HAL
open science

Contrôle intuitif de la synthèse sonore d'interactions solidiennes : vers les métaphores sonores

Simon Conan

To cite this version:

Simon Conan. Contrôle intuitif de la synthèse sonore d'interactions solidiennes : vers les métaphores sonores. Acoustique [physics.class-ph]. Ecole Centrale Marseille, 2014. Français. NNT : 2014ECDM0012 . tel-01152851

HAL Id: tel-01152851

https://theses.hal.science/tel-01152851

Submitted on 18 May 2015

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

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

École Doctorale : Sciences pour l'ingénieur : Mécanique, Physique, Micro et Nanoélectronique (ED353)

CNRS - LABORATOIRE DE MECANIQUE ET D'ACOUSTIQUE (UPR7051)

THÈSE DE DOCTORAT

pour obtenir le grade de

DOCTEUR de l'ÉCOLE CENTRALE de MARSEILLE

Discipline : Acoustique

Contrôle intuitif de la synthèse sonore d'interactions solidiennes – vers les métaphores sonores –

par

CONAN Simon

Directeurs de thèse : KRONLAND-MARTINET Richard, ARAMAKI Mitsuko

Soutenue le 3 Décembre 2014
devant le jury composé de :

DEPALLE Philippe	Prof. - McGill University	Rapporteur
SERAFIN Stefania	Prof. - Aalborg Universitet	Rapportrice
ROCCHESSO Davide	Prof. - Università IUAV di Venezia	Examinateur
TORRESANI Bruno	Prof. - Aix-Marseille Université	Examinateur
ARAMAKI Mitsuko	Doc. - CNRS-LMA	Directrice de thèse
KRONLAND-MARTINET Richard	Doc. - CNRS-LMA	Directeur de thèse
DERRIEN Olivier	MCF - Université de Toulon	Invité
YSTAD Sølvi	Doc. - CNRS-LMA	Invitée

Remerciements

Lorsqu'on arrive au moment d'écrire cette partie de la thèse, c'est en général plutôt bon signe. Lorsqu'on y arrive, heureux d'en être au moment où il s'agit de rendre la pareille, ça l'est d'autant plus. Ça n'est pourtant pas une tâche facile, tant de gens et de choses ces dernières années...

Mes premiers remerciements vont naturellement à mon "triumvirat" d'encadrement, qui a toujours su être bienveillant et à l'écoute, présent sans être pesant. Votre confiance et la liberté que vous m'avez accordées m'ont permis (je l'espère !), de mener à bien ces travaux dans des conditions idéales. En particulier, merci Richard (Kronland-Martinet) pour le "contrôle haut-niveau" de la thèse, et à Mitsuko (Aramaki) et Sølvi (Ystad) pour les heures innombrables passées à me lire, relire et re-relire. J'espère qu'encore beaucoup d'étudiants auront la chance d'être encadrés par des personnes aussi humaines et chaleureuses que vous.

J'aimerais également remercier toutes les personnes avec qui j'ai eu l'occasion de travailler durant ces trois années, et en particulier Olivier (Derrien) pour ne m'avoir jamais abandonné à "traiter les signaux", jamais avare en temps ni en explications pédagogiques. Un acknowledgement particulier à Thierry (Voinier), entre autres pour les cours informels de "signal processing", j'aurai bien voulu renvoyer l'ascenseur mais je ne sais malheureusement pas installer de baignoire, je demeure donc en reste pour de futurs travaux Bandolais !

Merci également à mon jury "international", de Château-Gombert à Montréal en passant par Venise et Aalborg, pour avoir pris le temps de relire et critiquer mes travaux. J'ai beaucoup apprécié l'échange que nous avons pu avoir lors de la soutenance et ensuite.

Si je n'avais qu'un paragraphe personnel à faire, ce serait bien pour Etienne (Thoret), mon binôme de thèse (Knoll Dupond og Tott Dupont !), pour le soutien mutuel permanent, les quantités de discussions philosophico-scientifiques (ou scientifico-philosophiques je ne sais pas très bien) toujours enrichissantes, et puis aussi tout simplement pour être un type bien, un vrai pote, un mec avec qui c'est agréable de bosser. Comme dirait un grand philosophe "pourvu que ça dure".

Un grand merci également à Charles "Mad MAX/MSP" (Gondre) pour ton immense aide technique, je ne suis pas sûr que le synthé roulerait aussi bien sans toi.

Merci également à Sébastien "Danger Suduf l'Induss" (Denjean) pour avoir allumé le feu de notre rude hiver Marseillais 2013.

Merci spécial à Soizic "Blanc-Lim" (Terrien) pour tout ce que tu m'apportes depuis déjà un certain temps.

Ces années passées dans le bâtiment AA n'auraient pas été aussi agréables sans toutes les personnes qui y sont et y ont été, doctorants, post-docs, stagiaires, les personnes qui passent dire bonjour ou boire un café... Un salut particulier à mes colocataires et ex-colocataires de bureau, Lennie "Brousse" (Gandemer), Jocelyn "Mocap" (Rozé), et Adrien "Lamantin" (Sirdey). Un remerciement spécial à Charly (Verron) qui

m'a initié aux joies de la synthèse sonore, c'est un peu à cause de toi si je suis là maintenant ! Une bonne bise également à Anaïk (Olivero), Pyo (Michaud), Quentin (Mesnildrey), Gauthier (Real) -au sens large-, Jean-Baptiste (Doc), Fabrice (Silva), Gaëtan (Parseihian), Sami (Karkar), le gars Clément (François) et Adrien (Merer). Un petit coucou également à Christophe (Vergez) pour les discussions "pongistiques", ainsi qu'à Erick (Ogam) pour la bonne humeur permanente, même (et surtout !) lorsqu'on parle de crâne, de fémur et d'Obama. Le bâtiment AA ne serait pas aussi gai sans les passages fréquents de Michèle (Laurent), notre gestionnaire de choc, toujours arrangeante – merci ! Un salut chaleureux également aux collègues du bâtiment P, en particulier Jacques (Chatron), Sabine (Meunier), Olivier (Macherey) et Sophie (Savel) pour les discussions toujours sympathiques. Merci enfin à "fleur du pays" et "torréfaction Noailles" pour leur soutien "psychotropique", en particulier durant la rédaction.

Ces années passées hors du bâtiment AA n'auraient pas été aussi délicieuses sans tous les amis, de Marseille et d'ailleurs. Tout d'abord un énorme merci au 28, mon lieu de villégiature principal, le meilleur dancefloor disco Marseillais, et tous ses habitants parce que vous êtes "OK" ! Un salut particulier à Harold (Omer), mon "compagnon d'infortune" Marseillais, j'espère continuer à déconner encore longtemps avec toi. Un petit coucou aux divers groupes de potes, le club Mickey, les Caëristes, le Netto Mx Crew, les copains de centrale (big up particulier à Simon "numéro 3" (Benacchio) qui a soutenu le même jour que moi, gros style). Un grand merci aux groupes de musique que j'ai traversés et qui m'ont permis de bien me défouler, Le Pompier Poney Club (ma fanfare de cœur), Zirkon, les Frères Pourcel. Enfin, Marseille ne serait pas ce qu'elle est sans la Plaine et ses bars, une dédicace particulière à Saïd du Bar de la Plaine, notre sas de décompression durant la rédaction, ainsi qu'à Gérard du Point Bar pour m'avoir permis de faire cette superbe soirée de fin de thèse.

Pour finir, merci Maman Nelly et Papa Philippe ("pas fier mais heureux" !) pour m'avoir soutenu durant mes études, sans forcer, ainsi qu'à Parâtre Jacquot et Marâtre Odile. Et merci aussi aux (nombreux) frères et sœurs, en particulier à Anjela et Manu pour être venus me soutenir le jour de la soutenance (et les banderolles ? !).

"Tout ça me plaisait. J'avais envie de couiner,
de faire des bruits bizarres, des bruits inédits."
John Fante*, La route de Los Angeles*

Résumé

Un des enjeux actuels de la synthèse sonore est le contrôle perceptif (i.e. à partir d'évocations) des processus de synthèse. En effet, les modèles de synthèse sonore dépendent généralement d'un grand nombre de paramètres de bas niveau dont la manipulation nécessite une expertise des processus génératifs. Disposer de contrôles perceptifs sur un synthétiseur offre cependant beaucoup d'avantages en permettant de générer les sons à partir d'une description du ressenti et en offrant à des utilisateurs non-experts la possibilité de créer et de contrôler des sons intuitivement. Un tel contrôle n'est pas immédiat et se base sur des hypothèses fortes liées à notre perception, notamment la présence de morphologies acoustiques, dénommées "invariants", responsables de l'identification d'un évènement sonore.

Cette thèse aborde cette problématique en se focalisant sur les invariants liés à l'action responsable de la génération des sons. Elle s'articule suivant deux parties. La première a pour but d'identifier des invariants responsables de la reconnaissance de certaines interactions continues : le frottement, le grattement et le roulement. Le but est de mettre en œuvre un modèle de synthèse temps-réel contrôlable intuitivement et permettant d'effectuer des transitions perceptives continues entre ces différents types d'interactions (e.g. transformer progressivement un son de frottement en un son de roulement). Ce modèle s'inscrira dans le cadre du paradigme "action-objet" qui stipule que chaque son résulte d'une action (e.g. gratter) sur un objet (e.g. une plaque en bois). Ce paradigme s'adapte naturellement à une approche de la synthèse par modèle source-filtre, où l'information sur l'objet est contenue dans le "filtre", et l'information sur l'action dans la "source". Pour ce faire, diverses approches sont abordées : études de modèles physiques, approches phénoménologiques et tests perceptifs sur des sons enregistrés et synthétisés.

La seconde partie de la thèse concerne le concept de "métaphores sonores" en élargissant la notion d'objet à des textures sonores variées. La question posée est la suivante : étant donnée une texture sonore quelconque, est-il possible de modifier ses propriétés intrinsèques pour qu'elle évoque une interaction particulière comme un frottement ou un roulement par exemple ? Pour créer ces métaphores, un processus de synthèse croisée est utilisé dans lequel la partie "source" est basée sur les morphologies sonores des actions précédemment identifiées et la partie "filtre" restitue les propriétés de la texture. L'ensemble de ces travaux ainsi que le paradigme choisi offre dès lors de nouvelles perspectives pour la constitution d'un véritable langage des sons.

Site internet

Ce document est accompagné d'un site internet, comprenant des exemples sonores et vidéos de démonstration des outils de synthèse développés au cours de cette thèse. Le symbole ($\sharp$) dans le texte indique que des exemples sonores sont disponibles sur la page qui se trouve à l'adresse suivante :

`http://www.lma.cnrs-mrs.fr/~kronland/TheseSConan/`

Table des matières

Remerciements **i**

Résumé **iv**

Site internet **v**

Introduction **1**

I Contexte et enjeux de la thèse **3**
A Synthèse sonore . . . 3
A.1 Modèles de signaux . . . 5
A.2 Modèles physiques . . . 7
B Contrôle de la synthèse sonore . . . 8
B.1 Contrôle du timbre . . . 10
B.2 Contrôle des attributs perceptifs de la source sonore . . . 10
C Domaines d'utilisation de la synthèse sonore . . . 12
D Paradigme d'étude . . . 13
D.1 Approche écologique de la perception . . . 14
D.2 Invariants perceptifs : comment le son nous informe ? . . . 16
D.3 Paradigme *action-objet* . . . 18
D.4 Modèle de synthèse adopté . . . 19
E Enjeux de la thèse . . . 21
E.1 A partir des sons d'interactions entre objets solides... . . . 21
E.2 ... vers les métaphores sonores . . . 22
E.3 Méthodologie et organisation du document . . . 23

II Synthèse et contrôle haut-niveau de sons de roulement **25**
A Synthèse et perception des sons de roulement : état de l'art . . . 25
A.1 Synthèse de sons de roulement . . . 26
A.1.1 Les modèles basés sur la physique . . . 26
A.1.2 Les modèles empiriques de signaux . . . 28
A.1.3 Les modèles basés sur des schémas d'analyse/synthèse 29
A.2 Perception des sons de roulement . . . 29
B Mise en évidence d'un invariant transformationnel du roulement . . . 31
B.1 Sujets . . . 32
B.2 Stimuli . . . 32
B.3 Protocole . . . 33
B.4 Résultats . . . 33
B.5 Discussion . . . 33
C Modélisation de l'invariant transformationnel . . . 34

C.1 Caractérisation de la force d'interaction 35
C.2 Schéma d'analyse/synthèse de la séquence d'impact 36
C.3 Indice pour la perception de la vitesse de roulement 38
C.4 Modélisation de la forme de l'impact 39
C.5 Estimation des paramètres 41
D Stratégie de contrôle intuitif 43
D.1 Contrôle de la taille de la bille 43
D.2 Contrôle de la vitesse de la bille 44
D.3 Contrôle de la rugosité de la surface 44
E Evaluation perceptive de la stratégie de contrôle intuitif 46
E.1 Sujets 46
E.2 Stimuli 46
E.3 Protocole 46
E.4 Résultats 47
E.5 Discussion 47
F Discussion générale 48

III Extension du Modèle à d'Autres Interactions et Stratégie de Contrôle du Synthétiseur 51
A Sons de friction : état de l'art 52
A.1 Synthèse de sons de friction 52
A.1.1 Synthèse de sons de friction linéaire 52
A.1.2 Synthèse de sons de friction non-linéaire 53
A.2 Perception et utilisation des sons de friction 53
B Etude perceptive des interactions "frotter" et "gratter" 54
B.1 Catégorisation perceptive de sons de friction enregistrés 54
B.1.1 Sujets 55
B.1.2 Stimuli 55
B.1.3 Protocole 55
B.1.4 Résultats 56
B.1.5 Discussion 56
B.2 Analyse qualitative des sons catégorisés 57
C Contrôle perceptif des actions "frotter" et "gratter" 60
C.1 Description du contrôle 61
C.2 Validation par synthèse 62
C.2.1 Sujets 62
C.2.2 Stimuli 62
C.2.3 Protocole 63
C.2.4 Résultats 63
C.2.5 Discussion 63
D Modèle générique de sons d'interactions continues 65
D.1 Description des paramètres du modèle 65
D.2 Morphologie des différentes interactions 66
D.2.1 Morphologie de l'interaction "gratter" 66
D.2.2 Morphologie de l'interaction "frotter" 67
D.2.3 Morphologie de l'interaction "rouler" 67
D.3 Stratégie de navigation dans l'espace des actions 67
D.3.1 Définition des prototypes 67
D.3.2 Espace sonore des interactions 68

E Perspectives d'élargissement de l'espace sonore des interactions : vers la friction non-linéaire . . . 69
E.1 Modèle de synthèse source-filtre de sons de friction non-linéaire 70
E.2 Construction du modèle d'excitation pour les actions "rouler", "frotter" et "gratter" en synthèse additive . . . 73
E.2.1 Modèle simplifié . . . 73
E.2.2 Application aux interactions rouler, frotter et gratter . . 75
F Discussion générale . . . 76

IV Métaphores Sonores 78
A Les textures sonores : définition . . . 79
B Synthèse de textures sonores : état de l'art . . . 80
B.1 Synthèse soustractive de textures sonores . . . 81
B.2 Autres méthodes de synthèse de textures sonores . . . 82
C Proposition d'un modèle d'analyse/synthèse de textures sonores . . . 83
C.1 Définition du modèle . . . 84
C.2 Estimation d'un modèle AR . . . 84
C.3 Estimation d'un modèle AR en sous-bandes . . . 87
D Création de métaphores sonores . . . 91
D.1 Contribution de l'interaction . . . 91
D.2 Exemples de métaphores sonores . . . 93
E Construction du corpus sonore pour les tests perceptifs . . . 94
E.1 Choix des textures sonores et resynthèse . . . 94
E.2 Choix des interactions . . . 95
F Expérience 1 : métaphores sonores vs. mélange des sons . . . 97
F.1 Sujets . . . 97
F.2 Stimuli . . . 97
F.3 Protocole . . . 98
F.4 Résultats . . . 98
F.5 Discussion . . . 99
G Expérience 2 : reconnaissance des interactions dans la métaphore . . . 100
G.1 Sujets . . . 100
G.2 Stimuli . . . 100
G.3 Protocole . . . 100
G.4 Résultats . . . 100
G.5 Discussion . . . 102
H Expérience 3 : reconnaissance des textures originales dans la métaphore 103
H.1 Sujets . . . 103
H.2 Stimuli . . . 103
H.3 Protocole . . . 103
H.4 Résultats . . . 104
H.5 Discussion . . . 105
I Discussion générale . . . 105

Conclusion et Perspectives 107

A Modélisation des densités de probabilité des séries d'impacts du roulement 127

B Fonction de détection d'attaques 129

C Crédits corpus de textures sonores 131

D Publications associées à la thèse **132**

Introduction

Cette thèse s'est intéressée à la synthèse sonore et au contrôle perceptif (i.e. à partir d'évocations) de celle-ci. Les modèles de synthèse sonore dépendant généralement d'un grand nombre de paramètres de bas niveau dont la manipulation nécessite une expertise des processus génératifs. Disposer de contrôles perceptifs sur un synthétiseur offre cependant beaucoup d'avantages en permettant de générer les sons à partir d'une description du ressenti et en offrant à des utilisateurs non-experts la possibilité de créer et de contrôler des sons intuitivement. Le contrôle haut-niveau de la synthèse de sons ayant un contenu sémiotique particulier est donc un enjeu actuel et présente un grand intérêt, notamment pour les designers sonores, ces outils permettant de créer et manipuler les sons pour des applications variées. Parmi ces applications on retrouve par exemple le son pour les jeux vidéos (Böttcher, 2013; Lloyd *et al.*, 2011), le design sonore (Farnell, 2010; Susini *et al.*, 2014), la sonification (Hermann *et al.*, 2011; Dubus et Bresin, 2013) et la réalité virtuelle/augmentée pour la réhabilitation motrice (Danna *et al.*, 2013; Rodger *et al.*, 2014). Jusqu'à présent, la majeure partie de ces applications utilisent des grandes banques de sons indexées par des attributs verbaux, mais cette approche à des limites. Premièrement, concevoir le son voulu (le son que l'on s'imagine) à partir d'échantillons d'une base de données peut être un processus long et fastidieux, et nécessite une grande expertise. Deuxièmement, tous les scénarii sonores doivent être prévus à l'avance, ne laissant pas de possibilités d'interactions imprévues entre l'utilisateur et son environnement. D'autres stratégies de sonification que l'utilisation de sons pré-enregistrés existent, mais se basent souvent sur la modification d'attributs sonores basiques (hauteur tonale, tempo, spatialisation, etc) (Dubus et Bresin, 2013) et permettent difficilement des manipulations subtiles du timbre selon des évocations induites par le son. La synthèse sonore évocative et contrôlable offre donc une alternative intéressante aux limitations des approches précédemment présentées.

Outre ces enjeux généraux, les travaux de cette thèse ont été plus spécifiquement développés dans le cadre du projet ANR MétaSon. Ce projet pluridisciplinaire réunit le Laboratoire de Mécanique et d'Acoustique, l'Institut de Mathématiques de Marseille, le Laboratoire de Neurosciences Cognitives ainsi que le partenaire industriel Peugeot-Citroën. Le projet MétaSon s'intéresse à définir des stratégies susceptibles d'informer l'humain sur les évolutions d'un système dynamique dans un contexte cognitif spécifique. Le son est une modalité sensorielle qui s'y prête tout à fait. Deux applications principales sont concernées par ce projet. La première est la sonification des véhicules électriques. En effet, à basse vitesse, les véhicules électriques sont silencieux et représentent donc un danger pour les piétons. De plus, le silence du véhicule influe sur la conduite (Denjean *et al.*, 2012, 2013). Il est donc nécessaire d'informer par le son, à l'intérieur comme à l'extérieur du véhicule, de la dynamique et du danger potentiel, tout en laissant une marge créatrice dans le design sonore, marque d'identité du véhicule. La seconde application concerne le diagnostic et la rééducation des troubles de l'écriture liés à la dysgraphie. Celle-ci, indépendante des capacités à lire et non liée à des

troubles psychologiques, désigne les difficultés à accomplir des gestes graphiques. Le son, par définition dynamique, est donc une modalité naturelle pour tenter de remédier à ces troubles en informant en temps-réel de la qualité du geste par la sonification de celui-ci. Il apparaît donc nécessaire de pouvoir proposer des outils de génération sonore contrôlables et évocatifs. Cependant la mise en place d'un tel contrôle perceptif n'est pas immédiat et se base sur des hypothèses fortes liées à notre perception, notamment la présence de morphologies acoustiques, dénommées "invariants", responsables de l'identification d'un évènement sonore.

On abordera cette problématique en se focalisant sur les invariants sonores liés à l'évocation de certaines actions, et plus particulièrement des interactions continues entre solides. Elle s'articule suivant deux parties. La première a pour but d'identifier des invariants permettant la reconnaissance de certaines interactions continues entre objets solides : le frottement, le grattement et le roulement. Le but est de mettre en œuvre un modèle de synthèse temps-réel contrôlable intuitivement et permettant d'effectuer des transitions perceptives continues entre ces différents types d'interactions. Pour ce faire, diverses approches sont abordées : études de modèles physiques, approches phénoménologiques et tests d'écoute sur des sons enregistrés et de synthèse. La seconde partie de la thèse concerne le concept de "métaphores sonores". La question posée est la suivante : étant donnée une texture sonore quelconque, est-il possible de modifier ses propriétés intrinsèques pour qu'elle s'anime d'un mouvement et évoque une interaction particulière telle qu'un frottement ou un roulement par exemple ? Pour effectuer ces conformations, on s'appuie alors sur les morphologies sonores précédemment identifiées.

Après avoir présenté un état de l'art des domaines concernés dans le chapitre I, les chapitres II et III se focalisent dans un premier temps sur la synthèse de sons évoquant différents types d'interactions continues (frotter, gratter, rouler) en se basant sur un paradigme action-objet qui stipule qu'on peut décrire les sons comme résultant d'une action (e.g. gratter) sur un objet (e.g. une plaque en bois). Ce paradigme s'adapte naturellement à une approche de la synthèse par modèle source-filtre, où l'information perceptive sur l'objet est contenue dans le "filtre", et l'information sur l'action dans la "source". Un modèle générique basé sur les invariants associés aux actions frotter, gratter, rouler et permettant d'effectuer des transitions perceptives continues entre ces 3 interactions (e.g. transformer progressivement un frottement en un roulement) est présenté. Ce modèle repose sur des propriétés statistiques particulières de micro-impacts entre l'excitateur (e.g. une bille qui roule) et le résonateur (e.g. une plaque en bois). Ces modèles d'interactions sont ensuite validés par des tests perceptifs.

Le chapitre IV traite quant à lui des métaphores sonores. Pour ce faire, on se base sur le principe de la synthèse croisée, en considérant la texture sonore comme résultante du résonateur (« objet ») du modèle, excité par un processus stochastique. La méthode d'estimation (LPC) sur des textures stationnaires (final d'orchestre, phonème, note de clarinette par exemple) est tout d'abord présentée. Les métaphores sonores sont ensuite créées à partir de croisements entre les textures et les invariants associés à plusieurs interactions continues. Enfin sur la base d'une série de tests perceptifs, nous montrerons l'avantage de notre méthode par rapport à une approche de "design sonore" plus classique consistant à mélanger les deux flux sonores. Enfin, on verra que la méthode proposée permet également de conserver les informations pertinentes pour permettre à la fois la reconnaissance de la texture et de l'interaction effectuée.

Chapitre I

Contexte et enjeux de la thèse

Sommaire

A	**Synthèse sonore**	**3**
B	**Contrôle de la synthèse sonore**	**8**
C	**Domaines d'utilisation de la synthèse sonore**	**12**
D	**Paradigme d'étude**	**13**
E	**Enjeux de la thèse**	**21**

Les travaux abordés dans cette thèse font appel à plusieurs domaines. Une grande partie de cette thèse s'intéresse à la synthèse sonore et au *contrôle intuitif* associé. L'idée du contrôle intuitif est de permettre à l'utilisateur (un designer sonore ou un musicien par exemple) de créer des sons et les manipuler de manière cohérente avec sa perception. On portera donc naturellement une attention particulière à la manière dont les sons sont perçus, et ce dès le développement des modèles de synthèse sonore et leur contrôle (Kronland-Martinet *et al.*, 2012). Cette thèse est donc nécessairement liée au domaine du design sonore. D'une part, on proposera des outils de synthèse de sons du quotidien contrôlables à haut-niveau, et en particulier d'interactions continues comme les sons de roulement ou de frottement. D'autre part, on proposera une méthode permettant d'enrichir des textures sonores afin qu'elles évoquent ces interactions particulières, créant ainsi ce qu'on appellera des *métaphores sonores*.

Dans ce chapitre, après avoir rappelé quelques généralités sur les principaux modèles de synthèse sonore, on abordera la notion de contrôle de la synthèse sonore et on fera l'état de l'art sur les différentes approches du contrôle. Quelques applications de la synthèse sonore (jeux vidéos, réhabilitation,...) seront également présentées.

Plus spécifique aux études de cette thèse et du groupe de recherche dans lequel ces travaux ont été développés, on adoptera un paradigme de synthèse qu'on appelle *action-objet*, inspiré de *l'approche écologique de la perception*. Cette théorie sera présentée, ainsi que le modèle de synthèse *source-filtre* qui en découle. On passera également en revue un certain nombre d'études sur la perception des sons du quotidien. Enfin, les enjeux particuliers de cette thèse seront exposés.

A Synthèse sonore

La synthèse sonore est aujourd'hui une discipline établie, trouvant ses racines dans les travaux pionniers de Max Mathews aux laboratoires Bell vers la fin des années 1950 (Mathews et Guttman, 1959; Mathews, 1963) (bien que des ordinateurs aient déjà joué

des sons au début des années 1950 (Hill, 1954), voir (Doornbusch, 2004) pour un résumé historique sur ce point). Les premières synthèses numériques étaient effectuées à partir d'oscillateurs, de formes d'ondes pré-calculées et lues à vitesse variable, ainsi que de filtres. Rapidement, des applications musicales ont vu le jour comme l'arrangement de *Bicycle Built for Two* par Max Mathews en 1961 ou la composition *Computer Suite from Little Boy* de Jean-Claude Risset en 1968.

Actuellement, on peut diviser les méthodes de synthèse en deux grandes classes : les modèles dits de "signaux" ou "abstraits", et les modèles dits "physiques". Bien qu'on puisse établir des liens entre certains modèles de signaux et certains modèles physiques, la philosophie de chacune de ces classes de modèles est très différente. Les modèles physiques s'attachent à décrire, modéliser et reproduire la cause physique à l'origine du son. La phase de modélisation se fait en général indépendamment de toutes considérations perceptives. Si le résultat sonore du modèle ne satisfait pas l'expérimentateur, afin s'approcher d'un son plus "réaliste", ou plus "perceptivement satisfaisant", celui-ci tentera généralement de raffiner la description physique du phénomène. A l'inverse, les modèles de signaux sont souvent dits "abstraits" car ils ne partent généralement pas d'une modélisation de la physique du phénomène. Comme on le verra un peu plus loin, ces modèles partent généralement d'une représentation particulière des signaux (e.g. une somme de sinusoïdes). Le but est ensuite d'ajuster les paramètres pour obtenir le signal désiré. Ces paramètres peuvent être ajustés par un schéma d'analyse/synthèse sur des sons enregistrés où l'expérimentateur tentera de s'approcher au mieux du signal analysé. On peut ensuite juger du résultat de deux manières : soit d'un point de vue purement mathématique en mesurant un ou plusieurs critères (erreur de reconstruction par exemple), soit d'un point de vue perceptif en jugeant de la qualité de la synthèse (par des tests d'écoute par exemple). Une autre manière d'ajuster les paramètres d'un modèle de synthèse abstrait est d'aborder le problème d'un point de vue phénoménologique. En général, approcher le problème de cette manière va de paire avec le choix et la mise au point du modèle de synthèse sonore. L'idée d'une approche phénoménologique est d'essayer de considérer le phénomène produisant le son sous un angle plus perceptif, en se posant la question de ce qui est perceptivement important dans le son à synthétiser pour l'auditeur. Cette approche de la synthèse ne cherche donc pas à reproduire la "réalité physique", mais plutôt "l'effet perceptif".

Une autre différence entre les modèles physiques et de signaux se situe plutôt du point de vue des "possibilités sonores". En effet, la synthèse sonore par modélisation physique ne pourra reproduire "que" la physique (elle permet cependant de créer des sons inédits, voir par exemple les instruments proposés par Bilbao (2009)). Si on a par exemple modélisé parfaitement la physique du roulement et que les sons produits sont très satisfaisants d'un point de vue perceptif, on sera bloqué par la physique elle-même si l'on veut faire "rouler une fourchette en bois, mais qu'elle roule à la manière d'une bille". En effet, les modes propres d'une fourchette étant différemment répartis que ceux d'une sphère, il sera difficile de "tricher" en trouvant une sphère donnée ayant les mêmes modes que la fourchette que l'on désire faire rouler (cependant des travaux ont récemment été menés en ce sens par Villeneuve et Cadoz (2012) qui ont étudié le problème inverse suivant : "Etant donné un son, quel modèle physique peut le produire ?"). En revanche, si l'on a par exemple un modèle de signal qui nous permet de synthétiser des sons de roulement et prenant dans ses paramètres directement les modes propres de l'objet roulant, il est donc aisé de synthétiser ce son (cela ne veut pas pour autant dire que lorsqu'un auditeur entendra ce son il répondra quelque chose comme : "hum... on dirait une bille qui roule, mais elle sonne comme une fourchette" !).

La tâche devient encore plus ardue si l'on souhaite générer des sons sortant encore plus de l'ordinaire par modélisation physique, comme "frotter du feu" ou bien "faire rouler une bille liquide et métallique". Cette tâche est donc plus abordable par des modèles de signaux que des modèles physiques. Dans cette thèse, on utilisera des modèles venant de ces deux grandes classes, et on les décrira donc brièvement dans cette section.

A.1 Modèles de signaux

Il existe énormément de modèles de signaux différents, certains étant des combinaisons de plusieurs de ces modèles. On abordera dans cette section une partie des modèles de signaux les plus courants, et on choisit de séparer les modèles de synthèse sonore linéaires des modèles non-linéaires.

Modèles linéaires Un modèle très répandu est la synthèse additive, décrivant un son comme une somme de N sinusoïdes, c'est-à-dire que le son généré est :

$$s(t) = \sum_{k=1}^{N} \mathrm{A}_k(t) \sin\left(2\pi \int_0^t f_k(\tau)\mathrm{d}\tau + \phi_k\right) \tag{I.1}$$

où A_k et f_k sont respectivement l'amplitude et la fréquence instantanée de la $k^{\text{ième}}$ sinusoïde, ϕ_k est la phase de cette sinusoïde et t le temps. Si on considère les amplitudes et fréquences fixes, alors tout signal s observé sur un intervalle $[0, T]$ peut être décomposé en une somme infinie de sinusoïdes par décomposition en série de Fourier. En pratique, une approche estimant les paramètres par transformée de Fourier sur des trames successives du signal est souvent utilisée (McAulay et Quatieri, 1986; Serra, 1989), permettant ainsi une plus grande flexibilité par exemple pour appliquer des transformations au signal entre les phases d'analyse et de synthèse (comme par exemple l'étirement temporel d'une note d'instrument de musique sans modifier sa hauteur tonale ni son attaque). Serra (1989) considère de plus la partie résiduelle de la décomposition en somme de sinusoïdes comme un processus stochastique, ce qui lui permet de la modéliser comme un bruit blanc filtré. Une autre approche, très prisée des compositeurs entre autres et qu'on peut voir comme une généralisation de la synthèse additive, est la synthèse granulaire (utilisée et formalisée entre autre par Iannis Xenakis, Barry Truax, Curtis Roads...). Le principe est de juxtaposer des courts segments sonores, appelés grains, afin de créer des textures sonores complexes (Roads, 1985). On reviendra sur cette méthode dans le chapitre IV.

