

N° d'ordre : 3327

THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ BORDEAUX I

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET
D'INFORMATIQUE

Par **Matthias Robine**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

**ANALYSE DE LA PERFORMANCE MUSICALE ET
SYNTHÈSE SONORE RAPIDE**

Soutenue le : 13 Décembre 2006

Après avis des rapporteurs :

Daniel Arfib Directeur de Recherche
Caroline Traube Professeure adjointe

Devant la commission d'examen composée de :

Daniel Arfib	Directeur de Recherche	Rapporteur
Philippe Depalle	Professeur	Examineur
Myriam Desainte-Catherine	Professeur	Examineur
Sylvain Marchand	Maître de Conférences	Examineur
Robert Strandh	Professeur	Examineur
Caroline Traube	Professeure adjointe ...	Rapporteur

Analyse de la performance musicale et synthèse sonore rapide

Résumé :

Cette thèse d'informatique musicale explore d'une part la performance instrumentale et propose d'autre part des algorithmes de synthèse additive rapide.

Un état de l'art sur l'analyse du jeu instrumental est d'abord réalisé, explorant les différentes composantes de l'interprétation musicale. Une étude sur l'importance du doigté au piano est alors présentée. La performance pianistique est ainsi analysée pour mettre en évidence l'influence du doigté sur la performance. Le jeu académique au saxophone est aussi analysé, afin d'évaluer le niveau technique de saxophonistes en fonction de l'évolution de leurs paramètres sonores. Une méthode automatique de cette évaluation est alors proposée.

Dans une deuxième partie, nous explorons des algorithmes de synthèse additive rapide. Nous étudions d'abord la possibilité d'avoir recours à des techniques non linéaires. Nous présentons ensuite PASS, une nouvelle méthode où des polynômes remplacent des fonctions sinusoïdales pour accélérer la synthèse sonore. Une application de la méthode PASS est finalement présentée, permettant la synthèse rapide et réaliste de surfaces océaniques.

Discipline : Informatique

Mots-clés : jeu instrumental, composantes de l'interprétation musicale, synthèse sonore, synthèse additive rapide.

LaBRI,
Université Bordeaux 1,
351 cours de la Libération,
33405 Talence Cedex (FRANCE).

Musical Performance Analysis and Fast Sound Synthesis

Abstract:

This document deals in a first part with musical performance and proposes in a second part some algorithms for fast sound synthesis.

We begin by discussing the several parameters which could be extracted from a musical performance. We expose then our work on the piano fingering. A lot of biomechanic studies have previously highlight the influence of the physiology of a pianist in his performance, particularly due to his hands. This knowledge lead us to propose a new method of automatic fingering which uses dynamic fingering. We propose also to evaluate the technical level of a musical performer by analysing non expressive performances like scales. Our results are based on the analysis of alto saxophone performances, however the same approach can be used with other instruments. Our aim is to highlight the technical part in the performance by considering the evolution of the spectral parameters of the sound.

The second part of the document propose to study the use of non linear methods of sound synthesis for the additive synthesis. Then, we propose a new fast sound synthesis method using polynomials. This is an additive method where polynomials are used to approximate sine functions. Practical implementations show that this method called *Polynomial Additive Sound Synthesis* (PASS) is particularly efficient for low-frequency signals. We propose finally to adapt fast sound synthesis methods to the simulation of ocean waves.

Discipline: Computer-Science

Keywords: components and evaluation of the musical performance, sound synthesis, fast additive synthesis.

LaBRI,
Université Bordeaux 1,
351 cours de la Libération,
33405 Talence Cedex (FRANCE).

Remerciements

Je tiens à remercier toutes les personnes qui m’ont accompagné durant ces années de doctorat. Au sein du LaBRI, dans les équipes pédagogiques, avec mes amis ou dans ma famille, les supports ont tous été chaleureux et je vous en remercie. Ils ont chacun contribué à l’achèvement de ce document.

Robert Strandh, mon directeur de thèse, a ainsi été à l’écoute de mes idées. Il m’a fourni la matière scientifique qui s’est développée au long de ces dernières années. Sylvain Marchand aussi, qui m’a emmené sur les voies de la synthèse additive, et avec qui j’ai participé à ma première conférence internationale chez nos amis québécois. Merci à Mathieu Lagrange qui a participé activement aux travaux sur le saxophone que je présente dans ce mémoire.

Merci évidemment aux rapporteurs de mon manuscrit, Caroline Traube et Daniel Arfib, qui ont d’abord accepté cette tâche et qui ont apporté ensuite leur avis éclairé sur mes recherches. Un grand merci à Phillippe Depalle, Myriam Desainte-Catherine, Robert Strandh et Sylvain Marchand d’avoir constitué le remarquable jury de ma soutenance.

Je tiens à remercier l’équipe son du LaBRI qui m’a accueilli lors de ces travaux. L’échange scientifique y a toujours été enrichissant. Merci au SCRIME qui a notamment financé mes missions, et particulièrement à Myriam Desainte-Catherine qui a toujours un intérêt pour la chose musicale. Merci aussi à l’Université de Bordeaux et au ministère qui m’ont financé durant ces 3 années.

Ma thèse a également révélé ma passion pour l’enseignement supérieur. J’ai eu la chance de pouvoir enseigner dans différents établissements universitaires et je remercie vivement Marc Valat, Françoise Marc du Département Universitaire des Sciences d’Agen, Giuliaana Bianchi, Carole Blanc, Robert Strandh à l’Université de Bordeaux 1 ainsi que mes collègues de l’IUT Bordeaux 1 de leur confiance et de leur accompagnement dans ces riches expériences humaines et pédagogiques.

Les amis et collègues ont bien sûr été très importants durant ces années au LaBRI. Je remercie particulièrement l’inénarrable Joachim Pouderoux de tous ces moments passés ensemble et de sa générosité, notamment concernant son expertise dans l’informatique ; Alexandre Pinlou aussi, qui m’a soutenu quand le prompt clignotait mystérieusement le premier jour de mes études d’informatique et qui a partagé avec moi l’expérience d’une thèse, tout comme Martin Raspaud ; Jocelyn Fréchet, Fanny Chevalier et Florian Iragne avec qui j’ai partagé un bureau agréable avec des discussions ouvertes. Un merci chaleureux également à Pierre Hanna, avec qui nous partageons des perspectives passionnantes de recherche et qui m’a fait confiance dans son projet Simbals.

Enfin, je voudrais souligner le soutien fort et permanent de ma famille qui a joué un rôle très important dans l'aboutissement de ce travail. Je pense notamment à la belle Delphine et aux deux trésors que sont Noémie et Camille.

Table des matières

Remerciements	v
Introduction	1
I Notions d’informatique musicale	5
1 Onde sonore et perception	9
1.1 Éléments physiques	9
1.2 Perception auditive	10
1.2.1 L’oreille	10
1.2.2 Psychoacoustique	11
1.2.3 Paramètres sonores	12
1.2.4 Propriétés perceptives	14
1.2.5 Masquages	15
1.2.6 Paramètres perceptifs	17
2 Modèles de sons	19
2.1 Domaine temporel	20
2.1.1 Échantillonnage	20
2.1.2 Quantification	21
2.2 Analyse sinusoïdale	22
2.3 Domaine fréquentiel	23
2.3.1 Spectre	23
2.3.2 Propriétés spectrales	23
2.3.3 Enveloppe spectrale	25
2.4 Synthèse additive	25
2.4.1 Résonateur numérique	27
2.4.2 Transformée de Fourier inverse	28
3 Musique instrumentale	31
3.1 Notation musicale	32
3.2 Format MIDI	32
3.3 Interprétation musicale	33
3.4 Paramètres musicaux	34
3.4.1 Rythme	34

3.4.2	Hauteur de note	34
3.4.3	Nuance	35
3.4.4	Timbre	36
3.4.5	Vibrato et tremolo	36
3.4.6	Articulation	37
3.4.7	Phrasé	37
II	Analyse et évaluation du jeu instrumental	39
4	Analyse du jeu instrumental	43
4.1	Paramètres de la performance musicale	44
4.1.1	Composante physique	44
4.1.2	Composante technique	44
4.1.3	Composante expressive	45
4.2	Analyse des paramètres sonores et musicaux	47
4.3	Analyse des écarts à un modèle	47
4.3.1	Modèle rythmique	47
4.3.2	La partition comme modèle	48
4.3.3	L'évolution théorique des paramètres sonores	49
4.4	Influence physique	49
4.4.1	Modèles physiques d'instruments	50
4.4.2	Physiologie de l'instrumentiste	50
4.5	Technique	50
4.6	Expressivité	52
4.6.1	Différenciation et reconnaissance d'interprètes	52
4.6.2	Humanisation d'une performance automatique	52
4.7	Éléments de pédagogie musicale	53
4.7.1	Contexte académique	53
4.7.2	Exercices instrumentaux de base	53
4.7.3	Niveau technique et expressivité	54
4.7.4	Évaluation en pédagogie musicale	54
5	Le doigté au piano	57
5.1	Le piano	57
5.2	Le doigté au piano	58
5.3	Biomécanique et piano	58
5.4	Doigté automatique	59
5.4.1	Transitions	63
5.4.2	Matrices de pondération	63
5.4.3	Doigté dynamique	65
5.5	Analyse de la performance	66
5.5.1	Protocole d'enregistrement et analyse	66
5.5.2	Reconnaissance du doigté	67
5.5.3	Perspectives	70

6	Évaluation de la performance au saxophone	73
6.1	Éléments de pédagogie pour le saxophone	74
6.1.1	Le saxophone	74
6.1.2	Contexte académique	74
6.1.3	Exercices de base	76
6.1.4	Évaluation du niveau technique	77
6.2	Protocole d'enregistrement des saxophonistes	78
6.3	Analyse de l'évolution des paramètres du son	79
6.4	Métriques d'évaluation	79
6.4.1	Écart type pondéré	79
6.4.2	Analyse glissante	81
6.4.3	Métriques pour les sons droits	81
6.4.4	Ambitus de nuance	81
6.4.5	Métriques pour les sons filés	83
6.4.6	Métriques pour les sons vibrés	83
6.5	Résultats	86
6.5.1	Conversion des métriques aux notes	86
6.5.2	Interprétation des résultats	86
6.6	Évaluation du niveau technique général	88
6.6.1	Note technique générale	88
6.6.2	Classement automatique des instrumentistes	91
6.7	Perspectives	91
6.7.1	Saxophone tutor	91
6.7.2	Généralisation de la méthode à d'autres instruments	93
III	Synthèse Sonore Rapide	97
7	Techniques non linéaires pour la synthèse des sons harmoniques	101
7.1	Modulation d'amplitude	102
7.1.1	Modulation d'amplitude RM	102
7.1.2	Modulation d'amplitude AM	103
7.1.3	Résultats	105
7.2	Modulation de fréquence	105
7.2.1	FM formantique	105
7.2.2	Autres techniques de modulation de fréquence	110
7.2.3	Résultats	110
7.3	Distorsion et polynômes	111
7.3.1	Distorsion	111
7.3.2	Polynômes de Tchebycheff	113
7.3.3	Polynômes quelconques	115
7.3.4	Résultats	115
7.4	Des fonctions remarquables	115
7.4.1	Valeur absolue	115
7.4.2	Développements limités	118
7.4.3	Résultats	121

7.5	Formes analytiques	123
7.5.1	Peigne harmonique	123
7.5.2	Synthèse DSF	123
7.5.3	Résultats	126
7.6	Conclusion	126
8	Synthèse additive polynomiale	129
8.1	Évolution polynomiale des paramètres sonores	129
8.2	Polynomial Additive Sound Synthesis (PASS)	130
8.2.1	Approximation polynomiale des fonctions sinusoïdales	130
8.2.2	Calcul incrémental des polynômes	134
8.2.3	Générateur polynomial	134
8.2.4	Structure de données	136
8.2.5	Changement des paramètres sonores	138
8.2.6	Complexité	138
8.2.7	Résultats	140
8.3	Méthode hybride	140
8.4	Application à la synthèse de surfaces océaniques	144
8.4.1	Simulation réaliste d'océans	144
8.4.2	Modèle de surface océanique	144
8.4.3	Utilisation des méthodes sonores	145
8.4.4	Complexité des méthodes	147
8.4.5	Implémentation dans la bibliothèque Aqua	147
8.4.6	Résultats	147
	Conclusion	151
	Bibliographie	152

Liste des tableaux

1	24 bandes critiques de fréquence en Hertz.	17
2	Principaux types d'événements MIDI.	33
3	Correspondance entre la hauteur de note et la fréquence du signal sonore correspondant.	35
4	Exemple de nuances.	36
5	Résultats pour la note Sol aigu.	87
6	Résultats de <i>Pierre</i> , un saxophoniste de niveau moyen.	88
7	Résultats de <i>Shang</i> , un saxophoniste confirmé.	89
8	Approximation de $ x $ par un polynôme de degré 10 en fonction de l'amplitude a , pour $x = a \cos(2\pi ft)$	117
9	Développement limité en 0 d'ordre 5 de la fonction $\ln(1 + a \cos(2\pi ft))$ en fonction de l'amplitude a du signal d'entrée.	121
10	Développements limités en 0 des principales fonctions usuelles.	122
11	Principales formules analytiques utilisables pour la synthèse DSF.	125
12	Erreur de l'approximation polynomiale du signal d'un partiel.	132
13	Comparaison du temps de calcul du résonateur numérique et de la méthode PASS pour 5 secondes de synthèse sonore, et pour différentes fréquences d'échantillonnage du son.	141
14	Comparaison du temps de calcul du résonateur numérique et de la méthode PASS pour 5 secondes de synthèse sonore, avec une fréquence d'échantillonnage $F_e = 44100$ Hz.	141
15	Temps de calcul d'une grille uniforme avec la méthode sincos, le résonateur numérique et la méthode PASS, pour différents nombres de points, nombres de vagues, largeurs de grille et vitesses de vent.	148

Table des figures

1	Schéma de l'oreille.	11
2	La cochlée.	12
3	Masquages.	16
4	Courbes isosoniques.	18
5	Chaîne de traitement du son.	19
6	Numérisation du son.	21
7	Représentation temporelle d'un signal sinusoïdal pur et sa représentation spectrale.	24
8	Spectre obtenu par transformée de Fourier.	24
9	Spectre d'un son harmonique, enveloppe spectrale et formants.	26
10	Spectres des fenêtres de Bartlett et de Hann.	29
11	Échelle des fréquences de variation des paramètres.	31
12	Le <i>Performance Worm</i> de Langner et Goebl propose une visualisation de la performance musicale expressive.	46
13	Les robots musiciens.	51
14	Numérotation des doigts au piano.	59
15	Gamme de Do majeur pour le piano, avec indication du doigté conventionnel.	60
16	Travaux biomécaniques de Ortmann [Ort62] sur le piano.	60
17	Travaux biomécaniques de McKenzie et VanEerd [KV90] sur le doigté au piano : intervalle moyen entre les notes.	61
18	Travaux biomécaniques de McKenzie et VanEerd [KV90] sur le doigté au piano : vitesse d'appui sur les touches.	62
19	Notation pour les transitions entre les notes du clavier.	64
20	Coefficient de <i>staccato</i> pour une gamme de Do majeur jouée au piano à la main droite.	68
21	Valeurs de vélocité MIDI pour une gamme de Do majeur jouée au piano à la main droite.	69
22	Méthode de reconnaissance de doigté à partir de la performance.	71
23	Saxophone alto en Mi bémol, avec indication des clés.	75
24	Exercices réalisés par les saxophonistes pendant leur enregistrement.	78
25	Vecteurs de fréquence et d'amplitude d'un son de saxophone extraits avec le logiciel Praat [BW06].	80
26	Vecteurs de fréquence et d'amplitude pour un son filé joué par deux saxophonistes différents.	82
27	Vecteur d'amplitude et vecteur linéaire par morceaux pour deux sons filés joués par deux saxophonistes différents.	84

28	Vecteur de fréquence du son vibré d'un saxophoniste expert, évolution de la fréquence et de l'amplitude du vibrato.	85
29	Notes techniques générales obtenues par les saxophonistes.	90
30	Dendrogramme du classement des saxophonistes.	92
31	Visualisation des résultats de Lilian en fonction du type d'exercice.	94
32	Visualisation des résultats de Paul en fonction de la note jouée.	95
33	Signal bipolaire et unipolaire.	103
34	Exemple de modulation avec un signal modulant de fréquence $f_m = 400$ Hz et un signal porteur de fréquence $f_c = 1000$ Hz.	104
35	Application de la technique AM sur les maxima de l'enveloppe spectrale d'un signal.	106
36	Exemples de fonctions de Bessel de première espèce.	108
37	Synthèse FM en fonction de l'indice de modulation I	109
38	Synthèse FM pour un signal de voix.	112
39	Synthèse par distorsion. Exemple de quatre fonctions de transfert.	113
40	Application de la fonction non linéaire $g(x) = x^2 + x^3$ sur le signal dont le spectre est constitué de deux composantes sinusoïdales.	116
41	Application de la fonction $ x $ sur les maxima de l'enveloppe spectrale d'un signal.	118
42	Extension d'un signal à bande limitée par la fonction $ x $	119
43	Exemples de synthèse DSF.	124
44	Principe général de la méthode PASS.	130
45	Approximation polynomiale d'un signal sinusoïdal.	133
46	Les événements dans la méthode PASS.	135
47	Méthodes <i>delete-max</i> et <i>insert</i> pour une file de priorité implémentée par un tas.	137
48	Méthode <i>replace</i> pour une file de priorité implémentée par un tas.	137
49	Changement du paramètre d'amplitude quand le signal est minimal ou maximal.	138
50	Changement du paramètre de fréquence quand le signal est minimal ou maximal.	139
51	Comparaison du temps de calcul du résonateur numérique et de la méthode PASS pour 5 secondes de synthèse sonore. Le temps de calcul est fonction du nombre de partiels, avec une fréquence moyenne fixe.	142
52	Comparaison du temps de calcul du résonateur numérique et de la méthode PASS pour 5 secondes de synthèse sonore. Le temps de calcul est fonction de la fréquence moyenne des partiels, avec un nombre fixé de partiels.	143
53	Grille adaptative avec une résolution d'écran de 1024×768 pixels, 256×192 points sur la grille, 400 vagues et une vitesse de vent de $5 \text{ m}\cdot\text{s}^{-1}$	150
54	Surface océanique de haute qualité : 800 vagues avec une grille de 1024×768 points.	150

Introduction

Cette thèse d’informatique musicale présente nos travaux dans le domaine de l’analyse de la performance musicale, et dans celui de la synthèse sonore rapide. Composé de trois parties principales, ce document a pour objectifs d’introduire les notions utiles pour la compréhension du sujet, de parcourir les techniques existantes dans différents domaines, et de présenter nos travaux et nos résultats. La partie I introduit des notions sur la physique de l’onde sonore, sur les modèles de sons et la musique instrumentale. La suite de ce document se base alors sur ces éléments généraux. Nous présentons ensuite dans la partie II nos recherches sur la performance musicale instrumentale, de son analyse à son évaluation. Nous nous intéressons plus particulièrement au doigté des pianistes et à la technique des saxophonistes. La partie III relate quant à elle nos travaux menés dans le domaine de la synthèse additive. Il s’agit de trouver des méthodes rapides de synthèse, éventuellement pour des cas particuliers de signaux sonores, tels que les signaux harmoniques. Des techniques de synthèse non linéaires ont ainsi été étudiées et évaluées, et une nouvelle méthode de synthèse additive rapide est proposée.

I - Notions d’informatique musicale

Cette partie introduit quelques représentations et modélisations pour le son et la musique. Le chapitre 1 présente ainsi les propriétés de la physique du son, d’une source sonore à la perception de l’onde sonore qu’elle génère. Nous y étudions la physique et les propriétés de l’appareil auditif humain. Nous explicitons aussi les paramètres sonores qui permettent de décrire le son. La modélisation sonore, et notamment la modélisation sinusoïdale, est introduite dans le chapitre 2. Après avoir expliqué la technique de numérisation d’une onde sonore analogique, nous suivons la chaîne de traitement du son dans le modèle sinusoïdal. Nous traitons donc de l’analyse sonore par transformée de Fourier, puis des notions de *spectre*, *enveloppe spectrale* ou *formant*. Nous présentons aussi le principe de synthèse additive qui est à la base des modèles spectraux. Les techniques rapides pour ce type de synthèse sont étudiées et comparées entre elles. Cette section nous sert de référence pour nos travaux sur la synthèse additive, exposés ultérieurement. Afin d’introduire la seconde partie de ce document, orientée vers l’analyse de la performance musicale, le chapitre 3 donne des notions de musique instrumentale académique. Nous y discutons de la notation musicale, et de l’interprétation instrumentale notamment. Nous définissons des paramètres musicaux importants, tels que le phrasé ou l’articulation. Cette partie sert de base aux méthodes d’analyse et d’évaluation de la performance musicale dans les chapitres suivants.

II - Analyse et évaluation de la performance musicale

Cette partie présente nos recherches sur la performance musicale, et plus précisément sur la performance instrumentale académique. Nous proposons dans le chapitre 4 de décomposer la performance instrumentale en trois parties principales : physique, technique et expressive. Nous recensons alors les méthodes existantes spécialisées dans chacune de ces parties, et nous présentons leurs techniques d'analyse. Nous donnons ensuite des indications de pédagogie instrumentale sur l'évaluation de la performance dans un contexte académique. L'identification et la séparation des paramètres physiques, techniques et expressifs de la performance et le recours à des connaissances pédagogiques instrumentales posent les bases de méthodes d'analyse et d'évaluation spécifiques, proches de l'évaluation humaine de la performance.

Nous exposons dans le chapitre 5 les recherches menées sur le doigté au piano. Après une présentation de l'instrument et de ses contraintes physiques, nous présentons les travaux existants sur la main et les doigts du pianiste dans des domaines tels que la biomécanique, la musicologie ou l'informatique. Une méthode de doigté automatique est alors proposée, qui s'appuie sur la programmation dynamique et les contraintes biomécaniques du jeu pianistique. Après avoir reproduit des expériences menées en biomécanique sur l'exécution de gammes au piano, nous proposons de mettre en évidence l'empreinte physico-technique de la main du pianiste dans la performance instrumentale technique. Ce constat sert de base à notre proposition de méthode de reconnaissance de doigté à partir de la performance.

Le chapitre 6 présente nos travaux sur l'évaluation technique de saxophonistes. Nous commençons dans cette partie par décrire la physique du saxophone, puis la pédagogie académique de cet instrument. Nous listons les différents exercices pratiqués par les instrumentistes au cours de leur apprentissage, et les différentes méthodes d'évaluation de leurs performances. Cela nous permet de justifier le protocole et les objectifs d'une nouvelle méthode automatique d'évaluation. Nous avons mené une étude sur une trentaine de saxophonistes. En suivant l'évolution des paramètres spectraux des sons des différents interprètes lors d'exercices techniques imposés, nous proposons une méthode automatique pour évaluer leurs niveaux techniques. Nous utilisons pour cela des métriques d'évaluation combinant des données spectrales du son, et qui correspondent aux critères techniques d'évaluation utilisés par les professeurs de saxophone. Les résultats de notre méthode d'évaluation correspondent bien aux niveaux académiques des saxophonistes observés. Cette méthode nous permet également de visualiser la performance technique d'un élève en fonction du niveau de sa classe instrumentale, et de lui attribuer une note technique générale. Nous concluons ce chapitre en proposant une méthode automatisée de tutorisation de l'apprentissage technique du saxophone.

III - Synthèse sonore rapide

Des années 1960 au début des années 1980 se sont développées de nouvelles techniques de synthèse sonore. Les limites des systèmes informatiques de l'époque ont encouragé la recherche de synthèses non linéaires pouvant générer des signaux sonores complexes avec peu de moyens de calcul. Aujourd'hui nous avons une puissance de calcul bien supérieure, mais les sons traités sont également de plus en plus complexes. Cela nécessite des méthodes de synthèse sonore toujours plus rapides. Pour le cas particulier des sources quasi

harmoniques (cas de la voix ou de nombreux instruments), nous avons alors étudié dans le chapitre 7 la possibilité de recourir à des techniques non linéaires pour la synthèse additive, et nous avons répertorié celles qui peuvent engendrer des structures harmoniques. Le signal généré par ces méthodes est cependant déterminé par des règles mathématiques, son enveloppe spectrale est alors contrôlée par les fonctions mathématiques génératrices des techniques. Cette signature sonore rend difficile la synthèse d'un son harmonique quelconque. Nous concluons donc dans ce chapitre que les techniques non linéaires étudiées ne sont pas assez souples pour synthétiser un son naturel harmonique avec la même qualité qu'une synthèse additive classique.

Nous présentons dans le chapitre 8 une nouvelle méthode rapide de synthèse sonore additive appelée PASS (*Polynomial Additive Sound Synthesis*) qui utilise des polynômes pour approcher et remplacer les fonctions sinusoïdales des oscillateurs. La synthèse additive consiste à additionner les signaux de nombreux oscillateurs. Habituellement, un échantillon est calculé pour chaque oscillateur. Ces échantillons sont ensuite additionnés pour générer un échantillon du signal sonore. Le temps de calcul de la synthèse est donc proportionnel au nombre d'oscillateurs et à la fréquence d'échantillonnage. En utilisant des polynômes pour approcher les oscillateurs, nous pouvons additionner les coefficients polynomiaux pour obtenir les coefficients d'un unique polynôme générateur. Le degré de ce polynôme ne dépend pas du nombre d'oscillateurs, et son évaluation régulière produit les échantillons du son. Cependant, comme l'approximation polynomiale proposée n'est valable que sur une partie de la période de chaque oscillateur, il faut gérer des événements d'actualisation des coefficients polynomiaux du générateur. Nous proposons alors l'utilisation d'une structure de donnée efficace. Nous présentons des résultats comparatifs de notre méthode et de celle du résonateur numérique. Nous constatons que PASS est particulièrement efficace pour les signaux avec de nombreux oscillateurs à basse fréquence, ou quand la fréquence d'échantillonnage du son est très élevée.

Nous proposons finalement une application à la méthode PASS, pour la synthèse temps-réel et réaliste de surfaces océaniques. La création de surfaces océaniques réalistes est une problématique très proche de la synthèse additive. Il s'agit de synthétiser le plus vite possible un très grand nombre de vagues, chaque vague de l'océan étant l'équivalent d'un oscillateur dans le son. Nous avons implémenté les méthodes de synthèse sonore additive les plus rapides dans un logiciel de visualisation réaliste de surfaces océaniques. Nous présentons les résultats de nos tests, avec une amélioration sensible des temps de calcul par rapport aux méthodes existantes. En utilisant un rendu adaptatif de la surface océanique et un filtrage des vagues en fonction de leur longueur d'onde, nous obtenons ainsi une visualisation réaliste et quasiment temps-réel de surfaces océaniques avec PASS.

Première partie

Notions d'informatique musicale

Cette partie présente des notions d’informatique musicale. Nous parcourons entre son et musique une hiérarchie de niveaux de représentation, des paramètres sonores aux paramètres musicaux. Certaines figures de cette partie sont extraites du cours “Analyse et synthèse du son musical” dispensé par Sylvain Marchand à l’Université de Bordeaux.

Nous commençons par l’explication dans le chapitre 1 des phénomènes physiques et psychoacoustiques qui participent à la perception du son. Une source sonore en vibration produit une onde sonore qui peut être captée par l’oreille humaine. Le son est alors le résultat de phénomènes physiques et psychoacoustiques complexes. Nous listons les paramètres qui caractérisent le son.

Le son peut également être capté par un microphone, et numérisé. Cette étape est présentée dans le chapitre 2. Une modélisation du son autorise l’observation ou la transformation des paramètres du modèle. Nous introduisons la modélisation sinusoïdale, et l’analyse par transformée de Fourier qui permet de modéliser le son dans le domaine fréquentiel. Les propriétés spectrales du modèle sont explicitées, et mises en relation avec les paramètres perceptifs. Une synthèse sonore donne un son réel à partir de son modèle. Nous nous attardons sur la synthèse additive, à la base des modèles spectraux, et nous présentons les méthodes rapides pour cette synthèse.

Nous présentons des notions de musique instrumentale dans le chapitre 3, dont la notation musicale et sa représentation symbolique dans un fichier informatique. Après un bref historique de l’interprétation instrumentale, nous listons les principaux paramètres musicaux qui déterminent la performance instrumentale.

Chapitre 1

Onde sonore et perception

Une source sonore en vibration émet un son dont les propriétés physiques sont traitées chez les humains par un mécanisme psychoacoustique complexe. La perception du son dépend donc de ce filtre. Le mot *son* est ambigu : il désigne aussi bien la vibration physique, en acoustique, que la sensation que cette vibration procure, en psychoacoustique. Nous choisissons donc d'utiliser dans ce document le terme d'*onde sonore* pour la vibration physique, le mot *son* nous servant à décrire la sensation auditive provoquée par la perception d'une onde sonore.

Dans ce chapitre, nous présentons d'abord, dans la section 1.1, les principes physiques de la production et la propagation d'une onde sonore. La section 1.2 explique ensuite comment cette onde est perçue par un humain, avec des considérations physiologiques et psychoacoustiques. Les paramètres sonores et perceptifs qui permettent de décrire un son sont alors listés.

1.1 Éléments physiques

L'onde sonore est produite par une source sonore en vibration, *i.e.* un corps qui oscille autour d'une position d'équilibre. Il y a par exemple deux grandes familles de sources sonores instrumentales : les sources à excitation entretenue, comme la voix ou le violon, et celles à impulsion initiale, dans le cas des percussions par exemple (cloche, gong, *etc.*).

Une source sonore produit une perturbation de la pression dans son milieu environnant, ce qui donne naissance à une onde de pression. La propagation de l'onde est sphérique si la source est fixe, ponctuelle, omnidirectionnelle et si le milieu environnant est homogène. En pratique elle est le plus souvent directionnelle. L'ébranlement de la matière se caractérise par une variation de pression se propageant de proche en proche, la transmission de l'énergie se faisant sans déplacement de particule. Les variations de pression se propagent dans l'air à la *vitesse du son* (environ $343 \text{ m}\cdot\text{s}^{-1}$ pour une pression équivalente à une atmosphère (1013 hPa) et une température de 20 degrés Celsius). En fonction de son environnement, et pour chaque obstacle rencontré (par exemple les murs d'une pièce), l'onde sonore change de direction – c'est la propriété de réflexion –, et perd éventuellement de l'intensité, en fonction de la propriété d'absorption du matériau de l'obstacle rencontré. Des phénomènes de diffraction de l'onde peuvent aussi subvenir.

La nature des vibrations de l'onde sonore peut être périodique ou non-périodique, ou

encore une combinaison de ces deux natures. Les variations périodiques suivent un motif particulier qui se répète dans le temps, ce qui donne une sensation de hauteur. Ce n'est en général pas le cas des variations non-périodiques. Les sons naturels sont presque toujours semi-périodiques : ils sont une combinaison de variations périodiques et non-périodiques, générant une certaine forme d'onde. La durée du motif récurrent de la forme d'onde est sa période, et le nombre de répétitions de cette période en une seconde nous donne sa fréquence. Si nous prenons l'exemple de la production d'une onde sonore sinusoïdale (onde dite *pure*), les caractéristiques de fréquence (et donc la période), d'amplitude et de phase de la sinusoïde sont des paramètres physiques déterminants pour la perception de cette onde. C'est aussi le cas des propriétés physiques des ondes complexes.

1.2 Perception auditive

La perception auditive est la sensation provoquée par la réception d'une onde sonore. Elle est bien sûr dépendante du fonctionnement de l'oreille, et des mécanismes psychoacoustiques complexes qui interprètent les signaux reçus par celle-ci. Quatre paramètres principaux permettent de décrire la perception d'un son : sa durée, son intensité, sa hauteur et son timbre. Les unités de mesure et les échelles de valeurs de chaque paramètre peuvent différer, en fonction du degré de considération des phénomènes perceptifs.

1.2.1 L'oreille

L'appareil auditif comporte trois parties principales, comme l'illustre la figure 1 : l'oreille externe, l'oreille moyenne et l'oreille interne. L'oreille externe est le point de départ du mécanisme physiologique de l'audition. Elle est principalement constituée du *pavillon*, la grande partie qui sert à capter et à concentrer les ondes sonores. Du pavillon, le son suit le *canal auditif externe*, un tube qui conduit à l'oreille moyenne. L'oreille moyenne comprend le *tympan* et les *osselets*, quatre très petits os (le marteau, l'enclume, l'os lenticulaire et l'étrier). Le son est le résultat de vibrations de l'air dans le conduit auditif, ce qui fait vibrer le tympan. Les vibrations du tympan sont ensuite transmises aux osselets, puis à l'oreille interne *via* la *fenêtre ovale*. La chaîne des osselets est située dans la *caisse du tympan*. Celle-ci est une petite cavité communiquant avec l'extérieur par la *trompe d'Eustache*, qui assure l'équilibre des pressions des deux côtés du tympan.

L'oreille interne contient le *vestibule*, organe de l'équilibre responsable de la perception de la position angulaire de la tête et de son accélération, et l'organe de l'ouïe, la *cochlée*. L'appareil vestibulaire est constitué de trois canaux semi-circulaires, disposés orthogonalement dans les trois plans. Ils sont remplis d'un liquide, et lorsque l'oreille est soumise à un mouvement, l'inertie de ce liquide rend le mouvement détectable par des *cellules ciliées* (cellules sensorielles coiffées de structures filamenteuses). La disposition des trois canaux en trois plans orthogonaux permet de détecter la position angulaire de la tête dans toutes les directions possibles. La cochlée est un organe creux rempli également d'un liquide appelé *endolymphe*. Elle est tapissée de cellules ciliées, coiffées de *stéréocils* groupés en une *touffe ciliaire* libre de vibrer. Ces cellules sont disposées le long de la *membrane basilaire* qui vient partitionner la cochlée en deux chambres. L'ensemble des cellules ciliées et des membranes associées constituent l'*organe de Corti*. La membrane basilaire et les cellules ciliées qu'elle porte sont mises en mouvement par les vibrations transmises au travers de

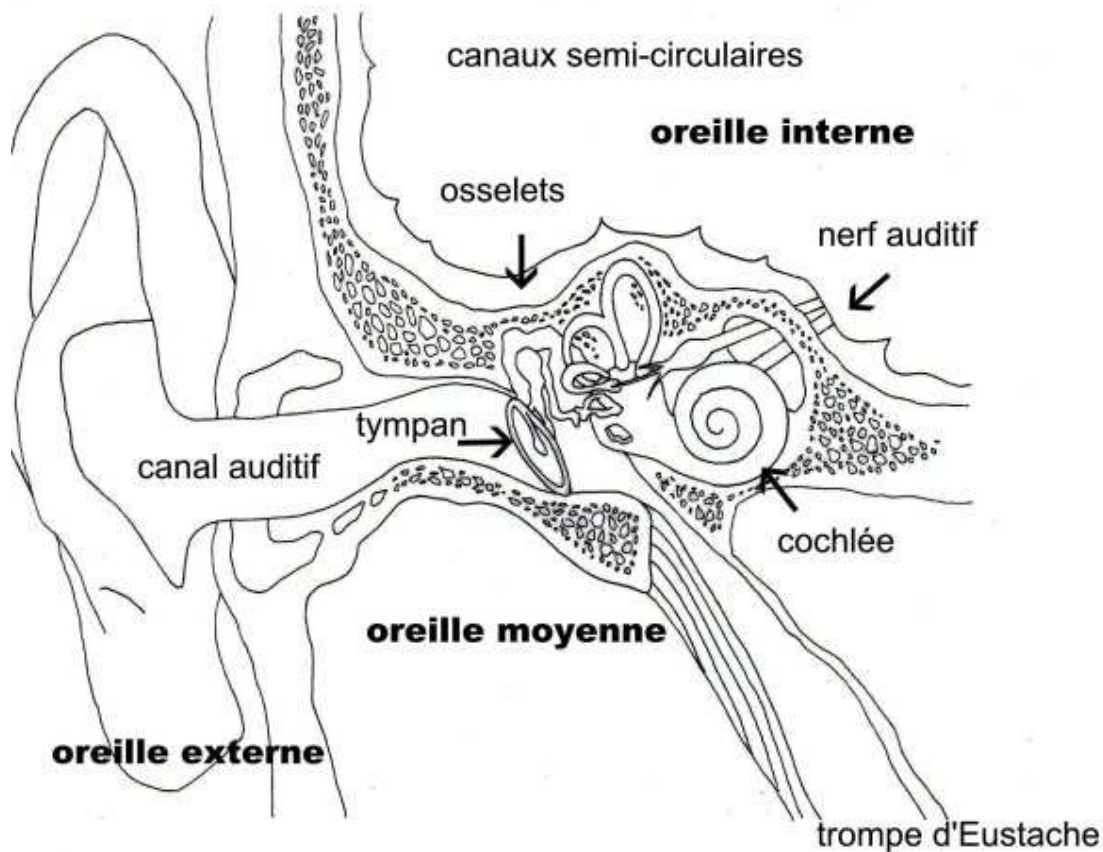


FIG. 1 – Schéma de l'oreille (image du Centre d'Information sur la Surdit  d'Aquitaine).

l'oreille moyenne. B k sy [vB60] a montr  que plus la fr quence d'un signal per u est  lev e, plus la position du maximum de d formation de la membrane basilaire est  loign e de la base de la cochl e, comme l'illustre la figure 2. L'oreille op re une d composition spectrale du son, *i.e.* chaque cellule r pond pr f rentiellement   une certaine fr quence pour permettre au cerveau de diff rencier la hauteur des sons. Les cellules cili es les plus proches de la base de la cochl e r pondent pr f rentiellement aux aigus, et celles situ es en son *apex* (dernier tour de la cochl e) r pondent aux basses fr quences. Ce sont les cellules cili es qui font la transduction m cano lectrique : elles transforment un mouvement de leur touffe ciliaire en signal nerveux. Conduit par le nerf auditif, celui-ci est interpr t  par le cerveau comme un son de la hauteur tonale correspondant   la cellule excit e.

1.2.2 Psychoacoustique

La psychoacoustique a pour objet l' tude exp rimentale des relations quantitatives entre les ondes acoustiques mesurables physiquement et les r ponses de l'ensemble du syst me auditif (sensations et perceptions auditives). L'une des premi res observations de la psychoacoustique est qu'il n'y a pas de relation bijective entre les param tres physiques

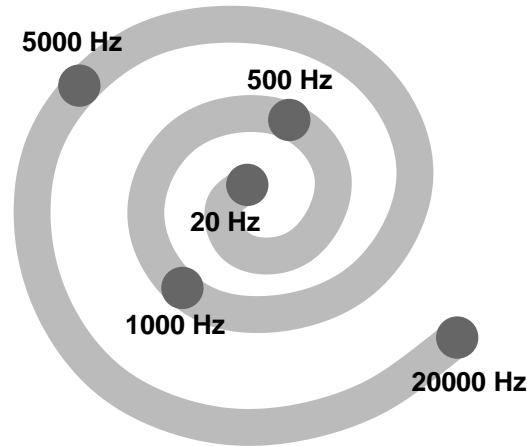


FIG. 2 – La cochlée. Il y a une correspondance entre la position des récepteurs cochléaires (la base du limaçon est le point le plus éloigné de la fenêtre de contact avec l'oreille moyenne) et la fréquence des vibrations sonores perçues. Les récepteurs cochléaires (ou mécanorécepteurs) sont des cellules ciliées de la membrane basilaire sensibles à de très légères inclinaisons de leur cils.

des ondes sonores et les sensations qu'elles produisent. Par exemple, si une augmentation de la fréquence d'une vibration sinusoïdale entraîne une augmentation de la hauteur perçue, elle peut aussi donner lieu à une variation de l'intensité perçue.

Comme nous l'avons vu avec le fonctionnement de l'oreille décrit dans la section précédente, nous entendons un son lorsque des vibrations de l'air ambiant atteignent notre tympan et le mettent en mouvement, dans des conditions d'amplitude et de fréquence telles que cette stimulation mécanique y provoque un phénomène bio-électrique. Le traitement de l'information contenue dans ce phénomène se poursuit à travers différents relais jusqu'au cortex cérébral, le résultat étant la perception du son. L'analyse de la chaîne fonctionnelle qui va de la vibration du tympan à la perception a notamment été conduite par Békésy [vB60], dont les travaux sont à l'origine de l'expansion que connaît la physiologie de l'audition, et donc la psychoacoustique.

1.2.3 Paramètres sonores

Nous avons expliqué dans la section précédente que les paramètres perceptifs du son évoluent différemment des propriétés physiques de l'onde sonore. La loi de Fechner par exemple, qui s'applique à tout organe sensoriel, indique que la sensation est proportionnelle au logarithme de l'excitation : nous percevons les paramètres sonores sur des échelles logarithmiques.

Nous pouvons décrire le son par quatre caractéristiques perceptives principales, en fonction de propriétés physiques telles que la fréquence ou l'amplitude d'une onde sonore : la durée, la hauteur, le volume et le timbre. Le paramètre de durée étant très proche pour l'onde sonore et pour le son perçu, nous nous concentrons sur les autres paramètres.

Fréquence et hauteur

La cochlée décompose l'onde sonore en composantes sinusoïdales élémentaires, de fréquences distinctes. Si une onde sonore ne contient qu'une seule fréquence, c'est une onde sinusoïdale, ou une onde sonore *pure*. De tels sons sont rarement naturels, et les ondes sonores contiennent en général de nombreuses fréquences. La hauteur est un paramètre perceptif relatif au paramètre physique de fréquence, et plus précisément à la fréquence fondamentale. Cette dernière peut être la fréquence la plus basse perçue par la cochlée, mais aussi un effet de la combinaison des autres fréquences, alors qu'elle n'est pas présente dans le signal (cas de la fondamentale absente). Les fréquences et les hauteurs d'un son peuvent être données en Hertz. Lorsque le nombre de périodes par seconde d'une onde sonore décroît, la sensation de hauteur du son décroît également.

Intensité et volume

L'intensité d'un son est la caractéristique nous permettant de distinguer un son fort d'un son faible. C'est un paramètre physique proportionnel au carré de l'amplitude de l'onde sonore, qui a pour unité le Watt par mètre carré ($\text{W}\cdot\text{m}^{-2}$). L'oreille perçoit l'intensité sur une échelle logarithmique, pour donner des valeurs de volume. L'unité Bel est alors utilisée pour la mesurer, le décibel (dB) étant la dixième partie du bel. Le niveau de référence pour définir le décibel acoustique est le seuil d'audibilité d'une fréquence de 1000 Hz, *i.e.* la valeur d'intensité nécessaire à une onde sonore pure de fréquence 1000 Hz pour être perçue. L'intensité acoustique de référence est alors $I_0 = 10^{-12}\text{W}\cdot\text{m}^{-2}$. Cette intensité de référence correspond à la variation d'une pression de référence $P_0 = 2\cdot 10^{-5}\text{Pa}$. Les valeurs de référence I_0 et P_0 sont utilisées pour définir l'échelle des décibels SPL (*Sound Pressure Level*). La formule qui permet d'obtenir la valeur du volume en décibels à partir des valeurs d'intensité I ou de pression P est alors :

$$\text{dB}_{\text{SPL}} = 10 \log_{10} \left(\frac{I}{I_0} \right) = 20 \log_{10} \left(\frac{P}{P_0} \right) \quad (1)$$

Plus la variation de pression est grande, et plus le volume sonore est important. Quand l'intensité d'une onde sonore est multipliée par 10, sa valeur en décibels est augmentée de 10 dB. L'échelle des décibels est ainsi une échelle de comparaison. Cela donne une importance relative à la valeur de référence utilisée.

Le volume d'un son peut aussi être exprimé en fonction de l'amplitude du signal, en utilisant une amplitude de référence $A_0 = 10^{-6}$. Les volumes audibles sont dans ce cas approximativement compris entre 0 et 120 dB. La valeur en décibels du volume v en fonction de l'amplitude a du signal, et l'inverse, sont données par les équations suivantes :

$$v = 20 \log_{10} \left(\frac{a}{A_0} \right) \quad (2)$$

$$a = A_0 10^{v/(20\text{dB})} \quad (3)$$

où A_0 est l'amplitude de référence pour 0 dB.

Timbre

Le timbre d'un son est un paramètre complexe, défini par défaut : c'est ce qui différencie deux sons de même durée, de même hauteur et de même volume. Le timbre des instruments

de musique est par exemple la couleur différente donnée au son par une clarinette et une trompette qui jouent la même note, avec un volume et une durée comparables.

Le timbre est notamment déterminé par la répartition en fréquence, la combinaison et les interactions des différentes composantes fréquentielles présentes dans l'onde sonore. La phase d'attaque d'un son, composée de *transitoires d'attaque*, qui correspond au début de la production d'une onde sonore par une source, et qui confère au son un caractère d'explosivité ou non, participe également de façon importante à son timbre.

1.2.4 Propriétés perceptives

Si les grandeurs physiques d'une onde sonore sont simples à analyser, le processus psychoacoustique d'interprétation humaine de ces valeurs est beaucoup plus complexe. Quand le volume d'un son augmente par exemple, la sensation de hauteur peut aussi varier. La hauteur semble augmenter ou baisser pour respectivement les hautes ou les basses fréquences de l'onde sonore. Pour les signaux sonores très courts, la hauteur perçue est plus basse que dans le cas d'un son long, avec un son de même fréquence et de même timbre. Ces interdépendances entre les paramètres sonores sont souvent liées à des propriétés perceptives de l'oreille. La considération de ces propriétés amène à considérer des paramètres perceptifs dont les valeurs peuvent être très différentes de celles des paramètres sonores.

Seuils de l'audition

Le *seuil de sensibilité* de l'oreille varie approximativement entre 20 Hz et 16000 Hz. Cet intervalle se réduit inexorablement avec l'âge. Du fait de sa géométrie et de la nature de ses parois, l'oreille externe ne transmet pas également toutes les fréquences. L'onde sonore est décomposée fréquemment en composantes sinusoïdales par la cochlée. Les composantes dont les amplitudes sont inférieures au *seuil d'audition* (voir figure 3) ne sont pas audibles. Le seuil d'audition \mathcal{S}_a dépend de la fréquence f en Hertz des composantes, et peut s'approcher ainsi d'après [ZF90] :

$$\mathcal{S}_a(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4 \quad (4)$$

Le maximum de sensibilité auditive concerne la principale bande de fréquence utilisée par la voix. Une sinusoïde de fréquence 3000 Hz est perçue plus forte qu'une sinusoïde de même amplitude mais de fréquence 5000 Hz ou 500 Hz.

D'autres seuils sont également importants pour l'audition. Le *seuil de douleur* correspond à un volume de 120 dB pour un son pur à 1000 Hz. Au-delà de ce seuil, l'oreille peut subir des dommages physiques irréversibles. Le *temps d'intégration* de l'oreille varie de 50 à 100 ms suivant l'intensité. Le *seuil temporel de reconnaissance de la hauteur* est défini à 1/100e de seconde en moyenne. Un son plus bref est perçu sans hauteur précise et il est alors qualifié de *claquement* par les acousticiens.

Bandes critiques

Les humains peuvent percevoir dans le signal sonore des fréquences allant approximativement de 20 Hz à 16000 Hz, et identifier plus de 600 hauteurs différentes. Il n'est pas possible de différencier deux fréquences quand elles sont trop proches. Le seuil de différenciation de deux fréquences est noté Δ_f . Lorsque deux fréquences sont éloignées, le système

auditif arrive à les différencier, mais si elles sont dans le même intervalle Δ_f , l'oreille peut avoir un comportement non linéaire (les deux sons ne sont pas perçus séparément mais combinés). L'intensité perçue de la somme des deux fréquences est alors moins importante que la somme des deux intensités, alors que si les fréquences sont éloignées, les sensations d'intensité s'ajoutent.

Les intervalles Δ_f sont liés aux propriétés physiques de la membrane basilaire et forment des bandes critiques de fréquence. 24 bandes suffisent à couvrir le spectre audible, elles sont données dans le tableau 1. Les largeurs de bande Δ_f peuvent être exprimées en fonction de la fréquence centrale f en Hertz :

$$\Delta_f = 25 + 75 \left(1 + 1.4 \left(\frac{f}{1000} \right)^2 \right)^{0.69} \quad (5)$$

Plusieurs modélisations de ces bandes critiques ont été proposées, dont une est particulièrement utilisée : l'échelle Bark, d'après Barkhausen. Cette échelle est adaptée pour la représentation des fréquences, car elle est proche de la perception. Sur cette échelle, les bandes critiques sont ainsi de largeur constante. Il y a des formules qui permettent de passer de l'échelle linéaire en Hertz à l'échelle Bark, et inversement. Si f est une valeur de fréquence en Hertz, et v son équivalent en fréquence Bark, f et v satisfont alors les relations :

$$v = \begin{cases} f/100 & \text{si } f \leq 500 \\ 9 + 4 \log_2(f/1000) & \text{si } f > 500 \end{cases} \quad (6)$$

$$f = \begin{cases} 100 v & \text{si } v \leq 5 \\ 1000 \cdot 2^{(v-9)/4} & \text{si } v > 5 \end{cases} \quad (7)$$

Une autre échelle proche de la perception est l'échelle ERB (*Equivalent Rectangular Bandwidth*). Le passage d'une fréquence f en Hertz vers une fréquence v en ERB, et l'inverse, sont donnés par les équations suivantes d'après [MG83] :

$$v = 11.17268 \log \left(1 + \frac{46.06538 f}{f + 14678.49} \right) \quad (8)$$

$$f = \frac{676170.4}{47.06538 - e^{0.08950404v}} - 14678.49 \quad (9)$$

1.2.5 Masquages

Un masquage se produit quand une composante fréquentielle du son n'est pas audible en raison de la présence d'une autre composante fréquentielle. Par exemple, si l'amplitude d'une composante fréquentielle est sous le seuil d'audition, elle est inaudible. De la même manière, une sinusoïde de fréquence f_M et d'amplitude a_M peut masquer une autre sinusoïde de fréquence f_m et d'amplitude a_m . L'effet de masquage est maximal quand f_M et f_m sont proches, et si $a_M > a_m$. Le seuil de masquage généré par une sinusoïde est proche d'un triangle avec des échelles proches de la perception (fréquences en Bark et amplitudes en dB par exemple). Un exemple de ce type de masquage est donné par la figure 3. Il y a aussi des effets de masquage temporel. Le post-masquage est l'effet masquant persistant d'un son disparu. Même quand le son n'est plus présent, son masque reste effectif quelques millisecondes, avant de décroître. Il existe aussi un pré-masquage, assez limité : l'effet masquant d'un son agit quelques millisecondes avant l'arrivée du son lui-même.

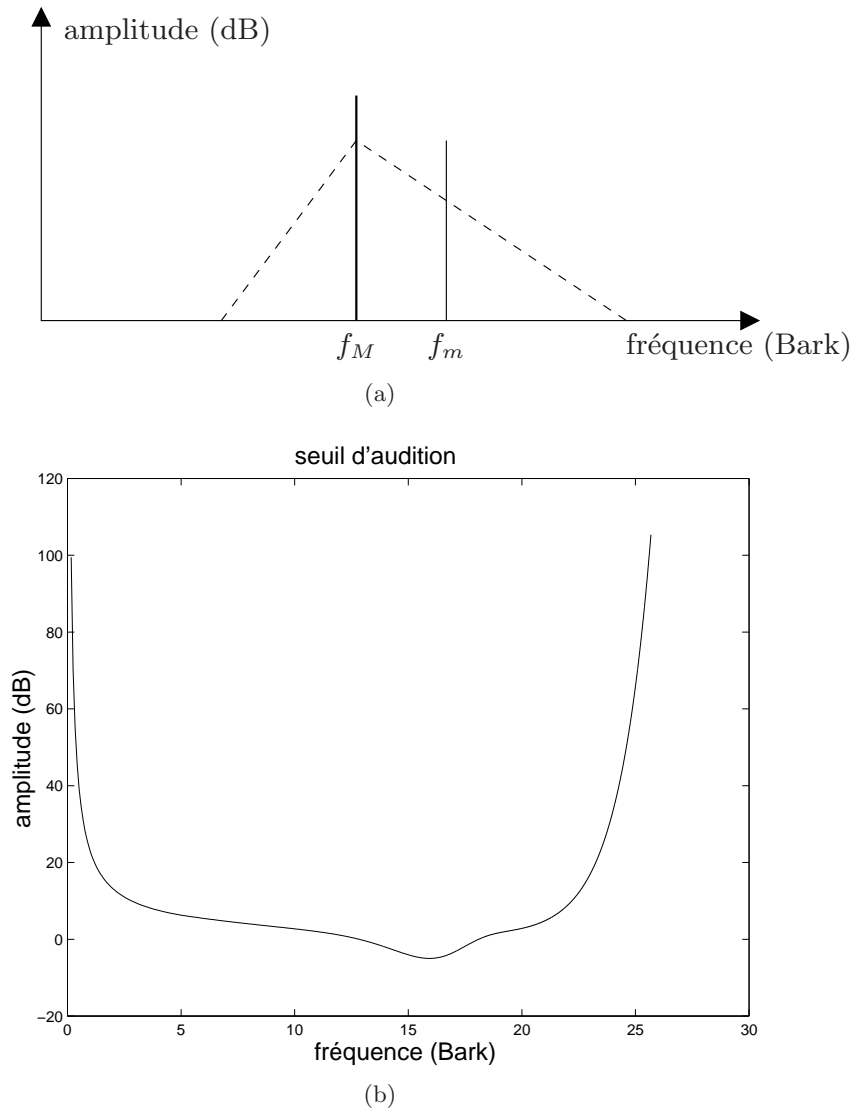


FIG. 3 – Masquages. (a) Masquage fréquentiel, cas où M et m sont deux sinusoïdes de fréquences respectives f_M et f_m , et d'amplitudes respectives a_M et a_m telles que $a_M > a_m$. Si f_m est proche de f_M , le son m peut être masqué par le son M (si le pic représentant m se trouve sous la courbe de masquage triangulaire due à M). Le triangle de masquage a une pente gauche de 27 dB par Bark , et une pente droite de -15 dB par Bark . Le sommet du triangle se situe à 10 dB sous le pic du partiel correspondant à M . (b) Seuil d'audition : la courbe représente le seuil de perception de l'amplitude pour une oreille humaine en parfait état. Pour chaque fréquence, le seuil de perception est différent : les fréquences les mieux perçues (la courbe avoisine le 0 dB) se situent dans la gamme moyenne entre 1000 et 3000 Hz . C'est aussi dans cette gamme que la dynamique de sensation est la plus grande. Si le pic représentant une sinusoïde se trouve sous le seuil d'audition, la sinusoïde est inaudible.

n°	f_{inf}	f_c	f_{sup}	Δ_f	n°	f_{inf}	f_c	f_{sup}	Δ_f
1	20	50	100	80	13	1720	1850	2000	280
2	100	150	200	100	14	2000	2150	2320	320
3	200	250	300	100	15	2320	2500	2700	380
4	300	350	400	100	16	2700	2900	3150	450
5	400	450	510	110	17	3150	3400	3700	550
6	510	570	630	120	18	3700	4000	4400	700
7	630	700	770	140	19	4400	4800	5300	900
8	770	840	920	150	20	5300	5800	6400	1100
9	920	1000	1080	160	21	6400	7000	7700	1600
10	1080	1170	1270	190	22	7700	8500	9500	1800
11	1270	1370	1480	210	23	9500	10500	12000	2500
12	1480	1600	1720	240	24	12000	13500	15500	3500

TAB. 1 – 24 bandes critiques de fréquence en Hertz. La fréquence limite inférieure f_{inf} , la fréquence limite supérieure f_{sup} , la fréquence centrale f_c et la largeur de la bande Δ_f sont indiquées pour chaque bande.

1.2.6 Paramètres perceptifs

La prise en compte des propriétés perceptives de seuil ou de masquage sur les paramètres de durée, de hauteur et de volume d'un son donne les paramètres psychoacoustiques respectivement appelés chronie, tonie, et sonie du son. Le timbre est un paramètre psychoacoustique également très étudié. De nombreux descripteurs du timbre d'un son peuvent être énumérés, tels la brillance, la couleur ou la rugosité.

Prenons l'exemple de la sonie, associée au volume d'un son. À volume égal, les sons à basse ou haute fréquence ont une sonie inférieure aux sons à fréquence moyenne, en fonction du seuil d'audition. Ainsi, pour paraître aussi intense qu'un son de 1 000 Hz à 43 dB, un son de 100 Hz ou un son de 10 000 Hz doivent avoir un niveau de 63 dB. Ces trois sons ont alors un niveau d'isotonie de 43 phones. Le phone est l'unité qui sert à exprimer le niveau d'isotonie. Les courbes isotoniques furent mesurées pour la première fois en 1933 par Fletcher et Munsen. Robinson et Dadson effectuèrent en 1956 de nouvelles mesures, considérées comme plus précises. Celles-ci furent la base du standard ISO226, qui fut utilisé jusqu'en 2003. En raison des divergences entre ces mesures, l'organisation ISO (*International Organization for Standardization*) a actualisé la norme en 2003. La figure 4 donne des exemples de courbes isotoniques qui suivent la nouvelle norme ISO226-2003.

