THESE de DOCTORAT de l'UNIVERSITE PARIS 6

Spécialité :

Acoustique, Traitement de signal, Informatique Appliqués à la Musique

Sujet de la thèse :

# PERCEPTION DE RUGOSITE PSYCHOACOUSTIQUE : D'UN ATTRIBUT ELEMENTAIRE DE L'AUDITION A L'ECOUTE MUSICALE

présentée par M. Daniel Pressnitzer pour obtenir le grade de DOCTEUR de l'UNIVERSITE PARIS 6 devant le jury composé de MM. :

Stephen McAdams, Directeur
Pascal Challande
Laurent Demany, Rapporteur
Joël Gilbert
Jean-Claude Risset, Rapporteur

# Remerciements

*Mes remerciements vont tout d'abord très sincèrement à Stephen McAdams, directeur de cette thèse, pour son enthousiasme scientifique, son ouverture d'esprit et ses encouragements constants.*

*Je tiens aussi à remercier Jean-Claude Risset, directeur de la formation doctorale Atiam. La création de cette formation sous son impulsion a seule rendu ce travail possible.*

*Merci à Laurent Demany pour son intérêt et ses encouragements, ainsi que pour avoir accepté d'être rapporteur de cette thèse.*

*Merci à Pascal Challande, président du jury, ainsi qu'à Joel Gilbert pour avoir accepté de juger ce travail.*

*Merci à Hugues Vinet, directeur scientifique de l'Ircam, pour m'avoir accueilli au sein d'un environnement aussi stimulant.*

*Merci à Suzanne Winsberg pour avoir éclairé en quelque phrases certaines idées statistiques de base auxquelles j'avais obstinément résisté pendant toute une scolarité.*

*Merci à Joshua Fineberg pour nous avoir permis de disséquer un bout de sa musique.*

*Merci à toutes les personnes remarquables que j'ai eu la chance de cotoyer dans l'équipe Perception et Cognition musicales : Mondher Ayari, Michael Beauvois, Sylvie Benoit, Caroline Bey, Sophie Donnadieu, Anne Faure, Olivier Houix, Cécile Marin, Vincent Roussarie et Patrick Susini.*

*Merci à Philippe Depalle et Marie-Thérèse Join pour leur engagement généreux dans la formation ATIAM.*

*Merci à tous les membres de l'Ircam qui se sont de près ou de loin intéressés à ce travail, ou que j'ai à un moment ou à un autre accablés de questions : Gérard Assayag, Olivier Warusfel, René Caussé, Xavier Rodet, Marie-Hélène Serra, Laurent Ghys et Jean-Paul Coulon. Merci à Rodolphe Bailly et Carolyn Rodgers pour m'avoir permis d'assouvir mon goût déraisonnable pour les vieux articles.*

*Merci à MiJin Park, Didier Lechenne, et tout le collectif "Comme chacun de nous était plusieurs, cela faisait déjà beaucoup de monde" pour m'avoir permis de partager leurs aventures artistiques. Merci aussi à Jacopo Baboni-Schilingi pour donner un visage plus humain à la musique contemporaine !*

*Merci à Bennett Smith pour le tiroir en haut à gauche de son bureau.*

*Merci à Laurent Cerveau, Philippe Derogis, Peter Hanappe et Marc-Pierre Verge pour avoir rendu la vie nocturne à l'Ircam — imprévisible.*

*Merci à Holle Kirchner pour avoir partagé chaque étape de ce travail.*

*Enfin, merci à ma famille, Dodie, Joëlle et Gil Pressnitzer, pour leur affection et leur soutien.*

# Résumé

Le thème de cette thèse est l'étude expérimentale de l'une des facettes de notre perception des sons, appelée rugosité psychoacoustique. La question des causes du percept est tout d'abord posée. L'intensité de la sensation de rugosité évoquée par un son donné semble dépendre d'un grand nombre de paramètres acoustiques. Les théories actuelles, regroupées sous les noms d'approches "spectrale" ou "temporelle", s'opposent quant aux paramètres pertinents. Ces approches sont confrontées au cours de deux expériences grâce à des manipulations de phase de sons synthétiques. Les résultats conduisent à rejeter l'approche spectrale et à proposer une révision de l'approche temporelle, basée sur un modèle computationnel inspiré de mécanismes physiologiques. Le lien entre les sons élémentaires traditionnellement utilisés dans le laboratoire et la rugosité évoquée par des sons complexes est ensuite abordé, grâce à une troisième expérience portant sur la rugosité de l'addition de sons élémentaires. Il apparaît que la rugosité de sons complexes ne peut pas simplement être estimée par la somme des rugosités des sons les constituant. Un mécanisme sensible à la cohérence entre enveloppes temporelles doit être prise en compte. Un modèle fonctionnel de l'effet est proposé. Enfin, la rugosité est étudiée dans une situation d'écoute musicale, avec des sons instrumentaux. Une théorie proposée par Helmholtz relie les notions de consonance et dissonance dans la musique tonale occidentale à la rugosité des accords habituellement employés. Deux nouvelles expériences proposent une extension des théories de Helmholtz à un contexte musical non tonal. L'étude du percept de rugosité illustre la variété des mécanismes sensoriels et cognitifs qu'il convient de prendre en compte si l'on désire comprendre ce qui se passe entre l'instant ou un son parvient à nos oreilles et le moment ou nous pouvons "l'entendre".

# Abstract

The subject of this thesis is the experimental study of one aspect of sound perception called psychoacoustic roughness. The question of the causes of the percept is first raised. The intensity of the roughness sensation seems to depend on a wide range of acoustical parameters. The existing theories, grouped under the terms "spectral" and "temporal" approaches, diverge according to which parameters are really relevant. These approaches are confronted in two experiments thanks to phase manipulation of synthetic sounds. The results lead to a rejection of the spectral approach and to a modification of the temporal approach, based on a computational model inspired by physiological mechanisms. The link between the sounds traditionnaly used in the laboratory and the roughness produced by complex sounds is considered in a third experiment on the roughness of the addition of elementary tones. It appears that roughness of complex sounds cannot be simply deduced from the roughnesses of their constituent sounds. A mechanism sensitive to correlation between enveloppes has to be considered. A functionnal model of the effect is proposed. Finally, roughness is studied in a musical listening situation, with instrumental sounds. A theory proposed by Helmholtz links the notions of consonance and dissonance in Western tonal music with the roughness of the chords usually employed. Two new experiments propose an extension of Helmholtz's theory to a non tonal context. The study of roughness illustrates the variety of sensorial and cognitive mechanisms to be taken into account if one wants to understand what happens between the instant when a sound reaches our ears and the moment we can "hear" it.

# Table des matières

**1 Introduction** **14**
1.1 Contexte de l'étude . . . . . 15
1.1.1 Sensations élémentaires . . . . . 15
1.1.2 Attributs perceptifs de sons complexes . . . . . 15
1.1.3 Ecoute musicale . . . . . 16
1.2 Plan de l'étude . . . . . 16

**Partie Théorique** **17**

**2 Perception de rugosité** **18**
2.1 Description originale de la rugosité . . . . . 19
2.1.1 Rugosité de la superposition de deux sons purs . . . . . 19
2.1.2 Questions ouvertes par la théorie de Helmholtz . . . . . 22
2.2 La rugosité comme dimension du timbre . . . . . 23
2.2.1 Description des dimensions du timbre . . . . . 24
2.2.2 Similarité perceptive . . . . . 25
2.2.3 Corrélats verbaux de la rugosité . . . . . 26
2.3 Approche spectrale . . . . . 28
2.3.1 Rugosité d'une paire de sons purs . . . . . 28
2.3.2 Influence des paramètres acoustiques spectraux . . . . . 30
2.3.3 Interprétations . . . . . 32
2.3.4 Hypothèse Spectrale . . . . . 34
2.4 Approche temporelle . . . . . 35
2.4.1 Rugosité de sons purs modulés en amplitude . . . . . 35
2.4.2 Influence des paramètres acoustiques temporels . . . . . 36
2.4.3 Interprétations . . . . . 39
2.4.4 Hypothèse Temporelle . . . . . 43
2.5 Lien entre les approches spectrales et temporelles . . . . . 43
2.5.1 Convergence des indices acoustiques . . . . . 43
2.5.2 Divergences entre les approches : influence de la phase . . . 44
2.6 Conclusion . . . . . 47

**3 Rugosité de sons quelconques 48**
3.1 Combinaison de rugosités . . . . . 48
3.1.1 Rugosités partielles . . . . . 48
3.1.2 Eléments en faveur d'une addition entre régions fréquentielles 49
3.1.3 Éléments opposés à une addition entre régions fréquentielles 52
3.1.4 Résultats expérimentaux contradictoires . . . . . 56
3.1.5 Bilan sur la combinaison de rugosités . . . . . 57
3.2 Estimation de rugosité de sons quelconques . . . . . 58
3.2.1 Modèle analytique de Helmholtz . . . . . 58
3.2.2 Modèles spectraux . . . . . 61
3.2.3 Modèles temporels . . . . . 62
3.3 Conclusion . . . . . 65

**4 Perception de consonance musicale 66**
4.1 Consonance et dissonance . . . . . 67
4.1.1 L'énigme posée par Pythagore . . . . . 67
4.1.2 Consonance abstraite . . . . . 68
4.1.3 Consonance d'intervalles . . . . . 69
4.1.4 Consonance fonctionnelle . . . . . 70
4.1.5 Vocabulaire . . . . . 71
4.2 Théories psychoacoustiques . . . . . 72
4.2.1 Théorie de Helmholtz basée sur la rugosité . . . . . 72
4.2.2 Autres théories psychoacoustiques de la consonance . . . . 79
4.3 Théories cognitives . . . . . 82
4.3.1 Hiérarchies tonales . . . . . 82
4.3.2 Théorie Générative de la Musique Tonale . . . . . 86
4.4 Évaluations dans le cadre de la musique tonale . . . . . 87
4.4.1 Perception de consonance d'intervalles . . . . . 87
4.4.2 Perception de mouvements de tension et détente . . . . . 89
4.4.3 La notion composite de consonance . . . . . 90
4.5 Extensions . . . . . 91
4.5.1 Connaissances implicites et indices sensoriels . . . . . 91
4.5.2 Gammes et tempéraments . . . . . 93
4.5.3 Sons inharmoniques . . . . . 95
4.5.4 Discussion . . . . . 98
4.6 Conclusion . . . . . 99

**Partie Expérimentale** **100**

**5 Influence de la phase** **101**
5.1 Introduction . . . 102
5.1.1 Rappels sur les approches spectrales et temporelles . . . 102
5.1.2 Influence possible de la phase . . . 103
5.2 Expérience 1 . . . 104
5.2.1 Hypothèses expérimentales . . . 104
5.2.2 Méthode . . . 105
5.2.3 Résultats . . . 110
5.3 Discussion . . . 110
5.3.1 Influence de la valeur *rms* de l'enveloppe sur la rugosité . 110
5.3.2 Sons de combinaison . . . 112
5.3.3 Facteurs temporels . . . 115
5.4 Dérivations analytiques . . . 118
5.4.1 Expression analytique d'enveloppes par transformée de Hilbert . . . 118
5.4.2 Application des formules analytiques aux sons pAM . . . 122
5.4.3 Effet du filtrage auditif sur les sons pAM . . . 125
5.5 Conclusion . . . 128

**6 Influence de la forme de modulation** **129**
6.1 Introduction . . . 130
6.2 Expérience 2 . . . 131
6.2.1 Hypothèses expérimentales . . . 131
6.2.2 Méthode . . . 133
6.2.3 Résultats . . . 135
6.3 Discussion . . . 137
6.3.1 Influence de la profondeur de modulation . . . 137
6.3.2 Influence de la forme de l'enveloppe . . . 137
6.4 Discussion générale des Expériences 1 et 2 . . . 138
6.4.1 Effet de la dispersion au niveau de la membrane basilaire . 138
6.4.2 Asymétrie temporelle . . . 140
6.4.3 Effet des mécanismes d'adaptation . . . 141
6.4.4 Rugosité et perception d'enveloppe . . . 145
6.5 Application possible aux sons musicaux . . . 146
6.6 Conclusion . . . 148

**7 Influence de la cohérence de modulation** **149**
7.1 Introduction . . . 150
7.1.1 Divergences autour de l'effet de cohérence d'enveloppe . . 150
7.1.2 Modulation effective des sons SAM co- et anti-phasiques . 151
7.2 Modulation d'amplitude irrégulière . . . 153

7.2.1 Sons jAM . . . . . 153
7.2.2 Expérience préliminaire . . . . . 155
7.2.3 Addition des sons jAM . . . . . 157
7.3 Expérience 3 . . . . . 157
7.3.1 Hypothèses expérimentales . . . . . 157
7.3.2 Méthode . . . . . 159
7.3.3 Résultats . . . . . 161
7.4 Discussion . . . . . 163
7.4.1 Effet de la cohérence d'enveloppe . . . . . 163
7.4.2 Evolution de la rugosité avec l'écart fréquentiel . . . . . 167
7.4.3 Comparaison d'enveloppes à travers les canaux auditifs . . 169
7.5 Conclusion . . . . . 170

**8 Tension et rugosité 172**
8.1 Introduction . . . . . 173
8.1.1 Tension musicale dans un contexte tonal . . . . . 173
8.1.2 Contexte non tonal . . . . . 174
8.1.3 Rugosité et consonance musicale . . . . . 176
8.2 Expérience 4 . . . . . 178
8.2.1 Méthode . . . . . 178
8.2.2 Résultats . . . . . 182
8.2.3 Analyses supplémentaires . . . . . 184
8.3 Expérience 5 . . . . . 187
8.3.1 Fusion et spécificités . . . . . 187
8.3.2 Méthode . . . . . 190
8.3.3 Résultats . . . . . 190
8.3.4 Réplication des analyses supplémentaires . . . . . 192
8.4 Modélisation . . . . . 193
8.5 Discussion générale . . . . . 195
8.5.1 Tension et rugosité . . . . . 195
8.5.2 Comparaison entre Expériences 4 et 5 . . . . . 197
8.5.3 Remarques sur la procédure expérimentale et les méthodes d'analyse . . . . . 198
8.5.4 Perception de tension musicale dans une musique non tonale 199
8.6 Conclusion . . . . . 201

**9 Conclusion 202**
9.1 Synthèse des résultats obtenus . . . . . 202
9.2 Mécanismes temporels dans la perception auditive . . . . . 204
9.2.1 Rugosité et modulations d'amplitude . . . . . 204
9.2.2 Modélisation . . . . . 204
9.3 Psychoacoustique et écoute musicale . . . . . 205
9.4 Epilogue . . . . . 206

**Annexes** **208**

**A Méthode expérimentale et analyses statistiques** **209**
A.1 Plan expérimental . . . 209
A.1.1 Méthodes couramment employées . . . 209
A.1.2 Méthode de jugements binaires par comparaisons par paires 211
A.2 Analyse Bradley-Terry-Luce . . . 212
A.2.1 Principe . . . 212
A.2.2 Calcul des valeurs BTL . . . 213
A.3 Analyse par bootstrap . . . 214
A.3.1 Motivations du choix de la technique de bootstrap . . . 214
A.3.2 Principe . . . 215
A.3.3 Application aux valeurs BTL . . . 217
A.4 Analyse par composantes principales . . . 218
A.4.1 Principe . . . 218
A.4.2 Application à l'estimation de cohérence entre auditeurs . . . 218
A.5 Conclusion . . . 220

**B Implémentation d'un modèle d'estimation de rugosité** **221**
B.1 Introduction . . . 221
B.2 Description du modèle . . . 221
B.3 Évaluation . . . 223
B.4 Remarques sur les différentes étapes . . . 224
B.5 Conclusion . . . 225

**C Liste des abbréviations utilisées** **226**

**Références bibliographiques** **228**
Bibliographie . . . 228
Publications liées à la thèse . . . 245

# Table des figures

2.1 La double sirène employée par Helmholtz . . . . . 20
2.2 Les qualités perceptives d'une paire de son purs . . . . . 21
2.3 Dyade de sons purs . . . . . 29
2.4 Rugosité d'intervalles selon Guthrie et Morrill (1928) . . . . . 30
2.5 Intervalles sans rugosité et intervalles de rugosité maximum pour une dyade de sons purs . . . . . 31
2.6 Rugosité d'une dyade de sons purs en fonction de l'écart fréquentiel normalisé par la bande critique . . . . . 33
2.7 Son pur sinusoïdalement modulé en amplitude . . . . . 35
2.8 Influence de la profondeur de modulation sur la rugosité d'un son SAM . . . . . 36
2.9 Influence de la fréquence de modulation sur la rugosité de sons SAM 38
2.10 Influence de la fréquence de modulation et de la fréquence centrale sur la rugosité de sons SAM . . . . . 39
2.11 Fréquence de modulation sans rugosité et de rugosité maximale pour les sons SAM . . . . . 42
2.12 Comparaison de la rugosité d'un son SAM et d'un son FM en fonction de la fréquence de modulation . . . . . 45

3.1 Rugosité de sons FM en fonction de la fréquence de modulation . 51
3.2 Rugosité de bruit blanc sinusoïdalement modulé en amplitude . . 52
3.3 Conditions co-phasiques et anti-phasiques . . . . . 53
3.4 Rugosité de la superposition de deux sons SAM . . . . . 54
3.5 Rugosité de bruit filtrés . . . . . 55
3.6 Modèle de Helmholtz (1877) . . . . . 59
3.7 Calculs de rugosité d'intervalles musicaux selon Helmholtz (1877) 60
3.8 Le principe du modèle spectral de Hutchinson et Knopoff (1978) . 61
3.9 Le principe du modèle temporel de Aures (1985) . . . . . 64

4.1 Le monocorde . . . . . 67
4.2 Coïncidences entre partiels de sons harmoniques . . . . . 73
4.3 Influence du registre sur la rugosité d'intervalles . . . . . 74
4.4 Influence du spectre sur la rugosité d'intervalles . . . . . 76
4.5 Hiérarchies tonales perçues pour des accords de trois notes . . . . 83
4.6 Relations entre théories perceptives et "consonance musicale" . . 92
4.7 Rugosité d'intervalles entre sons de la gamme de Pierce (1966) . . 95

5.1 Sons pAM . . . . . 106
5.2 Résultats de l'Expérience 1 . . . . . 109
5.3 Résultat de l'expérience de contrôle de l'effet de bande critique . . 111
5.4 Effet possible du premier produit de distorsion cubique . . . . . 113
5.5 Filtrage des sons pAM . . . . . 117
5.6 Addition dans le plan complexe de deux composantes d'amplitudes inégales . . . . . 121
5.7 Construction graphique de l'enveloppe des sons pAM . . . . . 124
5.8 Construction graphique de l'enveloppe des sons pAM filtrés . . . . 126
5.9 Réponse en amplitude et en phase du filtre gammatone . . . . . 127

6.1 Stimuli de l'Expérience 2 . . . . . 134
6.2 Résultats de l'Expérience 2 . . . . . 136
6.3 Effet de l'adaptation et de la compression-sons pAM . . . . . 143
6.4 Effet de l'adaptation et de la compression-modulation asymétrique 144

7.1 Composition spectrale de l'addition de deux sons SAM . . . . . 152
7.2 Addition de deux sons SAM dans les conditions co-phasique et anti-phasique . . . . . 152
7.3 Sons jAM . . . . . 155
7.4 Résultats de l'expérience préliminaire . . . . . 156
7.5 Composition spectrale de l'addition de deux sons jAM. . . . . 158
7.6 Addition de deux sons jAM dans les conditions corrélés et non-corrélés . . . . . 158
7.7 Résultats de l'Expérience 3 . . . . . 162
7.8 Effet de la cohérence d'enveloppe . . . . . 163
7.9 Estimation des méthodes de combinaison entre rugosités partielles 166
7.10 Interprétation de l'influence de $\Delta_f$ sur la rugosité globale . . . . . 168

8.1 Partition de l'extrait de *Streamlines* (J. Fineberg) étudié dans les Expériences 4 et 5 . . . . . 175
8.2 L'échelle de tension calculée par l'algorithme du compositeur . . . 177
8.3 Onde temporelle et spectre d'amplitude du timbre orchestral $T_7$ . 179
8.4 Résultats de l'Expérience 4 . . . . . 183
8.5 Analyse par composantes principales des résultats de l'Expérience 4 186
8.6 Onde temporelle et spectre à long terme du timbre orchestral $T_7$ après remixage . . . . . 189
8.7 Résultats de l'Expérience 5 . . . . . 191
8.8 Estimation de rugosité par modèle pour les huit timbres . . . . . 194

B.1 Principe du modèle de rugosité implémenté . . . . . 222
B.2 Évaluation du modèle . . . . . 224

# Liste des tableaux

4.1 Les consonances pythagoriciennes, VIème siècle av. J.C. . . . . 69
4.2 La classification des consonances selon Jean de Garlande, XIIIème siècle . . . . . 70
4.3 La classification des consonances selon Jean de Murs, fin du XIVème siècle . . . . . 71
4.4 La classification des consonances selon Helmholtz (1877) . . . . . 77

# Chapitre 1

# Introduction

*"Dans ces ténèbres, ce vide et ce silence qui entourent le ciel et la terre, seuls planent quelques êtres intérieurs, colorés et sonores."*

G. T. Fechner (1879)

Nous sommes confrontés, dans notre vie quotidienne, à une foule de stimulations acoustiques. Malgré la grande complexité de cet environnement sonore, nous pouvons généralement en extraire des informations utiles : nous parvenons à reconnaître la cause d'un bruit soudain, à suivre un message verbal que nous communique un interlocuteur, et même à éprouver du plaisir en écoutant de la musique dont les sonorités ne représentent pourtant rien de concret. La mécanique et l'acoustique permettent de décrire les modalités de production de tout ces évènements, leur propagation, leurs réflections... Toutefois, pour comprendre comment ces phénomènes purement physiques font de nous des êtres "colorés et sonores", il est indispensable de s'intéresser aux particularités de la perception auditive humaine (McAdams et Bigand, 1994).

Le thème de notre recherche concerne l'un des aspects de cette perception. A l'écoute de certains sons, comme ceux obtenus en soufflant dans un sifflet d'arbitre ou en appuyant simultanément sur deux touches adjacentes d'un piano, nous éprouvons une sensation particulière qui a été nommée "rugosité" (von Helmholtz, 1877). Cette qualité sonore, au nom légèrement effrayant, permet en fait d'aborder plusieurs niveaux des problématiques liées à la perception auditive.

## 1.1 Contexte de l'étude

### 1.1.1 Sensations élémentaires

Pour étudier les sensations évoquées par un son donné, comme l'intensité ou la hauteur, des paradigmes définis par la psychophysique peuvent être utilisés (Fechner, 1860). Une perspective intéressante consiste ensuite à tenter d'expliquer les effets observés en rapport avec ce que nous savons de la physiologie du système auditif (Pickles, 1988).

Pour ce faire, l'étude de la rugosité passe d'abord par des stimuli "de laboratoire", pour lesquels l'influence de divers paramètres acoustiques sur l'intensité de la sensation est explorée. Plus qu'un simple adjectif pour décrire qualitativement une expérience d'écoute ponctuelle, nous verrons que plusieurs éléments indiquent que la perception de rugosité est l'un des aspects élémentaires de notre perception des sons. L'étude de ce premier aspect de la rugosité s'inscrit dans la perspective plus générale des efforts faits pour mieux comprendre les mécanismes fondamentaux de la perception auditive.

### 1.1.2 Attributs perceptifs de sons complexes

Notre perception des sons complexes est en général différente de la somme des sensations provoquées par leurs parties considérées isolément. Par exemple, une note de violoncelle, physiquement composée de plusieurs partiels superposés, n'est pas perçue comme une collection de hauteurs mais comme *un* son, avec une hauteur bien définie. Plusieurs types de mécanismes interviennent pour nous permettre d'organiser les évènements sonores en représentations psychologiques cohérentes (Bregman, 1990). Ce sont ces représentations, aussi appelées "images auditives", qui une fois formées possèdent des "attributs émergeants" comme l'intensité ou la hauteur (McAdams, 1984).

La rugosité est l'un de ces attributs. Ainsi, la transition entre l'étude de la sensation élémentaire et la compréhension de la rugosité de sons complexes se doit de considérer l'influence de mécanismes susceptibles de combiner un grand nombre d'informations, pour finalement aboutir à une image auditive possédant certaines caractéristiques perceptives.

### 1.1.3 Ecoute musicale

La description du percept de rugosité a été établie à l'origine dans le cadre d'une théorie de la consonance et de la dissonance musicale. Ces notions, au cœur des pratiques des musiciens, ont suscité un nombre étonnant de tentatives d'explications scientifiques depuis les premiers philosophes grecs jusqu'à nos jours (Tenney, 1988). Le percept élémentaire de rugosité peut être envisagé au sein de l'activité complexe qu'est l'écoute musicale comme un support à des traitements cognitifs de plus haut niveau, pouvant éventuellement aboutir à la perception de dissonance. Sa pertinence doit dans ce cas être établie avec des stimuli possédant une certaine richesse acoustique, susceptibles de provoquer de tels traitements. L'étude de la rugosité peut alors contribuer au long dialogue entre scientifiques et musiciens concernant la consonance, et peut même présenter un intérêt renouvelé du fait des approches originales adoptées par certains compositeurs contemporains.

## 1.2 Plan de l'étude

Nous avons souhaité aborder ces différents aspects de la perception de rugosité au cours de cette thèse, de manière à conserver une vue d'ensemble du phénomène.

La partie théorique sera donc consacrée à établir les problématiques retenues. Le Chapitre 2 présentera les travaux existant portant sur la rugosité en tant que percept élémentaire. Le Chapitre 3 présentera d'autres données sur des sons qui paraîtront sans doute au lecteur encore bien élémentaires, mais dont le but est pourtant d'aboutir à l'estimation de rugosité des sons complexes. Le Chapitre 4 proposera une tentative de définition de la consonance musicale et examinera les théories perceptives la concernant, dont celle basée sur la rugosité.

La partie expérimentale reprendra ces problématiques. Les Chapitres 5 et 6 seront consacrés à l'exposition de nouvelles données modifiant notre compréhension des modalités élémentaires de perception de rugosité. Le Chapitre 7 établira certains mécanismes supplémentaires à prendre en compte pour estimer la rugosité de sons quelconques. Enfin, le Chapitre 8 proposera une extension des théories de la consonance basées sur la rugosité à un contexte musical non tonal.

# Partie Théorique

# Chapitre 2

# Perception de rugosité

*Les études ayant contribué à la compréhension de la perception de rugosité sont examinées.*

Les études concernant la perception de rugosité sont nombreuses et leurs approches parfois contradictoires, ce qui rend toute tentative de définition simple soit partielle, soit incorrecte. La première description du phénomène apparaît dans un ouvrage regroupant une théorie physiologique de l'audition et une théorie musicale (von Helmholtz, 1877). Si la notion de rugosité a été introduite à partir des introspections d'un auteur particulier, un ensemble de résultats indiquent désormais qu'elle est un attribut élémentaire de la perception du timbre des sons complexes. Il est alors possible de s'interroger sur les mécanismes auditifs qui en sont la cause. C'est en ce sens que nous organiserons notre synthèse de la littérature sur le sujet. Les résultats expérimentaux disponibles peuvent être classés en deux grandes approches, l'une basée sur la manipulation de la composition spectrale des sons et l'autre sur celle de leurs paramètres temporels. Malgré leurs nombreux points communs, ces approches trahissent des divergences dans l'interprétation à donner au phénomène.

# 2.1 Description originale de la rugosité

## 2.1.1 Rugosité de la superposition de deux sons purs

### 2.1.1.1 Problème acoustique

Une question apparemment simple est à l'origine de la première description systématique du percept de rugosité : que se passe-t-il lorsque deux sons purs sont produits simultanément ? Cette question est posée au sein de l'ouvrage de Helmholtz (1877) sur la perception auditive. En tant qu'acousticien, il propose une réponse indiscutable : si les deux sons sont de fréquences différentes, ils coexistent dans le milieu de propagation, l'air, et l'onde de pression qui en résulte possède des fluctuations d'amplitude appelées battements. Néanmoins, en plus d'être un acousticien, Helmholtz est aussi un physiologiste intéressé par la perception. En écoutant la superposition de deux sons purs dont il fait varier l'écart fréquentiel grâce à un dispositif ingénieux (la double sirène, Figure 2.1), il distingue trois zones de qualités perceptives très différentes (Figure 2.2).

### 2.1.1.2 Qualités perceptives

Si les deux sons sont de fréquences légèrement différentes, leur superposition produit la perception d'un son pur à la fréquence moyenne des sons originaux dont l'amplitude fluctue lentement. Ce phénomène est connu sous le nom de battements, bien connu des accordeurs de piano. Les battements peuvent être entendus clairement jusqu'à 4 à 6 Hz.

Si l'on augmente l'écart fréquentiel, les battements s'accélèrent. Au delà d'une dizaine de Hz, leur perception s'estompe : il devient par exemple impossible de les compter. Néanmoins, Helmholtz affirme que nous continuons à avoir *conscience* de la présence de ces battements rapides. Cette impression donne au son un aspect "intermittent, dur, discordant, rugueux". Ce dernier adjectif, *rauhig* en Allemand, sera finalement retenu par Helmholtz et les auteurs ultérieurs pour décrire la sensation éprouvée[1]. Selon Helmholtz, des battements de l'ordre de 33 Hz produisent une rugosité maximale.

1. Une traduction plus appropriée en Français serait sans doute "rauque", donnant le substantif "raucité" (Demany, communication personnelle). Malheureusement, le mot de rugosité semble être rentré dans l'usage.

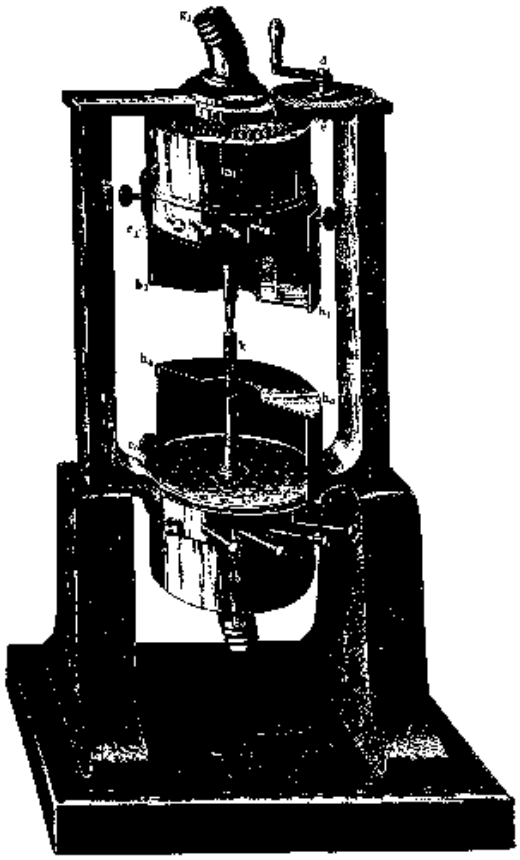

FIG. 2.1 – La double sirène employée par Helmholtz (1877). De l'air sous pression arrive par les tuyaux au sommet et à la base de la sirène. A chaque arrivée d'air, un disque immobile percé et un disque rotatif percé lui aussi sont superposés. Les disques rotatifs sont montés sur un même axe. L'alternance des coïncidences entre les trous des disques immobiles et mobiles provoquent une variation périodique de pression dont la fréquence dépend de la vitesse de rotation de l'axe. La manivelle visible en haut et à droite permet de faire tourner le disque "fixe" supérieur à faible vitesse et ainsi désaccorder les deux sirènes. Les résonateurs montés sur les disques sont choisis pour favoriser le premier partiel. Cette sirène semble avoir fortement marqué le compositeur Edgar Varèse en tant qu'instrument de synthèse sonore expérimentale (de la Motte-Haber, 1994).

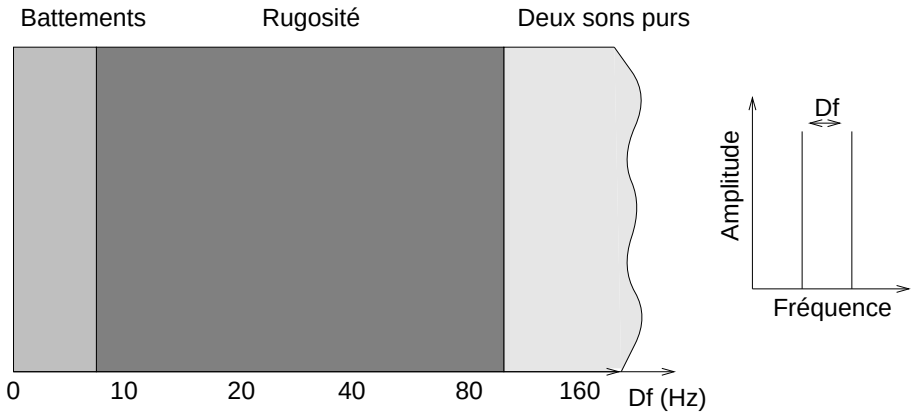


FIG. 2.2 – Les qualités perceptives d'une paire de son purs en fonction de leur écart fréquentiel. Les frontières entre les régions ne sont pas abruptes, notamment la frontière supérieure entre rugosité et deux sons distincts dépend du registre. D'après Helmholtz (1877)

Si l'écart fréquentiel entre les sons est encore augmenté, la conscience des battements et la sensation de rugosité diminuent jusqu'à disparaître. La perception de deux sons purs à des fréquences différentes, aux amplitudes constantes, apparaît alors. La limite supérieure d'audibilité de rugosité dépend fortement du registre considéré, plus les sons étant aigus et plus la fréquence des battements à la limite perceptibles augmente. Helmholtz affirme entendre, dans les conditions les plus favorables, jusqu'à 132 battements par seconde.

#### 2.1.1.3 Interprétation proposée par Helmholtz

Nous ne pouvons pas, bien entendu, proposer dès maintenant une interprétation complète et satisfaisante des causes du phénomène de rugosité. Si tel était le cas, une bonne partie du travail qui suit serait inutile ! Néanmoins, l'interprétation qu'en propose Helmholtz permet de soulever la plupart des problématiques qui se dégageront des études ultérieures.

L'existence des trois zones de qualités perceptives différentes est interprétée en s'appuyant sur une théorie de la physiologie du système auditif. Cette théorie nouvelle (à l'époque) considère que la cochlée, grâce à la mise en vibration de la membrane basilaire, est le siège de la sensation auditive. En partant d'observations physiologiques, l'hypothèse est faite que l'oreille se comporte comme une série de résonateurs signalant la présence des fréquences composant le signal acoustique.

La perception des battements lents lors de la superposition de deux sons purs à un faible écart fréquentiel peut alors être expliquée si les sons excitent le même résonateur : les battements, visibles dans l'onde acoustique, sont alors effectivement codés physiologiquement. La différence de nature entre la sensation lorsque l'on passe des battements à la rugosité, quand l'écart fréquentiel augmente, est justifiée par des considérations d'ordre psychologique (une "confusion" de la perception des deux sons purs) et d'ordre sensoriel (une plus grande excitation des fibres nerveuses). La disparition de la sensation de rugosité fait enfin l'objet de deux autres conjectures. La première est que si l'écart fréquentiel est trop large, les deux sons ne peuvent exciter les mêmes résonateurs ; la seconde est que, par analogie avec la vision, une fatigue perceptive résulterait dans l'intégration des fluctuations en une sensation continue.

### 2.1.2 Questions ouvertes par la théorie de Helmholtz

La théorie physiologique de Helmholtz est conforme avec le codage tonotopique effectivement réalisé par le système auditif périphérique — à ceci près que la mécanique de la membrane basilaire est plus complexe que celle suggéré par l'image d'une série de résonateurs juxtaposés (Davis, 1983; Pickles, 1988), et que certains des aspects temporels du codage concernant notamment la perception de hauteur (Demany, 1989) sont négligés. La perception de battements lorsque cette décomposition est mise en défaut représente néanmoins une première exposition du concept de bande critique. L'existence d'une telle bande critique a été largement confirmée depuis par une masse de données tant physiologiques que psychophysiques (Fletcher, 1940; Zwicker, Flottorp et Stevens, 1957; Scharf, 1970). Le lien entre la rugosité et ces nouvelles données mérite donc d'être réexaminé. Nous y reviendrons en section 2.3.3.

Une ambiguïté sur la nature même du phénomène peut être remarquée dans la théorie. La rugosité est définie comme le résultat de battements entre sons purs. Quelle est donc la cause réelle du phénomène, la *proximité fréquentielle* entre sons purs ou les *battements* qui en résultent ? Plusieurs éléments présents dans l'ouvrage de Helmholtz permettent de penser que celui-ci était conscient de cette ambiguïté. En marge des observations concernant les dyades de sons purs, Helmholtz réalise deux expériences révélatrices. Tout d'abord, il bouche tous les trous d'un disque tournant de sirène sauf un, sur lequel il ajuste une anche double.

En mettant la sirène en marche, il obtient un son complexe modulé en amplitude à la vitesse de rotation du disque. La sensation de rugosité est présente pour certaines vitesses de rotations. Ensuite, il écoute le son d'un diapason derrière un disque de sirène en rotation : il module donc en amplitude un son pur. Là encore, le résultat peut être rugueux. Ces expériences ne permettent néanmoins pas de répondre à la question posée, car les aspects spectraux et temporels sont covariants entre modulation d'amplitude et battements entre sons purs. Cette ambiguïté va se retrouver dans les études expérimentales à venir, et c'est autour d'elle que nous organiserons leur présentation (sections 2.3, 2.4).

Le passage de la rugosité définie pour des sons purs et la rugosité de sons complexes n'est pas traité explicitement. Lors des calculs destinés à alimenter la discussion sur les intervalles musicaux, Helmholtz se ramène toujours à la rugosité entre deux sons purs en isolant des paires partiels au sein des sons complexes. Conscient des problèmes posés par une telle méthode, il traite la combinaison à effectuer entre ces calculs de façon elliptique. Nous exposerons ces problèmes dans le Chapitre 3.

Finalement, l'argumentation sur la différence qualitative entre rugosité et battements est conclue en affirmant que la sensation de rugosité est naturellement "désagréable". Ceci permet ensuite d'établir les bases d'une théorie de la consonance musicale. Les justifications données par Helmholtz du caractère désagréable sont peu convaincantes, un passage abrupt de considérations physiologiques à une appréciation qualitative étant délicat. Néanmoins, ce qualificatif de "désagréable" a été repris et utilisé avec succès par nombre d'auteurs ultérieurs pour recueillir des jugements expérimentaux. Nous aborderons les questions que soulèvent ce point dans la section suivante ainsi que dans le Chapitre 4, consacré à la notion de consonance musicale.

## 2.2 La rugosité comme dimension du timbre

La rugosité est censée caractériser un attribut de la perception des sons de façon indépendante de leur hauteur, durée ou intensité. En tant que telle, elle ne peut être qu'un élément du timbre. Reste maintenant à savoir si elle est réellement une *dimension* du timbre, ou si au contraire ce mot n'est qu'un adjectif de plus pour décrire tel ou tel son particulier. La théorie de Helmholtz est le résultat des

introspections, réflexions et déductions de l'auteur, sans doute influencées par sa théorie musicale. Il convient donc de s'interroger sur la validité de la rugosité en tant que percept élémentaire.

### 2.2.1 Description des dimensions du timbre

Un survol rapide de la littérature classique consacrée au timbre est dans un premier temps légèrement décourageant. En appliquant la technique des différentiateurs sémantiques[2] à un ensemble d'adjectifs censés décrire les attributs du timbre, von Bismarck montre que la dimension "rugueux-lisse" est peu ou pas appropriée pour décrire le corpus de sons synthétiques qu'il étudie (von Bismarck, 1974). Ce résultat indique que le mot qui a été choisi plus ou moins arbitrairement par Helmholtz n'évoque rien de spécial chez la majorité des auditeurs (ce qui n'est pas une surprise pour qui a tenté de convaincre ces mêmes auditeurs de passer une expérience sur le sujet !).

Les études qui utilisent des techniques d'analyse multidimensionelle sont mieux à même de dégager des dimensions perceptives pertinentes sans passer par le vocabulaire. En effet, c'est à partir du traitement de jugements de similarité que les dimensions sont extraites. Les études employant de telles méthodes appliquées au timbre des instruments musicaux ont systématiquement mis en évidence une dimension liée au temps d'attaque, une autre liée à la répartition des composantes fréquentielles, mais jamais de dimension liée de près ou de loin à la rugosité (Plomp, 1970; Grey, 1977; McAdams, Winsberg, Donnadieu, De Soete et Krimphoff, 1995). Ceci découle en fait de la sélection du corpus de stimuli : des sons instrumentaux ou synthétiques représentant une note isolée, jouée à une hauteur moyenne. Ils ont donc tous une rugosité nulle ou très faible, et une éventuelle dimension de rugosité ne peut ressortir des analyses[3].

Dans d'autres conditions de sélection d'échantillons sonores, la rugosité est néanmoins apparue comme une dimension du timbre. Björk (1985) a étudié par la technique des différentiateurs sémantiques un corpus de sons de l'environne-

2. Un ensemble de jugements sur des échelles bipolaires opposant deux adjectifs (Osgood, Suci et Tannenbaum, 1957).

3. La rugosité peut néanmoins selon nous être pertinente pour décrire le timbre des instruments de musique, notamment pour des notes graves, des sons multiphoniques ou réalisés dans des modes de jeu tel que le flatterzunge ou chanté/joué. Aucun élément expérimental ne conforte cependant cette proposition pour l'instant.

ment (allant du bruit de la mer à un ensemble de cris d'alerte de différentes espèces animales). Il trouve alors que le facteur principal des jugements, pour 84 sujets, est un attribut regroupant des adjectifs tels que tendu, pressant, déplaisant qui est fortement corrélé avec une échelle de rugosité estimée. La rugosité est aussi apparue corrélée au désagrément occasionné par certains sons d'automobiles (Hashimoto et Hatano, 1994). Les modèles utilisés pour évaluer la qualité des sons environnementaux ou industriels comportent d'ailleurs généralement une estimation de rugosité (Fastl, 1997).

### 2.2.2 Similarité perceptive

Une étude d'un intérêt particulier pour établir le statut de la rugosité en tant que dimension élémentaire est celle de Schellenberg et Trainor (1996). Les auteurs s'interrogent sur la similarité perçue entre intervalles musicaux obtenus par la superposition de deux sons complexes. Des séquences préliminaires d'intervalles successifs, aux fondamentales variables mais de même nature (des quintes, 3 tons et demi), sont présentées aux auditeurs. Un intervalle de test, soit identique à l'un des intervalles de la séquence, soit plus étroit de un demi-ton (formant maintenant un triton, 3 tons), soit plus étroit de un ton (formant une quarte, 2 tons et demi) est ensuite proposé. La tâche des auditeurs consiste à déterminer si l'intervalle de test était ou non présent dans la séquence préliminaire. L'hypothèse est que plus le test est perçu comme similaire à un intervalle de quinte, plus les confusions avec un élément de la séquence sont probables.

Les 28 sujets participant à l'expérience ont montré un taux de détection des tritons supérieur à celui des quartes. Ceci indique que bien que l'indice lié à l'écart fréquentiel permettant de différencier le test de la séquence soit deux fois plus grand pour la quarte que pour le triton, elle a plus souvent été jugée par erreur comme appartenant à la séquence. Il semble en fait que les jugements aient été faits sur la base de la rugosité des accords : la quinte et la quarte sont peu rugueuses alors que le triton est très rugueux (von Helmholtz, 1877; Plomp et Levelt, 1965).

Néanmoins, il est aussi possible de supposer une influence de l'apprentissage implicite des auditeurs, tous étant sans doute familiers avec l'harmonie tonale occidentale. Selon ce système, la quinte et la quarte sont plutôt consonantes alors que le triton est très dissonant. Les auditeurs auraient donc pu juger la similarité

en fonction d'un degré de consonance conventionnel. L'expérience a été répétée avec des bébés de 7 mois, chez qui la perception des hiérarchies tonales n'est pas encore présente (Trainor et Trehub, 1994). Avec une procédure légèrement modifiée (*operant head-turn conditioning)*, les résultats obtenus sont essentiellement les mêmes. Une similarité plus forte est perçue entre quinte et quarte qu'entre quinte et triton.

Un résultat similaire avait en fait déjà été obtenu dans d'autres circonstances. Les étourneaux (*Sturnus vulgaris)* peuvent apprendre à discriminer entre stimuli acoustiques et généraliser ce qu'ils ont appris si une dimension perceptive a pu être extraite (Hulse, Takeuchi et Braaten, 1992). En présentant des accords de trois notes à ces oiseaux, il a été montré que la seule dimension permettant d'expliquer les généralisations effectuées était celle de "consonance" des accords (Hulse, Bernard et Braaten, 1995). Invoquer une connaissance culturelle de la consonance semble ici encore plus improbable. La dimension de rugosité, corrélée à celle de consonance pour ces intervalles, a pu par contre être la base de la généralisation. En effet, le système auditif des oiseaux ressemble à celui des mammifères sous plusieurs aspects et notamment en ce qui concerne la perception de modulation d'amplitude (Langner, 1992). Enfin, mentionnons des expériences sur la perception de trains d'impulsions ou de paires de sons purs par les poissons rouges *(Carassius auratus)* qui ont aussi démontré la perception d'une dimension identifiée comme étant la rugosité (Fay, 1994, 1996).

Ces études indiquent clairement l'existence d'un attribut auditif de base correspondant à la rugosité. La perception de cet attribut n'est pas lié à un apprentissage, musical ou autre, vu sa perception par les nouveau-nés. Le fait qu'il soit aussi perçu par d'autres espèces animales est loin d'être anecdotique. Ceci prouve que la rugosité, à l'instar de la perception de hauteur qui a été mise en évidence avec les mêmes techniques chez ces différentes espèces, possède un intérêt "écologique" et est vraisemblablement liée à des mécanismes physiologiques élémentaires. Dès lors, la question de tenter de comprendre quels sont ces mécanismes sous-jacents prend tout son sens.

### 2.2.3 Corrélats verbaux de la rugosité

Les études classiques portant sur la rugosité ont employé différentes consignes verbales pour recueillir les jugements des auditeurs. En effet, demander à des su-

jets musiciens ou non de juger la rugosité d'un ensemble de sons peut apparaître malaisé, la plupart d'entre eux n'ayant jamais pensé spontanément à qualifier un quelconque son de rugueux (von Békésy, 1960; Faure, 1994). D'autres étiquettes verbales ont donc été employées. Le plus souvent ont été recueillis soit des jugements "d'agrément", du fait du caractère désagréable de la rugosité selon Helmholtz, soit des jugements de "consonance", la rugosité étant corrélée à la dissonance des accords musicaux.

Dans un contexte musical, ces deux mots n'ont pas la même connotation. La notion même de consonance semble délicate à définir (voir section 4.1). En demandant à des musiciens d'ordonner des intervalles tempérés, Guernesey (1928) a montré que les accords jugés les plus consonants n'étaient pas forcément les plus agréables. Toutefois, une autre étude a démontré une forte corrélation entre jugements de consonance et d'agrément pour des non-musiciens, avec notamment un fort agrément perçu pour l'octave et la quinte ce qui n'est pas le cas en général pour les musiciens (van de Geer, Levelt et Plomp, 1962). Il est clair qu'une influence culturelle liée à l'apprentissage ressort de ces études, les musiciens (ou mélomanes) trouvant agréables des intervalles dissonants et ennuyeux des intervalles très consonants.

L'évaluation d'intervalles arbitraires non basés sur des rapports mais sur des différences de fréquences, faits de sons purs, permet d'obtenir des jugements de consonance concordants entre musiciens et non-musiciens (Guthrie et Morril, 1928). De plus, les échelles "d'agrément" et de "consonance" sont dans ce cas confondues (Figure 2.4). Il semble que ces deux mots, ainsi que les autres qui ont pu être employés, reflètent une même dimension générale dite "d'évaluation" (van de Geer, Levelt et Plomp, 1962; Plomp et Levelt, 1965; Björk, 1985). Cette dimension perceptive est susceptible d'un consensus entre auditeurs, elle peut dans certaines conditions être mesurée par les mots de "consonance" ou "d'agrément", sans pour autant préjuger du degré de consonance ou d'agrément préféré par chaque auditeur.

La seule dimension sensorielle qui permet d'interpréter dans un cadre cohérent tous les résultats que nous allons maintenant présenter est celle de rugosité. Nous pensons donc que, en dépit des différents adjectifs utilisés pour transmettre la consigne aux auditeurs, une mesure de rugosité a en fait été recueillie et nous la

considérerons comme telle. L'accord entre résultats des différentes études malgré les différences de consignes est un autre élément qui justifiera cette hypothèse.

## 2.3 Approche spectrale

### 2.3.1 Rugosité d'une paire de sons purs

#### 2.3.1.1 Le stimulus de base

Un grand nombre d'études se sont penchées sur la rugosité produite par une dyade de sons purs d'amplitudes similaires. En notant $f_1$ et $f_2$ les fréquences des deux sons purs, ce stimulus est de la forme :

$$Dy(t) = \cos(2\pi f_1 t) + \cos(2\pi f_2 t) \tag{2.1}$$

De façon équivalente, en faisant apparaître $D_f = f_2 - f_1$ et $f_{moy} = (f_1 + f_2)/2$ :

$$\begin{aligned} Dy(t) &= 2\cos(2\pi \frac{D_f}{2} t)\cos(2\pi f_{moy} t) \\ &= \underbrace{2|\cos(2\pi D_f t)|}_{E(t)}\cos(2\pi f_{moy} t) \end{aligned} \tag{2.2}$$

L'enveloppe $E(t)$ de $Dy(t)$ est de période $1/D_f$, mais sa dérivée est discontinue quand elle passe par zéro. Ce stimulus est représenté en Figure 2.3. Deux paramètres principaux peuvent ici être variés indépendamment : l'écart fréquentiel $D_f$, qui impose la fréquence des battements, et la fréquence moyenne $f_{moy}$.

#### 2.3.1.2 Méthodes

Diverses méthodes ont été employées avec ce type de stimuli pour parvenir à une mesure que nous estimons, pour les raisons indiquées plus haut, être liée à la rugosité. Mayer (1874, 1875, 1894) a utilisé des diapasons spécialement conçus pour donner un son aussi pur que possible et a demandé à ses sujets (entre 1 et 12 selon les conditions, tous auditeurs expérimentés) de déterminer "le plus petit intervalle consonant". Cross et Goodwin (1893) ont employé des tuyaux excités par une source de pression et ont effectué une mesure similaire, en ajoutant une

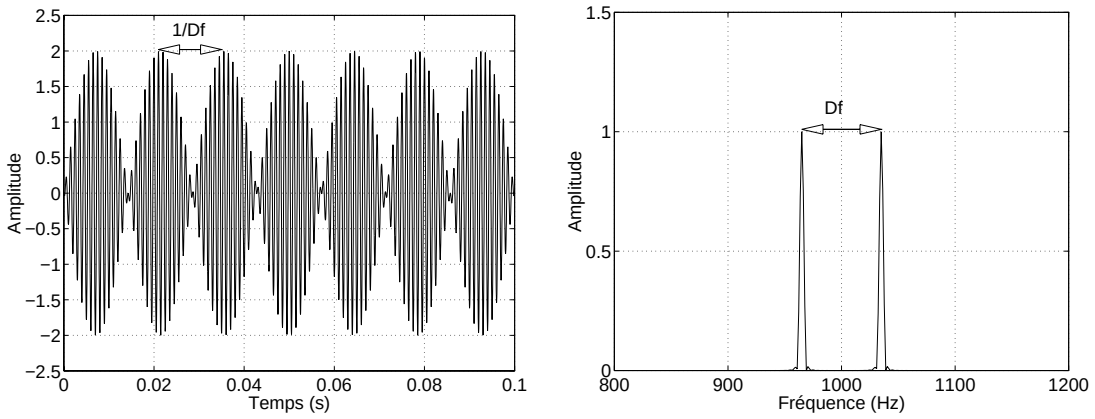


FIG. 2.3 – Forme d'onde temporelle et spectre d'amplitude d'une dyade de sons purs. L'écart fréquentiel $D_f$ impose la période des battements.

mesure de l'intervalle "le plus dur" *(harsh)* pour un seul sujet. Kaestner (1909) a employé des diapasons et recueilli des jugements de comparaison par paires : pour trente intervalles, tous avec une fondamentale de 320 Hz, chaque stimulus est comparé avec tous les autres et la proportion où chaque intervalle est jugé comme "plus plaisant" est comptabilisée, pour 5 sujets. Guthrie et Morrill (1928) ont utilisé des sifflets et ont fait juger 44 intervalles arbitraires par deux groupes de 381 et 372 sujets. Un groupe a donné des jugements de consonance, l'autre d'agrément, la proportion de jugements de consonance ou d'agrément est la mesure considérée. Plomp et Levelt (1965) ont utilisé des oscillateurs diffusés sur haut-parleurs et ont recueilli un jugement de consonance sur une échelle numérique pour 5 groupes indépendants d'une dizaine de sujets chacun. Plomp et Steeneken (1968) ont employé une méthode d'ajustement de fréquence pour obtenir des estimations de l'intervalle produisant une rugosité maximale et du premier intervalle où les sons n'interfèrent plus perceptivement, pour 20 sujets, en utilisant des oscillateurs présentés au casque dont la sonie était contrôlée. Kameoka et Kuryiagawa (1969a) ont utilisé des intervalles produits par des oscillateurs sur haut-parleurs et une méthode de comparaisons de consonance par paires incomplète (chaque jugement de consonance est une comparaison entre stimuli sur une échelle numérique, toutes les paires ne sont pas testées) pour plusieurs groupes d'une vingtaine de sujets chacun en moyenne.

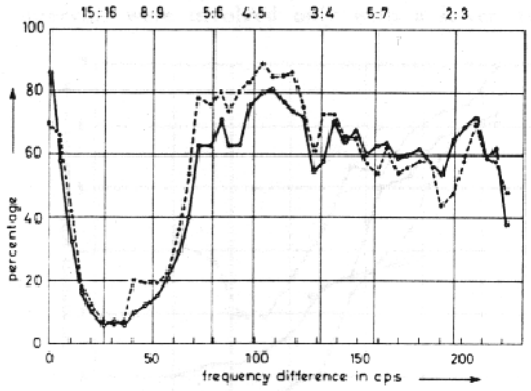


FIG. 2.4 – Échelles de consonance (trait plein, 381 sujets) et d'agrément (*pleasantness*, traits pointillés, 372 sujets) pour 44 intervalles entre deux sons purs. Les deux mesures correspondent à l'inverse de la rugosité, qui augmente à partir de l'unisson, atteint un maximum puis décroît pour se maintenir à un niveau quasi-nul. D'après Guthrie et Morrill (1928).

## 2.3.2 Influence des paramètres acoustiques spectraux

### 2.3.2.1 Influence de l'écart fréquentiel et de la fréquence moyenne

Toutes les études ayant recueilli une estimation de rugosité en fonction de l'écart fréquentiel $D_f$ ont confirmé la description qualitative de Helmholtz. En partant d'une rugosité nulle pour l'unisson, un maximum de rugosité est observé puis la rugosité diminue progressivement jusqu'à une disparaître (voir par exemple Figure 2.4).

L'intervalle pour lequel la disparition de rugosité est observée dépend du registre considéré, soit avec nos notations de $f_{moy}$. L'intervalle provoquant une rugosité maximale dépend lui aussi de $f_{moy}$. De manière à pouvoir comparer les résultats des différentes études, nous avons choisi de nous intéresser à ces deux mesures, quand celles-ci étaient disponibles : le premier intervalle pour lequel la rugosité disparaît, et l'intervalle provoquant une rugosité maximale. Nous les représentons en Figure 2.5(a) et (b).

L'intervalle de disparition de rugosité en fonction du registre est représenté pour six études différentes. Une simple inspection visuelle indique un très fort accord entre les différents auteurs. Globalement, l'intervalle provoquant la disparition de rugosité augmente avec le registre. Le seul point aberrant est celui fourni par Plomp et Levelt (1965) pour 125 Hz. Ceci est probablement dû au

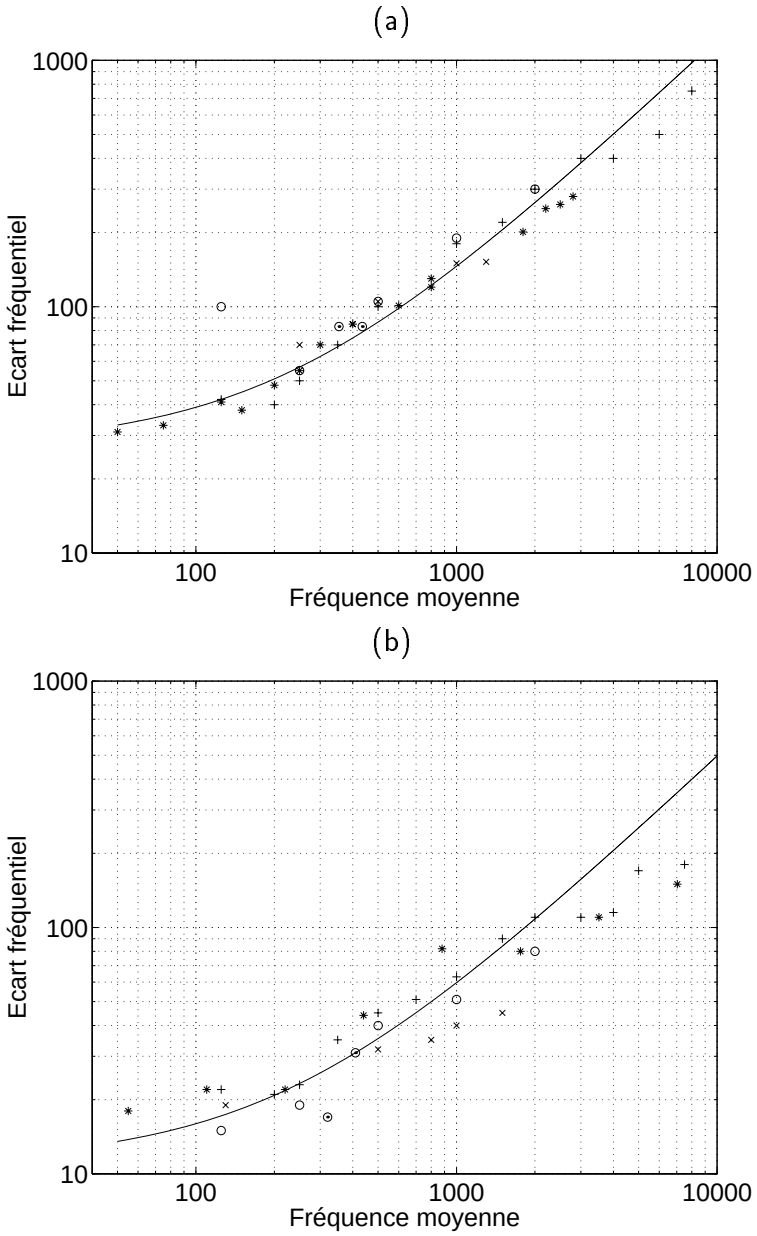


FIG. 2.5 – *(a)* : le plus petit intervalle sans rugosité en fonction de la fréquence moyenne. ∘ : Plomp et Levelt (1965) ; + : Plomp et Steeneken (1968) ; ∗ : Mayer (1894) ; × : Cross et Goodwin (1893) ; ⊙ : Guthrie et Morril (1928) et Kaestner (1909).
*(b)* : intervalle de rugosité maximale. ∘ : Plomp et Levelt (1965) ; + : Plomp et Steeneken (1968) ; ∗ : Kameoka et Kuryiagawa (1969a) ; × : Cross et Goodwin (1893) ; ⊙ : Guthrie et Morril (1928) et Kaestner (1909).
Les traits pleins représentent 110 % (a) et 45 % (b) de la bande critique mesurée en ERB (voir texte).

fait que dans cette étude le niveau en dB SPL des composantes était maintenu constant : la sonie de la composante la plus grave pour une fréquence moyenne de 125 Hz approche alors le seuil d'audition. Dans une étude ultérieure, le contrôle de la sonie en lieu et place du niveau physique fournit une valeur cohérente avec les autres auteurs (Plomp et Steeneken, 1968).

L'intervalle qui provoque une rugosité maximale est représenté pour six études différentes. Là encore, l'accord entre les études est satisfaisant, quoique la dispersion des valeurs soit plus importante.

### 2.3.3 Interprétations

#### 2.3.3.1 Ajustement des données aux mesures de bande critique

L'hypothèse d'un lien entre rugosité et bande critique, suggérée par Helmholtz, a été reprise par Greenwod (1961) et Plomp et Levelt (1965). Nous avons utilisé une mesure courante de bande critique, obtenue à partir d'expériences de masquage (Patterson, 1976), pour illustrer ce lien. Une fraction de la largeur de la bande critique aux différentes fréquences moyennes, mesurée en ERB (Glasberg et Moore, 1990), a été superposée aux données de la Figure 2.5. La fraction a été choisie par ajustement visuel.

Pour ce qui est de l'intervalle de rugosité maximale, une fraction de 110 % de la bande critique en ERB produit un ajustement très satisfaisant. L'ajustement de la rugosité maximale à 45 % de la bande critique en ERB est légèrement moins bon, du fait de la dispersion plus marquée des données. Cette fraction est de plus en désaccord avec celle de 25 % proposée par Plomp et Levelt (1965). Remarquons néanmoins que ceux-ci ont utilisée une mesure différente de bande critique (Zwicker, Flottorp et Stevens, 1957) et qu'ils indiquent que cette fraction est une règle qualitative. L'idée principale qui ressort de ces ajustements est que la rugosité de deux sons purs dépend du rapport entre leur écart fréquentiel et la largeur de la bande critique (Figure 2.6).

Un désaccord peut toutefois être noté entre les valeurs observées et la bande critique en ERB pour les fréquences centrales élevées, notamment pour l'intervalle de rugosité maximum (Plomp et Steeneken, 1968; Greenwood, 1991a).

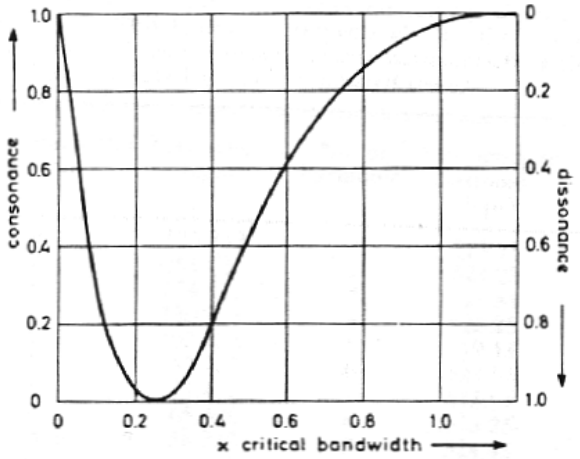


FIG. 2.6 – Rugosité d'une dyade de sons purs en fonction de l'écart fréquentiel normalisé par la bande critique. L'ordonnée représente l'inverse de la rugosité. L'influence de la fréquence moyenne se traduit par une modification de la largeur de la bande critique et donc de l'échelle en abscisse. D'après Plomp et Levelt (1965).

#### 2.3.3.2 La rugosité comme mesure de la bande critique

Le lien entre rugosité et bande critique est indiqué par la concordance des données obtenues indépendamment pour les deux phénomènes. Il est de plus en parfait accord avec la théorie du filtrage auditif pour ce qui est du premier intervalle sans rugosité. Une interaction entre les composantes fréquentielles est en effet nécessaire pour produire une rugosité : pour des faibles séparations fréquentielles mais une écoute dichotique, aucune rugosité n'est perçue (Licklider, Webster et Hedlun, 1950). Un phénomène supplémentaire doit éventuellement être pris en compte pour les très larges séparations fréquentielles, du fait de la présence de sons de combinaison pouvant interagir avec une composante physique (von Helmholtz, 1877; Plomp, 1976). Ceci ne change pas le fait que le premier intervalle sans rugosité doit forcément être atteint quand les composantes fréquentielles commencent à être résolues. Nous discuterons la possibilité que cet intervalle soit atteint avant (notamment pour les fréquences moyennes élevées, comme remarqué en Figure 2.5) en examinant les résultats de l'approche temporelle (section 2.4).

L'accord entre les différentes études de la Figure 2.5 est tout à fait remarquable si l'on considère les disparités dans leur obtention : pratiquement une

dizaine de laboratoires indépendants, employant des dispositifs divers, avec des consignes différentes, des plan expérimentaux couvrant un éventail de méthodes psychophysiques allant de l'estimation directe de grandeur aux comparaisons par paires selon plusieurs variantes, pour aboutir sur plus d'un siècle à une échelle qui sera confirmée par la physiologie et un ensemble d'autres mesures psychoacoustiques. Les jugements de consonance ont été les premières mesures fiables de bande critique sur une large étendue fréquentielle. Ces jugements, que l'on aurait pu croire indissociablement liés à l'arbitraire musical, ont été l'une des méthodes les plus robustes et les plus efficaces pour accéder à la grandeur fondamentale pour l'audition qu'est la bande critique (Greenwood, 1961, 1991a,b).

### 2.3.4 Hypothèse Spectrale

L'approche spectrale débouche ainsi sur une conception du phénomène basée sur les paramètres du son tels que fréquences et amplitudes des partiels. Des formules analytiques de calcul de rugosité de deux sons purs sont proposées à partir de ces paramètres (Plomp et Levelt, 1965; Kameoka et Kuriyagawa, 1969a). Ces formules seront intégrées dans des modèles plus généraux visant à évaluer la rugosité de sons quelconques, qui se basent eux aussi sur la représentation spectrale des sons en considérant chaque partiel comme un son pur (Chapitre 3). Les principes de base qui découlent de cette approche peuvent donc être résumés comme suit :

- la rugosité peut être prédite à partir de la composition spectrale des sons ;
- pour deux sons purs, la rugosité disparaît pour un écart fréquentiel supérieur à une largeur de la bande critique à la fréquence moyenne considérée ;
- pour deux sons purs, la rugosité maximale est atteinte pour un écart approximativement égal à une fraction constante de la largeur de la bande critique à la fréquence moyenne considérée.

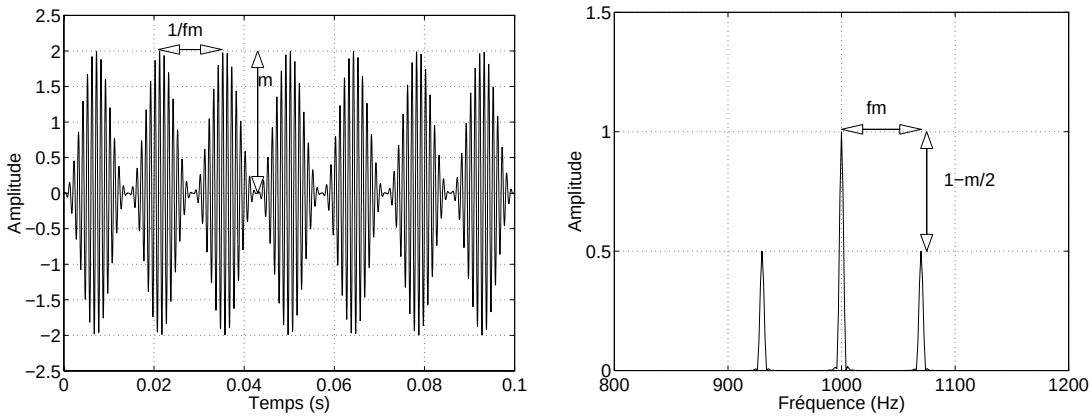


FIG. 2.7 – Forme d'onde temporelle et spectre d'amplitude d'un son pur sinusoïdalement modulé en amplitude. La fréquence de modulation $f_m$ impose l'écart entre les composantes spectrales.

# 2.4 Approche temporelle

## 2.4.1 Rugosité de sons purs modulés en amplitude

### 2.4.1.1 Le stimulus de base

Le stimulus caractéristique de l'approche temporelle est le son pur modulé sinusoïdalement en amplitude (SAM). Il peut être décrit par :

$$SAM(t) = \underbrace{(1 + m * \cos(2\pi f_m t)}_{E(t)} \cos(2\pi f_c t) \tag{2.3}$$

Ce type de stimulus permet de faire varier $f_c$, la fréquence centrale (ou fréquence porteuse), $f_m$ la fréquence de modulation et $m$, la profondeur de modulation. L'enveloppe $E(t)$ est ici une sinusoïde d'amplitude $m$ et oscillant autour d'une valeur constante égale à 1 (Figure 2.7).

### 2.4.1.2 Méthodes

Mayer (1874, 1875, 1894) a étudié des diapasons écoutés à travers un disque de sirène tournant. La vitesse de rotation pour lequel le sentiment d'interruption disparaît et celle qui produit la dissonance maximale sont mesurées pour 3 sujets. Terhardt a étudié des SAM produits par oscillateurs selon une variété de méthodes (Terhardt, 1968b,a, 1974a). La première consiste à ajuster la profon-

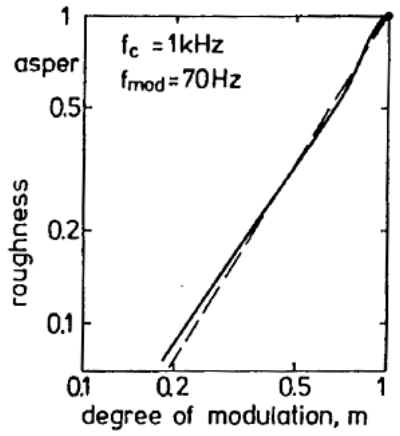


FIG. 2.8 – Influence de la profondeur de modulation sur la rugosité d'un son SAM. La rugosité augmente avec la profondeur de modulation suivant une loi, correspondant ici à $r = m^{1,6}$. D'après Zwicker et Fastl (1990).

deur de modulation $m_v$[4] d'un son de comparaison jusqu'à percevoir une rugosité équivalente entre le son testé (invariant) et le son de comparaison (avec $m_v$ ajustable). En effet, nous allons voir qu'un lien monotone a été trouvé entre rugosité et profondeur de modulation (section 2.4.2.1). Une autre possibilité est la méthode de fractionnement ou production de rapport pour laquelle les sujets ajustent la profondeur de modulation du son test ($m_t$ variable) pour qu'elle soit égale à la moitié ou au double de la rugosité d'un son de comparaison (invariant). La plupart des auteurs suivants utiliseront ces même méthodes (Vogel, 1974; Aures, 1984; Zwicker et Fastl, 1990).