Un autre type très répandu de modèle de synthèse sonore est la synthèse soustractive, également appelée source-filtre (voir par exemple (Roads, 1998)). Cette technique permet notamment d'obtenir des spectres très riches à un coût souvent moindre que par la synthèse purement additive, en partant d'un son très riche comme un bruit blanc ou un train d'impulsions et en façonnant son spectre par filtrage :

$$s(t) = [e * h]\,(t) \tag{I.2}$$

où $e(t)$ est le signal source et $h(t)$ la réponse impulsionnelle du filtre. Ces modèles sont particulièrement adaptés à la modélisation de la parole, en considérant que les cordes vocales produisent un train d'impulsion dans le cas d'un signal voisé ou un bruit blanc dans le cas d'un signal non voisé, puis que ce signal vient ensuite être filtré par un filtre représentant les caractéristiques du conduit vocal. Le schéma sur la figure I.1 illustre le principe de cette méthode. Cette modélisation suppose donc un découplage de l'excitateur et du résonateur, alors que même dans le cas des systèmes linéaires, la connexion

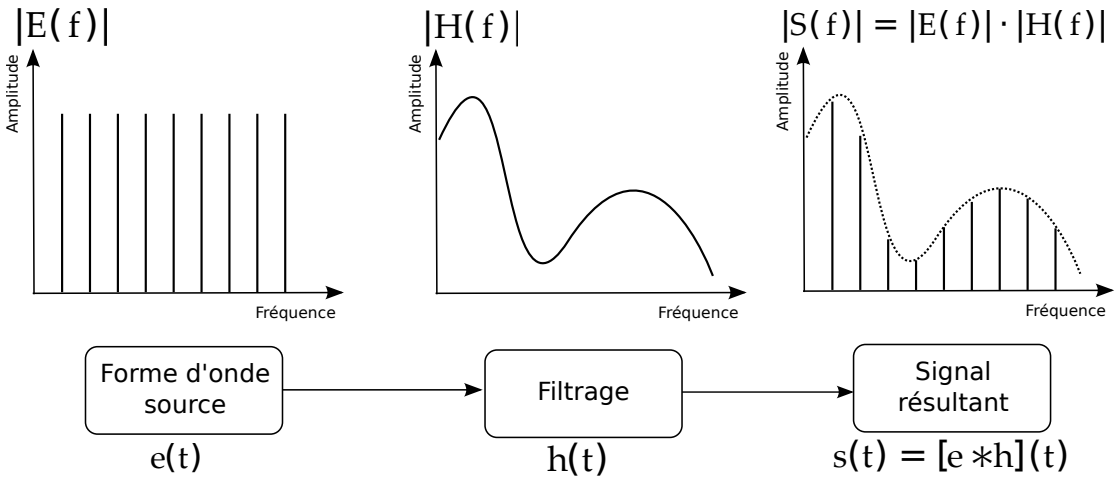

FIGURE I.1 – Principe de la synthèse soustractive.

de deux objets modifie leur comportement (pour être rigoureux, on ne peut normalement pas parler de notions de fréquences individuelles de chaque objet lorsque ceux-ci sont couplés, voir par exemple (Bilbao, 2009)). Cependant on peut parfois supposer que le couplage entre les deux objets est faible, ou perceptivement peu important si on s'intéresse purement à l'effet sonore produit et non à la simulation de la "réalité physique". L'effet perceptif du couplage entre les deux objets peut également être simulé directement dans l'excitation comme on le verra en fin de chapitre III, dans la partie E.1. La synthèse soustractive sera la base des travaux développés dans cette thèse.

Modèles non-linéaires Une méthode non-linéaire classique est la synthèse par modulation de fréquence (Chowning, 1973) qui permet d'obtenir des sons aux spectres très riches en modulant la fréquence d'une sinusoïde par une autre sinusoïde, donc de manière extrêmement peu coûteuse pour un ordinateur :

$$s(t) = A_0 \sin\left(2\pi f_0 t + I \sin\left(2\pi f_1 t\right)\right) \tag{I.3}$$

où f_0 est la fréquence de la porteuse, f_1 est la fréquence de modulation, I l'index de modulation et A_0 l'amplitude. On peut montrer que le spectre de s est alors composé de la porteuse f_0 et de composantes à $f_0 + \alpha f_1$, où $\alpha \in \{..., -2, -1, 0, 1, 2, ...\}$, les amplitudes de chacune de ces composantes dépendant de I. Le problème majeur de ce modèle de synthèse est sa non-prédictibilité : une petite variation des paramètres peut changer radicalement le son résultant. Un autre problème majeur est la mise en place d'une méthode d'analyse permettant la calibration des paramètres du modèle à partir de signaux enregistrés (des études montrent cependant que pour certains types de signaux, il est possible d'estimer les paramètres du modèle, voir par exemple (Justice, 1979; Payne, 1987; Delprat *et al.*, 1990; Horner *et al.*, 1993)).

Dans le même esprit, la synthèse par modulation d'amplitude consiste à moduler l'amplitude d'une sinusoïde de fréquence f_0 par une autre de fréquence f_1 :

$$s(t) = \left(A_0 + A_1 \cos\left(2\pi f_1 t\right)\right) \cos\left(2\pi f_0 t\right) \tag{I.4}$$

et créer ainsi un son ayant un spectre présentant une raie centrale à f_0 et deux raies latérales à $f_0 - f_1$ et $f_0 + f_1$.

A.2 Modèles physiques

Comme expliqué plus haut, la philosophie de la synthèse par modèle physique est radicalement différente des méthodes par modèles de signaux. L'idée de base est de décrire le fonctionnement physique du système étudié, par exemple un instrument de musique. Si on cherche à reproduire le son d'une guitare, on modélisera par des équations différentielles couplées, selon le degré de raffinement voulu, le déplacement de la corde, l'interaction avec le chevalet, la caisse de résonance, l'interaction du doigt avec la corde et sa position, le rayonnement dans l'air... Comparativement aux modèles de signaux, les modèles physiques sont généralement plus coûteux en temps de calcul et ne peuvent pas tous être exécutés en temps réel, en particulier lorsqu'une fréquence d'échantillonnage élevée est désirée, ce qui est important pour la synthèse sonore. De plus, les modèles physiques sont moins flexibles que les modèles de signaux, et généralement une modélisation physique est dédiée à un type de sons particulier (par exemple les sons de guitare), là où par exemple un modèle de synthèse additive peut permettre de synthétiser une très large classe de sons (e.g. aussi bien la guitare que le violon ou la clarinette). Cependant les modèles physiques permettent généralement de synthétiser des sons d'une très grande qualité, comme par exemple les modèles proposés par Bilbao (2009) [1].

Les premières simulations directes des équations aux dérivées partielles modélisant un phénomène physique de production de son pour la synthèse sonore datent de la fin des années 1960 (Ruiz, 1969), et ont depuis considérablement avancé, notamment grâce à l'augmentation constante des capacités de calcul des ordinateurs ainsi qu'à l'amélioration des connaissances. Le principe consiste à discrétiser les dérivées partielles en temps et en espace des équations décrivant le phénomène par des approximations aux différences finies (Ascher et Petzold, 1998). Cette discrétisation permet d'obtenir un schéma numérique donnant accès au déplacement en chaque point (discret) du système étudié à l'instant $t + \Delta_t$ en fonction des déplacements à l'instant t, où Δ_t est le pas d'intégration temporel (i.e. on accède à la solution en tout point toutes les Δ_t secondes).

Dans les modèles physiques, l'approche de la synthèse sonore par guide d'onde a été grandement exploitée, en partie grâce aux travaux de Julius Smith (Smith, 1992, 2014). Cette approche, moins coûteuse en temps de calcul que la simulation directe, est particulièrement efficace lorsque la partie vibrante du phénomène considéré peut être modélisée comme un milieu à une seule dimension (par exemple les cordes, certains instruments à vent). L'idée de base consiste à considérer les solutions de l'équation des ondes à une dimension, qui peut s'écrire comme deux ondes se propageant en sens opposé et n'interagissant pas entre elles, pouvant ainsi être simplement simulées par deux lignes à retard (idée premièrement proposée par Kelly et Lochbaum (1962) pour la synthèse de parole et reprise entre autres par Karplus et Strong (1983) pour la synthèse de sons de cordes pincées). Cette ligne à retard est initialisée par des nombres aléatoires et terminée par un filtre passe-bas. Le coût d'un tel algorithme est très faible : il est du même ordre qu'un oscillateur sinusoïdal obtenu par lecture de table d'onde, mais permet d'obtenir un spectre harmonique plutôt qu'une seule composante. Des extensions à 2 et 3 dimensions pour modéliser par exemple des membranes (Laird, 2001) ou des réponses impulsionnelles de salles (Murphy *et al.*, 2007) ont également été proposées.

Cadoz *et al.* (1984) ont eux proposé des outils de synthèse sonore basés sur la simulation d'éléments mécaniques simples, tel des systèmes masse-ressort, reliés entre eux. La correspondance avec des objets physiques réels (comme une plaque ou un tube) n'est

1. Exemples sonores disponibles ici : `http://www2.ph.ed.ac.uk/~sbilbao/nsstop.html`

pas directe, mais cette méthode de synthèse permet une grande modularité (on peut facilement connecter différents blocs afin d'obtenir des résultats sonores intéressants), et des travaux sur l'interactivité grâce à des contrôleurs externes facilitent l'usage (Leonard *et al.*, 2013).

Enfin, la synthèse modale considère le comportement vibratoire de l'objet comme résultant de la contribution de plusieurs modes oscillants à une seule fréquence et indépendants entre eux (Adrien, 1991). Chaque mode vibratoire est décrit par une équation différentielle du second ordre dont la solution homogène est une sinusoïde exponentiellement amortie, et le déplacement global permettant de remonter au son généré est simplement la somme des vibrations décrites par chaque oscillateur. Différentes conditions initiales ou fonctions d'excitations peuvent être considérées selon le phénomène modélisé. Les travaux de J. M. Adrien sur la synthèse modale ont débouché sur l'environnement graphique Modalys [2] qui permet de connecter des objets physiques simples (cordes, plaques, tubes, etc) et les faire interagir entre eux.

Au-delà des modèles de synthèse sonore en tant que tels, un point important est le contrôle associé à ces modèles. En effet, un modèle de synthèse sonore, en particulier pour un utilisateur peu expérimenté, est compliqué à utiliser pour obtenir le son souhaité, le son "imaginé" par l'utilisateur. Des recherches se penchent donc depuis un certain nombre d'années sur des méthodes permettant de contrôler plus aisément les modèles de synthèse sonore. La section suivante a pour but de détailler l'état de l'art sur les recherches en contrôle des processus de synthèse.

B Contrôle de la synthèse sonore

Dans le domaine de la synthèse sonore, on appelle "contrôle" l'ensemble des possibilités que l'utilisateur du synthétiseur aura pour modifier le son produit. On peut d'emblée distinguer deux grands types de contrôles. Le contrôle "gestuel" s'intéresse à la manière dont le geste de l'utilisateur va agir sur le modèle de synthèse, et le contrôle "intuitif" ou "perceptif" qui étudie comment les paramètres que manipulera l'utilisateur vont affecter ce qu'évoquera le son, en terme de qualité sonore (contrôle du timbre) ou de propriétés de la source sonore. Cette phase est cruciale dans le développement d'un synthétiseur, et ce particulièrement pour les modèles de signaux (les modèles physiques prenant directement en entrée des paramètres "compréhensibles" par l'utilisateur, du moins si celui-ci est familier avec la physique). Prenons l'exemple de la synthèse additive. L'utilisateur peut générer des sons très riches en sommant un grand nombre de sinusoïdes et en modulant leurs fréquences et amplitudes. Cependant, on arrive rapidement à plusieurs centaines de paramètres, et réussir à modifier un à un ces paramètres pour obtenir le résultat sonore désiré demande une grande expertise. Bien que l'approche par "essai-erreur" mène à des résultats sonores intéressants par accident, celle-ci n'est pas toujours souhaitable. Il est donc intéressant de relier des ensembles de paramètres de synthèse "bas-niveau" (e.g. les fréquences et amplitudes des oscillateurs dans le cas de la synthèse additive) à un plus petit nombre de paramètres "haut-niveau" interprétables plus facilement par l'utilisateur. Définir les relations adéquates entre les paramètres bas-niveau et haut-niveau est appelé *mapping* dans le domaine de la synthèse sonore. En tant que verbe, "to map" signifie cartographier, ce qui montre bien l'idée de relation entre un grand nombre de données (provenant de satellites dans le cas d'une carte) difficilement compréhensibles par l'utilisateur

2. `http://forumnet.ircam.fr/fr/product/modalys/`

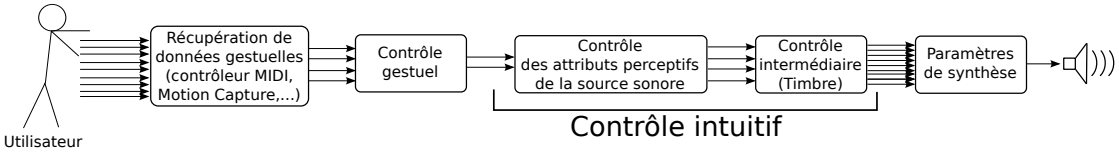

FIGURE I.2 – Représentation des différents niveaux de mapping, depuis le geste de l'utilisateur jusqu'au modèle de synthèse sonore. Les différents niveaux ne sont pas nécessairement tous présents dans un synthétiseur.

et un ensemble moins grand de données (par exemple les contours de la côte maritime sur la carte) interprétables et utilisables plus facilement par l'utilisateur. Cette idée de cartographie illustre également l'idée de différents niveaux de compréhension, d'interprétation et d'expertise requises pour lire différents types de cartes. Si mon but est d'aller faire une randonnée intensive sur le massif de la Sainte-Baume sans emprunter les chemins balisés mais sans vouloir me perdre, j'opterais sûrement pour une carte assez précise représentant les lignes de niveau, les petits chemins, mais qui nécessite une certaine expertise pour l'utiliser. Si en revanche je désire aller en voiture de Marseille jusque la magnifique ville de Morlaix, j'opterais plutôt pour une carte bien moins précise représentant la France et ne montrant pas les détails de la carte précédente en les groupant dans des ensembles plus facilement interprétables comme des autoroutes et des villes. On peut faire l'analogie avec le contrôle de la synthèse sonore, dont on a représenté les différents niveaux sur la figure I.2 (tous les niveaux ne sont pas nécessairement présents dans un modèle de synthèse contrôlable). Le contrôle direct des paramètres "bas-niveau" du synthétiseur (le plus à droite sur la figure) est ce que l'on retrouve dans la majeure partie des synthétiseurs du commerce où en général une longue pratique est nécessaire avant d'arriver à une bonne maîtrise. Un bon exemple de synthétiseur commercial donnant un aperçu des différents niveaux de contrôle (ou tout du moins des différents niveaux de représentation des caractéristiques d'un son) est Massive de chez Native Instruments [3]. Ce synthétiseur permet de contrôler les paramètres de synthèse bas-niveau (filtres, oscillateurs, etc), mais classe également les réglages prédéfinis par des attributs intermédiaires liés au timbre du son (avec une interprétation très libre et subjective de la notion de timbre par les développeurs) avec des adjectifs comme "gras", "chaud" ou "brillant" (bien qu'on ne puisse pas contrôler directement ces attributs, ils sont en lien avec l'étape "contrôle intermédiaire" de la figure I.2), et également par style musical, qu'on peut voir comme une hiérarchie supérieure au timbre (par exemple la musique new-wave va plutôt contenir des sonorités "froides"). D'autres synthétiseurs du commerce s'axent plus sur des contrôles haut-niveau, avec par exemple Chromaphone d'Applied Acoustics Systems [4] qui permet de connecter différents types d'objets entre eux (plaques, tubes,...) et les impacter dans un but musical, ou bien les synthétiseurs de sons d'environnement développés par AudioGaming [5] donnant des outils de contrôle intuitifs pour les designers sonores (où par exemple dans le synthétiseur de sons de pas on peut contrôler des attributs comme le type de chaussure, la démarche, le type de sol,...).

Enfin, les recherches en contrôle gestuel de la synthèse sonore visent à permettre une utilisation des modèles de synthèse de manière plus expressive, grâce à l'étude de nouvelles interfaces pour le contrôle et la recherche de mappings intéressants entre les données issues de capteurs gestuels et les paramètres du modèle de synthèse (voir par

3. `http://www.native-instruments.com/`
4. `https://www.applied-acoustics.com/chromaphone/overview/`
5. `http://www.audiogaming.net/`

exemple (Wanderley et Depalle, 2004) ou le numéro spécial à venir du Computer Music Journal dédié aux avancées récentes dans ce domaine (Wanderley et Malloch, 2014)). Un grand nombre de travaux ont lieu dans ce domaine à tel point qu'une conférence spéciale sur le contrôle gestuel "New Interfaces for Musical Expression" existe depuis une dizaine d'années. Cet aspect ne sera pas abordé dans cette thèse et ne sera donc pas développé davantage dans l'état de l'art. Dans la suite de cette section on détaillera donc brièvement la littérature sur le contrôle du timbre, puis on traitera l'état de l'art sur le contrôle des attributs perceptifs de la source sonore.

B.1 Contrôle du timbre

Wessel (1979) fait partie des premiers à avoir l'idée d'utiliser les résultats obtenus grâce à des études sur le timbre (qui sont relativement nombreuses, voir entre autres Grey (1977); Risset et Mathews (1969)) afin de contrôler la synthèse sonore. Dans le but de contrôler un processus de synthèse additive de sons d'instruments musicaux, David Wessel mit en place une série de tests perceptifs sur des sons d'instruments (de dissemblance et de ressemblance entre sons) dont les résultats ont permis de construire un espace de timbre, mettant en avant deux dimensions principales qui sont la répartition spectrale de l'énergie et la nature de l'attaque. Ainsi, l'auteur suggère de contrôler la synthèse sonore en naviguant dans cet espace de timbre et ainsi passer de manière continue et régulière entre différents sons d'instruments par l'interpolation des paramètres de synthèse dans cet espace. D'autres études ont suivis, notamment Schindler (1984) qui laisse une grande liberté à l'utilisateur en le laissant définir des enveloppes dynamiques des partiels et en proposant des méthodes de réduction de données, ou Desainte-Catherine et Marchand (1999) qui proposent le modèle "Structured Additive Synthesis", défini par quatre paramètres qu'on peut visualiser et modifier : une amplitude globale, une fréquence fondamentale, une "couleur" qui correspond à l'enveloppe spectrale et une "distorsion" qui correspond au fait que les harmoniques ne sont pas exactement les multiples de la fondamentale.

D'autres auteurs ont opté pour des approches d'apprentissage. Ainsi Gounaropoulos et Johnson (2006) proposent tout d'abord un jugement subjectif d'un corpus de sons par des sujets sur un nombre fini de termes utilisés couramment par les musiciens (chaud, brillant, dur,...), puis des méthodes d'apprentissage par réseaux de neurones sont appliquées aux résultats afin de permettre l'ajustement automatique des paramètres d'un modèle de synthèse directement à partir des adjectifs décrivant la qualité sonore (i.e. pour que l'utilisateur puisse directement définir quelque chose comme "je désire un son très chaud et un peu boisé"). Dans le même esprit, Le Groux et Verschure (2008) proposent d'effectuer une méthode d'apprentissage sur une analyse en composantes principales de l'évolution des composantes spectrales afin de resynthétiser des sons en conservant leur timbre mais en modifiant sonie et hauteur tonale ; les auteurs insistent sur le fait que le modèle proposé est assez générique pour contrôler d'autres caractéristiques du timbre comme la brillance. Enfin, Hoffman et Cook (2006, 2007) ont formalisé une généralisation de ces différentes approches dans ce qu'ils appellent "Feature-based synthesis".

B.2 Contrôle des attributs perceptifs de la source sonore

On pourrait supposer que les modèles permettant le meilleur contrôle intuitif de l'évocation qui soit sont les modèles physiques, car ils décrivent directement le phénomène sous-jacent la production du son. Cependant, au-delà du fait qu'ils sont en général plus coûteux en temps de calcul que les modèles de signaux, cette supposition

n'est pas entièrement vraie. Dans le cas d'une plaque par exemple, les paramètres que pourra définir l'utilisateur sont entre autres le module d'Young, le coefficient de Poisson ou la masse volumique, qui sont des paramètres difficilement appréhensibles par un non physicien et qui ne sont pas reliés directement à des évocations, par exemple du matériau. En effet, la réalité physique n'est pas nécessairement la réalité perceptive : il n'est pas dit qu'en imposant les valeurs des coefficients de l'aluminium trouvées dans des abaques dans le modèle physique (de plaque ou de barre par exemple) que le son évoquera nécessairement un objet métallique (de même que le son réellement produit par un objet métallique ne sera pas nécessairement perçu comme tel). De plus, à l'état de nos connaissances actuelles, ces modèles ne permettent pas de générer tous les sons souhaités (par exemple, il semble complexe d'envisager la modélisation physique d'une scène de pluie).

Dans le cas de la synthèse additive de sons d'impacts, Aramaki *et al.* (2009b, 2011) proposent un contrôle intuitif du matériau perçu. La calibration du contrôle haut-niveau s'est basée à la fois sur des tests perceptifs et sur des considérations physiques. Ce synthétiseur a également l'intérêt de permettre des transitions perceptives continues entre les différents matériaux, permettant de créer des sons intermédiaires, ce qui a un intérêt notamment pour étudier le fonctionnement du système auditif (Aramaki *et al.*, 2009a; Micoulaud-Franchi *et al.*, 2011). Ces transitions continues sont plus simples à effectuer sur des modèles de signaux que sur des modèles physiques : en mettant en avant les dimensions permettant de discriminer les matériaux, on peut interpoler les paramètres de synthèse pour passer d'un matériau à l'autre selon ces dimensions (de la même manière que l'espace de timbre proposé par Wessel (1979) pour les instruments de musique). Ce synthétiseur offre de plus l'avantage de contrôler le processus de synthèse à différents niveaux, aussi bien en manipulant des évocations (matériau) que des paramètres intermédiaires (entre l'évocation et les paramètres de synthèse) liés au timbre. Des travaux sont actuellement poursuivis pour étendre ce contrôle à la forme perçue de l'objet impacté (Rakovec *et al.*, 2013). On reviendra sur ce synthétiseur plusieurs fois au cours de cette thèse.

Le projet *The Sounding Object* (Rocchesso et Fontana, 2003), achevé il y a une dizaine d'années mais dont les idées continuent à être développées, a permis une nette avancée dans le domaine du contrôle de la synthèse sonore des sons du quotidien. Ces travaux ont été suivis par le projet *Closed* [6] dont le but était d'approfondir les modèles de synthèse développés dans le cadre de *The Sounding Object*, et d'évaluer la pertinence de ces outils pour le design sonore. Ces projets ont notamment permis le développement du *Sound Design Toolkit* [7] qui propose la synthèse et le contrôle de sons solidiens et de liquide. Dans cet outil de design sonore, les différents modèles sont classés hiérarchiquement (Rath *et al.*, 2003) d'une manière proche de celle proposée par Gaver (1993b) : les "modèles bas-niveau" (impact, friction, bulle) permettent de dériver des événements basiques (roulement, froissement) et des processus dérivés de hiérarchie plus élevée (bris d'objet, rebond, éclaboussement). Ces modèles de base ont par exemple permis de développer des applications interactives de réalité virtuelle où le son produit par différents types de sols peut être synthétisé et renvoyé en temps-réel au sujet pendant que celui-ci marche (Nordahl *et al.*, 2011b,a). Sur le contrôle des sons de pas, Cook (2002) a également proposé une interface de contrôle à haut-niveau (nommée "Bill's Gait") d'un modèle physiquement informé. Basés sur les travaux de Farnell (2010), Verron *et al.* (2010) ont proposé un synthétiseur spatialisé de sons d'environnement comme le feu, la pluie, les bruits de pas, le vent... Ce synthétiseur permet à l'utilisateur de contrô-

6. `http://closed.ircam.fr/`
7. `http://www.soundobject.org/SDT/`

ler intuitivement le son d'une source par des contrôles comme l'intensité ou le taux de ruissellement pour les sons de pluie, ou bien la force ou la froidure pour les sons de vent. Il permet également de contrôler la spatialisation et l'extension des différentes sources sonores. Actuellement, le projet *SkAT-VG* [8] s'attache à lever des verrous sur le contrôle de la synthèse, en tâchant de mettre en place des outils de synthèse proposant à l'utilisateur de pouvoir ébaucher des sons en utilisant des imitations vocales et gestuelles, de la même manière qu'on ébauche généralement une idée de façon très naturelle en dessinant. Le projet *MétaSon* dans lequel s'inscrivent ces travaux de thèse s'intéresse également au contrôle des attributs perceptifs des sources sonores et à la possibilité de créer des sons inédits appelés "métaphores sonores" (ce concept sera décrit dans la fin de ce chapitre ainsi que dans le chapitre IV). Une partie de ce projet se penche également sur la modélisation mathématique et l'estimation de la dynamique des sons, qui sont abordés comme des translations et dilatations dans le plan temps-fréquence. Ces travaux permettent par exemple d'estimer la dynamique du son produit par une voiture qui accélère, de lui enlever cette dynamique, lui appliquer une autre dynamique... (Omer et Torrésani, 2013)

Cette "dynamique" du son nous amène enfin aux études qui se sont intéressées à l'évocation du mouvement dans la synthèse sonore monophonique. Merer *et al.* (2013) se sont attachés aux caractéristiques acoustiques responsables de l'évocation d'un mouvement dans un son monophonique. Pour cela, des sons abstraits (sons dont on ne peut remonter à la source l'ayant produit) ont été utilisés afin de minimiser la médiation à des références cognitives ou culturelles, et par conséquent se focaliser sur les attributs propres au son. Des tests de catégorisation et de caractérisation graphique des sons (i.e. les sujets devaient dessiner le mouvement évoqué par le son), couplés à une analyse des signaux ont permis de mettre en avant certaines propriétés du son qui évoquent différents types de mouvement (chuter, tourner,...). Basée sur les descripteurs mis en avant par les tests et l'analyse des signaux, une interface de contrôle du mouvement évoqué par la synthèse sonore a été développée et validée par un test perceptif. Thoret *et al.* (2014); Thoret (2014) ont eux proposé un contrôle évoquant la fluidité du geste produit pour un modèle de synthèse de sons de friction (Van Den Doel *et al.*, 2001). Les auteurs ont pour cela pris en compte dans leur mapping des contraintes bio-mécaniques du geste humain, i.e. la loi en "1/3" reliant courbure et vitesse du geste (Lacquaniti *et al.*, 1983).

La synthèse sonore, et en particulier lorsque l'on dispose de modèles contrôlables comme on l'a vu ici, est un outil parfaitement adapté à la transmission d'informations. La section suivante présente quelques études ayant utilisé la synthèse sonore pour diverses applications.

C Domaines d'utilisation de la synthèse sonore

Au-delà des applications musicales évidentes qui furent les premières utilisations de la synthèse sonore, celle-ci est d'un grand intérêt pour plusieurs domaines. D'un point de vue fondamental, la synthèse sonore est d'intérêt pour étudier la perception auditive. Aramaki *et al.* (2009a) ont par exemple étudié l'activité cérébrale lors de l'écoute de continua entre matériaux sur des sons d'impacts générés par un modèle de synthèse contrôlable à haut-niveau. Ces mêmes sons ont été utilisés pour mettre en évidence les différences perceptives entre schizophrènes et non-schizophrènes (Micoulaud-

8. `http://skatvg.iuav.it/`

Franchi *et al.*, 2011). McDermott et Simoncelli (2011); McDermott *et al.* (2013) ont quant à eux étudié la perception des textures sonores, et en particulier la manière dont elles sont probablement codées au niveau cérébral grâce à un modèle de synthèse inspiré par la physiologie du système auditif. Geffen *et al.* (2011) ont eux étudié la manière dont sont probablement codés au niveau cérébral les sons de liquides en utilisant un modèle de synthèse proposé par van den Doel (2004).

La synthèse sonore peut également trouver des applications importantes dans le design sonore. En effet, Susini *et al.* (2014) proposent la définition suivante du design sonore :

> *A "sound design" approach is implemented in order to create "new" sounds with the intention that they will be heard in a given context of use. By new sounds, we mean sounds that cannot be found in existing sound databases, or cannot be recorded or, at least, cannot be directly used without being modified.*

La synthèse sonore, et en particulier si elle est contrôlable intuitivement, peut permettre de créer ces "sons nouveaux", notamment pour des applications en "design des interactions sonores" (*Sonic Interaction Design*), où l'idée est "d'explorer les moyens par lesquels le son peut être utilisé pour transmettre des informations, du sens, ainsi que des qualités esthétiques et émotionnelles dans des contextes interactifs" (Franinović et Serafin, 2013), notamment avec des objets du quotidien "augmentés" d'un retour sonore interactif. La synthèse sonore peut également être intéressante d'un point de vue industriel, pour la sonification des voitures électriques par exemple. Le mapping des paramètres de synthèse avec les paramètres issus du contrôle du véhicule (vitesse, accélération, etc) peut être mis en oeuvre sur la base d'études en simulateur de l'influence du bruit moteur sur la conduite (Denjean *et al.*, 2012, 2013). Un autre exemple de design sonore en situation interactive vient du monde du jeu vidéo, qui a pendant de nombreuses années effectué ses bruitages à partir de sons pré-enregistrés, et s'intéresse de plus en plus à la synthèse sonore, souvent appelée "procedural audio" dans ce domaine (Böttcher, 2013). En effet, la seule utilisation de banques de sons lors du design sonore d'un jeu nécessite de prévoir chaque scène à l'avance. Cependant, l'utilisation d'un moteur physique de synthèse sonore n'est pas toujours exploitable en temps réel et nécessite souvent des pré-calculs (Zheng et James, 2010, 2011). Afin de combler le fossé entre l'utilisation d'un moteur de synthèse physique et l'approche classique, Picard *et al.* (2009) ont par exemple proposé une méthode de synthèse granulaire. Basé sur un synthétiseur de sons d'environnements spatialisés (Verron, 2010; Verron *et al.*, 2010), Verron et Drettakis (2012) ont quant à eux proposé tout un moteur de synthèse directement relié à au moteur graphique du jeu vidéo.

Enfin, la synthèse sonore trouve également des applications pour la réhabilitation motrice. Rodger *et al.* (2014) ont étudié l'effet de l'apport de sons de pas synthétiques pour aider des patients atteint de la maladie de Parkinson à améliorer leur démarche. Les effets positifs montrés encouragent à envisager davantage la synthèse sonore dans des processus de rééducation. Danna *et al.* (2013) ont quant à eux montré que l'ajout de sons de friction dans une tâche d'écriture peut aider à la réhabilitation de la dysgraphie.

D Paradigme d'étude

Les travaux de cette thèse ont été effectués selon un paradigme d'étude qui a été proposé pour mettre en oeuvre un contrôle intuitif par évocation (cf section B.2). Cette idée de contrôle intuitif de la synthèse sonore est intimement liée à la façon dont nous percevons les sons et dont ils nous informent. Initialement proposé par Gaver (1993b,a),

un son du quotidien peut être décrit naturellement par une action effectuée sur un objet. Il propose également de considérer séparément l'objet produisant le son de l'action effectuée sur cet objet et ayant causée le son, et ainsi considérer l'action et l'objet comme indépendants. Il justifie cette séparation par le fait qu'un même objet puisse être frappé, écrasé ou gratté ; inversement, il est possible de faire rouler différents objets. Cette description des sons en terme d'actions est confortée par les travaux de VanDerveer (1979) qui montrent que les auditeurs décrivent spontanément les actions ayant produit ces sons. De plus, Lemaitre et Heller (2013) ont récemment montré que la description d'un son en terme d'interaction spécifique (e.g. gratter) était plus rapide et plus précise qu'une description à d'autres niveaux comme la catégorie générique d'interaction englobant les interactions spécifiques (e.g. friction ou déformation) ou la manière dont a été effectuée l'action (e.g. frotter rapidement). Cette description du son en terme "d'interaction" et "d'objet" est pertinente d'un point de vue perceptif mais elle peut l'être aussi d'un point de vue physique dans certains cas : un physicien qui doit par exemple décrire la manière dont est produit un son d'interaction entre un petit objet et une surface (un roulement ou une friction par exemple) décrira sûrement d'une part les propriétés mécaniques de la surface sur laquelle interagit le petit objet et d'autre part les équations du mouvement menant à ce son, et éventuellement le couplage entre les deux corps interagissant. Ainsi, le paradigme d'étude proposé pour le contrôle haut-niveau se basera sur l'approche écologique de la perception qui suppose l'existence *d'invariants* perceptifs indépendants liés à l'action ou à l'objet. Dans la section qui suit, on décrira l'approche écologique de la perception qui suggère ce paradigme *action-objet*, puis on passera en revue les études qui se sont intéressées à la manière dont le son nous informe, mettant en lumière certains de ces invariants perceptifs. Enfin, on décrira le paradigme *action-objet* et le modèle de synthèse adopté.