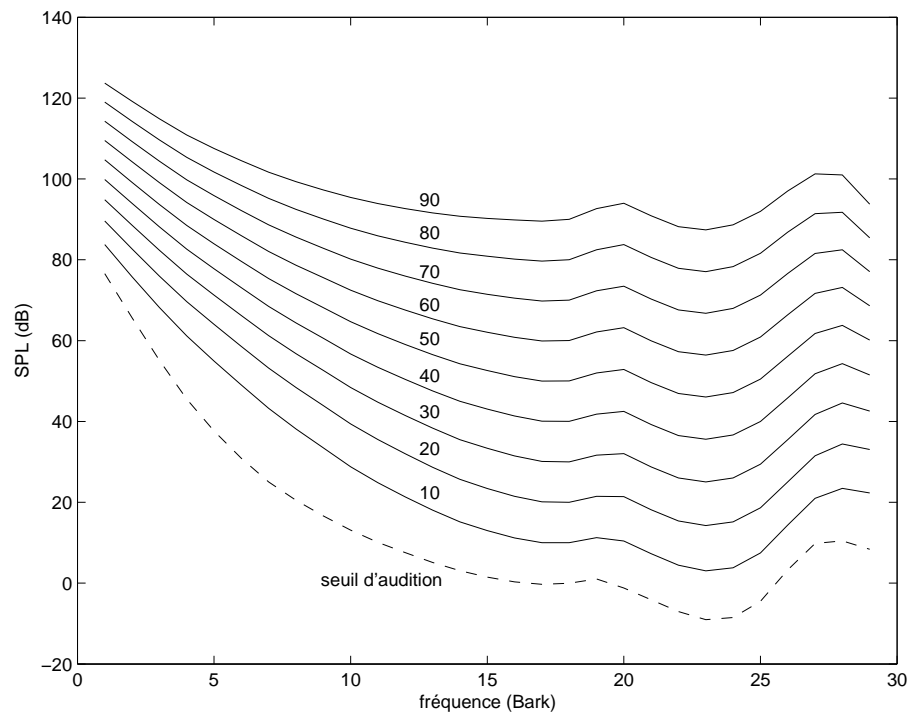


FIG. 4 – Courbes isophoniques : représentation de la sensibilité de l'oreille humaine aux différentes fréquences et à différentes intensités. La courbe inférieure en ligne pleine est celle de valeur 10 phons, les courbes ont ensuite un espacement de 10 phons jusqu'à la courbe de 90 phons.

Chapitre 2

Modèles de sons

Ce chapitre présente les différentes étapes de la modélisation d'un son en informatique. Pour être traité par des moyens informatiques, le signal sonore analogique doit d'abord être numérisé. Une phase d'analyse permet ensuite de représenter le son sous la forme d'un modèle mathématique, avec des paramètres. La modélisation facilite les opérations de transformation sur le son, les paramètres du modèle sont modifiables. Une synthèse sonore permet finalement de redonner un son réel à partir de sa modélisation. L'ensemble de ces opérations représente la chaîne de traitement du son, elle est symbolisée par la figure 5. La différence entre le son réel et le son modélisé s'appelle l'erreur de modélisation. Dans de nombreux cas, il faut faire un compromis entre la concision du modèle – le nombre de paramètres nécessaires pour expliquer l'observation – et l'amplitude de l'erreur de modélisation. Le choix du modèle se fait en fonction des transformations qu'il autorise sur ses paramètres.

Il y a deux familles de modèles sonores : les modèles physiques et les modèles de signaux. Les modèles physiques consistent à mettre en équations la source sonore et calculer une solution. Les modèles de signaux ne s'intéressent qu'à la mesure du signal acoustique, sans réellement se préoccuper de la nature de la production. Ce chapitre se concentre sur la famille des modèles de signaux, et plus spécialement sur le modèle sinusoïdal. Le modèle sinusoïdal paramètre le son au niveau de l'oreille interne, et il est ainsi proche de la perception.

Nous commençons par décrire, dans la section 2.1, comment un signal sonore naturel et continu peut être représenté numériquement dans le domaine temporel. La méthode

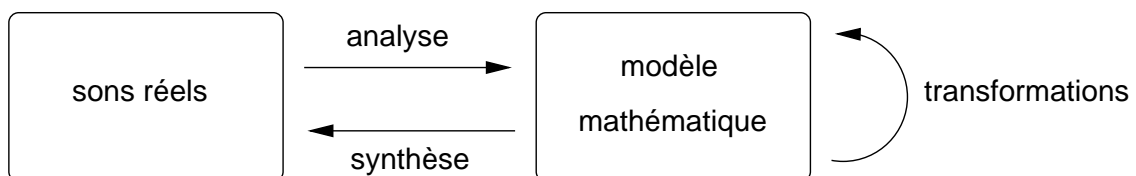


FIG. 5 – Chaîne de traitement du son. Le son est analysé pour donner un modèle mathématique, qui peut être transformé. Une synthèse permet de repasser dans le domaine physique pour donner un son réel.

d'analyse sinusoïdale présentée dans la section 2.2 permet de passer dans un domaine fréquentiel, pour modéliser le son. La section 2.3 décrit les paramètres et les représentations d'un son dans ce modèle. Les transformations sonores dans le modèle ne sont pas l'objet de ce document, de nombreux exemples peuvent être trouvés dans les travaux de Verfaillie [Ver03]. Nous présentons des techniques de synthèse additive dans la section 2.4, qui permettent de revenir dans le domaine temporel.

2.1 Domaine temporel

Un signal audio peut être représenté en temps continu par les variations d'amplitude de l'onde sonore en fonction du temps, c'est une représentation temporelle. Le signal est dans ce cas modélisé comme une fonction continue de son amplitude en fonction du temps mesuré en secondes. Il peut donc être noté $s(t)$, où $t \in \mathbb{R}$. L'opération de numérisation se réalise en deux étapes : l'échantillonnage et la quantification.

2.1.1 Échantillonnage

L'échantillonnage consiste à passer d'un signal à temps continu (un signal électrique, un signal acoustique, ...) en une suite discrète de valeurs. La conversion d'un signal analogique en un signal numérique se fait grâce à un convertisseur analogique-numérique ou ADC (*Analog-to-Digital Converter*). Ce convertisseur peut mesurer des milliers de fois par seconde la valeur du signal et convertir cette valeur en un nombre binaire. Les valeurs mesurées à intervalles réguliers constituent alors les échantillons du son numérique. La période d'échantillonnage est la période de temps séparant deux échantillons successifs. La fréquence d'échantillonnage F_e s'exprime en Hertz, et correspond à l'inverse de la période d'échantillonnage. Elle détermine la précision temporelle de la conversion. Une illustration de l'opération d'échantillonnage est donnée par la figure 6.

Le théorème de Shannon-Nyquist, théorème fondamental de l'audio numérique, indique que pour être représenté correctement, un signal de fréquence maximale f_{\max} doit être échantillonné avec une fréquence au moins deux fois supérieure : $F_e > 2f_{\max}$. Puisque la plage fréquentielle de l'oreille humaine va approximativement de 20 Hz à 16000 Hz, un système de conversion doit prélever des échantillons à un taux au moins égal à 32000 fois par seconde pour éviter les dégradations audibles. La fréquence d'échantillonnage du standard disque compact (CD) est ainsi égale à 44100 Hz.

L'échantillonnage correspond à la périodisation de la plage de fréquence allant de 0 Hz jusqu'à la fréquence de Nyquist (demi-fréquence d'échantillonnage) : toute fréquence supérieure va se retrouver dans cette plage de fréquence, c'est le repliement haute-fréquence (des composantes de haute fréquence se trouvent alors par erreur représentées dans les basses fréquences). Les fréquences négatives vont également revenir dans cette plage, c'est le repliement basse-fréquence. Cela ajoute au signal sonore numérisé des composantes fréquentielles qui ne sont pas présentes dans le signal analogique.

La reconstruction analogique du signal numérique est possible si les variations de celui-ci sont assez lentes, ou réciproquement si la période d'échantillonnage est assez fine. Elle utilise un convertisseur numérique-analogique (DAC pour *Digital-to-Analog Converter*).

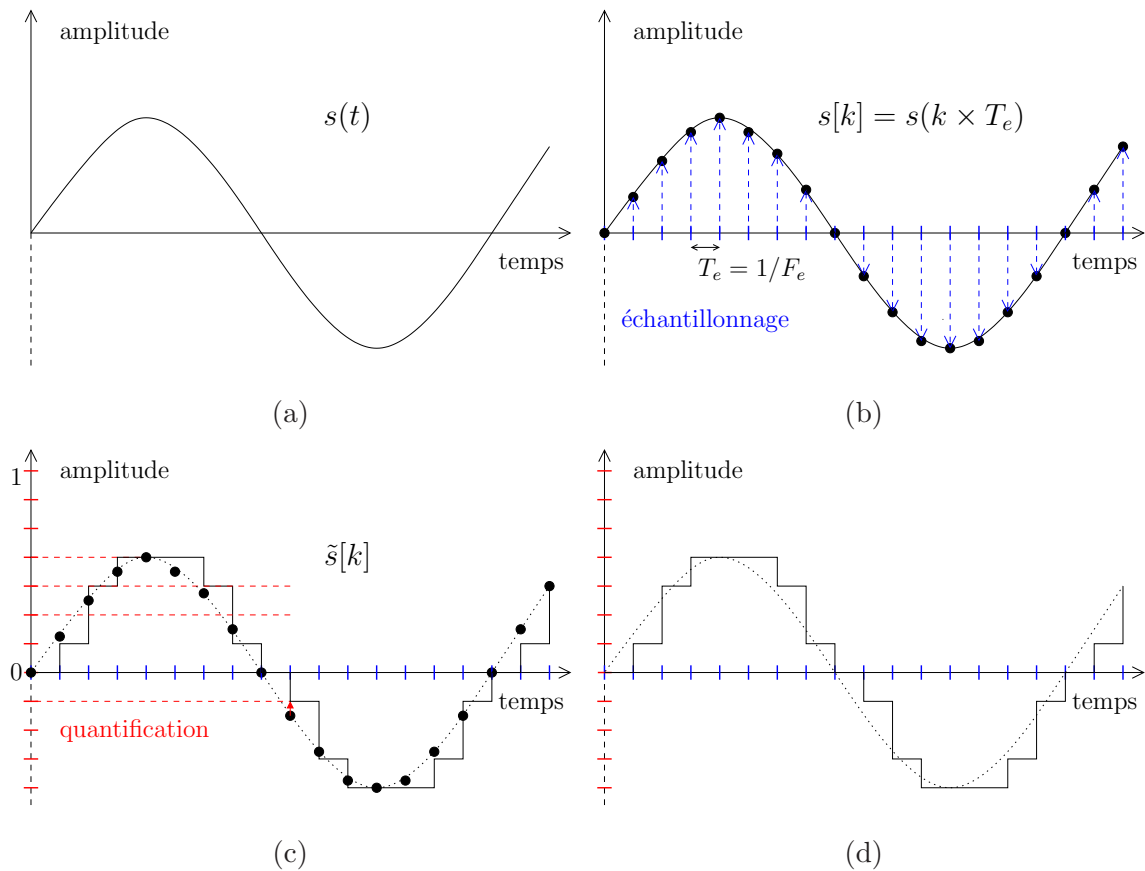


FIG. 6 – Numérisation du son. (a) Représentation temporelle continue du son. (b) Échantillonnage avec une fréquence d'échantillonnage F_e , inverse de la période T_e . (c) Quantification sur i bits, avec 2^i valeurs possibles. (d) Représentation numérique du son par paliers.

2.1.2 Quantification

La seconde étape d'une conversion analogique-numérique est la quantification, illustrée par la figure 6. La représentation numérique d'un échantillon utilise un nombre fini de bits. Le nombre de bits utilisés détermine la précision en amplitude, la dynamique de la conversion. Pour que le signal soit fidèlement reproduit, il doit non seulement être échantillonné avec une fréquence suffisante, mais aussi quantifié avec un nombre suffisant de bits, pour que l'amplitude de chaque valeur soit la plus précise possible. Si les valeurs sont quantifiées dans un ambitus trop faible, elles diffèrent significativement des valeurs du signal analogique, et cela ajoute un bruit de quantification au signal. La plupart des systèmes numériques fonctionnent avec des quantifications sur 16 bits, ce qui donne une plage dynamique d'environ 96 dB (l'oreille est sensible à 50 dB de dynamique dans les graves, et jusqu'à 120 dB de dynamique aux alentours de 3 kHz).

2.2 Analyse sinusoïdale

Le théorème de Fourier indique que toute fonction périodique peut être modélisée sous la forme d'une somme de sinusoïdes d'amplitudes données et de fréquences en relation harmonique. Pour des sons quasi-stationnaires, ces amplitudes et fréquences évoluent lentement dans le temps, contrôlant un ensemble d'oscillateurs pseudo-sinusoïdaux appelés partiels du son. C'est la représentation de McAulay-Quatieri [MQ86] pour les signaux de parole, également utilisée par Serra et Smith [SS90] pour des signaux musicaux. Dans cette représentation, un signal audio s peut donc s'écrire :

$$s(t) = \sum_{i=1}^N a_i(t) \sin(\phi_i(t)) \quad (10)$$

où N est le nombre de partiels dans le son, et :

$$\phi_i(t) = \phi_i(0) + 2\pi \int_0^t f_i(u) du \quad (11)$$

Les fonctions $f_i(t)$, $a_i(t)$ et $\phi_i(t)$ sont respectivement la fréquence, l'amplitude et la phase du i -ième partiel du son.

Cette décomposition peut s'effectuer par transformée de Fourier. La transformée de Fourier est une opération mathématique qui, appliquée au signal sonore, donne les paramètres des sinusoïdes qui le composent. Elle permet donc de passer de la représentation temporelle – amplitude en fonction du temps – à la représentation spectrale – amplitude en fonction de la fréquence. Le signal est alors représenté par fenêtre d'analyse temporelle dans le domaine fréquentiel, les amplitudes de ses composantes élémentaires sinusoïdales sont exprimées en fonction de leur fréquence. La transformée de Fourier donne le spectre complexe $S(f)$ d'un signal $s(t)$ continu par la relation :

$$S(f) = \int_{-\infty}^{+\infty} s(t) e^{-i2\pi ft} dt \quad (12)$$

Il y a une version discrète de la transformée de Fourier, applicable sur une partie d'un signal sonore discret $s[k] = s(k T_e)$, avec T_e la période d'échantillonnage du signal et K le nombre d'échantillons du son :

$$S[m] = \sum_{k=0}^{K-1} s[k] e^{-i\frac{2\pi}{K} km} \quad (0 \leq m < K) \quad (13)$$

Le spectre discret obtenu est échantillonné en fréquence par pas de (F_e/K) Hz, *i.e.* les valeurs de fréquence analysées sont toutes multiples de cette fréquence. Pour améliorer l'analyse, le signal sonore peut être multiplié par le signal d'une fenêtre de pondération w avant d'effectuer la transformée de Fourier. Le spectre s'exprime alors sous la forme :

$$S[m] = \frac{2}{\sum w[k]} \sum_{k=-\infty}^{+\infty} w[k] s[k] e^{-i\frac{2\pi}{K} km} \quad (0 \leq m < K) \quad (14)$$

L'équation (13) est équivalente à l'équation (14) si la fenêtre de pondération w considérée est la fenêtre rectangulaire de taille K :

$$w[k] = \begin{cases} 1 & \text{si } 0 \leq k < K \\ 0 & \text{sinon} \end{cases} \quad (15)$$

Une fenêtre de pondération permet de calculer la somme de Fourier avec un nombre fini de termes. Différents types de fenêtres peuvent être utilisés pour effectuer une analyse dans le modèle sinusoïdal, comme la fenêtre de Bartlett ou la fenêtre de Hann par exemple. Le type de fenêtre utilisé a une influence sur l'imprécision de l'analyse. La figure 10 montre deux types différents de fenêtres d'analyse, et un exemple de l'application de la transformée de Fourier avec fenêtre de pondération sur un signal sinusoïdal pur : plusieurs composantes sont extraites, à cause de l'imprécision de l'analyse. Diverses méthodes présentées dans Marchand [Mar00] ou Lagrange [Lag04] peuvent alors corriger cette erreur.

2.3 Domaine fréquentiel

L'analyse du signal sonore permet de passer du domaine temporel au domaine fréquentiel, appelé aussi domaine spectral. Cela reproduit le mécanisme de la cochlée expliqué dans la section 1.2. Nous avons maintenant un spectre pour chaque trame (vecteur) de valeurs du signal numérique. C'est la représentation spectrale du son.

2.3.1 Spectre

Si le signal analysé est un son pur, de fréquence et d'amplitude constantes sur une fenêtre d'analyse, il peut s'exprimer sous la forme :

$$s(t) = a \sin(2\pi ft + \phi) \quad (16)$$

où a est l'amplitude, f la fréquence et ϕ la phase du signal. Un tel signal est alors représenté par un trait dans le spectre, comme le montre la figure 7 : c'est un partiel. La figure 8 donne l'exemple d'un spectre obtenu par l'analyse d'un son plus complexe. Les maxima locaux (ou pics) dans le spectre sont les partiels.

2.3.2 Propriétés spectrales

L'analyse sonore dans le modèle sinusoïdal est proche de la perception, et plus précisément du fonctionnement de la cochlée. Les propriétés du spectre sont alors en relation avec les paramètres perceptifs du son, et notamment le timbre présenté dans la section 1.2.3.

La périodicité, phénomène temporel, est le principal phénomène physique en relation avec la perception de hauteur. D'un point de vue fréquentiel, et donc dans le spectre, la périodicité d'un son entraîne une répartition harmonique de ses partiels, *i.e.* les partiels sont placés dans le spectre à des fréquences multiples de la fréquence fondamentale. Fourier exprime au début du 19ème siècle que "tout mouvement périodique complexe se décompose en une somme de mouvements périodiques simples (sinusoïdes) appelés harmoniques, et dont les fréquences sont des multiples entiers de la fréquence la plus basse, appelée fondamentale".

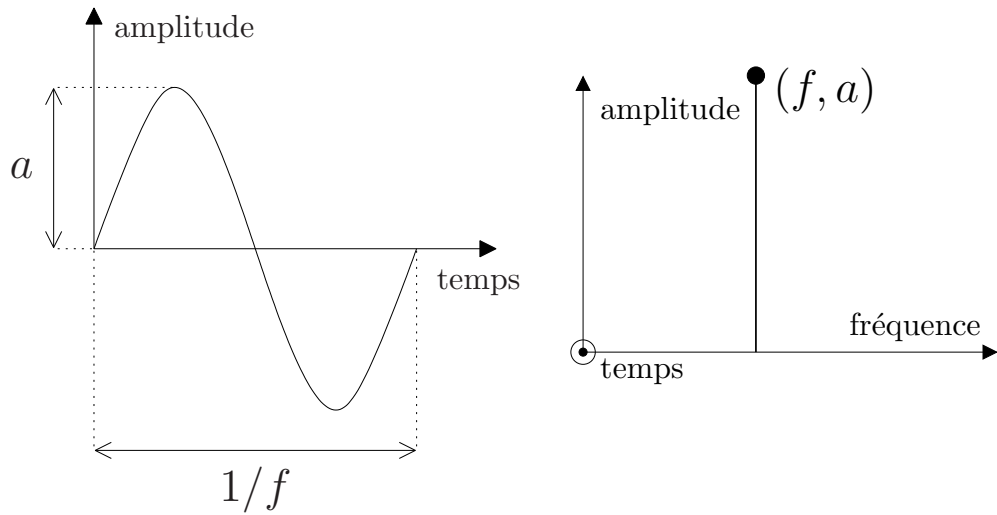


FIG. 7 – Représentation temporelle d'un signal sinusoïdal pur $s(t) = a \sin(2\pi ft)$ à gauche, et sa représentation spectrale à droite.

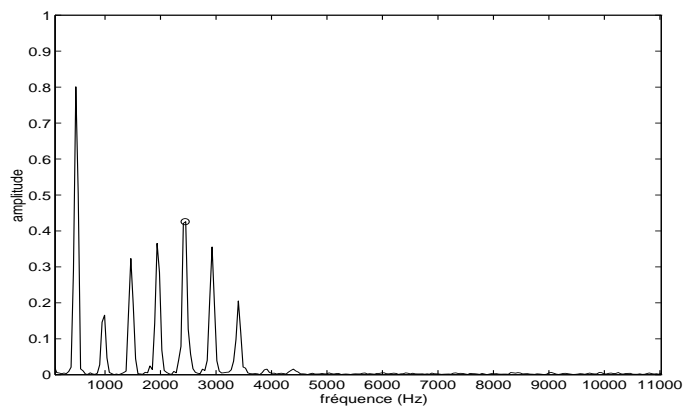


FIG. 8 – Spectre obtenu par transformée de Fourier. Les pics du spectre sont les partiels, situés en fréquence en abscisse et avec une amplitude indiquée en ordonnée.

Toute la différence entre deux sonorités périodiques de même hauteur est fonction du rang des harmoniques présents, ainsi que de leurs amplitudes. Si la fréquence fondamentale n'est pas la fréquence du premier partiel harmonique, nous sommes dans le cas de la fondamentale absente. Cela peut provenir d'une sonorité nasillarde, comme celle d'un hautbois. Si la fréquence fondamentale d'un son n'est pas l'écartement entre les partiels, nous sommes dans le cas où il manque des partiels harmoniques dans le son. Dans le cas de la clarinette, il manque les partiels d'ordre pair, caractéristique d'un son creux. La fréquence fondamentale d'un son n'est pas forcément un maximum d'énergie du spectre. La perception du maximum d'énergie spectrale est à mettre en rapport avec un autre phénomène de perception de la hauteur, dit de hauteur spectrale par opposition à la hauteur tonale.

2.3.3 Enveloppe spectrale

Dans le cas d'un son harmonique, l'ensemble des amplitudes du spectre matérialisent l'enveloppe spectrale du son : c'est la forme lissée du spectre du signal, la couleur du son. Les bosses du spectre sont appelées les formants. Ce sont les parties les plus importantes de l'enveloppe spectrale sur le plan de la perception. C'est notamment grâce aux formants que nous pouvons distinguer les voyelles de la voix. Les notions d'enveloppe et de formant sont illustrées par la figure 9.

2.4 Synthèse additive

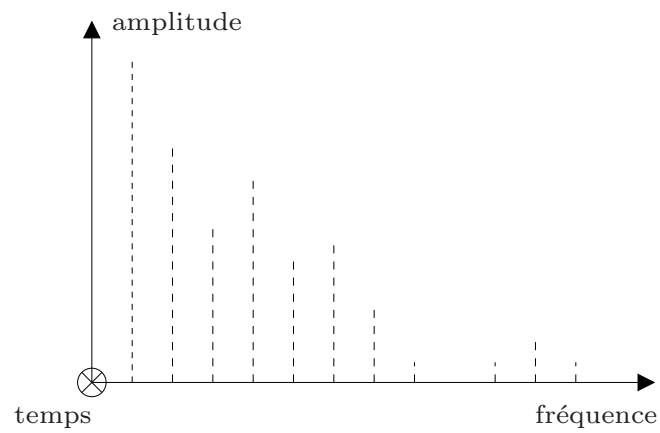
Les modèles spectraux paramètrent le son au niveau de l'oreille interne et sont ainsi proches de la perception. À la base de ces modèles se trouve la synthèse additive, qui permet de reproduire fidèlement de nombreux sons naturels et de les manipuler. La structure de base de cette synthèse est le partiel. Les modèles spectraux fondés sur la synthèse additive nécessitent le calcul d'un grand nombre d'oscillateurs sinusoidaux à l'aide d'algorithmes très efficaces. La synthèse additive, présentée par exemple par Moorer [Moo77], est donc à l'origine de la représentation spectrale des sons, et se base sur le théorème de Fourier.

La synthèse additive consiste pour chaque partiel du son à calculer son signal correspondant dans le domaine temporel, et à additionner ces valeurs. Le signal $s(t)$ peut alors s'exprimer ainsi :

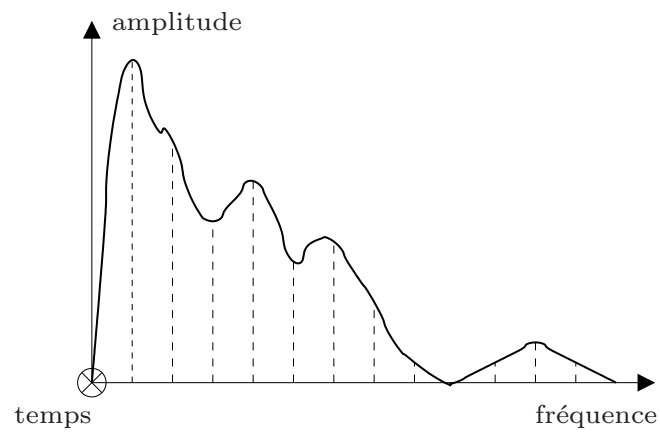
$$s(t) = \sum_{i=1}^N a_i(t) \sin(\Phi_i(t)) \quad (17)$$

La synthèse de sinusoides dans un modèle stationnaire à court terme consiste à générer une somme de sinusoides dont les paramètres d'amplitude et de fréquence sont constants dans la trame de synthèse. Dans le cas d'un signal stationnaire, a_i est donc constante, et $\Phi_i(t) = 2\pi f_i t + \phi_i$. Les paramètres f_i , a_i et ϕ_i sont, respectivement la fréquence, l'amplitude et la phase du i -ième partiel. ϕ_i est la phase initiale du partiel i au début de la trame de synthèse. La synthèse additive consiste alors à calculer par fenêtre de synthèse le signal :

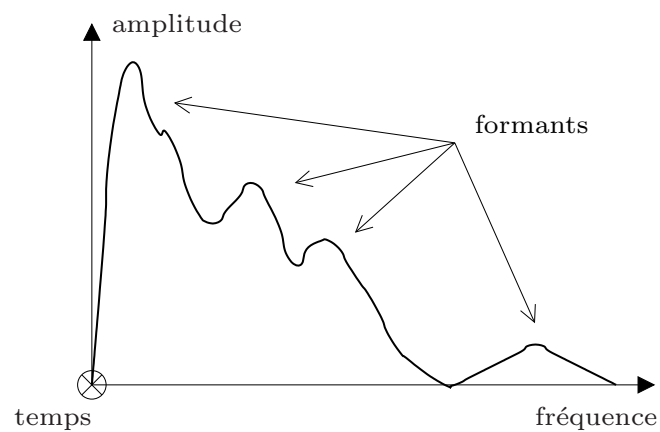
$$s(t) = \sum_{i=1}^N a_i \sin(2\pi f_i t + \phi_i) \quad (18)$$



(a)



(b)



(c)

FIG. 9 – (a) Spectre d'un son harmonique. (b) L'enveloppe spectrale est la forme lissée du spectre. (c) Un formant est un maximum d'énergie dans le spectre, une bosse de l'enveloppe spectrale.

Traditionnellement, un échantillon est calculé par partiel, et ces échantillons sont additionnés pour donner un échantillon du son. La complexité de la synthèse additive est alors proportionnelle au produit du nombre de partiels par la fréquence d'échantillonnage. Mais utiliser la fonction sinus pour calculer les échantillons de chaque partiel, ceci pour chaque échantillon du son, est très coûteux en temps de calcul. Utiliser la synthèse additive pour synthétiser un orchestre symphonique est alors un défi, que nous souhaiterions relever en temps réel. C'est pourquoi nous avons besoin d'algorithmes rapides de synthèse additive. Nous présentons dans la section 2.4.1 un algorithme incrémental, le résonateur numérique, qui permet de calculer le signal de chaque oscillateur de façon optimale, et qui est donc très efficace pour la synthèse additive. Une autre méthode pour ce type de synthèse est présentée dans la section 2.4.2. Elle utilise la transformée de Fourier inverse pour passer du domaine fréquentiel au domaine temporel trame par trame.

2.4.1 Résonateur numérique

Pour que la synthèse additive soit rapide, il faut calculer le plus efficacement possible les échantillons temporels pour chacune des sinusoïdes :

$$s[k] = a \sin(k\Delta_\phi + \phi) \quad \text{où} \quad \Delta_\phi = \frac{2\pi f}{F_e} \quad (19)$$

avec a , f , et ϕ , respectivement l'amplitude, la fréquence et la phase initiale d'un partiel, et F_e la fréquence d'échantillonnage du son.

Les contributions des différentes sinusoïdes sont ensuite additionnées pour obtenir les échantillons du son sur une trame temporelle. Comme la pulsation Δ_ϕ est constante au cours du temps, un algorithme récursif de calcul de la fonction sinus peut être mis en œuvre.

La méthode du résonateur numérique, utilisée par Gordon et Smith [GS85], ou encore par Smith et Cook [SC92], est très efficace pour la synthèse additive. Elle permet de calculer le signal de chaque partiel avec un nombre optimal d'opérations. Les valeurs de la fonction sinus sont données par un algorithme incrémental, ce qui évite les appels coûteux à la fonction pour chaque partiel. Marchand et Strandh [MS99, Mar00] proposent également d'utiliser le résonateur numérique pour la synthèse additive. Le résonateur est initialisé pour chaque partiel comme indiqué par l'équation (20), avec F_e la fréquence d'échantillonnage du son, a , f , et ϕ respectivement l'amplitude, la fréquence, et la phase initiale du partiel, et Δ_ϕ l'incrément de phase :

$$\begin{cases} \Delta_\phi & = 2\pi f / F_e \\ s[0] & = a \sin(\phi) \\ s[1] & = a \sin(\phi + \Delta_\phi) \\ C & = 2 \cos(\Delta_\phi) \\ s[k+1] & = C \cdot s[k] - s[k-1] \end{cases} \quad (20)$$

Le calcul incrémental d'un échantillon correspondant à un partiel requiert une addition et une multiplication. Les échantillons ainsi calculés sont ensuite additionnés pour calculer un échantillon du son.

La prise en compte de propriétés acoustiques permet à Lagrange et Marchand [LM01] d'accélérer la synthèse. En considérant le seuil d'audition et les masquages présentés dans la section 1.2.4, des partiels non audibles peuvent être ignorés. Le résonateur numérique a alors moins de partiels à traiter, la synthèse est accélérée.

2.4.2 Transformée de Fourier inverse

Dans le but de synthétiser efficacement de nombreuses sinusoïdes simultanément, Freed, Rodet et Depalle [FRD92, FRD93] ont proposé d'utiliser la transformée de Fourier inverse. Il s'agit de reconstruire le spectre à court terme du son à un instant t , en ajoutant la contribution de chaque partiel, et d'appliquer ensuite la transformée de Fourier inverse (FFT^{-1}) trame par trame pour obtenir la représentation temporelle du son, comme une sorte de vocodeur de phase inverse. Si la transformée de Fourier inverse est rapide, l'étape de reconstruction du spectre à court terme est cependant délicate. La figure 10 donne un exemple des partiels à générer dans le spectre pour synthétiser une seule sinusoïde, en tenant compte de la fenêtre utilisée pour la transformée de Fourier inverse.

Le meilleur gain de complexité s'obtient quand le nombre d'oscillateurs est grand en comparaison du nombre d'échantillons à calculer pour chaque trame. Cette approche est très intéressante, parce que la complexité n'est alors plus dépendante du produit nombre d'oscillateurs par fréquence d'échantillonnage. Cependant, les paramètres sonores doivent être fixés sur un intervalle de temps pour effectuer la transformée de Fourier inverse. Le contrôle des paramètres de la synthèse est donc moins souple qu'avec le résonateur numérique présenté dans la section précédente. Marchand [Mar00] a comparé en 2000 les deux méthodes, et estimé que l'utilisation du résonateur numérique était légèrement avantageuse. Meine et Purnhagen [MP02] ont conclu 2002 que la transformée de Fourier inverse était plus rapide. En fait, les deux méthodes sont proches, et leur comparaison dépend surtout du contexte (processeur, système d'exploitation, ...) et de détails d'implémentation.

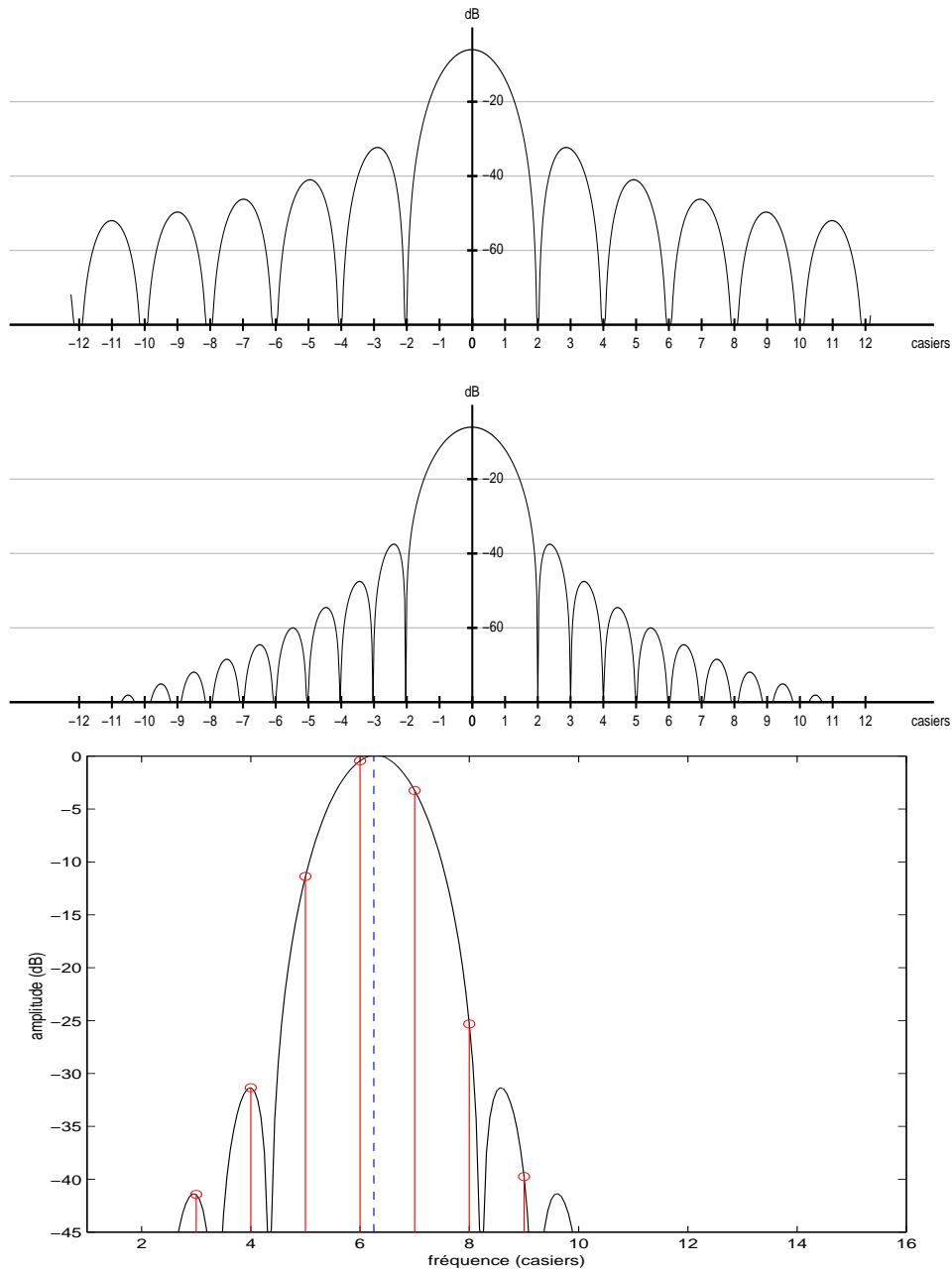


FIG. 10 – Spectres des fenêtres de Bartlett (en haut) et de Hann (au milieu). Ces fenêtres peuvent servir de fenêtre de base pour la transformée de Fourier ou la transformée de Fourier inverse. L'analyse sonore d'un signal sinusoïdal pur devrait donner un seul partiel dans le spectre (en bas, partiel en pointillés), mais l'étalement du pic d'analyse dû au fenêtrage entraîne la création de nombreuses valeurs (en bas, cercles) dans le domaine fréquentiel. Pour effectuer une synthèse sonore additive, il faut reconstruire un tel spectre discret pour chaque composante sinusoïdale à générer dans le son.

Chapitre 3

Musique instrumentale

La musique instrumentale est l'ensemble des éléments de la musique qui concernent la pratique instrumentale, de la notation musicale à la performance. Une partition musicale ou sa représentation symbolique et informatique sont des exemples de notation musicale. La performance instrumentale concerne notamment les domaines de l'interprétation et de l'analyse des paramètres musicaux dans le son.

La diversité des sons naît de la complexité des paramètres musicaux, et le son n'est qu'une composante bas-niveau de la musique instrumentale. Desainte-Catherine et Marchand [DCM99] proposent d'unifier cette dualité musique-son sur une même échelle de contrôle, en fonction de la fréquence de contrôle des paramètres sonores. La figure 11 donne une illustration de cette échelle. Le contrôle associé à la notation musicale est compris entre 0 à 8 Hz et concerne les durées des notes, les mesures, les phrases ou même les formes musicales. La plage de fréquence de 8 à 20 Hz est dédiée au contrôle des paramètres sonores d'amplitude ou de fréquence. Les plages supérieures concernent l'audition. Cette échelle, restreinte au contrôle des paramètres, donne une idée de la hiérarchisation générale musique-son. Avec des connaissances sur la musique instrumentale, nous pouvons alors espérer analyser dans le son, non seulement les évolutions des paramètres sonores, mais aussi repérer des structures musicales ou des volontés d'interprétation expressive.

Nous donnons dans ce chapitre quelques notions sur la musique instrumentale, sa notation, son interprétation et ses paramètres. Nous commençons dans la section 3.1 par présenter la notation musicale occidentale. La norme MIDI, qui propose une représentation informatique pour cette notation, est introduite dans la section 3.2. Des éléments sur l'histoire et les techniques de l'interprétation instrumentale sont ensuite donnés dans la section 3.3. Enfin, les principaux paramètres de la performance musicale sont listés dans la section 3.4.

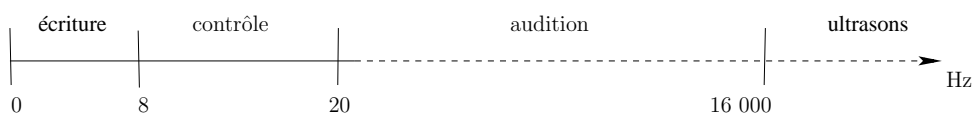


FIG. 11 – Échelle des fréquences de variation des paramètres.

3.1 Notation musicale

La notation musicale est vivante, et aujourd'hui encore, de nouveaux symboles sont constamment proposés pour représenter des sons instrumentaux, notamment dans la musique contemporaine. Massin et Massin [MM98] précisent que la musique est un langage organisé, et que la grammaire musicale, comme celle des langages parlés, a beaucoup évolué au cours de l'histoire musicale. Mais le code a toujours suivi l'expérience : en musique, l'intuition des compositeurs et leurs découvertes ont toujours précédé la codification du langage. C'est pourquoi cette grammaire est en perpétuelle mutation.

Nous nous concentrons alors sur la notation musicale académique (occidentale et néo-classique). Les premiers documents que nous pouvons consulter sur la notation musicale occidentale datent environ du 3ème siècle avant notre ère et proviennent de la Grèce. L'apparition de l'imprimerie pendant la Renaissance est une étape importante qui mène vers une standardisation de la notation musicale telle que nous la connaissons. La notation répond à une conception de la musique orientée vers la fixation de certaines propriétés du son, en priorité la hauteur et la durée. Elle peut être imprécise lorsqu'il est question de caractéristiques telles que le timbre ou l'intensité. L'ambiguïté des signes dont se sert le musicien pour transmettre ses idées musicales représente aussi une chance, car cela permet de s'adapter à des contextes stylistiques et personnels différents.

De nombreux symboles peuvent être utilisés pour la notation musicale. Il est intéressant de noter que la plupart des valeurs qu'ils symbolisent (valeurs de durée, d'intensité, ...) sont relatives entre elles, et n'ont pas de correspondance directe avec des valeurs dans le monde physique. Ainsi une note *piano* doit être jouée doucement, il n'est pas indiqué quelle est la valeur de son amplitude. Le plus important est de respecter l'échelle des valeurs. Ainsi une note *piano*, quelle que soit son intensité, doit être jouée moins fort qu'une note avec une nuance *mezzo forte*. Le rapport entre les différentes valeurs de durée, de hauteur ou d'intensité des notes est donc plus important que les valeurs elles-mêmes.

3.2 Format MIDI

Le *Musical Instrument Digital Interface*, ou MIDI, est un protocole de communication et de commande permettant l'échange de données entre instruments de musique électroniques pouvant être des ordinateurs. Apparue en 1982, il est géré par un comité international, l'*International Midi Association*. Sous le terme MIDI sont regroupées plusieurs normes, relatives au protocole logique, à l'interface physique, au format de fichier et à l'attribution des sons.

Divers types de messages surviennent entre deux instruments MIDI connectés, qui témoignent du passage des événements MIDI en cours. L'un des plus communs est le message *Note On* : lorsqu'une note est enfoncée sur un clavier MIDI (ou un équivalent), un message *Note On* est envoyé avec des informations sur le moment de l'événement, le port MIDI utilisé, le canal MIDI de l'événement (de 1 à 16), le type d'événement (*Note On*), la note envoyée (de la note 0 à la note 127), la vitesse de l'attaque, ou vélocité. Les hauteurs de note représentables vont de Do-2 à Sol8, avec Sol3 la note Sol de la *clé de Sol*, avec une résolution d'un 1/2 ton. La hauteur peut être réglée par demi-tons, avec une précision d'1/4096 de demi-ton. Le format MIDI peut alors être utilisé pour jouer des notes dans des gammes non tempérées (avec des hauteurs de note différentes de celle

Note On	Indication qu'une touche est enfoncée. La vitesse d'enfoncement, la vélocité, est codifiée de 0 à 127 afin de transmettre la nuance.
Note Off	Indication qu'une touche est relâchée.
Aftertouch	Correspondance avec la pression effectuée sur la touche. La valeur va de 0 à 127.
Pitch Bend	Indication d'une action sur la molette <i>Pitch Bend</i> . Cette molette joue sur la hauteur des notes jouées.
Program Change	Changement de programme, qui correspond à la sélection d'un son sur un synthétiseur. Les valeurs théoriques sont de 16384, soit 128 banques de 128 sons.
Control Change	Changement de contrôle correspondant aux réglages des paramètres sonores d'un synthétiseur (volume, sustain, ...).
Systèmes Exclusifs (SysEx)	Code propre à chaque constructeur, permettant la programmation en profondeur de chaque synthétiseur.
MIDI Clock	Horloge de synchronisation entre 2 séquenceurs MIDI.

TAB. 2 – Principaux types d'événements MIDI.

d'un piano standard). Les principaux événements MIDI sont donnés dans le tableau 2. Des méta-événements peuvent être insérés dans un fichier MIDI, comme l'armature d'une portée (qui concerne la tonalité) ou une indication de tempo.

Le format MIDI permet la représentation symbolique et informatique d'une partition musicale. En effet, la plupart des symboles écrits sur une partition sont représentables par des messages MIDI. Mais toute la relativité de la notation musicale est alors perdue : les nuances sont par exemple associées à des valeurs fixes, ainsi que les hauteurs de note. La performance musicale peut également être enregistrée dans un fichier MIDI, *via* une interface MIDI. Le niveau symbolique de cette représentation évite alors le recours à une analyse sonore pour obtenir des informations sur des macro-événements tels que le début et la fin des notes par exemple. La partition et son interprétation peuvent donc toutes deux être représentées avec le format MIDI. L'analyse de la performance, qui s'effectue principalement par comparaison entre partition et interprétation, en est facilitée.

3.3 Interprétation musicale

Lors d'une interprétation musicale, un musicien joue une partition et donne ainsi corps à la notation musicale. L'interprète assouplit généralement les règles de notation de la durée ou de la hauteur des sons, en prenant un certain nombre de libertés par rapport à l'écrit. Les paramètres musicaux non codifiés sur la partition, tels que le vibrato ou le timbre, sont utilisés dans un but expressif. L'interprétation d'une partition est alors plus ou moins éloignée de sa notation, en fonction de la technique de l'interprète, mais surtout en fonction de son expressivité. Il est cependant difficile de savoir quelle part d'initiative l'interprète a par rapport à la notation, d'autant que cela évolue avec l'histoire de la musique.

Dorian [Dor42] retrace l'histoire de l'interprétation instrumentale depuis la Renaissance. À l'époque baroque, qui accorde à l'interprète une place centrale, une grande part

du jeu musical reste dépendant du *bon goût* dicté par les maîtres de musique en matière d'interprétation. Mais l'interprète est roi, et il peut ajouter de nombreux ornements non présents dans la notation musicale. Après le 18ème siècle, l'activité de l'interprète se détache de plus en plus de celle du compositeur, leurs techniques gagnent en spécificité. La part de liberté qui revient à l'interprète s'amenuise, il devient de plus en plus soumis à la fidélité du texte écrit. Les improvisations ou fioritures précédemment ajoutées sont maintenant évitées. L'invention du métronome par Maelzel en 1859, suivie de celle du diapason, impose des normes pour le tempo et l'accord des instruments, et finit par contraindre l'interprète à une lecture exacte de la notation musicale.

En informatique musicale, Widmer *et al.* [WDG⁺03] cherchent aujourd'hui le paramètre d'Horowitz, du nom du célèbre pianiste russe Vladimir Horowitz, *i.e.* ce qui caractérise le jeu d'un interprète en le rendant unique. Les écarts observés entre une performance instrumentale et la notation musicale d'une œuvre détermine des paramètres musicaux d'interprétation.

3.4 Paramètres musicaux

Le *son instrumental* est la perception d'une onde sonore produite par un instrument. Suivant le niveau d'analyse, les paramètres analysés peuvent être sonores ou bien des paramètres de l'interprétation instrumentale. Ainsi, si Marchand [Mar01a] propose d'utiliser une méthode efficace avec transformée de Fourier pour retrouver les paramètres sonores tels que la hauteur des sons, Marchand et Raspaud [MR04] proposent alors de ré-analyser ces résultats, de nouveau avec une transformée de Fourier, pour trouver des paramètres sonores de plus haut niveau, comme le paramètre de contrôle du vibrato instrumental. L'analyse sonore d'un son instrumental peut également permettre d'évaluer la technicité d'un interprète, d'analyser son expressivité, ou de mettre en évidence la structure musicale d'un morceau. Pour cela, nous donnons quelques indications sur les paramètres importants de la performance instrumentale.

3.4.1 Rythme

Le rythme est un paramètre essentiel de la musique instrumentale, il régit les structures temporelles. Il s'inscrit dans un mouvement général d'un morceau, le *tempo*, et à l'intérieur de *mesures*. Le rythme définit la longueur des notes, leur rapport en durée. Ces durées sont évaluées en temps, unité temporelle relative au tempo, au mouvement général. Ainsi la ronde vaut 4 temps, la noire 1 temps ou la croche 1/2 temps. Le rythme écrit peut être très affecté par une interprétation instrumentale, par exemple lorsqu'un interprète joue un morceau en *swing*. Le *swing* a pour principal effet de remplacer un rythme avec des notes de durées égales par un autre avec des notes de durées inégales.

3.4.2 Hauteur de note

La hauteur d'une note est déterminée sur une partition par sa position sur la portée musicale, en fonction de la clé utilisée pour la portée (clé de Sol, clé de Fa, ...). Suivant sa fréquence et la tessiture de l'instrument joué, cette hauteur peut être considérée comme grave, médium ou aiguë. La hauteur d'une note est associée à une fréquence de référence,

note	f	note	f
Do	261,6 Hz	Fa #	370,0 Hz
Do #	277,2 Hz	Sol	392,0 Hz
Ré	293,7 Hz	Sol #	415,3 Hz
Ré #	311,1 Hz	La	440 Hz
Mi	329,7 Hz	La #	466,2 Hz
Fa	349,2 Hz	Si	493,9 Hz

TAB. 3 – Correspondance entre la hauteur de note et la fréquence f du signal sonore correspondant, pour une octave, et dans le cas d'un diapason à 440 Hz.

donnée par rapport au La du diapason par exemple. Toutes les notes d'un instrument sont alors accordées en fonction de cette référence. La relation suivante donne la fréquence d'une note en fonction de sa hauteur H :

$$F(H) = F_0 \cdot 2^{(\text{oct}-3) + \frac{\text{ton}-10}{12}} \quad (21)$$

avec F_0 la fréquence de référence pour le La de l'octave 3, oct l'octave de la note dont la fréquence est cherchée et ton le numéro du demi-ton de la note cherchée dans la gamme (Do a le numéro 1, Do # le numéro 2, La le numéro 10, ...)

Si généralement la fréquence de référence choisie est le La à 440 Hz, cette correspondance peut varier. La musique baroque peut par exemple se jouer comme à l'époque, avec un diapason à 415 Hz. Le tableau 3 donne les correspondances des hauteurs de note avec la fréquence du signal sonore correspondant, dans le cas d'un diapason à 440 Hz.

La justesse est un paramètre musical important, qui se rapporte à la hauteur des notes. Il y a trois critères qui peuvent participer à la justesse d'une note :

- la correspondance ou non de la hauteur de la note avec une échelle de fréquence de référence. Cette échelle peut être établie à partir du La d'un diapason à 440 Hz par exemple. Dans ce cas, la note La est dite fausse si sa fréquence n'est pas égale à 440 Hz.
- la constance de la fréquence au cours de la durée de la note (non vibrée). Si la note tenue n'a pas une fréquence constante dans le temps, elle est fausse.
- le respect des intervalles entre les notes. La justesse est évaluée en fonction de la justesse de l'intervalle depuis la note précédente, et plus généralement en fonction de tous les intervalles dans un passage musical donné. Suivant le tempérament choisi (valeurs des intervalles dans la gamme), la justesse n'est pas la même. La différence entre l'échelle tempérée du piano par exemple, et l'échelle naturelle – qui différencie par exemple un Sol # d'un La b , au contraire du piano – peut donner une impression de fausseté à l'une des deux échelles.

Le paramètre de justesse est très relatif. C'est pourquoi c'est un des paramètres les plus difficiles à maîtriser, avec le timbre, pour les instruments entretenus.

3.4.3 Nuance

Dans la musique occidentale, les nuances sont l'ensemble de signes notés sur une partition ayant pour fonction d'indiquer l'intensité relative d'une note, d'une phrase, ou encore

terme italien	abréviation	traduction
<i>pianissimo</i>	<i>pp</i>	très faible
<i>piano</i>	<i>p</i>	faible
<i>mezzo piano</i>	<i>mp</i>	moyennement faible
<i>mezzo forte</i>	<i>mf</i>	moyennement fort
<i>forte</i>	<i>f</i>	fort
<i>fortissimo</i>	<i>ff</i>	très fort

TAB. 4 – *Exemple de nuances.*

d'un passage entier d'une œuvre musicale. La succession des nuances et l'interprétation qu'en fait le musicien restituent la dynamique de l'œuvre. Dès le 18ème siècle, les compositeurs prennent l'habitude de noter les nuances au moyen de divers termes italiens – la plupart du temps, seules leurs abréviations sont utilisées – et de quelques autres signes, placés au-dessus ou au-dessous de la portée. Le tableau 4 indique les nuances les plus utilisées pour la notation musicale. Il s'agit des nuances *pianissimo* à *fortissimo*, qui fixent l'intensité des notes. Les dynamiques de nuance sont des indications de variation de la nuance sur un passage musical. Le *crescendo* est ainsi une dynamique de nuance qui augmente continuellement une nuance, pendant le temps du *crescendo*, allant de la nuance de départ jusqu'à atteindre la nuance d'arrivée. Le *decrescendo* procède de la même façon, mais en abaissant la nuance. Les nuances utilisant le préfixe *subito*, comme *subito forte*, créent un fort contraste avec les nuances précédentes.

3.4.4 Timbre

Comme nous l'avons vu dans la section 1.2.3, le timbre d'un son est d'abord défini par défaut : c'est ce qui différencie le son de deux instruments qui jouent la même note, avec la même nuance. C'est le paramètre d'interprétation le plus riche et sûrement le moins bien décrit. McAdams *et al.* [MWD⁺95] ont étudié la perception de ce paramètre complexe, afin de le représenter dans un espace multi-dimensionnel, et ainsi mieux le décrire.

3.4.5 Vibrato et tremolo

Le vibrato est un effet appliqué à une note de musique, au moyen d'un vibré manuel, parfois grâce à un accessoire, au souffle ou à la voix. Il s'agit de provoquer une variation de la hauteur du son autour de sa hauteur moyenne. D'après Londeix [Lon70], le vibrato est "l'ondulation légère et régulière du son, pratiquée intentionnellement, suivant l'intérêt musical, pour renforcer, au moins conventionnellement, la valeur expressive des notes". Il ajoute que "le vibrato est un moyen merveilleux de rendre plus convaincante une phrase musicale, plus pathétique un accent, à condition que ce vibrato soit maîtrisé et de qualité. Il n'y a pas de vitesse absolue de vibrato. Chacun doit affirmer son goût et sa personnalité, par la plus ou moins grande amplitude de l'ondulation (qui peut être variable selon la phrase à jouer), par la plus ou moins grande fréquence des ondulations (qui peut être variable selon le style de l'œuvre)". Le vibrato est donc un paramètre expressif employé par les musiciens sur de nombreux instruments différents, et dont Verfaillie *et al.* [VGD05]

fait l'état de l'art. Ce paramètre musical est rarement noté sur une partition, ce qui laisse la liberté à l'interprète de le personnaliser et de le rendre très expressif.

Alors que le vibrato est une variation de hauteur, le tremolo est une variation de l'intensité de la note autour d'une valeur moyenne en conservant la hauteur de départ.

3.4.6 Articulation

“L'articulation est l'art instrumental d'imiter les inflexions de la voix humaine”, d'après l'encyclopédie Wikipedia¹. Cela consiste essentiellement dans la maîtrise des attaques des notes, ainsi que dans la gestion des micro-silences entre les notes (silences d'articulations) qui donnent toute leur clarté au message musical. Tous les contrastes d'appuis, d'accents et d'enchaînements des notes participent à l'articulation d'un passage musical. À l'époque de la musique ancienne, alors que les nuances étaient beaucoup moins développées qu'aujourd'hui, l'articulation était un des moyens principaux d'expression.

Nous pouvons distinguer l'effet de *legato*, indiqué par une ligne courbe englobant les notes d'une mélodie devant être interprétées de façon liée (pas de micro-pauses entre les notes). Au contraire, un jeu *staccato* désigne un type de phrasé dans lequel les notes des motifs et des phrases musicales doivent être exécutées avec des suspensions entre elles, jouées *détaché*. Ces notes détachées sont indiquées au moyen d'un point placé au-dessus ou au-dessous de la tête de la note.

3.4.7 Phrasé

Dans la musique occidentale, le phrasé indique la façon d'exécuter les différentes phrases musicales. Une phrase musicale est un ensemble de mesures d'une partition qui constituent un groupe homogène, une ligne mélodique. La phrase musicale organise la durée de la musique comme les phrases d'un texte. De la même façon qu'un texte n'est pas qu'une succession de mots, une mélodie ne doit pas être une simple succession de notes, mais plutôt un enchaînement de phrases musicales. Des silences écrits séparent souvent les phrases successives, et même parfois les divers motifs de chaque phrase. Les phrases musicales sont souvent ponctuées par une cadence. En harmonie classique, une cadence est une progression harmonique de deux accords, destinée, par son caractère conclusif, à marquer la fin d'une pièce ou d'une phrase musicale. Les cadences constituent la respiration du discours musical.

¹[http://fr.wikipedia.org/wiki/Articulation_\(musique\)](http://fr.wikipedia.org/wiki/Articulation_(musique))

Deuxième partie

Analyse et évaluation du jeu
instrumental

L'analyse de la performance musicale est un sujet transverse, traité par de nombreuses études de biomécanique, de psychoacoustique, de musicologie ou d'informatique. Nous nous limitons dans ce document à l'étude de la performance instrumentale dans un contexte académique.

Nous proposons de décomposer les différentes influences sur la performance musicale en trois composantes principales : physique, technique et expressive. Nous donnons dans le chapitre 4 des exemples de l'influence de ces différentes parties sur la performance. Les études existant sur l'analyse de la performance instrumentale sont présentées, et classées en fonction de leur méthode d'analyse des paramètres de la performance. Nous donnons aussi des éléments pédagogiques sur la pratique des exercices techniques instrumentaux et sur l'évaluation instrumentale dans un contexte académique. Ces éléments vont nous guider dans les chapitres suivants.

