

UNIVERSITE DE LA MEDITERRANEE - AIX-MARSEILLE II

ECOLE DOCTORALE SCIENCES POUR L'INGENIEUR :
MECANIQUE, PHYSIQUE, MICRO ET NANOELECTRONIQUE (ED 353)

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE LA MEDITERRANEE
AIX-MARSEILLE II

Discipline : Acoustique, Traitement du signal et Informatique Appliqués à la Musique (ATIAM)

Présentée et soutenue publiquement par

Mathieu BARTHET

Le jeudi 18 décembre 2008

De l'interprète à l'auditeur : une analyse acoustique et perceptive du timbre musical

Co-directeurs de thèse : Dr. Richard KRONLAND-MARTINET

Dr. Sølvi YSTAD

Dr. Philippe GUILLEMAIN

JURY :	Pr. Stephen McADAMS	Rapporteur
	Dr. Laurent DAUDET	Rapporteur
	Pr. Kristoffer JENSEN	Examinateur, Président
	Dr. Richard KRONLAND-MARTINET	Co-directeur de thèse
	Dr. Sølvi YSTAD	Co-directeur de thèse
	Dr. Philippe GUILLEMAIN	Co-directeur de thèse

A mes grands-parents Henri et Jacqueline, Teddy et Colette, pour tout cet amour partagé, et les passions qu'ils m'ont transmises.

A mes parents Frédéric et Valérie, et à ma soeur Chloé, pour leur soutien, leur amour, et leurs doux timbres de voix.

“On remarqua, dans la plage “Dîner au motel”, la sonorité étrange de la trompette de Miles. Un fragment de peau se détacha à un moment de sa lèvre pour se coincer dans l’embouchure. Pareil à ces peintres qui doivent parfois au hasard la qualité plastique de leur pâte, Miles accueillit volontiers ce nouvel élément d’un jeu “inouï” au sens littéral du mot, jamais entendu.”

Boris Vian décrivant la séance d’enregistrement de la musique du film *Ascenseur pour l’échafaud*, composée par Miles Davis en 1957.

“It is the most fascinating of all fields in the psychology of music^a, for exact scientific work which has the most far-reaching significance in the understanding of music, musical education, the evaluation and construction of musical instruments, and musical criticism. It certainly is one of four cornerstones of the psychology of musical esthetics. But for the reason stated, this is all a virgin field, quite uncharted but certain to be explored and possessed by the musical world in the immediate future.”

^aNDLR : à propos de l’étude du timbre

[[Seashore, 1938](#), p. 100]

Résumé

Découvrir quels sont les paramètres acoustiques corrélés à l'expression musicale revêt un intérêt fondamental afin d'améliorer notre compréhension de la perception musicale, et trouve de multiples applications en synthèse sonore. L'interprétation musicale est un acte au cours duquel un instrumentiste traduit le signal notationnel d'un compositeur en sons musicaux, tout en impliquant son propre jugement. Si la plupart des études sur l'interprétation musicale portent sur le rythme, la dynamique, ou les hauteurs, bien moins se consacrent au timbre. Cet attribut perceptuel du son, à la fois lié à l'identité de la source sonore et à la qualité sonore, semble pourtant être un paramètre musical essentiel. L'objet de cette thèse, dont l'approche repose sur l'analyse-synthèse, est de mieux comprendre le rôle du timbre dans l'interprétation musicale.

Une étude de dissemblance réalisée sur des sons de synthèse de clarinette, obtenus à partir d'un modèle physique, a permis d'évaluer l'influence du contrôle instrumental (pression d'alimentation et pince) sur les timbres produits par l'instrument. Des enregistrements d'extraits musicaux joués un grand nombre de fois par un même clarinettiste professionnel sur un instrument naturel selon différentes intentions musicales ("scolaire" et "expressive") ont ensuite été analysés. Les mécanismes de transmission de l'expression musicale ont ainsi pu être étudiés au travers de changements de timbre, de rythme et de dynamique. Certaines variations de timbre (variations de qualité sonore au sein des notes et entre les différentes notes) sont reproduites de manière systématique par l'interprète lorsque son intention musicale est la même. La nature de ces variations change lorsque l'intention expressive change, ce qui tend à prouver que les musiciens agissent sur certaines dimensions du timbre afin de varier leur expression. Deux expériences perceptives complémentaires ont révélé que la nature des évolutions temporelles de la brillance des notes influe sur les préférences musicales des auditeurs. La qualité musicale de séquences inexpressives, produites sur des instruments de type entretenu à l'aide d'échantillonneurs, a notamment pu être améliorée de manière significative grâce à un contrôle de la brillance par filtrage dynamique.

L'ensemble de ces travaux appuie l'idée que les variations morphologiques temporelles de timbre (par ex. variations temporelles de brillance) constituent l'un des vecteurs de l'expression musicale.

Discipline : Acoustique, Traitement du signal et Informatique appliqués à la Musique.

Mots-clefs : timbre, interprétation musicale, expression musicale, perception musicale, analyse-synthèse, transformation sonore, modélisation physique, clarinette.

Laboratoire :

CNRS - Laboratoire de Mécanique et d'Acoustique (UPR 7051)
31 chemin Joseph-Aiguier
13402 Marseille Cedex 20
France

Abstract

English title : "From performer to listener : an acoustical and perceptual analysis of musical timbre"

Discovering the acoustical parameters responsible for musical expression is of fundamental interest to improve our understanding of musical perception, and finds numerous applications in sound synthesis. Musical interpretation is an act during which the performer traduces the notational signal from the composer while involving his/her own judgement. If many studies on musical interpretation deal with the role of timing, dynamics or pitch, rather few focus on the one of timbre. Nevertheless, this perceptual attribute referring to both the identity of the sound source and the sound quality, seems to be an essential musical parameter. The goal of this work, based on an analysis-synthesis approach, is to better understand the role of timbre in musical interpretation.

Dissimilarity judgments of synthesized clarinet tones obtained from a physical model, allowed to characterize the influence of the control parameters of the instrument (blowing pressure and player's lip force on the reed) on the generated timbres. Recordings of musical excerpts played a large number of times by the same professional clarinetist on a natural instrument according to different musical intentions ("scholastic" and "expressive") were then analyzed. The transmission process of musical expression was hence studied through changes in timbre, rhythm, and dynamics. Some timbre variations (sound quality variations within tones or between tones) are systematically reproduced by the performer when his/her expressive intention is the same. The nature of these variations is modified when the intention changes, which tends to prove that musicians act on certain timbre dimensions to vary their expression. Two complementary perceptual experiments revealed that the evolutions of the tones' brightness across time influence listeners' musical preferences. The musical quality of inexpressive sequences, generated on self-sustained instruments with samplers, could significantly be improved thanks to the control of brightness by dynamic filtering.

This work supports the idea that the temporal morphological variations of timbre (e.g. brightness temporal variations) are involved in musical expression.

Key-words : timbre, musical interpretation, music performance, music perception, analysis-synthesis, sound transformation, physical modeling, clarinet.

Remerciements

Je tiens tout d'abord à remercier chaleureusement mes co-directeurs de thèse Richard, Sølvi, et Philippe. Merci de m'avoir permis de poursuivre la passion que j'éprouve pour la musique et les sciences en me proposant cette thèse au sein de l'équipe S2M. Merci de m'avoir transmis avec patience et pédagogie une partie de vos savoir et savoir-faire en matière d'analyse-synthèse, perception sonore, et modélisation physique, et ce dans un climat de travail toujours si agréable. Merci à Sølvi pour toutes ces réunions passées à se questionner sur la notion de timbre sans pour autant finir timbrés, à établir des protocoles expérimentaux, à écouter et ré-écouter des sons et tâcher d'en comprendre le sens, ainsi que pour ces nombreuses corrections et relectures d'articles. Merci à Richard pour ses riches idées et ses sages conseils, ses cours de signal éclairant, et son humour qui réserve parfois des surprises... Merci à Philippe pour cet appel lors de mon stage de Master 2 destiné à me parler d'un certain projet de thèse pluridisciplinaire autour des notions de timbre et d'interprétation musicale, pour m'avoir appris à faire des sons de clarinette avec des 0 et des 1, et pour ces nombreux travaux pratiques sur la table d'expérimentation de bruits d'impacts.

Je tiens également à remercier mes rapporteurs et examinateur, Stephen McAdams, Laurent Daudet et Kristoffer Jensen pour le temps qu'ils ont investi dans l'évaluation de mon travail de thèse, et leurs précieuses remarques et corrections.

J'ai éprouvé beaucoup de plaisir à travailler au sein de l'équipe S2M du Laboratoire de Mécanique et d'Acoustique. Il y règne une ambiance de travail très conviviale, qui est non sans être liée à la gentillesse et la bonne humeur de ses différents acteurs : chercheurs, ingénieurs, thésards, post-docs, stagiaires, et sa chère secrétaire. Je tiens à ce titre à les remercier tout aussi chaudement que sait le faire notre bien-aimé barbecue, ami rassembleur, dont les douces émanations nous ont bien souvent procuré satisfaction.

Merci à Thierry dont la gentillesse égale son expertise en Max-MSP, à Mitsuko pour toutes ces discussions intéressantes, ses conseils avisés sur mes articles, ses coups de main de maître en analyses statistiques, à Christophe et Jean qui répondent dès qu'on sonne (lire consonnes...) et que l'on a une question, à Didier pour nos échanges stimulants sur la question de l'interprétation musicale, et pour m'avoir donné envie d'écouter de la flûte traversière. Merci à Fabrice, qui de latex à la mécanique du point (pour certains TD...), en passant par le rayonnement de la clarinette, aura toujours su m'éclairer avec brio et générosité, à Jonathan pour avoir partagé de sacrés moments avant, pendant et après des conférences internationales..., à Marie-Céline, pour nos discussions sur les tests perceptifs, et notre perception des calanques lors de cette fameuse ballade à Morgiou pleine de cailloux, à Thibaud, pour sa sympathie et ses produits corses (!), à Adrien, qui de collègue de bureau est devenu collègue tout cours, merci pour ses nombreux services et son esprit d'équipe, à Benjamin que je questionne dès 8h00 du matin sur l'orthogonalisation. Merci à Michèle pour sa gentillesse, son efficacité et pour rendre simple ce qui pourrait être compliqué.

Je tiens également à remercier Sabine Meunier et Sophie Savel de l'équipe APIM pour leurs conseils sur les protocoles d'expériences perceptives et leur gentillesse. Je remercie également

chaleureusement Jean-Claude Risset, qui m’a partagé son expertise sur le timbre et orienté vers des articles très instructifs lors de sympathiques entretiens.

Je voudrais saluer l’aide précieuse des statisticiens Henri Burnet de l’Institut des Neurosciences Cognitives De la Méditerranée, et Jean-Pierre Durbec du Centre d’Océanologie de Marseille, qui m’ont ouvert leurs portes avec amabilité et permis de sélectionner les méthodes statistiques adaptées aux traitements de mes données.

Je suis très reconnaissant envers le clarinettiste professionnel et si bon pédagogue Claude Crousier du Conservatoire National de Région de Marseille, sans qui ce travail n’aurait pu avoir lieu. En effet, ce sont ses notes, qui émises et enregistrées dans la chambre anéchoïque du LMA, ont nourri une bonne partie de mes recherches sur l’exploration du timbre dans l’interprétation musicale. Merci pour ton implication dans ce projet, et les intéressantes discussions que nous avons pu avoir sur la clarinette et son contrôle. Ce sont également les participants des expériences perceptives que j’ai menées que je souhaite remercier : merci à ceux du LMA, du GRIM (Groupe de Recherche et d’Improvisation Musicales), et du Train en Marche.

Mireille Besson et Daniele Schön de l’Institut des Neurosciences Cognitives De la Méditerranée ont su étendre mon horizon scientifique aux sciences cognitives et à la psychologie de la musique lors de précieux échanges. Je tiens à les en remercier.

Je tiens à remercier l’Ecole Doctorale Sciences pour l’Ingénieur : Mécanique, Physique, Micro et Nanoélectronique (ED 353), qui offre notamment à ses étudiants la possibilité de réaliser de très intéressantes formations (comme la communication scientifique en anglais) ; je tiens à saluer le travail, la compétence et l’amabilité de ses anciens et nouveaux responsables (M. Suquet, Mme Pullino, M. Saurel, M. Campion). Je souhaiterais également remercier la Société Française d’Acoustique, qui m’a accordé une bourse afin de participer à un congrès international, et organise des journées de travail prolifiques, comme celles du Groupe Perception Sonore, l’Agence Nationale de la Recherche, qui au travers du financement du projet Sensons (ANR, JC05-41996) dirigé par notre équipe, m’a permis d’exercer mes recherches dans de très bonnes conditions (achat de matériel comme des microphones, possibilité de participer à des congrès internationaux, etc.).

En rapport avec l’enseignement, je tiens vivement à remercier Françoise Briolle, avec qui ce fut un réel plaisir de travailler en tant que chargé de TP en traitement du signal au sein de l’IUT Réseau et Télécommunications de Luminy pendant plusieurs années. Je suis également très reconnaissant à Frédéric Lebon qui m’a permis d’avoir un poste d’Attaché Temporaire d’Enseignement et Recherche au sein de l’Université de Provence, dont je retire une riche expérience.

Je remercie ma chère famille qui m’a toujours épaulé, et tout particulièrement mon père pour la relecture de ce manuscrit entre deux avions lors de quelques voyages en Bulgarie.

Une énorme dédicace à mon entourage d’amis qui rend si belle la vie : aux parisiens de la “Boulard Team”, mes frères et soeurs de primaires ! (Alex, Aurélien, Julien, Bassem, Perrine, les deux Ben, et mon ami de toujours Michaël), qui même à distance sont toujours dans mes pensées, à mes proches d’Henri IV (Cécile, Stéphanie, Erwan Coco Robicheau, avec qui j’ai discuté de la définition du timbre jusqu’à ce que lune remplace soleil), à “l’Ist crew”, friand de belotes, escapades chez Georges, et surtout racontages d’exclus (Adrien, Erwan, Fatima, Jim, Aniss, Clément, Enrique, et Xavier, sans oublier le petit nouveau Moriu), aux copains musiciens de la Terrasse (Patrick, Bernard et Farid) pour toutes ces notes aux Trois Chapeaux, à celui qui m’a fait découvrir comment faire crier sa Stratocaster, Mickey, aux collègues du LAM (Adrien, Pierre et Maëva) avec qui nous avons fait nos premiers pas dans ce monde de l’acoustique musicale, aux

marseillais intercontinentaux de la Maison Du Bonheur (Anne, Ana Laura, Sarah, Valérie, Fred, Jehan-Ju, et Hanes), pour lesquels amour, humour et gentillesse rentrent en résonance, aux Zékoutzy Pourvoir (cf. MDB !) pour ces répétitions inoubliables dans le garage du 34 Saint-Bazile, aux potes de DEA (Tim, Jehan-Ju bis, Thomas, Fab, Didier, et Simone) avec qui parcourir les sols de Château-Gombert, Saint-Yvoine, Montpellier, ou Paris tout en découvrant l'acoustique fut un bonheur intarissable, à ceux du CNRS (Céline, Bastien, Romain), pour nos moments partagés sur la Plaine (...), aux zicos de Kiwan (Luccio, Fabien et Tim bis) pour nos impros locaux, aux acteurs du Train En Marche sur le devant de la scène marseillaise (Jeremy, France, Pli, Pauline, Julien, Vincent, Pascal), à mes chers voisins du Baignoir et d'un peu plus loin (Manu², Antoine, Thomas, Pierrick, Nico, et Valentine), dont je n'oublierai jamais le soutien pendant ma rédaction de thèse autour de bons repas et de bonnes rigolades ; à celle qui, durant toutes ces lignes n'a cessé d'être présente dans mon coeur, à celle qui m'a soutenu et encouragé durant toutes ces années, à celle qui me plaît pour sa beauté au sens large, à mon ange là près de moi, Angela.

Enfin, une spéciale dédicace à Aurélien et Elimar, qui se passaient les bagues autour des doigts à Paris, alors que je soutenais mes travaux de thèse à Marseille. Longue vie d'amour à vous !

Table des matières

1	Introduction	1
1.1	Timbre et musique	1
1.1.1	L'ère contemporaine ou le timbre comme nouvelle fonction du langage musical	1
1.1.2	L'avènement de la synthèse sonore	2
1.1.3	De la synthèse sonore à la perception des sons musicaux	4
1.1.3.1	La structure "hauteur-timbre" de Risset	4
1.1.3.2	Echelle chromatique tempérée, série harmonique, série inharmonique	5
1.1.3.3	L'analyse par la synthèse	9
1.1.4	De l'utilisation du timbre dans les musiques traditionnelles	10
1.2	Sémiologie musicale	16
1.2.1	La musique, ou l'art de l'ouïe	16
1.2.2	La tripartition de Molino et Nattiez	17
1.2.3	La question du sens musical	17
1.3	Écoute ou écoutes ?	19
1.3.1	Qu'écoutent les spécialistes ?	19
1.3.2	Écouter, Ouïr, Entendre, Comprendre	21
1.4	Du compositeur à l'auditeur	24
1.4.1	Définition de l'interprétation musicale	24
1.4.2	Chaîne de communication des musiques écrites	26
1.4.3	Du rôle du timbre dans l'interprétation musicale	28
1.5	Aspects méthodologiques	30
1.6	Organisation de la thèse	32
2	Les différentes facettes du timbre	33
2.1	De l'identification à la qualification des sons	33
2.1.1	Lien causal du timbre à la source sonore	33
2.1.2	Le paradoxe du timbre	36
2.1.3	Définition générale du timbre schaefferienne : le concept de permanence / variation	36
2.1.3.1	La registration et la facture instrumentale, sources de variations de timbre "abstraites"	37
2.1.3.2	Le jeu instrumental, source de variations de timbre "concrètes"	40
2.1.4	Le timbre en tant que caractéristique sonore	42
2.1.5	Bilan : les deux écoutes du timbre	43
2.1.6	Les définitions standard du timbre en psychoacoustique	46
2.1.7	L'aspect morphologique temporel du timbre	48
2.2	Les corrélations entre acoustique et perception du timbre	49
2.2.1	La conception classique du timbre	50
2.2.2	La théorie des formants	50

2.2.3	Relations de phase entre les composantes harmoniques	52
2.2.4	Le timbre, un attribut multidimensionnel du son	53
2.2.4.1	Méthodologie d'analyse de la représentation perceptive du timbre	54
2.2.4.1.1	Sélection ou élaboration des stimuli	54
2.2.4.1.2	Sélection des participants	55
2.2.4.1.3	Test de jugements de dissimilarité (ou de dissemblance)	55
2.2.4.1.4	Représentations des jugements de dissimilarité (analyse multidimensionnelle de proximité et classification ascendante hiérarchique)	56
2.2.4.1.5	Interprétation psychoacoustique de la structure	59
2.2.4.1.6	Vérification de l'interprétation dans des situations musicales	60
2.2.4.2	Les espaces de timbre de Grey et McAdams <i>et al.</i>	60
2.2.4.2.1	L'espace de timbre de Grey	60
2.2.4.2.2	L'espace de timbre de McAdams <i>et al.</i>	60
2.2.5	Les descripteurs de timbre et de signal, corrélats physiques de la perception	62
2.2.5.1	Notations et boîte à outils MATIMBRE	63
2.2.5.2	L'enveloppe temporelle d'amplitude (ENV)	66
2.2.5.2.1	Calcul de l'enveloppe crête à partir du signal analytique	66
2.2.5.2.2	Enveloppe RMS (Root Mean Square)	67
2.2.5.3	Les descripteurs temporels	68
2.2.5.3.1	Le temps d'attaque (AT)	68
2.2.5.3.2	Le Centre de Gravité Temporel (TC)	70
2.2.5.4	Les descripteurs spectraux et/ou spectro-temporels	70
2.2.5.4.1	Le Centre de Gravité Spectral (SC)	71
2.2.5.4.2	L'Étalement Spectral (SS)	77
2.2.5.4.3	Le Coefficient d'Asymétrie Spectrale (SSK)	77
2.2.5.4.4	Le Coefficient d'Aplatissement Spectral (SKU)	78
2.2.5.4.5	Le Roulement Spectral (SRO)	78
2.2.5.4.6	Le Flux Spectral (SF)	79
2.2.5.5	Analyse/Synthèse additive de sons (quasi-)harmoniques	80
2.2.5.6	Les descripteurs "harmoniques"	81
2.2.5.6.1	Le Centre de Gravité Spectral "harmonique" (HSC)	81
2.2.5.6.2	Les Centres de Gravité Spectraux Impair (OSC) et Pair (ESC)	82
2.2.5.6.3	L'Irrégularité Spectrale (IRR)	82
2.2.5.6.4	Le rapport d'énergie Impair/Pair (OER)	83
2.2.5.6.5	Le Tristimulus (TR1, TR2, TR3)	83
2.2.5.6.6	L'Inharmonicité (INH)	84
2.2.5.7	Les MFCC (Mel-Frequency Cepstral Coefficients)	84
2.2.5.8	Echelle d'amplitude (linéaire, puissance)	85
2.2.5.9	Limites des descripteurs de timbre	86
3	Clarinette et modélisation physique	87
3.1	Origines	87
3.2	Fonctionnement général de l'instrument	88
3.3	Principaux facteurs constitutifs du timbre de la clarinette	90
3.3.1	Le système bec-anche	90
3.3.2	Le tuyau fermé/ouvert : un résonateur quart d'onde	91
3.3.3	Le couplage non-linéaire entre l'excitateur et le résonateur	93
3.3.4	Effets des trous latéraux ouverts	96

3.3.5	Le pavillon	98
3.3.6	La dissipation d'énergie	98
3.3.7	L'inharmonicité du résonateur	98
3.3.8	Rôle de l'instrumentiste	99
3.4	Synthèse par modélisation physique	100
3.4.1	Modèle physique simplifié	101
3.4.1.1	Hypothèses générales	101
3.4.1.2	Description géométrique du système et notations	102
3.4.1.3	L'anche	102
3.4.1.4	Les résonateurs (instrument et conduit vocal)	103
3.4.1.5	Couplage non linéaire entre l'excitateur et le résonateur	105
3.4.1.6	Variables adimensionnées	105
3.4.1.7	Le rayonnement	106
3.4.1.8	Résumé du modèle complet	108
3.4.2	Construction du modèle de synthèse	108
3.4.2.1	Modélisation numérique de l'impédance d'entrée	108
3.4.2.2	Modélisation numérique du rayonnement	110
3.4.3	Exemples de simulation	112
4	Du contrôle de la clarinette au timbre	117
4.1	Introduction	117
4.2	Méthodes	117
4.2.1	Corpus sonore	117
4.2.1.1	Mesures de la pression d'alimentation	118
4.2.1.2	Espace des paramètres de contrôle	119
4.2.2	Sélection des stimuli	120
4.2.3	Sélection des participants	121
4.2.4	Appareillage	122
4.2.5	Procédure	122
4.2.6	Analyses statistiques	122
4.2.6.1	Homogénéité des réponses des participants	122
4.2.6.2	Analyse multidimensionnelle de proximité (MDS)	122
4.2.6.3	Classification ascendante hiérarchique	123
4.2.7	Descripteurs de timbre	123
4.3	Résultats	124
4.3.1	Homogénéité des réponses des participants	124
4.3.2	Espace de timbre perceptif	124
4.3.2.1	Analyse multidimensionnelle de proximité (MDS)	124
4.3.2.2	Espace de timbre perceptif	125
4.3.3	Corrélat mécaniques de l'espace de timbre perceptif	126
4.3.4	Corrélat acoustiques de l'espace de timbre perceptif	129
4.3.4.1	Corrélat acoustiques de la première dimension	129
4.3.4.2	Corrélat acoustiques de la seconde dimension	135
4.3.4.3	Corrélat acoustiques de la troisième dimension	136
4.3.5	Analyse des questionnaires	137
4.4	Conclusions : vers des descripteurs de timbre adaptés	142

5	Relation entre l'intention expressive et le timbre	145
5.1	Introduction	145
5.2	Corpus sonore	148
5.2.1	Interprétations de l'extrait de Bach	148
5.2.2	Interprétations de l'extrait de Mozart	148
5.2.3	Dispositifs d'enregistrement	150
5.3	Segmentation	151
5.3.1	Pré-traitement et réduction du problème	152
5.3.1.1	Estimation de la fréquence fondamentale	152
5.3.1.2	Calcul de la fonction de détection	154
5.3.1.3	Détection des pics	154
5.3.1.4	Validation et correction d'erreurs	154
5.3.1.5	Performance de la méthode	154
5.4	Choix des descripteurs	156
5.4.1	Descripteurs de timbre	157
5.4.2	Descripteurs rythmiques	158
5.4.3	Descripteur de la dynamique	158
5.4.4	Descripteur de la hauteur	158
5.5	Calcul et synchronisation des descripteurs	159
5.5.1	Calcul des descripteurs	159
5.5.2	Synchronisation des descripteurs	160
5.6	Méthodes d'analyses statistiques	161
5.6.1	Mesures des dissimilarités entre descripteurs	161
5.6.2	Classification ascendante hiérarchique	161
5.6.3	Analyses de variance (ANOVA)	161
5.7	Résultats	162
5.7.1	Interprétations de l'extrait de Bach	162
5.7.1.1	Dissimilarité et classification ascendante hiérarchique	162
5.7.1.2	Analyses de variance	164
5.7.2	Interprétations de l'extrait de Mozart	172
5.7.2.1	Matrices de dissimilarité	172
5.7.2.2	Comparaison de la moyenne des descripteurs	174
5.8	Discussion	178
6	Influence perceptive des variations de timbre	181
6.1	Introduction	181
6.2	Expérience 1 : d'un jeu expressif vers un jeu inexpressif	182
6.2.1	Stimuli	182
6.2.1.1	Corpus sonore	182
6.2.1.2	Transformations sonores	182
6.2.1.3	Elaboration des stimuli	185
6.2.2	Participants	187
6.2.3	Appareillage	187
6.2.4	Procédure : test de jugement de préférence	187
6.2.5	Méthodes statistiques	188
6.2.5.1	Représentation des données perceptives	188
6.2.5.2	Homogénéité des réponses	189
6.2.5.3	Analyses de variance	189
6.2.5.4	Classification Ascendante Hiérarchique (CAH)	189
6.2.6	Résultats	190
6.2.6.1	Homogénéité des réponses	190

6.2.6.2	Influence de l'extrait musical	190
6.2.6.3	Influence des transformations	190
6.2.6.4	Classification ascendante hiérarchique	193
6.2.7	Discussion	193
6.3	Expérience 2 : d'un jeu inexpressif vers un jeu expressif	195
6.3.1	Stimuli	196
6.3.1.1	Corpus sonore	196
6.3.1.2	Contrôle de la brillance par filtrage dynamique	196
6.3.1.3	Elaboration des stimuli	199
6.3.2	Participants	200
6.3.3	Appareillage	200
6.3.4	Procédure : test de jugement de dissimilarité et de préférence	200
6.3.5	Méthodes statistiques	201
6.3.5.1	Analyse des dissimilarités	201
6.3.5.2	Analyses des préférences	201
6.3.6	Résultats	202
6.3.6.1	Analyses des dissimilarités	202
6.3.6.2	Analyses des préférences	204
6.3.6.3	Analyse des questionnaires	206
6.3.7	Discussion	209
7	Conclusions et perspectives	211
7.1	Espace de timbre de la clarinette	211
7.2	Mesures des variations de timbre	213
7.3	Analyse perceptive de l'effet des variations de brillance	213
7.4	Lien entre les trois chapitres expérimentaux	214
7.5	Perspectives	214
	Annexes	215
A	Liste des Exemples Sonores	217
B	Echelle, tempérament et perception des hauteurs	219
C	Description du bec de clarinette	223
D	Constantes du modèle physique de clarinette	225
E	Cartographie du jeu instrumental de la clarinette	227
F	Compléments des analyses statistiques	229
F.1	Concernant le chapitre 4	229
F.2	Concernant le chapitre 5	231
F.2.1	Description de la méthode séquentielle de Holm :	231
F.2.2	Interprétations de l'extrait de Bach	232
F.2.3	Interprétations de l'extrait de Mozart	233
F.3	Concernant le chapitre 6 - Expérience 1	234
F.4	Concernant le chapitre 6 - Expérience 2	234
G	Liste de publications	237
	Bibliographie	239

Chapitre 1

Introduction

Nous allons au cours de ce chapitre poser les jalons de notre réflexion et donner quelques définitions nécessaires afin d'aborder le vaste sujet de la perception musicale et plus particulièrement de ses relations avec le timbre musical. Il serait bien fade de parler de musique et de sons sans en offrir à l'ouïe quelques échantillons, aussi nous vous proposerons tout au long du texte et afin d'illustrer les propos tenus quelques exemples sonores qui sont présents sur le disque compact accompagnant le document.

1.1 Timbre et musique

1.1.1 L'ère contemporaine ou le timbre comme nouvelle fonction du langage musical

Si nous n'avons encore que peu de recul sur la musique du XX^e siècle, et encore moins sur celle du XXI^e, il est tout de même un fait marquant que l'on peut d'ores et déjà constater, celui du rejet massif du système traditionnel tonal¹ chez les compositeurs contemporains de musique occidentale dite "savante". Non que cela signifie que la musique traditionnelle ait été fustigée, loin de là et fort heureusement, mais qu'au profit de l'évolution musicale, la domination des hauteurs et des durées dans laquelle elle est ancrée depuis la Renaissance s'est vue remise en question. Parmi les grandes révolutions, s'opère notamment la recherche du son, du timbre musical pour lui-même. Si la définition du timbre - "étrange et multiple paramètre", comme le dépeint Boulez [Boulez, 1963, p. 37] - mérite que l'on s'y attarde, et fait d'ailleurs l'objet du chapitre 2, retenons dès à présent que ce que l'on dénomme par timbre dans le son musical n'est ni hauteur, ni durée, ni intensité. Dès 1911, Schönberg, l'un des premiers instigateurs de l'émergence du timbre comme fonction centrale du langage musical laisse entrevoir sa vision avant-gardiste dans son Traité sur l'harmonie : "Le son devient perceptible grâce à la couleur sonore, dont l'une des dimensions est la hauteur... Ainsi, si, à partir de couleurs sonores différenciées en fonction de la hauteur, l'on peut créer des motifs - que l'on appelle "mélodies" -, des progressions dont la cohérence s'apparente à celle des processus cognitifs, on doit pouvoir également construire des progressions à partir des couleurs de l'autre dimension, à partir de ce que l'on peut appeler des progressions de la simple "couleur sonore", dont les relations suivraient une logique tout à fait équivalente à celle qui nous satisfait dans les mélodies de hauteurs." [Schoenberg, 1911]. Il est l'inventeur de la *Klangfarben-melodie* (littéralement mélodie de couleurs sonores), dans laquelle, comme le décrit Cadoz, "les instruments et leur jeu se combinent en des processus complexes minutieusement organisés et

¹système harmonique où les relations entretenues par les accords sont régies par la tonalité et où un accord particulier, dit de tonique, prend plus de poids que les autres. "La tonalité a été utilisée par les compositeurs classiques du début du XVII^e siècle jusqu'au début XX^e siècle (époques baroque, classique, et romantique), dans le jazz à partir du ragtime jusqu'au hard bop des années 60, et dans une bonne partie de la musique populaire passée et actuelle." [Siron, 2004].

maîtrisés par l'écriture, pour collaborer à la réalisation d'une "identité de timbre spécifique", elle-même évolutive" [Cadoz, 1991]. L'Exemple Sonore n° 1 est extrait de la troisième des pièces pour orchestre opus 16 de Schönberg, *Farben*. Dans cette pièce, la clarinette perd son identité propre pour se fondre dans une texture sonore au timbre spécifique mise en jeu par l'association sonore de tous les instruments à la fois. On conçoit bien par là l'ingénu travail de recherche compositionnelle ayant permis de parvenir à cet effet, ce dernier ne pouvant s'expliquer simplement par le fait que les notes émises par les différents instruments soient synchrones. Selon le procédé utilisé par Schönberg, le rôle des accords² n'est plus d'amener une fonction harmonique, mais d'amener une sonorité, une masse sonore, un timbre. Les choix d'instrumentations conduisent à créer des changements de couleurs sur un même accord tenu. Il est également remarquable de constater que la trame des durées se révèle non par des changements de hauteurs, mais par des changements de timbre. Un phénomène similaire est présent dans *Chronochromie* de Messiaen (cf. Exemple Sonore n° 2), bien que dans cet exemple des changements de hauteurs aient lieu en parallèle de ceux de timbres. Comme le souligne Lerdahl au sujet de cette oeuvre, "le timbre n'est donc plus uniquement une "couleur", mais il devient un élément porteur de forme essentiel." [Lerdahl, 1991]. La fin de l'Exemple Sonore n° 2 laisse entendre les réactions mitigées du public présent lors de la représentation de la pièce donnée en 1962 par l'orchestre national de la RTF (Radiodiffusion Télévision Française) et dont une partie semble manifestement hermétique à l'écriture de Messiaen. Car s'il est bien quelque chose qui est également remis en question en musique à l'époque contemporaine - et dans l'art, d'une manière générale - c'est la notion d'esthétique [Tolstoï, 1931]. Cela peut certes être troublant pour l'auditeur qui perd ses repères, jusque là habitué à une conception du Beau fondée sur les notions de joliesse et d'harmonie. Avec la complexification du matériau musical, les compositeurs développent de nouveaux langages musicaux (par ex. dodécaphonisme, sérialisme). Les notations traditionnelles deviennent alors insuffisantes et de nouveaux modes de représentations voient le jour (cf. figure 1.1).

1.1.2 L'avènement de la synthèse sonore

Non sans être liée au développement du rôle du timbre en musique, une autre grande révolution s'opère au XX^e siècle avec l'avènement de la synthèse sonore. En 1957, Max Mathews, l'un des pères de l'informatique musicale, développe le premier logiciel de synthèse numérique du son, baptisé *Music 1*. Dans [Gayou, 2007], nous est confié une amusante anecdote sur les origines de sa découverte. C'est lors d'un concert, après une mauvaise interprétation d'un récital de piano de Schnabel que John Pierce alors directeur de Mathews lui glissa : "Max, un ordinateur peut faire mieux que ça ! Arrête un moment tes recherches sur la voix et écris-nous un programme de musique !", ce dans quoi Mathews, passionné de musique, ne tardera pas à s'engager. Si l'interprète de la pièce de Schnabel n'a peut-être pas satisfait son public de l'époque, elle aura au moins suscité le grand bouleversement de la synthèse par ordinateur ! La grande richesse de l'ordinateur en musique provient du fait qu'à lui seul, il est à même de générer tous les effets sonores possibles, tout du moins sur le plan théorique. Ceci est lié au théorème de Shannon qui sous-entend que l'on peut produire tous les sons perceptibles par l'oreille humaine à l'aide d'échantillons numériques. L'ordinateur apparaît alors comme un "instrument universel" qui permet tant de reproduire les sons existants, que d'en produire de nouveaux, des sons jusqu'alors "inouïs". Ceci ne manque pas d'inspirer les musiciens et chercheurs qui entrevoient rapidement les multiples possibilités offertes par l'outil numérique. L'on ne peut manquer de citer à ce propos Jean-Claude Risset, l'un des pionniers de la musique spectrale³ : "l'ordinateur, utilisé comme outil de synthèse, permet d'élaborer à volonté la structure du son et de pousser le travail de composition jusqu'au niveau de la microstructure : le musicien qui ne se contente pas de composer avec des sons peut envisager de composer le son lui-même." [Risset, 1991]. Comme le rappelle John Chowning, les paramètres traditionnels de

²"entité harmonique perçue globalement, formée d'au moins trois sons joués simultanément." [Siron, 2004]

³expression utilisée par Dufourt pour décrire les compositions de Risset.

$\text{♩} = 120$ Très dynamique

mf

S.

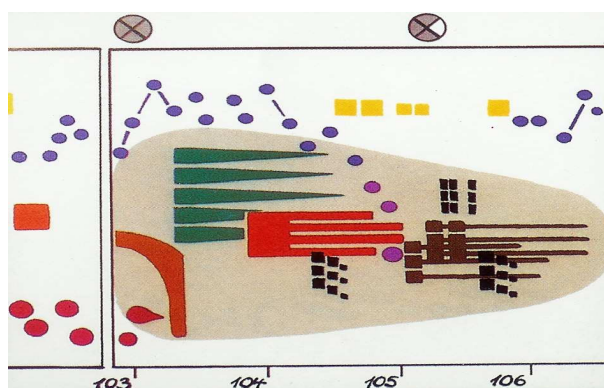
1. Mi - rá lo que me en-con - tré ————— por an - dar a - ga - cha -
 2. Las mu - cha - chas de mi tier - ra no son bo - ni - tas ni

A.

1. Mi - rá lo que me en-con - tré, lo que en-con - tré al an - dar a - ga - cha -
 2. Las mu - cha - chas de mi tier - ra no son bo - ni' bo - ni - tas ni

B.

(a) Chanson populaire du Venezuela



(b) Artikulation de Ligeti (Hörpartitur de Wehinger)

FIG. 1.1 – Exemples de partitions traditionnelle (a), et contemporaine (b)

composition dans la musique occidentale sont la hauteur, qu'elle soit utilisée en succession (mélodie) ou en simultanéité (accord ou harmonie), le temps (rythme, tempo), la dynamique⁴, le timbre (au niveau de l'instrumentation), et parfois la localisation comme dans la musique à plusieurs chœurs [Chowning, 2007]. L'art de l'instrumentation, comme nous l'a montré l'exemple de *Farben* cité ci-dessus, alloue déjà au compositeur de subtiles possibilités d'agissements sur le timbre orchestral, mais celles-ci restent limitées. Les partiels, constituants des sons instrumentaux, et qui peuvent former des séries harmoniques (par ex. instruments à vent tel que la clarinette, orgue), quasi-harmoniques (par ex. cordes, bois à corps coniques), ou inharmoniques (par ex. cloches, gongs), ne peuvent être qu'en partie modifiés par le contrôle instrumental, ou pour certains instruments comme les cuivres, par des dispositifs annexes tels que les sourdines. Dans tous les cas, les timbres engendrés évoluent dans un univers limité et déterminé par la physique de l'instrument et les gestes instrumentaux du musicien. Ce qui vient d'être dit pourrait laisser sous-entendre un côté restrictif aux possibilités offertes par les instruments traditionnels. Loin s'en faut, car tant l'instrument que l'action du musicien sont variables dans le temps ; les instruments sont faits de matières "vivantes" (ne dit-on pas d'une guitare qu'elle a "bien vieilli" ?, le piano est connu pour se désaccorder lorsqu'il est placé à proximité d'un radiateur, l'anche de la clarinette réagit différemment selon l'hygrométrie, etc.), et le musicien par nature, n'accomplira jamais deux fois exactement le même geste instrumental. Là réside justement, la beauté de la relation instrumentale.

⁴Ce que l'on entend par dynamique en musique correspond à la variation d'intensité (également appelée *nuance*) des sons au cours du temps.

1.1.3 De la synthèse sonore à la perception des sons musicaux

1.1.3.1 La structure "hauteur-timbre" de Risset

Parmi les nombreux procédés compositionnels mis au point par Risset, il en est un, celui de la structure "hauteur-timbre" ou prolongation de l'harmonie dans le timbre, qui illustre bien les nouveaux potentiels offerts par la synthèse sonore. Ce procédé, illustré par la figure 1.2, est utilisé au début de sa pièce *Mutations* (cf. Exemple Sonore n° 3). Dans cet exemple, comme l'explique Chowning [Chowning, 2007, p. 27], Risset "composa une brève suite de hauteurs qu'on peut écouter d'abord en succession (mélodie), puis simultanément (harmonie), enfin comme des partiels qui évoluent dans le temps, et qu'on perçoit comme un timbre". Nous sommes aux portes de la psychoacoustique musicale.

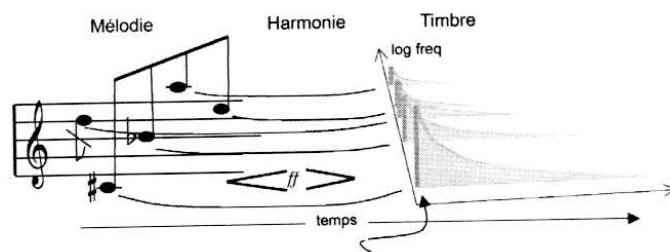


FIG. 1.2 – Représentation de la structure "hauteur-timbre" utilisée par Jean-Claude Risset au début de *Mutations* (figure de John Chowning [Chowning, 2007])

La première partie de l'extrait est constituée d'une succession de notes qui sont des sons purs⁵. Si nous les entendons comme telles, c'est qu'elles sont jouées de manière asynchrones et très brèves⁶. Dès lors que les sons coexistent ensemble dans le temps, l'on perçoit non plus des notes isolées, mais un accord, qui constitue l'harmonie. A en croire ces considérations, toute association verticale de partiels évoluant dans le temps serait alors à l'origine du concept d'accord. Il n'en est rien, il s'agit d'une condition nécessaire, mais non suffisante. La troisième partie de l'extrait vient nous le démontrer. A l'instant indiqué par la flèche sur la figure 1.2, les sons purs, constituants de l'accord sont attaqués tous ensemble de manière synchrone. Notre système auditif procède alors à la *fusion* des composantes et nous les entendons non pas comme des hauteurs, mais comme le timbre d'un son unique, une note de gong. Il ne suffit pas non plus de se restreindre à la seule phase d'attaque pour expliquer ce phénomène. Le fait que les composantes décroissent de façon similaire y participe également. Une explication du phénomène peut provenir du fait que "l'oreille" possède alors suffisamment d'éléments pour retrouver une causalité physique : une phase d'excitation (le bref instant où la mailloche rentre en contact avec le gong), et une phase de résonance (la structure mécanique métallique vibrant sur ses modes de résonances). Si le son de gong est bien reproduit artificiellement (synthétisé), deux éléments rentrent encore en jeu. Premièrement, il y a le fait que les différentes composantes ne meurent pas toutes aux même instants. Comme le souligne Risset, "les composantes ont des durées différentes : si elles étaient synchrones, on obtiendrait un son de carillon "électronique" " [Risset, 1991, p. 250]. Les différentes lois d'extinction des composantes spectrales sont directement liées à la physique de la structure vibrante. Aramaki et Kronland-Martinet ont montré qu'en jouant sur les lois d'amortissements des composantes spectrales, on est à même de reproduire des sons d'impacts provenant de structures aux matériaux bien différents

⁵un son est qualifié de pur s'il ne possède qu'une seule composante spectrale (ou partiel), par opposition aux sons complexes qui en possèdent plusieurs. [NDLR : Expression à ne pas confondre avec son envers, un "pur son", qui dans le contexte actuel est souvent utilisé par les rappeurs pour désigner une chanson dont ils approuvent la forte valence émotionnelle !]

⁶Il s'agit là de l'écriture "horizontale", ou mélodique, par opposition à l'écriture "verticale", ou harmonique. On remarquera, au passage, que les termes horizontal et vertical sont employés par référence à la notation traditionnelle.

(métal, bois, verre, etc.) [Aramaki et Kronland-Martinet, 2006]. Le deuxième ingrédient nécessaire à la reproduction fidèle du son de gong tient en la relation inharmonique qui existe entre les composantes spectrales mises en jeu. Ainsi, les relations de fréquences entre les composantes du son agissent sur le timbre perçu. Ceci n'étant pas des plus intuitifs, voici une bonne occasion de rappeler quelques bases de (psycho)acoustique musicale qui nous resserviront par la suite. Remontons pour ce faire quelques instants dans le passé afin de retracer l'origine des échelles musicales et des notions de séries harmonique et inharmonique.

1.1.3.2 Echelle chromatique tempérée, série harmonique, série inharmonique

Perception des hauteurs Depuis le milieu du XIX^e , l'échelle de référence de la musique occidentale est l'échelle chromatique tempérée au sein de laquelle l'intervalle⁷ d'octave, qui correspond à un doublement de la fréquence physique, est divisé en douze demi-tons égaux (cf. figure B.1 de l'Annexe B). Ainsi vient d'être défini le tempérament égal pour lequel le plus petit intervalle est le demi-ton tempéré qui représente $1/12^e$ d'octave. Mais comment établir ces douze parts égales d'octave ? Est-ce là juste un problème d'arithmétique, ou bien la perception n'est-elle pas également à prendre en compte ? Nous vient à tous l'image du gâteau découpé avec équité afin de satisfaire de la manière la plus juste possible l'appétit de tous les convives ! Mais l'équité des intervalles est d'un autre ordre en musique et ne peut être seulement un problème d'arithmétique, car les lois de la perception sonore entrent en jeu. Il suffit pour s'en rendre compte (visuellement !) d'observer la disposition des touches sur une guitare⁸). Une correspondance existe en effet entre les douze parts d'octave dont nous parlions et les largeurs des touches de la guitare, schématisées sur la figure B.2 de l'Annexe B. En appuyant la corde contre la touche à l'aide du doigt, le musicien change ainsi la longueur de la corde vibrante, longueur qui est, en première approximation, inversement proportionnelle à la fréquence fondamentale de la note émise. La figure B.2 montre que la largeur des touches n'est pas constante lorsque l'on se déplace du grave vers l'aigu (lorsque l'on réduit la longueur vibrante de la corde) : les distances deviennent de plus en plus petites. Ceci n'est pas sans raisons et découle de notre perception. En effet, afin de bâtir des intervalles de hauteurs qui procurent des sensations similaires quel que soit le registre⁹, ce sont les *rappports* de fréquences entre les sons, et non leurs différences, qui doivent être conservés. Afin de rester proche du fonctionnement de notre système auditif, les transformations sur les intervalles s'établissent donc sur les rapports de fréquences de manière logarithmique et non linéaire¹⁰. La figure B.3 de l'Annexe B montre deux opérations réalisées sur un intervalle initial de quarte, l'une préserve les différences de fréquences des deux composantes au cours du temps (cf. Exemple Sonore n° 4), l'autre préserve leurs rapports de fréquences et réalise ainsi un isomorphisme de l'intervalle initial dans le temps (cf. Exemple Sonore n° 5). La nature des intervalles musicaux est donc directement reliée à des faits perceptifs et ne découle pas d'une construction de l'esprit *a priori*, ce que la définition de l'échelle chromatique tempérée donnée précédemment tendait à faire croire.

Le système pythagoricien En Occident, c'est au VI^e siècle avant J.-C que le mathématicien et philosophe grec Pythagore indique que les vibrations des cordes des instruments de musique produisent un son harmonieux seulement si le *ratio* des longueurs des cordes est un nombre en-

⁷On appelle intervalle, le matériel mélodique et harmonique le plus simple et qui réfère à la distance de hauteurs entre deux sons.

⁸non "fretless" de préférence ! L'expression anglaise "fretless" signifie sans frettes, les frettes étant de fines barres métalliques transversales au manche de la guitare et permettant de séparer les touches, zones du manche où le guitariste pose ses doigts pour émettre les différentes notes.

⁹"partie de l'étendue totale d'un ensemble de sons (extrême grave, grave, médium, aigu et suraigu). Sur beaucoup d'instruments et sur la voix, un registre se caractérise par une homogénéité de sonorité. Les changements de registre jouent un important rôle expressif." [Siron, 2004].

¹⁰additionner des intervalles revient à multiplier les rapports de fréquences, soustraire des intervalles, revient à les diviser.

tier petit. Comme le rappelle Schaeffer citant Chailley, Pythagore, ne “s’est pas fixé pour but de déterminer les intervalles produits par des rapports préétablis mais, il a été frappé de la coïncidence entre des intervalles déjà reconnus, et les rapports de longueurs de corde.” [Schaeffer, 1966, p. 606]. Revenons brièvement sur la genèse de l’échelle chromatique tempérée et de la gamme diatonique. Pythagore a fondé par l’addition des quintes un système acoustique cyclique permettant d’obtenir une gamme heptatonique (à sept sons) et qui porte d’ailleurs son nom : la gamme diatonique¹¹ pythagoricienne. Avec un tel système, un empilement de douze quintes ramené au sein d’un octave (par division du nombre adéquat de puissance de 2) ne permet pas de retomber exactement sur la note de départ ; l’écart obtenu est appelé le comma pythagoricien.

Le système zarlinien L’introduction de tierces permettant des constructions harmoniques intéressantes prend son essor à la Renaissance, grâce aux travaux du Vénitien Zarlino. Le système acoustique qu’il développe implique une hiérarchie des notes autour d’un centre tonal. Cependant, le système zarlinien est mal approprié aux changements de centre tonal (modulations), car lorsque la tonique change, la hauteur des notes varie également (ce qui explique pourquoi les pièces écrites pour clavecin, instrument dont l’origine remonte vers 1400 [Fletcher et Rossing, 1998, p. 340] et accordé suivant le système zarlinien, ne “sonnent” bien que dans une tonalité précise).

Le système tempéré Le tempérament égal dont nous avons parlé ci-dessus représente une solution intermédiaire entre le système mélodique pythagoricien et le système harmonique zarlinien et présente l’avantage de permettre des modulations de tonalité (justement dans les douze tonalités de l’échelle chromatique tempérée, cf. Le Clavier Bien Tempéré de Bach). Ceci est cependant réalisé aux dépens de la justesse de quelques notes, notamment celles des quartes et des quintes¹², comme le montre la figure B.4 de l’Annexe B, qui présente une comparaison des intervalles les plus usuels dans les différents systèmes pythagoricien, zarlinien et tempéré. Ainsi, nous avons vu quelques clefs permettant de comprendre pourquoi le tempérament égal s’est imposé dans la musique occidentale. Les instruments dont il sera question dans nos expériences reposent sur le tempérament égal. En partant de la conservation des rapports de fréquences et du principe que la gamme chromatique sous-tend un doublement de la fréquence fondamentale en douze demi-tons, la loi qui régit les intervalles de l’échelle chromatique tempérée peut s’exprimer ainsi : *une note située à un intervalle de d demi-tons d’une autre note a une fréquence $2^{\frac{d}{12}}$ plus élevée que celle de l’autre note*. Si nous raisonnons en terme de longueur de corde vibrante pour un instrument tel que la guitare, ce sont les rapports de longueur de corde vibrante qui doivent être invariants, puisque en première approximation, la longueur de corde vibrante est inversement proportionnelle à la fréquence fondamentale de la note émise¹³. Si nous ajoutons la condition que la moitié de la corde doit être atteinte en douze cases, nous avons là un moyen de calculer la largeur des touches et par la même occasion accompli une (infime) partie du travail du luthier. La figure B.5 de l’Annexe B présente une confrontation entre les valeurs théoriques des intervalles de l’échelle chromatique tempérée et des valeurs d’intervalles mesurées sur un instrument à tempérament égal, la guitare classique (les valeurs des intervalles ont été obtenues à partir de mesures de la largeur des touches (cf. figure B.2 de l’Annexe B).

¹¹le terme diatonique signifie que les mouvements mélodiques et harmoniques sont formés à partir des notes d’une seule gamme, et hiérarchisés autour d’un note prépondérante, la tonique.

¹²Les systèmes de Pythagore et de Zarlino présentent tous deux des quarte et quinte justes, respectivement définies par les rapports $\frac{4}{3}$ et $\frac{3}{2}$.

¹³La loi qui régit la longueur de corde vibrante L et la fréquence f_1 du premier mode de vibration transverse dans le cas d’une corde idéale, dénuée de raideur est : $f_1 = \frac{1}{2L} \sqrt{\frac{T}{\mu}}$, où T est la tension de la corde, et μ sa masse linéique. La fréquence fondamentale f_1 est donc inversement proportionnelle à la longueur L .

Séries harmonique et inharmonique Nous venons de rappeler la définition du tempérament égal, mais qu'en est-il de la série harmonique. Le terme de série harmonique¹⁴ voit le jour au *XVII^e* siècle grâce aux travaux du physicien français Sauveur. La découverte de la série harmonique amène des bases théoriques afin de confirmer les règles empiriques de formation des accords, et permet donc d'expliquer les constatations de Pythagore indiquant que les vibrations des cordes des instruments de musique produisent un son harmonieux seulement si le ratio des longueurs des cordes est un nombre entier petit. En effet, les notions de consonance/dissonance entre les sons instrumentaux sont directement liées aux nombres d'harmoniques qu'ils ont en commun (pour de plus amples informations, voir l'illustration donnée par Schaeffer dans le cas d'un système zarlinien [Schaeffer, 1966, p. 618]). On peut se demander si les fréquences des composantes d'une série harmonique correspondent aux fréquences de sons issus du tempérament égal. Comme le montre la figure 1.3, cela dépend des harmoniques considérés. Hormis les intervalles d'octaves pour lesquels la correspondance est parfaite, il existe de subtiles et parfois mêmes larges différences : par ex. pour les 3^e, 6^e, et 12^e composantes de la série harmonique, associées à des quintes dans le système tempéré, les écarts de fréquences sont faibles, d'environ 2 cents, mais pour la 13^e composante harmonique, l'écart fréquentiel par rapport à la note la plus proche dans le système tempéré (une 13^e, ou 6^{te} si on la ramène à l'intérieur de l'octave) est grand, d'environ 60 cents.

¹⁴Dans une série harmonique, les composants sont des multiples entiers de la fréquence la plus basse, dite fréquence fondamentale. La série harmonique se caractérise donc par la série entière (1, 2, 3, 4, 5, etc.). Par opposition, toute série sera dite inharmonique si les fréquences des composants sonores sont dans des rapports non rationnels.

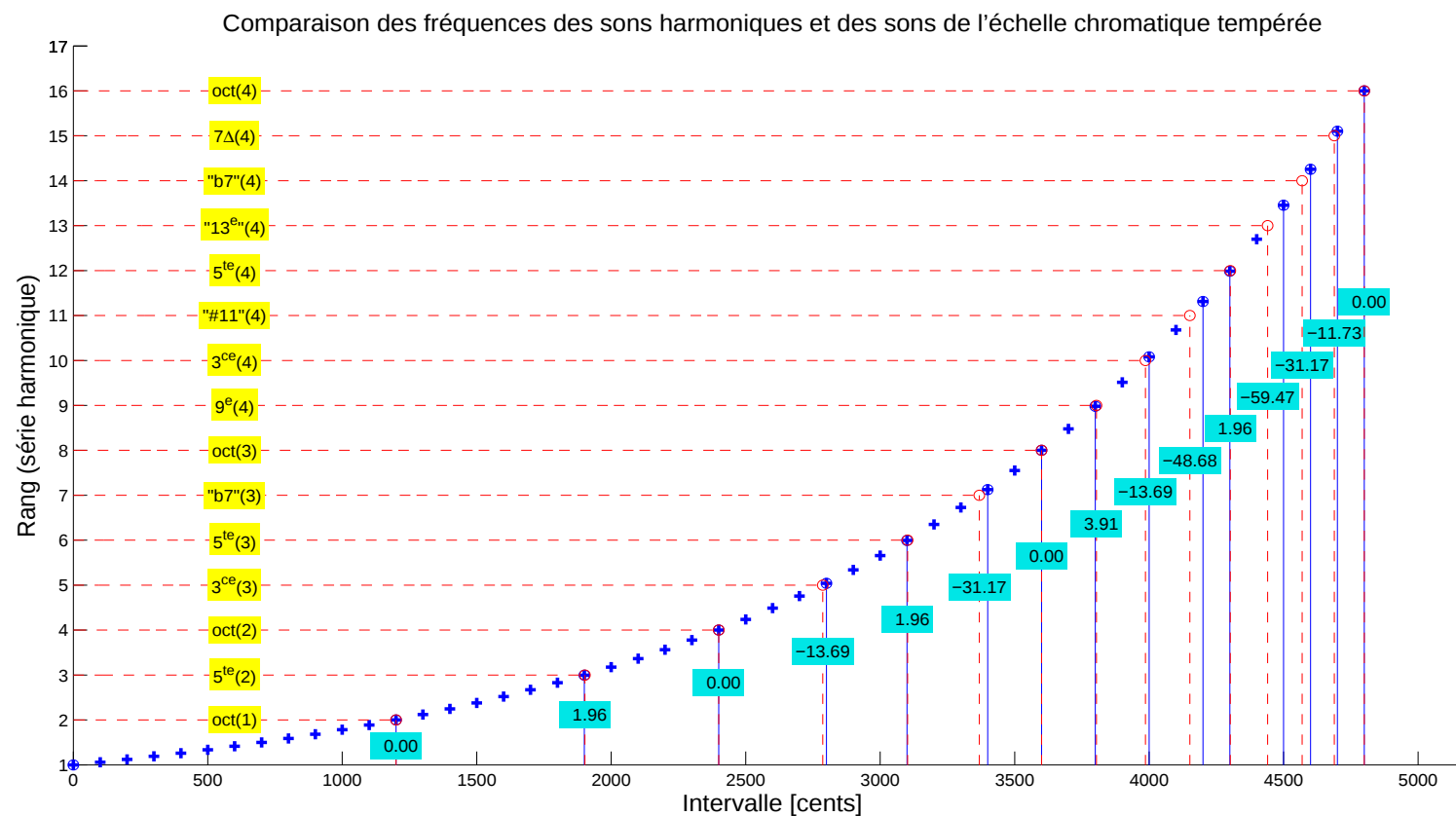


FIG. 1.3 – Comparaison des fréquences des sons harmoniques (ronds) et des sons de l'échelle chromatique tempérée (croix). Les écarts entre les fréquences des harmoniques et celles des notes les plus proches dans le système tempéré (exprimés en cents) sont présentés sous les points de la courbe (intervalles tempérés / rangs harmoniques). Les correspondances entre les harmoniques et les intervalles du système tempéré associés sont indiquées sur la gauche à côté des rangs harmoniques. Le numéro entre parenthèses réfère à l'indice d'octave. Lorsque l'intervalle est entouré de guillemets, cela signifie que les écarts de fréquences sont importants. Ainsi, l'on voit que le 3^e harmonique correspond à quelques cents près (1.96) à une note située une quinte plus un octave au-dessus de la note fondamentale selon l'échelle tempérée (le 3^e harmonique est également communément appelé 12^e et revêt une grande importance dans le cas de la clarinette, comme nous le verrons au chapitre 3).

1.1.3.3 L'analyse par la synthèse

Ayant définies ces notions importantes, revenons à présent à la question du timbre et à l'exemple musical de Risset, extrait de *Mutations*. Les notes utilisées au-dessus du $C\sharp^{15}$ sont successivement le Bb (sixte), le D (seconde bémol, oct. 2), le E (tierce bémol, oct. 2) et le A (quinte augmentée, oct. 2). Vues de manière verticale (harmonique), ces notes forment un accord altéré très tendu¹⁶ $C\sharp^{-13(9b\ 5+)^{17}}$. Aucune des notes supérieures au $C\sharp$ ne fait partie de sa série harmonique (cf. figure 1.3). La série en présence est donc inharmonique, caractérisée, si l'on se repose sur le système tempéré, par les rapports successifs : 1, 1.68, 2.11, 2.38, 3.17. L'inharmonicité apparaît donc comme l'un des ingrédients essentiels pour reproduire le timbre si caractéristique du gong. On pourra comparer les deux sons des Exemples Sonores *n°s* 6 et 7, tous deux générés à partir d'une enveloppe d'amplitude correspondant à des objets sonores de type "percussif", constituée d'une phase d'attaque brève, suivie d'une phase de décroissance exponentielle, mais utilisant respectivement une série inharmonique, fondée sur les rapports présentés ci-dessus, et une série harmonique 1, 2, 3, 4, 5. Notons que dans ces exemples les composantes spectrales ont toutes des durées identiques, ce qui ne permet pas la reproduction parfaite d'un son de type gong dans le cas de l'Exemple Sonore *n°* 6). La figure 1.4 illustre le procédé utilisé par le compositeur au début de sa pièce *Mutations*.

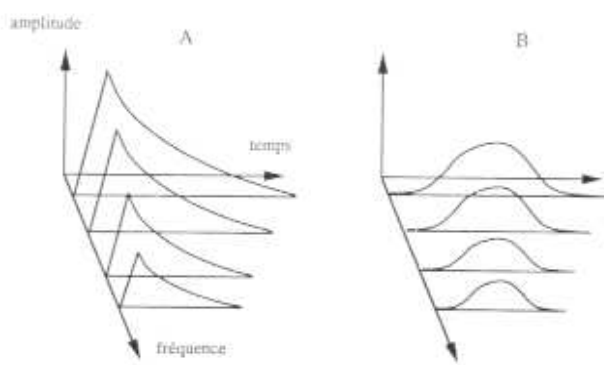


FIG. 1.4 – A : son inharmonique dont les composantes spectrales possèdent des enveloppes synchrones de type attaque/décroissance : son de type "gong" (correspond à la 3^e partie de l'extrait de *Mutations*). B : son inharmonique dont les composantes spectrales possèdent des enveloppes asynchrones de type gaussienne, les maxima d'amplitude ne sont pas atteints aux mêmes instants : l'impression qui en découle est celle d'une texture fluide aux composantes dispersées (correspond à la 2^e partie de l'extrait de *Mutations*). Figure issue de [Risset, 1991, p. 250].

Cet exemple permet de constater que l'on ne peut pas parler de timbre sans parler de la composition harmonique des sons. Timbre et hauteurs ne sont donc pas orthogonaux et semblent s'influencer mutuellement. Il n'est pas un hasard si l'on parle en musique de la "couleur" d'un accord. Celle-ci est étroitement liée aux notes qui composent l'accord et donc aux différents intervalles mis en jeu, qui comme nous avons pu le voir - au travers d'exemples - agissent sur le timbre perçu. Nous avons également pu constater que la nature des évolutions temporelles des amplitudes des composantes spectrales agit sur la sensation de fusion. Les aspects temporels influent donc aussi

¹⁵Le système américain de notations des notes de musique est utilisé dans le document. Le do est représenté par C, le ré, par D, le mi par E, etc.

¹⁶L'alternance de *tensions* et de *résolutions* (détentes) est fondamentale en musique. Elle constitue la notion de cadences que l'on retrouve communément dans de nombreux genres musicaux, classique, jazz, variété, etc.

¹⁷Une notation des accords utilisée en jazz est ici employée : le signe $-$ signifie mineur, les altérations sont indiquées entre parenthèses, $9b$ désigne la neuvième bémol, et $5+$ désigne la quinte augmentée.

sur le timbre. Nous aurons l'occasion de revenir plus en détails sur ces points au chapitre 2. Enfin, et c'est là un point important, l'analyse de l'extrait sonore de Risset et les conséquences que nous en avons pu en tirer sur la perception illustre bien la démarche de *l'analyse par la synthèse* [Risset et Mathews, 1969], l'un des fondements de la méthodologie employée dans cette thèse. La figure 1.5 résume cette démarche.

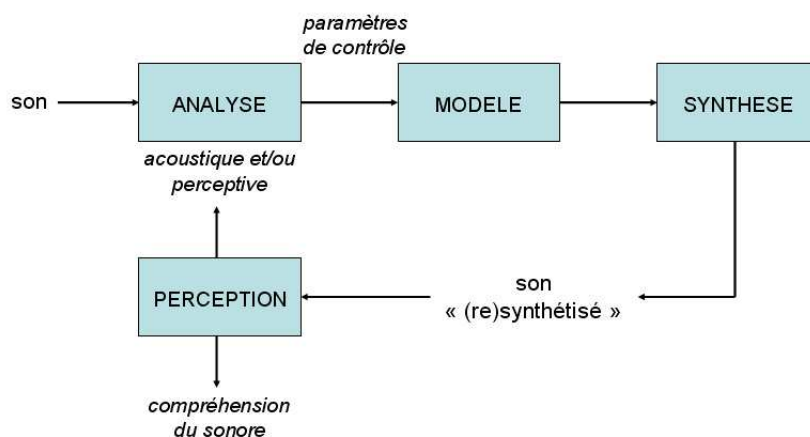


FIG. 1.5 – La démarche d'analyse par la synthèse

1.1.4 De l'utilisation du timbre dans les musiques traditionnelles

L'intérêt croissant porté au timbre au XX^e siècle tant chez les musiciens que chez les scientifiques, ce parallélisme n'étant certainement pas le fruit du hasard, amène à se poser la question suivante : est-ce que l'importance du timbre ne se révèle que dans les musiques contemporaines, de par l'invention de nouveaux systèmes compositionnels et de production (synthèse), ou est-ce qu'au contraire son rôle n'a-t-il pas toujours été essentiel, mais jusqu'alors masqué dans le cadre du système tonal occidental par la domination des hauteurs, des durées et des intensités ? L'univers sonore rendu possible grâce à la synthèse semble inépuisable, mais comme le dit Schaeffer, les machines aussi ont peut être leurs propres limites, "peut-être n'atteindront-elles jamais la subtilité et la sensibilité d'un son artisanal" [Schaeffer, 1966, p. 631]. Nous montrerons dans cette thèse que le timbre ne peut être astreint qu'au simple rôle de moyen, de matérialité sonore, résultat d'une causalité physique, permettant de traduire les notes d'une partition musicale. Quelques exemples instructifs sur l'utilisation du timbre dans les musiques instrumentales méritent d'être cités. En effet, si la mélodie, l'harmonie, et le rythme semblent façonner les musiques traditionnelles, de subtiles utilisations du timbre peuvent parfois s'y cacher.

Musique traditionnelle vocale Remontons donc l'espace d'un instant au Moyen Âge avec un chant populaire polyphonique écrit dans la langue d'Oc : *El cant del boier* (traduction : La chanson du bouvier¹⁸). Il s'agit d'un chant profane (c'est à dire non religieux), qui a d'ailleurs pour origine le pays cathare. L'Exemple Sonore n° 8 correspond au premier couplet de la chanson d'après une interprétation de la chorale *Les Voix Là* (Marseille). Les paroles associées à cet extrait (traduites en français) et sa partition sont respectivement données par les figures 1.6 et 1.7. La caractéristique commune et surprenante de tous les couplets est la présence des voyelles chantées "A, É, I, O, U". Ces voyelles n'ayant a priori pas de significations intrinsèques (nous verrons plus tard en quoi ceci doit être mis en rapport avec des références culturelles), elles semblent être utilisées

¹⁸le bouvier est celui, ou celle qui s'occupe des boeufs, les garde, conduit leur attelage.

par le compositeur pour leurs différences de couleurs sonores, leurs différents timbres. Même si le passage d'une voyelle à l'autre est couplé à un mouvement descendant d'intervalles de quarte jusqu'au "U" final à partir duquel la hauteur des voix remonte, il n'y a pas moins là un bel exemple d'utilisation des variations de timbre en musique. Les phonéticiens ont en effet montré que les variations de la qualité des voyelles sont perçues de manière analogue à des variations de timbre [Slawson, 1968]. Comme le souligne Risset citant Mathews : "A ceux qui doutent de la possibilité d'articuler un discours musical à partir du timbre, Max Mathews rappelle qu'il existe un système de communication élaboré qui se fonde sur les différenciations de timbre : la parole !" [Risset, 1991, p. 258-259]. La qualité des voyelles et le timbre ont des corrélats acoustiques en commun : les formants.

Dans le cadre de la voix parlée ou chantée, le conduit vocal (larynx, pharynx, bouche et nez pour les sons nasaux) agit comme un résonateur qui filtre la "voix-source"¹⁹, ou son émis par le flux d'air provenant des poumons et qui engendre la vibration des cordes vocales [Sundberg, 1995]. Ce que l'on appelle formant dans le cadre de la voix correspond aux 4 ou 5 résonances importantes du conduit vocal²⁰. La forme du conduit vocal peut être modifiée suivant la position des organes articulatoires (lèvres, mâchoire, langue, et larynx). Tout changement de section du conduit vocal entraîne un déplacement des fréquences de chaque formant. Ceci détermine le son des voyelles engendrées. A titre indicatif, si l'on assimile le conduit vocal à un cylindre parfait (d'une longueur de 17.5 cm pour un adulte mâle), fermé au niveau de la glotte et ouvert au niveau des lèvres, les fréquences des quatre premiers formants seraient environ 500, 1500, 2500, 3500 Hz (autrement dit une série d'harmoniques impairs, comme celle associée au résonateur cylindrique de la clarinette de type "fermé/ouvert", que nous analyserons au chapitre 3).

1. Quand le bouvier vient de labourer, (bis) Il plante son aiguillon, A, E, I, O, U ! Il plante son aiguillon.	5. Quand je serai morte, enterre-moi, (bis) Au plus profond de la cave, A, E, I, O, U ! Au plus profond de la cave.
2. Il trouve sa femme au pied du feu, (bis) Triste et inconsolable, A, E, I, O, U ! Triste et inconsolable.	6. Et les romains (ndlr: pèlerins) qui passeront, (bis) Prendront l'eau signée (ndlr: bénite), A, E, I, O, U ! Prendront l'eau signée.
3. Si tu es malade, dis-le-moi, (bis) Je te ferai un potage, A, E, I, O, U ! Je te ferai un potage.	7. Et ils diront celle qui est morte ici, (bis) C'est la pauvre Jeanne, A, E, I, O, U ! C'est la pauvre Jeanne.
4. Avec une betterave, un navet, (bis) Une alouette maigre, A, E, I, O, U ! Une alouette maigre.	8. Qu'elle s'en est allée au Paradis, (bis) Au ciel avec ses chèvres, A, E, I, O, U ! Au ciel avec ses chèvres.

FIG. 1.6 – Paroles de la chanson "Lo cant del boier" (traduction en français d'après [Plantevin, 2008]).

¹⁹Le spectre de la voix source est constitué d'une fréquence fondamentale et d'un grand nombre d'harmoniques dont les amplitudes décroissent de manière uniforme avec la fréquence (pente d'environ -12dB / octave). La fréquence fondamentale des sons parlés ou chantés est fixée par la fréquence de vibration des cordes vocales.

²⁰Par extension, un formant correspond d'une manière plus générale à une zone spectrale où l'énergie est prononcée (bosse dans le spectre). A l'inverse, un anti-formant correspond à une zone spectrale où l'énergie est faible (creux dans le spectre).

Lo boier

chant 1
chant 2
voies

quand lo bo- ié- ér ven de lau- rar quand lo bo- ié- ér ven

Quand lo bo- ier ven de lau- rar quand lo bo- ier ven

chant 1
chant 2
voies

de lau- rar plan ta son a gu lha- da a

de lau- rar plan- ta son a- gu- lha- a da a a a

chant 1
chant 2
voies

plan ta son a gu

é é i i o o u u u plan- ta son- a- gu-

utilisation des variations de timbre au travers des voyelles A, E, I, O, U

chant 1
chant 2
voies

lha- a- da tro-

lha- a- a- a- a- a- da

FIG. 1.7 – Partition du premier couplet du champ populaire occitan datant du Moyen Age “Lo cant del boier” (La chanson du bouvier) : un exemple d’utilisation des variations de timbre au niveau compositionnel.

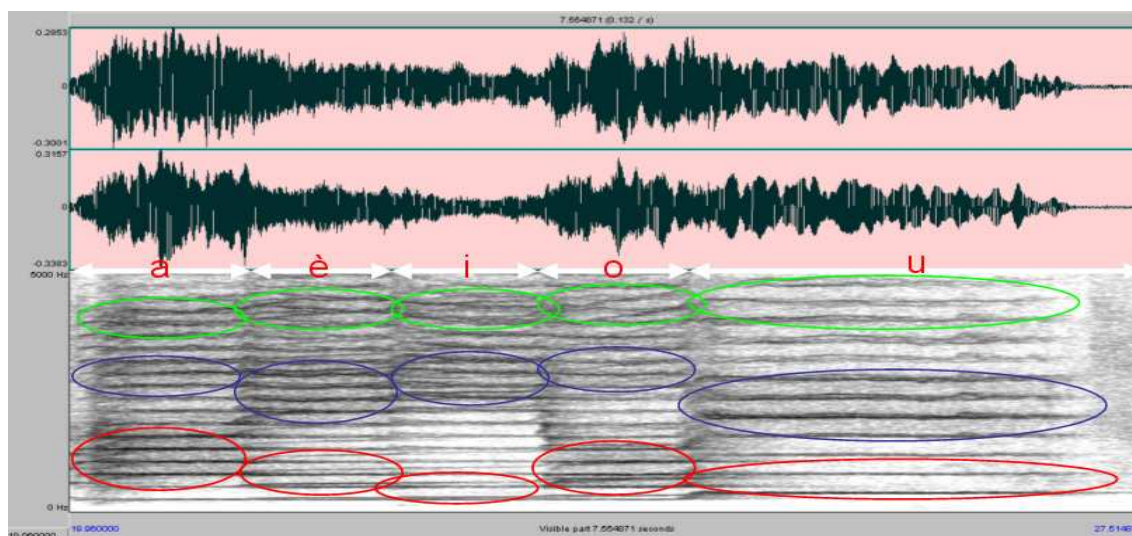


FIG. 1.8 – Signal temporel (haut) et spectrogramme (bas) de la partie correspondant aux voyelles A, E, I, O, U. Les analyses ont été effectuées avec le logiciel *Praat* (les fenêtres d’analyse sont des gaussiennes d’une durée de 100 ms).

Le spectrogramme correspondant au passage des voyelles est présenté sur la figure 1.8. Les positions qualitatives des principales zones formantiques associées à la production des différentes voyelles sont également indiquées sur la figure 1.8. Notons que comme il s’agit d’un chant polyphonique à cinq voix, ces zones formantiques sont le fruit de l’association des différentes voix. Les déplacements fréquentiels des formants lors du passage d’une voyelle à l’autre sont corrélés aux changements de couleurs sonores.

Il s’avère que la symbolique de cette chanson, dont les paroles peuvent paraître anodines au premier abord, est en réalité profonde. Dans une analyse proposée par Plantevin [Plantevin, 2008], nous apprenons en effet que ce chant revêtait une fonction bien particulière du temps des attaques des catholiques romains contre les cathares, celle de pouvoir s’avertir d’une vallée à l’autre. Si l’on peut rester dubitatif face à cette interprétation, l’explication qui suit lui confère une certaine authenticité. Les voyelles de la chanson étaient souvent chantées dans un ordre différent par les cathares. En jouant sur l’ordre des voyelles, les cathares pouvaient ainsi communiquer différents messages de manière codée. A l’aide des combinaisons des 5 voyelles A, E, I, O, U, ils avaient ainsi à leur disposition $5! = 120$ messages différents ! Les voyelles utilisées sont aussi les initiales de la devise des rois d’Aragon : “Austri Est Imperare Orbi Universum” (l’empire du monde appartient au midi). La femme malade dont il est question dans le 2^e couplet serait en réalité - ou est devenue lors des guerres - un symbole représentant l’église cathare attaquée ! Il semble maintenant que les voyelles utilisées dans ce chant ne sont pas complètement dénuées de significations. Elles n’ont peut-être pas seulement été utilisées pour la joliesse de leurs couleurs comme nous l’avions naïvement supposé plus tôt. Voilà que le timbre qui juste à présent revêtait une fonction purement musicale devient un subtil stratagème de défense contre l’assaillant. L’auteur n’en cache pas sa surprise !

Musiques traditionnelles asiatiques Si les notations de la musique traditionnelle occidentale omettent le timbre sinon par référence à l’instrument, ce n’est pas le cas de toutes les musiques traditionnelles. Le rôle du timbre est en effet prédominant dans certaines musiques traditionnelles asiatiques. En Chine, un système de notation très complexe comprenant plus d’une centaine de symboles a été développé pour représenter les différentes techniques de jeu permettant d’agir sur

le timbre du *Qin*, une ancienne cithare chinoise à sept cordes [Traube, 2004]. Comme nous l'apprend François Picard, "L'art du *qin* se trouve intimement lié à la richesse de timbres permise par l'obtention de la même hauteur sur des cordes différentes. Très peu de joueurs, encore de nos jours pourraient, à l'instar des pianistes, déchiffrer à vue une partition dans son tempo. La complexité des instructions, corollaire de celle des choix à effectuer pour jouer une note, implique un lent travail de mise en place." [Picard, 2003, p. 29]. La tablature pour cithare *Qin* présentée figure 1.9 montre les symboles utilisés pour indiquer à l'interprète que la note doit être produite avec une corde à vide²¹. Etant donné que les différentes cordes du *Qin*, traditionnellement faites de soie torsadée, n'ont pas le même diamètre, il en résulte de subtiles différences de timbre pour des sons de même hauteur produits sur des cordes distinctes.

Partition synoptique
Shijing yuepu, 1788, *biao* 12

sol abaissé aigu	吉	應鍾清羽立宮林鍾清變徵主調爲高五字調
Diatonique sur la base du système des Cinq sons	清變徵	
Gongche pour flûtes : droites, de Pan, globulaires, transversales	簫排簫 壎 篪	
Gongche pour flûtes traversières et orgues à bouche	笙 笛	
Chromatique pour carillons de cloches et de pierres	磬 鐘 林 鍾	
corde 4 à vide : attaque de l'index droit vers l'extérieur	琴 荷	
Tablature pour cithares <i>Qin</i>		
corde 20 corde 15 attaques des index vers l'extérieur	瑟 螭	
Tablature pour cithares <i>Se</i>		

le Si chromatique (sixte aiguë) comme fondamentale,
le Sol chromatique (quinte abaissée aiguë) comme tonalité principale,
soit tonalité du caractère de La aigu

FIG. 1.9 – Partition synoptique pour différents instruments traditionnels chinois dont la cithare *Qin* (d'après [Picard, 2003, p. 36]).

Les Japonais, nous révèle Yoshihiko Tokumaru, sont sensibles au timbre : "Une telle sensibilité est liée à la tendance qu'ont les Japonais à négliger les hauteurs absolues et les intervalles dans leur perception" [Tokumaru, 1991, p. 90]. Elle est probablement également en lien avec la nature de leur langue orale qui repose sur de subtiles variations sonores. La tendance qu'ont les Japonais à négliger les hauteurs se retrouve dans la façon d'entendre les sons non musicaux. Les Japonais représentent par exemple les cris des oiseaux et des insectes par des combinaisons de voyelles et de consonnes (et donc de timbres, comme nous en avons discuté ci-dessus), et non des relations d'intervalles : "Un *matumusi*, grillon du pin, est censé chanter "tin ti ro rin", alors qu'un *kôrogi*, grillon, aurait pour chant "ke ro ke ro ke". Ici les distinctions entre /t/ et /r/ - grillon du pin

²¹i.e. aucun doigt ne vient presser la corde ; ceci implique un choix compositionnel lié au timbre, car un son de même hauteur aurait pu être obtenu sur une autre corde en raccourcissant la longueur de corde vibrante par pression du doigt.

- et celles entre /k/ et /r/ - grillon - montrent le souci de distinguer les timbres parmi les insectes." [Tokumaru, 1991, p. 91].

Dans certaines pièces, le **syamisen (shamisen)**, instruments à trois cordes pincées, attaquées au plectre (intermédiaire du banjo et de la guitare)²², doit imiter les appels des grillons du pin. Afin d'effectuer les choix exacts de timbres lors de l'interprétation d'une pièce, les musiciens utilisent une notation orale baptisée *kuti syamisen*²³. Comme pour le Qin chinois, les sons produits par les cordes à vide sont distingués des autres pour leur qualité de timbre, ce que l'on peut remarquer sur la partition de la figure 1.10 : "Par exemple, le son si de la troisième corde à vide est baptisé "ten", alors qu'un son de même hauteur obtenu sur la deuxième corde pressée a pour nom "ton" " [Tokumaru, 1991, p. 94]. L'Exemple Sonore n° 9 (extrait de la pièce *Kageboshi* interprétée par Kikuko Satoh, de l'ensemble *Yamato*) permettra de s'imprégner de la sonorité de cet instrument dont les particularités sont rappelées dans ces propos de Landy : "La qualité du son de cet instrument léger n'a pas d'équivalent exact en Occident. Elle est essentielle pour les Japonais qui attachent beaucoup d'importance à ses variations de timbres. L'oreille occidentale ne s'y accoutume pas facilement. On a même pu voir dans le son sec et raclé de l'instrument "l'antithèse des sons harmonieux généralement utilisés par la musique occidentale du XIX^e siècle" (Cunningham)" [Landy, 1996, p. 124].

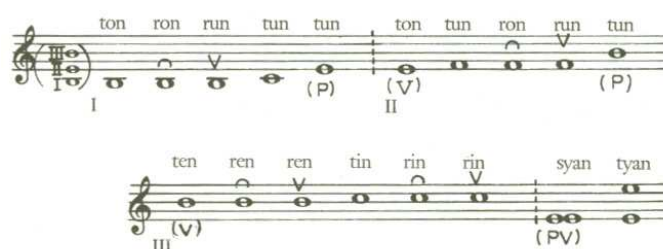


FIG. 1.10 – Exemple de partition pour shamisen. La notation orale (*kuti syamisen*) est indiquée au-dessus des notes. Le (p) signifie position pressée, et le (v) signifie corde à vide. On remarquera les deux sons de même hauteur mais de timbres différents symbolisés par "tun" et "ton" (figure issue de [Tokumaru, 1991, p. 94]).

Dans le cas du **ko-tsuzumi (ou kotudumi)**, petit tambour d'épaule japonais en forme de sablier, utilisé dans le théâtre de *Nô* (drame musical dansé), un système précis de classement des timbres a été établi. Ce système met en évidence quatre sortes de timbres, obtenus selon différents types de frappes et de tensions de la membrane de la percussion, sur laquelle l'instrumentiste peut agir au moyen de ficelles (*kasira* : haut et fort, *kan* : haut et faible, *otu* : bas et bien résonant, *hodo* : plus bas et fréquemment plus faible). Une notation orale, le *kuti syôga*, existe comme pour le shamisen. Chacun des signes de timbres mentionnés précédemment possèdent donc des équivalents oraux (respectivement /ta/, /ti/, /po/, et /pu/). Il faut également souligner que pour cette percussion, les changements de timbre s'accompagnent généralement de changements de hauteur, même si cette dernière n'est pas bien déterminée, à l'inverse des sons harmoniques. L'Exemple Sonore n° 10, extrait d'un enregistrement de la musique d'une pièce de théâtre *Nô* (*Shakkyô* : *Le Pont en pierres*), permet de se rendre compte que les sons du tambour *ko-tsuzumi* constituent la trame rythmique

²²L'ancêtre du shamisen, le "sanshien" chinois, était tendu de la peau des serpents de grande taille connus dans ces régions. Les Japonais n'ayant pas de peau de serpent de la taille suffisante remplacèrent cette dernière par des peaux de chats (ou de chiens sur les instruments bon marché) ; "Les peaux choisies par les Japonais donnent un son plus fort et plus clair. Elles réagissent mieux aux touches et aux glissando délicats." ([Landy, 1996, p. 119]).

²³Cette notation orale représente également un élément pédagogique, permettant au maître d'apprendre une phrase à son élève.

permettant d'assurer la structure musicale de la pièce. Les voix que l'on entend dans l'extrait suggèrent l'ambiance dans laquelle l'acteur principal, qui prend l'apparence de l'"Esprit du Lion", effectue une danse très énergique. Notons au passage une autre particularité du théâtre de Nô qui a des répercussions sur le timbre de la voix de l'acteur principal. Les acteurs utilisent des masques, d'une part pour l'esthétique visuelle, mais aussi par la nécessité d'interpréter deux rôles, et surtout des rôles de femmes (les acteurs sont tous des hommes). D'autre part, au niveau acoustique, ces masques ont pour conséquence de transformer le timbre de la voix des acteurs en le rendant plus grave car ils coupent certaines fréquences aiguës. La diction est ainsi rendue non réaliste, ce qui est propice à l'univers théâtral du Nô, où le personnage principal connaît souvent une transformation entre les deux parties de la pièce (forme réincarnée puis aspect véritable) [Tamba, 1988].

Le **shakuhachi** japonais, une lourde flûte droite de bambou constituée d'un tuyau ouvert aux deux extrémités (du type flûte traversière), est un instrument réputé pour ses fines modulations de hauteur et de timbre. L'instrument a été introduit au Japon par les prêtres bouddhistes qui le ramenèrent de Chine. Il dispose de quatre trous antérieurs et d'un trou postérieur qui le destinent à produire la gamme japonaise pentatonique. Un bon instrumentiste peut agir sur la hauteur (changements d'un quart de ton ou moins) et le timbre des sons grâce à des mouvements de tête et à différentes techniques de respiration. L'[Exemple Sonore n° 11](#) est extrait de l'une des oeuvres fondamentales du répertoire du shakuhachi, *Kokuu*, qui porte la symbolique de l'harmonie parfaite au sein de l'univers (l'interprétation du morceau est de Nisshaku Yonsun de l'ensemble *Yamato*).

Les quelques exemples cités ci-dessus nous ont permis de mettre en évidence différentes utilisations du timbre dans les musiques chantées ou instrumentales traditionnelles. Nous avons vu qu'elles sont en lien étroit avec la facture instrumentale et la technique de jeu de l'instrumentiste. Nous retiendrons cette phrase de Picard : "L'esthétique traditionnelle se trouve dans le son, donné par la nature et sculpté par l'homme, structurant le temps" [Picard, 2003, p. 6]. Ceci met en exergue le rôle primordial que tient le musicien lors de son interprétation d'une oeuvre écrite. L'interprète n'est pas qu'un traducteur, il est celui qui sculpte le son, et peut par là développer sa personnalité musicale. Tokumaru en parlant des joueurs de ko-tsuzumi le fait bien valoir : "Il va de soi que les techniciens de cet instrument, comme ceux des autres instruments, veulent développer leur originalité dans le timbre. S'ils y réussissent, les auditeurs reconnaissent les musiciens par une seule attaque" [Tokumaru, 1991, p. 93]. Ainsi, nous décelons que l'utilisation du timbre peut tant se révéler au niveau du compositeur, que dans les choix et la maîtrise de l'interprète.

Au bout de la chaîne de communication musicale se trouve l'auditeur dont l'oreille sera sensible ou non aux structures musicales et aux finesses de sonorités. Avant d'aborder la question du rôle du timbre dans l'interprétation musicale, il paraît donc intéressant de se donner un cadre d'analyse général qui considère les différents acteurs du musical : compositeur, interprète, auditeur. Une discipline s'y consacre : la sémiologie musicale.

1.2 Sémiologie musicale

1.2.1 La musique, ou l'art de l'ouïe

Il paraît intéressant afin d'aborder la vaste question de la sémiologie musicale de replacer la musique dans son contexte, celui de l'art. Léon Tolstoï définit l'art comme "une forme de l'activité humaine consistant pour un homme, à transmettre à autrui ses sentiments, consciemment et volontairement, par le moyen de certains signes extérieurs" [Tolstoï, 1931, p. 57]. L'art peut donc se voir comme un système de signes, portant un sens et cherchant à communiquer une signification. La musique en tant qu'"art de combiner les sons" [Larousse, 1957] est donc par essence un moyen

de communication²⁴.

1.2.2 La tripartition de Molino et Nattiez

Si cet aspect de la musique semble aller de soi, la question du sens du message qu'elle véhicule est elle, en revanche, un sujet bien complexe. En se reposant sur des disciplines telles que la linguistique et la phonétique, Molino et Nattiez fondent la sémiologie musicale dans les années 70. Dans [Nattiez, 1975], Nattiez propose un cadre théorique d'analyse du musical depuis sa production, jusqu'à sa perception. Comme le remarque Nattiez, "les phénomènes musicaux sont de nature sémiologique parce qu'ils renvoient à quelque chose d'autre dans le monde extérieur ou dans la pensée de ceux qui les utilisent." (ibidem). Il définit la sémiologie musicale comme "l'entreprise qui prendra en charge la spécificité de la manière dont la musique devient un *fait symbolique pour ses utilisateurs* : le compositeur, l'interprète, l'auditeur, le musicologue." (ibidem, p. 50). Cette analyse s'instaure sur la conception sémiologique de Molino comportant trois pôles : le **niveau "poiétique"** (étymologiquement : faire), qui fait référence aux stratégies de production du message musical, le **niveau "neutre"** qui concerne le message lui-même dans sa réalité matérielle (la partition, l'oeuvre jouée), et le **niveau "esthétique"** (étymologiquement : faculté de percevoir) qui s'attache à décrire les stratégies de réception du message (ibidem).

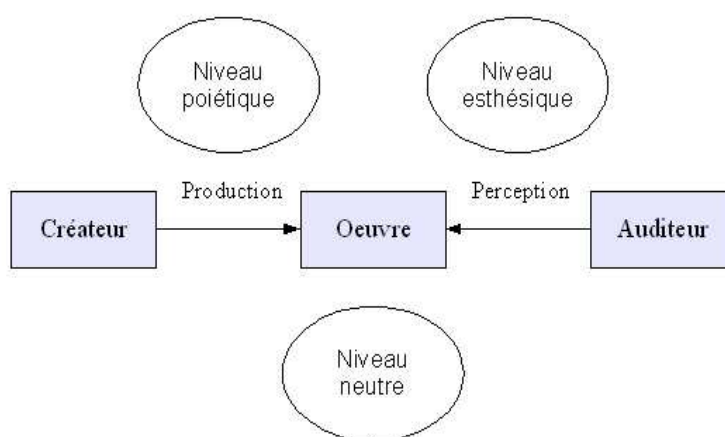


FIG. 1.11 – Sémiologie musicale : la tripartition de Molino et Nattiez (selon [Nattiez, 1975, p. 52]).

1.2.3 La question du sens musical

D'après la tripartition, schématisée figure 1.11, il paraît clair que le sens musical est à double niveau, d'un côté celui du créateur (compositeur dans le cas de musiques écrites ou instrumentiste dans le cas de musiques improvisées), de l'autre celui de l'auditeur. Selon Tolstoï, l'art est fondé sur "l'aptitude de l'homme à éprouver les sentiments éprouvés par un autre homme" [Tolstoï, 1931]. Cependant, les différences de significations que peut avoir l'objet d'art selon le point de vue du créateur ou selon le point de vue du récepteur semblent être négligées dans cette conception de l'art. En effet, le domaine de l'esthétique ne correspond pas nécessairement à celui de la poiétique. Il paraît clair qu'une oeuvre musicale n'a pas un pouvoir expressif unique. Cette citation de Descartes rapportée par Nattiez le souligne : "La même chose qui fait envie de danser à quelques-uns, peut donner envie de pleurer aux autres" [Nattiez, 1975, p. 131]. Le principe que

²⁴Il est d'ailleurs amusant de constater que les premiers programmes de synthèse musicale naquirent dans un laboratoire de télécommunication, le *Bell Labs*.

sous-tend la vision de Tolstoï, selon qui la communication artistique repose sur une transmission de sentiments, est loin de faire l'unanimité. La position de Schönberg, par exemple, est toute autre ("La musique parle dans sa propre langue de matières purement musicales"), et semble rejoindre celle de Stravinsky ("Je considère la musique par essence impuissante à exprimer quoi que ce soit : un sentiment, une attitude, un état psychologique, un phénomène de la nature, etc. L'expression n'a jamais été la propriété immanente de la musique. La raison d'être de celle-ci n'est d'aucune façon conditionnée par celle-là. Si comme c'est presque toujours le cas, la musique paraît exprimer quelque chose, ce n'est qu'une illusion et non pas réalité. C'est simplement un élément additionnel que, par convention tacite et invétérée, nous lui avons prêté, imposé comme une étiquette, un protocole, bref, une tenue et que, par accoutumance ou inconscience, nous sommes arrivés à confondre avec son essence" [Stravinsky, 1971]). Les Chinois ont une conception analogue de la musique : "La substance de la musique réside, pour les Chinois, dans le son. Substance-corps (*ti*) s'oppose à fonction (*yong*), comme nature (*xing*) à passion (*qing*). La musique n'a pas de fonction et n'exprime aucun sentiment. Elle équivaut à une résonance (*ganying*), réponse spontanée, mise en mouvement (*dong*) de l'air, des souffles. Elle est son." [Picard, 2003, p. 8]. Pour les Japonais, il semble que la musique soit comprise différemment puisque dans le théâtre de Nô, les chœurs, formés de chants à l'unisson, sont employés afin de décrire le paysage et l'époque dans lesquels s'instaurent la pièce, mais également les sentiments de l'acteur principal. Grâce à cette "description" chantée, les décors, éléments de représentations réalistes sont supprimés, seule la musique fonde l'univers abstrait dans lequel les acteurs évoluent²⁵ [Tamba, 1988].

Il existe donc des différences d'opinions notables concernant tant l'essence de la musique que ce qu'elle véhicule. Pour les uns, elle transmet directement les sentiments du créateur, pour les autres, la traduction de la musique en termes d'émotions est absurde, la musique ne peut être que la musique ! Nattiez de mentionner les intéressantes distinctions proposées par Meyer au sujet de ces divergences : "il y a d'une part les *absolutistes* qui considèrent que la signification musicale repose exclusivement dans les rapports entre les éléments constitutifs de l'oeuvre elle-même, et de l'autre, les *référentialistes* pour lesquels il ne peut y avoir de signification que par renvoi à un monde extramusical de concepts, d'actions, d'états émotionnels et de caractères. Mais cette première dichotomie se double d'une autre qui ne lui correspond pas : les *formalistes* qui, selon l'auteur, n'admettent pas que la musique puisse provoquer des réponses affectives - elle a une signification intrinsèque que lui donne le jeu de ses formes -, et les *expressionnistes* qui admettent l'existence des sentiments. Mais alors que les formalistes sont nécessairement absolutistes, les expressionnistes sont soit absolutistes - si, pour eux, l'expression est dans la musique elle-même -, soit référentialistes si l'expression s'explique par un renvoi au monde extérieur." [Nattiez, 1975, p. 132]. Selon ces dichotomies, nous situerions Stravinsky, pour qui la musique a une signification *intrinsèque*, du côté des "formalistes", et Tolstoï, qui croit en la signification *extrinsèque* de l'art en général, du côté des "expressionnistes". Ainsi, la position de Stravinsky, qui peut paraître provocante, semble être une réaction au point de vue référentialiste qui exclut toute signification *intrinsèque* à la musique et ne la considère qu'au travers des sentiments qu'elle suscite. Sans doute veut-il mettre là en exergue le côté noble de la musique, comme si son existence et sa signification portaient en eux quelque chose qui dépassait de loin la condition humaine. Comme le souligne Nattiez, Stravinski en fin de compte ne doit pas nier que la musique peut susciter des sentiments (ceci serait tout de même aberrant venant de la part du compositeur du *Sacre du printemps*), mais indique que si tel est le cas, ce ne peut-être que par des *associations* et non de manière *immanente*. Encore une fois, ressurgit la nécessaire distinction entre l'oeuvre prise telle quelle, au niveau neutre, et la perception que l'on en a, et qui peut se caractériser par la description d'états psychologiques.

²⁵L'émotion dans le théâtre de Nô n'est pas seulement suscitée par la musique, mais par sa conjonction avec les récitatifs, les danses, les costumes des acteurs, etc.

Faute de pouvoir aller plus loin dans le complexe et passionnant débat sur les conceptions intrinsèque ou au contraire extrinsèque du sens musical, qui dépasse le cadre de ce travail, nous donnerons, en guise de conclusion, notre position. Les dichotomies proposées par Meyer ont certes le mérite d'éclaircir les différents points de vue, mais elles sont aussi leurs limites, en ce sens qu'elles sous-tendent des catégories dont les frontières semblent être définies de manière exclusive. Il nous semble que sur certains aspects, ni les formalistes, ni les expressionnistes n'ont tort. On pourrait peut-être nous reprocher là un excès de diplomatie, mais cette position est confortée par le fait que la perception musicale n'est pas quelque chose d'immuable, sur le plan individuel d'une part, et sur le plan collectif, d'une autre, comme nous le verrons par la suite lorsque nous aborderons la question de l'écoute. La musique ne peut-elle s'écouter pour elle-même ? Si nous devons inéluctablement ressentir des états émotionnels de mélancolie, de joie, de tristesse, à chaque écoute de telle ou telle oeuvre, ne serait-ce pas fatigant ? Ne pas ressentir d'émotions spécifiques à l'écoute d'une musique ne nous paraît pas négatif en soi, cela ne voudrait pas dire que la musique est fade ou dénuée d'expression, mais que simplement nous *voulons* l'écouter comme telle. La musique est peut-être "le langage moins le sens" comme la décrit Lévi-Strauss, qui met bien là en évidence que la signification musicale ne peut se comparer à celle du langage, mais ceci ne l'empêche-t-elle pas de porter, de manière immanente, une signification plus abstraite qui, quelque part dans notre fort intérieur, peut - si on la *comprend* - faire sens ? De l'autre côté, il paraît si manifeste que la musique peut provoquer en nous des sentiments, être la source d'émotions (parfois allant jusqu'au rire ou aux larmes, cf. citation de Descartes) que soutenir le contraire paraît absurde ; qui n'a jamais connu la sensation de "chair de poule" à l'écoute d'une pièce musicale, chez soi, en concert, parfois même déclenchée par surprise, de manière si subite, par un simple petit fragment sonore de quelques secondes !

1.3 Écoute ou écoutes ?

Ainsi, la question du sens musical nous amène-t-elle naturellement à celle de l'écoute. Ces deux questions sont directement liées. Ne sommes-nous pas, auditeurs, au bout de la chaîne de communication exposée sur la figure 1.11 ? Ce que l'un recherche lors de l'écoute musicale, l'autre le recherche-t-il également ? L'écoute n'est-elle pas conditionnée par nos connaissances, notre culture, l'entraînement de notre oreille à déceler les structures des phénomènes musicaux ? Afin d'amener quelques éléments de réponses à ces questions, analysons tout d'abord les propos de certains compositeurs, musicologues, et psychologues.

1.3.1 Qu'écoutent les spécialistes ?

C'est en ces termes que John Cage, lors d'un entretien réalisé dans le cadre du documentaire *Ecoute* [Sebestik, 2005], s'exprime à propos de l'écoute : "Quand j'entends ce qu'on appelle la musique, il me semble que quelqu'un parle et parle de ses sentiments ou de sa conception des relations, mais quand j'entends le bruit de la circulation, ici sur la 6^e avenue par exemple, je n'ai pas l'impression que quelqu'un parle, mais plutôt que le son agit. Et j'aime l'activité du son. Il devient plus fort et plus doux, plus élevé et plus bas, plus long et plus court, et tout cela me satisfait pleinement. Je n'ai pas besoin que le son me parle. [...] Les gens supposent qu'écouter, c'est plus qu'écouter. Et parfois, ils parlent d'écoute intérieure ou de signification du son. Quand je parle de musique, les gens finissent par comprendre que je parle de son qui ne signifie rien, qui n'est pas intérieur, mais seulement extérieur. Ceux qui comprennent cela disent : "vous voulez dire que ce ne sont que des sons ?", pensant que ce qui n'est qu'un son est inutile. Moi, j'aime les sons pour eux-mêmes, je n'ai pas besoin qu'ils soient autre chose, je ne veux pas qu'ils soient psychologiques, je ne veux pas qu'un son prétende qu'il est un baquet, qu'il est président, ou qu'il est

amoureux d'un autre son (NDLR : rires de Cage !). Je veux seulement que ce soit un son. Et je ne suis pas si bête. Un philosophe allemand très célèbre, Emmanuel Kant, a dit qu'il y avait deux choses qui n'ont rien à signifier : l'une est la musique, l'autre, le rire. *Elles ne doivent rien signifier pour nous donner un plaisir profond*. L'expérience sonore que je préfère à toute est l'expérience du silence. Et le silence, presque partout dans le monde aujourd'hui, c'est la circulation. Si vous écoutez Beethoven ou Mozart, vous voyez qu'ils sont toujours pareils. Mais écoutez la circulation, elle est toujours différente" [Sebestik, 2005]. Nous retrouvons chez Cage des éléments du point de vue formaliste de Stravinsky. Tous deux considèrent la musique, ou les sons d'une manière générale pour Cage, pour eux-mêmes, pour les *formes* qu'ils peuvent prendre au cours du temps. La signification que ces sons peuvent véhiculer n'est pas au cœur de son écoute. Nombreux seraient ceux qui, à l'écoute de bruits de circulation, diraient "j'entends une voiture, des bruits de moteurs diesel, un klaxon", etc. Cage considère les objets sonores à un niveau plus abstrait en tâchant de les dégager de toute référence à son intérieur. Ceci n'est pas sans être lié avec ce qu'il recherche avant tout dans l'écoute musicale et qui lui procure la plus grande satisfaction : la surprise, la nouveauté. Ce type d'écoute coïncide avec ce que Schaeffer appelle *l'écoute acousmatique*, i.e. une écoute qui se concentre sur les effets induits par les sons et non leurs causes.

Les propos du cinéaste et compositeur Michel Fano présentent des analogies avec ceux de Cage : "Beaucoup de gens quand ils écoutent de la musique se réfèrent à des événements de leur vie affective, de leur vie intérieure qui finalement n'ont pas grand rapport avec l'essence même de la musique, qui n'est pas faite pour raconter autre chose qu'elle même (NDLR : encore une fois le point de vue formaliste). Donc une écoute musicale sérieuse, vraie, passe d'abord par l'oubli de soi et par *l'attention vers l'autre*, c'est à dire vers le texte musical dans une relation qui n'est pas sans rappeler pour moi la relation amoureuse par exemple." (ibidem). Si l'on exagère un peu le discours de Fano, il existerait donc deux types d'écoute musicale, une écoute "égoïste", centrée vers soi-même, et une écoute plus altruiste, neutre, vraie, totalement dévouée à la musique qui selon ses propos devient presque personnifiée.

L'écoute présente encore d'autres facettes, et peut même poser problème comme l'avoue le musicologue Célestin Deliège : "[...] toute écoute m'insatisfait. Je parlais de cette écoute dans la salle de concert, mais je suis encore moins satisfait par une écoute de disque. Pourquoi ? Parce que par le disque l'écoute est naturellement complètement figée. Vous avez un disque, vous êtes heureux, un peu déçus par endroits de ce que fait l'interprète et puis vous réécoutez le disque, réentendez toujours la même chose, vous n'avez plus de surprises, c'est toujours pareil, c'est pour ça que personnellement, j'ai très peu de disques. La partition pour moi, c'est *l'écoute silencieuse*, c'est l'écoute qui m'apporte le plus [...]. Le privilège de cette écoute est d'en ralentir le tempo, combien de fois je n'estime pas, quand je suis dans une salle de concert, que je n'ai pas perçu exactement ce que j'aurais voulu entendre parce que naturellement le contenu de la musique est extrêmement fugitif [...]. Les choses vont parfois très vite, on n'entend pas toutes les notes, on entend un ensemble, une structure, mais pas la note, et bien moi, j'ai besoin de la partition pour voir la note. Je sais que Strauss ne le voulait pas, mais il les a écrites et je veux les voir." (ibidem). Les remarques de Deliège amènent une question nouvelle : l'écoute a-t-elle besoin du sonore ? Son expérience montre que celui qui maîtrise suffisamment la lecture du langage musical peut très bien se passer du sonore. Bien entendu, il ne peut y avoir équivalence entre l'écoute réelle d'un phénomène sonore, et ce qu'il intitule "*l'écoute silencieuse*". Mais sur le plan théorique, une écoute silencieuse peut très bien être envisagée à un niveau cognitif. En effet, rappelons brièvement le phénomène de réception du son.

Notre oreille est sensible à l'énergie acoustique qui lui parvient grâce au tympan. Cette énergie acoustique change plusieurs fois de nature lors de son parcours au sein du système auditif. Lors de son trajet de l'oreille externe vers la cochlée, l'énergie acoustique est convertie en énergie

mécanique lors de la transmission par les osselets, puis en énergie électrique après traitement par la membrane basilaire. Les signaux codés par la membrane basilaire sous forme d'influx nerveux stimulent ensuite des zones du cerveau situées dans le cortex auditif. Le circuit de la perception auditive n'est pas constitué que d'un simple "trajet aller" vers le cerveau, ce dernier envoyant également des informations à l'oreille (ce qui nous sert de défense par exemple lorsque l'on est surpris par un bruit intense qui pourrait dépasser le seuil de douleur).

Une des finalités des vibrations sonores qui nous parviennent est la stimulation du cortex auditif. Ne peut-on pas après tout exciter ces zones à partir d'autres modalités sensorielles, en l'occurrence, la vision, lors de la lecture d'une partition musicale ? En d'autres termes, la notion d'écoute "silencieuse" est-elle vraiment sensée ? Des études de neurosciences ont montré que les informations musicales visuelles semblent stimuler le système auditif sitôt après 100 ms de traitement [Schön et Besson, 2005]. Lecture musicale et perception musicale ne sont donc pas deux phénomènes antinomiques. Il faut reconnaître que la sensation de Deliège est parfaitement fondée et que le phénomène d'écoute peut se voir décorrélié du sonore !

Pour le psychologue Stephen Mc Adams, l'écoute musicale est implicitement liée aux connaissances que l'on a acquise depuis l'enfance, et donc par la même occasion à notre culture : "Un enfant est exposé à la musique dans sa culture et réussit à comprendre la nature de cette musique [...]. Donc il y a une sorte d'acquisition complètement automatique par la simple exposition à un environnement sonore qui est structuré et cette structure est très complexe [...]. La musique est structurelle. En comprenant au fur et à mesure la structure, cela donne lieu à des attentes qui vont réaliser la partie de la structure qui n'est pas encore jouée. Ça veut dire qu'on a des connaissances, ce qu'on appelle en psychologie des *connaissances abstraites* sur les systèmes musicaux, sur les formes qu'on rencontre habituellement dans le cas de la musique, et surtout de la musique tonale. Il y a un certain nombre de formes qui sont utilisées à plusieurs reprises à travers le répertoire de la musique et, on reconnaît ces choses-là.". Puis, au sujet de la musique contemporaine : "On rompt complètement l'évolution, on décide de façon intellectuelle qu'on va changer le système et il y a une sorte de divorce entre les grammaires de l'écoute qu'on a acquises à travers l'évolution de la culture et la grammaire de la composition, donc il y a une sorte de cassure entre ces deux grammaires, et on peut se poser la question : est-ce que le cerveau humain est capable de prendre le relais à travers une génération ou deux, etc., de commencer à comprendre, ou est-ce que les compositeurs, certains compositeurs de notre temps, ont vraiment fait une sorte de cassure avec les possibilités du cerveau, donc les possibilités biologiques. C'est une autre façon de poser la question : est-ce qu'il y a des limites ?" (ibidem). Il est vrai qu'avec la musique contemporaine, les structures se complexifient (mais ce qui est complexe, n'est-il pas juste ce qui n'est pas encore compris, limpide ?) : les systèmes de hauteurs utilisés dans le système tonal et fondés comme nous avons pu le voir à partir de *constatations empiriques découlant de notre perception* (la consonance et la dissonance des intervalles régissent leurs fonctions harmoniques) se voient remplacés par des *systèmes conçus a priori*, sur le papier, d'après une démarche essentiellement intellectuelle (cf. sérialisme de Schönberg par exemple). La partition d'*Artikulation* de Ligeti figure 1.1 n'a-t-elle pas l'air beaucoup plus abstraite, beaucoup plus mystérieuse que celle de la chanson populaire située juste au-dessus ? Mais peut-être est-ce l'auteur, qui en disant cela biaise l'observation, puisqu'il est guidé par ses connaissances propres. Peut-être, comme le sous-entend Mc Adams, nous faudra-t-il du temps pour apprivoiser l'inconnu. La réponse à cette question, seule le futur l'amènera.

1.3.2 Écouter, Ouïr, Entendre, Comprendre

S'il est une chose qui découle de l'analyse du discours de ces différents spécialistes, compositeurs (Cage, Fano), musicologue (Deliège), et psychologue (Mc Adams), c'est bien que *l'écoute*

musicale est multiple. Tel y recherche la surprise, tel autre va même jusqu'à l'aborder comme une relation amoureuse, et tel autre se trouve complètement insatisfait de l'écoute sonore et y préférera une écoute "silencieuse". Le travail considérable de Pierre Schaeffer dont nous n'avons encore que peu parlé jusqu'à présent amène quelques éclaircissements sur la question de l'écoute. Comme nous pouvons le voir sur le tableau présenté figure 1.12 calquée sur celle du *Traité des objets musicaux* [Schaeffer, 1966, p. 113], Schaeffer distingue quatre *fonctions* de l'écoute : écouter, ouïr, entendre, comprendre.

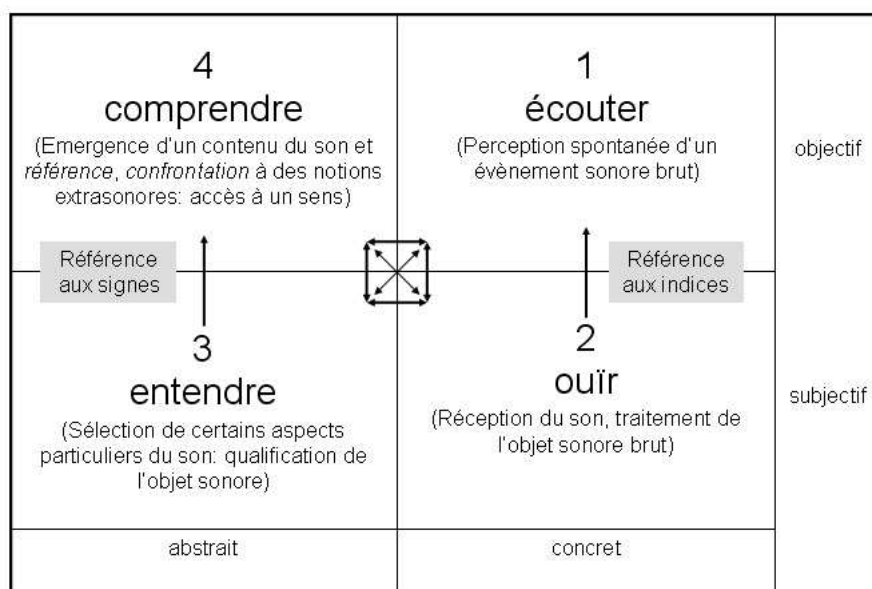


FIG. 1.12 – Les quatre fonctions de l'écoute selon Schaeffer (d'après [Schaeffer, 1966, p. 113]). Nous avons symbolisé les possibles échanges fonctionnels entre les différents secteurs par les flèches au centre du tableau.

Les significations de ces termes sont les suivantes :

- “1. Écouter, c'est prêter l'oreille, s'intéresser à. Je me dirige activement vers quelqu'un ou vers quelque chose qui m'est décrit ou signalé par un son.
2. Ouïr, c'est percevoir par l'oreille. Par opposition à écouter qui correspond à l'attitude la plus active, ce que j'ouïs, c'est ce qui m'est donné dans la perception.
3. D'entendre, nous retiendrons le sens étymologique : “avoir une intention”. Ce que j'entends, ce qui m'est manifeste, est fonction de cette intention.
4. Comprendre, prendre avec soi, et dans une double relation avec écouter et entendre. Je comprends ce que je visais dans mon écoute, grâce à ce que j'ai choisi d'entendre. Mais, réciproquement, ce que j'ai déjà compris dirige mon écoute, informe ce que j'entends.” (ibidem, p. 104).

Les quatre fonctions de l'écoute selon Schaeffer Le tableau n'a pas pour objet de proposer une quelconque chronologie des processus mis en jeu par notre système perceptif lors de l'écoute, mais de dégager différentes fonctions de l'oreille qui n'ont pas toutes les mêmes finalités. Schaeffer le rappelle, la perception auditive, jusqu'à un certain degré, reste un phénomène instantané ; lorsque quelqu'un frappe à la porte, notre réaction est immédiate, nous tournons la tête vers cette dernière, source des vibrations sonores qui nous parviennent. Les distinctions entre ces différentes fonctions de l'oreille sont subtiles et pas forcément évidentes à saisir, aussi partons-nous du langage pour les éclaircir. Si *écouter* implique une intention active de l'auditeur (“prêter l'oreille”), *ouïr* paraît révéler quelque chose de plus immanent, propre à la fonction biologique de l'audition.

La différence est saillante dans cette citation de Pacuvius décrivant son attitude réservée à l'égard des astrologues : "Il vaut mieux les ouïr que les écouter" (ibidem, p. 103). Écouter un événement sonore correspond sans doute au rôle le plus primitif de la perception, signaler quelque chose (le signal d'alarme qui avertit d'un danger par exemple). Le son est alors traité comme un indice permettant d'identifier l'événement à son contexte causal. Ouïr en revanche n'a trait non pas à l'identification de la source sonore, mais à "l'objet sonore brut". La distinction entre écouter et ouïr, pris dans le sens d'entendre, est illustrée par ces propos : "à chaque répétition d'un son enregistré, j'écoute le même objet : bien que je ne l'entende jamais pareillement, que d'inconnu il devienne familier, que j'en perçoive successivement divers aspects, qu'il ne soit donc jamais pareil, je l'identifie toujours comme cet objet-ci bien déterminé" (ibidem, p. 115). L'écoute de l'objet sonore brut (en 2.) s'organise donc en indices (en 1.), tous deux référant à un donné concret. La fonction en 3. révèle une sélection subjective de l'auditeur face à tel ou tel aspect particulier du son, il s'agit d'une qualification abstraite de l'objet sonore. Cette fonction met en avant le fait que de par notre faculté de pouvoir sélectionner dans l'écoute musicale ce que nous souhaitons, nous n'écoutons pas tous de la même manière (par ex. l'écoute de celui qui se concentrera sur le phrasé du guitariste dans tel trio jazz ne correspondra pas à l'écoute de celui qui sera subjugué par les roulements du batteur). Enfin, les données récoltées en 3. pourront être traitées comme des *signes* et prendront éventuellement un *sens* (par ex. "je comprends ce que tu me dis", dans le cas du sens véhiculé par la parole). Concernant la notion de sens dont il est question dans la fonction 4. *comprendre* de l'oreille, Schaeffer souligne bien la distinction entre langage et musique, et parle alors dans le cadre d'une écoute musicale, de "valeurs musicales donnant accès à un sens musical".

