée à une dimension spectro-temporelle que Krumhansl [Kru89] a appelée « flux spectral ». Il s'agit d'une combinaison du degré de synchronisme de l'attaque des différentes harmoniques et du degré de fluctuation de l'enveloppe spectrale pendant la durée du son. La troisième dimension semble inhérente à la présence relative de vibrations transitoires inharmoniques de fréquences élevées dans la partie initiale du son avant l'établissement de la partie harmonique. Cette dimension peut être associée au descripteur « qualité d'attaque » [McA94b].

Les études ultérieures tendent à retrouver les deux premières dimensions [Gre77] [KMW94] [MWD⁺95]. Par contre, le corrélat acoustique de la troisième dimension est moins bien défini et varie suivant les études. Cette troisième dimension est associée au « flux spectral » [MWD⁺95] ou aux irrégularités de la structure fine du spectre [KMW94]. Elle semble plus sensible au corpus de sons sélectionné, aux groupes de sujets.

Il faut ajouter à ces dimensions perceptives des traits propres à certains instruments appelés « spécificités », comme par exemple le rebond du sautereau d'un clavecin [MWD⁺95].

Ainsi, la figure 2.2 représente l'espace des timbres sur trois dimensions de l'étude de McAdams et coll. [MWD⁺95].

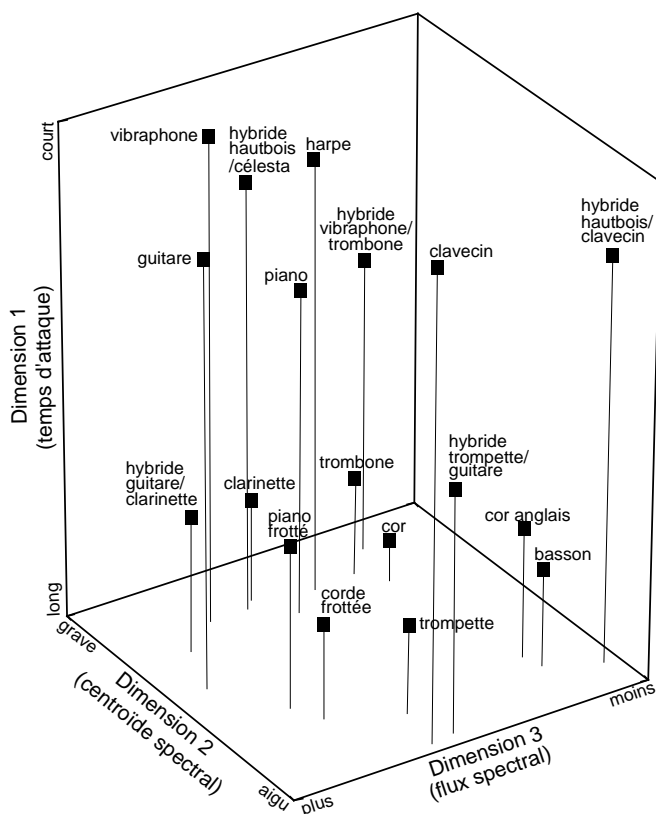


Figure 2.2 - Espace des timbres en trois dimensions [MWD⁺95] : le modèle spatial avec spécificités et cinq classes latentes de sujets est dérivé de jugements de dissemblance sur les 18 timbres par 84 sujets. Les corrélats acoustiques des dimensions perceptives sont indiqués entre parenthèses.

■ **Corrélat acoustique** À partir de l'espace perceptif tridimensionnel du timbre de C.L. Krumhansl [Kru89], [Kri93] et [KMW94] ont déterminé

à partir des trois dimensions perceptives telles que la brillance, la rapidité d'attaque et le flux spectral, des paramètres acoustiques corrélés à ces attributs. Ces travaux tentent de systématiser ces paramètres physiques afin de proposer des algorithmes de calcul sur les sons.

2.4.3 Limites de cette approche

Un des problèmes de cette approche est la recherche de paramètres physiques par simple corrélation avec les dimensions perceptives. Ainsi, il se peut qu'il existe un paramètre acoustique, non considéré, qui soit plus pertinent pour décrire une dimension perceptive.

Généralement, l'approche traditionnelle psychophysique reste une approche analytique. Il peut lui être reproché d'utiliser en majorité des stimuli artificiels : les sons sont généralement composés des primitives sonores simplifiées dont les paramètres sont contrôlés (amplitude, fréquence, amortissement, ...). Mais cette restriction rend possible un contrôle très fin des sons et donc de l'information acoustique disponible. Cette approche permet ainsi de tester des hypothèses sur les propriétés acoustiques sous-jacentes aux phénomènes perceptifs.

Par contre, elle ne permet pas de mettre en lumière ce qui peut orienter dans le son la perception des propriétés d'une source, l'instanciation d'une catégorie de sons. Elle considère uniquement le point de vue du percevant en faisant abstraction de la structure de la source sonore. C'est une limite lorsque nous considérons l'étude des propriétés audibles des sources sonores. C'est pourquoi nous présentons maintenant l'approche écologique qui met l'emphase sur la source.

2.5 Acoustique écologique

2.5.1 Introduction

L'acoustique écologique a pour but d'étudier les capacités de « perception de la source sonore » sur la base du son et d'identifier dans la structure acoustique ce qui permet ces capacités [CWTss]. Ces études se focalisent donc sur la perception des propriétés des sources sonores.

Par exemple, en nous focalisant sur les propriétés de la source plutôt que sur les propriétés du son, nous entendons un objet en verre de petite taille et creux tombant sur une surface dure plutôt qu'un son bref, fort et grave.

Le fait de juxtaposer les termes « acoustique » et « écologique » est né du désir de lier la description physique des sources sonores, à la perception

des propriétés pertinentes pour notre interaction avec l'environnement.

Il est de tradition d'associer les processus de connaissance des propriétés de l'environnement à la vision [CWTss]. La vision nous permet avec beaucoup de précision d'estimer, par exemple, la forme ou la taille d'un objet. Ainsi grâce, à la précision de la vision, nous aurions conscience des propriétés environnementales nous entourant. L'hypothèse avancée par ce courant théorique, est de penser qu'à travers le son nous avons cette même acuité.

Nous introduirons ultérieurement les aspects purement théoriques de la théorie écologique, voir §2.6.2.1, p. 32, dans le but de la comparer aux théories plus classiques de la perception et de dégager les limites de cette approche.

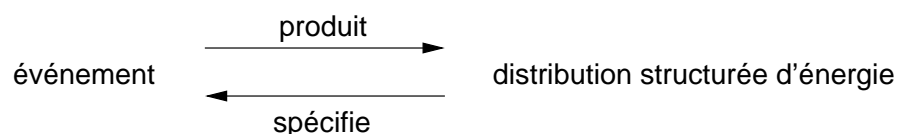
2.5.2 Spécification des événements

Dans l'optique d'une approche « écologique », il est primordial de caractériser la nature de l'événement produit et sa structure, afin de spécifier l'information par laquelle nous allons saisir les propriétés de cet événement. En effet, la viabilité de cette théorie dépend de la démonstration que les patterns d'énergie qui stimulent les sens contiennent une spécification de l'environnement [MC81].

Pour cela, nous devons étudier quelles sont les propriétés mécaniques et acoustiques des sources sonores permettant une spécification potentielle de leurs propriétés, telles que leur forme, leur taille, leur matériau ...

Parallèlement à cette spécification physique des événements sonores, nous devons montrer quelles sont les capacités des auditeurs à percevoir les propriétés des sources sonores.

■ **Patterns d'énergie** Une source sonore mise en vibration (stimulus distal), produit une onde sonore (stimulus proximal) dont le pattern d'énergie est structuré par cet événement. C'est ce pattern d'énergie qui atteint notre système auditif. Ainsi, la perturbation mécanique de la source est aussi ce que nous percevons et ce qui spécifie la source comme le montre le schéma ci-dessous [CWTss].



Une objection à cela est de dire que tous les objets ou événements différents ne structurent pas l'information dans l'environnement de façon différenciée. Mais, comme le remarque Fowler [Fow91] le fait de percevoir le monde n'implique pas que nous percevions de manière différente toutes choses

physiquement différentes. Deux événements perçus comme différents doivent structurer le médium (l'air dans le cas de l'audition) différemment et ces différences doivent être détectées par le système perceptif approprié. Des auteurs comme Fowler et J.J. Gibson pensent que ces deux conditions sont presque toujours réunies pour les propriétés pertinentes des objets et des événements avec lesquels nous interagissons dans l'environnement.

L'exemple donné par Casey [Cas98] sur les bruiteurs de films (artistes « Foley » du nom du précurseur) reflète bien cette remarque. À partir d'un petit ensemble d'objets et d'outils, ces artistes peuvent bruiteur en direct des effets sonores, illustrant un film ou une pièce de radio. Ainsi, à partir de sources sonores différentes nous pouvons « tromper » l'auditeur en donnant l'information essentielle et nécessaire d'un son qui spécifie l'événement voulu. Ces artistes semblent ainsi avoir saisi les composantes essentielles des sons, permettant de recréer l'information pertinente spécifiant les propriétés des sources. Connaissant ces composantes essentielles des sons, ils peuvent simuler différentes sources sonores à l'aide d'objets simples.

Dans le film « Monty Python & The Holy Grail », dans de nombreuses scènes du film apparaît un cavalier. Lors de sa première apparition nous le voyons en un plan serré qui s'élargit pour le montrer sans cheval mimant sa chevauchée. Ce cavalier est accompagné d'un écuyer bruiteur les pas du « cheval » à l'aide de noix de coco frappées qui s'entrechoquent. Les Monty Python utilisent ces mêmes bruitages avec beaucoup d'humour en les rendant cette fois-ci visibles.

2.5.3 Propriétés physiques potentielles

Un objet physique mis en vibration peut être décrit par les attributs propres à l'objet tels que sa configuration, son matériau mais aussi par les interactions auxquelles il est soumis lors de son excitation. En réduisant les propriétés d'un objet vibrant à cette description, nous pouvons dégager les propriétés physiques potentielles engagées dans le processus de reconnaissance de la source. Bien sûr, cette description ne tient pas compte des sources telles que les liquides, les sons avec une structure temporelle complexe ou les sons aérodynamiques tels que Gaver [Gav88] [Gav93] a pu les décrire dans le cadre de cette démarche. Néanmoins, dans notre étude portant sur des objets vibrants tels que des barres de métal frappées, il nous semble opportun de spécifier ces objets.

Afin d'étudier les propriétés perceptives des sources sonores, il convient de connaître les différentes contributions temporelles et spectrales qui peuvent être engagées. Le tableau 2.1 introduit les différentes propriétés d'un objet vibrant liées à l'objet lui-même et à son excitation, ainsi que leurs effets

temporels et fréquentiels. Par exemple, en frappant un marteau sur une barre nous induisons un type d'interaction qui conduit à des changements sur le spectre et l'amplitude, suivant la nature du marteau et la force de frappe. La barre en bois, de part son matériau, produit un spectre dont l'énergie est principalement dans les fréquences graves et s'éteint rapidement, à l'opposé d'un son d'une barre en métal.

Cette approche nous permet de décrire la source sonore, voir §2.3.2, p. 13, ainsi que l'onde sonore, son corollaire acoustique. Dans le chapitre 4, nous présenterons des méthodes d'analyse permettant de dégager les propriétés mécaniques et acoustiques des objets mis en vibration.

Attributs de la source	Effets sur le son
Interaction	
type	amplitude, spectre
force	amplitude, bande passante du spectre
Matériau	
force de rappel	fréquences
densité	fréquences
amortissement interne	variation de l'amplitude et des fréquences
homogénéité	effets complexes sur l'amplitude et les fréquences
Configuration	
forme	fréquences, pattern spectral
taille	fréquences, bande passante
cavité résonnante	pattern spectral

Tableau 2.1 - Composantes de certaines sources sonores familières [Gav88]

2.5.3.1 Variabilité

Les relations que nous venons de présenter entre les attributs de la source et l'information acoustique concomitante montre un lien « idéal ». Généralement, cette information acoustique peut comporter une grande part de variabilité qui est intrinsèque à la nature changeante de l'environnement. A différents moments, un même événement sonore peut produire des ondes acoustiques différentes. L'environnement impose des transformations supplémentaires sur l'onde physique en plus des variations dues à la source [Han89]. Malgré cette variabilité, l'acoustique écologique part de l'hypothèse qu'il existe de l'information invariante qui permet de spécifier la source sonore.

2.5.3.2 Les invariants

Comme nous venons de le voir, nous devons montrer que les patterns d'énergie qui stimulent les sens contiennent une spécification de l'environnement sonore. Cette spécification sera décrite en termes d'invariant. Les invariants peuvent être définis comme des patterns de stimulation dans le temps et l'espace qui restent inchangés à travers certaines transformations [MC81].

Nous allons détailler ici deux types d'invariants qui semblent particulièrement importants dans le cadre de l'étude des propriétés des sources sonores : les invariants structuraux et les invariants transformationnels. Ces invariants doivent être mis en relation avec le phénomène de constance perceptive que nous introduirons dans la partie sur « la reconnaissance des événements sonores » §2.6, p. 31. Ce phénomène décrit le fait que la perception de la source sonore peut être invariante (par exemple reconnaissance du violon), même si le stimulus proximal ne l'est pas [Han89]. Dans le cadre de la théorie écologique, les invariants sont à chercher au niveau du stimulus proximal, du pattern d'énergie spécifiant la source sonore et non pas dans des mécanismes correctifs de l'organisme percevant [MC81].

Invariants structuraux \triangleright Ils décrivent les aspects de la structure sonore qui sont « partagés par des objets qui, à un niveau de description particulier, peuvent être considérés comme identiques » p. 25 [MC81]. Ce sont des propriétés structurelles du stimulus proximal qui restent les mêmes, bien que d'autres propriétés changent.

Invariants transformationnels \triangleright Ces invariants vont décrire les changements qui agissent sur le stimulus distal. C'est le style de changement au niveau du stimulus proximal qui spécifie les changements dans ou du stimulus distal.

Un violon sera toujours identifiable comme tel, qu'il joue dans le registre grave ou aigu, que ses cordes soient pincées ou frottées avec l'archet et qu'il soit l'oeuvre de Stradivarius ou un instrument fabriqué en usine. Ainsi les invariants structuraux vont spécifier la nature de l'événement sonore.

Par exemple, un invariant transformationnel peut être un pincement de corde (c'est-à-dire la façon dont la corde est excitée), reconnaissable qu'il s'agisse d'une corde de guitare, de violon, de clavecin ou de basse. L'effet Doppler peut spécifier la dynamique de la source produisant du son car lorsque la source est stationnaire, la fréquence est constante, quand la source sonore s'éloigne, sa fréquence diminue. C'est le pattern de changement en amplitude et en fréquence qui devient invariant. Nous retrouvons des phénomènes similaires en vision avec les patterns de quelqu'un qui marche. Ce phénomène est

utilisé par le « motion capture » au cinéma pour augmenter le réalisme des mouvements de personnages de synthèse.

Les invariants structuraux sont donc les propriétés qui indiquent la nature de la source ou du groupe de sources sonores participant à un événement, tandis que les invariants transformationnels sont les configurations de changement du stimulus qui indiquent la nature de la modification subie par la source sonore [McA94b].

Les expériences que nous allons présenter dans la section suivante tentent de mettre en lumière ces invariants qui spécifient les propriétés des sources et permettent ainsi leur reconnaissance perceptive.

2.5.4 Propriétés audibles des sources sonores

Nous allons présenter dans cette partie une synthèse de différentes études qui contribuent à la compréhension de la perception des propriétés des sources sonores.

Dès 1636, Mersenne [Mer75] fut un précurseur dans ce domaine. Dans son livre « Harmonie universelle »^[a] sont introduits les prémices de l'acoustique écologique. Mersenne compare les sons et la lumière afin de comprendre « les actions dans lesquelles l'oreille est plus subtile que l'oeil, & conséquemment toutes les rencontres où le Son est plus subtil que la lumière ; il faudra faire un dénombrement de tout ce que l'on peut ouïr & apprendre en ténèbres & de nuit, & de ce qui peut entrer dans l'esprit par le moyen de la seule oreille »^[b] p. 19 [Mer75].

En effet dans son livre il aborde le thème « De la nature & des proprietz du Son » où il indique que « Le Son repreſente la grandeur & les autres qualitez des corps par leſquels il eſt produit » p. 19 [Mer75], et sa proposition XII qui tente de « Determiner la difference des ſons que font les differens metaux, & les differentes especes de bois, & de pierres de mſme grandeur » p. 180.

Nous voyons en lui un précurseur de ces études qui mettent en relation les propriétés des sources sonores telles que le matériau et la géométrie et leurs corollaires acoustiques. Par cette approche, il propose des propriétés acoustiques potentielles engagées dans la perception des propriétés des sources sonores.

Nous présentons par la suite différentes expériences menées qui sont classées par thématiques, en introduisant tout d'abord des études portant sur l'identification de sources et la perception des interactions. Ensuite nous nous

^[a]Merci à Alain de Cheveigné pour sa contribution bibliographique sur Mersenne.

^[b]Ceci est une traduction personnelle de l'ancien français au français contemporain.

focaliserons sur des études spécifiques consacrées à la perception par l'auditeur des matériaux et enfin à celle de la géométrie, rentrant dans le cadre de l'étude de la « précision métrique » de l'audition. Volontairement, les études liées au jugement de la qualité sonore des événements – le gêne ou l'agrément – ne sont pas abordées, car elles ne rentrent pas directement dans la compréhension des propriétés des sons qui véhiculent l'information décrivant les aspects des sources.

2.5.4.1 Identification de sources

L'étude pionnière dans ce domaine est celle de Warren & Verbrugge [WV84] sur l'habileté des auditeurs à distinguer des événements sonores de « casses » et de « rebonds » de récipients en verre. Dans un premier temps, les sons réels ont été très bien identifiés comme des événements sonores de « casses » ou de « rebonds ». Afin d'étudier la perception de la structure temporelle de ces événements que ces auteurs pensent impliquée dans l'identification, ils ont synthétisé ces deux événements en générant un train quasi-périodique d'impulsions acoustiques amorties provenant de l'enregistrement de quatre bouts de verres rebondissant. L'événement de « rebonds » a été réalisé en rendant synchrone les quatre impulsions acoustiques différentes, contrairement à l'événement de « casses » dont les impulsions acoustiques sont désynchronisées. Le bruit original de casse a été ajouté en début du signal. Les résultats montrent une très bonne performance à identifier ces deux types d'événements, même si le bruit de casse est enlevé dans l'événement de « casses ». Cette étude est très intéressante par sa démarche de combiner sons réels et sons de synthèse simplifiant l'événement. Elle met ainsi en lumière des invariants transformationnels de type temporel qui semblent pertinents dans la différenciation perceptive de ces événements.

B.H. Repp, quant à lui, met en parallèle les applaudissements et la parole [Rep87]. Pour lui les applaudissements sont un instrument de communication et reflètent comme la parole, la configuration d'excitation de la source. Dans cette étude, il est demandé à un groupe d'auditeurs de reconnaître les applaudissements enregistrés des membres de ce même groupe qui travaillent ensemble. Les résultats montrent que les participants n'arrivent pas à reconnaître la personne qui applaudit même si leur propre reconnaissance est meilleure. Lorsque les participants identifient le genre de la personne qui applaudit, ils se trompent systématiquement dans le sens de stéréotypes (des applaudissements aigus, rapides et d'amplitude faible sont associés à une femme). L'auteur a ensuite demandé de reconnaître différentes configurations de mains. Ces résultats sont très bons lorsque les sons proviennent de la même personne mais se dégradent lorsque des configurations de mains

de personnes différentes sont mélangées. Les analyses acoustiques permettent d'expliquer les différences de configurations, mais n'ont pas été mises en relation avec les phénomènes physiques. Ces résultats montrent qu'il n'y a pas d'extraction d'invariants propres à la configuration des mains.

Li et coll. [LLP91] se sont penchés sur la reconnaissance du genre de personne marchant sur une surface dure. Les pas réguliers de 16 personnes portant le même style de chaussures ont été enregistrés. Une première étude montre une bonne reconnaissance des genres des marcheurs. Les analyses spectrales sur les sons, montrent que les jugements d'identification du genre masculin sont caractérisés notamment par une énergie plus importante dans les basses fréquences. Les jugements d'identification du genre féminin sont liés aux sons dont l'énergie est plus importante dans les fréquences élevées. Une dernière série d'expérience de reconnaissance des genres a confirmé la pertinence de ces paramètres par modification des sons originaux. Comme pour l'expérience précédente, les résultats montrent une tendance forte à l'utilisation de stéréotypes et d'idiosyncrasies de la part des auditeurs. Cette tendance rend difficile l'étude des invariants pour les propriétés mécaniques humaines. C'est pourquoi une grande partie des travaux sur l'acoustique écologique s'est focalisée sur l'étude des propriétés des événements mécaniques [CWTss].

2.5.4.2 Perception des interactions

Freed [Fre90] a étudié la perception de la dureté relative de six maillets différents frappant des casseroles de quatre tailles différentes. La dureté physique des maillets varie du métal, au caoutchouc recouvert de feutre. Les casseroles ont été frappées à l'aide de chacun de ces maillets en contrôlant la vitesse de l'impact. Les participants de l'expérience ont jugé la dureté perçue sur une échelle unidimensionnelle. Parallèlement à ces expériences perceptives, Freed a proposé des paramètres acoustiques de prévision du timbre, basés sur une représentation auditive périphérique du signal acoustique. Bien que ces paramètres varient en fonction de la dureté physique du maillet ainsi que de la taille de la casserole, les jugements perceptifs effectués se sont avérés indépendants de la taille des casseroles. Ces résultats semblent montrer que les auditeurs extraient une information propre au maillet, même si d'un point de vue physique il est difficile de séparer la contribution de la casserole, de celle du maillet. Nous voyons par cette étude que nous devons affiner nos outils d'analyse, afin de caractériser ces invariants propres au stimulus proximal.

2.5.4.3 Perception des matériaux

Il existe une littérature plus abondante dans l'étude de la perception des matériaux. Dans certaines études, la perception du matériau est couplée avec la perception des dimensions géométriques des objets. Nous présentons ces dernières études dans la section suivante.

Lutfi & Oh [LO97] ont étudié la discrimination de changements de matériau de barres frappées. Ces travaux utilisent des sons de synthèse construits sur la base de principes simples qui décrivent la vibration de barres encastées. Les auteurs ont fait varier deux paramètres physiques (densité, module d'Young) pour faire varier la composition de matériaux de verre et de métal. Les sons résultants sont composés de trois sinusoïdes amorties variant en intensité, fréquence et amortissement. Il est demandé à l'auditeur de juger si le son provient d'une barre en métal ou en verre. La bonne réponse est donnée à chaque fois. Les résultats montrent une mauvaise discrimination des matériaux et une forte corrélation avec les variations en fréquence pour le métal et pour le verre. Pour le verre, un poids est donné aussi à l'amortissement. Pour les auteurs, la discrimination des matériaux est pauvre du fait du poids perceptif important donné aux fréquences des sons. Une critique majeure est la construction des sons basée uniquement sur un petit nombre de composantes fréquentielles qui poussent l'auditeur à se focaliser sur la fréquence fondamentale.

Klatzky et coll. [KPK00] ont eux aussi utilisé les modèles physiques afin d'étudier la reconnaissance de matériaux de barres. Ils se sont focalisés sur l'élasticité, paramètre important de la spécification physique des matériaux. Leur modèle est celui de Wildes & Warren [WW88] qui décrit les différents paramètres qui varient d'un matériau à l'autre et plus particulièrement la modélisation de l'amortissement. De la même façon que Lutfi et coll., les sons sont construits sur la base de sinusoïdes exponentiellement amorties mais de façon plus complexe. Deux tâches perceptives ont été menées avec des auditeurs, une première tâche de jugement de similarité perçue entre les différents matériaux et une tâche d'identification à choix forcé du matériau (caoutchouc, verre, bois et métal). Les résultats montrent que contrairement à l'expérience de Lutfi et coll. [LO97] les auditeurs utilisent les fréquences ainsi que l'amortissement temporel avec une prédominance pour ce dernier. Ces résultats peuvent être expliqués par la variance plus importante de l'amortissement temporel du fait des différents matériaux. Cette variance plus importante, tend à rendre plus saillant ce paramètre dans cette expérience, comparativement à celle de Lutfi et coll. [LO97].

Roussarie [Rou99] a étudié la relation, entre les variations de paramètres mécaniques telles que la densité, l'élasticité et les pertes énergétiques de

barres et de plaques, et la perception de ces événements. L'utilisation de modèles informatiques [Dou96] [Lam97] a permis un contrôle très fin de ces paramètres. Dans une première série d'expériences, il est demandé aux auditeurs de juger la dissemblance perceptive entre des sons de barres à section constante mais dont la densité et un facteur d'amortissement sont variables. Les dimensions perceptives qui sous-tendent ces jugements perceptifs sont très proches des dimensions mécaniques et leurs corollaires acoustiques. Des résultats similaires se retrouvent avec les plaques. Ces résultats permettent de formaliser les paramètres mécaniques qui sous-tendent la perception des matériaux.

2.5.4.4 Perception de la géométrie

Gaver [Gav88], comme nous l'avons déjà mentionné, a amorcé le développement d'un certain nombre d'études sur la perception des sources sonores. Gaver [Gav88] a étudié l'habileté des auditeurs à juger le matériau et la longueur de barres. Dans une première expérience, les auditeurs ont très bien identifié le matériau de sons de barre (de bois et de métal enregistrés) indépendamment des variations de longueur. Dans une deuxième expérience, les auditeurs ont jugé la longueur relative des barres, indépendamment des variations du matériau, à partir de ces mêmes enregistrements. Les résultats individuels montrent que la plupart des auditeurs n'ont pas été performants dans le jugement de la longueur relative et qu'il diffère du jugement simple de hauteur du son. Il existe une interaction entre la perception du matériau et la longueur. Par exemple, les plus petites barres ont été jugées plus petites dans le cas du métal par rapport au bois. Pour les barres les plus longues, le résultat est inverse. L'interprétation proposée par Gaver est que les sons d'une barre de métal et de bois pour une même longueur, sont plus dissemblables que des sons produits par des barres courtes et longues de même matériau. Le même type d'expérience a été effectué sur la base de sons de synthèse, même s'ils sont proches des expériences avec sons réels, il n'a pas été tenté de dégager des paramètres acoustiques permettant de prédire ces résultats.

Carello et coll. [CAKP98] ont étudié plus spécifiquement le jugement de longueur de baguettes de bois par le son. Des baguettes de différentes longueurs tombaient sur une surface rigide. Les participants ne voyaient pas les baguettes tomber et jugeaient la longueur en réglant la position d'une surface visible devant eux. Ce même paradigme a été utilisé pour un corpus de baguettes longues et un autre de baguettes courtes. Les résultats montrent que les participants arrivent à juger la longueur relative des baguettes. La longueur perçue des baguettes a toujours été sous-estimée dans le cas du premier corpus de baguettes et surestimées dans le cas des baguettes plus petites. Des

études acoustiques n'ont pas permis de mettre en évidence des paramètres acoustiques permettant d'expliquer les deux résultats. La longueur perçue, des baguettes les plus longues, semble liée à la durée du son, alors que l'amplitude des sons semble être un meilleur candidat pour les barres courtes. Aucun paramètre d'évolution temporelle pouvant décrire les rebonds de la baguette n'a été proposé qui pourrait être une information potentielle sur leur longueur. Les auteurs ont proposé un critère mécanique, lié au tenseur d'inertie, pour expliquer les résultats mais il n'a été proposé aucun lien entre ce paramètre mécanique et son résultat acoustique.

Une série d'études menées par Kunkler-Peck & Turvey [KPT00], est basée sur la reconnaissance des formes géométriques de plaques fines de différentes formes et matériaux, suspendues derrière un rideau et frappées à l'aide d'un marteau pendulaire. Dans une première série d'expériences (1 et 2), les sujets ont jugé la hauteur et la largeur de plaques de même surface en métal. Pour la deuxième expérience, le matériau a varié (plexiglas et bois, en plus du métal) en gardant les mêmes variations géométriques que l'expérience précédente. Les résultats indiquent que les participants sont capables de juger perceptivement la longueur et la largeur relatives des plaques de manière indépendante. Les jugements des dimensions sont du même ordre de grandeur que les dimensions réelles et les rangs sont préservés. Même en faisant varier le matériau, les auditeurs ont montré une bonne acuité à estimer les dimensions des plaques même si les dimensions perçues diminuaient suivant le matériau (métal \neq bois \neq plexiglas). Deux explications peuvent être données : la position de départ du pendule est la même quelque soit les matériaux et le type de matériau induit une diminution en fréquence de la même façon. Les fréquences de vibration suivant la longueur et la largeur des plaques semblent être de bons prédicteurs des résultats. Dans une troisième expérience, les auditeurs ont identifié la forme des plaques (rectangle, triangle et cercle) au dessus du niveau du hasard. Dans une quatrième expérience, les mêmes formes sont déclinées en trois matériaux comme précédemment (métal, bois, plexiglas). Les participants jugeaient la forme et le matériau. Les résultats montrent que les participants discriminent les formes à un niveau bien au-dessus du hasard et qu'ils discriminent parfaitement les différents matériaux. De plus, ils ont observé des interactions entre matériau et forme : les sujets avaient tendance à associer un matériau particulier avec une forme particulière : bois et cercle, métal et triangle, plexiglas et rectangle. Par contre les auteurs, n'ont pas relié ces résultats à des paramètres acoustiques et/ou mécaniques.

Les travaux de Lakatos et coll. [LMC97] nous intéressent tout particulièrement puisqu'ils ont été à la base de nos travaux. Dans cette étude, les auteurs ont cherché à évaluer l'habileté des auditeurs à discriminer la forme

géométrique d'objets résonnants sur la base de leurs attributs auditifs. Les objets étudiés sont des barres de métal et de bois frappées dont seule la section varie. Pour chaque matériau (bois et métal), les expériences impliquent une tâche d'appariement intermodal entre les sons des barres enregistrées et la représentation visuelle de leurs sections. Les analyses multidimensionnelles (voir §3.4.2) pour l'expérience sur les barres de métal, révèlent un espace à deux dimensions perceptives et pour les barres de bois, un espace à une dimension. Pour les barres de métal, la première dimension est corrélée avec le rapport épaisseur/largeur et la deuxième dimension avec une mesure du centre de gravité spectral. Pour les barres en bois, la première dimension est elle aussi le rapport épaisseur/largeur mais avec une corrélation plus faible. Pour les barres de métal, il existe une séparation perceptive nette, entre les blocs et les plaques, basée sur les deux dimensions. Les résultats de ces expériences ont été associés à des analyses acoustiques ainsi que des mesures d'accélérométrie afin de caractériser les modes de vibration des barres. Ces analyses montrent que la présence des modes de vibration de flexion suivant la largeur et l'épaisseur et que les modes de torsion sont proches des valeurs théoriques. Dans le cas du bois, ces modes sont parfois absents ou très amortis. Ainsi les fréquences liées aux modes de flexion suivant l'épaisseur et la largeur, et les fréquences de torsion donnent chacune une information sur le rapport épaisseur/largeur et peuvent être des informations potentielles utilisées par les auditeurs. Mais les auteurs n'ont pas proposé une validation de ces invariants potentiels. Les résultats de cette expérience seront repris ultérieurement.

2.5.4.5 Synthèse sur ces études

Les études que nous venons de présenter montrent un panel de démarches intéressantes pour l'étude des propriétés audibles des sources sonores. Ces études permettent de pointer les limites de certaines approches. Par exemple, les expériences basées sur la perception d'action d'effecteurs humains (la marche, les applaudissements) semblent nettement parasitées par les stéréotypes et les idiosyncrasies des personnes, rendant difficile la caractérisation de ces sources [CWTss].

L'étude des phénomènes mécaniques semble un meilleur candidat pour l'étude des invariants structuraux et transformationnels des événements sonores. Dans ce cadre, l'utilisation de la synthèse sonore apporte parfois des résultats contradictoires lorsque les sons produits s'éloignent des sons naturels et tendent à mettre en valeur artificiellement certains paramètres limitant la portée de ces résultats.

Deux voies semblent prometteuses. Une première approche est l'utilisa-

tion de modèles physiques de synthèse sonore permettant un contrôle fin des paramètres mécaniques. Cette démarche permet de faire le lien entre, variation des paramètres mécaniques et des propriétés acoustiques concomitantes et la perception des événements sonores. Une autre approche est basée sur les sons réels et l'étude de la perception de ces événements sonores. Dans ce cas, une attention particulière est donnée à l'interprétation des résultats par des analyses acoustiques et mécaniques. Ces analyses permettent de proposer des sons « schématiques » dont l'information potentielle a été contrôlée afin de tester des hypothèses dans le même cadre expérimental que l'expérience sur les sons réels.

En abordant le thème de l'acoustique écologique, nous avons étudié les liens entre les propriétés de la source et la façon dont elles pouvaient être perçues, ce que nous avons appelé : « propriétés audibles de la source ». Nous voyons dans ces études, que généralement, l'idée implicite est de montrer ce qui peut nous permettre la reconnaissance de ces événements sonores. Nous allons maintenant introduire la reconnaissance des sources sonores afin de montrer qu'elles peuvent être les limites de cette approche si nous restons uniquement dans le cadre d'une théorie écologique radicale. Pour cela, nous introduirons une approche plus classique qui nous semble pertinente pour la reconnaissance des sources sonores.

2.6 La reconnaissance des sources sonores

2.6.1 Introduction

En interagissant dans notre environnement, nous sommes confrontés à une multitude d'informations (visuelles, auditives ...). Nos systèmes sensoriels sont ainsi bombardés d'informations auxquelles nous devons donner sens. Donner sens à cette information implique des processus perceptifs de reconnaissance et d'identification des sources sonores nous entourant.

Au vu de cette multitude d'informations, nous devons décider quelles parties de l'information sensorielle proviennent d'un même événement par exemple sonore. Ainsi comme le remarque A. Bregman :

« L'analyse des scènes auditives est le processus réunissant en une unité perceptive l'ensemble des signaux provenant, dans une période de temps donnée, d'une seule source sonore de l'environnement » p. 12 [Bre94]

Cette étape est essentielle car la reconnaissance des sources et des événements sonores ne peut avoir lieu qu'après avoir séparé les informations perceptives propres à chaque source.

La reconnaissance des sources sonores fait référence à la mémoire des événements, à leur expérience, c'est-à-dire à la reconnaissance d'une source sonore par l'appariement de sa forme physique, acoustique, à sa représentation en mémoire [Bou96] et à sa représentation interne. La reconnaissance des sources sonores est aussi impliquée lors de la comparaison d'événements sonores. La reconnaissance et l'identification des sources sonores sont des activités importantes de notre vie quotidienne. Elles constituent des processus centraux dans l'élaboration de nos connaissances.

Nous allons présenter tout d'abord deux théories qui divergent sur ce que représente le processus de perception et qui nous intéressent particulièrement dans le cadre de la reconnaissance des sources sonores. Nous introduirons ensuite les différentes étapes qui vont rendre compte de la reconnaissance, en exposant deux modèles qui rendent compte de ces divergences et qui font le lien avec l'acoustique écologique.

2.6.2 Deux approches de la perception

Lorsque nous avons introduit la perception et la cognition, nous avons fait mention de l'aspect écologique de la perception, c'est-à-dire le rôle de l'interaction « humain-environnement » dans la perception, voir §2.2.4, p. 9. Deux approches théoriques divergent sur le rôle de cette interaction et son implication dans la perception et la reconnaissance des sources sonores : la perception directe et la perception indirecte.

« If percepts are constructed, why are they usually accurate? Surely perceiving is not just a lucky way of having mental images! The answer must lie in the kind and quality of the optical information available to the perceiver. The information must be specific enough in most cases to ensure that the constructed percept is true to the real object. But if this is admitted, the notion of “construction” seems almost superfluous. » p. 18 [Nei76]

2.6.2.1 La perception directe

La perception directe est associée à la théorie écologique développée par Gibson [Gib66] [Gib79] et reprise par d'autres auteurs [MC81]. Elle postule qu'il existe une spécification précise des objets, des sources sonores, des lieux et des événements dans la stimulation sensorielle et qui sont disponibles pour l'organisme. Cette approche est appelée « théorie écologique » car elle met l'accent sur le système « organisme et environnement ». Elle cherche à savoir ce que les organismes doivent connaître de leur environnement et comment il peut être connu. L'organisme évoluant dans un environnement riche en

2.6. LA RECONNAISSANCE DES SOURCES SONORES

information, son système perceptif devient alors façonné, « accordé », à l'information disponible. Les organismes percevant saisissent alors l'information invariante fournie par le monde environnant. Gibson [Gib66] [Gib79] abandonne la vue traditionnelle du stimulus, c'est-à-dire des variables physiques de bas niveaux. Il redéfinit la notion de stimulus avec une vue plus psychologique qui implique, non pas de l'énergie, mais de l'information. Il définit l'information comme la structure qui va spécifier l'environnement à un animal.

2.6.2.2 La perception indirecte

La perception indirecte fait référence aux théories du traitement de l'information [LN77]. Contrairement à la théorie écologique, cette approche considère que les sens reçoivent une description fréquemment inadéquate du monde. L'entrée sensorielle ne donne pas une information complète des objets, des sources sonores ou des événements. La perception implique donc des traitements internes qui permettent de transformer l'entrée sensorielle en une information élaborée. Ces processus internes nous permettent d'être conscient indirectement du monde externe, sous la forme de représentations internes.

La qualité perceptive du percept provient de la personne qui perçoit. A l'exception des Gestaltistes (lois innées d'organisation), la perception repose sur la mémoire. Il faut donc des processus psychologiques pour construire des arrangements spatiaux et de forme et reconstituer les moments discrets en des événements.

2.6.2.3 Comparaison des théories

Il existe des divergences fondamentales entre ces deux théories qui sont résumées dans le tableau 2.2.

Caractéristiques	Perception directe	Perception indirecte
Système	organisme-environnement	organisme
Perception	détection de l'information	perception médiatisée
Stimulation	très riche	entrée pauvre

Tableau 2.2 - Comparaison de la perception directe et indirecte [MC81]

■ **Stimulation** La première divergence est sur la stimulation. Elle est riche dans le cadre de la théorie écologique et spécifie la source. Dans le cadre de la perception indirecte, elle est pauvre et doit impliquer une reconstruction interne du sens. Cette différence implique que dans le cadre de la théorie

écologique, la notion de représentation interne est rejetée alors que pour la perception indirecte, la perception est médiatisée.

■ **Actif** Il apparaît deux approches fondamentalement différentes du terme « actif ». Pour la perception indirecte, « actif » est employé dans le sens d'une construction d'une signification contrairement à la perception directe où « actif » renvoie à une exploration, une saisie (« pick-up » en anglais) de l'information.

■ **Constance perceptive** La constance perceptive nomme les phénomènes pour lesquels les propriétés perceptives d'un événement sonore restent constantes même s'il existe des changements significatifs au niveau du stimulus proximal (image rétinienne, ondes sonores au niveau de l'oreille). Suivant ces deux théories, deux stratégies opposées sont développées.

Perception indirecte P La constance perceptive est créée par correction du système nerveux. Les invariants dans ce cas sont associés à des processus perceptifs.

Perception directe P La constance perceptive est associée aux propriétés de la stimulation qui sont invariantes. Nous avons déjà introduit cette notion d'invariant dans le cadre de l'acoustique écologique, voir §2.5.3.2, p. 23.

2.6.2.4 Notre cadre théorique

La perception directe a été vivement critiquée, voir l'article de Ullman [Ull80] intitulé « Against direct perception ». Une des critiques majeures est que les processus qui sont impliqués dans l'identification des invariants sont beaucoup plus complexes que ce qu'a pu décrire Gibson. De plus, le refus de la médiation des représentations internes, a été utilisé par les détracteurs de cette théorie pour démontrer l'impasse de cette approche.

Pourtant, cette approche nous paraît intéressante, car elle a permis de resituer la perception dans le cadre d'une interaction entre l'humain et son environnement. L'acoustique écologique s'est inspirée de cette théorie pour offrir une nouvelle approche qui tente de spécifier les invariants impliqués dans la perception des propriétés des sources sonores. Cette nouvelle démarche semble être l'écho de la remarque impertinente de Gibson [Gib66] :

« If the physicist can specify the type of mechanical disturbance and the psychologist can obtain corresponding judgments, this type of psychophysics should prove to be more interesting than the psychophysics of abstract loudness and pitch » p. 89 [Gib66].

La grande majorité des études sur l'acoustique écologique à l'exception des instigateurs tels que Carello et coll. [CWTss] adopte une approche hybride qui tente de spécifier les invariants dans le stimulus par l'acoustique et la mécanique. Ces études intègrent les théories de traitement de l'information afin de comprendre les processus sous-jacents à la perception des propriétés des sources sonores et à leur reconnaissance. Nous nous placerons dans ce cadre théorique.

Dans la suite du texte, nous présentons les différentes étapes de traitement perceptif impliquées dans la reconnaissance des sources sonores. Ce modèle de traitement appartient aux théories du traitement de l'information.

2.6.3 Etapes de la reconnaissance

Le schéma 2.3 présente différentes étapes perceptives susceptibles d'intervenir et d'interagir lors de la reconnaissance d'événements sonores. La reconnaissance se distingue de l'identification qui consiste à pouvoir nommer la source sonore.

Le processus sous-jacent à la reconnaissance d'une source sonore est de le « re-connaître », c'est-à-dire de lui trouver un représentant en mémoire à long terme qui va lui donner un statut d'événement connu. Par le recours à la mémoire à long terme, la reconnaissance est une phase d'activation des connaissances [Tij01], c'est une simulation interne de l'événement sonore.

La reconnaissance par la simulation et l'identification qui peut suivre, dépend du contexte de présentation de l'événement sonore, de l'organisation des connaissances en mémoire à long terme et de la tâche en cours.

Ce modèle s'écarte d'un modèle purement ascendant, hiérarchique car il interagit avec des processus descendants à différents niveaux du processus. Il existe une interpénétration entre ces deux types de processus. Les processus descendants correspondent à des attentes liées aux schémas mentaux activés par le contexte.

2.6.3.1 Etapes de traitement

Nous détaillons ici les différentes étapes de traitement.

■ **Transduction sensorielle** Il ne nous semble pas pertinent dans le cadre de cette thèse de revenir sur la transduction sensorielle, le lecteur pourra se référer à l'excellente introduction de B.J. Moore, [Moo97] ou [Bag99] et [McA94a] pour des introductions en langue française. C'est une représentation « brute » du signal acoustique dans le système nerveux auditif péri-

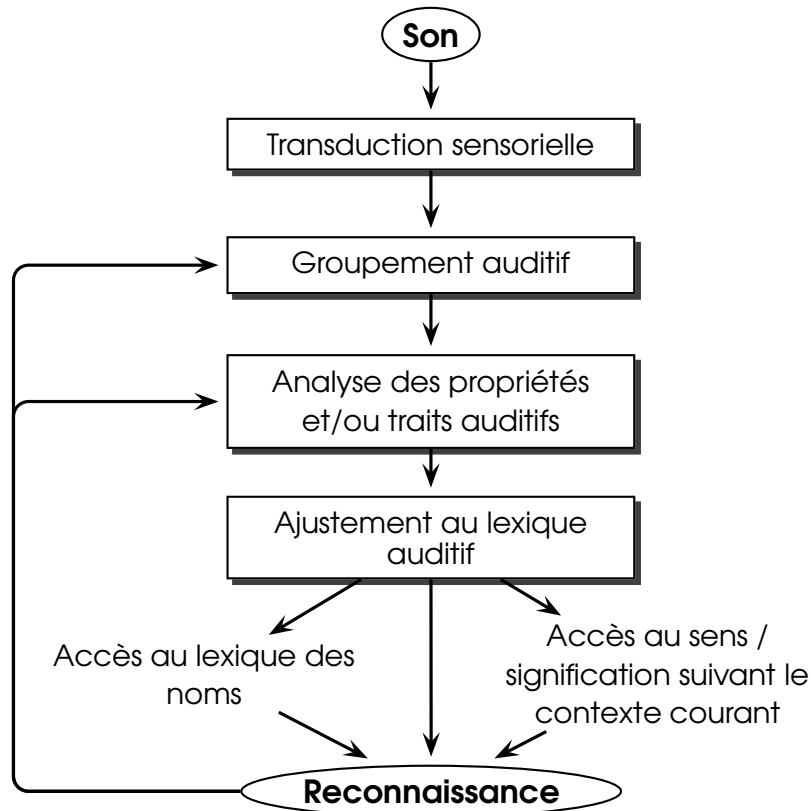


Figure 2.3 - Schéma des étapes de traitement qui interviennent dans la reconnaissance et l'identification des sources sonores. Cette disposition n'implique pas un traitement séquentiel de toutes les étapes ; de même certaines étapes regroupent plusieurs traitements. D'après [McA94b].

phérique [McA94a], c'est-à-dire une décomposition temporelle et fréquentielle de l'information nerveuse auditive.

■ **Groupements auditifs** Les processus auditifs de groupement permettent de séparer les contributions de plusieurs sources lorsque nous sommes plongés dans une scène sonore complexe. Les propriétés d'un événement sonore ne peuvent être analysées comme telles tant que ses éléments n'ont pas été intégrés en une unité perceptive et séparés de ceux des autres événements sonores. Nos connaissances sur le contexte de la scène sonore peuvent déjà intervenir dans la facilitation de ce traitement.

Ces processus font référence à ce que Bregman [Bre90b] a désigné sous un terme général : « analyse primaire de la scène auditive » qui regroupent des principes [Bre94] tels que :

« Il est extrêmement rare que des sons n'ayant aucun rapport entre eux

2.6. LA RECONNAISSANCE DES SOURCES SONORES

démarrant exactement au même moment » p. 19; « Les propriétés d'un son isolé tendent à se modifier de façon continue et lente. » p. 21; « Lorsqu'un corps vibre à une période répétée, ses vibrations donnent naissance à un pattern acoustique dont les fréquences des composants sont des multiples d'une même fréquence fondamentale. » p. 31 et « La plupart des modifications qui surviennent dans un signal acoustique affecteront tous les composants du son résultant, de manière identique et simultanée. » p. 33.

Ces différentes régularités, qui ne sont pas toujours robustes, peuvent être pertinentes lorsqu'il existe une convergence entre celles-ci, mettant ainsi en échec les ambiguïtés ou incongruités.

Dans le cadre d'expériences en laboratoire, les sources sont généralement présentées seules lorsque nous travaillons plus spécifiquement sur la reconnaissance des sources sonores. Dans ce cas, ces processus ne rentrent pas spécifiquement en jeu.

■ **Analyse des propriétés et/ou traits auditifs** Suite aux processus de groupement auditif, chaque événement peut être analysé suivant des attributs auditifs pertinents tels que : la sonie, la durée subjective, la hauteur, le timbre...

C'est à ce niveau que peuvent se refléter les invariants structuraux ou transformationnels, voir §2.5.3.2, p. 23. Les propriétés et les traits auditifs vont être déterminés par la nature de la structure résonnante qui entre en vibration (p. ex. tableau 2.1, p. 22). Nous voyons ici le rôle important des études sur l'acoustique écologique qui permettent de faire le lien entre ces invariants propres à l'onde sonore et la façon dont le système auditif les extrait.

■ **Lexique auditif** A partir de la description de chaque source par ses attributs auditifs, nous pouvons accéder au lexique auditif, c'est-à-dire aux catégories d'événements sonores qui vont nous permettre la reconnaissance, voire l'identification des sources sonores. L'accès à ce lexique auditif sera développé de manière détaillée dans la section « Les catégories », §2.7, p. 40. Cet accès peut amener directement à la reconnaissance de l'événement sonore, sans systématiquement passer par une étape intermédiaire d'accès au lexique des noms.

2.6.4 Discussion du modèle

Le modèle que nous venons de présenter s'inscrit dans la théorie du traitement de l'information. Pour remettre ce modèle dans une perspective éco-

logique, nous le mettons en parallèle avec le modèle « d'interprétation des sons » de Richards [Ric88]. Son approche est une transposition du modèle de Marr [Mar82], de reconnaissance visuelle des objets, à l'audition, en intégrant les spécificités propres à l'audition. Il propose, lui aussi, trois niveaux successifs de représentations internes : une esquisse primaire auditive, une esquisse $2_{1/2}$ -D, et une esquisse 3 -D. Même si l'utilisation de termes directement liés à la modalité visuelle n'est pas forcément appropriée, les trois niveaux sont à rapprocher du modèle que nous avons présenté.

Le premier niveau de représentation peut s'approcher de l'étape de transduction sensorielle et de l'étape de groupement auditif réunies. D'un point de vue computationnelle, elle est à mettre en lien avec « l'image auditive » du signal acoustique donnée par des modèles auditifs [PAG95] [Pat00]. Le deuxième niveau de représentation est intéressant car il suppose que nous ayons une représentation mentale des sources sonores sous la forme d'une décomposition perceptive de la source suivant quatre composantes majeures : 1) la source de puissance P ; 2) l'oscillateur O ; 3) le résonateur R ; 4) le coupleur C, appelé modèle P.O.R.C. Cette décomposition en modèle P.O.R.C est proposée comme un modèle général de compréhension du comportement acoustique des sources sonores. A partir d'une collection de propriétés perceptives qui capture la topologie basique de la source sonore (matériau, forme, excitation, ...), il faut créer une représentation de la source sonore, c'est le troisième niveau. Ce dernier niveau de représentation semble plus problématique pour l'auteur, de part sa terminologie et ce qu'elle représente. Richard [Ric88] pense qu'à partir de la décomposition perceptive en modèle P.O.R.C, nous pouvons déjà avoir une représentation typique de la source, par exemple, sous forme d'une organisation hiérarchique de ses propriétés. Il n'est donc pas utile de passer par ce dernier niveau de représentation.

Cette analyse peut sembler proche de la théorie écologique qui, dans son approche la plus radicale [CWTss], postule que le fait de saisir les propriétés audibles des sources implique leur reconnaissance sans passer par des étapes médiatisées et de comparaison avec des représentations déjà en mémoire.

« ..., the ecological approach asserts that input, once properly construed, is rich and lawful and specific to its source directly, without elaboration by internal mechanisms. » p.2 [CWTss].

De notre point de vue, un événement sonore n'a pas de statut en soi s'il n'est pas associé ou comparé à des événements déjà rencontrés afin d'amener à sa reconnaissance ou non s'il semble inconnu. Ainsi une simple représentation des propriétés audibles de la source ne permet pas cela. De plus les situations auxquelles nous sommes confrontés ne se produisent jamais exactement

2.6. LA RECONNAISSANCE DES SOURCES SONORES

de la même façon. Un enregistrement simple de ces expériences perceptives ne serait d'aucune utilité pour aborder des problèmes présents ou anticiper le futur [Est94]. Même si les événements sonores sont uniques, occasionnels, les organismes ne les traitent pas de manière unique, Ils répondent à ces stimulations sur la base d'apprentissages passés, en les comparant à des stimulations ayant le même statut perceptif ou cognitif. Nous voyons ici que la mémoire joue un rôle essentiel dans ces processus, car son organisation fait que l'information acquise par les expériences passées peut être applicable aux situations présentes. L'essence de l'organisation de la mémoire est la « la catégorisation » [Est94], notion que nous abordons dans la suite de ce texte.

2.7 Catégories et événements sonores

2.7.1 Introduction

Comme nous venons de le voir, la tâche essentielle d'un système perceptif est de donner au cerveau les moyens de s'appropriier le monde physique, de dialoguer avec lui. Ce monde peut être vu comme un monde continu et un monde de l'infiniment variable [Sch90]. Il faut donc trouver un moyen de condenser l'information véhiculée par un événement sonore ou un objet afin de lui associer une forme interne dans laquelle la variabilité aura été fortement réduite. Cette variabilité peut être réduite par les processus de groupements auditifs et par l'analyse des propriétés et des traits auditifs, voir §2.6.3.1. Ces processus nous permettront de mettre en relation cette forme interne avec nos connaissances présentes du monde externe par la catégorisation.

2.7.2 Définition de la catégorisation

La « catégorisation » peut être décrite comme une conduite adaptative fondamentale par laquelle nous « découpons » le réel physique et social. Sa fonction cognitive est la création de catégories (d'objets, d'événements sonores, d'individus, etc.) nécessaires à la transition du continu au discret [HKK⁺98]. Les catégories correspondent à un type de représentation interne, à savoir qu'elles entretiennent une relation de dépendance avec les événements sonores ou les objets dans leur environnement [Meu01]. Ce passage, du continu au discret, nous permet d'organiser nos connaissances de manière productive, sous forme de catégories, de relations, où dans des structures plus complexes.

Nous sommes constamment à juger si un événement sonore est une instance d'une catégorie. Cette opération cognitive est importante d'un point de vue « écologique ». Elle implique que nous puissions connaître à tout moment si un objet perçu peut être pertinent dans un contexte donné : prise en compte d'un danger suivant le contexte d'apparition de l'objet ; dans un autre registre, la reconnaissance d'un objet pouvant être utilisé comme instrument de percussion.

Contrairement à la notion de « classification » (partition d'une collection d'événements sonores en groupes), la « catégorisation » implique en plus, que la connaissance de la catégorie à laquelle un événement sonore appartient, nous renseigne sur les propriétés de cet événement [Est94]. L'acte de catégorisation comporte donc deux aspects : un aspect « inductif », qui consiste à assigner un événement sonore à une catégorie à partir de certaines caractéristiques de celui-ci et un aspect « déductif », qui est l'utilisation de

l'appartenance d'un événement à une catégorie afin d'associer à celui-ci les caractéristiques de cette catégorie.

2.7.3 Catégorisation et mémoire

La catégorisation renseigne sur les traitements cognitifs (quels critères et quels types de traitement le système utilise pour catégoriser) et sur la manière dont les connaissances sont structurées en mémoire à long terme (MLT) [Tij01]. La plupart des modèles de la MLT représentent les connaissances sous forme de relations entre catégories. Les représentations internes des événements sonores sont intégrées dans des schémas complexes qui les lient par des relations au sein de la mémoire sémantique [EK00].

La mémoire à long terme contient un nombre important d'informations permanentes de nature différente, sous forme imagée (modale, liée à une modalité sensorielle) ou propositionnelle (amodale). Nous pouvons distinguer en MLT, la mémoire sémantique et la mémoire épisodique [Tul72], toutes deux associées à la mémoire déclarative. La mémoire procédurale, quant à elle, fait référence aux savoirs-faire.

Mémoire épisodique \triangleright Elle fait référence au stockage des événements spécifiques se déroulant dans un endroit particulier et à un temps donné.

Mémoire sémantique \triangleright Mémoire où sont emmagasinées les connaissances générales sous forme de concepts abstraits et généraux. C'est notamment le système par lequel l'individu stocke sa connaissance du monde.

Les recherches sur la mémoire se sont avant tout centrées sur la distinction entre ces deux types. Cette vision dichotomique tend à se nuancer maintenant. Les recherches s'orientent actuellement vers les similarités et les connexions entre celles-ci [EK00].

Nous nous focaliserons plus particulièrement sur le contenu de la mémoire sémantique car elle est directement associée à l'organisation catégorielle de nos connaissances sur les événements sonores.

2.7.4 But de la catégorisation

Un des buts premiers de la catégorisation est de nous permettre de traiter de manière productive et informative une grande quantité d'informations.

Un nombre limité de catégories ne donnera accès qu'à une connaissance très générale, pauvre d'un point de vue informatif et ne nous permettant pas d'interagir au quotidien. Au contraire, plus il existera un grand nombre de catégories spécifiques, plus les membres de ces catégories seront homogènes et partageront des propriétés communes. Une telle organisation catégorielle

fournit une grande source de détails, mais propose uniquement de l'information spécifique qui ne permet pas d'être transférée à d'autres situations.

Ces deux cas extrêmes ne nous permettent pas d'aller au-delà de l'information fournie. La catégorisation en soi n'est pas utile si elle ne permet pas d'inférer sur les propriétés d'un objet ou d'un événement sonore qui ne sont pas directement accessibles. Ce processus d'inférence est rendu possible grâce à l'exploitation de la structure catégorielle de nos connaissances. Nous devons donc développer des processus de généralisation, d'abstraction et d'inférence, afin de dépasser le conflit entre un système économe en ressources cognitives et une organisation productrice d'information.

Cet équilibre entre économie cognitive et informativité se développe par l'utilisation de stratégies cognitives telles que :

- z les classes d'équivalence,
- z l'exploitation des régularités de notre environnement,
- z l'exploitation de niveaux d'abstraction privilégiés.

2.7.4.1 Classes d'équivalence

« ..., one purpose of categorization is to reduce the infinite differences among stimuli to behaviorally and cognitively usable proportions. It is to the organism's advantage not to differentiate one stimulus from others when that differentiation is irrelevant to the purposes at hand. » p. 29 [Ros78].

Une catégorie existe lorsque des objets ou des événements sonores sont traités de façon équivalente. Dans un contexte donné, les caractéristiques qui distinguent les objets, ne sont plus prises en compte [Kom92]. La cohérence fait référence à cette propriété qui doit lier les instances d'une catégorie conceptuelle [Kom92].

La cohérence des classes d'équivalence peut être abordée sous deux aspects, un aspect lié aux origines des catégories et un autre lié à la nature des catégories. Certains auteurs différencient le concept qui se réfère à une idée possédée mentalement ou à une notion, alors qu'une catégorie fait référence à l'ensemble des entités qui sont groupées ensemble [GK03]. Par exemple le concept de *chien*^[a] est tout état psychologique qui fait référence aux pensées sur les chiens. La catégorie *chien* consiste en toutes les entités dans le monde réel qui sont catégorisées de manière appropriée comme des chiens. Nous voyons que ces deux approches expliquent ce qui unit l'extension (la référence) d'un mot ou le concept, à la catégorie de choses auxquelles il s'applique.

^[a]Lorsque nous utilisons un changement de police de caractère dans ce document, celui-ci fait référence au concept et non à sa dénomination.

2.7. CATÉGORIES ET ÉVÉNEMENTS SONORES

■ **Origines des catégories** Nous pouvons distinguer trois origines aux catégories [And91].

Étiquette linguistique ⊢ Une étiquette linguistique associée à une catégorie est la preuve de son existence. Deux objets distincts associés à cette catégorie sont traités de façon équivalente. Par exemple, lorsque nous parlons du concept de *chien* et que nous disons qu'un berger allemand et un teckel (entités externes) sont des *chiens*, implicitement nous considérons qu'ils appartiennent à la même catégorie. Cette définition peut paraître circulaire, mais elle est indispensable à la communication entre individus.

Recouvrement entre objets ⊢ Lorsque nous observons que des objets se recouvrent de par leur similarité superficielle, ils sont regroupés dans une catégorie même en absence d'une étiquette linguistique dans un contexte donné.

Fonction similaire ⊢ Des objets peuvent servir la même fonction et, de par leurs attributs fonctionnels, sont traités de façon similaire au sein d'une catégorie. Par exemple, le concept de *percussion* peut regrouper au sein d'une même catégorie des objets concrets de formes très différentes tels qu'un djembé ou un bidon d'huile.

Ces différentes origines (linguistique, superficielle, fonctionnelle) ne s'opposent pas. Chacune éclaire une nature prédictive des catégories.

■ **Nature des catégories** De façon différente, nous pouvons ranger les différents types de catégories suivant le type de similarité qui les fonde [GK03], c'est-à-dire une ressemblance entre les items sur des propriétés directement accessibles aux sens, ou bien sur des propriétés plus cognitives, et leurs relations aux entités du monde externe (par ordre décroissant) :

Catégories naturelles ⊢ Elles regroupent les choses naturelles que nous retrouvons dans l'environnement tels que les chiens, les fleurs ...

Catégories artificielles ⊢ Elles regroupent les objets fabriqués par l'homme : les instruments de musique, les voitures ...

Catégories ad hoc ⊢ Elles sont créées à des fins spécifiques dans des occasions particulières [Bar83] : les choses que nous emmenons pour camper, les choses à enlever d'une maison en feu.

Concepts abstraits ⊢ Ces concepts n'ont pas de référence directe avec des entités extérieures. Il peut s'agir par exemple, de schémas abstraits impliqués dans la résolution de problèmes, de concepts tels que la beauté ou l'art.

Cette typologie met l'accent sur la similarité comme principe d'élaboration des catégories. En effet, les objets appartenant aux catégories naturelles et artificielles possèdent de nombreuses propriétés superficielles communes, contrairement aux deux dernières. Dans la section §2.8, p. 58, nous reviendrons plus longuement sur le concept de similarité.

■ **Inférence** Les classes d'équivalence au-delà de la sauvegarde des ressources cognitives ont une nature prédictive.

Les catégories permettent de généraliser nos expériences. Si nous reconnaissons qu'un objet ou événement sonore appartient à une catégorie, nous pouvons prédire les caractéristiques de celui-ci à partir des instances de cette catégorie. La catégorisation permet d'aller au-delà de l'information qui est fournie. En connaissant un item particulier d'une catégorie contextuelle, une personne peut déduire que les instances de cette catégorie peuvent lui ressembler et ainsi inférer sur les propriétés de celles-ci. Ce processus de déduction est à la base du comportement adaptatif dans la vie de tous les jours.

Ce processus semble difficile lorsque nous manipulons des catégories ad hoc ou des schémas abstraits. Dans ce cas, les instances des catégories peuvent être reliées, autrement que par la similarité superficielle, rendant difficile l'inférence. Mais grâce aux classes d'équivalence, des choses qui sont perceptivement dissemblables peuvent être traitées de manière équivalente [GK03].

2.7.4.2 Redondance structurelle

Les stimuli « simples », qui peuvent être utilisés pour des expériences en laboratoire, possèdent une structure intrinsèque. Le monde que nous percevons peut être vu, lui, comme un ensemble structuré. Ses propriétés ont des cooccurrences qui n'apparaissent pas de façon équiprobable dans la nature [Ros78].

Les objets naturels que nous percevons possèdent des structures fortement corrélées entre elles. Il existe des cooccurrences de propriétés qui sont plus fréquentes que d'autres : il est assez rare que nous percevions des organismes possédant à la fois les propriétés de fourrure, de plumes et d'ailes.

Dans notre environnement, nous pouvons noter que la cooccurrence des propriétés « plumes, ailes » - un pigeon - est nettement supérieure à la cooccurrence des propriétés « fourrure, ailes » - une chauve-souris - [Ros78] [MR81]. Dans le domaine sonore, nous pouvons juger de la nature récurrente de certains événements sonores dans notre vie quotidienne dont certains ont une fréquence écologique plus élevée que d'autres [Bal93]. Nous associons aussi, plus généralement, un son brillant, riche en fréquences élevées et dont les amortissements sont lents, à un matériau métallique qu'à du bois

2.7. CATÉGORIES ET ÉVÉNEMENTS SONORES

[Gav88]. Nous voyons ici la contribution des études en acoustique écologique qui permettent de mettre en lumière certaines invariances et régularités que le système auditif extrait.

Ainsi, les processus d'organisation de nos connaissances doivent pouvoir utiliser ces régularités, afin de privilégier certaines structures catégorielles qui semblent plus informatives que d'autres. Ces régularités reflètent une structure intrinsèque du monde perçu.

Nous remarquons ici qu'il existe une certaine écologie de l'organisation de nos connaissances, par le biais d'une organisation catégorielle. Cette écologie se traduit par la prise en compte de la structure environnementale.

2.7.4.3 Niveaux d'abstraction, niveau de base

Des chercheurs en anthropologie, [Ber78], qui étudiaient des catégories zoologiques et biologiques, ont mis en évidence que la classification des plantes utilisée par les Mayas Tzeltal du Mexique correspondait, à des catégories qui se situaient à un niveau particulier d'une classification hiérarchique scientifique (taxinomie) des plantes : le rang générique. Ce résultat apparaît dans d'autres études (voir [Atr98]).

Ce rang générique semble la meilleure partition pour résumer les catégories de façon informative et permettre ainsi une exploitation optimale de ces connaissances dans la vie quotidienne.

Dans le domaine de la psychologie et de la représentation des objets concrets, Rosch et coll. [RMG⁺76] ont montré que nous utilisons des niveaux d'abstractions privilégiés de nos connaissances : un niveau superordonné (p. ex. le mobilier), un niveau intermédiaire (p. ex. les chaises), et un niveau subordonné (p. ex. une chaise de cuisine). Ces différents niveaux d'abstraction s'organisent dans une structure hiérarchique, une taxinomie, dont les catégories sont reliées les unes aux autres par un système d'inclusion.

Ils ont pu découvrir, qu'au niveau très général, les objets partagent peu de propriétés, contrairement au niveau spécifique où beaucoup de propriétés sont communes.

Au niveau intermédiaire, les objets possèdent un nombre important d'attributs communs et il existe une discontinuité maximale entre les catégories du même niveau. A ce niveau, les objets possèdent des propriétés spéciales que les autres niveaux n'ont pas. Les catégories naturelles correspondant à ce niveau seraient le meilleur miroir de la structure corrélée de l'environnement [Ros78].

Ce niveau appelé, « niveau de base » de catégorisation, est le niveau d'abstraction qui représente le meilleur compromis entre le nombre et l'infor-

mativité des catégories. Ce niveau de base se retrouve dans d'autres domaines tels que les catégories psychiatriques [MR81].

■ **Prédiction du niveau de base** Rosch et coll. [RMG⁺76] ont proposé un indice probabiliste de prédiction (*cue validity*) qui permet d'indiquer le niveau de base potentiel en fonction des propriétés partagées par les objets des catégories d'un niveau hiérarchique considéré. Cet indice peut être mis en relation avec le modèle de contraste de Tversky [Tve77] (voir §3.3.3, p. 81).

Les catégories de niveau supérieur (p. ex. mobilier, véhicule) auront un indice de prédiction nettement inférieur à celui du niveau de base (p. ex. chaise, voiture) car les items possèdent peu d'attributs en commun, voire même plus d'attributs distincts que communs. De même, les catégories de niveau inférieur (p. ex. chaise longue, cabriolet) auront un indice de prédiction inférieur au niveau de base, car cette fois-ci, les items partagent de nombreux attributs avec les catégories contrastées (différentes) de ce même niveau d'abstraction. Le niveau de base est associé au maximum de cet indice de prédiction [Ros78].

■ **Niveau de base : niveau opérationnel** Suite à différents travaux (voir [Ros78] et [MR81] pour une revue détaillée), quatre définitions opérationnelles ont pu être associées au niveau de base :

- z Un nombre important d'attributs communs sont énumérés pour ces catégories.
- z Les programmes moteurs pour interagir avec ou manipuler les instances de la catégorie sont comparables.
- z Les membres de la catégorie ont une forme assez similaire pour que nous puissions reconnaître une forme moyenne entre les objets de cette catégorie.
- z Le nom de la catégorie est le premier qui vient à l'esprit en la présence d'un objet. De plus les objets de cette catégorie sont reconnus le plus rapidement. Les jeunes enfants ont une forte tendance à nommer tous les objets à leur niveau de base. Par exemple, tous les objets à quatre pattes sont des chiens.

■ **Remarques** Les principes cognitifs qui déterminent le niveau de base semblent universaux. Lakoff [Lak87] pense que l'universalité est garantie, car nous partageons à travers les cultures des expériences semblables qui sont structurées par la façon dont le monde est, par la façon dont nous sommes. Nous retrouvons les mêmes niveaux de base dans différentes cultures, car

les humains partagent les mêmes capacités pour la perception de la forme, l'imagerie mentale et le mouvement moteur. Ces comportements de base varient lorsqu'ils sont sous-utilisés dans une culture, ou au contraire lorsqu'ils sont développés à un niveau d'expertise.

Des variabilités existent suivant le domaine, la culture et l'expertise quant au niveau d'abstraction auquel se situe le niveau de base [MR81]. Par exemple, le niveau de base pour des taxinomies biologiques correspond au niveau supérieur, contrairement aux objets matériels. De même, il existe une forte variation dans laquelle se situe le niveau de base, suivant l'expertise dans le domaine considéré [SPB00]. Ainsi, ce qui peut être classé comme niveau de base pour un novice, peut se retrouver à un niveau supérieur d'abstraction chez un expert.

2.7.5 Théories des catégories

Une idée sous-jacente aux recherches sur la catégorisation est de tenter de comprendre comment les concepts et catégories sont représentés mentalement. A défaut de pouvoir y répondre directement, l'étude de la catégorisation s'est focalisée sur les processus qui opèrent sur les catégories. Une question importante est de comprendre les processus qui opèrent lorsque nous écoutons un son, et que nous associons ce son à la catégorie des sons de saxophone, plutôt qu'à celle de la catégorie des sons de voiture.

Nous abordons ici les différentes approches théoriques qui tentent d'expliquer ce processus. Généralement, dans la littérature quatre grandes théories sont abordées : la vue classique, la théorie du prototype (air de famille), celle des exemplaires et celle basée sur les explications. A partir de la synthèse théorique de Smith et Medin [MS81], des auteurs ont proposé une réactualisation de ces théories tout en présentant de nouvelles approches telles que des théories hybrides : Komatsu [Kom92], Goldstone et Kersten [GK03]. Il existe parfois des différences dans la façon de « catégoriser » ces différentes théories, par exemple certains modèles de la théorie des exemplaires sont vus comme des théories qui définissent les catégories sous forme de frontières [GK03].

Les trois premières théories (classique, prototype, exemplaires) font mention explicite de la similarité, contrairement à la dernière théorie basée sur les explications. Ces trois théories suggèrent que les objets ou les événements sonores soient classifiés comme des instances d'une catégorie, en vertu des attributs qu'ils partagent avec une description abstraite (logique ou prototypique) de la catégorie, ou bien avec les instances de la catégorie (exemplaires) [Kom92]. Dans le cadre de nos travaux, nous nous focaliserons avant tout sur la catégorisation basée sur la similarité, c'est pourquoi nous n'aborderons pas

ici l'approche basée sur les explications qui semblent plus pertinente pour des concepts abstraits (voir [Kom92] pour une introduction).

Ces différentes approches théoriques de la définition des concepts sont représentées schématiquement sur la figure 2.4 (avec en plus la notion de frontière).

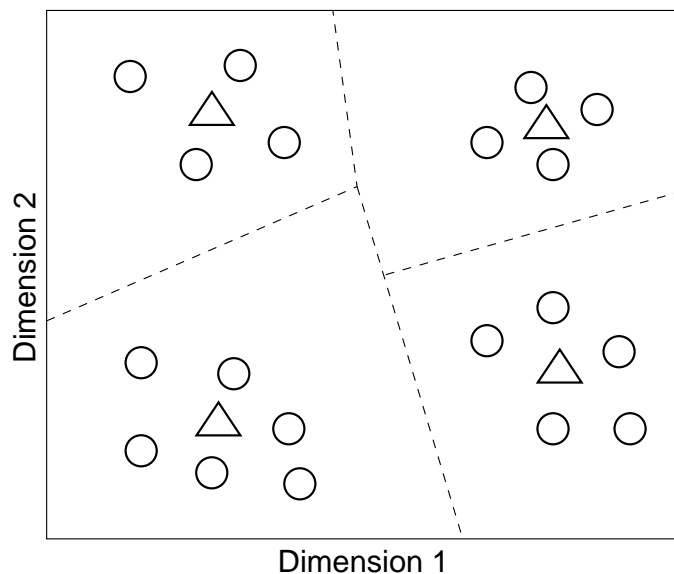


Figure 2.4 - Différentes théories suggèrent que les catégories soient représentées par des exemplaires individuels (les cercles), par une description abstraite - logique ou prototypique - de chaque catégorie (les triangles) ou bien par les frontières entre catégories (symbolisés par des lignes discontinues); adaptée de [GK03]. Les catégories sont ici représentées de manière arbitraire dans un espace à deux dimensions.

Il est intéressant de noter que nous retrouvons ce même principe d'équilibre entre économie cognitive et informativité à l'intérieur même d'une catégorie [Kom92]. En effet, soit une catégorie entière peut être représentée par une simple représentation qui abstrait les propriétés générales de l'ensemble des instances de cette catégorie, soit au contraire une catégorie mentale peut impliquer plusieurs représentations. Les différentes théories des catégories se retrouvent entre ces deux extrêmes.

2.7.5.1 Théorie classique

■ **Approche logique** Cette théorie doit être associée aux travaux de la logique qui définissent l'appartenance d'un objet ou d'un événement sonore à une catégorie, par un ensemble de propriétés qui sont nécessaires et suffisantes. Elle est aussi appelée théorie classique [SM81], en référence aux travaux

2.7. CATÉGORIES ET ÉVÉNEMENTS SONORES

d'Aristote qui cherchait déjà à définir, à dénommer et à classifier la structure de la réalité par l'utilisation de la logique [LT98].

Par exemple, un triangle équilatéral : (1) est une figure fermée, (2) avec trois côtés, (3) qui a tous les côtés de même longueur et (4) qui est construit uniquement de segments de ligne. Ainsi, toutes les figures qui possèdent ces quatre propriétés sont des triangles équilatéraux, et chaque triangle équilatéral a ces quatre propriétés.

Une catégorie, dans ce cas, est définie par deux notions : « l'intension » et « l'extension », [Fre52] cité dans [EK00]. L'intension d'un concept correspond à l'ensemble des propriétés (attributs définis) qui définissent l'appartenance à une catégorie conceptuelle. L'extension est l'ensemble des entités qui sont membres d'une catégorie conceptuelle. D'un point de vue sémantique, l'extension est la relation qui lie le terme à son référent concret dans le monde.

Cette définition a des conséquences importantes puisqu'elle implique que l'appartenance et la condition d'appartenance à une catégorie est sans ambiguïté, un objet ou un événement sonore appartenant ou non à une catégorie. De ce fait, les attributs qui ne sont pas nécessaires à l'appartenance à la catégorie ne sont pas pris en compte. Ces propriétés impliquent que les catégories subordonnées (p. ex. les professeurs) héritent de toutes les propriétés des catégories superordonnées (p. ex. les humains). Ceci ne permet pas d'expliquer pourquoi il existe des niveaux d'abstraction privilégiés, voir §2.7.4.3, p. 45. Ces contraintes fortes ont été les points d'achoppement de cette théorie.

■ **Catégories artificielles** L'approche classique tend à caractériser des catégories artificielles construites pour l'expérience avec une structure logique (p. ex. les études de Bruner et coll. [BGA56] cités dans [Gar87]). Ainsi, les expériences qui utilisent ce type de stimuli artificiels induisent les sujets à découvrir les règles de construction des stimuli plutôt que la structure intrinsèque des catégories. Nous remarquerons aussi, qu'une telle approche utilisant des stimuli artificiels limite la portée des résultats quant à l'étude de la catégorisation des objets naturels. De plus, il est difficile de savoir si de tels modèles peuvent être appliqués à d'autres situations où les règles et les catégories ne sont pas aussi précisément définies [MS78].

■ **Limites de la théorie** A partir des années 1950, cette théorie a souffert de critiques, suite aux remarques de Wittgenstein [Wit53] qui a montré que les instances de certaines catégories ne partageaient pas les mêmes propriétés ou qu'un concept était parfois difficile à définir. Par exemple, les instances de la catégorie « jeu » ne partagent pas des propriétés communes ou ne possèdent pas d'attributs définis. Le concept de jeu reste dans ce cas difficile

à définir (p. ex. « activité qui engage le plaisir », « activité avec certaines règles », ...). Ainsi, il est parfois difficile de donner une définition consciente d'un concept ou bien de pouvoir verbaliser les attributs définissant le concept (p. ex. le concept de célibataire).

Ces résultats se retrouvent lorsque nous demandons à des sujets de juger l'appartenance à une catégorie (p. ex. « Une tomate est-elle un fruit ou un légume ? »). Il s'avère qu'il existe de fortes variations intra- et inter-individuelles [EK00]. Des résultats semblables sont trouvés lorsqu'il est demandé à des personnes de donner une définition d'un concept. Ces résultats montrent que les catégories n'ont pas toujours des frontières très claires.