## 2.4.2 Influence des paramètres acoustiques temporels

### 2.4.2.1 Influence de la profondeur de modulation

La rugosité dépend directement de la profondeur de modulation. Une profondeur de modulation nulle ne produit ni battements ni rugosité. Pour les fréquences de modulation qui peuvent produire une rugosité non nulle, la rugosité augmente avec la profondeur de modulation. La rugosité maximale est atteinte pour $m = 1,2$ puis des profondeurs de modulation supérieures provoquent une diminution de rugosité (Mathes et Miller, 1947; Aures, 1984). Néanmoins, pour

4. Le $v$ signifie *Vergleichung*, comparaison en Allemand. Nous conservons cette notation pour faciliter la lecture des Figures.

les profondeurs de modulations supérieures à 1 la forme de l'enveloppe change radicalement et la signification à donner à une telle mesure est ambiguë.

Le résultat important est que la rugosité augmente de façon monotone lorsque la profondeur de modulation passe de 0 à 1 (Terhardt, 1968a). La dépendance quantitative de la rugosité sur la profondeur de modulation est alors du type :

$$r \sim m^p \tag{2.4}$$

Cette équation a été proposée par différents auteurs (Terhardt, 1968a; Vogel, 1975; Guirao et Garavila, 1976; Zwicker et Fastl, 1990). L'exposant exact semble lui dépendre des méthodes de mesure, tout en étant compris entre 1,4 et 2. Zwicker et Fastl (1990) proposent une valeur de 1,6 (Figure 2.8).

#### 2.4.2.2 Influence de la fréquence de modulation et de la fréquence porteuse

Pour $m$ fixé et non nul, la rugosité est influencée par la fréquence de modulation. Une fréquence de modulation nulle produit un son pur sans rugosité, puis si la fréquence de modulation augmente la rugosité augmente, atteint un maximum puis décroît et disparaît (Mayer, 1894; Terhardt, 1968a; Vogel, 1974; Aures, 1984).

Intéressons-nous d'abord aux valeurs des fréquences centrales et de modulation $f_c$ et $f_m$ provoquant une rugosité maximale. Mayer (1874, 1875, 1894) obtient des mesures sur une région fréquentielle couvrant les fréquences porteuses de 64 Hz à 1024 Hz et démontre un accroissement régulier de la fréquence de modulation correspondant au maximum de rugosité. Terhardt (1968a, 1970) poursuit ces mesures au dessus de 1 kHz et montre que si la fréquence $f_m$ produisant le maximum de rugosité croît effectivement avec $f_c$ au dessous de 2 kHz, elle reste constante et égale à 80 Hz pour des fréquences porteuses plus élevées. Des résultats plus complets sont proposés par Aures (1984), et reproduits en Figure 2.9. La caractéristique passe-bande de la perception de rugosité en fonction de $f_m$ pour $f_c$ fixée apparaît ici clairement. Seule la mesure à 125 Hz présente un maximum secondaire, mais remarquons qu'à cette fréquence de modulation la composante grave du son SAM (située à $f_c - f_m$) n'est plus audible, le stimulus testé change alors de nature.

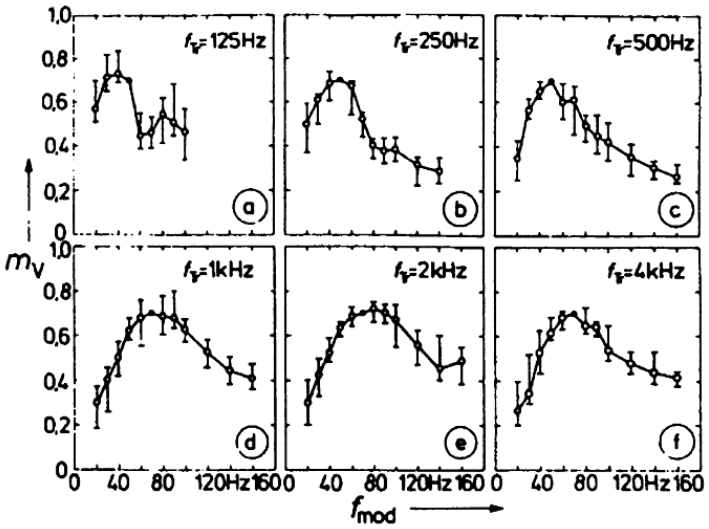


FIG. 2.9 – Influence de la fréquence de modulation $f_m$ sur la rugosité de sons SAM à différentes fréquences centrales. Pour chaque $f_c$, $f_m$ est en abscisse et la profondeur de modulation $m_v$ ajustée d'un son de comparaison modulé à une fréquence de référence (le seul point sans écart-type sur les courbes) est en ordonnée. La rugosité perçue correspond approximativement au carré de cette profondeur ajustée. D'après Aures (1985).

La fréquence porteuse influence aussi l'intensité du maximum de rugosité. La rugosité maximale atteinte augmente avec $f_c$ jusqu'à 1 kHz, puis diminue (Terhardt, 1968a; Vogel, 1974; Aures, 1984). Cette influence de $f_m$ et $f_c$ sur la rugosité a été résumée par Zwicker et Fastl (1990) (Figure 2.10). Il est malheureusement impossible de savoir si ce résumé comprend de nouvelles données apportées par les auteurs ou s'il constitue simplement une synthèse des données de la littérature. Néanmoins, cette "normalisation" de la rugosité de sons modulés en amplitude est considérée comme la référence pour étalonner les modèles computationnels de rugosité (Daniel et Weber, 1997). Le maximum de rugosité qu'il est possible d'obtenir pour des sons SAM est déterminé comme correspondant à $f_c = 1$ kHz et $f_m = 70$ Hz. Cette rugosité est choisie comme l'unité d'une échelle de mesure baptisée *asper*.

#### 2.4.2.3 Influence du niveau

L'influence du niveau de présentation sur la perception de rugosité de sons modulés en amplitude semble être faible. Terhardt (1968b) mesure un doublement

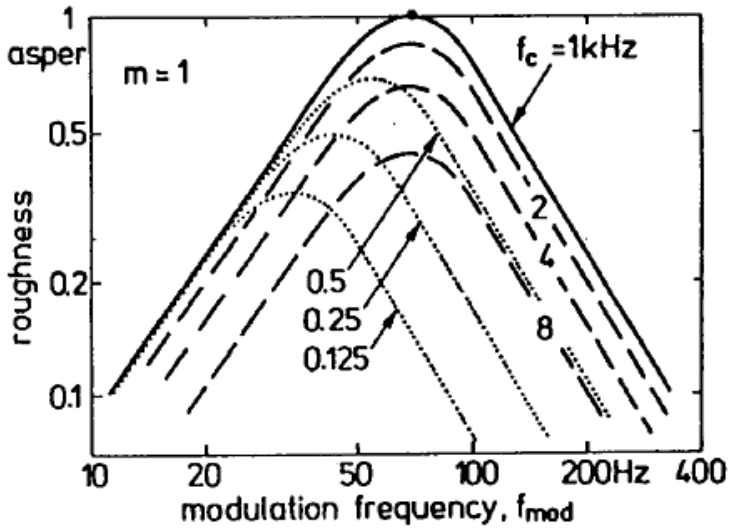


FIG. 2.10 – Rugosité d'un son SAM en fonction de la fréquence de modulation $f_m$ avec comme paramètre la fréquence centrale $f_c$ et. D'après Zwicker et Fastl (1990).

de rugosité pour une différence de 20 dB et $f_c = 1$ kHz, $f_m = 70$ Hz. Le contrôle de l'influence de l'ordre de présentation et de l'écoute binaurale ou monaurale montre que l'effet du niveau est généralement plus faible, le doublement pour 20 dB étant en fait le cas où le niveau a le plus d'influence (Terhardt, 1974a).

## 2.4.3 Interprétations

### 2.4.3.1 Degré de modulation effective

L'approche temporelle considère que la perception de fluctuations d'amplitude est la cause de la rugosité des sons SAM. Divers mécanismes interviennent pour extraire un "degré de modulation effective", directement lié à la rugosité, de l'enveloppe physique (Terhardt, 1974a; Aures, 1985b).

Le filtrage auditif impose une première limitation à la perception de rugosité : si les trois composantes des SAM sont résolues, la modulation d'amplitude ne peut être transmise et le degré de modulation effective est faible. L'influence de la bande critique est donc considérée de façon similaire que pour l'approche spectrale. L'approche temporelle rajoute toutefois un second facteur limitant : la fréquence des battements. Selon Terhardt (1974a), le système auditif se com-

porterait comme un filtre passe-bas vis à vis de la transmission de l'enveloppe. Ainsi, des battements trop rapides seraient intégrés et ne provoqueraient pas non plus de degré de modulation effective important. Une fois l'enveloppe "effective" extraite, il suffit d'estimer son index de modulation pour obtenir la rugosité. La faible influence du niveau global incite Terhardt à proposer le calcul d'une profondeur de modulation relative, obtenue en normalisant l'amplitude de l'enveloppe par rapport à celle du signal.

En résumé, pour une profondeur de modulation donnée, la rugosité tend à augmenter quand la fréquence de modulation augmente. Toutefois, quand la fréquence de modulation augmente la "profondeur de modulation effective" tend à diminuer du fait de deux facteurs : la bande critique et une caractéristique passe-bas de la modulation d'amplitude. Le premier facteur est limitant aux fréquences centrales graves, où la bande critique est étroite, alors que le deuxième facteur intervient aux fréquences centrales élevées. Ces deux mécanismes distincts permettent d'expliquer l'influence complexe de la fréquence de modulation en fonction de la fréquence centrale sur la rugosité des SAM (Figure 2.10). La fréquence $f_m$ provoquant le maximum de rugosité augmente d'abord avec $f_c$, car la bande critique s'élargit, puis reste constante et égale à 70 Hz quand la bande critique n'est plus limitante.

L'introduction du facteur limitant pour $f_m$ trop élevée permet d'expliquer le désaccord remarqué précédemment entre la position des maxima de rugosité par rapport à la bande critique pour les dyades de sons purs de fréquences moyennes élevées (Figure 2.5). Notons que certains auteurs, cités dans l'approche spectrale, envisageaient la possibilité de ce deuxième facteur sans que leurs théories ne puissent le prendre en compte (Plomp et Steeneken, 1968). Helmholtz (1877) avait lui trouvé un moyen détourné de l'inclure dans ses calculs (section 3.2.1). Une caractéristique passe-bas, équivalente à ce facteur, a depuis été observée pour les seuils de détection de modulation d'amplitude (Viemeister, 1977, 1979).

#### 2.4.3.2 Comparaisons entre résultats expérimentaux

Nous pouvons résumer les résultats obtenus pour les SAM en représentant la fréquence de modulation pour laquelle la rugosité disparaît et celle où la rugosité maximale est observée (Figure 2.11). Pour faciliter les comparaisons avec

l'approche spectrale, les fractions de bande critique en ERB représentées en Figure 2.5 sont reproduites.

Le nombre d'études est nettement inférieur à celui disponible pour l'approche spectrale. Les données de Mayer (1894) pour la disparition de rugosité rejoignent une valeur égale à 110 % d'une bande critique en ERB. Celles de Terhardt (1970) sont plutôt concordantes avec la bande critique définie par Zwicker (1961). Les raisons d'un tel désaccord ne sont pas entièrement claires. Nous interprétons ceci en soulignant que pour les sons SAM, trois composantes fréquentielles interagissent. Même si seulement deux d'entre elles sont espacées de moins d'une bande critique, une rugosité peut encore être produite. La situation est plus complexe que pour une dyade de sons purs et dans ce cas, l'addition des contributions des différentes paires de partiels pourrait peut-être expliquer l'accord médiocre entre la disparition de rugosité et la bande critique. Pour les fréquences élevées, l'influence du deuxième facteur devient prépondérante et limite l'intervalle à une valeur inférieure à une bande critique.

L'intervalle de rugosité maximale ne correspond pas non plus à une fraction constante de la bande critique, et ceci même pour les fréquences graves où elle est censée être le facteur limitant. Le maximum de rugosité est en fait assez large (Figure 2.9). La plupart des points dans un intervalle d'une trentaine de Hz autour du point choisi comme étant le maximum ne présentent pas de différences significative du point de vue statistique du fait des écarts types observés. Vogel (1974), en employant une méthode de comparaison de stimuli ayant des fréquences centrales différentes (et donc plus délicate à réaliser) obtient pour 5 auditeurs une rugosité à peu près équivalente pour toutes les fréquences centrales de 1 à 4 kHz et les fréquences de modulation 30 à 100 Hz. Les positions précises et "normalisées" par Zwicker et Fastl (1990) des maxima de rugosité représentent sans doute une simplification commode, mais doivent être considérées comme le centre d'une large région de forte rugosité pour les sons SAM.

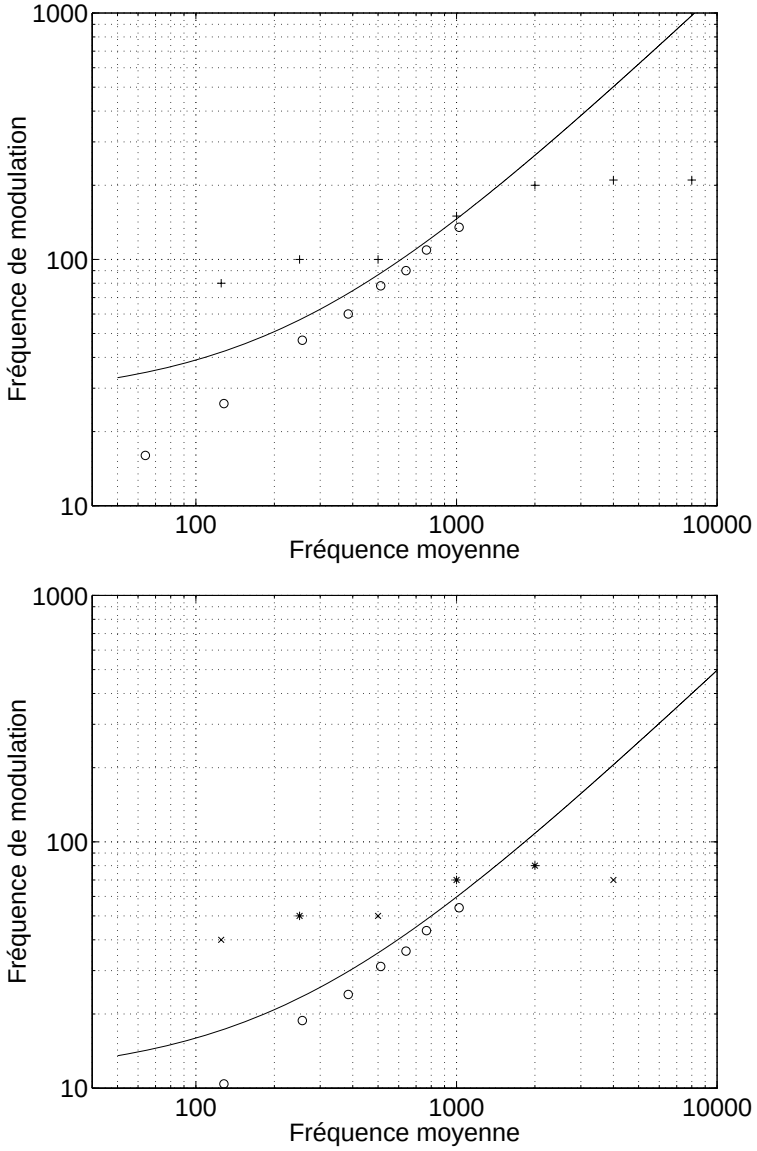


FIG. 2.11 – En haut : la plus petite fréquence de modulation sans rugosité en fonction de la fréquence centrale. ∘ : Mayer (1894) ; + : Terhardt (1972). Le trait plein correspond à 110 % de la bande critique en ERB. En bas : fréquence de modulation provoquant une rugosité maximale. ∘ : Mayer (1894) ; + : Terhardt (1972) ; × : Aures (1984). Le trait plein correspond à 45 % de la bande critique en ERB.

### 2.4.4 Hypothèse Temporelle

L'approche temporelle aboutit à une explication du phénomène de rugosité basée sur les paramètres des modulations d'amplitude effectivement perçues. Cette interprétation peut être résumée comme suit :

- la rugosité peut être prédite à partir de l'enveloppe temporelle des signaux issus de chaque filtre auditif ;
- la rugosité tend à augmenter avec la fréquence de modulation et avec la profondeur de modulation relative ;
- la rugosité est limitée par la largeur de la bande critique et par une caractéristique passe-bas de la perception d'enveloppe.

## 2.5 Lien entre les approches spectrales et temporelles

### 2.5.1 Convergence des indices acoustiques

Les deux approches que nous venons de séparer sur la base des stimuli qu'elles ont employés ont abouti à des interprétations basées soit dans le domaine spectral soit dans le domaine temporel. Elle sont néanmoins fortement corrélées. En effet, la manipulation de l'écart fréquentiel de la dyade de sons purs (approche spectrale, Figures 2.3) modifie l'enveloppe temporelle du signal, alors que la modification de la fréquence de modulation d'un son SAM (approche temporelle, Figure 2.7) change l'écart entre ses composantes fréquentielles. La première question à se poser est celle de la cohérence entre les résultats obtenus.

La rugosité de dyades de sons purs et de sons SAM a été directement comparée (Terhardt, 1974a). La rugosité d'un son SAM de fréquence de modulation 40 Hz et de profondeur de modulation variable a ajustée à celle d'une dyade de sons purs avec un écart fréquentiel de 40 Hz et de fréquence moyenne égale à la fréquence centrale du son SAM. L'ajustement a été réalisé pour une profondeur de modulation du son SAM entre 0,6 et 0,7, alors que la dyade possède une enveloppe dont la profondeur de modulation est maximale. Ceci est interprété par l'action du filtrage passe-bas de l'enveloppe, qui “adoucirait” la discontinuité de

l'enveloppe de la dyade (équation 2.1). L'amplitude de la première composante spectrale de l'expression de l'enveloppe est de 2/3 — cohérente avec la profondeur de modulation ajustée.

En ajoutant à cette comparaison directe la comparaison indirecte faite sur les intervalles de fin de rugosité et de rugosité maximale obtenus par les deux approches (Figures 2.5 et 2.11, discutée en section 2.4.3.2), il est est clair que dans les deux approches le même phénomène est mesuré et mis en jeu. Les approches spectrales et temporelles sont donc fortement convergentes, du fait de la corrélation entre les paramètres acoustiques des stimuli considérés jusqu'ici.

## 2.5.2 Divergences entre les approches : influence de la phase

### 2.5.2.1 Comparaison SAM/FM

Il existe un moyen simple de dissocier les indices acoustiques spectraux et temporel que nous venons d'évoquer : il suffit de manipuler la phase de sons possédant le même spectre d'amplitude. L'une des méthodes possible consiste à exploiter le lien entre modulation de fréquence et modulation d'amplitude. L'expression générale d'un son modulé en fréquence est donné par :

$$FM(t) = \sin[2\pi f_c t + \underbrace{\frac{\Delta_f}{f_m}}_{\beta} \sin(2\pi f_m t + \Phi)] \tag{2.5}$$

L'index de modulation $\beta$ est le rapport entre l'écart fréquentiel $\Delta_f$ couvert par la modulation et la fréquence de modulation $f_m$. Cet index n'a en général rien à voir avec la profondeur de modulation $m$ des sons SAM, un son FM ayant une enveloppe plate. Toutefois, si $\beta$ est faible, il est possible de montrer par un développement de Taylor que le spectre d'amplitude du son FM est très similaire à celui du son SAM avec $m = \beta$ (Hartmann, 1997). Les phases respectives des différents composants sont néanmoins différentes (la phase de la bande latérale grave est décalée de $\pi$ radians pour le son FM). Ainsi, la comparaison SAM-FM avec $\beta$ faible permet, en maintenant les indices spectraux approximativement constants, d'opposer deux enveloppes temporelles distinctes.

Terhardt (1974a) a comparé la rugosité produite par des sons SAM et FM. Les stimuli choisis étaient tous de fréquence porteuse égale à 1 kHz, avec $\beta = 0,5$, et de fréquence de modulation variable. La tâche des sujets consistait à ajuster

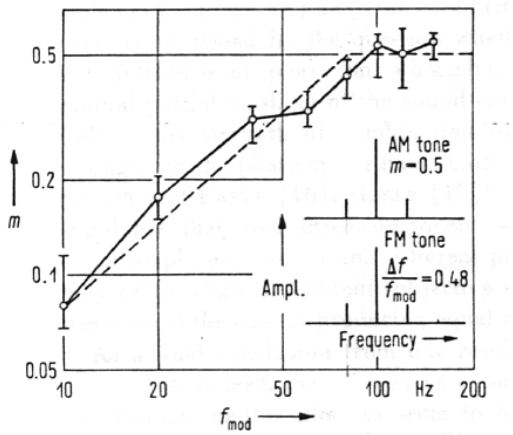


FIG. 2.12 – Comparaison de la rugosité d'un son SAM et d'un son FM en fonction de la fréquence de modulation. Pour chaque fréquence de modulation, la profondeur de modulation ajustée pour le son SAM est représentée. Si la phase n'influençait pas la rugosité, une ligne horizontale à $m = 0,5$ devrait être observée (voir texte). D'après Terhardt (1974).

le degré de modulation $m$ du son SAM pour obtenir la même rugosité que le son FM. Les résultats sont reproduits en Figure 2.12. Si seule l'amplitude des composantes spectrales déterminait la rugosité, tous les ajustements devraient être faits à $m = 0,5$. Ceci n'est pas observé : le son FM est généralement moins rugueux que le son SAM.

Terhardt interprète ceci grâce au degré de modulation effective. Trois régions peuvent être distinguées au vu des résultats. Si la fréquence de modulation est faible ($f_m < 80$ Hz), l'influence de la phase se traduit par une différence entre les enveloppes des stimuli. En effet, les sons sont contenus dans une bande critique et leur enveloppe acoustique est seulement légèrement altérée par le filtrage auditif. Dans ce cas, le son FM a une enveloppe quasi-constante alors que le son SAM est modulé. Il est jugé moins rugueux. Si la fréquence de modulation augmente ($80 < f_m < 160$ Hz), seules deux des composantes spectrales de chaque son sont séparées par moins d'une bande critique. Dans ce cas, les enveloppes sont altérées et la rugosité des sons devient comparable. Enfin, pour $f_m > 160$ Hz, toutes les composantes sont résolues et la rugosité des deux sons est nulle (mesure impossible).

#### 2.5.2.2 Comparaison SAM/QFM

L'influence de la phase sur la perception de sons avec un spectre d'amplitude rigoureusement constant a été abordée par Mathes et Miller (1947). Par synthèse analogique, trois composantes spectrales correspondant au spectre d'amplitude d'un son SAM ont été générées puis la phase d'une composante variée pour passer du son SAM à l'approximation d'un son FM. Ce son, ayant le spectre d'amplitude d'un son SAM mais le spectre de phase d'un son FM, est traditionnellement appelé Quasi-FM. Il possède une enveloppe qui n'est pas rigoureusement constante mais tout de même largement moins modulée que le son SAM correspondant.

La comparaison entre SAM et QFM indique que pour des fréquences de modulation entre 25 et 75 Hz, le son SAM est plus rugueux que le son QFM. L'explication proposée est aussi basée sur l'idée de sélectivité fréquentielle, permettant ou non la perception de l'enveloppe.

#### 2.5.2.3 Insuffisance des données sur l'effet de la phase

Les expériences sur l'influence de la phase dans la perception de rugosité sont cruciales pour permettre d'évaluer les approches spectrales et temporelles. En effet, par ces manipulations il est possible d'opposer les indices spectraux et temporels qui sont habituellement fortement covariants. Les résultats de Terhardt (1974a) et de Mathes et Miller (1947) vont dans le sens de l'hypothèse temporelle. Plusieurs points problématiques ressortent néanmoins à l'examen de ces données.

La comparaison SAM/FM de Terhardt (1974a) illustre l'effet de la bande critique sur le degré de modulation effective de manière ambiguë. La profondeur de modulation $m$ du son SAM doit être de plus en plus grande pour égaler la rugosité du son FM. Ceci traduit d'une part que l'efficacité de $m$ en terme de degré de modulation effective diminue lorsque les composantes deviennent résolues, mais aussi peut-être que le son FM devient plus rugueux. Lorsque seulement deux composantes fréquentielles sont à l'intérieur d'une même bande critique, les filtres auditifs adjacents peuvent devenir des sources indépendantes de rugosité. La rugosité de sons étendus fréquentiellement sur plus d'une bande critique est l'objet du Chapitre 3. Cette possibilité rend plus complexe l'interprétation des données.

Le spectre d'amplitude des sons FM et SAM employés par Terhardt (1974a) n'est pas rigoureusement identique. L'expérience de Mathes et Miller (1947) com-

pare des sons aux spectres d'amplitude identiques. Malheureusement, les auteurs ne donnent qu'une description qualitative des différences de rugosité entre SAM et QFM, sans réellement apporter de données quantitatives.

Enfin et surtout, pour Terhardt comme pour Mathes et Miller, seules deux conditions de phases "extrêmes" ont réellement été comparées : un son SAM opposé à un son FM (ou QFM). Les différents points expérimentaux de la Figure 2.12 permettent par exemple d'évaluer l'influence de la fréquence de modulation sur l'effet de phase, mais ne contribuent pas directement à qualifier cet effet de phase. Chaque point est une mesure indépendante de cet effet, avec seulement deux valeurs extrêmes du paramètre. Il est d'une part possible que ces deux conditions de phase aient conduit à des configurations particulières des sons de combinaison, permettant une interprétation spectrale des résultats (Buunen, Festen, Bilsen et van den Brink, 1974). D'autre part, la comparaison entre seulement deux conditions extrêmes peut faire qu'une partie des effets possibles de la phase n'aient pas pu être observés.

## 2.6 Conclusion

La rugosité, telle que définie par Helmholtz (1877), est l'effet particulier produit par la perception de battements rapides entre sons purs. Cet attribut du timbre des sons est perçu dès la naissance chez l'homme, ainsi que par d'autres espèces. En tant que tel, il semble posséder une certaine importance "écologique" et être lié à des mécanismes auditifs élémentaires. La définition de Helmholtz est double, et cette ambiguïté se retrouve chez les auteurs ultérieurs qui ont adopté des approches que nous avons nommées "spectrales" et "temporelles". L'approche spectrale a permis de souligner le lien entre rugosité et bande critique. L'approche temporelle a introduit la notion de profondeur de modulation effective de l'enveloppe. La confrontation des deux approches, notamment grâce à l'étude d'effets de phase, produit des résultats allant dans la direction de l'hypothèse temporelle. Néanmoins, ces résultats comportent des lacunes rendant toute conclusion délicate. Le Chapitre 5 se proposera d'apporter de nouveaux éléments sur ces effets de phase.

# Chapitre 3

# Rugosité de sons quelconques

*Les modalités de combinaison de rugosité au sein de sons complexes sont évoquées. Les modèles de calcul existants sont présentés.*

L'étude des causes de la perception de rugosité nous a amené jusqu'ici à considérer des stimuli hautement simplifiés, des dyades de sons purs ou des sons modulés en amplitude. La plupart des sons naturels ou musicaux diffèrent radicalement de ces stimuli sous plusieurs aspects. Parmi ces différences, l'une des plus importantes est sans doute leur étendue fréquentielle. Les contributions de régions fréquentielles éloignées les unes des autres doivent alors généralement être combinées. Les modalités de cette combinaison font l'objet de deux hypothèses distinctes : une simple addition entre régions ou bien l'intervention de mécanismes plus complexes. Les modèles explicites de calcul de rugosité sont donc amenés à faire un choix, leur présentation à la fin de ce Chapitre sera l'occasion de résumer les diverses hypothèses présentées jusqu'ici.

## 3.1 Combinaison de rugosités

### 3.1.1 Rugosités partielles

Les dyades de sons purs ou les sons SAM provoquant une certaine rugosité ont une étendue fréquentielle physique inférieure ou égale à une largeur de bande

critique. Il est donc possible pour comprendre leur perception de ne considérer que le résultat du filtrage auditif autour de leurs fréquences centrales. La plupart des sons réels couvrent eux une large bande fréquentielle. Néanmoins, sous l'effet du filtrage auditif, ces sons vont être décomposés en sous-bandes correspondant aux bandes critiques. Une manière de relier la rugosité de tels sons avec les résultats précédents est donc de considérer successivement chacune des sous-bandes et de définir pour chacune d'elle une "rugosité partielle".

La combinaison de ces rugosités partielles devient alors le problème clé. Les résultats expérimentaux disponibles sont contradictoires, une partie d'entre eux tendant à prouver que les effets des rugosités partielles s'additionnent simplement alors que d'autres indiquent que cette addition peut être mise en défaut. Nous allons donc organiser la revue des résultats concernant la rugosité de sons autres que les dyades ou les sons SAM autour de cette opposition.

### 3.1.2 Eléments en faveur d'une addition entre régions fréquentielles

#### 3.1.2.1 Masquage de rugosité

Les sons SAM étudiés par l'approche temporelle ont une étendue fréquentielle physique inférieure à une bande critique. Il est bien évident qu'en fonction du niveau de présentation, leurs patterns d'excitation le long de la membrane basilaire peuvent s'étendre sur plus d'une bande critique. Dans ce cas, le problème de la combinaison de rugosités se pose déjà. L'importance des régions fréquentielles éloignées de la fréquence centrale du son SAM dans le percept global de rugosité a été estimée par des expériences de "masquage de rugosité" (Vogel, 1974).

Les stimuli étudiés sont composés de l'addition d'un son SAM et d'un son masquant (un son pur de niveau élevé ou une bande de bruit). Dans tous les cas, les résultats indiquent une diminution de rugosité pour les stimuli "masqués". Les conditions les plus intéressantes sont toutefois celles ou le masquant est relativement éloigné du son SAM et situé au-dessus de sa fréquence centrale. Dans ce cas, du fait de l'asymétrie des patterns de masquage (Egan et Hake, 1950), le masquant ne modifie que la partie supérieure du pattern d'excitation du SAM. La réduction de rugosité qui est néanmoins observée montre que l'ensemble du pat-

tern d'excitation contribue à la rugosité globale. Les contributions des différentes régions semblent additives.

#### 3.1.2.2 Trains d'impulsions

Un train d'impulsions est le son périodique ayant la plus large étendue spectrale possible, car constitué théoriquement d'une infinité de partiels d'égale amplitude correspondant aux harmoniques de la fréquence de répétition. Terhardt (1974a) a étudié la rugosité des trains d'impulsions.

Ces déterminations se sont avérées délicates. En effet, même à des fréquences très graves de l'ordre de 20 Hz, le train d'impulsions est rugueux du fait de l'interférence entre les partiels de rangs harmoniques élevés. Si n'importe quelle région fréquentielle dans laquelle existe des composantes irrésolues est atténuée par filtrage, la rugosité diminue. La rugosité est donc qualitativement composée des contributions de l'ensemble des régions fréquentielles (Terhardt, 1974a).

#### 3.1.2.3 Sons modulés en fréquence

Les sons modulés en fréquence (FM, Equation 2.5) peuvent eux aussi couvrir une large étendue fréquentielle. Si l'index de modulation a une valeur élevée, le spectre du son FM s'enrichit d'un grand nombre de raies et sa fréquence balaye régulièrement plusieurs bandes critiques.

La rugosité de sons FM a été étudiée par Kemp (1982). Une méthode d'estimation directe de grandeur avec 8 sujets a été employée. La fréquence de modulation a une influence sur la rugosité. Pour une fréquence centrale de $f_c = 1600$ Hz, la courbe de rugosité présente une caractéristique passe-bande avec un maximum autour de 50 Hz. Une influence de l'excursion en fréquence est aussi observée : les stimuli avec un $\Delta_f$ faible ne provoquent pas de rugosité, puis la rugosité augmente avec $\Delta_f$. Ces résultats sont résumés en Figure 3.1.

Ces résultats présentent certaines similarités avec ceux observés pour les sons SAM. Ceci peut être rapproché du fait que la modulation de fréquence introduit une modulation d'amplitude dans les différents canaux auditifs (McAdams et Rodet, 1988). En effet, pour chaque période de modulation, le son FM balaye une région fréquentielle égale à $\Delta_f$. Si $\Delta_f$ est très inférieur à une bande critique, le balayage n'introduit qu'une faible fluctuation dans les régions correspondant aux limites du pattern d'excitation et la rugosité observée est nulle. Si par contre $\Delta_f$

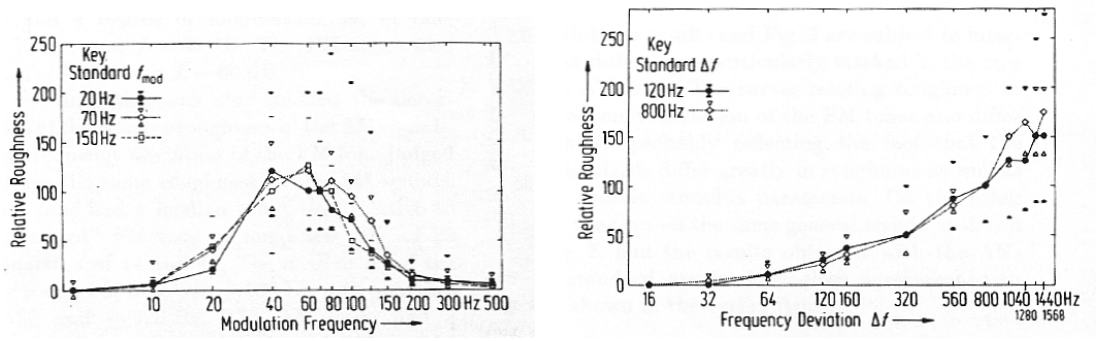


FIG. 3.1 – Rugosité de sons FM en fonction de la fréquence de modulation $f_m$ pour $f_c = 1600$ Hz et $\Delta_f = 800$ Hz (à gauche) et de l'excursion en fréquence $\Delta_f$ pour $f_c = 1600$ Hz et $f_m = 70$ Hz (à droite). D'après Kemp (1982).

couvre plusieurs bandes critiques, une large fluctuation d'amplitude à la fréquence $f_m$ est observée dans chacune d'entre elles. Si l'on admet une addition entre les rugosités partielles, l'influence de $f_m$ dans chaque canal est alors semblable à celle observée pour les sons SAM.

#### 3.1.2.4 Bruit large bande modulé

Fastl (1977) a étudié la rugosité produite par un bruit large bande modulé sinusoïdalement en amplitude. Ce bruit était un bruit blanc atténué de 3 dB/oct à partir de 700 Hz pour être "uniformément masquant". Pour six sujets et avec une méthode d'estimation directe, Fastl montre qu'un bruit modulé en amplitude peut produire une rugosité appréciable, supérieure même à celle provoquée par les sons SAM. La rugosité dépend de la fréquence de modulation. Une caractéristique passe-bande est observée avec un maximum de rugosité pour $f_m = 70$ Hz. La profondeur de modulation influence aussi la rugosité selon une loi d'exposant (Figure 3.2).

La ressemblance entre l'influence des paramètres acoustiques de la modulation sur la rugosité pour le bruit modulé et les sons purs modulés indique un rapport possible entre les deux types de phénomènes. L'hypothèse de rugosités partielles liées aux enveloppes dans chaque canal auditif combinée à l'addition des contributions de chaque canal permettrait d'expliquer les résultats observés.

Ces résultats indiquent aussi qu'un son sans structure spectrale à long terme peut être rugueux. Ceci est un élément supplémentaire contre l'hypothèse spec-

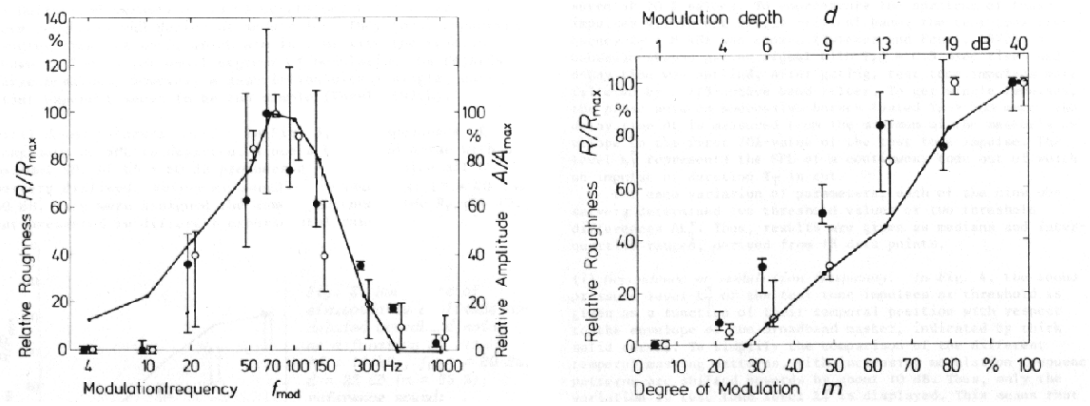


FIG. 3.2 – Rugosité de bruit blanc sinusoïdalement modulé en amplitude en fonction de la fréquence de modulation (à gauche) et de la profondeur de modulation (à droite). D'après Fastl (1977).

trale. Il convient néanmoins de prendre quelques précautions dans son interprétation car la simple addition de rugosités partielles produites par des indices temporels dans chacun des canaux auditifs est contestée.

### 3.1.3 Éléments opposés à une addition entre régions fréquentielles

#### 3.1.3.1 Superposition de sons SAM

La rugosité d'un son SAM isolé a été largement étudiée. Certains auteurs ont étudié la rugosité de deux sons SAM présentés simultanément pour tenter de comprendre l'additivité de rugosités partielles.

Mathes et Miller (1947) ont superposé deux sons SAM de même fréquence de modulation, de même profondeur de modulation mais situés à des fréquences centrales différentes. Ils affirment alors que la rugosité globale dépend des relations de phase entre enveloppes : si les deux sons sont modulés de façon co-phasique, c'est-à-dire si leurs enveloppes atteignent leurs maxima et minima aux mêmes instants, la rugosité est supérieure à celle obtenue si les enveloppes sont anti-phasiques, c'est-à-dire quand à un maximum de l'une correspond un minimum de l'autre (Figure 3.3). Cet effet est d'autant plus important que la différence entre fréquences centrales est faible. Ceci va à l'encontre de l'hypothèse d'addi-

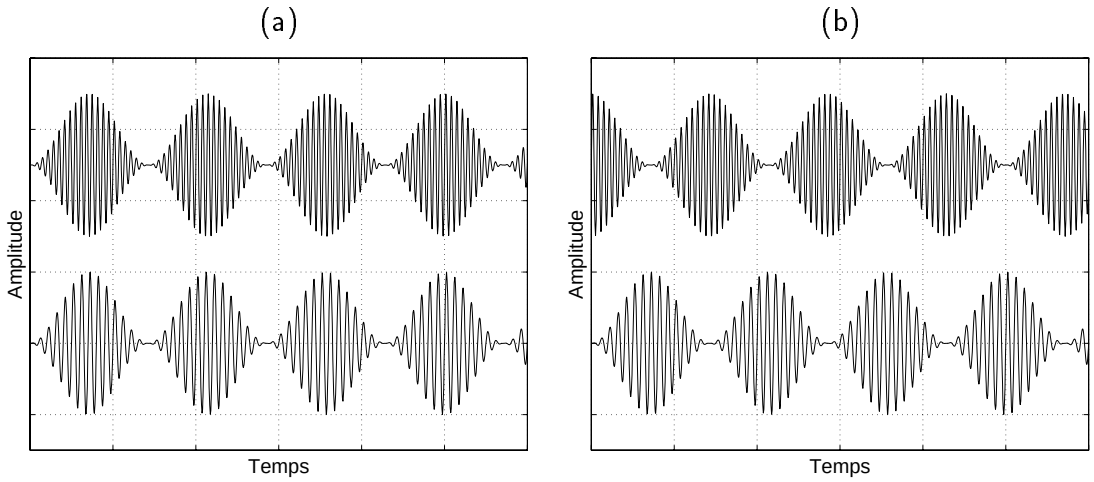


FIG. 3.3 – Illustration des conditions co-phasiques (a) et anti-phasiques (b) pour la superposition de deux sons SAM. Entre les deux sons SAM additionnés les fréquences porteuses varient, la fréquence de modulation reste constante et la phase de la modulation varie.

tion, car les rugosités partielles sont théoriquement inchangées entre les cas co- et anti-phasiques.

Terhardt (1974a) a repris le même genre de stimuli pour recueillir des données quantitatives sur cet effet, en superposant deux sons SAM notés ici $SAM_1$ et $SAM_2$. La fréquence de modulation est fixée à $f_{m1} = f_{m2} = 40$ Hz. Les fréquences centrales, associées par couples, sont comprises entre 3.8 et 2 kHz pour $f_{c1}$ et 4.2 et 7 kHz pour $f_{c2}$. Ceci permet de faire varier l'écart fréquentiel $f_{c2} - f_{c1}$ de 400 à 5000 Hz suivant les couples tout en conservant une fréquence moyenne $(f_{c1} + f_{c2})/2$ de 4 kHz. Les sons $SAM_1$ et $SAM_2$ aux différentes fréquences centrales sont d'abord comparés à une référence commune dans le but d'ajuster leurs profondeurs de modulation pour qu'ils produisent tous isolément une rugosité identique. L'addition des $SAM_1$ et $SAM_2$ est ensuite comparée à la référence, dont la profondeur de modulation est maintenant variable. La référence est telle que $f_{c,ref} = 4000$ Hz, $f_{m,ref} = 40$ Hz. Les conditions d'enveloppes co-phasiques et anti-phasiques sont testées, de façon similaire à Mathes et Miller (1947). Si l'addition des deux sons $SAM_1$ et $SAM_2$ se traduisait par une addition de leurs rugosités respectives, la rugosité totale par rapport à la référence devrait doubler. Du fait des valeurs numériques retenues, ce doublement correspond à une profondeur ajustée de $m_{(1+2)} = \sqrt{2}/2$.

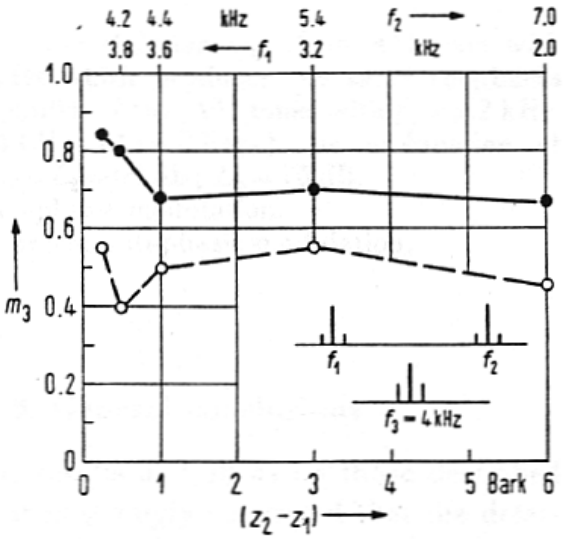


FIG. 3.4 – Superposition de deux sons SAM. La fréquence moyenne de l'ensemble des deux sons est maintenue constante et égale à 4 kHz. Leur écart fréquentiel est représenté en abscisse (en Bark en bas et en Hertz en haut). Les traits pleins représentent les enveloppes co-phasiques, les traits pointillés les enveloppes anti-phasiques (voir texte). D'après Terhardt (1974).

La profondeur de modulation ajustée est présentée en Figure 3.4 (5 sujets). L'examen des résultats montre que l'addition des rugosités partielles (m = 0,7) n'est observée que pour les sons SAM écartés de plus d'une bande critique et d'enveloppes co-phasiques. Les conditions co-phasiques sont plus rugueuses que les conditions anti-phasiques pour toutes les séparations. Dans le cas d'enveloppes anti-phasiques et si les sons sont écartés de plus d'une bande critique, la rugosité des deux sons est égale à celle d'un seul d'entre eux. Dans le cas où les sons sont écartés de moins d'une bande critique, la différence semble être plus marquée. Ces effets complexes sont donc en opposition avec une simple addition de rugosités partielles.

#### 3.1.3.2 Rugosité de bruits filtrés

Un signal aléatoire possède des fluctuations intrinsèques qui se reflètent dans son enveloppe. Pour un bruit blanc, ces fluctuations aléatoires se compensent et l'enveloppe est pratiquement plate. En revanche, pour un bruit filtré, l'enveloppe présente des maxima et des minima bien définis (Hartmann, 1997). Pour caractériser cette enveloppe, une fréquence de modulation équivalente peut être calculée

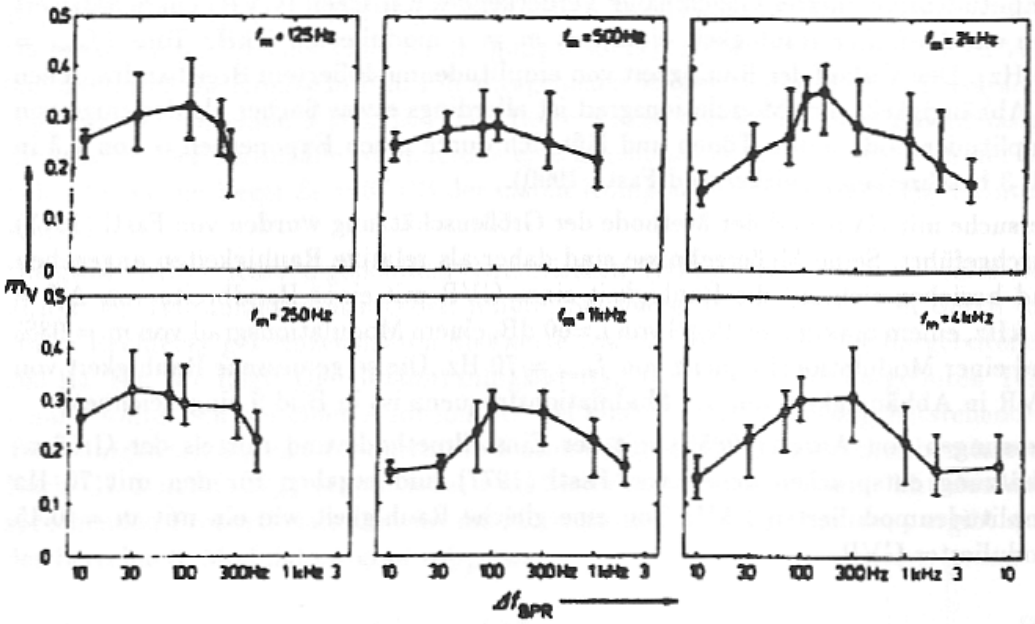


FIG. 3.5 – Influence de la largeur de bande et de la fréquence centrale de la bande sur la rugosité de bruit filtrés. Pour chaque fréquence centrale, la largeur de bande est en abscisse et la profondeur de modulation ajustée d'un son de comparaison est en ordonnée. D'après Aures (1985).

en estimant la moyenne des intervalles de temps entre maxima. Il est clair que les modulations d'enveloppe plus rapides que la demi-largeur de bande vont être atténuées par le filtrage. On peut ainsi montrer que la fréquence de modulation équivalente d'un bruit blanc filtré est de $f_{m,eq} = 0,64.\Delta_B$, où $\Delta_B$ est la bande passante du filtrage (Aures, 1985b).

Aures (1985b) a étudié la rugosité de bruits filtrés en faisant varier la fréquence centrale du filtrage ainsi que la largeur de la bande passante. Les résultats, obtenus par ajustement de la profondeur de modulation d'un son SAM de référence, sont repris en Figure 3.5. La rugosité dépend de la largeur de bande : une bande trop étroite ou trop large ne produit que peu ou pas de rugosité. La rugosité maximale est atteinte pour une bande de largeur approximativement égale à 100 Hz.

En partant de la largeur de bande la plus faible (10 Hz), il est aisé d'interpréter l'augmentation de rugosité avec $\Delta_B$ selon l'hypothèse temporelle : plus la bande s'élargit et plus des fluctuations rapides apparaissent, provoquant de la rugosité. Le maximum de rugosité est observé pour une modulation équivalente de 64 Hz. Ainsi, des signaux qui comportent des modulations d'amplitude de l'ordre de

70 Hz, même si elles sont combinées avec des modulations plus lentes, produisent une rugosité appréciable.

La diminution de rugosité pour les largeurs de bande supérieures à 100 Hz est plus surprenante. Nous savons que le système auditif réalise une décomposition en bandes de fréquence du signal. Ainsi, pour des largeurs de bandes supérieures à une bande critique, le signal sera décomposé en sous-bandes. Tout se passe donc comme si l'on ajoutait côte à côte des bandes de bruits filtrées avec une largeur égale à la bande critique à la fréquence centrale considérée. Des fluctuations comportant des fréquences susceptibles de produire une rugosité se retrouvent donc sur un grand nombre de canaux auditifs quand on élargit le filtrage, et pourtant la rugosité diminue au lieu d'augmenter. Ce résultat est intuitivement connu : le bruit blanc non modulé produit une rugosité pratiquement nulle, alors qu'une addition des fluctuations qu'il provoque dans chaque bande critique devrait conduire à une très forte rugosité. La diminution observée va donc à l'opposé de l'hypothèse de sommation des rugosités partielles entre les différentes bandes critiques.

### 3.1.4 Résultats expérimentaux contradictoires

Les expériences permettant le contrôle du plus grand nombre de paramètres en vue de comprendre l'influence de la cohérence entre enveloppes temporelles dans la combinaison de rugosité sont celles concernant la superposition de deux sons SAM. Vogel (1974) a tenté de reproduire ce type d'expérience. La même tendance est observée, à savoir que les modulations anti-phasiques sont moins rugueuses que les modulations co-phasiques. Si l'effet est observé pour toutes les séparations fréquentielles testées, il est relativement faible et au vu des écarts-type il n'est statistiquement significatif pour aucune séparation considérée indépendamment.

Aures (1985b) a réalisé une expérience similaire en étudiant l'addition de deux sons SAM, modulés à 40 Hz, l'un centré à 300 Hz et l'autre centré à 3000 Hz. La profondeur de modulation du son SAM le plus grave a été fixée, celle du plus aigu est variée pour obtenir un plus grand nombre de points expérimentaux. Les conditions co- et anti-phasiques sont opposées. Aucune différence n'est observée entre ces conditions. En augmentant le niveau de la composante grave, un léger effet de la phase relative des enveloppes apparaît. Cet effet devient important si les fréquences porteuses sont proches.

L'apparition de la différence entre co- et anti-phasique pour des niveaux élevés ou des écarts fréquentiels faibles est interprétée par le plus grand recouvrement entre patterns d'excitation dans ces conditions. En effet, si les enveloppes sont en opposition de phase, leur superposition va avoir tendance à s'annuler. Les zones fréquentielles où les patterns d'excitation des deux SAM se recouvrent comportent alors moins de fluctuations pouvant contribuer à la rugosité globale.

Cette interprétation des données de Aures (1985b) ne peut être étendue à celles de Terhardt (1974a). Pour le point de plus grande différence fréquentielle étudié par Terhardt (5 kHz), l'effet est encore présent alors que les patterns d'excitation ne peuvent pas se recouvrir. Il existe donc une divergence entre ces deux ensembles de résultats, ceux de Vogel (1974) se trouvant entre les deux.

### 3.1.5 Bilan sur la combinaison de rugosités

Les résultats expérimentaux présentés ne permettent pas de définir les modalités de combinaison de rugosités partielles. Qualitativement, l'addition des contributions de différentes régions fréquentielles est supportée par de nombreuses expériences (Vogel, 1974; Terhardt, 1974a; Kemp, 1982; Fastl, 1977). Toutefois, l'addition rigoureuse n'est parfois pas observée. En ce qui concerne la rugosité des sons FM, supposée créée par la combinaison des contributions des différents canaux fréquentiels, elle n'est pas en rapport direct avec le nombre de bandes critiques couvertes par la largeur de de modulation (Kemp, 1982).

Des situations où l'addition qualitative n'est elle-même plus valable ont été démontrées. Lors de l'addition de sons SAM dans les conditions anti-phasiques, l'ajout de rugosités partielles laisse la rugosité globale inchangée. Pour les bruits filtrés l'ajout de rugosités partielles diminue la rugosité globale. Un point commun relie ces derniers résultats : il semble que la cohérence des fluctuations entre les bandes critiques joue un rôle dans l'intégration des rugosités partielles. En effet, si l'on calcule un indice qui mesure la cohérence entre enveloppes (comme un coefficient d'intercorrélation), cet indice sera supérieur dans le cas co-phasique que anti-phasique. Pour un bruit large bande, les modulations introduites par le filtrage auditif ne sont pas corrélées. Ainsi, il semble que l'addition de rugosité soit qualitativement vérifiée lorsque les enveloppes sont cohérentes, alors qu'elle est fausse lorsque les enveloppes sont incohérentes.

Cette idée générale, avancée par plusieurs auteurs (Mathes et Miller, 1947; Terhardt, 1974a; Vogel, 1974), est néanmoins la source de divergences car une partie au moins de l'effet semble être lié à un recouvrement entre patterns auditifs. Nous proposerons de nouvelles données sur ce sujet dans le Chapitre 7. Auparavant, la présentation des modèles de calcul de rugosité de sons quelconques va nous permettre de synthétiser les hypothèses adoptées par les différentes approches.

## 3.2 Estimation de rugosité de sons quelconques

### 3.2.1 Modèle analytique de Helmholtz

#### 3.2.1.1 Cas de deux sons purs

Helmholtz propose le premier un modèle analytique de calcul de rugosité (von Helmholtz, 1877). Les modalités précises du calcul sont rejetées en annexe de l'ouvrage original, et les détails en sont assez complexes. Nous allons en présenter un bref résumé. L'idée de base est que la rugosité d'une paire de son purs est déterminée par le produit de deux fonctions, l'une reliée à l'énergie des battements qu'ils provoquent le long de la membrane basilaire et l'autre à une "rugosité propre" qui exprime le fait que la plus grande rugosité est obtenue selon Helmholtz pour des battements de l'ordre de 30 Hz. L'énergie des battements est dérivée d'équations simplifiées du mouvement de la membrane basilaire, la fonction de rugosité propre est arbitraire. Quelques illustrations graphiques proposées en Figure 3.6 permettent de comprendre que les deux facteurs capturent commodément diverses influences remarquées expérimentalement. L'énergie des battements est en fait un facteur spectral, qui limite la rugosité en fonction de l'écart fréquentiel et qui dépend du registre. La rugosité propre, elle, est plutôt un facteur temporel qui relie rugosité et vitesse des battements.

Dans le cas de sons complexes, tous les partiels peuvent interagir entre eux et donner lieu à un ensemble de rugosités partielles. De manière à quantifier la rugosité produite par deux sons complexes, Helmholtz réplique le calcul pour deux sons purs à tous les couples de partiels possibles. Pour ceci, Helmholtz utilise ses propres calculs sur le mouvement de la corde frottée du violon et injecte les fréquences et amplitudes relatives des partiels dans les formules analytiques. Le résultat est représenté en Figure 3.7, chaque trait représentant la rugosité d'une

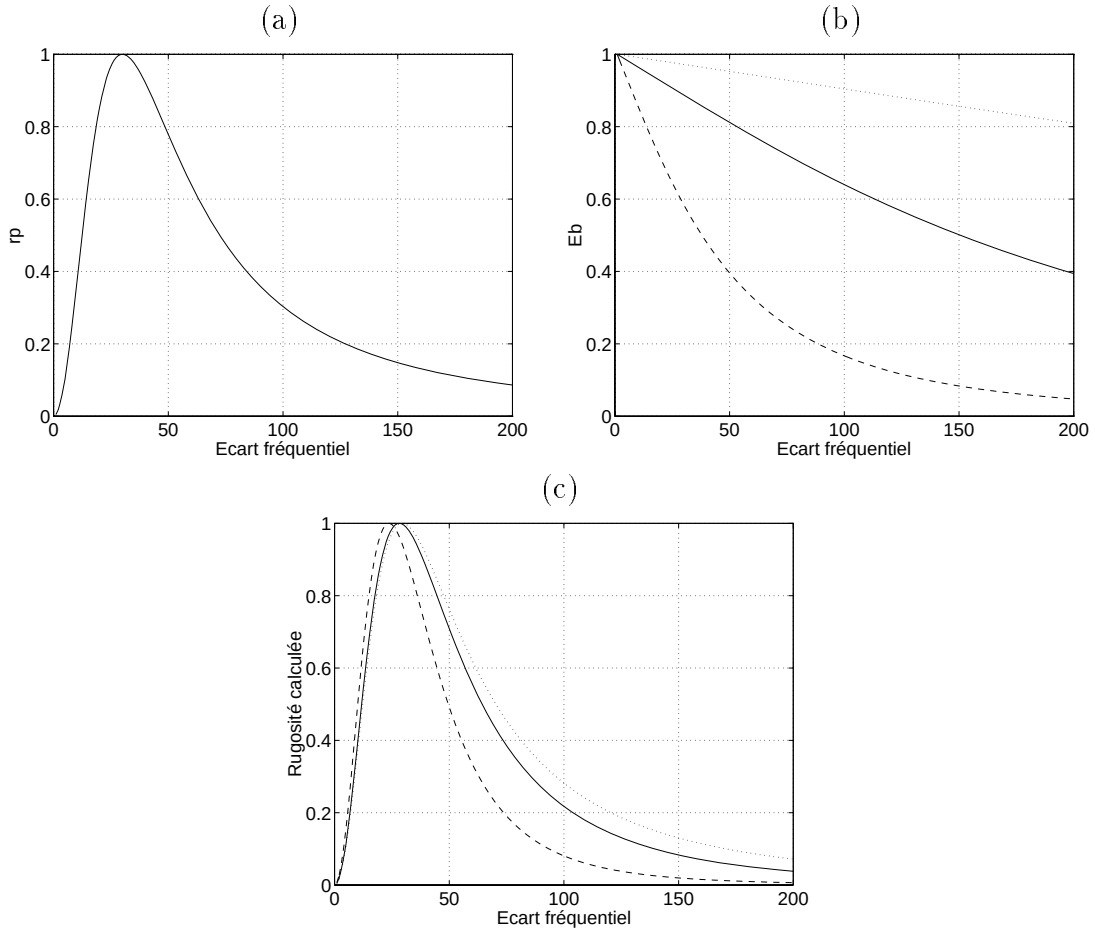


FIG. 3.6 – Illustration des l'influence des deux facteurs de la formule analytique de calcul de rugosité selon Helmholtz (1877). La fonction de rugosité propre est donnée en fonction de l'écart fréquentiel (a). L'énergie des battements (b) est donné pour trois fondamentales différentes (trait plein : Do milieu, 262 Hz ; trait interrompu : 4 octaves plus grave ; trait pointillé : 4 octaves plus aigu). La multiplication des deux termes reproduit la disparition plus rapide de rugosité dans le registre grave (c).

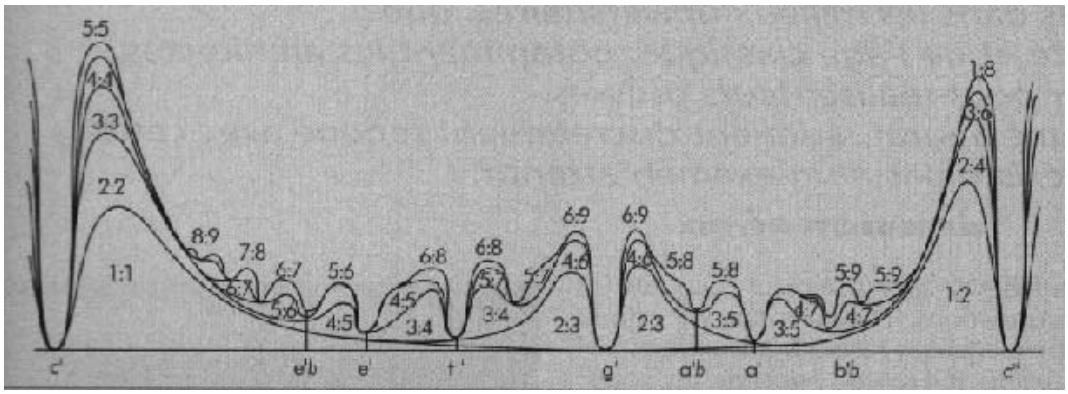


FIG. 3.7 – Application de la formule analytique de calcul de rugosité aux paires de partiels de deux sons à la composition spectrale proche de celle du violon. Le son grave est fixé à Do milieu (264 Hz) et le son le plus aigu varie continuement sur un octave. Chaque trait représente la rugosité d'une paire de partiels repérée par une fraction (p. ex. 1 : 2 indique la rugosité produite par le fondamental du son aigu et l'octave du son grave). D'après Helmholtz (1877)

paire de partiels. Remarquons que Helmholtz se garde bien d'additionner les rugosités partielles provoquées par chaque paire d'harmoniques : il laisse implicitement au lecteur la responsabilité d'effectuer mentalement cette opération !

L'observation de ce graphe démontre que certains intervalles se situent à des creux de rugosité : parmi ceux-ci l'unisson, l'octave, la quinte. A l'inverse, la seconde mineure, la septième majeure ou le triton possèdent de nombreuses paires de partiels susceptibles de produire une rugosité.

Plusieurs contradictions ou négligences sont présentes dans le modèle de Helmholtz (Bailhache, 1986). Ce modèle imparfait démontre de plus des désaccords quantitatifs avec les données expérimentales ultérieures (Plomp et Levelt, 1965 par exemple). Il capture néanmoins les principaux traits de la perception de rugosité. Helmholtz en convient d'ailleurs : "Bien que la rigueur de cette théorie laisse encore beaucoup à désirer, elle nous suffit pour faire voir que notre hypothèse peut réellement expliquer la répartition des dissonances et des consonances, telle qu'elles apparaissent dans la nature". La discussion de cette formule — lapidaire — sera la matière du Chapitre 4.

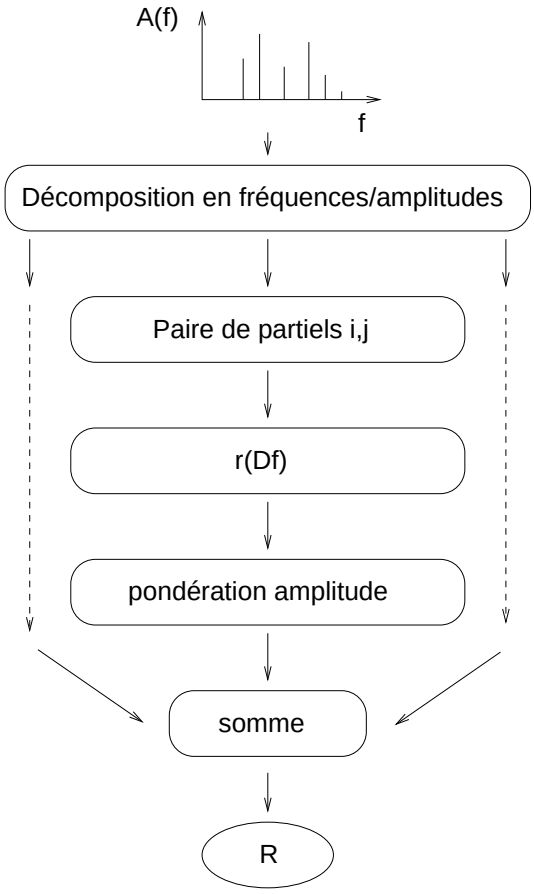


FIG. 3.8 – Le principe du modèle spectral de Hutchinson et Knopoff (1978). A partir de la composition spectrale du son à évaluer (densité spectrale de puissance, par exemple) les rugosités de toutes les paires de partiels sont calculées et ajoutées après pondération des amplitudes.

### 3.2.2 Modèles spectraux

Plusieurs modèles explicites de calcul de rugosité ont été proposés à la suite des résultats de l'approche spectrale. Leur principe, suggéré par Plomp et Levelt (1965), consiste à additionner les rugosités partielles des couples de partiels formant la représentation spectrale des sons complexes (Hutchinson et Knopoff, 1978; Sethares, 1993). Ces implémentations sont très similaires, nous présenterons le modèle de Hutchinson et Knopoff (1978) car il a été utilisé dans des études qui nous intéresseront plus tard (Vos, 1986; Bigand, Parncutt et Lerdahl, 1996).

Le principe de ce modèle est schématisé en Figure 3.8. Une liste de fréquences et d'amplitudes $(f_i, a_i)$ représente le son à analyser. Pour chaque paire possible

de composantes $(i, j)$, une rugosité partielle $r_{ij}$ est calculée[1] en fonction de leur différence fréquentielle selon les résultats de Plomp et Levelt (1965). Toutes les rugosités partielles, pondérées par les amplitudes des composants respectifs, sont ensuite additionnées pour fournir la rugosité totale. La rugosité estimée $r$ pour un son comportant $N$ composantes de fréquences et d'amplitudes notées $(f_i, a_i)$ est donnée en Equation 3.1.

$$r = \frac{\Sigma_{i,j=1}^{N} a_i a_j r_{i,j}}{\Sigma_{i=1}^{N} a_i^2} \tag{3.1}$$

avec

$$\begin{cases} r_{i,j} & = & (2.7183 * \delta f_{norm} * exp^{-\delta f_{norm}})^2 \\ \delta f_{norm} & = & \frac{f_i - f_j}{cb_{i,j} * 0,25} \\ cb_{i,j} & = & 1,72 * (\frac{f_i + f_j}{2})^{0,65} \end{cases} \tag{3.2}$$

Ce modèle ajusté aux données de Plomp et Levelt (1965) pour des sons purs fournit qualitativement les mêmes résultats que celui de Helmholtz pour les sons complexes, avec un coût de calcul faible. Les modalités d'addition des rugosités partielles en fonction du niveaux des composantes sont arbitraires. Des raffinements ont été proposés pour corriger ce point (Kameoka et Kuriyagawa, 1969b). Les deux versions du modèle ont toutefois échoué à prédire de façon fine la rugosité d'intervalles de sons complexes (Vos, 1986). Les modèles spectraux ne peuvent de plus pas prédire les effets de phase (Mathes et Miller, 1947; Terhardt, 1974a; Hutchinson et Knopoff, 1978), ni la rugosité des bruit modulés en amplitude (Fastl, 1977) ou filtrés (Aures, 1985b).

### 3.2.3 Modèles temporels

Les modèles dérivant de l'hypothèse temporelle estiment la rugosité à partir de l'onde de pression des sons à analyser. Une simulation du filtrage auditif est réalisée, et la rugosité est déduite des fluctuations d'enveloppe présentes dans chacun des canaux auditif. Terhardt n'a pas à notre connaissance proposé de modèle explicite. Un modèle inspiré de l'hypothèse temporelle et utilisé dans l'industrie a été développé par Zwicker et Fastl (1990). Un autre modèle a été défini

1. Le modèle original emploie des tables, nous présentons la formule analytique ajustée par Parncutt (communication personnelle).

par Aures (1985b) et amélioré par Daniel et Weber (1997). Nous présenterons ce dernier modèle du fait de la disponibilité de données publiées le concernant.

Le principe du modèle est illustré en Figure 3.9. La première étape de traitement consiste à décomposer un signal sonore en bandes critiques. Le filtrage réalisé est non-linéaire, il dépend de l'amplitude pour reproduire les effets d'élargissement du pattern d'excitation avec le niveau. Le signal issu de chaque filtre est ensuite redressé, puis filtré passe-bande. Ce filtrage a pour effet d'extraire l'enveloppe du signal redressé et de la pondérer selon les fréquences "utiles" à la rugosité. Un filtre différent est utilisé selon les canaux. La modulation effective est ensuite calculée en extrayant la valeur $rms$ de l'enveloppe divisée par la valeur moyenne. Un coefficient de pondération est ensuite appliqué aux modulations effectives pour reproduire la baisse de rugosité aux hautes et basses fréquences. Enfin, chaque rugosité partielle est pondérée par un coefficient d'intercorrélation avec la bande critique voisine, et l'ensemble ajouté pour donner la rugosité finale.

Le modèle fournit dans l'ensemble une très bonne reproduction de la plupart des données disponibles, notamment dans sa version la plus récente (Daniel et Weber, 1997). Il comporte néanmoins nombre de paramètres arbitraires, comme la forme des filtrages passe-bande ajustés aux résultats concernant la dépendance de rugosité sur la fréquence de modulation, le coefficient de pondération pour reproduire l'effet de la fréquence centrale sur la rugosité, et le coefficient d'intercorrélation final.

Cette dernière étape d'intercorrélation reflète une hypothèse faite sur l'addition des rugosités partielles. Elle est suffisante pour limiter la rugosité du bruit blanc, mais ne reproduit pas l'effet (contesté par Aures) de la phase des enveloppes pour les larges séparations fréquentielles. Les données disponibles sont toutefois insuffisantes pour évaluer le statut de cette étape d'intercorrélation : simple commodité de calcul ou reflet d'un mécanisme fondamental? Le fait que ce modèle soit très efficace pour reproduire la plupart des données existantes suggère qu'il serait intéressant de tester directement les hypothèses qu'il adopte, de manière à évaluer ses possibilités de généralisation aux sons pour lesquels il n'a pas été calibré.