Écoutes naturelle/culturelle versus écoutes banale/spécialisée L'auteur propose également deux dichotomies caractérisant les tendances de l'écoute. D'une part, il distingue l'écoute *naturelle* de l'écoute *culturelle* : "Par écoute *naturelle*, nous voulons décrire la tendance prioritaire et primitive à se servir du son pour renseigner sur l'événement" (ibidem, p. 120). Il s'agit d'une écoute commune à tous les hommes quelle que soit leur civilisation. Les fonctions utilisées dans ce type d'écoute sont celles de la partie droite du tableau de la figure 1.12. A l'opposé, lors d'une écoute *culturelle*, "on se détourne ainsi délibérément (sans cesser de l'entendre) de l'événement sonore et des circonstances qu'il révèle relativement à son émission, pour s'attacher au message, à la signification, aux *valeurs* dont le son est porteur" (ibidem, p. 121). On conçoit bien que cette écoute peut varier d'une collectivité à l'autre, et même d'un individu à l'autre puisqu'elle est fonction d'une culture, de connaissances. A l'inverse, ce type d'écoute est donc davantage orientée vers le secteur 4 du tableau.

L'autre couple amené par Schaeffer oppose l'écoute *banale* de l'écoute *spécialisée*, ou pratique. Il s'agit là de "marquer la différence de compétence dans l'écoute, de qualité dans l'attention, et aussi la confusion des intentions de l'écoute banale, alors que l'écoute spécialisée choisit délibérément, dans la masse des choses à écouter, ce qu'elle veut entendre et élucider" (ibidem, p. 121). L'écoute banale, représentée dans le haut du tableau (secteurs 1. et 4.) reste superficielle ("l'auditeur n'a pas de curiosité ni de référence particulières", ibidem. p. 123), elle ne cherche pas à tirer des indices, analyser, trouver des qualifications au sonore, ce qui en revanche, est le propre de l'écoute spécialisée (secteurs 2 et 3). On pourra remarquer que souvent, en situation de concert, les personnes non musiciennes, n'ayant ou ne croyant pas avoir suffisamment "d'oreille", se tournent vers leurs complices musiciens afin de connaître leur jugement de valeur sur le jeu de tel ou tel instrumentiste. Mais l'écoute spécialisée ne devient pas pour autant la plus objective. Les quelques exemples donnés par Schaeffer le mette bien en avant : "Ainsi le phonéticien oublie le sens des mots pour n'entendre que leurs éléments phonétiques : le médecin ne se sert du "33, 33..." que pour en déduire l'état des poumons de son patient" (ibidem).

Bilan Au vu de ces descriptions et de celles de la section précédente, plusieurs conclusions se dégagent. L'écoute n'est pas quelque chose de figée, elle dépend fortement d'un contexte. Ce contexte est entre autres déterminé par notre culture (D'où vient-on ? Quelles sont nos connaissances abstraites, quelles sont celles acquises consciemment ?). Rappelons à ce propos la sensibilité de l'oreille japonaise pour le timbre, révélée par Tokumaru et dont nous avons parlé au paragraphe 1.1.4. Mais le contexte est également déterminé par notre spécialisation (par ex. si je suis acousticien, je remarquerai peut-être le travail de timbre d'un compositeur, ayant d'une part déjà connaissance de l'existence de la notion de timbre et d'autre part ayant éduqué mon oreille à cet effet, aspects qui ne viennent pas de manière innée), et notre intention dans l'écoute (par ex. je peux écouter une musique comme elle est, ou bien rechercher la façon dont tel compositeur utilise le contrepoint, je peux écouter les bruits de la nature pour eux-mêmes ou bien chercher à analyser quel motif mélodique utilise cet oiseau), le dernier point se résumant bien en ces termes : je suis libre de mon écoute (à partir du moment où j'ai connaissance de cette liberté !). Selon la nature de ce contexte, les quatre fonctions schaefferiennes de l'écoute ne sont pas utilisées dans les mêmes proportions. Semblent maintenant émerger de possibles explications sur les sources de divergences d'opinions sur le sens musical. Une personne qui attachera beaucoup d'importance aux valeurs, et au sens aura pour finalité de comprendre le musical (secteur 4). Cela peut certes requérir du temps, du moins plus que n'en offre le "temps-réel", la musique est fugitive comme le dit Deliège. Ce dernier se voit donc insatisfait de la situation de concert où quelque part tout va trop vite pour une analyse musicale approfondie et préfère à ce titre voyager dans la partition musicale à sa guise afin d'y déceler et de comprendre les structures et procédés compositionnels sous-jacents. Cage semble moins user de la fonction comprendre que de celle d'entendre, non pas qu'il ne puisse pas comprendre le sens véhiculé par les sons de circulation dont il nous parle, mais peut-être parce qu'il choisit et préfère limiter son écoute à l'entendre (secteur 3), dans le but de se laisser surprendre, de garder une naïveté par rapport au son, de le recevoir de manière brute (ouïr).

1.4 Du compositeur à l'auditeur

À la fin de la section 1.1.4, nous avons évoqué la relation instrumentale et l'importance du rôle de l'interprète lors de l'exécution de musiques écrites. La situation d'interprétation musicale qui n'apparaît pas de manière explicite dans le découpage proposé par Molino et Nattiez (cf. figure 1.11) semble pourtant se joindre à l'activité compositionnelle, tout du moins en partie, dans la constitution du niveau poétique (création). En effet, comme l'indique Cadoz, "dans le contexte traditionnel, toute oeuvre fait l'objet, à la fin du processus, d'une ultime intervention créative, celle du ou des interprètes, qui se saisissent d'un matériau particulier : la combinaison partition-instrument(s), pour donner une "exécution". Cette exécution n'est pas une traduction, la partition ne détermine pas tout, et l'interprète s'implique dans une relation qui est bien une relation de création dans la mesure où il apporte sa propre contribution. Le système interprète-partition-instrument est plus que la somme de ses parties." [Cadoz, 1991, p. 44].

1.4.1 Définition de l'interprétation musicale

Examinons la définition proposée par Scholes dans son encyclopédie de la musique *The Oxford Companion To Music* : "Interpretation in music is simply the act of performance with the implication that in this act the performer's judgement and personality necessarily have their share. Just as there is no means by which a dramatist can so write his play as to dictate to the actors the precise way in which his lines are to be spoken, so there is no means by which a composer can so notate his compositions as to dictate to the performer the precise way in which they are to be sung or played. Thus, no two performers adopt exactly similar treatments of any given composition, and the difference of speed, intensity, etc., produce a difference of effect similar to the slight difference of emphasis and shade of meaning in the declamation of a speech by two actors. There have

been even cases where a composer has admitted that a performer's interpretation was not merely different from but better than the one he himself had in mind when composing." [Scholes, 1960] (traduction : L'interprétation en musique est simplement l'acte d'exécution avec l'implication que le jugement et la personnalité de l'exécutant y prennent nécessairement part. Au même titre qu'il n'existe pas de moyens par lesquels un dramaturge puisse écrire sa pièce en dictant aux acteurs de manière précise la façon dont ses lignes doivent être prononcées, il n'existe pas de moyens par lesquels un compositeur puisse noter ses compositions en indiquant à l'interprète la façon précise dont elles doivent être chantées ou jouées. Ainsi, deux interprètes n'adopteront jamais exactement les mêmes traitements pour une composition donnée, et la différence de rapidité, intensité, etc., produisent une différence d'effet similaire à la légère différence d'accent et de nuance de signification dans la déclamation d'un texte par deux acteurs. Il y a déjà eu des situations où un compositeur a admis que l'interprétation d'un musicien n'était pas simplement différente mais meilleure que celle qu'il avait en tête en composant.). On sent bien de par cette description que compositeur et interprète sont deux acteurs complices dans l'activité de production du musical. Ceci justifie amplement la prise en compte de l'interprète au niveau poétique d'une sémiologie musicale. Le langage met d'ailleurs bien en exergue l'existence d'une différence entre interprétation et exécution, terme qui tend à assigner au musicien le simple rôle de reproduction mécanique de l'œuvre écrite. Leurs équivalents anglais *interprétation* et *performance* sont souvent confondus. La majeure partie de la littérature anglophone sur le sujet de l'interprétation musicale (cf. par exemple, dans les domaines de la psychologie de la musique et de l'acoustique musicale, l'article très enrichissant de Gabrielsson [Gabrielsson, 1999]) emploie le terme *performance*, mais la confusion des termes est légitime. En effet, on ne peut guère concevoir d'exécution qui ne soit interprétée pour la simple et bonne raison que le musicien n'est pas un robot²⁶. Pour Siron, interpréter, c'est "rendre vivante²⁷ une musique fixée, notée sur une partition ou mémorisée" [Siron, 2004]. Il y a bien là l'idée de donner vie à la forme écrite en l'actualisant dans le présent. Dans certaines musiques, les choix issus du jeu d'un interprète viennent clairement enrichir la composition originelle : "Les musicologues chinois complètent la notation traditionnelle par la transcription d'un maître donné, dont l'identité est indiquée ; ils combinent sur la même page tablature des doigtés et portée, incluant les indications de durées, de tempos et d'intensités. L'enseignement oral fournissait ces données dans le cadre esthétique et social de l'école dont faisait partie le maître." [Picard, 2003]. Le même phénomène se retrouve dans le jazz où l'on trouve de manière récurrente plusieurs versions d'un même thème pour un standard, qui sont des transcriptions issues des différentes interprétations données par tel ou tel joueur émérite. Ceci est d'ailleurs fortement lié au fait que le jazz soit un genre musical où l'improvisation tient une place très importante.

L'interprète a donc un double rôle, celui de décoder le signal notationnel du compositeur et celui de l'enrichir selon son propre jugement. En effet, la partition n'indique pas tout. D'une part, les notations tendent à discrétiser des paramètres qui dans la relation instrumentale peuvent pour certains instruments varier de manière continue (e.g la hauteur des notes sur une portée). D'autre part, des codes d'interprétation liés à une tradition, un style de musique particulier, sont parfois implicites. Par exemple, le *swing* est devenu, à partir de conventions culturelles, une caractéristique intrinsèque du jazz. Pour "swinguer", des groupes de notes notées "binaires" devront être pensées et jouées "ternaires" (par ex. deux croches seront jouées comme un "triolet noire-croche"). Si l'information n'est pas marquée telle quelle sur une partition, c'est que d'une part elle alourdirait la notation, et que d'autre part, elle n'est peut-être tout simplement pas traduisible parfaitement. Comme le souligne Duke Ellington dans sa composition *It Don't Mean A Thing (If It Ain't Got That Swing)* ("Ça ne sert à rien s'il n'y a pas le swing"), le non respect de ce code stylistique implicite corrompt la qualité de l'interprétation musicale, et peut même aller jusqu'à faire

²⁶De nos jours, même les robots musiciens deviennent de vrais interprètes (cf. le robot trompettiste de Toyota [Toyota, 2008]) !

²⁷C'est nous qui soulignons.

perdre à la musique le sens recherché par le compositeur. En effet, le fait qu'un musicien *interprète* une oeuvre écrite n'est pas nécessairement gage de musicalité. Tout débutant qui interprète ses premières oeuvres d'études laissera échapper une fausse note au détour d'une phrase, omettra de marquer tel ou tel accent, jouera de manière trop "mécanique" et "scolaire", et ne fera pas vivre la musique à sa juste valeur. Selon l'art, le talent et l'expérience du musicien, une interprétation pourra donc être "expressive" ou "inexpressive", "vivante" ou "mécanique", "fidèle" à l'oeuvre, ou "incongrue", "originale" ou "banale", etc. L'impression que laissera une interprétation à l'auditeur relève de l'esthétique et est par là fonction des cultures, des inclinations, des psychologies (etc.) individuelles. N'omettons pas non plus que pour certains, la musique peut très bien se dispenser de la forme sonore pour faire sens. Comme nous l'avons vu au paragraphe 1.3.1, à un niveau cognitif, la lecture d'une partition peut pour les musiciens expérimentés donner naissance à une écoute, voire une interprétation "intérieure". Lors d'un tel processus, interprète et auditeur ne sont alors qu'une seule et même personne.

La définition de Scholes indique également qu'**au travers d'une interprétation, le musicien développe une personnalité musicale**. C'est dans ses choix d'utilisation des paramètres musicaux (dont seuls le tempo et la dynamique sont cités explicitement dans la définition de Scholes) qu'il se distinguera d'un autre interprète. La sonorité du musicien paraît également déterminante eu égard à son identité musicale. Voici pour exemple comment deux clarinettes de traditions différentes, Jack Brymer, issu du classique, et Benny Goodman, issu du jazz, interprètent le *Concerto pour Clarinette et Orchestre* (KV. 622) de Mozart (cf. respectivement, les [Exemples Sonores 12](#) et [13](#)). Les timbres des deux instrumentistes sont complètement différents ; le son velouté et doux de Brymer s'oppose à celui plus brillant, voire nasillard de Goodman. Il en résulte un profond changement dans la perception de l'interprétation musicale.

1.4.2 Chaîne de communication des musiques écrites

Analysons désormais la chaîne de communication musicale en tenant compte du rôle de l'interprète. Un schéma général de communication musicale, appliqué au cas des musiques écrites, est proposé figure 1.4.2. Il n'a pas la prétention de représenter de manière exhaustive une telle chaîne qui est très complexe, mais tâche de donner une vision synthétique rappelant les différents pôles de la sémiologie musicale (poiétique, neutre, esthétique), et les différents niveaux d'analyses accessibles au chercheur. Considérer une telle chaîne de communication est un sujet interdisciplinaire. Les principales disciplines concernées sont indiquées sur le schéma.

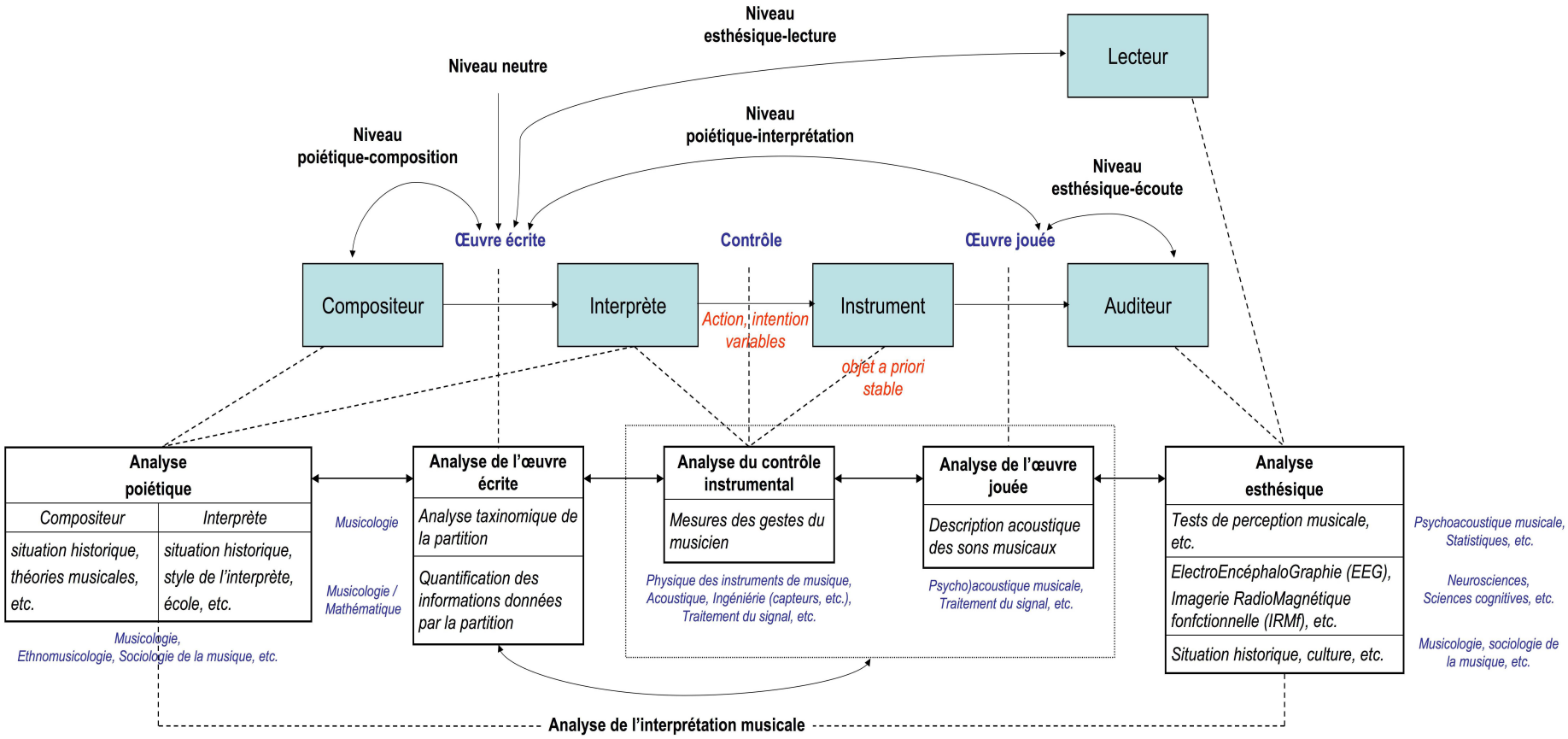


FIG. 1.13 – Chaîne de communication musicale dans le cas des musiques écrites.

Rappelons que le domaine de la poïétique s'attache à caractériser ce qui relève de la conception de l'œuvre musicale (notamment "les conditions philosophiques, sociologiques, psychologiques, historiques, esthétiques, matérielles qui motivent ou conditionnent le créateur jusqu'à ce que l'œuvre soit considérée comme achevée" [Nattiez, 1975, p. 52]). Dans la tripartition proposée par Nattiez, l'œuvre écrite et l'œuvre jouée s'inscrivent toutes deux au niveau neutre, ce dernier se voulant indépendant de toute logique de production ou perceptuelle. Notre conception n'est pas exactement la même puisque nous considérons que l'œuvre jouée (interprétée) contient également des éléments liés au processus de création. Aussi, selon notre conception, seule l'œuvre écrite s'inscrit dans le niveau neutre. L'œuvre jouée fait également partie du niveau poïétique. Ceci est justifié pour les raisons exposées ci-dessus, par le fait que l'interprétation musicale consiste à traduire et à compléter le travail du compositeur. Nous distinguerons donc deux niveaux de poïétiques : la poïétique-compositionnelle et la poïétique-interprétationnelle. **L'analyse de l'interprétation musicale prend alors forme dans la confrontation des analyses de l'œuvre écrite (niveau neutre), et des analyses du jeu de l'interprète, accessibles d'une part au travers de mesures du contrôle instrumental, et d'autre part au travers de descriptions sonologiques de l'œuvre jouée.** Une analyse esthétique approfondie se devra prendre en compte l'ensemble de la chaîne de traitement du sonore, dans ses allers-venues depuis l'oreille jusqu'au cerveau, mêlant des disciplines telles que la musicologie, la psychologie de la musique, la psychoacoustique musicale, les neurosciences, les sciences cognitives, etc.

Face à l'étendue et à la complexité d'une telle chaîne de communication, on conçoit bien qu'il nous a fallu faire certains choix méthodologiques dans le but d'étudier le rôle du timbre dans l'interprétation musicale. Notre démarche qui relève essentiellement de l'acoustique et de la psychoacoustique (ou perception) musicale, sera détaillée dans la section 1.5.

1.4.3 Du rôle du timbre dans l'interprétation musicale

Si les paramètres du jeu instrumental liés au temps et à la dynamique (intensité) ont fait l'objet de nombre d'études en psychologie de la musique et analyse de l'interprétation musicale (cf. par exemple [Sundberg *et al.*, 1989; Repp, 1992; Palmer, 1997; Widmer et Goebel, 2004; De Poli, 2006]), le rôle du timbre, objet d'étude du présent travail, n'a été que très peu abordé. Ceci n'est pas sans être lié avec le fait que le timbre est une notion complexe qui ne possède toujours pas de définition rigoureuse, comme nous le verrons au chapitre 2, et reste par conséquent difficilement mesurable. Il n'en est pas moins l'une des variables essentielles des sons musicaux, du moins c'est ce que nous tâcherons de démontrer. Déjà en 1938, Carl E. Seashore, l'un des pionniers des recherches sur l'interprétation musicale avec ses collaborateurs de l'Université de l'Iowa, s'exprimait ainsi à propos du timbre : "timbre as a fourth attribute of tone is by far the most important aspect of tone and introduces the largest number of problems and variables" [Seashore, 1938].

Les travaux initiés par Johan Sundberg au laboratoire d'acoustique musicale du KTH (Royal Institute of Technology, Suède) couplés à ceux de ses collaborateurs Lars Frydén, Anders Friberg, et Roberto Bresin, ont conduits à l'élaboration d'un programme de modélisation de l'interprétation musicale pour les musiques instrumentales reposant sur un système de règles établies à partir de la théorie musicale et ajustées selon une démarche d'analyse par la synthèse. Ce programme baptisé *Director Musices*²⁸, applique une palette d'une douzaine de règles à un fichier d'entrée MIDI (Musical Instrument Digital Interface) de manière à imiter le jeu expressif d'un vrai interprète. L'ajout de l'expression musicale²⁹ se fait essentiellement au travers de paramètres temporels (par ex. durées des notes, articulation), de dynamique, et de hauteurs [Mathews *et al.*, 2003]. Nous avons appliqué le programme à un fichier MIDI correspondant au début de l'Allemande de la

²⁸téléchargeable à l'adresse indiquée en référence [Sundberg *et al.*, 2008].

²⁹"art de rendre vivantes les émotions et sentiments "contenus" dans la musique." [Siron, 2004]

deuxième *Suite de Bach pour Violoncelle* (BWV. 1008)³⁰. L'**Exemple Sonore n° 14** correspond à la version de l'extrait rigoureusement fidèle aux indications de la partition, tandis que l'**Exemple Sonore n° 15** correspond à la version obtenue après application des règles de Director Musices (réglages par défaut). Les deux exemples ont été générés par un échantillonneur employant des sons de violoncelle. On peut constater que le processus d'interprétation automatique ajoute dans ce cas à la partition initiale principalement des variations rythmiques, et use d'un principe très souvent employé en musique, le ralentissement (*ritardando*) final pour marquer la fin de la phrase musicale. À titre de comparaison, l'**Exemple Sonore n° 16** correspond à l'interprétation de cet extrait musical de Bach par le violoncelliste Mischa Maisky. Ces exemples suggèrent qu'afin de parvenir à une synthèse imitative de l'interprétation musicale, il semble également primordial de prendre en compte - pour certains instruments - les *variations de timbre*³¹ produites par l'interprète.

Facture instrumentale et geste du musicien Comme nous l'avons déjà évoqué à la fin du paragraphe 1.1.4, **deux facteurs importants agissent sur le timbre dans la relation instrumentale : la facture instrumentale, et le geste du musicien**. Nous soulignerons, à titre d'exemple et en référence à la facture instrumentale du shamisen japonais, que le plectre utilisé par le musicien pour gratter la corde, et qui peut paraître accessoire pour le néophyte, revêt en réalité une grande importance quant à la sonorité de l'instrument : "Pour faire, de la main droite, sonner leur instrument, les exécutants de Chine et des Ryukyu (NDLR : archipel d'îles japonaises) utilisaient des ongles d'ivoire ou d'os. Les Japonais préfèrent un plectre qui a la forme d'une feuille de *gingko biloba*, c'est-à-dire celle d'un éventail élargi et galbé, que l'on tient par un manche épais, mais aminci pour faciliter une bonne prise de la main droite." [Landy, 1996, p. 122]. Après une description du plectre du shamisen, l'auteur réfère au contrôle de l'instrument : "L'élément capital du jeu reste l'angle d'attaque selon lequel le plectre accroche ou frappe les cordes." (ibidem). Nous pouvons voir le plectre du shamisen sur la photo de la figure 1.14, qui présente un joueur de shamisen en pleine concentration, et permet de saisir avec quelle consécration cet interprète aborde le geste instrumental.

Au même titre que c'est par le geste que le pongiste pourra vaincre son adversaire, c'est le geste qui permettra au musicien de parvenir à ses fins, de produire un son qui conquerra son auditeur. Pour les instruments à cordes pincés tels la guitare, vient s'ajouter à la nature du plectre utilisé et à l'angle d'attaque de la corde, un paramètre capital pour le timbre : la position du point de pincement de la corde [Traube *et al.*, 2003]. Les **Exemples Sonores 17 et 18** sont deux courts extraits joués à la guitare acoustique pour lesquels le musicien agit intentionnellement sur les variations de timbre entre des notes ou groupes de notes successives³², par action sur la position du point de pincement. On entend notamment dans l'**Exemple Sonore n° 17** une même mélodie jouée à quatre positions différentes sur le manche. La sonorité passe alors du "métallique" lorsque les attaques sont proches du chevalet au "chaud" et "rond", lorsque ces dernières sont effectuées à un niveau proche de la moitié de la corde.

À l'opposé de la guitare, les instruments auto-entretenus tels que la clarinette permettent à l'instrumentiste de produire des variations de timbre durant la production même d'une note. C'est avec virtuosité que le clarinettiste contemporain Ned Rothenberg en dispose dans sa pièce *Nail on the Head* (cf. **Exemple Sonore n° 19**). Un même motif mélodique répété à une cadence célèbre se voit transformer au travers de micro-modulations de timbre et de hauteur. On notera le changement

³⁰Cet extrait est utilisé pour analyser la nature des variations de timbre dans l'interprétation musicale au chapitre 5.

³¹Nous définirons avec plus de précisions la notion de *variations de timbre* au chapitre 2. Notons dès à présent que nous l'employons ici afin de référer aux variations de la qualité sonore au cours du temps.

³²et non durant leur production, ce qui n'est pas réalisable à la guitare, sans effets additionnels.



FIG. 1.14 – Joueur de shamisen (photo issue de [Landy, 1996]).

de brillance progressif au début de l'extrait, et celui si soudain s'immiscant entre les 22 et 23^e secondes.

1.5 Aspects méthodologiques

Les travaux exposés dans cette thèse se consacrent à l'exploration du rôle du timbre dans l'interprétation musicale tant au niveau de la production musicale (poiétique) qu'au niveau de sa perception (esthétique). Notons qu'ils s'inscrivent au sein du projet pluridisciplinaire "Sensons" [SENSONS, 2008], dirigé par notre équipe. Ce projet rassemble, autour de la thématique générale du *sens des sons*, des études relevant de la psychoacoustique, des neurosciences cognitives, et de l'analyse-synthèse.

D'un point de vue méthodologique, nous nous sommes concentrés sur un instrument auto-entretenu monodique³³, la clarinette, reconnu pour ses qualités expressives et la richesse de ses timbres. En effet, comme nous avons pu le remarquer grâce aux Exemples Sonores de clarinette cités précédemment (12, 13 et 19), cet instrument offre au musicien la possibilité de modifier le timbre des sons générés pendant la production même des notes. Nous avons également limité notre champ d'investigation à des extraits musicaux provenant des répertoires Baroque (Bach) et Classique (Mozart), suffisamment riches pour alimenter ces premières recherches sur l'influence du timbre dans l'interprétation. Il est important de spécifier que nous n'avons pas abordé la question de l'image spatiale dans la notion de timbre. Nous avons cherché à identifier des effets de timbre robustes qui restent valables dans différentes situations d'écoute. Ce choix méthodologique part du constat qu'il est en effet possible d'apprécier une interprétation musicale malgré des conditions d'écoute bien différentes (par exemple salle de concert, haut-parleurs, casques, écoute dans une

³³qui ne peut jouer qu'une seule hauteur à la fois.

voiture), où le timbre et l'image spatiale de l'instrument ne sont pas restitués de la même manière. Les différentes expériences perceptives réalisées dans le cadre de cette thèse ont été menées en cabine audiométrique et avec une écoute au casque. Comme rapporté dans [Lavandier, 2005, p. 18], de nombreux travaux ont montré que les expériences perceptives réalisées au casque ou en écoute directe par des haut-parleurs dans des cabines audiométriques fournissaient des résultats similaires concernant la restitution du timbre.

Nous avons dans un premier temps utilisé un modèle de synthèse de clarinette reposant sur la modélisation physique de manière à comprendre les liens entre le fonctionnement acoustique de l'instrument et les timbres produits. En effet, de par leur conception, les modèles de synthèse reposant sur la modélisation physique permettent de générer des sons pour différentes combinaisons des paramètres de contrôle de l'instrument, et ce de manière reproductible. Les principaux paramètres de contrôle du modèle sont corrélés à la pression d'alimentation, et à la force exercée par l'instrumentiste sur l'anche, également appelée *pince*³⁴. Une expérience perceptive de jugements de dissimilarité a été menée afin d'analyser la représentation perceptive des différents timbres produits par le modèle, et de déterminer des corrélats acoustiques qui permettent d'en prédire la structure. Ces analyses ont permis de mettre en évidence des descripteurs temporels (par ex. Temps d'Attaque) et spectraux (par ex. Centre de Gravité Spectral, Rapport Impair/Pair), fortement caractéristiques des différents timbres de l'instrument. Les modèles de synthèse reposant sur la modélisation physique sont d'une grande utilité pour étudier le fondement de la relation entre contrôle instrumental et son, et permettent de mieux cibler les descripteurs de timbre susceptibles de varier au cours du jeu sur un instrument naturel. Cependant, ils restent généralement trop simplifiés pour reproduire les subtiles possibilités offertes par les instruments naturels. Les contrôleurs numériques auxquels ils sont associés (par ex. *Yamaha WX-5* pour la clarinette³⁵) ne sont que des ersatz des vrais instruments, que les musiciens doivent à leur tour apprivoiser. Pour ces raisons et afin d'étudier le timbre en situation d'interprétation musicale, nous avons par la suite dirigé nos recherches sur l'instrument naturel.

Dans un second temps, des analyses sonologiques de différentes interprétations musicales jouées par un clarinettiste professionnel ont été effectuées afin de déterminer les corrélats acoustiques de l'expression musicale. Nous avons développé une méthodologie d'analyse de l'expression musicale à partir de séquences musicales monophoniques. Cette dernière repose sur des représentations temps-fréquence des signaux (Transformée de Fourier glissante et analyse additive). Elle comprend une étape préliminaire de segmentation des notes, puis une étape de caractérisation objective de paramètres liés au timbre, au rythme, à la dynamique et aux hauteurs. Des analyses statistiques (par ex. analyse de variance, classification ascendante hiérarchique) ont été menées afin d'étudier d'une part si les variations de timbre produites par un interprète sont reproduites systématiquement lorsque son intention expressive est inchangée, et d'autre part si leur nature change lors de la comparaison entre des interprétations "scolaires" (neutres et inexpressives) et des interprétations expressives.

Nous avons par la suite cherché à évaluer l'importance des variations temporelles de timbre au niveau perceptif selon un paradigme d'analyse par la synthèse. Lors d'une première expérience, nous avons évalué l'influence perceptive de trois réductions de l'expression musicale portant sur le timbre, le rythme et la dynamique. Un modèle d'analyse-synthèse additif (de type sinusoïdal + bruit) a été employé pour effectuer les transformations de timbre (gel de la brillance des notes), et de rythme (modification de la durée des notes). La réduction de dynamique a été réalisée grâce à un filtre compresseur. Les analyses statistiques des jugements de préférence mesurés sur un ensemble de 20 participants indiquent que la suppression des variations temporelles de brillance des notes est

³⁴Le fonctionnement du modèle et ses paramètres de contrôle sont décrits au chapitre 3.

³⁵Ce contrôleur apparaît sur les photos de la figure 4.1 du chapitre 4

la transformation qui induit la plus grande perte d'expression musicale. Une seconde expérience complémentaire à la première a été conduite en vue d'examiner la possibilité d'améliorer l'expression d'une interprétation en agissant sur l'évolution temporelle de la brillance des notes. Une procédure de contrôle dynamique de la brillance par filtrage dynamique a été développée de façon à pouvoir appliquer les variations de brillance issues d'une interprétation à une autre interprétation. Il nous a ainsi été possible d'ajuster les variations de brillance de séquences inexpressives générées à l'aide d'échantillonneurs pour différents instruments (clarinette, flûte piccolo, violoncelle, sitar, guitare) sur celles d'une interprétation de clarinette expressive. 20 participants ont ensuite suivi une tâche de comparaison par paires des différentes séquences au cours de laquelle il leur était demandé de juger les dissemblances entre les séquences en terme d'interprétation, et de sélectionner les séquences préférées. Les analyses statistiques des mesures de préférence montrent que l'application des variations temporelles de brillance issues d'un jeu expressif améliore la qualité musicale des séquences générées à partir des échantillonneurs de manière fortement significative pour certains instruments. Cette amélioration s'opère de façon prononcée pour les instruments entretenus (clarinette, piccolo et violoncelle), et de façon beaucoup moins prononcée pour ceux de type non-entretenus (guitare et sitar).

1.6 Organisation de la thèse

Le chapitre 2 est consacré à la notion de timbre. Nous examinerons en quoi ce paramètre du son présente un caractère dual qui renvoie tantôt à l'identification des sources sonores (aspect désigné par *timbre-identité* dans [Marozeau, 2004]), tantôt à leur qualification (*timbre-qualité* (*ibidem*)). En se reposant sur les principaux résultats de la littérature, nous décrirons une méthodologie d'exploration du timbre visant à déterminer les corrélats psychoacoustiques de la représentation perceptive de différents timbre (appelés à ce titre *descripteurs de timbre*). A la fin de ce chapitre, nous présenterons un certain nombre de descripteurs de timbre, temporels, spectraux et spectro-temporels, que nous avons été amenés à employer lors des différentes expériences de cette thèse. Leurs formulations et les méthodes de calcul associées seront détaillées.

Au cours du chapitre 3, nous décrirons le fonctionnement général de la clarinette et les principaux facteurs acoustiques à l'origine de son timbre. Nous présenterons ensuite le modèle de synthèse de clarinette que nous avons utilisé dans l'expérience du chapitre 4. Ce dernier repose sur un modèle physique simplifié qui reproduit de manière réaliste le comportement de l'instrument en fonction de ses paramètres de contrôle (corrélés à la pression, et à la pince).

Le chapitre 4 présente l'étude de dissemblance des sons de synthèse de clarinette, réalisée en vue d'explorer la relation entre le contrôle instrumental et la perception des timbres produits par l'instrument.

Dans le chapitre 5, nous présenterons les méthodes et les résultats des analyses sonologiques d'interprétations musicales de clarinette. Nous montrerons comment l'intention expressive de l'interprète influe sur les variations de timbre.

Les deux études perceptives complémentaires réalisées afin de mieux comprendre l'influence des variations temporelles de brillance sur l'expression musicale seront exposées au chapitre 6.

Enfin, le chapitre 7 sera voué aux conclusions de ce travail, et aux perspectives auxquelles il donne lieu.

Chapitre 2

Les différentes facettes du timbre

La notion de timbre mérite à ce stade quelques éclaircissements. En effet, le premier chapitre nous a permis de nous rendre compte que le mot timbre pouvait être utilisé dans des contextes bien différents ; un compositeur parlera volontiers des *timbres* de l'orchestre par référence aux différents instruments qui le composent, tandis qu'on décrira souvent un instrumentiste au travers de son *timbre*, par renvoi à ce son si spécifique qui le caractérisera parmi tant d'autres, fruit d'années de travail et de recherches. D'autre part, il est clair que l'on peut également parler du timbre d'un son sans le rapporter à un instrument déterminé, mais plutôt en l'appréciant comme une caractéristique propre de ce son, fait mis en exergue avec l'avènement de la synthèse sonore. Nous voyons donc que suivant le contexte, la notion de timbre ne peut être prise selon les mêmes acceptions. Nous tâcherons au cours de ce chapitre de fournir quelques clefs permettant d'en distinguer les différentes facettes, ainsi que d'en préciser quelque peu la nature.

Nous commencerons par rappeler quelques-unes des précieuses réflexions de Schaeffer sur les corrélations entre acoustique et perception du timbre musical, réflexions desquelles il proposa une définition générale du timbre reposant sur le concept de permanence/variation [Schaeffer, 1966]. Nous analyserons ensuite les définitions standards du timbre qui sont employées couramment en psychoacoustique, non sans être controversées. Nous montrerons à ce titre quelques-unes de leurs limites. Enfin, nous présenterons la méthodologie et les résultats fondamentaux d'études portant sur la recherche d'une structure psychologique multidimensionnelle du timbre. Nous fournirons à cette occasion un glossaire des principaux *descripteurs de timbre*, paramètres psychoacoustiques qui rendent compte de certaines de ses propriétés.

2.1 La dualité du timbre : de l'identification à la qualification des sons

Comme le souligne Cadoz, "s'il faut parler du timbre, c'est en tant qu'*attribut perceptuel*¹. Ce qu'il faut entendre par là, c'est que sa valeur, sa fonction finale, ce sont l'oreille et l'intelligence musicale qui en décident, quel que soit le dispositif de production." [Cadoz, 1991, p. 17]. La notion de timbre semble donc intrinsèquement liée à la perception. Comme nous le verrons par la suite, si le timbre peut présenter plusieurs facettes, c'est parce que la perception, elle aussi, est multiple.

2.1.1 Lien causal du timbre à la source sonore

Dans son *Traité des objets musicaux*, Schaeffer propose une première définition empirique du timbre, qui comme le précise l'auteur, n'est qu'une lapalissade : "ce à quoi on reconnaît que divers sons proviennent du même instrument." [Schaeffer, 1966, p. 55]. Cette définition sous-tend la

¹C'est nous qui soulignons.

fonction primaire du timbre qui est de référer à la source sonore (par ex. les *timbres* de l'orchestre). L'origine de cette fonction provient très certainement de la signification étymologique du timbre : le mot vient du nom grec *tympanon*, qui désignait un instrument de musique, une sorte de tambour à cordes² [Risset, 2003]. Par extension, la *peau de timbre* correspond à la membrane inférieure d'un tambour de forme cylindrique à deux peaux ; il est possible de modifier le *timbre* du tambour en agissant, grâce à des clefs, sur une corde qui vibre contre la peau³ [CNRTL, 2008]. La caisse claire (*snare drum* en anglais), élément central des batteries, percussions largement utilisées dans les styles musicaux jazz et rock, dérive très certainement de ce type de tambour. Autrefois en boyau, la corde tendue en double contre la peau inférieure du tambour est de nos jours substituée par une série de cordelettes métalliques (on en trouve par exemple en acier, en bronze, en cuivre) également tendues sur les peaux de timbre des caisses claires (cf. photos de la figure 2.1). Elles ont pour effet de donner à la percussion sa couleur caractéristique, en ajoutant de la brillance au son. Ce n'est d'ailleurs pas un hasard si elle porte le nom de caisse *claire*. Dans l'appellation anglaise, ce n'est pas l'effet qui est mis en valeur mais sa cause, *snare* désigne les cordelettes. En français, effet et cause sont entrelacés : le *timbre de la caisse claire* sert également à désigner les cordelettes ! Il est d'ailleurs possible, au biais d'un levier, d'ôter son timbre (cordelettes) à la caisse claire (en désolidarisant les cordelettes de la peau).

Le mot timbre est également associé à un autre instrument, le *glockenspiel*, que l'on retrouve derrière les noms de carillon, jeu de clochettes, ou *jeu de timbres* [Scholes, 1960, p. 788], le mot désignant alors des lames qui, lorsqu'elles sont mises en action par frappement, produisent un son aigu et cristallin.

Mais peut-être que son utilisation la plus classique était celle de désigner une *cloche* que l'on frappait avec un petit marteau pour appeler⁴. Dans certaines langues (par ex. espagnol), nos sonnettes électriques modernes ont parfois conservé ce nom, comme en témoigne la photo de la figure 2.2.

Enfin, comme le souligne Schaeffer, si le timbre musical ne peut être traité au premier abord, autrement que par un renvoi à l'instrument, une "marque de fabrique" du phénomène sonore, cela ne doit sans doute pas être étranger à l'autre signification importante que revêt le mot, celle de désigner la marque ou le cachet apposé sur un document pour en garantir l'origine (par ex. le *timbre* de la poste). La photo de la figure 2.3 a le mérite de rassembler - non sans un brin d'humour - les différentes facettes du timbre que nous venons de mentionner.

De par cette vertu d'*identification* de la source sonore, le timbre s'empreint donc d'une nature *typologique*⁵, et ouvre les portes de la classification instrumentale. Cette fonctionnalité du timbre est judicieusement mis en avant dans l'expression *timbre-identité*, introduite par Marozeau *et al.* [Marozeau, 2004].

²Définition du tympanon d'après le Centre National de Ressources Textuelles et Lexicales : "instrument composé d'une caisse trapézoïdale sur laquelle sont tendues des cordes de laiton que l'on frappe avec deux baguettes d'acier ou deux petits maillets de bois, et qui fut très répandu jusqu'au XVIII^e s. [citation] "Il fallait rester là jusqu'au soir [à écouter les musiciens tziganes] devant un verre de vin de Hongrie, la gorge serrée, les yeux fous, tout le corps secoué au battement nerveux du tympanon." (A. Daudet, *Trente ans Paris*, 1888, p. 291)." [CNRTL, 2008].

³La dualité du timbre, objet de la première partie de ce chapitre, apparaît déjà : le mot timbre est utilisé tantôt pour désigner l'instrument (le tambour), tantôt pour caractériser le son qu'il produit.

⁴[citation] : "Ali ! cria t-il puis il frappa un coup sur le timbre de cuivre. Ali parut." (Dumas, père, *Monte-Cristo*, t. 1, 1846, p. 676)" [CNRTL, 2008].

⁵"Typologie : science de l'analyse et de la description des formes typiques d'une réalité complexe, permettant la classification." [CNRTL, 2008].



FIG. 2.1 – *Peau de timbre* et *timbre* d'une caisse claire de batterie.



FIG. 2.2 – Sonnette d'entrée ou *timbre* au Mexique.



FIG. 2.3 – *Timbres* d'instruments de musique.

2.1.2 Le paradoxe du timbre

Même si le lien causal primitif du timbre à la source sonore est primordial, nous voyons bien qu'il ne peut se suffire à lui-même. En effet, Schaeffer met rapidement en exergue ce paradoxe remarquable que sous-tend l'utilisation du terme timbre "qui veut à la fois que les instruments aient *un* timbre, et que chaque objet sonore qu'on en tire ait, pourtant, *son* timbre particulier." [Schaeffer, 1966, p. 233]. C'est bien ce dont il était question lorsque nous décrivions, au chapitre 1, les subtilités timbrales offertes par le shamisen japonais, et dont seul le musicien expert saura tirer profit. Anticipons donc les réflexions des prochaines sections et désignons par *timbre-qualité*, expression employée dans [Marozeau, 2004], cet aspect de l'écoute qui permet au timbre d'être non plus compris comme un indice, signature de la cause sonore, mais également comme une caractéristique propre du son, alors analysé selon ses effets. Le paradoxe précédent évoqué par Schaeffer se voit quelque peu éclairci si l'on raisonne ainsi : un instrument dispose d'un *timbre-identité* propre (ce qui fait que l'on est capable de reconnaître l'instrument parmi d'autres), mais peut disposer d'une large palette de *timbre-qualités*, dont l'obtention est étroitement liée à la nature du contrôle instrumental. Il est important de noter que les notions de timbre-identité et de timbre-qualité ne sont pas indépendantes l'une de l'autre (par ex. si l'on varie le timbre-identité, i.e. si l'on change d'instrument, intrinsèquement le timbre-qualité en sera changé), mais leur intérêt est de mettre en valeur des fonctions différentes de l'écoute (cf. figure 1.12). Il nous paraît important de tenir compte du fait que la notion de timbre-identité, qui renvoie aux causes productrices du son, peut s'établir suivant différents niveaux. Ainsi, nous désignerons par *timbre-identité de l'instrument* l'aspect de la perception qui nous permet de reconnaître un instrument parmi d'autres. Nous désignerons par *timbre-identité du geste* l'aspect de la perception qui nous permet de reconnaître le geste à l'origine du timbre-qualité particulier d'un instrument.

2.1.3 Définition générale du timbre schaefferienne : le concept de permanence / variation

L'auteur du *Traité des objets musicaux* propose alors une définition plus générale du timbre instrumental : "une variation musicale assouplissant et "compensant" une permanence causale." [Schaeffer, 1966, p. 239]. Il existe selon Schaeffer deux types de sources de variations du timbre instrumental venant compléter les effets sonores induits inéluctablement par la physique de l'instrument ainsi que la nature générale du procédé d'excitation qui lui est associé :

- **les sources de variations dites "abstraites"** provenant de la *registration*⁶, que l'on peut définir comme l'art (relevant de la lutherie) d'élaborer la sonorité de l'instrument le long de sa tessiture. Bien souvent, des variations de timbre existent aussi entre plusieurs instruments du même type. Nous ajouterons donc aux sources de variations abstraites ce qui a trait, en termes de facture, aux choix esthétiques entrepris par les luthiers et qui concernent la sonorité générale d'un instrument (telle guitare ne sonnera pas de la même manière que telle autre). Les variations "abstraites" sont donc liées aux notions de *timbre-identité de l'instrument* et de *timbre-qualité* ;
- **les sources de variations dites "concrètes"**, résultant du contrôle instrumental. Les variations "concrètes" sont liées aux notions de *timbre-identité du geste* et de *timbre-qualité*.

Examinons maintenant plus précisément ce qui se cache derrière ces notions.

⁶"registre : partie de l'étendue totale d'un ensemble de sons (*extrême grave, grave, médium, aigu et suraigu*). Sur beaucoup d'instruments et sur la voix, un registre se caractérise par une homogénéité de sonorité. Les *changements de registre* jouent un important rôle expressif." [Siron, 2004].

2.1.3.1 La registration et la facture instrumentale, sources de variations de timbre "abs-traites"

Registration Il est bien connu que les instruments de musique disposent de différents registres. Si ces derniers opèrent une division de la tessiture d'un instrument en fonction des hauteurs, ils ne reposent pas moins sur la notion de timbre. En effet, c'est par l'homogénéité des timbres des notes qui le composent que les registres s'organisent. Ainsi, au timbre chaud de la clarinette dans le registre grave⁷, s'oppose le timbre extrêmement brillant, voire agressif, de ses registres plus aigus, baptisés clairon⁸ et altissimo. Nous verrons au prochain chapitre consacré à la clarinette comment ces variations de timbre inter-registres peuvent s'expliquer par le fonctionnement acoustique de l'instrument. La sonorité des différents registres est donc liée à la physique de l'instrument. Il ne faut pas perdre de vue que ce qui détermine la facture instrumentale est le fruit d'un long travail traversant les époques, ajusté par essais et erreurs, et nourrit du savoir-faire des luthiers. Ainsi l'exprime Schaeffer : "A la permanence causale (corde frappée) [NDLR : cas du piano], fait pendant une certaine variation musicale des effets, voulue par le luthier, dosée en fonction d'exigences artistiques⁹, et obtenue mécaniquement par des moyens divers : doublage ou triplage des cordes dans l'aigu, filage des cordes graves¹⁰, couplage et résonance due à la table d'harmonie, feutrage plus ou moins épais des marteaux, etc." (ibidem, p. 238). C'est donc en fonction de critères musicaux dont le développement se fait au travers d'une longue évolution, que les instruments connaissent leur facture actuelle. Bien sûr, aux critères de sonorité s'ajoutent des éléments portant, entre autres, sur la jouabilité et l'intonation¹¹ (cf. par exemple le système de cléage *Boehm* pour les instruments à vent).

Ecoles de facture instrumentale C'est aussi pourquoi nous trouvons, pour de nombreux instruments, différentes écoles de facture instrumentale. "Les facteurs de clavecin ont élaboré deux méthodes générales de construction. L'école italienne a notamment construit des instruments de petites dimensions : la longueur de la corde donnant le *do* médium est comprise entre 25 et 27 centimètres ; le pincement est rapide ; la sonorité claire, pénétrante et résonante. Les facteurs flamands, en revanche, fabriquaient des instruments plus massifs et plus grands (la longueur de la corde du même *do* atteignait 35 centimètres). La sonorité de ces clavecins, ronde et suave, était très différente de celle des clavecins italiens." [Kottick *et al.*, 1995, p. 41]. Et d'apprendre un peu plus loin dans l'article de Kottick "L'oreille a une autre particularité qui fait du clavecin un instrument à la sonorité riche : le système auditif "entend" bien les sons plus aigus d'un instrument, mais mal les sons graves. Pour compenser, les instruments de musique utilisent bien plus d'énergie pour produire les notes graves que les notes aiguës. Ainsi un bon clavecin doit être plus puissant dans les graves afin de produire une sonorité apparemment uniforme sur tout le registre de l'instrument. C'est d'un équilibre raffiné entre les cordes, le bois, l'air, l'oreille et le cerveau que naît la "sonorité" du clavecin." (ibidem, p. 48). Une autre particularité est à noter concernant certains clavecins, celle de posséder deux séries de cordes, auxquelles sont associées deux séries de sauteraux, dont les plectres viennent pincer les cordes à des endroits différents [Fletcher et Rossing, 1998, p. 341]. Comme nous avons pu le voir au sujet de la guitare, pincer une corde à des positions différentes induit des différences de timbre notables. Aussi ce procédé est utilisé dans le cas du clavecin pour

⁷registre également appelé *chalumeau*, d'après le nom d'un instrument à anche simple dont l'apparition a précédé celle de la clarinette (cf. chapitre 3).

⁸La sonorité des notes du registre clairon se rapproche de celle de la *clarine*, petite trompette du *XVIII^e* siècle, qui donna son nom à la clarinette.

⁹C'est nous qui soulignons.

¹⁰Le filage des cordes graves de piano leur confère une masse linéique plus importante, ceci afin de produire des notes graves tout en gardant une longueur de corde raisonnable.

¹¹"art de poser le corps d'un son et d'en placer la hauteur avec précision. Difficile sur beaucoup d'instruments, elle peut être précise (*juste*) ou imprécise (*fausse*). Joue souvent un caractère expressif." [Siron, 2004]

conférer au son de l'instrument une variété de contenu harmonique.

A l'instar du clavecin, il existe également pour la clarinette deux écoles principales de facture instrumentale, qui vont de pair avec deux techniques de jeu (cf. section 2.1.3.2 qui traite des variations de timbre "concrètes" dues au contrôle instrumental) et pôles sonores différents. L'école "allemande" propose des becs présentant une longue table¹² et des anches conséquemment longues. Elle est associée à des sons plutôt "sourds". A l'inverse, le système de construction de l'école "française" consiste en des becs et des anches plus courts. Il est associé à des sons plutôt "brillants". Ceci peut être l'un des facteurs expliquant les différences de sonorité des clarinettes de Brymer et Goodman dans leurs interprétations du concerto de Mozart (cf. [Exemple Sonores n^{os} 12 et 13](#) dont il est question à la section 1.4.1).

Règles d'identité timbrale Face à ces variations de timbre liées à l'essence même de l'instrument, il est légitime de se demander ce qui conduit alors l'auditeur à conférer aux sons provenant des différents registres instrumentaux une identité commune. De par une démarche analytique s'attachant à comprendre ce qui établit l'identité instrumentale du piano le long de sa tessiture, Schaeffer aboutit aux constatations suivantes : à une dynamique¹³ d'autant plus raide¹⁴ que la tessiture est aiguë, s'associe une complexité harmonique d'autant plus riche que la tessiture est grave, aspect dont il formule les effets, qui peuvent paraître paradoxaux, ainsi : "le timbre harmonique¹⁵ du piano est non seulement plus riche, mais plus clair pour le registre grave, et il est plus pauvre et plus sombre pour le registre aigu" [Schaeffer, 2005, p. 48]. Il souligne ainsi que la richesse spectrale est d'autant plus élevée que les notes sont graves. Il propose alors une règle d'ensemble, illustrée par la figure 2.4, à laquelle semblent obéir les timbres particuliers des notes du piano :

$$\text{Raideur dynamique} \times \text{Richesse harmonique} = \text{constante}$$

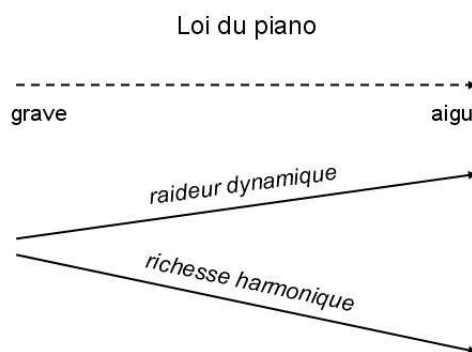


FIG. 2.4 – Loi d'identité timbrale du piano le long de sa tessiture selon Schaeffer.

Il est important de noter que le fait que les registres graves soient plus riches que les registres aigus est vrai pour la plupart des instruments, ceci étant lié au fait que les pertes (par frottements, thermiques, etc.) augmentent avec la fréquence.

¹²La table d'un bec de clarinette correspond à la zone sur laquelle l'anche vient s'enrouler lors de ses oscillations (cf. figure C.1 de l'Annexe C).

¹³Rappelons que la dynamique en musique réfère aux variations d'intensité du son au cours du temps. Sur le plan physique, la dynamique est donc liée à l'enveloppe d'amplitude (ou d'énergie) temporelle du son, qui peut-être obtenue de différentes façons (par ex. enveloppe RMS (Root Mean Square), cf. section 2.2.5).

¹⁴La raideur dynamique fait référence à l'inclinaison de la pente de l'enveloppe temporelle d'énergie d'une note.

¹⁵Par cette formulation, Schaeffer fait référence aux aspects spectraux du timbre, par opposition aux aspects temporels (comme l'attaque).

Dès lors, nous nous rendons compte que le timbre ne peut se définir comme un paramètre monodimensionnel, comme c'est le cas de la hauteur, qui pour les sons à "hauteur déterminée" est quantifiable en première approximation par l'écart de fréquence entre les différents harmoniques, qui généralement correspond à la fréquence fondamentale. Le timbre, lui, ne pourra se définir qu'en tant qu'attribut perceptuel référant à plusieurs paramètres physiques. On le décrira alors comme une **variable multidimensionnelle**. Les études mentionnées ci-dessus témoignent déjà de la dépendance du timbre envers des aspects du son qui relèvent tant du domaine temporel (par ex. attaque) que du domaine spectral (par ex. contenu harmonique du son) : on parlera alors de dépendance *spectro-temporelle*. La compréhension des fondements de l'identité instrumentale est complexe. Il semble que l'identité instrumentale est régie par l'existence de *relations entre paramètres dont l'évolution est corrélée* (par ex. loi du piano). Ainsi, Risset a montré que la "signature" acoustique des sons de trompette résidait en grande partie dans la relation entre l'intensité et la richesse harmonique des sons, ceci permettant de reproduire l'effet "cuivré", propre à ce type de sons [Risset et Mathews, 1969] : une augmentation de l'intensité est couplée à une augmentation de la proportion d'harmoniques de rangs élevés (de fréquences élevées). En se reposant sur l'analyse de sons naturels, Risset a montré que l'on pouvait reproduire les sons de trompette de manière efficace en accordant leur contenu harmonique sur l'amplitude du premier harmonique, selon une certaine loi, telle que décrite par la figure 2.5.

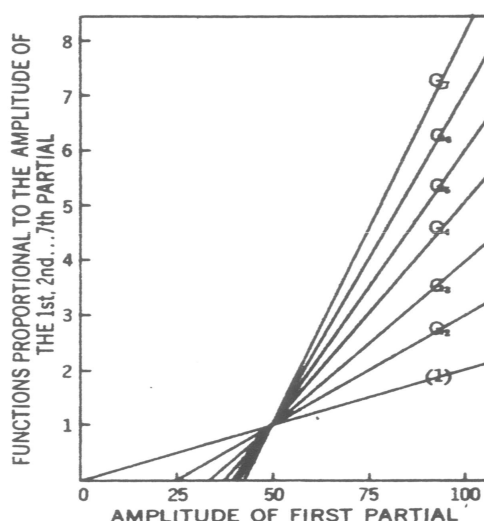


FIG. 2.5 – Synthèse de sons de trompette : loi d'évolution de l'amplitude des composantes spectrales (G_i) en fonction de l'amplitude du premier partiel. Figure issue de [Risset et Mathews, 1969].

Notons cependant que cette loi d'évolution n'est pas suffisante pour expliquer la particularité de certains timbres instrumentaux car elle est commune à de nombreux instruments. Ceci est une conséquence naturelle des non-linéarités associées à la production des sons instrumentaux [Fletcher et Rossing, 1998, p. 564]. Nous verrons en effet au chapitre 3 qu'une relation analogue existe pour la clarinette (telle que l'illustre la figure 3.3, issue des travaux de Benade).

Nous avons donc vu au travers de quelques exemples combien la facture instrumentale, dont l'un des éléments est la registration, peut être fondamentale tant pour la qualité sonore d'un instrument, que pour son identité même. Ces variations de timbre ont été désignées comme des variations *abstraites*, dans la mesure où une fois l'instrument donné, existant, elles en font partie intégrante, et deviennent donc transparentes pour l'instrumentiste. Cependant, nous souhaitons souligner que la dichotomie variations de timbre abstraites / concrètes n'est pas exclusive. En effet, il existe des

instruments pour lesquels le musicien peut de par les choix qu'il opère, aller chercher un "sous-instrument" contenu au sein du premier, et qui présentera une sonorité originale. C'est le cas de l'orgue, dont les jeux de mutations permettent à l'organiste de façonner le timbre d'une note par la sélection des harmoniques qui devront être joués ([Risset, 1991, p. 248], [Fletcher et Rossing, 1998, p. 566])¹⁶, de l'accordéon qui possède des boutons permettant de passer d'un timbre à l'autre, mais encore de la guitare électrique, qui offre à l'instrumentiste une variété de choix de sons, allant du "chaud" au "criard", selon les types de micros (grave, médium, aigu) ou combinaison de micros utilisés afin de restituer la vibration de la corde¹⁷. Nous évoquerons également à ce sujet les éléments additionnels qui peuvent s'ajouter à l'instrument initial pour en transformer la sonorité (par ex. la sourdine pour la trompette, mise au devant de la scène jazz par Miles Davis, la cabine *Leslie* couplée à l'orgue *Hammond*, ajoutant au son des variations spectro-temporelles résultant d'effets Doppler, sans compter la spatialisation sonore qu'elle entraîne).

2.1.3.2 Le jeu instrumental, source de variations de timbre "concrètes"

Comme déjà évoqué dans l'introduction, **pour certains instruments, la relation instrumentale, née de l'association musicien/instrument, permet d'engendrer des variations de timbre**, malgré une facture instrumentale qui elle est a priori stable.

Mentionnons ces propos de Schaeffer au sujet du contrôle instrumental : "Même en observant les règles du jeu [NDLR : i.e. dans des conditions de gestes instrumentaux traditionnels], il peut délivrer des variétés d'objets où émerge, en propre, la *facture*. Dans l'instrument le plus stéréotypé, le piano, on admet qu'il existe un "toucher" propre au pianiste. [...] A fortiori, un violoniste ou un flûtiste sont en mesure de tirer de l'instrument une variété d'objets s'insérant cependant dans les mêmes registres ou possédant la même forme : objets liés, piqués, pizz, vibrato, etc., mais où domine leur personnalité, leur "son", comme on dit, car il ne suffirait pas non plus de parler de leur "timbre" " [Schaeffer, 1966, p. 56]. Nous voyons bien d'une part que les timbres des sons instrumentaux sont ductiles¹⁸, et d'autre part qu'il émerge au travers du timbre une seconde forme d'identité, celle de l'instrumentiste¹⁹. Outre de caractériser un instrument, l'oreille²⁰ (une oreille avertie dirons-nous) peut être à même de reconnaître en quelques notes le son d'un musicien dont elle s'est éprise, ou même au contraire qu'elle abhorre. Effectivement de par leur son (mais aussi leur jeu, n'oublions pas que la musique implique généralement des *phrases*), les musiciens se forgent une identité sonore spécifique. Nous supposons que ce fait est d'autant plus avéré pour les instruments qui engagent de manière directe ou indirecte le conduit vocal (par ex. voix, instruments à vent), et donc la morphologie interne si spécifique à chaque individu. Un fan de Miles Davis n'identifierait-il pas son trompettiste préféré simplement à l'issue de quelques-unes de ses notes ? Quel passionné de rock, ne devinerait pas Jimi Hendrix derrière ce son de guitare électrique si unique²¹ ? Bien sûr la recherche d'une identité sonore nouvelle, voire insolite ou excentrique sera certainement plus saillante dans des styles de musique comme le jazz, ou le rock, où les solistes sont souvent mis au devant du groupe ou de l'orchestre que dans la musique classique, où certaines contraintes esthétiques imposent au soliste d'unir son son avec celui de l'orchestre.

¹⁶Risset souligne d'ailleurs à ce propos que le principe fondamental de la synthèse additive avait déjà été compris par les facteurs d'orgue du Moyen-Âge !

¹⁷Nous pourrions l'appeler à cette occasion guitare *éclectique* !

¹⁸dont la forme est modifiable, changeante.

¹⁹Il est fort probable que d'autres facteurs musicaux que le timbre participent également à l'identité instrumentale

²⁰Nous sommes conscients de cet abus de langage, que l'on trouve d'ailleurs couramment, et qui est de parler de l'oreille comme seule protagoniste de la perception (e.g "oreille d'or"), alors que nous savons bien que le cerveau lui aussi a son mot à dire dans la perception.

²¹Et pourtant, paradoxalement, il n'y pas plus courant que l'instrument utilisé par ledit musicien, la fameuse *Fender Stratocaster*, qui n'en doutons pas a été vendue à des millions d'exemplaires. Il faut cependant noter que dans cet exemple, les effets additionnels utilisés par le musicien (par ex. distorsion, *flanger*) tiennent également une place très importante dans son identité sonore.

Techniques de jeu Nous avons mentionné les différentes écoles de facture de la clarinette, et le fait qu'elles y trouvent des pendants en termes de techniques de jeu. Comme nous l'apprend le clarinettiste et découvreur Pierre André Taillard²², avec une technique "allemande", le clarinettiste prend beaucoup de bec et place sa lèvre inférieure près de la naissance de la table. La pression de la lèvre qu'il exerce sur l'anche est très modérée et il utilise principalement la pression de l'air pour la courber et l'amener à osciller. Le timbre est contrôlé en grande partie par la pression d'air et la configuration du conduit vocal. En revanche, avec une technique "française", le clarinettiste place sa lèvre nettement après la naissance de la table et comprime assez fortement l'anche sur la table. La majorité du contrôle de la nuance et de la sonorité se fait par l'intermédiaire de la pression exercée sur l'anche. Selon le clarinettiste Claude Crousier, avec qui nous avons travaillé, les notions d'école allemande et d'école française fondées sur des traditions dans le jeu et des améliorations apportées à la clarinette propres à ces deux pays, ont changé de nos jours. En effet, les tendances s'étant diversifiées, il n'existerait plus une "école française de France", ou une "école allemande d'Allemagne". Le choix de la technique de jeu adoptée par un clarinettiste découle de son choix d'oreille. Ainsi, il préfère parler du son choisi par le clarinettiste. Les "sons plutôt sourds" seraient des "sons plutôt allemands" et les "sons plutôt clairs", des "sons plutôt français". Claude Crousier est convaincu qu'avec le même bec et la même anche, il est possible d'obtenir ces deux pôles sonores à la clarinette, "sourd" ou bien "clair". Le clarinettiste Jack Brymer partage d'ailleurs le même avis : "D'un point de vue strictement scientifique, les deux types de clarinette [allemande et française] sont virtuellement identiques car ce sont les fervents de l'un ou de l'autre type d'instrument qui décident de la façon dont la clarinette doit sonner en fonction de différences acoustiques infimes." ([Brymer, 1979], p. 21). Dans les interprétations du Concerto de Mozart par les clarinettistes Brymer et Goodman (cf. [Exemple Sonores n^{os} 12 et 13](#)), les différences de timbres entre les deux instrumentistes proviennent probablement de telles différences de techniques de jeu, en plus d'être liées à de possibles changements de facture instrumentale²³. Les deux pôles sonores de la clarinette, l'un sourd, l'autre clair, sont d'ailleurs ici bien représentés.

Certains traités de musique ne se trompent pas sur la question du timbre, ainsi Siron le caractérise comme étant "En rapport avec la couleur et avec le mode de production d'un son, caractéristique de chaque instrument, voire de chaque instrumentiste." [Siron, 2004, p. 736]. **Nous voyons bien là une autre facette du timbre-identité, celle qui porte non plus l'empreinte de l'instrument, mais celle de l'instrumentiste même.**

Références aux variations de timbre dans les systèmes notationnels Le système de notations de la musique traditionnelle occidentale ne réserve que très peu d'indications au timbre. L'indication de l'instrument pour lequel est écrit la pièce est bien sûr une référence directe au timbre-identité, mais même ce dernier n'est pas toujours présent. En effet, comme nous l'apprend Risset, de nombreuses pièces baroques ont été écrites pour "divers instruments" sans autres indications d'instrumentations [Risset, 2003]. Il n'y pas d'équivalent pour le timbre de ce que représentent les notes pour les hauteurs. En effet, il n'existe pas de notations qui symbolisent de manière précise les évolutions de timbre-qualité dans le temps. Il existe cependant des notations qui, par référence aux différents modes de jeu d'un instrument spécifique, sont liées au timbre-qualité. C'est le cas par exemple des symboles employés dans les partitions synoptiques des musiques traditionnelles asiatiques, et dont nous avons parlé section 1.1.4. Pour le violon, nous trouverons des indications telles que *sul ponticello* (près du chevalet), *spiccato* (avec l'archet qui rebondit), *col legno* (avec l'archet retourné, en jouant avec le bois), *con ou senza sordini* (avec ou sans sourdine), qui engagent différentes sonorités de l'instrument [Risset et Mathews, 1969] [Risset, 2003].

²²dans un entretien rapporté dans la thèse de Claudia Fritz [Fritz, 2004].

²³Dans la considération de ces effets, nous supposons ici d'ordre secondaire les phénomènes liés à l'acoustique de la salle de concert, ou encore à la chaîne électroacoustique impliquée dans l'enregistrement des exécutions musicales des deux instrumentistes.

Dans la musique contemporaine les indications des compositeurs sont parfois très détaillées, comme l'on pourra en juger par ces instructions de William Russel, sur la partition de son oeuvre *Fugue for Height Percussion Instruments* (1933) :

“‘Handkerchief over drum head’ [of snare drum] ; ‘Sweep wire brush stick across head near rim’ [timpani]. ‘Scratch strings, lengthwise, along winding, with a coin held like a banjo pick’ [piano] ; ‘Pizzicato on the strings of the piano with fingertips. ‘Place a piece of paper over drum head’ [snare drum]. ‘Pizzicato with back of finger nail’ [piano]. ‘Rub resined glove or cloth over a snare drum stick.’” [Scholes, 1960, p. 787].

Certaines de ces instructions précisent en détail le dessein sonore du compositeur, et ont donc pour finalité de désigner des variations de timbre-qualité. Mais si elles le font, c’est toujours de manière indirecte, par des descriptions causales (par ex. “mouchoir sur le tambour [de la caisse claire]”), et non par une description des effets (qui, par rapport à cet exemple, auraient pu être : *caisse claire : son feutré*).

La figure 2.6 tente de synthétiser les différents facteurs pouvant agir sur le timbre instrumental que nous venons de décrire.

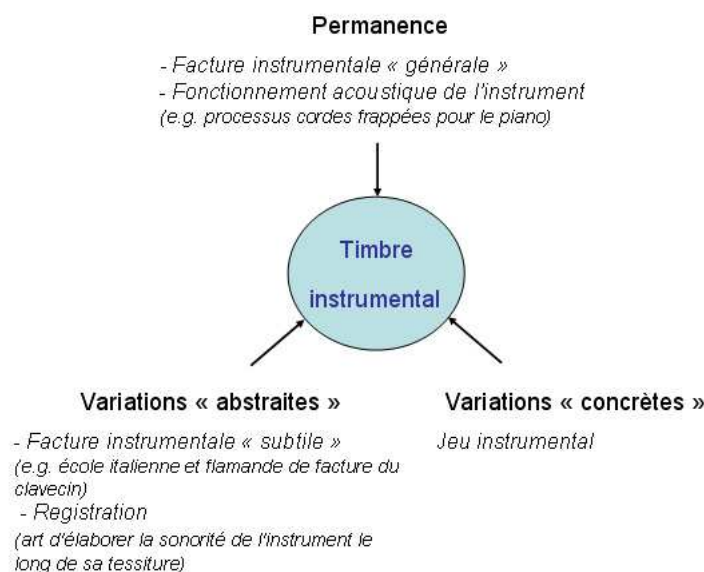


FIG. 2.6 – Timbre instrumental : concept de permanence / variation.

2.1.4 Le timbre en tant que caractéristique sonore

Puisque le timbre est variable pour un même instrument en fonction des registres, et au-delà, en fonction des instrumentistes, nous sentons bien que l'on ne peut limiter sa description à l'identification de causalités (*timbre-identité*), mais que notre perception peut également se concentrer sur les *effets*, sur la couleur sonore, propre aux qualifications (par ex. un son “rond”, “chaud”, “clair”, “nasillard”, etc.). Schaeffer remarque en effet que l'on peut faire allusion au “*timbre d'un son sans le rapporter clairement à un instrument déterminé, mais plutôt en le considérant comme une caractéristique donnée propre de ce son, perçue pour elle-même.*” [Schaeffer, 1966, p. 232]. Comme nous l'avons déjà mentionné ci-dessus, cette facette du timbre sera désignée par *timbre-qualité* [Marozeau, 2004].

Le timbre ou la couleur sonore C'est d'ailleurs cet aspect du timbre qui prévaut parfois lorsqu'on tâche de le définir. Nous citerons à ce titre Scholes : "Timbre means *tone quality*²⁴ - coarse or smooth, ringing or more subtly penetrating, 'scarlet' like that of the trumpet, 'rich brown' like that of the cello, or 'silver' like that of the flute. These colour analogies come naturally to every mind, as does the metaphorical term now become a commonplace as synonym for 'timbre' - tone colour ; the German for 'timbre' is *Klangfarbe* - literally 'sound-colour'." [Scholes, 1960, p. 1028] (traduction : Le timbre signifie la *qualité sonore* - rugueuse ou lisse, éclatante ou plus subtilement pénétrante, 'écarlate' comme celle de la trompette, 'brun riche' comme celle du violoncelle, ou 'argentée' comme celle de la flûte. Ces analogies de couleur viennent naturellement à l'esprit, au même titre que le terme métaphorique couleur sonore, dont l'utilisation en tant que synonyme de 'timbre' est maintenant banale ; En Allemand, le mot timbre est désigné par *Klangfarbe* - littéralement 'couleur sonore'). Nous nuancerons cependant la définition de Scholes qui instaure que timbre et qualité sonore sont synonymes. La qualité sonore nous paraît être quelque chose de plus général que le timbre puisqu'elle peut être affectée par d'autres facteurs que le timbre, des changements de hauteur ou d'intensité, par exemple.

Typologie et Morphologie Avec l'élaboration d'un solfège des objets musicaux s'adonnant à décrire les effets du sonore plus que leurs causes, Schaeffer introduit les notions de *forme* et de *matière* : "ainsi défini, le timbre d'un objet n'est pas autre chose que sa forme et sa matière sonores, sa complète description, dans les limites des sons que peut produire un instrument donné, compte tenu de toutes les variations de facture qu'il permet." [Schaeffer, 1966, p. 232]. A l'aspect *typologique* du timbre évoqué plus haut, lié à la classification des instruments de musique reposant sur la manière dont les sons sont produits (cf. par ex. la classification de Sachs/Hornbostel), se double donc un aspect *morphologique*²⁵, qui a trait à la description de la forme du son. Typologie et morphologie ne sont pas indépendantes, loin s'en faut, elles se "renvoient l'une à l'autre, ne peuvent s'élaborer que par approximations successives" [Schaeffer, 2005, p. 144].

Caractère fonctionnel du timbre en musique C'est bien de son aspect ayant trait à la qualité sonore que s'est dégagé au *XX^e* siècle, le nouveau caractère fonctionnel du timbre en musique. Rappelons à ce propos l'exemple de *Farben* de Schönberg (cf. l'[Exemple Sonore n° 1](#) de la section 1.1), où l'identité instrumentale se perd au profit de la couleur sonore générale de l'orchestre, et de son évolution dans le temps. Il en va de même en musique concrète²⁶, où force est de constater que l'auditeur coupé de sa compétence à identifier les sources traditionnelles sonores, se doit de développer une oreille sensible aux qualités du contenu.

2.1.5 Bilan : les deux écoutes du timbre

La dualité du timbre (identité et qualité) est intimement liée à deux fonctions de la perception auditive, celle d'identifier les sources sonores et donc de traiter le son en tant qu'indices, et celles de les qualifier (cf. section 1.3.2). La figure 2.7 tente de synthétiser les différentes notions que nous venons de voir. La perception du timbre renvoie tantôt au pôle des causes (*timbre-identité de l'instrument* et *timbre-identité du geste*), tantôt au pôle des effets (*timbre-qualité*).

²⁴C'est nous qui soulignons. Notons au passage que l'anglais fait judicieusement la distinction entre note musicale, désignée comme en français par 'note', et le son associé, qui est désigné par 'tone'.

²⁵"Morphologie : science, étude des formes.", d'après [CNRTL, 2008].

²⁶Musique élaborée par Schaeffer et ses collaborateurs du Groupe de Recherches Musicales (GRM) de Paris dans les années 60, et reposant sur le précepte de "ne plus recourir aux causes pour confronter les effets, et [à] découvrir dans les effets de la sonorité les critères de l'objet." [Schaeffer, 2005, p. 138]. L'"anonymat de la bande magnétique" joue alors le même rôle que la "tenture de Pythagore qui masquait l'orateur, voilait son geste, ne laissait émerger que le sens." (ibidem).

L'hypothèse "fonctionnaliste" Lorsque l'on tente de comprendre pourquoi notre système auditif a développé de telles capacités d'identifications des sources sonores et des événements qui leurs sont liés (une simple petite maladresse ou accident de l'instrumentiste ne passe pas inaperçue, cf. la citation de Boris Vian décrivant sur la sonorité de Miles Davis dans l'épigraphie de la thèse), l'hypothèse "écologique" ou "fonctionnaliste" préconisée par Gibson [Gibson, 1966], et approfondie au cas des sons entre autres par Bregman [Bregman, 1994], McAdams [McAdams, 1994] et Risset [Risset et Wessel, 1999], est souvent évoquée. Comme le rapporte Risset, "Suivant ce point de vue, les sens - l'audition en particulier - auraient évolués pour extraire au mieux, à partir des signaux physiques qui lui parviennent, des informations sur l'environnement utiles à la survie." [Risset, 1994, p. 103]. Quoique l'écoute musicale soit tout de même bien loin d'avoir à faire avec la survie²⁷, on peut en effet concevoir que progressivement le système auditif se soit spécialisé de tel sorte qu'il puisse entendre les subtilités des sons musicaux, et notamment distinguer les différents timbres instrumentaux.

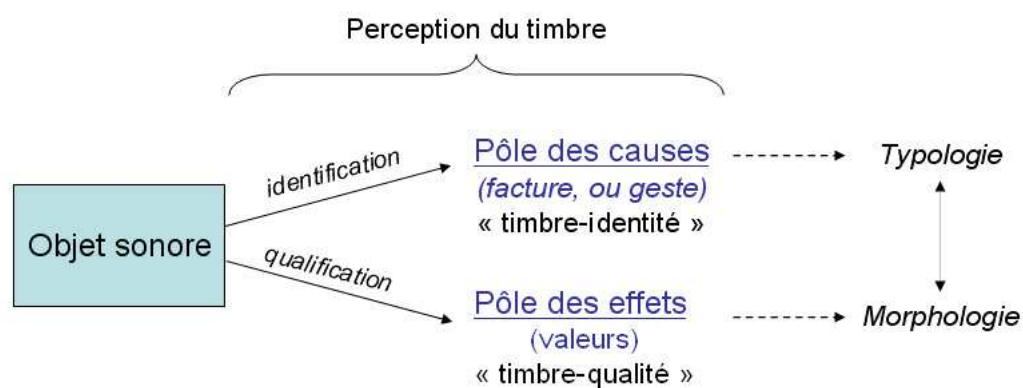


FIG. 2.7 – Schéma synoptique des deux écoutes du timbre rendant compte de la dualité identification/qualification des sources sonores.

Timbre et catégorisation cognitive Les théories de la catégorisation cognitive expliquent le caractère biface du timbre. D'après Castellengo et Dubois, les sons musicaux peuvent être catégorisés selon un premier point de vue, à partir des "catégories de sources qui les ont produits ou dont ils émanent", catégorisation que l'on peut alors qualifier d'*extensionnelle* (à partir des objets), et qui est liée aux processus d'identifications des sources, telles les classifications des odeurs (par ex. odeur de pomme) [Castellengo et Dubois, 2005]. Cette catégorisation fait référence à la notion de *timbre-identité* que nous avons définie plus haut. D'autre part, les sons peuvent être catégorisés "comme sons, et alors être décrits de manière *intensionnelle* (en compréhension) en fonction des propriétés qui les caractérisent, qui donc les qualifient telles les couleurs (le bleu du ciel)." (ibidem). Cette catégorisation est en rapport avec la notion de *timbre-qualité* décrite précédemment. D'autres auteurs parlent encore de *modes de perception source et interprétatif* (*source and interpretative modes of perception*) [Hajda et al., 1997].

Castellengo et Dubois ont mené une expérience pédagogique²⁸ afin de mieux comprendre comment les auditeurs emploient ces deux facettes de l'écoute du timbre, et ce en fonction du degré d'expertise auditive et des acquis. L'une des questions à l'origine de l'étude était de savoir si l'identification des sources prévalait sur leur qualification. L'expérience est simple : faire écouter

²⁷L'exemple d'*el cant del boier*, de la section 1.1.4 pourra être pris comme l'exception qui confirme la règle !

²⁸Le cadre de l'expérience était celui d'un cours, les participants de l'expérience étant également élèves.

une séquence musicale en ayant inversé le sens de lecture.

Il est connu que par un tel procédé l'aspect identitaire du timbre est bouleversé (cf. les Exemples Sonores n^{os} 24 et 25, qui sont les versions d'une même séquence musicale jouée respectivement en sens de lecture normal, et en sens inverse). Ceci permet de prendre conscience de l'importance du facteur temporel dans la perception du timbre (et prouve que le spectre à lui seul ne détermine pas le timbre, puisqu'une telle transformation ne modifie pas le spectre moyen à long-terme)²⁹.

La tâche des participants, issus de trois populations différentes (élèves de conservatoire, de master d'acoustique, et ingénieur) consistait à qualifier le(s) son qu'ils avaient entendu(s). Il s'avère que malgré la nature de la tâche, les auditeurs experts en analyse sonore musicale (ceux du conservatoire) ont plutôt favorisé une démarche de reconnaissance de l'instrument (par ex. "instruments à cordes frottées"). Les étudiants habitués aux technologies du traitement du signal (ceux du master d'acoustique) ont réussi à identifier la nature de la transformation de manière plus fréquente que les autres participants, employant même parfois des termes techniques (par ex. "reverse effet sur une pièce musicale"). Enfin, c'est le groupe dont l'éducation musicale est la moins poussée (les élèves ingénieurs) qui quelque part a le mieux répondu à la tâche, en fournissant "des qualifications renvoyant tant à une description "naturelle" du son qu'à des effets produits exprimés sous formes d'adjectifs déverbaux" (par ex. "haché"). **Force est de conclure que l'écoute du timbre musical relève alternativement des deux aspects - identité et qualité - évoqués précédemment, et que ces deux processus sont mis en oeuvre de manière différente selon la culture et les connaissances des sujets (éducation musicale, scientifique, technique, etc.).**

Les résultats de cette expérience amènent également une preuve de ce que nous évoquions de manière intuitive à la section 1.2.3 au sujet des conceptions "intrinsèques" ou au contraire "extrinsèques" de la musique. Nous soutenions en effet le caractère versatile de l'écoute, ressortant tant à un niveau individuel que dans la confrontation des écoutes de divers individus, et par la même justifions l'incohérence des clivages d'une conceptualisation du sens musical exclusive, prônant soit son caractère intrinsèque contenu dans le son lui-même, soit son caractère extrinsèque, fonctionnant par associations de sentiments et émotions chez l'auditeur. Il nous paraissait logique de conjecturer que ces deux caractères ne pouvaient que se côtoyer chez l'auditeur, et que leur répartition devait singulièrement dépendre des goûts et de l'éducation musicale. L'analyse sémantique des différentes formes lexicales utilisées par les participants dans l'expérience sur le timbre de Castellengo et Dubois ont permis aux auteurs de recenser trois différents types de qualifications selon qu'elles concernent :

- les propriétés "intrinsèques" du son (discontinu, clair, ...);
- son mode de production (un peu accéléré, modifié, ...);
- les effets produits (endormant) ou les évocations suscitées (triste)" (ibidem).

La grande richesse de la perception auditive est donc de pouvoir jongler entre ces différentes facultés d'identification (par la recherche d'indices causaux, cf. figure 1.12) ou de qualification des effets sonores, de considérer le son pour lui-même (niveau intrinsèque), ou par ce qu'il évoque et le sens qu'il peut prendre (niveau extrinsèque).

²⁹La modification ne se limite d'ailleurs pas qu'à modifier le timbre. De nombreux paramètres sont aussi perturbés : dynamique, hauteurs, durées, etc.

2.1.6 Les définitions standard du timbre en psychoacoustique

Nous aurions pu commencer ce chapitre en présentant les définitions classiques du timbre utilisée en psychoacoustique, mais nous ne l'avons pas fait jugeant plus riche l'approche de Schaeffer sur le plan musical, approche qui reste, certes, très conceptuelle. En effet, il n'existe pas encore à ce jour de définition rigoureuse du timbre le décrivant de façon univoque. Le timbre reste "défini" par ce qu'il n'est pas et non par ce qu'il est.

Les définitions de l'ANSI et de l'AFNOR L'Association américaine de normalisation (ANSI) définit le timbre comme suit :

"Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar ; Timbre depends primarily upon the spectrum of the stimulus, but it also depends on the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus." [ANSI, 1960],

(traduction : le timbre est l'attribut de la sensation auditive qui permet à l'auditeur de juger que deux sons présentés de manière similaire et ayant les mêmes sonies³⁰ et hauteurs sont dissimilaires. Le timbre dépend principalement du spectre du stimulus, mais il dépend également de la forme d'onde, de la pression acoustique, de la localisation fréquentielle du spectre, et des caractéristiques temporelles du stimulus.).

La définition proposée par l'Association Française de Normalisation (AFNOR) est très semblable :

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

Les ambiguïtés de ces définitions Ces définitions sont controversées dans la littérature. Elles présentent en effet plusieurs ambiguïtés. Premièrement, il nous semble en toute logique qu'il n'y ait pas qu'un seul attribut (ou caractère) de la sensation auditive qui permette de discriminer deux sons de mêmes hauteur, sonie et intensité. En effet, le son lui-même, pris dans sa globalité (tous paramètres confondus), peut induire une sensation auditive qui rend possible sa distinction avec un autre son. Les définitions ne sont donc pas univoques.

D'autre part, les définitions reposent sur une condition suffisante mais non nécessaire. Cette condition suffisante peut se formuler ainsi : *si deux sons ont les mêmes hauteurs, sonies et durées et qu'ils sont perçus différemment alors ce qui les distinguent sont leurs timbres*. Notons qu'il est également possible de distinguer les timbres de deux sons ayant des hauteurs, intensités et durées différentes. Il est en effet rare que les sons instrumentaux naturels possèdent des sonies et des hauteurs identiques ; nous sommes pourtant capables de les discriminer d'après leurs timbres (si tel n'était pas le cas, nous ne pourrions pas distinguer les différents instruments des morceaux de musique que nous écoutons). De plus, il nous est également possible de percevoir des différences de timbre pour des sons qui n'ont pas de hauteur déterminée (par ex. sons de cloches et de gongs, bruits de pas dans la neige et dans les graviers, etc.).

³⁰La sonie correspond à la sensation d'intensité perçue. Nous veillerons à ne pas la confondre avec l'intensité, prise au sens physique, ou la dynamique, terme employé en musique pour référer aux variations d'intensité.

Ce qui est complexe avec le timbre, c'est qu'il dépend des attributs du son dont on essaye de l'affranchir afin de le définir (hauteur et sonie). Marozeau *et al.* ont en effet montré que la fréquence fondamentale³¹, influe sur la perception du timbre (notamment la sensation de brillance³²) [Marozeau, 2004; Marozeau et de Cheveigné, 2007]. Au delà d'un certain niveau, des modifications de sonie peuvent également engendrer des modifications du timbre perçu [Fletcher, 1934]. Il est intéressant de remarquer que des relations de covariances entre timbre, sonie et hauteur existent souvent de manière naturelle dans les sons instrumentaux. Mentionnons à ce sujet, le *crescendo* de clarinette, lors duquel la richesse spectrale du son augmente, à mesure que l'intensité augmente, et le timbre cuivré de la trompette, caractéristique des sons intenses de l'instrument [Risset et Mathews, 1969]. Outre son utilisation dans l'expression musicale, le *vibrato*³³ est également une signature acoustique très caractéristique de certains instruments (par ex. la voix et le violon). Ne serait-ce pas alors dans ces subtiles variations de hauteur que se définissent certains timbres de voix ? Pourquoi chercher alors à décorréler la hauteur de la notion de timbre ? Face à ces corrélations, ce n'est peut-être pas par hasard qu'un compositeur tel que Schönberg considère que la hauteur n'est qu'une dimension de la couleur sonore (cf. la citation mentionnée à la section 1.1).

Les descriptions des corrélats physiques de l'attribut perceptuel de timbre restent vagues dans les définitions (par ex. "Le timbre dépend principalement du spectre", "Physiquement, le timbre est lié à la composition spectrale d'un son et à son évolution."). Il n'est pas fait mention de manière explicite de la structure harmonique ou inharmonique des sons, qui est pourtant, comme nous l'avons vu à la section 1.1.3.3, déterminante pour le timbre (cf. Exemples Sonores *n*^{os} 6 dont la structure est inharmonique, et 7 dont la structure est harmonique). Il est également connu que l'inharmonicité des cordes métalliques de piano, dont la raideur est importante, explique l'une des spécificités du timbre de l'instrument.

Bregman a relevé d'autres inconvénients de la définition de l'ANSI [Bregman, 1994, p. 92]. Le fait que les sons doivent être présentés de "manière similaire" exclut des situations pourtant importantes. Par exemple, le fait de présenter un son disposant d'un certain timbre à l'oreille gauche, et présenter un son doté d'un autre timbre à l'oreille droite, nous empêche-t-il de juger leur dissemblance ? Plus ennuyeux encore, si l'on s'intéresse à des sons itératifs. Des sons répétés 20 fois par secondes présentent bien une différence de timbre par rapport à des sons répétés 40 fois par secondes, pourtant si l'on s'en tient rigoureusement à ce que propose la définition, ils ne seraient pas présentés à l'auditeur de manière identique. Bregman va même jusqu'à dire, sûrement un peu par provocation face à ces illogismes, que la définition de l'ANSI devrait être : "We do not know how to define timbre, but it is not loudness and it is not pitch." (ibidem, p. 93) (traduction : Nous ne savons pas comment définir le timbre, mais ce n'est pas la sonie et ce n'est pas la hauteur.) ! Et de proposer que tant que le timbre ne sera pas bien défini, il serait peut-être préférable de se contenter de parler des caractéristiques du son qui agissent sur l'intégration séquentielle³⁴.

Comme nous venons de le voir, la complexe notion de timbre reste toujours mal comprise et les définitions standard incomplètes. Le fait de considérer que le timbre n'est ni hauteur, ni sonie, ni durée est probablement trop restrictif. Cependant, explorer le timbre à hauteur, sonie et durée fixés

³¹En général la fréquence fondamentale est fortement corrélée à la sensation de hauteur perçue. Nous veillerons cependant à ne pas confondre fréquence fondamentale et hauteur, puisque cette dernière dépend également d'autres aspects de la distribution spectrale (cf. par ex. [Terhardt *et al.*, 1982b]).

³²La brillance est corrélée au Centre de Gravité Spectral comme nous le verrons à la section 2.2.5

³³Le vibrato correspond à une modulation de fréquence des composantes spectrales, s'opérant à une fréquence faible (environ comprise entre 2 et 7 Hz).

³⁴L'intégration séquentielle rend compte du fait de séparer ou au contraire fusionner des flux auditifs. Nous avons vu que le timbre ne prend forme que lorsque notre système perceptif procède à la fusion des composants du son (cf. section 1.1.3.1). Nous décrirons à la section 2.2.5 des travaux de Wessel ayant permis de montrer que la brillance peut influencer l'intégration séquentielle.

n'en reste pas moins un bon point de départ. Nous verrons dans la section 2.2 que de nombreuses études ont permis de mettre en lumière des paramètres psychoacoustiques qui semblent être de bons *prédicteurs* de notre perception du timbre.

2.1.7 L'aspect morphologique temporel du timbre

Il est un autre aspect occulté par les définitions standard et essentiel à l'égard de la musique, celui des possibles évolutions du timbre au cours du temps. Il s'agit pourtant d'un caractère morphologique important. En effet, le timbre n'y apparaît que comme un caractère global du son (qui fait que l'on est capable de discriminer, par exemple, les sons d'une trompette et d'un luth, malgré des hauteurs, sonies et durées fixées). Il nous paraît également important de considérer le timbre selon un autre horizon temporel. En effet, nous avons mené une expérience qui met en évidence qu'au sein d'une même note d'instrument de musique des différences de timbre audibles peuvent se produire³⁵. Nous nous sommes pour ce faire reposés sur un modèle d'analyse-synthèse additif et une transformation sonore permettant de figer dans le temps le Centre de Gravité Spectral d'un son tout en gardant inchangés les autres aspects du son (hauteur, intensité, durée)³⁶. Nous avons appliqué la transformation à un son naturel de clarinette (cf. [Exemple Sonore n° 26](#)) afin de synthétiser deux versions modifiées du son originel pour lesquelles le Centre de Gravité Spectral a respectivement été figé suivant :

- sa valeur "initiale" après la phase d'attaque (cf. [Exemple sonore n° 27](#)), et
- sa valeur "finale" juste avant l'extinction (cf. [Exemple Sonore n° 28](#)).

De par ce procédé, nous rallongeons donc dans le temps un phénomène qui détenait un caractère instantané. L'oreille a ainsi davantage de "temps" pour entendre les différences de timbre ayant lieu entre le début et la fin du son originel. L'[Exemple Sonore n° 27](#) a un timbre plutôt "sourd" (son Centre de Gravité Spectral est faible), tandis que l'[Exemple Sonore n° 28](#) révèle un timbre plutôt "brillant". Une écoute attentive de la note permet en effet d'entendre une augmentation de la brillance pendant son déroulement. **Cet exemple nous montre que si fugitives qu'elles soient, des variations de timbre peuvent avoir lieu au sein même d'un seul son instrumental.**

Ces observations nous ont conduits à définir trois concepts :

- **Le Timbre-Qualité Instantané**, abstraction qui réfère à la sensation de timbre-qualité que l'on serait à même de percevoir à un instant donné d'un son, si notre système auditif le permettait ;
- **La Variation Temporelle de Timbre**, qui correspond aux variations temporelles du Timbre-Qualité Instantané ;
- **La Forme Temporelle de Timbre**, qui correspond à l'effet perceptuel résultant des variations temporelles du Timbre-Qualité Instantané.

Les notions de timbre et sonance selon Seashore Il est intéressant de remarquer que Seashore désignait par timbre la qualité sonore instantanée ("Physically the timbre of the tone is a cross section of the tone quality for the moment represented by the duration of one vibration in the sound." [[Seashore, 1938](#), p. 97] (traduction : Physiquement, le timbre d'un son est la coupe transversale de la qualité sonore sur un instant correspondant à la durée d'une vibration dans le son),

³⁵C'est l'aspect *timbre-qualité* qui est ici considéré.

³⁶La description du modèle d'analyse-synthèse additif est donnée à la section 2.2.5.5. La transformation est décrite au chapitre 6. La définition du Centre de Gravité Spectral est donnée à la section 2.2.5.

et ne lui conférait que des attributs spectraux³⁷ : “Timbre is that characteristic of a tone which depends upon its harmonic structure as modified by absolute pitch and total intensity. The harmonic structure is expressed in terms of the number, distribution, and relative intensity of its partials.” (ibidem) (traduction : Le timbre correspond aux caractéristiques d’un son qui dépendent de sa structure harmonique telle que modifiée par la hauteur absolue et l’intensité totale. La structure harmonique se traduit en termes du nombre, de la distribution et de l’intensité relative de ses partiels.). Et il désignait par *sonance*, terme emprunté à Metfessel [Metfessel, 1926], les changements de la qualité sonore au cours du temps, dont certains sont induits par le timbre (“sonance is that aspect of tone quality which results from fluctuations in pitch, intensity, time, and timbre within a tone.” [Seashore, 1938, p. 108] (traduction : la sonance correspond à l’aspect de la qualité sonore déterminé par les fluctuations de hauteur, d’intensité, de temps et de timbre au sein d’un son.”).

Nous nous intéressons dans ce travail à l’interprétation musicale où l’évolution temporelle est primordiale. En effet, généralement la musique ne peut être conçue que comme un processus temporel constitué de sons/notes, de phrases, de mouvements (etc.), qui façonnent la structure du discours musical. Nous avons cherché si des liens existent entre la façon dont l’interprète exécute un morceau et les Formes Temporelles de Timbre³⁸ (cf. chapitre 5). Du point de vue de l’auditeur, nous avons étudié comment certaines Formes Temporelles de Timbre influencent la perception musicale (cf. chapitre 6).

2.2 Les corrélations entre acoustique et perception du timbre

Jusqu’ici nous avons beaucoup parlé des différentes acceptions du timbre suivant ses conditions d’emploi ou d’interprétation, mais moins de ce qui permet de le décrire. Les mesures acoustiques offrent des informations sur la structure du son qu’il est judicieux de mettre en relation avec notre façon de le percevoir. C’est donc au travers de *corrélations* entre les domaines de la physique et de la perception que pourront se dégager des descriptions morphologiques et typologiques du timbre. Il s’agira alors de ne pas confondre psychologie et “réalité” physique (par ex. la perception du temps et la mesure de la durée sont deux choses différentes), mais de les mettre en correspondance, comme le souligne Schaeffer : “ce qui est sur la bande (A) produit un effet physique (B), ainsi qu’un effet sonore ou musical (C) : on peut dire au mieux que l’effet musical C est *en corrélation* avec l’effet physique (B), puisqu’ils ont tous deux la même origine A. La perception et la mesure apparaissent sur le même plan, et non l’une subordonnée à l’autre ; on ne peut prétendre que B explique C ; tout au plus l’accompagne t-il.” [Schaeffer, 1966, p. 146].

Nous commencerons par rappeler quelques résultats fondamentaux sur la structure du timbre et sa perception, puis nous présenterons les principales découvertes qui ont permis de proposer une représentation perceptive multidimensionnelle du timbre (aussi appelée *espace de timbre*). Le but n’étant pas ici de dresser une liste exhaustive des travaux effectués sur le timbre, nous invitons le lecteur désireux d’obtenir de plus amples informations, à consulter l’article général de Hajda et al. [Hajda et al., 1997], qui offre une précieuse revue des études psychoacoustiques menées sur le timbre jusqu’en 1997, ainsi que de leurs aspects méthodologiques, et celui de Risset et Wessel sur l’exploration du timbre par analyse et synthèse, disponible en anglais [Risset et Wessel, 1999], mais également en français [Risset et Wessel, 1991]³⁹.

³⁷ cf. conception classique du timbre au paragraphe, 2.2.1.

³⁸ Pour des raisons de simplicité, nous parlerons aussi de *variations de timbre*.

³⁹ La version anglaise de l’article est cependant plus complète puisqu’elle a été actualisée dans l’édition de 1999 du livre *Psychology of music*.

2.2.1 La conception classique du timbre

La vision d'Helmholtz Selon la conception classique, le timbre ne serait caractérisé que par le spectre de fréquence à long terme, c'est à dire, dans le cas de sons quasi périodiques, les seuls considérés comme musicaux par Helmholtz (1877), de la proportion relative des différents harmoniques [Helmholtz, 1954]. Selon lui, seuls les aspects stationnaires du son, associés au régime permanent des instruments, déterminaient la qualité sonore musicale. Comme nous l'avons vu au paragraphe 2.1.7, Seashore assimilait aussi timbre et spectre, cependant et paradoxalement, il ne niait pas mais au contraire défendait le fait que l'évolution du spectre au cours du temps participait également au timbre des sons musicaux (aspect qu'il désignait par *sonance de timbre*).

Or, il est désormais un fait avéré que malgré son implication sur le timbre, le spectre à lui seul ne détermine pas tous les aspects de ce dernier. Une expérience simple pour s'en convaincre consiste à renverser le sens de lecture d'un son ou d'une séquence musicale (cf. l'étude de Castellengo et Dubois mentionnée à la section 2.1.5, et les séquences musicales jouées en sens normal et inverse dans les Exemples Sonores *n^{os}* 24 et 25) : la qualité sonore en est totalement perturbée, au point que l'identification de la(es) source(s) sonore(s) ne soit souvent plus possible, et ce bien que le spectre de Fourier à long-terme, lui, soit conservé. **L'expérience montre ainsi la dépendance du timbre envers des facteurs temporels.**

Une autre observation montre qu'il n'y a pas équivalence entre timbre et spectre, puisque lorsque celui-ci est altéré par des distorsions causées par le moyen de transmission du son, nous sommes toujours à même de reconnaître le timbre d'un instrument. Ainsi on reconnaîtra le piano même au travers d'un téléphone, dont la très faible bande passante (qui ne s'étend que d'environ 300 à 3400 Hz) réduit drastiquement l'information contenue dans le spectre originel de l'instrument. Dans de tels cas, nous ne pourrions pas non plus nier que si ce qui se rapporte à l'identité instrumentale n'est pas bouleversé, en revanche la qualité du timbre aura changé. Comme le souligne Risset, **"Le spectre est bien un paramètre sensible, mais on ne peut l'assimiler au timbre, signature auditive de l'instrument. Avant que la haute fidélité ne soit la règle, on pouvait identifier les instruments de musique enregistrés, malgré d'importantes distorsions linéaires bouleversant les spectres : il faut bien que l'identité des sources tienne à des traits plus robustes."** [Risset, 1991, p. 245].

2.2.2 La théorie des formants

L'assimilation du timbre au spectre induisait à croire que la nature du timbre dépendait uniquement du dosage des partiels (*overtones* en anglais), constituants structurels des sons musicaux. Suivant cet entendement, une transposition (*pitch shifting* en anglais) sans modification de timbre pourrait être réalisée par simple translation spectrale, en conservant la répartition d'énergie entre les partiels du son d'origine.

Mise en évidence de l'importance des formants Montrons par une expérience simple que le timbre n'est pas uniquement déterminé par la répartition d'énergie des composantes spectrales. Tentons donc de transposer une séquence musicale par translation fréquentielle du spectre, en conservant les rapports d'énergie des différents composantes spectrales des sons originels. *Le vocodeur de phase* permet d'effectuer ces traitements sans modifier les durées des sons considérés (cf. par ex. [Portnoff, 1976])⁴⁰.

⁴⁰Une transposition du grave à l'aigu, réalisée par une accélération de la vitesse de lecture des échantillons sonores, conduit intrinsèquement à raccourcir la durée de la séquence initiale. À l'inverse, une transposition de l'aigu au grave, effectuée par ralentissement de la vitesse de lecture, entraîne un allongement de la durée.

Le vocodeur de phase, dont l'origine remonte aux années 70, est issu des travaux sur la synthèse vocale de Homer Dudley, menés dans les années 40 aux laboratoires Bell. Ce dernier a inventé le "Vocoder" (Voice Operated reCORDER), premier système de synthèse vocale. **Le vocodeur de phase, comme d'une manière générale les techniques d'analyse-synthèse, repose sur trois étapes principales, l'analyse, consistant à obtenir une représentation d'un son, la modification éventuelle de cette représentation, et la synthèse, pour obtenir de nouveau un son, qui pourra soit reproduire fidèlement le son analysé (synthèse imitative), soit être une version modifiée du son originel.** La représentation utilisée peut être une représentation temps-fréquence, comme celle obtenue grâce à la Transformée de Fourier à Court Terme (TFCT), ou son pendant discret, la Transformée de Fourier Discrète à Court Terme (TFDCT).

Nous avons utilisé un vocodeur de phase afin de réaliser l'expérience de transposition sur une séquence musicale chantée (la séquence utilisée correspond à l'[Exemple Sonore n° 20](#), extrait du morceau *Tom's dinner* de la chanteuse américaine Suzanne Vega)⁴¹. L'[Exemple Sonore n° 21](#) correspond à une transposition vers le grave, d'un facteur⁴² égal à 0.5 (octave inférieur), tandis que les [Exemples Sonores n°s 22 et 23](#) correspondent à des transpositions vers l'aigu, de facteurs respectivement égaux à 1.2, et 2 (octave supérieur). Bien que nous ne disposions pas de séquences naturelles produites par la chanteuse dans ces différentes tessitures comme éléments de comparaison (si tant est qu'elles aient été réalisables), il est difficilement imaginable qu'elles aient pu ressembler aux séquences transposées artificiellement. Par nos transformations, c'est non seulement le timbre de la voix de la chanteuse qui est perturbé, mais son identité même. La transposition à l'octave inférieur transforme radicalement la voix originale de femme en une voix d'homme à la musicalité incertaine, tandis que les transpositions vers l'aigu lui confèrent un caractère comique, non sans rappeler ces voix que nous avons coutumes d'entendre dans les dessins animés⁴³. Cette expérience nous montre qu'une transposition de hauteur par simple translation fréquentielle de la distribution spectrale ne conserve pas le timbre de la voix. Ceci est lié au fonctionnement acoustique de l'appareil vocal, dont nous avons discuté à la section 1.1.4. En effet, le fonctionnement de l'appareil vocal repose sur la vibration des cordes vocales, qui détermine la fréquence fondamentale du son, et la phonation, de laquelle découle des *zones formantiques spectrales, pratiquement stables lors de la production de différentes hauteurs*. Les formants du spectre définissent une *enveloppe spectrale* déterminante pour le timbre. La réalisation d'une transposition de voix fidèle à une identité timbrale requiert la conservation, non pas des rapports d'énergie entre les composantes spectrales, mais celle d'une enveloppe spectrale fixe, sorte de "signature" acoustique de l'articulation du conduit vocal du chanteur. Il est intéressant de remarquer que de tels types de transformations ont déjà été élaborés avec succès (cf. par ex. [\[Pallone, 2003\]](#), où l'auteur introduit le concept de "*transposition fréquentielle sous contraintes perceptives*", contraintes destinées notamment à la conservation du timbre).

Les travaux du phonéticien Slawson ont en effet démontré que l'identité timbrale des sons voisés⁴⁴, relevait d'une enveloppe spectrale fixe, et donc de formants fixes : "The popular notion that a particular timbre depends upon the presence of certain overtones (if that notion is interpreted as the "relative pitch" theory of timbre) is seen [...] to lead not to invariance but to large differences in musical timbre with changes in fundamental frequency. The "fixed pitch" or formant theory of

⁴¹ Les séquences transformées ont été générées à l'aide de fonctions *Matlab*, associées au livre *DAFx : Digital Audio Effects* [\[Zölzer, 2002a\]](#), et disponibles à l'adresse indiquée dans la référence [\[Zölzer, 2002b\]](#).

⁴² Le facteur de transposition correspond au rapport entre les fréquences modifiées et les fréquences originales.

⁴³ Ceci met en évidence l'une des fonctionnalités offertes par des systèmes tels que le vocodeur de phase, celle de produire des effets audionumériques, dont les intérêts musicaux ou audiovisuels relèveront alors du domaine de la création artistique.

⁴⁴ Le voisement est une qualité de certains sons de la parole ou du chant. Toutes les voyelles sont voisées. En revanche, certaines consonnes sont voisées (par ex. [b], [d], [g]), tandis que d'autres ne le sont pas (par ex. [p], [t], [k]). Il s'agit dans ce dernier cas de sons bruités.

timbre is seen in those same results to give much better predictions of the minimum differences in musical timbre with changes in fundamental frequency. The results [...] suggest that the formant theory may have to be modified slightly. A precise determination of minimum differences in musical timbre may require a small shift of the lower resonances, or possibly the whole spectrum envelope, when the fundamental frequency changes drastically." [Slawson, 1968, p. 100] (traduction : "La conception classique qu'un timbre spécifique relève de la présence de certains partiels (si cette notion est interprétée comme la théorie du timbre selon la "hauteur relative") implique [...] non pas une invariance mais de larges différences de timbre musical lors de changements de la fréquence fondamentale. La théorie du timbre selon la "hauteur fixe", ou théorie des formants, présente en regard des mêmes résultats de bien meilleures prédictions des différences minimales de timbre musical, lors de changements de fréquence fondamentale. Les résultats [...] suggèrent que la théorie des formants doit être légèrement modifiée. Une détermination précise des différences minimales de timbre musical semble devoir nécessiter une petite translation des résonances les plus basses, ou possiblement de toute l'enveloppe spectrale, lorsque la fréquence fondamentale change drastiquement.").

D'une manière générale, les structures formantiques des sons d'instruments de musique sont liées aux résonances engendrées par un ou des éléments de la chaîne de production du son (par ex. caisse de résonance de la guitare, embouchure de la trompette, conduit vocal pour la voix). Il paraît légitime que du point de vue de la perception, l'oreille utilise ces structures pour discriminer certains sons (afin de différencier la voix d'un adulte de celle d'un enfant par exemple). Les résultats de Slawson (cf. citation ci-dessus) montrent également que la théorie des formants a des limites. Lorsque les transpositions sont importantes (facteur 1.5), l'identité des voyelles est mieux préservée lorsque les formants sont déplacés d'un facteur de 10 à 12 %, relatif au changement de fréquence fondamentale (ibidem). Nous avons effectivement vu au paragraphe 1.1.4, d'après les travaux de Sundberg, que d'un point de vue causal (forme du conduit vocal), la position des formants n'était pas la même pour des voyelles chantées dans différents registres.

2.2.3 Relations de phase entre les composantes harmoniques

Si la description du timbre ne peut se limiter au spectre, elle ne peut pas non plus se limiter à la forme d'onde. Plomp a en effet montré que des ondes périodiques qui présentent des relations de phase différentes entre les composantes harmoniques ne peuvent pas être discriminées par le système auditif [Plomp, 1976] (constatation d'ailleurs déjà réalisée par Ohm (1789-1854) [Risset, 2003]). Faisons-en l'expérience. Nous avons généré deux signaux, un premier de forme d'onde "carrée", ne contenant que des harmoniques impairs (cf. [Exemple Sonore n° 29](#)), et un second possédant la même structure harmonique que le premier, à ceci près que les harmoniques adjacents sont en opposition de phase (cf. [Exemple Sonore n° 30](#)). Afin d'obtenir la même perception d'intensité, l'amplitude du signal carré a été réduite d'un facteur 2 relativement à celle du second signal. En effet, le fait d'additionner des composantes hors phase induit une diminution de l'énergie par rapport au premier cas, où toutes les composantes harmoniques sont en phase. Comme nous pouvons le constater d'après la figure 2.8, les formes d'ondes des deux signaux sont totalement différentes. Pourtant, les deux signaux générées ont le même timbre ; **l'oreille semble donc insensible aux relations de phase entre les composantes harmoniques dans le cas de sons strictement périodiques**. Le point de vue "écologique" ou "fonctionnaliste" prôné par Gibson [Gibson, 1966] permet de fournir une justification plausible à cet aspect de l'audition : **"L'oreille est sourde aux relations de phase, car ces relations sont complètement perturbées dans tout environnement un peu réverbérant ; elle est au contraire extrêmement sensible aux aspects de fréquences, ce paramètre physique restant généralement très stable lors de la propagation du son entre source et auditeur."** [Risset, 2003].

Notons que dans le cas de sons musicaux harmoniques, les relations de phase entre les composantes peuvent avoir une influence sur la perception du timbre (cf. par ex. [Dubnov et Rodet, 1997] et [Andersen et Jensen, 2004]).

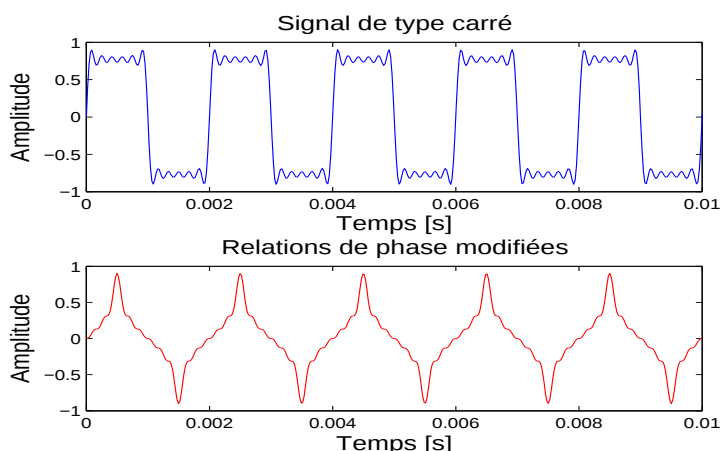


FIG. 2.8 – Insensibilité du système auditif aux relations de phase entre les composantes harmoniques dans le cas de sons périodiques. Les deux formes d’onde représentées sont périodiques de même fréquence fondamentale (500 Hz) et sont le résultats de l’addition de sinusoides dont les fréquences suivent celles d’une même série d’harmoniques impairs (1, 3, 5, ..., 11). Pour la première forme d’onde (signal du haut), les composantes harmoniques sont toutes en phase. En revanche, pour la seconde (signal du bas), les composantes harmoniques adjacentes (i.e. 1 et 3, 3 et 5, etc.) sont en opposition de phase. Les formes d’onde sont différentes, pourtant les signaux sonores associés (cf. respectivement les Exemple Sonores n^{os} 29 et 30) sont identiques à l’oreille (toutefois, après ajustement de la sonie).

Alors que des formes d’ondes différentes peuvent être associées à un même timbre, on peut s’interroger sur la “contraposée” : est-ce que des timbres différents peuvent être associés à des formes d’ondes identiques⁴⁵. Nous dirons que généralement pour un même individu et selon le même contexte, deux sons de formes d’ondes identiques (ou un même son présenté deux fois) procurent la même sensation auditive (et a fortiori la même sensation de timbre). En revanche, si le contexte psychologique change, il nous semble qu’un même son peut être entendu (dans le sens de compris) différemment (en raison d’aspects qui, entre autres, pourraient être liés au timbre). Les propos que nous avons tenus sur l’écoute musicale (par ex. Mozart pourra tantôt mettre l’auditeur dans un état d’éveil, tantôt l’ennuyer, voire l’agacer, selon le contexte psychologique), le fait que nous sommes capables de mémoriser et de comprendre des structures musicales de manière implicite et donc de les aborder de manières différentes lors d’écoutes ultérieures, tendent à l’affirmer. Mais alors, nous nous rendons compte que la source des différences interviendrait davantage à un niveau cognitif, au travers d’évocations suscitées par le son, qu’à un niveau perceptif strict.

2.2.4 Le timbre, un attribut multidimensionnel du son

Olson (1967) [Olson, 1967] et Schaeffer (1966) [Schaeffer, 1966] ont montré l’importance de l’attaque, du profil dynamique et de la richesse harmonique. Ainsi Schaeffer transformait-il un son de piano en un son de flûte et réciproquement un son de flûte en un son de piano par des manipulations portant uniquement sur la dynamique des sons [Schaeffer, 2005, p. 42]. Et de conclure :

⁴⁵Nous ne parlons pas ici de différences inter-individuelles. Nous savons que de telles différences existent par exemple pour la sensation d’intensité, puisque les seuils d’audition sont variables selon les individus (voire même entre les deux oreilles d’un même individu), et leur âge [Zwicker et Fastl, 1990]. Les seuils différentiels de timbre (brillance) obtenus dans [Kendall et Carterette, 1996] s’avèrent plus élevés pour les non-musiciens que les musiciens.

“Voici à présent que la dynamique (que les musiciens appellent nuances et les physiciens niveau) risque de devenir l’un des facteurs du timbre, secret de la matière sonore.” (ibidem). Ces résultats prouvent le caractère multidimensionnel du timbre. Il reste encore à le préciser plus finement.