2.7.5.2 Théorie du prototype

Le développement de la théorie du prototype fait suite aux limites rencontrées par l'approche classique des catégories considérée comme trop rigide. L'approche théorique associée à « l'air de famille » [Wit53], ou généralement appelée « théorie du prototype », montre que la cohésion des catégories est due à l'air de famille entre leurs membres. Les membres d'une catégorie sont reliés les uns aux autres, sans qu'il existe une propriété commune à tous. Dans ce cas le principe d'appartenance à une catégorie est un processus d'appariement et non plus une vérification de conditions nécessaires et suffisantes.

Rosch et Mervis [RM75] ont proposé une notion fondamentale : les catégories ont une description centrale, un « prototype », qui représente l'ensemble de la catégorie. Ces prototypes correspondent à des représentants « moyens » extraits des différentes instances rencontrées. Ainsi, la représentation interne d'un objet ou d'un événement sonore devient l'instance d'une catégorie s'il est plus similaire au prototype de cette catégorie qu'aux prototypes des autres catégories.

■ **Structure prototypique d'une catégorie** Pour les catégories d'objets concrets, E. Rosch [Ros78] fait l'hypothèse que les prototypes se développent grâce aux mêmes principes que ceux qui gouvernent la formation des catégories perceptives (couleurs, formes élémentaires), c'est-à-dire autour d'éléments qui ont un statut perceptif ou cognitif particulier.

Rosch et Mervis [RM75] ont montré que, plus une instance d'une catégorie est considérée comme prototypique, plus celle-ci partage des attributs avec les membres de la catégorie, et moins elle en partage avec ceux des catégories contrastées. Les prototypes d'une catégorie sont ceux qui reflètent le plus la structure redondante de la catégorie entière [Ros78]. Le prototype se définit

2.7. CATÉGORIES ET ÉVÉNEMENTS SONORES

donc comme l'exemplaire qui condense les propriétés les plus saillantes de la catégorie.

Des résultats similaires sont trouvés lorsque des stimuli artificiels sont utilisés. Posner et Keele [PK68] ont utilisé des stimuli artificiels fabriqués sur la base de neuf points variant autour de positions qui formaient le prototype de chaque catégorie. Ils ont montré que les prototypes n'ayant jamais été vus avant, étaient mieux catégorisés que des nouveaux exemplaires créés à partir de distortions de ces prototypes. Les personnes ont pu apprendre ces catégories en extrayant une structure interne : le prototype. Cette généralisation est facilitée si les personnes ont été entraînées avec un nombre important d'items pour chaque catégorie.

Dans le cas de structures artificielles, le prototype précède la catégorie, contrairement à la plupart des domaines de connaissance, où les prototypes sont formés à travers l'apprentissage et le traitement de l'information [RM75].

La caractéristique principale du concept de prototype est que les personnes arrivent spontanément à lister les attributs qui sont vrais pour une catégorie donnée, alors qu'ils trouvent difficile de construire une définition explicite d'un concept à partir de ses attributs [Ham95].

■ **Typicalité** Les personnes peuvent juger certaines instances d'une catégorie plus typiques que d'autres. Par exemple, Wittgenstein [Wit53] a argumenté que certains exemplaires étaient de meilleurs exemplaires de la catégorie jeu que d'autres, par exemple le « jeu d'échec » est un meilleur représentant de la catégorie « jeu » que le « jeu de dés ». Ses observations ont été confirmées par des expériences [Ros75] [MR81]. La théorie classique ne permet pas d'expliquer ce phénomène puisque toute instance d'une catégorie peut être le représentant d'une catégorie.

Cette caractéristique des catégories conceptuelles montre qu'il n'existe pas de frontière claire telle que la théorie classique le postule, mais plutôt une gradation d'appartenance à une catégorie.

La typicalité influence différents processus [MR81] [Ros78] :

Temps de traitement ⊢ Le temps de réaction des sujets, lorsqu'ils jugent l'appartenance d'un item à une catégorie, est d'autant plus rapide si celui-ci a été jugé comme plus prototypique.

Apprentissage ⊢ Les enfants apprennent l'appartenance à une catégorie des membres prototypiques avant les autres membres.

Ordre et probabilité ⊢ Lorsqu'il est demandé aux sujets d'énumérer les membres d'une catégorie, les membres prototypiques apparaissent en premier.

Référence cognitive ▷ Les membres les plus typiques servent de référence cognitive, lorsque par exemple nous disons qu'une ellipse est presque un cercle plutôt que le contraire [Tve77].

■ **Indice de ressemblance** Suivant les auteurs, il existe deux façons de définir l'indice de ressemblance (ou l'air de famille selon Wittgenstein [Wit53]) [Kom92]. Premièrement, l'indice de ressemblance d'une instance à une catégorie augmente lorsque la similarité entre l'instance et les autres membres de la catégorie augmente et lorsque la similarité avec les membres des catégories contrastées diminue. Dans ce cas, l'air de famille décrit une relation entre les instances d'une catégorie, mais ne fait pas référence à sa représentation mentale. Cette approche peut être reliée à la théorie des exemplaires, §2.7.5.3. Dans la seconde approche, l'indice de ressemblance augmente comme la similarité entre une instance et la tendance centrale de la catégorie [Ham95]. Dans ce sens, il décrit une relation entre les instances et une représentation mentale, le prototype. Généralement, la théorie du prototype fait référence à cette dernière définition.

■ **Limites du modèle** L'approche développée par E. Rosch est avant tout une théorie structurelle, mais il est parfois difficile pour des concepts abstraits de spécifier l'ensemble des règles ou croyances qui définissent l'appartenance à une catégorie conceptuelle. Par exemple, cette approche modélise difficilement les processus de catégorisation impliqués lors de la formation de catégories « ad hoc » [Bar83].

De même, cette approche est limitée dans la façon de voir comment les connaissances sur les concepts sont utilisées dans le processus de catégorisation. En effet, les personnes n'utilisent pas simplement les attributs seuls mais aussi les relations ou la corrélation qui existent entre ces attributs [GMG91].

E. Rosch [Ros78] pose elle-même les limites de la théorie en expliquant que les prototypes ne spécifient, ni le type de représentation, ni les modèles de traitement. Elle impose plutôt des contraintes associées aux catégories.

Il est difficile de ne pas tenir compte de l'influence du contexte lors du processus de catégorisation. Il est clair qu'il existe des effets de contexte qui influencent le niveau d'abstraction auquel l'objet est considéré, mais aussi le choix des items qui sont nommés, appris, listés ou bien attendus dans une catégorie [Ros78]. Il est aussi important de reconsidérer les notions de niveau de base, de prototype dans le cadre d'un contexte plus naturel qu'une situation de laboratoire.

■ **Extension du modèle** Généralement, il existe une vue restrictive de ce qu'est un prototype. Les auteurs supposent souvent que les prototypes ne contiennent que des caractéristiques perceptives superficielles d'objets ou d'événements sonores, ou directement observables. La similarité serait dans ce cas uniquement calculée sur la base de ceux-ci. Bien sûr, toutes ces caractéristiques sensorielles existent, mais il n'y a aucune raison pour que d'autres caractéristiques ne soient pas représentées par un prototype. Par exemple, certains concepts abstraits possèdent des structures prototypiques tels que « la Science » ou « l'Art » [Ham81].

Théorie explicite J.A. Hampton [Ham93] a proposé une version « explicite » de la théorie du prototype afin de la formaliser de manière opérationnelle. L'appartenance d'un item à une catégorie est déterminée par le calcul d'une mesure de similarité d'un item au prototype. Nous voyons, dans ce cas, que le prototype a une définition explicite. La mesure de similarité est basée sur le degré de concordance entre les caractéristiques pondérées de cet item et celles du prototype. En plus de la pondération associée aux différents traits, il existe un critère de seuil lié à cette échelle de similarité. Le seuil à partir duquel un item appartient à une catégorie, permet d'expliquer les variations entre sujets, suivant le contexte [Ham95]. L'instabilité du processus de catégorisation peut être expliqué par le seuil et par le poids accordé aux différents traits suivant le contexte. Généralement, cette instabilité est liée aux instances les moins typiques. Une des puissances de ce modèle est d'expliquer pourquoi des items peuvent appartenir clairement à une catégorie si le poids associé aux caractéristiques est nettement supérieur au seuil. Dans ce cas (p. ex. les nombres impairs), le poids est suffisamment élevé pour qu'il n'y ait aucune ambiguïté quand à l'appartenance à la catégorie mais il existe tout de même des variations qui reflètent la typicalité des membres.

2.7.5.3 Théorie des exemplaires

Suite aux limites de « l'approche classique », la théorie de « l'air de famille » a permis de rendre compte de l'effet de typicalité au sein d'une catégorie. Il existe une autre vue, la théorie des « exemplaires » ou de « l'instance » qui tend à assouplir certaines contraintes associées aux deux théories précédentes. Dans l'approche prototypique, les attributs uniques d'un item qui ne contribuent pas à la tendance centrale ou à l'air de famille ne sont pas inclus dans la représentation prototypique. Contrairement à cette approche, la théorie de l'instance permet aux attributs qui sont spécifiés par les instances d'une catégorie de ne pas obligatoirement être partagés par toutes les instances. Ainsi, un exemplaire d'une catégorie peut avoir des attributs qui

lui sont propres. Pour résumer, les trois approches (classique, air de famille, de l'instance) ont des critères d'appartenance à une catégorie allant respectivement d'une définition logique contraignante à une définition « souple ».

■ **Modèle du contexte** La version simple de la théorie des exemplaires, le modèle du contexte (Context model), a été présentée par Medin et Schaffer [MS78]. Le modèle stipule que nous représentons les catégories par le stockage des exemplaires en mémoire et nous classons les objets ou les événements sonores sur la base de leur similarité avec les exemplaires en mémoire [NJ00]. Ainsi, la similarité d'un item à une catégorie est la somme des similarités entre l'instanciation de cet objet et toutes les instances de chaque catégorie pertinente [GK03]. Cette approche a été motivée par l'étude des relations entre les processus de catégorisation, d'identification et de reconnaissance [Nos92].

Une des caractéristiques de ce modèle est que les similarités entre les exemplaires, peuvent être modifiées par l'attention sélective. Les propriétés des stimuli peuvent différer de part leur saillance, possédant par exemple un poids associé à chacune des propriétés particulières (dans le cas de dimensions d'objet comme la taille, la position, la brillance) [Est94]. Les attributs les plus saillants attirent d'autant plus l'attention. Ces attributs peuvent être pondérés – par un « poids attentionnel » – de manière appropriée suivant le contexte, ce qui a donné le nom de « context model ».

Abstraction L'approche de l'instance ne réfute pas qu'il y ait abstraction des propriétés des instances d'une catégorie. Par contre, il n'existe pas d'abstraction de la tendance centrale à travers toutes les instances de la catégorie. Il existe deux approches de la représentation des exemplaires. Une approche utilise une représentation sous forme de traits ou de valeurs d'attributs [MS78], par exemple le modèle de Estes qui représente une catégorie sous forme de tableau de propriétés propres à chaque exemplaire [Est94]. Une deuxième approche est de représenter les exemplaires comme des points dans un espace psychologique multidimensionnel [Nos92].

L'abstraction a lieu lors de l'utilisation du concept (jugements de typicalité, d'appartenance d'un objet à une catégorie, raisonnement) et non pendant l'apprentissage, contrairement à la théorie prototypique où l'abstraction fait partie intégrante du concept et peut parfois préexister au concept. En effet, les prototypes peuvent préexister aux catégories lorsque ceux-ci ont des bases physiologiques (couleurs, formes) ou lorsque nos systèmes perceptifs reflètent les partitions naturelles des objets du monde externe. Mais dans beaucoup d'autres domaines, les prototypes sont formés à travers les processus d'ap-

prentissage et de traitement de l'information [RM75].

L'approche de l'instance ne requiert pas que toutes les informations sur les instances soient retenues. Certaines instances ou des détails spécifiques peuvent être oubliés au cours du temps, ou bien non codés initialement suivant l'attention [MS78]. Plutôt que d'enregistrer passivement les événements de manière distincte, il nous semble naturel de combiner les événements qui font référence à la même entité [GK03].

Typicalité et tendance centrale Les décisions pour catégoriser sont basées sur la similarité, entre l'objet qui est catégorisé et tous les exemplaires de chaque catégorie pertinente. Lorsqu'un item devient plus similaire aux exemplaires d'une catégorie A, ou moins similaire aux exemplaires des autres catégories, la probabilité de placer cet item dans la catégorie A augmente. Les jugements catégoriels peuvent changer suivant l'individu ou le contexte lorsqu'un item est aussi proche de deux ensembles d'exemplaires, car l'appartenance à une catégorie est basée sur des règles probabilistes. Les items varient en typicalité par rapport à une catégorie de la même façon que leur similarité à l'ensemble des exemplaires agrégés [GK03]. La typicalité peut s'expliquer aussi par le nombre d'instances associées en mémoire pour une catégorie. Ainsi un rouge-gorge est beaucoup plus typique qu'un pingouin car nous sommes confrontés à beaucoup plus de rouge-gorges que de pingouins. Nous avons donc plus d'instances du premier en mémoire [EK00]. Nous pouvons expliquer les gradients de typicalité de la même façon. Sans qu'il y ait une réelle existence d'un prototype, une tendance centrale peut émerger de manière cumulative, à travers les instances d'un concept. Le prototype sera l'exemplaire le plus similaire à l'ensemble des instances de la catégorie. Par exemple, pour l'expérience de Posner et Keele, §2.7.5.2, le prototype d'une catégorie sera en moyenne plus similaire aux exemplaires qui ont servi pendant l'entraînement que les exemplaires nouveaux car ces derniers sont des distortions des prototypes [GK03].

■ **Extensions du modèle** Suite à l'introduction du modèle de contexte de Medin et Schaffer [MS78], différentes extensions du modèle ont été proposées : le modèle du contexte généralisé (GCM), la théorie de reconnaissance généralisée (GRT) qui sont à la base de nombreuses variantes (voir [UJR98] pour une introduction aux différents modèles).

Nosofsky [Nos86] a généralisé (GCM : General Context Model) la version originale du modèle de contexte en utilisant explicitement un modèle de distance géométrique qui se rapproche des modèles multidimensionnels (voir §3.4.2, p. 85), pour estimer la similarité entre un événement sonore et les

instances d'une catégorie, sur la base des propriétés disponibles au niveau de l'événement sonore. L'appartenance à une catégorie est estimée par une probabilité conditionnelle d'appartenance.

La théorie générale de la reconnaissance (GRT : General Recognition Theory) a été proposée par Ashby et Perrin [AP88] [Ash92a] [Per92]. Ce modèle peut être vu comme une extension multidimensionnelle de la théorie de la détection du signal. La différence fondamentale avec la GCM est que les catégories sont associées ici à des régions dans l'espace psychologique. Ce modèle est parfois lié aux théories basées sur les frontières de catégories [GK03], plutôt que la théorie des exemplaires. Et contrairement aux prototypes qui représentent la tendance centrale d'une catégorie, cette approche définit les catégories par leur périphérie. Ces deux approches sont en accord avec les résultats de la perception catégorielle (voir [Har87] pour une revue sur ces recherches).

2.7.5.4 Bases communes de la catégorisation

Malgré les différences apparentes entre les diverses théories, elles reposent sur trois hypothèses communes : la représentation, la recherche d'information et le choix d'une réponse [Ash92b] [AM93].

■ **La représentation** En ce qui concerne la représentation interne d'un stimulus et des catégories contrastées, la plupart des théories supposent que l'effet perceptif de la présentation d'un stimulus peut être représenté par un point dans un espace multidimensionnel. Ce postulat est parfois controversé, certains auteurs comme Tversky [Tve77] (voir le modèle de contraste §3.3.3, p. 81) adoptent plutôt une représentation non-numérique, sous forme d'ensemble de traits. Les théories diffèrent sur l'existence de variabilité dans les percepts associés à un stimulus seul et sur la façon dont les catégories peuvent être représentées. Dans le cas de la GRT, la représentation interne des exemplaires est vue comme des distributions probabilistes multivariées.

■ **Recherche d'information** Ce processus décrit l'information exacte qui doit être collectée avant qu'une réponse puisse être donnée. La théorie des exemplaires postule que le sujet doit calculer la similarité entre la représentation du stimulus et la représentation en mémoire de tous les exemplaires de chaque catégorie. Dans le cas de la théorie du prototype, la probabilité que l'instanciation d'un item appartienne à une catégorie est déterminée par la similarité de cet item, à chacun des prototypes des catégories. La similarité est généralement associée à une fonction monotone décroissante de la distance psychologique entre les points représentant l'item et le prototype.

■ **Choix d'une réponse** Cette hypothèse décrit comment un sujet sélectionne une réponse après que l'information pertinente ait été collectée. Les différences se situent dans le type de réponse donnée, déterministe ou probabiliste. Une réponse déterministe donnera toujours la même réponse pour une information donnée alors qu'un processus probabiliste non. Dans l'approche classique, la réponse est déterministe de par la définition logique d'appartenance à une catégorie. Les théories de l'instance postulent que l'appartenance à une catégorie est probabiliste, (modèle de choix de similarité). Par contre pour la théorie explicite du prototype, la réponse est « déterministe » suivant un critère de seuil assurant l'appartenance à une catégorie. La GRT fonctionne sur un critère de frontière plutôt que de seuil, en intégrant un bruit perceptif. Ashby et Maddox [AM93] montrent que les règles déterministes et probabilistes sont équivalentes, dans le sens où le modèle de choix de similarité permet de modéliser aussi bien le comportement d'un sujet donnant des réponses fortement variables que faiblement variables.

2.7.6 Discussion

Les différentes théories que nous venons de présenter possèdent leurs propres caractéristiques quant à l'appartenance à la catégorie, sa représentation... Il existe maintenant des approches qui tentent de combiner les bénéfices de chacune d'elles. Généralement, ces théories se combinent en associant une représentation abstraite (règles ou prototypes) avec des instances individuelles [NJ00]. Par exemple, Erickson et Kruschke [EK98] ont développé une théorie modulaire (ATRIUM) qui implique à la fois un modèle basé sur les règles et une approche basée sur les exemplaires. Le modèle RULEX proposé par Nosofsky et coll. [NPM94] est basé sur des règles simples, logiques et sur un stockage des exceptions occasionnelles à ces règles. Smith et coll. [SMM97] ont associé un modèle prototypique à un modèle d'exemplaires. Dans ce modèle, le prototype est vu comme un objet idéalisé qui est composé des valeurs des dimensions observés le plus fréquemment chez les membres d'une catégorie. De la même façon que précédemment, les exemplaires reflètent les exceptions. Malgré les différences rencontrées, ces différents modèles utilisent à chaque fois la notion de similarité entre un objet ou événement sonore à catégoriser et les catégories auxquelles il peut appartenir. Nous allons nous focaliser sur la similarité pour la définir de manière plus précise et la rendre ainsi « opérationnelle ».

2.8 Similarité

2.8.1 Introduction

Lors de l'étude des différentes théories de la catégorisation, nous avons fait mention de la similarité. L'idée intuitive, et qui a été reprise dans de nombreuses théories de la cognition, est que nous classons ensemble les choses que nous trouvons similaires :

« Remembering accurately events that happened in the past would help our present activities very little if we were unable to judge the similarities between past events and present situations so as to have a basis for deciding on the relevance of things remembered for present tasks. » p. 15 [Est94].

Ainsi, lorsque nous demandons à des auditeurs d'identifier une source sonore qui n'est pas apparente, il est fréquent d'entendre : « ce son ressemble à ... », il nous paraît naturel d'essayer de reconnaître un son par comparaison à nos expériences antérieures.

Les différentes approches théoriques de la catégorisation que nous avons introduites postulent que la façon d'affecter un item à une catégorie, dépend de la similarité entre cet item et cette catégorie. Suivant les théories, l'appartenance à une catégorie est associée aux conditions nécessaires et suffisantes (théorie classique), à la relation perceptive entre l'item et « le/les prototype/s » (théorie prototypique), à la relation perceptive entre l'item et « les exemplaires » (théorie de l'instance).

Mais un certain nombre de théoriciens ont questionné le degré avec lequel la notion de similarité est suffisamment définie et contrainte pour servir de seule explication au processus de catégorisation [Gol94]. Les objections envers la similarité sont généralement :

- z qu'elle est flexible et indéterminée (quelle mesure de similarité ? quelles propriétés sont impliquées dans cette mesure ?),
- z qu'elle est trop ancrée dans la perception pour expliquer des concepts abstraits.

2.8.2 Dissociation entre similarité et catégorisation

Certains auteurs ont montré l'existence d'une dissociation entre jugement de similarité et catégorisation. Rips et Collins [RC93] ont demandé à des groupes différents de juger soit explicitement la similarité entre des instances individuelles et des membres de catégories. Dans une autre condition, les participants ont jugé la probabilité que les mêmes instances soient membres de ces catégories.

Les résultats montrent que les sujets jugent la similarité plutôt sur la tendance centrale de la catégorie, alors que les jugements d'appartenance sont liés à la distribution de l'information au sein des catégories. D'autres expériences montrent que notre connaissance, sur la variabilité à l'intérieur d'une catégorie, peut influencer une décision d'appartenance mais pas un jugement de similarité [Rip89].

Le jugement catégoriel dépend de connaissances génétiques, biologiques, historiques... qui permettent de mettre en lumière des propriétés causales cachées, des relations aux autres entités. Les jugements de similarité dépendent, quant à eux, des propriétés intrinsèques aux items (la taille, la forme ...) [RC93].

Le jugement de similarité apparaît plus focalisé sur le stimulus et la similarité impliquée dans la catégorisation plutôt basée sur la description de la catégorie [Urd98]. Le jugement de similarité répondrait à la question : « A quoi ressemble ce que je perçois ? », le jugement de catégorisation répondrait à la question « Puis-je mettre ce que je perçois en relation avec ce que je connais ? ». La similarité serait plus guidée par les données et la catégorisation par les théories, nos connaissances.

La définition de la similarité doit donc être assez flexible, afin de s'adapter à des idiosyncrasies, mais doit être en même temps stable pour servir de fondement pour la cognition. La stabilité sans la flexibilité donne des jugements qui ne sont pas vraiment adaptés à des demandes spécifiques. La flexibilité sans la stabilité produit des jugements qui ne bénéficient pas de l'information qui reste constante à travers les situations [Gol95]. Nous voyons le besoin de réconcilier ces deux vues, afin de les intégrer dans une même approche.

Nous présentons tout d'abord les aspects flexibles de la similarité, pour aborder ensuite une approche théorique qui nous semble réconcilier ces approches divergentes.

2.8.3 Flexibilité de la similarité

Il apparaît que des mêmes objets peuvent être membres de catégories différentes en fonction du contexte, par exemple, un saint-bernard peut être membre des catégories : chien, animal de compagnie, sauveteur, chien de garde ...

Si nous regroupons des items exclusivement suivant leur air de ressemblance, il est étonnant qu'ils puissent appartenir à plusieurs classes à la fois. L'exemple précédent montre que différentes propriétés sont susceptibles d'être impliquées dans la mesure de similarité. De ce fait, la similarité entre des objets peut être radicalement différente suivant les propriétés utilisées.

Au contraire pour des concepts qui sont explicitement définis (p. ex. définition du triangle équilatéral §2.7.5.1, p. 48), la similarité entre deux items d'une même catégorie peut être réduite à la définition d'appartenance. La similarité, dans ce cas, est réduite à la simple définition de conditions nécessaires et suffisantes qui rend caduque la notion de similarité [Hamon].

Ceci pose un problème de définition de la similarité. Dire que « un objet α est similaire à un objet β » doit être alors assorti de la mention : « suivant la (ou les) propriété(s) X », [Goo72] cité dans [MGG93].

Il convient donc de spécifier quelles sont les propriétés pertinentes qui sont intégrées dans la mesure de similarité. Or, à première vue, il semble exister une circularité liée à l'utilisation de la similarité comme explication théorique des catégories. La similarité semble alors présupposer la catégorisation de l'item [MS95].

■ **Similarité et connaissances** Nos connaissances peuvent nous aider à spécifier les propriétés pertinentes d'un événement sonore, dans un contexte donné ou bien à focaliser notre attention sur certaines propriétés. Une autre possibilité est que ces connaissances permettent de lier ensemble les propriétés d'un concept, lorsqu'elles ont été identifiées. Des études ont montré que lorsque la structure catégorielle est cohérente avec nos connaissances, l'apprentissage de ces catégories est facilité, car nos connaissances font émerger les relations entre les propriétés [MS95]. Par contre, lorsqu'il existe une incohérence entre les deux, l'apprentissage devient difficile.

De la même façon, nous pouvons inférer grâce à nos connaissances certaines propriétés qui ne sont pas directement accessibles. Les relations entre attributs peuvent faire émerger des attentes sur les autres attributs possibles. Par exemple lorsque nous évoquons le concept de *prédateur*, nous inférons aisément sur le fait que celui-ci doit être, entre autre, rapide et avoir des dents acérées.

« One important use of knowledge is in specifying the relevant features in a domain » p. 129 [MS95].

■ **Similarité et effet de contexte** Généralement, une approche de la similarité est basée sur un ensemble fixe de propriétés, voir §3.3, p. 74. Chacune des entités qui est comparée est représentée sous forme de propriétés. Ces propriétés sont déterminées avant la comparaison [Gol95]. Cette approche a l'intérêt d'être modulaire mais elle est valable uniquement si les propriétés ne dépendent pas l'une de l'autre ou du contexte.

Or la similarité dépend du contexte et du cadre de référence. En général, l'espace des propriétés pertinentes n'est pas spécifié de manière explicite mais plutôt par le contexte [Tve77]. Ainsi, la similarité n'est plus vue simplement comme une relation entre deux objets, mais comme une relation entre deux objets et un contexte [GMH97]. L'effet de contexte, dans ce cas influence aussi bien la catégorisation que la similarité. Tversky [Tve77] et Tversky et Gati [TG78] ont montré que la saillance perceptive des propriétés des items peut varier suivant le contexte, par exemple, suivant les variations dans le corpus d'objets étudiés (p. ex. en enlevant ou en ajoutant des items).

■ **Catégorisation suivant le but** Barsalou [Bar83] a montré l'existence de catégories appelées « ad hoc » telle que la catégorie « des choses à enlever d'une maison en feu », c'est-à-dire un type particulier de catégorie que nous formons dans un but précis. La cohérence, entre les instances de la catégorie, acquiert du sens, uniquement dans un contexte particulier. Par contre, il n'existe pas forcément une grande similarité superficielle entre les attributs perceptifs des membres de cette catégorie. Par exemple, dans ce cas, « un bébé » et « des bijoux » sont associés à cette catégorie bien qu'il existe peu de ressemblances entre ces deux items.

Anderson [And91] va à l'encontre de ces explications en affirmant qu'il existe une confusion entre l'étiquetage linguistique et les catégories. Pour lui, le fait de créer une phrase nominale appropriée ne veut pas dire que ses référents deviennent des catégories mentales. Nous pouvons raisonner sur des catégories « ad hoc », mais ce raisonnement implique une inférence causale et non une inférence catégorielle. Hampton parle lui de similarité « tautologique », car rien n'explique comment ces catégories sont contruites [Hamon].

2.8.4 Similarité base de la catégorisation

Nous avons vu précédemment qu'un grand nombre de modèles de la catégorisation sont basés sur la similarité. La flexibilité de la similarité rend difficilement généralisables ces différentes théories si nous admettons une dichotomie entre similarité « pure », basée sur l'information perceptive, et une approche purement explicative de la catégorisation, basée par exemple sur des théories scientifiques. Dans la continuité de cette idée, Goldstone et coll. [GB98] ont proposé que la pensée conceptuelle soit enracinée dans la similarité perceptive. Dans ce cas, les processus perceptifs guident la construction des concepts abstraits, même si le lien n'est pas forcément direct. La similarité et les concepts ont leurs racines dans la perception et sont fortement reliés aux mécanismes perceptifs dans leurs implémentations.

2.8.4.1 Contraintes sur la similarité

Bien que nous ayons montré les effets de contexte et de nos connaissances sur la similarité, il faut tout de même indiquer, que la similarité possède un noyau stable qui est relativement indépendant du contexte [Gol94]. Des études montrent que même lorsque des règles sont imposées lors de la catégorisation, les personnes ne peuvent pas ignorer les similarités superficielles. Un exemple connu est l'interférence de Stroop [Str35] cité dans [Gol94] : les personnes mettent plus de temps à nommer la couleur d'un mot si le nom est une couleur différente comparativement à un mot neutre (p. ex. le mot « rouge » écrit en jaune vs. le mot « roue » écrit en jaune).

Lors d'un jugement de similarité, le processus de comparaison entre items impose lui aussi des contraintes lorsque nous ne pouvons pas déterminer les propriétés pertinentes des items pris séparément [Gol94]. Lorsqu'un item possède des propriétés ambiguës, celles-ci sont déterminées lorsque celui-ci est mis en comparaison avec d'autres items [MGG93].

Il existe un certain nombre de données expérimentales qui montrent que les interactions entre items comparés, les variables associées à la tâche et aux stimuli, servent à résoudre des cas incertains dans un jugement de similarité. Ces contraintes sont importantes, car elles illustrent le fait que « contraint » n'est pas la même chose que « fixé et statique » [Gol94]. Si nous connaissons les deux items qui sont comparés et que nous connaissons les variables de la tâche, ou de la situation, nous pouvons prédire quelles peuvent être les propriétés qui seront importantes lors du processus de comparaison et donc du jugement de similarité. La perception de la similarité doit être vue comme un processus dynamique et constructif.

Par exemple, si nous jugeons la similarité entre les Etats-Unis et l'Angleterre sur une échelle de 20, il est probable qu'une petite partie de nos connaissances sur ces deux pays sera activée (p. ex. le sport pratiqué, histoire commune des migrations). Dans ce cas, l'information accessible pour un concept sera étendue et testée pour son application sur les autres concepts [MGG93].

2.8.4.2 Différents types de similarité

Par exemple, il nous semble naturel de penser que la similarité des ordinateurs aux téléviseurs n'est pas du même type que la similarité entre les ordinateurs et les bouliers [Gol95]. Dans un cas, nous pouvons restreindre la similarité aux propriétés directement accessibles aux sens, pour reprendre notre exemple, l'ordinateur et le téléviseur ont dans les deux cas un écran de visualisation. A l'opposé, la similarité peut être intégrée dans un système

d'abstraction qui met en jeu notre expertise, nos connaissances ou bien un but spécifique. Par exemple, nous pensons la similarité entre l'ordinateur et le boulier en faisant référence à la fonction de calcul de ces deux objets. Dans cette comparaison, nous voyons que seule une connaissance, à priori, de ces objets nous permet de les mettre en relation et de les trouver similaires. Ces deux exemples englobent-ils la même conception de la similarité ?

Afin de reconcilier stabilité et flexibilité de la similarité, il convient de faire une distinction entre la similarité par défaut et celle dirigée par nos connaissances [Gol95]. La similarité par défaut est utilisée lorsque peu d'informations spécifiques sont connues sur les items. La comparaison se fait sur les informations directement accessibles. La similarité dirigée est utilisée dans des cadres plus informatifs, lorsque nous observons des dépendances contextuelles entre les items, sur la base de nos connaissances, de notre expertise et de nos croyances. Certains auteurs parlent de similarité de « surface » et de similarité « profonde » [Hamon].

Nous allons détailler les différentes caractéristiques propres à ces deux types de similarité qui sont résumées dans le tableau 2.3.

Sophistication \triangleright Par définition, la similarité par défaut est moins sophistiquée que celle dirigée. Elle intervient plus lorsque nous effectuons des comparaisons rapides, plus systématiquement chez les enfants ou lorsque nous effectuons une tâche de manière distraite. Nous nous focalisons sur des propriétés qui sont directement accessibles par nos sens. La similarité dirigée étant plus contextuelle, elle demande une élaboration propre à la situation comme dans l'exemple que nous avons donné sur la comparaison entre les Etats-Unis et l'Angleterre.

Information \triangleright Si nous avons peu d'information sur les items que nous comparons ou la propriété qui nous intéresse, la similarité par défaut intervient plus particulièrement.

Catégories \triangleright La similarité par défaut, qui est particulièrement stable, produit des catégories référentielles que nous utilisons quotidiennement. Au contraire, la similarité dirigée est utilisée pour créer dynamiquement des catégories. Ainsi les taxinomies, voir §2.7.4.3, sont des structures qui semblent être utilisées de manière automatique sur la base de similarités stables, contrairement aux catégories ad hoc, voir §2.8.3, qui sont générées suivant le contexte.

Construction \triangleright La similarité dirigée est activée sur le court-terme dans le cadre de situations particulières, de manière temporaire. La similarité par défaut semble ancrée chez l'organisme percevant. Sur la base de pré-supposés perceptifs, la similarité par défaut s'adapte avec l'expérience,

tout au long de la vie. A la place d'une dichotomie entre ces deux similarités, il faut plutôt y voir une continuité car une similarité dirigée peut se systématiser avec le temps et l'expérience en une similarité par défaut.

Inférence \triangleright Les deux similarités ont un rôle différent dans le processus d'inférence. La similarité par défaut produit des inférences d'ordre général qui sont basées directement sur l'observation. La similarité dirigée produit des inférences spécifiques qui sont utilisées lorsque la similarité superficielle est relativisée par nos connaissances ou nos idiosyncrasies.

Par exemple, la similarité par défaut nous permet de voir la similarité entre les crabes et les homards. La similarité dirigée peut créer des connexions plus lointaines, en liant les crabes et les araignées sous la catégorie « arthropode ».

La similarité par défaut sert de fondement à la similarité dirigée, mais si celle-ci suit régulièrement un pattern systématique, elle peut devenir avec le temps, une similarité par défaut. Par exemple, lorsque nous écoutons des musiques qui sortent du cadre tonal occidental, nous percevons parfois cette musique comme dissonante. Le plaisir vient alors par une intégration et une compréhension implicites de la structure et des sonorités au fil des écoutes.

Dimensions comparées	Similarité par défaut	Similarité dirigée
Sophistication	simple	complexe
Information	peu d'information	information substantielle
Catégories	taxinomie	ad hoc, métaphores
Construction	long terme	court terme
Inférence	générale	spécifique

Tableau 2.3 - Caractéristiques de la similarité par défaut et de la similarité dirigée [Gol95].

2.8.4.3 Fondement de la catégorisation

Même s'il est important de spécifier les différents degrés d'élaboration de la similarité, nous voyons que la similarité peut servir de fondement à la catégorisation.

Généralement, les objections contre la similarité sont souvent associées à la confusion entre, les concepts comme constructions culturelles, et les concepts comme éléments de représentation mentale [Hamon]. Le concept comme construction culturelle est ce qu'une culture a pu développer et faire évoluer à travers des générations de pensée et de découvertes. Des experts, dans chaque culture, avec une responsabilité sociale reconnue apprennent

ces concepts des générations précédentes, les perfectionnent parfois et les transmettent aux générations futures.

Lorsque dans de nombreux cas où la catégorisation ne demande pas une correction reflétant les normes culturelles, chaque individu utilise ses propres schémas de représentation des concepts et défend leur validité. Dans la vie quotidienne, nous continuons à mettre ensemble les choses qui nous semblent similaires (de manière superficielle ou profonde) [Hamon].

Comme nous avons pu le voir, la similarité ne doit pas être restreinte à ce qui est directement accessible. La similarité doit être vue comme une intégration de différentes propriétés (perçues de façon holistique ou analytique) et qui prennent sens dans le cadre d'un contexte et de nos connaissances.

« La catégorisation inclut certes le traitement des propriétés perceptibles des objets mais celles-ci ne font sens qu'articulées sur une théorie qui, en leur accordant un statut soit essentiel, soit accidentel, en détermine l'appartenance catégorielle » p. 10 [Dub97].

2.9 Discussion

Dans ce chapitre, nous avons introduit un cadre théorique pour l'étude de la perception des propriétés des sources sonores en lien avec leur reconnaissance de ces événements sonores.

Nous avons présenté deux approches théoriques qui contribuent à éclairer différemment la reconnaissance des sources sonores. Une première approche est associée à la théorie du traitement de l'information, basée sur l'étude des propriétés perceptives des événements sonores. Elle met l'emphase sur les traitements internes qui permettent de former une représentation de l'événement sonore, sur la base de propriétés perceptives, telle que le timbre. Cette approche pourrait relier information sensorielle et propriétés mécaniques par une approche constructiviste et inférentielle ou associative. Une deuxième approche, la théorie écologique met l'emphase sur la perception des propriétés audibles des sources sonores. Elle postule que les invariants structuraux et transformationnels des événements sonores donnent une information structurée, un pattern d'énergie. Cette information spécifie de manière suffisante les propriétés des sources sonores. Elle nous permet la reconnaissance des sources sonores par une saisie de l'information et non pas par des processus de traitement et de représentation interne de l'événement sonore.

Dans le cadre de cette thèse, nous prenons une approche intermédiaire entre ces deux cadres théoriques. L'acoustique écologique, telle que nous l'entendons, étudie le lien entre les propriétés physiques des sources sonores et la

représentation auditive de ces propriétés. Nous associons les propriétés audibles des sources sonores aux propriétés perceptives étudiées par la psychophysique. Sur la base d'un modèle de reconnaissance des événements sonores, nous cherchons à connaître les propriétés audibles qui permettent l'accès au lexique auditif de nos connaissances sur notre environnement sonore.

La catégorisation, et la similarité qui est la base de ce processus, nous permettent d'étudier les propriétés sous-jacentes à la formation de catégories de sons et ainsi de mettre en lumière les propriétés impliquées dans la reconnaissance des sources sonores. Le but de cette thèse n'est pas de comparer les différentes théories de la catégorisation d'un point de vue expérimental, mais d'utiliser ce processus pour remonter aux propriétés audibles des sources sonores.

Nous présentons dans le chapitre suivant des techniques d'analyse des données de similarité, dans un terme plus général de proximité. Elles nous permettent de représenter la similarité suivant différents modèles (de représentation graphique ou de calcul de la similarité) que nous mettons en perspective avec cette partie théorique. Ainsi, la similarité est pour nous un point de départ important pour révéler les processus perceptifs impliqués dans la catégorisation de sources sonores et la perception de leurs propriétés audibles.

Chapitre 3

Représentation des proximités



3.1 Introduction

Nous avons décrit dans le précédent chapitre le rôle de la similarité dans les processus de catégorisation. La catégorisation nous semble importante pour refléter la structure perceptive ou cognitive de l'organisation de notre environnement sonore. En nous intéressant à la perception des sources sonores, nous cherchons à connaître ce que le système auditif extrait comme invariant au niveau du stimulus proximal, afin de refléter les propriétés de ces sources. Cette extraction nous semble indispensable à la reconnaissance des sources sonores. L'étude des catégories nous permet d'étudier les propriétés sous-jacentes aux catégories et ainsi de comprendre quelles sont les propriétés potentielles qui peuvent être impliquées dans la reconnaissance des sources sonores.

La similarité entre les événements sonores nous semblent donc un point d'étude important pour atteindre cet objectif. L'étude de la similarité s'en-globe dans un cadre plus large appelé « étude des proximités » qui permet d'aborder la proximité entre des stimuli à différents niveaux de description comme nous l'avons déjà introduit, voir §2.3.2, p. 13, c'est-à-dire physique,

perceptif/cognitif...

Ainsi dans cette partie, nous introduisons différentes techniques nous permettant l'analyse de nos données expérimentales présentées dans les chapitres 5 et 6. Nous les confronterons avec des analyses mécaniques et acoustiques de nos stimuli, voir chapitre 4, afin de tester nos hypothèses théoriques.

Ce chapitre a pour but d'introduire différentes techniques, allant du calcul des proximités entre stimuli, à leurs analyses et à leurs représentations. L'analyse des proximités nous permet de visualiser et d'interpréter l'information pertinente contenue dans les données.

3.1.1 Notion de proximité

Le terme de proximité entre stimuli englobe : la notion de similarité, de dissemblance, d'intercorrélation entre stimuli, terminologie suggérée par Shepard [She62a] [She62b], c'est-à-dire toute variable qui fait référence à une notion de promiscuité ou de distance entre deux items de même type. Dans notre cas, les stimuli étudiés sont des événements sonores.

L'analyse des proximités nous permet d'expliquer les relations de similarité observées entre ces événements sonores, de capturer leurs structures perceptives sous-jacentes [Tve77]. La production de jugements de proximité par des auditeurs sur des événements sonores, des expériences de classification de sons peuvent permettre de refléter ces processus perceptifs. Mais ces analyses peuvent être effectuées sur la proximité entre des événements sonores sur une base physique ou acoustique, pour les mettre en correspondance avec les résultats perceptifs.

3.1.2 Modèles

Dans ce cadre théorique, il existe des modèles cognitifs de la proximité qui décrivent de façon rigoureuse la base structurelle qui sous-tend les jugements de similarité [VMS95]. Il existe des techniques d'analyse de données, développées dans le cadre de ces modèles, telles que les analyses multidimensionnelles ou les représentations arborées.

Le résultat de l'analyse des proximités est contraint par le choix du modèle de représentation de ces données. La représentation des proximités est une étape essentielle dans l'analyse des données expérimentales. C'est pourquoi, il est essentiel de comprendre le fondement de ces méthodes.

N'oublions pas que ces analyses ont pour objectif de faciliter l'interprétation grâce à une réduction statistique des données.

3.1.3 Synopsis d'une analyse

Le schéma de Kruskal (figure 3.1 [Kru77]) réactualisé, met en lumière les relations entre les différents modules d'analyse et les différents types de données que nous abordons. Il permet de faire le lien entre les différentes étapes de traitement et d'analyse.

L'étape \circ permet de faire le lien entre une représentation multivariée et une représentation de proximité. Cette relation est bidirectionnelle. Une analyse multidimensionnelle peut être ainsi vue comme une procédure permettant de transformer des données de proximité en données multivariées^[a]. Ceci reste une première approximation.

L'étape δ , quant à elle, fait le lien entre les données de proximité et leurs représentations sous forme de classes. De la même façon cette relation est bidirectionnelle. Cette analyse peut être, par exemple, une analyse par groupement hiérarchique permettant de représenter les données sous la forme d'une structure emboîtée. Le passage de données de classement à une matrice de proximité est une étape indispensable lors de l'analyse d'expériences de classification.

Par contre les étapes λ et \sim sont unidirectionnelles. L'étape \sim regroupe les étapes \circ et δ , agencée de manière séquentielle. Elle permet de passer de données multivariées à une représentation par classes. L'étape λ , contrairement à ce qu'avancait Kruskal, est maintenant largement exploitée. Une des techniques utilisée est basée sur des modèles gaussiens de représentations des classes. Elle permet une grande flexibilité quant à la forme et à la variance de ces classes, voir [JMF99] et [DHS01] pour une introduction.

3.2 Les proximités

3.2.1 Introduction

Une matrice de proximité permet de représenter pour chaque paire de stimuli d'un corpus, la distance, la similarité ou la dissemblance entre ces stimuli. Cette matrice peut être dérivée de trois types de données [Kru77] :

- z les données de proximité δ ;
- z les données de classement λ ;
- z des descripteurs perceptifs ou physiques \circ .

^[a] À plusieurs variables

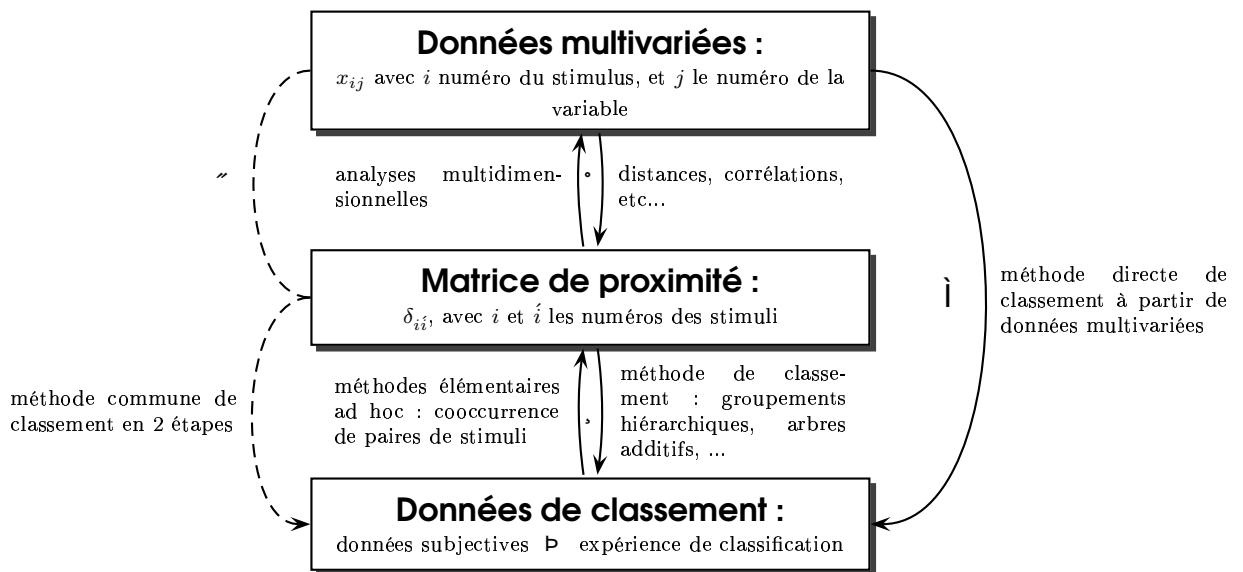


Figure 3.1 - Adaptation du schéma de Kruskal [Kru77]

Suivant le contexte expérimental, la matrice de proximité peut être obtenue directement, comme c'est le cas pour des jugements de similarité. Dans le cas d'une expérience de classification de stimuli, il faut transformer les partitions produites par les participants, en une matrice de proximité reflétant la structure de celles-ci. Mais il faut parfois passer par une étape de calcul des proximités à partir des descriptions physiques ou perceptives des stimuli. Dans ce cas, il existe différents modèles suivant la nature des descripteurs qui peuvent impliquer des processus perceptifs différents.

3.2.2 Jugements de proximité

Suivant certains paradigmes expérimentaux, nous pouvons travailler directement sur une matrice de proximité ([BG97] chap. 6, [SRY81] chap. 2 et 3, pour une introduction dans le domaine du goût et de l'olfaction).

Comparaison par paires P Les participants jugent directement la similarité ou la dissemblance de paires de stimuli, comme dans le cas des études sur le timbre des instruments de musique que nous avons déjà abordées, voir §2.4, p. 16.

Ordonnance P Les participants trient un ensemble de stimuli dans un ordre donné (en fonction d'un critère donné par l'expérimentateur). Pour alléger ce protocole, il est demandé au début d'effectuer deux groupes : le premier correspondant aux paires similaires et l'autre groupe correspondant aux paires dissemblables. Pour chaque groupe, la procé-

dure est la même, tant que la tâche ne devient pas impossible. Cette méthode peut se rapprocher des jugements de triades : trois stimuli $\{i, j, k\}$ sont présentés aux sujets et ils doivent former : 1) soit un couple avec les deux stimuli les plus similaires ; 2) soit deux couples dont l'un sera composé des stimuli les plus similaires et l'autre des plus dissemblables.

Stimulus de référence \triangleright Parmi n stimuli, un est utilisé comme référence, il est demandé aux sujets de juger tous les $(n-1)$ stimuli avec la référence. Chaque stimulus devient à son tour la référence. Par contre, cette méthode demande une analyse particulière pour comparer les jugements qui n'ont pas la même référence.

Différentiel sémantique \triangleright Pour chaque stimulus, les participants font un jugement sur des échelles utilisant des couples d'adjectifs : « précis-imprécis », « brillant-sourd »...

A chaque fois, le but de ces expériences est de découvrir les propriétés sous-jacentes à cette mesure de similarité. L'intérêt de ces expériences est de n'avoir aucun a priori sur les propriétés qui sont mises en jeu (nombre, nature).

3.2.3 Expérience de classification

Lors d'une expérience de classification, il est demandé à des participants de partager un ensemble de stimuli, des sons par exemple, en groupes distincts de sons similaires. Ce paradigme peut être adapté facilement en variant uniquement la consigne donnée aux participants. Il existe d'autres protocoles, où les stimuli peuvent appartenir à deux groupes en même temps.

Cette méthode peut être utilisée même avec des enfants et il a été rapporté que les participants trouvent ce paradigme plaisant [Mil69]. De plus, si un grand nombre de stimuli est étudié, cette méthode peut être plus efficace que des jugements par paires, qui demanderaient un nombre trop conséquent d'essais ([RK75] cité dans [Daw96]).

La matrice de proximité est obtenue de manière semi-directe. Les occurrences des paires de sons regroupées dans une même classe sont calculées. Ainsi, plus une paire de sons aura été placée dans une même classe par les participants, plus ces deux sons seront considérés comme similaires, [Kru77] et [Daw96]. D'un point de vue mathématique, le complémentaire de cette similarité^[a] a des propriétés métriques, sur lesquelles nous reviendrons (§3.3.2.2), qui sont particulièrement pertinentes pour les méthodes d'analyse des proximités [Mil69].

^[a]La fréquence des paires qui n'ont jamais été mises ensembles : la dissemblance.

Le tableau 3.1, exemple tiré de [Daw96], nous montre une simulation de différentes partitions réalisées par deux groupes de sujets sur un ensemble de quatre stimuli $\{A, B, C, D\}$. Les vecteurs de fréquence d'occurrences (F) correspondent aux différentes façons de partager cet ensemble de quatre stimuli (chaque classe est séparée par un tiret). Le nombre de fois que ces patterns ont pu être observés est indiqué pour chaque groupe de sujets. Les cooccurrences de paires et de triplets de stimuli observés sont mentionnées.

Alors que le nombre de cooccurrences de paires est le même, celui des triplets varie entre les deux groupes. Même si nous voyons que nous perdons de l'information, il existe peu de techniques pour tenir compte de toutes les cooccurrences de triplets, de quadruplets... Dans l'idéal, nous pourrions énumérer toutes les partitions possibles d'un ensemble de n stimuli, sans perdre d'information. Malheureusement résumer et représenter cette information est mathématiquement très complexe, et il n'existe aucune technique, à l'heure actuelle, pour représenter ce type de données.

Daws [Daw96] a proposé une nouvelle façon de construire les proximités entre stimuli, afin de prendre en compte des proximités de paires et de triplets. Mais cette technique ne permet pas de visualiser directement cette information et reste très marginale. C'est pourquoi, comme dans toutes les autres études [Maf99] [Guy96] utilisant le paradigme de classification, nous utiliserons l'information sur les cooccurrences de paires de sons.

3.2.3.1 Formalisme et exemple

La dissemblance entre une paire de stimuli $\{i, j\}$ peut s'écrire sous la forme :

$$\delta(i, j) = \begin{cases} 1 & \{i, j\} \text{ n'ont pas été regroupés dans la même classe} \\ 0 & \text{cas contraire} \end{cases} \quad (3.0)$$

Nous donnons ici un exemple d'une expérience de classification libre^[a]. Trois participants virtuels ont classé six sons. Le résultat de ces classifications est donné pour chaque sujet :

Sujet 1 :	Sujet 2 :	Sujet 2 :
Classe 1 P 1 2 3	Classe 1 P 1 3	Classe 1 P 1 2 3
Classe 2 P 4 5 6	Classe 2 P 4 6	Classe 2 P 4 5
	Classe 3 P 2 5	Classe 3 P 6

Ceci donne pour chaque sujet une matrice de dissemblance (matrice d'incidence) entre chaque paire de sons :

^[a]Le nombre de classes n'est pas imposé.

3.2. LES PROXIMITÉS

	Groupe 1		Groupe 2	
Vecteurs de fréquences (F)	ABCD	0	ABCD	0
	ABC-D	5	ABC-D	1
	ABD-C	0	ABD-C	0
	ACD-B	0	ACD-B	0
	A-BCD	1	A-BCD	2
	AB-CD	0	AB-CD	1
	AC-BD	1	AC-BD	2
	AD-BC	0	AD-BC	0
	AB-C-D	1	AB-C-D	4
	AC-B-D	0	AC-B-D	3
	A-BC-D	1	A-BC-D	4
	AD-B-C	0	AD-B-C	0
	A-BD-C	2	A-BD-C	0
	A-B-CD	2	A-B-CD	0
A-B-C-D	5	A-B-C-D	1	
Cooccurrences de paires	AB	6	AB	6
	AC	6	AC	6
	BC	7	BC	7
	AD	0	AD	0
	BD	4	BD	4
	CD	3	CD	3
Cooccurrences de triplets	ABC	5	ABC	1
	ABD	0	ABD	0
	ACD	0	ACD	0
	BCD	1	BCD	2

Tableau 3.1 - Simulation de résultats d'expérience de classification de 4 stimuli {A,B,C,D} pour deux groupes de sujets.

son	Sujet 1						son	Sujet 2						son	Sujet 3					
	1	2	3	4	5	6		1	2	3	4	5	6		1	2	3	4	5	6
1	0	0	0	1	1	1	1	0	1	0	1	1	1	1	0	0	0	1	1	1
2	0	0	0	1	1	1	2	1	0	1	1	0	1	2	0	0	0	1	1	1
3	0	0	0	1	1	1	3	0	1	0	1	1	1	3	0	0	0	1	1	1
4	1	1	1	0	0	0	4	1	1	1	0	1	0	4	1	1	1	0	0	1
5	1	1	1	0	0	0	5	1	0	1	1	0	1	5	1	1	1	0	0	1
6	1	1	1	0	0	0	6	1	1	1	0	1	0	6	1	1	1	1	1	0

La matrice globale de dissemblance correspond à l'agrégation des matrices

individuelles d'incidence. Elle correspond à la moyenne des cooccurrences de paires de sons sur l'ensemble des participants. Dans chaque cellule de la matrice est notée la fréquence avec laquelle deux sons ont été placés dans des groupes différents.

Soit la matrice globale des dissemblances entre paires sur les 3 sujets :

son	Matrice globale					
	1	2	3	4	5	6
1	0.00	0.33	0.00	1.00	1.00	1.00
2	0.33	0.00	0.33	1.00	0.67	1.00
3	0.00	0.33	0.00	1.00	1.00	1.00
4	1.00	1.00	1.00	0.00	0.33	0.33
5	1.00	0.67	1.00	0.33	0.00	0.67
6	1.00	1.00	1.00	0.33	0.67	0.00

$$\text{Matrice globale} = 1/N \sum_N \delta_s(i, j) \quad (3.1)$$

avec N le nombre de sujets et $\delta_s(i, j)$ les matrices individuelles d'incidence.

Cette matrice a des propriétés que nous développerons dans §3.3.2.2, p. 80.

3.2.4 Remarque

Comme nous avons pu le voir sur le schéma de Kruskal, ces différentes analyses peuvent être complémentaires. Par exemple, il peut être intéressant quand on a obtenu des résultats par une méthode « directe », de les comparer avec des analyses basées sur les propriétés physiques des sons afin d'en modéliser les stratégies perceptives.

3.3 Différents modèles de la proximité

3.3.1 Introduction

Les modèles formels de similarité que nous présentons par la suite utilisent explicitement les propriétés perceptives des stimuli mis en comparaison. De ce fait, les propriétés des stimuli ont une influence importante sur le type de processus mis en jeu par les auditeurs, lors de tâches perceptives [Gar78]. Nous devons alors tenter de différencier les types de propriétés permettant de décrire les stimuli du point de vue perceptif.

Bien sûr, il existe une infinité de propriétés qui peuvent être associées à un stimuli. Notre démarche est de spécifier les propriétés des stimuli qui sont pertinentes d'un point de vue psychologique.

3.3.1.1 Typologie des propriétés

Il existe dans la littérature une distinction faite entre différentes propriétés perceptives d'un stimulus. La sémantique liée à la description des différents aspects d'un stimulus, c'est-à-dire de ses propriétés, peut parfois être confuse. Garner [Gar78] a différencié le vocabulaire associé aux propriétés des stimuli. Il propose de distinguer tout d'abord les propriétés « constituantes », des propriétés « globales ».

Les attributs d'un stimulus sont leurs propriétés constituantes, une propriété qui nous permet de définir un stimulus de manière particulière sans qu'elle soit synonyme du stimulus. Les propriétés globales, quant à elles, sont associées à un ensemble d'attributs ou encore à des propriétés émergentes non-dissociables en attributs. Un stimulus possède aussi bien des propriétés constituantes que globales, elles décrivent alors différents aspects d'un même stimulus. Nous nous focaliserons sur les propriétés constituantes car elles rentrent directement en jeu dans les modèles de similarité que nous présentons.

Lors de l'étude de la perception de nos événements sonores, nous tenterons de spécifier les différents types de propriétés mis en jeu dans les tâches expérimentales que nous avons développées.

Les propriétés peuvent être subdivisées en deux classes : les dimensions et les traits (en anglais : « features »).

Dimension \triangleright Une dimension est un attribut d'un stimulus qui peut prendre plusieurs niveaux et existe toujours. Les différents niveaux possibles sont mutuellement exclusifs. Les dimensions quantitatives et qualitatives se distinguent par le rôle du zéro qui est un niveau pour une dimension quantitative et qui correspond à la non-existence d'une dimension pour une dimension qualitative. La sonie ou la brillance sont deux exemples de dimensions.

Trait \triangleright Un trait est un attribut d'un stimulus qui existe ou n'existe pas. Dans le cas de son existence, contrairement à une dimension, le trait n'existe que sur un seul niveau. Le trait peut être traité comme un élément dissociable.

3.3.1.2 Implications

Il existe une différence cruciale entre les traits et les dimensions. L'ensemble des dimensions est connu pour chaque stimulus car il existe toujours une valeur positive associée à chaque dimension. Par contre, l'ensemble des traits d'un stimulus est connu uniquement lorsque ceux-ci existent. Quand

un trait n'existe pas pour un stimulus, il n'y a aucune information dans le stimulus qui indique la potentialité de ce trait [Gar78].

De plus, les propriétés des stimuli n'impliquent pas automatiquement que l'organisme les traite de la même façon, suivant des traits ou des dimensions. Ainsi, nous pouvons construire des stimuli sur la base de dimensions qui seront perçues comme des traits. De plus, il n'existe pas forcément une relation directe entre les dimensions physiques, leurs nombres et les dimensions perceptives d'un même corpus de stimuli [GT82].

Lors de l'analyse des résultats, nous devons tenir compte des propriétés de ces attributs, afin de voir quel modèle de similarité peut être le plus pertinent dans la représentation des processus perceptifs.

A partir des descripteurs perceptifs des sons, il existe différentes façons de calculer la proximité^[a]. A partir de ces distinctions, trois grands types de modèle peuvent être distingués :

Géométriques ou spatiaux \triangleright la proximité perceptive entre les stimuli est représentée par une distance géométrique entre ceux-ci [She62a] [She62b] [Kru78].

En termes de traits ou de caractères \triangleright les similitudes entre stimuli sont exprimées en termes de traits communs ou spécifiques [Tve77].

Type transformationnel \triangleright la similarité psychologique, entre deux stimuli comparés, est représentée par une fonction monotone décroissante. Elle est fonction du nombre d'opérations cognitives qui transforment un stimulus en un stimulus de référence afin qu'ils soient perceptivement identiques [Ima77] cité dans [Gol99].

Nous nous sommes focalisés ici avant tout sur les propriétés perceptives des événements sonores. Un événement sonore comme nous l'avons déjà introduit, voir §2.3.2, peut être décrit suivant ces propriétés physiques et perceptives. Les modèles que nous détaillons par la suite peuvent s'étendre aux propriétés physiques afin de mettre en relation les processus perceptifs de jugement de similarité et leurs propriétés physiques concomitantes.

3.3.2 Approche géométrique

L'analyse théorique des relations de similarité a été nettement dominée par les modèles géométriques [Tve77]. Dans ces modèles, la similarité perceptive entre deux stimuli est représentée par une distance. Généralement, cette approche est associée à la famille des distances de Minkowski qui inclut la distance euclidienne.

^[a]Ce qui est différent du type de modèle de représentation de ces proximités.

3.3.2.1 Calcul des distances

Le calcul de distance dans cette section est orienté pour des valeurs quantitatives continues même si nous introduirons des possibilités offertes pour des descripteurs mixtes (voir sect : 3.3.2.1).

Une fonction non-négative d définie sur toutes paires d'un ensemble de stimuli, est une « distance » ou une « métrique » si, et seulement si, elle suit ces quatre axiomes :

$$\begin{aligned}
 \text{si } i = j & \Leftrightarrow d(i, j) = 0 && (1 : \text{Minimalité}) \\
 \text{si } i \neq j & : d(i, j) > 0 && (2 : \text{Positivité}) \\
 \forall i, j & : d(i, j) = d(j, i) && (3 : \text{Symétrie}) \\
 \forall i, j, k & : d(i, k) \leq d(i, j) + d(j, k) && (4 : \text{Inégalité triangulaire})
 \end{aligned} \tag{3.1}$$

L'inégalité triangulaire, équation 3.3.2.1, dit essentiellement que d'aller d'un point i à un point j est plus court que de faire un détour par le stimulus k . Une distance est dite semi-métrique lorsqu'elle ne satisfait que les axiomes métriques mais pas l'inégalité triangulaire.

■ **Distance de Minkowski** Nous présentons la distance de Minkowski car elle rentre dans des modèles de représentation des proximités que nous présenterons ultérieurement. La distance de Minkowski entre deux stimuli peut être définie comme :

$$d(i, j) = \left[\sum_{f=1}^p |x_{if} - x_{jf}|^r \right]^{1/r}; \tag{3.1}$$

avec $r \geq 1$. Pour $r = 2$, nous retrouvons la distance euclidienne, pour $r = 1$ la distance de Manhattan, « city block »^[a] et pour $r = \infty$, la distance de Chebychev.

De manière générale, une distance peut s'écrire sous la forme :

$$d_{ij} \simeq F\left(\sum_{f=1}^p w_f z_{ijf}\right); \tag{3.1}$$

avec z_{ijf} qui mesure la différence entre les $i^{\text{ième}}$ et $j^{\text{ième}}$ stimuli sur la $f^{\text{ième}}$ variable, F est une fonction linéaire ou monotone. Les valeurs w_f , $\{f = 1, \dots, p\}$, sont des poids associés à chaque variable; généralement, ils sont

^[a] De la même façon que nous marchons d'un bâtiment i à un bâtiment j dans une ville où les rues sont perpendiculaires.

typiquement normalisés à 1. Un poids peut être spécifié pour chaque variable si une d'entre elles est plus saillante que les autres.

Il existe d'autres distances qui ne rentrent pas dans le cadre des distances de Minkowski telles que :

Distance de Mahalanobis \triangleright C'est une distance généralisée qui tient compte de la corrélation entre les descripteurs et qui est indépendante de l'échelle de chaque descripteur. Cette mesure calcule la distance entre deux points dans un espace dont les axes ne sont pas nécessairement orthogonaux, pour tenir compte des corrélations entre descripteurs.

Corrélation \triangleright Matrice de corrélation des résultats inter-individuels.

■ **Distance hétérogène** Lors de l'étude d'un ensemble d'items, il est parfois nécessaire de prendre en compte des descripteurs hétérogènes, afin de calculer la distance entre ces items (exemple : étude d'instruments de musique, date, pays, taille ...). Le mélange de variables de type continu et nominal, contraint à l'utilisation et à la redéfinition de nouvelles distances qui associent ces variables hétérogènes.

Il existe des techniques de calcul de distance, propres à chaque type de variable. Ces différents types de variables peuvent être mélangés : par exemple, des items peuvent être décrits par des variables quantitatives, qualitatives et même binaires. Certains algorithmes permettent aussi de travailler avec des données absentes. Les articles [KR90] [LL98] [Cor71] [WM97] présentent de manière exhaustive différents types de distance qui incluent des variables hétéroclites.

3.3.2.2 Remarques sur les distances

■ **Intégrable/séparable** Comme nous venons de le voir, des stimuli peuvent être décrits dans une structure dimensionnelle par les distances généralisées de Minkowski. La représentation des similarités perceptives entre stimuli par ces distances est dépendante du type de distance associée à ces stimuli. Dans ce contexte nous examinons les relations entre stimuli dans un espace psychologique, voir §3.4.2, p. 85.

La distance euclidienne ($r = 2$, équation 3.3.2.1) est pertinente lorsque les auditeurs jugent la similarité entre les stimuli comme la distance minimum entre ceux-ci, indépendamment des dimensions des stimuli. Pour la distance de Manhattan ($r = 1$, équation 3.3.2.1), les sujets semblent utiliser la somme des distances absolues sur chaque dimension des stimuli. La distance de Manhattan suggère qu'il existe une indépendance perceptive entre les dimensions alors que pour la distance euclidienne, il existe une interdépendance entre ces dimensions.

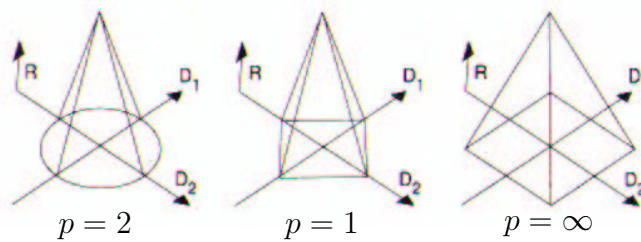
3.3. DIFFÉRENTS MODÈLES DE LA PROXIMITÉ

La distance euclidienne reste invariante par rotation dans l'espace, contrairement à la distance de Manhattan. Cet effet peut être testé perceptivement en manipulant les dimensions.

Les propriétés de ces distances accentuent la différence entre ces deux types de propriété des stimuli.

Garner [Gar74] a ainsi défini les propriétés des dimensions que produit la distance euclidienne comme « intégrables » (perception des couleurs,...) et les propriétés complémentaires liées à la métrique de Manhattan comme « séparables » (taille, variations géométriques ...). Le lecteur pourra se référer à ses travaux [Gar74] sur le paradigme de « classification restreinte » qui lui a permis d'examiner les relations perceptives entre différentes dimensions. Bien sûr, il ne faut pas voir cette séparation comme dichotomique mais plutôt comme un continuum entre dimensions « séparables » et « intégrables ».

Cross [Cro65], cité dans [BG97], a proposé une formalisation assez proche. Les distances euclidiennes sont vues comme un modèle « d'excitation » - figure 3.2 - car un stimulus se différencie d'un autre dans toutes les directions de l'espace psychologique. Dans le cas d'un espace de Manhattan un stimulus se différencie d'un autre par la somme des excitations sur chacune des dimensions. C'est un modèle de « discrimination ». Le modèle de « dominance » fait référence au cas où $r = \infty$, la différence de force de réaction entre deux stimuli est due seulement à la dimension sur laquelle ces deux stimuli diffèrent le plus.



Modèle d'excitation Modèle de discrimination Modèle de dominance

Figure 3.2 - Modèle de Cross dans un espace $2D$, avec $D1$ et $D2$ les dimensions. R fait référence à la force de réaction à un stimulus sur ces deux dimensions, d'après [BG97]

■ **Similarité/dissemblance** Pour exprimer que x et y se ressemblent plus que z et t , nous choisirons un indice de similarité tel que $s(x, y) > s(z, t)$, ou un indice de dissemblance tel que $d(x, y) < d(z, t)$. Il s'agit d'une relation de préordonnance associée à d . L'utilisation du terme de ressemblance fait référence aux expériences de jugements de similarité entre deux stimuli i et j . Nous définissons un indice de dissemblance sur un ensemble I comme une

application d de $I \times I$ dans \mathcal{R}^+ :

$$\begin{aligned} I \times I &\xrightarrow{d} \mathcal{R}^+ \\ (i, j) &\longrightarrow d(i, j) \end{aligned} \quad (3.1)$$

telle que :

$$\begin{aligned} d(i, j) &= d(j, i) \quad \forall (i, j) \in I \times I \\ d(i, i) &= 0 \quad \forall i \in I \end{aligned} \quad (3.1)$$

Dans la suite du chapitre, nous considérons systématiquement les dissemblances ; elles sont employées prioritairement dans les programmes d'analyse. Un indice de similarité s peut toujours être associé à un indice de dissemblance par la relation :

$$d(x, y) = s_{max} - s(x, y) \quad (3.1)$$

avec s_{max} la similarité maximale.

L'indice de dissemblance, tel que nous venons de le définir est peu contraignant, mais il a le désavantage de permettre des incohérences. Par exemple nous pouvons avoir en même temps :

$$\left\{ \begin{array}{l} i \neq j \\ d(i, j) = 0 \\ \text{il existe } k \text{ tel que } d(i, k) \neq d(j, k) \end{array} \right. \quad (3.1)$$

C'est pourquoi les propriétés (1) et/ou (4) de l'équation 3.3.2.1 sont généralement imposées aux indices de dissemblance. Dans le cas de la propriété (1) les dissemblances définissent des « indices de distance », lorsque la propriété (4) est vérifiée ces indices représentent des « écarts ». Naturellement lorsque les propriétés (1) et (4) sont réunies nous retrouvons la définition d'une « distance » [CP76] [DO74].

■ **Groupement et distance** Miller [Mil69] a montré que les données de classement, c'est-à-dire le calcul des cooccurrences de paires, satisfont les critères métriques qui définissent une distance. Une matrice individuelle d'incidence contient 0 pour tous les items mis ensemble et 1 dans le cas contraire. Chaque item est nécessairement mis avec lui même, c'est-à-dire que $d(i, i) = 0$. La matrice est symétrique, $d(i, j) = d(j, i)$. L'inégalité triangulaire est satisfaite puisqu'un participant ne peut pas mettre l'item i avec l'item j et l'item j avec l'item k sans que k soit mis avec l'item i . Ainsi, les matrices individuelles de dissemblance sont des métriques. Lorsque nous sommes toutes les matrices individuelles pour obtenir la matrice globale, nous conservons ces propriétés métriques.

3.3.3 Modèle de contraste

Tversky [Tve77] a proposé une nouvelle formulation de la similarité sur la base d'une description des stimuli en termes de traits. Les stimuli étudiés sont représentés par des collections de traits perceptifs. La similarité entre les stimuli est décrite par un processus d'appariement entre les traits et est explicitée sous la forme d'une combinaison linéaire des mesures des traits communs et distinctifs entre les stimuli.

3.3.3.1 Critiques des modèles métriques

Les modèles d'analyse de la proximité représentent majoritairement les stimuli dans un espace métrique, qu'ils soient des modèles géométriques (dans un espace avec coordonnées) ou les représentations arborées qui impliquent que les données respectent, elles aussi, les axiomes métriques.

Tversky [Tve77] a montré que les axiomes métriques, équation 3.3.2.1, ne sont pas toujours pertinents pour la représentation des jugements de similarité. La représentation dimensionnelle est appropriée pour certains stimuli, par exemple le timbre. Par contre, les traits qualitatifs paraissent plus pertinents que des dimensions quantitative pour la représentation des visages, des pays ...

■ **Critique des axiomes métriques** Nous présentons ici des critiques que Tversky a porté sur les modèles métriques de similarité. Il a développé de manière conséquente ces critiques dans de nombreux articles [Tve77] [TG82]. Nous présentons ici uniquement les plus récurrentes d'entre elles.

Minimalité Elle implique, que la probabilité de juger la similarité d'un stimulus avec lui-même, est égale pour tous les stimuli. Mais il peut être montré, que la probabilité de juger deux stimuli comme « identiques » plutôt que « différents », n'est pas la même suivant les stimuli (ex : objets complexes).

Symétrie La relation de similarité entre deux items n'est pas toujours symétrique. Il existe des données perceptives montrant qu'un item servant de référence apparaît, moins similaire à un item moins saillant que l'inverse. Par exemple l'ellipse est plus similaire à un cercle qu'un cercle à une ellipse, le cercle étant plus prototypique que l'ellipse, voir §2.7.5.2, p. 50.

Inégalité triangulaire Cette relation peut être discutée lorsque le type de jugement de similarité n'est pas explicité et que celui-ci met en jeu des processus de comparaison différents suivant les paires d'items. Par exemple,

lorsque nous jugeons la similarité entre les pays nous pouvons introduire des proximités géométriques (la Jamaïque est similaire à Cuba) vs politique/économique (Cuba est similaire à la Russie^[a]) mais la Jamaïque et la Russie ne sont pas pour autant similaires. Ceci montre que la similarité conceptuelle n'est pas nécessairement transitive.

3.3.3.2 Appariement de traits

Tversky [Tve77] a proposé une alternative aux modèles métriques et multi-dimensionnelles de la similarité en développant une approche théorique basée sur l'appariement de traits. Cette approche permet d'analyser et d'expliquer :

- z le rôle des traits communs et distinctifs,
- z les relations entre les jugements de similarité et de dissemblance,
- z la présence d'asymétries dans les jugements,
- z les effets de contexte (expérimental).

■ **Les traits** Nous considérons un ensemble d'items étudiés : $\Delta = \{a, b, c, \dots\}$. Chaque item dans Δ est caractérisé par un ensemble de traits ou d'attributs, noté A, B, C qui représentent les ensembles de traits associés respectivement aux items a, b, c . Les traits peuvent correspondre à des composantes comme les yeux, la bouche, à des propriétés concrètes comme la couleur ou la taille mais aussi à des propriétés abstraites comme le système politique, le type de culture Il est important de noter que l'ensemble des données associées à un item peut être très riche en contenu et complexe dans sa forme. Par exemple, ces données peuvent être de l'ordre de l'apparence de l'item, de sa fonction, de sa relation avec les autres items, toutes les propriétés qui peuvent être déduites de nos connaissances générales sur le monde. Face à une tâche particulière, il convient de noter que nous sélectionnons dans cet ensemble de données, un sous-ensemble de traits ou d'attributs qui sont pertinents pour effectuer cette tâche. Ceci implique que la représentation d'un item sous forme d'une collection de traits peut être vue comme un processus d'extraction et de collection. Communément le terme de trait fait référence à des descripteurs binaires ou nominaux. Dans le cadre de ce modèle la notion de trait est étendue aux variables ordinales ou cardinales.

■ **Formalisation** La similarité observée entre un item a et un item b , notée $S(a, b)$ est exprimée comme une fonction de leurs traits communs et distinctifs (voir figure 3.3) :

^[a]à la date de l'article [Tve77] !

3.3. DIFFÉRENTS MODÈLES DE LA PROXIMITÉ

- z $A \cap B$ \triangleright les traits communs aux items a et b
- z $A - B$ \triangleright les traits propres à l'item a
- z $B - A$ \triangleright les traits propres à l'item b

La définition de cette fonction s'apparente à *l'appariement*. Cette similarité est traitée comme une mesure ordinale. Si $S(a, b) > S(c, d)$ alors a est plus similaire à b , que l'est c à d . Pour définir une fonction d'appariement, Tversky montre que cette fonction, en plus de la notion d'appariement, doit obéir à l'axiome de monotonie, c'est-à-dire que la similarité s'accroît avec l'addition de traits communs ou avec la suppression de traits distinctifs.

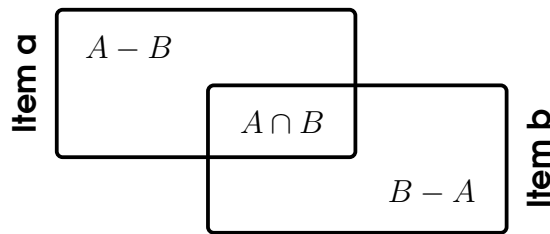


Figure 3.3 - Modèle de contraste, avec a, b les items et A, B les ensembles d'attributs respectifs, $A - B$ les attributs propres à a (respectivement $B - A$ les attributs propres à b) et $A \cap B$ les attributs communs. D'après [Tve77].

Modèle de contraste La similarité perceptive entre deux items a et b peut alors être exprimée sous la forme d'une fonction :

$$S(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \quad (3.1)$$

avec θ, α et $\beta \geq 0$ et f une fonction mesurant la saillance des traits communs ou distinctifs. S et f sont des échelles d'intervalles. Les paramètres θ, α et β dépendent de la tâche ou du contexte. Le modèle de contraste explique la similarité entre les items par une différence pondérée entre les mesures de leurs traits communs et distinctifs. f mesure la contribution des traits (communs ou distinctifs) à la similarité entre les items. Cette contribution est à formaliser suivant les propriétés mises en jeu lors de la tâche.