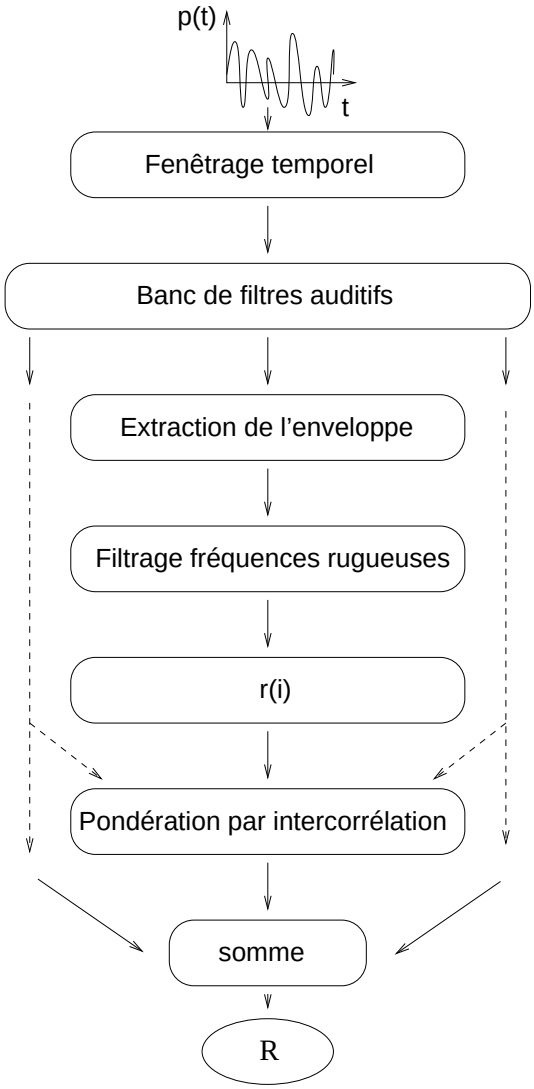


FIG. 3.9 – Le principe du modèle temporel de Aures (1985) et Daniel et Weber (1997). A partir de l'onde temporelle du son à évaluer, les enveloppes "effectives" dans toutes les bandes critiques sont obtenues, les rugosités partielles calculées et ajoutées après pondération par l'intercorrélation entre canaux adjacents.

## 3.3 Conclusion

L'extension des résultats obtenus pour des dyades de sons purs ou pour des SAM à l'estimation de rugosité de sons quelconques passe par une compréhension des modalités de combinaison des rugosités partielles à travers les canaux auditifs. Les données expérimentales montrent que si dans certaines conditions, les rugosités partielles s'additionnent, des situations où la cohérence entre enveloppes temporelles influence la combinaison de rugosité sont aussi observées. L'interprétation de l'effet de cohérence d'enveloppe est source de divergence entre les auteurs, certaines données apparaissant directement contradictoires. Les modèles de rugosité adoptent donc chacun des hypothèses différentes pour permettre l'évaluation des sons quelconques. Le modèle de Aures (1985b) notamment permet de reproduire la plupart des données disponibles. Néanmoins, l'hypothèse retenue pour combiner les rugosités partielles entre les canaux ne peut être justifiée ou infirmée sur la base des résultats expérimentaux disponibles. Nous nous proposons d'apporter dans le Chapitre 7 de nouvelles données à ce sujet.

# Chapitre 4

# Perception de consonance musicale

*Les théories perceptives de la consonance, dont celle liée à la rugosité, sont présentées. Les perspectives d'extension hors du contexte tonal sont soulignées.*

La dimension psychoacoustique de rugosité a été définie à l'origine pour proposer une base perceptive à l'harmonie musicale (von Helmholtz, 1877). Il est clair que les questions soulevées par une telle théorie vont être de nature très différente de celles qui nous intéressaient il y a quelques pages seulement. Une tentative de définition préliminaire du mot "consonance" sera donc proposée selon une perspective historique, en soulignant que la notion a en fait évolué. Nous présenterons ensuite les théories perceptives de la consonance, que nous séparerons en deux groupes : les théories psychoacoustiques, dont celle basée sur la rugosité, et les théories cognitives. Pour la musique harmonique tonale, ces théories permettent de décrire plusieurs aspects complémentaires de la notion de consonance qui sont étroitement imbriqués. Les tentatives d'extension hors du contexte tonal seront enfin évoquées.

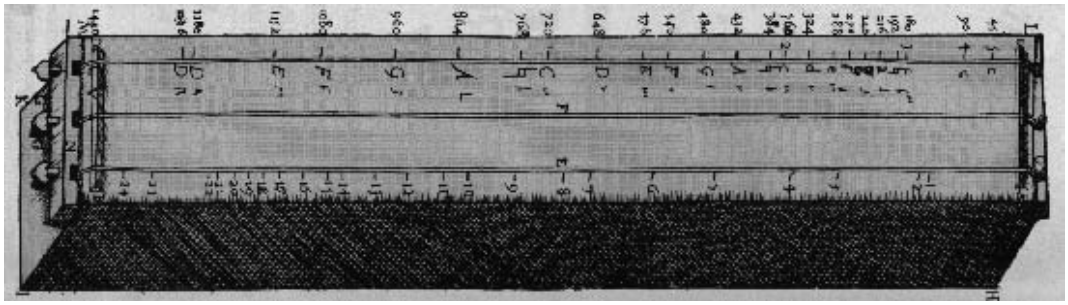

FIG. 4.1 – Le monocorde reproduit dans l'Harmonie Universelle de Mersenne (1636). D'après Assayag et Cholleton (1995).

## 4.1 Consonance et dissonance

### 4.1.1 L'énigme posée par Pythagore

Helmholtz affirme dans son ouvrage, avec un certain aplomb, avoir résolu grâce au concept de rugosité "l'énigme posée il y a 2500 ans à la science par Pythagore", à savoir le problème de la consonance musicale. Cette énigme est habituellement exprimée sous la forme d'une anecdote (James, 1993) — dont "l'authenticité" n'a que peu d'importance ici.

Alors qu'il se promenait dans la cité, Pythagore passa devant l'échoppe d'un forgeron. Les artisans, affairés à leur tâche, frappaient avec de lourds maillets diverses pièces métalliques. Ce qui aurait dû produire un vacarme épouvantable évoquait pourtant une impression "harmonieuse". Intrigué, Pythagore pénétra dans l'échoppe et découvrit que les masses des différents maillets entretenaient un rapport numérique particulier. Il construisit alors divers dispositifs reproduisant ce qu'il avait observé : de ces expérimentations est né le monocorde, instrument central à l'enseignement et aux spéculations sur la musique pendant tout le Moyen Âge – et toujours apprécié des professeurs d'acoustique (Figure 4.1). Ce dispositif permet, en ajustant à volonté la longueur d'une corde vibrante tout en conservant une corde de longueur fixe, de produire un ensemble d'intervalles musicaux. Pythagore retrouve l'impression "harmonieuse" en ajustant la longueur des cordes à des rapports numériques particuliers.

L'énigme posée par Pythagore est alors la suivante : parmi l'ensemble des intervalles possibles entre deux sons, seul un petit groupe se distingue de façon

qualitative des autres. Ces intervalles, appelés "consonants" [1], correspondent à des rapports numériques simples s'ils sont identifiés par les longueurs de cordes permettant de les obtenir (2:1, 3:2, ou 4:3). Pourquoi en est-il ainsi?

Cette énigme a suscité une masse étonnante de théories, qui la placent au cœur d'un dialogue incessant entre scientifiques et musiciens (de la Motte-Haber, 1994; Fichet, 1995). De fait, cette anecdote semble être le récit de la première expérience scientifique connue du monde occidental : à partir d'une observation naturelle, un dispositif expérimental a été élaboré, qui a servi à bâtir une interprétation théorique centrée sur le concept abstrait de nombre (James, 1993). Cette théorie a aussi produit une invention, la gamme musicale dite pythagoricienne. Ce dialogue privilégié peut néanmoins être compromis par une certaine confusion sémantique autour du mot "consonance", qui est de plus fortement connoté esthétiquement. Le meilleur moyen à notre sens de clarifier ceci est d'évoquer l'évolution historique de la signification du mot consonance (Tenney, 1988).

### 4.1.2 Consonance abstraite

Le système élaboré par Pythagore pour expliquer la place particulière des intervalles consonants repose tout entier sur une mystique du nombre (James, 1993; Knobloch, 1992). La consonance musicale est expliquée dans un cadre philosophique général, dont elle est l'une des manifestations les plus éclatantes. Les intervalles pythagoriciens sont établis à partir des rapports de puissances des nombres symboliques que sont 1, 2, 3 et 4. Les rapports les plus simples qu'il est possible de former avec ces nombres sont 1:2, l'octave, 2:3, la quinte, et 3:4, la quarte, qui sont donc les consonances (Tableau 4.1). Aristote, en critiquant cette vue dans sa "Métaphysique", résumera ceci en rapportant que les pythagoriciens "supposaient que les éléments des nombres sont les éléments de toutes choses, et que l'univers tout entier est une échelle musicale et un nombre" (Meyer, 1992). Une première définition de la consonance est liée à cette ligne de pensée, généralement désignée sous le nom "d'harmonie des sphères".

**Consonance 1** *La consonance est une affinité abstraite entre notes.*

La consonance ainsi définie n'est ni mélodique, ni harmonique. L'affinité abstraite entre notes s'explique par un système généralement lié à une cosmologie et

1. Ce mot latin signifie "qui sonne avec", son équivalent grec est symphonique.

| | |
|---|---|
| Consonances | octave, quinte, quarte |
| Dissonances | autres intervalles |

TAB. 4.1 – Les consonances pythagoriciennes, VIème siècle av. J.C.

basé sur des rapports numériques (Meyer, 1992; Knobloch, 1992; James, 1993; Assayag et Cholleton, 1995). La production de ces rapports par des instruments de musique n'en donne qu'un reflet forcément imparfait, ne pouvant en aucun cas infléchir la théorie.

### 4.1.3 Consonance d'intervalles

Une deuxième acceptation du concept de consonance et de dissonance apparaît à partir du Xème siècle avec l'essor de la polyphonie en occident (Tenney, 1988). Les rencontres "verticales" des entre notes ont donné aux musiciens un nouveau paramètre à organiser, en plus des hauteurs et durées définissant les mélodies : "l'aspect" des intervalles de notes simultanées. La dichotomie entre intervalles consonants que sont l'octave, la quinte et la quarte et les autres, qui sont tous considérés dissonants, devient clairement insuffisante au regard des expérimentations des musiciens.

Plus d'une vingtaine de traités sont identifiables entre le XIème et le XIVème siècle qui proposent une nouvelle classification des intervalles et une redéfinition de la notion de consonance. Jean de Garlande ("De mensurabilis musice", env. 1250) distingue par exemple six degrés distincts (Table 4.2). Les consonances pythagoriciennes restent des consonances, mais une échelle plus fine est établie. Les intervalles de tierces deviennent (imparfaitement) consonants. Cette évolution est accompagnée par une profusion de nouvelles théories (James, 1993), où l'on retrouve entre autres les noms de Kepler, Galilée, Leibniz ou Euler. Ce classement fait l'objet d'un très fort consensus entre les traités (Tenney, 1988). Résumons donc par une deuxième définition cette nouvelle notion de consonance :

**Consonance 2** *La consonance est une échelle représentant la qualité d'un ensemble de notes simultanées hors de tout contexte.*

La qualité de consonance est ici liée à l'effet des sons sur l'oreille, impliquant une notion d'"agrément". L'idée d'"agrément" doit être comprise comme "une échelle d'agrément susceptible de consensus" (voir section 2.2.3), à distinguer

| | |
|---|---|
| Consonances parfaites | unisson, octave |
| médianes | quinte, quarte |
| imparfaites | tierce majeure et mineure |
| Dissonances imparfaites | sixte majeure et septième mineure |
| médianes | seconde majeure et sixte mineure |
| parfaites | seconde mineure, triton, septième majeure |

TAB. 4.2 – La classification des consonances selon Jean de Garlande, ("De mensurabilis musice", env. 1250).

des préférences individuelles (Parncutt et Strasburger, 1994). Par exemple, les intervalles les plus consonants ne sont pas les seuls utilisés par les musiciens, bien au contraire l'échelle de consonance est un moyen expressif.

### 4.1.4 Consonance fonctionnelle

Une autre évolution dans le traitement des consonances débute dès la fin du XIVème siècle. En effet, les traités musicaux comme ceux de Jean de Murs ou Philippe de Vitry ne reconnaissent alors plus six mais seulement trois degrés de consonance (Tableau 4.3).

Ce changement intervient parallèlement à l'établissement des règles du contrepoint, qui sont destinées à organiser la coexistence entre d'une part la conduite des mélodies dans chaque voix, et d'autre part les rencontres entre notes qui en découlent. L'évolution du statut de la quarte en est un exemple : l'une des premières consonances pour Pythagore et pour les théoriciens du Moyen Âge, elle est maintenant passée au statut de dissonance car son emploi au vu des règles du contrepoint l'apparente aux dissonances (von Helmholtz, 1877; Tenney, 1988). Ainsi, les nouveaux traités l'assimilent à une dissonance, certes particulière, opérant ainsi un glissement de sens d'une "consonance d'intervalle" vers une "dissonance fonctionnelle".

Ce glissement devient de plus en plus marqué avec l'affirmation progressive de la tonalité. Le ton, la note centrale autour de laquelle les autres s'organisent, définit une gamme particulière et une hiérarchie à l'intérieur de la gamme. Le ton établit donc un contexte puissant, qui est le plus souvent exprimé par la basse fondamentale des accords (Rameau, 1722). Un intervalle ou même une note isolée peuvent alors devenir consonant ou dissonant selon le contexte impliqué par le

| | |
|---|---|
| Consonances parfaites | unisson, octave, quinte |
| imparfaites | tierce majeure et mineure, sixte majeure et mineure |
| Dissonances | (quarte), seconde majeure et mineure, septième majeure et mineure, triton |

TAB. 4.3 – La classification des consonances selon Jean de Murs ("Ars Contrapuncti") ou Philippe de Vitry ("Ars contrapunctus"), fin du XIVème siècle. D'après Tenney (1988).

son fondamental. La dernière acceptation de la notion de consonance que nous distinguerons est donc la suivante :

**Consonance 3** *La consonance est la fonction d'une ou plusieurs notes par rapport à un contexte.*

Rameau avance une autre idée associée à la consonance fonctionnelle, celle de mouvements de tension et détente (Shirlaw, 1917; Tenney, 1988). Une dissonance dans un contexte va "réclamer" (de manière singulièrement anthropomorphique, comme le note Tenney) une résolution vers une consonance. Cette résolution de dissonance vers consonance se traduit, selon Rameau, par un mouvement de détente, et l'inverse par une tension.

### 4.1.5 Vocabulaire

Un rapide survol historique nous a permis de distinguer trois acceptations différentes de la notion de consonance : la consonance abstraite, la consonance d'intervalles et la consonance fonctionnelle. Nous nous y référerons désormais comme aux Consonances 1, 2 et 3. Le lien entre cette classification et celle de Tenney (1988) est étroit, à la différence qu'il distingue deux notions supplémentaires : ce que nous avons appelé consonance fonctionnelle est réparti sur deux périodes historiques, la première allant du XIVème siècle jusqu'à Rameau et la deuxième de Rameau à nos jours. De plus, Tenney distingue une notion de consonance spécialement adaptée à la théorie de Helmholtz. Nous pensons que ces distinctions sont entièrement fondées du point de vue musicologique mais ne sont pas nécessaires dans le cadre de notre étude. En effet, un "dosage" entre les trois larges notions que nous venons de définir nous suffira pour discuter la place et l'objet des théories perceptives de la consonance que nous allons présenter.

## 4.2 Théories psychoacoustiques

### 4.2.1 Théorie de Helmholtz basée sur la rugosité

#### 4.2.1.1 Rugosité des intervalles musicaux

La théorie de la consonance de Helmholtz (1877) est basée sur l'estimation de la rugosité d'intervalles formés par deux sons complexes (modèle présenté au Chapitre 3). Certains écarts fréquentiels se distinguent par des "creux" de rugosité, alors que d'autres démontrent une large rugosité. Les creux correspondent aux intervalles d'unisson, d'octave, de quinte, de quarte. Au contraire, les régions de forte rugosité sont obtenues pour des écarts correspondant aux intervalles de secondes, septièmes et triton (Figure 3.7).

La faible rugosité de certains intervalles est due à la coïncidence entre les partiels des sons complexes qui les composent. En effet, deux partiels qui coïncident ne peuvent provoquer ni battements ni rugosité : plus les coïncidences sont nombreuses, plus les sources potentielles de rugosité sont éliminées. Les rapports numériques employés pour désigner les intervalles, qui correspondent au rapport entre fréquences fondamentales, permettent de prévoir ces coïncidences : par exemple pour la quinte (3 : 2) le troisième partiel de la note grave coïncide avec le deuxième partiel de la note aiguë. Ceci est illustré en Figure 4.2.

Les sons utilisés pour le calcul de Helmholtz sont harmoniques et constitués de partiels dont l'amplitude décroît avec le rang. La rugosité dépendant de l'amplitude et de la fréquence moyenne du couple de partiels considérés, les coïncidences entre partiels de rangs faibles produiront une réduction de rugosité plus efficace que les coïncidences entre rangs élevés (von Helmholtz, 1877). Ceci implique qu'une rugosité minimale sera obtenue pour les rapports faisant intervenir les petits nombres entiers. L'apparition des intervalles simples dans les "creux" de rugosité est ainsi expliquée. Réciproquement, si un intervalle dont les partiels sont coïncidants est désaccordé, c'est-à-dire si l'on s'éloigne d'un rapport simple, un grand nombre de partiels d'amplitude importante vont se trouver proches et produire de la rugosité. Un intervalle de faible rugosité est donc entouré d'intervalles à forte rugosité, ce qui explique les résultats concernant les secondes, septièmes et triton.

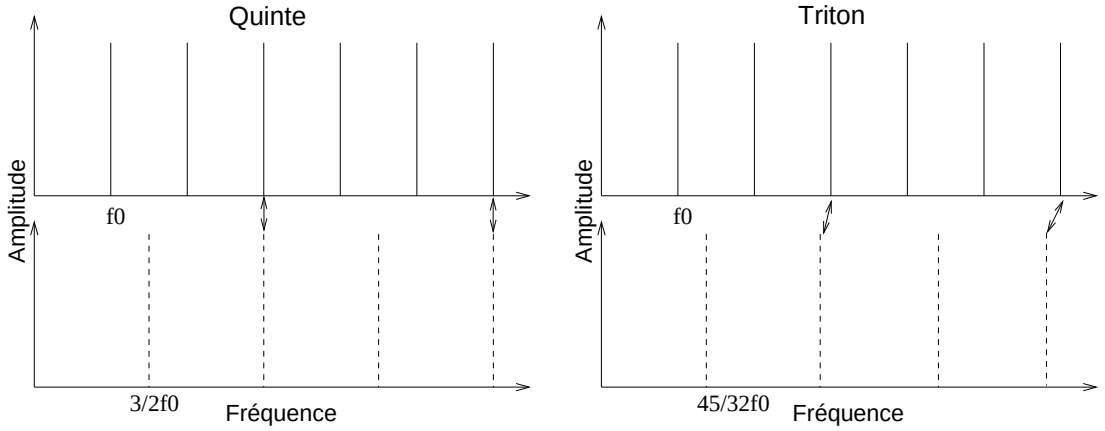


FIG. 4.2 – Intervalles entre sons harmoniques. Pour la quinte, un partiel sur deux du son aigu coïncide avec un partiel sur trois du son grave. Les partiels non coïncidants sont bien séparés de leurs voisins et ne produisent pas plus de rugosité. Si le son aigu est diminué d'un demi-ton, l'intervalle devient un triton. Les partiels qui étaient coïncidants se retrouvent adjacents et peuvent provoquer de la rugosité.

#### 4.2.1.2 Influence du registre sur la rugosité des intervalles musicaux

Il est impossible de calculer la rugosité d'un intervalle entre deux sons complexes à partir du seul rapport numérique de leurs fréquences fondamentales (qui détermine pourtant entièrement le nom de l'intervalle). La rugosité dépend de l'écart fréquentiel absolu entre partiels, or un rapport donné ne produit pas les mêmes écarts selon le registre. De plus, la rugosité produite par deux sons purs pour un écart donné varie elle aussi avec le registre (Chapitre 2).

L'influence du registre sur la rugosité des intervalles est illustrée grâce à des simulations en Figure 4.3. Globalement, la rugosité augmente si le registre est grave. Ceci est lié au phénomène de bande critique, qui est un facteur limitant pour la rugosité. La bande critique se rétrécit pour les registres graves, mais ce rétrécissement est plus lent que celui de l'écart absolu représenté par un même intervalle. Un intervalle de tierce majeure, par exemple, est plus large qu'une bande critique dans le registre aigu et donc peu rugueux alors qu'il est inférieur à une bande critique et rugueux dans le registre grave.

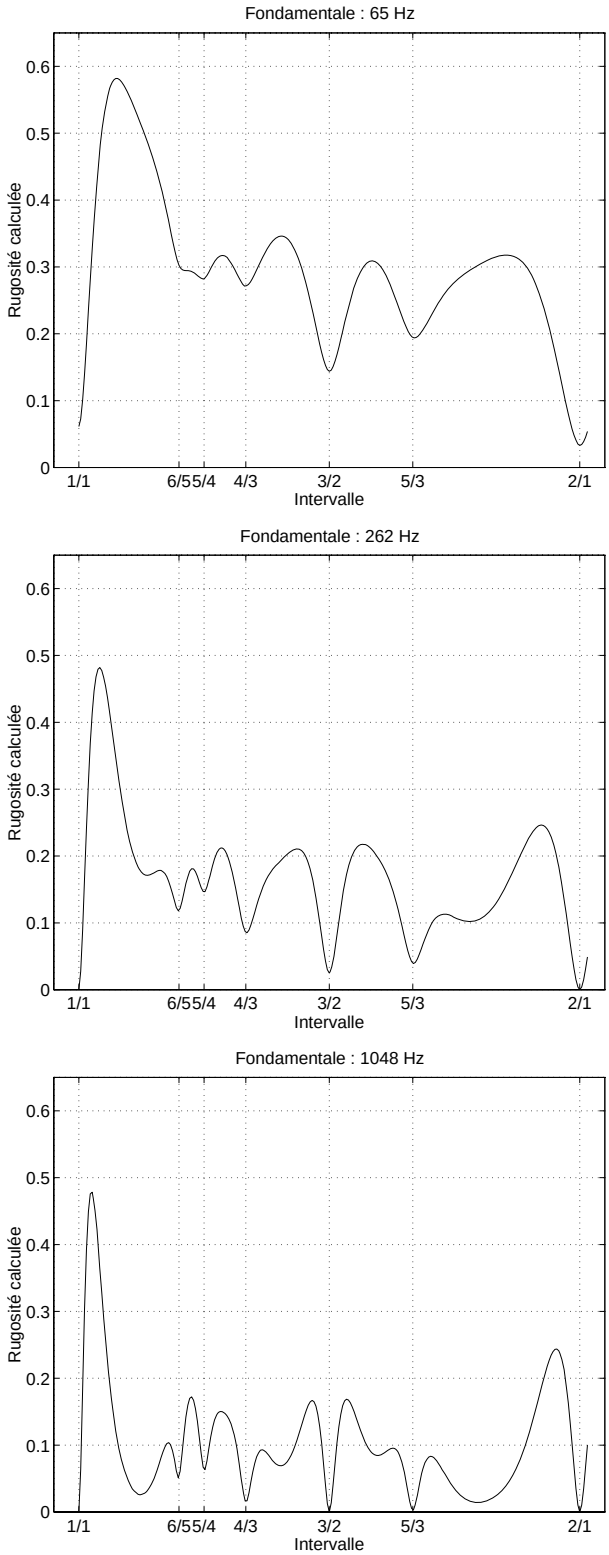


FIG. 4.3 – Rugosité des intervalles musicaux compris dans un octave pour les fondamentales du son le plus grave égales à 65 Hz, 262 Hz ou 1048 Hz. Les estimations ont été obtenues à partir du modèle de Hutchinson et Knopoff (1978) avec deux sons complexes comportant six partiels d'égale amplitude.

#### 4.2.1.3 Influence du nombre d'harmoniques sur la rugosité des intervalles musicaux

Le nombre et l'amplitude des harmoniques présents dans deux sons complexes influencent la rugosité des intervalles qu'ils peuvent former (Figure 4.4). Deux sons purs ne produiront une rugosité que pour les intervalles fréquentiels inférieurs à la bande critique dans le registre considéré. Plus l'on rajoute de partiels et plus les couples de partiels pouvant interagir sont nombreux. Les rapports simples apparaissent alors sous la forme de "creux" des estimations de rugosité du fait des coïncidences. Ceci indique qu'un même intervalle joué avec des timbres différents peut produire des rugosités différentes, pourvu que la différence de timbre se traduise dans les propriétés spectrales à long terme des sons. Notamment, plus un spectre est riche et plus certains rapports se distinguent parmi les intervalles musicaux possibles.

#### 4.2.1.4 Consonance des intervalles musicaux

Helmholtz propose que la consonance musicale est liée à l'absence de rugosité. La distinction entre consonances et dissonances serait donc due à une différence de qualité perceptive. Dans les détails, la théorie est bien plus raffinée qu'une simple identification entre rugosité et dissonance.

La rugosité est un critère continu qui dépend en toute rigueur de chaque réalisation sonore. Les classements d'intervalles selon leur consonance que nous avons présentés en section 4.1 représentent eux une qualité attachée à une classe de réalisations sonores, la superposition de sons harmoniques entretenant un certain rapport fréquentiel. De manière à relier le critère continu de rugosité et le classement des intervalles musicaux, Helmholtz fait deux hypothèses : une consonance sera considérée d'autant meilleure qu'elle est libre de rugosité sur tout le registre disponible, et bien contrastée par rapport aux intervalles voisins.

Quatre classes de consonances sont ainsi définies : les consonances absolues pour lesquelles tous les partiels sont coïncidants ; les consonances parfaites, pas ou peu rugueuses quel que soit le registre ; les consonance moyennes plus ou moins rugueuses selon le registre mais bien séparées des intervalles voisins ; les consonances imparfaites pour lesquelles les partiels coïncidants sont faibles ou absents et qui sont donc moins bien définies par rapport aux intervalles voisins. La répar-

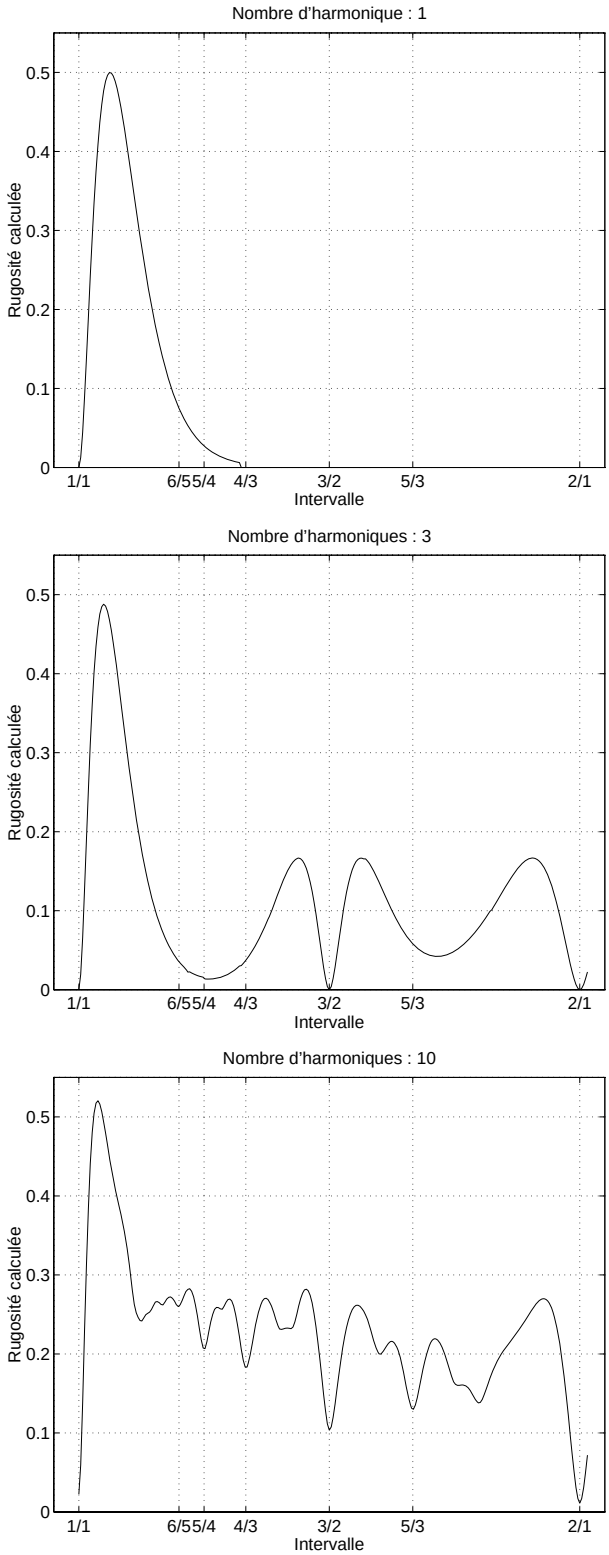


FIG. 4.4 – Rugosité des intervalles musicaux compris dans un octave pour des sons purs, des sons complexes composés de 3 partiels harmoniques d'égale amplitude, des sons complexes composés de 10 partiels harmoniques d'égale amplitude. Les estimations ont été obtenues à partir du modèle de Hutchinson et Knopoff (1978).

| | |
|---|---|
| Consonances absolues | octave |
| parfaites | quinte, quarte |
| moyennes | tierce majeure et sixte majeure |
| imparfaites | tierce mineure et sixte mineure |
| Dissonances | secondes et septièmes, triton |

TAB. 4.4 – La classification des intervalles selon leur consonance d'après Helmholtz (1877).

tition des intervalles dans ces classes est alors obtenue en examinant les résultats du modèle de rugosité (Figure 3.7). Nous les reproduisons en Tableau 4.4.

Ce classement est très proche de ceux que nous avons présentés au début de ce Chapitre dans la section consacrée à la consonance d'intervalles ou Consonance 2 (Tableau 4.2). Les extrêmes de ces classements sont notamment identiques.

A l'époque où ce classement est établi, la quarte n'est plus considérée comme une consonance parfaite (Tableau 4.3). Helmholtz en est conscient, et il propose que la quarte est une consonance parfaite en ce qui concerne le critère de rugosité mais qu'elle partage les propriétés des dissonances dans les règles de conduite des voix.

La rugosité propose une gradation fine voire continue des consonances alors que les musiciens de l'époque de Helmholtz considèrent plutôt des catégories. Cet état de fait est aussi reconnu, et justifié par des arguments similaires à ceux évoqués au sujet de la quarte. Helmholtz insiste sur le fait qu'il importe d'avoir à l'esprit l'ensemble des règles du système tonal pour comprendre la notion de consonance musicale. La rugosité n'est dans sa théorie que l'une des contraintes imposées aux utilisations musicales des intervalles. Ces contraintes peuvent être d'ordre fonctionnel et imposer un changement de statut de certains intervalles, ou une catégorisation entre consonances et dissonances.

#### 4.2.1.5 Discussion de l'approche empiriste

L'approche de Helmholtz se distingue de celles qui l'ont précédée par le parti-pris résolument empiriste de l'auteur. La musique dépend de l'auditeur qui la perçoit. Cette idée est en opposition avec l'idée d'une harmonie éternelle et immuable extérieure tant à l'auditeur qu'à la musique elle-même. Les accords ne sont par exemple plus considérés comme de simples relations numériques abstraites, mais incarnés dans une réalisation sonore. Cette théorie est donc radicalement

opposée aux théories métaphysiques de la consonance, le thème de l'harmonie des sphères. Helmholtz s'oppose aussi aux théories purement mathématiques. Le nombre ne peut être qu'une conséquence et non une cause, il doit être reflété dans un phénomène physique puis perçu pour pouvoir servir de base à une théorie de la consonance. Helmholtz critique par exemple Euler et sa théorie basée sur la reconnaissance de l'ordre exprimé par les rapports simples (Hellegouarch, 1992). Qu'en est-il lorsque ces intervalles sont légèrement modifiés, comme c'est par exemple le cas pour les intervalles du tempérament égal? Les rapports numériques deviennent alors horriblement compliqués et notre "organe de l'ordre"devrait souffrir atrocement. Pour l'approche basée sur la rugosité, ces déviations ne constituent pas un problème majeur du fait du peu de rugosité introduite par de si faibles écarts – et de plus du fait que par une "heureuse coïncidence" le tempérament est effectués sur les intervalles les moins sensibles en terme de rugosité (Vos, 1986). Les théories d'inspiration purement physiques sont plus proches de celle de Helmholtz, qui les discute abondamment, mais finalement considérées par celui-ci comme incomplètes. La seule présence de la série harmonique dans les sons naturels ne peut suffire à justifier une perception privilégiée des rapports numériques abstraits. Cette perception doit passer par la médiation de mécanismes physiologiques.

La théorie basée sur la perception de rugosité propose donc qu'une dimension élémentaire, liée à la physiologie du système auditif, permet d'expliquer une partie de la notion de consonance musicale, à savoir l'aspect perceptif des intervalles hors de tout contexte (ce qui se rapproche de ce que nous avons appelé Consonance 2). Cette échelle de rugosité, si elle est réellement imposée par la physique et la physiologie du système auditif, est universelle. Par contre, les systèmes musicaux qui peuvent s'appuyer sur une telle échelle sont eux multiples et dépendants d'un nombre considérable d'autres facteurs. Il serait faux de faire à Helmholtz le procès d'une tentative de réduction de l'harmonie à un système mécanique. Pour preuve, une partie de son ouvrage est consacrée à l'évaluation de la part des contraintes liées à la rugosité pour une variété de styles musicaux occidentaux ou non. Cette critique pourtant habituelle de la théorie est donc largement injustifiée. Laissons-le pour terminer plaider sa cause sur ce point précis :

> De là il s'ensuit, – et cette proposition n'est pas assez présente même de nos jours dans les esprits des théoriciens et historiens de la mu-

sique – *que le système des Gammes, Modes, et Tissus Harmoniques ne repose pas seulement sur des lois naturelles inaltérables mais est aussi, au moins en partie, le résultat de principes esthétiques, qui ont déjà changé et changeront encore, avec le développement progressif de l'humanité.* [pp. 235, traduit de l'édition anglaise, ses italiques.]"

### 4.2.2 Autres théories psychoacoustiques de la consonance

#### 4.2.2.1 Consonance sensorielle

La théorie de Helmholtz a abouti à la création d'une nouvelle notion, la "consonance sensorielle" [2] (Plomp et Levelt, 1965). La consonance sensorielle est proche du concept musical de Consonance 2, mais l'explication qui en est donnée repose explicitement sur des phénomènes perceptifs de bas niveau comme la rugosité. Toutefois, consonance sensorielle et rugosité ne sont pas identifiables.

Des jugements de dyades de sons purs sur des échelles sémantiques ont indiqué que la dissonance sensorielle était liée à la rugosité pour un registre moyen à élevé, et à la perception de fluctuations pour le registre grave (Rakowski, 1992). Un corpus de stimuli différent a permis à Aures de déterminer un indice dit "d'euphonie" qui intègre en plus de la rugosité des estimations de sonie, d'acuité et de tonalité (Aures, 1985a). La sonie correspond à l'intensité subjective comme définie par Zwicker et Scharf (1965), l'acuité à une estimation de centre de gravité spectrale définie par von Bismarck (1974) et la tonalité distingue les sons avec une hauteur des bruits. Un son possède un degré d'euphonie d'autant plus grand qu'il est peu rugueux, peu fort, avec une acuité faible et une forte tonalité.

Les paramètres trouvés par Rakowski ou les ingrédients du degré d'euphonie de Aures démontrent une dépendance sur le corpus de sons utilisés. Ces études indiquent que la notion de consonance sensorielle ne peut être identifiée à celle de rugosité. La rugosité est un paramètre y participant dans une large part, aux côté d'autres paramètres psychoacoustiques pouvant varier selon les contextes.

2. L'expression originale de Plomp et Levelt est *tonal consonance*, consonance liée au son. Du fait des confusions qu'elle peut entraîner, cette expression a été remplacée par "consonance sensorielle" (Aures, 1985a; Parncutt et Strasburger, 1994).

#### 4.2.2.2 Fusion perceptive

Une autre théorie de la consonance (Consonance 2) est celle de la fusion perceptive (Stumpf, 1898). La consonance perçue est ici rapportée à un nouvel indice, le "degré de fusion". Selon Stumpf, ce qui fait qu'un intervalle est consonant est qu'il est perçu comme un tout. Le degré de fusion peut être mesuré expérimentalement par la difficulté qu'ont les auditeurs à déterminer le nombre de notes d'un accord.

Une expérience mesurant le degré de fusion des intervalles tempérés a été réalisée (DeWitt et Crowder, 1987). La tâche s'est avérée trop simple, et sauf pour les intervalles d'octave les sujets étaient parfaitement capables de déterminer le nombre réel de notes présentes, sans même démontrer de différences notables entre les temps de réaction. Il semble que la notion de fusion proposée par Stumpf soit une conséquence à la fois de l'absence de rugosité et de principes plus généraux de l'organisation des scènes auditive (Bregman, 1990). Elle ne peut néanmoins expliquer la consonance des intervalles musicaux traditionnels, aucun d'entre eux à part l'octave ne remplissant réellement les conditions de "fusion".

#### 4.2.2.3 Hauteur virtuelle

Un modèle de consonance musicale comportant deux paramètres psychoacoustiques distincts a été formulé par Terhardt (Terhardt, 1970, 1974b, 1984). Le premier paramètre est celui de rugosité, le second est celui de "hauteur virtuelle".

Le paramètre de hauteur virtuelle est défini à partir d'un modèle de la perception de hauteur des sons complexes (Terhardt, 1974b). Selon ce modèle, deux modes de perception de hauteur sont disponibles : le mode analytique et le mode synthétique. Le mode analytique tente d'extraire les hauteurs spectrales présentes dans un son complexe, qui ne correspondent pas exactement aux fréquences acoustiques du fait de mécanismes périphériques. A partir de ces hauteurs spectrales, le mode synthétique génère un ensemble de hauteurs probables par des procédés de plus haut niveau. Ces procédés sont dépendants de notre apprentissage, notamment de notre habitude de percevoir des sons harmoniques comme la voix humaine. Le mode synthétique revient en quelque sorte à estimer si les partiels composant le son complexe ne pourraient pas être expliqués par une série harmonique provenant d'une fondamentale unique, présente ou absente. La qualité de l'ajustement est une estimation de la saillance de la basse fondamentale perçue,

et plus cette saillance est élevée meilleure est la consonance. Terhardt propose donc que la théorie de Rameau (1722) est relayée par le phénomène perceptif de la fondamentale absente.

Le modèle de consonance proposé par Terhardt tente d'établir deux corrélats psychoacoustiques, dont la perception est par ailleurs avérée par des expériences antérieures, à deux facettes de la consonance. Ces deux principes sont relativement indépendants, ainsi peuvent-ils se renforcer mutuellement ou au contraire être mis en conflit. Une valeur "d'agrément" est déterminée par la rugosité (Consonance 2), l'implication d'une basse fondamentale est estimée par la saillance de hauteur virtuelle (plutôt Consonance 3).

#### 4.2.2.4 Communalité et proximité de hauteurs

Le modèle de Terhardt (1974b) a été implémenté et enrichi par Parncutt (Parncutt, 1989; Parncutt et Strasburger, 1994). Cette implémentation permet le calcul d'indices de rugosité (Hutchinson et Knopoff, 1978), de tonalité (Aures, 1985a), et de saillance de la basse fondamentale (Terhardt, 1974b). Néanmoins, en plus de ces indices psychoacoustiques qui caractérisent individuellement chaque accord, Parncutt propose des indices qui relient deux accords successifs.

Le modèle travaille sur une suite de listes de fréquences et d'amplitudes qui représentent des évènements sonores successifs. Ces listes sont d'abord transformées en hauteurs spectrales en simulant certains traitements périphériques (Terhardt, 1974b). Les hauteurs virtuelles sont ensuite évaluées en testant tous les gabarits harmoniques possibles par rapport aux hauteurs spectrales (les fréquences, dans le modèle, sont approximées au demi-ton). Pour chaque sonorité, un vecteur qui associe aux 120 demi-tons compris entre 16 Hz et 16 kHz une valeur de saillance de hauteur virtuelle est ainsi obtenu.

Les paramètres qui relient deux évènements entre eux se basent sur la comparaison des vecteurs de hauteurs virtuelles. Cette comparaison prend deux formes : une corrélation entre les deux vecteurs pour la "communalité de hauteur", une estimation de la probabilité de percevoir les différences entre les deux vecteurs pour la "proximité de hauteur". La communalité de hauteur mesure la similarité entre hauteurs virtuelles, la proximité de hauteur mesure les aspects plutôt mélodiques de mouvements entre les voix. La complémentarité des notions peut être illustrée simplement : la résolution par demi-tons des notes dissonantes d'un

accord vers un accord consonant est un exemple de faible communalité de hauteur mais de grande proximité de hauteurs. A l'inverse, un accord suivi par son inversion représente une forte communalité mais une faible proximité (Parncutt et Strasburger, 1994). Le paramètre de communalité a été testé expérimentalement à propos de jugements de relation *(relatedness)* et les résultats sont apparus satisfaisants (Thompson et Parncutt, 1997).

L'ensemble du modèle de Parncutt permet finalement de prendre en compte plusieurs aspects de la consonance musicale tout en restant lié à des phénomènes de bas niveau. La Consonance 2 est abordée avec le calcul de rugosité, la Consonance 3 avec la basse fondamentale et le lien entre accords successifs. L'apprentissage éventuel de l'auditeur n'intervient que de façon minimale, simplement pour justifier la perception des hauteurs virtuelles.

## 4.3 Théories cognitives

### 4.3.1 Hiérarchies tonales

#### 4.3.1.1 Existence de hiérarchies entre notes

L'approche cognitive se distingue de l'approche psychoacoustique par le fait qu'elle considère des éléments extérieurs au message acoustique pour proposer des théories de la consonance musicale. La problématique de la consonance est généralement abordée par le biais de structures de connaissances abstraites dont disposeraient les auditeurs pour comprendre et apprécier la musique (Krumhansl, 1983; Lerdahl et Jackendoff, 1983). La réalité de ces structures de connaissances peut être soumise à l'expérience, ce qui justifie que nous les présentions dans le cadre des théories perceptives.

Dans le cadre de la musique occidentale, une tonalité définit une série de notes, la gamme diatonique, choisie parmi les douze notes disponibles dans un octave, la gamme chromatique. En plus de favoriser les notes de la gamme diatonique par rapport aux autres, la théorie musicale établit aussi une hiérarchie entre elles : la tonique sera la plus importante, suivie de la dominante (la quinte) et ainsi de suite. Selon la tonalité et donc le contexte, les évènements sonores ne se verront pas attribuer la même importance

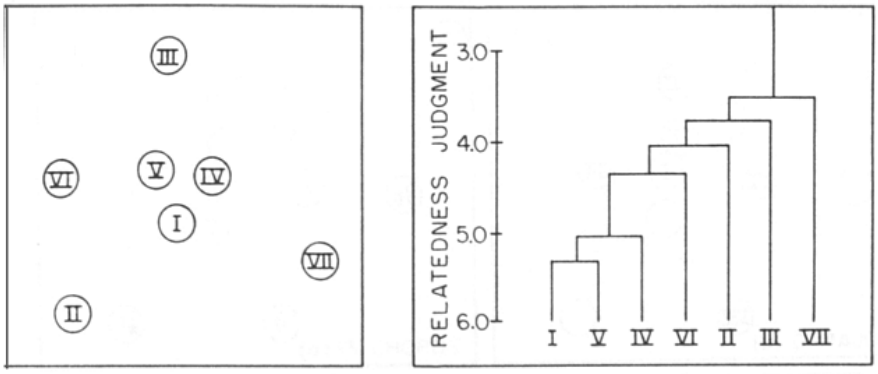


FIG. 4.5 – Hiérarchies tonales perçues pour des accords de trois notes. A droite, représentation des résultats de l'analyse mutidimensionelle. A gauche, même données analysées par clusters hiérarchiques. D'après Krumhansl (1983).

La perception de structures hiérarchiques entre les notes de la gamme chromatique a été étudiée par Krumhansl (1979, 1982,). Une méthode dite de "note sonde" a été développée : les sujets doivent juger de l'adéquation *(goodness of fit)* d'une note sonde après avoir été confrontés à un contexte. Par exemple, ils entendent une gamme de Do majeur puis une note choisie aléatoirement dans la gamme chromatique et doivent juger de l'adéquation de cette dernière note.

L'existence d'une représentation mentale des relations hiérarchiques définies par la théorie musicale entre les notes composant une gamme majeure ou mineure a été démontrée (Krumhansl et Shepard, 1979; Krumhansl et Kessler, 1982). Pour des notes isolées suivant une gamme majeure, les jugements d'adéquation les plus élevés ont été obtenus pour les notes diatoniques par rapport aux notes chromatiques, et parmi les notes diatonique une prédominance de la tonique, de la dominante, de la tierce majeure et de la quarte a été observée.

#### 4.3.1.2 Existence de hiérarchies entre accords

De la même manière que la théorie musicale attribue une importance privilégiée à la note constituant la tonique d'une gamme, l'accord formé par l'empilement de tierces successives à partir de cette tonique possède aussi une importance particulière. En Do majeur, cet accord dit parfait majeur et noté I est constitué des trois notes Do, Mi et Sol.

Des jugements de similarités entre les 7 triades qu'il est possible de former en choisissant pour fondamentales les degrés successifs d'une gamme majeure ont démontré un regroupement des degrés I, IV et V par rapport aux autres, ce qui correspond aux usages de la théorie harmonique tonale (Krumhansl, Bharucha et Castellano, 1982). La nature des deux dimensions est toutefois incertaine : le regroupement des accords structurellement importants a été considéré comme une indication qu'ils ont pu être jugés par rapport à leur fonction dans un contexte diatonique, mais rien n'indique qu'un critère acoustique n'est pas en fait la cause de cet arrangement : les accords majeurs sont au centre, les autres à l'extérieur. En imposant un contexte avant chaque jugement de similarité, il a été montré que l'espace perceptif obtenu est modifié (Bharucha et Krumhansl, 1983). Dans ce cas, la présence d'un contexte a bien modifié la perception d'accords tout en laissant leur consonance sensorielle inchangée. D'autres analyses sur le même principe ont indiqué que les relations entre tonalités, en plus de celles à l'intérieur d'une tonalité, peuvent aussi influencer la perception des accords (Krumhansl et Kessler, 1982).

#### 4.3.1.3 Interprétations

Des hiérarchies tonales entre évènements sonores, notes ou accords, sont donc perçues. Leur base n'est pas entièrement sensorielle : leur perception est totalement absente quelques mois après la naissance (Trainor et Trehub, 1994), elle est influencées par l'apprentissage mais est établie dès six ans (Krumhansl et Keil, 1982).

La première hypothèse est que cette connaissance résulte d'une exposition à la musique occidentale tonale et d'une assimilation implicite de ses conventions . Les relations hiérarchiques entre évènements décrites par la théorie musicale sont en effet étroitement corrélées avec leur fréquence d'emploi (Krumhansl, 1990). Une image pour désigner cette hypothèse est, humoristiquement, celle du "lavage de cerveau" (Mathews et Pierce, 1980). Il a ainsi été possible d'entrainer des réseaux neuronaux pour reproduire les attentes d'auditeurs liées aux hiérarchies tonales (Bharucha et Olney, 1989). La théorie du lavage de cerveau exprime sans doute une part de bon sens reconnue qui considère que chaque auditeur possède une culture musicale façonnée par son environnement.

Une autre interprétation est inspirée des théories de la linguistique issues des travaux de Chomsky (Chomsky, 1965). Pour Chomsky, l'apprentissage du langage ne peut pas se faire simplement par un apprentissage associatif, certaines règles et mécanismes d'ordre cognitif doivent être présents pour permettre cet apprentissage. Krumhansl propose que de telles règles, au niveau de la perception de la musique, ont pu imposer des contraintes à la structure que les musiciens ont donné à la tonalité. La perception des hiérarchies tonales, en plus d'être influencée par l'apprentissage, pourrait être facilité par ces règles cognitives abstraites (Krumhansl, 1983, 1990).

#### 4.3.1.4 Dissonance cognitive

L'existence possible de règles cognitives pertinentes pour l'écoute musicale rajoute un aspect à la perception des mouvements de tension et détente, qui ne seraient plus uniquement liés à la consonance. En effet, une tension ou "dissonance cognitive" peut être provoquée si de telles règles sont enfreintes (Mandler, 1984). Meyer (1956) avait ainsi proposé l'application de règles de "bonne formation", développées sur le modèle de celle des psychologues de la Gestalt dans le domaine de la vision (Koffka, 1935), à la perception de tension et détente musicales. Un événement qui s'écarterait de l'attente générée par les règles de bonne formation (poursuite du contour mélodique, changement de tonalité brusque, etc...) provoquerait une tension. Cette démarche a été complétée par Narmour (1990) qui propose un modèle explicite de prédiction de tension et détente mélodique basé sur un ensemble de règles de bonne formation.

L'existence de hiérarchies tonales et l'interprétation basée sur des règles cognitives montre que la perception de la consonance d'un accord dans un certain contexte dépend non seulement de ses caractéristiques perceptives mais aussi de structures psychologiques et cognitives abstraites. L'influence de l'apprentissage sur la perception des hiérarchies indique de plus que la culture musicale de l'auditeur peut interagir avec ces règles.

### 4.3.2 Théorie Générative de la Musique Tonale (Lerdahl et Jackendoff, 1983)

#### 4.3.2.1 Principe

Les hiérarchies tonales établissent l'existence de facteurs modifiant la perception d'un événement sonore donné en fonction de son contexte. Le contexte, dans les résultats expérimentaux rapportés, consistait en une gamme ou une courte séquence d'accords. La situation d'écoute musicale est bien plus complexe, car un contexte non ambigu ne peut bien souvent pas être défini à chaque instant. Les possibilités offertes par cette ambiguïté sont même l'une des richesses du système tonal.

L'objectif de la *Théorie Générative de la Musique Tonale* est de décrire les intuitions d'un auditeur "idéal" lorsqu'il écoute une pièce de musique tonale occidentale (Lerdahl et Jackendoff, 1983). Ces intuitions consistent principalement, selon les auteurs, en la perception d'une succession de mouvements de tension et de détente. La perception de ces mouvements découle de deux étapes principales (la version originale de la théorie ne séparait pas aussi clairement les deux étapes, mais nous considérons la version révisée par Lerdahl, 1989).

La première étape est celle de réduction de la trame temporelle. Le message musical tel que noté sur la partition est segmenté en unités cohérentes par un ensemble de règles mélodico-rythmiques. Une phrase mélodique peut par exemple constituer un des segments de la trame. Cette trame est ensuite réduite. Les règles qui permettent la réduction sont appelées règles de stabilité : les évènements les moins stables comme par exemple les notes de passage sont progressivement éliminés. Les règles de stabilité reposent dans une large mesure sur une formalisation des hiérarchies tonales évoquées plus haut, les éléments les moins importants hiérarchiquement étant considérés les moins stables. Différents niveaux de réduction sont extraits jusqu'à ne parvenir finalement qu'à un seul élément générateur.

Les éléments dégagés à chaque niveau de réduction sont ensuite mis en relation les uns avec les autres grâce à la deuxième étape dite de réduction prolongationelle. Entre deux éléments existe un lien, une prolongation, et ces liens peuvent être répartis en deux grands groupes, les tensions ou les détentes. Pour évaluer leur nature, les relations de stabilité de l'étape précédente sont utilisées : un élément stable suivi par un élément moins stable à un certain niveau de réduction

produit une prolongation de type tension, l'inverse une détente. Le résultats de cette analyse sont habituellement représentés sous la forme d'un graphe arborescent reliant les évènements entre eux aux divers niveaux de réduction. La forme des branches de l'arbre exprime les intuitions de l'auditeur prévues par l'analyse. Cette représentation souligne le fait que pour la théorie, les mouvements de tension et détente sont ressentis à diverses échelles temporelles : entre éléments adjacents dans la partition bien sûr mais aussi entre le début et la fin de la pièce par exemple.

#### 4.3.2.2 Réduction prolongationelle et consonance

Le détail de la théorie est naturellement plus complexe, notamment du fait de l'existence de règles transformationelles entre les niveaux de réduction, l'interaction possibles entre les règles aux différents niveaux, la possibilité d'analyse récursive. Ce bref aperçu permet néanmoins de souligner l'apport de cette théorie quand à la problématique de la consonance. La consonance de chaque événement est ici déterminée, à plusieurs niveaux, selon un contexte global établi par le déroulement temporel de l'œuvre musicale. Néanmoins, la notion de consonance sensorielle reste pertinente au niveau de la formulation des règles de stabilité (Lerdahl et Jackendoff, 1983; Lerdahl, 1989).

## 4.4 Évaluations dans le cadre de la musique tonale

### 4.4.1 Perception de consonance d'intervalles

#### 4.4.1.1 Classements d'experts

Les théories perceptives de la consonance que nous venons de présenter peuvent et doivent même par nature être soumises à l'expérimentation. L'un des premiers aspects ayant fait l'objet de tests est le classement des intervalles de deux notes compris dans un octave selon leur consonance.

La méthode d'obtention de tels classements est importante par rapport aux interprétations à leur donner. Malmberg (1918) a utilisé un groupe "d'experts", des musiciens ou auditeurs expérimentés. Des classements de fusion, de douceur,

de pureté, de mélange *(fusion, smoothness, purity, blending)* ont été demandés. Un classement n'a été retenu qu'après qu'il fasse l'unanimité entre les 12 experts. Tout les adjectifs sauf celui de fusion ont permis d'aboutir à un même classement, interprété comme une échelle de consonance et identique à celui présenté en Tableau 4.4.

La concordance entre le classement de Malmberg et les prévisions de la théorie de la consonance basée sur la rugosité est un élément en sa faveur cité par Plomp et Levelt (1965). Toutefois, des jugements d'experts ne permettent pas de préciser dans quelle mesure la corrélation observée n'est pas accidentelle et si le classement de Malmberg n'est pas en fait le reflet d'un apprentissage musical.

#### 4.4.1.2 Influence de l'apprentissage musical et comparaisons interculturelles

Une étude s'est directement penchée sur l'influence de la familiarité culturelle avec la musique occidentale sur les jugements de consonance d'intervalles de deux notes (Butler et Daston, 1963).

Des comparaisons par paires de consonance entre intervalles ont d'abord été demandés à 53 étudiants américains. Il est apparu que les jugements des auditeurs, non sélectionnés pour leurs capacités musicales, n'étaient absolument pas faits au hasard. L'échelle obtenue était identique à celle des auditeurs experts de Malmberg. Il peut être argumenté ici que bien que non experts, les étudiants étaient familiers avec la musique occidentale.

L'expérience a été poursuivie avec 308 auditeurs japonais. Différents sous-groupes ont été établis, sur la base des préférence musicales (musique occidentale ou traditionnelle) et sur la capacité absolue à discriminer les intervalles occidentaux. Aucune différence n'a été observée pour les jugements de consonance entre les sous-groupes, et l'échelle obtenue ressemblait encore une fois fortement à celle des experts de Malmberg.

Ces études convergent vers la conclusion que la consonance d'intervalles hors de tout contexte a une réalité perceptive, en dehors des considérations d'apprentissage implicite ou explicite. Cette consonance serait donc une consonance sensorielle. L'échelle obtenue peut de plus être expliquée par la rugosité de ces intervalles. Néanmoins, du fait de la petite variété des stimuli testés (toujours les

même 12 intervalles !) l'influence potentielle d'autres facteurs sensoriels ne peut être exclue.

### 4.4.2 Perception de mouvements de tension et détente

#### 4.4.2.1 Mouvements de tension et détente

Le lien possible entre la consonance d'intervalles hors de tout contexte et l'écoute musicale est la notion de mouvements de tension et détente. Nous avons déjà rencontré ces mouvements à plusieurs reprises. Ils ont été relié à la notion de consonance fonctionnelle ou Consonance 3 par les théories de Rameau. La *Théorie Générative de la Musique Tonale* se fixe pour but de révéler un tel réseau sous-jacent de tension et de détente qui serait présent à travers toute œuvre tonale. En fait, l'identification de schémas de tension et détente est le point de départ de la plupart des analyses harmoniques classique de la musique occidentale (Schenker, 1935). Une étude récente a tenté d'évaluer la perception de mouvements de tension et détente dans de courtes séquences d'accord, et la part prise dans cette perception par différentes théories de la consonance que nous avons présentées (Bigand, Parncutt et Lerdahl, 1996).

#### 4.4.2.2 Présentation de l'étude de Bigand et al., 1996

Le but de l'étude était de déterminer le degré de tension d'un accord inclus dans une courte séquence. Pour ce faire, trois accords successifs de quatre notes chacun étaient présentés. La séquence des trois accords était toujours de la forme Do Majeur - X -Do Majeur. La tâche des sujets était de noter sur une échelle numérique la tension de l'accord X. Cinquante variétés d'accords ont été testées : X pouvait être diatonique ou non à Do Majeur, contenir une septième ou non, être en position renversée ou non. Les sujets étaient séparés en deux groupes selon leur éducation musicale.

Les prédictions de différentes approches ont été établies pour l'ensemble des accords testés. Le camp "cognitif" était représenté par les règles proposées par Lerdahl dans le cadre des raffinements successifs de la Théorie Générative de la Musique Tonale (Lerdahl, 1988). Ces règles peuvent être considérées comme la formalisation des résultats portant sur les hiérarchies tonales (Krumhansl, 1983). Pour le camp "psychoacoustique", le modèle de Parncutt comportant une es-

timation de rugosité (selon Hutchinson et Knopoff, 1978) et une estimation de communalité de hauteur ont été convoqués (Parncutt, 1989). Une nouvelle mesure de mouvement horizontal, basée sur une simplification de l'indice de proximité de hauteur, a été considérée.

Les résultats expérimentaux ont démontré – une influence combinée des différents facteurs. Les effets de hiérarchie tonale ont été significatifs pour tous les auditeurs, mais de façon plus marquée pour les musiciens. Un effet de l'apprentissage explicite est donc démontré. Le modèle de rugosité a permis de prédire que les accords comportant une tierce mineure et ceux comportant une septième ont été jugés globalement plus tendus. Des effets dus aux mouvements mélodiques entre les voix ont aussi été démontrés, avec une importance moindre pour les musiciens.

Une analyse par régression multiple entre les prédictions des divers paramètres et les résultats montre que globalement, une combinaison des indices de hiérarchie tonale, de communalité de hauteur, de mouvement horizontal et de rugosité permet d'expliquer les résultats. Il est à noter que le modèle de rugosité ne tient pas une place importante dans la régression statistique. Ce modèle comporte trop d'imperfections pour distinguer finement entre les 50 accords testés. Les auteurs conservent néanmoins le critère de rugosité dans leur modèle de tension car il peut seul prédire qualitativement les résultats liés aux tierces mineures et aux septièmes.

Il apparaît donc qu'un grand nombre de facteurs interviennent dans la perception des mouvements de tension et détente de courtes séquences d'accords apparentées à la musique tonale : des facteurs psychoacoustiques déterminant la consonance sensorielle de l'événement isolé et sa relation avec ceux qui l'entourent mais aussi des règles cognitives portant sur la connaissance des hiérarchies tonales, acquises au moins en partie par un apprentissage culturel.

### 4.4.3 La notion composite de consonance

La notion de consonance est, comme l'on aurait pu s'y attendre, une notion complexe comportant de multiples facettes. D'une part, comme nous l'avons souligné en section 4.1, elle comporte diverses acceptations historiques. Toutes les théories dites de la consonance n'ont donc pas forcément le même objet. D'autre part, l'influence convergente de différents facteurs observée expérimenta-

lement indique que ces théories sont chacune susceptibles d'expliquer une partie du concept, de façon plus ou moins importante selon le matériau musical. Ces théories ne sont donc pas exclusives, elles coexistent pour tenter de capturer la complexité de la notion de consonance dans la musique tonale.

Pour clarifier ceci, nous proposons de résumer les liens entre les théories perceptives de la consonance par une figure hautement schématique. Si l'on distingue d'une part les notions de Consonance 2 et de Consonance 3 et d'autre part les approches psychoacoustiques et cognitives, nous pouvons placer les différents modèles et théories évoqués au cours de ce Chapitre[3] approximativement comme indiqué en Figure 4.6.

Les résultats des études expérimentales destinées à évaluer l'importance de ces facteurs indiquent que tous peuvent être pertinents. Ils interagissent sans doute de façon complexe pendant l'expérience d'écoute de l'auditeur. Certains peuvent être considérés comme étant à la base d'autres (la hauteur virtuelle et la communalité de hauteur, la rugosité et l'établissement de conventions arbitraires...). Nous avons de fait épargné au lecteur les multiples flèches qui pourraient se superposer à une telle figure, qu'il se sente libre d'en imaginer autant de combinaisons que nécessaires !

## 4.5 Extensions

### 4.5.1 Connaissances implicites et indices sensoriels

Certaines des composantes de la consonance musicale avérées dans l'écoute de la musique tonale reposent sur un apprentissage des auditeurs. Même les auditeurs dits non musiciens ont internalisé inconsciemment la plupart des grands principes des hiérarchies tonales (Bigand, 1989; Bigand, Parncutt et Lerdahl, 1996). Dans le cas où ces auditeurs sont confrontés à une musique dont les principes ne respectent pas les règles du système tonal, ces connaissances implicites seront inopérantes. L'expression de variations de consonance ou de mouvements de tension et détente dans un tel contexte, si elle est recherchée, doit emprunter d'autres moyens. L'un d'eux consiste à s'appuyer sur les indices définis par les théories psychoacoustiques de la consonance.

3. La notion de Consonance 1 ou consonance abstraite n'est pas vraiment concernée par la discussion des théories perceptives, sauf accident...

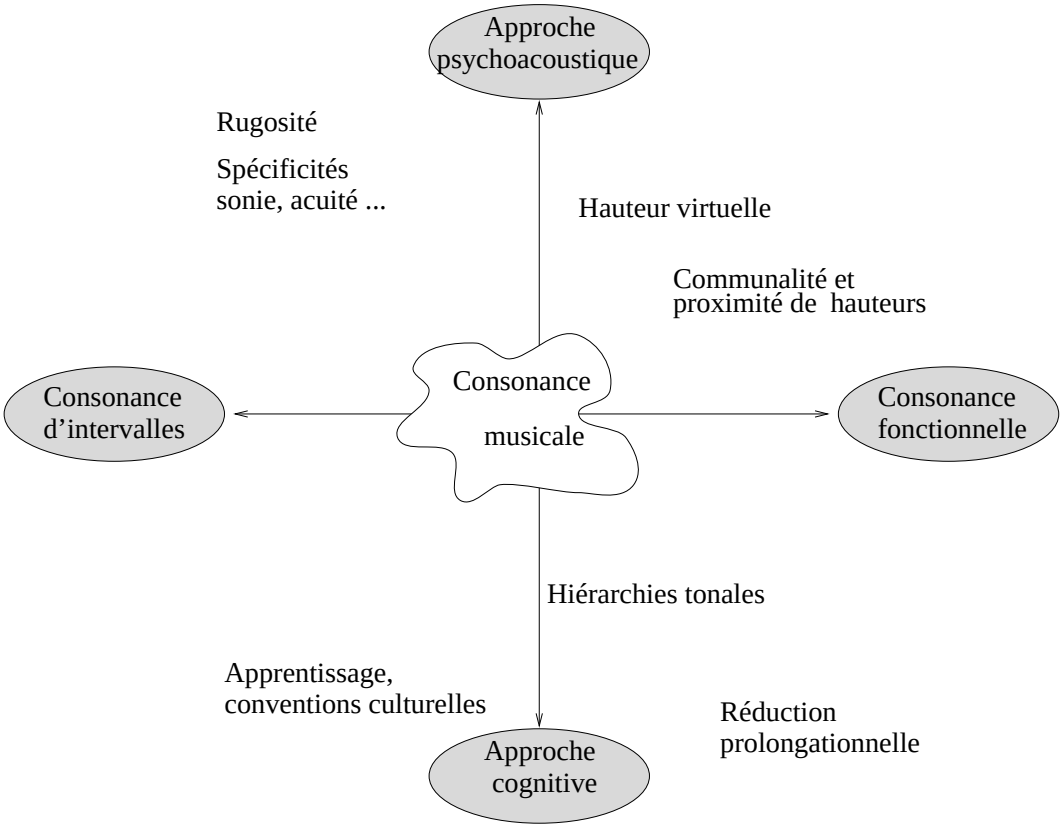


FIG. 4.6 – Représentation schématique des relations entre les théories perceptives de la consonance et la notion complexe de "consonance musicale".

La rugosité est l'un de ces indices indépendant du style ou des matériaux musicaux employés. Elle se prête donc à l'exploration des possibilités d'extension de la notion de consonance à des musiques non conventionnelles, notamment en ce qui concerne l'emploi du nouveau vocabulaire offert par la synthèse sonore (Risset, 1994). Il était possible d'entrevoir ce développement au sein même de l'ouvrage de Helmholtz. Critiquant la vue de Rameau comme quoi la série harmonique fournit une base "naturelle" à l'harmonie musicale, Helmholtz note : "En outre, en frappant des verges, des cloches, des membranes, en soufflant dans des cavités, Rameau aurait pu obtenir un grand nombre d'accords dissonants, tout autres que ceux qu'ils trouvait avec les cordes et les autres instruments de musique. Il aurait fallu, pourtant, les considérer comme naturels." Reconnaissons tout de même que l'auteur ne poussera pas cette réflexion bien loin du fait de son goût manifeste pour la musique harmonique tonale !

## 4.5.2 Gammes et tempéraments

### 4.5.2.1 Tempérament d'intervalles

La rugosité des intervalles musicaux varie avec leur tempérament, car les coïncidences exactes entre partiels n'apparaissent que pour certains rapports bien précis. Vos (1986) a montré que pour des quintes ou des tierces l'accord jugé le plus pur est l'intervalle juste. Le tempérament dégrade plus la pureté pour la quinte que pour la tierce. Ces résultats sont interprétables en termes de rugosité : l'intervalle juste est celui qui produit le moins de rugosité, et le tempérament de la quinte entraîne plus de rugosité que celui de la tierce. Des données supplémentaires indiquent aussi une légère contribution de l'écart fréquentiel absolu (Vos, 1986).

Mathews, Pierce et Roberts (1987) ont approfondi ces résultats. L'effet du tempérament sur des accords majeurs (soit trois notes dans le rapport 4:5:6) et mineurs (rapport 10:12:15) a été testée. Ces accords sont largement utilisés dans la musique tonale et bien connus des auditeurs. Un grand nombre de partiels d'amplitudes non négligeables coïncident pour l'accord majeur, un moins grand nombre pour l'accord mineur. Deux autres types d'accords ont aussi été testés, avec des rapports 3:5:7 et 5:7:9. Ces accords ne sont pas familiers pour les auditeurs, et pour cause : les notes qui les composent ne font pas partie de

la gamme chromatique. Néanmoins, comme l'accord majeur, ils comportent un grand nombre de partiels coïncidants. Les résultats des jugements de "justesse" *(in-tune)* distinguent des sujets appelés "purs", qui préfèrent l'absence de tempérament, et ceux appelés "riches" qui préfèrent les accords un peu désaccordés de part et d'autre de l'intervalle juste. Le pattern de réponse riche est plus marqué pour les accords majeurs et non-traditionnels et moins marqué pour les accords mineurs.

Les effets du type d'accord sur le pattern "pur" sont entièrement cohérents avec les résultats de Vos (1986). Le pattern "riche" réclame une nouvelle interprétation. Il ne peut pas être expliqué par la reconnaissance du tempérament égal, puisqu'il est aussi observé pour les accords non-traditionnels. La rugosité permet d'interpréter l'ensemble des données. En effet, une différence de rugosité existe toujours entre accords justes et tempérés, qu'ils soient traditionnels ou non. Les résultats indiqueraient donc que certains auditeurs préfèrent les intervalles sans rugosité, "purs", d'autres avec un peu de rugosité, "riches". La perception de rugosité semble être l'unique critère de jugement, mais de façon intéressante l'interprétation qui en est faite dépend du goût personnel (Mathews, Pierce et Roberts, 1987).

#### 4.5.2.2 Liens entre gammes et spectres des sons

Si le spectre des sons n'est plus harmonique, les relations numériques provoquant la coïncidence de partiels sont bouleversées. Une première démonstration de ceci est la gamme à 8 tons également espacés à l'intérieur de l'octave de Pierce (1966, 1984). Associés à cette gamme, par synthèse numérique, Pierce génère des sons dont les partiels correspondent à un degré sur deux de la gamme. Ce système induit de singulières lois concernant la rugosité des intervalles : en partant d'une note quelconque comme fondamentale, l'intervalle formé avec le degré suivant est très rugueux, l'intervalle avec le degré d'après non rugueux, et ainsi de suite (Figure 4.7). Cette gamme a été utilisée dans une étude appelée "Canon à huit sons".