D.1 Approche écologique de la perception

L'approche écologique de la perception est due à Gibson (1966, 1979) [9]. Afin d'expliquer cette théorie, également appelée *perception directe*, il est tout d'abord important de décrire brièvement ce que l'on appelle *perception indirecte* à laquelle s'oppose la théorie de Gibson. La perception indirecte est généralement associée à la théorie de l'information appauvrie. Chaque jour, l'être humain est confronté à un grand nombre de stimuli (visuels, sonores, tactiles, olfactifs...), et certains stimuli, ou combinaisons de stimuli aboutissent à une perception extrêmement riche par rapport à l'information reçue. C'est typiquement le cas de la photographie, qui est une représentation de l'espace à deux dimensions et dont nous allons systématiquement reconstituer la dimension manquante. Pour la perception des stimuli dynamiques, cette théorie prend en compte que les stimuli sont reçus sous la forme d'une succession d'échantillons, éventuellement déformés par les capteurs sensoriels (déformations dues à la rétine pour l'image, distorsion du son par le filtrage cochléaire,...). Cette suite d'échantillons subira ensuite une succession d'opérations afin de recréer du sens pour devenir finalement un événement perçu (Michaels et Carello, 1981).

L'approche écologique s'oppose radicalement à l'approche de l'information appauvrie, car elle rejette toutes les études "de laboratoire" où les stimuli sont des reproductions artificielles de stimuli naturels (d'où la dénomination "écologique", l'écologie étant la science qui étudie les êtres vivants dans leur milieu et les interactions entre eux). Gibson a proposé sa théorie pour la vision, et propose que la perception est contrainte

9. Voir le livre de Michaels et Carello (1981) pour une introduction plus accessible à la théorie de Gibson

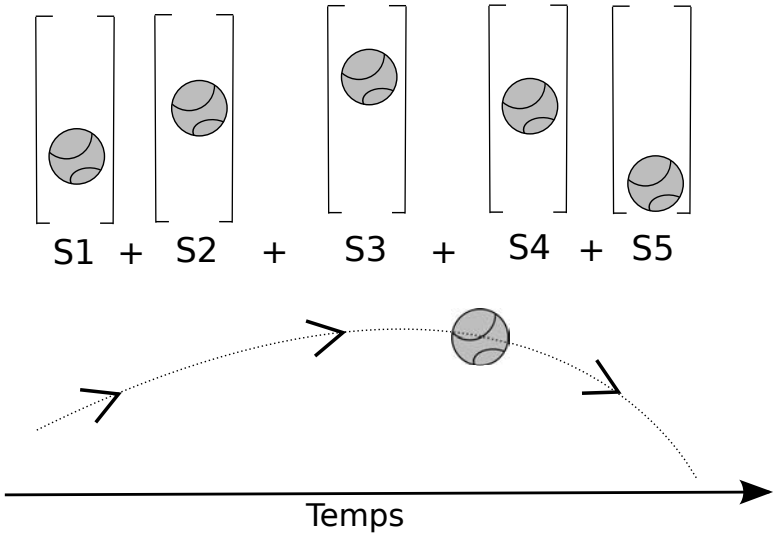

FIGURE I.3 – Caricature de la différence entre approche traditionnelle (haut) et écologique (bas) de la perception. L'approche traditionnelle considère que l'événement est décomposé en une suite d'instants, chacun étant décrit par son propre stimulus S_i ; l'événement est perçu en rassemblant ensemble les stimuli et en reconstituant la dynamique. Dans l'approche écologique, l'information est à travers le temps et ainsi coexiste avec l'événement. La tâche de celui qui perçoit est de détecter l'événement spécifié par l'information (d'après Michaels et Carello (1981)).

par notre interaction avec l'environnement. Il propose ainsi de redéfinir totalement la notion de stimulus, en général décrit par des données physiques primaires comme la fréquence et le niveau pour un son, qui ignore selon lui la véritable information contenue dans le stimulus. En particulier, il rejette la notion d'échantillonnage vue précédemment et propose que l'information doit être détectée dans son ensemble et en particulier en intégrant le temps, et le stimulus est ainsi rattaché à un événement particulier. Cette opposition entre perception directe et indirecte, dans le cas de la vision, peut être caricaturée par la figure I.3. Ainsi Gibson met de côté les adeptes du traitement de l'information : percevoir (ou même décrire) ne revient pas à décrire cette succession d'événements mais détecter l'information dans son ensemble, un genre de motif global, une morphologie.

Au-delà de la validité psychologique de l'une ou l'autre de ces approches, c'est cette description de l'information comme un tout, comme un motif global, qui nous semble particulièrement intéressante. En effet, Gibson propose que la reconnaissance des événements visuels est possible grâce à des structures invariantes contenues dans le flux sensoriel. Concernant la perception auditive, cette approche a été exploitée en premier par Warren et Verbrugge (1984), et formalisée plus tard par McAdams (1993). Cette approche suppose donc l'existence de *structures invariantes* qui portent l'information nécessaire à la reconnaissance des événements sonores. Ces supposés invariants sont divisés en deux catégories : les *invariants structurels* qui mènent à la reconnaissance des propriétés physiques de l'objet sonore (par exemple son matériau, sa forme, etc) et les *invariants transformationnels* qui décrivent le type de changement ou l'action effectuée sur l'objet. Ainsi d'après Michaels et Carello (1981) : "if an event is something happening to a thing, the *something happening* is presumed to be specified by *transformational invariants* while the *thing* that it is happening to is presumed to be described by *structural invariants*". Par exemple, une corde vibrante produit un spectre particulier

qui permet à l'auditeur de la reconnaître, qu'elle soit frottée (e.g. violon), pincée (e.g. guitare) ou frappée (e.g. piano). De la même manière, il est possible d'entendre si un cylindre rebondit, frotte ou roule quelque soit son matériau (Lemaitre et Heller, 2012).

D.2 Invariants perceptifs : comment le son nous informe ?

Un grand nombre d'études traitent de la façon dont nous percevons les sons "réels" (i.e. on exclut ici les études psychoacoustiques utilisant souvent des tons purs, complexes harmoniques ou bruits large bande qui n'ont pas pour but d'évoquer des concepts "haut-niveau"), et en particulier aux informations auxquelles un auditeur peut remonter grâce à ces sons. Bien qu'elles ne soient pas nécessairement placées dans le contexte de l'approche écologique de la perception, ces études sont d'intérêt pour nos travaux. On traitera majoritairement dans cette partie la perception des sons du quotidien autres que musicaux et vocaux.

Gaver (1993b) a proposé une taxonomie empirique des événements sonores du quotidien : ceux-ci peuvent être séparés en trois grandes classes représentant "l'état de la matière", i.e. les sons solides, liquides et gazeux. Chacune de ces catégories comprend des événements basiques (e.g. impact ou friction pour les solides, explosion ou vent pour les gaz, goutte ou éclaboussement pour les liquides), desquels découlent des événements plus complexes (e.g. les vagues découlent des éclaboussements, les bris d'objets des impacts,...), et les trois grandes catégories se recoupent pour former les événements hybrides (e.g. l'explosion d'une bouteille pleine de liquide). Bien que Gaver n'ait pas validé la taxonomie qu'il propose, les études de Lemaitre *et al.* (2010) et Houix *et al.* (2012) ont permis de confirmer partiellement cette classification, dévoilant une catégorie supplémentaire par rapport aux trois états de la matière qui est les sons provenant de machines. Cette taxonomie a de plus été confirmée par l'étude de la classification d'imitation vocales de sons du quotidien qui est similaire à la classification des sons originaux (Lemaitre *et al.*, 2011). L'imitation vocale a par ailleurs été montrée comme intéressante pour décrire les sons, notamment lorsque ceux-ci ne sont pas identifiables et facilement descriptibles par des verbes (Lemaitre et Rocchesso, 2014), et également pour décrire et caractériser les sons d'accélérations de voitures (i.e. si le son évoque une voiture sportive, une berline etc) (Sciabica, 2011).

Une grande partie des études de la littérature sur la capacité auditive à remonter aux propriétés mécaniques des sources ayant produit le son s'est focalisée sur les sons produits par des solides, et en particulier la capacité à reconnaître le matériau d'un objet impacté. Les premiers travaux sont dus à Gaver (1988) et ont montré la capacité à reconnaître les sons de barres en métal des sons de barres en bois impactées et de différentes tailles, avec des résultats similaires pour des sons enregistrés ou de synthèse. D'autres études ont confirmé la capacité à discriminer les matériaux d'objets impactés entre les catégories grossières (e.g. bois et plastique qui sont peu résonants par rapport à verre et métal qui sont plus résonants) mais ont pointé des confusions à l'intérieur de ces catégories grossières (e.g. bois confondu avec plastique), voir entre autres (Lutfi et Oh, 1997; Avanzini et Rocchesso, 2001a; Tucker et Brown, 2002; Giordano et McAdams, 2006). En particulier, il a été montré que les sons d'impacts contiennent suffisamment d'information pour identifier le matériau (Wildes et Richards, 1988), et que la perception du matériau est principalement reliée à l'amortissement en fonction de la fréquence des différentes composantes spectrales (Tucker et Brown, 2002; Klatzky *et al.*, 2000; Giordano et McAdams, 2006) et à la rugosité (Aramaki *et al.*, 2009a). Un point intéressant dans l'étude de Giordano et McAdams (2006) est qu'elle montre également une dépendance entre le matériau perçu et la fréquence fondamentale du son.

En particulier, les sons de petits objets métalliques sont associés à des objets en verre. Une explication possible, et cohérente avec l'hypothèse de Gibson qui est que notre perception est contrainte par notre interaction avec l'environnement, est que ces sons sont acoustiquement proches or nous sommes plus habitués à entendre des impacts sur des petits verres que sur des petites plaques en métal, ainsi notre système perceptif ne se base plus sur les indices d'amortissements. En effet, même si l'on aime beaucoup le vin, on ne trinque jamais avec des verres de deux mètres de diamètre mais avec des verres de taille classique (mais beaucoup de fois dans le cas d'un amateur de vin), et on infère le matériau du fait de cette régularité statistique dans l'environnement acoustique du quotidien (e.g. trinquer avec des verres vs frapper des casseroles en métal, qui font un son plus grave que les verres). Récemment, ce type de régularités statistiques dans l'environnement acoustique quotidien qui contraignent notre perception a été étudié par Parise *et al.* (2014), en particulier sur la connotation spatiale associée à la fréquence des sons : un son est haut ou bas, les mélodies montent ou descendent, etc. Leur étude montre que, statistiquement, les sons aigus sont à une élévation plus haute que les graves dans notre environnement quotidien. Il a également été montré que l'on pouvait remonter à la taille des objets impactés par le son qu'ils produisent (Lakatos *et al.*, 1997; Carello *et al.*, 1998; Grassi, 2005), et dans une certaine mesure à leur forme (Lakatos *et al.*, 1997; Kunkler-Peck et Turvey, 2000; Rakovec *et al.*, 2013). La capacité auditive à percevoir la dureté des matériaux impactés a également été étudiée par Freed (1990) et Giordano et Petrini (2003), qui montrent entre autres que les auditeurs sont capables de déterminer si l'objet a été impacté par un excitateur plutôt mou ou dur. Le nombre d'études sur la perception des actions, et donc sur les *invariants transformationnels*, est moins important. Warren et Verbrugge (1984) ont montré qu'il était possible de prédire à partir du rythme d'une série d'impact si un verre rebondit ou se brise. En effet, les rebonds présentent une régularité temporelle qui n'existe pas lorsqu'un objet se brise. Ils ont ainsi montré qu'en présentant à des auditeurs la superposition de 4 rebonds différents et désynchronisés entre eux, alors ceux-ci percevaient un bris de verre. Stoelinga (2007) et Houben *et al.* (2004) ont quant à eux montré la capacité auditive à percevoir la vitesse de billes roulantes. Enfin Lemaitre et Heller (2012) ont montré que la perception de l'action effectuée par un cylindre (impact, rebond, roulement ou frottement) est très robuste quelque soit son matériau (verre, plastique, métal ou bois).

Un autre exemple d'invariant, qu'on peut considérer comme structurel, nous vient de la parole. En effet, il a été montré que la capacité à reconnaître les différentes voyelles est fortement reliée à la fréquence des deux premiers formants (Peterson et Barney, 1952), ce qui permet ainsi de comprendre différents locuteurs, et de comprendre ce que dit une personne qu'elle ait la voix claire (parole "voisée") ou enrouée (parole "non-voisée"). Ces fréquences de formants sont tellement caractéristiques de la parole que Remez *et al.* (1981) ont montré qu'en créant un signal composé de trois sinusoïdes qui suivent les fréquences centrales des trois premiers formants, la parole est toujours compréhensible.

D'autres auteurs ont étudié la capacité auditive à remonter à d'autres types d'informations. Repp (1987) a par exemple étudié la perception des applaudissements. Pour ce faire, il a enregistré les applaudissements de plusieurs sujets, ceux-ci se connaissant tous. Chaque sujet écoutait ensuite les applaudissements et devait retrouver qui applaudissait. Les performances ont été relativement médiocres. Cependant, des analyses plus poussées ont permis de montrer une cohérence dans le jugement du sexe de l'applaudisseur, bien que ce jugement ne reflète pas nécessairement du sexe réel de l'applaudisseur : les applaudissements associés à des personnes de sexe féminin sont les sons aigus et de rythme rapide, et ceux associés à des applaudisseurs de sexe mas-

culin sont généralement les sons plus graves et avec un rythme plus lent. On peut donc constater par cette étude que la perception peut être biaisée par certains stéréotypes culturels. De même, Li *et al.* (1991) ont montré la capacité auditive à distinguer le sexe d'une personne via les sons de pas. Cabe et Pittenger (2000) ont quant à eux montré la capacité des auditeurs à prédire le temps nécessaire pour remplir de liquide un récipient uniquement grâce au son produit. Toujours sur les sons de liquides, Velasco *et al.* (2013) ont montré que l'écoute de sons de remplissage d'un liquide dans un verre permet d'estimer si le liquide est froid ou chaud. D'un point de vue plus neuroscientifique, Geffen *et al.* (2011) ont étudié la manière dont sont probablement codés les sons de liquides au niveau du système auditif. Comme beaucoup de sons naturels (Voss et Clarke, 1975; Attias et Schreiner, 1997), les sons de liquides présentent une structure invariante d'échelle. Ainsi, Geffen *et al.* (2011) ont montré que les enregistrements de sons d'eau présentent un spectre de puissance en $1/f$, où f est la fréquence, et sont perçus comme naturels et évoquant toujours des sons d'eau quelque soit la vitesse de lecture du son (ils sont donc invariants d'échelle). De plus, des sons d'eau synthétiques constitués d'une somme de sinus glissants (un sinus glissant bref et bien calibré reproduit le son d'une goutte, cf par exemple (van den Doel, 2004)) dont la distribution est contrôlée, sont perçus comme naturels et ressemblant à de l'eau seulement s'ils respectent une structure invariante d'échelle.

On voit donc à travers ces études que l'audition est une modalité sensorielle permettant de remonter à un grand nombre d'informations. La manière dont nous percevons les sons est importante car elle est reliée à la façon dont nous interagissons avec le monde. En effet, les informations auxquelles l'auditeur a accès grâce à l'audition lui donnent les moyens d'inférer certaines propriétés sur les sources ayant produit ces sons, et ainsi d'agir en conséquence. Castiello *et al.* (2010) et Sedda *et al.* (2011) ont par exemple montré que les sons sont utilisés dans la planification des gestes de préhension d'objets. En effet, les sons évoquant des actions activent des aires motrices du cerveau tandis que d'autres types de sons ne les activent pas (Pizzamiglio *et al.*, 2005). Les informations auditives sont tellement importantes qu'un joueur de tennis expérimenté voit ses performances baisser s'il est privé du son (Takeuchi, 1993), et que la modification des sons produits par l'interaction avec un objet perturbe l'expérience haptique (Zampini *et al.*, 2003; Zampini et Spence, 2004; Spence et Zampini, 2006).

La synthèse sonore, et en particulier lorsque l'on dispose de modèles contrôlables, est donc un outil parfaitement adapté à la transmission d'informations, comme on l'a vu dans la partie C. Dans la partie suivante, on proposera un paradigme général pour la stratégie de contrôle, fondé sur la description du son en invariants structurels et transformationnels. Cette stratégie de contrôle conduira à un modèle de synthèse générique qui sera utilisé tout au long de cette thèse.

D.3 Paradigme *action-objet*

Sur la base de ces catégories d'invariants perceptifs, les invariants transformationnels et structurels, les stratégies de contrôle intuitif par évocation suivent un schéma commun, le paradigme *action-objet*. Un schéma conceptuel idéal de ce paradigme est présenté sur la partie supérieure de la figure I.4 (le modèle de synthèse sur la partie inférieure est décrit dans la section suivante). Ainsi, on aimerait idéalement pouvoir combiner différentes actions à différents objets, et créer éventuellement des combinaisons inédites comme rouler sur de l'eau ou bien souffler sur une plaque. L'idée est donc de ne pas devoir décrire explicitement le couplage physique entre l'action et l'objet. A

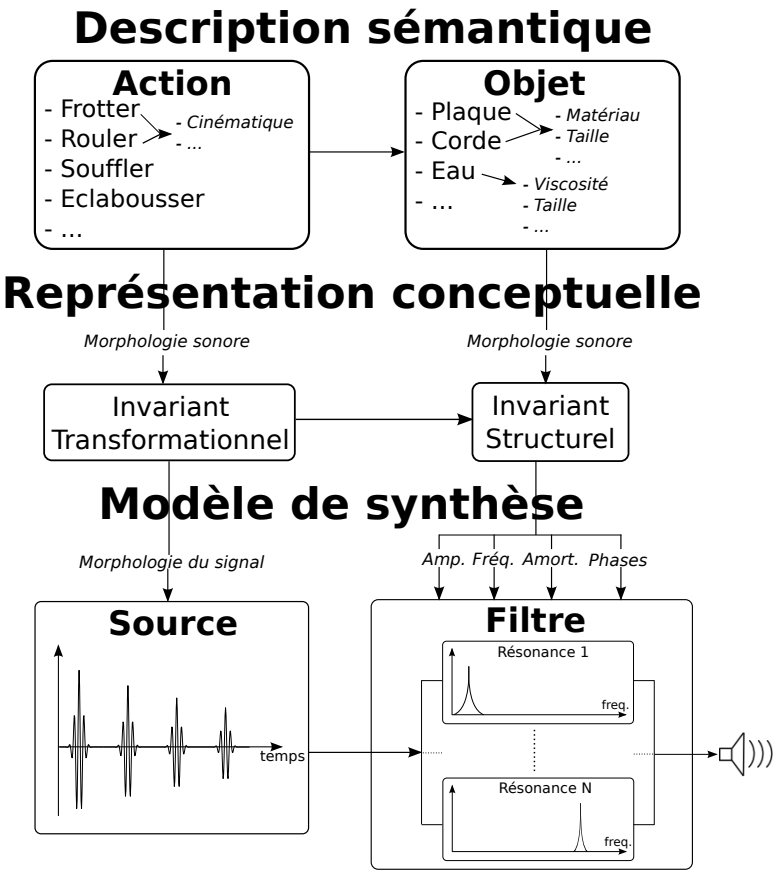

FIGURE I.4 – Paradigme *action-objet* et modèle source-filtre associé.

chaque action ou objet est associée une morphologie sonore particulière (l'invariant structurel ou transformationnel) qui le caractérise, et les sons sont générés directement à partir d'une description sémantique de la source sonore. Le choix et le nombre de paramètres de contrôles haut-niveau proposés (labels accessibles) est un aspect propre à la finalité d'un tel outil. Ces labels émanent généralement des catégories perceptives mises en évidence par des études perceptives. Le contrôle associé nécessitera de définir un mapping avec la méthode de synthèse choisie, décrite dans la section suivante.

D.4 Modèle de synthèse adopté

Un modèle de synthèse bien adapté à la séparation des caractéristiques de l'objet d'une part et de l'action d'autre part est le modèle source-filtre. On se place dans le cas des sons solidiens, qui sont les sons majoritairement abordés dans cette thèse. On choisit de représenter les caractéristiques de l'objet, i.e. les invariants structurels, dans la partie filtre du modèle. Pour un objet donné, ces caractéristiques ne changent pas quelque soit l'action effectuée. Les caractéristiques de l'action, i.e. les invariants transformationnels, seront contenues dans la partie source. Ainsi, sur un objet donné, on peut effectuer différentes actions et inversement, le son résultant étant une simple convolution entre le signal source et la réponse impulsionnelle du filtre. Le modèle source-filtre ne simule donc pas de phénomène de couplage entre la source et le filtre. Il est cependant important de noter que le modèle source-filtre ne limite pas les possibilités de synthèse sonore à la simulation des phénomènes linéaires. Effectivement, les effets non-linéaires perceptivement importants peuvent être pris en compte dans le

terme source. Cela a par exemple été proposé par Bensa *et al.* (2004) pour la synthèse de sons de piano ou pour la synthèse sonore de flûte traversière par Ystad et Voinier (2001), la synthèse de sons d'instruments générés par la friction non-linéaire comme la scie musicale (Serafin *et al.*, 2002), mais également pour la synthèse de sons de friction non-linéaire comme les sons de couinements d'un doigt sur de la vaisselle humide (Thoret *et al.*, 2013).

Afin d'implémenter la partie objet du modèle en synthèse soustractive, chaque composante de la réponse impulsionnelle de l'objet considéré est modélisée par un filtre passe-bande résonant. En pratique, on utilise les filtres proposés par Mathews et Smith (2003), qui sont obtenues par deux oscillateurs couplés :

$$\begin{cases} x(n+1) = x_1 x(n) - y_1 y(n) + u(n) \\ y(n+1) = y_1 x(n) + x_1 y(n) \end{cases} \tag{I.5}$$

avec

$$\begin{cases} x_1 = e^{-\frac{1}{\tau f_s}} \cos\left(2\pi \frac{f}{f_s}\right) = R\cos(\theta) \\ y_1 = e^{-\frac{1}{\tau f_s}} \sin\left(2\pi \frac{f}{f_s}\right) = R\sin(\theta) \end{cases} \tag{I.6}$$

où f est la fréquence de résonance du filtre, τ le temps en secondes pour décroître de $1/e$, f_s la fréquence d'échantillonnage et $u(n)$ est le signal en entrée du filtre. Ainsi si $u(n)=\delta(n)$, alors $x(n)$ est un cosinus exponentiellement amorti et $y(n)$ un sinus exponentiellement amorti, et en combinant $x(n)$ et $y(n)$ on peut contrôler la phase de la sinusoïde générée. La fonction de transfert en z du filtre est alors (pour la sinusoïde) :

$$H(z) = \frac{Y(z)}{U(z)} = \frac{R\sin(\theta)z^{-2}}{1 - 2R\cos(\theta)z^{-1} + R^2 z^{-2}} \tag{I.7}$$

Le filtre est normalisé en le multipliant par $(1-R^2)/R$ afin d'avoir un gain à la résonance identique quelque soit la fréquence et la résonance (approximation vraie pour les filtres "très résonants", i.e. quand $R \to 1$, voir la démonstration dans (Parker et Bovermann, 2013) par exemple). Enfin, afin de se rapprocher du vrai cas des sinusoïdes amorties, dont la fonction de transfert est une fonction Lorentzienne qui a pour module approximativement $\tau/2$ aux alentours de la fréquence de résonance, le filtre est également normalisé par cette valeur. Ainsi, la réponse impulsionnelle de l'objet résonant est implémentée comme un banc de filtres résonants en parallèle, et on peut directement ajuster les paramètres grâce à une méthode d'analyse type ESPRIT (Roy et Kailath, 1989; Badeau *et al.*, 2002) par exemple. On voit donc que grâce à cette approche, n'importe quel signal source peut être utilisé en entrée du banc de filtres résonants. A noter qu'un algorithme similaire a déjà été proposé par van den Doel et Pai (2003). Un autre modèle proche et très utilisé en synthèse musicale est celui des "Filtres d'Ondes Formantiques" (FOFs) développé par Rodet *et al.* (1984) et qui, plutôt que de considérer simplement des exponentielles amorties propose de multiplier celles-ci par un arche de cosinus.

La méthode de synthèse soustractive permet une généralisation des travaux précédents sur les sons d'impacts, et notamment les travaux sur leur contrôle intuitif développés par (Aramaki et Kronland-Martinet, 2006; Aramaki *et al.*, 2009b, 2011), qui proposaient une méthode de synthèse additive. En synthèse additive, les fréquences des oscillateurs correspondent aux fréquences propres de l'objet résonant. Dans le schéma de synthèse proposé par les auteurs, du bruit peut être ajouté à la sortie des oscillateurs afin de simuler la partie stochastique de l'impact. Ensuite, le signal sinusoïdes+bruit est filtré en différentes bandes fréquentielles selon l'échelle de Bark (Zwicker et Fastl,

1990), puis chacune de ces bandes se voit appliquer une enveloppe temporelle d'amplitude afin de prendre en compte la dépendance fréquentielle de l'amortissement. Ce processus de synthèse est intéressant pour simuler des impacts seuls, mais n'est pas adapté à des interactions plus complexes avec l'objet. Par exemple, pour synthétiser le rebond d'un objet, le signal sinusoïdes+bruit doit être déclenché de manière précise, tout comme les enveloppes variant dans le temps des bancs de filtres. Même si cela est encore envisageable, on comprend que cela devient très complexe et coûteux pour des interactions continues plus complexes comme le roulement ou le frottement.

Le modèle de synthèse source-filtre rattaché à ce paradigme action-objet est présenté sur le bas de la figure I.4. Un des intérêts majeurs du modèle proposé par (Aramaki et Kronland-Martinet, 2006; Aramaki *et al.*, 2009b, 2011), outre le contrôle intuitif du matériau perçu, est la possibilité d'effectuer des transitions continues entre différents matériaux, permettant ainsi de créer des perceptions ambigües qui peuvent par exemple être intéressantes pour étudier le fonctionnement du système auditif, mais également intéressantes en terme de design sonore. Dans cette thèse, on s'intéressera particulièrement aux transitions perceptives continues entre différentes actions et donc spécifiquement à la modélisation de termes sources permettant l'évocation de différentes actions ainsi qu'au contrôle intuitif associé au modèle de synthèse. On s'intéressera également à la partie "objet" dans le dernier chapitre. Dans la section qui suit, on présentera spécifiquement le cadre des travaux réalisés au cours de cette thèse.

E Enjeux de la thèse

Dans cette thèse, on s'intéressera aux sons produits par des actions entre des solides, qu'on présentera et dont on justifiera le choix dans la section qui suit. La possibilité de modifier des textures sonores grâce aux modèles de synthèse de sons d'actions entre solides développés, afin que ces textures évoquent des actions spécifiques, a également été étudiée dans cette thèse. Ce concept, qu'on appelle *métaphores sonores*, sera également présenté ensuite.

E.1 A partir des sons d'interactions entre objets solides...

Comme on l'a vu précédemment, la taxonomie décrivant les sons du quotidien proposée par Gaver (1993b) a été partiellement confirmée par plusieurs études (Lemaitre *et al.*, 2010, 2011; Houix *et al.*, 2012), qui mettent en avant trois grandes classes représentant l'état de la matière (solide, liquide, gazeux) et une quatrième qui est les sons produits par des machines. Houix *et al.* (2012) ont poussé l'étude en particulier sur les sons solides. Il a été demandé à 30 sujets de classer et décrire un corpus de sons en se focalisant sur les actions causant les sons indépendamment de la source sonore. Une analyse des résultats en groupes hiérarchiques a permis d'identifier un premier niveau séparant les sons en 2 groupes, d'une part les sons produits par la déformation d'un seul objet (e.g. froissement ou déchirement) et les sons produits par l'interaction entre plusieurs objets (e.g. impact ou frottement). Les déformations révèlent ensuite deux sous-groupes qui sont les déformations sans destruction (e.g. froissement) et celles avec (e.g. déchirement), et de même les interactions se divisent en deux sous-groupes qui sont les interactions discrètes (e.g. impact) et les interactions continues (e.g. frottement). Enfin, au niveau le plus bas se trouve la description spécifique de chaque interaction.

Dans cette thèse, nous avons choisi de nous focaliser sur le sous-groupe des interactions continues. Trois interactions continues spécifiques seront étudiées : "rouler", "frotter" et "gratter". Dans la littérature, les sons de roulement ont été étudiés tant du

point de vue perceptif (Houben *et al.*, 2001; Houben, 2002; Houben *et al.*, 2004, 2005; Stoelinga, 2007) que du point de vue de la synthèse sonore, par modèle physique (Rath et Rocchesso, 2005; Stoelinga, 2007; Stoelinga et Chaigne, 2007), par modèle de signal phénoménologique (Hermes, 1998; Van Den Doel *et al.*, 2001) et par schéma d'analyse/synthèse sur des enregistrements réels (Lagrange *et al.*, 2010; Lee *et al.*, 2010). En ce qui concerne les sons liés aux interactions "frotter" et "gratter", il n'existe pas à notre connaissance d'études dans la littérature ayant différencié ces deux interactions, les deux termes semblant être utilisés sans distinction particulière. Gaver (1993a) a proposé un premier modèle de signal phénoménologique de sons de frottements, qui a ensuite été amélioré par Van Den Doel *et al.* (2001). Comme décrit par Lagrange *et al.* (2010), le modèle d'analyse/synthèse qu'ils proposent est suffisamment générique pour être utilisé pour l'ensembles des sons d'interactions continues y compris les sons de frottements. Enfin d'un point de vue perceptif, les sons de frottements ont été étudiés par Thoret *et al.* (2014); Thoret (2014), et en particulier leur relation avec la perception des mouvements biologiques.

Les études liées au roulement seront détaillées dans le chapitre II, consacré à cette interaction. On s'attachera à y développer un modèle de signal, en se basant sur l'étude d'un modèle physique, et on proposera un contrôle intuitif du modèle de signal. Les études liées au frottement seront détaillées dans le chapitre III, consacré à l'étude des différences perceptives entre les interactions "frotter" et "gratter", ainsi qu'à la mise en place d'un modèle de synthèse permettant de synthétiser ces deux interactions et du contrôle intuitif associé. On présentera en fin de ce chapitre un synthétiseur prenant également en compte le roulement et permettant d'effectuer des transitions perceptives continues entre les trois interactions étudiées, de la même manière que les transitions continues entre matériaux comme proposé par (Aramaki et Kronland-Martinet, 2006; Aramaki *et al.*, 2009b, 2011).