En parallèle aux démarches d’analyse par la synthèse (cf. par ex. [Risset et Mathews, 1969; Moorer, 1977]), et expérimentales (cf. par ex. [Schaeffer, 1966]), une méthodologie d’analyse de la représentation perceptive du timbre musical a été développée dans les années 70 (cf. par ex. [Grey, 1977; Wessel, 1978; McAdams, 1999]). Nous en décrivons les principes fondamentaux dans la prochaine section. La méthodologie repose sur des campagnes de mesures perceptives, au cours desquelles les jugements de participants sont collectés (généralement des jugements de dissimilarité entre paires de sons), puis d’analyses statistiques des données obtenues. En effet, puisque le timbre est un attribut multidimensionnel du son, l’analyse de données perceptives à partir d’analyses statistiques multivariées (cf. par ex. [Dillon et Goldstein, 1984]) semble particulièrement adaptée pour explorer les aspects complexes de la perception du timbre. S’inscriront parmi les techniques statistiques les plus utilisées, l’analyse multidimensionnelle de proximité [Shepard, 1962b,a; Kruskal, 1964a,b], souvent intitulée analyse MDS, en référence à l’anglais *MultiDimensional Scaling*, et la classification ascendante hiérarchique (*hierarchical cluster analysis* en anglais) [Sneath, 1957; Ward, 1963], que nous désignerons aussi dans ce document par *analyse en clusters*. D’un point de vue général, il est remarquable de constater le lien étroit entre les avancées des recherches sur le son, celles des statistiques et l’avènement de l’ordinateur rendant possible l’exécution d’algorithmes complexes de manière très rapide.

2.2.4.1 Méthodologie d’analyse de la représentation perceptive du timbre

Une description complète de la méthodologie pourra être trouvée dans les articles fondamentaux de Grey [Grey, 1977], Wessel [Wessel, 1978], et McAdams [McAdams, 1999]. La figure 2.9 issue de [McAdams, 1999] résume la démarche d’obtention d’un espace de timbre et de son interprétation psychophysique. Nous en rappellerons ici les principes :

2.2.4.1.1 Sélection ou élaboration des stimuli Il convient dans un premier temps d’établir le corpus sonore utilisé lors du test perceptif. Il pourra s’agir **soit de sons issus d’instruments naturels, soit de sons synthétiques**. Grey prône l’utilisation de sons de synthèse dans l’exploration du timbre musical pour plusieurs raisons. D’une part, certains modèles de synthèse rendent possibles la modification des amplitudes et fréquences instantanées des composantes spectrales, ainsi que des durées. Ceci permet notamment d’égaler de manière indépendante la hauteur, la sonie et la durée des sons générés, afin que ces propriétés du son n’influencent pas les jugements des auditeurs, et que les différences entre les stimuli relèvent exclusivement du timbre, conformément à la définition standard (cf. section 2.1.6). D’autre part, des simplifications n’altérant que peu la reconnaissance des timbres instrumentaux peuvent être effectuées (par ex. approximation des enveloppes par segments de droite [Moorer, 1977]), ce qui selon Grey peut faciliter par la suite l’interprétation des résultats.

Le nombre de stimuli constituant le corpus sonore est également à déterminer, tout en gardant à l’esprit qu’afin que les techniques d’analyses statistiques puissent fournir des représentations significatives, un nombre minimal de sons est requis. Il n’existe pas de règles théoriques permettant de définir le nombre minimal de sons à utiliser. Wessel nous apprend que dans la plupart des recherches psychologiques reposant sur l’analyse multidimensionnelle de proximité, un nombre de 10 objets (des sons dans notre cas) semblent suffisants pour deux dimensions, et 15 pour trois dimensions [Wessel, 1978]. D’autre part, le nombre de stimuli devra rester suffisamment petit, de sorte que la durée totale de l’expérience perceptive reste raisonnable pour les participants, dont le maintien de la concentration tout au long du test est primordial.

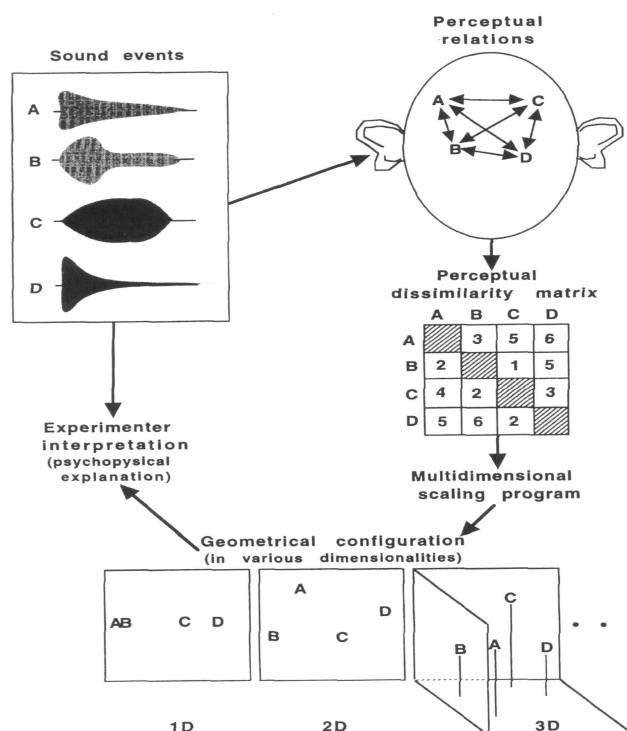


FIG. 2.9 – Méthodologie d'étude de la structure perceptive sous-jacente à la perception du timbre : de l'espace de représentation multidimensionnel à son interprétation psychophysique. Figure de McAdams, issue de [McAdams, 1999, p. 87].

L'ambitus de variations entre les stimuli est aussi un facteur important. Lorsque la plage de variation est trop grande, des différences subtiles entre les sons pourront passer inaperçues tant au niveau de l'auditeur qui ne les mettra pas en évidence dans ses jugements, focalisant son attention sur les différences dominantes, qu'au niveau de la méthode statistique qui pourra alors converger vers des solutions dégénérées [Wessel, 1978].

2.2.4.1.2 Sélection des participants Le choix des participants sera réalisé en fonction de leurs audition, sexe, culture, âge, expérience musicale, etc. Tantôt l'expérimentateur souhaitera comparer les jugements de musiciens et non musiciens, tantôt la nature de la tâche requerra une écoute spécialisée⁴⁶ apte à déceler de subtiles variations musicales, auquel cas il conviendra de sélectionner uniquement des participants ayant des qualifications musicales.

Pour des raisons similaires à celles évoquées ci-dessus, **le nombre de participants** devra être suffisamment grand pour que les analyses statistiques aient un sens. Disposer d'un nombre élevé de participant permet notamment de minimiser les effets d'éventuelles erreurs de mesures (liées à un manque de concentration par exemple). Les expériences menées sur le timbre font souvent état d'une vingtaine de participants, nombre également choisi de manière empirique.

2.2.4.1.3 Test de jugements de dissimilarité (ou de dissemblance) La méthodologie repose dans la plupart des cas sur la collecte de *mesures des dissimilarités* ("distances perceptives") entre paires de stimuli différents. Les mesures s'opèrent généralement sur une échelle allant de "Très

⁴⁶La notion d'écoute spécialisée a été définie à la section 1.3.2

dissemblables" à "Très semblables". En pratique, les mesures de dissimilarités sont associées à des valeurs numériques discrètes (par ex. sur une échelle de 0 à 1 par pas d'un centième). Il convient de signaler à ce sujet que l'information utile n'est pas contenue dans les valeurs absolues des dissimilarités mais dans leurs valeurs relatives. Nous reviendrons sur la notion d'échelle de mesure dans le paragraphe suivant.

2.2.4.1.4 Représentations des jugements de dissimilarité (analyse multidimensionnelle de proximité et classification ascendante hiérarchique) L'analyse multidimensionnelle de proximité (MDS) permet de former une représentation des objets (en l'occurrence les sons) dans un espace géométrique euclidien, de telle sorte que leurs positions relatives dans l'espace reflètent les valeurs des distances perçues. La dimension d'un tel espace se veut être bien inférieure au nombre total de stimuli utilisés afin de clarifier la nature des relations perceptives entre les stimuli. Les références [McAdams, 1999] et [Caclin, 2004] proposent des descriptions des principaux algorithmes d'analyse MDS dans le cadre de l'étude du timbre musical. Mentionnons à ce titre :

- l'algorithme classique MDSCAL défini par Kruskal [Kruskal, 1964a,b]. Il a par exemple été utilisé par Wessel dans [Wessel, 1979] ;
- l'algorithme INDSCAL développé par Carroll et Chang [Carroll et Chang, 1970], qui prend en compte dans le calcul des distances perceptives de possibles différences inter-individuelles (au moyen de poids, relatifs à chaque participant, et chaque dimension). Cet algorithme a par exemple été utilisé par Grey dans [Grey, 1977] ;
- l'algorithme CLASCAL établi par Winsberg et De Soete [Winsberg et De Soete, 1993], qui permet de déterminer des "classes latentes" de sujets en fonction de leurs stratégies de réponses. La méthode a par exemple été utilisée par Caclin *et al.* dans [Caclin *et al.*, 2005] ;
- l'algorithme EXSCAL issu des travaux de Winsberg et Carroll [Winsberg et Carroll, 1988], qui lui ne prend pas en compte les spécificités individuelles, mais en revanche tient compte d'éventuelles spécificités que peuvent présenter certains sons par rapport à d'autres (par ex. la faible présence d'harmoniques pairs pour la clarinette). Un exemple d'utilisation de l'algorithme pourra être trouvé dans [Krumhansl, 1989] ;
- une extension au modèle CLASCAL, qui réunit tant les spécificités pour les sons, que les poids pour les classes de sujets. Cette extension a par exemple été employée par McAdams *et al.* dans [McAdams *et al.*, 1995].

Notons que l'analyse MDS a le même objectif que l'Analyse en Composantes Principales (ACP) [Saporta, 1990, p. 183], i.e. trouver une configuration de N objets dans un espace euclidien de dimension faible, mais les données initiales diffèrent : dans l'ACP, ce sont des variables décrivant les objets qui sont connues, tandis que dans la MDS, les données de départ sont des distances ou des dissimilarités entre les objets. Si l'on dispose d'une véritable distance euclidienne entre les objets, alors la MDS peut être vue comme un cas particulier de l'ACP (et la formule de Torgerson, sur laquelle repose l'ACP, peut être utilisée). Montrons à ce titre pourquoi les distances perceptives entre sons peuvent être considérées comme des dissimilarités. Pour cela, comparons les propriétés des distances, des dissimilarités, et des distances perceptives.

Soit E l'ensemble des S objets (stimuli) à classer. Une distance d est une application de $E \times E$ dans R^+ telle que (cf. [Saporta, 1990, p. 243]) :

$$\begin{cases} (1) d(i, j) = d(j, i) \\ (2) d(i, j) \geq 0 \\ (3) d(i, j) = 0 \Leftrightarrow i = j \\ (4) d(i, j) \leq d(i, k) + d(k, j) \end{cases} \quad (2.1)$$

Analysons quelles propriétés des distances sont vérifiées par les distances perceptives entre stimuli. Pour ce faire, le couple (i, j) désignera une paire de stimuli. Une distance perceptive ne respectera pas nécessairement la propriété (1). Le fait qu'elle ne soit pas vérifiée indique un effet d'ordre de présentation des stimuli. Dans une étude de Grey sur les dissemblances de timbre entre 16 sons d'instruments de l'orchestre, l'auteur a relevé des différences dans les mesures de distances perceptives selon l'ordre de présentation des stimuli [Grey, 1977]. Le coefficient de corrélation de Pearson⁴⁷ global (i.e. moyenné sur tous les participants), calculé entre les paires de demi-matrices de dissimilarité (l'une correspondant à l'ordre (i, j) et l'autre à l'ordre (j, i)) était de 0.905 (et non 1). Ces différences ne se sont cependant pas révélées significatives, et par conséquent, les deux demi-matrices ont été moyennées avant de procéder à l'analyse multidimensionnelle de proximité. Notons qu'il se peut que les différences constatées par Grey soient également dues à un manque de régularité dans les réponses. En effet, il se peut que des différences aient également été constatées pour des mesures de distances perceptives répétées avec un même ordre de présentation des stimuli. Du reste, pour de nombreuses études, l'effet d'ordre s'est avéré non significatif. En pratique, il est courant dans les tests de dissemblance reposant sur la comparaison par paires d'effectuer les mesures de distances perceptives sur les $\frac{S \times (S - 1)}{2}$ combinaisons (sans ordre et sans répétition)⁴⁸ des S stimuli, puis de tirer aléatoirement l'ordre de présentation des stimuli. Ceci présente deux avantages. D'une part, cela permet de réduire considérablement le nombre de comparaisons à effectuer (il y en a $S \times (S - 1)$ lorsque l'effet d'ordre est pris en compte) et donc la durée de l'expérience perceptive. D'autre part, la considération des mesures des distances perceptives moyennes au sein de l'échantillon d'individus permet alors de moyenner l'éventuel effet d'ordre, puis de considérer la propriété (1) comme vraie.

La propriété (2) est respectée par les distances perceptives. En revanche, la propriété (3) ne l'est pas ; deux sons différents peuvent être perçus comme étant identiques. Enfin, les distances perceptives ne se conforment pas non plus à la propriété (4), l'inégalité triangulaire ne prenant pas nécessairement acte en matière de perception.

D'autre part, une dissimilarité se définit au travers des propriétés suivantes :

$$\begin{cases} (1) d(i, j) = d(j, i) \\ (2) d(i, j) \geq 0 \\ (3) d(i, i) = 0 \end{cases} \quad (2.2)$$

Au vu des raisons mentionnées ci-dessus, il est légitime de considérer, à un certain stade de l'analyse, une distance perceptive comme une dissimilarité. La propriété (3) traduit le fait que l'on admet que la distance perceptive entre un son et lui-même est nulle. De fait, les paires de sons identiques ne sont en général jamais présentées dans les tests de dissemblance (cela permet d'éviter S mesures supplémentaires). Notons toutefois que mesurer les distances perçues entre deux

⁴⁷La définition du coefficient de corrélation de Pearson est donnée à la section 2.2.5.

⁴⁸Rappelons que le nombre de combinaisons (sans ordre et sans répétitions) de 2 objets parmi N est égal à $C_S^2 = \binom{S}{2}$, où C est le coefficient binomial.

sons identiques peut s'avérer être un moyen d'évaluer la consistance des réponses des participants...

Les dissimilarités entre les stimuli sont évaluées sur une échelle de nature *ordinaire*⁴⁹. Considérons à ce titre la figure 2.10. Les intervalles entre les différents nombres de l'échelle n'ont pas nécessairement de sens, d'une part car les jugements perceptifs sont réalisés de manière relative, et d'autre part car la perception des dissemblances n'est sûrement pas linéaire le long de l'échelle. La nature perceptive d'un intervalle à un endroit de l'échelle (par ex. I_1) ne correspond peut être pas à celle d'un intervalle de même taille, situé à un autre endroit de l'échelle (par ex. I_2).

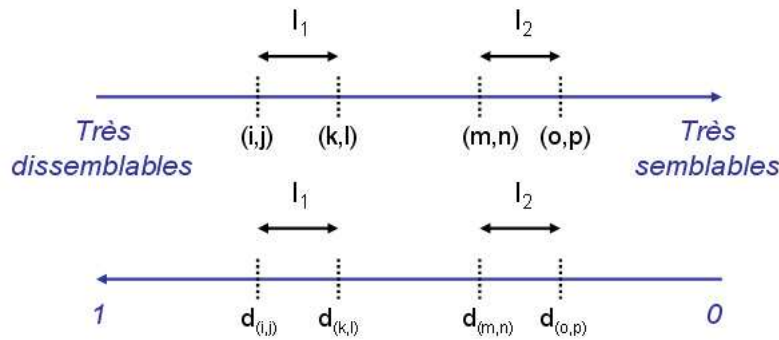


FIG. 2.10 – Echelle de mesure de dissimilarité entre sons : une échelle de type *ordinaire*. Les symboles $(i,j) \dots (o,p)$ représentent des couples de sons, $d_{(i,j)} \dots d_{(o,p)}$ les dissimilarités associées. Les deux intervalles I_1 et I_2 , bien que quantitativement identiques n'ont pas nécessairement le même sens au niveau perceptif. C'est l'ordre qui compte : $d_{(o,p)} < d_{(m,n)} < d_{(k,l)} < d_{(i,j)}$.

Dans les algorithmes de MDS dits *non-métriques*, les distances δ entre les points de l'espace euclidien délivré par l'analyse sont déterminées de manière à respecter au mieux l'ordre défini par les dissimilarités d ⁵⁰. Si $d_{(k,l)} < d_{(i,j)}$, les algorithmes cherchent à obtenir des distances entre objets telles que $\delta_{(k,l)} < \delta_{(i,j)}$, et ce pour un maximum d'objets. Notons que des algorithmes de MDS *métriques* existent également, qui cherchent à conserver les distances et non les rangs. La conservation des rangs est réalisée au moyen de transformations monotones croissantes des dissimilarités $M(d_{(i,j)})$ tout en minimisant une quantité (*goodness-of-fit criterion* en anglais), qui dans le cas de l'algorithme MDSCAL développé par Kruskal est appelé *stress* (de Kruskal). Le stress de Kruskal est donné par (cf. [Saporta, 1990, p. 183]) :

$$stress_{Kruskal} = \sqrt{\frac{\sum_{i,j} \left(\delta_{(i,j)} - M(d_{(i,j)}) \right)^2}{\sum_{i,j} \delta_{(i,j)}^2}} \quad (2.3)$$

C'est également cette quantité qui permettra de caractériser la "qualité" de la configuration MDS obtenue, ainsi que d'apprécier la perte d'informations subie (rôle que joue le pourcentage totale d'inertie expliquée dans le cas de l'ACP).

Pour l'algorithme MDSCAL, que nous avons utilisé lors de certaines études exposées dans cette thèse (cf. chapitres 4 et 6), c'est à l'utilisateur de spécifier le nombre de dimensions de la

⁴⁹Pour une discussion sur les échelles de mesures, se reporter par exemple au très bon ouvrage de Siegel et Castellan [Siegel et John Castellan, 1988], p. 23-32.

⁵⁰Notons que des algorithmes de MDS *métriques* existent également. Ces derniers cherchent à conserver les distances définies par les données d'entrée, et non leurs rangs.

configuration recherchée. Ce nombre de dimensions se détermine généralement de manière empirique à l'aide des différents critères suivants (cf. [Dillon et Goldstein, 1984, p. 145]) :

- les diagrammes de Shepard (représentation des distances en fonction des dissimilarités) qui permettent de vérifier la conservation des rangs entre les valeurs de dissimilarités et les distances obtenues dans l'espace (afin de respecter les relations d'ordre, la fonction liant distance et dissimilarités doit être une fonction monotone croissante) ;
- la "règle du coude" (*scree-elbow criterion*), qui suppose que la représentation du *stress* en fonction du nombre de dimensions testées présente un coude au niveau d'un nombre de dimensions adéquat (l'ajout de dimensions supplémentaires à l'espace ne faisant alors pratiquement plus évoluer le *stress*) ;
- un critère reposant sur la valeur du *stress*, qui consiste à considérer que les configurations dont la valeur de *stress* est inférieure à 0.1 offrent une représentation des données raisonnable [Krzanowski, 2000]. Cependant, la valeur du seuil de *stress* de 0.1 (donnée par Kendall) est empirique. Elle provient de constatations réalisées *a posteriori* à partir d'analyses de cas réels. Ce critère ne peut se suffire à lui-même, car on ne peut pas donner de seuil de manière générale, sans tenir compte de la taille du tableau, et de la force des corrélations entre variables [Saporta, 1990, p. 173].
- l'interprétation des résultats par l'expérimentateur, qui seul peut se rendre compte du bien-fondé de la représentation.

Notons également que la solution des méthodes MDS est définie à une transformation orthogonale près (rotation, symétrie, etc.) [Saporta, 1990, p. 184].

Souvent utilisée en complément de la MDS, la **classification ascendante hiérarchique (ou analyse en clusters)** permet de regrouper des objets en classes (groupes, ou clusters) qui présentent des différences systématiques. Une classification ascendante hiérarchique consiste donc en un ensemble de partitions de E (l'espace des S objets) en classes de moins en moins fines obtenues par le regroupement successifs de parties⁵¹. Elle se représente souvent par un *dendrogramme*, ou arbre de classification. Plusieurs algorithmes existent également pour ces méthodes selon la stratégie d'agrégation utilisée (cf. [Saporta, 1990, p. 256], et [Dillon et Goldstein, 1984, p. 168]) : le saut minimum (*single linkage* en anglais), qui consiste à considérer la plus petite distance entre éléments de deux parties, son inverse, le saut maximum (*complete linkage*), la technique de Ward, reposant sur la perte d'inertie entre deux classes résultant de leur regroupement, en sont quelques exemples. Ces méthodes sont particulièrement adaptées lorsque la représentation perceptive sous-jacente à un corpus sonore n'est pas de type continue. C'est notamment le cas lorsqu'il existe de larges différences entre les stimuli considérés.

2.2.4.1.5 Interprétation psychoacoustique de la structure Il revient au chercheur d'interpréter la configuration géométrique des stimuli délivrée par la MDS, ou la classification obtenue suite à une analyse en clusters. Cette interprétation pourra être réalisée en cherchant des facteurs acoustiques ou psychoacoustiques permettant d'expliquer l'ordre des stimuli suivant les différentes dimensions de l'espace de représentation de l'analyse MDS, et/ou rendant compte, dans le cas d'analyses en clusters, du regroupement des stimuli similaires dans des classes. Généralement, le point de départ de l'interprétation psychoacoustique consiste à écouter tous les sons répartis le long d'une dimension de l'espace de représentation, en tâchant de déceler ce qui change de manière systématique. Les paramètres psychoacoustiques qui présentent de fortes corrélations linéaires avec les coordonnées des sons le long d'une dimension sont couramment appelés *descripteurs* de timbre [Krimphoff et al., 1994] (cf. section 2.2.5). Il est important de garder en tête

⁵¹Le terme *ascendante* provient du fait que la méthode procède, par un processus itératif, au regroupement des objets les plus proches, jusqu'en celui des objets, ou groupes d'objets les moins proches.

que de tels descripteurs peuvent être considérés comme de bons prédicteurs de la structure de la représentation perceptive des sons, mais ceci ne veut pas dire pour autant qu'ils expliquent la façon dont nous percevons (*prédire n'est pas expliquer*).

2.2.4.1.6 Vérification de l'interprétation dans des situations musicales Tel que mentionné par Wessel dans [Wessel, 1978], il est également une étape de la méthodologie d'étude du timbre essentielle, mais pourtant rarement considérée, celle de la vérification de la pertinence des descripteurs de timbre découverts dans des situations musicales. En effet ne perdons pas de vue qu'outre leur caractère fondamental, les recherches sur le timbre ont aussi des applications pour le contrôle de l'expressivité des sons dans la composition et l'interprétation (cf. par ex. [Wessel, 1978] et [Jensen, 2002]). On peut s'enquérir entre autres de l'aptitude des espaces de timbre à prédire comment effectuer des *transpositions de timbre* (considérées au même titre qu'il existe des transpositions de hauteurs). Ehresman et Wessel, dans [Ehresman et Wessel, 1978], ainsi que McAdams et Cunibile, dans [McAdams et Cunibile, 1992], ont notamment étudié si des translations vectorielles similaires dans un espace de timbre (même orientation, sens, et longueur) procuraient des perceptions d'intervalles de timbre similaires. Les résultats tendent à soutenir le modèle vectoriel d'intervalles de timbre, même s'ils sont nuancés dans le cas de [McAdams et Cunibile, 1992], par le fait que les mesures associées à certains participants (non musiciens) des expériences perceptives y adhéraient bien moins que d'autres (compositeurs).

2.2.4.2 Les espaces de timbre de Grey et McAdams *et al.*

2.2.4.2.1 L'espace de timbre de Grey Dans son étude sur les dissemblances de timbre entre 16 notes isolées d'instruments de l'orchestre reproduites par synthèse⁵², Grey obtint une configuration MDS (méthode INDSCAL) à 3 dimensions. Ces dernières sont, selon l'auteur, reliées avec les facteurs suivants :

- Dimension I : la distribution d'énergie spectrale ;
- Dimension II : la synchronicité des harmoniques de rangs élevés au cours des transitoires d'attaque, et de manière associée, le degré de fluctuation spectrale au cours du temps ;
- Dimension III : la présence d'énergie haute fréquence de faible amplitude pendant l'attaque.

La figure 2.11 tâche de résumer les différents facteurs d'interprétation de l'espace de timbre obtenu par Grey. A ceux qui restent sceptiques sur le fait que le timbre d'un même instrument peut présenter de fortes variations, nous signalerons que dans l'étude de Grey, les distances perceptives entre différents sons d'un même instrument, produits selon différents modes de jeu, sont parfois plus grandes que les distances perceptives entre des sons provenant d'instruments différents ; c'est le cas des deux sons de saxophones *piano* et *mezzo forte*, ainsi que des deux sons de violoncelle *muted sul tasto* (étouffé près des touches) et *sul ponticello* (près du chevalet), qui sont situés à des positions opposées le long de la dimension I reliée à la distribution spectrale (ce qui s'interprète par leurs largeurs de bande différentes). De plus, les résultats d'une analyse de classification ascendante hiérarchique montrent que les trois sons de violoncelle (les deux mentionnés ci-dessus et un son joué normalement) n'appartiennent pas à un même groupe, le son joué *sul tasto* étant jugé plus proche d'un son de flûte.

2.2.4.2.2 L'espace de timbre de McAdams *et al.* Dans une de leurs études sur le timbre, McAdams et ses collaborateurs ont utilisé la synthèse par modulation de fréquence (aussi connue sous le nom de synthèse FM, d'après l'anglais *Frequency Modulation*) [Chowning, 1973] afin de

⁵²Dans l'étude de Grey, les sons proviennent du hautbois, du cor anglais, du basson, de clarinettes, de saxophones, de flûte, de trompette, du cor français, du trombone, et du violoncelle. Certaines notes correspondent à des modes de jeu différents.

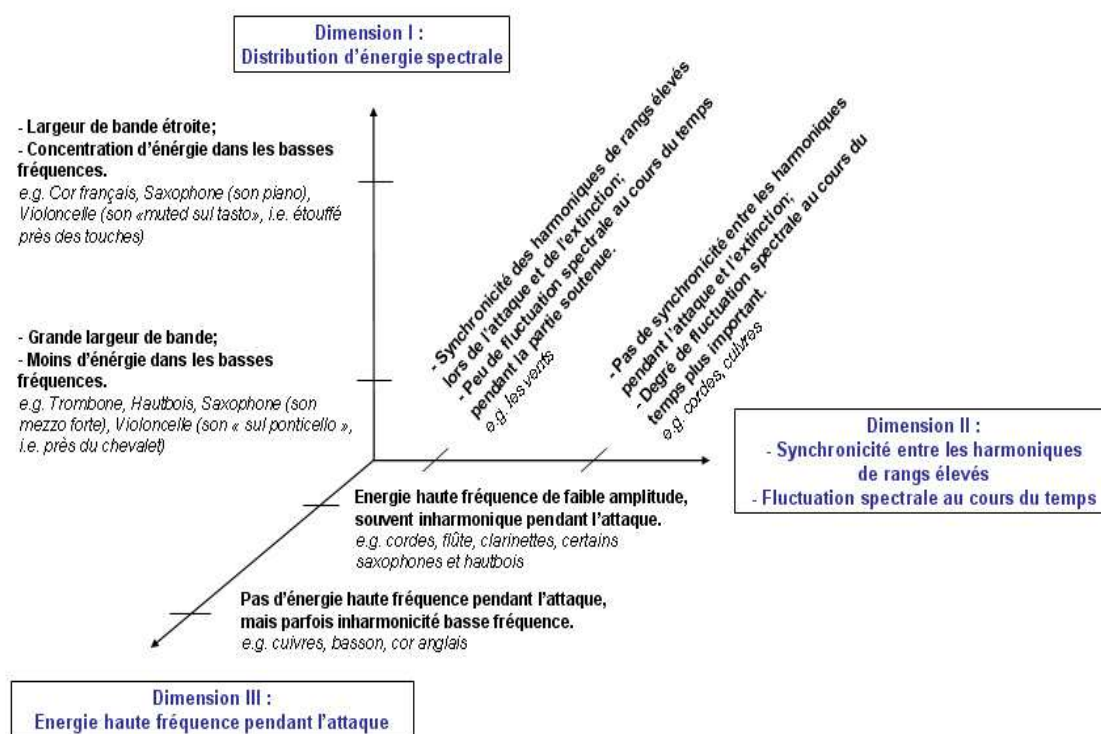


FIG. 2.11 – Schéma synthétique de l'espace de timbre à 3 dimensions de Grey indiquant les facteurs psychoacoustiques permettant d'en expliquer la structure.

produire 18 sons instrumentaux, incluant tant des imitations d'instruments que des instruments hybrides⁵³.

Comme l'illustre la figure 2.12 issue de [McAdams, 1999], la configuration MDS obtenue (méthode CLASCAL) est également à 3 dimensions. Les corrélats acoustiques qui prédisent le mieux la répartition des stimuli le long des différentes dimensions sont :

- Dimension 1 : le logarithme du temps d'attaque, rendant compte du temps de montée de l'énergie dans le son ;
- Dimension 2 : le Centre de Gravité Spectral, rendant compte de la brillance du son ;
- Dimension 3 : le Flux Spectral, rendant compte des fluctuations spectrales au cours du temps.

Nous allons définir ces descripteurs précisément dans la prochaine section. Il est intéressant de remarquer que les résultats des études de Grey et McAdams *et al.*, que nous venons de rappeler, ont de nombreux points en commun. Un espace perceptif de timbre à 3 dimensions a été obtenu dans les deux cas⁵⁴. La dimension I de l'espace de McAdams *et al.*, concernant l'attaque des sons, semble être reliée à la dimension III de l'espace de Grey. La dimension II de l'espace de McAdams *et al.*, associée au Centre de Gravité Spectral, est corrélée à la dimension I de l'espace de Grey (pour les sons instrumentaux, un lien fort existe entre la valeur du Centre de Gravité Spectral et la largeur de bande des sons). La dimension III de l'espace de McAdams *et al.* est similaire à la

⁵³Les instruments utilisés sont le cor français, la trompette, le trombone, la harpe, le vibraphone, le clavecin, le cor anglais, le basson, la clarinette, la guitare, un son de corde frottée, le piano. Les instruments hybrides sont la *trumpar* (trompette/guitare), l'*oboleste* (hautbois/celesta), le *striano* (corde frottée/piano), le *vibrone* (vibraphone/trombone), l'*obochord* (hautbois/clavecin) et la *guitarnet* (guitare/clarinette).

⁵⁴De fait, la majeure partie des études sur les dissemblances de timbre révèlent des espaces de timbre à 2, ou 3 dimensions.

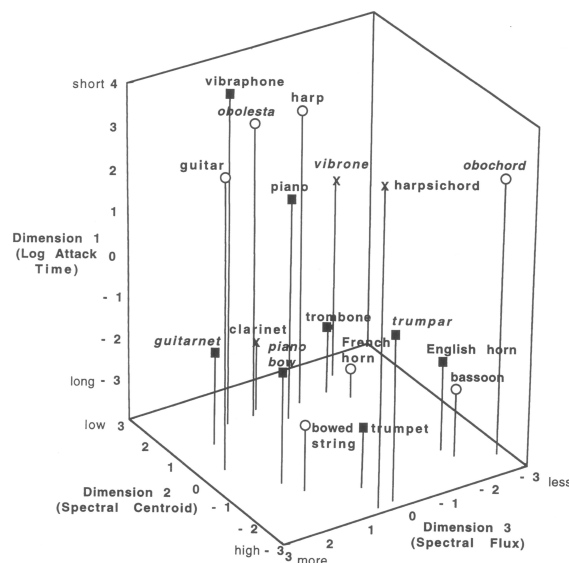


FIG. 2.12 – Espace de timbre à 3 dimensions obtenu par McAdams *et al.*. La figure indique les corrélats acoustiques (*descripteurs de timbre*) ayant présentés les meilleures prédictions de la position des stimuli le long des différentes dimensions. Figure de McAdams *et al.*, issue de [McAdams, 1999].

dimension II de l'espace de Grey, liée à la synchronicité des différents composantes spectrales, et à leurs fluctuations temporelles.

2.2.5 Les descripteurs de timbre et de signal, corrélats physiques de la perception

Les *descripteurs de timbre* sont des mesures objectives qui ont pour but de caractériser le timbre. Ils sont obtenus à partir d'algorithmes de calcul reposant sur l'analyse des sons et d'éventuels modèles physiologiques de l'audition. Certains d'entre eux s'avèrent être de bons prédicteurs de la structure de la représentation psychologique sous-jacente à la perception du timbre, i.e. de l'*espace de timbre*. Ils sont à ce titre utilisés pour l'identification et la classification automatique des timbres instrumentaux, et d'une manière plus générale, pour l'extraction de caractéristiques musicales (*Music Information Retrieval* en anglais). Certains systèmes de codage audio numérique (par ex. MPEG 4 [Tzanetakis, 2002]), et certaines normes de description audio dont le but est de faciliter l'indexation et la recherche de documents sonores dans des bases de données (par ex. MPEG 7 [Kim *et al.*, 2005]) reposent sur des descripteurs de timbre. Nous ferons l'emploi de descripteurs de timbre au chapitre 4, afin d'interpréter la structure de l'espace perceptif obtenu à l'issue d'une étude de dissemblance sur des sons de synthèse de clarinette, et aux chapitres 5 et 6, afin d'analyser dans quelle mesure le timbre est corrélé à l'expression musicale.

Une description de nombreux descripteurs audio pourra être trouvée dans [Peeters, 2004]. Nous présenterons ici les principaux descripteurs de timbre rencontrés dans la littérature. Ces derniers peuvent être divisés en trois catégories : **les descripteurs temporels**, **les descripteurs spectraux**, et **les descripteurs spectro-temporels**. Notons que ces distinctions ne réfèrent pas aux techniques de calcul associées mais bien aux aspects que les descripteurs reflètent. Dans certains cas, les calculs peuvent d'ailleurs être effectués aussi bien dans le domaine temporel que dans le domaine spectral. Par exemple, le Centre de Gravité Spectral peut être calculé de manière spectrale, mais également de manière temporelle. Inversement, l'enveloppe temporelle d'énergie peut être calculée dans le domaine spectral, mais également dans le domaine temporel. Il est im-

portant de remarquer que si un descripteur spectral est considéré dans son évolution au cours du temps, il deviendra dès lors un descripteur spectro-temporel. Associé à ce découpage, nous ferons également la dichotomie importante entre **les descripteurs instantanés**, et **les descripteurs globaux**. Aussi, en fonction de l'horizon temporel d'analyse considéré, un même descripteur pourra revêtir un caractère instantané (définition à court-terme) ou bien un caractère global (définition à long-terme). En effet, si l'évolution temporelle d'un descripteur instantané est caractérisée par des statistiques descriptives (par ex. moyenne, variance) sur une échelle de temps suffisamment longue (par ex. totalité du son, partie entretenue du son, etc.), le nouveau descripteur obtenu pourra pour fonction de caractériser de manière globale le son ou une partie du son. D'autres descripteurs n'auront pas l'ambivalence instantané/global, et ne pourront être définis que globalement, comme caractéristique de tout le signal (le temps d'attaque d'un son, par exemple).

Comme nous le verrons, les formulations retenues ici pour le calcul des descripteurs, reposent pour la plupart sur une représentation temps-fréquence obtenue par la Transformée de Fourier calculée à Court Terme TFCT (*Short Time Fourier Transform STFT* en anglais)⁵⁵. En pratique, nous travaillons sur des signaux échantillonnés, et utilisons une estimation numérique de la transformée, la Transformée de Fourier Discrète à Court Terme (TFDCT).

2.2.5.1 Notations et boîte à outils MATIMBRE

Nous désignerons par F_e (respectivement T_e) la fréquence (respectivement période) d'échantillonnage des signaux. $x(m)$ sera généralement utilisé pour désigner un signal temporel à valeurs discrètes réelles. $x(t)$ sera généralement utilisé pour désigner un signal temporel à valeurs continues réelles.

La Transformée de Fourier Discrète à Court Terme du signal échantillonné $x(m)$ sera notée $\{X_n(k)\}_{k=1,\dots,N}$, où n désigne l'horizon temporel d'analyse, k est un entier qui fait référence aux coefficients de la Transformée de Fourier Discrète (TFD), et N est le nombre de points de calcul de la TFD. En pratique, un nombre de points de calcul pair sera généralement choisi afin d'optimiser le temps de calcul de l'algorithme de Transformée de Fourier Rapide (TFR) (*Fast Fourier Transform* (FFT), en anglais). Rappelons que la Transformée de Fourier glissante correspond à une succession de TFD sur des segments ou trames temporelles (*frame* en anglais) extraites du signal temporel $x(m)$ par fenêtrage. La segmentation du signal x est réalisée au moyen d'une fenêtre d'analyse w que l'on fait glisser le long de x . Chaque segment fenêtré $x_n(m)$ est obtenu ainsi :

$$x_n(m) = w(m - n)x(m) \quad (2.4)$$

où w désigne la fenêtre d'analyse utilisée d'une longueur de W échantillons. $w(m - n)$ correspond donc à la fenêtre centrée sur l'horizon temporel d'analyse n . Afin d'être précis en temps, on peut faire glisser les fenêtres en les décalant d'un nombre faible d'échantillons, inférieur à la taille de la fenêtre. Il y a dans ce cas du recouvrement (échantillons en commun) entre les trames adjacentes. Nous noterons $x_\tau(t)$ le signal continu associé à $x_n(m)$.

Il sera supposé que les coefficients de la TFD associés aux fréquences positives (incluant la fréquence nulle et la fréquence de Nyquist) sont ceux obtenus pour $k = 1, \dots, 1 + \frac{N}{2}$ si N est pair, et pour $k = 1, \dots, \frac{N}{2}$ si N est impair.

Les amplitudes et phases de la TFDCT seront respectivement notées $A_n(k)$ et $\phi_n(k)$. La fréquence (positive) $f(k)$ associée à la k^{eme} composante spectrale est obtenue par l'équation suivante :

⁵⁵La Transformée de Fourier à Court Terme est également appelée Transformée de Fourier glissante.

$$f(k) = (k - 1) \frac{F_e}{N} \quad (2.5)$$

où $k = 1, \dots, 1 + \frac{N}{2}$, lorsque N est pair. Ainsi $f(1)$ correspond à la fréquence nulle, et $f(1 + \frac{N}{2})$ correspond à la fréquence de Nyquist $\frac{F_e}{2}$. Le rapport $\Delta f = \frac{F_e}{N}$ correspond à la résolution fréquentielle de l'analyse. Notons que N ne doit pas nécessairement être égal à la taille de la fenêtre W . Il peut lui être supérieur afin d'obtenir un meilleur échantillonnage fréquentiel. Il est important de remarquer que la valeur de la fréquence $f(k)$ ne dépend pas du temps. En effet, pour la TFDCT la résolution fréquentielle est invariante au cours du temps (le pavage temps-fréquence est fixe). Il s'ensuit que la localisation fréquentielle des composantes spectrales analysées est constante au cours du temps.

Les terminologies employées pour les descripteurs seront données au fur et à mesure de leurs présentations. Notons dès à présent que celle-ci repose sur les noms anglais des descripteurs. Par exemple, le Centre de Gravité Spectral sera désigné par SC, en raison de la terminologie anglaise Spectral Centroid.

La figure 2.13 présente la méthodologie générale d'évaluation des descripteurs. La boîte à outils MATIMBRE, dédiée à l'analyse du signal et au calcul de descripteurs de timbre au sein de l'environnement Matlab, a été développée à cet effet. Notons qu'il existe d'autres boîtes outils qui permettent l'analyse du timbre (cf. par exemple les boîtes à outils Psysound [Cabrerá et al., 2008], et Mirtoolbox [Lartillot et Toivainen, 2007]). L'un des intérêts du développement de la boîte à outils MATIMBRE est de disposer de fonctions de calcul des descripteurs de timbre reposant sur les définitions standard, qui peuvent être adaptées à l'analyse de l'interprétation musicale (par exemple par l'ajout d'un paramètre régularisateur qui permet de réduire l'influence du bruit au moment des transitions entre notes pour certains descripteurs, ou l'emploi de différentes échelles d'amplitude, points qui seront discutés dans la suite de la section 2.2.5). La boîte à outils MATIMBRE s'accompagne d'une documentation et de nombreuses fonctions d'exemples, permettant de présenter son fonctionnement à l'utilisateur.

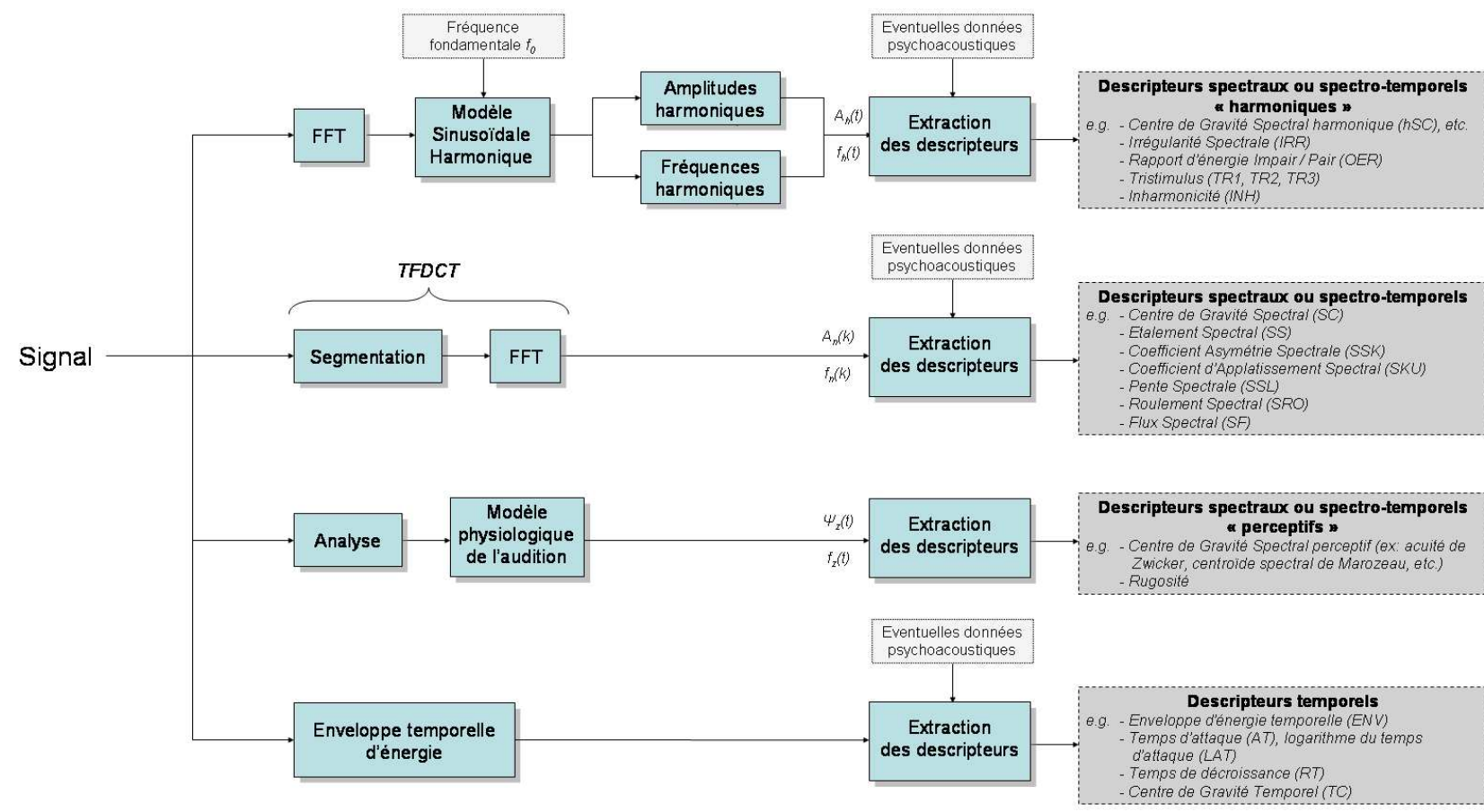


FIG. 2.13 – Méthodologie d’évaluation des descripteurs de timbre et de signal dans le cas de signaux (quasi-)harmoniques. La nomenclature des descripteurs est celle utilisée dans le manuscrit.

2.2.5.2 L'enveloppe temporelle d'amplitude (ENV)

L'enveloppe temporelle d'amplitude rend compte de la variation d'amplitude d'un signal au cours du temps. Il s'agit donc d'un descripteur instantané. Différentes méthodes de calcul sont possibles. Nous exposerons celles qui sont les plus répandues.

2.2.5.2.1 Calcul de l'enveloppe crête à partir du signal analytique Pour un signal $x(t)$ donné, il existe une infinité de couples $[\tilde{a}(t), \tilde{\phi}(t)]$, tels que (cf. [Picinbono, 1997]) :

$$x(t) = \tilde{a}(t) \cos[\tilde{\phi}(t)] \quad (2.6)$$

où $\tilde{a}(t)$ est une fonction de modulation d'amplitude, et $\tilde{\phi}(t)$ est une fonction de modulation de phase. Même si la confusion est souvent faite, il n'est pas rigoureux de considérer les fonctions $\tilde{a}(t)$ et $\tilde{\phi}(t)$ comme les amplitudes et fréquences instantanées du signal puisqu'elles ne sont pas définies de manière unique.

Le signal analytique $z(t)$ associé à un signal $x(t)$ à valeurs réelles, permet de définir un unique couple $[a(t), \phi(t)]$, dit canonique, tel que :

$$z(t) = a(t) \exp[j\phi(t)] \quad (2.7)$$

où $\exp$ est la fonction exponentielle, et j est l'unité imaginaire. Le signal analytique permet de définir de façon rigoureuse (univoque) l'amplitude instantanée $a(t)$ et la phase instantanée $\phi(t)$ associées au signal $x(t)$ en généralisant au cas non stationnaire les notions d'amplitudes et de phase dans un cas stationnaire. En effet, l'amplitude et la phase d'un signal stationnaire monochromatique correspondent au module et à la dérivée de la phase du signal analytique.

Le signal analytique $z(t)$ associé à un signal à valeurs réelles $x(t)$ peut s'écrire :

$$z(t) = x(t) + j\mathcal{H}\{x(t)\} \quad (2.8)$$

La partie imaginaire du signal analytique correspond à la transformée de Hilbert $\mathcal{H}$ du signal auquel il est associé. Dans le domaine temporel, la transformée de Hilbert $\mathcal{H}$ du signal $x(t)$ est obtenue par convolution avec la fonction $h(t) = \frac{1}{\pi t}$:

$$\mathcal{H}\{x(t)\} = (h * x)(t) \quad (2.9)$$

Dans le domaine fréquentiel, cette opération revient à filtrer le signal $x(t)$ par un filtre de réponse en fréquence $\theta(\omega)$, défini par (cf. [Max et Lacoume, 2000, p. 37]) :

$$\theta(\omega) = \begin{cases} 2 & \text{si } \omega \geq 0 \\ 0 & \text{si } \omega < 0 \end{cases} \quad (2.10)$$

Notons que le signal analytique $z(t)$ ne peut être une fonction réelle puisque sa transformée de Fourier $Z(\omega)$ est nulle pour les fréquences négatives.

En prenant le problème à l'inverse, si l'on connaît le signal analytique $z(t)$ associé à un signal $x(t)$, il est alors possible d'obtenir le signal $x(t)$, le couple canonique $[a(t), \phi(t)]$, où $a(t)$ et $\phi(t)$ sont respectivement l'amplitude et la phase instantanée, de la manière suivante :

$$\begin{cases} x(t) = \text{Re}[z(t)] \\ a(t) = |z(t)| \\ \phi(t) = \arg(z(t)) \end{cases} \quad (2.11)$$

où $\text{Re}[z(t)]$, $|z(t)|$, et $\arg(z(t))$ sont respectivement la partie réelle, le module et l'argument du complexe $z(t)$. Notons que l'amplitude instantanée $a(t)$ est par définition positive (module de $z(t)$), et que la phase instantanée $\phi(t)$ est définie modulo 2π . Nous verrons à la section 2.2.5.5 que cette procédure peut avoir des applications intéressantes pour l'analyse et la synthèse sonore.

Concernant le calcul de l'enveloppe d'amplitude d'un signal, il peut s'avérer utile de ne conserver que les modulations basses fréquences de l'amplitude instantanée $a(t)$. En effet, la résolution temporelle du système auditif fait que nous sommes essentiellement sensibles aux variations "lentes" de l'amplitude⁵⁶. Pour ce faire, cette dernière peut être filtrée par un filtre passe-bas. Aussi, si nous notons $h_{PB}(t)$ la réponse impulsionnelle du filtre passe-bas, l'enveloppe d'amplitude temporelle $ENV(t)$ du signal $x(t)$ peut être obtenue par l'équation suivante :

$$ENV(t) = (a * h_{PB})(t) \quad (2.12)$$

Un filtre passe-bas avec une fréquence de coupure de 5 Hz est utilisé dans [Peeters, 2004]. Notons que l'enveloppe alors obtenue est une enveloppe crête (qui suit les maxima du signal analysé).

2.2.5.2.2 Enveloppe RMS (Root Mean Square) Une autre méthode classique de détermination de l'enveloppe consiste à calculer les valeurs quadratiques moyennes ou valeurs RMS (Root Mean Square) instantanées du signal. Les valeurs RMS instantanées d'un signal $x(t)$ correspondent aux valeurs RMS associées aux différentes trames du signal $x_\tau(t)$. Pour chaque trame $x_\tau(t)$, la valeur RMS correspond à la racine carrée de la puissance moyenne calculée sur la durée T de la trame. Puisque la Transformée de Fourier est une isométrie, les valeurs RMS peuvent être calculées aussi bien dans le domaine temporel que dans le domaine fréquentiel (de Fourier). En effet, d'après le théorème de Parseval :

$$\left(\int_{\tau - \frac{T}{2}}^{\tau + \frac{T}{2}} |x_\tau(t)|^2 dt \right)^{\frac{1}{2}} = \left(C \int_{-\infty}^{+\infty} |x_\tau(\omega)|^2 d\omega \right)^{\frac{1}{2}} \quad (2.13)$$

où T est la durée de la fenêtre d'analyse centrée en τ , et de valeurs nulles à l'extérieur de l'intervalle $[\tau - \frac{T}{2}; \tau + \frac{T}{2}]$. C est une constante dont la valeur est liée à la convention choisie pour la Transformée de Fourier (en général, $C = \frac{1}{2\pi}$).

Dans le domaine fréquentiel, l'enveloppe d'amplitude RMS, notée $ENV(n)$, est donnée par l'équation suivante (cf. par ex. [Tzanetakis, 2002, p. 37]) :

$$ENV(n) = \sqrt{\frac{\sum_{k=1}^K A_n(k)^2}{K}} \quad (2.14)$$

où $K = 1 + \frac{N}{2}$, lorsque N est pair, et $K = \frac{N}{2}$, lorsque N est impair. Pour une onde sinusoïdale monochromatique d'amplitude crête A , l'amplitude RMS est égale à $A \frac{\sqrt{2}}{2}$. Remarquons

⁵⁶Les mesures obtenues dans [Moore et al., 1988] indiquent que la résolution temporelle du système auditif est d'environ 8 ms (pour des signaux de 500 et 2000 Hz). Une telle résolution revient à ne pouvoir percevoir que des variations d'amplitudes dont la fréquence ne dépasse pas 125 Hz.

que contrairement à l'amplitude crête, l'amplitude RMS présente l'avantage de ne pas être définie de manière ambiguë lorsque la forme d'onde du signal analysé n'est pas symétrique (dans le cas de l'amplitude crête, la valeur dépend de la référence que l'on choisit : la valeur maximale positive par rapport à la moyenne, la valeur maximale négative par rapport à la moyenne, ou la demi-somme des valeurs maximales positive et négative). Comme les signaux échantillonnés sont rarement symétriques, et a fortiori les signaux échantillonnés provenant d'enregistrements de sons instrumentaux, nous avons préféré utiliser l'enveloppe d'amplitude RMS pour caractériser la dynamique au cours du temps plutôt qu'une enveloppe d'amplitude crête dérivée du signal analytique.

2.2.5.3 Les descripteurs temporels

2.2.5.3.1 Le temps d'attaque (AT) Au niveau acoustique, l'attaque réfère à la portion instable d'un signal comprise entre son amorce (*onset* en anglais) et le moment où le système mécanique vibratoire⁵⁷ parvient à un état stable lorsqu'une onde stationnaire est établie (cet état stable est alors appelé régime permanent, par opposition au régime transitoire qui correspond à la phase d'excitation du système). Le temps d'attaque rend compte du temps de montée de l'énergie pendant le régime transitoire. Il s'agit donc d'un descripteur global.

Il n'existe pas à ce jour de méthodes robustes de calcul qui seraient valables pour tous types de sons, dont les dynamiques sont souvent bien différentes (par ex. sons percussifs, sons entretenus). Il se peut très bien qu'un son (issu de la synthèse, ou du travail sur bande magnétique par exemple) ne dispose pas à proprement parler d'attaque (c'est le cas lorsque l'on coupe la portion initiale d'un son d'instrument, de sorte qu'il démarre directement dans sa partie entretenue, ou décroissante). Dans ses travaux sur la perception de l'attaque, Schaeffer a employé le terme d'*anamorphose temporelle*. Il met là en exergue les déformations qu'il peut exister entre les domaines de la physique et de la perception ; "les instants de notre écoute ne coïncident pas avec les centimètres de la bande magnétique." [Schaeffer, 2005, p. 38]. Il a montré que la perception de l'attaque est liée à la forme générale des sons, à leur dynamique. Le fait que la perception de l'attaque dépende de ce qu'il advient après peut paraître paradoxal ! En effet, il avait remarqué que si l'on coupait les premiers instants d'une note de piano issue du registre grave, l'identification de l'instrument n'en était que faiblement perturbée. De telles notes présentent après l'attaque physique, une dynamique qui est sensiblement une droite dont la pente est constante et faible. En revanche, si l'on ôte une portion suffisamment longue au début d'un son provenant du registre aigu du piano, l'identité instrumentale peut être fortement altérée (en coupant 1,5 s un son de piano aigu devient flûté). Pour les notes des registres aigus du piano, la dynamique après la phase d'attaque physique, ne correspond pas à une pente régulière (par ex. le A4 correspondant à la Figure 7 du *Traité des objets musicaux* [Schaeffer, 1966, p. 216] présente successivement décroissance, aplatissement, creux et même ressaut). De plus, son allure générale est beaucoup plus raide (cf. loi du piano exposée section 2.1.3.1). De ses expériences, Schaeffer a établi deux lois sur la perception des attaques qu'il nous semble utile de rappeler :

“- 1^{ère} loi : pour les sons entretenus, de façon générale, l'oreille est sensible, pour qualifier sa perception de la raideur d'attaque, à la façon dont l'énergie sonore apparaît dans le temps ;
- 2^{ème} loi : pour les sons à attaque percutée ou pincée suivie de résonance, l'oreille est sensible, pour qualifier sa perception de raideur d'attaque, à la façon dont l'énergie disparaît plus encore qu'à celle dont elle apparaît." (ibidem, p. 226 et 229).

Gordon a confirmé les expériences de Schaeffer en montrant que bien que deux sons soient synchrones au niveau de leur amorce, leurs attaques pouvaient ne pas être perçues comme simul-

⁵⁷ dans le cas d'un son d'instrument de musique, par exemple.

tanées [Gordon, 1987]. L'instant correspondant à l'amorce physique d'un son et l'instant du début de la perception d'attaque doivent donc être distingués (plus le temps de montée est long, plus la distinction est prononcée). D'autres conclusions se sont dégagées de son étude :

- Lorsque le temps de montée d'un son est rapide, la perception de son temps d'attaque est déterminée essentiellement par des facteurs liés à l'amplitude ;
- Lorsque le temps de montée d'un son est long, la perception de l'attaque est plus facilement influençable par des facteurs spectraux (par ex. non synchronicité entre les harmoniques de faibles rangs et les harmoniques de rangs élevés). Il en résulte qu'elle est plus confuse chez les auditeurs. En effet, il paraît bien difficile de saisir la notion d'attaque dans le cas d'un *crescendo* de clarinette progressant lentement du *pianissimo* au *fortissimo*.

Plusieurs modèles de temps d'attaque reposant sur l'enveloppe d'amplitude ont été testés par Gordon [Gordon, 1987] : (1) le temps mis pour parvenir à la valeur maximale de l'enveloppe d'amplitude, (2) le temps que met l'enveloppe pour dépasser un seuil absolu, (3) le temps que met l'enveloppe pour dépasser un seuil relatif à sa valeur maximale (qui peut être exprimé comme pourcentage de la valeur maximale ou en un nombre de décibels en dessous du maximum), (4) repose sur un seuil absolu comme dans (2), mais par rapport à l'enveloppe d'énergie.

Comme nous pouvons le constater, la plupart de ces modèles sollicitent la définition d'un seuil. Le modèle (1) n'offre pas de bonnes performances. En revanche, les trois autres modèles présentent de bonnes prédictions de la perception, mais peuvent cependant faire des erreurs de plus de 10 ms (par rapport au temps d'attaque perçu), ce qui est au dessus de la résolution temporelle du système auditif (environ 5 ms). Le modèle (3) reposant sur un pourcentage de l'amplitude maximale s'est avéré marcher le mieux avec un seuil de 5.8 % du maximum (valeur qui diffère singulièrement des 17.8 % initialement proposés par Vos et Rasch [Vos et Rasch, 1981]). Les seuils dépendent donc manifestement des sons utilisés dans les expériences perceptives.

Le calcul du descripteur de temps d'attaque AT est réalisé selon la formule proposée dans [Krimphoff et al., 1994] qui le définit comme l'intervalle de temps entre un instant qui prédit le début de la perception du son t_{deba} , et l'instant t_{max} lorsque l'enveloppe atteint sa valeur maximale. Le seuil $seuil_{deba}$, associé à t_{deba} n'est pas le seuil d'audition absolu, mais dépend linéairement de l'amplitude maximale [Vos et Rasch, 1981]. De plus, il permet de s'affranchir d'une éventuelle présence de bruit dans le signal. Cependant, comme nous avons vu que l'instant correspondant au maximum de l'enveloppe (cf. modèle (1) de Gordon ci-dessus) n'était pas le plus adéquat, nous préférons utiliser un deuxième seuil relatif à ce maximum t_{fina} (cf. modèle (3) de Gordon ci-dessus), comme proposé également dans [Peeters, 2004]. Ainsi :

$$AT = t_{fina} - t_{deba} \quad (2.15)$$

où t_{deba} et t_{fina} correspondent respectivement aux temps mis par l'enveloppe d'amplitude pour atteindre les seuils de début d'attaque $seuil_{deba}$, et de fin d'attaque $seuil_{fina}$. Les valeurs de seuils pourront être exprimées en fonction d'une certaine quantité de décibels en dessous de l'amplitude maximale (Krimphoff et al. ont testé des seuils de début d'attaque allant de -2 dB à -23 dB de l'amplitude maximale [Krimphoff et al., 1994]). Les seuils peuvent également être exprimés par un pourcentage de l'amplitude maximale (Peeters propose des seuils de début et de fin d'attaque correspondant respectivement à $seuil_{deba} = 20\%$, et $seuil_{fina} = 90\%$ de la valeur d'amplitude maximale de l'enveloppe [Peeters, 2004]). En raison des anamorphoses temporelles pour reprendre l'expression de Schaeffer, et de la nature empirique de la définition du descripteur de temps d'attaque (utilisation de seuils), on se gardera de confondre le temps de fin d'attaque avec l'instant physique où une éventuelle onde stationnaire se met en place dans le système vibratoire.

Le temps d'attaque est souvent exprimé sur une échelle logarithmique. Le descripteur associé, que l'on notera LAT , est souvent l'une des caractéristiques acoustiques retenues pour interpréter les espaces de timbre. En effet, Krimphoff *et al.* [Krimphoff *et al.*, 1994] ont trouvé que ce descripteur était fortement corrélé [$r=0.94$, $p<0.01$] avec la première dimension de l'espace de Krumhansl [Krumhansl, 1989], obtenu sur un set de 21 timbres synthétiques (obtenus par Wessel *et al.* par synthèse FM [Wessel *et al.*, 1987]), associés soit à des sons naturels instrumentaux, soit à des sons hybrides (par ex. piano "frotté"). McAdams *et al.* ont signalé une corrélation aussi forte [$r=-0.94$, $p<0.0001$, soit 88% de la variance expliquée] avec la première dimension de la configuration MDS obtenue suite à leur expérience [McAdams *et al.*, 1995]. Le logarithme du temps d'attaque LAT est défini par :

$$LAT = \log_{10}(AT) \quad (2.16)$$

où la fonction $\log_{10}$ est le logarithme en base 10.

Le temps de décroissance (RT) Dans le cas de sons entretenus instrumentaux, il existe après la phase d'attaque une phase de régime permanent, pendant laquelle l'énergie nécessaire aux auto-oscillations du système acoustique est fournie de manière plus ou moins continue par l'instrumentiste. Le début de la phase de décroissance est lié au moment où l'énergie n'est plus fournie au système vibratoire. L'onde stationnaire est alors amortie en raison des différentes formes de pertes que subit la pression acoustique (par ex. pertes visco-thermiques). Il en résulte un déclin régulier de l'amplitude du son.

Il est possible, de la même façon que pour le temps d'attaque, de définir un descripteur global qui caractérise le temps de décroissance, ou temps que met l'énergie à se dissiper à la fin d'un son. En se reposant sur la même méthodologie que pour le temps d'attaque, le temps de décroissance peut être calculé par :

$$RT = t_{find} - t_{debd} \quad (2.17)$$

où t_{debd} et t_{find} correspondent respectivement aux temps pour lesquels l'enveloppe d'amplitude atteint les seuils de début de décroissance $seuil_{debd}$, et de fin de décroissance $seuil_{find}$. Nous avons généralement utilisé des seuils tels que $seuil_{debd} = 90\%$, et $seuil_{find} = 20\%$ du maximum de l'enveloppe d'amplitude, considérée dans la partie décroissante du son.

Notons qu'il existe des méthodes de détection automatique des début et fin de phases d'attaque et de décroissance (cf. par ex. [Jensen, 1999a]).

2.2.5.3.2 Le Centre de Gravité Temporel (TC) Le Centre de Gravité Temporel est la moyenne pondérée par le temps de l'enveloppe d'amplitude (ou d'énergie). C'est donc un descripteur global du son. Il rend compte de la zone temporelle où l'énergie est concentrée. Ce descripteur permet de distinguer des sons percussifs (pour lesquels l'attaque est très brève) de sons entretenus (pour lesquels l'énergie apparaît en général plus progressivement). Il se calcule comme suit :

$$TC = \frac{1}{F_e} \frac{\sum_{n=1}^{N_x} n ENV(n)}{\sum_{n=1}^{N_x} ENV(n)} \quad (2.18)$$

où N_x est le nombre d'échantillons du signal x considéré. TC est homogène à un temps.

2.2.5.4 Les descripteurs spectraux et/ou spectro-temporels

Les descripteurs spectraux visent à caractériser objectivement la répartition d'énergie des composantes spectrales. Ils sont souvent utilisés afin d'interpréter la structure des espaces perceptifs

de timbre. Dans un tel cas, leur objectif est de quantifier les aspects importants du spectre par une seule donnée numérique, qui pourra alors être confrontée avec la position des stimuli le long des dimensions de l'espace de timbre. Nous présenterons la plupart du temps des formulations des descripteurs spectraux à court-terme, qui ainsi considérés, prendront un caractère spectro-temporel (ce pourquoi nous avons regroupé les descripteurs spectraux et les descripteurs spectro-temporels dans une seule et même section).

2.2.5.4.1 Le Centre de Gravité Spectral (SC) Grey et Gordon sont les premiers à introduire, en 1978, le Centre de Gravité Spectral (SC) dans le contexte de l'étude du timbre musical [Grey et Gordon, 1978]. Ce dernier est directement corrélé à la sensation de "brillance" d'un son. Plus un son est "mat", plus son Centre de Gravité Spectral sera faible, plus il est "brillant", plus au contraire ce dernier sera élevé. Les spectres riches (i.e. contenant un grand nombre de composantes spectrales) seront caractérisés par des SC élevés et correspondront à des sons plutôt "métalliques". Cependant la réciproque n'est pas toujours vraie, SC peut être élevé alors que le spectre est pauvre (c'est le cas d'un son aigu de flûte ou de clarinette par exemple).

Considérons le spectre d'amplitude comme une variable aléatoire réelle définie sur un ensemble Ω , dont les valeurs sont les fréquences physiques x , et les amplitudes normalisées sont les probabilités d'observer ces fréquences $dP_X(x)$. Le spectre peut alors être caractérisé par ses différents moments. Le Centre de Gravité Spectral, μ , représente le moment d'ordre 1 (également appelé moyenne de la distribution) :

$$\mu = \int_{\Omega} x \, dP_X(x) \quad (2.19)$$

où :

- x représente les données observées, i.e. les fréquences, et
- $dP_X(x)$ est la probabilité d'observer x , compte-tenu des amplitudes $A(x)$, telle que $dP_X(x) = \frac{A(x)}{\int_{\Omega} A(x)}$.

Différentes formulations du Centre de Gravité Spectral peuvent être trouvées dans la littérature selon le type de représentation spectrale utilisé, les types d'échelles d'amplitude et de fréquence considérés, et l'éventuelle prise en compte de modèles physiologiques de l'audition (cf. [Marozeau, 2004, p. 79] pour une revue des différentes formulations du descripteur de Centre de Gravité Spectral utilisant de tels modèles). Parmi les représentations spectrales possibles, on distinguera la représentation temps-fréquence obtenue avec la TFDCT, qui implique de calculer SC à partir des coefficients de TFD effectuées à courts-termes, ou la représentation reposant sur un modèle sinusoïdal harmonique, qui implique de calculer SC à partir des amplitudes et fréquences instantanées des composantes harmoniques. La discrétisation de la dimension fréquentielle des représentations spectrales peut être effectuée suivant une échelle linéaire (Δf constant), une échelle logarithmique ($\frac{\Delta f}{f}$ constant), l'échelle des *mels* [Stevens et al., 1937], une échelle reposant sur les bandes critiques (mesurées en Barks [Fletcher, 1940] [Zwicker et Feldtkeller, 1981], ou selon leur équivalent, en ERB (Equivalent Rectangular Band) [Moore et Glasberg, 1983]). Les amplitudes des représentations spectrales peuvent être caractérisées suivant une échelle linéaire, l'échelle des décibels (dB), une échelle de puissance, une échelle dérivée de la sonie partielle [Zwicker et Fastl, 1990], etc.

L'intention première de Grey et Gordon était de vérifier la pertinence de l'interprétation physique de la dimension I de l'espace de timbre obtenu par Grey [Grey, 1977] (cf. section 2.2.4.2), corrélée à la nature de la distribution d'énergie spectrale (et notamment la largeur de bande). Ils ont comparé les facultés de différentes méthodes de calcul du Centre de Gravité Spectral à prédire les positions des stimuli le long de cette dimension [Grey et Gordon, 1978]. Les différences

entre ces méthodes sont liées à la manière dont la distribution d'énergie spectrale statique est estimée. Puisque les sons considérés étaient harmoniques (rappelons qu'ils ont été obtenus par synthèse additive), la représentation spectrale utilisée par les auteurs repose sur un spectre de raies harmoniques, défini : (1) en prenant les amplitudes maximales des enveloppes d'amplitude des composantes harmoniques au cours du temps, (2) les moyennes temporelles des enveloppes d'amplitude, et (3) les moyennes temporelles des enveloppes d'énergie. Deux unités différentes pour le spectre de raies ont été employées (linéaire et décibel). D'autre part, en s'inspirant des travaux de Zwicker et Scharf sur la sonie [Zwicker et Scharf, 1965], ils ont dérivé du spectre de raies obtenu deux autres formulations de l'énergie spectrale : un pattern d'excitation périphérique, et une fonction de sonie. Les confrontations des corrélations obtenues pour des espaces de timbre issus de différentes expériences perceptives les ont amenés aux conclusions suivantes (nous indiquons entre parenthèses la corrélation obtenue avec la dimension I de l'espace de Grey) :

- les amplitudes temporelles moyennes (modèle 2) fournissent de meilleurs résultats que les amplitudes crêtes (modèle 1), ainsi que l'énergie moyenne (modèle 3) ;
- la représentation linéaire ($r = 0.93$) induit une meilleure prédiction que la représentation en décibels ($r = 0.83$) ;
- la fonction de sonie fournit de meilleurs résultats ($r = 0.94$) que le pattern d'excitation périphérique ($r = 0.91$) ;
- la fonction de sonie permet une légère amélioration des corrélations ($r = 0.94$) par rapport à la représentation linéaire physique ($r = 0.93$).

Dans l'étude de Grey et Gordon, les deux meilleures représentations pour le calcul du Centre de Gravité Spectral sont donc la fonction de sonie dérivée du modèle de Zwicker et Scharf, et la représentation linéaire classique. Les travaux de Marozeau *et al.* tendent à confirmer que la prise en compte d'un modèle auditif adéquat permet d'améliorer les prédictions [Marozeau, 2004]. Le descripteur défini par Marozeau *et al.* repose sur celui proposé par Zwicker et Fastl, baptisé "acuité" [Zwicker et Fastl, 1990]. Le calcul de l'*acuité* repose sur le découpage du spectre en 24 bandes, de largeur 1 Bark. Son unité est l'*acum*. L'évaluation de l'amplitude spectrale est réalisée au moyen de la sonie spécifique, i.e. sonie en fonction de la fréquence sur une échelle de Barks, elle-même pondérée par une fonction qui "sert à prendre en compte le fait que l'acuité d'un bruit de faible bande croît fortement, de façon surprenante, pour des fréquences centrales élevées" [Zwicker et Fastl, 1990]. Le descripteur de Marozeau *et al.* apporte certaines améliorations à l'acuité de Zwicker et Fastl, en utilisant le modèle de sonie de Moore et Glasberg [Moore et Glasberg, 1996]. Un filtrage passe-bande, coupant les basses et les hautes fréquences, modélise les filtrages effectués par l'oreille externe et moyenne. La sélectivité fréquentielle de l'oreille est ensuite prise en compte par un banc de filtres passe-bandes ("Gammatones"). La puissance instantanée est calculée, traitée, puis mise à la puissance 0.3 (valeur qui découle de la loi de Stevens [Stevens, 1957]), afin de modéliser la fonction de sonie partielle représentant l'amplitude spectrale. Le Centre de Gravité Spectral est déterminé en pondérant la quantité obtenue par la sonie instantanée, i.e. la somme à travers toutes les bandes des sonies partielles. Son unité est exprimée en taux d'ERB (*ERB-rate*). Marozeau et de Cheveigné ont également proposé une version de ce descripteur visant à tenir compte de l'effet de la fréquence fondamentale sur la perception de la brillance [Marozeau et de Cheveigné, 2007]. La modification consiste à soustraire au Centre de Gravité Spectral, converti au préalable en Hz, la fréquence fondamentale du signal considéré, puis à le reconvertir en *ERB-rate*. Si le descripteur semble bien fonctionner, au sens prédictif, pour des stimuli simples et calibrés (synthèse additive), il n'a toutefois pas encore été validé à l'égard de sons musicaux, ou environnementaux.

Notons que dans l'étude de Grey et Gordon, si la prise en compte du modèle de sonie de Zwicker et Scharf apporte une amélioration par rapport à la représentation du spectre d'amplitude linéaire moyen ($r = 0.94$ par rapport à $r = 0.93$), cette dernière reste cependant subtile. De fait,

un grand nombre de formulations du Centre de Gravité Spectral rencontrées dans les études sur le timbre, ne prennent pas en compte de variables psychophysiologiques (cf. par ex. [Beauchamp, 1982; McAdams *et al.*, 1995; Kendall *et al.*, 1999]). De plus, les descripteurs reposant sur des modèles physiologiques sont difficilement inversibles, ce qui dans une optique de contrôle du timbre par analyse-synthèse est problématique. Au vu de ces considérations, nous avons opté pour une représentation classique du spectre d'amplitude (linéaire ou puissance). Les sons musicaux étant en général non stationnaires (facteur d'autant plus marqué si l'on s'intéresse non plus à des sons isolés mais à des phrases ou des séquences musicales), il semble judicieux de s'intéresser à l'évolution temporelle du Centre de Gravité Spectral instantané $SC(t)$. De plus, Krimphoff *et al.* ont montré qu'un calcul du SC global $\overline{SC}(t)$ à partir de la moyenne des SC instantanés, dérivés de la TFDCT, offrait de meilleures prédictions qu'un calcul reposant sur les amplitudes maximales ou moyennes (cf. modèles (1) et (2) de Grey et Gordon ci-dessus). Ils ont pu ainsi expliquer 88 % de la variance [$r(20)=0.94$, $p<0.01$] de la première dimension de l'espace Krumhansl [Krumhansl, 1989].

Dans le cas de sons quasi-périodiques (la plupart des sons instrumentaux, hormis les sons percussifs de types gongs ou cloches), une première approche consiste à ne considérer que le spectre (quasi-)harmonique. Cette approche correspond à la conception classique du timbre (cf. section 2.2.1) considérant que seule la partie stationnaire des sons musicaux n'a de répercussion sur la "qualité" musicale. Nous intitulerons **Centre de Gravité Spectral Harmonique (HSC)** le descripteur associé à cette approche. Sa formulation peut s'établir ainsi :

$$HSC(t) = \frac{\sum_{h=1}^H f_h(t) A_h(t)}{\sum_{h=1}^H A_h(t)} \quad (2.20)$$

où h représente le rang des composantes harmoniques, H est le nombre total de composantes considérées, $A_h(t)$ et $f_h(t)$ sont respectivement les amplitude et phase instantanées de la h^{eme} composante harmonique. L'unité du descripteur HSC est le Hz.

Toutefois, les signaux naturels ne sont rarement, voire jamais, déterministes. Outre les phases d'attaques et de décroissance instationnaires, il existe la plupart du temps une composante stochastique large-bande, qui elle aussi apporte sa contribution à la distribution d'énergie spectrale. C'est le cas des sons de clarinette par exemple, où le souffle du musicien génère un bruit filtré par le résonateur, qui très certainement contribue à l'identité timbrale de l'instrument. Une représentation des sons de type déterministe/stochastique telle que proposée par Smith et Serra [Smith, III] semble donc plus appropriée. Un modèle similaire a été employé par exemple par Ystad [Ystad, 1998] pour la synthèse de sons de flûte traversière, ainsi que dans le cadre de ce travail pour la synthèse de sons de clarinette (cf. section 2.2.5.5). Afin de tenir compte aussi bien des parties déterministe et stochastique des sons instrumentaux, le Centre de Gravité Spectral peut être calculé à partir des coefficients de la TFD. Selon cette approche, le **Centre de Gravité Spectral (SC)** défini "à court terme", i.e. à partir de TFD effectuées sur des trames ce courtes durées, obtenues par fenêtrage du signal analysé (cf. section 2.2.5.1), est donné par l'équation suivante :

$$SC(n) = \frac{\sum_{k=1}^K f(k) A_n(k)}{\sum_{k=1}^K A_n(k)} \quad (2.21)$$

où $A_n(k)$ est le k^{eme} coefficient de la TFD associée à la trame du signal analysé $x_n(m)$, centrée sur l'horizon temporel d'analyse n , et $f(k)$ est la fréquence associée à la k^{eme} composante spectrale (cf. section 2.2.5.1). K correspond à l'indice de la dernière composante spectrale considérée. Si tout le spectre est considéré et que le nombre de points de calcul N de la TFD est pair, alors $K = 1 + \frac{N}{2}$ (prendre un $K < 1 + \frac{N}{2}$ permet de calculer SC en ne considérant qu'une seule partie

du spectre).

La définition 2.21 présente un inconvénient lorsque le bruit⁵⁸ est prédominant pour de très faibles amplitudes. Cette situation se rencontre au sein d'une même note, au niveau des phases d'attaque et de décroissance du son, ainsi qu'au sein d'une séquence musicale, au niveau des transitions entre deux notes détachées. Beauchamp a proposé l'ajout d'un seuil, au dénominateur de l'expression 2.21, dont le rôle est de stabiliser numériquement la valeur de SC lorsque le dénominateur devient très faible [Beauchamp, 1982]. Considérant cette amélioration l'expression de SC devient :

$$SC(n) = \frac{\sum_{k=1}^K f(k) A_n(k)}{b_0 + \sum_{k=1}^K A_n(k)} \quad (2.22)$$

où b_0 est le paramètre de seuil. Remarquons que lorsque $b_0 = 0$, l'expression 2.22 est équivalente à l'expression 2.21. Une comparaison des valeurs du Centre de Gravité Spectral, calculé pour un crescendo de clarinette avec et sans prise en compte du seuil stabilisateur b_0 , est proposée dans le chapitre 3, à la figure 3.4.

En pratique, les seuils d'amplitude spectrale sont généralement donnés en décibels. Si l'on désire une dynamique de dyn dB (i.e. la différence maximale d'amplitude entre les raies spectrales considérées correspond à dyn dB), le seuil d'amplitude linéaire b_0 correspondant pourra être obtenu par :

$$b_0 = M \exp\left(-\frac{dyn \times \ln(10)}{20}\right) \quad (2.23)$$

où $M = \max_{n,k} (A_n(k))$ est l'amplitude maximale de la TFDCT associée au son, et $\ln$ est le logarithme népérien. Une dynamique de 60 dB paraît être un bon compromis afin de conserver l'information utile et s'affranchir du bruit parasite.

Il existe également une estimation de SC obtenue dans le domaine temporel à partir de l'enveloppe RMS du signal et de l'enveloppe RMS de sa dérivée [Verfaille, 2003]. Cette dernière peut avoir des avantages dans les applications temps-réel.

Il incombera à l'expérimentateur, l'entreprise de dériver du Centre de Gravité Spectral instantané, tel que formulé par l'équation 2.22, des statistiques descriptives (par ex. moyenne, variance), qui permettront de considérer certains aspects du descripteur à un niveau plus global. Comme nous l'avons vu, le Centre de Gravité Spectral moyen $\overline{SC(n)}$ est souvent employé afin d'interpréter la structure des espaces perceptifs de timbre.

Propriétés du Centre de Gravité Spectral Dans le cas des sons d'instruments musicaux, le Centre de Gravité Spectral s'avère souvent covarier avec l'amplitude RMS des signaux. Ceci est lié au fonctionnement acoustique des instruments, une augmentation de l'intensité allant généralement de pair avec un accroissement de la richesse spectrale, aspect sur lequel nous reviendrons dans le cas de la clarinette dans les chapitres 3 et 4. En revanche, il ne faut pas perdre de vue que sur le plan mathématique, SC est indépendant de l'amplitude RMS [Beauchamp, 1982]. La démonstration en est simple. Si nous multiplions un signal par un certain gain, l'amplitude RMS changera, tandis que SC restera inchangé, puisque le dénominateur de son expression présente un terme de normalisation (cf. équation 2.22). Inversement, on peut très bien concevoir deux signaux qui disposent d'une même amplitude RMS, mais de différents SC ; par exemple, un signal dont

⁵⁸Le mot bruit est pris ici dans une différente acception que celle utilisée quelques lignes plus haut, il s'agit ici du bruit à un niveau "signal", et non du bruit constitutif des sons instrumentaux, tel que le bruit de souffle d'un clarinettiste.

l'enveloppe spectrale décroîtrait linéairement dans la bande de fréquence $[0; f_c]$ avec la pente $-\alpha$ présenterait un SC plus faible qu'un signal dont l'enveloppe spectrale croîtrait linéairement dans la bande de fréquence $[\frac{F_e}{2} - f_c; \frac{F_e}{2}]$ avec la pente α . Cependant, les deux signaux auraient la même amplitude RMS. L'indépendance de SC avec l'intensité dans la formulation du descripteur est d'importance puisqu'en général des descripteurs indépendants de l'intensité et de la hauteur sont recherchés pour décrire le timbre.

Robustesse du Centre de Gravité Spectral Le Centre de Gravité Spectral est le seul descripteur qui apparaît de manière récurrente dans les études du timbre musical en tant que bon prédicteur de notre perception (par ex. [Grey et Gordon, 1978; Wessel, 1978; McAdams *et al.*, 1995, 1999; Kendall *et al.*, 1999; Caclin *et al.*, 2005]).

De plus, il semble également très bien adapté à la caractérisation des différentes qualités de timbre qui peuvent être produites sur un même instrument, pour des notes obtenues selon différents modes de jeu :

- Dans l'étude de Grey, il traduit bien les différences entre des sons de saxophone joués *piano* (faible SC, son plutôt "sourde") et *forte* (SC élevé, son plutôt "riche" et "brillant"), ainsi que celles entre des sons de violoncelle obtenus près de la touche (son plutôt "sourde"), "normalement", et près du chevalet (son plutôt "métallique"). Notons que lorsque les enveloppes spectrales des instruments sont permutées (par ex. trompette $\leftrightarrow$ trombone), tout en conservant les largeurs de bande initiales (seules les amplitudes des harmoniques communs sont changées), les positions des sons le long de la dimension I corrélée avec SC, sont également permutées [Grey et Gordon, 1978]. Vis-à-vis de cette dimension, ceci tend donc à conférer une importance plus grande au critère associé au Centre de Gravité Spectral qu'en celui associé à la largeur de bande ;
- Il permet également de caractériser des sons de guitare produits pour différents points de pincement de la corde (un son joué près du chevalet sonne "métallique", tandis qu'un son joué près des touches sonne plus "sourde") [Traube *et al.*, 2003] ;
- Nous verrons au chapitre 4 qu'il rend aussi bien compte des distinctions entre différents timbres de clarinette, obtenus pour diverses combinaisons de pression d'alimentation et de pince de l'instrumentiste (les sons associés à une faible pression et une forte pince tendent à être peu intenses et "sourds", tandis que ceux associés à des fortes pressions et une pince faible tendent à être intenses et "brillants").

Le Centre de Gravité Spectral ne semble pas seulement relever du timbre-qualité, mais semble également façonner le timbre-identité d'un instrument⁵⁹. En effet, l'étude sur les sons de trompette menée par Risset prouve que l'un des ingrédients clefs de ce cuivre réside dans la relation entre l'intensité et la richesse spectrale [Risset, 1966], relation avec laquelle le Centre de Gravité Spectral est corrélé (cf. section 2.5).

Dans le contexte de la synthèse sonore, Beauchamp [Beauchamp, 1982] a montré que le Centre de Gravité Spectral instantané pouvait s'avérer utile afin d'ajuster automatiquement les paramètres de contrôle d'un modèle de synthèse de sons instrumentaux reposant sur la distorsion non-linéaire et le filtrage du signal. La méthode proposée par l'auteur lui a permis de resynthétiser des sons de cornets et de saxophone avec une erreur moyenne (estimée à partir des différences spectrales instantanées) allant de 4 à 26 %, tandis que les erreurs obtenues avec la technique de synthèse FM se situaient entre 36 et 72 %.

Les études sur le timbre mentionnées précédemment reposent sur des notes d'instruments considérées de manière isolée. D'un point de vue général, il paraît important d'analyser si les des-

⁵⁹Les notions de timbre-qualité et de timbre-identité ont été définies dans la section 2.1.

cripteurs de timbre permettent de prédire l'influence du timbre sur la perception musicale (6^{ème} étape de la méthodologie exposée à la section 2.2.4.1). Wessel [Wessel, 1978] a démontré que des changements de brillance entre les notes d'une séquence musicale pouvaient modifier la perception de l'organisation temporelle de la séquence. La figure 2.14 illustre son expérience. La séquence musicale est formée par la répétition d'un motif de trois notes ascendantes. Les notes représentées par un même symbole (x ou o) possèdent une même brillance. Lorsque les différences de brillance entre les notes sont faibles, la séquence est perçue comme la répétition d'un triplet ascendant. En revanche, lorsque les différences de brillance entre les notes sont larges, l'oreille tend à suivre les sons qui ont la même brillance, et la plupart des auditeurs perçoivent alors deux lignes mélodiques descendantes, constituées par les triplets de notes de même brillance, chaque ligne ayant sa propre identité timbrale. Ce type d'effet est répertorié sous le nom de *ségrégation auditive* dans la littérature [Bregman, 1994].

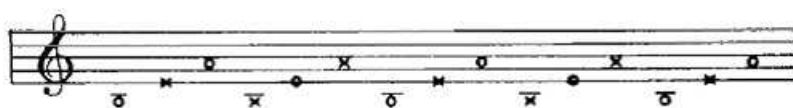


FIG. 2.14 – Démonstration de l'influence de la brillance sur la ségrégation de flux auditif par Wessel [Wessel, 1978]. Les notes symbolisées par o et x ont différents niveaux de brillance. Lorsque les différences sont grandes l'intégration séquentielle est modifiée, le système auditif perçoit une succession de lignes descendantes entrelacées dont les notes ont des timbres similaires.

Ainsi, des modifications de timbre (brillance) peuvent transformer l'intégration séquentielle d'une séquence musicale, au même titre que les hauteurs, comme c'est le cas pour un procédé compositionnel connu, le *contrepoint*. Largement utilisé dans la musique baroque, le contrepoint est l'art de superposer plusieurs lignes mélodiques indépendantes et de les faire progresser ensemble [Siron, 2004, p. 729]. Nécessitant souvent une certaine virtuosité de l'instrumentiste, Bach l'utilisait afin de créer une sensation de polyphonie à partir d'instruments monodiques. Dans le contrepoint, la ségrégation des voix est réalisée grâce à des facteurs qui renforcent les liens horizontaux entre certaines notes, tout en affaiblissant les liens verticaux entre d'autres notes [Bregman, 1994, p. 496]. Si les principaux facteurs qui induisent la ségrégation des voix reposent sur la hauteur et le rythme⁶⁰, l'expérience de Wessel prouve que le timbre peut également y participer. Dans les *sonates en trio pour orgue* de Bach, les trois différentes voix doivent être jouées sur des claviers différents, ou des pédaliers ayant des réglages de leviers différents [Bregman, 1994, p. 498]. Le fait que les trois lignes disposent de timbres différents garantit ainsi qu'elles puissent être distinguées et suivies par les auditeurs, même si les hauteurs se croisent.

Pour toutes les raisons mentionnées précédemment, nous avons porté une attention particulière à l'influence des variations de brillance sur la perception musicale. La plupart des études se concentrent sur des différences de brillance entre différentes notes. Nous avons choisi d'étudier l'effet des variations temporelles de brillance durant la production même des sons musicaux⁶¹ (cf. section 2.1.7) ; Les questions suivantes seront abordées dans les chapitres 5 et 6 : Est-ce que les interprètes reproduisent fidèlement les variations de brillance lorsqu'ils jouent selon une même intention musicale ? Est-ce que ces variations sont modifiées lorsque l'intention expressive change ? Quelles sont leurs influences sur les préférences musicales des auditeurs ?

⁶⁰ par ex. utiliser des petits intervalles de hauteurs pour les notes associées à une même voix, éviter les attaques synchrones (typiquement en alternant dans le temps les notes des deux voix), éviter l'utilisation de mouvements parallèles, éviter de croiser les lignes mélodiques.

⁶¹ Il s'agit de variations du timbre-qualité.

Limites du Centre de Gravité Spectral Bien entendu, le Centre de Gravité Spectral à lui seul ne détermine pas le timbre. Des sons ayant des Centres de Gravité Spectraux égaux mais des timbres différents peuvent être générés de manière simple. Une condition suffisante (mais non nécessaire) à l'égalité des Centres de Gravité Spectraux de deux sons est que leur distribution d'énergie spectrale soit symétrique et de même axe de symétrie. Les [Exemples Sonores n^{os} 31 et 32](#) ont été générés suivant ce principe. Leurs timbres sont dissemblables, pourtant ils disposent de Centres de Gravité Spectraux identiques. Le facteur qui est à la source des variations de timbre entre ces deux sons est l'irrégularité spectrale, facteur que nous décrirons dans la section [2.2.5.6](#). Le premier son présente en effet une enveloppe spectrale plate, tandis que le second présente une enveloppe spectrale en forme de "chapeau".

2.2.5.4.2 L'Étalement Spectral (SS) Nous avons vu que du point de vue statistique, le Centre de Gravité Spectral peut être vu comme le moment d'ordre 1 μ_1 (moyenne) de la distribution d'énergie spectrale X (cf. équation [2.19](#)). Selon cette même approche, l'Étalement Spectral, correspond à la variance σ^2 , ou moment centré d'ordre 2, de la distribution d'énergie spectrale. L'Étalement Spectral est donc une mesure de la dispersion de l'énergie spectrale autour de sa moyenne, le Centre de Gravité Spectral. Sa formulation générale peut s'établir ainsi :

$$\sigma^2 = \int_{\Omega} (x - \mu_1)^2 dP_X(x) \quad (2.24)$$

où les quantités sont les mêmes que dans l'équation [2.19](#). La racine carrée de la variance, σ , correspond à l'écart-type de la distribution X .

De la même manière que pour SC (cf. équation [2.22](#)), nous définirons l'Étalement Spectral à court-terme (SS), à partir de la TFDCT. Un seuil stabilisateur b_0 a été ajouté au dénominateur de l'expression classique de ce descripteur, pour les mêmes raisons que pour SC. L'Étalement Spectral à court-terme (SS) est donné par :

$$SS(n) = \frac{\sum_{k=1}^K (f(k) - SC(n))^2 A_n(k)}{b_0 + \sum_{k=1}^K A_n(k)} \quad (2.25)$$

où $SC(n)$ correspond au Centre de Gravité Spectral à l'horizon temporel d'analyse n , et les autres quantités sont définies comme dans l'équation [2.22](#). L'Étalement Spectral est donc homogène au carré de la fréquence (son unité est le Hz^2). Sa racine carrée, i.e. l'écart-type de la distribution spectrale, est elle homogène à une fréquence.

En pratique, pour les sons instrumentaux, l'Étalement Spectral et le Centre de Gravité Spectral varient souvent conjointement ; en général, plus un son est riche, plus il est brillant, et par la même occasion plus la dispersion autour de la moyenne de la distribution spectrale est grande. Dans [\[Marozeau, 2004, p. 84\]](#), Marozeau signale une forte corrélation ($r = 0.9$) entre l'Étalement Spectral et la troisième dimension de l'espace de McAdams *et al.* [\[McAdams et al., 1995\]](#), plus élevée que celle obtenue par les auteurs avec un descripteur spectro-temporel, le Flux Spectral (cf. définition ci-dessous), qui n'explique que 29% de la variance [$r=0.54$, $p<0.05$].

2.2.5.4.3 Le Coefficient d'Asymétrie Spectrale (SSK) Le coefficient d'asymétrie γ_1 (*skewness* en anglais) permet de caractériser la forme d'une distribution statistique. Il est calculé à partir du moment centré d'ordre 3, noté μ_3 , et de l'écart-type σ :

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\int_{\Omega} (x - \mu_1)^3 dP_X(x)}{\sigma^3} \quad (2.26)$$

où les quantités sont définies comme dans l'équation 2.24. Le coefficient d'asymétrie est un nombre sans dimension. Il indique le degré d'asymétrie de la distribution :

- si $\gamma_1 < 0$, l'énergie est concentrée à droite de la moyenne,
- si $\gamma_1 = 0$, la distribution est centrée sur la moyenne, et
- si $\gamma_1 > 0$, l'énergie est concentrée à gauche de la moyenne.

En pratique, nous calculerons le Coefficient d'Asymétrie Spectrale à court-terme (SSK) ainsi :

$$SSK(n) = \frac{1}{SS(n)^{\frac{3}{2}}} \frac{\sum_{k=1}^K (f(k) - SC(n))^3 A_n(k)}{b_0 + \sum_{k=1}^K A_n(k)} \quad (2.27)$$

où les quantités sont définies comme dans l'équation 2.25.

2.2.5.4.4 Le Coefficient d'Aplatissement Spectral (SKU) Le coefficient d'aplatissement γ_2 (*kurtosis* en anglais) d'une distribution statistique donne une mesure de son aplatissement autour de sa moyenne. Il est calculé à partir du moment centré d'ordre 4, noté μ_4 , et de l'écart-type σ :

$$\gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{\int_{\Omega} (x - \mu_1)^4 dP_X(x)}{\sigma^4} \quad (2.28)$$

où les quantités sont définies comme dans l'équation 2.24. Le coefficient d'aplatissement est un nombre sans dimension, qui est toujours supérieur à 1 (cf. [Saporta, 1990, p. 27]). Il indique le degré de sélectivité fréquentielle de la distribution autour de sa moyenne, et mesure l'importance des "queues" de distribution :

- $\gamma_2 = 3$ pour une distribution normale,
- $\gamma_2 < 3$ pour une distribution plus plate,
- $\gamma_2 > 3$ pour une distribution plus sélective (cf. [Peeters, 2004]).

En pratique, nous calculerons le Coefficient d'Aplatissement Spectral à court-terme (SKU) ainsi :

$$SKU(n) = \frac{1}{SS(n)^2} \frac{\sum_{k=1}^K (f(k) - SC(n))^4 A_n(k)}{b_0 + \sum_{k=1}^K A_n(k)} \quad (2.29)$$

où les quantités sont définies comme dans l'équation 2.25.

2.2.5.4.5 Le Roulement Spectral (SRO) Le Roulement Spectral est la fréquence telle que 95% de l'énergie du spectre soit contenue en dessous [Peeters, 2004] (on trouve également des formulations du Roulement Spectral reposant sur un seuil de 85%). Il peut représenter un moyen de détecter le seuil fréquentiel à partir duquel le signal ne contient plus que du bruit. Nous définissons le Roulement Spectral à court-terme SRO comme la fréquence $f(R)$ associée au R^{eme} coefficient de la TFD, telle qu'à l'horizon temporel d'analyse n :

$$\sum_{k=1}^R A_n(k)^2 = 0.95 \sum_{k=1}^K A_n(k)^2 \quad (2.30)$$

où $K = 1 + \frac{N}{2}$ lorsque N est pair, et les autres quantités sont définies comme dans l'équation 2.25.

2.2.5.4.6 Le Flux Spectral (SF) Le Flux Spectral est une mesure du degré de variation du spectre d'amplitude au cours du temps. Introduit qualitativement par Grey afin d'interpréter la seconde dimension de l'espace de timbre obtenu dans son étude [Grey, 1977] (cf. section 2.2.4.2), ainsi que par Krumhansl [Krumhansl, 1989], pour expliquer la troisième dimension de sa configuration MDS, la première mesure quantitative du Flux Spectral voit le jour grâce aux travaux de Krimphoff *et al.* [Krimphoff *et al.*, 1994] et McAdams *et al.* [McAdams *et al.*, 1995]. Le Flux Spectral correspond à la corrélation moyenne entre les spectres d'amplitude obtenus pour des trames temporelles adjacentes $x_n(m)$ et $x_{n-1}(m)$, lors de la TFDCT. Le Flux Spectral est donc un descripteur global du son. McAdams *et al.* propose la mesure suivante, reposant sur le coefficient de corrélation linéaire de Pearson :

$$SF = \frac{1}{N_w} \sum_{n=2}^{N_w} |r_{n,n-1}| \quad (2.31)$$

où N_w représente le nombre total de trames (i.e. nombre total de fenêtres d'analyse), et $r_{n,n-1}$ est le coefficient de corrélation de Pearson entre les spectres d'amplitude adjacents $A_n(k)$ et $A_{n-1}(k)$, respectivement pris aux horizons temporels d'analyse n et $n - 1$.

Le coefficient de corrélation de Pearson mesure le caractère plus ou moins linéaire d'un nuage de points (cf. [Saporta, 1990, p. 126]). Il est défini comme suit :

$$r = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \quad (2.32)$$

où $\bar{x}$ et $\bar{y}$ sont respectivement les moyennes des variables x et y , et s_x et s_y leurs écarts-types, définis, rappelons le, par le moment centré d'ordre 2 :

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.33)$$

$$s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.34)$$

Le numérateur de l'expression 2.32 est la covariance observée pour des signaux comparés sans déphasage. Une formulation équivalente du coefficient de corrélation de Pearson est donc la suivante :

$$r = \frac{1}{n} \frac{C_{xy}(0)}{\sqrt{C_{xx}(0)C_{yy}(0)}} \quad (2.35)$$

où C_{xy} désigne la fonction de covariance croisée (i.e. corrélation croisée centrée), et C_{xx} et C_{yy} sont les fonctions d'autocovariance de x et y . Les valeurs de ces fonctions sont prises en 0, de sorte à obtenir une mesure de liaison entre les signaux x et y en phase deux à deux. Le coefficient de corrélation de Pearson varie par définition entre -1 et 1. Lorsque sa valeur absolue est égale à 1, il existe une relation linéaire exacte entre x et y , telle que $\forall i \ ax_i + by_i + c = 0$, où a et b sont des réels.

Le fait de prendre la valeur absolue de r dans l'expression du Flux Spectral donnée par l'équation 2.31 entraîne ce dernier à ne varier qu'entre 0 et 1. Les cas extrêmes sont les suivants :

- si $SF = 0$, alors les corrélations entre spectres d'amplitudes adjacents sont nulles, i.e. les comportements des composantes spectrales au cours du temps sont très dissemblables ;

- si $SF = 1$, alors les corrélations entre spectres d'amplitudes adjacents sont maximales (elles expriment une relation linéaire soit positive, soit négative), i.e. les comportements des composantes spectrales au cours du temps sont synchrones.

Le descripteur de Flux Spectral s'est avéré être relativement bien corrélé ($[r=0.54, p<0.05]$, 29% de la variance expliquée) avec la troisième dimension de l'espace de McAdams *et al.* (cf. [McAdams *et al.*, 1995]). Cependant, nous avons vu ci-dessus que Marozeau signalait une corrélation bien plus importante de cette dimension avec l'Étalement Spectral [Marozeau, 2004, p. 84]. La corrélation du Flux Spectral avec la troisième dimension de l'espace de timbre de Krumhansl est également moyenne, contrairement à ce que présupposait l'auteur. Krimphoff *et al.* ont en effet montré qu'il ne rendait compte que de 34% de la variance le long de cette dimension, contre les 73% expliqués par l'Irrégularité Spectrale [Krimphoff *et al.*, 1994]. Cependant, rien n'écarterait le fait que d'autres formulations du Flux Spectral ne pourraient pas conduire à de meilleurs résultats.

2.2.5.5 Analyse/Synthèse additive de sons (quasi-)harmoniques

Certains types de descripteurs s'attachent à décrire le comportement de la structure harmonique des sons instrumentaux. Ils se reposent donc, non pas sur tout le spectre, mais seulement sur l'information relative aux composantes harmoniques⁶². Nous les appellerons à ce titre *descripteurs "harmoniques"*. Leur calcul nécessite de connaître les amplitudes et fréquences des différentes composantes harmoniques.

L'analyse et la synthèse sonore sont généralement des processus symétriques, i.e. les paramètres issus de l'analyse servent également de paramètres de contrôle au modèle. Nous en profiterons donc pour introduire dans cette section le modèle d'analyse-synthèse additive utilisé dans plusieurs de nos études, afin d'analyser, et/ou resynthétiser et/ou transformer des sons musicaux harmoniques. Une description des techniques classiques d'analyse-synthèse additive pourra être trouvée dans la référence [Roads, 1998]. Rappelons que la technique d'analyse-synthèse additive est bien adaptée aux sons de clarinette puisque ceux-ci sont essentiellement harmoniques. Il est donc logique d'en analyser la structure à partir d'une décomposition élémentaire en sinusoïdes dont les fréquences sont dans des rapports entiers (ou presque) les unes par rapport aux autres.

La technique d'analyse-synthèse additive que nous avons utilisée repose sur la même heuristique que la Synthèse par Modèles Spectraux (SMS) proposée dans [Smith, III], qui consiste à décomposer un son suivant une partie *déterministe* (composantes à bande étroite), et une partie *stochastique* (par définition large bande). Suivant un tel modèle, un signal $s(t)$ peut s'écrire ainsi :

$$s(t) = \sum_{h=1}^H A_h(t) \cos[2\pi \int_0^t f_h(\tau) d\tau + \phi_h(0)] + r(t) \quad (2.36)$$

où $A_h(t)$ et $f_h(t)$ sont respectivement les amplitudes et fréquences instantanées de la h^{eme} composante harmonique parmi H . $\phi_h(0)$ représente la phase de la composante h à l'instant initial, et $r(t)$ est la partie stochastique du signal, également appelée *bruit* ou *résidu*. Notons que cette méthode de synthèse est particulièrement adaptée pour modifier certains aspects du son (par ex. la durée par expansion ou compression temporelle, mais également le timbre, cf. chapitre 6), puisque les amplitudes et les fréquences instantanées peuvent être contrôlées de manière individuelles. Notons que dans le cas de la clarinette, la structure harmonique des sons est complétée par des bruits auxiliaires large bande (par exemple, les bruits de souffle ou de clefs), qui contribuent aussi à l'identité de l'instrument. Ces derniers sont contenus dans le résidu.

⁶²Les observations restent valables pour des *composantes quasi-harmoniques*.

L'étape d'analyse de notre modèle d'analyse-synthèse additive n'est pas effectuée de la même façon que dans la Synthèse par Modèles Spectraux [Smith, III]. La technique d'analyse employée dans la SMS repose sur le suivi des partiels dans le plan temps-fréquence. Dans notre cas, le son est analysé par un banc de filtres passes-bandes dont les fréquences centrales sont ajustées de façon à correspondre aux fréquences des différents composantes quasi-harmoniques. Ceci permet d'associer à chaque composante harmonique h , un signal analytique à bande étroite $z_h(t)$. Nous baptiserons à ce titre cette technique *Synthèse par Signal Analytique* ou SAS (*Synthesis based on Analytic Signal*). Les amplitudes $A_h(t)$ et phases $\phi_h(t)$ instantanées des composantes quasi-harmoniques sont ensuite dérivées des signaux analytiques comme dans les équations 2.11. La fréquence instantanée $f_h(t)$ associée à la h^{eme} composante quasi-harmonique est obtenue par :

$$f_h(t) = \frac{1}{2\pi} \frac{d\phi_h(t)}{dt} \quad (2.37)$$

En pratique, dans le cas de signaux discrets, la dérivée de la phase est approximée par le rapport entre la différence des phases successives et l'intervalle temporel les séparant, i.e. la période d'échantillonnage T_e . Ceci implique au préalable de dérouler la phase (qui sinon est définie $[2\pi]$). La fréquence instantanée $f_h(n)$ à l'instant n découle alors de l'équation suivante :

$$f_h(n) = \frac{F_e}{2\pi} (\phi_h(n+1) - \phi_h(n)) \quad (2.38)$$

où ϕ_h est la phase déroulée.

Il est important de signaler que les amplitudes et fréquences instantanées des composantes n'auront un sens physique et perceptif pour la composante (i.e. son amplitude peut être associée à l'intensité, et sa fréquence à la hauteur) que si les signaux qui leurs ont été associées sont bien quasi-monochromatiques, i.e. que les bandes étroites dans lesquelles sont définies les signaux analytiques ne contiennent bien qu'une seule composante. Le résidu $r(t)$ est obtenu par soustraction du son originel avec la partie déterministe dans le domaine temporel (il n'est pas effectué de modélisation du bruit par un bruit blanc gaussien comme pour la SMS). Puisque les signaux analytiques sont à bandes étroites, les éventuelles variations transitoires des amplitudes instantanées $A_h(t)$ (pouvant avoir lieu pendant la phase d'attaque) sont lissées. De par la procédure, le résidu peut donc contenir, en plus du bruit large bande du signal, une partie des attaques des composantes harmoniques analysées. La SAS permet de suivre d'éventuelles modulations de fréquence d'une composante, à condition que ces dernières restent bien inférieures à la largeur des filtres passes-bandes (égale à l'intervalle de fréquence entre chaque composante, i.e. la fréquence fondamentale f_0). Des modulations fréquentielles induites par des effets musicaux comme le *vibrato*, typiquement de l'ordre de quelques Hertz, sont donc parfaitement détectables. Cependant, il est important de noter que si la valeur de la fréquence de modulation est "grande", une modulation d'amplitude (trémolo) due à l'analyse pourra apparaître dans l'amplitude instantanée, en raison de la forme de la réponse fréquentielle du filtre passe-bande servant à localiser les différentes composantes. Les séquences musicales analysées dans le cadre de ce travail ne contiennent pas ou très peu de notes avec vibrato (cf. chapitres 5 et 6). Lorsque c'est le cas, ce dernier reste très subtil. Le modèle d'analyse-synthèse SAS que nous venons de décrire s'avère donc très bien adapté aux sons étudiés. Des exemples sonores seront présentés au cours du chapitre 6.

2.2.5.6 Les descripteurs "harmoniques"

2.2.5.6.1 Le Centre de Gravité Spectral "harmonique" (HSC) L'expression du Centre de Gravité Spectral "harmonique" (HSC) est donnée par l'équation 2.20.

2.2.5.6.2 Les Centres de Gravité Spectraux Impair (OSC) et Pair (ESC) Puisque certains instruments peuvent présenter des différences de comportements entre les composantes harmoniques impaires et paires, il peut être intéressant de calculer de manière séparée les Centres de Gravité Spectraux relatifs aux distributions spectrales de chaque type de composante. Les Centres de Gravité Spectraux Impair (OSC) et Pair (ESC) instantanés sont définis de la manière suivante [Barthet *et al.*, 2005] :

$$\begin{aligned} OSC(t) &= \frac{\sum_{h=0}^{\lceil \frac{H}{2} \rceil - 1} f_{2h+1}(t) A_{2h+1}(t)}{b_0 + \sum_{h=1}^H A_h(t)} \\ ESC(t) &= \frac{\sum_{h=1}^{\lfloor \frac{H}{2} \rfloor} f_{2h}(t) A_{2h}(t)}{b_0 + \sum_{h=1}^H A_h(t)} \end{aligned} \quad (2.39)$$

où $\lfloor \cdot \rfloor$ et $\lceil \cdot \rceil$ désignent respectivement les parties entières par défaut et par excès (utilisées afin de considérer un nombre correct d'harmoniques quelque soit H). b_0 est le seuil introduit dans l'expression 2.22, et les autre quantités sont définies comme dans l'équation 2.20.

2.2.5.6.3 L'Irrégularité Spectrale (IRR) L'Irrégularité Spectrale (IRR) est un descripteur qui permet de décrire le degré d'irrégularité de l'enveloppe spectrale [Krimphoff *et al.*, 1994]. L'Irrégularité Spectrale est un nombre sans dimension. Plus sa valeur est grande, plus l'irrégularité de l'enveloppe spectrale est grande.

Le paramètre d'Irrégularité Spectrale, initialement proposé par Krimphoff *et al.*, mesure les écarts des amplitudes des harmoniques par rapport à une enveloppe spectrale globale, calculée en prenant la moyenne de trois raies adjacentes dont les amplitudes sont exprimées en décibels. Ce descripteur, noté ici *IRRKRI*, est bâti selon une échelle logarithmique. Son expression en fonction du temps est donnée par :

$$IRRKRI(t) = \log_{10} \left(\sum_{h=2}^{H-1} \left| 20 \log_{10}(A_h(t)) - \frac{20 \log_{10}(A_{h+1}(t)) + 20 \log_{10}(A_h(t)) + 20 \log_{10}(A_{h-1}(t))}{3} \right| \right) \quad (2.40)$$

où $\log_{10}$ est le logarithme en base 10, et les autre quantités sont définies comme dans l'équation 2.20.

L'Irrégularité Spectrale *IRRKRI* s'est révélée être fortement corrélée [$r(20)=-0.85$, $p<0.01$] avec la troisième dimension perceptive de l'espace de Krumhansl *et al.*, dont elle explique 73% de la variance [Krimphoff *et al.*, 1994]. Ce descripteur permet notamment de discriminer des instruments dont les composantes harmoniques présentent toutes des amplitudes élevées (par ex. trompette), de ceux dont les amplitudes des harmoniques présentent des disparités d'une composante à l'autre (par ex. clarinette, qui dispose principalement d'harmoniques impairs).

Toutefois, on notera que si l'enveloppe spectrale est parfaitement plate (cas difficilement probable pour les sons naturels, mais concevable pour des signaux complexes synthétiques dont les amplitudes des différents harmoniques seraient toutes égales), le descripteur, tel que calculé par l'expression 2.40, reste indéterminé. Il tend vers moins l'infini. Kendall et Carterette ont proposé une mesure de l'Irrégularité Spectrale sur une échelle linéaire, notée *IRRKEN*, qui permet d'éviter cette incohérence [Kendall et Carterette, 1996] :

$$IRRKEN(t) = \sum_{h=2}^{H-1} \left| A_h(t) - \frac{A_{h+1}(t) + A_h(t) + A_{h-1}(t)}{3} \right| \quad (2.41)$$

où les quantités sont définies comme dans l'équation 2.20.

Jensen a également établi une autre définition en utilisant le rapport de la somme des carrés des différences d'amplitudes des composantes adjacentes, et de l'énergie totale (cf. [Jensen, 1999b, p. 94]). L'utilisation d'un seuil permettant de stabiliser numériquement la valeur du descripteur lorsque le bruit prédomine peut s'avérer intéressante. Aussi, nous emploierons la formulation suivante pour l'Irrégularité Spectrale instantanée *IRRJEN*, dérivée de la définition de Jensen :

$$IRRJEN(t) = \frac{\sum_{h=1}^{H-1} (A_h(t) - A_{h+1}(t))^2}{b_0^2 + \sum_{h=1}^H A_h(t)^2} \quad (2.42)$$

où b_0 est le seuil introduit dans l'expression 2.22, et les autres quantités sont définies comme dans l'équation 2.20.

2.2.5.6.4 Le rapport d'énergie Impair/Pair (OER) Plus spécifique encore que l'Irrégularité Spectrale, le rapport d'énergie Impair/Pair rend compte des différences d'énergie entre les composantes harmoniques impaires et paires. Il s'adaptera particulièrement à la caractérisation de sons provenant d'instruments comme la clarinette, dont on sait que le résonateur cylindrique "fermé/ouvert" ne laisse passer qu'en grande partie les harmoniques impairs (cf. chapitre 3), les harmoniques pairs n'étant présents que dans une moindre mesure [Benade et Kouzoupis, 1988]. Le rapport d'énergie Impair/Pair (OER), ici exprimé en fonction du temps, est un nombre sans dimension dérivé du rapport entre l'énergie totale des composantes impaires et l'énergie totale des composantes paires (cf. par ex. [Peeters, 2004]) :

$$OER(t) = \frac{b_0 + \sum_{h=0}^{\lfloor \frac{H}{2} \rfloor - 1} A_{2h+1}(t)}{b_0 + \sum_{h=1}^{\lfloor \frac{H}{2} \rfloor} A_{2h}(t)} \quad (2.43)$$

où $\lfloor \cdot \rfloor$ désigne la partie entière par défaut, b_0 est le seuil introduit dans l'expression 2.22, et les autres quantités sont définies comme dans l'équation 2.20. Le nombre total d'harmoniques H sera supposé pair afin qu'un nombre identique d'harmoniques impairs et d'harmoniques pairs soit considéré. En pratique, si tel n'est pas le cas, l'utilisation de la partie entière de $\frac{H}{2}$ permet de rendre cohérent le calcul du descripteur. Si $OER < 1$, les harmoniques pairs sont dominants, tandis que si $OER > 1$, les harmoniques impairs sont dominants.

2.2.5.6.5 Le Tristimulus (TR1, TR2, TR3) Le Tristimulus, défini par Pollard et Jansson dans la référence [Pollard et Jansson, 1982], offre une caractérisation de la distribution d'énergie spectrale par trois coordonnées. Il a été introduit comme l'équivalent pour le timbre des attributs de la couleur pour la vision (la couleur peut être décrite entre autres par trois paramètres R, V, B, spécifiant respectivement les quantités de Rouge, Vert, et Bleu) [Jensen, 1999b]. Pollard et Jansson ont employé le Tristimulus afin d'étudier le comportement transitoire des sons musicaux. Ystad l'a utilisé en tant que critère psychoacoustique permettant d'ajuster les paramètres de contrôle d'un modèle de synthèse reposant sur la distorsion non-linéaire, destiné à reproduire la partie déterministe de sons de flûte traversière (cf. [Ystad, 1998, p. 79]). Il s'est avéré, dans cette application, plus pertinent que le Centre de Gravité Spectral, car il prend en compte l'évolution des premières composantes harmoniques. En effet, ceci joue un rôle important sur le plan perceptif dans le cas de la flûte, car ces composantes varient fortement en fonction de la pression. Les trois paramètres

adimensionnés définissant le Tristimulus sont :

- TR_1 , qui rend compte de l'énergie relative du composant fondamental par rapport à l'énergie totale,
- TR_2 , qui rend compte de l'énergie relative des harmoniques 2, 3, et 4 par rapport à l'énergie totale, et
- TR_3 , qui rend compte de l'énergie relative des composantes harmoniques supérieures (5 à H) par rapport à l'énergie totale.

