



L'effet de la fréquence fondamentale sur le timbre

Jeremy Marozeau

► To cite this version:

Jeremy Marozeau. L'effet de la fréquence fondamentale sur le timbre. Autre. Université Pierre et Marie Curie - Paris VI, 2004. Français. NNT : . tel-00008742

HAL Id: tel-00008742

<https://theses.hal.science/tel-00008742>

Submitted on 17 Aug 2001

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

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

THÈSE de DOCTORAT de l'UNIVERSITÉ
de Pierre et Marie Curie, PARIS VI

Spécialité :
Science Mécanique, Acoustique et Électronique

Présentée par
M. Jeremy MAROZEAU

pour obtenir le grade de
DOCTEUR de L'UNIVERSITÉ
de Pierre et Marie Curie, PARIS VI

Sujet de la thèse

L'effet de la fréquence fondamentale
sur le timbre

Thèse soutenue le 25 octobre 2004 devant un jury composé de :

Georges Canévet, DR CNRS, LMA Marseille (Rapporteur)
Laurent Demany, DR CNRS, Université de Bordeaux II (Rapporteur)
Alain de Cheveigné, DR CNRS, IRCAM Paris (Directeur de thèse)
Benoit Fabre, PR, Université de Paris VI (Président du jury)
Emmanuel Bigand, PR, Université de Bourgogne (Examineur)
Stephen McAdams, PR, Université de McGill Montréal (Examineur)
David Wessel, PR, Université de Californie Berkeley (Examineur)

Résumé

Cette thèse est dédiée à l'étude de l'effet de la fréquence fondamentale (F_0) sur le timbre. Elle relate sept expériences utilisant des stimuli issus d'instruments naturels ou synthétiques.

Les expériences effectuées consistaient à demander à des sujets de juger des dissemblances de timbre entre des paires de stimuli sans tenir compte de leur éventuelle différence de hauteur. Les expériences I-II et III utilisaient des stimuli issus d'instruments de musique de l'orchestre occidental. Les autres expériences utilisaient des stimuli synthétisés par addition de sons purs. L'expérience I présentait des paires de sons appariés en F_0 . L'expérience II les présentait avec une différence constante de F_0 . L'expérience III a fait varier dans la même session cette différence de F_0 (ΔF_0). L'expérience IV utilisait des stimuli synthétiques ne variant que par leur enveloppe temporelle et avec les mêmes ΔF_0 que celles de l'Exp. III. L'expérience V utilisait des stimuli ne variant que par leur centroïde spectral et avec les mêmes ΔF_0 que celles de l'Exp. III et IV. L'expérience VI utilisait également de stimuli ne variant que par leur centroïde spectral mais avec des ΔF_0 plus importantes. Finalement l'expérience VII utilisait de stimuli ne variant que par leur étalement spectral et avec les mêmes ΔF_0 que celles que l'Exp. VI.

Les résultats montrent que les sujets sont capables de juger une différence de timbre entre deux stimuli avec un écart important de F_0 (du moins jusqu'à 18 demi-tons). Toutefois, une quantité proportionnelle à la ΔF_0 , mais faible par rapport au timbre, est ajoutée aux jugements de dissemblance, ayant pour conséquence l'apparition d'une dimension prédite par la F_0 .

En première approximation, le timbre des instruments est stable face aux variations de la fréquence fondamentale. En deuxième approximation, des variations sont observées, mais elles restent faibles comparées aux différences entre les instruments. Ces variations peuvent être prédictibles par l'analyse du signal.

Aucun effet significatif de la F_0 sur la dimension du timbre prédite par l'impulsivité n'a été observé. Cette dimension est donc considérée comme stable face à la F_0 .

La dimension du timbre prédite par le centroïde spectral se déplace de façon uniforme vers le côté faible en fonction de la F_0 . Le descripteur a donc été corrigé afin qu'il intègre cet effet.

Un léger effet significatif et dépendant du timbre a été observé sur la dimension perceptive liée à l'étalement spectral. Cette dimension reste toutefois relativement stable face à la F_0 .

Summary

This dissertation examines the effect of the fundamental frequency (F_0) on the timbre of a musical instrument. Seven experiments are reported, that were run with natural or synthetic sounds.

In these experiments, subjects were asked to judge timbre dissimilarities between two stimuli while ignoring their possible pitch difference. Experiments I, II and III used stimuli from natural instruments of the occidental orchestra. The other four experiments used stimuli synthesized by pure tones addition.

Experiment I presented pairs of stimuli with the same F_0 . Experiment II presented them with a constant difference in F_0 . Experiment III varied in the same session these F_0 differences (ΔF_0). Experiment IV used synthetic stimuli changing only by their temporal envelope and using the same ΔF_0 as in experiment III. Experiment V used stimuli changing only by their spectral centroid and using the same ΔF_0 as in experiments III and IV. Experiment VI used also stimuli changing only by their spectral centroid, but with a wider set of ΔF_0 . Finally, experiment VII used stimuli changing only by their spectral spread and using the same ΔF_0 as in experiment VI.

The results show that subjects are able to judge a timbre dissimilarity between two stimuli with a large F_0 difference. However, a certain quantity, proportional to the ΔF_0 and small compared to the timbre difference, is added to dissimilarity judgments, having for consequence the appearance of a dimension predicted by F_0 .

As a first approximation, the fundamental frequency has no effect on the timbre of musical instruments. In a second approximation, small effects are observed, but they are weak compared to the differences between instruments. These variations can be predicted from a specific analysis of the signal.

No significant effect of F_0 was found on the timbre dimension predicted by the impulsiveness. This dimension can be considered as independent of F_0 .

The timbre dimension, predicted by the spectral centroid, shifts itself with the F_0 . Its descriptor has been corrected to integrate this effect.

A small and significant effect of F_0 has been found on the timbre dimension predicted by the spectral spread.

Remerciements

Mes premiers remerciements vont à celle qui m'a soutenu, épaulé, encouragé, qui a corrigé, et recorrige et rererecorrige ce document. Ma dame, merci, simplement pour tout.

Je remercie ensuite sincèrement mon directeur de thèse Alain de Cheveigné, qui m'a appris un métier, mon métier. Au-delà des aspects techniques, j'ai appris avec lui les côtes nobles de la recherche comme la rigueur, l'honnêteté et la persévérance.

Grand merci à Anne Caclin, compagne de labeur, pour ses multiples corrections.

Mes remerciements vont également à Stephen McAdams pour m'avoir accueilli au sein de son équipe et m'avoir fait découvrir le monde de la psychoacoustique.

Un chaleureux merci à David Wessel pour ses conseils avisés et son accueil au CNMAT.

Merci également à Sabine Meunier et Georges Canevet, pour avoir recueilli un pauvre thésard en fin de thèse au sein de son équipe au LMA.

Merci Don Patrick Susini pour avoir aidé mon intégration à Marseille et m'avoir fait découvrir sa passion pour le design sonore.

Je remercie bien sûr tous mes collègues de l'Ircam, les demoiselles : Anne (encore), Aline, Alice. Les demoiseaux Guillaume le roi des Pouet-Pouets, Julien le duc des Loubards, Olivier le sage de l'Excelsior et Daniel le marchand de Carrensac. Les Madames : Suzanne et Sylvie ; et les Messieurs : Benneth et Gérard.

Je remercie également collègues de l'équipe APIM du LMA : Florence, Sophie, Michelle, Merce et Guy. Plus particulièrement Isabelle et Mathieu pour les apéros-thèse du mois d'Août.

Je remercie particulièrement l'employé SNCF de la gare de Hyères qui a récupéré l'unique exemplaire de cette thèse corrigée et oubliée un beau jour dans le

TGV.

Je remercie bien sûr énormément mes parents adorés qui même dans l'incompréhension la plus totale m'ont soutenu à tous moments.

Merci également à la centaine de personnes qui sont venues courageusement passer une ou plusieurs de mes expériences. Sans cette double centaines d'oreilles rien n'aurait été possible.

Et enfin merci à : Miles, Coltrane, Mingus, Duke, Herbie, Tom Zé, Chico...

encore à fofie, et à notre futur.

Table des matières

Remerciements	5
1 Introduction	11
1.1 Définitions du timbre	11
1.2 Objectifs de la thèse	12
1.3 Intérêts de la recherche	13
1.4 Plan du document	14
2 La psychoacoustique du timbre	15
2.1 De la psychophysique à la psychoacoustique	15
2.2 Les relations psychoacoustiques <i>primaires</i>	17
2.3 Les relations psychoacoustiques <i>secondaires</i>	18
2.4 La hauteur	20
2.5 Le timbre	24
2.6 Effet de la F_0 sur le timbre	28
2.7 Questions abordées dans cette thèse	32
3 Méthodes	35
3.1 Stimuli	35
3.2 Tâche et dispositif expérimental	37
3.3 Analyses	38
4 Expérience I. Dissemblance de timbre à F_0 égale	45
4.1 Méthodes	45
4.2 Résultats et analyses	46
4.3 Implications pour le timbre	52
4.4 Conclusions	53

5	Expérience II. Dissemblance de timbre à F_0 inégale	55
5.1	Méthodes	56
5.2	Résultats et analyses	56
5.3	Implications pour le timbre	61
5.4	Conclusions	63
6	Expérience III. Dissemblance de timbre à différence de F_0 variable	65
6.1	Méthodes	65
6.2	Résultats et analyses	66
6.3	Implications pour le timbre	73
6.4	Conclusions	73
7	Descripteurs de signal	75
7.1	Introduction	75
7.2	Les descripteurs d'impulsivité	76
7.3	Les descripteurs de centroïde spectral	79
7.4	Les descripteurs d'étalement spectral	83
7.5	Les autres descripteurs pertinents pour le timbre	84
7.6	Comment améliorer les descripteurs	87
7.7	Conclusions	89
8	Expérience V. Le centroïde spectral	91
8.1	Méthodes	92
8.2	Résultats et analyses	94
8.3	Implications pour le timbre	101
8.4	Conclusions	102
9	Expérience VI. Le centroïde spectral au-delà de l'octave	103
9.1	Méthodes	103
9.2	Résultats et analyses	105
9.3	Implications pour le timbre	110
9.4	Conclusions	113
10	Expérience VII. L'étalement spectral	115
10.1	Méthodes	115
10.2	Résultats et analyses	116
10.3	Implications pour le timbre	119
10.4	Conclusions	121

11	Expérience IV. L'impulsivité	123
11.1	Méthodes	123
11.2	Résultats et analyses	125
11.3	Implications pour le timbre	128
11.4	Conclusions	130
12	Discussion générale	131
12.1	Psychoacoustique du timbre	131
12.2	Effets de la F_0	133
12.3	Qu'avons-nous appris sur le comportement des instruments selon la note jouée ?	134
12.4	Descripteurs - Intérêt pour les applications	135
12.5	Acquis méthodologiques	136
12.6	Critique, perspectives	137
13	Résumé des Conclusions	141
A	Statistiques des descripteurs pour les instruments utilisés	143
A.1	Impulsivité	143
A.2	Centroïde spectral	145
A.3	Etalement spectral	146
B	Analyse multidimensionnelle de proximité	149
B.1	La MDS classique	149
B.2	Les différents algorithmes de MDS	151
B.3	Les choix de la représentation	154
C	Tessitures des instruments de l'orchestre	157
D	Lexiques	159
D.1	Qualités perceptives	159
D.2	Caractéristique physique	160
D.3	Divers	161

Chapitre 1

Introduction

1.1 Définitions du timbre

Avant de se lancer dans une étude sur le timbre, il semble essentiel de se poser la question de sa définition. Ce mot est souvent utilisé par bon nombre de personnes venant d'horizons différents : musiciens, acousticiens, linguistes, psychoacousticiens... Les définitions sont diverses et peuvent prendre un sens différent suivant le contexte. Toutefois, trois usages principaux, souvent confondus, ressortent :

Le timbre-identité Le timbre peut servir à désigner les propriétés permettant la reconnaissance d'une catégorie de source sonore. Cette notion est utilisée par exemple lorsque l'on parle *du timbre d'une guitare, d'un violon, d'une voix d'homme...* Le timbre caractérise ici une impression sonore générale : toutes les guitares ont le même timbre, quelles que soient les caractéristiques de l'instrument, du musicien, de la nuance ou de la note jouée.

Le timbre-individualité Le timbre peut être utilisé afin de caractériser une source particulière. On parle par exemple *du timbre de sa première guitare, de la voix d'une personne, d'un musicien...* Le timbre caractérise encore une impression globale mais limitée à une source unique ou d'une association : source plus musicien.

Le timbre-qualité Le timbre est utilisé en psychoacoustique afin de désigner une qualité perceptive au même titre que la hauteur ou la sonie. Dans ce sens, il caractérise la perception d'un son par d'autres qualités que la hauteur, la durée,

la sonie ou la localisation. On parle par exemple *du timbre d'un son, d'une note d'un instrument, d'une voyelle...*

Les deux premiers usages sont très proches. Ils se réfèrent à des classes de sons préétablies et laissent la possibilité à une certaine variabilité à l'intérieur de chaque classe (les différentes notes ou réalisations de notes d'un même instrument, etc.). Ils se différencient par le niveau de détail des classes (i.e. le timbre de la guitare vs le timbre d'une guitare). Le troisième usage est nécessaire pour l'analyse des 2 autres. En effet, l'identité d'un instrument ou d'une personne se base sur l'ensemble des timbres-qualité émis par cet instrument. Une clarinette peut, par exemple, suivant sa tessiture, son mode de jeu, ses nuances, émettre des sons ayant des timbres-qualité très différents. Toutefois, comme il est possible dans bien des cas de reconnaître une clarinette à partir d'une simple note, la confusion entre les trois usages a souvent été faite. Les notes issues d'un même instrument possèdent des caractéristiques communes et des caractéristiques qui leur sont propres. La variation du timbre-qualité d'un instrument selon la note jouée, l'intensité, etc. reflète la physique de l'instrument. Cette fonction de variation peut donc servir d'indice de reconnaissance du type d'un instrument, d'un instrument ou d'un individu particulier.

Dans ce document, nous utiliserons le mot "timbre" suivant la définition du timbre-qualité. Cette définition correspond à l'esprit de la définition de l'ANSI, retranscrite dans l'annexe D et que nous commenterons dans le Chapitre 2.

1.2 Objectifs de la thèse

Notre étude est consacrée à l'effet des variations du timbre des sons d'instruments naturels ou synthétiques en fonction de la fréquence fondamentale (F_0). Cette question est intéressante à plusieurs titres. Comme nous venons de le voir, cette variation peut être une caractéristique importante pour la reconnaissance d'un instrument. Comprendre comment le timbre varie avec la F_0 est donc un préambule de l'étude sur le timbre-identité. Pour mettre en évidence la contribution des variations de timbre-qualité à l'identité d'un instrument, il est nécessaire de combiner deux axes de recherche : le premier, acoustique, où les aspects mécaniques des instruments influençant le timbre du son seraient étudiés en fonction de la F_0 de l'excitateur ; le deuxième, psychoacoustique, où l'influence de la F_0 sur le timbre serait étudiée comme a pu l'être par exemple l'influence de l'intensité sur la hauteur. Notre étude se limite uniquement à ce deuxième axe. Seul un bref aperçu de l'étendue des variations acoustiques, résumées par les descripteurs du

signal, sera exposé dans l'annexe A. Une fois l'axe perceptif étudié, les résultats de l'axe acoustique pourront être convenablement interprétés.

1.3 Intérêts de la recherche

Comprendre comment le timbre varie avec la F_0 revêt trois intérêts principaux.

Intérêt scientifique Cette question entre directement dans le domaine de recherche en psychoacoustique. Cette science, discutée plus en détail au Chapitre 2, étudie les relations entre un phénomène acoustique et la perception auditive qu'elle crée. La psychoacoustique du timbre a été étudiée en mettant en relation, d'une part, les analyses multidimensionnelles censées révéler des dimensions perceptives, d'autre part, les analyses physiques. Ces expériences ont toutes été faites à F_0 constante ; l'étude de la dépendance du timbre sur la F_0 restait à faire.

Intérêt pour la conception de descripteurs du timbre Les études sur le timbre ont permis la conception de plusieurs descripteurs du signal capables de prédire la perception du timbre d'un stimulus. Toutefois, ces études, ne faisant pas varier la F_0 , n'ont pas permis à ces descripteurs d'intégrer ce paramètre. De nouvelles études sur le timbre comportant différentes F_0 permettraient d'améliorer ces descripteurs.

Intérêt musical Le timbre a longtemps été négligé dans la musique occidentale. Le compositeur se contentait généralement d'indiquer le nom de l'instrument ainsi que quelques rares indications comme "*sul tasto*" ou "*pizzicato*". De nos jours, notamment avec l'arrivée de la musique concrète puis électroacoustique, les compositeurs ont pris conscience que le timbre pouvait faire partie de la composition musicale au même titre que la hauteur. Etudier comment le timbre varie avec la F_0 , permet également de mieux comprendre comment la hauteur et le timbre interagissent. Cette question devient importante lorsque ces deux qualités varient à l'intérieur d'un même morceau de musique.

Applications possibles Les descripteurs de signal qui prédisent le timbre sont utiles pour les applications d'indexation par le contenu de bases de données audio ou multimédia. Comprendre comment le timbre-qualité varie avec la F_0 permet

également de mieux caractériser le timbre-identité d'une catégorie et donc d'améliorer les algorithmes de reconnaissance de source.

1.4 Plan du document

Ce document commence par deux chapitres théoriques. Le premier, Chapitre 2, expose les bases de la psychoacoustique, ainsi que certaines études classiques sur le timbre. Le suivant, Chapitre 3, résume les différentes méthodes d'analyse utilisées dans cette thèse.

Les premiers chapitres expérimentaux, Chapitres 4 à 6, décrivent une série de trois expériences sur la variation du timbre de stimuli issus d'instruments de musique en fonction de la F_0 .

Le Chapitre 7 est consacré aux descripteurs du signal et à leur conception.

Les Chapitres 8 et 9 décrivent des expériences effectuées avec des stimuli synthétiques portant sur l'effet de la F_0 sur la dimension du timbre prédite par le centroïde spectral. Le Chapitre 10 reprend l'expérience du Chapitre 9 mais avec la dimension du timbre prédite par l'étalement spectral. Le Chapitre 11 reprend l'expérience du Chapitre 8 mais avec la dimension du timbre prédite par l'impulsivité.

Le Chapitre 12 discute des résultats de l'ensemble des expériences, et le Chapitre 13 conclut.

L'annexe A montre les variations des descripteurs le long de la tessiture des instruments de l'orchestre occidental. L'annexe B consiste en un bref rappel de l'analyse multidimensionnelle de proximité. Cette technique statistique sera utilisée fréquemment durant cette étude. L'annexe C rappelle la tessiture des instruments de l'orchestre occidental. Enfin, l'Annexe D donne plusieurs définitions de notions importantes.

Chapitre 2

La psychoacoustique du timbre

Notre étude portant sur l'étude psychoacoustique de la relation secondaire entre la F_0 et le timbre, ce chapitre pose certaines bases théoriques. Il commence par rappeler ce qu'est la psychophysique et la psychoacoustique. Puis il passe rapidement en revue les relations primaires et secondaires. Deux sections seront ensuite accordées à la hauteur et au timbre. Enfin, les études sur la relation entre la F_0 et timbre seront discutées.

2.1 De la psychophysique à la psychoacoustique

2.1.1 La psychophysique

La psychoacoustique est une branche d'une science plus générale appelée psychophysique, considérée comme la porte d'entrée de la psychologie dans les sciences (voir Bagot, 1999, et Gescheider, 1976, pour un historique complet). En effet, avant elle, de grands savants comme Kant estimaient que puisque l'on ne peut pas appliquer les mathématiques à l'âme, la psychologie n'est et ne sera jamais une science. En 1860, Fechner montre le contraire dans son livre "Elemente der Psychophysik". Il y jette les bases de cette nouvelle science, qui a pour vocation d'étudier les relations entre les stimuli physiques et les sensations qu'ils induisent. Il s'appuie notamment sur les travaux d'un physiologiste allemand du début du 19ème siècle : Ernst Weber. Celui-ci est célèbre pour sa loi émise en 1831 qui stipule que la quantité minimum d'intensité¹ ΔI à ajouter à une intensité

¹Le terme intensité est utilisé ici dans le sens générique de quantité d'une grandeur physique.

I déjà présente afin qu'une variation soit perceptible est proportionnelle à cette intensité. Cela implique que plus l'intensité d'un stimulus est élevée, plus la quantité à ajouter pour qu'un changement soit perçu doit également être élevée. Fechner propose une loi reliant la sensation perçue avec le logarithme de l'intensité. Avec l'amélioration des méthodes de mesure et l'élaboration de meilleurs protocoles expérimentaux, cette loi a été de plus en plus critiquée. Stevens (1975) consacre un chapitre de son livre *Psychophysics* à décrire Fechner. Il propose une nouvelle loi reliant la sensation perçue avec une fonction de puissance de l'intensité. Cette loi est notamment utilisée afin de prédire une sensation de sonie à partir de l'intensité d'un son pur.

2.1.2 La psychoacoustique

La psychoacoustique s'intéresse spécifiquement aux relations entre un phénomène acoustique et la perception auditive qu'il engendre. Un son est produit par la vibration d'une source, puis est capté par l'auditeur à travers son appareil auditif. Celui-ci transforme le son par des filtrages successifs et transmet l'information au système nerveux. Cet influx est finalement traduit en une sensation auditive. Les expériences psychoacoustiques tentent de quantifier ces sensations par des questions précises destinées à aider l'auditeur à les exprimer.

Dans le cas classique, avec des sensations comme la sonie, les réponses sont représentées sur une échelle. Le stimulus est ensuite analysé et caractérisé par une mesure physique. Les deux types de mesures sont ensuite comparées. Cela permet de déterminer l'influence de la variation d'un facteur physique sur une perception auditive et d'aider à la compréhension du fonctionnement de certains processus perceptifs.

Dans le cas plus complexe, avec des sensations comme le timbre, un son peut véhiculer plusieurs qualités auditives, les réponses sont donc décomposées en plusieurs dimensions "psychologiques" orthogonales ou non. Le stimulus est ensuite caractérisé par plusieurs mesures que l'on représente sur des dimensions "physiques". Finalement, les deux types de dimensions sont comparés. Certaines analyses statistiques permettent même de quantifier l'effet de chaque caractéristique physique sur la sensation considérée.

2.2 Les relations psychoacoustiques *primaires*

Prenons l'exemple d'un son pur de durée limitée, définie entre 0 et T par l'équation

$$s(t) = A \times \sin(2\pi ft + \phi), \quad (2.1)$$

où $s(t)$ représente le signal de pression à l'instant t , f sa fréquence, ϕ sa phase et A son amplitude. En variant chacun de ces paramètres d'une quantité suffisante et de façon indépendante, la perception de ce son sera modifiée. Une variation de l'amplitude entraîne une différence de sonie ; une variation de fréquence, une différence de hauteur ; et une variation de durée, une différence de durée perceptive. Chacune de ces grandeurs physiques a une influence directe sur une qualité perceptive. Ces relations sont qualifiées de *primaires*, et commentées dans la suite de cette section.

La sonie et l'intensité acoustique Stevens (1955) propose afin d'estimer la sonie une fonction de la puissance 0,3 de l'intensité du signal : $S_{sones} = (1/15.849) \times (I/I_0)^{0.3}$, où, S_{sones} représente la sonie estimée en sones, I l'intensité du signal, et I_0 l'intensité de référence. Cette loi permet une règle simple : *la sonie est doublée tout les 10 dB*. Une unité perceptive de sonie nommée "sones" est définie en prenant pour référence la sonie d'un son pur de 1000 Hz à un niveau de 40 dB SPL (Hartmann, 1997)

La hauteur et la fréquence La hauteur peut être prédite par la fréquence fondamentale (F_0), correspondant pour les sons périodiques à l'inverse de leur période. Dans la pratique l'extraction de la F_0 n'est pas toujours une tâche triviale (voir de Cheveigné et Kawahara, 2002, pour une revue des différentes méthodes). Afin de mesurer la sensation de hauteur des sons purs, Stevens et Volkmann (1937) ont construit une échelle appelée "mel" en demandant aux sujets d'estimer une hauteur deux fois plus et deux fois moins élevée. Par convention, une hauteur de 1000 mels correspond à la hauteur perçue d'un son pur de 1000 Hz. Cette échelle indique une relation linéaire pour des fréquences entre 1000 et 10000 Hz. Pour les basses fréquences (en dessous de 1000 Hz), la hauteur décroît plus rapidement que la fréquence, et pour des hautes fréquences elle décroît moins rapidement. Toutefois, cette échelle fut souvent critiquée comme l'explique Hartmann (1997) à cause, notamment, de sa différence avec les échelles musicales connues et la physiologie de la cochlée.

La durée perçue et la durée physique La relation entre la durée perçue (subjective) et la durée physique (objective) peut être approchée par une droite de pente 1. Stevens (1975) montre, plus précisément, que cette relation suit une fonction de puissance de 1,1. Creelman (1962) démontre que la loi de Weber est respectée avec une constante de 0,125. Zwicker et Fastl (1990) proposent de fixer une unité de durée perceptive nommée 'dura'. Un dura, par définition, correspond à la sensation de durée induite par un son pur de 1 kHz à 60 dB SPL avec une durée de une seconde. Puis, par une tâche consistant à doubler et réduire de moitié la durée, ils ont pu construire une relation entre la durée subjective et la durée physique d'un son de 1 kHz.

2.3 Les relations psychoacoustiques *secondaires*

Outre la dépendance d'une qualité perceptive sur le paramètre physique qui l'affecte de façon primaire, un effet secondaire sur les autres qualités peut être perçu. Nous qualifierons ces relations de *secondaires*. Cette section fait un tour d'horizon de ces différentes relations. L'effet de la F_0 sur le timbre étant le thème principal de cette thèse nous lui avons dédié la section 2.6.

2.3.1 Effets secondaires sur la sonie

Nous avons vu que la sonie dépendait de façon primaire de l'intensité. Mais elle dépend de façon secondaire de la fréquence, de la répartition spectrale et de la durée du signal. L'effet de la fréquence sur la sonie d'un son pur est décrit dans les graphes d'isotonie (Fletcher et Munson, 1933; Robinson et Dadson, 1956) indiquant l'intensité à donner à un son pur pour obtenir un niveau de sonie donné. Cet effet peut en partie être expliqué par les caractéristiques physiologiques de l'oreille externe et moyenne (Killion, 1977; Puria et coll., 1997). On ne peut toutefois pas exclure d'autres facteurs.

De nouveaux phénomènes apparaissent en considérant les sons complexes. Par exemple, la sonie d'un son composé de deux harmoniques dépend fortement de l'écart de fréquence séparant les deux harmoniques. Si cette différence est supérieure à une largeur de bande critique alors la sonie totale est équivalente à la somme des sonies partielles dues à chaque son pur séparément. Si cette différence diminue et que les deux sons purs tombent dans la même bande critique alors la sonie totale se calcule à partir de la fonction de puissance de la somme de l'intensité des deux sons purs. La sonie peut donc être influencée par la largeur de bande

ou plus généralement par la répartition spectrale de l'énergie. Elle peut être prédite par un descripteur plus complexe faisant intervenir les puissances dans l'ensemble des bandes critiques (Zwicker et Fastl, 1990; Moore et Glasberg, 1996).

La sonie dépend aussi de la durée. En effet, lorsque la durée d'un stimulus est inférieure à une seconde, celle-ci affecte la sonie de façon monotone. Munson (1947) a étudié cette relation et a modélisé la sonie Ψ ainsi : $\Psi = \Psi_{max} \times E_s(t)$, où Ψ_{max} est la sonie d'un son de même amplitude et de durée infinie et où $E_s(t)$ est une fonction intégrative (Hartmann, 1997). Pedersen et coll. (1977) démontrent que cette relation suit une loi exponentielle dont la constante de temps est de 80 ms.

2.3.2 Effets secondaires sur la hauteur

L'effet secondaire de l'intensité sur la hauteur d'un son pur a été étudié en détail par Stevens (1935). Il constata que non seulement la hauteur des sons à basse fréquence (en dessous de 1000 Hz) diminuait avec une augmentation de l'intensité (Fletcher, 1934), mais que pour les sons de fréquence élevée (au-dessus de 4000 Hz) la hauteur augmentait avec l'intensité. Ces résultats furent souvent remis en cause. Vershuure et van Meeteren (1975) ont montré qu'ils n'étaient valides que sur des résultats moyennés sur les sujets et que l'effet de l'intensité sur la hauteur pouvait varier énormément d'un sujet à l'autre. Burns (1982) montra que cet effet pouvait également varier pour un même sujet d'une oreille à l'autre.

Un faible effet de la durée physique du signal sur la hauteur a été montré par Doughty et Garner (1948).

Comme nous le verrons dans la section 2.4.3, une variation de la distribution de l'énergie dans l'enveloppe spectrale, résumée par le centroïde spectral, peut facilement faire varier la perception de la hauteur. Cet effet reflète l'existence de deux hauteurs ; la première est liée à la F_0 et la seconde à la position sur l'axe fréquentielle d'une concentration de puissance (de Cheveigné, 2004). Il se manifeste principalement à travers des erreurs d'octave (Demany, 1989). Cependant, des effets plus subtils peuvent également avoir lieu (Singh et Hirsh, 1992).

L'enveloppe temporelle influence, comme nous le verrons plus loin, directement le timbre. Mais elle peut également influencer la hauteur. Cet effet pouvant paraître surprenant a été démontré par Hartmann (1978). Quinze sujets devaient trouver la hauteur la plus élevée entre des paires de stimuli. Ces paires étaient composées de deux sinus (appariés en sonie), ayant deux enveloppes temporelles différentes. La première était carrée et pouvait être considérée comme l'enveloppe d'un son entretenu. La seconde suivait une exponentielle décroissante et pouvait

être considérée comme l'enveloppe d'un son impulsif. Les résultats montraient que, à F_0 équivalent, les sons avec une enveloppe temporelle exponentielle possédaient une hauteur tonale plus élevée (de 0,7 à 2,6 %) que les sons à enveloppe carrée.

2.3.3 Effets secondaires sur la durée perceptive

Creelman (1962) a étudié la durée perceptive de signaux en fonction de l'intensité du bruit de fond dans lesquels ils sont immergés. Il a montré que pour un rapport signal sur bruit faible, la capacité à détecter des différences de durée grandissait rapidement avec l'intensité du signal, mais que cette dépendance devenait négligeable lorsque le signal devenait fort et clair par rapport au bruit. Nous pouvons donc conclure que l'intensité, sauf conditions particulières, n'a pas d'effet sur la durée perçue.

2.4 La hauteur

Cette thèse traitant des effets secondaires de la F_0 sur le timbre, il est utile de passer en revue certains aspects de la hauteur, corrélat primaire de la F_0 . Nous verrons ainsi qu'il est possible que la hauteur et le timbre aient certaines caractéristiques en commun.

2.4.1 Les dimensions de la hauteur d'un son pur : la tonie et le chroma

Jusqu'ici nous avons considéré la hauteur comme une qualité perceptive unidimensionnelle, descriptible par une grandeur physique unique : la F_0 . Cependant, les choses ne sont pas aussi simples. Deux sons séparés par une octave peuvent être perçus comme ayant une hauteur plus proche que deux sons séparés par une autre relation harmonique (comme le triton par exemple) bien que la différence de fréquence fondamentale ΔF_0 soit plus grande dans le premier cas que dans le deuxième. Cela remet en cause le caractère unidimensionnel de la hauteur. Bachem (1950) affirme qu'elle est composée de deux dimensions que nous nommerons par les termes de la tonie et du chroma comme le recommande Demany (1989). Comme lui, nous utiliserons le terme de hauteur tonale afin de désigner la sensation de hauteur "globale", correspondant à ce que les Anglo-saxons appellent "pitch".

La tonie L'AFNOR définit la tonie comme le "*Caractère de la sensation auditive lié à la fréquence des sons*". Trois notes sont ajoutées :

- *Le "mel" a été proposé comme unité de tonie ;*
- *Pour une même fréquence, elle peut varier légèrement avec la pression acoustique ;*
- *Elle peut être caractérisée par la fréquence d'un son pur soit à un niveau de pression acoustique conventionnel, soit à un niveau d'isotonie conventionnel, qui devront être normalisés, donnant à l'auditeur moyen une sensation de tonie comparable.*

La première note montre que la tonie correspond à la qualité mesurée par Stevens et Volkmann (1937). La deuxième note indique un effet secondaire de l'intensité sur la tonie. La troisième note rappelle que la tonie peut être prédite par la F_0 .

Le chroma Bien que les gammes musicales varient énormément d'une culture à l'autre, l'octave, correspondant selon l'échelle occidentale à douze demi-tons, reste un écart commun à tous. Il peut être obtenu en multipliant la F_0 par 2 (du moins en première approximation voir Shepard, 1982, pour une revue). La deuxième octave (vingt quatre demi-tons) peut être obtenue en multipliant la F_0 par 4, la troisième en la multipliant par 8, etc. Tous les sons, différenciés d'une ou plusieurs octave, partagent une même qualité perceptive, on dit alors qu'ils ont le même chroma. De manière générale, Bachem (1950) avance que tous les sons dont la fréquence est dans un rapport 2 ont le même chroma.

2.4.2 Les dimensions de la hauteur tonale des sons complexes périodiques : la hauteur brute et la hauteur fondamentale

D'autres phénomènes perceptifs apparaissent en considérant les sons complexes périodiques. Ces sons constituent la majorité des sons émis par des sources induisant une sensation de hauteur comme la voix ou les instruments de musique. Ils sont composés d'harmoniques ayant des fréquences correspondant à $n \times F_0$ (avec $n = 1, 2, 3, 4, \dots$), où F_0 est une fréquence donnée nommée fréquence fondamentale. Un auditeur, devant apparier la hauteur d'un son pur et d'un son complexe, ajustera la fréquence du son pur assez précisément avec celle de la F_0 du son complexe, avec cependant parfois des erreurs d'octave (Demany, 1989). Terhardt (1974) oppose deux modes de perception de la hauteur tonale : le premier qualifié d'*analytique* est utilisé pour percevoir la hauteur des partielles séparément qu'il

nomme *hauteur spectrale* ; le second qualifié de *synthétique* consiste à intégrer plusieurs éléments, des harmoniques dans ce cas, pour créer une sensation unique selon la théorie du *Gestalt*, il la nomme *hauteur virtuelle*. Cette même notion est nommée *hauteur fondamentale* par Demany (1989).

Il est également possible de percevoir une hauteur correspondant à une fréquence fondamentale à partir d'un son ne contenant pas d'énergie dans cette même fréquence. Un son par exemple avec des harmoniques à 400, 500 et 600 Hz, sera perçu avec la même hauteur qu'un son pur de 100 Hz. On dit dans ce cas que la fondamentale est "absente", où "virtuelle", et sa hauteur "résiduelle" (Schouten, 1970). Si l'on compare ce son avec un autre composé de trois harmoniques aux fréquences : 100, 200 et 300 ayant également une hauteur fondamentale appariée avec un son pur à 100 Hz, on remarque que les deux stimuli diffèrent par une autre qualité perceptive. Le premier, avec la fondamentale absente, sera décrit par les sujets comme plus aigu et le second plus grave. Cette qualité est nommée par Demany (1989) la *hauteur brute*.

Dans les sons complexes, il est possible de faire varier indépendamment la hauteur fondamentale et la hauteur brute. Shepard (1964) utilise cette caractéristique pour créer sa fameuse illusion auditive de sons qui montent indéfiniment. Risset (1971) montre même que certains sujets peuvent être plus sensibles à la hauteur brute qu'à la hauteur fondamentale et vice-versa.

2.4.3 La représentation de la hauteur

Nous avons vu que la hauteur tonale d'un son pur est décomposable en au moins deux dimensions : la tonie et le chroma. Ces deux dimension peuvent être représentées par une configuration en spirale (Shepard, 1999; Risset, 1978), comme le montre la Figure 2.1.

Ueda et Ohgushi (1987) ont validé cette représentation par des tests psychoacoustiques. Ils ont obtenu par une analyse MDS un espace en spirales à partir de jugements de dissemblance entre toutes les paires extraites d'un corpus de stimuli composé de sons synthétiques variant principalement par leur F_0 . L'enveloppe spectrale de ces stimuli était basée sur ceux conçus par Shepard pour obtenir son illusion auditive de sons qui montent indéfiniment.

Patterson (1990) propose pour les sons complexes de représenter la hauteur non plus par une spirale mais par un cylindre. Son argumentation est la suivante : pour les sons complexes, il est possible de se déplacer continuellement d'une note vers son octave sans changer de chroma. Le domaine de la hauteur des notes de musique ressemble donc plus à une surface entourant un cylindre. Il propose à

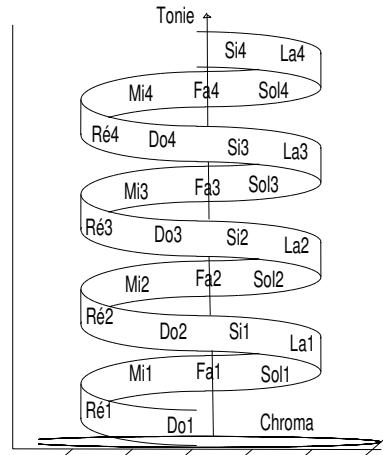


FIG. 2.1: Schéma représentant l'espace de la hauteur tonale, suivant deux dimensions, la première suivant un cercle, la hauteur chromatique, et la seconde verticale, la tonie. La combinaison des deux forme une spirale selon Shepard (1999) ou un cylindre selon Patterson (1990).

cette fin plusieurs méthodes :

- l'atténuation des harmoniques paires ;
- la translation de l'enveloppe spectrale ;
- le déphasage entre les harmoniques paires et impaires.

Il est intéressant de noter que les deux premières méthodes impliquent directement une variation du timbre. La translation de l'enveloppe spectrale, en particulier, influence de façon principale une dimension du timbre décrite par le centroïde spectral et souvent désignée par la brillance (Von Bismarck, 1974; Faure, 2000). Les deux notions sont donc fortement liées. Certains auteurs considèrent même que la hauteur brute est une dimension du timbre (voir Bagot 1999), ou à l'inverse que la brillance est une dimension de la hauteur (comme Hesse 1982, cité par Robinson 1993).

Une confusion supplémentaire apparaît, en considérant que puisque la hauteur fondamentale correspond à la hauteur d'un son pur de fréquence F_0 , celle-ci peut également être décomposée en chroma et en tonie. La tonie d'un son complexe peut être considérée comme une dimension discriminant des sons complexes de même chroma à différentes octaves. La frontière entre la tonie d'un son complexe et la hauteur brute est fine et compliqué. Robinson (1993) a toutefois montré que ces deux notions sont bien dissociables, comme le sont également le chroma et la

tonie d'un son complexe (Warren et coll., 2003). On pourrait donc considérer que la hauteur est composée de trois dimensions, plus ou moins fortement liées. Notre étude pourra contribuer à clarifier la question.

2.5 Le timbre

2.5.1 Difficultés propres à la recherche sur le timbre

L'étude du timbre ne peut s'effectuer avec les mêmes méthodes que celle de la hauteur, la sonie ou la durée pour plusieurs raisons. Premièrement, il semble que le timbre peut varier selon plusieurs dimensions perceptives en même temps et ne peut donc pas être décrit par une grandeur physique unique. Deuxièmement, la définition du timbre reste encore aujourd'hui assez vague. L'American Standards Institute (1960) propose la définition suivante : *"Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar"*, pouvant être traduite ainsi : *"Le timbre est la qualité perceptive utilisée par l'auditeur pour estimer la différence entre deux stimuli présentés dans les mêmes conditions avec la même sonie et la même hauteur"*. Une note accompagne cette définition : *"Timbre depends primarily upon the spectrum of the stimulus, but it also depends on the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus"* ou en français : *"Le timbre dépend premièrement du spectre fréquentiel du stimulus, mais également de la forme d'onde, de la pression acoustique, de la disposition des fréquences à l'intérieur du spectre et des caractéristiques temporelles du stimulus."*. Cette définition ne dit pas ce qu'est le timbre, mais ce qu'il n'est pas. Finalement, comme nous l'avons souligné lors de l'introduction, le sens même du mot timbre peut évoluer en fonction du contexte.

2.5.2 Du jugement de dissemblance aux configurations dans un espace MDS

Il existe plusieurs méthodes expérimentales destinées à étudier le timbre (voir Hajda et coll., 1997, pour une revue détaillée). Une des plus commune consiste à demander aux sujets de juger la dissemblance de timbre entre des stimuli. La définition du timbre résidant dans ce qui permet à deux stimuli d'être perçus comme différent, il semble naturel de demander aux sujets de quantifier cette différence.

Les stimuli sont généralement appariés en sonie (ou en niveau dBA suivant les études), en F_0 et en durée, afin de n'être différenciés que par leur timbre. La tâche consiste alors à juger la dissemblance entre les stimuli présentés à travers toutes les paires possibles, à partir d'un corpus expérimental. Les résultats de l'expérience sont recueillis à l'intérieur d'une matrice de dissemblance. La matrice de dissemblance, une fois obtenue, permet à l'analyse multidimensionnelle de proximité, MDS de l'anglais MultiDimensional Scaling, de représenter dans un espace euclidien les différents timbres éloignés les uns des autres en fonction de leur dissemblance. Cette technique reste très utilisée dans la recherche sur le timbre. Elle permet notamment de mettre en lumière les différentes dimensions qui le composent. La section suivante retrace différents espaces de timbre souvent cités dans la littérature. Le principe de la MDS est expliqué dans l'Annexe B et les techniques utilisées pour cette étude dans le Chapitre 3.

2.5.3 Les espaces de timbre

L'espace de Pols

Pols et coll. (1969) demandèrent à des auditeurs de juger les dissemblances entre 11 voyelles synthétiques, appariées en F_0 , par une méthode de comparaison triadique. L'analyse MDS de ces dissemblances a permis de représenter ces voyelles dans un espace perceptif à trois dimensions. Puis, ils ont construit un espace dans lequel les attributs physiques des stimuli sont représentés de la manière suivante : Pour chaque stimulus l'énergie par bande en tiers d'octave est calculée. Comme 18 filtres sont utilisés, une matrice 11×18 est obtenue. Par analyse en composantes principales, cette matrice est réduite à 3 dimensions expliquant 81,7 % de la variance. Après rotation de la solution MDS pour aligner ses axes sur les trois dimensions de la solution ACP, les corrélations entre dimensions perceptives et physiques sont bonnes (0,99, 0,97 et 0,74). Cette étude démontra qu'il était possible de prédire, par des attributs physiques, un jugement perceptif du timbre de sons complexes.

L'espace de Grey

La technique de Pols et coll. a été appliquée aux sons d'instruments de musique par Wedin et Goude (1972), Wessel (1973), Miller et Carterette (1975) et Grey (1977). L'espace créé par ce dernier est fréquemment cité et ses stimuli sont repris dans bon nombre d'expériences.

Grey (1977) a demandé à des sujets de juger la similarité de toutes les paires d'un corpus de 16 stimuli sur une échelle de 1 à 30. Ces stimuli ont été créés à partir des instruments de musique suivant : deux hautbois (joués par deux différents musiciens, sur deux instruments différents), un cor anglais, un cor français, un basson, une clarinette en *mi♭*, une clarinette basse, une flûte, deux saxophones alto (même instrument joué mezzo forte et piano), un saxophone soprano, une trompette, un trombone et trois violoncelles (un même instrument joué sul tasto, sul ponticello, et normalement). Chaque son a été analysé puis reconstruit par synthèse additive. Finalement, ils ont été ajustés afin d'être appariés en sonie, F_0 , et durée perceptive. Par une analyse MDS avec la technique INDSCAL (voir l'annexe B), il a représenté le timbre de ces stimuli dans un espace tridimensionnel comme le montre la figure 2.2.