3.3.4 Discussion sur les modèles

Nous avons présenté ici deux modèles de similarité, l'un basé sur la distance et plus particulièrement sur les distances de Minkowski et l'autre étant le modèle de contraste de Tversky. Les critiques que Tversky [Tve77] a adressé au modèle de distance sont de deux ordres.

La première critique que nous avons déjà introduite est liée aux critères métriques propres à une distance (voir équation 3.3.2.1). Il s'avère que ces critères métriques sont souvent respectés, par exemple, lors d'une expérience de classification ou d'un jugement de similarité. Le respect des critères métriques nous permettra d'utiliser des représentations sous forme spatiale, aussi bien que sous forme d'arbre.

La seconde critique de Tversky & Gati [TG82] [GT82] est liée spécifiquement à l'utilisation des distances de Minkowski que nous retrouvons dans les modèles spatiaux de représentation. Les distances de Minkowski (voir équation 3.3.2.1) impliquent les suppositions d'additivité interdimensionnelle et d'additivité segmentielle, voir [TG82] [GT82] pour une présentation complète. L'additivité interdimensionnelle implique que la dissimilarité entre des items le long d'une dimension, doit être indépendante des valeurs fixes des items sur une autre dimension. L'additivité segmentielle implique, quant à elle, que pour trois items O_i O_j O_k ordonnés sur dimension, la relation entre les dissemblances δ est : $\delta(O_i, O_k) = \delta(O_i, O_j) + \delta(O_j, O_k)$. Gati et Tversky ont montré dans une série d'articles [TG82] [GT82] que les critères propres aux distances de Minkowski ne sont pas toujours respectés et que ce type de distance ne permet pas toujours la représentation des données de proximité psychologique. De plus le modèle de contraste permet de calculer la similarité entre items, sur la base de leurs traits perceptifs et il a été largement utilisé.

Pour représenter ces différents types de similarité, nous abordons dans la suite du document des techniques de représentation des proximités. Ces techniques sont basées sur des modèles spatiaux et sur des modèles sous forme de graphe où les relations entre les items ne sont pas spatiales. Tout comme les modèles de similarité, le choix de la représentation dépend du type des données, nous compléterons ces remarques lorsque nous aurons comparé les différents modèles de représentation des proximités, §3.6, p. 111.

3.4 Représentation des proximités

3.4.1 Introduction

La représentation des proximités peut être schématiquement scindée en deux approches^[a] :

- z Les modèles spatiaux.
- z Les modèles sous forme de graphe.

^[a]Méthodes d'analyse revenant le plus dans la littérature.

Les modèles spatiaux représentent chaque item comme un point dans un espace avec coordonnées. Les distances entre les points reflètent les proximités psychologiques entre les items. C'est le cas, par exemple, des analyses multidimensionnelles.

Les graphes représentent chaque item par un nœud dans une structure de liens connectant ces items, typiquement un arbre (ces notions sont abordées ultérieurement) ou un réseau de neurones. Les relations entre les différents nœuds du graphe reflètent les proximités observées entre les items [Kru77].

3.4.2 Analyses multidimensionnelles

Nous exposons ici une introduction aux analyses multidimensionnelles (MDS pour « MultiDimensional Scaling ») afin de comprendre les spécificités de cette méthode et de pouvoir les confronter aux analyses par représentations arborées. Le lecteur pourra se référer au livre de Borg & Groenen [BG97] pour une présentation détaillée et actualisée.

L'analyse MDS est une méthode qui permet de représenter directement les relations de proximité entre items dans un espace multidimensionnel, avec peu de dimensions pour en permettre l'interprétation. Les relations entre les items, les dimensions perceptives qui sous-tendent leurs représentations, ne sont pas connues a priori ou seulement au travers d'hypothèses. La MDS permet alors de révéler les relations et les structures intrinsèques aux données.

L'essor des analyses MDS est lié à l'effort de modélisation des phénomènes psychologiques. La géométrie des analyses MDS a servi de base aux espaces de représentation mentale ("espace psychologique").

Nous pouvons distinguer 4 types d'utilisation des analyses MDS :

- z Méthode d'analyse exploratoire des données ;
- z Pour tester si certains critères permettant de distinguer des items se reflètent dans les différences de jugement entre ces items ;
- z Pour découvrir les dimensions qui sous-tendent les jugements de similarité ;
- z Modèle psychologique expliquant les jugements de dissemblance en termes de loi basée sur une fonction propre de distance. (Voir §3.3.2.2, p. 78 sur les dimensions intégrables ou séparables).

Les psychophysiciens par exemple, utilisent la MDS comme un modèle qui permet de relier les items dont les propriétés physiques sont connues avec leurs représentations perceptives ou cognitives. Dans ce cas, ils étudient comment les dimensions physiques connues sont représentées psychologiquement. Bien sûr, cette approche tente d'étendre à d'autres stimuli, un espace multidimensionnel, sur un ensemble défini d'items.

Une autre approche, en sociologie, étudie plutôt les corrélations entre des différences de comportement dans un contexte donné. Les types de comportement sont déjà codés et ce sont les relations entre ces comportements qui importent, par exemple, l'étude des attitudes à travers la contestation politique [Lev83] cité dans [BG97].

3.4.2.1 Synopsis d'une analyse

Une analyse MDS tente de représenter les proximités δ_{ij} par des distances entre des points dans une configuration X de m dimensions. Un système de coordonnées est associé à cette configuration spatiale, m axes perpendiculaires liés à une origine O . Ces axes sont des dimensions au sens d'une métrique ayant la même unité. Chaque point est ainsi décrit par un unique m -tuple correspondant aux projections du point sur chacun des m axes. Ce m -tuple est donc un vecteur de coordonnées dans un espace cartésien. La distance d_{ij} entre les points dans cet espace peut être calculée, par exemple, grâce à une distance de Minkowski (équation 3.3.2.1).

L'analyse MDS associe les proximités δ_{ij} avec les distances correspondantes $d_{ij}(X)$ d'un espace MDS, X tel que :

$$\delta_{ij} \xrightarrow{f} d_{ij}(X) \quad (3.1)$$

avec f qui spécifie le modèle de MDS. Le but est de faire correspondre les deux entités δ_{ij} et $d_{ij}(X)$. Cette fonction f peut être de type exponentielle ou logarithmique comme dans le cas de la distance généralisée, telle que Shepard [She87] l'a définie. De manière générale, elle décrit une transformation linéaire à variation lente.

A chaque moment de l'analyse, une erreur est calculée entre le modèle $d_{ij}(X)$ et les proximités δ_{ij} . Cette erreur peut être le « stress » (voir §3.4.5.4, p. 103). Le modèle est élaboré en minimisant cette erreur afin de calculer une géométrie optimale. Le lecteur pourra se référer aux livres de Borg et Groenen [BG97] et Schiffman et coll. [SRY81] pour les aspects calculatoires propres à la MDS. Plusieurs études ont introduit cette méthode et ces applications dans le domaine sonore [Don97] [McA94b].

3.4.2.2 Modèles de distance

Nous avons introduit la distance généralisée (équation 3.3.2.1, p. 77). Dans cette formulation la différence entre deux stimuli sur une variable peut être la distance euclidienne. Le poids introduit dans cette expression permet d'inclure un poids associé à chaque source. C'est le modèle INDSCAL de Carroll et Chang [CC70]. Winsberg & Soete [WDS93] ont entendu le modèle

précèdent en associant des classes latentes de sujets. Chaque sujet appartient à une classe latente, pour laquelle un poids spécifique est introduit pour chaque dimension dans la distance euclidienne pondérée. Il existe aussi des modèles où une dimension spécifique peut être introduite pour chaque item [WC89b].

3.4.3 Représentations arborées

Contrairement à l'approche précédente, nous cherchons à représenter les proximités entre items à l'intérieur d'un réseau de connexions. Les représentations sous forme de graphe tentent de définir les proximités entre items sous forme de relations plutôt que de dimensions perceptives. Ces relations peuvent être des relations d'appartenance à des catégories, des relations de typicalité ou de familiarité entre items (voir § 2). Cette représentation, même si elle se présente dans un plan bidimensionnelle, n'est pas un espace cartésien.

Il existe différents types de graphes avec chacun leurs particularités.

Dendrogramme ▷ Les items sont inscrits dans une structure hiérarchique de classes emboîtées, 3.5. Cette structure peut être construite en partageant un ensemble d'items en sous-ensembles de plus en plus fins. Nous cherchons à associer des propriétés perceptives ou acoustiques permettant de décrire ces sous-ensembles. Ces différents sous-ensembles peuvent représenter des catégories, allant du plus abstrait au plus spécifique, et décrire ainsi différents niveaux d'abstraction, voir §2.7.4.3. Cette représentation peut être un moyen de modéliser l'organisation de la mémoire sémantique.

Arbre additif ▷ Les relations entre items sont décrites par un réseau de chemins uniques entre ces items, figure 3.4. Cette représentation est utilisée dans des domaines très divers. Buneman a été un des premiers à l'introduire, pour représenter les filiations de textes provenant d'une même source et au développement très long comme les textes médiévaux [Bun71], cité dans [Abd90]. En biologie, il permet de comprendre la similarité entre espèces contemporaines, en fonction du nombre de séquences identiques dans la structure de l'ADN. Dans le domaine de la psychologie, il est associé aux travaux de Rosch (§ 2) sur la typicalité (gradient de représentativité des items) ou la familiarité entre items. Des auteurs ont étudié ces concepts dans le cadre de la représentation mentale des sons environnementaux à l'aide de ces représentations [Guy96] [Maf99].

Nous nous intéressons à ces deux représentations, de par leurs propriétés

respectives, associées à la représentation des catégories que nous avons développée dans le chapitre 2.

Avant d'étudier les représentations arborées, il convient d'introduire certaines notions liées à la théorie des graphes [Lop00] [BG88] [Ber73] [Gor96] [DSC96]. Cette théorie nous permet de formaliser mathématiquement les représentations arborées et ainsi de comprendre les particularités de chaque type de représentation.

3.4.3.1 Notion de graphe

Un graphe est un couple $G(V, E)$ déterminé par :

- z Un ensemble $V = \{v_1, v_2, \dots, v_n\}$ dont les éléments sont appelés « sommets » ou « nœuds ».
- z Un ensemble $E = \{(v_i, v_j) | v_i, v_j \in V\}$ qui connecte deux nœuds. Ces éléments sont appelés des « arcs ». Si l'arc n'est pas orienté, le couple $u = (v_i, v_j)$ est appelé une « arête ».

La notion d'orientation porte à confusion car tous les graphes sont orientés. Il convient de différencier l'orientation d'une arête lorsque celle-ci est bidirectionnelle contrairement à un arc qui est unidirectionnel, [Ber73]. Pour un arc $u = (v_i, v_j)$, on dira que v_i et v_j sont les extrémités de l'arc u . L'arc part de v_i et arrive à v_j . Pour une arête, l'orientation de l'arc ne joue aucun rôle. Dans ce cas, le graphe est « non orienté ».

Adjacence \triangleright Deux sommets sont adjacents s'ils sont joints par une arête. Deux arêtes sont adjacentes si elles ont au moins une extrémité commune.

Voisinage \triangleright Le voisinage d'un sommet v_i est l'ensemble des sommets adjacents à v_i .

Degré \triangleright Le degré d'un sommet est le nombre de sommets adjacents à un sommet v_i . Par exemple si un sommet v_i est entouré de 4 sommets adjacents, il est de degré 4. Une « feuille » ou « nœud terminal » est un sommet de degré 1 et un « nœud interne » de degré supérieur à 1.

3.4.3.2 Connexité des graphes

Chaîne et cycle \triangleright Une « chaîne » est une séquence d'arêtes telle que chaque arête possède une extrémité commune avec la suivante :

$$(v_a, v_b), (v_b, v_c), (v_c, v_d) \text{ est une chaîne.}$$

3.4. REPRÉSENTATION DES PROXIMITÉS

Un « cycle » est une chaîne qui contient au moins une arête, telle que toutes les arêtes de la séquence soient différentes et dont les extrémités coïncident :

$$(v_a, v_b), (v_b, v_c), (v_c, v_a) \text{ est un cycle.}$$

Remarque : la notion de chemin et circuit s'utilisent de la même façon que chaîne et cycle, pour un graphe orienté.

Connexe \triangleright Un graphe $G = (V, E)$ est « connexe » si $\forall i, j \in V$ avec $i \neq j$, il existe au moins une chaîne entre i et j .

Longueur du chemin et arbre valué \triangleright Dans un graphe « valué » ou un « réseau », un poids positif $L(v_i, v_j)$ est attribué à chaque arête (v_i, v_j) . Un graphe valué est un couple (G, L) formé d'un graphe $G = (V, E)$ et d'une fonction L , à valeurs strictement positives, définie sur E . $L(v_i, v_j)$ est appelée « longueur » ou « valuation » de l'arête (v_i, v_j) .

Chemin minimum \triangleright Le « chemin minimum » entre n'importe quels nœuds $(v_i, v_j) \in V$, noté $d_{min}(v_i, v_j)$ est défini comme le chemin de longueur minimum qui connecte les nœuds v_i et v_j . Les chemins minimaux satisfont l'inégalité triangulaire et sont symétriques. Si les chemins minimaux sont de longueur nulle pour toutes les paires de nœuds identiques et sont strictement positifs pour toutes les paires de nœuds distinctes, ils constituent une distance (§3.3.2.1).

3.4.3.3 Arbre

Un arbre peut être défini comme un graphe dont chaque paire de nœuds est connectée par un chemin unique. Il n'existe pas de chemin (v_i, v_i) , donc pas de cycle. Les arbres peuvent être valués ou non. Certains sommets peuvent être étiquetés (« sommet réels ») et d'autres pas (« sommet latents »).

X-arbre \triangleright Soit X un ensemble fini. Un X -arbre est un couple (G, f) formé d'un arbre $G = (V, E)$ et d'une fonction f de X dans V telle que pour tout $v \in V - f(X)$, le degré est ≥ 3 .

La fonction f est l'étiquetage du X -arbre. Les sommets dans $f(X)$ sont appelés sommets réels, les sommets dans $V - f(X)$ ^[a] sont appelés les sommets latents. Lorsque $f(X)$ est l'ensemble des feuilles de G , le X -arbre est dit « libre ». Lorsque l'ensemble X décrit un ensemble fini de n items, les feuilles du X -arbre représenteront les items. Les arbres que nous décrivons par la suite appartiennent à cette dernière catégorie.

^[a]L'ensemble des sommets moins ceux étiquetés.

3.4.3.4 Différents types d'arbre

Arbre additif \triangleright C'est un graphe valué non-orienté, connexe et acyclique dont toutes les paires sont connectées par un chemin « unique ». Il n'existe pas de connexion (v_i, v_i) .

Arbres ultramétriques \triangleright C'est un graphe valué orienté où chaque sommet possède un seul précédent sauf un qui n'en a pas : la racine. $\forall v \in V$, il existe un chemin unique de la racine à v . Un arbre « ultramétrique » est un arbre dont chaque nœud terminal est équidistant d'un nœud spécifique : la « racine » de l'arbre. Les poids sont définis au niveau des nœuds et non pas des arêtes.

Groupements additifs \triangleright Cette méthode a été introduite par Shepard et Arabie, [SA79]. Elle est appelée ADCLUS. Si nous considérons la similarité entre deux items S_{ij} , à chaque classe est associé un paramètre b_k . La similarité totale entre deux items i et j est la somme totale de tous les paramètres associés aux classes qui contiennent ces deux items. C'est une méthode simple, mais qui demande une implémentation complexe. Il n'existe que l'implémentation originale en Fortran^[a].

Arbre étendu \triangleright Corter et Tversky, [CT86], ont proposé les arbres étendus EXTREE^[b] pour représenter les traits communs entre items en complément des traits distincts (voir §3.3.3). Dans ce type d'arbre, chaque lien est marqué par un segment qui représente les traits communs d'items qui n'appartiennent pas forcément au même sous-arbre. Les longueurs des segments marqués n'interviennent pas dans le calcul des distances entre deux items.

3.4.3.5 Propriétés des distances d'arbre

Les distances d'arbres ultramétrique ou additif ont des propriétés uniques qui influencent leur lecture.

■ **Distance additive** La distance d'arbre additif satisfait « l'inégalité additive » ou « la condition des 4 points ». Soit d_{ij} une distance d'arbre additif entre les nœuds représentant les items i et j , alors :

$$d_{ij} + d_{kl} \leq \max(d_{ik} + d_{jl}, d_{il} + d_{jk}), \text{ pour tout } i, j, k \text{ et } l \quad (3.1)$$

c'est-à-dire que pour tous items i, j, k et l , les deux plus importantes des 3 sommes $(d_{ij} + d_{kl})$, $(d_{ik} + d_{jl})$, $(d_{il} + d_{jk})$ sont égales. Si un ensemble de

^[a]<http://www.netlib.org>

^[b] Terme original : « extended additive trees »

3.4. REPRÉSENTATION DES PROXIMITÉS

données satisfait la condition des 4 points, un arbre additif unique peut être construit qui représente parfaitement les dissemblances.

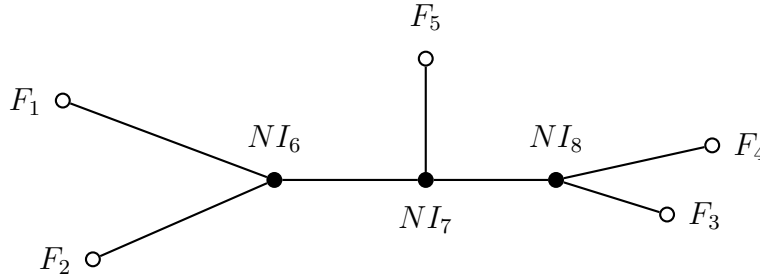


Figure 3.4 - Exemple d'arbre additif. Les nœuds internes sont représentés par des cercles pleins et les feuilles par des cercles vides. F fait référence aux feuilles de l'arbre, NI aux nœuds internes.

Sur la figure 3.4 la distance entre les nœuds F_1 et F_3 est égale à $w(F_1, NI_6) + w(NI_6, NI_7) + w(NI_7, NI_8) + w(NI_8, F_3)$, avec $w(i, j)$ le poids (ou longueur) associé à chaque arête (i, j) .

■ **Distance ultramétrique** Soit d_{ij} une distance d'arbre ultramétrique entre deux nœuds v_i et v_j représentant les items i et j alors :

$$\begin{aligned} d_{ii} &= 0, & d_{ij} &= d_{ji}, & \forall i, j \\ d_{ij} &\leq \max(d_{ik}, d_{kj}), & \forall i, j, \text{ et } k \end{aligned} \quad (3.1)$$

C'est « l'inégalité ultramétrique ». D'une manière moins formelle, nous pouvons dire que pour 3 items i, j et k , les 2 distances inter-items les plus larges, d_{ik} et d_{jk} , sont égales entre elles. Elles forment des triangles isocèles dont la plus petite distance est la base. Si les données d'une matrice de dissemblance satisfont l'inégalité ultramétrique, ces données peuvent être représentées par un arbre ultramétrique unique. Sur la figure 3.5 le sommet R_9 représente la racine. Si l'arbre est ultramétrique nous pouvons écrire que :

$$\begin{aligned} w(F_1, NI_7) + w(NI_7, R_9) &= w(F_2, NI_7) + w(NI_7, R_9) = \\ w(F_3, NI_8) + w(NI_8, R_9) &= w(F_4, NI_6) + w(NI_6, NI_8) + w(NI_8, NI_9) = \\ w(F_5, NI_6) + w(NI_6, NI_8) &+ w(NI_8, R_9). \end{aligned} \quad (3.1)$$

La formulation de cette équation peut porter à confusion car la hauteur de fusion de classes dans un arbre ultramétrique n'est pas associée aux arêtes mais aux nœuds internes. Toutefois, cette écriture permet de montrer l'équidistance des feuilles à la racine.

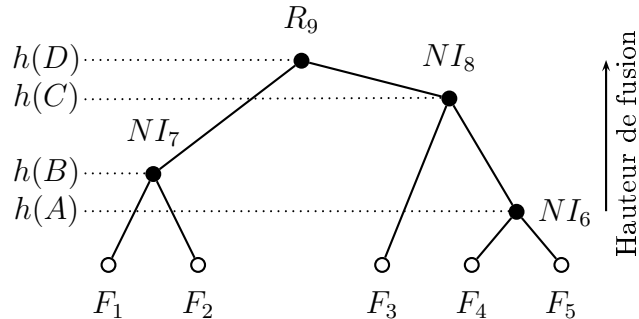


Figure 3.5 - Exemple d'arbre ultramétrique. Les nœuds internes sont représentés par des cercles pleins et les feuilles par des cercles vides. F fait référence aux feuilles de l'arbre, NI aux nœuds internes, R représente la racine de l'arbre. Les hauteurs de fusion des éléments sont mentionnées par h .

■ **Remarques** La condition des quatre points est incluse dans la condition ultramétrique ([CB94], p. 44).

La propriété qui définit une distance ultramétrique est un cas particulier de la distance de Minkowski lorsque dans l'équation 3.3.2.1, $r = \infty$ (p. 504 [CP76]). Holman [Hol72] cité dans [BG88] a montré que toute ultramétrique sur un X -arbre est une distance euclidienne, donc représentable dans un espace de dimension $(n-1)$, avec n le nombre d'items. Si des données satisfont la propriété ultramétrique, une représentation dans un espace euclidien est envisageable, mais devient dans ce cas difficilement interprétable.

3.4.3.6 Topologie

La topologie de l'arbre et les « poids » attachés aux « nœuds internes » d'un arbre ultramétrique sont choisis afin que les distances d'arbres ultramétriques soient le plus proches des dissemblances observées. Une représentation ultramétrique d'une dissemblance définit une représentation hiérarchique de l'ensemble des items.

Dans une représentation d'arbre additif des dissemblances, les items sont représentés par des nœuds terminaux dont la topologie et les « poids », attachés aux « arêtes », sont choisis afin que les distances d'arbre additif soient le plus proches des dissemblances observées.

3.4.4 Arbre additif

Nous décrivons dans cette partie l'algorithme original décrit par Sattah et Tversky [ST77]. Si des données satisfont l'inégalité additive, équation 3.4.3.5, un arbre additif unique peut être construit.

3.4. REPRÉSENTATION DES PROXIMITÉS

L'algorithme, implémenté par Corter [Cor82] ainsi que celui de Barthélemy et Guénoche [BG88], doit résoudre deux problèmes :

- z Construire un arbre avec la structure la plus appropriée aux données de dissemblance.
- z Estimer les longueurs des arêtes.

3.4.4.1 Construction

Dans un arbre additif, un ensemble de quatre items distincts $\{x, y, u, v\}$ peut se retrouver dans une des configurations de la figure 3.6. Nous pouvons exprimer les distances additives correspondantes à chacune des configurations :

$$\begin{aligned}
 d(x, y) + d(u, v) &< d(x, u) + d(y, v) = d(x, v) + d(y, u) & (i) \\
 d(x, v) + d(y, u) &< d(x, u) + d(y, v) = d(x, y) + d(u, v) & (ii) \\
 d(x, u) + d(y, v) &< d(x, y) + d(u, v) = d(x, v) + d(y, u) & (iii)
 \end{aligned}
 \tag{3.1}$$

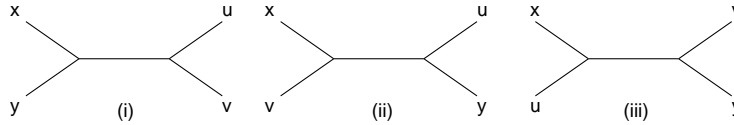


Figure 3.6 - Les trois configurations possibles d'arbre additif pour quatre items $\{x, y, u, v\}$.

Le but est de trouver la meilleure configuration sur la base des dissemblances observées. Pour n'importe quel quadruplet d'items nous pouvons écrire :

$$\delta(x, y) + \delta(u, v) \leq \delta(x, u) + \delta(y, v) \leq \delta(x, v) + \delta(y, u)
 \tag{3.1}$$

Il se trouve que dans ce cas, la meilleure configuration pour représenter ces dissemblances est la configuration (i). Ainsi x et y sont dits voisins si :

$$\delta(x, y) + \delta(u, v) \leq (\delta(x, u) + \delta(y, v), \delta(x, v) + \delta(y, u))
 \tag{3.1}$$

Pour chaque paire d'items $\{x, y\}$, ADDTREE examine tous les items $\{u, v\}$ et compte le nombre de quadruplets pour lesquels x et y sont voisins (score de voisinage). La paire $\{x, y\}$ avec le plus grand score est combinée pour former un nouvel élément z . La nouvelle dissemblance entre z et un autre élément u est égale à :

$$\begin{aligned}
 \frac{\delta(u, x) + \delta(u, y)}{2} & \quad \text{Sattah et Tversky, [ST77]} \\
 \frac{\delta(u, x) + \delta(u, y) - \delta(x, y)}{2} & \quad \text{Barthélemy et Guénoche [BG88]}
 \end{aligned}
 \tag{3.1}$$

3.4.4.2 Estimation

La construction de l'arbre est indépendante de l'estimation des longueurs d'arête. L'estimation des longueurs est faite sur le critère des moindres carrés. Il minimise l'écart entre les dissemblances et les distances d'arbre par l'expression :

$$\sum_{i < j} (\delta_{ij} - d_{ij})^2 \quad (3.1)$$

avec d la distance d'arbre et δ la dissemblance. Un arbre additif a $m \leq 2n - 3$ arêtes pour n items. Pour estimer la longueur des arêtes, l'algorithme doit résoudre l'équation $CX = \delta$, X étant le vecteur des arêtes inconnues. C est une matrice de taille $\binom{n}{2} \times m$ telle que :

$$c_{ij} = \begin{cases} 1 & \text{si la } i^{\text{ième}} \text{ distance d'arbre inclut la } i^{\text{ième}} \text{ arête} \\ 0 & \text{dans le cas contraire} \end{cases} \quad (3.1)$$

La solution aux moindres carrés de l'expression 3.4.4.2 est $X = (C^T C)^{-1} C^T \delta$ et nous donne la longueur des arêtes.

3.4.4.3 Validation

Pour la validation de l'arbre additif, le lecteur se rapportera à l'article de Guénoche et Garreta [GG99] pour le détail des critères métriques et topologiques. Maffiolo [Maf99] et Guyot [Guy96] présentent des travaux utilisant cette représentation dans le domaine de la perception sonore. Les mesures de distortion entre la matrice des dissemblances et la matrice des distances d'arbre sont les mêmes que dans le cas des groupements hiérarchiques. Nous présenterons ces mesures ultérieurement : §3.4.5.4, p. 102.

3.4.5 Arbre hiérarchique

Les classifications sont généralement sous la forme d'un arbre. Il existe aussi des diagrammes sous la forme de lignes étoilées^[a].

De manière générale, il existe un axe représentant les noeuds où ils sont soit des items, soit des noeuds internes (figure 3.5, l'axe horizontale). L'autre axe représente les branches avec la « hauteur » de fusion des items et/ou des classes (figure 3.5, l'axe vertical). La hauteur de fusion est associée aux noeuds internes.

^[a]terminologie anglaise : skyline plot.

Lorsque les feuilles de l'arbre, les nœuds terminaux, sont équidistants de la racine, nous parlons de dendrogramme. Cette condition d'équidistance des feuilles à la racine, implique que les classes soient emboîtées.

Il est important de rappeler qu'une représentation graphique peut être la même entre un arbre additif (tracé de forme hiérarchique) et un arbre hiérarchique. Ces représentations impliquent une lecture des distances arborées totalement différente comme celles que nous avons définies, §3.4.3.5, p. 90.

3.4.5.1 Hauteur de groupement et partition

■ **Hauteur de groupement** Une hauteur h peut être attribuée à chaque nœud interne qui relie deux éléments (les cercles sur la figure 3.5). Un élément peut être aussi bien un item qu'une classe.

Pour toutes parties qui ont en commun des items, par exemple les classes $C = \{F_3, F_4, F_5\}$ et $A = \{F_4, F_5\}$ de la figure 3.5, alors :

$$h(A) \leq h(C) \Leftrightarrow A \subseteq C, \quad (3.1)$$

Dans ce cas les hauteurs de fusion $h(A)$ et $h(C)$ sont respectivement associées au nœuds NI_6 et NI_8 de cette même figure. Cette condition définit une hiérarchie totale « stratifiée ». Elle implique un ordre des nœuds dans la lecture de cet arbre.

La quantité $h(i, j)$ est définie pour toute paire d'items (i, j) comme la hauteur du plus petit sous-ensemble contenant les items i et j . C'est une distance ultramétrique. Si deux éléments i et j sont agrégés, pour un élément z qui n'est pas encore agrégé alors $d(z, i) \leq d(i, j)$ et $d(z, j) \leq d(i, j)$, sinon (z, i) ou (z, j) devraient être agrégé à la place de (i, j) .

La valeur $h(i, j)$ indique la différence de position dans la classification hiérarchique entre les items i et j . Une faible valeur de h , par rapport à la hauteur maximale — $h(D)$ sur la figure 3.5 — montre qu'une paire d'items peut être regardée comme proches d'un point de vue de la similarité.

■ **Partition** Une partition correspond à l'ensemble des classes à un niveau donné de la hiérarchie. Les partitions qui sont compatibles avec une hiérarchie stratifiée sont celles que nous obtenons en coupant l'arbre hiérarchique par des droites horizontales à différentes hauteurs. Par exemple, l'arbre hiérarchique de la figure 3.5, contient une chaîne de 5 partitions notées P_i qui est détaillée sur la figure 3.7.

Dans un arbre hiérarchique, nous nous intéressons généralement à un nombre limité de niveaux qui nous informent sur la structure de nos données de départ. La pertinence du choix des niveaux peut être liée à des critères

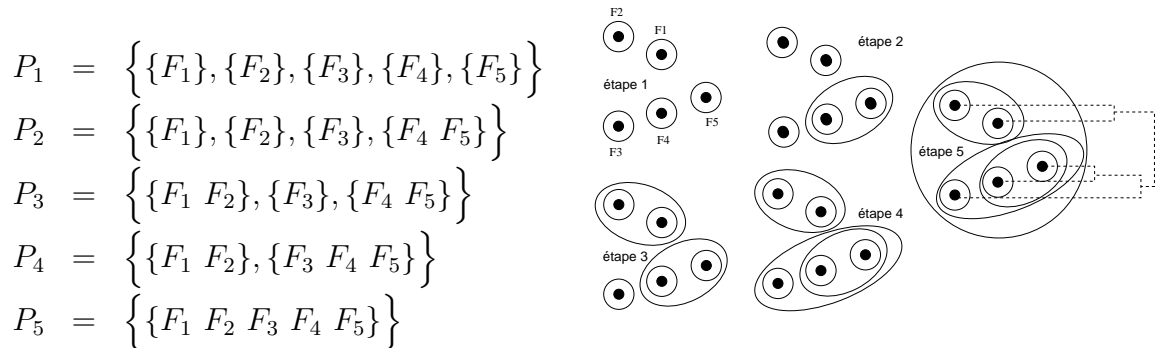


Figure 3.7 - Représentation visuelle de la construction d'un arbre hiérarchique (figure droite) associée à sa chaîne de 5 partitions P_i (figure gauche). L'arbre hiérarchique correspond à la figure 3.5.

internes de bonne représentation des données — voir §3.4.5.4 — mais aussi à des critères externes tels que les connaissances a priori sur l'ensemble de stimuli étudié. Nous aborderons une nouvelle technique pour estimer le niveau de partition le plus « stable » dans le cadre d'expériences de classification, §3.5.2.

■ **Matrice cophénétique** La matrice cophénétique [LL98] reprend la hauteur de fusion h entre deux items devenant membre d'une même classe. Ainsi, cette hauteur peut être définie pendant la formation des classes lors des groupements hiérarchiques. Chaque dendrogramme peut être représenté par une matrice unique dans laquelle la distance, pour chaque paire d'items, est représentée par cette hauteur de fusion. Cette matrice permet de mesurer, entre autre, la distortion entre la matrice de dissemblance et les distances d'arbre hiérarchique, §3.4.5.4. Cette matrice cophénétique est une matrice de distance ultramétrique.

Matrice cophénétique associée à la figure 3.5 :

Matrice cophénétique					
	F_1	F_2	F_3	F_4	F_5
F_1					
F_2	$h(B)$				
F_3	$h(D)$	$h(D)$			
F_4	$h(D)$	$h(D)$	$h(C)$		
F_5	$h(D)$	$h(D)$	$h(C)$	$h(A)$	

avec h la hauteur de fusion des classes. Nous avons omis la demi-matrice

triangulaire supérieure puisque cette matrice définit une distance (équation 3.4.3.5) et est donc symétrique.

3.4.5.2 Algorithmes

Il existe différentes façons d'obtenir des arbres indicés hiérarchiques à partir d'une matrice de dissemblance. Ces méthodes appartiennent au domaine de la classification automatique non supervisée, le nombre de classes n'étant pas imposé a priori. La plupart des méthodes sont séquentielles et consistent en la répétition d'une procédure donnée jusqu'à ce que tous les items soient traités et regroupés. Nous présenterons ici un algorithme de type agglomératif.

■ **Grouperments agglomératifs** Au départ, nous avons n classes : n singletons correspondant aux n items à classer. Le nombre de classes est réduit à chaque itération de l'algorithme, en regroupant la paire d'éléments^[a] les plus similaires. Une nouvelle valeur de distance est associée entre cette nouvelle classe formée et les classes restantes. Nous obtenons ainsi une hiérarchie de $(n - 1)$ niveaux de partition et non pas une partition de m classes arbitraires.

Nous pouvons résumer cette procédure sous la forme d'un algorithme :

- construire les n singletons \mathbf{P} n items ;
- construire la matrice des distances $D \mathbf{P}$ $d(i, j)$ de proximité ; opérer l'agrégation des deux items les plus proches. Première partition à $(n - 1)$ classes ;
- faire :
 - réactualiser les distances entre la nouvelle classe et les éléments restants (équation : 3.4.5.2) pour obtenir une matrice réduite D , en enlevant la paire d'éléments agglomérés et en laissant les autres distances inchangées ;
 - trouver D pour la paire d'éléments la plus proche ;
 - agglomérer ces deux éléments dans une nouvelle classe ;
- tant que : toutes les classes ne sont pas agglomérées dans une seule classe contenant les n items répéter 1 , [Gor96] [CP76] [LMP97].

Il existe différentes façons de recalculer la nouvelle distance suivant la méthode de groupement adoptée. Lorsque deux classes C_i et C_j sont regroupées, la distance entre la nouvelle classe formée $C_i \cup C_j$ et les autres classes

^[a]Items ou classes générés par l'algorithme.

restantes C_k est évaluée de façon récurrente. Elle s'écrit sous la forme :

$$D(C_i \cup C_j, C_k) = \alpha_i D(C_i, C_k) + \alpha_j D(C_j, C_k) + \beta D(C_i, C_j) + \gamma |D(C_i, C_k) - D(C_j, C_k)| + \delta_i h(C_i) + \delta_j h(C_j) + \epsilon h(C_k) \quad (3.1)$$

avec $h(C_i)$ qui représente la hauteur de la classe C_i dans le dendrogramme. Au début de l'itération, les classes sont des singletons :

$$C_i \equiv \{i\}, C_j \equiv \{j\}; i, j = \{1, \dots, n\} \text{ avec } D(C_i, C_j) = d(i, j) \quad (3.1)$$

La figure 3.8 montre les différentes relations de distance entre les classes qui sont introduites dans la formule de Lance-Williams [LW66] [LW67] [Jam78].

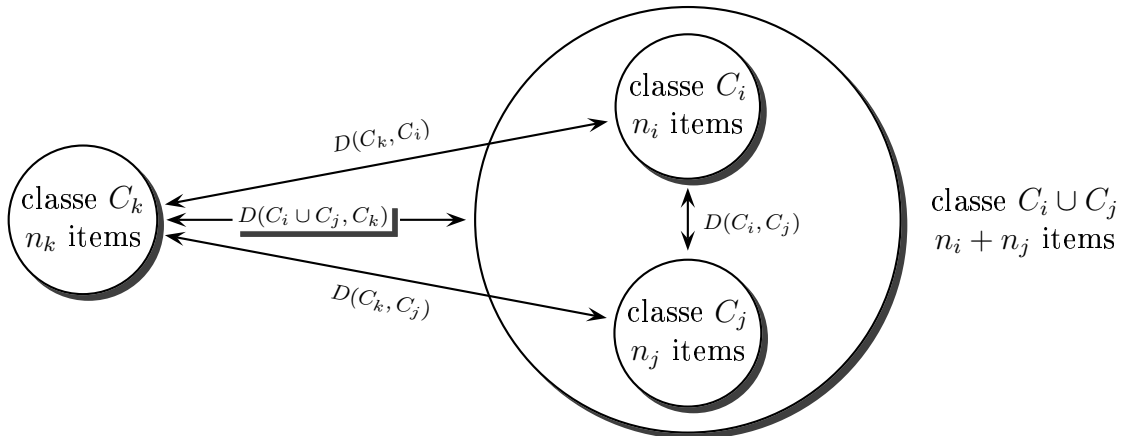


Figure 3.8 - Principe de l'algorithme de groupement agglomératif, avec C_i la classe i et n_i le nombre d'items dans cette classe, [LL98]

Le tableau 3.2 présente les différentes méthodes agglomératives suivant les paramètres utilisés dans la formule de Lance-Williams. Les notations regroupent celles utilisées dans [Gor96] et [LL98] qui reflètent les deux types de notations standard. Les articles de Gordon [Gor87] et [Gor96] font l'inventaire des articles associés à chaque méthode.

3.4.5.3 Choix d'une méthode de groupement

Nous venons de voir que les méthodes agglomératives sont contraintes par le choix des paramètres qui rentrent dans le calcul des nouvelles distances.

Si les données satisfont les conditions ultramétriques, les différentes méthodes donneront les mêmes résultats [Joh67]. Lorsque nous utilisons des données expérimentales, celles-ci ne sont pas parfaitement ultramétriques.

3.4. REPRÉSENTATION DES PROXIMITÉS

Méthode de groupement	Paramètres				
	α_i	β	γ	δ_i	ϵ
Liens simples (single linkage)	$1/2$	0	$-1/2$	0	0
Liens complets (complete linkage)	$1/2$	0	$1/2$	0	0
Association moyenne (UPGMA)	$\frac{n_i}{n_i+n_j}$	0	0	0	0
Poids proportionnels (WPGMA, McQuitty)	$1/2$	0	0	0	0
Moyenne de la dissemblance	$\frac{\binom{n_i+n_k}{2}}{\binom{n_+}{2}}$	$\frac{\binom{n_i+n_j}{2}}{\binom{n_+}{2}}$	0	$-\frac{\binom{n_i}{2}}{\binom{n_+}{2}}$	$\frac{\binom{n_k}{2}}{\binom{n_+}{2}}$
Somme des carrés	$\frac{n_i+n_k}{n_+}$	$\frac{n_i+n_j}{n_+}$	0	$\frac{-n_i}{n_+}$	$\frac{-n_k}{n_+}$
Somme incrémentale des carrés (Ward)	$\frac{n_i+n_k}{n_+}$	$\frac{-n_k}{n_+}$	0	0	0
Centroïde (UPGMC)	$\frac{n_i}{n_i+n_j}$	$\frac{-n_i n_j}{(n_i+n_j)^2}$	0	0	0
Médian (WPGMC)	$1/2$	$-1/4$	0	0	0
Flexible	$1/2(1 - \beta)$	β	0	0	0

Note : n_i est le nombre d'items dans la classe C_i ; $n_+ \equiv n_i + n_j + n_k$

Tableau 3.2 - Méthodes agglomératives suivant les paramètres $\{\alpha_i, \beta, \gamma, \delta_i, \epsilon\}$ de la formule de Lance-Williams (équation 3.4.5.2)

Lorsque nous construisons un arbre hiérarchique, nous transformons l'indice des proximités δ en une distance ultramétrique d_u . En première approche, il est judicieux de chercher la distance ultramétrique la plus proche de δ . Cette approche est malheureusement difficile à mettre en place, même si il existe des tentatives d'optimisation [DSC96] ou [Sri90]. Les méthodes agglomératives que nous venons de voir permettent d'obtenir des arbres hiérarchiques grâce à des stratégies heuristiques plus « simples » à mettre en œuvre.

Les différences entre les résultats obtenus par ces analyses, peuvent parfois dérouter. Nous avons vu précédemment un algorithme général tel que nous pouvons le rencontrer dans des applications statistiques. Nous abordons maintenant les méthodes telles qu'elles ont été proposées, afin de mettre en lumière les particularités de chaque méthode qui déterminent la morphologie des arbres.

■ **Groupements par liens simples** Cette procédure est une des premières avec les groupements par « liens complets » à avoir été proposée [Joh67]. Comme pour tous les algorithmes le but est de regrouper à chaque étape les éléments (classe ou item) les plus proches jusqu'à ce que tous les éléments soient regroupés. La nouvelle distance associée à cette fusion est choisie par le critère de distance minimale entre deux items i et j appartenant à chacune

des classes considérées A et B comme le montre la figure 3.9 (b) :

$$d_{min}(A, B) = \min_{\substack{i \in A \\ j \in B}} d(i, j) \quad (3.1)$$

Cette méthode s'appelle « groupements par liens simples »^[a] car à chaque phase de groupement les deux classes sont liées par le lien le plus court entre deux items propres à chaque classe. Dans cette méthode, chaque item dans une classe est plus similaire à tous les items de la classe qu'aux items qui ne sont pas dans la classe.

Du fait de sa formulation, elle est aussi appelée méthode du « plus proche voisin »^[b]. Elle ne demande aucun recalcul numérique des distances et est basée uniquement sur les liens entre items.

Cette méthode conduit souvent à regrouper dans la même classe des éléments qui sont très éloignés lorsque nous regardons les proximités de départ et provoque ainsi un « effet de chaîne » lorsqu'il existe des items intermédiaires entre des classes. Cette propriété rend la méthode intéressante pour mettre en évidence des gradients tels que l'ordination. Cette procédure de groupement est sensible aux bruits dans les données [Mil96] et [Cor71]. Le bruit peut être associé à des différences interindividuelles dans l'organisation des partitions lors d'expériences de classification. C'est pourquoi il est préférable d'utiliser une méthode plus robuste à ces différences puisque nous cherchons avant tout une représentation consensuelle des données de classification.

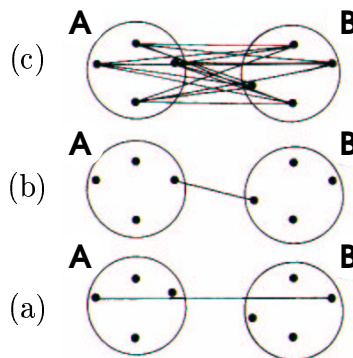


Figure 3.9 - Calcul des nouvelles distances pour (a) « Liens complets », (b) « Liens simples », (c) « groupements moyens » entre deux classes A et B . A et B représentent les classes à regroupées. [KR90].

^[a]«Single linkage»

^[b]« Nearest neighbour. »

■ **Groupements par liens complets** Contrairement à la méthode précédente, la nouvelle distance associée à la fusion entre deux classes A et B est estimée en prenant la distance maximale entre deux items i et j , appartenant respectivement aux deux classes A et B , voir figure 3.9 (a), telle que :

$$d_{max}(A, B) = \max_{\substack{i \in A \\ j \in B}} d(i, j) \quad (3.1)$$

Au contraire de la méthode précédente, celle-ci tend à produire des classes plus compactes et homogènes. Lorsqu'une classe tend à grandir, il devient de plus en plus difficile pour les nouveaux items de la joindre, puisque l'appartenance à la classe se fait à l'unanimité [LL98].

Lance et Williams [LW67] parlent de dilatation de l'espace des items contrairement à la méthode précédente qui le contracte. Cette stratégie peut être pertinente dans le cas où nous voulons contraster les discontinuités dans un ensemble d'items.

■ **Association moyenne** A chaque pas de l'algorithme, nous cherchons les deux éléments les plus proches pour les joindre dans une même classe. Le calcul des nouvelles distances est simple. Soient deux classes A et B avec n_a et n_b leur nombre d'items respectifs. La distance entre les deux classes est définie par la moyenne de toutes les distances entre les items des deux classes. Nous pouvons l'écrire sous cette forme :

$$d(A, B) = \frac{1}{n_a n_b} \sum_{\substack{i \in A \\ j \in B}} d(i, j) \quad (3.1)$$

voir figure 3.9 schéma (c).

Contrairement aux méthodes précédentes qui conservent en mémoire les dissemblances de départ, cette méthode met à jour et réduit la matrice des distances à chaque pas. A chaque fusion d'éléments (classe ou items), ceux-ci sont remplacés par une seule entité dans la matrice.

Du fait du recalcul des distances à chaque pas de l'algorithme, il faut extraire indirectement une matrice cophénétique en suivant le processus agglomératif pour représenter les niveaux de fusion entre chaque paire d'items, §3.4.5.1. La dissemblance entre groupes qui se joignent progresse de manière monotone.

■ **Variantes** Il existe des variantes du groupement moyen. La première est le groupement à poids proportionnels qui inclut une pondération dans le cas où des sous-ensembles d'items de tailles différentes sont perceptivement plus

homogènes. Dans ce cas, le calcul des nouvelles distances est pondéré par la taille des groupes. Les méthodes de groupement centroïde et median sont basées sur le calcul des centroïdes des classes utilisant une distance euclidienne. La méthode de groupement médian inclut une pondération comme la méthode à poids proportionnels [KR90]. Ces méthodes conservent l'espace de référence.

La méthode de Ward [War63] est reliée aux méthodes « groupement centroïde » et « groupement median » car elle est associée à une représentation géométrique liée aux centres de gravité des classes. Elle semble tout à fait pertinente pour des données basées sur des échelles d'intervalles ou explicitement extraites d'un espace euclidien. Le but est de minimiser à chaque étape la variance à l'intérieur de chaque classe et de maximiser la variance entre les classes. Le lecteur pourra trouver une revue détaillée de ces différentes méthodes dans le livre de Legendre & Legendre [LL98].

■ **Choix de la méthode** Des études par simulation numérique ont montré que la méthode par association moyenne semble la plus robuste dans toutes les situations. Ces études partent de données générées dont la structure de groupes est choisie, auxquelles du bruit est ajouté pour tester la robustesse des méthodes. Le lecteur pourra se référer aux travaux de Milligan [Mil80] [Mil96] [CM96], sur la relation entre les erreurs ajoutées dans les données et les algorithmes de groupement. Ces résultats sont confirmés par des critères donnés par Kaufman et Rousseeuw [KR90] pour sélectionner une méthode de groupement.

La méthode Ward peut être pertinente si la matrice des distances de départ est clairement un espace euclidien.

Notre but est d'analyser des résultats de classification libre afin de trouver la partition optimale représentant la structure des données. Nous n'avons aucun a priori quant à une représentation géométrique des données. De plus Breckenridge, [Bre90a] dans [Mil96] a montré que le groupement par association moyenne a des bonnes propriétés de réplification.

C'est pourquoi la méthode de groupement selon l'association moyenne semble pertinente pour nos analyses.

3.4.5.4 Validation

Dans les sections précédentes nous avons tenté de présenter différentes approches de l'analyse par groupements hiérarchiques. Milligan [Mil96], dégage les différentes étapes cruciales dans l'analyse de groupement. Nous reprenons les trois dernières étapes propres à l'analyse :

- z Le choix de la méthode de groupement.

- z Le nombre de groupes sélectionnés représentant la partition optimale dans l'arbre hiérarchique.
- z L'interprétation, le test et la réplication des résultats.

Nous nous intéressons maintenant aux deux dernières étapes et présentons différents critères qui semblent montrer un consensus entre les auteurs afin de tester :

- z La relation entre proximités et distances ultramétriques associées.
- z Le choix de la partition optimale ou du nombre de groupes optimal.

■ **Mesures de distortion** Des données peuvent être représentées de manière adéquate par une classification hiérarchique s'il existe une faible distortion dans la transformation des proximités de départ en distances ultramétriques [Gor87]. Nous résumons ici des mesures de distortion.

Corrélation cophénétique C'est le coefficient de corrélation de Pearson calculé entre la matrice cophénétiq

ue, §3.4.5.1, et la matrice originale des proximités [LL98] [Gor96] [Cor71]. Cette corrélation donne une mesure de la relation linéaire entre ces deux matrices.

Il peut être préférable d'utiliser des coefficients de corrélation basés sur les rangs, comme le coefficient de corrélation de Kendall ou le coefficient de corrélation de Spearman, si l'intérêt réside plus dans la structure géométrique du dendrogramme que dans la longueur des branches [LL98].

Il n'existe pas de test de significativité, puisque la matrice cophénétiq

ue n'est pas indépendante de la matrice de dissemblance. Pour calculer la corrélation entre deux matrices indépendantes, le lecteur pourra se référer aux tests de Mantel qui utilise les permutations [LL98].

Distance de Gower Ce critère est retrouvé systématiquement dans les algorithmes d'optimisation de distances ultramétriques. Il correspond au stress brut de Kruskal. Il est utilisé comme mesure de concordance dans les analyses multidimensionnelles non-métriques. Il varie sur $[0, \infty)$. Une faible valeur indique une bonne correspondance.

$$\sigma_r = \sum_{i < j} [d_u(i, j) - \delta(i, j)]^2 \quad (3.1)$$

avec d_u la distance ultramétrique et $\delta(i, j)$ les dissemblances de départ entre un item i et un item j .

Stress de Kruskal Il existe deux versions du stress de Kruskal [Kru64]
Stress-1 :

$$\sigma_1^2 = \frac{\sum_{i < j} [d_u(i, j) - \delta(i, j)]^2}{\sum_{i < j} d_u^2(i, j)} \quad (3.1)$$

Stress-2 :

$$\sigma_2^2 = \frac{\sum_{i < j} [d_u(i, j) - \delta(i, j)]^2}{\sum_{i < j} [d_u^2(i, j) - \bar{d}_u]^2} \quad (3.1)$$

avec \bar{d}_u la distance ultramétrique et $\delta(i, j)$ les dissemblances de départ entre un item i et un item j et \bar{d}_u la moyenne des distances ultramétriques.

■ **Partition optimale** Gordon [Gor87] fait remarquer que le nombre de niveaux dans un arbre ($n - 1$, pour n items) peut être confus pour l'interprétation. En effet, une partition avec des classes trop importantes donnent peu d'information quant à l'organisation perceptive des items. Inversement, une partition constituée de petits groupes donnent trop de détails. Généralement nous ne nous intéressons qu'à un petit nombre de niveaux pour interpréter les résultats [Mil69].

De ce fait, il est intéressant de pouvoir dégager une partition optimale, c'est-à-dire de déterminer le nombre de groupes pertinents dans un arbre hiérarchique. Il existe différentes stratégies qui sont présentées dans l'article de Milligan et Cooper [MC85]. Nous ne présenterons qu'une partie de ces critères en sélectionnant ceux que nous retrouvons le plus dans la littérature. Ces critères ne sont pas absolus. Ils permettent d'appuyer et d'aider le choix d'une partition.

Indices de niveau C'est un critère très simple qui examine la progression des niveaux de fusion entre les classes lors de groupements agglomératifs, [MC85]. S'il existe une structure de classe avec un fort contraste, nous devons observer au début des hauteurs de fusion très rapprochées qui impliquent des similarités fortes, puis une augmentation brusque du niveau de fusion. Les classes se rejoignent à niveau de distance élevé. La rupture dans la pente de progression des niveaux de fusion peut nous donner une première estimation pour couper l'arbre hiérarchique, afin de trouver une partition optimale.

Corrélation bisériale de point A chaque niveau de la hiérarchie, nous regardons la partition correspondante. Pour une partition donnée nous codons une matrice cophénétique binaire : lorsqu'une paire d'items est regroupée dans la même classe $d_u(ij) = 1$, et 0 dans le cas contraire. Pour chaque partition correspondant à un niveau donné, nous renouvelons cette procédure. Le coefficient de corrélation bisériale de point est calculé entre chaque

matrice cophénétique dérivée d'une partition et la matrice de dissemblance de départ. La partition, dans la hiérarchie qui maximise la corrélation avec les données de départ, peut être considérée comme la partition localement optimale [Mil81] [LL98] [SS93] [LL94].

3.5 Comparaison des partitions

Le but de cette section est de présenter un critère, l'indice de Rand, qui permet de mesurer la correspondance entre deux partitions, c'est-à-dire entre deux résultats d'une classification sur un même ensemble d'items.

Premièrement, ce critère nous permet de comparer le résultat de chaque participant d'une même expérience de classification. Ainsi, nous pourrions obtenir une mesure de concordance inter-participants pour une expérience de classification. Nous pouvons alors effectuer une analyse par groupements hiérarchiques, afin de trouver des groupes de participants qui répondent de manière similaire ou bien écarter des personnes répondant très différemment.

De plus, ce critère nous permet de développer une technique pour trouver le niveau hiérarchique dans une analyse par groupements hiérarchiques qui représente le mieux la partition moyenne entre les participants.

3.5.1 Indice de Rand

Soit un ensemble $S\{O_1, \dots, O_i, \dots, O_n\}$ de n items avec O_i l'item i . $U = \{u_1, \dots, u_R\}$ et $V = \{v_1, \dots, v_C\}$ représentent deux partitions différentes de l'ensemble S . Comme dans le cas de deux sujets qui partagent un même ensemble d'items. Les parties de U sont des sous-ensembles^[a] de S tel que l'union de tous les u_i soit égale à l'ensemble S . Le raisonnement est le même pour V .

Soit n_{ij} le nombre d'items qui sont communs aux classes u_i et v_j . L'information sur le recouvrement des classes entre les différentes partitions U et V peut être écrite sous la forme d'une table de contingence. Les nombres $n_{i.}$ et $n_{.j}$ font référence respectivement au nombre d'items dans les classes u_i (ligne i) et v_j (colonne j) comme dans le tableau 3.3.

^[a] Comme nous l'avons vu dans l'exemple du tableau 3.1.

		Partition V						
Classes		v_1	v_2	\cdot	v_j	\cdot	v_C	Sommes
Partition U	u_1	n_{11}	n_{12}	\cdot	n_{1j}	\cdot	n_{1C}	$n_{1\cdot}$
	u_2	n_{21}	n_{22}	\cdot	n_{2j}	\cdot	n_{2C}	$n_{2\cdot}$
	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
	u_i	n_{i1}	n_{i2}	\cdot	n_{ij}	\cdot	n_{iC}	$n_{i\cdot}$
	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
	u_R	n_{R1}	n_{R2}	\cdot	\cdot	\cdot	n_{RC}	$n_{R\cdot}$
Sommes		$n_{\cdot 1}$	$n_{\cdot 2}$	\cdot	$n_{\cdot j}$	\cdot	$n_{\cdot C}$	$n_{\cdot\cdot} = n$

Tableau 3.3 - Notation pour comparer deux partitions

La mesure de correspondance entre les partitions U et V est basée sur la façon dont les paires d'items sont classées dans ce tableau de contingences. Quatre types de classement de ces paires d'items, parmi les $\binom{n}{2}$, ont été recensés^[a], [HA85] [KG99] [SS96] [DL00] :

type (a) \mathfrak{P} Les items de la paire considérée sont placés dans la même classe dans U et puis dans V .

type (b) \mathfrak{P} Les items de la paire sont placés dans la même classe dans U et dans des classes différentes dans V .

type (c) \mathfrak{P} Les items de la paire sont placés dans des classes différentes dans U et dans la même classe dans V .

type (d) \mathfrak{P} Les items de la paire sont placés dans des classes différentes dans U puis dans V .

A partir du tableau des contingences nous pouvons exprimer (a),(b),(c) et (d) en fonction de n , le nombre d'items, n_i , n_j et n_{ij} , [HA85] :

$$\begin{aligned}
\text{type (a) : } & \frac{1}{2} \sum_{i=1}^R \sum_{j=1}^C n_{ij}(n_{ij} - 1) \\
\text{type (b) : } & \frac{1}{2} (n^2 + \sum_{i=1}^R \sum_{j=1}^C n_{ij}^2 - (\sum_{i=1}^R n_i^2 + \sum_{j=1}^C n_j^2)) \\
\text{type (c) : } & \frac{1}{2} (\sum_{j=1}^C n_j^2 - \sum_{i=1}^R \sum_{j=1}^C n_{ij}^2) \\
\text{type (d) : } & \frac{1}{2} (\sum_{i=1}^R n_i^2 - \sum_{i=1}^R \sum_{j=1}^C n_{ij}^2)
\end{aligned} \tag{3.1}$$

Les types (a) et (d) peuvent être interprétés comme un accord dans la classification d'une paire d'items et les types (b) et (c) représentent les désaccords entre les deux partitions.

^[a] Attention : notations conformes à l'article [KG99].

L'indice de Rand R_d [Ran71] qui calcule l'agrément entre deux partitions s'écrit :

$$R_d = [(a) + (d)]/\binom{n}{2} = \left\{ \binom{n}{2} + \sum_{i=1}^R \sum_{j=1}^C n_{ij}^2 - \frac{1}{2} \left(\sum_{i=1}^R n_i^2 + \sum_{j=1}^C n_j^2 \right) \right\} / \binom{n}{2} \quad (3.1)$$

avec $\binom{n}{2}$ = le coefficient binomial d'indice n .

Lorsqu'il existe une cohérence parfaite entre les deux partitions l'indice de « Rand » est égal à 1.0 et lorsqu'il existe une indépendance totale entre les deux partitions, il tend vers 0.

Hubert et Arabie [HA85] ont proposé une version ajustée de l'indice de Rand car celui-ci n'est pas forcément nul même lorsque deux partitions comparées sont construites de manière aléatoire. L'indice de Rand ajusté R_a est calculé sur la base de l'indice de Rand standard, auquel on soustrait un indice de Rand calculé sur deux partitions générées aléatoirement ayant les contraintes suivantes : le même nombre de classes et le même nombre d'items dans chaque classe que dans les deux partitions comparées.

L'indice de Rand ajusté, R_a , corrigé par le hasard, s'écrit sous sa forme générale :

$$R_a = \frac{\sum_{i=1}^R \sum_{j=1}^C \binom{n_{ij}}{2} - \sum_{i=1}^R \binom{n_i}{2} \sum_{j=1}^C \binom{n_j}{2} / \binom{n_{ij}}{2}}{1/2[\sum_{i=1}^R \binom{n_i}{2} + \sum_{j=1}^C \binom{n_j}{2}] - \sum_{i=1}^R \binom{n_i}{2} \sum_{j=1}^C \binom{n_j}{2} / \binom{n_{ij}}{2}} \quad (3.1)$$

Les notions sont les mêmes que pour l'équation 3.5.1 et cet indice est également borné entre 0 et 1.

Milligan [Mil96] a testé différents indices pour évaluer l'agrément entre deux partitions dérivées d'analyses de groupement, avec des nombres différents de classes. L'indice de Rand ajusté donne d'excellents résultats et est largement adopté.

3.5.2 Partition optimale : une nouvelle approche

Nous avons précédemment décrit des critères pour nous guider dans le choix de la partition optimale, voir §3.4.5.4. Ces critères sont directement liés à l'exploitation de la structure de l'arbre hiérarchique et/ou de la matrice de proximité de départ.

Dans le cadre d'expériences de classification d'items, nous observons une variabilité intersujets. Ces différences peuvent résider dans le nombre de classes effectuées par les participants et/ou les items classés dans ces classes. Nous devons donc déterminer la partition de classes qui semble être la plus commune entre les sujets.

Il n'existe pas de méthode directe d'estimation de la partition optimale pour des données de classification. Une estimation de la partition optimale directement extraite des données de classement de différents participants, est un problème calculatoire extrêmement complexe. C'est pourquoi dans le cadre de cette thèse, nous avons développé une méthode originale qui combine l'utilisation des arbres hiérarchiques et des techniques de bootstrap pour estimer la stabilité d'un test statistique, l'indice de Rand ajusté (§3.5).

3.5.2.1 Introduction au bootstrap

Le bootstrap est une technique qui permet d'effectuer des inférences statistiques [ET93]. Le bootstrap repose sur la simulation de données à partir d'un nombre limité d'observations. Il permet d'estimer l'erreur standard d'un test statistique lorsqu'il n'existe aucun a priori sur la distribution des observations. Généralement, pour des tests statistiques dont la distribution des observations est connue, par exemple normale ou gaussienne, nous pouvons estimer les seuils significatifs pour ces tests, ce qui est impossible directement dans le cas où la distribution des observations est inconnue.

■ **Présentation** Le principe est de partir d'un échantillon de N observations, $\mathbf{x} = (x_1, x_2, \dots, x_N)$, pour lequel nous appliquons un test statistique $s(\mathbf{x})$, par exemple la moyenne ou la médiane. Le but est de créer B échantillons bootstrap, $\mathbf{x}_b^* = (x_1^*, x_2^*, \dots, x_N^*)$, avec b le numéro de l'échantillon bootstrap. Chaque \mathbf{x}_b^* est constitué en tirant avec remise N valeurs parmi l'échantillon observé \mathbf{x} de départ. Un échantillon bootstrap peut être par exemple : $\mathbf{x}^* = (x_3, x_1, x_3, x_8, \dots, x_5)$, ainsi certaines valeurs apparaissent zéro fois, d'autres une fois, d'autres deux fois, Pour chaque échantillon bootstrap \mathbf{x}_b^* , nous calculons une statistique $s(\mathbf{x}_b^*)$. Pour estimer l'erreur standard, nous calculons tout d'abord la moyenne de la statistique, $\bar{s}_{boot} = \sum_{b=1}^B s(\mathbf{x}_b^*)/B$, sur l'ensemble des valeurs $s(\mathbf{x}_b^*)$ estimées sur les échantillons bootstrap. L'écart type des répliques bootstrap $\sigma(s_{boot}) = \{\sum_{b=1}^B (s(\mathbf{x}_b^*) - \bar{s}_{boot})^2 / (B-1)\}^{1/2}$, permet alors d'évaluer l'erreur standard du test statistique sur l'échantillon de départ.

Le nombre de bootstraps est empirique. Généralement 50 tirages donnent un début d'information, $B=200$ permet d'estimer l'erreur-standard et $B=500$ permet d'évaluer les intervalles de confiance [ET93].

3.5.2.2 Application du bootstrap

La partition optimale correspond aux classes les plus récurrentes entre les classifications des participants. Nous avons appliqué la technique du boot-

strap pour estimer la partition optimale sur un arbre hiérarchique de référence. Cet arbre représente les données de cooccurrences sur l'ensemble des participants.

Le but est de calculer l'indice de Rand ajusté entre un niveau de partition de l'arbre hiérarchique de référence et un niveau équivalent de partition d'un arbre hiérarchique bootstrap, voir figure 3.7 p. 96, pour les niveaux de partition. Cet arbre hiérarchique bootstrap est créé à partir des données originales ré-échantillonnées avec remise. L'indice de Rand ajusté est calculé à différents niveaux équivalents de partition, c'est-à-dire pour un niveau de 2 classes, 3 classes, Cette procédure est répétée pour chaque échantillon bootstrap. Si des classes apparaissent comme stables à travers les participants, nous devons trouver un niveau où les classes de l'arbre de référence sont concordantes avec les classes des arbres bootstrap à un niveau donné.

Par ce moyen, nous essayons de trouver le niveau hiérarchique et sa partition optimale associée, les plus stables parmi l'ensemble des participants. La partition optimale est estimée comme le niveau où l'indice de Rand ajusté est le plus élevé en moyenne sur l'ensemble des tirages bootstrap et dont l'écart-type des répliques bootstrap est le plus faible.

Il nous a semblé judicieux de définir le même niveau de partition entre un arbre de référence et un arbre bootstrap par le même nombre de classes dans chacune des partitions (référence et bootstrap). Ce critère est lié au but recherché puisque nous cherchons à trouver les classes communes sur l'ensemble des participants.

Nous présentons en détail les différentes étapes de l'estimation de la partition optimale qui sont résumées sur les figures 3.11, p. 114 et 3.12, p. 115.

■ **Matrices d'incidence** Tout d'abord chaque classification effectuée par chacun des N participants est transformée en une matrice individuelle d'incidence (nombre de fois que deux sons ont été mis ensemble). C'est notre échantillon d'observations de départ correspondant aux N matrices d'incidence.

■ **Arbre de référence et partitions associées** Dans un premier temps nous construisons la matrice de cooccurrences sur l'ensemble des N participants à partir des matrices d'incidence. Nous effectuons une analyse par groupements hiérarchiques à partir de la matrice de cooccurrences, c'est l'arbre de référence. Nous associons à cet arbre l'ensemble des partitions, (2 classes, 3 classes, ...) comme dans l'exemple de la figure 3.7 p. 96, correspondant aux différents niveaux d'emboîtement des classes.

■ **Bootstrap** A partir des N matrices d'incidence de départ, nous effectuons B tirages aléatoires avec remise pour obtenir B échantillons bootstrap de matrices d'incidence. Pour chaque échantillon bootstrap de matrices d'incidences, nous associons une matrice de cooccurrences bootstrap. A chaque tirage nous créons un arbre hiérarchique bootstrap à partir de la matrice de cooccurrences bootstrap, avec l'ensemble des partitions associées (2 classes, 3 classes, ...), voir figure 3.11.

Pour chaque tirage nous calculons les indices Rand ajustés entre les mêmes niveaux de partition de l'arbre de référence et les niveaux de partition de l'arbre bootstrap.

A la fin du bootstrap, nous avons donc pour chaque niveau de partition (2 classes, 3 classes, ..., n classes) la distribution des indices de Rand ajustés entre les partitions de l'arbre de référence et les partitions des B arbres bootstrap.

■ **Partition optimale** Dans un premier temps, nous calculons la moyenne \bar{R}_a sur la distribution bootstrap des indices de Rand ajustés, pour chaque niveau de partition (2 classes, 3 classes, ...). De la même façon, nous calculons l'écart type $\sigma(R_a)$ sur la distribution bootstrap des indices de Rand ajustés. La figure 3.10 représente l'évolution de \bar{R}_a et $\sigma(R_a)$ pour des partitions de 2 à 15 classes sur un exemple de données du chapitre 5. Sur cette même figure, nous confrontons ce résultat avec un calcul des corrélations bisérialles entre la matrice cophénétique (calculée à différents niveaux de hiérarchie comme par exemple sur la figure 5.6) et la matrice de dissemblance de départ.

Pour estimer la partition optimale, nous regardons l'évolution de la moyenne \bar{R}_a et de l'écart type associé, suivant le nombre de classes dans chaque niveau de partition. Le nombre de classes optimal correspond au maximum de l'indice de Rand bootstrap et au minimum de l'écart type. Par exemple sur la figure 3.10, nous observons un maximum de \bar{R}_a et un minimum de $\sigma(R_a)$ pour une partition de 4 classes. La corrélation bisériale donne aussi une indication de 4 classes, mais de manière moins précise car calculée à partir des matrices cophénétiques binaires, contrairement à l'indice de Rand calculé directement sur les partitions.

La partition optimale correspond donc au nombre de classes les plus stables à travers les données de classification des participants. Ce nombre de classes optimum permet de couper l'arbre hiérarchique de référence au niveau correspondant à la partition optimale.

Une application de cette méthode est donnée lors de l'analyse de nos résultats expérimentaux chapitre 5.

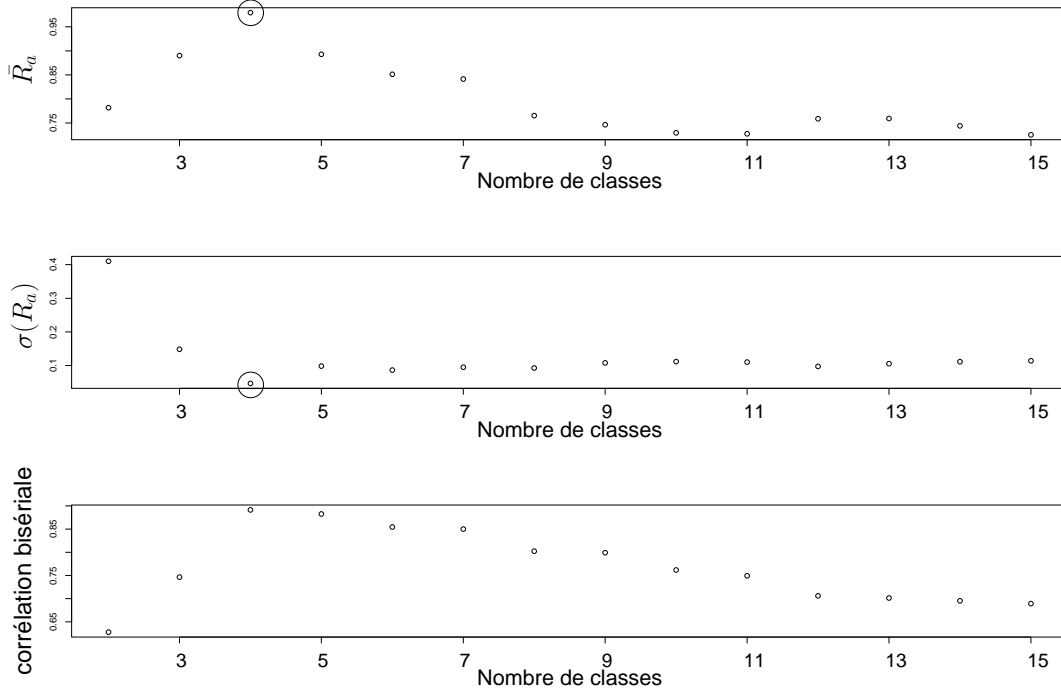


Figure 3.10 - Variation de \bar{R}_a et $\sigma(R_a)$ et de la corrélation bisériale suivant le nombre de classes (de 2 à 15), pour un bootstrap de 1000 tirages à partir de 30 matrices d'incidence tirées de l'expérience de classification « orientée » du chapitre 5. La valeur maximale de \bar{R}_a et la valeur minimale de $\sigma(R_a)$, toutes les deux entourées, correspondent à une partition de 4 classes dans cet exemple.

3.6 Discussion

Nous avons abordé dans ce chapitre deux notions associées à la proximité : sa modélisation et sa représentation. La modélisation formelle des proximités est généralement basée sur deux modèles distincts : les modèles géométriques et le modèle de contraste. Les modèles géométriques calculent la proximité entre deux items sous la forme d'une distance de Minkowski, la plus souvent euclidienne. Le modèle de contraste estime la proximité entre deux items par une fonction additive de leurs traits communs pondérés par les traits distinctifs propres à chaque item. D'un point de vue perceptif, ces deux modèles sont directement à associer aux différents types de propriétés perceptives des stimuli : les dimensions (modèle géométrique) ou les traits (modèle de contraste).

L'intérêt de ces deux modèles est de pouvoir s'appliquer de façon similaire, aux propriétés physiques des stimuli, permettant alors de prédire les

propriétés physiques sous-jacentes à la similarité perceptive. Dans le chapitre 5, nous présenterons une application du modèle de contraste sur la base de traits et de propriétés acoustiques. Ce modèle appliqué soit à des données perceptives, soit à des données acoustiques, nous permettra de modéliser les processus perceptifs sous-jacents aux tâches de classification mises en œuvre dans le chapitre 5.

Généralement, nous ne connaissons pas a priori les propriétés perceptives ou cognitives qui sous-tendent les processus psychologiques. Les expériences, qu'elles soient des tâches de classification (chapitre 5) ou des jugements de dissemblance (chapitre 6), nous permettent d'obtenir une mesure de similarité psychologique entre les sons. La représentation des proximités permet alors d'inférer le type de similarité impliqué, similarité de surface ou similarité profonde §2.8, ainsi que les propriétés liées aux processus perceptifs ou cognitifs impliqués dans les tâches expérimentales.

Nous avons introduit deux approches pour représenter les données de proximité, les modèles spatiaux (MDS) et les modèles arborés (hiérarchique et additif). Les modèles spatiaux représentent directement les proximités entre des items dans un espace multidimensionnel. Les dimensions de l'espace sont considérées comme les dimensions perceptives qui sous-tendent les jugements de proximité. Dans les représentations arborées, il n'y a pas de notion d'espace géométrique. Pour les arbres hiérarchiques, la similarité entre les items prend la forme de relations hiérarchiques imposées, alors que pour les arbres additifs, la similarité est une mesure de filiation entre les items.

Les différentes façons de représenter les proximités permettent d'éclairer différemment la structure perceptive ou cognitive des données. Par exemple, les données catégorielles résultant par exemple de tâches de classification se représentent mieux par des arbres, car les modèles spatiaux se prêtent mal aux relations hiérarchiques [SSG97] entre les items. De plus, si les propriétés impliquées dans les processus perceptifs s'apparentent à des traits, les arbres sont plus à même de représenter ce type de données. Un arbre additif peut être vu comme une représentation basée sur des traits distinctifs, contrairement aux arbres hiérarchiques basés sur des traits communs [CT86]. Les arbres additifs permettent de représenter, en plus, les effets de typicalité des items qui peuvent être parfois pertinents [Bar91]. Pour la représentation arborée, nous utiliserons de manière privilégiée les arbres hiérarchiques, car nous avons développé des techniques d'analyse des données de classification dérivées de ce type de représentation.

Les techniques d'analyse et de représentation des proximités que nous venons d'introduire nous servent à analyser les résultats de nos expériences présentés dans les parties expérimentales §5 et §6. Les analyses statistiques seront mises en relation avec les analyses acoustiques et mécaniques des sons

que nous expliquons dans la prochaine section.

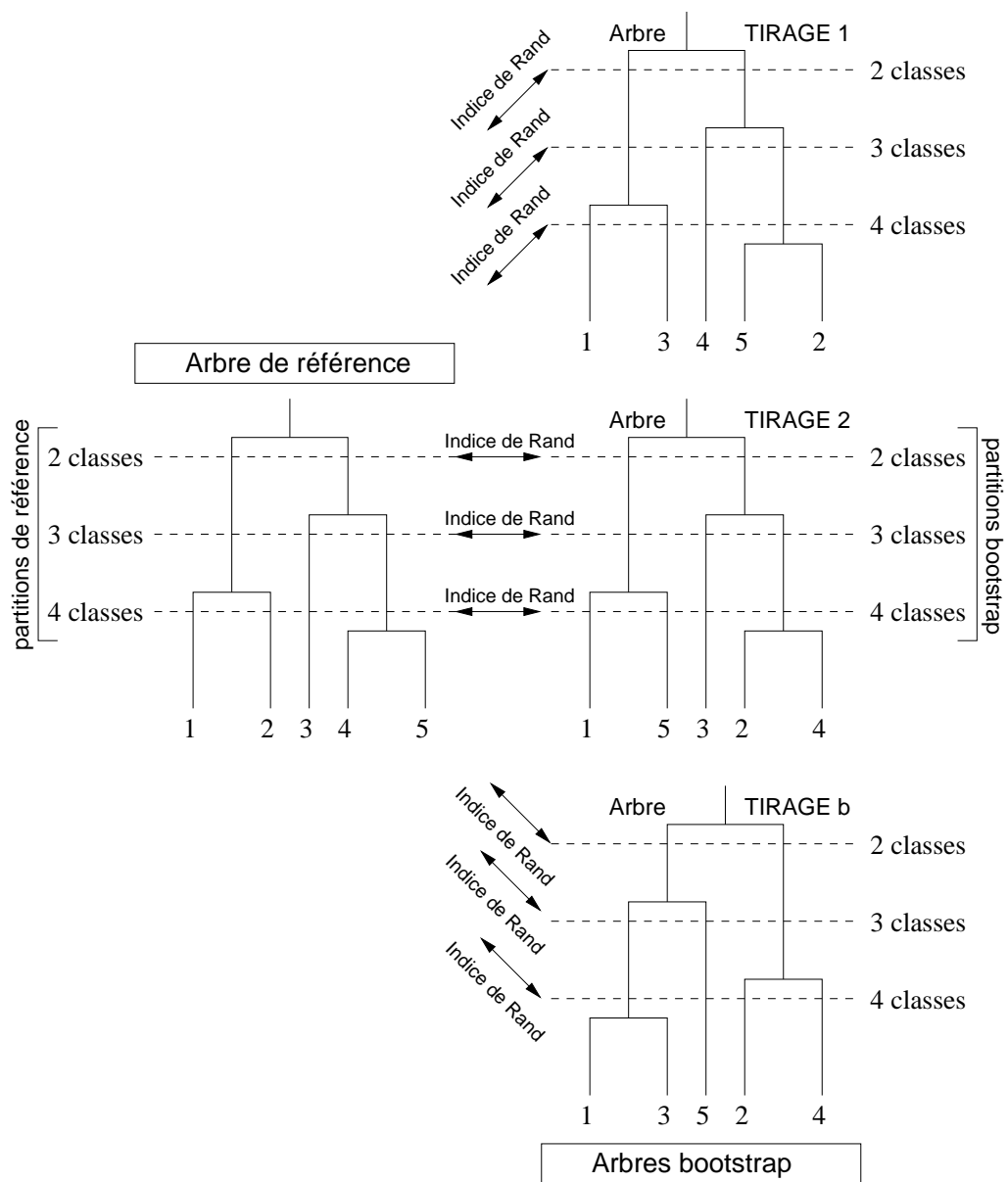


Figure 3.11 - Méthode de bootstrap appliquée à la comparaison des partitions entre l'arbre de référence et les arbres bootstrap par l'indice de Rand. Les partitions sont ici limitées à 2, 3 et 4 classes.