Le chapitre 5 présente nos travaux sur le doigté pianistique. Après une introduction à propos du piano et du doigté concernant cet instrument, nous présentons des études de biomécanique qui mettent en évidence les contraintes physiques de la main au piano. Nous proposons aussi une méthode de doigté automatique d'une mélodie. À partir de ces éléments, nous jetons les bases d'une nouvelle méthode d'analyse de la performance pianistique, qui se propose de reconnaître le doigté utilisé par un interprète. Cette méthode utilise la programmation dynamique pour comparer la performance avec une base d'apprentissage, en intégrant des considérations biomécaniques.

Suite aux travaux menés avec Lagrange [RL06], nous présentons dans le chapitre 6 nos recherches sur l'évaluation technique de saxophonistes. Nous commençons par décrire le saxophone et sa pédagogie académique. Nous proposons une nouvelle méthode automatique pour l'évaluation technique de saxophonistes, à partir de l'analyse de leurs performances. En fonction de leur réussite à des exercices instrumentaux simples, nous expliquons comment nous pouvons leur attribuer une note technique générale, les interclasser et visualiser leurs résultats.

Chapitre 4

Analyse du jeu instrumental

L'analyse de la performance musicale est complexe. Outre les méthodes mises en œuvre, ce sont surtout la multiplicité et l'interdépendance des paramètres à extraire qui rend l'analyse délicate. Dans le cas de la performance instrumentale, le résultat sonore change en fonction de la physiologie de l'instrumentiste, de ses gestes, de sa technique, de son expressivité ou de son humeur par exemple. Nous nous demandons si une analyse sonore peut retrouver ces informations, et dans ce cas quels paramètres sonores il faut alors observer dans la performance.

Palmer [Pal97] et Gabrielsson [Gab99] retracent l'historique des différentes méthodes utilisées pour analyser la performance musicale. Elles concernent de nombreux domaines de recherche : biomécanique, synthèse sonore, musicologie, psychologie ou encore intelligence artificielle. Il s'agit d'analyser les paramètres sonores, ou encore des événements MIDI, afin de modéliser la performance. L'évaluation de la performance se base en général sur une comparaison à un modèle, qui peut être une partition, un rythme, ou l'évolution théorique d'un paramètre musical (nuance, vibrato, ...). Ce sont les écarts à ce modèle qui permettent d'évaluer et de comparer les performances. Une modélisation de la performance est alors possible, et l'application en synthèse sonore de cette modélisation peut par exemple permettre d'interpréter une partition par un ordinateur, et d'humaniser ainsi une performance automatisée.

Toute la difficulté est d'isoler les paramètres dans le modèle de performance. Si le contexte n'est pas clairement identifié lors de l'analyse, il est difficile de déterminer si un écart de rythme par rapport à un modèle est causé par le niveau technique insuffisant de l'interprète, par sa volonté expressive, sa fatigue, ou la conséquence de l'utilisation d'un instrument particulier. Au piano par exemple, dans le cas d'une mélodie accompagnée, il a été observé que les notes de la mélodie ne sont pas synchrones avec les accords, et surviennent environ 30 ms avant : c'est le *melody lead*. Cela a été successivement interprété comme une volonté expressive de ressortir la mélodie de l'accompagnement, par Palmer [Pal96], puis analysé comme un artefact de vélocité, phénomène physique provenant du piano, par Repp [Rep96] puis Goebel [Goe01]. Kopiez [KBGA03] donne une autre illustration de cette ambiguïté, puisqu'il faut savoir que le tempo et les rythmes instables du pianiste ayant suivi son protocole n'est pas causé par des insuffisances techniques, mais par le jeu en continu d'une pièce de Satie pendant plus de 15 heures !

Ce chapitre présente les différentes approches de la performance instrumentale, de

l'analyse à la synthèse. Nous donnons également des éléments de pédagogie sur l'enseignement instrumental académique, qui peuvent être utiles pour construire des méthodes et protocoles permettant d'observer un aspect précis de la performance. Nous commençons dans la section 4.1 par énumérer les paramètres de la performance musicale. La section 4.2 explique quels paramètres sonores ou musicaux doivent être suivis dans la performance. Les écarts de ces paramètres à un modèle permettent d'évaluer et de modéliser la performance musicale (section 4.3). Les sections 4.4, 4.5 et 4.6 indiquent quels facteurs respectivement physiques, techniques et expressifs peuvent influencer sur la production du son. Des éléments pédagogiques sur l'enseignement académique sont donnés dans la section 4.7, afin d'expliquer notamment quelles informations supplémentaires peuvent être obtenues à partir de la performance instrumentale.

4.1 Paramètres de la performance musicale

Si la musique était exactement jouée comme elle est écrite, les ordinateurs seraient des virtuoses, et la performance musicale serait robotique. Quand ils interprètent une œuvre, les musiciens dévient en effet de façon significative de ce qui est écrit sur leur partition. Seashore et Metfessel [SM25] précisent que les ressources illimitées de l'art vocal ou instrumental proviennent des déviations artistiques effectuées par rapport au son pur, à l'exact, au parfait, au précis, au rigide. Mais les déviations ne sont pas toujours artistiques. Nous proposons alors de décomposer les différentes influences sur la performance instrumentale en trois composantes principales : la composante physique, la composante technique et la composante expressive.

4.1.1 Composante physique

La composante physique s'appuie sur le couple instrument-instrumentiste, qui impose à la performance instrumentale des contraintes physiques évidentes. La nécessité pour l'instrumentiste d'interagir de façon physique avec l'instrument fait directement intervenir sa physiologie. Ainsi, les mains d'un pianiste, les lèvres d'une flûtiste, l'air expiré d'un clarinetiste, le larynx d'un chanteur ou le bras d'une violoniste sont autant de parties physiques très importantes dans l'exécution instrumentale, et ont un impact certain sur la performance sonore. L'influence physique de l'instrument est aussi naturelle. Les propriétés acoustiques de la forme conique du saxophone, la caisse de résonance d'une guitare ou l'éloignement des marteaux d'un piano aux cordes influent directement sur le son produit, soit en donnant un timbre particulier, ou bien en provoquant des décalages temporels dans la performance. Cette composante physique de la performance est modélisée principalement en biomécanique pour l'instrumentiste et en physique, acoustique et informatique pour les instruments, en particulier avec les modèles physiques d'instruments. Une modélisation physique complète de la performance instrumentale doit unifier ces recherches pour présenter un modèle instrumentiste-instrument.

4.1.2 Composante technique

La composante technique dans la performance musicale est peu étudiée. Pourtant, l'instrumentiste est forcément imparfait, même si les défauts techniques d'un virtuose

peuvent être difficiles à détecter. Alors il y a dans le son de la performance instrumentale des influences du niveau technique de l'interprète. Le contrôle de la justesse du son dans le temps pour un saxophone, la difficulté d'un enchaînement de doigtés au piano (où la technicité essaye de compenser une influence physique contraignante), la maîtrise d'une dynamique linéaire de nuances au violon sont autant d'empreintes de la technicité dans le son. Dans un contexte pédagogique, il existe des méthodes d'apprentissage instrumental automatisé qui proposent une analyse technique de la performance. Mais de telles méthodes ne détectent en général que des erreurs grossières, telles qu'une erreur de note ou de rythme.

4.1.3 Composante expressive

La composante expressive est tout ce qui n'est lié ni au physique, ni au technique dans la performance. Le vibrato qui se resserre en fin de note pour un chanteur, des appuis très marqués au piano sans qu'ils soient indiqués sur la partition, un ralenti dans une cadence à la clarinette, parmi tant d'autres intentions, sont des aspects expressifs de la performance instrumentale. Certaines méthodes recherchent alors les règles d'interprétation qui régissent la performance de l'interprète. Les différences d'interprétation peuvent être le reflet des différentes analyses musicales d'une même partition, ou bien l'interprète transmet plutôt une émotion personnelle, hors des règles d'analyse musicale ou de style. La performance musicale expressive peut être modélisée par des règles, qui peuvent découler de l'expertise de musiciens professionnels, d'une analyse de la partition ou de l'analyse de nombreuses interprétations.

Différentes méthodes de visualisation peuvent mettre en lumière des aspects de la performance expressive. Langner et Goebel, puis Dixon, proposent ainsi le *Performance Worm* [LG03, DGW02, DGW05]. La performance est visualisée selon le tempo et la sonie. Des disques représentant chacun la mesure d'une partition forment un serpent qui représente l'interprétation générale d'un morceau (figure 12).

Le *Royal Institute of Technology* (KTH) de Stockholm propose des règles pour la performance expressive [FFBS91, Fri91, Fri95, FSF95]. Une de ces règles établit par exemple que dans une séquence de huit notes, une note sur deux peut être jouée un peu plus longtemps. Cette règle de *notes inégales* [Vei77] est utilisée dans le jazz et la musique baroque. D'après le KTH, il y a trois types différents de règles pour la performance musicale expressive :

- Les règles de différenciation, qui accentuent les différences entre les notes de la gamme (Do, Ré, Mi, ...) et entre les durées de notes (croche, double-croche, ...).
- Les règles de groupement, qui gèrent l'appartenance d'une note à un groupe de notes. En musique, cette appartenance existe à différents niveaux, le phrasé en est un exemple. Les effets possibles sont l'insertion de micro-pauses entre les notes n'appartenant pas au même groupe, ou un ralentissement du rythme en fin de groupe.
- Les règles d'ensemble, qui conservent un ordre dans les ensembles de notes. Elles synchronisent les différentes notes dans chaque voix pour répondre aux exigences des autres règles, en raccourcissant ou allongeant certaines notes. Elles gèrent aussi les accords de notes et les intervalles.

Devant la complexité et la diversité des paramètres d'une performance musicale, il est utile de se concentrer sur quelques paramètres. L'analyse est facilitée si le contexte complexe est clairement identifié lors de la performance (lieu d'enregistrement, physique

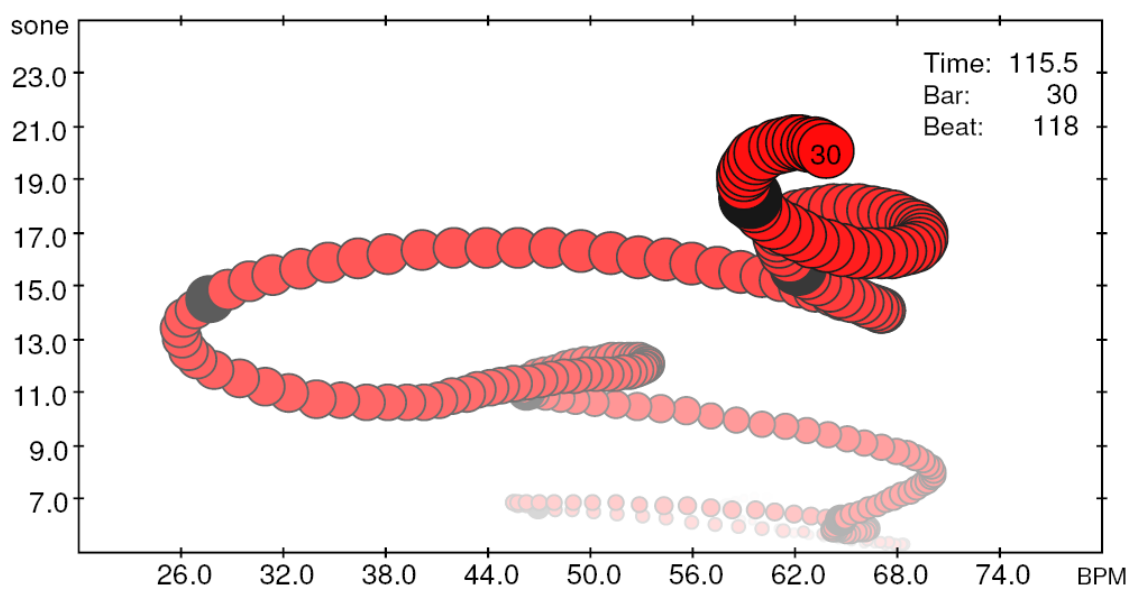


FIG. 12 – Le Performance Worm de Langner et Goebel propose une visualisation de la performance musicale expressive [DGW05]. Des disques, représentant chacun une mesure de la partition interprétée, sont placés sur un espace à deux dimensions. En fonction du tempo en abscisse et de la sonie en ordonnée, les disques forment un serpent représentant l'interprétation générale d'un morceau.

de l'instrument, physiologie de l'instrumentiste, niveau technique de l'interprète et de la partition, ...).

4.2 Analyse des paramètres sonores et musicaux

Pour pouvoir analyser la performance musicale, il faut d'abord avoir analysé les paramètres sonores ou musicaux de la performance. Les indications de tempo, les événements temporels, la dynamique des paramètres d'amplitude ou de vibrato, ou l'articulation sont des informations essentielles à l'évaluation et à la modélisation de la performance. L'analyse sonore est donc souvent la base de l'étude sur la performance musicale. La détection fiable des hauteurs, des amplitudes et des événements temporels de début et de fin des notes est nécessaire pour analyser finement les variations entre interprètes, par exemple. C'est en ce sens que Scheirer [Sch98], Bello [Bel03], ou Dixon [Dix04, Dix06] proposent des algorithmes de détection de hauteur et de début de note.

Le format MIDI est également très utilisé pour l'analyse de la performance. Il a l'avantage de proposer une représentation symbolique de la performance, avec des mesures de la vélocité, des événements de début et fin de note, et de nombreux autres paramètres décrits dans la section 3.2. En utilisant ce format, la précision des événements temporels n'est plus liée à la qualité de l'analyse sonore. Les informations MIDI peuvent être recoupées avec les données analysées depuis les fichiers audio pour une plus grande fiabilité. Les pianos numériques Bösendorfer sont pour cela un excellent moyen d'étude, puisqu'ils combinent la qualité d'un piano de concert à queue et des capteurs de type MIDI d'une très grande précision (en fait le format de données est proche du MIDI mais il est différent car il a été spécialement développé pour ce modèle de clavier). C'est l'instrument de référence avec le Disklavier de Yamaha pour les études sur la performance musicale au piano. En outre, quasiment tous les claviers numériques possèdent aujourd'hui une sortie pour les données MIDI. Comme le répertoire du piano est vaste, et que les outils sont à disposition, nous remarquons que cet instrument focalise de nombreuses études sur la performance. L'inconvénient du format MIDI est qu'il n'y a justement pas d'équipements MIDI disponibles pour tous les instruments. La plupart des familles d'instruments disposent d'instruments MIDI spécifiques, peu répandus, et qui ne sont pas équivalents à l'instrument acoustique, comme l'instrument MIDI à vent WX5 de Yamaha. Les études sur les instruments autres que le piano sont donc très dépendantes de la qualité des méthodes d'extraction des paramètres sonores.

4.3 Analyse des écarts à un modèle

L'analyse de la performance musicale est souvent réalisée par comparaison à un modèle. Le modèle peut être dans ce cas un rythme théorique, une partition, une dynamique théorique de nuances. Ce sont les écarts à ce modèle qui caractérisent la performance.

4.3.1 Modèle rythmique

Un moyen d'analyser la performance d'un point de vue physique ou technique est d'utiliser comme modèle des exercices instrumentaux de base dont le rythme est régulier. C'est

le cas des gammes ou des arpèges par exemple. Le rythme est alors défini par l'exercice, et nous pouvons mesurer l'écart de la performance à ce rythme.

Les mesures biomécaniques de Mac Kenzie et VanEerd [KV90] sur l'exécution de gammes au piano, et l'étude biomécanique du jeu au piano d'Ortmann [Ort62] suivent cette voie. C'est la base du travail préliminaire que nous avons effectué avec le piano, présenté dans le chapitre suivant.

4.3.2 La partition comme modèle

La méthode d'analyse de la performance la plus fréquente consiste à comparer l'interprétation instrumentale avec la partition de la pièce jouée. La partition de musique donne des indications de tempo, de rythme, les hauteurs des notes, les nuances, les effets de tempo, les articulations, *etc.* Tous ces paramètres musicaux théoriques peuvent être comparés aux données analysées à partir de la performance. Suivant le niveau technique d'un interprète, les écarts observés par rapport à la partition peuvent être volontaires ou non. Pour les études sur l'expressivité, il est communément supposé que l'influence physique ou technique est négligeable, en étudiant les performances d'interprètes professionnels. Pour étudier la performance d'un élève, il est au contraire supposé (ou demandé) qu'il ne soit pas expressif, afin de seulement détecter les erreurs techniques commises.

Dixon [Dix03] analyse automatiquement la performance expressive à partir d'enregistrements audio de pièces musicales connues. Il utilise la partition sous la forme d'un fichier MIDI pour aider l'algorithme. Le système d'alignement de partition proposé tente de faire correspondre les données audio à la partition, note par note, en cherchant les événements temporels près de leur date théorique indiquée par la partition. Cela permet de mesurer les choix subtils d'interprétation qui différencient les grands interprètes. Scheirer [Sch98] utilise en plus des notions d'analyse musicale. Dans un but pédagogique, les logiciels de tutorisation de l'enseignement instrumental comme *Piano Tutor* [DSJ⁺90, DSJ⁺93] proposent de comparer la performance à une partition afin de détecter les erreurs commises. Des exercices adaptés peuvent alors être proposés. C'est une analyse technique de la performance.

L'analyse de la performance fondée sur la comparaison avec la partition soulève cependant de nombreuses questions. Avec une telle méthode, il n'y a pas de prise en compte de la physique de l'instrument, alors que des difficultés techniques spécifiques sont associées à chaque instrument. Un léger écart temporel lors du démarrage d'une note grave au saxophone, par un élève moyen, peut être analysé comme une erreur technique, alors qu'il n'y aura pas de retard constaté pour un pianiste débutant (pas de difficulté de production du son au piano). Si la méthode d'analyse ne prend pas en compte les spécificités des instruments, il n'est pas possible de différencier l'influence physique et l'influence technique dans la performance. Il est donc très difficile de construire une méthode générale – pour plusieurs instruments différents – d'évaluation technique par exemple. Une autre question se pose sur le caractère volontaire ou non des écarts à la partition. Si les écarts sont volontaires (swing rythmique, ralenti, ...), ils sont expressifs, sinon ils sont influencés par la technique ou la physique. Il faut alors que le contexte d'enregistrement de la performance soit toujours bien connu pour éviter une erreur d'analyse, et principalement le niveau technique des instrumentistes. Des difficultés apparaissent également pour la comparaison des amplitudes de la performance avec les nuances indiquées sur la partition.

Décider si une note est bien jouée *mezzo forte* à partir de sa valeur d'amplitude est très risqué. De plus, les valeurs d'amplitude peuvent énormément varier en fonction du niveau sonore d'enregistrement de la performance. Le problème est le même avec la justesse, qui ne peut pas être évaluée par rapport à la partition, mais plutôt comme une évolution de la fréquence en fonction du temps.

4.3.3 L'évolution théorique des paramètres sonores

Afin d'analyser une performance instrumentale, il est possible d'étudier l'évolution des paramètres sonores, et la comparer à l'évolution théorique attendue. Afin d'étudier la justesse par exemple, l'évolution en fréquence peut être comparée à une valeur de fréquence constante, car une note juste ne doit pas voir sa fréquence varier dans le temps. La valeur de la constante de référence peut soit être la valeur initiale en fréquence de la note, soit la valeur moyenne des différentes valeurs en fréquence de la note. L'expressivité d'un vibrato peut être mesurée en fonction de l'évolution de sa fréquence et de son amplitude.

Pour analyser les nuances de la performance, au moins deux échelles de temps peuvent être utilisées. Pendant la durée d'une note, l'évolution de l'amplitude peut être évaluée en fonction de l'évolution attendue (*crescendo*, nuance fixe, ...). Les écarts à l'évolution attendue donnent des indications qui peuvent être à la fois physiques, techniques ou expressives. L'évolution des amplitudes sur plusieurs notes consécutives est aussi très intéressante, puisqu'il est alors possible de la comparer avec la dynamique attendue par la partition. L'évaluation est relative : il faut essayer d'établir une relation entre l'échelle des amplitudes utilisées par l'interprète et l'échelle des nuances dans la partition, quelles que soient les valeurs d'amplitude.

Bretos et Sundberg [BS03] ont ainsi mesuré l'évolution des paramètres de vibrato dans de longues notes chantées *crescendo* par dix sopranos différentes. Deux longues notes aiguës ont été choisies parmi dix enregistrements d'un aria de l'opéra *Aïda* de Verdi. Les deux notes sont chantées sans accompagnement instrumental, avec un *crescendo*. Ils en concluent que la fréquence du vibrato a tendance à s'accélérer à la fin de la note chantée. Jensen [Jen02] propose une étude sur l'évolution de la fréquence et de l'amplitude des sons instrumentaux. Il constate que les variations d'intensité influent sur l'enveloppe du spectre fréquentiel, et il propose un changement d'enveloppe modélisé en fonction de la fréquence. Nous avons aussi proposé avec Lagrange [RL06] de suivre les évolutions de la fréquence et de l'amplitude de sons tenus instrumentaux, pour une évaluation technique de saxophonistes cette fois. Nous présentons ce travail dans le chapitre 6.

4.4 Influence physique

Les travaux concernant l'influence de la physique de l'instrument sur la performance sont assez nombreux en informatique, notamment dans le domaine de la modélisation physique des instruments. Au contraire, l'impact de la physiologie du musicien sur sa performance, domaine bien connu en biomécanique, est beaucoup moins considéré.

4.4.1 Modèles physiques d'instruments

La performance est dépendante des caractéristiques structurelles de l'instrument et de ses accessoires. Au saxophone, la forme du bec, les nervures du bois de l'anche – sa *force* –, ou la forme conique du corps de l'instrument sont autant de facteurs influant sur la performance sonore. La synthèse par modèle physique consiste à produire un son à partir de la description physique d'un objet et de la manière dont il est utilisé pour produire un son. Il s'agit généralement d'une mise en équation simplifiée des contraintes physiques liées à l'instrument.

Fletcher et Rossing [FR98] présentent une étude générale sur les familles d'instruments et les modèles physiques associés. Dans ce domaine, de nombreux modèles sont proposés, spécifiques aux différentes excitations qui produisent le son (corde, lèvres, anche, ...). Ainsi Vergez et Rodet [VR01] proposent par exemple un modèle de trompette, avec un modèle des lèvres du trompettiste. Avanzini *et al.* [ABB⁺01] exposent la problématique de la modélisation du piano, avec notamment l'excitation de la corde et l'interaction marteau-corde. Ystad [Yst99, YV01b, YV01a] introduit un modèle de flûte en temps réel, prenant en compte à la fois l'aspect perceptif et l'aspect physique du son produit à la flûte.

4.4.2 Physiologie de l'instrumentiste

Il est évident que la physiologie de l'interprète influe sur sa performance. Certains modèles physiques évoqués ci-dessus prennent en compte des données physiologiques de l'instrumentiste, telles que les lèvres du flûtiste par exemple. Nous distinguons la physiologie de l'instrumentiste de ses gestes. Le volume d'air expirable, une main avec des doigts courts, les propriétés du larynx sont des données physiologiques, alors que la direction de l'air expiré, ou les mouvements de la main sont des gestes musicaux.

Fuks [Fuk97] explique par exemple comment la composition de l'air soufflé par un musicien varie en fonction du temps de l'expiration, et altère la performance des instrumentistes à vent. La biomécanique du musicien a souvent été étudiée, notamment à travers l'aventure des robots musiciens décrite par Kapur [Kap05]. Après des études sur la musique, la physique, l'anatomie et la mécanique, Vaucanson est le premier à présenter en 1738 un automate musicien, le *flûteur automate*. L'université de Waseda propose aujourd'hui le robot flûtiste WF-4 [COI⁺04] qui respecte de nombreuses contraintes physiques (figure 13).

4.5 Technique

La technique instrumentale est d'abord liée à la maîtrise de l'interface instrumentiste-instrument, une surface de contrôle faite de touches au piano, de clés à la clarinette, de cordes au violon, de pistons à la trompette, *etc.* Un musicien doit souvent travailler techniquement à dépasser les difficultés liées à sa physiologie ou à la physique de son instrument, comme avec la position de la main et du poignet au piano par exemple. Il doit maîtriser les gestes instrumentaux pour contrôler la production du son. Le geste instrumental est la manipulation et le jeu technique d'un instrument. Les paramètres du geste sont par exemple la vitesse d'un filet d'air, l'emplacement d'un doigt ou la pression d'un archet sur une corde. Les variations de ces paramètres ont un effet sur le timbre et sont clairement

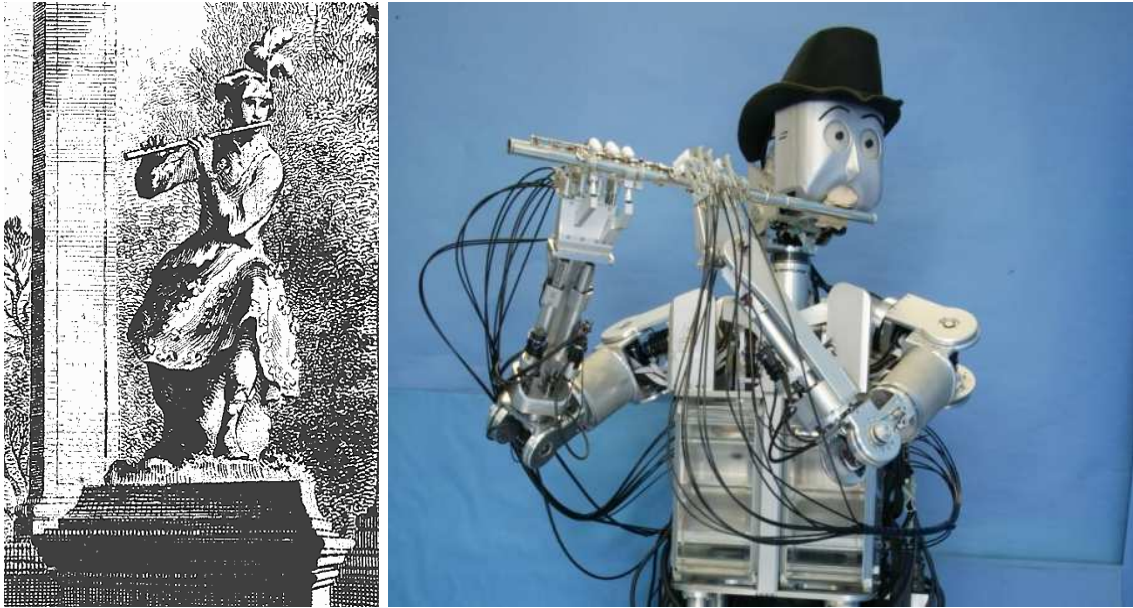


FIG. 13 – *Les robots musiciens.* À gauche, une illustration du flûteur automate¹, le premier automate musicien, présenté par Vaucanson en 1738 à Paris. À droite, le robot flûtiste WF-4², développé à l’université de Waseda, au Japon.

perçus par un auditeur entraîné, comme un professeur de musique. De nombreuses années de développement du comportement moteur sont nécessaires pour contrôler l’instrument de façon à produire intentionnellement des sons d’une certaine qualité, ou avec un certain timbre.

Il est possible de retrouver à partir de l’enregistrement d’une performance technique non expressive des éléments techniques du jeu instrumental. Mais les paramètres physiques présentés dans la section précédente influent également sur la performance. En observant l’évolution des paramètres de la performance plutôt que leurs valeurs instantanées, il est possible d’isoler les paramètres techniques de la performance.

Traube et Smith [TS01] proposent de retrouver le doigté et les points de pincement des cordes à la guitare, à partir d’un enregistrement audio. Puis Traube *et al.* [TDW03] ajoutent une approche multi-niveaux pour l’extraction de paramètres du geste instrumental, fondée sur une connaissance de la physique de l’instrument d’une part, et sur la connaissance de la perception du timbre de la guitare classique d’autre part. Dans le chapitre 5, nous proposons d’utiliser les nombreux travaux de biomécanique traitant du doigté et des gestes de la main au piano afin de retrouver le doigté utilisé par un interprète, à partir de sa performance pianistique.

¹http://www.bernard-pin.com/agrand/agr_auto_fluteur.htm

²<http://www.takanishi.mech.waseda.ac.jp/research/index.htm>

4.6 Expressivité

Le paramètre d'expressivité dans la performance musicale est très étudié, parce qu'il est considéré comme le facteur artistique de la performance. Suivant les méthodes utilisées, son étude permet de reconnaître des musiciens à partir de leur interprétation, ou bien d'humaniser la lecture d'une partition par un ordinateur.

4.6.1 Différenciation et reconnaissance d'interprètes

Les travaux de Langner et Goebel [LG03] ou Scheirer [Sch98] expliquent comment différencier deux performances pianistiques en utilisant les déviations temporelles et l'intensité dans la performance. Une modélisation individualisée de la performance expressive est alors possible, où les paramètres capturent le style de l'interprète. Stamatatos et Widmer [Sta01, SW02] utilisent également les déviations de la performance de pianistes par rapport à une partition pour décrire leur style, et proposent ensuite de les identifier dans une autre performance.

4.6.2 Humanisation d'une performance automatique

Les différentes modélisations de la performance expressive introduites dans la section 4.1 peuvent être utilisées à des fins de synthèse sonore réaliste. Il peut s'agir suivant les cas d'humaniser automatiquement une performance, de jouer une partition dans le style d'un célèbre interprète, ou d'appliquer des règles d'analyse musicale pour interpréter une partition.

Dixon *et al.* [DGW05] proposent une interface pour manipuler la performance musicale expressive en temps réel. Les utilisateurs peuvent contrôler les variations de tempo et d'intensité pour produire des interprétations expressives. Purbrick [Pur00] présente *EASY (Expression Articulation SYnthesizer)*, qui permet la synthèse de performances expressives en variant l'articulation de la pièce instrumentale. Clynes introduit le *SuperConductor system* [Cly98], qui accompagne l'utilisateur dans l'interprétation d'une partition. Cela permet de produire une performance expressive sans avoir à contrôler techniquement l'interface d'un instrument.

Tanguiane [Tan92] affirme que la performance peut être décrite par des règles d'analyse musicale. Une simulation de la performance humaine fondée sur l'analyse de la partition à jouer est alors proposée. Johnson [Joh91] a travaillé avec des interprètes professionnels pour établir des règles automatisées d'interprétation. Le *SasEx system*, proposé par Arcos *et al.* [AdMS97], utilise une technique de *Case-Based Reasoning* (voir par exemple Aamodt et Plaza [AP94]). Cela consiste à comparer de nombreuses performances musicales à une partition afin d'établir des règles, par apprentissage. Les règles peuvent ensuite être appliquées à d'autres partitions, si le contexte est proche du contexte d'apprentissage (même instrument, même style). La méthode de synthèse associe des techniques de transformation spectrale dans le modèle SMS (*Spectral Modeling Synthesis*) [Ser97] et des connaissances musicologiques comme le préconise Narmour [Nar90].

Les règles expressives du KTH présentées dans la section 4.1 sont implémentées dans le programme *Director Musices* [BFS02]. Ce programme permet de paramétrer les effets de ces règles, pour interpréter une partition et humaniser la performance musicale. Friberg *et al.* [FBM⁺03] proposent d'utiliser le *Radio Baton* [BM97], piloté par le logiciel

Conductor [Mat89], pour contrôler gestuellement les paramètres majeurs de la synthèse expressive (principalement le tempo et les nuances). Le logiciel *Director Musices* gère alors les paramètres expressifs plus fins. Bresin et Friberg [BF99, BF00b, BF00a] se concentrent sur l'émotion que peut engendrer l'écoute de la performance. Ils proposent un paramétrage des règles d'interprétation expressive du KTH dans le logiciel *Director Musices* pour synthétiser des performances tristes, gaies ou coléreuses !

4.7 Éléments de pédagogie musicale

Cette section introduit des éléments de pédagogie instrumentale dans un contexte académique. Afin de trouver les méthodes les plus adaptées pour l'analyse de la performance musicale, il est en effet intéressant de connaître les différentes pratiques instrumentales, et les méthodes existantes d'évaluation.

4.7.1 Contexte académique

Louis XIV crée en 1669 l'Académie royale de musique. Rattachée à la maison du roi, elle devient plus tard, d'une part le premier conservatoire national supérieur de musique et, d'autre part, l'Opéra de Paris. C'est en France le début de l'enseignement académique.

Au fil de l'histoire de la musique, les compositeurs sont de moins en moins interprètes, l'interprétation devient alors un art à part entière. Les époques et les styles passant, les techniques instrumentales s'enrichissent. Les conservatoires sont alors garants de la conservation de toutes les techniques instrumentales. L'enseignement académique construit des bases techniques et expressives, telles que chaque instrumentiste doit savoir interpréter une œuvre classique. La maîtrise exigée de la justesse, du timbre du son, des nuances, amènent tous les instrumentistes académiques à pratiquer des exercices instrumentaux de base très similaires.

4.7.2 Exercices instrumentaux de base

Les exercices instrumentaux de base, communs à tous les instruments dans un contexte académique, sont généralement :

- *Les sons tenus*. Il s'agit de contrôler les paramètres sonores produits par l'instrument (justesse, nuance, ...) en maintenant longtemps le jeu d'une note. C'est un exercice technique pratiqué par les instruments entretenus.
- *Les gammes, ou les arpèges*. L'exercice de base de l'apprentissage académique. Les gammes permettent notamment de travailler sur l'interaction avec l'instrument. C'est un exercice gestuel technique, où la régularité et la rapidité d'exécution sont éprouvés.
- *Les exercices d'articulation, les exercices mécaniques*. Ils consistent à maîtriser des aspects techniques de l'instrument, par une répétition d'exercices courts et spécifiques. L'articulation d'un passage musical est l'ensemble des transitions entre les notes : principalement les attaques (début de note) et la gestion des micro-silences entre les notes. Les exercices mécaniques imposent la répétition de petites cellules de notes d'une difficulté technique exigeante.

- *La lecture à vue.* Il s’agit de jouer pour la première fois une partition, et de l’interpréter. Les paramètres de la performance musicale sont donc musicaux (maîtrise du solfège), physiques, techniques et expressifs.
- *Les études techniques.* Ce sont des pièces musicales composées pour travailler techniquement des points particuliers. Elles peuvent être interprétées de manière très expressive.

Ces différents exercices sont complétés par l’interprétation de pièces instrumentales expressives. Il est intéressant de noter la particularité de chaque exercice, dont la performance peut apporter des informations très différentes et complémentaires sur l’interprète. Les gammes et les arpèges sont intéressants pour étudier par exemple l’influence de la technique de doigté sur la performance musicale. Ces exercices ont en général un doigté conventionnel, et ils sont très régulièrement pratiqués par les instrumentistes. De plus, quel que soit le niveau technique d’un instrumentiste, il est confronté à une limite technique de vitesse d’exécution observable, qui lui est propre. Ce sont donc des exercices très adaptés à l’évaluation technique.

4.7.3 Niveau technique et expressivité

Les influences techniques et expressives dans la performance peuvent se confondre. Dans les premiers mois d’apprentissage de son instrument, l’élève se concentre surtout sur sa technique, et les informations analysées à partir de la performance musicale sont donc essentiellement techniques. Puis les défauts techniques s’estompent progressivement – sans jamais disparaître totalement – et l’expressivité de l’interprète devient le paramètre principal de la performance.

Dans un contexte académique, le niveau général des élèves instrumentistes se divise en classes de niveau technique qui peuvent se nommer : *préparatoire* pour les premières années, *élémentaire* et *moyen* pour des niveaux intermédiaires de deux ou trois années chacun, *fin d’études* qui termine le premier cycle d’apprentissage, essentiellement technique, *supérieur* qui est un niveau d’exigence très élevé, et *spécialisé* pour les interprètes professionnels. Chaque niveau a son exigence technique et expressive, le passage du niveau *fin d’études* étant considéré comme un pallier technique très important. L’expressivité de l’interprète se construit naturellement lentement dans le temps, mais ne peut éclore réellement qu’avec un niveau technique sûr.

4.7.4 Évaluation en pédagogie musicale

L’évaluation est une partie essentielle de l’enseignement instrumental académique. Pour comprendre et éventuellement simuler automatiquement l’évaluation humaine de la performance instrumentale, nous commençons par exposer les méthodes traditionnellement employées dans les écoles de musique. Nous expliquons pourquoi cette évaluation est en général relative, et nous présentons des méthodes existantes d’évaluation automatique de la performance dans un contexte d’apprentissage instrumental.

Méthodes d’évaluation

Les conservatoires et écoles de musique organisent annuellement des examens de fin d’année pour évaluer leurs élèves. Les élèves y participent la plupart du temps chaque

année, parfois uniquement en fin de cycle (un cycle est une classe de niveau d'une durée de deux ou trois ans). Cet examen consiste en général à jouer une ou plusieurs pièces instrumentales, il est évalué par un jury. L'évaluation de fin d'année est jugée sur des critères physiques, techniques et expressifs. Le paramètre physique évalué concerne en général la qualité de l'instrument joué. Le jury peut également relever une position inappropriée du musicien pour l'exécution instrumentale. Lors de l'examen, les erreurs techniques sont détectées (rythme, justesse, ...), et l'expressivité est aussi évaluée (dynamique de nuance, vibrato, ...). La notion complexe de timbre est également décisive, mêlant propriétés physiques et techniques de l'instrument. Le jury essaye d'apprécier chaque aspect de la performance instrumentale indépendamment, et peut estimer par exemple que :

- “l'élève est intéressant, mais il doit changer d'instrument pour progresser” ;
- “l'élève est techniquement doué, mais son expressivité est pauvre” ;
- “la performance a des défauts techniques, mais l'élève est très expressif”.

Dans les conservatoires, l'instrument est souvent exigé d'excellente qualité, et l'élève se soumet à un examen technique pendant l'année. L'examen de fin d'année peut alors se concentrer sur le paramètre expressif de la performance.

Wrigley [Wri05] présente d'autres méthodes d'évaluation de la performance musicale, beaucoup moins académiques. Ainsi le test de *Watkins-Farnum Performance Scale* (WFPS) [WF54] est un test pendant lequel les instrumentistes doivent jouer des passages musicaux de plus en plus difficiles. Les conclusions de l'évaluation sont données en fonction du nombre d'erreurs commises, et de la difficulté technique atteinte. Sur le même principe, il existe aussi un test pour les clarinettes (*Clarinet Performance Rating Scale* [Abe73]), pour les tubas (*Euphonium-Tuba Performance Rating Scale* [Ber89]), et bien d'autres tests spécifiques à chaque instrument.

Évaluation relative

Nous étudions ici la question de la relativité de l'évaluation en général, et celle d'une performance musicale en particulier. Le fonctionnement de l'évaluation dans les milieux scolaires et universitaires montre que le classement des élèves prime la plupart du temps sur la note qu'ils obtiennent. La notation est alors relative au niveau général de la classe, aux meilleurs éléments qui la composent. L'évaluation instrumentale procède généralement de la même façon. Un élève est bien sûr évalué selon ses qualités intrinsèques, mais il est surtout classé parmi les élèves de sa classe instrumentale.

L'évaluation relative intègre les difficultés techniques spécifiques à chaque instrument. Si un saxophoniste a plus de difficultés pour jouer des notes graves avec son instrument, cela influe sur sa performance, mais son classement parmi la classe ne sera pas altéré : la difficulté est la même pour tous les saxophonistes. Il en est de même pour la justesse au violon par exemple. L'évaluation relative permet donc à une méthode d'analyse d'être généralisable à plusieurs instruments.

Évaluation automatisée

L'évaluation automatique d'une performance tente généralement d'être fidèle à ce qu'un professionnel de l'instrument analysé aurait perçu. L'objectif est souvent de proposer dans un contexte pédagogique des outils de tutorisation virtuelle. Ainsi, de nombreux contenus multimédia tentent d'automatiser l'enseignement instrumental. Le principal défaut de

ces méthodes est de ne pas offrir de retour à l'élève instrumentiste sur sa performance, l'élève décide alors de changer de leçon quand il estime avoir réussi. Sans analyse de la performance, ces méthodes ne peuvent pas être individualisées.

Certaines méthodes, comme *Piano Tutor* [DSJ⁺90, DSJ⁺93] ou la méthode présentée par le projet *IMUTUS* [SHA04, FLO⁺04], proposent des environnements interactifs avec retour pour l'instrumentiste sur sa performance musicale après analyse de son interprétation sonore. Le projet *i-Maestro* (*Interactive Multimedia Environment for technology enhanced music education and creative collaborative composition and performance*) [Con05] vise également au développement de solutions innovantes pour l'éducation musicale, dans le domaine théorique aussi bien que pratique. Il souhaite développer de nouveaux environnements d'auto-apprentissage interactifs.

Nous remarquons cependant que ces logiciels s'adressent en priorité à des instrumentistes de faible niveau technique, puisqu'il s'agit par exemple de savoir si l'interprète s'est trompé de note ou de rythme. Les erreurs détectées sont grossières. Nous pouvons imaginer dans une même démarche détecter des erreurs de plus haut niveau technique : c'est ce que nous proposons pour les instruments entretenus dans le chapitre 6.

Chapitre 5

Le doigté au piano

Nous présentons dans ce chapitre nos recherches sur le doigté au piano. Ce domaine est transverse puisqu'il concerne au moins la musicologie, la biomécanique, et l'informatique. Nous nous sommes attachés dans un premier temps à étudier les différents travaux existant dans ces domaines. À partir de la reproduction de quelques expériences biomécaniques sur le piano, nous présentons nos travaux préliminaires sur les méthodes de doigté automatique de partition et de reconnaissance du doigté à partir de la performance pianistique.

Après une brève présentation du piano dans la section 5.1, nous soulignons dans la section 5.2 l'importance et l'influence du choix du doigté dans la performance pianistique. Nous exposons ensuite, section 5.3, les différents travaux de biomécanique qui analysent cette influence dans la performance. Nous proposons alors un algorithme pour trouver le doigté d'une mélodie (section 5.4), et nous montrons dans la section 5.5 comment retrouver le doigté utilisé par un pianiste à partir de sa performance.

5.1 Le piano

Par l'étendue de son répertoire et par son histoire musicale, le piano est souvent considéré comme le roi des instruments. C'est un instrument de musique à clavier et à cordes frappées, classé parmi la famille des cordes. Le son est produit par les cordes tendues sur un cadre rigide, au-dessus de la table d'harmonie. Elles sont frappées par des marteaux couverts de feutre, actionnés par l'enfoncement des touches du clavier. La vibration des cordes est stoppée par un étouffoir lorsque la touche du clavier est relâchée.

Closson [Clo44] retrace l'historique du piano. Créé en 1709 par l'italien Cristofori sous l'appellation *piano-forte*, le piano naît de l'évolution du clavicorde et du tympanon. Vandervellen [Van95] indique les grandes étapes de la facture du piano. Pour notre étude sur le doigté, nous nous concentrons sur le clavier, dont les évolutions jusqu'au clavier moderne sont détaillées par Haury [Hau99]. Le clavier standard du piano moderne est composé de 88 touches (sauf exceptions, comme les claviers d'étude par exemple), séparé en 52 touches blanches et 36 touches noires.

5.2 Le doigté au piano

Le doigté consiste dans l'art de choisir la séquence de doigts à utiliser pour jouer une série de notes. Il y a au piano de nombreuses possibilités de combinaisons, ce qui contraste avec la plupart des autres instruments. Pour les instruments à vent par exemple, les clés ou pistons à actionner sont en général associés à un doigt particulier. Dans le cas du piano, chacun des doigts peut potentiellement appuyer sur n'importe laquelle des touches du clavier, il n'y a pas de doigté standard pour jouer une note. La notation du doigté au piano associe le nombre 1 au pouce de chaque main, et numérote successivement les doigts en partant du pouce jusqu'à 5, comme illustré par la figure 14.

Pour sauter d'une note à une autre, la main se déplace sur le clavier selon des trajectoires courbées et asymétriques. D'après Parncutt [PT02], le doigté optimal est alors un compromis entre les contraintes de nature physique (structure du clavier), anatomique (structure de la main), motrice (coordination des doigts), cognitive (mémoire pour les patrons complexes), et interprétative (structure, émotion). Un bon doigté permet à l'interprète de jouer un passage le plus confortablement possible, permettant d'interpréter ce passage avec la plus grande aisance.

Si les éditeurs de musique indiquent parfois un doigté pour certains passages, Debussy ne mettait pas de doigté dans la musique qu'il éditait. Comme il indique dans la préface de ses douze études pour piano, c'est à l'interprète de découvrir quel doigté lui convient le mieux, et non à l'éditeur de partitions d'influencer le jeu du pianiste par un doigté suggéré. Excepté dans les livres pour débutants, les doigtés de base sont généralement évidents et ne sont donc pas indiqués.

Clarke *et al.* [CPRS97] ont interrogé des pianistes professionnels sur les paramètres qui les amènent au choix d'un doigté particulier. Les paramètres étudiés sont les considérations techniques, l'influence de l'interprétation, les indications de doigté fournies par l'éditeur, le lien avec les conditions d'exécution et le poids de l'enseignement. Les interprètes estiment qu'un doigté se fonde en premier sur l'interprétation. Martin [Mar01b] donne des exemples de doigtés pour illustrer ce fait. Les pianistes utilisent les doigtés types (fournis par l'éditeur par exemple) essentiellement pour *déchiffrer* (jouer pour la première fois) un morceau, et les modifient ensuite pour personnaliser leur interprétation.

5.3 Biomécanique et piano

De nombreuses études de biomécanique se sont attachées à décrire le fonctionnement des doigts, de la main et du bras du pianiste. Cet aspect physique a une influence sur la performance, à travers le doigté notamment. Ainsi les travaux de Ortmann [Ort62], plus récemment ceux de Wagner [Wag88], Bejjani *et al* [BFX⁺89] ou Harding *et al.* [HBH93], s'intéressent aux mouvements de la main et des doigts du pianiste. Bros et Papillon [BP01] proposent une méthode d'éducation posturale progressive pour les pianistes en fonction de cette connaissance.

Les travaux de biomécanique de Ortmann [Ort62] et McKenzie et VanEerd [KV90] montrent l'influence que peut avoir le choix du doigté sur la performance musicale, en analysant la performance en fonction des doigts utilisés. Ils utilisent notamment des exercices non expressifs à base de gammes, où le doigté est conventionnel. Un exemple de doigté conventionnel pour la gamme de Do majeur au piano est donné par la figure 15.

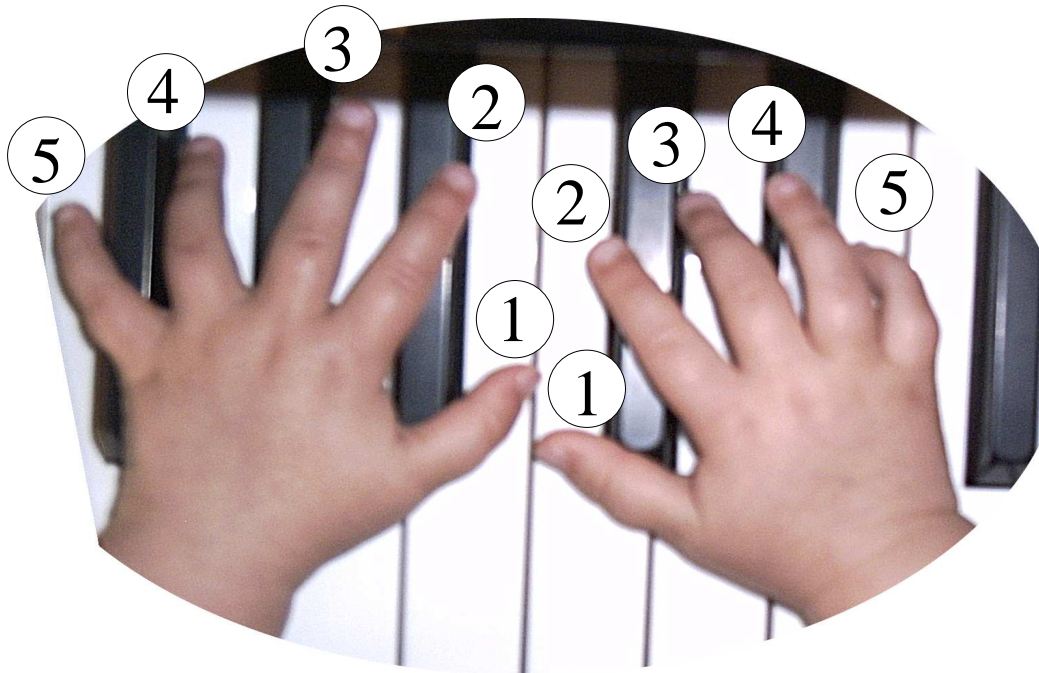


FIG. 14 – Numérotation des doigts au piano. La numérotation commence par le pouce, doigté 1, pour la main droite comme pour la main gauche.

Ortmann a photographié les trajectoires des doigts durant l'exécution des gammes, comme le montre la figure 16. Nous pouvons constater qu'il y a des motifs répétés dans le mouvement des doigts. McKenzie et VanEerd ont également réalisé des expériences avec des pianistes jouant des gammes à différents tempos et pour les deux mains. Une partie de leurs résultats est illustrée par les figures 17 et 18. En fonction des doigts utilisés, et souvent lors du passage du pouce, il y a des irrégularités dans les intervalles de temps entre les notes (intervalle qui devrait être constant pour une gamme) et dans les nuances (mesurées par McKenzie et VanEerd en vélocité MIDI). Les déviations observées dans la performance d'un passage technique rapide caractérisent les doigts utilisés : il y a une sorte d'empreinte de la main dans la performance.

5.4 Doigté automatique

Nous cherchons à trouver un doigté possible pour jouer une séquence de notes. Des recherches ont précédemment été menées par Sayegh [Say89] et Radicioni [RAL04a, RAL04b, RL05a, RL05b] pour la guitare, ou par Parncutt *et al.* [PSC⁺97] pour le piano, dont la méthode a été raffinée par Jacobs [Jac01]. Les différents modèles présentés permettent de générer une suite de positions de doigts à partir d'une mélodie sur une partition. Des pénalités sont attribuées en fonction de la difficulté de chaque transition élémentaire entre deux notes. Le doigté résultant est la combinaison la plus confortable des positions successives pour les doigts.

Le modèle proposé par Parncutt *et al.* [PSC⁺97] prend en considération les contraintes

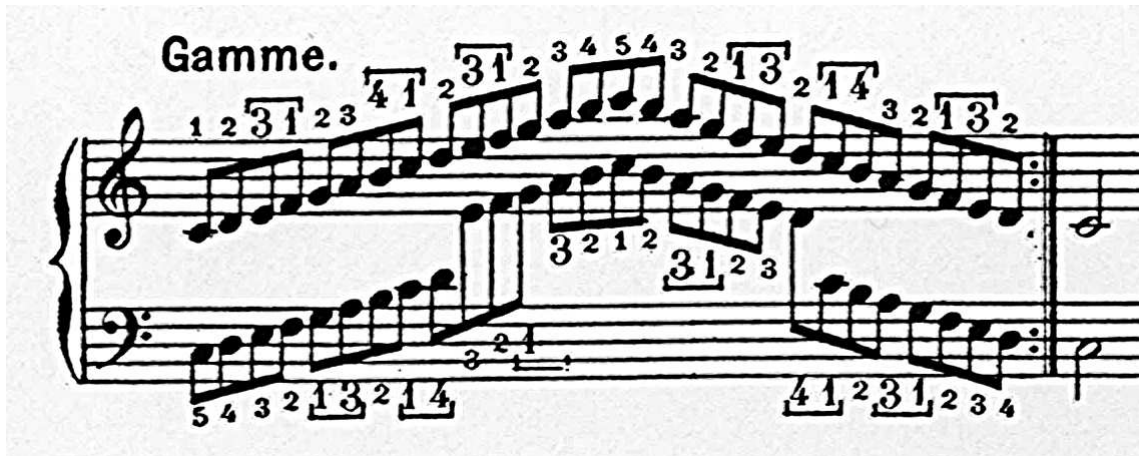
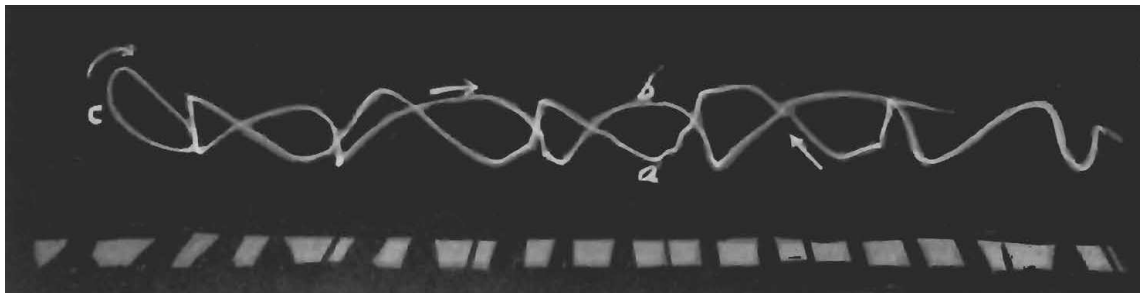
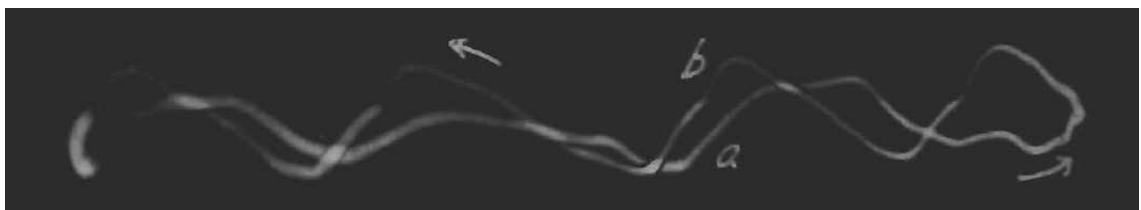


FIG. 15 – Gamme de Do majeur pour le piano, avec indication du doigté conventionnel (extrait de la méthode Descaves [Des50]).

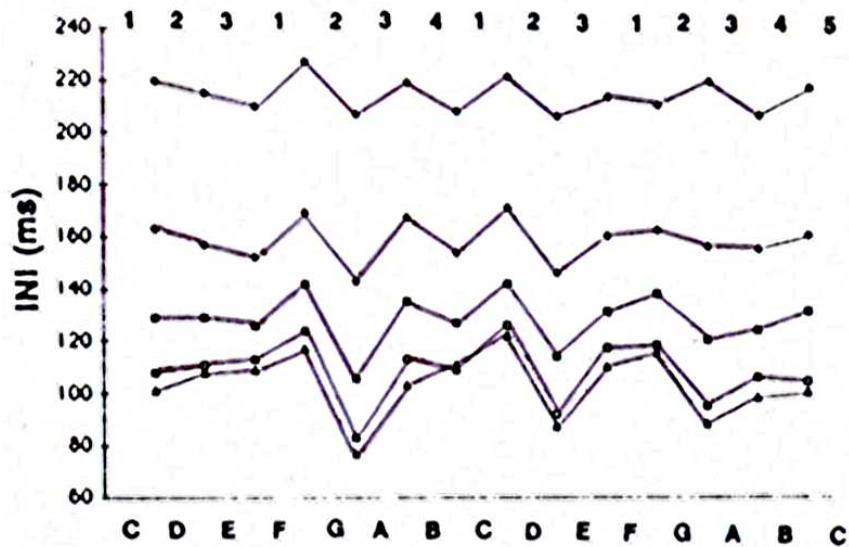


(a)

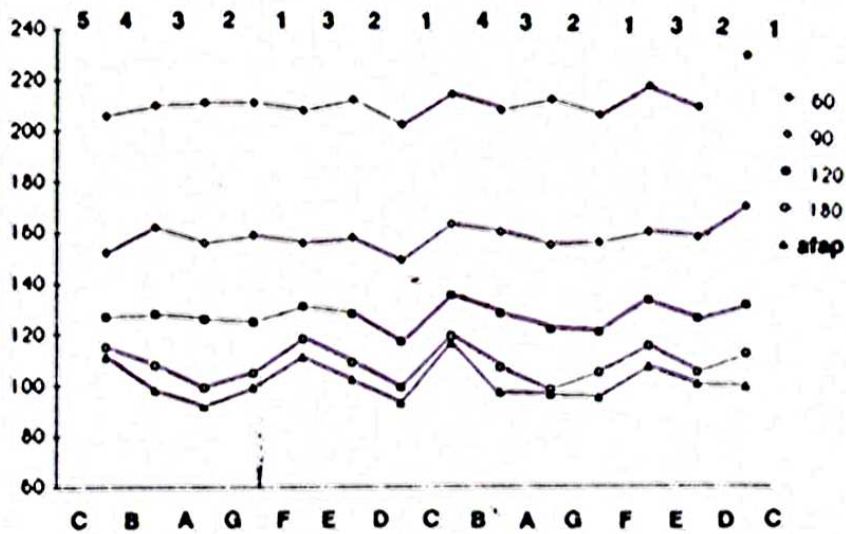


(b)

FIG. 16 – Travaux biomécaniques de Ortmann [Ort62] sur le piano. Une lumière a été attachée au doigt d'un pianiste et les trajectoires sont photographiées. (a) Trajet du 3ème doigt de la main droite dans une gamme ascendante, puis descendante, avec un tempo allegro (4 notes par pulsation pour environ 168 pulsations par minute). Le photographe est face au pianiste. (b) Mouvement du pouce dans une gamme rapide, ascendante et descendante. Le photographe est derrière le pianiste, sur sa gauche. Nous remarquons que chaque doigt a une trajectoire propre, avec des motifs répétés dans son déplacement.