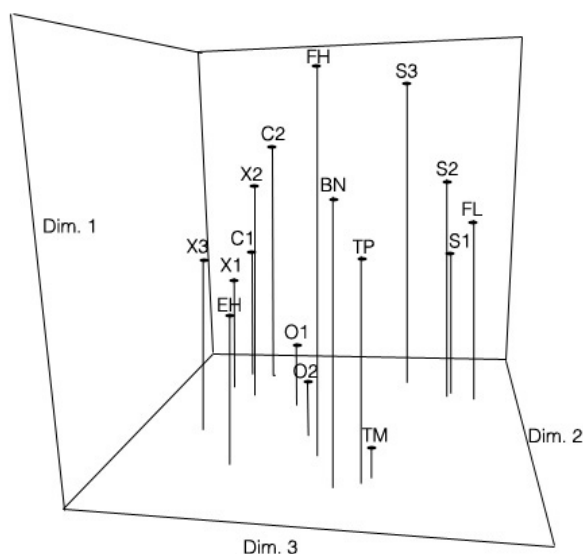


FIG. 2.2: *Espace de timbre tridimensionnel issue de Grey (1977). Les abréviations sont les suivantes : O1, O2 les hautbois ; C1, C2 les clarinettes, X1, X2, X3 les saxophones ; EH le cor anglais ; FH le cor français, S1, S2, S3 les violoncelles ; TP la trompette ; TM le trombone ; FL la flûte, et BN le basson*

Afin de déterminer les aspects physiques corrélés aux dimensions perceptives, le spectrogramme de chaque stimulus a été analysé. La première dimension a été interprétée comme étant liée à la distribution du spectre d'énergie. Elle distingue

la trompette et le saxophone 3 du trombone. Les premiers ont une faible largeur de bande spectrale, concentrée sur les basses fréquences. Le trombone, au contraire, possède un large spectre. La seconde dimension a été interprétée comme étant liée à la synchronie des harmoniques durant l'attaque. Elle distingue en effet les bois (possédant un temps d'oscillation rapide), des cordes et des cuivres. La dernière dimension a été interprétée comme liée à la présence d'énergies à haute fréquence, inharmoniques, et de basse amplitude durant l'attaque. Elle discrimine les instruments comme le violoncelle, la flûte et la clarinette, des instruments comme les cuivres possédant une énergie inharmonique durant l'attaque située dans les basses fréquences.

Cette expérience a mis en avant au moins deux paramètres physiques, bien qu'encore assez mal définis, corrélés avec des dimensions perceptives du timbre : la qualité d'attaque et la distribution des harmoniques dans le spectre.

L'espace de McAdams

Dans les années 1980, sous l'impulsion de David Wessel, puis de Stephen McAdams, l'Ircam entreprit plusieurs recherches sur la perception du timbre. De nouveaux espaces furent obtenus à partir de stimuli créés par synthèse FM, basés sur l'enveloppe spectrale et temporelle d'instruments naturels (Wessel, 1979; Krumhansl, 1989; McAdams et coll., 1995; Donnadieu, 1997). Ces expériences ont permis de trouver les corrélats physiques des dimensions perceptives : le logarithme du temps d'attaque, le centre de gravité spectral, et la fluctuation spectrale (Krimphoff et coll., 1994). Des améliorations ont été apportées sur les méthodes d'analyse multidimensionnelles de proximité, comme la notion de spécificité et de classe latente (Winsberg et Carroll, 1989; Winsberg et De Soete, 1993), voir l'annexe B. Les nouvelles techniques de synthèse des sons ont permis notamment de créer des instruments hybrides, comme le vibrone (mélange du vibraphone et du trombone), ou la guitarnet (mélange de la guitare et la trompette). Le but était de vérifier si l'instrument hybride se situait à mi-chemin entre ces deux composants. L'étude de McAdams montre, par exemple, un stimulus considéré comme un 'piano frotté' partageant bien la même position que le piano, le long de la dimension décrite par le centre de gravité, mais se situant à son opposé sur la dimension décrite par le temps d'attaque (voir figure 2.3).

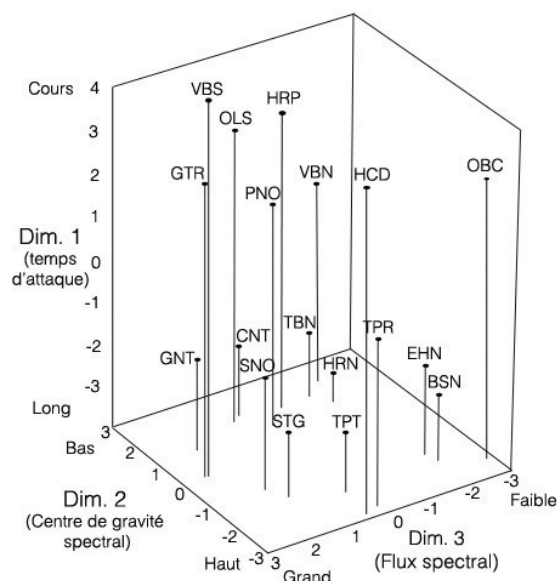


FIG. 2.3: Espace de timbre tridimensionnel issue de McAdams (1995). Les abréviations sont les suivantes : hrn le cor d'harmonie ; tpt la trompette, tbn le trombone, hrp la harpe, tpr le trumpar (trompette/guitare), ols le oboleste (hautbois/celesta), vbs le vibraphone, sno le striano (corde frottée/piano), hcd le clavecin, ehn le cor anglais, bsn le basson, cnt la clarinette, vbn le Vibrone (vibraphone/trombone), obc le obochord (haubois/clavecine), gtr la guitar, stg les cordes frottées, pno le piano, gnt la guitarnet (guitare/clarinette)

2.6 Effet de la F_0 sur le timbre

2.6.1 Effet de la F_0 sur une tâche de discrimination de timbres

La tâche de discrimination de deux timbres est rendue plus difficile s'ils sont jugés avec deux F_0 différentes. Parallèlement, la capacité à discriminer deux hauteurs se dégrade également en fonction de la différence de répartition spectrale qui les sépare. Cet effet peut être interprété comme le résultat d'une interaction entre la hauteur et le timbre et, par conséquent, on peut formuler l'hypothèse que ces deux attributs perceptifs ne sont pas traités de façon indépendante par le cerveau.

Beal (1985) a demandé à des auditeurs si deux accords joués soit à la guitare, soit au piano, soit au clavecin étaient identiques. La tâche a été effectuée

sans problème, avec un taux d'erreur quasi-nul, lorsque les deux accords étaient joués sur le même instrument. Cependant, lorsque la discrimination devait se faire entre deux instruments différents, et à fortiori deux timbres différents, le taux d'erreur montait de façon importante (jusqu'à 30 % pour les musiciens, et 64 % pour les non musiciens.) Dans une seconde expérience, Beal a demandé aux auditeurs de juger si deux accords étaient joués sur le même instrument. Les résultats montrèrent que les auditeurs faisaient plus d'erreurs lorsque les accords étaient différents. Toutefois, l'effet était moindre.

Melara et Marks (1990) ont étudié la relation entre le timbre et la hauteur en utilisant la tâche de classification de Garner (1974) avec quatre stimuli construits à partir de deux enveloppes spectrales et deux F_0 différentes. Les auditeurs devaient discriminer soit les deux timbres, soit les deux hauteurs, la qualité non jugée étant soit stable soit variable. Dans les deux cas, la discrimination du timbre et de la hauteur, le taux d'erreur augmentait si la qualité non jugée variait aléatoirement. Le temps de réaction augmentait également, passant de 345 ms, pour la discrimination du timbre à hauteur fixe, à 407 ms lorsque la hauteur variait, et de 345 ms pour la discrimination de la hauteur à timbre fixe, à 392 ms en variant le timbre. Krumhansl et Iverson (1992) refirent les mêmes expériences, mais avec des sons musicaux et obtinrent des conclusions similaires.

Moore et Glasberg (1990) montrèrent que plus les timbres sont différents (i.e. moins ils contiennent d'harmoniques en commun), plus leur différence de F_0 doit être grande afin qu'elle soit perceptible.

Semal et Demany (1991) effectuèrent des expériences en se basant sur un paradigme de reconnaissance de hauteur de Deutsch (1972) dans lequel l'auditeur devait juger si deux sons étaient identiques ou différents. Les sons utilisés étaient uniquement des sons purs dont seule la fréquence variait. Les sujets devaient donc indiquer si une éventuelle différence de hauteur était perçue. Entre la présentation de ces deux sons, les expérimentateurs introduisirent des sons "perturbateurs" que l'auditeur ne devait pas prendre en considération pour effectuer son jugement. Ces perturbateurs pouvaient être de plusieurs formes : 1) silence ; 2) des sons purs ayant une F_0 aléatoire mais proche du son de base ; 3) des sons purs mais avec une F_0 éloignée des sons testés ; 4) des sons complexes avec une F_0 équivalente aux sons du 2) ; 5) des sons complexes mais des F_0 trois fois plus basses que pour 2). Les sons 2) et 4) pouvaient être considérés comme perçus avec les mêmes hauteurs mais des timbres différents, et il en va de même pour les sons 3) et 5). Les résultats montrèrent que seuls les "perturbateurs" avec des F_0 proches (2 et 4) des sons testés affectaient les résultats. La variation de timbre n'ayant pas apporté de perturbations supplémentaires, on peut donc penser que le timbre est stocké

indépendamment de la hauteur dans la mémoire auditive à court terme.

Krumhansl et Iverson (1992) continuèrent ce même type d'études mais avec un paradigme légèrement différent. Un stimulus (la cible) devant être mémorisé était intégré dans une séquence de sons (le contexte). L'auditeur devait être attentif soit à un éventuel changement de timbre, soit un changement de hauteur. Les résultats ont été observés en fonction du contexte, qui pouvait varier soit en timbre soit en hauteur. Si les performances chutaient lorsque le contexte variait par la qualité non testée, alors les deux qualités interagissaient. Les résultats montrèrent qu'il n'y a pas de dégradation du score pour une tâche de mémorisation de la hauteur lorsque le timbre du contexte change, et vice-versa. Cette étude montre un autre résultat intéressant : lorsque le contexte varie en hauteur, la tâche de reconnaissance de hauteur est affectée, mais une variation de timbre n'affecte pas la tâche de reconnaissance de timbre. Ces résultats suggèrent que la hauteur est perçue de façon relative, alors que le timbre est perçu de façon absolue. Des résultats similaires ont été trouvés par Starr et Pitt (1997).

Citons également des expériences traitant de l'effet du timbre sur la perception d'une mélodie. Celle-ci est la résultante de plusieurs variations de hauteurs associées au même flux. Le timbre est un des éléments principaux dans le processus de fusion (McAdams et Bregman, 1979). En effet, deux sons avec un timbre similaire seront associés à une même source. Wessel (1979) a montré qu'en variant un paramètre du timbre la mélodie pouvait changer. Tout en préservant une structure de trois notes, il a introduit des différences de timbre en variant le centroïde spectral d'une note sur deux. La perception d'une mélodie unique s'est transformée progressivement en deux mélodies distinctes.

2.6.2 L'enveloppe spectrale et la F_0

Le timbre dépend de la forme de l'enveloppe spectrale. Celle-ci peut être définie comme un profil modelant l'énergie de chaque harmonique selon sa fréquence ou son rang. En parole, elle est souvent caractérisée par ses formants. Les solutions MDS ont confirmé sa grande influence sur le timbre. Une variation de cette enveloppe entraîne principalement une variation de timbre. Mais comment doit-elle varier pour conserver un timbre constant lorsque la F_0 est modifiée ? Deux hypothèses sont possibles : 1) elle reste fixe et l'énergie de chaque harmonique s'adapte en fonction de sa fréquence sur cette enveloppe ; 2) elle subit une dilatation ou une contraction (suivant que la F_0 augmente ou descend), afin de conserver le rapport d'énergie entre les harmoniques. Cela revient à se demander laquelle de ces deux caractéristiques doit rester constante : l'enveloppe spectrale ou l'énergie

sur chaque harmonique.

Slawson (1968) étudia cette question en demandant à des auditeurs d'ajuster la fréquence des deux premiers formants afin que deux sons, différenciés par une octave, soient perçus avec des timbres similaires. Le meilleur résultat fut obtenu par une augmentation de 10 % de ces fréquences.

Plomp (1970) demanda à des auditeurs d'estimer la dissemblance entre 9 stimuli, créés par la combinaison de trois F_0 à 200, 250, 320 Hz et de trois enveloppes spectrales constituées d'un unique formant à 2000, 2500 et 3200 Hz. Cette expérience montra que les auditeurs trouvaient plus similaires les stimuli ayant la même enveloppe spectrale. Ces dissemblances furent ensuite représentées sur un espace à deux dimensions. La première dimension était parfaitement corrélée à la fréquence du formant de l'enveloppe spectrale (le timbre) et la deuxième était corrélée avec la F_0 . Ces résultats furent confirmés par Mencl (1998) dans une expérience similaire. Ils accréditent l'hypothèse de l'enveloppe spectrale fixe.

2.6.3 Effet de la F_0 sur l'espace de timbre

Miller et Carterette (1975) ont demandé aux sujets d'effectuer un jugement de dissemblance entre tous les éléments d'un corpus de 27 stimuli, créés à partir de toutes les combinaisons possibles de trois F_0 s (200, 400, 800 Hz), de trois enveloppes temporelles, et de trois enveloppes spectrales. Les réponses ont été analysées par une MDS. Un espace tridimensionnel a été obtenu par une analyse INDSCAL (voir annexe B) où le poids que les sujets accordaient à chaque dimension était calculé. L'analyse a révélé que la première dimension contenait quasiment toute la variance. Cette dimension étant fortement corrélée à la F_0 , ils en ont conclu que les différences de hauteur avaient masqué les différences plus subtiles du timbre. Ils refirent donc les mêmes expériences en appariant les stimuli en F_0 . Depuis, les expériences de dissemblance sur le timbre se font principalement à F_0 constante.

2.6.4 Effet de la F_0 sur le timbre-identité

Nous avons émis l'hypothèse dans l'Introduction que l'identification d'un instrument de musique pouvait se faire, en partie, sur la base de la *variation* du timbre (timbre-qualité) en fonction de la note émise. Si c'est le cas, le taux d'identification devrait augmenter avec l'étendue de la F_0 des notes présentées. Le spectre du violon, par exemple, est sujet à des variations complexes dues à ses différents modes de résonance (Fletcher et Rossing, 1998). La répartition de l'énergie sur

chacune de ses harmoniques varie donc fortement avec la F_0 . Le timbre des notes des instruments à vent peut également changer rapidement entre les notes graves (avec le trou de registre fermé) et les notes aiguës (avec le trou de registre ouvert). Ces caractéristiques apparaissent uniquement si l'instrument est joué sur toute sa tessiture (Risset et Wessel, 1999). La trajectoire d'un instrument dans l'espace de timbres (qualité) pourrait ainsi contribuer à son timbre (identité). Le violon est ainsi difficilement identifiable à partir d'une synthèse effectuée par un clavier bon marché, qui ne fait que transposer une enveloppe spectrale avec une différentes F_0 .

Sandell et Chronopoulos (1997) montrèrent qu'un sujet entraîné à reconnaître un instrument améliore ses performances en fonction de la tessiture des notes couverte durant son entraînement. Handel et Erickson (2001) montrèrent que les sujets étaient capables d'associer deux notes au même instrument, si leur différence de F_0 restaient en dessous de l'octave. Nous avons vu dans la section 2.6.2 que pour qu'un timbre reste inchangé, son enveloppe spectrale doit être fixe. McAdams et Rodet (1988) ont montré que l'ajout d'un vibrato pouvait améliorer la perception de cette enveloppe.

Les descripteurs du timbre étant élaborés à partir d'espaces de stimuli ayant tous une même F_0 , ce paramètre est resté constant lors de leur élaboration (Krimphoff et coll., 1994; Misdariis et coll., 1998; Peeters, 1999). Cependant, comme le timbre varie en fonction de sa tessiture, les algorithmes destinés à la reconnaissance automatique d'instruments de musique, devraient logiquement intégrer ce paramètre. Kitahara et coll. (2003) ont montré que, si la F_0 était prise en considération dans les algorithmes, le taux de reconnaissance augmentait considérablement.

2.7 Questions abordées dans cette thèse

Problématiques L'espace MDS se révèle être un outil privilégié dans l'étude du timbre car il permet d'abord de le décomposer en dimensions psychologiques, puis de lui associer des descripteurs physiques pertinents. Cependant, nous avons vu que depuis Miller et Carterette (1975) les expériences de dissemblance ne s'effectuent plus qu'avec des stimuli appariés en F_0 . Pourtant, un espace contenant des timbres issus de stimuli ayant des F_0 différentes pourrait être une source d'information fort utile. Les compositeurs seraient également intéressés par connaître la différence entre les timbres de deux sons joués à des notes différentes. Il serait également possible d'étudier les différences entre le timbre des notes issues

d'un même instrument. Les descripteurs pourraient être améliorés en intégrant le paramètre F_0 .

Afin de créer un tel espace, il est nécessaire de demander aux sujets d'effectuer un jugement sur des timbres à F_0 différentes. L'ajout de ce paramètre pourrait induire plusieurs difficultés. Premièrement, comme dans les expériences de Miller et Carterette (1975), les dissemblances pourraient être dominées par la hauteur, rendant l'espace MDS peu pertinent. Deuxièmement, si la hauteur et le timbre se révèlent être deux notions inséparables, alors les comparaisons entre timbres seraient possibles seulement à hauteur appariée. Troisièmement, il est possible que des dimensions, considérées dans certains cas comme appartenant au timbre et dans d'autres à la hauteur (comme la hauteur brute et la brillance), viennent perturber les jugements.

Afin de contourner le premier problème (prédominance de la hauteur), nous avons introduit dans des expériences de dissemblance de timbre une consigne supplémentaire : le sujet devra désormais faire abstraction d'une éventuelle différence de hauteur et ne baser son jugement que sur le timbre. Pour le deuxième problème (timbre et hauteur inséparable), les études décrites dans la Section 2.6.1, nous laissent espérer que ce ne sera pas le cas. Le troisième problème (confusion des dimensions) semble être le plus délicat. Une attention particulière sera portée lors de la création des stimuli afin que la hauteur tonale des stimuli ne soit pas ambiguë (i.e. qu'il n'y ait pas d'erreur d'octave par exemple).

Hypothèses Ceci devrait amener à la création d'un espace contenant des timbres créés à partir de stimuli ayant des F_0 différentes. Plusieurs hypothèses peuvent être émises sur leur position en fonction de leur F_0 :

1. Invariance de la position du timbre ;
2. Déplacement isométrique gardant les positions relatives invariantes ;
3. Déplacement non isométrique.

La première hypothèse implique que la variation du timbre d'un stimulus en fonction de la F_0 est négligeable comparée aux différences des timbres contenus dans l'espace. La seconde hypothèse admet une rotation ou un déplacement uniforme dans un sens commun à tous les timbres. La troisième hypothèse permet un déplacement propre à chaque timbre dans une direction arbitraire. Nos expériences ont été effectuées afin de pouvoir choisir parmi ces trois hypothèses.

Pour résumer, nous avons vu dans ce chapitre que les relations psychoacoustiques pouvaient être répertoriées en deux catégories nommées *primaire* et *secondaire*. Chacune de ces relations a été brièvement commentée. Cette thèse étant destinée à étudier l'effet secondaire de la fréquence fondamentale sur le timbre, la notion de hauteur ainsi que ses dimensions ont été exposées. En outre, la définition du timbre a été commentée. Nous avons passé en revue quelques expériences classiques représentant le timbre dans des espaces multidimensionnels, pour arriver finalement aux études traitant de l'effet de la F_0 sur le timbre. Le chapitre suivant résume les différentes méthodes statistiques utilisées dans cette thèse.

Chapitre 3

Méthodes

Ce chapitre présente le protocole expérimental et les méthodes statistiques utilisés dans cette thèse.

3.1 Stimuli

Les expériences utilisaient comme stimuli des sons d'instruments naturels (Exp. I à III) ou de synthèse (Exp. IV à VII). Cette section présente les stimuli des trois premières expériences, les autres seront décrites dans le chapitre 8.

Dix instruments naturels et deux de synthèse ont été utilisés. Chaque instrument était joué sur trois notes différentes : si_2 (à 247 Hz), $do\#_3$ (à 277 Hz), et sib_3 (à 466 Hz)¹. Ces notes ont été choisies afin de pouvoir explorer une petite différence de F_0 (deux demi-tons ou une seconde majeure en harmonie : entre si_2 et $do\#_3$) et une différence plus importante (onze demi-tons ou une septième majeure entre si_2 et sib_3). L'intervalle de onze demi-tons est supérieur à celui de deux en termes de F_0 , mais inférieur en termes de chroma. Il représente environ le tiers de la tessiture typique d'un instrument de l'orchestre, voir l'Annexe C. Compte tenu du chevauchement des tessitures des différents instruments, l'intervalle de onze demi-tons correspond à l'intervalle maximum qu'on peut former avec des notes communes aux instruments choisis.

Les sons des instruments naturels étaient issus de la base de données : Studio En Ligne (IRCAM, 2000). Cette base de données présente l'avantage sur celle

¹Ce document utilise la notation française dans laquelle le la_3 correspond à 440 Hz (A4 en notation anglo-saxonne).

de McGill (Opolko et Wapnick, 1989), souvent utilisée dans les expériences psychoacoustiques, que chaque note est jouée par le même musicien, sur le même instrument et enregistrée avec les mêmes micros dans une même pièce. Le Tableau 3.1 donne la liste des instruments, naturels et de synthèse, avec leurs abréviations.

TAB. 3.1: Tableau récapitulatif des instruments utilisés lors des Exp. I-III.

Instruments	abréviations	particularités
guitare	Gu	le si_2 est joué sur la corde de mi grave, le $do\#_3$ sur la corde de la et le $si\flat_3$ sur la corde de ré.
harpe	Hr	
violon pizzicato	Vp	le si_2 et $do\#_3$ sont joués sur la corde de sol, le $si\flat_3$ sur la corde de ré.
violon frotté	Vf	les notes sont jouées sur les mêmes cordes que pour le violon pizzicato
contrebasse frottée	Cb	toutes les notes sont jouées sur la corde de sol
Synthèse A	SA	instrument de synthèse à enveloppe spectrale fixe dérivée du saxophone alto
Synthèse B	SB	même commentaire que pour SA
hautbois	Hb	
clarinette	Cl	en $si\flat$
flûte traversière	Fl	en do
cor	Cr	en fa
trompette	Tr	en do

Ces instruments ont été choisis afin que chaque famille soit représentée : les cordes pincées (Gu, Hr, Vp), les cordes frottées (Vf, Cb), les bois (Hb, Fl), et les cuivres (Cr, Tr). L'étude de McAdams et coll. (1995) comportait des sons hybrides dont certaines caractéristiques physiques étaient les mêmes, tandis que d'autres différaient fortement (comme le "piano frotté" et le "piano normal" par exemple). Le Vp et Vf ont été choisis dans ce même esprit. Ces deux instruments diffèrent par leur qualité d'attaque et se ressemblent par leur spectre.

Deux instruments de synthèse ont été créés à partir de l'enveloppe spectrale d'une note de saxophone jouée mezzo forte sur le si_2 . Le premier, SA, utilise cette enveloppe pour synthétiser des sons à si_2 , $do\#_3$ et $si\flat_3$. Le second, SB, utilise

une enveloppe dilatée d'un facteur de 1,89 pour synthétiser des sons à ces mêmes notes. Le rapport de 1,89 correspond au rapport de F_0 entre si_2 et si_3 , le spectre de SB à si_3 est donc la version dilatée du spectre de SA à si_2 . Les instruments SA et SB ont été conçus pour déterminer si le timbre dépend de l'enveloppe spectrale définie en terme de fréquence ou de rang harmonique.

Tous les stimuli (sauf Vp) ont été tronqués à une durée de 1,5 s par l'application d'une rampe cosinusoidale de 0,2 s démarrant à 1,3 s. Les notes issues du violon pizzicato ayant une durée inférieure à 1,5 secondes, la portion de leur décroissance a été étirée à l'aide d'un logiciel de manipulation proposé par le système Studio En Ligne afin qu'ils atteignent la durée de 1,5 s. La trompette, le cor, et la harpe ont été légèrement filtrés dans la région de 10 kHz afin d'éliminer certains bruits parasites dus aux conditions d'enregistrement. Les stimuli ont été égalisés en sonie par une expérience d'ajustement de niveau avec celui de Cb, pris comme stimulus de référence, et présenté à 60 dBA. Cette tâche a été effectuée par 6 sujets ne participant pas aux autres expériences. Les stimuli ont été échantillonnés à une fréquence de 44100 Hz avec 16 bits de résolution. Ils ont été présentés aux sujets à travers un casque Sennheiser 520 II.

Le choix de stimuli d'instruments naturels est motivé par la volonté d'assurer à notre étude une pertinence musicale. Il a l'inconvénient de confondre deux effets possibles de la F_0 sur le timbre : des effets génériques qui affecteraient le timbre de tout son dont la F_0 change, et des effets spécifiques à des instruments particuliers, dus à des variations dans les processus de production en fonction de la note jouée. Dans notre étude, nous comptons sur les techniques d'analyse post-hoc pour faire la part de ces deux types d'effet.

3.2 Tâche et dispositif expérimental

La tâche consistait à estimer la dissemblance de timbre entre les sons d'une paire en ajustant un curseur sur une échelle continue. Dans les instructions données aux sujets, le timbre était défini comme "la quatrième composante d'un son, les trois autres étant : la hauteur, la sonie (ou l'intensité perçue), et la durée". Ils avaient pour instruction d'utiliser toute l'échelle et de ne pas tenir compte des éventuelles différences de hauteur, durée, sonie ou "bruit d'enregistrement". La consigne stipulait également que le jugement devait se faire uniquement sur les qualités sonores perçues. Si l'instrument de musique dont était issu un stimulus était reconnu, cette information ne devait pas être prise en considération dans le jugement. Les expériences se sont déroulées dans une cabine audiométrique pour-

vue d'un écran et d'une souris d'ordinateur ainsi que d'une sortie audio sur un casque. L'environnement expérimental a été conçu avec le logiciel PsiExp (Smith, 1995). L'interface était composée d'un curseur pouvant se déplacer avec la souris sur une échelle continue allant de similaire (codé 0) à différent (codé 1), et deux boutons, permettant soit de réécouter les sons, soit de valider le choix. Une fois le jugement validé, l'interface jouait une nouvelle paire choisie de façon aléatoire. Avant chaque session, les sujets avaient la possibilité d'écouter tous les stimuli utilisés dans le sous-groupe pour avoir une idée de l'ordre de grandeur des dissemblances présentées.

La plupart des sujets ont été recrutés parmi les visiteurs des journées Portes Ouvertes de l'IRCAM. Ils avaient des profils très variés et étaient pour la majorité extérieurs au monde de la psychoacoustique. Certains ont effectué plusieurs expériences, comme ceux des Exp. I et III, d'autres n'en ont fait qu'une seule. Ils ont été indemnisés 9,15 euros de l'heure pour leur participation.

3.3 Analyses

Les scores de dissemblance pour chaque sujet et session sont placés dans une matrice de dimension $n \times n$, où n représente le nombre de stimuli et l'élément ij la dissemblance entre les stimuli i et j . Selon les expériences, nous obtenons deux cas de figure : 1) l'ordre de présentation n'est pas considéré (si la paire ij est testée, alors la paire ji ne le sera pas), alors seule la partie inférieure de la matrice est remplie ; 2) toutes les paires sont présentées, alors la matrice est entièrement remplie. Ces données sont analysées par plusieurs techniques statistiques qui s'appliquent, soit à l'ensemble des matrices pour tous les sujets (analyse en cluster, ANOVA), soit à des matrices moyennées entre les sujets (coefficient de corrélation entre matrices, MDS).

3.3.1 L'analyse en cluster

L'analyse en cluster est une méthode statistique permettant d'extraire d'un jeu de données les grandes classes regroupant les objets entre eux. L'analyse se base sur une matrice de proximité entre les objets considérés. Elle commence par grouper les deux objets les plus proches, puis par itération va hiérarchiser les autres sous la forme d'un arbre, appelé également dendrogramme. Cette représentation permet ainsi d'extraire les grandes classes d'objets regroupées le long des même branches (voir Houix, 2003, pour une revue détaillée). L'analyse de cluster s'ap-

plique aux matrices de dissemblance de chaque sujet et a pour but de grouper les sujets en classes. Cette analyse sera appliquée à chaque expérience pour écarter les sujets dont les réponses sont fortement atypiques.

3.3.2 Le coefficient de corrélation

Le coefficient de corrélation sert à mesurer la similitude entre résultats dans différentes conditions. Il est calculé entre les matrices moyennées des sujets. Lorsque les matrices sont incomplètes, les corrélations entre ces matrices ne sont effectuées que sur la partie présente. Un test permet d'estimer la probabilité que la valeur observée soit le fruit du hasard. La corrélation est jugée significative si cette probabilité estimée est inférieure à un seuil (5 %).

3.3.3 L'analyse de variance (ANOVA)

L'ANOVA (mesures répétées) est utilisée pour comparer différentes conditions. Elle est appliquée à l'ensemble des matrices de dissemblance de tous les sujets et conditions à tester. Elle sert : (1) à dire si les différences observées sont statistiquement significatives et (2) à quantifier leur grandeur (facteur R^2). Le facteur R^2 permettra par exemple d'estimer l'importance des variations de dissemblance entre deux F_0 différentes, en les comparant aux différences entre paires d'instruments à F_0 égale. Les effets non significatifs ou dont le R^2 est faible pourront être écartés de la discussion.

3.3.4 L'analyse multidimensionnelle de proximité (MDS)

L'analyse multidimensionnelle de proximité, abrégée MDS de l'anglais MultiDimensional Scaling, est un outil statistique permettant de représenter un corpus de stimuli dans un espace de faible dimension, typiquement 2, 3 ou 4. A partir des dissemblances jugées et résumées dans une matrice, comme nous l'avons décrite plus haut, une configuration spatiale est obtenue où tous les stimuli sont représentés. Cette section retrace rapidement les différents algorithmes que nous utiliserons. Les détails techniques de la MDS sont exposés dans l'annexe B.

MDSCAL

L'algorithme MDSCAL (MultiDimensionnal SCALing) est le modèle classique de la MDS. Il tente de représenter au mieux un ensemble de dissemblances

dans une configuration spatiale par les distances inter-stimuli : deux stimuli jugés très dissemblables seront placés à deux opposés de l'espace ; deux stimuli jugés très semblables seront placés côte à côte. L'algorithme ne donnant en sortie qu'une matrice de distances, la configuration n'est définie ni en position absolue, ni en orientation. Par convention, l'espace est centré sur l'origine, et l'orientation des axes coïncide avec les directions de variance maximale grâce à une analyse en composante principale (ACP)

Cet algorithme sera utilisé pour analyser les résultats de l'Exp. V, Chapitre 8, à partir d'une implémentation sous MATLAB de l'algorithme de SMACOF (Scaling by Majorizing a Complicated Function) décrite dans Borg et Groenen (1997).

MDSCAL avec rotation procustéenne

MDSCAL donne une solution dont l'orientation spatiale est indéterminée ; toute orientation est acceptable à la place de l'orientation par défaut issue de l'ACP. Si on dispose d'indices qui suggèrent une orientation particulière, on peut les utiliser pour orienter l'affichage de la solution. Dans notre cas, les études précédentes et nos propres investigations indiquent que des descripteurs de signal (centroïde spectral, etc.) ont une valeur prédictive, et nous les utiliserons pour imposer une orientation aux solutions MDSCAL par une rotation dite procustéenne. Cette rotation ne modifie pas la solution MDS, elle permet simplement de l'examiner sous un angle propice. Toutefois, elle risque de diriger leur interprétation vers une hypothèse préétablie. Les coefficients de corrélation entre descripteurs et projections sur les axes donnent a posteriori une indication de la validité du choix d'orientation. Si des valeurs élevées sont obtenues, cela suggère à la fois la pertinence de la rotation, des descripteurs, ainsi que de leur aptitude à prédire chacun une dimension de la solution MDS

Cette rotation sera appliquée pour analyser les résultats de l'Exp. I, Chapitre 4.

EXSCAL

L'algorithme EXSCAL (EXtended mdSCAL), élaboré par Winsberg et Carroll (1989), permet à chaque stimulus de varier sur une dimension qui lui est propre. Ainsi, si un stimulus possède une caractéristique particulière, celle-ci ne nécessite pas de dimension supplémentaire. Cet algorithme est particulièrement adapté pour

un corpus expérimental contenant des sons naturels dont un (ou plusieurs) diffère significativement par un aspect qui lui est typique.

EXSCAL sera utilisé pour analyser les résultats de Exp. III, Chapitre 6. La procédure de choix de modèle aboutira néanmoins à une solution sans spécificité.

INDSCAL et CLASCAL

L'algorithme INDSCAL (INdividual Differences SCALing) proposé par Carroll et Chang (1970) repose sur l'hypothèse que tous les sujets utilisent les mêmes dimensions mais ne les pondèrent pas de la même façon. Cette hypothèse permet d'obtenir un meilleur modèle du comportement des sujets et offre aussi une orientation plausible de la solution MDS. En effet, la rotation qui maximise la variation interindividuelle sur les axes de l'espace doit, selon l'hypothèse de l'algorithme INDSCAL, aligner ces axes sur des dimensions perceptives. A la différence d'une solution MDSCAL avec rotation procustéenne sur un descripteur, l'orientation d'une solution INDSCAL dépend uniquement de la structure des réponses des sujets. Elle ne dépend pas d'un descripteur. Si on constate qu'une dimension d'une solution INDSCAL est fortement corrélée avec un descripteur, alors la dimension perceptive et le descripteur qui le prédit peuvent tous deux être considérés comme particulièrement pertinents.

L'algorithme CLASCAL, élaboré par Winsberg et De Soete (1993), reprend le principe de INDSCAL, à la différence près que les poids sur les dimensions ne sont plus attribués à chaque individu mais à des classes d'individus ayant des patterns de réponses semblables.

Le choix entre les deux algorithmes se fera suivant le nombre de sujets. En dessous de quinze, les classes ne sont pas bien définies, il est donc préférable d'utiliser INDSCAL, et au-delà CLASCAL.

Le faible nombre de sujets présents dans nos expériences ne permet pas à CLASCAL d'être exploité au mieux. Il sera toutefois utilisé pour ré-analyser les résultats de l'Exp. III, Chapitre 12.

3.3.5 La méthode des ancrés

La méthode des "ancres" a pour but de déterminer si le timbre d'un instrument donné a changé significativement d'une F_0 à l'autre. Elle s'appuie sur une analyse de variance appliquée aux sous-ensembles des jugements de dissemblance qui impliquent cet instrument.

Afin d'en illustrer le principe, prenons un instrument X , dont sont issus deux notes à des F_0 différentes. Nous essayons de savoir si leurs timbres X_1 et X_2 sont statistiquement différents. Dans un premier temps, nous pouvons tester si la dissemblance $[X_1X_2]$ diffère de zéro. Malheureusement, l'échelle de réponses est bornée à zéro et l'effet de bord qui en résulte risque de produire systématiquement une dissemblance moyenne non-nulle. Une seconde approche consiste à utiliser un autre instrument Y ("l'ancre") et de tester si les dissemblances $[X_1Y]$ et $[X_2Y]$ diffèrent. Si l'analyse montre qu'elles ne diffèrent pas, on peut conclure que X_1 et X_2 se trouvent sur une hypersphère centrée sur Y (voir figure 3.1). Supposons que cette analyse soit répétée pour un deuxième ancre Z et que $[X_1Z]$ et $[X_2Z]$ ne diffèrent pas, alors X_1 et X_2 se trouvent aussi sur une hypersphère centrée sur Z . L'intersection d'hypersphères centrées sur Y et Z se trouve dans un plan *orthogonal* à la direction YZ . En répétant cette analyse, en utilisant successivement les autres instruments comme ancres, et en supposant qu'aucune comparaison ne révèle de différence significative, alors on peut conclure que le segment $[X_1X_2]$ est soit *nul*, soit orthogonal à l'espace contenant toutes les ancres, c'est à dire (si les ancres ont été bien choisies) *orthogonal à l'espace des timbres*. Si à l'inverse la différence est significative (après correction pour tests répétés) on pourra dire que l'instrument a "bougé" dans l'espace des timbres. On pourra même estimer la direction du déplacement dans un espace MDS à partir du signe de l'effet.

En pratique, la procédure est un peu plus complexe. Chaque ancre Y existe en deux versions, Y_1 et Y_2 , qui diffèrent par leur F_0 . L'une ou l'autre pourrait servir d'ancre. A supposer qu'on choisisse Y_1 , alors des deux dissemblances $[X_1Y_1]$ et $[X_2Y_1]$ à comparer, l'une contient des stimuli de même F_0 et l'autre non. Ceci risque de biaiser la comparaison (par exemple si le sujet n'ignore pas parfaitement les différences de hauteur). Le choix de Y_2 produirait le biais inverse. Cependant, en ajoutant les dissemblances pour ces deux ancres et en comparant les sommes, les biais s'annulent :

$$[X_1Y_1] + [X_1Y_2] = [X_2Y_1] + [X_2Y_2]. \quad (3.1)$$

Pour un instrument donné, si cette équation est vraie en prenant tous les instruments comme ancres, alors son timbre (identité) pourra être considéré comme stable par rapport à la F_0 .

De façon concrète, pour chaque instrument X à tester, on prend tous les jugements entre cet instrument et les autres instruments Y (les ancres). Pour chaque X_i ($i = 1, 2$), on forme $[X_iY_1] + [X_iY_2]$ (Eq. 3.1). Les données sont alors soumises à une ANOVA avec les facteurs F_0 (celle de X) et ancre. L'absence d'effet de F_0 ou

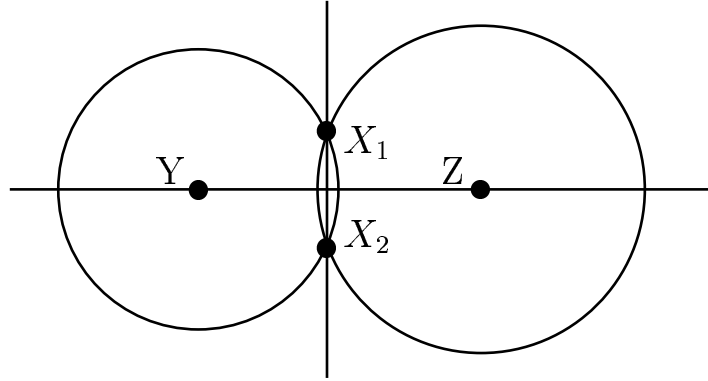


FIG. 3.1: *Experience III. Schéma illustrant la méthode des “ancres”, permettant d’analyser si deux timbres sont significativement différents. Le plan représente un hypothétique espace de timbre bidimensionnel. Y et Z sont deux instruments “ancres”, et X_1 et X_2 sont deux timbres, issu de l’instrument X, dont la différence est étudiée. Si la dissemblance entre X et Y reste invariante malgré le changement de F_0 de X, alors X_1 et X_2 se positionnent sur un cercle centré en Y. Le même raisonnement est possible entre X et Z. Dans le cas où les deux sont vrais alors les timbres de X ne peuvent qu’être positionnés sur une ligne orthogonale à $[YZ]$.*

d’interaction $F_0 \times$ ancre signale la stabilité de l’instrument X. Cette analyse est répétée pour chaque instrument.

La méthode des ancres est utilisée dans l’analyse des résultats de l’Exp. III pour se prononcer sur le changement de timbre de chaque instrument.

Chapitre 4

Expérience I. Dissemblance de timbre à F_0 égale

Cette expérience est une réplique des expériences classiques (Wessel, 1973; Grey, 1977; McAdams et coll., 1995). Les sujets avaient pour tâche de juger la dissemblance de timbre entre des sons instrumentaux de même F_0 . Cette expérience apporte une nouveauté par rapport aux expériences classiques, car elle est répétée à trois F_0 différentes, avec les mêmes sujets et instruments. L'expérience se divise en trois sessions, 'a', 'b', 'c', chacune à une F_0 différente.

4.1 Méthodes

Les stimuli de cette expérience sont décrits en Sect. 3.1. Dans chaque session toutes les paires de sons d'instruments distincts de même F_0 étaient présentées. La F_0 variait de session en session : si_2 pour la session 'a', $do\#_3$ pour la session 'b' et $si\flat_3$ pour la session 'c'. Il y avait 12 instruments et 66 paires de sons par session.

Vingt-sept sujets ont participé à cette expérience, dont 14 hommes et 13 femmes, 15 musiciens et 12 non musiciens. Ils étaient âgés de 22 à 30 ans, et aucun n'a dit présenter de troubles auditifs. Afin d'être catégorisé comme "musicien", le sujet devait avoir pratiqué un instrument de musique pendant les trois dernières années. Un tiers des sujets a effectué les sessions dans l'ordre "abc", un tiers dans l'ordre "cba", et un tiers l'ordre "bca". Cet ordre n'a pas été parfaitement contrebalancé. Il aurait fallu pour cela que le deuxième groupe passe les sessions dans l'ordre "cab". Cependant, la session 'c' apparaissant dans les trois positions

possibles, un éventuel effet d'ordre pourra être décelé par une analyse de variance (ANOVA).

Les sujets ont passé les trois sessions la même journée, séparées par des pauses de 5 minutes. L'expérience totale durait environ une heure.

4.2 Résultats et analyses

4.2.1 Effets de sujet, ordre et expertise instrumentale

Les coefficients de corrélation entre les matrices de dissemblance de chaque sujet ont été calculés puis soumis à une analyse en cluster utilisant une méthode "moyenne" (Kaufman et Rousseeuw, 1990), voir Chapitre 3. Cette analyse, combinée avec la même faite pour l'Exp. II, nous permis d'écarter trois sujets dont les réponses étaient très faiblement corrélées avec les autres. Les résultats qui suivent sont ceux des 24 sujets restants.

Afin de déceler un effet d'ordre éventuel, une ANOVA a été faite sur les résultats de la session 'c' avec les facteurs ordre (3) et paires d'instruments (66). Le nombre de sujets n'étant pas le même pour les trois ordres, un facteur correctif a dû être appliqué sur le degré de liberté du facteur résiduel (Abdi, 1987; Wonnacott et Wonnacott, 1990). Ni l'effet d'ordre [$F(2, 21) < 1$] ni son interaction avec les paires [$F(130, 1365) < 1$] n'étaient significatifs.

Une autre ANOVA a été effectuée pour chaque session avec les facteurs expertise musicale (2) et paires d'instruments (66). Aucun effet d'expertise musicale n'est apparu, ni sous la forme d'un effet principal [$F(1, 65) < 1$], ni sous la forme d'une interaction avec le facteur paire d'instruments [$F(65, 1430) < 1$].

4.2.2 Matrices de dissemblance

Les scores de dissemblance pour chaque sujet et session sont placés dans une matrice de dimension $n \times n$, où n représente le nombre de stimuli. Puisque l'ordre de présentation des stimuli dans chaque paire n'est pas considéré, seule la partie inférieure de la matrice est remplie.

La figure 4.1 montre la matrice de la session 'a', moyennée entre tous les sujets.