E.2 ... vers les métaphores sonores

Outre les possibilités de synthèse précédemment décrites, le paradigme d'étude dans lequel s'inscrivent ces travaux ouvre également des voies d'exploration vers de nouvelles sonorités ("inouïes") en combinant virtuellement des actions à des objets qui n'y sont pas naturellement associés (e.g. combinaison de l'action "rouler" et "liquide", ou bien de "frotter" et de "vent"). Dans la même idée, il est possible de substituer "l'objet" à une "texture sonore", (e.g. un choeur tenant un accord ou une nappe de synthétiseur), équivalent d'une matière. Cela nous amène à la définition de "métaphores sonores" dans la mesure où ces combinaisons inédites induisent des évocations plus abstraites que les événements du quotidien, souvent recherchées pour leurs caractéristiques esthétiques. La définition de *Métaphore* donnée par le Larousse est la suivante :

> *Emploi d'un terme concret pour exprimer une notion abstraite par substitution analogique, sans qu'il y ait d'élément introduisant formellement une comparaison.*

On peut recadrer cette définition le contexte de la linguistique introduit par De Saussure (1916). Ferdinand De Saussure propose qu'un *signe* est composé du *signifié* qui est le concept, la représentation mentale associée à ce signe et du *signifiant* qui est l'image acoustique d'un mot et qui désigne le signifié. Ces travaux sont la base de la sémiotique, qui étudie les signes et leur signification, ne se restreignant pas au seul cadre du langage mais à l'ensemble des signes de toutes les modalités sensorielles. Ainsi, on peut voir la métaphore comme une modification ou une substitution du signifiant. La métaphore peut entre autres avoir pour but d'aider à la conceptualisation : par exemple, on peut difficilement parler de l'amour sans le concevoir comme une force physique

(*être attiré par quelqu'un*, *avoir un coup de foudre*...) ou encore comme de la folie (*il est fou d'elle*...) (Moriceau, 2003). La métaphore peut aussi avoir un but plus poétique, en vue de donner une image plus riche ("Pâle dans *son lit vert* où la lumière pleut", Arthur Rimbaud, *Le dormeur du val* ; ici le *lit vert* symbolise l'herbe).

On retrouve souvent ce terme employé dans la littérature sur la sonification (Hermann *et al.*, 2011). Dans la sonification, on emploie généralement ce terme lorsqu'une grandeur (e.g. une distance ou une température) est représentée par un son ayant un attribut (e.g. hauteur tonale ou rythme) variant avec elle, permettant ainsi de (mieux) comprendre la grandeur via l'audition, ou bien de substituer une représentation visuelle par une représentation auditive. On retrouve également ce terme dans la littérature sur la musique électroacoustique (et dans la musique en général). Ainsi, d'après Field (2000), les bruits de pas dans une pièce électroacoustique vont souvent être une métaphore du voyage, les sons de portes qui s'ouvrent et se ferment peuvent symboliser l'entrée et la sortie dans une partie de la pièce électroacoustique.

Dans le cadre des travaux présentés dans cette thèse, on voudra modifier des *textures sonores* (e.g. la pluie, un final d'orchestre, une nappe de synthétiseur...) afin de leur faire évoquer une interaction continue particulière (roulement ou frottement par exemple). Ces travaux seront présentés dans le chapitre IV, où les textures sonores seront clairement définies. L'idée sera donc d'évoquer des notions très abstraites comme "Faire rouler le final du requiem de Mozart", "Faire couiner une nappe de synthétiseur" ou encore "Frotter la pluie". Il est difficile de rattacher un signifié aux signifiants associés aux phrases précédemment citées. On proposera donc de passer par des signifiants qui ne sont plus dans le domaine linguistique. Pour ce faire, on considérera la texture sonore comme un objet sur lequel on peut interagir, et ainsi on représentera les invariants structurels de la texture sonore dans la partie filtre du modèle. Puis on pourra faire évoquer différentes interactions à la texture en entrant dans la partie filtre un signal source caractérisant l'invariant transformationnel lié à l'action. On peut donc voir cette proposition de schéma de métaphores sonores comme un remplacement du signifiant, bien qu'il ne clarifie pas nécessairement le sens signifié. Dans le chapitre IV, ces métaphores seront évaluées perceptivement, ce qui permettra également de tester la robustesse des invariants.

L'idée de ce type d'outils est de proposer une méthode pour des designer sonores et musiciens permettant de modifier le sens véhiculé par une texture sonore. Ainsi, afin de véhiculer un message particulier, l'utilisateur pourrait, grâce aux propositions faites dans cette thèse, enrichir la texture sonore de manière intuitive. Cette proposition est en phase avec la définition du design sonore donnée par Susini *et al.* (2014) (cf section C), qui est de proposer la création de sons "nouveaux", dans le sens où l'on ne peut pas les trouver dans des banques de sons existantes ou qui ne peuvent pas être enregistrés.

E.3 Méthodologie et organisation du document

La figure I.5 présente l'organisation de la thèse ainsi que la méthodologie qui sera adoptée. Les chapitres II et III seront dédiés à la mise en place de modèles de synthèse de sons d'interactions entre solides. Pour ce faire, on étudiera un modèle physique (chapitre II) ainsi que des enregistrements de sons d'interactions (chapitre III). La méthodologie employée sera la même. On supposera qu'il existe un invariant transformationnel qui permet de reconnaître une interaction particulière, et on validera cette hypothèse par un test perceptif. Des analyses qualitatives et quantitatives des signaux seront effectuées afin d'identifier les invariants structurels propres à chaque action étudiée. Ces structures invariantes seront par la suite modélisées et une stratégie de contrôle intui-

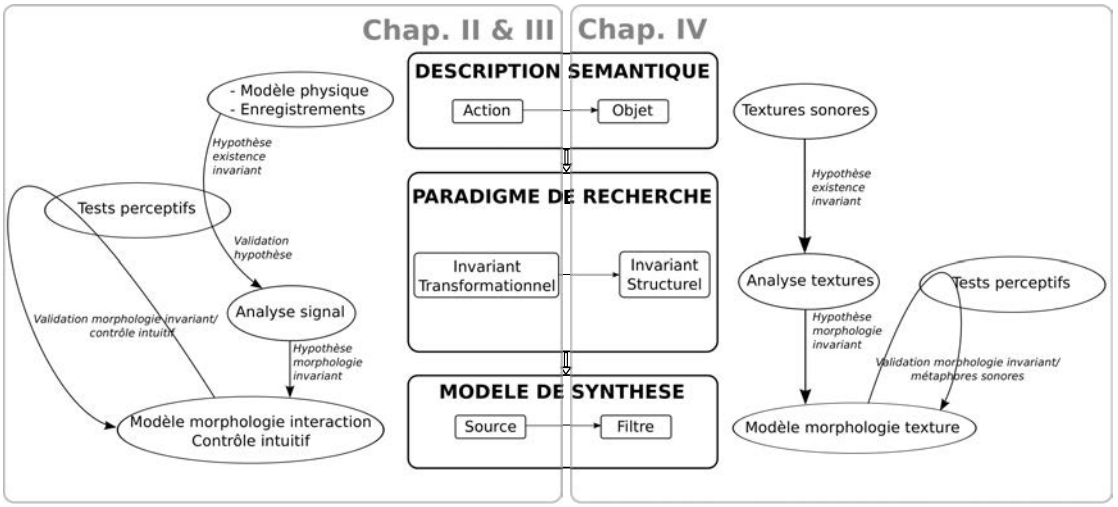

FIGURE I.5 – Méthodologie de la thèse et organisation du document.

tif sera proposée. Ces invariants, ainsi que les stratégies de contrôle associées seront validées par des tests perceptifs. Pour les métaphores sonores (chapitre IV), on se restreindra à une certaine classe de textures sonores. On supposera que la morphologie d'une texture sonore peut être représentée dans la partie filtre du modèle (invariant structurel). On proposera une méthode d'analyse qui permettra d'estimer "l'objet" texture, sur lequel on pourra interagir grâce aux modèles mis en place dans les chapitres précédents, puis une série de tests perceptifs permettra de valider le concept de métaphores sonores. Dans le dernier chapitre, on proposera quelques perspectives sur ces travaux.

Chapitre II

Synthèse et contrôle haut-niveau de sons de roulement

Sommaire

A	**Synthèse et perception des sons de roulement : état de l'art**	**25**
B	**Mise en évidence d'un invariant transformationnel du roulement**	**31**
C	**Modélisation de l'invariant transformationnel**	**34**
D	**Stratégie de contrôle intuitif**	**43**
E	**Evaluation perceptive de la stratégie de contrôle intuitif**	**46**
F	**Discussion générale**	**48**

Ce chapitre est basé sur les travaux présentés dans l'article de journal (Conan, Derrien, Aramaki, Ystad, et Kronland-Martinet, 2014a).

Dans cette partie, on décrira tout d'abord un modèle de synthèse de sons d'objets roulants, tel une bille. Après avoir passé en revue la littérature sur le sujet, on étudiera un modèle basé sur la physique, afin de mettre en évidence un invariant transformationnel lié au roulement : la force d'interaction entre la bille et la surface sur laquelle elle roule. Un test perceptif nous permettra de valider cet invariant et de montrer qu'il peut effectivement être utilisé comme excitation dans notre modèle source-filtre. On proposera ensuite une modélisation de cette force d'interaction, en constatant qu'elle peut être considérée comme une suite d'impacts ayant des propriétés statistiques particulières. Ce modèle sera établi grâce à un schéma d'analyse/synthèse sur la série d'impacts obtenue par simulation, ainsi qu'une modélisation de la forme de ces impacts, basée sur la physique. Comme nous le verrons dans le chapitre suivant, il sera judicieux d'avoir établi un modèle "signal" de cet invariant, car son caractère générique permettra de l'étendre à d'autres sons d'interactions, tels que le frottement et le grattement. Enfin, des contrôles haut-niveau du synthétiseur seront proposés pour la taille et la vitesse de la bille, ainsi que pour la rugosité de la surface sur laquelle l'objet roule. Ces contrôles seront validés par un test perceptif.

A Synthèse et perception des sons de roulement : état de l'art

Dans cette partie, nous passerons en revue les différents modèles de synthèse de sons de roulement proposés dans la littérature. On détaillera ensuite les quelques études sur la perception de ces sons.

A.1 Synthèse de sons de roulement

Les différentes méthodes de synthèse de sons de roulement existant dans la littérature seront ici présentées en prenant soin de séparer trois grandes catégories de modèles : les modèles basés sur la physique, les modèles de signaux empiriques, et les modèles basés sur des schémas d'analyse/synthèse.

A.1.1 Les modèles basés sur la physique

Bien que le phénomène physique sous-jacent à la production du son par des objets roulants ne soit pas clairement élucidé (Stoelinga, 2007), on trouve plusieurs approches fondées sur la physique dans la littérature.

Stoelinga et Chaigne (2007) ont proposé d'adapter le modèle décrit par Chaigne et Lambourg (2001) et Lambourg *et al.* (2001), qui permet la simulation d'un impact simple sur une plaque, afin de pouvoir considérer des excitations dont la position varie au cours du temps. Ce modèle considère les vibrations en flexion d'une plaque mince rectangulaire (modèle de Kirchhoff-Love (Morse et Ingard, 1968)) couplée à un impacteur via la loi de contact de Hertz (voir par exemple (Falcon *et al.*, 1998)). Le problème est résolu en utilisant une méthode numérique basée sur les différences finies. Pour le roulement, Stoelinga et Chaigne considèrent que la plaque présente des irrégularités de surface à une échelle microscopique et que l'impacteur (la bille) parcourt ces aspérités au cours de sa trajectoire. Les aspérités sont distribuées selon un profil de variation de hauteur qui suit un bruit blanc uniforme, agissant ainsi comme une perturbation dans le terme de compression de la loi de contact de Hertz. Ainsi, la force d'interaction F_c entre la bille et la plaque, obéissant à la loi de contact de Hertz, devient :

$$F_c = \begin{cases} k \left[\underbrace{R - \big(\eta(t) - W_p(x(t), y(t), t) - W_s(x(t), y(t))\big)}_{\mathrm{Y}} \right]^{3/2} & \\ \qquad\qquad \text{pour } \mathrm{Y} > 0, \\ 0 \qquad \text{sinon,} \end{cases} \tag{II.1}$$

où $(x(t), y(t))$ sont les coordonnées de la bille sur la plaque paramétrisées par le temps t, R le rayon de la bille, η le déplacement vertical de son centre de gravité, W_p le déplacement vertical de la plaque (i.e. ses vibrations en flexion), W_s le profil vertical de surface (simulé par un bruit blanc uniforme) et k la constante élastique de Hertz.

Rath et Rocchesso (2005) ont également proposé un modèle basé sur la physique pour synthétiser des sons de roulement. Cette méthode diffère de celle proposée par Stoelinga et Chaigne par l'utilisation d'une composante dissipative dans le modèle de contact (Hunt et Crossley, 1975). Ce modèle de contact fut pour la première fois utilisé dans un contexte de synthèse sonore par Avanzini et Rocchesso (2001b) pour générer des sons d'impacts. Il faut néanmoins noter que le modèle de Hunt et Crossley considère une surface de contact non-infinitésimale, et l'utilisation d'un tel modèle non-linéaire pour l'interaction du roulement peut être critiquée d'un point de vue physique. En effet, comme l'interaction entre la bille et les aspérités sont modélisées comme des micro-contacts, il peut sembler raisonnable de considérer un contact ponctuel et ainsi d'utiliser une loi de Hooke linéaire pour modéliser la force de contact. Cependant, la pertinence d'un point de vue perceptif du modèle décrit par Rath et Rochesso a été montrée à travers plusieurs tests perceptifs (Rath, 2004; Rath et Rocchesso, 2005). Une autre différence avec le modèle proposé par Stoelinga et Chaigne vient du fait que l'objet résonant (la plaque sur laquelle la bille roule par exemple) est modélisé

par un ensemble de N oscillateurs linéaires du second ordre (systèmes masse-ressort-amortisseur), tandis que Chaigne et Stoelinga discrétisent directement l'équation physique de la plaque vibrante. Chaque oscillateur représente un mode propre de l'objet résonant, i.e. sa fréquence de résonance et son amortissement (on a affaire à un modèle de synthèse modale (Adrien, 1991)). Le modèle peut être formalisé comme ceci :

$$\begin{cases} x = x_e - \sum_{i=1}^{N} x_r^{(i)} \\ \ddot{x}_r^{(i)} + g_r^{(i)} \dot{x}_r^{(i)} + \left[\omega_r^{(i)}\right]^2 x_r^{(i)} = \frac{1}{m_r^{(i)}} f(x, \dot{x}), i \in [\![1, N]\!] \\ \ddot{x}_e = g - \frac{1}{m_e} f(x, \dot{x}) \end{cases} \tag{II.2}$$

avec

$$f(x, \dot{x}) = \begin{cases} kx^\alpha + \lambda x^\alpha \dot{x} = kx^\alpha (1 + \mu \dot{x}) & , \quad x > 0 \\ 0 & , \quad x \leq 0 \end{cases} \tag{II.3}$$

Les paramètres de l'excitateur (par exemple une bille) et de l'objet résonant (par exemple une plaque ou une corde) sont marqués respectivement par les indices e et r. x_e et m_e sont le déplacement vertical et la masse de l'excitateur, respectivement. $\omega_r^{(i)}$ et $g_r^{(i)}$ sont respectivement la fréquence propre et l'amortissement du $i^{\text{ème}}$ mode propre de l'objet résonant, et $m_r^{(i)}$ la "masse" du $i^{\text{ème}}$ mode[1], qui contrôle les propriétés inertielles de l'oscillateur. $x_r^{(i)}$ est le déplacement vertical du $i^{\text{ème}}$ oscillateur. g est la constante gravitationnelle (9.81 m/s^2). La force d'interaction est représentée par $f(x, \dot{x})$. La constante de raideur du contact k est définie comme :

$$k = \frac{4}{3}\sqrt{R}\left(\frac{1 - \nu_e^2}{E_e} + \frac{1 - \nu_r^2}{E_r}\right)^{-1} \tag{II.4}$$

où E et ν sont respectivement le module de Young et le coefficient de Poisson et R est le rayon de la bille. λ est la constante d'amortissement (on définit ici $\lambda = \mu/k$), et l'exposant non-linéaire α rend compte de la géométrie locale à l'interface de contact (selon la théorie de Hertz, on prendra $\alpha = 3/2$ dans le reste de ce document).

Dans le but d'adapter ce modèle de contact non-linéaire à l'interaction de roulement, un profil de surface irrégulier est introduit. D'après des considérations physiques, la surface $\mathfrak{S}_r$ est supposée imparfaite, i.e. pas parfaitement lisse à une échelle microscopique, comme c'est le cas pour des surfaces réelles (Sayles et Thomas, 1978). Ainsi, la bille suit le profil de surface et impacte certaines aspérités, en fonction de sa taille et de celle de l'aspérité. La vitesse de la bille parallèlement à la surface est forcée et il est considéré dans ce modèle qu'elle n'est pas affectée par l'interaction. Le concept du modèle est schématisé sur la figure II.1. Ainsi, à mesure que la bille se déplace le long de la surface, un déplacement vertical (marqué x_{offset} sur la figure II.1) est ajouté à la variable de distance verticale x. Par conséquent, le modèle simule une bille qui rebondit localement et dont l'énergie est perturbée au cours du temps par le terme x_{offset}, évoquant ainsi une interaction de roulement. On peut également voir ce modèle comme une bille rebondissant sur une surface dont on fait aléatoirement varier la hauteur, comme dans l'expérience décrite par Luck et Mehta (1993) où est étudié le comportement dynamique d'une bille sur une plateforme vibrante.

1. Dans le modèle original de synthèse de sons de roulement (Rath et Rocchesso, 2005), les auteurs ne prennent pas en compte la position de la bille sur l'objet résonant. Néanmoins, comme proposé par Avanzini *et al.* (2002), cet effet (i.e. l'accentuation et l'atténuation de certains modes résonants en fonction de la position de l'excitateur sur l'objet résonant) peut être pris en compte dans le terme $m_r^{(i)}$.

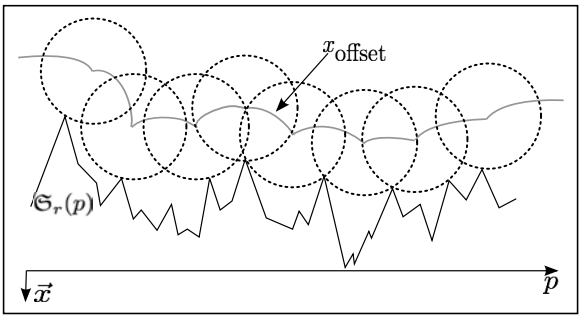

FIGURE II.1 – Terme de perturbation x_{offset}, déterminé par le profil de surface $\mathfrak{S}_r$ et la taille de la bille. Adapté de Rath et Rocchesso (2005).

A.1.2 Les modèles empiriques de signaux

Plusieurs auteurs ont proposé des modèles de signaux dans le but de synthétiser des sons de roulement, en se basant sur des considérations phénoménologiques. Ces modèles sont principalement basés sur des schémas source/filtre, présupposant ainsi que l'excitateur et le résonateur ne sont physiquement pas couplés (i.e. que les vibrations du résonateur n'influencent pas le comportement de l'excitateur). Ainsi, le processus peut être modélisé par un signal d'excitation (représentant la force appliquée au résonateur par l'excitateur) filtré par la réponse impulsionnelle de l'objet résonant. De telles méthodes n'imitent pas la physique mais ont plutôt pour but de reproduire les effets sonores pertinents d'un point de vue perceptif pour évoquer un son de roulement.

Le schéma de synthèse sonore baptisé "FoleyAutomatic", proposé par Van Den Doel *et al.* (2001) permet de reproduire différentes interactions continues telles que le frottement (nous y reviendrons dans le chapitre suivant) ou le roulement en utilisant une approche source/filtre. Le procédé de synthèse général consiste à lire un bruit à une vitesse variable proportionnelle à la vitesse relative entre les deux objets en interaction, le bruit modélisant la rugosité de la surface résonante. Les auteurs considèrent un modèle de surface fractale, connu pour bien modéliser les surfaces rugueuses réelles (Sayles et Thomas, 1978; Zahouani *et al.*, 1998). Les surfaces fractales ont une densité spectrale de puissance proportionnelle à ω^β ($\beta < 0$), où β est le paramètre contrôlant la rugosité de la surface (la surface est de plus en plus lisse à mesure que $|\beta|$ augmente). Pour la synthèse de sons de roulement, les auteurs proposent une étape de filtrage supplémentaire qui renforce l'enveloppe spectrale près du premier mode de l'objet résonant, en lien avec les observations faites par Hermes (1998). Ce dernier fait l'hypothèse que le contact entre la bille et la surface est plus "doux" pour le roulement que pour la friction. Pour prendre en compte cette hypothèse, les attaques franches des impacts modélisés généralement comme une somme de sinusoïdes exponentiellement amorties sont remplacées par des attaques plus graduelles. Ainsi, Hermes modélise chaque impact dans le cas du roulement par une somme de réponses impulsionnelles de type gammatone [2], i.e. $s(t) = \sum_{i=1}^{N} a_i t^{\gamma-1} \exp(-t/\tau_i) \sin(2\pi f_i t)$, avec $\gamma = 2$. Ce modèle d'impact est ensuite convolué à une série temporelle d'impulsions suivant une distribution de Poisson. Dans le but de renforcer l'impression d'entendre un objet roulant, l'enveloppe du son est modulée. Cela peut être justifié par le fait qu'un objet roulant n'est ni parfaitement sphérique ni homogène. Cette idée est également exploitée par Rath et Rocchesso (2005). Houben (2002) a par ailleurs montré la pertinence de ces modulations d'ampli-

2. Habituellement, les filtres gammatones sont principalement utilisés pour modéliser le filtrage cochléaire (Katsiamis *et al.*, 2007)

tude d'un point de vue perceptif : l'ajout de modulations d'amplitude artificielles à des sons de roulement enregistrés modifie la perception de la vitesse et de la taille de l'objet roulant.

A.1.3 Les modèles basés sur des schémas d'analyse/synthèse

Lagrange *et al.* (2010) et Lee *et al.* (2010) ont proposé des méthodes d'analyse/synthèse pour les sons générés par des objets roulants. Les deux approches considèrent un modèle source/filtre. L'idée générale est d'examiner des sons enregistrés dans le but d'extraire les paramètres de la partie "filtre" du modèle (qui caractérise les résonances des objets en interaction) et un signal utilisé comme excitation (terme "source") de la partie "filtre". Les deux modèles supposent que le son résultant du roulement est caractérisé par une suite de micro-impacts sur une surface résonante. Cependant, ces deux méthodes diffèrent dans la manière d'estimer les paramètres de la partie filtre et d'effectuer la détection des micro-impacts.

Lee *et al.* (2010) proposent d'estimer d'abord les instants où ont lieu les impacts en filtrant passe-haut le signal, puis en utilisant l'enveloppe d'énergie de ce signal filtré pour détecter les maxima, supposés comme étant les instants où ont lieu les impacts. Les paramètres de la partie filtre sont ensuite estimés pour chaque impact isolé filtré en sous-bandes. Dans chaque sous-bande, les zéros sont d'abord estimés puis enlevés du signal par filtrage inverse. Ces zéros modélisent les raies de suppression d'énergie qu'on peut observer sur des sons de roulement enregistrés lorsque la bille parcourt une surface finie (une plaque par exemple) (Stoelinga *et al.*, 2003), et qui sont dues au fait que les modes de l'objet résonant sont excités et atténués différemment en fonction du point d'excitation. Le fait d'estimer les zéros puis de les enlever du signal par filtrage inverse a également pour but de réduire l'ordre de la prédiction linéaire (LPC) utilisée par la suite pour estimer les pôles résonants. Une fois tous ces paramètres estimés, la synthèse peut être effectuée.

Lagrange *et al.* (2010) proposent d'extraire les paramètres de la partie filtre liée à l'objet résonant en utilisant la méthode haute-résolution ESPRIT (Roy et Kailath, 1989; Badeau *et al.*, 2002). L'estimation ESPRIT est effectuée sur des courtes trames du signal, qui sont déterminées par un critère sur le signal faisant en sorte que l'objet résonant soit le plus possible en régime "libre" (i.e. sur une partie où l'excitateur n'agit plus sur l'objet résonant, hormis le moment de contact au début de l'impact). Le signal source est ensuite déduit par filtrage inverse (déconvolution) dans le domaine fréquentiel. Chaque amplitude et instant d'impact est déterminé sur le signal source et l'excitation est ensuite codée en modélisant la distribution des amplitudes et des intervalles temporels entre impacts successifs respectivement par des distributions exponentielle et gamma. Chaque impact est ensuite modélisé en utilisant une fenêtre de Meixner (Schuijers *et al.*, 2003).

A.2 Perception des sons de roulement

La thèse de Houben (2002) (cf également (Houben *et al.*, 2004)) s'est intéressée à la capacité auditive à déterminer la taille et la vitesse de billes en bois roulant sur des surfaces en bois par l'écoute d'enregistrements. Pour les tests, les enregistrements présentant des modulations d'amplitudes ou des "irrégularités" (dues par exemple à un rebond de la bille) ont été exclus, et tous les sons ont été normalisés au même niveau en dB SPL afin que les sujets ne puissent se baser sur cet indice acoustique. Pour l'évaluation de la taille, les sujets devaient choisir la bille la plus grosse dans un test de comparaison entre 2 sons produits par des billes de tailles adjacentes (i.e. entre deux billes

de tailles successives, les diamètres choisis étant 22, 25, 35, 45, 55, 68 et 83 mm). Les résultats moyennés sur les huit sujets ayant participé à l'expérience montrent une bonne capacité à distinguer la bille la plus grosse entre deux sons. Pour évaluer la perception de la vitesse, 7 vitesses différentes ont été enregistrées (0.36, 0.50, 0.63, 0.71, 0.79, 0.87, et 0.93 m/s), et toutes les paires ont été comparées (la bille est à taille constante). Les résultats montrent encore une fois que les sujets sont capables de déterminer la bille la plus rapide. Des expériences supplémentaires montrent par la suite un fort effet d'interaction entre la taille est la vitesse, et que lorsque les deux covarient, les performances dans le jugement de la taille et de la vitesse diminuent (performances nettement plus dégradées pour la vitesse). Une étude des propriétés spectrales et temporelles sur lesquelles les auditeurs ont pu se baser est ensuite effectuée. Aucun descripteur temporel n'a été mis en évidence. En revanche, les auteurs calculent le "barycentre de sonie spécifique" (*centroid of the specific loudness*), qui décrit dans quelle bande auditive ERB l'intensité sonore est la plus élevée (Zwicker et Fastl, 1990). Ce descripteur permet de très bien prédire la discrimination de la taille de la bille. L'effet en fonction de la vitesse est beaucoup moins marqué. Par la suite, Houben *et al.* (2005) ont monté des expériences sur des stimuli transformés, en "mélangeant" les contenus spectraux et temporels de différents stimuli, dans le but d'évaluer l'importance de chaque information pour la perception de la taille et de la vitesse des billes roulantes. Les résultats montrent que si les sujets doivent discriminer la bille la plus grosse, ils choisissent effectivement un son présentant le spectre d'une grosse bille. Si en revanche les sujets doivent déterminer la bille la plus rapide, ils se basent plutôt sur le contenu spectral d'une petite bille. Les informations temporelles ne montrent ainsi que peu d'effet comparativement aux informations spectrales. Les auteurs en concluent donc que l'information de taille est prédominante sur l'information de vitesse, et qu'elle est basée sur des indices spectraux. Ces expériences confirment de plus l'effet d'interaction entre la taille et la vitesse, notamment que la perception de la vitesse est grandement influencée par la taille de la bille. Houben (2002) a ensuite évalué la modification de la perception de la taille et de la vitesse des billes roulantes en ajoutant une modulation d'amplitude artificielle aux enregistrements. Cette modulation dépend de la taille et de la vitesse linéaire de la bille (la fréquence de la modulation est proportionnelle à la vitesse linéaire et inversement proportionnelle à la taille). Les résultats généraux sont les suivants : **A/** si la vitesse angulaire de la bille (i.e. la modulation d'amplitude) est cohérente avec la vitesse linéaire (i.e. la vitesse physique, "réelle" des enregistrements), alors **(1)** on a une amélioration dans le jugement de la vitesse par rapport au cas sans modulation, **(2)** le jugement de la taille ne change pas significativement par rapport au cas sans modulation ; **B/** si la vitesse angulaire de la bille est incohérente avec la vitesse linéaire, alors la modulation d'amplitude influence la perception de la taille et de la vitesse.

La thèse de Stoelinga (2007) s'est quant à elle consacrée sur la perception des sons de roulements enregistrés et synthétiques. Dans une première expérience, la capacité à distinguer si une bille roule depuis le bord d'une plaque vers le centre ou inversement a été étudiée sur des stimuli enregistrés et de synthèse. Les stimuli de synthèse ont été réalisés en imposant le spectre moyen d'un son de roulement enregistré à un bruit blanc, puis en appliquant un filtrage en peigne variable dans le temps qui simule l'effet sonore induit par la variation de la position de la bille sur la plaque, comme proposé dans (Stoelinga *et al.*, 2003). Les résultats montrent que les sujets sont capables de discriminer les deux conditions (vers le centre par rapport à vers le bord), mais qu'ils ne sont cependant pas capables de réassocier cette différence à la direction de la bille. C. Stoelinga a ensuite répliqué et confirmé les expériences de (Houben, 2002; Houben *et al.*, 2004) qui montrent que dans un test de comparaison par paire sur des sons de

billes roulant à différentes vitesses (à taille constante) ou de tailles différentes (à vitesse constante), il est possible de discriminer la bille la plus rapide (la question posée par les auteurs était légèrement différente : "quelle bille a parcouru la plus grande distance ?") ou la plus grosse, respectivement. Par le même protocole, ils ont en revanche également montré que les sujets n'étaient pas capables de différencier la plaque la plus fine sur laquelle roule une bille. Les auteurs ont ensuite proposé, pour la taille et la vitesse, une expérience de jugement sur une échelle métrique absolue. Bien que les jugements des sujets ne correspondent pas à la véritable taille ou vitesse de l'objet ayant produit le son, l'étude des fonctions psychométriques obtenues a montré que les sujets sont capables de se construire une échelle qui leur est propre et sur laquelle ils effectuent leurs jugements. Ces résultats viennent donc appuyer les tests de comparaison précédents. Afin de compléter ces études, Stoelinga (2007) a également conduit des tests de dissemblance par comparaison entre paires de sons (des enregistrements ont été effectués pour tous les croisements possibles de 3 tailles de billes, 3 vitesses de billes et 3 épaisseurs de plaques). Pour chaque paire, les sujets ont dû juger de la différence entre les 2 sons sur une échelle de 1 ("pas de différences audibles") à 5 ("très différents"). Une analyse multidimensionnelle ("Multidimensional scaling" ou MDS) montre que l'épaisseur de la plaque se trouve dans une dimension orthogonale à la vitesse et à la taille de la bille. Les différences dues aux variations de l'épaisseur de la plaque sont donc perçues mais éludées. Ainsi, bien que les auditeurs ne soient pas capables de remonter à l'épaisseur de la plaque, Stoelinga suggère que cette information est pré-traitée afin de l'exclure pour juger de la taille et de la vitesse de la bille qui sont peut-être plus importantes. C. Stoelinga propose une analogie intéressante avec la perception de la parole. En effet, le système auditif humain permet de percevoir la parole dans un grand nombre d'environnements différents, en particulier dans différentes salles qui filtrent la parole différemment, causant parfois de grandes modifications au signal de parole original. Cette capacité à ignorer le filtrage permet de se focaliser sur l'événement sonore d'intérêt pour l'auditeur. De la même manière, il suggère que le procédé pour la perception des propriétés du roulement (vitesse et taille) fait abstraction de l'effet induit par l'épaisseur de la plaque. La seule différence est que la plaque est partie intégrante du système générant le son, et qu'il ne peut y avoir de son sans la plaque. Ainsi, on ne peut pas écouter le son de roulement sans le filtre, tandis qu'on peut écouter la voix de quelqu'un sans le filtre de la pièce, c'est-à-dire dehors.