-
- Construire les N matrices individuelles d'incidence MI ;
 - , Construire la matrice de cooccurrences MC à partir des N matrices d'incidence MI ;
 - İ Analyse arborée (arbre de référence) sur la matrice de cooccurrences MC ;
 - ˆ L'arbre de référence est coupé à des niveaux correspondant à : 2 classes, 3 classes, ..., ces niveaux sont appelés : partitions de l'arbre de référence ;
- ↪ faire :
- ∨ Tirage aléatoire avec remise de N matrices d'incidence (\widetilde{MI}) ;
 - ∨ Calcul de la matrice de co-occurrences \widetilde{MC} à partir de cet échantillonnage de N matrices d'incidence \widetilde{MI} ;
 - ∨ Analyse arborée (arbre bootstrap) sur la matrice de co-occurrences \widetilde{MC} ;
 - ∨ L'arbre bootstrap est coupé à des niveaux correspondant à : 2 classes, 3 classes, ..., ces niveaux sont appelés partitions de l'arbre de bootstrap ;
 - ∨ Calcul des indices de Rand ajustés R_a entre les partitions de l'arbre de référence et les partitions de l'arbre bootstrap pour les mêmes niveaux de partition (2 classes, 3 classes, ...) ;
- ∨ Pour chacun des B tirages.
- Moyennes – \bar{R}_a – et écarts types – $\sigma(R_a)$ – sur l'ensemble des B tirages des indices de Rand sur chacun des niveaux de partition ;
 - Ñ Partition optimale : valeur maximale des \bar{R}_a et valeur minimale des $\sigma(R_a)$ sur les différents niveaux de partition.
-

Figure 3.12 - Résumé des différentes étapes pour estimer la partition optimale.

Chapitre 4

Analyse des signaux et modélisation mécanique



4.1 Introduction

Dans cette partie, nous exposons des techniques de traitement du signal qui nous permettent de caractériser les propriétés acoustiques des sons de barre de métal utilisés pour nos expériences. Par la suite nous mettrons en relation ces propriétés avec les différentes analyses de données expérimentales (chapitre 5 et 6). En complément de l'analyse des signaux, nous présentons la méthode des éléments finis qui nous permettra de modéliser et d'estimer les différents modes de vibrations des barres de métal. Ces modélisations mécaniques, conjointement aux analyses de signaux, nous serviront pour la synthèse de nouveaux sons utilisés dans la dernière série d'expériences (chapitre 6).

4.2 Analyses acoustiques

Nous présentons ici deux méthodes de traitement du signal qui sont particulièrement pertinentes pour l'étude des signaux percussifs comme les sons de barre de métal. Ces deux méthodes permettent d'extraire, pour chaque partiel, la fréquence et l'évolution de l'amortissement au cours du temps.

La première méthode présentée est de type paramétrique qui impose que la structure sous-jacente au signal soit sous la forme de sinusoides exponentiellement amorties. Cette hypothèse permet d'estimer dans un premier temps les fréquences et les amortissements associés puis dans un deuxième temps, l'amplitude et la phase de chaque partiel.

La deuxième méthode est une technique non-paramétrique qui contrairement à la première, est basée sur la transformée de Fourier. En calculant le spectre sur différentes parties du signal, les pics fréquentiels émergeant dans chacun des spectres associés aux fenêtres d'analyse permettent de calculer l'amortissement et la fréquence associés à chaque pic. Cette méthode permet ensuite de synthétiser directement des sons à partir de ces analyses grâce à une technique utilisant les formes d'onde formantiques.

4.2.1 Méthode paramétrique

Il s'agit d'une méthode d'analyse du signal de haute résolution. Ces techniques sont très souvent utilisées dans le domaine de la détection de direction d'arrivée d'un signal au niveau d'une antenne. De plus, elles sont particulièrement adaptées aux sons non-entretenus, de type percussif.

Ces méthodes, contrairement à la transformée de Fourier, supposent que le signal qui est observé possède une structure sous-jacente particulière qui influence directement les propriétés du signal [Lar93]. Cette structure dirige ainsi le type d'analyse du signal considéré.

4.2.1.1 Matrix-pencil

La méthode paramétrique dite de « matrix-pencil »^[a] associée à ces modèles, est une analyse en deux étapes. Une première étape permet d'estimer les fréquences ainsi que leurs décroissances temporelles. Dans un deuxième temps, cette méthode estime les amplitudes ainsi que les phases associées à ces fréquences. Cette méthode est présentée de manière détaillée dans l'article de J. Laroche, [Lar93]. Le tutoriel de J. Kusuma [Kus02] présente différentes méthodes paramétriques (ESPRIT, MUSIC).

^[a]Le terme français « faisceau de matrices » est très rarement employé.

■ **Fréquences et amortissements** Soit un signal x_n discrétisé dans le temps, dont la fréquence d'échantillonnage est F_s Hz. La méthode « matrix pencil » approche le signal original x_n par une somme de L sinusoides exponentiellement amorties de fréquence f_i , ayant comme amplitude initiale A_i , comme phase initiale ϕ_i et comme amortissement initial α_i . Le signal approché s'exprime sous la forme :

$$x_n = \sum_{i=1}^L A_i e^{-\alpha_i n} \cos(2\pi n f_i + \phi_i) \quad (4.0)$$

A partir de N échantillons consécutifs du signal dont les fréquences et amortissements seront extraits, deux matrices $(N - p) \times (p)$, avec p arbitraire, sont construites. Elles sont des partitions du signal à des instants différents. Cette construction peut s'apparenter à une autocorrélation. Du fait de l'hypothèse de la structure sous-jacente au signal, il existe une relation entre ces deux matrices qui permet de calculer des vecteurs propres ainsi que des valeurs propres. Les fréquences ainsi que les amortissements sont estimés à partir de ces deux derniers. Le nombre de fréquences contenues dans le signal est lié à l'évolution des valeurs propres calculées lorsqu'il n'existe pas d'a priori quant au nombre de sinusoides. La courbe de décroissance des valeurs propres, appelée généralement éboulis des valeurs propres, donne une approximation du nombre de sinusoides lorsque les valeurs propres chutent brutalement avant de se stabiliser.

Cette méthode d'estimation des paramètres du signal exploitent l'invariance rotationnelle des matrices (la méthode ESPRIT [RK89]) et des techniques de calcul matriciel tels que la matrice inverse de Moore-Penrose. Ces techniques ne seront pas présentées ici car elles sortent du cadre de cette thèse. Le lecteur pourra se familiariser avec ces techniques de traitement du signal grâce aux deux références déjà mentionnées [Kus02] et [Lar93].

Pour illustrer le principe de cette méthode, nous prendrons un exemple [Kus02] d'un signal \mathbf{x} de N observations qui est modélisé par une expression simple : $x_n = A_1 \exp(j\omega_1 n)$ avec A_1 un nombre complexe représentant l'amplitude et la phase de la composante fréquentielle de pulsation $\omega_1 = 2\pi f_1$. Le vecteur d'observations \mathbf{x} de longueur N s'exprime sous deux formes :

$$\begin{aligned} \mathbf{x} &= [x_0, x_1, \dots, x_{N-1}], && \text{le signal connu} \\ &= A_1 \times [1, e^{j\omega_1}, e^{j\omega_1^2}, \dots, e^{j\omega_1(N-1)}], && \text{le modèle de signal} \end{aligned} \quad (4.0)$$

Le signal connu est maintenant séparé en deux vecteurs de cette façon :

$$\begin{aligned}
 \mathbf{x} &= \underbrace{[x_0, x_1, \dots, x_{N-2}, x_{N-1}]}_{\mathbf{S}_1} \\
 &= [x_0, \underbrace{x_1, \dots, x_{N-2}, x_{N-1}}_{\mathbf{S}_2}]
 \end{aligned} \tag{4.0}$$

Nous pouvons alors écrire que : $\mathbf{S}_2 = e^{j\omega_1}\mathbf{S}_1$, expression permettant de calculer ω_1 connaissant \mathbf{S}_1 et \mathbf{S}_2 . Si le signal est bruité, une technique de moyennage peut être utilisée afin d'obtenir une meilleure estimation de ω_1 . Sur cet exemple simple, nous voyons le principe général de l'estimation des fréquences. La méthode ESPRIT généralise cette approche à plusieurs sinusoides. Le signal est segmenté en deux matrices correspondant à deux sous-partitions d'échantillons contigus se recouvrant comme dans l'exemple précédent. La relation entre ces deux matrices du signal est telle que la matrice qui les lie correspond à une matrice unité des $e^{j\omega_n}$, avec n le nombre de sinusoides. Les techniques de calcul matriciel permettent de remonter à cette matrice et de calculer les vecteurs et valeurs propres.

■ **Amplitudes et phases** L'amortissement et la fréquence de chaque sinusoides étant connus, l'estimation de son amplitude ainsi que de sa phase devient un problème linéaire [Lar93]. L'amplitude et la phase sont estimées dans un deuxième temps en minimisant l'erreur quadratique au carré entre le signal original et un signal recréé. Ce signal est créé à partir de sinusoides dont seules les amplitudes et les phases sont inconnues.

■ **Utilisation** Le point délicat dans cette méthode est le nombre de sinusoides prises en compte dans le modèle, qui peuvent influencer l'estimation des paramètres. De plus cette méthode est limitée à des signaux dont le nombre d'échantillons est faible (1000-3000). Il existe différents moyens de pallier à ces carences.

Une première approche est d'estimer les pics dans le spectre dont les amplitudes sont supérieures à un seuil (à fixer) et d'associer le nombre de pics au nombre de sinusoides à chercher. Une autre méthode consiste à effectuer un filtrage du signal autour d'une fréquence considérée et de lui appliquer la méthode paramétrique en ne considérant qu'une seule sinusoides. Dans tous les cas, il est préférable de surestimer le nombre de sinusoides en éliminant les fréquences dont les amplitudes sont très faibles.

Cette méthode ne s'appliquant uniquement sur un nombre d'échantillons faible (de l'ordre de 1000), il est parfois nécessaire de passer par une démodulation ainsi qu'un sous-échantillonnage pour réduire le nombre de points et

effectuer ce type d'analyse. Une autre solution [Izm00] est de décomposer le signal en fenêtres temporelles de courte durée et d'appliquer cette méthode sur chaque fenêtre temporelle afin de s'affranchir de la contrainte de la durée du signal.

4.2.2 Analyse en Modèles de Résonance : modRes

L'analyse en modèle de résonances est une méthode développée dans le cadre de l'analyse et de la synthèse de la voix [Rod97] et [Mar99]. Cette analyse considère que les pics qui seront extraits des spectres du signal sont des résonances du corps vibrant qui a produit le signal. C'est un modèle simple pour des systèmes à oscillation libre tels qu'une barre de métal frappée. Les résonances dans le cas d'une barre de métal, sont les différentes fréquences modales de la barre.

De la même façon que précédemment, le signal est implicitement vu comme composé de résonances dont chaque partiel possède son propre amortissement exponentiellement amorti. Contrairement à la technique paramétrique précédente, l'analyse est basée sur des transformées de Fourier sur différentes portions du signal. A partir des pics fréquentiels qui se retrouvent sur différentes fenêtres du signal, leurs amplitudes ainsi que leurs décroissances temporelles sont estimées.

4.2.2.1 Présentation de la méthode

■ **Synopsis** A chaque étape, deux fenêtres temporelles sont choisies sur le son, voir fig. 4.1. Les spectres sont estimés sur ces deux fenêtres par transformée de Fourier. Les pics sont extraits sur le spectre calculé sur chacune des deux fenêtres puis appariés. Le seuil d'appariement entre fréquences peut être ajusté afin d'affiner la séparation entre les pics. Lorsqu'un appariement est possible (suivant ce seuil), la paire de pics est associée à un modèle de résonance du corps vibrant ayant produit le son analysé. La position en fréquence ainsi que l'évolution en amplitude des pics (entre les deux fenêtres temporelles) permet de déduire un taux de décroissance par pic. Cette première opération produit un modèle « élémentaire », sur la figure 4.2 le modèle e_0 . Tout le processus itératif est décrit sous la forme d'un schéma, figure 4.2.

Le modèle élémentaire courant (e_i , pour l'étape i) est confronté au modèle de résonances obtenu lors de la précédente étape d'analyse (m_i), sauf pour la première itération qui produit un modèle initial (e_0). Les résonances modélisées sont appariées lorsque les pics du modèle élémentaire courant et du modèle précédent sont suffisamment proches en fréquence. Leurs paramètres sont modifiés en fonction des anciennes et des nouvelles valeurs. Ces

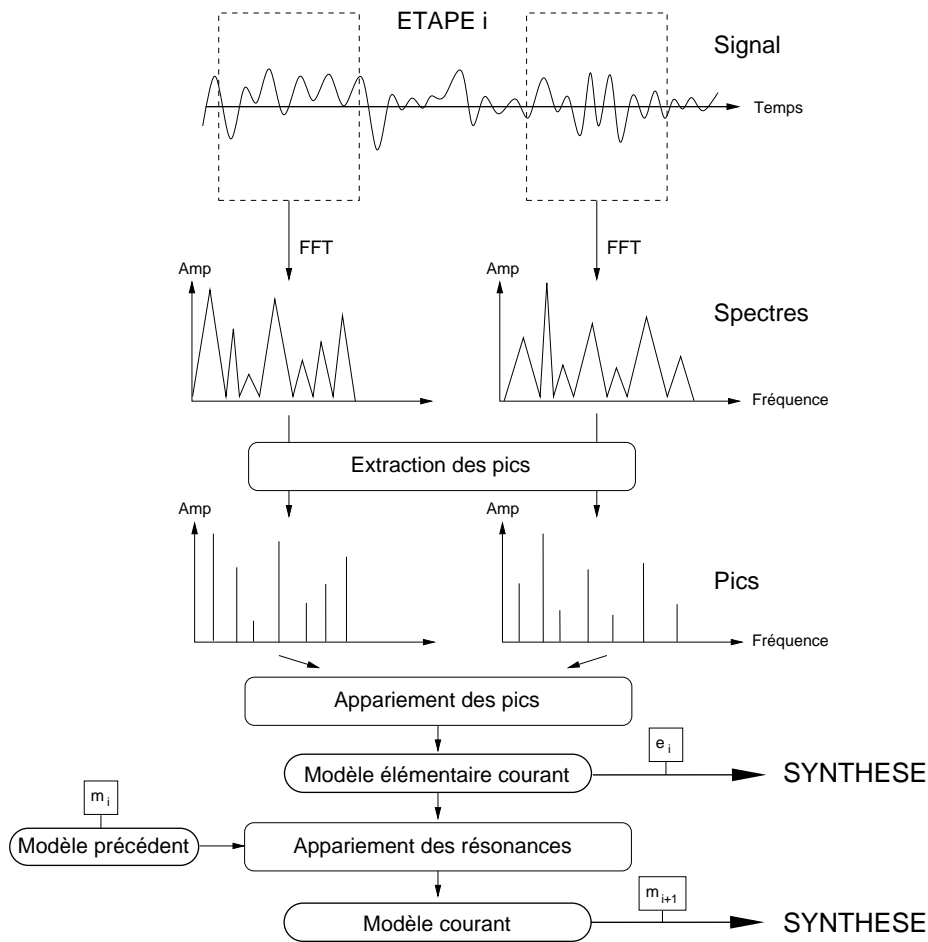


Figure 4.1 - Synopsis de l'analyse/synthèse en modèles de résonance

appariements produisent un modèle « fusionné » (m_{i+1}) de résonances qui est repris dans l'analyse suivante. Lorsque l'appariement n'est pas possible, les résonances sont conservées ou rajoutées sans modification, le modèle est dans ce cas appelé modèle « enrichi » (m_{i+1}) et constitue le produit final d'une étape d'analyse i . Après N étapes nous obtenons le modèle définitif m_N .

■ **Détails sur l'analyse** L'analyse commence avec deux fenêtres de courte durée appliquées sur le son. Lorsque le son possède une attaque rapide, il est possible de placer la première fenêtre avant le son afin d'affiner l'analyse de l'attaque. A chaque nouvelle étape, la taille des fenêtres et leur écartement sont augmentés. Ceci permet de gagner en résolution fréquentielle et d'affiner l'estimation du taux de décroissance des résonances. L'utilisateur peut fournir à chaque étape un certain nombre de paramètres (taille, type et position des fenêtres, ...) , et peut écouter le son correspondant à l'analyse élémentaire

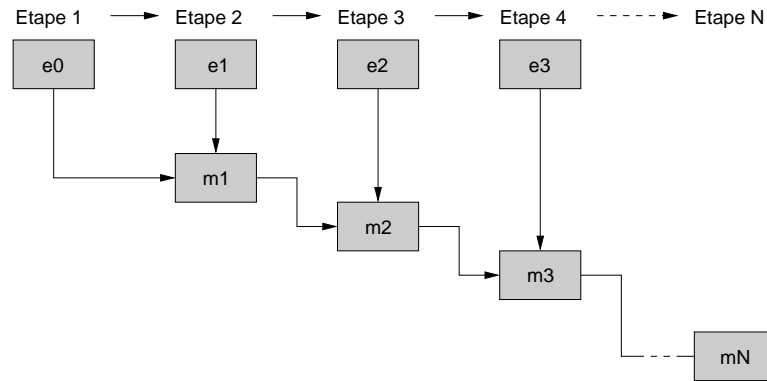


Figure 4.2 - Schéma du déroulement des analyses. Pour une étape i e_i correspond à un modèle élémentaire courant, m_{i-1} correspond à un modèle enrichi et m_i à un modèle fusionné. Après plusieurs étapes, N , l'analyse estime un modèle de résonances m_N

courante, ainsi que celui correspondant au résultat global (le modèle fusionné « enrichi ») courant. L'analyse complète est terminée, généralement après 3 à 8 étapes lorsque le résultat est jugé satisfaisant à l'oreille.

4.2.3 Comparaison des deux méthodes

La différence entre la méthode « matrix pencil » et la méthode modRes porte principalement sur le caractère paramétrique de la première. En effet, la méthode « matrix pencil » part de l'hypothèse que le signal peut être vu comme une somme de sinusoides exponentiellement amorties. Le nombre de sinusoides est fixé a priori. Pour l'analyse modRes, il n'existe pas de contrainte sur le nombre de sinusoides. De même il n'y a pas de limitation quant à la longueur du signal ou la complexité du calcul, ce qui rend modRes très robuste [Rod97]. Nous pouvons noter que même si modRes n'est pas une méthode paramétrique, implicitement le signal est vu comme une somme de résonances pouvant s'apparenter à des sinusoides amorties pour l'estimation des amortissements. Ceci ne rentre pas directement dans le calcul. Par contre la méthode modRes n'estime pas la phase contrairement à ESPRIT, ce qui peut être problématique dans certains cas. Dans le cadre de cette thèse nous avons utilisé ces deux outils, et plus particulièrement modRes dans le cadre de la synthèse par formes d'onde formantiques que nous introduisons dans la section suivante.

4.3 Synthèse par FOF

Dans le but de synthétiser des sons de barre à partir des analyses précédentes par modèle de résonances, nous avons utilisé le programme CHANT [RPB84]. Ce programme permet de synthétiser un son en utilisant des formes d'onde formantiques. Ces dernières tentent de modéliser les résonances à la fois du point de vue temporel et fréquentiel de la source produisant le son.

4.3.1 Introduction

Le développement de CHANT est directement lié à la synthèse de son de la voix chantée [RPB84] [BR89]. Il a été par la suite généralisé à la synthèse d'autres signaux qui peuvent être analysés sous la forme de résonances de la source sonore.

Le fonctionnement de CHANT est dérivé de l'étude du mécanisme de la voix. Nous introduisons succinctement le fonctionnement de l'appareil vocal, considéré ici comme un système source-filtre. Pour une revue complète de la problématique, voir [RPB85].

L'appareil vocal consiste en une source d'énergie et un système de résonances. Les poumons génèrent un flot d'air continu qui est converti en signal alternatif par une modulation périodique (p. ex. dans le cas des voyelles au niveau des cordes vocales) ou chaotique (p. ex. dans le cas des consonnes au niveau des lèvres). Ce signal peut être filtré par les résonances du système (larynx-pharynx-bouche), qui donne une couleur caractéristique au phonème.

CHANT simule ce fonctionnement par un modèle de type excitation-résonance. A chaque formant dans le cadre de la voix ou résonance dans le cadre général, est associée une réponse de base simulée avec une technique de synthèse spécifique, les Formes d'Onde Formantiques (FOF). CHANT génère le son résultant en additionnant les FOF correspondant à chaque formant pour une source pseudo-périodique donnée. Cette source peut être impulsive si le son simulé est percussif.

De la même manière, si nous considérons un son de barre de métal, chaque fréquence modale est liée à un mode de vibration spécifique (torsion, flexion, compression) comme une résonance de la barre. Chacune des résonances associées à un son de barre de métal est modélisée par une FOF. Les FOF sont particulièrement pertinentes de par leurs structures exponentiellement amorties, présentées dans la section suivante.

4.3.2 Formes d'Onde Formantiques

Les FOF ont la structure d'une sinusoïde exponentiellement amortie. Le signal h_k associé est modélisé par l'expression $h(k) = \sum_{i=1}^n s_i(k)$, avec $s_i(k)$ la FOF i et k l'index de l'échantillon. De plus les sons naturels n'ayant pas d'excitation instantanée, une fonction apporte le contrôle de ce paramètre (le transitoire d'attaque), la FOF s'écrit alors de la façon suivante :

$$s_i(k) = \begin{cases} 0 & \text{pour } k < 0 \\ 1/2(1 - \cos[\beta_i k])A_i e^{-\alpha_i k} \sin[\omega_i k + \Phi_i] & \text{pour } 0 \leq k \leq \pi/\beta_i \\ e^{-\alpha_i k} \sin(\omega_i k + \Phi_i) & \text{pour } \pi/\beta_i < k \end{cases} \quad (4.0)$$

avec β_i qui détermine le temps d'attaque du partiel, α_i l'amortissement, $\omega_i = 2\pi f_i$ la fréquence de résonance, A_i le gain de la résonance. La figure 4.3 (1) représente la FOF.

Le spectre de la FOF, figure 4.3 (2), présente un maximum qui peut être ajusté directement par les paramètres suivants : ω_i est la fréquence centrale du maximum, $\alpha_i \pi$ est la largeur de bande à $-3dB$ et β_i règle la largeur des jupes ou la pente de l'attaque. La figure 4.3 (2) quant à elle montre l'influence de β_i sur le spectre pour différentes valeurs.

La synthèse du son composé de FOF ayant chacune des paramètres d'amplitude, de fréquence, d'amortissement, se fait en excitant chacune des ondes formantiques par une excitation virtuelle unitaire. La figure 4.3 (3) représente une synthèse d'un son de barre de métal par une excitation en parallèle de différentes FOF.

4.4. MÉTHODE DES ÉLÉMENTS FINIS

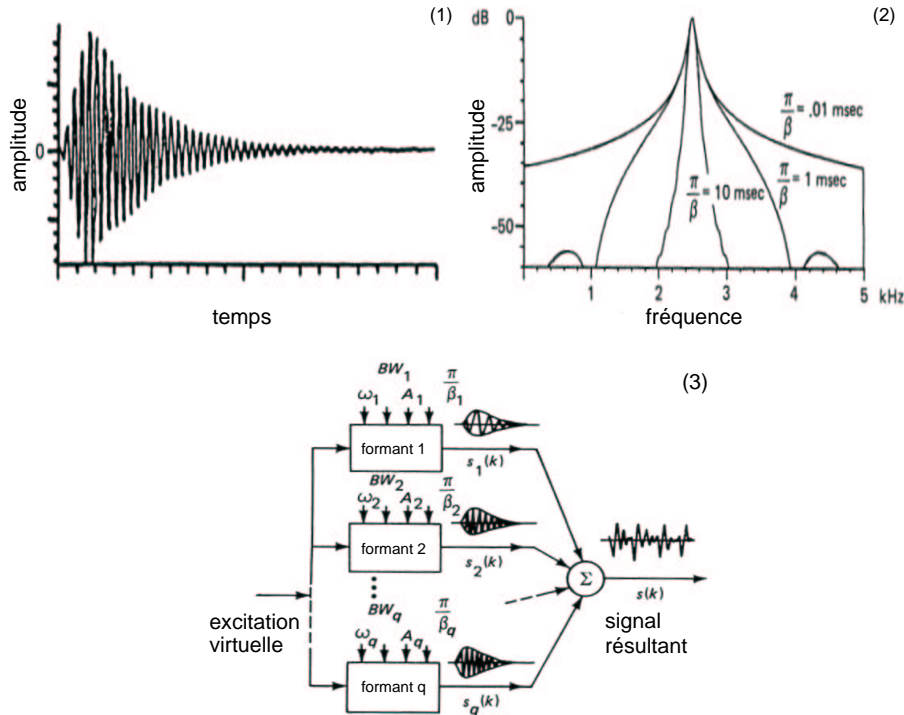


Figure 4.3 - La figure (1) représente la forme d'onde formantique (FOF) décrite par l'équation 4.3.2. La figure (2) représente le spectre d'une FOF pour différentes valeurs du temps d'attaque β . Enfin la figure (3) représente une synthèse en parallèle de plusieurs FOF produisant le signal final. Figures tirées de [BR89].

4.4 Méthode des éléments finis

Pour l'analyse physique et la simulation sonore, nous avons modélisé les vibrations des barres de métal utilisées lors de nos expériences de classification (§5, p. 132) et de représentation (§6, p. 177) des sources sonores.

Dans le cas d'une barre de métal, nous disposons des outils théoriques pour écrire les équations des différents modes de vibration de la barre : mouvement longitudinal, de torsion ou de flexion. Si nous considérons un modèle de barre pour lequel nous ne tenons pas compte des effets secondaires comme l'inertie de rotation ou l'effet de cisaillement, il est aisé dans le cadre du modèle de Euler-Bernoulli de calculer analytiquement les fréquences modales d'une barre pour ces différents mouvements. La mise en équation de ces différents mouvements ainsi que leur résolution analytique sont présentées en annexe §A, p. 222.

Néanmoins, nous avons utilisé la modalisation par éléments finis pour vi-

ser deux buts principaux. Le premier but est de déterminer les fréquences modales pour chacune des barres de manière plus précise que les estimations par une résolution des équations classiques du mouvement. Un second but est de visualiser les déformées modales propres à chacune des barres et d'associer le type de mode de vibration (longitudinal, flexion, torsion) à chacune des fréquences modales. Ces deux informations nous permettront de mettre en correspondance les fréquences estimées par nos analyses acoustiques avec les fréquences modales et le type de modes de vibration associé. Nous utiliserons ces informations pour les analyses de nos expériences ainsi que pour la synthèse de nouveaux sons.

4.4.1 Mouvements des barres

Afin de spécifier le comportement mécanique des objets résonants tels que les barres de métal, nous nous intéressons aux équations du mouvement de ces objets. Ces équations rentrent dans le cadre de la mécanique des systèmes continus (par opposition aux systèmes « masses-ressorts » qui eux comportent un nombre défini de degrés de liberté).

De manière générale, le comportement des structures mécaniques reste difficilement calculable de manière analytique de par la géométrie, le matériau, les conditions aux limites de la structure (c'est à dire la fixation de l'objet ...) [LBDH86]. Il est nécessaire d'avoir recours dans ce cas à des méthodes donnant des solutions numériques, telles que la méthode des éléments finis

Même si notre objet est simple, une barre, il nous a semblé judicieux d'utiliser la méthode des éléments finis afin de pouvoir identifier facilement, par la visualisation, le type de mode de vibration associé à chaque fréquence modale. Ainsi la modélisation en trois dimensions du comportement de chacune des barres de métal nous permet de calculer les fréquences de manière plus précise et de visualiser les déformations propres à chacun des modes de vibration.

4.4.2 Principe général des éléments finis

Lorsque nous étudions un problème physique, nous cherchons à décrire son comportement physique, afin de comprendre son fonctionnement, le prédire. Pour cela nous devons trouver un champ^[a] (scalaire, vectoriel, ou tensoriel) que nous notons $\mathbf{u}(M, t)$, qui doit remplir un certain nombre de conditions [Gua02]. Ce champ qui peut être vectoriel comme le déplacement, doit

^[a]Région de l'espace dans laquelle un phénomène physique peut être observé.

satisfaire à un ensemble d'équations aux dérivées partielles et d'équations ordinaires en tout point M d'un domaine Ω et à tout instant $t^{[a]}$. De plus ce champ doit respecter des conditions aux limites (pouvant varier en fonction du temps) sur la frontière $\partial\Omega$ du domaine. Les conditions aux limites sont des relations ou des valeurs imposées à \mathbf{u} et/ou à ses dérivées sur la frontière.

Face à un problème réel, nous pouvons généralement trouver un modèle physique pour lequel nous posons les équations et trouvons les conditions aux limites. Malheureusement il n'existe pas toujours une approche mathématique permettant d'obtenir une résolution analytique [Gua02]. Dans ce cas, il faut introduire des méthodes d'approximation en physique qui ont pour objectif de remplacer un problème mathématique défini sur un milieu continu (équations différentielles ou intégrales) par un problème mathématique discret (équation matricielle) tel que la méthode des éléments finis. Ce problème devient un problème de dimension finie que nous savons résoudre [Hou98].

La méthode des éléments finis revient à décrire, en une première approche, le comportement d'une structure continue sous la forme d'une structure maillée dont le comportement des éléments est connu et calculable.

La méthode des éléments finis du fait de la discrétisation de la structure étudiée demande des ressources de calculs qui augmentent rapidement avec la complexité de la structure. Le traitement se fait donc de manière informatique.

4.4.3 Etapes de la méthode

Nous pouvons résumer la méthode des éléments finis en différentes étapes [LBDH86] [Gua02] [Hou98] :

- z discrétisation du milieu continu en sous-domaines ;
- z construction de l'approximation nodale par sous domaine ;
- z discrétisation du problème physique ;
- z assemblage des matrices élémentaires, prise en compte des conditions limites ;
- z résolution du système d'équations par une technique numérique.

Ces étapes sont détaillées dans les sections suivantes.

^[a] Si le temps t n'apparaît pas comme variable, le problème est dit stationnaire. Sinon c'est un problème d'évolution.

■ **Principe général** Le domaine continu étudié est découpé en sous-domaines. La méthode des éléments finis consiste alors à chercher une solution approchée de la solution exacte sous la forme d'un champ $\tilde{F}(M, t)$ défini par morceaux sur des sous-domaines de Ω .

Les n sous-domaines Ω_i doivent être tels que leur union $\bigcup_{i=1}^n \Omega_i = \Omega$ et $\tilde{\Omega}_i \cap \tilde{\Omega}_j = \emptyset \quad \forall i \neq j$, où $\tilde{\Omega}_i$ désigne l'intérieur de Ω_i . Autrement dit, les Ω_i sont une partition de Ω [Gua02].

4.4.3.1 Discrétisation géométrique

Le domaine matériel continu est discrétisé en sous-domaines qui sont des mailles de dimensions finies. Un domaine peut être une structure continue baignant dans un milieu différent (p. ex. une coque dans un fluide), ou une structure continue comportant des milieux différents (p. ex. une poutre en matériau multicouche). Dans le cas des barres, le domaine peut être ramené à la structure si nous ne tenons pas compte du rayonnement de la barre. Par souci de simplicité nous associerons le domaine à la structure.

Les mailles sont des formes géométriques simples qui couvrent la structure pour s'approcher idéalement de la structure étudiée. Le but est alors de minimiser les erreurs de calcul. Les mailles simples de dimension finie sont appelées éléments finis car elles sont réunies en des points situés sur leur contour appelés points nodaux (ou nœuds) qui assurent la continuité entre ces éléments. Le nombre de nœuds (multiplié par les degrés de liberté propres à chaque nœud) contenus dans les mailles de la structure donne le nombre de degrés de liberté de celle-ci.

La figure 4.4 montre un exemple de discrétisation d'une structure cubique en éléments finis sous la forme d'éléments finis cubiques. Contrairement à cet exemple, l'intérêt du type de méthode adopté est de mailler des structures géométriques beaucoup plus complexes. Ceci implique qu'il existe toujours une erreur de discrétisation géométrique du domaine étudié qui doit être minimisée, ce qui n'est pas le cas ici.

4.4.3.2 Approximation nodale

Le but de l'approximation nodale est d'approcher le champ \mathbf{u} (dans notre cas le vecteur déplacement), par une approximation \mathbf{u}^* du champ dans chaque élément fini Ω_i . Cette approximation au niveau de chaque élément fini est effectuée pour un nombre fini de valeurs correspondant à leurs nœuds.

L'approximation du champ local \mathbf{u}_i au niveau de chaque élément fini se fait par interpolation (généralement sous la forme polynomiale) entre les

4.4. MÉTHODE DES ÉLÉMENTS FINIS

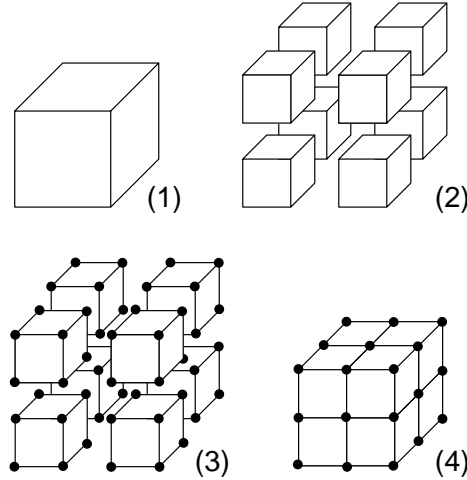


Figure 4.4 - La figure (1) représente la structure à mailler, ici un cube. La figure (2) montrent les 8 éléments finis, ici des mailles cubiques qui divisent le domaine. Afin de parler d'éléments finis, il faut associer à chaque maille des nœuds (représentés par des points noirs) montrés sur la figure (3). La figure (4) représente la discrétisation complète de la structure de départ. Nous remarquons qu'il existe une continuité inter-éléments, les éléments finis possédant des nœuds communs.

valeurs du champ des nœuds. Cette étape est appelée approximation nodale du champ :

$$\{u_{(M)}^*\} = [N_{(M)}]\{u_i\}, \text{ en tout point } M \text{ du sous-domaine } \Omega_i \quad (4.0)$$

avec $\{u_{(M)}^*\}$ le champ vectoriel local approché en tout point M de l'élément Ω_i , $[N]$ la matrice des fonctions d'interpolation de l'élément et $\{u_i\}$ les variables nodales relatives aux nœuds d'interpolation de l'élément.

Le but de l'interpolation est de donner une valeur approchée du champ local en tout point de la maille considérée en ne connaissant que les valeurs d'une solution du champ aux nœuds de l'élément.

La matrice des fonctions d'interpolation est construite à partir d'une approximation générale sur une base de fonctions connues indépendantes (en général polynomiales) que nous ne détaillerons pas ici. Cette base est choisie suivant le nombre de dimensions de la structure et le type d'interpolation, linéaire ou quadratique. En connaissant la forme de l'élément fini, nous pouvons exprimer cette forme générale sous une base de fonctions polynomiales. Les programmes d'éléments finis ont déjà des fonctions polynomiales déterminées suivant le type de maillage et d'éléments finis utilisés. Généralement cette phase est « transparente » pour l'utilisateur.

Le champ global, représentant le champ en tout point de la structure, est une juxtaposition de ces champs locaux. Une des contraintes imposées est la continuité entre les sous-domaines, ce qui implique une continuité entre les nœuds des éléments qui se juxtaposent.

Il faut remarquer que le maillage s'effectue dans un repère dont les points ont des coordonnées. Ceci implique que les nœuds ainsi que les interpolations sont ramenés aux coordonnées du repère réel.

4.4.3.3 Discrétisation du problème physique

Afin de poser le problème à résoudre sur l'ensemble du domaine considéré, il faut trouver comment poser le problème sur chaque élément fini. Il faut d'abord poser le problème physique sous une forme intégrable qui correspond à ce qui est appelé « forme variationnelle » du problème. Dans le cadre d'une barre, cette forme est l'équation du mouvement. Le but de cette étape est de définir en chaque élément les formes matricielles de la loi de comportement de la structure.

Sur chaque élément que nous appelons D_e , l'équation du comportement s'écrit sous forme matricielle :

$$[M_e]\{\ddot{u}_n\} + [K_e]\{u_n\} = \{F_{de}\}, \quad \forall D_e \quad (4.0)$$

avec $[M_e]$ la matrice des masses élémentaires correspondant aux masses ramenées sur les nœuds, u_n les variables nodales (p. ex. un champ vectoriel), $[K_e]$ la matrice de raideur élémentaire et $\{F_{de}\}$ les efforts donnés. Dans le cas où les liaisons entre les éléments ne sont pas parfaites, il conviendrait d'ajouter des forces inconnues au niveau de chaque élément. La variable nodale est exprimée sur la base de l'approximation polynomiale. Dans l'exemple ci-dessus, nous avons négligé l'amortissement qui peut être introduit dans la première partie de l'expression du comportement 4.4.3.3 par l'expression $[R_e]\{\dot{u}_n\}$ avec $[R_e]$ la matrice des amortissements élémentaires.

4.4.3.4 Assemblage et conditions aux limites

L'assemblage des matrices définies sur chaque élément permet d'exprimer l'équation du mouvement sur l'ensemble de la structure.

La matrice de masse globale $[M]$ est exprimée sous la forme d'une somme des matrices des masses élémentaires $[M_e]$. Il en est de même pour la matrice globale de raideur $[K]$. La matrice des efforts donnés, F_d , est également décrite par la somme des travaux virtuels sur chaque élément.

Le principe des travaux virtuels s'écrit alors :

$$[M]\{\ddot{U}\} + [K]\{U\} = \{F_d\} \quad (4.0)$$

avec U le champ global.

La résolution du problème physique devient alors un problème de calcul matriciel. Pour résoudre le système d'équations, il faut tenir compte des différentes conditions aux limites. Les fréquences et les modes associés sont déduits de la recherche des solutions de l'équation différentielle 4.4.3.4 par des calculs de valeurs et de vecteurs propres en posant par exemple le champ sous une forme exponentielle complexe.

Le calcul pouvant être extrêmement long, il existe suivant le nombre de degrés de liberté des techniques matricielles qui permettent de minimiser le temps de calcul.

4.5 Discussion

Dans cette partie nous avons présenté différents outils de traitement de signal et de modélisation mécanique nous permettant d'analyser à la fois du point de vue vibratoire et du signal, le comportement des barres que nous utiliserons pour nos expériences.

Les analyses des nos signaux utiliseront parallèlement les deux méthodes ESPRIT et modRes. Les analyses modRes seront plus particulièrement appliquées pour la série d'expériences sur la représentation des sources sonores en intégrant nos analyses par méthodes des éléments finis pour la synthèse de nouveaux sons. Nous avons développé ainsi une approche originale de la synthèse sonore en utilisant parallèlement des techniques de traitement du signal et la modélisation par éléments finis des barres.

Conjointement aux analyses des données expérimentales, ces différentes techniques d'analyse du signal rendront possible la comparaison entre modèles perceptifs des processus perceptifs et modèles acoustiques prédisant ces processus.

Chapitre 5

Classification des sources sonores



5.1 Introduction

5.1.1 Cadre d'étude

Dans ce chapitre, nous présentons différentes expériences qui ont pour but d'étudier la relation entre perception des propriétés audibles d'événements sonores simples, tels que des barres de métal frappées, et leur reconnaissance.

Pour étudier cette problématique, nous sommes partis d'un objet simple tel qu'une barre de métal excitée en différents points. Nous voulons connaître si les auditeurs extraient des invariants structuraux propres à chaque barre, malgré les changements de points d'excitation. Grâce à cela, nous essayons de spécifier les propriétés audibles d'une source, une barre de métal frappée, qui permettent sa reconnaissance.

L'utilisation d'objets résonants simples permet de se placer à un niveau assez général de description afin de tenter de généraliser les résultats à d'autres sources [Gav88]. De plus, une bonne connaissance mécanique de ces événe-

ments nous permet, de lier le phénomène perceptif au phénomène physique dans une même démarche. Nous cherchons donc à connaître si les auditeurs ont la capacité à extraire des invariants structuraux propres à la source sonore amenant ainsi à sa reconnaissance.

Nous avons tout d'abord mené des expériences de classification de sons de barres de métal. Les sons utilisés proviennent de barres excitées en différents points. Nous avons utilisé ce protocole de classification sous trois consignes expérimentales différentes, afin de juger des effets de contexte. Nous avons procédé à différentes analyses acoustiques sur les sons afin de retrouver l'information utilisée par les auditeurs, lors des tâches de classification. Sur la base de ces analyses, nous avons fait une expérience d'appariement afin d'estimer la représentation perceptive de chaque barre. Suite à cette expérience, nous avons modélisé les résultats des classifications à partir de ces représentations perceptives afin de valider nos hypothèses. Dans la suite, du texte nous détaillons ces différentes étapes.

5.1.2 Hypothèses

Une barre peut être décrite par ses différents attributs mécaniques : sa géométrie, son matériau, son type d'excitation, comme nous l'avons déjà introduit, voir §2.1, p. 22. Ces différents attributs physiques de la source influencent l'aspect temporel, spectral et spectro-temporel du stimulus proximal. Dans notre cas, seules vont varier la position d'excitation ainsi que la section (épaisseur/largeur). Les barres sont du même matériau – de l'acier laminé – et la longueur est commune à toutes les barres. Nous pouvons penser que dans notre cas, la forme de l'objet, notamment la section géométrique de la barre serait un facteur déterminant de son identité sonore comme l'ont déjà montré les études de Lakatos et coll. [LMC97]. L'ensemble de ces attributs invariants peut être une source potentielle d'information pour la reconnaissance de la source sonore et donc pour son identité sonore.

Nous partons de l'hypothèse que l'invariance potentielle peut être le pattern spectral du signal de chaque barre, qui est lié à la contribution de ses différents modes. En excitant la barre en différents points, nous pondérons ces différentes contributions spectrales en tentant de maintenir ce pattern spectral. Ainsi chaque barre peut être caractérisée par un pattern spectral composé des différentes fréquences modales de vibration dont l'amplitude varie suivant la position d'excitation.

Pour étudier les invariants potentiels engagés dans la reconnaissance de ces sources sonores, nous utilisons un paradigme de classification libre des sons. Nous demandons aux participants de chaque expérience de regrouper des sons suivant une consigne. Nous voulons connaître si les classifications

effectuées par les participants reflètent des catégories auditives communes. Nous voulons voir si ces catégories sont propres à chaque barre. Ces catégories nous permettront d'inférer les propriétés perceptives qui reflètent les invariants acoustiques, voire mécaniques, propres à chaque barre.

En faisant varier les consignes des expériences, tout en gardant le même protocole (mêmes sons et tâche de classification), nous étudierons les variations du contexte sur la nature des processus perceptifs ou cognitifs mis en jeu. Nous pouvons ainsi étudier les différents types de similarité (voir §2.8.4.2, p. 62) impliqués dans le processus de catégorisation.

5.2 Méthode

5.2.1 Présentation des stimuli

Le corpus de sons est constitué de six barres parallélépipédiques en acier laminé ayant des largeurs et des épaisseurs distinctes mais dont la longueur est constante. La figure 5.1 indique la distribution des dimensions des barres. Le choix de ce corpus de barres a été effectué grâce aux travaux de Lakatos et coll. [LMC97] afin d'obtenir un continuum perceptif entre les différentes barres. Les barres sont un sous-ensemble des stimuli utilisés dans leur expérience. Chaque barre est posée sur deux liens élastiques, voir figure 5.1. Cette position s'apparente au mode de fixation des lames de bois d'un xylophone. En première approximation, nous pouvons associer cette position à une condition « libre-libre » pour le mouvement de la barre. La condition « libre-libre » revient à négliger les supports de fixation de la barre mise en vibration.

Chaque barre est excitée en cinq points, voir figure 5.1, afin de privilégier certains modes de vibration :

- z modes de vibration de flexion suivant l'épaisseur : positions 1 & 2
- z modes de vibration de flexion suivant la largeur : positions 4 & 5
- z et le mode de torsion : position 3

Les positions de frappe 1 et 4 privilégient la mise en vibration du mode fondamental et les positions de frappe 2 et 5 plus particulièrement les modes d'ordre élevé. La figure 5.2 montre les cinq spectrogrammes correspondant à la barre 80 x 30 excitée à ses 5 points. Les différents spectrogrammes permettent de voir les différentes contributions fréquentielles liées à chaque mode de vibration (flexion suivant l'épaisseur et suivant la largeur) ainsi que le mode de torsion. Nous pouvons voir la présence des fréquences modales des trois modes de vibration pour chaque son, bien que l'amplitude ainsi que

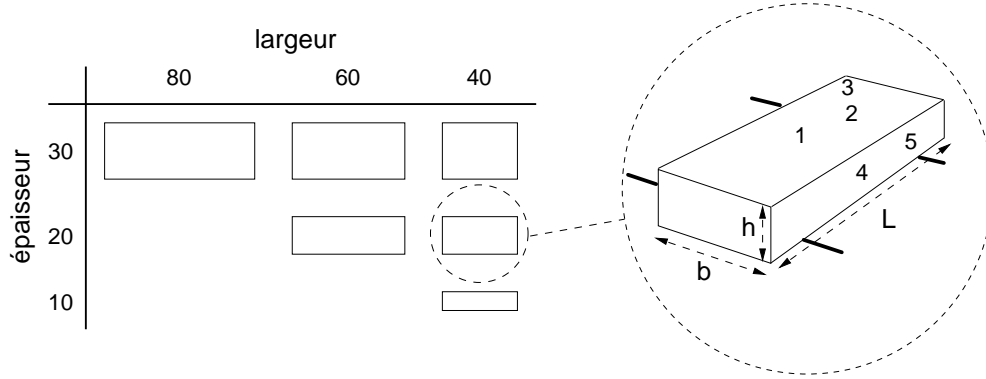


Figure 5.1 - Epaisseur (h) et largeur (b) des six barres, ces dimensions sont exprimées en millimètre. La barre ($h : 20$ mm et $b : 40$ mm) est représentée avec les 5 positions d'excitation, L représente la longueur commune à toutes les barres de 300 mm. Les deux liens élastiques sur lesquels la barre est posée, sont représentés par des traits gras. Nous prendrons comme convention de noter une barre de $h : 20$ mm et $b : 40$ mm excitée au point 5 sous la forme : $40 \times 20 \cdot 5$, c'est-à-dire $b \times h \cdot \text{position}$.

l'amortissement de chaque fréquence varie suivant les positions d'excitation.

Nous pouvons remarquer que pour les modes de flexion d'une barre libre-libre, il n'existe pas de relation harmonique entre les partiels et la fréquence fondamentale :

$$f_{Flex}(n) = \frac{\pi h}{8L^2\sqrt{12}} \sqrt{\frac{E}{\rho}} (3.011^2, 5^2, 7^2, \dots, (2n + 1)^2) \quad (5.0)$$

avec 3.011^2 le coefficient multiplicatif pour le premier partiel et pour $n \geq 2$, le coefficient devient une expression générale $(2n + 1)^2$ avec n le numéro du partiel.

Pour les modes de torsion d'une barre libre-libre, au contraire, il existe une relation harmonique entre les partiels et la fréquence fondamentale :

$$f_{Tors}(n) = \frac{n\alpha}{2L} \sqrt{\frac{E}{2\rho(1 + \nu)}} \quad (5.0)$$

avec L la longueur, h la largeur ou l'épaisseur, E le module de Young, ρ la densité, ν le coefficient de Poisson, α fonction monotone du rapport épaisseur/largeur et n le numéro du mode. Le développement des calculs associés aux équations du mouvement d'une barre est présenté en annexe §A, p. 222.

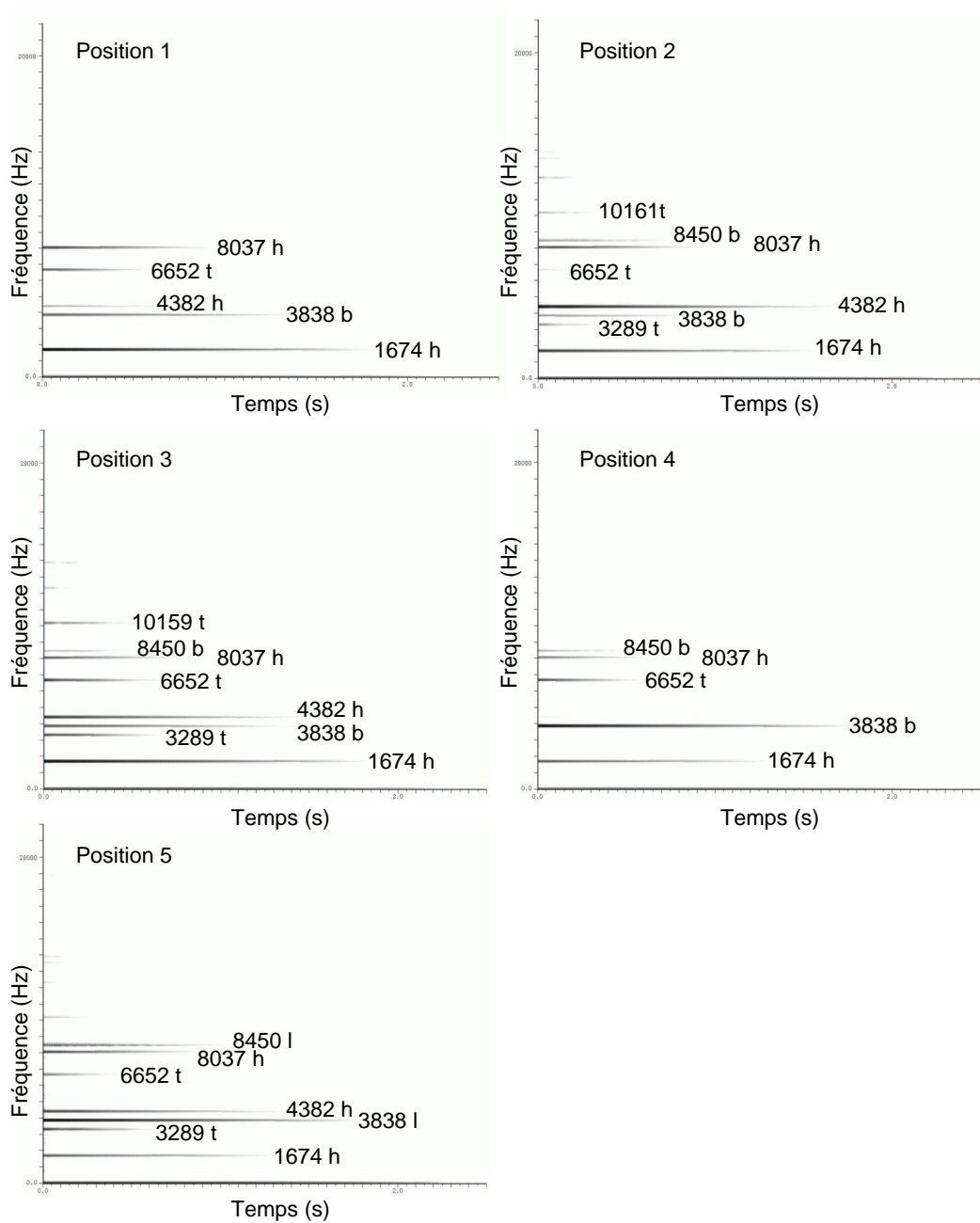


Figure 5.2 - Spectrogrammes de la barre 80×30 excitée aux 5 positions. Un spectrogramme représente la variation des fréquences en fonction du temps, l'amplitude est représentée par la noirceur de la raie spectrale. Les fréquences les plus saillantes sont marquées à côté des raies. Chaque fréquence est associée à son mode de vibration : modes de vibration de flexion (h suivant l'épaisseur, b suivant la largeur), mode de torsion t. Les modes de vibration sont estimés par la méthode des éléments finis.

5.2.2 Enregistrements

Chacun des 30 sons (6 barres x 5 positions) a été enregistré en chambre anéchoïque directement sur le disque dur d'une station NeXt grâce au logiciel Max [IRC93] à une fréquence d'échantillonnage de 44.1 kHz en 16 bit. Le microphone (Schöeps série Colette, capsule MK6 en position omnidirectionnelle) servant à l'enregistrement a été placé à 8 cm de hauteur du centre de chaque barre. Les barres sont excitées à l'aide d'un marteau d'impact Brüel & Kjær Type 8202 nous permettant de visualiser le signal d'excitation et donc de contrôler la force de l'impact sur un oscilloscope Helwett Packard (3561A Dynamic signal analyzer).

5.2.3 Protocole expérimental

Les auditeurs visualisent sur l'écran d'un ordinateur 30 « boîtes sons » correspondant aux 6 barres et leurs 5 positions de frappe, voir figure 5.3. Chaque boîte peut être déplacée ou bien jouée de manière interactive sur l'écran pour former les classes de son.

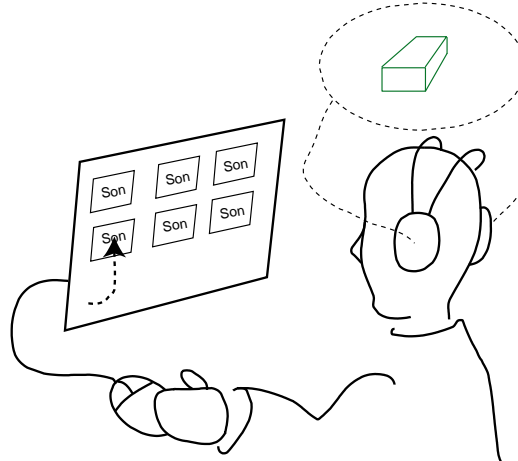


Figure 5.3 - Interface d'expérimentation pour la classification des sons. Les auditeurs peuvent déplacer les sons et les écouter de manière interactive.

Trois instructions différentes ont été données à trois groupes de participants différents pour effectuer la tâche de classification :

- z Regroupez à l'écran les boîtes correspondant aux sons qui vous paraissent similaires. « Classification non-orientée » [1]
- z Regroupez les sons qui ont été produits par la même barre. « Classification orientée » [2]

-
- z Faites 6 classes avec 5 sons chacune correspondant aux 6 barres x 5 positions d'excitation. « Classification contrainte » [3]

Avant la deuxième tâche de classification, nous avons procédé à une démonstration sonore d'une barre frappée en différents points sans spécifier les positions d'excitation. Cette barre possède les mêmes caractéristiques que les autres (même matériau, même longueur), mais elle n'est pas incluse dans le corpus de notre expérience. Pour la troisième tâche, nous avons effectué la même démonstration en spécifiant cette fois-ci, les 5 positions d'excitation utilisées lors de l'enregistrement des sons.

■ **Hypothèses liées aux consignes** Ces trois consignes nous permettent de tester la manière dont les auditeurs se représentent ces événements sonores en fonction d'une orientation plus ou moins contrainte de la tâche de classification. Dans un premier temps, la condition non-orientée a été présentée. La question était : sans aucune information donnée a priori sur les sources, les auditeurs extraient-ils naturellement des informations propres à chaque barre ? (perception directe). Dans ce cas, les informations pertinentes seraient les invariants au niveau du stimulus proximal qui spécifient le stimulus distal sans ambiguïté. Puisque les résultats ne révélaient pas un groupement conforme aux barres, il semblait nécessaire de donner une information de plus en plus complète sur les sources – conditions « orientée » et « contrainte » – afin que les auditeurs puissent traiter, voire inférer l'information auditive pertinente spécifiant chaque barre (perception indirecte). Les invariants seraient cette fois-ci à relier à des processus perceptifs de construction de l'identité de barre. Il se peut également que les résultats révèlent des processus cognitifs de stratégie inférentielle à partir des données perceptives. La variation des consignes nous donne la possibilité d'étudier ces différentes hypothèses.

5.2.4 Matériel

Tous les participants ont passé l'expérience dans une cabine à double paroi Soluna S1. Les sons ont été diffusés grâce au logiciel Max [IRC93] sur une station NeXT équipée d'une carte ISPW. La conversion numérique/analogique est gérée par un convertisseur ProPort. Les sons sont amplifiés par un amplificateur stéréo Canford et diffusés en écoute diotique – même signal aux deux oreilles – à travers un casque Sennheiser HD 520 II. Les sons ont été diffusés à un niveau de 70 dB. L'interface de réponse ainsi que la collecte des résultats sont gérées par le logiciel PsiExp [Smi95].

5.2.5 Participants

Les participants sont répartis en trois groupes différents correspondant aux trois conditions expérimentales. Ils proviennent tous de la région parisienne et ont été recrutés à l'aide d'une base de données dans laquelle ils se sont inscrits. Ils ont été rémunérés.

5.3 Analyse des classifications

Suite aux expériences de classification, nous avons mis en œuvre des analyses afin de rendre compte des classes effectuées par les participants. L'analyse par groupements hiérarchiques des données de classification nous permet de comprendre la structure des classes formées par les participants. L'analyse des données de classification s'effectue en plusieurs étapes successives :

- Analyse des données brutes.
- Transformation des partitions individuelles en matrices d'incidence.
- Analyse des différences interindividuelles à partir des partitions effectuées par les participants.
- Matrice globale de cooccurrences par moyennage des matrices d'incidence individuelles pour chaque expérience de classification.
- Analyse par groupements hiérarchiques.
- Estimation de la partition optimale pour chacune des trois expériences de classification et analyse de la structure hiérarchique des classes.

Le détail des méthodes d'analyse est donné dans le chapitre sur les proximités, §3.2.3, p. 71 et §3.4, p. 84.

5.3.1 Analyse des données brutes

Les résultats des classifications sont sous la forme de partitions effectuées par les participants pour chacune des trois conditions. Une partition est l'ensemble des classes et leurs sons associés. Dans un premier temps, il est intéressant d'examiner la distribution du nombre de classes faites par les participants suivant les différentes conditions, voir figure 5.4. Il faut noter que la distribution des classes pour la condition « contrainte » n'est pas représentée sur cette figure, car le nombre de classes imposé est six. En moyenne, pour la condition « non-orientée », le nombre de classes est de $\mu = 6.2$, variant de 4 à 15 classes. Dans le cas de la condition « orientée », la moyenne est de $\mu = 8.3$ et le nombre de classes varie entre 2 et 13 classes. Un test de Student entre

le nombre de classes effectuées par chaque participant entre les deux conditions montre qu'une différence significative ($t(df) = 56.276, p = 0.03$) existe entre les deux moyennes. Nous observons donc une tendance à faire moins de classes pour la condition « non-orientée » que pour la condition « orientée » et qui ne correspond pas au nombre des 6 barres utilisées pour produire les sons.

Dans le cas de la condition « non-orientée », le fait de ne donner aucune indication mise à part la similarité perceptive, semble indiquer que les participants différencient les sons beaucoup plus que dans la condition « orientée ». Cette stratégie doit être mise en regard avec les analyses de la structure des classifications que nous présentons dans la suite du texte. De plus, même si pour la condition « non-orientée », la moyenne est de 6 classes, nous devons regarder si ce nombre de classes reflète des classes communes à travers les participants.

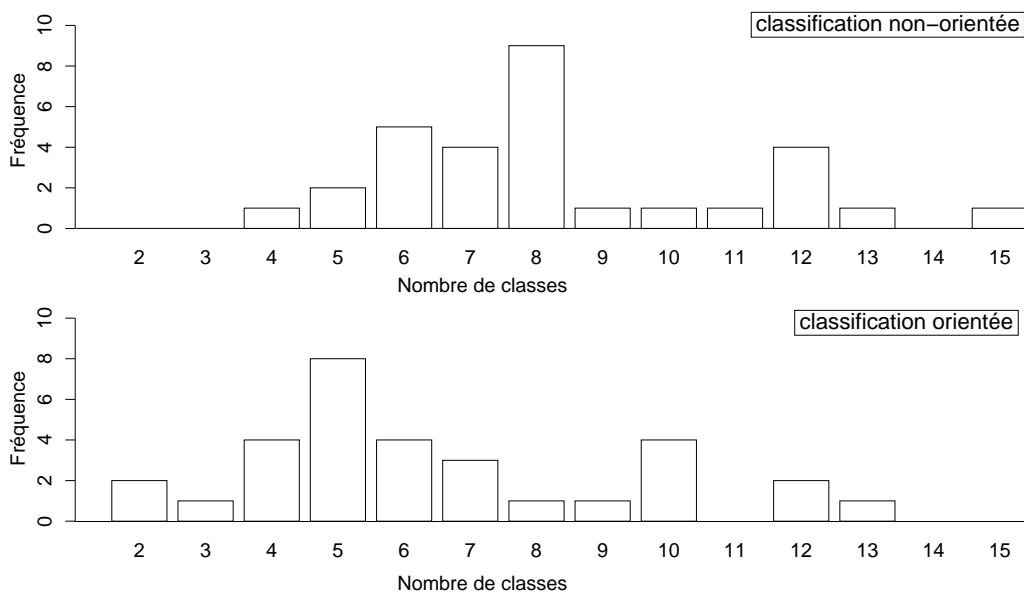


Figure 5.4 - Nombre de classes effectuées par les participants pour la condition « non-orientée » et la condition « orientée ». La condition « contrainte » n'a pas été représentée car le nombre de 6 classes est imposé.

5.3.2 Analyse des différences interindividuelles

Nous allons tout d'abord étudier les différences interindividuelles entre les classifications effectuées par les participants. Cette analyse sert à identifier des groupes de personnes ayant des patterns de réponses similaires mais aussi

des individus effectuant des classifications très différentes par rapport aux autres participants. Elle nous permet de mettre en valeur des effets possibles d'expertise tels que la pratique musicale ou l'ingénierie sonore ou bien des profils de réponses.

Pour mesurer la concordance entre les partitions individuelles pour chacune des trois conditions, nous avons utilisé l'indice de Rand ajusté, voir §3.5, p. 105, qui nous permet de calculer directement la concordance sur les partitions (donc les données brutes) plutôt que de calculer une corrélation entre les matrices individuelles d'incidence. Cette mesure de concordance est calculée entre chaque paire de participants afin d'obtenir une matrice de dissemblance entre les classifications individuelles. Cette matrice a été obtenue pour chacune des trois conditions.

En utilisant l'indice de Rand standard, les individus qui sortent de ces analyses sont généralement ceux qui ont effectué une partition dont le nombre de classe diffère nettement des autres (par exemple le plus petit nombre et le plus grand nombre). L'indice du nombre de classes n'est pas un facteur pertinent dans la mesure où il ne reflète pas forcément une tendance à répondre très différemment des autres individus. Comme nous le verrons ces personnes peuvent se trouver à des niveaux différents de catégorisation.

De manière générale, pour les trois conditions, aucun groupe tranché d'individus ne se dégage des analyses par groupements hiérarchiques (méthode moyenne), voir figure 5.5. Nous n'observons pas de groupes dont les dissemblances à l'intérieur de chacun soient faibles et dont les dissemblances entre les groupes soient élevées (groupes se joignant à un niveau élevé de l'arbre hiérarchique). Pour les trois conditions, il existe une continuité dans les niveaux de regroupement des classes.

Dans les conditions « non-orientée » et « orientée », quelques participants isolés semblent se démarquer de chaque ensemble de participants : 3 pour la condition « non-orientée » (sujet 5, sujet 1 et sujet 25), 3 pour la condition « orientée » (sujet 21, sujet 28, sujet 5). Cette démarcation se traduit par une discontinuité faible dans les niveaux de regroupements pour chacun des arbres hiérarchiques, mais qui reste difficilement interprétable. Pour compléter ces analyses, les arbres additifs sur les matrices de concordance associées aux deux conditions (voir annexe §C.1, p. 236) ne permettent, pas elles, aussi d'écarter ces quelques participants au vu de la structure compacte de ces arbres. C'est pourquoi, nous avons fait le choix de garder les résultats de l'ensemble des participants à ces trois expériences de classification.

Ces analyses montrent qu'il existe une variabilité entre les résultats des participants mais qui ne permet pas de les discriminer.

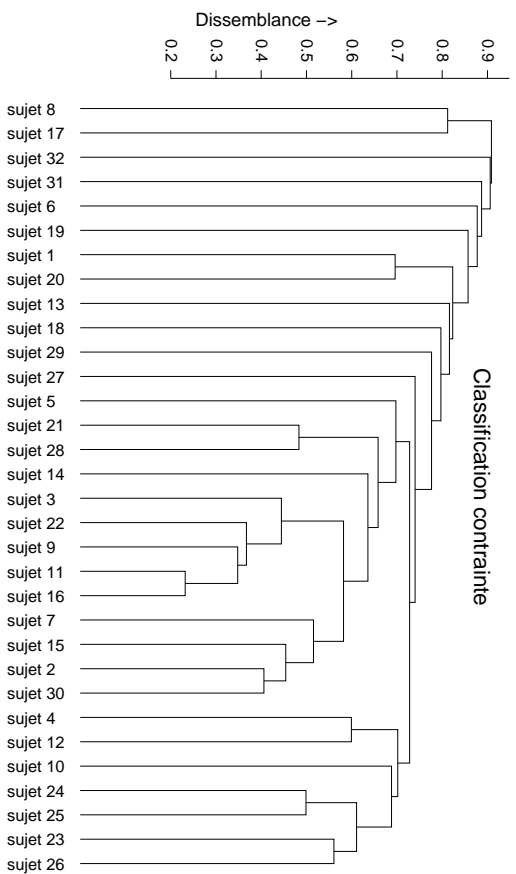
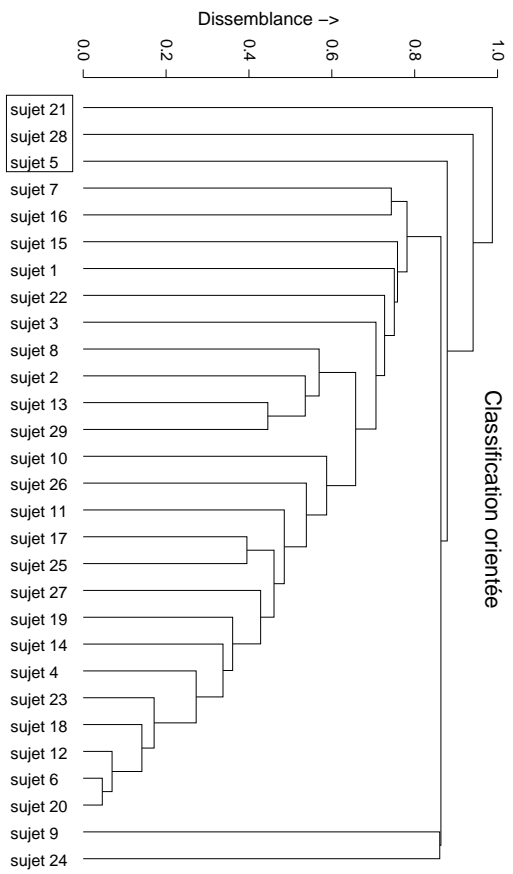
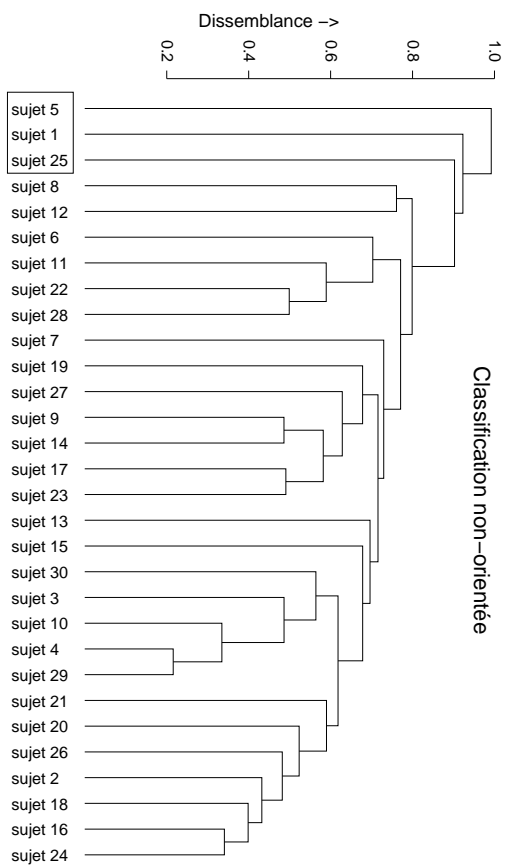


Figure 5.5 - Groupements hiérarchiques sur les matrices interindividuelles des différentes conditions : classifications non-orientée, orientée, contrainte. Les sujets, dont les partitions semblent s'écarter des autres participants, sont encadrés.

5.3.3 Analyse des groupements collectifs

Suite aux analyses précédentes, nous avons construit pour chacune des 3 conditions une matrice globale résumant la similarité entre chaque paire de sons. A partir de chacune des matrices globales, nous avons procédé à une analyse par groupements hiérarchiques (méthode moyenne) pour chaque condition. Le choix de cette analyse a déjà été introduit dans le chapitre §3, p. 67.

Cette analyse représente les données de classifications sous la forme de classes emboîtées. Nous partons de l'hypothèse que cette hiérarchie est le reflet du niveau de catégorisation dans lequel l'auditeur se place, voir §2.7.4.3. En effet, lors d'une tâche de classification, un auditeur peut se placer à un niveau d'abstraction très général en créant un nombre de classes peu élevé. Ces catégories générales sont peu informatives dans le sens où les objets partagent peu de propriétés – niveau superordonné –, voir §2.7.4.3 et la figure 5.6. Une autre personne peut former beaucoup de groupes et au contraire se placer à un niveau d'abstraction spécifique – niveau subordonné – où les objets partagent beaucoup de propriétés communes mais peuvent être difficilement différenciés. Nous cherchons pour chaque condition le niveau de base, le niveau intermédiaire entre ces deux niveaux d'abstraction, nous permettant de résumer la catégorisation effectuée par les auditeurs au niveau le plus « utile » pour la perception quotidienne. Cette partition optimale reflète les propriétés pertinentes extraites par les auditeurs lors de leur tâche de classification.

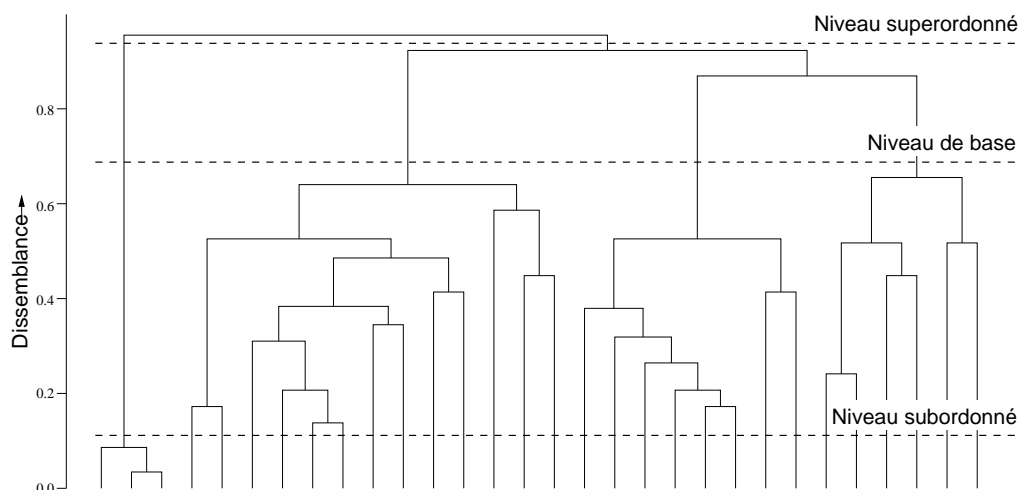


Figure 5.6 - Exemple de niveaux d'abstraction potentiels visualisés sur arbre par groupements hiérarchiques.

5.3.3.1 Analyse globale

Pour chacune des trois conditions, nous avons analysé les classifications par groupements hiérarchiques sur les trois matrices de dissemblance. Les groupements hiérarchiques pour chaque condition sont présentés sur les figures 5.7, 5.8 et 5.9, dont nous faisons une analyse détaillée.

■ **Mesure de distortion** Pour mesurer la bonne représentation des données de départ, nous regardons la distortion entre ces arbres et les matrices de cooccurrences de départ. Elle est évaluée par une mesure de corrélation cophénétique, voir §3.4.5.1. La corrélation cophénétique est calculée entre les matrices de cooccurrences et les distances ultramétriques des arbres hiérarchiques. Le tableau 5.1 résume les corrélations pour chacune des trois conditions. Les trois corrélations sont élevées même si pour la condition « contrainte », celle-ci est plus faible. Dans cette condition, le nombre de classes à effectuer est imposé aux participants. Il semble que cette contrainte induise plus de variabilité dans les réponses et se traduise par une représentation moins bonne des données de départ. Cette remarque doit être complétée par les analyses de la structure des arbres.

Condition	Non-orientée	Orientée	Contrainte
Corr. cophénétique	0.93	0.96	0.91

Tableau 5.1 - Corrélation cophénétique entre les distances ultramétriques de chaque arbre et les matrices de dissimilarité respectives.

■ **Analyse de la structure** L'analyse de la structure de chacun des trois arbres hiérarchiques nous permet d'évaluer :

- z L'émergence de catégories perceptives dans les classifications effectuées par les auditeurs par estimation notamment d'une partition optimale.
- z La structuration de ces catégories.
- z Les propriétés qui sous-tendent ces catégories.

Nous avons évalué la structure des hiérarchies suivant deux critères : un critère statistique et un critère de correspondance à la catégorie géométrique de l'objet vibrant, pour chacune des trois expériences.