Chaque carré représente la dissemblance entre deux instruments. Les trois premières colonnes correspondent aux instruments impulsifs (Gu, Hr ou Vp) et les

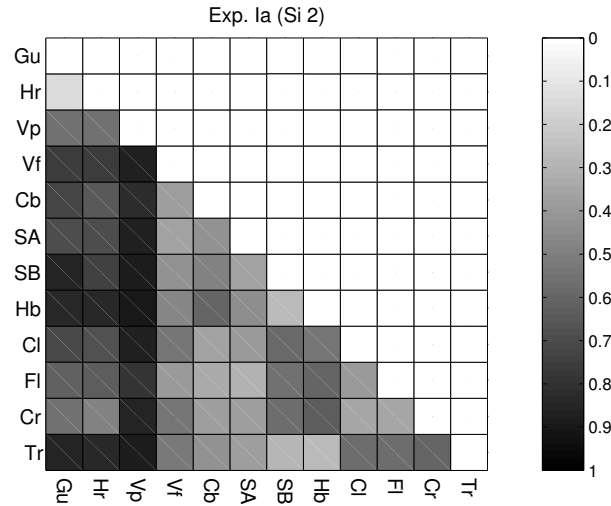


FIG. 4.1: Matrice de dissemblance pour la session 'a' de l'expérience I. Chaque carré représente la dissemblance entre deux instruments. Plus un carré est foncé plus la dissemblance a été jugée importante.

autres aux instruments entretenus (Vf, Cb, SA, SB, Hb, Cl, Fl, Cr, Tr). On distingue deux triangles correspondant aux comparaisons entre instruments impulsifs (en haut à gauche) ou entretenus (en bas à droite), et un rectangle correspondant aux comparaisons entre impulsifs et entretenus. La couleur plus sombre du rectangle indique que la dissemblance est plus grande entre les deux classes qu'à l'intérieur de chaque classe. La figure 4.2 montre les trois matrices moyennes de chaque session.

Les trois matrices sont visuellement semblables. Cette similitude est confirmée par le calcul des coefficients de corrélation entre les parties triangulaires inférieures des matrices moyennes : 0,88 entre les sessions 'a' et 'b' ; 0,81 entre les sessions 'b' et 'c' et 0,89 entre les sessions 'a' et 'c' ($d.d.l. = 64$, $p < 0,001$ pour les trois coefficients).

4.2.3 Analyses de variance

Afin de quantifier les effets de la F_0 , une ANOVA à mesures répétées a été effectuée avec les facteurs F_0 (3) et paires d'instruments (66). L'ANOVA permet, d'une part, de repérer les effets qui pourraient être dus au hasard et de les écarter

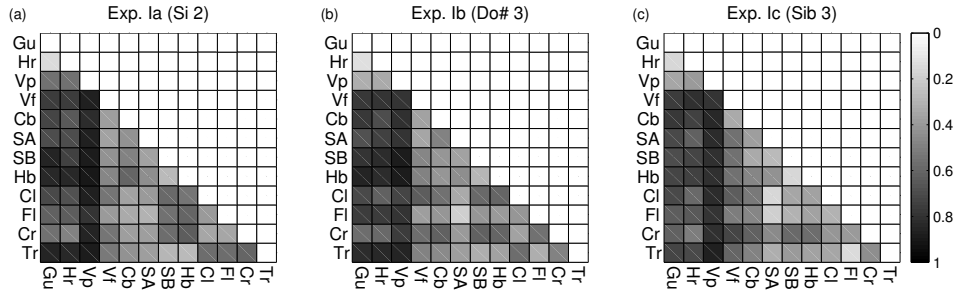


FIG. 4.2: Matrices de dissemblance pour les trois sessions de l'expérience I. Les commentaires sont les mêmes que pour la figure 4.1.

de la discussion et, d'autre part, de quantifier la grandeur des effets qui restent. Les résultats sont résumés dans le tableau 4.1. Tous les effets sont significatifs.

TAB. 4.1: Tableau ANOVA pour Exp. I. Les abréviations sont les suivantes : *S* pour Sujets, F_0 pour Fréquence fondamentale, *PI* pour Paires d'Instruments, d.d.l. pour degré de liberté, *SC* pour Somme des Carrés, *CM* pour Carré Moyen, *F* pour le rapport de Fisher expliqué, ϵ pour la correction de Greenhouse-Geisser appliquée aux d.d.l., p.c. pour la probabilité critique et R^2 pour le pourcentage de variance induit par chaque effet (en additionnant aux R^2 du tableau les 48,5 % dus à la variance inter-sujets on obtient 100 %).

Source	d.d.l.	SC	CM	F	ϵ	p.c.	R^2
S	23	32,02	1,39				
F_0	2	0,93	0,46	5,72	0,90	0,008	0,3
$F_0 \times S$	46	3,73	0,08				
PI	65	167,65	2,58	53,97	0,10	0,0001	46,6
$PI \times S$	1495	71,45	0,05				
$F_0 \times PI$	130	16,53	0,13	5,65	0,122	0,0001	4,6
$F_0 \times PI \times S$	2990	67,29	0,02				

Le pourcentage de la variance induite par chaque effet est exprimé par le coefficient R^2 (Wonnacott et Wonnacott, 1990). L'effet principal des paires d'instruments, représentant la part de dissemblance inter-instruments qui reste fixe à

travers les F_0 , est d'environ 47 % de la variance. La somme de l'effet principal F_0 et de l'interaction, représentant la part de dissemblance qui varie à travers les F_0 , contient moins de 5 % de la variance. Les différences inter-instruments sont donc bien plus importantes que les différences intra-instruments dues à la F_0 .

La grande proportion de variance inter-instruments reflète en partie la forte dissemblance entre impulsifs et entretenus. Il est donc intéressant d'examiner les effets à l'intérieur de trois sous-groupes : impulsifs, entretenus, et mixtes. Des ANOVA ont été appliquées à chacun de ces sous-groupes. Le Tableau 4.2 indique la probabilité critique pour chaque effet et sous-groupe, et le Tableau 4.3 récapitule les R^2 des sources de variance. L'effet principal des paires d'instruments et de l'interaction sont significatifs. L'effet principal de la F_0 est marginal ou non significatif. A l'intérieur de chaque sous-groupe, la variation de dissemblance avec F_0 (première ligne de Tab. 4.3) est plus faible que la dissemblance inter-instruments (2e ligne de Tab. 4.3). Cependant le rapport entre l'effet des paires d'instruments et les effets liés à la F_0 est plus faible pour chaque sous-groupe (2,6, 1,7, 4,5 pour impulsifs, entretenus et mixtes respectivement) que pour l'ANOVA globale (9,5). Toutefois, il reste relativement élevé.

TAB. 4.2: *Tableau résumant les résultats des ANOVA partielles avec la probabilité critique (p.c.) pour les facteurs F_0 , paires d'instrument (PI), et leur interaction.*

Source	p.c.			
	Tous	Impulsifs	Entretenus	Mixtes
F_0	0,008	0,0012	0,0804	0,0586
PI	0,0001	0,0001	0,0001	0,0001
$F_0 \times PI$	0,0001	0,0019	0,0001	0,0005

La section suivante présente les analyses MDS qui permettent de passer des dissemblances entre timbres à une représentation spatiale des timbres

4.2.4 Analyses multidimensionnelle de proximité

Pour chaque session une solution MDS de type MDSCAL a été calculée à partir de la matrice de dissemblance moyennée sur tous les sujets. Le nombre de stimuli étant relativement faible (12), le choix du nombre de dimensions était restreint à une ou deux (Chap. B, équation B.9). Puis, par une mesure de stress, le

TAB. 4.3: Pourcentage de variance induit par chaque facteur de l'Exp. I pour chaque groupe d'instruments. La première ligne (relative à F_0) est la somme des R^2 dus à l'effet principal F_0 et son interaction avec les paires d'instruments. La troisième ligne représente la variance due aux différences inter-sujets

Source	R^2 (%)			
	Tous	Impulsifs	Entretenus	Mixtes
F_0 et interactions	4,9	10,2	9,5	3,8
PI	46,6	26,2	16,4	17,1
sujets et interactions	48,5	63,6	74,1	79,1
total	100	100	100	100

choix s'est arrêté sur une représentation à deux dimensions. L'algorithme MDS ne fixant pas d'orientation particulière, une configuration physique de référence a été créée afin de servir de cible à leur rotation (voir chapitre 3). Nous nous sommes basés sur l'espace de timbre de McAdams et coll. (1995) qui montrait des dimensions perceptives corrélées avec le logarithme de temps d'attaque et le centre de gravité spectral. Nous avons cependant utilisé des descripteurs légèrement différents (impulsivité et centroïde spectral), voir chapitre 7. Le même espace de référence, dérivé des descripteurs appliqués aux stimuli de la session 'a', a été utilisé pour les trois sessions. La figure 4.3 montre la solution MDS de la session 'a'. Chaque stimulus est représenté par l'étiquette de son instrument (voir tableau 3.1). Les stimuli jugés similaires sont proches les uns des autres, comme par exemple le hautbois, la trompette et le SynthB. A l'opposé, les stimuli jugés dissemblables sont éloignés les uns par rapport aux autres. Les instruments impulsifs jugés similaires entre eux et dissemblables avec les autres forment ainsi un petit groupe isolé.

La figure 4.4 représente les solutions MDS des trois sessions.

Ces solutions représentent bien les données car leur stress est bas (0,11, 0,11 et 0,10) et le pourcentage de variance expliqué par le modèle est élevé (94 %, 94 % et 95 % pour les sessions 'a', 'b', et 'c', respectivement).

Les trois espaces ont un aspect similaire. Les corrélations entre les projections sur la première dimension sont élevées : 0,93 entre la session 'a' et 'b', 0,96 entre la session 'b' et 'c' et 0,93 entre la session 'a' et 'c'. Pour la deuxième dimension, les corrélations sont également élevées entre la session 'b' et 'c' (0,93) et plus

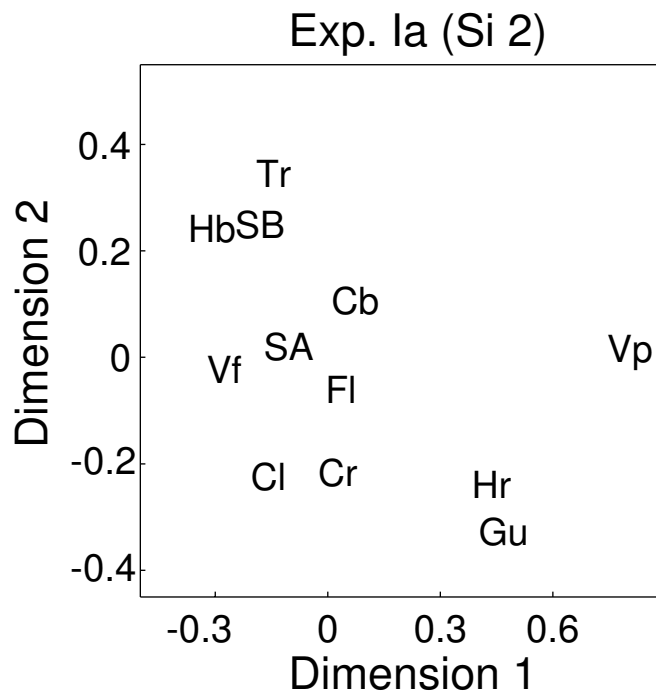


FIG. 4.3: Solution MDS pour l'Exp. I, pour la session 'a'. Pour le sens des étiquettes voir Tab. 3.1).

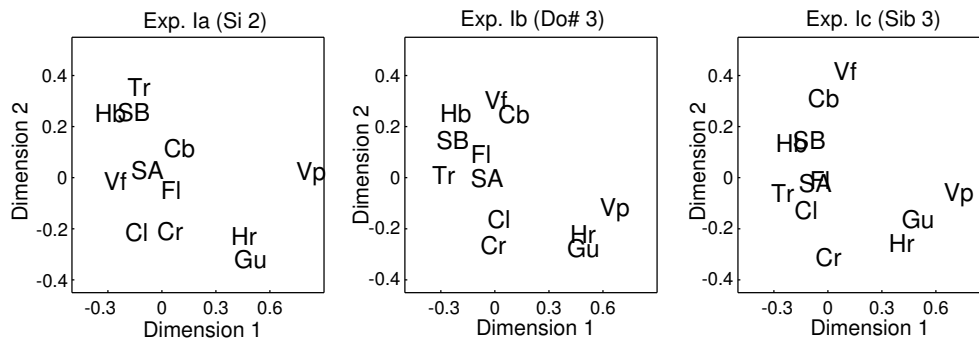


FIG. 4.4: Espace bidimensionnel de timbre de l'Exp. I, pour les sessions 'a' (gauche), 'b' (centre) et 'c' (droite). Voir la légende de la figure 4.3.

faibles entre les sessions 'a' et 'b' (0,67) et entre les sessions 'a' et 'c' (0,51). Les variations de timbre se font donc principalement selon la deuxième dimension, alignée sur le centroïde spectral.

Le chapitre 7 traitera des relations entre les dimensions perceptives et physiques.

4.3 Implications pour le timbre

Cette expérience est une réplique des études classiques (Wessel, 1973; Grey, 1977; McAdams et coll., 1995). Les solutions MDS obtenues sont cohérentes dans les grandes lignes avec celles obtenues dans ces études. Comme dans l'étude de McAdams et coll. (1995), les dimensions principales sont bien corrélées avec des descripteurs reflétant l'impulsivité de l'enveloppe temporelle et le centre de gravité de l'enveloppe spectrale (voir Ch. 7).

L'Exp. I apporte une nouveauté par rapport à ces études en répétant la tâche à trois F_0 avec les mêmes instruments et sujets, alors que les expériences de Wessel, Grey et McAdams se sont faites avec la même F_0 de 311 Hz. L'Exp. I permet donc de tester la généralité des solutions MDS et de mesurer leur stabilité lorsque la note varie.

Les résultats peuvent se décrire à deux niveaux d'approximation, la première donne les grandes lignes, la deuxième le détail. En *première approximation*, le timbre paraît stable. Les matrices de dissemblance (Fig. 4.2) et les solutions MDS (Fig. 4.4) sont semblables dans les grandes lignes. Les mesures de variance (R^2) issues des ANOVA montrent que la part de dissemblance qui reste stable (effet de paires d'instruments) est plus grande que la part qui varie (effet de F_0 et interaction). Les coefficients de corrélation entre sessions sont élevés, aussi bien pour les matrices de dissemblance que pour les solutions MDS. Cette stabilité est en partie le reflet de la dissemblance entre impulsifs et soutenus, qui a peu de raisons de changer avec la F_0 . Cependant, la conclusion reste globalement valide lorsqu'on examine séparément les paires d'impulsifs, soutenus ou mixtes.

En termes des hypothèses énoncées à la fin du Chapitre 2, les résultats permettent d'écarter l'hypothèse 3 (déplacements arbitraires des instruments dans l'espace des timbres). Ils sont compatibles avec l'hypothèse 1 (absence de déplacement) ou 2 (déplacement isométrique). Cette dernière ne peut pas être écartée car l'expérience n'a pas comparé les timbres d'une F_0 à l'autre. Rien ne permet donc d'écarter l'hypothèse que tous les timbres se sont déplacés en bloc (par translation ou rotation) dans l'espace des timbres. En outre, les sujets avaient pour

consigne d'utiliser l'ensemble de l'échelle de dissemblance pour chaque session. On ne peut donc pas non plus écarter l'hypothèse d'une contraction ou expansion de la configuration dans l'espace de timbre. L'Exp. II a pour but de lever ces ambiguïtés.

En *deuxième approximation*, des différences entre les matrices de dissemblance sont visibles. Les ANOVA indiquent que ces différences sont statistiquement significatives. Des différences entre les solutions MDS sont également apparentes. L'examen des coefficients de corrélation (élevés pour la première dimension, plus faibles pour la deuxième) indique que les variations concernent principalement la dimension qui s'aligne sur un descripteur spectral (le centroïde spectral), et moins la dimension qui s'aligne sur un descripteur temporel (l'impulsivité). En termes des hypothèses du Chapitre 2, on retient donc l'hypothèse 3 (déplacements arbitraires) en seconde approximation.

Sachant que les variations de timbre sont significatives, il serait souhaitable de les étudier de plus près, instrument par instrument. Les solutions MDS révèlent des déplacements d'instruments d'une F_0 à l'autre. Malheureusement, ces déplacements sont difficilement interprétables pour trois raisons : (a) la position d'un instrument dans la solution MDS est une fonction conjointe de tous les facteurs et ne reflète pas que cet instrument ; (b) nous n'avons pas de test de significativité des déplacements ; et (c) comme il vient d'être rappelé, les solutions MDS sont définies à une translation, rotation et homothétie près. Les Exp. II et III permettront une analyse plus fine des changements de timbre d'instruments particuliers.

4.4 Conclusions

Chaque session de cette expérience, prise séparément, réplique les expériences classiques de dissemblance de timbre (Wessel, 1973; Grey, 1977 ou McAdams et coll., 1995). Les résultats obtenus ici sont globalement cohérents avec ces études. Notre expérience permet en outre de mesurer les changements de timbre d'une F_0 à l'autre. En première approximation, les timbres semblent varier assez peu d'une F_0 à l'autre, mais l'expérience ne permet pas d'écarter l'hypothèse d'une variation "en bloc", dont les effets n'apparaîtraient pas dans les analyses. En deuxième approximation, on note des variations significatives du timbre de certains instruments, mais l'expérience ne permet pas de dire lesquelles, ni de les qualifier en détail. Les expériences suivantes lèveront certaines de ces ambiguïtés. Pour commencer, l'Exp. II étend l'Exp. I (avec les mêmes sujets et instruments) en comparant les timbres entre instruments avec des F_0 différentes.

Chapitre 5

Expérience II. Dissemblance de timbre à F_0 inégale

L'expérience II était une réplique de l'Exp. I avec une différence de F_0 (ΔF_0) fixe entre les instruments de chaque paire : 2 demi-tons dans la session 'a', et 11 dans la session 'b'. Les sujets, instruments et F_0 utilisés étaient les mêmes que pour Exp. I. L'Exp. II nous a donc permis d'étudier les effets de l'introduction d'une différence de F_0 entre deux sons sur le jugement de leur dissemblance de timbre. Si ces effets ne sont pas trop perturbant, la comparaison de leur timbre restera possible. Ceci a permis de tester l'hypothèse d'une dérive systématique du timbre avec la F_0 non testée lors de l'Exp. I (hypothèse 2, déplacement isométrique). En entreprenant cette expérience nous n'étions pas certains de son succès. Miller et Carterette (1975) ont déjà effectué une expérience de dissemblance entre des paires non-appariées en F_0 . Ils ont trouvé que les sujets basaient leur jugement en grande partie sur la différence de hauteur, et négligeaient ainsi les différences de timbre plus subtiles. Notre expérience se différencie de la leur par la consigne. Les sujets de Miller et Carterette devaient juger de la dissemblance sans consigne particulière, alors que les nôtres avaient pour consigne de ne pas tenir compte de la hauteur et de se concentrer sur le timbre. Toutefois, nous ne savions pas avant l'expérience si cette consigne était réalisable par les sujets.

Les instruments, les F_0 et les sujets étaient les mêmes que dans l'Exp. I. Cette expérience a donc permis de mesurer l'impact de la différence de F_0 dans les jugements de dissemblance de timbre.

5.1 Méthodes

Cette expérience se divisait en deux sessions nommées 'a' et 'b'. Dans chaque session, les sujets avaient pour consigne de juger une dissemblance de timbre (sans tenir compte de la différence de hauteur) entre deux sons ayant une différence de F_0 constante. Les stimuli utilisés dans la session 'a' étaient ceux des sessions 'a' et 'b' de l'Exp. I (avec des F_0 à si_2 et $do\#_3$). Les stimuli utilisés dans la session 'b' étaient ceux des sessions 'a' et 'c' de l'Exp. I (avec des F_0 à si_2 et $si\flat_3$). Les paires de stimuli présentaient une différence constante de F_0 de deux demi-tons (une seconde majeure) pour la session 'a', et de 11 demi-tons (une septième majeure) pour la session 'b'. Toutes les combinaisons de F_0 et d'instruments étant testées, chaque session contenait un total de 144 paires. A l'intérieur d'une même session, l'ordre des F_0 (grave-aiguë ou aiguë-grave) était toujours le même afin d'aider les sujets à ignorer la différence de hauteur. Les sujets étaient les mêmes que ceux de l'Exp. I et ont été divisés en quatre groupes différant par l'ordre des sessions ("ab" versus "ba") et par la présentation des F_0 dans chaque session (grave en premier versus aiguë en premier). La proportion de musiciens et de non musiciens était approximativement la même dans chaque groupe. Les sujets ont passé les deux sessions dans la même journée (une semaine après l'Exp. I), avec une pause de 10 minutes entre les sessions. L'expérience a duré environ 2 heures.

5.2 Résultats et analyses

5.2.1 Effets de sujet, ordre et expertise instrumentale

Les trois sujets exclus des analyses de l'Exp. I ont montré une faible corrélation avec le reste de sujets. Leurs données ont donc été également retirées pour le reste de ces analyses.

Des ANOVA préliminaires n'ont pas révélé d'effet significatif de l'expertise musicale, de l'ordre de passation des sessions, ni de l'ordre des F_0 . Les quatre groupes de sujets ont été réunis.

5.2.2 Matrices de dissemblance

La figure 5.1 représente la matrice de dissemblance moyennée sur les sujets, pour la session 'a'.

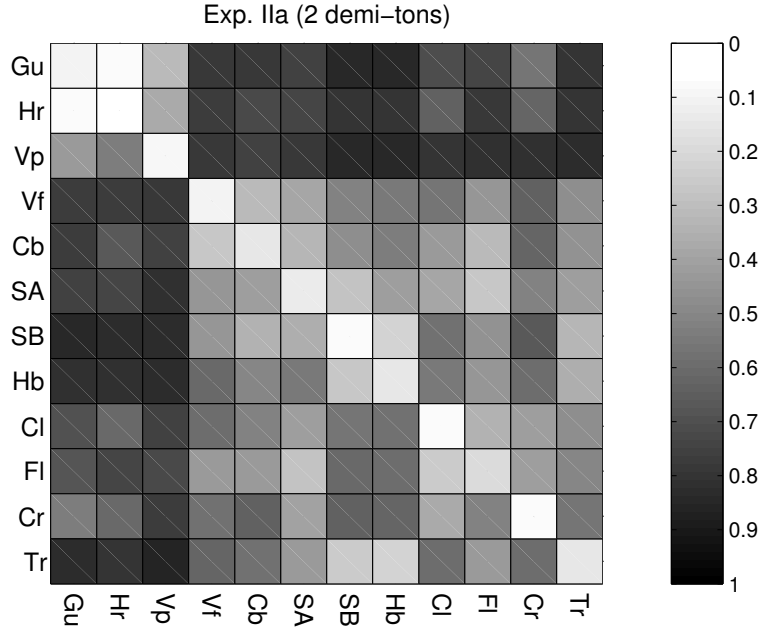


FIG. 5.1: Matrice de dissemblance pour la session 'a' de l'expérience II. Les instruments en abscisse possèdent une F_0 à si_2 , et ceux en ordonnée une F_0 à $do\#_3$. La diagonale contient les paires présentant deux fois le même instrument à des F_0 différentes.

Contrairement à la matrice de la figure 4.1, celle-ci est complète car toutes les paires ont été testées. Les stimuli en abscisse possèdent une F_0 à si_2 , et ceux en ordonnée une F_0 à $do\#_3$. Chaque intersection représente donc la dissemblance entre deux stimuli à deux F_0 différentes. La diagonale contient les dissemblances entre les paires constituées du même instrument. Cette figure montre trois aspects importants :

Premièrement, les valeurs le long de la diagonale sont particulièrement faibles. Une analyse de variance sur cette diagonale (ainsi que celle de la session 'b') est exposée dans la section 5.2.3.

Deuxièmement, la matrice est symétrique : la dissemblance entre un instrument X à si_2 et un instrument Y à $do\#_3$ diffère peu de la dissemblance entre Y à si_2 et X à $do\#_3$.

Troisièmement, le triangle inférieur de la matrice est très semblable à celui de l'Exp. Ia, figure 4.1.

Une matrice de dissemblance moyenne a également été calculée pour la session 'b'. La figure 5.2 représente les deux matrices de l'expérience.

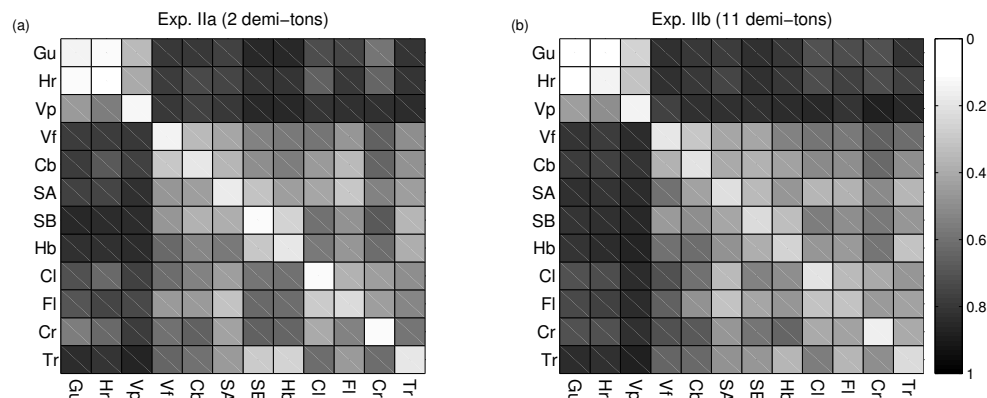


FIG. 5.2: Matrices de dissemblance pour les deux sessions de l'expérience II. Le panneau de gauche représente la matrice de la session 'a', avec en abscisse des instruments joués à si_2 , et en ordonnée des instruments joués à $do\#_3$. Le panneau de droite représente la matrice de la session 'b' avec, en abscisse, des instruments à si_2 et en ordonnée des instruments à $si\flat_3$.

Les deux matrices moyennes sont, à quelques nuances près, identiques. La matrice Ib contient notamment les trois aspects évoqués plus haut (diagonale faible, symétrie et similitude avec les matrices de l'Exp. I). Les deux matrices sont corrélées avec un coefficient de 0,95 (avec $d.d.l. = 142$, et $p < 0,001$). On peut cependant noter des valeurs plus sombres sur la diagonale de 'b' que sur la diagonale de 'a'.

Pour chaque matrice (moyennée sur les sujets), le triangle supérieur a été réfléchi par rapport à la diagonale et moyenné avec le triangle inférieur, et la moyenne comparée avec les matrices de l'Exp. I. La session 'a' de l'Exp. II est bien corrélée avec les sessions 'a' et 'b' de l'Exp. I avec des coefficients de 0,93 et 0,95 respectivement (avec $d.d.l. = 64$, et $p < 0,001$ pour les deux corrélations). La session 'b' de l'Exp. II est également bien corrélée avec les sessions 'a' et 'c' de l'Exp. I avec des coefficients de 0,86 et 0,94, respectivement (avec $d.d.l. = 64$, et $p < 0,001$ pour les deux corrélations).

5.2.3 Analyses de variance

Les triangles supérieurs des matrices de dissemblance de chaque sujet ont été réfléchis par rapport à la diagonale et soumis, avec les triangles inférieurs, à une ANOVA à mesures répétées avec les facteurs paires d'instruments (66), ordre de F_0 , triangle supérieur vs inférieur, (2) et ΔF_0 (2). Les résultats sont résumés dans le tableau 5.1. Le facteur d'ordre F_0 (triangle du bas vs. triangle du haut) n'est pas interprétable en lui-même car il dépend de la convention qui régit le placement des instruments dans le tableau. Cependant, il est inclus dans les analyses afin de quantifier l'ensemble des effets dus aux facteurs dépendants de la F_0 .

TAB. 5.1: Tableau d'ANOVA pour l'Exp. II. *S* : Sujets, ΔF_0 : différence de fréquence fondamentale, *O* : Ordre de présentation de F_0 , *PI* : Paires d'Instruments, d.d.l : degré de liberté. *SC* : Somme des Carrés, *CM* : Carré Moyen, *F* : rapport de Fisher expliqué, ϵ : correction de Greenhouse-Geisser appliqué aux d.d.l, p.c. : probabilité critique, R^2 : pourcentage de variance induite par chaque effet.

Source	d.d.l.	SS	MS	F	ϵ	p.c.	R^2
S	23	54,5	2,37				
ΔF_0	1	1,01	1,01	5,31	1	0,03	0,2
$\Delta F_0 \times S$	23	4,37	0,19				
O	1	0,69	0,69	10,48	1	0,003	0,1
$O \times S$	23	1,52	0,07				
PI	65	221,4	3,41	51,5	0,08	0,0001	43,8
$PI \times S$	1495	98,9	0,07				
$\Delta F_0 \times O$	1	0,23	0,23	7,54	1	0,012	0,04
$\Delta F_0 \times O \times S$	23	0,68	0,03				
$\Delta F_0 \times PI$	65	5,58	0,09	3,62	0,21	0,0001	1,1
$\Delta F_0 \times PI \times S$	1495	35,45	0,02				
$O \times PI$	65	4,14	0,06	2,3	0,2	0,006	0,8
$O \times PI \times S$	1495	41,33	0,03				
$\Delta F_0 \times O \times PI$	65	2,47	0,04	1,71	0,19	0,06	0,5
$\Delta F_0 \times O \times PI \times S$	1495	33,28	0,02				

Tous les effets principaux sont significatifs ainsi que leur interaction deux à deux. Seule leur interaction à trois n'est pas significative. Pour estimer l'ampleur

de chacun de ces effets, chaque facteur est quantifié par le coefficient R^2 (dernière colonne du tableau 5.1). L'effet dû aux paires d'instruments représente 43,8% de la variance totale alors que les effets dus à la F_0 totalisent 2,8%. Comme dans l'Exp. I, il apparaît que les dissemblances de timbre ne dépendent de la F_0 qu'à un degré limité.

Les ANOVA ont été répétées pour les sous-matrices correspondant aux paires d'instruments impulsifs, entretenus ou mixtes. Le tableau 5.2 indique le pourcentage de la variance attribué à chaque effet pour la totalité des données et, quand les dissemblances sont restreintes aux paires contenant uniquement des instruments impulsifs, entretenus ou mixtes (avec un impulsif et un entretenu).

TAB. 5.2: Pourcentage de variance induit pour chaque facteur séparément et pour chaque sous-groupe. La première ligne (relative à F_0) est la somme des R^2 due à l'effet principal de la ΔF_0 , de l'ordre de présentation et de leurs interactions avec les paires. L'avant-dernière ligne représente la variance due aux différences inter-sujets.

Source	R^2 (%)			
	Tous	Impulsifs	Entretenus	Mixtes
ΔF_0 et interactions	2,8	6,3	4,6	2,8
PI	43,8	29,2	15,2	9,1
sujets et interactions	53,4	64,5	80,2	88,1
total	100	100	100	100

Le rapport entre la variance associée aux paires d'instruments (part de la dissemblance inter-instruments qui ne change pas avec la F_0) et la variance associée à la F_0 (l'effet principal et son interaction avec les paires d'instruments représentant la part de dissemblance qui change avec la F_0) est plus faible à l'intérieur des sous-groupes, en particulier pour les paires d'instruments entretenus. Cependant, il reste largement supérieur à 1.

Des analyses supplémentaires ont également été effectuées sur les valeurs des diagonales. Une ANOVA avec les facteurs instruments (12) et ΔF_0 (2), montre un effet significatif de ces facteurs [$F(11, 253) = 5,3$, $\varepsilon = 0,43$, $p = 0,0001$ et $F(1, 23) = 16,1$, $p = 0,0005$ respectivement], mais pas de leur interaction. Ces résultats suggèrent que les différences de F_0 affectent les jugements de dissem-

blance et que cet effet augmente avec la ΔF_0 , tout en étant indépendant de l'instrument. Les valeurs moyennes sur les diagonales sont de 0,11 et 0,20 pour la session 'a' et 'b' respectivement.

Les valeurs le long de la diagonale sont faibles comparées aux autres. Elles possèdent une dissemblance moyenne de 0,16. La plus grande dissemblance pour une paire contenant deux fois le même instrument (la flûte avec 0,25) est plus petite que toutes les autres dissemblances entre les paires contenant deux instruments différents (à l'exception de la paire Gu-Hr avec un score de 0,1).

5.2.4 Analyses multidimensionnelles de proximité

Les matrices de dissemblance de Exp. II ne supportent pas une analyse MDS car elles n'incluent pas toutes les comparaisons entre stimuli (manquent les comparaisons à F_0 fixe). L'analyse serait techniquement possible si on l'appliquait aux sous-matrices triangulaires (ou à leur moyenne), mais elle nous apprendrait peu par rapport aux analyses déjà faites dans l'Exp. I ou qui seront faites dans l'Exp. III. L'analyse serait aussi possible appliquée à une matrice composite, assemblée par exemple à partir des matrices triangulaires des Exp. Ia et Ib et la matrice carrée de l'Exp. IIa. Cependant, les différences entre expériences (degré d'entraînement des sujets, jeu de stimuli) rendraient la valeur d'une telle analyse critiquable. Aucune analyse MDS n'a donc été effectuée sur les données de l'Exp II.

5.3 Implications pour le timbre

La forme de la matrice de dissemblance permet de faire des inférences concernant les changements de timbre d'une F_0 à l'autre. En particulier, elle permet de choisir entre les deux hypothèses, invariance et déplacement isométrique, que l'Exp. I ne permettait pas de départager. Dans l'hypothèse d'invariance, la matrice devrait avoir trois propriétés. Premièrement, les valeurs le long de sa diagonale devraient être nulles. Deuxièmement, la partie délimitée par le triangle inférieur de la matrice devrait être similaire à celles observées lors de l'Exp. I. Troisièmement, la matrice devrait être symétrique par rapport à cette diagonale.

La première propriété sera vraie si les sujets jugent que la dissemblance entre deux stimuli issus du même instrument est faible. Toutefois, les valeurs non-nulles observées peuvent être attribuées à un effet de bord de l'échelle de réponses limitée à zéro. Les sujets ne pouvant donner des réponses négatives, la variabilité des réponses implique une moyenne non-nulle. Cet effet est malheureusement délicat

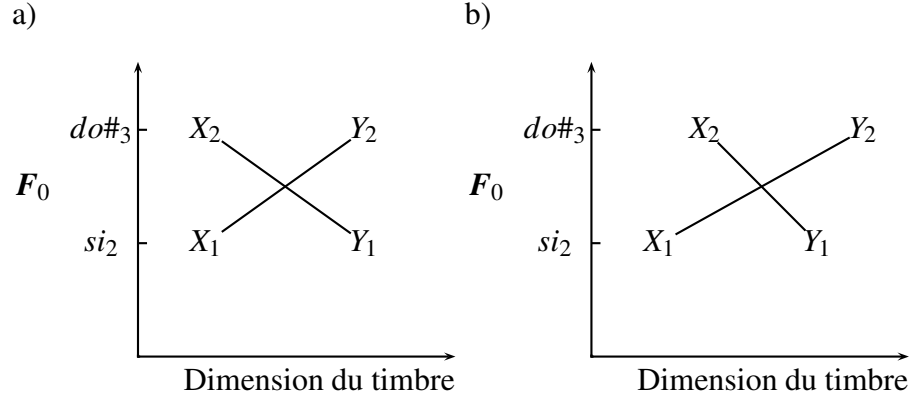


FIG. 5.3: Illustration de l'effet hypothétique de la F_0 le long d'une dimension du timbre (en abscisse) sur deux instruments X et Y . Les instruments sont représentés à deux F_0 différentes par X_1 , Y_1 et X_2 , Y_2 , respectivement. Le panneau de gauche illustre l'hypothèse 1 (invariance) et celui de droite l'hypothèse 2 (variation isométrique) qui implique $\overline{X_1Y_2} \neq \overline{X_2Y_1}$.

à quantifier. Cependant, comme nous l'avons vu les valeurs le long des diagonales sont faibles comparées aux autres.

La deuxième propriété sera vraie si les sujets jugent aussi dissemblables deux paires composées des mêmes instruments, indifféremment de la ΔF_0 qu'elles revêtent.

Afin de comprendre la deuxième propriété, considérons deux instruments (X et Y) différant le long d'une dimension du timbre (représentée en abscisse dans la figure 5.3). Les positions de ces deux instruments le long de cette dimension sont représentées à deux F_0 différentes par X_1 , Y_1 et X_2 , Y_2 , respectivement. L'Exp. I nous apprend que les distances $\overline{X_1Y_1}$ et $\overline{X_2Y_2}$ sont approximativement équivalentes. Si, de surcroît, les timbres des instruments restent invariant sur cette dimension avec un changement de F_0 (hypothèse 1), alors la distance $\overline{X_1Y_2}$ devra être équivalente à la distance $\overline{X_2Y_1}$ [figure 5.3.(a)]. Si, par contre, le timbre des instruments se déplace uniformément le long de cette dimension (hypothèse 2), alors les deux distances $\overline{X_1Y_2}$, $\overline{X_2Y_1}$ seront différentes [figure 5.3.(b)]. L'égalité de ces distances implique donc, soit un déplacement du timbre nul, soit un déplacement orthogo-

nal à la dimension séparant X et Y . Supposons que cette égalité reste vraie pour toutes les distances entre les instruments, cela implique que le timbre ne s'est pas déplacé sur les dimensions occupées par le corpus expérimental. La symétrie, si elle est observée, implique donc l'invariance du timbre face à la F_0 (hypothèse 1).

Les analyses ont révélé que les valeurs des diagonales étaient faibles, les matrices symétriques et corrélées avec celles de l'Exp. I. Nous pouvons donc en conclure qu'il n'y a pas eu de déplacement isométrique entre les timbres des instruments représentés dans les solutions MDS des trois sessions de l'Exp. I.

5.4 Conclusions

Cette expérience montre que les sujets sont capables de juger une dissemblance de timbre entre des sons non-appariés en F_0 . Malgré la différence de hauteur, ils ont effectué la tâche de façon similaire que dans l'Exp. I. Cette expérience montre également que l'hypothèse d'un déplacement isométrique de l'espace de timbre peut être réfutée, car les matrices se sont révélées symétriques et les diagonales petites. Ce résultat affine les conclusions de l'Exp. I en montrant que les instruments gardent relativement bien leur position dans l'espace de timbre lorsque la F_0 change. Enfin, cette expérience montre, également en accord avec l'Exp. I, qu'un effet significatif de la F_0 apparaît en deuxième approximation. Toutefois, cet effet reste faible.

Le fait que les sujets soient capables de juger des timbres en ignorant une différence de F_0 ouvre une perspective particulièrement intéressante pour ce type d'études.

Les matrices obtenues dans Exp. II sont incomplètes et ne permettent pas de construire des solutions MDS avec les positions des instruments aux différents F_0 . L'expérience III présentera aux sujets un ensemble complet de paires (pour un ensemble plus petit d'instruments), ce qui rendra possible cette analyse. Elle permettra de dire si les sujets, qui dans Exp. II ont réussi à ignorer une différence de hauteur fixe, pourront encore ignorer cette différence de hauteur si elle est variable. Enfin l'Exp. III nous permettra d'appliquer une méthode d'analyse originale, la méthode des "ancres" (Chapitre 3), pour analyser les changements de timbre instrument par instrument.

Chapitre 6

Expérience III. Dissemblance de timbre à différence de F_0 variable

L'Exp. I mesurait les dissemblances de timbre à F_0 fixe et l'Exp. II des dissemblances de timbres à *différence de F_0* (ΔF_0) fixe. Dans l'expérience III, les sujets ont jugé les dissemblances de timbre entre des paires d'instruments, avec et sans ΔF_0 . Son but était d'étendre et de généraliser les résultats des Exp. I et II, et en particulier de voir si les sujets étaient capables d'effectuer un jugement de timbre cohérent entre des stimuli avec une différence de hauteur qui variait au cours de l'expérience. La différence de F_0 variable aurait pu augmenter la saillance des différences de hauteur et ainsi perturber les jugements de timbre. Toutes les paires de sons étaient comparées, il était donc possible d'appliquer une analyse MDS pour étudier les déplacements dans l'espace de timbre (ce que ne permettait pas Exp. II), ainsi qu'une nouvelle méthode (celle des "ancres") permettant d'estimer si certains de ces déplacements sont statistiquement significatifs. Enfin, elle a fourni une base expérimentale pour le développement des descripteurs (chapitre 7).

6.1 Méthodes

Afin que l'expérience garde une durée raisonnable, neuf instruments ont été sélectionnés parmi les douze que comportaient les expériences précédentes. Ces instruments sont : Gu, Hr, Vp, Vf, SA, Hb, Cl, Cr, et Tr. Les trois F_0 utilisées étant identiques à celles des expériences précédentes. Le corpus de sons comportait donc 27 stimuli. L'expérience se divisait en deux sessions : 'a' et 'b'. Chaque session utilisait les neuf instruments à deux F_0 différentes : le *si*₂ et *do*_{#3} pour la

session 'a', et le si_2 et $si\flat_3$ pour la session 'b'. Elles utilisaient ainsi les mêmes ΔF_0 que celles des sessions 'a' et 'b' de l'Exp. II. La consigne restait identique à celle de l'Exp. II. Chaque session était composée de 153 paires. Les sujets avaient la possibilité de prendre une pause de 10 minutes au milieu. Chaque session durait environ une heure.

Le groupe de sujets de la session 'a' était composé de 25 personnes âgées de 19 à 30 ans, dont 15 hommes et 10 femmes, 13 musiciens et 12 non musiciens. Aucun d'eux n'avait participé aux expériences précédentes. Le groupe de sujets de la session 'b' était composé de 18 personnes âgées de 19 à 30 ans, dont 10 hommes et 8 femmes, 9 musiciens et 9 non musiciens, 11 d'entre eux ont participé à la session 'a'.

6.2 Résultats et analyses

6.2.1 Effets de sujet et expertise instrumentale

La cohérence entre les sujets a été testée à l'aide d'une analyse en cluster avec une méthode 'moyenne'. L'analyse a révélé que, pour la session 'a', trois sujets étaient faiblement corrélés avec les autres. Ils ont donc été exclus des analyses. Aucun sujet de la session 'b' n'a été exclu. Pour les deux sessions, des ANOVA n'ont révélé aucun effet significatif de l'expertise musicale.

6.2.2 Matrices de dissemblance

Les scores de dissemblance sont placés dans le triangle inférieur d'une matrice. La figure 6.1 représente la matrice de la session 'a'.

Cette matrice est constituée de trois parties : le triangle en haut à gauche contient toutes les paires d'instruments à si_2 ; le triangle en bas à droite, les paires d'instruments à $do\#_3$ (ou à $si\flat_3$ dans la session 'b') ; et le carré en bas à gauche, toutes les paires d'instruments ayant des F_0 différentes. Si l'on considère en plus que le carré du bas est constitué également de deux triangles symétriques par rapport à la diagonale, on obtient au total quatre triangles de même forme. Ils sont d'une grande similitude. Leur corrélation deux à deux est toujours supérieure à 0,95 ($df = 34$, $p < 0,0001$). Ce résultat montre que la différence de F_0 variable n'a pas perturbé les jugements de dissemblance de timbres. La figure 6.2 représente les matrices des deux sessions.

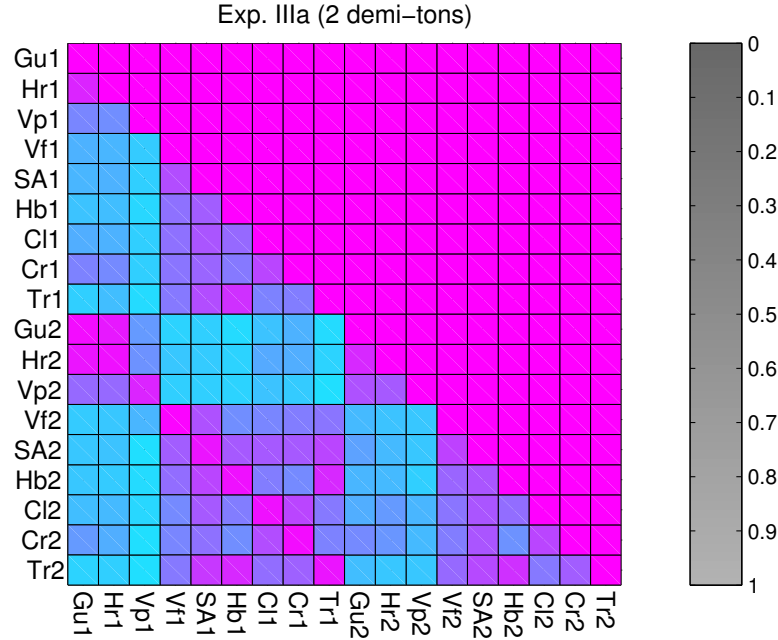


FIG. 6.1: Matrices de dissemblance de la session 'a' de l'expérience III. Sur les étiquettes des axes, les indices 1 et 2 indiquent si l'instrument est joué à si_2 ou à $do\#_3$.