(a)



(b)

FIG. 17 – Travaux biomécaniques de McKenzie et VanEerd [KV90] sur le doigté au piano. L'intervalle moyen entre les notes en ms est donné en fonction du tempo et de la note, pour une gamme de Do majeur (a) ascendante à la main droite et (b) descendante à la main droite. Le doigté est indiqué en haut, et l'unité de tempo est le nombre de pulsations par minute. Les différents tempos utilisés sont : 60, 90, 120, 180 pulsations par minute et "aussi vite que possible" (afap). Les notes sont indiquées en notation américaine, avec une correspondance A-La, B-Si, C-Do, ... Nous constatons que les évolutions des différentes courbes sont proches, et que le passage du pouce (enchaînement 3-1 ou 4-1 en montant et 1-4 ou 1-3 en descendant) a un effet significatif sur les intervalles.

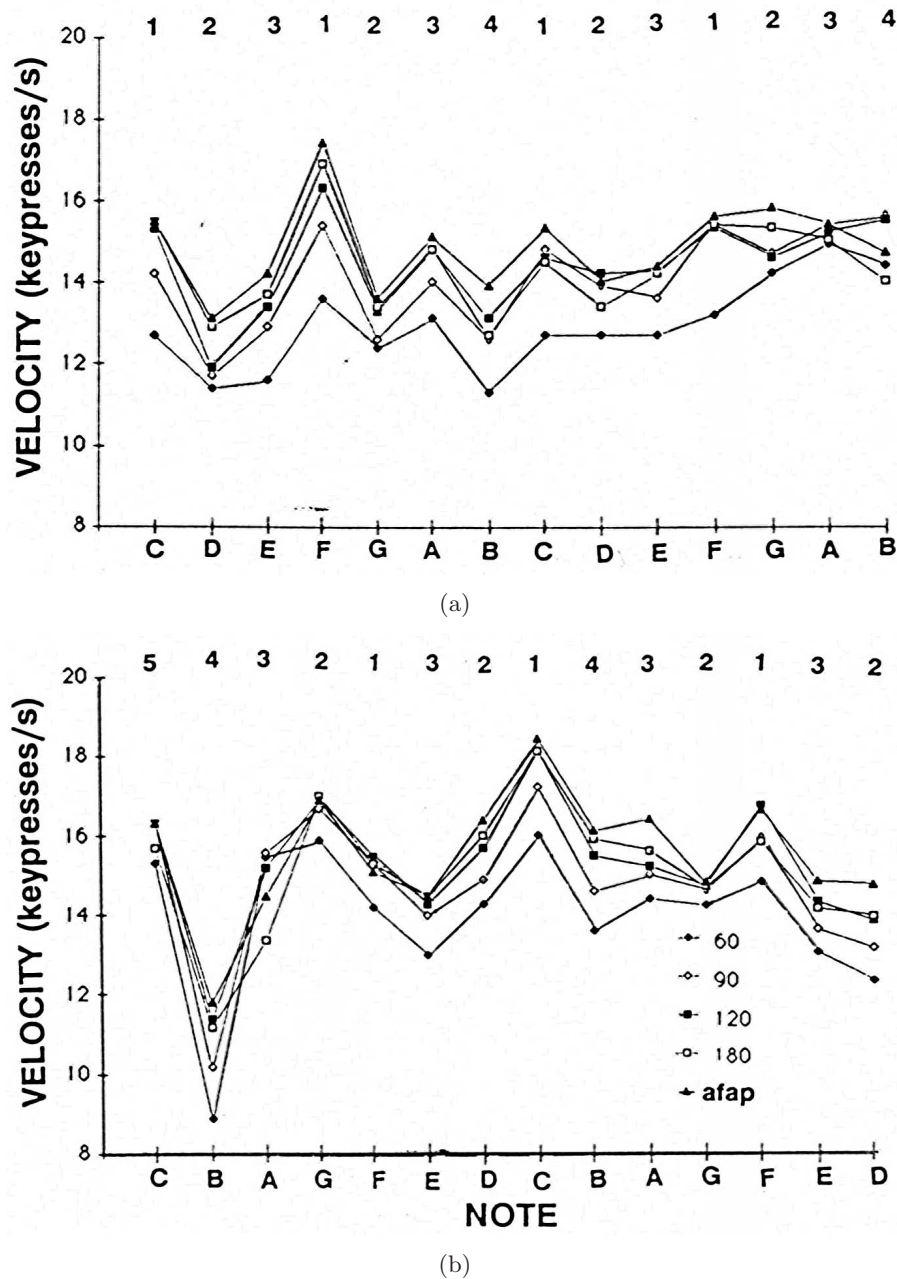


FIG. 18 – Travaux biomécaniques de McKenzie et VanEerd [KV90] sur le doigté au piano. La vitesse d'appui sur les touches (en rapport direct avec l'intensité du son) est donnée en fonction du temps, pour une gamme de Do majeur (a) ascendante à la main droite et (b) descendante à la main droite. Le doigté est indiqué en haut, et l'unité de temps est le nombre de pulsations par minute. Les différents tempos utilisés sont : 60, 90, 120, 180 pulsations par minute et "aussi vite que possible" (afap). Les notes sont indiquées en notation américaine, avec une correspondance A-La, B-Si, C-Do, ... Nous constatons que les courbes sont proches, et que le passage du pouce est souvent associé à un pic de vélocité.

ergonomiques rencontrées par les pianistes. Le doigté est généré en deux étapes : premièrement, tous les doigtés possibles sont énumérés, et deuxièmement, chaque suite de doigts est pondérée en fonction des difficultés qu'elle engendre. La pondération est fondée sur l'application de douze règles de bon doigté, participant chacune à une difficulté générale. Le doigté de coût minimal est alors le doigté conseillé par la méthode. L'algorithme de doigté pour la guitare proposé par Sayegh [Say89] utilise une représentation fondée sur les graphes, avec les doigts comme sommets, reliés par des arêtes dont le poids est associé à la difficulté de la transition entre notes. Trouver le meilleur doigté est équivalent à la recherche du chemin de coût minimal dans le graphe. Radicioni *et al.* [RAL04a, RAL04b, RL05a, RL05b] ont repris ces travaux en ajoutant notamment la caractérisation des contraintes biomécaniques qui servent à pondérer les arêtes du graphe.

Nous proposons également une méthode de doigté automatique d'un passage musical rapide et technique, fondée sur la programmation dynamique [Bel57]. Nous excluons donc les paramètres d'expressivité du doigté, conformément aux pianistes qui choisissent leur doigté uniquement sur des critères biomécaniques quand le passage est techniquement très difficile. Il s'agit de trouver le chemin de coût minimal dans un graphe où les sommets sont des doigts, comme Sayegh [Say89] le propose, mais en utilisant cette fois une méthode de programmation dynamique. À l'instar de Parncutt *et al.* [PSC⁺97], le modèle est restreint à l'étude du doigté de la main droite, pour le fragment d'une mélodie.

5.4.1 Transitions

Nous disposons d'une partition, ou du fichier MIDI correspondant. À partir de la suite de notes à jouer, nous proposons de construire la liste des transitions à effectuer successivement entre les touches du clavier. Les touches blanches du clavier sont notées b , les noires n . Un chiffre est associé à la transition, il représente le minimum du nombre de frontières de touche traversées sur le clavier entre les deux notes de la transition. Ce chiffre est positif quand l'intervalle est ascendant (vers la droite du clavier), négatif sinon. Ainsi un enchaînement (Do, Ré, Mi) amène à la création de la liste de transitions : (bb1, bb1), c'est-à-dire deux fois la transition d'une touche blanche à une autre touche blanche, avec un écart d'une touche sur le clavier. De la même façon, l'enchaînement (Do, Si bémol, La) génère la suite de transitions (bn-2, nb-1). La figure 19 illustre cette notation.

5.4.2 Matrices de pondération

Une transition entre deux notes peut être effectuée en utilisant différentes combinaisons de doigts. Nous voulons trouver le meilleur doigté possible pour un passage rapide donné. Il faut alors convenir de ce qui rend un doigté meilleur qu'un autre. Comme nous l'avons décrit dans la section 5.2, le doigté est fortement lié à la biomécanique de l'interprète. Certains enchaînements sont difficiles à réaliser, voire impossibles. Nous proposons dans un premier temps d'utiliser des matrices de transitions M , dont les valeurs sont indexées sur la fréquence d'utilisation des différentes transitions dans les doigtés conventionnels (par exemple le doigté des gammes, des arpèges). $M(1, 2)$ est alors le coût de la transition du doigt 1 vers le doigt 2. En convenant qu'il est impossible de réaliser au piano des intervalles supérieurs à un intervalle de douzième (intervalle de 19 demi-tons) sans saut de main, le nombre de matrices nécessaires est fini et limité.

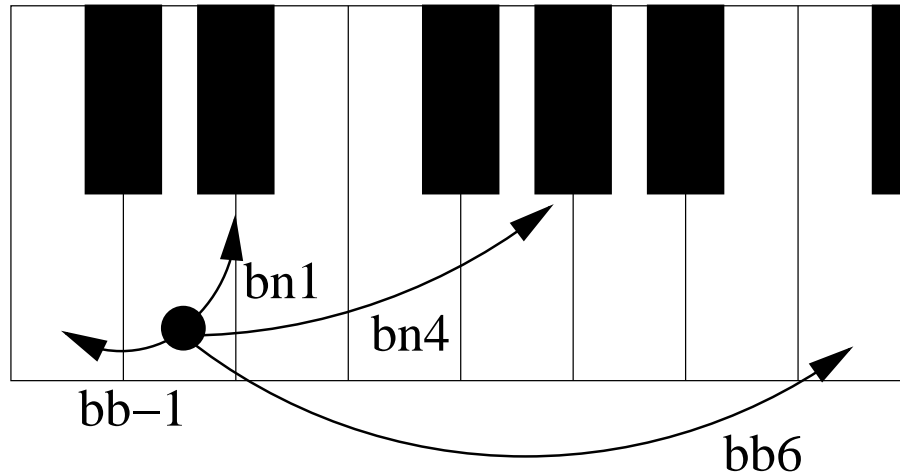


FIG. 19 – Notation pour les transitions entre les notes du clavier. Les notes blanches sont notées b , les notes noires n . L'écart entre les notes est indiqué par une valeur positive vers la droite du clavier, qui représente le minimum du nombre de frontières de touche traversées sur le clavier entre les deux notes de la transition. Ainsi $bb6$ représente la transition d'une note blanche à une autre, avec une distance de 6 frontières traversées, soit une transition ascendante Ré-Do par exemple.

M_{bb1} est la matrice de coût pour la transition d'une note blanche vers la note blanche suivante à droite du clavier :

$$M_{bb1} = \begin{pmatrix} 1000 & 1 & 2 & 10 & 100 \\ 100 & 1000 & 1 & 10 & 100 \\ 5 & 1000 & 1000 & 1 & 100 \\ 5 & 1000 & 1000 & 1000 & 1 \\ 1000 & 1000 & 1000 & 1000 & 1000 \end{pmatrix} \quad (22)$$

Les valeurs de M_{bb1} sont indicatives. La valeur 1000 est arbitrairement élevée et indique : "quasiment impossible". Ce coût est attribué pour la répétition d'un même doigt pour jouer les deux notes de la transition (diagonale de la matrice). Les doigtés les moins coûteux sont alors : 1-2, 2-3, 3-4 ou 4-5. De la même façon, nous pouvons donner un exemple pour M_{bn-1} :

$$M_{bn-1} = \begin{pmatrix} 1000 & 1 & 2 & 100 & 1000 \\ 1000 & 1000 & 10 & 1000 & 1000 \\ 1000 & 1 & 1000 & 1000 & 1000 \\ 1000 & 5 & 1 & 1000 & 1000 \\ 1000 & 50 & 20 & 5 & 1000 \end{pmatrix} \quad (23)$$

Nous avons utilisé des valeurs de matrice indicatives, fondées sur la connaissance des doigtés conventionnels, eux-mêmes issus de propriétés biomécaniques. Il est possible d'établir d'autres contraintes, afin d'établir par exemple le meilleur doigté qui n'utilise pas le doigt 3. Pour cela, il suffit de pondérer avec une très grande valeur toutes les transitions passant par ce doigt, et il ne sera pas présent dans le doigté optimal. La motivation de la pondération des matrices peut ainsi être variée. Elle peut être conventionnelle, mais

aussi individualisée par exemple, avec des poids adaptés à l'anatomie d'un interprète particulier. Les valeurs des matrices peuvent également être calculées par apprentissage, en fonction de la fréquence des doigtés réellement utilisés par les interprètes professionnels ou les professeurs. Le doigté optimal dépend donc des contraintes choisies pour pondérer les matrices.

5.4.3 Doigté dynamique

Nous souhaitons trouver automatiquement le meilleur doigté pour un passage musical de plusieurs notes d'une mélodie – sans accords, c'est-à-dire sans notes simultanées –, en utilisant les matrices de pondération présentées ci-dessus. Nous avons précédemment indiqué comment construire une suite de transitions à partir d'une suite de notes. Nous proposons d'utiliser la programmation dynamique présentée par Belmann [Bel57] pour déduire le doigté optimal à partir de ces transitions. La programmation dynamique permet de restreindre les possibilités de doigté à chaque étape de la méthode, et éviter ainsi une quantité exponentielle de possibilités non optimales. Cela permet de trouver le doigté dont le coût est minimal, en fonction des coûts élémentaires de chaque transition.

Le coût d'un doigté, c'est la combinaison des coûts élémentaires de chaque transition traversée. Le coût minimal dépend alors de la fonction de combinaison choisie. Nous avons dans un premier temps utilisé la fonction *somme*, le coût d'un chemin est alors la somme des coûts élémentaires, et le doigté optimal est celui dont la somme des coûts est la plus faible. Mais nous pouvons aussi bien choisir une autre fonction. La fonction *maximum* permet, par exemple, d'éviter un doigté dont toutes les transitions sont très faciles, mais avec une transition presque impossible à réaliser au milieu. Le doigté optimal est dans ce cas celui dont le maximum des coûts des transitions traversées est le plus faible (c'est une fonction *minmax*).

Nous construisons un graphe avec des numéros de doigts comme sommets, et des transitions pondérées comme arêtes. Une chaîne intéressante du graphe est donc une suite de doigts et des transitions telles que la série de notes voulue puisse être jouée. Voici la méthode proposée pour la construction du graphe, et la détermination du doigté optimal :

1. Nous démarrons la construction du graphe avec pour sommets origines 5 numéros de doigts distincts. Ce sont les doigtés possibles pour la première note.
2. Tant qu'il reste des transitions à effectuer, faire :
 - (a) Chaque sommet d'arrivée est relié par des arêtes à 5 nouveaux sommets, les 5 positions de doigts possibles pour la note suivante.
 - (b) Les nouvelles arêtes de transition sont pondérées en fonction de la transition à effectuer. Le coût de cette transition est estimé en fonction du doigt d'origine et du doigt d'arrivée. Ainsi, pour une transition bb1 du doigt 2 au doigt 3, le coût de la transition est donné par $M_{bb1}(2, 3)$.
 - (c) Nous conservons uniquement pour chaque sommet d'arrivée le chemin de coût minimal en provenance des sommets origines. Il n'y a plus que cinq chemins possibles.
3. Il y a 5 chemins possibles des sommets origines aux sommets d'arrivée. Ils peuvent être départagés en imposant le doigt utilisé pour la note de début ou la note de fin,

ou bien en choisissant finalement le chemin dont le coût est minimal. C'est le doigté optimal.

5.5 Analyse de la performance

Connaître le doigté particulier utilisé par un interprète, à partir de sa performance, offrirait des alternatives de doigté aux pianistes pour leur propre interprétation. Pour que l'information du doigté soit associé à la performance, Parncutt [Par95] propose d'utiliser la vidéo pendant la performance pour enregistrer les doigts utilisés, ou bien l'utilisation d'un gant, pour capter la pression des doigts. Mais ce dispositif étant peu utilisé, cette information n'est pas disponible pour la plupart des interprétations pianistiques. Nous proposons alors une méthode pour retrouver le doigté utilisé par un pianiste, à partir de l'analyse de sa performance.

5.5.1 Protocole d'enregistrement et analyse

Nous avons commencé par reproduire une expérience menée par MacKenzie et VanEerd [KV90]. Nous avons demandé à un pianiste de jouer des gammes avec la main droite, sans expressivité, à des tempos différents. Nous avons analysé les déviations temporelles observées entre la performance musicale et le rythme théorique, et les écarts d'amplitude.

Des données MIDI ont été enregistrées lors de la performance de gammes par un pianiste. Le clavier électrique utilisé disposait d'un *touché dynamique* (reproduction de la résistance des touches d'un piano acoustique), et d'une interface MIDI. Le pianiste a joué la gamme de Do majeur avec la main droite uniquement, sur 4 octaves ascendantes et 4 octaves descendantes, en suivant les doigtés conventionnels indiqués par la figure 15. La gamme de Do majeur a été jouée 5 fois par indication de tempo, pour 5 tempos différents, avec 4 notes jouées pour 60, 72, 84, 90 et 100 pulsations par minute. Le tempo a été donné avec un métronome, avec deux indications par pulsation (battue à la croche). Il y a eu 20 octaves ascendantes enregistrées par indication de tempo, et 20 autres descendantes. Nous avons établi des moyennes par octave ascendante, par octave descendante et par tempo à partir de ces 20 octaves.

Nous nous sommes intéressés à deux paramètres en particulier, le coefficient de *staccato* d'une part, et la vélocité (en rapport avec l'amplitude) d'autre part. Le coefficient de *staccato* est le rapport entre la durée d'une note et l'intervalle de temps entre le début de cette note et le début de la suivante. Ce paramètre est illustré, en fonction des doigts utilisés et du tempo demandé, par la figure 20. Nous constatons que les courbes de l'évolution octave par octave de ce coefficient en fonction du tempo sont très proches. Comme pour les travaux de McKenzie et VanEerd, illustrés par la figure 17, nous pouvons observer que le passage du pouce dans la gamme influe directement sur le coefficient de *staccato*, qui diminue pour la note jouée avant celle utilisant le doigté 1. Cela est dû à la préparation du passage du pouce, qui oblige à lever le doigt précédant le pouce assez tôt, pour permettre au pouce d'appuyer sur la touche du clavier au bon moment. Le deuxième paramètre observé est la vélocité, en fonction des doigts utilisés et en fonction du tempo. Les résultats de cette analyse sont présentés par la figure 21. Dans la gamme ascendante, l'utilisation du pouce est marquée par une augmentation de la vélocité, alors que c'est l'effet inverse dans la partie descendante de la gamme. Nous retrouvons alors une empreinte physico-technique

de la main de l'interprète dans la performance : empreinte physique, car elle est liée aux contraintes biomécaniques de la main, et empreinte technique, car un interprète atténue ces contraintes physiques par sa technique et son entraînement.

5.5.2 Reconnaissance du doigté

Nous proposons une méthode pour reconnaître le doigté utilisé dans une performance pianistique, pour des passages rapides et techniques. Cette méthode utilise la programmation dynamique, et s'appuie sur l'observation du coefficient de *staccato* et de la vitesse des notes.

Comme dans la section 5.4.3, nous construisons un graphe de doigté, avec des transitions comme arêtes et des numéros de doigts comme sommets. Les transitions sont déduites de la série de notes jouée par l'interprète. Nous commençons par générer tous les doigtés possibles pour la séquence des deux premières notes du passage. Un filtrage physique peut alors être effectué, excluant les transitions impossibles à réaliser (nous sommes dans un contexte d'enchaînement rapide de notes) : un enchaînement Do-Ré (bb1) ne peut pas être joué en utilisant le doigté 5-1 par exemple.

Nous proposons d'utiliser ensuite une formule de corrélation pour pondérer les transitions en fonction des similitudes entre les valeurs analysées et des valeurs acquises par apprentissage. Cette formule compare les valeurs du coefficient de *staccato* et de vitesse observées dans la performance et celles d'une base de données obtenue par apprentissage. Comme nous avons pu l'observer dans la section précédente, ces deux paramètres varient en fonction des doigts utilisés lors de la transition. Comparer des valeurs obtenues par apprentissage et les valeurs observées de ces deux paramètres peut nous servir à reconnaître la séquence de doigts utilisée pour deux notes d'une transition. Nous pouvons ainsi nous baser sur la différence de valeurs entre le coefficient de *staccato* observé et le coefficient théorique. Pour ne pas tenir compte du signe de la différence, nous proposons d'utiliser le carré de la différence. La différence de vitesse entre les deux doigts de la transition est également un paramètre important. Nous proposons de diminuer la valeur de la corrélation, en la multipliant par un coefficient $\alpha < 1$, si la différence de vitesse entre les deux notes de la transition est du même signe que celle attendue, et de la multiplier par un coefficient $\beta > 1$ sinon. La formule de corrélation C_t que nous proposons d'utiliser, avec t une transition donnée, peut alors s'exprimer ainsi :

$$C_t(d_1, d_2) = v_t(d_1, d_2) \cdot (cs - cs_t(d_1, d_2))^2 \quad (24)$$

où t dénote la transition, cs est le coefficient de *staccato* analysé, $cs_t(d_1, d_2)$ est le coefficient de *staccato* enregistré pour la transition t , du doigt d_1 au doigt d_2 , et :

$$v_t(d_1, d_2) = \begin{cases} \alpha & \text{si } dv_t(d_1, d_2) \cdot dv \geq 0 \\ \beta & \text{si } dv_t(d_1, d_2) \cdot dv < 0 \end{cases} \quad (25)$$

avec α et β des constantes telles que $\alpha < 1$ et $\beta > 1$, et $dv_t(d_1, d_2)$ la différence de vitesse observée dans la base de données entre les doigts d_1 et d_2 .

Cette formule de corrélation nous permet d'attribuer un poids à chaque arête du graphe. Cette étape dépend fortement de la formule de corrélation utilisée. La programmation dynamique est alors employée, comme dans la section 5.4.3, pour restreindre les possibilités

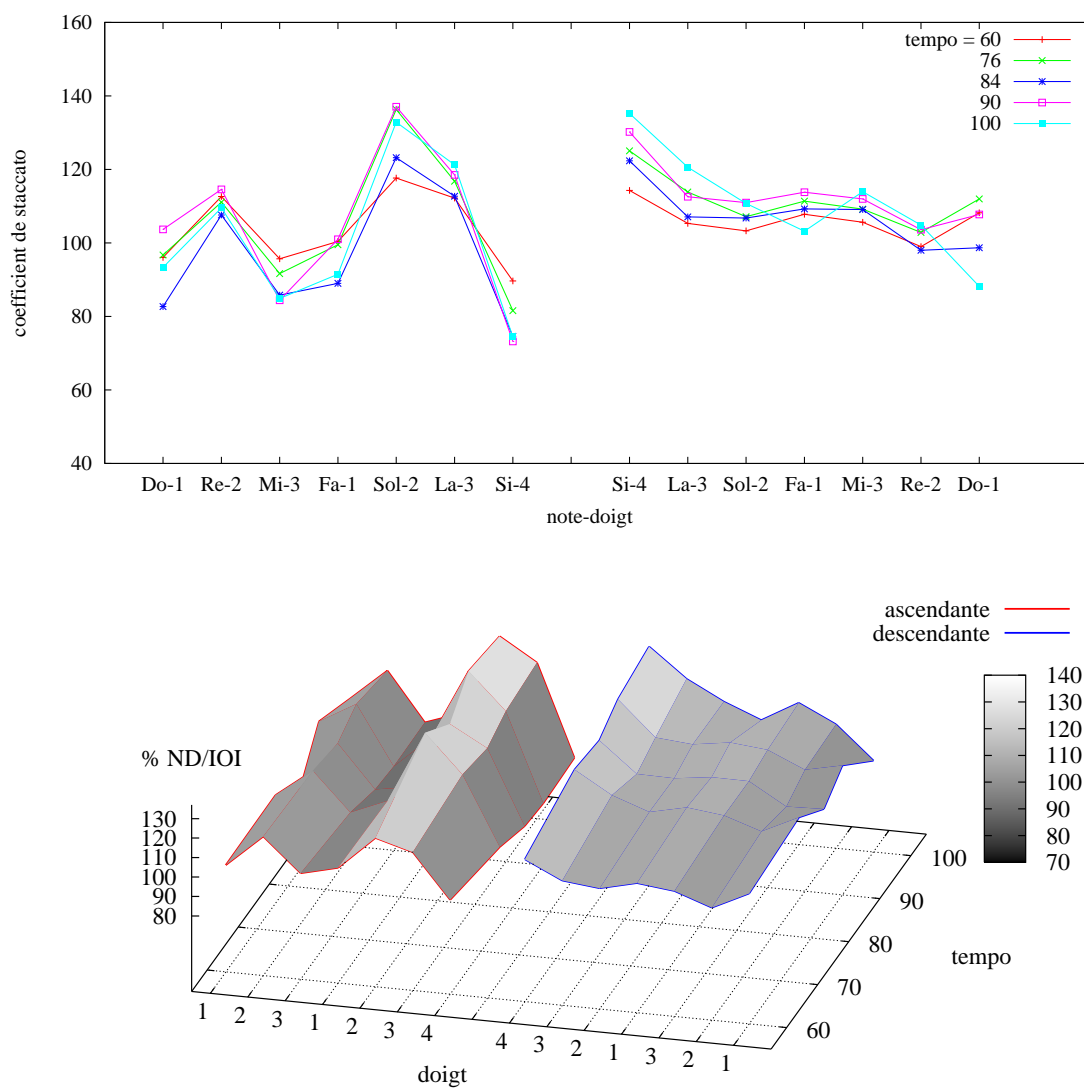


FIG. 20 – Coefficient de staccato pour une gamme de Do majeur jouée au piano à la main droite. Le coefficient est le rapport en pourcentage entre la durée de la note (ND) indiquée en abscisse et l'intervalle de temps séparant le début de cette note et la suivante (IOI). En haut, les résultats sont visualisés en 2 dimensions, et en bas en 3 dimensions. Les courbes de gauche indiquent les résultats par octave de la partie ascendante de la gamme, et celles de droite la partie descendante. Nous remarquons une baisse des valeurs du coefficient avant le passage du pouce (avant un doigté 1).

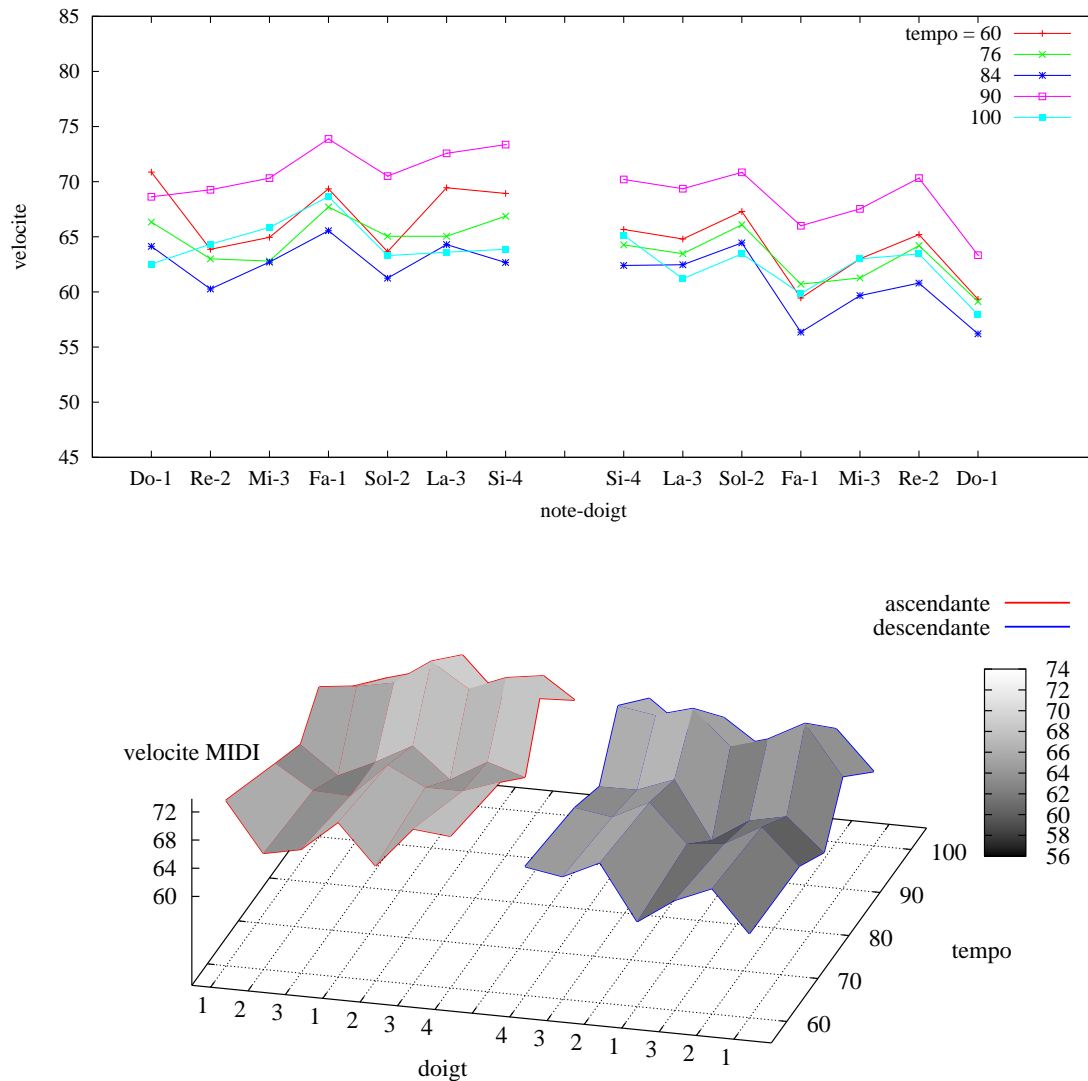


FIG. 21 – Valeurs de vélocité MIDI pour une gamme de Do majeur jouée au piano à la main droite. En haut, les résultats sont visualisés en 2 dimensions, et en bas en 3 dimensions. Les courbes de gauche indiquent les résultats par octave de la partie ascendante de la gamme, et celles de droite la partie descendante. Nous remarquons une augmentation de la vélocité pour les notes jouées avec le doigté 1 dans la partie ascendante de la gamme, et l'effet inverse dans la partie descendante.

de doigté à chaque étape. Cela permet de ne conserver que le meilleur chemin pour chaque sommet d'arrivée, en fonction des pondérations traversées précédemment. La même méthode est appliquée pour chaque transition, jusqu'à la dernière note jouée. À la fin de la méthode, le doigté proposé est le chemin optimal dans le graphe. Une étape de la méthode est illustrée par la figure 22.

5.5.3 Perspectives

À partir d'études menées dans différents domaines (biomécanique, musicologie, informatique), nous avons proposé des méthodes de doigté automatique et de reconnaissance de doigté. Cependant, ces travaux sont préliminaires, et ils mériteraient d'être mieux évalués. Nous n'avons pas encore assez de données – performances pianistiques non expressives – pour mener une évaluation approfondie.

Nous projetons donc de tester la méthode de doigté automatique et de comparer les résultats avec des doigtés conventionnels sur des pièces variées. Nous envisageons aussi de travailler avec différents interprètes pour tester notre méthode de reconnaissance de doigté sur des exercices techniques, et également sur des pièces instrumentales. La formule de corrélation, proposée dans la section 5.5.2, est pour le moment assez naïve. Elle pourrait ainsi être améliorée, par exemple en tenant compte des écarts types, en plus de la moyenne, des valeurs enregistrées par apprentissage.

Stamatatos et Widmer [SW02] proposent de caractériser les interprètes, pour pouvoir les identifier à travers leur performance. Peut-être que l'empreinte physico-technique mise en évidence dans cette partie est également un élément à prendre en compte pour reconnaître un interprète à partir de sa performance.

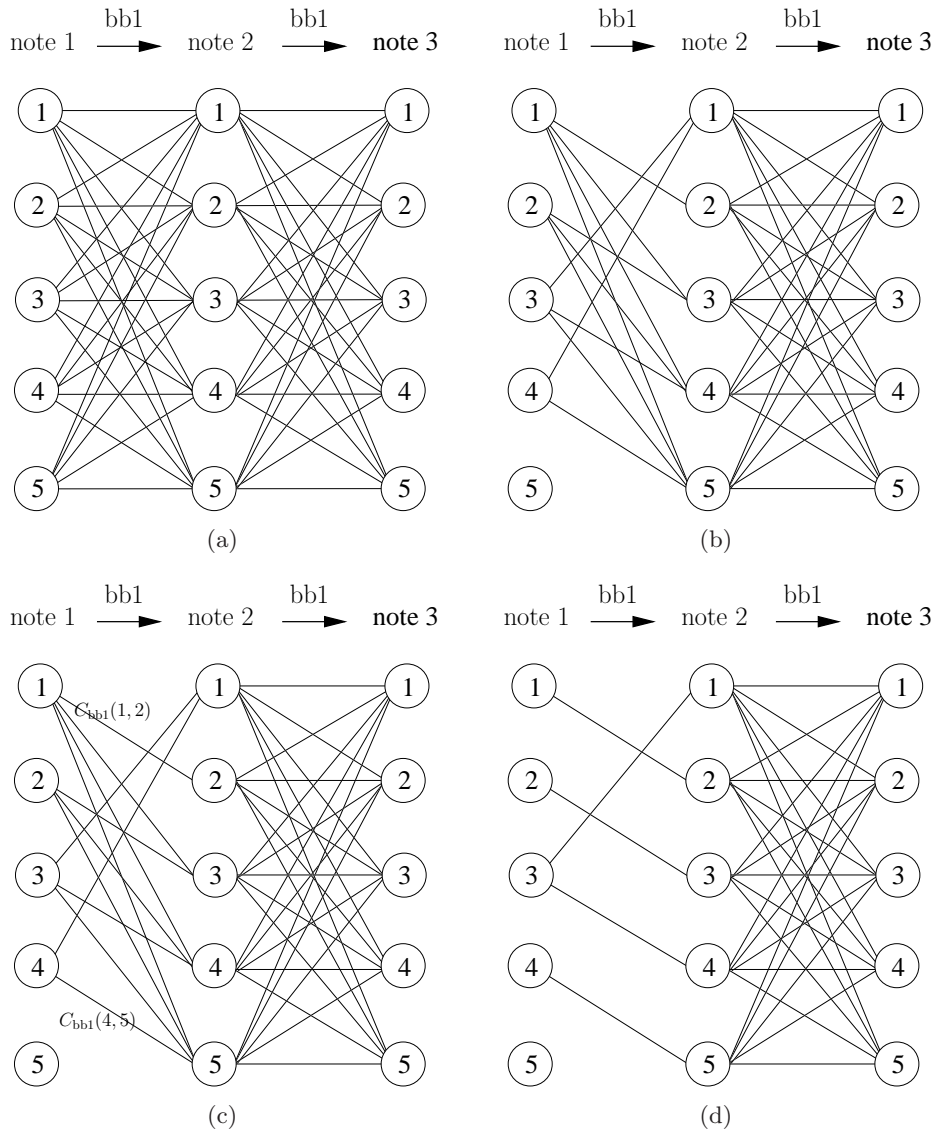


FIG. 22 – Méthode de reconnaissance de doigté à partir de la performance. (a) Doigté à trouver pour une séquence de 3 notes. Nous devons d'abord trouver le doigté utilisé pour l'intervalle bb1 entre les 2 premières notes. (b) Les transitions impossibles peuvent être supprimées pour le premier intervalle bb1 (d'après des éléments de biomécanique et / ou de pédagogie du piano). (c) Un poids est attribué à chaque arête par corrélation avec une base de données construite par apprentissage. Pour cet exemple nous n'avons indiqué que 2 valeurs de corrélation par souci de clarté, entre les doigts 1 et 2, et entre 4 et 5. (d) Résultat du processus de programmation dynamique : si la deuxième note est jouée avec le 4ème doigt, c'est probablement que la première était jouée avec le 3ème. Le même procédé est maintenant appliqué pour la transition de la note 2 à 3, et ce, jusqu'à la dernière note.

Chapitre 6

Évaluation de la performance au saxophone

Nous proposons une méthode pour évaluer le niveau technique d'un saxophoniste en analysant ses performances musicales. Nos résultats sont fondés sur l'étude du saxophone alto, bien qu'une méthode semblable puisse être employée pour évaluer tout instrument entretenu (autres instruments à vent, cordes frottées, ...). Notre travail repose sur une connaissance de l'instrument et de sa pédagogie dans un contexte académique, et il s'appuie aussi sur l'étude préalable menée avec Lagrange [RL06]. Nous proposons d'utiliser des exercices musicaux simples et courts qui permettent de différencier une large palette de niveaux techniques. Pour se concentrer sur l'aspect technique, nous excluons du protocole de la méthode les éléments expressifs du jeu instrumental. Par imitation de la pédagogie instrumentale, nous proposons que l'évaluation soit relative à un groupe d'élèves ou un niveau académique supposé. Cette méthode peut être étendue afin de tutoriser un élève, en lui proposant de corriger les défauts techniques identifiés par une évaluation automatisée.

De nombreux facteurs influent sur l'évolution des paramètres sonores de la performance musicale. Pour cette étude sur les saxophonistes, nous ne nous intéressons pas à l'influence physique sur la performance : ni à la physique de l'instrument, ni à la physiologie de l'interprète. Haas [Haa01] propose par exemple SALTO, un modèle physique de saxophone qui applique des contraintes physiques de l'instrument pour modeler le son, et Fuks [Fuk97] explique comment la composition de l'air expiré influe sur la hauteur d'une note. Le couple physiologie de l'interprète / instrument donne au son une partie de son timbre, une couleur, un état. Nous nous intéressons plutôt à l'évolution et au contrôle de cet état.

Pour nous concentrer sur la technique du musicien, nous analysons l'évolution des paramètres sonores au cours du temps. Le son analysé provient de la performance d'exercices techniques et non expressifs imposés. Les exercices choisis sont des sons droits et filés, vibrés ou non. Ce sont des exercices de base au saxophone, qui évaluent notamment la capacité à contrôler la pression d'air, l'attaque du son et la justesse. Même s'ils sont simples, qu'ils évaluent une partie seulement des capacités techniques d'un saxophoniste, la réussite de ces exercices est fortement corrélée avec le niveau technique général de l'instrumentiste dans un contexte académique. Nous montrons que l'étude de l'évolution des paramètres du son lors de ces exercices suffit à différencier un très large éventail de niveaux techniques, de l'instrumentiste débutant à l'expert.

Nous donnons d'abord dans la section 6.1 des éléments de pédagogie sur le saxophone, pour justifier le choix des exercices techniques ou des métriques utilisés par notre méthode d'évaluation. Après avoir décrit les conditions d'enregistrement des saxophonistes dans la section 6.2, nous expliquons comment nous analysons les fichiers audio dans la section 6.3. Nous introduisons alors les métriques utilisées (section 6.4), et nous donnons les résultats obtenus dans la section 6.5. Enfin, nous proposons une automatisation de la méthode d'évaluation (section 6.6), une extension vers une tutorisation automatisée, et la généralisation possible à d'autres instruments (section 6.7).

6.1 Éléments de pédagogie pour le saxophone

Nous donnons ici des éléments sur l'enseignement académique du saxophone, et sur l'évaluation de la performance technique pour cet instrument. Cela nous permettra de construire une méthode automatique d'évaluation technique de saxophonistes à partir d'une analyse sonore de l'instrument, où le protocole d'enregistrement et les métriques utilisées pour l'analyse seront adaptées à l'instrument.

6.1.1 Le saxophone

Adolphe Sax invente le saxophone en 1840. C'est un instrument à vent en cuivre, de la famille des bois en raison de son anche. Il est de forme conique parabolique, percé de trous que l'on peut obturer par des tampons de peau. Il se compose d'un pavillon évasé, d'une culasse et d'un corps dans lequel vient s'emboîter le bocal. Le bec portant une anche s'adapte sur le bocal. Le son est produit par une anche simple, languette de roseau fixée sous le bec par une ligature. C'est en battant contre le bec taillé en biseau au contact du souffle de l'instrumentiste, que se produit le son. Une illustration de l'instrument et de ses clés est donnée par la figure 23.

La famille des saxophones compte sept membres. De la note la plus grave du saxophone contrebasse à la plus aiguë du soprano, il y a une énorme échelle de hauteur (celle du piano moins une douzaine de notes). L'instrument comporte trois registres : grave, médium et aigu, sur un ambitus de deux octaves et une quinte. Dans de nombreuses musiques contemporaines et actuelles, le registre suraigu est aussi utilisé. Les notes suraiguës sont obtenues à partir de doigtés spéciaux permettant de faire sonner une harmonique particulière.

Les saxophones les plus couramment employés sont ceux qui forment le *quatuor de saxophones* : le saxophone soprano en Si bémol, le saxophone alto en Mi bémol, le saxophone ténor en Si bémol et le saxophone baryton en Mi bémol. Les saxophones sont des instruments transpositeurs : la note indiquée sur une partition n'est pas la même que celle entendue. Ainsi lorsque le saxophone alto joue un Do, c'est un Mi bémol qui est entendu (d'où sa désignation : saxophone alto en Mi bémol).

6.1.2 Contexte académique

Notre travail s'effectue dans un contexte académique. L'enseignement académique du saxophone prit son essor en 1942, quand Marcel Mule [Rou82, Thi04] fût nommé professeur au Conservatoire National Supérieur de Musique de Paris. Mule reprenait alors la classe de

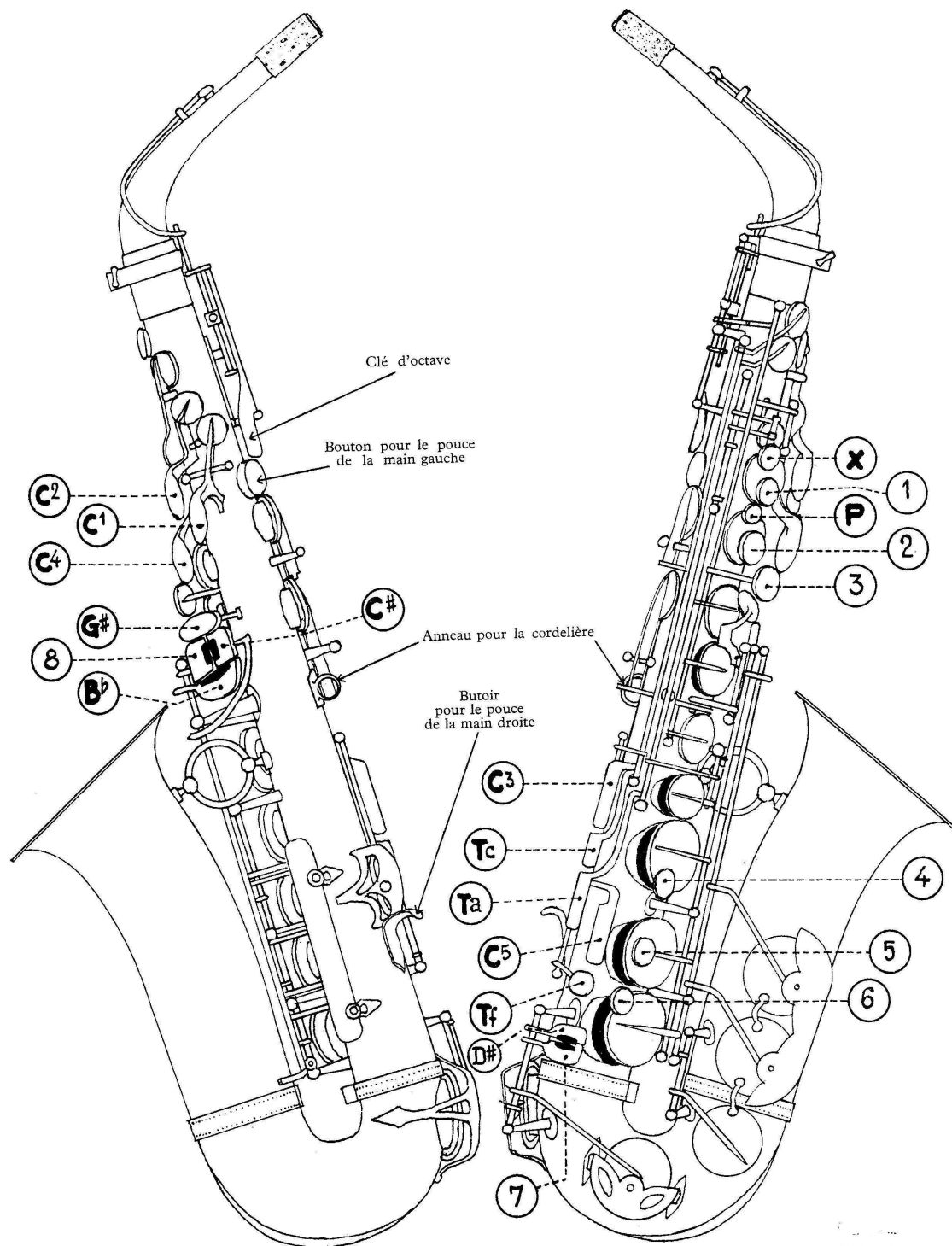


FIG. 23 – Saxophone alto en Mi bémol, avec indication des clés (illustration extraite de la méthode Londeix [Lon70]).

saxophone fermée depuis le lointain départ d’Adolphe Sax en 1870. Il a formé les professeurs de saxophone les plus influents en France, comme Daniel Deffayet ou Jean-Marie Londeix, mais aussi à l’étranger. Les traces de son enseignement – vibrato systématique, insistance sur l’aspect technique, phrasé classique – sont encore perceptibles dans l’enseignement du saxophone actuel, partout dans le monde.

6.1.3 Exercices de base

Comme pour tous les instruments, il y a des exercices de base au saxophone. Ces exercices sont utilisés pour apprendre l’instrument, progresser techniquement, mais aussi pour *chauffer* l’instrument (démarrage d’une session de jeu instrumental, où l’instrument se réchauffe réellement) ou pour entretenir le niveau technique de l’instrumentiste. Nous nous intéressons principalement ici aux exercices mécaniques, aux sons droits, filés et vibrés, aux gammes et aux arpèges. Ce sont des exercices techniques qui peuvent être joués sans expressivité, et qui sont pratiqués très régulièrement par les saxophonistes de tous niveaux. Durant l’apprentissage du saxophone, ces exercices de base sont souvent complétés par la pratique d’études techniques, qui mêlent technicité et expressivité (les “Études Variées” de Mule [Mul50] par exemple), et surtout par l’interprétation d’œuvres instrumentales, pour lesquelles la technicité est entièrement au service de l’expressivité.

Sons tenus

Au saxophone, les *sons tenus* sont des exercices de base pour apprendre notamment à contrôler sa *colonne d’air* (pression d’air fournie par l’instrumentiste), et à maîtriser certains aspects du jeu au saxophone – attaques, nuances – sur toutes les notes de la tessiture. Les *sons tenus droits*, ou plus simplement *sons droits* consistent à jouer une note à une certaine nuance fixe durant un temps assez long. Les *sons filés* imposent de jouer la note avec une dynamique de nuance *crescendo* (en augmentant l’amplitude), puis *decrescendo* (en diminuant l’amplitude), avec le meilleur *ambitus de nuance* possible (étendue entre la plus faible et la plus forte nuance jouée). Enfin, les *sons vibrés* permettent de travailler la régularité et la fréquence du vibrato.

La gestion de la colonne d’air influe de façon importante sur la performance des sons tenus, particulièrement sur les nuances et la justesse. C’est un facteur essentiel du contrôle de l’instrument qui peut refléter le niveau technique général d’un saxophoniste.

Le vibrato

D’après Gilbert *et al.* [GST05], il y a deux familles de vibrato au saxophone : le vibrato *à la mâchoire* et le vibrato *à l’air*. Dans le premier cas, le vibrato consiste en un mouvement de la mâchoire, qui permet de contrarier la vibration de l’anche et de former une ondulation dans le son. La deuxième technique ne fait pas appel à la mâchoire mais à la cavité buccale, en particulier à la langue. D’intensité plus faible, ce vibrato est plutôt utilisé en début et en fin de phrase pour pallier les limitations de l’instrument. A faible intensité, le vibrato de mâchoire est en effet difficile à produire. L’introduction de souffle et la perte de son ne permettent plus à la mâchoire d’avoir suffisamment de force pour contrarier la vibration de l’anche. Le musicien peut utiliser les deux systèmes de vibrato sur un même morceau, selon l’intensité et la partie de la phrase musicale.

Quelle que soit la technique de vibrato utilisée, l'apprentissage académique impose une régularité en fréquence et en amplitude pour les variations. Une fois cette base acquise, les paramètres du vibrato peuvent varier dans un but expressif. D'après les recherches effectuées sur le vibrato par Marcel Mule, Jean-Marie Londeix conseille dans le 3^{ème} cahier du "Saxophone en Jouant" [Lon70] de travailler le vibrato à la vitesse de base d'environ 300 ondulations à la minute, *i.e.* 4 ondulations par pulsation pour un tempo de 75-76 pulsations par minute, 3 ondulations à 100, 5 ondulations à 60, ...

Gammes et arpèges

Les gammes sont bien sûr, comme pour tous les instruments, un exercice imposé au saxophone. Elles peuvent être jouées dans plusieurs modes, les plus courants étant le mode majeur et le mode mineur harmonique. Elles peuvent être jouées conjointement (les notes de la gamme se suivent) ou par intervalles (Do, Mi, Ré, Fa, ... par exemple pour la gamme de Do majeur en tierces). Le cahier de gammes de Londeix [Lon62] est une référence pour cet exercice.

Un saxophoniste pratique également des arpèges, dépliant sur toute sa tessiture – parfois augmentée par l'utilisation de notes suraiguës – différents accords de gammes. Ces accords peuvent être majeurs, mineurs, de septième de dominante ou de septième diminuée par exemple.

Exercices mécaniques

Les exercices mécaniques sont des formules concises et limitées en nombre de notes et en durée. Le saxophoniste doit les répéter un grand nombre de fois. Ils sont destinés à faire progresser l'instrumentiste sur un point technique précis, à corriger des défauts techniques. Les instrumentistes travaillent souvent les passages techniques difficiles d'une œuvre instrumentale sous la forme d'exercices mécaniques. Londeix précise dans ses "Cahiers d'Exercices Mécaniques" [Lon77] qu'ils donnent l'habitude à l'élève d'attaquer la difficulté en face. Dans ce cahier, le protocole est très strict : métronome, tempo, durée d'exécution ; il n'y a aucune expressivité dans ce type d'exercice.

Dans le cadre d'une méthode automatique d'évaluation et de tutorisation de saxophonistes, des exercices mécaniques ciblés seraient efficaces pour corriger les défauts techniques détectés.

6.1.4 Évaluation du niveau technique

De nombreux conservatoires ou écoles de musique imposent à leurs élèves un examen technique. Il consiste à vérifier que le saxophoniste possède le niveau technique requis par son niveau académique. Il peut s'agir par exemple de jouer devant un jury une gamme choisie au hasard, ou de *déchiffrer* (jouer pour la première fois une partition) un passage technique exigeant. Il est également fréquent que le professeur demande à un élève la réalisation d'exercices techniques à chaque début de cours, afin d'évaluer régulièrement la progression de sa technique.

L'évaluation technique est relative sur au moins trois points, car la réussite de l'exercice est jugée :



FIG. 24 – Exercices réalisés par les saxophonistes pendant leur enregistrement. L'exemple est donné pour la note Do médium. L'instrumentiste doit jouer un son droit piano, un autre mezzo forte et un troisième forte. Il doit ensuite jouer un son filé, et finir pour cette hauteur de note avec un son droit vibré. Il n'y a pas de consigne sur la durée des notes.

- en fonction des difficultés liées à l'instrument : un exercice dans le registre grave du saxophone est plus difficile à réaliser que dans le registre médium par exemple ;
- en fonction de la difficulté propre à l'exercice : les gammes ne sont pas toutes de difficulté équivalente ;
- en fonction du niveau technique général de la classe de saxophones (groupe d'élèves) dans laquelle l'instrumentiste se trouve, du niveau académique qu'il est supposé avoir.

Si nous voulons automatiser l'évaluation du niveau technique d'un saxophoniste à partir d'un enregistrement sonore de ses exercices, il faut tenir compte de cet environnement. Si des instrumentistes d'une même classe jouent les mêmes exercices avec les mêmes instruments, la performance de l'exercice peut alors être évaluée uniquement en fonction du niveau général de la classe (en supposant l'effectif de la classe suffisant).

6.2 Protocole d'enregistrement des saxophonistes

Nous avons établi un protocole d'enregistrement adapté à l'évaluation technique des saxophonistes à partir des exercices techniques qu'ils pratiquent régulièrement. Comme notre méthode est fondée sur l'analyse de l'évolution des paramètres sonores, nous avons choisi des exercices à base de sons tenus. Ce choix est justifié, d'après le professeur de saxophone Jean-Marie Londeix [Umb00], car la maîtrise de la colonne d'air et la maîtrise de la justesse pour toutes les notes du saxophone sont deux éléments très corrélés avec le niveau technique général du saxophoniste dans un contexte académique.

Les enregistrements des saxophonistes ont eu lieu au Conservatoire National de Région de Bordeaux et à l'École de Musique de Talence, en France. Une trentaine de saxophonistes ont été enregistrés, de tous niveaux techniques, incluant les professeurs. Ils ont joué des sons tenus, sans consigne de durée, sur 6 hauteurs de note différentes : Si grave, Fa grave, Do médium, Sol aigu, Ré aigu et La suraigu.

Il y avait 5 exercices à réaliser par note :

- un son droit avec une nuance *piano* (faible amplitude),
- un son droit avec une nuance *mezzo forte* (amplitude moyenne),
- un son droit avec une nuance *forte* (forte amplitude),
- un son filé en *crescendo decrescendo* (amplitude de 0 à fort, puis de fort à 0),
- un son vibré.

Un exemple d'exercice demandé est donné par la figure 24.

Les données audio ont été enregistrées avec un microphone Sony ECM-MS907 relié à la carte son d'un ordinateur portable. Le format PCM a été choisi pour les fichiers audio, avec une fréquence d'échantillonnage de 44100 Hz et une quantification sur 16 bits. L'ensemble des fichiers audio forment une base de données d'environ 900 fichiers.

6.3 Analyse de l'évolution des paramètres du son

Nous souhaitons évaluer des instrumentistes jouant des sons tenus, en considérant leur capacité à contrôler leur pression d'air. Nous cherchons les paramètres du son qui conditionnent l'évaluation humaine et qui peuvent être analysés de façon fiable et robuste. Nous nous concentrons alors sur l'évolution de la fréquence fondamentale dans le temps, qui concerne perceptivement la justesse pour les musiciens, et sur l'évolution de l'amplitude, qui correspond à la perception des nuances. Nous avons utilisé des techniques d'estimation de fréquence : la fréquence fondamentale F_0 a été extraite par auto-corrélation à l'aide du logiciel d'analyse Praat, développé par Paul Boersma [Boe93, Boe01]. Initialement destinée à l'analyse de la voix, c'est une technique qui fonctionne pour des sons monophoniques. Tous les instruments mélodiques pouvant jouer un son monophonique, notre méthode d'évaluation reste donc généralisable pour tout instrument entretenu.

Avec une technique d'estimation de fréquence, la fréquence fondamentale peut être erronée si l'amplitude est trop faible, comme l'illustre la figure 25. En outre, la valeur de la fréquence fondamentale peut aussi être erronée à cause de sauts d'octaves, problème identifié de la méthode. Nous indiquerons dans la suite comment nous avons traité ces difficultés. Les paramètres analysés sont exprimés sur des échelles proches de la perception. Ainsi les fréquences s'expriment sur l'échelle des ERB, et l'amplitude s'exprime en décibels. Dans les sections suivantes nous notons F et A les vecteurs de fréquence et d'amplitude sur l'échelle proche de la perception, et f et a sinon.

6.4 Métriques d'évaluation

Nous introduisons les métriques utilisées pour évaluer la capacité d'un saxophoniste à contrôler les nuances, la justesse et la régularité du vibrato pendant la production du son.