B Mise en évidence d'un invariant transformationnel du roulement

Dans cette partie, on se focalisera sur les caractéristiques perceptives du signal responsables de l'évocation de l'interaction de roulement. La force d'interaction bille-surface calculée grâce au modèle proposé par Rath et Rocchesso (2005) (cf partie A.1.1) évoque, selon nous, fortement le roulement et contient donc sûrement de l'information pertinente sur cette interaction. On suppose donc qu'elle est un invariant transformationnel du roulement et qu'elle porte par conséquent l'information pertinente liée à l'évocation du roulement, indépendamment des propriétés du résonateur. Afin de tester cette hypothèse, on étudiera des sons de synthèse plutôt que des enregistrements, car la synthèse sonore permet de créer des stimuli contrôlables de manière très précise.

Le modèle de synthèse choisi est celui proposé par Rath et Rocchesso (2005). Premièrement, de notre point de vue, ce modèle produit les sons les plus convaincants (♯). Il a été montré par un test perceptif que ce modèle produisait des sons sponta-

nément reconnus par les auditeurs comme évoquant un objet roulant (Rath, 2004). De plus, ce modèle permet de se focaliser sur l'excitation, i.e. la force d'interaction, en réduisant la contribution de l'objet résonant. On considèrera donc par la suite des forces d'interaction calculées sur des surfaces rigides, i.e. sans contribution du résonateur. Le résonateur sera pris en compte dans un deuxième temps dans le processus, par convolution avec sa réponse impulsionnelle. Les équations (II.2) et (II.3), perturbées au cours du temps par le terme x_{offset} sont discrétisées puis résolues grâce à un schéma explicite de Runge-Kutta d'ordre quatre (RK4) (Ascher et Petzold, 1998), connu pour être une méthode précise pour résoudre des équations différentielles. Comme montré par Papetti *et al.* (2011), les solutions obtenues par la méthode RK4 sur un modèle physique d'une bille rebondissant (et utilisant le même modèle de contact que (II.3)) sont toujours proches des solutions analytiques. La surface est discrétisée avec une résolution spatiale de 0.1 mm et la fréquence d'échantillonnage temporelle est de f_s=44.1 kHz.

Dans la suite, on mettra en œuvre un test perceptif pour vérifier si, pour certaines combinaisons de paramètres, l'information perceptive pertinente liée à l'évocation du roulement est effectivement reliée à la force d'interaction. Le test étudiera également la contribution des caractéristiques modales dans l'évocation du roulement.

B.1 Sujets

Dix-sept sujets ont pris part à l'étude : 12 hommes, 5 femmes, âgés de 30 ans en moyenne (écart-type : 8.7 ans), membres du laboratoire. Aucun d'eux n'avait eu connaissance des stimuli ni ne présentait de troubles auditifs.

B.2 Stimuli

D'après les équations (II.2) et (II.3), on peut voir que durant le contact entre l'objet roulant et la surface ($x > 0$), le couple de coefficients $(\frac{k}{m}, \frac{\lambda}{k}) \stackrel{\text{def}}{=} (\kappa, \mu)$ paramétrise le comportement de la force d'interaction. Dans le but d'étudier si certaines combinaisons de ces coefficients permettent l'évocation du roulement, des forces d'interaction ont été générées en échantillonnant régulièrement l'espace des paramètres (κ, μ) pour toutes les combinaisons possibles entre $\kappa = [\kappa_1, \kappa_2, \kappa_3] = \left[5 \cdot 10^7, 5 \cdot 10^9, 5 \cdot 10^{11}\right]$ et $\mu = [\mu_1, \mu_2, \mu_3] = [0.1, 1, 10]$. Les valeurs de κ ont été choisies par rapport à la littérature : Stoelinga et Chaigne (2007) utilisent des valeurs entre 10^4 et 10^9, tandis que Falcon *et al.* (1998) utilisent des valeurs de l'ordre de 10^{12} N.m$^{-\alpha}$.kg^{-1}. Pour chaque paire, des simulations de 5 secondes ont été effectuées sur deux surfaces fractales différentes ($\beta = -0.5$ et $\beta = -1$). Dix-huit stimuli associés aux forces d'interactions seules ont ainsi été générés.

Dans le but d'étudier la contribution d'un objet résonant à la perception du roulement, 18 stimuli supplémentaires ont été obtenus en convoluant les forces seules à une réponse impulsionnelle évoquant un objet métallique (obtenu d'après la méthode proposée par Aramaki *et al.* (2011)), portant à 36 au total le nombre de stimuli pour le test. Pour toutes les simulations, une vitesse constante de 20 cm/s a été imposée, un rayon de bille constant de 1 cm et une hauteur d'aspérité maximum de 0.1 μm. Une enveloppe d'amplitude a été appliquée à tous les stimuli (une gaussienne a été empiriquement choisie), afin d'évoquer une source approchant l'auditeur puis s'en éloignant ($\sharp$).

B.3 Protocole

Le test s'est déroulé dans un bureau calme, sur un ordinateur portable standard avec un casque Sennheiser HD-650. Une interface graphique spécifique a été développée sous Max/MSP[3]. Un jugement subjectif de la "sensation de roulement" était demandé aux participants. A chaque stimulus, l'auditeur devait évaluer à quel point le son lui évoquait le son d'un objet roulant, grâce à un curseur allant de 0 à 100. Des descripteurs verbaux pour la "sensation de roulement" ont été utilisés, comme proposé par Murphy *et al.* (2011), allant de "Pas du tout comme un roulement" (0), "Un peu comme un roulement" (25), "A peu près comme un roulement" (50), "Presque comme un roulement" (75) à "Exactement comme un roulement" (100). Avant l'évaluation, 6 sons différents étaient présentés au sujet pour le familiariser avec le type de stimuli du test. Durant l'évaluation, les 36 stimuli ont été présentés dans un ordre aléatoire pour éviter l'effet de l'ordre de présentation.

B.4 Résultats

Les données ont été analysées en utilisant une analyse de la variance à mesures répétées (ANOVA) avec comme facteurs l'irrégularité de la surface β (2 niveaux), la contribution du résonateur (2 niveaux), κ (3 niveaux) et μ (3 niveaux). On considère les effets significatifs si la valeur de p est inférieure ou égale à 0.05. Un effet principal significatif a été trouvé pour la contribution du résonateur ($F(1,16)$=16, p=0.001). La "sensation de roulement" moyenne est de 40.9% pour les forces seules et de 53.7% pour les forces convoluées, i.e. environ 13% de plus quand la force est convoluée avec la réponse impulsionnelle d'un objet résonant. L'effet de surface n'a pas été significatif ($p = 0.13$). En revanche, l'interaction $\kappa \times \mu$ a été significative ($F(4,64)$=5.08, p=0.001). Les sons générés avec le jeu de paramètres (κ_2, μ_2) ont obtenu les notes les plus hautes, et ont donc été perçus les plus proches du roulement. Ce jeu de paramètres diffère significativement de toutes les autres combinaisons ($p < 0.01$ au moins) sauf pour le couple (κ_2, μ_3) ($p = 0.07$). Sur la figure II.2, l'évaluation de la "sensation de roulement" pour les forces seules (i.e. non convoluées avec la réponse impulsionnelle de l'objet résonant), moyennée sur les sujets et les deux surfaces différentes, est présentée dans le plan (κ, μ).

B.5 Discussion

Les résultats de ce test perceptif montrent que les sons générés avec le couple (κ_2, μ_2) ont été plus souvent perçus comme produits par un objet roulant que les autres combinaisons (67.2%, cf figure II.2). Les jugements de "sensation de roulement" sont déjà assez élevés pour les sons basés uniquement sur la force d'interaction, ce qui tend à montrer que cette force d'interaction est un indice acoustique pertinent pour la perception du roulement. De plus, convoluer cette force avec la réponse impulsionnelle d'un objet résonant améliore significativement la "sensation de roulement" (jugement à 83.5% pour le couple (κ_2, μ_2), qui obtient également la meilleure note dans la condition où la force est convoluée à une réponse impulsionnelle). On pourrait avancer que le test perceptif est biaisé du fait qu'il était explicitement dit aux sujets qu'ils allaient entendre des "sons de roulement". Cependant, après l'expérience, la plupart des sujets rapportaient que certains sons évoquaient une petite bille roulant sur une surface dure (béton, carrelage) et d'autres sur une plaque métallique, ainsi ils étaient capables d'imaginer l'action et même un scénario d'après le son présenté. On peut ainsi conclure que

3. `http://cycling74.com/`

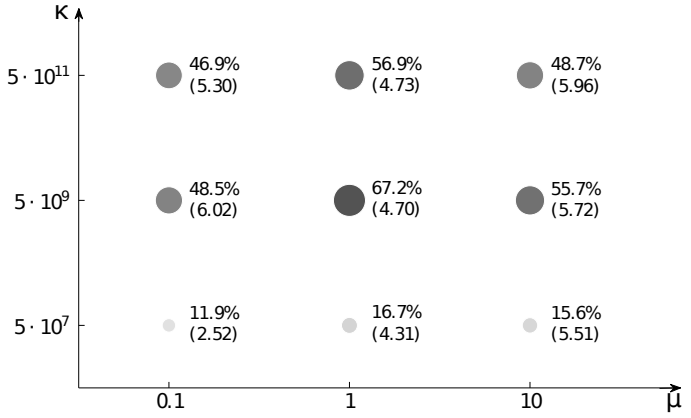

FIGURE II.2 – Résultats du test perceptif pour les sons correspondants aux forces seules (pas de convolution avec la réponse impulsionnelle d'un objet résonant). Le jugement de la "sensation de roulement" moyenne (erreur type de la moyenne entre parenthèse) est indiqué pour chaque couple (κ, μ). L'échelle de gris et la taille des cercles est proportionnelle au jugement moyen.

la force d'interaction a un impact fort sur la perception de la "sensation de roulement". Cette expérience a mis en évidence une zone privilégiée dans l'espace (κ, μ), centrée autour de (κ_2, μ_2), dans laquelle la perception du roulement est manifeste. On utilisera donc par la suite le couple de paramètres (κ_2, μ_2) et des combinaisons alentours comme référence pour le reste de ce chapitre.

C Modélisation de l'invariant transformationnel

Le test perceptif précédent a permis de révéler que la force d'interaction f elle-même était porteuse de l'information auditive pertinente pour l'évocation du roulement. De plus, convoluer cette force avec la réponse impulsionnelle d'un objet vibrant renforçait l'évocation du roulement. Dans cette partie, on s'attachera à proposer une modélisation d'un point de vue signal de cette force de roulement. D'un point de vue signal, cette force d'interaction peut être considérée comme une suite d'impacts caractérisés par des statistiques d'amplitude et d'instants de déclenchement particuliers. De plus, ces impacts ont une forme spécifique dépendant de leur amplitude, que nous modéliserons. Modéliser cette série d'impacts implique d'analyser et reproduire ces statistiques. Comme on le verra dans le chapitre suivant, le fait de proposer un modèle de signal "générique" permettra d'étendre le modèle à d'autres types d'interactions continues.

Dans cette partie, on proposera un modèle de synthèse incluant un schéma d'analyse/synthèse ainsi que deux processus supplémentaires :

— une modulation d'amplitude qui a été montrée comme étant un indice pertinent pour la perception de la vitesse de l'objet roulant (cf A.2)
— un modèle d'impact isolé qui prend en compte la dépendance de la durée de contact (i.e. la durée de l'impact) avec l'amplitude

Dans la suite de cette partie, une caractérisation d'un point de vue signal ainsi que les principales contributions du modèle de synthèse proposé seront données. Plus loin, les détails sur l'estimation des paramètres seront fournis.

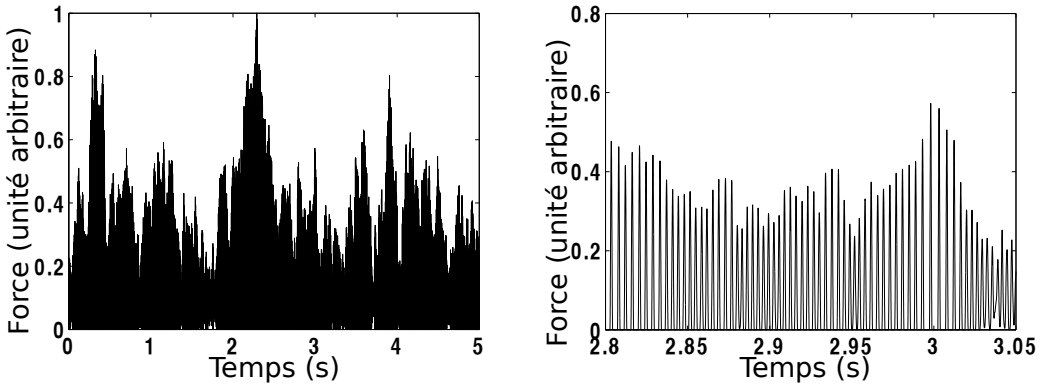

FIGURE II.3 – Exemple typique de comportement de la force d'interaction non-linéaire perçue comme un roulement. Le zoom sur la figure de droite permet de constater que f peut être considérée comme une suite d'impacts.

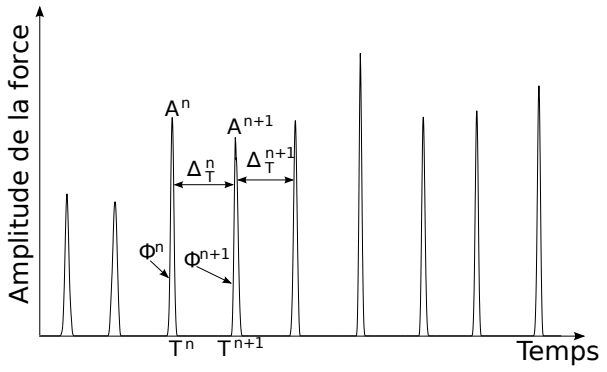

FIGURE II.4 – Notations du modèle décrit par l'équation (II.5). Le zoom a été effectué sur une très courte période (de l'ordre de 10 ms) de la force présentée figure II.3.

C.1 Caractérisation de la force d'interaction

On se focalise ici sur la force d'interaction f dans le but de caractériser son comportement d'un point de vue signal. Comme on peut le voir sur un exemple typique de comportement de la force d'interaction non-linéaire perçue comme un roulement (figures II.3 et II.4), on peut considérer f comme une succession de micro-impacts. On peut formaliser cette observation par le modèle suivant :

$$f(t) = \sum_n A^n \, \phi^n(t - T^n) \tag{II.5}$$

où A^n et T^n sont respectivement l'amplitude maximum et l'instant de déclenchement du $n^{\text{ième}}$ impact. Ils caractérisent ce qu'on appellera par la suite la "séquence d'impacts" et seront considérés comme des variables aléatoires. ϕ^n représente ce qu'on appellera par la suite la "forme d'impact". Du fait du comportement non-linéaire de la force d'interaction, la "forme d'impact" varie avec l'index n. De tels modèles ont déjà été proposés dans la littérature et on en décrira un dans la partie C.4.

Modéliser les propriétés statistiques de la séquence d'impacts est un problème plus complexe. Pour ce faire, on suppose que la bille roule sur une surface "infinie" et ayant des propriétés statistiques constantes. Par conséquent, à vitesse constante, le proces-

sus stochastique modélisant la série d'impacts est stationnaire, i.e. ses propriétés statistiques ne dépendent pas de l'origine temporelle. Il est donc plus pertinent de considérer l'intervalle temporel entre deux impacts successifs plutôt que la position temporelle absolue de chacun des impacts. Ainsi, on introduit la série $\Delta_T^n = T^{n+1} - T^n$ (cf figure II.4). La séquence d'impacts est donc caractérisée par la série temporelle bi-dimensionnelle (A^n, Δ_T^n).

Les comportements statistiques des séries A^n et Δ_T^n pour plusieurs simulations de la force d'interaction (simulées d'après les équations (II.2) et (II.3)) ont été étudiés. Les valeurs des paramètres physiques pour les simulations ont été choisies de manière à être cohérentes avec les résultats obtenus dans le test perceptif sur la sensation de roulement (cf partie B) :

— Vitesse de la bille : 20 cm/s
— Rayon de la bille : 1 cm
— $\kappa = \left[10^9, 5{\cdot}10^9, 10^{10}\right]$ N.m$^{-\alpha}$.kg^{-1}
— $\mu = [0.5, 1, 2]$ s.m^{-1}
— $\beta = [0, -0.5, -1]$
— Hauteur maximale des aspérités : 0.1 μm

Les séries A^n et Δ_T^n ont été extraites en détectant les maxima des forces simulées. Par définition, A et Δ_T ont des valeurs strictement positives, et donc des valeurs moyennes strictement positives. Pour analyser ces processus stochastiques, on calcule les variables centrées de A et Δ_T en leur soustrayant leur valeur moyenne. On note ${}^cA^n = A^n - \mu_A$ et ${}^c\Delta_T^n = \Delta_T^n - \mu_{\Delta_T}$, où μ_X est la moyenne du processus X. On évalue ensuite les fonctions d'autocorrélation $C_{({}^cA^n, {}^cA^n)}(k)$ et $C_{({}^c\Delta_T^n, {}^c\Delta_T^n)}(k)$ de ${}^cA^n$ et ${}^c\Delta_T^n$ respectivement. Ces fonctions, représentées sur la figure II.5 (lignes du milieu et du bas), montrent que les deux processus ne sont clairement pas blancs, i.e. que la valeur X^n dépend des valeurs passées X^{n-i} ($i < n$). L'observation d'un tel processus dont l'état présent dépend des états précédents est cohérent avec la physique, car la série d'impacts est structurée dans le temps : le modèle de roulement est dérivé de celui de rebond, qui présente une structure organisée (les impacts successifs sont de plus en plus rapprochés dans le temps et d'amplitude de plus en plus faible, cf A.1.1). La corrélation $C_{({}^cA^n, {}^c\Delta_T^n)}(k)$ entre ${}^cA^n$ et ${}^c\Delta_T^n$ est également représentée figure II.5. On peut constater que les deux variables aléatoires sont fortement corrélées, ce qui est également cohérent avec la physique : les impacts de faible amplitude A (respectivement de grande amplitude) sont généralement suivis d'un rebond rapproché (respectivement distant), caractérisé par un petit (respectivement un grand) intervalle Δ_T. Cette structure particulière des fonctions de corrélations caractérise probablement l'interaction de roulement, et sa reproduction sera cruciale pour effectuer une synthèse convaincante.

C.2 Schéma d'analyse/synthèse de la séquence d'impact

Le diagramme du schéma d'analyse/synthèse de la série d'impacts est représenté sur la figure II.6. Le schéma de synthèse représente le processus de synthèse complet, i.e. après la synthèse des séries A^n et Δ_T^n, l'amplitude est modulée et un modèle de forme d'impact est appliqué. Ces deux étapes sont décrites dans les parties C.3 et C.4. Ci-après, on décrira les points principaux du processus d'analyse/synthèse. Les détails concernant l'estimation des paramètres sont donnés C.5.

Schéma d'analyse Comme décrit dans la partie C.1, les séries A^n et Δ_T^n sont tout d'abord extraites de simulations numériques du modèle de Rath et Rocchesso (2005) et centrées en leur soustrayant leur valeur moyenne. L'étape d'analyse consiste à sup-

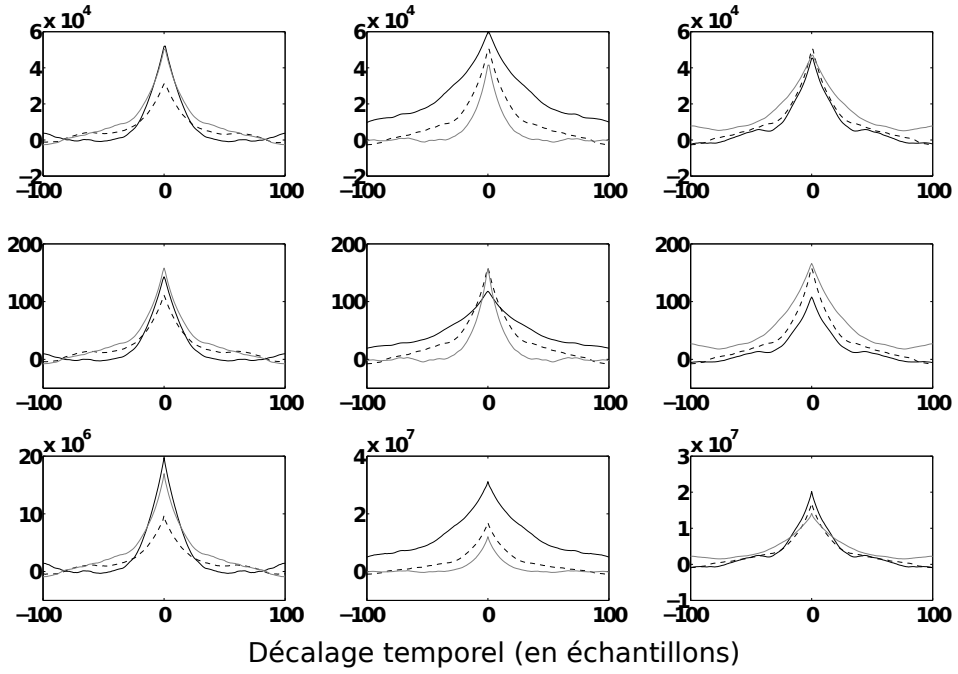

FIGURE II.5 – Fonctions de corrélations de séquence d'impacts d'une bille roulant sur une surface fractale pour différents jeux de paramètres (κ, μ, β). *Ligne du haut :* Corrélation entre ${}^cA^n$ et ${}^c\Delta_T^n$. *Ligne du milieu :* Autocorrélation de ${}^cA^n$. *Ligne du bas :* Autocorrélation de ${}^c\Delta_T^n$. *Colonne de gauche :* κ=10^9 (tirets noirs), $5{\cdot}10^9$ (gris), 10^{10} (trait plein noir) N.m$^{-\alpha}$.kg^{-1}. *Colonne de milieu :* μ=0.5 (trait plein noir), 1 (tirets noirs), 2 (gris) s.m^{-1}. *Colonne de droite :* β=0 (trait plein noir), -0.5 (tirets noirs), -1 (gris). Quand ils ne varient pas, les paramètres sont : κ=$5 \cdot 10^9$ N.m$^{-\alpha}$.kg^{-1}, μ=1 s.m^{-1} et β=-0.5.

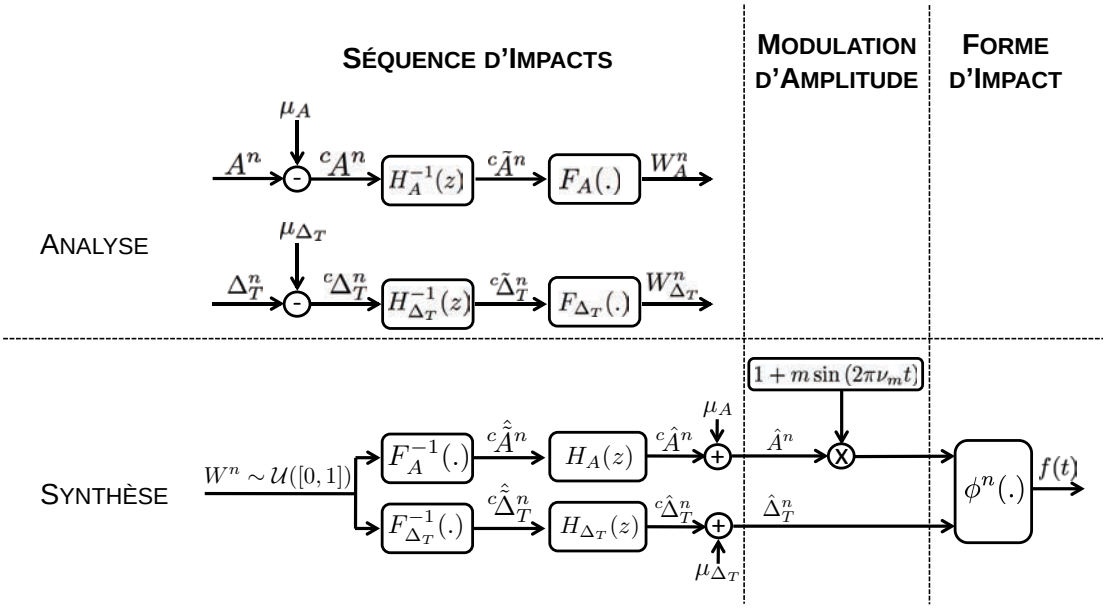

FIGURE II.6 – Haut : Schéma d'analyse des séries (A^n, Δ_T^n). Bas : Schéma de synthèse global.

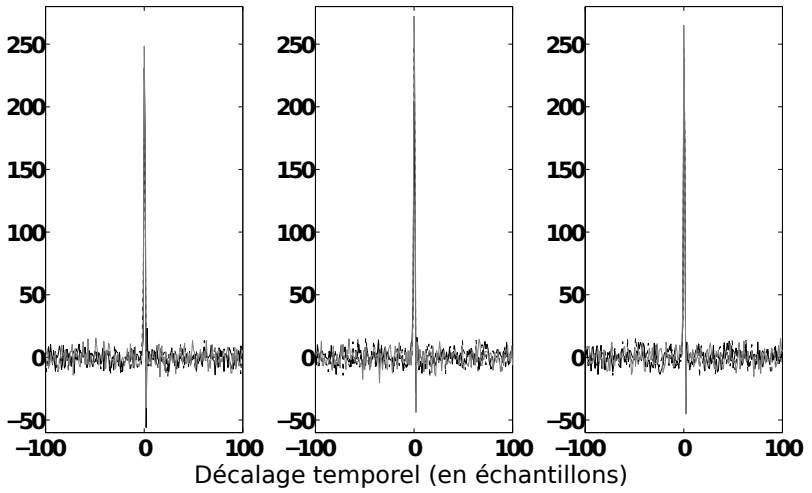

FIGURE II.7 – Corrélations croisées entre les sorties W_A^n and $W_{\Delta_T}^n$ du schéma d'analyse. Les paramètres et les couleurs utilisées sont les mêmes que sur la figure II.5.

primer l'auto-corrélation en estimant les filtres blanchisseurs $H_A^{-1}(z)$ et $H_{\Delta_T}^{-1}(z)$. Les versions blanchies des séries temporelles sont nommées ${}^c\tilde{A}^n$ and ${}^c\tilde{\Delta}_T^n$. Les fonctions de répartition de ${}^c\tilde{A}^n$ and ${}^c\tilde{\Delta}_T^n$ (nommées respectivement $F_A(.)$ and $F_{\Delta_T}(.)$) sont estimées dans le but de transformer ces séries temporelles afin qu'elles suivent une distribution uniforme. Ces dernières sont nommées W_A^n et $W_{\Delta_T}^n$.

Enfin, la corrélation $C_{(W_A^n, W_{\Delta_T}^n)}$ entre les signaux résiduels W_A^n et $W_{\Delta_T}^n$ est estimée. On a représenté $C_{(W_A^n, W_{\Delta_T}^n)}$ sur la figure II.7. Etant donnée l'autocorrélation à long terme de A^n et Δ_T^n (cf figure II.5), on suppose que $C_{(W_A^n, W_{\Delta_T}^n)}$ est proche d'une fonction delta de Dirac, en d'autres termes que les séries W_A^n et $W_{\Delta_T}^n$ sont totalement corrélées et représentent un seul et même processus stochastique (ce comportement restant vrai pour toutes les combinaisons de paramètres considérées). Cette hypothèse va dans le sens de la physique : la seule variable aléatoire dans le modèle physique considéré (Rath et Rocchesso, 2005) (décrit en partie A.1.1) est le profil de surface. Le reste du modèle est totalement déterministe.

Schéma de synthèse Pour un processus X (i.e. A ou Δ_T), le schéma de synthèse est obtenu en inversant le schéma d'analyse. Comme expliqué dans la partie précédente, le même processus blanc $W \sim \mathcal{U}([0,1])$ est utilisé pour synthétiser ${}^c\hat{A}$ et ${}^c\hat{\Delta}_T$. ${}^c\hat{X}^n$ peut donc être modélisé par :

$$ {}^c\hat{X}(t) = \left[F_X^{-1}(W) * h_X \right](t) \tag{II.6} $$

où le symbole $\hat{}$ représente le "processus estimé" (i.e. le processus simulé), h_X est la réponse impulsionnelle du filtre inverse H_X, et $*$ est le produit de convolution. Ici t représente le temps discret, déterminé en fonction de la fréquence d'échantillonnage.

C.3 Indice pour la perception de la vitesse de roulement

Des simulations numériques avec modèle physique ont conduit à la conclusion que la modification de la vitesse de la bille affecte faiblement la sensation perceptive de vitesse ($\sharp$). D'un point de vue perceptif, l'asymétrie d'un objet roulant, entraînant des modulations d'amplitude dans le son produit, contribue à la perception de sa vitesse. En

effet, comme montré par Houben (2002), la perception de la vitesse d'une bille roulante est majoritairement véhiculée par des indices temporels (cf partie A.2). Néanmoins, la physique des objets roulants asymétriques est complexe (voir par exemple la thèse de Theron (2008)) et le lien entre le processus physique et la perception du roulement n'a pas été montré. Cependant certains auteurs ont supposé que les objets roulants ne sont jamais parfaitement homogènes ni parfaitement symétriques, et ont proposé une simple modulation sinusoïdale de la force d'interaction (Hermes, 1998; Rath et Rocchesso, 2005). D'après ces études, on introduit donc dans le modèle de synthèse une modulation d'amplitude, et l'équation II.5 devient :

$$f(t) = [1 + m \sin(2\pi\nu_m t)] \sum_n A^n \, \phi^n(t - T^n) \tag{II.7}$$

avec $m \in [0, 1]$ la profondeur de modulation. La fréquence de modulation suit :

$$\nu_m \propto \frac{v}{R} \tag{II.8}$$

où v est la vitesse transversale de la bille et R son rayon.

C.4 Modélisation de la forme de l'impact

Dans cette partie, on se focalisera sur la modélisation de la forme des impacts ϕ^n. Un modèle simplifié défini par une fonction cosinus surélevé est proposé par Van Den Doel *et al.* (2001) :

$$\phi^n(t) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\frac{2\pi t}{t_0^n}\right)\right] & , \quad t \in \left[-\frac{t_0^n}{2}, \frac{t_0^n}{2}\right] \\ 0 & , \quad \text{sinon} \end{cases} \tag{II.9}$$

où t_0^n est la durée de l'impact. Dans notre modèle, t_0^n n'est pas constant de par le comportement non-linéaire du modèle de force d'interaction. Il n'est cependant pas considéré comme une variable aléatoire indépendante, ainsi on suppose que t_0^n est déduit de la séquence d'impacts via une loi déterministe présentée plus loin. Afin de simplifier les notations, on se focalisera sur un impact isolé et on omettra l'indice n.

Le second terme $\lambda x^\alpha \dot{x}$ dans le modèle de force d'interaction de l'équation (II.3) est connu pour introduire de l'hystérésis (voir par exemple (Avanzini et Rocchesso, 2004)), de telle sorte que plus la vitesse d'impact est grande, plus ce dernier a une forme asymétrique. Dans le but de valider le modèle d'impact proposé équation (II.9), on examine tout d'abord la gamme de vitesses d'impact présente dans une interaction de roulement type. Plusieurs simulations d'une bille roulante à différentes vitesses (de 10 cm/s à 100 cm/s) ont été réalisées pour toutes les combinaisons de paramètres κ, μ et β considérées figure II.5. La figure II.8 représente la distribution des vitesses d'impacts obtenues. Celles-ci n'excèdent pas 5 cm/s. Pour des vitesses d'impacts inférieures à 5 cm/s (avec les valeurs de κ et μ proposées figure II.5), le modèle proposé (II.9) s'ajuste sur les impacts simulés avec l'équation (II.3) avec un coefficient de corrélation $0.90 < R^2 < 0.99$. Par conséquent, le modèle d'impact en cosinus surélevé (II.9) convient à notre étude.