Cette figure montre que les deux matrices sont semblables. Elles sont corrélées avec un coefficient de 0,93 ($df = 151$, $p < 0,0001$). La matrice 'b' comporte les caractéristiques évoquées plus haut. En la décomposant également en quatre triangle, on observe que leur corrélation deux à deux est toujours supérieure à 0,85 ($df = 34$, $p < 0,0001$). Nous pouvons en conclure que même pour une ΔF_0 de 11 demi-tons, la F_0 ne perturbe pas profondément les jugements de dissemblance.

6.2.3 Comparaison avec Exp. I et II

La figure 6.2 souligne également de grandes similitudes avec les matrices des expériences précédentes dont on aurait enlevé les trois instruments non utilisés dans cette expérience : Cb, SB, et Fl. Les triangles contenant les paires sans ΔF_0 ressemblent aux matrices de l'Exp. I. Le carré contenant les paires avec une ΔF_0 ressemble aux matrices de l'Exp. II. Il possède notamment les trois aspects évo-

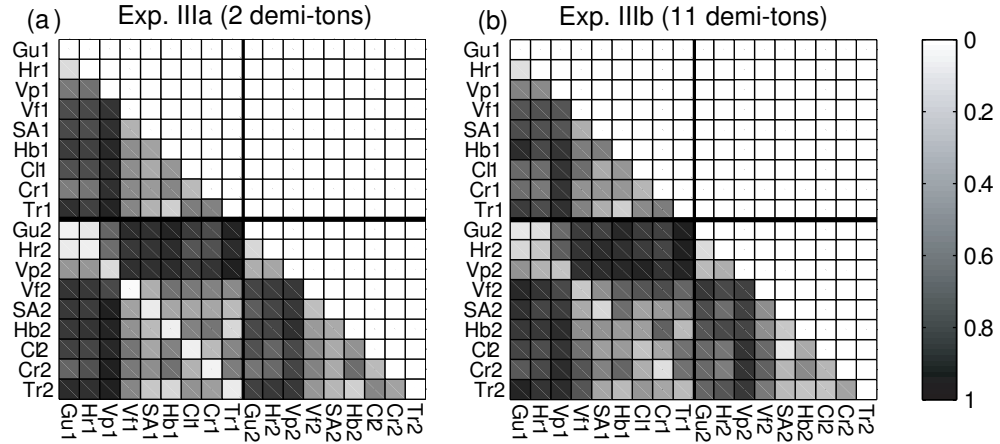


FIG. 6.2: Matrices de dissemblance pour les deux sessions de l'expérience III. Le panneau de gauche montre la matrice de la session 'a', les commentaires sont les mêmes que pour la figure 6.1. Le panneau de droite représente la matrice de la session 'b'. Les étiquettes 1 et 2 indiquent que le stimulus à une F_0 à si_2 ou si_3 , respectivement.

qués dans le chapitre 5 : la diagonale est claire, les deux triangles sont symétriques, chaque triangle est semblable à ceux obtenus avec des paires sans ΔF_0 .

Afin de quantifier ces ressemblances un coefficient de corrélation a été calculé entre les deux matrices de cette expérience et des matrices composées à partir des matrices de dissemblance des Exp. I et II (dont les trois instruments absents lors l'Exp. III ont été retirés). Les coefficients de corrélation sont élevés avec 0,95 entre la matrice de la session 'a' et la matrice composée des Exp. Ia, Ib et IIa, et 0,92 entre la matrice de la session 'b' et la matrice composée des Exp. Ia, Ic et IIb ($df = 151$; $p > 0,001$ pour les deux corrélations).

6.2.4 Analyses de variance globale

Comme pour les expériences précédentes une ANOVA a été effectuée pour chaque session. Pour cela, le carré a été divisé en deux triangles, dont celui du haut a été inversé, afin d'en obtenir quatre présentant des paires d'instruments identiques (mais à des F_0 différentes). Les quatre triangles représentent donc les quatre combinaisons de F_0 entre paires. Ils ont été soumis à cette ANOVA avec les facteurs F_0 , les triangles (4) et paires d'instruments (36). Le facteur principal F_0

était significatif [$F(3,63) = 16,39$, $p < 0,0001$ pour la session 'a' et $F(3,51) = 17,68$, $p < 0,0001$ pour la session 'b'], de même que pour les paires d'instruments [$F(35,735) = 62,33$, $p < 0,0001$, pour la session 'a' et $F(35,595) = 42,05$, $p < 0,0001$ pour la session 'b'], ainsi que les interactions [$F(105,2205) = 2,09$, $p < 0,0001$ pour la session 'a' et $F(105,785) = 2,95$, $p < 0,0001$ pour la session 'b']. Toutefois, les effets significatifs dus à la F_0 ont été tempérés par la mesure de R^2 . Celle-ci, comme pour les expériences précédentes, nous informe que les effets dus à la F_0 contribuent faiblement à la variance (2,5 % et 6,2 % pour la session 'a' et 'b', respectivement) contrairement à l'effet dû aux paires d'instruments (49 % et 36 % pour les sessions 'a' et 'b', respectivement). Comme pour les expériences précédentes, les timbres issus d'un même instrument étaient relativement stables malgré une différence de F_0 . Des effets existent cependant, mais ils sont faibles, et dépendent de l'instrument (interaction significative lors des ANOVA). La section suivante décrit un nouveau type d'analyses qui permet de mieux étudier les effets instrument par instrument.

6.2.5 Analyses de variance partielle : la méthode des ancres

Le déplacement de chacun des neuf instruments a été étudié par la méthode des ancres, décrite au chapitre 3. Afin de tester l'hypothèse d'un changement significatif du timbre de l'instrument, les deux termes de l'équation 3.1 sont soumises à une ANOVA à mesures répétées avec les facteurs ancres (8) et F_0 (2) (représentant les deux parties de l'équation). Pour chaque instrument, les huit autres faisaient office d'ancres. Neuf ANOVA, une par instrument, ont donc été effectuées.

La session 'a' Les R^2 des facteurs F_0 et interactions sont résumés dans le tableau 6.1. Seuls les effets significatifs sont montrés. L'effet principal des ancres, toujours significatif, n'est pas représenté dans le tableau. Pour la trompette, l'effet principal de la F_0 est significatif mais faible (une mesure de R^2 nous le montre). Pour le violon pizzicato, l'effet de F_0 et son interaction avec les ancres sont significatifs. Le timbre des sons issus de cet instrument change donc avec la F_0 . Pour tous les autres instruments aucun effet significatif dû à la F_0 et ses interactions avec les ancres n'a été trouvé.

La session 'b' Pour trois instruments (Gu, Hr et Cr) aucun effet significatif n'a été trouvé. L'effet de la F_0 sur la clarinette est significatif mais faible. Les effets

TAB. 6.1: Tableau récapitulatif des tailles des effets significatifs à 5% pour les facteurs de F_0 ainsi que leur interaction avec les ancrés.

	R^2 (%)			
	2 demitons		11 demitons	
Instrument :	F_0	$F_0 \times$ ancre	F_0	$F_0 \times$ ancre
Gu	-	-	-	-
Hr	-	-	-	-
Vp	2,79	2,51	-	4,28
Vf	-	-	1,17	-
SA	-	-	0,55	1,52
Hb	-	-	1,23	1,92
Cl	-	-	0,59	-
Cr	-	-	-	-
Tr	0,40	-	1,66	1,92

sont significatifs et conséquents pour le reste des instruments : Vp, Vf, SA, Hb et Tr.

Le nombre d'instruments dont le timbre change avec la F_0 grandit donc avec la ΔF_0 .

6.2.6 Analyses multidimensionnelles de proximité

Pour chaque session, une solution MDS a été extraite des données. Le choix du nombre de dimensions étant plus délicat à déterminer, en partie à cause du nombre de stimuli, les MDS ont été calculées avec le logiciel EXSCAL. La fonction du maximum de vraisemblance indiquait une solution à deux et trois dimensions sans spécificité pour les sessions 'a' et 'b', respectivement. Toutefois, l'étude de la solution à 4 dimensions de la session 'b' révèle une caractéristique particulièrement pertinente pour l'étude de la relation entre la F_0 et le timbre. Cette solution a donc été choisie. Afin d'appliquer le même traitement aux deux sessions, la même représentation à 4 dimensions a été utilisée pour la session 'a'. Chaque solution a ensuite subi une rotation procustéenne vers une configuration cible physique composée à partir des quatre (deux de plus que dans l'Exp.I) descripteurs suivants :

l'impulsivité, le centroïde spectral, l'étalement spectral et la F_0 . Ces descripteurs seront détaillés au chapitre 7.

Par souci de clarté, nous commencerons par commenter la solution de la session 'b' représentée figure 6.3. Chaque point représente un instrument à si_2 et l'autre extrémité de la ligne le même instrument à si_3 . Les points noirs représentent les instruments dont l'analyse, par la méthode des ancres a révélé une différence significative entre leurs timbres. Le déplacement observé pour ces instruments doit être considéré comme significatif. Pour les autres, ce déplacement peut être dû au bruit expérimental.

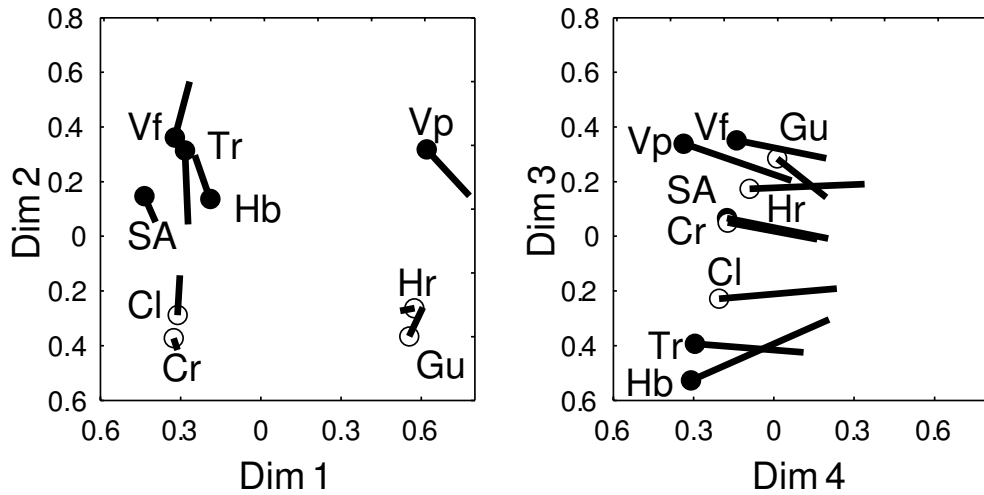


FIG. 6.3: Espace de timbre de l'Exp. III pour la session 'b'. Le panneau de gauche représente la première et la deuxième dimension, celui de droite la troisième et la quatrième. L'étiquette représente la position de l'instrument à si_2 et l'extrémité de la ligne la position du même instrument si_3 . Les points remplis indiquent les instruments dont les timbres changent de façon significative selon la méthode des ancres (voir chapitre 3).

Le panneau de gauche représente les deux premières dimensions. La première discrimine parfaitement les instruments impulsifs et entretenus. La deuxième range les instruments selon leur centroïde spectral. La longueur du trait noir représente la différence de position entre deux stimuli issus du même instrument. A l'exception de Cl, tous les instruments dont l'analyse par la méthode des ancres a révélé une différence significative, montrent une plus grande différence de position que

les instruments considérés comme stables.

Le panneau de droite représente les deux autres dimensions. La troisième dimension ordonne les stimuli selon leur valeur d'étalement spectral, décrit au chapitre 7. La quatrième dimension est particulièrement intéressante car elle se montre fortement corrélée à la F_0 . Ce résultat montre que les sujets ont systématiquement ajouté une certaine quantité à leur jugement de dissemblance pour des paires non appariées en F_0 . Toutefois, cette dimension étant orthogonale aux autres, liées au timbre, elle ne les influence pas directement.

La figure 6.4 représente la solution de la session 'a'.

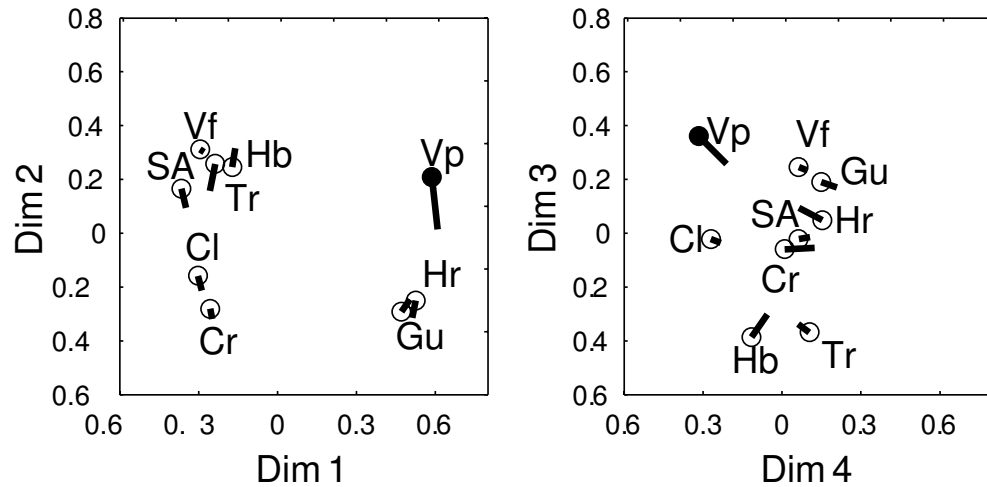


FIG. 6.4: Espaces de timbre de l'Exp. III pour la session 'a'. Les commentaires sont les mêmes que pour la figure 6.3.

Les différences de timbre inter-instruments sont ici beaucoup plus faibles. Seul Vp montre un déplacement conséquent. Ces résultats sont en accord avec les analyses par la méthode des ancres. Les trois premières dimensions des deux sessions sont fortement corrélées entre elles (0,99, 0,95, et 0,94 respectivement). La quatrième dimension montre une corrélation plus faible : 0,31.

Afin de pouvoir comparer cet espace avec ceux obtenus lors de l'Exp. I, les configurations de la session Ia et Ib ont été rassemblées après leur avoir enlevé les trois instruments non utilisés dans cette expérience. Cette nouvelle configuration "composite" a ensuite été utilisée pour effectuer des corrélations avec la configuration de la session IIIa. Les deux dimensions sont fortement corrélées (0,92 et 0,90 pour la première et deuxième dimension, respectivement).

Le même traitement a été appliqué aux solutions de la session Ia et Ic, pour les corrélérer avec la configuration de la session IIIb. Les deux dimensions sont également fortement corrélées (0,92 et 0,87 pour la première et deuxième dimension, respectivement). Ce résultat montre, comme celui de l'Exp. II, que malgré une différence de hauteur les sujets jugent le timbre de la même façon.

6.3 Implications pour le timbre

Comme dans les Exp. I et II, les ANOVA globales ont montré que les variations de timbre d'une F_0 à l'autre sont faibles mais significatives. L'Exp III a étendu ce résultat avec une ΔF_0 variable (et donc potentiellement plus saillante). Elle a aussi permis de partager les instruments entre ceux dont le timbre "bouge" et ceux dont il reste fixe, grâce à la méthode des ancres. L'absence apparente de tout changement de timbre pour certains instruments est intéressante. Elle suggère soit qu'il n'existe pas de dépendance systématique du timbre sur la F_0 , soit que ces instruments ont été conçus pour compenser cette dépendance.

La méthode des ancres permet de dire que le timbre d'un instrument particulier a changé, ce que ne permettaient pas les autres méthodes. L'analyse MDS permet ensuite de visualiser ce changement dans un espace de timbre.

Les solutions MDS retenues comportent quatre dimensions. Les deux premières correspondent à celles trouvées dans l'Exp. I. La troisième est bien corrélée avec un descripteur d'étalement spectral, Chapitre 7. La quatrième est corrélée dans la session 'b' seulement avec la F_0 . Cette quatrième dimension correspond probablement à la hauteur, que les sujets n'ont pas réussi à ignorer complètement. Si c'est le cas, seules les trois premières dimensions correspondent à la définition du timbre.

6.4 Conclusions

Les résultats de cette expérience sont tous cohérents avec ceux obtenus lors des deux premières expériences. Les patterns de réponses de toutes les matrices de dissemblance sont corrélés, ainsi que les solutions MDS. Ceci permet d'affirmer que les sujets sont capables d'ignorer une différence de hauteur, même variable, lorsqu'ils jugent une différence de timbre. Il serait tentant de conclure à la parfaite indépendance entre le timbre et la hauteur. Toutefois la quatrième dimension de l'espace IIIb, fortement corrélée avec la F_0 , nous oblige à nuancer cette conclu-

sion. La F_0 introduit, il est vrai, un effet léger lors des jugements de dissemblance de timbre.

En intégrant les résultats des trois premières expériences, nous pouvons conclure qu'en première approximation, le timbre est relativement stable face à la fréquence fondamentale. Mais, en deuxième approximation, on décèle de petits changements significatifs et non isométriques. L'analyse par la méthode des ancres permet de discriminer les instruments ayant un timbre qui change de ceux qui restent stables.

L'un des grands intérêts des études sur le timbre est de pouvoir concevoir des indicateurs physiques capables de prédire ou de décrire la position d'un stimulus sur une dimension perceptive donnée. Dans les études précédentes, les solutions MDS étaient constituées uniquement de stimuli appariés en F_0 . Ce facteur n'a donc jamais été intégré dans la conception de ces descripteurs. Cette expérience permet, avec un espace contenant des timbres non appariés en F_0 , d'ajouter des contraintes importantes amenant à une meilleure conception de ces descripteurs.

Si l'analyse par la méthode des ancres a estimé que les deux timbres d'un instrument n'étaient pas significativement différents, alors les descripteurs des deux stimuli devront garder cette même stabilité. Dans le cas contraire, les descripteurs devront prédire ces changements.

Le chapitre suivant étudie les différents descripteurs existant et explique comment les adapter à ces nouvelles contraintes.

Chapitre 7

Descripteurs de signal

7.1 Introduction

L'analyse multidimensionnelle de proximité permet de mettre en évidence certaines dimensions importantes du timbre. Dans un deuxième temps, il est intéressant d'essayer de mettre en correspondance objectivement chacune de ces dimensions par des mesures physiques. Ces mesures amènent à l'élaboration de descripteurs du signal. Chaque descripteur est une quantité objective issue de l'étude du signal acoustique. Il a pour vocation de pouvoir prédire la position d'un stimulus sur les dimensions d'un espace de timbre. Il en existe une infinité, mais seulement un nombre restreint est utile car les dimensions importantes du timbre semblent être en nombre réduit et beaucoup de descripteurs sont des variantes ou des combinaisons d'autres descripteurs. Ils permettent l'élaboration de la psychophysique du timbre, et ainsi de pouvoir définir le timbre, non plus par la "négative" (tout ce qui n'est pas la hauteur, la sonie...), mais par la "positive" (le timbre peut être prédit par l'impulsivité, le centroïde spectral...). Ces descripteurs sont très utilisés dans diverses applications comme l'indexation de bases de données ou les effets audionumériques. Ils ont également une grande valeur informative pour les compositeurs.

7.2 Les descripteurs d'impulsivité

7.2.1 Revue des descripteurs d'impulsivité et de temps d'attaque

Grey (1977) proposa, afin de décrire deux de ces dimensions perceptives, la synchronie entre les harmoniques aiguës et les harmoniques graves, ainsi que la présence de hautes fréquences à basse amplitude lors de l'attaque. Cependant, il ne donna pas plus de détails ni de méthodes afin de quantifier de façon objective ces deux attributs. Wessel (1978) lia la deuxième dimension de son espace à ce qu'il appela '*quality of the bite in the attack*'. Il la fit varier en changeant le temps d'attaque de ses sons. Krimphoff (1993) étudia en détail les moyens de calculer les descripteurs physiques des dimensions perceptives. Il proposa, pour la dimension liée à l'attaque, d'utiliser le "logarithme du temps d'attaque". Ce descripteur est corrélé avec un coefficient de -0,94 avec la deuxième dimension perceptive de l'espace de Krumhansl (1989). Il est également corrélé avec le même coefficient à la première dimension de l'espace de McAdams et coll. (1995). Il définit le logarithme du temps d'attaque (*lta*) ainsi :

$$lta = \log_{10}(t_{max} - t_{seuil}), \quad (7.1)$$

où t_{max} est l'instant où l'enveloppe temporelle du signal atteint sa valeur maximale et t_{seuil} l'instant où elle dépasse le seuil de 2 % du niveau maximum. Afin de calculer ces instants, une bonne estimation de l'enveloppe temporelle est nécessaire. L'auteur la définit en additionnant la valeur absolue de l'amplitude des 40 premières harmoniques.

Iverson et Krumhansl (1993) créèrent trois configurations MDS à partir d'un même corpus d'instruments. La première représentait les dissemblances entre les sons entiers ; la deuxième, les dissemblances entre les transitoires d'attaques ; et la troisième, les dissemblances entre les stimuli amputés de leur attaque. Étonnamment, les trois configurations étaient très similaires. Les auteurs en ont déduit que les attributs dynamiques du timbre ne sont pas uniquement présents durant l'attaque, mais également tout le long du signal, et que plusieurs attributs physiques peuvent contribuer à la même dimension perceptive.

En considérant cette étude, Marozeau et coll. (2003) reprirent une idée de Susini (1999) et montrèrent que pour des stimuli relativement longs (plus d'une seconde), le temps de décroissance influence également la dimension du timbre liée aux qualités temporelles du signal. Celle-ci, pouvant être prédite par une mesure

de l'impulsivité, se définit comme la durée durant laquelle l'enveloppe temporelle de la puissance du signal est au-dessus d'un seuil de 40 %, divisé par la durée durant laquelle cette enveloppe est supérieure au seuil de 10 %. Cette enveloppe est calculée en lissant le signal au carré par une convolution avec une fenêtre rectangulaire de 8 ms. Il est possible, pour améliorer l'efficacité du lissage, d'utiliser une fenêtre rectangulaire dont la longueur correspond précisément à la période fondamentale du signal. Toutefois, cela implique que le signal soit harmonique et que la détection de sa F_0 se fasse sans erreur. Afin que la valeur de l'impulsivité soit positivement corrélée avec les stimuli, le rapport calculé est soustrait de la valeur 1 (i.e. un son impulsif a une valeur proche de 1 et un son entretenu de 0).

Toutefois, une réserve peut être émise sur l'implémentation proposée par Marozeau et coll. (2003). Ce descripteur est le résultat d'un rapport. Un son qui subirait une importante dilatation temporelle parfaitement uniforme, conserverait une même valeur d'impulsivité, alors que le son sera perçu, en fonction de la magnitude de cette dilatation, comme étant de plus en plus entretenu. Afin de résoudre ce problème, l'impulsivité peut être considérée simplement comme la durée durant laquelle l'enveloppe temporelle de la puissance du signal est au-dessus d'un seuil de 40 %. L'unité de ce descripteur correspond par conséquent aux secondes. Cette définition présente juste un léger inconvénient car il associe aux sons très impulsifs une valeur faible et aux sons entretenus une valeur élevée. Un travail en profondeur sur la conception d'un bon descripteur temporel serait nécessaire. Mais cela dépasse le cadre de cette thèse.

7.2.2 Descripteur choisi

Le descripteur se basant sur le logarithme du temps d'attaque possède plusieurs inconvénients majeurs apparaissant lors de son implémentation sur des sons issus d'échantillons naturels :

1. Le maximum du son ne représente pas toujours l'instant final de l'attaque, comme certaines notes de violon par exemple ;
2. Le seuil de 2 % peut être facilement dépassé par un léger bruit parasite ;
3. Le calcul de l'enveloppe demande une estimation de la F_0 et de ses partielles, il est donc difficilement applicable aux sons non harmoniques ;
4. Depuis Stevens nous savons que la perception et la physique sont liées par une fonction de puissance et non logarithmique, cette dernière semble donc moins adaptée.

Le descripteur de l'impulsivité est plus robuste car il n'impose pas une estimation d'un temps d'attaque précis. Il est également plus général car il prend en considération l'attaque et la décroissance. Nous utiliserons donc dans ce document le descripteur d'impulsivité "simplifié" (ou le rapport n'est plus considéré mais seulement la durée au dessus du seuil).

7.2.3 Corrélations avec les solutions MDS

Le tableau 7.1 indique les corrélations entre la première dimension des solutions MDS de l'Exp I et III et les descripteurs de logarithme de temps d'attaque et d'impulsivité. Ces descripteurs ont également été testés avec la première dimension de l'espace de McAdams et coll. (1995). Cet espace, ayant été effectué avec un grand nombre de sujet et le logiciel CLASCAL, est orienté afin que ces dimensions présentent une validité psychologiques (Chapitre 3)..

TAB. 7.1: *Tableau récapitulatif de la valeur absolue des corrélations entre les descripteurs de logarithme de temps d'attaque et d'impulsivité avec les projections sur la première dimension des solutions de l'Exp. I, III. et McAdams et coll. (1995) (d.d.l. = 10, pour l'Exp. I et d.d.l. = 16 pour l'Exp. III et McAdams et coll. 1995) toutes les corrélations sont significatives ($p < 0.001$). Les corrélations sont données à une décimale près afin de se garder d'une interprétation trompeuse sur une précision non-significative. Les corrélations sont arrondies à la décimale la plus proche (p.ex. 0,96 $\rightarrow$ 1,0).*

Expériences	Logarithme du temps d'attaque	Impulsivité
Exp. Ia	0,8	0,9
Exp. Ib	0,9	0,9
Exp. Ic	0,8	0,9
Exp. IIa	1,0	1,0
Exp. IIb	1,0	1,0
McAdams (1995)	0,9	0,9

Les corrélations sont toutes élevées. L'impulsivité semble donner de meilleurs scores. Toutefois, le faible degré de liberté des données limite la précision des coefficients. L'impulsivité sera donc préférée à cause de sa robustesse et de sa facilité de calcul, évoquées dans le paragraphe précédent, et non sur des meilleurs corrélations.

7.3 Les descripteurs de centroïde spectral

7.3.1 Revue des descripteurs de centroïde spectral

La perception du timbre a souvent été comparée à la perception visuelle. Nous parlons volontiers de la couleur d'un son ou du "Tone Farbe" de Helmholtz. La brillance, attribut très usité en perception visuelle, permet de discerner les tons ternes des tons brillants. La psychoacoustique s'est rapidement approprié ce terme afin de désigner une qualité de timbre. Elle permet de distinguer des sons allant du mat au brillant. Cette qualité perceptive se retrouve dans toutes les expériences psychoacoustiques sur le timbre, quel que soit le type de sons employé, des voyelles aux sons musicaux en passant par des sons d'avertisseurs (Lemaitre, 2004), ou même de ventilateurs (Susini et coll., 2001). Cette qualité est directement liée à la forme de la distribution de l'énergie à l'intérieur du spectre de puissance.

Afin de quantifier la distribution de l'énergie à l'intérieur d'un spectre par des valeurs simples, il est possible de l'assimiler à une distribution statistique simple. Celle-ci peut être décrite par sa moyenne et sa variance. Ces deux valeurs appliquées à une distribution spectrale prennent les termes de "centre de gravité spectral" et "étalement spectral". Le premier tient une place importante en psychoacoustique depuis que Grey et Gordon (1978) ont montré qu'il décrivait bien la brillance d'un stimulus. La littérature présente plusieurs façons de le calculer. La façon la plus simple est la suivante :

$$\bar{f} = \frac{\sum_i f_i \times A_i}{\sum_i A_i}, \quad (7.2)$$

où f_i représente la fréquence de la i ème raie d'un spectre, A_i l'amplitude de la i ème raie. Cette mesure nous donnant une fréquence correspondant au centre de gravité de ce spectre, celle-ci est représentée par le symbole : $\bar{f}$. Nous dénommerons cette solution "CGS de base". Les variantes proposées se basent principalement sur cette équation mais en changeant les composantes. Par exemple Kendall et coll. (1999), proposent non plus de calculer le centre de gravité à partir de toutes les fréquences du spectre mais seulement à partir des harmoniques $\bar{k}$:

$$\bar{k} = \frac{\sum_k k \times A_k}{\sum_k A_k}, \quad (7.3)$$

où k représente la k ème harmonique, A_k l'amplitude de cette harmonique. Cette équation demande, afin d'être calculée, un algorithme complexe d'extraction de

partielle. Nous dénommerons cette solution "CGS de rang". Pour des sons harmoniques, les deux équations sont proportionnelles : $\bar{f} = F_o \times \bar{k}$. Les expériences de dissemblance "classique" de timbres, n'utilisant que des stimuli appariés en F_o , ne peuvent discriminer les résultats de deux équations parfaitement corrélées.

La pondération par l'amplitude des raies spectrales ou des harmoniques est loin du fonctionnement physiologique de l'oreille. Des pondérations par l'énergie ou le logarithme de l'amplitude ont également été proposées (Krimphoff et coll., 1994).

Le descripteur de Zwicker et Fastl

Zwicker et Fastl (1990) développèrent un nouveau descripteur dans le même esprit. Ce descripteur a pour particularité d'être basé sur des données physiologiques de l'audition. Son unité est l'acum et possède pour référence un bruit à largeur de bande de 1 bark centré à 1 kHz produisant une acuité de 1 acum. Il est basé sur le modèle de la sonie et les premières étapes sont identiques.

Le spectre du signal est découpé en 24 bandes, z , de largeur de 1 bark (voir Hartmann (1993) pour plus de détail). Le niveau à l'intérieur de chacune de ces bandes est ensuite utilisé pour calculer les sonies de coeurs. L'effet du filtrage de l'oreille externe et moyenne, ainsi que le bruit de fond de l'oreille interne sont compensés par des atténuations dans chacune de ces bandes en fonction de la fréquence. Le masquage fréquentiel est pris en considération par le calcul des sonies flancs. Ces différentes opérations aboutissent à l'évaluation de la sonie spécifique, $N'(z)$, (sonie en fonction de la fréquence sur une échelle de Barks). La sonie totale, N , correspond simplement à l'intégration de cette sonie spécifique : $N = \int_0^{24} N' dz$. L'acuité, A , se définit ainsi :

$$S = 0,11 \frac{\int_0^{24} N' \times g(z) \times z dz}{\int_0^{24} N' dz}, \quad (7.4)$$

où $g(z)$ est une fonction de pondération équivalente à 1 de 0 à 16 barks, puis qui augmente de façon exponentielle pour atteindre 4 à 24 barks. Cette fonction sert "*à prendre en compte le fait que l'acuité d'un bruit de faible bande croît fortement, de façon surprenante, pour des fréquences centrales élevées*" (Zwicker et Fastl, 1990).

Notons finalement que le nom "d'acuité" prête à confusion, car d'une part il fait fortement référence à une notion de hauteur, et de l'autre il est également utilisé comme la dénomination d'une dimension perceptive pouvant être assimilée

à la brillance. Ce même type de confusion a souvent été fait pour la sonie et l'algorithme qui la prédit.

Le descripteur de centroïde spectral

Nous utiliserons dans ce document un descripteur qui nous est propre. Ce descripteur est dans l'esprit de celui de Zwicker et Fastl, mais certaines améliorations inspirées du modèle de sonie de Moore et Glasberg (1996) y sont intégrées. Le signal est premièrement filtré par un filtre passe-bande coupant les basses et les hautes fréquences, imitant ainsi le filtrage effectué par l'oreille externe et moyenne (Killion, 1977). Puis il est filtré à travers un banc de filtres passe-bandes, nommés "gammatones", (Patterson et coll., 1995; Slaney, 1993) imitant la sélectivité fréquentielle de l'oreille. Chaque bande est espacée d'un demi-ERBrate ($0,5 \times z$) entre 25 Hz et 19 kHz (Hartmann, 1997). La puissance instantanée est ensuite calculée pour chaque bande. Elle est lissée en l'additionnant à elle-même retardée de $1/4f_c$ où f_c est la fréquence centrale d'une bande. Puis elle est convoluée par une fenêtre rectangulaire de 8 ms, correspondant en première approximation au temps d'intégration temporelle mesuré par Plack et Moore (1990). Ces signaux de puissance lissés sont ensuite mis à la puissance 0,3 afin de modéliser la sensation de sonie partielle dans chaque bande. Un centroïde spectral instantané $\bar{z}(t)$ est calculé ainsi :

$$\bar{z}(t) = \frac{\sum_z z \times \psi_z(t)}{\sum_z \psi_z(t)}, \quad (7.5)$$

où $\psi_z(t)$ représente la sonie partielle de la bande z à l'instant t . Enfin, le centroïde spectral instantané est pondéré par la sonie instantanée (somme à travers toutes les bandes des sonies partielles) et moyenné afin d'obtenir un descripteur unique $\bar{z}$ caractérisant le signal entier.

7.3.2 Choix du descripteur

Le choix du modèle a été effectué par des critères de robustesse, de simplicité, ainsi que de validation par rapport aux phénomènes physiologiques. Le centroïde spectral, par ses caractéristiques, remplit bien ces critères :

- La sommation de l'énergie à l'intérieur des bandes se fait à partir d'un spectre de puissance garantissant ainsi l'indépendance par rapport à des paramètres tels que la taille des fenêtres d'analyse.

- Le regroupement sur une largeur de bandes de 1 ERB correspond bien à la sélectivité fréquentielle de l'oreille.
- Aucun algorithme de détection d'harmonique ni de F_0 n'est nécessaire.
- Le niveau de chaque bande subit une transformation non-linéaire (puissance 0,3) afin d'approcher la sonie estimée dans chaque bande.
- Le centroïde spectral est exprimé selon un axe en ERBrate ou en Hz.

7.3.3 Corrélations avec les solutions MDS

Le tableau 7.2 montre les corrélations entre la deuxième dimension des solutions MDS de l'Exp I, III et McAdams et coll. (1995) avec les descripteurs mentionnés.

TAB. 7.2: *Tableau récapitulatif de la valeur absolue des corrélations, à une décimale de précision, entre les descripteurs et les projections sur la deuxième dimension des solutions de l'Exp. I, III et McAdams et coll. (1995) (d.d.l. = 10, pour l'Exp. I et d.d.l. = 16 pour l'Exp. III et McAdams et coll. 1995) toutes les corrélations sont significatives ($p < 0.001$).*

Expériences	CGS de base	CGS de rang	Acuité	centroïde spectral
Exp. Ia	0,9	0,9	0,9	0,9
Exp. Ib	0,9	0,9	0,9	0,9
Exp. Ic	0,9	0,9	0,9	0,9
Exp. IIa	1,0	0,8	1,0	1,0
Exp. IIb	1,0	0,6	1,0	1,0
McAdams (1995)	0,9	0,9	0,8	0,9

Tous les descripteurs sont bien corrélés avec la deuxième dimension du timbre, seul le CGS de rang décrit nettement moins bien la deuxième dimension de l'Exp. III.

7.4 Les descripteurs d'étalement spectral

7.4.1 Petite revue des descripteurs d'étalement spectral

L'étalement spectral, $\tilde{f}$ est complémentaire du CGS. Il ressemble à l'écart type d'une distribution statistique :

$$\tilde{f} = \sqrt{\frac{\sum_i (f_i - \bar{f})^2 \times A_i}{\sum_i f_i}}, \quad (7.6)$$

où f_i , A_i et $\bar{f}$ représentent les mêmes notions que dans l'équation 7.2. Chaque paramètre est soumis aux mêmes choix que lors du calcul du CGS et peut également être modélisé de façon à rendre compte de la physiologie de l'oreille.

Le descripteur employé

Le descripteur que nous avons utilisé se base sur la sonie partielle, calculée comme nous l'avons expliqué, pour notre descripteur de centroïde spectral. L'étalement spectral $\tilde{z}$ se définit donc ainsi :

$$\tilde{z} = \sqrt{\frac{\sum_z (z - \bar{z})^2 \times \psi_z}{\sum_z \psi_z}}, \quad (7.7)$$

où z , ψ_i et $\bar{z}$ représentent les mêmes notions que dans l'équation 7.5.

7.4.2 Choix du descripteur

Ce descripteur a l'avantage d'être complémentaire avec le centroïde spectral. Il partage les mêmes critères de robustesse et de bonnes correspondances avec les données physiologiques. Il est théoriquement indépendant du centroïde spectral. Toutefois, dans la pratique, le spectre des instruments naturels étant rarement symétrique, il évolue souvent dans le même sens. La trompette par exemple va augmenter sa brillance par l'apparition d'énergie dans les hautes fréquences. Ce phénomène aura pour conséquence une augmentation du centroïde spectral et de l'étalement spectral.

7.4.3 Corrélations avec les solutions MDS

Ce descripteur est corrélé avec un coefficient de 0,9 ($d.d.l. = 16$ et $p < 0.001$) avec la troisième dimension de l'Exp. III pour les deux sessions. Il est également corrélé avec un coefficient de 0,9 avec la troisième dimension de McAdams et coll. (1995). On peut donc penser que ce descripteur décrit mieux cette dimension psychologique que le flux spectral, proposé par McAdams, corrélé à 0,5.

7.5 Les autres descripteurs pertinents pour le timbre

La littérature propose bon nombre d'autres descripteurs, cette section va en passer certains en revue.

7.5.1 Le vibrato

Les stimuli utilisés lors d'expériences de dissemblance sont, pour des raisons pratiques, souvent relativement courts (voir section 2). Les variations temporelles n'ont le temps de se manifester réellement que durant l'attaque ou l'extinction. Cependant, lorsqu'un musicien joue une note d'une certaine durée (en général supérieure à la seconde), il est rare que celle-ci garde un aspect temporel stable. Le vibrato est par exemple une des techniques les plus usitées par les musiciens afin de donner "vie" à une note. Il a fait l'objet de nombreuses modélisations comme par exemple Mellody et Wakefield (2000) ou Ewert et coll. (2002). Ces modèles se basent principalement sur la variation de l'enveloppe temporelle, et sur ses fréquences associées.

7.5.2 La régularité spectrale

Les études sur les descripteurs de Krimphoff et coll. (1994) ont démontré que la troisième dimension de l'espace de timbre de Krumhansl (1989) était corrélée à la régularité spectrale. Cette dimension permet de distinguer les sons ayant une enveloppe spectrale régulière des sons dont l'enveloppe spectrale est "accidentée". La clarinette présentant un rapport élevé entre ses harmoniques paires et impaires, reste un exemple très cité. Faure (2000) propose, afin de décrire la sensation associée, le terme de "richesse". Krimphoff, afin de la quantifier, propose une mesure

de "Déviation Spectrale" définie ainsi :

$$Reg = \sum_{k=1}^N |20\log(A_k) - \frac{1}{3}(20\log(A_{k-1}) + 20\log(A_k) + 20\log(A_{k+1}))|, \quad (7.8)$$

où A_k représente l'amplitude de la k ème harmonique.

La même technique que pour le centroïde spectral et l'étalement spectral peut être utilisée afin de rendre ce descripteur plus proche de la physiologie de l'oreille.

$$\tilde{z} = \sum_{z=1}^Z |\psi_z - \frac{1}{3}(\psi_{z-1} + \psi_z + \psi_{z+1})|, \quad (7.9)$$

où Z représente le nombre total de bande filtre, z et ψ_z représentent les mêmes notions que dans l'équation 7.5. De récentes études ont montré la pertinence de ce descripteur (Lemaitre, 2004). Toutefois, un travail supplémentaire sur ce descripteur est nécessaire afin, par exemple, de le rendre insensible au nombre de filtres utilisés.

7.5.3 La fluctuation spectrale

L'espace de timbre de McAdams et coll. (1995) montre une corrélation entre la fluctuation spectrale et la troisième dimension perceptive. Ils proposent de la définir comme la moyenne des corrélations entre les spectres d'amplitudes de fenêtres temporelles adjacentes ; plus le degré de variation est faible plus la corrélation est grande. Ce descripteur, bien qu'étant bien corrélé avec des espaces créés à partir de signaux synthétiques, est très sensible à deux paramètres de base. Premièrement, le pas temporel ; plus le pas est faible, plus les différences seront faibles car le spectre n'aura pas le temps d'évoluer significativement. Ce descripteur risque donc d'ignorer les variations à basses fréquences. A l'opposé, avec un pas temporel trop grand, le spectre d'une fenêtre temporelle donnée moyenne trop fortement le signal dans le temps et le descripteur risque d'ignorer les variations hautes fréquences. Deuxièmement, ce descripteur nécessite une recherche des harmoniques autour des fréquences multiples de F_0 . Un seuil doit donc être fixé afin de discriminer l'énergie d'une harmonique du bruit. Si le seuil est faible, de nombreuses d'harmoniques vont être choisies, le nombre de points de la corrélation grandira et affaiblira la contribution de variations significatives que pourraient avoir les premières harmoniques. Si, dans l'autre sens, le seuil est trop élevé, le nombre d'harmoniques retenues sera faible. La corrélation se faisant sur peu de points,

sa significativité sera très réduite. Finalement, remarquons que, par définition, le terme est maladroit ; en effet plus la valeur de flux spectral est élevée, moins le signal fluctue, l'inverse aurait été plus approprié.

Scheirer et Slaney (1997) proposent, afin de discriminer un signal musical d'un signal de parole, d'estimer le flux spectral en calculant la norme au carré de la différence entre les amplitudes de l'enveloppe spectrale, fenêtre par fenêtre :

$$FS = \sum_{k=1}^K \sqrt{\sum_{t=1}^{(T-1)} \frac{(Ak_{t+1} - Ak_t)^2}{T-1}}, \quad (7.10)$$

où Ak_t est la moyenne de l'amplitude de toutes les harmoniques au temps t . Cette définition corrige bien les imperfections de la précédente en n'utilisant plus la corrélation. De surcroît, la définition est plus adaptée ; plus le flux spectral est grand, plus le signal fluctue. Cependant ce descripteur reste très dépendant du niveau global du stimulus. En effet si FS correspond au flux spectral d'un stimulus, ce même stimulus multiplié par une constante, donc avec un rapport harmonique inchangé, verra son flux spectral multiplié par cette même constante. Pire, si l'enveloppe temporelle de ce stimulus varie avec le temps, son flux spectral sera grandement affecté. Un sinus à 1 kHz modulé en amplitude par un sinus à 5 Hz possédera un flux spectral non négligeable. Finalement, tout comme le flux spectral de McAdams et coll., il est fortement dépendant du pas temporel utilisé.

7.5.4 La rugosité

La rugosité est une sensation auditive induite par des battements de même bande de fréquence. Cette sensation est particulièrement complexe à prédire. Différents modèles perceptifs ont toutefois été conçus (Zwicker et Fastl, 1990, par exemple). Pressnitzer (1994) propose le modèle suivant : le signal est d'abord filtré par un passe-bande imitant les caractéristiques de l'oreille externe et moyenne. Puis il est décomposé en sous-bandes de fréquence par un gammatone (Patterson et coll., 1995; Slaney, 1993). Chaque sous-bande est redressée (seule la partie positive est gardée), puis compressée par une puissance de 0,5. Les enveloppes temporelles de chaque sous-bande sont ensuite extraites par un filtre passe-bas, et transformées en spectre fréquentiel. Un filtre passe-bande est finalement appliqué afin de ne garder pour spectre que les fréquences de modulation contribuant à la rugosité. Les énergies de ces fréquences donnent ainsi les rugosités partielles du signal.