Cette exploration a été généralisée par Sethares (1993). L'hypothèse est formulée que pour chaque type de spectre, il existe une ou plusieurs gammes adaptées qui correspondent aux minima locaux de la fonction de rugosité. A partir de cette hypothèse, la gamme juste pour les sons harmoniques mais aussi des

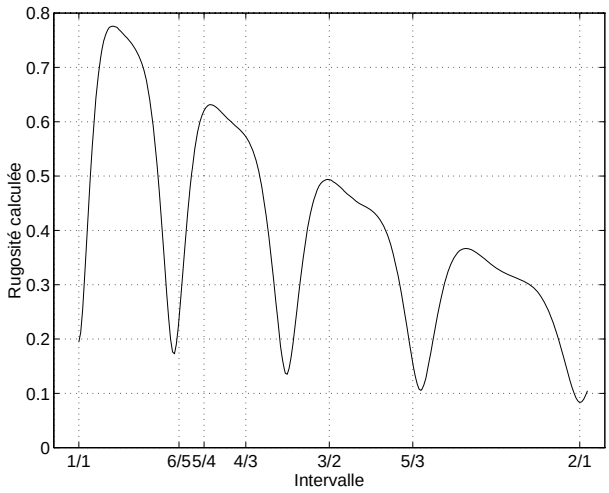


FIG. 4.7 – Rugosité d'intervalles estimée par le modèle de Hutchinson et Knopoff pour les sons complexes correspondants à la gamme à huit tons de Pierce. Les échelles de la gamme sont situés alternativement sur les creux et les bosses.

gammes adaptées aux sons avec un spectre harmonique étiré ou inharmoniques sont dérivées (Sethares, 1993, 1994, 1997).

### 4.5.3 Sons inharmoniques

#### 4.5.3.1 Consonance d'intervalles de sons inharmoniques

Le lien entre rugosité et gammes est un aspect des extensions des théories de la consonance. Une autre question est la possibilité ou non de retrouver des caractéristiques de l'harmonie avec des sons non harmoniques.

Slaymaker (1970) a examiné les caractéristiques de sons obtenus par étirement ou compression de la série harmonique, pour lesquels "l'octave" entre le deuxième partiel et le premier partiel n'est plus exactement dans un rapport de 2. L'étude est qualitative, les jugements perceptifs étant ceux "des gens se trouvant là" au bon moment ! Slaymaker note que si les sons inharmoniques sont arrangés en respectant les coïncidences entre partiels, il est possible de produire une impression de consonance même pour avec des sons fortement étirés. La consonance est néanmoins "étrange", cette étude informelle constituant selon les mots de l'auteur "un coup d'œil vers un nouveau monde musical".

Ces résultats qualitatifs ont été complétés par Geary (1980). Les sons qu'il emploie sont astucieusement construits. Il s'agit de sons contractés dont le rap-

port entre le premier et le deuxième partiel est de $\sqrt{2}$. Ce rapport non rationnel fait qu'aucune fondamentale absente ne peut être mathématiquement ajustée aux sons. De plus, grâce à ce spectre particulier, les rapports d'intervalles de la musique tonale correspondant à la quarte, la quinte et la sixte provoquent la présence de partiels adjacents et donc de rugosité. A l'inverse, les intervalles (approximatifs) de quinte augmentée et de triton provoquent la coïncidence des partiels. Les indices de consonance liés au rapport entre fréquences fondamentales et ceux liés à la rugosités sont donc inversés par rapport aux intervalles habituels composés de sons harmoniques. Les jugements par paires, établis sur 14 sujets à l'éducation musicale contrastée montrent que la consonance est significativement jugée plus importante pour les intervalles provoquant peu de rugosité. L'éducation musicale a peu d'influence. Ainsi, dans ces conditions expérimentales de présentation d'intervalles isolés, la rugosité semble déterminer les relations de consonance entre sons inharmoniques même si cette consonance n'a pas lieu pour les rapports fréquentiels connus culturellement des auditeurs.

#### 4.5.3.2 Tentatives d'harmonies "inharmoniques"

Ce résultat est encourageant, mais la Figure 4.6 nous permet de deviner que, puisque la consonance d'intervalles isolés n'est pas le seul déterminant de la consonance musicale, les problèmes que vont devoir aborder les compositeurs désireux d'utiliser des sons inharmoniques ne sont pas épuisés (Pierce, 1984) ! Une exploration de quelques complications supplémentaires a été entreprise par Mathews et Pierce (1980). Dans cet article important, trois critères de perception de l'harmonie sont mis en opposition : la rugosité, la perception d'une basse fondamentale, et la reconnaissance d'intervalles due au "lavage de cerveau". Plusieurs expériences ont été réalisées pour explorer ces trois facettes de l'harmonie tonale.

Dans la première expérience, la reconnaissance de tonalité a été testée avec des sons harmoniques et inharmoniques. Trois séquences M, T et X sont présentées aux auditeurs sous la forme MXTX. M est une suite d'accords établissant clairement, du point de vue théorique, une tonalité (une cadence par exemple). T est la même séquence que M transposée dans une autre tonalité, différente soit par le choix de la tonique soit en opposant les modes majeur et mineur. X est soit dans la tonalité de M, soit de T. La tâche des auditeurs est de reconnaître la tonalité de X. Deux versions de l'expérience sont réalisées : dans l'une, les sons

sont harmoniques et les intervalles sont ceux de la musique tonale. Dans l'autre, le spectre des sons est étiré (rapport 2,4 au lieu de 2) et les intervalles étirés de manière à préserver les coïncidences entre partiels pour les rapports d'"octave" ou de "quinte" étirés, etc. Dans ces conditions, les critères de rugosité ne changent pas entre les versions harmoniques et étirées, en revanche la perception de basse fondamentale est présente pour la version harmonique et absente pour la version étirée. Les résultats indiquent que les sujets, dans tous les cas, parviennent à exécuter la tâche au-dessus du seuil de hasard. Ainsi, l'absence de basse fondamentale n'a pas empêché les jugements. Le critère de rugosité semble avoir pu donner des indices suffisants pour la reconnaissance de tonalité.

Une deuxième expérience a consisté à évaluer le degré d'achèvement d'une succession de deux accords. Les sujets ont comparé des cadences à des anticadences (V I ou II V). Dans ce cas, comme prévu par l'harmonie tonale, la cadence a donné une impression de finalité bien supérieure à celle de l'anticadence quand les accords étaient composés de sons harmoniques. Pour la version étirée, les résultats sont différents : aucune des séquences n'a donné de sentiment de finalité. Il semble que l'absence de basse fondamentale ait ici empêché la perception de finalité. Une vérification a consisté à modifier les spectres des sons harmoniques, en supprimant certains partiels, pour que les deux accords aient exactement la même rugosité. Dans ce cas, malgré l'absence de l'indice de rugosité, le sentiment de finalité de la cadence de sons harmoniques persiste pour tous les auditeurs et de façon plus marquée pour les musiciens. Ceci indique que les effets combinés de la basse fondamentale et de l'apprentissage culturel peuvent être suffisants pour reconnaître la finalité de cadences.

Un troisième point a été abordé grâce à la gamme à huit sons de Pierce. Cette gamme permet une forte différence de rugosité entre ses accords dissonants et consonants employés dans les "cadences", plus importante que pour la cadence harmonique. Dans ce cas, les "cadences" (très rugueux vers non rugueux) ont été jugé significativement plus finales que les anticadences (peu rugueux vers très rugueux). Les critères de basse fondamentale ou d'acculturation sont ici inopérants, le critère de rugosité peut donc aussi suffire à transmettre une impression de finalité.

### 4.5.4 Discussion

Les résultats de Mathews et Pierce sont intéressants à plus d'un titre : ils posent clairement la complexité du problème de la consonance musicale lorsque l'on décide de s'écarter du système tonal. Dans ce système, les influences de divers critères perceptifs et d'apprentissage à long terme se combinent, elles ont été codifiées et assimilées par les musiciens au long de siècles de spéculation et de tâtonnements pour définir un moyen d'expression d'une grande richesse. Pourtant, dès que les sons deviennent inharmoniques, l'expérience accumulée devient inopérante et la compréhension des actions combinées de ces influences est loin d'être achevée. Cette compréhension serait néanmoins appréciable, par exemple pour tirer parti de la "ductilité" nouvelle (Risset, 1994) du matériau sonore offerte par la synthèse.

Une piste semble néanmoins intéressante : la rugosité, facteur psychoacoustique largement étudié par ailleurs et indépendant de l'acculturation des auditeurs, semble déterminer dans une large part la consonance d'intervalles hors de tout contexte. La dernière expérience de Mathews et Pierce montre qu'elle peut même permettre dans certaines conditions d'établir un sentiment de finalité entre deux accords. Ce résultat est d'importance : le sentiment de finalité associé à la cadence joue un rôle central dans l'expression des mouvements de tension et détente, considérés comme l'un des éléments essentiels de l'écoute musicale (Lerdahl et Jackendoff, 1983). La question de la possibilité d'expression de mouvement de tension et détente hors du contexte tonal explorée par Mathews et Pierce mérite donc d'être poursuivie. Plusieurs questions sont soulevées par leurs résultats.

Tout d'abord, les sons utilisés ont été générés par un programme de synthèse additive qui permet de contrôler rigoureusement fréquences, phases et amplitudes des composantes spectrales des sons. Ceci a pour effet de favoriser les effets liés à la rugosité : en effet, la rugosité diminue de façon dramatique si les composantes spectrales qui interagissent ne sont pas de fréquences fixes et d'amplitudes stables et équivalentes (von Békésy, 1960). Lors de la production sonore à partir d'instruments acoustiques, les fréquences et amplitudes possèdent de riches fluctuations intrinsèques. La maîtrise grandissante des techniques de synthèse sonore autorise aussi l'introduction de telles fluctuations dans les sons synthétiques. L'extension éventuelle des résultats à un contexte employant des sons plus riches réclame donc de nouvelles expérimentations. Parmi les questions encore sans réponse, il est pos-

sible de se demander si des accords faits de sons instrumentaux parviendront à produire des différences de rugosité suffisantes pour exprimer des mouvements de tension et détente sans l'aide des autres indices disponibles habituellement.

Une deuxième interrogation soulevée par Mathews et Pierce eux-mêmes est liée à la problématique de l'organisation des scènes auditives. En effet, en plus de satisfaire de façon convergente aux critères de rugosité, de basse fondamentale et d'acculturation, les sons harmoniques facilitent l'intégration de leurs différents partiels en un percept unique (Bregman, 1990). Pour certains sons inharmoniques, il n'est pas rare de voir quelques partiels se détacher du son global et modifier ainsi radicalement la perception des qualités émergeantes (McAdams, 1984). Les sons inharmoniques ont alors peut-être été organisés en différentes images auditives, ce qui a pu affecter la perception de rugosité (Mathews, Pierce et Roberts, 1987). De tels effets devraient être encore plus marqués pour des sons instrumentaux, qui ont tendance à produire chacun une image individuelle, et que l'on voudrait regrouper en accords homogènes. Si l'on désire donc étudier les jugements de rugosité ou de tension d'accords faits de sons naturels, une influence des indices de groupement perceptif est à contrôler.

## 4.6 Conclusion

Ce chapitre nous a permis d'examiner diverses théories perceptives qui ont été élaborées à propos de la consonance musicale. Certaines adoptent une approche psychoacoustique, en déduisant directement du signal acoustique et de son traitement perceptif des indices de consonance. L'approche cognitive fait intervenir des hiérarchies connues des auditeurs, dont l'apprentissage est éventuellement facilité par des mécanismes cognitifs. Leurs influences combinées au sein du système tonal a été soulignée. Les théories psychoacoustiques ne sont par nature pas limitées à un système musical particulier. Parmi elles, la rugosité a été considérée pour l'extension de la notion d'harmonie hors du système tonal. Deux expériences poursuivant cette idée sont proposées au Chapitre 8.

# Partie Expérimentale

# Chapitre 5

# Influence de la phase dans la perception de rugosité

*Les approches spectrales et temporelles sont opposées grâce à des manipulations de phase. Une influence supplémentaire de la forme de l'enveloppe après filtrage auditif est proposée.*

Deux approches ont été adoptées pour étudier la perception de rugosité, l'une basée sur les aspects spectraux du signal acoustique et l'autre sur ses aspects temporels (Chapitre 2). En approfondissant une méthode proposée par plusieurs auteurs (Mathes et Miller, 1947; Terhardt, 1974a; Smoorenburg, 1970), nous allons opposer ces approches. Des données expérimentales sont recueillies en manipulant la phase de la composante centrale de signaux à trois composantes. Les résultats indiquent que la phase a une influence importante sur la perception de rugosité. Cette influence peut en partie être expliquée par une modification de l'enveloppe des signaux acoustiques, conformément à l'hypothèse temporelle. Toutefois, d'autres aspects des données amènent à formuler une nouvelle hypothèse basée sur l'influence possible de la forme de l'enveloppe après transformation par les mécanismes auditifs périphériques.

# 5.1 Introduction

## 5.1.1 Rappels sur les approches spectrales et temporelles

La définition originale de Helmholtz (1877) concernant la rugosité possède deux facettes, comme nous l'avons remarqué en section 2.1.2. Si la rugosité est l'attribut perceptif causé par des battements entre composantes fréquentielles proches, doit on considérer les battements temporels ou la présence de composantes fréquentielles comme la base du percept ? Cette distinction prend une importance toute particulière à la lumière des récentes études développementales (Schellenberg et Trainor, 1996) et comparatives (Hulse, Bernard et Braaten, 1995; Fay, 1996) qui indiquent que la perception de rugosité dérive vraisemblablement de mécanismes élémentaires. L'opposition entre approches spectrale et temporelle, superflue au niveau purement acoustique, implique en fait une interprétation du phénomène basée sur des mécanismes auditifs de natures bien distinctes.

Les études que nous avons regroupées sous le nom d'approche spectrale ont étudié l'influence du spectre d'amplitude des sons sur leur rugosité (Cross et Goodwin, 1893; Mayer, 1894; Kaestner, 1909; Plomp et Levelt, 1965; Kameoka et Kuriyagawa, 1969a). La présence de composantes à l'intérieur d'une bande critique a été considérée comme la cause du percept (section 2.3). A la suite de ces résultats, des modèles de calcul de rugosité ont été proposé se basant sur la composition spectrale des sons à évaluer (section 3.2.2).

L'autre approche distinguée a étudié l'influence de paramètres temporels sur la rugosité (Terhardt, 1970, 1974a; Vogel, 1974; Fastl, 1977; Aures, 1985b). L'interprétation proposée est alors que la rugosité dépend des fluctuations d'enveloppe à l'intérieur de chaque bande critique (section 2.4). Ces résultats ont inspiré un autre type de modèles, qui estiment la rugosité en extrayant une "profondeur de modulation effective" (section 3.2.3).

Les paramètres spectraux et temporels des stimuli utilisés par les deux approches sont fortement covariants : deux composantes fréquentielles donnant lieu à des battements d'amplitude, ou une modulation d'amplitude introduisant des composantes fréquentielles (Figures 2.3 et 2.7 ; section 2.5). Une méthode proposée par différents auteurs pour opposer ces indices consiste à manipuler la phase de sons complexes en laissant leur spectre d'amplitude constant. Des com-

paraisons entres sons SAM et QFM (Mathes et Miller, 1947) ou sons SAM et FM (Terhardt, 1974a) ont apporté des éléments qui penchent en faveur d'une interprétation temporelle, l'enveloppe semblant être le facteur déterminant pour la rugosité plutôt que la composition spectrale (section 2.5.2). Nous avons déjà mentionné les insuffisances quantitatives de ces données qui ne permettent pas de déterminer entièrement l'effet possible de la phase (section 2.5.2). De fait, ces insuffisances ont été soulignées au moins en une occasion bien précise.

### 5.1.2 Influence possible de la phase

*Smoorenburg:* I think it is dangerous to take the modulation depth as the only criterion for roughness because that implies that phase as such does not play a part. I can show you some stimuli with the same envelope, but with rather different roughness.

(...)

Do you agree that such a change [in phase itself] may result in a difference in roughness?

(...)

*Zwicker:* Is the envelope the same? How can you then change the phase?

*Terhardt:* I don't have any results referring to such a situation.

Proceedings of the International Symposium on Hearing (pp. 287-290)
Driebergen, 23-27 June 1969.

A la suite d'une communication de Terhardt (1970) où l'enveloppe temporelle était proposée comme base à la perception de rugosité, une discussion visiblement animée a pris place entre plusieurs "grands noms" de la psychoacoustique. Durant cette discussion, dont un court extrait est reproduit ci-dessus, Smoorenburg a contesté l'interprétation des données de Terhardt. Il affirme notamment que deux sons ayant même spectre d'amplitude et même enveloppe peuvent produire des rugosités différentes. Ainsi, selon lui, "il est dangereux de ne considérer que la profondeur de modulation [d'amplitude] comme critère de rugosité". La discussion ne précise pas quels critères Smoorenburg souhaitait ajouter, mais du fait des travaux de l'auteur il semble qu'il faisait référence aux sons de combinaison introduits par les non-linéarités du système auditif (Smoorenburg, 1972a,b).

Cette explication avait déjà été avancée par Goldstein (1967a) et a été récemment reprise par Greenwood (1991b) pour expliquer une partie des différences perceptives entre sons SAM et QFM. Une ou plusieurs composantes spectrales supplémentaires étant introduites par les produits de distorsion, une interprétation alternative des résultats concernant l'influence de la phase sur la rugosité devient possible.

Aucun résultat expérimental n'a été publié par la suite, à notre connaissance, pour confirmer ou infirmer les affirmations de Smoorenburg. L'expérience que nous allons maintenant présenter étudie l'influence de la phase sur la perception de rugosité de stimuli qui incluent les sons SAM et QFM et qui permettent de tester directement la remarque de Smoorenburg.

## 5.2 Expérience 1

### 5.2.1 Hypothèses expérimentales

Considérons trois composantes spectrales de fréquences $(f_c - f_m), (f_c)$ et $(f_c + f_m)$, avec des amplitudes relatives de $1/2, 1$ et $1/2$. Pour étudier l'influence des relations de phase entre ces trois composantes, nous pouvons varier la phase à l'origine de la composante centrale, notée $\phi$ (équation 5.1) :

$$pAM(t) = \frac{1}{2}\cos[2\pi(f_c - f_m)t] + \cos(2\pi f_c t + \phi) + \frac{1}{2}\cos[2\pi(f_c + f_m)t] \qquad (5.1)$$

Pour $\phi = 0$, un son pur à la fréquence centrale $f_c$ modulé sinusoïdalement en amplitude à $f_m$, avec une profondeur de modulation de 100%, est obtenu. En faisant systématiquement varier $\phi$, nous obtenons une famille de sons ayant tous un même spectre d'amplitude mais des structures temporelles différentes. Nous appellerons ces sons "pseudo-Amplitude Modulated" (pAM). Si seules les relations entre composantes fréquentielles déterminent la rugosité, tous les sons pAM de spectres d'amplitude identiques devraient être jugés comme également rugueux.

Lorsque $\phi$ varie, la forme de l'enveloppe d'amplitude du son pAM varie en général (Figure 5.1, l'expression analytique de l'enveloppe sera dérivée en section 5.4.2.1). Quand $\phi$ passe de 0 à $\pi/2$, l'enveloppe "s'aplatit" alors que le son passe

d'un son SAM à un son QFM. L'un des moyens de quantifier cette variation est de calculer la valeur $rms$ de l'enveloppe. Pour l'hypothèse temporelle, cette grandeur est supposée déterminer la rugosité si le son est contenu dans une même bande critique et si une seule fréquence de modulation est impliquée. En faisant varier $\phi$ systématiquement, nous pouvons explorer de façon quantitative le lien entre la valeur $rms$ de l'enveloppe et la rugosité pour un spectre d'amplitude donné.

Il est aussi possible d'obtenir un même spectre d'amplitude et une même enveloppe temporelle mais une structure temporelle fine différente en considérant les sons pAM correspondant à $\phi = \pm\varphi$, avec $\varphi \in ]0, \pi/2]$. Si seule la valeur $rms$ de l'enveloppe contribue à la rugosité, les sons pAM correspondant aux phases de signe opposé $\phi = \pm\varphi$ devraient être jugés identiquement.

## 5.2.2 Méthode

### 5.2.2.1 Sujets

Deux groupes de 15 auditeurs ont été testés en deux sessions comprenant des fréquences centrales ($f_c$) différentes. Les groupes étaient constitués de 17 hommes et 13 femmes âgés de 24 à 32 ans ($\overline{M} = 27$ ans). Les auditeurs ont déclaré n'avoir aucun problème d'audition. Ils n'étaient pas sélectionnés sur la base d'aptitudes musicales ou d'expérience antérieure avec des tests psychoacoustiques. La plupart d'entre eux ont été recrutés à partir d'une base de donnée de volontaires et n'avaient jamais participé à une expérience psychoacoustique.

### 5.2.2.2 Stimuli

Sept séries de sons pAM ont été utilisées pour lesquelles une fréquence centrale et une fréquence de "modulation" ont été associées : $(f_c, f_m) = (125$ Hz ; 30 Hz), (250 ; 40), (500 ; 50), (1000 ; 70), (2000 ; 70), (4000 ; 70) et (8000 ; 70). Dans chaque cas, $f_m$ a été choisie de façon à correspondre à la fréquence de modulation produisant une rugosité maximum dans le cas d'un son pur de fréquence $f_c$ sinusoïdalement modulé en amplitude à la fréquence $f_m$ (Zwicker et Fastl, 1990). La phase $\phi$ de la composante centrale a été variée de $-\pi/2$ à $+\pi/2$ par pas de $\pi/6$. Pour chaque série de sons pAM, sept stimuli différents ont donc été produits ($\phi = -\pi/2; -\pi/3; -\pi/6; 0; \pi/6; \pi/3; \pi/2$). Avec ces stimuli, quatre conditions distinctes de valeur $rms$ de l'enveloppe correspondant aux quatre valeurs abso-

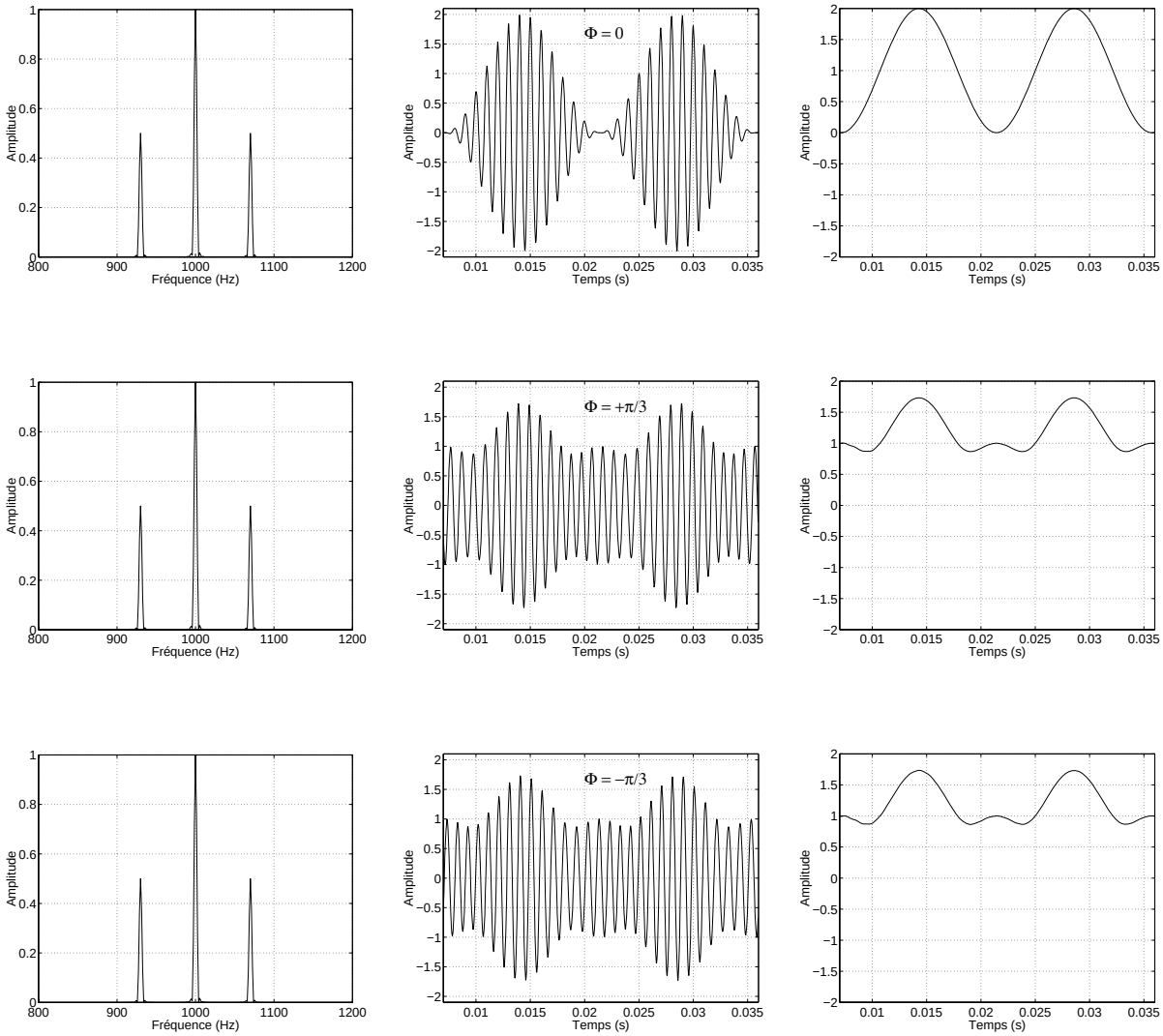


FIG. 5.1 – Spectre d'amplitude, forme d'onde temporelle et enveloppe temporelle des sons pAM pour les conditions $\phi = 0, \phi = \pi/3, \phi = -\pi/3$. Le spectre d'amplitude reste constant, l'enveloppe varie si $|\phi|$ est modifié, les conditions $\pm|\phi|$ ont la même enveloppe temporelle mais des structures temporelles fines différentes (même si visuellement les deux ondes semblent très similaires).

lues distinctes de $\phi$ sont obtenues. Chaque son avait une durée de 1 seconde, et avait un fondu au silence (demi-période de cosinus) de 50 ms à l'attaque et à l'extinction.

#### 5.2.2.3 Matériel

Les stimuli ont été générés numériquement par synthèse additive avec une résolution de 16 bits et un taux d'échantillonage de 44.1 kHz. Les signaux étaient reproduits sur une station d'informatique musicale NeXT à travers une carte de traitement de signal numérique ISPW (Lindemann, Dechelle, Smith et Starkier, 1991), des convertisseurs analogique/numérique Pro IO et un amplificateur Canford. Ils étaient présentés à travers un casque Sennheiser HD 520 II. Le niveau de présentation, tel que mesuré par un sonomètre Bruel & Kjaer 2209, était fixé à 60 dBA. Ce niveau est un niveau standard pour la plupart des expériences sur la perception de rugosité. Les produits de distorsion de l'ensemble de la chaîne de reproduction sonore, tels que mesurés par un analyseur de spectre Bruel & Kjaer, étaient de niveau inférieur à 60 dB par rapport au signal. Les sujets étaient assis dans une cabine audiométrique à double paroi Soluna S1. La conduite de l'expérience ainsi que la collecte des données par l'intermédiaire d'une interface graphique étaient assurées par un programme écrit dans l'environnement PsiExp sur la station NeXT (Smith, 1995).

#### 5.2.2.4 Procédure

Les auditeurs étaient informés qu'ils allaient devoir juger un attribut sonore différent de la hauteur ou de l'intensité, appelé "rugosité". La notion de rugosité était alors introduite à l'aide d'exemples sonores. Ces exemples étaient des sons SAM aux fréquences utilisées dans l'expérience. Les auditeurs pouvaient à l'aide d'une interface faire varier leur profondeur de modulation de façon continue, modifiant ainsi de façon concomitante le spectre d'amplitude et l'enveloppe du son. De tels exemples permettent de passer progressivement d'un son de rugosité nulle à des sons de forte rugosité en laissant d'autres dimensions perceptives comme l'intensité ou la hauteur aussi constantes que possible. Une fois que les auditeurs déclaraient avoir compris ce qu'était la dimension "rugosité" et après 1 à 3 essais de familiarisation sans renforcement, l'expérience pouvait débuter. Les stimuli étaient alors présentés en blocs séparés (en ordre aléatoire) correspondant

chacun à une fréquence centrale. Dans chaque bloc, les 42 paires possibles de sons non-identiques dans un ordre comme dans l'autre étaient présentées en ordre aléatoire. Pour chaque essai expérimental, les auditeurs devaient décider parmi les sons de la paire lequel était le plus rugueux (2 Alternative Forced-Choice). Avant de répondre, les sujets pouvaient, s'ils le désiraient, réécouter la paire autant que nécessaire. Chaque bloc durait approximativement 10 minutes. Les auditeurs avaient 5 blocs par session expérimentale.

#### 5.2.2.5 Analyses statistiques

La méthode employée pour traiter les données expérimentales, identique pour toutes les expériences que nous présenterons, est décrite en détail dans l'Annexe A. L'analyse consiste à employer la méthode de Bradley-Terry-Luce (BTL) pour transformer en une échelle linéaire les jugements de comparaisons binaires (David, 1988). Plus un son a été choisi comme "plus rugueux" à travers ses comparaisons avec les autres sons, plus la valeur obtenue par la méthode BTL est élevée. L'échelle représente donc le contraste perçu entre les sons pAM d'un bloc concernant la dimension rugosité. Chaque point de l'échelle est déterminé par 6 comparaisons avec les autres sons dans 2 ordres et pour 15 sujets, soient 180 jugements expérimentaux.

Pour tester si les différences de rugosité observées sur l'échelle sont significatives, elles doivent être comparées aux écarts-type. Néanmoins, l'estimation de l'écart-type des données BTL ne peut être faite de manière classique. La technique de bootstrap a donc été choisie, une distribution empirique de valeurs BTL étant générée par rééchantillonage avec remise à partir des données originales pour estimer l'écart-type (Efron et Tibshirani, 1993). Notons que les intervalles obtenus de cette manière ne sont pas forcément centrés sur les valeurs BTL. La significativité à un certain risque est testée à partir des intervalles de confiance dérivés des écarts-type. Nous considérerons par la suite qu'une différence entre deux valeurs BTL est significative si les deux intervalles définis par $1,65\sigma$ autour de chaque valeur ne se recouvrent pas, ce qui correspond à un risque de première espèce de $p < 0.05$ (Efron et Tibshirani, 1993).

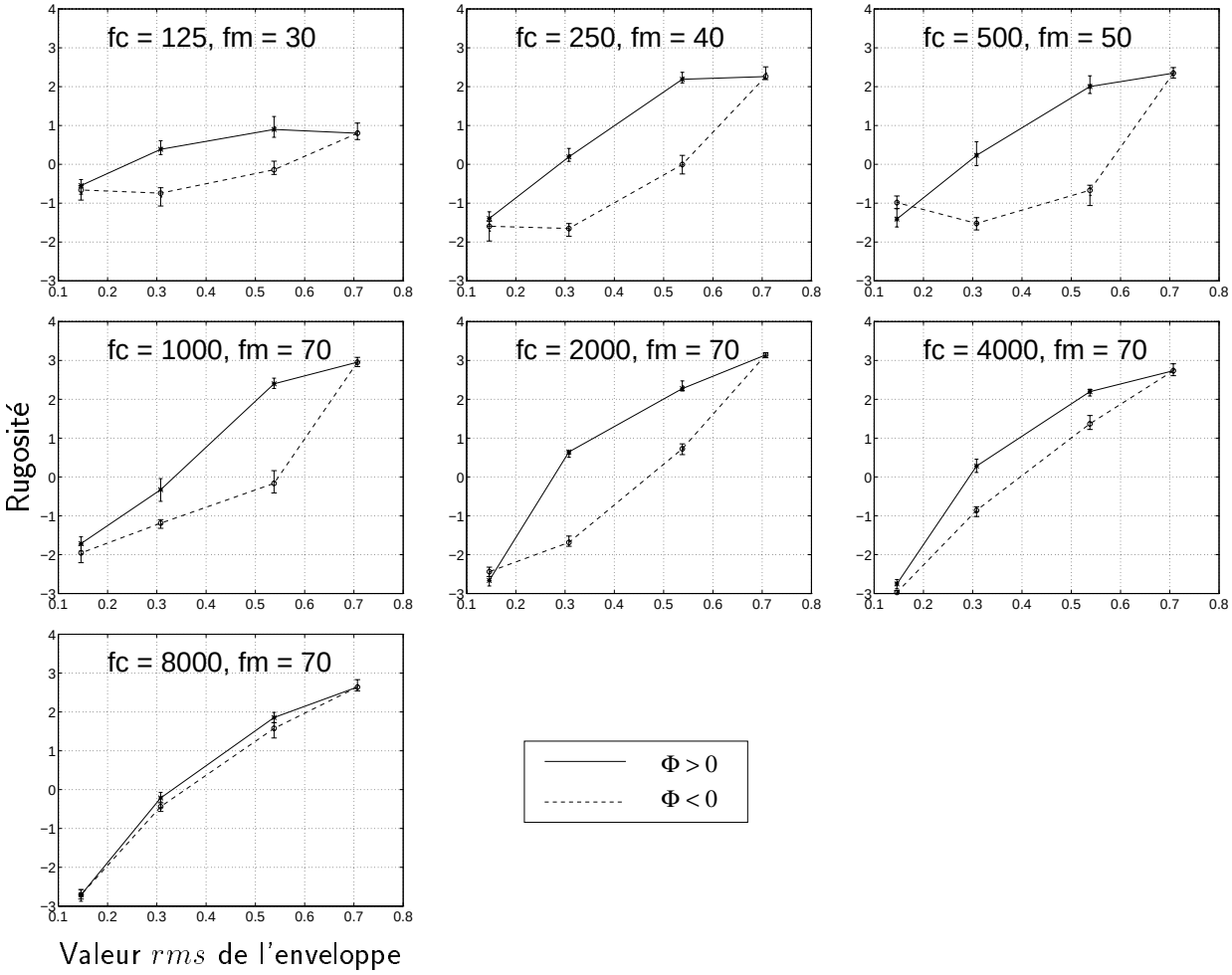


FIG. 5.2 – Résultats pour les sept séries de stimuli correspondant aux différentes fréquences centrales. Les valeurs de rugosité estimées par la méthode BTL sont données en fonction de la valeur $rms$ de l'enveloppe d'amplitude (la plus grande valeur $rms$ correspond à $\phi = 0$, la plus faible à $\phi = \pm\pi/2$). Les traits pleins représentent les conditions $\phi > 0$ et les pointillés $\phi < 0$.

### 5.2.3 Résultats

Les résultats pour les sept séries de stimuli sont présentés Figure 5.2. Pour faciliter les interprétations, la rugosité perçue est indiquée en fonction de la valeur $rms$ de l'enveloppe. Cette valeur est le reflet direct de la variable indépendante $\phi$ : la valeur $rms$ la plus faible est obtenue avec $|\phi| = \pi/2$, puis croît lorsque $|\phi|$ décroît.

Le facteur le plus évident contribuant à la perception de rugosité est la valeur $rms$ de l'enveloppe, une plus grande valeur $rms$ provoquant plus de rugosité. Cet effet est similaire en amplitude à travers la plupart des conditions, sauf pour la condition $f_c = 125$ Hz où le contraste global est notablement moins marqué. Néanmoins, pour les $f_c$ inférieures ou égales à 4 kHz un fort effet du signe de la phase est aussi observé pour des sons ayant un même spectre d'amplitude et une même enveloppe. Cet effet est cohérent pour les différences de phase intermédiaires ($|\phi| = \pi/6, \pi/3$) : les stimuli avec $\phi > 0$ sont jugés significativement plus rugueux que ceux avec $\phi < 0$. Cet effet disparaît pour toutes les fréquences centrales à $\phi = \pm\pi/2$, et pour toutes les phases à $f_c = 8$ kHz. Un autre trait notable des résultats est la faible taille des écarts-type, indiquant un fort agrément entre sujets.

## 5.3 Discussion

### 5.3.1 Influence de la valeur $rms$ de l'enveloppe sur la rugosité

#### 5.3.1.1 Effet de bande critique

La rugosité a été influencée par la valeur $rms$ de l'enveloppe pour un spectre d'amplitude donné : la rugosité augmente globalement avec la valeur $rms$ à l'intérieur de chaque condition $(f_c, f_m)$. Cette augmentation est toutefois moins marquée pour $f_c = 125$ Hz. Nous interprétons ceci comme un effet de bande critique. Pour $f_c = 2$ kHz, les stimuli pAM couvrent 140 Hz, ce qui est nettement inférieur à la largeur de la bande critique auditive, estimée par sa bande rectangulaire équivalente (ERB = 240 Hz ; Glasberg et Moore, 1990). Par contre, pour $f_c = 125$ Hz les stimuli couvrent 60 Hz, ce qui est supérieur à la bande critique de 38 Hz à

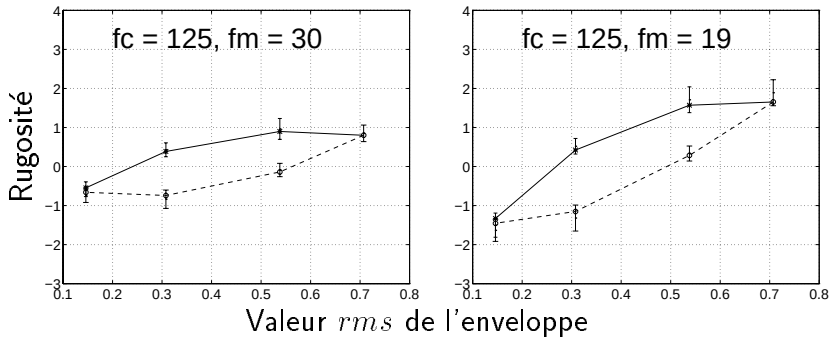


FIG. 5.3 – Rugosité en fonction de la valeur $rms$ de l'enveloppe d'amplitude pour $f_c = 125$ Hz. A gauche, reproduction des résultats de la Figure 5.2, à droite résultats de l'expérience de contrôle.

cette fréquence centrale. Dans ce cas, les interactions ont principalement lieu entre composantes adjacentes (séparées de 30 Hz).

De manière à tester cette interprétation, un bloc de contrôle a été rajouté dans l'expérience pour certains (15) sujets. Dans ce bloc, la fréquence de modulation a été choisie telle que le rapport $f_m$/ERB soit identique à celui utilisé pour la condition $f_c = 1000$ Hz. Une série de sons pAM avec $f_c = 125$ Hz, $f_m = 19$ Hz a donc été générée. Les résultats de cette expérience de contrôle sont reproduits en Figure 5.3. Le contraste global augmente par rapport à la condition originale, ce qui confirme notre hypothèse. La plus grande variabilité des données reflète sans doute la faible rugosité produite par une fréquence de modulation aussi basse : les jugements expérimentaux deviennent alors plus difficiles.

Le contraste réduit observé dans les données originales suggère que l'effet de l'enveloppe globale des sons pAM, quand elle est manipulée par des relations de phase, est d'autant plus marqué que les trois composantes peuvent interagir à l'intérieur d'une même bande critique. Ce résultat est cohérent avec les données disponibles sur les effets de phase monauraux observés avec des stimuli similaires (Mathes et Miller, 1947; Goldstein, 1967b; Terhardt, 1974a; Buunen, 1975).

#### 5.3.1.2 Période de l'enveloppe

Un aspect des stimuli autre que la valeur $rms$ de l'enveloppe peut être proposé pour expliquer une partie de la modification de rugosité perçue quand $|\phi|$ varie. Dans les conditions où $|\phi| = \pi/2$, la période de l'enveloppe est divisée par

deux (voir l'expression analytique en section 5.4.2.1). Ce changement fait que la fréquence de modulation apparente des sons pAM devient plus élevée que celle originalement choisie pour produire une rugosité maximale. Ceci pourrait contribuer à la faible rugosité observée. Néanmoins, deux éléments distincts suggèrent que la plus grande contribution dans les jugements de rugosité provient des variations dans la valeur $rms$ de l'enveloppe et non de la période de cette enveloppe. Pour toutes les conditions de fréquences centrales, la division de la période par deux n'est complètement présente que pour $|\phi| = \pi/2$. Pour les autres valeurs de $\phi$, la période reste constante et pourtant la rugosité varie parallèlement à la valeur $rms$ de l'enveloppe. Pour l'expérience de contrôle avec $f_c = 125$ Hz et $f_m = 19$ Hz, la fréquence de modulation est inférieure à la fréquence de modulation produisant une rugosité maximale. Dans ce cas précis, la division de la période de l'enveloppe conduit à une fréquence de modulation apparente de 38 Hz, ce qui devrait produire plus de rugosité. Ceci n'est pas observé et les valeurs de rugosité quand la division de période est présente sont toujours faibles, en accord avec la faible valeur $rms$ de l'enveloppe (Figure 5.3).

#### 5.3.1.3 Evaluation de l'hypothèse temporelle

Lorsque la valeur $rms$ de l'enveloppe diminue, une diminution de rugosité est aussi observée malgré le fait que la composition spectrale des stimuli reste constante. Cette variation est d'autant plus forte que l'étendue fréquentielle des stimuli est faible devant la largeur de la bande critique. Tout ces aspects des résultats sont en bon accord avec l'hypothèse du lien entre la rugosité et le degré de modulation effective de l'enveloppe (Terhardt, 1974a). Néanmoins, l'influence significative de la phase pour une enveloppe et un spectre d'amplitude donnés n'est pas attendue avec une telle hypothèse. La suite de la discussion a pour objet d'évaluer plusieurs facteurs qui pourraient sous-tendre cette observation nouvelle.

### 5.3.2 Sons de combinaison

#### 5.3.2.1 Produit de distorsion cubique de premier ordre

Les sons de combinaison générés par les trois composants des stimuli pAM ont pu jouer un rôle dans les jugements de rugosité. Parmi ces sons de combinaison,

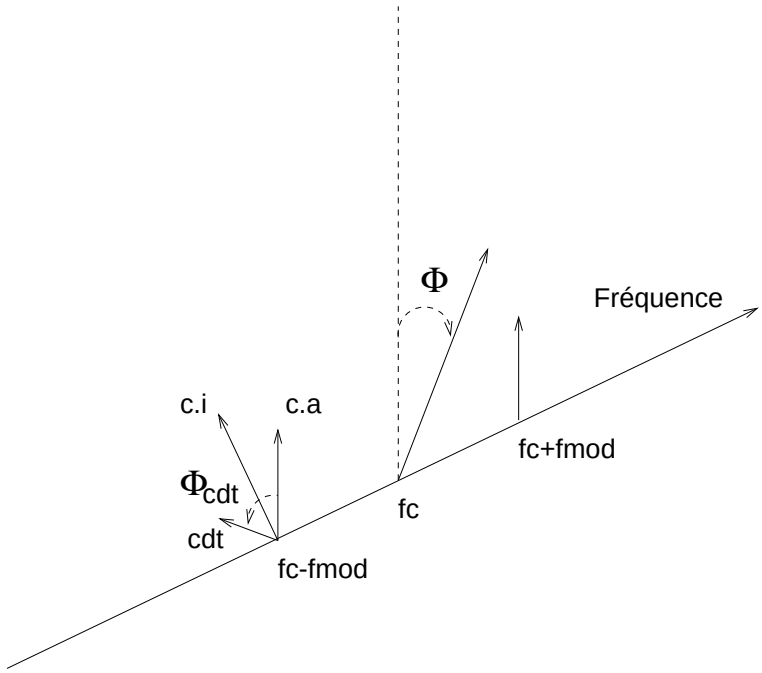


FIG. 5.4 – Représentation dans l'espace des phases des sons pAM et influence du produit de distorsion cubique. Chaque composante est représentée par un vecteur dont la position sur l'axe horizontal correspond à sa fréquence et l'angle à sa phase. Le produit de distorsion (c.d.t) s'ajoute à la composante acoustique (c.a) pour produire la composante interne (c.i).

de nombreuses études ont montré que le plus important était le premier produit de distorsion cubique (Goldstein, 1967a, 1970; Plomp, 1976; Greenwood, 1991b). Lorsque deux composantes acoustiques (appelées les primaires) situées à $f_1, f_2$ avec $f_1 < f_2$ sont présentées simultanément au système auditif, un produit de distorsion (*cubic distortion tone* ou CDT) situé à $2f_1 - f_2$ est généré. Dans le cas des sons pAM, si les composantes situées à $f_c$ et $f_c + f_m$ sont considérées comme étant les primaires, un CDT est généré à la fréquence $2(f_c) - (f_c + f_m) = f_c - f_m$. Il coïncide donc en fréquence avec l'une des composantes physiques du son pAM et peut potentiellement interagir avec elle[1].

Les CDT générés par des stimuli relativement similaires ont été étudiés de façon approfondie (Buunen, Festen, Bilsen et van den Brink, 1974; Buunen, 1976). Les stimuli employés étaient constitués de trois composantes fréquentielles régulièrement espacées (comme pour les sons pAM), représentant des harmoniques

1. Ce type de produit de distorsion ne peut bien entendu pas être masqué sans affecter le stimulus lui-même.

élevés d'une fondamentale absente et avec les amplitudes des deux composantes supérieures égales entre elles pour favoriser l'apparition du CDT (contrairement au cas des sons pAM). La composante grave était choisie avec une amplitude de -10 dB par rapport aux primaires, soit d'un ordre de grandeur similaire aux niveaux attendus pour le CDT. Dans ce cas, l'interaction entre la composante acoustique et le CDT sont accentuées. Cette interaction peut être comprise comme une somme vectorielle (voir Figure 5.4). L'hypothèse testée est que si pour une certaine configuration de phase des primaires le produit de distorsion cubique et la composante acoustique s'annulent, la hauteur résiduelle du stimulus doit devenir moins saillante. Les résultats ont indiqué que, effectivement, la phase des primaires a une influence sur la saillance de la hauteur résiduelle, et qu'une certaine condition de phase produit une hauteur résiduelle moins saillante que les autres. Ceci a permis aux auteurs de conclure que la phase du CDT change avec celle des primaires et que l'intensité de la "composante interne" perçue à la fréquence $f_c - f_m$ en est affectée.

#### 5.3.2.2 Influence du CDT sur la rugosité des sons pAM

Suivant ce résultat, les différences de rugosité observées entre sons pAM de mêmes enveloppes et spectres d'amplitude physiques pourraient être attribuées à l'influence de la phase des primaires sur la phase du CDT. La phase modifierait alors le "spectre interne" et donc la rugosité perçue[2].

Un élément des données obtenues avec les sons pAM est cohérent avec cette hypothèse. Une loi a été dérivée pour relier la variation de phase du son de combinaison avec celle de la phase des primaires (Buunen, Festen, Bilsen et van den Brink, 1974). Cette loi est donnée en équation 5.2 :

$$\delta\phi_{CT} = 2\delta\phi_{f_1} - \delta\phi_{f_2} \tag{5.2}$$

Dans le cas des sons pAM, $\delta\phi_{f_1} = \phi$ et $\delta\phi_{f_2} = 0$. Ainsi, pour les deux conditions de phase $\phi = \pm\pi/2$, les changements de phase du CDT sont de $\delta\phi_{CT} = \pm\pi$. De telles valeurs font que la contribution du CDT dans la somme vectorielle est inchangée entre les conditions $\phi = \pm\pi/2$. De fait, aucune différence significative de rugosité n'a été observée entre les conditions $\phi = \pm\pi/2$.

2. Remarquons d'ailleurs que Buunen, dans son travail de thèse (1976), note de façon informelle une différence de rugosité entre ses stimuli.

#### 5.3.2.3 Limites de l'interprétation basée sur les CDT

Cette interprétation pose toutefois deux problèmes. Tout d'abord, l'effet de la phase en tant que telle disparaît lorsque $f_c = 8$ kHz, alors que les produits de distorsions sont vraisemblablement encore présents. Buunen (1976) a notamment pu réaliser des enregistrements physiologiques démontrant leur présence dans ses stimuli à trois composantes. Notons toutefois qu'il n'a pu effectuer les mesures psychophysiques correspondantes. Greenwood (1991a) indique quant à lui que l'amplitude du CDT diminue pour les régions fréquentielles élevées, mais ne remet pas en cause son existence.

Un autre problème est la grande variabilité inter-sujets observée dans la dépendance de la phase du son de combinaison avec la phase des primaires (Buunen, Festen, Bilsen et van den Brink, 1974; Buunen, 1976). En effet, si la variation de phase a pu être résumée dans l'équation 5.2, la phase absolue du CDT varie grandement entre les sujets testés. Les auteurs ont par exemple dû présenter des données individuelles pour mettre l'effet en évidence. Cette variabilité inter-sujets a été de nouveau observée et soulignée par Zurek et Leishowitz (1976) dans des conditions similaires. Ceci signifie que la relation de phase particulière entre primaires qui cause une atténuation maximale de la "composante interne" dépend de chaque sujet.

En résumé, il est probable que les sons pAM aient été accompagnés de produits de distorsion, dont le plus important est le CDT qui interagit avec les primaires de façon différente en fonction de $\phi$. Toutefois, la disparition de l'effet de phase pour $f_c = 8$ kHz d'une part et surtout la cohérence des données à travers les auditeurs et les conditions d'autre part semblent indiquer que les modifications dues aux CDT n'ont pas été l'élément principal sur lequel les jugements de rugosité ont été établis.

### 5.3.3 Facteurs temporels

Des facteurs temporels peuvent être examinés pour interpréter les différences de rugosité entre les conditions de phases positives et négatives. Même si ces conditions produisent des spectres d'amplitude et des enveloppes temporelles identiques, les structures temporelles fines des stimuli ne sont *pas* identiques. De manière à estimer dans une première approche l'effet de ces différences tempo-

relles fines après le filtrage auditif, un modèle computationnel a été utilisé. La décomposition fréquentielle en bandes critiques au niveau de la membrane basilaire a été simulée par un banc de filtres "gammatone" dont la forme, les fréquences centrales et les largeurs ont été choisies de manière à reproduire des données expérimentales de masquage (Patterson, 1976; Glasberg et Moore, 1990; Patterson, Allerhand et Giguère, 1995).

La région fréquentielle produisant une rugosité maximale selon l'hypothèse temporelle pour les sons pAM centrés à $f_c = 1000$ Hz a été déterminée grâce à un modèle de calcul de rugosité présenté en Annexe B. Ce modèle calcule la valeur $rms$ d'une enveloppe efficace après le filtrage auditif[3]. Le filtre produisant le plus de rugosité, selon ce modèle, est celui centré à 1064 Hz. Plus généralement, du fait de l'étalement des patterns d'excitation des sons purs vers les hautes fréquences, les interactions maximales entre composantes ont lieu au-dessus de la fréquence centrale du son pAM. Les filtres gammatone, linéaires et symétriques, sous-estiment cette asymétrie mais leur élargissement en fonction de la fréquence centrale permet néanmoins d'en capturer une partie (Moore, 1997).

Le signal issu du filtre gammatone centré à 1064 Hz est présenté en Figure 5.5, pour les différentes conditions de phase. Un examen de cette Figure montre que les modulations introduites par les conditions de phase opposées ont des formes différentes. Dans les conditions de phases positives, l'enveloppe augmente de façon abrupte puis décroît progressivement, alors que dans les conditions de phases négatives l'enveloppe augmente progressivement puis décroît de manière abrupte. Cette asymétrie n'est pas observée pour $\phi = \pm\pi/2$, et d'autres simulations montrent qu'elle disparaît pour les fréquences centrales élevées. Nous allons démontrer analytiquement dans la section 5.4.2.2 que cette asymétrie dans l'enveloppe, visible dans une simulation numérique, est en fait la conséquence inévitable de l'atténuation de la composante grave des sons pAM et s'observe donc dans tous les filtres auditifs situés au dessus de $f_c$. La dérivation analytique nous permettra aussi de montrer que l'asymétrie doit disparaître pour les conditions $\phi = \pm\pi/2$ pour toutes les fréquences centrales, ainsi que pour toutes les phases si la fréquence centrale augmente. L'asymétrie est donc présente quand des différences

3. Pour déterminer le canal le plus actif, la région de 800 à 1200 Hz a été examinée avec une densité de 16 filtres gammatone par ERB. Les autres paramètres du modèle sont ceux de l'Annexe B.

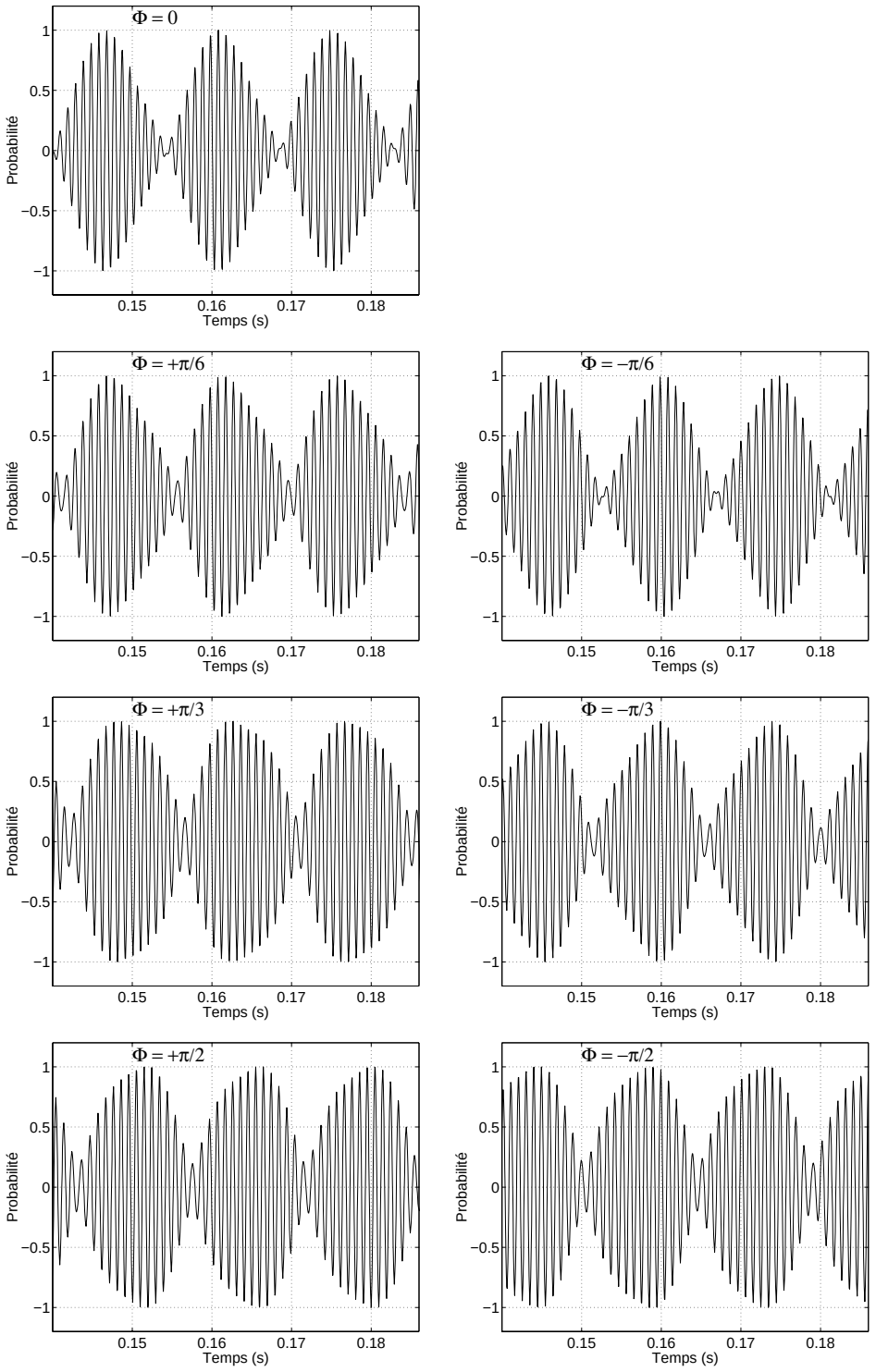


FIG. 5.5 – Effet d'un filtre gammatone centré à 1064 Hz sur les stimuli pAM dans la condition $f_c = 1000$ Hz. Après filtrage, la forme de l'enveloppe est différente entre les condition $\phi > 0$ et $\phi < 0$, sauf pour $\phi = \pm\pi/2$. Les unités d'amplitude sont linéaires mais arbitraires.

de rugosité ont été expérimentalement observées et absente quand les rugosités ont été jugées identiques.

Il est important de noter que la valeur $rms$ de l'enveloppe de modulation après le filtrage auditif reste la même en dépit des asymétries introduites. Il est possible de vérifier ceci dans les simulations et nous le démontrerons dans les dérivations analytiques. L'identité des valeurs $rms$ empêche ici toute interprétation des résultats dans les termes de l'hypothèse temporelle classique. L'exploration des différences de structures temporelles fines des sons pAM, telles qu'accentuées par le filtrage auditif, suggère donc que la forme de l'enveloppe pourrait avoir une influence sur les jugements de rugosité — pour un spectre d'amplitude et une valeur $rms$ de l'enveloppe donnés.

Une autre expérience (Expérience 2) a été construite pour tester directement cette hypothèse. Elle sera présentée au Chapitre 6. Ce chapitre contiendra aussi une discussion générale des résultats des Expériences 1 et 2. Auparavant, la fin du présent chapitre sera consacrée à la dérivation analytique de l'enveloppe des sons pAM ainsi que des effets du filtrage auditif sur ceux-ci, de manière à confirmer les effets déjà décrits à propos des simulations numériques.

## 5.4 Dérivations analytiques

### 5.4.1 Expression analytique d'enveloppes par transformée de Hilbert

#### 5.4.1.1 Cas général

Considérons le cas général d'un signal constitué de la somme de N sons purs de fréquences $f_n = \omega_n/2\pi$ ayant chacun une amplitude et une phase à l'origine quelconques :

$$x(t) = \sum_{n=1}^{N} A_n \cos(\omega_n t + \phi_n) \tag{5.3}$$

Pour extraire l'enveloppe de $x(t)$, une méthode classique consiste à calculer le signal analytique associé à $x(t)$, noté $\tilde{x}(t)$, et à en prendre le module. Le signal analytique associé à un signal réel a pour partie réelle le signal réel, et pour partie imaginaire la transformée de Hilbert du signal réel, soit :

$$\tilde{x}(t) = x(t) + i\mathcal{H}[x(t)] \tag{5.4}$$

avec

$$\mathcal{H}[x(t)] = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x(u)}{t-u} du \tag{5.5}$$

Ici, le signal analytique se calcule simplement :

$$\tilde{x}(t) \quad = \quad \sum_{n=1}^{N} A_n e^{i(\omega_n t + \phi_n)} \tag{5.6}$$

Pour obtenir l'enveloppe du signal, que nous notons $E(t)$, il suffit de prendre le module du signal analytique $\tilde{x}(t)$. Après transformations, $E(t)^2$ s'exprime (Hartmann, 1997) :

$$E(t)^2 = \sum_{n=1}^{N} A_n^2 + \sum_{\substack{m>n \\ n=1}}^{N} 2A_n A_m \cos[(\omega_n - \omega_m)t + (\phi_n - \phi_m)] \tag{5.7}$$

Cette formule est très générale, et donne l'expression analytique de l'enveloppe de tout signal exprimé sous la forme de l'équation 5.3.

#### 5.4.1.2 Cas de signaux à trois composantes du type sons pAM

Dans le cas de signaux ne comprenant que trois composantes spectrales situées à $f_c - f_m, f_c$ et $f_c + f_m$, l'équation 5.7 peut se simplifier. Pour l'instant, nous allons considérer les amplitudes des composantes comme des variables notées $A_1, A_2$ et $A_3$. Pour plus de concision dans la notation, nous utiliserons les vitesses angulaire $\omega_c = 2\pi f_c$ et $\omega_m = 2\pi f_m$ pour représenter les fréquences. Les stimuli pAM peuvent alors s'écrire :

$$pAM(t) = A_1 \cos[(\omega_c - \omega_m)t] + A_2 \cos(\omega_c t + \phi) + A_3 \cos[(\omega_c + \omega_m)t] \tag{5.8}$$

Dans ce cas, l'expression de l'enveloppe est :

$$\begin{aligned} E(t)^2 &= \underbrace{A_1^2 + A_2^2 + A_3^2}_{dc} + \underbrace{2A_1A_3\cos(2\omega_m t)}_{a(t)} \\ &\quad + \underbrace{2[A_1A_2\cos(\omega_m t + \phi) + A_2A_3\cos(\omega_m t - \phi)]}_{b(t)} \end{aligned} \tag{5.9}$$

Nous pouvons décomposer l'équation 5.9 en une somme de trois termes : une composante continue que nous appellerons $dc$, une composante sinusoïdale de fréquence $2\omega_m$ appelée $a(t)$, et la somme de deux composantes de fréquence $\omega_m$ appelée $b(t)$.

Le terme $b(t)$ est la somme de deux cosinus de même fréquence $w_m$ mais avec des amplitudes et phases différentes. Nous savons alors que le signal résultant sera de même fréquence $w_m$, reste à déterminer son amplitude et sa phase. Exprimons donc $b(t)$ avec la notation d'Euler :

$$\begin{aligned} b(t) &= 2A_1A_2e^{i(\omega_m t+\phi)} + 2A_2A_3e^{i(\omega_m t-\phi)} \\ &= 2A_2e^{i\omega_m t}(A_1e^{i\phi} + A_3e^{-i\phi}) \end{aligned} \tag{5.10}$$

Notons

$$\begin{aligned} A_1e^{i\phi} + A_3e^{-i\phi} &= \vec{v} \\ &= \alpha e^{i\theta} \end{aligned} \tag{5.11}$$

Par addition dans le plan complexe nous obtenons (voir figure 5.6) :

$$\begin{aligned} \mathcal{R}(\vec{v}) &= (A_1 + A_3)\cos(\phi) \\ \mathcal{I}m(\vec{v}) &= (A_1 - A_3)\sin(\phi) \end{aligned} \tag{5.12}$$

D'où, par de simples transformations le module de $\vec{v}$ :

$$\alpha = \sqrt{(A_1 - A_3)^2 + 4A_1A_3\cos^2(\phi)} \tag{5.13}$$

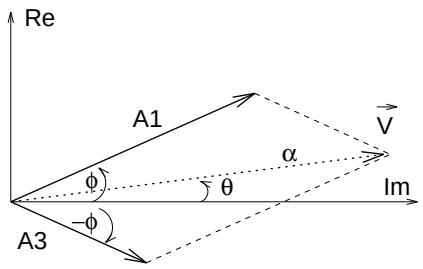


FIG. 5.6 – Addition dans le plan complexe des deux composantes d' amplitudes inégales composant $\vec{v}$.

ainsi que son argument :

$$\begin{cases} \theta = \arctan[\frac{A_1 - A_3}{A_1 + A_3} tan(\phi)] & (\phi \neq \pm\pi/2) \\ \theta = sgn(A_1 - A_3)\phi & (\phi = \pm\pi/2) \end{cases} \tag{5.14}$$

Ainsi, $b(t)$ est un cosinus de fréquence $w_m$, dont l'amplitude et la phase dépendent de $A_1, A_3$ et $\phi$.

Finalement, nous pouvons écrire l'expression analytique de l'enveloppe du signal décrit dans l'équation 5.8 :

$$\left\| \begin{array}{lcl} E(t)^2 & = & dc + a(t) + b(t) \\ dc & = & A_1^2 + A_2^2 + A_3^2 \\ a(t) & = & 2A_1A_3\cos(2\omega_m t) \\ b(t) & = & 2A_2\sqrt{(A_1 - A_3)^2 + 4A_1A_3\cos^2(\phi)}cos(\omega_m t + \theta) \\ & & \begin{cases} \theta = \arctan[\frac{A_1 - A_3}{A_1 + A_3} tan(\phi)] & (\phi \neq \pm\pi/2) \\ \theta = sgn(A_1 - A_3)\phi & (\phi = \pm\pi/2) \end{cases} \end{array} \right. \tag{5.15}$$

### 5.4.2 Application des formules analytiques aux sons pAM

#### 5.4.2.1 Enveloppe des sons pAM

L'enveloppe des sons pAM peut être obtenue grâce à l'équation 5.15 en fixant $A_1 = 1/2, A_2 = 1$ et $A_3 = 1/2$. Une simplification majeure alors vient du fait que, dans le cas des sons pAM, les amplitudes des composantes graves et aiguës sont égales ($A_1 = A_3$). Dans ce cas, l'équation 5.14 indique que :

$$\forall \phi, \theta = 0 \tag{5.16}$$

D'où par simplification de 5.15 :

$$\left\| \begin{array}{lcl} E(t)^2 & = & dc + a(t) + b(t) \\ dc & = & 3/2 \\ a(t) & = & 1/2\cos(2\omega_m t) \\ b(t) & = & 2|\cos(\phi)|\cos(\omega_m t) \end{array} \right. \tag{5.17}$$

De manière à comprendre l'influence de $\phi$ sur la forme de l'enveloppe, examinons les contributions de chacun des trois termes qui participent au carré de $E(t)$. Cette influence est illustrée en Figure 5.7. Le terme $dc$ est simplement une valeur continue. Le terme $a(t)$ est une composante sinusoïdale de fréquence $2f_m$ qui ne dépend pas de $\phi$. Le terme $b(t)$ est une autre composante sinusoïdale, de fréquence $f_m$, dont l'amplitude dépend de $\phi$. Si $\phi = 0$, $b(t)$ a une amplitude maximum. Dans ce cas, l'équation 5.17 se réduit à une modulation cosinusoïdale. En effet :

$$\begin{array}{rcl} E(t)^2 & = & 3/2 + 1/2\cos(2\omega_m t) + 2\cos(\omega_m t) \\ & = & 3/2 + \cos^2(\omega_m t) - 1/2 + 2\cos(\omega_m t) \\ & = & (1 + cos(\omega_m t))^2 \end{array} \tag{5.18}$$

La valeur $rms$ de cette enveloppe est donc $\sqrt{2}/2$. Si $\phi = \pm\pi/2$, le terme $b(t)$ disparaît et seuls restent le terme $a(t)$ et la composante continue $dc$. Dans ce cas, la valeur $rms$ de l'enveloppe n'est plus que de $(1 - \sqrt{2}/2)/2$ et sa période et divisée par deux. Notons enfin que l'amplitude de $b(t)$ ne dépend pas du signe

de $\phi$ : deux phases de même valeur absolue mais de signes opposés produiront exactement la même enveloppe d'amplitude.

#### 5.4.2.2 Effet de l'atténuation de la composante grave

Examinons l'influence d'un filtre passe-bande idéal qui serait à phase linéaire et qui atténuerait la composante grave des sons pAM de moitié. Dans ce cas, l'enveloppe du signal résultant est donné par l'équation 5.15 avec $A_1 = 1/4, A_2 = 1, A_3 = 1/2$. Son expression est :

$$\left\| \begin{array}{lcl} E(t)^2 & = & dc + a(t) + b(t) \\ dc & = & 11/8 \\ a(t) & = & 1/4\cos(2\omega_m t) \\ b(t) & = & \sqrt{1/4 + 2\cos^2(\phi)}\cos(\omega_m t + \theta) \\ \theta & = & \begin{cases} \arctan[-1/3tan(\phi)] & \text{si } (\phi \neq \pm\pi/2) \\ -\phi & \text{si } (\phi = \pm\pi/2) \end{cases} \end{array} \right. \tag{5.19}$$

L'élément important à remarquer ici, au delà des valeurs numériques particulières à l'exemple choisi, est que la phase de $b(t)$ dépend de $\phi$. Si $\phi > 0$, alors $\theta < 0$ et $b(t)$ va être légèrement retardé par rapport à $a(t)$. Ce retard va se traduire par une asymétrie dans la forme de l'enveloppe. Comme $b(t)$ a une amplitude plus importante que $a(t)$, dans ce cas l'asymétrie consistera en une croissance rapide et une décroissance lente. Au contraire, si $\phi < 0$, $b(t)$ est légèrement en avance et l'enveloppe a une une croissance lente et une décroissance rapide. Les constructions graphiques proposées en Figure 5.8 permettent d'illustrer ceci.

Cette enveloppe a une autre propriété intéressante. Dans les cas $\phi = +\pi/2$ et $\phi = -\pi/2$, le déphasage de $b(t)$ est respectivement de $\theta_{\pi/2} = -\pi/2$ et $\theta_{-\pi/2} = +\pi/2$. La différence en ce qui concerne la position relative de $a(t)$ et de $b(t)$ entre ces conditions est donc de $\theta_{-\pi/2} - \theta_{+\pi/2} = +\pi$, soit une demi-période de $b(t)$. Or une demi-période de $b(t)$ correspond à une période entière de $a(t)$ : la forme de l'enveloppe est donc inchangée, elle est seulement décalée dans le temps. La forme de l'enveloppe ne dépend donc pas du signe de $\phi$ dans le cas particulier $\phi = \pm\pi/2$.