La non-linéarité α dans le modèle de contact (II.3) introduit une dépendance entre la durée de l'impact et la vitesse à laquelle la bille impacte la surface (Falcon *et al.*, 1998; Chaigne et Doutaut, 1997; Avanzini et Rocchesso, 2001b). Avanzini et Rocchesso (2001b) ont montré que la durée du contact t_0 est donnée par :

$$t_0 = \underbrace{\left(\frac{m_e}{k}\right)^{\frac{1}{\alpha+1}}}_{\mathcal{A}} \underbrace{\left(\frac{\mu^2}{\alpha+1}\right)^{\frac{\alpha}{\alpha+1}}}_{\mathcal{B}} \underbrace{\int_{v_{out}}^{v_{in}} \frac{dv}{(1+\mu v)\left[-\mu\,(v - v_{in}) + \log\left|\frac{1+\mu v}{1+\mu v_{in}}\right|\right]^{\frac{\alpha}{\alpha+1}}}}_{\mathcal{C}} \tag{II.10}$$

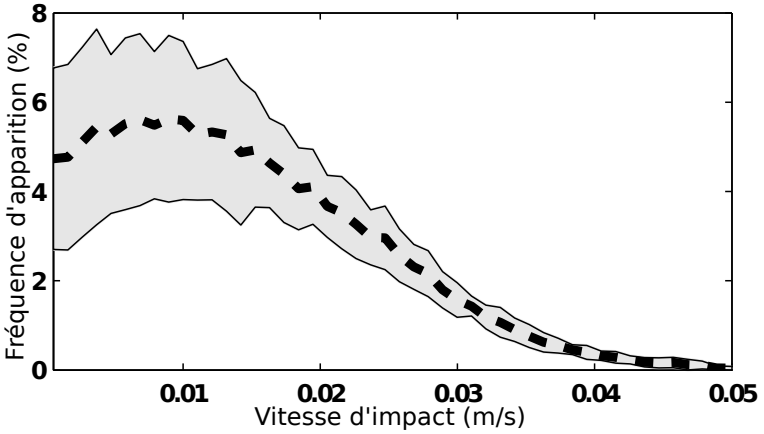

FIGURE II.8 – Distribution simulée des vitesses d'impacts durant l'interaction de roulement. Ligne noire pointillée : moyenne. Aire grisée : écart type.

où v_{in} et v_{out} sont respectivement les vitesses au début et à la fin de l'impact. On peut montrer que v_{out} est une fonction dépendant de μ et v_{in}, ainsi l'intégrale $\mathcal{C}$ ne dépend que de μ et v_{in}. Avanzini et Rocchesso (2004) ont montré que la durée de contact est très peu (voire pas du tout) affectée par μ, tandis que v_{in} a une grande influence sur t_0. D'après les développements analytiques de Chaigne et Doutaut (1997) qui montrent que $t_0 \propto v_{in}^{-1/5}$ (dans leur étude, les auteurs ne considèrent pas le terme dissipatif μ qui est considéré nul), on propose de modéliser la durée du contact comme proportionnelle à une loi puissance de la vitesse d'impact. Comme μ n'affecte pas le terme $\mathcal{B}\times\mathcal{C}$ dans (II.10), la durée de contact peut être réécrite comme :

$$t_0 = \zeta \cdot v_{in}^{-\theta} \tag{II.11}$$

où ζ est une constante déterminée par la masse de la bille m_e, la raideur de contact k et la non-linéarité α. La régression linéaire du modèle (II.9) sur des données de simulations pour divers paramètres du modèle physique donnent $\theta \approx 0.23$ (coefficient de corrélation $R^2 > 0.99$). Comme le montre l'expression de t_0 (II.10), augmenter la masse de la bille (m_e) revient à augmenter t_0, tandis qu'augmenter la dureté de la surface k diminue t_0.

Néanmoins, la vitesse d'impact v_{in} ne peut pas être directement estimée d'après l'observation de la force d'interaction. Ce modèle n'est donc pas adapté à notre étude. Une solution consiste à modéliser la durée de l'impact comme fonction de l'amplitude d'impact en utilisant la même loi mais différents paramètres :

$$t_0 = \zeta' . A^{-\theta'} \tag{II.12}$$

La vitesse d'impact et son amplitude sont évidemment liées : plus la vitesse de l'impact sera grande, plus son amplitude le sera également. Il n'y a cependant pas d'évidence théorique montrant une relation linéaire entre A et v_{in}. On supposera néanmoins que la non-linéarité est suffisamment faible pour que notre modèle soit valide. Notons cependant que du fait de la souplesse de l'objet résonant impacté, Avanzini et Rocchesso (2004) ont montré que la durée de contact est toujours supérieure lorsque l'impacteur est couplé à un tel objet, et qu'elle augmente avec la masse des résonateurs ($m_r^{(i)}$ dans (II.2)) (Avanzini et Rocchesso, 2001b). Néanmoins, comme l'a montré le test subjectif décrit partie B, la force d'interaction f seule est pertinente d'un point de vue perceptif pour l'interaction de roulement, et le but de cette étude n'est pas de reproduire à l'identique le modèle physique dans ses moindres subtilités.

C.5 Estimation des paramètres

Ici on décrira la manière d'estimer les paramètres de notre modèle de signal à partir de forces d'interactions obtenues par le modèle physique présenté en partie A.1.1. Ces paramètres seront ensuite utilisés pour la synthèse. On considère deux catégories : les paramètres qui décrivent la séquence d'impacts et ceux décrivant la forme des impacts.

Paramètres de la séquence d'impacts Comme détaillé en partie C.2, la séquence d'impacts est caractérisée par les valeurs moyennes μ_X, les filtres blanchisseurs inverses $H_X(z)$ et les fonctions de répartition inverses F_X^{-1}, X étant l'amplitude A ou l'intervalle entre deux impacts successifs Δ_T. Les valeurs moyennes peuvent être aisément obtenues par l'estimateur empirique de la moyenne $\mu_X = (1/N)\sum_{n=1}^{N} X^n$. Pour les filtres inverses, on suppose que les deux séries A et Δ_T suivent un modèle Auto-Régressif à Moyenne Ajustée (ARMA) :

$$X(z) \approx H_X(z)\tilde{X}(z) \quad , \quad H_X(z) = \frac{1+\sum_{i=1}^{p} b_i z^{-i}}{1+\sum_{i=1}^{q} a_i z^{-i}} \tag{II.13}$$

La stratégie usuelle permettant d'optimiser les filtres blanchisseurs consiste à minimiser l'énergie du signal résiduel $\tilde{X}$ en fonction des coefficients $\boldsymbol{a} = (a_i)_{i\in[1,q]}$ et $\boldsymbol{b} = (b_i)_{i\in[1,p]}$:

$$(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}) = \underset{\boldsymbol{a},\boldsymbol{b}}{\operatorname{argmin}} \left(\sum_{n=1}^{N} \left[h_X^{-1} * X \right]^2 (t) \right) \tag{II.14}$$

où le symbole ˆ signifie valeur estimée. L'optimisation est effectuée via une stratégie itérative de Gauss-Newton (Ninness *et al.*, 2005) [4].

L'application de cette méthode sur les séries A et Δ_T pour plusieurs simulations du modèle physique nous a montré qu'un seul pôle et un seul zéro ($p = q = 1$) sont suffisants pour blanchir les processus de manière satisfaisante. Ce résultat est appuyé par l'observation des fonctions d'autocorrélation de ${}^c\tilde{A}^n$ et ${}^c\tilde{\Delta}_T^n$ figure II.9. On peut constater que pour toutes les combinaisons de paramètres utilisées, les fonctions d'autocorrélation sont très proches d'une distribution de Dirac, montrant ainsi le succès du blanchiment des signaux.

Les fonctions de répartition F_X sont directement reliées aux densités de probabilités de ${}^c\tilde{X}$. Comme on peut le voir figure II.10, ${}^c\tilde{A}^n$ et ${}^c\tilde{\Delta}_T^n$ suivent approximativement une distribution Gaussienne centrée. Les variances σ_X^2 peuvent être aisément estimées via une méthode des moindres carrés. Le bruit blanc uniforme W et la fonction de répartition inverse peuvent donc être remplacées dans le modèle par un bruit blanc Gaussien centré réduit, multiplié par les écarts types σ_A et σ_{Δ_T}, afin de générer directement les processus ${}^c\tilde{A}^n$ et ${}^c\tilde{\Delta}_T^n$.

Paramètres de la forme des impacts Dans le modèle proposé en C.4, la durée des impacts t_0 dépend de l'amplitude A et cette fonction a deux paramètres : θ' et ζ' (cf equation (II.12)). Lorsque l'on ajuste le modèle de t_0 en fonction de A sur des simulations du modèle physique, on observe que pour certaines réalisations du roulement, on obtient des courbes ajustées aberrantes dues à des points "hors modèle" (ces points peuvent par exemple être dus à une mauvaise estimation de la durée de quelques impacts si deux impacts se recouvrent partiellement).

4. Méthode implémentée dans la fonction ARMAX de MATLAB `http://www.mathworks.fr/fr/help/ident/ref/armax.html`.

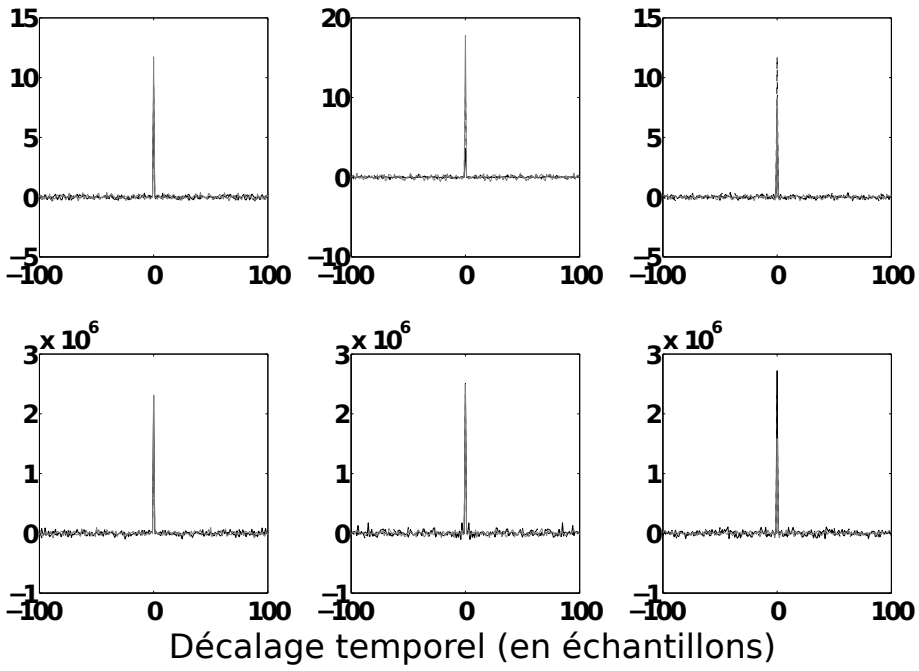

FIGURE II.9 – Autocorrélation des signaux blanchis ${}^{c}\tilde{A}^{n}$ *(ligne supérieure)* and ${}^{c}\tilde{\Delta}_{T}^{n}$ *(ligne inférieure)* pour des filtres blanchisseurs à 1 pôle-1 zéro. Les paramètres et couleurs utilisés sont les mêmes que figure II.5.

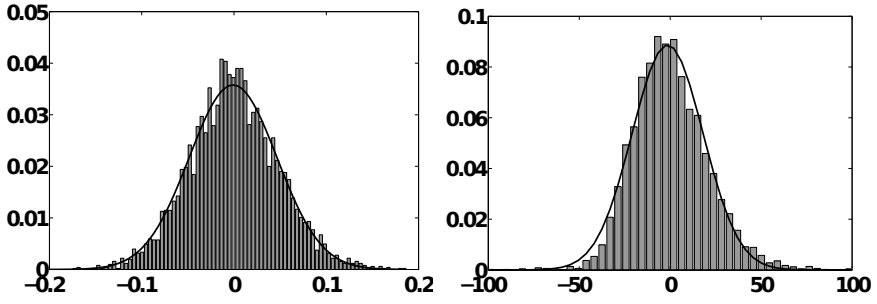

FIGURE II.10 – Barres : densité de probabilité estimée de ${}^{c}\tilde{A}$ *(gauche)* et ${}^{c}\tilde{\Delta}_{T}$ *(droite)*. Courbe noire : estimation par les moindres carrés des lois gaussiennes. Pour rester lisible, seules les estimations avec les paramètres κ=5·10^{9} N.m$^{-\alpha}$.kg^{-1}, μ=1 s.m^{-1} and β=−0.5 sont représentées. Les estimations pour les autres jeux de paramètres sont disponibles en annexe A.

Afin d'éviter ces situations, on estime tout d'abord la loi sur des simulations d'impacts isolés (où on n'a pas de points hors modèle). Cela nous donne des valeurs optimales pour θ' et ζ'. Pour l'interaction de roulement, on conserve θ' et on effectue l'optimisation seulement sur ζ'.

Ici, on donne les résultats pour une simulation avec les paramètres κ=2·10^9 N.m$^{-\alpha}$ .kg^{-1} et μ=1 s.m^{-1}, un rayon de bille de 1 cm, une vitesse de roulement de 20 cm/s et un exposant de la surface fractale β=−1. La première optimisation donne θ'=0.29 (R^2=0.93). L'optimisation pour la courbe de roulement donne R^2=0.70.

D Stratégie de contrôle intuitif

Dans cette partie, on s'attachera à décrire la stratégie de contrôle intuitif du modèle de synthèse de sons de roulement présenté précédemment. On se focalise sur le contrôle de trois attributs perceptifs : la taille et la vitesse de la bille, ainsi que la rugosité de la surface. Bien que la masse de la bille puisse également être considérée, on ne la prendra pas en compte, en supposant qu'il y a un lien direct entre la taille perçue et la masse perçue. Pour ces trois contrôles haut-niveau, les relations entre les paramètres physiques et les paramètres de synthèse seront présentées. La gamme de variation des paramètres de synthèse dans le mapping proposé sera également discutée.

D.1 Contrôle de la taille de la bille

La taille de la bille influence directement la durée de l'impact t_0. D'après l'équation (II.10), on peut voir que t_0 dépend de la masse de la bille m_e, de la raideur de contact k et du terme dissipatif μ, m_e et k dépendants du rayon de la bille. De plus, le paramètre μ influe très peu sur t_0. En rappelant que $\alpha = 3/2$, en exprimant la masse de la bille en fonction de sa densité et en rappelant que $k \propto \sqrt{R}$ (équation (II.4)), on montre que t_0 est directement proportionnel à R (pour des propriétés matérielles constantes de l'excitateur et du résonateur). Ainsi, on peut exprimer (II.11) comme :

$$t_0 \propto R \cdot v_{in}^{-0.23} \tag{II.15}$$

t_0 peut ainsi être exprimé en fonction du rayon de la bille et de l'amplitude de l'impact (cf (II.12)) :

$$t_0 \propto R \cdot A^{-0.29} \tag{II.16}$$

A partir de cette expression, on définit le contrôle haut-niveau pour la taille perçue S, défini comme le rayon normalisé $S \in [0.1, 1]$, influençant directement la durée de l'impact avec le mapping :

$$t_0 = 7.88 \times 10^{-4}\, S\, A^{-0.29}\ ,\ [\mathrm{s}] \tag{II.17}$$

La constante multiplicative provient de l'analyse du modèle, puis a été raffinée arbitrairement afin d'obtenir un mapping satisfaisant. Ainsi plus S est élevé, plus la durée de t_0 est longue et plus la taille perçue est grande. La perception de la taille sera donc majoritairement reliée à des indices spectraux : en effet, augmenter (resp. diminuer) t_0 revient à diminuer (resp. augmenter) la fréquence de coupure d'un filtre passe-bas. Ceci est cohérent avec les études de (Houben, 2002; Houben *et al.*, 2004, 2005), qui montrent que la perception de la taille de billes roulantes est majoritairement due à des indices spectraux (cf partie A.2). De plus le sens de variation de la fréquence de coupure en fonction de la taille est cohérent avec leur étude (voir en particulier la densité spectrale de puissance moyenne d'enregistrements de roulements pour plusieurs tailles de billes sur la figure 7 de (Houben *et al.*, 2004)).

D.2 Contrôle de la vitesse de la bille

Comme vu dans la partie A.2, la perception de la vitesse peut être améliorée grâce par une modulation d'amplitude sur la force d'interaction, i.e. plus la bille est rapide (respectivement lente), plus la modulation d'amplitude est rapide (respectivement lente). On définit ainsi un paramètre haut-niveau pour la vitesse perçue $V \in [0.1, 1]$, agissant sur la fréquence ν_m de la modulation d'amplitude (cf équation (II.8)) avec le mapping suivant :

$$\nu_m = 3\,\frac{V}{S} \tag{II.18}$$

La constante multiplicative est ici arbitraire. La profondeur de modulation est fixée à $m = 0.3$ de sorte que l'effet soit audible mais paraisse toujours naturel.

D.3 Contrôle de la rugosité de la surface

Dans le modèle physique, la rugosité de la surface fractale est contrôlée par l'exposant β. Lorsque l'on change β, on observe de grands changements dans la variation des paramètres de synthèse liés à la séquence d'impacts, i.e. les statistiques des séries A et Δ_T qui sont contrôlées respectivement par (μ_A,σ_A) et (μ_{Δ_T},σ_{Δ_T}), ainsi que par les coefficients des filtres $H_A(z)$ et $H_{\Delta_T}(z)$. Dans la suite, on va donc étudier la variation de ces paramètres de synthèse en fonction de β. On se base pour cela sur des simulations du modèle physique avec les paramètres suivants : κ=$\frac{1}{3}\cdot 10^{10}$ N.m$^{-\alpha}$.kg^{-1}, μ=1 s.m^{-1}, R=1 cm, v=20 cm/s et une taille d'aspérités maximum de 0.1 μm. Le paramètre de rugosité de la surface β varie entre -1.5 (surface lisse) et 0 (surface rugueuse). Le jeu de paramètres (κ, μ) a été choisi d'après les résultats de l'expérience perceptive présentée précédemment, puis raffiné par une séance d'écoute informelle.

Dans la colonne de gauche de la figure II.11, on a représenté les variances estimées σ_A et σ_{Δ_T} en fonction de β. On constate que σ_A ne présente pas de comportement particulier, tandis que σ_{Δ_T} augmente avec β. Les variations de σ_{Δ_T} semblent cohérentes avec la physique : une surface rugueuse comporte peu de corrélation entre les aspérités successives et implique donc une grande variabilité de durée entre les impacts successifs. Dans la colonne du milieu, on a représenté les variations des coefficients des filtres $H_A(z)$ et $H_{\Delta_T}(z)$ en fonction de β. Pour les deux séries A et Δ_T, les coefficients a_1 et b_1 augmentent avec β. Dans la colonne de droite, les moyennes μ_A and μ_{Δ_T} estimées sont représentées en fonction de β. On peut voir que μ_A ne présente pas de comportement particulier, tandis que μ_{Δ_T} augmente avec la rugosité de la surface. Les variations de μ_{Δ_T} semblent cohérentes avec la physique : une surface rugueuse va induire de plus grands intervalles entre impacts.

En modifiant simultanément l'ensemble des paramètres d'après les observations précédentes, on peut reproduire des sons évoquant différentes rugosités de surface, allant d'une surface extrêmement lisse (quasiment comme des crissements, par exemple une bille glissant sur une vitre) à une surface extrêmement chaotique (comme une bille roulant sur du béton endommagé et rencontrant beaucoup de grandes aspérités).

On définit ainsi un contrôle haut-niveau ρ contrôlant la rugosité de la surface et variant de 0 (surface lisse) à 1 (surface rugueuse). Ce contrôle est différent du paramètre β de la surface fractale dans le modèle physique, qui varie entre 0 (surface rugueuse) et $-\infty$ (surface parfaitement lisse). Le paramètre normalisé ρ est plus naturel dans le contexte du contrôle haut-niveau. On suppose un mapping linéaire entre ρ et les paramètres de synthèse. Les valeurs extrêmes ont été choisies d'après l'analyse précédente (cf figure II.11), et sont résumées dans le tableau II.1.

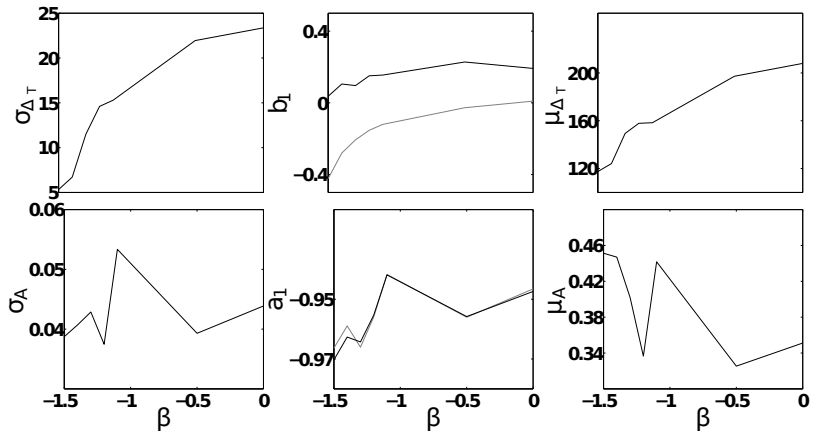

FIGURE II.11 – Variances (colonne de gauche) et valeurs moyennes (colonne de droite) des séries A et Δ_T du modèle de signal proposé en fonction du paramètre de rugosité de la surface β du modèle physique (β=−1.5 : surface lisse. β=0 : surface rugueuse). Colonne du milieu : Coefficients des filtres blanchisseurs inverses de la série d'amplitude A (en noir) et de la série des intervalles de temps Δ_T (en gris). Echelle de μ_{Δ_T} et σ_{Δ_T} en échantillons (fréquence d'échantillonnage : 44100 Hz).

TABLE II.1 – Valeurs extrêmes pour le mapping linéaire entre le contrôle de rugosité de surface ρ et les paramètres de synthèse liés à la série d'impacts (amplitude A et intervalle entre impacts Δ_T).

		$\rho_{min} = 0$	$\rho_{max} = 1$
A	σ_A	0.04	0.04
	a_1	−0.97	−0.93
	b_1	0.07	0.32
	μ_A	0.43	0.27
$\mathbf{\Delta_T}$	σ_{Δ_T}	0.19 (ms)	0.85 (ms)
	a_1	−0.97	−0.93
	b_1	−0.34	0.35
	μ_{Δ_T}	3.1 (ms)	6.4 (ms)

E Evaluation perceptive de la stratégie de contrôle intuitif

Dans cette partie, on présentera un test perceptif permettant d'évaluer le mapping proposé entre les paramètres bas-niveau du modèle de synthèse et les contrôles haut-niveau.

E.1 Sujets

Quatorze participants ont pris part à l'expérience : 9 hommes, 5 femmes, 29 ans en moyenne (écart-type : 9.9 ans). Tous étaient des volontaires du laboratoire et aucun d'entre eux n'avait eu connaissance des stimuli avant l'expérience, ni ne présentait de troubles auditifs.

E.2 Stimuli

Cinq valeurs pour chaque paramètre de contrôle haut-niveau ont été considérées. Les stimuli duraient 3s et ont été normalisés par rapport à leur amplitude maximum. Pour chaque contrôle haut-niveau testé, 15 stimuli (3 jeux de 5 stimuli) ont été synthétisés.

Les 5 valeurs à juger étaient : $S = [0.1, 0.3, 0.5, 0.7, 0.9]$, $V = [0.1, 0.3, 0.5, 0.8, 1]$, et $\rho = [0, 0.25, 0.5, 0.8, 1]$. Pour le jugement de la taille de la bille S, le contrôle de rugosité de surface ρ était fixé à 0.5 et trois vitesses V différentes ont été utilisées (0.1, 0.5 et 1). Pour le jugement de vitesse V, ρ était fixé à 0.75 et trois tailles S différentes ont été utilisées (0.1, 0.3 et 0.5). Pour le jugement de la rugosité de surface ρ, V était fixé à 0.5 et trois tailles S différentes ont été utilisées (0.1, 0.5 et 0.9) ($\sharp$).

E.3 Protocole

Le test s'est déroulé dans une pièce calme, sur un ordinateur portable équipé d'un casque Sennheiser HD-650. Une interface graphique a été spécialement développée sous MAX/MSP. L'évaluation subjective a été effectuée via un test de comparaison par paire (test $A - B$). Les sons ont été évalués par bloc dans lequel seul était testé le contrôle haut-niveau (e.g. pour l'évaluation de la taille, seul le paramètre de taille variait au sein d'un bloc, et les paramètres de vélocité étaient différents dans chacun des 3 blocs, cf E.2). Cette séparation en bloc a été effectuée pour éviter les effets d'interaction entre différents facteurs. En effet, Houben *et al.* (2004) ont montré que lorsque la taille et la vitesse varient entre deux sons à comparer, cela affectait la capacité à discriminer la bille la plus rapide ou la plus grosse. Pour chaque bloc de 5 stimuli, 10 paires correspondant à toutes les combinaisons possibles des 5 stimuli (les paires identiques n'étaient pas testées) étaient présentées au sujet. Ainsi pour chaque contrôle haut-niveau un total de 30 paires (i.e. 3 jeux de 10 paires) ont été présentées.

Chaque contrôle haut-niveau a donc été évalué successivement (3 sessions indépendantes comprenant chacune 3 blocs). Dans chaque session, les blocs ainsi que les paires à l'intérieur de chaque bloc étaient présentées dans un ordre aléatoire. Les participants pouvaient écouter la paire $A - B$ autant de fois que souhaité et choisir une des 3 possibilités suivantes grâce à l'interface graphique (la phrase explicative dépendant du contrôle haut-niveau testé) :

- **icône** $A>B$
  - "*A est plus grand que B*"
  - ou "*A est plus rapide que B*"
  - ou "*A roule sur une surface plus rugueuse que B*"

— **icône** $A{=}B$
  — "*A est aussi grand que B*"
  — ou "*A est aussi rapide que B*"
  — ou "*A roule sur une surface aussi rugueuse que B*"
— **icône** $B{>}A$
  — "*B est plus grand que A*"
  — ou "*B est plus rapide que A*"
  — ou "*B roule sur une surface plus rugueuse que A*"

Chaque session était précédée de 3 essais d'entraînement, afin de familiariser les participants aux sons, à la tâche et à l'interface. Les participants n'avaient pas de retour sur leur réponse, ni durant l'entraînement ni durant le test, et n'étaient pas informés que seul le contrôle haut-niveau testé variait entre les stimuli d'un même bloc. Les 3 sessions étaient effectuées dans un ordre différent pour chaque sujet afin d'éviter un effet de l'ordre de présentation.

E.4 Résultats

Pour chaque contrôle haut-niveau, les données ont été collectées dans une matrice M_p^s de taille 5×5, pour chaque participant p et pour chaque bloc s (donc au total 3 matrices par sujet pour chaque contrôle haut-niveau). Les matrices étaient remplies de la manière suivante : dans le bloc s, si le sujet p répond $A > B$ pour les sons A et B correspondant respectivement à la $i^{\text{ième}}$ et à la $j^{\text{ième}}$ valeur du paramètre de contrôle, l'élément $M_p^s(j,i)$ ($i^{\text{ième}}$ colonne, $j^{\text{ième}}$ ligne) est mis à 1 ; si le sujet répond $B > A$, l'élément $M_p^s(i,j)$ est mis à 1 ; si le sujet répond $A = B$, l'élément $M_p^s(j,i)$ est mis à 0.5 si $i > j$, ou $M_p^s(i,j)$ si $j > i$. La diagonale de chaque matrice est mise à 0.5 car les paires identiques n'ont pas été testées. Un vecteur reflétant une échelle subjective est obtenu en sommant les lignes de la matrice M_p^s, comme proposé par David (1988). Afin de se focaliser sur le contrôle haut-niveau testé, les 3 vecteurs obtenus pour les 3 blocs ont été moyennés. Les scores relatifs ainsi obtenus pour chaque participant p reflètent la vitesse, la taille et la rugosité relative perçues en fonction de la valeur du contrôle haut-niveau. Les résultats moyennés sur les sujets sont représentés figure II.12.

Les résultats ont été analysés par une ANOVA à mesures répétées sur les scores relatifs pour chaque contrôle haut-niveau, avec la valeur du paramètre de contrôle comme facteur intra-sujets (5 niveaux). Pour toutes les analyses statistiques, les effets sont considérés significatifs si $p \leq 0.05$. Lorsque le facteur est significatif, des tests de comparaison post-hoc (Tukey) sur les sons 2 à 2 sont effectués.

L'effet du contrôle haut-niveau S s'avère significatif ($F(4,52){=}98.895$, $p{<}0.001$), et les 5 valeurs de taille testés diffèrent significativement entre elles ($p < 0.001$ pour toutes). Deuxièmement, l'effet du contrôle haut-niveau V s'avère significatif ($F(4,52){=}45.804$, $p{<}0.001$). Les 5 valeurs de vitesses diffèrent significativement entre elles ($p < 0.05$), excepté entre la $2^{\text{ème}}$ et la $3^{\text{ème}}$ ($p = 0.867$). Enfin, l'effet du contrôle haut-niveau ρ s'avère également significatif ($F(4,52){=}47.491$, $p{<}0.001$), et les 5 valeurs de rugosité testées diffèrent significativement entre elles ($p < 0.01$ pour toutes).

E.5 Discussion

Les résultats du test montrent que les contrôles haut-niveau proposés produisent les effets désirés. Bien que les sujets n'étaient pas informés sur les indices auditifs à utiliser, on a pu faire les observations suivantes. Pour le contrôle de la taille, les sujets ont naturellement associé les sons de plus basse fréquence (du fait des durées d'impacts plus longues) et/ou les modulations d'amplitude plus lentes aux plus grandes billes.

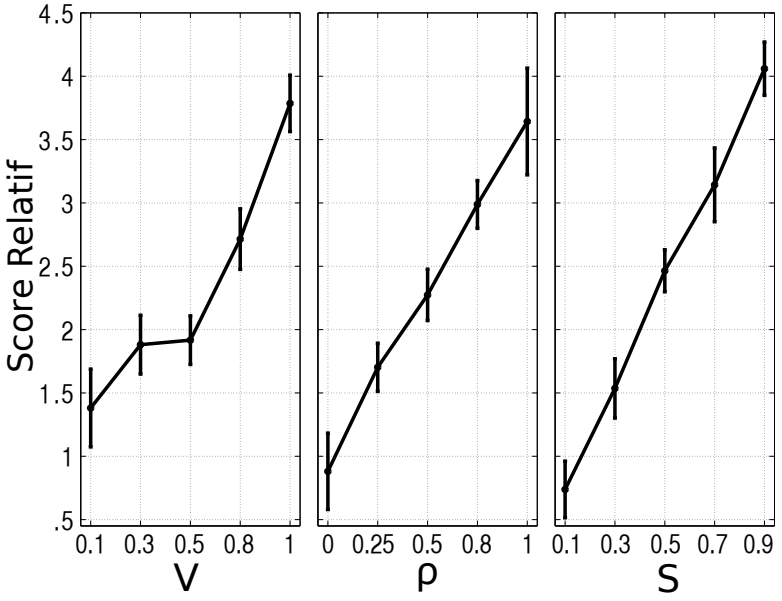

FIGURE II.12 – Scores relatifs (avec intervalles de confiance à 95%) moyennés sur les 14 sujets, pour les 5 valeurs de chaque contrôle haut-niveau : vitesse de la bille (gauche), rugosité de la surface (milieu) et taille de la bille (droite).

Pour le contrôle de la rugosité de la surface, les participants ont rapporté à la suite de l'expérience que les surfaces extrêmement lisses évoquaient une bille qui crissait, tandis que pour les surfaces très rugueuses ils entendaient des impacts séparés car la bille rebondissait beaucoup. Pour le contrôle de la vitesse, les résultats sont également largement satisfaisants et cohérents avec les résultats de Houben (2002) montrant que les modulations d'amplitudes influencent la perception de la vitesse. Néanmoins, on ne trouve pas de différence significative de vitesse entre les pas 2 et 3. Cette confusion est peut-être due à la "discrétisation" linéaire utilisée du contrôle de vitesse V qui ne correspond pas nécessairement à la perception (qui est, elle, souvent non-linéaire). Une expérience de calibration formelle serait nécessaire afin d'ajuster le mapping pour que celui-ci propose des intervalles de variations cohérents avec la perception.