5.3. ANALYSE DES CLASSIFICATIONS

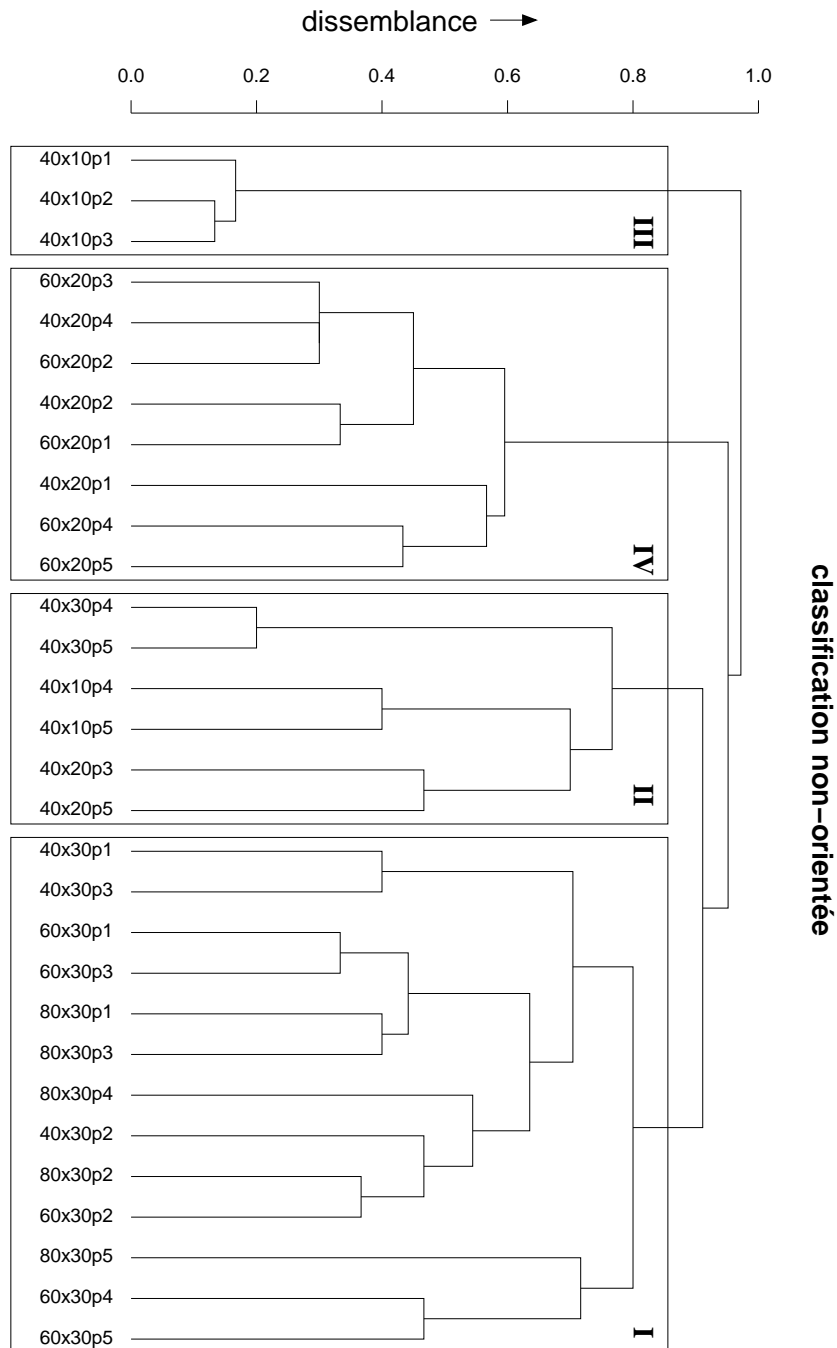


Figure 5.7 - Analyse par groupements hiérarchiques (méthode moyenne) pour la condition non-orientée. Les rectangles et les chiffres romains associés indiquent les 4 classes du niveau de base. Ces informations sont expliquées dans la suite du texte.

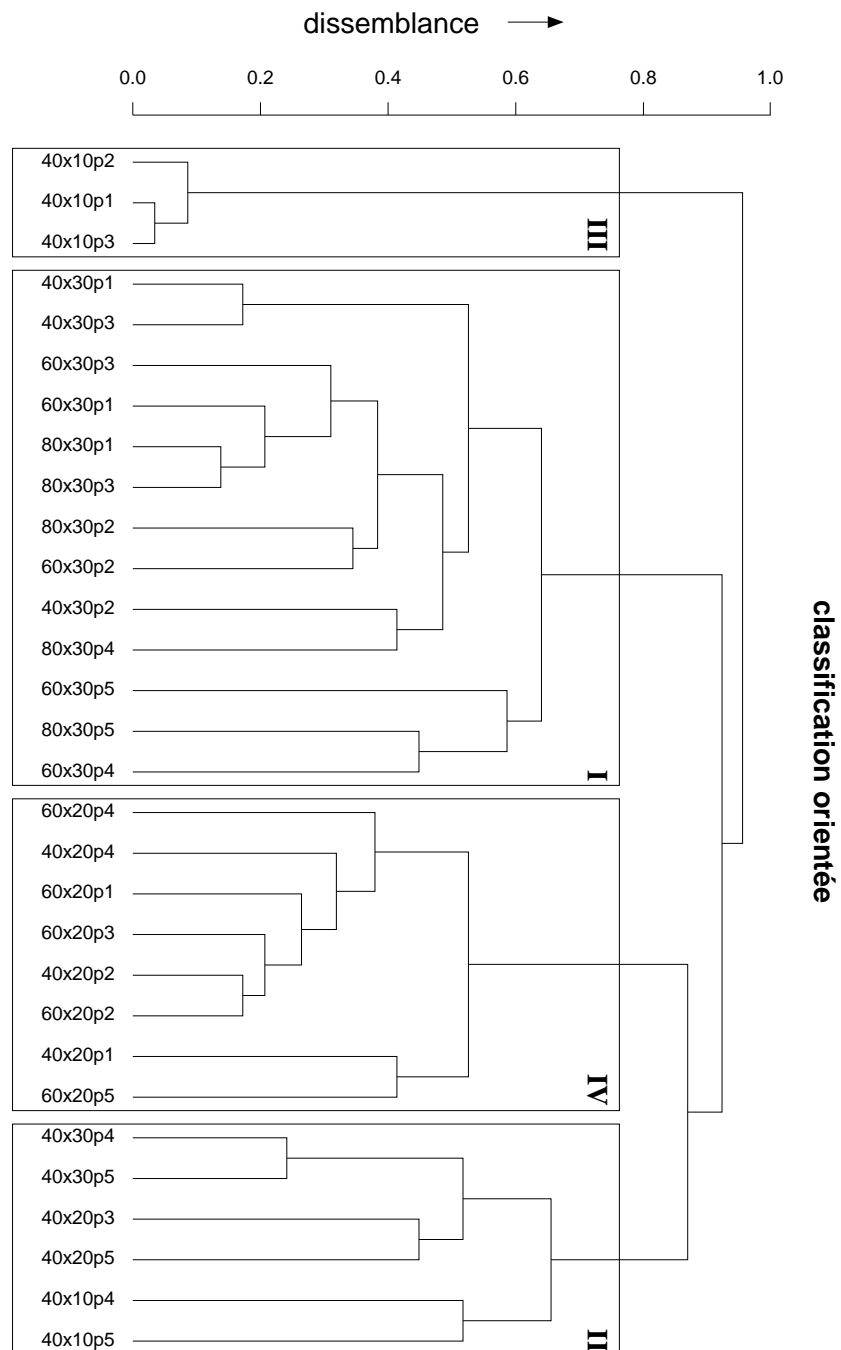


Figure 5.8 - Analyse par groupements hiérarchiques (méthode moyenne) pour la condition non-orientée. Les rectangles et les chiffres romains associés indiquent les 4 classes du niveau de base. Ces informations sont expliquées dans la suite du texte.

5.3. ANALYSE DES CLASSIFICATIONS

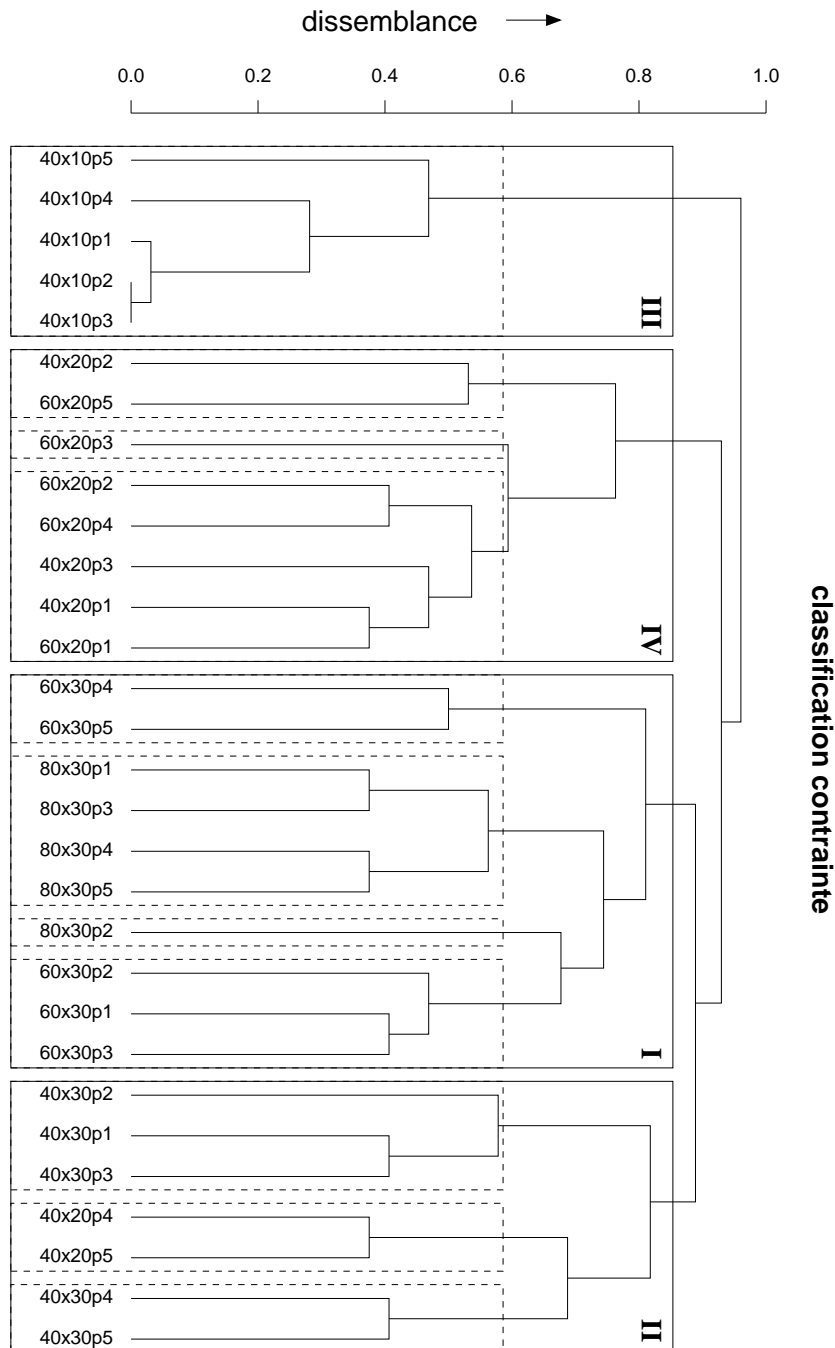


Figure 5.9 - Analyse par groupements hiérarchiques (méthode moyenne) pour la condition non-orientée. Les rectangles et les chiffres romains associés indiquent les 4 classes retenues. Les 11 classes du niveau de base sont représentées par des rectangles dont les traits sont discontinus. Ces informations sont expliquées dans la suite du texte.

■ **Classes émergentes** Dans un premier temps, nous voulons connaître s'il existe une partition commune émergente, à travers un groupe de participants pour chacune des conditions. Cette partition optimale refléterait des catégories qui représentent le niveau de base. Le but est de savoir si les auditeurs ont pu extraire des invariants structuraux propres à chaque barre et donc si les classes formées vont refléter des catégories propres à chacune des six barres.

A partir d'un arbre hiérarchique, des classes émergentes existent lorsque les dissemblances entre les membres propres à chaque classe sont faibles comparativement aux dissemblances entre les classes considérées. Dans ce cas, nous observons des classes dont les membres se regroupent à un niveau faible dans la hiérarchie de l'arbre et se rejoignent entre elles à un niveau nettement plus haut de dissemblance.

Il est souvent délicat d'estimer le niveau optimal de cette façon, notamment pour la condition « contrainte », figure 5.9. C'est pourquoi nous avons proposé une nouvelle méthode statistique pour estimer la partition optimale, car il existe peu de solutions proposées.

■ **Partition optimale et niveau de base** Nous avons développé une nouvelle méthode statistique afin d'estimer la partition optimale, voir §3.4.5.4 pour une présentation détaillée. Cette technique est basée sur le bootstrap et l'indice de Rand ajusté, permettant de comparer des partitions.

Les figures 5.10, 5.11 et 5.12, représentent pour chacune des conditions, l'évolution de l'indice de Rand ajusté (moyenne et écart type) à des niveaux de hiérarchie différents, c'est à dire à un niveau de partition de 2 classes, 3 classes ..., 15 classes.

La valeur maximale de la moyenne de l'indice de Rand ajusté et le minimum de son écart type, nous donnent une estimation de la partition optimale. Ces deux critères indiquent qu'à ce niveau de partition, les classes sont les plus récurrentes à travers l'ensemble des participants. Cette partition de classes est une estimation de la partition optimale.

Les résultats de cette méthode montrent que pour les conditions « orientée » et « non-orientée », un niveau de partition de quatre classes émerge nettement des classifications. Par contre, pour la condition « contrainte », les deux indices suggèrent une partition optimale comportant un plus grand nombre de classes, autour de 11 classes (de 10 à 12). Dans les trois conditions, nous ne retrouvons pas un nombre de 6 classes correspondant aux barres.

Cette analyse montre que dans la condition « contrainte », il existe une grande variabilité entre les participants puisqu'ils ont chacun effectué 6 classes de 5 sons. Cette variabilité peut être attribuée à des stratégies va-

5.3. ANALYSE DES CLASSIFICATIONS

riables entre chaque participant. Par contre, comme l'ont montré les analyses des différences interindividuelles, ces stratégies ne peuvent pas être associées à des groupes de participants ayant une stratégie propre. Pour les conditions « orientée » et « non-orientée », le nombre de 4 classes ne correspond pas aux moyennes du nombre de classes effectuées par les participants. Dans la condition « orientée », le nombre moyen de classes est de 6 contrairement à la condition « non-orientée » qui est de 8. Ceci indique que dans la condition « orientée », les classes effectuées par les participants sont plus stables que dans les autres conditions. Ainsi le niveau hiérarchique de la partition de ces 4 classes est inférieur aux deux autres conditions, voir les figures 5.8, 5.7 et 5.9.

Nous retiendrons de ces analyses, un niveau de partition optimale de 4 classes pour les conditions « orientée » et « non-orientée » et un niveau de partition optimale entre 10 et 12 classes pour la condition « contrainte ». Nous allons maintenant mettre en relation ces partitions avec la structure perceptive de celle-ci.

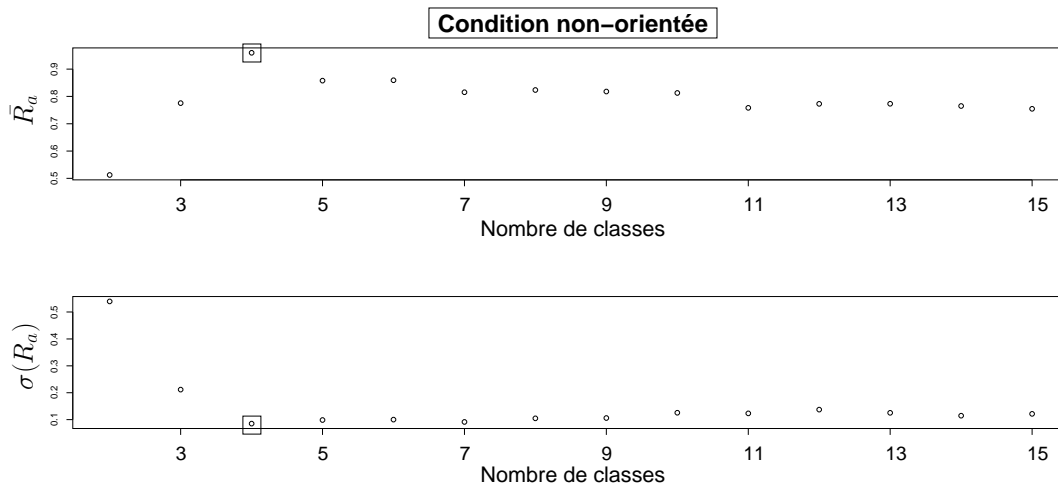


Figure 5.10 - Variation de la moyenne de l'indice de Rand ajusté \bar{R}_a et de son écart type $\sigma(R_a)$ suivant le nombre de classes (de 2 à 15) pour un bootstrap de 1000 tirages à partir des 29 matrices d'incidence. La valeur maximale de \bar{R}_a et la valeur minimale de $\sigma(R_a)$, toutes les deux encadrées, correspondent à une partition optimale de 4 classes.

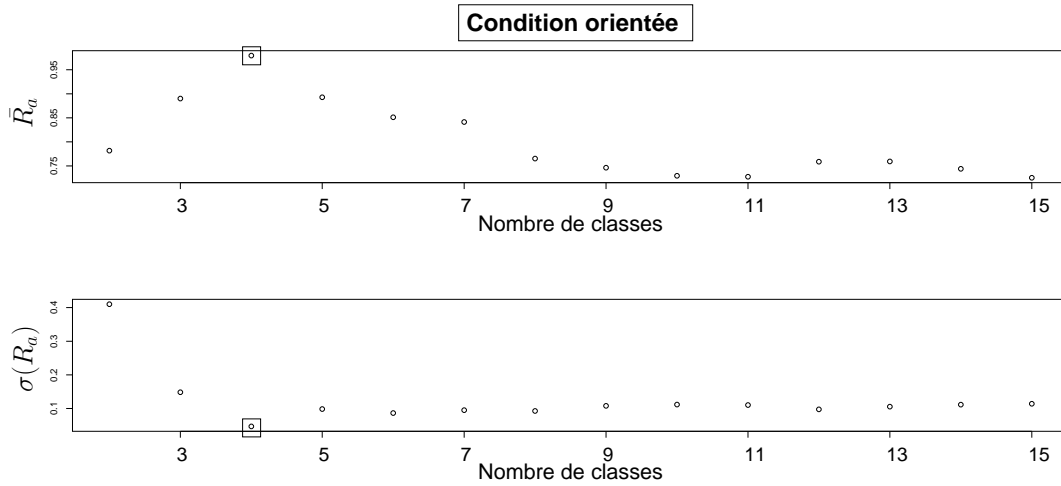


Figure 5.11 - Variation de la moyenne de l'indice de Rand ajusté \bar{R}_a et son écart type $\sigma(R_a)$ suivant le nombre de classes (de 2 à 15) pour un bootstrap de 1000 tirages à partir des 30 matrices d'incidence. La valeur maximale de \bar{R}_a et la valeur minimale de $\sigma(R_a)$, toutes les deux encadrées, correspondent à une partition optimale de 4 classes.

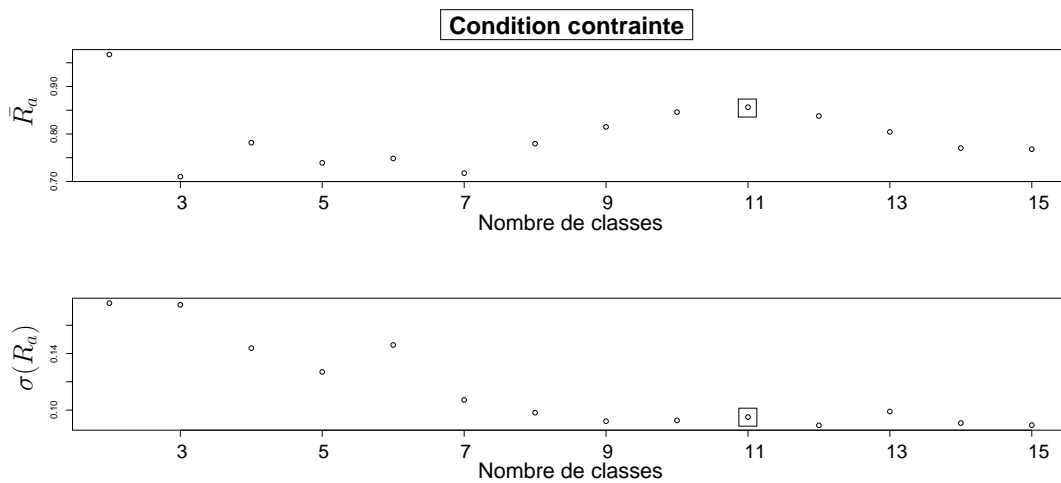


Figure 5.12 - Variation de la moyenne de l'indice de Rand \bar{R}_a et son écart type $\sigma(R_a)$ et de la corrélation bisériale suivant le nombre de classes (de 2 à 15) pour un bootstrap de 1000 tirages à partir des 32 matrices d'incidence tirées de l'expérience de classification « orientée » du chapitre 5. La valeur maximale de \bar{R}_a et la valeur minimale de $\sigma(R_a)$, toutes les deux encadrées, correspondent à une partition optimale de 11 classes.

5.3.3.2 Catégories et géométrie

L'analyse précédente nous donne une estimation statistique de la partition optimale. Nous devons la confronter à des critères externes, c'est-à-dire à l'information que nous avons sur nos stimuli pour nous permettre d'interpréter la structure de ces partitions optimales. Nous comparons ici les résultats des différentes conditions.

Dans un premier temps, nous comparons la structure des trois conditions à un niveau de partition de 4 classes même si, pour la condition « contrainte », celle-ci ne correspond pas à la partition optimale. Les classes associées à ce niveau sont représentées par des rectangles pleins groupant les sons propres à chacune des classes sur les figures 5.7, 5.8 et 5.9. Pour les conditions « non-orientée » et « orientée », cette partition peut être associée au niveau de base dans le processus de catégorisation des sons, sur l'ensemble des participants. Dans un deuxième temps, nous étudierons plus particulièrement le niveau de partition optimale propre à la classification « contrainte ». Pour la condition « contrainte », les classes associées à ce niveau hiérarchique sont représentées par des rectangles dont les traits sont discontinus, voir figure 5.9.

■ **Synthèse des résultats** Afin de comparer les différentes conditions, nous avons résumé les groupements hiérarchiques de chaque condition en prenant comme référence le niveau de partition de 4 classes. Nous avons donc coupé chacun des trois arbres hiérarchiques au niveau de 4 classes pour en extraire la structure.

Par contre, il est difficile d'inclure dans cette représentation les niveaux inférieurs des arbres hiérarchiques correspondant à la structure interne de ces 4 classes. En effet aucun critère ne nous permet de couper à un niveau arbitraire les arbres hiérarchiques sur ces niveaux inférieurs. Néanmoins, pour la condition « contrainte » nous avons inclus la partition optimale correspondant à un niveau de 11 classes. Nous analyserons séparément cette information.

Cette structure hiérarchique réduite est projetée sur la matrice des stimuli afin de lier la structure des classes à la géométrie et aux positions d'excitation des 6 barres. Cette synthèse des résultats se trouve sur la figure 5.13.

■ **Catégories** Si nous observons la structure des 4 classes pour chacune des trois conditions, il existe une forte correspondance entre ces groupements perceptifs et la géométrie des barres. En effet, ces catégories sont liées soit à des épaisseurs soit à des largeurs communes des barres qui ont produit les sons. Nous préférons parler maintenant de catégories, car ces groupements perceptifs sont liés à la structure géométrique des barres, donc aux propriétés physiques de la source sonore.

Les quatre catégories peuvent être décrites de la façon suivante :

Catégorie I ▸ Les sons associés à cette catégorie proviennent de barres qui ont en commun une épaisseur de 30 mm.

Catégorie II ▸ Les sons associés à cette catégorie proviennent de barres qui ont en commun une largeur de 40 mm.

Catégorie III ▸ Nous retrouvons dans cette catégorie des sons issus de barres d'épaisseur 20 mm.

Catégorie IV ▸ Cette catégorie est composée de sons provenant de la barre d'épaisseur 10 mm.

Nous détaillons ci-dessous l'analyse de ces catégories.

5.3.3.3 Discussion

La première remarque que nous pouvons faire est le fait que les auditeurs n'ont pas regroupé les sons provenant d'une même barre. Il n'existe pas 6 catégories correspondant aux 6 barres \times 5 positions (à l'exception de la barre 40×10 pour la condition « contrainte »). De manière générale, les résultats des trois classifications donnent des résultats, au niveau de la partition optimale, très proches même, s'il existe des différences dans la structure fine des 4 catégories retenues entre les trois classifications. Nous pouvons remarquer tout de même que, pour la condition « contrainte », certaines des quatre catégories retenues varient par rapport aux deux autres classifications.

■ **Structure interne aux classes** L'analyse des hiérarchies pour chaque condition nous donne la possibilité d'explorer les structures sous-jacentes aux classifications.

Condition « non-orientée » et « orientée » Il existe une structure interne pour les conditions « non-orientée » et « orientée » très semblable. En observant la figure 5.13 pour ces deux conditions, nous remarquons que la catégorie **III** se démarque très nettement des autres classes. Elle est associée à une barre fine d'épaisseur 10 mm. De par son épaisseur, elle se rapproche d'une plaque (voir Lakatos et coll. [LMC97]) et possède une fréquence fondamentale beaucoup plus basse que les autres barres. De plus, cette barre est la seule de cette épaisseur. Il existe un fort découplage entre les modes de vibration transverse suivant l'épaisseur et suivant la largeur de cette barre. Ceci implique une prédominance des fréquences modales transverses suivant l'épaisseur, lorsque la barre est excitée sur la face (et de même pour la largeur et l'excitation sur le côté).

5.3. ANALYSE DES CLASSIFICATIONS

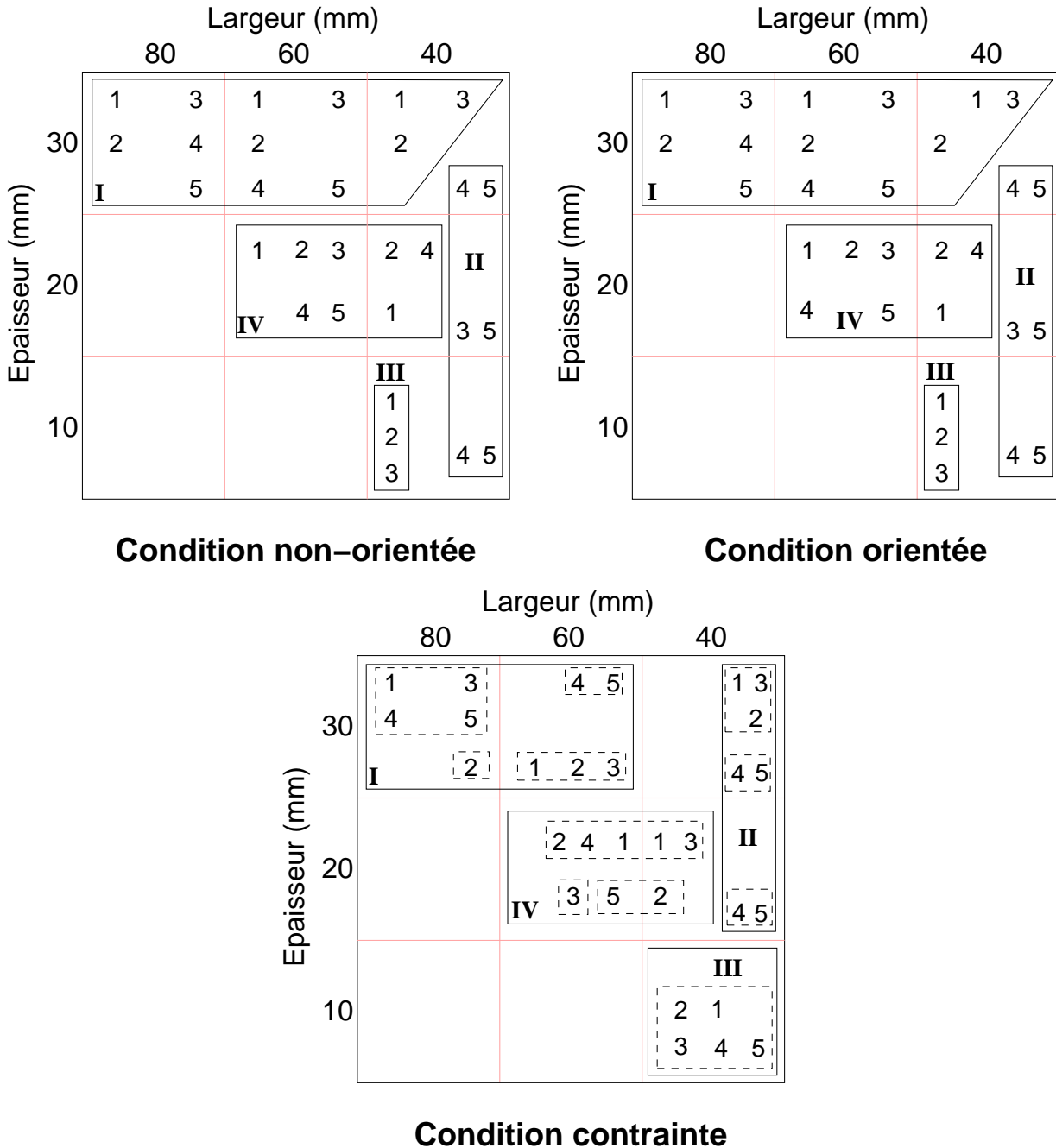


Figure 5.13 - Le résultat de l'analyse par groupements hiérarchiques pour chacune des trois classifications est superposé à la matrice. 5.1 p. 135. Chaque carré représente une barre, les numéros correspondent aux 5 positions d'excitation. Les rectangles et les chiffres romains associés indiquent les 4 classes de la partition optimale pour les conditions « non-orientée » et « orientée ». De la même façon, pour la condition « contrainte » nous avons indiqué la partition de 4 classes ainsi que la partition optimale de 11 classes représentées par des rectangles dont les traits sont discontinus.

La catégorie **II** regroupe, quant à elle, des barres de largeur 40 mm. Les positions d'excitation (4 et 5, mise à part le son 40×20.3) indiquent que les fréquences modales qui ont été privilégiées correspondent aux modes de vibration suivant la largeur 40 mm. Pour les barres 40×20 mm et 40×30, il existe un découplage conséquent entre les modes de vibration suivant la largeur et l'épaisseur, comme pour la barre 40×10 mm. Ainsi nous retrouvons les sons associés à une largeur et une épaisseur données pour une même barre dans des catégories différentes.

Pour la catégorie **IV**, nous observons que pour la barre 60×20 mm, les 5 positions d'excitation sont regroupées. Par contre, tous les sons liés aux excitations, suivant l'épaisseur, sont dans une sous-catégorie différente de ceux liés aux excitations suivant la largeur, si nous regardons les niveaux hiérarchiques inférieurs à la partition optimale. Les sons de cette barre étant associés à la barre 40×20 (position 1 2 4, suivant l'épaisseur), nous pouvons penser que les fréquences, liées au mode de vibration suivant l'épaisseur, jouent un rôle prépondérant d'un point de vue perceptif.

De la même façon que pour la catégorie précédente, la catégorie **I** regroupe les sons des barres 80×30 et 60×30 pour les 5 positions d'excitation. Cette catégorie est propre à des barres qui partagent la même épaisseur de 30 mm. Nous retrouvons les sous-catégories liées à la position de frappe sur la face ou sur le côté (à l'exception de la position 4 pour la barre 80×30).

Condition « contrainte » Les résultats montrent que les auditeurs n'ont pas regroupé les sons par barre, même en demandant spécifiquement d'effectuer la classification sur cette base. Comme pour les deux premières conditions les catégories semblent articulées autour de dimensions privilégiées (épaisseur 30 mm pour la catégorie **III**, 20 mm pour la catégorie **IV** ; largeur 40 mm pour la catégorie **II**). Mais il faut noter que contrairement aux deux premières classifications, la barre 40×10 se retrouve dans une catégorie propre et que la barre 40×30 est cette fois-ci associée suivant ces 5 positions d'excitation dans la même catégorie **II**. Même s'il semble qu'il y ait une légère influence de la consigne, les fréquences liées aux modes de vibration suivant des dimensions privilégiées (largeur ou épaisseur) semble jouer encore un rôle prépondérant. Par contre, pour les barres 40×10 et 40×30, il semble que l'information suivant les deux dimensions soit utilisée puisque nous retrouvons l'ensemble des positions d'excitation dans une même catégorie pour chacune des deux barres. Cette information liée aux deux dimensions géométriques, est comme l'ont montré Lakatos et coll. [LMC97], une source potentielle d'information pour la représentation de la géométrie des barres. Si nous regardons les valeurs du centre de gravité spectral (zone fréquentielle

où l'énergie est maximale) ou des paramètres liés à la résonance du son, ces paramètres ne permettent pas d'expliquer la catégorie propre à la barre 40×10 .

Nous pouvons souligner que, le fait d'imposer le nombre de classes correspondant aux 6 barres et leurs 5 positions d'excitation, a rendu la tâche difficile. Ce constat se reflète dans les commentaires des participants après cette tâche de classification. Cette difficulté, liée au fait d'avoir imposé le nombre de classes, implique un recouvrement moins important des partitions entre les participants. Ce constat se retrouve dans les analyses, aussi bien par arbre hiérarchique que par arbre additif (annexe §C.6, p. 241) et par analyse multidimensionnelle (annexe §C.9, p. 244), montrant dans les trois cas des structures beaucoup moins catégorielles. Même en orientant les auditeurs vers la source, ils n'ont pas réussi à extraire systématiquement du stimulus proximal des invariants propres à chaque barre, la source distale. Nous pouvons observer cet effet sur la structure de la partition optimale de 11 classes qui présente une discrimination plus importante des 4 catégories. Mais cette structure reste liée aux positions de frappe suivant l'épaisseur ou la largeur.

Comme dans les deux premières conditions, le rôle des modes de torsion (en principe favorisés par la position d'excitation 3) semble minime comparativement aux modes de flexion.

■ Remarques

Similarité Les résultats de ces trois expériences sont intéressantes du point de vue du type de similarité impliqué lors de la tâche de classification. En effet, suivant les trois consignes, nous avons tenté d'orienter les auditeurs d'abord vers une écoute des qualités intrinsèques aux sons (sans faire référence à la source), puis de plus en plus vers la source, vers une écoute des propriétés audibles des sources, en leurs donnant de plus en plus d'informations sur ces sources.

Ces résultats semblent indiquer que les auditeurs se sont avant tout focalisés sur les similarités de surface, puisque les analyses montrent une variation faible entre les résultats des expériences. La similarité, impliquée dans les tâches de classification, semble robuste, quel que soit le contexte expérimental. Ainsi, les résultats de la classification « contrainte » montrent que la similarité de surface semble irréprouvable, car même en familiarisant les participants avec des sons de barre, il existe peu de variations.

Les sons de barres de métal ne montrent pas une identité causale forte qui permettrait de guider l'auditeur dans une représentation mentale de la source

ou d'extraire des invariants structuraux. Les participants ont commenté les classes et les sons de manière informelle après l'expérience. Même s'ils ont pu associer ces sons à des sons de cloches ou des sons percussifs, il semble qu'ils n'appartiennent pas à une catégorie d'objets à part entière, ayant une valeur sémantique forte.

Catégories Dans la partie théorique, nous avons introduit différentes théories des catégories, (classique, prototypique, exemplaire). Il semble difficile de trancher sur le type d'organisation qui sous-tend les quatre catégories. La construction d'arbres additifs (voir annexe §C.2, p. 239) ne révèle pas de structure prototypique comme d'autres études ont pu le montrer [Maf99] [GCF97]. Nous pouvons penser que les sons utilisés ne véhiculent pas un sens sémantique fort, contrairement à des sons environnementaux qui induiraient une organisation des catégories autour de prototypes.

Même si l'analyse par groupements hiérarchiques impose une structure de classes, l'analyse par arbre additif confirme l'émergence de catégories. La représentation sous forme d'arbre semble nettement plus adaptée pour nos résultats car des analyses multidimensionnelles (modèle de distance euclidienne en 2 dimensions, annexe p. 242) ne dégagent aucune dimension telle que la résonance/décroissance temporelle ou le centre de gravité spectrale. Ces propriétés pourraient être des dimensions pertinentes pour ce type de sons percussifs. Ce résultat peut s'expliquer par la structure catégorielle de nos résultats. De plus, la structure des catégories apparaît plutôt s'organiser autour d'un modèle de similarité, basé sur les traits liés aux différentes hauteurs associées aux fréquences modales des barres, plutôt que de dimension.

5.3.4 Hypothèses

Suite à ces expériences, plusieurs hypothèses se dégagent. Les résultats indiquent que les auditeurs ont associé les sons à des dimensions privilégiées (épaisseur et largeur), plutôt qu'aux différentes barres. Les résultats de la condition « non-orientée » montrent que les auditeurs n'extraient pas une information propre à chaque barre, sur la base du stimulus proximal. Dans les conditions « orientée » et « contrainte », une information de plus en plus complète sur les sources, n'a pas permis d'inférer l'information auditive propre à chaque source. Ces résultats sont contraires aux conclusions des expériences de Lakatos et coll. [LMC97] qui ont montré que nous arrivons à extraire une information propre à la section géométrique des barres, sur la base d'une information sur le rapport épaisseur sur largeur. Nous retrouvons ce résultat dans la condition « contrainte » uniquement pour la barre 40×10 mm et de manière moins tranchée pour la barre 40×30. Il semble donc qu'il

5.4. MODÈLE BASÉ SUR LES ANALYSES ACOUSTIQUES

n'y ait pas une extraction directe des propriétés audibles liées à la section géométrique de la barre. Même si, dans la condition « contrainte », les auditeurs ont utilisé une information, pour certaines barres, qui semble liée aux deux dimensions, plutôt qu'à une dimension privilégiée.

Lors des tâches de classification, les participants semblent plutôt effectuer leur jugement sur la base des propriétés perceptives qui véhiculent une information soit sur la largeur soit sur l'épaisseur. Cette information peut être liée aux fréquences modales de flexion suivant la largeur ou l'épaisseur car elles sont proportionnelles respectivement à la largeur et à l'épaisseur. Suivant les positions d'excitation, nous avons privilégié les modes de flexion suivant l'épaisseur (positions 1, 2 et 3) ou suivant la largeur (positions 4 et 5) qui impliquent des fréquences modales privilégiées, suivant ces dimensions. Il apparaît donc que les auditeurs se focalisent sur les hauteurs proéminentes entendues dans chaque son qui sont provoquées par ces fréquences modales.

Dans le tableau 5.2, nous résumons les hauteurs qui peuvent être associées aux sons de chaque catégorie. Ces hauteurs sont à mettre en correspondance avec les fréquences du premier mode de flexion, suivant la largeur ou l'épaisseur. Les résultats des analyses acoustiques sont détaillés en annexe §C.4, p. 245.

Catégorie	fréquence	hauteur associée
I	1671 Hz (flexion épaisseur)	Sol # 5
II	2181 Hz (flexion largeur)	Do # 6
III	571 Hz (flexion épaisseur)	Do # 4
IV	1130 Hz (flexion épaisseur)	Do # 5

Tableau 5.2 - Correspondance entre les catégories et les hauteurs des sons associés à ces catégories. Nous avons indiqué la première fréquence modale des modes de flexion.

C'est pourquoi il nous a semblé intéressant de nous focaliser sur l'écoute analytique des différentes contributions fréquentielles associées aux différents modes de vibration, donc sous forme de traits plutôt que de dimensions. Le but est de voir si ces différentes hauteurs, entendues dans chacun des sons, pouvaient servir de cadre de comparaison (via le modèle de contraste, §3.3.3, p. 81) entre les sons, lors de la formation des classes.

5.4 Modèle basé sur les analyses acoustiques

D'après le résultat des expériences précédentes, nous posons l'hypothèse que les auditeurs utilisent des hauteurs qui sont liées aux fréquences des

premiers modes de vibration pour comparer les différents sons. Dans un premier temps, nous avons voulu étudier quelle était l'information fréquentielle disponible dans chacun des 30 sons.

Grâce à la méthode modRes nous avons pu estimer les fréquences modales de chacun des 30 sons de barres. Nous avons donc, pour chaque son, la liste des fréquences modales. A partir de l'analyse spectrale des sons, nous avons construit un modèle d'appariement de traits dérivé du modèle de contraste de Tversky, voir §3.3.3, p. 81. Notre modèle de contraste est basé uniquement sur les traits communs.

Dans notre modèle, chaque composante fréquentielle est traitée comme un trait. Ce choix est dérivé de l'hypothèse d'une écoute analytique, de la part des participants des différentes hauteurs liées à ces fréquences, dans chacun des sons. Cette écoute peut être la base de la comparaison des sons lors de la classification. Ainsi, plus deux sons ont des traits fréquents communs, plus ces sons seront considérés comme similaires.

Pour calculer la similarité entre chaque paire de sons, nous avons estimé le nombre de fréquences communes pour chacune des paires. Pour cela, nous avons comparé chacune des fréquences d'un son avec chacune des fréquences de l'autre son comme le montre la figure 5.14. A chaque fois qu'une fréquence est commune aux deux sons, un score égal à 1 est associé à la paire de sons. L'intervalle de comparaison est de 1% de la fréquence considérée, ce qui s'approche du seuil différentiel fréquentiel (la plus petite différence fréquentielle perçue). Le seuil différentiel fréquentiel varie en fonction de la fréquence entre 0.15 % à 0.85 % pour des sujets entraînés [WJG77] et entre 0.6 % et 2 % pour des sujets non entraînés [Sin85].

Par exemple sur la figure 5.14, les fréquences de 1671 Hz (son a) et 1650 Hz (son b) sont considérées comme communes, de même pour les fréquences 2143 Hz (son a) et 2150 Hz (son b). La similarité entre ces deux sons est égale à 2.

A la fin de la comparaison des deux sons, un score égal au nombre de fréquences communes donne la similarité entre ces deux sons. Le nombre de fréquences à comparer est généralement différent pour les deux sons. C'est pourquoi les fréquences communes sont pondérées par le nombre total de fréquences comparées afin de normaliser les résultats. Cette normalisation est nécessaire afin de transformer cette mesure de similarité en distance. Nous avons effectué ces comparaisons pour toutes les paires possibles, à partir des 30 sons.

Afin de comparer ce modèle aux résultats de classification, nous avons procédé à une même analyse par groupements hiérarchiques sur la base de cette mesure de similarité.

Cette analyse dérivée de cette matrice de similarité correspond à chaque

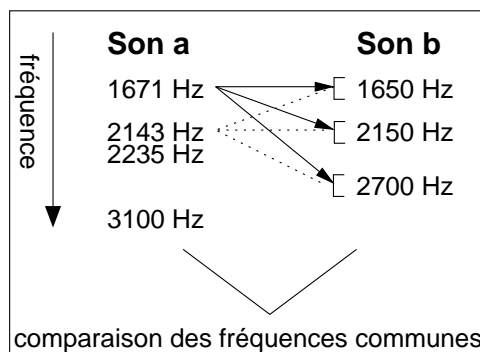


Figure 5.14 - Principe de comparaison entre les traits fréquentiels pour chaque paire de sons a et b. L'intervalle de comparaison, symbolisé par \square est de 1 % de la fréquence de référence.

barre individuelle, voir la figure 5.15. Cette structure indique que l'information acoustique nécessaire pour effectuer la tâche est présente mais que les participants ne l'utilisent pas telle quelle ou que l'information n'est pas présente d'un point de vue perceptif (par exemple si la fréquence considérée a une amplitude faible ou bien si l'amortissement est trop rapide). Pour la condition « contrainte », ce modèle pourrait expliquer la catégorie propre à la barre 40×10 mm et montrer l'influence des contributions spectrales, suivant les modes de flexion épaisseur et largeur, pour la barre 40×30 mm.

Il faut noter que l'appariement des traits n'est pas pondéré par l'amplitude et/ou l'amortissement des fréquences modales. Nous avons préféré opter dans un premier temps pour un modèle simple qui nous a permis de montrer que les fréquences modales pouvaient être une source d'information potentielle pour effectuer les tâches de classification. De plus, comme nous l'avons déjà mentionné, le pattern spectral de chacune des barres apparaît comme un bon prédicteur de la géométrie des barres.

Pour aller plus loin dans nos analyses, nous avons mené une expérience afin d'estimer quelles étaient les hauteurs audibles pour chacun des sons.

5.5 Expérience d'appariement

Suite aux analyses des expériences précédentes nous avons demandé à des auditeurs d'estimer la hauteur et la sonie pour chacune des hauteurs entendues dans chacun des sons de barre. Cette expérience nous permet d'élaborer un modèle de réponse, à partir de ces différentes hauteurs. Comme précédemment, il est élaboré sur la base du modèle de contraste de Tversky, [Tve77]. Cette modélisation sera ensuite comparée aux résultats expérimentaux des

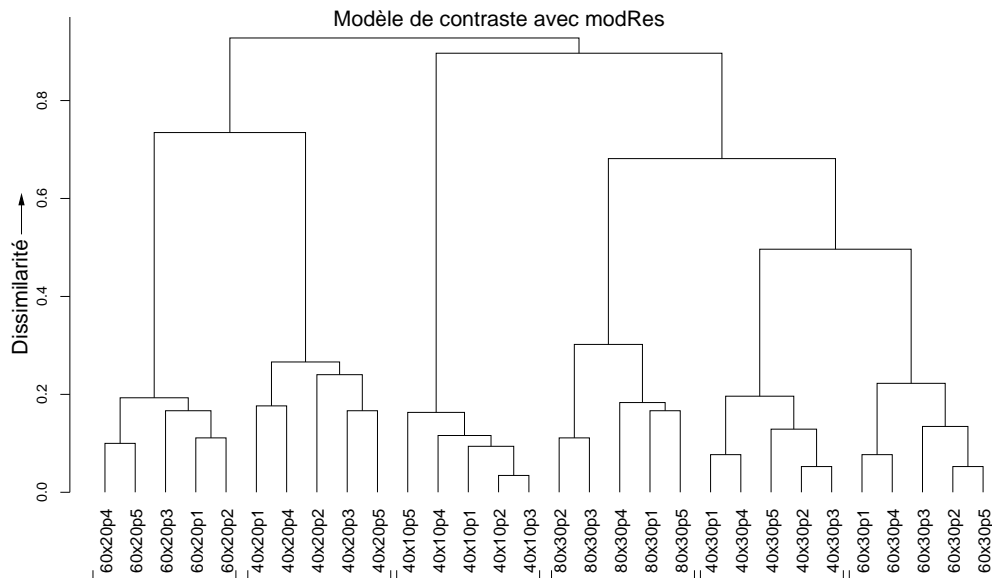


Figure 5.15 - Analyse par groupements hiérarchiques (groupement moyen) des similarités entre les 30 sons, calculées sur la base du modèle de Tversky. Ce modèle tient compte uniquement des traits communs entre chaque paire de sons ($A \cap B$ dans le modèle). Les fréquences modales de chacun des 30 sons ont été estimées à partir d'une analyse modRes.

expériences de classification.

5.5.1 Protocole expérimental

5.5.1.1 Participants

15 personnes ont participé à cette expérience (6 hommes et 9 femmes). Elles ont été recrutées dans la région parisienne et ont été rémunérées pour leur participation. La plupart d'entre elles ont été contactées grâce à une base de données sur laquelle elles se sont inscrites pour participer à des expériences de psychoacoustique, les autres personnes travaillant à l'Ircam. Les auditeurs ont déclaré n'avoir aucun problème d'audition.

5.5.1.2 Stimuli

Les sons de barre pour cette expérience sont les mêmes que ceux utilisés lors des expériences de classification. Ils nous serviront de sons de référence. Un son pur, ajusté aux sons de référence par les auditeurs, est créé avec une fonction sinus exponentiellement amortie (amortissement de -6 dB à la

moitié de l'amplitude maximale). L'attaque du son est une rampe linéaire de 20 ms. L'attaque et l'amortissement, fixes, quelle que soit la fréquence, ont été estimés par rapport à une valeur moyenne sur les analyses acoustiques. Seules la fréquence de la fonction sinus ainsi que l'amplitude peuvent varier. Les sons ont une durée proche des sons de barres, c'est-à-dire maximum 2.5 s.

5.5.1.3 Protocole

■ **Appariement** Les participants écoutent les sons de façon dichotique, voir figure 5.16. Le son pur est diffusé à l'oreille gauche. Les auditeurs ajustent sa fréquence ainsi que son amplitude afin d'apparier sa hauteur et sa sonie avec l'une des hauteurs entendues dans le son de barre diffusé à l'oreille droite. Ils ont dû effectuer cet ajustement pour chacune des hauteurs entendues dans chacun des 30 sons. Nous avons préféré diffuser les sons de manière dichotique pour ne pas avoir d'interaction au niveau de la membrane basilaire, entre ces deux sons percussifs.

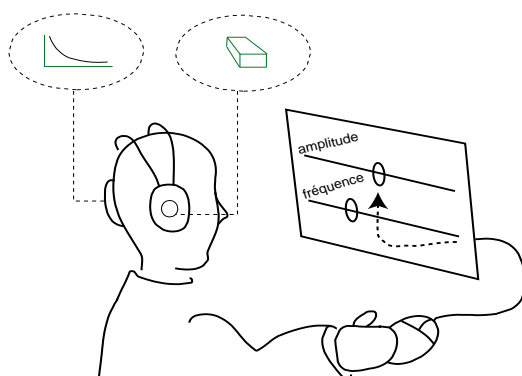


Figure 5.16 - Interface pour l'appariement des sons en hauteur. Les sons sont diffusés de manière dichotique, à l'oreille gauche les participants écoutent la sinusoïde amortie et à l'oreille droite le son de barre. Des curseurs permettent de régler la fréquence et l'amplitude du son pur afin que sa hauteur et sa sonie soient équivalentes à la hauteur entendue dans le son de barre.

Dans un premier temps, les participants règlent grossièrement la fréquence ainsi que l'amplitude du son pur, à l'aide de curseurs distincts. Ensuite, ils peuvent régler finement sa fréquence ainsi que son amplitude sur deux autres curseurs distincts. Ils ont la possibilité d'écouter le son pur, seul, ainsi que le son de barre, seul. Lorsque l'auditeur règle au moins deux sons purs, il peut écouter un arpège de ces sons purs. Cet arpège permet de comparer les sons purs appariés avec le son de référence.

Avant de commencer ces appariements, les participants ont entendu les 30 sons de barres afin de se familiariser avec ces sons.

Comme l'expérience est longue (d'une durée moyenne de 1h30), nous avons vérifié que les auditeurs ne subissaient aucune augmentation temporaire du seuil d'audition, en effectuant un audiogramme avant et après l'expérience. Aucune baisse de seuil auditif n'a été observée.

■ **Matériels** L'expérience s'est déroulée dans une cabine insonorisée à double paroi Soluna S1. Les sons ont été diffusés grâce au logiciel Max [IRC93] sur une station NeXT équipée d'une carte ISPW. La conversion numérique vers analogique est gérée par un convertisseur ProPort. Les sons sont alors amplifiés par un amplificateur stéréo Canford et diffusés en écoute dichotique à travers un casque Sennheiser HD 520 II. Les sons ont été diffusés à un niveau de 70 dB. L'interface de réponse ainsi que la collecte des résultats, sont gérées par le logiciel PsiExp [Smi95].

■ **Diplacousie** L'écoute dichotique peut provoquer de la diplacousie binaurale. Dans ce cas, un son de fréquence fixe va évoquer une sensation de hauteur différente entre les oreilles gauche et droite. Afin de mesurer cet effet, les participants ont ajusté la fréquence d'un son pur – diffusé dans l'oreille gauche – afin d'apparier sa hauteur avec celle évoquée par un son pur de fréquence et d'amplitude fixes, diffusé à l'oreille droite. Cette partie a eu lieu avant l'expérience principale, elle a permis en plus aux participants de se familiariser avec le dispositif expérimental.

Pour tenir compte de la diplacousie due à ce protocole expérimental, nous avons effectué pour chaque personne une mesure de diplacousie sur 5 sons purs de fréquences différentes (640, 720, 960, 1010, 1290 Hz). Les résultats sont résumés dans le tableau 5.3. Nous observons un effet faible de la diplacousie, sauf pour la fréquence de 1290 Hz. Cet effet est comparable aux seuils différentiels fréquentiels que nous avons déjà mentionnés, voir §5.4. Par contre, nous utiliserons comme référence les seuils différentiels mentionnés dans la littérature pour notre modèle de contraste que nous développons dans la suite du texte. Nous faisons ce choix car le seuil élevé pour la fréquence de 1290 Hz est difficilement explicable et que cette expérience de contrôle a été effectuée sur un petit nombre de données et de répliques.

5.5.2 Transformation des données

Cette expérience d'appariement nous a permis d'associer à chacune des hauteurs entendues pour chacun des 30 sons de barre, les différentes fréquences et amplitudes ajustées par chaque participant.

5.5. EXPÉRIENCE D'APPARIEMENT

Généralement, les participants ont apparié un maximum de 3 à 4 fréquences pour chaque son, mais pour la plupart des sons, deux fréquences ont été ajustées, voir figure 5.17. Suite aux expériences précédentes, notre hypothèse est de penser que ces fréquences appariées vont être, au niveau du stimulus proximal, une source d'information potentielle utilisée par les auditeurs pour effectuer leur classification.

5.5.2.1 Distribution collective des hauteurs

Notre but est de comparer les différents sons sur la base de ces appariements dans le cadre du modèle de contraste, §3.3.3, p. 81. Pour cela, nous avons cumulé, à chacun des 30 sons, les appariements en fréquence et en amplitude, associés à ces sons sur l'ensemble des participants.

De la même façon que le modèle parcimonieux précédent – basé sur les analyses acoustiques – l'idée, sous-jacente au modèle, est de donner un poids plus important aux fréquences proches qui ont été appariées.

Deux problèmes sont apparus en vue d'utiliser ces appariements pour notre modèle de contraste. Le premier vient de la très grande variabilité intra- et inter-individuelle pour les appariements en amplitude. Il nous a semblé donc peu pertinent d'intégrer ces données dans notre modèle. Même si l'information sur les amplitudes des composantes n'a pu être intégrée dans notre modèle, le réglage en amplitude a permis aux participants d'ajuster les fréquences de manière plus confortable.

Le second problème est associé à la variabilité naturelle interindividuelle entre les fréquences qui semblent ajustées pour une même hauteur. Pour résoudre ce problème, nous avons utilisé une fonction de lissage qui nous permet de transformer les appariements fréquentiels en y intégrant la variabilité interindividuelle. Cette fonction de lissage induit le cumul des fréquences proches, d'un point de vue perceptif, en donnant ainsi une saillance plus importante aux fréquences vraisemblablement appariées à une même hauteur.

Fréq. de référence, (Hz)	640	720	960	1010	1290
Fréq. ajustée moyenne, (Hz)	645	733	967	1026	1373
Ecart type (Hz)	25	88	20	47	271
Diplacousie (%)	0.78	1.80	0.73	1.58	6.43

Tableau 5.3 - Mesure de diplacousie moyenne entre une fréquence de référence (oreille droite) et sa fréquence ajustée (oreille gauche). La déviation est exprimée en % de la fréquence de référence.

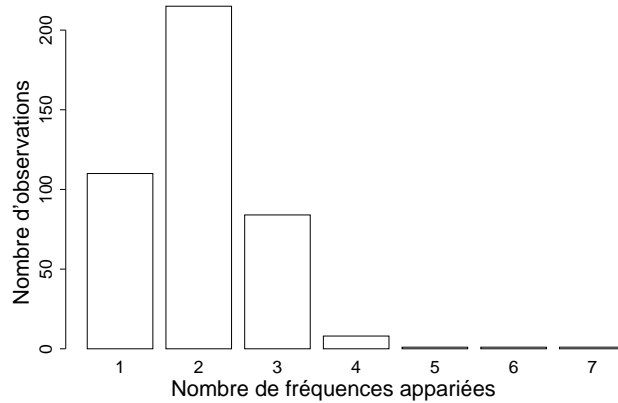


Figure 5.17 - Distribution du nombre d'appariements de fréquences avec les 30 sons cumulés sur l'ensemble des participants.

■ **Lissage** La fonction de lissage (équation 5.1) est une exponentielle symétrique qui est centrée à chaque fois sur la fréquence appariée considérée et dont le maximum de cette fréquence est égal à 1, voir la figure 5.18. La largeur de la fonction (à 0.75 du maximum de l'amplitude) est proportionnelle à la fréquence appariée pour tenir compte de l'augmentation du seuil différentiel avec la fréquence. Le seuil de tolérance est aussi intégré. Cette fonction prend en entrée un vecteur correspondant à la gamme de fréquences qui nous intéresse, à une fréquence d'échantillonnage que nous fixons. Le processus est récursif, la fonction étant appliquée pour chaque fréquence d'appariement d'un son donné. Le spectre d'appariement pour un son est le cumul de l'application de cette fonction à chacune de ces fréquences appariées f_a .

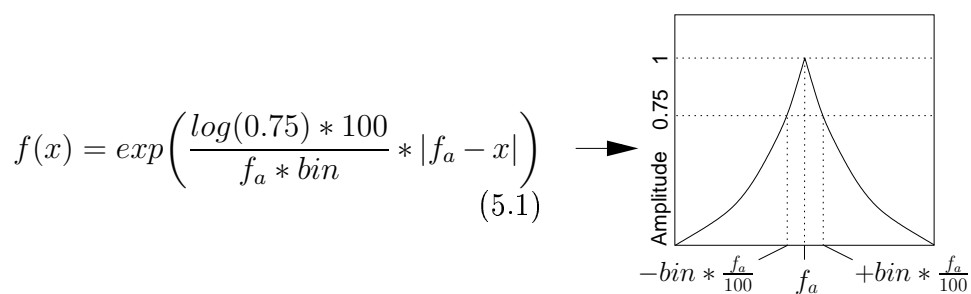


Figure 5.18 - Fonction de lissage (équation 5.1), avec f_a la fréquence considérée à lisser, x le vecteur d'entrée et $2 * bin$ la largeur de la fonction à 0.75 du maximum de l'amplitude. La figure représente la fonction qui prend la valeur maximum à la fréquence appariée considérée f_a et qui a une largeur proportionnelle à celle-ci et à la tolérance, bin , associée.

Cette fonction cumule l'information, pour des fréquences proches (associée à une même hauteur entendue). Ainsi, plus des fréquences proches ont été appariées (de l'ordre du pour-cent), plus la valeur de la fonction (la saillance)

devient importante autour de ces fréquences.

Sur la figure 5.19, sont représentées deux « distributions des hauteurs » que nous avons créées. Nous utilisons cet abus de langage, car nous les considérons comme la représentation de l'ensemble des hauteurs entendues de façon collective par les sujets, même si les données correspondantes sont des fréquences.

La première distribution des hauteurs (1) représente le résultat de la fonction de lissage, pour un ensemble de 5 fréquences. Comme nous l'avons déjà mentionné, la distribution (1) montre qu'aux fréquences d'appariement, la fonction prend ses maxima égaux à 1. c'est-à-dire que pour chacune de ces fréquences la saillance est de 1. La distribution des hauteurs (2) montre le résultat du lissage pour 7 fréquences proches allant de 2145 à 2265 Hz. Nous remarquons dans ce cas que la saillance autour de ces fréquences prend un maximum supérieur à 4.

La figure 5.20 montre, pour le son de la barre 40×20 excitée en position 5, la transformation des données fréquentielles d'appariement (ici dans une gamme de 400 à 5000 Hz) en sa distribution des hauteurs. La première figure représente le spectre moyen du même son calculé par la transformée de Fourier. Nous pouvons remarquer une très bonne correspondance entre le spectre physique et la distribution des hauteurs pour les deux premières fréquences (1139/1135 et 2193/2210 Hz). Une comparaison plus systématique, entre les distributions de hauteurs et les fréquences modales de chaque son montre, une bonne convergence entre les deux.

Le modèle de contraste, appliqué aux analyses acoustiques, nous a permis de montrer que le pattern spectral de chaque son pouvait être un invariant potentiel utilisé pour retrouver les sons provenant d'une même barre. Les distributions des hauteurs mettent en lumière les différentes fréquences – associées aux hauteurs entendues dans chacun des sons – qui peuvent être impliquées dans les jugements de similarité susceptibles d'être mis en jeu dans les expériences de classification. Nous allons par la suite appliquer notre modèle de contraste à ces distributions de hauteurs.

5.5.3 Modèle de contraste

Ainsi, comme nous venons de le voir, nous avons construit les distributions des hauteurs associées à chacun des 30 sons. Cette étape de transformation des données expérimentales nous permet d'obtenir des « spectres d'appariement » qui vont être utilisés dans notre modèle de contraste.

Le modèle de contraste de Tversky permet d'exprimer la similarité entre deux objets sur la base de leurs traits communs et de leurs traits propres, voir §3.3.3, p. 81. Nous voulons calculer la similarité entre chaque paire de

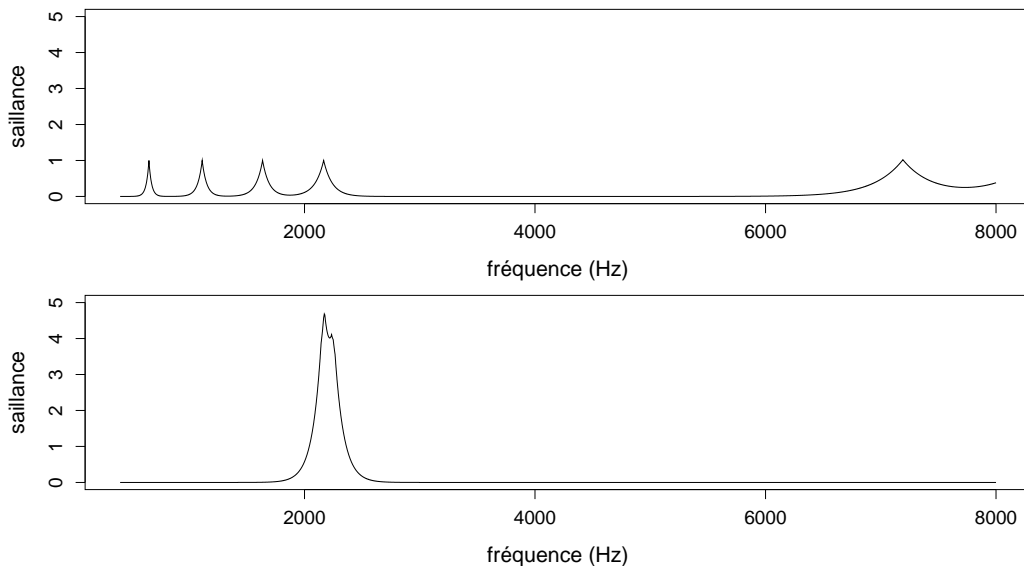


Figure 5.19 - La figure (1) représente la distribution des hauteurs pour 5 fréquences f_a : $\{650, 1112, 1636, 2166, 7192\}$ après application de la fonction de lissage. La figure (2) représente la distribution des hauteurs pour 7 fréquences f_a proches : $\{2144, 2166, 2172, 2175, 2235, 2249, 2265 \text{ Hz}\}$. Dans les deux cas l'intervalle de tolérance prend une valeur de $bin = 1$.

sons sur la base de ces distributions de hauteurs. Ceci nous permet de modéliser un processus possible de réponses de la part des auditeurs, lors des tâches de classification, sur la base d'une représentation auditive des sons de barre. Dans le cadre de ces expériences, nous considérons que les fréquences appariées associées aux différentes hauteurs entendues dans chacun des sons, sont les traits qui rentrent en jeu dans ce modèle. Ainsi, nous posons l'hypothèse que les auditeurs comparent les différents sons, sur la base des hauteurs propres et communes, à travers les différents sons.

5.5.3.1 Présentation du modèle

Afin de comparer deux à deux les différentes distributions de hauteurs, nous avons développé une version intégrative du modèle de contraste. Sur la figure 5.21, deux distributions de hauteurs ont été superposées (barres 80×30.2 et 40×30.2). Nous avons mis en valeur les surfaces des deux spectres qui leur sont communes, ainsi que les surfaces propres à chaque spectre. Le calcul de la similarité entre deux sons est effectué à partir de ces surfaces.

Plutôt que de considérer directement les fréquences appariées comme les traits de notre modèle de contraste, nous avons préféré un modèle intégratif

5.5. EXPÉRIENCE D'APPARIEMENT

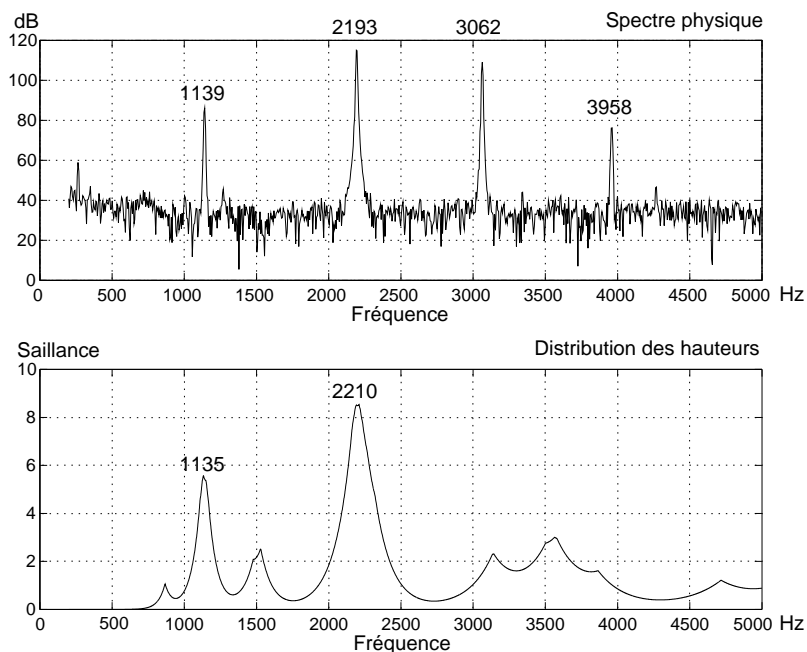


Figure 5.20 - Comparaison du spectre physique de la barre 40×20.5 et de sa distribution des hauteurs associées. Pour le spectre physique l'amplitude est exprimée en dB, le poids des fréquences dans le cas de la distribution des hauteurs est exprimé par la valeur de saillance (eq. 5.1).

qui tient compte de la variabilité interindividuelle. En effet les arbres hiérarchiques des différentes conditions sont construits à partir des matrices de cooccurrences sur l'ensemble des auditeurs.

■ **Description du modèle** Afin d'évaluer les différentes surfaces qui nous intéressent, nous avons procédé à l'évaluation des intégrales des distributions des hauteurs. Comme il n'existe pas de résolution analytique pour calculer directement ces intégrales, nous avons donc procédé à des intégrations numériques. Le principe de l'intégration numérique est de remplacer le calcul de l'intégrale considérée par son approximation, par des surfaces plus faciles à évaluer. Pour le calcul de l'intégrale d'une fonction, la surface est calculée de manière approchée, sur différentes bandes successives, plus ou moins étroites, sur l'ensemble de la fonction, voir figure 5.22.

Il existe des méthodes plus ou moins fines d'approximation des intégrales par intégration numérique, voir [PTVF92]. Nous avons utilisé la méthode la plus simple appelée « méthode des trapèzes » (voir équation 5.2 et sa figure 5.22 associée) par rapport à d'autres méthodes telles que la méthode de Romberg qui tente de minimiser l'erreur. Voir [PTVF92] pour une introduction

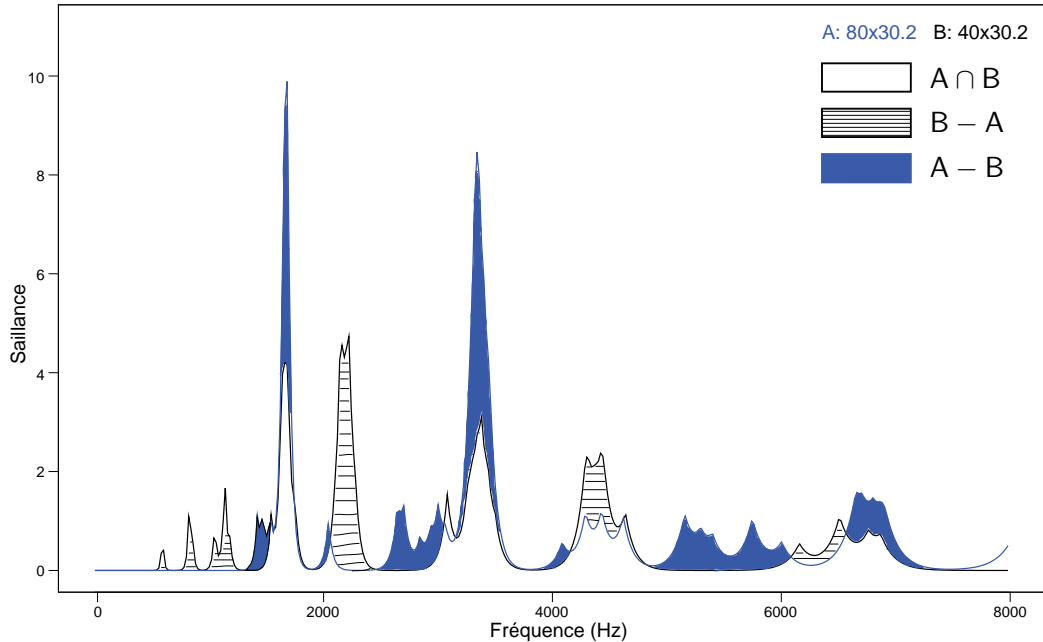


Figure 5.21 - Comparaison de deux distributions de hauteurs (sons de barre 80×30.2 et 40×30.2) sur la base du modèle de contraste de Tversky. Les zones grisées représentent les surfaces propres à chaque distribution (foncée pour le son 80×30.2 , hachurée pour le son 40×30.2). Les surfaces blanches représentent les surfaces communes aux deux spectres. L'axe des abscisses représente la gamme fréquentielle (0-8000 Hz), l'axe des ordonnées représente la saillance des fréquences.

de ces méthodes.

Le choix de cette méthode vient de notre façon d'estimer les surfaces communes et propres aux distributions de hauteurs afin de comparer deux sons, dans le cadre de notre modèle de contraste. La méthode des trapèzes nous permet, sur une bande donnée de la fonction (comme sur la figure 5.22), de pouvoir estimer facilement les surfaces communes ou propres aux deux fonctions. La figure 5.23 montre différents cas de calcul des intégrations, par la méthode des trapèzes : seule une des distributions a une valeur nulle, dans la bande fréquentielle considérée (cas 1 ou cas 2), une des distributions a des valeurs de saillance toujours supérieures à l'autre (cas 3 ou 4), cas (5) où une des distributions a des valeurs de saillance supérieures à l'autre distribution et qui s'inverse dans une même bande de fréquences considérées.

Nous voyons par ces exemples que l'estimation des surfaces propres et communes est facilitée par l'utilisation d'une méthode simple telle que la méthode des trapèzes. L'algorithme propre au calcul de ces différentes intégrations est présenté en annexe §C.5.1.1, p. 246.

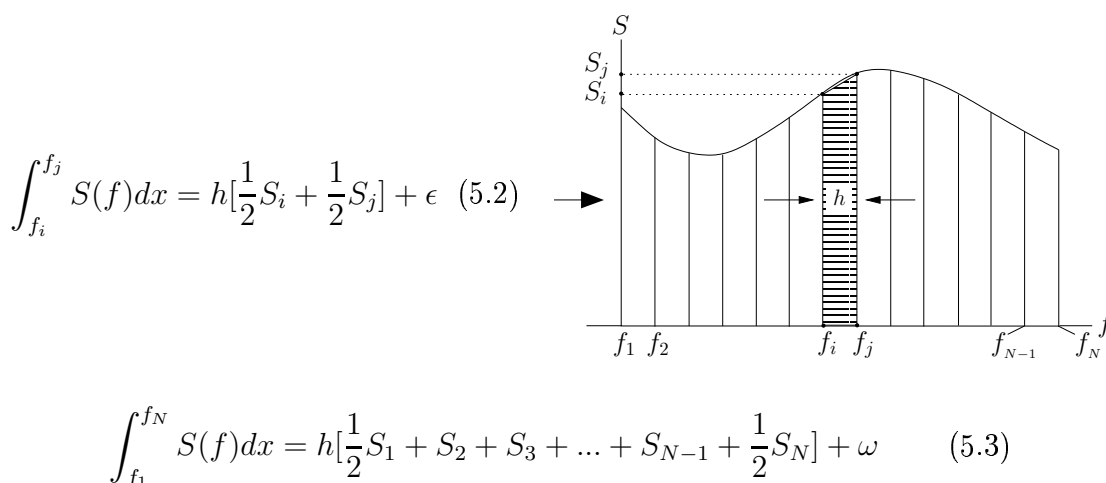


Figure 5.22 - L'équation 5.2 décrit l'approximation d'une intégrale par la méthode des trapèzes. La figure représente une distribution pour laquelle l'intégrale (surface hachurée) est calculée entre les points f_i et f_j . S_i et S_j sont les valeurs de la distribution $S(f)$ aux points respectifs f_i et f_j , h la largeur de bande équivaut à $|f_i - f_j|$, ϵ est l'erreur d'approximation par la méthode des trapèzes. L'équation 5.3 est le calcul de l'intégrale de la distribution $S(f)$ entre les points f_1 et f_N . Le calcul de l'intégrale totale est estimée sur toutes les bandes de largeur constante h de la fonction. ω représente l'erreur sur l'estimation de l'intégrale par intégration numérique.

■ **Evaluation** Dans un premier temps, nous avons évalué la pertinence de nos calculs par l'estimation des intégrales propres et communes entre deux fonctions. Pour cela, nous avons construit deux fonctions « triangle » qui ont la particularité d'avoir, sur la même gamme de variations des zones propres et communes, voir la figure 5.24. La trivialité de ces fonctions a pour but de permettre un calcul analytique, à priori, de ces différentes intégrales ($A \cap B$, $A - B$, $B - A$). Le tableau associé montre les calculs des intégrales théoriques et estimées par nos calculs.

Nous pouvons observer une très bonne approximation de ces différentes intégrales, nous observons en effet une erreur bien inférieure à 1 %. Même si nous utilisons la méthode des trapèzes, en prenant des intervalles de calcul extrêmement fins, nous pouvons ainsi minimiser le taux d'erreur.

5.5.3.2 Mise en œuvre

Le modèle de contraste que nous venons de présenter nous permet de calculer la similarité entre toutes les paires des 30 sons de barre. Nous avons appliqué notre modèle intégratif de contraste sur l'ensemble des distributions de hauteurs dérivées de notre expérience d'appariement. Nous détaillons par la suite l'application des différentes étapes associées à la mise en œuvre de

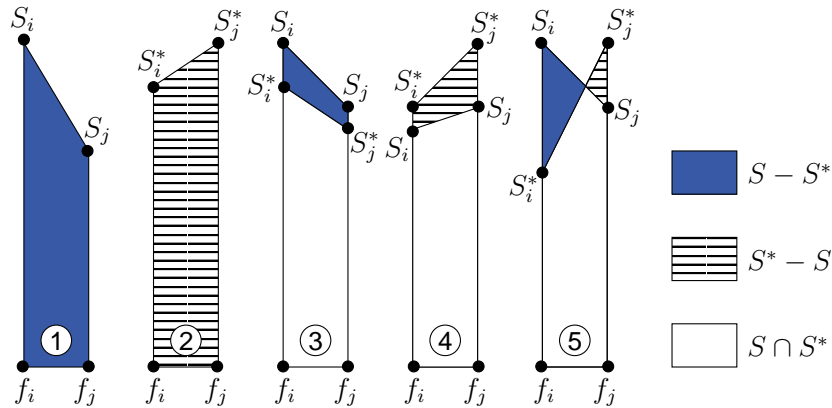


Figure 5.23 - 5 cas de comparaison des surfaces de deux distributions S et S^* dans une bande $|f_i - f_j|$. Le spectre S prend les valeurs S_i et S_j en f_i et f_j , de la même façon pour la fonction S^* , f_i^* et f_j^* . Les surfaces blanches représentent les surfaces communes aux deux distributions S et S^* . Les surfaces hachurées sont les surfaces propres à la distribution S et les surfaces sombres, celles propres à la distribution S^* .

ce modèle.