L'influence de $\phi$ que nous venons de décrire ne dépend pas des valeurs numériques particulières choisies pour l'illustrer. Cette influence sera toujours quali-

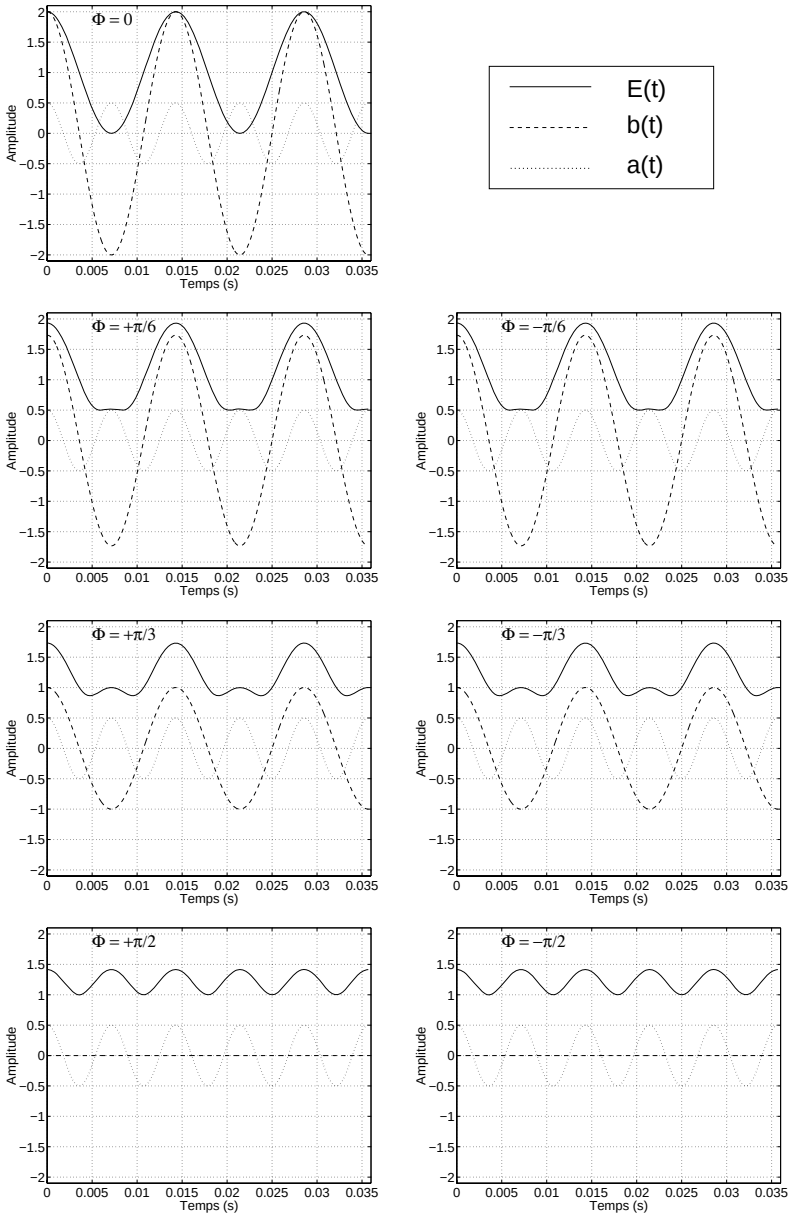


FIG. 5.7 – Construction graphique de l'enveloppe des sons pAM pour les valeurs de $\phi$ employées expérimentalement. $a(t)$ est représenté en trait pointillé, $b(t)$ en trait hachuré et $E(t)$ en trait plein (voir équation 5.9). Les conditions $\phi$ positif ou $\phi$ négatif, produisent la même enveloppe, car $b(t)$ ne dépend que de la valeur absolue de $\phi$. Si $\phi$ augmente, l'amplitude de $b(t)$ diminue et l'enveloppe s'aplatit.

tativement similaire si $A_1 < A_3$, c'est-à-dire si la composante grave du son pAM est plus atténuée que la composante aiguë.

### 5.4.3 Effet du filtrage auditif sur les sons pAM

L'exemple que nous venons de considérer illustre qualitativement ce qui se passe pour les canaux auditifs de fréquence centrale supérieure à la fréquence centrale du son pAM. Dans ces canaux, du fait du filtrage passe bande centré au dessus du stimulus, la composante grave du son pAM est plus atténuée que la composante aiguë (une atténuation supplémentaire de la composante centrale ne change pas le raisonnement).

A la différence de cet exemple, le filtrage gammatone comme le filtrage auditif réel ne sont pas à phase linéaire (Smith, Sieben, Kohlrausch et Schroeder, 1986; Kohlrausch et Sander, 1995; Giguère et Woodland, 1994). Toutefois, la faible étendue fréquentielle des sons pAM permet de faire une linéarisation de ce filtrage autour de la fréquence centrale du stimulus sans introduire une erreur trop élevée. Ceci est illustré en Figure 5.9 pour le cas du filtre gammatone, qui est à phase minimale, et sera discuté en section 6.4 pour ce qui est du filtrage "réel". Une telle linéarisation a été faite dans un cas similaire par Goldstein (1967b). Notre hypothèse ici est que tout filtrage physiquement raisonnable permettra de faire cette linéarisation pour les sons pAM. Ce filtrage, s'il atténue la composante grave des sons pAM par rapport à la composante aiguë, produira alors forcément une asymétrie telle que celle que nous avons décrite.

Dans les canaux auditifs situés au-dessous de la fréquence centrale du son pAM, le filtrage auditif a pour effet d'atténuer la composante aiguë par rapport à la composante grave. L'asymétrie est dans ce cas inversée. Toutefois, du fait de l'étalement des patterns d'excitation des sons purs vers les hautes fréquences, ces canaux contiendront bien moins d'activité. Au niveau de l'activité globale, notre hypothèse est donc que le filtrage en bandes critiques a pour effet de redresser le front de la modulation d'amplitude pour les conditions de phase positive et de l'adoucir pour les conditions de phase négative.

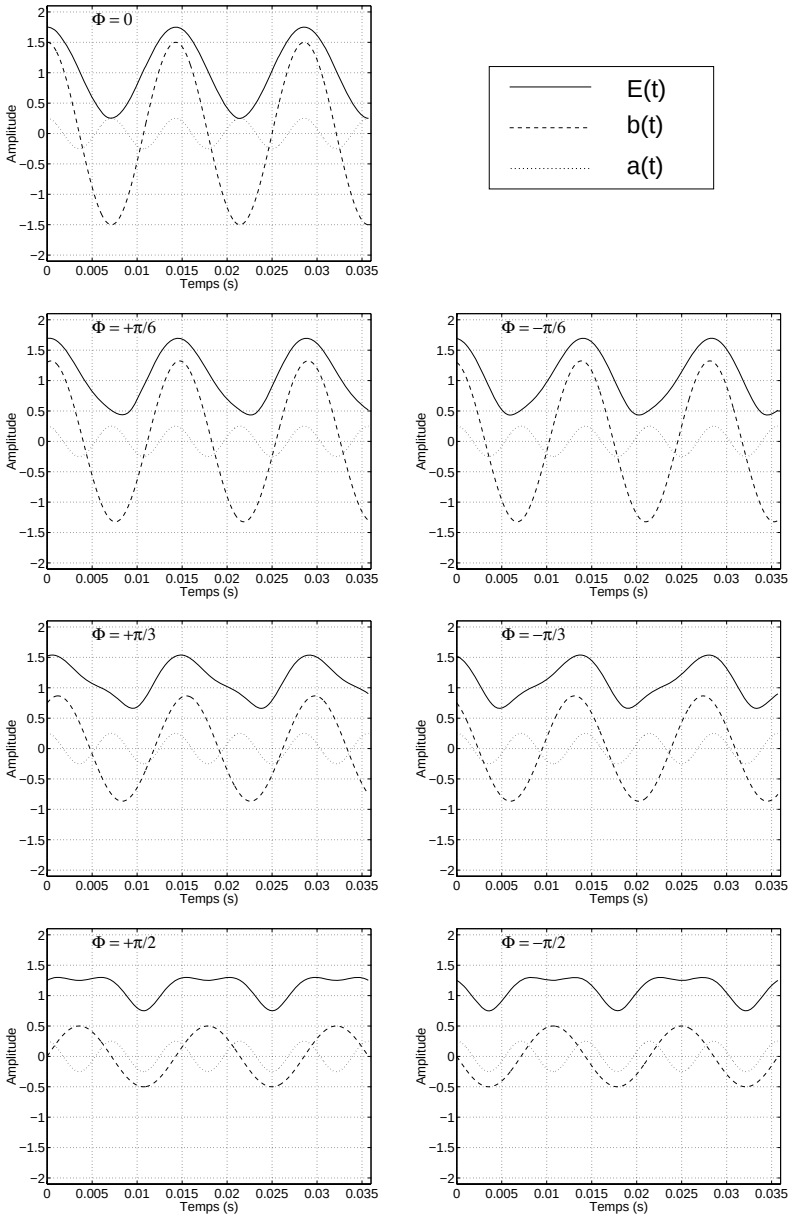


FIG. 5.8 – Construction graphique de l'enveloppe des sons pAM aux amplitudes modifiées : $a(t)$ est représenté en trait pointillé, $b(t)$ en trait hachuré et $E(t)$ en trait plein. Une asymétrie apparaît entre $\phi$ positif et négatif car la phase de $b(t)$ dépend de $\phi$. Cette asymétrie disparaît pour $\phi = \pm\pi/2$ (voir texte).

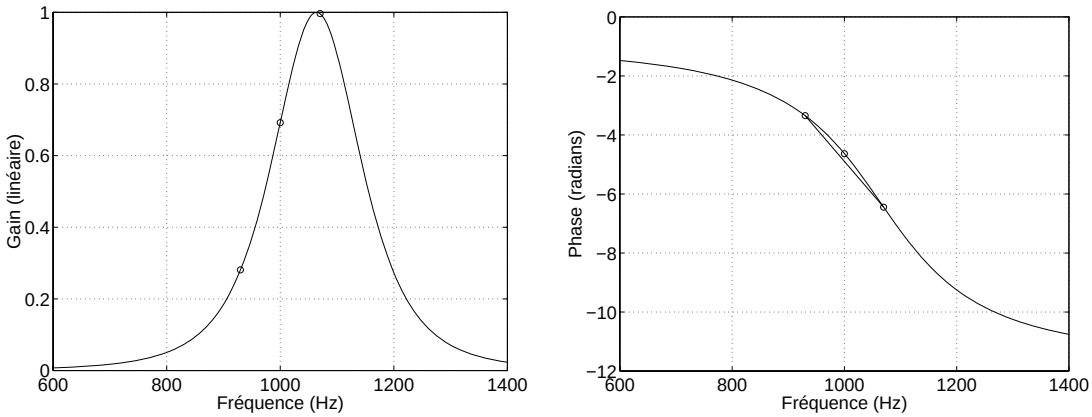


FIG. 5.9 – Réponse en amplitude (a) et en phase (b) du filtre gammatone centré à 1064 Hz autour de sa fréquence de résonance et approximation linéaire pour la réponse en phase. Les fréquences présentes dans le son pAM sont indiquées par des points. L'hypothèse de phase linéaire est raisonnable pour un signal à bande aussi étroite que les sons pAM.

#### 5.4.3.1 Conclusion des dérivations analytiques

Les dérivations analytiques nous ont permis de montrer que tout filtrage pouvant être approximé par un filtrage à phase linéaire et qui atténue la composante grave des sons pAM par rapport à la composante aiguë introduit une asymétrie dans l'enveloppe telle que celle illustrée par les simulations de la Figure 5.5. Cette asymétrie entre phases positives et négatives disparaît pour $\phi = \pm\pi/2$ quelle que soit $f_c$. Si $f_c$ augmente, les valeurs particulières de $f_c$ et de $f_m$ choisies dans l'expérience font que l'asymétrie disparaît progressivement car dans ce cas, la largeur des bandes critiques devient très supérieure à celle des stimuli (qui reste constante et égale à 140 Hz pour $f_c \geq 1000$). Par exemple, pour $f_c = 8000$ Hz, le filtre situé à 8000 Hz aurait une ERB de 889 Hz. Il s'ensuit que ces filtres auditifs ne modifient que très peu le rapport d'amplitude entre composantes et n'introduisent qu'une très faible asymétrie. Ainsi, les dérivations analytiques permettent de prévoir une asymétrie introduite par le filtrage auditif quand les rugosités ont été jugées différentes et une absence d'asymétrie quand les rugosités ont été jugées identiques. De plus, ces asymétries ne modifient pas la valeur $rms$ de l'enveloppe : en effet, les différences de forme sont entièrement dues aux relations de phase entre composantes de l'enveloppe.

## 5.5 Conclusion

L'influence de la phase sur la perception de rugosité de sons ayant le spectre d'amplitude de sons SAM a été étudiée expérimentalement. Il a été démontré que la phase de la composante centrale, quand elle modifie l'enveloppe sans changer le spectre d'amplitude, a une influence importante sur la perception de rugosité. Ce résultat confirme et étend ceux obtenus par Mathes et Miller (1947) et Terhardt (1974a). Toutefois, il est aussi apparu que la phase peut aussi avoir une influence sur la perception de rugosité même si elle laisse l'enveloppe constante. Ce résultat nouveau semble ne pas pouvoir être interprété de façon complète dans les cadres théoriques existants. Nous avons proposé pour l'expliquer une hypothèse basée sur une asymétrie de l'enveloppe introduite par le filtrage auditif. Cette hypothèse, basée sur des simulations numériques et des dérivations analytiques, va être testée expérimentalement dans le prochain Chapitre.

# Chapitre 6

# Influence de la forme de l'enveloppe de modulation

*L'influence de la forme de l'enveloppe sur la perception de rugosité est démontrée expérimentalement. Ceci autorise une interprétation commune pour les Expériences 1 et 2.*

La rugosité de sons purs modulés en amplitude par des enveloppes ayant le même spectre d'amplitude, la même valeur $rms$, mais des formes différentes, est étudiée. Pour les modèles existants, de tels sons devraient produire une même rugosité. Néanmoins, les résultats de l'Expérience 1 suggèrent qu'il pourrait en être autrement. Une modulation similaire à une onde "en dent de scie" a été utilisée et comparée à son inversion temporelle. Les sons avec des enveloppes ayant une croissance abrupte et une décroissance lente ont été jugés comme plus rugueux. Ce résultat est cohérent avec l'hypothèse formulée pour interpréter les données de l'Expérience 1. Les données combinées des deux expériences sont donc de nouveau discutées, et la pertinence des effets observés est estimée dans la perspective de modélisation de la rugosité de sons musicaux.

## 6.1 Introduction

Il n'existe pas, à notre connaissance, de données expérimentales directement pertinentes quant au problème de l'influence potentielle de la forme de l'enveloppe sur la perception de rugosité. Les stimuli utilisés dans les études que nous avons regroupées sous le nom d'approche spectrale ont été pensés en fonction de leurs représentations spectrales d'amplitude. Les stimuli utilisés par l'approche temporelle étaient quasi-exclusivement sinusoïdalement modulés en amplitude. Pour les seules études faisant exception à cette règle – Guirao et Garavila, (1976), avec une modulation rectangulaire, Kemp (1982) avec des stimuli modulés en fréquence, Aures (1985b) avec des bruits filtrés – la forme d'enveloppe n'a pas été considérée comme un paramètre en tant que tel. Les prévisions des deux approches concernant la question de la forme d'enveloppe sont donc à dériver indirectement des modèles de calcul qu'elles proposent.

Pour ce qui est de l'approche spectrale (Plomp et Levelt, 1965; Kameoka et Kuriyagawa, 1969a; Hutchinson et Knopoff, 1978), la forme de l'enveloppe est par nature considérée comme non pertinente. En effet, du fait de l'hypothèse que seuls les paramètres spectraux sont nécessaires pour pouvoir estimer la perception de rugosité, le paramètre essentiellement "temporel" qu'est la forme d'enveloppe ne peut être directement pris en compte. Indirectement, si une différence de forme d'enveloppe se traduit par une différence dans le spectre d'amplitude du stimulus, ces modèles pourront prédire une différence de rugosité. Néanmoins, pour une composition spectrale donnée ces modèles prédiront une seule et unique rugosité.

Dans les modèles temporels, le calcul du degré de modulation effective prend en compte dans une certaine mesure à la fois le spectre d'amplitude du son et les relations de phase présentes entre ses composantes (Terhardt, 1974a; Aures, 1985b; Daniel et Weber, 1997). Prenons pour simplifier l'exemple de sons pouvant être considérés comme limités à l'intérieur d'un seul canal auditif. Le calcul de la rugosité revient alors à extraire une valeur $rms$ de l'enveloppe acoustique filtrée (voir section 3.2.3). Un changement de forme de l'enveloppe peut affecter la rugosité estimée si ce changement est accompagné d'une variation de valeur $rms$, et si cette variation est conservée après le filtrage des fréquences rugueuses. Cette caractéristique des modèles temporels a permis d'expliquer une partie des résultats de l'Expérience 1. Néanmoins, si la forme de l'enveloppe change alors qu'une même valeur $rms$ est conservée — comme par exemple au cours de l'inver-

sion temporelle de signaux asymétriques dans le temps — les modèles temporels ne prédisent aucune variation de rugosité.

En résumé, la forme de l'enveloppe est prise très indirectement en compte par l'approche spectrale à travers les modifications du spectre d'amplitude, et indirectement en compte par l'approche temporelle par les effets combinés de la composition spectrale et des relations de phase sur la valeur $rms$. Aucune des deux approches ne prédit de différences de rugosité pour des signaux avec des enveloppes de même composition spectrale et de même valeur $rms$ mais avec des formes différentes, comme il est possible d'en obtenir en comparant un signal asymétrique dans le temps et son inversion.

## 6.2 Expérience 2

### 6.2.1 Hypothèses expérimentales

Cette expérience se propose de comparer la rugosité perçue pour des sons purs modulés en amplitude dont la forme de modulation est asymétrique. Une modulation ressemblant à une onde " en dent de scie", qui présente une croissance progressive et une décroissance abrupte, a été choisie. La synthèse de sons modulés par une telle enveloppe et les mêmes sons renversés dans le temps permet alors de comparer des stimuli avec un même spectre d'amplitude, une même valeur $rms$ de l'enveloppe, mais dont les formes d'enveloppes présentent des différences similaires à celles qui ont supposées avoir été introduites par le filtrage auditif des stimuli de l'Expérience 1. L'intérêt de ces stimuli en relation avec ceux de l'Expérience 1 est qu'il n'est maintenant plus nécessaire d'inférer les actions du filtrage auditif pour obtenir une asymétrie d'enveloppe, car cette asymétrie est présente dans les signaux acoustiques (et préservée après filtrage auditif, voir la Discussion générale).

Nous avons choisi de ne pas utiliser une véritable onde en dent de scie pour moduler nos stimuli car le spectre d'amplitude d'une telle onde consiste théoriquement en une infinité de raies spectrales, avec une décroissance de 6 dB par octave. En plus des problèmes de repliement que l'échantillonage d'une telle onde peut causer (Stilson et Smith, 1996), les auditeurs pourraient éventuellement utiliser des indices spectraux à travers les canaux auditifs pour faire leurs jugements.

Nous avons donc construit les signaux destinés à devenir les enveloppes des stimuli par troncation du spectre d'amplitude d'une onde en dent de scie idéale. Pour une fréquence centrale $f_c$ donnée, seules les composantes spectrales situées à une distance fréquentielle de moins de 1/2 ERB de $f_c$ ont été considérées. Ceci permet de limiter autant que possible les indices spectraux présent dans les stimuli après modulation. Une technique de synthèse additive a permis de générer une version "dent de scie" de l'enveloppe des stimuli, qui a ensuite été multipliée par la porteuse moyennant une certaine profondeur de modulation. Les équations 6.1 et 6.2 résument ce procédé :

$$\begin{cases} E_{ds}(t) & = & \sum_{n=1}^{n=N} \frac{1}{n} \cos(2\pi n f_m t - \pi/2) \\ N.f_m & \leq & \frac{1}{2}\mathrm{ERB}(f_c) \end{cases} \tag{6.1}$$

$$x_{ds}(t) = (1 + m.\frac{E_{ds}(t)}{max[E_{ds}(t)]}).sin(2\pi f_c t) \tag{6.2}$$

Pour obtenir la version "dent de scie inversée", il suffit de prendre l'inversion temporelle de l'enveloppe, soit $E_{ids}(t) = E_{ds}(-t)$. Par de simples transformations trigonométriques, il est possible de montrer que ceci est dans notre cas équivalent à inverser le signe des phases de chaque composante :

$$\begin{cases} E_{ids}(t) & = & \sum_{n=1}^{n=N} \frac{1}{n} \cos(2\pi n f_m t + \pi/2) \\ N.f_m & \leq & \frac{1}{2}\mathrm{ERB}(f_c) \end{cases} \tag{6.3}$$

$$x_{ids}(t) = (1 + m.\frac{E_{ids}(t)}{max[E_{ids}(t)]}).sin(2\pi f_c t) \tag{6.4}$$

Des illustrations de ces stimuli correspondant aux conditions expérimentales seront proposées en Figure 6.1. De manière à se placer dans une situation aussi similaire que possible à celle utilisée pour recueillir les données de l'Expérience 1, différentes fréquences centrales $f_c$ ont été testées. Pour chaque condition, la valeur $rms$ de l'enveloppe a été variée en modifiant la profondeur de modulation $m$. Selon les approches spectrales et temporelles, les stimuli possédant une plus grande profondeur de modulation devraient être plus rugueux à la fois du fait de l'augmentation du niveau des composantes de modulation par rapport à la porteuse (approche spectrale) et du fait de l'accroissement de la valeur $rms$ de l'enveloppe (approche temporelle).

Pour chaque valeur $rms$ de l'enveloppe, deux versions des stimuli appelées "dent de scie" et "dent de scie inversée" ont été produites. Selon les approches spectrales et temporelles, les versions "dent de scie" et "dent de scie inversée" qui ont une même composition spectrale et une même valeur $rms$ d'enveloppe devraient avoir la même rugosité. Selon l'hypothèse que nous avons proposée pour interpréter les résultats de l'Expérience 1, les stimuli "dent de scie inversée" qui ont une croissance rapide et une décroissance lente devraient être jugés plus rugueux que les stimuli "dent de scie" qui ont une croissance lente et une décroissance rapide.

### 6.2.2 Méthode

#### 6.2.2.1 Sujets

Dix auditeurs ont participé à l'expérience. Le groupe d'auditeurs consistait en 6 hommes et 4 femmes âgés de 24 à 45 ans ($\overline{M} = 28ans$). Ils ont été recrutés de manière similaire à l'Expérience 1 et n'étaient donc pas sélectionnés sur la base d'aptitudes musicales ou d'expérience antérieure avec des tests psychoacoustiques.

#### 6.2.2.2 Stimuli

Les sons décrits dans les équations 6.2 et 6.4 ont été employés. La fréquence de modulation $f_m$ a été choisie égale à 70 Hz par analogie avec la fréquence de modulation produisant une rugosité maximale pour les sons SAM. La première fréquence centrale pour laquelle plus d'une composante pouvait être retenue moyennant le critère d'étendue fréquentielle limitée était $f_c = 2500$ Hz. Trois conditions ont été retenues : $(f_c, f_m) = (2500$ Hz ; 70 Hz), (5000 ; 70), (10000 ; 70). Pour chaque condition, trois profondeurs de modulation ont été utilisées : $m = 0,4; 0,6; 0,8$. Enfin, les versions "dent de scie" et "dent de scie inversée" ont été inclues dans chaque condition. Les stimuli "dent de scie" et "dent de scie inversée" obtenus pour chaque $f_c$ sont représentés en Figure 6.1.

En résumé, les sons ainsi obtenus sont tous restreints à l'intérieur d'une bande critique. A chaque couple $(f_c, f_m)$ sont associés six stimuli, produisant trois valeurs $rms$ d'enveloppe différentes correspondant aux trois profondeurs de modulation. Pour chaque valeur $rms$ deux formes asymétriques sont produites avec le même spectre d'amplitude.

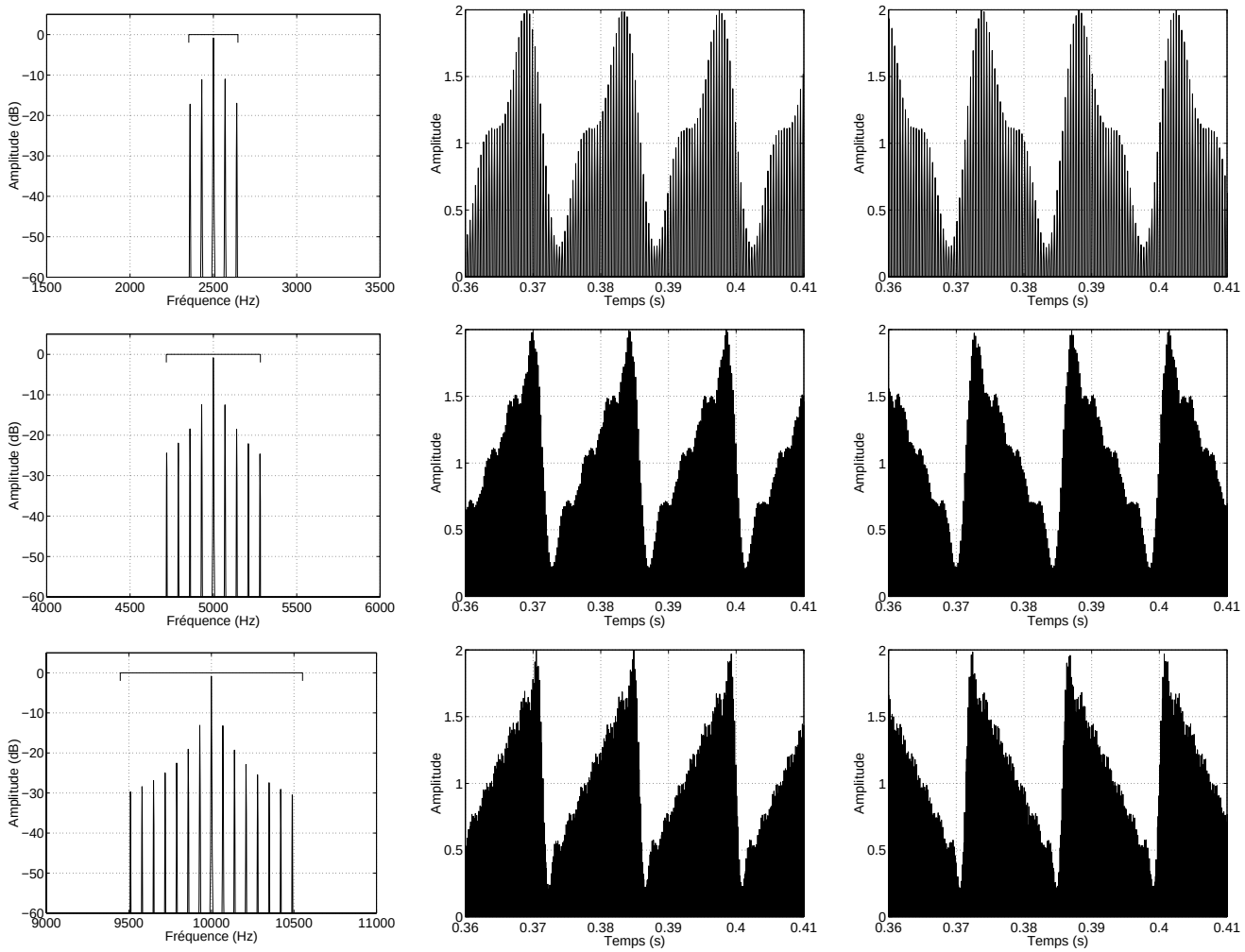


FIG. 6.1 – Stimuli de l'Expérience 2. Les spectres d'amplitude et les ondes "dent de scie" et "dent de scie inversée" des stimuli avec une profondeur de modulation de $m = 0,8$ sont donnés pour $f_c = 2500$ Hz (en haut), $f_c = 5000$ Hz (au milieu), et $f_c = 10000$ Hz (en bas). Dans tous les cas, la fréquence de modulation est de $f_m = 70$ Hz. La bande critique correspondante mesurée en ERB est indiquée par un trait horizontal au dessus des spectres d'amplitude. Le nombre de composantes spectrales augmente avec $f_c$. Par voie de conséquence, l'asymétrie est de plus en plus prononcée.

#### 6.2.2.3 Matériel

Le matériel a été maintenu strictement identique par rapport à l'Expérience 1, à la seule différence que le casque utilisé est maintenant un Sennheiser 420, l'autre étant devenu "indisponible" entre les deux expériences. Le niveau de présentation est de nouveau de 60 dBA.

#### 6.2.2.4 Procédure

Une procédure identique à celle de l'Expérience 1 a été adoptée. La notion de rugosité était introduite à l'aide de sons sinusoïdalement modulés en amplitude similaires à ceux employés pour l'Expérience 1. Ici encore, les auditeurs pouvaient faire varier leur profondeur de modulation de façon continue. Bien que la différence entre les exemples et les sons expérimentaux était plus marquée dans cette expérience, aucun sujet n'a signalé de difficulté à comprendre les instructions. Les stimuli étaient présentés en blocs séparés pour chaque fréquence centrale. Dans chaque bloc, les 30 paires possibles de sons non-identiques dans un ordre comme dans l'autre étaient présentées en ordre aléatoire. Pour chaque paire, les auditeurs devaient décider quel son était le plus rugueux. Chaque bloc durait moins de 10 mn et les auditeurs avaient 3 blocs par session expérimentale.

#### 6.2.2.5 Analyses statistiques

Les analyses de l'expérience précédente ont été exactement reproduites avec les nouvelles données (BTL et bootstrap). Dans cette nouvelle expérience, chaque point de l'échelle de rugosité perçue est le résultat de 70 jugements expérimentaux.

### 6.2.3 Résultats

Les résultats sont présentés en Figure 6.2. L'influence de la valeur $rms$ de l'enveloppe, directement liée à la profondeur de modulation $m$, est significative pour toutes les fréquences porteuses : une plus grande valeur $rms$ produit plus de rugosité. Une influence significative de la forme de l'enveloppe est aussi observée : pour un même spectre d'amplitude et pour une même valeur $rms$ de l'enveloppe, la condition "dent de scie inversée" est systématiquement jugée plus rugueuse que la condition "dent de scie". Cet effet important est significatif à toutes les fréquences centrales, mais il est moins marqué pour $f_c = 2500$ Hz.

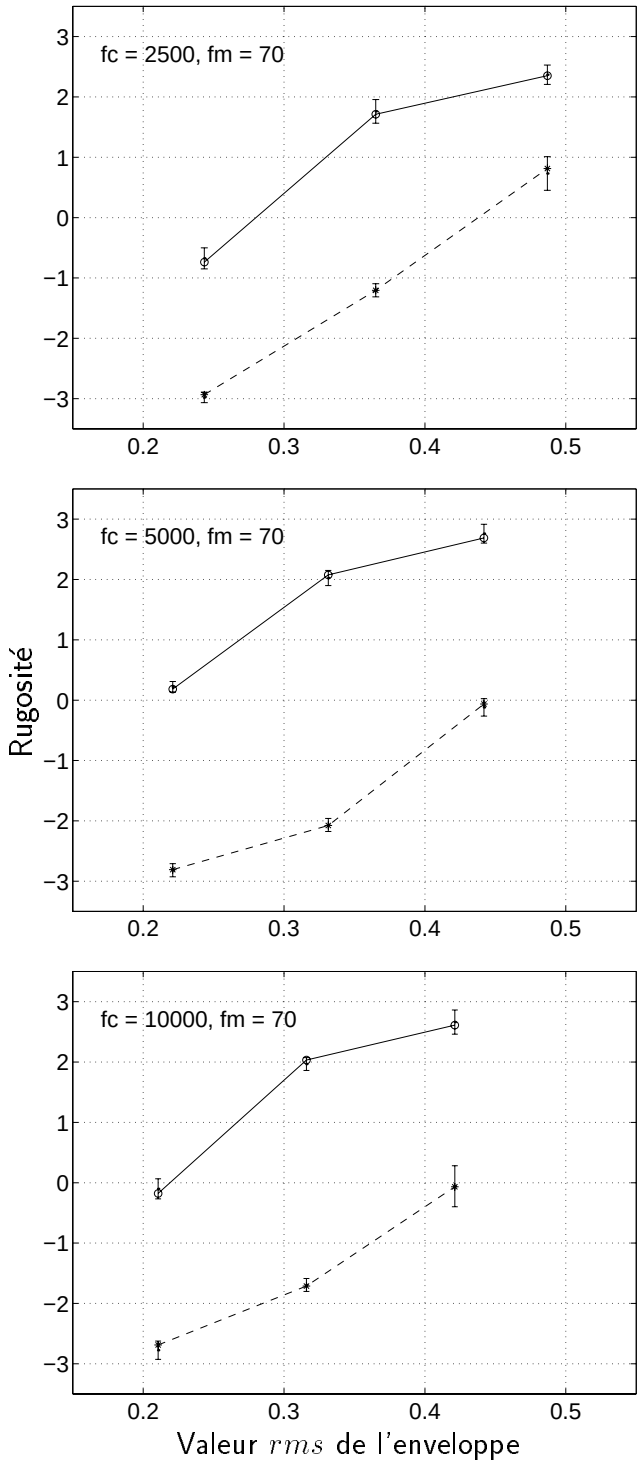


FIG. 6.2 – Résultats de l'Expérience 2 pour les trois fréquences centrales. La valeur $rms$ de l'enveloppe est en abscisse, la valeur de rugosité estimée par la méthode BTL en ordonnée. Les traits plein représentent les modulations "dent de scie inversée", les traits pointillés les modulations "dent de scie".

## 6.3 Discussion

### 6.3.1 Influence de la profondeur de modulation

L'augmentation de la profondeur de modulation reflétée dans la valeur $rms$ de l'enveloppe a eu pour effet une augmentation de la rugosité perçue. Rappelons que contrairement à l'expérience précédente, les stimuli utilisés ici n'ont pas le même spectre d'amplitude entre les conditions de valeurs $rms$ distinctes. L'augmentation de valeur $rms$ est obtenue en augmentant la profondeur de modulation, ce qui a aussi pour effet d'accroître l'amplitude des différentes composantes spectrales. L'augmentation de rugosité avec la valeur $rms$ de l'enveloppe est donc compatible avec une interprétation spectrale ou temporelle des résultats.

### 6.3.2 Influence de la forme de l'enveloppe

Contrairement aux hypothèses spectrales et temporelles, la forme de l'enveloppe a une influence significative à toutes les fréquences centrales testées. Cet effet est cohérent avec les résultats de l'Expérience 1 : les conditions "dent de scie inversée", qui présentent une croissance rapide et une décroissance lente, sont jugées systématiquement plus rugueuses que les conditions "dent de scie" où la croissance est lente et la décroissance rapide.

Cet effet est accentué pour les fréquences porteuses aiguës. Ceci est aussi cohérent avec nos hypothèses. L'effet de phase disparaissait dans l'Expérience 1 pour $f_c = 8000$ Hz. Dans ce cas, l'asymétrie supposée introduite par le filtrage auditif disparaissait aussi. Ici, l'asymétrie n'a pas à être introduite par le filtrage auditif car elle est imposée dans les stimuli eux-mêmes. L'effet de phase (ou de façon équivalente d'inversion temporelle) est alors encore observé à $f_c = 10000$ Hz. Cet effet est même moins prononcé pour la fréquence centrale la plus basse : dans ce cas, du fait du faible nombre de composantes spectrales, l'asymétrie de l'onde acoustique est moins claire, ce qui a eu pour conséquence une réduction de l'effet perceptif (Figure 6.1). Tout ceci suggère que les effets de phase observés dans cette Expérience 2 sont liés à l'enveloppe et non à la structure fine des stimuli qui serait perdue du fait de la perte de verrouillage de phase *(phase locking)* pour les fréquences centrales élevées.

## 6.4 Discussion générale des Expériences 1 et 2

### 6.4.1 Effet de la dispersion au niveau de la membrane basilaire

#### 6.4.1.1 Filtrage gammatone et filtrage "réel"

Nous avons utilisé de manière classique un banc de filtres "gammatone" pour estimer les effets du filtrage auditif sur les stimuli de l'Expérience 1. Le banc de filtres gammatone n'est en fait pas le meilleur moyen d'étudier les effets de la phase sur la perception auditive. En effet, le filtre gammatone est à phase minimale, ce qui signifie que sa réponse en phase peut se déduire intégralement de sa réponse en fréquence. Intuitivement, tous les déphasages introduits sont ceux "dus"aux modifications d'amplitude. Il existe néanmoins un ensemble de données qui indiquent que le filtrage "réel" reflète des effets de dispersion de phase liés à la propagation des ondes le long de la membrane basilaire.

Des expériences comparant le pouvoir de masquage de sons complexes harmoniques dont les phases étaient arrangées selon la loi de Schroeder (1970) ont notamment mis en évidence ces propriétés dispersives dans une situation pouvant être rapprochée des conditions de l'Expérience 1. La loi de phase de Schroeder permet de générer des signaux harmoniques avec un facteur-crête relativement faible. Les signaux étant harmoniques, l'inversion des phases (conditions S+ et S-) ne change pas l'enveloppe acoustique. Il a pourtant été démontré que les propriétés de masquage des sons S+ et S- sont différentes. En termes de seuils absolus de détection de sons purs, les signaux S+ ont un pouvoir masquant nettement inférieur à celui des sons S- (Smith, Sieben, Kohlrausch et Schroeder, 1986). Un modèle computationnel de propagation le long de la membrane basilaire, incluant des propriétés dispersives, a permis de fournir des éléments pour interpréter ce résultat (Strube, 1985). Du fait de la dispersion, les enveloppes des signaux masquants après propagation sont différentes pour les conditions S+ et S- : l'enveloppe du signal S+ possède un facteur-crête supérieur. Les auditeurs pourraient alors, dans cette condition, "écouter dans les vallées" de l'enveloppe pour détecter le masqué, ce qui est impossible pour les conditions S-. Cette hypothèse a été confirmée par la mesure des périodogrammes de masquage *(masking period pattern)* des sons S+ et S- (Kohlrausch et Sander, 1995).

D'autres différences entres sons S+ et S- ont été exhibées en ce qui concerne le masquage temporel pro-actif (Carlyon et Datta, 1997a). Dans ce cas, le son S+ qui possède une enveloppe avec un large facteur-crête semble atténué par un mécanisme de compression rapide, ce qui réduit son pouvoir masquant. De façon intéressante, un effet similaire a été mesuré avec des sons asymétriques dans le temps (Carlyon, 1996). Ces deux derniers ensembles de résultats ont pu être reproduits à l'aide d'un modèle computationnel (Giguère et Woodland, 1994) dérivé de celui de Strube (1985).

L'ensemble de ces résultats impliquent que le filtrage effectué par la membrane basilaire n'est pas à phase linéaire ou à phase minimale. L'effet de la dispersion de phase, qui modifie l'enveloppe des stimuli, semble pouvoir être reproduit correctement par des modèles computationnels.

#### 6.4.1.2 Effet de la dispersion sur les stimuli utilisés

Les stimuli utilisés dans les Expériences 1 et 2 ont été soumis au modèle computationnel qui a permis l'explication des différences entre sons S+ et S- et entre sons asymétriques (Giguère et Woodland, 1994). Le modèle a été utilisé tel qu'implémenté au sein de la plateforme AIM (Patterson, Allerhand et Giguère, 1995), en incluant la modification du facteur de qualité suggérée par Carlyon (1996, 1997a). L'enveloppe des signaux présents dans chacun des canaux a été extraite par transformée de Hilbert, et sa valeur $rms$ calculée. La somme des valeurs $rms$ a ensuite été considérée. Cet indice, basé sur une simplification de l'hypothèse temporelle, permet de déterminer si la dispersion due au filtrage auditif affecte globalement les propriétés de l'enveloppe qui sont supposées déterminer la rugosité.

Aucun effet systématique de la phase sur la valeur $rms$ de l'enveloppe n'a pu être mis en évidence lors de la comparaison entre conditions de phase positive ou négative (Expérience 1) ou entre conditions "dent de scie" et "dent de scie inversée" (Expérience 2). Les différences observées étaient inférieures à 1 %, et des cas où la différence allait à l'encontre des résultats expérimentaux ont pu être observées.

Ce résultat numérique peut être interprété qualitativement. Pour ce qui est des sons pAM, l'absence d'influence de la dispersion s'explique par le faible nombre de composantes utilisées : un nombre de composantes supérieur à 5 a été trouvé

nécessaire pour observer les différences entre sons S+ et S- (Carlyon et Datta, 1997b). D'autre part, l'étendue fréquentielle des sons pAM est limitée et plus les composantes sont proches, moins les différences de phase introduites par la propagation sont marquées. Certains sons de l'Expérience 2 ont plus de 5 composantes, mais leur étendue fréquentielle reste limitée par rapport à leur fréquence centrale. Ceci est à rapprocher de l'argument similaire évoqué lors de la discussion des dérivations analytiques en section 5.4.3. L'hypothèse que nous soutenons est que dans le cas de sons à bande étroite, un filtrage à phase linéaire, à phase minimale ou un filtrage réaliste quant à la dispersion de phase au niveau de la membrane basilaire donnent un résultat quasiment équivalent (Goldstein, 1967b).

Insistons pour terminer sur le fait que l'asymétrie dans l'onde résultante est tout de même présente dans les résultats fournis par le modèle dispersif, et ceci pour les raisons de modification du rapport des amplitudes évoquées plus haut. Nos résultats ne sont donc en aucun cas contradictoires avec ce modèle et plus généralement avec les données indiquant la dispersion de phase du filtrage imposé par les premières étapes du traitement auditif. Simplement, l'examen de l'effet de cette dispersion indique qu'elle ne peut constituer une base suffisante pour interpréter les résultats des Expériences 1 et 2.

### 6.4.2 Asymétrie temporelle

L'Expérience 2 a employé des stimuli présentant une asymétrie temporelle. Les conséquences d'une telle asymétrie, sans particulièrement faire référence à la rugosité, ont été étudiées par Patterson (1994a,b). Les stimuli employés dans ces études étaient des sons purs modulés par des enveloppes constituées par la répétition d'exponentielles croissantes ou décroissantes. Les sons ainsi obtenus sont donc modulés périodiquement en amplitude de façon asymétrique : la modulation exponentielle peut être vue comme une "dent de scie" en coordonnées logarithmiques.

Il a été démontré que les conditions d'enveloppe avec une croissance lente et une décroissance abrupte *(ramped)* induisaient la perception d'une composante sinusoïdale prononcée à la fréquence porteuse, alors que les conditions de croissance abrupte et de décroissance lente *(damped)* induisaient cette perception de manière plus faible. Ce résultat a été obtenu avec différentes fréquences porteuses (de 400 à 4800 Hz) et différentes fréquences de modulation (de 20 à 120 Hz). L'interpré-

tation proposée par l'auteur est que la différence entre les conditions *ramped* et *damped* réside au niveau de la distribution des intervalles temporels à travers les canaux auditifs : les conditions *damped* font résonner un grand nombre de filtres, mais chacun à leur fréquence propre, alors que la condition *ramped* impose à ces filtres la fréquence de la porteuse du stimulus. L'effet de retournement temporel sur le timbre de ces sons modulés en amplitude de façon asymétrique s'explique donc par l'accumulation d'indices de structure temporelle fine entre les canaux auditifs. Une démonstration supplémentaire de ceci est que l'effet diminue avec les fréquences porteuses élevées, du fait de la perte de verrouillage de phase.

Une répartition différente des intervalles temporels entre les canaux ne semble pas expliquer les différences perceptives entre stimuli de l'Expérience 2. En effet, l'effet de phase observé se maintient (et augmente) pour les hautes fréquences, ce qui indique qu'il ne peut être uniquement basé sur des indices de structure temporelle fine. Notre hypothèse est que dans ce cas les auditeurs ont "écouté l'enveloppe" plutôt que la porteuse pour faire leurs jugements. Sans exclure la coexistence avec les indices de structure temporelle fine entre canaux pour les fréquences centrales graves, des indices intra-canaux au niveau de l'enveloppe d'amplitude semblent pouvoir fournir une base suffisante pour permettre d'établir les jugements de rugosité.

### 6.4.3 Effet des mécanismes d'adaptation

L'interprétation des données des Expériences 1 et 2 que nous proposons suggère que la forme de l'enveloppe d'amplitude temporelle après avoir pris en compte les effets du filtrage auditif influence la perception de rugosité. Bien entendu, ce filtrage est seulement une étape très précoce des traitements auditifs. Les effets de ce filtrage, et notamment l'asymétrie introduite, doivent être mis en perspective avec les traitements ultérieurs. Un niveau de représentation du stimulus acoustique particulièrement intéressant est celui de l'activité au niveau du nerf auditif, car cette étape constitue un passage obligé pour les informations auditives à partir duquel elles se projettent vers un grand nombre de zones cérébrales. De plus, la plupart des mécanismes susceptibles d'intervenir entre le stimulus acoustique et cette représentation commencent à être suffisamment connus pour être modélisés informatiquement (Meddis et Hewitt, 1991; Seneff, 1988; Patterson, Allerhand et Giguère, 1995).

La transformation principale des stimuli des Expériences 1 et 2 entre le filtrage auditif et l'activité au niveau du nerf auditif est due aux mécanismes d'adaptation, caractéristiques de la transduction au niveau des cellules ciliées internes. De tels mécanismes accentuent le début de chaque cycle de modulation, et pourraient donc transformer de manière radicale les asymétries soit introduites par le filtrage auditif (Expérience 1) soit présentes dans le stimulus acoustique lui-même (Expérience 2). Un modèle computationnel des traitements auditifs périphériques a été utilisé pour évaluer les effets de ces mécanismes d'adaptation (Meddis et Hewitt, 1991).

Le modèle choisi comporte une étape de préfiltrage large bande simulant l'action de l'oreille externe et moyenne, un banc de filtres gammatone reprenant les effets du filtrage auditif au niveau de la membrane basilaire, puis un modèle de cellules ciliées internes. Ce modèle a permis de reproduire de façon quantitative la plupart des effets des mécanismes d'adaptation observés physiologiquement (Meddis, 1986, 1988). Le modèle a été utilisé dans son implémentation la plus récente au sein de l'environnement DSAM (O'Mard et Meddis, communication personnelle). Les paramètres standard des différents modules ont été utilisés. Le premier pré-filtrage n'a que très peu d'importance sur le résultat des simulations. Le filtrage gammatone influence l'asymétrie introduite pour les stimuli de l'Expérience 1. Les paramètres du modèle de Meddis ont été choisis pour représenter l'activité d'une fibre à haut niveau de décharge spontanée (Meddis, 1988). La manipulation de ces paramètres modifient l'influence du niveau de présentation sur les résultats. Si aucune valeur raisonnable ne semble susceptible d'inverser les tendances observées dans les simulations qui vont être maintenant présentées, l'intensité des effets peut toutefois varier grandement. Ces simulations ne peuvent donc donner qu'une première idée des transformations subies par les stimuli après le filtrage auditif.

La Figure 6.3 représente le résultat d'une simulation pour des sons pAM de l'Expérience 1. Les conditions comparées sont $\phi = \pm\pi/3$ pour $(f_c, f_m) = (1000, 70)$ Hz. Cette simulation représente l'étape de traitement suivant la Figure 5.5 (qui avait servit à représenter l'asymétrie après filtrage auditif). L'examen de la forme de l'enveloppe indique une variation avec le signe de la phase. Pour $\phi > 0$, le début de chaque cycle de modulation présente un pic suivit par

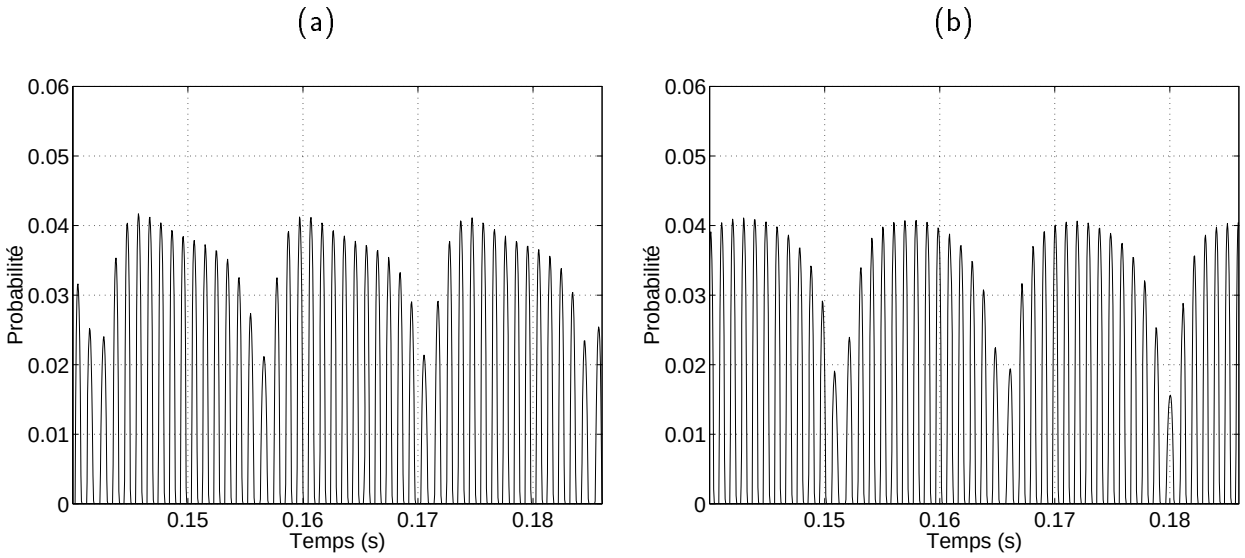


FIG. 6.3 – Simulation de probabilités de décharge au niveau du nerf auditif par le modèle de Meddis (1988). Sons pAM avec $(f_c, f_m) = (1000, 70\text{Hz})$, $\phi = +\pi/3$ (a) et $\phi = -\pi/3$ (b). La probabilité de décharge dans le canal auditif centré à 1064 Hz est représentée en fonction du temps. La faible valeur de probabilité est celle fournie par le modèle, mais rappelons qu'une cinquantaine de générateurs aléatoires sont généralement associés pour transformer les probabilités en train de décharges.

une décroissance de la fonction de probabilité de décharge. Pour $\phi < 0$, le pic de probabilité au début de chaque cycle de modulation est pratiquement absent.

Ces différences de forme se comprennent en considérant l'effet des mécanismes d'adaptation sur l'asymétrie introduite par le filtrage auditif : la condition $\phi > 0$ présentait une asymétrie de type croissance rapide et décroissance lente. La croissance abrupte fait que la probabilité de décharge au début de chaque cycle est forte, puis du fait à la fois de la décroissance du stimulus et de l'entrée en jeu des mécanismes d'adaptation rapide cette probabilité décroît rapidement. Pour ce qui est de $\phi < 0$, le début de chaque cycle est toujours accentué, mais l'adaptation qui s'ensuit est contrecarrée par l'augmentation graduelle d'intensité du stimulus : une sorte de "plateau" se forme après le pic. Vu que la différence de forme est due à l'asymétrie induite par le filtrage auditif, elle disparaît pour les conditions $\phi = \pm\pi/2$ ainsi que pour les fréquences centrales élevées.

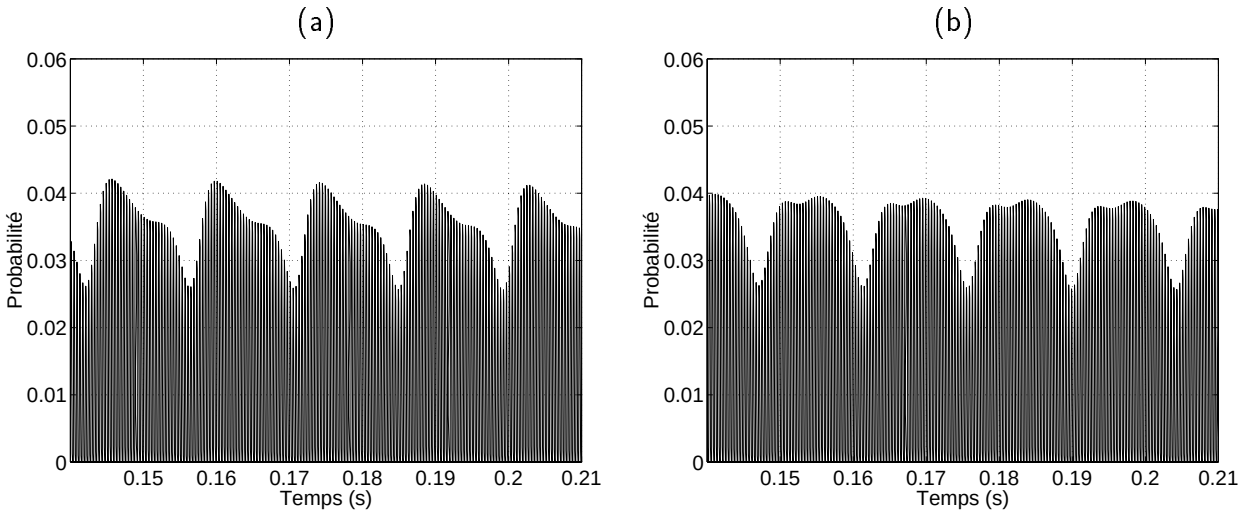


FIG. 6.4 – Simulation de probabilités de décharge à un niveau de traitement correspondant au nerf auditif par le modèle de Meddis (1988). Stimuli de l'Expérience 2 avec $(f_c, f_m) = (2500,\ 70\text{Hz})$, m = 0,8. Conditions "dent de scie inversée" (a) et "dent de scie" (b). La probabilité de décharge dans le canal auditif centré à 2500 Hz est représentée en fonction du temps.

Les mêmes simulations[1] ont été effectuées pour les stimuli de l'Expérience 2. Les résultats pour la condition $(f_c, f_m) = (2500, 70)$ Hz et $m = 0,8$ sont représentés en Figure 6.4. La différence de forme introduite par les conditions "dent de scie" et "dent de scie inversée" est très similaire à celle observée pour les conditions de phase positive et négative de l'Expérience 1. Le stimulus "dent de scie inversée" présente un pic au début de chaque cycle de modulation suivi par un décroissance rapide. La condition "dent de scie" a pour effet de remonter le niveau de probabilité de décharge du plateau suivant le pic. La même explication en termes de mécanismes d'adaptation peut être apportée.

Les différences de forme de l'enveloppe observées dans les Figures 6.3 et 6.4 ne s'accompagnent néanmoins pas de variations de la valeur $rms$ de l'enveloppe. Les simulations ont été systématiquement répétées pour toutes les conditions des Expériences 1 et 2 et les valeurs $rms$ des enveloppes présentes dans 32 canaux auditifs distribués de 100 à 10000 Hz extraites. Certaines variations de l'ordre de quelques pourcents sont apparues, mais sans corrélation apparente avec les résultats des expériences perceptives.

1. La région fréquentielle produisant une rugosité maximale a été déterminé par le modèle de l'Annexe B. Le filtre directement centré à 2500 Hz est celui qui produit ici le plus de rugosité.

### 6.4.4 Rugosité et perception d'enveloppe

Plusieurs hypothèses ont été examinées pour interpréter les résultats des Expériences 1 et 2. Reprenons donc ces différentes hypothèses en relation avec les théories perceptives de la rugosité.

La première hypothèse possible est celle de l'approche spectrale, à savoir que le spectre d'amplitude des stimuli acoustiques détermine la rugosité perçue. Les différences de rugosité entre les sons pAM de l'Expérience 1 ainsi que les différences entre une modulation en "dent de scie" et son retournement temporel dans l'Expérience 2 indiquent que cette hypothèse est clairement insuffisante.

Il est possible de raffiner cette hypothèse en prenant en compte la présence de produits de distorsion. L'influence de ces produits de distorsion, notamment du premier produit de distorsion cubique, ne peut être écartée de l'interprétation des résultats de l'Expérience 1, même si la stabilité des résultats et la disparition de l'effet aux fréquences centrales élevées sont des éléments allant plutôt à l'encontre d'une telle interprétation. Une influence systématique de ces sons de combinaison sur les stimuli de l'Expérience 2 est encore moins probable : les relations de phase entre n'importe quel groupe de trois composantes qui n'inclut pas la composante centrale ne changent pas entre les conditions "dent de scie" et "dent de scie inverse".

La troisième hypothèse évoquée est celle du lien entre la rugosité et le degré de modulation effective après les effets du filtrage auditif. Cette hypothèse permet d'expliquer l'influence globale de la valeur $rms$ de l'enveloppe sur les stimuli des Expériences 1 et 2, mais elle échoue à prédire les différences de rugosité entre conditions de phase positive et négative pour l'Expérience 1, et entre conditions "dent de scie" et "dent de scie inversée" pour l'Expérience 2. Les modèles temporels existants ne peuvent donc pas reproduire ces aspects de nos résultats.

Une quatrième hypothèse concerne un raffinement des modèles temporels qui prendrait en compte de façon plus réaliste les effets du filtrage auditif. Notamment, du fait de la dispersion au niveau de la membrane basilaire, les relations de phase entre composantes de stimuli ayant la même enveloppe acoustique pourraient conduire à un changement dans la valeur $rms$ de l'enveloppe après transmission le long de la membrane basilaire. Une telle amélioration ne peut qu'être pertinente pour les sons à large bande. Néanmoins, ces effets ne sont pas suffisants pour interpréter les différences perceptives entre pAM de phases de signes

opposés ou entre les sons "dent de scie" et "dent de scie inversée", du fait de la faible étendue fréquentielle des stimuli.

Un autre raffinement des modèles temporels consisterait à prendre en compte l'étape ultérieure du traitement auditif, à savoir les mécanismes d'adaptation liés à la transduction du signal au niveau des cellules ciliées internes. Les simulations informatiques dont nous disposons indiquent qu'il est effectivement possible de remarquer des différences de formes d'enveloppe là où des différences perceptives de rugosité ont été jugées, et que ces différences disparaissent là où les même rugosités ont été jugées. Ceci est valable pour toutes les conditions des Expériences 1 et 2. Toutefois, si l'on en croit les résultats du modèle de cellule ciliée utilisé, ces différences ne se résument pas à une variation de valeur $rms$ de l'enveloppe des probabilités de décharge.

L'hypothèse que nous proposons donc est que l'analyse de l'enveloppe après le filtrage auditif peut fournir la base à une explication unique de l'ensemble des données présentées au cours des Expériences 1 et 2. Cette analyse ne peut pas se faire en extrayant simplement une valeur $rms$ de l'enveloppe obtenue par filtrage du signal redressé, la méthode adoptée par les modèles existants. La rugosité dépendrait plutôt des fluctuations d'une "enveloppe perçue", dérivée de l'enveloppe acoustique en prenant en compte les effets du filtrage auditif, les effets d'adaptation, puis un mécanisme sensible à la forme de cette enveloppe. La nature d'un tel mécanisme reste à déterminer, et ouvre un ensemble de questions quant à la perception des modulations d'amplitude par le système auditif (Viemeister, 1979; Frisina, Smith et Chamberlain, 1990; Langner, 1992; Lorenzi, Micheyl et Berthommier, 1995).

## 6.5 Application possible aux sons musicaux

A la lumière de ces nouveaux résultats expérimentaux, l'estimation de rugosité de sons complexes en situation musicale peut être de nouveau abordée.

Les modèles spectraux sont généralement bien adaptés à la notation musicale. Une partition, où les évènements sont représentés par une liste de notes et de dynamiques, peut simplement être transformée en une liste de fréquences et d'amplitude. Ceci a été exploité notamment dans le cadre d'environnements fournissant aux compositeurs le calcul de divers indices psychoacoustiques à par-

tir d'une notation de type musical (Parncutt et Strasburger, 1994). Les modèles spectraux trouvent donc leur utilité dans le cadre de la prévision à partir d'une partition, les effets combinés de l'interprétation et de l'acoustique de la salle étant sans doute plus importants que les effets de phase observés ici. Ces modèles peuvent donc, moyennant un certain apprentissage quant à leur utilisation, fournir des résultats correspondant à une certaine réalité perceptive (Pressnitzer, McAdams, Winsberg et Fineberg, 1996).

Toutefois, ces modèles prédiraient exactement la même rugosité pour tous les stimuli testés pour une même condition $f_c$ dans l'Expérience 1 (soit une ligne horizontale pour tous les graphes de la Figure 5.2). Si l'on est intéressé par de plus subtiles différences entre sons enregistrés ou synthétisés, les modèles temporels sont donc mieux adaptés. Ces modèles reproduiraient, moyennant un ajustement de leurs paramètres, la dépendance globale de rugosité sur la valeur $rms$ de l'enveloppe.

Ces modèles échoueraient toutefois à reproduire les différences observées entre conditions de phase positives et négatives, ou entre sons "dent de scie" et "dent de scie inversée". Cet échec porte-t-il à conséquence? Il peut en effet être remarqué que dans des conditions naturelles d'écoute en champ libre, l'effet des différences de phase est fortement estompé par la propagation et les réflections multiples du son (Risset, 1988). Ainsi, une transposition directe des résultats concernant les effets de phase que nous avons étudiés sur la rugosité dans un contexte musical n'aurait que peu ou pas de sens. Toutefois, l'étude de l'influence des effets de phase dans une situation de laboratoire est un moyen efficace de gagner quelques données sur les mécanismes fondamentaux impliqués dans la perception de rugosité, ce qui pourrait devenir déterminant pour tenter de comprendre la rugosité de sons musicaux complexes. Notre proposition n'est pas que la phase en tant que telle doit être rajoutée à la liste (déjà bien longue) des paramètres acoustiques ayant une influence sur la rugosité, mais que seul un modèle basé sur la simulation du système auditif périphérique et inspiré des mécanismes mis ici en évidence sera capable d'inclure de manière naturelle toutes les dépendances de la rugosité sur divers paramètres acoustiques, dont la phase, mais aussi d'autres qui ne manqueront pas d'être exhibés et qui seront peut-être plus "résistants" aux conditions d'écoute naturelle.

## 6.6 Conclusion

L'influence de la forme de l'enveloppe sur la perception de sons complexes a été étudiée expérimentalement. Lors de la modulation de sons purs par une enveloppe asymétrique, les sons ayant une enveloppe présentant une croissance rapide et une décroissance lente sont jugés plus rugueux que les sons présentant une croissance lente et une décroissance rapide. Les résultats combinés des Expériences 1 et 2 étayent l'hypothèse que l'enveloppe temporelle après prise en compte des traitements auditifs périphériques que sont le filtrage en bandes critiques et les mécanismes d'adaptation détermine la rugosité. Toutefois, contrairement à ce qui est suggéré par les modèles existants, l'extraction de l'enveloppe par filtrage passe-bas ou passe-bande et l'estimation d'une valeur $rms$ ne suffisent pas à prédire la rugosité perçue. Un nouveau mécanisme sensible à la forme de l'enveloppe doit être recherché, et devrait être inclus dans les modèles susceptibles de prévoir la rugosité de sons quelconques, dont les sons musicaux.

# Chapitre 7

# Influence de la cohérence de modulation

*La rugosité produite par l'addition de deux sons purs modulés de façon irrégulière démontre un effet de cohérence d'enveloppe, dû à des comparaisons entre canaux auditifs.*

La perception de sons complexes suppose la combinaison d'informations à travers une large étendue fréquentielle. En ce qui concerne la perception de rugosité, plusieurs études suggèrent une influence de la cohérence des modulations présentes dans les canaux auditifs : plus cette cohérence est élevée et plus la combinaison de rugosité semble efficace (Mathes et Miller, 1947; Terhardt, 1974a). Cet effet peut être expliqué soit par une comparaison entre canaux, soit par l'addition destructive entre patterns d'excitation. Nous avons étudié l'addition de stimuli modulés en amplitude par une enveloppe irrégulière. L'interférence de ces stimuli est bien moins destructive que celle observée pour ceux habituellement utilisés. L'effet de cohérence d'enveloppe est de nouveau observé, sans qu'il puisse être attribué aux recouvrement entre patterns d'excitation.

## 7.1 Introduction

### 7.1.1 Divergences autour de l'effet de cohérence d'enveloppe

La plupart des sons de notre environnement, dont les sons musicaux, comportent de l'énergie dans une large bande de fréquence. Dans ce cas, le filtrage périphérique a en général pour effet d'introduire des modulations d'amplitude dans de nombreux canaux auditifs. Avant d'envisager de relier les résultats obtenus avec des sons simplifiés à la perception de rugosité de sons quelconques, une étude de la combinaison des rugosités éventuellement présentes dans plusieurs canaux auditifs semble donc pertinente.

Une revue de la littérature sur ce sujet a été proposée au Chapitre 3. Il est apparu que les données disponibles peuvent être vues selon deux modèles distincts. D'une part, un certain nombre de résultats concernant des sons à large bande ont été interprétés comme une simple addition de rugosité entre canaux auditifs (Fastl, 1977; Kemp, 1982). D'autre part, des expériences sur la rugosité produite par l'addition de deux sons SAM ont démontré un effet de la cohérence entre les enveloppes présentes dans des canaux auditifs éloignés (Mathes et Miller, 1947; Terhardt, 1974a).

Ce dernier résultat semble indiquer la présence d'un mécanisme de comparaison inter-canaux, qui selon le degré de cohérence entre enveloppes permet ou ne permet pas l'addition des rugosités partielles. Cette interprétation a néanmoins été contestée : reprenant les stimuli utilisés pour démontrer l'effet de cohérence, Aures (1985b) a démontré une dépendance de l'effet sur le niveau de présentation ainsi que sur l'écart fréquentiel entre les deux sons SAM. L'interprétation alternative qu'il propose est que les interférences destructives entre patterns d'excitation produits par les sons SAM sont en fait la cause de l'effet de cohérence d'enveloppe. Ainsi, la seule combinaison des rugosités partielles présentes à l'intérieur de chaque canal auditif permettrait de rendre compte de l'effet de cohérence observé. Cette interprétation est rendue possible par certaines particularités des stimuli employés.

### 7.1.2 Modulation effective des sons SAM co- et anti-phasiques

Les études de Mathes et Miller (1947), Terhardt (1974a), Vogel (1974) et Aures (1985b) ont employé comme stimuli l'addition de deux sons SAM. Le degré de cohérence entre enveloppes était varié en opposant les conditions co-phasiques et anti-phasiques (voir section 3.1.3.1). Le spectre d'amplitude de tels stimuli est présenté en Figure 7.1 pour le cas où la première fréquence porteuse est égale à 4000 Hz et où la deuxième fréquence porteuse est éloignée d'une demi-bande critique mesurée en ERB ($f_{c1} = 4000$ Hz, $f_{c2} = 4233$ Hz). Les deux fréquences de modulation sont identiques et égales à $f_m = 50$ Hz. Cet exemple est choisi car il correspond au cas où les deux sons SAM peuvent effectivement interagir malgré le filtrage auditif. L'activité présente dans la zone d'interaction ressemblerait à l'addition physique des ondes temporelles, que nous allons considérer.

La condition co-phasique est obtenue en choisissant toutes les phases à l'origine égales à 0. L'onde temporelle obtenue est présentée en Figure 7.2(a). Il est possible de remarquer plusieurs périodicités dans l'enveloppe résultante. La première périodicité est de durée égale à 20 ms. Elle correspond à la fréquence de modulation de chacun des sons SAM. Une deuxième périodicité plus rapide est imposée par les battements entre les deux fréquences porteuses (ici $4,3$ ms).

La condition anti-phasique est obtenue en retardant la modulation de la composante supérieure d'une demi-période, pour que les minima de l'enveloppe de l'un des sons SAM correspondent aux maxima de l'autre. Dans ce cas, l'onde temporelle est telle que représentée en Figure 7.2(b). Du fait des relations de phase entre composantes de modulation, la périodicité à 20 ms devient quasiment invisible, seule reste notable la périodicité rapide due aux battements entre porteuses.

Les théories temporelles relient la perception de rugosité au calcul d'une profondeur de modulation effective, pour laquelle les fréquences de modulation qui ne sont pas perçues comme provoquant de la rugosité sont atténuées. Même si les résultats des Expériences 1 et 2 indiquent que cette approche peut être insuffisante dans certains cas, ces mêmes expériences ont montré qu'elle constitue une première approximation généralement instructive. Le calcul de la profondeur de modulation effective peut se résumer ici à un filtrage passe-bas de l'onde temporelle rectifiée. Le résultat de ce filtrage est représenté par un trait pointillé sur la Figure 7.2. Il est manifeste que pour la condition co-phasique l'enveloppe ainsi

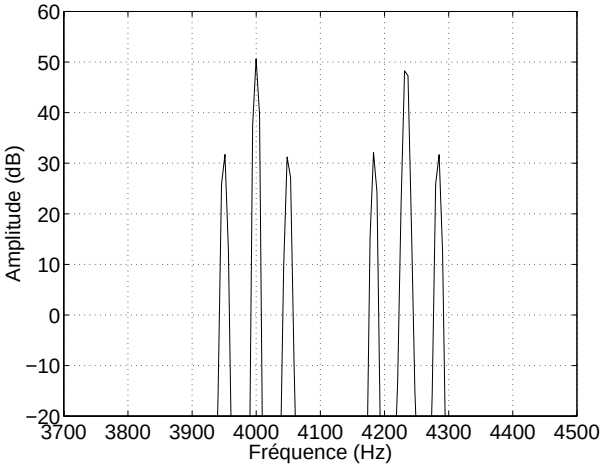


FIG. 7.1 – Composition spectrale de l'addition de deux sons SAM. Les interactions entre composantes se combinent pour définir l'enveloppe du signal.