La prise en compte de la variation de l'amplitude des modes propres de la surface résonante en fonction de la position de la bille pourrait peut-être améliorer la perception de la vitesse, car cela produit des effets audibles. Stoelinga (2007) a en effet montré que les sujets étaient capable de discriminer le son d'un bille roulant du centre vers le bord d'une plaque de celui d'une bille roulant du bord vers le centre d'une plaque (cf partie A.2). Ils se sont donc basés sur cet indice sonore (car, dans l'expérience proposée, cet indice acoustique était le seul disponible pour la discrimination). Néanmoins, les auditeurs n'étaient pas capables d'associer à ces variations spectrales une direction précise de la trajectoire de la bille.

F Discussion générale

Dans ce chapitre, on s'est intéressé à la mise en œuvre d'un modèle de synthèse de sons de roulement. Le but de cette étude était d'identifier des caractéristiques du signal responsables de la perception de l'interaction "rouler" puis de proposer un modèle de signal relativement générique (pour pouvoir permettre par la suite d'utiliser ce modèle pour la synthèse d'autres interactions), compatible avec la synthèse sonore temps-réel

et offrant des contrôles haut-niveau.

Dans la première partie de l'étude, un test perceptif a permis d'identifier la morphologie du signal responsable de la perception de l'interaction "rouler". Dans ce but, un modèle physique (Rath et Rocchesso, 2005) a été utilisé pour générer les stimuli. Ce modèle est particulièrement intéressant, car il produit des sons reconnus spontanément comme évoquant un objet roulant (Rath, 2004), et permet d'évaluer la pertinence perceptive des différents sources physiques intervenant dans la production du son. Il est par exemple possible d'évaluer l'influence de l'objet résonant, en simulant des surfaces plus ou moins rigides ou en le supprimant. La suppression du résonateur a permis d'évaluer la force d'interaction non-linéaire seule dans un test perceptif et de montrer que cette force a un grand impact sur la perception de la "sensation de roulement". Le test perceptif a également révélé que convoluer cette force d'interaction à la réponse impulsionnelle d'un objet augmente cette "sensation de roulement". Si on se replace dans le cadre de l'approche écologique de la perception, et en particulier dans le paradigme action-objet (cf chapitre I partie D.3), on peut considérer cette force comme un invariant transformationnel du roulement. Il est probable que le système perceptif humain extraie une information de ce type, quelque soit le matériau du résonateur sur lequel la bille roule (sauf si celui-ci est "trop résonant", i.e. peu amorti ; des tests informels nous ont en effet permis de constater qu'on avait du mal à reconnaître l'interaction lorsque celle-ci est effectuée sur un matériau très résonant). On peut rapprocher ça des suppositions de Stoelinga (2007), qui propose que le système auditif exclut les propriétés liées à l'épaisseur de la plaque sur laquelle le matériau roule pour remonter directement aux informations de taille et de vitesse de la bille (cf partie A.2). Ainsi, le système perceptif peut remonter à l'action en faisant abstraction du matériau. Ceci est corroboré par l'étude de Lemaitre et Heller (2012), qui montre la perception de l'action est robuste quelque soit le matériau, tandis que la perception du matériau est fragile.

On a vu que la structure de l'invariant transformationnel liée au roulement peut être considérée comme une série d'impacts ayant des propriétés statistiques particulières. En se basant sur cette observation, on a proposé un modèle de signal pour cet invariant, grâce à un schéma d'analyse/synthèse complet. Plus particulièrement, une méthode permettant d'estimer et simuler les statistiques de la séquence d'impacts, ainsi qu'un modèle paramétrique d'impact (Van Den Doel *et al.*, 2001) dont la durée en fonction de l'amplitude ont été modélisés. Afin d'améliorer la perception de la vitesse, comme proposé par Houben (2002), une modulation d'amplitude de cette série d'impact a été ajoutée.

Enfin des contrôles haut-niveau (vitesse et taille de la bille, et rugosité de la surface sur laquelle la bille roule) ont été proposés et validés grâce à un test perceptif. De plus, le modèle de synthèse proposé permettant de générer des sons perceptivement satisfaisants (♯) et sa faible complexité permettant une implémentation temps-réel, il est donc tout à fait compatible pour une intégration dans des applications de réalité virtuelle. Les contrôles haut-niveau peuvent être directement reliés à des paramètres décrivant une scène provenant d'un environnement graphique (jeu vidéo) par exemple. Cette application est actuellement testée dans le cadre du projet ANR PHYSIS[5] qui s'intéresse à l'analyse, transformation et synthèse sonore temps-réel pour les environnements virtuels interactifs et la réalité augmentée.

Comme proposé par plusieurs auteurs (Stoelinga *et al.*, 2003; Murphy *et al.*, 2011), les sons synthétisés peuvent être rendus plus réalistes en prenant en compte des effets supplémentaires, comme la variation de l'amplitude des modes au cours du temps de

5. `http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2[CODE]=ANR-12-CORD-0006`

la structure résonante qui dépend de la position de la bille sur la surface (pour une bille roulant sur une surface de dimensions finies). Actuellement cet effet est pris en compte dans le synthétiseur en considérant plusieurs filtrages en peigne dont les fréquences fondamentales dépendent de la distance de la bille par rapport aux bords de l'objet résonant (méthode des sources images, comme utilisée en acoustique des salles (Allen et Berkley, 1978)). Cependant, afin d'obtenir un effet perceptif encore plus convaincant, des phénomènes plus complexes supplémentaires doivent sûrement être pris en compte, comme la dispersion, qui peut être simulée par un filtrage en peigne inharmonique (Stoelinga *et al.*, 2003). On peut constater que l'effet réel à prendre en compte est effectivement plus complexe en injectant l'excitation proposée dans un schéma aux différences finies d'un modèle physique de plaque où cet effet est naturellement pris en compte (Bilbao, 2009) (♯).

Enfin, on pourrait imaginer calibrer ce modèle par des méthodes d'analyse/synthèse sur des enregistrements comme proposé par Lagrange *et al.* (2010) et par Lee *et al.* (2010). On peut même penser à utiliser le modèle développé dans ce chapitre afin de pré-informer le modèle d'analyse, par rapport à la forme des impacts et les relations statistiques entre les impacts successifs. Pour l'analyse, une approche de type "matching-pursuit" serait également intéressante à étudier (Mallat et Zhang, 1993; Daudet, 2006).

Dans le chapitre suivant on s'intéressera à d'autres interactions continues : "frotter" et "gratter". Le modèle de signal proposé dans ce chapitre étant relativement générique, on l'adaptera à ces autres interactions. Une stratégie générale de contrôle intuitif, permettant notamment d'effectuer des transitions continues entre les différentes interactions sera proposée.

Chapitre III

Extension du Modèle à d'Autres Interactions et Stratégie de Contrôle du Synthétiseur

Sommaire

A	**Sons de friction : état de l'art**	**52**
B	**Etude perceptive des interactions "frotter" et "gratter"**	**54**
C	**Contrôle perceptif des actions "frotter" et "gratter"**	**60**
D	**Modèle générique de sons d'interactions continues**	**65**
E	**Perspectives d'élargissement de l'espace sonore des interactions : vers la friction non-linéaire**	**69**
F	**Discussion générale**	**76**

Ce chapitre est basé sur les travaux présentés dans l'article de journal (Conan, Thoret, Aramaki, Derrien, Gondre, Ystad, et Kronland-Martinet, 2014b), ainsi que dans les actes de conférences (Conan, Aramaki, Kronland-Martinet, Thoret, et Ystad, 2012; Conan, Thoret, Aramaki, Derrien, Gondre, Kronland-Martinet, et Ystad, 2013).

Dans ce chapitre, nous nous proposons d'élargir notre investigation et d'étendre le modèle de synthèse décrit précédemment à d'autres interactions continues telles que "frotter" et "gratter". Le modèle est en effet assez générique pour envisager d'être adapté à ces nouvelles interactions. Après avoir passé en revue la littérature sur les modèles de synthèse de sons de friction (linéaire et non-linéaire), on présentera les résultats d'un test de catégorisation forcée sur des sons de friction enregistrés. Les résultats du test nous permettront d'émettre une hypothèse sur les différences morphologiques du signal sonore responsables de la distinction de ces deux interactions. L'hypothèse sera validée en proposant la modification d'un modèle de synthèse de sons de frottements existant, permettant ainsi de passer continument de "frotter" à "gratter", puis en utilisant les sons générés par ce modèle dans un test perceptif. Enfin, un modèle générique pour les trois interactions "frotter", "rouler" et "gratter" et une stratégie de contrôle associée, qui permet des transitions perceptives continues entre ces trois interactions, seront proposés. Des perspectives d'élargissement de cet "espace sonore des interactions", basées sur des travaux préliminaires, seront également présentées en fin de chapitre.

A Sons de friction : état de l'art

Les sons de friction résultent du glissement entre deux corps solides. Ces sons sont d'une très grande diversité et vont du son du violon à celui des freins d'automobile, en passant par la stridulation de certains insectes, la main qui frotte sur le canapé ou encore le crissement d'une craie sur un tableau (voir (Akay, 2002) pour une revue détaillée sur les sons de friction). Dans cette partie, on passera tout d'abord en revue l'état de l'art sur la synthèse de sons de friction, puis on verra quelques études perceptives ayant trait à ces sons. On choisit de séparer les sons de friction en deux grandes classes, qui dépendent majoritairement du niveau de pression exercé entre les deux solides en contact (Akay, 2002) : les sons de *friction linéaire* et les sons de *friction non-linéaire*. Les sons de friction linéaire sont générés lorsque les deux solides en interaction sont très peu couplés, et que la source principale de bruit est due aux petits chocs entre les aspérités microscopiques des deux surfaces. Ben Abdelounis *et al.* (2010) ont montré que c'était le cas pour des solides soumis à des charges en pression faibles. C'est par exemple le cas pour des sons de frottement, grattement ou glissement. On a dans ce cas à faire à des sons plutôt bruités. Les sons de friction non-linéaire sont eux produits lorsque les solides en interaction sont fortement couplés, par exemple lorsque qu'une forte pression est imposée sur les solides en contact. Dans ce cas, la source principale de bruit n'est plus due à des micro-chocs comme dans le cas linéaire, mais à des instabilités mécaniques, qui peuvent émerger même avec des surfaces parfaitement lisses. C'est le cas de sons comme les grincements de portes ou les crissements de freins par exemple.

A.1 Synthèse de sons de friction

A.1.1 Synthèse de sons de friction linéaire

Basés sur les travaux de Gaver (1993a), Van Den Doel *et al.* (2001) ont proposé un modèle phénoménologique permettant de synthétiser des sons de friction linéaire, tel que le son produit lorsque l'on frotte un mur ou lorsqu'un objet glisse le long d'un plan incliné. Ce modèle suppose (de manière cohérente avec les travaux postérieurs de Ben Abdelounis *et al.* (2010)) que le son généré lorsque l'on frotte deux objets solides l'un contre l'autre est le résultat de micro-contacts entre les aspérités des 2 surfaces en contact. Les auteurs proposent donc de considérer que les surfaces sont imparfaites à une échelle microscopique, et qu'on peut donc assimiler leur profil (vertical) à un bruit. Ce profil de surface est simulé par un bruit fractal, qui a été montré comme décrivant bien les surfaces réelles (Sayles et Thomas, 1978; Zahouani *et al.*, 1998). Ce bruit vient ensuite être "lu" à une vitesse variable proportionnelle à la vitesse relative entre les 2 solides en interaction. Le signal obtenu est utilisé pour exciter l'objet résonant qui est décrit par un modèle de synthèse modale (Adrien, 1991). On utilisera une version modifiée de ce modèle plus loin dans ce chapitre. Pour des applications de la synthèse sonore aux jeux vidéos, Ren *et al.* (2010) utilisent une approche similaire, mais en décrivant les surfaces selon plusieurs niveaux de détails : un niveau macroscopique qui correspond à la géométrie des objets en contact, un niveau mésoscopique qui correspond à l'échelle du pixel et représente des irrégularités sur les objets, et enfin un niveau microscopique qui est le niveau représenté par le bruit fractal dans le modèle de Van Den Doel *et al.* (2001). Le modèle d'analyse/synthèse proposé par Lagrange *et al.* (2010) et présenté dans le chapitre précédent (cf partie A.1.3), dû à sa généricité, permet également l'analyse/synthèse de sons de friction. Enfin, Ben Abdelounis *et al.* (2011) ont quant à eux proposé une modélisation en éléments finis de surfaces rugueuses en contact. Leur mo-

dèle n'avait cependant pas pour but la synthèse sonore, mais la prédiction du niveau sonore en fonction des propriétés mécaniques des surfaces en contact.

A.1.2 Synthèse de sons de friction non-linéaire

La friction non-linéaire a lieu lorsque les deux objets en contact interagissent fortement entre eux, c'est-à-dire qu'un fort couplage a lieu entre les deux : les vibrations de l'objet 1 modifient le comportement de l'objet 2, tandis que celles de l'objet 2 modifient le comportement de l'objet 1 ! Ces sons ont été étudiés majoritairement dans le cadre de la reproduction de sons d'instruments de musique, tels le violon ou le bol Tibétain (Serafin, 2004), mais permettent également de reproduire des sons non-musicaux, tels les crissements de freins ou d'essuie-glace (Elmaian, 2013; Elmaian *et al.*, 2014), ou encore les grincements de portes (Avanzini *et al.*, 2005). Contrairement aux sons de friction linéaire vus précédemment, les sons de friction non-linéaire sont plutôt tonaux. L'approche pour la synthèse sonore consiste le plus souvent à discrétiser les équations physiques du modèle. Ces modèles permettent la production d'une large variété de sons d'excellente qualité. Des violonistes expérimentés ont par exemple jugé le modèle proposé par Serafin (2004) tout à fait jouable grâce à un archet électronique et sonnant naturellement (Young et Serafin, 2003), ce qui montre la pertinence de la modélisation physique. Ils peuvent cependant être difficiles à contrôler dans certains cas : du fait de leur non-linéarité, on observe des phénomènes de bifurcation entre différents régimes de vibration des objets en interaction selon les valeurs des paramètres, phénomènes difficilement prédictibles et contrôlables. Pour un utilisateur non expérimenté sur ce type de modèle souhaitant par exemple obtenir le son d'une porte qui grince ou d'un verre qui chante, il sera ardu d'ajuster tous les paramètres physiques du modèle.

Afin d'aider au contrôle de la synthèse de sons de friction non-linéaire, Thoret *et al.* (2013) ont proposé un modèle source-filtre, en identifiant les effets des non-linéarités pour différents types de régimes (couinement, grincement de porte, auto-oscillation d'un verre qui chante...) et en les intégrant au niveau du terme source. Ce terme source est modélisé comme une somme harmonique de sinusoïdes, dont on vient faire varier fréquence fondamentale et amplitude d'après des observations inspirées de la physique ainsi que des observations sur des signaux enregistrés. Ainsi, le couplage entre les deux objets interagissant est directement pris en compte dans le terme source, et permet ainsi une plus grande contrôlabilité du modèle en proposant à l'utilisateur d'effectuer lui-même le mapping entre la vitesse et la pression du geste avec les différents régimes vibratoires (couinement, auto-oscillations etc). On reviendra sur ce modèle en fin de chapitre, et ce modèle sera également utilisé dans le chapitre IV. La prise en compte de la non-linéarité dans le terme source d'un modèle source-filtre pour la synthèse de sons de friction non-linéaire a également été exploitée auparavant par Serafin *et al.* (2002) afin de synthétiser des sons de scie musicale et de glassharmonica. Dans le cas des modèles source-filtre, le filtre permet de prendre en compte les modes de l'objet résonant.

A.2 Perception et utilisation des sons de friction

Divers auteurs ont considéré les sons de friction dans des études perceptives. Avanzini *et al.* (2004) ont par exemple utilisé des sons de friction dans une tâche audio-visuelle interactive, où le modèle proposé par Avanzini *et al.* (2005) était contrôlé en temps-réel par le geste de l'utilisateur. Le but de l'expérience était de déterminer si la modalité auditive peut substituer le retour haptique dans une tâche d'évaluation des propriétés inertielles d'un objet virtuel manipulé par l'utilisateur. Le même modèle de

synthèse de sons de friction a également été utilisé par Serafin *et al.* (2013) afin d'évaluer l'apport d'un retour sonore sur une tâche d'équilibre. Danna *et al.* (2013) ont utilisé des sons de frottements de synthèse (Van Den Doel *et al.*, 2001) afin de sonifier en temps-réel des variables pertinentes de l'écriture manuscrite (pression instantanée du stylo sur le papier et vitesse tangentielle instantanée). Le but de leur expérience était d'évaluer l'apport d'un retour sonore sur l'activité d'écriture pour la réhabilitation de la dysgraphie, qui est une pathologie affectant les capacités motrices fines provoquant ainsi une extrême difficulté à accomplir la tâche d'écriture, ces troubles intervenant indépendamment des capacités à lire et n'étant pas liés à un un trouble psychologique (Hamstra-Bletz et Blöte, 1993; Smits-Engelsman *et al.*, 2001).

D'un point de vue plus fondamental, Thoret *et al.* (2014) ont étudié la perception auditive des mouvements biologiques. Ici encore, le modèle de synthèse de sons de frottements proposé par Van Den Doel *et al.* (2001) est utilisé pour générer les stimuli, ainsi que des enregistrements d'écriture manuscrite. Les résultats de cette étude montrent que, à l'instar d'autres modalités (e.g. en vision (Viviani et Stucchi, 1989, 1992) ou en proprioception (Viviani *et al.*, 1997)), la perception auditive des mouvements biologiques est contrainte par la loi dite en "1/3" (Lacquaniti *et al.*, 1983).

Enfin, d'un point de vue plus artistique, Heinrichs et McPherson (2014) ont étudié à travers des tests perceptifs l'influence de différents mapping des paramètres de synthèse d'un modèle de sons de friction non-linéaire (plus particulièrement de la synthèse de sons de grincements de portes) sur l'expressivité, dans le but de permettre à l'utilisateur un meilleur contrôle.

A travers ces exemples, on peut voir que les sons de friction et en particulier les modèles de synthèse contrôlables permettant de les générer, semblent d'un grand intérêt tant pour des applications artistiques que de réhabilitation motrice, mais sont également utiles pour étudier le fonctionnement du système auditif d'un point de vue plus fondamental. Ils ont souvent été utilisés pour transmettre une information liée au mouvement (soit pour remonter à des propriétés du traitement de l'information auditive (Thoret *et al.*, 2014), soit dans une tâche de guidage (Avanzini *et al.*, 2004; Serafin *et al.*, 2013; Danna *et al.*, 2013)).

Dans la littérature, il n'existe pas à notre connaissance d'études ayant différencié les interactions "frotter" et "gratter", que ce soit du point de vue de la perception ou du point de vue de la synthèse. Le but de la partie qui suit est d'étudier ces différences.

B Etude perceptive des interactions "frotter" et "gratter"

Dans la littérature en synthèse sonore, les interactions "frotter" et "gratter" ne sont jamais distinguées à notre connaissance. Ici, on montrera dans un premier temps qu'il existe une différence perceptive entre ces deux interactions grâce à un test perceptif sur des sons de friction enregistrés. Puis, en regard des résultats du test, on effectuera une analyse qualitative sur les signaux. Cette analyse qualitative nous permettra de supposer une différence entre les morphologies des signaux évoquant l'une ou l'autre de ces interactions.

B.1 Catégorisation perceptive de sons de friction enregistrés

Comme explicité précédemment, le but de cette partie est de mettre en évidence la capacité de distinguer des sons associés à l'action "frotter" de sons associés à l'action "gratter". D'un point de vue phénoménologique, on suppose que la surface de contact

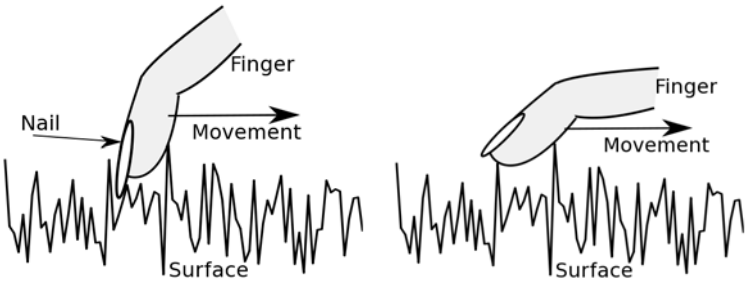

FIGURE III.1 – Un ongle qui gratte une surface (gauche), un doigt qui frotte une surface (droite). L'axe vertical représente la hauteur des aspérités de la surface (exagérée pour la clarté de la figure) ; l'axe horizontal représente la direction dans laquelle le doigt se déplace.

est plus étendue pour l'interaction "frotter" (interaction avec la pulpe des doigts par exemple) que pour l'interaction "gratter" (interaction avec les ongles par exemple) pour laquelle l'interacteur "voit mieux" la surface. En d'autres termes, on suppose que les sons associés au "frottement" sont issus d'un balayage superficiel de la surface par l'interacteur, tandis que ceux associés au "grattement" sont issus d'un balayage en profondeur de la surface (voir figure III.1 pour une schématisation du concept). Le test perceptif présenté par la suite permettra d'évaluer à travers un protocole de catégorisation forcée s'il existe effectivement une discrimination cohérente à travers les sujets de sons d'interactions entre ces deux catégories.

B.1.1 Sujets

Quatorze sujets ont participé à l'expérience : 4 femmes, 10 hommes, 30 ans en moyenne (écart-type : 12 ans). Aucun d'entre eux n'avait eu connaissance des stimuli avant l'expérience ni ne présentait de troubles auditifs.

B.1.2 Stimuli

Vingt enregistrements monophoniques de sons d'interactions sur 10 surfaces différentes ont été effectués (fréquence d'échantillonnage de 44100 Hz). Pour tester l'hypothèse précédente sur la différence de nature entre les deux interactions "frotter" et "gratter", 2 sons ont été enregistrés sur chacune des 10 surfaces testées (entre autres : des murs lisses et rugueux, différents types de papier de verre, un tapis...), l'un en interagissant avec la pulpe des doigts, l'autre avec les ongles. Pour chaque enregistrement, l'expérimentateur conservait au mieux la vitesse de son geste (♯).

B.1.3 Protocole

Le test s'est déroulé dans une pièce calme, sur un ordinateur portable muni d'un casque Sennheiser HD-650. Les sujets étaient informés qu'ils allaient devoir classer les sons dans une des 2 catégories, "frottement" ou "grattement", grâce à l'interface développée spécialement sous MAX/MSP (voir figure III.2). Avant le début du test, les 20 stimuli étaient joués une fois. Durant la tâche de catégorisation, aucune contrainte de temps n'était imposée et les sujets pouvaient écouter les sons autant de fois que désiré.

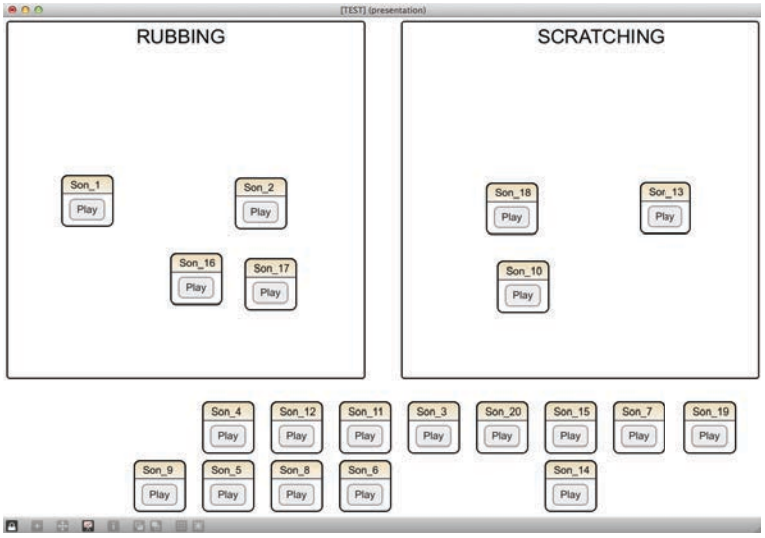

FIGURE III.2 – Interface pour le test de catégorisation forcée des sons de friction enregistrés.

B.1.4 Résultats

Pour chaque sujet, la valeur (arbitraire) 1 est affectée aux sons associés à la catégorie "gratter", et la valeur 0 aux sons associés à la catégorie "frotter". Pour chaque son, ces valeurs sont moyennées sur les sujets, permettant ainsi d'obtenir un pourcentage de classification de chaque son.

Un test d'adéquation du χ^2 (par rapport à l'hypothèse nulle qui est que le son n'a pas été plus associé à une catégorie qu'à une autre, i.e. que le son a été statistiquement autant associé à frotter qu'à gratter) est effectué pour chaque son afin d'évaluer s'il a été ou non associé à une catégorie. Quatorze sons ont été associés significativement à l'une ou l'autre des catégories (un risque de 5% a été choisi), en particulier 8 à l'interaction frotter et 6 à l'interaction gratter (cf figure III.3). Le stimulus associé le moins significativement est le 19 à 78.6% ($\chi^2_{df=1} = 4.57\,;p = 0.032$), les stimuli associés à 100% sont les numéros 3, 5, 7, 9, 11, 12, 16, 17 et 20 ($\chi^2_{df=1} = 14\,;p < 0.001$).

Les 6 sons associés à l'interaction gratter (numérotés 8, 11, 14, 16, 17 et 19) sont bien des sons pour lesquelles les enregistrements ont été effectués en interagissant avec les ongles. Les 8 sons associés à l'interaction frotter (numérotés 3, 4, 5, 7, 9, 12, 15 et 20) sont bien des sons pour lesquelles les enregistrements ont été effectués en interagissant avec la pulpe des doigts, hormis le son 4. On constate donc que l'hypothèse émise lors de l'enregistrement des stimuli semble être valide.

B.1.5 Discussion

Deux ensembles de sons peuvent être déterminés : les sons associés significativement à l'une ou l'autre des 2 interactions, et ceux qui mènent à une perception plus ambigüe. Le fort taux d'association de certains sons dans chacune des 2 catégories nous permet de conclure que, d'un point de vue de la perception auditive, il existe bien une différence sémiotique entre les interactions "frotter" et "gratter". Cette distinction est globalement cohérente avec l'hypothèse faite lors de l'enregistrement, i.e. que les sons produits lorsque l'on interagit avec les ongles sont catégorisés comme "grattement", tandis que ceux produits lorsque l'on interagit avec la pulpe des doigts sont catégo-

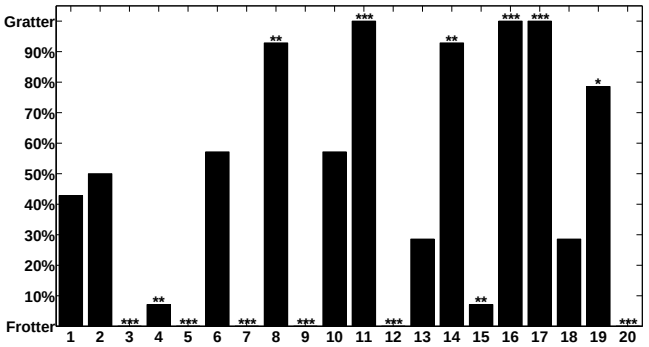

FIGURE III.3 – Résultats du test subjectif sur les sons enregistrés. Le numéro de chaque son est représenté sur l'axe horizontal, l'axe vertical représente le pourcentage d'association à l'interaction gratter. Une étoile signifie que le son a été associé significativement à l'une des deux interactions avec $p < 0.05$, deux étoiles avec $p < 0.01$, trois étoiles avec $p < 0.001$.

risés comme "frottement". L'ambiguité observée pour certains sons (i.e. le fait qu'ils n'aient pas été associés unanimement à l'une ou l'autre des catégories) montre que cette distinction n'est pas totalement catégorielle.

B.2 Analyse qualitative des sons catégorisés

Les spectrogrammes des sons significativement associés aux interactions gratter et frotter sont représentés respectivement sur les figures III.4 et III.5. Une première observation de ces représentations nous permet de constater que les sons associés au frottement présentent plutôt des évolutions lentes par rapport à leur durée dans le domaine temps-fréquence. A l'inverse, les sons associés au grattement présentent majoritairement des évolutions rapides dans le domaine temps-fréquence. On suppose que les variations lentes sont plutôt dues à des caractéristiques "macroscopiques" de l'interaction, telle la variation de vitesse et/ou de pression du geste ayant produit les sons. Ces évolutions lentes, qu'on peut voir qualitativement au cours du temps comme une dilatation puis une contraction de l'énergie, se retrouvent également sur certains sons associés au grattement. Cette observation sur des sons enregistrés est cohérente avec le modèle phénoménologique proposé par Van Den Doel *et al.* (2001) pour la synthèse de sons de friction, qui considère un bruit lu à une vitesse proportionnelle à la vitesse relative entre les deux solides (cf partie A.1.1). En effet, la lecture plus rapide (resp. plus lente) d'un signal dans le domaine temporel correspond à une dilatation (resp. une contraction) dans le domaine fréquentiel.

On suppose que les évolutions rapides qu'on peut distinguer sur les représentations temps-fréquence sont dues à des micro-événements lors de l'interaction. Sous chaque représentation temps-fréquence, on a représenté la fonction de détection d'attaques calculée à partir de transformées de Fourier à court-terme du signal (Bello *et al.*, 2004). Cet indicateur utilisé pour extraire des informations sur des signaux musicaux permet de combiner les informations de variations de phase et d'énergie afin de rendre compte de changements brusques dans le signal, synonymes d'attaques (percussives ou harmoniques). Son fonctionnement est décrit en annexe B. Ce type de descripteur permet d'effecteur des tâches de *Music Information Retrieval* comme la détection de tempo et est donc adapté à des signaux musicaux, qui contiennent en général un petit nombre d'at-

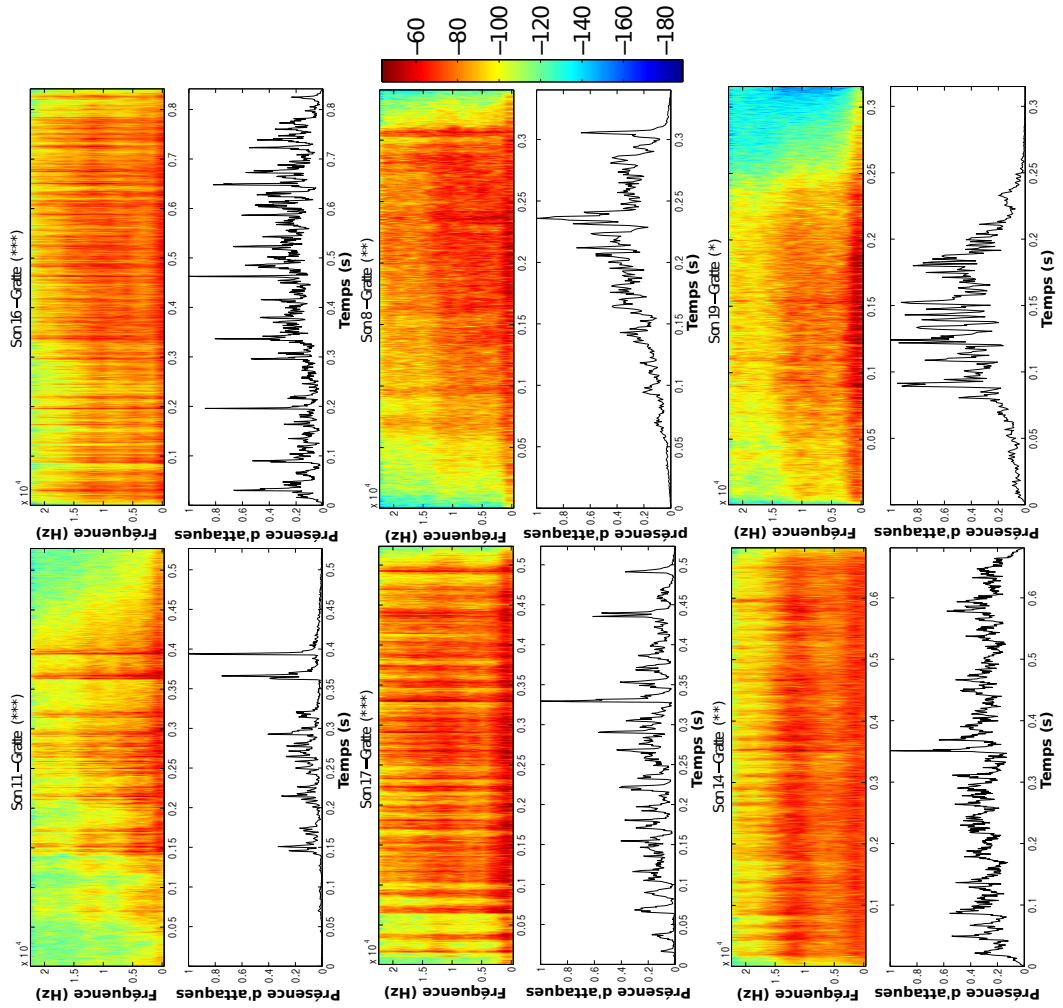

FIGURE III.4 – Représentations temps-fréquence (en dB) et fonction de détection d'attaques pour les sons associés à l'interaction gratter. Paramètres utilisés pour la représentation temps-fréquence : Transformées de Fourier discrètes sur des fenêtres de Blackmann-Harris de 64 points (fréquence d'échantillonnage des signaux de 44100 Hz), avec un recouvrement temporel de 90 %. Paramètres utilisés pour la fonction de détection d'attaques identiques, excepté l'utilisation d'un recouvrement des fenêtres de 50 %. La petite taille de fenêtre a été choisie afin de bien mettre en valeur les variations temporelles rapides. Dans le titre de chaque son, la valeur p est reportée (1, 2 ou 3 étoiles comme décrit dans la figure III.3).