Au moment de la rédaction de ce document, nous n'avons pas encore connaissance d'espaces de timbre de stimuli issus d'instruments de musique ayant une dimension corrélée avec la rugosité. Cependant, pour des études sur le timbre des sons issus de bruits naturels (Bjork, 1985) ou industriels (Lemaitre, 2004), ils apparaissent plus fréquemment. C'est sans doute dû au fait que la rugosité étant considérée comme désagréable, les facteurs se sont évertués à la supprimer des instruments de musique. Cette conception est en train de changer, notamment avec la popularisation de la musique concrète et électronique.

7.6 Comment améliorer les descripteurs

L'Exp. III permet, notamment avec l'analyse par la méthode des ancres, de pouvoir améliorer ces descripteurs en y ajoutant trois contraintes supplémentaires :

1. A chaque F_0 , le descripteur doit prédire convenablement la position du stimulus sur la dimension perceptive lui correspondant ;
2. Un instrument dont le timbre ne varie pas avec la F_0 doit conserver un descripteur invariant ;
3. Les descripteurs associés aux instruments dont le timbre varie avec la F_0 , doit varier de façon correspondante.

Les figures 7.1 et 7.2 montrent les quatre descripteurs choisis en fonction de la position de chaque stimulus sur les dimensions de l'espace pour la session 'a' et 'b' de l'Exp. III, respectivement.

Ces descripteurs ont, pour la plupart, été proposés à partir d'études sur le timbre faites à F_0 constante. Dans leur élaboration, la F_0 n'a jamais été considérée comme un paramètre. Ce n'est plus le cas avec les descripteurs utilisés pour les espaces de l'Exp. III, contenant des stimuli dont les F_0 sont différentes.

Lors des précédentes expériences, les stimuli ayant tous la même F_0 , les descripteurs de CGS de base et CGS de rang ne pouvaient être dissociés car ils étaient parfaitement corrélés entre eux. L'espace de l'Exp. III. nous permet de choisir entre ces deux méthodes car il contient des stimuli à plusieurs F_0 . Afin de se persuader du résultat, il suffit d'observer la figure du nuage de points. Le rapport de F_0 entre les stimuli à si_2 et à si_3 étant approximativement de 2, cela implique que tout les stimuli à si_3 aient un descripteur deux fois plus petit. La corrélation baisse légèrement pour la session 'a', où le rapport de F_0 est proche de 1, (de 0,9 à 0,8), mais elle baisse significativement pour la session 'b', où le rapport est proche

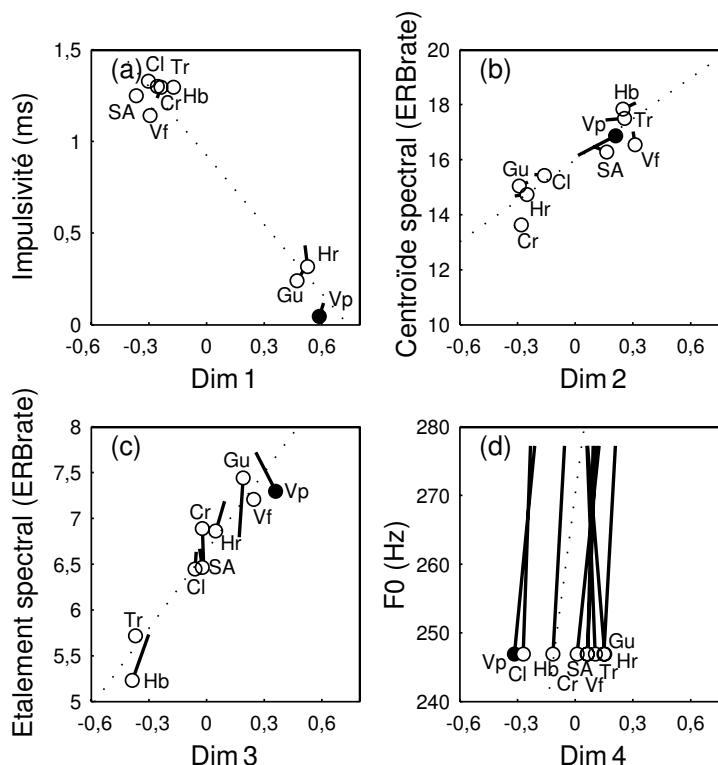


FIG. 7.1: Nuage de points (scatterplot) entre les descripteurs et les dimensions perceptives qu'ils prédisent pour l'Exp. IIIa. (a) l'impulsivité versus la dimension 1 ; (b) le centroïde spectral versus la dimension 2 ; (c) l'étalement spectral versus la dimension 3 ; (d) la F_0 versus la dimension 4. Pour chaque instrument l'étiquette représente sa position à si_2 , l'autre extrémité de la ligne sa position à $do\#_3$. Les points en noir indiquent que l'instrument "bouge" selon la méthode des ancres. La ligne en pointillé représente la droite de régression.

de 2 (de 0,9 à 0,6). Cette expérience nous permet d'affirmer que le CGS de rang n'est pas adapté.

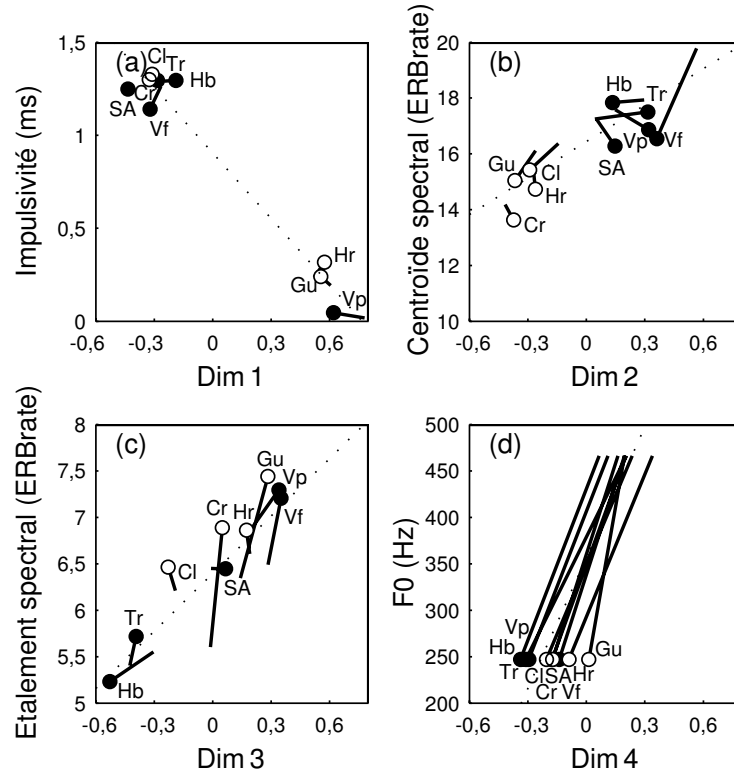


FIG. 7.2: Nuage de points (scatterplot) entre les descripteurs et les dimensions perceptives qu'ils prédisent pour l'Exp. IIIb (voir la légende figure 7.1)

7.7 Conclusions

Ce chapitre nous a permis de choisir trois descripteurs fiables et robustes : l'impulsivité, le centroïde spectral, et l'étalement spectral. Ils sont bien corrélés avec les dimensions des solutions MDS des Exp. I et III.

Les expériences suivantes se basent sur ces descripteurs afin d'étudier l'effet de la F_0 sur le timbre dimension par dimension.

Chapitre 8

Expérience V. Le centroïde spectral

Les expériences I-III ont étudié l'effet de la F_0 sur le timbre des notes issues d'instruments de musique. Une certaine stabilité a été observée, ainsi que des petites variations spécifiques aux instruments. Le timbre des instruments naturels pouvait être affecté par une variation de F_0 de deux façons : 1) par une dépendance possible du timbre de tout stimulus sur sa fréquence fondamentale ; 2) et/ou par une dépendance propre à chaque instrument, due à des variations des processus physiques selon la note jouée. Les expériences IV-VII utilisent des stimuli synthétiques dans le but d'éliminer ce deuxième facteur.

Les expériences I-III utilisaient dans le même corpus, des instruments impulsifs et entretenus. Il y avait une forte présomption que la stabilité du timbre face à la F_0 soit due, en grande partie, à l'importante discrimination entre ces deux groupes d'instruments. Des analyses ont été faites systématiquement à l'intérieur des sous-groupes, montrant que la stabilité était conservée. Toutefois, ces analyses ont mis en lumière que la stabilité du timbre était plus faible à l'intérieur des sous-groupes. Ceci appelle à une nouvelle étude avec un corpus expérimental n'ayant pas cette discrimination.

Le timbre est un attribut multidimensionnel. Ce caractère amène de nombreuses complications lors de son étude. Il semblerait maintenant plus favorable de ne plus considérer le timbre dans son entité mais dimension par dimension. L'expérience IV étudie l'effet de la F_0 sur la dimension décrite par l'impulsivité, les expériences V-VI l'étudient sur la dimension décrite par le centroïde spectral, enfin, l'expérience VII, sur celle décrite par l'étalement spectral. Les stimuli, créés synthétiquement, ne varient qu'en F_0 et sur la dimension concernée.

Pour des raisons de clarté, ce chapitre commence par présenter l'Exp. V. Puis nous décrirons l'Exp. VI au Chapitre 9, et l'Exp. VII au Chapitre 10. L'expérience

IV sera finalement présentée en Chapitre 11.

8.1 Méthodes

8.1.1 Stimuli

Les stimuli sont volontairement basés sur un modèle de synthèse très simple pour qu'ils aient des grandeurs de descripteurs bien définies. Tous les sons possèdent la même durée physique de 0,5 secondes et environ le même niveau de 75 dBA. De plus, avant chaque expérience, un ajustement de la sonie est effectué expérimentalement par des sujets ne participant pas aux autres expériences. Tous les stimuli sont échantillonnés à 44.1 kHz avec une résolution de 16 bits. La synthèse est effectuée par addition de sinus avec le logiciel MATLAB.

L'enveloppe temporelle L'enveloppe temporelle reprend la forme la plus typique des enveloppes de synthèse. Elle est, comme le montre la figure 8.1, de la forme ADSR (Attack Decay Sustain Release), utilisée par les premiers synthétiseurs (Roads, 1996). Elle est formée d'une rampe croissante (Attack) de 50 ms, d'une rampe décroissante (Decay) de 22,5 ms, d'une partie stable entretenue (Sustain) de 382,5 ms et d'une nouvelle rampe décroissante d'extinction (Release) de 45 ms.

L'enveloppe spectrale L'enveloppe spectrale suit également le modèle très simple d'une distribution gaussienne, pouvant varier selon deux paramètres : sa moyenne et son écart type. Le premier agit sur le centroïde spectral et le second sur l'étalement spectral. Afin de rester dans l'esprit des modèles perceptifs, l'enveloppe spectrale considérée montre des caractéristiques proches de la perception. Son axe fréquentiel suit l'échelle ERBrate et son énergie est compressée par une puissance de 0,3. Une fois ces conditions réunies, la forme de son spectre suit une distribution gaussienne. La figure 8.2 montre le spectre de trois stimuli de même F_0 , variant uniquement par leur centroïde spectral.

Nous avons vu avec l'instrument SA qu'un spectre fixe ne garantissait pas forcément des descripteurs fixes et donc un timbre fixe. La contrainte ici était donc de donner la même valeur de descripteurs aux stimuli ayant des F_0 différentes. Les stimuli n'étaient pas limités en nombre d'harmoniques, ou du moins jusqu'à la moitié de leur fréquence d'échantillonnage (selon le théorème de Shannon). Mais

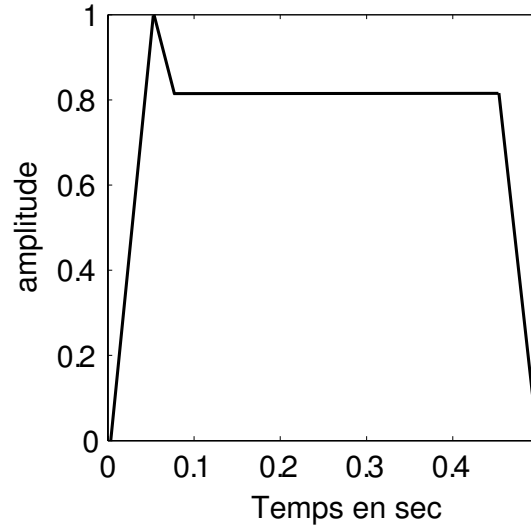


FIG. 8.1: Enveloppe temporelle des stimuli de l'Exp. V, avec comme paramètres 50, 22,5, 382,5 et 45 ms correspondant respectivement aux durées d'attaque, de décroissance, de soutien et d'extinction.

le fait qu'ils ne possèdent pas d'énergie à une fréquence inférieure à leur F_0 a provoqué une dissymétrie de l'enveloppe spectrale ayant pour conséquences qu'une variation de moyenne pouvait affecter l'étalement spectral et qu'un changement d'écart type pouvait affecter le centroïde spectral. Afin de prendre en considération ce problème, les stimuli ont été effectués par pas successifs : 1) la moyenne était adaptée afin d'obtenir le centroïde spectral cible ; 2) l'écart type était adapté afin d'obtenir l'étalement spectral cible ; 3) la moyenne était réadaptée avec le nouvel écart type afin d'obtenir de nouveau le centroïde spectral cible. Ces opérations garantissaient que les stimuli possèdent bien les grandeurs de descripteurs désirées.

Cette expérience comportait au total 24 stimuli issus de la combinaison de trois F_0 : si_2 , $do\#_3$ et $si\flat_3$; et de huit valeurs de centroïde spectral : 15, 16, 17, 18, 19, 20, 21, 22 ERBrate. Tous les stimuli avaient l'enveloppe temporelle mentionnée plus haut et une même valeur d'étalement spectral à 5 ERBrate.

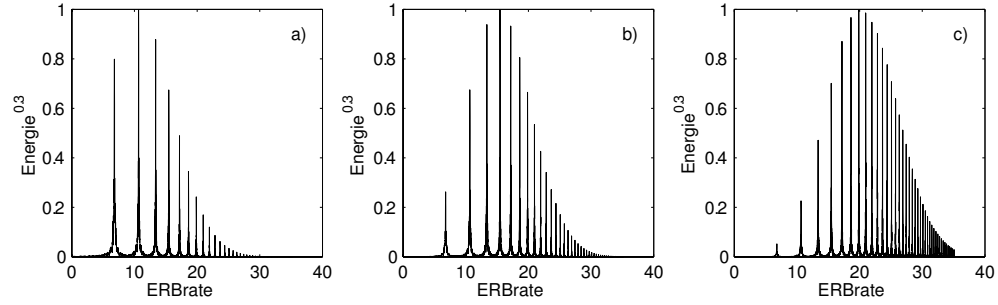


FIG. 8.2: Trois spectres de stimuli de même étalement spectral ne variant que par leur centroïde spectral à : a) à 15, b) 18, et c) 22 ERBrate

8.1.2 Tâche

La consigne, l'interface, ainsi que le déroulement de cette expérience étaient semblables à ceux de l'Exp. III. La session 'a' présentait tous les stimuli ayant une F_0 à si_2 et $do\#_3$, la session 'b' les stimuli avec une F_0 à si_2 et $si\flat_3$.

8.1.3 Sujets

Quatorze sujets ont participé à cette expérience, dont 7 femmes et 7 hommes, 7 musiciens et 7 non musiciens. Aucun d'entre eux n'avait participé aux premières expériences. L'expérience durait environ une heure. L'ordre des sessions 'a' et 'b' était contrebalancé.

8.2 Résultats et analyses

8.2.1 Effets de sujet, ordre et expertise instrumentale

La cohérence entre les sujets a été testée à l'aide d'une analyse en cluster avec une méthode 'moyenne'. L'analyse a révélé que, pour la session 'a', les sujets 2, 6, et 13 étaient faiblement corrélés avec les autres. Les sujets 6 et 13 étaient encore faiblement corrélés lors la session 'b'. Ces trois sujets ont donc été exclus du reste des analyses. Des ANOVA n'ont pas révélé d'effet de l'ordre de passation des sessions ni de l'expertise musicale.

8.2.2 Matrices de dissemblance

La matrice de dissemblance de la session 'a', moyennée à travers tous les sujets, est représentée sur le panneau de gauche de la figure 8.3.

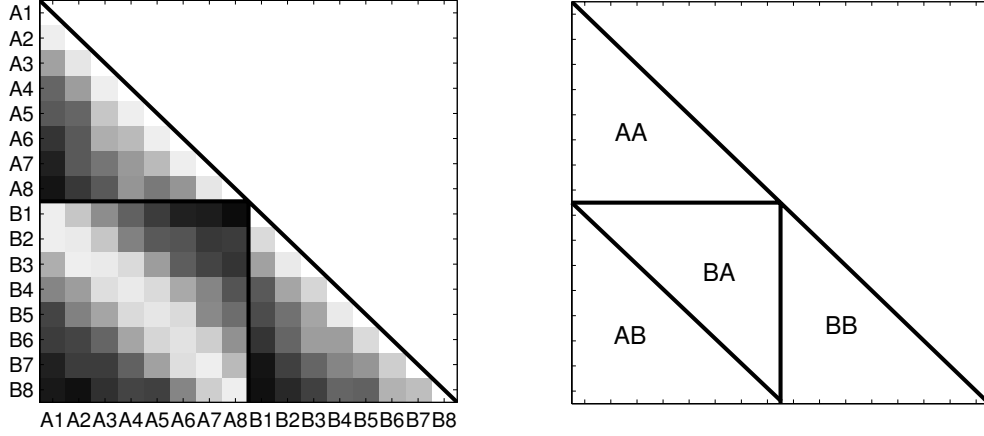


FIG. 8.3: *Panneau gauche : Matrice de dissemblance de l'Exp. V, pour la session 'a'. Chaque stimulus est représenté par une lettre et un chiffre. La lettre correspond à sa F_0 : A = si_2 , B = $do\#_3$; le chiffre sa valeur de centroïde spectral : 15, 16, 17, 18, 19, 20, 21, 22 ERBrate numérotée de 1 à 8, respectivement. Panneau droit : Division schématique d'une matrice de dissemblance en quatre sous-ensembles de données.*

Cette matrice de dissemblance peut être décomposée en quatre sous-ensembles de données regroupés chacun dans un triangle comme le montre le panneau de droite de la figure 8.3. Chaque triangle est constitué de paires avec une relation particulière. Le premier, AA, comporte toutes les paires ayant des stimuli à si_2 . Le deuxième, BB, comporte toutes les paires avec des stimuli à $do\#_3$ (ou $si\#_3$ si l'on considère la matrice de la session 'b'). Le troisième et le quatrième, AB et BA, comportent toutes les paires composées de deux stimuli avec des F_0 différentes.

Ces quatre triangles sont très semblables. Ils sont tous constitués d'un dégradé continu entre des valeurs fortes dans l'angle droit du triangle contenant les paires jugées les plus dissemblables : [A1-A8, A1-B8, B1-A8, B1-B8], et l'hypoténuse du triangle contenant les paires jugées les plus semblables : [A1-A2, A2-A3, etc.]. Leur coefficient de corrélation deux à deux est toujours supérieur à 0,95 ($d.d.l. = 26, p < 0,001$) ce qui confirme cette similitude. Le carré comportant les triangles AB et BA comporte les trois mêmes aspects que les matrices de l'Exp. II, Chapitre

5, c'est à dire : la symétrie, la diagonale faible et la similitude avec les matrices contenant des paires appariées en F_0 , AA et BB.

La figure 8.4 montre les matrices de dissemblance des deux sessions.

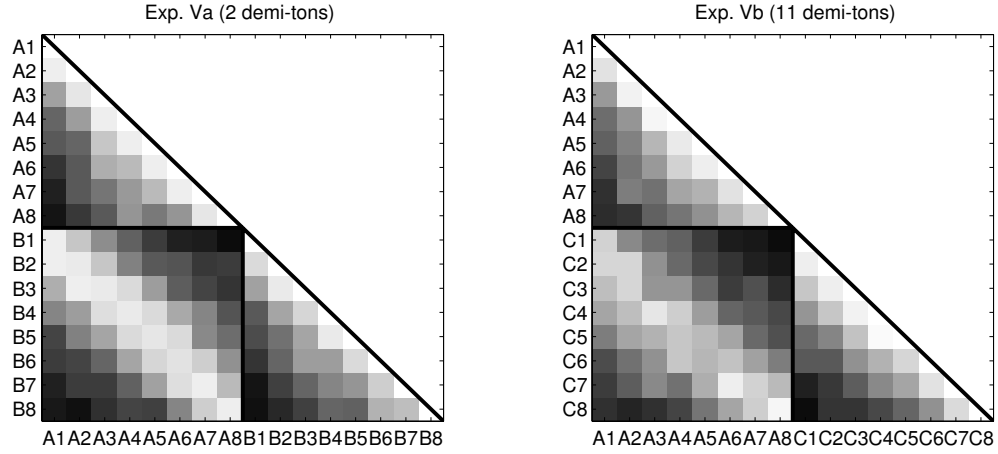


FIG. 8.4: *Panneau gauche : Matrice de dissemblance de la session 'a', même légende que figure 8.3. Panneau droit : Matrice de dissemblance de la session 'b'. La lettre A correspond à une F_0 à si_2 et la lettre C une F_0 à si_3 ; chaque chiffre correspond à la même valeur de centroïde spectral que dans la figure 8.3.*

Ces deux matrices sont également très proches et sont corrélées avec un coefficient de 0,9 ($d.d.l. = 118$, $p < 0,001$). Toutefois, la diagonale contenant des paires avec deux stimuli de même centroïde spectral mais différenciés en F_0 est plus dissemblable pour la session 'b', avec une moyenne à 0,19, que pour la session 'a' avec une moyenne à 0,05. La symétrie du carré semble également moins fortement marquée.

8.2.3 Analyses de variance

Ces deux matrices ont été soumises à une ANOVA à mesures répétées avec les facteurs session représentée par la ΔF_0 (2) et paires d'instruments¹ (120). Les résultats sont résumés dans le tableau 8.1. L'effet principal ΔF_0 n'est pas significatif, au contraire, de l'effet des paires d'instruments et de l'interaction entre les

¹Le terme "instrument" sera utilisé dans de cette expérience et celle qui suit pour désigner un ensemble de stimuli ayant la même valeur de centroïde spectral.

8.2. RÉSULTATS ET ANALYSES

deux facteurs. Ceci atteste d'une certaine influence de la ΔF_0 , et donc de la F_0 , sur la tâche. Cet effet peut être quantifié grâce à une mesure de R^2 . La somme de la variance due à l'effet de ΔF_0 est inférieure à 3 %, son influence est donc faible.

TAB. 8.1: Tableau ANOVA entre la session 'a' et 'b' de l'Exp. V. Les abréviations sont les suivantes : S pour sujets, ΔF_0 pour la différence de fréquence fondamentale, PI pour les paires d'instruments, d.d.l. pour degré de liberté, SC pour Somme des Carrés, CM pour Carré Moyen, F pour le rapport F, p.c. pour la probabilité critique et R^2 pour le pourcentage de variance induit par chaque effet.

Source	d.d.l.	SC	CM	F	p.c.	R2
S	10	35.57	3,56			
ΔF_0	1	0,00	0,00	0,02	0,8815	0,0
$\Delta F_0 \times S$	10	1,00	0,10			
PI	119	184.74	1,55	31,32	0,0000	54,91
$PI \times S$	1190	58,98	0,05			
$PI \times \Delta F_0$	119	10,00	0,08	2,17	0,0000	2,97
$PI \times \Delta F_0 \times S$	1190	46,14	0,04			

Trois autres ANOVA ont ensuite été effectuées sur les sous-ensembles à l'intérieur de chaque session. La première teste les deux triangles contenant les paires appariées en F_0 , AA vs BB ; la deuxième teste les deux triangles du carré, AB vs BA ; la troisième compare le triangle AA moyenné avec BB contre le triangle AB moyenné avec BA afin de mesurer les dissemblances avec et sans ΔF_0 . Dans le but d'éviter un faux positif possible dû au nombre élevé d'analyses faites sur les mêmes données le seuil de significativité a été abaissé à 0,5 %, pour ces ANOVA et pour celles qui suivent dans cette expérience. Le tableau 8.2 résume ces ANOVA par les mesures de R^2 des effets significatifs pour les sessions 'a' et 'b', séparément.

L'effet principal des paires d'instruments est toujours significatif. L'effet principal de la F_0 est significatif entre les triangles AA et BB de la session 'a'. Pour la session 'b', les interactions le sont également. Dans les deux sessions la moyenne du triangle BB (0,49 et 0,44 pour la session 'a' et 'b', respectivement) est supérieure à la moyenne du triangle AA (0,43 et 0,41). La différence de moyenne est faible et ne semble pas augmenter avec la ΔF_0 .

TAB. 8.2: Tableau récapitulatif des R^2 pour les ANOVA intra-sessions, pour les effets significatifs à 0,5 %

	Session 'a'			Session 'b'		
	F_0	PI	$F_0 \times \text{PI}$		$F_0 \text{ PI}$	$F_0 \times \text{PI}$
AA vs BB	0,73	55,85	-	-	54,71	2,93
AB vs BA	-	55,90	-	10,38	33,95	-
[AA+BB] vs [AB+BA]	-	64,41	-	2,83	55,47	-

L'effet de la F_0 entre les deux triangles du carré, AB et BA, est un bon indicateur de sa symétrie. Seul le carré de la session 'b' peut être considéré comme significativement asymétrique.

Pour la session 'a', aucun effet significatif dû à la F_0 n'est constaté sur les dissemblances avec et sans ΔF_0 . Pour la session 'b' les interactions sont significatives. Dans tous les cas le R^2 montre que les effets dus à la F_0 sont bien plus faibles que les effets dus au centroïde spectral.

8.2.4 Analyses multidimensionnelles de proximité

Les stimuli n'ayant, a priori pas de spécificités et aucune classe de sujets ne semblant se dégager de l'analyse en cluster, une simple analyse MDS, type MDSCAL, utilisant l'algorithme de SMACOF (voir annexe B) a été effectuée. Une représentation à deux dimensions a été choisie en se basant sur la méthode du "coude" tracé à partir des mesures de stress. La figure 8.5 nous montre les configurations obtenues respectivement pour les sessions 'a' et 'b'. Les stimuli à si_2 sont représentés par un carré, ceux à $do\#_3$ ou sib_3 par une étoile. Chaque stimulus est relié par un trait aux deux stimuli donc le centroïde spectral est le plus proche et à même F_0 , ainsi qu'au stimulus doté du même centroïde spectral mais avec une F_0 différente.

Chaque MDS prend la forme particulière dite du *fer à cheval* (pour plus de détail voir Buja et Swayne, 2001). Cette configuration a pu être provoquée par une sous-estimation des grandes distances sur l'échelle des dissemblances (Schonne- mann, 1983). En effet, un score de 0 est attribué aux paires jugées *très similaires* et un score de 1 est attribué à celles jugées *très différentes*. Ces deux extrêmes devraient correspondre, dans un espace, à une distance nulle et une distance infi-

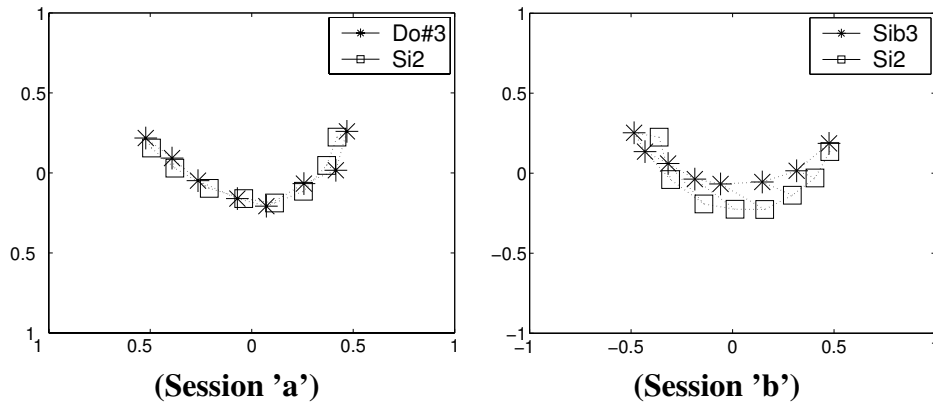


FIG. 8.5: Solution MDS de l'Exp. V. Les carrés représentent les stimuli à si_2 et les étoiles à $do\#_3$ et sib_3 pour la session 'a' et 'b', respectivement.

nie, respectivement. Si le score de 0 renvoie bien à une distance nulle, le score de 1 renvoie, lui, à une distance finie, et peut donc contribuer à une sous-estimation de la représentation des grandes dissemblances. Afin de contourner ce problème, une fonction arc-tangente hyperbolique peut être utilisée pour adapter les dissemblances à ce type de jugement (voir annexe B). Cette fonction, comme le montre la figure 8.6, peut être approchée par une droite de pente 1 pour des valeurs en abscisse allant de 0 à 0,5, puis par une exponentiellement tendant vers l'infini pour une valeur en abscisse de 1. Elle laisse donc inchangées les dissemblances inférieures à 0,5 et augmente exponentiellement les dissemblances supérieures.

La figure 8.7 montre les solutions MDS obtenues en utilisant cette fonction pour les sessions 'a' et 'b'. Les solutions ont également subi une rotation spécifique. Afin de pouvoir orienter les solutions pour que chaque dimension corresponde aux dimensions psychologiques, deux méthodes sont possibles : La première consiste à utiliser des algorithmes de MDS orientant l'espace à partir des poids donnés par les sujets sur chaque dimension, voir Chapitre 2. Dans cette expérience, le nombre de sujet étant faible, l'algorithme CLASCAL n'a pas extrait de grandes différences entre les classes de sujets. L'orientation proposée était donc peu précise. La deuxième méthode consiste à orienter les axes par rapport à des descripteurs dont la littérature a montré qu'ils étaient corrélés avec les dimensions psychologiques. McAdams et coll. (1995) ont montré à l'aide de CLASCAL et 88 sujets que le centroïde spectral décrivait bien une dimension du timbre. Toutefois, cette méthode pose un problème car cette expérience comporte des stimuli à plu-

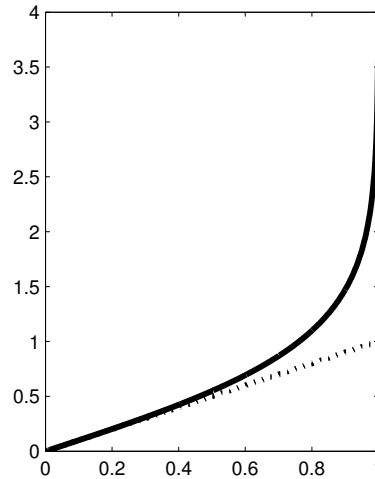


FIG. 8.6: *Fonction arc-tangente hyperbolique, assimilable à une droite de pente 1 (en pointillé) pour des abscisses entre 0 et 0,5 ; puis se comportant de façon exponentielle jusqu'à atteindre l'infini pour une abscisse de 1*

sieurs F_0 et, dans l'hypothèse d'une variation "isométrique" du timbre avec la F_0 , des stimuli avec le même centroïde spectral et des F_0 différents peuvent ne plus partager la même position sur cette dimension psychologique. Nous avons donc pris le parti d'orienter la solution afin d'optimiser la corrélation non pas entre tous les stimuli et leur centroïde spectral, mais seulement entre les stimuli à si_2 et leur centroïde spectral.

La session 'a' est unidimensionnelle (les stimuli à si_2 et $do\#_3$ sont alignés) et la session 'b' est bidimensionnelle. La répartition des stimuli aigus (à $do\#_3$ et $si\flat_3$), symbolisés par des étoiles, semble légèrement plus étendue que celle des stimuli graves (à si_2) symbolisés par les carrés. Pour la session 'b', les stimuli à $si\flat_3$ semblent être décalés systématiquement vers la gauche par rapport aux stimuli à si_2 . Sur la première dimension, les stimuli sont régulièrement espacés selon la valeur de leur centroïde spectral. Cet aspect indique que la perception de cette dimension est proportionnelle au centroïde spectral et donc que celui-ci est un bon descripteur. Sur la seconde dimension de la session 'b', les stimuli sont discriminés selon leur F_0 .

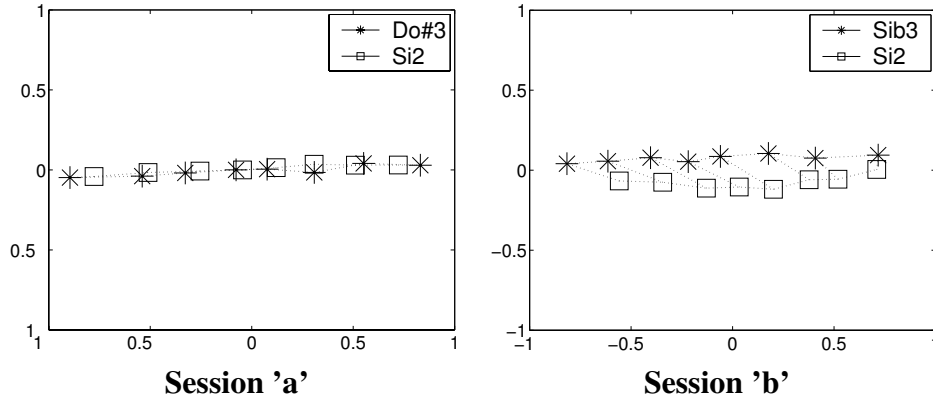


FIG. 8.7: Solutions MDS de l'Exp. V à partir de l'arc-tangente hyperbolique des données de dissemblance

8.3 Implications pour le timbre

Les analyses ANOVA permettent d'évaluer si les caractéristiques des solutions MDS sont significatives.

Expansion Les analyses ANOVA sur les sous-matrices à F_0 constante (AA vs BB) estiment si les paires appariées à une certaine F_0 ont été jugées globalement plus ou moins dissemblables que celles jugées à une autre F_0 . Si tel est le cas, cela doit se manifester sur la MDS par une expansion des stimuli en fonction de leur F_0 . Les résultats de ces analyses ont montré un effet significatif pour les deux sessions. On peut donc en conclure que l'élargissement de la répartition des stimuli avec une F_0 élevée est significatif. Cet élargissement peut être considéré comme uniforme pour la session 'a' (Effet principal de la F_0 significatif) et comme dépendant du centroïde spectral pour la session 'b' (effet de l'interaction significatif). Il serait tentant d'y voir un effet systématique. L'Exp. VI (chapitre suivant) nous apprendra que cela n'est pas le cas.

Déplacement isométrique Les analyses ANOVA entre les deux triangles avec une ΔF_0 , AB vs BA, estiment la symétrie du carré. Nous avons vu lors de l'Exp. II, Chapitre 5, que la symétrie de ces matrices impliquait que les solutions MDS n'avaient pas subi un déplacement isométrique le long des dimensions du timbre. Les mêmes conclusions peuvent être tirées des résultats de ces ANOVA. Un effet

lié à la F_0 s'est révélé significatif pour la session 'b'. On peut donc en conclure que pour cette session le déplacement vers la gauche observé le long de la première dimension était significatif. L'Exp. VI (chapitre suivant) montre que cet effet est systématique et grandit avec la F_0 .

La deuxième dimension L'ANOVA sur les triangles entre les paires avec et sans ΔF_0 estime si la dissemblance $[X_1 Y_1]$ est significativement différente de $[X_1 Y_2]$, ou de $[X_2 Y_1]$. Si tel est le cas, on peut penser que la position de X_1 est différente de X_2 , et que celle de Y_1 est différente de Y_2 . Cela se manifestera dans la solution MDS par deux groupes discriminés selon leur F_0 . L'effet de la F_0 ne s'est révélé significatif que pour la session 'b'. On peut en conclure que cette session est bidimensionnelle et la session 'a' unidimensionnelle. L'Exp. VI (chapitre suivant) montre que cet effet grandit avec la ΔF_0 .

8.4 Conclusions

Dans cette expérience, les stimuli n'étaient différenciés que selon deux paramètres physiques : leur centroïde spectral et leur F_0 . La session 'a', avec une faible ΔF_0 , a permis d'obtenir un espace unidimensionnel. La session 'b', avec une ΔF_0 plus importante, a produit quant à elle un espace bidimensionnel. Ces deux espaces possèdent une dimension bien décrite par le centroïde spectral.

Cette expérience montre des résultats attribuables à un effet de la F_0 sur le timbre : 1) pour une faible ΔF_0 , la dimension liée à la brillance subit une expansion uniforme ; 2) pour une différence plus importante, ajoutée à l'expansion, la dimension se décale légèrement vers la droite ; 3) une dimension prédite par la F_0 apparaît avec une ΔF_0 importante.

Toutefois, les ΔF_0 utilisées au cours de cette expérience restent toujours en dessous de deux limites : 1) l'octave sur l'échelle des tonies ; 2) 2 demi-tons sur l'échelle des chroma. La première est une limite importante selon Handel et Erickson (2001). La seconde montre que les grands écarts de chroma n'ont pas été explorés et leur hypothétique effet sur le timbre non plus. L'expérience suivante va tester si les effets dévoilés lors de cette expérience peuvent se généraliser pour des écarts de F_0 plus importants sur l'échelle de tonies et de chroma.

Chapitre 9

Expérience VI. Le centroïde spectral au-delà de l'octave

L'expérience V a montré des effets de la F_0 sur le timbre se manifestant par une dilatation et un déplacement uniforme de la position des stimuli le long de la dimension décrite par le centroïde spectral. Le premier effet apparaît pour les deux ΔF_0 testées et le second seulement pour la ΔF_0 de 11 demi-tons. A cet écart, une deuxième dimension liée à la F_0 apparaît également. Ces résultats montrent que l'effet de la F_0 est différent selon l'écart testé. Cet écart reste toujours limité en dessous de l'octave. Une fois passée cette limite, de nouveaux phénomènes pourraient apparaître. L'expérience suivante va tester ce cas de figure en augmentant les écarts testés.

9.1 Méthodes

9.1.1 Choix des F_0

Deux nouvelles F_0 , le fa_3 et le fa_4 correspondant respectivement à 349 et 698 Hz et deux F_0 déjà utilisées, le si_2 et le $si\flat_3$, ont été choisies afin d'explorer les résultats suivant trois échelles de hauteurs différentes. Le fa_4 permet de tester l'effet de la F_0 sur le timbre au-delà de la limite de l'octave (voir figure 9.1a). Le fa_3 permet, en comparaison avec les autres F_0 , de tester diverses relations de chroma et consonances.

Les ΔF_0 précédentes ne dépassaient pas la limite de 2 demi-tons sur le cercle des chroma (voir figure 9.1b). L'écart si_2-fa_3 permet de tester la plus grande dis-

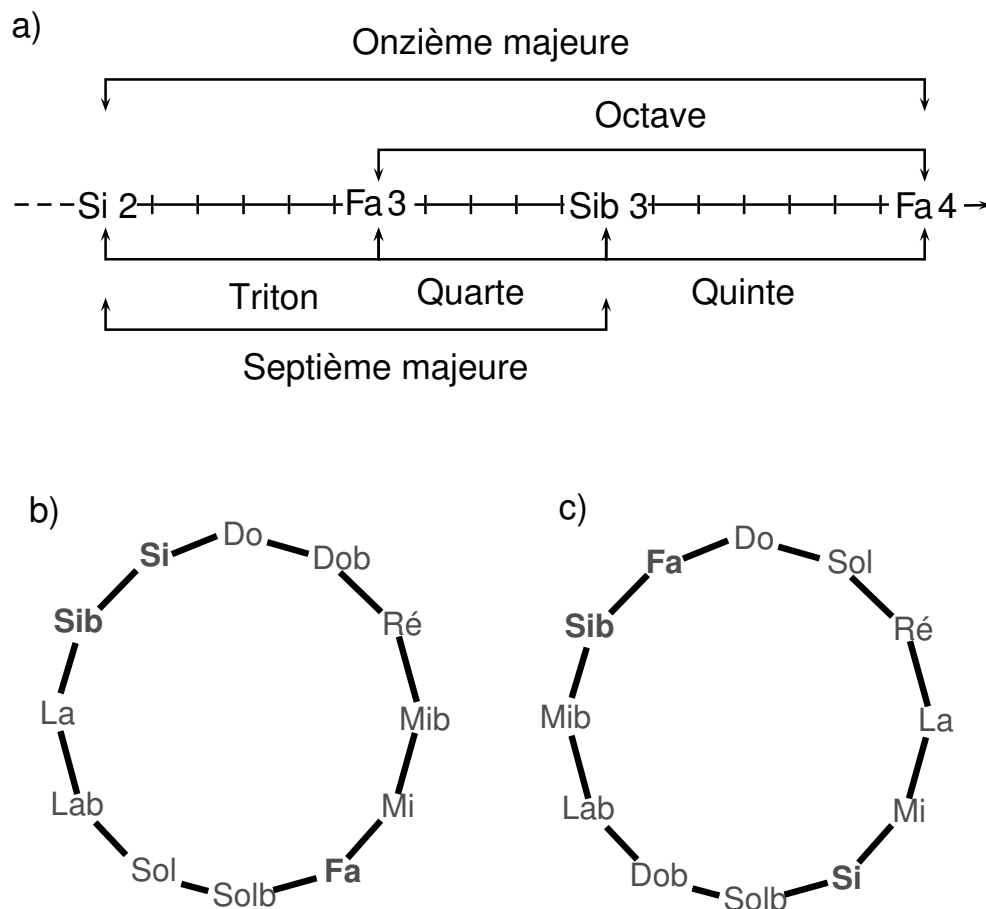


FIG. 9.1: Panneau 'a' : Schéma représentant tous les écarts possibles entre les nouvelles F_0 . Panneau 'b' : Le cercle des chroma. Panneau 'c' : Le cercle des quintes.

tance possible sur ce cercle : le triton.

Puisque les expériences consistent à présenter des paires de sons harmoniques, une troisième échelle doit être considérée : l'échelle de consonances (voir figure 9.1c). Celle-ci permet de prédire si des sons sont perçus de façon plus ou moins consonante (Shepard, 1999). Cette notion peut être prédite par la distance séparant les chroma de deux sons sur le cercle des quintes. L'écart fa_3-fa_4 nous permet de tester une distance nulle, et l'écart si_2-fa_3 une distance maximale sur ce cercle.

9.1.2 Tâche

La consigne de cette expérience était similaire à celle des expériences précédentes (Exp. I-V). Seuls les stimuli différaient. Le corpus expérimental était composé de 24 stimuli synthétisés à partir de la combinaison des 4 F_0 exposées plus haut (si_2 , fa_3 , $si\flat_3$ et fa_4) et de 6 centroïdes spectraux, à 17, 18, 19, 20, 21 et 22 ERBrate. Leur méthode de synthèse était la même que pour les stimuli de l'Exp. V décrite dans le Chapitre 8. Ils possédaient les mêmes caractéristiques temporelles et le même étalement spectral que ces stimuli. L'expérience comportait une unique session.

Quatorze sujets y ont participé dont 7 musiciens et 7 non musiciens, 7 femmes et 7 hommes. L'expérience comportant un nombre important de paires à juger (256), les sujets pouvaient prendre une pause au milieu. L'expérience durait en moyenne une heure et demi.

9.2 Résultats et analyses

9.2.1 Effets de sujet et expertise instrumentale

La cohérence entre les sujets a été testée avec une analyse en cluster. Aucun sujet ne se distinguant particulièrement, ils ont tous été pris en compte dans les analyses. L'ANOVA n'a trouvé aucun effet significatif de l'expertise musicale.