Les valeurs instantanées de TR_1 , TR_2 , et TR_3 sont données par les expressions suivantes :

$$\begin{cases} TR_1(t) = \frac{A_1(t)}{b_0 + \sum_{h=1}^H A_h(t)} \\ TR_2(t) = \frac{A_2(t) + A_3(t) + A_4(t)}{b_0 + \sum_{h=1}^H A_h(t)} \\ TR_3(t) = \frac{\sum_{h=5}^H A_h(t)}{b_0 + \sum_{h=1}^H A_h(t)} \end{cases} \quad (2.44)$$

où b_0 est le seuil introduit dans l'expression 2.22, et les autres quantités sont définies comme dans l'équation 2.20. Notons que dans le cas où le seuil b_0 est nul, la somme des trois paramètres de Tristimulus est égale à 1.

2.2.5.6.6 L'Inharmonicité (INH) Les descripteurs présentés ci-dessus visaient à décrire certaines propriétés du spectre d'amplitude. La localisation fréquentielle des composantes spectrales est également un critère déterminant de l'identité instrumentale, comme nous l'avons souligné à la section 1.1.3.3 (comparaison de sons harmoniques et inharmoniques). L'Inharmonicité représente la divergence des fréquences des composantes spectrales d'un signal par rapport à celles d'un signal purement harmonique. Nous utiliserons la formulation donnée dans [Peeters, 2004], qui repose sur la valeur de la fréquence fondamentale f_0 , et pondère les divergences fréquentielles par l'énergie des composantes spectrales. Son expression temporelle est la suivante :

$$INH(t) = \frac{1}{f_0(t)} \frac{\sum_{h=1}^H |f_h(t) - hf_0(t)| A_h(t)^2}{b_0^2 + \sum_{h=1}^H A_h(t)^2} \quad (2.45)$$

où $f_0(t)$ représente la valeur de la fréquence fondamentale instantanée, b_0 est le seuil introduit dans l'expression 2.22, et les autres quantités sont définies comme dans l'équation 2.20. Le terme $|f_h(t) - hf_0(t)|$ étant au plus égal à $f_0(t)$, l'inharmonicité varie entre 0 (signal parfaitement harmonique) et 1 (signal inharmonique), lorsque b_0 est nul.

2.2.5.7 Les MFCC (Mel-Frequency Cepstral Coefficients)

Les MFCC (Mel-Frequency Cepstral Coefficients) [Davis et Mermelstein, 1980] sont une représentation du cepstre [Bogert *et al.*, 1963] reposant sur une échelle de fréquences qui rend compte de la perception humaine, l'échelle Mel [Stevens *et al.*, 1937]⁶³. Les MFCC associés à un signal sont obtenus en appliquant la Transformée en Cosinus Discrète au logarithme du module de la Transformée de Fourier du signal, filtré par un banc de filtres triangulaires, dont les fréquences centrales sont espacées de manières égales selon l'échelle Mel.

⁶³L'échelle Mel a été conçue de sorte qu'une variation constante en mels soit perçue par les auditeurs comme une variation constante de fréquence (en Hz), et en posant le fait que 1000 mels correspondent à 1000 Hz.

Les MFCC s'avèrent être efficaces pour le traitement et la reconnaissance de la parole, puisqu'ils permettent de décrire l'enveloppe spectrale avec un nombre restreint de paramètres (13 coefficients sont généralement employés pour la représentation de la parole). Dans [Tzanetakis, 2002], l'auteur montre qu'une représentation de la musique avec les 5 premiers coefficients MFCC offre de bonnes performances dans une tâche de reconnaissance automatique du genre musical. Dans [Terasawa *et al.*, 2006], les auteurs proposent un modèle de timbre reposant sur 13 coefficients MFCC. Ce modèle a l'avantage de représenter le timbre dans un espace dont les dimensions sont orthogonales. Cependant, il concerne uniquement des sons stationnaires et ne permet pas de prendre en compte les aspects temporels ou spectro-temporels liés au timbre. Les auteurs n'offrent pas de comparaison de leur modèle avec d'autres modèles de timbre, notamment ceux issus d'études de dissemblance et d'analyses statistiques multivariées (cf. section 2.2.4), et reposant sur les descripteurs classiques (Temps d'Attaque, Centre de Gravité Spectral, etc.).

De par leur définition, les MFCC semblent bien adaptés aux instruments dont le fonctionnement acoustique repose sur un modèle source/filtre (comme la voix), ou plus généralement aux instruments dont le spectre présente des formants. En effet, pour la voix (parole/chant), la nature du timbre est en grande partie liée à la localisation fréquentielle de zones formantiques qui varient selon la forme du conduit vocal (cf. par exemple les études de Sundberg [Sundberg, 1995]). Ceci explique probablement pourquoi les MFCC, qui caractérisent les formants du spectre, sont utilisés pour l'identification automatique de différents chanteurs. En revanche, le fonctionnement acoustique de la clarinette n'est pas du type source/filtre. Le spectre externe de l'instrument est principalement déterminé par les résonances du conduit cylindrique, qui suivent une série d'harmoniques impairs, et le couplage non-linéaire avec l'excitateur, qui explique les différences de richesse spectrale et l'apparition d'harmoniques pairs (cf. chapitre 3). Pour les différentes raisons mentionnées ci-dessus, nous n'avons pas utilisé les MFCC pour caractériser les variations de timbre-qualité de la clarinette. De plus, dans la perspective de reproduire des variations de timbre spécifiques liées à l'interprétation (afin d'établir un modèle d'interprétation musicale ou d'étudier l'influence des variations au niveau perceptif), il semble très complexe de contrôler les valeurs des MFCC par problème inverse.

2.2.5.8 Echelle d'amplitude (linéaire, puissance)

Comme nous l'avons mentionné précédemment, certains descripteurs de timbre peuvent être calculés suivant différents types d'échelles d'amplitude. Les formulations des descripteurs ont été présentées pour une échelle de type linéaire. Toutefois, une échelle de puissance peut être utilisée en remplaçant dans les formules concernées les amplitudes des composantes spectrales $A_n(k)$, ou les amplitudes instantanées des composantes harmoniques $A_h(t)$ par leurs carrés respectifs $A_n(k)^2$, et $A_h(t)^2$. Lorsque le seuil stabilisateur b_0 est utilisé, sa valeur devra alors aussi être remplacée par b_0^2 .

Notons qu'une échelle linéaire consiste à considérer avec équité chaque composante spectrale, tandis qu'une échelle de puissance tend à accorder plus de poids aux composantes spectrales dominantes. Par conséquent, une échelle de puissance sera moins sensible au bruit (dans la mesure où ce dernier est bien large bande et de faible amplitude). Lors de l'interprétation de la structure des espaces de timbre perceptifs à l'aide de descripteurs, il peut être intéressant d'utiliser les différents types d'échelles d'amplitude sans *a priori* (comme dans la référence [Grey et Gordon, 1978]), afin de déterminer l'échelle qui semble correspondre au mieux avec notre perception. Nous avons suivi ce principe dans l'étude de dissemblance de timbres de clarinette, exposée au chapitre 4.

2.2.5.9 Limites des descripteurs de timbre

Même si les études psychoacoustiques sur les dissemblances de timbre sont en accord sur certains points, il existe également des divergences qui tendent à montrer que les résultats obtenus dépendent des stimuli utilisés dans les expériences. Par exemple, dans les études où les stimuli sont courts (moins d'une seconde, comme dans l'étude de Grey [Grey, 1977]), l'attaque devient perceptivement saillante, peut-être car elle procure de l'information utile en l'absence d'un long régime permanent. Cependant, dans les études utilisant des sons entretenus plus longs (par ex. [Wessel, 1973]), les caractéristiques du régime permanent semblent prédominer.

Nous tenons à insister sur le fait que la méthodologie reposant sur l'analyse multidimensionnelle de proximité et l'interprétation psychophysique des dimensions des configurations géométriques obtenues permet tout au plus de déterminer des corrélats physiques de la perception. En effet, rien ne permet à ce jour de prouver l'hypothèse que les descripteurs de timbre expliquent réellement la nature de la structure psychologique sous-jacente à la discrimination de sons de timbres différents. Le Centre de Gravité Spectral est le seul descripteur que l'on retrouve communément dans la plupart des études. Il semble à ce titre rendre compte de manière robuste d'un attribut marquant pour la perception, la brillance du son. Il représente alors un bon moyen potentiel d'identifier et de classer différents instruments de manière automatique. Mais il est difficile de savoir s'il explique réellement les processus mis en jeu dans la perception auditive pour parvenir à de telles fins. Le cerveau disposerait-il d'un calculateur de Centre de Gravité Spectral enfoui quelque part dans le cortex ? Il est difficile de répondre à cette question. Nous soutenons l'hypothèse que c'est la confluence d'un ensemble de facteurs qui permet au système auditif de percevoir les subtilités timbrales. Les différents descripteurs psychoacoustiques de timbre décrits ci-dessus ne tiennent pas compte d'éventuels traitements cognitifs tel que le recours à la mémoire. Les corrélations avec les dimensions perceptives des espaces se déterminent à partir de descripteurs globaux, qui résument parfois l'information spectro-temporelle en une seule donnée quantifiable. Mais le système auditif dispose-t-il d'un tel intégrateur temporel ? Nous avons vu que bien plus que la valeur globale de tel ou tel paramètre, ce sont les relations temporelles entre différents facteurs qui semblent jouer un rôle prééminent sur l'identité et de la qualité sonore instrumentale (cf. étude des sons de trompette de Risset [Risset, 1966]).

Les descripteurs de timbre n'en sont pas moins intéressants dans la mesure où ils peuvent tout de même être la source de modèles prédictifs de notre perception. S'ils peuvent *prédire* notre perception, cela ne veut pas dire qu'ils *expliquent* la façon dont nous fonctionnons.

Chapitre 3

Clarinette et modélisation physique

La clarinette est un instrument qui a plus de trois siècles. Elle a et continue de séduire de nombreux compositeurs. Sa versatilité, sa grande richesse sonore, ne sont pas étrangers au fait qu'on la retrouve dans des styles musicaux très variés (par ex. classique, jazz, musique traditionnelle bretonne, klezmer, grecque). Elle représente également un sujet d'étude passionnant pour les physiciens puisque la compréhension de son fonctionnement complexe représente une belle application de disciplines telles que l'acoustique et la mécanique des fluides.

Au cours de ce chapitre, nous commencerons par rappeler les origines historiques de la clarinette, ainsi que son fonctionnement général. Nous présenterons les principaux facteurs physiques expliquant la nature de son timbre, ainsi que les paramètres de contrôle permettant à l'instrumentiste de le faire varier. Afin d'obtenir de plus amples informations, nous invitons le lecteur à consulter les articles fondamentaux de Wilson et Beavers [Wilson et Beavers, 1974], Benade et Kouzoupis [Benade et Kouzoupis, 1988], ainsi que ceux généraux de Kergomard [Kergomard, 1991] et Laloë (Franck et Susanne) [Laloë et Laloë, 1995]. Nous nous intéresserons enfin à la synthèse par modélisation physique (simplifiée), qui dans le cadre de ce travail représente un outil aidant à éclaircir certains aspects des complexes relations entre contrôle, timbre et perception.

3.1 Origines

La clarinette est un instrument à vent à anche simple de la famille des bois¹. Elle est issue des travaux du facteur allemand Johann Christoph Denner (1655-1707) qui la conçoit vers 1700 (cf. photo de la figure 3.1(a)). Certains la voient comme une évolution du *chalumeau*², "instrument rudimentaire des paysans" [Brymer, 1979, p. 26], dont le nom assigné à tort aux premières clarinettes, sans doute en raison des fortes ressemblances physiques et sonores entre les deux instruments, s'est vu ensuite utilisé pour désigner son registre grave.

Il est intéressant de constater que c'est pour son timbre que la clarinette s'est très tôt fait remarquer des compositeurs : "[...] dès son apparition au *XVII^e* siècle, elle a immédiatement complété les bois de l'orchestre [NDLR : la flûte, le hautbois et le basson sont ses aînés], ajoutant à la palette dont disposaient les compositeurs, une couleur que beaucoup d'entre eux recherchaient depuis longtemps : c'est l'une des ses plus importantes fonctions dans l'orchestre." [Brymer, 1979,

¹Contrairement à ce qu'il est souvent coutume de penser, ce n'est non pas le matériau du corps de l'instrument qui vaut à la clarinette son appartenance à la famille des bois, mais celle de son "moteur", l'anche, une fine lamelle de roseau. Si l'ébène est certes l'un des matériaux phares utilisés pour fabriquer le corps de l'instrument, on trouve également des clarinettes modernes en plastique ou matériaux composites, d'autres en verre, ou bien même métal. Remarquons que ceci vaut également pour le saxophone, dont le corps est généralement en laiton ou en cuivre.

²Le terme chalumeau vient du latin *calamus*, petite anche, ou du grec *calane*, pipeau de roseau [Brymer, 1979, p. 27].

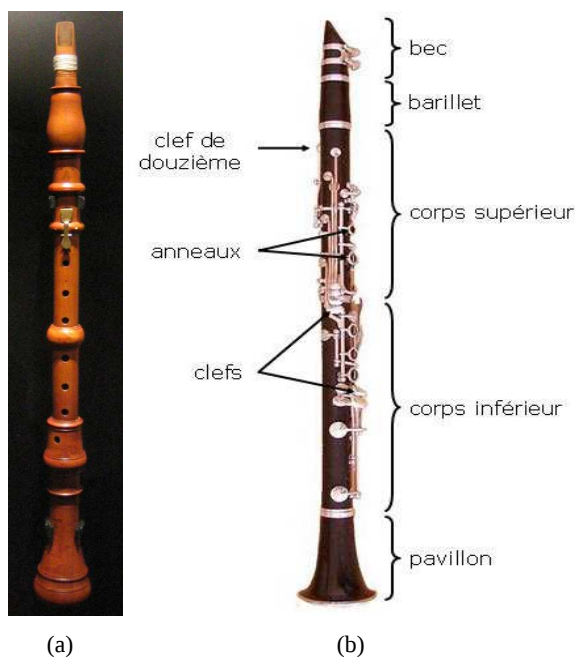


FIG. 3.1 – (a) Clarinette ancienne du facteur Johann Christoph Denner (Musée des instruments de musique de Bruxelles), (b) Description d’une clarinette moderne avec un système de clefs de type “Boehm”.

p. 17]. Mozart, l’auteur des célèbres *Quintette pour Clarinette et Cordes KV. 581*³, et *Concerto pour Clarinette KV. 622*, s’exprimait en effet ainsi dans une lettre qu’il écrivit à son père : “Oh, si seulement nous avions aussi des clarinettes - vous ne pouvez imaginer l’effet splendide d’une symphonie avec flûtes, hautbois et clarinettes.”.

La clarinette se distingue parmi les bois en raison de sa large étendue, presque 4 octaves complètes, dont les subdivisions classiques sont **les registres chalumeau, de gorge, clairon et aigu**, mais aussi car elle possède un éventail de nuances très large, “du *ppp* presque inaudible” au “*fff* qui sonne comme une trompette ; elle possède une variété plus subtile de qualité sonore, de la douceur veloutée à la dureté de l’acier” [Brymer, 1979, p. 18]. La clarinette moderne la plus utilisée est la clarinette soprano en $B\flat$ disposant du système de clefs “Boehm”. Ce système de clefs a été inventé par le bavarois Theobald Boehm pour les flûtes. Il a été adapté en 1839 à la clarinette par le français Klosé (cf. photo de la figure 3.1(b)). Les avantages de ce système concernent la jouabilité de l’instrument, puisqu’il permet à l’instrumentiste de couvrir des trous ordinairement inaccessibles au moyen d’une série de clefs à anneaux, et très probablement la justesse et le timbre de l’instrument (la taille des trous augmente).

3.2 Fonctionnement général de l’instrument

Comme le montre la figure 3.1(b), la clarinette est constituée : d’un bec sur lequel l’anche, fine lamelle flexible généralement en roseau, est fixée au moyen d’une ligature, d’un barillet qui relie le bec au corps de l’instrument de perce approximativement cylindrique (en deux parties), et enfin d’un pavillon. Du point de vue acoustique, la clarinette peut être vue comme l’association d’un **excitateur** (ou générateur acoustique), l’anche vibrante simple, et d’un **résonateur**, la “colonne” d’air contenue dans l’instrument. Le **couplage entre l’excitateur et le résonateur** est assuré par

³L’un des extraits musicaux utilisés dans l’expérience perceptive décrite au chapitre 6 est issu de cette oeuvre.

l'écoulement d'air au travers de l'ouverture entre l'anche et le bec (la figure C.1 de l'Annexe C offre une description du bec de clarinette). La clarinette appartient aux instruments *entretenus*⁴, pour lesquels le couplage entre l'excitateur et le résonateur dure autant que le son, hormis durant la courte phase d'extinction de ce dernier⁵.

Pour les instruments à vent, l'air est directement mis en vibration sans intermédiaire mécanique (contrairement, par exemple, au cas de la guitare, pour laquelle la mise en vibration de l'air provient des oscillations de la corde qui sont transmises à l'air contenu dans la caisse de résonance par l'intermédiaire du chevalet et de la table d'harmonie). La forme intérieure du tuyau, donc celle de la "colonne d'air", dont les modes propres de vibration sont excités, joue le rôle prépondérant au niveau de la hauteur ainsi que de la sonorité de l'instrument [Kergomard, 1991, p. 230].

Pour engendrer un son, l'instrumentiste impose une surpression statique dans sa bouche. Sous certaines conditions, celle-ci est transformée par l'instrument en une pression oscillant suffisamment rapidement pour correspondre à des sons audibles. Cette transformation est rendue possible grâce à l'anche⁶ dont les oscillations, augmentant ou diminuant l'ouverture du canal d'anche, viennent "hacher" le débit d'air du souffle de l'instrumentiste, ce que l'on appelle "l'effet valve". Lors d'un fonctionnement normal de l'instrument, la fréquence de vibration de l'anche dépend peu de ses caractéristiques propres (masse et raideur) et de l'amortissement que lui confère l'instrumentiste avec sa lèvre inférieure, mais surtout de la position du premier trou latéral ouvert (ou bien de l'extrémité du pavillon, pour la note la plus grave obtenue quand tous les trous sont fermés). Ceci provient du fait, que la principale fréquence propre de vibration de l'anche est très supérieure à la fréquence de jeu.

Lorsque l'onde de pression se propageant dans le tuyau du bec vers son extrémité rencontre un trou ouvert, une certaine quantité d'air s'échappe vers l'extérieur, ce qui crée une dépression dans le tuyau. Cette dépression se propage alors en direction inverse, c'est-à-dire du trou ouvert vers le bec. Il y a donc réflexion de l'onde de pression avec changement de signe. Lorsque la dépression atteint l'anche, cette dernière déjà contrainte à refermer le canal d'anche sous le joug de la pression de la bouche, s'applique encore plus fortement sur le bec, la pression interne étant plus faible⁷. L'onde de pression se réfléchit alors au niveau de l'extrémité du tuyau presque fermée par l'anche, mais cette fois sans changement de signe. La dépression se propage jusqu'au trou latéral ouvert, et se réfléchit pour la troisième fois, sous forme d'une surpression. Lorsque cette surpression parvient à l'anche, elle force cette dernière à s'ouvrir permettant à une nouvelle bouffée d'air de s'introduire dans le tuyau. Ceci vient entretenir une oscillation dite permanente ; le cycle est alors complet. Dans le cas d'un tuyau parfaitement cylindrique sans perturbations (par ex. pavillon, trous latéraux, etc.), La durée d'un cycle fondamental correspond donc au temps mis par l'onde sonore pour parcourir quatre fois la distance L qui sépare l'anche du premier trou latéral ouvert

⁴Dans le cas des instruments à vent, on parle également d'instruments *auto-entretenus* ou *auto-oscillants*, pour insister sur le fait que les oscillations proviennent de la nature du système, et non de l'extérieur.

⁵Par opposition, pour les instruments *non-entretenus* comme les cordes pincées, le couplage n'intervient qu'au début du son.

⁶L'anche peut être considérée à ce titre comme le "moteur", le générateur acoustique de l'instrument.

⁷Le mouvement des anches, ou plus généralement des valves contrôlées par pression (par ex. anches, dans le cas des bois, ou lèvres dans le cas des cuivres) dépend de la différence de pression qui existe de part et d'autre. Parmi les instruments à vent, les fonctionnements peuvent être différents. D'après les observations d'Helmholtz sur le fonctionnement des anches [Helmholtz, 1954] et la classification proposée par Bouasse [Bouasse, 1929], Fletcher et Rossing présentent un système de classification des valves qui repose sur un doublet de symboles (σ_1, σ_2) [Fletcher et Rossing, 1998, p. 402]. σ_1 est égal à +1 si un excès de pression appliqué à l'entrée de la valve tend à l'ouvrir, et au contraire à -1 s'il tend à la fermer. Réciproquement, σ_2 est égal à +1 si un excès de pression en sortie de la valve tend à l'ouvrir, et à -1, s'il tend à la fermer. Ainsi, suivant cette classification, l'anche vibrante de clarinette correspond au couple (-,+).

(ou de l'extrémité de l'instrument), i.e. $\frac{4L}{c}$, où c est la célérité du son⁸.

3.3 Principaux facteurs constitutifs du timbre de la clarinette

3.3.1 Le système bec-anche

Le bec - Le bec joue un rôle important sur le timbre de l'instrument. Pour un instrumentiste, le choix d'un bec est quelque chose de très personnel, qui se fait en relation avec sa morphologie, son écoute, la recherche d'un son précis correspondant à une spécialisation musicale, mais aussi les possibilités musicales qu'il permet (par ex. legato, détaché, facilité d'émission des notes, etc.).

La majorité des becs utilisés de nos jours sont en ébonite (caoutchouc vulcanisé dont la couleur proche de l'ébène est à l'origine de son nom). Au cours des siècles, différents essais ont été réalisés avec des matériaux comme le bois, le verre, la porcelaine, le cristal. Cependant, comme pour le tuyau, la perce d'un bec semble avoir une importance bien plus grande sur son comportement que le matériau lui-même. La facture d'un bec a une telle importance sur le son que des becs provenant d'une même série, ayant donc théoriquement les mêmes côtes, peuvent présenter des variations dans leur réponse. Outre les dimensions et la forme intérieure du bec, la forme incurvée de sa table (cf. la figure C.1 de l'Annexe C) joue également un rôle essentiel. Elle permet à l'anche de s'y "enrouler" progressivement lors de ses oscillations [Ollivier, 2002], ce qui limite l'apparition d'harmoniques élevés. En effet, plus le mouvement de l'anche est brutal entre les positions ouverte et fermée (par ex. situation d'anche battante où l'anche vient taper le bord du bec, en effectuant d'éventuels rebonds), plus les harmoniques élevés seront favorisés [Laloë et Laloë, 1995, p. 80]. En revanche, pour des petites oscillations (à faibles pressions), le mouvement de l'anche pourra être quasi-sinusoïdal et engendrera un son plutôt pur (peu de composantes harmoniques).

Pour les enfants qui débutent, et en raison de leur alimentation en air peu soutenue, les becs préconisés sont ceux qui présentent une faible cambrure ("becs pas trop ouverts"⁹). Ces becs dits "faciles" permettent d'éviter que les jeunes clarinettes aient à forcer et se fatiguent.

L'anche - Les anches de roseau¹⁰ sont très disparates. Au sein d'une boîte d'anches de même type, les différences peuvent être importantes. Ceci représente un problème pour les musiciens. De plus, les réactions des anches varient selon l'hygrométrie¹¹ et l'altitude. Selon le clarinetiste Claude Crousier, la recherche d'une anche convenable résulte d'une adéquation entre le son désiré par l'instrumentiste, sa morphologie, le bec utilisé, le style musical, et le lieu où il joue.

La fréquence propre de vibration de l'anche est une limite pour la fréquence de jeu, fait déjà observé par Helmholtz [Helmholtz, 1954, p. 390]. Wilson et Beavers ont montré de manière théorique et expérimentale que l'anche devait être fortement amortie afin que la résonance la plus basse du tuyau soit favorisée [Wilson et Beavers, 1974]. Ceci est réalisé en situation de jeu par la force de pression qu'exerce l'instrumentiste sur l'anche, au moyen de ses dents et de sa lèvre inférieure (cf. section 3.3.8). Thomson a montré que selon l'amortissement dû à la lèvre de l'instrumentiste, la fréquence de résonance de l'anche pouvait varier de 2000 à 3000 Hz [Thompson, 1979]. Si l'anche n'est pas suffisamment amortie, alors c'est sa fréquence propre de vibration qui domine

⁸La fréquence de vibration du mode fondamental correspondant est donc $f_1 = \frac{c}{4L}$. Nous verrons qu'elle diffère légèrement de la fréquence de jeu, déterminant la hauteur du son perçu.

⁹Un bec ouvert est un bec pour lequel la cambrure de la table est grande, ce qui conduit à avoir une large ouverture en bout d'anche en l'absence de tension d'embouchure de l'instrumentiste.

¹⁰Le roseau est une plante dont le sud de la France est fécond ; le fabricant *Vandoren* dispose de célèbres roselières dans le Var.

¹¹L'hygrométrie caractérise l'humidité relative de l'air.

son mouvement oscillant, et cela induit une vibration sonore de fréquence très élevée et généralement désagréable (le “canard d’anche”).

Les caractéristiques physiques de l’anche sont aussi importantes pour le timbre de l’instrument. Il existe des anches de différentes “forces”. Les anches “faibles”, minces et très élastiques, facilitent l’émission du son¹². Elles oscillent avec une grande amplitude, ce qui entraîne un son riche en harmoniques élevés, plutôt “acide” [Laloë et Laloë, 1995, p. 84]. A l’opposé, les anches “fortes” sont plus rigides et nécessitent plus d’efforts à l’instrumentiste¹³. Elles offrent en contrepartie une sonorité plus “pleine” et “homogène” (ibidem). Les anches plastifiées (par ex. du fabricant *Rico*) ont l’avantage de présenter moins d’imprégnation à la salive et de durer plus longtemps mais n’ont pas conquis la majorité des instrumentistes. Si elles sont utilisées par les musiciens de jazz, elles le sont bien moins par les musiciens de classique. Signalons au passage le système original de languettes en plastique aluminisé *Claripatch*, inventé par le clarinettiste suisse Pierre-André Taillard. Ce système vient s’intercaler entre l’anche et le bec, en vue d’améliorer la qualité sonore [Taillard, 2004], et permettre de continuer à utiliser de vieilles anches.

3.3.2 Le tuyau fermé/ouvert : un résonateur quart d’onde

Comme nous l’avons vu, la vibration de l’anche dépend de la pression renvoyée par le tuyau. Il convient donc de caractériser ce dernier. Le tuyau de la clarinette est ouvert au niveau du pavillon, mais il est presque fermé au niveau de l’embouchure. En effet, la fente entre l’anche et le bec, qui varie lors du mouvement de l’anche, reste cependant toujours étroite devant la section transversale du tuyau. Ceci entraîne une réflexion de l’onde de pression au niveau de l’embouchure quasiment identique à celle qui se produirait si le tuyau était complètement fermé.

Afin d’étudier de manière simplifiée le fonctionnement acoustique du tuyau, prétendons que ce dernier est parfaitement cylindrique et sans pertes, que tous les trous situés avant le premier trou ouvert sont fermés, et que l’extrémité est complètement fermée au niveau de l’embouchure. Au niveau du pavillon, le fait que la clarinette soit ouverte conduit la pression totale à être approximativement égale à la pression atmosphérique. Ceci signifie que la pression acoustique¹⁴ est quasiment nulle à la sortie du tuyau. Il y a donc un *noeud de pression* à la sortie du tuyau. La vitesse des molécules d’air, libres de se déplacer de part et d’autre du pavillon est en revanche maximale. Le noeud de pression à la sortie du tuyau se double donc d’un *ventre de vitesse*. Ceci traduit une *condition aux limites de Dirichlet*. A l’entrée de la clarinette (considérée rappelons-le comme fermée), le phénomène inverse se produit : c’est la vitesse acoustique qui s’annule (*noeud de vitesse*), ce qui correspond à une *condition aux limites de Neumann*. La vitesse oscillant en quadrature avec la pression¹⁵, cette dernière est maximale à l’entrée de l’instrument. Il y a donc un *ventre de pression* à l’entrée de l’instrument.

Remarquons que la plus petite distance entre un maximum et un zéro pour une onde sinusoïdale correspond à un quart de la longueur d’onde. On dit que le résonateur de type fermé/ouvert, tel

¹²Ce sont celles que l’on conseille en général aux clarinettistes débutants.

¹³Pour les anches, les expressions *faibles* et *fortes* sont parfois utilisées pour désigner un tout autre phénomène ; Une anche est dite *forte* lorsque sa fréquence propre fixe la fréquence d’auto-oscillation des vibrations sonores (par ex. cas des lèvres pour la trompette, ou des cordes vocales pour la voix). A l’inverse, elle est dite *faible* lorsque c’est l’une des résonances du résonateur, auquel elle est couplée, qui fixe la fréquence d’auto-oscillation (par ex. cas de la clarinette pour un mode de jeu traditionnel sans “canards”, ces derniers étant justement émis lorsque la fréquence propre de l’anche domine (cf. section 3.3.8)).

¹⁴La pression acoustique correspond à une variation de pression. Si l’on note $\mathbf{P}(\mathbf{r}, t)$ la pression de l’air au point de coordonnées $\mathbf{r}$ (les caractères gras sont utilisés ici pour indiquer des vecteurs), à l’instant t , et P_0 , la pression statique atmosphérique, la pression acoustique au bout du tuyau $\mathbf{p}(\mathbf{r}, t)$ est définie par : $\mathbf{p}(\mathbf{r}, t) = \mathbf{P}(\mathbf{r}, t) - P_0$.

¹⁵Les ondes de pression se déplacent des zones de fortes pressions (vitesses faibles) vers celles de basses pressions (vitesses élevées).

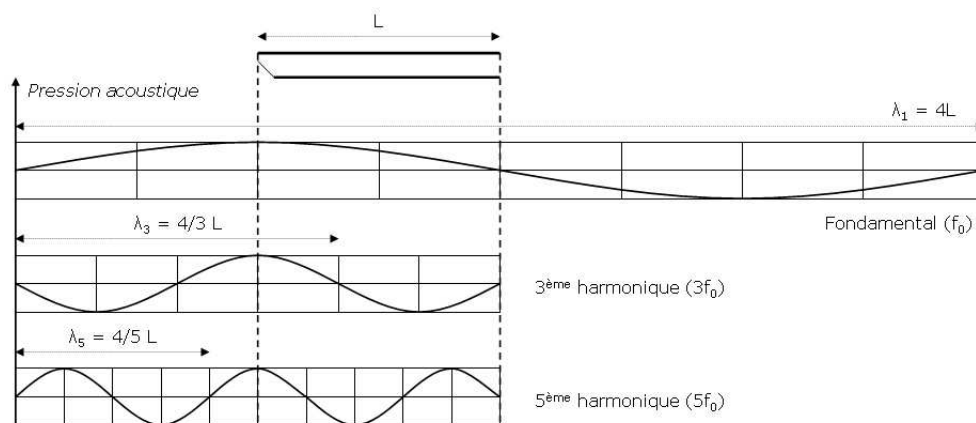


FIG. 3.2 – Tuyau sonore de type clarinette : une extrémité “fermée”, l’autre ouverte. Représentation schématique de la pression acoustique pour les modes propres du résonateur n^{os} 1, 3, et 5.

que celui de la clarinette, est un **résonateur quart d’onde**. Ceci indique que l’onde permanente de plus basse fréquence (harmonique fondamental), satisfaisant les conditions aux limites du tuyau, dispose d’une longueur d’onde égale à quatre fois la longueur du tuyau, comme représenté en haut de la figure 3.2 (et donc de fréquence $f_1 = \frac{c}{4L}$). Ceci corrobore la description du cycle de vibration du mode fondamental donnée à la section 3.2. Les autres vibrations vérifiant les conditions aux limites sont les harmoniques impairs successifs au fondamental (figure 3.2). **Les fréquences des modes propres d’un tuyau de type fermé/ouvert suivent une série d’harmoniques impairs (1, 3, 5, 7, etc.).** Ceci est un critère déterminant pour le timbre de l’instrument. Pour les multiples pairs de la fréquence fondamentale ($2f_1, 4f_1$, etc.), c’est une sous-pression qui arrive au niveau de l’anche à l’instant où une surpression provenant de la cavité buccale devrait commencer sa propagation. Il y a alors opposition entre la réaction en pression du tuyau et l’excitation, ce qui explique pourquoi ces modes de vibrations ne peuvent résonner dans le tuyau. Le fait que les harmoniques pairs de pression ne peuvent résonner dans le tuyau ne veut pas dire qu’ils ne sont pas à même de propager dans le tuyau. En effet, si le spectre de pression interne de la clarinette contient très peu d’harmoniques pairs, ils peuvent être présents de façon non négligeable dans le spectre externe. Nous verrons que leur existence est liée au couplage non-linéaire entre l’anche et le résonateur.

La réaction du tuyau peut être caractérisée par une quantité appelée **impédance acoustique**, égale, en un point et une fréquence donnés, au rapport entre la pression que l’on doit imposer en ce point pour faire passer dans le tuyau une onde sonore vibrant à cette fréquence, et le débit acoustique correspondant¹⁶. Une grande valeur de l’impédance indique que la réaction en pression du résonateur sur l’anche sera forte à la fréquence de vibration considérée. De fait, les résonances du tuyau, qui correspondent aux valeurs élevées de l’impédance (ou pics), sont si fortes qu’elles contrôlent la vibration de l’anche. Dans des modes de jeu normaux, l’instrument ne peut donc émettre des notes qu’à des fréquences proches des fréquences de résonance du tuyau, qui comme nous l’avons vu reposent sur une série d’harmoniques impairs. Nous présenterons dans la prochaine section des impédances d’entrée du résonateur de clarinette simulées (cf. section 3.4.2.1). Nous pourrions constater qu’aux fréquences harmoniques paires, l’impédance acoustique est presque nulle. L’impédance acoustique représente une véritable signature de la forme du tuyau.

¹⁶Le débit acoustique au travers d’une surface est le produit de l’aire de cette surface par la vitesse particulaire.

Le matériau ne semble avoir que peu d'effet sur les vibrations de l'air interne à l'instrument, qui à niveau très élevé sont souvent supérieures à 150 dB (que les clarinettes soient en ébène, ou en métal, elles ne changent pas pour autant de typologie). Ceci étant, il s'agit là d'un sujet controversé, et il ne paraît pas illogique de penser que d'éventuelles vibrations des parois ainsi que la rugosité de leur face interne peuvent influencer sur les oscillations. Ainsi, il semble que les parois métalliques favorisent certains harmoniques aigus, tandis que des parois en bois rugueuses augmentent l'absorption, et par conséquent nuisent à l'oscillation (cf. [Kergomard, 1991, p. 231]).

3.3.3 Le couplage non-linéaire entre l'excitateur et le résonateur

La naissance de l'auto-oscillation dans l'instrument est due à un couplage non linéaire entre l'excitateur et le résonateur. Comme le souligne Laloë, "Bien des caractéristiques de l'ensemble bec-anche (par exemple les variations de la longueur vibrante de l'anche dues à la courbure de la "table" du bec) engendrent en effet de telles non-linéarités. De plus, même si le bout de l'anche bougeait de façon parfaitement sinusoïdale, ce ne serait pas nécessairement le cas pour le débit d'air que l'anche admet dans le résonateur, car le débit d'air n'est pas proportionnel à l'ouverture de l'anche. Il existe donc de multiples raisons pour que l'anche, actionnée par une pression à la fréquence f , puisse fournir un débit à des fréquences multiples plus élevées $2f$, $3f$, etc. Rien ne lui interdit de produire des vibrations à des fréquences où elle n'est nullement excitée par la pression dans le résonateur. C'est ainsi que s'explique son aptitude à engendrer des fréquences pour lesquelles l'impédance du résonateur sonore est très faible, par exemple le deuxième harmonique, qui n'a rien d'aussi mystérieux qu'on le voit parfois écrit !" [Laloë et Laloë, 1995, p. 84].

Worman [Worman, 1971], Benade [Benade, 1976], et Thompson [Thompson, 1979], ont étudié les effets des non-linéarités sur le spectre interne de la clarinette. Il paraît intéressant de rappeler à ce stade les **lois de Worman-Benade** :

1. Pour une anche contrôlée en pression, à des niveaux modérés (de faibles à moyens), et en négligeant les forces de Bernoulli présentes au bout de l'anche, l'amplitude de la pression interne P_n de la n^{eme} composante harmonique d'un son est reliée à l'amplitude P_1 du composant fondamental selon la formule suivante :

$$P_n = kP_1^n \quad (3.1)$$

où k est une constante [Benade, 1976, p. 441]. La relation exprime le fait que lorsqu'un clarinettiste joue crescendo, l'amplitude du n^{eme} composant harmonique croît proportionnellement à la puissance n de celle du fondamental (par ex. si le fondamental est augmenté de 10 dB, l'harmonique 3 sera augmenté de 30 dB¹⁷). Cet effet a une répercussion importante pour le timbre de l'instrument puisqu'à mesure qu'un musicien augmente sa pression d'alimentation, d'une part l'intensité sonore augmente, et d'autre part, l'amplitude des harmoniques d'ordres élevés augmentent, et ce plus rapidement que celle du fondamental (cf. figure 3.3)¹⁸. Ceci se traduit perceptivement par une augmentation de la brillance du son à mesure que l'intensité augmente, comme le révélera un peu plus loin l'analyse d'un crescendo de clarinette. A l'opposé, lorsque l'on joue pianissimo, le son ne présente pratiquement aucun harmonique au-dessus du fondamental.

La loi précédente a toutefois ses limites, ce qui amène Benade à cette deuxième règle :

¹⁷En effet, $20\log_{10}(x)^n = n \times 20\log_{10}(x)$

¹⁸Notons ici que nous avons extrapolé les résultats portant sur le spectre interne de la clarinette sur son spectre externe. En effet, sauf cas exceptionnel où un élément additionnel tel qu'une sourdine, connue pour rejeter certaines fréquences, est connectée au bout du tuyau, le renforcement d'un harmonique du son interne se retrouve dans le son externe [Kergomard, 1991, p. 233].

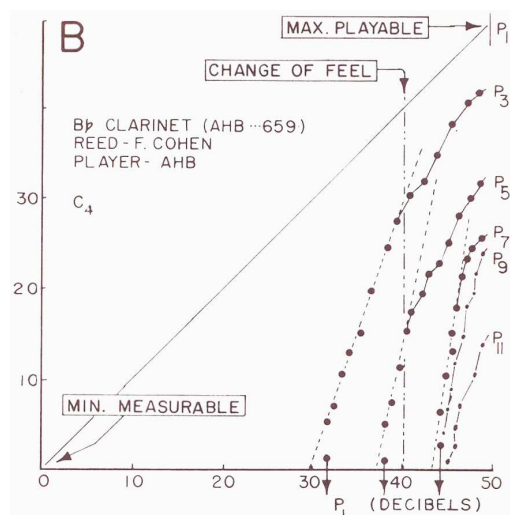


FIG. 3.3 – Mesures du spectre interne de clarinette : relations entre les amplitudes des partiels impairs d'ordres élevés et le fondamental (exprimées en dB SPL) lors d'un crescendo de clarinette (note C4). Les lignes en pointillés représentent les amplitudes théoriques prédites par la loi de Worman-Benade. Figure de Benade, issue de [Benade, 1976, p. 443].

2. La relation exprimée par l'équation 3.1 n'est valable que lorsque l'amplitude du mouvement de l'anche est suffisamment faible de tel sorte que le débit d'air qu'elle laisse parvenir dans le tuyau ne s'annule jamais totalement. Lorsque la pression d'alimentation atteint le seuil où l'anche ferme complètement le canal d'anche durant une portion de chaque cycle de ses oscillations (**situation d'anche battante**), "le musicien perçoit un changement, l'auditeur perçoit un changement de qualité sonore, et les partiels d'ordre élevés tendent à augmenter de manière parallèle à l'augmentation du fondamental." (traduction) [Benade, 1976, p. 442].

La figure 3.3 présente les relations mesurées (trait plein) et théoriques (traits pointillés) entre les amplitudes des partiels d'ordres élevés et leur fondamental lors d'un crescendo de clarinette. Ainsi, l'on se rend bien compte que plus le niveau augmente, plus le nombre de partiels d'ordres élevés augmentent : lors du crescendo, le son est au départ quasiment sinusoïdal (seul le fondamental est présent), puis à partir de 30 dB et de manière progressive, les autres partiels font leur apparition. D'autre part, l'augmentation relative d'amplitude par rapport au fondamental croît à mesure que l'ordre du partiel augmente (les pentes deviennent plus raides), ce qui confirme - jusqu'à un certain point - les prédictions de la loi de Worman-Benade (représentées par les lignes en pointillés). Cependant, au delà d'un certain niveau de jeu (indiqué par "change of feel"), la théorie n'est plus adéquate, et l'évolution de l'amplitude des partiels semble en effet se faire de manière parallèle à celle du fondamental.

Il est intéressant de remarquer que cette relation, entre niveau de jeu et amplitude des partiels, présente des analogies avec celle proposé par Risset dans le cas de la trompette, d'après l'analyse de sons réels (cf. figure 2.5). Nous avons vu que cette relation était l'un des éléments fondamental à prendre en compte en vue d'une synthèse réaliste de ce type de sons. Ce n'est pas pour rien que clarinette et trompette appartiennent toutes deux à une même famille, celle des vents. Plus précisément, elles ont en commun la nature de leur générateur acoustique, une valve contrôlée par pression, anche simple de roseau dans le cas de la clarinette (bois), anche "lippale" (les lèvres de l'instrumentiste) dans le cas de la trompette. **Le couplage non-linéaire entre l'excitateur et le résonateur apparaît donc à deux niveaux, d'une part, comme un élément constitutif d'une**

d'identité timbrale (cf. permanence de Schaeffer)¹⁹, et d'autre part, comme la clef de modifications notables de la qualité sonore (cf. variation de Schaeffer), offrant en l'occurrence au musicien des degrés de liberté pour façonner le timbre des sons pendant leur production. Si la figure 3.3 a permis de faire ces constats visuellement, il manque à l'oreille d'en être convaincue. Ainsi, analysons le crescendo de clarinette de l'Exemple Sonore n° 33, joué par un clarinettiste professionnel et enregistré en studio.

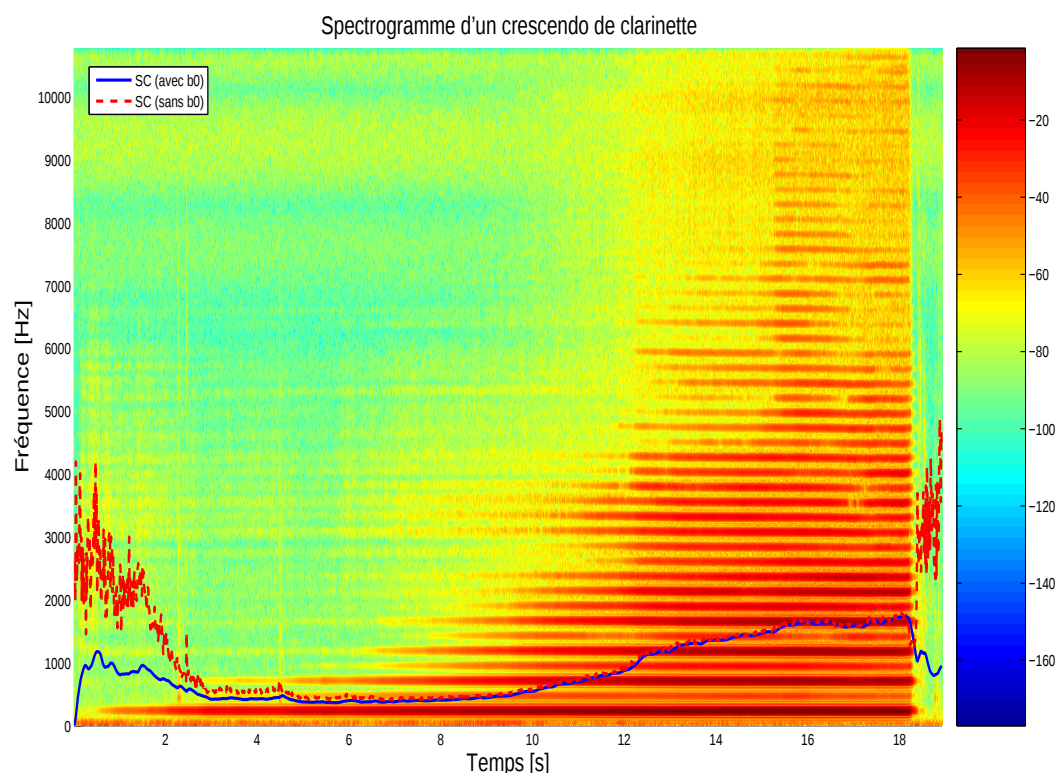


FIG. 3.4 – Spectrogramme d'un crescendo de clarinette, joué par un clarinettiste professionnel ($f_{jeu} \approx 237\text{ Hz}$, ce qui correspond à un $A\sharp 3$ un peu haut). L'évolution du Centre de Gravité Spectral (SC) au cours du temps est superposée à la représentation temps-fréquence. Deux formulations du descripteur SC sont comparées (cf. section 2.2.5.4.1), celle avec le seuil stabilisateur b_0 (trait plein), et celle sans seuil (traits pointillés). L'analyse a été effectuée avec des fenêtres de Hann d'environ 20 ms. Les amplitudes, normalisées par rapport à l'amplitude maximale, sont exprimées en décibels (dB).

La figure 3.4 permet de se rendre compte de l'évolution dynamique du spectre externe de la clarinette au cours d'un crescendo. Le son reste quasiment sinusoïdal lors des 3 premières secondes, puis les composantes harmoniques apparaissent successivement. Les harmoniques impairs apparaissent généralement avant les harmoniques pairs, preuve de la prédisposition de l'instrument à produire des auto-oscillations à ces fréquences. Remarquons la dominance de leurs amplitudes sur celles des harmoniques pairs, tout du moins jusqu'à environ 2 kHz, fréquence à partir de laquelle les énergies relatives des harmoniques pairs et impairs sont semblables. Ceci est peut être lié avec

¹⁹Le tuyau de l'instrument est un autre facteur d'identité timbrale important. Le spectre de la clarinette dispose en effet essentiellement d'harmoniques impairs en raison de la configuration de son résonateur, un tuyau cylindrique de type fermé/ouvert, tandis que celui de la trompette présente une série d'harmoniques complète, en raison de la conicité de son résonateur.

la fréquence propre de résonance de l'anche (qui peut se situer vers 2 kHz), et/ou au réseau de trous latéraux ouverts. Notons que les creux vers 6500, 8200 et 10000 Hz à la fin du son peuvent être dus à un phénomène d'interférences destructives de l'onde sonore dues à des réflexions dans le studio d'enregistrement (toutefois peu réverbérant). Il est également possible que ces creux soient liés au réseau de trous fermés, qui joue le rôle d'un filtre en peigne réjecteur.

L'analyse de ce crescendo permet également de constater la sensibilité du Centre de Gravité Spectral au bruit si l'on n'utilise pas le terme stabilisateur présent au dénominateur de l'expression 2.22 (cf. courbe en pointillés). Pour cette formulation du descripteur, lorsque le bruit est dominant, les valeurs de SC sont très élevées (environ 2500 Hz) et non significatives puisque le son est à peine perceptible (ceci est le cas au début du son jusqu'à environ 2.5 s, et à la fin, à partir de 18 s). La prise en compte du seuil b_0 permet de pallier en partie ce problème (courbe en trait plein), puisqu'il fait tendre la valeur du descripteur vers une valeur plus faible, lorsqu'il y a peu d'énergie dans le signal. Notons cependant que le terme b_0 a également pour effet de légèrement abaisser la valeur du descripteur SC lorsque le signal contient de l'énergie (la courbe de SC avec b_0 reste toujours en dessous de celle de SC sans b_0). L'écart entre les courbes restent cependant faible relativement à la valeur du descripteur (environ 50 Hz). Du reste, le rôle de ce descripteur n'est pas de prédire de manière absolue le percept de brillance, mais de traduire les variations relatives de brillance au cours du temps. Concernant le possible rôle du timbre dans l'expression musicale, nous soutenons l'hypothèse que ce sont les variations relatives du Centre de Gravité Spectral qui sont importantes du point de vue perceptif, plus encore que les valeurs absolues. En effet, le Centre de Gravité Spectral dépend du spectre externe qui est à même d'être modifié par de nombreux facteurs liés à l'enregistrement (par ex. position du microphone), ou l'environnement acoustique (par ex. salle réverbérante, chambre anéchoïque). Il n'est pas illogique de penser que la forme des variations dynamiques du spectre sont elles moins sensibles à ces facteurs. La formulation du descripteur SC utilisant le seuil stabilisateur b_0 sera par la suite employée (notamment au chapitre 5 lors de comparaisons des variations de SC pour différentes interprétations musicales).

L'évolution du Centre de Gravité Spectral (SC avec b_0) est une signature de la forte augmentation de brillance qui se produit au fur et à mesure que l'intensité augmente. Ce dernier passe d'environ 290 Hz, au temps $t \approx 1s$, à environ 1800 Hz, au temps $t \approx 18s$, ce qui représente un écart de plus de 1500 Hz, soit une augmentation de plus de 500 % entre ses valeurs "initiale" et "finale" (le CGS n'est pas comparable à la hauteur d'une note, mais cependant rappelons à titre indicatif que l'oreille est capable de percevoir des écarts fréquentiels de l'ordre de 1 Hz [Wier et al., 1977]²⁰). Cet exemple traduit bien le type de changement de *timbre-qualité* que l'instrumentiste est à même d'engendrer lors de la production d'une note.

3.3.4 Effets des trous latéraux ouverts

Benade a montré qu'un réseau de trous latéraux ouverts jouait le rôle d'un filtre passe-bas en réflexion (soit un filtre passe-haut en transmission) [Benade, 1976]. Les basses fréquences sont en partie réfléchies au voisinage du premier trou ouvert, et en partie transmises à l'extérieur. Au delà d'une certaine limite, appelée **fréquence de coupure du réseau de trous latéraux**, que nous noterons f_c^{reseau} , les fréquences ne sont pas réfléchies au niveau du premier trou ouvert. Les fréquences élevées continuent à se propager au travers du réseau de trous latéraux, puis atteignent l'extrémité de l'instrument, où elles sont éventuellement réfléchies en direction de l'embouchure, ou transmises à l'extérieur. Ce phénomène a deux conséquences importantes :

²⁰Le seuil de discrimination fréquentiel dépend entre autres de la fréquence, du niveau, ainsi que du contenu harmonique des sons. La valeur de 1 Hz correspond à un son pur de 20 Hz, présenté aux auditeurs à un niveau de 40 dB [Wier et al., 1977].

- Concernant la hauteur des notes, déterminée par les résonances basses fréquences, et notamment celle du mode fondamental, **tout se passe comme si le tuyau était coupé au niveau du premier trou ouvert**. L'ouverture ou la fermeture des trous, qui modifie la longueur de tuyau effective, permet en effet au musicien de changer la hauteur des notes produites.

- **Il est pratiquement impossible d'émettre des notes dont la fréquence fondamentale est située au dessus de la fréquence de coupure $f_c^{\text{réseau}}$ (environ 1500 Hz pour une clarinette moderne)**, ce qui du reste est également appuyé par le fait que la fréquence propre de vibration de l'anche (entre 2000 et 3000 Hz selon les mesures de Thomson [Thompson, 1979]) agisse comme une fréquence de coupure pour la fréquence de jeu. L'impédance d'entrée d'un tuyau sonore percé de trous latéraux devient très faible à partir de la fréquence de coupure du réseau de trous latéraux. En effet, pour les fréquences supérieures à $f_c^{\text{réseau}}$, le tuyau n'exerce qu'une réaction négligeable sur l'anche, ce qui ne favorise guère la production des oscillations correspondantes. Benade a montré que dans le cas simplifié d'un réseau de trous réguliers (trous de même dimensions et espacement), la fréquence de coupure du réseau est à peu près constante quelques soient les notes sur les deux premiers registres de la clarinette²¹ [Benade et Kouzoupis, 1988]. Elle dépend du diamètre relatif des trous latéraux par rapport à celui du conduit cylindrique, de la hauteur des cheminées²², et de la distance entre les trous [Benade, 1976, p. 449]. On conçoit alors l'importance de ces paramètres pour le timbre de l'instrument. Leur ajustement, qui est fonction de critères musicaux (justesse, timbre, etc.), revient au savoir-faire des facteurs d'instruments.

Concernant le son externe, les trous latéraux sont essentiels puisqu'ils rayonnent une partie importante de la puissance sonore, **le son étant principalement émis par le premier, ou les deux premiers, trous latéraux ouverts en venant du bec (hormis pour la note la plus grave qui ne peut s'échapper que par le pavillon)**²³. Lorsque plusieurs trous sont ouverts, il peut se produire, pour certaines fréquences, des phénomènes d'interférences complexes. En effet, dans certaines zones de l'espace, la puissance sonore à une fréquence donnée peut s'annuler, tandis qu'elle peut être renforcée dans d'autres. Par exemple, dans le cas simplifié où le réseau de trous est régulier, en un point situé à égale distance de deux trous consécutifs, l'effet de ces deux trous s'annule pour les vibrations de longueur d'onde égale au double de l'intervalle entre les deux trous, puisque ces derniers rayonnent l'onde avec les mêmes amplitudes mais en opposition de phase [Kergomard, 1991, p. 233].

Enfin, mentionnons également le rôle primordial du **trou de registre** qui a une action tant sur la hauteur de la note que sur le timbre. En effet, lorsque l'instrumentiste ouvre ce trou, la clarinette passe sur son registre clairon, situé à un octave plus une quinte (soit une douzième²⁴) au-dessus des notes initiales : on dit à ce titre que **la clarinette quinte**. En effet, contrairement aux autres instruments à anche, la clarinette ne peut pas octavier. Ceci résulte directement de la géométrie de son tuyau dont découle une impédance acoustique favorisant les harmoniques impairs (à l'inverse, les tuyaux coniques fermé/ouvert tels que ceux du saxophone et de la trompette, ou bien les tuyaux cylindriques ouvert/ouvert tels que ceux de la flûte traversière, présentent des pics d'impédances pour toutes les fréquences harmoniques). Le deuxième pic d'impédance du résonateur de clarinette correspond comme nous l'avons vu au troisième harmonique, dont la note correspondante est une

²¹La notion de registre ici est celle employé par les physiciens : le premier registre correspond aux notes associées au premier mode de résonance du résonateur (de fréquence f_1), tandis que le second registre correspond aux notes associées au deuxième mode de résonance du résonateur (de fréquence $f_2 \approx 3f_1$) [Debut, 2004, p. 16].

²²La cheminée d'un trou latéral correspond au canal cylindrique creusé dans les parois de l'instruments, par lequel l'air s'échappe.

²³Le rayonnement de la clarinette est donc très variable en fonction des registres.

²⁴La clef permettant au musicien d'ouvrir le trou de registre est d'ailleurs appelée clef de douzième (cf. photo de la figure 3.1(b)).

douzième (le lecteur pourra se reporter à la figure 1.3 du chapitre 1 pour une correspondance entre rangs harmoniques et notes musicales associées). Le trou de registre, de plus petite dimension qu'un trou latéral, ne perturbe pas suffisamment la propagation des ondes pour entraîner une réflexion aussi intense que celle d'un trou latéral. Cependant, cette réflexion est telle qu'elle favorise un fonctionnement associé au deuxième pic d'impédance, de fréquence triple du fondamental, tout en amortissant et déplaçant fréquemment le mode de vibration fondamental qui se trouve ainsi défavorisé.

3.3.5 Le pavillon

Les notes les plus graves de la clarinette, pour lesquelles il n'y a pas ou peu de trous ouverts, ne sont pas soumises au filtrage passe-bas (en réflexion) résultant du réseau de trous latéraux. **Afin que la sonorité de l'instrument reste homogène au travers des différents registres, le pavillon, évasement terminal de l'instrument, produit un effet similaire à celui d'un réseau de trou latéral (filtre passe-bas en réflexion).** Il tend en effet à transmettre les fréquences élevées vers l'extérieur (au dessus d'une certaine fréquence de coupure f_c^{pavillon}), mais à réfléchir vers l'arrière les fréquences basses. Outre l'effet de l'évasement du pavillon, l'importante rupture de section entre l'extrémité ouverte d'un tuyau et l'espace dans lequel il rayonne entraîne les basses fréquences à se réfléchir bien plus que les hautes fréquences, qui du fait de leurs faibles longueurs d'onde sont transmises plus facilement vers l'extérieur.

3.3.6 La dissipation d'énergie

Les pertes par dissipation sont liées à plusieurs phénomènes. Nous avons vu qu'au niveau des trous latéraux ouverts, ou au niveau du pavillon, une partie de l'énergie est rayonnée vers l'extérieur, émettant ainsi l'onde sonore souhaitée. Il est remarquable de constater que la **perte d'énergie par rayonnement, sans laquelle l'instrument perdrait tout intérêt, ne représente qu'une très faible partie de l'énergie acoustique mise en jeu (environ un pour cent)**. Les pertes par rayonnement sont d'autant plus fortes que la fréquence est élevée et le rayon de sortie du tuyau est grand.

La dissipation d'énergie à l'intérieur du tuyau est surtout fixée par la viscosité de l'air ainsi que sa conduction calorifique. Elle a lieu essentiellement près des parois de l'instrument par frottements visqueux (les molécules d'air près des parois frottent les unes avec les autres, ce qui produit de la chaleur). Ces pertes, couramment appelées **pertes visco-thermiques** sont d'autant plus fortes que la fréquence est élevée et le rayon du tuyau est étroit.

3.3.7 L'inharmonicité du résonateur

Si le résonateur de la clarinette était purement cylindrique et sans pertes, ses résonances seraient harmoniques (la seconde fréquence de résonance serait exactement $f_2 = 3f_1$, la troisième exactement $f_3 = 5f_1$, etc.). Cependant un examen précis de la géométrie de la perce du tuyau révèle des irrégularités par rapport à un cylindre parfait (cf. par ex. [Debut, 2004, p. 17]). Ces irrégularités comptent parmi les facteurs à l'origine d'une légère inharmonicité du résonateur de l'instrument (leurs effets dépend de la fréquence). Ceci est souligné par Laloë, "Si son tuyau sonore était parfaitement cylindrique, les cinq premières de ses résonances au moins seraient effectivement bien alignées. Hélas, l'évasement du pavillon, la constriction du bec, les trous latéraux sont des "perturbateurs" qui détruisent ce bel ordre harmonique [NDLR : ajoutons à ces causes, le trou de registre, le gradient de température, la dispersion visco-thermique [Debut, 2004]]. [...] sur beaucoup d'instruments (en tout cas ceux que nous avons mesurés), les pics d'impédances des trois notes les plus graves de l'instrument sont mal alignés, les fréquences de résonance étant plus basses que trois fois, cinq fois, etc., celle du premier pic ; c'est la raison pour laquelle les notes

correspondantes baissent d'autant plus que l'on joue fort. L'effet reste cependant faible, de l'ordre de 0,5 pour cent en fréquence." [Laloë et Laloë, 1995, p. 84].

Mentionnons également que la **fréquence de jeu** diffèrent légèrement de celle du pic d'impédance d'entrée du résonateur correspondant. A faible intensité, la théorie montre que les fréquences de jeu sont déterminées par les zéros de la partie imaginaire de l'admittance d'entrée du résonateur [Grand et al., 1996] (elles sont proches des fréquences des maxima du module de l'impédance d'entrée, mais légèrement plus basses). L'anche a également une petite influence sur la fréquence de jeu pour différentes raisons. D'une part, en raison du débit volumique créée par le mouvement de l'anche qui s'additionne à celui produit par la différence de pression au travers de l'ouverture du canal d'anche. D'autre part, à hautes fréquences, la fréquence de résonance de l'anche, couplée à son amortissement, tendent à "tirer" la fréquence de jeu. Ces effets peuvent être pris en compte au moyen de **corrections de longueur** ΔL^{25} , à peu près indépendantes de la fréquence, et consistant à considérer un tuyau légèrement plus grand que sa longueur physique réelle [Nederveen, 1969; Dalmont et al., 1995]. Les effets des perturbations liées à la géométrie du tuyau peuvent également être pris en compte par des corrections de longueur. Concernant l'effet du pavillon, Levine et Schwinger ont proposé une correction de longueur $\Delta L \approx 0.61a$, où a est le rayon du tuyau, valable en basse fréquences²⁶ [Levine et Schwinger, 1948]. Concernant l'effet des trous latéraux ouverts, Benade a calculé des formules simplifiées de corrections de longueur dépendant de la fréquence [Benade, 1976, p. 450].

La clarinette est connue pour ses problèmes d'intonation (notamment les douzièmes correspondants aux quatre notes les plus graves restent légèrement trop grandes, de 20 à 30 cents [Debut, 2004, p. 15]. Debut et al. ont étudié si les relations d'(in)harmonicité entre les pics de résonance de l'instrument entraînées par les perturbations du tuyau et les différentes pertes, permettaient d'expliquer les défauts de justesse observés par les musiciens [Debut, 2004]. Ils concluent que les différentes sources d'inharmonicité de la clarinette se compensent globalement sur l'ensemble du registre, ce qui a amené les auteurs à supposer que les problèmes d'intonation ne découlent pas du résonateur mais proviennent des effets d'anche et/ou de l'interaction avec le conduit vocal.

3.3.8 Rôle de l'instrumentiste

L'instrumentiste contrôle de nombreux paramètres qui agissent sur le couplage entre l'anche et le tuyau sonore, et par conséquent influent sur le timbre des sons. Mentionnons les principaux :

La pression d'alimentation statique est l'un des paramètres de contrôle les plus importants pour le timbre (cf. couplage non-linéaire entre l'excitateur et le résonateur). Elle détermine également (avec la pince) le niveau sonore.

La force de pression (pince) exercée par l'instrumentiste au moyen de sa lèvre inférieure et de ses dents limite la longueur vibrante de l'anche, et contrôle son amortissement (lorsque l'anche n'est pas suffisamment amortie, cela conduit à des sons très aigus liés à sa propre résonance, les "canards", sons de prédilection de tout débutant !). Les dents du clarinettiste qui permettent d'affermir la lèvre participent également à l'amortissement. Ces facteurs sont donc déterminants pour la fréquence de jeu. D'autre part, la pince permet également de moduler un paramètre important pour le timbre, l'ouverture du canal d'anche, qui contrôle la quantité d'air qui rentre dans l'instrument, et influe sur la dynamique de l'anche (cf. par ex. situation d'anche battante).

²⁵La longueur effective du tuyau L_{eff} est alors telle que : $L_{eff} = L + \Delta L$, où L est la longueur réelle entre l'entrée de l'instrument et le premier trou ouvert (ou pavillon).

²⁶i.e. $ka \ll 1$, avec $k = \frac{2\pi}{\lambda}$, où k désigne le nombre d'ondes, et λ la longueur d'onde.

Le conduit vocal est l'un des éléments permettant au clarinettiste d'affiner le couplage avec l'instrument, afin d'ajuster l'intonation, de produire des effets (par ex. vibrato), ou bien encore de modifier le timbre. Si son rôle important est depuis longtemps connu des musiciens, le sujet est encore mal compris des physiciens, et fait l'objet de nombreuses recherches. Fritz et Wolfe ont par exemple montré que différentes combinaisons du conduit vocal (par ex. "aw", "ee") entraînaient des différences notables dans les distributions d'énergie spectrale des sons résultants [Fritz et Wolfe, 2004]. Le modèle physique de synthèse de clarinette que nous avons utilisé, et dont la description fait l'objet de la prochaine partie, prend en compte un modèle simplifié de conduit vocal [Guillemain, 2007].

La langue permet à l'instrumentiste de produire une grande variété d'attaques en contrôlant l'arrivée d'air et le moment où l'anche est relâchée.

Le choix du matériel et le style de jeu - Il revient au musicien de choisir un matériel qui corresponde au son recherché, tout en restant en adéquation avec sa morphologie. Nous avons vu que le système bec-anche était à ce sujet primordial (cf. section 3.3.1). Les deux pôles sont l'anche forte avec un bec plutôt "ouvert", ou au contraire, l'anche souple, avec un bec plutôt "fermé". Dans la première situation, qui correspondrait à un style "allemand" (cf. 2.1.3.2), l'instrumentiste peut jouer une grande partie de la dynamique de jeu sans que le bout de l'anche ne ferme complètement l'ouverture à son extrémité. Il y a alors d'importants changements de timbre lorsque la pression d'alimentation varie [Benade, 1976, p. 446]. Les sons correspondants sont globalement plutôt "sourds". Selon Benade, dans la deuxième situation, qui correspondrait à un style "français", l'ouverture du canal d'anche se ferme pendant une portion importante de chaque cycle de ses oscillations (ibidem). Ceci entraîne une augmentation importante de l'énergie des harmoniques de rangs élevés. Les sons correspondants sont globalement plus "brillants" que dans la première situation. Lorsque le niveau de dynamique change, Benade souligne que le spectre interne connaît dans ce cas moins de variations (région à droite de la ligne verticale indiquant "change of feel", sur la figure 3.3). Comme nous l'avons déjà mentionné en parlant du timbre, la plupart des clarinettistes optent pour un compromis entre ces deux situations extrêmes.

Dans un cadre plus général, l'Annexe E rapporte une analyse des différents facteurs impliqués dans la relation instrumentale, selon le clarinettiste Claude Crousier.

3.4 Modèle de synthèse de clarinette reposant sur un modèle physique simplifié

Si les bases de la physique de la clarinette sont à ce jour maîtrisées, il n'en reste pas moins une multitude de phénomènes complexes à mieux comprendre (par ex. régimes d'oscillations pour des pressions d'alimentation élevées, non-linéarités de l'anche, régimes transitoires, rayonnement, influence dynamique du conduit vocal, intonation). On conçoit dès lors que l'exploitation de modèles de synthèse reposant sur un modèle physique simplifié ne permettra que de déblayer dans ses grandes lignes les relations entre contrôle de l'instrument et timbre généré. Ceci n'ôte cependant pas tout l'intérêt que l'on peut y porter.

De nombreuses méthodes de calcul des auto-oscillations d'instruments à anche simple tel que la clarinette ont été proposées. Les approches fréquentielles telle que celle proposée par Worman [Worman, 1971], ainsi que celle développée par Gilbert et Kergomard, reposant sur la méthode numérique d'équilibrage harmonique [Gilbert et Kergomard, 1989], permettent d'accéder aux solutions stationnaires du régime périodique. Elles sont utiles pour étudier la hauteur de la note et l'évolution du spectre en fonction des paramètres du modèle, mais ne permettent pas d'étudier la

stabilité des solutions. Les approches temporelles, telle que celle de Schumacher [Schumacher, 1981], qui utilise une fonction de réflexion, permettent d'analyser la stabilité des solutions et les transitoires d'attaque, essentiels au niveau perceptif. D'autres méthodes ont été mises au point dans le but d'effectuer les calculs en temps-réel. Nous citerons les méthodes par guide d'onde numérique (cf. par ex. [Smith, III]), et la modélisation numérique d'ondes (cf. par ex. [van Walstijn et Campbell, 2003]), reposant sur la formulation des ondes progressives et régressives²⁷. Le modèle physique de synthèse développé par Guillemain *et al.*, dont l'une des applications est la synthèse sonore numérique contrôlable en temps-réel, n'utilise que les variables physiques (pression et débit acoustiques) (cf. [Guillemain *et al.*, 2005]). Le modèle que nous avons utilisé repose sur une variation du modèle précédent, qui considère également le conduit vocal de l'instrumentiste [Guillemain, 2007]. Le schéma de synthèse associé repose sur un modèle d'impédance numérique, ce qui permet de prendre en compte facilement différentes géométries de résonateur (ceci, associé à des modifications de la modélisation de l'anche, permet de réaliser la synthèse de différents instruments à vent, par ex. clarinette, saxophone, hautbois, trompette, au moyen d'un même modèle générique). Nous avons employé une version basique de ce modèle qui fonctionne hors temps-réel (implémentation sous *Matlab*), suffisante pour générer des sons de manière isolée (cf. étude de dissemblance de timbres exposée au chapitre 4). Nous décrivons dans cette section les principes généraux du modèle utilisé.

3.4.1 Modèle physique simplifié

3.4.1.1 Hypothèses générales

Le fluide (air) est supposé réel (visqueux) et incompressible. L'écoulement est supposé laminaire (le nombre de Reynolds²⁸ associé à l'écoulement est environ égal à 500 [Kergomard, 1995, p. 250], irrotationnel, unidimensionnel. Une propagation linéaire en ondes planes selon la direction longitudinale du tuyau de l'instrument est considérée. Les forces de pesanteur agissant sur l'air sont négligées. Le modèle ne prend pas en compte de réseaux de trous latéraux. Ceux-ci, comme nous l'avons vu dans la section précédente, sont pourtant très importants pour les amplitudes des spectres interne et externe de l'instrument (ils induisent un effet de fréquence de coupure) [Benade et Kouzoupis, 1988]. Nous verrons dans la section 3.4.2.1 que le schéma de synthèse utilisé surestime les pertes visco-thermiques, ce qui permet - indirectement - de pallier l'absence de prise en compte de réseaux de trous latéraux. Pour les mêmes raisons, et ainsi que décrit dans les sections 3.4.1.4 et 3.4.2.1, la formulation de l'impédance d'entrée de l'instrument sera simplifiée en supposant que l'impédance de rayonnement Z_s à la sortie de l'instrument est nulle. Par rapport aux oscillations ayant lieu à l'intérieur de l'instrument, l'influence du rayonnement sera approchée par une simple correction de longueur $\Delta L \approx 0.61R$, où R est le rayon du conduit cylindrique modélisant le corps de l'instrument [Levine et Schwinger, 1948]. En revanche, concernant le spectre externe, nous utiliserons une approximation de l'impédance de rayonnement telle que proposée par Silva *et al.* [Silva *et al.*, 2008], d'après les travaux de Levine et Schwinger [Levine et Schwinger, 1948], et Norris et Sheng [Norris et Sheng, 1989] (cf. sections 3.4.1.7 et 3.4.2.2). Du point de vue de la dispersion, les effets des réseaux de trous latéraux et du pavillon seront négligés, notre but premier étant d'étudier le timbre de l'instrument, et non la justesse de l'instrument²⁹.

²⁷Les ondes progressives sont celles qui s'éloignent de la source (de l'embouchure vers un trou ouvert et/ou pavillon), tandis que les ondes régressives sont celles qui s'en rapprochent.

²⁸Le nombre de Reynolds Re est un nombre sans dimension qui représente le rapport entre les forces d'inertie et les forces visqueuses. Pour les faibles valeurs de Re (inférieures à 2000), les forces de viscosité sont prépondérantes, et l'écoulement est *laminaire*, i.e. il ne présente pas trop de variations spatiales ou temporelles, contrairement à un écoulement *turbulent*.

²⁹Notons toutefois que les effets de dispersion peuvent être également un élément constitutif du timbre, mais nous les jugeons d'ordre second.

Bien que simplifié le modèle utilisé est toutefois plus évolué que les modèles classiques sans pertes, ou celui dit de “Raman”³⁰ qui ignore la dispersion, et suppose que les pertes par dissipation, incluant celles par rayonnement sont indépendantes de la fréquence (cf. [Dalmont *et al.*, 2005]).

3.4.1.2 Description géométrique du système et notations

La figure 3.5 présente une description de la géométrie du système, ainsi que des variables physiques qui lui sont associées. Le corps de l’instrument est considéré comme parfaitement cylindrique de longueur L et de rayon R . L’aire de la section transverse de l’instrument sera notée S ($S = \pi R^2$).

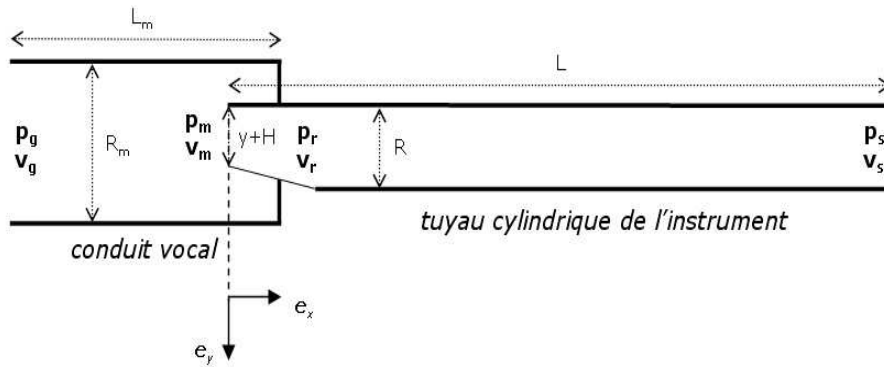


FIG. 3.5 – Description géométrique du modèle et des variables physiques employées.

La pression de la bouche $p_m(t)$ résulte d’une pression de souffle $p_g(t)$ imposée à l’entrée du conduit vocal, au niveau de la glotte. Le conduit vocal est également modélisé par un tuyau cylindrique, de rayon R_m et de longueur L_m . L’aire de la section transverse du conduit vocal sera notée S_m ($S_m = \pi R_m^2$). La pression $p_r(t)$ correspond à la pression à l’entrée du corps de l’instrument (en sortie du canal d’anche). $v_g(t)$, $v_m(t)$ et $v_r(t)$ sont respectivement les vitesses acoustiques au niveau de la glotte, de la bouche, et de l’entrée du corps de l’instrument. Les débits acoustiques associés seront notés $u_g(t)$, $u_m(t)$ et $u_r(t)$ ³¹.

H correspond à la hauteur de l’ouverture du canal d’anche lorsque cette dernière est au repos. Le déplacement de l’anche sera noté $y(t)$. Il vaut 0 lorsque l’anche est à l’équilibre, et $-H$ lorsque l’anche obture totalement l’ouverture du canal d’anche.

3.4.1.3 L’anche

Le modèle d’anche repose sur une représentation simple de type masse/ressort/amortissement (système à un degré de liberté). Cette représentation soutient l’hypothèse que le mouvement de l’anche est à une dimension, selon la direction transverse verticale. Cette hypothèse s’avère fondée pour des modes de jeux normaux et des oscillations de faibles amplitudes. Nous citerons à ce propos Backus, dont les observations du mouvement de l’anche grâce à une méthode photo-électrique, n’ont pas révélé de mouvements transverses horizontaux : “These observations showed that as far as could be seen visually there were no transverse oscillations of the reed under proper conditions. Under certain maladjustments the loud squeals (NDLR : “canards”) obtained showed occasionally some evidence of transverse vibrations, but under proper adjustment the end of the

³⁰Le nom est issu des travaux de l’auteur sur les oscillations des cordes frottées.

³¹Rappelons que débit u_i et vitesse acoustique v_i sont reliés par l’aire de la section transverse S_i du conduit dans lequel a lieu l’écoulement : $u_i = S_i v_i$.

reed moved as a unit with no perceptible phase shift from one side to the other." [Backus, 1963, p. 807].

Le déplacement de l'anche $y(t)$ par rapport à sa position d'équilibre $y = 0$, lorsqu'elle soumise à la différence de pression $\Delta p(t) = p_m(t) - p_r(t)$, est donné par l'équation de dynamique suivante (cf. par ex. [Wilson et Beavers, 1974]) :

$$\mu_r \frac{d^2 y(t)}{dt^2} + C_r \frac{dy(t)}{dt} + K_r y(t) = -\Delta p(t) \quad (3.2)$$

où μ_r , C_r et K_r sont respectivement la masse surfacique, l'amortissement, et la raideur de l'anche (égale au quotient entre la différence de pression $p_m - p_r$ et le déplacement $-y$, lorsque la charge est statique). Nous introduirons également la plus basse fréquence propre de l'anche f_r ,

et sa fréquence angulaire associée $\omega_r = 2\pi f_r = \sqrt{\frac{K_r}{\mu_r}}$. Le facteur de qualité de l'anche s'écrit

$Q_r = \frac{\mu_r \omega_r}{C_r} = \frac{\sqrt{\mu_r K_r}}{C_r}$. Des valeurs typiques des caractéristiques physiques de l'anche sont données dans l'Annexe D.

La pression de placage statique de l'anche p_M (pression pour laquelle le canal d'anche se ferme) est donnée par [Kergomard, 1995] :

$$p_M = \mu_r H \omega_r^2 = K_r H \quad (3.3)$$

p_M sera supposée constante.

Traduite dans le domaine fréquentiel, l'équation 3.2 donne la fonction de transfert linéaire de l'anche :

$$\frac{Y(\omega)}{\Delta P(\omega)} = \frac{\omega_r^2}{\omega_r^2 - \omega^2 + j \frac{\omega \omega_r}{Q_r}} \quad (3.4)$$

Notons que les caractéristiques physiques de l'anche sont considérées ici comme constantes. Cette approximation omet de possibles modifications dynamiques de l'amortissement, et de la raideur de l'anche (et donc de sa fréquence de résonance) selon la tension, et la position de la lèvres de l'instrumentiste. Toutefois, l'évolution des caractéristiques physiques de l'anche en fonction de la force d'appui de l'instrumentiste n'est pas encore connue.

3.4.1.4 Les résonateurs (instrument et conduit vocal)

Instrument Le tuyau de l'instrument est considéré comme parfaitement cylindrique (cf. figure 3.5). De plus, nous supposons que son rayon R est grand devant l'épaisseur de la couche limite (écoulement laminaire). La théorie classique de Kirchoff permet d'exprimer **les pertes par dissipation visco-thermiques**, ayant lieu lors de la propagation des ondes dans le tuyau (cf. par ex. [Pierce, 1981]). Ces pertes sont dépendantes de la fréquence (en $\sqrt{\omega}$), et sont prises en compte par le nombre d'onde complexe $k'(\omega)$, défini pour une onde plane de fréquence angulaire ω par :

$$\begin{cases} k'(\omega) = k(\omega) - j^{\frac{3}{2}} \eta c \sqrt{\omega}, \text{ avec} \\ \eta = \frac{1}{R c^{\frac{3}{2}}} (\sqrt{l_v} + (\frac{C_p}{C_v} - 1) \sqrt{l_t}) \end{cases} \quad (3.5)$$

où $k(\omega) = \frac{\omega}{c}$ est le nombre d'onde classique, j est l'unité imaginaire, c est la vitesse du son dans l'air, l_v et l_t sont les longueurs des couches limites visqueuses et thermiques, C_p est la chaleur spécifique à pression constante, et C_v la chaleur spécifique à volume constant. Des valeurs typiques de ces différents paramètres pourront être trouvées dans l'Annexe D.

Le résonateur de l'instrument est caractérisé par son impédance d'entrée $Z_r(\omega)$, qui relie de manière linéaire les transformées de Fourier $P_r(\omega)$ et $U_r(\omega)$ des pression et débit d'entrée du résonateur de l'instrument $p_r(t)$, et $u_r(t)$. Dans le cas d'un conduit cylindrique et sous l'hypothèse que la pression p_s à la sortie de l'instrument est nulle, cette dernière est donnée par :

$$Z_r(\omega) = jZ_c \tan(k'(\omega)L) \quad (3.6)$$

où $Z_c = \frac{\rho c}{S}$ est l'impédance caractéristique du tuyau, ρ étant la densité moyenne de l'air (cf. Annexe D).

Conduit vocal En raison de la vocation du modèle proposé dans [Guillemain, 2007] à la synthèse temps-réel, et pour des raisons de simplicité, le conduit vocal de l'instrumentiste est également considéré comme étant parfaitement cylindrique (cf. figure 3.5). Les pertes visco-thermiques, prises en compte dans le nombre d'onde complexe $k'_m(\omega)$, associé aux ondes se propageant dans le conduit vocal, repose également sur l'expression 3.5. Cependant, le rayon utilisé dans le modèle de pertes n'est pas le rayon géométrique R_m du modèle de conduit vocal, mais un rayon plus petit R_p et ajustable, dont le rôle est de considérer des pertes additionnelles dues à la nature des tissus humains, mous et absorbants³². Il est également supposé que la valeur du nombre d'onde k'_m est nulle à la fréquence zéro, de sorte que la composante statique de la pression de souffle $p_g(t)$ soit transmise entièrement à la pression de la bouche $p_m(t)$. En notant $P_g(\omega)$, et $U_g(\omega)$ les transformées de Fourier de $p_g(t)$ et $u_g(t)$, le formalisme en ligne de transmission permet d'écrire :

$$\begin{cases} P_g(\omega) &= \cos(k'_m(\omega)L_m)P_m(\omega) + jZ_m \sin(k'_m(\omega)L_m)U_m(\omega) \\ U_g(\omega) &= \frac{j}{Z_m} \sin(k'_m(\omega)L_m)P_m(\omega) + \cos(k'_m(\omega)L_m)U_m(\omega) \end{cases} \quad (3.7)$$

où $Z_m = \frac{\rho c}{S_m}$ est l'impédance caractéristique du tuyau modélisant le conduit vocal. Ceci conduit à :

$$P_m(\omega) = \frac{2e^{-jk'_m(\omega)L_m}}{1 + e^{-2jk'_m(\omega)L_m}} P_g(\omega) - jZ_m \tan(k'_m(\omega)L_m) U_m(\omega) \quad (3.8)$$

qui montre que les résonances du conduit vocal sont celles d'un résonateur quart d'onde³³.

Le modèle fait l'hypothèse que les paramètres du conduit vocal R_m , L_m , et R_p sont constants au cours du temps. Ceci permet déjà d'obtenir différentes configurations du conduit vocal de manière statique. Cependant, il paraît clair que dans des situations réelles, l'instrumentiste peut être amené à modifier son conduit vocal au cours du temps, en vue d'agir sur la hauteur, ou sur le timbre (cf. section 3.3.8). Concernant la synthèse temps-réel, la prise en compte de telles modifications dynamiques du conduit vocal pourrait s'avérer très intéressante. Afin qu'elles puissent être effectives, il faudrait également disposer de contrôleurs qui soient capables de les mesurer, ce qui

³²Les pertes visco-thermiques sont en effet plus élevées dans un tuyau étroit que dans un tuyau large, car l'épaisseur de la couche limite relative au rayon du tuyau est plus grande, i.e. l'amortissement aux parois est plus grand [Fletcher et Rossing, 1998, p. 202].

³³En effet, les pôles sont obtenus pour $1 + e^{-2jk'_m(\omega)L_m} = 0 \Rightarrow 2k'_m(\omega)L_m = (2n+1)\pi \Rightarrow \omega_m = (2n+1)\frac{\pi c}{2L_m}$, avec n entier.

n'est actuellement pas encore le cas. Des recherches sur ce sujet sont actuellement menées au sein de l'équipe S2M.

3.4.1.5 Couplage non linéaire entre l'excitateur et le résonateur

Modèle classique - Le modèle classique de couplage entre l'anche vibrante simple et le résonateur de l'instrument repose sur la relation de Bernoulli stationnaire (application du théorème de la conservation de l'énergie cinétique). Appliquée à l'anche, au niveau de laquelle se forme un jet lorsqu'elle est soumise à une pression statique de la bouche p_m , la relation de Bernoulli s'écrit :

$$p_m + \frac{1}{2}\rho v_m^2 = p_r(t) + \frac{1}{2}\rho v_r^2(t) \quad (3.9)$$

Le modèle classique repose également sur les deux hypothèses suivantes :

Hypothèse 1 - L'énergie cinétique du jet qui rentre dans l'instrument est complètement dissipée par turbulence lors de son développement dans l'embouchure, soit $v_m = 0$. L'équation 3.9 implique donc :

$$\Delta p(t) = p_m - p(t) = \frac{1}{2}\rho \left(\frac{u_r(t)}{S_{jet}} \right)^2 \quad (3.10)$$

où $u_r(t)$ est le débit dans l'embouchure, proportionnel à l'aire de la section transverse du jet S_{jet} .

Hypothèse 2 - L'aire de la section transverse du jet S_{jet} est supposée être égale à l'aire de l'ouverture du canal d'anche $S_y(t)$, elle même supposée rectangulaire et proportionnelle à la hauteur de l'ouverture de la fente $y(t)$:

$$S_{jet}(t) = S_y(t) = w(y(t) + H) \quad (3.11)$$

où w est la largeur de l'anche (cf. Annexe D). Le débit dans l'embouchure est alors donné par :

$$u_r(t) = S_y(t) \sqrt{\frac{2\Delta p(t)}{\rho}} \quad (3.12)$$

La relation 3.12 est appelée la **caractéristique non-linéaire de l'embouchure**.

Modèle utilisé - Le modèle utilisé ne fait pas la première des deux hypothèses décrites précédemment, i.e. la vitesse $v_m(t)$ au niveau de la bouche du musicien n'est plus négligée. La relation de Bernoulli conduit alors à deux équations l'une exprimant le débit $u_r(t)$ à l'entrée du résonateur pour un jet entrant dans l'instrument, et l'autre l'exprimant pour un jet entrant dans la bouche (cf. [Guillemain, 2007]). Cependant, pour des raisons de simplicité, l'ouverture du canal d'anche est supposée négligeable devant celle de la bouche, et devant la section du tuyau de l'instrument (i.e. $\frac{wH}{S_m} \approx 0$ et $\frac{wH}{S} \approx 0$). Ceci conduit à considérer une caractéristique non-linéaire du même type que celle formulée par l'équation 3.12, mais avec $\Delta p(t) = p_m(t) - p_r(t)$.

3.4.1.6 Variables adimensionnées

Tel que proposé dans [Wilson et Beavers, 1974], puis dans [Kergomard, 1995], le modèle utilise des variables adimensionnées $\tilde{p}_g, \tilde{u}_g, \tilde{p}_m, \tilde{u}_m, \tilde{p}_r, \tilde{u}_r$. Elles sont définies à partir des variables physiques par les relations suivantes :

$$\begin{cases} \tilde{p}_{g,m,r}(t) &= \frac{p_{g,m,r}(t)}{p_M} \\ \tilde{u}_{g,m,r}(t) &= Z_c \frac{u_{g,m,r}(t)}{p_M} \end{cases} \quad (3.13)$$

Les versions adimensionnées des impédances acoustiques à l'entrée $\tilde{Z}_r$, et à la sortie du résonateur $\tilde{Z}_s$, correspondent au quotient des impédances acoustiques Z_r et Z_s par l'impédance caractéristique du résonateur de l'instrument Z_c .

Principaux paramètres de contrôle du modèle Outre la longueur L du conduit cylindrique déterminant la hauteur de la note émise, le modèle dispose de deux autres paramètres de contrôle importants. Le premier est la **pression d'alimentation adimensionnée** $\tilde{p}_g(t)$ (au niveau de la glotte) tel que définie dans l'équation 3.13. Nous la noterons également $\gamma(t)$. Le second, tel que proposé dans la référence [Kergomard, 1995], est un **paramètre adimensionné caractéristique de l'embouchure du musicien**, qui sera noté $\zeta(t)$. $\zeta(t)$ est notamment corrélé à l'ouverture du canal d'anche à l'équilibre H (ce qui permet de traduire la tension d'embouchure de l'instrumentiste). Il est défini par :

$$\zeta(t) = wH(t)Z_c \sqrt{\frac{2}{\rho p_M}} \quad (3.14)$$

La formulation adimensionnée de l'ouverture du canal d'anche $S_y(t)$ sera notée $W(t)$.

3.4.1.7 Le rayonnement

Si l'on suppose que le rayonnement est nul, le coefficient de réflexion $R_s(\omega)$ à l'extrémité du tuyau cylindrique vaut -1 (quelque soit leur fréquence, les ondes acoustiques sont totalement réfléchies au bout du tuyau, avec une inversion de la phase). A l'inverse, si l'on prend en compte la dissipation et la dispersion entraînées par le rayonnement, le coefficient de réflexion $R_s(\omega)$ peut s'écrire (cf. [Silva et al., 2008]) :

$$R_s(\omega) = -|R_s(\omega)|e^{-2jk(\omega)\Delta L(\omega)} = \frac{Z_s(\omega) - 1}{Z_s(\omega) + 1} \quad (3.15)$$

où $|R_s(\omega)|$ exprime les pertes par dissipation dues au rayonnement, $e^{-2jk(\omega)\Delta L(\omega)}$ concerne la dispersion due au rayonnement ($\Delta L(\omega)$ est un terme de correction de longueur qui dépend de la fréquence), et $Z_s(\omega)$ est l'impédance de rayonnement. Cette dernière peut s'exprimer ainsi :

$$Z_s(\omega) = \frac{R_s(\omega) + 1}{R_s(\omega) - 1} = j \tan(k(\omega)\Delta L(\omega) - j\frac{1}{2} \ln|R_s(\omega)|) \quad (3.16)$$

où $\ln$ est le logarithme népérien.

Dans l'hypothèse d'un rayonnement monopolaire ("point source élémentaire"), la pression externe $\tilde{p}_{ext}(t)$ s'exprime de manière classique comme la dérivée temporelle du débit $\tilde{u}_s(t)$ à la sortie de l'instrument (cf. par ex. [Chaigne, 2001, p. 113]) :

$$\tilde{p}_{ext}(t) = \frac{d}{dt} \tilde{u}_s(t) \quad (3.17)$$

Exprimons donc le débit de sortie en fonction de l'impédance de rayonnement (ou du coefficient de réflexion). Nous omettrons ici de noter les dépendances en ω ou en temps t pour ne pas alourdir les expressions.

$$\begin{cases} \tilde{P}_s &= \tilde{Z}_s \tilde{U}_s \\ \Leftrightarrow \tilde{P}_s + \tilde{U}_s &= (1 + \tilde{Z}_s) \tilde{U}_s \end{cases} \quad (3.18)$$

Or, la pression $\tilde{p}_s$ correspond à la somme d'une onde plane progressive $\tilde{p}_s^+$, et d'une onde plane régressive $\tilde{p}_s^-$, ce qui conduit à écrire :

$$\begin{cases} \tilde{p}_s &= \tilde{p}_s^+ + \tilde{p}_s^- \\ \tilde{u}_s &= \tilde{p}_s^+ - \tilde{p}_s^- \end{cases} \quad (3.19)$$

$$\Rightarrow \tilde{p}_s + \tilde{u}_s = 2\tilde{p}_s^+ \quad (3.20)$$

Or,

$$\tilde{p}_s^+ = \tilde{p}_r^+ e^{-jk'L} \quad (3.21)$$

où le terme exponentiel correspond à un aller dans le tuyau. Similairement à la relation 3.20, $\tilde{p}_r + \tilde{u}_r = 2\tilde{p}_r^+$, donc :

$$\tilde{p}_s + \tilde{u}_s = (\tilde{p}_r + \tilde{u}_r) e^{-jk'L} \quad (3.22)$$

La relation 3.18 s'écrit alors :

$$\begin{cases} (1 + \tilde{Z}_s) \tilde{U}_s &= (\tilde{P}_r + \tilde{U}_r) e^{-jk'L} \\ \Rightarrow \tilde{U}_s &= \frac{\tilde{P}_r + \tilde{U}_r}{1 + \tilde{Z}_s} e^{-jk'L} \end{cases} \quad (3.23)$$

L'équation 3.17 exprimée dans le domaine fréquentiel, ainsi que l'équation 3.23, conduisent à l'expression suivante de la pression externe :

$$\tilde{P}_{ext}(\omega) = j\omega \tilde{U}_s(\omega) = j\omega \frac{\tilde{P}_r(\omega) + \tilde{U}_r(\omega)}{1 + \tilde{Z}_s(\omega)} e^{-jk'(\omega)L} \quad (3.24)$$

Du point de vue perceptif, le terme $e^{-jk'(\omega)L}$ pourra être négligé puisqu'il correspond au retard et aux pertes subies par la pression acoustique au cours d'un simple aller dans le tuyau depuis l'embouchure jusqu'à son extrémité. De plus, en exprimant $1 + Z_s$ en fonction de R_s , d'après l'expression de gauche de l'équation 3.16, nous obtenons :

$$\tilde{P}_{ext}(\omega) = \frac{1}{2} j\omega (1 - R_s(\omega)) (\tilde{P}_r(\omega) + \tilde{U}_r(\omega)) \quad (3.25)$$

Nous avons utilisé les formules approchées de calcul du module du coefficient de réflexion à la sortie du tuyau, et de la correction de longueur proposées par Silva *et al.* [Silva *et al.*, 2008], d'après les résultats de Levine et Schwinger [Levine et Schwinger, 1948].

Remarquons que si l'on ne tient pas compte du rayonnement ($Z_s(\omega) = 0$), ce qui revient à dire que la réflexion en bout de tube est totale (le coefficient de réflexion $R_s(\omega)$ vaut -1), l'expression 3.25 se simplifie en :

$$\tilde{P}_{ext}(\omega) = j\omega (\tilde{P}_r(\omega) + \tilde{U}_r(\omega)) \quad (3.26)$$

Une comparaison des effets des relations 3.25 (avec rayonnement), et 3.26 (sans rayonnement) sur la pression externe est présentée à la section 3.4.2.2.

3.4.1.8 Résumé du modèle complet

Le modèle physique complet est régi par :

- La pression d'alimentation $p_g(t)$;
- Le système d'équations 3.7 décrivant la propagation de la pression et du débit dans le conduit vocal ;
- L'impédance d'entrée de l'instrument $Z_r(\omega)$, définie par la relation 3.6 ;
- L'équation de la dynamique d'anche 3.2 ;
- La caractéristique non linéaire $u_r(t) = F(S_y(t), \Delta p(t))$ (équation 3.12), où F représente la fonction non linéaire ;
- La pression externe donnée par la relation 3.25.

Les formulations équivalentes de ces relations reposant sur les variables adimensionnées, présentées à la section 3.4.1.6, sont données dans la référence [Guillemain, 2007].

3.4.2 Construction du modèle de synthèse

3.4.2.1 Modélisation numérique de l'impédance d'entrée

L'impédance d'entrée adimensionnée $\tilde{Z}_r(\omega)$, définie équation 3.6, peut s'écrire sous la forme :

$$\tilde{Z}_r(\omega) = \frac{1 + R_r(\omega)}{1 - R_r(\omega)} \quad (3.27)$$

où $R_r(\omega) = -e^{-2ik'(\omega)L}$ est le coefficient de réflexion à l'entrée du résonateur, qui décrit la propagation et la dissipation visco-thermique. Ce coefficient de réflexion est modélisé par un filtre passe bas numérique du premier ordre dont la transformée en z , $F(z)$, est donnée par :

$$F(z) = -\frac{b_0 z^{-D}}{1 - a_1 z^{-1}} \quad (3.28)$$

avec $z = e^{j\frac{\omega}{F_e}}$, où F_e est la fréquence d'échantillonnage, et $D = \lfloor F_e \frac{2L}{c} \rfloor$ est le retard que met l'onde pour parcourir un aller-retour dans le tuyau. Les coefficients du filtre sont définis tels que les amplitudes des deux premiers pics d'impédance, situés aux fréquences f_1 et f_2 , du modèle numérique correspondent à celles du modèle continu (cf. figure 3.6). Le retard de groupe du filtre numérique est compensé grâce à une correction de longueur³⁴.

Comme on peut le constater sur la figure 3.6, le modèle numérique du coefficient de réflexion R_r surestime largement les pertes dissipatives pour les fréquences élevées. Paradoxalement, ceci peut être vu comme un avantage puisque cela permet de "tenir compte" de pertes dissipatives additionnelles du tuyau qui ne sont pas modélisées (par ex. pertes dissipatives dues au rayonnement, aux réseaux de trous latéraux, etc.), qui comme nous l'avons rappelé ont un important rôle aux fréquences élevées (diverses fréquences de coupure).

L'influence de la modélisation numérique du coefficient de réflexion peut être également observée sur le module de l'impédance d'entrée du résonateur de l'instrument (cf. figure 3.7). Conformément aux effets observés pour le coefficient de réflexion, le modèle numérique d'impédance d'entrée du résonateur présente des modes d'ordres élevés bien plus amortis que celui du modèle continu sans impédance d'entrée terminale (les pics sont plus larges et moins élevés). Le filtre a également pour effet de réhausser les zéros du module de l'impédance d'entrée par rapport au

³⁴Notons qu'il ne s'agit pas là d'une correction de longueur liée à la dispersion causée par des perturbations du tuyau.

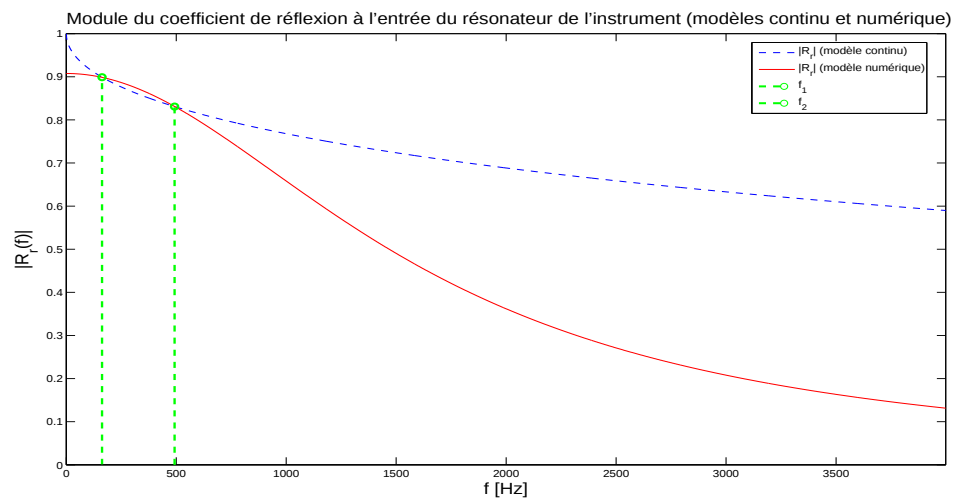


FIG. 3.6 – Module du coefficient de réflexion $|R_r(f)|$ à l'entrée du résonateur de l'instrument : comparaison des modèles continu sans impédance de rayonnement terminale, et numérique.

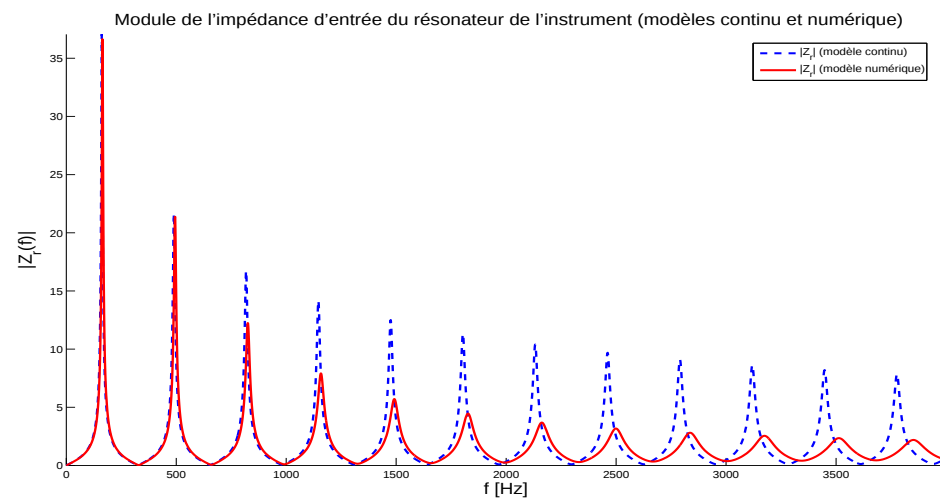


FIG. 3.7 – Module de l'impédance d'entrée $|Z_r(f)|$ du résonateur de l'instrument : comparaison des modèles continu et numérique ($f_1 \approx 164.8 \text{ Hz}$ (E3, écrit $F\sharp 3$)).

modèle continu (et ce dès 2000 Hz environ), tel que le fait l'effet d'un réseau de trous latéraux ou d'un pavillon. A partir de cette fréquence, l'impédance d'entrée est alors non négligeable aux fréquences des harmoniques pairs. Bien que son amplitude reste faible, ceci est l'un des éléments qui explique la présence d'harmoniques pairs à hautes fréquences dans le spectre des sons générés par l'instrument. Les harmoniques pairs à basses fréquences dans le spectre externe proviennent essentiellement du débit (cf. équation 3.25).

La figure 3.8 présente les modules des impédances d'entrée numériques sans et avec prise en compte du rayonnement. Cette comparaison permet de montrer que les pertes par rayonnement sont négligeables devant les pertes induites par le modèle numérique, ce qui explique pourquoi nous ne les considérons pas dans le calcul de l'impédance d'entrée.

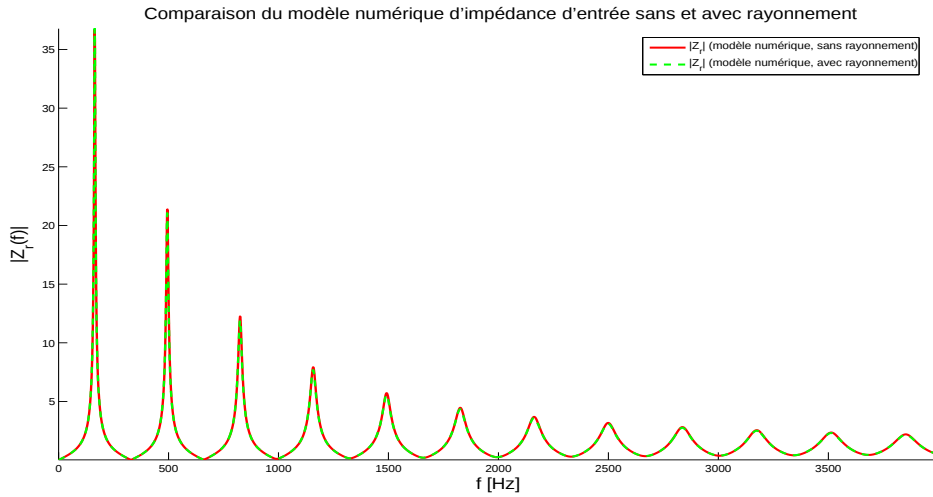


FIG. 3.8 – Effet de la prise en compte du rayonnement sur le modèle numérique d'impédance d'entrée. En trait plein : $|Z_r|$ en considérant $Z_s = 0$, en traits pointillés : $|Z_r|$ en considérant $Z_s \neq 0$.

Le même type de filtre que celui décrit par la relation 3.28 est utilisé pour modéliser les pertes et la propagation dans le conduit vocal.

3.4.2.2 Modélisation numérique du rayonnement

Exprimées avec la transformée en z , les formulations de la pression externe avec et sans prise en compte du rayonnement (cf. équations 3.25 et 3.26) deviennent :

$$P_{ext}(z) = T_s(z)(P_s(z) + U_s(z)) \quad (3.29)$$

où $T_s(z)$ est une fonction de transfert de transmission telle que :

$$T_s(z) = \begin{cases} \frac{F_e}{2}(1 - z^{-1})(1 - R_s(z))(P_s(z) + U_s(z)) , & \text{avec rayonnement} \\ F_e(1 - z^{-1})(P_s(z) + U_s(z)) , & \text{sans rayonnement} \end{cases} \quad (3.30)$$

Le terme $1 - z^{-1}$ est l'approximation numérique la plus simple de la dérivée, exprimée comme la différence entre les échantillons à l'instant n et les échantillons à l'instant $n - 1$.

Le filtre utilisé pour modéliser le coefficient de réflexion R_s à la sortie du résonateur de l'instrument est un filtre passe-bas d'ordre 1, défini tel que :

$$R_s(z) = -\frac{1 - a_s - b_s + b_s z^{-1}}{1 - a_s z^{-1}} \quad (3.31)$$

La figure 3.9 présente une comparaison des modèles continu (selon [Silva et al., 2008]) et numérique du coefficient de réflexion de sortie. Le filtre numérique dont la fonction de transfert est donnée équation 3.31 permet de réaliser une très bonne approximation du modèle continu. Notons cependant une divergence entre les deux courbes à partir d'environ 18 kHz, ce qui n'a *a priori* pas d'effets indésirés, puisque les ondes sonores de fréquences élevées sont déjà fortement atténuées par les pertes dissipatives visco-thermiques (cf. impédance d'entrée sur la figure 3.7). Tel que prévu, les pertes augmentent avec la fréquence (le coefficient de réflexion devient plus faible),

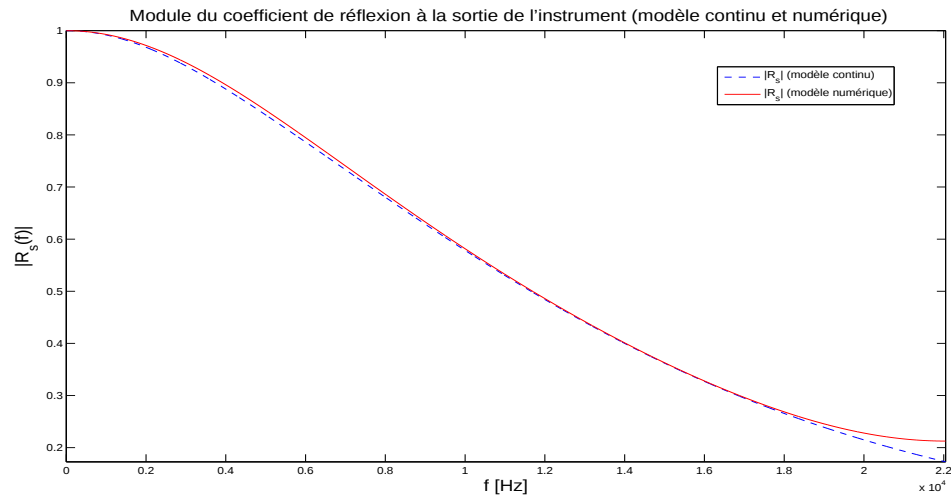


FIG. 3.9 – Module du coefficient de réflexion $|R_s(f)|$ à la sortie du résonateur de l'instrument : comparaison des modèles continu (selon [Silva *et al.*, 2008]) et numérique (cf. équation 3.31).

les ondes de fréquences élevées étant plus facilement transmises vers l'extérieur du fait de leurs faibles longueurs d'ondes. A l'opposé, les ondes de basses fréquences tendent à être réfléchies complètement avec changement de signe (en 0, le coefficient de réflexion vaut bien -1).

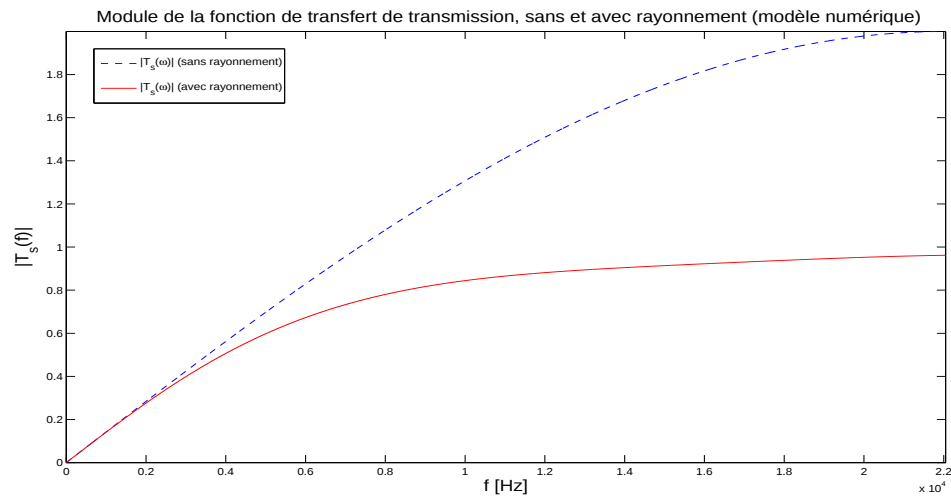


FIG. 3.10 – Influence du rayonnement sur le module de la fonction de transfert de transmission T_s (modèle numérique). En traits pointillés : sans rayonnement, en trait plein : avec rayonnement.

Comme l'illustre la figure 3.10, lorsque les pertes dissipatives par rayonnement ne sont pas prises en compte, les fréquences élevées sont surestimées (cf. fonction de transmission en traits pointillés). Le modèle de rayonnement (cf. première des deux équations 3.30) permet d'éviter cet effet indésirable pour le spectre externe (cf. fonction de transmission en trait plein sur la figure 3.10), et ainsi de se rapprocher du fonctionnement réel de l'instrument.

3.4.3 Exemples de simulation

En vue de présenter certains aspects du comportement du modèle, nous avons simulé un crescendo de clarinette, allant du *piano* (*p*) au *fortissimo* (*ff*). Le son correspondant est l'[Exemple Sonore n° 34](#). Cet exemple sonore n'a pas été synthétisé de façon à reproduire le crescendo naturel analysé à la section 3.4.1.5 ([Exemple Sonore n° 33](#)). Cependant, il pourra être intéressant de comparer qualitativement les spectres obtenus. Le son synthétisé est beaucoup plus court (6 s) que le son naturel, mais dispose d'une fréquence de jeu similaire ($f_{jeu} \approx 240\text{ Hz}$).

Pour simuler un crescendo de manière très simplifiée, nous avons généré un profil de pression d'alimentation adimensionnée $\gamma(t)$ qui augmente au cours du temps et maintenu fixe le paramètre d'ouverture du canal d'anche à l'équilibre (soit $\zeta(t) = c^{te} = 0.2$). Les paramètres de contrôles utilisés sont représentés dans le haut de la figure 3.11. Les phases d'attaques et de relâchement du profil de pression d'alimentation sont modélisées par des branches de tangentes hyperboliques. Dans la portion soutenue, le crescendo est simulé par une évolution linéaire de la pression d'alimentation (de $\gamma_1 = 0.4$ à $\gamma_2 = 0.8$). Le bas de la figure 3.11 représente la pression externe obtenue.

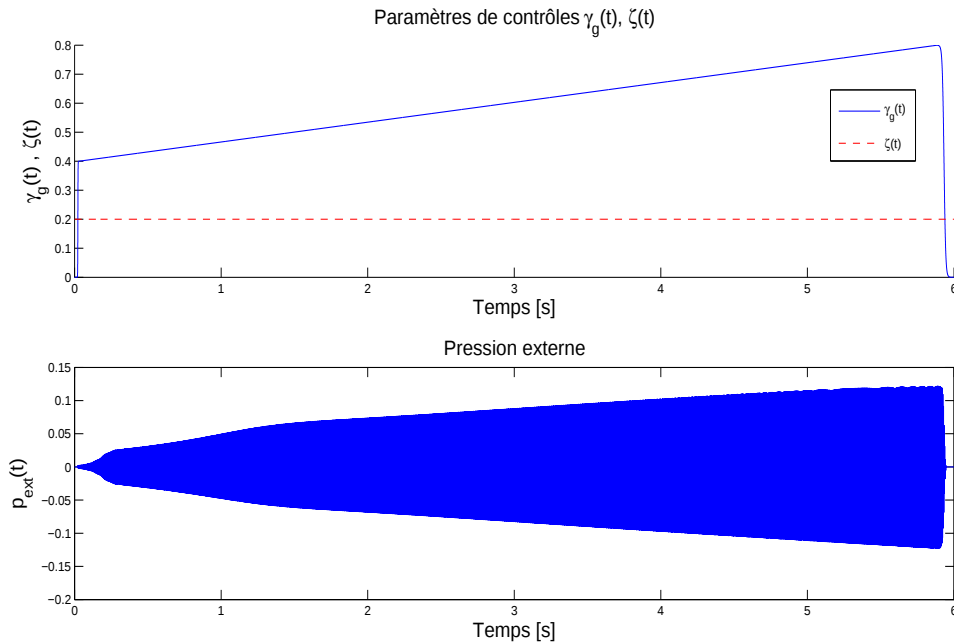


FIG. 3.11 – En haut : paramètres de contrôles $\gamma(t)$ (trait plein) et $\zeta(t)$ (traits pointillés) utilisés afin de générer un crescendo par synthèse. En bas : pression externe.

Le spectrogramme et Centre de Gravité Spectral instantané correspondant au son simulé sont représentés figure 3.12. Nous noterons que comme pour le son naturel, les harmoniques impairs tendent à apparaître en premier, et leurs amplitudes tendent à dominer celles des harmoniques pairs. Les harmoniques pairs semblent être cependant bien plus présents dans le son synthétique que dans le son naturel. Les harmoniques de rangs élevés apparaissent également de manière progressive, au fur et à mesure que le niveau sonore augmente. Ceci se traduit par une augmentation de la brillance du son au cours du temps, caractérisée par l'augmentation du Centre de Gravité Spectral, qui évolue d'environ 300 Hz jusqu'à 2000 Hz durant la progression du son. D'une manière globale, le spectre du son synthétique est beaucoup plus riche que le spectre du son naturel. Ceci est très probablement dû au fait certaines pertes ne sont pas prises en compte dans le modèle

simplifié, bien que les pertes visco-thermiques soient déjà surestimées. Notons que les paramètres de contrôle utilisés ici ont été générés artificiellement, ce qui permet difficilement le rapprochement avec des sons naturels générés par de vrais instrumentistes. Dans le prochain chapitre, nous utiliserons en revanche un profil de pression mesuré.