■ **Distributions de hauteurs** Dans un premier temps, nous avons créé, pour chacun des 30 sons, une distribution de hauteurs, à partir des données expérimentales. Nous avons restreint les distributions à une gamme de fréquences allant de 400 à 8000 Hz, au vu de la gamme de variation des appariements fréquentiels effectués par les participants.

Afin d'intégrer la variabilité interindividuelle entre les appariements des participants, nous avons pris un intervalle de 1% (c'est-à-dire pour une valeur de bin de 0.5%, voir équation 5.1). L'intervalle de tolérance que nous prenons est dans l'ordre de grandeur des seuils différentiels en fréquence que nous avons déjà mentionné. Contrairement à ces expériences, les sons étant inharmoniques, nous avons préféré prendre la borne supérieure de 1 %. Les spectres sont échantillonnés à 1 Hz. Ce choix est en accord avec notre intervalle de tolérance.

■ **Intégration et normalisation** A partir de l'ensemble des 30 spectres d'appariement, nous avons calculé la similarité entre chaque paire de sons, sur la base du modèle de contraste.

Nous n'avons pas utilisé directement le modèle de contraste sous sa forme standard, telle que nous l'avons déjà présentée. La difficulté avec le modèle de contraste, dans notre cas, est le fait de comparer des sons dont le nombre de fréquences appariées varie suivant les sons considérés. Le modèle de contraste peut ainsi produire une similarité plus importante, dans le cas où deux sons

5.5. EXPÉRIENCE D'APPARIEMENT

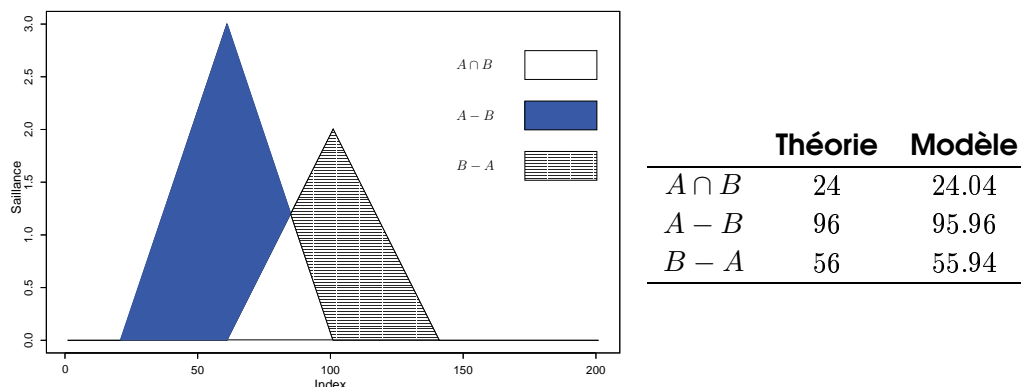


Figure 5.24 - Evaluation des calculs des intégrales communes aux deux fonctions triangles ($A \cap B$) et des intégrales propres à chaque fonction ($A - B$ ou $B - A$). Le tableau associé à cette figure présente le calcul théorique ainsi que l'estimation par notre méthode de ces différentes intégrales.

ont de nombreuses fréquences appariées. Ainsi l'ensemble des traits comparés est large, par rapport à deux sons qui n'ont qu'un petit nombre de fréquences communes mais dont l'ensemble des traits comparés est restreint. Par contre, d'un point de vue perceptif, ces deux paires de sons peuvent avoir une similarité équivalente, puisque partageant dans les deux cas un nombre de traits communs plus important que les traits distinctifs.

C'est pourquoi, nous avons opté pour le modèle de rapport (équation 5.5.3.2, [Tve77]) qui permet de normaliser la similarité entre 0 et 1. Cette normalisation est nécessaire lorsque nous voulons analyser nos données par groupements hiérarchiques comme nous l'avons fait pour les expériences de classification, voir §5.3.3, p. 143.

La similarité $S(a, b)$ entre deux objets a et b s'exprime sous la forme :

$$S(a, b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)} \quad (5.3)$$

avec $A - B$ les traits propres à \mathbf{a} (respectivement $B - A$ les traits propres à \mathbf{b}) et $A \cap B$ les traits communs. f est une fonction mesurant la saillance des traits communs ou distinctifs. S et f sont des échelles d'intervalle. Les paramètres α et $\beta \geq 0$ dépendent de la tâche ou du contexte.

Dans le cadre de nos expériences de classification, la tâche est non directionnelle (nous n'orientons pas les participants à comparer le son a au son b plutôt que de comparer le son b au son a). Nous pouvons donc supposer que $S(a, b) = S(b, a)$. Cette symétrie de la similarité implique alors que $\alpha = \beta$ [TG78].

5.5.3.3 Résultats

A partir du modèle de rapport (équation 5.5.3.2), nous avons calculé la similarité entre chaque paire des 30 sons. Dans notre cas, la fonction f qui permet de mesurer la saillance des traits fréquentiels est l'intégration des surfaces propres et communes des distributions de hauteurs. Ainsi, nous avons calculé la similarité entre chaque paire de sons grâce à l'expression suivante du modèle de rapport :

$$S(a, b) = \frac{\int_{A \cap B}}{\int_{A \cap B} + \alpha(\int_{A-B} + \int_{B-A})} \quad (5.3)$$

avec \int l'intégrale des spectres propres à a et/ou b .

Suite aux calculs de similarité entre les sons à partir des distributions de hauteurs, nous avons effectué une analyse par groupements hiérarchiques (groupement moyen). Nous voulons comparer l'estimation des similarités par notre modèle aux similarités résultant des trois expériences de classification. Nous présentons sur la figure 5.25, le résultat de notre modèle d'appariement, avec une valeur de $\alpha = 1$, conjointement aux résultats de la condition « orientée ». Nous avons représenté uniquement les quatre catégories qui correspondent à la partition optimale de l'arbre hiérarchique (voir la figure C.10 en annexe p. 248 pour l'arbre original) et qui nous permettent de les confronter aux catégories qui se dégagent des trois expériences de classification.

5.5.3.4 Analyses

L'estimation des similarités à partir de notre modèle montre des résultats pour les catégories principales très proches des résultats des trois classifications. Ainsi, nous retrouvons les quatre catégories correspondant à des dimensions privilégiées (épaisseur \blacktriangleright catégories I, IV et III, largeur \blacktriangleright catégorie II). Les différences, dans les niveaux inférieurs de la hiérarchie des arbres, semblent difficilement interprétables. Par contre, ce modèle ne permet pas d'expliquer la catégorie propre à la barre 40×10 mm ni le regroupement des sons de la barre 40×30, pour la condition « contrainte ».

Nous devons remarquer que notre modèle de rapport est basé essentiellement sur les traits communs. Le modèle étant intégratif, une valeur de $\alpha = 1$ fait que l'expression $\int_{A \cap B} + (\int_{A-B} + \int_{B-A})$ reste constante. Dans ce modèle de contraste, les traits communs sont associés aux hauteurs communes entre les sons de barre qui permettent d'expliquer les 4 catégories pour les conditions « non-orientée » et « orientée ». Ces hauteurs proéminentes sont liées aux fréquences modales de flexion, suivant l'épaisseur ou la largeur. L'introduction de traits distinctifs avait pour but d'expliquer les différences de

5.5. EXPÉRIENCE D'APPARIEMENT

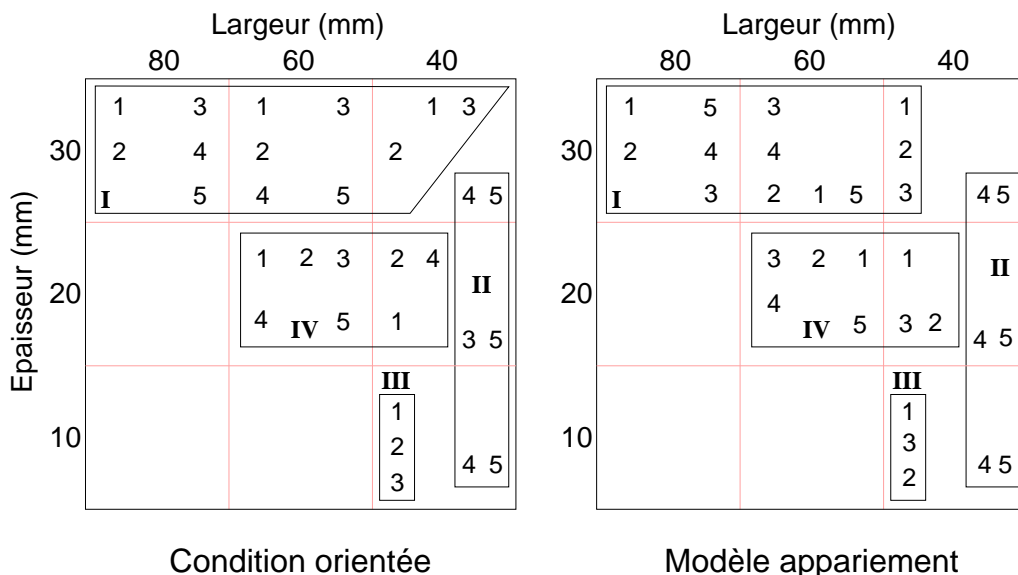


Figure 5.25 - Comparaison des classes obtenues dans la condition « orientée » et issues du modèle de contraste dérivée des distributions de hauteurs.

groupement dans la condition « contrainte ». Dans ce cas, nous pensons que pour certains sons (barres 40×10 et 40×30), les hauteurs proéminentes étaient mises en comparaison avec des hauteurs moins saillantes (les traits distinctifs) liées aux fréquences modales de flexion, suivant l'autre dimension. Par exemple, pour les sons de la barre 40×10, les fréquences associées au mode de flexion, suivant la largeur, pourraient jouer ce rôle discriminant par rapport aux autres sons. Dans le cas de la barre 40×30, les traits communs liés aux fréquences de flexion, suivant l'épaisseur, pour les conditions « non-orientée » et « orientée », deviendraient des traits distinctifs, dans la condition « contrainte ». En effet, dans cette condition, les sons liés à l'excitation, suivant l'épaisseur, sont groupés dans la catégorie liée à la largeur, pour la barre 40×30.

Le rôle des traits distinctifs nous a semblé important car le modèle de contraste, appliqué aux fréquences estimées par des analyses acoustiques, a montré que les fréquences communes, pouvaient à elles seules, expliquer un regroupement des sons par barre, dans le cas où les hauteurs associées à ces fréquences seraient toutes perçues.

Nous avons fait varier la valeur de α pour essayer d'interpréter les résultats de chacune des expériences de classification. Le but est d'observer l'influence des traits distinctifs dans les processus de comparaison, propres à chaque tâche de classification. Pour cela, nous avons effectué des régressions multiples

entre les composantes du modèle (traits communs et traits distinctifs) et chacune des matrices de proximité. Nous n'avons pas réussi à trouver des valeurs de α expliquant les variations de la structure fine des catégories et à expliquer les variations propres à la condition « contrainte ». Notre modèle semble donc expliquer le rôle des hauteurs communes entre les sons mais ne permet pas de prédire le rôle des hauteurs distinctives. Ceci peut s'expliquer par le nombre faible de fréquences appariées (en moyenne 2), pour chacun des sons, par l'ensemble des participants. Dans cette expérience, les sujets semblent s'être focalisés sur les hauteurs liées aux fréquences modales de flexion des dimensions, privilégiées par la position de frappe sur la barre.

Il apparaît donc que ce modèle, basé uniquement sur les fréquences associées aux hauteurs perçues, n'explique pas tous les processus perceptifs impliqués dans les tâches de classification. Pour la condition « contrainte », il existe une forte variabilité entre les sujets, du fait du nombre de classes et de sons imposés. Les participants ont pu utiliser différents critères perceptifs qui ne rentrent pas dans le cadre de ce modèle mais qui, du fait de cette variabilité, ne ressortent pas clairement.

Malgré tout, notre modèle de rapport nous a permis de mettre en évidence le fait que les participants comparent les différents sons sur la base des hauteurs spectrales entendues dans chacun des sons. Ces différentes hauteurs spectrales sont à relier à leurs fréquences concomitantes, associées à des modes de vibration propres. Ces modes de vibration de flexion (suivant l'épaisseur ou la largeur) induisent des fréquences qui sont directement liées aux dimensions de la barre.

Par contre, le modèle montre que les auditeurs se focalisent sur la hauteur la plus saillante qui est liée à la fréquence modale de flexion, suivant l'épaisseur ou suivant la largeur. Nous voyons donc qu'à partir de sources similaires aux expériences de Lakatos et coll. [LMC97], les auditeurs se focalisent sur des propriétés différentes pour effectuer cette tâche.

5.6 Discussion

Cette série d'expériences nous montre que les auditeurs n'extraient pas, du stimulus proximal, des invariants propres à chaque barre, sur la base des fréquences modales. Ils se focalisent avant tout sur des dimensions géométriques (épaisseur ou largeur) privilégiées. Ces jugements sont effectués sur la base des hauteurs spectrales, liées aux fréquences de vibration, suivant ces dimensions.

Nous avons voulu mettre en lumière la signature spectrale de chaque barre en l'excitant de différentes manières. Ceci n'a pas permis de spécifier

la signature sonore de l'objet qui aurait pu être une propriété invariante du stimulus proximal.

Contrairement aux résultats d'autres auteurs [LMC97] [KPT00], il semble que les auditeurs n'aient pas réussi à encoder les propriétés spectrales et temporelles des sons, dans des représentations mentales qui reflètent les propriétés physiques et géométriques des sources (telles que la section géométrique d'une barre [LMC97] ou les dimensions d'une plaque [KPT00]). Même si les auditeurs se sont focalisés sur ces dimensions géométriques privilégiées, ce processus semble être la conséquence de processus de comparaison des sons, sur la base d'une similarité de surface (la hauteur), en faisant abstraction de la source.

D'un point de vue théorique, nous voulions comparer les théories de la perception directe et celles du traitement de l'information. En ne donnant aucune information sur la source (condition « non-orientée »), nous avons cherché à voir si les auditeurs extraient naturellement, au niveau du stimulus proximal l'information associée à chaque barre (perception directe). Dans les conditions « orientée » et « contrainte », nous avons fourni une information de plus en plus complète sur les sources, pour inférer des traitements et des représentations propres à chaque source. Il apparaît que les auditeurs n'extraient pas des invariants structuraux propres à chaque barre dans la condition « non-orientée ». Dans les conditions « orientée » et « contrainte », les auditeurs ne construisent pas une représentation perceptive de la géométrie de la barre qui permettrait le groupement des sons par barre. Dans les trois conditions, les auditeurs ont effectué leur jugement, sur la base d'une similarité directement accessible à partir des attributs auditifs élémentaires. Ce processus est effectué par comparaison des hauteurs qui reflètent les fréquences modales privilégiées par l'excitation de la barre suivant la face ou le côté.

J.J. Gibson fait une remarque qui nous semble importante à ce sujet : « les vibrations (sonores) portent une signification externe et objective dans la mesure où elles spécifient la source » [Gib66], p. 8. Dans le cadre de nos expériences, il semble que les auditeurs soient restés à une écoute musicale des sons, sans que ces sons véhiculent une information sur les propriétés de la source. Dans une démarche d'acoustique écologique, il semble difficile de conclure, pour cette étude, sur une écoute des attributs audibles de la source.

Suite à ces résultats, il nous a semblé important de reprendre l'étude de Lakatos et coll. [LMC97] qui a inspiré notre problématique. Les résultats de cette étude montrent que les auditeurs discriminent auditivement des sons de barres, variant uniquement par leurs sections géométriques. L'information utilisée est liée au rapport épaisseur/largeur, correspondant à la section géométrique de la barre. Leurs résultats n'ont pas permis de déterminer le

type d'information utilisée qui peut provenir aussi bien des fréquences modales liées aux modes de flexion (suivant l'épaisseur et la largeur) ou aux fréquences modales de torsion directement proportionnelles à ce rapport. Pour nos expériences, nous avons essayé de privilégier différents modes de vibration, par rapport à d'autres suivant l'endroit d'excitation. Dans notre cas, le découplage des modes n'a peut-être pas permis aux auditeurs d'utiliser une information acoustique leur permettant de se représenter la section géométrique de la barre. De plus nos résultats n'ont pas permis de mesurer l'importance des modes de torsion, du fait peut-être du mode de fixation utilisé. Nous allons tenter dans une nouvelle série d'expériences d'étudier l'importance de ces contributions modales, dans la représentation auditive de la géométrie d'une barre de métal.

Chapitre 6

Représentation des sources sonores



6.1 Introduction

Une des expériences qui a guidé nos travaux est celle de Lakatos et coll. [LMC97] sur la représentation (mentale) des caractéristiques audibles de la source. Leurs travaux ont tenté de spécifier les invariants liés aux propriétés d'un corps résonnant (des barres de métal et de bois) qui permettent d'associer une collection de sons, variant suivant leur timbre et leur hauteur, à une source sonore unique.

L'étude de Lakatos et coll. n'a pas permis de conclure précisément sur la contribution spécifique des différents modes de vibration, (flexion, torsion ...) dans la représentation auditive des dimensions géométriques de la source sonore, sur la base de propriétés acoustiques. A partir d'un sous-ensemble du même corpus de barres que pour leurs expériences, il semble que, dans notre cas, les auditeurs n'associent pas les sons à des catégories propres à chaque barre mais à des dimensions géométriques particulières. Nos résultats sont donc contraires à l'expérience de Lakatos et coll. pour laquelle les

participants discriminent les barres sur la base d'une représentation auditive des différences de section entre ces barres. Leur étude repose sur une tâche d'appariement des informations auditives à une représentation visuelles des objets, contrairement à notre première série d'expériences impliquant des tâches de classification.

Suite à ces questions restées en suspens, il nous a semblé important de revenir sur les expériences qui ont impulsé nos travaux, c'est-à-dire sur la contribution effective des différents modes de vibration (torsion, flexion) sur la représentation de la géométrie d'un objet résonnant simple, tel qu'une barre de métal. De la même façon que pour nos expériences de classification, nous nous focaliserons sur l'étude des barres de métal.

6.2 Représentation d'une source sonore

6.2.1 Rappel des travaux antérieurs

Lakatos et coll. [LMC97] ont cherché à évaluer la capacité des auditeurs à discriminer la forme géométrique d'objets résonnants, sur la base de leurs attributs auditifs.

Les objets étudiés sont des barres rectangulaires de métal et de bois frappées dont la section seule varie. Les expériences impliquent une tâche d'appariement intermodal entre les sons des barres (une expérience avec les barres de métal et une autre avec les barres de bois) enregistrés et la représentation visuelle de leurs sections. Une présentation détaillée du protocole expérimental est donnée ultérieurement.

Les analyses multidimensionnelles, représentant les résultats pour les barres de métal, révèlent un espace à deux dimensions et pour les barres de bois, un espace à une dimension. Pour les barres de métal, la première dimension est corrélée avec le rapport épaisseur/largeur (h/b) et la deuxième dimension avec une mesure du centre de gravité spectral. Pour les barres en bois, la seule dimension est également corrélée avec le rapport épaisseur/largeur.

L'espace multidimensionnel représentant les données de l'expérience des barres de métal, montre une séparation perceptive nette entre les blocs et les plaques basée sur les deux dimensions. Les résultats sur les barres de bois montrent une corrélation avec le rapport h/b nettement inférieure à l'expérience sur les barres de métal. Les résultats de ces expériences ont été associés à des analyses acoustiques ainsi qu'à des mesures d'accélérométrie afin de caractériser les modes de vibration des barres. Ces analyses montrent la présence des modes de flexion suivant, la largeur et l'épaisseur, ainsi que des modes de torsion, en bon accord avec les valeurs théoriques pouvant servir

de corrélats fréquentiels aux dimensions perceptives.

6.2.2 Hypothèses

Lorsqu'une barre est mise en vibration, les deux types de modes de vibration (flexion et torsion) produisent des fréquences modales qui sont directement liées :

- z à la largeur de la barre, pour une vibration de flexion, suivant cette dimension ;
- z à l'épaisseur de la barre, dans le cas d'une vibration de flexion, suivant cette dimension ;
- z au rapport épaisseur/largeur de la barre, pour les fréquences modales de torsion.

si nous faisons abstraction de la contribution des modes de vibration longitudinaux.

Nous voyons donc ici que pour les modes de flexion (épaisseur et largeur), la comparaison des fréquences modales donne indirectement une information sur la section de la barre, contrairement aux modes de torsion qui donnent directement cette information.

Nous remarquons aussi que, dans le cas de la flexion, les fréquences modales sont inharmoniques, contrairement aux fréquences modales de torsion, voir équations 5.2.1 et 5.2.1 p. 135. Deux processus perceptifs peuvent rentrer alors en jeu, comme bases de représentation auditive des dimensions géométriques : la comparaison perceptive des hauteurs spectrales, liée aux fréquences modales de flexion, suivant l'épaisseur et la largeur et une focalisation sur la série harmonique associée aux hauteurs liées aux fréquences modales de torsion.

Afin de tester les contributions propres à ces différents modes de vibration, nous avons réutilisé le paradigme expérimental de Lakatos et coll. [LMC97] en sélectionnant, dans les sons originaux, les fréquences propres à chaque mode (flexion ou torsion) et en resynthétisant des sons dont l'information fréquentielle a été réduite, pour les barres de métal.

6.3 Construction des stimuli

Nous allons tout d'abord présenter les sons originaux qui ont été utilisés dans l'étude de Lakatos et coll. Ces sons nous serviront par une double analyse (traitement du signal et mécanique) à sélectionner les informations propres aux différents modes de vibration, en vue de la synthèse de nouveaux sons.

6.3.1 Sons originaux

Les 12 barres qui ont été utilisées pour cette étude ont été enregistrées par les auteurs ([LMC97]) en chambre anéchoïque, grâce à deux microphones Schoeps sur un DAT Sony 670, à un taux d'échantillonnage 48 kHz en 16 bits. Les barres sont suspendues à un support, par le biais d'un crochet soudé sur celles-ci. Pour les expériences précédentes le crochet a été enlevé, du fait de la présence de la résonance du crochet dans les signaux. Les deux microphones ont été positionnés à un mètre du centre de chaque barre et séparés de 30 cm. Les barres ont été frappées en leur milieu, suivant l'épaisseur. Cette position d'excitation correspond à la position (1), dans le cadre des expériences précédentes, §5.2, p. 134. La figure 6.1 montre la distribution des dimensions des barres ainsi que la position d'excitation. Afin d'obtenir une excitation homogène entre les barres, plusieurs enregistrements ont été effectués pour chaque barre. Le choix s'est porté sur les sons dont l'excitation est uniforme sur l'ensemble des 12 sons de barre.

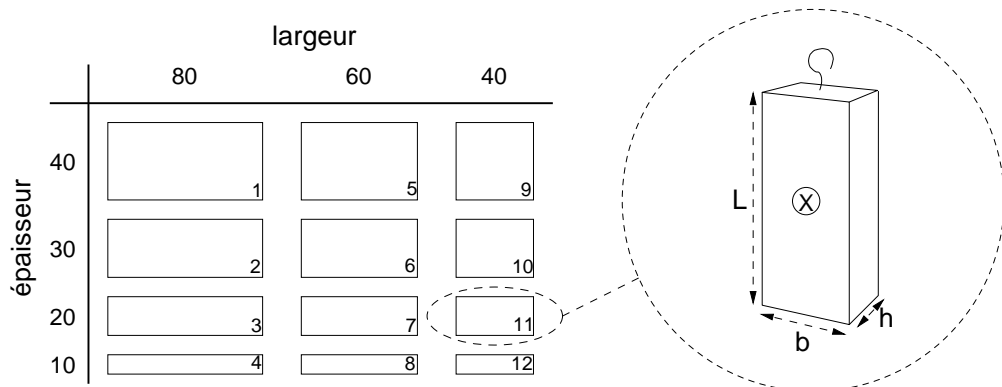


Figure 6.1 - Epaisseur (h) et largeur (b) des 12 barres, ces dimensions sont exprimées en millimètre. La barre (h : 20 mm et b : 40 mm) est représentée pour l'exemple avec son point d'excitation, notée (x), commun entre toutes les barres. Ce point est situé au milieu de la barre et l'excitation a lieu suivant l'épaisseur. L représente la longueur des barres (30 cm) qui est commune à toutes les barres. Nous prendrons comme convention de noter une barre de h : 20 mm et b : 40 mm sous la forme : 40 x 20.

6.3.2 Analyse et synthèse

Nous avons voulu séparer pour chacun des 12 sons de barre, les différentes contributions fréquentielles propres aux modes de flexion et de torsion. Ceci

nous permet de créer des sons dont l'information fréquentielle est sélectionnée suivant ces modes de vibration : modes de flexion seuls, modes de torsion seuls et des sons comprenant ces deux types d'information modale. Nous voulons ainsi créer des sons schématiques (parfois appelé « sound cartoon », [Gav88]) de barres de métal par synthèse. Le terme schématique est utilisé dans le sens où l'information initiale a été réduite mais nous voulons que cette réduction informationnelle reste pertinente, d'un point de vue perceptif. Nous contrôlons la composition fréquentielle, afin de tester nos hypothèses sur la contribution des différents types de modes.

Contrairement à d'autres études utilisant uniquement des sons de synthèse [LO97] [KPK00] [Rou99], il nous a paru préférable de partir de sons naturels, résultant d'un processus mécanique, pour rester au plus près des conditions de l'expérience originale.

Comme nous l'avons déjà mentionné, §2.5.4.5, p. 30, il existe deux approches, dans l'étude des propriétés des sources basées sur les sons de synthèse. La première approche est liée à des modèles physiques complexes qui fournissent un contrôle précis des paramètres physiques. Il a été montré que certains de ces paramètres sont liés à la perception des matériaux, voir [Rou99] ou [KPK00]. Généralement ces études se sont focalisées sur la perception du matériau. Même si ces modèles proposent un paramétrage précis permettant de changer la nature du matériau, il semble difficile, dans notre cas, de pouvoir sélectionner l'information modale nous intéressant.

D'autres méthodes sont basées essentiellement sur un modèle de signal représenté sous la forme de sinusoides exponentiellement amorties. Ces études, comme précédemment, se sont avant tout focalisées sur la perception des matériaux. Elles montrent une certaine divergence quant au rôle et au modèle du type d'amortissement, voir [LO97] et [KPK00]. Du fait du choix parfois arbitraire des fréquences, il est difficile pour nous d'utiliser directement un tel modèle.

Dans notre cas nous adopterons une démarche hybride en nous basant sur l'analyse des signaux et la modélisation par éléments finis des barres. La synthèse, par contre, utilisera les formes d'ondes formantiques qui sont un modèle particulièrement pertinent pour des sons percussifs. Ces différentes techniques ont été présentées dans la partie §4, p. 116.

La figure 6.2 schématise la démarche de construction des sons que nous avons adoptée : analyse des sons par modèles de résonance (étape 1), sélection des fréquences modales par comparaison aux modes propres calculés par éléments finis (étape 2) et enfin resynthèse des sons par formes d'ondes formantiques (étape 3).

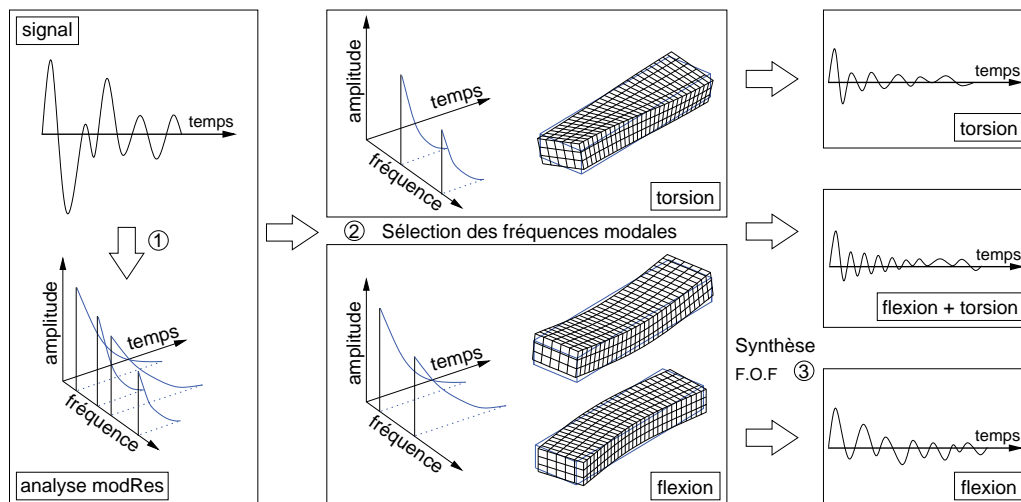


Figure 6.2 - Synopsis de l'analyse et de la resynthèse des sons. La première étape est l'analyse du signal par modèles de résonance. La deuxième étape correspond à la sélection des fréquences modales à l'aide des modèles de vibration des barres. La troisième étape correspond à la synthèse des nouveaux sons.

6.3.2.1 Analyse par modèles de résonance

Dans un premier temps (étape 1), nous avons effectué des analyses en modèles de résonance « modRes » pour chacun des 12 sons. Comme nous l'avons déjà présenté, §4.2.2.1, p. 120 pour une présentation détaillée, cette analyse permet d'estimer chaque mode de résonance du corps ayant produit le son analysé. Un mode de résonance est caractérisé par sa fréquence, son amplitude, son temps d'attaque, sa largeur de bande correspondant à la décroissance.

Le programme d'analyse, en modèles de résonance, utilise le format SDIF [WFR⁺98] de description du son qui permet d'enregistrer différentes représentations, quelles soient dans le domaine spectral ou temporel. Il permet de décrire différents types de signaux : transformée de Fourier à court terme sur des fenêtres, pistes de suivi de sinusoides, suivis de pics spectraux dans le temps ... Nous présentons en annexe, §D.1, p. 250, un exemple de fichier en sortie d'analyse modRes qui nous sert, par la suite, pour la synthèse avec le programme CHANT.

Le format SDIF permet d'avoir un format standard qui le rend compatible avec de nombreux programmes et notamment le programme CHANT de synthèse par F.O.F. Grâce à ce format, chaque son est décrit par ses différentes résonances dont les valeurs décrites précédemment peut être faci-

lement reparamétrées manuellement. La facilité d'utilisation de ce format de description nous a permis de manipuler les différents paramètres, en fonction du choix des fréquences modales, effectué grâce à l'analyse mécanique du comportement de chaque barre.

6.3.2.2 Modélisation du comportement des barres

Afin de modéliser le comportement des barres, nous avons estimé les modes propres des 12 barres par une modélisation par éléments finis. Ceci nous a permis d'identifier les différents modes de vibration des barres grâce à la visualisation des déformées modales.

Le problème physique a été simplifié puisque la modélisation du comportement de la barre ne servait pas directement à la synthèse du son. Nous avons étudié le comportement de chaque barre en ignorant son rayonnement dans un milieu aérien ainsi que l'impact de l'excitation et la déformation locale concomitante de la structure. De plus, nous avons posé l'hypothèse raisonnable que la barre est un matériau isotrope, c'est-à-dire un matériau dans lequel les propriétés physiques sont identiques dans toutes les directions, matériau macroscopiquement homogène, même si c'est de l'acier laminé. L'acier laminé est un matériau légèrement orthotrope plus dense dans le sens de l'épaisseur que dans le sens de largeur et de la longueur. Nous nous sommes mis dans le cas d'une barre en vibration « libre-libre » qui n'implique aucune contrainte sur les frontières de la barre, ce qui revient à négliger les supports de fixation de la barre.

Pour nos travaux, nous avons utilisé le programme d'éléments finis CASTEM2001 [VCA88] qui gère aussi bien le maillage (par script) que les calculs numériques.

■ **Déroulement** Le programme crée en premier lieu un maillage défini par un script qui décrit la structure, le matériau, les conditions aux limites. Ensuite le programme calcule les modes propres de vibration et extrait les fréquences propres. Le programme permet par la suite de visualiser et d'animer les déformées modales. Un script utilisé dans CASTEM est présenté en annexe, voir §D.2, p. 251.

■ **Maillage** Nous avons maillé la structure de chaque barre de façon systématique en utilisant comme maille volumique de référence un cube à 20 noeuds (sommets et milieux des arêtes, CU20 en code CASTEM2001). Le maillage est fait indirectement par « déplacement » [Gua02]. Le programme commence par créer une ligne maillée qui se déplace par translation pour former un maillage surfacique de quadrangles. A partir de la surface, le volume

est créé de la même façon par extrusion. Chaque élément défini est associé à ses nœuds. Les fonctions d'interpolation sont de type quadratique du fait de l'élément fini choisi. Nous avons choisi d'utiliser 6 éléments finis dans le sens de la largeur et de l'épaisseur et 30 éléments dans le sens de la longueur.

La finesse du maillage est retenue en observant la variation des fréquences modales, lorsque nous augmentons le nombre d'éléments finis. Cette augmentation du nombre d'éléments finis est arrêtée lorsque les fréquences se stabilisent. La taille du maillage doit être cohérente avec le domaine fréquentiel et sa résolution associée afin qu'ils soient compatibles avec le domaine des fréquences audibles.

La figure 6.3 présente des exemples de maillages ainsi que les déformées modales des premiers modes de flexion suivant l'épaisseur et la largeur ainsi que le mode de torsion.

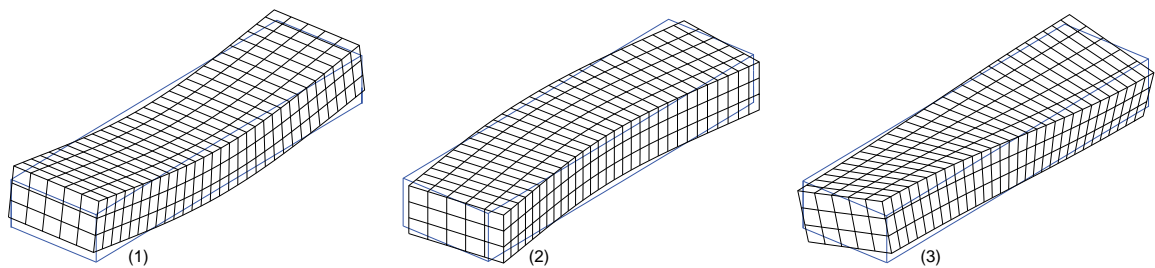


Figure 6.3 - Modélisation par éléments finis des modes propres d'une barre de métal 80x40 grâce au programme CASTEM2001 : (1) les déformées modales du premier mode de flexion suivant l'épaisseur et (2) suivant la largeur ; (3) les déformées modales du premier mode de torsion. Pour cette illustration graphique, le maillage de la barre est constitué de 5 éléments cubiques suivant la largeur, 3 suivant l'épaisseur et de 30 suivant la longueur. Pour les analyses, nous avons maillé avec 6 éléments pour la largeur et pour l'épaisseur.

■ **Choix des paramètres** Nous avons choisi les paramètres indiqués dans le tableau 6.1, correspondant à l'acier, à partir de valeurs mentionnées dans la littérature [FR91] p. 234. Nous avons adapté de manière heuristique ces valeurs afin de s'approcher au mieux des résultats de l'analyse des signaux précédents, notamment le module d'Young qui semblait trop élevé pour sa valeur standard (21000). De plus les données physiques de ce matériau ne sont pas disponibles auprès du distributeur.

6.3.2.3 Sélection des fréquences modales

Le but de ces analyses et de ces modélisations mécaniques est de pouvoir séparer les contributions fréquentielles des modes de flexion et des modes de

6.3. CONSTRUCTION DES STIMULI

Paramètres physiques	Valeurs
E (code CASTEM \triangleright YOP)	20300
ρ (code CASTEM \triangleright RHOP)	7.8 (kg/dm^3)
ν (code CASTEM \triangleright NUP)	0.29

Tableau 6.1 - E le module d'Young relié à l'élasticité du matériau, ρ la densité du matériau et ν le coefficient de Poisson correspondant au rapport entre les deux déformations linéaires relatives respectivement perpendiculaire et parallèle à la direction d'une sollicitation uniaxiale.

torsion afin de créer de nouveaux sons. Les fréquences ont été sélectionnées en comparant l'analyse du signal en modèles de résonance avec les fréquences modales estimées par CASTEM. La visualisation des déformées modales, associées à chaque fréquence nous a permis d'identifier la famille de mode (flexion ou torsion). Le tableau 6.2 montre la comparaison des fréquences. Nous voyons une très bonne correspondance entre les estimations et les valeurs réelles des fréquences. Il faut remarquer que cette correspondance diverge parfois de l'ordre du pourcentage, suivant les barres. Ces divergences peuvent être dues aux variations intra-barres du matériau. De plus, l'analyse en modRes nous a donné parfois beaucoup plus de modèles de résonance que le nombre réel de résonances. Généralement ces résonances supplémentaires sont très proches d'une résonance réelle (variation de l'ordre du pour-cent), mais d'amplitude nettement inférieure. Dans ce cas nous avons procédé à une simplification des données en gardant uniquement la résonance de plus grande amplitude.

Comme la longueur est commune à toutes les barres (30 cm), nous avons enlevé les fréquences modales spécifiques aux modes de vibration longitudinaux.

6.3.2.4 Synthèse

Ainsi, pour chaque barre, nous avons séparé les fréquences modales de flexion (épaisseur + largeur) des fréquences modales de torsion. A partir des fichiers de données, nous avons synthétisé des sons contenant l'information fréquentielle propre aux modes de flexion, d'autres contenant l'information fréquentielle propre aux modes de torsion et enfin des sons contenant les deux.

Nous avons utilisé pour cela le programme CHANT [RPB84] qui permet de synthétiser un son à partir des fréquences de résonance sélectionnées. Nous avons contrôlé à chaque fois la qualité perceptive des sons avec l'information des deux modes de vibration pour s'assurer de la correspondance avec les sons naturels d'origine. Les sons sont échantillonnés à 44.1 kHz en 16 bit.

Mode	F modRes (Hz)	F estimée (Hz)	DéviatiOn	Amp	Amor (s^{-1})
e	2184.94	2200.6	0.71 %	0.01409	5.67
w	3888.44	3847.3	1.07 %	0.00026	4.67
t	3992.53	3926.2	1.66 %	0.00384	1.02
e	5556.20	5551.0	0.09 %	0.00614	0.91
t	8006.55	7876.5	1.62 %	0.04868	2.68
w	8489.03	8475.1	0.16 %	0.00097	2.16
e	9861.51	9816.2	0.46 %	0.11931	14.92
t	12077.40	11883	1.61 %	0.00283	2.60
w	13587.30	13452	0.99 %	0.00050	2.89
e	14630.40	14547	0.57 %	0.00471	4.76
t	16244.30	15980	1.62 %	0.00232	2.51

Tableau 6.2 - Comparaison des fréquences modales estimées (F estimée) à l'aide des éléments finis avec les fréquences obtenues par modRes (F modRes) pour une barre (80x40 mm). Pour chaque fréquence, le mode de vibration est indiqué (mode de flexion suivant l'épaisseur e, suivant la largeur w et mode de torsion t). La déviation entre la fréquence estimée et la fréquence de résonance est indiquée ainsi que l'amplitude (Amp) et l'amortissement (Amor), la constante de la décroissance temporelle propre à chaque fréquence.

Nous avons donc un corpus de 36 sons de synthèse (12 sons « flexion », 12 sons « torsion », 12 sons « flexion+torsion » basé sur le corpus des 12 sons originaux. Sur chacun des sons, nous avons appliqué une atténuation linéaire de 25 ms, à la fin du son, pour que la durée de chaque son soit de 2.5 s.

6.4 Expériences

Les différentes séries de sons que nous avons créées, nous permettent de tester le type d'information fréquentielle auquel les auditeurs sont sensibles pour se représenter la géométrie des barres. Comme l'a montré l'étude de Lakatos et coll. [LMC97], les auditeurs semblent utiliser des propriétés liées aux rapports des dimensions « épaisseur/largeur » des barres. Le pattern spectral de chaque barre est composé de fréquences liées aux modes de flexion, de torsion et longitudinaux, les barres ayant la même longueur, cette dernière information est invariante entre les barres. Les fréquences liées aux modes de flexion ainsi que les fréquences liées aux modes de torsion sont deux informations de l'objet proximal pouvant être à la source de la discrimination perceptive de la section géométrique des barres. Nous allons tester maintenant ces différentes contributions modales, dans cette tâche d'appariement intermodal.

6.4.1 Protocole expérimental

Nous avons mené une série de quatre expériences utilisant le même paradigme expérimental que l'étude de Lakatos et coll. [LMC97]. Les quatre expériences correspondent aux trois corpus de sons que nous venons de créer par sa synthèse ainsi que les sons originaux que nous avons à disposition :

Expérience 1 ▷ Sons originaux de l'expérience de Lakatos et coll. [LMC97].
condition originale

Expérience 2 ▷ Sons de synthèse contenant l'information fréquentielle des modes de flexion suivant la largeur et suivant l'épaisseur et des modes de torsion. condition « flexion+torsion »

Expérience 3 ▷ Sons de synthèse contenant l'information fréquentielle des modes de flexion suivant la largeur et suivant l'épaisseur.
condition « flexion »

Expérience 4 ▷ Sons de synthèse contenant l'information fréquentielle des modes de torsion. condition « torsion »

6.4.1.1 Paradigme intermodal

Nous avons utilisé un paradigme dit de choix forcé à deux alternatives (appelé généralement par son acronyme anglais : 2AFC) [LMC97]. Pour chacune des quatre expériences et à chaque essai, les auditeurs entendent une séquence de deux sons. Conjointement au début du premier son apparaissent deux paires de rectangles sur l'écran de l'ordinateur. Chaque paire de rectangles

représente le rapport exact épaisseur/largeur (mis à l'échelle de l'écran) des deux barres qui ont produit les deux sons joués. Les deux ordres possibles sont présentés comme le montre la figure 6.4. L'auditeur doit alors choisir l'ordre (visuel) correspondant à la séquence sonore qu'il vient d'entendre. Du fait de la tâche, un ordre visuel correspondait à chaque fois à la bonne réponse, les auditeurs avaient donc une chance sur deux de réussir.

Nous avons repris les figures utilisées dans l'expérience de Lakatos et coll. [LMC97]. Pour chaque expérience nous avons utilisé les 12 sons correspondant aux 12 barres, voir §6.3.1, p. 180, pour les dimensions géométriques des sections des barres. Chaque paire de sons est présentée dans un ordre A ou dans un ordre B, par contre la figure associée est la même pour les deux ordres sonores.

Il a été demandé aux auditeurs d'utiliser n'importe quelle information perçue dans chacun des deux sons. Les auditeurs ont pu réécouter à chaque fois la séquence sonore le nombre de fois qu'ils désiraient avant de répondre.

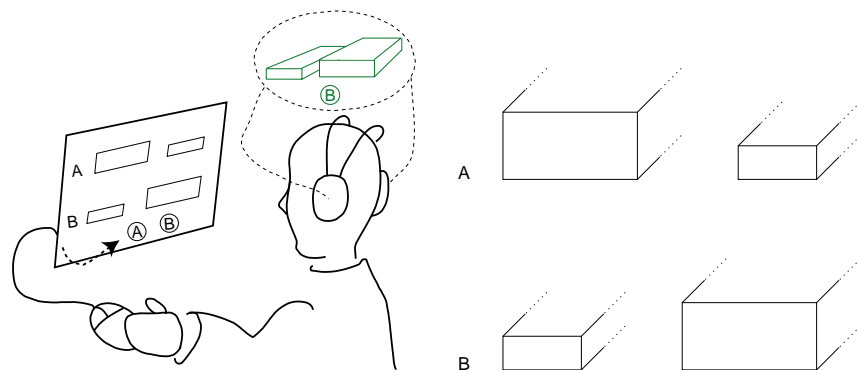


Figure 6.4 - Protocole expérimental d'appariement intermodal entre une paire de figures représentant deux barres dans un ordre A ou dans l'ordre inverse B et une séquence de deux sons produits par ces deux barres présentée simultanément. Les auditeurs doivent indiquer à quel ordre visuel A ou B correspond la séquence sonore. Une des réponses est la bonne à chaque fois.

6.4.2 Déroulement

Nous avons tenté de nous mettre dans les mêmes conditions expérimentales que l'expérience originale afin de pouvoir comparer les résultats.

Avant chacune des 4 expériences, les auditeurs ont pu manipuler un ensemble de 5 barres (longueur 15 cm : 80x40 mm, 40x40 mm ; longueur 20 cm : 80x30 mm, 60x30 mm, 40x30 mm) qui n'étaient pas présentes dans le corpus des barres des expériences pour se familiariser avec les objets. Nous

avons choisi ces dimensions afin de ne pas être exactement dans le même cas que les expériences. Par contre le marteau que nous avons utilisé, est le même que pour l'enregistrement des sons originaux. Grâce à cette manipulation des barres, les auditeurs ont pu se familiariser avec les interactions « marteau-barre ». Cette séquence dure en moyenne 5 minutes.

Ensuite les auditeurs se sont entraînés sur une série de 20 essais dans les mêmes conditions que l'expérience. Les 20 essais sont tirés au hasard. Comme dans l'expérience, aucun retour n'est donné quant à leur réponse.

Suite à cette familiarisation avec la tâche, chaque participant a passé l'expérience sur un seul bloc de 132 paires de sons correspondant aux 132 possibilités de créer une paire de sons différents avec les 12 sons de barre. L'ordre des séquences sonores A-B ou B-A ainsi que l'agencement des différentes paires visuelles sont contrôlés de manière aléatoire. Les paires regroupant la même barre ont été exclues. Cette partie dure généralement 20 minutes.

Pour chacune des 4 expériences, seul le corpus des 12 sons a changé. De plus, les groupes de participants sont différents pour chacune des 4 conditions : sons originaux, sons « flexion+torsion », sons « flexion » et sons « torsion ».

6.4.3 Remarques

Les auteurs de l'expérience originale ont utilisé ce paradigme plutôt qu'un paradigme standard de discrimination car une étude informelle a montré que les auditeurs réussissent parfaitement à discriminer deux sons de barres de section différente [LMC97]. Pour ces auteurs cette tâche mesure l'habileté à correctement apparier un changement acoustique donné avec un changement géométrique donné de la représentation de la section de l'objet. C'est une mesure d'identification relative plutôt qu'une discrimination.

Pour notre série d'expériences nous avons demandé à chaque participant d'expliquer les stratégies et les critères perceptifs utilisés. Cet entretien a eu lieu après la passation. Il nous a semblé important de procéder à cet entretien afin de nous guider dans l'analyse des résultats, notamment pour trouver des corrélats acoustiques aux propriétés perceptives utilisées pour effectuer la tâche.

6.4.4 Matériel

Tous les participants ont passé l'expérience dans une cabine audiométrique IAC. Contrairement à l'expérience originale, nous avons diffusé les sons au casque de manière dichotique (Sennheiser HD 520 II) plutôt que sur des haut-parleurs. L'interface de l'expérience utilise le programme PsiExp

[Smi95] et Max MSP [IRC93] sur une station NeXt. Elle correspond à l'interface de l'expérience originale, le but étant d'être proches des mêmes conditions. La conversion numérique/analogique est gérée par un convertisseur Pro-Port. Les sons ont été amplifiés par un amplificateur stéréo Yamaha P2075. Les sons ont été calibrés à un niveau de diffusion de 70 dB maximum en crête.

6.4.5 Participants

Quatre groupes de participants ont été recrutés pour passer les quatre expériences. Pour la réplique de l'expérience originale, 25 personnes ont participé (13 hommes et 12 femmes, pour le groupe homme : 6 non musiciens et 8 musiciens, pour le groupe femme : 5 musiciennes et 7 non musiciennes). De manière arbitraire, nous avons considéré les musiciens comme des personnes pratiquant la musique d'un point de vue professionnel ainsi que celles pratiquant régulièrement de la musique. Nous avons inclus dans ce groupe les ingénieurs du son. Pour l'expérience utilisant les sons de synthèse « flexion + torsion », 26 personnes ont été recrutées : 15 femmes (7 musiciennes, 8 non musiciennes) et 11 hommes (6 musiciens, 5 non musiciens). Le groupe de l'expérience avec les sons de synthèse « flexion » comporte 28 personnes : 13 femmes (5 musiciennes, 8 non musiciennes), 15 hommes (7 non musiciens, 8 musiciens). Pour la dernière expérience avec les sons de synthèse « torsion », 20 personnes ont participé : 11 hommes (5 musiciens, 6 non musiciens), 9 femmes (5 musiciennes, 4 non musiciennes). Ces participants ont été recrutés à partir d'une base de données à laquelle ils se sont inscrits et à l'intérieur de l'Ircam. Tous les participants ont déclaré n'avoir aucun problème d'audition.

6.5 Analyses des données

Nous allons dans un premier temps présenter les données brutes correspondant aux taux de bonnes réponses pour chacune des 4 expériences. Dans un deuxième temps, nous introduirons des analyses en arbres hiérarchiques et multidimensionnelles, dérivées des données brutes, afin de dégager la structure perceptive sous-jacente au processus d'appariement. Pour cela, le taux de bonnes réponses a été transformé en une mesure de dissemblance.

6.5.1 Analyse des données brutes

Une première remarque est que pour chacune des quatre expériences, nous trouvons des taux globaux de bonnes réponses inférieurs à 50 % pour certains

6.5. ANALYSES DES DONNÉES

participants et supérieurs à 50 % pour d'autres. Le tableau 6.3 résume la distribution des participants pour chaque condition.

condition	original	flexion+torsion	flexion	torsion
< 50 %	6 (24 %)	11 (44 %)	14 (50 %)	7 (35 %)
≥ 50 %	19 (76 %)	15 (56 %)	14 (50 %)	13 (65 %)

Tableau 6.3 - Répartition des participants suivant leur taux de bonnes réponses < 50 % ou ≥ 50 % pour les différentes conditions.

Contrairement à l'expérience originale, pour laquelle 5 participants sur 60 seulement avait un taux de bonnes réponses inférieur à 75 %, nous observons des taux inférieurs à 50 %. Cette observation est importante car elle semble indiquer que les participants utilisent des stratégies qui peuvent donner des patterns de réponses inverses, c'est-à-dire des appariements contraires (choisant l'ordre A-B alors les sons étaient dans l'ordre B-A). Nous retrouvons ces inversions quel que soit le type d'information contenu dans les sons.

Du fait de ces inversions, il ne serait pas utile d'effectuer un moyennage directe des données pour chacun des groupes de participants. Nous avons donc séparé pour chacune des expériences les résultats des participants qui ont des scores < 50 % et ≥ 50 %. Le tableau 6.4 présente la moyenne des taux de bonnes réponses pour chacun des deux groupes et ceci dans chacune des 4 conditions.

condition	originale	flexion+torsion	flexion	torsion
moyenne (groupe < 50) %	29.3	36.7	32.8	36.0
moyenne (groupe ≥ 50) %	72.5	66.6	75.3	73.5

Tableau 6.4 - Moyenne des taux de bonnes réponses sur l'ensemble des participants pour chacun des groupes (< 50 % et ≥ 50 %) pour 4 conditions.

Dans chacune des conditions nous remarquons que les scores sont en-dessous ou au-dessus du hasard (50 % puisque les participants ont deux possibilités pour répondre). Ces résultats montrent que certains auditeurs ont réussi à apparier de la bonne façon les sons avec leurs représentations visuelles mais que d'autres ont systématiquement effectué un jugement inverse. Nous voyons donc ici que des stratégies opposées semblent intervenir dans chacune des conditions. Nous remarquons aussi des pourcentages similaires, même si le type d'information contenu dans les sons varie. Par contre, pour la condition « flexion+torsion » les scores sont inférieurs aux autres conditions. Nous n'avons pas observé de pattern systématique de réponses suivant le genre ou la pratique musicale.

Au regard de ces premières analyses sur les données brutes, il semble difficile de mesurer la contribution respective des modes de flexion et des modes de torsion dans la représentation auditive des dimensions géométriques de la barre. En effet, les scores, pour la condition « flexion » sont du même ordre de grandeur que la condition « torsion ». Nous devons donc poursuivre nos analyses afin de voir si les mêmes propriétés perceptives (et leurs propriétés acoustiques concomitantes) sont engagées dans le processus d'appariement. Dans tous les cas, une information pertinente semble extraite par les participants pour effectuer la tâche.

Pour affiner nos analyses, nous transformons les données afin de pouvoir utiliser de la même manière, les données des participants dont les scores sont $< 50\%$, et ceux dont les scores sont $\geq 50\%$. Pour cela, nous transformons ces données en une mesure de dissemblance entre les paires de sons. D'un point de vue perceptif, nous associons un choix au hasard (près de 50%), à une confusion perceptive entre deux sons de barre et un appariement aléatoire à leurs représentations visuelles associées. D'un point de vue cognitif, ce score peut s'expliquer par l'incapacité à appliquer une stratégie d'association audiovisuelle. Cette confusion entre les sons implique une distance nulle entre ces sons, du point de vue de la tâche d'appariement audiovisuel amenant à un appariement aléatoire. Les scores loin de 50% reflètent une distinction forte entre les deux sons et un appariement systématique à leurs représentations visuelles associées. Ce jugement reflète alors une dissemblance perceptive élevée et/ou un taux d'association audiovisuelle élevé entre ces deux sons qui permet cet appariement systématique. Notons que cette mesure considère comme équivalents les jugements associant un changement d'une section, à une autre à un changement de hauteur montant ou descendant.

6.5.2 Transformation des données

Les données globales sont représentées sous la forme d'une matrice de proportions, qui regroupe le nombre de fois qu'une paire de sons i et j a été associée avec sa bonne paire visuelle (bonne identification). Cette matrice de proportion a été transformée en une échelle de dissemblance, voir fig 6.5.

Cette transformation des proportions globales, en une mesure de dissemblance, permet d'étudier les dimensions perceptives sous-jacentes qui contribuent au processus d'appariement. Ainsi, une réponse au hasard implique que les sources sonores n'ont pas été discriminées : nous pouvons donc l'associer à une dissemblance nulle entre ces deux sources. Dans le cas d'un appariement systématique sur l'ensemble des auditeurs entre les sons et la représentation visuelle, leur dissemblance est maximale.

Cette transformation met en valeur la discrimination entre les représen-

$$\delta(i, j) = 2 * |prop(i, j) - 0.5| \quad (6.1)$$

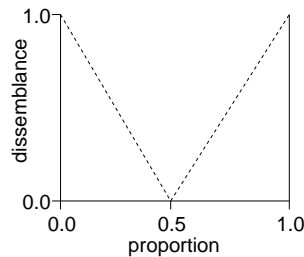


Figure 6.5 - Transformation des proportions en dissemblance. Par exemple, la confusion entre deux objets ($prop = 0.5$) est transformée en une dissemblance nulle.

tations des sources. Dans la cas d'un appariement toujours inverse entre représentations et sons, la forte dissemblance entre les sources sonores sera associée à une forte discrimination entre celles-ci et non plus à une bonne identification de la représentation géométrique des sources.

Pour chacune des expériences, nous avons dans un premier temps calculé une matrice de proportion pour les participants dont les scores sont inférieurs à 50 % et une autre pour les participants dont les scores sont supérieurs ou égaux à 50 %. Chaque matrice ($< 50\%$ et $\geq 50\%$) est transformée ensuite en une matrice de dissemblance. La matrice globale de dissemblance pour chacune des quatre conditions est la moyenne de ces deux matrices ($< 50\%$ et $\geq 50\%$). Cette matrice de dissemblance nous permet d'utiliser une représentation de ces données, par une analyse multidimensionnelle ou un arbre hiérarchique.

6.5.3 Analyse des dissemblances

Avant d'analyser les données pour dégager les propriétés sous-jacentes au processus d'appariement, nous avons voulu étudier les paires de sons pour lesquelles la dissemblance est maximum, c'est-à-dire pour les paires de sons toujours discriminées ; et au contraire pour les paires dont la dissemblance est minimale, c'est-à-dire pour les paires de sons qui ne sont jamais discriminées. Cette première analyse a pour but de relier ces patterns de réponse, aux géométries des barres associées à ces paires. Pour cela, nous avons analysé chacune des 4 matrices de dissemblance par groupements hiérarchiques. Ces analyses nous donnent la possibilité de visualiser les paires de sons très similaires (niveau bas de la hiérarchie) et les paires très différentes dont les groupes se rejoignent à un niveau de dissemblance plus élevé (niveau haut de la hiérarchie). Les analyses sont présentées dans la figure 6.6 pour les conditions originales et « flexion + torsion » et dans la figure 6.7 pour les conditions « flexion » et « torsion ».

6.5.3.1 Remarques générales

Pour les quatre conditions, nous observons une démarcation entre le groupe de barres associé à des plaques et celui associé à des blocs. Le groupe des « plaques » peut être vu comme une covariation entre l'augmentation du rapport épaisseur sur largeur et une diminution de l'épaisseur (10 mm ou 20 mm). Pour le groupe des « blocs » les rapports épaisseur/largeur sont plus faibles. Ce résultat est en accord avec l'étude originale [LMC97].

Les expériences, utilisant les sons originaux ainsi que les sons de synthèse avec l'information des modes « torsion + flexion » donnent, une structure très proche dans cette démarcation « plaques » / « blocs ». A l'intérieur du groupe « plaque », nous remarquons une très bonne cohérence dans la structure des sous-groupes qui semble plutôt liée à des épaisseurs communes. Pour la condition originale, nous observons pour le groupe des « blocs » une structure plus continue dont les dissemblances entre les paires semblent augmenter avec le rapport largeur/épaisseur. Par contre, pour la condition « torsion+flexion », nous retrouvons pour le groupe « blocs » une structure beaucoup plus catégorielle où semblent covarier l'épaisseur et le rapport épaisseur/largeur. Il n'existe pas une concordance directe, pour ce groupe, entre les sons originaux et les sons de synthèse « torsion+flexion ».

Pour la condition « flexion », il apparaît très nettement une organisation principalement basée sur l'épaisseur. Mise à part la barre 40x20 mm, nous pouvons observer une démarcation entre les plaques et les blocs et dans chacun de ces sous-groupes, des classes d'épaisseur : 30 mm et 40 mm pour le groupe des « plaques », 20 et 10 mm pour le groupe des « blocs ». Pour cette condition, le son de la barre 40x20 mm se démarque de tous les autres sons, il a la particularité d'avoir un centre de gravité spectral nettement plus élevé que les autres sons.

Pour la condition « torsion », il semble difficile de trouver, à première vue, une structure liée directement à la variation épaisseur/largeur puisque seule l'information liée aux fréquences modales de torsion est présente. Dans cette condition, le son de la barre 80x10 se dégage des autres dans cette analyse. Le son a la particularité d'avoir une fréquence fondamentale la plus basse (1274 Hz) par rapport aux autres barres ainsi que le centre de gravité spectral le plus bas. Généralement, les sons de synthèse contenant l'information de torsion ont la particularité d'être beaucoup plus aigus que les sons ayant l'information de flexion.

Suite à ces premières analyses, nous pouvons remarquer que pour la condition « flexion », un simple jugement de hauteur tonale basée sur la fréquence fondamentale du mode de flexion par rapport à l'épaisseur, peut être une information pertinente pour réussir la tâche d'appariement. De plus, l'utili-

sation de la hauteur tonale pourrait expliquer les scores inférieurs à 50 % qui montre un appariement cohérent mais inverse à la géométrie de l'objet. En effet, les participants peuvent associer une barre plate à une hauteur tonale basse, suivant leur propre représentation mentale de la géométrie de la source, mais d'autres participants peuvent associer de manière inverse une hauteur tonale basse à une barre épaisse. Pour la condition « torsion », nous ne pouvons pas dégager une utilisation unique de la hauteur, puisque les confusions entre les barres ne sont pas forcément directement liées à des rapports épaisseur sur largeur proches, les fréquences de torsion étant proportionnelles à ce rapport.

Afin d'aller plus loin dans ces analyses, nous avons procédé pour chacune des conditions à des analyses multidimensionnelles afin d'étudier une autre représentation de la structure perceptive qui sous-tend ces résultats.

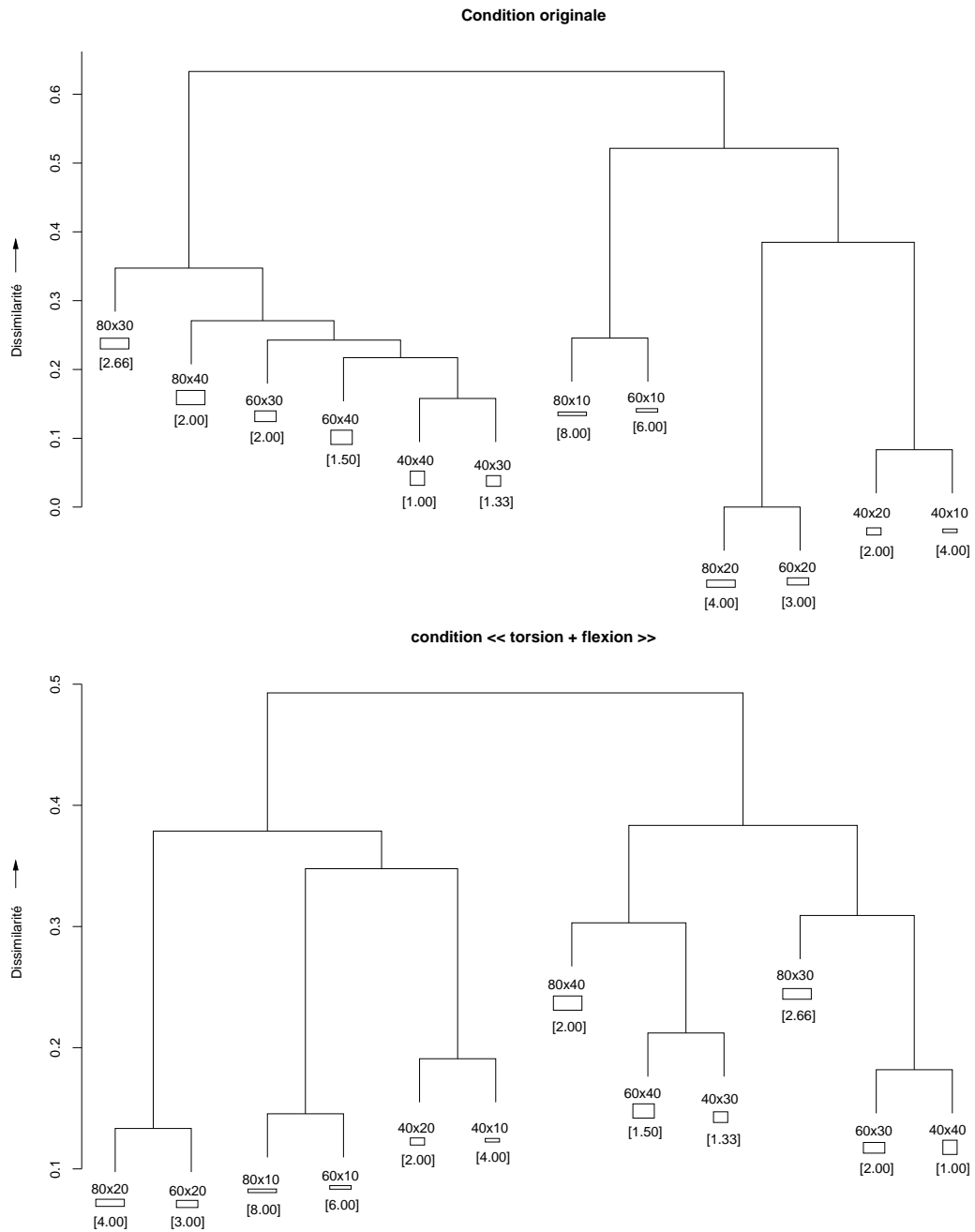


Figure 6.6 - Groupements hiérarchiques (méthode moyenne) sur les matrices de dissemblance pour les conditions originale et « torsion + flexion ». Dans chaque condition, nous avons indiqué les dimensions de chaque barre, la représentation graphique de sa section ainsi que le rapport épaisseur/largeur (entre crochets).

6.5. ANALYSES DES DONNÉES

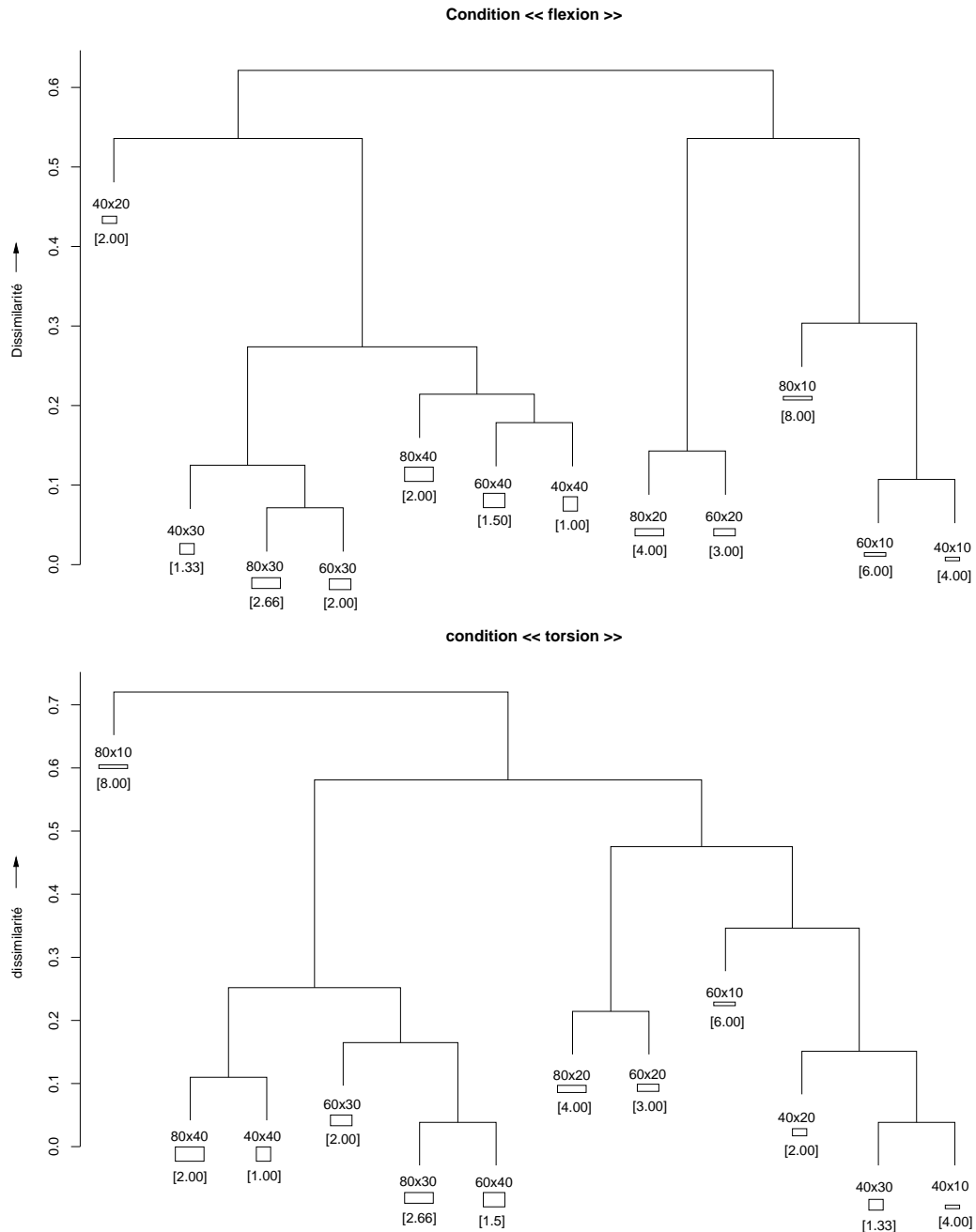


Figure 6.7 - Groupements hiérarchiques (méthode moyenne) sur les matrices de dissemblance pour les conditions « flexion » et « torsion ». Dans chaque condition, nous avons indiqué les dimensions de chaque barre, la représentation graphique de sa section ainsi que le rapport épaisseur/largeur (entre crochets).

6.5.4 Analyses multidimensionnelles

Les matrices de dissemblance à l'exception de la diagonale, ont été soumises à une analyse multidimensionnelle quasi non-métrique EXSCAL [WC89a], voir §3.4.2, p. 85. Cette analyse est quasi-non-métrique car au modèle de distance euclidienne est ajouté un calcul possible des spécificités pour chaque stimulus. L'expression de la distance est :

$$d(i, j) = \left[\sum_{f=1}^p |x_{if} - x_{jf}|^2 + s_i + s_j \right]^{1/2} \quad (6.1)$$

avec $d(i, j)$ la distance estimée par le modèle MDS entre les stimuli i et j , x_{if} la coordonnée du stimuli i le long de la dimension f et s_i la spécificité du stimulus i .

La spécificité peut caractériser une propriété unique, parfois appelé trait, d'un stimulus ou une dimension le long de laquelle seul ce stimulus varie. La variation de la grandeur de cette spécificité n'est associée à aucune des dimensions partagées par tous les stimuli dans l'espace MDS.

Nous avons utilisé l'indice de BIC pour choisir l'espace le plus approprié pour représenter les stimuli, c'est-à-dire le nombre de dimensions, avec ou sans spécificités, voir [Sch78] pour une explication de cet indice. Généralement, la valeur du BIC la plus faible donne le meilleur modèle.

Le tableau 6.5 présente les différentes valeurs de l'indice de BIC. Pour l'ensemble des expériences, les valeurs les plus faibles du BIC, sont plus faibles pour les modèles MDS sans spécificités. Les sons étant percussifs et du même matériau, ils sont assez homogènes. De plus, pour l'ensemble des expériences, un modèle à deux dimensions semble le plus adapté pour décrire les données de chaque expérience, sauf dans le cas de la condition « flexion » pour laquelle il reste une indétermination (2 ou 3 dimensions). Nous n'avons pas pu tester des modèles à une seule dimension. Le nombre de dimension est cohérent avec l'expérience de Lakatos et coll. [LMC97]. Pour la condition « flexion », nous avons fait le choix de restreindre le modèle MDS à deux dimensions pour le comparer aux autres conditions. Le choix des deux dimensions a été confirmé par les analyses en deux dimension ou trois dimensions.

L'expérience avec les sons originaux nous servira d'espace de référence pour orienter les espaces perceptifs des trois conditions de synthèse pour les comparer un à un. Pour cela, nous avons effectué des rotations procrustéennes de chaque espace des conditions utilisant des sons de synthèse par rapport à l'espace des sons originaux. La rotation procrustéenne permet de tourner une matrice par rapport à une matrice de référence, en minimisant la somme des différences au carré, afin de maximiser la similarité entre ces deux matrices.

6.5. ANALYSES DES DONNÉES

dimension	Sans spécificité	Avec spécificités
	BIC	BIC
Condition originale		
2	<u>-15.6</u>	22.8
3	-2.5	46.9
4	28.1	78.6
Condition « flexion + torsion »		
2	<u>-16.8</u>	3.0
3	-12.2	26.6
4	-3.5	46.5
Condition « flexion »		
2	<u>-33.6</u>	-9.4
3	<u>-33.8</u>	13.8
4	-9.9	39.9
Condition « torsion »		
2	<u>-43.7</u>	-4.5
3	-26.4	21.2
4	4.1	4.1

Tableau 6.5 - Valeurs de l'indice de BIC pour chacune des conditions suivant le nombre de dimensions. Une valeur faible du BIC indique un bon modèle pour représenter les données. Les valeurs les plus faibles sont soulignées.

Les matrices sont, pour nous, les dimensions de chaque son pour chacun des espaces. Les rotations sont possibles car le modèle de distance est invariant par rotation.

6.5.4.1 Espaces perceptifs

Les trois espaces des conditions « torsion + flexion », « flexion » ainsi que « torsion » sont représentés sur les figures 6.8 et 6.9. Nous avons aussi ajouté l'espace des données originales de l'expérience de Lakatos et coll. [LMC97] sur la figure 6.8. Les espaces des trois conditions de synthèse, ainsi que l'espace des données de Lakatos et coll., ont été tournés afin de les placer par rapport à notre espace perceptif des sons originaux.

■ **Remarques générale** Pour les trois conditions utilisant des sons de synthèse, nous retrouvons les classes de « plaques » et les classes de « blocs » de la même façon que nos analyses par groupements hiérarchiques. Il est intéressant de noter que la structure spatiale du groupe des « plaques » (partie gauche des espaces) est relativement semblable quelques soient les

conditions, mis à part un son pour la condition « flexion ». Ce son correspond à la barre 40×20 et ressort de la même façon sur l'arbre hiérarchique de la condition « flexion » par rapport aux autres sons et possède une particularité liée au centre de gravité spectrale que nous avons déjà mentionnée. Par contre nous observons beaucoup plus de variations de la structure spatiale dans la classe des « blocs » surtout au niveau de la deuxième dimension. Nous pouvons observer que les espaces sont plus compacts pour les conditions avec les sons de synthèse qu'avec les sons originaux. Cet effet est dû à certains sons pour chaque condition qui perturbe la rotation des espaces (40×20 et 40×40, condition « flexion + torsion » ; 40×20, condition « flexion » ; 40×30 condition, « torsion »).

Espace de Lakatos et coll. Nous avons représenté l'espace des données originales de Lakatos et coll. afin de le comparer aux résultats de notre réplique de cette expérience, voir figure 6.8. Nous remarquons que les deux espaces ont une structure spatiale très proche qui atteste de la bonne réplique de cette expérience. Les dimensions respectives de chaque espace sont corrélées à 0.92 (dimension 1) et 0.94 (dimension 2) après rotation de l'espace de Lakatos et coll. vers l'espace des sons originaux. Il faut noter que les espaces sont proches, même si, dans la réplique de cette expérience (condition « sons originaux »), le nombre d'inversion est plus élevé (scores inférieurs à 50 %). Cette réplique est importante car elle nous permet de dire que les conditions d'expérimentation sont semblables à l'expérience originale. Nous pouvons donc tester nos différentes conditions de sons de synthèse suivant le type d'information contenu dans ces sons par rapport à l'expérience originale.

Condition « flexion + torsion » Il n'existe pas une relation directe entre notre espace de référence avec les sons originaux et l'espace avec les sons de synthèse contenant les deux informations fréquentielles de torsion et de flexion. Cette divergence se traduit par des scores plus faibles dans cette condition. Suite à la bonne réplique de l'espace avec les sons originaux par rapport à l'expérience originale, nous ne pouvons pas imputer cette divergence à une variation des conditions expérimentales. Il apparaît plutôt que les sons de synthèse semblent contenir une information acoustique qui varie par rapport aux sons originaux. Ces sons sont construits sur la base des deux contributions fréquentielles : flexion et torsion. Il se peut que le fait d'avoir réduit l'information acoustique à ces seules contributions fréquentielles induit des variations par rapport aux sons originaux. Ces variations pourraient faire émerger des informations acoustiques déjà présentes dans les sons originaux mais sur lesquels les auditeurs ne pouvaient pas focaliser. Ou bien c'est une

information acoustique induite par cette synthèse que les auditeurs utilisent maintenant.

Condition « flexion » et « torsion » Il est intéressant de remarquer que lorsque l'information fréquentielle, liée aux modes de flexion et aux modes de torsion, est séparée, les espaces propres à chaque condition « flexion » et « torsion » sont plus proches de l'espace des sons originaux. Ceci se traduit par une variation plus faible de la structure spatiale dans les deux cas par rapport à l'espace des sons originaux. Les scores associés à ces deux conditions sont d'ailleurs du même ordre de grandeur que pour la condition avec les sons originaux. Cela montre que l'information contenue dans les sons de synthèse, propres aux modes de flexion ou modes de torsion, est une source potentielle d'information pour effectuer la tâche d'appariement audiovisuel. Même si, lorsque ces deux informations sont mises ensemble, les auditeurs ont plus de difficultés à extraire une information pertinente pour la tâche.

Afin d'affiner nos analyses, nous allons maintenant mettre en relation les dimensions perceptives de chaque espace avec des corrélats physiques. Ceci nous permettra d'expliquer l'information acoustique utilisée par les participants dans chacune des conditions.