6.4.1 Écart type pondéré

Pour jouer correctement un son droit, le saxophoniste doit maintenir constantes la fréquence et l'amplitude du son. Dans le but d'évaluer la qualité de sa performance, nous utilisons alors naturellement l'écart-type, ou déviation standard, notée σ :

$$\sigma(X) = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (X(i) - \bar{X})^2} \quad (26)$$

où X est le vecteur de données de longueur N , et la moyenne de X est :

$$\bar{X} = \frac{1}{N} \sum_{i=0}^{N-1} X(i) \quad (27)$$

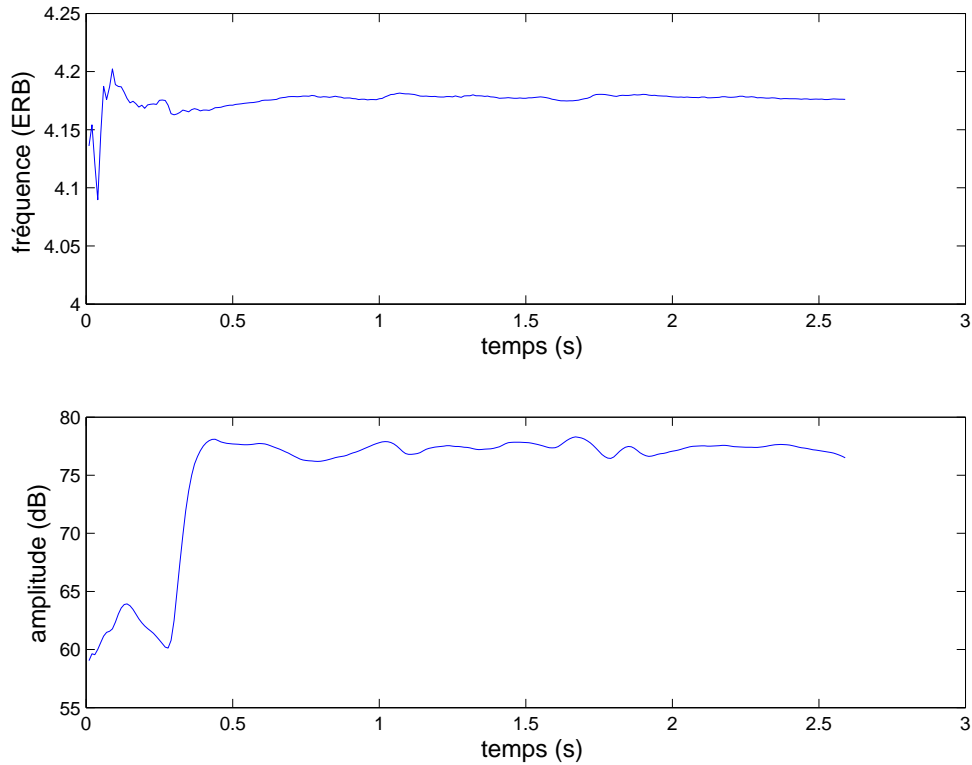


FIG. 25 – Vecteurs de fréquence et d’amplitude d’un son de saxophone extraits avec le logiciel Praat [BW06]. Avant l’attaque du son, l’estimation des valeurs de fréquence est perturbée par la faible amplitude.

Si l’amplitude du son est très forte, une légère déviation du paramètre de fréquence est tout de suite audible. Au contraire, si l’amplitude est très faible, de fortes variations en fréquence peuvent ne pas être perçues. Nous proposons donc de pondérer cet écart type par la valeur de l’amplitude, en utilisant l’écart type pondéré wd :

$$wd(X) = \sqrt{\frac{1}{N\bar{a}} \sum_{i=0}^{N-1} a(i) (X(i) - \bar{X})^2} \quad (28)$$

où a est le vecteur d’amplitude sur une échelle linéaire, de même taille que le vecteur X .

Cette pondération est également utile pour minimiser les problèmes dus à la technique d’estimation de fréquence, quand l’amplitude est très faible (figure 25). Nous pourrions plutôt utiliser un seuil d’amplitude, pour ignorer les parties de trop faible amplitude. Mais la phase d’attaque est très importante pour évaluer un instrumentiste, et elle pourrait être endommagée par cette opération. En utilisant l’écart type pondéré par l’amplitude, nous pouvons considérer l’évolution des paramètres sonores sur toute la durée du son.

6.4.2 Analyse glissante

Aucune consigne n'a été donnée aux instrumentistes sur la durée des sons, ces durées peuvent alors être très différentes. Pour comparer les écarts en amplitude et en fréquence de différents saxophonistes sur un même intervalle de temps, nous proposons une analyse glissante pour l'écart type pondéré présentée ci-dessus. L'analyse glissante *swd* permet également de comparer les valeurs des paramètres sonores à des moyennes locales, ce qui donne une meilleure évaluation de la déviation :

$$\text{swd}(X) = \frac{1}{K} \sum_{i=0}^{K-1} \text{wd}(X[i\Delta, \dots, i\Delta + 2\Delta]) \quad (29)$$

où Δ est la taille du décalage et $K = \lfloor N/\Delta \rfloor$. Le décalage Δ est égal à la moitié de la taille du vecteur afin de réaliser un recouvrement.

Le choix de la taille de la fenêtre d'analyse est déterminant. Une taille trop petite, et nous prenons en compte de très petites déviations qui ne sont pas perceptibles. D'un autre côté, une trop grande fenêtre d'analyse nous amène à considérer des déviations très lentes. Dans ce deuxième cas, les lentes variations sont pénalisées, alors qu'elles ne sont pas perçues comme gênantes. Sur la figure 26, la performance serait ainsi pénalisée, alors que l'évolution des paramètres de la courbe épaisse est satisfaisante. Nous avons choisi pour nos travaux d'utiliser une fenêtre de temps de 80 ms.

6.4.3 Métriques pour les sons droits

Le saxophoniste qui joue un son droit attaque le son avec une fréquence et une amplitude données. Idéalement, ces paramètres doivent rester constants jusqu'à la fin du son. Nous pouvons alors utiliser une analyse glissante avec écart type pondéré pour construire nos métriques. Les fréquences et les amplitudes étant différentes pour chaque exercice, nous divisons les résultats par les valeurs moyennes des paramètres. Les résultats des métriques sont alors tous comparables, quelles que soient la note jouée et la nuance imposée. Nous obtenons deux métriques, d_f pour évaluer la performance technique du son droit en fréquence, et d_a pour l'amplitude :

$$d_f = \frac{\text{swd}(F)}{\bar{F}} \quad (30)$$

$$d_a = \frac{\text{swd}(A)}{\bar{A}} \quad (31)$$

6.4.4 Ambitus de nuance

L'ambitus de nuance est le coefficient multiplicateur entre l'amplitude du son joué le moins fort et celle du son le plus fort. Pour les sons droits, il s'agit du coefficient entre les valeurs d'amplitude des sons droits *piano* et celles des sons droits joués *forte*. Nous notons cet ambitus α . Plus α est grand, plus le saxophoniste contrôle son échelle de nuances, sa pression d'air. C'est une qualité technique de maîtrise de l'instrument. Mais les exercices proposés sont alors techniquement plus difficiles : un son droit à très faible amplitude est bien plus difficile à produire qu'avec une amplitude moyenne. Pour évaluer correctement

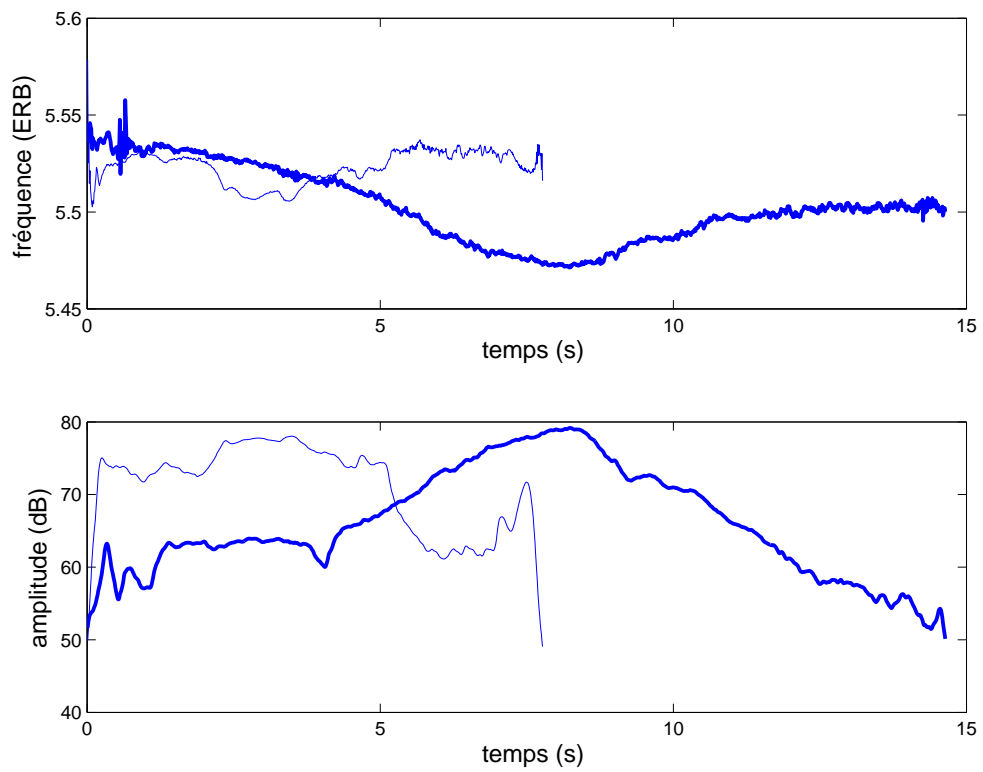


FIG. 26 – Vecteurs de fréquence et d'amplitude pour un son filé joué par deux saxophonistes différents. En gras, la courbe est celle d'un expert, l'autre est celle d'un saxophoniste de niveau élémentaire.

un instrumentiste, il convient alors d'observer conjointement les résultats donnés par les métriques mais aussi la valeur de l'ambitus α .

6.4.5 Métriques pour les sons filés

La fréquence d'un son filé doit être constante. La métrique de la fréquence d_f est alors la même que pour les sons droits. Lors d'un son filé idéal, l'amplitude du son démarre de 0, augmente linéairement (sur une échelle proche de la perception) pour atteindre une valeur maximale d'amplitude M , puis elle décroît linéairement jusqu'à 0. L'indice du maximum d'amplitude M dans le vecteur d'amplitude est noté m . Lors des enregistrements, les saxophonistes ont eu pour consigne de diminuer l'amplitude dès qu'ils avaient atteint le maximum, pour former idéalement deux pentes linéaires, dans une échelle proche de la perception. Nous proposons alors pour la métrique de l'amplitude de se référer à l'évolution linéaire L ainsi décrite :

$$L(i) = \begin{cases} a_1 \cdot i + A(0) & \text{si } i < m \\ a_2 \cdot (i - m) + M & \text{sinon} \end{cases} \quad (32)$$

avec N le nombre de valeurs d'amplitude analysées dans le son. Les coefficients a_1 et a_2 sont les coefficients respectivement de la partie linéaire croissante et décroissante, tels que :

$$\begin{cases} a_1 & = & \frac{M-A(0)}{m} \\ a_2 & = & \frac{A(N-1)-M}{N-1-m} \end{cases} \quad (33)$$

Soit D le vecteur différence entre le vecteur d'amplitude jouée A et le vecteur d'amplitude souhaitée L . Comme nous utilisons une échelle proche de la perception, D est un vecteur de ratios d'amplitude en échelle linéaire. Deux exemples de vecteurs L , de vecteurs d'amplitude A , et de leurs différences D sont donnés par la figure 27, pour deux saxophonistes de niveaux techniques différents.

La métrique proposée pour l'amplitude est l'analyse glissante avec écart type pondéré sur D . Comme l'objectif de l'exercice est de créer une forte dynamique de nuance, nous proposons de diviser les résultats par l'ambitus de nuance ($M - \min(A)$). La métrique $d_{<>}$ proposée pour l'amplitude est donc :

$$d_{<>} = \frac{\text{swd}(D)}{(M - \min(A))} \quad (34)$$

6.4.6 Métriques pour les sons vibrés

L'évolution de la fréquence pendant un son vibré est représenté par la figure 28, c'est une trajectoire pseudo-sinusoidale. Le vibrato est traditionnellement enseigné aux saxophonistes classiques avec 4 oscillations par temps pour environ 72 pulsations à la minute (voir section 6.1.3). La fréquence du vibrato est dans ce cas proche de 4.8 Hz. Nous avons alors proposé dans [RL06] de considérer la déviation des mesures de la fréquence du vibrato par rapport à cette valeur théorique comme métrique de fréquence pour les sons vibrés. La fréquence de ce vibrato peut cependant devenir plus personnelle, varier selon la pièce jouée ou par expressivité. La comparaison avec une valeur fixe ne tient pas compte de cela. Par

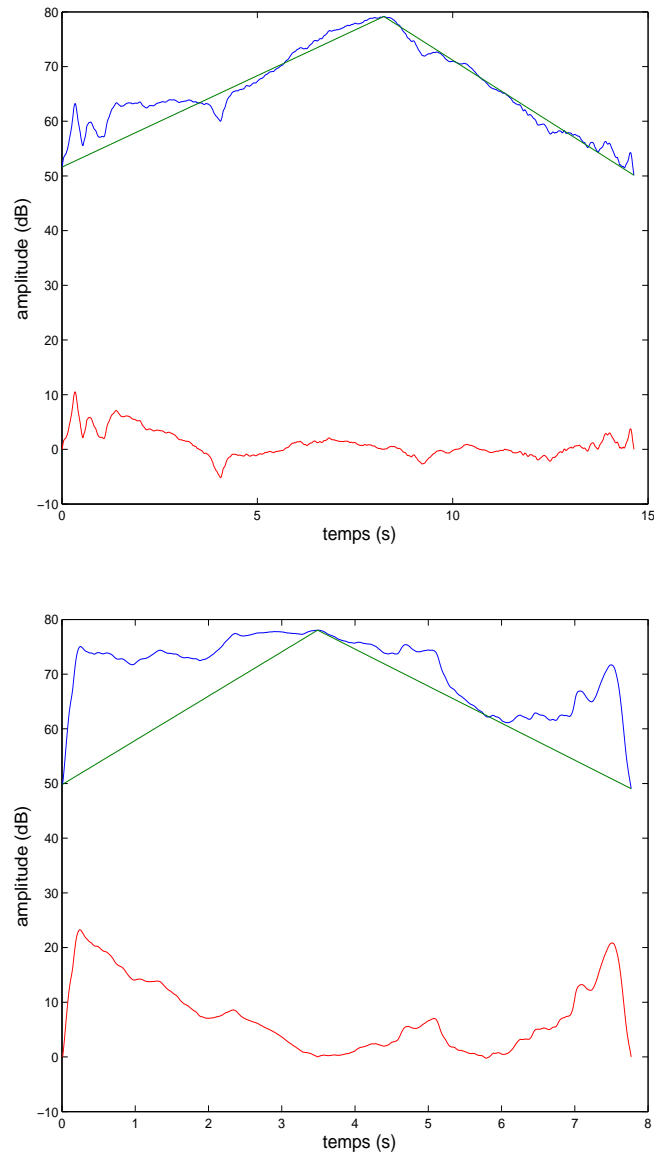


FIG. 27 – Sur chaque figure, il y a le vecteur d'amplitude A en haut, correspondant au jeu d'un son filé au saxophone, avec le vecteur linéaire par morceaux L associé. La courbe du bas représente le vecteur D , différence des vecteurs A et L . C'est cette différence qui sert de base au calcul de la métrique d'amplitude. Sur la première figure, le saxophoniste est classé comme expert alors que la deuxième figure présente la courbe d'un élève de niveau élémentaire.

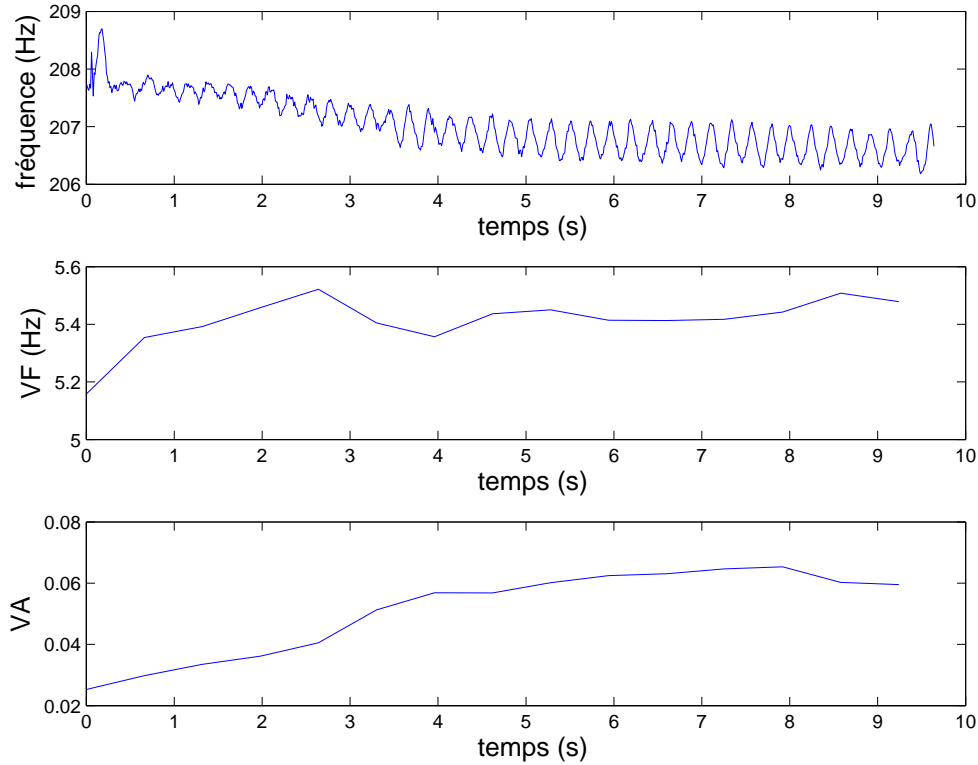


FIG. 28 – La figure du haut représente le vecteur F de fréquence du son vibré d'un saxophoniste expert. Au milieu, les valeurs du vecteurs VF tracent l'évolution de la fréquence du vibrato. En bas, ce sont les valeurs du vecteur VA qui donnent l'évolution de l'amplitude du vibrato.

contre, tout saxophoniste doit pouvoir faire vibrer le son de manière régulière en fréquence et en amplitude de fréquence. C'est cette régularité que nous évaluons maintenant, quelles que soient les valeurs des paramètres.

Nous suivons l'évolution en fréquence et en amplitude du vibrato dans le temps, en appliquant une analyse fréquentielle glissante sur le vecteur $(f - \bar{f})$. Pour chaque analyse fréquentielle, nous utilisons un intervalle de temps équivalent à 4 vibrations à 4.8 Hz, une fenêtre de Hann et une transformée de Fourier de 4096 points avec bourrage de zéros. Pour une fenêtre i donnée, la valeur et l'indice du maximum du spectre indiquent respectivement la variation d'amplitude $VA(i)$ et de fréquence $VF(i)$ du vibrato en fréquence. Nous proposons alors les métriques de fréquence d_{vf} et d'amplitude d_{va} du vibrato définies par :

$$d_{vf} = swd(VF) \quad (35)$$

$$d_{va} = swd(VA) \quad (36)$$

6.5 Résultats

Dans cette partie nous expliquons comment des notes techniques sont données aux saxophonistes à partir de l'application des différentes métriques vues dans la section précédente. Nous donnons des exemples possibles d'interprétation pédagogique des résultats.

6.5.1 Conversion des métriques aux notes

Nous proposons de convertir en notes techniques les résultats obtenus par l'application des métriques. Ces notes doivent pouvoir être interprétées plus facilement que les résultats bruts. Nous pourrions normaliser le résultat de chaque métrique et appliquer un coefficient multiplicateur, pour donner une note entre 0 et 100 par exemple pour chaque exercice. Mais nous voulons que l'évaluation technique soit relative à la performance générale de la classe de saxophone (les instrumentistes ont tous joué avec le même type de saxophone et ils ont réalisé les mêmes exercices techniques).

Nous avons groupé les saxophonistes dans des classes en fonction de leur niveau académique. Les notes techniques ont alors été données en fonction de la moyenne de la classe des saxophonistes confirmés. Cette classe regroupe des élèves de haut niveau et des professeurs. C'est une simulation du fonctionnement de l'enseignement instrumental académique, où les élèves de meilleur niveau servent de référence technique. La classe des saxophonistes experts aurait également pu servir de référence, mais son effectif est trop faible (3 éléments, contre 7 pour la classe des confirmés). La note de référence est établie à 100. Comme les résultats issus des métriques sont des valeurs d'erreur, nous proposons de retenir l'inverse de leur valeur. Nous multiplions cette valeur par 100, avant de la diviser par la moyenne des notes des instrumentistes de la classe des saxophonistes confirmés.

Nous séparons résultats d'amplitude et résultats de fréquence. Nous utilisons pour les résultats d'amplitude les métriques définies dans la section 6.4 : d_a , $d_{<>}$ et d_{va} pour calculer les notes techniques respectivement pour les sons droits, les sons filés et les sons vibrés. Les notes de fréquence sont obtenues en appliquant les métriques d_f , d_f et d_{vf} respectivement pour les sons droits, les sons filés et les sons vibrés.

6.5.2 Interprétation des résultats

Les saxophonistes ont été groupés en cinq classes – *débutants*, *élémentaires*, *moyens*, *confirmés*, *experts* – en fonction de leur niveau académique dans leurs écoles de musique. Les notes obtenues avec les métriques proposées reflètent bien ce classement : les classes obtiennent des notes homogènes, avec des écarts-types modérés.

Le tableau 5 montre les notes obtenues pour la note Sol aigu. Ces notes obtenues sont corrélées avec le classement académique : les *experts* obtiennent par exemple pour un son tenu *mezzo forte* une note d'amplitude de 108, les *confirmés* une note de 100, les *moyens* 61, les *élémentaires* 46 et les *débutants* 39. Nous pouvons également noter que tous les niveaux inférieurs à celui de la classe *confirmés* ont une certaine difficulté à maîtriser la justesse sur les sons droits *piano* et *mezzo forte* (faibles résultats en fréquence).

Les résultats de Pierre et Shang sont présentés dans les tableaux 6 et 7. Pierre est un élève de niveau moyen de l'École de Musique de Talence et Shang est un saxophoniste confirmé du Conservatoire National de Région de Bordeaux. Nous pouvons interpréter les résultats qu'ils obtiennent d'un point de vue technique et pédagogique. Les résultats de

Résultats d'Amplitude						
	α	p	mf	f	<>	vibrato
experts (3)	21	124 (12)	108 (30)	197 (100)	111 (13)	144 (42)
confirmés (8)	19	100 (33)	100 (28)	100 (25)	100 (23)	100 (57)
moyens (6)	20	55 (24)	61 (14)	67 (17)	57 (11)	48 (13)
élémentaires (5)	10	54 (13)	46 (5)	53 (12)	43 (12)	19 (34)
débutants (8)	9	50 (27)	39 (20)	47 (19)	35 (9)	19 (0)

Résultats de Fréquence						
	α	p	mf	f	<>	vibrato
experts (3)	21	146 (44)	100 (26)	136 (47)	127 (33)	130 (18)
confirmés (8)	19	100 (36)	100 (34)	100 (51)	100 (37)	100 (45)
moyens (6)	20	48 (18)	57 (19)	63 (19)	62 (20)	57 (31)
élémentaires (5)	10	33 (12)	39 (14)	37 (8)	32 (3)	14 (25)
débutants (8)	9	35 (19)	32 (15)	34 (17)	40 (19)	16 (0)

TAB. 5 – Résultats pour la note Sol aigu. α est l'ambitus de nuance (coefficient d'amplitude entre le son droit piano et le son droit forte ; p, mf, et f correspondent respectivement aux sons droits joués piano, mezzo forte et forte ; la notation <> correspond au son filé, et vibrato au son vibré. 5 classes d'instrumentistes sont représentées (le nombre d'instrumentistes par classe est indiqué entre parenthèses). La classe confirmés sert de référence 100 pour donner une note relative aux performances individuelles. Les résultats sont des notes, avec l'écart type de la classe entre parenthèses. Les classes sont assez homogènes, avec des valeurs d'écart type modérées. Surtout, les notes obtenues correspondent bien au niveau technique supposé, alors que les classes sont composées en fonction du niveau académique des instrumentistes. Ainsi les notes d'amplitude pour le son tenu mezzo forte sont parfaitement corrélées avec le niveau académique.

Résultats d'Amplitude					
note	α	p	mf	f	<>
Si grave	22	21	36	28	44
Fa grave	31	66	69	47	50
Do médium	26	46	32	54	67
Sol aigu	21	28	48	50	59
Ré aigu	37	31	43	50	56

Résultats de Fréquence					
note	α	p	mf	f	<>
Si grave	22	55	79	67	90
Fa grave	31	30	65	86	89
Do médium	26	57	76	131	107
Sol aigu	21	46	68	78	87
Ré aigu	37	55	85	122	113

TAB. 6 – Résultats de Pierre, un saxophoniste de niveau moyen. Les résultats en fréquence sont assez bons, voire très bons pour les sons tenus forte. Pierre doit progresser dans le contrôle du son à faible amplitude, dans le grave notamment. Il n'a pas réalisé les exercices avec vibrato.

Pierre par exemple (tableau 6), montrent qu'il contrôle bien sa justesse, particulièrement pour les sons tenus *forte*. Mais il doit progresser dans le contrôle du son à faible amplitude, dans le grave notamment, et apprendre le vibrato. Shang (tableau 7) pourrait avoir un meilleur ambitus de nuances. Il maîtrise néanmoins très bien l'amplitude des sons graves. Il doit être plus attentif au paramètre de justesse, car il obtient un ensemble de résultats assez faibles en fréquence, à part dans le registre médium.

Avec un nombre limité d'exercices techniques et des métriques adaptées, nous arrivons à évaluer techniquement un saxophoniste relativement à des saxophonistes confirmés. Les résultats sont faciles à interpréter, et révèlent les facilités ou difficultés techniques.

6.6 Évaluation du niveau technique général

Chaque métrique proposée précédemment est adaptée à un aspect technique particulier. Il est donc possible d'évaluer les interprètes sur différents aspects techniques. Mais il est également intéressant de donner une note technique générale qui reflète le niveau technique général d'un instrumentiste. Cela nous permet en plus d'interclasser des performances, ou des saxophonistes différents. Encore une fois, c'est le mécanisme qu'appliquent les professeurs de musique, en combinant (parfois difficilement) les différents aspects techniques d'une performance pour interclasser des élèves d'une même classe.

6.6.1 Note technique générale

Pour obtenir une note technique générale, il faut combiner les résultats obtenus par l'instrumentiste pour les différents points techniques observés. Les résultats obtenus avec

Résultats d'Amplitude						
note	α	p	mf	f	<>	vibrato
Si grave	15	167	92	104	85	86
Fa grave	34	127	93	81	94	160
Do médium	15	105	91	111	92	121
Sol aigu	19	92	114	125	119	40
Ré aigu	23	142	80	85	107	114

Résultats de Fréquence						
note	α	p	mf	f	<>	vibrato
Si grave	15	72	59	64	53	92
Fa grave	34	36	62	78	78	91
Do médium	15	146	84	108	98	62
Sol aigu	19	93	81	75	80	116
Ré aigu	23	66	51	89	81	189

TAB. 7 – Résultats de Shang, un saxophoniste confirmé. Les résultats en amplitude sont très bons, avec un remarquable contrôle dans le grave, même si l'ambitus de nuance α pourrait être plus élevé à son niveau. Shang doit par contre travailler la justesse, car il a un ensemble de résultats assez faibles en fréquence, à part dans le médium.

l'exercice du vibrato et celui de la note La suraiguë ne sont pas pris en compte, car trop peu d'instrumentistes les ont réalisés correctement. L'ambitus de nuance est un paramètre qui donne une information pédagogique importante, sa valeur est donc intégrée à la note générale.

Nous supposons alors que les résultats évoluent logarithmiquement en fonction du niveau technique de l'instrumentiste : il est d'autant plus dur de progresser que les résultats sont déjà bons. Le bénéfice de cette approche est illustré par la figure 29. Ainsi, dans le but de mieux différencier les niveaux, nous considérons pour classer les instrumentistes le logarithme de leurs résultats. Pour combiner les différents résultats provenant des métriques, nous donnons le même poids à chacun. Le résultat du $k^{\text{ème}}$ saxophoniste obtenu avec la métrique d est ainsi calculé :

$$\tilde{d}(k) = \frac{\log(d(k)) - \overline{\log(d)}}{\sigma(\log(d))} \quad (37)$$

avec \bar{x} et $\sigma(x)$ respectivement la moyenne et l'écart type de x .

Les résultats $\tilde{d}(k)$ sont centrés-réduits. Ils peuvent alors être combinés, leurs écarts-types et leurs moyennes valant respectivement 1 et 0. La note générale obtenue par un instrumentiste est alors la moyenne de ces résultats. Cette note est comprise entre -1 et 1. Comme nous pouvons l'observer sur la figure 29, les notes évoluent quasi-linéairement en fonction du niveau technique évalué. Ce comportement linéaire est utile pour un classement correct des instrumentistes, car il permet un classement plus régulier.

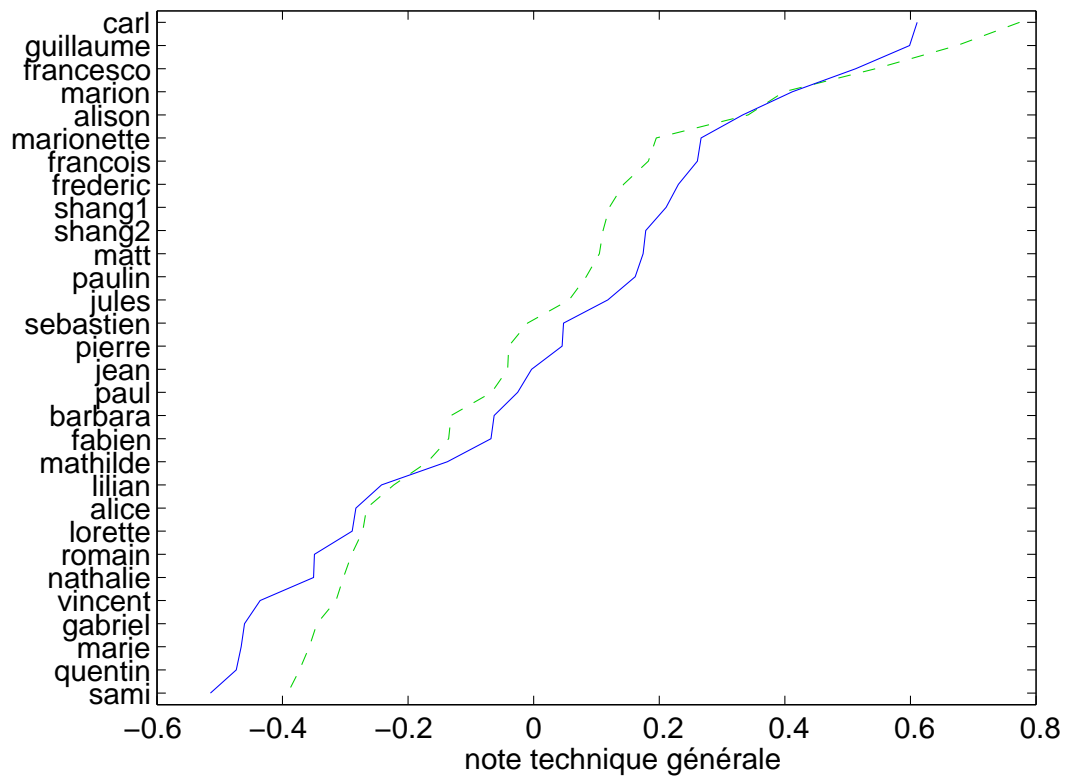


FIG. 29 – Notes techniques générales obtenues par les saxophonistes. En pointillés, les notes sont données directement avec les résultats issus des métriques. La ligne pleine indique les notes calculées avec la version logarithmique des résultats. Ces dernières sont utilisées pour un classement automatique des instrumentistes, car elles permettent un classement plus régulier.

6.6.2 Classement automatique des instrumentistes

Si la note technique générale est déjà un élément pédagogique intéressant, elle nous permet aussi d'envisager un classement automatique des instrumentistes dans des classes de niveau. Pour les résultats obtenus précédemment, nous avons classé les saxophonistes en fonction de leur niveau académique, évalué chaque année par leurs professeurs. Nous avons retenu cinq niveaux différents, de *débutants* à *experts*.

La technique de groupage hiérarchique [Joh67] est un algorithme de classement qui produit une série de partitions des instrumentistes. La première partition consiste en des singletons, *i.e.* une classe par instrumentiste. La dernière correspond à une seule classe contenant tous les musiciens. À chaque étape, la méthode de classement groupe les deux classes les plus proches en fonction de leur note générale. Il s'agit donc pour la première étape de regrouper les deux éléments qui ont la note la plus proche.

Le classement hiérarchique peut être représenté par un diagramme à deux dimensions, appelé *dendogramme*, qui illustre la fusion pour chaque étape du classement (voir figure 30). La longueur de la barre verticale qui relie deux classes est calculée en fonction de la distance entre les deux classes. Nous avons choisi pour notre étude un partitionnement en cinq classes, de *débutants* à *experts*. Ce choix peut être représenté par une coupe à une certaine étape dans le dendogramme, représentée sur la figure 30 par une ligne en pointillés. Le nombre de classes peut ainsi être facilement changé.

Le partitionnement automatique en fonction des notes techniques générales correspond assez bien au niveau académique supposé des interprètes. La grande majorité des saxophonistes se trouvent dans la classe correspondant à leur niveau académique. Pour chaque métrique considérée, l'écart-type du classement automatique est inférieur ou égal à celui obtenu par un classement académique. Cela démontre en partie la pertinence de la note technique générale pour comparer et interclasser automatiquement des instrumentistes. Si une performance est ajoutée à une évaluation déjà faite, les notes sont recalculées, et les classes ré-établies.

6.7 Perspectives

Nous avons proposé une méthode automatisée d'évaluation technique de saxophonistes qui :

- évalue indépendamment plusieurs points techniques en suivant l'évolution des paramètres sonores,
- interclasse différentes performances selon les différents résultats,
- donne une note technique générale à la performance.

Cette méthode propose aux saxophonistes d'estimer leur niveau technique, d'identifier leurs défauts afin de les corriger. Cela offre une information pédagogique à l'instrumentiste, un retour automatique.

6.7.1 Saxophone tutor

À partir de notre méthode automatisée d'évaluation technique et de classement, il est simple de proposer une méthode de tutorisation technique pour les saxophonistes :

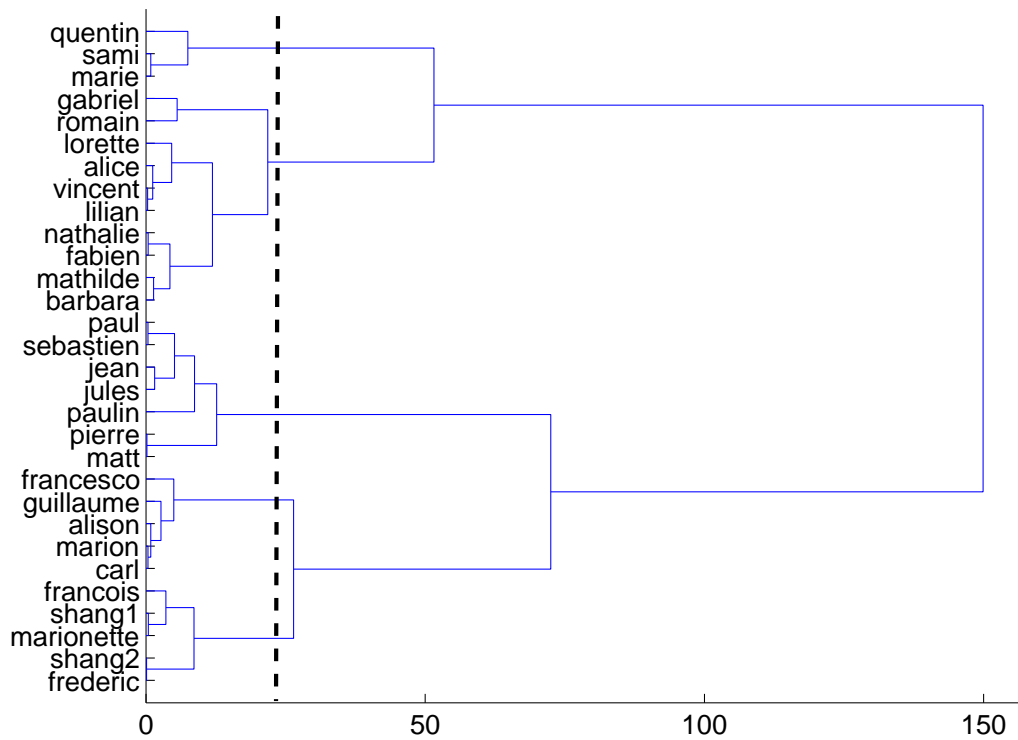


FIG. 30 – Dendrogramme du classement des saxophonistes. Les instrumentistes sont classés en fonction de leur note technique générale. Les classes de niveau technique sont obtenues en coupant le dendrogramme à une certaine profondeur représentée par la ligne pointillée.

- Le saxophoniste réalise des exercices simples de sons tenus, autant de fois qu’il le souhaite.
- Ses notes sont affichées, par exercice, et visualisées. Les figures 31 et 32 donnent des exemple de visualisations possibles.
- Des exercices spécifiques sont proposés par rapport aux défauts techniques identifiés.
- Une note technique générale est donnée. Elle sert de base à un classement technique relatif aux autres performances.

Un même saxophoniste peut ainsi s’évaluer régulièrement, et vérifier que ses anciennes performances sont moins bien classées.

6.7.2 Généralisation de la méthode à d’autres instruments

Notre méthode d’évaluation technique est généralisable à tous les instruments entretenus. Avec ce type d’instruments, le musicien doit maîtriser l’évolution des paramètres sonores dans le temps. C’est le cas de tous les instruments à vent, mais aussi du violon et son archet, et même de la voix. Il faut choisir les exercices adéquats pour chaque instrument, qui serviront de base au protocole. Les phases d’attaque du son pourraient être mieux observées qu’avec la méthode actuelle. Une méthode de tutorisation automatisée peut alors être mise en place.

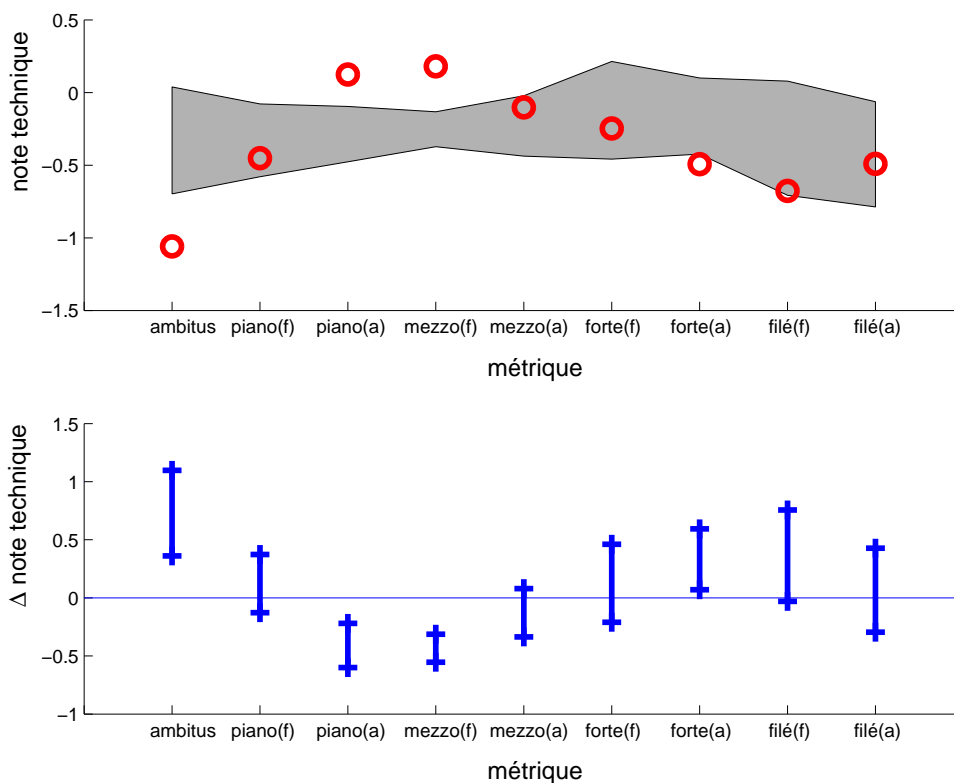


FIG. 31 – Visualisation des résultats de Lilian en fonction du type d'exercice. En haut, une première visualisation qui donne les notes sous forme de cercle pour les différentes métriques. *piano(f)* donne le résultat de la performance en fréquence du son droit piano, *filé(a)* celui de la performance en amplitude du son filé. La zone grisée est délimitée par deux bornes : la ligne inférieure représente la moyenne de la classe à laquelle appartient Lilian, pour chaque exercice, et la ligne supérieure représente la moyenne de la classe supérieure. En bas, c'est une autre possibilité de visualisation des mêmes données : le trait horizontal représente le niveau technique de Lilian pour chaque exercice. Ce sont les barres verticales qui représentent par leur borne inférieure et supérieure respectivement les niveaux moyens de la classe courante et celui de la classe supérieure.

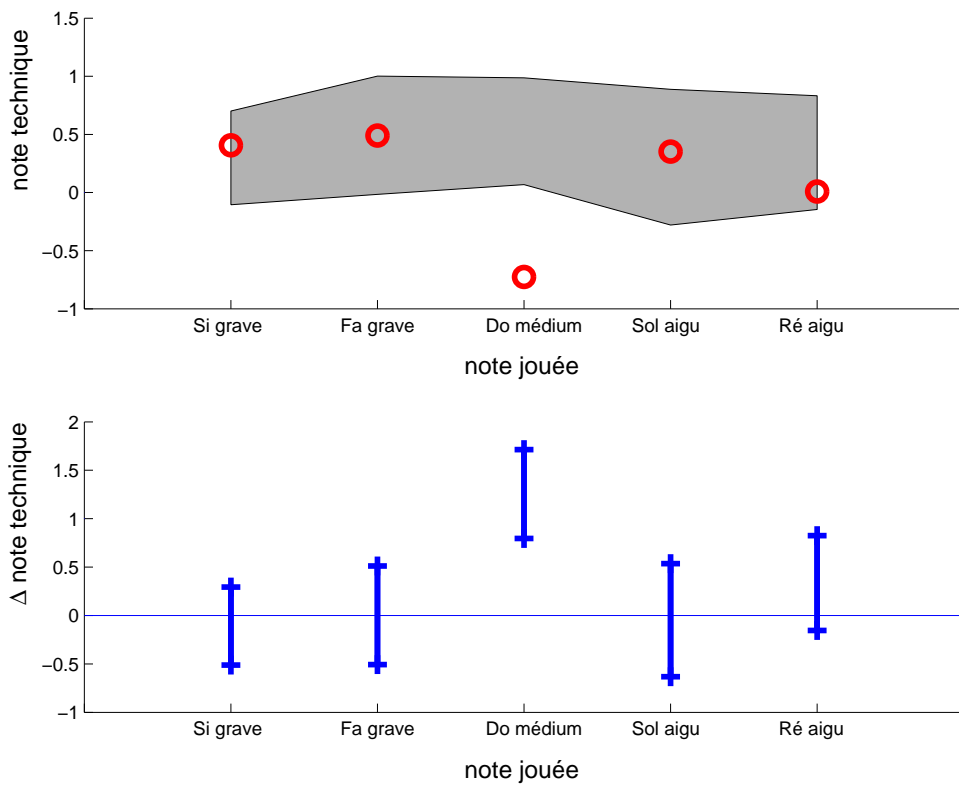


FIG. 32 – Visualisation des résultats de Paul en fonction de la note jouée. Comme les résultats sont relatifs, nous pouvons interpréter que Paul est d'un bon niveau technique par rapport à sa classe (il est un peu au-dessus des bornes inférieures de la zone grisée ou des barres verticales). Il n'a pas atteint encore le niveau technique supérieur. Par contre il doit progresser spécifiquement sur le contrôle en justesse et en amplitude de son médium, que ceux de sa classe maîtrisent beaucoup mieux que lui.

Troisième partie

Synthèse Sonore Rapide

Nous présentons dans cette partie nos recherches sur des méthodes rapides de synthèse additive. Les modèles spectraux paramètrent le son au niveau de l'oreille interne et sont ainsi proches de la perception. À la base de ces modèles se trouve la synthèse additive, qui permet de reproduire fidèlement de nombreux sons naturels et de les manipuler. La structure de base de cette synthèse est le partiel, oscillateur quasi-sinusoïdal dont la fréquence et l'amplitude évoluent lentement dans le temps. La synthèse additive consiste à produire des échantillons d'un son en additionnant les signaux élémentaires de chaque oscillateur.

Si la synthèse additive est très souple, puisqu'elle permet de contrôler individuellement les paramètres sonores des oscillateurs, elle peut difficilement être utilisée pour traiter des sons particulièrement complexes, de spectre riche, car elle nécessite un temps de calcul important dans ce cas. Les méthodes traditionnellement utilisées pour ce type de synthèse sont le résonateur numérique ou la transformée de Fourier inverse. Même s'il existe des algorithmes efficaces, diminuer encore le temps de calcul pour la synthèse permet dans un contexte temps-réel de laisser plus de temps aux autres étapes du traitement du son, l'analyse et la transformation sonore.

Le chapitre 7 présente nos recherches sur des techniques non linéaires de synthèse sonore. Ces techniques anciennes permettent de générer des sons complexes et harmoniques avec peu de calculs. Nous voulons savoir si elles peuvent remplacer avantageusement les méthodes classiques pour la synthèse additive des sons harmoniques. Nous étudions successivement les différentes techniques non linéaires de modulation, de distorsion. Nous testons aussi l'utilisation de fonctions d'après leurs développements en séries de Taylor, et la synthèse par formes closes.

Le chapitre 8 expose la recherche menée avec Marchand et Strandh [RSM06] sur une nouvelle méthode de synthèse sonore additive. Nommée PASS (*Polynomial Additive Sound Synthesis*), elle utilise des polynômes pour remplacer les fonctions sinusoïdales des oscillateurs. Enfin, nous présentons avec Fréchet [RF06] une application possible à la méthode PASS dans le domaine de la synthèse réaliste de surfaces océaniques.

Chapitre 7

Techniques non linéaires pour la synthèse des sons harmoniques

Nous avons présenté dans le chapitre 2 les algorithmes de synthèse additive les plus rapides. Nous souhaitons maintenant étudier des techniques plus spécifiques, afin d'accélérer la synthèse additive pour des cas particuliers. Nous nous limitons alors à la famille des sons harmoniques, qui contient de nombreux sons d'instruments (flûte, saxophone, ...) et les voyelles de la voix par exemple. Nous avons alors répertorié différentes techniques non linéaires qui peuvent engendrer des structures harmoniques. L'objectif est de déterminer si elles peuvent produire un son très proche de celui produit par une synthèse additive linéaire, mais beaucoup plus rapidement, *i.e.* avec une complexité inférieure.

Une fonction g est linéaire si et seulement si $g(\lambda x + \mu y) = \lambda g(x) + \mu g(y)$ pour tout x, y , et pour toutes constantes réelles λ et μ , elle n'est pas linéaire (ou non linéaire) sinon. Dans le cas de la synthèse sonore, Roads [Roa96] indique qu'il n'y a pas de théorie générale possible pour les systèmes non linéaires. En pratique une synthèse sonore non linéaire consiste à produire un grand nombre de partiels à partir d'un nombre réduit d'oscillateurs. Comme il y a des fréquences en sortie qui ne sont pas en entrée, la cohérence psychoacoustique n'est cependant pas respectée (sauf si les composantes générées sont inaudibles).

La modulation d'amplitude permet par exemple de générer $2N$ partiels avec $N + 1$ oscillations sinusoïdales, et la modulation de fréquence permet en combinant seulement deux oscillateurs d'en générer une infinité. Autres pistes possibles, les fonctions $g(x) = \ln(1 + x)$ ou $g(x) = |x|$ qui, appliquées à une sinusoïde $x = \sin(2\pi ft)$, engendrent une infinité d'oscillations sinusoïdales couvrant la totalité de l'étendue du spectre audible. Les fréquences des oscillations générées sont toutes multiples de la fréquence de la sinusoïde d'origine, ce qui confère au son engendré une structure harmonique. Les amplitudes des partiels du son sont cependant dictées par les expressions mathématiques génératrices, et sont très difficilement contrôlables.

Nous avons retenu plusieurs critères d'évaluation pour les techniques proposées. Le premier critère porte sur la conservation de la structure harmonique. Cela consiste principalement à observer les éventuels repliements spectraux que peuvent générer les méthodes. La complexité des méthodes étudiées est un paramètre également très important, puisque nous cherchons des techniques rapides pour synthétiser de nombreux partiels. Enfin nous

avons observé le spectre fréquentiel du signal synthétisé, pour le comparer à celui du signal synthétisé par une méthode additive linéaire. Le moyen de comparaison retenu ici est la conservation de l'enveloppe spectrale, et notamment des formants, dont nous connaissons l'importance pour la perception (voir section 2.3.3). Comme les partiels du son ne sont pas synthétisés individuellement, chaque méthode est alors appréciée en fonction du degré de contrôle des amplitudes des partiels du son généré. Par commodité et suivant la méthode, nous utiliserons indifféremment dans ce chapitre la fonction sinus ou la fonction cosinus pour décrire le signal d'un oscillateur (il ne s'agit que d'un changement du paramètre de phase).

Nous étudions dans ce chapitre successivement les différentes techniques de modulation : modulations d'amplitude dans la section 7.1 et modulations de fréquence dans la section 7.2. Puis nous nous intéressons aux techniques de distorsion dans la section 7.3. Nous répertorions dans la section 7.4 des fonctions dont les développements en série de Taylor sont remarquables pour la synthèse sonore. La synthèse par formes analytiques est alors présentée (section 7.5), et la section 7.6 donne une conclusion générale à cette étude.

7.1 Modulation d'amplitude

Pour obtenir un phénomène de modulation, il faut mettre au moins deux signaux sinusoïdaux en présence. L'idée est d'utiliser un des deux signaux pour faire varier, moduler, un des paramètres de l'autre signal : la sortie du premier signal devient un des paramètres du second. La terminologie utilisée en synthèse par modulation est celle introduite lors des premières utilisations du principe de modulation de signaux, avec la transmission radio. C'est ainsi que l'on parle toujours de signal porteur et de signal modulateur, ce dernier étant le signal contenant l'information à transmettre.

Les deux méthodes de modulation d'amplitude étudiées sont la modulation en anneau (RM pour *Ring Modulation*) et la modulation d'amplitude (AM pour *Amplitude Modulation*). La technique RM module deux signaux bipolaires, alors que la technique AM module un signal bipolaire avec un signal unipolaire. Un signal bipolaire, comme la plupart des signaux audio, est un signal qui a dans le domaine temporel des valeurs d'amplitude positives et négatives. Le signal unipolaire n'a que des valeurs d'amplitude positives. Cette notion est illustrée par la figure 33.

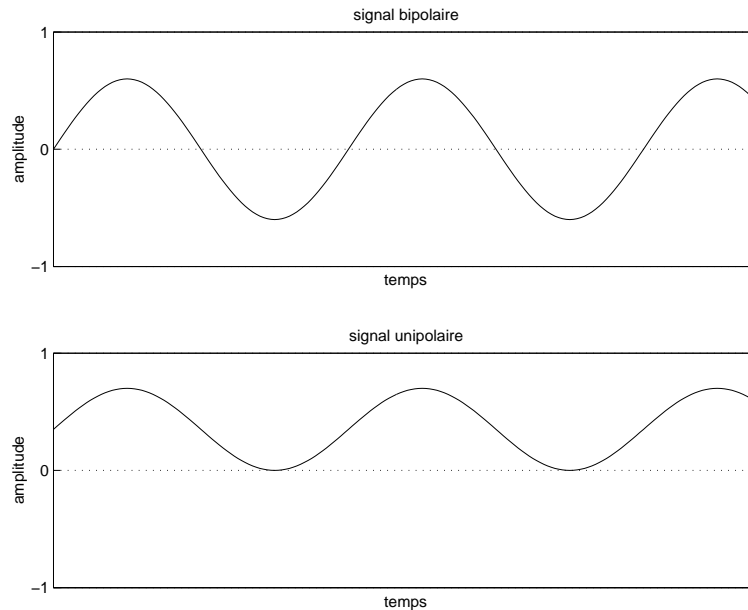
7.1.1 Modulation d'amplitude RM

Une des premières techniques employées par la musique électronique a été la modulation en anneau, ou *Ring Modulation*. Deux signaux bipolaires sont multipliés, un signal porteur s_c par un signal modulateur s_m :

$$s(t) = s_c(t) \cdot s_m(t) \quad (38)$$

La formule trigonométrique (39) permet une réécriture de l'expression du signal :

$$\cos \alpha \cos \beta = \frac{1}{2}(\cos(\alpha + \beta) + \cos(\alpha - \beta)) \quad (39)$$

FIG. 33 – *Signal bipolaire et unipolaire.*

Avec un signal porteur s_c et un signal modulateur s_m , de fréquences respectives f_c et f_m , le signal peut alors être exprimé ainsi :

$$s(t) = \frac{1}{2}(\cos(2\pi(f_c + f_m)t) + \cos(2\pi(f_c - f_m)t)) \quad (40)$$

Le signal résultant de l'application d'une synthèse RM ne contient pas de composante sinusoïdale de fréquence égale à celle du signal porteur ou à celle du signal modulateur, mais à leur somme ($f_c + f_m$) et à leur différence ($f_c - f_m$). L'amplitude de ces composantes vaut la moitié de la multiplication des amplitudes des signaux porteur et modulateur. Un exemple est donné par la figure 34 avec $f_c = 1000$ Hz et $f_m = 400$ Hz.

Si le signal porteur est complexe et le signal modulant sinusoïdal pur, la modulation en anneau remplace chaque partiel du signal porteur par une paire de composantes sinusoïdales, la somme et la différence de la fréquence du partiel et celle du signal modulant. Dans le cas où les signaux multipliés sont tous les deux complexes, le spectre résultant est encore plus riche, puisqu'il contient les sommes et différences de toutes les fréquences mises en présence, mais le son résultant de la modulation en anneau peut alors être inharmonique.

7.1.2 Modulation d'amplitude AM

La méthode de modulation d'amplitude AM est similaire à la modulation en anneau, sauf que le signal modulateur est unipolaire au lieu d'être bipolaire. Si le signal modulant est d'amplitude a_m et de fréquence f_m , et le signal porteur d'amplitude a_c et de fréquence f_c , alors le signal modulé $s(t)$ peut s'exprimer ainsi :

$$s(t) = a \cos(2\pi f_c t) \text{ avec } a = (a_c + a_m \cos(2\pi f_m t)) \quad (41)$$

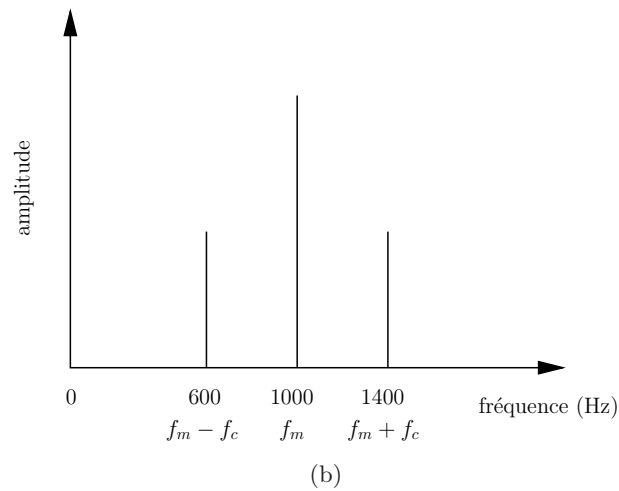
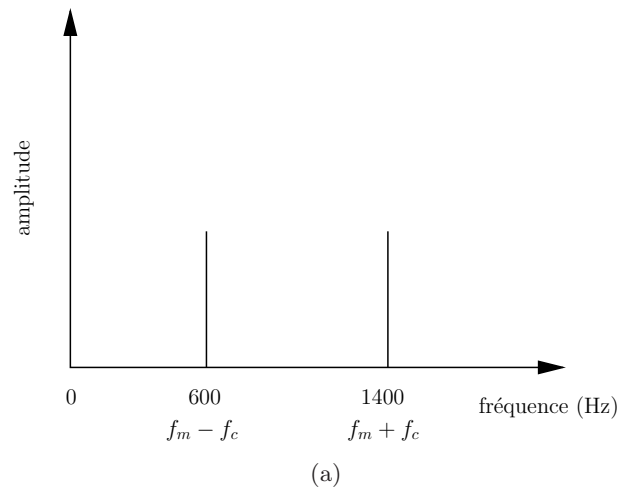


FIG. 34 – Exemple de modulation avec un signal modulant de fréquence $f_m = 400$ Hz et un signal porteur de fréquence $f_c = 1000$ Hz. Les figures représentent les spectres des signaux synthétisés avec (a) la modulation RM et (b) la modulation AM. Dans les deux cas, deux partiels sont générés, à 600 Hz et 1400 Hz, ajoutés à celui situé à la fréquence porteuse pour la modulation AM.