9.2.2 Matrices de dissemblance

La figure 9.2 représente la matrice moyennée à travers tous les sujets. Chaque lettre représente une F_0 utilisée : A) si_2 , B) fa_3 , C) $si\flat_3$ et D) fa_4 . Chaque chiffre représente une valeur de centroïde spectral dans l'ordre : 17, 18, 19, 20, 21 et 22. Cette matrice peut être décomposée en sous-matrice de type triangle ou carré. La sous-matrice de type triangle AA par exemple contient toutes les paires appariées en F_0 à si_2 ; la sous-matrice de type carré AB contient les paires ayant un stimulus à si_2 et un à fa_3 , etc. Chaque sous-matrice, comme celles de l'Exp. V, montre une augmentation graduelle de la dissemblance lorsqu'on s'écarte de la diagonale et donc que la ΔF_0 grandit.

Chaque sous-matrice carrée peut être divisée selon sa diagonale en deux triangles. Les paires contenues dans chaque triangle inférieur (BA, CA,..., CB...) ont un point commun : *à l'intérieur de chaque paire le stimulus doté de la F_0 la plus*

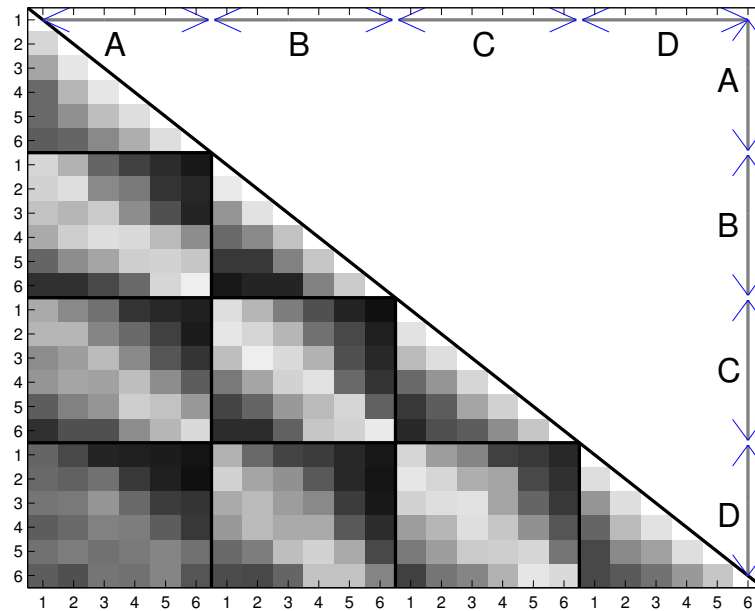


FIG. 9.2: Matrice de dissemblance de l'Exp. VI. Chaque lettre représente une F_0 utilisée : A) si_2 , B) fa_3 , C) si_3 et D) fa_4 . Chaque chiffre représente une valeur de centroïde spectral dans l'ordre : 17, 18, 19, 20, 21 et 22 ERBrate.

faible possède également le plus faible centroïde spectral (A1-B2, A1-B3,..., C4-D7...). Pour les paires du triangle supérieur (AB, AC,..., BC...), on constate l'inverse : à l'intérieur de chaque paire le stimulus avec la plus faible F_0 possède le centroïde spectral le plus élevé (A2-B1, A3-B1,..., C7-D4...). Le tableau 9.1 résume toutes les dissemblances moyennes pour chacun des ces triangles.

TAB. 9.1: Tableau récapitulatif des moyennes des sous-matrices de l'Exp. VI.

F_0	A (si_2)	B (fa_3)	C (si_3)	D (fa_4)
A	0,34	0,59	0,64	0,76
B	0,40	0,50	0,63	0,73
C	0,46	0,41	0,43	0,58
D	0,56	0,39	0,31	0,39

Ce tableau montre que les triangles inférieurs ont toujours des moyennes plus faibles que les triangles supérieurs. La symétrie des carrés est donc loin d'être parfaite. Ce tableau montre également que les dissemblances moyennes augmentent majoritairement avec la ΔF_0 .

9.2.3 Analyses de variance

Une ANOVA à mesure répétée sur les sous-matrices ne contenant que les paires appariées en F_0 (AA, BB, CC et DD) a été effectuée avec les facteurs F_0 (4) et paires d'instruments (15). Le tableau 9.2 retrace les résultats de cette analyse.

TAB. 9.2: ANOVA entre les sous-matrices contenant les paires appariées en F_0 (AA, BB, CC et DD) de l'Exp. VI

Facteur	d.d.l.	SC	CM	F	p.c.	R2
S	13	7,93	0,61			
F_0	3	2,74	0,91	25,19	0,0000	3,21
$F_0 \times S$	39	1,41	0,04			
PI	14	41,87	2,99	74,18	0,0000	48,98
$PI \times S$	182	7,34	0,04			
$F_0 \times PI$	42	2,52	0,06	1,51	0,023	2,95
$F_0 \times PI \times S$	546	21,67	0,04			

Tous les effets sont significatifs, toutefois, la contribution à la variance totale des effets dus à la F_0 reste faible avec moins de 7 %.

Des ANOVA ont été effectuées entre chacune de ces sous-matrices, deux à deux (AA vs BB, AA vs CC,...). Le tableau 9.3 représente le R^2 des effets significatifs à 0,5 %.

Ce tableau nous apporte trois informations importantes : Premièrement, il confirme l'information visuelle selon laquelle toutes les sous-matrices à F_0 constantes sont très proches. En effet, les effets significatifs dus à la F_0 ne contribuent que faiblement à la variance totale existant entre les sous-matrices testées. L'effet le plus important se situe entre la sous-matrice à si_2 et la sous-matrice à fa_3 , les effets liés à la F_0 représentant près de 10 % de la variance. Deuxièmement, aucun effet significatif dû à la F_0 n'a été trouvé entre la sous-matrice à si_2 et celle

CHAPITRE 9. EXPÉRIENCE VI. LE CENTROÏDE SPECTRAL AU-DELÀ
DE L'OCTAVE

TAB. 9.3: ANOVA entre toutes les sous-matrices à même F_0 , deux à deux (AA vs BB, AA vs CC,...), seules les R^2 des effets significatifs à 0,5 % sont représentés.

	A (si_2)			B (fa_3)			C ($si\flat_3$)		
	F_0	PI	$F_0 \times \text{PI}$	F_0	PI	$F_0 \times \text{PI}$	F_0	PI	$F_0 \times \text{PI}$
B	5,9	49,85	3,79						
C	2,05	48,00	-	1,07	58,35	-			
D	-	44,54	-	2,64	51,71	2,94	-	48,84	-

à fa_4 , malgré l'écart important de 18 demi-tons (une fois et demi l'octave). Troisièmement, l'effet observée sur la session 'b' de l'Exp. V se retrouve bien entre la sous-matrice à si_2 et celle à $si\flat_3$. Toutefois, cet effet ne semble pas s'accroître avec la ΔF_0 ; si c'était le cas les effets entre les sous-matrices à si_2 et fa_4 dus à la F_0 seraient significatifs.

Des ANOVA ont également été effectuées afin de comparer les deux triangles de chaque sous-matrice carrée (AB vs BA, AC vs CA...). Le tableau 9.4 représente le R^2 des effets significatifs.

TAB. 9.4: ANOVA entre les deux triangles des sous-matrices non appariées en F_0 (AB vs BA, AC vs CA,...), seules les R^2 des effets significatifs à 0,5 % sont représentés.

	A (si_2)			B (fa_3)			C ($si\flat_3$)		
	F_0	PI	$F_0 \times \text{PI}$	F_0	PI	$F_0 \times \text{PI}$	F_0	PI	$F_0 \times \text{PI}$
B	8,52	42,87	5,11						
C	9,5	28,49	-	11,39	43,41	3,92			
D	12,47	8,31	-	27,45	21,17	3,85	17,45	29,18	-

Ce tableau montre que l'effet de la F_0 est toujours significatif et conséquent, même peut-être plus important que l'effet des paires d'instruments. Ce résultat confirme que les différences de moyennes décrites dans le tableau 9.1 entre les triangles inférieurs et supérieurs sont significatives.

9.2.4 Analyses multidimensionnelles de proximité

La figure 9.3 représente la solution MDS, type MDSCAL, à deux dimensions. Les stimuli à si_2 sont représentés par des carrés, ceux à fa_3 par des triangles, ceux à sib_3 par des étoiles, et ceux à fa_4 par des losanges. Chaque stimulus est relié par un trait aux deux stimuli dont le centroïde spectral est le plus proche et ayant une même F_0 , ainsi qu'aux deux stimuli dont la F_0 est la plus proche et ayant un même centroïde spectral. Les données de dissemblance ont été transformées grâce à une fonction arc-tangente hyperbolique et la solution a été orientée afin que la corrélation entre les stimuli à si_2 et leur centroïde spectral soit maximum.

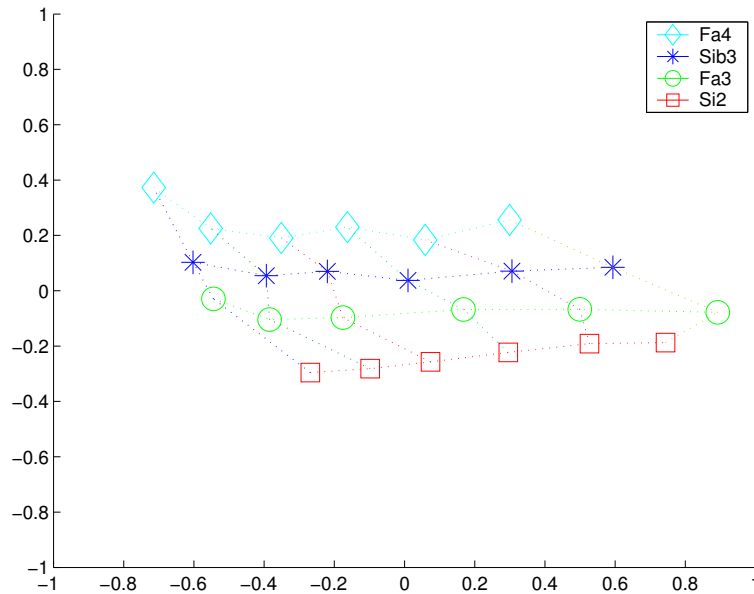


FIG. 9.3: Solution MDS à deux dimensions de l'Exp. VI. Les stimuli à si_2 sont représentés par des carrés, ceux à fa_3 par des cercles, ceux à sib_3 par des étoiles, et ceux à fa_4 par des losanges.

Les stimuli sont clairement organisés en quatre couches parallèles réparties sur deux dimensions. Chaque couche contient des stimuli à même F_0 . À l'intérieur de chaque couche, les stimuli sont ordonnés proportionnellement selon leur centroïde spectral. Les couches à fa_3 et sib_3 sont en légère expansion par rapport à celles à si_2 . La couche à fa_4 arrête cette expansion pour revenir à la même taille que celle à si_2 . Chaque couche semble se décaler vers la gauche en fonction de la F_0 des

stimuli qu'elles contient.

9.3 Implications pour le timbre

Expansion Les analyses ANOVA sur les sous-matrices à F_0 constante deux à deux (AA vs BB, BB vs CC...), tableau 9.3, indiquent si une couche est significativement différente d'une autre. Le tableau 9.1 des dissemblances moyennes indique si cette différence correspond à un élargissement ou à une contraction de la couche. L'effet significatif entre AA et BB et la plus grande dissemblance moyenne de BB montrent que la couche à fa_3 est plus étendue que celle à si_2 . Les effets entre AA et CC et entre BB et CC et les moyennes montrent que la couche à si_3 se contracte par rapport à fa_3 mais reste toujours plus étendue que celle à si_2 . L'effet non significatif entre CC et DD montre que la couche à fa_4 n'est pas significativement différente de celle à si_3 . L'effet non significatif entre AA et DD montre que la couche à fa_4 n'est pas non-plus significativement différente de celle à si_2 (qui elle est différente de celle à si_3). Ce résultat, en première vue contradictoire, montre que l'étendu de la couche à fa_4 se situe entre celle à si_2 et celle à si_3 . L'effet de la F_0 n'est donc pas systématique et il est, par conséquent, délicat à commenter.

Déplacement isométrique Les analyses ANOVA entre les deux triangles des sous-matrices non appariées en F_0 (AB vs BA, AC vs CA...) permettent d'estimer si les déplacements des couches sont significatifs. Les résultats, tableau 9.4, montrent que la F_0 a un effet significatif sur toutes les sous-matrices. Leurs déplacements ne sont donc pas dus au hasard. Il y a bien un effet systématique de la F_0 sur le timbre.

F_0 , chroma, consonance La solution MDS contient une deuxième dimension liée à la hauteur. Cette expérience permet de déterminer l'échelle selon laquelle la F_0 affecte le timbre. En effet, en supposant une influence suivant les chroma, les stimuli à fa_3 et fa_4 auraient été confondus sur la deuxième dimension. Les stimuli à si_2 et si_3 , éloignés d'un chroma, auraient également été très proches. Enfin, les stimuli si_2 et fa_3 , étant séparés par un triton, auraient été à deux positions extrêmes (figure 9.4b). Maintenant, en supposant l'échelle de consonance, les stimuli à fa_3 et fa_4 auraient à nouveau partagé la même position, le si_3 leur étant éloigné d'une quinte, aurait également été très proche et finalement le si_2 aurait été éloigné de tous car opposé dans le cercle des quintes 9.4c).

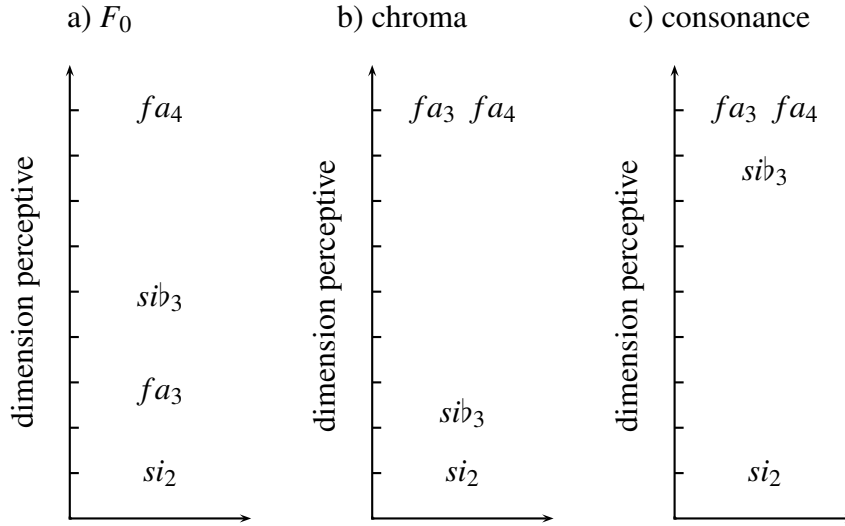


FIG. 9.4: Schéma représentant les trois ordres possibles le long de la seconde dimension selon la dimension de la hauteur avec laquelle elle est liée : a) la F_0 ; b) le chroma ; 3) la consonance.

La deuxième dimension ordonnant les stimuli selon leur F_0 permet d'attribuer l'effet de la F_0 suivant cette échelle (Figure 9.4a).

Amélioration du descripteur de centroïde spectral La solution MDS montre un déplacement significatif vers le côté faible de la dimension corrélée au centroïde spectral. Le panneau a) de la figure 9.5 montre le nuage de points le long de la première dimension, et le centroïde spectral de chaque stimulus. Selon l'axe physique, on retrouve les 6 valeurs de centroïde spectral utilisées pour leur fabrication. Cependant, le long de l'axe perceptif les positions ne suivent pas parfaitement ces valeurs. La corrélation est toutefois très bonne avec un coefficient de 0,91 et donc 82% de variance expliquée ($d.d.l. = 22$; $p < 0,001$).

Une amélioration de ce descripteur est proposée en repassant dans un premier temps des ERBrate aux Hertz par la formule de Hartmann (1997) :

$$\bar{f} = \frac{e^{\frac{\bar{z}}{9,26}} - 1}{0,00437}, \quad (9.1)$$

où $\bar{z}$ est une valeur en ERBrate de notre modèle et $\bar{f}$ sa valeur correspondante en

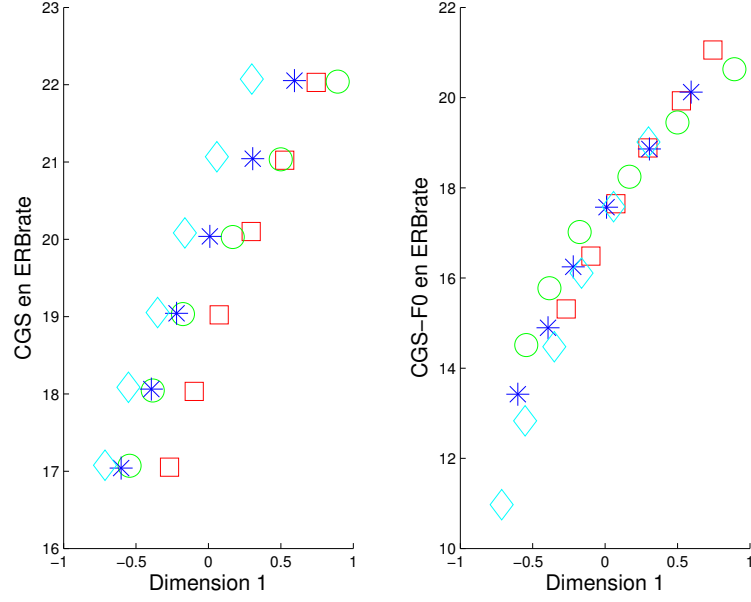


FIG. 9.5: Nuages de points entre la position des stimuli le long de la première dimension pour des descripteurs possibles : a) le centroïde spectral ; b) le centroïde spectral- F_0

Hz. Puis, dans un deuxième temps, la valeur de la F_0 du stimulus est soustraite à $\bar{f}$:

$$\bar{f}_{corr} = \bar{f} - F_0, \quad (9.2)$$

où $\bar{f}_{corr}$ correspond à la nouvelle valeur corrigée de notre modèle. Enfin la valeur est de nouveau retransformée en ERBrate ainsi :

$$\bar{z}_{corr} = 9.26 \times \log(0.00437 * \bar{f}_{corr} + 1), \quad (9.3)$$

Le panneau b) de la figure 9.5 montre le nuage de points entre la première dimension perceptive, et les $\bar{f}_{corr}$. On observe ainsi que les valeurs sont beaucoup mieux alignées le long de la droite de régression. Cet effet est quantifiable par un coefficient de corrélation de 0,97 expliquant plus de 93 % de la variance ($d.d.l. = 22$; $p < 0,001$). Pratiquement, cette correction semble améliorer la prédiction de la différence de timbre entre des stimuli non appariés en F_0 . Théoriquement, cette idée a déjà été évoquée dans le domaine du traitement de la parole (Traunmuller,

1981; Hoemeke et Diehl, 1994), où la distance entre le premier formant et la fréquence fondamentale est proposée comme un paramètre important de la qualité des voyelles. Rappelons également les travaux de Slawson (1968) qui montrèrent que, pour que deux sons séparés d'une octave soient appariés en timbre, leurs deux premiers formants doivent être déplacés de 10 %. Ce résultat montre bien qu'une enveloppe fixe, impliquant un centroïde spectral également quasi-fixe, n'est pas la meilleure solution pour appairer les timbres de deux sons.

9.4 Conclusions

Les résultats de cette expérience montrent que la comparaison de timbre est possible au moins jusqu'à une différence de 18 demi-tons, correspondant à une fois et demi l'octave. Toutefois, les dissemblances moyennes augmentent proportionnellement avec la ΔF_0 . Cette augmentation se manifeste dans la solution MDS par une dimension corrélée avec la F_0 et orthogonale à la dimension du timbre décrite par le centroïde spectral. Sur cette dimension, les stimuli sont ordonnés selon leur F_0 et non selon leur chroma ou consonance. La solution MDS laisse apparaître un effet systématique de la F_0 sur la dimension perceptive liée au centroïde spectral. Cet effet peut être prédit par une nouvelle définition du centroïde spectral améliorée. Enfin, l'expansion en fonction de F_0 observée lors de l'Exp. V ne se généralise pas, elle est donc difficilement interprétable.

Chapitre 10

Expérience VII. L'étalement spectral

L'expérience III avait révélé une troisième dimension du timbre corrélée avec le descripteur de l'étalement spectral. Une expérience similaire à l'Exp. VI a donc été effectuée en ne faisant varier que ce descripteur.

10.1 Méthodes

Le corpus était composé de vingt-quatre stimuli issus de la combinaison de six valeurs d'étalement spectral et quatre F_0 . Les stimuli ont été synthétisés par la même méthode que les stimuli des Exp. V et VI. Les différentes valeurs d'étalement spectral ont été atteintes en faisant varier l'écart type d'une enveloppe spectrale en forme de distribution Gaussienne, comme le montre la figure 10.1.

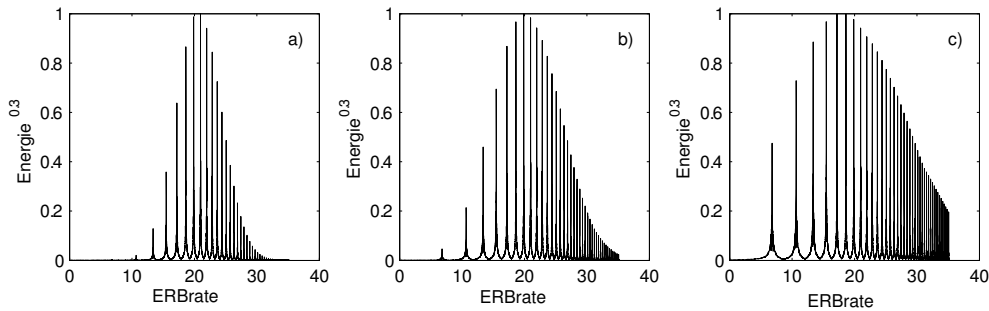


FIG. 10.1: Trois spectres de stimuli de même centroïde spectral ne variant que par leur étalement spectral à : a) 4, b) 5, et c) 6,5 ERBrate

Les stimuli avaient un centroïde spectral de 22 ERBrate, ainsi qu'une enveloppe temporelle et des F_0 identiques à celles de l'Exp. VI. Les 6 valeurs d'étalement spectral étaient : 4, 4,5, 5, 5,5, 6 et 6,5 ERBrate.

Les 14 sujets qui ont passé cette expérience étaient composés du même nombre de musiciens et non musiciens, de 9 femmes et de 5 hommes. L'expérience comportait 256 paires à juger. Les sujets pouvaient prendre une pause au milieu. L'expérience durait en moyenne une heure et demi.

10.2 Résultats et analyses

10.2.1 Effets de sujet et expertise instrumentale

Les corrélations entre les matrices de dissemblance des sujets ont été soumises à une analyse en cluster. Cette analyse a montré que les sujets 4 et 10 formaient une classe séparée, ils ont donc été éliminés du reste des analyses. Aucun effet de l'expertise musicale n'a été trouvé par l'ANOVA. Les données ont donc été moyennées pour les 12 sujets restant, musiciens et non-musiciens.

10.2.2 Matrices de dissemblance

La matrice de dissemblance moyennée à travers les sujets est représentée figure 10.2. Cette matrice possède un aspect similaire à celle de l'Exp. V. Elle peut être décomposée de la même manière en sous-matrices. Chaque sous-matrice montre une augmentation graduelle de la dissemblance lorsqu'on s'écarte de la diagonale.

Le tableau 10.1 représente les dissemblances moyennes dans chaque sous-matrice. Ce tableau indique que les triangles inférieurs ont toujours des moyennes plus faibles que les triangles supérieurs. Les carrés ne semblent donc pas symétriques.

10.2.3 Analyses de variance

Le tableau 10.2 retrace les résultats de l'ANOVA à mesure répétée sur les sous-matrices ne contenant que les paires appariées en F_0 (AA, BB, CC et DD) avec les facteurs F_0 (4) et paires d'instruments¹ (15).

¹Le terme "instrument" sera utilisé lors de cette expérience pour désigner un ensemble de stimuli ayant la même valeur d'étalement spectral.

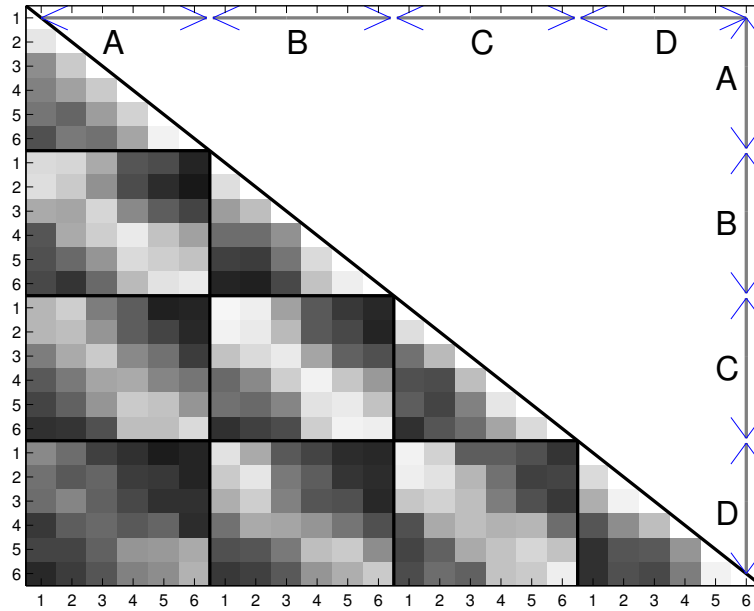


FIG. 10.2: Matrice de dissemblance de l'Exp. VIII. Chaque lettre représente une F_0 utilisée : A) si_2 , B) fa_3 , C) do_3 et D) fa_4 . Chaque chiffre représente une valeur d'étalement spectral dans l'ordre : 4, 4,5, 5, 5,5, 6 et 6,5 ERBrate.

Tous les effets sont significatifs. La contribution à la variance totale des effets dus à la F_0 reste toutefois faible avec un total de R^2 autour de 7 %.

Le tableau 10.3 représente le R^2 des effets significatifs à 0,5 % des ANOVA effectuées entre chacune de ces sous-matrices deux à deux (AA vs BB, AA vs CC...). Les effets dus à la F_0 sont toujours significatifs mais faibles, sauf entre les sous-matrices BB (fa_3) et CC (si_3) pour lesquelles seul l'effet dû aux paires d'instruments est significatif.

Le tableau 10.4 représente le R^2 des effets significatifs des ANOVA effectuées afin de comparer les deux triangles des chaque sous-matrice carrée (AB vs BA, AC vs CA...). Ce tableau montre que l'effet de la F_0 ou son interaction est significatif mais faible pour les sous-matrices AB (si_2-fa_3), BC (fa_3-si_3), BD (fa_3-fa_4) et CD (si_3-fa_4). Pour les sous-matrices AC (si_2-si_3) et AD (si_2-fa_4), seul l'effet des paires d'instruments est significatif.

TAB. 10.1: Tableau récapitulatif des moyennes des sous-matrices de l'Exp. VI.

F_0	A (si_2)	B(fa_3)	C(sib_3)	D (fa_4)
A	0,38	0,55	0,59	0,73
B	0,42	0,49	0,51	0,67
C	0,49	0,39	0,47	0,55
D	0,62	0,47	0,40	0,48

TAB. 10.2: ANOVA entre les sous-matrices contenant les paires appariées en F_0 (AA, BB, CC et DD) de l'Exp. VII

Facteur	d.d.l.	SC	CM	F	p.c.	R2
S	5.44	11	0.49			
F_0	1.32	3	0.44	6.66	0.0012	1.76
$F_0 \times S$	2.18	33	0.07			
PI	35.33	14	2.52	48.73	0.0000	47.02
$PI \times S$	7.98	154	0.05			
$F_0 \times PI$	4.16	42	0.10	2.44	0.0000	5.54
$F_0 \times PI \times S$	18.74	462	0.04			

10.2.4 Analyses multidimensionnelles de proximité

La solution MDS, de type MDSCAL à deux dimensions a été calculée. Elle est représentée figure 10.3. Les données de dissemblance ont été transformées à l'aide d'une fonction arc-tangente hyperbolique et la solution a été orientée afin que la corrélation entre les stimuli à si_2 et leur étalement spectral soit maximum.

Les stimuli sont organisés en 4 couches discriminées selon leur F_0 . A l'intérieur de chaque couche, les stimuli sont bien ordonnés selon leur étalement spectral, toutefois, la répartition est loin d'être linéaire. Les couches sont approximativement parallèles. Nous pouvons noter néanmoins que la couche à fa_3 s'entremêle avec celle à sib_3 . Un léger déplacement vers la gauche semble apparaître avec la F_0 .

10.3. IMPLICATIONS POUR LE TIMBRE

TAB. 10.3: ANOVA entre toutes les sous-matrices à même F_0 , deux à deux (AA vs BB, AA vs CC,...) de l'Exp. VII. Seules les R^2 des effets significatifs à 0,5 % sont représentés.

	A (si_2)			B (fa_3)			C (sib_3)		
	F_0	PI	$F_0 \times PI$	F_0	PI	$F_0 \times PI$	F_0	PI	$F_0 \times PI$
B	2,83	47,78	3,56						
C	2,18	44,98	-	-	56,22	-			
D	2,28	43,12	3,47	-	53,58	4,27	-	48,49	5,59

TAB. 10.4: ANOVA entre les deux triangles des sous-matrices non appariées en F_0 (AB vs BA, AC vs CA...) de l'Exp. VI.

	A (si_2)			B (fa_3)			C (sib_3)		
	F_0	PI	$F_0 \times PI$	F_0	PI	$F_0 \times PI$	F_0	PI	$F_0 \times PI$
B (fa_3)	3,86	48,01	-						
C (sib_3)	-	40,12	-	3,01	51,69	-			
D (fa_4)	-	21,67	-	9,85	27,50	-	5,74	36,87	-

10.3 Implications pour le timbre

Expansion Les analyses ANOVA sur les sous-matrices à F_0 constante deux à deux (AA vs BB, BB vs CC...), tableau 10.3, combinées avec les dissemblances moyennes, tableau 10.1, nous permettent de dire que la couche à si_2 est globalement moins étendue que les trois autres. Celles-ci ne changent globalement pas entre-elles (pas d'effet significatif du facteur principal F_0). La F_0 n'a pas un effet systématique impliquant une dilatation des couches.

Déplacement isométrique Les analyses ANOVA entre les deux triangles des sous-matrices non appariées en F_0 (AB vs BA, AC vs CA...), estiment que la couche à fa_3 s'est déplacée significativement par rapport à si_2 . Les couches à sib_3

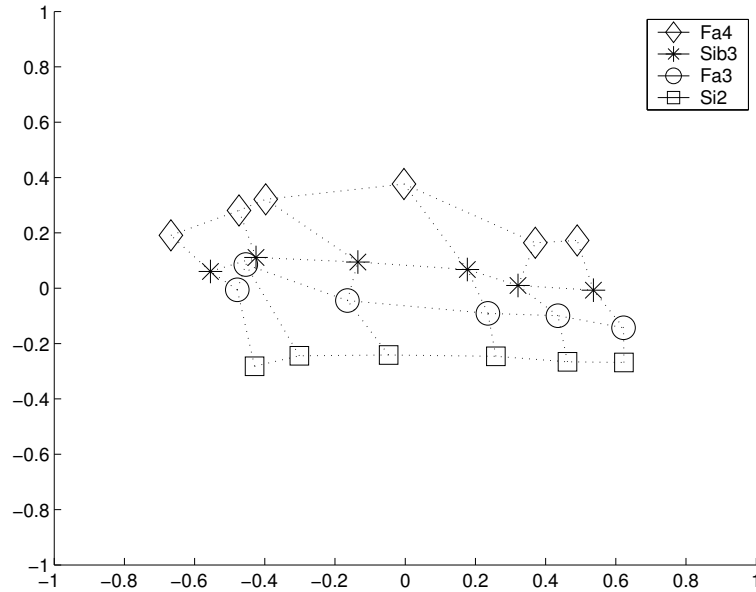


FIG. 10.3: Solution MDS à deux dimensions de l'Exp. VII. Les stimuli à si_2 sont représentés par des carrés, ceux à fa_3 par des cercles, ceux à sib_3 par des étoiles, et ceux à fa_4 par des losanges.

et fa_4 ne révèlent pas de déplacements significatifs par rapport à si_2 et, logiquement, se déplacent par rapport à fa_3 , ou nous devrions plutôt dire que fa_3 se déplace par rapport à sib_3 et fa_4 . Toutefois, les ANOVA soulignent une contradiction en estimant que la couche sib_3 s'est déplacée par rapport à fa_4 . Notons également que ces déplacements n'apparaissent pas de façon flagrante sur la solution MDS. L'interprétation de ces résultats doit donc être considérée avec prudence.

F_0 , chroma, consonance Les quatre couches apparaissent clairement dans l'ordre des F_0 . Aucun effet du chroma ni de la consonance n'a donc été observé.

Amélioration du descripteur d'étalement spectral La première dimension est corrélée avec un coefficient de 0,97 ($d.d.l. = 22$, $p < 0,001$) avec l'étalement spectral, comme le montre la figure 10.4 retraçant le nuage de points entre les deux dimensions (physique et perceptive).

Aucune nouvelle amélioration n'a été trouvée afin de prendre en considération les éventuels déplacements des couches de stimuli.

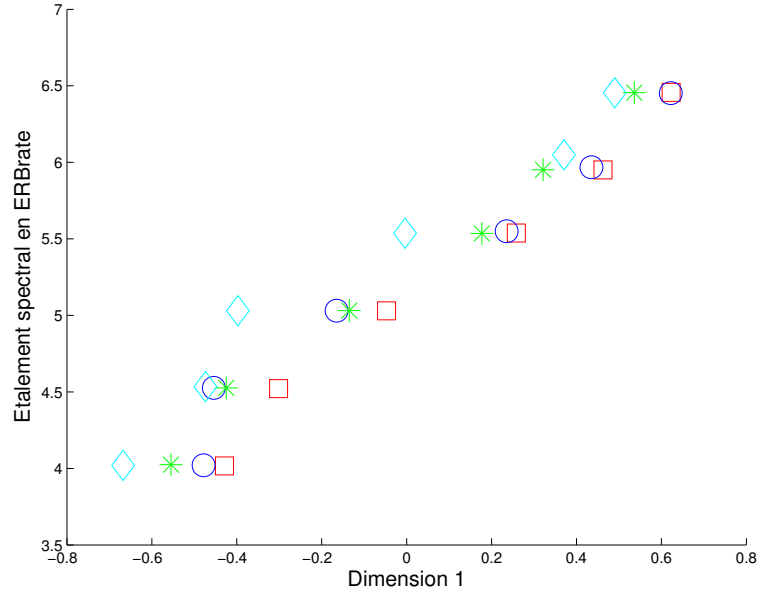


FIG. 10.4: Nuages de points entre la position des stimuli le long de la première dimension et l'étalement spectral

10.4 Conclusions

Cette expérience montre qu'il est possible de faire des comparaisons de timbre avec des stimuli ayant jusqu'à une octave et demi de différence. Toutefois, les sujets ajoutent à leur jugement une composante de dissemblance proportionnelle avec la ΔF_0 comme dans l'Exp. VI. Cette composante se manifeste par une dimension orthogonale à celle du timbre et corrélée avec la F_0 . La structure de la solution MDS est assez bien ordonnée. Elle est composée de quatre couches discriminées par leur F_0 . A l'intérieur de chaque couche, les stimuli sont ordonnés suivant leur étalement spectral.

Chapitre 11

Expérience IV. L'impulsivité

Cette expérience a été effectuée avant l'Exp. V, avec le même protocole expérimental, elle est composée de deux sessions contenant chacune deux F_0 . Elle est dédiée à l'étude de l'effet de la F_0 sur la dimension du timbre décrite par l'impulsivité.

11.1 Méthodes

Cette expérience utilisait 24 stimuli créés à partir de la combinaison de trois F_0 , si_2 , $do\#_3$ et $si\flat_3$ et de huit enveloppes temporelles. Tous les stimuli étaient synthétisés avec une enveloppe spectrale en forme de distribution gaussienne et possédaient la même valeur de centroïde spectral de 17 ERBrate et la même valeur d'étalement spectral de 5 ERBrate.

Les stimuli avaient une enveloppe temporelle de type ADSR. Afin de faire varier le plus simplement possible l'impulsivité, le temps d'attaque de cette enveloppe a été pris comme unique paramètre. La figure 11.1 représente trois enveloppes variant principalement par leur temps d'attaque.

Le tableau 11.1 indique les différents temps d'attaque utilisés et l'impulsivité calculée correspondant.

Les temps d'attaque ont été choisis afin que leur logarithme soit régulièrement espacé. Ce choix a été motivé par l'étude de McAdams et coll. (1995) qui a montré que l'une des dimensions perceptives du timbre était corrélée avec le logarithme du temps d'attaque des stimuli. Toutefois, ce choix se révèle après coup quelque peu maladroit car il met l'accent sur un descripteur légèrement différent de celui que nous désirons étudier. Cependant, cette incohérence n'entache pas les

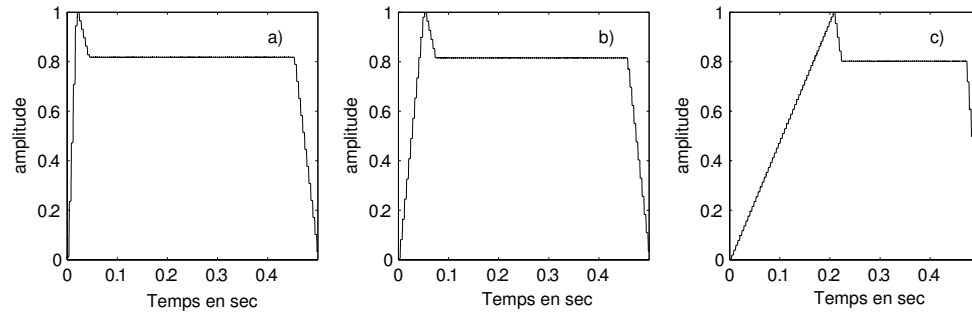


FIG. 11.1: Trois enveloppes temporelles variant principalement par leur temps d'attaque : 1) à 25 ms, 2) 50 ms, et 3) 20 ms

TAB. 11.1: Tableau résumant les valeurs de temps d'attaque choisies ainsi que l'impulsivité correspondante en ms.

Temps d'attaque	Impulsivité
20	403
25	424.5
35	440
50	450
70	458
105	463
145	467
205	470

conclusions de l'expérience.

La tâche, l'interface, ainsi que le déroulement de cette expérience sont semblables à ceux l'Exp. III et V. Dans la session 'a' tous les stimuli ayant une F_0 à si_2 et $do\#_3$ étaient présentés par paires. Dans la session 'b' tous les stimuli ayant une F_0 à si_2 et $si\flat_3$ étaient présentés par paires.

Quatorze sujets ont participé à cette expérience, dont 7 femmes et 7 hommes, 7 musiciens et 7 non musiciens. Aucun d'entre eux n'avait participé aux premières expériences. L'expérience durait environ une heure. L'ordre des sessions 'a' et 'b' était contrebalancé.

11.2 Résultats et analyses

11.2.1 Effets de sujet, ordre et expertise instrumentale

Afin de constater la cohérence des profils de réponses des sujets, une analyse en cluster avec une méthode moyenne a été effectuée à partir des corrélations entre les matrices de dissemblance individuelles. Cette analyse nous a clairement montré que les données des sujets : 1, 10 et 14 étaient très faiblement corrélés avec le reste des données pour la session 'a'. Les sujets 1 et 10 restaient faiblement corrélés pour la session 'b', ainsi que les sujets 7 et 3. Les sujets 1, 3, 7, 10 et 14 ont donc été exclus du reste des analyses. Cette analyse, couplée à une ANOVA, a permis également de constater qu'il n'y avait pas d'effet de l'ordre de passation ni de l'expertise musicale. Les neuf sujets non exclus ont donc été regroupés et moyennés.

11.2.2 Matrices de dissemblance

La matrice de dissemblance de la session 'a', moyennée à travers tous les sujets non écartés, est représentée sur la figure 11.2.

Comme pour les matrices de l'Exp. V, cette matrice peut être décomposée en quatre sous-ensembles de données regroupés chacun dans un triangle, schématisés dans la figure 8.3. Les quatre triangles sont semblables. Ils révèlent tous une augmentation graduelle de leurs dissemblances au fur et mesure que l'on s'écarte de la diagonale. Cette similitude est confirmée par leur coefficient de corrélation deux à deux toujours supérieur à 0,95 ($d.d.l. = 26, p < 0,001$).

La figure 11.3 représente les matrices de dissemblance des deux sessions.

Ces deux matrices sont corrélées avec un coefficient de 0,90 ($d.d.l. = 118, p < 0,001$). Elles sont, en première approximation, proches. Toutefois, en deuxième approximation, on peut noter que le carré de la session 'a' est beaucoup plus symétrique que celui de la session 'b'. Le paragraphe suivant examine ces symétries par des ANOVA.

11.2.3 Analyses de variance

Ces deux matrices ont été soumises à une ANOVA à mesures répétées avec les facteurs ΔF_0 (2) et paires d'instruments¹ (120). Le tableau 11.2 résume les

¹Le terme "instrument" sera utilisé lors de cette expérience pour désigner un ensemble de stimuli ayant la même enveloppe temporelle.

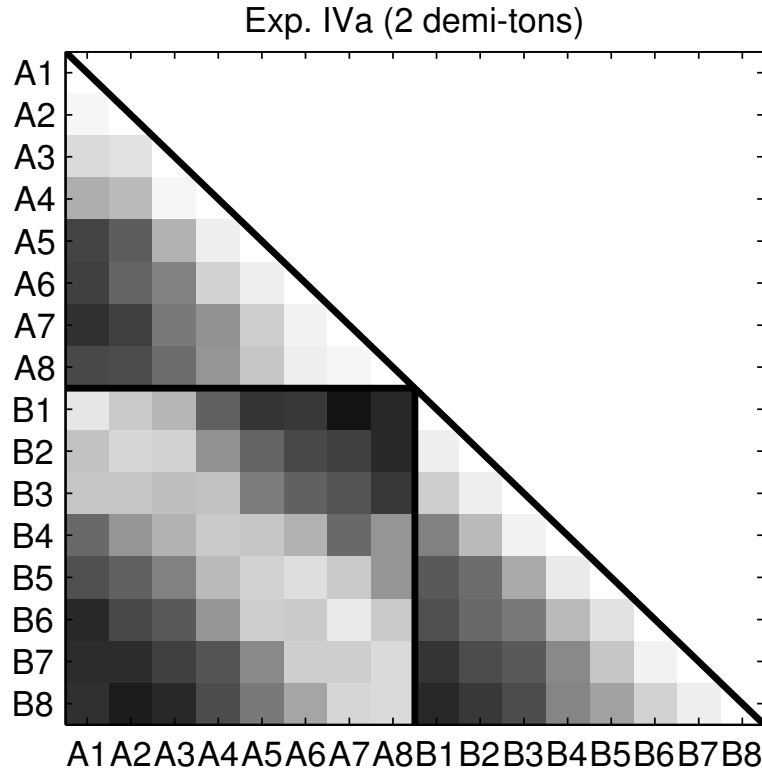


FIG. 11.2: Matrice de dissemblance de l'Exp. IV, pour la session 'a'. Chaque stimulus est représenté par une lettre et un chiffre. La lettre correspond à sa F_0 : A = si_2 , B = $do\#_3$; chaque chiffre correspond à une valeur de temps d'attaque : 20, 25, 35, 50, 70, 105, 145, 205 ms représentée par les chiffres de 1 à 8, respectivement.

résultats. L'effet principal de la ΔF_0 n'est pas significatif, mais son interaction avec les paires d'instruments l'est. Toutefois, cet effet contribue peu à la variance totale des résultats, avec un R^2 autour de 3 %.

Les trois ANOVA sur les parties séparées des matrices, réalisées pour l'Exp. V, ont également été effectuées pour cette expérience. La première teste le triangle AA contre BB ; la deuxième, le triangle AB contre BA ; la troisième, la moyenne des triangles AA et BB contre la moyenne des triangle AB et BA. Le tableau 11.3 résume ces ANOVA par les mesures de R^2 des effets significatifs, pour les sessions 'a' et 'b', séparément.