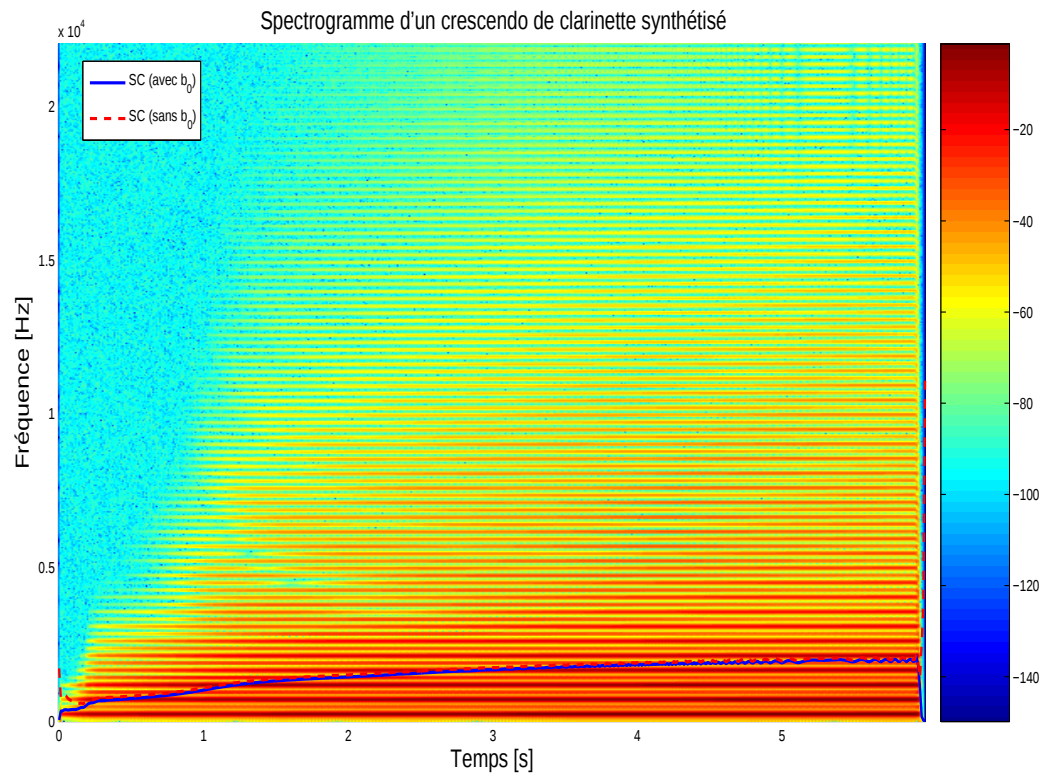


FIG. 3.12 – Spectrogramme d'un crescendo de clarinette synthétisé ($f_{jeu} \approx 240Hz$). L'évolution du Centre de Gravité Spectral (SC) au cours du temps est superposée à la représentation temps-fréquence. Deux formulations du descripteur SC sont comparées, celle avec le seuil stabilisateur b_0 (trait plein), et celle sans seuil (traits pointillés). L'analyse a été effectuée avec des fenêtres de Hann, d'environ 20 ms. Les amplitudes, normalisées relativement à l'amplitude maximale, sont exprimées en décibels (dB).

La figure 3.13 représente la caractéristique non linéaire $u_r = F(\Delta p)$ pour quelques cycles d'oscillations correspondant aux nuances *piano* ($t \approx 40ms$), *mezzo forte* ($t \approx 1.4s$), et *fortissimo* ($t \approx 5s$). Hormis pendant la phase d'attaque, pour des faibles variations de pression et un faible débit, la relation entre débit et pression est approximativement linéaire (cf. phase *piano*). Ceci correspond à des oscillations quasi-sinusoidales des variables physiques telles que l'ouverture du canal d'anche, le débit, et les pressions interne et externe (cf. figure 3.14). Lorsque la pression d'alimentation augmente (cf. phase *mezzo forte*), la plage de variations de pression devient plus importante. La relation entre débit et pression n'est plus linéaire, ce qui entraîne des oscillations plus "chahutées", dont le spectre contient davantage d'harmoniques aigus (cf. figure 3.15). Lorsque la pression d'alimentation dépasse un certain seuil (cf. phase *fortissimo*), les amplitudes d'oscillation de l'anche sont telles qu'elle ferme l'embouchure durant une portion (environ une demi-période) de son cycle d'oscillation (ce qui correspond à la situation d'anche battante). Tel que le montre la figure 3.16, ceci a pour conséquence d'annuler le débit dans l'embouchure. Sa forme d'onde n'est plus symétrique, mais "rognée" d'un côté, et ce approximativement toutes

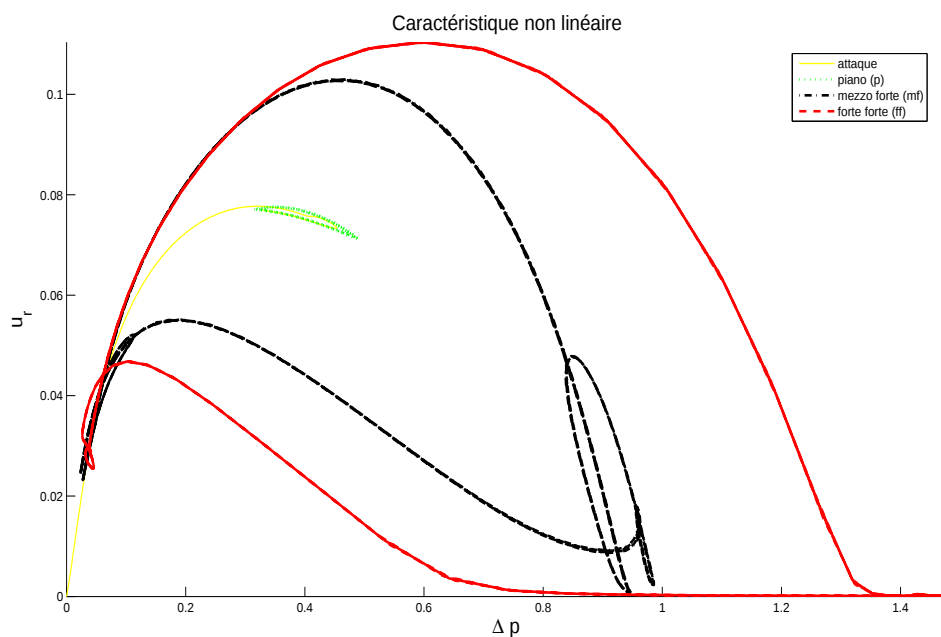


FIG. 3.13 – Caractéristique non linéaire $u_r = F(\Delta p)$. La caractéristique non-linéaire est représentée sur des portions du signal (d'environ 5 périodes d'oscillations) correspondant à différents niveaux de dynamique : attaque (trait plein -), *piano* (pointillés ...), *mezzo forte* (alternance traits pleins et pointillés -.), *fortissimo* (traits pointillés - - -).

les demi-périodes. Il s'ensuit que la pression externe contient encore plus d'harmoniques de rangs élevés, comme le confirme le spectrogramme.

L'[Exemple Sonore n° 35](#) présente une simulation d'une courte séquence musicale, obtenue par juxtaposition de différentes notes générées de manière isolée. Le "canard numérique" que l'on peut entendre sur l'une des notes de cette séquence indique, qu'à la manière d'un vrai instrument, le modèle produit de tels sons si le contrôle n'est pas adéquat.

De par le fait que le modèle physique sur lequel repose la synthèse est simplifié, les sons générés présentent des différences avec les sons naturels (ils sont par exemple plus riche spectralement). Cependant, comme nous avons pu le constater, le modèle de synthèse reproduit bien le comportement dynamique de l'instrument, tel que traduit par la caractéristique non-linéaire. Ceci garantit d'une part l'identité timbrale des sons de clarinette, et d'autre part une reproduction fidèle des variations de timbre-qualité entraînées par les variations des principaux paramètres de contrôle de l'instrument (pression et pince). De plus, un tel modèle permet de générer des sons de hauteurs et durées similaires de manière reproductible. Il s'avère pour cela particulièrement intéressant afin d'étudier la relation entre les contrôles de l'instrument, les timbres produits par le modèle, et leur perception. Dans l'étude exposée au prochain chapitre, nous nous attacherons à analyser et interpréter la structure de l'espace de timbre perceptif du modèle.

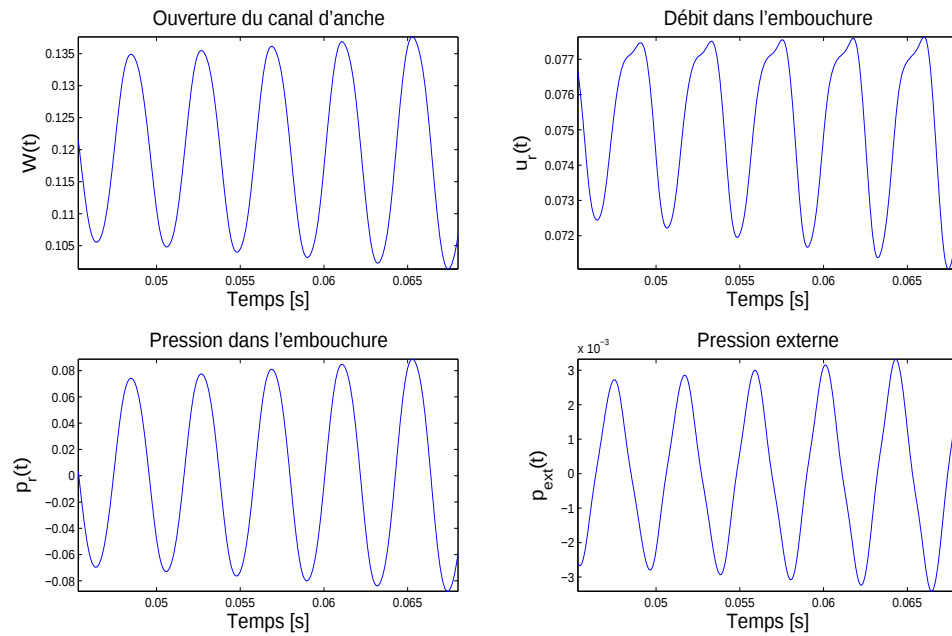


FIG. 3.14 – Environ 5 cycles d'oscillations de l'ouverture du canal d'anche (en haut, à gauche), du débit dans l'embouchure (en haut, à droite), des pressions dans l'embouchure (en bas, à gauche) et externe (en bas, à droite), dans une situation de jeu *piano*.

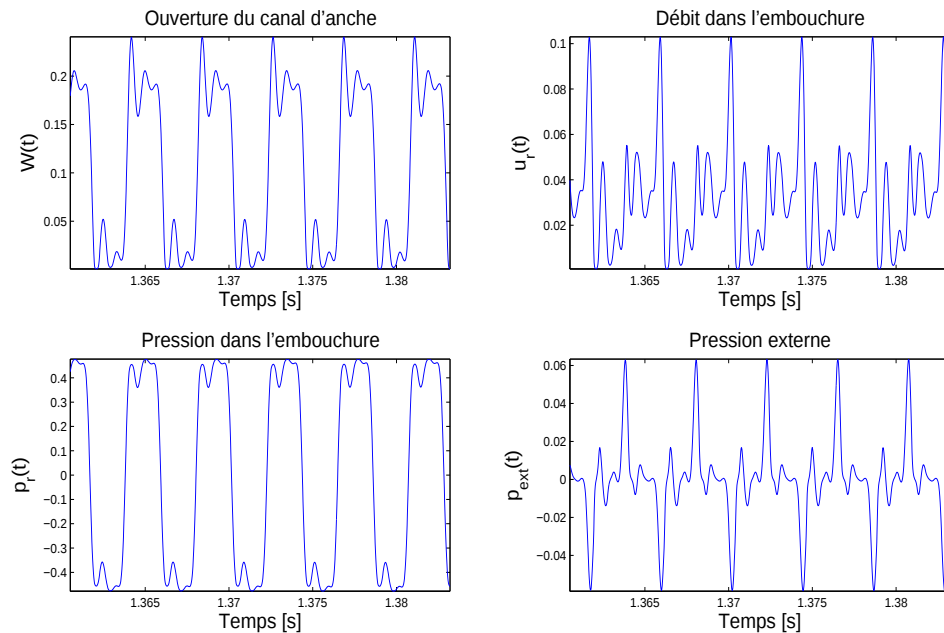


FIG. 3.15 – Environ 5 cycles d'oscillations de l'ouverture du canal d'anche (en haut, à gauche), du débit dans l'embouchure (en haut, à droite), des pressions dans l'embouchure (en bas, à gauche) et externe (en bas, à droite), dans une situation de jeu *mezzo forte*.

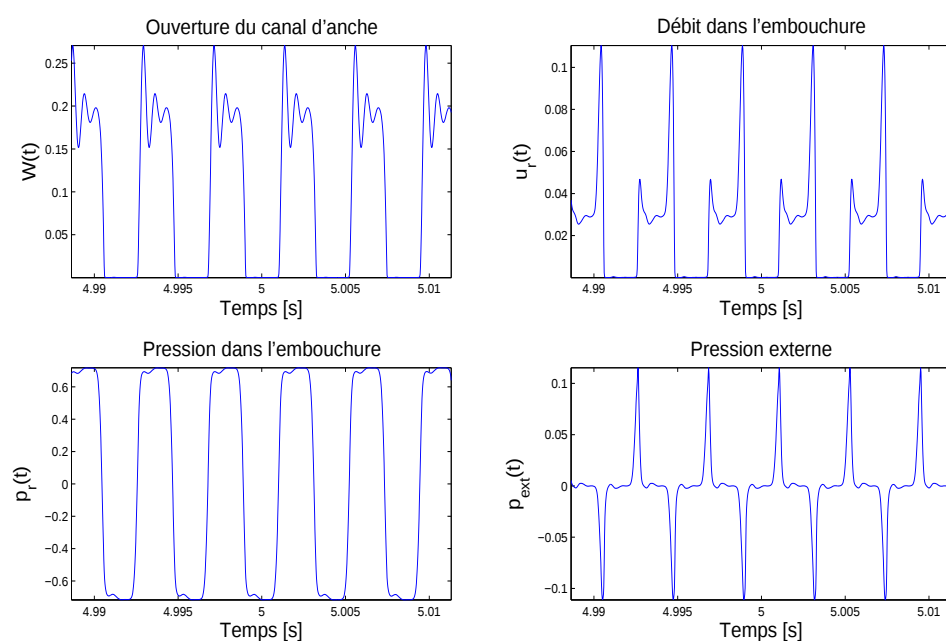


FIG. 3.16 – Environ 5 cycles d'oscillations de l'ouverture du canal d'anche (en haut, à gauche), du débit dans l'embouchure (en haut, à droite), des pressions dans l'embouchure (en bas, à gauche) et externe (en bas, à droite), dans une situation de jeu *fortissimo*.

Chapitre 4

Du contrôle de la clarinette à la perception du timbre

Certains des travaux exposés dans ce chapitre sont présentés dans l'article [Barthet *et al.*, 2009c].

4.1 Introduction

Le modèle décrit dans le chapitre précédent a été utilisé afin d'étudier la relation entre le contrôle de l'instrument et le timbre des sons générés. Nous avons mené une étude perceptive de dissemblance de timbre à partir d'un ensemble de sons représentatifs des différents timbres produits par le modèle. Nous présenterons dans un premier temps les méthodes expérimentales employées lors de l'étude de dissemblance (par ex. génération des stimuli, procédure du test perceptif, analyses statistiques). La structure de l'espace de timbre perceptif obtenu à l'issue de l'expérience sera ensuite analysée à partir des descripteurs de timbre présentés dans la section 2.2.5, et des paramètres de contrôle du modèle.

4.2 Méthodes

Une description de la méthodologie employée pour les études de dissemblance de timbre a été donnée dans la section 2.2.4.1. Nous décrivons ici son application au cas de sons de clarinette générés avec le modèle de synthèse décrit dans le chapitre précédent.

4.2.1 Corpus sonore

Nous nous sommes concentrés sur l'influence des deux paramètres de contrôle principaux de la clarinette, i.e. la pression d'alimentation, et la force de pression (également appelée pince) appliquée par l'instrumentiste sur l'anche au moyen de sa lèvre inférieure. Comme nous l'avons vu au chapitre précédent, cette dernière a pour effet de modifier la position de l'anche au repos et donc l'ouverture du canal d'anche. Les paramètres de contrôle adimensionnés du modèle γ et ζ sont respectivement corrélés à la pression d'alimentation et à la pince (cf. section 3.4.1.6). Afin de se concentrer sur des conditions de jeu ordinaires, le problème a été restreint à l'emploi d'une pince constante au cours du temps (i.e. $\zeta(t) = c^{te}$). En effet, les variations de la pince pendant la production des notes sont surtout employées afin de produire des effets particuliers, tels que le vibrato. En vue de produire des sons réalistes, un profil de pression d'alimentation a été mesuré lors de la production d'une note par un musicien en situation de jeu.

4.2.1.1 Mesures de la pression d'alimentation

Nous avons demandé à un instrumentiste de produire un son soutenu relativement court sur la deuxième note la plus grave de la clarinette (le $E3$, $f_0 \approx 164.81$ Hz), en employant le contrôleur numérique *Yamaha WX-5*, adapté aux instruments à vent numériques, et le modèle de synthèse temps-réel (implémenté dans l'environnement *MAX/MSP*), dont la version simplifiée a été présentée au chapitre précédent. La note $E3$ a été choisie car son doigté entraîne l'ouverture d'un seul trou latéral (le premier en remontant du pavillon vers le bec). En effet, le modèle simplifié décrit à la section 3.4, qui ne prend pas en compte de réseau de trous latéraux, est dans ces conditions bien adapté.

Deux mesures différentes de la pression d'alimentation ont été effectuées. La première a été réalisée à l'aide d'une sonde de pression placée à l'intérieur de la bouche du musicien. La sonde de pression, un tuyau de très faible section (environ 0.5 mm), est reliée à un capteur de pression, de type transducteur piezzo-électrique¹, monté sur un circuit imprimé. Ce dispositif a été développé par Didier Ferrand et Arnaud Farcy dans le cadre du stage de ce dernier, portant sur l'influence du conduit vocal dans le jeu du saxophone [Farcy, 2008]. L'acquisition des données provenant du capteur de pression est réalisée grâce au système *dSpace*, à une fréquence d'échantillonnage de 25 kHz, et sur 16 bits (32768 valeurs). La seconde mesure de pression provient du capteur de pression du contrôleur numérique *Yamaha WX-5*. Le contrôleur, qui fonctionne suivant la norme MIDI (Musical Instrument Digital Interface), renvoie des informations environ toutes les 10 ms, et codées sur 7 bits (128 valeurs). Les photos de la figure 4.2 illustrent l'ensemble du dispositif.

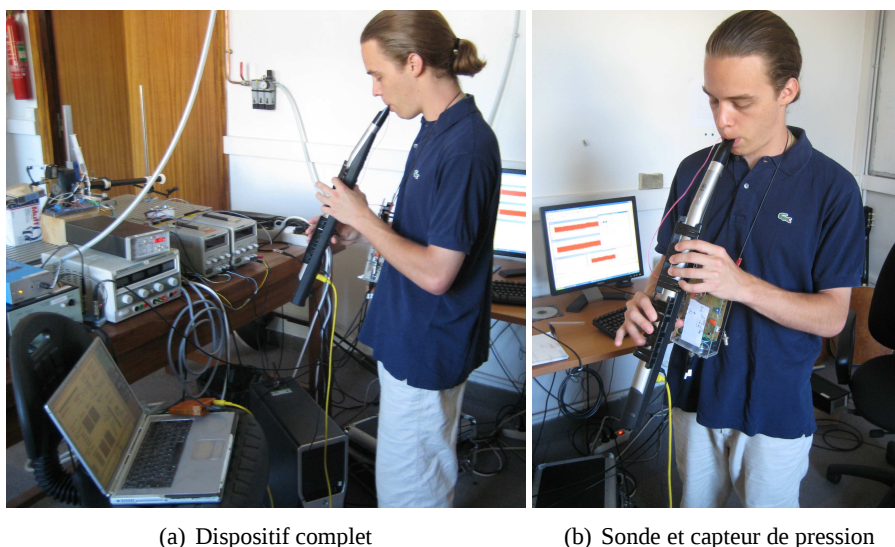


FIG. 4.1 – (a) Mesure de la pression d'alimentation en situation de jeu sur le modèle de synthèse temps-réel de clarinette. L'instrumentiste contrôle le modèle de synthèse, implémenté dans l'environnement *MAX/MSP*, via un contrôleur numérique *Yamaha WX-5*. (b) La pression est mesurée par le contrôleur numérique *Yamaha WX-5*, ainsi que par un capteur relié à une sonde de pression située dans la bouche de l'instrumentiste.

La figure 4.2 présente les deux mesures du profil temporel de pression d'alimentation. La mesure du capteur de pression a été lissée au moyen d'un filtre numérique passe-bas. Les deux mesures ont été ré-échantillonnées à 44.1 kHz et normalisées de manière à correspondre aux va-

¹La référence du capteur de pression est ASCX05DN. Ces capteurs sont linéaires jusqu'à une pression de 5 Psi, soit environ 0.35 bars (1 Psi = 0,069bar).

leurs du paramètre de contrôle adimensionné γ . La montée de pression mesurée dans la bouche par la sonde correspond jusqu'à $t \approx 0.45s$ à l'établissement de la pression à l'entrée du canal d'anche, obstrué par la langue de l'instrumentiste, avant le départ du son. Ceci explique pourquoi le capteur du WX-5 ne sort rien jusqu'à l'instant où la langue part. La comparaison des courbes montrent que la discrétisation MIDI dans le cas du WX-5 entraîne des variations de pression par paliers. Ces variations sont audibles lors d'une écoute attentive des sons produits avec une telle mesure. Nous avons pour ces raisons pris comme référence, le profil de pression mesuré grâce à la sonde et au capteur de pression. Il sera noté par la suite $\gamma^{ref}(t)$. Il est intéressant de constater que l'allure de la pression d'alimentation statique peut être bien décrite par un modèle du type ADSR (Attaque, Décroissance, Soutien, Relâchement).

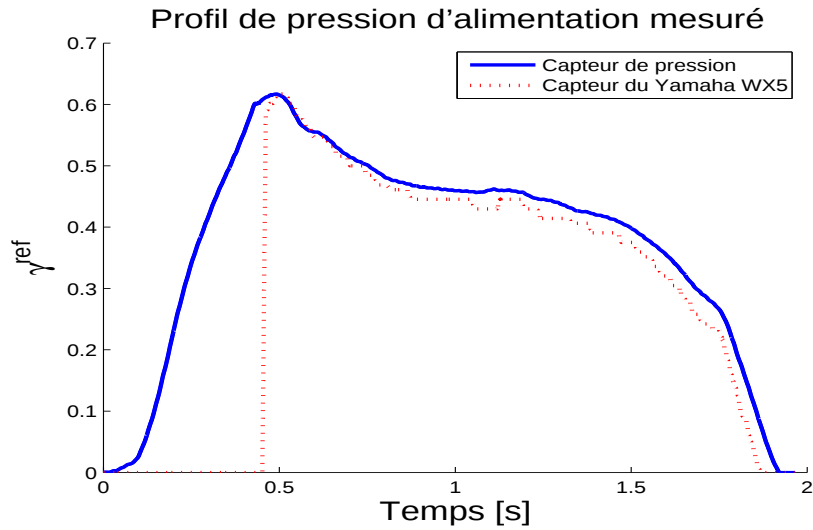


FIG. 4.2 – Profils de pression mesurés par la sonde et le capteur de pression γ^{ref} (trait plein), et par le contrôleur Yamaha WX-5 (traits pointillés).

4.2.1.2 Espace des paramètres de contrôle

Le profil de pression de référence $\gamma^{ref}(t)$ a ensuite été modifié par dilatation/contraction en amplitude, de manière à faire varier sa valeur maximale dans une certaine plage. Il est pour ce faire normalisé, puis multiplié par un facteur, noté γ_m . Les différents contrôles de pression $\gamma(t)$ sont obtenus de la manière suivante :

$$\gamma(t) = \gamma_m \frac{\gamma^{ref}(t)}{\gamma_m^{ref}} \quad (4.1)$$

où γ_m^{ref} est la valeur maximale du profil de référence $\gamma^{ref}(t)$, et γ_m est la valeur maximale du nouveau profil $\gamma(t)$.

Nous avons généré 150 sons en sélectionnant 10 valeurs γ_m dans l'intervalle $0.53 \leq \gamma_m \leq 0.80$ (valeurs maximales du profil $\gamma(t)$), et 15 valeurs de ζ dans l'intervalle $0.1 \leq \zeta \leq 0.4$. Un nombre plus élevé de valeurs de ζ a été choisi car, comme nous le verrons par la suite, de fines variations de ce paramètre peuvent entraîner des changements de timbre importants. Les différents profils de $\gamma(t)$ et $\zeta(t)$ qui ont été utilisés sont exposés figure 4.3.

Les plages de variation des paramètres ont été choisies de manière large afin d'obtenir des sons allant du *pianissimo* au *fortissimo*. Tel que le montre la figure 4.4, pour des couples (γ, ζ)

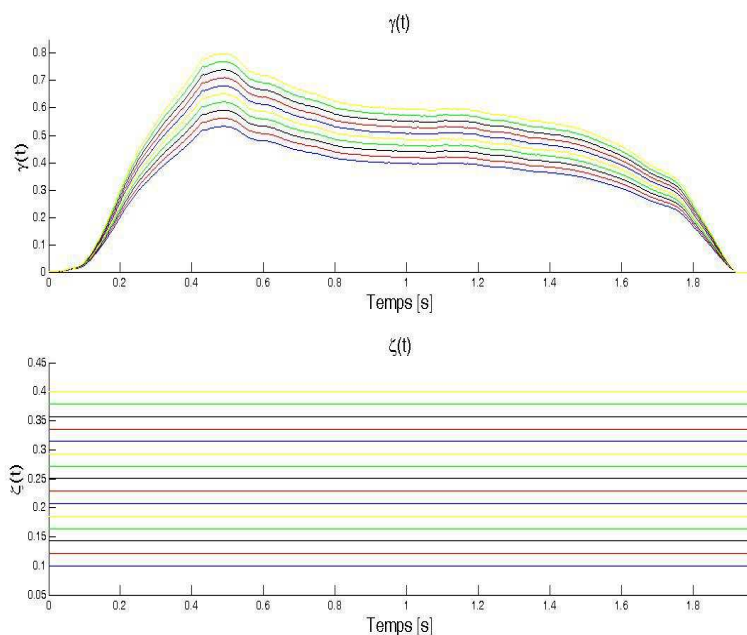


FIG. 4.3 – Profils temporels de γ (haut) et de ζ (bas) pour les 150 sons générés.

trop petits, le seuil de démarrage des auto-oscillations n'est pas atteint, et aucun son n'est produit. A l'inverse, pour des couples (γ, ζ) trop grands, les sons générés quintoient. Pour ces sons, le troisième harmonique (situé une quinte plus un octave au dessus du fondamental) tend à prendre autant voire plus d'importance que le fondamental (cf. section 3.2). Pour des valeurs de (γ, ζ) "raisonnables", les sons générés correspondent à des conditions normales de jeu. C'est le cas de 92 sons parmi les 150 générés au total. Leurs paramètres sont définis par la zone située au milieu sur la figure 4.4. Tous les sons ont une durée similaire d'environ 2 s, et sont tous perçus de même hauteur (E3).

4.2.2 Sélection des stimuli

15 sons ont été sélectionnés parmi ceux du corpus décrit précédemment, en vue de bien représenter la palette des différents timbres produits par le modèle de synthèse, dans le cas des contrôles utilisés. Notons que nous avons restreint ces choix aux sons associés à des conditions normales de jeu (pas de "canards", ni de sons qui quintoient). Différentes combinaisons des paramètres γ_m et ζ sont ainsi représentées pour des valeurs faibles, moyennes et élevées (cf. figure 4.4).

Les différents sons choisis ont par la suite été égalisés en sonie afin que l'intensité ne compte pas parmi les critères de discrimination. Plusieurs procédés d'égalisation de la sonie ont été testés. Un procédé d'ajustement automatique de la sonie a été mis au point. Ce dernier consiste à déterminer le gain permettant d'ajuster la sonie d'un son sur celle d'un autre par un procédé d'optimisation. Une fonction *Matlab* développée par Isabelle Boulet et Georges Canevet a été utilisée à cet effet. Le modèle de sonie employé est le modèle de Zwicker, qui repose sur un calcul des sonies spécifiques sur les bandes critiques (0 à 24 Barks) [Zwicker et Fastl, 1990]. Cependant, le procédé d'ajustement automatique de la sonie n'a pas offert de résultats convaincants pour les sons de synthèse de clarinette sélectionnés. Ceci est peut-être dû au fait que les sons présentent de très larges différences de richesse spectrale, et que le modèle ne compense pas assez l'influence de la brillance sur la sonie (ibidem). Pour cette raison, un procédé d'égalisation de la sonie des sons

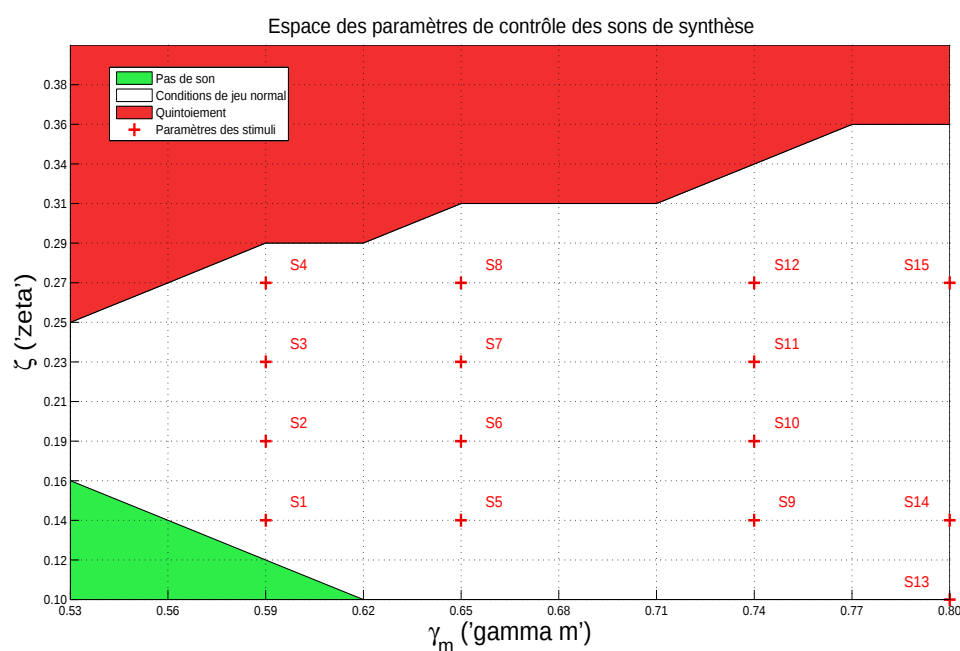


FIG. 4.4 – Nature des sons obtenus avec le modèle en fonction des paramètres de contrôle γ et ζ . Trois différentes zones sont mises en évidence : une première correspondant à des paramètres de contrôle en dessous du seuil d’auto-oscillation de l’instrument (pas de son), une seconde correspondant à une situation de jeu normale, et une troisième correspondant à des sons qui quintoient. Les valeurs des paramètres de contrôle associés aux stimuli choisis dans l’expérience perceptive sont également indiquées (cf. section 4.2.2). Notons qu’à partir de $\gamma_m = 0.62$, la frontière de seuil d’oscillation n’a pas été explorée. Il est possible qu’elle soit atteinte pour des valeurs de ζ inférieures à 0.1.

réalisé à l’écoute, par les expérimentateurs, a été préféré.

Une description qualitative des paramètres de contrôle associés aux stimuli est donnée dans le tableau 4.2.2.

Nomenclature	Description
S1-S4	γ_m faible, ζ faible à élevé
S5-S8	γ_m moyen, ζ faible à élevé
S9-S12	γ_m élevé, ζ faible à élevé
S13-S15	γ_m très élevé, ζ très faible à élevé

TAB. 4.1 – Description qualitative des stimuli en fonction des valeurs des paramètres de contrôle γ_m et ζ auxquelles ils sont associés.

4.2.3 Sélection des participants

Le test perceptif a été conduit par un groupe de 16 participants (10 hommes et 6 femmes âgés de 23 à 47 ans). 10 d’entre eux pratiquent un instrument de musique de manière régulière et/ou ont reçu une formation musicale. Aucun participant n’a signalé avoir de problèmes auditifs.

4.2.4 Appareillage

L'expérience s'est déroulée dans une cabine audiométrique isolée acoustiquement. Un programme de test de jugement de dissimilarité avec une interface utilisateur graphique a été développé à ce titre dans l'environnement *Matlab*. Le programme a été exécuté sur un ordinateur *Apple iMAC G4*. Les sons, également joués par l'ordinateur, ont été diffusés aux participants au moyen d'un casque *STAX SRM-310* disposant d'un amplificateur dédié.

4.2.5 Procédure

L'expérience repose sur des jugements de dissimilarité entre paires de sons (cf. section 2.2.4.1 pour de plus amples informations sur ce type de procédure).

Deux phases préliminaires ont précédé l'expérience. La première est une **phase d'écoute** lors de laquelle les auditeurs doivent écouter l'ensemble des stimuli. Ces derniers apparaissent dans un ordre aléatoire sur l'écran d'ordinateur. Leur lecture peut être déclenchée à la souris en cliquant sur le bouton approprié. Il a été demandé aux participants d'écouter tous les sons lors de cette étape d'une part afin de s'y familiariser, et d'autre part afin de prendre conscience de l'ambitus des différences qu'ils peuvent présenter. La seconde étape est une **phase d'apprentissage**. Quelques paires de stimuli sont présentées aux participants afin qu'ils s'exercent à effectuer la tâche demandée, et par la même occasion s'habituent à l'interface utilisateur. Durant l'expérience, les participants doivent juger les dissimilarités entre paires de sons sur une échelle allant de "Très semblables" à "Très dissemblables". Les paires de sons identiques ne sont pas prises en compte. L'ordre de présentation de deux sons donnés (par ex. [1,2] ou [2,1]) est établi de manière aléatoire afin de prendre en compte, d'un point de vue statistique, un éventuel effet d'ordre (cf. section 2.2.4.1.4). Pour les mêmes raisons l'ordre de présentation des différentes paires est aussi déterminé de manière aléatoire (par ex. [2,1], puis [10,14], etc.). L'expérience compte un total de 105 paires ($\frac{15 \times 14}{2}$).

A la fin de l'expérience, les participants ont été invités à noter sur un questionnaire les principaux critères utilisés lors de la discrimination des sons.

4.2.6 Analyses statistiques

4.2.6.1 Homogénéité des réponses des participants

Une classification ascendante hiérarchique (analyse en cluster) des mesures de dissimilarités des participants a été menée dans un premier temps. Elle a pour but de déterminer si certains participants présentent des réponses systématiquement différentes de celles des autres. Dans un tel cas, il est alors préférable de mener les analyses statistiques des données sur des groupes distincts, ou bien d'en exclure, si cela est justifié, les réponses de certains participants. Les mesures de distance entre les réponses des participants sont obtenues à partir du coefficient de corrélation de Pearson (selon son complément par rapport à 1).

4.2.6.2 Analyse multidimensionnelle de proximité (MDS)

Pour un échantillon considéré homogène, les matrices de dissimilarité individuelles ont été moyennées afin d'obtenir une matrice de dissimilarité globale, représentative de l'échantillon. Cette dernière est ensuite analysée selon une technique d'analyse multidimensionnelle de proximité non-métrique (MDSCAL) (voir par ex. [Dillon et Goldstein, 1984]).

4.2.6.3 Classification ascendante hiérarchique

Une classification ascendante hiérarchique a également été effectuée à partir de la matrice de dissimilarité globale. Elle permet de révéler les catégories de sons qui présentent des différences systématiques.

4.2.7 Descripteurs de timbre

Afin de trouver des prédicteurs acoustiques de la structure perceptive sous-jacente à la discrimination des différents sons, un certain nombre de descripteurs de timbre a été employé. Le tableau 4.2 en précise la liste, ainsi que leur nomenclature. La description et la formulation de ces paramètres peuvent être trouvées au chapitre 2, à la section 2.2.5.

Type et nomenclature	Nom du descripteur (français)	Nom du descripteur (anglais)
Temporels		
AT	Temps d'Attaque	Attack Time
LAT	Logarithme du Temps d'Attaque	Logarithm of the Attack Time
RT	Temps de Relâchement	Release Time
TC	Centre de Gravité Temporel	Temporal Centroid
Spectraux ou spectro-temporels		
SC	Centre de Gravité Spectral	Spectral Centroid
SS	Étalement Spectral	Spectral Spread
SSK	Asymétrie Spectrale	Spectral Skewness
SKU	Coefficient d'Aplatissement Spectral	Spectral Kurtosis
SRO	Roulement Spectral	Spectral Roll-off
SF	Flux Spectral	Spectral Flux
SFAT	Flux Spectral pendant l'attaque	Spectral Flux Attack
Harmoniques		
HSC	Centre de Gravité Spectral Harmonique	Harmonic Spectral Centroid
OSC	Centre de Gravité Spectral Impair	Odd Spectral Centroid
ESC	Centre de Gravité Spectral Pair	Even Spectral Centroid
OER	Rapport Impair/Pair	Odd/Even Ratio
IRRKRI	Irrégularité Spectrale (Krimphoff)	Spectral Irregularity (Krimphoff)
IRRKEN	Irrégularité Spectrale (Kendall)	Spectral Irregularity (Kendall)
IRRJEN	Irrégularité Spectrale (Jensen)	Spectral Irregularity (Jensen)
TR1	Tristimulus (1 ^{er} coefficient)	Tristimulus (1 st coefficient)
TR2	Tristimulus (2 ^{eme} coefficient)	Tristimulus (2 nd coefficient)
TR3	Tristimulus (3 ^{eme} coefficient)	Tristimulus (3 rd coefficient)
INH	Inharmonicité	Inharmonicity

TAB. 4.2 – Nomenclature et noms des descripteurs de timbre utilisés (cf. également section 2.2.5).

Le calcul de certains descripteurs a été effectué suivant deux échelles d'amplitude, l'une linéaire et l'autre de puissance (cf. section 2.2.5). Ainsi, nous pourrions déterminer pour chaque descripteur l'échelle qui offre la meilleure adéquation avec les données perceptives. Les seuils de début et de fin d'attaque, utilisés pour le calcul du temps d'attaque, sont respectivement $seuil_{deba} = 10\%$ et $seuil_{fina} = 90\%$ de la valeur maximale de l'enveloppe RMS. Les seuils de début et de fin de la phase de relâchement, utilisés pour le calcul du temps de relâchement sont respectivement $seuil_{debr} = 50\%$ et $seuil_{finr} = 10\%$ de la valeur maximale de l'enveloppe RMS.

En fonction des valeurs de γ_m et de ζ les seuils de démarrage et d'extinction des auto-oscillations peuvent varier. Il s'ensuit de très fines différences de durée entre les sons. Un descripteur de la durée des sons DUR a donc été ajouté afin d'étudier ce critère au même titre que les critères de timbre. Nous avons préféré ne pas ré-ajuster la durée des sons d'une part car les différences sont très subtiles, et d'autre part, car cet aspect du modèle physique traduit une réponse naturelle de l'instrument à son contrôle. Du reste, à l'issue de l'expérience, aucun participant n'a reporté s'être reposé sur la durée des sons.

Les descripteurs spectraux ont été calculés à partir de la Transformée de Fourier à fenêtre glissante. Les fenêtres utilisées sont des fenêtres de *Hann* de 1024 points (environ 20 ms à 44.1 kHz). Un recouvrement de 50% de la taille de la fenêtre a été employé (512 points). Le nombre de points de calcul de la Transformée de Fourier Discrète est de 8192. La résolution fréquentielle est donc d'environ 43 Hz (fréquence inférieure aux notes de plus basses fréquences analysées). La discrétisation du spectre s'opère à une résolution d'environ 5 Hz.

Les descripteurs harmoniques ont été déterminés à partir des amplitudes et fréquences instantanés des composantes harmoniques, déterminés selon la procédure d'analyse-synthèse additive SAS, décrite à la section 2.2.5.5.

Les descripteurs spectraux définis à court-terme (issus de la TFDCT) et les descripteurs harmoniques instantanés ont par la suite été moyennés entre le début de la phase attaque et la fin de la phase de relâchement en vue d'obtenir une mesure globale caractéristique des sons.

4.3 Résultats

4.3.1 Homogénéité des réponses des participants

La figure 4.5 présente les résultats de la classification ascendante hiérarchique des réponses des participants. Cette dernière ne révèle pas de différences systématiques entre les musiciens et les non musiciens. Trois groupes principaux ressortent de l'analyse, mais les distances entre les groupes restent cependant modérées. Les données de tous les participants ont donc été conservées par la suite.

4.3.2 Espace de timbre perceptif

4.3.2.1 Analyse multidimensionnelle de proximité (MDS)

Une analyse multidimensionnelle de proximité de type non-métrique (MDSCAL) a été effectuée sur la matrice de dissimilarité globale. La figure 4.6 montre l'évolution du stress² en fonction du nombre de dimensions de la configuration MDS. Afin de déterminer un nombre de dimensions satisfaisant, nous avons utilisé plusieurs critères de manière combinée (cf. section 2.2.4.1.4) : l'analyse de l'évolution du stress en fonction du nombre de dimensions (recherche du "coude"), l'analyse du diagramme de Shepard (cf. section 2.2.4.1 et la référence [Dillon et Goldstein, 1984, p. 145]), une règle empirique qui consiste à dire qu'une valeur de stress inférieure à 0.1 offre une bonne représentation des données [Krzanowski, 2000] (la valeur seuil de 0.1 donnée par Kendall provient de constatations *a posteriori* réalisées à partir d'analyses de cas réels), et l'interprétation de la configuration obtenue. La courbe d'évolution du stress (figure 4.6) présente un point d'inflexion pour 3 dimensions. Le "coude" semble être atteint pour 6 dimensions. Cependant, nous avons choisi une configuration MDS à 3 dimensions pour plusieurs raisons. D'une part, la valeur

²Le stress utilisé est le stress de Kruskal (cf. section 2.2.4.1).

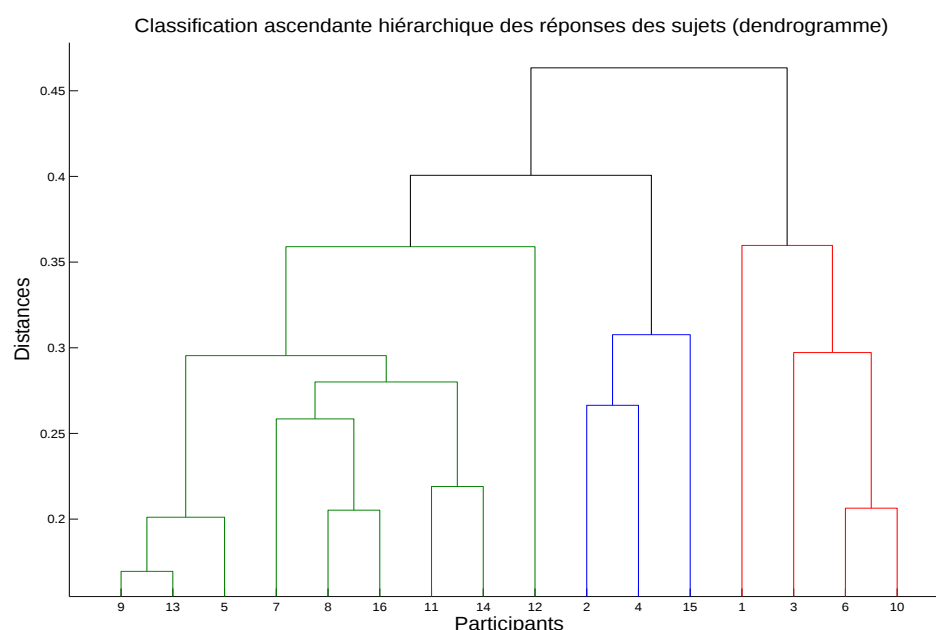


FIG. 4.5 – Représentation en dendrogramme de la classification ascendante hiérarchique des distances entre les réponses des participants. La technique de regroupement utilisée est le lien complet (voir par ex. [Saporta, 1990]).

de stress obtenue pour 3 dimensions est raisonnablement “bonne” ($stress \approx 0.05$, valeur inférieure au seuil de 0.1, qui reste certes arbitraire). D’autre part, d’après le diagramme de Shepard présenté figure 4.7, la fonction qui lie les distances inter-stimuli, dérivées de la configuration MDS, et les disparités (obtenues par transformation monotone des dissimilarités des participants) est bien une fonction monotone croissante. Ceci indique que les relations d’ordre entre les dissimilarités sont globalement bien conservées. Notons toutefois qu’il existe une légère distorsion des données (autrement, dissimilarités et distances seraient parfaitement alignées le long de la première bissectrice). Enfin, l’interprétation des dimensions élevées (supérieures à 3) dans les espaces MDS obtenus pour 4, 5 et 6 dimensions est difficile. L’écoute des différents stimuli répartis le long de ces dimensions dans les configurations spatiales associées ne permet pas de dégager d’évolutions systématiques de manière claire. Aucun des paramètres mécaniques (contrôle), ni des descripteurs de timbre utilisés dans les prochaines sections, ne s’est avéré être corrélé de manière significative avec les dimensions d’ordre supérieures à 3 dans les configurations à 4, 5 et 6 dimensions³.

4.3.2.2 Espace de timbre perceptif

Les projections de la configuration MDS sur ses première et deuxième dimensions, ainsi que sur ses première et troisième dimensions sont représentées sur la figure 4.8. Afin d’évaluer l’influence du contrôle sur la position des stimuli dans l’espace perceptif (cf. section 4.3.3), les stimuli ont été reliés par γ_m identiques (figures 4.8(a) et 4.8(b)), et par ζ identiques (figures 4.8(c) et 4.8(d)).

Analysons maintenant la structure de cet espace en fonction des paramètres mécaniques (paramètres de contrôle) et psychoacoustiques (descripteurs de timbre) associés aux sons.

³Notons toutefois qu’il se pourrait que d’autres types de descripteurs soient corrélés avec ces dimensions, mais ceci n’ôterait pas le fait que les relations entre les sons le long de ces dimensions ne soient pas claires au niveau perceptif.

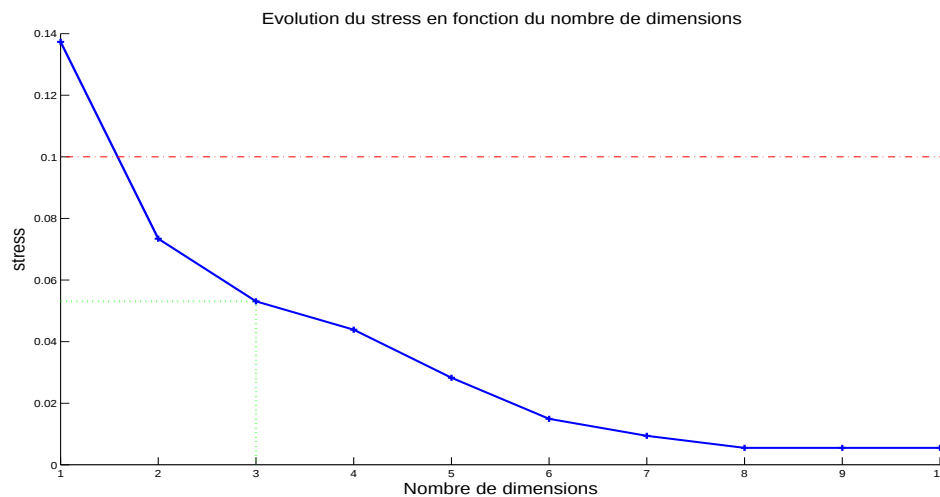


FIG. 4.6 – Evolution du stress (de Kruskal) en fonction du nombre de dimensions de la configuration MDS. Une configuration à 3 dimensions est retenue.

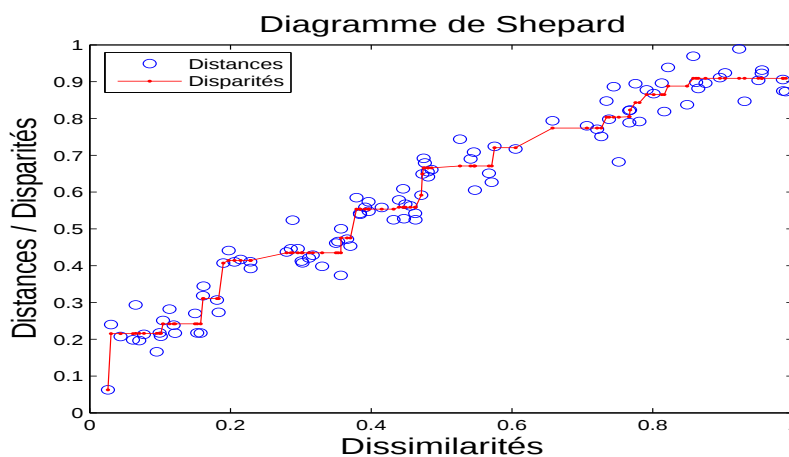
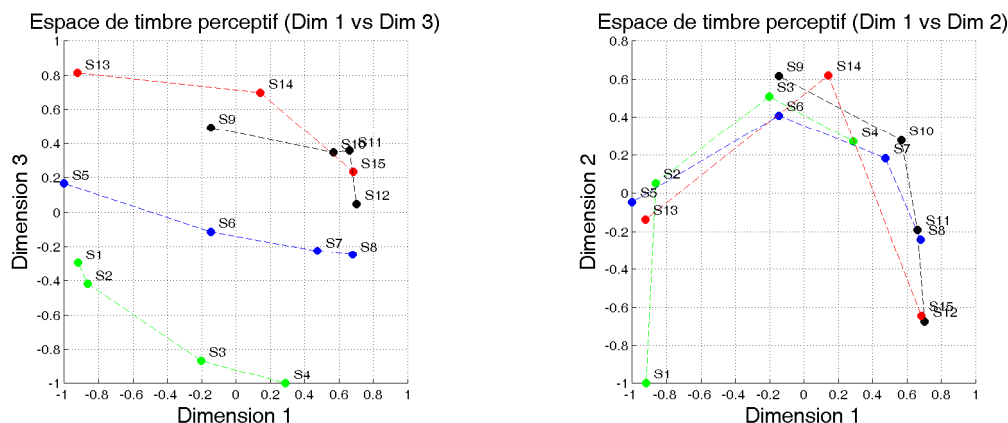


FIG. 4.7 – Diagramme de Shepard associé à l'analyse multidimensionnelle de proximité non-métrique. Ce diagramme présente la relation entre les distances inter-stimuli issues de la configuration MDS et les dissimilarités dérivées des réponses des participants. La figure indique également les disparités, obtenues par transformation monotone croissante des dissimilarités dans les procédures d'analyse multidimensionnelle de type non-métrique.

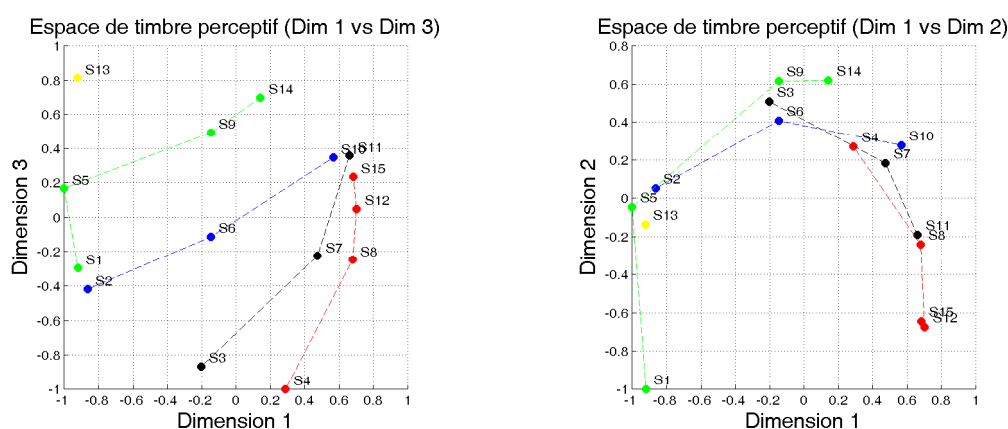
4.3.3 Corrélatifs mécaniques de l'espace de timbre perceptif

Il est intéressant de constater une certaine similarité de l'évolution de la position des stimuli dans l'espace perceptif lorsque l'un des paramètres de contrôle est fixé et que l'autre varie. Les observations suivantes peuvent être faites pour les première et troisième dimensions :

- Quelque soit γ_m , lorsque ζ augmente, la position des stimuli se décale vers la droite le long de la première dimension, et tendent à décroître le long de la troisième dimension (figure 4.8(a)). Plus γ_m est grand, plus les stimuli se situent vers le haut le long de la troisième dimension.
- Quelque soit ζ , lorsque γ_m augmente, la position des stimuli se décale vers la droite le long de la première dimension, et se décalent vers le haut le long de la troisième dimension (figure 4.8(c)).



(a) Dimension 1 vs Dimension 3 (lien : isovaleurs de γ_m) (b) Dimension 1 vs Dimension 2 (lien : isovaleurs de γ_m)



(c) Dimension 1 vs Dimension 3 (lien : isovaleurs de ζ) (d) Dimension 1 vs Dimension 2 (lien : isovaleurs de ζ)

FIG. 4.8 – Projections de la configuration MDS sur ses différentes dimensions. Sur les figures (a) et (b), les sons de même γ_m sont reliés. Sur les figures (c) et (d), les sons de même ζ sont reliés.

On notera toutefois que le décalage vers la droite le long de la première dimension s'estompe pour les fortes valeurs de γ_m et ζ (S11, S12, S15).

Afin d'interpréter de manière plus précise ces évolutions, les corrélations linéaires (r de Pearson) entre les valeurs des paramètres de contrôle γ_m et ζ utilisés pour générer les stimuli, et les coordonnées de ces derniers dans l'espace perceptif ont été calculées. Elles sont répertoriées dans le tableau 4.3.3.

Paramètres de contrôle	Dimension 1	Dimension 2	Dimension 3
γ_m	-	-	0.88***
ζ	0.76***	-	-0.54*

TAB. 4.3 – Coefficients de corrélations r de Pearson (d.d.l. = 13) entre les coordonnées des stimuli dans l'espace perceptif et les valeurs des paramètres de contrôle γ_m et ζ associés aux sons. Les valeurs qui ne sont pas significatives ne sont pas reportées (le signe - est utilisé le cas échéant). Les probabilités p que les mesures soient indépendantes sont indiquées de la manière suivante : * $p < .05$, ** $p < .01$, *** $p < .001$.

Il est intéressant de constater que le paramètre ζ , lié à l'ouverture du canal d'anche à l'équilibre, est corrélé avec la première dimension de l'espace perceptif [$r(13) = 0.76$, $p < 0.001$]. Cette

corrélation traduit le fait que lorsque ζ augmente, les stimuli se déplacent de gauche à droite le long de la première dimension, comme le montre la figure 4.8(a). Le paramètre γ_m , relié à la pression d'alimentation, n'est pas corrélé de manière significative avec la première dimension. En revanche, il est corrélé de manière fortement significative avec la troisième dimension de l'espace perceptif [$r(13)=0.88$, $p<0.001$]. Sur la figure 4.8(c), les stimuli se déplacent en effet vers le haut le long de la troisième dimension lorsque γ_m augmente. ζ n'est que faiblement corrélé avec la troisième dimension. Les relations entre les paramètres de contrôle et la deuxième dimension sont plus complexes. Aucun des deux paramètres de contrôle n'est corrélé de manière linéaire avec la deuxième dimension. Les figures 4.8(b) et 4.8(d) laissent à penser que les relations entre γ_m , ζ et les coordonnées des stimuli le long de la deuxième dimension sont non linéaires (une forme en cloche semble apparaître de manière systématique lorsque un des paramètres est fixé et que l'autre varie). Il serait intéressant d'explorer, dans les perspectives de ce travail, les corrélations de cette dimension avec des combinaisons linéaires ou non-linéaires des paramètres de contrôle.

Le dendrogramme associé à l'analyse en cluster de la matrice de dissimilarité globale entre les stimuli est représenté figure 4.9. L'analyse révèle trois sous-catégories principales. La première (située à droite sur la figure) est constituée par les stimuli correspondants à des valeurs de γ et ζ faibles, ou à une valeur de γ très élevée et une valeur de ζ très faible (S1, S2, S5, S13). La seconde (située au milieu sur la figure) est constituée par les stimuli ayant été générés pour des couples (γ, ζ) de valeurs moyennes (S3, S6, S9, S10, S14). La troisième (située à gauche sur la figure) rassemble les stimuli associés à une valeur de γ faible et une valeur de ζ élevée, ou des valeurs de γ et ζ toutes les deux élevées (S4, S7, S8, S11, S12, S15).

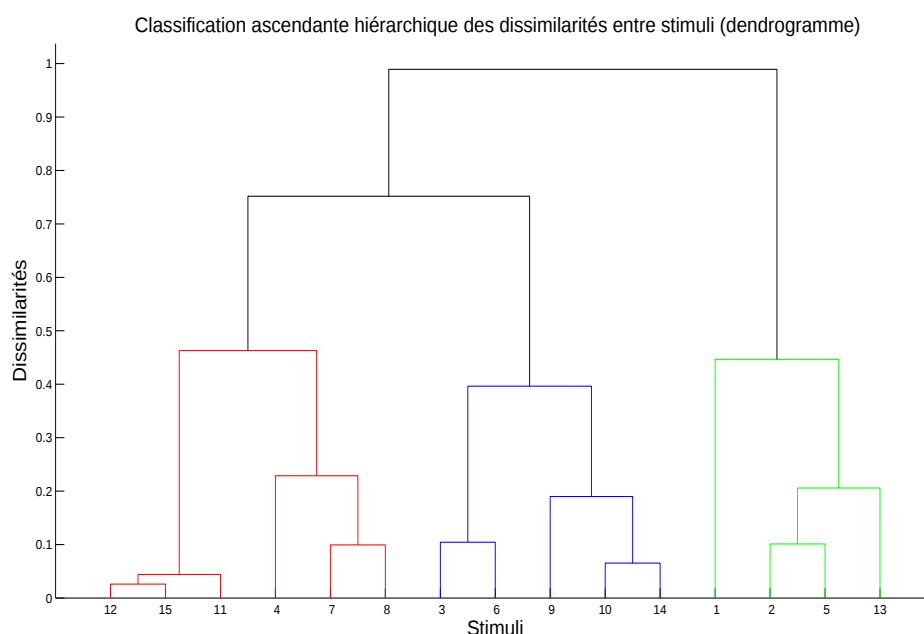


FIG. 4.9 – Représentation en dendrogramme de la classification ascendante hiérarchique effectuée à partir de la matrice de dissimilarité globale des stimuli. La technique de regroupement utilisée est le lien complet.

4.3.4 Corrélats acoustiques de l'espace de timbre perceptif

Le tableau 4.3.4 présente les corrélations entre les coordonnées des sons dans l'espace perceptif et les descripteurs de timbre employés (cf. section 4.2.7). Les probabilités (p) que les coordonnées perceptives et les valeurs des descripteurs soient indépendantes sont également indiquées. Ceci permettra de mesurer le degré de significativité des correspondances établies entre mesures perceptives et objectives. Lorsque les descripteurs de timbre ont à la fois été calculés suivant une échelle d'amplitude linéaire et une échelle de puissance, les deux coefficients de corrélations correspondants sont reportés dans le tableau.

La figure F.1 de l'annexe F.1 présente les intercorrélations entre les principaux descripteurs de timbre (AT, TC, SC, SS, SFAT, OER, IRRJEN, TR1, TR2, TR3) et les paramètres de contrôle (γ_m et ζ), qui permettent d'expliquer la répartition des sons dans l'espace de timbre perceptif.

Comme cela était prévisible pour les sons de clarinette, qui sont de type harmonique, le descripteur d'inharmonicité s'est avéré être toujours inférieur à 10^{-5} (valeur adimensionnée), i.e. environ 2×10^{-3} Hz, rapporté à la fréquence fondamentale des sons. Cette valeur est bien inférieure à notre seuil de discrimination fréquentiel (environ 1 Hz en basses fréquences [Wier *et al.*, 1977]). Nous avons donc exclu ce descripteur de l'analyse des corrélations.

4.3.4.1 Corrélats acoustiques de la première dimension

La première dimension de l'espace de timbre perceptif est corrélée de manière fortement significative avec le **Logarithme du Temps d'Attaque LAT**, et avec des descripteurs spectraux, tels que les différents descripteurs de **Centre de Gravité Spectral (SC, HSC, OSC)**, et le troisième coefficient de **Tristimulus (TR3)**. Ce dernier rend compte de l'énergie relative des composantes harmoniques de rangs élevés. Nous allons voir pourquoi dans le cadre des sons analysés tous ces descripteurs sont covariants. Notons que ces descripteurs ne sont pas du tout corrélés aux autres dimensions de la configuration MDS, ce qui est intéressant vis-à-vis de l'interprétation de la structure de l'espace perceptif.

Remarquons également les fortes corrélations de la première dimension avec le **Flux Spectral SFAT** calculé sur la phase d'attaque (le Flux Spectral global ne lui est que faiblement corrélé), ainsi qu'avec l'Asymétrie Spectrale, et le Roulement Spectral.

L'emploi d'une échelle d'amplitude de puissance augmente la valeur de corrélation du Centre de Gravité Spectral SC ($[r(13)=0.90, p<0.001]$ pour une échelle linéaire, et $[r(13)=0.94, p<0.001]$ pour une échelle de puissance). Le même phénomène se produit pour les Centres de Gravité Spectraux "harmonique" HSC et "impair" OSC ($[r(13)=0.91, p<0.001]$ pour une échelle linéaire, et $[r(13)=0.94, p<0.001]$ pour une échelle de puissance). Pour ce type de sons, il paraît donc intéressant d'utiliser une échelle de puissance, qui accorde un poids plus important aux composantes harmoniques dominantes (harmoniques de rangs faibles).

La figure 4.10, illustre la forte corrélation entre la répartition des stimuli le long de la première dimension et le Centre de Gravité Spectral, calculé selon une échelle de puissance. Ces résultats mettent en avant le fait que les auditeurs semblent être particulièrement sensibles aux différences de brillance entre les sons.

D'après les figures 4.8(a) et 4.8(c), la première dimension de l'espace de timbre perceptif sépare des sons correspondants à des couples (γ_m, ζ) de faibles valeurs, ou ayant une valeur de γ_m élevée et une valeur de ζ très faible (S1, S2, S5, S13), de sons correspondants à des couples (γ_m, ζ) de valeurs élevées (S8, S11, S12, S15). Les sons situés à l'extrême gauche le long de cette

Descripteurs	Dimension 1		Dimension 2		Dimension 3	
Echelle	linéaire puissance		linéaire puissance		linéaire puissance	
AT	-0.93***		-		-	
LAT	-0.95***		-		-	
RT	-		-		-0.78***	
TC	-		-		0.80***	
SC	0.90***	0.94***	-	-	-	-
SS	0.63*	0.88*	-0.63*	-	-	-
SSK	-0.93***	-0.72**	-	-0.54*	-	-
SKU	-0.88***	-0.77***	-	-	-	-
SRO	0.92***		-	-	-	-
SF	-	-0.58*	-	-	-	-
SFAT	-0.95***		-	-	-	-
HSC	0.91***	0.94***	-	-	-	-
OSC	0.91***	0.94***	-	-	-	-
ESC	0.90***	0.81***	-	-	-	-
OER	-	0.84***	-	-	0.87***	-
IRRKRI	-0.79***		-		-	
IRRKEN	-		0.53*		0.69**	
IRRJEN	0.79***		-		-	
TR1	-0.92***	-0.88***	-	-	-	-
TR2	-0.71**	0.62*	0.67**	0.62*	-	-
TR3	0.95***	0.95***	-	-	-	-

TAB. 4.4 – Coefficients de corrélations r de Pearson (d.d.l. = 13) entre les coordonnées des stimuli dans l'espace perceptif et les valeurs des descripteurs de timbre employés. Lorsque deux échelles d'amplitudes différentes (linéaire et puissance) ont été employées pour calculer les descripteurs, les deux valeurs de corrélations correspondantes sont reportées. Dans le cas contraire, la valeur de corrélation associée au descripteur est disposée au centre de la colonne. Les valeurs qui ne sont pas significatives ne sont pas reportées (le signe - est utilisé le cas échéant). Les probabilités p que les mesures soient indépendantes sont indiquées de la manière suivante : * $p < .05$, ** $p < .01$, *** $p < .001$. Pour chaque dimension, les deux plus fortes mesures de corrélations sont signalées en gras.

dimension disposent d'une attaque longue et sont plutôt sourds (cf. [Exemple Sonores \$n^{os}\$ 36 et 37](#) qui correspondent respectivement aux sons S2 et S13). A l'inverse, les sons situés à l'extrême droite de cette dimension disposent d'une attaque brève et sont très brillants (cf. [Exemples Sonores \$n^{os}\$ 38 et 39](#) qui correspondent respectivement aux sons S8 et S15). Ceci est en adéquation avec le fait que cette dimension est fortement corrélée avec le Temps d'Attaque, et des descripteurs rendant compte de la richesse spectrale (Centre de Gravité Spectral, Tristimulus TR3, Asymétrie et Roulement Spectral). En effet, compte-tenu de la façon dont les sons ont été produits, la durée de la phase d'attaque et la richesse spectrale sont des facteurs covariants. De manière générale, lorsque l'ouverture du canal d'anche suffisamment grande, plus la pression est élevée, plus les auto-oscillations démarreront rapidement (attaque brève), et le son sera riche en harmoniques de rangs élevés (brillant).

Les figures [4.11\(a\)](#) et [4.11\(b\)](#) présentent les formes d'ondes des signaux S2 (γ_m et ζ faibles) et S8 (γ_m et ζ élevés), ainsi que certains descripteurs temporels (positions temporelles du début et de la fin de l'attaque, du début et de la fin de la phase de relâchement, ainsi que Centre de Gravité Temporel). Les figures [4.12\(a\)](#) et [4.12\(b\)](#) montrent les spectrogrammes correspondants.

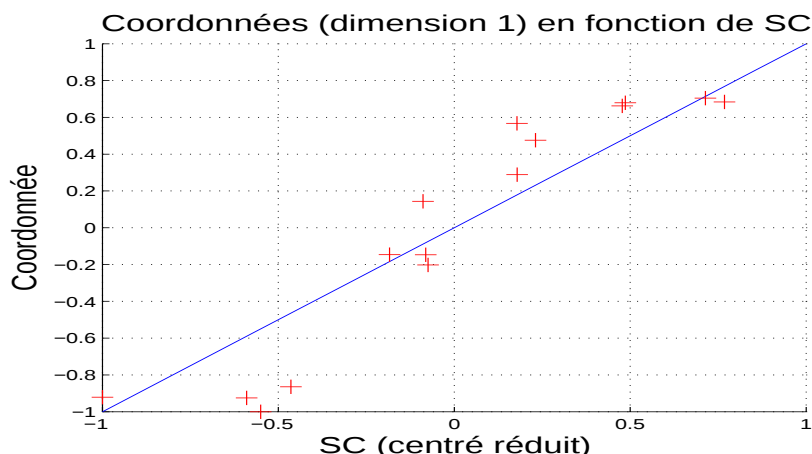


FIG. 4.10 – Coordonnées des stimuli le long de la première dimension en fonction du Centre de Gravité Spectral (centré réduit).

La comparaison des formes d'ondes des signaux révèlent une légère différence de durée entre les deux sons (environ 0.2 s). Le descripteur de durée DUR présente une corrélation relativement élevée [$r(13)=0.78$, $p<0.001$] avec la première dimension. Lorsque la pression et l'ouverture du canal d'anche sont tous deux élevés, la partie soutenue des sons est plus longue, et la phase de relâchement arrive légèrement plus tard. Lors de l'expérience ce critère a donc potentiellement pu être un facteur de discrimination des sons. Cependant, aucun participant n'a mentionné avoir utilisé la durée des sons comme critère de dissimilarité. De fait, la corrélation du descripteur de durée avec la première dimension reste bien plus faible que celles présentées par le Logarithme du Temps d'Attaque LAT [$r(13)=-0.95$, $p<.001$] ou le Centre de Gravité Spectral SC [$r(13)=0.94$, $p<.001$].

Comme on peut le constater sur les spectrogrammes 4.12(a) et 4.12(b), les phases d'amorces des différentes composantes ne s'opèrent pas de manière synchrone. Les quatre premiers harmoniques impairs apparaissent en premier (1, 3, 5, 7), puis leurs succèdent les harmoniques pairs (2, 4, 6). Les composantes harmoniques démarrent beaucoup plus rapidement lorsque la pression est élevée et la pince faible (i.e. ζ élevé, cf. la figure 4.12(b), associée au son S8) que lorsque la pression est faible et la pince forte (i.e. ζ faible, cf. la figure 4.12(a), associée au son S2). Ceci explique une valeur du Flux Spectral, calculé le long de la phase d'attaque, plus faible lorsque l'attaque est brève et riche en harmoniques (tel que c'est le cas pour le son S8, pour lequel SFAT = 0.76) que lorsque l'attaque est lente et pauvre en harmoniques (tel que c'est le cas pour le son S2, pour lequel SFAT = 0.88)⁴. Il n'est donc pas surprenant que ce descripteur soit également fortement corrélé avec les positions des stimuli le long de la première dimension (cf. tableau 4.3.4).

Les figures 4.13(a) et 4.13(b) permettent de comparer les valeurs de plusieurs descripteurs spectraux pour les sons S2 (γ_m et ζ faibles), et S8 (γ_m et ζ élevés). Le Centre de Gravité Spectral est beaucoup plus élevé pour le son S8 que pour le son S2 (cf. figure 4.13(a)). Le spectre est en effet globalement plus riche (cf. Roulement Spectral, figure 4.13(b)). De même, l'Étalement Spectral indique que la dispersion de la distribution spectrale autour de sa moyenne (le Centre de Gravité Spectral) est plus grande pour S8 que pour S2. Pour les deux sons, l'Asymétrie Spectrale est positive, ce qui indique que l'énergie de la distribution spectrale est concentrée vers les fréquences basses, par rapport à sa moyenne. En revanche, cet aspect est moins prononcé pour le son

⁴Rappelons que plus la valeur du Flux Spectral est élevée, plus la corrélation entre spectres successifs est grande.

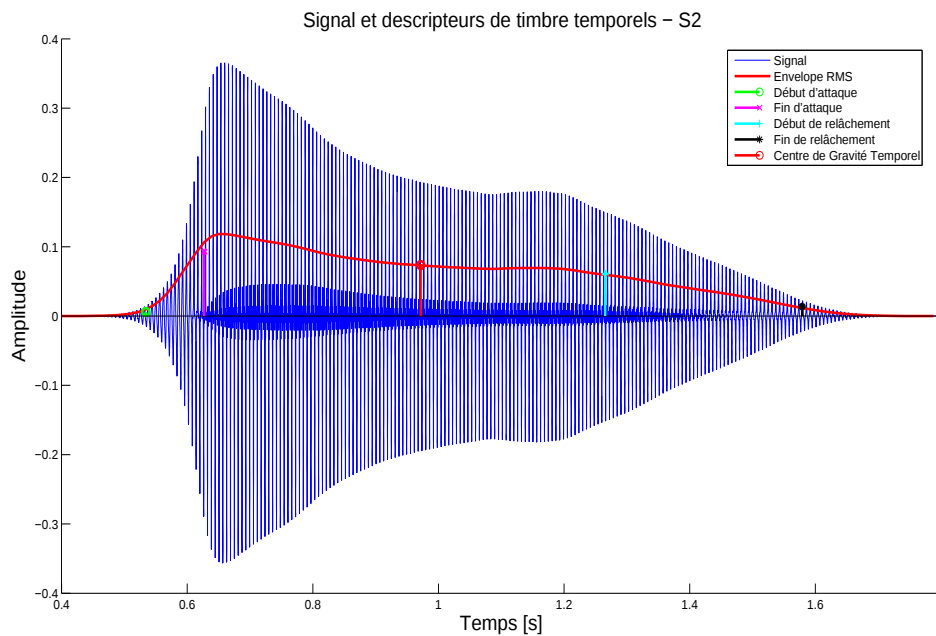
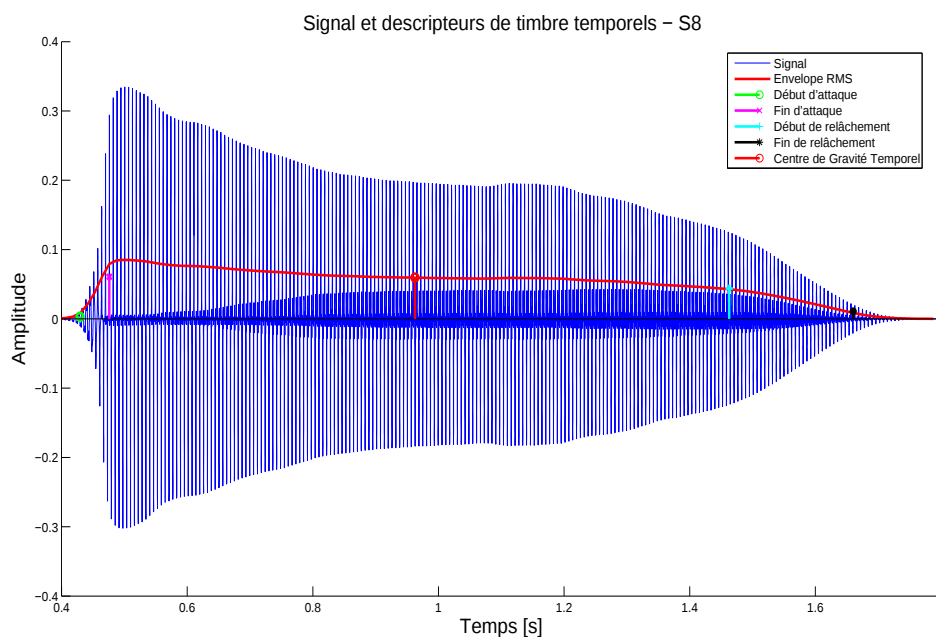
(a) Signal S2 (γ_m et ζ faibles)(b) Signal S8 (γ_m et ζ élevés)

FIG. 4.11 – (a) Forme d'onde du stimulus S2 (obtenu pour des valeurs de γ_m et ζ faibles), et (b) forme d'onde du stimulus S8 (obtenu pour des valeurs de γ_m et ζ élevées).

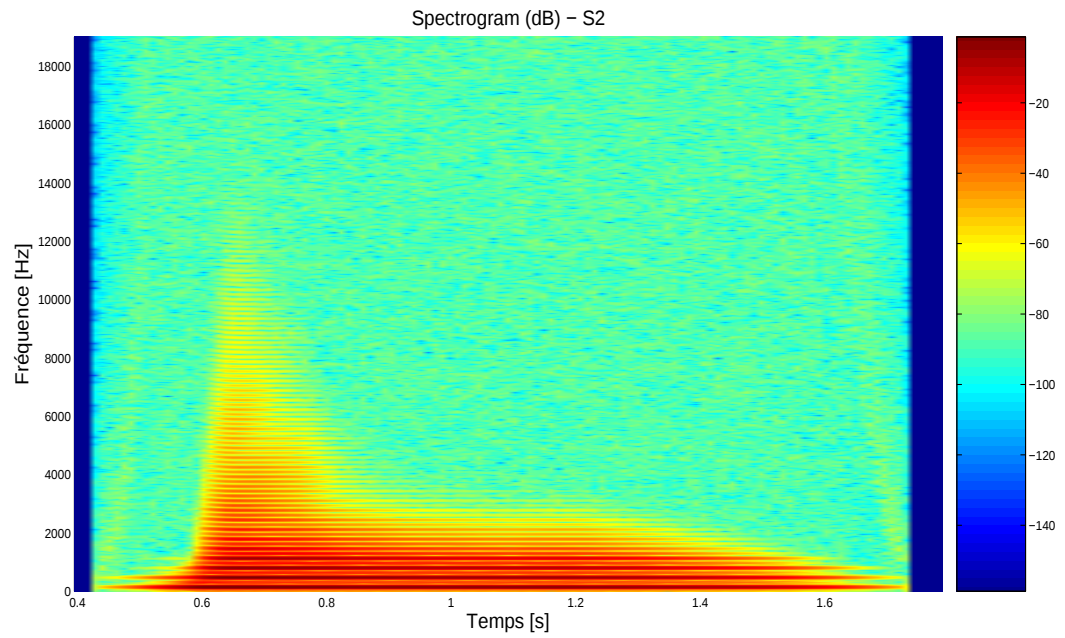
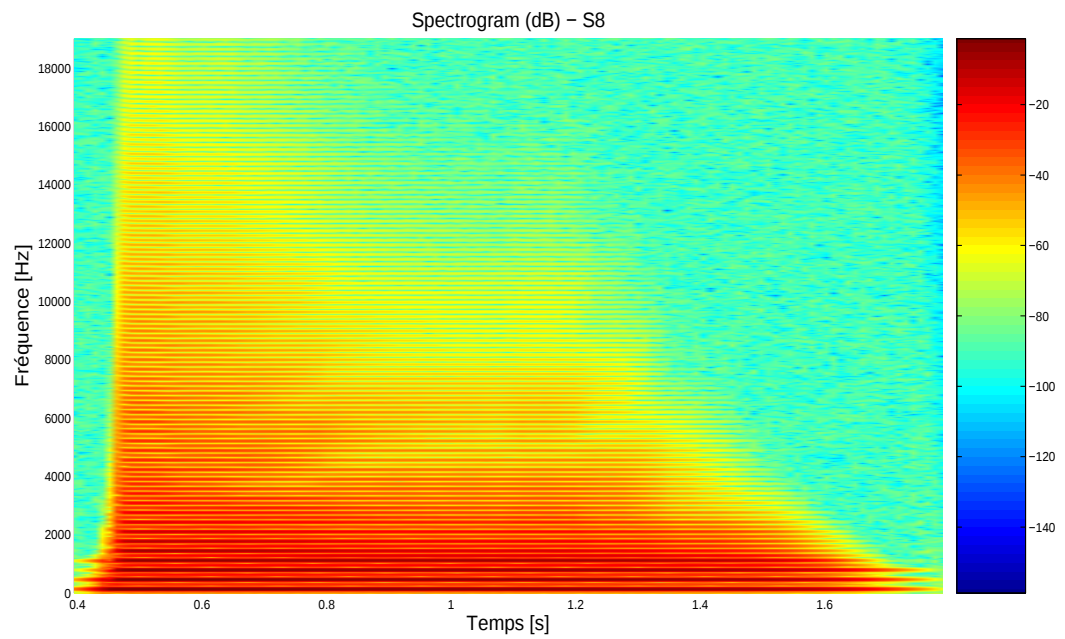
(a) Spectrogramme S2 (γ_m et ζ faibles)(b) Spectrogramme S8 (γ_m et ζ élevés)

FIG. 4.12 – (a) Spectrogramme du stimulus S2 (obtenu pour des valeurs de γ_m et ζ faibles), et (b) spectrogramme du stimulus S8 (obtenu pour des valeurs de γ_m et ζ élevées).

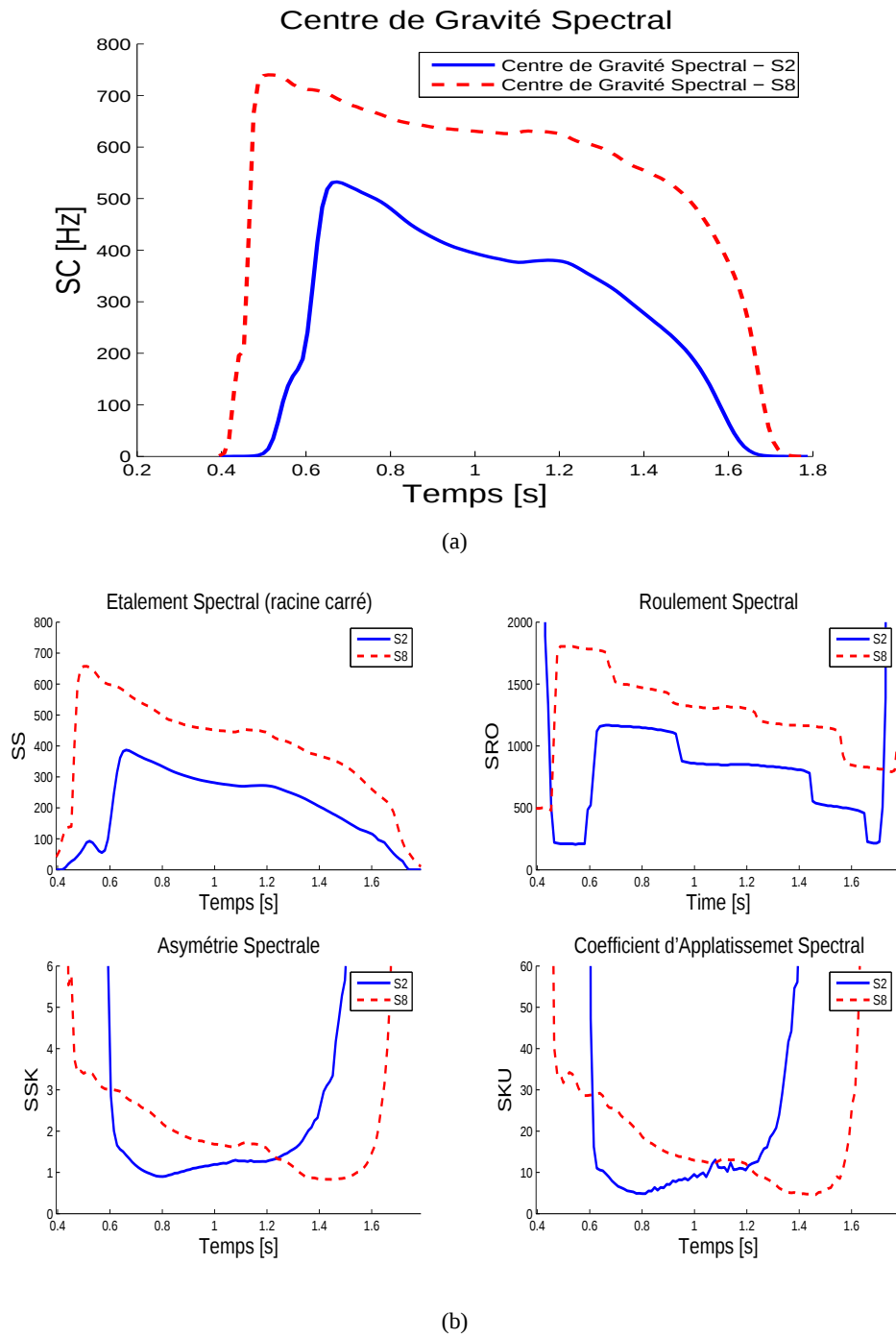


FIG. 4.13 – (a) Centres de Gravité Spectraux des sons S2 (obtenu pour des valeurs de γ_m et ζ faibles) et S8 (obtenu pour des valeurs de γ_m et ζ élevées) - (b) Etallement Spectral (SS), Roulement Spectral (SRO), Asymétrie Spectrale (SSK) et Coefficient d'Aplatissement Spectral (SKU) calculés pour les sons S2 et S8. Une échelle de puissance a été utilisée. Notons que lorsque le bruit prédomine (au début et à la fin du son), les valeurs numériques de certains descripteurs peuvent devenir très élevées, et ce malgré le terme stabilisateur b_0 (cf. section 2.2.5).

S8 dont la quantité élevée d'harmoniques aigus induit un étirement de la forme de la distribution spectrale vers les fréquences élevées. La forte dominance des harmoniques de rangs faibles dans le cas du son S8 explique aussi pourquoi la distribution spectrale est globalement plus sélective que pour le son S2 (cf. Coefficient d'Aplatissement Spectral, figure 4.13(b)).

La figure 4.3.4.1 présente les valeurs de Centre de Gravité Spectral des 92 sons correspondants à des conditions normales de jeu (cf. figure 4.4 pour la valeur des paramètres de contrôle). **Il est intéressant de remarquer que lorsque γ_m et ζ augmentent, le Centre de Gravité Spectral des sons simulés par le modèle augmente de manière monotone et croissante.** Ceci montre que la clarinette permet d'obtenir des sons avec une large variété de brillance. D'autre part, des sons de Centre de Gravité Spectraux identiques peuvent être obtenus pour des couples de pression et pince différents.

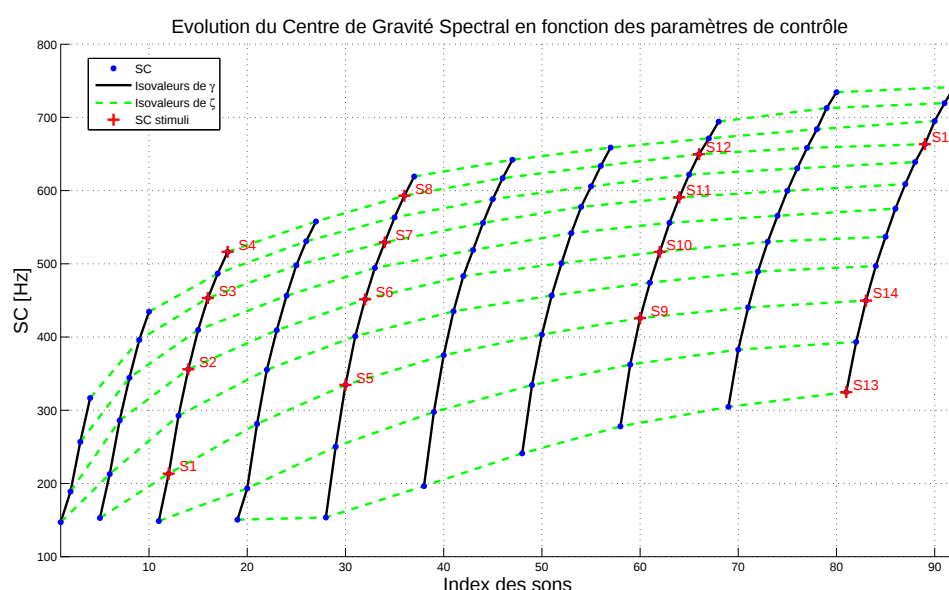


FIG. 4.14 – Evolution du Centre de Gravité Spectral (échelle de puissance) lorsque les paramètres de contrôle γ et ζ augmentent. Les courbes en trait plein indiquent les isovaleurs de γ (γ augmente de gauche à droite). Les courbes en pointillés indiquent les isovaleurs de ζ (ζ augmente de bas en haut).

4.3.4.2 Corrélats acoustiques de la seconde dimension

Les deux descripteurs les mieux corrélés avec la deuxième dimension de l'espace perceptif sont le **deuxième coefficient de tristimulus TR2** linéaire [$r(13)=0.67$, $p<0.01$], et l'**Étalement Spectral SS** linéaire [$r(13)=-0.63$, $p<0.05$]. Il est intéressant de remarquer, d'après la figure 4.8(b), qu'à quelques exceptions près, la position des stimuli le long de la deuxième dimension croît, atteint un maximum, puis décroît au fur et à mesure que ζ augmente, et ce quelque soit γ_m . Indépendamment de γ_m , à mesure que ζ augmente, l'énergie est donc distribuée des basses fréquences (TR1 est élevé lorsque ζ est faible, car le descripteur est fortement anticorrélé avec la dimension 1, cf. tableau 4.3.4, et figure 4.8(a)), aux fréquences moyennes (TR2, qui rend compte de l'énergie relative des harmoniques 2 à 4, atteint un maximum), puis aux hautes fréquences (TR3 est élevé lorsque ζ est élevé, car le descripteur est fortement corrélé à la dimension 1).

Cependant, les corrélations de TR2 et SS sont relativement faibles. De plus, ces descripteurs sont également corrélés de manière significative (et parfois de manière plus importante) avec la première dimension. Ces descripteurs ne s'avèrent donc pas idéaux pour trouver un espace de timbre objectif dont les dimensions seraient orthogonales. Le descripteur TR2 n'est en effet peut être pas le plus adapté pour les sons de clarinette, dont les harmoniques impairs et pairs peuvent présenter des différences de comportement, puisque son calcul repose sur la considération de deux harmoniques pairs (2 et 4), mais d'un seul harmonique impair (3). Nous allons voir dans la prochaine section que le rapport Impair/Pair s'avère être plus intéressant pour caractériser certaines spécificités du timbre de cet instrument.

4.3.4.3 Corrélats acoustiques de la troisième dimension

La troisième dimension s'avère être bien corrélée avec le **Rapport Impair/Pair OER**, calculé sur une échelle d'amplitude linéaire [$r(13)=0.87$, $p<.001$], et le **Centre de Gravité Temporel TC** [$r(13)=0.80$, $p<.001$], qui est le barycentre temporel de l'enveloppe RMS. La position des stimuli le long de la troisième dimension diminue à γ_m fixé lorsque ζ augmente (cf. figure 4.8(a)). De plus, lorsque γ_m augmente, la position des stimuli augmente le long de cette dimension. Comme le montrent les figures 4.15(a) et 4.15(b), le Rapport Impair/Pair et le Centre de Gravité Temporel présentent globalement une évolution de ce type en fonction des paramètres γ et ζ .

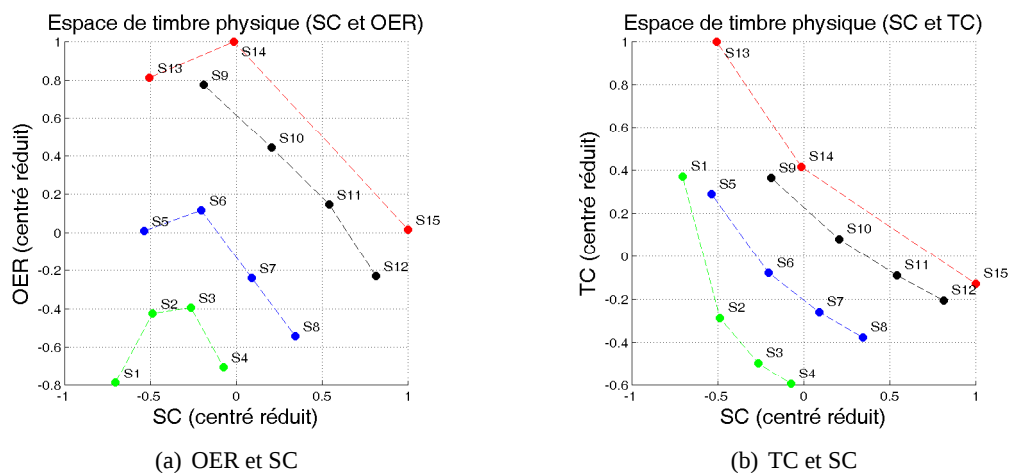


FIG. 4.15 – Espaces de timbre physique : (a) Rapport Pair/Impair (linéaire) en fonction du Centre de Gravité Spectral (puissance) - (b) Centre de Gravité Temporel en fonction du Centre de Gravité Spectral (puissance). Les sons de même γ_m sont reliés. Les valeurs des descripteurs sont centrées réduites.

Nous allons voir en quoi le Rapport Impair/Pair permet de mettre en évidence le net changement de timbre qui se produit lors de **la situation d'anche battante**, pendant laquelle l'anche vient taper la table du bec pendant une partie de son cycle d'oscillation [Barthet et al., 2005]. Ceci a pour effet de fermer quasi-totalement le canal d'anche pendant une partie du cycle d'oscillations. La figure 4.3.4.3 montre l'évolution temporelle de l'ouverture du canal d'anche adimensionnée $W(t)$ (cf. section 3.4.1.6) pour des sons générés à partir d'une valeur de γ_m fixe ($\gamma_m = 0.59$), et pour ζ variant de 0.12 à 0.27. Les sons n^{os} 2, 4, 6 et 8 correspondent aux stimuli S1, S2, S3, S4 de l'expérience. A partir d'une certaine valeur de ζ ($\zeta = 0.21$, son 5), l'ouverture du canal d'anche s'annule sur une portion de son cycle d'oscillation. Le phénomène s'accroît à mesure que ζ augmente. Lorsque cette situation se produit, le débit dans l'embouchure s'annule également. Ceci induit l'augmentation des harmoniques de rangs élevés, et notamment celle des

harmoniques pairs, qui sont alors favorisés. Il est intéressant de constater que ce phénomène apparaît lorsque l'on suit l'évolution du Rapport Impair/Pair linéaire (cf. 4.17(a)). En effet, ce dernier atteint sa valeur maximale pour le son $n^{\circ} 5$, pour lequel la situation d'anche battante commence à se produire. Le descripteur OER décroît ensuite aux valeurs de ζ plus élevées. Remarquons qu'une diminution de la valeur du Rapport Impair/Pair traduit une augmentation plus rapide de l'énergie des harmoniques pairs par rapport à celle des harmoniques impairs. C'est ce qui se produit lorsque l'anche commence à battre. Comme on peut le voir sur la figure 4.3.4.3, dans le cas de ces sons, lorsque le phénomène d'anche battante a lieu, il se déroule pendant un moment plus ou moins long, juste après l'attaque des sons. C'est ce qui explique la richesse spectrale de certains des stimuli de l'expérience pendant leur phase d'attaque, et également le fait que le Centre de Gravité Temporel tende à diminuer (i.e. se rapprocher de l'attaque) lorsque ζ augmente (cf. figure 4.15(b)).

Notons également que dans le cas des sons analysés, le Rapport Impair/Pair calculé selon une échelle de puissance ne laisse pas apparaître ce phénomène aussi clairement que lorsque l'échelle linéaire est employée (cf. figure 4.17(b)). Ce dernier tend à se stabiliser à partir d'une certaine valeur de ζ , mais ne diminue pas. En effet, l'échelle de puissance tend à accorder plus de poids aux harmoniques de rangs faibles qui eux restent dominants et sont peu affectés par le phénomène d'anche battante. Ceci explique que le Rapport Impair/Pair calculé selon les puissances soit bien corrélé avec la dimension 1 de l'espace perceptif [$r(13)=0.84$, $p<.001$]. Ce dernier évolue de manière monotone croissante en fonction de γ et ζ , telles les positions des stimuli correspondants le long de cette dimension.

Il est intéressant de souligner que cette cassure dans le comportement du Rapport Impair/Pair apparaît pour différentes combinaisons des valeurs des paramètres de contrôle du modèle, comme l'illustre la figure 4.3.4.3.

La pertinence des analyses précédentes a été vérifiée pour un son naturel de clarinette. Nous avons calculé le Rapport Impair/Pair du crescendo de clarinette dont le spectrogramme est représenté sur la figure 3.4 de la section 3.3.3 (le son correspond à l'Exemple Sonore $n^{\circ} 33$). Comme le montre la figure 4.19, le Rapport Impair/Pair atteint une valeur maximale à l'instant $t \approx 15.8s$). En effet, à partir de cette valeur l'énergie relative des harmoniques pairs (EA) augmente, tandis que celle des harmoniques impairs (OA) diminue. L'augmentation des harmoniques pairs se produit dans les fréquences élevées comme le montre le spectrogramme. A cet instant se produit un changement de timbre audible. Nous avons extrait la portion du son où se produit le phénomène (cf. Exemple Sonore $n^{\circ} 40$). L'augmentation rapide des harmoniques pairs de rangs élevés produit une sorte de grésillement très bref dans les aigus. Ce dernier s'estompe dès lors que les harmoniques pairs diminuent (à l'instant $t \approx 17s$), après que l'énergie relative des harmoniques pairs (EA) ait atteint sa valeur maximale. Nous attribuons ce grésillement à la situation d'anche battante. Le Rapport Impair/Pair semble très intéressant afin de prédire ce type de phénomène sur les sons de clarinette. Il apporte une information complémentaire au Centre de Gravité Spectral. En effet, le Centre de Gravité Spectral ne présente pas la phase décroissante que connaît le Rapport Impair/Pair, mais au contraire continue d'augmenter lorsque ce dernier diminue. Notons que pour le son naturel, ces observations sont valables indépendamment du type d'échelle utilisée pour le calcul des descripteurs (linéaire ou puissance).

4.3.5 Analyse des questionnaires

L'analyse des questionnaires indiquant les stratégies utilisées par les participants fait ressortir les trois critères principaux suivants : la **brillance** des sons, la nature de leur **attaque**, et la perception de **tension**. Le tableau 4.5 répertorie les différentes expressions employées par les participants en fonction de ces catégories.

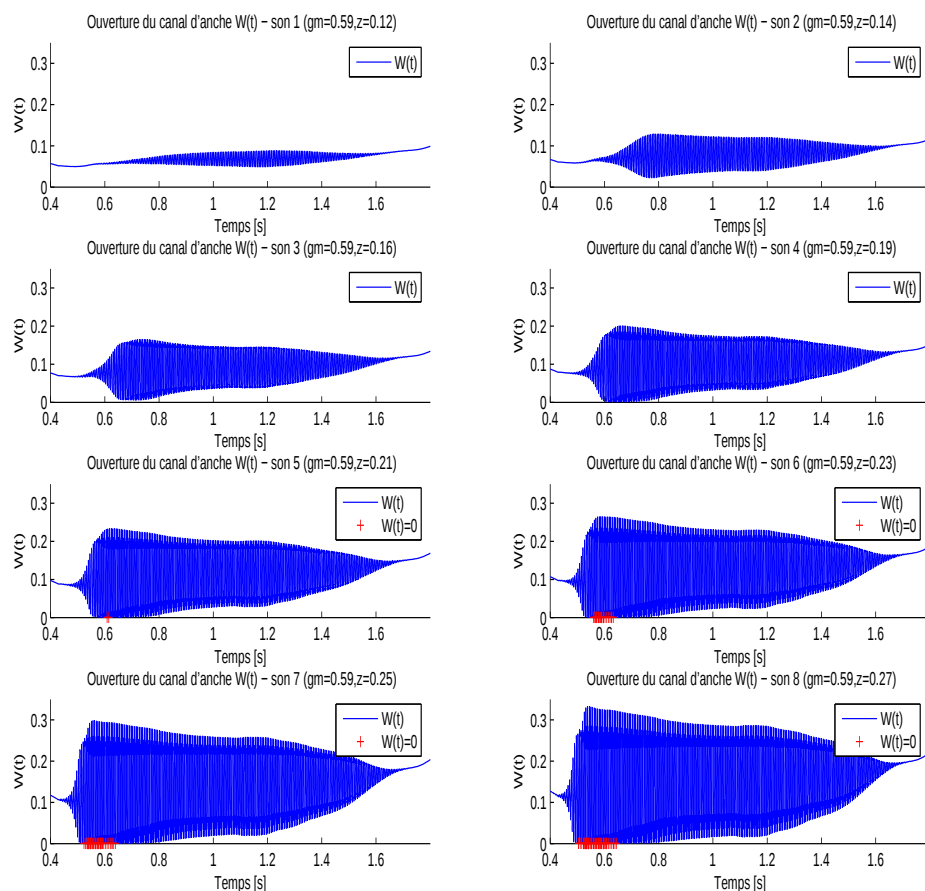


FIG. 4.16 – Evolution temporelle de l'ouverture du canal d'anche $W(t)$ pour différentes valeurs de ζ (noté z sur la figure), à γ_m fixé (noté gm sur la figure). Les instants pour lesquels l'ouverture s'annule sont indiqués par des croix.

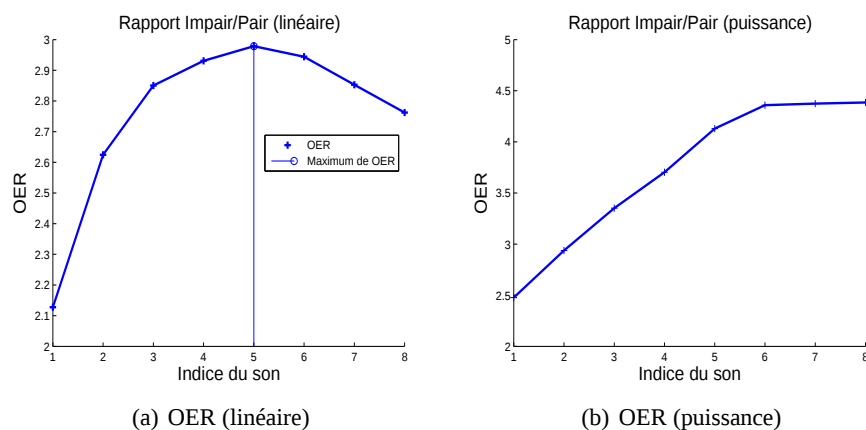


FIG. 4.17 – Rapport Impair/Pair des sons associés à la figure 4.3.4.3 (γ est fixé, et ζ augmente). La figure (a) présente la formulation linéaire, la figure (b) présente celle reposant sur la puissance.

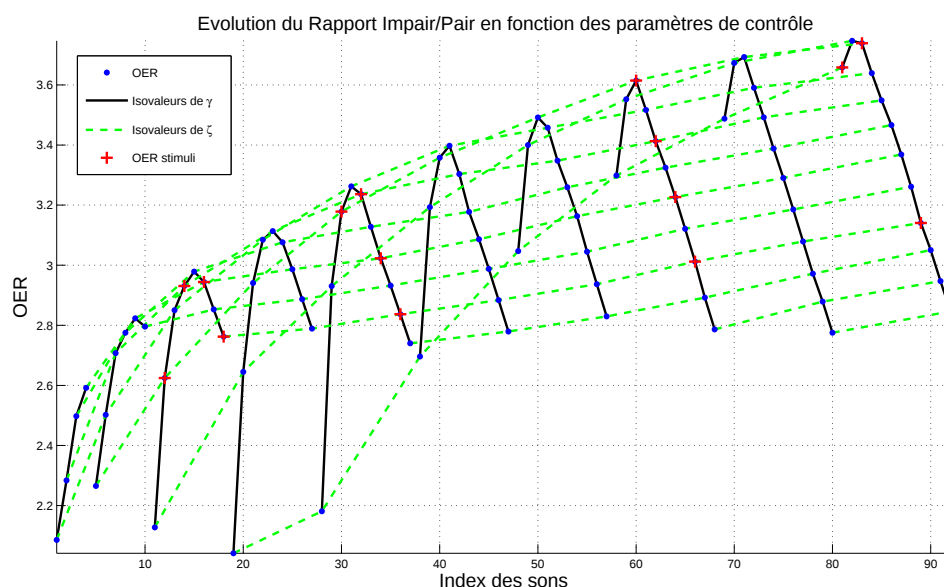


FIG. 4.18 – Evolution du Rapport Impair/Pair (linéaire) lorsque les paramètres de contrôle γ et ζ augmentent. Les courbes en trait plein indiquent les isovaleurs de γ (γ augmente de gauche à droite). Les courbes en traits pointillés indiquent les isovaleurs de ζ (ζ augmente de bas en haut).

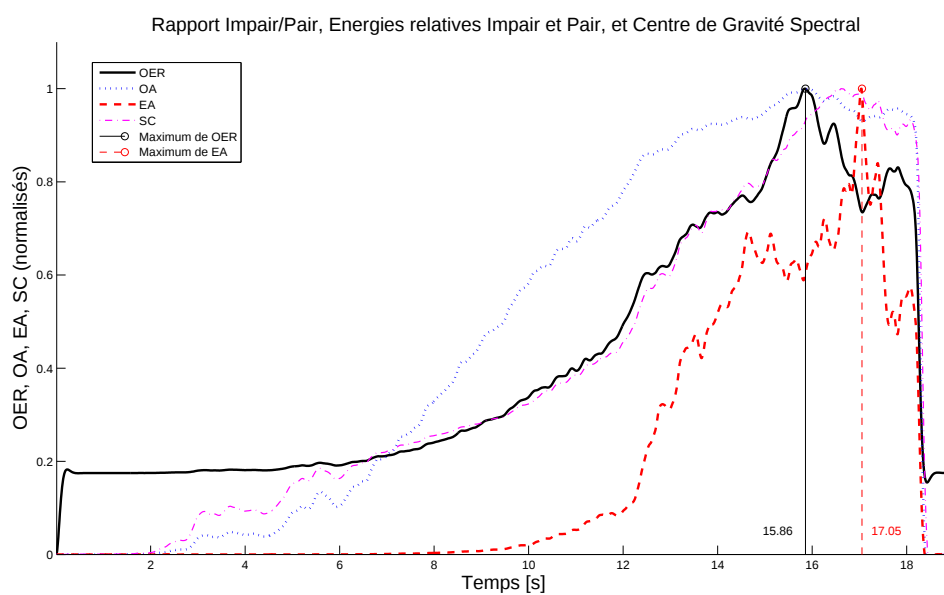


FIG. 4.19 – Analyse d'un son naturel de clarinette : évolutions temporelles du Rapport Impair/Pair (OER), des énergies relatives des harmoniques impairs (OA) et pairs (EA), et du Centre de Gravité Spectral (SC). Le son analysé correspond à l'Exemple Sonore n° 33. Les descripteurs ont été calculés sur une échelle de puissance (des résultats analogues sont obtenus d'après une échelle linéaire).

15 participants sur 16 ont employé des expressions faisant référence de manière directe (par ex. "brillant", "nasillard") ou indirecte ("riche", "puissance des harmoniques aigus") aux sensations

de *brillance* ou de *nasalité*. 10 participants sur 16 ont mentionné s'être reposé également sur la nature de l'attaque (par ex. durée, intensité). Il n'est donc pas étonnant que la première dimension de l'espace de timbre perceptif soit fortement corrélée avec des descripteurs décrivant la nature de l'attaque (le Temps d'Attaque logarithmique ou linéaire, qui en caractérise la durée, le Flux Spectral calculé pendant la phase d'attaque qui met en évidence la non-synchronicité de l'amorçage des composantes), et des descripteurs caractérisant la richesse spectrale (par ex. le Centre de Gravité Spectral, le troisième coefficient de Tristimulus TR3). Certains de ces descripteurs, notamment le Centre de Gravité Spectral, apparaissent de manière récurrente dans les études sur la discrimination de timbres (par ex. [Grey, 1977; Krumhansl, 1989; McAdams *et al.*, 1995; Caclin *et al.*, 2005]). Kendall et Carterette ont montré que le Centre de Gravité Spectral est fortement corrélé avec la perception de la *nasalité* (également appelée *brillance* ou *acuité*) [Kendall et Carterette, 1993]. Dans l'étude de McAdams *et al.*, menée à partir de sons provenant de différents instruments de l'orchestre, le Logarithme du Temps d'Attaque et le Centre de Gravité Spectral sont les meilleures prédicteurs des deux premières dimensions de l'espace perceptif obtenu [McAdams *et al.*, 1999]. Pour les sons que nous avons utilisés, attaque et richesse spectrale covarient : les sons à l'attaque lente sont également les plus sourds, et à l'inverse les sons à l'attaque rapide sont également les plus brillants. Il est donc normal que le Temps d'Attaque et le Centre de Gravité Spectral soient associés à la même dimension. Notons que ceci est intrinsèquement lié à la façon dont les sons ont été générés (un même profil de pression d'alimentation dont l'amplitude maximale varie, et différentes ouvertures du canal d'anche), mais que ceci ne veut pas dire que l'attaque et la richesse spectrale sont nécessairement covariants pour tous les sons de clarinette.

D'autre part, il est intéressant de remarquer que certains participants ont discriminé les sons en fonction de la sensation de tension qu'ils procurent (par ex. "doux", "agressif"). Il n'est pas illogique de penser que les sons à l'attaque lente et au timbre sourd sont jugés comme étant doux, et qu'à l'inverse, les sons à l'attaque brève et très brillants sont jugés comme étant agressifs. Zwicker a montré que la sensation de gêne occasionnée par des sons de même sonie augmente lorsque l'acuité (Centre de Gravité Spectral calculé à partir d'un modèle physiologique de l'audition) augmente. Le lien manifeste entre timbre et tension est particulièrement intéressant sur le plan musical. En effet, la musique est généralement constituée d'une succession de phases de tensions et de détentes (cf. la traditionnelle progression harmonique II/V/I, par exemple). Des études montrent que la sensation de tension est corrélée avec la réponse émotionnelle des auditeurs [Vines *et al.*, 2006]. Nous faisons l'hypothèse que les musiciens utilisent les variations de timbre afin de jouer sur cette dimension de tension, et que du point de vue de l'auditeur, ceci entraîne des modifications de la réponse émotionnelle. Il serait intéressant de confronter des mesures continues de la brillance à des mesures continues de tension, en situation musicale.

Certaines expressions semblent également faire référence à la situation d'anche battante ("richesse des sons", "grésillement", "intensité de l'attaque"), qui comme nous l'avons vu est mise en évidence par le rapport Impair/Pair. Ce dernier est la signature de la forte augmentation des harmoniques pairs de rangs élevés, qui intervient peu après la phase d'attaque, dans le cas des stimuli étudiés. L'irrégularité de l'enveloppe spectrale est également l'un des facteurs déterminant du timbre (cf. Exemples Sonores 31 et 32 de la section 2.2.5.4). Kendall et Carterette ont montré que des sons ayant un même Centre de Gravité Spectral mais des irrégularités spectrales différentes sont très bien discriminés [Kendall et Carterette, 1996]. Ces analyses sont consistantes avec la forte corrélation entre Rapport Impair/Pair et la troisième dimension de l'espace perceptif. Le Rapport Impair/Pair traduit une subtilité du timbre de la clarinette, mais à laquelle nous sommes très sensible.

	Critères		
Participant	Brillance	Attaque	Tension
P1	“nasillard”		“doux”
P2	“riche et nasillard”	“attaque plus ou moins molle”	
P3	“puissance des harmoniques aigus”		
P4	“son sourd (mat) / son brillant”	“douceur de l’attaque”	“douceur de l’attaque”
P5		“attaque”	
P6	“nasillard”	“attaque”	
P7	“richesse du son dans les aigus”	“dynamique du transitoire” “longueur du transitoire”	
P8			“son agressif vs son plus doux”
P9	“son perçant, son brillant”	“attaque”	“son perçant”
P10	“notes [...] rondes” / notes “nasillardes”		
P11	“brillance du son après l’attaque”	“temps d’attaque”, “intensité de l’attaque”	
P12	“grésillement”		
P13	“richesse des sons (plus d’harmoniques aigus)”		
P14	“Intensité du grésillement”		
P15	“son rond, plein versus étriqué” “brillance”	“attaque (molle ou brève et intense)”	
P16	“brillance”	“attaque (ou plus généralement transitoire)”	

TAB. 4.5 – Expressions utilisées par les participants pour décrire les stratégies de discrimination des sons. Elles sont regroupées en trois catégories principales : brillance, attaque, tension. Ces catégories ne sont pas nécessairement indépendantes. Certaines expressions peuvent donc se retrouver dans plusieurs catégories à la fois.

L'espace de timbre perceptif à trois dimensions associé à cette étude, ainsi que les principaux corrélats acoustiques déterminés, sont représentés sur la figure 4.20.

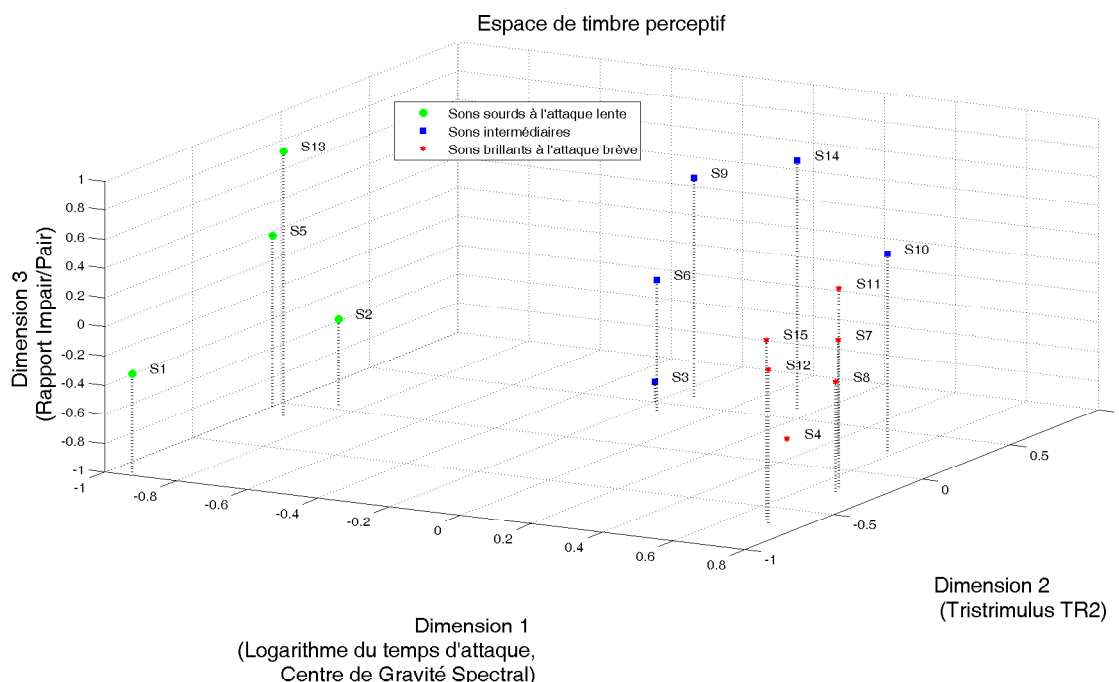


FIG. 4.20 – Espace de timbre perceptif des sons de synthèse de clarinette et corrélats acoustiques. Les trois principales catégories de sons déterminées par l'analyse en cluster sont représentées (sons sourds à l'attaque lente, sons intermédiaires, sons brillants à l'attaque brève).