Il vient alors :

$$s(t) = (a_c + a_m \cos(2\pi f_m t)) \cos(2\pi f_c t) \quad (42)$$

$$= a_c \cos(2\pi f_c t) + a_m (\cos(2\pi f_m t) \cos(2\pi f_c t)) \quad (43)$$

D'où :

$$s(t) = a_c \cos(2\pi f_c t) + \frac{a_m}{2} (\cos(2\pi(f_c + f_m)t) + \cos(2\pi(f_c - f_m)t)) \quad (44)$$

Le spectre du signal modulé contient la fréquence porteuse f_c et deux autres composantes de fréquences $(f_c + f_m)$ et $(f_c - f_m)$. Un formant composé de trois partiels est alors généré. La figure 34 illustre cela par un exemple avec une fréquence porteuse de 1000 Hz et une fréquence modulante de 400 Hz. Le spectre du signal résultant de la synthèse AM contient un partiel de même fréquence que celle de la porteuse, à 1000 Hz, ainsi que des composantes latérales à 600 et 1400 Hz. Dans l'équation (44), le coefficient a_m est le facteur de modulation qui permet d'ajuster l'amplitude des composantes latérales par rapport à la composante centrale. Quand ce facteur vaut 1, le signal porteur et le signal modulant ont la même amplitude, l'amplitude des composantes latérales vaut alors la moitié de l'amplitude de la composante centrale.

7.1.3 Résultats

La modulation AM a pour avantage sur la modulation RM de conserver dans le signal modulé une composante sinusoïdale de fréquence identique à celle du signal porteur. Pour réaliser une synthèse additive non linéaire, la modulation AM peut être appliquée sur les partiels correspondant aux maxima locaux de l'enveloppe spectrale du signal à synthétiser, pour tenter de reconstruire des formants. Deux composantes fréquentielles sont ajoutées pour chaque formant avec une seule modulation AM (figure 35). Les formants sont alors composés de trois partiels, ce qui est assez pauvre.

7.2 Modulation de fréquence

La modulation de fréquence (FM pour *Frequency Modulation*) est à l'origine une technique de transmission de signal analogique qui utilise la phase d'une sinusoïde autour d'une fréquence porteuse comme support de l'information. Elle permet de générer un ensemble potentiellement infini de partiels dans un son à partir de seulement deux oscillateurs.

7.2.1 FM formantique

Chowning [Cho73] a proposé d'utiliser et de contrôler la modulation de fréquence à des fins de synthèse sonore. L'expression d'un signal de synthèse FM est donnée par l'équation :

$$s(t) = a_c \sin(2\pi f_c t + I \sin(2\pi f_m t)) \quad (45)$$

où a_c et f_c sont l'amplitude et la fréquence du signal porteur, f_m la fréquence du signal modulateur, et I l'indice de modulation. En utilisant le principe mathématique suivant :

$$\sin(\theta + a \sin(\beta)) = J_0(a) \sin(\theta) + \sum_{n=1}^{+\infty} J_n(a) (\sin(\theta + n\beta) + (-1)^n \sin(\theta - n\beta)) \quad (46)$$

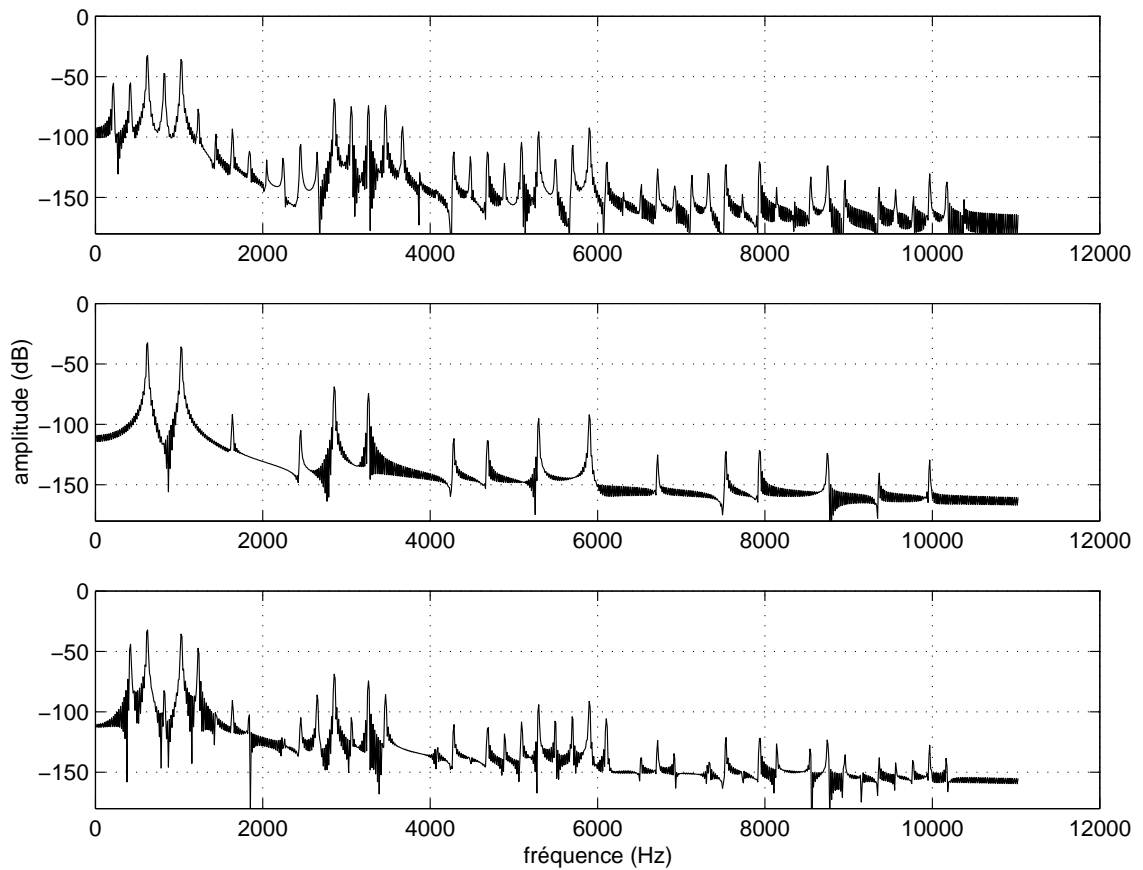


FIG. 35 – Application de la technique AM sur les maxima de l'enveloppe spectrale d'un signal. Il y a en haut le spectre d'un signal de voix obtenu par synthèse linéaire, puis au milieu les partiels correspondant au maxima locaux de l'enveloppe spectrale (un maximum local est ici un partiel, tel que son amplitude est supérieure ou égale à celle de ses deux voisins). En bas se trouve le spectre du signal obtenu par l'application d'une synthèse AM sur le signal composé des partiels maxima, avec la fréquence fondamentale du signal comme modulante. Le spectre généré est assez pauvre en partiels : 3 partiels seulement sont générés par formant. L'enveloppe spectrale reste cependant proche de celle du signal synthétisé linéairement.

où $J_n(a)$ est la fonction de Bessel de première espèce d'ordre n au point a , De Poli [Pol83] a donné une expression équivalente à l'équation (45) qui montre que les partiels générés dépendent des fonctions de Bessel :

$$s(t) = \sum_{n=-\infty}^{+\infty} J_n(I) \sin((2\pi f_c + 2n\pi f_m)t) \quad (47)$$

Les fonctions de Bessel J_k de première espèce sont les solutions de l'équation différentielle de Bessel :

$$x^2 y'' + xy' + (x^2 - k^2)y = 0, \quad k \geq 0 \quad (48)$$

Elles sont données par l'équation :

$$J_n(x) = \left(\frac{x}{2}\right)^n \sum_{k=0}^{+\infty} \frac{(-1)^k}{k!(n+k)!} \left(\frac{x}{2}\right)^{2k} \quad (n \in \mathbb{N}) \quad (49)$$

Enfin, les fonctions de Bessel peuvent être calculées par récurrence :

$$J_{k+1}(x) = \frac{2k}{x} J_k(x) - J_{k-1}(x) \quad (50)$$

Quelques fonctions de Bessel sont présentées sur la figure 36.

Le spectre d'un signal synthétisé par la méthode FM est constitué d'un partiel central à la fréquence porteuse f_c et d'une infinité de partiels situés aux fréquences $(f_c + kf_m)$, pour tout $k \in \mathbb{Z}$, disposés symétriquement autour de la fréquence centrale. Les partiels sont espacés uniformément en fréquence. Si la fréquence porteuse est un multiple de la fréquence de modulation, le signal généré est alors harmonique, avec la fréquence de modulation comme fréquence fondamentale. La synthèse est alors formantique, et la hauteur perçue correspond à la fréquence de modulation. Ainsi, si $f_c = pf_m$, nous obtenons :

$$s(t) = \sum_{n=-\infty}^{+\infty} J_n(I) \cos(2\pi(p+n)f_m t) \quad (51)$$

Si le signal modulé contient plusieurs partiels, des bandes latérales apparaissent autour de chacun des partiels. Il en résulte un enrichissement considérable du spectre. Les fonctions de Bessel J_n tendent vers 0 quand n tend vers l'infini. En conséquence, l'énergie du spectre est concentrée autour de la fréquence porteuse. Si $I = 0$, l'amplitude de la porteuse est maximale, et il n'y a pas de bande latérale (l'effet de modulation est annulé). En augmentant la valeur de I , l'amplitude de la porteuse diminue, et des bandes latérales apparaissent. La figure 37 montre le spectre résultant d'une modulation de fréquence, pour différentes valeurs de l'indice I . De Poli [Pol83] a estimé que le nombre de composantes d'amplitude significative dans les bandes latérales est fonction de cet indice I et vaut $(I + 1)$. L'indice de modulation contrôle donc le nombre de partiels produits, et l'enveloppe spectrale du signal synthétisé. Avec une fréquence porteuse $f_c = 800$ Hz et une fréquence modulante $f_m = 200$ Hz, les bandes latérales contiennent des composantes de fréquence 600 Hz ($f_c - f_m$), 1000 Hz ($f_c + f_m$), 400 Hz ($f_c - 2f_m$), 1200 Hz ($f_c + 2f_m$), 200 Hz ($f_c - 3f_m$), 1400 Hz ($f_c + 3f_m$), ...

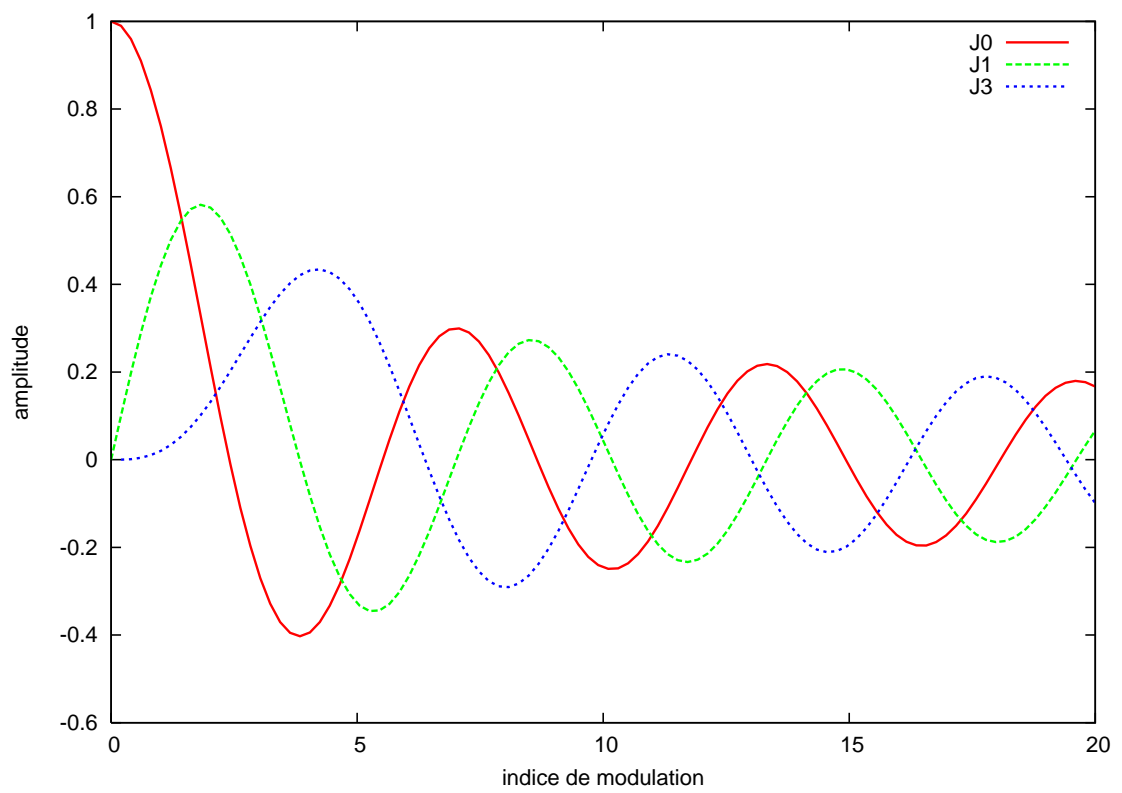


FIG. 36 – Exemples de fonctions de Bessel de première espèce.

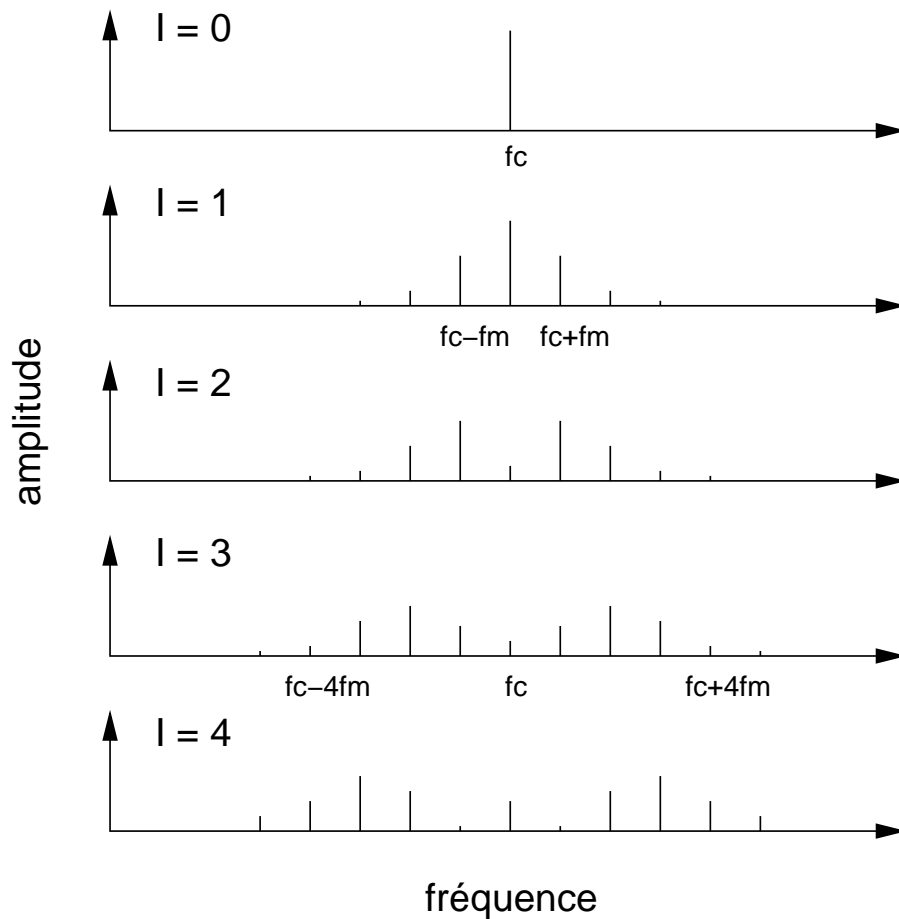


FIG. 37 – Synthèse FM en fonction de l'indice de modulation I . Lorsque $I = 0$, il n'y a pas de modulation. Lorsque I croît, des bandes latérales apparaissent de part et d'autre de la fréquence porteuse. Pour $I = 4$, chaque bande latérale contient $(4 + 1) = 5$ composantes significatives en amplitude. Les autres composantes sont négligeables. Le partiel généré à la fréquence porteuse n'est pas toujours le partiel de plus forte amplitude. (D'après [Cho73]).

La synthèse s'effectue dans le domaine numérique. Il peut y avoir un repliement dans le spectre des partiels de fréquences négatives, et ceux dont la fréquence est au-delà de la fréquence de Nyquist. En raison de cet effet d'*aliasing*, des partiels harmoniques mal atténués dans la partie haute du spectre sont susceptibles de se replier en partiels inharmoniques avec ceux déjà présents dans le spectre. La structure harmonique n'est alors pas conservée. Certains rapports (f_c/f_m), ou des indices de modulation très élevés, génèrent des fréquences de repliement qui viennent s'ajouter au spectre résultant, souvent avec des inversions de phase.

7.2.2 Autres techniques de modulation de fréquence

Il y a plusieurs variations possibles à la méthode FM. Plusieurs signaux porteurs peuvent être utilisés, avec un seul signal modulateur, afin de générer des bandes latérales dans des régions particulières du spectre, et ainsi constituer des formants dans une structure harmonique de même fréquence fondamentale. Voici une formulation de la modulation de fréquences à multiple porteuses, ou MCFM (*Multiple Carrier Frequency Modulation*) :

$$s(t) = a^{w_0} \sin(2\pi f_0 t + (I_1 \sin(2\pi f_m t))) \dots + a^{w_n} \sin(2\pi f_n t + (I_n \sin(2\pi f_m t))) \quad (52)$$

avec a une constante telle que $0 < a \leq 1$, w_n le poids de la $n^{\text{ième}}$ fréquence porteuse, f_0 la fréquence fondamentale du signal harmonique à synthétiser, f_n multiple de f_0 , f_m la fréquence de modulation généralement égale à f_0 , et I_n l'indice de modulation du signal porteur de fréquence f_n .

Cette synthèse est intéressante car elle génère des formants en s'appuyant sur des fréquences porteuses, et ceci avec une complexité limitée. Le signal généré est harmonique, car la fréquence de modulation est la même pour tous les formants. En 1980, Chowning propose ainsi d'utiliser la MCFM pour simuler la voix et compose l'œuvre *Phoné* en 1981. Cette méthode est alors parfaitement adaptée pour créer un signal harmonique dont le spectre est composé de formants, mais cela paraît plus difficile de l'utiliser pour synthétiser le spectre d'un son existant, car il y a de fortes contraintes sur les amplitudes des composantes. La relation entre les amplitudes des partiels générés explique la signature sonore caractéristique de la synthèse FM.

La *feedback FM* est une technique de modulation de fréquence très utilisée, brevetée par Yamaha pour ses synthétiseurs. Le principe est de réaliser une modulation de fréquence en réinjectant une partie du signal produit dans les paramètres de la synthèse. Les variations des amplitudes des partiels du signal générés en fonction de l'indice de modulation sont plus contrôlables qu'avec une synthèse FM classique. De plus, cette méthode de synthèse atténue l'effet sonore caractéristique de la synthèse FM.

7.2.3 Résultats

Il y a deux approches principales pour réaliser une synthèse additive d'un son harmonique avec une technique de modulation de fréquence. Pour reconstituer l'enveloppe du spectre, et plus précisément les formants, nous pouvons utiliser la modulation localement (formant par formant) sur un signal sinusoïdal pur, ou bien globalement sur un signal complexe (par exemple un signal composé des partiels maxima locaux de l'enveloppe spectrale).

Dans le premier cas, il faut changer de fréquence porteuse pour chaque formant. Nous pouvons utiliser la synthèse FM formantique, en utilisant la fréquence fondamentale du signal à synthétiser comme fréquence modulante. Il est préférable de recourir alors à la *feedback FM* qui autorise un meilleur contrôle sur les amplitudes des composantes du son. Mais peu de partiels sont produits si le coefficient de modulation est petit, et la technique manque alors d'intérêt. Si l'indice de modulation est grand, nous ne sommes pas assurés que le partiel généré à la fréquence porteuse soit celui de plus forte amplitude, comme l'illustre la figure 37 avec $I = 4$. Le formant, et donc l'enveloppe spectrale, sont déformés, ils ne correspondent pas au spectre du signal généré linéairement.

Dans le second cas, il s'agit d'effectuer une synthèse FM globale, en utilisant la fréquence fondamentale du signal à synthétiser comme fréquence du signal modulateur. Nous pouvons utiliser la MCFM avec comme signaux porteurs les signaux correspondant aux maxima locaux de l'enveloppe spectrale. Nous pouvons alors observer des inversions de phase qui se produisent quand deux maxima sont trop proches, ce qui entraîne des "trous" dans le spectre. Utiliser les voisins des maxima quand ceux-ci sont trop proches en fréquence donne dans le cas de nos expériences de meilleurs résultats (figure 38). La méthode est cependant empirique et non généralisable.

L'agencement particulier des partiels donne cependant un timbre caractéristique au son produit, ce qui ne permet pas de représenter tous les sons harmoniques. Le résultat de la FM est un signal dont les formants sont centrés sur la fréquence porteuse (si l'indice de modulation est assez petit), les fonctions de Bessel étant symétriques. Or les formants ne sont généralement pas centrés sur un partiel. Ce décalage a un effet sur la perception du son, même s'il est difficile à mesurer. Certaines méthodes permettent de retrouver la fréquence centrale réelle du formant, mais cela nécessite des calculs supplémentaires et la fréquence centrale retrouvée n'est plus multiple de la fréquence fondamentale. C'est également le cas de tous les partiels générés. Le son devient alors inharmonique.

7.3 Distorsion et polynômes

L'illustration la plus populaire des effets de la distorsion non linéaire est celle de l'amplificateur de guitare poussé au-delà de ses capacités d'amplification. Un amplificateur peut amplifier un signal linéairement sans déformer la forme d'onde jusqu'à une certaine limite. Au-delà de cette limite, l'amplificateur sature, ce qui se traduit par un enrichissement du spectre dans les hautes fréquences, c'est une distorsion.

7.3.1 Distorsion

La distorsion non linéaire peut être appliquée à un signal audio. Elle peut aussi être utilisée comme outil de synthèse sonore. Il s'agit de la *Waveshaping Synthesis*, technique de synthèse proposée par Risset [Ris69]. Pour comprendre l'effet de la distorsion, nous pouvons considérer la fonction de transfert de distorsion comme un miroir déformant. Le miroir est incliné avec un certain angle, le sujet reflété (le signal d'entrée) se trouve en-dessous du miroir et l'observateur (le signal de sortie) se trouve en face. L'angle du miroir détermine si l'image est reflétée de façon agrandie ou de façon diminuée. En informatique, cette fonction de transfert est traditionnellement stockée en mémoire dans un tableau.

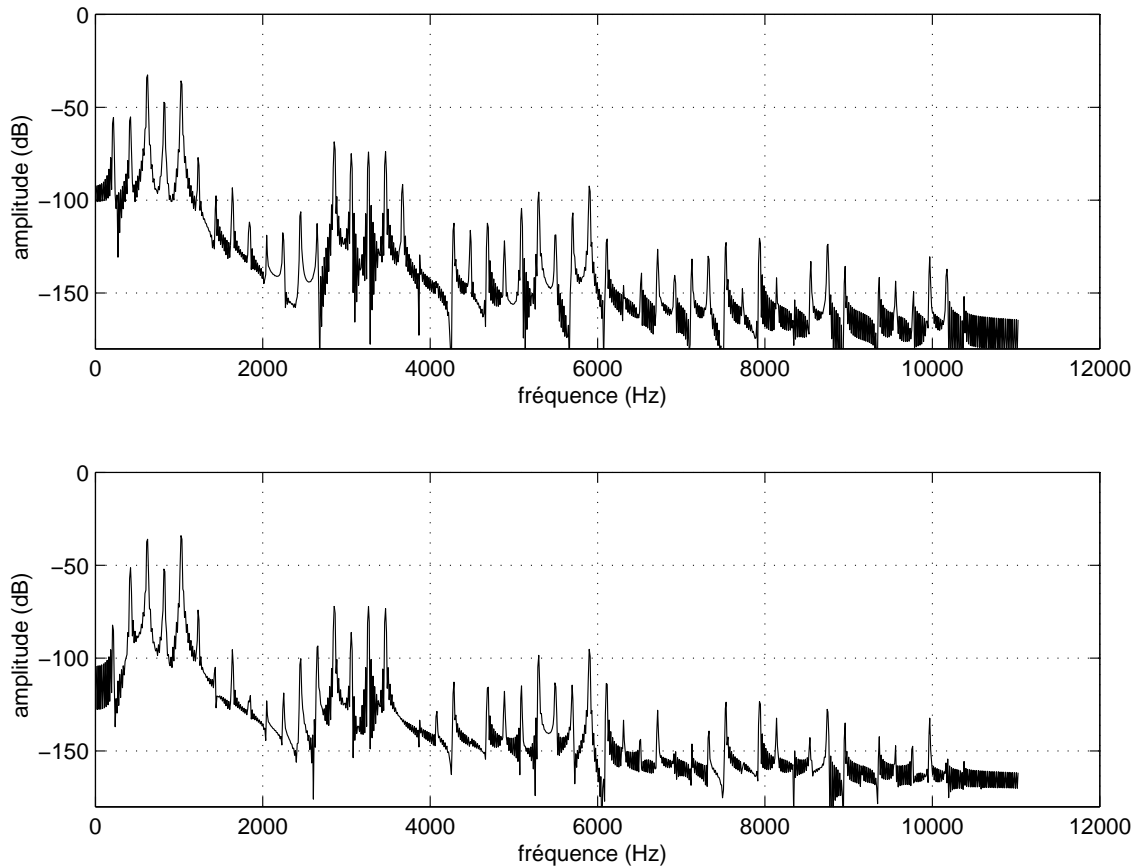


FIG. 38 – Synthèse FM pour un signal de voix. Il y a en haut le spectre du signal de voix généré linéairement, et en bas le spectre du signal résultant de l'application de la FM. La FM est appliquée sur chaque partiel correspondant à un maximum local de l'enveloppe spectrale du signal, sauf si deux partiels ainsi décrits sont trop proches. Dans ce cas, la synthèse FM est appliquée au partiel voisin de fréquence supérieure. Cette méthode empirique permet d'éviter des trous dans le spectre.

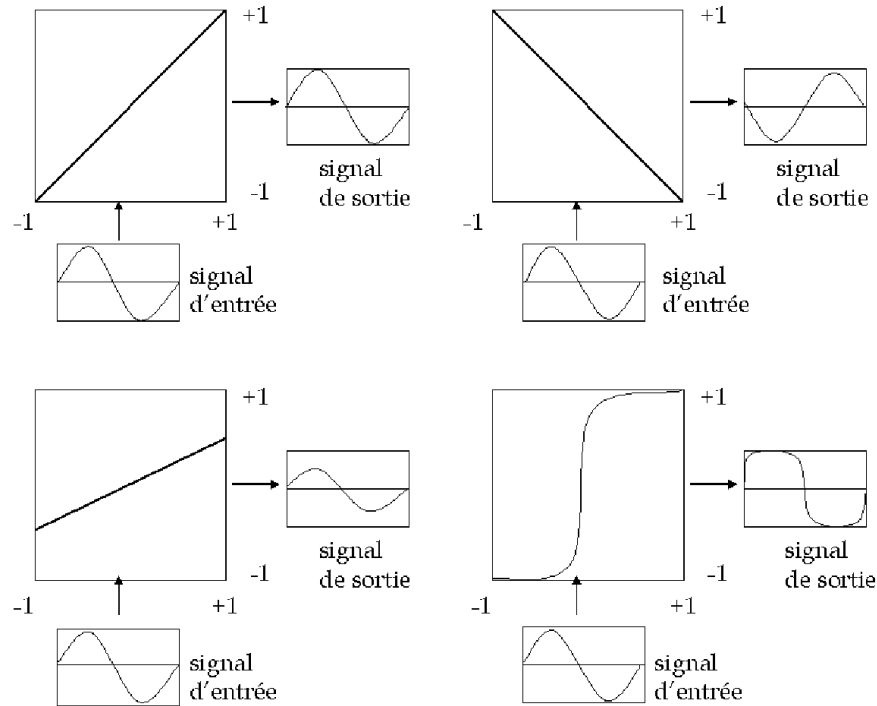


FIG. 39 – Synthèse par distorsion : exemple de quatre fonctions de transfert. (D'après [Roa96]).

La fonction de transfert W détermine donc la forme du signal de sortie. Le contenu spectral du signal dépend de la forme de W et de l'amplitude du signal d'entrée. Si le signal d'entrée a une amplitude maximale (prenant des valeurs entre -1 à $+1$), toute la fonction de transfert est utilisée. Si le signal d'entrée a une amplitude plus faible, seule la portion centrale de la fonction de transfert est alors utilisée. La figure 39 donne l'exemple de quatre fonctions de transfert.

7.3.2 Polynômes de Tchebycheff

La technique de distorsion qui nous intéresse particulièrement est celle introduite par Arfib [Arf79] et Le Brun [Bru79]. Ils ont proposé d'utiliser des fonctions de transfert fondées sur les polynômes de Tchebycheff. Avec un signal d'entrée cosinusoidal et une fonction de transfert composée par des polynômes de Tchebycheff, il est possible de prédire exactement le contenu harmonique du signal de sortie.

Les polynômes de Tchebycheff T_k possèdent la propriété suivante :

$$T_k(\cos(\theta)) = \cos(k\theta) \quad (53)$$

et ils vérifient une relation de récurrence simple :

$$\begin{cases} T_0(x) = 1 \\ T_1(x) = x \\ T_n(x) = 2x \cdot T_{n-1}(x) - T_{n-2}(x) \quad \forall n \geq 2 \end{cases} \quad (54)$$

Les huit premiers polynômes de Tchebycheff sont alors, pour $x = \cos(\theta)$:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \\ T_5(x) &= 16x^5 - 20x^3 + 5x \\ T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1 \\ T_7(x) &= 64x^7 - 112x^5 + 56x^3 - 7x \\ T_8(x) &= 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1 \end{aligned} \quad (55)$$

Avec un seul polynôme de Tchebycheff d'ordre k comme fonction de transfert, la fréquence d'un signal généré est k fois la fréquence du signal sinusoïdal d'entrée, ou signal fondamental. C'est donc la $k^{\text{ième}}$ harmonique, l'harmonique de rang k , qui est générée. La fonction de distorsion est une combinaison linéaire de polynômes de Tchebycheff. Ainsi, si T_k est un polynôme de Tchebycheff de degré k , une fonction de transfert W appliquée à une fonction cosinusoidale peut être décrite par :

$$W = \sum_{k=0}^K \alpha_k T_k(\cos(\theta)) = \sum_{k=0}^K \alpha_k \cos(k \cdot \theta) \quad (56)$$

Voici un exemple de fonction de transfert :

$$W = T_0 + \frac{T_2}{2} + \frac{T_3}{6} \quad (57)$$

W ajoute à un signal fondamental d'amplitude a_0 une seconde harmonique d'amplitude $(a_0/2)$, où a_0 est l'amplitude de la fondamentale, et une troisième harmonique d'amplitude $(a_0/6)$.

Le nombre de multiplications et d'additions nécessaires lors de la synthèse sonore dépend du degré du polynôme choisi comme fonction de transfert. En utilisant l'algorithme de Hörner, d additions et d multiplications sont nécessaires pour calculer un polynôme de degré d . Ce calcul doit être fait pour chaque partiel du son. En effet, si les polynômes de Tchebycheff ont un effet non linéaire sur le signal, la distorsion utilise une combinaison linéaire de polynômes. La synthèse, équivalente en qualité à la synthèse linéaire, n'est alors pas intéressante par sa complexité.

Nous remarquons maintenant que lorsque l'amplitude du signal cosinusoidale est inférieure à 1, plusieurs harmoniques peuvent être générées avec une fonction polynomiale de Tchebycheff. Ainsi, avec un tel polynôme de degré 5, trois composantes harmoniques (dont la fondamentale) ont une amplitude non nulle :

$$T_5(a \cos(\theta)) = a^5 \cos(5\theta) + 5a^3(a^2 - 1) \cos(3\theta) + 5a(a^2 - 1)(2a^2 - 1) \cos(\theta) \quad (58)$$

Comme ce ne sont pas les propriétés de la famille des polynômes de Tchebycheff qui sont utilisées ici, n'importe quel polynôme peut alors être utilisé.

7.3.3 Polynômes quelconques

Nous proposons d'utiliser un polynôme quelconque comme fonction de distorsion. En observant l'équation trigonométrique (39), nous remarquons que l'application d'une fonction polynomiale à un signal sinusoïdal pur produit un spectre harmonique à bande limitée par la valeur de $(d \cdot f)$, avec d le degré du polynôme et f la fréquence du signal d'entrée. Cette propriété est utile pour éviter les repliements spectraux. Nous ne sommes pas limités à l'utilisation d'une famille de polynômes, comme celle de Tchebycheff, et cette liberté nous donne plus de contrôle sur les amplitudes des composantes générées dans le son.

Avec un polynôme comme fonction de distorsion, il y a intermodulation quand le signal d'entrée est complexe. L'intermodulation se produit lorsque deux ou plusieurs composantes fréquentielles interagissent entre elles. Le spectre du signal synthétisé par cette méthode contient non seulement les harmoniques des composantes fréquentielles du signal passé en paramètre, mais également des composantes de fréquences égales aux différentes sommes ou différences des fréquences en entrée (un exemple est donné par la figure 40). Mais le contrôle de cette intermodulation est difficile quand les signaux d'entrée sont complexes. D'autre part, nous remarquons que si la technique est appliquée à un signal simple (une sinusoïde par exemple), seuls des partiels de fréquences supérieures sont ajoutées au signal, ce qui rend impossible la génération de n'importe quel formant.

7.3.4 Résultats

La distorsion utilisant les polynômes de Tchebycheff est une technique non linéaire qui est équivalente en qualité à une synthèse additive linéaire, quand elle prend une seule composante sinusoïdale en entrée, mais elle n'est pas moins complexe (en nombre d'opérations). En utilisant des polynômes quelconques pour effectuer une distorsion, de nombreuses harmoniques sont créées, mais l'intermodulation est très difficilement contrôlable et ne permet pas de former l'enveloppe spectrale souhaitée.

7.4 Des fonctions remarquables

Nous avons étudié dans cette partie des fonctions dont les propriétés mathématiques offrent un intérêt dans le cadre de la synthèse sonore des sons harmoniques.

7.4.1 Valeur absolue

Dans le domaine de la synthèse sonore, la fonction valeur absolue est principalement utilisée pour étendre des signaux téléphoniques à bande de fréquence limitée. Dans ce contexte, Makhoul [MB79] propose d'utiliser la fonction :

$$s(t) = \frac{1}{2} ((1 + \alpha)|x(t)| + (1 - \alpha)x(t)), \text{ pour } 0 \leq \alpha \leq 1 \quad (59)$$

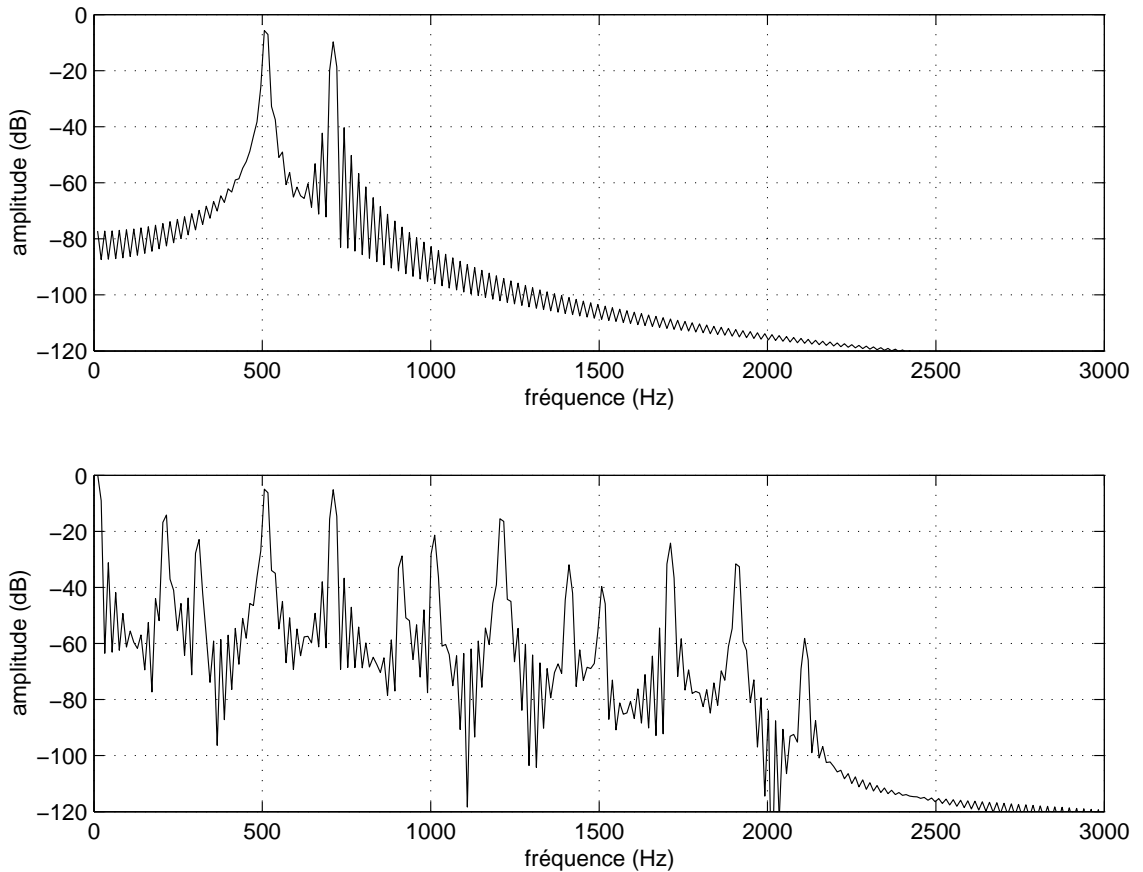


FIG. 40 – Le signal dont le spectre est représenté en haut est constitué de deux composantes sinusoïdales aux fréquences $f_a = 500$ Hz et $f_b = 700$ Hz. Le spectre représenté en bas est celui du même signal auquel a été appliqué la fonction non linéaire $g(x) = x^2 + x^3$. Le degré 2 de la fonction polynomiale génère des composantes de fréquence $2f_a$ et $2f_b$, ainsi que des composantes en $f_a \pm f_b$. Le degré 3 de la fonction génère des composantes en $3f_a$ et $3f_b$, ainsi que des composantes en $2f_a \pm f_b$ et $f_a \pm 2f_b$. À partir des deux fréquences f_a et f_b , 4 composantes multiples de ces deux fréquences sont générées et 6 composantes d'intermodulation et une valeur constante.

	0	f	$2 \cdot f$	$4 \cdot f$	$6 \cdot f$	$8 \cdot f$	$10 \cdot f$
0	0	0	0	0	0	0	0
0.1	1.29	0	1.92	0.76	0.13	0	0
0.2	2.61	0	3.87	1.51	0.25	0	0
0.3	3.95	0	5.82	2.25	0.38	0	0
0.4	5.25	0	7.71	2.97	0.50	0	0
0.5	6.43	0	9.42	3.61	0.62	0	0
0.6	7.36	0	10.75	4.12	0.72	0	0
0.7	7.79	0	11.34	4.33	0.78	-0.01	0
0.8	7.30	0	10.54	4.00	0.73	-0.02	0
0.9	5.22	0	7.38	2.71	0.51	-0.04	0.01
1	0.62	0	0.42	-0.10	0.05	-0.03	0.03

TAB. 8 – Approximation de $|x|$ par un polynôme de degré 10 en fonction de l'amplitude a , pour $x = a \cos(2\pi ft)$. Les composantes fréquentielles du son généré sont des harmoniques de la fréquence f . Les amplitudes des composantes sont données dans le tableau en fonction de leur fréquence (lignes) et de la valeur de l'amplitude a (colonnes). Nous remarquons que des valeurs d'amplitudes sont attribuées à la fréquence 0, ce qui est sans incidence sur la perception du son généré.

et Weinstein [Wei75] propose la formule suivante pour l'enrichissement de signaux à bande limitée :

$$s(t) = \begin{cases} x(t), & x \geq 0 \\ (1/2)|x(t)|, & x < 0 \end{cases} \quad (60)$$

L'intérêt de la fonction valeur absolue est explicité par Collen [Col02]. Il indique que la fonction $|x|$ peut être approchée sur $[-1; 1]$ par une série de Taylor d'ordre 10, c'est-à-dire un polynôme de degré 10 tel que :

$$|x| \approx 4.77x^2 - 18.8x^4 + 40.43x^6 - 40x^8 + 14.6x^{10} \quad (61)$$

Ainsi, quand $x = \cos(\theta)$, l'équation précédente devient :

$$|\cos(\theta)| \approx 0.625 + 0.425 \cos(2 \cdot \theta) - 0.975 \cos(4 \cdot \theta) + \quad (62)$$

$$0.466 \cos(6 \cdot \theta) - 0.273 \cos(8 \cdot \theta) + 0.285 \cos(10 \cdot \theta) \quad (63)$$

Pour une sinusoïde passée en entrée, 5 composantes sinusoïdales sont générées. Les valeurs de leurs amplitudes sont données en fonction de l'amplitude du signal d'entrée par le tableau 8. La constante que produit un développement limité déplace le signal temporel qui n'est plus centré en 0. Cela n'a pas d'effet sur la perception du son par l'oreille humaine mais peut amener une saturation. Il faut alors normaliser la sortie du calcul, en déduisant la constante de tous les échantillons du son.

La valeur absolue d'un signal sinusoïdal pur génère des harmoniques de rang pair, dont les amplitudes dépendent des valeurs de l'amplitude de la sinusoïde en entrée. Cela pose deux problèmes pour la synthèse de formants : d'une part les harmoniques ne sont pas toutes présentes (il faut utiliser la moitié de la fréquence fondamentale dans ce cas), et d'autre part les fréquences générées sont toutes supérieures à la fréquence de la sinusoïde

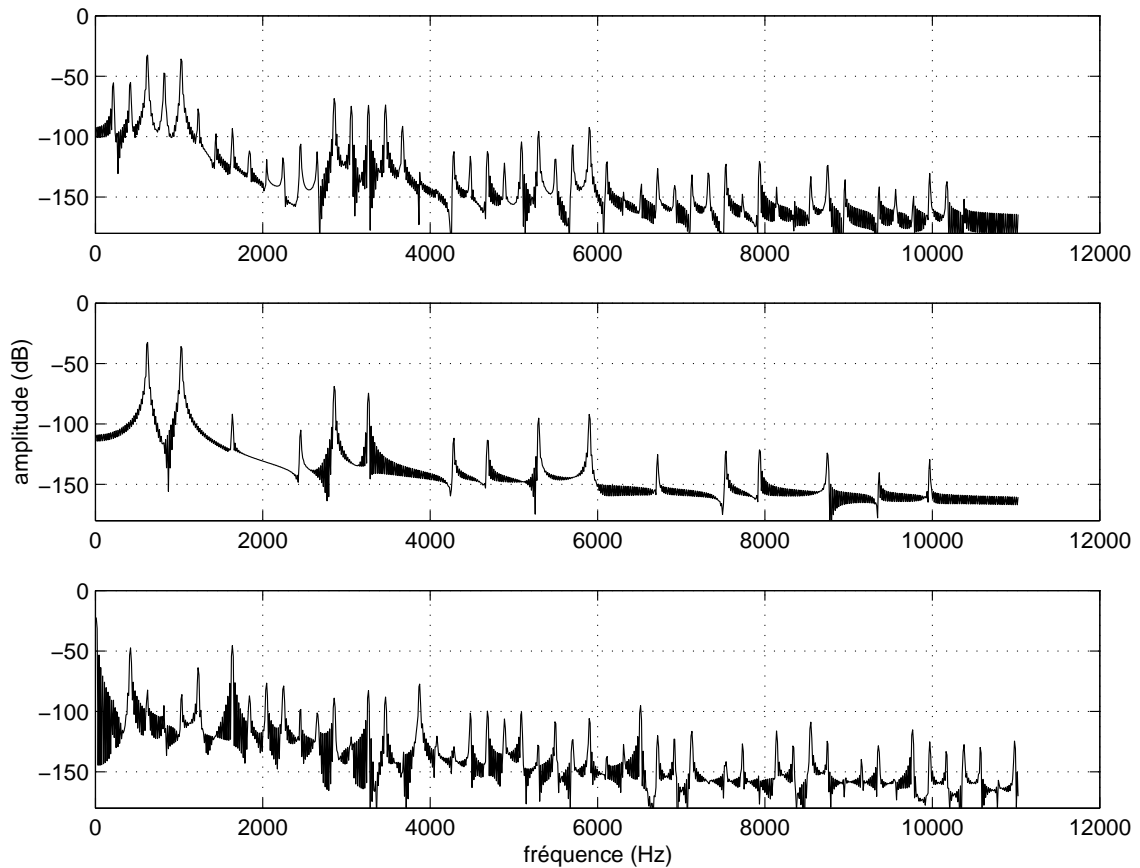


FIG. 41 – Application de la fonction $|x|$ sur les maxima de l'enveloppe spectrale d'un signal. Il y a en haut le spectre d'un signal de voix obtenu par synthèse linéaire, au milieu les maxima locaux de l'enveloppe de ce spectre. En bas se trouve le spectre du signal obtenu par l'application d'une valeur absolue sur le signal composé des partiels maxima. La structure harmonique du signal est préservée, le spectre est enrichi de façon significative. Les deux premiers formants sont perdus, l'enveloppe spectrale n'est pas la même qu'avec une synthèse linéaire.

en entrée, avec des amplitudes décroissantes. Or la plupart des formants ne peuvent pas être générés avec seulement des partiels décroissants en amplitude.

Nous pouvons appliquer la valeur absolue sur un signal complexe. La figure 41 donne des résultats sur un signal composé des partiels maxima de l'enveloppe spectrale du signal à synthétiser, et la figure 42 sur le signal à synthétiser limité en haute fréquence. Le principal résultat est l'enrichissement considérable du spectre, avec conservation de la structure harmonique du son. Mais l'enveloppe spectrale n'est pas préservée, et le contrôle des amplitudes des partiels générés est très faible.

7.4.2 Développements limités

Les propriétés non linéaires de la fonction valeur absolue, dont la conservation de la structure harmonique, sont liées au développement limité de cette fonction en 0. Nous

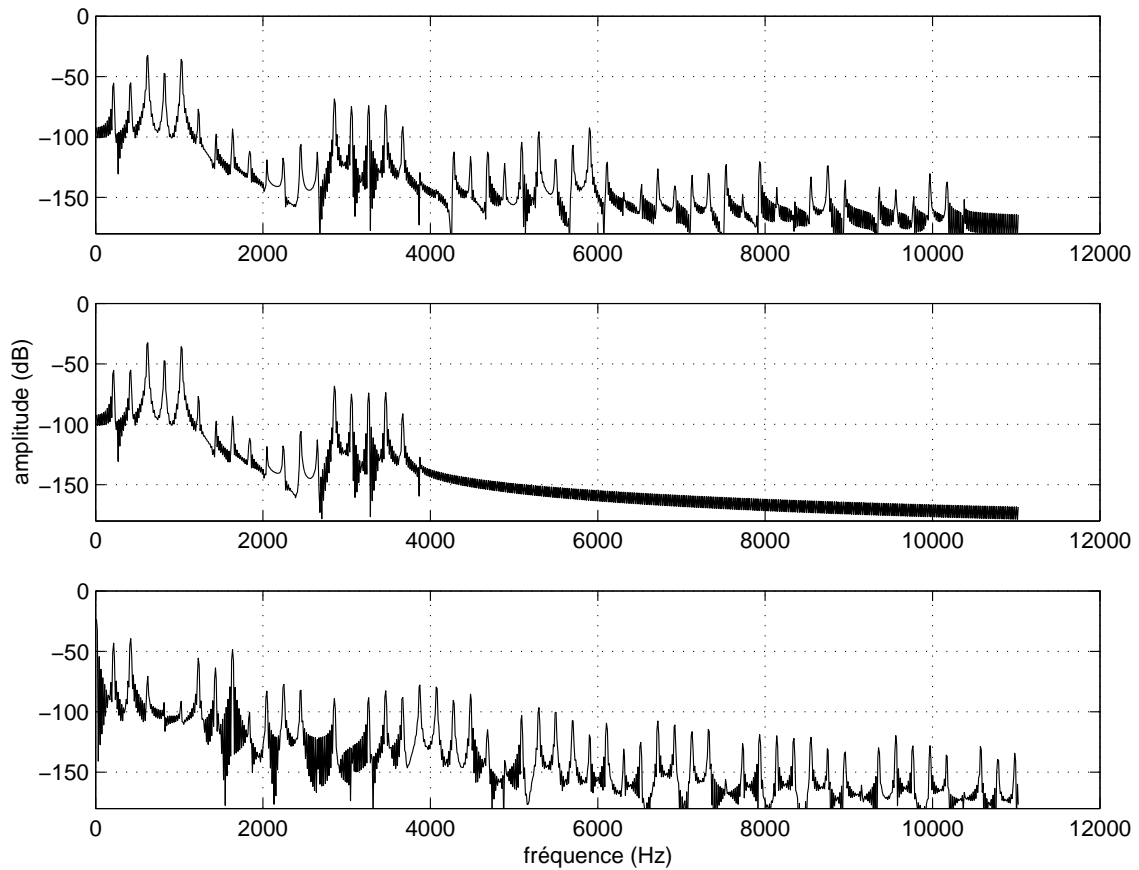


FIG. 42 – *Extension d'un signal à bande limitée par la fonction $|x|$. Il y a en haut le spectre d'un signal de voix obtenu par synthèse linéaire, puis au milieu le spectre du même signal limité en haute fréquence à 4000 Hz. En bas se trouve le spectre du signal obtenu par l'application d'une valeur absolue sur le signal à bande limitée. L'extension du signal ne crée aucune cassure dans le spectre à la fréquence de coupure et conserve la structure harmonique. L'enveloppe spectrale n'est cependant pas respectée et les formants dont la fréquence centrale est supérieure à la fréquence de coupure sont perdus. Des partiels de haute fréquence sont générés alors qu'ils n'apparaissent pas dans le spectre du signal généré linéairement. Cela entraîne des repliements haute fréquence et donc un bruit, même si la faible valeur des amplitudes de ces partiels limite cet effet.*

avons alors étudié d'autres fonctions, dont le développement limité en 0 peut être utilisé en synthèse sonore.

Commençons par définir le développement limité d'une fonction en série de Taylor. Soit $g : I \rightarrow \mathbb{R}$ une fonction n -fois dérivable et soit x_0 un point quelconque de l'intervalle I . Alors la fonction g possède un développement limité à l'ordre n au point x_0 tel que :

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \frac{g''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{g^n(x_0)}{n!}(x - x_0)^n + (x - x_0)^n \varepsilon(x) \quad (64)$$

avec $\lim_{x \rightarrow x_0} \varepsilon(x) = 0$.

Pour $g(x) = \ln(1 + x)$, avec $g :]-1; 1[\rightarrow \mathbb{R}$, nous avons donc :

$$\ln(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \dots + \frac{(-1)^{n-1}}{n}x^n + x^n \varepsilon(x) \quad (65)$$

avec $\lim_{x \rightarrow x_0} \varepsilon(x) = 0$.

Si $s(t) = \ln(1 + x(t))$ et $x(t) = a \cos(2\pi f_0 t)$, nous obtenons :

$$s(t) = a \cos(2\pi f_0 t) - \frac{a^2}{2} \cos^2(2\pi f_0 t) + \frac{a^3}{3} \cos^3(2\pi f_0 t) + \dots \quad (66)$$

En utilisant les propriétés trigonométriques suivantes :

$$\cos^2(x) = \frac{1}{2}(1 + \cos(2x)) \quad (67)$$

$$\cos^3(x) = \frac{3}{4} \cos(x) + \frac{1}{4} \cos(3x) \quad (68)$$

$$\cos^4(x) = \frac{3}{8} + \frac{1}{2} \cos(2x) + \frac{1}{8} \cos(4x) \quad (69)$$

l'équation (66) peut être développée et devient :

$$s(t) = -\left(\frac{a^2}{4} + \frac{3a^4}{32} + \dots\right) + \left(a + \frac{a^3}{4} + \dots\right) \cos(2\pi f_0 t) - \left(\frac{a^2}{4} + \frac{a^4}{8} + \dots\right) \cos(4\pi f_0 t) + \left(\frac{a^3}{12} + \dots\right) \cos(6\pi f_0 t) + \dots \quad (70)$$

Nous voyons alors apparaître des composantes harmoniques de la fréquence f_0 . Cette propriété mathématique de la fonction peut être utilisée comme méthode de synthèse. Le tableau 9 donne ainsi différentes amplitudes possibles pour les composantes harmoniques générées, en fonction de l'amplitude du signal d'entrée.

La fonction $\ln(1 + x)$ génère de nombreuses composantes sinusoïdales nouvelles, mais avec le phénomène de repliement spectral, des partiels non multiples de la fréquence fondamentale peuvent apparaître, et la structure harmonique du signal n'est alors pas préservée. Afin de limiter ce phénomène, il faut plutôt choisir des fonctions non linéaires qui ont un développement en série de Taylor d'ordre limité, dont les amplitudes ne sont plus significatives à partir d'un certain ordre du développement. Pour faire ce choix, le tableau 10

	const	f	$2 \cdot f$	$3 \cdot f$	$4 \cdot f$	$5 \cdot f$
0	0	0	0	0	0	0
0.1	0	0.10	0	0	0	0
0.2	-0.01	0.20	-0.01	0	0	0
0.3	-0.02	0.31	-0.02	0	0	0
0.4	-0.04	0.42	-0.04	0.01	0	0
0.5	-0.07	0.54	-0.07	0.01	0	0
0.6	-0.10	0.66	-0.11	0.02	-0.01	0
0.7	-0.15	0.81	-0.16	0.04	-0.01	0
0.8	-0.21	0.97	-0.23	0.06	-0.02	0
0.85	-0.25	1.06	-0.28	0.08	-0.03	0.01
0.9	-0.29	1.16	-0.33	0.10	-0.04	0.01
0.95	-0.34	1.26	-0.38	0.12	-0.05	0.01
0.99	-0.38	1.35	-0.44	0.14	-0.06	0.01

TAB. 9 – Développement limité en 0 d'ordre 5 de la fonction $\ln(1+a \cos(2\pi ft))$ en fonction de l'amplitude a du signal d'entrée. Les composantes fréquentielles du son généré sont des harmoniques de la fréquence f (lignes). Les amplitudes des composantes sont données dans le tableau en fonction de leur fréquence et de la valeur de l'amplitude a (colonnes).

donne les expressions des développements limités en 0 des principales fonctions usuelles. Regardons par exemple le développement limité en 0 de la fonction $g(x) = 1/(1-x)$:

$$g(x) = 1 + x + x^2 + x^3 + \dots + x^n + x^n \varepsilon(x) \quad (71)$$

avec $\lim_{x \rightarrow +\infty} \varepsilon(x) = 0$.

Cette expression permet de réaliser un peigne harmonique de partiels d'amplitude 1 qui pourrait ensuite être filtré pour former l'enveloppe voulue. Une infinité de composantes sinusoïdales sont générées à partir d'une seule sinusoïde. Mais comme il n'y a pas de décroissance d'amplitude dans les différents termes du développement, il y a un repliement haute fréquence infini et c'est en réalité un bruit blanc qui est synthétisé (plus précisément une séquence d'impulsions, le spectre considéré étant plat mais déterministe).

Les développements limités peuvent également se composer, dans le but de sculpter le spectre du signal généré. La somme, la multiplication ou le quotient (sous réserve de non-nullité du dénominateur) de deux fonctions qui possèdent un développement limité en 0 est une fonction qui possède également un développement limité en 0. Cela peut permettre d'obtenir un développement limité qui respecte une certaine pente de formant ou une forme d'enveloppe spectrale. Le contrôle des amplitudes des composantes sinusoïdales est cependant assez difficile, encore plus si la fonction ainsi construite est appliquée à un signal d'entrée complexe.