Aucun effet significatif lié à la F_0 entre AA et BB n'a été trouvé pour les

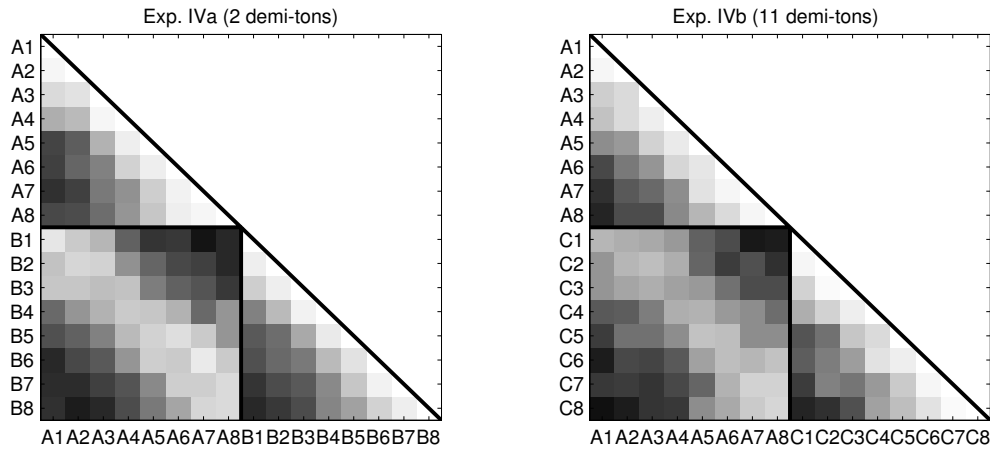


FIG. 11.3: *Panneau gauche : Matrice de dissemblance de la session 'a', même légende que figure 11.2. Panneau droit : Matrice de dissemblance de la session 'b'. La lettre A correspond à une F_0 à si_2 , et la lettre C une F_0 à si_3 ; chaque chiffre correspond à la même valeur de temps d'attaque que dans la figure 11.2.*

deux sessions. Aucun effet n'a non plus été trouvé entre AB et BA. Le seul effet de la F_0 se trouve pour la session 'b' entre les triangles avec et sans ΔF_0 . Ces résultats montrent la grande stabilité de la dimension perceptive du timbre décrite par l'impulsivité face à la F_0 .

11.2.4 Analyses multidimensionnelles de proximité

Les stimuli n'ayant, a priori, pas de spécificités et aucune classe de sujets ne semblant se dégager de l'analyse en cluster, un simple espace MDS de type MD-SCAL a été calculé. La figure 11.4 montre la solution obtenue pour la session 'a'. Les stimuli à si_2 sont représentés par un carré, ceux à $do\#_3$ par une étoile (ou si_3 quand nous considérons la session 'b'). Les dissemblances ont été transformées avec une fonction arc-tangente hyperbolique et la solution a été orientée afin que la projection sur la première dimension des stimuli à si_2 soit le plus corrélée possible avec leur impulsivité.

L'analyse MDS révèle une configuration quasi-unidimensionnelle. Tous les stimuli sont uniformément espacés, dans l'ordre de leur impulsivité, le long de la première dimension. Les stimuli à $do\#_3$ semblent légèrement plus hauts sur la deuxième dimension et plus étendus sur la première. Toutefois, comme nous le

TAB. 11.2: Tableau ANOVA pour Exp. IV. Les abréviations sont les suivantes : *S* pour sujets, ΔF_0 pour la différence de fréquence fondamentale, *PI* pour paire d'instruments, *d.d.l.* pour degré de liberté, *SC* pour Somme des Carrés, *CM* pour Carré Moyen, *F* pour le rapport de Fisher, *p.c.* pour la probabilité critique et R^2 pour le pourcentage de variance induit par chaque effet.

Source	d.d.l.	SC	CM	F	p.c.	R2
S	8	18,04	2,25			
ΔF_0	1	0,00	0,00	0,00	0,9672	0,00
$S \times \Delta F_0$	8	2,01	0,25			
PI	119	154,59	1,30	26,52	0,0000	58,41
$PI \times S$	952	46,63	0,05			
$PI \times \Delta F_0$	119	8,10	0,07	1,84	0,0000	3,06
$PI \times \Delta F_0 \times S$	952	35,31	0,04			

verrons, ces caractéristiques ne sont pas significatives.

La figure 11.5 représente les solutions des deux sessions.

La solution MDS de la session 'b' est proche de celle de la session 'a'. Toutefois, elle se différencie par une deuxième dimension beaucoup plus prononcée discriminant les stimuli par leur F_0 . Elle est donc composée de deux couches parallèles et superposées.

11.3 Implications pour le timbre

Les analyses ANOVA permettent d'évaluer si les caractéristiques des solutions MDS sont significatives.

Expansion Les analyses ANOVA sur les sous-matrices à F_0 constante (AA vs BB) ne se sont pas révélées significatives. Cela implique que les stimuli à si_2 n'ont pas été jugés globalement plus ou moins dissemblables que les stimuli à $do\#_3$ et $si\#_3$. L'expansion observée sur la solution de la session 'a' peut être due au hasard.

11.3. IMPLICATIONS POUR LE TIMBRE

TAB. 11.3: Tableau récapitulatif des R^2 pour les ANOVA intra-sessions, pour les effets significatifs à 0,5 %, pour les sous-matrices de l'Exp. IV

	Session 'a'			Session 'b'		
	F_0	PI	$F_0 \times \text{PI}$	F_0	PI	$F_0 \times \text{PI}$
AA vs BB	-	61,64	-	-	63,98	-
AB vs BA	-	58,58	-	-	41,58	-
[AA et BB] vs AB	-	68,70	-	6,98	58,99	2,89

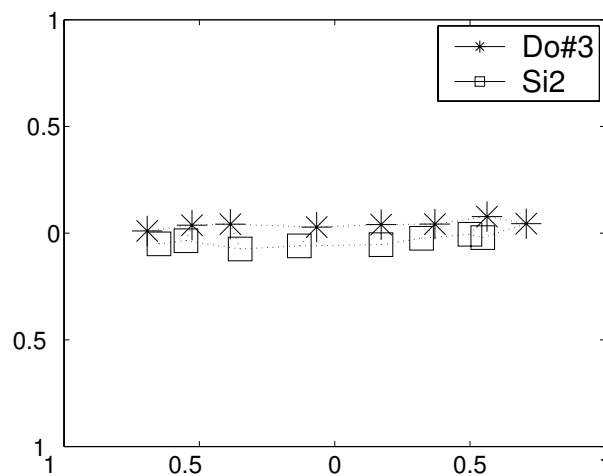


FIG. 11.4: Espace MDS de l'Exp. IV de la session 'a'. Les stimuli à si_2 sont représentés par un carré, ceux à $do\#_3$ par une étoile.

Déplacement isométrique Les analyses ANOVA entre les deux triangles avec une ΔF_0 , AB vs BA, estiment la symétrie des carrés. Aucun effet lié à la F_0 ne s'étant révélé significatif, on peut en conclure que les deux couches n'ont pas subi de déplacement isométrique le long de la première dimension (la seule liée au timbre).

La deuxième dimension La troisième ANOVA entre les triangles avec et sans ΔF_0 , AA et BB vs AB et BA, estime si les deux couches sont confondues. Seul

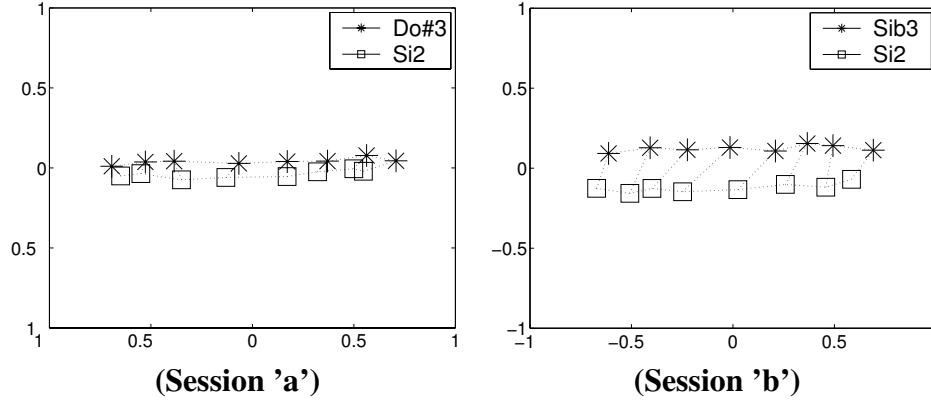


FIG. 11.5: Solution MDS de l'Exp. IV pour la session 'a' (panneau de gauche) et la session 'b' (panneau de droite).

l'ANOVA de la session 'b' révèle des effets significatifs pour le facteur F_0 . On peut en conclure que la session 'a' est unidimensionnelle et la session 'b' bidimensionnelle, et que cette dimension est liée à la F_0 .

11.4 Conclusions

Cette expérience confirme la stabilité de la dimension du timbre décrite par l'impulsivité. Le seul effet significatif se manifeste par l'apparition d'une dimension orthogonale à celle du timbre et corrélée avec la F_0 .

On peut noter que les stimuli sont disposés régulièrement le long de la première dimension. Cet aspect souligne la bonne validité du descripteur du logarithme d'attaque. On peut même se demander s'il ne décrit pas mieux la dimension perceptive que l'impulsivité qui, dans cette expérience, comporte des valeurs espacées de façon logarithmique. D'autres études sur ces descripteurs seraient nécessaires afin de pouvoir les partager, mais cela dépasse le sujet de cette étude.

Chapitre 12

Discussion générale

Ce chapitre vise à récapituler les acquis de l'ensemble des expériences et à discuter les résultats importants.

Afin que ces expériences soient pertinentes pour l'étude du timbre, il était crucial que les sujets puissent effectuer la tâche en ignorant bien les différences de hauteur dans leur estimation du timbre. Les résultats très similaires des Exp. I à III ont montré qu'ils en étaient en grande partie capables.

12.1 Psychoacoustique du timbre

Dans le même esprit que les études classiques, les Exp. I et III produisent des solutions MDS représentant les relations entre les timbres dans un espace à 2 et 4 dimensions. Chacune de ces dimensions est fortement corrélée avec un descripteur physique. Toutefois, les solutions MDS étant orientées dans ce sens, vers ces descripteurs, il est légitime de se demander si les dimensions ont une réalité psychologique.

La question a été abordée en utilisant le logiciel CLASCAL pour former une solution dont l'orientation des dimensions est déterminée sur la base des seules réponses, sans faire appel à des descripteurs. La figure 12.1 nous montre la solution obtenue pour la session 'b' de l'Exp. III.

Cette figure indique qu'au moins 3 des 4 dimensions de la solution présentée au Chapitre 6 correspondent bien à des dimensions de la solution CLASCAL. Les coefficients de corrélation entre les dimensions des solutions CLASCAL et MDSCAL sont respectivement de 1¹, 0,7, 0,5 et 1. En ce qui concerne les deux pre-

¹Les coefficients de corrélations de ce chapitre sont donnés à une décimale près et arrondis

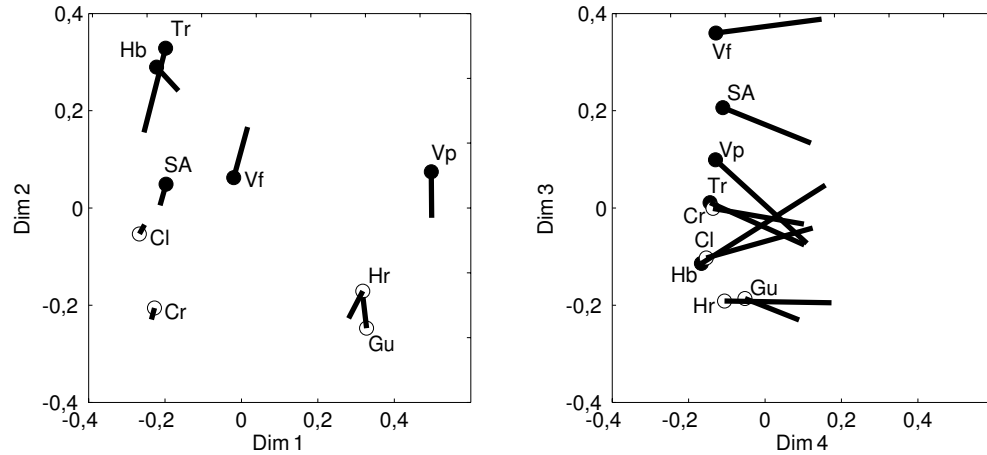


FIG. 12.1: Solution CLASCAL de l'Exp. III pour la session 'b'. Le panneau de gauche représente la première et la deuxième dimension, celui de droite la troisième et la quatrième. L'étiquette représente la position de l'instrument à si_2 et l'extrémité de la ligne la position du même instrument à sib_3 . Les points remplis indiquent les instruments dont les timbres changent de façon significative selon la méthode des ancres (voir Chapitre 3).

mières et la dernière, ces coefficients sont élevés, et suggèrent que les dimensions correspondantes des solutions MDS du Chapitre 6 ont bien une réalité psychologique. La corrélation est plus faible pour la troisième dimension (alignée sur l'étalement spectral dans la solution MDSCAL). Cela pourrait avoir deux explications. L'une serait l'imprécision de l'orientation offerte par CLASCAL, en particulier si le nombre de sujets est faible (ou si les poids qu'ils accordent à chaque dimension sont proches). L'autre serait que cette dimension est composite, et reflète en réalité plusieurs dimensions dont seule une est bien prédite par l'étalement spectral. Cela expliquerait pourquoi les études sont généralement en désaccord sur le descripteur correspondant à la troisième dimension.

Nous nous sommes donc basés sur la solution MDS de McAdams et coll. (1995) produite avec 88 sujets, et nous avons testé nos descripteurs dessus. Ceux-ci se sont montrés aussi performants, voire même plus, que ceux proposés par McAdams et coll. (les corrélations sont effectuées dans le Chapitre 3). Nous avons donc de bonnes raisons de penser que ces descripteurs sont pertinents pour prédire

vers la décimale la plus proche, afin d'éviter la sur-interprétation de ces coefficients.

des dimensions psychologiques.

Les expériences I à III utilisaient de stimuli issus d'instruments de l'orchestre. Bien que la consigne stipule explicitement que si les sujets reconnaissent l'instrument "source" ils ne doivent pas en tenir compte dans leur jugement, il est légitime de se demander s'ils ont effectué cette tâche convenablement. Nous avons, toutefois plusieurs raisons de penser que c'est le cas : premièrement, les instructions ; deuxièmement, le manque de différence entre musiciens et non-musiciens ; troisièmement, les dissemblances ne reflètent pas la structure en classe des instruments ; et quatrièmement, les dissemblances peuvent être bien prédites par les descripteurs. Sur cette base nous pouvons réfuter cette hypothèse et même proposer l'hypothèse inverse : l'identification d'un instrument se fait en grande partie sur la base du timbre, voire de ses variations avec la F_0 .

12.2 Effets de la F_0

Les résultats ont montré que la comparaison de timbres entre des stimuli avec des F_0 différentes est possible. La question de l'effet de la F_0 sur le timbre peut donc être abordée expérimentalement. En première approximation le timbre semble stable face à des variations de F_0 . Mais en deuxième approximation, certains effets, de types différents, peuvent être constatés :

Une dimension liée à la F_0 Les jugements de dissemblance sont affectés par une composante directement liée à la ΔF_0 . Les solutions MDS de l'Exp. IIb, IV et V ont montré que cet effet produit une nouvelle dimension corrélée avec F_0 et orthogonale à celles du timbre. Les Exp. VII et VIII ont confirmé l'existence de cette dimension sur une plus grande gamme de F_0 . Cette composante est négligeable pour une différence d'un ton, pourtant saillante perceptivement, puis devient proportionnelle à la ΔF_0 . Cet effet pourrait être imputable à l'influence de la hauteur dans les jugements de timbres malgré la consigne. Si tel est le cas, les Exp. VII et VIII ont montré que cet effet agit suivant la différence de hauteur tonale et non la différence de chroma ou la consonance.

Cette dimension est toutefois plus petite que ne le laissait entendre Miller et Carterette (1975). Elle n'a pas empêché la comparaison de timbres avec F_0 différentes. Cette contribution étant représentée sur une dimension orthogonale à celles du timbre, elle ne les pas perturbait. Toutefois, l'ajout d'une contribution constante à chaque dissemblance diminue la sensibilité de l'expérience.

La solution CLASCAL citée plus haut a attribué un poids sur chaque dimension pour trois classes différentes de l'Exp. IIIb. Il est intéressant de remarquer que l'une de ces trois classes n'attribue aucun poids à la quatrième dimension liée à la F_0 . Cela montre que les sujets de cette classe ont réussi à effectuer convenablement la tâche en ignorant les différences de hauteurs. Cette classe était constituée de 7 sujets, sur les 18 qui ont effectué l'expérience. Une seule personne de cette classe était non-musicienne.

Une variation dépendante de l'instrument L'analyse par la méthode des ancres a étudié si dans les solutions MDS de l'Exp. III le timbre de stimuli issus du même instrument gardaient la même position dans l'espace. Seul un instrument, Vp, sur 9 pour la session 'a', se révèle à deux positions différentes. Pour la session 'b' le nombre d'instruments qui "bougent" monte à cinq, Vp, Vf, SA, Hb et Tr. Toutefois, l'analyse n'exclut pas un déplacement orthogonal à cet espace (dû par exemple à la hauteur).

Un déplacement de la dimension décrite par le centroïde spectral Les ANOVA sur les sous-matrices des Exp. V et VI ont révélé un effet systématique de la F_0 sur le timbre d'instruments de synthèse créés avec un centroïde spectral fixe. Cet effet se manifeste par un déplacement isométrique de la dimension décrite par le centroïde spectral vers le côté "faible" au fur et à mesure que la F_0 augmente. Cela implique que pour que deux stimuli soient appariés en timbre, il ne faut pas qu'ils possèdent la même valeur de centroïde spectral mais la même valeur de centroïde spectral- F_0 .

On peut considérer que cette dimension du timbre est influencée non plus de façon primaire par le centroïde spectral mais par un autre descripteur correspondant au centroïde spectral- F_0 . En refaisant les Exp. V et VI en se basant sur ce nouveau descripteur, on devrait logiquement ne plus obtenir de déplacement isométrique de cette dimension. L'effet de la F_0 ne serait ainsi plus observé.

12.3 Qu'avons-nous appris sur le comportement des instruments selon la note jouée ?

La méthode des ancres a démontré que le timbre de certains instruments ne changeait pas avec la F_0 . Il est légitime de se demander si cette stabilité incombe à une volonté directe du facteur d'instruments. Pour les autres instruments, leurs

variations de timbre devraient, en principe, être visibles sur une représentation MDS. Toutefois, celle-ci comporte un bruit expérimental important qui est difficilement quantifiable. La figure 6.3 montre par exemple une différence de timbre plus importante pour Gu, estimé stable, que pour SA. On peut donc penser que le déplacement de Gu observé est imputable aux bruits expérimentaux. La méthode des ancres pourrait indiquer le sens du déplacement. En effet, en regardant les changements de dissemblance par rapport à chaque ancre, et en se référant à la solution MDS, on pourrait inférer la direction de changement par rapport aux dimensions du timbre. Ces possibilités n'ont pas été explorées à fond car elles sortent légèrement du cadre de cette thèse. Cependant, les techniques expérimentales et conceptuelles mises en place constituent un outil performant pour l'étude approfondie du comportement du timbre de chaque instrument en fonction de la note.

12.4 Descripteurs - Intérêt pour les applications

Dans l'esprit des expériences classiques sur le timbre, des descripteurs ont été proposés afin de décrire les dimensions trouvées. Une attention particulière a été accordée lors de leur conception à leur robustesse et leur pertinence physiologique. Ces critères étaient au moins aussi importants que le résultat des corrélations avec les dimensions trouvées. Les expériences ont permis de tester leur validité sur une gamme importante de F_0 . En première approximation, ils ont été validés par leur bonne prédiction des dimensions. Ils se sont également montrés stables pour des timbres dont l'analyse par la méthode des ancres a estimé qu'ils partageaient la même position dans l'espace. En deuxième approximation, les Exp. V et VI ont montré qu'il était possible d'améliorer le centroïde spectral afin qu'il prédise mieux les dissemblances de timbres non appariés en F_0 . Aucune amélioration n'a été proposée pour le descripteur d'étalement spectral.

Des études complémentaires seraient nécessaires pour affiner ces descripteurs. Leur validation demande à être poursuivie sur des signaux différents. Il serait bon également d'améliorer le descripteur d'impulsivité. Enfin, une nouvelle exploration de la troisième dimension serait grandement nécessaire.

Ces travaux se révèlent d'une grande utilité pour les applications de type indexation de bases de données sonores ou recherche par le contenu.

12.5 Acquis méthodologiques

Le jugement de timbre de stimuli non appariés en F_0 Ces expériences ont montré que les sujets étaient capables d’ignorer la hauteur d’un son afin de ne juger que son timbre. Ce résultat pourrait paraître trivial au premier abord. En effet, il semble usuel de comparer deux sons de hauteur différente. Tout le monde est capable de dire qu’un son de trompette est différent d’un son de violon, par exemple. Mais lorsqu’il s’agit de juger finement et sur une échelle continue, la tâche se complique. Dans les expériences de Miller et Carterette (1975), le poids que les sujets avaient mis sur la hauteur était si important qu’il masquait les différences plus subtiles du timbre, rendant inadéquante ce type d’expériences. Nos résultats montrent que si la consigne spécifie explicitement d’ignorer la hauteur, les sujets sont capables d’accomplir cette tâche. Du point de vue du protocole expérimental, ce résultat est peut-être le plus important de cette thèse, car il ouvre de nouvelles perspectives aux expériences de dissemblance et aux espaces MDS. En effet, désormais il n’est plus nécessaire d’apparier les stimuli en F_0 , avec les problèmes qui en découlent, afin de créer un espace de timbre.

Toutefois, à partir d’une ΔF_0 importante de cinq demi-tons, correspondant à la quarte des Exp. VI et VII, une dimension apparaît dans la solution MDS, pouvant être prédite par la F_0 . Cela implique que les sujets lors de leurs jugements surévaluent légèrement la dissemblance entre les paires non appariées en F_0 . Mais, d’une part, cette dimension est faible face à celle prédite par l’impulsivité, le centroïde spectral, et l’étalement spectral. Et d’autre part, elle est orthogonale aux dimensions du timbre. Elle ne les affecte donc pas en première approximation. En deuxième approximation, on constate que cette surévaluation diminue la sensibilité du jugement et peut donc affecter les dissemblances faibles. Nous avons vu plus haut que CLASCAL permettait d’extraire la capacité des sujets à ignorer la hauteur. Il serait donc envisageable d’utiliser uniquement ces sujets.

L’analyse multidimensionnelle de proximité Le résultat précédent nous a permis d’obtenir des solutions MDS comportant des stimuli à plusieurs F_0 . Lemaitre (2004) a déjà utilisé ce résultat dans sa thèse afin de créer un espace MDS contenant des timbres d’avertisseurs sonores avec des F_0 très diverses, sans dimensions liées à elles. Il est désormais envisageable de représenter dans un espace MDS le timbre des sons issus d’un même instrument mais à des F_0 différentes. Ou encore d’étudier la qualité des voix des hommes et des femmes dans un même espace.

L’Exp. V a montré que l’utilisation d’une fonction arc-tangente hyperbolique

pouvait améliorer la représentation en compensant le biais introduit par l'échelle bornée de l'interface.

Les analyses de variance Les ANOVA sont peu utilisées dans les études sur le timbre. Leur utilité est double. Elles permettent, d'une part, d'estimer si un effet agit significativement sur les jugements et d'autre part, de quantifier la taille de cet effet. Elles nous gardent de discuter des effets non significatifs ou faibles, même s'ils apparaissent sur les MDS. Elles nous ont permis, entre autre, de quantifier la part de variance due à la F_0 .

La méthode des ancres Cette nouvelle méthode d'analyse permet de tester si les différences de timbres des sons issus d'un même instrument étaient significatives ou, autrement dit, si l'instrument s'est déplacé vis à vis des autres instruments dans l'espace. Elle est efficace mais s'applique dans des conditions bien spécifiques : tous les stimuli doivent être testés dans les deux conditions (ici nous considérons deux F_0) et ils doivent bien échantillonner toutes les dimensions pertinentes du timbre. Lemaitre (2004) a utilisé ce résultat afin de tester si les relations entre les timbres des avertisseurs sonores variaient en fonction de paramètres d'enregistrement.

Le timbre dimension par dimension Les Exp. IV à VII ont exploré le timbre suivant une approche consistant à ne plus considérer toutes les dimensions du timbre en même temps mais dimension par dimension. Cette approche permet de simplifier les hypothèses ainsi que les interprétations. Elle permet de répondre à une question précise par une expérience unique.

12.6 Critique, perspectives

L'ensemble limité des instruments et des F_0 testés Les résultats de nos expériences concernant les instruments de l'orchestre doivent être considérés uniquement dans l'ensemble réduit des instruments utilisés (onze) ainsi que la gamme réduite de F_0 (cinq). Il serait intéressant d'explorer d'autres types d'instruments, comme par exemple des instruments asiatiques ou africains. La gamme des F_0 peut être élargie avec notamment des sons dans les basses ou les hautes fréquences.

Les stimuli issus d'instruments de musique Les variations de timbre observées lors des Exp. I à III restent toujours délicates à expliquer car les paramètres physiques de chaque stimulus ne sont pas contrôlés. L'appartenance au même instrument ne garantit pas l'uniformité du timbre. Le changement de timbre peut donc être de deux ordres : un effet de la F_0 ou une variation physique du stimulus avec la note jouée. Des expériences contrôlant, par analyse-synthèse, les caractéristiques physiques des stimuli issus du même instrument pourraient être effectuées afin de bien séparer ces deux phénomènes.

L'analyse MDS L'analyse MDS, bien que pouvant se montrer très efficace, se révèle également sensible à plusieurs paramètres. Il n'est pas toujours aisé de choisir le modèle approprié et ce choix peut grandement influencer l'interprétation des résultats. Cette analyse comporte des limites pratiques : un nombre de stimuli trop grand augmente déraisonnablement la durée de l'expérience, un nombre trop faible limite le nombre de dimensions à représenter ; un nombre de sujets trop important devient vite une grande charge pour l'expérimentateur, un nombre trop faible limite la puissance statistique d'algorithmes tel CLASCAL.

L'espace de timbre La représentation dans un espace des relations de dissemblance entre des timbres permet de les décomposer en dimensions orthogonales. Elles impliquent que nous décomposions la perception du timbre en plusieurs qualités indépendantes. Ces dimensions ne correspondent pas forcément aux dimensions les plus saillantes, mais aux dimensions permettant la meilleure distinction entre les timbres des stimuli présentés. En considérant cela, il est possible qu'il existe des dimensions importantes du timbre jamais révélées par les MDS. Les dimensions révélées dépendent de la nature des stimuli, expliquant en partie pourquoi les différentes études sur le timbre ne s'accordent pas sur le descripteur de la troisième dimension. Lorsque des descripteurs solides pour les deux premières dimensions seront mis au point, il serait intéressant de synthétiser des stimuli avec des valeurs constantes de ces descripteurs puis de faire juger leurs dissemblances de timbre. Les solutions MDS résultantes devraient révéler les autres dimensions du timbre avec plus de précision.

Les descripteurs La conception des descripteurs a pour but de prédire une dimension psychologique par un algorithme unique. La validité des descripteurs est limitée par plusieurs facteurs comme, par exemple, la variabilité inter-individuelle. Un descripteur bon pour un sujet sera-t-il bon pour un autre ? En deuxième exemple

citons la variabilité intra-individuelle. La perception du timbre d'un son peut être influencée par divers paramètres extérieurs. Un même sujet peut ainsi percevoir le timbre du même son différemment suivant le contexte rendant ainsi difficile la conception d'un descripteur universel. Un vrai travail de fond serait nécessaire afin de les concevoir le mieux possible.

Chapitre 13

Résumé des Conclusions

Cette thèse s'intéresse à la façon dont le timbre musical varie selon la note jouée. Elle s'inscrit dans la tradition des études sur la psychoacoustique du timbre, fondées sur les jugements de dissemblance et l'analyse multidimensionnelle de proximité (MDS). A la différence des études précédentes, faites principalement avec une fréquence fondamentale (F_0) constante, nous avons introduit la F_0 comme paramètre. Sept expériences ont été réalisées ; trois avec des sons d'instruments de l'orchestre et quatre avec des stimuli synthétiques. Dans chaque expérience les sujets jugeaient la dissemblance de timbre entre des sons appariés ou non en F_0 . Les sujets avaient pour consigne d'ignorer les différences de hauteur.

L'un des résultats les plus importants montre que la comparaison du timbre des sons non-appariés en F_0 est possible. Les difficultés rapportées par Miller et Carterette (1975) et l'absence d'autres études de ce type, pouvaient faire craindre que la différence de hauteur perturberait les jugements de dissemblances de timbre. Ce ne fut pas le cas : les sujets ont répondu à peu près de la même façon avec ou sans différence de F_0 . Ce résultat ouvre la voie à des études plus complètes pour explorer le timbre des instruments sur toute leur tessiture.

Il découle des résultats obtenus que, en première approximation, le timbre des instruments de notre corpus varie peu avec la F_0 . Les effets de la F_0 sur le timbre sont donc faibles en comparaison avec les différences de timbre entre instruments. Les mesures de dissemblance à F_0 égal laissaient ouverte la possibilité que le timbre de tous les instruments varie en bloc (hypothèse de déplacement isométrique). Celles faites avec une différence de F_0 ont écarté cette hypothèse.

En deuxième approximation, les analyses ont montré des effets significatifs (bien que faibles) de la F_0 sur le timbre. Les analyses de variance ont permis de les quantifier. Une nouvelle technique d'analyse, la méthode des ancres, a permis de

distinguer les instruments pour lesquelles la variation de timbre était pratiquement nulle des instruments pour lesquels cette variation était significative. Les analyses MDS ont permis de caractériser ces changements en termes de déplacement dans un espace multidimensionnel de timbre.

Les solutions MDS ont révélé également deux autres effets systématiques de la F_0 . D'une part une dimension liée à la F_0 apparaît lorsque la différence de F_0 est suffisamment importante (entre 2 et 5 demi-tons). Cependant, cette dimension est orthogonale et faible par rapport à celles du timbre. D'autre part, la dimension du timbre décrite par le centroïde spectral subit un léger déplacement isométrique.

Les expériences ont permis d'améliorer les descripteurs du timbre en les testant à travers plusieurs F_0 . Une correction sur le descripteur du centroïde spectral peut être introduite afin de compenser le déplacement que subit la dimension qu'il prédit avec la F_0 .

Ces améliorations peuvent contribuer à l'optimisation des applications d'indexation par le contenu de bases de données audio ou multimédia, ainsi que les algorithmes de reconnaissance de source.

Nous avons également l'espoir que ces résultats puissent aider les compositeurs dans leur approche musicale sur le timbre. Cela ne serait qu'un juste retour des choses, la musique ayant inspiré et soutenu fortement cette thèse.

Annexe A

Statistiques des descripteurs pour les instruments utilisés

Cette annexe indique les gammes de variation des descripteurs des instruments de musique, à travers l'analyse de toutes les notes de la tessiture de dix-sept instruments de l'orchestre : l'alto, le basson, la contrebasse, la clarinette, le cor, la flûte, la guitare, la harpe, le hautbois, le piano, le saxophone alto, le trombone, la trompette, le tuba, le violoncelle, le violon frotté, et le violon pizzicato. Les notes des instruments analysées sont issues de la base de données de l'IRCAM : Studio En Ligne. Pour chaque instrument un unique musicien a joué toutes les notes de la tessiture avec la même nuance mezzo forte, le même instrument, dans la même salle et avec les mêmes microphones. Le niveau de chaque son a été adapté afin qu'il possède tous le même niveau rms.

A.1 Impulsivité

Le descripteur de l'impulsivité peut posséder théoriquement une valeur infinie. A titre anecdotique, John Cage a écrit une pièce pour orgue nommée ORGUE2 ASAP ayant une durée de 639 ou certaines notes ont une valeur d'impulsivité des signaux numériques dépassant plusieurs années. A l'opposé, la valeur d'impulsivité la plus faible est limitée par l'échantillonnage. En effet, elle ne peut être inférieure à un échantillon, comme ce serait le cas pour une impulsion de dirac. La figure A.1 trace toutes les valeurs d'impulsivité de chaque note analysée en fonction de sa F_0 . Pour des raisons de temps de calcul, les échantillons ont été tronqués après une seconde et demi. Cela ne nuit pas à l'interprétation des don-

nées, car au-delà d'une seconde, on peut considérer que le son n'est pas du tout impulsif et donc attribuer une valeur à ce descripteur au-delà n'est plus pertinent et devient même redondant avec un descripteur de durée.

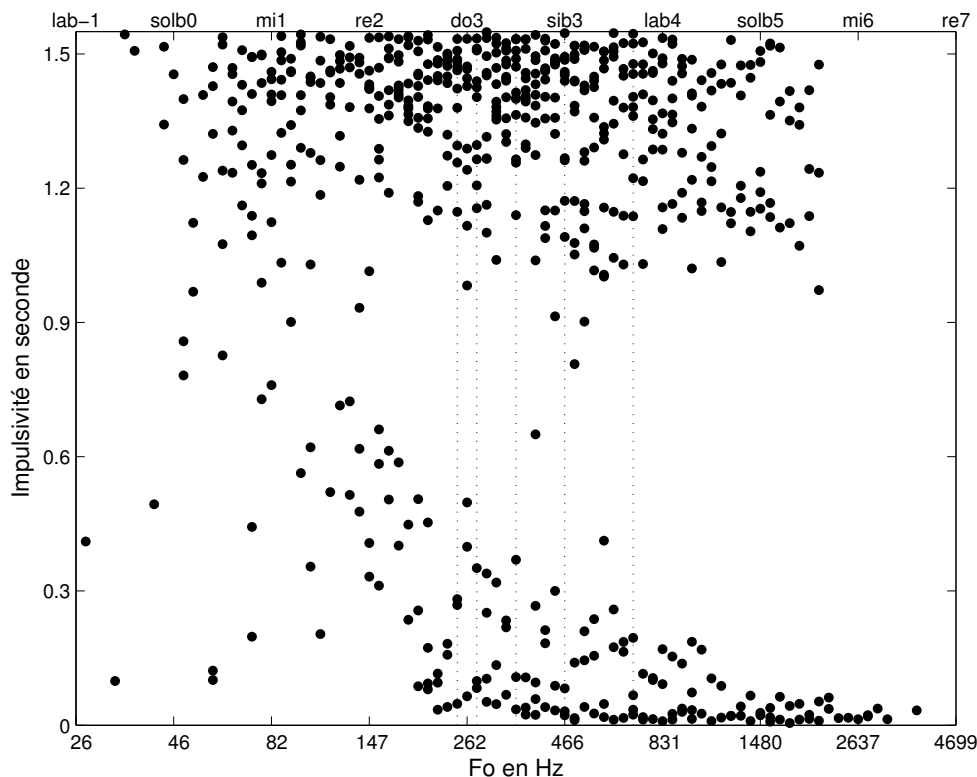


FIG. A.1: Nuages de points représentant l'impulsivité de chaque note issue des dix sept instruments analysés en fonction de leur F_0 . Les lignes verticales en pointillées correspondent aux valeurs de F_0 utilisées dans les expériences de cette thèse, à savoir dans l'ordre : si_2 , $do\#_3$, fa_3 , sib_3 et fa_4 .

L'impulsivité s'étend entre des valeurs proches de zéro et des valeurs proche d'une seconde et demi. On observe bien une ségrégation entre des sons impulsifs et entretenus. La F_0 ne semble influencer ni les sons entretenus, ni les impulsifs au-dessus de 300 Hz. On remarque, en dessous de cette fréquence, pour les impulsifs une légère diminution avec la F_0 . Cet effet est délicat à commenter car rien ne nous affirme qu'il n'est pas dû au hasard. Toutefois, pour les F_0 choisies lors de nos expériences cet effet n'apparaît pas encore clairement. Il est donc possible

que si nous avons utilisé des F_0 plus graves, la stabilité observée le long de la première dimension des Exp. I et III n'aurait pas été aussi forte.

A.2 Centroïde spectral

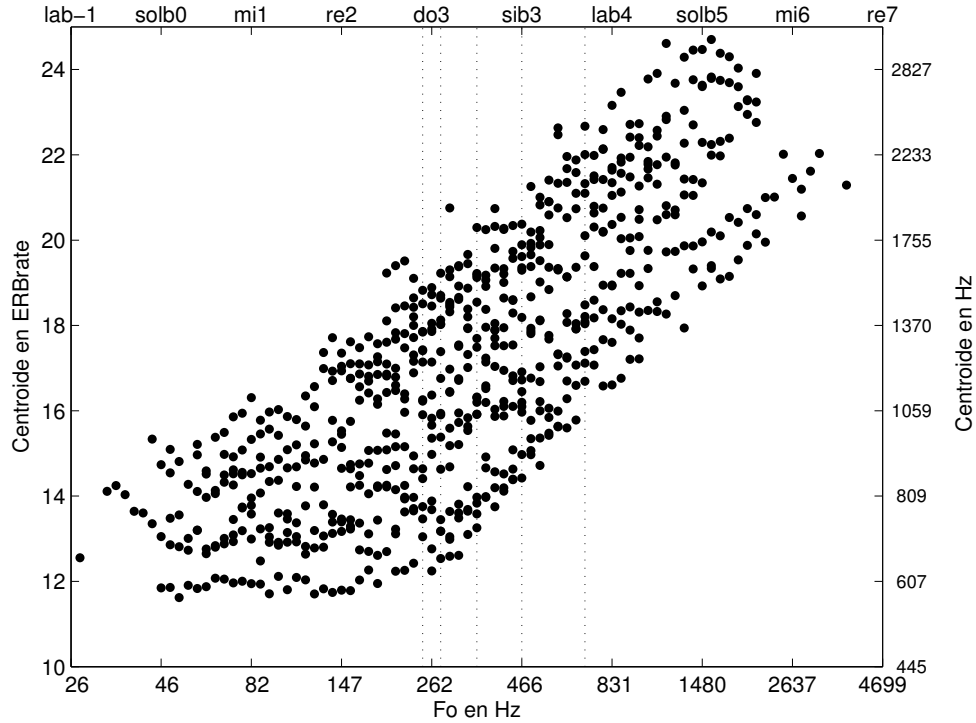


FIG. A.2: Nuage de points représentant le centroïde spectral en ERBrate et en Hz (échelle de droite) de chaque note issue des dix sept instruments analysés en fonction de leur F_0 .

Le centroïde spectral a été extrait des mêmes notes et est représenté figure A.2. Le centroïde spectral varie entre des valeurs de 12 ERBrate et 25 ERBrate. Il est intéressant de remarquer que les points suivent une tendance générale en augmentant avec la F_0 laissant vides deux zones opposées. La première, en haut à gauche, indique qu'il n'y a pas de son instrumental avec une F_0 et un centroïde spectral élevé. Ce type de son présente beaucoup d'harmoniques dans la même bande critique, impliquant une grande rugosité. Il est donc probable que ce phénomène ne

soit pas dû aux limites acoustiques, mais plutôt à une volonté du facteur de ne pas faire des instruments trop rugueux. La deuxième zone vide, en bas à droite, indique qu'il n'y a pas de son instrumental avec une F_0 et un centroïde spectral bas. Cela peut facilement se comprendre en remarquant que pour atteindre ce type de son, il est nécessaire que de l'énergie se situe en dessous de la fréquence fondamentale, ce qui est rarement le cas dans les sons harmoniques. Il est toutefois possible que du bruit, comme le souffle de la flûte, vienne exciter des basses fréquences et fasse descendre le centroïde spectral en dessous de la F_0 . Mais cet effet reste léger car le rapport signal sur bruit est particulièrement faible dans les instruments de musique.

Afin de mieux voir si le centroïde spectral évolue avec la F_0 , nous avons représenté le nuage de points avec seulement un instrument isolé, figure A.3. Pour des raisons de concision du document nous n'avons pas représenté les dix-sept instruments mais un seul par grande famille : le violoncelle pour les cordes frottées, la harpe pour les cordes pincées, le cor pour les cuivres et la flûte pour les bois.

Cette figure montre clairement que le centroïde spectral du violoncelle augmente de façon cohérente de 15 à 22 ERBrate avec la F_0 . La harpe possède un centroïde spectral stable dans les basses fréquences autour de 14 ERBrate. Puis il augmente progressivement à partir d'une F_0 à 260 Hz, pour atteindre une valeur de 22 ERBrate. Le cor contient un centroïde spectral relativement stable, autour de 14 ERBrate. On observe une légère augmentation avec la F_0 à partir de 300 Hz. La flûte possède un centroïde spectral qui augmente de façon cohérente avec la F_0 de 16 ERBrate à 24 ERBrate.

Chacune de ces courbes est représentative des instruments de sa famille. On peut en déduire que le centroïde spectral des instruments évolue bien avec la F_0 , du moins quand celle-ci est supérieure à 200 Hz. Il serait donc intéressant d'effectuer les mêmes types d'expériences mais avec des F_0 inférieures à cette limite.

A.3 Étalement spectral

L'étalement spectral a été extrait des mêmes notes et est représenté figure A.4. L'étalement spectral varie entre des valeurs de 3 ERBrate et 7 ERBrate. Les points qui forment le nuage sont uniformément espacés. Toutes les valeurs d'étalement sont donc, à priori, possibles quel que soit la F_0 . On n'observe aucun effet de la F_0 sur ce descripteur.

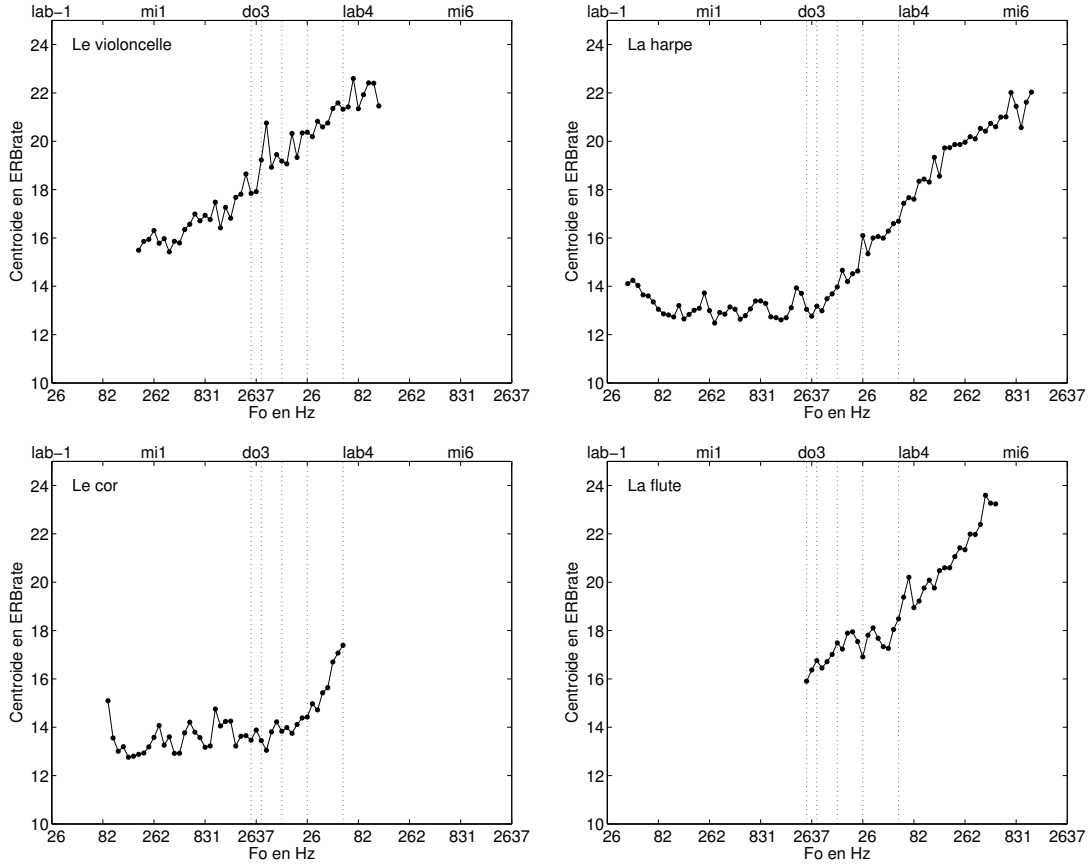


FIG. A.3: Nuages de points représentant le centroïde spectral en ERBrate de chaque note d'un instrument analysé en fonction de ces F_0 . Le panneau haut-gauche représente le violoncelle, le panneau haut-droite, la harpe, le panneau bas-gauche, le cor et le panneau bas-droite, la flûte.