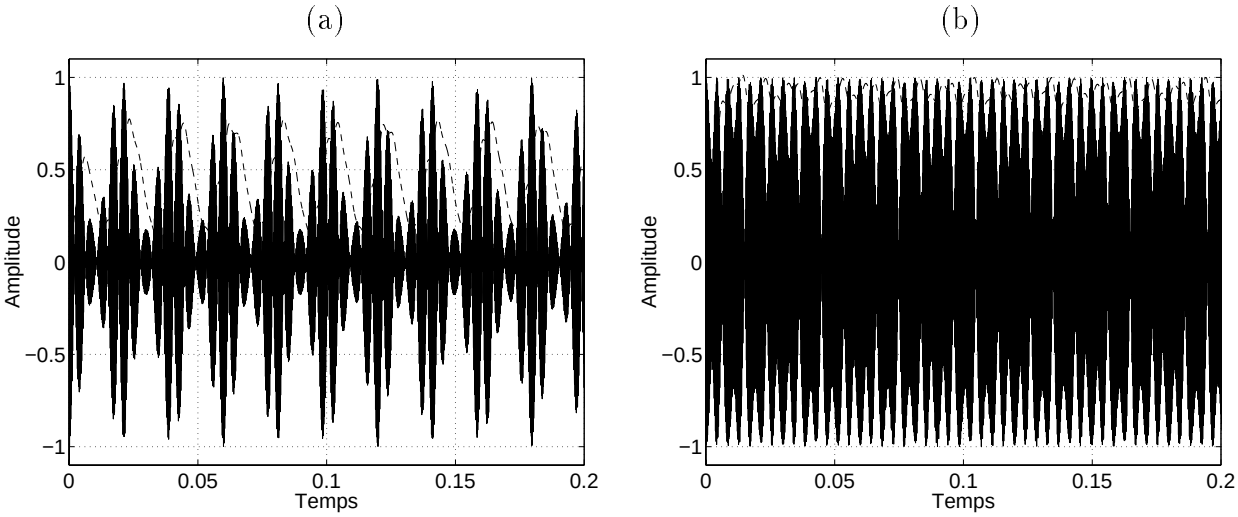


FIG. 7.2 – Onde temporelle produite par l'addition de deux sons SAM dans les conditions d'enveloppes co-phasique (a) et anti-phasique (b). Le trait pointillé représente une enveloppe extraite par filtrage passe-bas par un filtre de Butterworth d'ordre 2 (12 dB/oct) et de fréquence de coupure 70 Hz. La profondeur de modulation de chaque son SAM est de $m = 0,7$.

extraite possède une profondeur de modulation appréciable, alors qu'il n'en est pas de même pour la condition anti-phasique. La diminution de la valeur $rms$ de l'enveloppe est de l'ordre de 78 % entre les deux cas.

Ce calcul simplifié de profondeur de modulation effective est valable dans le cas particulier où les effets du filtrage auditif sont négligeables. Si les deux sons SAM sont plus éloignés en fréquence, le même raisonnement pourrait toutefois s'appliquer dans les canaux auditifs où les deux signaux interagissent. Ainsi, en plus d'introduire une différence de cohérence d'enveloppe entre canaux auditifs, les conditions co-phasiques et anti-phasiques provoquent donc des différences de profondeurs de modulation efficace à l'intérieur des canaux auditifs où les sons SAM interfèrent. Tout (Aures, 1985b) ou une partie (Terhardt, 1974a) de l'effet de cohérence d'enveloppe peut alors être attribué à ces diminutions de modulations efficaces. L'objet de l'Expérience 3 est de répliquer les études sur l'addition de sons purs modulés avec des stimuli pour lesquels les différences entres interférences constructives et destructives – inévitables – sont moins prononcées.

## 7.2 Modulation d'amplitude irrégulière

### 7.2.1 Sons jAM

Nous avons choisi d'étudier la rugosité produite par l'addition de sons purs modulés par une enveloppe dérivée d'un simple cosinus mais dont la périodicité est "bruitée". La méthode d'obtention de cette enveloppe consiste, à chaque cycle, à rallonger ou à raccourcir de quelques millisecondes la période. La construction du signal se fait en trois étapes que nous allons maintenant détailler.

Tout d'abord, la position temporelle $M_i$ des maxima de l'enveloppe est déterminée. Pour ce faire, une fréquence de modulation de référence $f_m$ est choisie, ainsi qu'un nombre $N$ de maxima. Les intervalles de temps $T_{ref}$ entre maxima successifs de la modulation sont, pour la fréquence de référence $f_m$ :

$$T_{ref} = 1/f_m \tag{7.1}$$

Les intervalles réels $T_i$ qui vont séparer les maxima $M_i, M_{i+1}$ sont obtenus par perturbation de $T_{ref}$ selon un tirage aléatoire, $U_{(0,1)}$, provenant d'une distribution uniforme sur l'intervalle $[0, 1]$. La perturbation est alternée pour chaque période :

tantôt l'intervalle est allongé, tantôt il est réduit. Un paramètre est introduit à ce niveau, la perturbation maximale possible est limitée à un certain pourcentage de la période de référence. Ce paramètre est noté $p_{max}$. On a alors :

$$T_i = T_{ref} + [(-1)^i p_{max}/f_m].U_{(0,1)} \quad i \in [1, N-1] \tag{7.2}$$

Il suffit alors d'intégrer les intervalles pour obtenir la position des maxima :

$$\begin{cases} M_1 = 0 \\ M_i = \sum_{j=1}^{i-1} T_i \quad i \in [2, N] \end{cases} \tag{7.3}$$

L'étape suivante consiste à construire une fonction d'enveloppe passant par tous ces maxima. Cette fonction a été définie par morceaux à partir d'une famille de fonctions $E_i(t)$, pour $i \in [0, N-2]$ :

$$t \in [M_i, M_{i+1}[ \quad E_i(t) = \cos(-2\pi.\frac{t-M_i}{M_{i+1}-M_i}) \tag{7.4}$$

Chaque fonction $E_i(t)$ est en fait une simple période de cosinus. Elle est égale à 1 quand $t$ est égal à $M_i$, décroît pour atteindre -1 quand $t$ est égal à $(M_{i+1}+M_i)/2$, puis croît et tend vers 1 quand $t$ tend vers $M_{i+1}$. La fonction $E(t)$ obtenue par la réunion des fonctions $E_i(t)$ est donc une fonction continue, définie sur $[0, M_{N-1}[$. Sa dérivée première est aussi continue, elle est en particulier égale à 0 pour tout les maxima $M_i$. Ses dérivées d'ordre supérieur, elles, ne sont pas continues.

Enfin, la dernière étape consiste à moduler un son pur à la fréquence $f_c$ par la fonction d'enveloppe $E(t)$ pour obtenir des sons que nous appellerons jAM (*jittered amplitude-modulated*).

Un exemple de son jAM est proposé en Figure 7.3. Les sons jAM ont plusieurs propriétés intéressantes en ce qui concerne la perception de rugosité. Tout d'abord, la valeur $rms$ de leur enveloppe est indépendante du tirage aléatoire et ne dépend que de la profondeur de modulation $m$. A la différence d'un bruit filtré de bande étroite, le facteur crête de l'onde temporelle est ici parfaitement contrôlé. L'enveloppe n'est certes pas périodique, mais la "période moyenne" de cette enveloppe est fixée par le choix de $f_m$. Cette période moyenne n'est pas due à une modulation lente au début du signal, par exemple, puis à une accélération de la modulation à la fin du signal : du fait de l'introduction des perturbations de manière alternée, la période oscille à chaque cycle autour de la période moyenne.

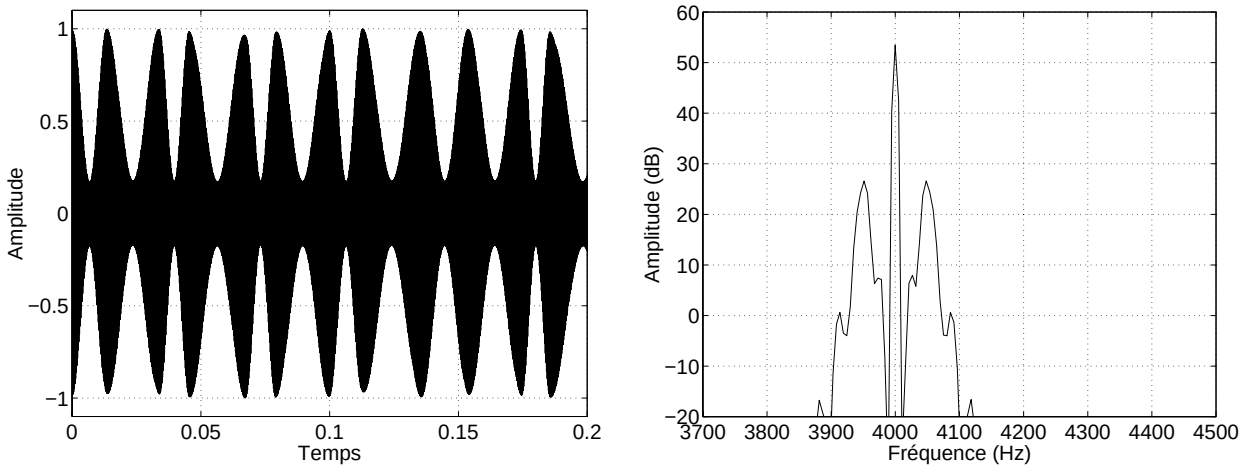


FIG. 7.3 – Onde temporelle et composition spectrale des sons jAM.

Le paramètre de perturbation maximale $p_{max}$ permet de borner les périodes possibles autour de la période de référence, par exemple pour limiter l'étendue fréquentielle du signal à une seule bande critique. A la limite, si $p_{max} = 0$, le son jAM n'est en fait qu'un son SAM à la fréquence porteuse $f_c$, modulé avec une profondeur $m$ à la fréquence $f_m$.

$$jAM(t) = [1 + m.E(t)]\cos(2\pi f_c t) \tag{7.5}$$

En ce qui concerne le but de ce chapitre, l'introduction d'un facteur aléatoire permet de générer deux jAM avec une même fréquence de modulation de référence, une même valeur $rms$ d'enveloppe, mais une faible cohérence entre enveloppes. Des valeurs d'autant plus faibles du coefficient d'intercorrélation entre enveloppes pourront être obtenues au fur et à mesure que $p_{max}$ augmente.

### 7.2.2 Expérience préliminaire

Avant d'envisager la rugosité de l'addition de sons jAM, il convient de s'interroger sur la rugosité qu'ils produisent isolément. Une écoute informelle nous a permis de faire l'hypothèse que l'introduction de la perturbation dans la modulation des sons jAM n'avait pas pour effet de réduire de façon dramatique leur rugosité. De manière à contrôler cette intuition, une expérience préliminaire a été réalisée.

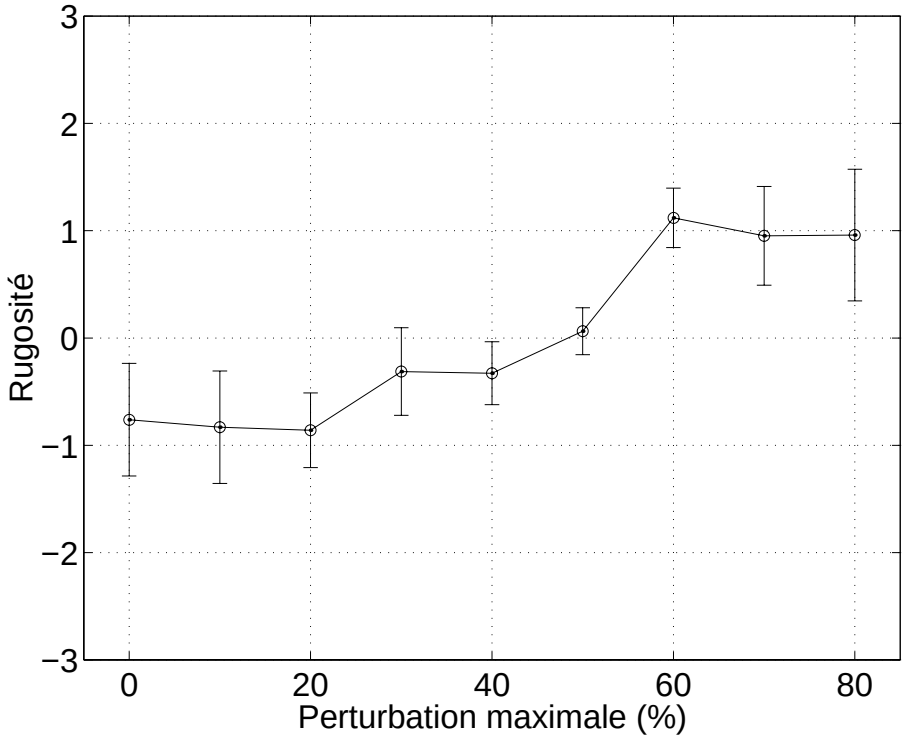


FIG. 7.4 – Résultats de l'expérience préliminaire. La rugosité de sons jAM de fréquence centrale $f_c = 4000$ Hz, de fréquence de modulation moyenne $f_m = 50$ Hz et de profondeur de modulation $m = 0,7$ est représentée en fonction du pourcentage de perturbation maximale $p_{max}$.

Un ensemble de sons jAM avec $(f_c, f_m) = (4000, 50$ Hz$)$ et $m = 0,7$ a été généré. Le paramètre $p_{max}$ contrôlant l'aspect aléatoire de l'enveloppe a été varié de 0 % à 80 %. Les stimuli de l'expérience ont été choisis parmi un ensemble de tirages aléatoires en imposant que la fréquence de modulation moyenne soit égale à 1 % près à $f_m$ et que la variances des périodes augmente de façon monotone avec $p_{max}$.

La méthode expérimentale désormais habituelle a été employée : des jugements de comparaison binaires de rugosité ont été effectués après l'écoute d'exemples constitués de sons SAM de profondeur de modulation variable. Les données ont été analysées par BTL et bootstrap. Six sujets ont participé à l'expérience.

Les résultats sont présentés en Figure 7.4. De faibles différences de rugosité sont observées entre les stimuli. La rugosité augmente globalement avec $p_{max}$, notamment pour les valeurs du paramètre supérieures ou égales à 60 %. Toutefois,

la taille des écarts-type fait qu'aucune différence observée entre deux stimuli considérés isolément n'est significative.

La conclusion que nous tirons de cette expérience préliminaire est que la rugosité des sons jAM est comparable ou légèrement supérieure à celle d'un son SAM de fréquence centrale et de fréquence de modulation équivalente (le point correspondant à $p_{max} = 0$ en Figure 7.4). Cette rugosité n'est pas négligeable, et nous pouvons utiliser les sons jAM pour étudier l'addition de rugosités partielles.

### 7.2.3 Addition des sons jAM

Considérons l'addition de deux sons jAM. Un situation similaire à celle présenté en section 7.1.2 peut être obtenue avec des sons jAM de paramètres $(f_{c1}, f_{m1}) = (4000, 50 \text{ Hz})$ et $(f_{c2}, f_{m2}) = (4233, 50 \text{ Hz})$, et une profondeur de modulation de $m = 0,7$. Comme les fréquences de modulation sont égales, l'enveloppe peut être calculée indifféremment pour les deux sons jAM. Si $p_{max}$ est non nul, par exemple égal à 50 %, deux tirages aléatoires permettent d'obtenir deux enveloppes $E_A(t)$ et $E_B(t)$ ayant un coefficient d'intercorrélation proche de 0.

Une condition corrélée équivalente à la condition co-phasique peut être obtenue en utilisant la même enveloppe $E_A(t)$ pour les deux sons jAM. A l'inverse, une condition non corrélée équivalente à la condition anti-phasique est obtenue en modulant le premier son jAM par $E_A(t)$ et le second par $E_B(t)$. Le résultat de ces additions est présenté en Figures 7.5 et 7.6.

La comparaison entre les Figures 7.6(a) et (b) indique que l'interférence entre sons jAM non corrélés n'annule pas entièrement les fluctuations d'enveloppe à la fréquence $f_m$. Notamment, si une profondeur de modulation effective est calculée de la même manière qu'en section 7.2, la réduction de valeur $rms$ observée n'est plus que de 30 %.

## 7.3 Expérience 3

### 7.3.1 Hypothèses expérimentales

Considérons l'addition de deux sons jAM tels que ceux décrits par les Figures 7.5 et 7.6. Leur fréquence de modulation moyenne est de $f_m = 50$ Hz. Malgré l'écart possible de période en période autour de la fréquence moyenne, ces

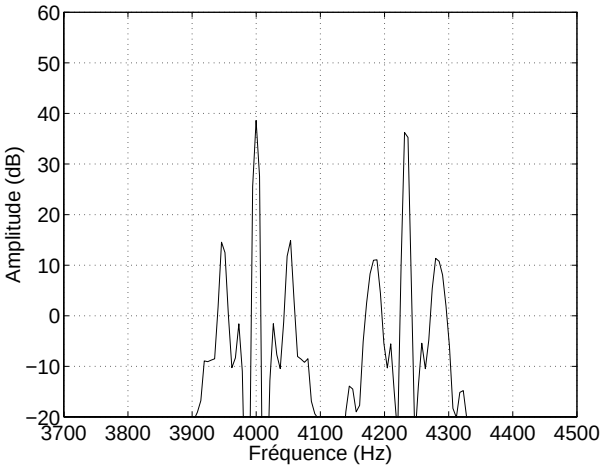


FIG. 7.5 – Composition spectrale de l'addition de deux sons jAM.

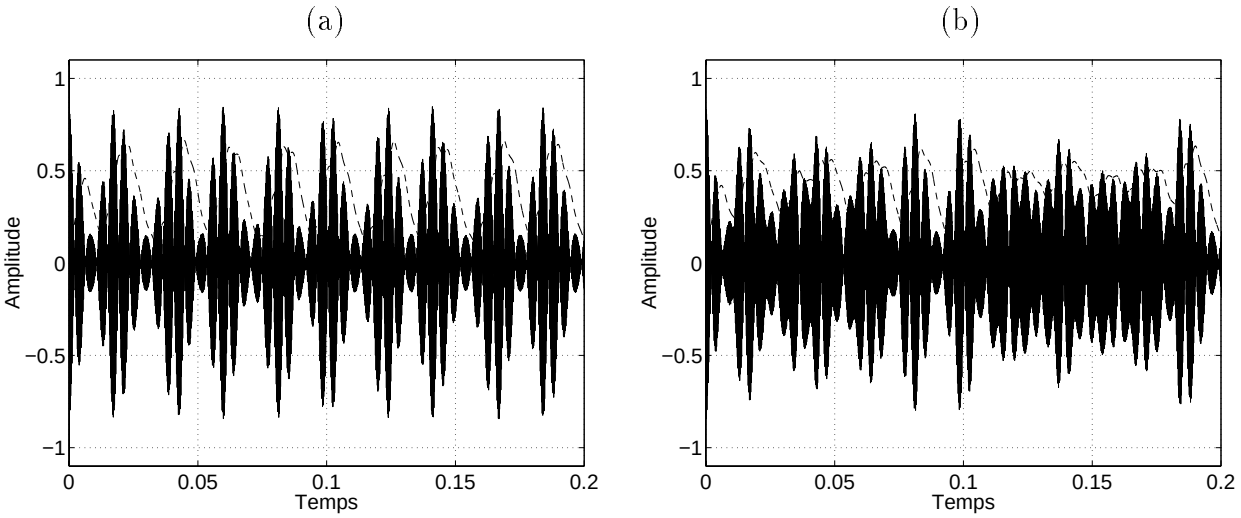


FIG. 7.6 – Onde temporelle produite par l'addition de deux sons jAM dans les conditions d'enveloppes corrélées (a) et non-corrélées (b). Le trait pointillé représente une enveloppe extraite par filtrage passe-bas à 70 Hz.

sons peuvent individuellement être considérés comme étant contenus à l'intérieur d'une même bande critique. En effet, la bande critique mesurée en ERB à 4000 Hz est de 456 Hz et cette bande s'élargit avec les fréquences croissantes. Il est possible de faire systématiquement varier l'écart fréquentiel entre les fréquences centrales $f_{c1}$ et $f_{c2}$. Au fur et à mesure que l'écart fréquentiel augmente, les interférences entre jAM diminuent jusqu'à éventuellement disparaître si les jAM sont traités dans des canaux auditifs distincts. En plus de varier cet écart fréquentiel, il est aussi possible d'opposer les conditions de modulations corrélées et non corrélées.

L'hypothèse d'addition des rugosités partielles prévoit que les seules différences observables entres conditions d'enveloppes corrélées ou non corrélées doivent provenir de la modification de la rugosité contenue dans les canaux où les sons jAM peuvent interagir. En particulier, l'effet de la cohérence d'enveloppe doit être le plus important pour les faibles séparations fréquentielles et diminuer jusqu'à disparaître pour les très larges séparations.

L'hypothèse de cohérence d'enveloppe prévoit à l'inverse une influence de la corrélation entre enveloppes même dans les conditions où les jAM sont résolus par le filtrage auditif. En revanche, ses prédictions quant au comportement de la combinaison de rugosité pour les faibles séparations fréquentielles ne sont pas explicites, du fait des problèmes liés à l'addition de sons SAM (Terhardt, 1974a).

## 7.3.2 Méthode

### 7.3.2.1 Sujets

Quinze auditeurs ont participé à l'expérience. Les auditeurs ont été recrutés de la même manière que pour les Expériences 1 et 2. Ils déclaraient n'avoir jamais eu aucun problème d'audition. La plupart d'entre eux ont été recrutés à partir d'une base de données de volontaires et n'avaient jamais participé à une expérience de psychoacoustique, mais certains avaient déjà participé à l'une de nos expériences.

### 7.3.2.2 Stimuli

Deux enveloppes de sons jAM ont été générées à partir de la procédure décrite en section 7.2.1. La fréquence de modulation moyenne choisie était de $f_m = 50$ Hz. La profondeur de modulation était de $0,7$. La perturbation maximale possible $p_{max}$ a été fixée à 50 %. Les tirages aléatoires ont été répétés jusqu'à obtenir

deux enveloppes dont la fréquence moyenne était de 50 Hz à 1 % près, dont les variances étaient similaires, et dont le coefficient de corrélation était faible (ici, $r = -0,04$).

Ces enveloppes ont été utilisées pour moduler des sons purs de fréquences variables. La fréquence centrale du son le plus grave a été fixée à 4000 Hz. La fréquence centrale du deuxième jAM a été variée de manière à représenter des écarts de $\Delta_{fc} = [0; 1/2; 1; 2; 4; 8]$ unités sur l'échelle des ERB ou *ERB-rate scale* (Moore et Glasberg, 1983; Glasberg et Moore, 1990). Une unité de cette échelle fournit une largeur rectangulaire équivalente à une bande critique calculée au centre de l'unité.

Six premiers stimuli ont été obtenus en modulant les deux fréquences porteuses par la même enveloppe. Ces stimuli représentent les conditions d'enveloppes corrélées. Six autres stimuli ont été obtenus en modulant la porteuse grave avec toujours la même enveloppe, et en modulant la porteuse aiguë par une enveloppe différente. Ces stimuli représentent les conditions non-corrélées. Les Figures 7.5 et 7.6 sont en fait un exemple de ces stimuli pour $\Delta_{fc} = 1/2$.

L'addition des sons jAM pour les autres écarts fréquentiels peut être déduite de ce même exemple. Aux faibles écarts, la plus grande partie de l'activité présente après filtrage auditif ressemblera aux Figures 7.6(a) et (b). Aux larges écarts, les sons jAM n'interagiront que très peu et après filtrage leur activité sera équivalente à deux sons jAM, tels que celui présenté en Figure 7.3, présents dans deux canaux indépendants. Aux écarts intermédiaires, une zone d'interaction sera entourée de deux sons jAM indépendants.

#### 7.3.2.3 Matériel

Les stimuli ont été générés numériquement avec une résolution de 16 bits et un taux d'échantillonage de 44.1 kHz. Le matériel de l'Expérience 1 a été utilisé (station d'informatique musicale NeXT, carte ISPW, convertisseurs Pro IO, amplificateur Canford, casque Sennheiser HD 520 II, cabine à double paroi Soluna S1). Tous les stimuli étaient normalisés en énergie avant présentation et le niveau mesuré par un sonomètre Bruel & Kjaer 2209 pour une séparation de $\Delta_{fc} = 1$ ERB était fixé à 60dBA. Ce niveau pouvait varier à travers les conditions, du fait du changement de composition spectrale des stimuli.

#### 7.3.2.4 Procédure

Une procédure similaire aux expériences précédentes a été employée. Les exemples habituels de sons SAM à profondeur de modulation variable ont été utilisés pour définir la notion de rugosité. Les 12 stimuli étaient alors présentés en ordre aléatoire selon toutes les paires possibles de sons non identiques dans un ordre comme dans l'autre. Pour chaque essai, les auditeurs devaient décider lequel des sons de la paire était le plus rugueux. Avant de répondre, ils pouvaient réécouter la paire autant que nécessaire. Les 132 jugements prenaient approximativement 30 mn.

#### 7.3.2.5 Analyses statistiques

La méthode d'analyse par BTL et bootstrap a été utilisée (Annexe A). Chaque point de l'échelle de rugosité perçue est ici le résultat de 330 jugements expérimentaux.

### 7.3.3 Résultats

Les résultats de l'Expérience 3 sont présentés en Figure 7.7. Globalement, le contraste entre enveloppes corrélées ou non a eu un effet marqué : les stimuli non-corrélés sont dans l'ensemble moins rugueux que les stimuli corrélés. Le pattern général d'évolution de la rugosité en fonction de l'écart fréquentiel est toutefois complexe. Il peut être divisé en trois parties.

- $\Delta_{fc} = 0$. La rugosité observée est assez faible par rapport aux autres conditions. Une différence significative existe entre enveloppes corrélées et non corrélées.

- $1/2 \leq \Delta_{fc} \leq 2$. Une forte rugosité est observée pour la condition $\Delta_{fc} = 1/2$, et ce pour les deux conditions qui ne présentent aucune différence significative. La rugosité décroît avec l'écart fréquentiel. La différence entre conditions augmente avec cet écart fréquentiel.

- $4 \leq \Delta_{fc} \leq 8$. La rugosité retrouve un niveau appréciable et reste relativement constante. Une différence est observée entre conditions, cette différence diminue avec $\Delta_{fc}$.

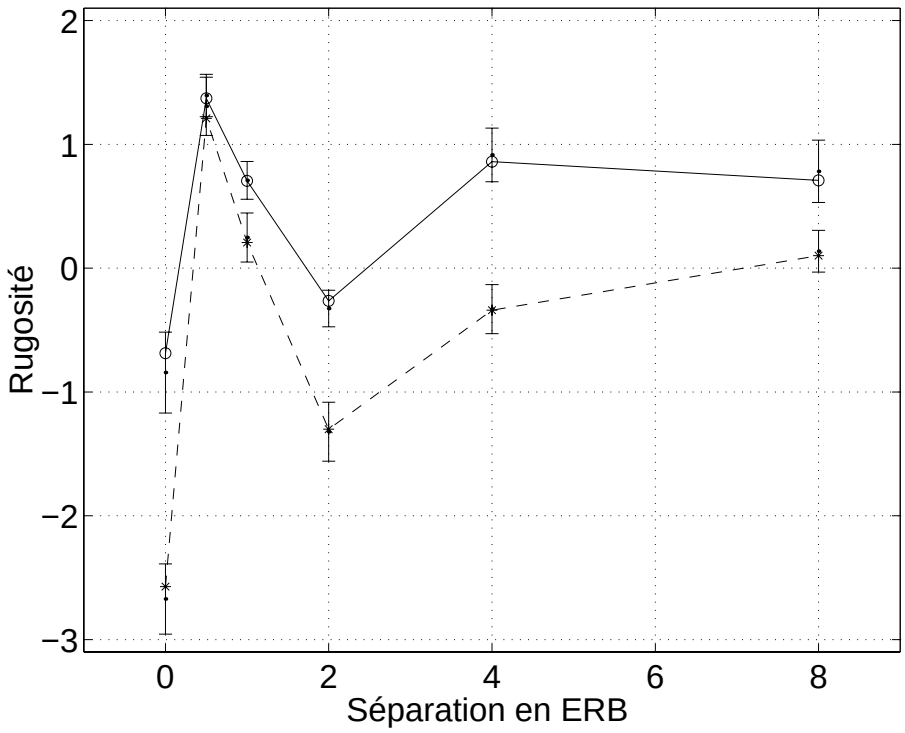


FIG. 7.7 – Résultats de l'Expérience 3. La rugosité perçue en ordonnée. L'écart fréquentiel mesuré sur l'échelle des ERB est en abscisse. Cet écart correspond à des fréquences de 4000, 4233, 4480, 5015, 6274 et 9772 Hz pour le son jAM aigu. Les traits pleins représentent les stimuli aux enveloppes corrélées, les traits pointillés les stimuli aux enveloppes non corrélées.

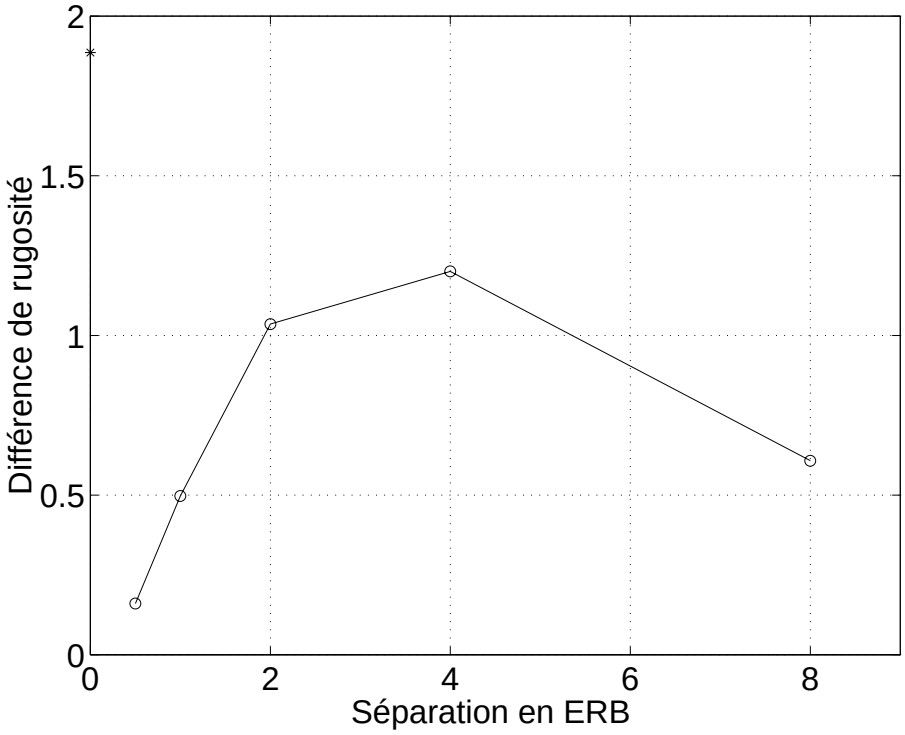


FIG. 7.8 – Effet de la cohérence d'enveloppe dans les résultats de l'Expérience 3. Les différences de rugosité entre conditions corrélées et non corrélées des données de la Figure 7.7 sont représentées en fonction de l'écart fréquentiel.

# 7.4 Discussion

## 7.4.1 Effet de la cohérence d'enveloppe

### 7.4.1.1 Caractérisation de l'effet

De manière à isoler l'effet de la cohérence d'enveloppe des autres aspects des données, les différences entre conditions corrélées et non corrélées de la Figure 7.7 ont été représentées en fonction de l'écart fréquentiel en Figure 7.8. Nous écartons pour l'instant de la discussion le point obtenu pour $\Delta_{fc} = 0$, qui représente le cas particulier où un seul son est présenté aux auditeurs, et qui sera traité en section 7.4.1.3.

L'effet d'enveloppe d'abord inexistant augmente rapidement avec la séparation fréquentielle, atteint son maximum pour une séparation de 4 ERB puis diminue lentement. Cet effet est significatif à $p < 0.05$ pour les écarts de 2 et 4 ERB, et à $p < 0.1$ pour l'écart de 8 ERB (selon les écarts-type de la Figure 7.7).

A la différence des résultats observés par les autres auteurs (Terhardt, 1974a; Vogel, 1974; Aures, 1985b), l'effet de cohérence est pratiquement absent pour les faibles séparations fréquentielles. Ceci indique que l'effet de cohérence que nous observons pour les larges séparations ne peut pas être dû aux recouvrements entre patterns d'excitation : en effet, quand ce recouvrement est maximal (faibles séparations), l'effet est absent, alors qu'il apparaît et se maintient pour de larges séparations qui ont vraisemblablement donné lieu à un recouvrement négligeable.

#### 7.4.1.2 Modalités de combinaison de rugosités partielles

Nous avons tenté de comparer quantitativement les prédictions faites par les différentes hypothèses émises pour la combinaison de rugosité. Pour ceci, un modèle d'estimation de rugosité a été employé. Ce modèle est décrit de façon détaillée en Annexe B. Il se base sur l'approche temporelle et est similaire dans son principe à celui de Aures (1985b). Un banc de filtres gammatone a servi à décomposer le signal en 36 canaux de fréquences centrales situées entre 2000 et 16000 Hz, de largeur 1 ERB, à raison de deux canaux par ERB. Les rugosités partielles ont été extraites par estimation de la valeur $rms$ des enveloppes efficaces présentes dans chacun des canaux (Annexe B).

La première hypothèse concernant la combinaison des rugosités partielles est celle de l'addition linéaire entre canaux (voir section 3). Dans ce cas, les seules différences possibles entre conditions corrélées et non corrélées sont celles dues aux interférences destructives à l'intérieur de chaque canal, aboutissant à une diminution de rugosité. La différence prévue entre conditions corrélées et non corrélées dans ce cas est proposée en Figure 7.9(a). Les effets des interférences, du fait du filtrage en bandes critiques, ne sont visibles que pour les écarts inférieurs à 2 ERB. Ceci est en désaccord complet avec l'effet observé.

Le modèle de Aures (1985b) comporte une étape d'intercorrélation entre canaux de manière à réduire la rugosité du bruit (section 3.2.3). Au moment de la sommation des rugosités partielles, une pondération obtenue grâce aux coefficients de corrélation avec les enveloppes des canaux adjacents est appliquée. En notant $r_i$ la rugosité partielle dans le canal $i$ et $c_{i,i+1}$ le coefficient de corrélation entre l'enveloppe présente dans le canal $i$ et celle présente dans le canal situé à une distance d'une bande critique, la rugosité totale $r$ s'obtient par (Daniel et Weber, 1997) :

$$r = \frac{1}{N} \sum_{i=2}^{N-1} (c_{i,i-1} c_{i,i+1})^2 r_i \tag{7.6}$$

Nous avons implémenté cette méthode de calcul à la suite des estimations de rugosités partielles. Le résultat des simulations est présenté en Figure 7.9(b). La disparition de l'effet de cohérence est plus progressive que dans le cas de l'addition simple. En effet, la prise en compte de la corrélation entre canaux adjacents réduit la rugosité pour deux sons jAM aux enveloppes incohérentes et de fréquences centrales proches. Les prévisions sont néanmoins toujours en désaccord avec les résultats, car cette méthode de calcul exclut l'existence d'un effet de cohérence pour les larges séparations.

Les résultats que nous avons obtenus semblent indiquer qu'une comparaison entre enveloppes est possible même lorsque les stimuli sont éloignés fréquentiellement. Une solution directe pour prendre en compte ceci consiste à considérer toutes les paires de canaux possibles et à effectuer l'addition de leurs rugosités partielles pondérées par la corrélation entre enveloppes effectives. Cette nouvelle méthode de calcul correspond à l'expression suivante :

$$r = \frac{1}{N(N-1)} \sum_{\substack{i,j=1 \\ j \geq i}}^{N} c_{i,j}(r_i + r_j) \tag{7.7}$$

Le résultat de cette méthode appliquée aux rugosités partielles obtenues par le modèle est présentée en Figure 7.9(c). La forme de la courbe est ici en bon accord avec les résultats expérimentaux. Cette méthode de calcul prévoit une influence de la cohérence maximale lorsque les stimuli sont partiellement résolus (entre 2 et 4 ERB) du fait des changements de corrélations entre canaux adjacents, et permet aussi de conserver une influence de la cohérence non négligeable lorsque les stimuli sont résolus, du fait de la non corrélation entre les deux régions rugueuses. La variation globale de l'effet selon l'écart fréquentiel est néanmoins sous-estimée.

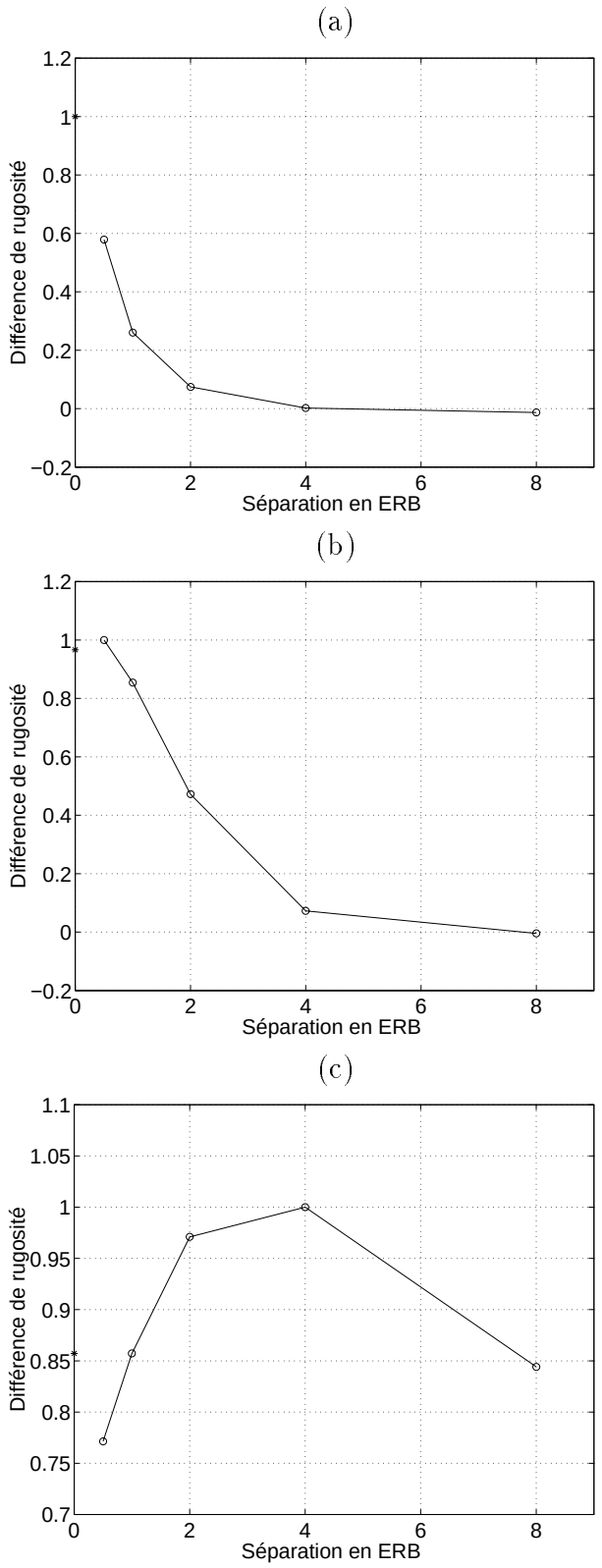


FIG. 7.9 – Estimation des méthodes de combinaison entre rugosités partielles (voir texte). Les différences sont normalisées, l'étoile représente l'effet prévu pour $\Delta_f = 0$. La somme est non pondérée (a), pondérée par le coefficient de corrélation entre canaux adjacents (b), pondérée par les corrélations entre tous les canaux (c). Cette dernière méthode est la seule en accord qualitatif avec l'effet de cohérence observé.

#### 7.4.1.3 Cas $\Delta_f = 0$

Le cas $\Delta_f = 0$ démontre un fort effet de cohérence d'enveloppe. Cet écart fréquentiel correspond au cas particulier où un seul son est présenté aux auditeurs. La comparaison entre conditions de cohérence est alors en fait la comparaison entre deux stimuli dont les valeurs $rms$ d'enveloppe varient, sans que la cohérence entre canaux ne soit affectée.

La Figure 7.9 indique que la simple somme entre valeurs $rms$ prédit effectivement un large effet. Cet effet est par contre sous-estimé par rapport aux autres séparations fréquentielles par la méthode de calcul que nous avons proposée. Ceci signifie sans doute que l'effet prévu par cette dernière méthode est généralement trop large : toutes les paires de canaux sont prises en compte et leurs effets simplement additionnés. Une légère incohérence, par exemple entre les extrémités du pattern d'excitation dans le cas $\Delta_f = 1/2$, produit déjà un fort effet par rapport à $\Delta_f = 0$. Une méthode intermédiaire serait sans doute souhaitable, qui prendrait en compte tous les canaux pour conserver la tendance observée en Figure 7.9(c) mais qui introduirait différentes pondérations pour "harmoniser" les prédictions faites pour un seul son et celle obtenues pour l'addition de plusieurs sons.

### 7.4.2 Evolution de la rugosité avec l'écart fréquentiel

L'évolution de la rugosité avec l'écart fréquentiel est relativement complexe. Nous interprétons les trois régions distinguées dans la description des résultats par la superposition de deux facteurs distincts.

Le premier facteur est lié au fait que l'apparition des rugosités partielles dépend du recouvrement ou non entre patterns d'excitations. Si deux sons rugueux sont additionnés à l'intérieur d'une même bande critique, leurs interférences donnent lieu à un nouveau signal avec une certaine enveloppe. La rugosité partielle dépend alors de cette enveloppe, elle est donc calculée *après* l'addition physique des stimuli. Au contraire, si les deux sons sont éloignés fréquentiellement, les rugosités partielles sont évaluées sur les enveloppes originales *puis* combinées (moyennant certains mécanismes dont la discussion est l'objet de cette expérience). L'addition de rugosités partielles dans deux canaux séparés serait alors plus efficace que la rugosité de l'addition des stimuli dans un même canal auditif (Kameoka et Kuriyagawa, 1969b). Une illustration de ceci est proposée

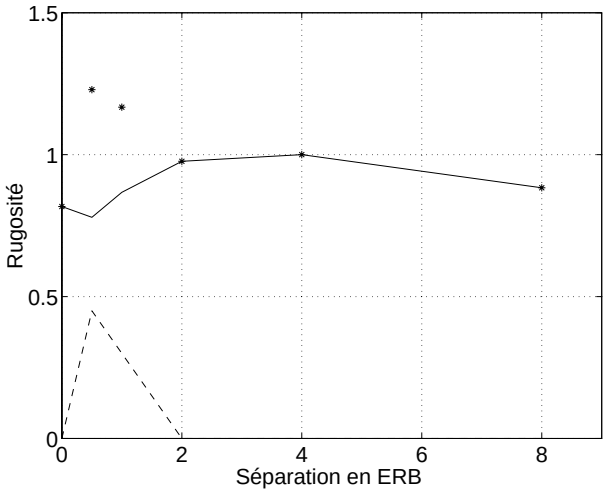


FIG. 7.10 – Interprétation de l'influence de $\Delta_f$ sur la rugosité globale. La rugosité calculée par la méthode d'addition après corrélation entre tous les canaux est proposée pour la condition d'enveloppe cohérente, en trait plein. Le facteur lié à la présence des battements supplémentaire est représenté en pointillés. Les étoiles représentent la simple addition de ces deux facteurs.

en Figure 7.10 pour la méthode de combinaison de rugosité avec corrélations entre tous les canaux, pour les stimuli avec enveloppes cohérentes. La rugosité augmente effectivement légèrement avec l'écart fréquentiel[1]. Ceci est à rapprocher des modalités de combinaison de sonie, où la sonie d'un bruit de puissance constante augmente avec sa largeur de bande si, et seulement si, celle-ci excède une bande critique (Fletcher, 1940; Zwicker et Scharf, 1965).

Le deuxième facteur est hypothétique et provient d'une écoute informelle des stimuli. Lorsque la séparation fréquentielle entre sons jAM est faible mais non nulle (1/2 ou 1 ERB), les battements entre leurs fréquences centrales viennent moduler les enveloppes obtenues (Figure 7.6 a). Ces battements sont théoriquement peu importants pour la rugosité, car ils sont de fréquence élevée et contribuent donc faiblement au calcul classique de la modulation effective. Néanmoins, il semble que ces battements aient produit une rugosité supplémentaire. La présence des patterns complexes qu'ils créent a l'intérieur de la modulation ont pu, selon nous, influencer la rugosité perçue, même si les fréquences de modulation impliquées ne produisent pas ou peu de rugosité si présentées isolément. Cette

1. Les trois méthodes donneraient ici des résultats similaires. La légère diminution observée pour $Delta_f = 8$ ERB est due à la diminution de rugosité du son aigu.

hypothèse, qui reste à vérifier par d'autres manipulations, est néanmoins à rapprocher des résultats des Expériences 1 et 2 sur l'influence de la forme de l'enveloppe. Ce deuxième facteur est représenté en Figure 7.10. Il est nul pour $\Delta_{fc} = 0$ car dans ce cas les battements supplémentaires sont absents, et disparaît lorsque les battements deviennent trop rapides ($\Delta_{fc} = 2$ ERB). La combinaison de ces deux facteurs permet d'interpréter qualitativement l'influence de l'écart fréquentiel sur la rugosité.

### 7.4.3 Comparaison d'enveloppes à travers les canaux auditifs

L'objet principal de cette expérience était d'estimer la présence ou non d'un mécanisme de comparaison à travers les canaux auditifs pour la perception de rugosité. Son existence, rendue probable par nos résultats, est en accord avec un certain nombre de données concernant la perception de modulations d'amplitude réparties dans plusieurs régions fréquentielles.

Richards (1987) a démontré que des auditeurs pouvaient distinguer entre la somme de deux sons purs modulés par un même bruit de bande étroite et la somme des deux mêmes sons purs modulés par deux bruits à bande étroite indépendants. Le modèle proposé pour expliquer ces résultats est basé sur la corrélation entre enveloppes.

Les expériences dites de démasquage par comodulation (*comodulation masking release* ou CMR) indiquent que la détectabilité d'un signal masqué par une bande de bruit modulée augmente si une deuxième bande de bruit modulée de façon cohérente est présentée dans une région fréquentielle éloignée (Hall et Grose, 1988). Plusieurs effets se combinent sans doute pour parvenir à ce démasquage. Néanmoins, le contrôle systématique de divers paramètres a montré qu'au moins une partie du CMR était due à une comparaison des enveloppes entre canaux auditifs (Schooneveldt et Moore, 1987).

Une expérience s'est spécifiquement penchée sur les modalités de traitement des modulations d'amplitude de sons SAM à travers les canaux auditifs (Yost et Sheft, 1989). En additionnant deux sons SAM, il a été démontré qu'il est possible de détecter une différence de phase entre les enveloppes pour différentes fréquences de modulation, fréquences centrales et profondeurs de modulation. Le seul paramètre influençant le seuil de détection est en fait l'écart fréquentiel entre

sons SAM. Les résultats peuvent être interprétées par un modèle de corrélation d'enveloppe uniquement dans les cas où les signaux sont traités indépendamment à travers des canaux auditifs distincts : Yost et Sheft notent les difficultés d'interprétation des données pour les faibles écarts entre SAM, du fait des battements supplémentaires. Ceci semble cohérent avec nos résultats.

Une deuxième expérience réalisée dans cette même étude démontre que la détection de modulation d'amplitude dans une région fréquentielle est affectée par la présence de modulation dans une région éloignée (*modulation detection interference* ou MDI). Ce résultat ne peut pas s'interpréter par la corrélation entre enveloppes, mais réclame plutôt un modèle du type de celui de Viemeister (1977) où l'extraction de la modulation d'amplitude est supposée réalisée après "l'agrégation" des informations provenant d'une large bande de canaux auditifs.

Il semble donc que non seulement un mécanisme de comparaison d'enveloppes entre canaux soit disponible pour l'audition, mais de plus que plusieurs stratégies soient possibles. Ceci est à rapprocher de l'importance de tels mécanismes pour la formation d'images auditives (McAdams, 1984). La cohérence de modulations d'amplitude de deux sons purs dans une zone fréquentielle qui correspond à la rugosité peut suffire à les faire fusionner en un seul percept (Bregman, Abramson, Doehring et Darwin, 1985), alors que leur incohérence peut induire une séparation du percept en deux images auditives (Bregman, Levitan et Liao, 1990). La capacité de détecter des similarités entres modulations présentes dans divers canaux est donc sans doute une adaptation à l'une des régularités de notre environnement : une telle cohérence trahit le fait que ces modulations ont de bonnes chances de provenir d'une même source acoustique. La méthode "crue" que nous avons proposée pour évaluer l'effet de cohérence dans le cadre de la perception de rugosité (section 7.4.1.2) représente donc sans doute plus une méthode de calcul, à raffiner et à valider avec un plus grand nombre de données, que le reflet du fonctionnement de ces mécanismes.

## 7.5 Conclusion

La rugosité de l'addition de sons modulés en amplitude par une enveloppe irrégulière a été étudiée. Les résultats confirment l'effet de cohérence d'enveloppes rapporté par certains auteurs (Mathes et Miller, 1947; Terhardt, 1974a). Cet effet

ne peut pas être interprété comme le résultat d'interactions à l'intérieur de chaque canal, un mécanisme de comparaison à travers les canaux auditifs doit être supposé. Cet effet de cohérence a pu être reproduit qualitativement en considérant les corrélations entre enveloppes des signaux présents dans tous les canaux auditifs. Les modalités précises d'addition de rugosité semblent néanmoins plus complexes : les résultats obtenus pour de faibles séparations fréquentielles indiquent que d'autres mécanismes, que nous pensons liés à la forme de l'enveloppe, entrent en jeu. Cette hypothèse reste à préciser, sans doute en combinaison avec les résultats des Expériences 1 et 2. L'effet de cohérence d'enveloppe observé rapproche la rugosité de problématiques plus générales, la comparaison de l'activité dans différents canaux étant par exemple un élément contribuant à l'organisation des scènes auditives.

# Chapitre 8

# Perception d'une tension musicale non tonale et sa relation avec la rugosité

*Des échelles de tension et de rugosité sont établies entre des timbres orchestraux. Les deux mesures sont corrélées sous certaines conditions.*

Parmi les sons complexes pouvant aboutir à la perception de rugosité, les sons musicaux occupent une place particulière. En effet, la rugosité a été proposée comme l'un des éléments participant à l'expression de dissonance et de tension dans le cadre de la musique tonale occidentale (Chapitre 4). Nous avons étudié la perception de tension musicale dans un contexte non familier pour la plupart des auditeurs, en examinant son lien avec la rugosité. Un ensemble d'accords extraits d'une pièce musicale non tonale ont été choisis et des jugements de tension et de rugosité ont été recueillis durant deux expériences. Différents degrés de ces attributs sont apparus entre les stimuli. La corrélation entre tension et rugosité semble être affectée par des "spécificités" qui apparaissent à l'analyse statistique des stratégies d'écoute et/ou de jugement. Une simulation à l'aide d'un modèle a été effectuée pour reproduire les échelles de rugosité. L'ensemble des résultats indique qu'une tension non tonale peut être perçue de façon stable sur la base de

différences de timbre et qu'elle est corrélée avec la rugosité, d'autant plus que la saillance d'autres attributs est réduite.

## 8.1 Introduction

### 8.1.1 Tension musicale dans un contexte tonal

L'impression d'une succession de moments de tension et de détente successifs est une expérience courante lorsque nous écoutons de la musique. L'étude de cette notion intuitive est en fait fondamentale pour la théorie musicale occidentale et elle a été source de multiples débats entre musiciens, musicologues et psychologues (Chapitre 4). Il n'existe pas de définition simple, universellement valide et largement acceptée de ce que signifient tension et détente dans un contexte musical. Pour rappeler la complexité de cette notion, considérons l'exemple élémentaire que constitue la cadence parfaite. Dans cette succession de deux évènements musicaux, un accord de septième de dominante basé sur le cinquième degré de la gamme majeure est considéré instable et dissonant, et appelle une résolution sur l'accord de tonique qui est stable et consonant. Il est généralement accepté que cette instabilité constitue une "tension", et sa résolution une "détente". La combinaison des deux constitue le motif bien connu qui peut être entendu comme la conclusion de la grande majorité des œuvres écrites aux 18ème et 19ème siècles. Malgré son caractère d'archétype, cet exemple illustre en fait plusieurs aspects de la notion de tension musicale.

La tension est supposée liée en partie à l'aspect intrinsèquement dissonant ou consonant des accords. L'accord de septième de dominante contient un triton, intervalle autrefois connu sous le nom de *diabolus in musica* – plutôt considéré dissonant, comme le lecteur l'aura deviné – et porte une certaine tension. L'accord de tonique contient seulement une quinte parfaite et une tierce majeure, deux intervalles consonants, et n'introduit peu ou pas de tension. Certaines théories basées explicitement sur la perception ont introduit le concept de rugosité pour expliquer une partie de cette consonance intrinsèque (von Helmholtz, 1877).

La tension musicale peut aussi être reliée à l'éveil d'une attente, la réalisation de cette attente amenant une détente, un retard dans sa réalisation accroissant la tension. De ce point de vue, la succession temporelle des deux accords de la cadence doit jouer un rôle. Les deux notes formant le triton dans l'accord de

septième sont instables et situées seulement à un demi-ton de la tonique et de la tierce de la triade majeure, qui elles sont stables. Ainsi, le triton a une tendance à se "résoudre" sur la tierce majeure (Meyer, 1956; Narmour, 1990).

L'éveil d'une telle attente pourrait aussi être liée à une habituation culturelle : le motif très connu de la cadence est reconnu par les auditeurs et la résolution est attendue. Au-delà de l'exemple simple de la cadence parfaite, il semble que les auditeurs de culture occidentale aient internalisé l'organisation hiérarchique de l'espace des hauteurs définie par les règles de la tonalité (Krumhansl, 1990; Bigand, 1993).

Dans le cadre tonal, de nombreux facteurs peuvent donc avoir une influence sur la perception de schémas de tension et de détente. Ces influences peuvent être convergentes, comme dans le cas de la cadence parfaite, ou contradictoires si le compositeur veut jouer avec des ambiguïtés structurelles. En tout état de cause, dans le cadre tonal un ensemble de règles établies et implicitement partagées par les auditeurs sont disponibles pour exprimer des mouvements de tension et de détente.

### 8.1.2 Contexte non tonal

Le problème est fort différent pour un compositeur ne souhaitant pas faire référence au système tonal. Prenons comme exemple la composition de la séquence qui a été choisie pour fournir la matière des Expériences 4 et 5 (la partition de cet extrait est proposée en Figure 8.1).

Des analyses spectrales de notes produites par un instrument acoustique ont été réalisées : dans ce cas, une contrebasse jouant avec de rapides changements de vibrato, de vitesse et de pression d'archet. Des ensembles de quatre à cinq partiels, sélectionnés à partir des partiels de plus forte amplitude à l'intérieur d'une certaine bande de fréquence, ont été isolés pour servir de "réservoir" de matériel musical. Huit accords, extraits de ce réservoir, ont été choisis par le compositeur puis agencés temporellement de manière à évoquer un contour de tension et de détente qui était désiré pour des raisons musicales. Les accords, en tant que résultats bruts d'une analyse acoustique, n'avaient aucune raison d'être conformes aux règles de l'harmonie tonale. De fait, il est impossible de leur trouver une fonction harmonique classique. Le compositeur a donc défini pour chacun d'eux une valeur de tension intrinsèque, par le biais d'un algorithme computationnel. Le

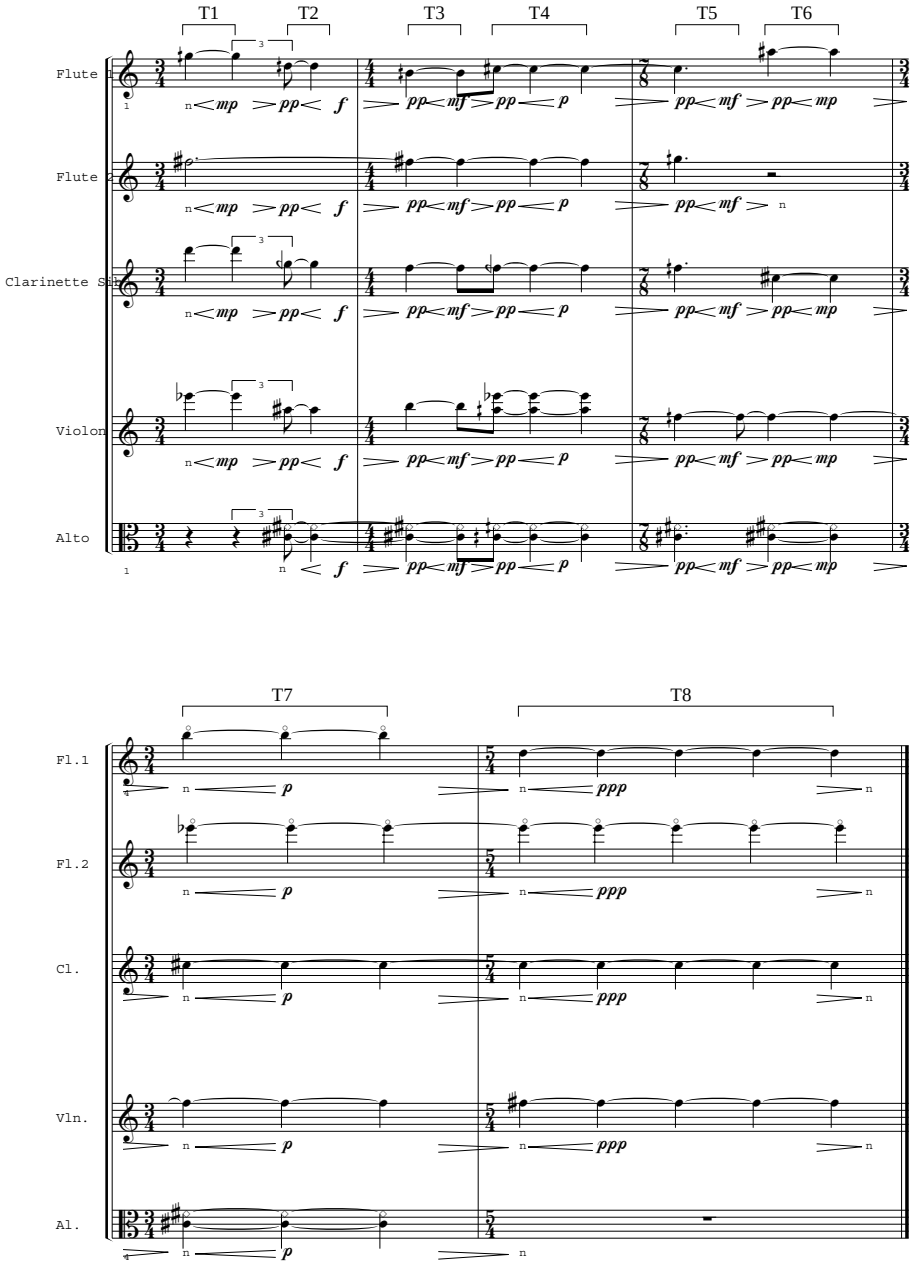


FIG. 8.1 – Partition de l'extrait de *Streamlines* (J. Fineberg) étudié dans les Expériences 4 et 5. Les accords indiqués en haut de la portée ($T_1$, etc.) ont été enregistrés individuellement, avec la même durée et le même profil dynamique. Ils constituent les huit "timbres orchestraux" utilisés comme stimuli.

plus grand commun diviseur des fréquences constituant chaque accord, moyennant une certaine tolérance (fixée à un seizième de ton) a été considéré comme une mesure qualitative de tension : plus bas était ce commun diviseur, plus tendu était supposé l'accord. La Figure 8.2 montre le résultat de cet algorithme appliqué aux huit accords. La forme de cette courbe a ensuite été utilisée dans la suite du processus de composition et son contour a été renforcé par d'autres paramètres musicaux.

Une telle approche soulève de nombreuses questions : premièrement, la notion de tension a-t-elle encore un sens pour un matériau musical qui ne possède aucune fonction tonale? Deuxièmement, dans la partition les mouvements sont soulignés par des différences de rythme et de dynamique, mais l'algorithme de tension utilisé est purement spectral et ne prend pas en compte ces indices supplémentaires. Pour être cohérent, les différences de timbre[1] entre accords devraient à elles seules être capables de transmettre les mouvements de tension prévus. Le premier but de cette étude était donc d'établir si des différences de tension, non basées sur des règles liées à la tonalité ou sur des indices de rythme ou de dynamique, peuvent effectivement être perçues par les auditeurs pour les "timbres orchestraux" choisis. Dans un contexte aussi peu familier pour eux, les auditeurs devraient alors utiliser des attributs psychoacoustiques de base – en attendant que les règles spécifiques à ce style soient éventuellement abstraites et apprises.

### 8.1.3 Rugosité et consonance musicale

La rugosité pourrait constituer un tel attribut psychoacoustique. D'une part, cet attribut semble être une dimension élémentaire de la perception auditive (voir la revue bibliographique du Chapitre 2). D'autre part, la rugosité relative des intervalles musicaux habituels a été démontrée corrélée à leur consonance (voir Chapitre 4). Cette corrélation a été interprété par les coïncidences entre partiels des sons harmoniques utilisés pour former les intervalles. Ce raisonnement a été étendu à des sons non harmoniques, et il a été vérifié que la relation entre rapports de fréquences et consonance perçue dépend effectivement du spectre des sons employés (Slaymaker, 1970; Geary, 1980). La possibilité d'expression de mouvements de tension et détente basés sur des différences de rugosité a aussi été envisagée (Mathews, Pierce et Roberts, 1987) : la rugosité est alors

1. Ces différences de timbre doivent de plus ici être visibles dans le spectre à long terme.

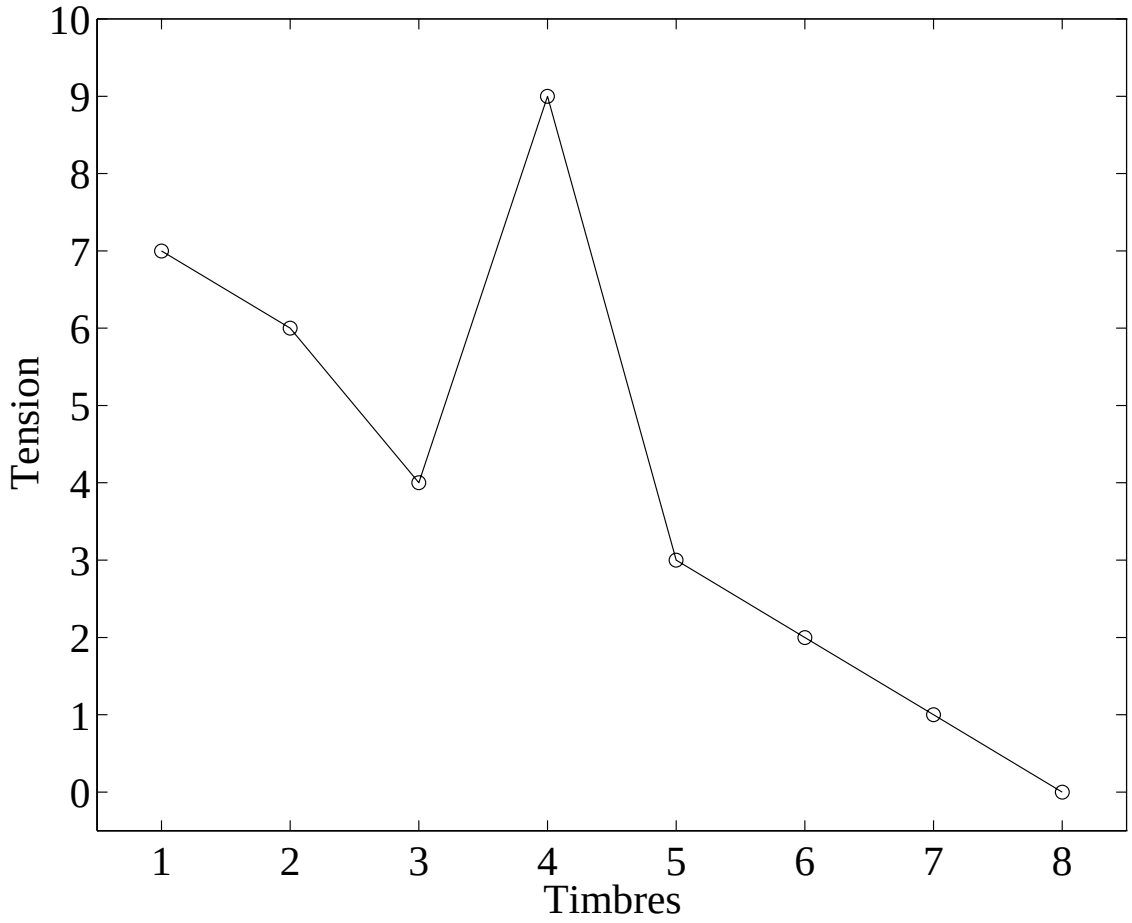


FIG. 8.2 – L'échelle de tension calculée par l'algorithme du compositeur. Les timbres ($T_1$, etc.) sont en abscisse dans un ordre correspondant à leur apparition dans la partition. L'échelle le long de l'abscisse n'est pas continue, les lignes entre les points ont été rajoutées pour révéler le profil de tension au cours du temps. L'échelle de tension est arbitraire et a été utilisée par le compositeur pour ordonner les timbres.

considérée comme une base sensorielle susceptible d'être utilisée pour imaginer des harmonies non tonales.

La plupart des résultats expérimentaux sur lesquels se basent les discussion du rôle de la rugosité au sein de l'écoute musicale ont été obtenus avec des stimuli synthétiques, et les conditions expérimentales employées pour les recueillir (la plupart du temps des jugements absolus de "consonance" ou d'"agrément") peuvent être critiquables dans un contexte musical réel (voir section 4.5.4). La contribution de la rugosité à la perception de tension dans un contexte musical a néanmoins été abordée dans une étude récente concernant de courtes séquences d'accords (Bigand, Parncutt et Lerdahl, 1996). La rugosité, avec d'autres facteurs tels que l'acculturation des auditeurs à la tonalité, a pris une part significative dans les jugements de tension. Malheureusement, dans cette étude la rugosité a seulement été estimée par un modèle informatique et non directement jugée. En plus d'étudier directement la perception de tension non tonale, le second but de ces expériences est de rassembler des jugements expérimentaux de rugosité pour des sons musicaux complexes. L'examen de la contribution de la rugosité à la perception de tension pourra alors être réexaminée.

## 8.2 Expérience 4

### 8.2.1 Méthode

#### 8.2.1.1 Stimuli

Une séquence de huit accords a été extraite de *Streamlines* (J. Fineberg, 1995). La partition de cette séquence, réorchestrée par le compositeur pour les besoins de l'expérience, est proposée en Figure 8.1. Les accords ont été joués par un orchestre de chambre (deux flûtes, une clarinette en Si bémol, un violon et un alto) composé de membres de l'ensemble *Itinéraire*. Ces accords, partageant un registre homogène, ont été enregistrés séparément avec une même durée et un même profil de dynamique (*dal niente* - *mf* - *dal niente*, soit croissant puis décroissant) pour une durée globale d'approximativement 10 s. Un exemple des caractéristiques spectrales et temporelles de l'un de ces stimuli est proposé en Figure 8.3. Les stimuli différaient donc principalement par leurs caractéristiques timbrales, et nous y feront désormais référence comme à des "timbres orchestraux".

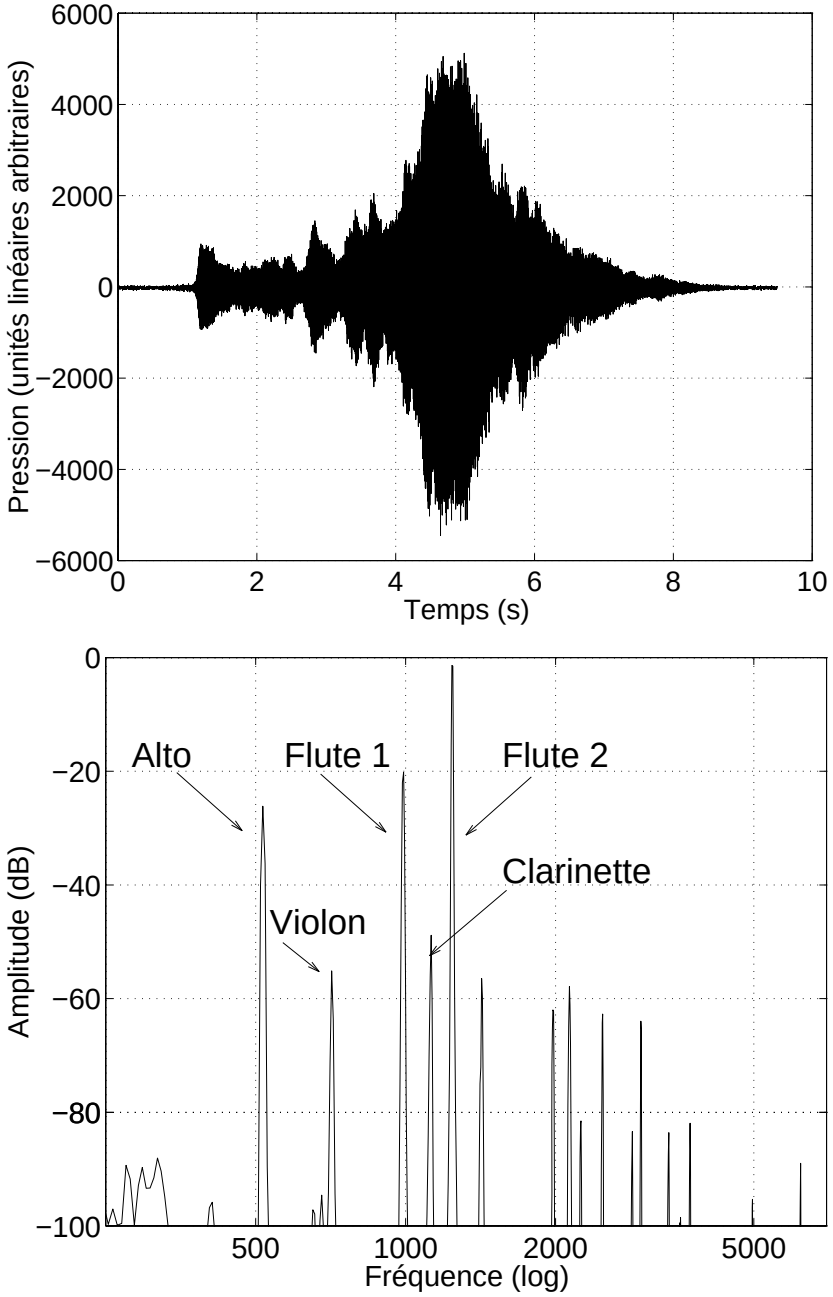


FIG. 8.3 – Onde temporelle et spectre d'amplitude (obtenu par méthode de Welch) du timbre orchestral $T_7$. Les unités de pression ainsi que la référence de l'échelle de décibels sont arbitraires. Les composantes spectrales correspondant aux fréquences fondamentales des différents instruments sont indiquées.