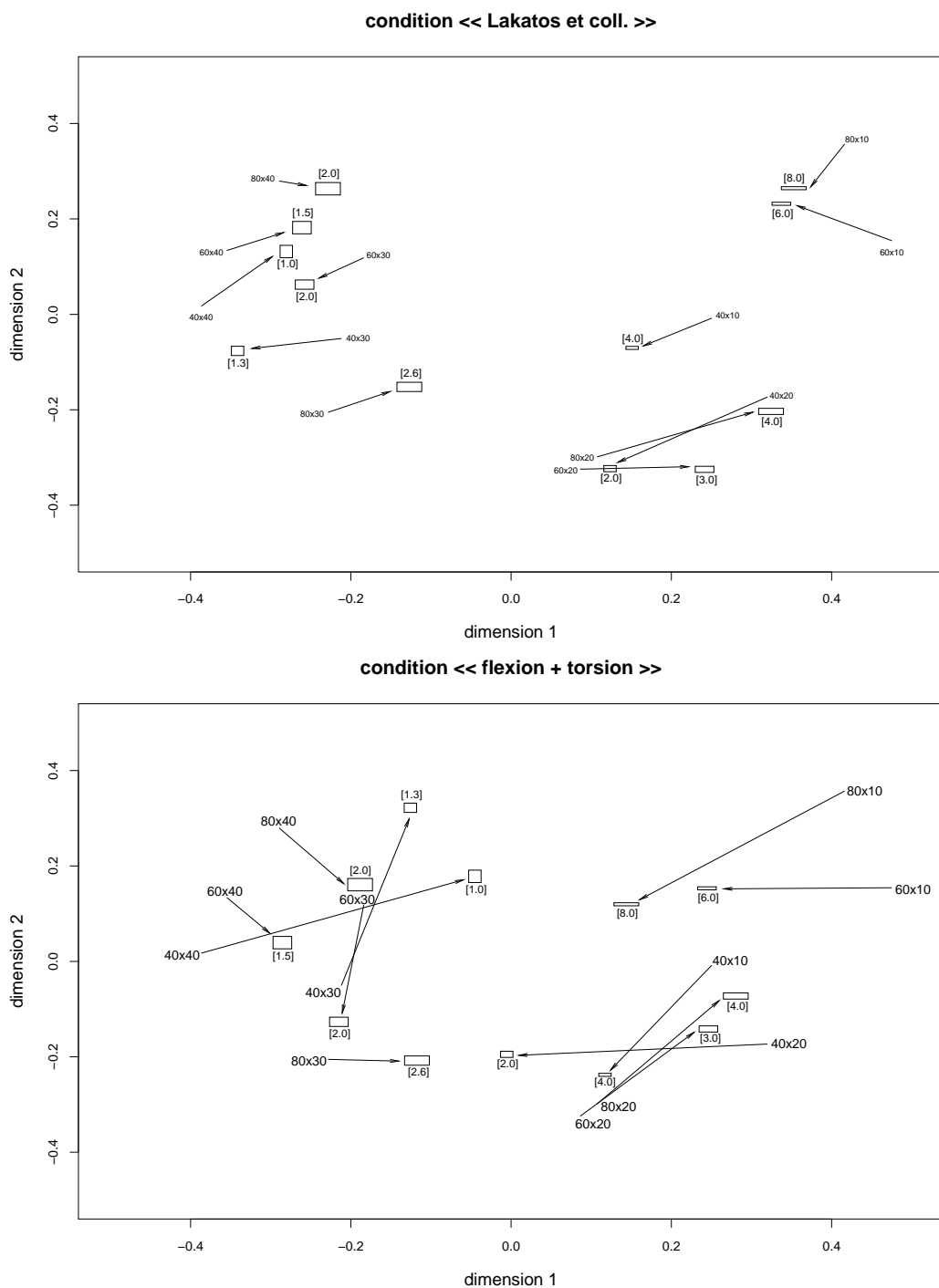


Figure 6.8 - Représentation de la solution en deux dimensions pour les conditions « Lakatos et coll. » et « flexion+torsion ». La condition « Lakatos et coll. » correspond aux résultats de l'expérience originale de Lakatos et coll. [LMC97]. Chaque solution a été tournée par rapport à l'espace deux dimensions des sons originaux. Le début de la flèche indique la position du stimulus dans l'espace avec les sons originaux (les dimensions de la barre sont indiquées). La fin de la flèche indique la position du stimulus de synthèse (la barre est représentée par son dessin). Le rapport épaisseur/largeur de chaque barre est noté entre crochets.

6.5. ANALYSES DES DONNÉES

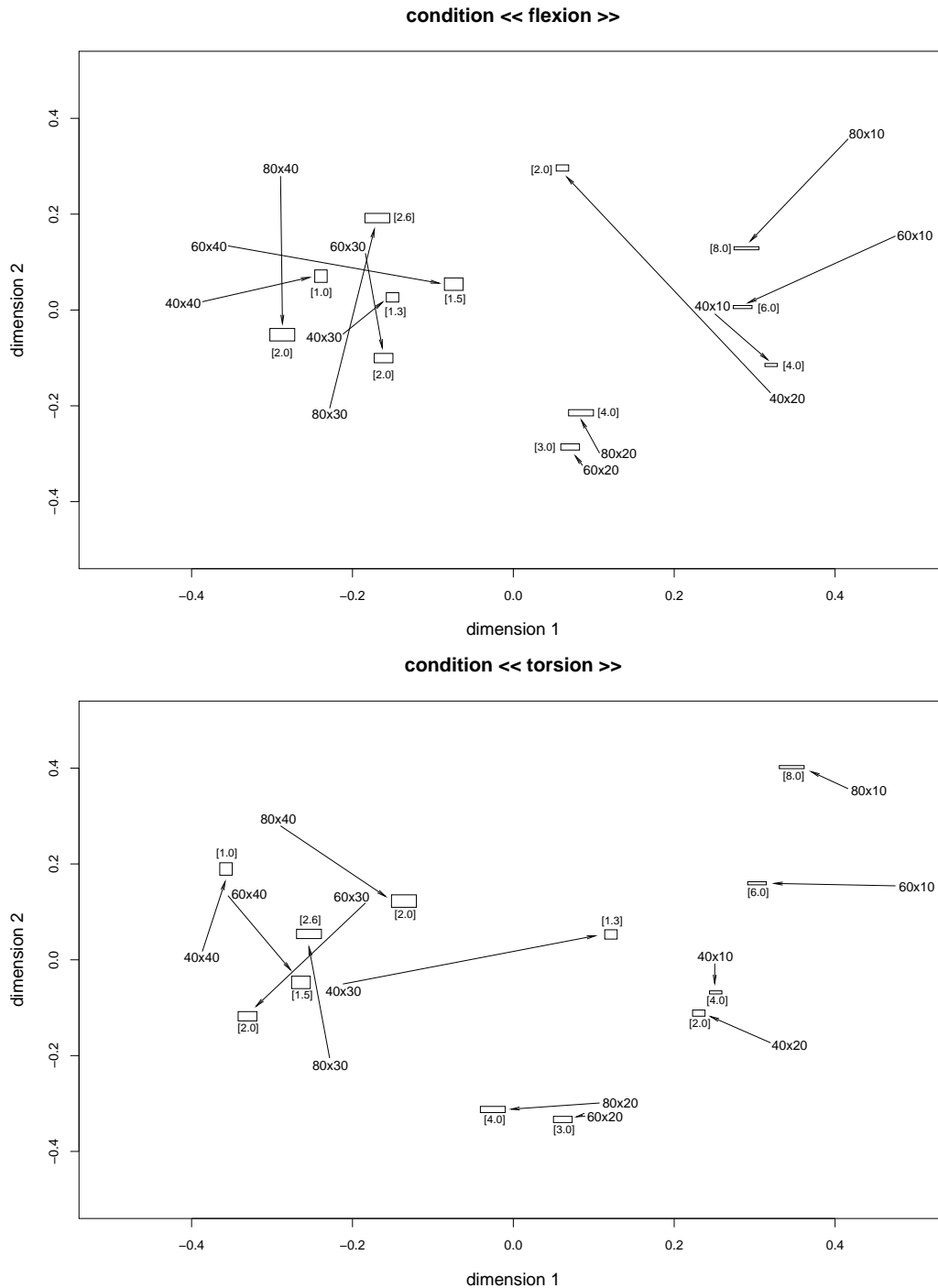


Figure 6.9 - Représentations des solutions en deux dimensions pour les conditions « flexion » et « torsion ». Chaque solution a été tournée par rapport à l'espace deux dimensions des sons originaux. Le début de la flèche indique la position du stimulus dans l'espace avec les sons originaux (les dimensions de la barre sont indiquées). La fin de la flèche indique la position du stimulus de synthèse (la barre est représentée par son dessin). Le rapport épaisseur/largeur de chaque barre est noté entre crochets.

6.5.5 Corrélats acoustiques potentiels

6.5.5.1 Verbalisations

Pour nous guider dans nos analyses, nous avons demandé à chaque participant de commenter les stratégies ainsi que les critères perceptifs utilisés lors de l'expérience. Cette verbalisation a lieu à la fin de l'expérience. Le but n'est pas de procéder à une étude sémantique complète, mais plutôt de trouver des critères revenant le plus souvent entre les participants et de sélectionner des indices acoustiques permettant d'expliquer ces critères perceptifs.

Pour les quatre expériences, les participants ont parlé implicitement ou explicitement de l'utilisation de la hauteur tonale (grave ou aiguë). Par exemple, certains ont expliqué qu'une barre peu épaisse produit un son plus grave qu'une barre plus épaisse. Certains participants ont exprimé un jugement inverse. Cette remarque est importante car le jugement des participants utilisant cette information, peut produire un appariement inverse systématique suivant la stratégie retenue. Nous pouvons mettre en relation cette indication avec une inversion systématique des appariements faits par certains sujets (scores bien inférieurs à 50 %).

Un deuxième critère est revenu très souvent. C'est la notion de résonance explicitée par des termes tels que « le son dure plus longtemps », « il résonne plus ». Généralement les participants ont associé directement la résonance à la forme de la barre. Par exemple, une grosse barre (épaisse et large) résonne plus longtemps. Parfois certains participants ont associé à la résonance des termes qui s'approchent de la brillance.

6.5.5.2 Paramètres acoustiques

Suite aux remarques formulées précédemment, nous nous sommes focalisés sur des indices liés à la hauteur tonale, à la résonance et à la brillance.

6.5.5.3 Hauteur

Un des paramètres importants est la hauteur tonale. La hauteur de chaque son peut être associée principalement à ses fréquences modales. Les barres sont excitées suivant l'épaisseur, la fréquence fondamentale est liée au mode de flexion suivant cette dimension. Elle joue un rôle prépondérant car elle est privilégiée par cette position de frappe et elle est inférieure à la fréquence fondamentale liée à la largeur, voir le tableau 6.6. Les fréquences modales liées aux modes de torsion peuvent être inférieures à la fréquence de flexion, suivant l'épaisseur, mais ont systématiquement une amplitude inférieure, du fait du mode d'excitation. Généralement les fréquences de torsion d'ordre

élevé sont les plus présentes. Ainsi, dans les conditions « sons-originaux », « flexion + torsion » et « flexion », la hauteur tonale de la barre peut être reliée à la fréquence fondamentale du mode de flexion, suivant l'épaisseur. Pour les sons de la condition « torsion », la hauteur tonale est directement liée à la fréquence fondamentale des modes de torsion. Comme ces sons sont harmoniques, même en l'absence de cette fréquence fondamentale, les auditeurs peuvent avoir la sensation de hauteur tonale directement liée à cette fréquence fondamentale.

fréquence mode de torsion		fréquence mode de flexion					
b/h	fréq (Hz)	b/h	fréq (Hz)	h (mm)	fréq (Hz)	b (mm)	fréq (Hz)
1	4936	3	3025	40	2190	80	3866
1.5	4563	4	2378	30	1610	60	3040
2	3993	6	1654	20	1118	40	2190
2.6	3284	8	1279	10	550		

Tableau 6.6 - Fréquences du premier mode de torsion suivant le rapport épaisseur/largeur (h/b) et du mode de flexion suivant l'épaisseur (h) ou la largeur (b). Analyses acoustiques effectuées par modRes.

6.5.5.4 Résonance

Pour spécifier la notion de résonance, deux paramètres nous ont semblé important : le temps de décroissance ainsi qu'un paramètre associé au centre de gravité temporel. Le premier paramètre peut être lié aux méthodes d'analyse que nous avons présentées, voir §4, p. 116. Pour un son percussif, après l'attaque, l'enveloppe temporelle peut être vue comme une fonction exponentielle décroissante. La décroissance temporelle (notée *decr*) est estimée en considérant la décroissance de l'enveloppe temporelle sous une forme exponentielle $exp(-at)$. Le paramètre α est appelé facteur d'amortissement [Rou99]. Plus le facteur d'amortissement augmente plus le son se dissipe rapidement. Ce facteur joue un rôle important dans la perception du matériau [Rou99] [KPK00].

Un autre paramètre important est le centre de gravité temporel qui semble approprié pour les sons de type percussif. Lakatos [Lak00] a montré que l'espace perceptif du timbre des instruments de musique percussifs peut être expliqué par trois dimensions. Peeter et coll. [PMH00] ont montré que les corrélats acoustiques de ces dimensions sont : le logarithme du temps d'attaque, le centroïde spectral et le centre de gravité spectral. Le centre de gravité temporel (CT) peut être calculé en estimant tout d'abord l'enveloppe temporelle du signal appelé *SEnv*. L'enveloppe temporelle *SEnv* est obtenue en

appliquant un filtre passe-bas sur le spectre du signal qui lisse l'évolution temporelle du signal.

L'expression de CT est :

$$CT = \frac{\sum_{n=1}^{\text{longueurde}(SEnv)} \frac{n}{sr} \cdot SEnv(n)}{\sum_{n=1}^{\text{longueurde}(SEnv)} SEnv(n)} \quad (6.1)$$

avec $SEnv$, l'enveloppe temporelle du signal, n le numéro de l'échantillon et sr la fréquence d'échantillonnage.

Le centre de gravité temporelle est directement lié à la résonance du son car plus le son résonne longtemps, plus le CT augmente.

6.5.5.5 Le centre de gravité spectral

Comme nous l'avons déjà mentionné lorsque nous avons présenté le timbre, voir §2.4, p. 16, le centre de gravité spectral associé à la notion de brillance est un prédicteur acoustique qui apparaît souvent dans les études sur la perception du timbre. Le centre de gravité spectral est la moyenne des fréquences du spectre pondérées par leurs amplitudes.

Le centre de gravité spectrale est extrait en déterminant tout d'abord le spectre de puissance sur le signal. Le centroïde spectral (CS) est calculé par l'expression :

$$CS = \frac{\sum_{k=1}^{\text{longueurdespectre}} f(k) \cdot S(k)}{\sum_{k=1}^{\text{longueurdespectre}} S(k)} \quad (6.1)$$

avec $f(k)$ la fréquence et le coefficient $S(k)$ du spectre de puissance associé au partiel k . Le centre de gravité spectral est calculé sur des petites fenêtres temporelles, puis moyenné.

Ce paramètre joue, lui aussi, un rôle important dans l'étude de la perception des matériaux [Rou99] [KPK00] et du timbre des sons percussifs [PMH00].

6.5.6 Analyses des corrélations

Pour chacun des espaces, nous avons corrélé les dimensions perceptives aux différents corrélats acoustiques qui semblent pertinents au regard des verbalisations des participants.

■ **Espace des sons originaux** Dans le tableau 6.7, nous présentons les différentes corrélations entre les paramètres physiques et les deux dimensions perceptives.

6.5. ANALYSES DES DONNÉES

Paramètre	Corrélation	
	Dim 1	Dim 2
h	-0.95	0.07
log(b/h)	0.83	0.15
CT	-0.61	-0.52
decr	-0.58	-0.64
CGS	-0.33	0.52

Tableau 6.7 - Corrélation entre les dimensions perceptives et les paramètres acoustiques : centre de gravité temporelle CT, centre de gravité spectral CGS et décroissance temporelle. h l'épaisseur et b la largeur. Condition « sons originaux ».

Contrairement à l'expérience originale, la dimension 1 semble mieux corrélée avec l'épaisseur qu'avec le rapport largeur/épaisseur. Sur la figure 6.10, nous avons tracé la variation entre la première dimension et l'épaisseur à gauche et le logarithme du rapport épaisseur/largeur à droite. A la vue de ces deux courbes, il semble difficile de trancher, même si la corrélation donne un avantage à l'épaisseur. Nous pouvons remarquer sur les deux courbes que la barre 40x20 semble venir perturber les données. En retirant les valeurs associées à cette barre, les corrélations augmentent pour les deux paramètres : épaisseur ($R = -0.97$) et $\log(W/H)$ ($R = -0.93$). Mais cela ne permet pas de départager ces deux paramètres. L'épaisseur est naturellement fortement corrélée avec la fréquence fondamentale du mode de flexion suivant l'épaisseur ($R = 0.99$) et la fréquence de torsion avec le logarithme du rapport épaisseur/largeur ($R = -0.98$). Ces deux informations peuvent donc expliquer la première dimension.

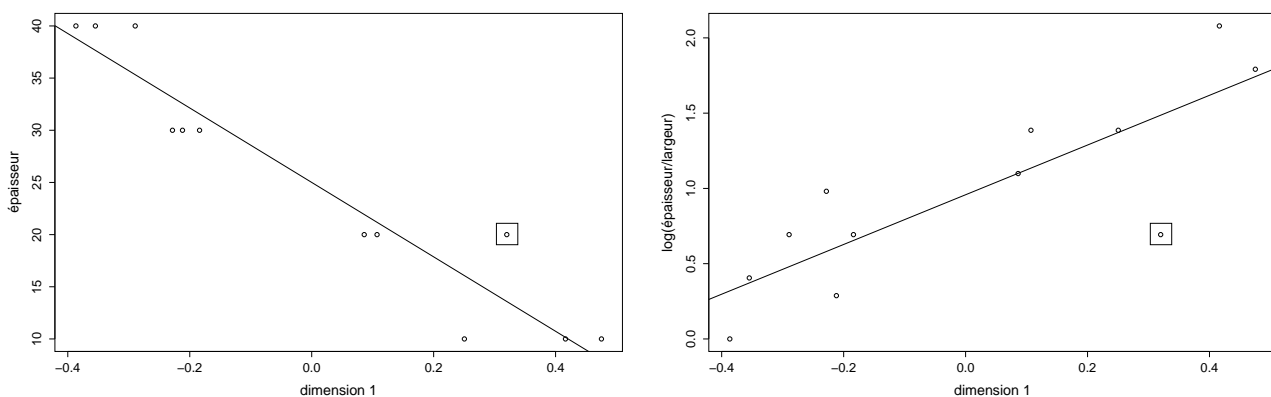


Figure 6.10 - Courbes représentant la variation de la dimension 1 avec l'épaisseur et la variation de la dimension 1 avec le logarithme du rapport épaisseur/largeur. Condition avec les sons originaux. La barre 40×20 est encadrée.

Pour la deuxième dimension, il n'y a pas de corrélation significative pour les paramètres tels que les centres de gravité spectral et temporel, ainsi que pour la décroissance. Nous avons tourné systématiquement l'espace des deux dimensions par rotation procrustéenne suivant deux axes liés aux paramètres physiques. Le première axe est l'épaisseur ou bien le logarithme du rapport épaisseur/largeur. Le deuxième axe est le CT ou la decr ou encore le CGS. Nous avons effectué toutes les combinaisons possibles. Les résultats montrent que pour le logarithme du rapport épaisseur/largeur, dans aucun cas nous n'observons une augmentation de la corrélation pour ce paramètre avec la dimension 1 de l'espace tourné. De la même façon, les paramètres CT ou CGS ou decr ne sont pas mieux corrélés avec la dimension 2 de l'espace tourné. Pour la rotation de l'espace impliquant comme axes l'épaisseur la décroissance temporelle, nous observons une augmentation entre l'axe liée la décroissance temporelle et la dimension 2 de l'espace tourné. En effet, lorsque nous tournons l'espace de référence vers l'épaisseur (dimension 1) et la décroissance temporelle (dimension 2), la corrélation entre l'épaisseur et la dimension 1 diminue à $R = 0.85$, mais la corrélation entre la décroissance temporelle et la dimension 2 est maintenant de $R = 0.85$. Par contre la décroissance temporelle est aussi corrélée légèrement avec la première dimensions à $R = 0.69$. Les figures de variation des dimensions avec ces deux paramètres sont présentées sur la figure 6.11. Une combinaison linéaire de ces paramètres ne donne pas de résultat.

Les résultats de ces analyses montrent qu'il est difficile de départager l'épaisseur et le rapport épaisseur/largeur pour expliquer la première dimension perceptive. Les rotations que nous avons effectuées, permettent de montrer que l'épaisseur semble une meilleure explication de la première dimension et que la décroissance temporelle tend à décrire la deuxième dimension. De plus, nous avons observé des inversions dans les résultats, pour cette condition. L'utilisation de la hauteur liée à la fréquence fondamentale du mode de flexion, suivant l'épaisseur, va dans le sens de nos analyses.

■ **Espace des sons « flexion + torsion »** Pour l'espace perceptif des sons contenant l'information fréquentielle de flexion et de torsion, nous présentons les corrélations entre les dimensions perceptives et les différents paramètres physiques dans le tableau ci-dessous :

Nous observons des corrélations beaucoup plus faibles que l'espace des sons originaux qui reflètent les résultats des données d'appariement. La première dimension est corrélée avec la dimension épaisseur. Il est intéressant de voir que le CGS est corrélé avec la dimension 1, même si la corrélation reste faible. Par contre, aucun des paramètres ne semble expliquer la deuxième

6.5. ANALYSES DES DONNÉES

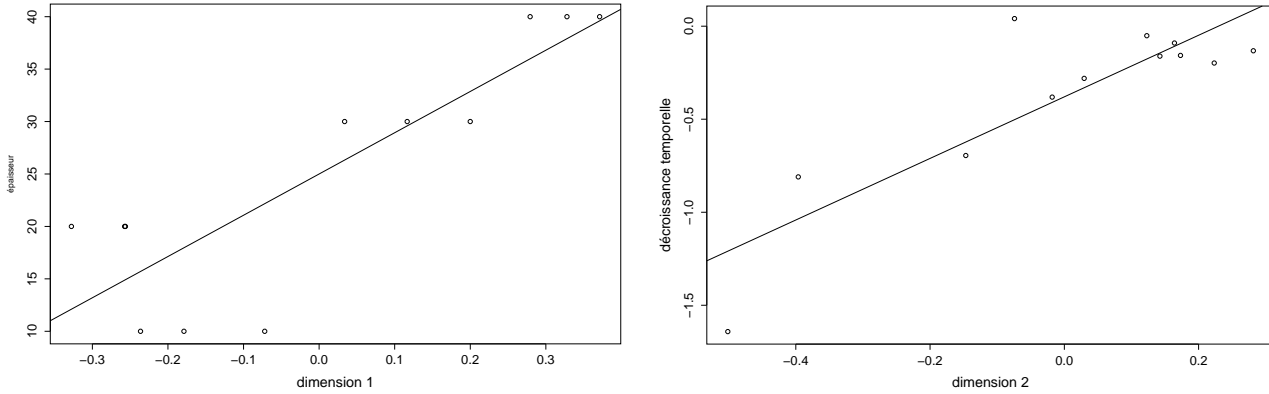


Figure 6.11 - Courbes représentant la variation de la dimension 1 avec l'épaisseur et la variation de la dimension 2 avec la décroissance. L'espace perceptif a été tourné vers ces deux dimensions. Condition avec les sons originaux.

Paramètre	Corrélation	
	Dim 1	Dim 2
h	-0.81	-0.27
$\log(b/h)$	0.67	0.30
CT	-0.27	-0.19
decr	-0.04	-0.29
CGS	-0.58	0.20

Tableau 6.8 - Corrélation entre les dimensions perceptives et les paramètres acoustiques : centre de gravité temporelle CT, centre de gravité spectral CGS et décroissance temporelle. Condition « flexion + torsion ».

dimension. Nous avons testé l'influence du centre de gravité spectral sur la première dimension, en effectuant une regression linéaire entre la première dimension et les deux paramètres épaisseur et CGS. La combinaison linéaire de l'épaisseur et du CGS : $f(\text{épaisseur}, \text{CGS})^{[a]}$, augmente la corrélation avec la dimension 1 ($R = 0.89$). Cette corrélation est significative, par régression, incrémentielle. qui traduit l'influence du CGS dans cette première dimension. La figure 6.12 montre les variations de la dimension 1 avec l'épaisseur ainsi qu'avec la combinaison linéaire de l'épaisseur et du CGS. Les résultats des expériences d'appariement sont plus faibles pour cette condition. Il se peut que la hauteur tonale, liée à la fréquence de flexion suivant l'épaisseur, soit utilisée par les auditeurs pour effectuer les appariements, mais que l'information liée au centre de gravité spectral des sons perturbe leurs réponses.

^[a] $-0.01436 * \text{épaisseur} - 0.00006357 * \text{CGS} + 7.779e - 01$

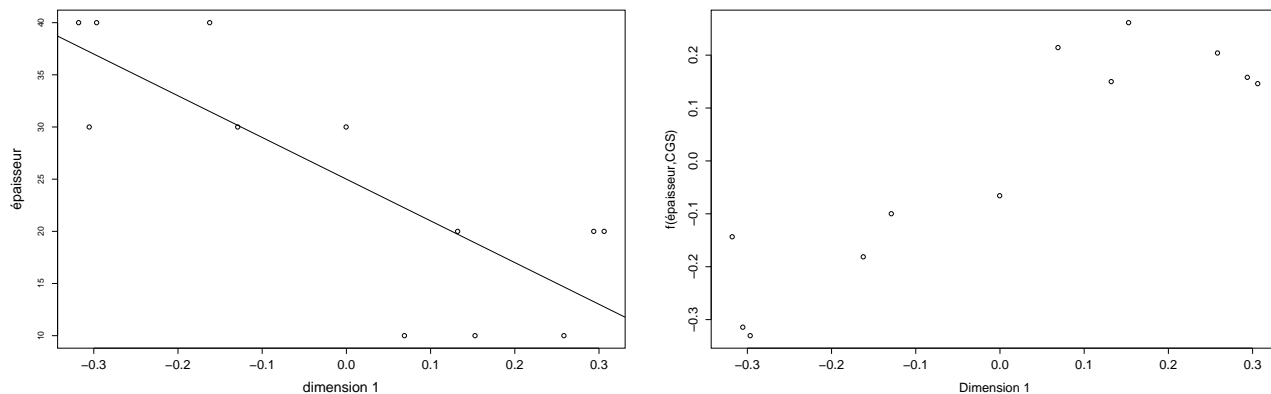


Figure 6.12 - Courbes représentant la variation de la dimension 1 avec l'épaisseur et la variation de la dimension 1 avec une combinaison linéaire de l'épaisseur et du CGS. Condition « flexion + torsion ».

■ **Espace des sons « flexion »** Les analyses par groupements hiérarchiques ainsi que l'espace perceptif montrent une forte dépendance des résultats vis à vis de l'épaisseur de la barre. Dans le tableau 6.9 résumant les corrélations entre dimensions perceptives et paramètres physiques, nous observons une forte corrélation entre la dimension 1 et l'épaisseur. Par contre, la dimension 2 ne semble être expliquée par aucun de ces paramètres. Comme précédemment, il nous a semblé intéressant d'effectuer des regressions linéaires entre l'épaisseur et les différents paramètres pour expliquer la première dimension. Nous avons trouvé une très bonne corrélation ($R = 0.97$) entre la dimension 1 et une combinaison de l'épaisseur et du centre de gravité temporel^[a]. Cette corrélation est significative par regression incrémentielle. Cette corrélation peut trouver une résonance dans les commentaires des participants qui ont souvent associé la notion de grave et aigu conjointement avec la résonance. La figure 6.13 montre les variations entre la dimension 1 et l'épaisseur ainsi qu'avec la combinaison linéaire entre l'épaisseur et le centre de gravité temporel. Même si l'épaisseur a un rôle prépondérant dans cette dimension, l'apport du centre de gravité temporelle est significatif. Nous n'avons pas trouvé cette influence pour le rapport épaisseur/largeur. Nous voyons que l'écoute de la hauteur, liée à la fréquence fondamentale du mode de flexion, suivant l'épaisseur, facilité par le centre de gravité temporel semblent, être des informations pertinentes pour réussir la discrimination entre les barres.

[a] $-0.02696 * \text{épaisseur} + 0.423426 * CT + 0.606938$

Paramètre	Corrélation	
	Dim 1	Dim 2
h	-0.93	0.16
log(b/h)	0.83	-0.10
CT	-0.14	0.08
decr	0.00	0.02
CGS	-0.34	-0.07

Tableau 6.9 - Corrélation entre les dimensions perceptives et les paramètres acoustiques : centre de gravité temporelle CT, centre de gravité spectral CGS et décroissance temporelle. Condition « flexion ».

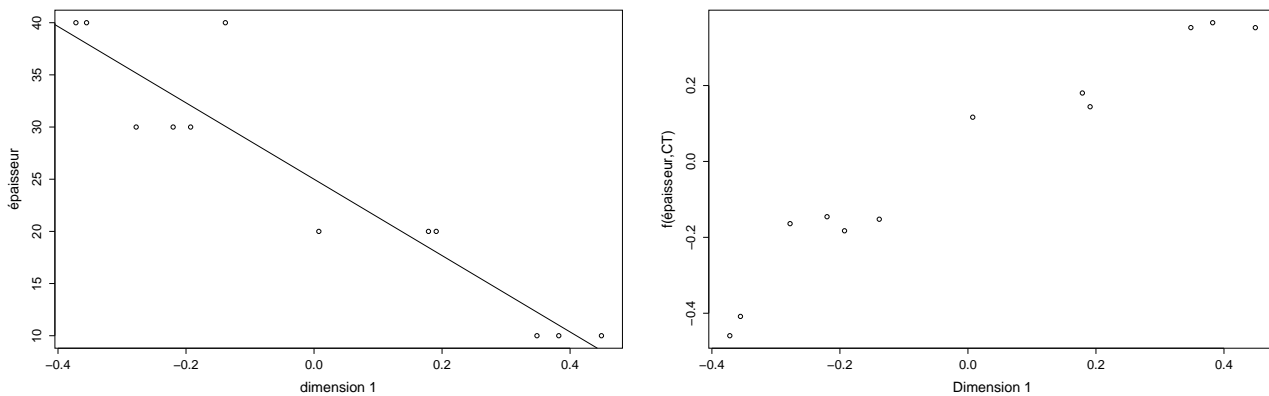


Figure 6.13 - Courbes représentant la variation de la dimension 1 avec l'épaisseur et la variation de la dimension 1 avec une combinaison linéaire de l'épaisseur et du CT. Condition « flexion ».

■ **Espace des sons « torsion »** Contrairement aux sons contenant l'information fréquentielle des modes de flexion, les sons « torsion » n'ont pas d'information liée à l'épaisseur ou la largeur, mais uniquement l'information fréquentielle liée au rapport épaisseur/largeur. C'est pourquoi dans le tableau 6.10 présentant les corrélations entre paramètres physiques et dimensions perceptives, l'épaisseur n'apparaît plus. Pour cette expérience nous observons une corrélation nettement moins élevée entre la dimension 1 et le premier paramètre physique, ici le logarithme du rapport épaisseur/largeur. La dimension 2 n'est expliquée par aucun des ces paramètres. Par contre il existe des corrélations assez élevées entre la première dimension et les paramètres liés à la résonance, tels que le centre de gravité temporel et la décroissance. Ces différentes corrélations élevées, nous ont orienté à faire des regressions linéaires entre chacun de ces paramètres (CT, decr, CGS) associé avec le logarithme du rapport épaisseur/largeur. Nous obtenons des corrélations significatives par regression incrémentielle pour une combinaison additive du

rapport avec la décroissance^[a], avec la dimension 1 ($R = 0.90$). Cette corrélation reste inférieure à celle obtenue pour les paramètres de l'espace des sons de synthèse « flexion ». Les participants ayant aussi bien réussi cette expérience que l'expérience précédente, l'information liée à cette combinaison linéaire peut être, elle aussi, une source d'explication.

Paramètre	Corrélation	
	Dim 1	Dim 2
log(b/h)	0.71	0.02
CT	-0.67	0.09
decr	-0.73	-0.14
CGS	-0.34	-0.01

Tableau 6.10 - Corrélation entre les dimensions perceptives et les paramètres acoustiques : centre de gravité temporelle CT, centre de gravité spectral CGS et décroissance temporelle. Condition « torsion »

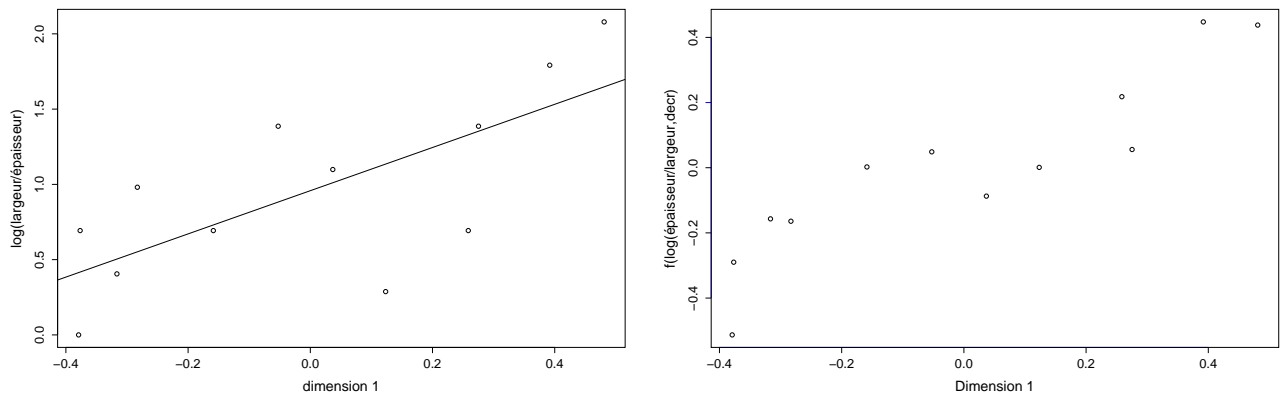


Figure 6.14 - Courbes représentant la variation de la dimension 1 avec le logarithme du rapport épaisseur/largeur et la variation de la dimension 1 avec une combinaison linéaire du logarithme du rapport épaisseur/largeur et de la décroissance. Condition « torsion ».

^[a] $0.27542 * \log(\text{épaisseur}/\text{largeur}) - 645.2 * \text{decr} - 0.55695$

6.6 Discussion

Dans cette partie, nous avons étudié l'apport des fréquences modales de flexion et de torsion dans la représentation de la section géométrique d'une barre. En effet, ces fréquences véhiculent une information potentielle sur la section géométrique des barres. Cette information est liée à la comparaison des fréquences modales associées à l'épaisseur et la largeur ou directement aux fréquences modales de torsion proportionnelles à la section géométrique.

Pour cela, nous avons synthétisé différents sons contenant l'information fréquentielle liée aux modes de flexion et/ou de torsion sur la base des sons originaux de l'expérience de Lakatos et coll. Nous avons repris leur paradigme d'appariement audiovisuel entre les sons de barres et leurs représentations visuelles. Quatre expériences utilisant ce même protocole ont été menées : condition avec les sons originaux, celles avec les sons contenant l'information fréquentielle des modes de flexion et torsion, des modes de flexion seuls et de torsion seuls.

Pour l'expérience utilisant les sons originaux, il nous est difficile de départager les contributions liées à l'épaisseur ou au rapport épaisseur/largeur pour expliquer les résultats des appariements. La première dimension de l'espace perceptif représentant ces données est corrélée de façon similaire à ces deux informations. La deuxième dimension associée au paramètre acoustique de décroissance temporelle peut contribuer à expliquer les résultats des appariements. Cette réplique de l'expérience avec les sons originaux fournit des résultats similaires à l'expérience de Lakatos et coll. Du fait de la corrélation moins importante du paramètre acoustique associé à la deuxième dimension (la décroissance temporelle), ceci peut expliquer la différence avec le résultat de l'expérience originale dont la deuxième dimension est corrélée avec le centre de gravité spectral. L'information acoustique associée à la deuxième dimension semble être plus difficile à mettre en lumière. De plus, dans la condition utilisant les sons de synthèse, les analyses n'ont dégagé aucune deuxième dimension, même en étendant nos analyses à d'autres paramètres acoustiques. Malgré tout, notamment pour la condition « torsion » et « flexion », les participants ont apparié de manière assez systématique les sons avec leurs représentations visuelles. Il apparaît donc que la deuxième dimension dans toutes les conditions, ne soit pas le reflet d'une information pertinente utilisée par les auditeurs dans cette tâche.

Dans le cas des sons de synthèse « flexion + torsion », les résultats s'écartent de l'expérience originale. L'espace perceptif associé aux résultats de l'expérience, présente des divergences dans la structure spatiale par rapport à l'espace de Lakatos et coll. Il semble que les sons de synthèse ne contiennent pas toute l'information acoustique extraite par les auditeurs dans les sons

originaux pour effectuer la tâche d'appariement.

Pour répondre à nos questions, il nous semble plus pertinent de comparer les résultats des expériences entre les conditions « flexion » et « torsion ». Pour les espaces associés aux résultats de ces expériences, nous avons pu dégager des critères acoustiques associés uniquement à la première dimension. Pour la condition « flexion », l'information pertinente est relié à une combinaison linéaire de l'épaisseur et du centre de gravité temporel. Il est à noter, que l'information liée à l'épaisseur reste le facteur explicatif le plus important. Pour la condition « torsion », les résultats peuvent être interprétés grâce à une combinaison entre le logarithme du rapport épaisseur/largeur et la décroissance. Pour ces deux expériences, les informations perceptives utilisées par les auditeurs pour apparier de manière systématique les sons avec leurs représentations visuelles associées sont différentes entre les deux conditions.

Deux remarques importantes peuvent éclairer ces résultats. Premièrement, l'amplitude des fréquences liée aux modes de torsion a été diffusée à un niveau sonore supérieur par rapport aux sons originaux. Ainsi, dans les sons originaux, la contribution des modes de résonance de torsion est à un niveau beaucoup plus faible que les modes de résonance liés aux modes de flexion. Dans la condition « torsion », nous avons donc focalisé artificiellement les auditeurs sur cette information.

Un deuxième problème est associé à un biais du paradigme expérimental développé par Lakatos et coll. [LMC97] que nous avons utilisé pour ces expériences. Ce biais doit être relié au choix de la distribution des dimensions d'épaisseur et de largeur des barres. Cette distribution est telle que lors de la présentation des paires graphiques des sections des barres, les participants ont peu d'essais dans lesquels les deux barres sont de même épaisseur. Ces essais particuliers sont au nombre de 24 sur les 132 paires visuelles des sections de barres. Ainsi dans 88 % des cas, les participants visualisent deux sections représentant des barres d'épaisseur différentes qu'ils doivent associer au bon ordre de la séquence sonore.

Les participants ont systématiquement mentionné l'utilisation de la hauteur tonale pour procéder à l'association de la séquence sonore avec la séquence visuelle. Ainsi, les sujets ont pu utiliser une stratégie qui consiste à affecter un changement montant ou descendant de hauteur perçue, à un agrandissement ou rétrécissement (ou vice versa) de l'épaisseur des barres représentées à l'écran. C'est une stratégie qui infère une relation audiovisuelle plutôt que de la percevoir directement. Ce résultat apparaît clairement dans la condition « flexion » pour laquelle la hauteur liée à l'épaisseur et la fréquence de flexion associée serait extraite par les participants pour effectuer la tâche. Cette stratégie permettrait d'expliquer les appariements inverses

systematiques effectués par certains participants.

Suite aux commentaires des participants et à nos résultats, il semble difficile d'affirmer que le son des barres véhicule une information permettant de se représenter mentalement les dimensions de la source.

D'autres auteurs [KPT00] ont montré que les auditeurs sont capables de juger les dimensions d'une plaque sur la base d'une pondération des fréquences modales liées à chaque dimension. Nos résultats semblent montrer plutôt une prééminence perceptive de la hauteur.

Chapitre 7

Conclusions



7.1 Introduction

La série d'expériences menée lors de cette thèse s'est inscrite dans une approche critique des théories de la perception directe et du traitement de l'information pour l'étude de la reconnaissance des événements sonores. Nous proposons un cadre théorique, l'acoustique écologique, qui s'inspire de ces deux approches. Les travaux antérieurs sur la perception du timbre se sont avant tout focalisés sur la recherche des dimensions perceptives sous-jacentes et leurs corrélats acoustiques. Cette approche est associée à la théorie du traitement de l'information car elle met l'emphasis sur des processus internes basés sur ces dimensions permettant la reconnaissance des événements sonores. Même si ces études cherchent implicitement des propriétés perceptives engagées dans la reconnaissance des sons, elles ne permettent pas d'expliquer comment nous sommes sensibles aux propriétés physiques des sources. La perception directe part du principe que nous sommes accordés à notre environnement, nous avons la capacité à extraire les invariants du stimulus proximal spécifiant les propriétés des sources sonores, leur reconnaissance est donc directe. L'acoustique écologique a pour but de relier spécifiquement

la perception auditive d'une source vibrante à ces propriétés physiques. Ces propriétés audibles des sources pourront nous donner une information sur la nature causale de l'événement sonore telle que son matériau, son type excitation ou sa géométrie et ainsi fournir une source d'information pour la reconnaissance de l'événement sonore.

Dans le cadre de cette thèse, nous nous sommes penchés sur un objet simple : une barre de métal. En frappant sur une barre de métal, nous avons voulu savoir si les auditeurs sont capables d'extraire une information acoustique invariante permettant de se représenter auditivement la géométrie de l'objet. Pour cela nous avons fait varier la section géométrique de différentes barres de même matériau en les excitant en différentes positions. Dans un deuxième temps nous avons tenté de spécifier les propriétés acoustiques (les fréquences modales en l'occurrence) nous permettant de nous représenter auditivement la section géométrique d'une barre.

Nous allons présenter une synthèse des résultats de nos expériences que nous mettrons ensuite en perspective.

7.2 Synthèse des résultats

7.2.1 Catégorisation auditives

Pour notre première série d'expériences, nous avons excité différentes barres de même longueur variant par leurs sections géométriques. Les expériences de Lakatos et coll. [LMC97] ont montré que les auditeurs se représentent auditivement la section d'une barre de métal sur la base d'informations acoustiques liées aux modes de vibration. En excitant en différents points une barre, nous avons cherché à éclairer de différentes manières le pattern spectral de la barre, sa signature acoustique. Nous cherchions à savoir si les auditeurs ont la capacité à extraire une information acoustique pertinente permettant d'associer les sons d'une barre frappée à différents endroits. Cette capacité doit être mise en relation avec la faculté de juger les dimensions d'un objet mis en vibration sur la base du son [LMC97] [CAKP98] [Gav88] [KPT00].

Pour cette série d'expériences, nous avons utilisé le paradigme de classification libre de sons. Ce paradigme a eu pour but de mettre en lumière des classes de sons pouvant refléter des catégories auditives. Ces catégories nous permettent de trouver les propriétés sous-jacentes à une organisation auditive d'un ensemble de sons. Cette méthode a permis dans d'autres recherches d'étudier les propriétés auditives liées à la représentation de sons environnementaux [Guy96] [Maf99]. Les techniques d'analyse des données de

classification libre nous ont permis de proposer une méthode originale afin de trouver les catégories pertinentes pour un ensemble d'auditeurs.

Les résultats de la première série d'expériences montrent que les auditeurs n'ont pas réussi à extraire des invariants acoustiques propres à chaque barre lorsque nous ne donnons aucune information sur les sources sonores. Dans le cas où nous fournissons une information plus complète sur la source, les auditeurs n'arrivent pas à inférer les propriétés audibles de la source sur la base d'une représentation de la section géométrique de chaque barre. Ainsi dans ces différentes conditions les auditeurs n'ont pas regroupé les sons propres à chaque barre.

Les catégories sont organisées suivant des dimensions privilégiées liées à l'épaisseur ou la largeur des barres. Les dimensions peuvent être associées à une écoute des hauteurs spectrales les plus saillantes liées aux fréquences modales de flexion suivant la largeur ou l'épaisseur.

Ces résultats ont suggéré que les auditeurs ont utilisé une similarité de surface, liée aux propriétés du signal acoustique, dans leur processus de classification.

Une expérience d'appariement nous a permis de connaître les différentes hauteurs spectrales entendues dans chacun des sons de barre. Cette expérience nous a fourni pour chaque son la distribution des hauteurs spectrales qui peuvent être impliquées dans les processus de comparaison lors des tâches de classification. A partir de ces données, nous avons proposé une mesure de similarité entre chaque paire de sons basée sur un modèle de contraste de Tversky [Tve77] pour expliquer les catégories effectuées par les auditeurs. Nous pensons que les auditeurs comparent les différents sons sur la base de traits perceptifs communs : les hauteurs spectrales. Un modèle similaire basé sur des analyses acoustiques a montré que l'information fréquentielle directement liée aux modes de vibration de flexion et de torsion pouvait permettre d'associer les sons provenant d'une même barre. Nous avons comparé les résultats de notre modèle dérivé de l'expérience d'appariement avec les résultats des expériences de classification. Ce modèle a révélé que les auditeurs utilisaient les hauteurs spectrales proéminentes liées aux modes vibratoires d'une dimension privilégiée afin de comparer et classifier ces sons. Il n'apparaît pas que les auditeurs utilisent l'information acoustique brute afin d'extraire des invariants perceptifs propres à chaque barre.

D'un point de vue théorique cette série d'expériences montre que les auditeurs n'ont pas la capacité à extraire des invariants structuraux propres à chaque barre. Ce résultat va à l'encontre de la théorie de la perception directe. Il apparaît aussi que les auditeurs ne se représentent pas la géométrie de la barre sur la base d'information a priori sur ces sources. Les jugements perceptifs semblent plutôt associés à des processus de comparaison de l'in-

formation directement accessible, c'est-à-dire les hauteurs tonales propres à chaque son. Même si ces hauteurs tonales sont liées aux fréquences modales de flexion propres à une dimension de la barre, nous ne pouvons pas affirmer qu'il y ait une représentation d'une dimension de la source.

7.2.2 Représentation mentale des sources

Pour notre deuxième série d'expériences, nous avons poursuivi l'étude de Lakatos et coll. [LMC97] sur la représentation mentale des sources. Nous avons cherché à étudier spécifiquement la contribution des modes de flexion suivant l'épaisseur et la largeur et des modes de torsion dans la représentation auditive de la section géométrique d'une barre. Ces deux informations fréquentielles donnent une information indirecte pour les modes de flexion (comparaison des fréquences de flexion liées à l'épaisseur et à la largeur) ou directement pour les modes de torsion (les rapports de fréquences sont liés de façon monotone à la section géométrique).

Nous avons utilisé pour cela des sons de synthèse contenant l'information qui nous semblait pertinente. Nous avons utilisé conjointement l'analyse par modèles de résonance et la modélisation des barres par éléments finis. Ces analyses nous ont permis pour chaque son de séparer les contributions fréquentielles propres aux modes de flexion, des contributions fréquentiels propres aux modes de torsion.

Nous avons testé différentes séries de sons : les sons originaux (enregistrés) de barres, des sons (synthétisés) contenant les deux informations fréquentielles liées aux modes de flexion et de torsion et des sons contenant uniquement l'information liée aux modes de flexion ou aux modes de torsion. Il a été demandé, comme pour l'expérience originale [LMC97], d'apparier le bon ordre des représentations graphiques des sections d'une paire de barres à une séquence sonore produite par les sons de la paire de barres.

Les résultats de ces expériences montrent que la réplique de l'expérience avec les sons originaux donne des résultats semblables à l'étude originale, même si dans notre cas l'utilisation d'une information directement liée à l'épaisseur peut aussi expliquer ce résultat. Cette information est la hauteur liée à la fréquence fondamentale du mode de flexion suivant l'épaisseur. L'utilisation d'une information liée à l'épaisseur se retrouve dans les résultats des expériences lorsque les sons contiennent uniquement l'information fréquentielle liée aux modes de flexion et dans une moindre mesure pour la condition « flexion + torsion ». Par contre la condition avec les sons de synthèse contenant l'information fréquentielle liée aux modes de torsion, donc liée au rapport épaisseur/largeur, permet également d'expliquer la discrimination des sections des barres.

Pour toutes les conditions, mise à part la condition « flexion + torsion », l'information fréquentielle est associée à un paramètre de résonance (centre de gravité temporel ou décroissance temporelle) qui semble ajouter une information pertinente. Ces paramètres de résonance se retrouvent dans d'autres études utilisant des sons percussifs [Rou99] [PMH00].

Les commentaires des participants paraissent confirmer nos résultats en indiquant l'utilisation systématique de la hauteur pour discriminer auditivement la section des barres. De plus il est apparu que le paradigme expérimental comportait un biais et permettait dans 88 % des cas de bien discriminer les barres uniquement sur la base de la hauteur tonale de chaque son. Ainsi certains participants associent le changement d'épaisseur à une direction de changement de la hauteur et d'autres à l'autre direction.

Cette seconde série d'expériences tend à montrer que les auditeurs mettent en place des stratégies perceptives afin d'associer un son avec sa représentation visuelle plutôt que de percevoir « directement » la géométrie par le son. Il ne semble pas exister d'association directe entre les sons et la perception de leurs formes comme pour notre première série d'expériences.

7.3 Réflexions

Les résultats de nos expériences montrent qu'il n'existe pas une perception directe des propriétés des sources dans ces différentes tâches. Les auditeurs élaborent plutôt des stratégies perceptives et/ou des réponses afin de comparer les dimensions des sources sonores plutôt que de saisir les invariants perceptifs spécifiant les sources. Ces résultats vont à l'encontre d'autres études montrant une capacité à juger auditivement les dimensions d'objets par exemple la longueur d'une baguette [Gav88] [CAKP98] ou les dimensions et formes de plaques [KPT00].

L'approche écologique tend à étudier les invariants acoustiques auxquels nous sommes sensibles et qui spécifieraient les propriétés physiques des sources sonores. Des études [CAKP98] [KPT00] proposant une approche écologique radicale, cherchent à expliquer parfois leurs résultats directement par les propriétés mécaniques sans chercher forcément un corrélat acoustique. Comme nous l'avons fait, il nous semble important de faire le lien entre les différents niveaux de description de l'événement sonore (mécanique, acoustique, perceptif) afin d'expliquer comment le son peut spécifier les propriétés de la source. D'autres études ont montré avec succès l'importance de paramètres physiques dans la perception du matériau d'une source sonore [Rou99] [KPK00]. Il semble que pour la perception des dimensions d'une source nous y soyons moins sensible.

Certains luthiers pour sélectionner le bois de fabrication de l'instrument de musique, frappent sur les troncs d'arbre et écoute le son. Dans l'écoute du son, ils semblent projeter la future sonorité de l'instrument, sa qualité, de la même façon que le mécanicien écoute le bruit du moteur pour connaître l'origine de la défaillance. Nous voyons que la relation entre les propriétés mécaniques de la source et la perception de celle-ci reste encore à explorer.

Annexe **A**

Mécanique



A.1 Théorie des barres

Nous allons décrire le comportement vibratoire d'une barre de section constante, de matériau isotrope comme le métal dans nos expériences. Pour accéder aux fréquences propres des différents modes de vibration de la barre nous devons prendre en compte les conditions aux limites. Dans notre cas nous considérons une barre « libre-libre », c'est-à-dire que nous supposons que cette barre ne possède pas de support de fixation. Nous faisons ici une synthèse des travaux, et des textes de [Dou96] ainsi que [KFC82], [Mor76], [GR93] et [FR91].

Nous pouvons considérer quatre différents types de modes vibratoires possibles : le longitudinal (compression/détentes suivant la longueur de la barre), de torsion (les sections de la barre tournent en opposition), et de flexion suivant l'épaisseur ou la largeur.

A.1.1 Vibrations longitudinales

Nous exposons ici un modèle unidimensionnel pour une barre dont les dimensions sont petites devant la longueur d'onde et dont les surfaces sont rigides. Nous pouvons considérer que le mouvement particulaire se fait suivant l'axe de vibration. De plus, sur une même section nous considérons en première approximation que l'ensemble des particules subit la même contrainte, ce qui implique que celle-ci soit appliquée uniformément sur toute la section.

Le déplacement longitudinal $\xi(x, t)$ d'une barre de section constante de densité linéique de masse ρ est relié au champ de contrainte $\sigma(x, t)$ par la loi de Hooke :

$$\sigma(x, t) = \frac{N(x, t)}{S} = E \frac{\partial \xi}{\partial x}(x, t); \quad (\text{A.0})$$

où E est le module de Young et $N(x, t)$ est l'effort axial dans la barre (cf. fig : A.1).

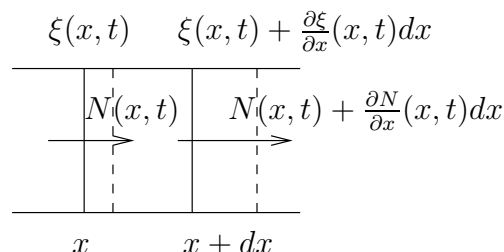


Figure A.1 - Élément différentiel d'une barre soumise à des contraintes axiales.

Si $N(x, t)$ représente l'effort axial à l'abscisse x , alors $N(x, t) + (\partial N(x, t) / \partial x) dx$

représente cette force à $x + dx$. Le bilan des forces sur une portion infinitésimale de longueur dx nous donne :

$$-N(x, t) + \left(N(x, t) + \frac{\partial N}{\partial x}(x, t)dx \right) = \frac{\partial N}{\partial x}(x, t)dx = \rho S dx \frac{\partial^2 \xi}{\partial t^2}(x, t); \quad (\text{A.0})$$

L'équation du mouvement de barre en vibration longitudinale s'écrit sous la forme :

$$\frac{\partial^2 \xi}{\partial t^2}(x, t) = c_L^2 \frac{\partial^2 \xi}{\partial x^2}(x, t) . \quad (\text{A.0})$$

avec $c_L = \sqrt{E/\rho}$ la vitesse de phase également non dispersive pour ces modes.

A.1.2 Vibrations de torsion

Le mode de torsion s'établit dans une barre lorsque celle-ci est soumise à des moments par rapport à son axe longitudinal. On repère le mouvement de rotation des sections transverses $S(x)$ par l'angle $\theta(x, t)$. Un élément dS de la section situé à une distance r de l'axe Ox se déplace d'un angle $\alpha = r\partial\theta/\partial x$.

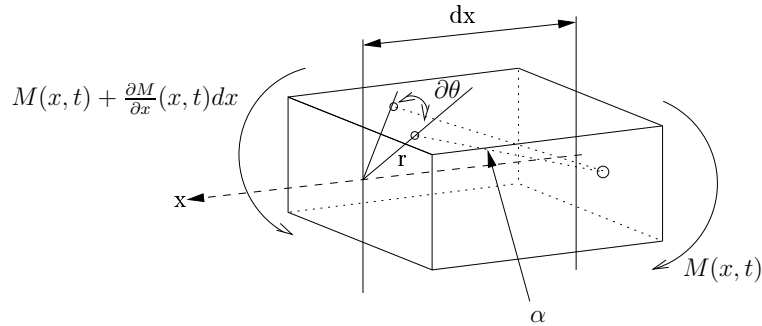


Figure A.2 - Élément différentiel d'une barre soumise à des moments de torsion

Ceci se traduit par une force de cisaillement $f(r, x, t)$ appliquée à dS que l'on associe à une contrainte de cisaillement telle que :

$$\sigma(r, x, t) = G\alpha(r, x, t) = Gr \frac{\partial \theta}{\partial x}(x, t) \quad (\text{A.0})$$

où G est le module de cisaillement. On obtient alors le moment de la force $f(r, x, t)$ par rapport à l'axe Ox , pour obtenir

$$dM(r, x, t) = r f(r, x, t) = Gr^2 \frac{\partial \theta}{\partial x}(x, t) dS; \quad (\text{A.0})$$

soit encore pour la section totale

$$M(x, t) = G \int r^2 \frac{\partial \theta}{\partial x}(x, t) dS = GK_T(x) \frac{\partial \theta}{\partial x}; \quad (\text{A.0})$$

où on a introduit $K_T(x)$, la raideur de torsion.

L'équation des ondes de torsion peut donc se construire comme suit :

$$[M(x, t) + \frac{\partial M}{\partial x}(x, t) dx] - M(x, t) = \frac{\partial M}{\partial x}(x, t) dx = \rho I(x) dx \frac{\partial^2 \theta}{\partial t^2}(x, t). \quad (\text{A.0})$$

Soit encore :

$$\rho I(x) \frac{\partial^2 \theta}{\partial t^2}(x, t) = G \frac{\partial}{\partial x} \left(K_T(x) \frac{\partial \theta}{\partial x}(x, t) \right). \quad (\text{A.0})$$

D'où l'équation des ondes de torsion pour une barre de section constante :

$$\frac{\partial^2 \theta}{\partial t^2}(x, t) + c_T^2 \frac{\partial^2 \theta}{\partial x^2}(x, t) \text{ avec } c_T = \sqrt{\frac{K_T G}{I \rho}}; \quad (\text{A.0})$$

La vitesse de phase est aussi non dispersive.

A.1.3 Vibrations de flexion

Nous introduirons ici le modèle d'Euler-Bernoulli. La flexion d'un élément dx de barre induit une tension et compression respectivement au dessus et en dessous d'un axe (appelé axe neutre) où les effets se compensent.

Dans un premier temps, il est considéré que les conséquences de la rotation des sections sont négligeables devant les déformations transversales. Les effets dûs au cisaillement ne sont pas pris en compte non plus. On suppose que les sections restent planes de la même manière que l'on considèrerait des contraintes uniformes pour les ondes longitudinales. Ceci se traduit suivant l'axe transversal de la barre par des forces de cisaillement, plus précisément appelées effort tranchant $f(x, t)$, auxquelles sont associés leurs moments par rapport au centre de l'élément différentiel. Elles sont le reflet des contraintes suivant l'axe longitudinal de la barre.

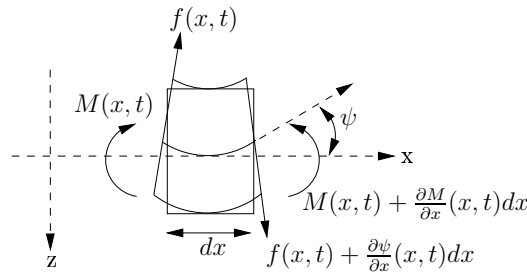


Figure A.3 - Élément différentiel d'une barre sous l'effet d'efforts tranchants

En effet, un petit élément de section dS situé à une distance r de l'axe neutre à un déplacement longitudinal $u = r\psi$ et, d'après la loi de Hooke, est soumis à une force $df = -EdSr\partial\psi/\partial x$, dont le moment associé est $dM = rdf$. Le moment total qui correspond à ces forces sur la section s'écrit de deux manières :

$$M(x, t) = -E \frac{\partial\psi}{\partial x} \int r^2 dS = -EI(x) \frac{\partial\psi}{\partial x} = -S(x) K_G^2(x) \frac{\partial\psi}{\partial x}; \quad (\text{A.0})$$

selon que l'on considère le moment géométrique (ou d'inertie) $I(x)$ ou le rapport de giration $K_G^2(x)$, définis. Pour une section rectangulaire, le moment géométrique est défini comme :

$$I(x) = \int r^2 dS = \frac{S^2(x)}{12} \left(\frac{h(x)}{b(x)} + \frac{b(x)}{h(x)} \right) \text{ et } K_G^2 = \frac{1}{S(x)} \int r^2 dS; \quad (\text{A.0})$$

$h(x)$ et $b(x)$ étant respectivement la hauteur et la largeur de la barre. Le plan principal de flexion qui nous intéresse ici est le plan (xOz) . Le moment d'inertie qui lui correspond est :

$$I(x) = \frac{l(x)h^3(x)}{12}. \quad (\text{A.0})$$

Pour la flexion dans le plan (xOy) , on aurait alors :

$$I(x) = \frac{h(x)l^3(x)}{12}. \quad (\text{A.0})$$

D'autre part, dans l'hypothèse des petits déplacements ($\psi \ll 1$), on a $\psi = \partial w/\partial x$, où on note $w(x, t)$ le déplacement transversal de la barre. Ainsi, il vient :

$$M(x, t) = -EI(x) \frac{\partial\psi}{\partial x}(x, t) = -EI(x) \frac{\partial^2 w}{\partial x^2}(x, t). \quad (\text{A.0})$$

Le bilan des forces sur la portion de barre (voir figure A.3) nous donne :

$$-f(x, t) + \left(f(x, t) + \frac{\partial f}{\partial x}(x, t) dx \right) = \frac{\partial f}{\partial x}(x, t) dx = \rho S dx \frac{\partial^2 w}{\partial t^2}(x, t). \quad (\text{A.0})$$

De même, le bilan des moments conduit à avoir :

$$M(x, t) - \left(M(x, t) + \frac{\partial M}{\partial x}(x, t) dx \right) + f(x, t) \frac{dx}{2} + \left(f(x, t) + \frac{\partial f}{\partial x}(x, t) dx \right) \frac{dx}{2} = 0. \quad (\text{A.0})$$

Soit au premier ordre en dx :

$$f(x, t) = \frac{\partial M}{\partial x}(x, t). \quad (\text{A.0})$$

A.2. RÉOLUTION DES ÉQUATIONS DE PROPAGATION

La forme générale de l'équation de propagation de flexion dans le cadre du modèle d'Euler-Bernoulli s'écrit alors :

$$\rho S(x) \frac{\partial^2 w}{\partial t^2}(x, t) = - \frac{\partial^2}{\partial x^2} \left[EI(x) \frac{\partial^2 w}{\partial x^2}(x, t) \right]. \quad (\text{A.0})$$

Pour une section constante, cette équation se réduit à :

$$\frac{\partial^2 w}{\partial t^2}(x, t) = - \frac{EI}{\rho S} \frac{\partial^4 w}{\partial x^4}(x, t). \quad (\text{A.0})$$

La relation de dispersion associée à (A.1.3) conduit à une vitesse de phase dispersive. Elle s'écrit :

$$\omega = \sqrt{\frac{EI}{\rho S}} k^2 = k c_F. \quad (\text{A.0})$$

A.2 Résolution des équations de propagation

Nous nous plaçons dans les cas où les conditions limites sont considérées comme « libre-libre ». Nous supposons de même que le mouvement est harmonique pour la résolution des équations (A.1.1), (A.1.2) et (A.1.3).

A.2.1 Vibrations longitudinales

En supposant que le mouvement est harmonique, soit $\xi(x, t) = U(x)e^{j\omega t}$, l'équation (A.1.1) devient :

$$U''(x) + \left(\frac{\omega}{C_L} \right)^2 U(x) = U''(x) + k^2 U(x) = 0; \quad (\text{A.0})$$

dont la solution générale s'écrit immédiatement comme :

$$U(x) = A_U \cos(kx) + B_U \sin(kx). \quad (\text{A.0})$$

L'absence de contrainte $\sigma(x, t)$ se traduit aux extrémités par la relation :

$$\int_S \sigma(x, t) dS = SE \frac{\partial \xi}{\partial x}(x, t) = O|_{x=0, L}; \quad (\text{A.0})$$

où S est la section de la barre. L'application des conditions aux limites (A.2.1) autorise alors l'écriture des fonctions et fréquences propres qui sont associées aux ondes longitudinales. Nous avons :

$$U_n(x) = A_U \cos(kx); \quad k_n = \frac{n\pi}{L}; \quad f_n = \frac{nc_L}{2L} \text{ pour } n \in |N^*. \quad (\text{A.0})$$

Ce type de vibration est donc harmonique.

A.2.2 Vibrations de torsion

Nous pouvons remarquer que les modes de vibration de torsion obéissent à une équation similaire à celle de la vibration longitudinale.

Nous pouvons écrire directement l'expression des fonctions et fréquences propres comme :

$$\Theta_n(x) = A_\Theta \cos(k_n x); k_n = \frac{n\pi}{L}; f_n = \frac{nc_T}{2L} \text{ pour } n \in \mathbb{N}^* . \quad (\text{A.0})$$

Nous pouvons noter que la vitesse de phase c_T dans le cas d'une section rectangulaire, vaut pour un rapport largeur sur épaisseur de 2 :

$$c_T = 0.74 \sqrt{\frac{G}{\rho}}; \quad (\text{A.0})$$

et si la largeur est 6 fois supérieure à l'épaisseur, nous avons :

$$c_T = \frac{\text{larg.}}{\text{ép.}} \sqrt{\frac{G}{\rho}} . \quad (\text{A.0})$$

Ces résultats nous seront utiles par la suite. Ce type de vibration est lui aussi harmonique.

A.2.3 Vibrations de flexion

Les conditions aux limites se caractérisent par une absence de moment et de force aux extrémités de la barre, qui s'expriment alors comme suit :

$$\frac{\partial^2 w}{\partial x^2}(x, t) = \frac{\partial^3 w}{\partial x^3}(x, t) = 0|_{x=0,L} . \quad (\text{A.0})$$

Pour un mouvement harmonique ($w(x, t) = W(x)e^{j\omega t}$), l'équation (A.1.3) devient :

$$W^{IV}(x) - k^4(x) = 0; \quad (\text{A.0})$$

dont la solution générale s'écrit :

$$W(x) = A_W \cosh(kx) + B_W \sinh(kx) + C_W \cos(kx) + D_W \sin(kx) . \quad (\text{A.0})$$

L'application des conditions aux limites conduit à l'équation :

$$2 - 2 \cosh(kl) \cos(kl) = 0 . \quad (\text{A.0})$$

Le calcul des zéros de l'équation (A.2.3), par résolution numérique ou graphique, donne la série des fréquences propres comme :

$$f_n = \frac{\pi K_G}{8L^2} \sqrt{\frac{E}{\rho}} [3.011^2, 5^2, 7^2, \dots, (2n+1)^2] . \quad (\text{A.0})$$

Notons que la série n'est pas harmonique.

Annexe **B**

Consignes

B.1 Classification : « non orientée »

- z Vous allez participer à une expérience de psychoacoustique.
- z Le principe de cette expérience est de regrouper les sons qui vous paraissent aller perceptivement ensemble, en autant de groupes que vous le désirez. Chaque groupe peut contenir de 1 à plusieurs sons. Chaque son doit appartenir à un groupe.
- z À l'écran apparaissent 30 boîtes "son" que vous pouvez écouter autant de fois que vous le souhaitez.
- z Pour écouter un son, il suffit de cliquer sur une de boites "son" à l'aide d'un des boutons de la souris.
- z Pour constituer un groupe, en déplaçant les boîtes "son", il faut appuyer sur la touche «**Command**» (en vert) ainsi que «**E**», puis cliquer sur la souris et sans relâcher, déplacer librement la boîte. Lorsque vous avez terminé l'opération, réappuyer sur les mêmes touches.
- z Conseil : au début écoutez tous les sons puis regroupez les sons qui vous paraissent aller tout de suite ensemble. Réécoutez l'ensemble des sons, puis affinez votre stratégie !
- z Bon courage !
- z En cas de problème n'hésitez pas à demander de l'aide.
- z Il vous sera demandé par la suite d'expliciter vos motivations ainsi que les indices que vous avez utilisés pour regrouper ces différents sons.

B.2 Classification : « orientée »

- z Vous allez participer à une expérience de psychoacoustique. Les sources sonores que vous écouterez sont des sons obtenus en frappant des barres de métal.
- z Le principe de cette expérience est de regrouper les sons qui proviennent de la même barre. Chaque groupe peut contenir de 1 à plusieurs sons. Chaque son doit appartenir à un groupe.
- z À l'écran apparaissent 30 boîtes "son" que vous pouvez écouter autant de fois que vous le souhaitez.
- z Pour écouter un son, il suffit de cliquer sur une de boîtes "son" à l'aide d'un des boutons de la souris.
- z Pour constituer un groupe, en déplaçant les boîtes "son", il faut appuyer sur la touche «**Command**» (en vert) ainsi que «**E**», puis cliquer sur la souris et sans relâcher, déplacer librement la boîte. Lorsque vous avez terminé l'opération, réappuyer sur les mêmes touches.
- z Conseil : au début écoutez tous les sons puis regroupez les sons qui vous paraissent aller tout de suite ensemble. Réécoutez l'ensemble des sons puis affinez votre stratégie!
- z Bon courage!
- z En cas de problème n'hésitez pas à demander de l'aide.
- z Il vous sera demandé par la suite d'expliciter vos motivations ainsi que les indices que vous avez utilisés pour regrouper ces différents sons.

B.3 Classification : « contrainte »

Les sources sonores que vous aller écouter sont des sons obtenus en frappant des barres de métal. Elles sont de même longueur avec des sections variables.

- z Le principe de cette expérience est de regrouper les sons qui proviennent d'une même barre. L'ensemble des sons provient de 6 barres, chaque barre étant frappée à 5 endroits distincts. (Voir exemple sonore). Chaque groupe des 6 que vous devez faire, correspond à une barre avec ses 5 positions.
- z À l'écran apparaissent 30 boîtes "son" (6 barres x 5 positions) que vous pouvez écouter autant de fois que vous le souhaitez.
- z Pour écouter un son il suffit de cliquer sur une de boîtes "son" à l'aide d'un des boutons de la souris.
- z Pour constituer un groupe, en déplaçant les boîtes "son", il faut appuyer sur la touche «**Command**» (en vert) ainsi que «**E**», puis cliquer sur la

souris et sans relâcher, déplacer librement la boîte. Lorsque vous avez terminé l'opération, réappuyer sur les mêmes touches.

- z Bon courage!
- z Il vous sera demandé par la suite de décrire les indices que vous avez utilisés pour regrouper ces différents sons.

B.4 Appariement

Première partie de l'expérience

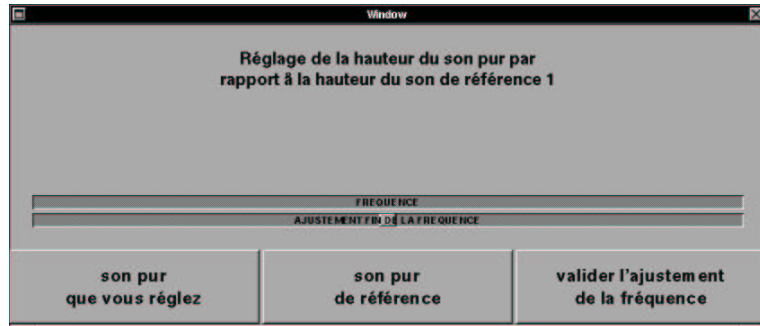
Principe général

Vous allez apparier en fréquence deux sons purs. Vous réglerez la fréquence d'un son entendu à l'oreille gauche pour que sa hauteur vous semble identique à celle d'un son pur de référence entendu à l'oreille droite. Cette première partie va nous permettre de mesurer la diplacousie, c'est-à-dire la différence de hauteur perçue entre une écoute monaurale (avec une seule oreille) et binaurale (avec les deux oreilles).

Méthode

Vous pouvez voir ci-dessous l'interface que vous allez utiliser dans la première partie de l'expérience. Il y a deux curseurs que vous pouvez bouger dans les deux sens. Le premier curseur (partie supérieure) permet de régler le son pur en fréquence, le curseur inférieur permet un réglage plus fin.

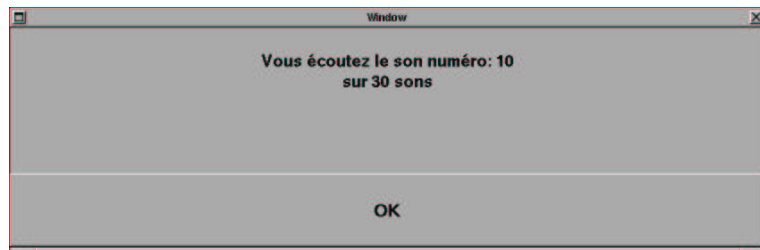
A chaque nouveau son pur de référence, le curseur supérieur "fréquence" est invisible dans le cadre. A partir du moment où vous cliquez dans le cadre, le curseur apparaît. A chaque fois que vous déplacez le curseur et que vous relâchez le bouton de la souris, le son pur de référence ainsi que le son pur réglé sont joués. Un bouton permet d'écouter seulement le son pur que vous êtes en train de régler. Le deuxième bouton permet d'écouter seulement le son de référence. Et le troisième valide le réglage effectué et passe au son de référence suivant.



Cette partie comporte 5 sons purs à régler.

Deuxième partie de l'expérience

Quand la première partie sera terminée, une fenêtre de transition vous invite, en cliquant sur « OK » à passer à la deuxième partie de l'expérience. Vous allez entendre maintenant 30 sons de barres de métal pour vous familiariser avec ces nouveaux sons.



Principe général

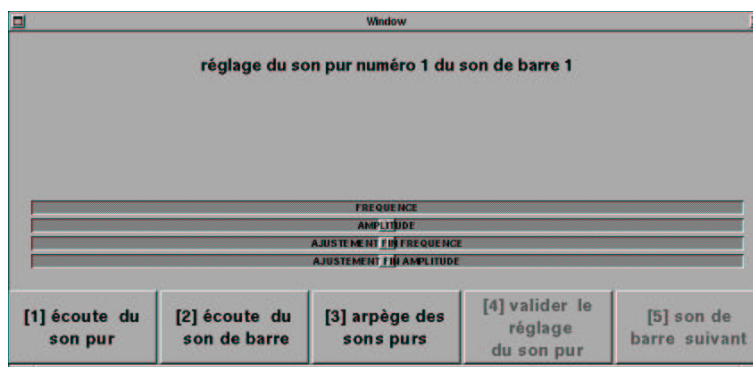
Dans cette partie nous vous demandons d'écouter attentivement ces sons. Pour chaque son de barre nous vous demandons d'écouter les différentes "notes" que vous entendez dans ce son. Pour chaque note "entendue" vous pouvez régler un son pur en fréquence ainsi qu'en amplitude (qui correspond à l'intensité perçue fort-faible) pour qu'il vous paraisse identique à cette note. Comme précédemment vous entendrez le son de référence : le son de la barre, dans l'oreille droite et le son pur que vous réglez dans l'oreille gauche.

Vous pouvez pour un même son de barre régler plusieurs sons purs si vous entendez différentes "notes" contenues dans le son (une sorte d'accord).

Méthode

Comme pour la première partie de l'expérience vous retrouvez une interface similaire. Nous avons deux types de curseur : un lié au réglage en fréquence du son pur et l'autre lié à son réglage en amplitude. A chaque fois il y a deux niveaux de réglage : un réglage "grossier" et un réglage fin. A chaque fois que vous déplacez le curseur et que vous relâchez le bouton de la souris, le son de la barre ainsi que le son pur réglé sont joués.

Un bouton permet d'écouter le son pur seul que vous réglez. Un autre permet d'écouter le son de barre uniquement. Vous pouvez quand vous avez réglé au moins deux sons purs écouter un arpège de ces sons purs. Les deux autres permettent de valider le son pur et de passer au son suivant.



Lorsque vous passez au son suivant, vous pouvez vous retrouver dans plusieurs cas. Pour chaque cas, la procédure à suivre est expliquée.

Bon courage !

Vous pouvez faire une pause pendant l'expérience. Vous pouvez la faire à tout moment puisque que c'est vous qui faites avancer l'expérience. En cas de problème vous pouvez me joindre au poste 15 51. En cas de mauvaise manipulation, si un écran autre que ceux présentés ci-dessus apparaît surtout ne faites rien et appelez moi !