7.4.3 Résultats

Les méthodes de synthèse utilisant les développements limités des fonctions ont deux avantages principaux. Elles ne sont pas complexes, l'application d'une fonction simple sur un signal sinusoïdal pur peut créer un nombre important de nouveaux partiels dans le

$$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + x^n \varepsilon(x)$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \dots + (-1)^n \frac{x^{2n}}{(2n)!} + x^{2n+1} \varepsilon(x)$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + (-1)^{n-1} \frac{x^{2n-1}}{(2n-1)!} + x^{2n} \varepsilon(x)$$

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \dots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} x^n + x^n \varepsilon(x)$$

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots + (-1)^n x^n + x^n \varepsilon(x)$$

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots + (-1)^{n-1} \frac{1}{n} x^n + x^n \varepsilon(x)$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots + x^n + x^n \varepsilon(x)$$

$$\ln(1-x) = -x - \frac{1}{2}x^2 - \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots - \frac{1}{n} x^n + x^n \varepsilon(x)$$

$$\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots + \frac{x^{2n}}{(2n)!} + x^{2n+1} \varepsilon(x)$$

$$\sinh(x) = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + \frac{x^{2n-1}}{(2n-1)!} + x^{2n} \varepsilon(x)$$

$$\arctan(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \dots + \frac{x^{2n-1}}{(2n-1)!} + x^{2n} \varepsilon(x)$$

$$\frac{1}{\sqrt{1-x^2}} = 1 + \frac{1}{2}x^2 + \frac{1.3}{2.4}x^4 + \dots + \frac{1.3.5\dots(2n-1)}{2.4.6\dots(2n)} x^{2n} + x^{2n+1} \varepsilon(x)$$

$$\arcsin(x) = x + \frac{1}{2} \frac{x^3}{3} + \frac{1.3}{2.4} \frac{x^5}{5} + \dots + \frac{1.3.5\dots(2n-1)}{2.4.6\dots(2n)} \frac{x^{2n+1}}{2n+1} + x^{2n+2} \varepsilon(x)$$

$$\arccos(x) = \frac{\pi}{2} - x - \frac{1}{2} \frac{x^3}{3} - \frac{1.3}{2.4} \frac{x^5}{5} - \dots - \frac{1.3.5\dots(2n-1)}{2.4.6\dots(2n)} \frac{x^{2n+1}}{2n+1} + x^{2n+2} \varepsilon(x)$$

TAB. 10 – Développements limités en 0 des principales fonctions usuelles.

son. Elles conservent également la structure harmonique du signal, si les amplitudes des composantes générées sont maîtrisées. C'est pourquoi elles ont été utilisées dans le cadre d'extension de bande téléphonique. Mais dans notre cadre de recherche, les séries de Taylor ne sont pas assez contrôlables pour effectuer une synthèse de qualité comparable à une synthèse additive linéaire.

7.5 Formes analytiques

Moorer [Moo76] propose d'utiliser les formules de sommation (DSF pour *Discrete Summation Formula*), ou formes analytiques, comme méthode de synthèse. En utilisant la relation :

$$2i \sin(x) = e^{ix} - e^{-ix} \quad (72)$$

et en appliquant la forme analytique que nous connaissons pour les suites géométriques :

$$\sum_{k=1}^{K-1} z^k = \frac{1 - z^K}{1 - z} \quad (z \neq 1) \quad (73)$$

Nous obtenons des formes analytiques trigonométriques. Nous en retiendrons tout d'abord deux :

$$\sum_{k=1}^n \sin(k\theta) = \frac{\sin((n+1)\frac{\theta}{2}) \sin(n\frac{\theta}{2})}{\sin(\frac{\theta}{2})} \quad (74)$$

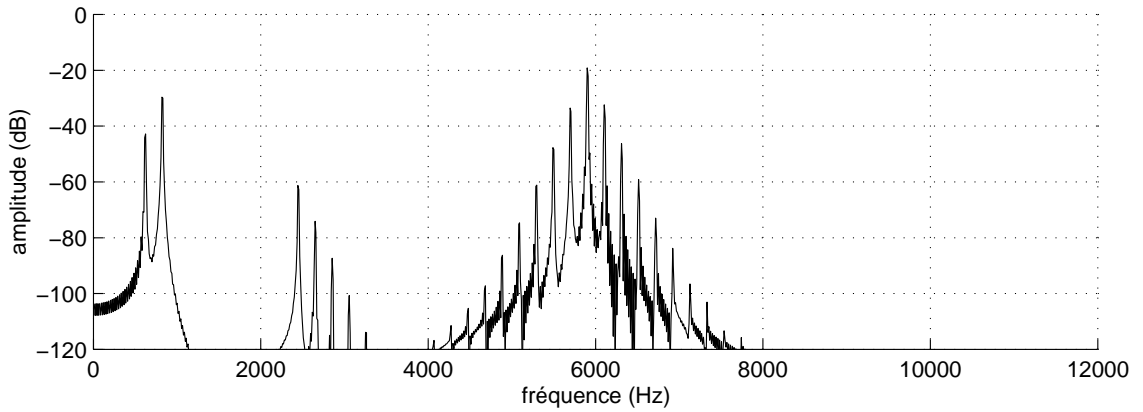
$$\sum_{k=0}^{n-1} a^k \sin(\theta + k\beta) = \frac{\sin(\theta) - a \sin(\theta - \beta) - a^n \sin(\theta + n\beta) + a^{n+1} \sin(\theta + (n-1)\beta)}{1 - 2a \cos(\beta) + a^2} \quad (75)$$

7.5.1 Peigne harmonique

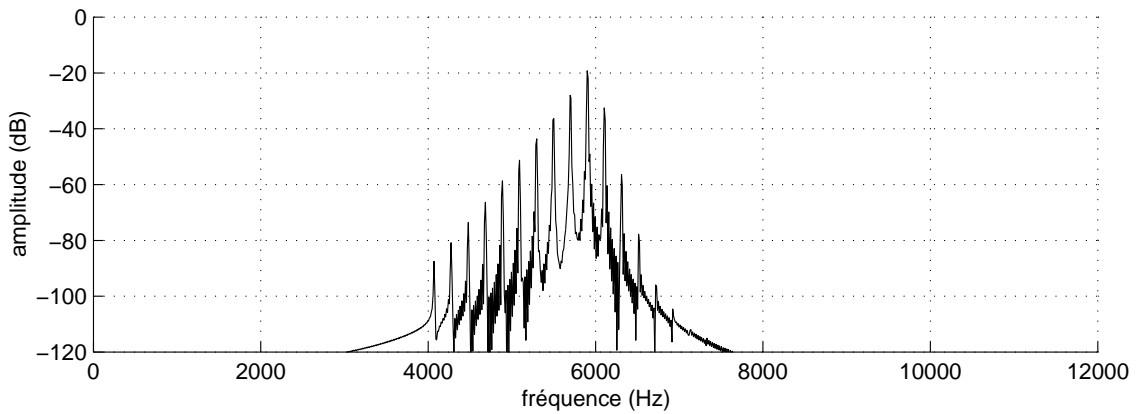
La formule (74) est la plus simple et nous intéresse pour plusieurs raisons. Seulement trois fonctions sinus sont nécessaires pour générer un nombre choisi de partiels (potentiellement très grand), et contrôler la bande de fréquence produite. En choisissant $\theta = 2\pi f_0$, le spectre du signal généré est un peigne harmonique de fréquence fondamentale f_0 , c'est à dire une série de partiels de même amplitude à des fréquences multiples de f_0 . Cette méthode ne permet pas de contrôler l'enveloppe spectrale du signal synthétisé. Le signal généré peut cependant être filtré, afin d'obtenir l'enveloppe souhaitée, ce qui peut s'avérer moins complexe qu'une synthèse additive linéaire.

7.5.2 Synthèse DSF

La formule (75) peut être utilisée en synthèse sonore. θ détermine la fréquence minimale générée, et N fixe le nombre de partiels. Quel que soit N , il n'y a toujours que 5 sinus à évaluer. C'est une technique qui peut être appliquée formant par formant, avec possibilité de représenter des formants asymétriques (figure 43). Deux synthèses DSF sont nécessaires par formant (pente gauche et droite du formant séparément). Les principales formules analytiques utilisables pour la synthèse DSF sont données dans le tableau 11.



(a)



(b)

FIG. 43 – Exemples de synthèse DSF. (a) 4 formes analytiques sont utilisées, 1 par pente de formant. Suivant les paramètres de la synthèse, et avec une complexité constante, nous pouvons choisir de générer la pente gauche d'un formant avec 2 partiels, la droite avec 4, ou synthétiser la pente gauche et la pente droite d'un même formant. (b) Exemple de synthèse d'un formant non symétrique.

$$\sum_{k=1}^n \sin(k\theta) = \frac{\sin((n+1)\frac{\theta}{2}) \sin(n\frac{\theta}{2})}{\sin(\frac{\theta}{2})}$$

$$\sum_{k=1}^n \cos(k\theta) = \frac{\cos((n+1)\frac{\theta}{2}) \sin(n\frac{\theta}{2})}{\sin(\frac{\theta}{2})}$$

$$\sum_{k=0}^{n-1} \sin(\theta + k\beta) = \frac{\sin(\theta + (n-1)\frac{\beta}{2}) \sin(n\frac{\beta}{2})}{\sin(\frac{\beta}{2})}$$

$$\sum_{k=0}^{n-1} \cos(\theta + k\beta) = \frac{\cos(\theta + (n-1)\frac{\beta}{2}) \sin(n\frac{\beta}{2})}{\sin(\frac{\beta}{2})}$$

$$\sum_{k=0}^{n-1} a^k \cos(k\theta) = \frac{(1 - a \cos(\theta))(1 - a^n \cos(n\theta)) + a^{n+1} \sin(\theta) \sin(n\theta)}{1 - 2a \cos(\theta) + a^2}$$

$$\sum_{k=0}^{n-1} a^k \sin(\theta + k\beta) = \frac{\sin(\theta) - a \sin(\theta - \beta) - a^n \sin(\theta + n\beta) + a^{n+1} \sin(\theta + (n-1)\beta)}{1 - 2a \cos(\beta) + a^2}$$

TAB. 11 – Principales formules analytiques utilisables pour la synthèse DSF.

7.5.3 Résultats

La synthèse DSF est une technique très intéressante car :

- le signal produit est à bande limitée, ce qui évite les repliements et contribue à conserver la structure harmonique du signal,
- le nombre d'opérations à effectuer dépend du nombre de formants que l'on veut synthétiser et non du nombre de partiels du signal,
- les formants du spectre du signal synthétisé peuvent être asymétriques.

Cependant les conclusions sont proches de celles établies pour la synthèse FM. Le timbre du son produit est caractéristique de la méthode : le signal a une enveloppe spectrale formée par des triangles (en échelle dB). Cette synthèse ne peut pas représenter aussi bien que la synthèse linéaire tout type de signal harmonique.

7.6 Conclusion

Nous avons répertorié des techniques non linéaires susceptibles de remplacer avantageusement la synthèse additive linéaire. Elles sont toutes capables de générer un signal harmonique, mais offrent peu de contrôle sur les amplitudes des composantes sinusoïdales générées. Pour synthétiser un signal, il y a deux principales méthodes non linéaires :

- soit synthétiser une partie du signal de façon linéaire – les maxima locaux de l'enveloppe spectrale du signal, ou une bande limitée en haute fréquence par exemple – et appliquer ensuite une méthode non linéaire d'enrichissement de ce signal,
- soit synthétiser le signal de façon locale au niveau spectral, en le générant formant par formant. Les composantes sinusoïdales d'amplitudes maximales dans l'enveloppe spectrale peuvent être choisies individuellement comme signal d'entrée. Cela nécessite l'application d'une technique non linéaire par formant (éventuellement des techniques différentes pour chacun des formants).

Dans le premier cas, la structure harmonique du signal est conservée, mais le contrôle des amplitudes des différentes composantes du signal n'est pas assez important pour arriver à un résultat satisfaisant, proche du signal généré linéairement. Les techniques d'enrichissement global ou d'extension de bande limitée sont donc à exclure. Il reste alors la possibilité de reconstituer l'enveloppe spectrale du son formant par formant. Nous avons expliqué que les techniques utilisant des polynômes sont trop complexes, et que les synthèses utilisant les développements limités en série de Taylor ne peuvent synthétiser que la pente droite des formants. Il reste deux types de synthèse étudiés : la synthèse par modulation et la synthèse DSF. Ces deux méthodes peuvent répondre partiellement à notre recherche : ce sont des techniques rapides, que nous pouvons centrer en fréquence (une production de bandes latérales est possible à gauche et à droite d'un partiel dans le spectre du son). Elles permettent donc de synthétiser des formants, afin d'obtenir l'enveloppe spectrale souhaitée pour le son généré. Mais les sons produits par ces techniques sont malheureusement caractéristiques de leurs fonctions mathématiques génératrices, elles ne peuvent pas représenter la diversité des sons harmoniques. Si pour un son particulier nous pouvons, après expérimentation, trouver des paramètres non linéaires efficaces pour la synthèse, la méthode ne peut être généralisée.

Nous avons vu qu'il existe une technique non linéaire de synthèse rapide, utilisant les formes analytiques, capable de synthétiser efficacement un peigne harmonique de sinus. Il

serait intéressant d'appliquer à ce signal des coefficients LPC (*Linear Predictive Coding*), de manière à sculpter le spectre du signal jusqu'à obtenir l'enveloppe spectrale souhaitée. Le nombre de coefficients utilisés déterminerait alors la complexité de la synthèse, et pourrait arbitrer le compromis entre le degré de dégradation et la rapidité de la synthèse.

En conclusion, si les techniques de synthèse non linéaires répertoriées sont capables de générer efficacement des sons avec des spectres riches, en conservant une structure harmonique, et si certaines d'entre elles sont avantageuses en complexité, elles ne peuvent cependant pas être comparées à une synthèse additive classique. Le spectre du signal généré par une méthode non linéaire est en effet déterminé par des règles mathématiques qui contrôlent son enveloppe. Chaque type de synthèse est lié à un son qui le caractérise. Cette signature sonore rend difficile la synthèse d'un son harmonique quelconque.

Chapitre 8

Synthèse additive polynomiale

Nous présentons dans ce chapitre une nouvelle méthode rapide de synthèse sonore additive appelée PASS (*Polynomial Additive Sound Synthesis*) qui utilise des polynômes pour approcher le signal des oscillateurs. Les fonctions sinusoïdales de chaque oscillateur sont approchées par morceaux polynomiaux, et nous devons gérer dans la méthode les changements fréquents des coefficients polynomiaux. Une structure de données efficace est alors proposée pour gérer ces changements et la production des échantillons du son.

La section 8.1 rappelle que des polynômes ont déjà été utilisés pour la synthèse additive, mais pour modéliser l'évolution des paramètres sonores. Nous détaillons ensuite le fonctionnement de la méthode PASS dans la section 8.2, et nous comparons les performances de cette nouvelle méthode avec celles du résonateur numérique présenté dans la section 2.4.1. La section 8.3 propose alors de combiner les avantages du PASS et de la méthode du résonateur dans une méthode hybride. Enfin, nous proposons dans la section 8.4 une application à la méthode PASS, pour synthétiser de façon réaliste et en temps réel des surfaces océaniques.

8.1 Évolution polynomiale des paramètres sonores

Les polynômes sont traditionnellement utilisés pour modéliser les paramètres du modèle sinusoïdal. Ainsi des polynômes de degré 3 servent à reconstruire les phases dans la méthode classique de McAulay-Quatieri [MQ86]. Ding et Qian [DQ97] proposent d'utiliser plutôt des polynômes de degré 4, et Girin *et al.* [GMdM⁺03] comparent les performances des méthodes en fonction de ce degré, toujours pour reconstruire le paramètre de phase, allant du degré 1 au degré 5.

Raspaud *et al.* [RMG05] proposent le modèle Poly-Sin : les paramètres sonores sont fonction d'une somme d'un polynôme et de fonctions sinusoïdales. La partie polynomiale modélise les variations lentes des paramètres – l'enveloppe –, tandis que la somme de sinusoïdes modélise les variations rapides. Le but n'est pas de rendre la synthèse sonore plus rapide, mais d'avoir une meilleure analyse des paramètres et de permettre des transformations sonores nouvelles.

Nous proposons d'utiliser les polynômes pour remplacer les signaux sinusoïdaux de chaque partiel – et non leurs paramètres séparément – dans un objectif de synthèse rapide.

8.2 Polynomial Additive Sound Synthesis (PASS)

Nous présentons une méthode de synthèse sonore rapide qui utilise des polynômes. Avec la synthèse additive classique, chaque oscillateur produit un échantillon, et tous les échantillons sont additionnés pour obtenir un échantillon du son. Le temps pris par la synthèse est alors proportionnel au produit du nombre d'oscillateurs par la fréquence d'échantillonnage. Nous proposons d'utiliser des polynômes pour remplacer l'utilisation des fonctions sinusoïdales. Il s'agit de trouver pour chaque partiel un nombre fini de polynômes qui l'approchent sur une partie périodique. Cette partie peut être une période du signal d'un partiel, mais également plusieurs périodes, ce qui est évidemment toujours périodique. Dans la suite de cette présentation, nous supposons néanmoins que la partie périodique est une période du signal d'un partiel. Nous utilisons alors les propriétés des polynômes pour éviter l'évaluation de chaque polynôme individuellement. Pour calculer la valeur d'un échantillon du signal, il suffit d'additionner les coefficients polynomiaux de même degré et d'évaluer ainsi un seul polynôme, appelé générateur. L'addition de polynômes est un polynôme de degré égal au maximum des degrés des polynômes additionnés. Le générateur est alors un polynôme de faible degré, indépendant du nombre d'oscillateurs dans le son. Le principe général de la méthode est illustré par la figure 44.

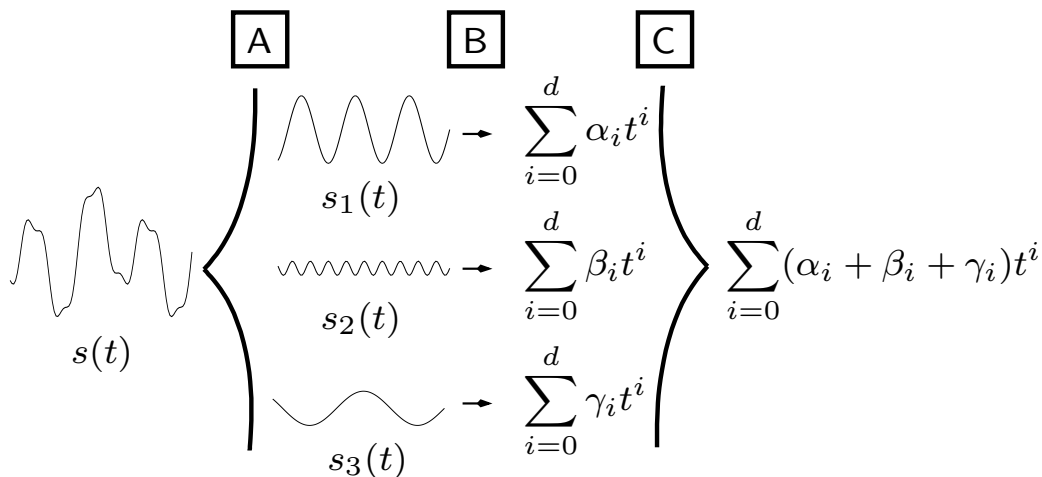


FIG. 44 – PASS. A : Un signal périodique peut être divisé en signaux sinusoïdaux élémentaires, les partiels (théorème de Fourier). B : Pour chaque partiel, un jeu de coefficients polynomiaux est calculé, pour approcher le signal sur une partie périodique. C : Les coefficients du polynôme générateur sont les sommes des coefficients des polynômes approchant les partiels. Les valeurs du générateur sont les échantillons du son synthétisé.

8.2.1 Approximation polynomiale des fonctions sinusoïdales

Le signal temporel $s_i(t)$ généré par un partiel i est défini par une fonction sinusoïdale, avec a_i , f_i et ϕ_i respectivement l'amplitude, la fréquence et la phase initiale du partiel i .

$$s_i(t) = a_i \sin(2\pi f_i t + \phi_i) \quad (76)$$

Nous proposons d'approcher cette fonction s_i par des polynômes. Pour calculer les coefficients polynomiaux de chaque partiel, nous commençons par approcher un signal unitaire u , d'amplitude $a = 1$, de fréquence $f = 1$, et de phase $\phi = 0$, *i.e.* :

$$u(t) = \sin(2\pi t) \quad (77)$$

Nous devons choisir le degré maximal d des polynômes, et l'intervalle à approcher de la partie périodique du signal. Nous appelons cet intervalle l'*intervalle de validité* I de l'approximation polynomiale. Si nous approchons une demi-période du signal u , alors $I = [0; 1/2[$. Les choix du degré maximal et de l'intervalle de validité sont fondamentaux, puisqu'ils déterminent la qualité de la synthèse et la complexité de la méthode.

Nous utilisons le rapport signal sur bruit (SNR pour *Signal-to-Noise Ratio*) pour mesurer la qualité de la synthèse. Le SNR évalue la performance de l'approche du signal sinusoïdal u par le polynôme U sur l'intervalle de validité du polynôme. Pour un intervalle I et un degré d donnés, la valeur du SNR est dans ce cas donnée par l'équation :

$$\text{SNR} = 10 \cdot \log_{10} \left(\frac{\int_I u^2(t) dt}{\int_I (u(t) - U(t))^2 dt} \right) \quad (78)$$

Pour optimiser la qualité de la synthèse, nous devons maximiser la valeur du SNR, c'est-à-dire trouver les coefficients qui minimisent le dénominateur de l'équation (78).

Le SNR donne une information générale sur le bruit dans le signal, mais n'indique pas par exemple si il y a des *clics* audibles. Nous voulons pourtant garantir une certaine propriété de continuité au signal sonore. L'approximation polynomiale est valide sur l'intervalle de validité, mais le signal global, reconstruit polynomialement par morceaux, doit aussi avoir certaines propriétés de continuité. Une propriété C^0 garantit une continuité en amplitude. Si nous notons C^k la continuité d'ordre k , plus k est grand et moins il y aura de discontinuités, et donc moins de probabilités d'introduire des *clics* dans le son. Ainsi une propriété de continuité C^1 est très fortement conseillée. Des contraintes nouvelles doivent alors être ajoutées dans la recherche des meilleurs coefficients polynomiaux, pour garantir certaines propriétés de continuité.

Ainsi, pour un polynôme U de degré 2 et un intervalle de validité $I = [0; 1/2[$, il suffit que $U(0) = U(1/2) = 0$ pour que le signal par morceaux soit C^1 continu. Les coefficients polynomiaux a_i qui minimisent le SNR sont trouvés en résolvant la minimisation du dénominateur du SNR qui vaut dans ce cas :

$$\int_0^{1/2} (\sin(2\pi t) - (a_0 + a_1 t + a_2 t^2))^2 dt \quad (79)$$

ce qui donne comme solution :

$$\begin{cases} a_0 = 0 \\ a_1 = 240/\pi^3 \\ a_2 = -480/\pi^3 \end{cases} \quad (80)$$

Pour approcher le signal unitaire sinusoïdal dans son ensemble, il faut alors utiliser, alternativement et périodiquement, les polynômes U_a sur une demi-période et U_b pour la seconde, tels que :

$$\begin{cases} U_a(t) = a_1 t + a_2 t^2 \\ U_b(t) = -a_1 t - a_2 t^2 \end{cases} \quad (81)$$

Le choix de l'intervalle de validité et du degré des polynômes a énormément d'influence sur la qualité de la synthèse, comme l'illustre le tableau 12. Nous voudrions alors allonger l'intervalle de validité, ce qui entraîne moins d'alternance de coefficients polynomiaux et améliore la performance (section 8.2.6), mais conjointement augmenter également le degré des polynômes, pour conserver un bon SNR. Nous verrons cependant dans la section 8.2.3 qu'un polynôme de haut degré peut amener la méthode à une instabilité numérique, ce qui nous limite en l'état de nos travaux à l'utilisation d'un degré polynomial inférieur ou égal à 4. D'après le tableau 12 et notre expérience, nous pouvons avancer que le choix d'un intervalle de validité $I = [0; 1/2[$ avec un degré polynomial $d = 2$ est particulièrement adapté pour une synthèse très rapide de qualité médiocre, et le choix de $I = [0; 1/2[$ avec $d = 4$ conditionne une synthèse rapide de bonne qualité. Nous pouvons aussi remarquer qu'avec un polynôme de degré 2 et une propriété de continuité C^1 , le polynôme générateur est en réalité infiniment dérivable.

Quels que soient les choix du degré polynomial ou de l'intervalle de validité, nous savons trouver les coefficients du polynôme $U(t)$ qui approche un signal sinusoïdal unitaire $u(t)$. Nous pouvons alors en déduire l'expression générale d'un polynôme approchant n'importe quel partiel. Pour un partiel i d'amplitude a_i , de fréquence f_i , et de phase initiale ϕ_i , le polynôme correspondant P_i est donné par :

$$P_i(t) = a_i U \left(f_i t + \frac{\phi_i}{2\pi} \right) \quad (82)$$

Les valeurs du SNR données dans le tableau 12 sont toujours les mêmes, quelles que soient les amplitudes ou fréquences des partiels. Le bruit engendré par cette approximation est particulier : il ajoute à la composante sinusoïdale à synthétiser des harmoniques d'amplitude faible, qui dépendent des choix de l'intervalle de validité I et du degré d du polynôme choisi. La figure 8.2.1 illustre l'approximation d'un partiel par un polynôme par morceaux, et le signal du bruit engendré.

I	d	C^0 SNR (dB)	C^1 SNR (dB)
$[0; 1/4[$	2	36	28
$[0; 1/4[$	3	57	28
$[0; 1/4[$	4	79	59
$[0; 1/4[$	5	102	59
$[0; 1/2[$	2	28	28
$[0; 1/2[$	3	28	28
$[0; 1/2[$	4	59	59
$[0; 1/2[$	5	59	59
$[0; 1[$	4	17	17
$[0; 1[$	5	42	42

TAB. 12 – Erreur de l'approximation polynomiale du signal d'un partiel. Le rapport signal sur bruit (SNR) est calculé en comparant l'erreur d'approximation ($u - U$) avec le signal u . Pour deux contraintes de continuité (C^0 et C^1), les valeurs du SNR sont exprimées en fonction de l'intervalle de validité I choisi, ainsi que du degré polynomial d .

Nous verrons dans la section 8.2.5 que les paramètres des partiels peuvent changer.

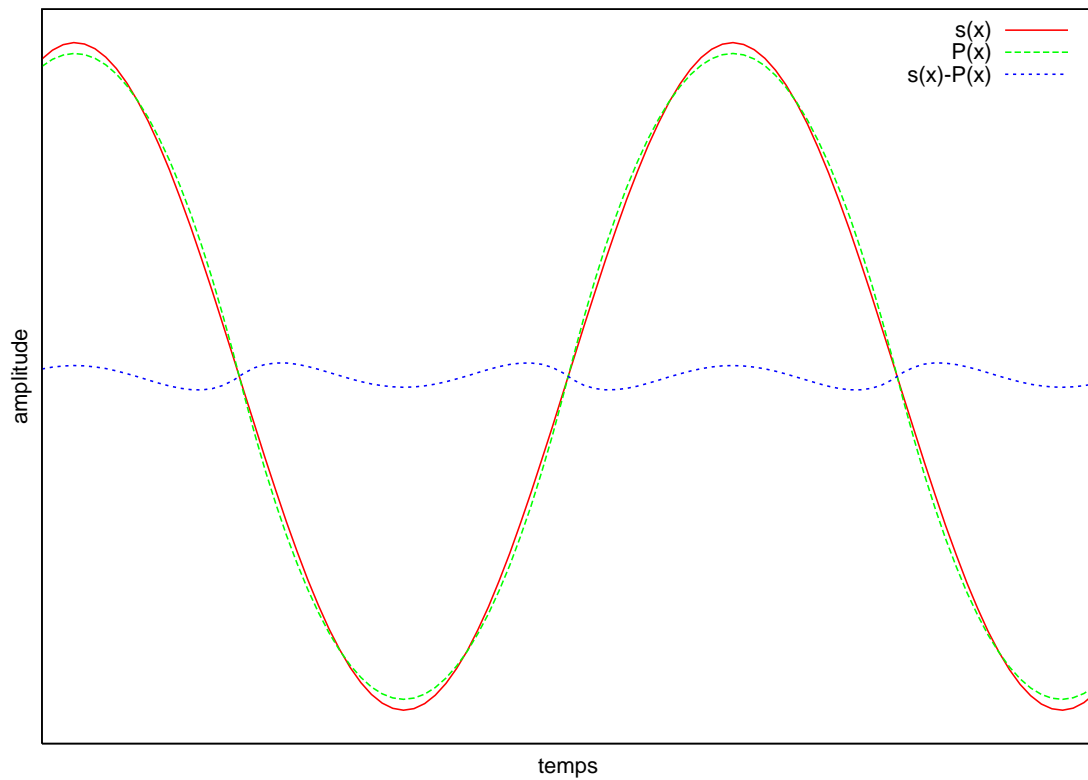


FIG. 45 – Approximation polynomiale d'un signal sinusoidal s . P est le polynôme par morceaux qui approche s , composé ici de 2 polynômes de degré 2 par période. La différence $s - P$ représente le bruit engendré par la méthode. Le bruit est périodique, il ajoute des harmoniques au partiel synthétisé.

Mais comme ils varient lentement dans le temps, nous les considérons constants sur de courtes durées. Sur ces intervalles de temps, les fonctions générées sont alors périodiques, et le signal d'un partiel doit être approché sur une période seulement. Si nous choisissons un intervalle de validité $I = [0; 1/2[$, il y a alors 2 ensembles de coefficients polynomiaux à trouver pour approcher le signal dans son intégralité.

Chaque ensemble de coefficients polynomiaux est valide sur un intervalle de la partie périodique d'un signal, l'intervalle de validité. Tant que le signal est périodique, nous utilisons les coefficients périodiquement. Pour réaliser la synthèse, ces coefficients doivent être changés régulièrement, en fonction de la durée de l'intervalle de validité, qui dépend de la fréquence du partiel.

8.2.2 Calcul incrémental des polynômes

Nous proposons d'utiliser une formule de Taylor pour calculer les valeurs d'un polynôme. Le théorème de Taylor avec reste intégral indique que si $n \in \mathbb{N}$, et si f est une fonction $(n+1)$ -dérivable sur $[a; b]$, alors :

$$f(b) = f(a) + \frac{(b-a)}{1!} f'(a) + \dots + \frac{(b-a)^n}{n!} f^n(a) + \int_a^b \frac{(b-a)^n}{n!} f^{n+1}(t) dt \quad (83)$$

Dans le cas particulier des polynômes, infiniment dérivables, le théorème de Taylor peut être appliqué. Et pour un polynôme de degré d , le reste intégral est nul si $n \geq d$. Un polynôme peut alors être évalué sans approximation à chaque instant, en utilisant sa précédente valeur et les valeurs de ses dérivées. Sa valeur en $t_0 + \Delta_t$ est donnée par :

$$P^k(t_0 + \Delta_t) = P^k(t_0) + \sum_{i=k+1}^d \frac{\Delta_t^{i-k}}{(i-k)!} P^i(t_0) \quad (84)$$

avec P^k la dérivée d'ordre k de la fonction polynomiale, P^0 étant le polynôme à évaluer. Le nombre de termes nécessaires dépend du degré du polynôme à évaluer (3 pour un polynôme de degré 2).

La méthode d'approximation polynomiale détaillée précédemment indique comment trouver les meilleurs coefficients polynomiaux. Nous pouvons en déduire la première valeur du polynôme et de ses dérivées. Puis nous utilisons l'équation (84) pour calculer les valeurs suivantes, avec un intervalle Δ_t correspondant au temps entre deux événements de mise à jour.

Nous savons que l'évaluation de fonctions polynomiales $P(x)$ pour de très grandes valeurs de x entraîne rapidement une imprécision numérique. Calculer incrémentalement les valeurs des polynômes évite ainsi d'utiliser de très grandes valeurs de temps. Cela permet surtout d'additionner des polynômes sans qu'ils soient nécessairement synchronisés sur un même intervalle de validité : cela autorise la somme de signaux déphasés.

8.2.3 Générateur polynomial

Utiliser un calcul incrémental pour chaque polynôme, et additionner ensuite les valeurs obtenues, ne serait pas plus avantageux que la méthode classique. Nous proposons plutôt d'utiliser un polynôme générateur, initialisé avec la somme de toutes les valeurs d'initialisation des polynômes approchant les partiels. Avec cette méthode, seul le générateur est

évalué à différents moments de la synthèse par un calcul incrémental, pour produire les échantillons du son.

Comme nous l'avons vu dans la section 8.2.1, les polynômes sont définis sur des intervalles de validité. Quand un polynôme atteint la fin de cet intervalle, le générateur doit être mis à jour : il faut soustraire les contributions polynomiales du partiel "expiré", et ajouter les nouvelles. Le calcul incrémental du générateur peut alors reprendre. Quand un échantillon doit être produit, il faut évaluer le générateur incrémentalement. Il y a donc deux types d'événements temporels lors de la synthèse :

- les événements de mise à jour des valeurs du polynôme et de ses dérivées ;
- les évaluations du polynôme et de ses dérivées.

La figure 46 donne un exemple d'une suite d'événements lors d'une synthèse avec 3 partiels.

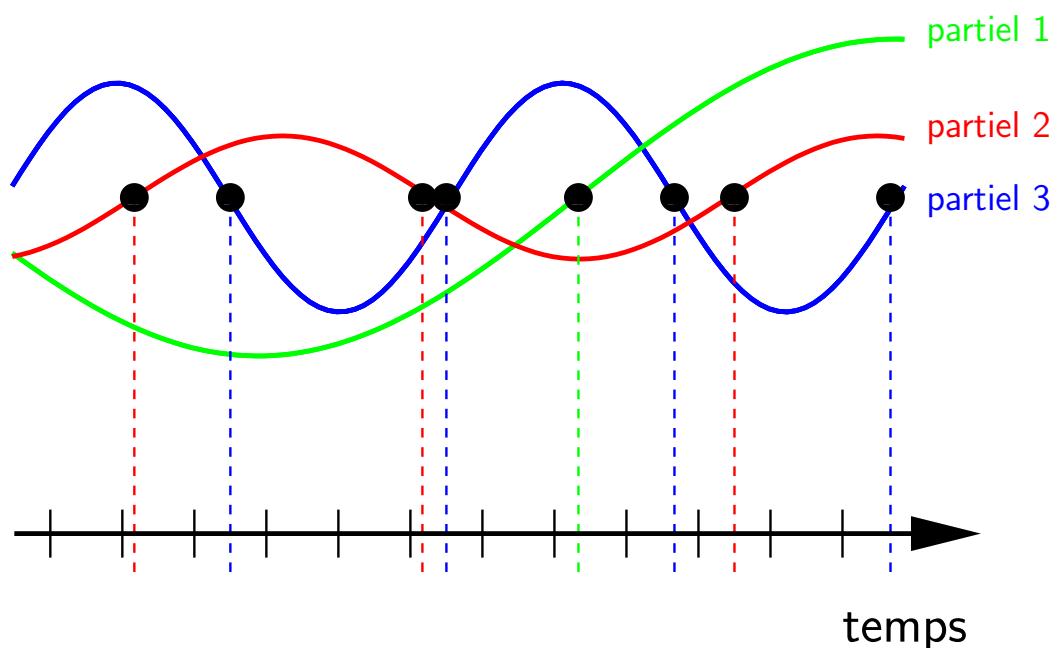


FIG. 46 – Les événements dans la méthode PASS. 3 partiels sont représentés par leurs signaux temporels polynomiaux qui approchent les fonctions sinusoidales. Deux types d'événements peuvent se produire dans le temps. Premièrement, les événements de mise à jour, représentés par les points sur les courbes, surviennent au moment d'actualiser les valeurs du polynôme générateur et de ses dérivées, quand l'approximation polynomiale d'un partiel a atteint la fin de l'intervalle de validité I (ici, $I = [0; 1/2[$). Deuxièmement, les événements d'évaluation du générateur surviennent quand un échantillon du son doit être généré. Le générateur est ainsi régulièrement mis à jour (lignes pointillées) et évalué (traits pleins) dans le temps.

L'évaluation incrémentale du générateur peut engendrer une instabilité numérique, due à la représentation machine des flottants. Des bibliothèques spécialisées pour manipuler des flottants sans erreurs ont été testées, comme la GNU Multiple Precision Arithmetic Library (GMP), mais cela augmente considérablement les temps de calcul. Nous préférons réinitialiser le générateur régulièrement avec une somme de fonctions sinusoidales. La durée des intervalles de temps avant ré-initialisation dépend principalement du degré choisi pour

l'approximation polynomiale.

La complexité de la méthode PASS est déterminée par la gestion des événements de mise à jour et d'évaluation du générateur. Pour optimiser cette gestion, nous proposons d'utiliser une structure de donnée adaptée.

8.2.4 Structure de données

Nous utilisons une file de priorité pour gérer les événements temporels de la méthode PASS. C'est un type abstrait de données muni des primitives suivantes :

- *insert* : ajoute un élément dans la file avec une certaine priorité ;
- *delete-max* : supprime l'élément de plus haute priorité de la file et renvoie sa valeur.

Au cours de la synthèse, des événements de mise à jour sont régulièrement insérés dans la file, ou bien supprimés et traités. L'implémentation classique d'une file de priorité est fondée sur la structure de tas (voir par exemple [AHU83]). Un tas est un arbre binaire qui satisfait deux contraintes :

- l'arbre est soit binaire complet, ou alors les sommets du dernier niveau de l'arbre sont rangés de gauche à droite ;
- chaque sommet est supérieur (en priorité) ou égal à chacun de ses fils. Le sommet du tas est ainsi toujours le premier élément à traiter.

Avec cette implémentation, les primitives sont exécutées avec une complexité d'ordre $O(\log(N))$, où N est le nombre d'éléments de la file.

La majorité du temps de calcul de la méthode PASS est dédié à la gestion de la file de priorité. Le nombre d'appels aux primitives du tas doit alors être optimisé. Notre première approche fut d'utiliser la primitive *delete-max* pour traiter l'élément de plus haute priorité, puis *insert* pour insérer le nouveau dans la file. Mais dans ce cas, le tas est réorganisé deux fois pour chaque paire d'opérations *delete-max* / *insert*. Un exemple est donné par la figure 47. Dans le but d'améliorer les performances, nous avons remplacé les primitives *delete-max* / *insert* par le couple *top* / *replace* :

- *top* : retourne l'élément de plus haute priorité de la file, sans le supprimer. Il est ainsi possible de traiter un événement sans le sortir de la file.
- *replace* : remplace l'élément de plus haute priorité de la file (le sommet du tas) par un élément à insérer. Le nouvel élément est placé au sommet du tas, avant que le tas ne soit réorganisé en descendant cet élément dans l'arbre en fonction de sa priorité.

Avec ces nouvelles méthodes, le tas n'est réorganisé qu'une fois par mise à jour. Cela avantage également la méthode PASS pour la gestion des hautes fréquences. En effet, les partiels de haute fréquence provoquent des mises à jour plus fréquentes, car la durée de leur intervalle de validité est plus courte. Avec la méthode *replace* et l'insertion d'un élément par le haut de la file, les événements temporels associés aux partiels de haute fréquence restent près du sommet du tas. Utiliser moins d'opérations de réorganisation du tas pour les mises à jour les plus fréquentes améliore la complexité de la méthode PASS. Les figures 47 et 48 donnent un exemple de gestion d'une file de priorité, où les méthodes *delete-max* / *insert* nécessitent 4 échanges d'éléments, alors que les méthodes *top* / *replace* seulement 1 échange.

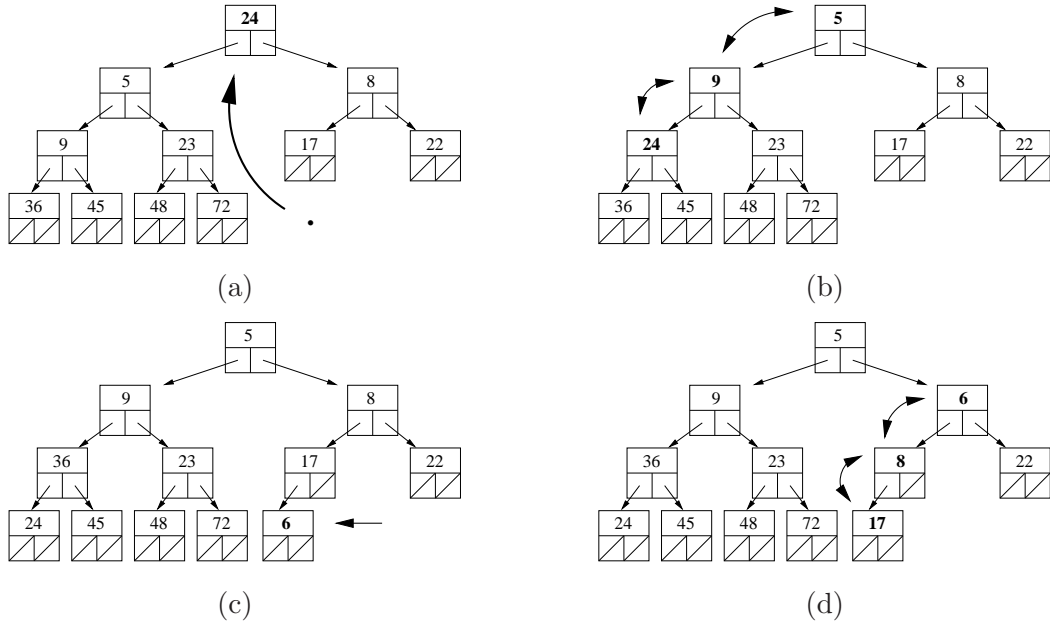


FIG. 47 – Méthodes delete-max et insert pour une file de priorité implémentée par un tas. (a) Suppression de l'élément de plus haute priorité. (b) Réorganisation du tas. (c) Insertion du nouvel élément par le bas. (d) Réorganisation du tas.

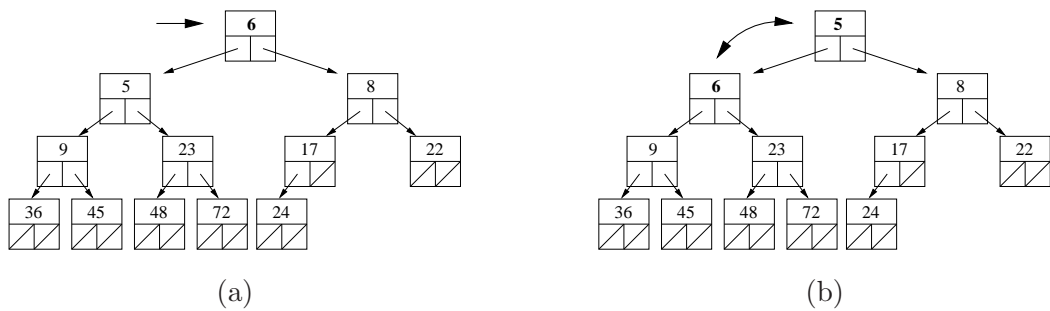


FIG. 48 – Méthode replace pour une file de priorité implémentée par un tas. (a) Remplacement de l'élément de plus haute priorité par l'élément à insérer. (b) Réorganisation du tas.

8.2.5 Changement des paramètres sonores

Nous considérons les paramètres sonores des partiels constants sur un court intervalle de temps. Mais ils changent, et ils doivent donc être régulièrement mis à jour. Strandh et Marchand indiquent dans [SM99] les meilleurs moments pour changer les paramètres d'un partiel. Pour un changement d'amplitude, le moment idéal est quand le signal est minimal, car la continuité du signal est préservée (figure 49). Le meilleur moment est au contraire au maximum du signal pour le changement de fréquence, car c'est la continuité de la dérivée du signal qui est ainsi préservée (figure 50).

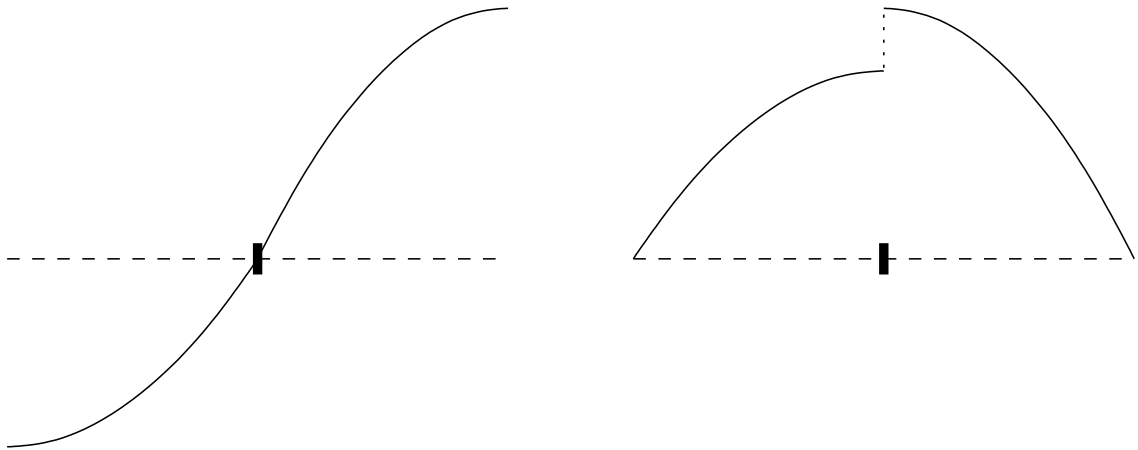


FIG. 49 – *Changement du paramètre d'amplitude quand le signal est minimal (à gauche) ou maximal (à droite). Le cas de gauche est préférable car il évite une discontinuité d'amplitude (clics).*

Changer les paramètres d'amplitude ou de fréquence d'un partiel revient à recalculer ses valeurs d'initialisation (valeurs polynomiales à la base d'une évaluation incrémentale, voir section 8.2.2). Et il faut aussi mettre à jour le générateur. Si le changement est fait à la fin de l'intervalle de validité, il a lieu au moment d'une mise à jour programmée, qui aurait eu lieu sans changement de paramètres. Le cas le plus favorable se présente avec un intervalle de validité $I = [0; 1/4[$. Dans ce cas, nous pouvons faire les changements aux moments optimaux (voir figures 49 et 50). Il n'y a donc pas de complexité ajoutée à la méthode. Dans tout autre cas, il faut insérer dans la file de priorité un événement supplémentaire de mise à jour, qui surviendra au moment optimal, en fonction du paramètre à changer. Mais peu d'événements de ce type doivent être ajoutés en comparaison du nombre de mises à jour classiques, ce qui limite l'impact du changement de paramètres sur la complexité générale de la méthode.

8.2.6 Complexité

La complexité de la méthode PASS dépend principalement de la gestion des événements, et donc de l'utilisation de la file de priorité. Soit p le nombre de polynômes nécessaires pour approcher un partiel sur une partie périodique. Alors la file de priorité va être sollicitée ($p \cdot f$) fois par partiel pour une seconde de synthèse sonore, avec f la fréquence du partiel (2000 fois par seconde pour 1 partiel de fréquence 1000 Hz et un intervalle de

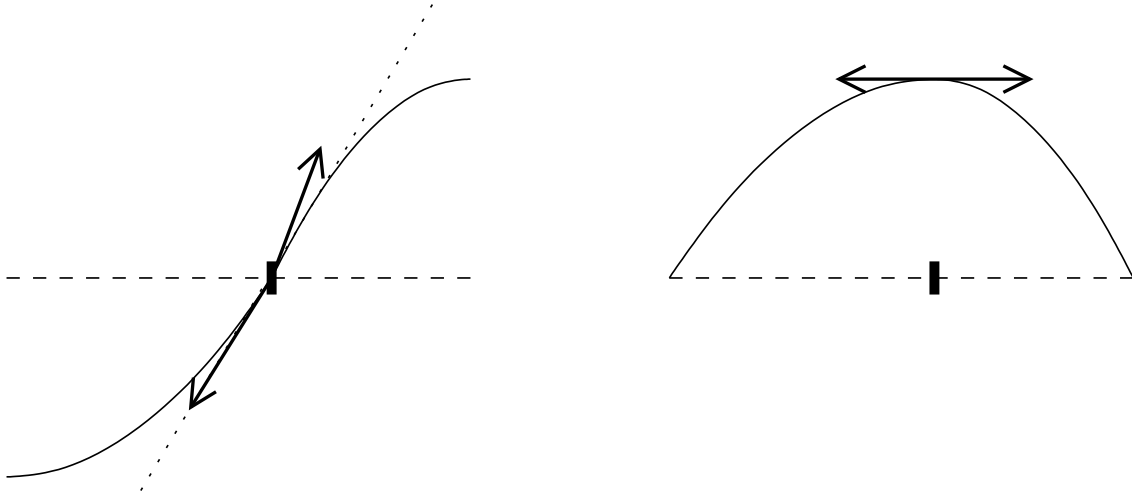


FIG. 50 – *Changement du paramètre de fréquence quand le signal est minimal (à gauche) ou maximal (à droite). Le cas de droite est préférable car il évite une discontinuité de la dérivée du signal (clics).*

validité $I = [0; 1/2[$). Nous pouvons en déduire le nombre total X d'appels à la file de priorité :

$$X = p\Delta_t \sum_{i=1}^N f_i = N\bar{f}p\Delta_t \quad (85)$$

où N est le nombre de partiels dans le son, f_i la fréquence du partiel i , \bar{f} la fréquence moyenne des partiels du son, et Δ_t la durée de synthèse.

En considérant l'ordre de complexité $O(\log(N))$ de chaque primitive de la file de priorité (fonction du nombre N d'éléments dans la file), nous pouvons exprimer la complexité C_1 correspondant à la gestion des événements de mise à jour :

$$C_1 = O(N \log(N) \bar{f} p \Delta_t) \quad (86)$$

Cette complexité est fortement dépendante de la fréquence moyenne des partiels du son. C'est pourquoi la méthode est plus efficace avec les basses fréquences. Si l'intervalle de validité I est doublé, la performance de la méthode PASS est en conséquence améliorée d'un facteur 2. Plus I est grand, meilleures sont les performances de la synthèse. Mais nous avons vu précédemment que pour conserver une qualité de synthèse satisfaisante lorsque I augmente, il faut également augmenter le degré des polynômes, et cela entraîne une instabilité numérique. Il doit y avoir un arbitrage entre stabilité et complexité de la méthode.

À la complexité C_1 s'ajoute C_2 , la complexité de production des échantillons du son. Pour chaque échantillon, il faut évaluer le générateur et ses dérivées. Ce calcul dépend donc du degré polynomial choisi pour la méthode. Avec F_e comme fréquence d'échantillonnage du son et d le degré polynomial choisi, nous obtenons :

$$C_2 = O(dF_e\Delta_t) \quad (87)$$

La complexité de la méthode PASS peut alors être exprimée ainsi :

$$C_{\text{PASS}} = O(\alpha N \log(N) \bar{f} p \Delta_t + d F_e \Delta_t) \quad (88)$$

avec α une constante dépendante de l'architecture de l'ordinateur.

Cette complexité est peu dépendante de la fréquence d'échantillonnage. Le temps de calcul n'est alors pas vraiment affecté par une augmentation de cette fréquence, comme le montre le tableau 13.

8.2.7 Résultats

Marchand [Mar00] a montré en 2000 que la méthode du résonateur numérique pouvait être plus rapide que l'utilisation de la transformée de Fourier inverse (FFT^{-1}). Plus récemment, Meine et Purnhagen [MP02] ont comparé les différentes méthodes de synthèse additive, et ont conclu que la méthode la plus rapide était aujourd'hui la FFT^{-1} . En fait, cela dépend surtout de détails d'implémentation, de l'ordinateur sur lequel est synthétisé le son, du nombre d'oscillateurs ou encore de la fréquence d'échantillonnage utilisée. Mais il est certain que la méthode du résonateur numérique a comme avantage un contrôle plus fin des paramètres sonores de chaque partiel, contrairement à la méthode FFT^{-1} . Comme c'est également le cas avec la méthode PASS, nous choisissons de comparer les performances de notre méthode avec celles du résonateur numérique.

Le choix de la fréquence d'échantillonnage du son affecte peu les performances de la méthode PASS, mais ce n'est pas le cas avec le résonateur numérique, dont voici une expression de la complexité C_{RN} :

$$C_{\text{RN}} = O(N F_e \Delta_t) \quad (89)$$

avec F_e la fréquence d'échantillonnage, N le nombre de partiels dans le son et Δ_t la durée de la synthèse. Nous remarquons que dans ce cas N et F_e sont multipliés. Cette différence de complexité entre le résonateur numérique et la méthode PASS est mise en évidence par le tableau 13.

La synthèse PASS est plus rapide que le résonateur numérique pour les basses fréquences, comme l'illustrent le tableau 14, et les figures 51 et 52. En utilisant des intervalles de validité couvrant les demi-périodes des partiels, et des polynômes de degré 2, la synthèse PASS est plus rapide avec 2500 partiels si la fréquence moyenne de ces partiels ne dépasse pas 300 Hz, et même 500 Hz si la fréquence d'échantillonnage est à 96 kHz. Une synthèse peut être réalisée en temps réel avec 5000 partiels de 150 Hz.

8.3 Méthode hybride

Nous proposons d'arbitrer entre complexité et stabilité numérique en combinant les avantages du résonateur numérique et de la méthode PASS. L'idée d'une synthèse mixte est de pouvoir changer de méthode de synthèse en fonction de la fréquence du partiel : pour les basses fréquences, la méthode PASS est préférée. Le seuil de fréquence garantit de toujours utiliser la méthode la plus performante, ce qui entraîne un gain de temps. De plus, l'instabilité numérique de la méthode PASS est limitée : en ne traitant que des basses fréquences, il y a moins de mises à jour, moins de calculs incrémentaux du polynôme

N	\bar{f}	F_e (Hz)	RN	PASS
4000	300	22050	3.2 s	6.6 s
4000	300	44100	6.3 s	6.6 s
4000	300	96000	13.7 s	6.6 s

TAB. 13 – Comparaison du temps de calcul du résonateur numérique (RN) et de la méthode PASS pour 5 secondes de synthèse sonore, et pour différentes fréquences d'échantillonnage du son. La méthode PASS est utilisée avec des polynômes de degré 2, et un intervalle de validité $I = [0; 1/2[$. N est le nombre de partiels, \bar{f} la fréquence moyenne de tous les partiels, et F_e est la fréquence d'échantillonnage du son. Les deux méthodes ont été implémentées en langage C, compilées avec le compilateur C GNU (gcc) version 4.0, et exécutées avec un processeur Intel Pentium 4 cadencé à 1.8 GHz.

N	\bar{f}	RN	PASS
2500	200	3.9 s	2.0 s
2500	300	3.9 s	3.0 s
2500	400	3.9 s	4.0 s
2500	500	3.9 s	5.0 s
5000	200	7.9 s	7.3 s
5000	300	7.9 s	10.6 s
5000	400	7.9 s	14.4 s

TAB. 14 – Comparaison du temps de calcul du résonateur numérique (RN) et de la méthode PASS pour 5 secondes de synthèse sonore, avec une fréquence d'échantillonnage $F_e = 44100$ Hz. La méthode PASS est utilisée avec des polynômes de degré 2, et un intervalle de validité $I = [0; 1/2[$. N est le nombre de partiels, \bar{f} la fréquence moyenne de tous les partiels. Les deux méthodes ont été implémentées en langage C, compilées avec le compilateur C GNU (gcc) version 4.0, et exécutées avec un processeur Intel Pentium 4 cadencé à 1.8 GHz.

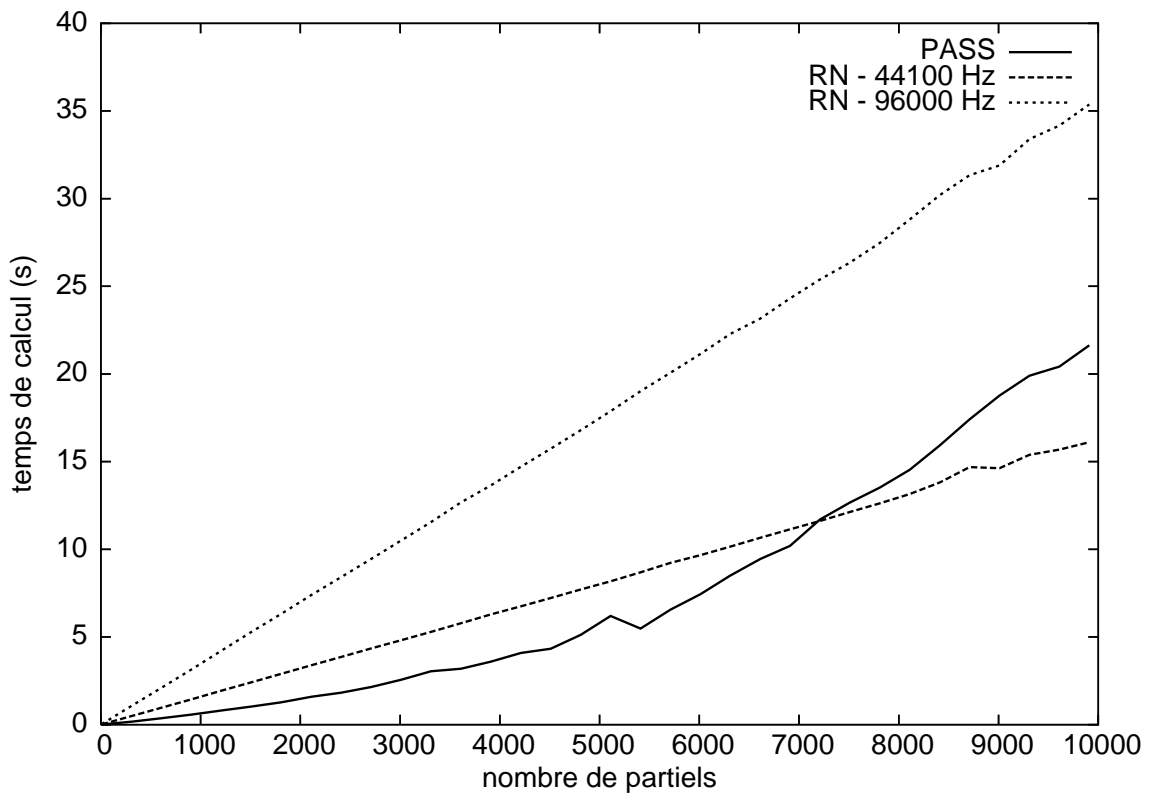


FIG. 51 – Comparaison du temps de calcul du résonateur numérique (RN) et de la méthode PASS pour 5 secondes de synthèse sonore. Le temps de calcul est fonction du nombre de partiels N , avec une fréquence moyenne fixe de $f = 200$ Hz. La méthode PASS est utilisée avec des polynômes de degré 2 et un intervalle de validité $I = [0; 1/2[$. Les deux méthodes ont été implémentées en langage C, compilées avec le compilateur C GNU (gcc) version 4.0, et exécutées avec un processeur PowerPC G4 cadencé à 1.25 GHz.