4.4 Conclusions : vers des descripteurs de timbre adaptés

L'étude perceptive de dissemblance de timbres de sons de synthèse de clarinette a révélé l'importance de descripteurs caractérisant l'attaque (par ex. Temps d'Attaque), la richesse spectrale (par ex. Centre de Gravité Spectral, Tristimulus TR3) et l'irrégularité spectrale (Rapport Impair/Pair) des sons. En effet, l'analyse en cluster des dissimilarités révèle que des sons de type *piano*, à l'attaque longue et au timbre sourd, sont discriminés de manière systématique des sons de type *forte*, à l'attaque brève et riche spectralement.

Des analyses de sons naturels ont révélé des comportements similaires à ceux du modèle de synthèse : augmentation de la brillance des sons lorsque la pression augmente, cassure dans le comportement du Rapport Impair/Pair lorsque le phénomène d'anche battante se produit. Loureiro *et al.* ont réalisé une étude sur la classification de 76 sons naturels de clarinette correspondants aux 19 notes du registre grave de l'instrument jouées pour différentes intensités (de *pp* à *ff*). Leurs résultats montrent également que le Centre de Gravité Spectral est l'attribut expliquant le mieux la catégorisation des sons en fonction de clusters établis à partir d'une Analyse en Composantes Principales des paramètres de la distribution spectrale [Loureiro *et al.*, 2004].

Les modèles de synthèse reposant sur la modélisation physique sont d'une grande utilité pour étudier le fondement de la relation entre le contrôle instrumental et le son. Le modèle de synthèse de clarinette nous a permis de mieux cibler les descripteurs de timbre susceptibles de varier au cours du jeu sur un instrument naturel, et qui se révèlent importants sur le plan perceptif. A la

lumière des résultats de l'étude de dissemblance, nous nous concentrerons par la suite sur le Temps d'Attaque, le Centre de Gravité Spectral, et le Rapport Impair/Pair pour caractériser le timbre de la clarinette. Cependant, les modèles de synthèse sont généralement trop simplifiés pour reproduire les subtiles possibilités offertes par les instruments naturels. Le contrôle dynamique du conduit vocal, qui n'est pas encore pris en compte dans les modèles physiques, semble notamment avoir de grandes répercussions sur le timbre [Fritz, 2004]. De plus, les contrôleurs numériques associés aux modèles de synthèse (par ex. Yamaha WX-5 pour la clarinette) ne sont que des ersatz des vrais instruments, auxquels les musiciens doivent s'accoutumer. Pour ces raisons et afin d'étudier le timbre en situation d'interprétation musicale, nous avons par la suite orienté nos recherches sur l'instrument naturel.

“Le timbre est matière, sonorité, mais aussi processus, forme : la forme découle de son destin, de ses aventures, de son inscription dans le temps - la musique est “ars bene movandi” qui doit absorber l’auditeur dans son devenir.”

[Risset, 1991, p. 259-260]

Chapitre 5

Relation entre l’intention expressive de l’interprète et les variations de timbre

Certains des travaux exposés dans ce chapitre sont présentés dans l’article [Barthet *et al.*, 2006] et dans l’article [Barthet *et al.*, 2009a].

5.1 Introduction

L’interprétation musicale est un acte au cours duquel l’interprète transforme le signal notationnel du compositeur en un signal acoustique destiné à être décodé par un auditeur (cf. section 1.4.1 et figure 1.4.2). Elle est le fruit d’une complexe interaction impliquant notamment l’intention du musicien, ses gestes de contrôle de l’instrument, et les possibilités acoustiques de ce dernier (cf. dans le cas de la clarinette, la cartographie du jeu instrumental présentée dans l’Annexe E). Nous soutenons l’hypothèse, comme nombre d’auteurs, que la musique est un langage spécifique au travers duquel l’interprète (dans le cas de musiques écrites) ou plus généralement l’instrumentiste (dans le cas de musiques improvisées) peut exprimer des sentiments. L’*expression musicale* réfère aux manifestations de ces sentiments dans l’œuvre jouée.

Découvrir quels sont les paramètres acoustiques corrélés à l’expression musicale revêt un intérêt fondamental afin d’améliorer notre compréhension de la perception musicale et par là-même, trouve de multiples applications en synthèse sonore. De bonnes revues des études sur l’interprétation musicale dans les domaines de la psychologie et de l’acoustique musicales pourront être trouvées dans les références [Gabrielsson, 1999] et [Palmer, 1997]. Seashore a été l’un des premiers à avancer que l’expression musicale pouvait se traduire par les déviations de certains paramètres musicaux (par ex. hauteur, rythme, dynamique, timbre) par rapport à un niveau “neutre”, régulier (“the artistic expression of feeling in music consists in aesthetic deviation from the regular - from pure tone, true pitch, even dynamics, metronomic time, rigid rhythms, etc.” [Seashore, 1938, p. 9]). L’expression en musique est en effet liée à la notion de variations (on parle alors de *variations expressives*). Le vibrato du chanteur (variation de hauteur autour d’une référence), les nuances (variations de l’intensité), les changements locaux du tempo (variations rythmiques) sont autant de facteurs qui en témoignent.

De nombreuses études ont démontré que les variations de hauteurs (par ex. vibrato), de rythme (par ex. position temporelle, durée des notes, articulation), et d’intensité, sont corrélées à l’intention expressive de l’interprète (cf. par ex. [Seashore, 1938; Gabrielsson, 1988; Friberg *et al.*, 1991; Repp, 1992; Todd, 1995; Palmer, 1996; Sundberg, 2000; Dixon *et al.*, 2002]). Ces résultats découlent d’analyses systématiques de descripteurs de l’expression musicale (par ex. mesures des variations de hauteurs, de la durée des notes jouées, de profils de variations de l’intensité) qui

montrent que les interprètes reproduisent très fidèlement les mêmes variations expressives selon une interprétation musicale donnée. La nature de ces variations s'avère être étroitement liée à la structure musicale, au point que des modèles prédicteurs de l'interprétation ont pu être établis. Les références [Widmer et Goebel, 2004] et [De Poli, 2006] proposent de bons historiques des études sur les modèles d'interprétation musicale. Certains reposent sur l'analyse par la mesure (par ex. [Todd, 1992, 1995; Windsor et al., 2006]), d'autres sur la théorie musicale et l'analyse par la synthèse (par ex. [Sundberg et al., 1989; Friberg, 1995; Bresin, 2000]), et d'autres encore sur l'apprentissage automatique (par ex. [Tobudic et Widmer, 2003; Goebel et al., 2004]).

Certaines règles d'interprétation semblent être communes à de nombreux interprètes (le ralentissement final du tempo (*ritardando*) à la fin d'une phrase musicale en est un exemple, cf. les Exemples Sonores *n^{os}* 14 et 15 de la section 1.4.3). En effet, certains auteurs soutiennent l'hypothèse que ces règles d'interprétation sont utilisées par les interprètes afin de communiquer la structure musicale de l'oeuvre à l'auditeur et lui fournir des clefs pour la décoder [Clarke, 1988]. S'il semble exister des façons communes d'interpréter et de rendre vivante la musique, de larges différences entre les interprètes ont également été relatées. Repp a montré que les variations de la durée des notes et de leurs positions temporelles pouvaient refléter le caractère original et personnel du jeu de certains pianistes célèbres [Repp, 1992].

Si la plupart des études sur l'interprétation musicale portent sur les hauteurs, le rythme et la dynamique, bien moins se consacrent au timbre. La complexité de cet attribut du son encore mal compris et le manque de notations musicales lui faisant référence¹ ne sont pas sans y être liés. Le timbre paraît pourtant être l'un des paramètres essentiels du son pouvant, pour certains instruments, être modifié par l'instrumentiste pendant la production même des notes. La faculté d'un instrument à produire des sons dont le timbre (au sens de qualité sonore) est variable dans le temps, dépend étroitement de ses spécificités acoustiques. Les instruments pour lesquels le couplage entre l'excitateur et le résonateur est maintenu pendant toute la durée du son (ou presque), i.e. les instruments aux oscillations auto-entretenues, sont propices aux variations de timbre. C'est le cas de la voix bien sûr, mais également des instruments à vent, des instruments à cordes frottées, etc. Dans une étude sur l'analyse de différentes interprétations d'un *aria*² de Verdi, Födermayr et Deutsch ont signalé que l'un des interprètes utilisait une subtile modification de timbre sur l'une des voyelles de la mélodie chantée afin de varier son expressivité [Födermayr et Deutsch, 1993]. Les chanteurs peuvent en effet modifier la géométrie de leur conduit vocal pour changer leur timbre (cf. par exemples les voyelles chantées "A/É/I/O/U" de l'Exemple Sonore *n^o* 8, section 1.1.4). Dans [Canazza et al., 1991], les auteurs signalent des différences de timbre (Centre de Gravité Spectral et Irrégularité Spectrale) lors de différentes interprétations musicales de clarinette. Comme nous l'avons montré au chapitre précédent, la clarinette est un instrument entretenu qui dispose d'une large palette de timbres. Nous avons notamment montré que les principaux contrôles de l'instrument (pression et pince) peuvent entraîner d'importants changements de la forme de la distribution spectrale des sons produits. Le Temps d'Attaque, le Centre de Gravité Spectral, le Rapport Impair/Pair se sont avérés être de bons corrélats psychoacoustiques de la discrimination des différents timbres de l'instrument.

L'étude de dissemblance décrite au chapitre 4 a été menée à partir de sons isolés. Les différences de timbre concernaient des différences de qualité sonore entre des sons distincts. Lors

¹Cette remarque vaut pour le système de notation traditionnel occidental. Nous avons vu que les systèmes de notations Chinois et Japonais emploient des symboles renvoyant au timbre, qui ont une importance déterminante pour la restitution fidèle de l'oeuvre écrite (cf. section 1.1.4). D'autre part, les musiques contemporaines occidentales se sont largement élargies sur la fonction musicale du timbre, et il n'est pas rare que certains compositeurs utilisent des indications - souvent personnelles - faisant référence au timbre.

²Mélodie expressive chantée, souvent rencontrée dans le contexte de l'opéra.

d'une analyse du timbre en situation musicale, la notion d'évolution temporelle de timbre prend toute son importance. En effet, la musique est un processus qui se définit dans le temps ; elle est généralement constituée d'une succession de notes, de phrases, de mouvements dont le dessein est de communiquer les idées musicales au cours de la composition. Il est dès lors présumable que le timbre pris dans l'acception de qualité sonore soit, au même titre que les hauteurs, le rythme ou la dynamique, un processus temporel porteur de formes détenant un rôle expressif. L'étude qui est présentée dans ce chapitre vise à démontrer que les *variations de timbre*³ sont utilisées par les interprètes afin de transmettre l'expression musicale. Dans cette étude, nous nous sommes concentrés sur le signal acoustique et non le contrôle de l'instrument, difficilement mesurable en situation de jeu sur un instrument naturel sans gêner l'instrumentiste⁴. Contrairement aux hauteurs et au rythme pour lesquels les indications de la partition fournissent une référence, il est difficile de définir un niveau "neutre" pour le timbre des sons instrumentaux. Tout son produit sur un instrument connaît nécessairement au cours de son déroulement dans le temps une phase d'attaque, une éventuelle portion soutenue, puis une décroissance. Des changements de qualité sonore "instantanés" s'ensuivent (cf. [Exemple Sonores n^{os} 26, 27, et 28](#)). A moins qu'une note de musique ne présente un réel défaut qui la rende désagréable à l'oreille, ou qui laisse traduire un contrôle involontaire et inexpérimenté de l'instrument (par ex. le canard de clarinette), une note prise seule, hors contexte, est difficilement qualifiable d'expressive ou d'inexpressive. La notion d'expression musicale et donc ses liens avec le timbre ne peut prendre sens qu'au sein d'un discours musical. Si la notion de *variation expressive de timbre* a un sens, nous supposons qu'elle se manifeste dans la façon dont l'interprète sculpte, façonne, le son des notes qu'il produit au cours du temps.

Nous avons donc développé une méthodologie permettant de comparer les variations temporelles de timbre produites par le musicien lors de différents types d'interprétations musicales. Nous avons utilisé comme référence de comparaison des variations de timbre, une interprétation musicale dite "scolaire", ou "neutre", pour laquelle l'interprète joue *a priori* de manière inexpressive. La méthodologie a été bâtie en fonction des deux questions suivantes :

- Est-ce que les variations de timbre produites par un interprète sont régulières lorsqu'une interprétation musicale est répétée selon une même intention expressive ?
- Est-ce que ces variations de timbre diffèrent lorsque l'intention expressive change ?

Afin de répondre à ces questions, nous avons étudié les corrélations entre des mesures objectives des variations de timbre sur un ensemble de séquences musicales, produites par un clarinettiste professionnel selon différentes intentions expressives. Les extraits musicaux utilisés sont issus de répertoires musicaux traditionnels, Baroque (*Suite de Bach n^o II*, BWV. 1008) et Classique (*Quintette pour Clarinettes et Cordes* de Mozart, KV. 581). Nous avons également confronté l'analyse du timbre à celle du rythme⁵, de la dynamique et de la hauteur. Il nous a fallu pour cela développer au préalable une méthodologie d'analyse de l'expression musicale à partir du signal audio. Cette méthodologie engage une procédure de segmentation des notes (détection de l'amorce et de la fin des notes), et le calcul de descripteurs adaptés.

Les descriptions du corpus sonore et de la méthode d'analyse de l'expression musicale font l'objet des premières sections de ce chapitre. Nous présentons ensuite les méthodes d'analyses statistiques employées afin de comparer les variations des descripteurs pour différentes interpré-

³Ce que nous désignons par variations de timbre correspond aux variations temporelles de qualité sonore des sons musicaux.

⁴Des recherches sur la mesure du contrôle des instruments naturels sont actuellement menées au sein du projet *Consonnes*, dirigé par notre équipe (cf. [\[CONSONNES, 2008\]](#)).

⁵Ce que nous entendons par rythme lorsque nous parlons d'expression musicale, concerne les fines déviations d'ordre rythmique effectuées par les interprètes. Le rythme moyen de la pièce musicale lui ne change pas, il est donné par la partition et reste un invariant.

tations. Dans la dernière partie, les résultats obtenus à l'aide de ces analyses seront présentés et discutés.

5.2 Corpus sonore

Deux extraits musicaux différents ont été utilisés lors de séances d'enregistrements réalisées avec un même clarinettiste professionnel. Les extraits proviennent des répertoires Baroque (Bach) et Classique (Mozart), styles musicaux traditionnels dans lesquels le sonorité tient une place primordiale, et où l'acte d'interprétation musicale est un véritable art. Afin de se concentrer sur le timbre, nous avons cherché des extraits où les effets liés aux hauteurs (par ex. le *glissando*, ou le *vibrato* très prisé dans le jazz) ne prennent pas une place importante dans l'interprétation.

5.2.1 Interprétations de l'extrait de Bach

Le premier extrait musical correspond à la phrase d'ouverture de l'*Allemande* (dance instrumentale) de la *Suite de Bach n° II* (BWV. 1008)⁶. La partition de cet extrait est donnée figure 5.1.



FIG. 5.1 – Partition de l'extrait de l'*Allemande* de la *Suite de Bach n° II* (BWV. 1008).

Le rythme de la phrase est binaire et ses notes appartiennent au registre le plus grave de la clarinette, le registre *chaleur* (cf. section 3.1). Les notations de la partition indiquent de jouer le mouvement dans un style "lourd" et "expressif". Un *ritardando* à la fin de la phrase est également indiqué. Une annotation personnelle de l'interprète indique une nuance sur la deuxième partie de la phrase (*decrescendo*).

Nous avons demandé au clarinettiste de jouer la phrase musicale selon deux interprétations différentes, une première de façon "scolaire" ou "mécanique" (inexpressive), et une seconde de façon expressive. Pour chaque interprétation, il a été demandé à l'interprète de répéter 20 fois la phrase sans changer d'intention musicale.

Un tempo de référence de 48 bpm (*beats per minute*) a été choisi. Lors des interprétations scolaires, le musicien disposait d'un métronome électronique qui lui était transmis au moyen d'une oreillette. Pour les interprétations expressives, le métronome n'était déclenché qu'avant les enregistrements, puis était ensuite coupé afin d'éviter des contraintes rythmiques. Les [Exemples Sonores n° 41](#) et [42](#) correspondent à deux des interprétations scolaires. Les [Exemples Sonores n° 43](#) et [44](#) correspondent à deux des interprétations expressives.

5.2.2 Interprétations de l'extrait de Mozart

Le premier extrait musical est assez court et le débit des notes est rapide (essentiellement des doubles-croches). Nous avons sélectionné un deuxième extrait plus long et au tempo plus lent. Il

⁶Cette Suite fait partie des *Suites de Bach pour Violoncelle* (BWV. 1007/12). Une adaptation pour clarinette de U. Delécluse a été employée par l'interprète.

correspond aux 17 premières mesures du *Larghetto* du *Quintette pour Clarinettes et Cordes* de Mozart (KV. 581). Sa partition est donnée figure 5.2.



FIG. 5.2 – Partition de l'extrait du *Larghetto* du *Quintette pour Clarinettes et Cordes* de Mozart (KV. 581).

Le rythme de l'extrait est ternaire et le mouvement est écrit pour le registre moyen de la clarinette, le *clairon* (cf. section 3.1). Nous avons choisi un mouvement *Larghetto* car de tels mouvements sont destinés à être joués de manière relativement lente. En effet, nous supposons que des notes longues (par ex. blanches, noires pointées) octroient à l'interprète un champ d'action plus grand sur le timbre que des notes courtes (par ex. doubles-croches). Ceci rejoint le point de vue de [Drake et Palmer, 1993], qui mentionne le fait que le contenu expressif de certaines notes est intrinsèquement contraint par des facteurs physiques (par ex. une courte durée), et/ou des facteurs liés à la structure musicale. Un tempo légèrement plus lent que pour le premier extrait a été choisi en conséquence (44 bpm).

Nous avons demandé au clarinettiste de jouer deux interprétations de manière scolaire, et quatre interprétations de manière expressive.

L'emploi du métronome lors de la première séance d'enregistrement (extrait de Bach) s'est avéré un peu gênant pour le musicien, même lors d'une interprétation scolaire. Pour cette raison, nous n'avons pas utilisé de métronome pendant les enregistrements de l'extrait de Mozart. Pour les deux types d'interprétation, le métronome n'a été utilisé qu'avant le début des enregistrements afin que la même pulsation générale soit gardée entre les différentes répétitions, et que les interprétations soient jouées sans contraintes rythmiques. Les Exemples Sonores *n*^{os} 45 et 46 correspondent aux deux interprétations scolaires. Les Exemples Sonores *n*^{os} 47 et 48 correspondent à deux des interprétations expressives.



FIG. 5.3 – Enregistrement de séquences musicales de clarinette en chambre anéchoïque. Deux types de microphones ont été utilisés, un système de microphones fixés sur le corps de l'instrument (*SD Systems LCM 82*), et un microphone en champ semi-lointain (*Neumann KM 183*).

5.2.3 Dispositifs d'enregistrement

Les enregistrements ont tous été effectués en chambre anéchoïque afin d'éviter des modifications du spectre externe de l'instrument dues aux réflexions ayant lieu dans des salles réverbérantes. Les séquences ont été enregistrées en numérique sur 16 bits, à une fréquence d'échantillonnage de 44.1 kHz. Puisque d'une manière globale le timbre peut être influencé par le microphone utilisé, et sa position par rapport à l'instrument, deux dispositifs différents ont été employés. Le premier dispositif est un système constitué de deux microphones fixés sur l'instrument (*SD Systems LCM 82* pour clarinette), l'un étant positionné au niveau du barillet, et l'autre au niveau du pavillon (cf. photo de la figure 5.3). Le second est un microphone omnidirectionnel *Neumann KM 183*, positionné à environ 1,50 m de l'instrument, et formant un angle droit avec le corps de l'instrument (la plupart des notes de la clarinette rayonnent au travers des trous latéraux ouverts, cf. section 3.3.4).

Pour les enregistrements de l'extrait de Bach, le timbre du registre grave de l'instrument a été jugé mieux restitué par les microphones *SD Systems* que par le microphone *Neumann*. Nous avons donc sélectionné les séquences correspondantes pour mener nos analyses. A titre de comparaison, l'[Exemple Sonore n° 43](#) correspond à un enregistrement de l'extrait de Bach effectué avec les microphones *SD Systems*, tandis que l'[Exemple Sonore n° 49](#) correspond à un enregistrement effectué avec le microphone *Neumann*. En revanche, pour les séquences musicales de l'extrait de Mozart, nous avons préféré les enregistrements réalisées avec le microphone *Neumann*. Ils ont été jugés plus fidèles au timbre de la clarinette dans son registre *clairon* que ceux réalisés avec les microphones *SD Systems*, qui sont très brillants et manque de fréquences basses. L'[Exemple Sonore n° 50](#) correspond à une séquence de l'extrait de Mozart enregistrée avec le microphone *SD Systems*, tandis que l'enregistrement de l'[Exemple Sonore n° 48](#) provient du microphone *Neumann*.

Ces choix revêtent certes un caractère subjectif, mais trouvent une explication logique, en

Extrait musical	<i>Allemande - Suite de Bach</i>	<i>Larghetto - Quintette de Mozart</i>
Rythme	binaire	ternaire
Durée (mesures)	1,8 mesures	17 mesures
(nombre de notes)	(27 notes, N1 à N27)	(75 notes, N1 à N75)
Nombre de phrases	1	4
Interprétations		
scolaires (I1)	20 répétitions (P1 à P20)	2 répétitions (P1 à P2)
expressives (I2)	20 répétitions (P21 à P40)	4 répétitions (P3 à P6)
Nombre total de séquences	40	6
Tempo de référence (bpm)	48	44
Microphones	SD Systems LCM 82	Neumann KM 183

TAB. 5.1 – Description du corpus sonore

lien avec le fonctionnement acoustique de l'instrument. En effet, pour l'extrait de Bach, les notes graves émises au travers du pavillon de l'instrument ne sont pas rayonnées dans la direction du microphone *Neumann*. Or, les vibrations sonores hautes fréquences sont très directives et se propagent donc essentiellement dans la prolongation du corps de l'instrument⁷. De plus, elles meurent plus rapidement que les basses fréquences en raison de leur plus grande sensibilité aux pertes visco-thermiques. Malgré le fait que le micro *Neumann* soit omnidirectionnel, il est logique qu'il ne capte pas bien la partie hautes fréquences des notes graves dans la configuration d'enregistrement associée à cette expérience. Il nous semble en effet que les séquences correspondantes manquent d'aigus, problème que ne connaissent pas celles enregistrées avec le dispositif *SD Systems*, puisque l'un des deux microphones de ce dernier est justement placé au niveau du pavillon. Le phénomène inverse se produit pour l'extrait de Mozart, puisqu'il n'engage que des notes du registre clairon, et donc pas de notes graves émises par le pavillon. Le dispositif *SD Systems* se trouve donc "privé de l'une de ses deux oreilles", et ne réceptionne les vibrations sonores qu'en grande partie par le microphone barillet. Il est logique que les enregistrements résultants manquent de basses fréquences, qui elles sont bien restituées par le microphone *Neumann*, placé en champ semi-lointain.

D'autre part, indépendamment de ces considérations, le fait que deux microphones différents aient été choisis pour les deux extraits différents n'est pas problématique pour les analyses qui vont suivre, car ces dernières consistent à *comparer* des interprétations enregistrées à chaque fois avec le même dispositif.

La tableau 5.1 présente une description synthétique du corpus sonore, ainsi que la nomenclature utilisée pour faire référence aux différentes séquences musicales enregistrées.

5.3 Segmentation

L'analyse de l'expression musicale requiert une étape préliminaire importante, celle de la segmentation des différentes notes des séquences musicales. Les notes représentent en effet les plus petits éléments structurels de la musique (dans le cas de la musique traditionnelle instrumentale, que nous traitons ici). Certains corrélats de l'expression rythmique sont définis à partir de la durée des notes jouées par l'interprète (cf. ci-dessous, à la section 5.4.2). Concernant le timbre, si nous attachons de l'importance à la caractérisation des variations temporelles "instantanées" du spectre (spectre à court-terme, en l'occurrence), il peut être également intéressant de les caractériser sur

⁷ce qui explique, par exemple, le son très aigu perçu lorsque l'on est dans "la ligne de mire" d'un trompettiste.

une autre unité temporelle d'analyse (notes ou groupe de notes par exemple).

La segmentation des notes se réalise au travers de la détection de leur amorce (*onset* en anglais) et de leur fin (*offset* en anglais). La référence [Bello *et al.*, 2005] fournit une très bonne revue de l'état de l'art des études sur la détection de l'amorce des notes. Les méthodes mises au point reposent généralement sur la localisation d'événements "transitoires" dans le signal. Un événement transitoire peut se traduire de différentes manières : une discontinuité soudaine et intense de l'énergie, un changement dans le spectre à court-terme ou dans les propriétés statistiques du signal, etc. Les méthodes de détection d'amorces peuvent donc être classifiées en fonction du type de description du signal sur lesquelles elles se reposent : l'enveloppe d'amplitude du signal, les amplitudes et phases spectrales issues de représentations temps-fréquence, les coefficients résultant d'une analyse temps-échelle, des modèles probabilistes de signaux, etc. [Bello *et al.*, 2005]. Bien que les méthodes développées offrent généralement de bonnes performances, il n'existe pas à ce jour de méthodes qui ne soient pas sujettes aux erreurs. Ceci n'est pas étonnant lorsqu'on considère la grande diversité que peuvent présenter les sons musicaux (sons à hauteurs déterminées et non percussifs comme ceux du violon, sons à hauteurs déterminées et percussifs, comme ceux du piano, sons polyphoniques dans le cas d'accords, mélanges complexes d'instruments, etc.). Les méthodes développées nécessitent toutes encore, au bout de la chaîne d'analyse, l'expérimentateur, dont l'oreille seule, peut valider les résultats obtenus.

Etant donné le grand nombre de notes à segmenter dans le corpus sonore (le nombre total de notes à segmenter dans le cas des interprétations de Bach est $27 \times 40 = 1080$, cf. tableau 5.1), nous avons développé une méthode semie-automatique⁸ de détection de leurs amorces et de leurs fins. Le problème de segmentation est ici facilité pour plusieurs raisons : la connaissance d'informations *a priori* provenant de la partition (la hauteur et la durée nominale des notes), la bonne qualité audio des enregistrements et le fait qu'il n'y ait pas de réverbération (enregistrement en chambre sourde), la présence de bruits de clefs (transitoires et à large bande) lors de certaines transitions entre notes (cf. les exemples sonores mentionnés ci-dessus). La méthode que nous avons mise au point repose sur la détection des instabilités de la fréquence fondamentale instantanée. En effet, dans le cas de séquence musicale issues d'instruments non percussifs comme la clarinette, un critère reposant seulement sur les variations d'amplitude n'est pas suffisant ; cette dernière varie peu lors de transitions entre des notes jouées de manière *legato*. Comme nous allons le voir, les variations de la fréquence fondamentale instantanée sont en revanche un bon indicateur des transitions entre les notes. Il est important de noter que le cahier des charges de la méthode que nous avons développée concerne le cadre des séquences musicales analysées dans cette étude, mais qu'elle pourrait également s'avérer intéressante afin de traiter d'autres types de séquences musicales. Un schéma synthétique décrivant ses différentes étapes (pré-traitement, réduction du problème, détection, validation/correction) est présenté figure 5.3.

5.3.1 Pré-traitement et réduction du problème

5.3.1.1 Estimation de la fréquence fondamentale

Dans [Gomez, 2002], l'auteur propose une classification des différentes méthodes d'obtention de la fréquence fondamentale instantanée. Les algorithmes peuvent être classés en fonction de leur domaine de traitement (temporel, fréquentiel), mais ce découpage n'est pas exclusif. En effet, certaines techniques peuvent être implémentées dans les deux domaines. C'est le cas de la technique utilisant la fonction d'autocorrélation du signal (pour un signal périodique, la détection du maximum de cette fonction permet de calculer la fréquence fondamentale). D'une manière générale, l'estimation de la fondamentale se heurte à des problèmes tels que la présence de bruit ou

⁸Un certain nombre de corrections sont effectuées à la main.

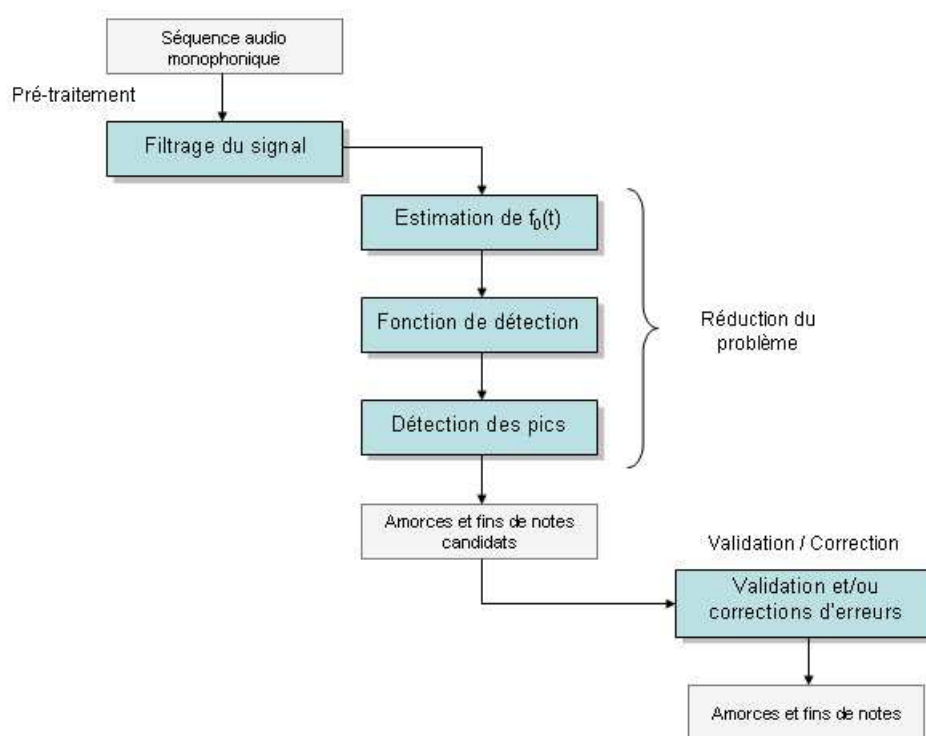


FIG. 5.4 – Méthode de segmentation des notes d’une séquence musicale monophonique reposant sur la détection des instabilités de la fréquence fondamentale instantanée $f_0(t)$.

la présence de partiels dont l’énergie est plus grande que celle du fondamental. Il s’ensuit généralement des erreurs d’estimation (par ex. globale : f_0 un octave trop bas, ou locale : sauts dans les valeurs estimées en raison d’un faible rapport signal/bruit). Nous avons justement utilisé le fait que la fréquence fondamentale instantanée soit mal définie lorsque le bruit prédomine, ou lors d’évènements transitoires (typiquement aux instants de transitions entre notes), pour détecter les instants initiaux des notes.

Nous avons utilisé le Logiciel d’Expertise Acoustique (LEA) développé par l’entreprise Genesis [Genesis, 2008] afin de générer des versions filtrées des séquences enregistrées ne contenant plus que la composante fondamentale. Le logiciel permet en effet d’effectuer la détection et l’estimation des composantes harmoniques d’un signal. La méthode utilisée par LEA découle des travaux de Jaillet et Torrèsani qui sont décrits dans [Jaillet, 2005]. Les méthodes usuelles reposent sur la recherche des crêtes dans le plan temps-fréquence de manière séquentielle, par la détection des pics spectraux, et la détermination des trajets temporels des partiels. La méthode proposée dans [Jaillet, 2005] repose sur une procédure d’optimisation globale qui s’opère directement dans une région définie du plan temps-fréquence. Les hypothèses de cette procédure sont que les composantes varient “lentement” dans le temps (relativement à l’unité d’analyse temporelle), et que leur rang harmonique est connu. Afin d’obtenir une bonne détection de la composante fondamentale, nous avons “aidé” l’algorithme, en traçant graphiquement un profil d’arêtes dans le plan temps-fréquence qui suit la composante fondamentale⁹. Ce profil sert ainsi de conditions initiales à l’algorithme d’optimisation qui détecte les maxima locaux dans une région définie autour du profil. Une fois la composante fondamentale détectée, un filtrage passe-bas est réalisé de manière à générer une nouvelle séquence audio contenant uniquement le fondamental. Puisque le signal ob-

⁹Cette fonctionnalité peut se réaliser à la souris directement à partir du spectrogramme dans LEA.

tenu est monochromatique, l'utilisation du signal analytique permet d'extraire la loi de modulation de fréquence de la composante (cf. section 2.2.5.5). La loi de modulation de fréquence obtenue présente des instabilités au niveau des transitions entre notes lorsque le bruit prédomine. Si cet aspect est gênant pour l'interprétation physique de la fréquence fondamentale, elle représente un avantage pour la détection des transitions entre notes. La fonction de modulation de fréquence fondamentale représente donc un bon candidat pour la détection des instants initiaux des notes.

5.3.1.2 Calcul de la fonction de détection

Les instabilités de la fréquence fondamentale instantanée apparaissent de manière amplifiée sur la fonction de modulation d'amplitude $A_{f_0}(t)$, également dérivée du signal analytique (cf. section 2.2.5.5). Nous avons donc choisi cette dernière comme fonction de détection des amorces et des fins de notes. La figure 5.5(a) présente la fréquence fondamentale instantanée $f_0(t)$ et la fonction de détection associée $A_{f_0}(t)$ pour l'une des séquences de clarinette de l'extrait de Bach. La figure 5.5(b) montre les résultats de la segmentation.

5.3.1.3 Détection des pics

La difficulté rencontrée lorsque l'on cherche à détecter les pics (maxima) d'une fonction est de tomber sur des maxima locaux de faibles amplitudes non désirés. Pour éviter ce problème, nous avons mis au point une technique de seuillage "adaptative" : un pic (candidat) ne sera considéré comme valide que si sa valeur est plus grande que la valeur du logarithme de la fonction de détection $A_{f_0}(t)$ en deux points situés dans le voisinage du pic candidat (à $t \pm \epsilon$). Ce principe permet de s'assurer que le pic candidat correspond bien à un pic recherché et non pas à un maximum local résultant des fluctuations locales de la fonction de détection dues au bruit. Le fait d'utiliser le logarithme permet de ne pas avoir à rentrer manuellement une valeur de seuil. C'est en ce sens que la technique de seuillage est qualifiée d'adaptative. Les instants $(t - \epsilon)$ et $(t + \epsilon)$ sont déterminés en fonction des informations *a priori* dont nous disposons, i.e. la durée nominale des notes de la séquence musicale. En effet, nous nous reposons sur le fait que la durée des notes jouées ne peut pas être inférieure à une certaine valeur déterminée en fonction de la durée nominale de la note la plus courte de la séquence. Cela permet de déterminer une zone temporelle autour d'un pic candidat qui ne contiendra pas l'un des autres pics recherchés. L'algorithme de détection des pics de la fonction $A_{f_0}(t)$ est présenté figure 5.6.

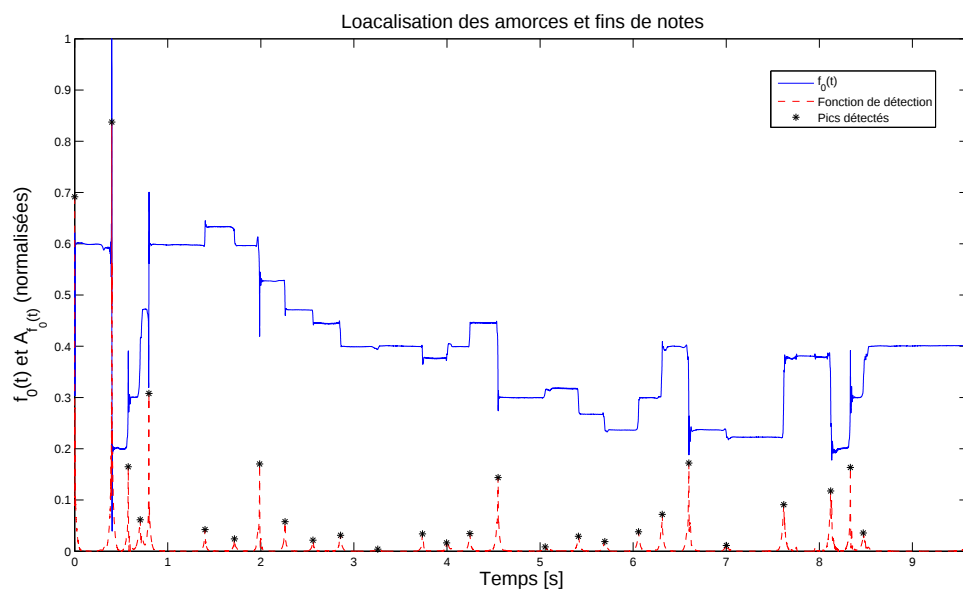
5.3.1.4 Validation et correction d'erreurs

Une fois que les pics ont été détectés une étape de validation et/ou de corrections d'éventuelles erreurs est nécessaire. Nous avons établi un critère automatique de détection d'erreurs grâce aux indications de la partition. Pour chaque note, le critère consiste à vérifier si la fréquence fondamentale moyenne et la durée de la note sont bien cohérentes par rapport à la partition, i.e. qu'elles ne s'écartent pas outre mesure des valeurs nominales. Dans le cas où des erreurs se produisent, nous avons effectué les corrections à la main¹⁰.

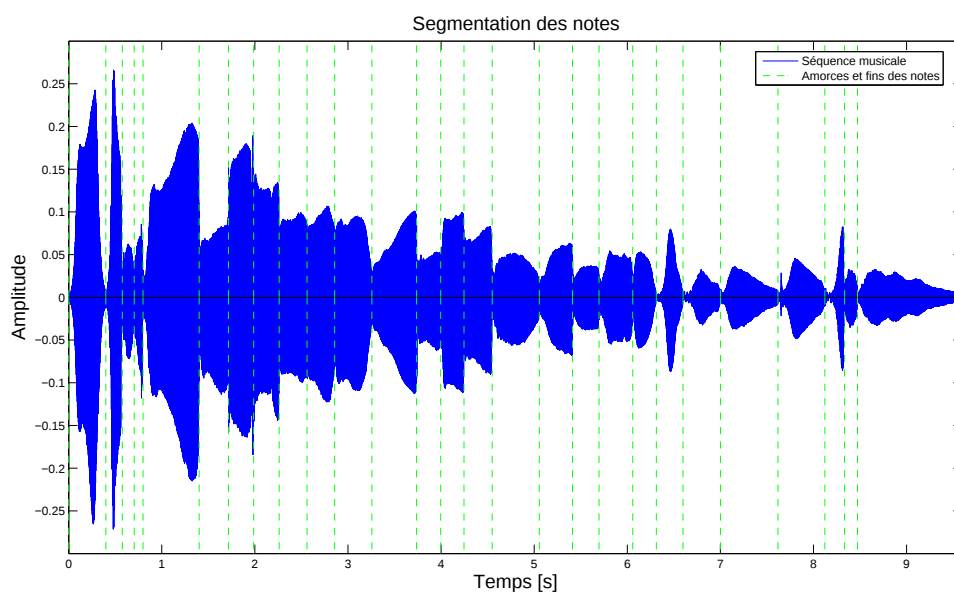
5.3.1.5 Performance de la méthode

Les performances des algorithmes de détection s'évaluent de manière classique en calculant le nombre de Vrais Positifs (VP), i.e. le nombre de bonnes détections, et le nombre de Faux Positifs (FP), i.e. le nombre de fausses détections. Il est à noter que le nombre de détections manquantes correspond à la différence entre le nombre de détections recherchées (le nombre de pics N_p dans notre cas) et le nombre de Vrais Positifs.

¹⁰ou plus précisément à l'oreille !



(a)



(b)

FIG. 5.5 – (a) Fréquence fondamentale instantanée $f_0(t)$ et fonction de détection associée $A_{f_0(t)}$. La séquence analysée est l'une des interprétations à la clarinette de l'extrait de Bach. Les pics détectés sont signalés par des étoiles. - (b) Exemple de segmentation des notes.

Les performances de la méthode de segmentation ont été évaluées sur le corpus de 40 interprétations de l'extrait provenant de la Suite de Bach. Le nombre total de notes contenues dans le corpus, et donc sur lesquelles nous avons évalué la méthode, est de 1080. Le tableau 5.2 présente les performances obtenues.

Il est à noter que dans notre cas, le nombre de Vrais et de Faux Positifs sont complémentaires (i.e. $VP+FP=100\%$). Ceci découle du fait que l'on impose à l'algorithme de détection de notes

```

[Initialisation du nombre de pics à détecter:  $N_p$ ]
 $p \leftarrow 1$ 
Répéter
- Localisation du maximum  $m$  de  $A_{f_0}(t)$  :  $t_{pic\_candidat}$ 
- Calcul des valeurs de  $20.\log[A_{f_0}(t)]$  pour deux instants  $t-\varepsilon$  et  $t+\varepsilon$ 
  au voisinage du pic candidat:  $a^-$  et  $a^+$ 
  [Le pic candidat est retenu s'il satisfait les conditions suivantes]
  - Si  $m > a^-$  et  $m > a^+$ 
    Alors faire
       $t_{pic}(p) \leftarrow t_{pic\_candidat}$ 
       $p \leftarrow p+1$ 
    Fin si
  - Remplissage de zéros de la zone analysée
     $A_{f_0}(t-\varepsilon : t+\varepsilon) \leftarrow 0$ 
Tant que  $p=N_p$ 
Tri des instants  $t_{pic}(p)$  par ordre croissant

```

FIG. 5.6 – Algorithme de détection des pics de la fonction $A_{f_0}(t)$.

	Interprétations scolaires	Interprétations expressives	Moyenne
Vrais Positifs (%)	98.21	98.57	98.39
Faux Positifs (%)	1.79	1.43	1.61

TAB. 5.2 – Performance de la méthode de segmentation (corpus des interprétations de Bach, 1080 notes au total).

un nombre prédéfini de pics à localiser. Lorsqu'un pic recherché n'est pas détecté, un pic erroné prend nécessairement sa place. Comme le montre le tableau 5.2, la méthode fournit d'excellentes performances (98.39 % de détections correctes en moyenne). Les rares fois où les transitions entre deux notes ne sont pas été détectées correspondent en général au cas où les deux notes qui se succèdent ont la même fréquence fondamentale et sont jouées de manière *legato* (par exemple, les deux *E*3, qui se situent vers $t = 3s$ sur la figure 5.5(a)). L'amplitude du pic de la fonction de détection est alors bien plus faible puisque la fréquence fondamentale varie peu, ce qui rend sa détection plus difficile.

5.4 Choix des descripteurs

Nous nous sommes concentrés sur quatre vecteurs (potentiels) de l'expression musicale, le timbre, le rythme, la dynamique et la hauteur. Des descripteurs adaptés à ces variables musicales ont été sélectionnés. Il est important d'insister sur le fait que notre but ici est de comparer les diverses occurrences de ces descripteurs issues d'un corpus d'interprétations musicales. Nous n'avons donc pas spécifiquement cherché à utiliser des descripteurs reposant sur des modèles physiologiques de l'audition (cf. acuité, sonie, modèles perceptifs de hauteur, etc.), généralement plus complexes à calculer, et surtout difficilement inversibles. Or, comme nous l'avons mentionné dans l'introduction, une perspective de ce travail est de pouvoir contrôler des descripteurs de timbre, qui auront au préalable prouvé leur pertinence, au travers d'analyses par la mesure et par la synthèse.

5.4.1 Descripteurs de timbre

Compte-tenu des résultats de l'étude de dissemblance exposés au chapitre 4 et de ceux provenant de la littérature sur le timbre, le timbre de la clarinette a été caractérisé par trois descripteurs rendant bien compte des principales variations de timbre possibles sur l'instrument : le Temps d'Attaque (AT), le Centre de Gravité Spectral (SC) et le Rapport Impair/Pair (OER). Afin d'obtenir de plus amples informations sur ces descripteurs, nous invitons le lecteur à consulter les sections 2.2.5 et le chapitre 4, où ils sont présentés. Leurs formulations sont données par les équations 2.15, 2.22, 2.43.

Nous attachons une importance particulière au Centre de Gravité Spectral pour son caractère général (on le retrouve dans pratiquement toutes les études sur le timbre), de par le fait qu'il est contrôlable dans le temps (le Temps d'Attaque n'est pas contrôlable dans le temps de la même manière), et d'autre part parce que nous supposons que ses variations sont liées à la notion de tension musicale, elle-même corrélée à l'émotion musicale (voir par ex. [Vines *et al.*, 2006]). Nous supposons que dans un contexte adapté, une augmentation de la brillance entraînera une augmentation de la tension, et réciproquement, une diminution de la brillance entraînera une diminution de la tension. D'autres auteurs partagent cette hypothèse, comme dans [Friberg, 2004], où l'émotion de "colère" est associée à un timbre brillant. Il est intéressant de noter que certains participants de l'expérience de dissemblance décrite au chapitre précédent ont utilisé de manière libre, et pour décrire de simples sons isolés, des verbalisations faisant référence à la sensation de tension ("doux"/"agressif"). Il est connu en musique que tension et hauteurs sont intimement liés : à un niveau harmonique (par ex. un accord dominant, i.e. contenant le triton $3^{ce}/7^b$ est tendu, et se résout sur un accord majeur ou mineur dans une cadence V/I), et à un niveau mélodique (par ex. un $C\sharp$ dans une tonalité de C majeur est une note singulière vis-à-vis de l'harmonie tonale traditionnelle, de par le fait qu'elle est dissonnante). Il est probable que le vecteur timbre, et notamment ses aspects liés à la brillance, soit utilisé par les interprètes¹¹ pour amplifier ou diminuer la tension à des instants précis dans la structure musicale.

Descripteurs de notes et descripteurs instantanés Rappelons que le Temps d'Attaque est un descripteur global, caractéristique unique d'un son. En revanche, les descripteurs spectraux et/ou spectro-temporels comme le Centre de Gravité Spectral et le Rapport Impair/Pair peuvent être définis de manière instantanée, à court-terme, ou de manière globale, selon l'unité temporelle d'analyse sélectionnée. Nous avons d'une part calculé ces descripteurs à court-terme, ce qui revient à considérer l'unité temporelle de l'analyse temps-fréquence (Transformée de Fourier à fenêtre glissante). Comme nous l'avons mentionné dans l'introduction, il peut aussi s'avérer intéressant de caractériser ces descripteurs sur une autre unité temporelle d'analyse. Nous avons choisi à ce titre le plus petit élément temporel de la structure musicale, la note. Nous avons donc dérivé du Centre de Gravité Spectral instantané deux descripteurs globaux définis au niveau de la note : le Centre de Gravité Spectral moyen (SCM), qui traduit la brillance moyenne d'une note, et "l'ambitus" du Centre de Gravité Spectral (SCA), qui traduit l'étendue des variations de brillance au cours d'une note. En effet, nous avons montré que ces dernières peuvent s'avérer importantes et audibles (cf. Exemple Sonores *n^{os}* 26, 27, et 28, de la section 2.1.7). Ces nouveaux descripteurs globaux sont définis par les équations suivantes :

$$\begin{cases} SCM &= \frac{1}{D} \sum_{t=t_{on}}^{t=t_{off}} SC(t) \\ SCA &= \underset{t_{on} \leq t \leq t_{off}}{Max} (SC(t)) - \underset{t_{on} \leq t \leq t_{off}}{Min} (SC(t)) \end{cases} \quad (5.1)$$

¹¹dans le cas d'instruments qui le permettent.

où D est la durée de la note considérée, t_{on} et t_{off} sont respectivement ses instants d'amorces (*onset*) et de fins (*offset*). $\max_{t_{on} \leq t \leq t_{off}} (SC(t))$ et $\min_{t_{on} \leq t \leq t_{off}} (SC(t))$ correspondent aux valeurs maximale et minimale du Centre de Gravité Spectral au cours de la note. Notons que ces deux descripteurs sont complémentaires : deux notes peuvent très bien avoir le même Centre de Gravité Spectral Moyen (SCM), mais présenter des ambitus de Centre de Gravité Spectraux (SCA) différents, et vice versa. L'ambitus du Centre de Gravité Spectral est une mesure corrélée à l'écart-type du Centre de Gravité Spectral au cours de la note. Dans une certaine mesure, les descripteurs SCM et SCA reviennent donc à considérer les moments d'ordre 1 et 2 de la série temporelle de SC au cours d'une note.

Le même processus a été effectué pour le Rapport Impair/Pair. Nous avons donc défini, à l'échelle d'une note, le Rapport Impair/Pair moyen (OERM), et l'ambitus du Rapport Impair/Pair (OERA). Leurs formulations découlent d'un raisonnement analogue à celui présenté au travers des équations 5.1.

5.4.2 Descripteurs rythmiques

De manière à caractériser les changements locaux de tempo, nous avons calculé un descripteur des variations rythmiques ΔDUR , défini comme la différence entre les durées des notes jouées par l'interprète (nous les appellerons *durées effectives*) et les durées des notes obtenues à partir d'une transcription des informations rythmiques de la partition (*durées nominales*). Ce descripteur est couramment employé dans la littérature comme corrélat de l'expression rythmique (cf. par ex. [Gabrielsson, 1988]). La formulation de ΔDUR est donnée par l'équation suivante :

$$\Delta DUR = D_{eff} - D_{nom} \quad (5.2)$$

où D_{eff} et D_{nom} sont respectivement les durées effectives et nominales des notes. ΔDUR est homogène à un temps. Il sera exprimé en secondes ou millisecondes.

Le tempo moyen $TEMPO$ des interprétations a également été calculé à partir des durées effectives des notes. Il est obtenu ainsi :

$$TEMPO = \frac{N_{temps}}{\sum_{n=1}^{N_{notes}} D_{eff}^n} \quad (5.3)$$

où N_{temps} est le nombre de temps total de l'extrait musical, N_{notes} est le nombre total de notes et éventuels silences, et D_{eff}^n est la durée effective de la note ou du silence n . Le tempo étant généralement exprimé en nombre de temps par minute (*beats per minute*), les durées effectives D_{eff}^n seront donc exprimées en minutes lors du calcul.

5.4.3 Descripteur de la dynamique

L'enveloppe RMS, notée $ENV(t)$, définie par l'équation 2.14, a été choisie pour caractériser les variations temporelles d'énergie acoustique. D'une manière analogue à celle présentée dans les équations 5.1, l'amplitude moyenne des notes (ENVM), et l'ambitus de l'enveloppe d'amplitude (ENVA) ont été dérivés de l'enveloppe d'amplitude instantanée $ENV(t)$.

5.4.4 Descripteur de la hauteur

La perception de la hauteur des sons complexes est un sujet délicat qui fait l'objet de nombreuses études (voir par ex. [Terhardt et al., 1982b,a]). Pour les sons quasi-harmoniques, la hauteur est étroitement liée à la fréquence fondamentale. Cependant, dans certains cas, elle peut être influencée par la forme de la distribution spectrale. Il est désormais avéré que certaines facettes du

timbre (notamment la richesse spectrale) influencent la hauteur, et que réciproquement la hauteur influence le timbre (cf. par ex. [Marozeau et de Cheveigné, 2007]).

Pour les sons de clarinette, qui contiennent majoritairement des harmoniques impairs, il est raisonnable de se reposer sur la fréquence fondamentale des notes pour décrire en première approximation leur hauteur. Notons que ceci ne reste qu'une approximation, puisque la présence d'harmoniques pairs d'ordres élevés, par exemple, pourrait être un facteur agissant sur le percept de hauteur. Cependant, ce modèle simple sera suffisant dans le cadre de cette étude.

La fréquence fondamentale instantanée des notes $F0(t)$ a été déterminée en utilisant le modèle d'analyse-synthèse paramétrique SAS décrit dans la section 2.2.5.5¹².

De par sa définition et de par la nature des sons analysés, le Centre de Gravité Spectral est étroitement lié à la fréquence fondamentale instantanée (cf. équation 2.22). En effet, pour les sons de clarinette du registre grave, le composant fondamental est en général celui dont l'énergie est la plus forte, et donc "attire" vers lui la moyenne de la distribution spectrale, i.e. le Centre de Gravité Spectral. Puisque nous observons des variations de Centre de Gravité Spectral, il sera important de les confronter aux variations de la fréquence fondamentale instantanée pour dissocier des changements de Centre de Gravité Spectraux dus à d'éventuelles variations de la fréquence fondamentale au sein des notes¹³, de ceux dus à l'intention du musicien.

Comme pour le Centre de Gravité Spectral et le Rapport Impair/Pair, nous avons dérivé de la fréquence fondamentale instantanée $F0(t)$, sa moyenne (F0M), et son ambitus de variation (F0A) pour les différentes notes.

5.5 Calcul et synchronisation des descripteurs

5.5.1 Calcul des descripteurs

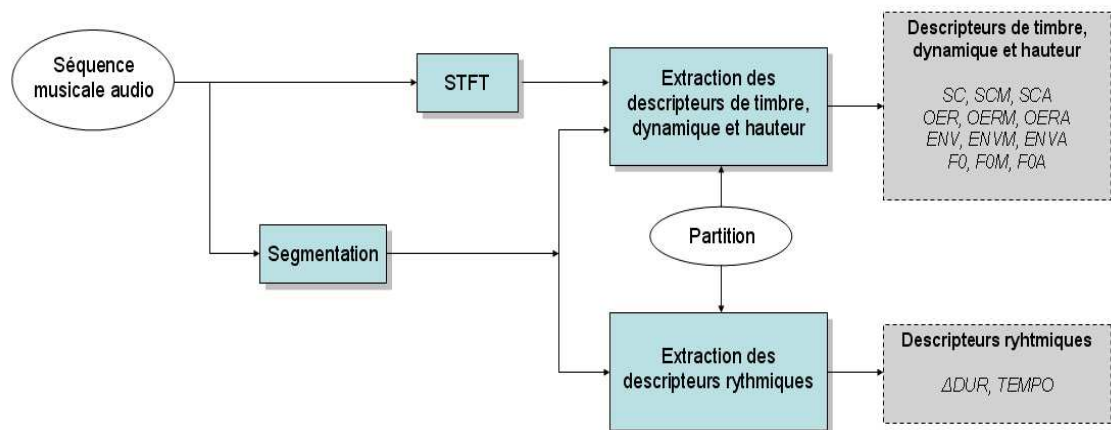


FIG. 5.7 – Méthodologie d'analyse sonologique de l'expression musicale. *STFT* : *Short Time Fourier Transform*.

¹²Signalons que la fréquence instantanée $F0(t)$ n'est pas la même que celle utilisée pour la détection des amorces et fins de notes. $F0(t)$ est issue de l'analyse additive des différentes composantes harmoniques, opérée sur chacune des notes issues de la segmentation.

¹³et du reste des fréquences des autres composants également, les changements ayant généralement lieu de manière analogue sur tous les composants.

La figure 5.7 donne un schéma synthétique de la méthodologie d'analyse sonologique de l'expression musicale développée.

Les descripteurs de timbre spectraux ont été calculés suivant une échelle d'amplitude de puissance afin de conférer un poids plus important aux harmoniques dominants, et ainsi rendre les descripteurs moins sensibles au bruit. De plus, nous avons vu que l'échelle de puissance a permis d'obtenir une corrélation plus forte du SC avec la dimension 1 de l'espace de timbre perceptif obtenu au chapitre 4. Une échelle d'amplitude linéaire s'est avérée plus intéressante pour le calcul du Rapport Impair/Pair pour les sons de synthèses de clarinette, mais nous avons constaté sur des sons naturels que le comportement du descripteur était semblable, que l'échelle soit linéaire ou de puissance. Lors de l'analyse du crescendo naturel (Exemple Sonore n° 33), l'échelle de puissance traduit plus nettement la cassure de OER lors du passage à la situation d'anche battante.

Comme le Centre de Gravité Spectral est dépendant des paramètres d'analyse (ceux de la TFDCT, i.e. nature et taille des fenêtres, et le paramètre stabilisateur b_0 , cf. équations 2.22), des paramètres identiques ont été utilisés afin de pouvoir comparer les descripteurs pour différentes occurrences. Des fenêtres de *Hann* de 1024 points (i.e. environ 20 ms) ont été employées, avec un recouvrement de 50% de la taille de la fenêtre.

Le Rapport Pair/Impair, qui est défini à la fréquence d'échantillonnage des signaux audio analysés selon notre méthode d'analyse (cf. équation 2.43), a été ré-échantillonné relativement à la résolution temporelle de la TFDCT, d'une part afin de réduire la quantité de données à traiter, et par conséquent les temps de calcul de nos algorithmes, et d'autre part afin de pouvoir le comparer avec le descripteur d'enveloppe d'amplitude RMS, défini à court-terme selon le pavage temporel de la TFDCT (cf. équation 2.14).

Les seuils utilisés pour le Temps d'Attaque sont les mêmes que dans l'expérience du chapitre 4. Les paramètres d'analyses employés pour extraire la fréquence fondamentale instantanée et l'enveloppe d'amplitude ont également été gardés constants lors de l'analyse des diverses séquences musicales.

5.5.2 Synchronisation des descripteurs

Les descripteurs globaux (AT, SCM, SCA, OERM, OERA, ENVM, ENVA, ΔDUR , F0M, F0A) calculés pour les différentes interprétations n'ont pas besoin d'être synchronisés entre les différentes interprétations puisqu'ils sont définis à l'échelle des notes. En revanche, une comparaison cohérente des variations de timbre, d'amplitude et de fréquence fondamentale instantanées requiert au préalable la synchronisation temporelle des descripteurs SC, OER, ENV, et F0. En effet, les variations rythmiques de l'interprète ne sont jamais exactement les mêmes (même quand l'intention expressive n'est pas modifiée), ce qui a pour conséquence de modifier les durées et positions temporelles des différentes notes. Pour chaque corpus d'interprétations, nous avons pris comme référence les durées et localisations temporelles moyennes des notes, calculées à partir de l'ensemble des séquences dont nous disposons. Une procédure de compression/expansion temporelle (*time-warping* en anglais) a été réalisée dans le but de synchroniser les descripteurs. L'interpolation par spline cubique a été utilisée à cet effet (fonction *spline* de *Matlab*). Les profils temporels des descripteurs ont ainsi été raccourcis ou allongés, de sorte que pour chaque note, leurs nouvelles durées correspondent aux durées moyennes des notes répétées par le musicien. Notons qu'une procédure similaire a été employée par Wanderley afin de comparer des profils de mouvements spatiaux de clarinettes lors de différentes interprétations [Wanderley, 2002].

5.6 Méthodes d'analyses statistiques

Pour chaque corpus musical (Bach et Mozart), différentes analyses statistiques ont été effectuées pour comparer les valeurs des descripteurs de timbre, de rythme, de dynamique et de hauteur entre les différentes interprétations.

5.6.1 Mesures des dissimilarités entre descripteurs

Une matrice de dissimilarité a été définie pour chaque descripteur afin de rendre compte des variations qu'ils peuvent présenter lors des différentes interprétations. Les dissimilarités ont été calculées en utilisant une métrique reposant sur une mesure de corrélation : la distance entre deux occurrences d'un même descripteur est égale à un moins le coefficient de corrélation de Pearson. Cette analyse a été effectuée aussi bien sur les descripteurs rythmiques définis à l'échelle des notes (AT, ΔDUR), que pour les descripteurs instantanés (SC, OER, ENV, F0).

5.6.2 Classification ascendante hiérarchique

De façon à déterminer si certains descripteurs permettent de discriminer les interprétations scolaires des interprétations expressives, une procédure de classification ascendante hiérarchique des matrices de dissimilarité (algorithme de lien complet) a ensuite été menée.

5.6.3 Analyses de variance (ANOVA)

Les analyses des dissimilarités décrites précédemment permettent de déterminer si les descripteurs présentent des différences à un niveau global. Elles n'aident pas à en revanche à les localiser au sein de la structure musicale. En effet, la métrique de distance utilisée repose sur l'ensemble des valeurs temporelles du descripteur.

Pour le corpus sonore relatif à l'extrait de Bach, qui contient un nombre suffisant d'échantillons représentant chaque type d'interprétation (20), des analyses de variance (ANOVA) à deux facteurs, l'intention expressive (deux modalités : scolaire et expressive) et la note (27 modalités, cf. tableau 5.1), ont été conduites pour chacun des descripteurs globaux (AT, SCM, SCA, OERM, OERA, ΔDUR , ENVM, ENVA, F0M, F0A). Ces ANOVA permettent d'évaluer s'il existe un effet d'interaction entre le type d'interprétation et la valeur des descripteurs définis à l'échelle des notes. Les cas échéant, des analyses *post hoc* de comparaisons multiples ont été menées de façon à localiser les notes ou groupes de notes qui présentent des différences significatives. Les comparaisons multiples ont été réalisées selon la procédure séquentielle de Holm-Sidak [Holm, 1979]. Cette dernière, décrite dans l'annexe F.2, s'avère être plus puissante (moins conservative) que d'autres types de tests de comparaisons multiples non séquentiels (par ex. Bonferroni, Sidak) [Ludbrook, 1998]. De plus, elle est adaptée aux comparaisons multiples relatives à un groupe de contrôle, qui dans cette étude est constitué des différentes interprétations scolaires. Ainsi, seules les comparaisons correspondant aux notes identiques dans la structure musicale sont effectuées : en d'autres termes, si l'on représente la série de notes relatives aux interprétations scolaires par (N_1 , N_2 , etc.), et celle relative aux interprétations expressives par (N_1' , N_2' , etc.), les comparaisons ne sont effectuées qu'entre les notes N_1 et N_1' , N_2 et N_2' , etc. Les autres comparaisons (par ex. N_1 avec N_2' , N_3' , etc.) n'auraient en effet pas de sens. D'une part, elles impliqueraient de comparer des notes qui ne correspondent pas aux mêmes instants dans la structure musicale, et d'autre part, les valeurs de certains descripteurs sont intrinsèquement affectées par la nature de la note (par ex. sa hauteur), ce qui en soi pourrait être à l'origine de différences significatives (par ex., pour les sons instrumentaux, la valeur globale du Centre de Gravité Spectral est en général étroitement liée à la fréquence fondamentale de la note). La procédure de Holm-Sidak a été effec-

tuée pour deux niveaux de significativité, 0.01 et 0.05.

Pour les interprétations de l'extrait de Mozart, une comparaison de la moyenne des descripteurs pour les interprétations scolaires et expressives est effectuée.

5.7 Résultats

5.7.1 Interprétations de l'extrait de Bach

5.7.1.1 Dissimilarité et classification ascendante hiérarchique

Dans les figures qui vont être présentées, les parties supérieures gauches des matrices de dissimilarité correspondent aux dissimilarités entre les 20 interprétations scolaires (P1 à P20), et les parties inférieures droites correspondent aux dissimilarités entre les 20 interprétations expressives (P21 à P40).

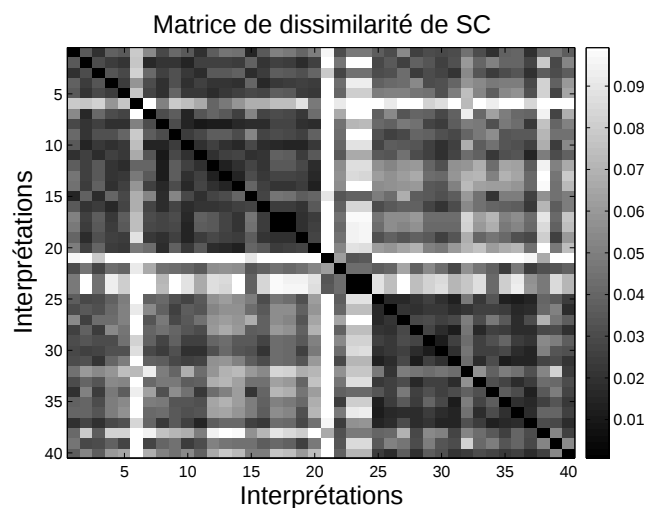
Les figures 5.8(a) à 5.8(c) montrent les matrices de dissimilarité obtenues pour le Centre de Gravité Spectral (a), les variations rythmiques (b), et l'enveloppe RMS (c). Hormis quelques exceptions, les plus faibles dissimilarités entre descripteurs apparaissent entre les interprétations jouées selon une même intention expressive (cf. les parties supérieures gauches et inférieures droites des matrices). En effet, les dissimilarités entre descripteurs sont plus élevées lorsque l'on compare les interprétations scolaires et les interprétations expressives (cf. les parties supérieures droites et inférieures gauches des matrices). Ces résultats sont confirmés par la classification ascendante hiérarchique dont les représentations sous forme de dendrogramme pourront être trouvées dans les figures F.2(a) à F.2(c) de l'Annexe F.2.

Pour le Centre de Gravité Spectral (SC), la classification ascendante hiérarchique fait ressortir deux groupes principaux, qui rassemblent chacun des interprétations jouées selon la même intention expressive (cf. figure F.2(a)). Ainsi, 16 des 20 interprétations scolaires sont rassemblées dans le sous-groupe de droite, et 16 des 20 interprétations expressives dans celui de gauche. Une des interprétations scolaires (P6), et trois des interprétations expressives (P21, P23, et P24) ont été classées à part. Ces résultats montrent d'une part que l'interprète tend à reproduire fidèlement les variations de brillance lorsqu'il joue selon une intention spécifique, et que d'autre part que la nature de ces variations change lorsque l'on passe d'un jeu scolaire à un jeu expressif.

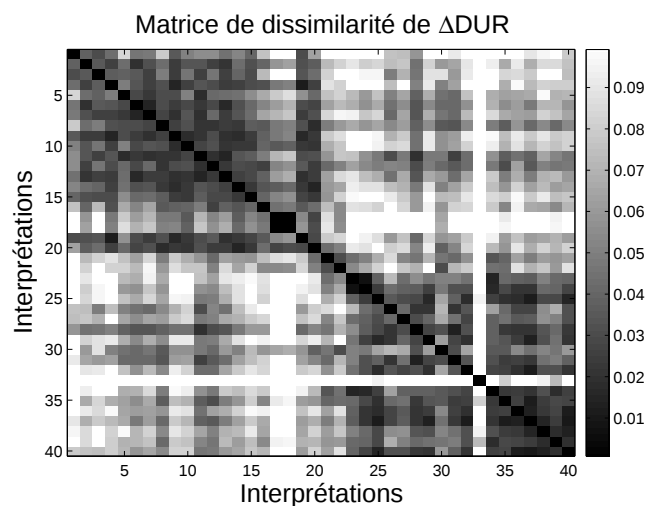
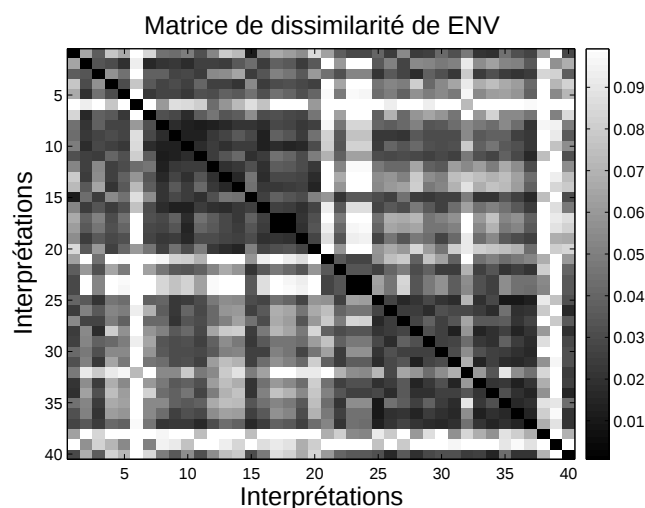
Pour les variations rythmiques (ΔDUR), les deux sous-groupes principaux de la classification ascendante hiérarchique discriminent également très bien les interprétations scolaires des interprétations expressives : 16 des 20 interprétations expressives appartiennent au sous-groupe de droite, tandis que les 20 interprétations scolaires se retrouvent dans le sous-groupe de gauche (cf. figure F.2(b)). Les variations rythmiques de trois des interprétations expressives (P21, P22, et P30) s'avèrent être plus proches de celles présentées par les interprétations scolaires (elles se retrouvent également dans le sous-groupe de gauche). Notons que l'une d'entre elle (P21) présente également des variations de SC différentes de celles des autres interprétations expressives. L'une des interprétations du corpus (P33) manifeste des variations rythmiques qui diffèrent systématiquement des autres.

Ces observations révèlent donc des différences de variations rythmiques entre les interprétations scolaires et expressives. Ceci n'est pas surprenant puisqu'un jeu scolaire est par définition "mécanique", c'est à dire que toutes les notes tendent à tomber sur les temps, et généralement, les variations rythmiques par rapport aux indications de la partition sont faibles¹⁴. De plus, les

¹⁴Notons que le fait qu'il y ait des variations rythmiques par rapport à la partition n'est pas un gage de qualité, encore



(a) Matrice de dissimilarité SC

(b) Matrice de dissimilarité ΔDUR 

(c) Matrice de dissimilarité ENV

FIG. 5.8 – Matrice de dissimilarité des variations de Centre de Gravité Spectral (SC) (a), des variations rythmiques (b), et des variations d'énergie acoustique (c) pour les 40 interprétations de l'extrait de Bach. Les 20 premières interprétations (P1 à P20) correspondent aux interprétations scolaires. Les 20 interprétations suivantes (P21 à P40) correspondent aux interprétations expressives. Plus les points sont foncés, moins les dissimilarités sont élevées.

	Interprétations scolaires	Interprétations expressives
Durées (moyenne, écart-type) en s	9.79 (0.13)	10.75 (0.21)
Tempo (moyenne, écart-type) en bpm	44.45 (0.57)	40.47 (0.78)

TAB. 5.3 – Moyennes et écarts-types des durées et tempi des interprétations scolaires et expressives de l'extrait de Bach. Les durées sont exprimées en secondes, les tempi en nombre de temps par minute (*bpm*).

interprétations scolaires ont été réalisées au métronome. De fait, comme l'indique le tableau 5.3, la durée moyenne des interprétations scolaires est inférieure d'une seconde environ à celle des interprétations expressives. Ceci va de pair avec un tempo moyen plus élevé pour le jeu scolaire (44.45 *bpm*) que pour le jeu expressif (40.47 *bpm*). Notons que le tempo associé au jeu scolaire est inférieur au tempo nominal de 48 *bpm* choisi comme référence. Nous en verrons l'explication plus loin en analysant plus précisément les variations rythmiques. Notons également que dans un jeu expressif les écarts-types des durées et tempi sont plus grands que pour un jeu scolaire, ce qui traduit très probablement le fait que l'interprète, n'étant plus contraint par le métronome, traite la dimension temporelle avec davantage de variabilité.

L'enveloppe d'amplitude globale permet également de différencier les interprétations scolaires des interprétations expressives (cf. figures 5.8(c) et F.2(c)). Le sous-groupe de gauche de la classification ascendante hiérarchique rassemble 13 des 20 interprétations scolaires, tandis que le sous-groupe de droite rassemble 16 des 20 des interprétations expressives. Les trois interprétations expressives (P21, P23, P24), se retrouvent comme pour le Centre de Gravité Spectral classées dans un groupe distinct.

Les dissimilarités globales de Temps d'Attaque, de Rapport Impair/Pair, et de fréquence fondamentale instantanée, ne permettent pas de discriminer les interprétations en fonction de l'intention expressive (les figures correspondant à ces descripteurs ne sont donc pas présentées). Les interprétations scolaires et expressives se retrouvent mélangées au sein de mêmes groupes lors de la classification ascendante hiérarchique des matrices de dissimilarités. Ceci ne veut pas dire que ces paramètres ne présentent jamais de différences systématiques entre les séquences scolaires et expressives, mais que ces dernières apparaissent probablement localement à l'échelle de quelques notes ou groupes de notes.

Les analyses précédentes ont permis de révéler des différences de comportement des descripteurs à un niveau global. Nous allons maintenant présenter les analyses de variance menées à un niveau local (notes ou groupes de notes).

5.7.1.2 Analyses de variance

Les résultats des analyses de variance (ANOVA) sur l'interaction entre le facteur d'intention expressive et le facteur relatif aux notes sont présentés dans le tableau 5.4 pour chacun des descripteurs AT, SCM, SCA, OERM, OERA, ΔDUR , ENVM, ENVA, F0M, F0A. Les tableaux 5.5 à 5.14 indiquent quels sont les notes qui présentent des différences significatives entre les interprétations scolaires et expressives, d'après les résultats des tests de comparaisons multiples (procédure de Holm-Sidak). Les figures 5.9 à 5.12 présentent une comparaison entre les descripteurs moyens calculés pour les 20 interprétations scolaires, et ceux calculés pour les 20 interprétations expressives. Les différences significatives entre interprétations, révélées par les tests de comparaisons multiples, sont mises en évidence sur les figures correspondantes (cf. marqueurs pour les descrip-

faut-il que ces dernières respectent la structure musicale et soient expressives.

teurs globaux, et zones grisées entre les deux courbes pour les descripteurs instantanés). Dans ce qui suit, les résultats seront analysés dans l'ordre suivant : hauteur, rythme, timbre, puis intensité.

Descripteurs	AT	SCM	SCA	OERM	OERA	ΔDUR
F(26;1026)	2.87***	6.33***	6.51***	6.47***	3.97***	16.80***
Descripteurs	ENVM	ENVA	F0M	F0A		
F(26;1026)	5.62***	1.64*	2.43***	2.62***		

TAB. 5.4 – Résultats des analyses de variance sur l'interaction entre le facteur d'intention expressive et le facteur relatif aux notes pour les différents descripteurs AT, SCM, SCA, OERM, OERA, ΔDUR , ENVM, ENVA, F0M, F0A. Les probabilités p qu'il n'y ait pas d'interaction entre les deux facteurs sont représentées par les notations suivantes : * $p < .05$, ** $p < .01$, *** $p < .001$.

Notes	N1	N5	N15	N17	N19	N21
t(52)	3.51**	6.27**	4.05**	3.19**	2.87**	2.33*

TAB. 5.5 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour le Temps d'Attaque (AT). Les comparaisons ont été effectuées pour chacune des 27 notes de l'extrait de Bach en considérant les 20 interprétations scolaires (groupe de contrôle) et les 20 interprétations expressives. Les notes sont représentées par leur rang au sein de la structure musicale. Seules les notes qui diffèrent significativement selon l'interprétation sont présentées dans le tableau. Le risque pris en concluant à une différence par le test t de Student (d.d.l.=52), corrigé pour des comparaisons multiples, est indiqué par les notations suivantes : * $p < 0.05$, ** $p < 0.01$.

Notes	N1	N5	N9	N10	N11	N12	N13	N14	N15	N21
t(52)	7.93**	2.83**	2.44*	2.79**	3.16**	2.41*	4.52**	4.94**	3.94**	2.81**

TAB. 5.6 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour le Centre de Gravité Spectral moyen (SCM). Légende : cf. tableau 5.5.

Notes	N1	N6	N7	N8	N12	N14	N15	N21	N22
t(52)	3.90**	7.93**	4.91**	2.61**	3.44**	3.37**	2.80**	3.30**	2.05*

TAB. 5.7 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour l'ambitus du Centre de Gravité Spectral (SCA). Légende : cf. tableau 5.5.

Notes	N6	N7	N11	N12	N13	N14	N15	N21	N24	N27
t(52)	5.08**	3.31**	4.66**	2.30*	6.65**	7.83**	3.96**	2.38*	3.72**	2.32*

TAB. 5.8 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour le Rapport Impair/Pair moyen (OERM). Légende : cf. tableau 5.5.

Notes	N1	N8	N11	N12	N13	N14	N15
t(52)	2.15*	2.60**	2.82**	6.46**	2.23*	2.87**	3.34**

TAB. 5.9 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour l'ambitus du Rapport Impair/Pair (OERA). Légende : cf. tableau 5.5.

Notes	N5	N6	N11	N12	N13	N16	N17	N18	N19	N22
t(52)	13.25**	7.23**	3.09**	2.25*	2.63**	4.11**	2.44*	2.49*	3.31*	2.88**
Notes	N23	N24	N27							
t(52)	5.26**	13.53**	13.15**							

TAB. 5.10 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour les variations rythmiques (ΔDUR). Légende : cf. tableau 5.5.

Notes	N1	N2	N6	N7	N8	N9	N10	N11	N12
t(52)	6.73**	2.85**	6.27**	3.52**	4.41**	4.49**	3.78**	3.40**	2.52*
Notes	N13	N14	N15	N16	N20	N21	N26		
t(52)	3.47**	3.70**	3.57**	2.42*	1.98*	2.36*	4.70**		

TAB. 5.11 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour l'enveloppe moyenne (ENVM). Légende : cf. tableau 5.5.

Notes	N5	N6	N7	N12	N14	N21	N24	N25	N27
t(52)	2.65**	2.22*	3.50**	2.40*	2.19*	2.08*	2.11*	3.67**	2.11*

TAB. 5.12 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour l'ambitus de l'enveloppe (ENVA). Légende : cf. tableau 5.5.

Notes	N4
t(52)	7.66**

TAB. 5.13 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour la fréquence fondamentale moyenne (F0M). Légende : cf. tableau 5.5.

Rang des notes	N1	N24
t(52)	7.55**	2.03*

TAB. 5.14 – Résultats des tests de comparaisons multiples (méthode de Holm-Sidak) pour l'ambitus de la fréquence fondamentale (F0A). Légende : cf. tableau 5.5.

Hauteur D'après le tableau 5.4, les résultats de l'ANOVA révèlent des effets significatifs de l'interprétation sur la fréquence fondamentale moyenne par note F0M [$F(26;1026)=2.43$, $p<0.001$] et sur l'ambitus de la fréquence fondamentale par note F0A [$F(26;1026)=2.62$, $p<0.001$]. Cependant, l'examen des résultats des tests de comparaisons multiples (cf. tableau 5.5) montre que seules une note pour F0M (N4), et deux pour F0A (N1 et N24) sont à l'origine de ces effets. De plus pour ces notes, les différences de F0M et F0A sont soit dues à des instabilités du descripteur F0 au moment de l'amorce et/ou de la fin de la note (donc un phénomène lié à l'analyse), soit elles sont vraiment très faibles. La différence maximale recensée sur la partie stable du son apparaît pour la première

note (N1) et est d'environ 1 Hertz, ce qui est à la limite de notre seuil de discrimination fréquentiel [Wier *et al.*, 1977]. Ces résultats montrent donc que dans le cadre des interprétations analysées, la hauteur n'est pas un critère de distinction entre les jeux scolaires et expressifs. Au niveau de l'interprète, la recherche de la stabilité de la hauteur à des fins de justesse a très probablement été primordiale.

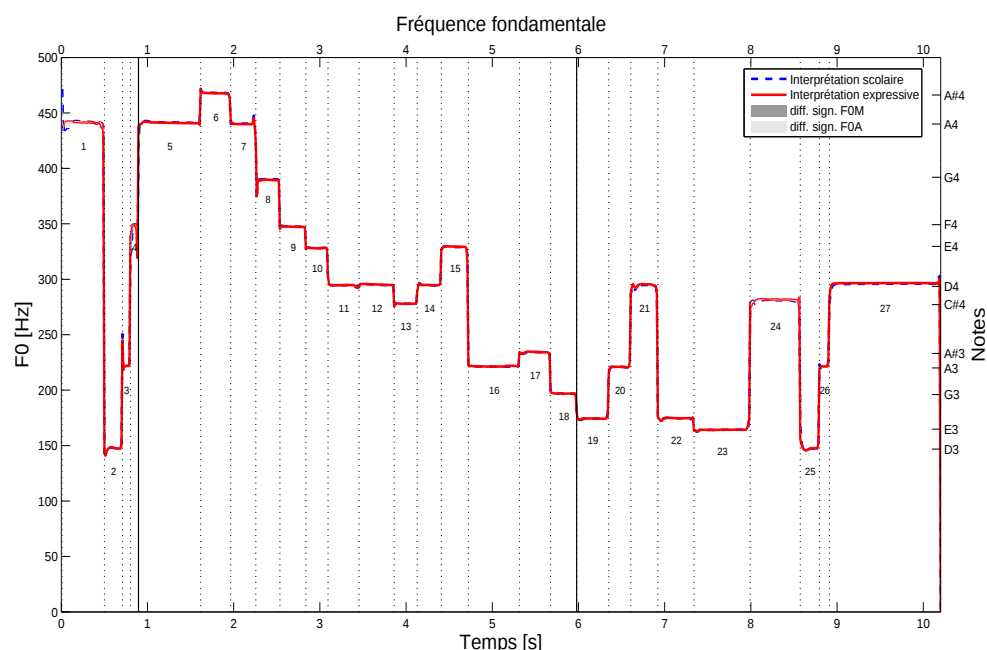


FIG. 5.9 – Fréquence fondamentale (F0) moyenne pour les 20 interprétations scolaires (ligne pointillée) et les 20 interprétations expressives (ligne solide) de l'extrait de Bach. Les lignes verticales pointillées indiquent les amorces et fins de notes. Les lignes verticales solides indiquent les débuts de mesure. Le rang des notes au sein de la structure musicale est également reporté.

Rythme Le type d'interprétation affecte de manière fortement significative le descripteur de variations rythmiques ΔDUR [$F(26;1026)=16.80$, $p<0.001$] (cf. tableau 5.4). En effet, tels que le tableau 5.10 et la figure 5.10 le montrent, un grand nombre de notes sont significativement plus longues pour les interprétations expressives que pour les interprétations scolaires (13 au total). Ceci est cohérent avec les effets signalés précédemment au sujet de la durée globale des séquences plus élevée, et du tempo légèrement plus lent, dans le cas d'un jeu expressif (cf. tableau 5.3). Cependant, il est intéressant de remarquer que la forme des variations rythmiques ΔDUR est similaire pour les deux interprétations. Mêmes les interprétations scolaires ne sont pas le résultat d'une traduction mécanique de la partition, il subsiste de légères variations rythmiques. Ceci explique pourquoi le tempo moyen des interprétations scolaires n'est pas exactement égal au tempo de référence. Le processus de raccourcir ou au contraire de rallonger les notes par rapport à la partition apparaît donc à la fois dans les interprétations scolaires et dans les interprétations expressives. Cependant, son effet est plus prononcé lorsque l'interprétation est expressive. C'est une constatation souvent relatée dans les études sur les variations rythmiques (cf. par ex. [Penel et Drake, 2004]). Par exemple, le *ritardando* final (N23 à N27) est employé dans les deux types d'interprétations, mais de manière amplifiée pour les interprétations expressives.

Timbre L'interaction entre le type d'interprétation et le Temps d'Attaque des notes est significative [$F(26;1026)=2.87$, $p<0.001$] (cf. tableau 5.4). Le tableau 5.5 et la figure 5.11(a) indiquent

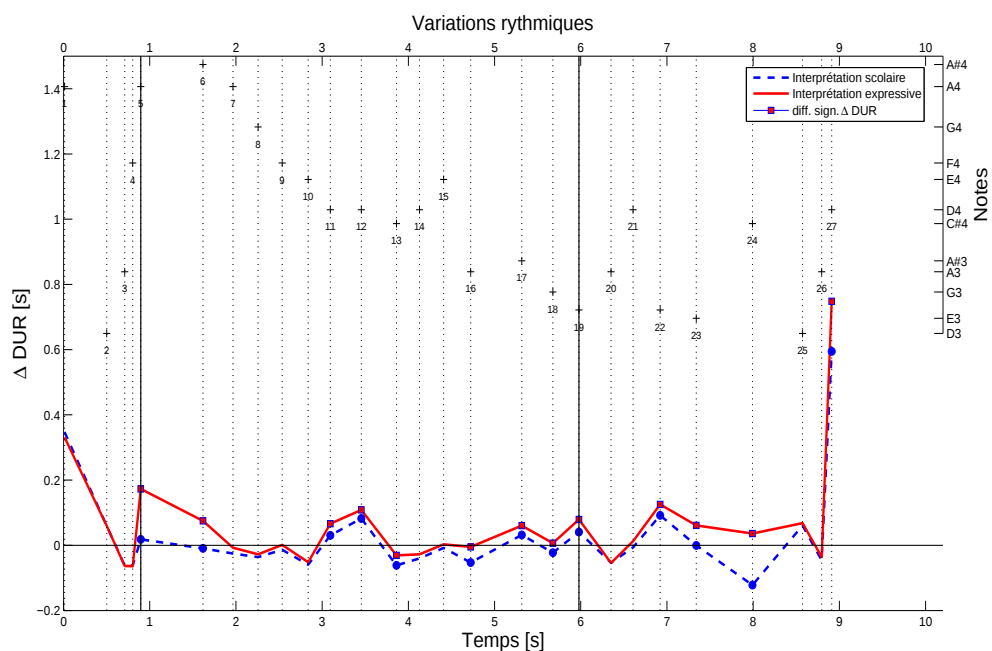
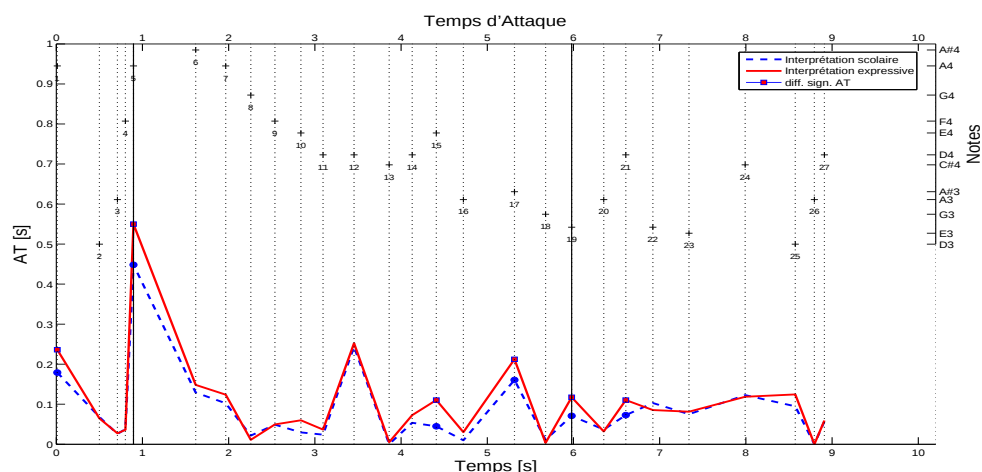


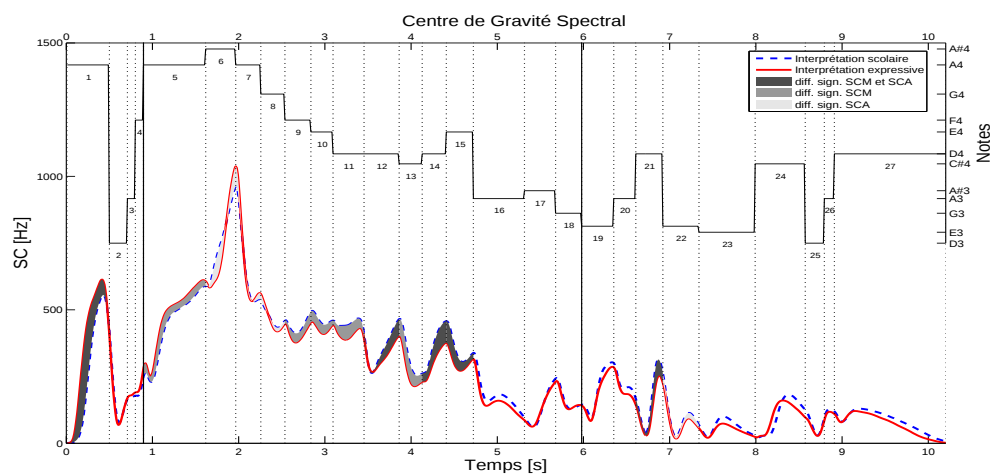
FIG. 5.10 – Variations rythmiques (ΔDUR) moyennes pour les 20 interprétations scolaires (ligne pointillée) et les 20 interprétations expressives (ligne solide) de l'extrait de Bach. Les notes ou groupes de notes pour lesquelles des différences systématiques ont été déterminées par l'ANOVA, sont mises en évidence (marqueurs ronds et carrés).

que plusieurs notes (6 au total) présentent des Temps d'Attaque significativement plus longs pour les interprétations expressives. De fait, le profil de Temps d'Attaque moyen tend à s'accorder sur celui des variations rythmiques : en général, plus les notes sont longues, plus leurs Temps d'Attaque sont élevés. La différence la plus grande, qui a lieu pour la cinquième note (N5), est de 100 ms, ce qui est largement au-dessus de la résolution temporelle du système auditif (environ 5 ms). Cependant, ne perdons pas de vue que le descripteur de Temps d'Attaque employé se fonde sur la détection de seuils relatifs au maximum de l'enveloppe d'amplitude des sons, et est par conséquent très sensible à la forme de l'enveloppe, et notamment à la localisation temporelle du maximum. Si ce dernier arrive à la fin d'une note de longue durée, AT tendra probablement à être plus grand que s'il arrive à la fin d'une note de courte durée. Ceci ne traduit peut-être pas d'une manière fidèle le percept d'attaque, qui pourrait potentiellement rester inchangé malgré une augmentation de la durée d'une note. Comme nous l'avons déjà mentionné dans la section 2.2.5.3.1, il n'existe pas à ce jour de modèle de temps d'attaque perceptif qui puisse rendre compte des nombreuses situations possibles dans un contexte musical. D'autre part, le concept d'attaque est ambigu pour des notes jouées de manière *legato* (d'autant plus que ces dernières sont brèves), comme c'est le cas de certaines notes du début de la phrase de Bach (dans la portion de N2 à N16).

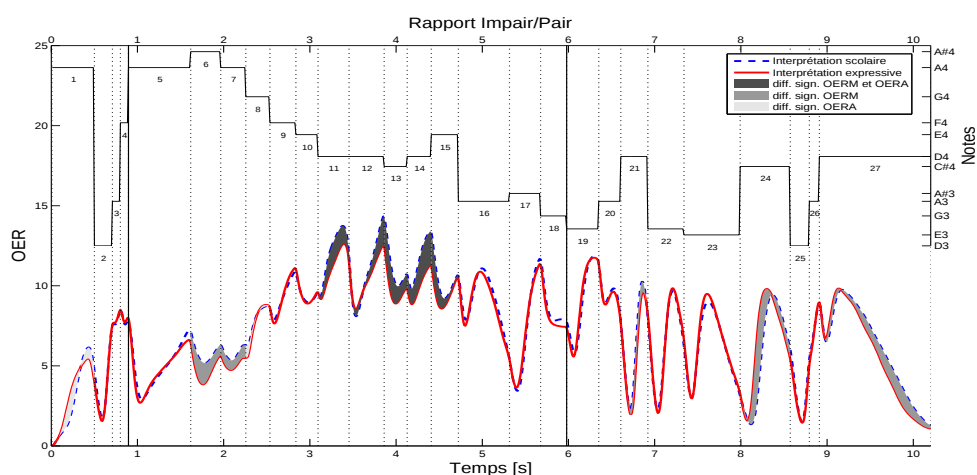
Comme indiqué dans le tableau 5.4, les ANOVA ont révélé des effets significatifs du type d'interprétation pour le Centre de Gravité Spectral moyen SCM [$F(26;1026)=6.33$, $p<0.001$] et l'ambitus du Centre de Gravité Spectral SCA [$F(26;1026)=6.51$, $p<0.001$]. Les tests de comparaisons multiples montrent que 14 des 27 notes de la séquence (soit plus de la moitié) présentent des différences significatives du comportement local du Centre de Gravité Spectral (moyenne et/ou ambitus) lorsque l'intention expressive change (cf. tableaux 5.6 et 5.7, ainsi que la figure 5.11(b)). Par exemple, la cinquième note (N5) de la séquence diffère, d'un jeu scolaire à un jeu expressif, par la valeur moyenne du Centre de Gravité Spectral SCM (mais pas par son ambitus SCA),



(a)



(b)



(c)

FIG. 5.11 – Temps d'Attaque AT (a), Centre de Gravité Spectral SC (b), Rapport Impair/Pair OER (c) moyens pour les 20 interprétations scolaires (ligne pointillée) et les 20 interprétations expressives (ligne solide) de l'extrait de Bach. Les notes ou groupes de notes pour lesquelles des différences systématiques ont été déterminées par l'ANOVA, sont mises en évidence (marqueurs ronds et carrés pour AT, zones grisées pour SC et OER). Les notes de la séquence musicale sont représentées dans la partie supérieure de la figure. Leur rang au sein de la structure musicale est également reporté.

la sixième note (N6) diffère par l'ambitus du Centre de Gravité Spectral SCA (mais pas par sa moyenne SCM), et la première note (N1) diffère à la fois par la valeur moyenne SCM et l'ambitus SCA du descripteur.

Dans certaines parties, le Centre de Gravité Spectral est plus élevé pour les interprétations scolaires que pour les interprétations expressives, dans d'autres parties, le phénomène inverse se présente. Il n'y a donc pas de règle basique qui se dégage du type *jeu expressif* $\Rightarrow$ *sons davantage brillants*, ou au contraire *jeu expressif* $\Rightarrow$ *sons moins brillants*, la nature des changements est subtile et semble dépendre de la structure musicale. La plus grande différence de valeur moyenne se produit pour la première note (N1) de la phrase. Elle est de 245.30 Hz, ce qui, rapporté à la fréquence fondamentale de la note (un A4, 440 Hz), donne une valeur de SC normalisée de 0.558, bien supérieure au seuil de différence moyen de Centre de Gravité Spectral de 0.117 obtenu dans [Kendall et Carterette, 1996]. Un passage entier constitué de 7 notes (N9-N15), est plus brillant pour les interprétations scolaires que pour les interprétations expressives (cf. tableau 5.6). Lorsque l'on écoute les séquences, ce passage semble plus doux et triste dans l'interprétation expressive que dans l'interprétation scolaire (cf. Exemples Sonores ^{nos} 41 et 43, par exemple).

Notons que les notes brèves (par ex. les appoggiatures N2, N3, N26) ne présentent pas de différences significatives de Centre de Gravité Spectral. Ceci tend à conforter l'hypothèse que lorsque les notes sont trop courtes, l'interprète ne dispose pas suffisamment de temps pour contrôler le timbre de manière expressive.

Comme prévu, le Centre de Gravité Spectral tend globalement à suivre les variations de hauteur. Mais le fait qu'aucune différence significative de fréquence fondamentale pertinente n'ait été constatée confirme que les différences significatives de Centre de Gravité Spectral sont bien dues à des variations de la forme de la distribution spectrale indépendantes de variations de fréquences des composantes. Il est d'ailleurs intéressant de constater qu'au sein d'une même phrase musicale des notes de même hauteur peuvent aussi avoir des valeurs de Centre de Gravité Spectraux très différentes. Par exemple, pour le groupe d'interprétations scolaires, les notes N11, N12, et N27 qui correspondent toutes à un D4 ($f_0 \approx 293.67$ Hz) ont des valeurs moyennes SCM (respectivement 448.93, 355.72, et 82.07 Hz)¹⁵ significativement différentes : N11 vs N12 [$t(351)=7.74$, $p<0.001$], N11 vs N27 [$t(351)=30.46$, $p<0.001$], et N12 vs N27 [$t(351)=21.45$, $p<0.001$]¹⁶. La forme du profil de variation du Centre de Gravité Spectral traduit bien ces différences. Ceci tend à confirmer que les variations de timbre-qualité dépendent de la position des notes dans la structure musicale.

Certaines notes présentent à la fois des différences systématiques de Temps d'Attaque et de Centre de Gravité Spectral. Par exemple, la première note de la première mesure (N5), située après l'anacrouse (N1) et les appoggiatures (N2 à N4), dure plus longtemps, a un Temps d'Attaque plus long et un Centre de Gravité Spectral plus élevé dans l'interprétation expressive. Cette note joue en l'occurrence un rôle important au sein de la structure musicale puisqu'elle commence la phrase musicale. Nous suggérons que le timbre et les variations rythmiques sont utilisés afin d'en accentuer l'importance, d'augmenter localement la tension musicale.

Comme pour le Temps d'Attaque et le Centre de Gravité Spectral, il y a une interaction entre l'intention expressive et le Rapport d'énergie Impair/Pair (au niveau de sa moyenne OERM [$F(26;1026)=6.47$, $p<0.001$], et au niveau de son ambitus OERA [$F(26;1026)=3.97$, $p<0.001$], cf.

¹⁵Le fait que la valeur de SCM puisse être inférieure à celle de f_0 est lié au terme stabilisateur b_0 de l'équation 2.22 qui tend globalement à abaisser la valeur du descripteur SC. Rappelons que ceci n'est pas gênant puisque l'on s'intéresse ici davantage aux valeurs relatives de SC (comparaisons).

¹⁶Ces résultats proviennent de tests de comparaisons multiples (méthode de Holm-Sidak) réalisés pour toutes les combinaisons 2 à 2 de notes à l'intérieur du même groupe d'interprétations.

tableau 5.4). Les tableaux 5.8 et 5.9 et la figure 5.11(c) permettent de localiser quelles sont les notes à la source de ces effets. 12 notes, parmi les 27 de la séquence musicale, présentent des différences significatives de Rapport Impair/Pair (moyenne OERM et/ou ambitus OERA). Pour certaines notes, les changements interviennent à la fois pour le Centre de Gravité Spectral et le Rapport Impair/Pair (le passage de N11 à N15 par exemple entre SCM et OERM). Cependant, les différences ne sont pas toujours synchrones avec celles du Centre de Gravité Spectral (par ex., les notes N6, N7, N24, N27 présentent des différences significatives de OERM, mais pas de SCM). Notons que le Rapport Impair/Pair est la plupart du temps supérieur à 1 (hormis pendant la phase d'attaque de la première note), ce qui traduit, comme prévu, la dominance des harmoniques impairs sur les harmoniques pairs pour les sons de clarinette.

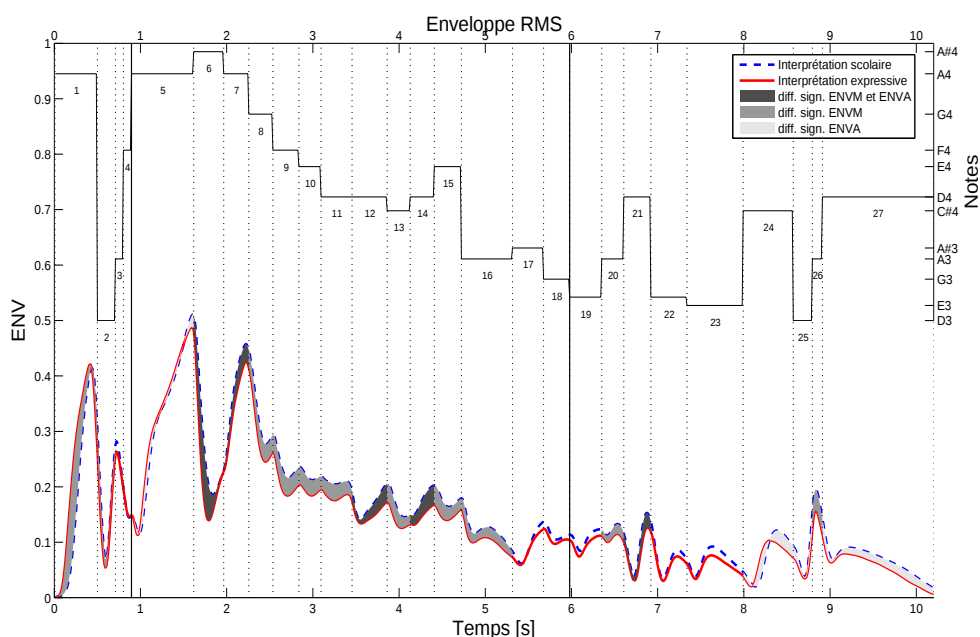


FIG. 5.12 – Enveloppe d'amplitude (ENV) moyenne pour les 20 interprétations scolaires (ligne pointillée) et les 20 interprétations expressives (ligne solide) de l'extrait de Bach.

Dynamique D'après le tableau 5.4, l'effet du type d'interprétation sur la moyenne ENVM de l'enveloppe RMS est significatif [$F(26;1026)=5.62, p<0.001$]. Il ne l'est que modérément pour son ambitus ENVA [$F(26;1026)=1.64, p<0.05$]. Les résultats des analyses *post hoc* rapportés dans les tableaux 5.11 et 5.12, ainsi que sur la figure 5.12, indiquent que 20 notes sur les 27 de la séquence présentent des différences significatives d'amplitude RMS moyenne ou d'ambitus d'amplitude. Ces changements sont souvent synchrones avec ceux du Centre de Gravité Spectral. Remarquons que pour les deux types d'interprétation, un *decrescendo* progressif a lieu à partir de 3 s jusqu'à la fin de la phrase (cf. figure 5.12). Examinons maintenant les corrélations entre les variations de timbre et la dynamique.

Corrélations entre timbre et dynamique La figure 5.13 montre les profils temporels moyens de Centre de Gravité Spectral, Rapport Impair/Pair et enveloppe RMS, issus des 20 interprétations expressives. Si le Centre de Gravité Spectral est globalement bien corrélé avec l'enveloppe RMS [$r^2(875) = 0.68, p < .0001$], le Rapport Impair/Pair présente un comportement plus complexe. En effet la corrélation globale linéaire entre OER et ENV est très faible [$r^2(875) = 0.007, p <$

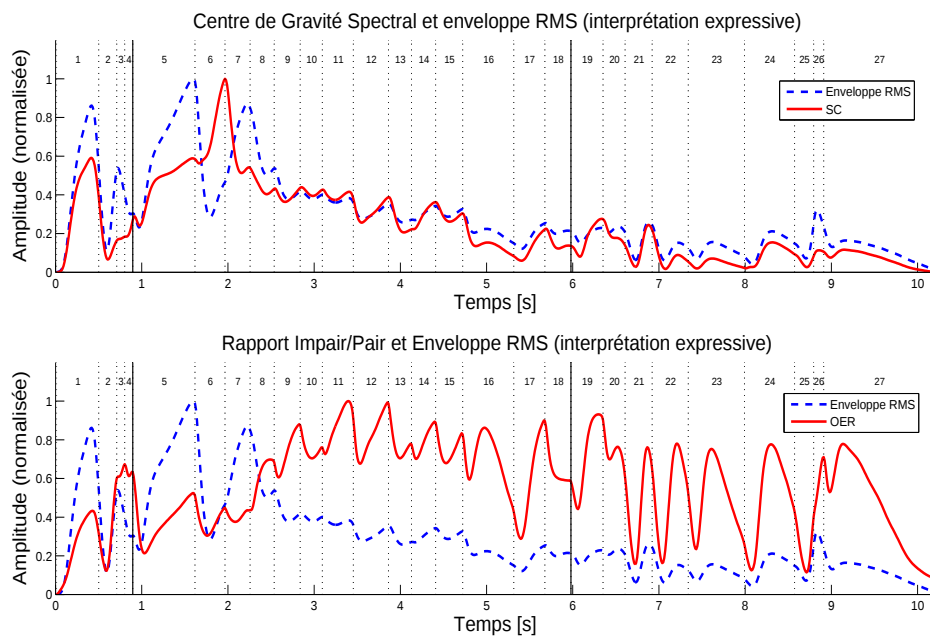


FIG. 5.13 – Comparaison du Centre de Gravité Spectral (SC) et du Rapport Impair/Pair (OER) avec l'enveloppe d'amplitude RMS (ENV) pour les interprétations expressives de l'extrait de Bach. SC, OER et ENV ont été normalisés.

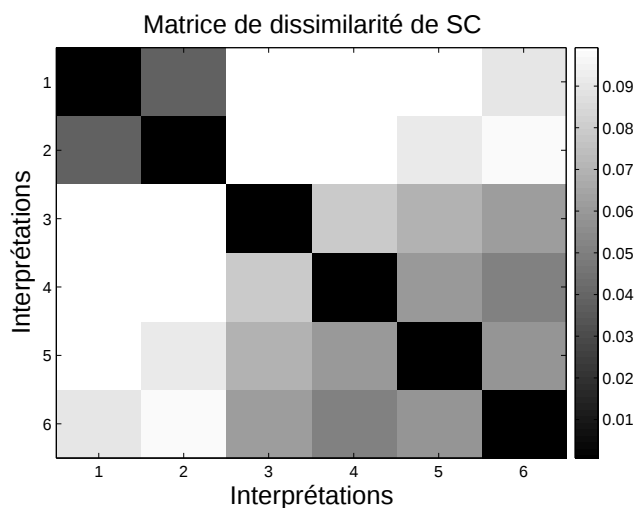
0.05]. Lors du *decrescendo*, SC décroît en phase avec ENV. En revanche, OER présente un comportement différent. Pour certaines notes, les corrélations sont élevées et pour d'autres elles sont faibles. La valeur globale du Rapport Impair/Pair devient notamment plus élevée lorsque l'énergie acoustique augmente. D'après la physique de l'instrument, lorsque le niveau de jeu est faible, les sons de clarinette sont quasiment sinusoïdaux et contiennent très peu d'harmoniques pairs. Ceci explique probablement l'augmentation de OER lors du *decrescendo*. Ce comportement non linéaire du descripteur par rapport à l'énergie acoustique explique pourquoi la corrélation linéaire avec l'amplitude RMS est faible.

5.7.2 Interprétations de l'extrait de Mozart

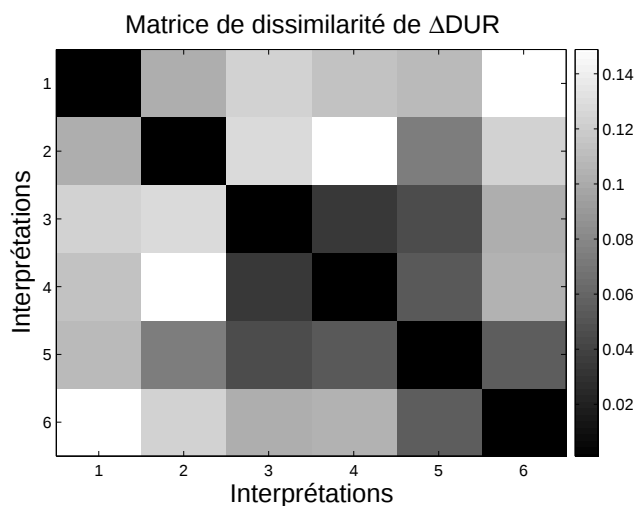
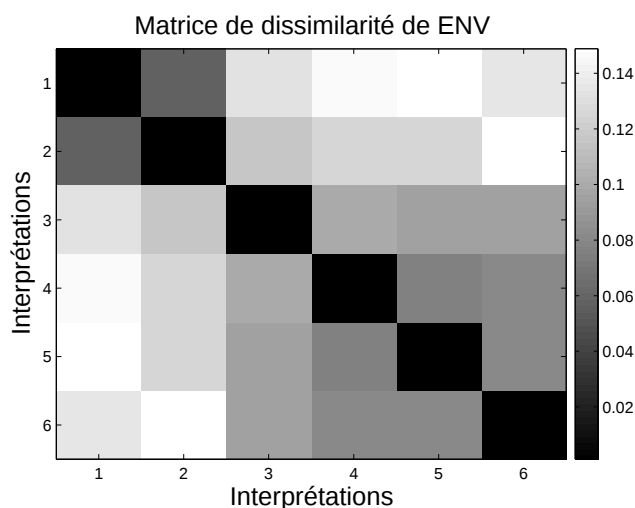
5.7.2.1 Matrices de dissimilarité

Les matrices de dissimilarité du Centre de Gravité Spectral (SC), des variations rythmiques (ΔDUR) et de l'enveloppe d'amplitude (ENV) sont représentées sur les figures 5.14(a) à 5.14(c). Les représentations en dendrogramme des classifications ascendantes hiérarchiques correspondantes sont données dans les figures F.3(a) à F.3(c).

Les interprétations scolaires (P1 et P2) diffèrent des interprétations expressives (P3 à P6), de par leurs variations de Centre de Gravité Spectral et par leurs variations de Rapport Impair/Pair. Les variations rythmiques semblent discriminer moins nettement les différents types d'interprétations que pour l'extrait de Bach. Cependant, les variations rythmiques des interprétations expressives P3 à P5 sont plus proches entre elles qu'avec celles des interprétations scolaires (P1 et P2).



(a) Matrice de dissimilarité SC

(b) Matrice de dissimilarité ΔDUR 

(c) Matrice de dissimilarité ENV

FIG. 5.14 – Matrice de dissimilarité des variations de Centre de Gravité Spectral (SC) (a), des variations rythmiques (b), et des variations d'énergie acoustique (c) pour les 6 interprétations de l'extrait de Mozart. Les 2 premières interprétations (P1 à P2) correspondent aux interprétations scolaires. Les 4 interprétations suivantes (P3 à P6) correspondent aux interprétations expressives. Plus les points sont foncés, moins les dissimilarités sont élevées.

5.7.2.2 Comparaison de la moyenne des descripteurs

La moyenne des différents descripteurs a été calculée en fonction de chaque type d'interprétation (P1 et P2 d'une part, et P3 à P6 d'autre part), en vue de comparer les paramètres de hauteur, de rythme, de timbre et de dynamique.

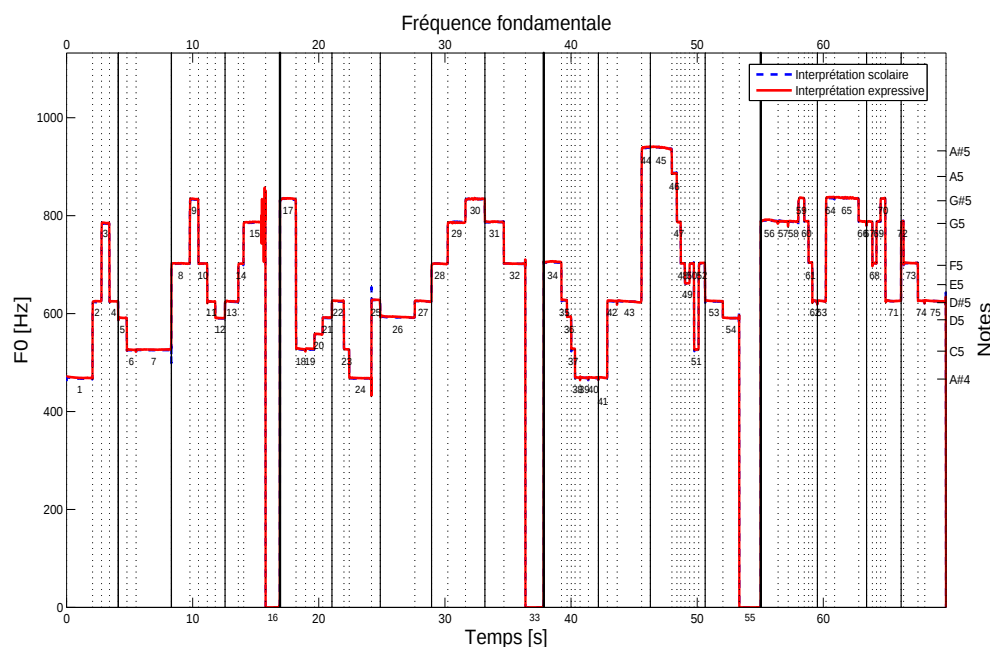


FIG. 5.15 – Fréquence fondamentale (F0) moyenne pour les 2 interprétations scolaires (ligne pointillée) et les 4 interprétations expressives (ligne solide) de l'extrait de Mozart. Les lignes verticales pointillées indiquent les amorces et fins de notes. Les lignes verticales solides indiquent les débuts de mesure. Le rang des notes au sein de la structure musicale est également reporté.

Hauteur Comme pour l'extrait de Bach, les fréquences fondamentales instantanées des interprétations scolaires et expressives sont quasiment identiques (cf. figure 5.15).