#### 8.2.1.2 Matériel

L'enregistrement s'est déroulé dans la salle de concert de l'IRCAM (l'Espace de Projection). La réverbération variable de cette salle était réglée à une valeur moyenne. Une paire de microphones Schoeps a été placée à 5 m des instrumentistes avec une élévation de 3 m. Les signaux stéréo ont été transférés sans modification à travers une table de mixage Neev V sur un enregistreur DAT Sony 7050. Ils ont ensuite été transférés et stockés numériquement sur disque dur à une fréquence d'échantillonage de 44.1 kHz et à une résolution de 16 bits à travers une interface ProTools Digidesign.

L'expérience elle même utilisait une cabine audiométrique à double paroi Soluna S1. Les sons étaient reproduits à partir d'une station d'informatique musicale NeXT à travers une carte de traitement de signal ISPW (Lindemann, Dechelle, Smith et Starkier, 1991), des convertisseurs numérique-analogique ProIO et un amplificateur de puissance Canford, et enfin un casque ouvert AKG K1000. Le niveau mesuré pendant la présentation des stimuli était approximativement de 80 dBA, ajustable à la demande des auditeurs pour permettre une situation d'écoute aussi confortable que possible. L'expérience était conduite par un programme écrit dans l'environnement PsiExp sur la station NeXT (Smith, 1995).

#### 8.2.1.3 Sujets

Le groupe expérimental était constitué de 30 sujets et ne comportait pas les auteurs. Les sujets étaient âgés de 17 à 45 ans ($M = 26$). Il provenaient d'horizons divers, certains travaillant à l'Ircam et d'autres recrutés par l'intermédiaire d'une base de donnée. Un effort particulier a été fait pour sélectionner des auditeurs aux habitudes musicales les plus diversifiées possibles. Le groupe comportait neuf musiciens professionnels (compositeurs ou instrumentistes ayant gagné de l'argent en faisant de la musique pendant l'année écoulée), huit musiciens amateurs (instrumentistes amateurs) et treize auditeurs non musiciens (n'ayant reçu aucune éducation musicale). Tous étaient payés pour leur participation[2]. Un auditeur non musicien n'a pas terminé l'expérience et a donc été retiré des résultats.

2. Ils devenaient ainsi auditeurs professionnels pour l'année à venir suivant nos critères...

#### 8.2.1.4 Procédure

L'expérience comportait deux parties, l'une concernant la tension l'autre concernant la rugosité. Durant la première partie, les huit timbres orchestraux étaient présentés par paires en ordre aléatoire et dans toutes les combinaisons possibles sans répétition (56 paires). Pour chaque paire, les auditeurs devaient faire un jugement de comparaison à choix forcé. La question à laquelle ils devaient répondre était : *"Entre les deux sons de cette paire, percevez-vous un mouvement de type "tension vers détente" ou plutôt "détente vers tension"*. Ils étaient explicitement incités à définir les notions de tension et de détente selon les critères qui leur paraissaient appropriés. Après trois à cinq essais de familiarisation sans renforcement, cette première partie durait approximativement 30 min.

Dans la deuxième partie, l'attention des auditeurs était focalisée sur un attribut particulier du son, la "rugosité". La dimension de rugosité était introduite par un exemple sonore caractéristique – un son pur de 1000 Hz modulé à 70 Hz dont la profondeur de modulation pouvait être manipulée à volonté par les sujets. Cet exemple ne produit aucune rugosité si la profondeur de modulation est de 0 % et la rugosité maximale qu'il est possible d'obtenir avec un seul son pur sinusoïdalement modulé en amplitude si cette profondeur de modulation est 100 % (Zwicker et Fastl, 1990). Les sujets étaient informés que la différence de qualité qu'il pouvaient entendre en manipulant l'interface graphique était une différence de rugosité, toutes choses égales par ailleurs. Une attention particulière a été portée à ne pas associer dans les instructions la rugosité avec un éventuel aspect "dissonant" ou "désagréable" des stimuli. Les 56 paires étaient en suite de nouveau présentées en ordre aléatoire et les sujets devaient faire une comparaison à choix forcé pour décider quel son était le plus rugueux.

Les deux parties de l'expérience étaient toujours effectuées dans cet ordre, jugements de tension puis jugements de rugosité, de manière à éviter tout biais dans les jugements de tension dû à une connaissance préalable de la notion de rugosité.

#### 8.2.1.5 Méthode d'analyse

Sur la base des jugements de comparaison, les huit timbres ont été arrangés sur une échelle linéaire autorisant une comparaison avec les prédictions du com-

positeur. La méthode habituelle d'analyse des résultats a été employée. Chaque point obtenu est le résultat de $7 * 2 * 29 = 406$ comparaisons expérimentales.

Le test de l'hypothèse nulle entre deux valeurs BTL est fait en comparant les intervalles de confiance attachées à chacune d'entre elle. Nous adopterons le critère statistique de significativité au risque $p < 0.05$ (voir Annexe A).

### 8.2.2 Résultats

Les résultats sont donnés en Figure 8.4. Considérons d'abord *l'échelle de tension*. La plus forte valeur de tension est obtenue par le timbre $T_1$ alors que la valeur la plus faible est produite par $T_5$. La différence entre ces valeurs extrêmes est supérieure d'un ordre de grandeur (dix fois) aux écarts-types attachés aux jugements. En utilisant le critère statistique basé sur les intervalles de confiance, la différence entre les extrêmes de l'échelle est donc significative. Ceci indique que tous les stimuli n'ont pas été jugés aléatoirement comme également "tendus". Un examen plus approfondi de l'échelle de tension indique que deux larges groupes de tension ont été perçus, l'un comportant les timbres $T_1, T_2, T_4, T_7$ comme étant plus tendus, et l'autre incluant les timbres $T_3, T_5, T_6, T_8$ comme étant moins tendus. Les différences à l'intérieur d'un groupe ne sont pas significatives, alors que les différences entre n'importe lesquels des timbres à travers les groupes sont significatives.

*L'échelle de rugosité* peut être analysée de manière similaire. La différence entre le timbre jugé le plus rugueux $T_2$ et le timbre jugé le moins rugueux $T_6$ est supérieure d'un ordre de grandeur aux écarts-type. Des différences significatives de rugosité ont donc été perçues entre sons orchestraux complexes. L'examen détaillé indique que les valeurs de rugosité sont plus régulièrement distribuées à travers les timbres que ne l'étaient les valeurs de tension.

La comparaison entre les deux échelles indique qu'elles diffèrent pour plusieurs timbres. Les différences sont significatives pour $T_3, T_5, T_6, T_7$. Spécifiquement, $T_6$ et $T_7$ sont jugés plus tendus que rugueux, alors que $T_3$ et $T_5$ sont jugés comme plus rugueux que tendus.

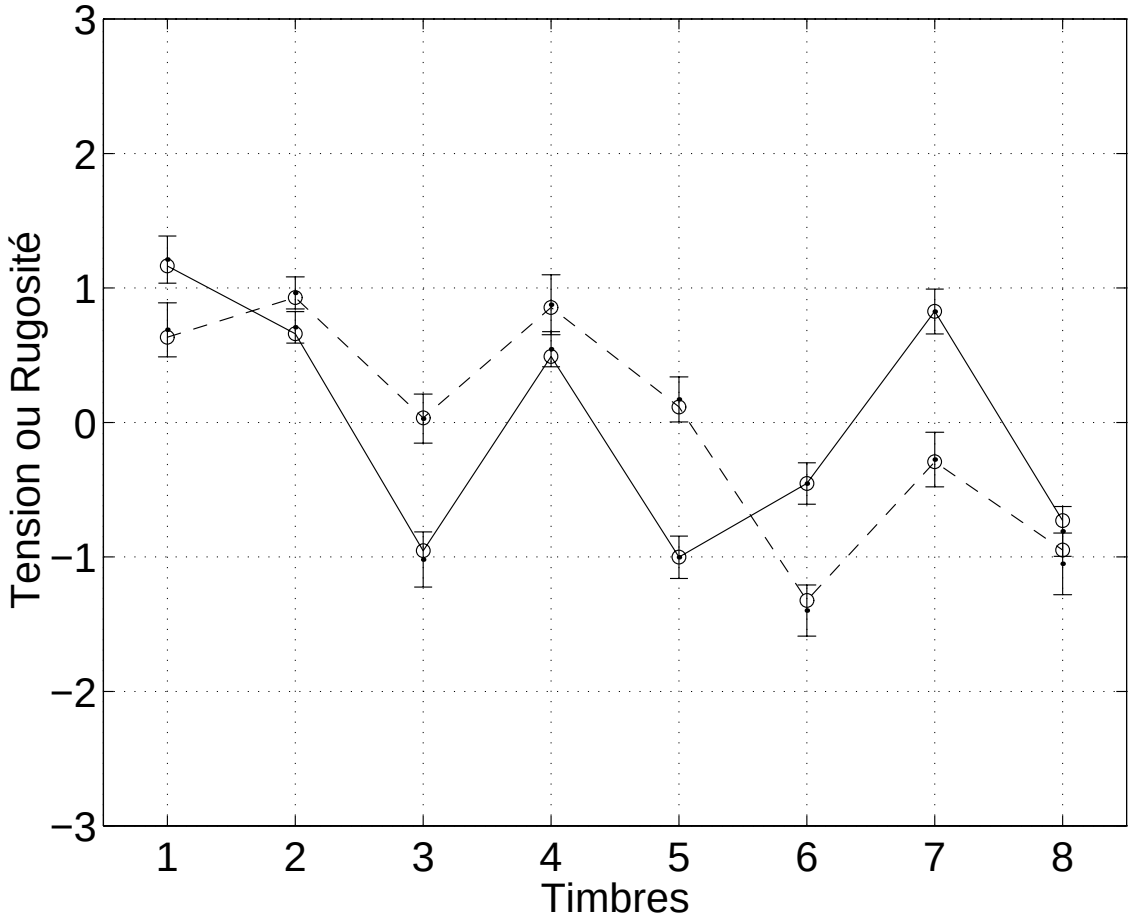


FIG. 8.4 – Échelles de tension (trait plein) et de rugosité (trait pointillés) pour l'Expérience 4. Les timbres sont en abscisse, les résultats de l'analyse BTL et les écarts-type estimés par bootstrap sont en ordonnée. Les cercles représentent la valeur BTL, les points la moyenne des réplications par bootstrap. L'échelle en abscisse n'est pas continue, les lignes entre les points ont été rajoutées pour plus de clarté. Les unités sont arbitraires et représentent le contraste perçu entre les stimuli.

### 8.2.3 Analyses supplémentaires

#### 8.2.3.1 Influence de l'ordre de présentation

L'influence éventuelle de l'ordre de présentation des timbres n'est pas prise en compte par l'analyse BTL : bien que chaque paire de timbres soit présentée dans les deux ordres séquentiels possibles au cours de jugements différents, les jugements pour une même paire sont regroupés pour calculer une proportion globale. Nous avons examiné si l'ordre de présentation avait en fait eu un effet général sur les jugements des auditeurs.

Un indice appelé coefficient de sensibilité à l'ordre (*cso*) a été défini comme le pourcentage d'essais pour lesquels le premier élément des paires a été choisi comme plus tendu ou rugueux. Si l'ordre de présentation n'a pas d'influence systématique sur les jugements, le premier élément d'une paire a autant de chance d'être "choisi" que le deuxième et le *cso* doit être égal à 50 %. Le test d'inférence statistique pour le *cso* a été défini comme suit : si la valeur 50 % est incluse dans l'intervalle de confiance à un risque donné, alors l'ordre est considéré comme n'ayant pas d'effet.

Le *cso* a été calculé pour tous les jugements de tension. Ce coefficient est contraint à l'intervalle $[0, 1]$ et donc l'emploi des écarts-type pour estimer un intervalle de confiance nécessiterait l'emploi d'une transformation. Pour éviter cela, nous avons choisi d'estimer l'intervalle de confiance à 90 % par la méthode des centiles. Un échantillon de bootstrap de 2000 échantillons a été généré, et l'intervalle de 90 % autour du cas central pris comme une estimation de l'intervalle de confiance (Efron et Tibshirani, 1993). L'ordre de présentation a été trouvé comme ayant eu un effet faible, mais significatif, sur les jugements : $cso = 0.45, cso \in [0.42, 0.47], p < 0.05$. L'ordre de présentation a aussi faiblement influencé les jugements de rugosité : $cso = 0.47, cso \in [0.46, 0.49], p < 0.05$.

Malgré la significativité observée, nous interprétons cette analyse comme une indication que l'ordre n'a eu que peu d'influence sur les jugements. En effet, la valeur moyenne du coefficient reste proche de 50 %. D'un point de vue pratique, ce biais est assez faible pour permettre l'emploi de la méthode BTL qui néglige l'influence de l'ordre.

#### 8.2.3.2 Stratégies d'écoute et/ou de jugement

L'étude statistique de la cohérence des jugements de tension recueillis entre les auditeurs est un moyen d'estimer s'ils ont tous répondus de la même manière, ou si différentes stratégies d'écoute et/ou de jugement sont apparues entre eux. Cette étude prend d'autant plus d'importance que le groupe expérimental comportait des musiciens professionnels, des amateurs, des auditeurs non musiciens, et qu'ils devaient juger une notion aussi complexe que la tension musicale sans que des critères explicites de jugement leur aient été suggérés. Un éventuel lien entre une stratégie de jugement et l'appartenance à l'une ou l'autre des catégories d'auditeurs serait l'indication d'une influence de l'éducation musicale sur les jugements expérimentaux.

La méthode employée pour étudier la cohérence des jugements à travers les auditeurs est l'analyse par composantes principales (ACP). Les 8 stimuli ont été jugés par 29 auditeurs. S'il n'y a aucune redondance à travers les données, 8 dimensions sont nécessaires pour les représenter. Toutefois, s'il y a en fait une certaine redondance, par exemple si les jugements des auditeurs sont corrélés entre eux, un plus petit nombre de composantes peut être suffisant. Le nombre de composantes nécessaires est trouvé en examinant la structure de la matrice de données après normalisation (Hotelling, 1933).

L'analyse a été faite a partir des jugements de tension de l'Expérience 4. Pour chaque timbre, un score a été défini comme le nombre de fois où il a été jugé plus tendu à travers tous les jugements d'un sujet donné. Nous interprétons chaque composante ou direction du nouvel espace comme représentant une stratégie d'écoute et/ou de jugement partagée par plusieurs auditeurs, ou en d'autres termes comme un "auditeur type". Il est alors possible de représenter les vecteurs des coefficients des auditeurs sur les composantes principales. Ces coefficients représentent la contribution de chaque auditeur à la composition de l'"auditeur-type" associé à la composante. Les valeurs attribuées aux différents timbres par ces "jugements-type" peuvent aussi être représentés dans le nouvel espace pour interpréter les axes obtenus (voir Annexe A).

La variabilité des scores représentant les jugements de tension a pu être expliquée par deux composantes expliquant 75 % de la variance à elles deux. Les autres composantes, expliquant moins de variance que les axes initiaux, ne sont pas considérées dans la suite de la discussion. L'examen des caractéristiques de

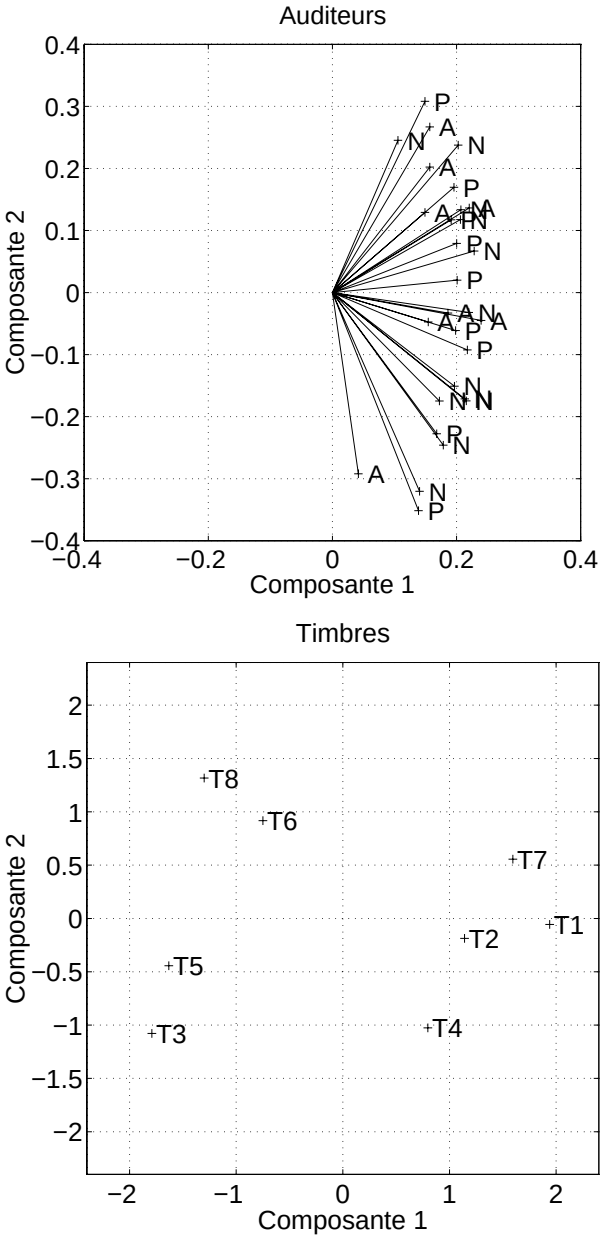


FIG. 8.5 – Analyse par composante principale des résultats de l'Expérience 4. Les coefficients attribués à chaque auditeur et valeurs attribuées à chaque timbre le long des deux premières composantes principales pour les jugements de tension sont représentés. Les musiciens professionnels sont notés P, les amateurs A et les auditeurs non musiciens N. Les timbres sont notés $T_1 - T_8$.

ces deux composantes montre que les auditeurs ont tous de larges poids de même signe sur la première dimension, qui explique 56 % de la variance (Figure 8.5). Ce résultat indique un fort consensus entre auditeurs. L'ordre de gauche à droite des timbres projetés sur cette dimension correspond au classement par tension croissante qui peut être dérivé de l'analyse BTL. Cette dimension, de par la forte variance expliquée et l'homogénéité des poids des sujets, est donc la dimension la plus importante des données et fournit la base des résultats de l'échelle BTL.

La seconde dimension différencie $T_8, T_6, T_7$ de $T_3, T_4$. Les coefficients des auditeurs sont régulièrement distribués le long de cette dimension, ce qui suggère des différences entre leurs stratégies. La distribution des sujets selon cette dimension ne montre aucune corrélation systématique avec l'éducation musicale (P, professionnel ; A, amateur ; N, non musicien). Cette dimension ne peut donc pas s'interpréter comme reflétant des différences explicites d'éducation musicale (telle que mesurée par notre critère). L'interprétation de cette deuxième dimension, qui explique nettement moins de variance que la première, sera discutée plus loin (section 8.5.2).

## 8.3 Expérience 5

### 8.3.1 Fusion et spécificités

Considérons de nouveau la façon dont les timbres ont été composés. Du fait de l'utilisation d'analyses spectrales de sons naturels, la partition contient de nombreuses notes qui entretiennent des rapports quasi-harmoniques. Par exemple, pour le timbre $T_7$ dont l'analyse était donnée en Figure 8.3, l'Alto et la Flûte 1 sont pratiquement espacés d'un octave. Les harmoniques des différents instruments ont aussi tendance à se superposer dans une large mesure. L'impression qui en résulte (et qui est centrale à l'esthétique du compositeur) est celle d'une "fusion" entre instruments. Il serait difficile de définir acoustiquement ce qui est impliqué par ce mot de fusion. Remarquons simplement de manière qualitative qu'une interrogation systématique des auditeurs à l'issue de l'Expérience 4 a permis de se rendre compte qu'aucun d'entre eux (même les musiciens professionnels) n'était capable de nommer les cinq instruments employés. L'impossibilité d'identifier des voyelles présentées simultanément a été démontrée comme liée à

l'absence de ségrégation dans le sens de l'analyse des scènes auditives (de Cheveigné, McAdams, Laroche et Rosenberg, 1995).

Toutefois, une écoute attentive des timbres permet de remarquer que des traits saillants sont présents dans certains des timbres étudiés. Les timbres $T_6, T_7, T_8$ contiennent une note de flûte aiguë, qui devait être jouée à une dynamique modérée (voir la partition en Figure 8.1). Dans l'interprétation réalisée, ces notes émergent du timbre global (voir par exemple l'amplitude de la composante "Flûte 2" dans la Figure 8.3). Ceci donne lieu à ce que nous appellerons une spécificité "de surface". Pour les timbres $T_3, T_4$, des intervalles évoquant une fonction harmonique peuvent être découverts dans la partition (par exemple pour $T_3$ une quinte parfaite entre les flûtes et une tierce presque majeure entre la clarinette et la seconde flûte). Ces intervalles évoquent une fonction syntaxique forte dans le système tonal. Dans l'intention originale du compositeur, ils auraient dû se fondre dans le timbre global. Néanmoins, le niveau relatif entre les instruments les a rendu saillants dans nos enregistrements. Ceci pourrait constituer ce que nous appellerons une spécificité "harmonique".

Une nouvelle expérience a été conçue pour laquelle des modifications des stimuli acoustiques ont été réalisées, sous la supervision du compositeur, pour parvenir à une meilleure "fusion". Un enregistrement multipiste a été utilisé pour produire une version des timbres pour laquelle le niveau relatif de chaque instrument ainsi que l'évolution de la dynamique dans le temps ont été modifiés. Là encore, aucun critère acoustique rigoureux ne peut être fourni pour décrire l'amélioration du degré de fusion : les timbres étaient retouchés sous le contrôle du compositeur, dans le but d'obtenir un résultat sonore plus proche de son intention originale. Néanmoins, nous proposons à titre d'exemple en Figure 8.6 les analyses acoustiques du timbre $T_7$ ainsi modifié. Il apparaît que dans ce cas, le remixage a eu pour effet de rétablir un équilibre approximatif entre instruments, notamment entre la Flûte 1, la Clarinette et la Flûte 2. L'onde temporelle permet aussi de remarquer que l'enveloppe temporelle est plus régulière, les attaques des différents instruments ayant été adoucies. Ceci retire un indice potentiel de ségrégation auditive (Bregman, 1990). Après le remixage des huit timbres, une procédure identique à l'Expérience 4 a été employée pour établir des échelles de tension et de rugosité entre les nouveaux stimuli.

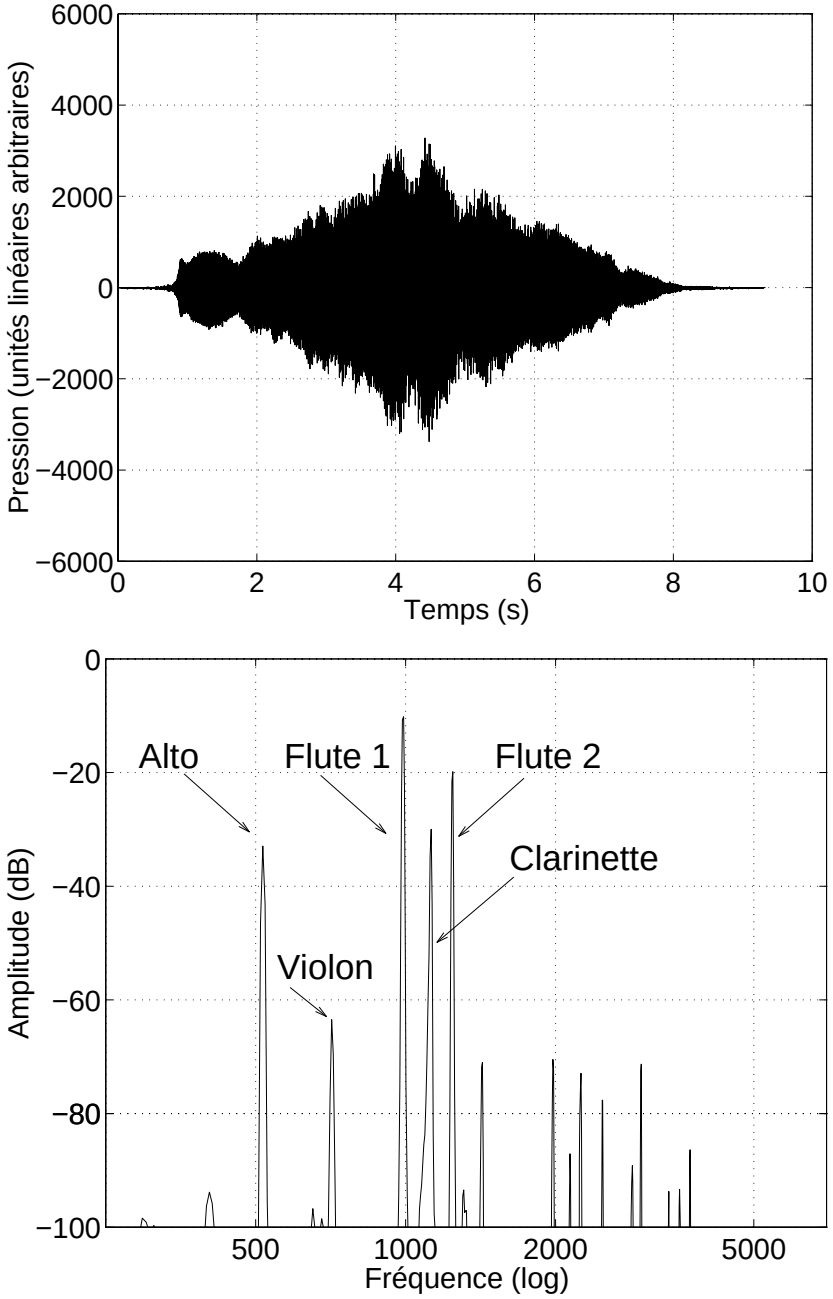


FIG. 8.6 – Onde temporelle et spectre à long terme du timbre orchestral $T_7$ après remixage utilisé dans l'Expérience 5. Les unités de pression ainsi que la référence de l'échelle de décibels sont arbitraires. Les composantes spectrales correspondant aux fréquences fondamentales des différents instruments sont indiquées.

## 8.3.2 Méthode

### 8.3.2.1 Stimuli

Une nouvelle version des timbres utilisés dans l'Expérience 4 a été réalisée (Figure 8.1). Un enregistrement multipiste réalisé simultanément avec l'enregistrement stéréo était disponible pour ces timbres. Les mêmes prises ont été utilisées. L'équilibre des volumes relatifs entre instruments ainsi que leurs évolutions temporelles ont été modifiés par remixage. Un effet de salle artificiel a été rajouté grâce à un processeur Lexicon pour se rapprocher de l'effet de salle naturel présent dans l'Expérience 4. Un exemple de timbre remixé est proposé en Figure 8.6.

### 8.3.2.2 Matériel

La prise de son multipiste a été réalisée avec 5 microphones cardioïde KM 140 Neuman, placés à 50 cm de chaque instrument. Le mixage a été réalisé manuellement sur la table Neev V à partir d'un lecteur multipiste 33.24 A Sony vers un DAT 7050 Sony. Le matériel de présentation des stimuli et de collecte des résultats est resté strictement identique à l'Expérience 4.

### 8.3.2.3 Sujets

Un groupe de dix sujets âgés de 18 à 45 ans ($M = 27$) a participé à l'expérience. Les sujets n'avaient pas participé à l'Expérience 4. Ils étaient recrutés de la même manière. Le groupe expérimental comportait deux musiciens professionnels, quatre musiciens amateurs et quatre auditeurs non musiciens. Ils étaient payés pour leur participation.

### 8.3.2.4 Procédure et analyses

La procédure expérimentale a été maintenue strictement identique par rapport à l'Expérience 4.

## 8.3.3 Résultats

Les méthodes d'analyse statistiques de l'Expérience 4 ont été reprises, et les résultats sont proposés en Figure 8.7. Pour les jugements de tension, l'analyse

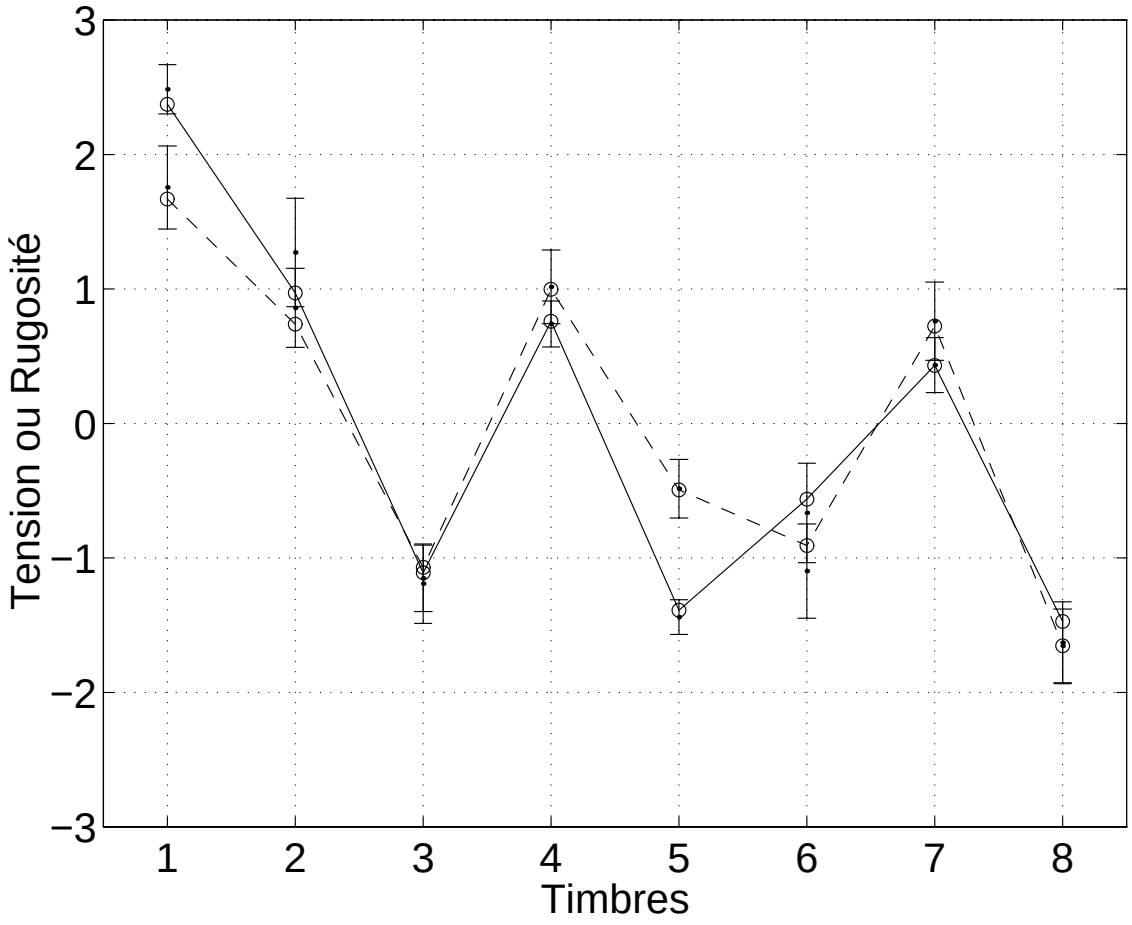


FIG. 8.7 – Échelles de tension (trait plein) et de rugosité (trait pointillés) pour l'Expérience 5. Les timbres sont en abscisse, les résultats de l'analyse BTL et les écarts-type estimés par bootstrap en ordonnée. Les cercles représentent la valeur BTL, les points la moyenne des réplications par bootstrap. L'échelles en abscisse n'est pas continue, les lignes entre les points ont été rajoutées pour plus de clarté. Les unités sont arbitraires et représentent le contraste perçu entre les stimuli.

BTL ainsi que le bootstrap montrent qu'une nouvelle fois des différences significatives ont été perçues entre les timbres. L'écart entre les stimuli situés aux valeurs extrêmes de l'échelle est plus accentué que dans l'Expérience 4, et toujours significatif à $p < 0.05$. La même tendance générale de réponse est observée, avec une distribution en deux larges groupes, à l'exception que désormais le timbre $T_1$ est perçu comme significativement plus tendu que tous les autres. Les auditeurs ont donc pu, dans cette nouvelle expérience, établir plus de deux niveaux de tension. A l'intérieur des groupes, le rang dans l'échelle de tension de $T_7$ et $T_8$ ont diminué alors que celui de $T_3$ a augmenté. Remarquons toutefois que l'échelle de tension diffère toujours de celle prévue par le compositeur, notamment pour les timbres $T_6, T_7$ et $T_8$.

Sur l'échelle de rugosité, $T_1$ devient le timbre le plus rugueux et $T_8$ le timbre le moins rugueux. L'écart qui les sépare est supérieur à celui observé dans l'Expérience 4, mais les écarts-type sont aussi plus importants. La différence observée est néanmoins toujours significative. Le même profil général de rugosité est observé, à la différence que les rangs sur l'échelle de rugosité de $T_6$ et $T_7$ ont augmenté.

Les conséquences de ces changements dans les deux échelles sont qu'elles se recouvrent désormais dans une large mesure. Un seul timbre, $T_5$, provoque une différence significative entre tension et rugosité jugées. Un coefficient de corrélation a été calculé pour décrire la ressemblance entre les échelles de tension et de rugosité. Le coefficient de corrélation est contraint à l'intervalle $[-1, 1]$ son intervalle de confiance à 90% a donc été estimé par bootstrap et centiles. Pour cette Expérience 5, la corrélation entre les deux échelles est de $r = 0.94,\ r \in [0.84\ 0.96]$. A titre de comparaison, la corrélation était de $r = 0.53,\ r \in [0.32\ 0.73]$ pour les échelles de tension et de rugosité de l'Expérience 4. A l'exception de la différence significative observée pour $T_5$ (pour laquelle nous ne proposons aucune interprétation), ces résultats indiquent que la perception de rugosité pourrait suffire pour expliquer la perception de tension des timbres remixés.

### 8.3.4 Réplication des analyses supplémentaires

Les analyses supplémentaires effectuées sur les résultats de l'Expérience 4 ont été répétées avec les données de l'Expérience 5. L'ordre de présentation tel que mesuré par le *cso* est apparu comme n'ayant d'influence significative ni sur les jugements de tension, ni sur les jugements de rugosité.

L'analyse par composantes principales effectuée pour caractériser les stratégies d'écoute et/ou de jugement a révélé pour les données de scores de tension une solution quasi-unidimensionelle. La première composante principale explique 73 % de la variance. Cette composante est une composante de consensus entre sujets. Elle reproduit les résultats de l'analyse BTL pour ce qui est de la tension jugée des timbres. La seconde composante principale explique 15 % de la variance, ce qui n'est pas beaucoup plus que n'importe laquelle des variables originales. Ceci indique que les jugements de cette Expérience 5 peuvent être presque exclusivement interprétés en termes de consensus.

## 8.4 Modélisation

Un modèle simple de calcul de rugosité a été testé pour tenter de reproduire les jugements expérimentaux (Hutchinson et Knopoff, 1978). Ce modèle a été présenté en section 3.2.2. Malgré ses nombreuses faiblesses théoriques, ce modèle a été testé car ses paramètres d'entrée sont du même type que ceux utilisés par le compositeur dans son algorithme original. D'une manière plus générale, ce modèle est bien adapté à la notation musicale où les accords sont représentés par une liste de notes [3].

Pour simuler nos résultats expérimentaux, les données fournies au modèle devraient être les notes écrites sur la partition, avec toutes leurs amplitudes égales. Néanmoins, un spectre doit aussi être associé à chaque note car les battements entre partiels sont une source essentielle de rugosité (Kameoka et Kuriyagawa, 1969b). Nous avons donc choisi d'associer à chaque fréquence fondamentale un spectre hypothétique composé de 4 partiels avec une décroissance de 12 dB/octave. Ceci représente bien sûr une schématisation des stimuli, mais demeure plausible si l'on considère le registre et les instruments employés (voir par exemple les Figures 8.3 et 8.6).

Les valeurs de rugosité calculées sont présentées en Figure 8.8. La corrélation entre l'analyse BTL des jugements de rugosité expérimentaux de l'Expérience 4 et la rugosité calculée est de ($r = 0.83$, $r \in [0.73\ 0.90]$, intervalles de confiance à 90% estimés par bootstrap et centiles). Pour l'échelle de rugosité expérimentalement

3. Ce modèle de rugosité a d'ailleurs été implémenté, grâce à G. Assayag, dans l'environnement d'aide à la composition *Patchwork* développé par l'Ircam et il a été utilisé par plusieurs compositeurs depuis.

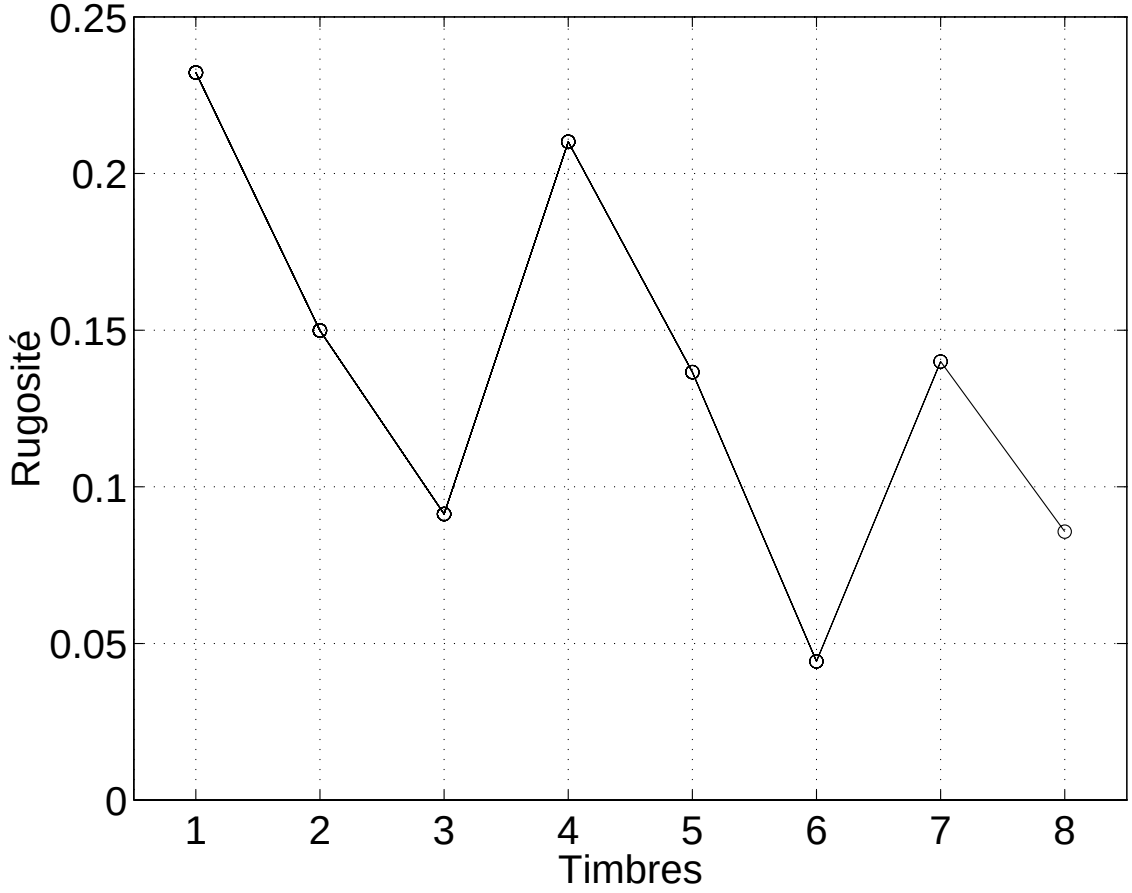


FIG. 8.8 – Estimation de rugosité par le modèle de Hutchinson et Knopoff (1978) pour les huit timbres. L'échelle relative de rugosité est arbitraire.

obtenue dans l'Expérience 5, la corrélation est de ($r = 0.88,\ r \in [0.81\ 0.94]$). Les estimations de rugosité sont donc généralement concordantes avec les simulations obtenues grâce au modèle.

Il est possible de comparer qualitativement les prévisions du modèle de rugosité avec l'algorithme original du compositeur dont les prévisions étaient données en Figure 8.2. Si l'on considère la rugosité comme un indicateur de tension, le modèle de rugosité est globalement meilleur pour prédire les jugements expérimentaux. Le timbre $T_7$, par exemple, était prévu comme évoquant une faible tension mais a été jugé comme à la fois tendu et rugueux. Le modèle reproduit ce résultat[4].

## 8.5 Discussion générale

Cette série d'expériences avait pour but d'examiner si des degrés de tension et de rugosité pouvaient être perçu pour un matériau musical complexe non-tonal. Nous allons dans un premier temps résumer les résultats obtenus. La comparaison des deux conditions expérimentales permettra ensuite de préciser les conditions du lien observé entre rugosité et tension. Le contexte de ces expériences, visant à se placer dans une situation musicale aussi "réaliste" que possible, sera ensuite discuté. Enfin, les résultats seront replacés dans le cadre des théories de tension musicale non tonale.

### 8.5.1 Tension et rugosité

L'Expérience 4 nous a permis d'exhiber des différences significatives de tension entre timbres orchestraux. La notion de "tension" – volontairement laissé floue

4. Les limites du modèle apparaissent toutefois clairement lorsque l'on veut mettre en évidence les différences entre les deux expériences. Une analyse acoustique des différents timbres a été réalisée de manière à en extraire une liste de partiels et d'amplitudes. Cette analyse a été réalisée, dans une perspective d'utilisation musicale, à l'aide d'outils disponibles pour les compositeurs. Une transformée de Fourier à court terme avec une longue fenêtre d'analyse et restreinte à la partie stationnaire du son a été réalisée avec le logiciel *AudioSculpt*. Un algorithme simple de suivi de partiels a ensuite été employé (Hanappe, 1996), et les fréquences et amplitudes exportées vers *Patchwork* dans lequel la rugosité a été calculée. Les corrélations respectives avec les résultats expérimentaux sont moins bonnes que celles obtenues sans les analyses acoustiques. Le modèle est donc mieux adapté à une utilisation "experte" pour laquelle des hypothèses simplificatrices sont faites sur le résultat sonore d'une partition, et ne parvient pas à prendre en compte les subtilités des signaux acoustiques réellement obtenus.

dans les instructions expérimentales – a donné lieu a un consensus entre auditeurs dans un contexte musical qui n'était pas familier à la plupart d'entre eux. La rugosité est aussi apparue comme un critère approprié pour juger les timbres. Toutefois, des différences significatives sont apparues entre les échelles, ce qui indique que des facteurs autres que la rugosité ont participé aux jugements de tension.

L'Expérience 5 a confirmé la perception de degrés de tension et de rugosité différents entre timbres orchestraux. Les deux échelles perceptives sont apparues pratiquement identiques. La corrélation observée indique que dans cette deuxième condition les jugements de rugosité pourraient suffire à eux seul pour rendre compte des jugements de tension, une plus grande rugosité induisant plus de tension. Bien entendu, une corrélation observée n'est en aucun cas la preuve d'un lien de causalité, mais d'autres éléments apparus lors des analyses supplémentaires réalisées sont cohérents avec l'hypothèse que la rugosité a constitué une base sensorielle pour les jugements de tension des timbres.

La succession temporelle entre les timbres n'a pas eu de grande influence sur les jugements de tension. Par conséquent, il semble que ce soient principalement les aspects intrinsèques des stimuli, à mettre en parallèle avec les concepts de consonance et dissonance intrinsèques pour la musique tonale, qui ont influencé les jugements de tension. La rugosité est considérée comme participant à la consonance intrinsèque dans un contexte tonal. Nos résultats pourraient donc contribuer à étendre cette interprétation à un contexte non tonal.

L'influence de l'acculturation implicite ou explicite des auditeurs (provenant tous d'un milieu "occidental") n'a pas une influence directe. Plus précisément, aucune corrélation n'a été trouvée entre les stratégies d'écoute et/ou de jugement et l'éducation musicale. La perception de rugosité, en tant qu'attribut sensoriel, est attendue comme ne dépendant pas de l'éducation musicale.

Finalement, la corrélation entre les échelles de tension et de rugosité n'est pas due à une confusion possible entre les mots "tension" et "rugosité". Dans les instructions expérimentales, la définition de "tension" a été laissée à l'appréciation des auditeurs, alors que la notion de rugosité a été définie par un exemple sonore. Les estimations de rugosité ont ensuite été obtenues par deux moyens distincts, à savoir une évaluation expérimentale et une modélisation informatique, et les résultats obtenus sont convergents. Enfin, les différences significatives observées

dans l'Expérience 4 entre les échelles de tension et de rugosité montrent que les auditeurs ont pu juger séparément les deux dimensions.

### 8.5.2 Comparaison entre Expériences 4 et 5

L'Expérience 4 a démontré des différences significatives entre les échelles de tension et de rugosité pour les stimuli étudiés, différences qui disparaissent généralement dans les résultats de l'Expérience 5. Nous interprétons cet effet comme basé sur les spécificités mentionnées en section 8.2.3.2. Le statut de ces spécificités est hypothétique. Néanmoins, une telle hypothèse nous permet d'interpréter plusieurs aspects des données recueillies.

Si l'on excepte $T_5$, tous les timbres qui ont montré une différence significative entre tension et rugosité dans l'Expérience 4 contenaient des spécificités. De plus, la nature des spécificités est cohérente avec les résultats : $T_6$, $T_7$ ont été jugés tendus mais moins rugueux et ils contiennent une note aiguë de flûte qui a pu induire une certaine tension sans provoquer de rugosité ; $T_3$ contenait des intervalles associés à la détente dans le système tonal et il a été jugé rugueux mais moins tendu. L'effet de la transformation effectuée sur les stimuli utilisés dans l'Expérience 5 peut être alors interprété comme ayant réduit la saillance perceptive de ces spécificités. Alors que moins de spécificités émergeaient des timbres, la rugosité est devenue le principal critère pour juger la tension.

L'analyse ACP réalisée pour étudier les stratégies d'écoute et/ou de jugement de l'Expérience 4 a révélé une composante secondaire de contraste entre les sujets. Cette composante oppose les timbres contenant une spécificité "de surface" à ceux contenant une spécificité "harmonique". Cette composante peut donc être interprétée comme un désaccord dans les stratégies d'écoute et/ou de jugements concernant l'impact des spécificités sur les jugements de tension. Après le remixage des timbres pour l'Expérience 5, cette composante secondaire a disparu. Le remixage ayant probablement eu pour effet l'amélioration de la fusion perceptive des timbres, cette influence de l'organisation des scènes auditives sur le lien entre tension et rugosité confirme l'hypothèse émise par Mathews (1987) évoquée au Chapitre 4.

Soulignons néanmoins que l'Expérience 5 n'est en aucun cas, à nos yeux, une version "améliorée" de l'Expérience 4 où d'éventuels "défauts de fabrication" auraient été éliminés des stimuli. Notre conclusion n'est pas que rugosité

et tension sont équivalentes : dans le cas général, comme le montre par exemple l'Expérience 4, l'influence de la rugosité se combine avec celle de nombreux autres facteurs pour participer à l'écoute attentive de l'auditeur et à ses impressions de tension et détente. Les transformations réalisées pour parvenir aux stimuli de l'Expérience 5, qui reviennent principalement à rendre plus homogènes entre eux les timbres comparés, permettent de montrer que la rugosité peut *éventuellement* exprimer à elle seule des mouvements de tension dans une situation qui reste musicalement réaliste. La comparaison entre les deux expériences permet en fait de mieux comprendre les conditions dans lesquelles la rugosité peut devenir un facteur prépondérant dans la perception de tension.

### 8.5.3 Remarques sur la procédure expérimentale et les méthodes d'analyse

Ces deux expériences ont tenté d'établir un lien entre un attribut psychoacoustique élémentaire, la rugosité, et la notion musicalement importante de tension. Le choix des stimuli reflète donc un compromis entre le réalisme musical et la limitation des variables expérimentales. D'un côté, les stimuli ont été extraits d'une œuvre musicale existante. Ils ont été enregistrés avec des instruments acoustiques, en situation de concert, et ont été remixés dans une seconde condition dans une situation habituellement rencontrée pour la préparation de musique enregistrée. La question de trouver une hiérarchie de tension a été motivée par une préoccupation rencontrée par le compositeur pendant l'écriture de la pièce. D'un autre côté, les timbres étaient homogènes en ce qui concerne le registre et la dynamique, ce qui peut représenter une limitation pour d'autres situations musicales. Par exemple, il est possible de s'interroger sur la saillance qu'auraient les indices de tension liés à la rugosité quand ceux-ci sont placés en compétition avec d'autres indices de différente nature, comme le rythme, la dynamique, ou l'utilisation de registres extrêmes. L'interaction entre facteurs mélodiques et harmoniques n'a pas non plus été étudiée ici. Ces interactions sont pourtant importantes dans le contexte de la musique polyphonique où les phénomènes de formation de flux ont été supposées influencer la perception de tension et de rugosité (Wright et Bregman, 1987).

La méthode de comparaison par paires a été utilisée pour éviter les jugements directs. Ces jugements auraient été difficiles à définir sans faire appel à des termes non forcément pertinents dans un contexte musical (agrément, etc.). De plus, la

complexité des stimuli aurait pu rendre la tâche délicate pour les auditeurs. Ce plan expérimental nous a amené à utiliser des méthodes d'analyse statistique originales : la technique de bootstrap a été employée pour estimer l'écart-type de valeurs BTL (voir Annexe A).

Ces analyses ont conduit à définir une seule valeur de tension pour chaque timbre. Bien entendu, ceci ne signifie pas qu'il existe une valeur "absolue" de tension intrinsèque pour les timbres étudiés. Il convient de souligner que la hiérarchie obtenue est relative. Les valeurs de tension estimées peuvent être pensées d'une façon similaire aux valeurs relatives de consonance et dissonance traditionnellement assignées aux intervalles utilisés dans la musique tonale. Bien que le niveau de dissonance "acceptable" ait varié au cours de l'histoire de la musique, ceci n'écarte pas la perception de consonance sensorielle comme une base potentielle à la consonance musicale (Bregman, 1991). Schoenberg (1911), par exemple, préférait appeler les "dissonances" des "consonances éloignées" !

### 8.5.4 Perception de tension musicale dans une musique non tonale

Les théories qui se sont intéressées à la perception de tension non tonale sont bien moins nombreuses que celles portant sur la tension tonale (voir Chapitre 4). Certains compositeurs ont proposé des théories de tension non tonale en relation étroite avec leurs propres styles musicaux, comme une codification de pratiques personnelles. Une théorie plus directement concernée par la vérification expérimentale est l'extension à la musique atonale de la *Théorie Générative de la musique tonale* faite par Lerdahl (1989).

La TGMT propose la réduction d'une pièce musicale, à différents niveaux, en unités structurelles grâce à la "réduction de la trame temporelle" (section 4.3.2). Cette réduction est faite selon des règles de stabilité qui sont basées sur les relations hiérarchiques qui existent dans l'espace des hauteurs lié à la tonalité (Krumhansl, 1990). La réduction est ensuite supposée fournir la base de la perception des mouvements de tension et détente ("réduction prolongationelle"). Pour la musique atonale, Lerdahl fait l'hypothèse que l'espace des hauteurs est "plat" et remplace les règles de stabilité par des règles de saillance pour obtenir des réductions similaires. Les théories de réduction temporelle tonales et atonales ont été toutes deux testées expérimentalement, avec des résultats contrastés (Dibben,

1994). Alors que les auditeurs étaient capables d'apparier une réduction avec son original pour des pièces tonales, les résultats étaient proches du hasard dans les conditions atonales. L'auteur propose différentes raisons pour ce dernier résultat, mais la conclusion suggérée est que l'incapacité d'apparier originaux et réductions est inhérente à la musique testée (des extraits des opus 11 et 19 de Schoenberg). Des niveaux de structure hiérarchiques seraient impossibles à extraire à l'écoute de musique atonale et les seuls mouvements de tension et détente qu'il est alors possible de percevoir dans ce contexte, à l'aide de simples associations entre évènements musicaux, sont donc d'importance structurelle moindre (Dibben, 1996).

Nos résultats permettent de nuancer cette conclusion. Les jugements expérimentaux de timbres orchestraux ont montré que des mouvements de tension et de détente pouvaient en fait être perçus par des auditeurs dans un contexte musical non relié à la tonalité. Une autre étude a donné un ensemble de résultats convergents avec ceux-ci, en démontrant un contraste aussi élevé pour les jugements de tension d'extraits de musique non tonale que tonale, avec de plus un effet significatif du timbre (Paraskeva et McAdams, 1997). Une différence entre nos expériences et celles de Dibben (1994, 1996) est qu'ici l'harmonie sous-jacente a été abordée en termes de timbre plutôt qu'en termes d'ensemble de hauteurs notées[5]. La qualité de timbre à elle seule, sans référence à l'organisation d'un espace de hauteurs, a été suffisante pour transmettre des mouvements de tension et de détente entre des paires de sons. La question est maintenant de savoir si les propriété locales que nous avons étudiées (comme la rugosité) sont capables de transmettre des mouvements de tension et de détente sur de plus longues échelles temporelles. Il serait d'un grand intérêt de tester si des originaux et des réductions, similaires à celles proposées par Lerdahl mais basée sur des règles de saillance de timbre incluant la rugosité, pourraient être appariées.

Le timbre comme dimension potentielle porteuse de forme a intéressé les compositeurs, depuis les expériences de Berlioz et Debussy jusqu'aux *Klangfarbenmelodien* de Schoenberg, et plus récemment avec les expériences dans le domaine de la synthèse sonore (Risset, 1988, 1994) et la tendance appelée musique spectrale (Murail, 1984; Pressnitzer et McAdams, 1998). La psychologie cognitive a ex-

5. Il convient de remarquer que dans la version originale de la TGMT (deuxième impression, page 297), le timbre était considéré comme un candidat pour participer à la perception de tension non tonale. Toutefois, les règles de saillance proposées par Lerdahl (1989) ont été défini en rapport à une collection de hauteurs et non à un événement sonore ayant un timbre homogène. Quoi qu'il en soit, elles ne comportaient pas la rugosité comme critère pertinent.

primé les conditions à partir desquelles un attribut multidimensionnel comme le timbre pourrait être la base de construction de forme musicale, principalement en termes de la possibilité de reconnaître et catégoriser des vecteurs dans un espace de timbre (McAdams, 1989; McAdams et Cunibile, 1992). Notre étude est une indication supplémentaire que le timbre pourrait aussi être un support privilégié pour l'harmonie, grâce à la dimension de rugosité.

## 8.6 Conclusion

Deux expériences ont montré que des mouvements de tension et détente ont pu être exprimés par des timbres orchestraux, alors qu'aucune fonction harmonique correspondant au système tonal occidental ne pouvait leur être attribuée. Divers facteurs peuvent influencer la perception de ces mouvements : des références tonales, auxquelles à la fois des musiciens et des non musiciens ont été sensibles, ou des caractéristiques de surface comme la présence de notes aiguës. Néanmoins, quand ces attributs ont été réduits, la rugosité est apparue fortement corrélée à la tension perçue. Ce résultat confirme et étend dans une situation musicale réaliste les théories plaçant la rugosité comme l'une des composantes de la perception de tension (Terhardt, 1974b; Mathews, Pierce et Roberts, 1987).

# Chapitre 9

# Conclusion

Les expériences qui ont été présentées au cours de cette thèse se sont intéressées à plusieurs niveaux de la perception de rugosité psychoacoustique : la nature des mécanismes liés à l'apparition et à l'intensité de la sensation a été mise en question ; la transition entre les résultats obtenus avec des sons simplifiés et l'estimation de rugosité de sons complexes a été considérée ; la place de la rugosité au sein de l'activité complexe qu'est l'écoute musicale a été évoquée. Nous souhaitons pour terminer résumer les principaux résultats obtenus, en soulignant les perspectives ouvertes par une meilleure compréhension du phénomène de rugosité dans le cadre des recherches sur la perception auditive.

## 9.1 Synthèse des résultats obtenus

Le percept élémentaire de rugosité peut être appréhendé selon deux grandes approches, que nous avons appelées "spectrale" (Plomp et Levelt, 1965; Kameoka et Kuriyagawa, 1969a) et "temporelle" (Terhardt, 1974a; Aures, 1985b). Ces approches ont été confrontées au cours des Expériences 1 et 2 grâce à des manipulations de phase. Pour des sons contenus à l'intérieur d'une même bande critique, il a été démontrée que la phase pouvait avoir une influence marquée sur la rugosité perçue. Une partie de l'effet observé peut être expliquée par des variations dans l'enveloppe des stimuli, ce qui renforce l'hypothèse temporelle contre l'hypothèse spectrale. Néanmoins, un effet de la phase a aussi été mis en évidence dans des conditions où la composition spectrale ainsi que le degré de modulation effective de l'enveloppe étaient maintenus constants. Cet effet a été interprété par

la présence d'une asymétrie de la forme de modulation, introduite par les effets du filtrage auditif (Expérience 1) ou imposée dans la construction des stimuli (Expérience 2). Une modulation avec une croissance abrupte et une décroissance lente semble plus rugueuse qu'une modulation avec une croissance lente et une décroissance abrupte. Ces résultats nouveaux nous ont conduit à une révision de l'hypothèse temporelle.

Pour faire le lien entre les stimuli employés dans l'étude des causes de la perception de rugosité et la rugosité de sons complexes, les modalités de combinaison de rugosités partielles éventuellement présentes dans plusieurs régions fréquentielles doivent être déterminées. Certains auteurs ont fait l'hypothèse d'une influence de la cohérence entre enveloppes (Mathes et Miller, 1947; Terhardt, 1974a), alors que d'autres attribuent cet effet à recouvrement entre patterns d'activité suivi d'une simple somme (Aures, 1985b). L'Expérience 3 a démontré que la rugosité de sons complexes ne pouvait être estimée par la simple somme des rugosités partielles considérées isolément. Un mécanisme de comparaison entre canaux auditifs, sensible à la cohérence des enveloppes temporelles, doit être pris en compte. Une première tentative de modélisation de l'effet a été proposée.

La rugosité des sons complexes que sont les accords musicaux possède une propriété remarquable : les accords les plus dissonants selon la théorie harmonique tonale occidentale sont justement les plus rugueux. Ceci a conduit à la formulation d'une théorie selon laquelle la rugosité fournirait une base sensorielle à la perception de dissonance musicale (von Helmholtz, 1877). Cette théorie permet d'envisager le rôle de la rugosité dans une situation d'écoute réaliste, où du statut de simple sensation elle devient partie prenante d'un ensemble de processus cognitifs de haut niveau : en effet, dans le contexte tonal, la rugosité se combine à de nombreuses autres influences — dont celles de l'apprentissage — pour participer à l'expression de mouvements de tension et détente (Bigand, Parncutt et Lerdahl, 1996). Les Expériences 4 et 5 ont montré que dans le cas d'un matériau musical n'ayant pas été composé selon les règles du système tonal, des mouvements de tension et détente pouvaient encore être perçus par des auditeurs musiciens ou non. La rugosité des timbres orchestraux utilisés a été trouvée corrélée aux mouvements de tension et de détente, plus de rugosité correspondant à plus de tension. La corrélation était d'autant plus marquée que les timbres étaient homogènes. Ces résultats vont dans le sens des théories considérant la rugosité comme

l'un des éléments contribuant à la perception de tension musicale, susceptible d'être pertinent hors du système tonal (Mathews, Pierce et Roberts, 1987).

## 9.2 Mécanismes temporels dans la perception auditive

### 9.2.1 Rugosité et modulations d'amplitude

Les résultats des Expériences 1, 2 et 3 indiquent que la rugosité est liée au traitement des fluctuations temporelles présentes dans les enveloppes des sons complexes. Ces fluctuations sont caractéristiques de nombreux signaux acoustiques. Pour des fréquences élevées, elles peuvent évoquer une sensation de hauteur dite résiduelle (Schouten, 1940; Licklider, 1954; Demany, 1989). Des périodicités plus longues aident à l'intelligibilité des signaux de parole (Drullman, Festen et Plomp, 1994). Les informations contenues dans les fluctuations d'enveloppes de fréquences entre 0 et 100 Hz peuvent même, si elles sont utilisées pour moduler des bandes de bruit, permettre à elles seules la compréhension d'un message parlé (Shannon, Zeng, Kamath, Wygonski et Ekelid, 1995). Enfin, à des fréquences inférieures la perception de rythme ou de prosodie est impliquée. Plusieurs mécanismes distincts sont sans doute utilisés dans les différentes régions de périodicité (Guttman et Julesz, 1963; Warren et Wrightson, 1981). La rugosité semble alors se situer à la limite inférieure de la perception de hauteur résiduelle (Terhardt, 1970). Cette constatation, issue de la simple comparaison des régions d'existence des deux sensations, mériterait sans doute d'être approfondie. L'étude des effets observés dans la perception de rugosité, comme les effets de forme d'enveloppe ou de cohérence entre canaux, dans le cadre de la perception d'autres attributs liés aux fluctuations d'amplitude permettrait peut-être d'apporter de nouveaux éléments à cette discussion.

### 9.2.2 Modélisation

Plusieurs modèles des traitements auditifs périphériques comportent désormais une étape d'analyse des enveloppes de modulation, en plus des étapes classiques de filtrage auditif et d'adaptation (Todd, 1996; Dau, Püschel et Kohlrausch,

1996). Un banc de filtres linéaires est utilisé pour extraire les modulations présentes dans chaque canal auditif. Cette méthode représente la simplification de données physiologiques sur la perception de modulations d'amplitude (Frisina, Smith et Chamberlain, 1990; Langner, 1992), elle a permis de reproduire une large variété de résultats expérimentaux (Dau, Püschel et Kohlrausch, 1996). Nos résultats sur l'influence de la phase et de la forme de l'enveloppe ne peuvent néanmoins pas être reproduits par de tels modèles. Ils indiquent donc que, au moins pour la perception de rugosité, l'un des aspects du codage de la modulation d'amplitude qui a pourtant son importance perceptive a été perdu dans cette simplification. Des modèles de neurones ont permis de reproduire de façon plus fidèle les observations physiologiques (Hewitt et Meddis, 1993; Lorenzi, Micheyl et Berthommier, 1995). L'application de tels modèles à la perception de rugosité, déjà envisagée de manière préliminaire (Pressnitzer, 1994), est sans doute une piste à explorer en vue de reproduire les influences observées. Une autre direction possible est le modèle de l'image auditive, dont le principe est susceptible de conserver les effets d'asymétrie que nous avons observés (Patterson, Allerhand et Giguère, 1995).

## 9.3 Psychoacoustique et écoute musicale

Le thème de la consonance musicale a suscité une masse étonnante de théories. Des philosophes grecs comme Pythagore ou Aristoxène de Tarente, des érudits du Moyen-Age comme Boèce, des physiciens comme Kepler, Galilée ou Mersenne, des mathématiciens comme Descartes, Leibniz, Euler, tous ont apporté leur contribution aux théories de l'harmonie. Ces contributions ont souvent été un reflet révélateur des conceptions scientifiques de leurs époques, et ont eu des motivations diverses.

L'étude de la rugosité psychoacoustique par Helmholtz s'inscrit dans cette longue tradition. La mise en parallèle d'un attribut élémentaire de la perception, enraciné dans des phénomènes physiologiques simples, et de l'écoute musicale, intimement liée à la créativité de chaque auditeur, peut bien entendue être longuement discutée. Nous nous contenterons ici de souligner qu'elle n'a pas pour but de valider ou d'infirmer telle ou telle réalisation musicale particulière, d'établir une sorte de critère de "perceptivement correct". Bien au contraire, la compré-

hension des particularités de la perception peut permettre de se les approprier pour éventuellement les intégrer dans de nouveaux langages musicaux comportant chacun leur nécessaire part d'arbitraire et de choix esthétiques.

L'étude des aspects élémentaires et musicaux de la rugosité doit à notre sens être vue plutôt comme un dialogue que comme une "application" de l'un vers l'autre. Nous avons adopté dans la présentation de nos résultats une perspective ascendante : des causes du phénomène à l'écoute en situation musicale. Il serait néanmoins tout aussi pertinent de reprendre l'ensemble de l'argumentaire en débutant par la question posée par la perception de tension musicale, et de redescendre jusqu'aux causes de la rugosité. Cette présentation serait d'ailleurs plus conforme à la chronologie réelle des expériences — les contraintes liées à l'étude musicale nous ayant par exemple amené à formuler le protocole expérimental retenu par la suite. Les interactions entre les différents étapes de cette étude ne peuvent toutes être explicitées ici, mais elles ont sans doute contribué à l'intérêt que nous avons personnellement trouvé à l'étude de cet attribut sonore (au nom toujours aussi effrayant).

## 9.4 Epilogue

Nous aimerions terminer en évoquant une coutume musicale de Nouvelle Guinée, pratiquée par les Kaluli de Papua. La cérémonie se déroule dans une hutte, où les membres d'un clan sont réunis. Des visiteurs de passage rentrent au crépuscule dans la hutte, en chantant des noms de lieux se référant au territoire de leurs hôtes. Ce chant produit des réactions particulières dans le public. Schieffelin (1979) cité dans Dowling et Harwood (1986) raconte

> "Après un moment, le public (les hôtes) deviennent très profondément émus. Certains éclatent en sanglots. Puis, en réaction au chagrin qu'on leur a fait éprouver, ils bondissent avec rage et brûlent les danseurs sur les épaules avec les torches utilisées pour éclairer la cérémonie. Les danseurs continuent leur performance sans montrer aucun signe de douleur. Les danses et les chants, ainsi que les sanglots et brûlures qui les accompagnent, continuent toute la nuit avec de brèves périodes de repos entre les chants." (p. 128, traduit de l'anglais).

L'évocation d'un nom de lieu auquel était peut être attaché un souvenir tragique a déclenché chez les auditeurs une telle réaction. Le fait est que le chanteur ne connaît que très peu de choses aux évocations qu'il utilise, étant totalement étranger au territoire de ses hôtes. Si seulement les quelques données concernant la perception auditive que nous avons rassemblées ici pouvaient contribuer à éviter de si embarrassantes situations...

# Annexes

# Annexe A

# Méthode expérimentale et analyses statistiques

Une même méthode expérimentale a été établie et développée pour toutes les expériences présentées dans cette thèse. Cette méthode repose sur le recueil de jugements de comparaisons par paires, la transformation de ces jugements en une échelle de valeurs par la méthode de Bradley-Terry-Luce (David, 1988), puis le test de la significativité des effets observés par une méthode de bootstrap (Efron et Tibshirani, 1993). Enfin, pour les expériences où des différences de stratégies basées sur l'éducation musicale des sujets pouvaient être considérées, une analyse par composantes principales a été réalisée. La présentation de cette méthode commune ainsi que les raisons de son choix ont été regroupés dans cette annexe.

## A.1 Plan expérimental

### A.1.1 Méthodes couramment employées

De nombreuses méthodes ont été utilisées pour obtenir des mesures de rugosité (voir Chapitre 2). Un rapide examen des techniques les plus courantes nous permet d'éclaircir les raisons du choix de la méthode finalement retenue.

La méthode la plus directe est celle d'estimation de grandeur. Suite à la présentation d'un stimulus, l'auditeur doit donner une valeur numérique correspondant à l'intensité de la sensation éprouvée (Stevens, 1957). Cette méthode a été largement employée pour recueillir des jugements de rugosité (Plomp et Levelt,

1965; Fastl, 1977; Kemp, 1982). Néanmoins, l'appariement d'une valeur numérique à une sensation représente un ensemble d'hypothèses fortes sur la nature de l'échelle à obtenir. De plus, l'auditeur doit pour faire son jugement établir une échelle mentale de la sensation à juger, se fixer une référence, et finalement décider d'une note chiffrée en restant cohérent avec ses jugements précédents. La difficulté de tels jugements pour la rugosité, même dans le cas de stimuli simples comme des sons purs modulés en fréquence, a été évoquée comme un problème par Kemp (1982). Les écarts-type obtenus avec cette méthode sont généralement assez élevés (voir les Figures 3.2 et 3.1 par exemple). Nous avons donc décidé de ne pas l'employer, du fait de la complexité des stimuli des Expériences 4 et 5 ainsi que de la finesse des différences étudiées pour les Expériences 1, 2 et 3.