B.5 Expérience « représentation »

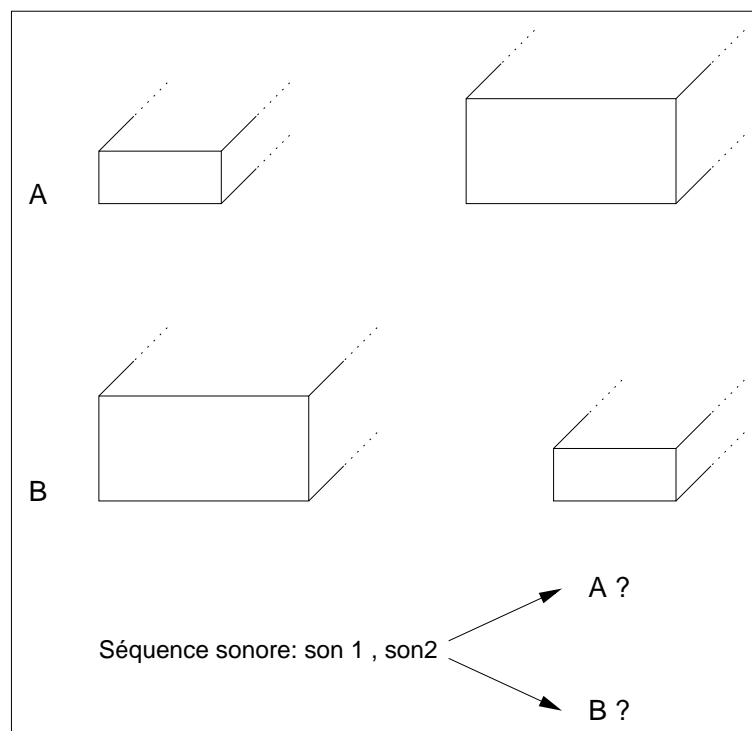
Cette expérience comprend plusieurs essais. A chaque essai, vous allez entendre une séquence de deux sons de barres de métal. Les barres ont la

même longueur mais ont des épaisseurs et des largeurs différentes. C'est-à-dire des sections différentes.

En même temps que la séquence sonore, vous allez voir apparaître deux paires de rectangles sur l'écran de l'ordinateur. Chaque paire est composée des deux mêmes rectangles dans un ordre différent (A ou B). Ces deux rectangles représentent les sections des deux barres avec lesquelles les sons de la séquence ont été produits. Chaque dessin de section est associé à un son de barre et tous les dessins sont représentés à la même échelle.

Vous allez donc déterminer à chaque fois quel ordre visuel, A ou B, correspond aux formes géométriques des sources entendues dans la séquence sonore, une des deux paires visuelles A ou B correspond toujours à la bonne réponse.

Pour effectuer cette tâche, vous allez utiliser n'importe quelle information perçue dans chacun des deux sons. Vous pouvez réécouter à chaque fois la séquence sonore.



Exemple d'essai

Annexe **C**

Expériences de classification



C.1 Concordance entre individus

C.1.1 Classification non-orientée

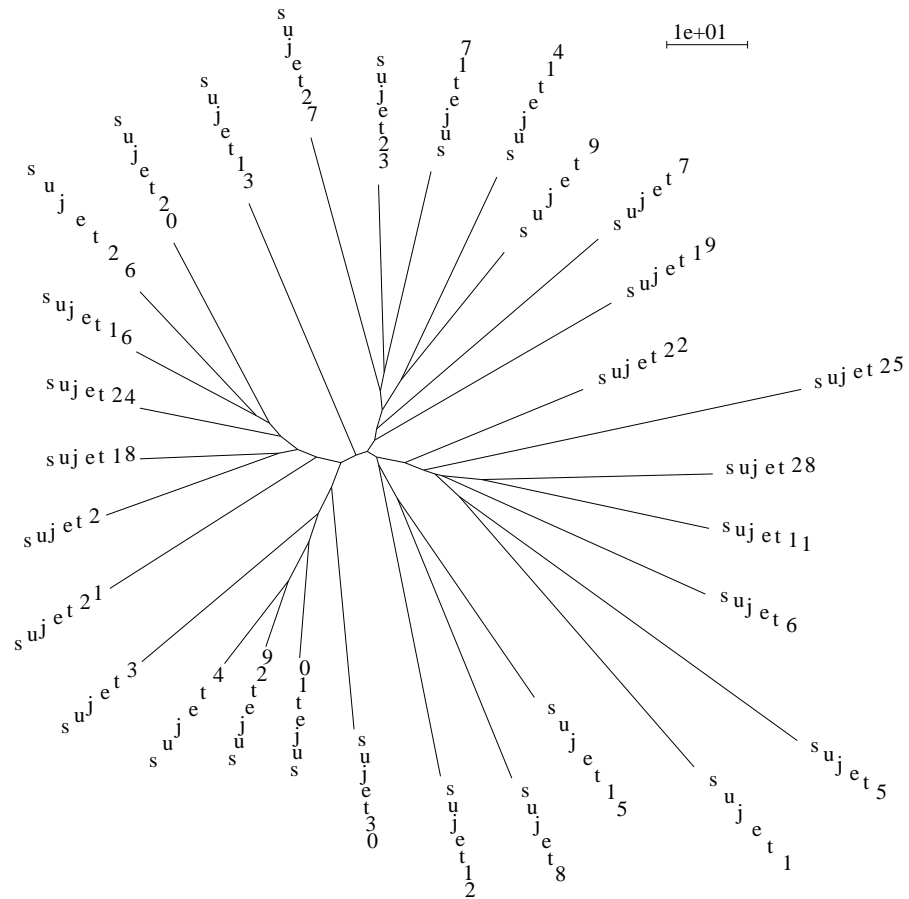


Figure C.1 - Arbre additif sur la matrice de concordance entre individus pour la classification non-orientée.

C.1.2 Classification orientée

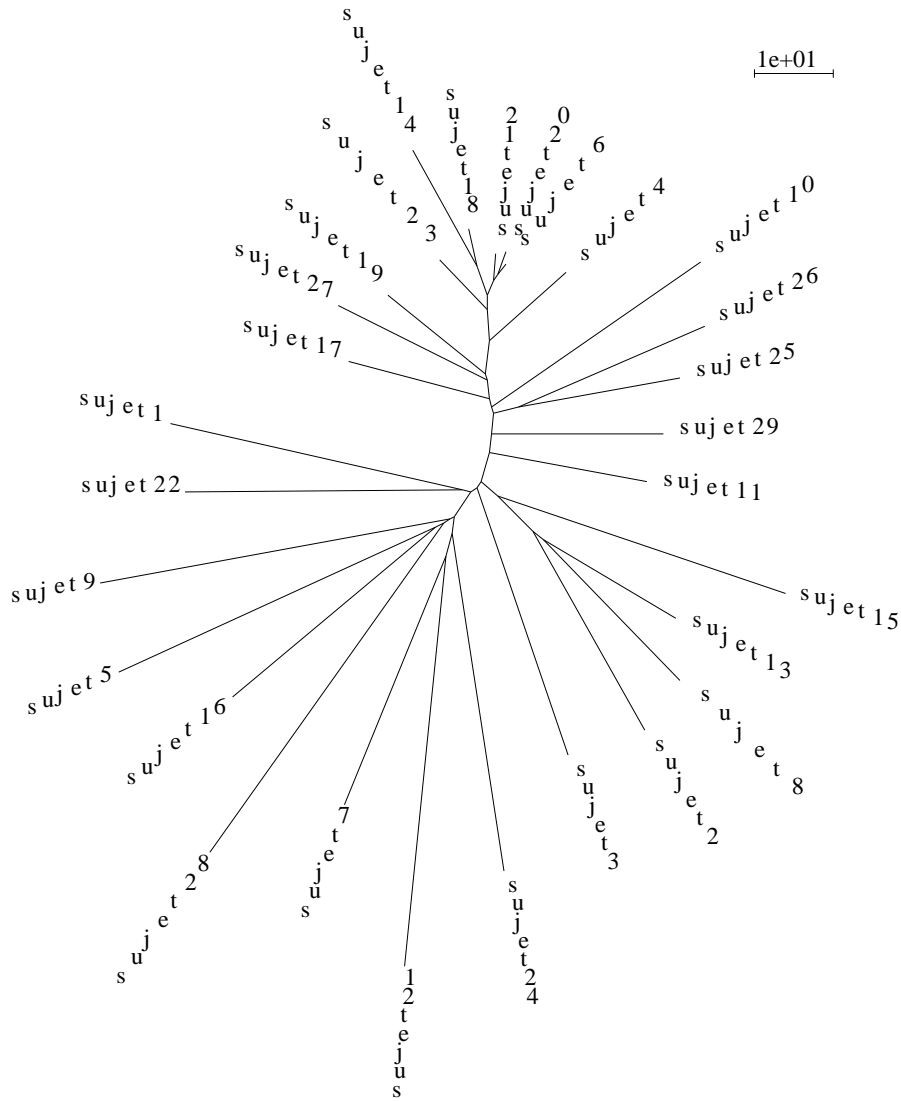


Figure C.2 - Arbre additif sur la matrice de concordance entre individus pour la classification orientée.

C.1.3 Classification contrainte

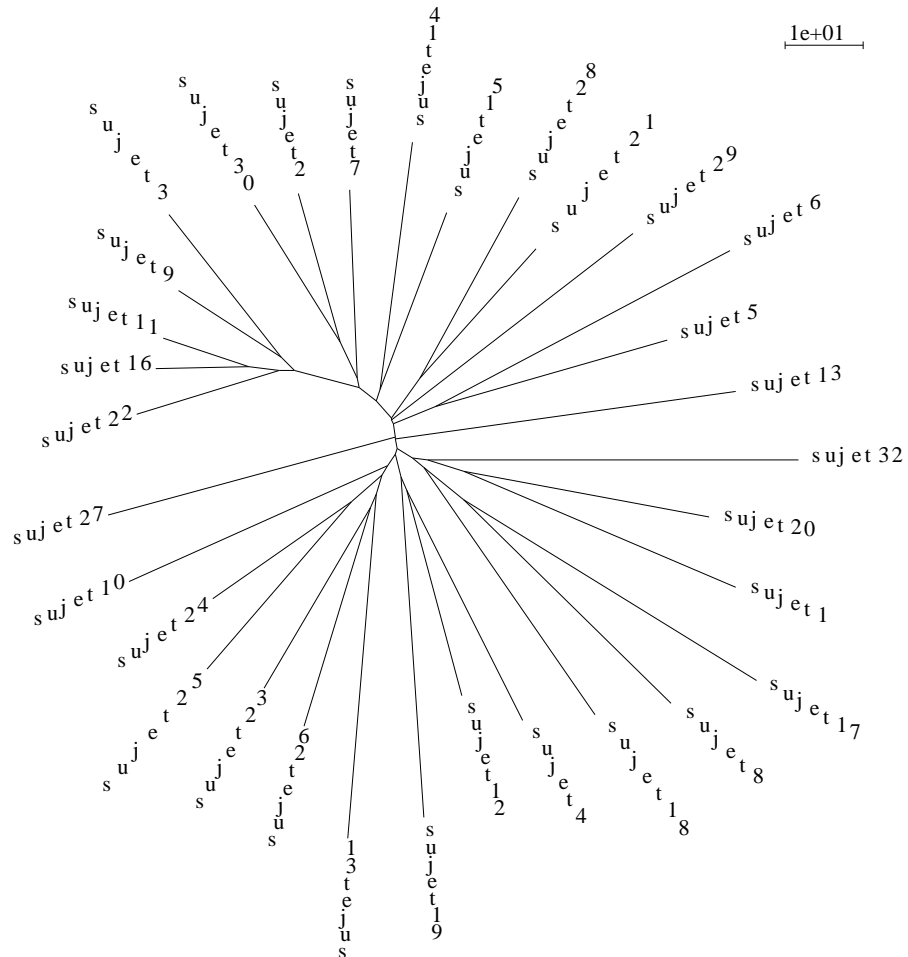


Figure C.3 - Arbre additif sur la matrice de concordance entre individus pour la classification contrainte.

C.2 Arbres additifs

C.2.1 Classification non-orientée

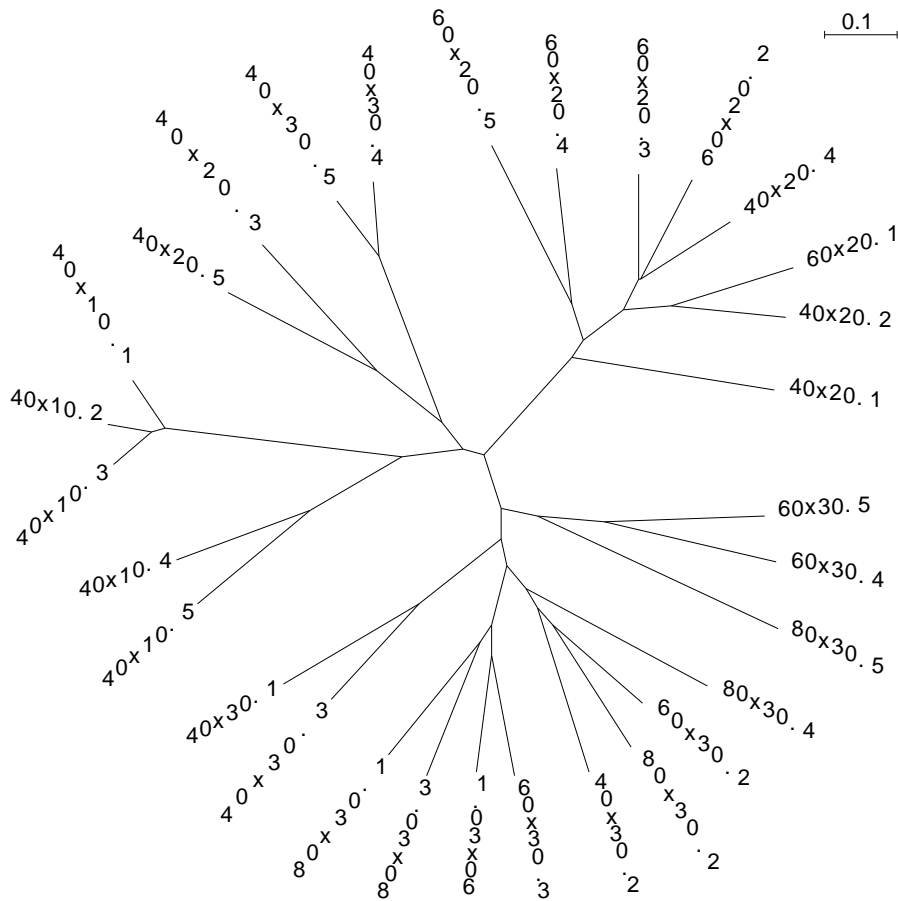


Figure C.4 - Arbre additif des résultats de la classification non-orientée.

C.2.2 Classification orientée

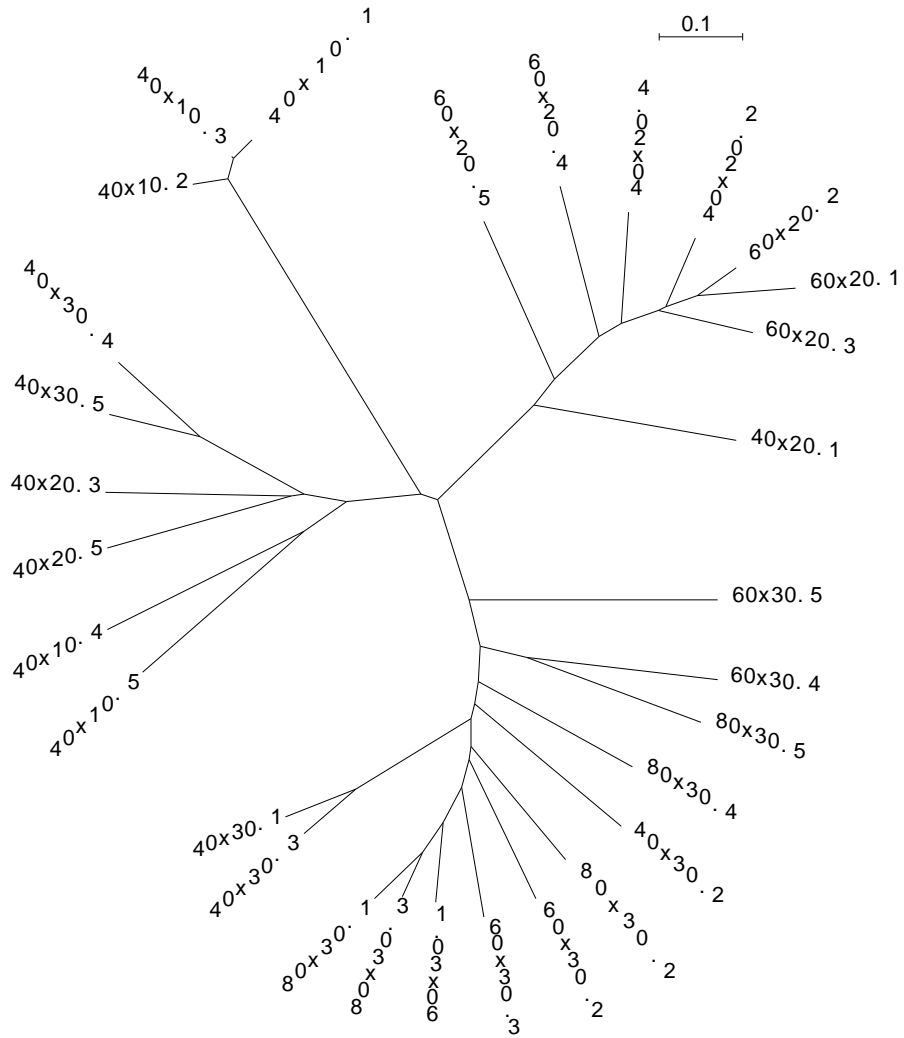


Figure C.5 - Arbre additif des résultats de la classification orientée.

C.2.3 Classification contrainte

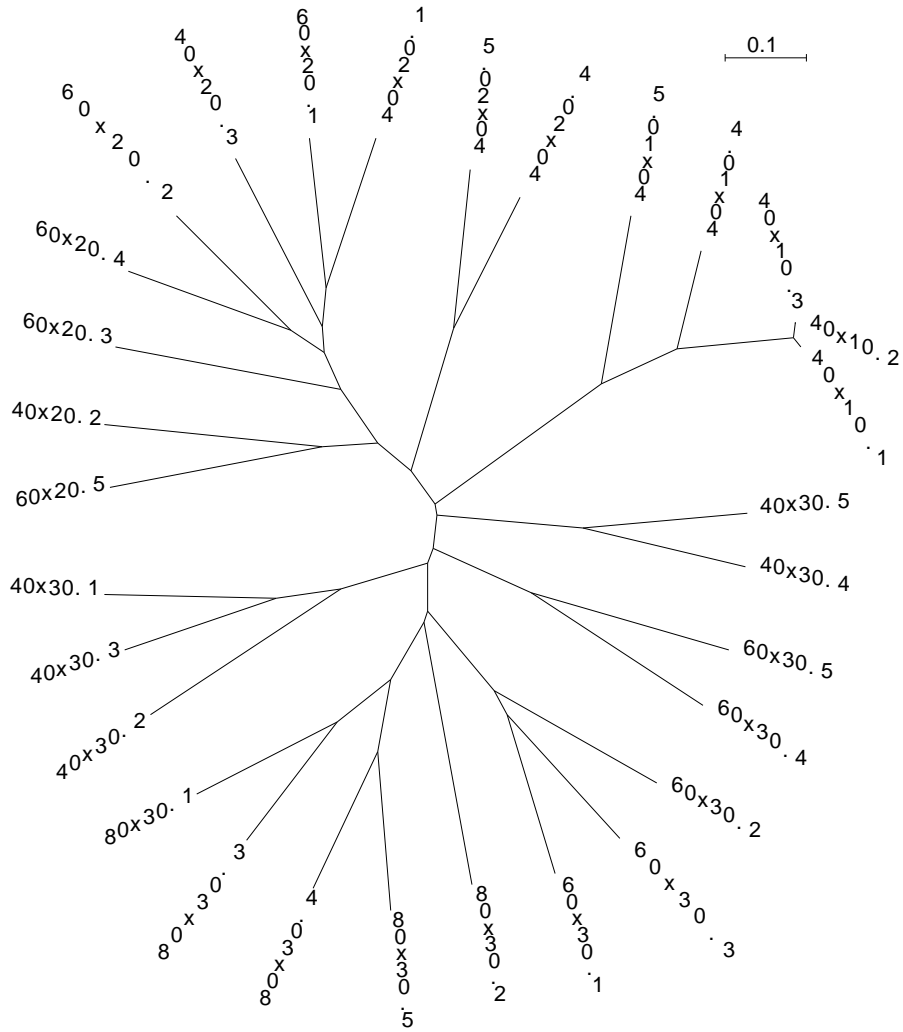


Figure C.6 - Arbre additif des résultats de la classification contrainte.

C.3 Analyses M.D.S

C.3.1 Classification non-orientée

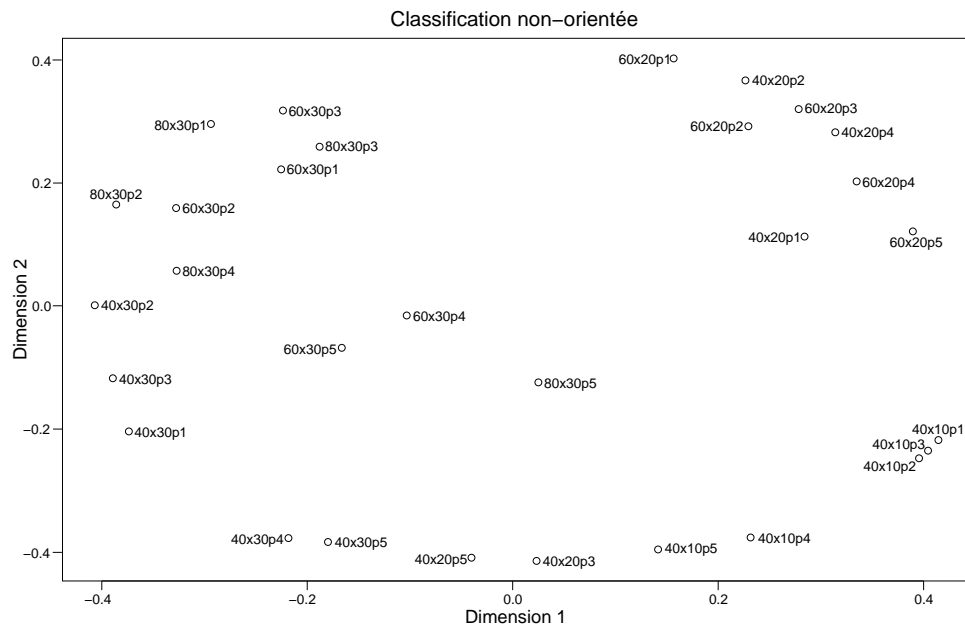


Figure C.7 - Analyse M.D.S (2 dimensions) des résultats de la classification non-orientée. Rotation procrustéenne par rapport à l'espace condition « orientée ».

C.3.2 Classification orientée

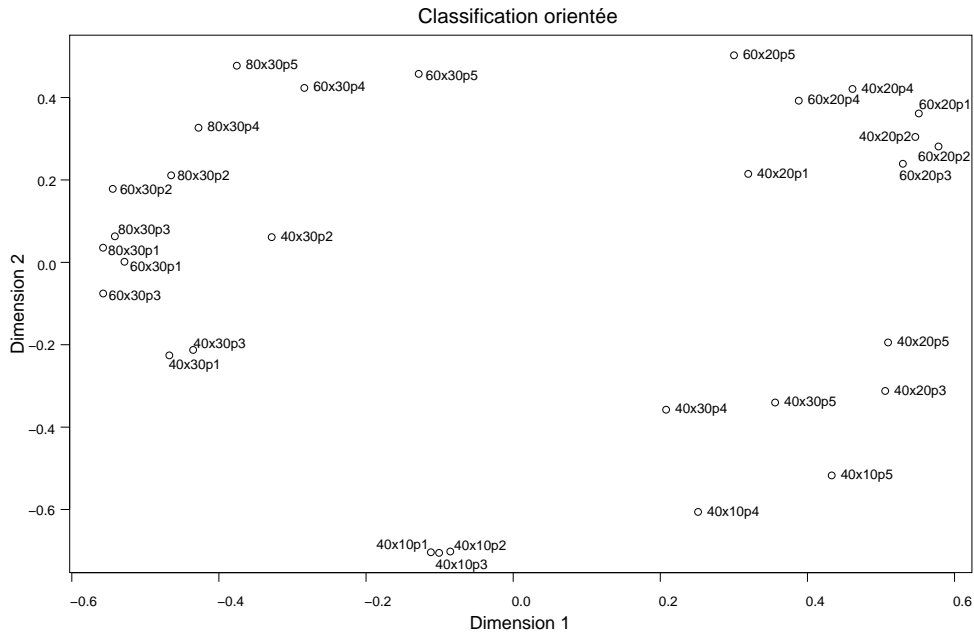


Figure C.8 - Analyse M.D.S (2 dimensions) des résultats de la classification orientée.

C.3.3 Classification contrainte

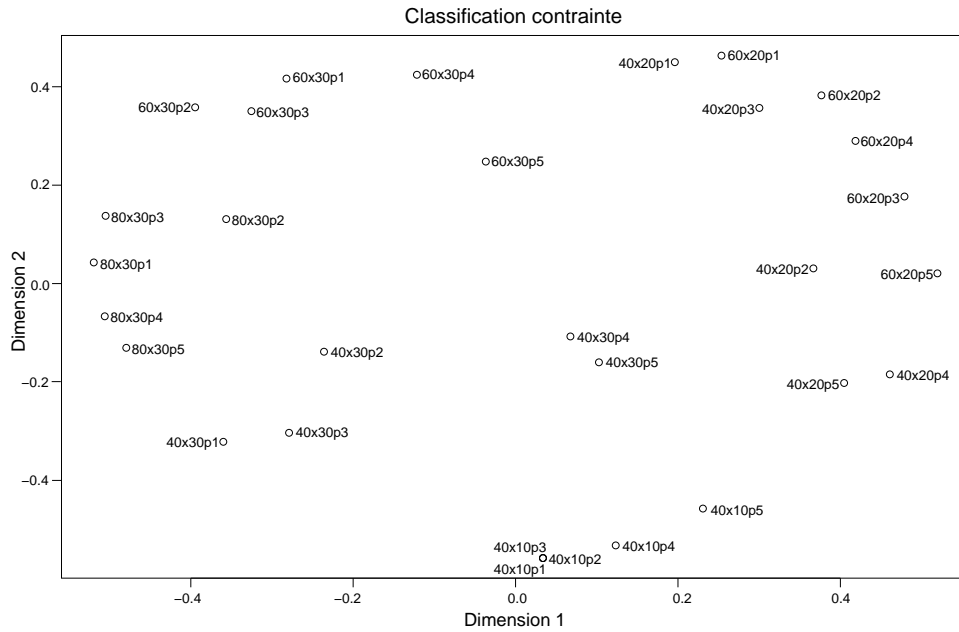


Figure C.9 - Analyse M.D.S (2 dimensions) des résultats de la classification contrainte. Rotation procrustéenne par rapport à l'espace condition « orientée ».

C.4 Analyses acoustiques

Valeurs théoriques

dimension (mm)	fréq. du 1 ^{er} mode de flexion (Hz)	rapport W./T.	fréq. du 1 ^{er} mode de torsion (Hz)
80	4370.8	4	2372.9
60	3278.0	3	2933.4
40	2185.4	8/3	3161.2
30	1639.0	2	3986.1
20	1092.7	4/3	4547.7
10	546.3		

Analyses acoustiques (méthode paramétrique)

dimension (mm)	fréq. du 1 ^{er} mode de flexion (Hz)	rapport W./T.	fréq. du 1 ^{er} mode de torsion (Hz)
80	4382.6	4	2371.8
60	3087.0/3074.3/3046.5	3	∅
40	2180.9/2193.7/2181.6	8/3	3289.2
30	1671.2/1674.6/1676.8	2	3959.4/3964.5
20	1139.6/1131.8	4/3	4717.6
10	571.1		

C.5 Modèle appariement

C.5.1 Modèle de contraste

C.5.1.1 Programme C du modèle de contraste

```
/*
  Calcul d'integration des spectres avec comparaison pas a pas
  des integrales des 2 spectres: A inter B, B - A, A - B (voir modele
  de contraste)
  pour l'integrer dans R -> compiler: R CMD SHLIB *.c -> *.so (shared librarie)
*/
#include<stdio.h>
/* variables d'entree
   [rowmat= nbre d'echantillons dans le spectre] [delta =pas reel d'integration]
   [Fe = freq d'echantillonnage] [ainterb amoinsb bmoinsa=voir emodle]
   [freqi= freq spectre i] [freqj= éfrq spectre j]
*/
void integration( int *rowmat , double *delta , double *Fe ,
double *freqi , double *freqj , double *ainterb , double *amoinb , double *bmoinsa){
  int k , ipas;
  ipas = *delta / *Fe;
  for(k = 0 ; k <= (*rowmat - ipas) ; k += ipas){
    /* cas particulier pour k = 0 */
    if(k == 0){
      /* atot[k] = 0.5 * *delta * (freqi[k] + freqi[ipas]);
      btot[k] = 0.5 * *delta * (freqj[k] + freqj[ipas]); */
      /* premiere condition */
      if((freqi[k] <= freqj[k] && freqi[ipas] <= freqj[ipas])
      || (freqi[k] >= freqj[k] && freqi[ipas] >= freqj[ipas])){
        if(freqi[k] <= freqj[k]){
          ainterb[k] = 0.5 * *delta * (freqi[k] + freqi[ipas]);
          amoinb[k] = 0;
          bmoinsa[k] = 0.5 * *delta * (freqj[k] + freqj[ipas]) - ainterb[k];
        }
        else{
          ainterb[k] = 0.5 * *delta * (freqj[k] + freqj[ipas]);
          amoinb[k] = 0.5 * *delta * (freqi[k] + freqi[ipas]) - ainterb[k];
          bmoinsa[k] = 0;
        }
      }
      /* deuxieme condition */
      if((freqi[k] < freqj[k] && freqi[ipas] > freqj[ipas])
      || (freqi[k] > freqj[k] && freqi[ipas] < freqj[ipas])){
        if(freqi[k] < freqj[k]){
          ainterb[k] = 0.5 * *delta * (freqi[k] + freqj[ipas]);
          amoinb[k] = 0.5 * *delta * (freqi[k] + freqi[ipas]) - ainterb[k];
          bmoinsa[k] = 0.5 * *delta * (freqj[k] + freqj[ipas]) - ainterb[k];
        }
        else{
          ainterb[k] = 0.5 * *delta * (freqj[k] + freqi[ipas]);
          amoinb[k] = 0.5 * *delta * (freqi[k] + freqi[ipas]) - ainterb[k];
          bmoinsa[k] = 0.5 * *delta * (freqj[k] + freqj[ipas]) - ainterb[k];
        }
      }
    }
  }
  /* pour k > 0 */
  else{
    /* atot[k] = 0.5 * *delta * (freqi[k] + freqi[k + ipas]);
    btot[k] = 0.5 * *delta * (freqj[k] + freqj[k + ipas]); */
  }
}
```

C.5. MODÈLE APPARIEMENT

```
/* meme chose que k = 0 sauf ici 0 -> k */
/* premiere condition */
if(((freqi[k] <= freqj[k] && freqi[k + ipas] <= freqj[k + ipas])
|| (freqi[k] >= freqj[k] && freqi[k + ipas] >= freqj[k + ipas]))){
    if(freqi[k] <= freqj[k]){
        ainterb[k] = 0.5 * *delta * (freqi[k] + freqi[k + ipas]);
        amoinsb[k] = 0;
        bmoinsa[k] = 0.5 * *delta * (freqj[k] + freqj[k + ipas]) - ainterb[k];
    }
    else{
        ainterb[k] = 0.5 * *delta * (freqj[k] + freqj[k + ipas]);
        amoinsb[k] = 0.5 * *delta * (freqi[k] + freqi[k + ipas]) - ainterb[k];
        bmoinsa[k] = 0;
    }
}
}
/* deuxieme condition */
if(((freqi[k] < freqj[k] && freqi[k + ipas] > freqj[k + ipas])
|| (freqi[k] > freqj[k] && freqi[k + ipas] < freqj[k + ipas]))){
    if(freqi[k] < freqj[k]){
        ainterb[k] = 0.5 * *delta * (freqi[k] + freqj[k + ipas]);
        amoinsb[k] = 0.5 * *delta * (freqi[k] + freqi[k + ipas]) - ainterb[k];
        bmoinsa[k] = 0.5 * *delta * (freqj[k] + freqj[k + ipas]) - ainterb[k];
    }
    else{
        ainterb[k] = 0.5 * *delta * (freqj[k] + freqi[k + ipas]);
        amoinsb[k] = 0.5 * *delta * (freqi[k] + freqi[k + ipas]) - ainterb[k];
        bmoinsa[k] = 0.5 * *delta * (freqj[k] + freqj[k + ipas]) - ainterb[k];
    }
}
}
}
}
```

C.5.1.2 Résultat du modèle de contraste

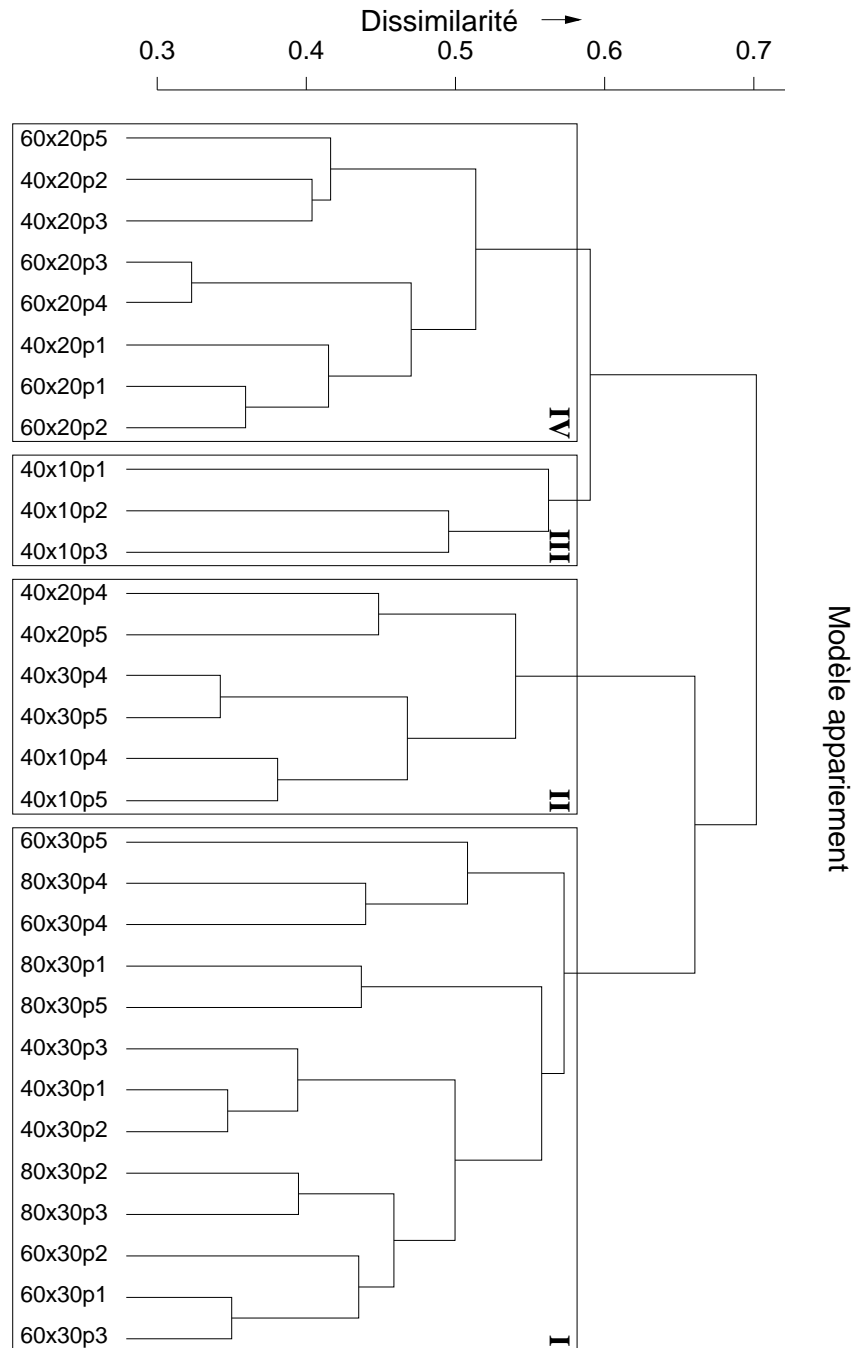


Figure C.10 - Analyse par groupements hiérarchiques moyens pour le modèle de Tversky dérivé de l'expérience d'appariement.

Annexe **D**

Représentation des sources sonores



D.2. CODE CASTEM POUR UNE BARRE

les informations sur la fréquence fondamentale (arbitraire dans notre cas pour des sons percussifs) et le nombre de canaux, ici un. Le son est donc monophonique.

FOF P Les 8 colonnes indiquent dans l'ordre : la fréquence, l'amplitude, l'amortissement (largeur de bande au maximum), la durée de l'excitation (l'attaque), le temps du début de l'atténuation, l'atténuation, la phase.

Voir la description des standards SDIF [EAS00].

D.2 Code CASTEM pour une barre

Script en langage GIBI qui permet de mailler et de calculer les modes propres et extraire les fréquences modales d'une barre de métal.

```
*-----*
*ANALYSE MODALE D'UNE BARRE VOLUMIQUE
*-----*

OPTI DIME 3 ELEM CU20;

* DONNEES
*-----
A1=0.3;
A2=0.08;
A3=0.04;
YOP=203.E9;
NUP= 0.29;
RHOP=7800;
VAL1=1;
VAL2=20000;
NF=15;
* MAILLAGE
*-----
P1= 0. 0. 0.;
P2= A2 0. 0.;
P3= 0. 0. A3;
P4= 0. A1 0.;
L1= DROITE 6 P1 P2;
S1= L1 TRAN 6 P3;
V1= S1 VOLU 30 'TRANS' P4;
* TRACE DU MAILLAGE
*-----
TRAC CACH V1;
*MODELE ET PROPRIETES DU MATERIAU
MOD1=MODEL V1 MECANIQUE ELASTIQUE CU20;
MAT1=MATER MOD1 YOUN YOP NU NUP RHO RHOP;
* MATRICES DE RIGIDITE ET DE MASSE
*-----
RIG1=RIGI MOD1 MAT1;
MAS1=MASS MOD1 MAT1;

*-----*
* ANALYSE MODALE
*-----*

MESS 'CALCUL DES 'NF 'PREMIERS MODES';
* CALCUL MODAL
```

```

*-----
TAB1=VIBR INTER VAL1 VAL2 'BASSE' NF RIG1 MASI 'IMPR' 'TBAS' ;
TAB2=TAB1.MODES;

* VALEURS DES FREQUENCES
*-----
I = 0;
REPETER BLOC1 NF;
I = I + 1;
MESS ' FREQUENCE ' I ' = '(TAB2.I).FREQUENCE' Hz';
FIN BLOC1;

* TRACE DES DEFORMEES MODALES
*-----
DEP11=(TAB2.1).DEFORMEE_MODAL;
DEFO=DEFO DEP11 V1 0. VERT;

MESS 'POUR VOIR LE IEME MODE';
MESS 'TAPEZ VOIR I;';
MESS 'POUR FINIR';
MESS 'TAPEZ FIN;';

DEBP VOIR I*ENTIER;
DEP1=(TAB2.I).DEFORMEE_MODAL;
DEF1=DEFO DEP1 V1 0.01 ROUGE;
TITRE 'DEFORMEES MODALES ' I ' A LA FREQUENCE '(TAB2.I).FREQUENCE' Hz';
MESS 'TAPEZ FIN DE TRACE';
MESS 'POUR VOIR L ANIMATION';
TRAC CACH ( DEF1 ET DEFO);
DEFO1 = ANIME 10 V1 DEP1 0.01;
TITRE 'ANIMATION DU MODE ' I ' A LA FREQUENCE '(TAB2.I).FREQUENCE' Hz';
MESS 'CLIQUEZ SUR ANIMATION';
MESS 'TAPEZ FIN DE TRACE';
TRAC CACH OSCIL DEFO1;
MESS 'POUR VOIR LE IEME MODE';
MESS 'TAPEZ VOIR I;';
MESS 'POUR FINIR';
MESS 'TAPEZ FIN;';
FINP;

*-----*

```


Bibliographie

- [Abd90] H. Abdi. Additive-tree representations. *Lecture Notes in Biomathematics*, 84 :43–59, 1990.
- [AHDS96] P. Arabie, L.J. Hubert, et G. De Soete, éditeurs. *Clustering and Classification*. World Scientific, 1^{re} édition, 1996.
- [AM93] F.G. Ashby et W.T. Maddox. Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37 :372–400, 1993.
- [And91] J.R. Anderson. The adaptative nature of human categorization. *Psychological Review*, 98(3) :409–429, 1991.
- [AP88] F.G. Ashby et N.A. Perrin. Toward a unified theory of similarity and recognition. *Psychological Review*, 95(1) :124–150, 1988.
- [ASA60] American Standards Association. Acoustical Terminology, S1.1-1960. New York, American Standards Association, 1960.
- [Ash92a] F.G. Ashby. Multidimensional models of categorization. Dans *Multidimensional Models of Perception and Cognition* [Ash92b], chapitre 16, pages 449–483.
- [Ash92b] F.G. Ashby, éditeur. *Multidimensional Models of Perception and Cognition*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1992.
- [Atr98] S. Atran. Scott folk biology and the anthropology of science : Cognitive universals and cultural particulars. *Behavioral & Brain Sciences*, 21(4) :547–569, 1998.
- [Bag99] J.-D. Bagot. *Information, sensation et perception*. Armand Colin, Paris, 2^e édition, 1999. 1^{re} édition 1996.
- [Bal93] J.A. Ballas. Common Factors in the Identification of an Assortment of Brief Everyday Sounds. *Journal of Experimental Psychology : Human Perception and Performance*, 19(2) :250–267, 1993.
- [Bar83] L.W. Barsalou. Ad hoc categories. *Memory & Cognition*, 11(3) :211–227, 1983.

-
- [Bar91] J.-P. Barthélemy. Similitude, arbres et typicalité. Dans D. Dubois, éditeur, *Sémantique et Cognition*, Collection Sciences du langage, pages 205–224. Editions du C.N.R.S., Paris, 1991.
- [Bar94] R. Barbaras. *La perception. Essai sur le sensible*. Optiques, Philosophie. Hatier, Paris, 1994.
- [Bar99] L.W. Barsalou. Perceptual symbol systems. *Behavioral & Brain Sciences*, 22 :577–660, 1999.
- [Ber70] G. Berkeley. *Trois Dialogues entre Hylas et Philonous*. Aubier, 1970. Traduction Leroy, A.
- [Ber73] C. Berge. *Graphes et hypergraphes*. Dunod Université, 2^e édition, 1973.
- [Ber78] B. Berlin. Ethnobiological classification. Dans E. Rosch et B.B. Lloyd, éditeurs, *Cognition and Categorization*, pages 9–26. Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.
- [BG88] J.-P. Barthélemy et A. Guénoche. *Les arbres et les représentations des proximités*. Masson, 1^{re} édition, 1988.
- [BG97] I Borg et P. Groenen. *Modern Multidimensional Scaling. Theory and Applications*. Springer-Verlag, New York, 1^{re} édition, 1997.
- [BGA56] J.S. Bruner, J. Goodnow, et G.A Austin. *A Study of Thinking*. Wiley, New York, 1956.
- [BH87] J.A. Ballas et J.H. Howard. Interpreting the language of environmental sounds. *Environment & Behavior*, 19(1) :91–114, January 1987.
- [Bou96] M. Boucart. *La reconnaissance des objets*. Presses Universitaires de Grenoble, 1996.
- [BR89] G. Bennett et X. Rodet. Synthesis of the Singing Voice. Dans M.V. Mathews et J.R. Pierce, éditeurs, *Current Directions in Computer Music Research*, pages 19–44, Cambridge, MA, 1989. MIT Press.
- [Bre90a] J.N. Breckenridge. Validating cluster analysis : Consistent replication and symmetry. Dans *Paper presented at the meeting of the Classification Society of North America*, Utah State University, UT, 1990.
- [Bre90b] A. Bregman. *Auditory scene analysis*. MIT Press, Cambridge, MA, 1990.
- [Bre94] A. Bregman. L’analyse des scènes auditives : l’audition dans des environnements complexes. Dans McAdams and Bigand [MB94], chapitre II, pages 11–39.

- [Bun71] P. Buneman. The recovery of trees from measures of dissimilarity. Dans F.R. Hodson, D.G. Kendall, et P. Tautu, éditeurs, *Mathematics in the Archaeological and Historical Sciences*. Edinburgh University Press, Edinburgh, 1971.
- [CAKP98] C. Carello, K.L. Anderson, et A.J. Kunkler-Peck. Perception of the length by sound. *Psychological Science*, 9(3) :211–214, 1998.
- [Cas98] M.A. Casey. *Auditory Group Theory : with Applications to Statistical Basis Methods for Structured Audio*. Thèse de doctorat, Massachusetts Institute of Technology Media Laboratory, February 1998. <http://xenia.media.mit.edu/~mkc/thesis/>.
- [CB94] F. Critchley et Fichet B. The partial order by inclusion of the principal classes of dissimilarity on a finite set, and some of their basic properties. Dans B. Van Cutsem, éditeur, *Classification and Dissimilarity Analysis*, Lecture Notes in Statistics, chapitre 2, pages 5–65. Springer-Verlag, 1994.
- [CC70] J.D. Carroll et J.J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, 35 :283–319, 1970.
- [CM96] R. Cheng et G.W. Milligan. Measuring the influence of individual data points in a cluster analysis. *Journal of Classification*, 13 :315–335, 1996.
- [Cor71] R.M. Cormack. A review of classification. *Journal of the Royal Statistical Society Series A*, pages 321–367, 1971.
- [Cor82] J.E. Corter. ADDTREE/P : A PASCAL program for fitting additive trees based on satch and tversky's ADDTREE algorithm. *Behavior Research Methods and Instrumentation*, 14(3) :353–354, 1982.
- [CP76] F. Cailliez et J.-P. Pages. *Introduction à l'analyse des données*. SMASH, 1^{re} édition, 1976.
- [Cro65] D.V. Cross. *Metric properties of multidimensional stimulus control*. Thèse de doctorat, University of Michigan, 1965.
- [CT86] J.E. Corter et A. Tversky. Extended similarity trees. *Psychometrika*, 51 :429–451, 1986.
- [CWTss] C. Carello, J.B. Wagman, et M.T. Turvey. Acoustic Specification of Object Properties. Dans J. Anderson et B. Anderson, éditeurs, *Moving Image Theory : Ecological Considerations*. Southern Illinois Press, in press.

-
- [Daw96] J.T. Daws. The analysis of free-sorting data : beyond pairwise cooccurrences. *Journal of Classification*, 13 :57–80, 1996.
- [DHS01] R.O. Duda, P.E. Hart, et D.G. Stork. *Pattern Classification*. John Wiley & Sons, 2^e édition, 2001.
- [DL00] S. Dolnicar et F. Leisch. Getting more out of binary data : Segmenting markets by bagged clustering. Working Paper 71, SFB “Adaptive Information Systems and Modeling in Economics and Management Science”, August 2000.
- [DO74] B.S. Duran et P.L. Odell. *Cluster Analysis, a Survey*. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 1974.
- [Don97] S. Donnadiou. *Représentation mentale du timbre des sons complexes et effets de contexte*. Thèse de doctorat, Université René Descartes - Paris V, 1997.
- [Dou96] V. Doutaut. *Etude expérimentale et simulations numériques d'instruments de percussion à clavier*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 1996.
- [DSC96] G. De Soete et J.D. Carroll. Tree and other network models for representing proximity data. Dans Arabie et al. [AHDS96], pages 157–197.
- [Dub97] D. Dubois, éditeur. *Catégorisation et Cognition : De la Perception au Discours*. Éditions Kimé, Paris, 1997.
- [EAS00] Ircam Equipe Analyse Synthèse. Standard SDIF Types. Document Ircam, Version 1.5, Août 2000. <http://www.ircam.fr/equipes/analyse-synthese/sdif/standard/types-main.html>.
- [EK98] M.A. Erickson et J.K. Kruschke. Rules and exemplars in category learning. *Journal of Experimental Psychology : General*, 127 :107–140, 1998.
- [EK00] M.W. Eysenck et M.T. Keane. *Cognitive Psychology. A Student's Handbook*. Psychology Press, 4^e édition, 2000.
- [Est94] W.K. Estes. *Classification and Cognition*. Oxford Psychology Series. Oxford University Press, 1^{re} édition, 1994.
- [ET93] B. Efron et R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1993.
- [Fau00] A. Faure. *Des sons aux mots : Comment parle-t-on du timbre musical ?* Thèse de doctorat, Ecoles des Hautes Etudes en Sciences Sociales, Paris, 2000.

- [Fow91] C.A. Fowler. Auditory perception is not special : We see the world, we feel the world, we hear the world. *J. Acoust. Soc. Am.*, 89(6) :2910–2915, 1991.
- [FR91] N.H. Fletcher et T.D. Rossing. *The Physics of Musical Instruments*. Springer-Verlag, New York, 1991.
- [Fre52] G. Frege. On sense and reference. Dans P. Geach et M. Black, éditeurs, *Translations from the philosophical writings of Gottlob Frege*. Basic Blackwell, Oxford, 1952.
- [Fre90] D. Freed. Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *J. Acoust. Soc. Am.*, 87(1) :311–322, 1990.
- [Gar74] W.R. Garner. *The Processing of Information and Structure*. The Experimental Psychology Series. Lawrence Erlbaum Associates, Hillsdale, NJ, 1974.
- [Gar78] W.R. Garner. Aspects of a stimulus : Features, dimensions, and configurations. Dans E. Rosch et B.B. Lloyd, éditeurs, *Cognition and Categorization*, pages 99–133. Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.
- [Gar87] H. Gardner. *The Mind's New Science*. BasicBooks, 2^e édition, 1987.
- [Gav88] W.W. Gaver. *Everyday Listening and Auditory Icons*. Thèse de doctorat, University of California, San Diego, 1988.
- [Gav93] W.W. Gaver. What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5 :1–29, 1993.
- [GB98] R.L. Goldstone et L.W. Barsalou. Reuniting perception and conception. *Cognition*, 65 :231–262, 1998.
- [GCF97] F. Guyot, M. Castellengo, et B. Fabre. Étude de la catégorisation d'un corpus de bruits domestiques. Dans Dubois [Dub97], chapitre 2^e, pages 41–58.
- [GG99] A. Guénoche et G. Garreta. Quelle confiance accorder à une représentation arborée? Dans *Actes des colloques de la Société Française de Classification*, Sept 1999.
- [Gib66] J.J. Gibson. *The Senses Considered as Perceptual Systems*. Houghton Mifflin Company, Boston, 1966.
- [Gib79] J.J. Gibson. *The ecological approach to visual perception*. Houghton Mifflin Company, Boston, 1979.

-
- [GK03] R.L. Goldstone et A Kersten. Concepts and Categories. Dans A.F. Healy et R.W. Proctor, éditeurs, *Comprehensive handbook of psychology, Volume Four : Experimental psychology*, pages 591–621. Wiley, New York, 2003.
- [GMG91] R.L. Goldstone, D.L. Medin, et D. Gentner. Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology*, 23 :222–262, 1991.
- [GMH97] R.L. Goldstone, D.L. Medin, et J. Halberstadt. Similarity in context. *Memory & Cognition*, 25(2) :237–255, 1997.
- [Gol94] R.L. Goldstone. The role of similarity in categorization : providing a groundwork. *Cognition*, 52 :125–157, 1994.
- [Gol95] R.L. Goldstone. Mainstream and avant-garde similarity. *Psychologica Belgica*, 35(2/3) :145–165, 1995.
- [Gol99] R.L. Goldstone. Similarity. Dans R.A. Wilson et F.C. Keil, éditeurs, *MIT encyclopedia of the cognitive science*, pages 763–765. MIT Press, Cambridge, MA, 1999.
- [Goo72] N. Goodman. Seven structures on similarity. Dans N. Goodman, éditeur, *Problems and Projects*, pages 437–447. Bobbs Merrill, New York, 1972.
- [Gor87] A.D. Gordon. A review of hierarchical classification. *Journal of the Royal Statistical Society Series A*, 150(part 2) :119–137, 1987.
- [Gor96] A.D. Gordon. Clustering and classification. Dans Arabie et al. [AHDS96].
- [GR93] M. Gérardin et D. Rixen. *Théorie des vibrations. Applications à la dynamique des structures*. Masson, 1993.
- [Gre77] J.M. Grey. Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, 61(5) :1270–1277, 1977.
- [GT82] I. Gati et A. Tversky. Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology : Human Perception and Performance*, 8(2) :325–340, 1982.
- [Gua02] J. Guarrigues. La méthode des éléments finis. Ecole Supérieure de Mécanique de Marseille, Janvier 2002. <http://esm2.imt-mrs.fr/gar/ef.html>.
- [Gui79] P. Guillaume. *La psychologie de la forme*. Flammarion, 1979.
- [Guy96] F. Guyot. *Etude de la perception sonore en termes de reconnaissance et d'appréciation qualitative : une approche par la catégorisation*. Thèse de doctorat, Université du Maine, Le Mans, 1996.

- [HA85] L. Hubert et P. Arabie. Comparing partitions. *Journal of Classification*, 2 :193–218, 1985.
- [Ham81] J.A. Hampton. An investigation of the nature of abstract concepts. *Memory and Cognition*, 9 :149–156, 1981.
- [Ham93] J.A. Hampton. Prototype models of concept representation. Dans I. van Mechelen, J.A. Hampton, R.S. Michalski, et P. Theuns, éditeurs, *Categories and concepts : Theoretical views and inductive data analysis*, London, 1993. Academic Press.
- [Ham95] J.A. Hampton. Testing the prototype theory of concepts. *Journal of Memory and Language*, 34 :686–708, 1995.
- [Hamon] J.A. Hampton. Similarity and categorization. City University London : <http://www.staff.city.ac.uk/hampton/papers.htm>, Prépublication.
- [Han89] S. Handel. Identification of Speakers, Instruments, and Environmental Events. Dans *Listening : An Introduction to the Perception of Auditory Events*, chapitre 8. M.I.T Press, Cambridge, Mass., 1989.
- [Har87] S. Harnad. *Categorical perception*. Cambridge University Press, Cambridge, 1987.
- [HKK⁺98] O. Houdé, D. Kayser, O. Koenig, J. Proust, et F. Rastier, éditeurs. *Vocabulaire des sciences cognitives*. Presses Universitaires de France, Paris, 1^{re} édition, 1998.
- [Hol72] E.W. Holman. The relation between hierarchical and euclidean models for psychological distances. *Psychometrika*, 37(4), 1972.
- [Hou98] H. Houdin. Méthode des éléments finis. Ecole Centrale de Nantes, 1998. <http://www.ec-nantes.fr/Fr/Enseignement/MecaMaterGC/Structures/EI2-TC/mef/mef.html>.
- [Ima77] S. Imai. Pattern similarity and cognitive transformations. *Acta Psychologica*, 41 :433–447, 1977.
- [IRC93] Documentation IRCAM. *Max*. IRCAM, 1993.
- [Izm00] O. Izmirli. Non-harmonic Sinusoidal Modeling Synthesis Using Short-Time High-resolution Parameter Analysis. Dans *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, December 7-9 2000.
- [Jam78] M. Jambu. *Classification automatique pour l'analyse des données (Tome 1)*. Dunod, Paris, 1978.
- [JMF99] A.K. Jain, M.N. Murty, et P.J. Flynn. Data clustering : a review. *ACM Computing Surveys*, 31(3) :264–321, 1999.

-
- [Joh67] S.C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3) :241–254, 1967.
- [Kac66] M. Kac. Can you hear the shape of a drum? *American Math. Monthly*, 73 :1–23, 1966.
- [KFC82] E.L. Kinstler, A.R. Frey, et A.B. Coppens. *Fundamentals of Acoustics*, chapitre 3. Wiley, 3^e édition, 1982.
- [KG99] A.M. Krieger et P.E. Green. A generalized rand-index method for consensus clustering of separate partitions of the same data base. *Journal of Classification*, 16 :63–89, 1999.
- [KMW94] J. Krimphoff, S. McAdams, et S. Winsberg. Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysiques. *Journal de Physique*, 4(c5) :625–628, 1994.
- [Kom92] L.K. Komatsu. Recent views of conceptual structure. *Psychological Bulletin*, 112(3) :500–526, 1992.
- [KPK00] L.R. Klatzky, D.K. Pai, et E.P. Krotkov. Perception of Material from Contact Sounds. *Presence*, 9(4) :399–410, 2000.
- [KPT00] A. J. Kunkler-Peck et M. T. Turvey. Hearing shape. *J. Experimental Psychology : Human Perception and Performance*, 26 :279–294, 2000.
- [KR90] L. Kaufman et P.J. Rousseeuw. *Findings Groups in Data. An Introduction to Cluster Analysis*. Wiley-Interscience, 1^{re} édition, 1990.
- [Kri93] J. Krimphoff. *Analyse acoustique et perception du timbre*. Mémoire de D.E.A d’Acoustique Appliquée, Université du Maine, 1993.
- [Kru64] J.B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1964.
- [Kru77] J. Kruskal. The relationship between multidimensional scaling and clustering. Dans J.V. Ryzin, éditeur, *Classification and clustering*, pages 17–44. Academic Press, 1977.
- [Kru78] C.L. Krumhansl. Concerning the applicability of geometric models to similarity data : The interrelationship between similarity and spatial density. *Journal of American Psychological Association*, pages 445–463, 1978.
- [Kru89] C.L. Krumhansl. Why is the musical timbre so hard to understand? Dans S. Nielzen et O. Olsson, éditeurs, *Structure and*

- perception of electroacoustic sound and music*, pages 43–53. Elsevier, Amsterdam, 1989.
- [Kus02] J. Kusuma. Parametric frequency estimation : ESPRIT and MUSIC. Tutorial MIT, 2002. <http://www.mit.edu/~kusuma/work.html>.
- [KVV01] M. Kubovy et D. Van Valkenburg. Auditory and visual objects. *Cognition*, 80 :97–126, 2001.
- [Lak87] G. Lakoff. *Women, Fire, and Dangerous Things*. Chicago University Press, Chicago, 1987.
- [Lak00] S. Lakatos. A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, 62 :1426–1439, 2000.
- [Lam97] C. Lambourg. *Modèle temporel pour la simulation numérique de plaques vibrantes - Application à la synthèse sonore*. Thèse de doctorat, Université du Maine, Le Mans, 1997.
- [Lar93] J. Laroche. The use of the matrix-pencil method for the spectrum analysis of musical signals. *J. Acoust. Soc. Am.*, 94(4) :1958–1965, 1993.
- [LBDH86] M. Lalanne, P. Berthier, et J. Der Hagopian. *Mécanique des vibrations linéaires*, chapitre « Calcul par éléments finis », pages 170–210. Masson, Paris, 2^e édition, 1986.
- [Lev83] S. Levy. A cross-cultural analysis of the structure and levels of attitudes towards acts of political protest. *Social Indicators Research*, 12 :281–309, 1983.
- [LL94] F.-J. Lapointe et P. Legendre. A classification of pure malt scotch whiskies. *Applied Statistics*, 43(1) :237–257, 1994.
- [LL98] P. Legendre et L. Legendre. *Numerical Ecology*. Amsterdam : Elsevier Science, 2^e édition, 1998.
- [LLP91] X. Li, R.J. Logan, et R.E. Pastore. Perception of acoustic source characteristics : Walking sounds. *J. Acoust. Soc. Am.*, 90(6) :3036–3049, 1991.
- [LMC97] S. Lakatos, S. McAdams, et R. Caussé. The representation of auditory source characteristics : Simple geometric form. *Perception and Psychophysics*, 59(8) :1180–1190, 1997.
- [LMP97] L. Lebart, A. Morineau, et M. Piron. *Statistique exploratoire multidimensionnelle*. Dunod, 2^e édition, 1997.
- [LN77] P.H. Lindsay et D.A. Norman. *Human Information Processing*. Academic Press, 2^e édition, 1977.

-
- [LO97] R.A. Lutfi et E. Oh. Auditory discrimination of material changes in a struck-clamped bar. *J. Acoust. Soc. Am.*, 102(6) :3647–3656, 1997.
- [Lop00] P. Lopez. Cours de graphes. Rapport technique, LAAS-CNRS, août 2000.
- [LT98] H. Lawson-Tancred. 7. ancient greek philosophy II : Aristotele. Dans A.C. Grayling, éditeur, *Philosophy 1. A guide through the subject*. Oxford University Press, 1998. New edition.
- [LW66] G.N. Lance et W.T. Williams. A generalized sorting strategy for computer classifications. *Nature*, 212 :218, 1966.
- [LW67] G.N. Lance et W.T. Williams. A general theory of classificatory sorting strategies : I. hierarchical systems. *Computer Journal*, 9 :373–380, 1967.
- [Maf99] V. Maffiolo. *De la caractérisation sémantique et acoustique de la qualité sonore de l’environnement urbain : structuration des représentations mentales et influence sur l’appréciation qualitative : application aux ambiances sonores de Paris*. Thèse de doctorat, Université du Maine, Le Mans, 1999.
- [Mar82] D. Marr. *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information*. Free Press, New York, 1982.
- [Mar99] F. Marti. Obtention de modèles de résonances. Documentation Analyse/Synthèse, version révisée d’après le document de P.-F. Baisnée de novembre 1990, Ircam, Paris, mai 1999. <http://www.ircam.fr/equipe/analyse-synthese/DOCUMENTATIONS/AnaModRes/anamod.html>.
- [MB94] S. McAdams et E. Bigand, éditeurs. *Penser les sons*. Presses Universitaires de France, Paris, 1994.
- [MC81] C.F. Michaels et C. Carello. *Direct Perception*. Prentice-Hall, Englewood Cliffs, NJ, 1981.
- [MC85] G.W. Milligan et M.C. Cooper. An examination of the procedures for determining the number of clusters in a data set. *Psychometrika*, 50 :159–179, 1985.
- [McA94a] S. McAdams. Audition : physiologie, perception et cognition. Dans M. Richelle, J. Requin, et M. Robert, éditeurs, *Traité de psychologie expérimentale*, pages 283–344. Presses Universitaires de France, Paris, 1994.

- [McA94b] S. McAdams. La reconnaissance de sources et d'événements sonores. Dans McAdams and Bigand [MB94], chapitre VI, pages 157–213.
- [Mer75] F.M. Mersenne. *Harmonie Universelle*. Editions du Centre Nationale de la Recherche Scientifique, Paris, 1975. Edition facsimilé, édition originale : 1636.
- [Meu01] J.G. Meunier. Trois types de représentations cognitives. Cahiers de Recherche N° 2001.1, Montréal, 2001. Laboratoire d'Analyse Cognitive de l'Information.
- [MGG93] D.L. Medin, R.L. Goldstone, et D. Gentner. Respects for similarity. *Psychological Review*, 100(2) :254–278, 1993.
- [Mil69] G.A. Miller. A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 6 :169–191, 1969.
- [Mil80] G.W. Milligan. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45 :325–342, 1980.
- [Mil81] G.W. Milligan. A monte carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika*, 46(2) :187–199, 1981.
- [Mil96] G.W. Milligan. Clustering validation : results and implications for applied analyses. Dans Arabie et al. [AHDS96], pages 341–375.
- [Moo97] B.C.J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, 4^e édition, 1997.
- [Mor76] P.M. Morse. *Vibrations of Sound*, chapitre 4. Americans Institute of Physics for the Acoustical Society of America, 1976.
- [MR81] C.B. Mervis et E. Rosch. Categorization of natural objects. *Ann. Rev. Psychol.*, 32 :89–115, 1981.
- [MS78] D.L. Medin et M.M. Schaffer. Context theory of classification learning. *Psychological Review*, 85 :207–238, 1978.
- [MS81] D.L. Medin et P.J. Schwanenflugel. Linear separability in classification learning. *Journal of Experimental Psychology : Human learning and Memory*, 7 :355–368, 1981.
- [MS95] G.L. Murphy et T.L. Spalding. Knowledge, similarity, and concept formation. *Psychologica Belgica*, 35(2/3) :127–144, 1995.
- [MWD⁺95] S. McAdams, S. Winsberg, S. Donnadiou, G. De Soete, et J. Krimphoff. Perceptual scaling of synthesized musical timbres :

-
- Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58 :177–192, 1995.
- [Nei76] U. Neisser. *Cognition and Reality*. W.H Freeman and Company, 1976.
- [NJ00] R.M. Nosofsky et M.K. Johansen. Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, 7(3) :375–402, 2000.
- [Nos86] R.M. Nosofsky. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology : General*, 115(1) :39–57, 1986.
- [Nos92] R.M. Nosofsky. Exemplar-based approach to relating categorization, identification, and recognition. Dans Ashby [Ash92b], chapitre 14, pages 363–393.
- [NPM94] R.M. Nosofsky, T.J. Palmeri, et S.C. McKinley. Rule-plus-exception model of classification learning. *Psychological Review*, 101 :53–79, 1994.
- [PAG95] R.D. Patterson, M.H. Allerhand, et C. Giguère. Time-domain modeling of peripheral auditory processing : A modular architecture and a software platform. *J. Acoust. Soc. Am.*, 98(4) :1890–1894, 1995.
- [Pat00] R.D. Patterson. Auditory images : How complex sounds are represented in the auditory system. *J. Acoust. Soc. Japan(E)*, 21(4) :183–190, 2000.
- [Per92] N.A. Perrin. Uniting identification, similarity and preference : General recognition theory. Dans Ashby [Ash92b], chapitre 6, pages 123–146.
- [PK68] M.I. Posner et S.W. Keele. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77 :352–363, 1968.
- [Pl070] R. Plomp. Timbre as a multidimensional attributes of complex tones. Dans R. Plomp et G.F. Smoorenburg, éditeurs, *Frequency analysis and periodicity detection in hearing*, pages 397–414. Sijthoff, Leiden, 1970.
- [PMH00] G. Peeters, S. McAdams, et P. Herrera. Instrument Sound Description in the Context of MPEG-7. Dans *Proceedings of the 2000 International Computer Music Conference*, pages 166–169, Berlin, Germany, August 2000.
- [PTVF92] W.H. Press, S.A. Teukolsky, W.T. Vetterling, et B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, MA, (USA), 2^e édition, 1992.

- [Ran71] W.M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66 :846–850, 1971.
- [RC93] L.J. Rips et A. Collins. Categories and resemblance. *Journal of Experimental Psychology : General*, 122(4) :468–486, 1993.
- [Rep87] B.H. Repp. The sound of two hands clapping : An exploratory study. *J. Acoust. Soc. Am.*, 81(4) :1100–1109, 1987.
- [Ric88] W. Richards. Sound interpretation. Dans W. Richards, éditeur, *Natural Computation*, pages 301–308. M.I.T Press, Mass., 1988.
- [Rip89] L.J. Rips. Similarity, typicality, and categorization. Dans S. Vosniadou et A. Ortony, éditeurs, *Similarity and analogical reasoning*, pages 21–59. Cambridge University Press, Cambridge : England, 1989.
- [RK75] S. Rosenberg et M.P. Kim. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10 :489–502, 1975.
- [RK89] R. Roy et T. Kailath. ESPRIT - estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acous. Speech Sig. Process.*, 37(7) :984–995, 1989.
- [RM75] E. Rosch et C.B. Mervis. Family resemblances : Studies in the internal structure of categories. *Cognitive Psychology*, 7 :573–605, 1975.
- [RMG⁺76] E. Rosch, C.B. Mervis, W.D. Gray, D.M. Johnson, et P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8 :382–439, 1976.
- [Rod97] X. Rodet. Musical Sound Signals Analysis/Synthesis : Sinusoidal+Residual and Elementary Waveform Models. Dans *Proceedings of the IEEE Time-Frequency and Time-Scale Workshop (TFTS'97)*, University of Warwick, Coventry, UK, 27th-29th August 1997.
- [Ros75] E. Rosch. Cognitive representations of semantic categories. *Journal of Experimental Psychology : General*, 104(3) :192–233, 1975.
- [Ros78] E. Rosch. Principles of categorization. Dans E. Rosch et B.B. Lloyd, éditeurs, *Cognition and Categorization*, chapitre 2, pages 27–48. Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.
- [Rou99] V. Roussarie. *Analyse perceptive de structures vibrantes simulées par modèle physique*. Thèse de doctorat, Université du Maine, Le Mans, 1999.

-
- [RPB84] X. Rodet, Y. Potard, et J.-P. Barrière. The CHANT Project : From the Synthesis of the Singing Voice to Synthesis in General. *Computer Music Journal*, 8(3) :15–31, 1984.
- [RPB85] X. Rodet, Y. Potard, et J.-P. Barrière. CHANT : de la synthèse de la voix chantée à la synthèse en général. Rapports de recherche 35/85, IRCAM, 1985. 21 p.
- [SA79] R.N. Shepard et P. Arabie. Additive clustering : Representation of similarities as combinaisons of discrete overlapping properties. *Psychological Review*, 86 :87–123, 1979.
- [Sch77] P. Schaeffer. *Traité des objets musicaux*. Editions du Seuil, Paris, nouvelle édition, 1977.
- [Sch78] G. Schwarz. Estimating the dimensions of a model. *Annals of Statistics*, 6 :461–464, 1978.
- [Sch90] J.-L. Schwartz. Perception de la parole : invariance et variabilité. Dans *Colloque de Physique C2, 1^{er} Congrès Français d’Acoustique*, volume 51, pages 461–470, Février 1990. Supplément au n°2.
- [She62a] R.N Shepard. Analysis of proximities : Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27 :125–140, 1962.
- [She62b] R.N Shepard. Analysis of proximities : Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27 :219–246, 1962.
- [She81] R.N. Shepard. Psychophysical complementarity. Dans M. Kubovy et J.R. Pomerantz, éditeurs, *Perceptual Organization*, chapitre 10, pages 279–341. Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.
- [She87] R.N. Shepard. Toward a universal law of generalization for psychological science. *Science*, 237 :1317–1323, 1987.
- [Sin85] R.N. Sinnot, J.M. and Aslin. Frequency and intensity discrimination in human infants and adults. *J. Acoust. Soc. Am.*, 78(6) :1986–1992, 1985.
- [SM81] E.E. Smith et D.L. Medin. *Categories and Concepts*. Harvard University Press, 1981.
- [Smi95] B. Smith. PsiExp : An environment for Psychoacoustics experimentation using IRCAM musical workstation. Dans *Musical Perception and Cognition Conference 95*, Berkeley, USA, 1995.

- [SMM97] J.D. Smith, M.J. Murray, et J.P. Minda. Straight talk about linear separability. *Journal of Experimental Psychology : Learning, Memory, & Cognition*, 23 :659–680, 1997.
- [SPB00] E. Sander, S. Poitrenaud, et A. Bastide. Catégorisation inclusive versus catégorisation exclusive. Séminaire de l'équipe "Processus inférentiels et modélisation", Université Paris 8, Novembre 2000.
- [Sri90] N. Sriram. Clique optimization : a method to construct parsimonious ultrametric trees from similarity data. *Journal of Classification*, 7 :33–52, 1990.
- [SRY81] S.S. Schiffman, M.L. Reynolds, et F.W. Young. *Introduction to Multidimensional Scaling. Theory, Methods and Applications*. Academic Press, Inc., 1981.
- [SS93] N. Sriram et L. Scott. Constructing optimal ultrametrics. *Journal of Classification*, 10 :241–268, 1993.
- [SS96] R. Saltstone et K. Stange. A computer program to calculate hubert and arabie's adjusted rand index. *Journal of Classification*, 13 :169–172, 1996.
- [SSG97] J. Spencer-Smith et R.L. Goldstone. The dynamics of similarity. *Bulletin of the Japanese Cognitive Science Society*, 4 :38–56, 1997.
- [ST77] S. Sattah et A. Tversky. Additive similarity trees. *Psychometrika*, 42(3) :319–345, 1977.
- [Str35] J.R. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18 :643–662, 1935.
- [TG78] A. Tversky et I. Gati. Studies of similarity. Dans E. Rosch et B.B. Lloyd, éditeurs, *Cognition and Categorization*, chapitre 4, pages 79–98. Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.
- [TG82] A. Tversky et I. Gati. Similarity, separability, and the triangle inequality. *Psychological Review*, 89(2) :123–154, 1982.
- [Tij01] C. Tijus. *Introduction à la psychologie cognitive*. Nathan, 2001.
- [Tul72] E. Tulving. Episodic and semantic memory. Dans E. Tulving et W. Donaldson, éditeurs, *Organisation of Memory*. Academic Press, London, 1972.
- [Tve77] A. Tversky. Features of similarity. *Psychological Review*, 84 :327–352, 1977.
- [UJR98] A. Unzicker, M. Jüttner, et I. Rentschler. Similarity-based models of human visual recognition. *Vision Res.*, 38 :2289–2305, 1998.

-
- [Ull80] S. Ullman. Against direct perception. *Behavioral & Brain Sciences*, 3 :373–415, 1980.
- [Urd98] I. Urdapilleta. *Catégorisation et implication des propriétés*. Thèse de doctorat, Université de Paris VIII - Saint Denis, 1998.
- [VCA88] P. Verpeaux, T. Charras, et Millard A. CASTEM 2000 : Une approche moderne du calcul des structures. Dans *Calcul des structures et intelligence artificielle*. Editions Pluralis, 1988.
- [VD79] N.J. Van Derveer. Ecological acoustics : Human perception of environmental sounds. Distertation Abstracts International, 40/09B, 4543. (University Microfilms No. 8004002), 1979.
- [vH89] H. von Helmholtz. *Optique physiologique. Tome II / Hermann von Helmholtz*, volume 2 de *Fac-sim. de l'éd. de Paris : V. Masson et fils, 1867. Trad. par Emile Javal et N. Th. Klein*. BNF : numérisation Gabay, J., Paris, 1989.
- [VMS95] I. Van Mechelen et G. Storms. Analysis of similarity data and Tversky's contrast model. *Psychologica Belgica*, 35(2/3) :85–102, 1995.
- [War63] J.H. Jr. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58 :236–244, 1963.
- [WC89a] S. Winsberg et J.D. Carroll. A quasi-nonmetric method for multidimensional scaling of multiway data via a restricted case of INDSCAL model. Dans R. Coppi et S. Bolasco, éditeurs, *Multiway data analysis*, pages 405–414. North-Holland, Amsterdam, 1989.
- [WC89b] S. Winsberg et J.D. Carroll. A quasi-nonmetric method for multidimensional scaling via an Euclidean model. *Psychometrika*, 54 :217–229, 1989.
- [WDS93] S. Winsberg et G. De Soete. A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika*, 58(2) :315–330, 1993.
- [Wes79] D.L. Wessel. Timbre space as a musical control structure. *Computer Music Journal*, 3(2) :45–52, 1979.
- [WFR⁺98] M. Wright, A. Freed, X. Rodet, D. Virolle, et R. Woehrmann. New Applications of the Sound Description Interchange Format. Dans *Proceedings of the 1998 International Computer Music Conference*, Ann Arbor, MI, USA, October 1998. ICMC 98.

- [Wit53] L. Wittgenstein. *Philosophical investigations*. Macmillan, New York, 1953.
- [WJG77] C.C. Wier, W. Jesteadt, et D.M Green. Frequency discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.*, 61(1) :178–184, 1977.
- [WM97] D.R. Wilson et T.R. Martinez. Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, 6 :1–34, 1997.
- [WV84] W.H. Warren et R.R. Verbrugge. Auditory perception of breaking and bouncing events : A case study in ecological acoustics. *J. Experimental Psychology : Human Perception and Performance*, 10 :704–712, 1984.
- [WW88] R.P. Wildes et Richards W.A. Recovering Material Properties from Sound. Dans Richards W.A., éditeur, *Natural Computation*, chapitre 25, pages 356–363. MIT Press, Cambridge, MA, 1988.