Rythme Les variations rythmiques présentent en revanche des différences de comportement (cf. figure 5.16). A un niveau global, elles tendent à être plus prononcées pour les interprétations expressives que pour les interprétations scolaires, comme pour l'extrait de Bach. En effet, comme le montre le tableau 5.15, la durée moyenne des interprétations expressives dépasse d'environ 5 s celle des interprétations scolaires ; le tempo moyen est également plus faible. Pour l'extrait de Bach, les différences de variations rythmiques entre les jeux scolaires et expressifs reposaient essentiellement sur l'ampleur des écarts par rapport à la référence (partition). Les distinctions sont ici plus complexes. Dans certains cas, les profils de variations rythmiques scolaires et expressifs ne se distinguent que par l'ampleur des variations (cf. mesures 3 à 5), dans d'autres les variations peuvent être en sens opposé (cf. mesure 2 et mesure 9).

Timbre Les Temps d'Attaque des notes sont généralement plus longs dans la version expressive, comme pour l'extrait de Bach (cf. figure 5.17(a)).

Les variations temporelles du Centre de Gravité Spectral et du Rapport Impair/Pair présentent de nettes différences entre les jeux scolaires et expressifs (figures 5.17(a) et 5.17(c)) pour des notes

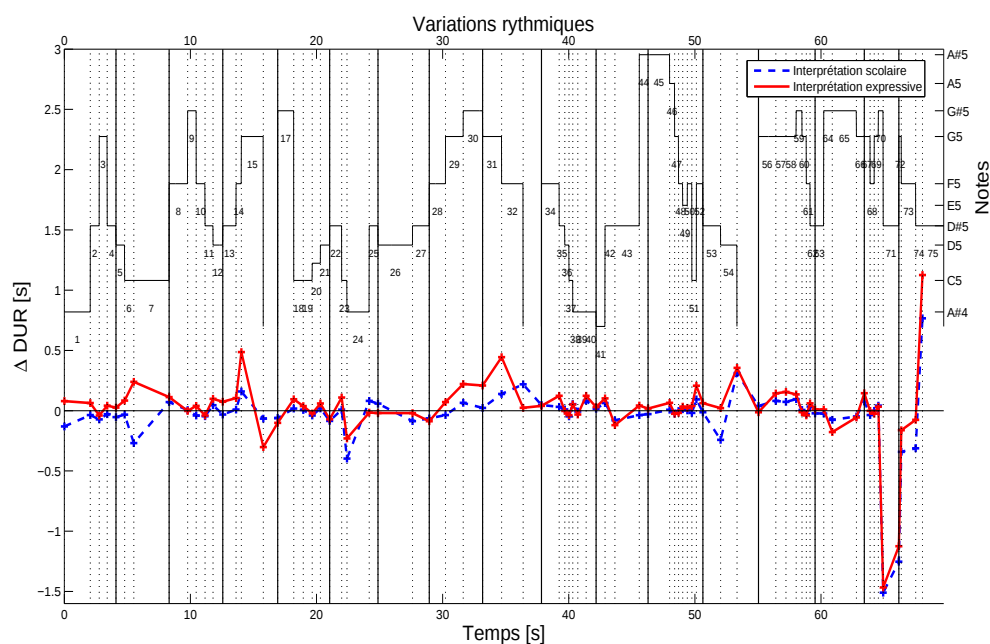


FIG. 5.16 – Variations rythmiques (ΔDUR) moyennes pour les 2 interprétations scolaires (ligne pointillée) et les 4 interprétations expressives (ligne solide) de l'extrait de Mozart.

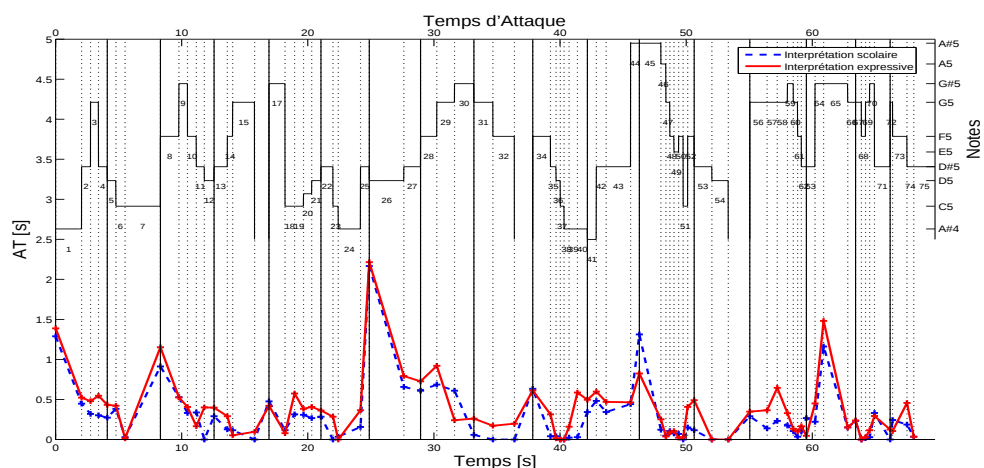
	Interprétations scolaires	Interprétations expressives
Durées (moyenne, écart-type) en s	67.80 (0.64)	72.46 (0.89)
Tempo (moyenne, écart-type) en bpm	45.13 (0.43)	42.23 (0.52)

TAB. 5.15 – Moyennes et écarts-types des durées et tempi des interprétations scolaires et expressives de l'extrait de Mozart. Les durées sont exprimées en secondes, les tempi en nombre de temps par minutes.

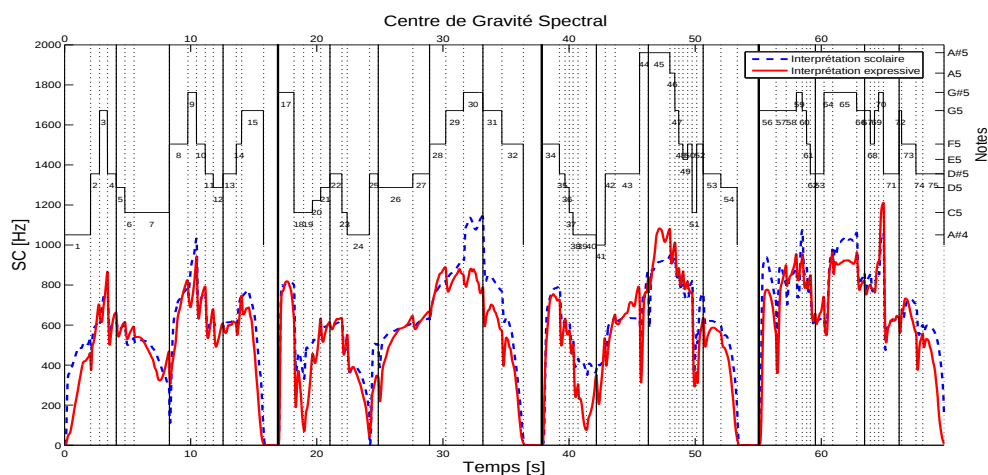
spécifiques ou des passages entiers. Par exemple, les évolutions temporelles de SC et OER sont très différentes pour la première note de la séquence. Dans la seconde phrase musicale de l'extrait, tout un passage qui s'étend de la 7^e à la 9^e mesure (N25 à N32) présente une forme temporelle de SC différente selon l'intention expressive : SC s'avère être moins élevé dans la version expressive sur quasiment tout le passage. Dans la troisième phrase musicale (mesure 12), c'est l'inverse qui apparaît, SC est plus élevé dans la version expressive. Ces différences de valeurs de SC sont parfois élevées. La plus grande différence de SC, qui apparaît pour la note N56 de la quatrième phrase musicale, est d'environ 420 Hz, ce qui correspond à un SC normalisé d'environ 0.54, bien au dessus du seuil de perception de différence de SC moyen de 0.117 (cf. [Kendall et Carterette, 1996]).

Dynamique La dynamique n'est également pas la même pour les interprétations scolaires et expressives (cf. figure 5.18). Il semble qu'il y ait plus de nuances d'intensité dans les versions expressives, les versions scolaires étant jouées légèrement plus fort et avec moins de variations entre les notes.

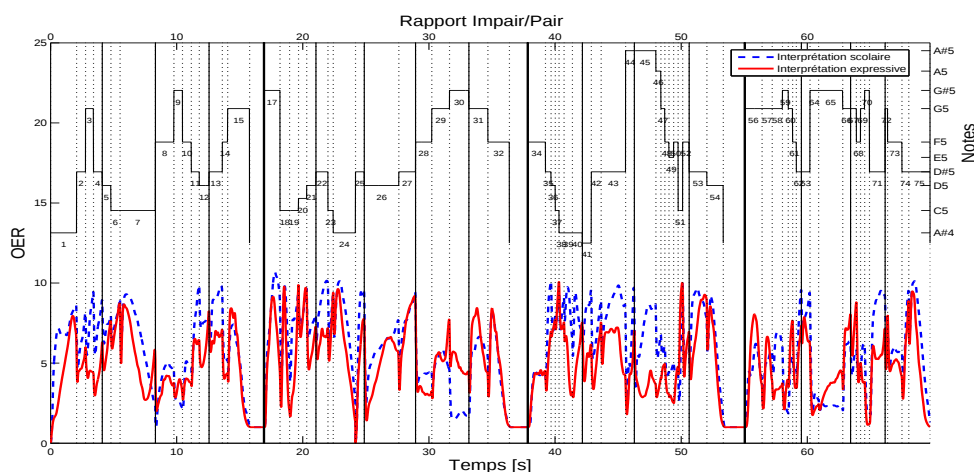
Corrélations entre timbre et dynamique Comme pour l'extrait de Bach, la corrélation linéaire entre le Centre de Gravité Spectral et l'enveloppe RMS est élevée quelque soit le type d'interprétation ($[r^2(6002) = 0.79, p < 0.0001]$ pour l'interprétation scolaire et $[r^2(6002) = 0.85, p <$



(a)



(b)



(c)

FIG. 5.17 – Temps d'Attaque AT (a), Centre de Gravité Spectral SC (b), Rapport Impair/Pair OER (c) moyens pour les 2 interprétations scolaires (ligne pointillée) et les 4 interprétations expressives (ligne solide) de l'extrait de Mozart.

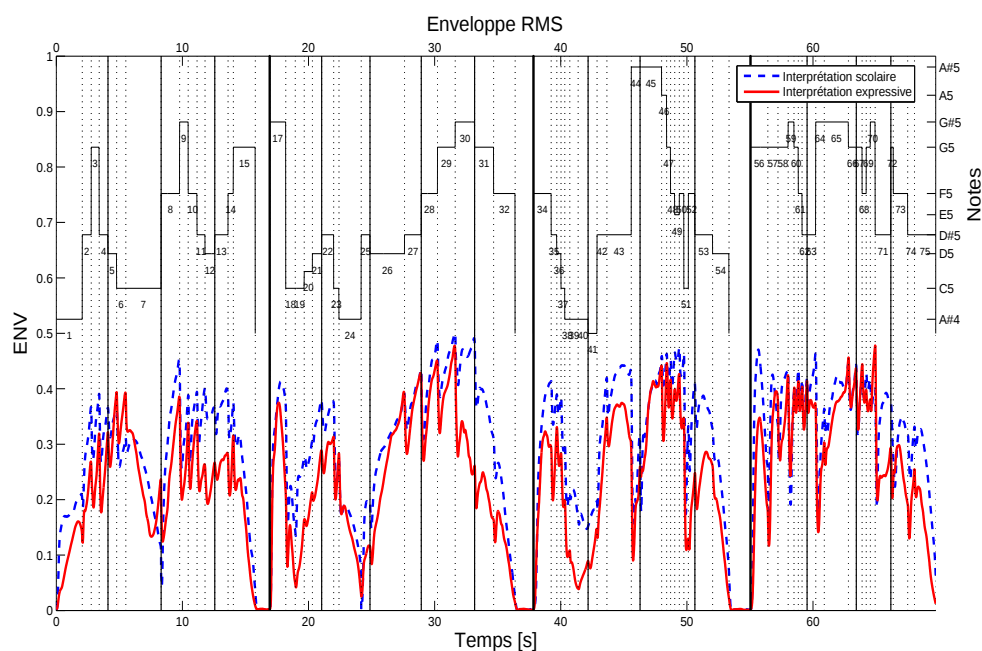


FIG. 5.18 – Enveloppe d'amplitude (ENV) moyenne pour les 2 interprétations scolaires (ligne pointillée) et les 4 interprétations expressives (ligne solide) de l'extrait de Mozart.

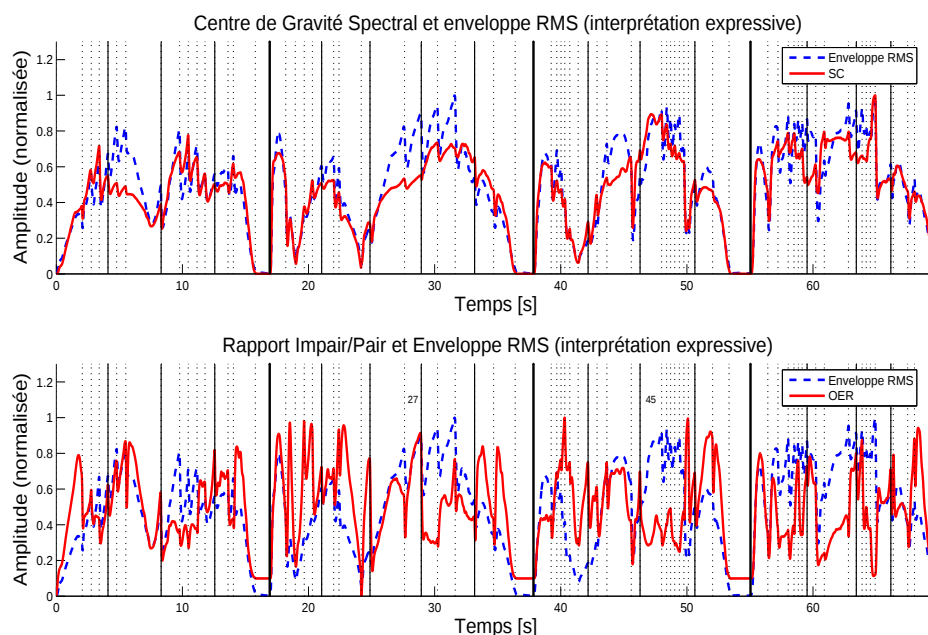


FIG. 5.19 – Comparaison du Centre de Gravité Spectral (SC) et du Rapport Impair/Pair (OER) avec l'enveloppe d'amplitude RMS (ENV) pour les interprétations expressives de l'extrait de Mozart. SC, OER et ENV ont été normalisés.

0.0001] pour l'interprétation expressive). La corrélation entre le Rapport Impair/Pair et l'enveloppe RMS ne l'est en revanche pas ($[r^2(6002) = 0.16, p < .0001]$ pour l'interprétation scolaire et $[r^2(6002) = 0.17, p < .0001]$ pour l'interprétation expressive). La figure 5.19 permet de

constater que la relation entre l'énergie acoustique et le Rapport Impair/Pair est non linéaire. Dans certaines parties, OER est bien corrélé avec l'enveloppe (par ex. note N27 de la seconde phrase), dans d'autres il présente des variations opposées (par ex. note N45 de la troisième phrase).

5.8 Discussion

Ces analyses ont permis de comparer le comportement de paramètres liés au timbre, au rythme, à la dynamique et à la hauteur pour des séquences de clarinette jouées de manière neutre (scolaire) ou bien expressive. Lorsque l'intention musicale est inchangée, des comportements très réguliers sont observés pour certains paramètres liés au timbre (Centre de Gravité Spectral, et parfois, Rapport Impair/Pair et Temps d'Attaque), au rythme, à la dynamique et à la hauteur. Ceci a déjà été constaté pour les variations rythmiques et la dynamique (par ex. [Palmer et Hutchins, 2006, p. 248]). Cette étude montre qu'il en est de même pour le timbre ce qui, à notre connaissance, n'a pas été reporté dans la littérature.

Si les interprètes (professionnels) reproduisent fidèlement certaines variations des variables acoustiques (timbre, durée, intensité, etc.) qui ne découlent pas directement des notations de l'œuvre écrite, il est fort probable que les choix effectués soient liés à la spécificité de l'interprétation et déterminent la façon dont le musicien exprime son intention. Ainsi, le fait que les mêmes variations de timbre apparaissent de manière systématique tend à suggérer qu'elles sont liées à l'intention musicale de l'interprète. Ceci serait démontré si différentes variations de timbre apparaissent lorsque l'intention musicale change. C'est justement ce que nous avons observé. Les diverses analyses statistiques employées (classification ascendante hiérarchique, analyse de variance) prouvent que des différences de timbre systématiques sont observables lorsque l'on compare des interprétations jouées de manière scolaire, et des interprétations jouées de manière expressive. Ces changements ont été constatés aussi bien pour des notes spécifiques (par ex. la note d'ouverture du mouvement) que des passages entiers de phrases musicales. Des différences significatives de Centre Gravité Spectral, et/ou Rapport Impair/Pair, et/ou Temps d'Attaque ont été constatées pour 18 notes sur les 27 de la séquence musicale de Bach. La nature des variations de timbre semble dépendre de la position des notes au sein de la structure musicale, et de leur durée. Les notes de durée très courtes n'ont en général pas présenté de différences significatives entre les interprétations. Il semble donc que les variations de timbre soient utilisées essentiellement sur des notes d'une durée suffisamment longue, pour lesquelles la partie tenue peut être façonnée au cours du temps. Ceci paraît cohérent du point de vue du contrôle de l'instrument. Il est d'ailleurs intéressant de remarquer que les interprétations expressives ont systématiquement été jouées plus lentement que les interprétations scolaires (tempo moyen de 44.45 / 40.47 *bpm* pour l'extrait de Bach, et 45.13 / 42.13 *bpm* pour l'extrait de Mozart). Ceci s'explique bien sûr du fait que les variations rythmiques sont plus prononcées (certaines notes sont très rallongées), mais peut-être est-ce là aussi un moyen pour l'interprète de disposer de plus de temps afin de contrôler le timbre de certaines notes. Pour l'extrait de Mozart, qui présente beaucoup plus de longues notes (blanches, noires, noires pointées) que pour l'extrait de Bach (essentiellement des doubles-croches), les différences de variations de Centre de Gravité Spectral et de Rapport Impair/Pair entre les jeux scolaire et expressif semblent encore plus prononcées.

Certains profils de variations de Centre de Gravité Spectral apparaissent de manière récurrente. Par exemple, les notes N1, N6, N12, et N14 de l'extrait de Bach présentent un profil d'évolution croissant au cours de la progression de la note (cf. figure 5.11(b)). Nous appellerons ce type de tendance d'évolution 'Faible Elevé' (FE). A l'opposé, les notes N16 et N27 présente une évolution du Centre de Gravité Spectral décroissante, que nous appellerons 'Elevé Faible' (EF). Les notes N22-24 présentent des phases pour lesquelles SC croît puis décroît successivement, ce qui sera noté 'Faible Elevé Faible' (FEF). La figure 5.20 présente des exemples d'évolutions typiques du

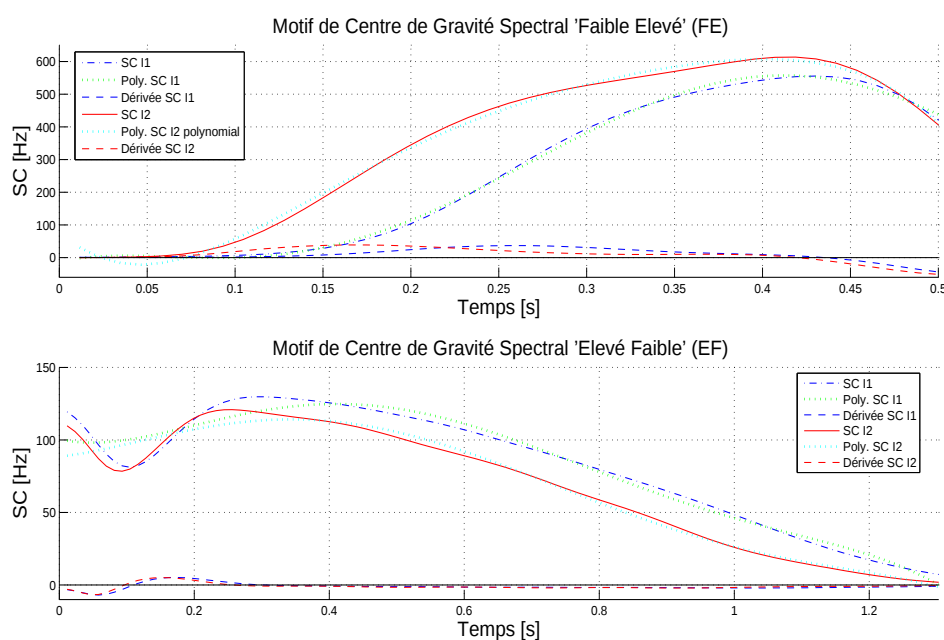


FIG. 5.20 – Exemples de profils d'évolution du Centre de Gravité Spectral 'Faible Elevé' (haut) et 'Elevé Faible' (bas) pour les interprétations scolaires (I1) et expressives (I2) de l'extrait de Bach. Le profil 'Faible Elevé' correspond à la note N1, tandis que le profil 'Elevé Faible' correspond à la note N27. Les dérivées de SC ainsi que des polynômes d'ordre 5 modélisant les courbes sont également présentés.

Centre de Gravité Spectral pour les deux types d'interprétations. Un profil 'Faible Elevé' correspondant à la première note de la séquence est présenté dans le haut de la figure, tandis qu'un profil 'Elevé Faible' correspondant à la dernière note de la séquence est présenté dans le bas de la figure. Les formes d'évolutions du profil FE ne sont pas les mêmes selon l'interprétation. La pente est plus élevée au début de la note pour la version expressive, ce qui se traduit par une dérivée du Centre de Gravité Spectral plus grande durant la première partie de la note. La forme du profil EF est en revanche similaire pour les deux interprétations (les différences de moyenne et d'ambitus ne sont pas significatives, cf. tableaux 5.6 et 5.7). Il est intéressant de noter que des polynômes d'ordre relativement faibles (ordre 5) modélisent fidèlement le comportement du descripteur SC (hormis dans certaines portions où ils deviennent négatifs).

Parmi les trois paramètres de timbre étudiés, le Centre de Gravité Spectral est le descripteur pour lequel le plus grand nombre de différences systématiques a été constaté (14 notes sur les 27 de l'extrait de Bach). Il est suivi par le Rapport Impair/Pair (12 notes) et le Temps d'Attaque (6 notes). Ces résultats ont compté parmi les facteurs qui nous ont amenés à nous concentrer par la suite sur le Centre de Gravité Spectral afin d'analyser, selon une procédure d'analyse par la synthèse, certains effets perceptifs liés aux variations de brillance. En effet, comme nous l'avons déjà mentionné, nous recherchons des paramètres de timbre généraux, qui ne soient pas trop spécifiques à un instrument (comme le Rapport Impair/Pair, qui est spécifique à la clarinette), mais dont l'importance au niveau de l'expression musicale pourrait éventuellement s'étendre à d'autres instruments (la flûte traversière par exemple, dont les sons présentent une irrégularité spectrale bien moindre que ceux de la clarinette [Ystad, 1998]).

Cette étude s'est reposée sur des musiques et modes de jeux traditionnels. Il serait intéressant d'explorer la fonction du timbre comme vecteur d'expression musicale pour d'autres musiques,

le Jazz par exemple, pour lequel le contrôle du timbre semble revêtir un rôle prépondérant, notamment pour le soliste qui improvise. Le timbre semble parfois dépasser l'identité instrumentale pour aller chercher l'identité même de l'instrumentiste, identifiable parmi tant d'autres, et en l'espace de quelques notes seulement (on pensera aux sons inimitables de John Coltrane, ou de Miles Davis). L'emploi en musique contemporaine de techniques de jeu complexes défiant l'acoustique de l'instrument marque également la volonté d'enrichir les palettes traditionnelles de timbres des instruments par de nouveaux timbres dont les formes façonnent le temps (cf. [Exemple Sonore n° 51](#) : effet *Flutterzhunge* à la clarinette).

Chapitre 6

Influence perceptive des variations de timbre

Certains des travaux exposés dans ce chapitre sont présentés dans l'article [Barthet *et al.*, 2008] et dans l'article [Barthet *et al.*, 2009b].

6.1 Introduction

Les analyses sonologiques d'interprétations musicales de clarinette décrites au chapitre précédent ont montré que les variations de timbre, de rythme et d'intensité produites par un interprète dépendaient de son intention expressive. Il semble alors intéressant de se demander comment ces variations affectent les jugements musicaux de l'auditeur. Dans ce chapitre, nous décrivons deux expériences perceptives établies dans le but de mieux comprendre l'influence des variations temporelles de timbre sur la perception musicale.

La démarche employée repose sur le paradigme d'analyse par la synthèse proposé par Risset et Wessel [Risset et Wessel, 1999] et appliqué par les auteurs à l'exploration du timbre (cf. section 1.1.3.3). Cette démarche se déploie en quatre phases : l'obtention d'une représentation d'un signal (analyse), la transformation de cette représentation, l'inversion de la représentation modifiée (synthèse), et l'analyse des effets de la transformation au niveau perceptif. L'évaluation des effets perceptifs permet alors de comprendre l'influence des paramètres qui en sont les causes.

Les deux expériences réalisées pour évaluer l'importance des variations temporelles de timbre sont complémentaires. La première expérience vise à analyser les effets induits par l'élimination des variations de timbre initialement contenues dans une interprétation musicale expressive. La deuxième expérience se rapporte au processus inverse, c'est-à-dire analyser les effets induits par l'ajout de variations de timbre issues d'une interprétation expressive à une interprétation inexpressive.

Comme nous avons pu le voir au cours des chapitres précédents, le timbre est un attribut complexe multidimensionnel du son. Nous nous sommes concentrés ici sur la brillance pour les différentes raisons que nous avons déjà mentionnées : son caractère général (le Centre de Gravité Spectral apparaît de manière récurrente comme un bon corrélat des espaces de timbre perceptif), son lien potentiel avec la notion de tension musicale, et comme nous l'avons vu au chapitre précédent, le fait que cette dimension du timbre semble être employée par les interprètes pour transmettre l'expression musicale.

6.2 Expérience 1 : d'un jeu expressif vers un jeu inexpressif

Dans cette expérience, nous évaluons l'influence de réductions des variations de timbre, de rythme, et d'intensité sur les préférences musicales d'auditeurs. Des transformations sonores visant à modifier chacun de ces paramètres de manière indépendante ou combinée ont été établies. La première partie de cette section vise à décrire ces transformations sonores et leur application à l'élaboration de stimuli. Nous présenterons ensuite le protocole de l'expérience perceptive, ainsi que les méthodes statistiques utilisées pour en analyser les données. Les résultats obtenus à l'issue de l'expérience et des analyses statistiques seront ensuite exposés puis discutés.

6.2.1 Stimuli

6.2.1.1 Corpus sonore

Comme nous le verrons plus en détail à la section 6.2.4, la procédure du test perceptif repose sur une comparaison par paires de différentes séquences musicales. Nous avons donc sélectionné des séquences musicales suffisamment longues pour qu'elles aient un sens musical, mais suffisamment courtes pour que la tâche de comparaison par paires ne devienne pas trop compliquée et que la durée de l'expérience reste raisonnable. Dans [Gabrielsson et Lindstrom, 1985], les auteurs mentionnent avoir employé avec succès des séquences d'une durée de 15 s dans une tâche de comparaison par paires, après avoir constaté que des séquences de 30 s étaient inappropriées pour une telle tâche.

Les séquences de référence choisies afin d'effectuer les transformations sonores proviennent des interprétations expressives des extraits de Bach et Mozart, analysées au chapitre 5. Les premières phrases musicales des enregistrements ont été sélectionnées de manière à disposer de séquences de durées d'environ 10 à 15 s.

6.2.1.2 Transformations sonores

Trois transformations ont été définies afin de modifier de manière indépendante le timbre, le rythme et la dynamique. Afin de réaliser les transformations de timbre et de rythme, nous nous sommes reposés sur le modèle paramétrique d'analyse-synthèse SAS, décrit à la section 2.2.5.5. Rappelons que ce dernier décompose un signal quasi-harmonique en une partie déterministe quasi-périodique, et une partie bruitée large-bande, appelée *résidu*. Ce modèle d'analyse-synthèse est particulièrement adapté pour les transformations sonores car ses paramètres, les amplitudes et fréquences instantanées des composantes harmoniques, peuvent être contrôlées de manière individuelle. Lorsque les paramètres d'analyse sont utilisés pour la synthèse sans avoir subis de transformations, ce modèle permet une reconstruction identique des séquences initiales. Le résidu est obtenu par soustraction du signal originel et du signal synthétisé sans modifications. Les Exemples Sonores ^{no} 52, 53, et 54 correspondent respectivement à une séquence musicale de clarinette (première phrase de l'extrait de Mozart), la resynthèse de la séquence, réalisée sans transformations, et le résidu seul. Notons que comme prévu, le résidu contient essentiellement des événements large bande, i.e. le bruit de souffle et certains bruits de clefs, qui apparaissent lors des changements de notes.

Afin de s'affranchir de phénomènes liés à la hauteur, les fréquences instantanées des composantes ont été gelées au cours du temps dans des rapports de fréquences entiers par rapport à la fréquence fondamentale moyenne. Dans le cas des séquences analysées, pour lesquelles, comme nous l'avons vu au chapitre 5, les variations de fréquence fondamentale instantanée sont très faibles, le gel des fréquences instantanées n'entraîne pas d'effets nuisibles audibles (cf. Exemple Sonore ^{no}

55 : resynthèse avec fréquences instantanées gelées au cours du temps).

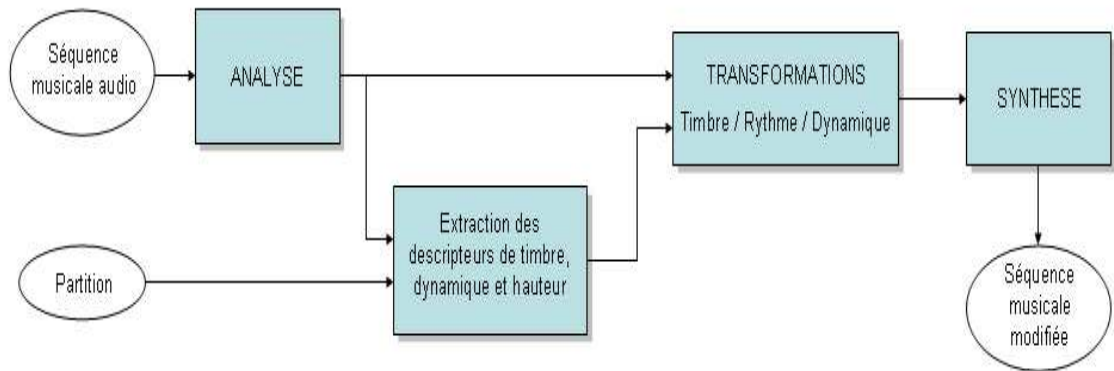


FIG. 6.1 – Méthodologie de transformation de l'expression musicale.

Une fois le modèle établi, l'élaboration de transformations est possible. Le schéma de la figure 6.1 résume le principe de transformation des variables liées à l'expression musicale, à partir d'une séquence musicale audio.

Transformation du timbre (T_T) Les variations temporelles de Centre de Gravité Spectral ont été éliminées grâce à une transformation supprimant le Flux Spectral (cf. section 2.2.5.4 pour sa définition). En effet, supprimer le Flux Spectral consiste à fixer la forme de l'enveloppe spectrale au cours du temps, ce qui a pour conséquence de figer la valeur du Centre de Gravité Spectral.

Afin de conserver les variations d'énergie acoustique au cours du temps, nous avons utilisé une transformation similaire à celle élaborée dans [McAdams *et al.*, 1999] afin d'évaluer l'influence de simplifications de la distribution spectrale sur la discrimination de timbres. Cette dernière consiste à figer les amplitudes instantanées des composantes harmoniques à leur valeur moyenne (calculée sur la partie entretenue du son), en utilisant un terme de pondération permettant de conserver la valeur de l'enveloppe RMS instantanée. Notons qu'au vu de ces traitements, ce sont le Centre de Gravité Spectral harmonique (HSC), et l'enveloppe RMS calculée sur la partie harmonique du son, qui sont ici considérés. Puisqu'une élimination drastique des variations de Centre de Gravité Spectral pourrait rendre les sons trop peu naturels, les micro-fluctuations des amplitudes instantanées (variations hautes fréquences de faibles amplitudes)¹ ont été conservées.

La nouvelle amplitude instantanée $\tilde{A}_h(t)$ de la h^{eme} composante harmonique, est obtenue selon les équations suivantes :

$$\begin{aligned}\tilde{A}_h(t) &= \beta_h(t) ENV(t) \\ \beta_h(t) &= \frac{\overline{A}_h + H_h(t)}{\sqrt{\sum_{h=1}^H [\overline{A}_h + H_h(t)]^2}}\end{aligned}\quad (6.1)$$

où $ENV(t)$ est l'enveloppe RMS, $\overline{A}_h$ est l'amplitude moyenne calculée sur la partie soutenue de l'amplitude instantanée originelle $A_h(t)$. $H_h(t)$ correspond aux micro-fluctuations hautes fréquences. Comme $H_h(t)$ est faible devant $\overline{A}_h$, le terme $\beta_h(t)$ est pratiquement constant au cours du

¹Ces micro-fluctuations de l'amplitude instantanée sont parfois appelées *shimmer* dans la littérature (cf. par ex. [Jensen, 1999b, p. 13]).

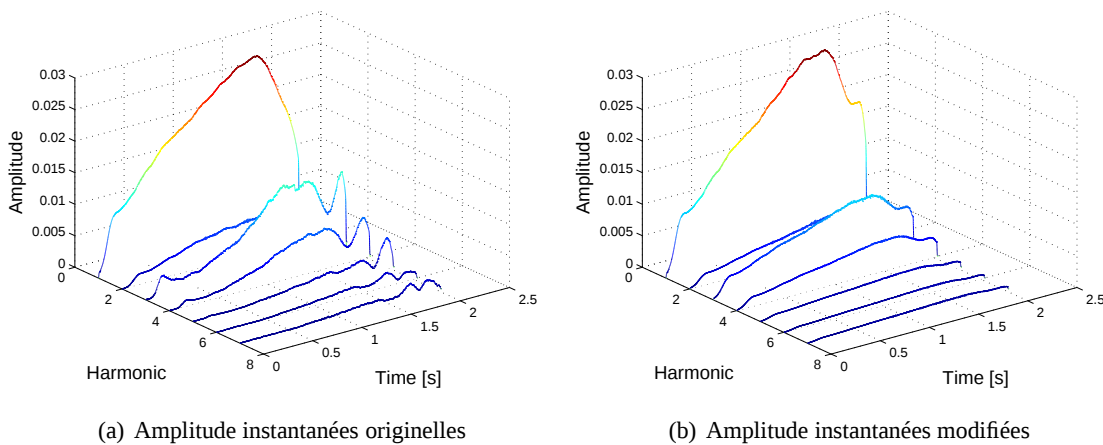


FIG. 6.2 – Transformations de timbre : gel du Flux Spectral (et donc de Centre de Gravité Spectral). (a) Amplitudes instantanées d'un son original de clarinette, et (b) Amplitudes instantanées modifiées. Seules les 7 premières composantes harmoniques sont représentées.

temps, mais fluctue légèrement autour de sa valeur moyenne. La forme des amplitudes instantanées modifiées est donc très semblable à celle de l'enveloppe RMS, à différents facteurs d'échelle près. Les figures 6.2(a) et 6.2(b) permettent de visualiser l'effet de la transformation sur les amplitudes instantanées d'une note de clarinette.

La transformation a pour effet de quasiment figer la valeur du Centre de Gravité Spectral au cours du temps. Comme le montrent les figures 6.3(a) et 6.3(b), le Centre de Gravité Spectral du son modifié fluctue légèrement autour de la valeur moyenne du Centre de Gravité Spectral du son original.

L'**Exemple Sonore n° 56** correspond à la resynthèse de l'une des interprétations de l'extrait de Bach, sans transformation, hormis le gel des fréquences instantanées (séquence M_0). L'**Exemple Sonore n° 57** correspond à la resynthèse avec le gel du Centre de Gravité Spectral au cours du temps (M_T).

Transformation du rythme (T_R) La suppression des déviations rythmiques introduites par l'interprète consiste à modifier les durées des notes de façon à ce qu'elles se conforment à une transcription exacte des indications de la partition. Ceci peut être réalisé sans affecter le timbre des sons par dilatation ou compression temporelle des amplitudes et fréquences instantanées.

Cette modification ne s'opère qu'après le transitoire d'attaque afin de préserver ce dernier. La fin du transitoire d'attaque est estimée en détectant le premier point d'inflexion significatif de l'enveloppe d'amplitude RMS (obtenu lorsque la dérivée seconde de l'enveloppe s'annule). Le Temps d'Attaque AT n'a pas été utilisé à cet effet, car il définit une durée, corrélée avec la perception de la durée d'attaque, et non l'instant de fin d'attaque. D'autre part, la notion d'instant de fin d'attaque utilisée dans le calcul de AT, reposant sur un pourcentage de l'amplitude maximale de l'enveloppe à atteindre, s'avère ne pas être ici appropriée. En effet, défini ainsi, l'instant de fin d'attaque tend à covarier avec la localisation temporelle de l'amplitude maximale, et s'avère surestimé lorsque le maximum est atteint vers la fin du son (par ex. *crescendo*).

Pour une note donnée, le coefficient de dilatation / compression α s'obtient ainsi :

$$\alpha = \frac{D'_{nom}}{D'_{eff}} \quad (6.2)$$

où D'_{nom} et D'_{eff} sont les durées nominales et effectives de la note auxquelles ont été retranchées la durée du transitoire d'attaque. Les valeurs des amplitudes $\widetilde{A}_h(t)$ et fréquences instantanées $\widetilde{f}_h(t)$ après le transitoire d'attaque sont obtenues d'après les équations suivantes :

$$\begin{aligned} \widetilde{A}_h(t) &= A_h(\alpha(t)) \\ \widetilde{f}_h(t) &= f_h(\alpha(t)) \end{aligned} \quad (6.3)$$

En pratique, les modifications sont réalisées en utilisant la fonction *spline* de Matlab qui permet de réaliser une interpolation des paramètres par splines cubiques.

Les figures 6.3(c) et 6.3(d) mettent en évidence l'effet de la suppression des déviations rythmiques au travers de la valeur du descripteur ΔDUR (cf. section 5.4.2) pour une séquence musicale expressive et pour la version modifiée.

L'Exemple Sonore n° 58 correspond à la resynthèse avec la suppression des déviations rythmiques initialement produites par l'interprète (M_R).

Transformation de la dynamique (T_D) De façon à limiter les variations de l'énergie acoustique, nous avons implémenté le compresseur et limiteur décrit dans [Zölzer, 1997]. La détection du niveau du signal est effectuée par un détecteur d'enveloppe utilisant les mesures crêtes. Un facteur de gain est ensuite utilisé pour ajuster l'amplitude du signal. Le compresseur / limiteur est contrôlé par les paramètres classiques de tels systèmes : le seuil de démarrage et le gain du limiteur, le seuil de démarrage et le gain du compresseur, ainsi que les temps d'attaque et de relâchement du système qui sont liés à la procédure de détection d'enveloppe. Bien que cette procédure soit non-linéaire et puisse théoriquement introduire de la distorsion harmonique, les paramètres de contrôle ont été choisis de façon à éviter des modifications de timbre audibles. Les temps d'attaque et de relâchement du système ont été fixés à 10 ms.

Un exemple de compression de la dynamique est donné sur les figures 6.3(e) et 6.3(f). La transformation a pour effet de réduire l'ambitus de variation de la dynamique.

Combinaisons des transformations Quatre combinaisons de ces transformations ont également été réalisées : la transformation du timbre et du rythme T_{TR} , la transformation du timbre et de la dynamique T_{TD} , la transformation du rythme et de la dynamique T_{RD} , et la transformation du timbre, du rythme et de la dynamique T_{TRD} .

6.2.1.3 Elaboration des stimuli

Les trois transformations de base (T_T , T_R , et T_D) ainsi que les quatre combinaisons (T_{TR} , T_{TD} , T_{RD} , et T_{TRD}) ont été appliquées aux séquences de clarinettes expressives sélectionnées. Les séquences ayant subi la compression de dynamique ont été égalisées en sonie relativement aux autres séquences. La resynthèse de la séquence de clarinette sans transformation a également été effectuée. Un léger effet de réverbération a été ajouté aux différentes séquences afin de reproduire l'influence naturelle d'une salle. Ainsi, pour chaque interprétation de clarinette sélectionnée, les 8 stimuli répertoriés dans le tableau 6.1 ont été générés.

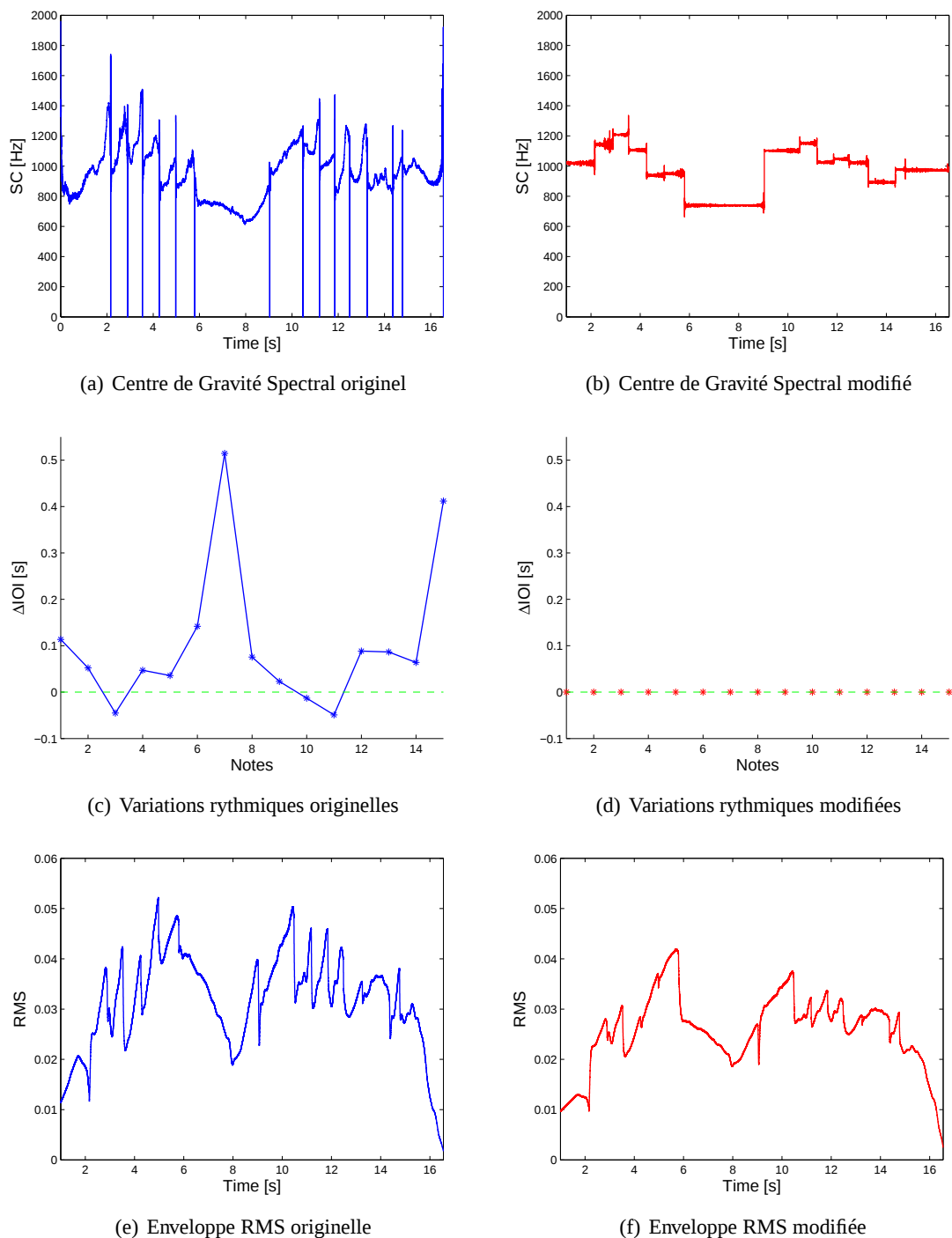


FIG. 6.3 – Variations expressives originelles (gauche) et modifiées (droite) pour l'une des interprétations de l'extrait de Mozart (première phrase).

Stimuli	Transformations	Description des transformations
M_0	-	Aucune transformation
M_T	T_T	Gel du Centre de Gravité Spectral
M_R	T_R	Élimination des déviations rythmiques
M_D	T_D	Compression de la dynamique
M_{TR}	T_{TR}	Combinaison de T_T et de T_R
M_{TD}	T_{TD}	Combinaison de T_T et de T_D
M_{RD}	T_{RD}	Combinaison de T_R et de T_D
M_{TRD}	T_{TRD}	Combinaison de T_T , T_R et T_D

TAB. 6.1 – Description des stimuli.

6.2.2 Participants

L'expérience perceptive a été conduite par 20 participants (14 hommes et 6 femmes, âgés de 19 à 50 ans). Des musiciens ont été choisis en raison de la tâche nécessitant de comparer des interprétations musicales. La plupart d'entre eux proviennent de l'ensemble d'improvisation du Groupe de Recherche et d'Improvisations Musicales (GRIM) de Marseille. Un panel varié d'instrumentistes a ainsi été représenté (clarinettiste, violoniste, pianiste, guitariste, etc.).

Le choix d'un nombre de 20 participants est lié aux résultats des analyses d'homogénéité des réponses au sein de l'échantillon (cf. sections 6.2.5.2 et 6.2.6.1).

6.2.3 Appareillage

L'appareillage est le même que celui de l'expérience de dissemblance décrite au chapitre 4 (cf. section 4.2.4).

6.2.4 Procédure : test de jugement de préférence

Nous avons choisi d'évaluer l'influence des transformations de l'expression musicale par des jugements de préférences issus de la comparaison par paires des différents stimuli. Les mesures de préférence ont par exemple été employées dans [Wapnick *et al.*, 1993] afin d'étudier la régularité des jugements lors d'évaluations d'interprétations musicales, et dans [Susini *et al.*, 1999] afin de déterminer des corrélats psychoacoustiques de la qualité sonore dans le cas de sons de voiture. Nous avons préféré mesurer les préférences selon un protocole de comparaison par paires plutôt que de demander aux participants de classer les séquences par ordre de préférence. En effet les mesures de préférences ne sont pas nécessairement transitives : s'il y a trois objets à comparer, A, B et C, il se peut que A soit préféré à B, B à C, mais que C soit préféré à A, ce qui poserait un problème dans le cas d'un test de classement par ordre de préférence.

L'expérience comprend deux sessions qui se succèdent, l'une associée à l'extrait de Bach (nous la désignerons par session 'Bach'), l'autre à l'extrait de Mozart (nous la désignerons par session 'Mozart'). L'ordre de passage des sessions a été alterné entre les différents participants.

Comme pour l'expérience de dissemblance décrite au chapitre 4, chaque session comprend les trois phases suivantes : écoute des stimuli, apprentissage de la tâche, et expérience. Les différents stimuli ont été comparés par paires. Pour chaque paire, les participants devaient sélectionner l'interprétation préférée. Ils avaient la possibilité de réécouter chacune des séquences. Chaque session comprend 28 paires, correspondant aux combinaisons non ordonnées et sans répétitions des 8 stimuli. Conséquemment, la durée des sessions est raisonnable (environ 20 mn). L'ordre de

présentation des différentes paires et l'ordre de présentation des stimuli au sein d'une paire ont été déterminés de manière aléatoire (cf. section 2.2.4.1.4).

A la fin du test, les participants étaient invités à décrire les critères expliquant leurs préférences sur un questionnaire.

6.2.5 Méthodes statistiques

Les données perceptives ont été analysées en utilisant des tests statistiques non paramétriques et des analyses multivariées. Décrivons dans un premier temps la représentation des données issues de l'expérience.

6.2.5.1 Représentation des données perceptives

Matrices de préférence individuelle et globale Les réponses d'un participant s sont contenues dans une matrice de préférence, notée P_s . Les éléments de P_s , notés $a_s(i, j)$, indiquent si le stimulus i a été préféré au stimulus j . Ils sont définis ainsi :

$$\forall (i, j) \in [1; N]_{\mathbb{N}}^2, \quad i \neq j \quad a_s(i, j) = \begin{cases} 1 & \text{lorsque le stimulus } i \text{ est préféré} \\ 0 & \text{lorsque le stimulus } j \text{ est préféré} \end{cases}$$

où N est le nombre total de stimuli ($N=8$). Comme les paires de stimuli identiques n'ont pas été employées, la diagonale de la matrice est fixée à 0. Les données de l'expérience remplissent la moitié des éléments de la matrice. Les autres éléments sont obtenus par complément par rapport à 1 de la façon suivante :

$$\forall (i, j) \in [1; N]_{\mathbb{N}}^2, \quad i \neq j \quad a_s(i, j) = 1 - a_s(j, i) \quad (6.4)$$

La matrice de préférence de l'échantillon, notée P , est définie comme la somme des matrices de préférences individuelles P_s :

$$P = \sum_{s=1}^S P_s \quad (6.5)$$

où S est le nombre total de participants dans l'échantillon ($S=20$). Ses éléments, notés $b(i, j)$ rendent compte du nombre de fois que le stimulus i a été préféré au stimulus j .

Notes d'interprétations Un autre moyen de représenter les données est de dériver des réponses des "notes" pour les différentes interprétations. Ces notes peuvent être obtenues en comptant le nombre total de fois qu'une interprétation a été préférée. Pour chaque participant s , elles sont calculées ainsi :

$$\forall i \in [1; N]_{\mathbb{N}} \quad R_s(i) = \sum_{\substack{j=1 \\ i \neq j}}^N P_s(i, j) \quad (6.6)$$

Etant donné le nombre total d'interprétations ($N=8$), les notes d'interprétations sont nécessairement comprises entre 0 et 7.

Les notes moyennes d'interprétation $R(i)$ des différentes séquences sont calculées sur l'échantillon considéré de la manière suivante :

$$\forall i \in [1; N]_{\mathbb{N}} \quad R(i) = \frac{1}{S} \sum_{s=1}^S R_s(i) \quad (6.7)$$

6.2.5.2 Homogénéité des réponses

Avant de procéder à l'analyse des préférences, il convient de vérifier l'homogénéité des réponses, i.e. le degré d'accord entre les participants. Nous avons pour cela utilisé le coefficient d'accord u de Kendall pour les comparaisons par paires (cf. [Siegel et John Castellan, 1988, p. 272]). Cette mesure d'association non paramétrique est définie par la formule suivante :

$$u = \frac{2 \sum_{i=1}^N \sum_{j=1}^N C_2^{b(i,j)}}{C_2^S C_2^N} - 1 \quad (6.8)$$

où C_n^p correspond au coefficient binomial²³. Rappelons que N est le nombre total de stimuli ($N = 8$), et que S est le nombre total de participants ($S = 20$). Lorsque S est pair (ce qui est notre cas), u peut varier de $\frac{-1}{S-1}$ à 1, cas pour lequel l'accord entre les participants est total. Dans [Siegel et John Castellan, 1988, p. 275], les auteurs définissent un index W_t calculé à partir de u et variant de 0 à 1.

La statistique u peut être vue comme une estimation du vrai degré d'accord ν dans la population. Le test statistique associé à u revient à tester l'hypothèse nulle $H_0 : \nu = 0$ qu'il n'y ait aucun accord entre les participants, contre l'alternative $H_1 : \nu \neq 0$, que le degré d'accord entre les participants soit supérieur à ce qu'il serait si les résultats des comparaisons par paires avaient été tirés de façon aléatoire. Comme le nombre de participants est grand ($S > 6$), une approximation de la distribution d'échantillonnage est utilisée. Elle est asymptotiquement distribuée comme une distribution χ^2 . Les résultats de ce test déterminent si la matrice de préférence globale de l'échantillon P reflète une tendance générale, auquel cas il pourra être intéressant de l'utiliser lors d'autres analyses.

6.2.5.3 Analyses de variance

Afin d'évaluer les influences de l'extrait musical ('Bach' ou 'Mozart') et des différentes transformations (8 modalités), les données ont été traitées au moyen d'une analyse de variance (ANOVA) à deux facteurs, à mesures répétées (chaque sujet a effectué les deux sessions 'Bach' et 'Mozart'). Les mesures utilisées sont les notes d'interprétation R_s dérivées des mesures de préférence (cf. section 6.2.5.1), qui sont de nature ordinale. Comme nous le verrons par la suite, les résultats de cette ANOVA globale révèlent qu'il n'y a pas de différence significative entre les deux extraits musicaux par rapport aux notes d'interprétation. Des ANOVA à un facteur (la transformation) à mesures répétées ont donc été conduites par la suite de manière individuelles pour les deux extraits 'Bach' et 'Mozart'. Afin d'évaluer l'influence des transformations, des tests de comparaisons multiples ont également été réalisés selon la procédure de Tukey (HSD "Honestly Significant Difference"). Cette procédure permet de déterminer les différences significatives entre les moyennes des notes d'interprétation pour toutes les combinaisons 2 à 2 des différentes transformations.

La séquence M_0 , obtenue sans transformation, servira de référence afin d'évaluer les conséquences des différentes réductions des paramètres d'expression musicale (timbre, rythme, dynamique).

6.2.5.4 Classification Ascendante Hiérarchique (CAH)

Puisque la nature des stimuli est d'ordre catégorielle (ils ont subi telle transformation ou ne l'ont pas subi), une méthode de classification ascendante hiérarchique (CAH) a également été uti-

²également noté $\binom{n}{p}$

³Dans [Siegel et John Castellan, 1988, p. 274], les auteurs fournissent également une formule simplifiée de l'expression présentée ici.

lisée pour analyser les données. Celle-ci a pour but de regrouper les différentes séquences selon leurs notes d'interprétation (cf. section 6.2.5.1). Les distances euclidiennes entre les notes d'interprétation $R(i)$ de l'échantillon ont été calculées pour chacune des 28 paires de séquences ($\frac{8 \times 7}{2}$). La méthode d'agrégation employée repose sur l'algorithme du lien complet (cf. par ex. [Saporta, 1990]). Ainsi, nous déterminerons si une relation peut être établie entre les groupes de séquences obtenus à l'issue de la CAH et les transformations auxquelles les séquences sont associées.

6.2.6 Résultats

6.2.6.1 Homogénéité des réponses

La figure F.4 de l'Annexe F.3 présente l'évolution de l'index W_t (qui peut varier entre 0 et 1), associé au coefficient d'accord de Kendall u , en fonction du nombre de participants aux expériences perceptives. Pour les deux sessions ('Bach' et 'Mozart'), l'index tend à se stabiliser au fur et à mesure que le nombre de participants augmente. Les évolutions de l'index devenant faibles à partir d'un nombre de participants suffisamment élevé, nous avons limité les expériences à 20 participants.

Les valeurs du coefficient d'accord de Kendall u et de l'index W_t sont reportées dans le tableau 6.2. Pour les deux sessions 'Bach' et 'Mozart', l'hypothèse nulle H_0 , qu'il n'y ait aucun accord entre les participants peut être rejetée avec un risque $\alpha < .001$. Nous pouvons en conclure qu'il y a un fort accord entre les participants vis-à-vis des séquences qu'ils préfèrent.

Session	Coefficient d'accord u	Index associé W_t	Statistique X^2	Degrés de liberté df
Bach	0.58 ($p < .001$)	0.60	338.20	28
Mozart	0.52 ($p < .001$)	0.54	303.80	28

TAB. 6.2 – Coefficient d'accord pour les deux sessions 'Bach' et 'Mozart'. Les valeurs de la statistique X^2 sont également indiquées. $df = 28$ indique le nombre de degrés de liberté.

Analysons maintenant si les réponses diffèrent en fonction de l'extrait musical choisi.

6.2.6.2 Influence de l'extrait musical

Les résultats de l'ANOVA à deux facteurs (l'extrait musical et les transformations) montrent que les différences entre les deux extraits 'Bach' et 'Mozart' ne sont pas suffisamment grandes pour exclure la possibilité qu'elles peuvent être dues au hasard. Il n'y a pas d'effet significatif de l'extrait sur les notes d'interprétation ($p=1.00$, les moyennes des notes d'interprétation sont identiques pour les deux extraits). L'interaction entre l'extrait et les transformations est par conséquent également non significatif [$F(7;133)=1.047$, $p=0.401$]. L'influence des transformations a donc été évalué individuellement pour chaque extrait.

6.2.6.3 Influence des transformations

Les ANOVA à un facteur (la transformation) indiquent des effets fortement significatifs des transformations pour les extraits 'Bach' [$F(7;133)=70.32$, $p<0.001$] et 'Mozart' [$F(7;133)=62.66$, $p<0.001$]. Les résultats des comparaisons multiples (Tukey) entre les transformations au sein de chaque session 'Bach' et 'Mozart' sont présentés dans les tableaux 6.3 et 6.4. Ils permettent de déterminer les paires de séquences pour lesquelles les différences de notes d'interprétation sont

significatives.

Les notes d'interprétation associées aux différentes séquences sont décrites par les diagrammes en boîte à moustaches⁴ de la figure 6.4.

	M_T	M_R	M_D	M_{TR}	M_{TD}	M_{RD}	M_{TRD}
M_0	13.30***	-	6.65***	15.92***	22.77***	8.87***	22.37***
M_T		9.67***	6.65***	-	9.47***	4.43*	9.07***
M_R			-	12.29***	19.14***	5.24**	18.74***
M_D				9.27***	16.12***	-	15.72***
M_{TR}					6.85***	7.05***	6.45***
M_{TD}						13.90***	-
M_{RD}							13.50***

TAB. 6.3 – Résultats des tests de comparaisons multiples (Tukey) pour le facteur lié aux transformations (8) au sein de la session 'Bach'. Le tableau présente les valeurs de la statistique q lorsqu'il y a des différences significatives. Lorsque les différences sont non significatives, le signe – est indiqué. Les probabilités de commettre une erreur de Type I sont représentées par : * $p < .05$, ** $p < .01$, *** $p < .001$.

	M_T	M_R	M_D	M_{TR}	M_{TD}	M_{RD}	M_{TRD}
M_0	14.52***	-	6.76***	13.53***	21.89***	7.36***	19.90***
M_T		13.13***	7.76***	-	6.37***	5.37**	5.41**
M_R			5.37**	12.14***	19.50***	5.97***	18.50***
M_D				6.76***	14.13***	-	13.13***
M_{TR}					7.36***	6.17***	6.37***
M_{TD}						13.53***	-
M_{RD}							12.53***

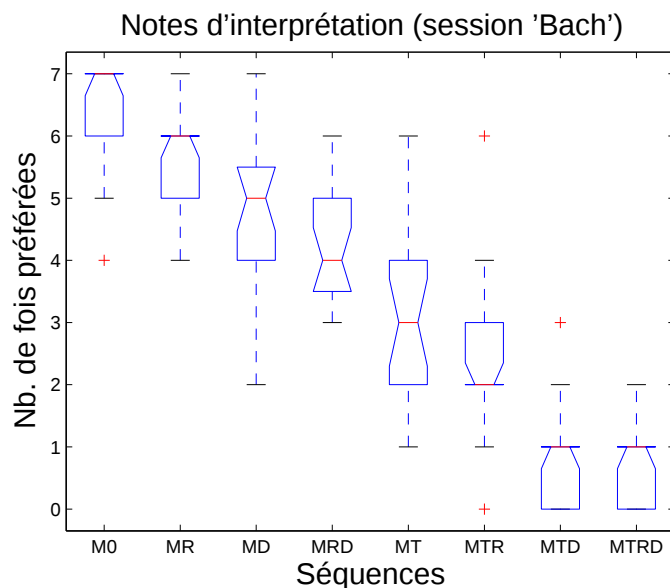
TAB. 6.4 – Résultats des tests de comparaisons multiples (Tukey) pour le facteur lié aux transformations (8) au sein de la session 'Mozart'. Légende : voir tableau 6.3.

Il n'est pas surprenant de constater que dans les deux sessions ('Bach' et 'Mozart') la séquence M_0 est celle qui a été le plus souvent préférée (cf. figure 6.4). En effet, cette séquence n'a pas subi les réductions des paramètres d'expression liés au timbre, au rythme, et à la dynamique. Réciproquement, il n'est pas surprenant non plus que la séquence ayant subi les trois réductions de timbre, de rythme et de dynamique M_{TRD} soit celle qui soit en moyenne la moins préférée.

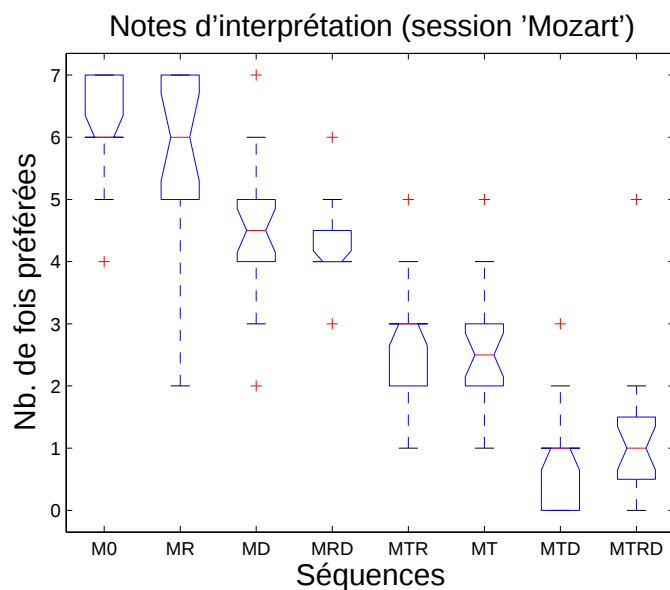
Outre la séquence de référence n'ayant subi aucune transformation, l'annulation des variations rythmiques est la transformation qui induit en moyenne le moins de perte de qualité musicale quelque soit la session. D'après les tableaux 6.3 et 6.4, il n'y a pas d'effet significatif de la transformation rythmique par rapport à la séquence de référence M_0 quelque soit l'extrait. Dans le cas de l'extrait de Bach, ceci est probablement relié au fait que la musique s'adapte bien à un rythme mécanique ; en effet, rappelons que l'extrait provient d'une *Allemande* qui est une danse instrumentale, ici de rythme binaire à quatre temps⁵. Concernant l'extrait de Mozart, la faible influence

⁴appelés également diagrammes en boîte de Tukey, cf. [Saporta, 1990, p. 115]

⁵On notera cette citation de Rousseau à propos de l'*Allemande* : "Sorte d'air ou de pièce de musique dont la musique est à quatre temps et se bat gravement." [Rousseau, 2007]



(a) Session 'Bach'



(b) Session 'Mozart'

FIG. 6.4 – Diagrammes en boîte à moustaches des notes d'interprétation pour les sessions 'Bach' (a) et 'Mozart' (b). Rappelons que les notes d'interprétation sont nécessairement comprises entre 0 et 7 (cf. section 6.2.5.1). Les séquences sont rangées de gauche à droite par ordre décroissant des valeurs médianes des notes d'interprétation. Les boîtes indiquent de bas en haut le premier quartile Q_1 , la médiane Q_2 , et le troisième quartile Q_3 . Les moustaches s'étendent de part et d'autre des boîtes jusqu'aux valeurs suivantes : en bas, jusqu'à $Q_1 - 1.5 \times (Q_3 - Q_1)$ s'il existe des valeurs encore plus petites, sinon jusqu'à la valeur minimale ; en haut, jusqu'à $Q_3 + 1.5 \times (Q_3 - Q_1)$ s'il existe des valeurs encore plus grandes, sinon jusqu'à la valeur maximale. Les valeurs atypiques sont spécifiées par des croix. Les encoches représentent une estimation de l'incertitude sur les médianes.

de l'annulation des variations rythmiques peut paraître plus étonnante étant donné que l'extrait au tempo plus lent est plus propice à ce type d'effets. Remarquons cependant que la variance de la note d'interprétation associée à la séquence M_R est beaucoup plus importante pour l'extrait de Mozart que pour l'extrait de Bach.

Après l'annulation des déviations rythmiques, la compression de dynamique est la transformation qui réduit le moins les préférences par rapport à la référence. Son effet par rapport à M_0 est cependant significatif (cf. tableaux 6.3 et 6.4).

Pour les deux sessions, les séquences qui ont subi la réduction de variations de timbre (M_T , M_{TD} , M_{TR} , M_{TRD}) sont systématiquement celles qui sont les moins préférées. Les tableaux 6.3 et 6.4 montrent que ce sont celles qui présentent les différences significatives les plus élevées par rapport à la référence M_0 . Le gel du Centre de Gravité Spectral induit donc une perte plus grande de qualité musicale que l'annulation des variations rythmiques ou la compression de dynamique. Il est intéressant de noter que l'effet d'annuler les variations de timbre est à lui seul plus dégradant du point de vue musical que la combinaison des réductions liées au rythme et à la dynamique (cf. note d'interprétation de M_T par rapport à celle de M_{RD}).

6.2.6.4 Classification ascendante hiérarchique

La figure 6.5 présente les résultats de la classification ascendante hiérarchique des différentes séquences en fonction de leurs notes d'interprétation. Les deux groupes principaux sont identiques pour les deux sessions. Ils rassemblent d'un côté toutes les séquences ayant subi la transformation liée au timbre ($M_T, M_{TD}, M_{TR}, M_{TRD}$), et de l'autre les séquences qui ne l'ont pas subi (M_0, M_R, M_D, M_{RD}). Ces résultats confirment l'effet prédominant de la transformation de timbre par rapport aux autres transformations.

6.2.7 Discussion

Les analyses de préférence qui viennent d'être présentées montrent que la sensibilité des participants aux réductions des variations expressives dépend des paramètres qui ont été modifiés (timbre et/ou rythme et/ou dynamique). La nature de l'extrait musical n'a pas changé de manière significative les préférences des participants. Parmi les sept réductions de l'expression musicale qui ont été effectuées, les effets les plus significatifs ont été entraînés par le gel du Centre de Gravité Spectral. Cette transformation de timbre induit en effet une plus grande perte de qualité musicale que l'annulation des variations rythmiques, la compression de la dynamique, ou leur combinaison. Dans [McAdams *et al.*, 1999], les auteurs ont étudié les effets de différentes simplifications spectro-temporelles sur la discrimination de sons instrumentaux isolés (par ex. clarinette, clavecin, trompette, etc.). Leur expérience montre que les auditeurs sont sensibles à un gel du Centre de Gravité Spectral (ou plus précisément du Flux Spectral). L'effet de cette simplification est toutefois moins significatif pour les sons de clarinette que pour les sons de trompette ce qui, comme le supposent les auteurs, est lié au fait que les sons de clarinette présentent un Flux Spectral plus faible que ceux de trompette [Grey, 1977].

Dans notre expérience, l'effet de cette transformation est étudié du point de vue de la préférence musicale. Les résultats montrent qu'elle induit une nette réduction des préférences musicales. Cette réduction peut s'expliquer selon trois hypothèses. La première hypothèse est que l'annulation des variations de Centre de Gravité Spectral au cours des notes entraîne une altération globale du *timbre-identité*⁶ de l'instrument. En effet, si les techniques d'analyse-synthèse

⁶aspect du timbre qui rend possible l'identification l'instrument, ou plus généralement celle de la source, cf. chapitre 2.

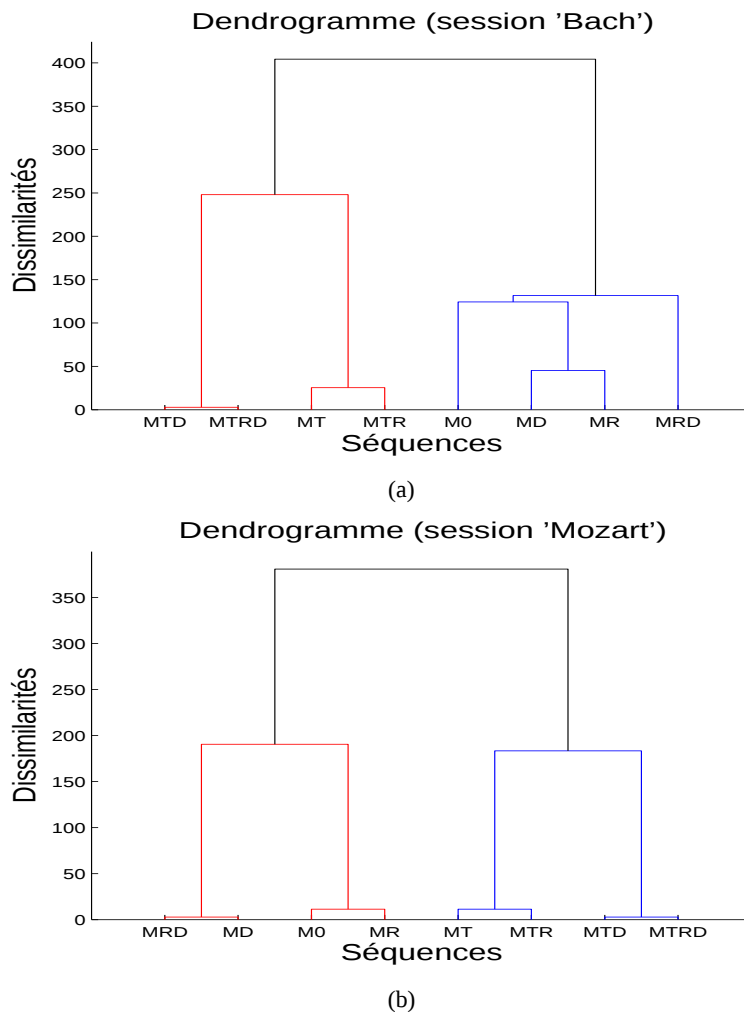


FIG. 6.5 – Représentations en dendrogrammes des classifications ascendantes hiérarchiques des différentes séquences pour les sessions 'Bach' (a) et 'Mozart' (b). La méthode d'agrégation utilisée repose sur l'algorithme du lien complet (cf. par ex. [Saporta, 1990]).

permettent de contrôler indépendamment le Centre de Gravité Spectral par rapport à d'autres dimensions sonores (par ex. l'énergie acoustique), ceci n'est peut-être pas acceptable du point de vue de l'identité timbrale des sons résultants (nous avons vu qu'il existait un lien fort entre Centre de Gravité Spectral et dynamique pour les sons de clarinette au chapitre 5). La deuxième hypothèse est que l'annulation des variations de Centre de Gravité Spectral est perçue comme une altération du *timbre-qualité* de l'instrument. Enfin, la troisième hypothèse se veut être au croisement des deux premières, i.e. que la transformation induit aussi bien un changement de *timbre-identité* qu'un changement de *timbre-qualité*. Nous avons fait écouter au clarinettiste, avec qui nous avons réalisé les enregistrements, les séquences ayant subi la transformation de timbre. Ce dernier a perçu qu'elles étaient différentes de son interprétation mais sans en être complètement certain. Il n'a pas évoqué le fait que ce n'était plus des séquences de clarinette, et au contraire a cherché des explications possibles des effets perçus. Parmi ces explications, il a mentionné un mauvais contrôle de l'instrument par l'interprète (par ex. un élève qui change son embouchure), un matériel usagé (par ex. une vieille anche), autant d'aspects expliquant le manque d'homogénéité globale de l'interprétation résultante. Ces remarques ne réfèrent pas à des changements de timbre-identité mais à des changements de timbre-qualité. D'après les questionnaires, les participants ont pour la plupart signalé avoir préféré les séquences avec des sons "vivants" par rapport aux séquences avec

des sons “statiques”. Ceci ne réfère pas non plus à des changements de timbre-identité, mais au contraire à la présence ou l’absence de variations de timbre-qualité et/ou d’intensité. Ces observations montrent que la transformation ne bouleverse pas profondément le timbre de l’instrument. L’aspect important qui ressort des analyses est que le manque de variations de timbre-qualité est déterminant dans la perception de l’expression musicale.

Abeles a travaillé sur le développement d’une échelle d’évaluation d’interprétation musicale à la clarinette [Abeles, 1973]. Ses résultats montrent que le timbre est l’un des plus importants facteurs utilisés par les professeurs de musique afin de juger les interprétations de clarinette. Au vu de ces analyses, il n’est pas si étonnant que les participants de notre expérience se soit essentiellement reposés sur le traitement qui affecte le timbre-qualité des sons.

Dans cette expérience, le complexe processus d’interprétation musicale est réduit à un modèle simple de type “linéaire” et “additif” (Timbre +/- Rythme +/- Dynamique). Ceci est une première étape afin de mieux comprendre les effets induits par chacun de ces paramètres. Cependant, il est probable, tant au niveau de l’interprète (dans le processus de production sonore) qu’au niveau de l’auditeur (dans le décodage du signal musical), que ces paramètres musicaux soient en interaction (par ex. une modification de timbre peut participer à la perception d’une modification de rythme). Lors d’une description de ces interprétations, le clarinettiste était persuadé d’avoir volontairement allongé la durée d’une note induisant la suivante à démarrer en retard. Les analyses révèlent que l’amorce de la seconde note démarre parfaitement sur le temps. Nous supposons que la sensation de retard provient du fait que la première des deux notes présente un *decrescendo* doublé d’une diminution de brillance, rendant incertain l’instant de sa fin. Cette anecdote met en avant la complexe interaction qui peut se produire au niveau perceptif entre les paramètres de timbre, de rythme et de dynamique, supposés indépendants sur le plan physique.

6.3 Expérience 2 : d’un jeu inexpressif vers un jeu expressif

La première expérience a permis de montrer que l’annulation des variations temporelles de la brillance des différentes notes d’une séquence musicale entraîne une diminution de sa qualité musicale. Dans l’expérience qui est présentée dans cette section, nous nous intéressons aux effets du processus inverse, c’est-à-dire savoir si l’application de variations expressives de brillance à une séquence initialement inexpressive peut en améliorer la qualité musicale.

Une technique de contrôle des variations temporelles de brillance par filtrage dynamique⁷ a été développée. Cette technique a été employée afin d’ajuster les variations de brillance de séquences musicales inexpressives en fonction d’un profil de variations de brillance issu d’un enregistrement d’une interprétation expressive. Les effets perceptifs de cette transformation de timbre ont ensuite été évalués grâce à des jugements de dissimilarité et de préférence mesurés pour un ensemble de 20 participants ayant à comparer différentes séquences musicales dont les variations de timbre avaient ou n’avaient pas été modifiées au préalable.

Nous décrirons dans un premier temps le corpus sonore employé ainsi que la méthodologie de contrôle dynamique de la brillance. Le protocole de l’expérience perceptive et les méthodes statistiques utilisées pour analyser les données seront ensuite exposés. Enfin, nous examinerons et discuterons des effets perceptifs induits par la transformation de brillance.

⁷ filtrage variable au cours du temps.

6.3.1 Stimuli

6.3.1.1 Corpus sonore

Un moyen simple d'obtenir des séquences musicales produites de façon similaire sur différents instruments est d'utiliser un échantillonneur. Certains échantillonneurs peuvent être contrôlés selon le protocole MIDI (Musical Instrument Digital Interface), qui transmet des informations relatives aux hauteurs et durées des notes, mais également relatives à leur intensité et à leur timbre (le paramètre qui contrôle ces aspects est appelé *vélocité*⁸).

Nous avons utilisé l'extrait musical provenant de l'*Allemande* de Bach (cf. chapitre 5 et section 6.2 du présent chapitre). Grâce aux analyses sonologiques décrites au chapitre précédent, nous avons défini une partition MIDI reproduisant les variations rythmiques d'une séquence interprétée de manière expressive. Afin de produire des séquences musicales inexpressives du point de vue du timbre, nous avons gardé constant dans le temps le paramètre MIDI de *vélocité*. Les variations d'intensité et de timbre au cours des notes se trouvent ainsi minimisées.

Cette partition MIDI a ensuite servi à contrôler un échantillonneur fonctionnant avec des sons instrumentaux naturels. Six instruments différents ont été employés : trois de type *entretenu* (la clarinette, le piccolo, et le violoncelle), et deux de type *non-entretenu* (la guitare, et la sitar). Les séquences musicales inexpressives ainsi générées ont ensuite subi une transformation de timbre portant sur l'évolution temporelle de la brillance. La prochaine section s'attache à décrire le principe de fonctionnement de la transformation de timbre. La figure 6.6 résume la procédure d'élaboration des stimuli.

6.3.1.2 Contrôle de la brillance par filtrage dynamique

L'un des atouts majeurs des modèles de synthèse sonore est d'offrir la possibilité de manipuler le timbre (cf. par ex. [Jensen, 1999b, p. 154]). Quelque soit la technique de synthèse employée, la manipulation du timbre découle de la modification des paramètres de contrôle du modèle. Dans le cas de la synthèse reposant sur la modélisation physique d'un instrument de musique, les paramètres de contrôle sont définis afin d'imiter le contrôle naturel de l'instrument (cf. chapitre 4). Dans le cas de la synthèse par modèle de signal *imitative*⁹, les paramètres de contrôle découlent d'une phase d'analyse au cours de laquelle les paramètres utilisés pour la resynthèse sont déterminés (par ex. analyse additive comme pour le modèle de synthèse additive employé au chapitre 5, détermination d'une fonction de distorsion non linéaire et d'un index de modulation pour la synthèse par distorsion non-linéaire, etc.). Le problème inverse consistant à déterminer les paramètres de contrôle des modèles à partir d'une description paramétrique du timbre (i.e. à partir de descripteurs de timbre, mais non du signal entier) est complexe. Ce type de procédure s'intègre dans le cadre plus général de la synthèse reposant sur des descripteurs (*feature-based synthesis* en anglais, cf. par ex. [Hoffman et Cook, 2007]). Dans [Beauchamp, 1982], le Centre de Gravité Spectral est utilisé afin d'ajuster l'un des paramètres de contrôle (index α) d'un modèle de synthèse fonctionnant suivant la distorsion non-linéaire et le filtrage passe-haut du signal de sortie. L'ajustement est possible car il existe une relation monotone croissante entre l'index α et le Centre de Gravité Spectral. La méthode est efficace et permet d'obtenir selon l'auteur de meilleurs performances pour la synthèse imitative que la synthèse par modulation de fréquence (*Frequency Modulation* en anglais). En revanche, elle nécessite en amont l'élaboration d'une fonction de distorsion dont la nature dépend de l'instrument étudié, et la connaissance de la relation entre le

⁸L'origine du nom de ce paramètre est lié au fonctionnement du piano. En effet, lorsqu'une touche de piano est enfoncée avec force, la vitesse avec laquelle le marteau vient frapper les cordes est plus importante que lorsque la touche est enfoncée doucement. Cela induit un son à la fois plus intense et plus riche spectralement.

⁹i.e. cherchant à reproduire un son naturel.

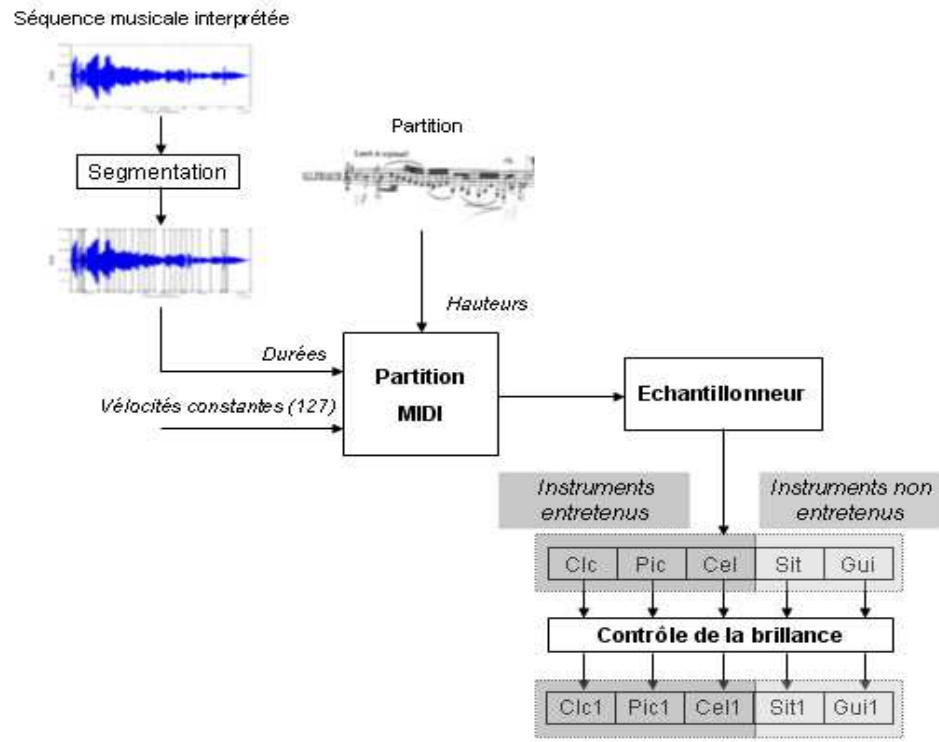


FIG. 6.6 – Méthode d'élaboration des stimuli. La nomenclature des séquences générées avec l'échantillonneur est la suivante : *Clc* (clarinette), *Pic* (piccolo), *Cel* (violoncelle), *Sit* (sitar), *Gui* (guitare).

paramètre de modulation α et le Centre de Gravité Spectral. Pour la synthèse additive, la détermination des amplitudes instantanées des composantes harmoniques à partir de la connaissance seule du descripteur de Centre de Gravité Spectral est un problème qui contient un nombre trop important d'inconnues (autant que de composantes harmoniques) pour être résolu de manière directe (pris tel quel, le problème admet une infinité de solutions). En effet, l'emploi de toute fonction du type $\beta_h(t) = \frac{N_h(t)}{\sqrt{\sum_{h=1}^H [N_h(t)]^2}}$ dans l'équation 6.1 permet potentiellement de modifier le Centre

de Gravité Spectral du son originel tout en conservant son enveloppe d'amplitude RMS, mais le problème est justement de déterminer les H inconnues $N_h(t)$ (H est le nombre de composantes harmoniques), qui permettent d'obtenir le Centre de Gravité Spectral désiré. Il ne peut y avoir de solution automatique à ce problème que si l'on se donne un modèle d'amplitude instantanée ne dépendant que de paramètres connus (par ex. le rang de la composante harmonique, et le Centre de Gravité Spectral). Dans [Jensen, 1999b, p. 88], Jensen propose un modèle d'enveloppe spectrale selon ce principe. Dans le domaine fréquentiel, le modèle d'enveloppe spectrale s'écrit sous la forme $A_h = B^{-h}$, où B peut être directement obtenu en fonction du Centre de Gravité Spectral. L'inversion de ce modèle dans le domaine temporel permet d'obtenir une fonction temporelle de contrôle de la brillance (appelée *Brightness Control Function* par l'auteur). Cette fonction représente un modèle intéressant afin d'obtenir un son avec un Centre de Gravité Spectral désiré, mais ne permet pas de réaliser à elle seule une synthèse imitative respectant l'identité timbrale d'un instrument (il faut pour cela disposer d'autres critères).

Au vu des difficultés que représente la résolution du problème inverse consistant à déterminer les paramètres de contrôles de modèle de synthèse pour obtenir une valeur du Centre de Gravité Spectral donnée et ce, dans le cadre d'un panel varié d'instruments de musique, nous n'avons pas

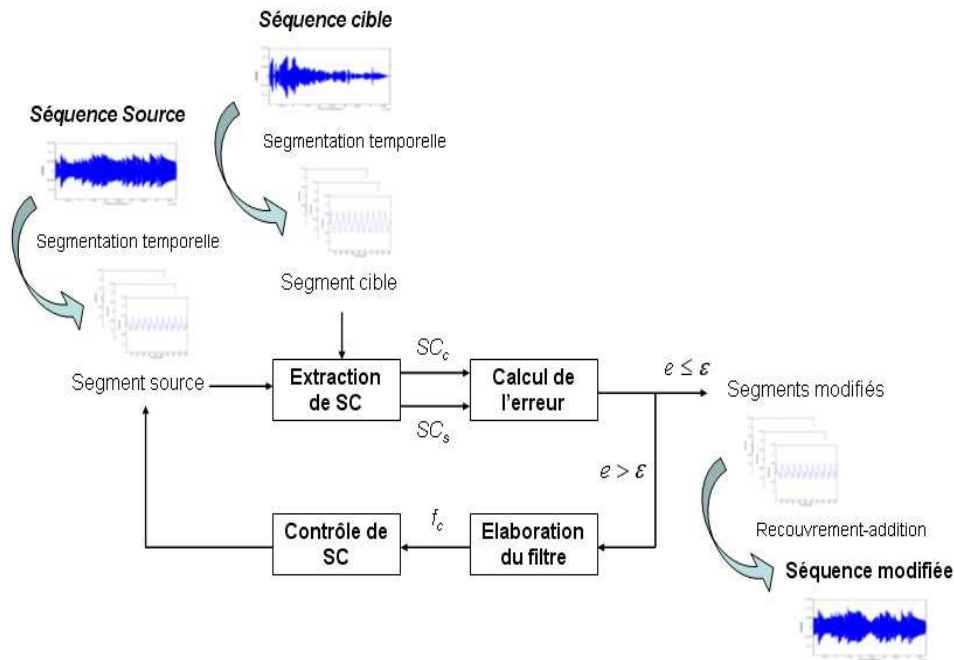


FIG. 6.7 – Méthodologie de contrôle dynamique de la brillance d'une séquence *source* à partir d'une séquence *cible*. Pour chaque segment temporel d'analyse, la fréquence de coupure f_c d'un filtre passe-bas est optimisée de manière à minimiser la fonction d'erreur e reposant sur la différence entre les Centres de Gravité Spectraux associés aux séquences source SC_s et cible SC_c . ϵ désigne le seuil d'erreur toléré.

utilisé dans cette étude de modèle de synthèse paramétrique. La méthodologie que nous avons développée repose sur un post-traitement de séquences musicales initialement générées à l'aide d'échantillonneurs. Son principe est fondé sur le fait que pour les sons instrumentaux, l'évolution de la brillance est généralement fortement corrélée à celle de la richesse spectrale (cf. le lien étroit entre le troisième coefficient de Tristimulus TR3 et le Centre de Gravité Spectral pour les sons de synthèse de clarinette étudiés au chapitre 4). Il paraît donc possible d'imiter la relation naturelle d'évolution de la brillance des sons instrumentaux en appliquant un filtrage passe-bas adéquat à un signal initialement riche. Robert Moog a ainsi pu réaliser la synthèse analogique soustractive de sons cuivrés en contrôlant la pente de coupure d'un filtre passe-bas [Risset, 1991, p. 245].

Notre méthodologie repose sur l'utilisation d'un filtre dynamique passe-bas servant à contrôler la brillance de séquences initialement riches en composantes spectrales. Ces dernières sont obtenues en fixant le paramètre de vélocité des échantillonneurs à sa valeur maximale (127), tel qu'indiqué sur la figure 6.6. L'ajustement du Centre de Gravité Spectral d'une séquence *source* sur celui d'une séquence *cible* s'opère par l'optimisation de la fréquence de coupure f_c d'un filtre dynamique passe-bas. Une procédure d'optimisation d'un filtre dynamique passe-bas a par exemple été utilisée dans [Aramaki et Kronland-Martinet, 2006] afin de déterminer les paramètres d'un modèle de synthèse de sons d'impacts.

La figure 6.7 présente les différentes étapes du contrôle dynamique de la brillance. Le processus s'établit à court-terme, i.e. sur des portions de signal de courte durée. En pratique, une fenêtre de Hann de 4096 points et un recouvrement de 50% de la taille de la fenêtre sont utilisés pour segmenter les signaux (la fréquence d'échantillonnage F_e des signaux est de 44.1 kHz). Ces paramètres ont été choisis de façon à ne contrôler que les variations lentes du Centre de Gravité Spectral (d'un maximum d'environ 10 Hz), susceptibles de découler des gestes de contrôle de l'instrumentiste.

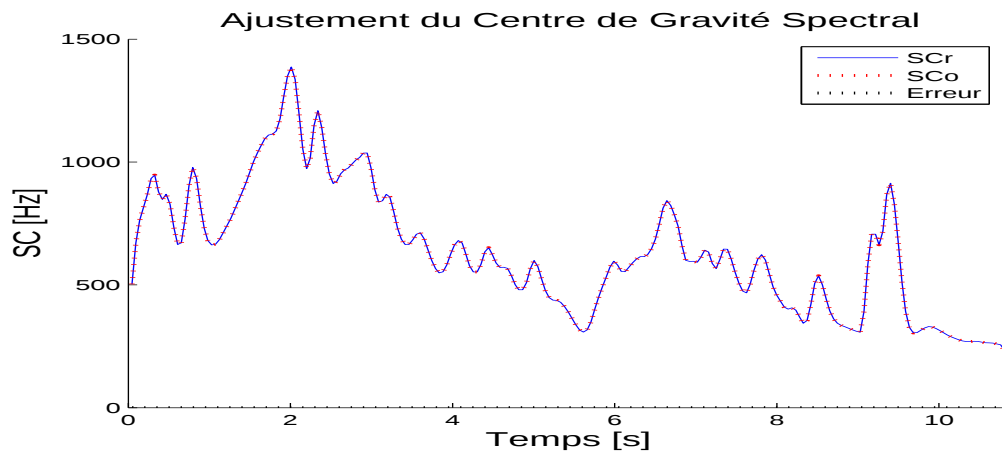


FIG. 6.8 – Résultats de la boucle d'optimisation utilisée dans la procédure de contrôle de la brillance : application à la séquence de clarinette (*Clc*) générée avec l'échantillonneur. Le Centre de Gravité Spectral instantané de la séquence modifiée (*SCo*) est quasiment égal à celui de la séquence cible (*SCr*).

Ceci permet également de ne pas altérer l'attaque des notes. Les Centres de Gravité Spectraux instantanés des séquences source SC_s et cible SC_c sont calculés pour chaque segment temporel d'analyse (selon l'équation 2.22). Le filtre passe-bas choisi pour filtrer les séquences est un filtre numérique passe-bas de type *Butterworth* d'ordre 2. La fréquence de coupure du filtre f_c est déterminée à chaque instant d'analyse afin de minimiser le critère d'optimisation $e(t)$ défini par l'équation suivante :

$$e(t) = |SC_s(t) - SC_c(t)| \quad (6.9)$$

Une technique d'optimisation non-linéaire avec contrainte a été utilisée (fonction *fmincon* de Matlab). Cette méthode a également été employée dans [Barthet, 2004] afin d'optimiser les paramètres géométriques d'un modèle simplifié de résonateur de trompette. L'emploi d'une contrainte permet de s'assurer que la fréquence de coupure du filtre reste bien inférieure à la fréquence de Nyquist, i.e. $0 < f_c < \frac{F_e}{2}$. Comme le gradient de la fonction e ne peut pas être obtenu de manière analytique, la matrice Hessienne du Lagrangien associée à e est approchée en utilisant une méthode de différences finies. La matrice Hessienne est mise à jour selon la méthode quasi-newtonienne BFGS (qui porte le nom de ses auteurs Broyden, Fletcher, Goldfarb, Shanno) [Goldfarb, 1970]. La boucle d'optimisation est arrêtée lorsque le critère d'erreur e devient inférieur ou égal à un seuil fixé à $\epsilon = 1 \times 10^{-15}$. Une fois que la fréquence de coupure optimale f_c^* est déterminée, le segment temporel analysé est filtré avec le filtre passe-bas correspondant. La séquence modifiée finale est obtenue selon un processus de recouvrement-addition (*overlapp-add* en anglais) des différents segments temporels filtrés.

Le résultat de l'ajustement dynamique du Centre de Gravité Spectral de la séquence de clarinette générée avec l'échantillonneur (*Clc*) en fonction de celui de la séquence de clarinette expressive (notée *Clh1*) est présenté sur la figure 6.8. Le processus d'optimisation fonctionne très bien, puisque comme la figure le montre, l'erreur entre les Centres de Gravité Spectraux source et cible est quasiment nulle.

6.3.1.3 Elaboration des stimuli

De manière à constituer un groupe de stimuli homogène, les séquences initiales (*Clc*, *Pic*, *Cel*, *Gui*, *Sit*), globalement très brillantes puisqu'elles ont été obtenues en fixant le paramètre de

Stimuli	Description
Clh1	Séquence expressive de clarinette (référence)
Clc0	Séquence de clarinette (échantillonneur)
Pic0	Séquence de piccolo (échantillonneur)
Cel0	Séquence de violoncelle (échantillonneur)
Sit0	Séquence de sitar (échantillonneur)
Gui0	Séquence de guitare (échantillonneur)
Clc1	Séquence Clc + Contrôle de brillance
Pic1	Séquence Pic + Contrôle de brillance
Cel1	Séquence Cel + Contrôle de brillance
Sit1	Séquence Sit + Contrôle de brillance
Gui1	Séquence Gui + Contrôle de brillance

TAB. 6.5 – Liste et description des stimuli.

vélocité des échantillonneurs à sa valeur la plus élevée, ont également subi un filtrage passe-bas. Pour chacune de ces séquences, la fréquence de coupure du filtre passe-bas (*Butterworth*, d'ordre 2) a été déterminée par une procédure d'optimisation identique à celle présentée ci-dessus, de façon à ajuster le Centre de Gravité Spectral moyen des séquences sur le Centre de Gravité Spectral moyen de la séquence de référence. Un léger effet de réverbération a été ajouté aux différentes séquences afin de reproduire l'influence naturelle d'une salle. Toutes les séquences ont ensuite été égalisées en sonie. Le tableau 6.5 donne la description des 11 stimuli ainsi élaborés.

Les *Exemples Sonores* *n*^{os} 59, 60, et 61 correspondent respectivement à la séquence expressive de clarinette *Clh1*, à la séquence de piccolo inexpressive *Pic0*, et à la version modifiée *Pic1*, après contrôle dynamique de la brillance en fonction de la séquence expressive *Clh1*.

6.3.2 Participants

20 participants (âgés de 23 à 50 ans) ont pris part à l'expérience perceptive. L'échantillon est composé de 14 hommes et 6 femmes, qui pour la plupart proviennent de notre laboratoire. Le groupe se répartit en 15 musiciens et 5 non-musiciens. Le nombre d'années de pratique musicale des musiciens est en moyenne de 12.7 ans. Tous les participants présentent une audition normale.

6.3.3 Appareillage

L'appareillage est identique à celui des expériences perceptives précédentes. La description en est donnée à la section 4.2.4.

6.3.4 Procédure : test de jugement de dissimilarité et de préférence

La procédure de l'expérience perceptive est bâtie selon le même schéma que l'expérience décrite dans la section précédente (cf. section 6.2.4), selon trois phases (écoute des stimuli, apprentissage de la tâche et expérience). Dans cette expérience, pour chacune des paires de séquences qui leurs sont présentées, les participants doivent juger les dissimilarités d'interprétation musicale entre les deux séquences, puis sélectionner celle qu'ils préfèrent. L'expérience comprend un total de 55 paires.

Dans les consignes de l'expérience, il a bien été spécifié aux participants que l'évaluation des dissimilarités devait porter sur les différences d'interprétation musicale et non des différences de timbre liées aux changements d'instruments. Ils étaient également avertis que les séquences étaient

synchronisées temporellement. Comme pour les précédentes expériences, un questionnaire leur a été remis afin qu'ils formulent les critères utilisés pour différencier les interprétations et déterminer leurs préférences.

6.3.5 Méthodes statistiques

Certaines des méthodes statistiques utilisées pour traiter les données de cette étude ont déjà été décrites dans l'expérience impliquant des mesures de dissemblance (chapitre 4), et celle impliquant des mesures de préférences (section 6.2.4 du présent chapitre). Nous ne les présenterons donc que succinctement dans ce qui suit.

6.3.5.1 Analyse des dissimilarités

Homogénéité des réponses Dans un premier temps, l'homogénéité des réponses dans l'échantillon est analysée grâce à la classification ascendante hiérarchique¹⁰ des distances¹¹ entre les réponses des participants. Il pourra ainsi être déterminé si les réponses des non-musiciens diffèrent systématiquement de celles des musiciens.

Analyses multivariées Les jugements de dissimilarité (d'interprétation) des participants sont ensuite moyennés et traités par une analyse multidimensionnelle de proximité non-métrique (MD-SCAL). Le but de l'analyse MDS n'est pas ici de vérifier une hypothèse de continuum perceptif, mais de profiter de la réduction de données qu'offre la méthode pour observer la répartition des séquences dans un espace perceptif, dérivé des différences d'interprétation entre les séquences.

Une classification hiérarchique de la matrice de dissimilarité moyenne de l'échantillon a également été conduite afin d'analyser si les séquences peuvent être séparées en classes distinctes en fonction de leurs spécificités. Si la transformation de timbre joue réellement un rôle au niveau de la sensation d'interprétation, nous nous attendons à observer deux classes de séquences, l'une constituée par les séquences présentant des variations inexpressives de brillance (séquences dont la nomenclature porte le chiffre 0), et l'autre constituée par les séquences présentant les variations expressives de brillance (séquences dont la nomenclature porte le chiffre 1).

6.3.5.2 Analyses des préférences

Représentation des données perceptives Comme à la section 6.2.5.1, les données individuelles sont représentées par des matrices de préférence P_s , et les données globales sont représentées par la matrice de préférence de l'échantillon P .

Des notes d'interprétation sont également dérivées des préférences individuelles. Compte-tenu du fait qu'il y a 11 stimuli au total, et que les comparaisons de séquences identiques n'ont pas été effectuées, les notes d'interprétations sont nécessairement comprises entre 0 et 10.

Coefficient d'accord de Kendall Le coefficient d'accord de Kendall u et l'index associé W_t ont également été déterminés à partir de la matrice de préférence globale P afin de déterminer le degré de régularité des réponses des participants.

Influence des transformations sur les préférences Une analyse de variance (ANOVA) à un facteur à mesures répétées a été effectuée afin d'évaluer l'effet du type de séquence (11 modalités) sur les notes d'interprétation. Des tests de comparaisons multiples (Tukey, HSD "Honestly

¹⁰algorithme de lien complet.

¹¹un moins le coefficient de corrélation de Pearson.

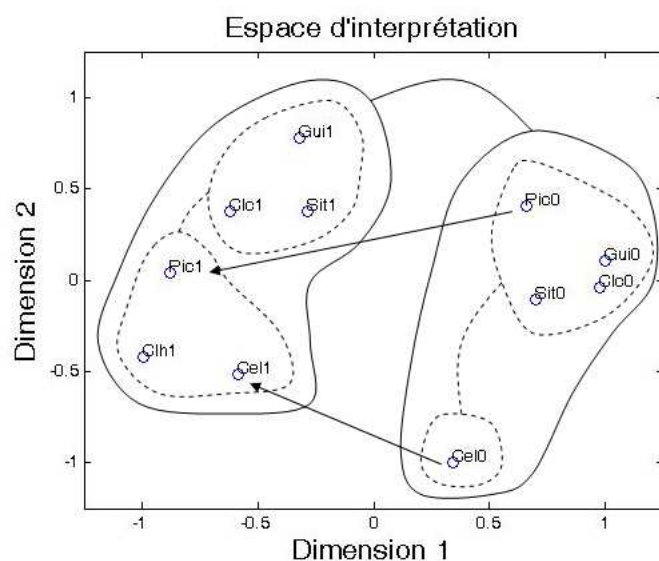


FIG. 6.9 – Projection de la configuration spatiale issue de l'analyse MDS sur ses 2 premières dimensions. Les deux premiers niveaux de regroupement de la classification ascendante hiérarchique sont superposés sur la figure (les lignes solides représentent les groupes du premier niveau, les lignes pointillées représentent les groupes du deuxième niveau). La forme des groupes n'a pas d'autre signification que celle de les délimiter. Les flèches mettent en évidence les changements de positions dans l'espace d'interprétation de la séquence de piccolo (*Pic0*) et du violoncelle (*Cel0*) lorsque le contrôle dynamique de la brillance est effectué selon le profil de variations de brillance de la séquence de clarinette expressive de référence (*Clh1*).

Significant Difference") ont ensuite été réalisés pour toutes les combinaisons 2 à 2 des différentes séquences. Ces comparaisons permettent de déterminer si la transformation de brillance induit une différence systématique de la note d'interprétation.

6.3.6 Résultats

6.3.6.1 Analyses des dissimilarités

Homogénéité des réponses Le dendrogramme associé à la classification ascendante hiérarchique des distances entre les réponses des participants est présenté sur la figure F.6 de l'Annexe F.4. Les participants non-musiciens (P6, P9, P10, P15, P17) ne se retrouvent pas dans une classe spécifique différente de celle des musiciens. Aussi, nous avons préféré conserver les données de tous les participants pour la suite des analyses.

Analyses multivariées L'analyse multidimensionnelle de proximité réalisée sur les mesures de dissimilarités moyennées sur tout l'échantillon révèle un espace perceptif d'interprétation à 3 dimensions ($stress \approx 0.07$). La courbe d'évolution du stress en fonction du nombre de dimensions, et le diagramme de Shepard associés à cette analyse MDS sont également présentés sur les figures F.7 et F.8 de l'Annexe F.4.

La projection de l'espace perceptif d'interprétation sur ses deux premières dimensions est représentée figure 6.9. La figure montre également les deux premiers niveaux de partitionnement de l'analyse en *clusters*. Le dendrogramme associé à cette dernière est présenté à la figure 6.10.

La première dimension de l'espace d'interprétation oppose les séquences inexpressives (symbolisées par "0") des séquences qui disposent des variations de brillance expressives (symbolisés

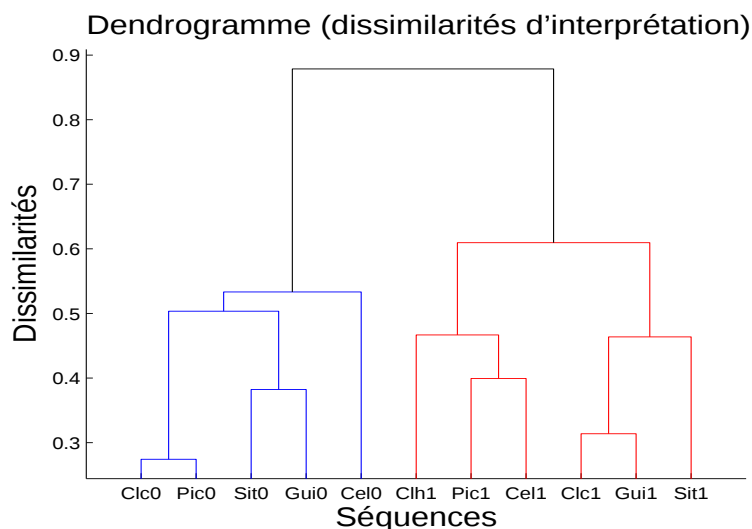


FIG. 6.10 – Représentation en dendrogramme de la classification ascendante hiérarchique des dissimilarités d'interprétation entre les séquences. L'algorithme de classification utilisé repose sur le lien complet (cf. par ex. [Saporta, 1990]).

par un "1"). L'analyse en clusters confirme cette opposition, puisque les deux groupes principaux rassemblent chacun les séquences possédant un même profil de brillance (d'un côté *Clh1*, *Pic1*, *Cel1*, *Clc1*, *Gui1*, *Sit1* et de l'autre *Clc0*, *Pic0*, *Sit0*, *Gui0*, *Cel0*, cf. figure 6.9 et 6.10). Le contrôle dynamique du timbre change donc de manière importante la nature de l'interprétation musicale.

Les deux classes principales de séquences mentionnées précédemment sont indépendantes du type d'instrument. Ceci montre que les participants ont bien répondu à la tâche puisqu'ils ne se sont pas reposés sur les dissimilarités globales de *timbre-identité* (entre différents instruments) pour évaluer les dissemblances d'interprétation, mais bien sur les modifications du *timbre-qualité*. Par exemple, bien que la clarinette et la guitare soient deux instruments aux timbres très différents, les interprétations produites sur ces instruments sont jugées très proches lorsqu'elles présentent la même évolution temporelle de la brillance (cf. la position des couples (*Clc1*, *Gui1*) et (*Clc0*, *Gui0*) dans l'espace perceptif).

L'application des variations de brillance de la séquence de référence (*Clh1*) produit des interprétations qui ressemblent davantage à celle de la référence. En effet, les séquences ayant subi la transformation de timbre sont systématiquement plus proches de la séquence de référence (*Clh1*) dans l'espace perceptif (figure 6.9). Les flèches superposées à la figure mettent en évidence cet effet pour le piccolo (*Pic0* → *Pic1*) et le violoncelle (*Cel0* → *Cel1*). L'effet est cependant moins prononcé pour les instruments à cordes pincées (la sitar et la guitare), ce sur quoi nous reviendrons plus tard, lors de l'analyse des préférences.

Comme le montre la figure 6.10, les groupes subordonnés issus de la procédure de partitionnement semblent être liés aux spécificités des instruments, lorsque ces derniers n'ont pas subi la transformation de brillance (les vents sont ensemble *Clc0* et *Pic0*, les cordes aussi *Sit0* et *Gui0*, et le violoncelle *Cel*, à part). Le fait qu'à un niveau local, les instruments de la même famille tendent à rester groupés dans l'espace d'interprétation est légitime. En effet, certains aspects des séquences musicales restent intrinsèquement liés aux spécificités acoustiques des sources. Cependant, lorsque la transformation est appliquée, ces liens tendent à se défaire : ainsi, on trouvera *Gui1*, plus proche de *Clc1*, que de *Sit1*, etc.

Nous avons vu que le contrôle de la seule dimension de brillance peut changer la perception de l'interprétation musicale. Nous allons maintenant étudier l'effet de la transformation de timbre sur les préférences musicales des auditeurs.

6.3.6.2 Analyses des préférences

Le coefficient d'accord u de Kendall calculé d'après la matrice de préférence globale de l'échantillon de 20 participants est $u = 0.21$ ($X^2 = 277, p < 0.00005, d.d.l. = 55$). L'index W_t associé est de $W_t = 0.25$. L'hypothèse que les réponses des participants ne soient pas concordantes peut donc être rejetée avec un risque $\alpha = 0.00005$. Les préférences des participants sont donc fortement corrélées. Ceci traduit le fait que la transformation de brillance induit un effet commun sur les préférences des participants.

L'évolution de l'index W_t en fonction du nombre de participants est présentée sur la figure F.5 de l'Annexe F.4. La figure montre que l'index évolue fortement lorsque le nombre de participants est faible, puis tend à se stabiliser lorsque 20 participants sont atteints.

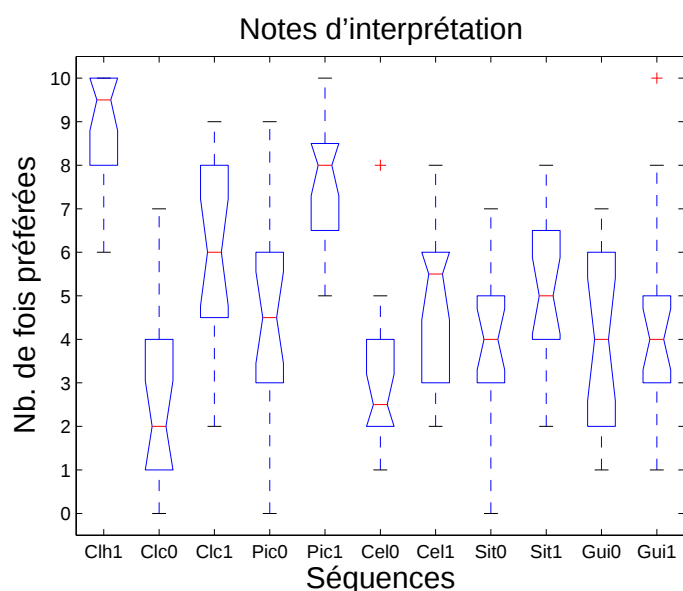


FIG. 6.11 – Diagrammes en boîte à moustaches des notes d'interprétation. Rappelons que les notes d'interprétation sont nécessairement comprises entre 0 et 10 (cf. section 6.3.5.2). La séquence expressive de référence *Clh1* est la première en partant de la gauche. Les séquences qui proviennent d'un même instrument sont ensuite placées côte à côte. Ainsi, il est rapidement possible de visualiser l'évolution des notes d'interprétation entre les séquences inexpressives (symbolisées par "0"), et les séquences ayant subi la transformation dynamique de brillance (symbolisées par "1"). Pour les informations concernant les boîtes à moustache, voir la légende de la figure 6.4.

	<i>Clc0</i>	<i>Clc1</i>	<i>Pic0</i>	<i>Pic1</i>	<i>Cel0</i>	<i>Cel1</i>	<i>Sit0</i>	<i>Sit1</i>	<i>Gui0</i>	<i>Gui1</i>
<i>Clh1</i>	13.68***	6.12***	9.68***	-	12.79***	8.68***	10.79***	7.90***	11.01***	8.79***
<i>Clc0</i>		7.57***	-	11.35***	-	-	-	5.79**	-	-
<i>Clc1</i>			-	-	6.68***	-	-	-	-	-
<i>Pic0</i>				7.34***	-	-	-	-	-	-
<i>Pic1</i>					10.46***	6.34***	8.46***	5.56***	8.68***	6.45***
<i>Cel0</i>						-	-	-	-	-
<i>Cel1</i>							-	-	-	-
<i>Sit0</i>								-	-	-
<i>Sit1</i>									-	-
<i>Gui0</i>										-

TAB. 6.6 – Résultats des tests de comparaisons multiples (Tukey) effectués pour déterminer les séquences (11 au total) qui présentent des notes d'interprétation significativement différentes. Les valeurs de la statistique q sont uniquement données lorsqu'il y a des différences significatives. Les valeurs non significatives sont indiquées par le signe $-$. $*p < .05$, $**p < .01$, $***p < .001$. Les valeurs en gras soulignent les comparaisons pour des séquences provenant d'un même instrument (*Clc0* et *Clc1*, etc.) sans et avec contrôle dynamique de la brillance.

Les distributions des notes d'interprétation sont caractérisées par le diagramme en boîte à moustaches de la figure 6.11. Le tableau 6.6 présente les résultats des comparaisons multiples (test de Tukey) effectuées entre les différentes séquences. Ces données permettent de signaler les séquences qui présentent des notes d'interprétation significativement différentes les unes des autres. La figure 6.11 et le tableau 6.6 indiquent que les modifications dynamiques de la brillance augmentent les notes d'interprétation des séquences de manière fortement significative pour la clarinette et le piccolo (cf. valeurs en gras dans le tableau 6.6). Il est remarquable de constater que la note d'interprétation de la séquence de piccolo modifiée *Pic1* (médiane = 8/10) est relativement proche de celle obtenue par la séquence de référence expressive *Clh1* (médiane = 9.5/10). Les différences de notes d'interprétation entre *Pic1* et *Clh1* ne sont d'ailleurs pas significatives (cf. tableau 6.6).

Instruments	<i>Clc</i>	<i>Pic</i>
Taux d'augmentation des notes d'interprétation (%)	200	78

TAB. 6.7 – Taux d'augmentation des valeurs médianes des notes d'interprétation lors de l'ajout du contrôle dynamique de la brillance pour la clarinette *Clc* et le piccolo *Pic*.

Les taux d'augmentation des valeurs médianes des notes entraînés par la transformation de brillance sont reportés dans le tableau 6.7 pour la clarinette et le piccolo. Ils sont exprimés comme un pourcentage d'augmentation par rapport à la note obtenue par les séquences inexpressives. Les augmentations sont fortes pour ces deux instruments auto-entretenus, et atteignent même 200% pour la séquence de clarinette *Clc1*. Ceci est probablement lié au fait que ces instruments produisent des sons entretenus de mêmes types que ceux de l'instrument de référence (instrument à vent). Les variations temporelles de timbre au cours des notes sont possibles sur ces instruments. Les auditeurs ont donc probablement trouvé les variations de timbre naturelles et en ont apprécié les effets sur l'expression musicale.

Pour le violoncelle, le changement de note d'interprétation après contrôle de la brillance n'est pas significatif, contrairement aux deux autres instruments entretenus (cf. tableau 6.6). Cependant, l'augmentation de la valeur médiane de la note d'interprétation est plus grande que pour les instruments non-entretenus (sitar et guitare). Pour ces derniers, les différences de notes d'interprétation avant et après transformation ne sont également pas significatives. Bien que les participants aient jugé les séquences *Cel0*, *Sit1* et *Guit1* comme plus proches de la référence que les versions inexpressives respectives, il est probable qu'ils aient jugé incohérente la transformation temporelle de brillance par rapport au fonctionnement acoustique de ces instruments. En effet, pour les instruments non-entretenus (guitare et sitar), le couplage entre l'excitateur et le résonateur intervient au démarrage du son, ce qui, sans effets additionnels, ne permet pas à l'instrumentiste un contrôle du timbre pendant la production des notes.

6.3.6.3 Analyse des questionnaires

Plusieurs points intéressants sont ressortis de l'analyse des questionnaires. Les participants ont souvent mentionné qu'une des séquences de l'expérience était différente des autres car elle paraissait plus naturelle (notamment de par la présence de "bruit de souffle" et "bruits de clefs"), et plus expressive. Ces observations signalent clairement la séquence naturelle de clarinette utilisée comme référence (*Clh1*). Les remarques coïncident avec les analyses statistiques : cette séquence est celle qui est en bas à gauche de l'espace d'interprétation issu de la MDS (figure 6.9) et a obtenu la note d'interprétation la plus élevée (la valeur de sa médiane est 9.5/10, cf. figure 6.11).