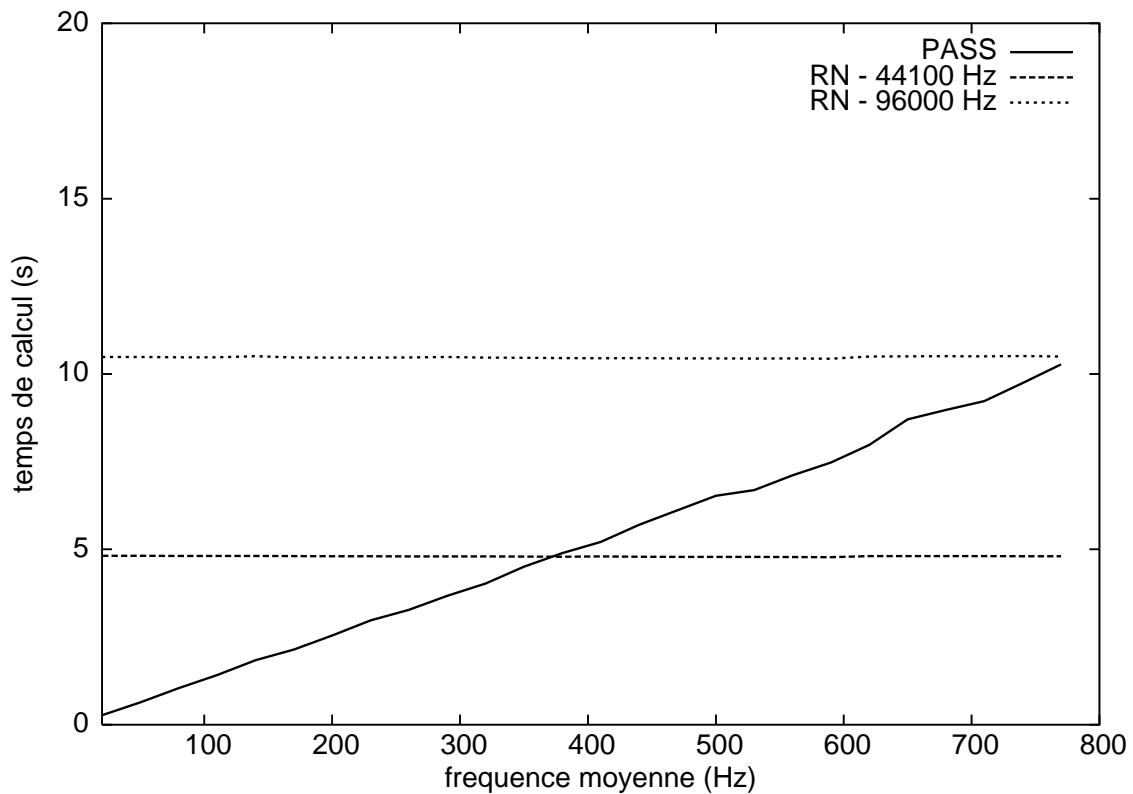


FIG. 52 – Comparaison du temps de calcul du résonateur numérique (RN) et de la méthode PASS pour 5 secondes de synthèse sonore. Le temps de calcul est fonction de la fréquence moyenne des partiels \bar{f} , avec un nombre fixé de partiels $N = 3000$. La méthode PASS est utilisée avec des polynômes de degré 2 et un intervalle de validité $I = [0; 1/2[$. Les deux méthodes ont été implémentées en langage C, compilées avec le compilateur C GNU (gcc) version 4.0, et exécutées avec un processeur PowerPC G4 cadencé à 1.25 GHz.

générateur. Avantage supplémentaire, la méthode du résonateur numérique peut utiliser la file de priorité du PASS pour gérer les moments optimaux de changement de paramètres sonores (voir section 8.2.5). Le choix de la méthode de synthèse peut être fait en fonction des nouveaux paramètres du partiel à chaque mise à jour. Comme nous connaissons à ce moment précis, pour les deux méthodes, l'amplitude, la fréquence et la phase du partiel mis à jour, le changement de méthode est très simple.

8.4 Application à la synthèse de surfaces océaniques

La proximité des problèmes posés par la synthèse sonore additive et ceux associés à la synthèse réaliste de surfaces océaniques nous a amené à utiliser la méthode PASS pour animer en temps réel des océans de vagues.

8.4.1 Simulation réaliste d'océans

Nous souhaitons tirer avantage des algorithmes de synthèse sonore additive pour animer de façon réaliste des surfaces océaniques en temps réel. Pour que le rendu soit réaliste, il faut bien détailler la surface océanique, en synthétisant des centaines de vagues. Des textures peuvent être utilisées pour simuler ce niveau de détail, dans des applications temps-réel tels que les jeux vidéos par exemple. Le rendu est alors rapide, mais la qualité de l'animation des vagues est médiocre. Dans ce cas, nous ne sommes plus dans un contexte réaliste.

Nous disposons de paramètres de vagues issus de mesures océanographiques pour différents états de mer. L'utilisation de la transformée de Fourier inverse offre alors une solution rapide de synthèse à partir de ces paramètres. Mais cette méthode manque de souplesse et peut difficilement produire un rendu réaliste, comme l'indique Fréchet [Fré06b]. D'un autre côté, les équations de Gerstner, présentées dans la section suivante, peuvent très bien rendre une surface océanique large et complexe. Mais comme elles nécessitent le calcul de nombreuses fonctions sinusoïdales, elles ne peuvent pas encore être utilisées pour calculer rapidement des surfaces de haute qualité, avec de nombreuses vagues.

Nous proposons alors d'utiliser les algorithmes de synthèse additive les plus rapides pour calculer et animer des surfaces océaniques dans le modèle de Gerstner. Ces méthodes sonores sont traditionnellement utilisées pour additionner de manière efficace beaucoup d'oscillateurs. En passant des variables de temps et de fréquence à des variables d'espace et de longueur d'onde de vague, les méthodes rapides de synthèse additive peuvent être utilisées pour calculer efficacement les coordonnées d'une surface océanique en mouvement. En utilisant un rendu adaptatif et un filtre *anti-aliasing*, nous proposons avec Fréchet [RF06] une nouvelle méthode de rendu de surface océanique réaliste et rapide.

8.4.2 Modèle de surface océanique

Les surfaces océaniques sont constituées de vagues. Nous considérons ici qu'une vague est une onde élémentaire, et non un train d'onde que désigne le terme "vague" usuel. Les paramètres les plus utilisés pour modéliser une telle vague sont l'amplitude a , la longueur d'onde λ , la fréquence f , le nombre d'onde $k = 2\pi/\lambda$ et la fréquence angulaire $\omega = 2\pi f$

(en eaux profondes, $\omega^2 = gk$, où g est l'accélération due à la gravité). Un vecteur de vague \vec{k} est un vecteur qui a pour norme k et pour direction le sens de propagation de la vague.

Le modèle utilisé traditionnellement pour la simulation d'océans est fondé sur les équations de Gerstner, introduites en informatique par Fournier et Reeves [FR86]. Ce modèle décrit la trajectoire des particules des vagues sur la surface de l'océan. Pour une vague en eau profonde, les particules suivent un cercle vertical autour de leur position au repos, l'ensemble des particules formant une trochoïde :

$$\begin{cases} \vec{x}(\vec{x}_0, t) = \vec{x}_0 + \sum_i \hat{k}_i a_i \sin(\vec{k}_i \cdot \vec{x}_0 - \omega_i t + \phi_i) \\ z(\vec{x}_0, t) = z_0 - \sum_i a_i \cos(\vec{k}_i \cdot \vec{x}_0 - \omega_i t + \phi_i) \end{cases} \quad (90)$$

où $\vec{x} = (x, y)$ est la position horizontale de la particule au temps t , z sa position verticale, $\vec{x}_0 = (x_0, y_0)$, (x_0, y_0, z_0) est la position de la particule au repos, \hat{k} est le vecteur unité de \vec{k} et ϕ est une phase aléatoire.

Le modèle de Gerstner permet d'animer une surface océanique de façon réaliste, mais nécessite pour cela un temps de calcul élevé. Pour le calcul d'un point de l'image, un couple de fonctions sinus / cosinus doit être calculé pour chaque vague. Et le rendu n'est réaliste qu'avec des centaines de vagues, et une interactivité de l'animation avec au moins 20 images par seconde (fps, pour *frames per second*). Pour diminuer le temps de calcul, Mastin *et al.* [MWM87] ou Tessorod [Tes01] utilisent la transformée de Fourier inverse (FFT^{-1}). L'utilisation de cet algorithme permet le rendu de surfaces océaniques en temps réel. Mais la méthode FFT^{-1} est difficilement utilisable pour un rendu réaliste, car la longueur de la grille et son nombre de points définissent l'ensemble de vecteurs de vagues utilisables, ce qui diminue la variété et le contrôle des surfaces océaniques synthétisées. C'est pourquoi nous n'utiliserons pas cette méthode pour nos comparaisons de performance.

Hinsinger *et al.* [HNC02] utilisent une grille adaptative pour réduire les calculs à la partie visible de la surface océanique. L'idée est de projeter une grille uniforme de l'écran sur la surface océanique au repos. Les équations de Gerstner sont alors uniquement évaluées pour les points de la grille projetée. De plus, les vagues de fréquence trop élevée sont supprimées, car elles entraînent de l'*aliasing*. Même si cette méthode est plus rapide qu'avec une grille uniforme, elle ne peut encore traiter qu'une centaine de vagues en temps réel.

En utilisant des méthodes de synthèse sonore additive, il est possible d'utiliser les équations de Gerstner avec des centaines de vagues, avec une fréquence d'affichage élevée. Combinées avec un rendu adaptatif et un filtre *anti-aliasing*, ces méthodes permettent une simulation réaliste de surfaces océaniques en temps réel.

8.4.3 Utilisation des méthodes sonores

Nous avons testé l'utilisation des méthodes de synthèse sonore additive rapide pour simuler une surface océanique. Pour conserver le contrôle des paramètres des vagues, et dans un souci de réalisme, nous n'utilisons pas la méthode de la transformée de Fourier inverse. Nous avons comparé les performances des méthodes PASS et du résonateur numérique à celle de la méthode classique.

Les méthodes de synthèse sonore additive sont efficaces pour additionner un nombre important de fonctions sinusoïdales. Au lieu de les utiliser pour synthétiser un signal sonore

en fonction du temps, nous proposons de les utiliser pour synthétiser les points d'une image d'océan, en fonction des coordonnées de l'image. Comme nous voulons synthétiser une grille de points d'une surface océanique, nous proposons de procéder ligne par ligne, et de réaliser une synthèse sonore par ligne. La synthèse se fait alors en fonction de la longueur de la ligne et non en fonction du temps.

Pour calculer la coordonnée en x d'un point de la ligne, nous utilisons les paramètres de chaque vague : \vec{k} , A , w et Φ , respectivement le vecteur de vague, l'amplitude, la fréquence et la phase de la vague. Les paramètres sonores d'un partiel correspondant à une vague, c'est-à-dire son amplitude a , sa fréquence f et sa phase initiale ϕ , sont calculés à partir des paramètres de la vague :

$$\begin{cases} a &= A\hat{k}_x \\ f &= \frac{k}{2\pi} \cos(\widehat{\vec{k}, \vec{L}}) \\ \phi &= \vec{k} \cdot \vec{x}_0 - \omega t + \Phi \end{cases} \quad (91)$$

où \hat{k}_x est défini par $\hat{\vec{k}} = (\hat{k}_x, \hat{k}_y)$, avec $\hat{\vec{k}}$ le vecteur unité de \vec{k} , et où \vec{L} est le vecteur directeur de la ligne de synthèse.

Comme nous avons les paramètres sonores, nous pouvons maintenant utiliser un algorithme de synthèse sonore le long d'une ligne de la grille, en utilisant la distance sur cette ligne Δ_L au lieu du temps. Δ_L est la distance sur la ligne depuis son premier point. Les valeurs de la coordonnée en x sont alors données par :

$$x = x_0 + \sum_{i=1}^N a_i \sin(2\pi f_i \Delta_L + \phi_i) \quad (92)$$

La même méthode est appliquée pour chaque coordonnée, et pour les dérivées nécessaires au calcul des normales. Seul le paramètre initial de phase change de ligne en ligne. Choisir de calculer toutes les fonctions sinusoïdales directement avec la fonction sinus est évidemment très coûteux en temps de calcul. C'est ce qui est fait traditionnellement avec les équations de Gerstner. Nous allons comparer cette méthode avec les performances du résonateur numérique ou de la méthode PASS.

Dans le cas du résonateur numérique, deux fonctions sinus par partiel et par ligne doivent d'abord être calculées, et il suffit ensuite d'une addition et d'une multiplication par partiel pour calculer la valeur suivante sur la ligne de synthèse. Pour utiliser la méthode PASS, il faut d'abord choisir un intervalle de validité sur lequel approcher la fonction sinus, et le degré du polynôme générateur (voir section 8.2.1). Cela influe fortement sur la qualité du rendu et sur le temps de calcul (donc sur la vitesse d'affichage). Nous avons choisi pour cette application d'approcher la fonction sinusoïdale sur des demi-périodes ($I = [0; 1/2[$), avec des polynômes de degré 2. Cela entraîne une erreur maximale de 4% par rapport au signal sinusoïdal, ce qui est négligeable pour un rendu avec des centaines de vagues. Comme les paramètres des vagues ne changent pas pendant la synthèse des lignes d'une grille de la surface océanique, l'algorithme PASS a été modifié. Nous utilisons des tableaux et un tri fusion pour gérer les événements de la synthèse, plutôt qu'une file de priorité implémentée par un tas (voir section 8.2.4).

8.4.4 Complexité des méthodes

Les complexités des méthodes du PASS et du résonateur numérique ont précédemment été exprimées par les équations (88) et (89). En appliquant un changement de paramètres, des paramètres d'une vague vers ceux d'un partiel comme indiqué par l'équation (91), avec L la longueur de la ligne d'océan à synthétiser et X le nombre de points sur cette ligne, nous obtenons :

$$C_{\text{PASS}} = c_1 \sum_{i=1}^N f_i \log NL + c_2 dX \quad (93)$$

$$C_{\text{RN}} = c_3 NX \quad (94)$$

où c_1 , c_2 , c_3 sont des constantes dépendantes de l'architecture de l'ordinateur.

Nous utilisons un rendu adaptatif selon le point de vue, comme Hinsinger *et al.* [HNC02], pour minimiser le nombre de points nécessaires pour l'étape de rendu, réduisant ainsi le temps de calcul d'une image. De plus, nous filtrons les vagues en fonction de leur longueur d'onde pour éviter les effets d'*aliasing*. Ces méthodes ont un impact certain sur la performance de la synthèse, même si les améliorations diffèrent selon la méthode employée. Pour la méthode du résonateur numérique, dépendante linéairement du nombre de points à synthétiser, les effets d'une réduction de points à synthétiser ou du nombre de vague sont évidents. Pour PASS, cela est plus difficile à évaluer, car les facteurs de complexité sont nombreux. Le fait qu'il y ait moins de points et moins de vagues diminue également le temps de calcul pour cette méthode. Mais la méthode *anti-aliasing* avantage la méthode PASS. En effet, ce filtrage concerne les vagues avec les fréquences les plus hautes. Or la méthode PASS est très sensible à la somme des fréquences, qui diminue alors fortement.

Une comparaison précise de la complexité des deux méthodes est délicate à effectuer, à cause des constantes qui dépendent de l'architecture des ordinateurs dans l'expression de la complexité théorique du PASS. Cependant, tous les tests effectués avec un rendu adaptatif indiquent que l'utilisation de la méthode PASS est plus efficace, pour un nombre très divers de vagues, de points à synthétiser, ou avec des conditions de vitesse de vent extrêmes (voir les résultats, section 8.4.6).

8.4.5 Implémentation dans la bibliothèque Aqua

Nous avons implémenté les méthodes du résonateur numérique et PASS dans la bibliothèque Aqua. *Aqua library* [Fré06a] est une bibliothèque C++ permettant de calculer la géométrie et les normales de surfaces océaniques, utilisable pour le rendu en temps réel et/ou réaliste. Le projet inclut un programme OpenGL de démonstration et s'appuie sur les recherches menées précédemment dans le domaine de la simulation réaliste et temps-réel de surfaces océaniques par Fréchet [Fré06b].

8.4.6 Résultats

Le tableau 15 montre un comparatif des temps de calcul d'une image de surface océanique pour les différentes méthodes. Le test a été effectué sur un PC avec un processeur Pentium 4 cadencé à 3 GHz. Pour la méthode classique, appelée ici *sincos*, nous avons

utilisé l'instruction *sincos* pour optimiser les calculs des fonctions sinusoïdales. Cela permet le calcul d'un sinus et d'un cosinus d'un même argument simultanément, réduisant le temps de calcul d'un facteur proche de 2.

Nous avons implémenté les différentes méthodes avec un rendu adaptatif. Mais comme les effets de ce rendu n'est pas le même pour les différents points de vue et les différentes méthodes, nous avons choisi pour les tests de synthétiser une grille uniforme de points. Les paramètres des vagues sont obtenus par un échantillonnage adaptatif d'un spectre JONSWAP, comme indiqué par Fréchet [Fré06b].

$X \times Y$	N	L	WS	SC	RN	PASS
		<i>m</i>	<i>m \cdot s^{-1}</i>	<i>ms</i>	<i>ms</i>	<i>ms</i>
128×128	200	50	5	360	100	45
128×128	200	50	15	360	100	20
128×128	200	200	5	360	100	120
128×128	200	200	15	360	100	30
128×128	400	50	5	730	210	105
128×128	400	50	15	730	210	50
128×128	400	200	5	730	210	265
128×128	400	200	15	730	210	70
256×256	200	50	5	1610	600	90
256×256	200	50	15	1610	600	50
256×256	200	200	5	1610	600	245
256×256	200	200	15	1610	600	65
256×256	400	50	5	3230	1150	210
256×256	400	50	15	3230	1150	95
256×256	400	200	5	3230	1150	535
256×256	400	200	15	3230	1150	140

TAB. 15 – Temps de calcul en ms d'une grille uniforme avec la méthode *sincos* (SC), le résonateur numérique (RN) et la méthode PASS, pour différents nombres de points ($X \times Y$), nombres de vagues (N), largeurs de grille (L) et vitesses de vent (WS).

Les algorithmes issus des méthodes sonores sont plus rapides que l'instruction *sincos* dans tous les cas. La méthode PASS est généralement la meilleure, mais elle est très sensible à la largeur de la grille et à la vitesse du vent. Nous pouvons maintenant animer des surfaces océaniques avec de nombreuses vagues et une fréquence d'affichage élevée, ce qui n'était pas possible avec la méthode classique.

À titre de comparaison, et pour éclairer les résultats donnés dans le tableau 15, le calcul d'une grille de 128×128 points et 256×256 points prend respectivement environ 10 ms et 60 ms avec la transformée de Fourier inverse. Comme indiqué précédemment, cette méthode est moins réaliste, mais plus rapide. Nous pouvons cependant noter que les temps obtenus avec les méthodes sonores sont assez proches de ceux obtenus avec une transformée de Fourier inverse.

Nous avons testé un rendu adaptatif avec une carte graphique ATI Radeon 9200, une résolution d'écran de 1024×768 pixels, et une grille de 128×96 points, 400 vagues et une vitesse de vent de $5 \text{ m} \cdot \text{s}^{-1}$. Avec la méthode *sincos* nous avons obtenu un nombre d'images

par seconde de 2 à 5 fps, en fonction du point de vue. Avec le résonateur numérique, nous avons atteint de 5 à 15 fps, et de 10 à 20 fps avec la méthode PASS. De plus, en multipliant par 2 la résolution de la grille, c'est-à-dire avec 4 fois plus de points, le nombre d'images par seconde avec PASS chute seulement de moitié (5 à 12 fps). La figure 53 représente une image rendue avec ces paramètres. Nous avons également testé le rendu d'une surface haute qualité (figure 54), avec 800 vagues et une grille de 1024×768 points. Le temps pris par la méthode sincos est de 1 min 10 s par image rendue, et moins de 2 s avec la méthode PASS. Cela représente une amélioration d'un facteur 35 dans ce cas.

Une de nos perspectives est d'adapter la méthode PASS à la synthèse de somme de trochoïdes, au lieu de décomposer l'équation paramétrique de la trochoïde en 3 fonctions sinusoïdales. Nous avons supposé jusqu'ici que les paramètres des vagues sont constants sur une ligne de la grille représentant la surface océanique. Mais il est possible avec PASS de faire changer les paramètres sans augmenter considérablement les temps de calcul. Il est donc envisageable de changer les paramètres des vagues à l'intérieur d'une image, pour rendre plusieurs états de mer sur un même plan d'eau. Nous pouvons remarquer enfin que tout signal peut être approché sur une partie par un polynôme. Ainsi, toute application qui doit additionner de nombreux signaux périodiques pourrait tirer avantage de la méthode PASS, initialement destinée à la synthèse sonore.



FIG. 53 – Grille adaptative avec une résolution d'écran de 1024×768 pixels, 256×192 points sur la grille, 400 vagues et une vitesse de vent de $5 \text{ m}\cdot\text{s}^{-1}$. Le nombre d'images par seconde va de 5 à 12 fps avec la méthode PASS (moins de 1 fps avec sincos).

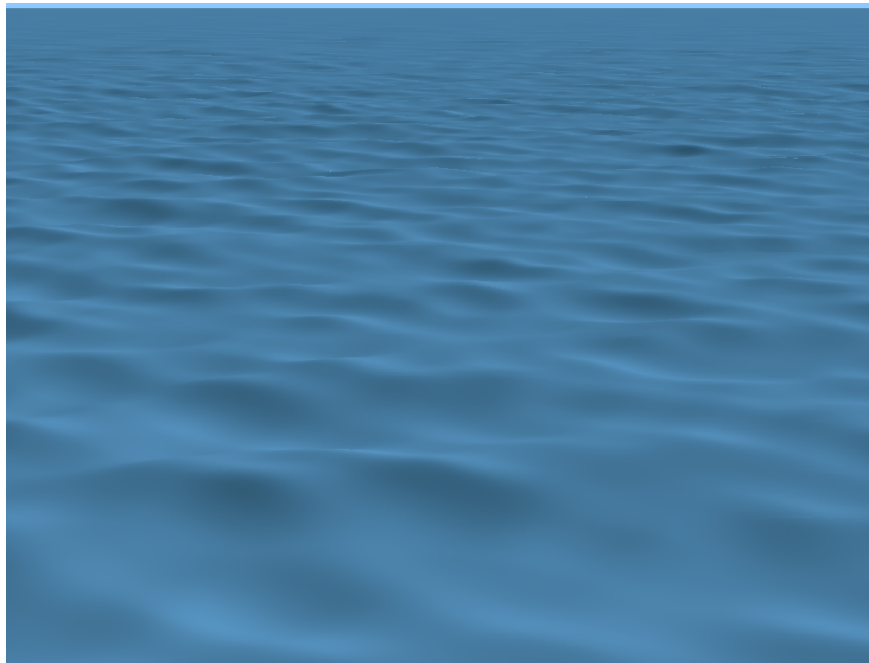


FIG. 54 – Surface océanique de haute qualité : 800 vagues avec une grille de 1024×768 points. Le temps de calcul est de moins de 2 s par image avec la méthode PASS alors qu'il est de 1 mn 10 s par image avec la méthode sincos.

Conclusion

Nous avons parcouru dans ce document le domaine de l'informatique musicale, au niveau symbolique et musical dans la partie II, et au niveau des paramètres sonores dans la partie III. Après avoir introduit quelques notions d'informatique musicale dans les premiers chapitres, nous nous sommes efforcés de faire un état de l'art de chaque domaine. Nous avons alors proposé de nouvelles méthodes pour la musique et pour le son. De nombreuses pistes s'ouvrent dans la continuité de ce travail.

Nous avons commencé dans le chapitre 4 par décomposer la performance instrumentale académique en trois parties : physique, technique, et expressive. Nous avons listé de nombreuses méthodes d'analyse et de synthèse de l'interprétation en fonction de ces critères. Cela nous donne une base claire pour établir des protocoles d'analyse de la performance instrumentale.

Nous avons ensuite étudié plus précisément la performance pianistique, dans le chapitre 5. Nous avons alors proposé des méthodes de doigté automatique d'une mélodie et de reconnaissance du doigté dans la performance qui utilisent la programmation dynamique. Ces méthodes n'ont cependant pas été suffisamment évaluées. Nous prévoyons d'éprouver les résultats donnés par notre méthode de doigté automatique sur des passages musicaux dont le doigté est connu, conventionnel. Ce serait l'occasion de tester l'efficacité de différentes fonctions de coût d'un chemin dans le graphe des transitions en fonction des coûts élémentaires de celles-ci. Le doigté simultané des deux mains pourrait être aussi étudié, en rajoutant des contraintes physiques dans le système. Notre travail étant préliminaire, les perspectives sont nombreuses pour la méthode de reconnaissance du doigté dans la performance. Nous devons d'abord disposer d'une base de données importante de performances pianistiques, incluant des pièces instrumentales non expressives – passages techniques rapides ou *traits* –, avec le doigté utilisé pour chacune d'entre elles. Notre méthode pourrait alors être évaluée en fonction de son taux de reconnaissance du doigté. Ce serait l'occasion d'ajuster la formule de corrélation encore naïve que nous avons avancée. La prise en compte de données statistiques sur une base d'apprentissage nous aiderait à mieux pondérer les transitions. Enfin, il serait intéressant de savoir si avec une telle méthode il est possible de reconnaître l'empreinte physico-technique d'une main donnée à partir d'une performance pianistique (éventuellement longue), et donc d'identifier un interprète.

Nos travaux sur l'évaluation technique de saxophonistes sont orientés vers la pédagogie de l'instrument. Nous avons abouti dans le chapitre 6 à la présentation d'une nouvelle méthode automatique pour l'évaluation technique de la performance instrumentale. Fondée sur des exercices musicaux simples et pratiqués par tous, elle permet d'évaluer des saxophonistes, et de les classer parmi d'autres interprètes. La démarche a été validée par des professeurs de saxophone, et les résultats de l'évaluation automatique correspondent

bien aux niveaux académiques des instrumentistes évalués. Nous avons alors imaginé une méthode de tutorisation automatique, qui intégrerait la méthode d'évaluation, le classement automatique et la visualisation des résultats. Pour proposer un logiciel complet d'auto-apprentissage par l'évaluation, il faudrait intégrer des exercices instrumentaux, ou des conseils pédagogiques, que nous proposerions à l'élève en fonction de ses résultats techniques. Étant donné que notre méthode d'évaluation est fondée sur l'étude de l'évolution des paramètres sonores dans le temps, elle peut naturellement être généralisée à tous les instruments entretenus (trompette, violon, ...), et même à la voix ; il serait intéressant de faire une étude pour vérifier cette assertion. Une autre perspective possible est l'étude de l'évolution du timbre : au lieu de suivre les évolutions de la fréquence fondamentale et de l'amplitude en fonction du temps, nous prévoyons d'observer celle des différents partiels du son. Il faut pour cela utiliser une analyse sinusoïdale, comme nous l'avons précédemment fait avec Lagrange [RL06], plutôt qu'une technique d'estimation de fréquence fondamentale. L'évolution de l'enveloppe spectrale d'un instrumentiste ainsi observée donnerait de nouvelles informations sur la technique de l'élève, mais aussi sur la performance et le timbre du saxophone en général. Les phases d'attaque du son sont très importantes dans l'interprétation instrumentale, comme nous l'a rappelé Jean-Marie Londeix [Umb00], et pourraient également être observées avec plus d'intérêt.

Après l'analyse musicale, nous avons exploré dans la partie III des algorithmes de synthèse... sonore. Afin d'accélérer la synthèse additive, nous avons d'abord répertorié et étudié des techniques non linéaires anciennes, puis nous avons proposé une nouvelle méthode de synthèse utilisant des polynômes.

Les différentes techniques non linéaires étudiées dans le chapitre 7 peuvent toutes générer des signaux sonores riches et harmoniques avec peu d'oscillateurs. Nous avons essayé de tirer partie de cette propriété pour remplacer avantageusement la synthèse additive. Mais les fonctions mathématiques génératrices déterminent la forme de l'enveloppe spectrale des sons produits. Ces techniques ne sont pas assez souples pour remplacer une synthèse additive classique pour tous les types de son. Cette étude a cependant permis de nous intéresser à la synthèse par formes analytiques. Cette technique peut générer avec très peu d'oscillateurs un peigne harmonique riche et à bande limitée choisie. Il serait intéressant, et à priori peu complexe, de filtrer le signal résultant d'une telle synthèse. En combinant les deux étapes, nous pouvons espérer obtenir dans le cas des sons harmoniques un résultat sonore équivalent à la synthèse additive, mais plus rapide que les méthodes classiques.

Nous avons présenté PASS dans le chapitre 8, une nouvelle méthode de synthèse additive polynomiale. Les résultats de la méthode sont encourageants, surtout quand les oscillateurs du signal ont des fréquences basses. Nous avons proposé de combiner PASS avec la méthode du résonateur numérique dans une méthode hybride, afin d'exploiter les meilleurs performances de chacune des méthodes en fonction des conditions de synthèse. Nous devons maintenant étudier les détails d'implémentation et les performances d'une telle méthode. L'algorithme du résonateur numérique, qui nous a servi de référence pour nos comparaisons, fixe en pratique les paramètres sonores sur de courtes fenêtres temporelles pour chaque oscillateur. Si nous autorisons PASS à faire de même, nous pouvons imaginer remplacer la gestion de la file de priorité des événements temporels par un tableau d'événements pré-calculé. Cette technique a déjà été utilisée, dans l'application du PASS à la synthèse réaliste de surfaces océaniques, et elle a donné des résultats très satisfaisants.

Bibliographie

- [ABB⁺01] Federico Avanzini, Balázs Bank, Gianpaolo Borin, Giovanni De Poli, Federico Fontana, and Davide Rocchesso. Musical instrument modeling : The case of the piano. In *Proceedings of the MOSART Workshop on Current Research Directions in Computer Music*, pages 124–133, Barcelona, Spain, 2001.
- [Abe73] Harold F. Abeles. Development and validation of clarinet performance adjudication scale. *Journal of Research in Music Education*, 21 :246–255, 1973.
- [AdMS97] Josep L. Arcos, Ramon L. de Mantaras, and Xavier Serra. Saxex : A case-based reasoning system for generating expressive musical performances. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 329–336, Thessaloniki, Greece, 1997.
- [AHU83] Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *Data Structures and Algorithms*, pages 392–407. Series in Computer Science and Information Processing. Addison-Wesley, 1983.
- [AP94] Agnar Aamodt and Enric Plaza. Case-based reasoning : Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications (AICom)*, 7(1) :39–59, 1994.
- [Arf79] Daniel Arfib. Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine waves. *Journal of the Audio Engineering Society*, 27(10) :757–768, 1979.
- [Bel57] Richard E. Bellman. *Dynamic Programming*. Princeton University Press, New Jersey, 1957.
- [Bel03] Juan P. Bello. *Towards the Automated Analysis of Simple Polyphonic Music : A Knowledge-based Approach*. PhD thesis, University of London, 2003.
- [Ber89] Martin J. Bergee. An objectively constructed rating scale for euphonium and tuba music performance. *Dialogue in Instrumental Music Education*, 13 :65–86, 1989.
- [BF99] Roberto Bresin and Anders Friberg. Synthesis and decoding of emotionally expressive music performance. In *Proceedings of the IEEE 1999 Systems, Man and Cybernetics Conference (SMC)*, pages 317–322, 1999.
- [BF00a] Roberto Bresin and Anders Friberg. Emotional coloring of computer-controlled music performances. *Computer Music Journal*, 24(4) :44–63, 2000.
- [BF00b] Roberto Bresin and Anders Friberg. Rule-based emotional colouring of music performance. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 364–367, 2000.

- [BFS02] Roberto Bresin, Anders Friberg, and Johan Sundberg. Director musices : The KTH performance rules system. In *Proceedings of SIGMUS-46*, pages 43–48, Kyoto, 2002.
- [BFX⁺89] Fadi J. Bejjani, Lawrence Ferrara, Naiming Xu, Concetta M. Tomaino, Lazaros Pavlidis, Junli Wu, and Jan Dommerholt. Comparison of three piano techniques as an implementation of a proposed experimental design. *Medical Problems of Performing Artists*, 4(3) :109, 1989.
- [BM97] Richard Boulanger and Max Mathews. The 1997 mathews radio-baton and improvisation modes. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 395–398, Thessaloniki, Greece, 1997.
- [Boe93] Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences, Amsterdam*, volume 17, pages 97–110, 1993.
- [Boe01] Paul Boersma. Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10) :341–345, 2001.
- [BP01] Catherine Bros and Marc Papillon. *La main du pianiste*. Collection Médecine des arts. Alexitère Editions, 2001. Méthode d'éducation posturale progressive.
- [Bru79] Marc Le Brun. Digital waveshaping synthesis. *Journal of the Audio Engineering Society*, 27(4) :250–266, 1979.
- [BS03] Jose Bretos and Johan Sundberg. Measurements of vibrato parameters in long sustained crescendo notes as sung by ten sopranos. *Journal of Voice*, 17(3) :343–352, 2003.
- [BW06] Paul Boersma and David Weenink. Praat : Doing phonetics by computer (version 4.4.28) [computer program], 2006. Retrieved in June, 2006, from <http://www.praat.org/>.
- [Cho73] John Chowning. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7) :526–534, 1973.
- [Clo44] Ernest Closson. *Histoire du Piano*. Éditions Universitaires, Bruxelles, 1944.
- [Cly98] Manfred Clynes. Superconductor, 1998. <http://www.superconductor.com>.
- [COI⁺04] Keisuke Chida, Isamu Okuma, Shuzo Isoda, Yukako Saisu, Kunimitsu Wakamatsu, Kazufumi Nishikawa, Jorge Solis, Hideaki Takanobu, and Atsuo Takanishi. Development of a new anthropomorphic flutist robot wf-4. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 152–157, 2004.
- [Col02] Patrice Collen. *Techniques d'Enrichissement de Spectre des Signaux Audio-numériques*. PhD thesis, École Nationale Supérieure des Télécommunications, 2002.
- [Con05] I-Maestro Consortium. I-maestro. Online. <http://www.i-maestro.org>, 2005.
- [CPRS97] Eric F. Clarke, Richard Parncutt, Matti Raekallio, and John A. Sloboda. Talking fingers : An interview study of pianists views on fingering. *Musicae Scientiae*, 1(1) :87–107, 1997.

- [DCM99] Myriam Desainte-Catherine and Sylvain Marchand. Vers un modèle pour unifier musique et son dans une composition multiéchelle. In *Proceedings of the Journées d'Informatique Musicale (JIM)*, pages 59–68, 1999.
- [Des50] Lucette Descaves. *Technique des Gammes, Arpèges, Tenues et Accords pour le Piano*, volume 1 of *Crescendo*. Gérard Billaudot, 1950.
- [DGW02] Simon Dixon, Werner Goebel, and Gerhard Widmer. The performance worm : Real time visualisation of expression based on langner's tempo-loudness animation. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 361–364, Göteborg, Sweden, 2002.
- [DGW05] Simon Dixon, Werner Goebel, and Gerhard Widmer. The 'air worm' : An interface for real-time manipulation of expressive music performance. In *Proceedings of the International Computer Music Conference (ICMC2005)*, pages 614–617, Barcelona, Spain, 2005.
- [Dix03] Simon Dixon. Towards automatic analysis of expressive performance. In *Proceedings of the Triennial Conference of the European Society for the Cognitive Sciences of Music*, pages 107–110, Hanover, Germany, 2003.
- [Dix04] Simon Dixon. *Computer Graphics and Multimedia : Applications, Problems, and Solutions*, chapter Analysis of Musical Content in Digital Audio, pages 214–235. J. DiMarco, 2004.
- [Dix06] Simon Dixon. Onset detection revisited. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2006.
- [Dor42] Frederick Dorian. *The History of Music in Performance : The Art of Musical Interpretation from the Renaissance to our Day*. Norton and Co., New York, 1942.
- [DQ97] Yinong Ding and Xiaoshu Qian. Processing of musical tones using a combined quadratic polynomial-phase sinusoid and residual (quasar) signal model. *Journal of the AES*, 45(7/8) :571–584, 1997.
- [DSJ⁺90] Roger B. Dannenberg, Marta Sanchez, Annabelle Joseph, Robert Joseph, Ronald Saul, and Peter Capell. A computer-based multi-media tutor for beginning piano students. *Journal of New Music Research*, 19(2-3) :155–173, 1990.
- [DSJ⁺93] Roger B. Dannenberg, Marta Sanchez, Annabelle Joseph, Robert Joseph, Ronald Saul, and Peter Capell. Results from the piano tutor project. In *Proceedings of the Fourth Biennial Arts and Technology Symposium*, pages 143–150, 1993.
- [FBM⁺03] Anders Friberg, Gerald Bennett, Max Mathews, Craig Sapp, and Johan Sundberg. A marriage of the director musices program and the conductor program. In *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03)*, volume 1, pages 13–16, Stockholm, Sweden, 2003.
- [FFBS91] Anders Friberg, Lars Fryden, Lars G. Bodin, and Johan Sundberg. Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal*, 15(2) :49–55, 1991.

- [FLO⁺04] Dominique Fober, Stéphane Letz, Yann Orlarey, Anders Askenfeld, Kjetil Hansen, and Erwin Schoonderwaldt. Imutus – an interactive music tuition system. In *Proceedings of the Sound and Music Computing conference (SMC)*, pages 97–103, Paris, 2004.
- [FR86] Alain Fournier and William T. Reeves. A simple model of ocean waves. *SIGGRAPH Computer Graphics*, 20(4) :75–84, 1986.
- [FR98] Neville H. Fletcher and Thomas D. Rossing. *The Physics of Musical Instruments*. Springer-Verlag, second edition, 1998.
- [FRD92] Adrian Freed, Xavier Rodet, and Philippe Depalle. Synthesis and control of hundreds of sinusoidal partials on a desktop computer without custom hardware. In *Proceedings of the International Conference on Signal Processing, Applications and Technology (ICSPAT)*, 1992.
- [FRD93] Adrian Freed, Xavier Rodet, and Philippe Depalle. Performance, synthesis and control of additive synthesis on a desktop computer using FFT^{-1} . In *Proceedings of the International Computer Music Conference (ICMC)*, Tokyo, Japan, 1993.
- [Fri91] Anders Friberg. Generative rules for music performance : A formal description of a rule system. *Computer Music Journal*, 15(2) :56–71, 1991.
- [Fri95] Anders Friberg. *A Quantitative Rule System for Musical Performance*. PhD thesis, Royal Institute of Technology, Sweden, 1995.
- [Fré06a] Jocelyn Fréchet. Aqua library, 2006. <http://www.nongnu.org/aqua/>.
- [Fré06b] Jocelyn Fréchet. Realistic simulation of ocean surface using wave spectra. In *Proceedings of the International Conference on Computer Graphics Theory and Applications (GRAPP)*, pages 76–83, 2006.
- [FSF95] Anders Friberg, Johan Sundberg, and Lars Fryden. The KTH rules for musical performance : Overview and recent additions. In *Proceedings of the International Congress on Acoustics (ICA)*, volume 3, pages 431–434, Trondheim, Norway, 1995.
- [Fuk97] Leonardo Fuks. Prediction and measurements of exhaled air effects in the pitch of wind instruments. *Proceedings of the Institute of Acoustics*, 19(5, Book 2) :373–378, 1997.
- [Gab99] Alf Gabrielsson. *The Psychology of Music*, chapter The Performance of Music, pages 501–602. Diana Deutsch, second edition, 1999.
- [GMdM⁺03] Laurent Girin, Sylvain Marchand, Josph di Martino, Axel Röbel, and Geofroy Peeters. Comparing the order of a polynomial phase model for the synthesis of quasi-harmonic audio signals. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 193–196, 2003.
- [Goe01] Werner Goebel. Melody lead in piano performance : Expressive device or artifact? *Journal of the Acoustical Society of America*, 110(1) :563–572, 2001.
- [GS85] John W. Gordon and Julius O. Smith. A sine generation algorithm for vlsi applications. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 165–168, 1985.

- [GST05] Joël Gilbert, Laurent Simon, and Jonathan Terroir. Vibrato of saxophones. *Acoustical Society of America Journal*, 118(4) :2649–2655, 2005.
- [Haa01] Joachim Haas. SALTO - a spectral domain saxophone synthesizer. In *Proceedings of MOSART Workshop on Current Research Directions in Computer Music*, Barcelona, 2001.
- [Hau99] Jean Haury. *Les Nouveaux Gestes de la Musique*, chapter Petite Histoire Illustrée de l’Interface Clavier, pages 93–110. Cultures Musicales. Éditions Parenthèses, 1999.
- [HBH93] David C. Harding, Kenneth D. Brandt, and Ben M. Hillberry. Finger joint force minimization in pianists using optimization techniques. *Journal of Biomechanics*, 26(12) :1403–1412, 1993.
- [HNC02] Damien Hinsinger, Fabrice Neyret, and Marie P. Cani. Interactive animation of ocean waves. In *Symposium on Computer Animation*, pages 161–166, 2002.
- [Jac01] J. Pieter Jacobs. Refinements to the ergonomic model for keyboard fingering of Parncutt, Sloboda, Clarke, Raekellio, and Desain. *Music Perception*, 18(4) :505–511, 2001.
- [Jen02] Kristoffer Jensen. Musical instruments parametric evolution. In *Proceedings of the International Symposium on Musical Acoustics (ISMA)*, Mexico, 2002.
- [Joh67] Stephen C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 2 :241–254, 1967.
- [Joh91] Margaret L. Johnson. Toward an expert system for expressive musical performance. *IEEE Computer*, 24(7) :30–34, 1991.
- [Kap05] Ajay Kapur. A history of robotic musical instruments. In *Proceedings of the International Computer Music Conference (ICMC)*, 2005.
- [KBGA03] Reinhard Kopiez, Marc Bangert, Werner Goebel, and Eckart Altenmüller. Tempo and loudness analysis of a continuous 28-hour performance of Erik Satie’s composition ”Vexations”. *Journal of New Music Research*, 32(3) :243–258, 2003.
- [KV90] Christine L. Mac Kenzie and Dwayne VanEerd. Rhythmic precision in piano scales : Motor psychophysics and motor programming. In M. Jeannerod, editor, *Proceedings of the Thirteenth Symposium on Attention and Performance*, pages 375–408, 1990. Hillsdale, Lawrence Erlbaum Associates.
- [Lag04] Mathieu Lagrange. *Sinusoidal Modeling of Polyphonic Sounds*. PhD thesis, University of Bordeaux 1, Bordeaux, France, 2004.
- [LG03] Jörg Langner and Werner Goebel. Visualizing expressive performance in tempo-loudness space. *Computer Music Journal*, 27(4) :69–83, 2003.
- [LM01] Mathieu Lagrange and Sylvain Marchand. Real-time additive synthesis of sound by taking advantage of psychoacoustics. In *Proceedings of the International Conference on Audio Effects (DAFx)*, pages 5–9, 2001.
- [Lon62] Jean-Marie Londeix. *Les Gammes Conjointes et en Intervalles pour Tous les Saxophones*. Éditions Henry Lemoine, Paris, 1962.

- [Lon70] Jean-Marie Londeix. *Le Saxophone en Jouant*, volume 3. Éditions Henry Lemoine, 1970.
- [Lon77] Jean-Marie Londeix. *Exercices Mécaniques pour tous les Saxophones*, volume 1. Éditions Henry Lemoine, Paris, 1977.
- [Mar00] Sylvain Marchand. *Sound Models for Computer Music (Analysis, Transformation, and Synthesis of Musical Sound)*. PhD thesis, University of Bordeaux 1, France, 2000.
- [Mar01a] Sylvain Marchand. An efficient pitch-tracking algorithm using a combination of Fourier transforms. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, pages 170–174, Limerick, Ireland, 2001.
- [Mar01b] Jean Martin. *Jouez avec Doigté... Deux Mains... Dix Doigts... Mais Tellement Plus*. Aléas, 2001.
- [Mat89] Max Mathews. *The Conductor Program and Mechanical Baton*. Current Directions in Computer Music Research. MIT press, Cambridge, Massachusetts, 1989.
- [MB79] John Makhoul and Michael Berouti. High frequency regeneration in speech coding systems. In *IEEE ICASSP*, pages 428–431, 1979.
- [MG83] Brian C.J. Moore and Brian R. Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74 :750–753, 1983.
- [MM98] Jean Massin and Brigitte Massin. *Histoire de la musique occidentale*. Les indispensables de la musique. Fayard, 1998. ISBN 2-213-02032-9.
- [Moo76] James A. Moorer. The synthesis of complex audio spectra by means of discrete summation formulae. *Journal of the Audio Engineering Society*, 24 :717–727, 1976.
- [Moo77] James A. Moorer. Signal processing aspects of computer music – a survey. *Computer Music Journal*, 1(1) :4–37, 1977.
- [MP02] Nikolaus Meine and Heiko Purnhagen. Fast sinusoid synthesis for MPEG-4 HILN parametric audio decoding. In *Proceedings of the International Conference on Audio Effects (DAFx)*, pages 239–244, 2002.
- [MQ86] Robert J. McAulay and Thomas F. Quatieri. Speech analysis / synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(4) :744–754, 1986.
- [MR04] Sylvain Marchand and Martin Raspaud. Enhanced time-stretching using order-2 sinusoidal modeling. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, pages 76–82, Naples, Italy, 2004.
- [MS99] Sylvain Marchand and Robert Strandh. InSpect and ReSpect : Spectral modeling, analysis and real-time synthesis software tools for researchers and composers. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 341–344, 1999.
- [Mul50] Marcel Mule. *Études Variées dans Toutes les Tonalités*. Alphonse Leduc, Paris, 1950.

- [MWD⁺95] Stephen McAdams, Suzanne Winsberg, Sophie Donnadiou, Geert De Soete, and Jochen Krimphoff. Perceptual scaling of synthesized musical timbres : Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58 :177–192, 1995.
- [MWM87] Gary A. Mastin, Peter A. Watterberg, and John F. Mareda. Fourier synthesis of ocean scenes. *IEEE Computer Graphics and Applications*, 7(3) :16–23, 1987.
- [Nar90] Eugène Narmour. *The Analysis and Cognition of Basic Melodic Structures : the Implicationrealization Model*. University of Chicago Press, Chicago, 1990. ISBN 0-226-56845-8.
- [Ort62] Otto Ortmann. *The Physiological Mechanics of Piano Technique*. Dutton, 1962.
- [Pal96] Caroline Palmer. On the assignment of structure in music performance. *Music Perception*, 14 :23–56, 1996.
- [Pal97] Caroline Palmer. Music performance. *Annual Review of Psychology*, 48 :115–138, 1997.
- [Par95] Richard Parncutt. Recording piano fingering in live performance. In *Proceedings of the KlangArt-Kongress*, pages 263–268, Osnabrück, Germany, 1995.
- [Pol83] Giovanni De Poli. A tutorial on digital sound synthesis techniques. *Computer Music Journal*, 7(2) :76–87, 1983.
- [PSC⁺97] Richard Parncutt, John A. Sloboda, Eric F. Clarke, Matti Raekallio, and Peter Desain. An ergonomic model of keyboard fingering for melodic fragments. *Music Perception*, 14(4) :341–382, 1997.
- [PT02] Richard Parncutt and Malcolm Troup. *Science and Psychology of Music Performance : Creative Strategies for Teaching and Learning*, chapter Piano, pages 285–302. R. Parncutt and G. E. McPherson, Oxford University Press, 2002.
- [Pur00] Jim Purbrick. Automatic synthesizer articulation. *Computer Music Journal*, 24(1) :20–31, 2000.
- [RAL04a] Daniele Radicioni, Luca Anselma, and Vincenzo Lombardo. An algorithm to compute fingering for string instruments. In *Proceedings of the National Congress of the Associazione Italiana di Scienze Cognitive*, Ivrea, Italy, 2004.
- [RAL04b] Daniele Radicioni, Luca Anselma, and Vincenzo Lombardo. A segmentation-based prototype to compute string instruments fingering. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM)*, Graz, Austria, 2004.
- [Rep96] Bruno H. Repp. Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America*, 100 :3917–3932, 1996.
- [RF06] Matthias Robine and Jocelyn Fréchet. Fast additive sound synthesis for real-time simulation of ocean surface. In *Proceedings of the International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 223–226, Budapest, Hungary, 2006.

- [Ris69] Jean-Claude Risset. *Catalog of Computer Synthesized Sound*. Murray Hill, Bell Telephone Laboratories, 1969.
- [RL05a] Daniele Radicioni and Vincenzo Lombardo. Computational model of chord fingering. In *Proceedings of the Annual Conference of the Cognitive Science Society*, Stresa, Italy, 2005.
- [RL05b] Daniele Radicioni and Vincenzo Lombardo. Guitar fingering for music performance. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 527–530, 2005.
- [RL06] Matthias Robine and Mathieu Lagrange. Evaluation of the technical level of saxophone performers by considering the evolution of spectral parameters of the sound. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 79–84, Victoria, Canada, 2006.
- [RMG05] Martin Raspaud, Sylvain Marchand, and Laurent Girin. A generalized polynomial and sinusoidal model for partial tracking and time stretching. In *Proceedings of the International Conference on Audio Effects (DAFx)*, pages 24–29, 2005.
- [Roa96] Curtis Roads. *The Computer Music Tutorial*. MIT Press, 1996.
- [Rou82] Eugène Rousseau. *Marcel Mule : sa Vie et le Saxophone*. Étoile Music, Shell Lake, États-Unis, 1982.
- [RSM06] Matthias Robine, Robert Strandh, and Sylvain Marchand. Fast additive sound synthesis using polynomials. In *Proceedings of the International Conference on Digital Audio Effects (DAFX)*, pages 181–186, Montréal, Canada, 2006.
- [Say89] Samir I. Sayegh. Fingering for string instrument with the optimum path paradigm. *Computer Music Journal*, 13(6) :76–84, 1989.
- [SC92] Julius O. Smith and Perry R. Cook. The second-order digital waveguide oscillator. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 150–153, 1992.
- [Sch98] Eric D. Scheirer. *Computational Auditory Scene Analysis*, chapter Using Musical Knowledge to Extract Expressive Performance Information from Audio Recordings. Lawrence Erlbaum, Rosenthal and Okuno, 1998.
- [Ser97] Xavier Serra. *Musical Signal Processing*, chapter Musical Sound Modeling with Sinusoids plus Noise, pages 91–122. Studies on New Music Research. Swets and Zeitlinger, C. Roads, S. T. Pope, A. Piccialli, G. De Poli, Lisse, the Netherlands, 1997.
- [SHA04] Erwin Schoonderwaldt, Kjetil Hansen, and Anders Askenfeld. Imutus – an interactive system for learning to play a musical instrument. In *Proceedings of the International Conference of Interactive Computer Aided Learning (ICL)*, pages 143–150, Villach, Austria, 2004.
- [SM25] Carl E. Seashore and Milton Metfessel. Deviation from the regular as an art principle. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 11/9, pages 538–542, 1925.

- [SM99] Robert Strandh and Sylvain Marchand. Real-time generation of sound from parameters of additive synthesis. In *Proceedings of the Journées d'Informatique Musicale (JIM)*, pages 83–88, 1999.
- [SS90] Xavier Serra and Julius O. Smith. Spectral modeling synthesis : A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4) :12–24, 1990.
- [Sta01] Efstathios Stamatatos. A computational model for discriminating music performers. In *Proceedings of the MOSART Workshop on Current Research Directions in Computer Music*, pages 65–69, 2001.
- [SW02] Efstathios Stamatatos and Gerhard Widmer. Music performer recognition using an ensemble of simple classifiers. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pages 335–339, 2002.
- [Tan92] Andranick Tanguiane. *Computer Representations and Models in Music*, chapter An Analytical Approach to Musical Performance, pages 122–142. Academic Press, London, 1992.
- [TDW03] Caroline Traube, Philippe Depalle, and Marcelo Wanderley. Indirect acquisition of instrumental gesture based on signal, physical and perceptual information. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 42–47, 2003.
- [Tes01] Jerry Tessendorf. Simulating ocean water. ACM SIGGRAPH Course Notes, 2001. Updated in 2004.
- [Thi04] Jean-Pierre Thiollet. *Sax, Mule and Co - Marcel Mule ou l'Éloquence du Son*. H and D, Paris, 2004. ISBN 2914266030.
- [TS01] Caroline Traube and Julius O. Smith. Extracting the fingering and the plucking points on a guitar string from a recording. In *Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 7–10, 2001.
- [Umb00] James C. Umble. *Jean-Marie Londeix : Master of the Modern Saxophone (Maitre du saxophone moderne)*. RONCORP Publications, Cherry Hills, USA, 2000. ISBN 0-939103-06-0, Dual language book (English and French).
- [Van95] Pascale Vandervellen. *Le Piano de Style en Europe : Des Origines à 1850 : Étude des Éléments Décoratifs et Mécaniques*. Pierre Mardaga, 1995. ISBN 2870095708.
- [vB60] Georg von Békésy. *Experiments in Hearing*. McGraw-Hill series in psychology. McGraw-Hill Book Companies, 1960.
- [Vei77] Jean-Claude Veilhan. *Les Règles de l'Interpretation Musicale à l'Époque Baroque (XVIIe-XVIIIe s.)*. Éditions musicales Alphonse Leduc, 1977. selon Bach, Brossard, Couperin, Hotteterre, Montéclair, Quantz, Rameau-d'Alembert, Rousseau, etc.
- [Ver03] Vincent Verfaillie. *Effets Audionumériques Adaptatifs : Théorie, Mise en Oeuvre et Applications en Création Musicale Numérique*. PhD thesis, Université Aix-Marseille II, 2003.

- [VGD05] Vincent Verfaillie, Catherine Guastavino, and Philippe Depalle. Perceptual evaluation of vibrato models. In *Proceedings of the Conference on Interdisciplinary Musicology*, Montreal, Canada, 2005.
- [VR01] Christophe Vergez and Xavier Rodet. Trumpet and trumpet player : Modélisation and simulation in a musical context. In *Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba, 2001.
- [Wag88] Christoph Wagner. The pianist's hand : Anthropometry and biomechanics. *Ergonomics*, 31(1) :97–131, 1988.
- [WDG⁺03] Gerhard Widmer, Simon Dixon, Werner Goebel, Elias Pampalk, and Asmir Tobudic. In search of the Horowitz factor. *AI Magazine*, 24(3) :111–130, 2003.
- [Wei75] Clifford J. Weinstein. A linear prediction vocoder with voice excitation. In *Proceedings of EASCON*, 1975.
- [WF54] J.G. Watkins and S.E. Farnum. *The Watkins-Farnum Performance Scale : A Standardised Achievement Test for All Band Instruments*. Hal Leonard Music, 1954.
- [Wri05] William J. Wrigley. *Improving Music Performance Assessment*. PhD thesis, Griffith University, Brisbane, Australia, 2005.
- [Yst99] Solvi Ystad. Identification and modeling of a flute source signal. In *Proceedings of the Workshop on Digital Audio Effects (DAFx)*, pages 187–190, Trondheim, Norvège, 1999.
- [YV01a] Solvi Ystad and Thierry Voinier. Analysis-synthesis of flute sounds using a non-linear digital waveguide model. In *Proceedings of the International Computer Music Conference (ICMC)*, Cuba, 2001.
- [YV01b] Solvi Ystad and Thierry Voinier. A virtually real flute. *Computer Music Journal*, 25(2) :13–25, 2001.
- [ZF90] Eberhard Zwicker and Hugo Fastl. *Psychoacoustics, Facts and Models*. Springer Verlag, 1990.