Une autre méthode largement employée pour l'estimation de rugosité est celle d'appariement. Dans ce cas, le jugement est fait en appariant deux niveaux de sensation en manipulant un paramètre physique d'un son de référence. Dans le cas de la rugosité, l'ajustement est généralement fait en modifiant la profondeur de modulation d'un son pur (Terhardt, 1968a; Vogel, 1974; Aures, 1985b). Cette méthode réclame une certaine similarité entre le son testé et la référence pour obtenir des résultats stables (Vogel, 1974). L'étude de stimuli musicaux complexes dans les Expériences 4 et 5 nous a donc conduit à l'écarter, ainsi que ses variantes (méthode de fractionnement ou de production de rapports), pour les jugements de rugosité. Quant aux jugements de tension pour ces mêmes expériences, l'appariement d'un quelconque paramètre acoustique ne peut pas être envisagé sous peine de biaiser totalement les résultats.

Enfin, des comparaisons par paires peuvent être utilisées (Kaestner, 1909; Kameoka et Kuriyagawa, 1969a; Vos, 1986). Diverses variantes ont été utilisées : en comparant deux stimuli, l'auditeur peut simplement indiquer lequel provoque un niveau de sensation plus élevé (jugement binaire, 2AFC), ou indiquer une valeur numérique reflétant l'écart perçu entre les stimuli. Le recueil de données par une méthode d'estimation de magnitude et par des jugements binaires par paires pour des intervalles musicaux a été comparé de façon directe (Vos, 1986). Du fait de la simplicité relative des jugements binaires par rapport à l'estimation de grandeur, Vos note que cette méthode permet une plus grande stabilité et une meilleure finesse dans les résultats obtenus. La méthode de jugements binaires par

comparaisons par paires est celle qui a été retenue pour recueillir les jugements de rugosité (Expériences 1, 2, 3, 4, 5) et de tension (Expériences 4, 5).

### A.1.2 Méthode de jugements binaires par comparaisons par paires

La méthode de comparaison binaire par paires consiste à présenter à chaque auditeur l'ensemble des stimuli à évaluer deux par deux et à lui demander une décision. Plusieurs plans expérimentaux peuvent être adoptés, comportant éventuellement des répétitions pour augmenter le nombre de jugements ou au contraire en omettant certaines paires pour réduire le temps d'expérience (Kameoka et Kuriyagawa, 1969a; David, 1988).

Nous avons pour notre part choisi de présenter aux sujets toutes les paires distinctes possibles dans un ordre comme dans l'autre, sans répétitions. Par exemple, pour trois objets $\{T_1, T_2, T_3\}$, ceci donnerait les 6 comparaisons $(T_1, T_2)$, $(T_2, T_1)$, $(T_1, T_3)$, $(T_3, T_1)$, $(T_2, T_3)$ et $(T_3, T_2)$. Généralement, $n$ objets donneront lieu à $n(n-1)$ comparaisons. L'absence de répétition permet de limiter le temps d'expérience, alors que l'inclusion des paires dans un ordre comme dans l'ordre permet de conserver la possibilité d'estimer si justement l'ordre de présentation a une influence.

Pour chaque paire, $(T_1, T_2)$ par exemple, les auditeurs doivent comparer les deux éléments selon une certaine dimension perceptive et décider si selon cette dimension $(T_1 > T_2)$ ou $(T_1 < T_2)$. Les comparaisons sont à choix forcé, c'est à dire que l'auditeur ne peut pas répondre $(T_1 = T_2)$, ou "je ne sais pas". En effet, cette possibilité pourrait être utilisée différemment par chaque auditeur, les plus intrépides ne l'utilisant jamais alors que les plus prudents attendraient une large différence perceptive pour ne plus l'utiliser. La prise en compte de ce seuil de décision individuel complique inutilement l'analyse des données : si deux objets sont effectivement tels que $(T_1 = T_2)$, alors avec suffisamment de jugements et d'auditeurs la probabilité de choisir l'un ou l'autre apparaîtra égale à 50 %.

Le recueil des données de comparaison ne suppose pas forcément que les stimuli ne diffèrent que selon une seule dimension perceptive. Si plusieurs dimensions varient, la tâche revient en fait à projeter les différences entre stimuli sur un axe unidimensionnel. La nature de l'échelle perceptive n'est elle aussi pas supposée *a priori*. Par exemple, la transitivité entre jugements n'est pas indispensable. Le

jeu du puits, des ciseaux, de la pierre et de la feuille de papier en est un exemple extrême : il est simple de comparer ces objets selon la règle du jeu, et pourtant aucune échelle linéaire ne permet de les ordonner (sinon il suffirait de jouer toujours la même chose pour être sûr de gagner)! Les hypothèses sur la nature de l'échelle se font au moment de l'analyse des données.

La seule véritable hypothèse délicate liée aux comparaisons par paires est que les auditeurs jugent effectivement les stimuli selon la dimension demandée par l'expérimentateur. Les consignes expérimentales sont donc importantes, et nous nous sommes autant que possible efforcés de faire reposer ces consignes sur des exemples auditifs non verbaux dans le cas de la rugosité, ou de laisser les auditeurs aussi libres que possible dans le cas de la tension. Une indication de la compréhension des consignes se trouve dans les données recueillies : si les auditeurs n'ont pas compris la tâche, il est probable que les données ne seront constituées que de bruit. Si les données sont au contraire stables, une dimension a été dégagée et jugée. Néanmoins, ce test n'est pas infaillible : l'obtention de bruit peut résulter d'une parfaite compréhension de la consigne mais de l'égalité des stimuli sur la dimension à juger, et un résultat stable peut provenir d'un agrément entre sujets pour juger une autre dimension des stimuli que celle demandée. Cette dernière hypothèse, bien que peu probable si le nombre de sujets est élevé, ne peut être totalement écartée. Les seules précautions à prendre pour l'éviter ne sont pas d'ordre statistique mais résident dans la conception globale de l'expérience, et notamment dans le choix des stimuli.

## A.2 Analyse Bradley-Terry-Luce

### A.2.1 Principe

La méthode d'analyse des données binaires de comparaisons par paires que nous avons choisie est celle du modèle de Bradley-Terry-Luce (Bradley, 1953). Ce modèle repose sur l'hypothèse que l'échelle qui sous-tend les jugements des auditeurs est une échelle continue et linéaire : les stimuli peuvent être représentés comme des points alignés sur une droite. D'autres modèles sont dérivés de cette hypothèse, le plus connu étant celui de Thurstone et Mosteller (1927). Ces modèles sont appelés *modèles linéaires* d'analyse des jugements de comparaison par paires.

### A.2.2 Calcul des valeurs BTL

Les stimuli jugés sont considérés selon cette hypothèse comme possédant une valeur "vraie" le long de la dimension étudiée. Prenons l'exemple d'un ensemble de timbres noté $\{T_1, T_2, ..., T_N\}$. Ces timbres sont supposés posséder les valeurs vraies $\{\theta_1, \theta_2, ..., \theta_N\}$ le long d'une certaine échelle (tension, rugosité). Ce sont ces valeurs que nous désirons estimer à partir des jugements de comparaison par paires.

Dans le formalisme des modèles linéaires, lors d'une comparaison expérimentale entre deux timbres $(T_i, T_j)$ les auditeurs ne comparent pas directement les valeurs vraies $(\theta_i, \theta_j)$, mais plutôt deux variables aléatoires associées $(t_i, t_j)$ dont les distributions sont centrées sur $\theta_i$ et $\theta_j$. Le jugement $T_i > T_j$ est observé si au moment du jugement $t_i > t_j$. Ainsi, même si les valeurs vraies $\theta_i$ et $\theta_j$ ne changent pas de jugement en jugement, du fait de la distribution aléatoire qui leur est associée la comparaison de deux mêmes timbres peut produire des "inversions" de jugements. La proportion de jugements $T_i > T_j$ est influencée à la fois par la forme des distributions aléatoires associées aux valeurs vraies, des distributions plus étalées favorisant les inversions, et par l'écart absolu entre les valeurs vraies, plus ces valeurs étant proches et plus les inversions étant probables.

Les modèles linéaires font des hypothèses sur la forme des distributions pour pouvoir estimer l'écart entre valeurs vraies. Le modèle de Bradley-Terry-Luce fait l'hypothèse d'une distribution ayant la forme d'une sécante hyperbolique et de variance constante pour tous les stimuli. Cette hypothèse est proche de celle de distribution normale faite par le modèle de Thurstone. Dans le cas BTL, la probabilité $\pi_{ij}$ que $T_i$ soit jugé supérieur à $T_j$ est donnée par l'équation A.1 :

$$\pi_{ij} = 1/2[1 + \tanh(\frac{\theta_i - \theta_j}{2})] \tag{A.1}$$

Cette probabilité $\pi_{ij}$ peut être estimée à partir des proportions $p_{ij}$ de jugements observés expérimentalement où $T_i$ a effectivement été jugé supérieur à $T_j$. Cette proportion est extraite des jugements de comparaison en considérant que les paires dans un ordre et dans l'autre constituent en fait une répétition du même jugement. L'étude de l'influence possible de l'ordre de présentation doit être faite séparément (voir par exemple section 8.2.3.1).

Le remplacement des probabilités par leurs estimées dans l'équation A.1 et l'inversion de cette équation ne permet que de déterminer le rapport entre les différences de valeurs vraies. Une condition de normalisation est nécessaire, pour le modèle BTL on impose à la somme des valeurs vraies d'être nulle. Moyennant cette normalisation, on obtient la formule des valeurs vraies estimées $\widehat{\theta}_i$ pour chaque timbre :

$$\widehat{\theta}_i = log[\prod_{i \neq j}(p_{ij}/p_{ji})] \tag{A.2}$$

L'équation A.2 permet de transformer les jugements expérimentaux de comparaison binaire en une échelle linéaire qui représente le contraste perçu entre stimuli selon la dimension à juger. Les valeurs ainsi obtenues sont celles représentées dans la plupart des Figures de la partie expérimentale.

## A.3 Analyse par bootstrap

### A.3.1 Motivations du choix de la technique de bootstrap

Le résultat de l'analyse BTL est une statistique descriptive de l'échantillon à notre disposition. La question qui se pose ensuite est quelle est la probabilité que les différences observées entre les valeurs BTL représentent de "véritables" différences perceptives ?

Reformulons notre question en termes statistiques. Nous avons deux échantillons de données que nous notons $x = \{x_1, x_2, ..., x_n\}$ et $y = \{y_1, y_2, ..., y_n\}$. Ces échantillons peuvent par exemple être les jugements des sujets pour deux stimuli différents. Nous calculons une statistique décrivant ces échantillons, soit $s(x)$ et $s(y)$. Cette statistique pourrait être la valeur BTL décrite précédemment. La question que nous nous posons est que si $x$ et $y$ sont considérés comme issus des distributions $X$ et $Y$, quelle est la probabilité que $s(X) = s(Y)$ ? Cette question constitue le test de l'hypothèse nulle.

La méthode habituellement utilisée pour répondre à cette question consiste à estimer la variabilité associée à la statistique considérée. Si la distribution de la statistique $s(x)$ quand l'échantillon $x$ est tiré au hasard dans la population $X$ est normale, alors l'écart-type $\sigma_s$ détermine l'intervalle autour de la moyenne $\overline{s(x)}$ dans lequel 68 % des valeurs de $s(x)$ se retrouveront. Il est alors possible de

déterminer si la différence observée entre $s(x)$ et $s(y)$ est due au hasard ou non, ceci pour un risque donné. Le test de l'hypothèse nulle réclame donc d'estimer les écarts-type associé aux statistiques $s(x)$ et $s(y)$.

Dans le cas courant où $s(.)$ est la statistique "moyenne", une solution simple est disponible car il est possible de montrer que la variance de $s(x)$ est directement liée à celle de la population $X$ par la relation $\sigma_s^2 = \sigma_X^2/n$. La variance de la population $X$ étant estimée en calculant la variance de l'échantillon $x$, la formule bien connue de l'écart-type autour de la moyenne est alors retrouvée.

Malheureusement, si $s(.)$ n'est pas la moyenne, il n'existe en général pas de formule permettant d'estimer son écart-type. Dans le cas de l'analyse BTL notamment, les méthodes analytiques classiques ne permettent pas, à notre connaissance, d'estimer l'écart-type associé à chacun des points $s(T_i)$ (David, 1988).

### A.3.2 Principe

La méthode de bootstrap a été utilisée pour estimer l'écart-type associé à chaque valeur BTL. Le principe général de cette méthode consiste à construire une distribution empirique, à partir des données expérimentales, pour estimer l'écart-type d'une statistique $s(x)$. Cette distribution empirique, notée $X^*$, est obtenue par ré-échantillonage avec remise des données contenues dans l'échantillon $x$ (Efron et Tibshirani, 1993).

La réalisation pratique de cette méthode est très simple. A partir de l'échantillon initial $x$ un ensemble d'échantillons de bootstrap $x^*$ sont calculés en recombinant les éléments de $x$. Par exemple, un échantillon de bootstrap pourrait se présenter sous la forme $x^{*1} = \{x_1^*, x_2^*, ..., x_n^*\} = \{x_2, x_5, x_5, x_7...\}$. Il est théoriquement possible de générer toutes les combinaisons pour décrire exactement la distribution $X^*$, en pratique un tirage aléatoire de $B$ échantillons de bootstrap avec $50 < B < 200$ se révèle suffisant. Pour chaque échantillon de bootstrap, la statistique $s(x^*)$ est calculée. La variance de la distribution des B valeurs de $s(x^*)$ sert finalement à estimer l'écart-type associé à $s(x)$ :

$$\widehat{\sigma_s} = \sqrt{\frac{\sum_{b=1}^{B}[s(x_b^*) - \overline{s(x^*)}]^2}{B-1}} \tag{A.3}$$

avec

$$\overline{s(x^*)} = 1/B \sum_{b=1}^{B} [s(x_b^*)] \tag{A.4}$$

La méthode de bootstrap ne réclame donc aucune hypothèse sur la statistique $s(x)$, ni même sur la normalité de sa distribution. La distribution empirique $X^*$ est générée numériquement à partir des données de l'échantillon $x$ puis utilisée pour estimer l'écart-type de $s(x)$ par rapport à la population $X$. Ce principe a été appelé principe de branchement, ou *plug-in*. L'estimation d'une certaine formule sur la population réelle $X$ étant impossible, elle est "branchée" sur une autre population estimant la population réelle. Ce principe n'est pas spécifique au bootstrap, l'estimation de la moyenne en est un exemple : la moyenne d'une population est estimée par un "branchement" de l'opérateur "moyenne" sur la distribution constituée par l'échantillon observé.

L'originalité de la méthode de bootstrap réside uniquement dans l'obtention de la distribution empirique sur laquelle la statistique désirée est "branchée". Le nom de la méthode vient en fait de la solution trouvée par le Baron de Münchhausen pour se sauver de la noyade alors qu'il était en train de couler sous l'eau d'un lac profond : il se souleva par ses lacets de bottes et put remonter ainsi tranquillement à la surface. Cette image résume, de façon un peu provocante, l'idée de base du bootstrap. Sans créer d'information supplémentaire, la distribution empirique tirée de l'échantillon initial permet de faire des inférences sur ce dernier. Un moyen intuitif de comprendre ceci est de considérer que l'une des possibilités d'observation de la distribution originale $X$ est l'ensemble des auditeurs qui sont venus participer à l'expérience, ce qui donne l'échantillon original $x_1, x_2, ..., x_n$. Néanmoins, avec la même distribution $X$, il aurait aussi été possible que l'auditeur 1 ne vienne pas, l'auditeur 2 vienne deux fois et ainsi de suite. Le bootstrap permet de tester la variabilité des statistiques tirées de l'échantillon réel par rapport à un ensemble d'autres échantillons possibles extrapolés de cet échantillon réel.

Il a été démontré que dans les cas où les méthodes statistiques classiques s'appliquent, le bootstrap fournit des estimations équivalentes. Par contre, en cas de non normalité ou de problèmes non solubles analytiquement, des simulations numériques ont montré que le bootstrap peut encore fournir des résultats corrects si l'échantillon est relativement représentatif de la distribution originale (Efron et Tibshirani, 1993).

### A.3.3 Application aux valeurs BTL

#### A.3.3.1 Estimation de l'écart-type

Pour estimer l'écart-type des valeurs BTL, 100 échantillons de bootstrap ont été générés. Ces échantillons peuvent être imaginés comme 100 groupes d'auditeurs tirés au hasard par ré-échantillonage avec remise du groupe initial. Pour chacun des échantillons de bootstrap, l'analyse BTL a été répétée. Le résultat de ces réplications est une distribution de 100 valeurs autour des valeurs originales. Nous avons alors calculé la moyenne de ces 100 valeurs, puis l'écart-type associé. Ces statistiques sont celles représentées sur toutes les Figures de la Partie Expérimentale : un point pour l'analyse BTL originale, et l'écart-type associé des réplications bootstrap de part et d'autre de la moyenne de ces réplications. Notons que l'écart-type n'est pas symétrique autour de la valeur BTL si elle diffère de la moyenne obtenue par bootstrap. Cette asymétrie est une indication, quand elle est présente, de la non normalité des données initiales.

Cette méthode est la plus classique permettant d'estimer un intervalle de confiance par bootstrap, et nous l'avons donc utilisée autant que possible (Expériences 1, 2, 3, 4, 5). Néanmoins, elle présente un inconvénient dans le cas où la statistique à estimer est bornée à un certain intervalle, comme dans le cas d'un coefficient de corrélation (Expériences 4 et 5). Si la statistique est bornée, l'estimation de l'intervalle de confiance par les écarts-type réclamerait une étape supplémentaire de transformation des données initiales. Dans ce cas, une méthode de calcul d'intervalle de confiance par centiles est généralement préférée (Efron et Tibshirani, 1993). Moyennant un nombre d'itérations de bootstrap bien plus élevé (2000), l'intervalle de confiance est simplement défini comme un certain pourcentage des cas observés autour du cas central. Nous avons donc utilisé cette deuxième méthode pour estimer les intervalles de confiance attachés aux coefficients de corrélation (Expériences 4, 5).

#### A.3.3.2 Test de l'hypothèse nulle

Le test de l'hypothèse nulle entre deux valeurs à un risque donné $p$ peut être fait en comparant les intervalles de confiance à $1-2p$ qui leur sont attachés (Efron et Tibshirani, 1993). Si les intervalles de confiance se recouvrent, l'hypothèse nulle ne peut être rejetée. Dans le cas d'une distribution normale, l'écart-type est

utilisé pour définir les intervalles de confiance : un intervalle de $1.645\sigma$ de part et d'autre du cas central défini un intervalle de confiance à 90%. Cette estimation reste valide si l'écart-type est obtenue par bootstrap (Efron et Tibshirani, 1993). Le critère statistique que nous avons donc retenu pour estimer la significativité d'une différence entre valeurs BTL au risque $p < 0.05$ est la comparaison de cette différence avec les intervalles de confiance définis à partir des écarts-type estimés, soit $1.645\hat{\sigma}$.

## A.4 Analyse par composantes principales

### A.4.1 Principe

L'analyse par composantes principales est une technique permettant la réduction des dimensions utilisées pour décrire un ensemble de données. L'application typique de cette méthode est l'évaluation d'un certain nombre d'objets selon diverses caractéristiques. Si tous les objets sont différents et si toutes les caractéristiques sont elles aussi indépendantes entre elles, le nombre des dimensions originales est effectivement nécessaire pour représenter les données. Néanmoins, si les données sont partiellement redondantes, par exemple si deux objets ont des caractéristiques fortement corrélées, il sera avantageux de décrire les données en omettant l'un d'eux. L'analyse par composantes principales (ou ACP) permet une telle simplification en examinant la structure de la matrice de corrélation entre les caractéristiques des objets.

### A.4.2 Application à l'estimation de cohérence entre auditeurs

Nous avons appliqué l'ACP dans le but d'estimer la cohérence entre les jugements des différents auditeurs pour les Expériences 4 et 5. Pour ce faire, les "objets"considérés sont les auditeurs et les caractéristiques leurs jugements expérimentaux. Nous avons donc défini une matrice $SC$ des scores attribués aux $n$ timbres par les $m$ auditeurs. L'élément $sc_{i,j}$ de cette matrice représente le nombre de fois où le timbre $T_i$ a été choisi à travers toutes les comparaisons par paires par l'auditeur $A_j$.

$$SC = \begin{array}{c|cccc} & A_1 & A_2 & ... & A_m \\ \hline T_1 & sc_{1,1} & sc_{2,1} & ... & ... \\ T_2 & sc_{1,2} & ... & ... & ... \\ ... & ... & ... & ... & ... \\ T_n & ... & ... & ... & sc_{n,m} \end{array} \tag{A.5}$$

L'analyse par composantes principales réclame d'abord de normaliser les vecteurs $A_1, A_2, ..., A_m$ de manière à ce qu'ils aient une moyenne nulle et une variance unité. La matrice de covariance $C$ des vecteurs ainsi normalisés est calculée : elle est égale à la matrice de corrélation entre les auditeurs "avant normalisation". Les vecteurs propres $V$ de la matrice $C$ sont ensuite calculés. Ces vecteurs sont orthogonaux, ce qui signifie qu'ils représentent chacun une dimension indépendante des données. De plus, la valeur propre qui leur est associée représente leur variance (Hotelling, 1933).

Prenons l'exemple du vecteur propre ayant la plus forte variance, noté $V_1 = (v_{11}, v_{12}, ...v_{1m})$. La première composante principale $CP_1$ peut alors être définie comme une combinaison linéaire des auditeurs :

$$CP_1 = v_{11}A_1 + v_{12}A_2 + ... + v_{1m}A_m \tag{A.6}$$

$CP_1$ représente une sorte "d'auditeur type", qui reflète la stratégie de jugement majoritaire parmi les auditeurs. $CP_1$ fournit le vecteur des scores de chacun des timbres selon cette stratégie type ; elle reflète la dimension la plus importante extraite de la variation des résultats. Ceci peut être rapporté à la technique de comparaison par paires où l'on demande aux auditeurs de "mentalement" projeter les différences entre objets sur une seule dimension (David, 1988). Un agrément éventuel entre l'échelle BTL et $CP_1$ constitue donc une indication que la consigne a été bien appliquée par les auditeurs.

Un autre renseignement fourni par $CP_1$ est la "composition"de l'auditeur type. Si tous les auditeurs sont également représentés, c'est-à-dire si les $v_{1i}$ sont de même ordre de grandeur, ceci signifie qu'un agrément s'est dégagé pour le jugement des timbres selon $CP_1$. Par contre, si les $v_{1i}$ sont d'ordre de grandeur différent, ou même de signe opposé, ceci signifie un désaccord entre les auditeurs pour juger les timbres selon $CP_1$.

Les mêmes interprétations peuvent être faites avec les autres vecteurs propres, chacun définissant une stratégie de jugement. Néanmoins, l'intérêt de l'ACP est de limiter si possible le nombre de composantes : si la variance associée à un vecteur propre est faible, la composante en question n'apporte que peu d'information sur les données. Le critère d'élimination des composantes n'est pas imposé par l'ACP, mais une règle classique consiste à ne conserver que les composantes qui expliquent plus de variance que l'une des variables originales prise indépendamment. S'il y a une redondance entre les données initiales, autrement dit dans notre cas si tous les auditeurs n'ont pas répondu indépendamment les un des autres, un nombre limité de composantes doit pouvoir expliquer la plus grande partie de la variance initiale.

## A.5 Conclusion

Les détails concernant la méthode expérimentale et les analyses statistiques des données ont été regroupés dans cette annexe. Les algorithmes décrits ont été implémentés informatiquement pour permettre leur application pratique. Le protocole général a été maintenu identique autant que possible à travers toutes les expériences réalisées.

# Annexe B

# Implémentation d'un modèle d'estimation de rugosité

## B.1 Introduction

Un modèle d'estimation de rugosité basé sur l'hypothèse temporelle a été implémenté, de manière à comparer les prédictions de cette approche avec certains de nos résultats expérimentaux. Il est notamment employé au Chapitre 7 pour opposer entre elles les différentes modalités de combinaisons de rugosités partielles.

## B.2 Description du modèle

Le son à analyser est représenté par ses variations de pression au cours du temps. Du fait de l'implémentation entièrement dans le domaine temporel, le calcul peut être fait sur tout ou une partie du son. Le principe du modèle, dont nous allons maintenant préciser les paramètres, est résumé en Figure B.1.

Le signal original est d'abord filtré passe-bande entre 450 et 1000 Hz, avec une atténuation de 6 dB/oct hors de la bande passante. Une décomposition en sous-bandes fréquentielles est ensuite réalisée par un banc de filtres gammatone. Les fréquences centrales des filtres ainsi que leurs largeurs de bande sont ajustées à l'échelle des ERB (Glasberg et Moore, 1990; Patterson, Allerhand et Giguère, 1995). Ces deux premières étapes sont réalisées dans la plateforme logicielle DSAM (O'Mard, Hewitt et Meddis, 1994).

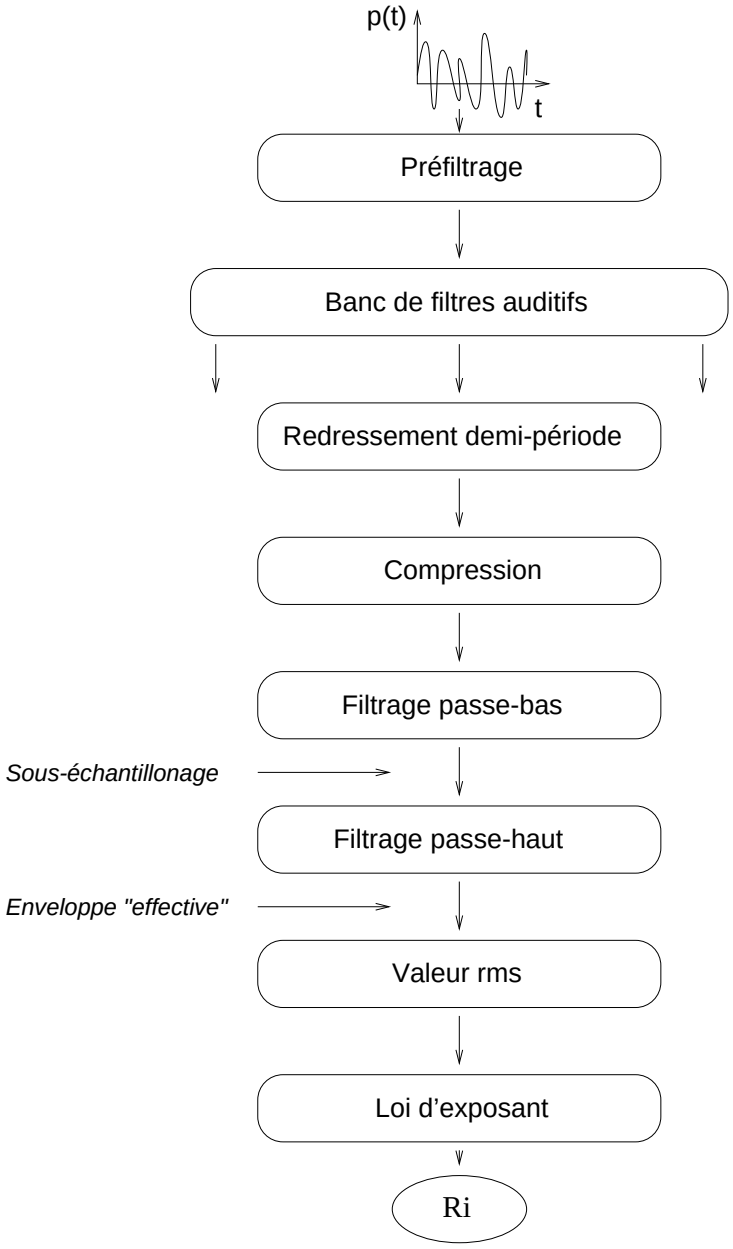


FIG. B.1 – Principe du modèle de rugosité. Le modèle implémente de façon directe les hypothèses de l'approche temporelle.

Chaque sous-bande, représentant un "canal" auditif, est ensuite traitée en parallèle par un programme Matlab. Le signal présent dans un canal donné est redressé demi-période (seule sa partie positive est conservée). Ce redressement est suivi d'une compression de type racine carrée (loi d'exposant à 0.5).

Un filtrage passe-bas est ensuite appliqué. Le filtre choisi est un filtre de Butterworth d'ordre 2 (12 dB/oct) de fréquence de coupure 70 Hz. Ce filtrage passe-bas étant appliqué à un signal à bande étroite (car issu d'un filtre auditif) redressé, il a pour conséquence l'extraction d'une enveloppe. La composition spectrale de cette enveloppe est limitée du fait du filtrage : les fréquences supérieures à 630 Hz sont atténuées de plus de 96 dB. La fréquence d'échantillonage originale du signal étant fixée à 44.1 kHz, nous avons décidé de sous-échantilloner cette enveloppe dans un facteur de 1/35 (nouvelle fréquence d'échantillonage de 1260 Hz). Cette manipulation permet de gagner en temps de calcul, et elle améliore la stabilité numérique du filtrage de l'étape suivante.

Un filtrage passe-haut est ensuite appliqué à l'enveloppe sous-échantillonée. Un filtre de Butterworth d'ordre 2 a été choisi, de fréquence de coupure 70 Hz. La combinaison de ce filtrage et du précédent fournit une "enveloppe efficace", où seules les fréquences de modulation contribuant à la rugosité (selon l'hypothèse temporelle) sont préservées.

La rugosité partielle du canal considéré est finalement obtenue en calculant la valeur $rms$ de l'enveloppe efficace, puis en élevant cette valeur à l'exposant 2. La combinaison des rugosités partielles à travers les canaux peut alors se faire de diverses manières, qui sont discutées au Chapitre 7.

## B.3 Évaluation

Une évaluation du modèle peut être réalisée en considérant ses prévisions pour la rugosité de sons SAM, les stimuli de base de l'approche temporelle. Le calcul a été fait pour un seul filtre centré sur la fréquence centrale du stimulus. Le résultat de ces simulations est présenté en Figure B.2. L'accord avec les résultats expérimentaux est qualitativement satisfaisant (Terhardt, 1970; Zwicker et Fastl, 1990; Daniel et Weber, 1997). Notamment, l'augmentation de rugosité avec la fréquence de modulation et l'entrée en jeu des deux facteurs limitant que sont la

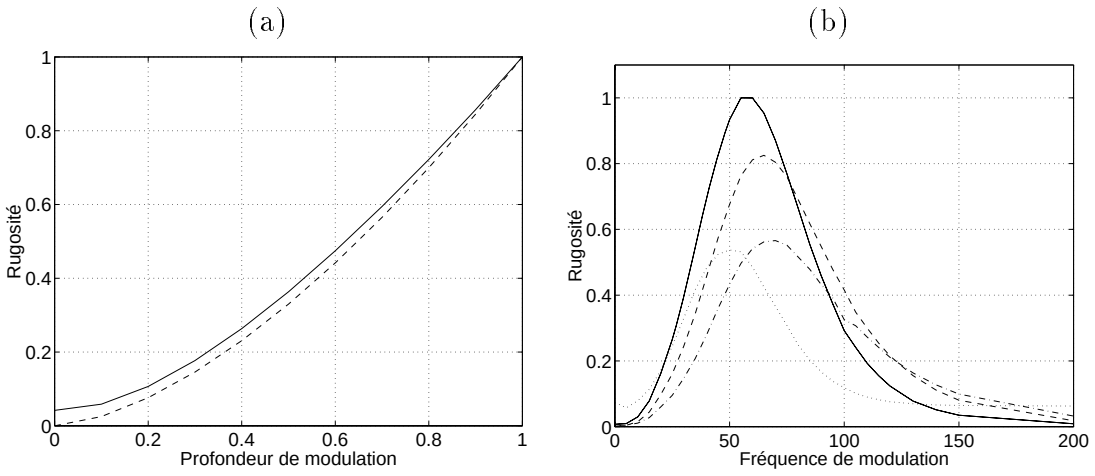


FIG. B.2 – Évaluation du modèle. (a) : rugosité calculée d'un son SAM de fréquence centrale 1000 Hz et de fréquence de modulation 70 Hz en fonction de la profondeur de modulation. En hachuré : la loi $m^{1,6}$ proposée par Zwicker et Fastl (1990). (b) : rugosité calculée de sons SAM à différentes fréquences centrales en fonction de la fréquence de modulation. Fréquence centrale de 1000 Hz : trait plein, 500 Hz : trait pointillé, 2000 Hz : trait hachuré, 4000 Hz : trait alterné.

bande critique et la caractéristique passe-bas de la modulation d'amplitude sont correctement reproduites (section 2.4).

## B.4 Remarques sur les différentes étapes

Le modèle que nous proposons diffère dans son implémentation de celui de Aures (1985b) et Daniel et Weber (1997), présenté en section 3.2.3. Pourtant, leurs principes sont similaires à quelques simplifications près.

Nous appliquons un pré-filtrage au signal avant de le décomposer en canaux fréquentiels. Ce filtrage est similaire dans son principe à la compensation des effets de l'audiogramme. Les fréquences de coupures choisies permettent de reproduire l'influence de la région fréquentielle sur la rugosité, et remplacent les coefficients correctifs appliqués en fin de calcul par Aures (1985b).

La simulation du filtrage auditif adoptée est linéaire. Elle représente une simplification efficace en termes de temps de calcul[1] et a été largement employée par d'autres auteurs (Meddis et Hewitt, 1991; Patterson, Allerhand et Giguère, 1995).

1. Le temps de calcul du modèle dans son ensemble est de l'ordre d'une dizaine de fois le temps réel sur une station Sun Sparc 10, pour une trentaine de canaux.

Cette simplification limite néanmoins notre implémentation à une utilisation à des niveaux modérés, pour lesquels les effets d'élargissement des filtres peuvent être négligés.

Aucune étape de compression n'est explicitement décrite dans le modèle de Aures. Néanmoins, le calcul des profondeurs de modulations efficaces est fait en normalisant la valeur *rms* de l'enveloppe par sa valeur continue, puis en "écrêtant" la valeur obtenue. Ceci revient en fait à appliquer une forte compression. Les deux étapes de redressement demi-période et de compression que nous avons choisies permettent de reproduire qualitativement ceci, en limitant par exemple l'influence du niveau sur la rugosité calculée. De plus, ces étapes ont été proposées par plusieurs auteurs comme une schématisation du résultat net des différentes compressions subies par un signal sonore du fait de divers mécanismes auditifs (Oxenham et Moore, 1994; Allen, 1996).

La forme du filtrage que nous avons choisie pour extraire l'enveloppe efficace est une manière directe d'implémenter les suggestions de Terhardt (1974a). Le nombre de paramètres s'en trouve réduit, au détriment de la possibilité d'ajustement aux données expérimentales. Ceci rend le modèle quantitativement moins précis que les implémentations de Aures (1985b) et Daniel et Weber (1997), sans toutefois changer la forme générale des prédictions.

Finalement, l'élévation de la valeur *rms* au carré est commune aux deux implémentations (Daniel et Weber, 1997).

## B.5 Conclusion

Un modèle simple de calcul de rugosité a été réalisé. Ce modèle peut être vu comme une implémentation "minimale" des hypothèses de l'approche temporelle. Son principe est très proche celui des modèles existants (Aures, 1985b; Daniel et Weber, 1997). Toutefois, le choix a été fait de limiter au maximum le nombre de paramètres pour se concentrer sur les différents mécanismes mis en jeu. Ceci a pour conséquence des prédictions moins précises pour les données expérimentales existantes. Cet outil est principalement destiné dans le cadre de cette thèse à évaluer les prédictions faites par l'approche temporelle — pour éclairer les raisons qui font que cette approche échoue parfois.

# Annexe C

# Liste des abbréviations utilisées

**ACP** Analyse par Composantes Principales. Méthode d'analyse statistique permettant de réduire le nombre de dimensions utilisées pour décrire un ensemble de données, décrite en Annexe A.

**Bark** Mesure de la sélectivité fréquentielle du système auditif introduite par Zwicker (1957). La largeur de la bande critique à la fréquence $f$ en Hz est selon cette mesure $Bark = 25 + 75(1 + 1,4f^2)^{0,69}$.

**BTL** Bradley-Terry-Luce. Méthode d'analyse statistique permettant de transformer des comparaisons par paires en une échelle de valeurs, décrite en Annexe A.

**CDT** Cubic Distortion Tone. Son de distorsion produit par des non linéarités liées au traitement auditif périphérique, décrit en section 5.3.2.

**dBA** Décibel A. Echelle logarithmique de mesure de pression correspondant aux dB SPL pondérés par la courbe d'isosonie à 60 phones. Cette échelle vise à traduire le fait qu'une même énergie acoustique ne produit pas la même intensité subjective selon la région fréquentielle où elle est présente.

**dB SPL** Décibel SPL. Echelle logarithmique de mesure de pression dont l'origine est fixée à 20 micro-Pascal.

**DSAM** Development System for Auditory Modelling. Plateforme logicielle destinée à faciliter l'élaboration de modèles des étapes de traitement auditif

périphérique. Pour plus d'informations et éventuellement télécharger le logiciel, voir:
`http://www.essex.ac.uk/psychology/hearinglab/index.html`

**ERB** Equivalent Rectangular Bandwidth. Mesure de la sélectivité fréquentielle du système auditif selon l'analogie avec un filtrage passe-bande. Une largeur de 1 ERB correspond à la largeur d'un filtre rectangulaire de gain unité passant autant de puissance que le "filtre" auditif centré à la même fréquence. La largeur de bande critique à la fréquence $f$ en Hz est donnée par $ERB = 24,7 + 0,108f$.

**FM** Frequency Modulated. Son modulé en fréquence.

**Hz** Hertz. Mesure de fréquence nommée selon Hertz, élève de Helmholtz.

**jAM** jittered-Amplitude Modulated. Désigne ici un son dérivé d'un son SAM mais dont la période de modulation est perturbée aléatoirement période-par-période. Voir description en section 7.2.1.

**pAM** pseudo-Amplitude Modulated. Désigne ici un son dont le spectre correspond à celui d'un son SAM avec une profondeur de modulation de 100 % mais dont le spectre de phase est manipulé. Voir description en section 5.2.1.

**QFM** Quasi Frequency Modulated. Son dont le spectre d'amplitude est celui d'un son SAM mais dont le spectre de phase est celui d'un son FM.

**SAM** Sinusoidally Amplitude Modulated. Son sinusoïdalement modulé en amplitude. Voir description en section 2.4.1.

# Références bibliographiques

# Bibliographie

Allen, J. B. (1996). "A review of active and passive basilar membrane cochlear mechanics", *J. Acoust. Soc. Am.* **99**, 2582.

Assayag, G. et Cholleton, J. P. (1995). "La musique et les nombres", *La Recherche* **26**(278), 804–810.

Aures, W. (1984). *Berechnugsverfahren für den Wolklang bebleibiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse.* PhD thesis, Technishes Universität München.

Aures, W. (1985a). "Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale [Calculation method of sensory euphony of sound signals]", *Acustica* **59**, 130–141.

Aures, W. (1985b). "Ein Berechnungsverfahren der Rauhigkeit [A roughness calculation method]", *Acustica* **58**, 268–281.

Bailhache, P. (1986). "Valeur actuelle de l'acoustique musicale de Helmholtz", *Rev. Hist. Sci.* **XXXIX**(4), 302–324.

von Békésy, G. (1960). *Experiments in Hearing.* New York: McGraw-Hill.

Bharucha, J. et Krumhansl, C. L. (1983). "The representation of harmonic structure in music: Hierarchies of stability as a function of context", *Cognition* **13**, 63–102.

Bharucha, J. J. et Olney, K. L. (1989). "Tonal cognition, artificial intelligence and neural nets", *Contemporary Music Review* **4**(1), 341–356.

Bigand, E. (1989). *Perception et Compréhension des Phrases Musicales.* PhD thesis, Université Paris X Nanterre.

Bigand, E. (1993). "The influence of implicit harmony, rhythm and musical training on the abstraction of tension-relaxation schemas in tonal music phrases", *Contemporary Music Review* **9**, 123–137.

Bigand, E., Parncutt, R. et Lerdahl, F. (1996). "Perception of musical tension in short chord sequences : the influence of harmonic function, sensory dissonance, horizontal motion, and musical training", *Perception and Psychophysics* **58**(1), 125–141.

von Bismarck, G. (1974). "Timbre of steady sounds: a factorial investigation of its verbal attribute", *Acustica* **30**, 146–159.

Björk, E. A. (1985). "The perceived quality of natural sounds", *Acustica* **57**, 186–188.

Bradley, R. A. (1953). "Some statistical methods in taste testing and quality evaluation", *Biometrics* **9**, 22–39.

Bregman, A. S. (1990). *Auditory Scene Analysis : The Perceptual Organisation of Sound.* Cambridge, MA: M.I.T Press.

Bregman, A. S. (1991). "Timbre, orchestration, dissonance, et organisation auditive", In Barrière, J. B. (Ed.), *Le timbre, Métaphore pour la Composition* (pp. 204–215). Paris: Christian Bourgois.

Bregman, A. S., Abramson, J., Doehring, P. et Darwin, C. J. (1985). "Spectral integration based on common amplitude modulation", *Perception and Psychophysics* **37**, 483–493.

Bregman, A. S., Levitan, R. et Liao, C. (1990). "Fusion of auditory components: Effects of the frequency of modulation", *Perception and Psychophysics* **47**(1), 68–73.

Butler, J. W. et Daston, P. G. (1963). "Musical consonance as musical preference: a cross-cultural study", *Journal of General Psychology* **79**, 129–142.

Buunen, T. J. F. (1975). "Two hypotheses on monaural phase effects", *Acustica* **34**, 98–105.

Buunen, T. J. F. (1976). *On the perception of phase differences in acoustic signals.* PhD thesis, Technische Hogeschool Delft.

Buunen, T. J. F., Festen, J. M., Bilsen, F. A. et van den Brink, G. (1974). "Phase effects in a three-component signal", *J. Acoust. Soc. Am.* **55**(2), 297–303.

Carlyon, R. P. (1996). "Spread of excitation produced by maskers with damped and ramped envelopes", *J. Acoust. Soc. Am.* **99**(6), 3647–3655.

Carlyon, R. P. et Datta, A. J. (1997a). "Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system", *J. Acoust. Soc. Am.* **101**(6), 3636–3647.

Carlyon, R. P. et Datta, A. J. (1997b). "Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components", *J. Acoust. Soc. Am.* **101**(6), 3648–3657.

de Cheveigné, A., McAdams, S., Laroche, J. et Rosenberg, M. (1995). "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement", *J. Acoust. Soc. Am.* **97**(6), 3736–3748.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, Mass.: M.I.T. Press.

Cross, C. R. et Goodwin, H. M. (1893). "Some considerations regarding Helmholtz's theory of consonance", *Proc. Acad. Sci. New Ser.* **19**, 1–12.

Daniel, P. et Weber, R. (1997). "Psychoacoustical Roughness: Implementation of an Optimized Model", *Acustica* **83**, 113–123.

Dau, T., Püschel, D. et Kohlrausch, A. (1996). "A quantitative model of the "effective" signal processing in the auditory system: I. Model structure.", *J. Acoust. Soc. Am.* **99**, 3615–3622.

David, H. A. (1988). *The Method of Paired Comparisons* (2nd. Ed.). Oxford University Press.

Davis, H. (1983). "An active process in cochlear mechanics", *Hearing Research* **9**(1), 79–90.

Demany, L. (1989). *Psychoacoustique et Perception Auditive*, Chapitre Perception de la hauteur tonale, (pp. 44–81). Paris: INSERM/EMI/SFA.

DeWitt, L. A. et Crowder, R. G. (1987). "Tonal fusion of consonant intervals: the oompf in Stumpf", *Perception and Psychophysics* **41**(1), 73–84.

Dibben, N. (1994). "The cognitive reality of hierarchical structures in tonal and atonal music", *Music Perception* **12**(1), 1–25.

Dibben, N. (1996). "Perceptual facilities in atonal music: Implications for the listening experience", In *International Conference on Music Perception and Cognition* (pp. 55–58). Montréal, Canada.

Dowling, W. J. et Harwood, D. L. (1986). *Music Cognition.* London: Academic Press.

Drullman, R., Festen, J. M. et Plomp, R. (1994). "Effect of temporal smearing on speech perception", *J. Acoust. Soc. Am.* **95**, 1053–1064.

Efron, B. et Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* London: Chapman & Hall.

Egan, J. P. et Hake, H. W. (1950). "On the masking pattern of a simple auditory stimulus", *J. Acoust. Soc. Am.* **22**, 622–630.

Fastl, H. (1977). "Roughness and temporal masking patterns of sinusoidally amplitude modulated broadband noise", In Evans, E. F. et Wilson, J. P. (Eds.), *Psychophysics and Physiology of Hearing* (pp. 403–415). London: Academic Press.

Fastl, H. (1997). "The psychoacoustics of Sound-Quality Evaluation", *Acustica* **83**. 754-764.

Faure, A. (1994). "Analyses de verbalisations obtenues en comparant des timbres musicaux". Master's thesis, DEA Sciences Cognitives, EHESS.

Fay, R. R. (1994). "Perception of temporal acoustic patterns by the goldfish (Carassius auratus)", *Hearing Research* **76**, 158–172.

Fay, R. R. (1996). "Perception of two-tone mixture by the goldfish", *J. Acoust. Soc. Am.* **100**(4), 2594.

Fechner, G. T. (1860). *Elements of Psychophysics.* London: Holt, Rinehart & Winston. H. E. Adler, Trans.

Fechner, G. T. (1879). *Die Tagesansicht gegenüber der Nachtansicht [Opinions diurnes et nocturnes sur le monde].* Leipzig: Breitkopf & Härtel.

Fichet, L. (1995). *Les Théories scientifiques de la musique aux XIXe et XXe siècles.* Paris: Librairie J. Vrin.

Fletcher, H. (1940). "Auditory patterns", *Rev. Mod. Phys.* **12**, 47–65.

Frisina, R. D., Smith, R. L. et Chamberlain, S. C. (1990). "Encoding of amplitude modulation detection in the gerbil cochlear nucleus: I. A hierarchy of enhancement", *Hearing Research* **44**, 90–122.

Geary, J. M. (1980). "Consonance and dissonance of pairs of inharmonic sounds", *J. Acoust. Soc. Am.* **67**(5), 1785–1789.

Giguère, C. et Woodland, P. C. (1994). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path", *J. Acoust. Soc. Am.* **95**(1), 331–342.

Glasberg, B. R. et Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data", *Hearing Research* **47**, 103–138.

Goldstein, J. L. (1967a). "Auditory nonlinearity", *J. Acoust. Soc. Am.* **41**(3), 676–689.

Goldstein, J. L. (1967b). "Auditory spectral filtering and monaural phase perception", *J. Acoust. Soc. Am.* **41**(2), 458–479.

Goldstein, J. L. (1970). "Aural combination tones", In Plomp, R. et Smoorenburg, G. (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 230–245). Leiden: A. W. Sijthoff.

Greenwood, D. D. (1961). "Critical bandwidth and the frequency coordinates of basilar membrane", *J. Acoust. Soc. Am.* **33**, 1344–1356.

Greenwood, D. D. (1991a). "Critical bandwidth and consonance in relation to cochlear frequency-position coordinates", *Hearing Research* **54**, 164–208.

Greenwood, D. D. (1991b). "Critical bandwidth and consonance: Their operational definitions in relation to cochlear nonlinearity and combination tones", *Hearing Research* **54**, 209–246.

Grey, J. M. (1977). "Multidimensional perceptual scaling of musical timbres", *J. Acoust. Soc. Am.* **61**, 1270–1277.

Guernesey, M. (1928). "The rôle of consonance and dissonance in music", *Am. J. Psychol.* **40**, 173–204.

Guirao, M. et Garavila, J. M. (1976). "Perceived roughness of amplitude modulated tones and noise", *J. Acoust. Soc. Am.* **60**(6), 1335–1338.

Guthrie, E. R. et Morril, H. (1928). "The fusion of non-musical intervals", *Am. J. Psychol.* **40**, 624–625.

Guttman, N. et Julesz, B. (1963). "Lower limits of auditory periodicity analysis", *J. Acoust. Soc. Am.* **35**, 610.

Hall, J. W. I. et Grose, J. H. (1988). "Comodulation masking release: Evidence for multiple cues", *J. Acoust. Soc. Am.* **84**, 1669–1675.

Hanappe, P. (1996). "Intégration des représentations Temps/Fréquence et des représentations musicales symboliques", In *Acte des Journées d'Informatique Musicale*. Caen, France.

Hartmann, W. M. (1997). *Signals, Sound, and Sensation*. Woodbury, New York: AIP Press.

Hashimoto, T. et Hatano, S. (1994). "Roughness level as a measure for estimating unpleasantness: modification of roughness level by modulation frequencies", In *Inter-Noise 94* (pp. 887–892).

Hellegouarch, Y. (1992). "L' "Essai d'une nouvelle théorie de la musique" de Leonhard Euler", In *Quadrivium : Musiques et sciences*. La Villette, Paris: IPMC.

von Helmholtz, H. L. F. (1877). *On the Sensations of Tone as the Physiological Basis for the Theory of Music.* 2nd. Ed. trans. A. J. Ellis (1885), from German 4th Ed., Dover, New York (1954).

Hewitt, M. J. et Meddis, R. (1993). "Regularity of cochlear nucleus stellate cells: A computational modeling study", *J. Acoust. Soc. Am.* **93**(6), 3390–3398.

Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology* **24**, 417–441, 498–520.

Hulse, S. H., Bernard, D. J. et Braaten, R. F. (1995). "Auditory discrimination of chord-based spectral structures by European starlings (Sturnus Vulgaris)", *J. Exp. Psych.: Gen* **124**(4), 409–423.

Hulse, S. H., Takeuchi, A. H. et Braaten, R. F. (1992). "Perceptual invariants in the comparative psychology of music", *Music Perception* **10**, 151–184.

Hutchinson, W. et Knopoff, L. (1978). "The acoustic component of Western consonance", *Interface* **7**, 1–29.

James, J. (1993). *The music of the spheres. Music, science and the natural order of the universe.* New York: Grove Press.

Kaestner, G. (1909). "Untersuchungen über den Gefühlseindruck unanalysierter Zweiklänge", *Psychol. Studien* **4**, 473–504.

Kameoka, A. et Kuriyagawa, M. (1969a). "Consonance theory part I", *J. Acoust. Soc. Am.* **45**, 1451–1458.

Kameoka, A. et Kuriyagawa, M. (1969b). "Consonance theory part II", *J. Acoust. Soc. Am.* **45**, 1459–1469.

Kemp, S. (1982). "Roughness of Frequency-Modulated Tones", *Acustica* **50**(3), 126–133.

Knobloch, E. (1992). "Rapports historiques entre musique, mathématique et cosmologie", In *Quadrivium : Musiques et sciences* (pp. 123–142). La Villette, Paris: IPMC.

Koffka, K. (1935). *Principles of Gestalt Psychology.* New York: Harcourt, Brace, & World.

Kohlrausch, A. et Sander, A. (1995). "Phase effects in masking related to dispersion in the inner ear. II. Masking period pattern of short sound targets", *J. Acoust. Soc. Am.* **97**(3), 1817–1829.

Krumhansl, C. L. (1983). "Perceptual structures for tonal music", *Music Perception* **1**, 28–62.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch.* Oxford: Oxford University Press.

Krumhansl, C. L., Bharucha, J. et Castellano, M. A. (1982). "Key distance effects on perceived harmonic structure in music", *Perception and Psychophysics* **32**, 96–108.

Krumhansl, C. L. et Keil, E. C. (1982). "Acquisition of the hierarchy of tonal functions in music", *Memory & Cognition* **10**, 243–251.

Krumhansl, C. L. et Kessler, E. J. (1982). "Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys", *Psychological Review* **89**, 334–368.

Krumhansl, C. L. et Shepard, R. N. (1979). "Quantification of the hierarchy of tonal functions within a diatoni context", *Journal of Experimental Psychology: Human Perception and Performance* **5**, 579–594.

Langner, G. (1992). "Periodicity coding in the auditory system", *Hearing Research* **60**, 115–142.

Lerdahl, F. (1988). "Tonal pitch space", *Music Perception* **5**, 315–345.

Lerdahl, F. (1989). "Atonal prolongational structure", *Contemporary Music Review* **4**, 65–87.

Lerdahl, F. et Jackendoff, R. (1983). *A Generative Theory of Tonal Music.* Cambridge, Mass.: M.I.T Press.

Licklider, J. C. R. (1954). "Periodicity and place pitch.", *J. Acoust. Soc. Am.* **26**, 945.

Licklider, J. C. R., Webster, J. C. et Hedlun, J. M. (1950). "On the frequency limits of binaural beats", *J. Acoust. Soc. Am.* **22**, 468–473.

Lindemann, E., Dechelle, F., Smith, B. et Starkier, M. (1991). "The Architecture of the IRCAM Musical Workstation", *Computer Music Journal* **15**(3), 41–49.

Lorenzi, C., Micheyl, C. et Berthommier, F. (1995). "Neural correlate of amplitude modulation detection", *Hearing Research* **90**, 219–227.

Malmberg, C. F. (1918). "The perception of consonance and dissonance", *Psychol. Monog.* **25**, 93–133.

Mandler, G. (1984). *Mind and Body.* New York: Norton.

Mathes, R. C. et Miller, R. L. (1947). "Phase effects in monaural phase perception", *J. Acoust. Soc. Am.* **19**, 780–797.

Mathews, M. V. et Pierce, J. R. (1980). "Harmony and nonharmonic partials", *J. Acoust. Soc. Am.* **68**(5), 1252–1257.

Mathews, M. V., Pierce, J. R. et Roberts, L. A. (1987). "Harmony and new scale", In Sundberg, J. (Ed.), *Harmony and Tonality* (pp. 59–84). Royal Swedish Academy of Music.

Mayer, A. M. (1874). "Art. XXI. Researches in Acoustics - No. 6.", *Am. Jour. Sci. - 3rd Series* **8**, 242–255.

Mayer, A. M. (1875). "Art. XXVIII. A redetermination of the constants of the law connecting the pitch of a sound with the duration of its residual sensation.", *Am. Jour. Sci. - 3rd Series* **9**, 267–269.

Mayer, A. M. (1894). "Researches in acoustics IX", *Phil. Mag.* **Series 5**(37), 259–288.

McAdams, S. (1984). "The Auditory image: A metaphor for musical and psychological research on auditory organisation", In Crozier, W. R. et Chapman,

A. J. (Eds.), *Cognitive Processes in the Perception of Art* (pp. 289–324). Amsterdam: North Holland.

McAdams, S. (1989). "Psychological constraints on form bearing dimensions in music", *Contemporary Music Review* **4**, 181–198.

McAdams, S. et Bigand, E. (Eds.). (1994). *Penser les sons.* Paris: Presses Universitaires de France.

McAdams, S. et Cunibile, J. (1992). "Perception of timbral analogies", *Philosophical Transactions of the Royal Society* **series B**(336), 383–389.

McAdams, S. et Rodet, X. (1988). "The role of FM-induced AM in dynamic spectral profile analysis", In *Basic Issues in Hearing.* London: Academic Press.

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.", *Psychological Research* **58**, 177–192.

Meddis, R. (1986). "Simulation of mechanical to neural transduction in the auditory receptor", *J. Acoust. Soc. Am.* **79**, 702–711.

Meddis, R. (1988). "Simulation of auditory-neural transduction: Further studies", *J. Acoust. Soc. Am.* **83**, 1056–1063.

Meddis, R. et Hewitt, M. J. (1991). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I : pitch identification", *J. Acoust. Soc. Am.* **89**(6), 2866–2882.

Meyer, C. (1992). "Mathématique et musique au Moyen Age", In *Quadrivium : Musiques et sciences* (pp. 107–121). La Villette, Paris: IPMC.

Meyer, L. B. (1956). *Emotion and Meaning in Music.* Toronto: Chicago Press.

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing, 4th ed.* London: Academic Press.

Moore, B. C. J. et Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns", *J. Acoust. Soc. Am.* **74**(3), 750–753.

de la Motte-Haber, H. (1994). *Principales théories scientifiques en psychologie de la musique : les paradigmes*, Chapitre I, (pp. 27–53). Paris: P.U.F.

Murail, T. (1984). "Spectra and Pixies", *Contemporary Music Review* **1 (1)**, 157–171.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures. The implication realisation model.* Chicago: University of Chicago Press.

O'Mard, L. P., Hewitt, M. J. et Meddis, R. (1994). *LUTEar, a flexible auditory simulation development computing system.* Loughborough: Loughborough University of Technology.

Osgood, C. E., Suci, X. et Tannenbaum, X. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Oxenham, A. J. et Moore, B. C. J. (1994). "Modeling the additivity of nonsimultaneous masking", *Hearing Research* **80**, 105–118.

Paraskeva, S. et McAdams, S. (1997). "Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas", In *Proceedings ICMC 1997* (pp. 438–441).

Parncutt, R. (1989). *Harmony: a psychoacoustical approach.* Berlin: Springer-Verlag.

Parncutt, R. et Strasburger, H. (1994). "Applying psychoacoustics in composition: "harmonic" progressions of "non-harmonic" sonorities", *Perspectives of New Music* **32**(2), 88–129.

Patterson, R. D. (1976). "Auditory filter shape derived with noise stimuli", *J. Acoust. Soc. Am.* **59**, 640–654.

Patterson, R. D. (1994a). "The sound of a sinusoid: Spectral models", *J. Acoust. Soc. Am.* **96**, 1409–1418.

Patterson, R. D. (1994b). "The sound of a sinusoid: Time-interval models", *J. Acoust. Soc. Am.* **96**, 1419–1428.

Patterson, R. D., Allerhand, M. H. et Giguère, C. (1995). "Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform", *J. Acoust. Soc. Am.* **98**, 1890–1894.

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing* (2nd. Ed.). London: Academic Press.

Pierce, J. R. (1966). "Attaining consonance in arbitrary scales", *J. Acoust. Soc. Am.* **40**, 249.

Pierce, J. R. (1984). *Le son musical.* Paris: Pour la science, Belin. Traduit de l'édition américaine de 1983.

Plomp, R. (1970). "Timbre as a multidimensional attribute of complex tones", In Plomp, R. et Smoorenburg, G. (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 397–414). Leiden: A. W. Sijthoff.

Plomp, R. (1976). *Aspects of Tone Sensation.* London: Academic Press.

Plomp, R. et Levelt, W. (1965). "Tonal consonance and critical bandwidth", *J. Acoust. Soc. Am.* **38**, 548–560.

Plomp, R. et Steeneken, H. J. M. (1968). "Interference between two simple tones", *J. Acoust. Soc. Am.* **43**(4), 883–884.

Pressnitzer, D. (1994). "Rugosité auditive de sons complexes : perception, physiologie et modélisation.". Master's thesis, DEA Atiam, Université Paris VI, Paris.

Pressnitzer, D. et McAdams, S. (1998). "Acoustics, psychoacoustics and spectral music", *Contemporary Music Review* (p. in press).

Pressnitzer, D., McAdams, S., Winsberg, S. et Fineberg, J. (1996). "Roughness and musical tension of orchestral timbres", In *Proceedings of the 4th International Conference on Music Perception and Cognition* (pp. 85–90). Montréal, Canada.

Rakowski, A. (1992). "Sensory dissonance as a dimension of timbre in music", In *Proceedings of the Second International Conference on Music Perception and Cognition* (p.$\tilde{6}$5). UCLA, Los Angeles.

Rameau, J. P. (1722). *Traité de l'harmonie réduite à ses principes naturels [Treatise on harmony reduced to its natural principles].* Trans. by P. Gosset (1971), Dover, New York.

Richards, V. (1987). "Monaural envelope correlation perception", *J. Acoust. Soc. Am.* **82**(5), 1621–1630.

Risset, J. C. (1988). "Perception, environnement, musiques", *InHarmoniques* **3**.

Risset, J. C. (1994). *Quelques aspects du timbre dans la musique contemporaine*, Chapitre III, (pp. 87–114). Paris: P.U.F.

Scharf, B. (1970). "Critical bands", In Tobias, J. V. (Ed.), *Foundations of Auditory Theory*, Volume 1. London: Academic press.

Schellenberg, E. G. et Trainor, L. J. (1996). "Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: Test of adult and infant listeners", *J. Acoust. Soc. Am.* **100**(5), 3321–3328.

Schenker, H. (1935). *Der Freie Satz.* Vienna: Universal Edition. Trans. E. Oster (1979) *Free Composition*, New York: Longman.

Schoenberg, A. (1911). *Harmonielehre [Theory of harmony].* trans. R. E. Carter, from German 3rd ed. (1922), Faber and Faber Ltd., London (1978).

Schooneveldt, G. P. et Moore, B. C. J. (1987). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band", *J. Acoust. Soc. Am.* **82**(6), 1944–1956.

Schouten, J. F. (1940). "The residue and the mechanism of hearing", *Proc. K. Ned. Akad. Wet.* **43**, 991–999.

Schroeder, M. R. (1970). "Synthesis of low peak-factor signals and binary sequences with low autocorrelation", *IEEE Trans. Inf. Theory* **16**, 85–89.

Seneff, S. (1988). "A joint synchrony/mean-rate model of auditory speech processing", *Journal of Phonetics* **16**(1), 55–76.

Sethares, W. A. (1993). "Local consonance and the relationship between timbre and scale", *J. Acoust. Soc. Am.* **94**(3), 1218–1228.

Sethares, W. A. (1994). "Adaptative tuning for musical scales", *J. Acoust. Soc. Am.* **96**(1), 10–18.

Sethares, W. A. (1997). "Specifying spectra for musical scales", *J. Acoust. Soc. Am.* **102**(4), 2422–2431.

Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J. et Ekelid, M. (1995). "Speech Recognition with Primarily Temporal Cues", *Science* **270**, 303–304.

Shirlaw, M. (1917). *The Theory of Harmony.* New York: Da Capo. reprinted (1969).

Slaymaker, F. H. (1970). "Chords from tones having stretched partials", *J. Acoust. Soc. Am.* **47**, 1569–1571.

Smith, B. (1995). "PsiExp: an Environment for Psychoacoustic Experimentation Using the IRCAM Musical Workstation", In *Society for Music Perception and Cognition Conference '95.* University of California, Berkeley.

Smith, B. K., Sieben, U. K., Kohlrausch, A. et Schroeder, M. R. (1986). "Phase effects in masking related to dispersion in the inner ear", *J. Acoust. Soc. Am.* **80**(6), 1631–1637.

Smoorenburg, G. F. (1970). "Comments on Terhardt' s "Frequency analysis and periodicity detection in the sensation of roughness and periodicity pitch" ", In Plomp, R. et Smoorenburg, G. (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 287–290). Leiden: A. W. Sijthoff.

Smoorenburg, G. F. (1972a). "Audibility region of combination tones", *J. Acoust. Soc. Am.* **52**, 603.

Smoorenburg, G. F. (1972b). "Combination tones and their origin", *J. Acoust. Soc. Am.* **52**, 615.

Stevens, S. S. (1957). "On the psychophysical law", *Psychol. Review* **64**, 153–181.

Stilson, T. et Smith, J. (1996). "Alias-free digital synthesis of classic analog waveform", In *Proceedings of 1996 International Computer Music Conference*. Hong Kong.

Strube, H. W. (1985). "A computationaly efficient basilar membrane model", *Acustica* **58**, 207–214.

Stumpf, C. (1898). "Konsonanz und Dissonanz", *Beiträge zur Akustik und Musikwissenschaft* **1**. (Leipzig: Johann Ambrosius Barth).

Tenney, J. (1988). *A History of Consonance and Dissonance*. New York: Excelsior Music Publishing.

Terhardt, E. (1968a). "Über akustiche Rauhigkeit und Schwangungsstärke", *Acustica* **20**, 215–224.

Terhardt, E. (1968b). "Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung", *Acustica* **20**, 210–214.

Terhardt, E. (1970). "Frequency analysis and periodicity detection in the sensation of roughness and periodicity pitch", In Plomp, R. et Smoorenburg, G. (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 278–287). Leiden: A. W. Sijthoff.

Terhardt, E. (1974a). "On the perception of periodic sound fluctuation (roughness)", *Acustica* **30**, 201–212.

Terhardt, E. (1974b). "Pitch, consonance and harmony", *J. Acoust. Soc. Am.* **55**, 1061–1069.

Terhardt, E. (1984). "The concept of musical consonance: a link between music and psychoacoustics", *Music Perception* **1**(3), 276–295.

Thompson, W. F. et Parncutt, R. (1997). "Perceptual judgments of triads and dyads: assessment of a psychoacoustical model", *Music Perception* **14**(3), 263–280.

Thurstone, L. L. (1927). "The method of paired comparisons for social values", *J. Abnorm. Soc. Psychol.* **21**, 384–400.

Todd, N. (1996). "Toward a theory of the central auditory system I: architecture", In *Proceedings of the 4th International Conference on Music Perception and Cognition* (pp. 173–178). Montréal, Canada.

Trainor, L. J. et Trehub, S. E. (1994). "Key membership and implied harmony in Western tonal music: developmental perspectives", *Perception and Psychophysics* **56**, 125–132.

van de Geer, J. P., Levelt, W. J. M. et Plomp, R. (1962). "The connotation of musical consonance", *Acta Psychol.* **20**, 308–319.

Viemeister, N. F. (1977). "Temporal factors in audition: a system analysis approach", In Evans, E. F. et Wilson, J. P. (Eds.), *Psychophysics and Physiology of Hearing* (pp. 419–429). London: Academic Press.

Viemeister, N. F. (1979). "Temporal Modulation Transfer Functions based upon Modulation Thresholds", *J. Acoust. Soc. Am.* **66**(5), 1364–1380.

Vogel, A. (1974). "Roughness and its relation to the time-pattern of psychoacoustical excitation", In Zwicker, E. et Terhardt, E. (Eds.), *Facts and models in hearing* (pp. 241–250). Berlin, Heidelberg: Springer.

Vogel, A. (1975). "Über den Zusammenhang zwischen Rauhigkeit und Modulationsgrad", *Acustica* **32**, 300.

Vos, J. (1986). "Purity ratings of tempered fifths and major thirds", *Music Perception* **3**(3), 221–258.

Warren, R. M. et Wrightson, J. M. (1981). "Stimuli producing conflicting temporal and spectral cues to frequency", *J. Acoust. Soc. Am.* **70**(4), 1020–1024.

Wright, J. K. et Bregman, A. (1987). "Auditory stream segregation and the control of dissonance in polyphonic music", *Contemporary Music Review* **2**(1), 63–92.

Yost, W. A. et Sheft, S. (1989). "Across-critical-band processing of amplitude-modulated tones", *J. Acoust. Soc. Am.* **85**(2), 848–857.

Zurek, P. M. et Leishowitz, B. H. (1976). "Measurement of the combination tones f2-f1 and 2f2-f1", *J. Acoust. Soc. Am.* **60**(1), 155–168.

Zwicker, E. (1961). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)", *J. Acoust. Soc. Am.* **33**, 248.

Zwicker, E. et Fastl, H. (1990). *Psychoacoustics, Facts and Models.* Berlin: Springer Verlag.

Zwicker, E., Flottorp, G. et Stevens, S. S. (1957). "Critical band width in loudness summation", *J. Acoust. Soc. Am.* **29**, 548–557.

Zwicker, E. et Scharf, B. (1965). "A model of loudness summation", *Psychol. Rev.* **72**, 3–26.

# Publications liées à la thèse

## Publications dans des revues a comité de lecture

Pressnitzer, D. et McAdams, S. (sous presse). "Acoustics, psychoacoustics and spectral music." *Contemporary Music Review.*

Pressnitzer, D., McAdams, S., Winsberg, S. et Fineberg, J. (sous presse). "Perception of musical tension for non-tonal orchestral timbres and its relation to psychoacoustic roughness.", *Perception and Psychophysics.*

## Publications soumises en cours d'expertise

Pressnitzer, D., McAdams, S. "Two phase effects in roughness perception." *Journal of the Acoustical Society of America*, accepté sous réserve de modifications.

## Chapitres d'ouvrages collectifs

Pressnitzer, D. et McAdams, S. (1998). "Phase effects in roughness perception". In A. Rees, A. Q. Summerfield, A. Palmer & R. Meddis (Eds.), *Psychophysical and Physiological Advances in Hearing*, Whurr Publishers, Londres.

## Résumés publiés dans des revues à comité de lecture

McAdams, S. et Pressnitzer, D. (1996). "Psychoacoustic factors contributing to musical tension in Western non-tonal music." *International Journal of Psychology,* 31 (3-4), 148.

Misdariis, N., Smith, B., Pressnitzer, D., Susini, P. et McAdams, S. (à paraître). "Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres." *Journal of the Acoustical Society of America.*

## Actes de congrès

Pressnitzer, D. and McAdams, S. (sous presse) "Summation of roughness across frequency regions". *Proceedings of the Oldenbourg Symposium on "Temporal processing in the auditory system: Psychophysics, physiology and models of hearing".*

Pressnitzer, D. et McAdams, S. (1997). "Influence of phase effects on roughness modeling", *Proceedings of the International Computer Music Conference*, pp. 31-34.

Pressnitzer, D. et McAdams, S. (1997). "Influence de la phase sur la perception de rugosité de sons complexes." *Actes du 4ème Congrès Français d'Acoustique,* (1), pp. 535-538.

Pressnitzer, D., McAdams, S., Winsberg, S. et Fineberg, J. (1996). "Roughness and tension of orchestral timbres." *Proceedings of the 4th International Conference on Music Perception and Cognition,* McGill University, Montreal, Canada, pp. 85-90.

Pressnitzer, D., McAdams, S., Winsberg, S. et Fineberg, J. (1996). "Tension et rugosité dans une musique non tonale." *Actes des Journées d'Informatique Musicale,* Caen, France, pp 147-156.

## Diffusion de connaissances

Pressnitzer, D. (1997). “Sons rugueux, sons tendus.” *Pour la Science* (240), pp. 34.

Pressnitzer, D. (1997). “L’approche du phénomène sonore”. In *La petite encyclopédie de la musique,* pp. 75-78. Paris : Éditions du Regard.