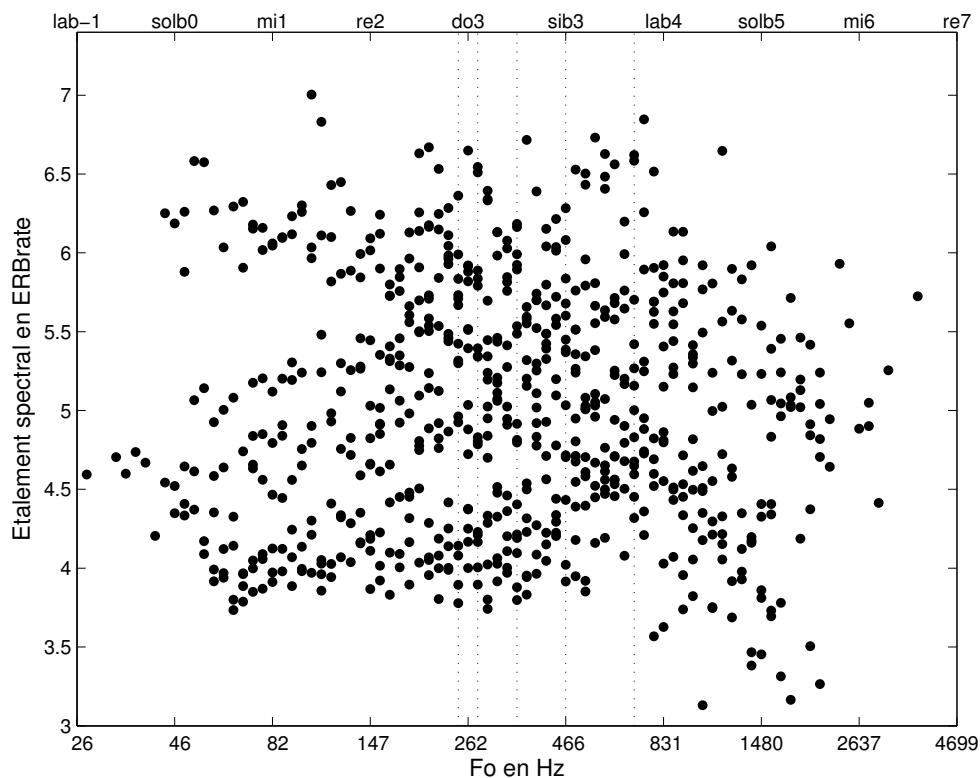


FIG. A.4: Nuages de points représentant l'étalement spectral en ERBrate de chaque note issue des dix sept instruments analysés en fonction de leur F_0 . Les lignes verticales en pointillées correspondent aux valeurs de F_0 utilisées dans les expériences de cette thèse, à savoir dans l'ordre : si_2 , $do\#_3$, fa_3 , sib_3 et fa_4 .

Annexe B

Analyse multidimensionnelle de proximité

L'analyse multidimensionnelle de proximité est l'une des méthodes les plus efficaces pour étudier le timbre. Elle est particulièrement adaptée à son caractère multidimensionnel et permet de dégager certaines de ses dimensions perceptives. Une fois le timbre décomposé, il est possible de trouver les caractéristiques physiques du signal qui l'influencent. Cette analyse sera dénommée MDS par abréviation de son nom anglais Multidimensional Scaling. Ce Chapitre lui est entièrement consacré et se base principalement sur les références suivantes : Borg et Groenen (1997) et Kruskal et Wish (1978)

B.1 La MDS classique

La MDS est une technique permettant de représenter des objets dans un espace multidimensionnel à partir de proximités (similarités ou dissemblances) estimées entre chacun de ces objets. L'espace ainsi obtenu contient n points correspondant aux n objets du corpus : *plus deux objets sont similaires, plus leurs points sont proches dans l'espace*. Cette technique permet de révéler les relations existant entre les objets du corpus.

Afin d'obtenir les coordonnées de chaque point dans cet espace, l'analyse MDS classique procède en trois étapes successives.

B.1.1 Des proximités aux disparités

Les proximités δ peuvent être recueillies de diverses manières. La première étape consiste à transformer, par une fonction f , ces données afin qu'elles puissent être adaptées au mieux pour une conversion vers des distances. Par exemple, si les proximités sont comprises entre -1 et 1, la fonction f pourrait simplement ajouter 1 à toutes les valeurs, afin d'éviter des données négatives ne pouvant pas correspondre à une distance. Ces proximités transformées sont nommées "disparités", $\hat{d}$. Suivant cette fonction f , deux types de MDS sont possibles :

Les MDS métriques Dans les MDS dites "métriques" la fonction f est formée d'opérations algébriques. Dans le cas le plus simple elle devient une transformation linéaire : $\hat{d}_{ij} = a + b \times \delta_{ij}$. Mais elle peut être tout aussi bien non-linéaire : logarithmique, exponentielle, cubique... Une infinité de fonctions peut être utilisée. Le Chapitre 8 montre qu'une fonction arc-tangente hyperbolique peut améliorer une représentation MDS en corrigeant les artefacts possibles dus à des jugements de dissemblance sur une échelle bornée.

Les MDS non métriques Les MDS non métriques imposent juste une relation ordinale entre les dissemblances, de telle sorte que :

$$\text{Si } \delta_{ij} < \delta_{kl} \text{ alors } \hat{d}_{ij} \leq \hat{d}_{kl}. \quad (\text{B.1})$$

Deux approches sont ensuite possibles dans le cas de dissemblances équivalentes : la première approche n'impose aucune contrainte sur les disparités correspondantes, la seconde impose qu'elles soient également équivalentes : si $\delta_{ij} = \delta_{kl}$ alors $\hat{d}_{ij} = \hat{d}_{kl}$.

B.1.2 Des disparités aux distances

L'étape suivante consiste à faire correspondre au mieux un ensemble de disparité à un ensemble de distances d définies à l'intérieur d'un espace de Minkowski à R dimensions ainsi :

$$d_{ij} = \left[\sum_{r=1}^R (X_{ir} - X_{jr})^n \right]^{\frac{1}{n}}, \quad (\text{B.2})$$

où X_{ir} et X_{jr} représentent les positions sur la dimension r des stimuli i et j , respectivement. Le paramètre n est dépendant de la métrique utilisée. Nous ne considérons dans ce document que le cas où n est équivalent à 2, définissant ainsi les

distances dans un espace Euclidien. L'ensemble de ces distances forme une configuration $\mathbf{X}$. Elles doivent suivre les trois propriétés classiques d'une norme :

- **La minimalité** : Cette propriété implique que la distance qui relie un point à lui-même doit être nulle et qu'aucune autre distance ne doit lui être inférieure (i.e. il ne peut exister de distances négatives). $d_{ij} \geq d_{ii} = 0$.
- **La symétrie** : La distance reliant le point i au point j est équivalente à la distance entre le point j et le point i : $d_{ij} = d_{ji}$
- **L'inégalité du triangle** : Afin que trois points (i , j et k) puissent être représentés dans un espace de plus d'une dimension, il est nécessaire que la plus grande des trois distances reliant ces points soit inférieure à la somme des deux autres : soit $d_{ik} > \max(d_{ij}, d_{jk}) \Rightarrow d_{ik} < d_{ij} + d_{jk}$

Une fois le nombre de dimensions fixé par l'utilisateur, une nouvelle contrainte doit être ajoutée afin d'obtenir des distances représentant le mieux les disparités. Kruskal propose une fonction qu'il nomme stress, σ . Les distances doivent minimiser cette fonction pour être optimum.

$$\sigma(\mathbf{X}) = \left[\frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2} \right]^{\frac{1}{2}} \quad (\text{B.3})$$

Afin d'obtenir ces distances, plusieurs algorithmes ont été proposés. Nous les verrons par la suite.

B.1.3 Des distances aux coordonnées

L'étape finale consiste à transformer ces distances en coordonnées. Les MDS classiques n'orientant pas les configurations, l'utilisateur est donc libre de la choisir. Par convention, l'ensemble est centré à l'origine, et l'orientation des axes est donnée par une analyse en composante principale. Le choix du nombre de dimensions a été pris lors de l'étape précédente.

B.2 Les différents algorithmes de MDS

Ces quarante dernières années ont vu la naissance de plusieurs nouveaux programmes de MDS, se basant sur de nouveaux modèles et prenant en considération d'autres paramètres. Cette section fait un rapide tour d'horizon de certains de ces programmes.

B.2.1 MDSCAL

MDSCAL (MultiDimensionnal SCALing) a été proposé par Torgerson. Ce modèle est considéré comme le modèle classique de la MDS. Il lie simplement les disparités à une métrique Euclidienne.

$$d_{ij} = \left[\sum_{r=1}^R (X_{ir} - X_{jr})^2 \right]^{\frac{1}{2}}. \quad (\text{B.4})$$

L'algorithme ne donnant en sortie que les distances entre les points, l'espace obtenu peut être représenté sous toutes les orientations possibles.

La première implémentation fut proposée par Torgerson en 1952. Par une suite d'opérations algébriques, une matrice de distance a été calculée à partir d'une matrice de disparité (pour les détails voir Borg et Groenen, 1997, pages : 207-209). Cette méthode fut populaire car elle permet d'obtenir une solution analytique, ne demandant donc pas d'itérations. Toutefois, la solution trouvée ne constitue pas la configuration optimale : celle qui permet d'obtenir une valeur de stress minimale. Kruskal, en 1964, proposa un algorithme permettant d'obtenir une meilleure configuration, par des itérations successives visant à minimiser la fonction de stress. Cet algorithme fut amélioré par De Leeuw et Heiser en 1977. Ils le nommèrent "SMACOF", pour *Scaling by Maximizing a Complicated Function*. Celui-ci donne de meilleurs résultats car il garantit une convergence monotone vers le minimum de la fonction de stress.

B.2.2 INDSCAL

INDSCAL proposé par Carroll et Chang (1970) attribue une échelle différente par sujet (INDividual Differences SCALing). Il repose sur l'hypothèse que chaque sujet utilise les mêmes dimensions mais ne les pondère pas de la même façon. Ces pondérations sont introduites dans l'équation B.4 par un facteur W_{kr} représentant le poids qu'attribue le sujet k à la dimension r :

$$d_{ijk} = \left[\sum_{r=1}^R W_{kr} (X_{ir} - X_{jr})^2 \right]^{\frac{1}{2}}. \quad (\text{B.5})$$

Cet algorithme délivre en sortie deux matrices. La première contient les coordonnées des stimuli à l'intérieur de l'espace psychologique décrit par les sujets. Le second, les pondérations de tous les participants sur chaque dimension. La

première permet, comme une MDS classique, d'extraire la structure des données. La seconde permet de quantifier et de qualifier les différences de jugement entre les participants. Cet algorithme fixe l'espace dans une orientation bien précise, permettant d'assimiler les dimensions trouvées à des dimensions psychologiques. L'inconvénient principal de cette méthode est qu'il fait apparaître un grand nombre de nouveaux degrés de liberté à déterminer.

B.2.3 CLASCAL

CLASCAL, élaboré par Winsberg et De Soete (1993), reprend le principe de INDSCAL, à la différence près que les poids sur les dimensions ne sont plus attribués à chaque individu mais à des classes d'individus ayant des patterns de réponses corrélés.

$$d_{ijt} = \left[\sum_{r=1}^R w_{tr} (X_{ir} - X_{jr})^2 \right]^{\frac{1}{2}}, \quad (\text{B.6})$$

où w_{tr} représente le poids attribué par la classe t sur la dimension r . Cela permet, comme INDSCAL, d'obtenir un espace MDS orienté. Toutefois, ici, le nombre de paramètres augmente raisonnablement. Ce modèle permet également d'extraire les grandes classes de sujets.

B.2.4 EXSCAL

EXSCAL (EXtended mdSCAL), élaboré par Winsberg et Carroll (1989), permet à chaque stimulus de varier sur une dimension qui lui est propre. Ainsi, si un stimulus possède une caractéristique particulière, cette spécificité n'engendrera pas de dimensions supplémentaires ne contenant qu'elle :

$$d_{ij} = \left[\sum_{r=1}^R (X_{ir} - X_{jr})^2 + S_i + S_j \right]^{\frac{1}{2}} \quad (\text{B.7})$$

où S_i et S_j représentent les spécificités des stimuli i et j , respectivement.

B.2.5 CLASCAL étendu

Winsberg et De Soete (1993) ont intégré à CLASCAL les améliorations apportées par EXSCAL en y incluant les spécificités.

$$d_{ijt} = \left[\sum_{r=1}^R w_{tr} (X_{ir} - X_{jr})^2 + v_t (S_i + S_j) \right]^{\frac{1}{2}} \quad (\text{B.8})$$

où v_t représente le poids accordé par la classe t sur les spécificités. Ce modèle a donné de bons résultats. La thèse de Donnadieu (1997) relate les améliorations qu'il apporte dans l'étude du timbre.

B.3 Les choix de la représentation

Nous avons vu dans la section B.1.2 que l'utilisateur devait choisir le nombre de dimensions de l'espace dans lequel ces données seront représentées. Ce choix est souvent très important dans l'interprétation des données. Il peut être motivé par des considérations théoriques ou pratiques.

Cette section présente deux méthodes différentes pour aider l'utilisateur dans son choix.

B.3.1 La méthode dite du "coude"

Nous avons vu que pour obtenir des distances à l'intérieur d'un espace avec un nombre de dimensions donné, la contrainte d'un stress minimal pouvait être ajoutée. Il est possible ensuite de comparer les différentes valeurs minimales de stress obtenues en fonction du nombre de dimensions choisies. Il semble logique que l'ajout d'une dimension fera toujours baisser le stress. Cependant, si à partir d'un certain nombre de dimensions, l'ajout d'une nouvelle dimension ne fait plus baisser le stress de façon conséquente, son apport sera jugé superflu, et elle sera associée au bruit. Ne pouvant être interprétée cette dimension n'a pas lieu d'être. Le nombre de dimensions choisi sera donc tel qu'elle ne sera pas représentée. Borg suggère de créer un graphique représentant le stress en fonction du nombre de dimensions afin de repérer un coude, indiquant l'ajout de dimensions ne contribuant pas à faire baisser le stress de façon conséquente. La figure B.1 montre la courbe de stress en fonction du nombre de dimensions pour l'analyse MDS de l'Exp. V, Chapitre 8. Un coude apparaît clairement autour de la deuxième dimension, car

B.3. LES CHOIX DE LA REPRÉSENTATION

la représentation à trois dimensions possède un stress seulement légèrement plus faible. La solution à deux dimensions a donc été privilégiée.

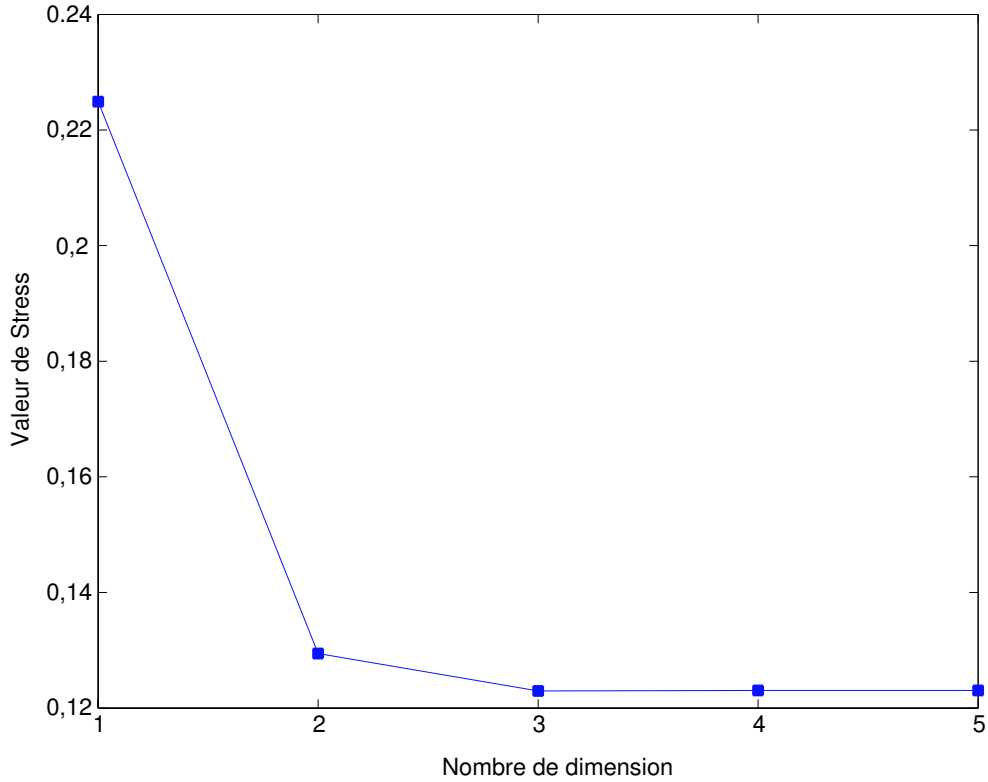


FIG. B.1: Courbe de stress de l'Exp. V en fonction du nombre de dimension

L'expérimentateur doit bien garder à l'esprit que l'ajout de dimensions rend la représentation plus sensible aux bruits expérimentaux. Kruskal et Wish (1978) proposent une règle simple et empirique qui restreint le nombre de dimensions à considérer en fonction du nombre de stimuli à représenter :

$$(n - 1) \geq 4R, \quad (\text{B.9})$$

où n est le nombre de stimuli et R le nombre de dimensions du modèle.

B.3.2 La méthode de maximum de vraisemblance

La fonction statistique de probabilité permet d'estimer les chances d'obtenir un certain échantillon à partir d'une population dont les paramètres sont connus, généralement la moyenne et l'écart type. La fonction de vraisemblance indique à quel degré il est vraisemblable que la population dont est extraite un échantillon donné possède certaines valeurs de paramètres.

Cette méthode statistique peut s'appliquer aux MDS afin de choisir le meilleur modèle et la meilleure configuration. La fonction de vraisemblance, $\hat{L}_\Omega$, indique pour un modèle Ω la probabilité d'obtenir une configuration $d(\mathbf{X})$ ($\pm \Delta d$) à partir des disparités, $\hat{\delta}$. En supposant que l'erreur suive une loi normale (Ramsay, 1997) d'écart type σ , cette fonction est proportionnelle au produit des probabilités de chaque point :

$$\hat{L}_\Omega \propto \prod_{i=1}^N \exp^{-0.5 \times (\frac{d - \hat{\delta}}{\sigma})^2} \Delta d. \quad (\text{B.10})$$

Cette mesure a été améliorée par Schwarz (1978) qui l'a nommée BIC pour Bayesian Information Criterion. Elle prend désormais en compte différents paramètres du modèle comme le nombre de ses degrés de liberté, v_Ω , le nombre de stimuli J , et le nombre de paires P .

$$BIC_\Omega = -2\text{Log}(\hat{L}_\Omega) + 2v_\Omega \text{Log}(N \frac{J(J-1)}{2}). \quad (\text{B.11})$$

Cette mesure est utilisée dans EXSCAL et CLASCAL afin de déterminer le nombre de dimensions, le nombre de classes et si le modèle doit inclure des spécificités.

Annexe C

Tessitures des instruments de l'orchestre

La figure C indique la tessiture des principaux instruments de musique occidentaux. La notation correspond à la notation française : $la_3 = 440$.

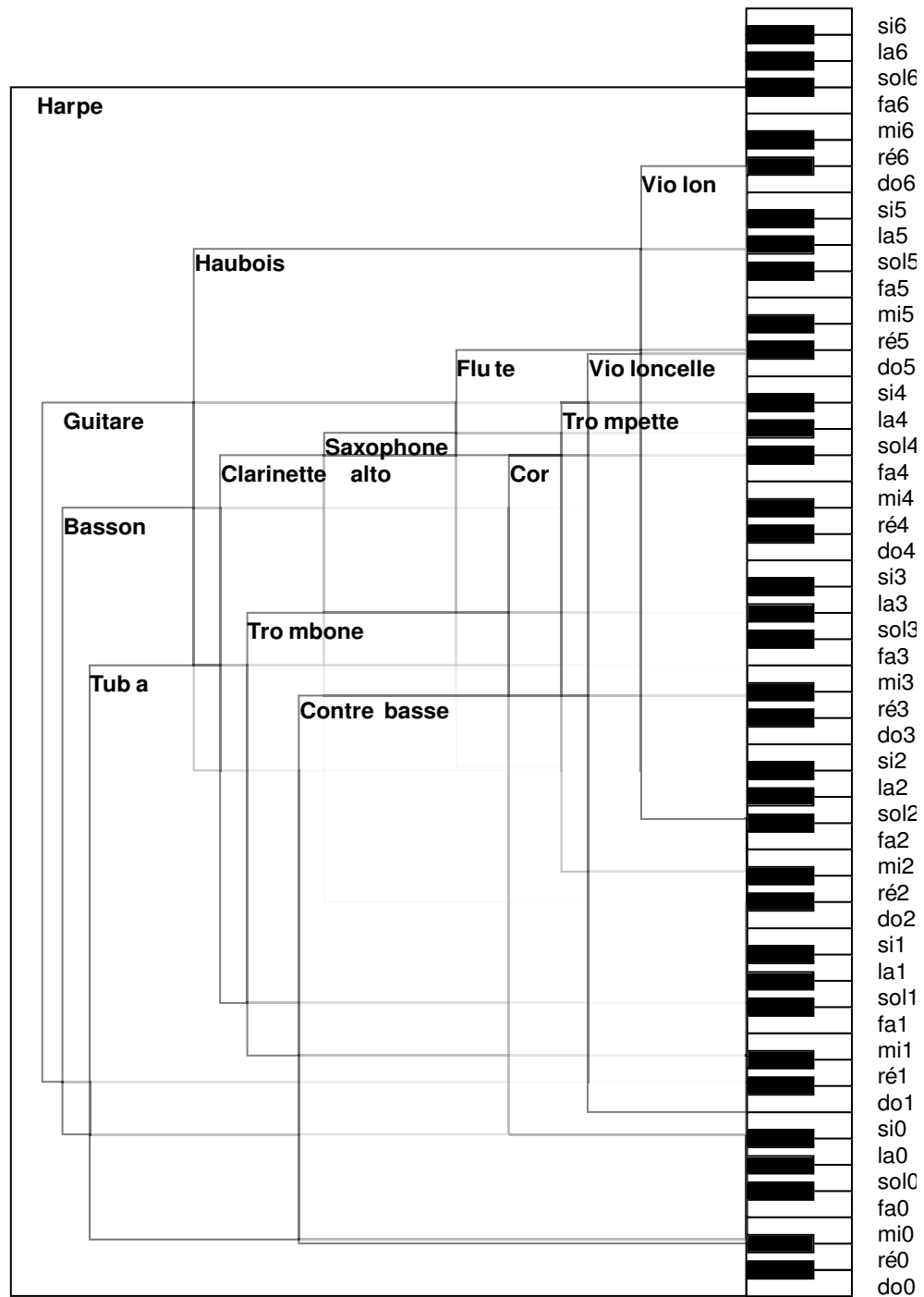


FIG. C.1: Figure illustrant les tessitures des instruments de l'orchestre occidentale

Annexe D

Lexiques

D.1 Qualités perceptives

Le timbre selon L'American Standards Institute (1960) *"Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar ; Timbre depends primarily upon the spectrum of the stimulus, but it also depends on the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus"*

Pouvant être traduit ainsi : *"Le timbre est la qualité perceptive utilisée par l'auditeur pour estimer la différence entre deux stimuli présentés dans les mêmes conditions avec la même sonie et la même hauteur. Le timbre dépend premièrement du spectre fréquentiel du stimulus, mais également de la forme d'onde, de la pression acoustique, de la disposition des fréquences à l'intérieur du spectre et des caractéristiques temporelles du stimulus."*

Le timbre selon l'Association Française de Normalisation (1977) *"Caractère de la sensation auditive qui différencie deux sons de même hauteur et de même intensité et qui permet la reconnaissance de l'origine ou de l'appartenance à un groupe de sons donnés. L'attaque, les composants du son etc. agissent sur le timbre. Physiquement, le timbre est lié à la composition spectrale d'un son et à son évolution"*.

Le timbre-identité Caractère de la sensation auditive qui permet la reconnaissance de l'appartenance à une catégorie de source donnée.

Le timbre-individualité Caractère de la sensation auditive qui permet la reconnaissance de l'appartenance à une source donnée.

Le timbre-qualité Caractère de la sensation auditive qui différencie deux sons de même hauteur et de même intensité. Il est influencé principalement par le centroïde spectral, l'impulsivité, et l'étalement spectral.

La hauteur selon l'Association Française de Normalisation (1977) *"Caractère de la sensation auditive lié à la fréquence d'un son périodique, qui fait dire que le son est aigu ou grave selon que cette fréquence est plus ou moins élevée. Par extension, la notion de hauteur peut s'appliquer avec une facilité plus ou moins grande à des sons non périodiques."*

La sonie Intensité subjective d'un son qui fait dire que le son est fort ou faible selon que la pression est plus ou moins élevée.

La sonie spécifique Répartition de la sonie en fonction de sa position sur une échelle ERBrate.

Brillance Caractère de la sensation auditive qui fait dire que le son est brillant ou mat selon la répartition spectrale de son énergie.

D.2 Caractéristique physique

La fréquence fondamentale, F_0 Inverse de la durée d'une période d'un son périodique.

Centroïde spectral Descripteur d'une dimension du timbre liée à la position moyenne sur une échelle ERBrate de la sonie spécifique.

Etalement spectral Descripteur d'une dimension du timbre liée à l'écart type de la sonie spécifique sur une échelle ERBrate.

Impulsivité Descripteur d'une dimension du timbre liée à la durée pendant laquelle l'enveloppe temporelle dépasse un seuil de 40 % de son amplitude max.

D.3 Divers

Relation psychoacoustique primaire Relation directe entre un phénomène physique et la qualité perceptive qu'il influence principalement, comme les relations : F_0 - hauteur, intensité - sonie, durée physique - durée perceptive.

Relation psychoacoustique secondaire Relation entre un phénomène physique et toute qualité perceptive autre que celle qu'il influence de façon primaire, comme les relations : F_0 - timbre, intensité - hauteur, durée physique - sonie.

Bibliographie

- Abdi, H. (1987). *Introduction au Traitement Statistique des Données Expérimentales*. PUG, Grenoble.
- AFNOR (1977). *Recueil des normes françaises de l'acoustique, Tome 1 (vocabulaire)*, NF S 30-107. Association Française de NORmalisation.
- ANSI (1960). *USA Standard Acoustical Terminology*. American National Standards Institute, New York.
- Bachem, A. (1950). Tone height and tone chroma as two different pitch qualities. *Acta Psychologica*, 7 : 80–88.
- Bagot, J. D. (1999). *Information, Sensation et Perception*. Armand Colin, Paris.
- Beal, A. L. (1985). The skill of recognizing musical structures. *Memory & Cognition*, 13 : 405–412.
- Bjork, E. A. (1985). The perceived quality of natural sounds. *Acustica*, 57 : 186–188.
- Borg, I. et Groenen, P. (1997). *Modern Multidimensional Scaling, Theory and Application*. Springer-Verlag, New York.
- Buja, A. et Swayne, D. (2001). *Visualization Methodology for Multidimensional Scaling*. Manuel d'utilisateur de XGvis.
- Burns, E. (1982). Pure-tone pitch anomalies. I. Pitch-intensity effects and diplacusis in normal ears. *J. Acoust. Soc. Am.*, 72 : 1394 – 1402.
- Caroll, J. D. et Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35 : 283–319.

- Creelman, C. D. (1962). Human discrimination of auditory duration. *J. Acoust. Soc. Am.*, 34 : 582–593.
- de Cheveigné, A. (2004). Pitch perception models. Edité par Plack, C. et Oxenham, A., dans *Pitch*, chapitre 6. Springer Verlag, New York.
- de Cheveigné, A. et Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.*, 111 : 1917–1930.
- Demany, L. (1989). Perception de la hauteur tonale. Dans *Psychoacoustique et Perception Auditive*, chapitre 2, pages 43–82. INSERM, Paris.
- Deutsch, D. (1972). Mapping of interactions in the pitch memory store. *Science*, 175 : 1020–1022.
- Donnadieu, S. (1997). *Représentation Mentale du Timbre des Sons Complexes et Effets de Contexte*. Thèse de Doctorat, Université Paris V.
- Doughty, J. M. et Garner, W. M. (1948). Pitch characteristics of short tones. II : Pitch as function of duration. *J. Experim. Psychol.*, 38 : 478–494.
- Ewert, S. D., Verhey, J. L., et Dau, T. (2002). Spectro-temporal processing in the envelope-frequency domain. *J. Acoust. Soc. Am.*, 112 : 2921–2931.
- Faure, A. (2000). *Des sons aux Mots, Comment Parle-t-on du Timbre Musical ?* Thèse de Doctorat, EHESS, Paris.
- Fletcher, H. (1934). Loudness, pitch and timbre of musical tones and their relations to the intensity, the frequency and the overtone structure. *J. Acoust. Soc. Am.*, 6 : 59–69.
- Fletcher, H. et Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *J. Acoust. Soc. Am.*, 5 : 82–108.
- Fletcher, N. H. et Rossing, T. D. (1998). *The Physics of Musical Instruments (deuxième édition)*. Springer-Verlag, New York.
- Garner, W. R. (1974). *The Processing of Information and Structure*. Potomac, Maryland).
- Gescheider, G. (1976). *Psychophysics. Method and Theory*. Lawrence Erlbaum Associates, New Jersey.

- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, 61 : 1270 – 1277.
- Grey, J. M. et Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *J. Acoust. Soc. Am.*, 63 : 1493–1500.
- Hajda, J., Kendall, R., Carterette, E., et Harshberger, M. (1997). Methodological issues in timbre research. Edité par Deliège, I. et Sloboda, J., dans *Perception and Cognition of Music*, chapitre 12, pages 253–306. Psychology Press, Londres.
- Handel, S. et Erickson, M. L. (2001). A rule of thumb : The bandwidth for timbre invariance is one octave. *Music Percept.*, 19 : 121–126.
- Hartmann, W. (1978). The effect of amplitude envelope on the pitch of sine wave tones. *J. Acoust. Soc. Am.*, 63 : 1105–1113.
- Hartmann, W. (1997). *Signals, Sound, and Sensation*. AIP Press, Woodbury, New York.
- Hartmann, W. M. (1993). Auditory demonstration on compact disk for large N. *J. Acoust. Soc. Am.*, 93 : 1–16.
- Hesse, H. P. (1982). The judgment of musical intervals. Edité par Clynes, M., dans *Music, Mind and Brain*, pages 217–225. Plenum Press, New York.
- Hoemeke, K. et Diehl, R. (1994). Perception of vowel height : The role of F_1 - F_0 distance. *J. Acoust. Soc. Am.*, 96 : 661 – 674.
- Houix, O. (2003). *Structures vibrantes et catégorisation auditive*. Thèse de Doctorat, Université du Maine, Le Mans, France.
- IRCAM (2000). Studio en ligne. [http ://www.ircam.fr](http://www.ircam.fr).
- Iverson, P. et Krumhansl, C. (1993). Isolating the dynamic attributes of musical timbre. *J. Acoust. Soc. Am.*, 94 : 2595–2603.
- Kaufman, L. et Rousseeuw, P. J. (1990). *Finding Groups in Data. An Introduction to Cluster Analysis*. Wiley-Interscience, Bruxelles.
- Kendall, R., Carterette, E., et Hajda, J. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Percept.*, 16 : 265–294.

- Killion, M. C. (1977). Revised estimate of minimum audible pressure : Where is the 'missing 6 dB' ? *J. Acoust. Soc. Am.*, 63 : 1501–1508.
- Kitahara, T., Goto, M., et Okuno, H. (2003). Musical instrument identification based on F_0 -dependent multivariate normal distribution. Dans *Proc. of the International Conference on Speech, Acoustics, and Signal Processing*, volume V, pages 421–424, Hong-Kong.
- Krimphoff, J. (1993). Analyse acoustique du timbre. Mémoire de DEA, Université du Maine/IRCAM. Dirigé par S. McAdams.
- Krimphoff, J., McAdams, S., et Winsberg, S. (1994). Caractérisation du timbre des sons complexes. Analyses acoustiques et quantification psychophysique. *Journal de physique*, 4 : 625–628.
- Krumhansl, C. L. (1989). Why is musical timbre so hard to understand ? Edité par Nielzen, S. et Olsson, O., dans *Structure and Perception of Electroacoustic Sound and Music*, pages 43–53. Excerpta Medica, New York.
- Krumhansl, C. L. et Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology - Human Perception and Performance*, 18 : 739–751.
- Kruskal, J. B. et Wish, M. (1978). *Multidimensional Scaling*. Sage, Californie.
- Lemaitre, G. (2004). *Comment concevoir de nouveaux sons pour les avertisseurs automobiles ? Une approche psychologique et acoustique*. Thèse de Doctorat, Université du Maine, Le Mans, France.
- Marozeau, J., P., de Cheveigné, A., McAdams, S., et Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *J. Acoust. Soc. Am.*, 114 : 2946–2957.
- McAdams, S. et Bregman, A. (1979). Hearing musical streams. *Comput. Music J.*, 3 : 26–60.
- McAdams, S. et Rodet, X. (1988). The role of FM-induced AM in dynamic spectral profile analysis. Edité par Duifhuis, H., Horst, J., et Wit, H., dans *Basic Issues in Hearing*, pages 359–369. Academic Press, Londres.

- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., et Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres : Common dimensions, specificities, and latent subject classes. *Psychol. Res.*, 58 : 177–192.
- Melara, R. D. et Marks, L. E. (1990). Interaction among auditory dimensions : Timbre, pitch and loudness. *Percept. and Psychophys.*, 48 : 169–178.
- Mellody, M. et Wakefield, G. H. (2000). The time-frequency characteristics of violin vibrato : Modal distribution analysis and synthesis. *J. Acoust. Soc. Am.*, 1007 : 598–611.
- Mencl, E. W. (1998). Spectral versus harmonic information for timbre : Pilot and experimental results. *J. Acoust. Soc. Am.*, 103 : 1879–1880.
- Miller, J. R. et Carterette, C. (1975). Perceptual space for musical structures. *J. Acoust. Soc. Am.*, 58 : 711–720.
- Misdariis, N., Smith, B., Pressnitzer, D., Susini, P., et McAdams, S. (1998). Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres. *J. Acoust. Soc. Am.*, 103 : 2812.
- Moore, B. C. J. et Glasberg, B. (1990). Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *J. Acoust. Soc. Am.*, 87 : 2163–2177.
- Moore, B. C. J. et Glasberg, B. R. A. (1996). A revision of zwicker's loudness model. *Acustica*, 82 : 335–345.
- Munson, W. A. (1947). The growth of auditory sensation. *J. Acoust. Soc. Am.*, 19 : 584–591.
- Opolko, F. et Wapnick, J. (1989). *McGill University master samples user's manual*.
- Patterson, R. D. (1990). The tone height of multi-harmonic sounds. *Music Percept.*, 8 : 203–214.
- Patterson, R. D., Robinson, K., Holdsworth, J. McKeown, D., Zhang, C., et Allerhand, M. (1995). Complex sounds and auditory images. Edité par Cazals, Y., Demany, L., et Horner, K., dans *Auditory Physiology and Perception*, pages 429–446. Pergamon Press, Oxford.

- Pedersen, O., Lyregaard, P., et Poulsen, T. (1977). The round robin test on evaluation of loudness level of impulse noise. Rapport Technique 22, Acoustics Laboratory, Copenhagen Technical University of Denmark.
- Peeters, G. (1999). Audio CE for instrument description (timbre similarity). Rapport Technique ISO/IEC JTC1/SC29/WG11/MEPG99/m5422, International Organization for Standardization.
- Plack, C. et Moore, B. (1990). Temporal window shape as a function of frequency and level. *J. Acoust. Soc. Am.*, 87 : 2178 – 2187.
- Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. Edité par Plomp, R. et Smoorenburg, G., dans *Frequency Analysis and Periodicity Detection in Hearing*, pages 397–414. A. W. Sijthoff, Leiden.
- Pols, J. C. W., Van Der Kamp, L. J., et Plomp, R. (1969). Perceptual and physical space of vowel sounds. *J. Acoust. Soc. Am.*, 46 : 458–467.
- Pressnitzer, D. (1994). *Rugosité auditive de sons complexes : perception, physiologie et modélisation*. Thèse de Doctorat, Université Paris 6, France.
- Puria, S., Peake, W. T., et Rosowski, J. J. (1997). Sound-pressure measurements in the cochlear vestibule of human-cadaver ears. *J. Acoust. Soc. Am.*, 101 : 2754–2770.
- Ramsay, J. O. (1997). Multiscale manual. Rapport technique, McGill University.
- Risset, J. C. (1971). Paradoxe de hauteur : Le concept de hauteur sonore n'est pas le même pour tout le monde. Dans *Proc. of Seventh International Congress on Acoustics*, pages 613–616, Budapest.
- Risset, J. C. (1978). Hauteur et timbre des sons. Rapport Technique 11, Rapport de l'Ircam, Paris.
- Risset, J. C. et Wessel, D. (1999). Exploration of timbre by analysis and synthesis. Edité par Deutsch, D., dans *The psychology of Music*, chapitre 5, pages 113–169. Academic Press Series, New York.
- Roads, C. (1996). *The Computer Music Tutorial*. The MIT Press, Cambridge, MA.

- Robinson, D. et Dadson, R. (1956). A re-determination of the equal loudness for pure tones. *Br. J. Appl. Phys.*, 7 : 166–181.
- Robinson, K. (1993). Brightness and octave position : Are changes in spectral envelope and in tone height perceptually equivalent? *Contemporary Music Review*, 9 : 83–85.
- Sandell, G. et Chronopoulos, M. (1997). Perceptual constancy of musical instrument timbres ; generalizing timbre knowledge across registers. Edité par Gabrielsson, A., dans *Proc. of the Third Triennial ESCOM Conference*, pages 222–227, Suède.
- Scheirer, E. et Slaney, M. (1997). Construction and evaluation of a robust multi-feature speech/music discriminator. Dans *Proc. of the International Conference on Speech, Acoustics, and Signal Processing*, pages 1331–1334, Munich. IEEE.
- Schonemann (1983). Some theory and results for matrices for bounded response scale. *Journal of Mathematical Psychology*, 27 : 311–324.
- Schouten, J. (1970). The residue revisited. Edité par Plomp, R. et Smoorenberg, G., dans *Frequency analysis and periodicity detection in hearing*, pages 41–58. Sijthoff, Leiden.
- Schwarz, G. (1978). Estimating dimension of a model. *Ann. Stat.*, 6 : 461–464.
- Semal, C. et Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *J. Acoust. Soc. Am.*, 89 : 2404–2410.
- Shepard, R., N. (1964). Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.*, 36 : 2346–2353.
- Shepard, R., N. (1982). Geometrical approximations to the structure of musical pitch. *Psy. Rev.*, 89 : 305–333.
- Shepard, R. (1999). Structural representation of musical pitch. Edité par Deutsch, D., dans *The psychology of Music*, chapitre 11, pages 343–390. Academic Press Series, New York.
- Singh, P. et Hirsh, I. (1992). Influence of spectral locus and F_0 changes on the pitch and timbre of complex tones. *J. Acoust. Soc. Am.*, 92 : 2650–2661.

- Slaney, M. (1993). An efficient implementation of the Patterson-Holdsworth auditory filter bank. *Apple Computer Technical Report*, 35.
- Slawson, A. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *J. Acoust. Soc. Am.*, 43 : 87–101.
- Smith, B. (1995). PsiExp : an environment for psychoacoustic experimentation using the IRCAM Musical Workstation. Edité par Wessel, D., dans *Society for Music Perception and Cognition Conference*, Berkeley. Université de Californie.
- Starr, G. E. et Pitt, M. A. (1997). Interference effects in short-term memory for timbre. *J. Acoust. Soc. Am.*, 102 : 486–494.
- Stevens, S. (1975). *Psychophysics*. Wiley-Interscience, New York.
- Stevens, S. S. (1935). The relation of pitch to intensity. *J. Acoust. Soc. Am.*, 6 : 150–154.
- Stevens, S. S. (1955). The measurement of loudness. *J. Acoust. Soc. Am.*, 27 : 815–829.
- Stevens, S. S. et Volkmann, J. (1937). A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Am.*, 8 : 185–190.
- Susini, P. (1999). *Perception Evolutive et Globale de Sons Non-Stationnaires*. Thèse de Doctorat, Université du Maine, Le Mans, France.
- Susini, P., Perry, I., Vieillard, S., Winsberg, S., McAdams, S., et Rodet, X. (2001). Sensory evaluation of air-conditioning noise : Sound design and psychoacoustic evaluation. Dans *Proc. of the International Congress on Acoustics*, Rome, Italie.
- Terhardt, E. (1974). Pitch, consonance, and harmony. *J. Acoust. Soc. Am.*, 55 : 1061 – 1069.
- Traunmuller, H. (1981). Perceptual dimension of openness in vowels. *J. Acoust. Soc. Am.*, 69 : 1465–1475.
- Ueda, R. D. et Ohgushi, K. (1987). Perceptual components of pitch : Spatial representation using a multidimensional scaling technique. *J. Acoust. Soc. Am.*, 82 : 1193–1200.

- Vershuure, J. et van Meeteren, A. (1975). The effect of intensity on pitch. *Acustica*, 32 : 33–44.
- Von Bismarck, G. (1974). Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30 : 159–172.
- Warren, J. D., Uppenkamp, S., Patterson, R. D., et Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc. Natl. Acad. Sci.*, 100 : 10038–10042.
- Wedin, L. et Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scand. J. Psych.*, 13 : 228–240.
- Wessel, D. (1973). Psychoacoustic and music : A report from Michigan state university. *PACE : Bulletin of the Computer Arts Society*, 30 : 1–2.
- Wessel, D. (1979). Timbre space as a musical control structure. *Comput. Music J.*, 3 : 45–52.
- Wessel, D. L. (1978). Timbre space as a musical control structure. Rapport Technique 12, Rapport de l'Ircam, Paris.
- Winsberg, S. et Carroll, J. D. (1989). A quasi-nonmetric method for multidimensional scaling via an extended euclidean model. *Psychometrika*, 54 : 217–229.
- Winsberg, S. et De Soete, G. (1993). A latent class approach to fitting the weighted euclidean model, CLASCAL. *Psychometrika*, 58 : 315–330.
- Wonnacott, T. H. et Wonnacott, R. J. (1990). *Introductory Statistics for Business and Economics, Fourth Edition*. John Wiley and Sons, New York.
- Zwicker, E. et Fastl, H. (1990). *Psychoacoustics : Facts and Models*. Springer-Verlag, New York.