Elle présente effectivement les différences significatives de notes d'interprétation les plus importantes avec les autres séquences (cf. tableau 6.6). Même après application de la transformation de brillance, les autres séquences sont jugées comme étant moins expressives que cette dernière (toutefois, les différences avec la séquence de piccolo modifiée *Pic1* ne sont pas significatives). Le fait que cette séquence de clarinette ait été enregistrée en chambre anéchoïque et en champ semi-lointain, explique la présence des bruits de souffle et de clefs qui, dans le cas d'enregistrements en champ lointain dans une salle réverbérante, peuvent être masqués. Il est intéressant de noter que ces bruits auxiliaires liés au jeu de la clarinette apparaissent dans ce contexte comme un attribut positif des sons contribuant à l'appréciation du jeu instrumental.

A l'inverse, une des séquences de violoncelle a souvent été décrite comme ne paraissant pas très naturelle. Il s'agit très probablement de la séquence *Cel0* qui est classée seule dans l'un des sous-groupes déterminés par l'analyse en clusters (cf. figures 6.9 et 6.10). Elle est en effet associée à une note d'interprétation relativement faible (cf. figure 6.11). Il est possible que le filtrage passe-bas effectué afin d'ajuster la brillance moyenne des séquences ait desservi le timbre initialement riche des sons de violoncelle. Cependant, il est intéressant de constater que le contrôle dynamique de la brillance produit une interprétation qui se rapproche de celle de la référence expressive (cf. figure 6.9).

Critères		
Timbre	Dynamique	Articulation
"brillance de certains sons"	"nuance"	"Fluidité"
"qualité sonore"	"variations d'amplitudes"	"Flux"
"chaleur du son"	"intensité sonore"	"Legato"
"Attaque"		"Attaque"
"force de fluctuation du son"		
"accents, position dans le temps"		
"forme du son (développement dans le temps, rond, plat, etc.)"		

TAB. 6.8 – Expressions utilisées par les participants pour décrire les stratégies de discrimination et de préférence des séquences musicales. Elles sont regroupées en trois catégories principales : timbre, dynamique, articulation. Ces catégories ne sont pas nécessairement indépendantes. Certaines expressions peuvent donc se retrouver dans plusieurs catégories à la fois.

Trois facteurs principaux expliquant les stratégies de discrimination et de préférence des participants sont ressortis de l'analyse des verbalisations libres des participants : **le timbre, la dynamique, et l'articulation**. Les expressions relatives à ces facteurs sont répertoriées dans le tableau 6.8.

Critère de timbre Parmi les expressions ayant trait au timbre, on pourra remarquer l'emploi d'un vocabulaire provenant d'autres modalités sensorielles ("brillance", "chaleur"), ce qui se rencontre souvent dans les verbalisations liées au timbre car le vocabulaire dont nous disposons pour le décrire de manière directe est relativement pauvre¹² [Faure, 2000]. D'autres expressions telles que "forme du son (développement dans le temps, rond, plat, etc.)" ou "force de fluctuation du son" ont été placées dans le tableau à la fois en tant que critère de timbre et en tant que critère de dynamique. La transformation de timbre consiste justement à modifier la *forme temporelle* du profil d'évolution de la brillance. En même temps, cette indication peut également faire référence à une sensation de modification d'intensité.

¹²Ceci vaut tout du moins pour les cultures occidentales.

Critère de dynamique Il est en effet plusieurs expressions en lien avec la notion d'intensité. Certaines peuvent, comme nous l'avons vu, aussi référer au timbre (par ex. "accents (position dans le temps)", "force de fluctuation du son"). Il est intéressant de constater que le vocabulaire utilisé peut provenir du domaine de la musique ("nuance", "accents"), mais également du domaine de l'acoustique ("amplitude", "intensité", "force"), ce qui ne doit pas sans être lié avec le fait que les participants disposent pour la plupart de connaissances et de formations dans ces deux domaines. Si la transformation de brillance n'a que très peu affecté l'intensité, caractérisée par l'amplitude RMS, l'analyse de ces expressions permet de constater qu'il n'en est pas de même pour l'intensité perçue. En effet, les expressions réfèrent clairement à des changements d'intensité et/ou de timbre. Les expériences de Zwicker sur l'acuité¹³ ont montré que des modifications de cette dernière pouvait modifier la sensation de gêne bien que les sons disposaient tous de la même sonie [Zwicker et Fastl, 1990]. Sans aller jusqu'à la notion de gêne dans le cadre de notre expérience, il semble que les modifications du Centre de Gravité Spectral aient modifié la sensation de sonie.

Critère d'articulation D'autres verbalisations renvoient à la notion musicale d'articulation (par ex. "fluidité", "legato"). Comme pour la dynamique, il semble donc que la modification temporelle du timbre n'induit pas qu'un changement perceptif de qualité sonore mais affecte également la dimension temporelle, si importante en musique. L'ajout de variations de brillance provenant d'une séquence de clarinette expressive semble améliorer les liaisons entre les notes et généralement produire une interprétation plus homogène. Ceci n'est pas étonnant compte-tenu du fait que les séquences inexpressives ont été générées par des échantillonneurs, et sont donc, dans une certaine mesure, la juxtaposition de sons isolés n'ayant pas de liens musicaux les uns avec les autres.

Malgré leur expérience musicale, certains participants ont signalé avoir trouvé l'expérience difficile. L'évaluation de l'interprétation musicale a déjà été signalée comme problématique ; dans [Whybrew, 1962], l'auteur stipule que dans une telle tâche les mesures employées sont subjectives et les jugements irréguliers même lorsque les évaluateurs sont des professeurs de musique professionnels ("the measures employed are typically subjective judgments based on irregular and uncontrolled observations"). Face à cette problématique certains auteurs tâchent de développer des échelles d'évaluation d'interprétation musicale reposant sur un ensemble de dimensions pré-établies [Abeles, 1973]. Dans notre cas, au dépend de la difficulté de la tâche, nous avons préféré ne pas utiliser d'échelles d'évaluation reposant sur un ensemble de descripteurs musicaux pré-définis ou d'échelles nominales à plusieurs niveaux ("différentiel sémantique") afin de ne pas biaiser la tâche d'évaluation par des connaissances *a priori*. Les difficultés éprouvées par certains participants pour effectuer les jugements de dissimilarité sont aussi probablement dues à la difficulté d'utiliser l'échelle de mesure ("Très semblables- -Très dissemblables") sur toute sa dynamique. En effet, les deux groupes principaux révélés par la classification ascendante hiérarchique montrent que d'une manière générale, les séquences ont soit été jugées très proches, soit très différentes, situations qui correspondent aux extrémités de l'échelle de mesures des dissimilarités. Ceci est du reste cohérent avec la façon dont les stimuli ont été générés dans cette expérience (séquences avec ou sans transformations). La complexité de la tâche est peut-être également liée au fait que les séquences sont relativement longues (environ 10-15 s). Les comparaisons de sons de longue durée sont moins simples que celles de sons de courtes durées. Rappelons que pour cette raison, les participants avaient la possibilité d'écouter autant de fois qu'ils le souhaitent, les séquences musicales d'une même paire. Supposons que nous sachions exactement comment "l'oreille" intègre l'information qu'elle reçoit au cours du temps, le jugement final peut être fortement influencé par les derniers événements entendus, en raison du processus de mémorisation [Susini et McAdams, 2000]. Un tel effet de récence n'a été constaté que pour un participant (non-musicien) qui a reporté avoir été plus sensible aux trois dernières notes de la séquence musicale.

¹³ Descripteur qui est analogue au Centre de Gravité Spectral, mais défini selon un modèle physiologique de l'audition.

Les verbalisations libres des participants donnent ainsi des clefs intéressantes pour comprendre les stratégies qu'ils ont utilisées afin de discriminer les interprétations et déterminer leurs préférences. Il est intéressant de voir que la modification temporelle de brillance n'affecte pas seulement une dimension sonologique liée au timbre, mais s'étend à d'autres dimensions musicales, telles que la dynamique ou l'articulation. Ceci semble corrélé avec le fait que la transformation agit sur les variations (*dynamiques*) du Centre de Gravité Spectral au cours du *temps*. D'après les points de vue des auditeurs et les résultats des analyses statistiques, le contrôle dynamique de brillance a globalement eu pour effet de transformer des interprétations inexpressives non homogènes et avec peu de variations en des interprétations plus expressives qui présentent davantage de variations, qui sont plus homogènes, et pour lesquelles l'articulation entre les notes est améliorée.

6.3.7 Discussion

Dans cette expérience, nous avons étudié l'effet d'appliquer les variations temporelles de brillance d'une interprétation expressive de clarinette à des séquences "neutres" ou inexpressives, générées avec des échantillonneurs. Les résultats de l'étude perceptive montrent que pour différents instruments, les auditeurs ont très bien discriminé les séquences en fonction de la nature de leur profil de variations temporelles de brillance. Les interprétations qui en moyenne ont été jugées les plus expressives sont celles ayant été traitées par le contrôle dynamique de brillance. Les améliorations de l'expressivité se sont avérées significatives pour deux des trois instruments entretenus utilisés dans l'expérience (clarinette, piccolo), instruments pour lesquels un contrôle du timbre pendant la production des notes est possible par l'interprète. Les faibles améliorations de la transformation de brillance pour les instruments non entretenus (sitar, guitare) tendent à montrer que cette dernière n'est pas adaptée à ces instruments, peut-être par désaccord avec leur fonctionnement acoustique, ce dernier n'offrant pas à l'interprète, dans des conditions normales de jeu, la possibilité de contrôler le timbre des sons lors de leur production¹⁴. Pour les instruments non-entretenus, les auditeurs ont sans doute perçu les séquences modifiées comme étant non naturelles, ce qui expliquerait pourquoi ces séquences n'ont pas été jugées meilleures que les séquences initiales. En effet, comme le souligne Lerdahl, les sons qui sont identifiables tendent à être plus appréciés que les sons auxquels nous ne sommes pas habitués [Lerdahl, 1991].

Ces résultats montrent que pour certains instruments entretenus, si l'on fait évoluer la brillance des notes d'une séquence musicale de façon appropriée au cours du temps, il est virtuellement possible d'améliorer l'expressivité de l'interprétation musicale. Sundberg et ses collaborateurs ont donné un exemple par la synthèse soulignant l'importance des subtiles variations rythmiques produites par les interprètes lors d'interprétations expressives. En effet, ces derniers montrent que lorsque l'on inverse la nature des variations rythmiques pour chaque note (un rallongement devenant un raccourcissement, et vice versa), cela produit des interprétations musicales complètement incongrues et absurdes. En s'inspirant de cette démonstration mais en l'appliquant au timbre, nous avons généré des séquences musicales pour lesquelles les variations de brillance ont été inversées au cours du temps pour chacune des notes de l'extrait (une augmentation de brillance devenant une diminution, et vice versa). Les Exemples Sonores n^{os} 62 et 63 correspondent respectivement à une séquence normale de clarinette, et à une modification de cette séquence avec les variations de brillance inversées au cours du temps pour chaque note de la séquence. Le résultat est clairement anti-musical et semble même opposé à un jeu naturel. Ceci est un exemple démontrant l'importance des formes temporelles du timbre musical.

¹⁴Notons d'ailleurs que l'impossibilité, à la guitare, de modifier la brillance des sons pendant leur production n'est peut-être pas étrangère au succès couru par la fameuse pédale d'effet "wah-wah". De fait, cet effet agit sur le spectre au moyen d'un filtre passe-bande, dont on peut contrôler la fréquence centrale, ce qui permet à l'instrumentiste de produire des variations de brillance en temps-réel. Bien entendu, si cet effet s'avère être fort approprié à certains styles musicaux (par ex., funk, blues), il l'est sans-doute bien moins pour d'autres.

Chapitre 7

Conclusions et perspectives

Les travaux rapportés dans cette thèse constituent une première exploration du rôle du timbre dans l'interprétation musicale. L'acte d'interprétation s'inscrit dans le cadre d'une chaîne de communication musicale dont les acteurs principaux sont le compositeur et l'interprète, à un niveau poïétique (création), et l'auditeur, à un niveau esthétique (perception)¹. Aussi, la question du timbre en musique a été abordée tour à tour sous l'angle poïétique (*comment le timbre est produit et manipulé par l'interprète ?*) et sous l'angle esthétique (*quel est son effet sur l'auditeur ?*).

Notre démarche s'est organisée selon les trois étapes suivantes :

- l'étude de l'influence des paramètres de contrôle d'un instrument numérique (clarinette) sur son timbre (chapitre 4). Cette étape s'est réalisée au travers de la détermination de l'espace de timbre perceptif de l'instrument, et son interprétation à l'aide de facteurs psychoacoustiques, corrélés aux différentes dimensions de l'espace ;
- les mesures des variations de timbre pour différents types d'interprétations musicales, afin d'étudier si les manipulations du timbre produites par l'interprète sont en lien avec son intention musicale (chapitre 5) ;
- l'évaluation des effets perceptifs induits par l'une des dimensions du timbre (la brillance) sur la qualité musicale, afin d'évaluer si le timbre peut être considéré pour l'auditeur comme un vecteur d'expression musicale (chapitre 6).

7.1 Espace de timbre de la clarinette

La première expérience analyse des distances perceptives entre les timbres de sons synthétiques de clarinette, obtenues à l'aide de jugements de dissemblance. Un espace de timbre à 3 dimensions semble bien exprimer la structure de la représentation perceptive de timbres produits pour différentes combinaisons de pressions et pinces. La première dimension est fortement corrélée à la nature de l'attaque : sa durée (Logarithme du Temps d'Attaque LAT [$r(13)=-0.95$, $p<0.001$], Temps d'Attaque AT [$r(13)=-0.93$, $p<0.001$]), la non-synchronicité des composantes harmoniques (Flux Spectral de l'attaque SFAT [$r(13)=-0.95$, $p<0.001$]), mais également à plusieurs aspects de la distribution spectrale : sa fréquence moyenne, le Centre de Gravité Spectral SC [$r(13)=0.94$, $p<0.001$], et un indicateur de la richesse spectrale, le Tristimulus TR3 [$r(13)=0.95$, $p<0.001$]. La deuxième dimension s'est avérée être corrélée avec des facteurs rendant compte, par rapport aux sons analysés, de l'évolution de l'énergie dans les fréquences moyennes (Tristimulus TR2 [$r(13)=0.67$, $p<0.01$]), Etalement Spectral SS [$r(13)=-0.63$, $p<0.05$]). La troisième dimen-

¹Les niveaux poïétique, neutre et esthétique sont les trois pôles de la tripartition de Molino et Nattiez, définie par ces derniers dans le cadre de la sémiologie musicale (cf. section 1.2).

sion est fortement corrélée avec le Rapport Impair/Pair OER [$r(13)=0.87, p<0.001$], et le Centre de Gravité Temporel TC [$r(13)=0.80, p<0.001$]. Il est intéressant de constater que certains de ces descripteurs (le Logarithme du Temps d'Attaque, et le Centre de Gravité Spectral) se sont souvent avérés être de bons prédicteurs des différences de timbre s'opérant à un autre niveau hiérarchique, entre des sons provenant d'instruments différents (cf. par ex. espace de timbre de McAdams *et al.* décrit à la section 2.2.4.2). D'autres corrélats acoustiques de l'espace perceptif sont reliés à des facettes du timbre plus spécifiques à la clarinette (par ex. le Rapport Impair/Pair, qui rend compte des différences de comportement entre les composantes harmoniques impaires et paires). Sur le plan de leurs définitions, le Temps d'Attaque et le Centre de Gravité Spectral sont des facteurs indépendants. Au niveau de traitement associé à une tâche de jugement de dissemblance (i.e. au terme de l'analyse sensorielle), ils semblent également être indépendants [Caclin, 2004]. Dans notre expérience, le Temps d'Attaque et le Centre de Gravité Spectral corrélaient tous deux à la même dimension de l'espace perceptif, mais ceci provient du fait que ces deux facteurs covarient (en sens inverse) pour les stimuli utilisés. En effet, la classification ascendante hiérarchique des dissimilarités a révélé deux grands groupes de sons, l'un correspondant à des sons de type *piano*, à l'attaque douce (AT long) et au timbre "sourd" (SC faible), l'autre à des sons de type *mezzo forte* ou *forte*, à l'attaque brève (AT faible), et au timbre "brillant" (SC élevé). Il est certainement possible de produire des variations de timbre à la clarinette en jouant de manière indépendante sur l'attaque ou la brillance. Cette hypothèse semble d'autant plus justifiée lorsque l'on considère l'aspect morphologique temporel du timbre. En effet, un son soutenu, stable, et un son joué crescendo pourraient très bien présenter tous deux le même type d'attaque, mais disposer d'une évolution temporelle de brillance complètement différente. Aussi, nous avons conservé à la fois le Temps d'Attaque, et le Centre de Gravité Spectral, afin de caractériser le timbre de la clarinette dans la suite de nos expériences.

L'étude de dissemblance a aussi permis de mettre en lumière une facette du timbre spécifique aux instruments à anche simple, la situation d'anche battante. En effet, des analyses de sons naturels et synthétiques ont montré que le Rapport Impair/Pair, fortement corrélé à la troisième dimension de l'espace perceptif, présente une "cassure" dans son évolution lorsque l'anche de l'instrument commence à battre contre la table du bec. Les théories physiques sur le couplage non-linéaire entre l'excitateur et le résonateur (cf. lois de Worman-Benade) ne permettent pas à ce jour de prédire quantitativement le comportement des composantes harmoniques pour des grandes oscillations, associées à des pressions d'alimentation élevées. Une théorie physique démontrera peut-être un jour quantitativement la cassure du rapport Impair/Pair que nous avons constaté lorsque l'anche commence à battre. Toujours est-il qu'à un niveau qualitatif, la physique explique bien ce phénomène. En effet, la situation d'anche battante entraîne l'annulation du débit dans l'embouchure sur des demi-périodes, ou presque, d'oscillation de la vibration sonore, ce qui ajoute une non-linéarité supplémentaire dans le couplage entre l'excitateur et le résonateur. L'ajout de cette non-linéarité entraîne l'apparition de toutes les composantes harmoniques dans le débit, et notamment celle des harmoniques pairs (le phénomène de fermeture du canal d'anche qui a lieu sur environ une demi-période de la vibration fondamentale de fréquence f_1 , induit des composantes dont la fréquence la plus basse est le double de celle de la vibration fondamentale $2f_1, 4f_1$, etc.). Ceci renforce donc le poids relatif des harmoniques pairs, peu présents aux faibles oscillations. En raison du rayonnement, les composantes harmoniques du débit (donc pairs et impairs) influencent, en plus de celles de la pression (essentiellement impairs), le spectre externe de l'instrument. Cette description qualitative du phénomène explique la cause de la cassure observable sur le Rapport Impair/Pair OER.

L'analyse des jugements de dissemblance de timbre nous a permis de déterminer des descripteurs traduisant de manière fidèle la façon dont nous percevons les principales variations de

timbre produites par l'instrument pour différents paramètres de contrôle² (par ex. Temps d'Attaque, Centre de Gravité Spectral, Rapport Impair/Pair). Le Rapport Impair/Pair s'est révélé être un bon indicateur du net changement de timbre de l'instrument qui se produit lors du passage à la situation d'anche battante.

7.2 Mesures des variations de timbre

Les analyses sonologiques d'interprétations musicales de clarinette jouées par un même musicien ont permis de mettre en évidence les deux résultats suivants :

- *Un clarinettiste (expert) reproduit de manière systématique certaines variations de timbre (Temps d'Attaque, Centre de Gravité Spectral, Rapport Impair/Pair) lors de répétitions d'interprétations produites selon la même intention musicale.*
- *La nature des variations de timbre change lorsque l'intention expressive de l'interprète change (par ex. brillance plus élevée, Temps d'Attaque plus long, etc.).*

Ces résultats tendent donc à prouver que les musiciens agissent sur certaines dimensions du timbre afin de varier leur expression, au même titre qu'ils agissent sur les aspects temporels (par ex. durées et placement temporel des notes), ou les aspects liés à l'intensité. Les facettes du timbre concernées relèvent tant d'aspects globaux caractéristiques d'une note (par ex. Temps d'Attaque), que d'aspects morphologiques temporels, caractéristiques d'une forme (fruit de la variation du timbre-qualité durant la production des notes, par ex. évolution du Centre de Gravité Spectral à court-terme).

7.3 Analyse perceptive de l'effet des variations de brillance

L'analyse par la synthèse des effets du gel des variations de brillance dans un contexte musical a révélé la forte sensibilité des auditeurs à ce type de réductions. Il s'avère que la transformation induit non seulement des modifications de timbre-qualité, mais réalise aussi, probablement, une translation du timbre-identité de l'instrument. En effet, même si les fluctuations spectrales sont moins importantes pour les sons de clarinette que pour les sons d'autres instruments, les composantes harmoniques de sons de clarinette détiennent chacune une dynamique temporelle qui leur est propre. Accorder cette dynamique individuelle selon une seule et même référence permettant de conserver l'amplitude RMS est un processus très réducteur. L'expérience montre que cette transformation diminue fortement les préférences musicales des auditeurs, bien plus que ne l'induit un jeu *métrologique* (pas de déviations rythmiques par rapport à la partition), et/ou limité en nuances. L'analyse des verbalisations libres des auditeurs au sujet de leurs critères décisionnels met en lumière une opposition entre des sons "statiques" et des sons "vivants", traduisant au vu des préférences des auditeurs, l'importance des *variations* temporelles de brillance pendant la production des notes.

Ces résultats sont confortés par ceux de la deuxième expérience pour laquelle les manipulations de timbre ne visent non pas à ôter aux sons originels leurs variations de brillance, mais à sculpter ces derniers au niveau spectral selon des formes temporelles de brillance spécifiques à un jeu expressif. Les résultats de l'analyse multidimensionnelle de proximité des degrés de similarité

²Les sons considérés dans cette expérience correspondent à des conditions traditionnelles de jeu. Mais, il est clair que la palette de timbres de l'instrument s'étend à d'autres types de sons comme : les canards, les sons qui quintoient, les sons multiphoniques, des effets tels que le Flatterzhunge (cf. [Exemple Sonore n° 51](#)), etc.

perçus entre les interprétations, et les analyses de variance des préférences des auditeurs montrent les deux points suivants :

- *Des interprétations musicales qui présentent des variations temporelles de brillance similaires tendent à être jugées proches, et ce indépendamment du type d'instrument (entretenu et non-entretenu).*

- *Un post-traitement consistant à contrôler dynamiquement les évolutions temporelles de brillance des notes d'une séquence musicale peut, dans un certain contexte, conduire à des améliorations significatives de sa qualité musicale (ceci a été constaté lorsque les variations de brillances sont contrôlées à partir de variations extraites d'un jeu expressif, et pour des séquences initiales générées à partir d'échantillonneurs fonctionnant avec des instruments de type entretenu).*

Lors d'une écoute musicale, le percept d'évolution de la brillance au cours du temps semble donc être déterminant pour l'interprétation dans le cas d'instruments qui permettent de contrôler cette dimension du timbre au cours du temps de manière naturelle (i.e. en adéquation avec le fonctionnement acoustique de l'instrument).

Les résultats de ces expériences légitiment la notion de *variations expressives de timbre*, puisqu'au niveau de l'auditeur, sous certaines conditions, elles peuvent entraîner l'amélioration des préférences musicales.

7.4 Lien entre les trois chapitres expérimentaux

L'étude de dissemblance de sons de clarinette du chapitre 4 prouve que le contrôle de l'instrument permet de réaliser des modifications de timbre, et que celles-ci sont discriminées par les auditeurs. De plus, elle nous a permis de dégager un certain nombre de descripteurs sur lesquels nous avons porté notre attention plus spécifiquement dans les analyses d'interprétations musicales. Les expériences du chapitre 5 ont montré que l'interprète agit sur le timbre pour jouer de manière expressive, mais les mesures effectuées n'indiquent pas si les variations de timbre sont importantes sur le plan perceptif dans un contexte musical. Les expériences perceptives du chapitre 6 ont permis de conclure sur ce point, en montrant que certaines des variations de timbre mises en évidence par la mesure, sont également très importantes du point de vue perceptif. De plus, ces expériences indiquent que les variations de timbre agissent sur la qualité de l'interprétation musicale pour d'autres types d'instruments que la clarinette, notamment ceux de type entretenu.

7.5 Perspectives

Les travaux sur l'analyse de l'interprétation musicale ont montré qu'il pouvait y avoir des différences significatives de timbre entre un jeu scolaire et un jeu expressif. Une première étape a été menée vers la caractérisation de la nature de ces différences à l'échelle de notes ou de groupes de notes (Centre de Gravité Spectral moyen SCM, ambitus du Centre de Gravité Spectral SCA, etc.). Il serait intéressant d'analyser des extraits musicaux plus longs afin d'analyser si des différences de timbre peuvent également être mises en évidence sur une échelle temporelle plus longue, par ex. entre le début et la fin des phrases musicales d'une pièce, entre plusieurs mouvements, pour un "pont" dans la musique rock, etc.). Il se peut en effet que les variations de timbre soient utilisées à un *méta-niveau* par les interprètes afin de marquer une différence d'expression entre deux passages différents.

D'autre part, en vue de l'élaboration de modèles prédictifs des variations de timbre reposant sur la structure musicale, il conviendrait d'analyser une large variété d'extraits musicaux et faire un rapprochement systématique des analyses avec la structure musicale. L'une des règles du modèle d'interprétation musicale décrit dans [Sundberg *et al.*, 1989] (*règles du KTH*) est baptisée par les auteurs *charge mélodique*. Cette dernière traduit le degré de singularité d'une note dans une pièce tonale, en se reposant sur le cycle des quintes. Selon cette règle, les singularités des notes de charges mélodiques élevées sont amplifiées par des opérations portant sur le niveau sonore, la durée, l'amplitude du vibrato. Nous nous permettons de spéculer sur le fait que la prise en compte de la brillance en plus de ces autres paramètres, permettrait probablement d'améliorer l'effet souhaité pour certains instruments (par ex. violoncelle, clarinette, etc.).

Nous nous sommes concentrés dans ce travail sur des répertoires musicaux traditionnels. Nous présumons que l'utilisation des variations de timbre par les musiciens revêt un caractère général qui n'est pas spécifique à un genre musical particulier. Il serait très intéressant d'analyser des séquences musicales de jazz par exemple, musique dont on sait que le timbre (peut être devrait-on parler plus généralement de sonorité) des solistes tient une place très importante.

Nos travaux se sont reposés sur les interprétations produites par *un seul interprète*. La comparaison d'interprétations d'un même morceau par *différents interprètes* (jouant d'un même instrument) permettrait d'étudier si les variations de timbre peuvent expliquer les similitudes ou au contraire les différences entre plusieurs interprètes (le cas frappant est bien entendu celui de la voix, puisque chacun d'entre nous dispose d'un timbre-identité qui lui est propre). Il conviendra dans une telle entreprise de décider si les interprètes doivent tous jouer sur le même instrument, auquel cas les analyses révéleraient des variations de timbre "concrètes" (liées au contrôle instrumental)³, ou s'ils peuvent employer leurs propres instruments, auquel cas les analyses révéleraient à la fois des variations de timbre "concrètes", et des variations de timbre "abstraites" (liées à la facture instrumentale). Concernant les sujets qui viennent d'être évoqués, un futur axe de recherche intéressant serait de développer des applications permettant d'identifier de manière automatique le style d'un interprète, à partir de critères liés aux variations de timbre.

D'autre part, les mesures des variations de timbre rapportées dans cette thèse n'ont été effectuées que sur *un seul instrument* (la clarinette). Afin d'identifier d'éventuels caractères généraux de l'utilisation du timbre comme critère d'expression, nous pourrions également mener des analyses d'interprétations musicales d'une même pièce, produites sur *différents instruments* (et *a fortiori* par différents interprètes), et comparer les évolutions de timbre entre elles. En ne restreignant pas non plus le timbre à une seule source sonore causale (par ex. un instrument), nous pourrions également envisager d'analyser le timbre naissant de la fusion des flux sonores produits par un ensemble d'instruments (par ex. le timbre de l'orchestre). Rememorons-nous à ce sujet l'*Exemple Sonore n° 1*, extrait de *Farben*, où Schönberg manipule subtilement les voix afin de donner naissance à de véritables *mélodies de couleurs*...

Enfin, un axe de recherche se concentrant sur le contrôle du timbre afin d'agir sur l'expression musicale semble revêtir un intérêt immédiat à la lumière des travaux exposés dans cette thèse. Un tel axe de recherche disposerait d'un intérêt d'ordre fondamental, car il permettrait le développement d'outils permettant de mieux comprendre les traitements associés à la perception musicale. D'autre part, il présenterait également un intérêt appliqué fort, en regard de la conception de nouveaux outils numériques d'aide à la création musicale, ou de celle d'effets audio-numériques temps-réel permettant au musicien d'explorer de nouvelles formes temporelles de timbre.

³Introduits par Schaeffer, les concepts de variations "concrètes" et "abstraites" de timbre réfèrent respectivement aux variations de timbre dues à la facture instrumentale, et à celles dues au contrôle instrumental du musicien (cf. section 2.1.3).

Annexe A

Liste des Exemples Sonores

Les Exemples Sonores mentionnés dans le document sont contenus dans le Disque Compact qui l'accompagne. Les numéros de pistes auxquels ils sont associés sont reportés dans les tableaux qui suivent. D'autre part, dans la version PDF¹ de ce manuscrit de thèse, des hyperliens permettent de déclencher automatiquement la lecture des Exemples Sonores. Il faut pour cela créer dans le répertoire contenant le fichier PDF, un sous-répertoire intitulé /sons, qui contient les Exemples Sonores (au format WAV²), avec la nomenclature suivante : *TheseMB_ExempleSonore_xx*, où xx indique le numéro du son (i.e. 01, 02, etc.).

Piste	Description de l'Exemple Sonore
1	Extrait de <i>Farben</i> (A. Schönberg)
2	Extrait de <i>Chronochromie</i> (O. Messiaen)
3	Extrait de <i>Mutations</i> (J.-C. Risset)
4	Son complexe avec différence de fréquence entre composantes fixe
5	Son complexe avec rapport de fréquence entre composantes fixe (quarte)
6	Son complexe inharmonique (de type gong)
7	Son complexe harmonique
8	Extrait d' <i>El cant del boier</i> (chorale Les Voix Là)
9	Séquence musicale de shamisen (extraite de <i>Kageboshi</i> , Satoh)
10	Séquence musicale de ko-tsuzumi (musique de Théâtre Nô, <i>Shakkyô : Le Pont en pierres</i>)
11	Séquence musicale de shakuhachi (extraite de <i>Kokuu</i> , Yonsun)
12	Extrait du <i>Concerto</i> de Mozart (interprète : J. Brymer)
13	Extrait du <i>Concerto</i> de Mozart (interprète : B. Goodman)
14	Extrait de l' <i>Allemande de la Suite II</i> de Bach (violoncelle MIDI)
15	Extrait de l' <i>Allemande de la Suite II</i> de Bach (avec les règles de <i>Director Musices</i>)
16	Extrait de l' <i>Allemande de la Suite II</i> de Bach (M. Maisky)
17	Séquence musicale de guitare (1) avec variations de timbre entre les notes (M. Barthet)
18	Séquence musicale de guitare (2) avec variations de timbre entre les notes (M. Barthet)
19	Séquence musicale de clarinette (extraite de <i>Nails on the head</i> , N. Rothenberg)
20	Séquence musicale chantée (extraite de <i>Tom's dinner</i> , S. Vega)
21	Transposition de l'Exemple Sonore 20 sans conservation du timbre (octave inférieur)
22	Transposition de l'Exemple Sonore 20 sans conservation du timbre (facteur = 1.2)
23	Transposition de l'Exemple Sonore 20 sans conservation du timbre (octave supérieur)
24	Extrait de <i>Miles</i> (M. Davis)
25	Lecture inversée de l'Exemple Sonore 24

¹Portable Document Format.

²WAVEform audio format.

Les Exemples Sonores mentionnés dans le document sont contenus dans le Disque Compact qui l'accompagne.

Piste	Description de l'Exemple Sonore
26	Son de clarinette (C. Crousier)
27	Resynthèse de l'Exemple Sonore 26 avec Centre de Gravité Spectral gelé au début
28	Resynthèse de l'Exemple Sonore 26 avec Centre de Gravité Spectral gelé à la fin
29	Signal complexe de forme d'onde carrée (harmoniques impairs en phase)
30	Signal complexe de forme carrée (harmoniques impairs hors phase)
31	Signal complexe avec enveloppe spectrale plate
32	Signal complexe avec enveloppe spectrale en forme de "chapeau"
33	Long crescendo de clarinette (C. Crousier)
34	Modèle de synthèse de clarinette : crescendo
35	Modèle de synthèse de clarinette : séquence musicale (avec "canard numérique")
36	Son de synthèse de clarinette : stimulus S2 (pression faible, et pince forte)
37	Son de synthèse de clarinette : stimulus S13 (pression très élevée, et pince très forte)
38	Son de synthèse de clarinette : stimulus S8 (pression moyenne, et pince faible)
39	Son de synthèse de clarinette : stimulus S15 (pression très élevée, et pince faible)
40	Son naturel de clarinette : situation d'anche battante (extrait de l'Exemple Sonore 33)
41	Extrait de l' <i>Allemande</i> de Bach - interprétation scolaire P5 (C. Crousier)
42	Extrait de l' <i>Allemande</i> de Bach - interprétation scolaire P18 (C. Crousier)
43	Extrait de l' <i>Allemande</i> de Bach - interprétation expressive P22 (C. Crousier)
44	Extrait de l' <i>Allemande</i> de Bach - interprétation expressive P27 (C. Crousier)
45	Extrait du <i>Larghetto du Quintette</i> de Mozart - interprétation scolaire P1 (C. Crousier)
46	Extrait du <i>Larghetto du Quintette</i> de Mozart - interprétation scolaire P2 (C. Crousier)
47	Extrait du <i>Larghetto du Quintette</i> de Mozart - interprétation expressive P4 (C. Crousier)
48	Extrait du <i>Larghetto du Quintette</i> de Mozart - interprétation expressive P5 (C. Crousier)
49	Extrait de l' <i>Allemande</i> de Bach - interprétation expressive P22 - Micro <i>Neumann</i>
50	Extrait du <i>Larghetto</i> de Mozart - interprétation expressive P5 - Micros <i>SD Systems</i>
51	Séquence de clarinette : effet <i>Flatterzunghe</i> (C. Crousier)
52	Phrase d'ouverture du <i>Larghetto du Concerto</i> de Mozart
53	Resynthèse de l'Exemple Sonore 52 avec résidu (modèle de synthèse SAS)
54	Résidu associé à la resynthèse de l'Exemple Sonore 52 (modèle de synthèse SAS)
55	Resynthèse de l'Exemple Sonore 52 avec $f_h(t) = h \times f_0(t)$ (modèle de synthèse SAS)
56	Resynthèse de l'extrait de Bach, sans transformations (M_0)
57	Resynthèse de l'extrait de Bach, avec gel du Centre de Gravité Spectral (M_T)
58	Resynthèse de l'extrait de Bach, avec variations rythmiques supprimées (M_R)
59	Extrait de l' <i>Allemande</i> de Bach - interprétation expressive
60	Séquence de piccolo inexpressive (piccolo : échantillonneur MIDI)
61	Séquence de piccolo après contrôle dynamique de la brillance
62	Extrait de l' <i>Allemande</i> de Bach (clarinette : échantillonneur MIDI)
63	Séquence de clarinette avec variations temporelles de brillance inversées

Annexe B

Echelle, tempérament et perception des hauteurs

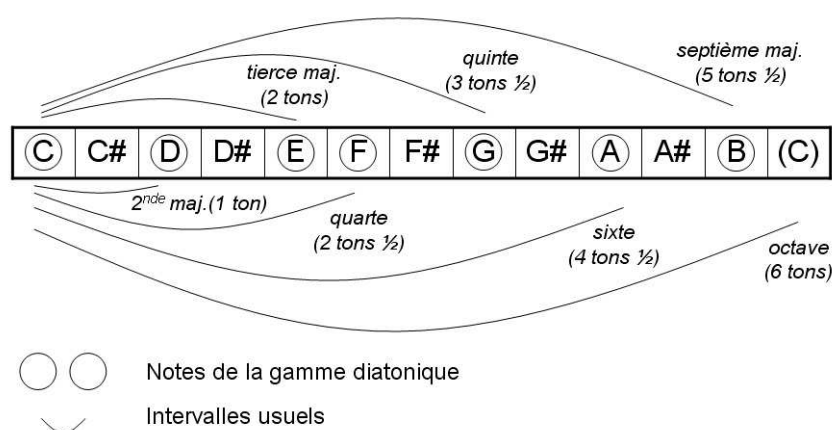


FIG. B.1 – Echelle chromatique et gamme diatonique tempérées.

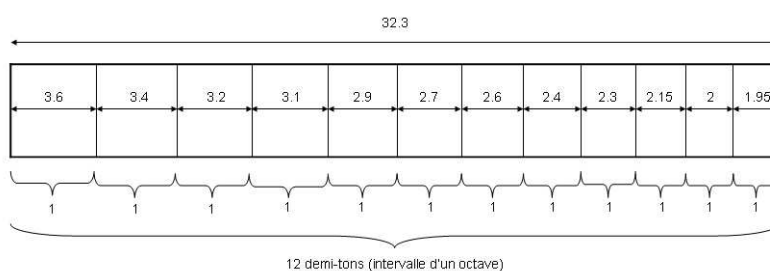


FIG. B.2 – Espacement des touches de la guitare au sein d'un octave. Les séparations verticales correspondent aux frettes, fines barres métalliques incrustées dans le manche afin de délimiter les touches. Les largeurs des touches sont données en cm. Les mesures ont été réalisées sur une guitare classique Yamaha G-245S.

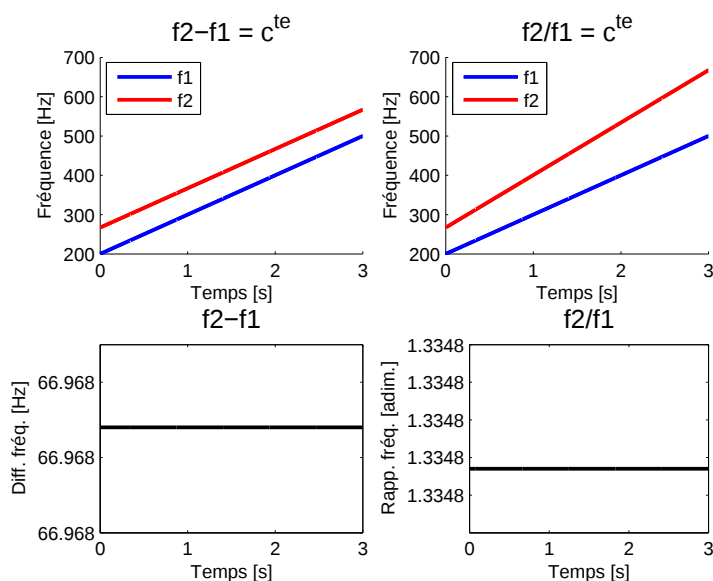


FIG. B.3 – Isomorphisme d'intervalle de hauteurs. Deux composantes de fréquences respectives f_1 et f_2 sont espacées de façon à former un intervalle de quarte (5 demi-tons tempérés, soit un rapport de fréquence de $2^{5/12}$) à l'instant initial du son. Les fréquences des composantes évoluent ensuite linéairement au cours du temps suivant deux procédés : à gauche, les différences de fréquences entre les composantes sont préservées (cf. [Exemple Sonore n° 4](#)), et à droite, les rapports de fréquences entre les composantes sont préservés (cf. [Exemple Sonore n° 5](#)). L'écoute des exemples sonores correspondants montre que la conservation des rapports de fréquences entraîne un isomorphisme des intervalles au niveau perceptif.

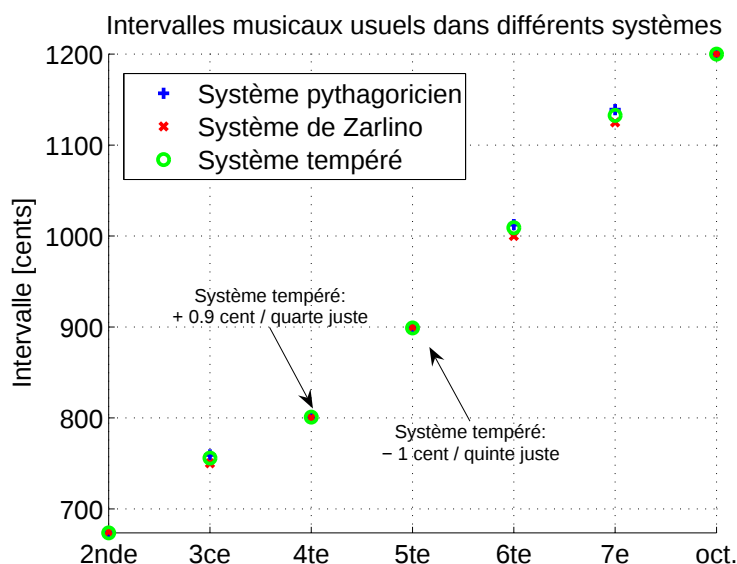


FIG. B.4 – Comparaison des intervalles usuels (exprimés en cents) dans les systèmes de Pythagore, de Zarlino et tempéré. Le cent est une unité d'intervalle définie telle que 1200 cents correspondent à un octave. Un demi-ton tempéré vaut donc 100 cents.

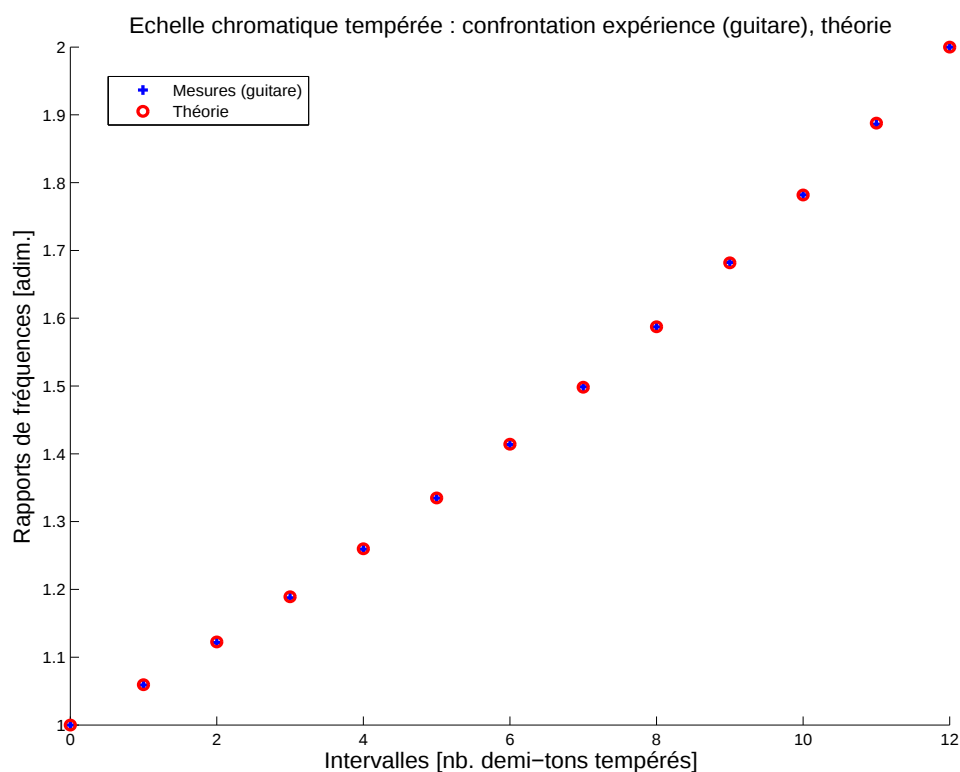


FIG. B.5 – Intervalles de l'échelle chromatique tempérée : confrontation entre les intervalles théoriques (loi $2^{d/12}$ où d est le nombre de demi-tons tempérés), et ceux déduits de la facture d'un instrument tempéré, la guitare, à partir de la largeur des touches (mesures réalisées sur une guitare classique Yamaha G-245S, cf. figure B.2). Notons que dans cette comparaison, il est supposé une relation linéaire, inversement proportionnelle entre la longueur de corde vibrante et la fréquence fondamentale de la note émise (cas d'une corde idéale), et que la fréquence du premier mode de vibration transverse de la corde n'est pas modifiée lors de la conversion de l'énergie mécanique en énergie acoustique, au travers des différents couplages impliqués dans la production du son émis par l'instrument, corde/chevalet, chevalet/table d'harmonie, table d'harmonie/cavité d'air. La courbe montre une bonne cohérence entre théorie et expérience, ce qui est plutôt de bonne augure pour le musicien !

Annexe C

Description du bec de clarinette



FIG. C.1 – Description du bec de clarinette (figure issue du magazine Vandoren N° 2 [Vandoren, 2000]).

Annexe D

Constantes du modèle physique de clarinette

Caractéristiques de l'anche

$$\left\{ \begin{array}{lcl} w & = & 10 \times 10^{-3} \quad m \\ \mu_r & = & 0.0231 \quad kg.m^{-2} \\ f_r & = & 2500 \quad Hz \\ Q_r & = & 5, \text{ soit :} \\ C_r & \approx & 72.57 \quad MKS \\ K_r & \approx & 5.7 \times 10^6 \quad Pa.m^{-1} \end{array} \right. \quad (D.1)$$

Paramètres géométriques du résonateur de l'instrument

$$R = 7 \times 10^{-3} \quad m \quad (D.2)$$

Constantes thermodynamiques

$$\left\{ \begin{array}{lcl} l_v & = & 4 \times 10^{-8} \quad MKS \\ l_t & = & 5.6 \times 10^{-8} \quad MKS \\ \frac{C_p}{C_v} & = & 1.4 \end{array} \right. \quad (D.3)$$

Autres constantes physiques

$$\left\{ \begin{array}{lcl} c & = & 340 \quad m.s^{-1} \\ \rho & = & 1.29 \quad MKS \end{array} \right. \quad (D.4)$$

Annexe E

Cartographie du jeu instrumental de la clarinette

En réponse à la question "Qu'est-ce qui est mis en jeu pour faire un son ?", le clarinettiste professionnel Claude Crousier a dressé une liste des facteurs impliqués dans la relation instrumentale :

1) Le corps

- a) Le souffle
 - technique de respiration - technique de souffle*
- b) L'embouchure
 - externe
 - lèvres, dents*
 - interne
 - phonation intérieure, c'est-à-dire la place du palais, l'arrière-gorge, le larynx, le bas de la langue, etc.*
- c) Les doigts
 - sur l'instrument
 - en train de boucher*
 - au dessus de l'instrument
 - comment sont-ils quand ils ne bouchent pas ?*
 - maintien de l'instrument dans la bouche
 - le bras, l'épaule, corps dans sa globalité*
- d) Le lien avec la terre
 - debout
 - assis
 - relation entre les pieds et la tête

2) L'intellect

- a) Connaissance de son corps
- b) Connaissance de son instrument
- c) Connaissance et analyse des styles musicaux
- d) La psychoacoustique

3) Le matériel

- a) L'instrument
- b) Le bec

- c) L'anche
- d) La ligature

4) Le lieu

- a) L'espace dans lequel on joue
- b) La place dans cet espace
- c) L'espace sans public et avec du public
- d) L'émission et la réception

5) Les conditions atmosphériques

- a) Température
- b) Hygrométrie
- c) Altitude

Avec, certes, un certain degré d'abstraction, le musicien perçoit donc son instrument comme faisant partie intégrante de lui-même. Se dessine alors une nouvelle entité musicien/instrument influencée par le monde extérieur (lieu, public, etc.), ce qui pour un clarinettiste va être déterminant aussi bien pour le choix de son matériel (l'anche, par exemple), que sur sa manière de contrôler son instrument (le souffle, par exemple).

Annexe F

Compléments des analyses statistiques

F.1 Concernant le chapitre 4

		Correlations											
		AT	TC	SC	SS	SFAT	OER	IRRJEN	TR1	TR2	TR3	GAMMA	ZETA
AT	Pearson Correlation	1	,664**	-,942**	-,825**	,952**	-,801**	-,780**	,963**	-,768**	-,952**	-,201	-,802**
	Sig. (2-tailed)	.	,007	,000	,000	,000	,000	,001	,000	,001	,000	,472	,000
	N	15	15	15	15	15	15	15	15	15	15	15	15
TC	Pearson Correlation	,664**	1	-,556*	-,416	,676**	-,142	-,075	,755**	-,278	-,581*	,555*	-,867**
	Sig. (2-tailed)	,007	.	,031	,123	,006	,614	,790	,001	,316	,023	,032	,000
	N	15	15	15	15	15	15	15	15	15	15	15	15
SC	Pearson Correlation	-,942**	-,556*	1	,955**	-,966**	,889**	,809**	-,947**	,686**	,998**	,349	,830**
	Sig. (2-tailed)	,000	,031	.	,000	,000	,000	,000	,000	,005	,000	,202	,000
	N	15	15	15	15	15	15	15	15	15	15	15	15
SS	Pearson Correlation	-,825**	-,416	,955**	1	-,897**	,894**	,768**	-,835**	,519*	,940**	,463	,766**
	Sig. (2-tailed)	,000	,123	,000	.	,000	,000	,001	,000	,047	,000	,082	,001
	N	15	15	15	15	15	15	15	15	15	15	15	15
SFAT	Pearson Correlation	,952**	,676**	-,966**	-,897**	1	-,780**	-,687**	,944**	-,595*	-,974**	-,165	-,901**
	Sig. (2-tailed)	,000	,006	,000	,000	.	,001	,005	,000	,019	,000	,557	,000
	N	15	15	15	15	15	15	15	15	15	15	15	15
OER	Pearson Correlation	-,801**	-,142	,889**	,894**	-,780**	1	,963**	-,738**	,743**	,874**	,712**	,493
	Sig. (2-tailed)	,000	,614	,000	,000	,001	.	,000	,002	,002	,000	,003	,062
	N	15	15	15	15	15	15	15	15	15	15	15	15
IRRJEN	Pearson Correlation	-,780**	-,075	,809**	,768**	-,687**	,963**	1	-,694**	,854**	,799**	,741**	,356
	Sig. (2-tailed)	,001	,790	,000	,001	,005	,000	.	,004	,000	,000	,002	,193
	N	15	15	15	15	15	15	15	15	15	15	15	15
TR1	Pearson Correlation	,963**	,755**	-,947**	-,835**	,944**	-,738**	-,694**	1	-,726**	-,956**	-,102	-,886**
	Sig. (2-tailed)	,000	,001	,000	,000	,000	,002	,004	.	,002	,000	,717	,000
	N	15	15	15	15	15	15	15	15	15	15	15	15
TR2	Pearson Correlation	-,768**	-,278	,686**	,519*	-,595*	,743**	,854**	-,726**	1	,691**	,417	,357
	Sig. (2-tailed)	,001	,316	,005	,047	,019	,002	,000	,002	.	,004	,122	,191
	N	15	15	15	15	15	15	15	15	15	15	15	15
TR3	Pearson Correlation	-,952**	-,581*	,998**	,940**	-,974**	,874**	,799**	-,956**	,691**	1	,314	,841**
	Sig. (2-tailed)	,000	,023	,000	,000	,000	,000	,000	,000	,004	.	,254	,000
	N	15	15	15	15	15	15	15	15	15	15	15	15
GAMMA	Pearson Correlation	-,201	,555*	,349	,463	-,165	,712**	,741**	-,102	,417	,314	1	-,183
	Sig. (2-tailed)	,472	,032	,202	,082	,557	,003	,002	,717	,122	,254	.	,514
	N	15	15	15	15	15	15	15	15	15	15	15	15
ZETA	Pearson Correlation	-,802**	-,867**	,830**	,766**	-,901**	,493	,356	-,886**	,357	,841**	-,183	1
	Sig. (2-tailed)	,000	,000	,000	,001	,000	,062	,193	,000	,191	,000	,514	.
	N	15	15	15	15	15	15	15	15	15	15	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

FIG. F.1 – Matrice de corrélations (r de Pearson, d.d.l. = 13) entre les principaux descripteurs de timbre et les paramètres de contrôle, caractérisant la répartition des sons de clarinette dans l'espace de timbre perceptif. Les descripteurs spectraux ont été calculés avec des amplitudes spectrales exprimées en puissance.

F.2 Concernant le chapitre 5

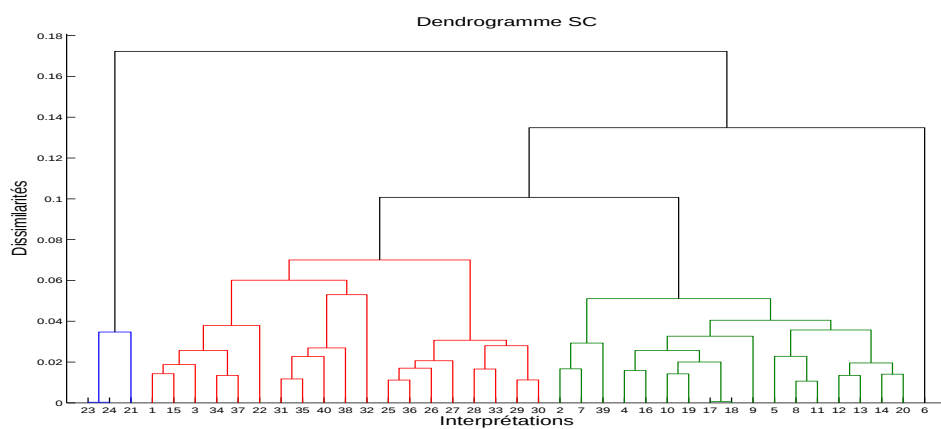
F.2.1 Description de la méthode séquentielle de Holm :

La procédure séquentielle de Holm [Holm, 1979], employée dans le cas de K comparaisons multiples, présente les étapes suivantes [Ludbrook, 1998] :

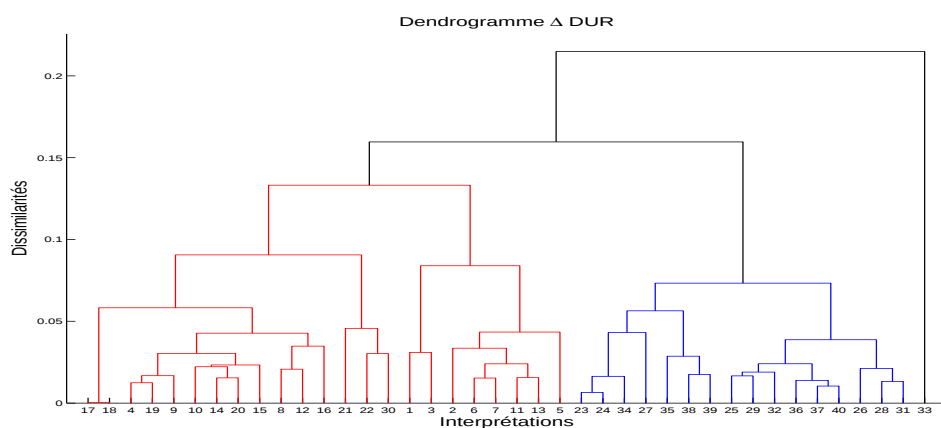
1. Effectuer toutes les comparaisons 2 à 2 par le test de Student.
2. Classer les comparaisons par ordre croissant des valeurs p obtenues.
3. Effectuer de nouveau les tests de Student en appliquant au risque $\alpha_{Student}$ la correction de
 - Bonferroni $\alpha_{Bonferroni} = m\alpha_{Student}$, ou celle de
 - Sidak (moins conservative) : $\alpha_{Sidak} = 1 - (1 - \alpha_{Student})^m$, en prenant m tel que :
 - $m = K$ (nombre total de comparaisons) pour la plus petite valeur p ,
 - $m = K - 1$ pour la valeur p immédiatement supérieur,
 - et ainsi de suite jusqu'à $m = 1$, correspondant à la plus grande valeur p initiale.

Si α_{Holm} pour la comparaison C est inférieur à α_{Holm} pour la comparaison précédente $C - 1$, alors α_{Holm} pour la comparaison C sera donné égal à α_{Holm} pour la comparaison $C-1$ (principe de monotonie).

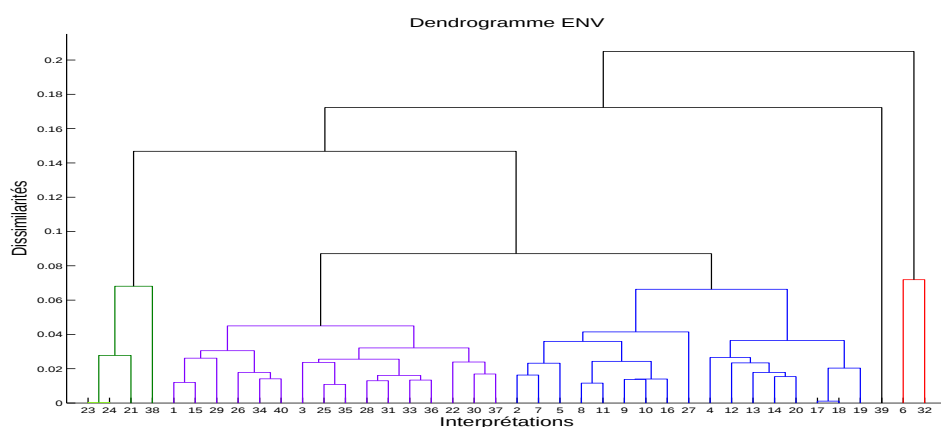
F.2.2 Interprétations de l'extrait de Bach



(a)



(b)



(c)

FIG. F.2 – Représentation en dendrogramme de la classification ascendante hiérarchique des dissimilarités de Centre de Gravité Spectral (a), variations rythmiques (b), et enveloppe RMS (c) pour le 40 interprétations de l'extrait de Bach. L'algorithme de classification utilisé est le lien complet.

F.2.3 Interprétations de l'extrait de Mozart

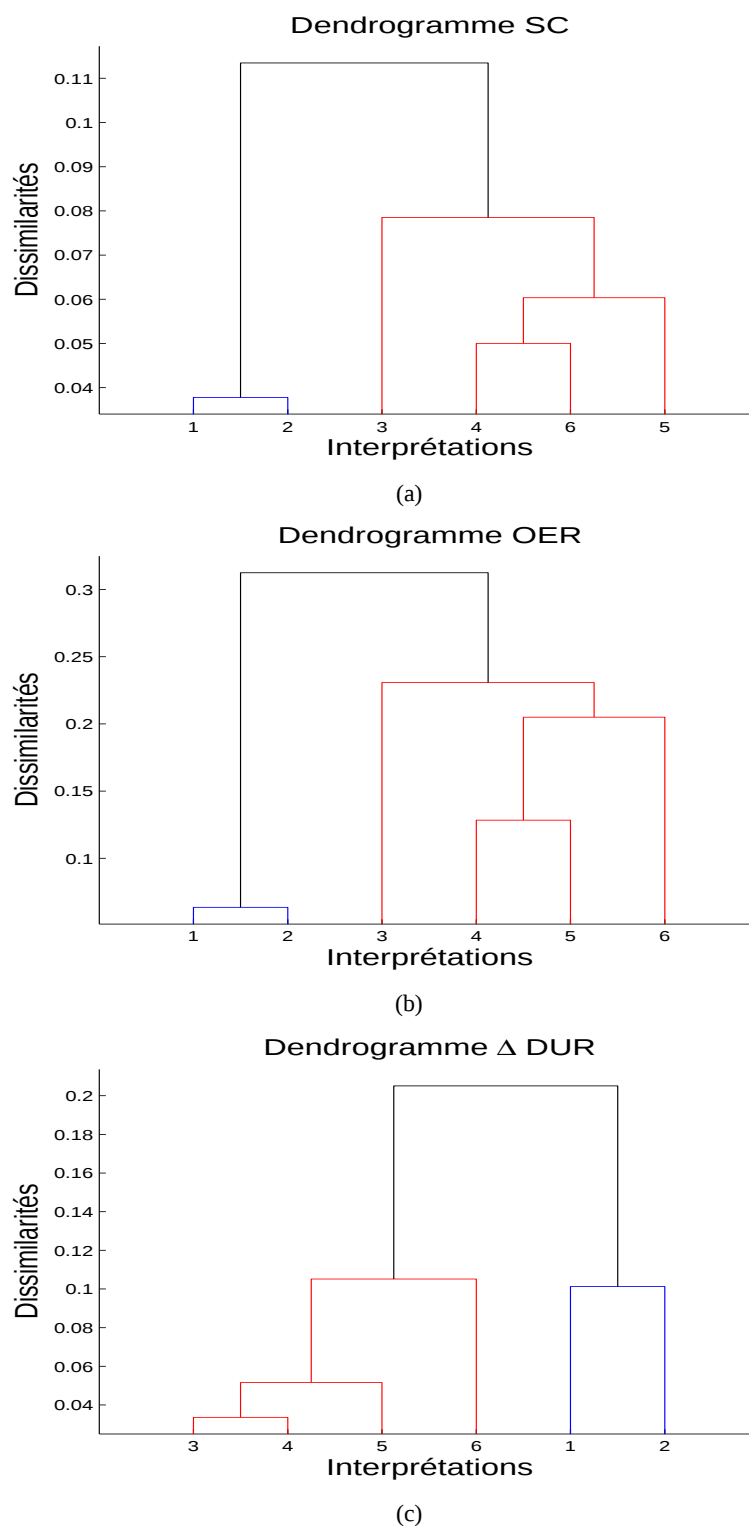


FIG. F.3 – Représentation en dendrogramme de la classification ascendante hiérarchique des dissimilarités de Centre de Gravité Spectral (a), variations rythmiques (b), et enveloppe RMS (c) pour le 6 interprétations de l'extrait de Mozart. L'algorithme de classification utilisé est le lien complet.

F.3 Concernant le chapitre 6 - Expérience 1

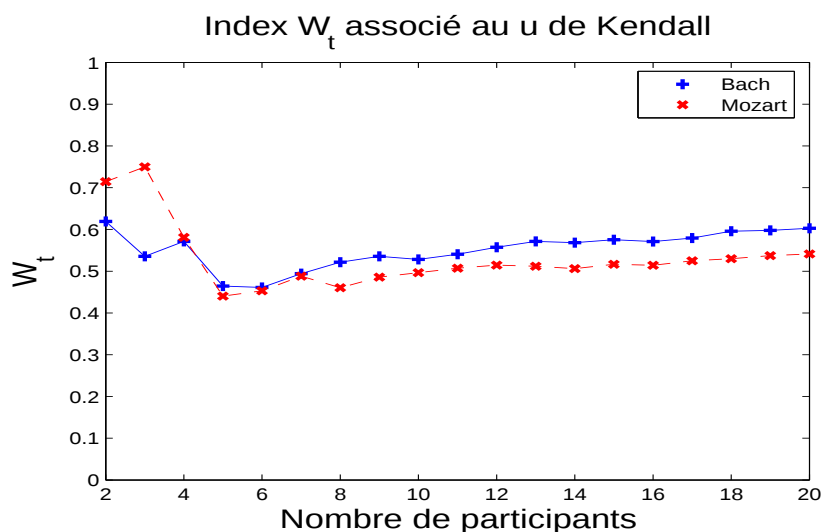


FIG. F.4 – Evolution de l'index d'accord W_t associé au u de Kendall (cf. [Siegel et John Castellan, 1988]) en fonction du nombre de participants pour les deux sessions 'Bach' (+) et 'Mozart' (x). La figure montre que l'index W_t (qui peut varier entre 0 et 1) tend à se stabiliser au fur et à mesure que le nombre de participants aux expériences perceptives augmente.

F.4 Concernant le chapitre 6 - Expérience 2

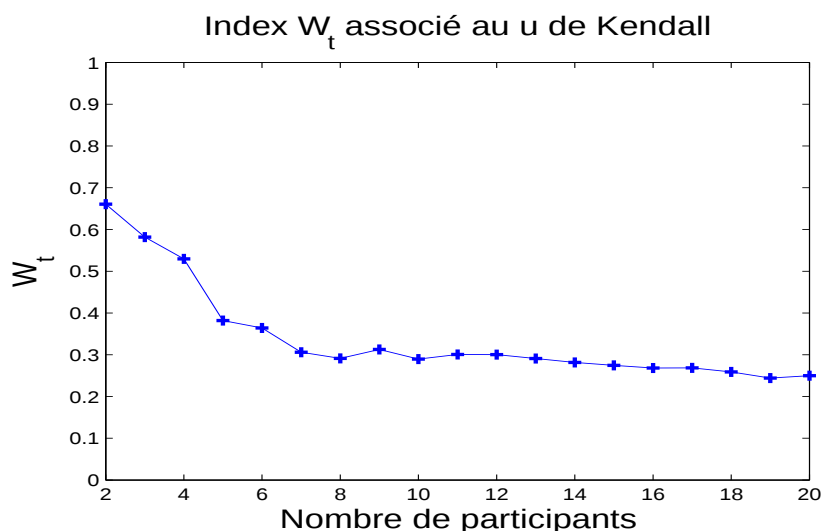


FIG. F.5 – Evolution de l'index d'accord W_t associé au u de Kendall (cf. [Siegel et John Castellan, 1988]) en fonction du nombre de participants. La figure montre que l'index W_t (qui peut varier entre 0 et 1) tend à se stabiliser au fur et à mesure que le nombre de participants aux expériences perceptives augmente.

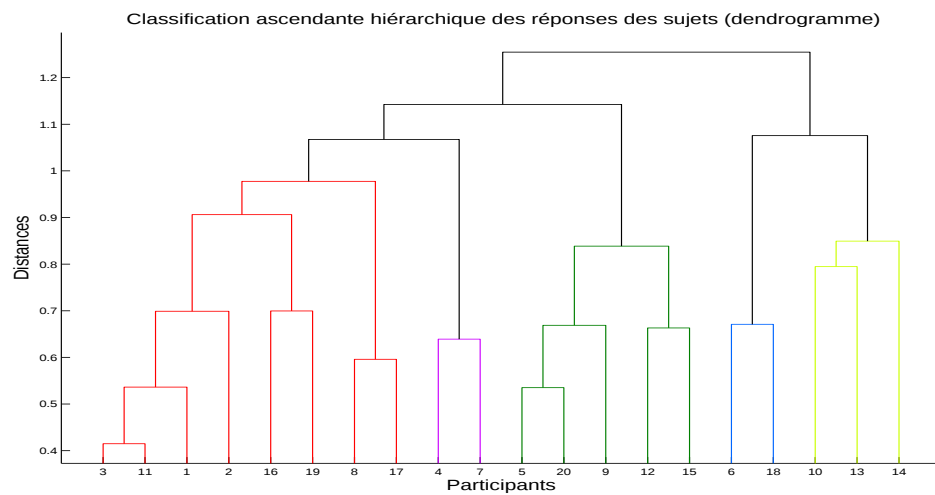


FIG. F.6 – Représentation en dendrogramme de la classification ascendante hiérarchique des distances entre les réponses des participants. L'algorithme de classification utilisé repose sur le lien complet (cf. par ex. [Saporta, 1990]).

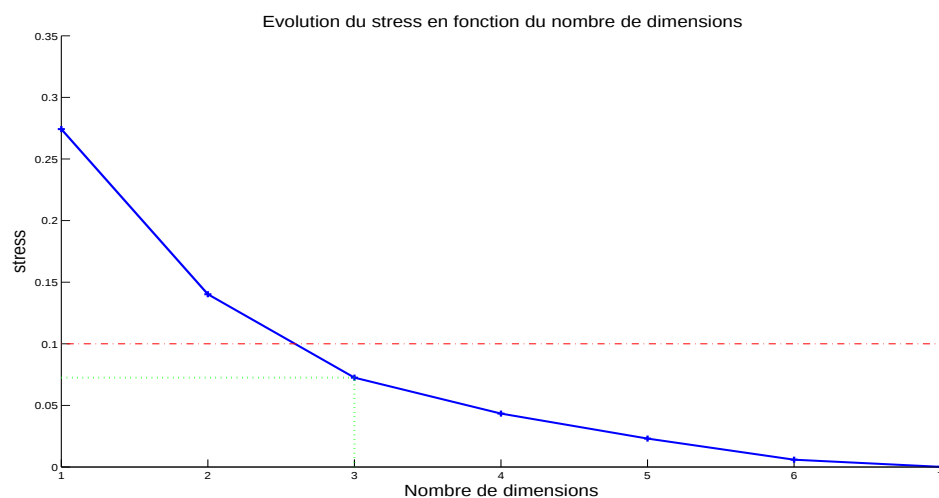


FIG. F.7 – Evolution du stress (de Kruskal) en fonction du nombre de dimensions de la configuration MDS (expérience 2). La courbe présente un coude au niveau d'un nombre de dimensions égal à 3 ($stress \approx 0.07$).

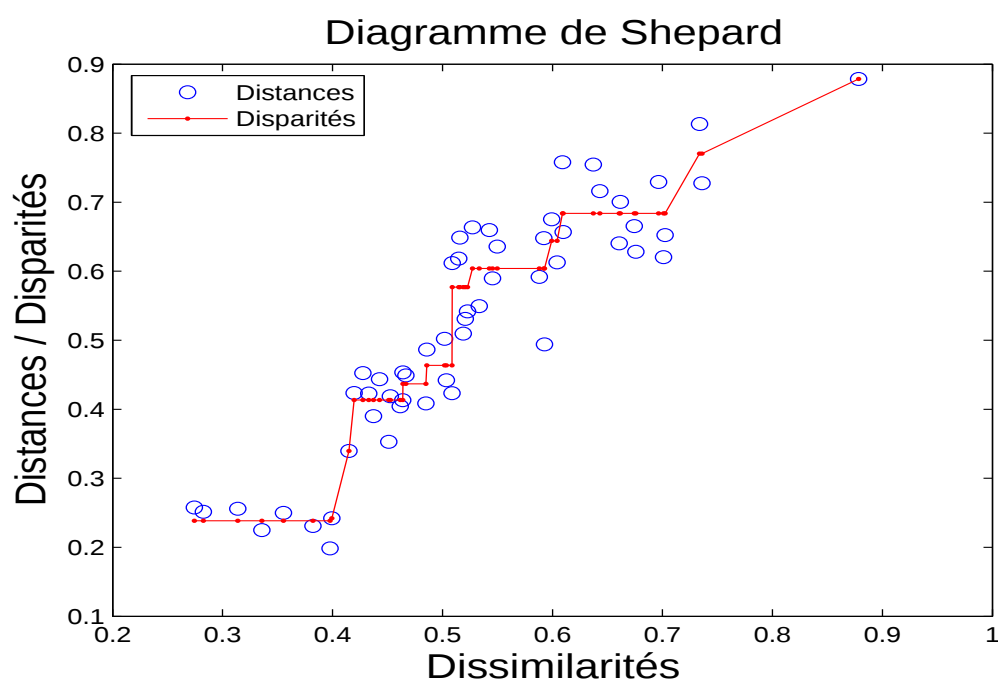


FIG. F.8 – Diagramme de Shepard associé à l'analyse multidimensionnelle de proximité non-métrique. Ce diagramme présente la relation entre les distances inter-stimuli issues de la configuration MDS et les dissimilarités dérivées des réponses des participants. La figure indique également les disparités obtenues par transformation monotone croissante des dissimilarités dans les procédures d'analyse multidimensionnelle de type non-métrique.

Annexe G

Liste de publications

Cette annexe présente les différentes publications de l'auteur relatives aux travaux exposés dans cette thèse.

Revues à comité de lecture :

Barthet, M., Guillemain, P., Kronland-Martinet, R., et Ystad, S., "From clarinet control to timbre perception", *Acta Acustica* (en révision)

Barthet, M., Depalle, P., Kronland-Martinet, R., et Ystad, S., "Analysis of the acoustical correlates of timbre in clarinet performance", *Music Perception* (en révision)

Barthet, M., Depalle, P., Kronland-Martinet, R., et Ystad, S., "Timbre as a vector of musical expression in clarinet performance", *Music Perception* (en révision)

Barthet, M., Kronland-Martinet, R., et Ystad, S., "Improving Musical Expressiveness by Time-Varying Brightness Shaping", *Computer Music Modeling and Retrieval : Sense of Sounds*, Springer-Verlag, 2008, pp. 313-336

Actes de conférence à comité de lecture :

Barthet, M., Guillemain, P., Kronland-Martinet, R., et Ystad, S., "Exploration of timbre variations in music performance", *J. Acoust. Soc. Am.* 123, 3663, 2008

Barthet, M., Depalle, P., Kronland-Martinet, R., et Ystad, S., "The effect of timbre in clarinet interpretation", *Proc. International Computer Music Conference (ICMC07)*, pp. 59-62, Copenhagen, Denmark, 2007

Barthet, M., Kronland-Martinet, R., et Ystad, S., "Consistency of timbre patterns in expressive music performance", *Proc. 9th International Conference on Digital Audio Effects (DAFx06)*, pp. 19-24, Montréal, Québec, Canada, 2006

Barthet, M., Guillemain, P., Kronland-Martinet, R., et Ystad, S., "On the relative influence of even and odd harmonics in clarinet timbre", *Proc. International Computer Music Conference (ICMC05)*, pp. 351-354, Barcelona, Spain, 2005

Posters :

Barthet, M., Guillemain, P., Kronland-Martinet, R., et Ystad, S., "Exploration of timbre variations in music performance", Acoustics'08, Paris, France, 2008

Barthet, M., Kronland-Martinet, R., et Ystad, S., "Evaluation perceptive d'une interprétation musicale en fonction de trois paramètres d'expression : le Rythme, l'Intensité et le Timbre", Journées fondatrices du Groupe Perception Sonore de la Société Française d'Acoustique, Lyon, France, 2007

Barthet, M., et Silva, F., "Les tribulations d'une corde de guitare", Espace Sciences de la Matière, Université de Provence, Fête de la Science, 2006

Barthet, M., Kronland-Martinet, R., et Ystad, S., "De la physique au timbre : applications au contrôle instrumental et à l'interprétation musicale", Journées des Jeunes Chercheurs en Audition, Acoustique musicale et Signal Audio (JJCAAS-05), Marseille, France, 2005

Bibliographie

- ABELES, H. F. (1973). Development and validation of a clarinet performance adjudication scale. *Journal of Research in Music Education*, 21(3):246–255.
- AFNOR (1977). *Recueil des normes françaises de l'acoustique, Tome 1 (vocabulaire), NF S 30-107*. Association Française de NORmalisation.
- ANDERSEN, T. H. et JENSEN, K. (2004). Importance and representation of phase in the sinusoidal model. *J. Audio Eng. Soc.*, 52(11):1157–1169.
- ANSI (1960). *USA Standard Acoustical Terminology*. American National Standards Institute, New York.
- ARAMAKI, M. et KRONLAND-MARTINET, R. (2006). Analysis-synthesis of impact sounds by real-time dynamic filtering. *IEEE Trans. on Acoust., Speech, and Sig. Proc.*, 14(2):695–705.
- BACKUS, J. (1963). Small-vibration theory of the clarinet. *J. Acoust. Soc. Am.*, 35(3):305–313.
- BARTHET, M. (2004). Optimisation des paramètres géométriques d'un modèle simplifié de résonateur de trompette. Mémoire de D.E.A., Université d'Aix-Marseille II, CNRS-LMA (S2M), Marseille.
- BARTHET, M., DEPALLE, P., KRONLAND-MARTINET, R. et YSTAD, S. (2009a). Analysis of the acoustical correlates of timbre in clarinet performance. *Music Perception (under revision)*.
- BARTHET, M., DEPALLE, P., KRONLAND-MARTINET, R. et YSTAD, S. (2009b). Timbre as a vector of musical expression in clarinet performance. *Music Perception (under revision)*.
- BARTHET, M., GUILLEMAIN, P., KRONLAND-MARTINET, R. et YSTAD, S. (2005). On the relative influence of even and odd harmonics in clarinet timbre. In *Proc. Int. Comp. Music Conf. (ICMC'05)*, pages 351–354, Barcelona, Spain.
- BARTHET, M., GUILLEMAIN, P., KRONLAND-MARTINET, R. et YSTAD, S. (2009c). From clarinet control to timbre perception. *Acta Acustica (under revision)*.
- BARTHET, M., KRONLAND-MARTINET, R. et YSTAD, S. (2006). Consistency of timbre patterns in expressive music performance. In *Proc. 9th Int. Conf. on Digital Audio Effects (DAFx06)*, pages 19–24, Montreal, Quebec, Canada.
- BARTHET, M., KRONLAND-MARTINET, R. et YSTAD, S. (2008). Improving musical expressiveness by time-varying brightness shaping. In KRONLAND-MARTINET, R., YSTAD, S. et JENSEN, K., éditeurs : *Sense of Sounds*, numéro 4969 de Lecture Notes in Computer Science (LNCS), chapitre Towards the Generation of Sense of Sounds, pages 313–336. Springer-Verlag, Berlin / Heidelberg.
- BEAUCHAMP, J. W. (1982). Synthesis by spectral amplitude and brightness matching of analyzed musical instrument tones. *Journal of the Audio Engineering Society*, 30(6):396–406.

- BELLO, J. P., DAUDET, L., ABDALLAH, S., DUXBURY, C., DAVIES, M. et SANDLER, M. B. (2005). A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*.
- BENADE, A. H. (1976). *The fundamental of musical acoustics*. Oxford University Press, New York.
- BENADE, A. H. et KOUZOUPI, S. N. (1988). The clarinet spectrum : Theory and experiment. *Journal of the Acoustical Society of America*, 83(1):292–304.
- BOGERT, B. P., HEALY, M. J. R. et TUKEY, J. W. (1963). *The Quefrency Analysis of Time Series for Echoes : Cepstrum, pseudo-Autocovariance, Cross-Cepstrum, and Saphe-Cracking*, chapitre 15, pages 209–243. Wiley, New York.
- BOUASSE, H. (1929). *Instruments à vent (Tome 1 et 2)*. Librairie Delagrave (1929 and 1930) and Albert Blanchard (1986), Paris.
- BOULEZ, P. (1963). *Penser la musique aujourd'hui*. TEL. Gallimard, Paris.
- BREGMAN, A. (1994). *Auditory scene analysis - The Perceptual Organization of Sound*. MIT Press, Cambridge.
- BRESIN, R. (2000). *Virtual Virtuosity : Studies in Automatic Music Performance*. Thèse de doctorat, Kungl Tekniska Högskolan (KTH), Stockholm, Sweden.
- BRYMER, J. (1979). *Clarinete*. Collection Yehudi Menuhin. Hatier, Paris.
- CABRERA, D., FERGUSON, S., RIZWI, F. et SCHUBERT, E. (2008). Psysound3 : a program for the analysis of sound recordings. In *Acoustics 08 Paris*, pages 1609–1613.
- CACLIN, A. (2004). *Interactions et indépendances entre dimensions du timbre des sons complexes, Approche psychophysique et électrophysiologique chez l'Humain*. Thèse de doctorat, Université Paris 6.
- CACLIN, A., MCADAMS, S., SMITH, B. K. et WINSBERG, S. (2005). Acoustic correlates of timbre space dimensions : A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, 118(1):471–482.
- CADOZ, C. (1991). Timbre et causalité. In CHRISTIAN BOURGOIS, éditeur : *Le timbre, métaphore pour la composition*, pages 17–46. I.R.C.A.M., Paris.
- CANAZZA, S., DE POLI, G., RINALDIN, S. et VIDOLIN, A. (1991). Sonological analysis of clarinet expressivity. In LEMAN, M., éditeur : *Music, Gestalt, and Computing : Studies in Cognitive and Systematic Musicology*, numéro 1317 de Lecture Notes in Artificial Intelligence (LNAI), pages 431–439. Springer-Verlag, Berlin / Heidelberg.
- CAROLL, J. D. et CHANG, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N -way generalization of 'eckart-young' decomposition. *Psychometrika*, 35:283–319.
- CASTELLENGO, M. et DUBOIS, D. (2005). Timbre ou timbres ? Propriété du signal, de l'instrument, ou construction cognitive ? (Timbre or timbres ? Property of the signal, the instrument, or cognitive construction ?). In *Proc. of the Conf. on Interdisciplinary Musicology (CIM05)*, Montréal, Québec, Canada.
- CHAIGNE, A. (2001). *Ondes acoustiques*. Éditions de l'École polytechnique, Paris.

- CHOWNING, J. (1973). The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society (Reprinted in Computer Music Journal, Vol. 1 N°2, pp. 46-54, 1977)*, 21:526–534.
- CHOWNING, J. (2007). Composer le son lui-même. In de MAULE, M., éditeur : *John Chowning*, numéro 7 de Portraits Polychromes, pages 23–28. Institut National de l’Audiovisuel, Paris.
- CLARKE, E. F. (1988). Generative principles in music performance. In SLOBODA, J. A., éditeur : *Generative processes in music. The psychology of performance, improvisation and composition*, Oxford Science Publications, pages 1–26. Oxford University Press, Oxford.
- CNRTL (2008). Centre national de ressources textuelles et lexicales. <http://www.cnrtl.fr/lexicographie/>.
- CONSONNES (2008). Site web du projet ANR Consonnes. <http://www.consonnes.cnrs-mrs.fr/>.
- DALMONT, J. P., GAZENGEL, B., GILBERT, J. et KERGOMARD, J. (1995). Some aspects of tuning and clean intonation in reed instruments. *Applied Acoustics*, 46:19–60.
- DALMONT, J.-P., GILBERT, J., KERGOMARD, J. et OLLIVIER, S. (2005). An analytical prediction of the oscillation and extinction thresholds of a clarinet. *J. Acoust. Soc. Am.*, 118(5):3294–3305.
- DAVIS, S. et MERMELSTEIN, P. (1980). Experiments in syllable-based recognition of continuous speech. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28:357–366.
- DE POLI, G. (2006). *Algorithms for Sound and Music Computing*, chapitre Expressiveness in music performance. Creative Commons.
- DEBUT, V. (2004). *Deux études d’un instrument de musique de type clarinette : Analyse des fréquences propres du résonateur et calcul des auto-oscillations par décomposition modale*. Thèse de doctorat, Université Aix-Marseille II.
- DILLON, W. R. et GOLDSTEIN, M. (1984). *Multivariate Analysis*. Wiley series in probability and mathematical statistics. John Wiley & Sons, New York.
- DIXON, S., GOEBL, W. et WIDMER, G. (2002). Real time tracking and visualisation of musical expression. In *Music and Artificial Intelligence : Second International Conference, ICMAI2002*, pages 58–68, Edinburgh, Scotland. Springer.
- DRAKE, C. et PALMER, C. (1993). Accent structures in music performance. *Music Perception*, 10:343–378.
- DUBNOV, S. et RODET, X. (1997). Statistical modeling of sound aperiodicities. In *Proc. of the International Computer Music Conference*, pages 43–50, Thessaloniki, Greece.
- EHRESMAN, D. et WESSEL, D. L. (1978). Perception of timbral analogies. Rapport technique 13, I.R.C.A.M., Paris.
- FARCY, A. (2008). Conception et mise en oeuvre d’un dispositif d’étude du rôle du conduit vocal dans le jeu du saxophone. Mémoire de D.E.A., Université d’Aix-Marseille II, CNRS-LMA (S2M), Marseille.
- FAURE, A. (2000). *Des sons aux mots, comment parle-t-on du timbre musical ? (From sounds to words, how do we speak of musical timbre ?)*. Thèse de doctorat, École des Hautes Etudes en Sciences Sociales, Paris.

- FLETCHER, H. (1934). Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure. *Journal of the Acoustical Society of America*, 6(2):59–69.
- FLETCHER, H. (1940). Auditory patterns. *Rev. Mod. Phys.*, 12:47–65.
- FLETCHER, N. H. et ROSSING, T. D. (1998). *The Physics of Musical Instruments*. Springer, Heidelberg, 2nd édition.
- FÖDERMAYR, F. et DEUTSCH, W. A. (1993). Parmi veder le lagrime. One aria, three interpretations. In *Proc. of the Stockholm Music Acoustics Conf.*, pages 96–107.
- FRIBERG, A. (1995). *A Quantative Rule System for Musical Performance*. Thèse de doctorat, Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm.
- FRIBERG, A. (2004). A fuzzy analyzer of emotional expression in music performance and body motion. In SUNDBERG, J. et BRUNSON, B., éditeurs : *Proceedings of Music and Music Science*, pages 1–13, Stockholm.
- FRIBERG, A., FRYDEN, L., BODIN, L. et SUNDBERG, J. (1991). Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal*, 15:49–55.
- FRITZ, C. (2004). *La clarinette et le clarinettiste : Influence du conduit vocal sur la production du son (The clarinet and the clarinetist : Influence of the vocal tract on the sound production)*. Thèse de doctorat, Université Paris 6 et University of New South Wales.
- FRITZ, C. et WOLFE, J. (2004). Acoustic impedance measurement of the clarinet players' airway. In *CFA/DAGA'04*, pages 101–102, Strasbourg.
- GABRIELSSON, A. (1988). Timing in music performance and its relations to music experience. In SLOBODA, J. A., éditeur : *Generative processes in music*, pages 27–51. Oxford University Press, Oxford. (Reprinted 2005).
- GABRIELSSON, A. (1999). The performance of music. In *Psychology of Music*. Academic Press, 2nd édition.
- GABRIELSSON, A. et LINDSTROM, B. (1985). Perceived sound quality of high-fidelity loudspeakers. *J. Audio Eng. Soc.*, 33(1):33–53.
- GAYOU, É. (2007). Entretien avec Max V. Mathews. In *Max Mathews*, numéro 11 de Portraits Polychromes, pages 7–26. Institut National de l'Audiovisuel, Paris.
- GENESIS (2008). Site web. <http://www.genesis-acoustics.com/>.
- GIBSON, J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin Company, Boston.
- GILBERT, J. et KERGMARD, J. (1989). Calculation of the steady-state oscillations of a clarinet using the harmonic balance technique. *J. Acoust. Soc. Am.*, 86(1):34–41.
- GOEBL, W., PAMPALK, E. et WIDMER, G. (2004). Exploring expressive performance trajectories : six famous pianists play six Chopin pieces. In LIPSCOMB, S. D., ASHLEY, R. et GJERDINGEN, R. O. & Webster, P., éditeurs : *Journal of Research in Music Education*, pages 505–509, Evanston, IL, USA. Adelaide, Australia : Causal Productions.
- GOLDFARB, D. (1970). A family of variable metric updates derived by variational means. *Mathematics of Computing*, 24:23–26.

- GOMEZ, E. (2002). Melodic description of audio signals for music content processing. Mémoire de D.E.A., Pompeu Fabra University, Barcelona.
- GORDON, J. W. (1987). The perceptual attack time of musical tones. *J. Acoust. Soc. Am.*, 82(1): 88–105.
- GRAND, N., GILBERT, J. et LALOË, F. (1996). Oscillation threshold of woodwind instruments. *Acustica*, 82:137–151.
- GREY, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, 61:1270–1277.
- GREY, J. W. et GORDON, J. W. (1978). Perception of spectral modifications on orchestral instrument tones. *Computer Music Journal*, 11(1):24–31.
- GUILLEMAIN, P. (2007). Some roles of the vocal tract in clarinet breath attacks : Natural sound analysis and model-based synthesis. *J. Acoust. Soc. Am.*, 121(4):2396–2406.
- GUILLEMAIN, P., KERGOMARD, J. et VOINIER, T. (2005). Real-time synthesis of clarinet-like instruments using digital impedance models. *J. Acoust. Soc. Am.*, 118(1):483–494.
- HAJDA, J. M., KENDALL, R. A., CARTERETTE, E. C. et HARSHBERGER, M. L. (1997). Methodological issues in timbre research. In DELIÉGE, I. et SLOBODA, J., éditeurs : *Perception and Cognition of Music*, pages 253–306. Psychology Press, New York, 2nd édition.
- HELMHOLTZ, H. v. (1954). *On the Sensations of Tone*. Dover. (from the works of 1877). Trad. anglaise avec notes et appendice de E.J. Ellis.
- HOFFMAN, M. et COOK, P. (2007). The featsynth framework for feature-based synthesis : design and applications. In *Proc. Int. Comp. Music Conf. (ICMC'07)*, Copenhagen, Denmark, volume 2, pages 184–187.
- HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6:65–70.
- JAILLET, F. (2005). *Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore*. Thèse de doctorat, Université de la Méditerranée - Aix-Marseille II.
- JENSEN, K. (1999a). Envelope model of isolated musical sounds. In *Proc. of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, NTNU, Trondheim.
- JENSEN, K. (1999b). *Timbre Models of Musical Sounds*. Thèse de doctorat, Department of Computer Science, University of Copenhagen.
- JENSEN, K. (2002). The timbre model - discrimination and expression. In *Proceedings of the Mosart Midterm Meeting, Esbjerg, Denmark*.
- KENDALL, R. et CARTERETTE, E. C. (1993). Verbal attributes of simultaneous wind instrument timbres : I. von bismarck's adjectives. II. adjective induced from piston's orchestraion. *Music Perception*, 10(4):445–468 ; 469–502.
- KENDALL, R. A. et CARTERETTE, E. C. (1996). Difference thresholds for timbre related to spectral centroid. In PENNYCOOK, B. et COSTA-GIOMI, E., éditeurs : *Proc. of the 4th International Conference on Music Perception and Cognition (ICMPC)*, pages 91–95, Montreal, Canada : Faculty of Music, McGill University.

- KENDALL, R. A., CARTERETTE, E. C. et HAJDA, J. M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, 16(3):327–364.
- KERGOMARD, J. (1991). Le timbre des instruments à anche. In CHRISTIAN BOURGOIS, éditeur : *Le timbre, métaphore pour la composition*, Musique / Passé / Présent, pages 224–235. I.R.C.A.M., Paris.
- KERGOMARD, J. (1995). Elementary considerations on reed-instrument oscillations. In *Mechanics of Musical Instruments*. Springer-Verlag, New York.
- KIM, H.-G., MOREAU, N. et SIKORA, T. (2005). *MPEG-7 Audio and beyond - Audio content indexing and retrieval*. John Wiley & Sons.
- KOTTICK, E., MARSHALL, K. et HENDRICKSON, T. (1995). L'acoustique du clavecin. In *Les instruments de l'orchestre*, pages 41–48. Pour la science Diffusion Belin, Paris.
- KRIMPHOFF, J., MCADAMS, S. et WINSBERG, S. (1994). Caractérisation du timbre des sons complexes, II Analyses acoustiques et quantification psychophysique (Characterization of complex sounds timbre, II Acoustical analyses and psychophysical quantification). *Journal de Physique IV, Colloque C5*, 4:625–628.
- KRUMHANS, C. L. (1989). Why is musical timbre so hard to understand? In NIELZÉN, S. et OLSSON, O., éditeurs : *Proc. of the Marcus Wallenberg Symposium held in Lund, Sweden*, pages 43–53, Amsterdam. Excerpta Medica.
- KRUSKAL, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29:1–27.
- KRUSKAL, J. B. (1964b). Nonmetric multidimensional scaling : a numerical method. *Psychometrika*, 29:115–129.
- KRZANOWSKI, W. J. (2000). *Principles of Multivariate Analysis : A User's Perspective*. Statistical Science. Oxford University Press, Oxford, Rev Sub édition.
- LALOË, F. et LALOË, S. (1995). La clarinette. In *Les instruments de l'orchestre*, Pour la science, pages 74–87. Belin, Paris.
- LANDY, P. (1996). *Musique du Japon*. Les Traditions Musicales de l'Institut International d'Études Comparatives de la Musique. Buchet/Chastel, Paris.
- LAROUSSE (1957). *Larousse Classique*. Librairie Larousse, Paris.
- LARTILLOT, O. et TOIVIAINEN, P. (2007). A matlab toolbox for musical feature extraction from audio. In *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France.
- LAVANDIER, M. (2005). *Différences entre enceintes acoustiques : une évaluation physique et perceptive*. Thèse de doctorat, Université d'Aix-Marseille II.
- LERDAHL, F. (1991). Les hiérarchies de timbre. In CHRISTIAN BOURGOIS, éditeur : *Le timbre, métaphore pour la composition*, Musique / Passé / Présent, pages 182–203. I.R.C.A.M., Paris.
- LEVINE, H. et SCHWINGER, J. (1948). On the radiation of sound from an unflanged circular pipe. *Physical Review*, 73(4):383–406.
- LOUREIRO, M. A., de PAULA, H. B. et YEHA, H. C. (2004). Timbre classification of a single instrument. In *ISMIR 2004 5th International Conference on Music Information Retrieval*, Barcelona, Spain. Audiovisual Institute, Universitat Pompeu Fabra.

- LUDBROOK, J. (1998). Multiple comparison procedures updated. *Clin. Exp. Pharmacol. Physiol.*, 25(12):1032–1037.
- MAROZEAU, J. (2004). *L'effet de la fréquence fondamentale sur le timbre (The effect of fundamental frequency on timbre)*. Thèse de doctorat, Université Pierre et Marie Curie, Paris VI.
- MAROZEAU, J. et de CHEVEIGNÉ, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *Journal of the Acoustical Society of America*, 121(1):383–387.
- MATHEWS, M., FRIBERG, A., BENNETT, G., SAPP, C. et SUNDBERG, J. (2003). A marriage of the Director Musices program and the conductor program. In *Proc. of the Stockholm Music Acoustics Conference (SMAC 03)*, Stockholm, Sweden.
- MAX, J. et LACOUME, J.-L. (2000). *Méthodes et techniques de traitement du signal*. Dunod, 5th édition.
- MCADAMS, S. (1994). La reconnaissance de sources et d'événements sonores (The recognition of sources and sound events). In MCADAMS, S. et BIGAND, E., éditeurs : *Penser les sons, Psychologie et sciences de la pensée*, pages 157–213. Presses Universitaires de France, Paris, 1st édition.
- MCADAMS, S. (1999). Perspectives on the contribution of timbre to musical structure. *Computer Music Journal*, 23(3):85–102.
- MCADAMS, S., BEAUCHAMP, J. W. et MENEGUZZI, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *Journal of the Acoustical Society of America*, 105(2):882–897.
- MCADAMS, S. et CUNIBILE, J. C. (1992). Perception of timbre analogies. *Philosophical Transactions of the Royal Society of London, Series B* 336:383–390.
- 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.
- METFESSEL, M. (1926). Sonance as a form of tonal fusion. *Pyschol. Rev.*, XXXIII:459–466.
- 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.
- MOORE, B. C. J., GLASBERG, B. R., PLACK, C. et BISWAS, A. K. (1988). The shape of the ear's temporal window. *J. Acoust. Soc. Am.*, 83(3):1102–1116.
- MOORE, B. C. J. et GLASBERG, B. R. A. (1996). A revision of Zwicker's loudness model. *Acustica*, 82:335–345.
- MOORER, J. A. (1977). Signal processing aspects of computer music – a survey. *Computer Music Journal*, 1(1):4–38.
- NATTIEZ, J.-J. (1975). *Fondements d'une sémiologie de la musique*. Union Générale d'Éditions, Paris.
- NEDERVEEN, C. J. (1969). *Acoustical aspects of woodwind instruments*. Northern Illinois University Press, Illinois.
- NORRIS, A. N. et SHENG, I. C. (1989). Acoustic radiation from a circular pipe with an infinite flange. *Journal of Sound and Vibration*, 135(1):85–93.

- OLLIVIER, S. (2002). *Contribution à l'étude des oscillations des instruments à vent à anche simple : validation d'un modèle élémentaire*. Thèse de doctorat, Université du Maine.
- OLSON, H. (1967). *Music, physics and engineering*. Dover, New York.
- PALLONE, G. (2003). *Dilatation et Transposition sous contraintes perceptives des signaux audio : application au transfert cinéma-vidéo*. Thèse de doctorat, Université Aix-Marseille II.
- PALMER, C. (1996). Anatomy of a performance : Sources of musical expression. *Music Perception*, 13:433–454.
- PALMER, C. (1997). Music performance. *Annual Review of Psychology*, 48:115–138.
- PALMER, C. et HUTCHINS, S. (2006). What is musical prosody ? In ROSS, B. H., éditeur : *The Psychology of Learning and Motivation*, volume 46, pages 245–278. Elsevier.
- PEETERS, G. (2004). A large set of audio features for sound description (similarity and description) in the cuidado project. Rapport technique, I.R.C.A.M., Paris.
- PENEL, A. et DRAKE, C. (2004). Timing variations in music performance : Musical communication, perceptual compensation, and/or motor control ? *Perception & Psychophysics*, 66(4):545–562.
- PICARD, F. (2003). *La musique chinoise*. YOU-FENG, Paris, 2nd édition.
- PICINBONO, B. (1997). On instantaneous amplitude and phase of signals. *IEEE Transactions on Signal Processing*, 45(3):552–560.
- PIERCE, A. D. (1981). *Acoustics*. McGraw-Hill, presently available from Acoust. Soc. Am., New York (1990), New York.
- PLANTEVIN, T. (2008). Une analyse de “lo cant del boier”. <http://www.zictrad.free.fr/>.
- PLOMP, R. (1976). *Aspects of tone sensation*. Academic Press, New York.
- POLLARD, H. et JANSSON, B. (1982). A tristimulus method for the specification of musical timbre. *Acta Acustica united with Acustica*, 51:162–71.
- PORTNOFF, Michael, R. (1976). Implementation of the digital phase vocoder using the fast fourier transform. *IEEE Trans. on Acoust., Speech and Sig.. Proc.*, ASSP-24(3):243–248.
- REPP, B. H. (1992). Diversity and commonality in music performance : An analysis of timing microstructure in Schumann's Träumerei. *Journal of the Acoustical Society of America*, 92(5): 2546–2568.
- RISSET, J.-C. (1966). Computer study of trumpet tones. *Bell Telephone Laboratories, Murray Hill N. J. and J. Acoust. Soc. Am.* 38, p. 912, 1965. (abstract only).
- RISSET, J.-C. (1991). Timbre et synthèse des sons. In CHRISTIAN BOURGOIS, éditeur : *Le timbre, métaphore pour la composition*, Musique / Passé / Présent, pages 239–260. I.R.C.A.M., Paris.
- RISSET, J.-C. (1994). Quelques aspects du timbre dans la musique contemporaine (A few aspects of timbre in contemporary music). In ZENATTI, A., éditeur : *Psychologie de la musique (Psychology of music)*, pages 87–114. Presses Universitaires de France, Paris, 1st édition.
- RISSET, J.-C. (2003). Le timbre. In NATTIEZ, J.-J., éditeur : *Musiques, une encyclopédie pour le XXI^e siècle*, volume Tome 2 - Les savoirs musicaux de Cité de la musique, pages 134–161. Actes Sud, Arles.

- RISSET, J.-C. et MATHEWS, M. (1969). Analysis of musical-instrument tones. *Physics Today*, 22(2):23–30.
- RISSET, J.-C. et WESSEL, D. (1991). Exploration du timbre par analyse et synthèse. In CHRISTIAN BOURGOIS, éditeur : *Le timbre, métaphore pour la composition*, Musique / Passé / Présent, pages 102–131. I.R.C.A.M., Paris.
- RISSET, J.-C. et WESSEL, D. L. (1999). Exploration of timbre by analysis and synthesis. In DEUTSCH, D., éditeur : *Psychology of Music*. Academic Press, 2nd édition.
- ROADS, C. (1998). *L'audionumérique (Traduit et adapté de l'anglais par Jean de Reydellet)*. Dunod.
- ROUSSEAU, J.-J. (2007). *Dictionnaire de musique*. Actes Sud, Arles. Fac-similé de l'édition de 1768.
- SAPORTA, G. (1990). *Probabilités Analyse des données et Statistique*. Editions Technip, 2nd édition.
- SCHAEFFER, P. (1966). *Traité des objets musicaux (Treaty of musical objects)*. Éditions du seuil.
- SCHAEFFER, P. (2005). *Solfège de l'objet sonore*. Institut National de l'Audiovisuel, Paris, 2ème édition.
- SCHOENBERG, A. (1911). *Harmonielehre*. Universal, Leipzig/Vienne ; trad. franç., 1983, Lattès, Paris.
- SCHOLES, P. A. (1960). *The Oxford Companion to Music*. Oxford University Press, Oxford, 2nd édition.
- SCHÖN, D. et BESSON, M. (2005). Visually induced auditory expectancy in music reading : A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 17(4):693–704.
- SCHUMACHER, R. T. (1981). Ab initio calculations of the oscillations of a clarinet. *Acustica*, 48:71–85.
- SEASHORE, C. E. (1938). *Psychology of Music*. McGraw-Hill - Reprinted 1967 by Dover Publications, New York.
- SEBESTIK, M. (2005). Dvd Ecoute. Arte vidéo.
- SENSONS (2008). Site web du projet ANR Sensons. <http://www.sensons.cnrs-mrs.fr/>.
- SHEPARD, R. N. (1962a). The analysis of proximities : multidimensional scaling with an unknown distance function. ii. *Psychometrika*, 27:219–246.
- SHEPARD, R. N. (1962b). The analysis of proximities : mutlidimensional scaling with an unknown distance function. i. *Psychometrika*, 27:125–140.
- SIEGEL, S. et JOHN CASTELLAN, Jr., N. (1988). *Non parametric statistics for the Behavioral Sciences*, chapitre Measures of association and their tests of significance, page 272. Statistic. McGraw-Hill International Editions, 2nd édition.
- SILVA, F., GUILLEMAIN, P., KERGOMARD, J., MALLARONI, B. et NORRIS, A. N. (2008). Approximation formulae for the acoustic radiation impedance of a cylindrical pipe. *Journal of Sound and Vibration*, 322:255–263.

- SIRON, J. (2004). *La Partition Intérieure*. Théories. Outre Mesure, Paris, 6th édition.
- SLAWSON, A. W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *J. Acoust. Soc. Am.*, 43(1).
- SMITH (III), J. O. (1998). Principles of digital waveguide models of musical instruments. In KAHRS, M. et BRANDENBURG, K., éditeurs : *Applications of DSP to Audio and Acoustics*, pages 417–466. Kluwer Academic Publishers, Dordrecht.
- SMITH (III), J. O. et SERRA, X. (1990). Spectral modeling synthesis : A sound analysis / synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4): 12–24.
- SNEATH, P. H. A. (1957). The application of computer to taxonomy. *Journal of General Microbiology*, 17:201–226.
- STEVENS, S. (1957). Concerning the form of the loudness function. *Journal of the Acoustical Society of America*, 29:603–606.
- STEVENS, S. S., VOLKMAN, J. et NEWMAN, E. (1937). A scale for the measurement of the psychological magnitude of pitch. *J. Acoust. Soc. Am.*, 8(3):185–190.
- STRAVINSKY, I. (1971). *Chroniques de ma vie*. Médiations-Gonthier, Paris.
- SUNDBERG, J. (1995). Le chant. In *Les instruments de l'orchestre*, pages 20–29. Pour la science Diffusion Belin, Paris.
- SUNDBERG, J. (2000). Grouping and differentiation two main principles in the performance of music. In *Integrated Human Brain Science : Theory, Method Application (Music)*, pages 299–314. Elsevier Science B.V.
- SUNDBERG, J., FRIBERG, A. et FRYDÉN, L. (1989). Rules for automated performance of ensemble music. *Contemporary Music Review*, 3:89–109.
- SUNDBERG, J., FRYDÉN, L., FRIBERG, A., BRESIN, R. et COLOMBO, V. (2008). Director musices. <http://www.speech.kth.se/music/performance/download/>.
- SUSINI, P. et MCADAMS, S. (2000). Effet de récence dans une tâche de jugement de la sonie. In *5ème Congrès français d'acoustique*, Lausanne, Suisse.
- SUSINI, P., MCADAMS, S. et WINSBERG, S. (1999). A multidimensional technique for sound quality assessment. *Acta Acustica*, 85:650–656.
- TAILLARD, P.-A. (2004). Claripatch. <http://www.claripatch.com/>.
- TAMBA, A. (1988). Cd japon : Musique du nô - shakkyô : Pont en pierres. Radio France.
- TERASAWA, H., SLANEY, M. et BERGER, J. (2006). A statistical model of timbre perception. In *ISCA Tutorial and Research Workshop on Statistical And Perceptual Audition (SAPA2006)*, pages 18–23, Pittsburgh.
- TERHARDT, E., STOLL, G. et SEEWANN, M. (1982a). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 71(3):679–688.
- TERHARDT, E., STOLL, G. et SEEWANN, M. (1982b). Pitch of complex signals according to virtual-pitch theory : Tests, examples, and predictions. *Journal of the Acoustical Society of America*, 71(3):671–678.

- THOMPSON, S. (1979). The effect of the reed resonance on the woodwind tone production. *J. Acoust. Soc. Am.*, 66(5):1299–1307.
- TOBUDIC, A. et WIDMER, G. (2003). Playing Mozart phrase by phrase. Rapport technique, ÖFAI-TR-2003-02.
- TODD, N. P. M. (1992). The dynamics of dynamics : A model of musical expression. *Journal of the Acoustical Society of America*, 91(6):3540–3550.
- TODD, N. P. M. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, 97(3):1940–1949.
- TOKUMARU, Y. (1991). Le timbre dans la musique japonaise. In CHRISTIAN BOURGOIS, éditeur : *Le timbre, métaphore pour la composition*, pages 90–97. I.R.C.A.M., Paris.
- TOLSTOÏ, L. (1931). *Qu'est-ce que l'art ?* Presses Universitaires de France, Paris.
- TOYOTA (2008). Page web sur le robot trompettiste toyota. <http://www.toyota.co.jp/en/special/robot>.
- TRAUBE, C. (2004). *An Interdisciplinary Study of the Timbre of the Classical Guitar*. Thèse de doctorat, Music Technology, Department of Theory, Faculty of Music, McGill University, Montreal, Canada.
- TRAUBE, C., DEPALLE, P. et WANDERLEY, M. (2003). Indirect acquisition of instrumental gesture based on signal, physical and perceptual information. In *Proc. of the 2003 Conference on New Interfaces for Musical Expression (NIME-03)*, Montreal, Canada.
- TZANETAKIS, G. (2002). *Manipulation, analysis and retrieval systems for audio signals*. Thèse de doctorat, Princeton University.
- van WALSTIJN, M. et CAMPBELL, M. (2003). Discrete-time modelling of woodwind instrument bores using wave variables. *J. Acoust. Soc. Am.*, 113:575–585.
- VANDOREN (2000). Vandoren magazine n° 2 - le choix d'un bec (première partie : clarinette).
- VERFAILLE, V. (2003). *Effets audio numériques adaptatifs : théorie, mise en oeuvre et usage en création musicale numérique*. Thèse de doctorat, Université Aix-Marseille II.
- VINES, B. W., KRUMHANS, C. L., WANDERLEY, M. M. et LEVITIN, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, 101:80–113.
- VOS, J. et RASCH, R. (1981). The perceptual onset of musical tones. *Percept. Psychophys.*, 29:323–335.
- WANDERLEY, M. (2002). Quantitative analysis of non-obvious performer gestures. In *Gesture and Sign Language in Human-Computer Interaction : International Gesture Workshop*, page 241. Springer, Berlin / Heidelberg.
- WAPNICK, J., FLOWERS, P., ALEGANT, M. et JASINSKAS, L. (1993). Consistency in piano performance evaluation. *Journal of Research in Music Education*, 41(4):282–292.
- WARD, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236–244.
- WESSEL, D. L. (1973). Psychoacoustics and music : A report from Michigan State University. *PACE : Bulletin of the Computer Arts Society*, 30:1–2.

- WESSEL, D. L. (1978). Timbre space as a musical control structure. *Rapport Ircam*, (12/78). (improved version of the initial article published in Vol. 3(2) of the *Computer Music Journal*, 1979).
- WESSEL, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3(2):45–52.
- WESSEL, D. L., BRISTOW, D. et SETTEL, Z. (1987). Control of phrasing and articulation in synthesis. In *Proc. of the 1987 International Computer Music Conference (ICMC)*, San Francisco. ICMA.
- WHYBREW, W. E. (1962). *Measurement and Evaluation in Music*, page 63. William C. Brown Company Publishers, Dubuque, Iowa.
- WIDMER, G. et GOEBL, W. (2004). Computational models of expressive music performance. *Journal of New Music Research*, 33(3):203–216.
- WIER, C. C., JESTEADTT, W. et GREEN, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.*, 6(1):178–184.
- WILSON, T. A. et BEAVERS, G. S. (1974). Operating modes of the clarinet. *J. Acoust. Soc. Am.*, 56(2):653–658.
- WINDSOR, W. L., DESAIN, P., PENEL, A. et BORKENT, M. (2006). A structurally guided method for the decomposition of expression in music performance. *Journal of the Acoustical Society of America*, 119(2):1182–1193.
- WINSBERG, S. et CAROLL, J. D. (1988). A quasi-nonmetric method for multidimensional scaling via an extended euclidean model. *Psychometrika*, 53:217–229.
- WINSBERG, S. et DE SOETE, G. (1993). A latent class approach to fitting the weighted euclidean model : Clascal. *Psychometrika*, 58:315–330.
- WORMAN, W.-E. (1971). *Self-sustained non-linear oscillations of medium amplitude in clarinet like systems*. Thèse de doctorat, Case Western Reserve University, Cleveland.
- YSTAD, S. (1998). *Sound modeling using a combination of physical and signal models*. Thèse de doctorat, Université Aix-Marseille II.
- ZÖLZER, U. (1997). *Dynamic range control*. John Wiley & Sons.
- ZÖLZER, U. (2002a). *DAFx Digital Audio Effects*, chapitre Time-frequency processing, page 237. John Wiley and Sons, Ltd.
- ZÖLZER, U. (2002b). Fonctions *Matlab* associées au livre *Digital Audio Effects*. http://www2.hsu-hh.de/ant/dafx2002/DAFX_Book_Page/matlab.html.
- ZWICKER, E. et FASTL, H. (1990). *Psychoacoustics : Facts and Models*. Springer Series in Information Sciences. Springer-Verlag, New York.
- ZWICKER, E. et FELDTKELLER, R. (1981). *Psychoacoustique*. Masson, Paris. Traduction par C. Sorin de l'ouvrage allemand "Das Ohr als Nachrichtenempfänger", Hirzel, Stuttgart, 1967.
- ZWICKER, E. et SCHARF, B. (1965). A model of loudness summation. *Psych. Rev.*, 72:3–26.

Résumé

Découvrir quels sont les paramètres acoustiques corrélés à l'expression musicale revêt un intérêt fondamental afin d'améliorer notre compréhension de la perception musicale, et trouve de multiples applications en synthèse sonore. L'interprétation musicale est un acte au cours duquel un instrumentiste traduit le signal notationnel d'un compositeur en sons musicaux, tout en impliquant son propre jugement. Si la plupart des études sur l'interprétation musicale portent sur le rythme, la dynamique, ou les hauteurs, bien moins se consacrent au timbre. Cet attribut perceptuel du son, à la fois lié à l'identité de la source sonore et à la qualité sonore, semble pourtant être un paramètre musical essentiel. L'objet de cette thèse, dont l'approche repose sur l'analyse-synthèse, est de mieux comprendre le rôle du timbre dans l'interprétation musicale.

Une étude de dissemblance réalisée sur des sons de synthèse de clarinette, obtenus à partir d'un modèle physique, a permis d'évaluer l'influence du contrôle instrumental (pression d'alimentation et pince) sur les timbres produits par l'instrument. Des enregistrements d'extraits musicaux joués un grand nombre de fois par un même clarinettiste professionnel sur un instrument naturel selon différentes intentions musicales ("scolaire" et "expressive") ont ensuite été analysés. Les mécanismes de transmission de l'expression musicale ont ainsi pu être étudiés au travers de changements de timbre, de rythme et de dynamique. Certaines variations de timbre (variations de qualité sonore au sein des notes et entre les différentes notes) sont reproduites de manière systématique par l'interprète lorsque son intention musicale est la même. La nature de ces variations change lorsque l'intention expressive change, ce qui tend à prouver que les musiciens agissent sur certaines dimensions du timbre afin de varier leur expression. Deux expériences perceptives complémentaires ont révélé que la nature des évolutions temporelles de la brillance des notes influe sur les préférences musicales des auditeurs. La qualité musicale de séquences inexpressives, produites sur des instruments de type entretenu à l'aide d'échantillonneurs, a notamment pu être améliorée de manière significative grâce à un contrôle de la brillance par filtrage dynamique.

L'ensemble de ces travaux appuie l'idée que les variations morphologiques temporelles de timbre (par ex. variations temporelles de brillance) constituent l'un des vecteurs de l'expression musicale.

Discipline : Acoustique, Traitement du signal et Informatique appliqués à la Musique.

Mots-clefs : timbre, interprétation musicale, expression musicale, perception musicale, analyse-synthèse, transformation sonore, modélisation physique, clarinette.

Laboratoire :

CNRS - Laboratoire de Mécanique et d'Acoustique (UPR 7051)
31 chemin Joseph-Aiguier
13402 Marseille Cedex 20
France