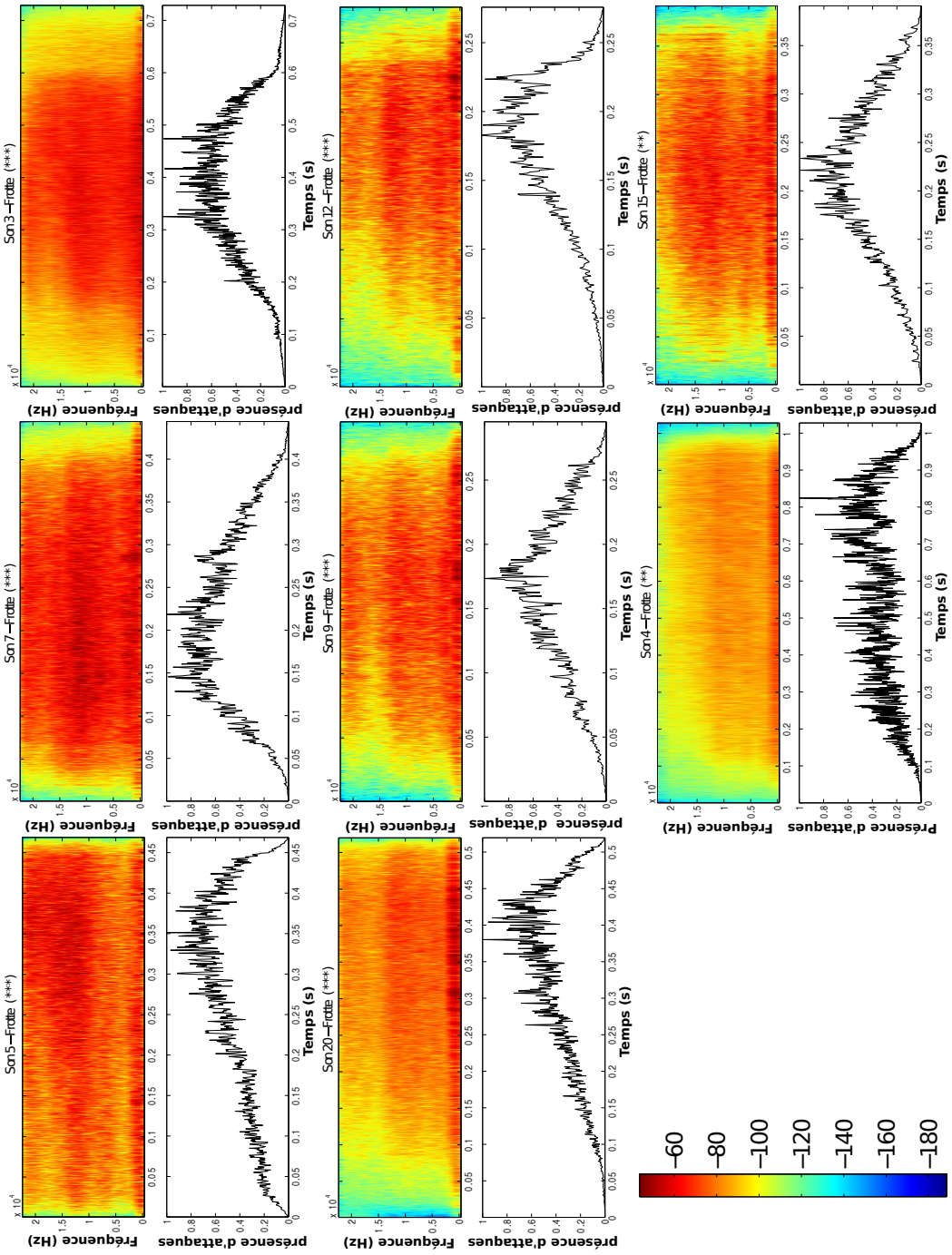

FIGURE III.5 – Représentations temps-fréquence (en dB) et fonction de détection d’attaques pour les sons associés à l’interaction frotter. Paramètres identiques à ceux de la figure III.4.

taques par seconde. Ici, on suppose que les sons de frottement ou de grattement sont dus à une succession de micro-impacts, comme suggéré par plusieures études (Akay, 2002; Ben Abdelounis *et al.*, 2010, 2011). Il est clair que si on fait cette hypothèse, vu la nature très bruitée des signaux (en particulier pour l'interaction frotter où on constate que la densité spectrale de puissance évolue lentement et ne laisse pas vraiment apparaître d'événements singuliers), on ne pourra pas détecter chaque micro-impact (ce qui pourrait permettre entre autre de remonter à l'excitation et ainsi de séparer les contributions "action" et "objet"). On peut cependant constater que parmi les fluctuations de la fonction de détection d'attaques on observe des événements singuliers, caractérisés par des pics abrupts, pour les sons associés à l'interaction gratter. Ces pics importants sont d'autant plus visibles pour les sons ayant été associés unanimement au grattement (numéro 11, 16 et 17). Au contraire, pour l'interaction frotter, on observe plutôt une fluctuation rapide de faible amplitude de la fonction au cours du temps autour de l'évolution lente de sa valeur moyenne.

On suppose donc qu'un indice acoustique responsable de la distinction entre ces deux interactions est cette présence d'impacts parcimonieux dans le cas de l'interaction gratter. On suppose de plus que pour le frottement, on a une densité temporelle d'impacts plus importante, qui mène ainsi a une densité spectrale évoluant de manière plus lente au cours du temps.

Dans la partie suivante, on proposera un modèle de synthèse de sons de friction prenant en compte à la fois cette micro-structure du signal permettant la distinction frotter/gratter, ainsi que les variations macroscopiques observées, qu'on suppose liées à la vitesse du geste ayant produit le son. Ce contrôle perceptif des interactions "frotter" et "gratter", basé sur ces observations, sera validé par la synthèse grâce à un test perceptif.

C Contrôle perceptif des actions "frotter" et "gratter"

Comme présenté en partie A.1.1, Van Den Doel *et al.* (2001) ont proposé pour leur modèle de synthèse de sons de friction linéaire la lecture d'un bruit fractal, qui représente la surface rugueuse, à une vitesse variable proportionnelle à la vitesse entre les 2 solides en contact. Si on considère le cas de l'interaction sur une surface fixe, la vitesse considérée est la vitesse du geste effectuant le frottement (cf figure III.1).

Plutôt que de considérer la lecture d'une table de bruit à vitesse variable, qui nécessite de prendre des précautions lors de l'interpolation du profil, on choisit la méthode suivante. Tout d'abord, on représente la surface rugueuse par un bruit blanc filtré passe-bas. En pratique, on utilise un filtre *biquad* (i.e. ayant 2 pôles et 2 zéros) passe-bas non résonant, qui d'après des tests informels donne un résultat perceptivement satisfaisant, et dont la fonction de transfert est :

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{III.1}$$

avec :

$$\begin{cases} b_0 = Gc^2 \\ b_1 = 2b_0 \\ b_2 = b_0 \\ a_1 = 2G(c^2 - 1) \\ a_2 = G\left(1 - \frac{c}{Q} + c^2\right) \\ c = \tan\left(\frac{\pi f_c}{fs}\right) \\ G = \frac{1}{1+\frac{c}{Q}+c^2} \end{cases} \tag{III.2}$$

où f_c est la fréquence de coupure du filtre, Q le facteur de qualité (fixé à $\frac{1}{\sqrt{2}}$ pour avoir un filtre non-résonant) et G permet de normaliser le gain du filtre quelque soit la fréquence.

La lecture d'un bruit stationnaire à vitesse variable revient approximativement à faire varier la fréquence de coupure du filtre proportionnellement à la vitesse. En effet, dans le cas continu, cela revient à considérer l'application qui à $x(t)$ associe $x(\alpha t)$, avec $\alpha > 1$ si on lit le signal plus rapidement, et $\alpha < 1$ si on le lit plus lentement. Dans le domaine de Fourier, cela revient à considérer l'application qui à $X(\omega)$ associe $\frac{1}{\alpha}X(\frac{\omega}{\alpha})$, où X est la transformée de Fourier de x et ω la pulsation. On voit donc que lorsqu'on lit plus lentement (resp. plus rapidement) le profil de surface, cela revient à contracter (resp. dilater) le spectre. Ainsi, faire varier la fréquence de coupure du filtre passe-bas proportionnellement à la vitesse du geste revient approximativement à effectuer ces dilatations-contractions. Pour être exactement dans le cas de la lecture à vitesse variable, il faudrait également faire varier la pente du filtre. Néanmoins, cette approximation donne des résultats perceptivement satisfaisants. Cette façon de voir dans le domaine fréquentiel est cohérente avec l'étude de Ye (2004) qui montre qu'en augmentant la vitesse d'un frottement, plus de modes haute fréquence des objets en interaction sont excités.

C.1 Description du contrôle

Comme on l'a vu précédemment, les signaux des sons évoquant l'interaction "gratter" présentent des différences notables par rapport aux sons évoquant l'interaction "frotter". On suppose qu'une différence majeure est la densité temporelle d'impacts : une surface frottée produit une grande densité temporelle d'impacts dans le signal sonore, tandis qu'une surface grattée en produit moins. Le point de vue phénoménologique d'une personne interagissant avec sa main sur une surface rugueuse a été exposé en partie B.1.

Dans le modèle de synthèse, ces différences peuvent être contrôlées en générant des suites d'impulsions d'amplitude différentes plus ou moins espacées dans le temps, qu'on viendra ensuite filtrer par le filtre précédemment présenté. Il existe différentes manières de modéliser ce type de bruit, voir par exemple les algorithmes proposés par Valimaki *et al.* (2013). Dans cet article, différents algorithmes permettant de générer des bruits blancs "parcimonieux" (i.e. comprenant beaucoup de valeurs nulles) sont comparés. Le but de l'étude est d'évaluer à travers un test perceptif si, pour différents paramètres de chacun de ces modèles, certains produisent des sons jugés plus "doux" qu'un bruit blanc Gaussien classique. Ici, on choisit de contrôler la densité et l'amplitude de la série d'impacts de la manière suivante : à chaque instant discret, l'impact est le résultat d'un processus de Bernoulli de paramètre $d \in [0,1]$, un impact étant produit uniquement si le résultat du tirage est 1. L'amplitude des impacts suit une loi uniforme

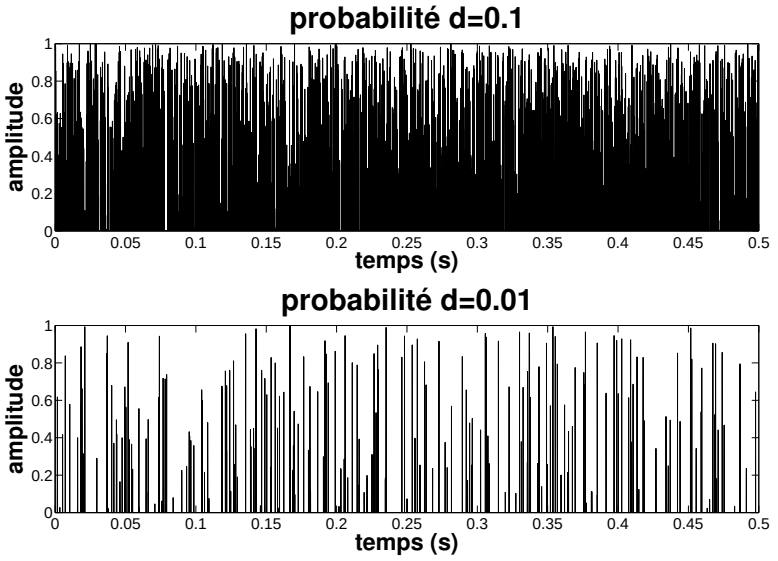

FIGURE III.6 – Haut : Grande densité temporelle d'impacts (d grand), correspondant à l'interaction frotter. Faible densité d'impacts (d faible), correspondant à l'interaction gratter.

sur l'intervalle [0,1]. Sur la figure III.6, deux bruits synthétisés en faisant varier la densité d sont représentés.

C.2 Validation par synthèse

Afin de valider le contrôle proposé permettant de distinguer l'interaction "frotter" de l'interaction "gratter", un test perceptif sur des sons de synthèse a été mis en place. Le but est d'étudier l'influence du paramètre d, contrôlant la densité temporelle d'impacts, sur la perception des interactions "frotter" ou "gratter".

C.2.1 Sujets

Trente-cinq participants ont passé l'expérience : 9 femmes, 26 hommes, moyenne d'âge de 30 ans (écart type : 12 ans). Six d'entre eux avaient déjà participé à l'expérience précédente sur les sons enregistrés. Aucun d'entre eux n'avait eu connaissance des stimuli avant l'expérience ni ne présentait de troubles auditifs.

C.2.2 Stimuli

Trente-et-un sons ont été synthétisés avec différentes densités temporelles d'impacts ($d \in [0.001,1]$, valeurs espacées logarithmiquement) grâce au modèle présenté précédemment. Un profil de vitesse enregistré à partir d'un geste réel (un trait rapide) sur une tablette graphique (cf figure III.7) a été utilisé pour contrôler la vitesse du geste dans le modèle (et donc la fréquence de coupure du filtre passe-bas). Pour ajouter du naturel aux stimuli, tous ont été convolués à une réponse impulsionnelle, évoquant un matériau dur type pierre, généré par le synthétiseur de sons d'impacts proposé par Aramaki *et al.* (2009b) (♯).

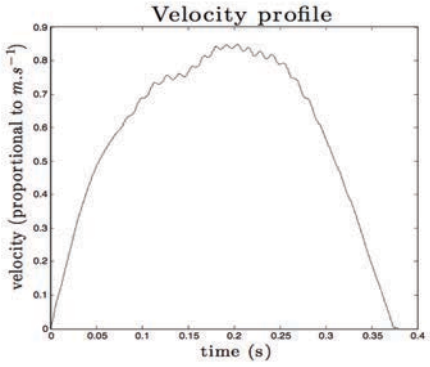

FIGURE III.7 – Profil de vitesse utilisé pour générer les stimuli.

C.2.3 Protocole

Un protocole de catégorisation à deux choix forcés a été choisi pour évaluer la pertinence du contrôle proposé. Avant que la session ne commence, chaque sujet entendait deux des stimuli du test ($d = 0.0073$ et $d = 0.91$), présentés dans un ordre aléatoire, afin de les familiariser avec le type de sons utilisés. Durant le test, les 31 stimuli ont été présentés dans un ordre aléatoire. Les participants pouvaient écouter 2 fois chaque stimulus, puis devaient l'associer à l'une des deux catégories proposées, "frotter" ou "gratter".

C.2.4 Résultats

Les résultats du test sont présentés sur la figure III.8 (figure du haut). Les résultats présentant une forme de sigmoïde, on peut les modéliser par une fonction logistique de la forme :

$$y = a + \frac{b}{1 + e^{-(d-c)\cdot g}} \tag{III.3}$$

où y est la probabilité de catégoriser le son dans l'interaction gratter, d la densité d'impacts, a l'asymptote basse, b la différence entre l'asymptote basse et l'asymptote haute, c la valeur de la densité d'impacts lorsque la sigmoïde atteint la moitié de son maximum et g la pente au point d'inflexion. Les paramètres ont été ajustés au sens des moindres carrés et sont résumés dans le tableau III.1. Ce type de modélisation a déjà été utilisé par exemple pour la catégorisation du matériau perçu dans des sons d'impacts pour des continua entre différents matériaux (Aramaki *et al.*, 2011; Micoulaud-Franchi *et al.*, 2011).

C.2.5 Discussion

Les résultats du test perceptif montrent que la variation de la densité temporelle d'impacts permet effectivement de contrôler l'interaction perçue, i.e. "frotter" ou "gratter". Pour des valeurs de $d > 0.1$ (i.e. pour un intervalle temporel moyen entre deux impacts successifs inférieur à 0.2 ms), les stimuli sont unanimement classifiés comme évoquant l'interaction frotter. Pour des valeurs de $d < 0.01$ (i.e. pour un intervalle temporel moyen entre deux impacts successifs supérieur à 2 ms), les stimuli sont unanimement classifiés comme évoquant l'interaction gratter. Pour un intervalle temporel moyen entre deux impacts successifs compris entre 0.2 ms et 2 ms, la perception est

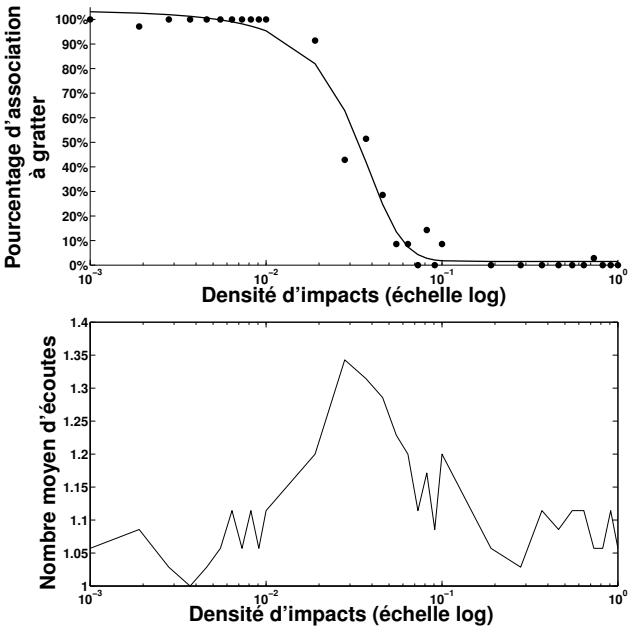

FIGURE III.8 – Haut : Résultats du test perceptif de catégorisation des sons de synthèse. En abscisse, la densité d'impacts sur un échelle logarithmique (unité temporelle : l'intervalle d'échantillonnage à 44100 Hz), en ordonnée le pourcentage d'association à l'interaction gratter. Les points représentent le pourcentage d'association à l'interaction gratter moyenné sur les sujets. La courbe représente la fonction sigmoïde estimée. Bas : Nombre moyen d'écoutes de chaque stimulus du test.

TABLE III.1 – Résultats de la régression de la sigmoïde pour le test de catégorisation des sons de synthèse. Pour chaque coefficient sont données la valeur (colonne 1), l'erreur-type (colonne 2), la statistique du test t par rapport à 0 (colonne 3) et la significativité p du t-test (colonne 4).

	estim.	err.	t	p
a	0.01	0.02	0.91	0.37
b	1.09	0.07	15.33	$< 10^{-3}$
c	0.03	0.002	12.28	$< 10^{-3}$
g	-86.42	14.74	-5.86	$< 10^{-3}$

plus ambiguë. Cette ambiguité perceptive se traduit par un nombre moyen d'écoute plus élevé pour ces stimuli (bas de la figure III.8).

Si on fait l'hypothèse que les auditeurs se basent sur le fait d'entendre des impacts plutôt distincts pour inférer que l'interaction est gratter, et inversement sur le fait d'entendre plutôt un son continu pour inférer que l'interaction est frotter, ces résultats sont cohérents avec la littérature sur la résolution auditive temporelle permettant de détecter des silences dans des signaux continus. En effet, les durées des silences nécessaires pour être détectables varient entre 2 ms et 20 ms selon le niveau et les propriétés spectrales des signaux (Verhey, 2010). Nos résultats sont néanmoins difficiles à comparer avec des études précises de la littérature étant donnée la complexité des stimuli utilisés pour notre test (convolution avec une réponse impulsionnelle du train d'impulsions, trains d'impulsions aléatoires, propriétés spectrales variables au cours du temps du fait de la fréquence de coupure variable du passe-bas...). Cependant, étant donnée la proximité des valeurs de densités temporelles d'impacts pour percevoir l'interaction gratter révélées par le test avec les résultats de la littérature, on peut supposer à juste titre qu'un des mécanismes cognitifs permettant de distinguer frottement et grattement est basé sur la résolution temporelle nécessaire pour distinguer des silences dans des signaux continus.

D Modèle générique de sons d'interactions continues

Les résultats présentés jusqu'ici ont permis de montrer que les interactions "frotter", "gratter" ainsi que "rouler" (cf chapitre II) sont régies par des statistiques particulières d'une suite d'impacts sur un objet résonant. En se basant sur des tests perceptifs ainsi qu'un modèle phénoménologique, on a montré qu'un contrôle de la densité temporelle d'impacts permet de distinguer les interactions frotter et gratter. En introduisant une structure plus particulière aux statistiques de ces séries d'impacts, i.e. des corrélations entre impacts ainsi qu'une durée particulière de ces impacts en fonction de leur amplitude, on reproduit la morphologie sonore responsable de la perception du roulement.

Il est donc possible de proposer un modèle génératif unique pour ces 3 interactions à partir de la formulation de la partie source f du modèle source-filtre, posée dans le chapitre II (équation (II.5)) :

$$f(t) = \sum_n A^n \, \phi^n(t - T^n) \tag{III.4}$$

La figure III.9 résume le processus de synthèse du terme source de notre modèle. Dans le cas des interaction frotter et gratter, la mesure des statistiques à imposer à la série d'impacts pour permettre l'évocation de l'une ou l'autre des interactions a été effectuée grâce aux tests perceptifs précédents ainsi que des observations empiriques sur des signaux enregistrés, tandis que pour l'interaction rouler, ces statistiques proviennent d'une analyse formelle d'un modèle physique. Dans cette partie, on récapitulera les différents paramètres de synthèse, puis on décrira les morphologies du signal source typiques de chaque interaction.

D.1 Description des paramètres du modèle

Les différents paramètres bas-niveau du modèle de synthèse générique sont :

— **Modèle d'impact.** Deux paramètres permettent le contrôle de la durée d'impact t_0 (équation (II.12)) dans le modèle d'impact choisi (équation (II.9)) : ζ' et θ'.

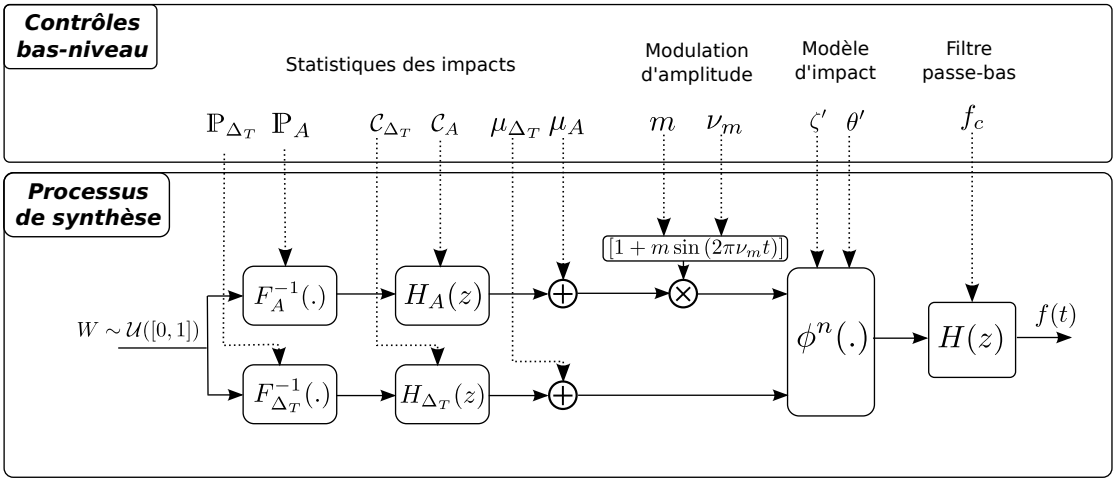

FIGURE III.9 – Schéma de synthèse général et contrôles bas-niveau.

— **Densités de probabilités.** Les densités de probabilités, notées $\mathbb{P}_A$ et $\mathbb{P}_{\Delta_T}$ pour respectivement A et Δ_T, sont échantillonnées en un nombre fini de valeurs discrètes et utilisées pour obtenir les fonctions de répartitions F_A et F_{Δ_T} par somme cumulée.
— **Filtres des séries.** Comme on l'a vu pour le roulement dans le chapitre II, ces filtres à 1 pôle-1 zéro permettent d'introduire de la corrélation entre les impacts, et plus précisément entre les séries d'amplitudes A et Δ_T. Chaque filtre est décrit par un jeu de 2 coefficients (a_1, b_1). On notera respectivement $\mathcal{C}_A$ et $\mathcal{C}_{\Delta_T}$ ces jeux de paramètres pour les filtres H_A et H_{Δ_T}.
— **Coefficients de décalage.** Ces coefficients μ_A et μ_{Δ_T} correspondent aux valeurs moyennes des séries A et Δ_T analysées et permettent, comme on l'a vu dans le chapitre II pour le roulement, de recentrer les séries d'amplitudes et d'intervalles temporels après l'étape de filtrage.
— **Modulation d'amplitude.** Les paramètres permettant de contrôler la modulation d'amplitude (cf équation (II.7)) sont la profondeur de modulation m et la fréquence de modulation ν_m (qui dépend de la vitesse et de la taille de l'objet roulant, cf équation (II.8)).
— **Filtre passe-bas.** Ce filtre H agit en dernière étape sur le signal source. Sa fréquence de coupure f_c est proportionnelle à la vitesse du geste (cf partie C).

En contrôlant ces différents paramètres bas-niveau du modèle de synthèse générique, on peut synthétiser différentes morphologies de signal source f, et ainsi évoquer différentes interactions. Ces morphologies sont décrites ci-après.

D.2 Morphologie des différentes interactions

D.2.1 Morphologie de l'interaction "gratter"

Comme on l'a vu précédemment, la morphologie du signal source qui permet d'évoquer l'interaction gratter est une suite parcimonieuse d'impacts, i.e. une suite d'impacts avec un intervalle temporel moyen entre les impacts de l'ordre de $\Delta_T > 2$ ms. Dans le test perceptif présenté en C.2 ayant permis d'obtenir ces résultats, les stimuli ont été synthétisés en considérant qu'à chaque instant discret on effectuait un tirage de Bernoulli de paramètre d et qu'un impact avait lieu lorsque l'on obtenait un "succès". Ce processus aléatoire peut être modélisé par une distribution géométrique qui définit la

probabilité que Δ_T prenne la valeur δ_t, i.e. que cette loi décrit le nombre de tirages de Bernoulli successifs nécessaires pour obtenir un "succès". Sa densité de probabilité est définie par : $\mathbb{P}_{\Delta_T}(\Delta_T = \delta_t) = (1-d)^{\delta_t - 1} d$. On verra plus loin qu'il est plus commode d'utiliser cette distribution. La densité de probabilité $\mathbb{P}_A$ de la suite d'amplitudes des impacts est toujours définie par une loi uniforme sur l'intervalle $[0, 1]$. Aucune corrélation entre les impacts n'est considérée, les coefficients des filtres $\mathcal{C}_A$ et $\mathcal{C}_{\Delta_T}$ sont donc nuls. Les paramètres μ_A et μ_{Δ_T} sont également nuls, et on considère $\theta' = 0$ (pas de dépendance entre l'amplitude de l'impact et sa durée).

D.2.2 Morphologie de l'interaction "frotter"

Dans le cas de l'interaction frotter, la série d'impacts doit avoir un intervalle temporel moyen entre les impacts de l'ordre de $\Delta_T < 0.2$ ms. Les paramètres de synthèse bas-niveau sont donc identiques à ceux de l'interaction gratter, excepté pour la densité de probabilité $\mathbb{P}_{\Delta_T}$ de la série Δ_T. Lorsque le paramètre d approche 1, la distribution géométrique tend vers un Dirac centré en $\delta_t = 1$ (l'unité temporelle est ici l'intervalle d'échantillonnage). Comme $\mathbb{P}_A$ suit toujours une loi uniforme, le bruit alors synthétisé est blanc et uniforme.

D.2.3 Morphologie de l'interaction "rouler"

Comme vu dans le chapitre II, $\mathbb{P}_A$ et $\mathbb{P}_{\Delta_T}$ sont des distributions Gaussiennes pour l'interaction rouler. Les variances de ces distributions, ainsi que les coefficients des filtres $\mathcal{C}_A$ et $\mathcal{C}_{\Delta_T}$ et les paramètres μ_A et μ_{Δ_T} permettent de contrôler la rugosité perçue de la surface sur laquelle l'objet roule (cf partie D.3). Le paramètre θ' est fixé à 0.29 d'après les simulations numériques effectuées dans le chapitre précédent.

D.3 Stratégie de navigation dans l'espace des actions

Les morphologies des différentes interactions ayant été définies, on peut maintenant proposer une stratégie de navigation entre elles. Pour chaque interaction, on proposera un prototype représentatif d'une morphologie dont on donnera les paramètres, et on présentera les différents contrôles proposés pour chacune de ces interactions. La stratégie de contrôle adoptée est inspirée de celle proposée par Aramaki *et al.* (2011) pour le contrôle intuitif du matériau perçu dans la synthèse de sons d'impacts, présentée dans l'état de l'art (cf parties B.2 et D.4).

D.3.1 Définition des prototypes

Pour chaque interaction, on définira donc un son prototype, qui est représenté par un jeu particulier de l'ensemble des paramètres $\mathfrak{P} = \{\mathbb{P}_A, \mathbb{P}_{\Delta_T}, \mathcal{C}_A, \mathcal{C}_{\Delta_T}, \mu_A, \mu_{\Delta_T}, m, \nu_m, \zeta', \theta', f_c\}$ représentant sans ambiguité la morphologie sonore associée. On nommera respectivement $\mathfrak{P}_{\text{scratch}}$, $\mathfrak{P}_{\text{rub}}$ et $\mathfrak{P}_{\text{roll}}$ les prototypes des interactions gratter, frotter et rouler. Les paramètres de ces prototypes seront ensuite interpolés entre eux pour passer d'une interaction à l'autre.

Prototypes des interactions "frotter" et "gratter" Le seul paramètre de synthèse différenciant ces deux interactions est la densité d'impacts d, fixée à $5 \cdot 10^{-3}$ pour gratter et 1 pour frotter, valeurs choisies d'après les résultats du test perceptif présenté en C.2. La vitesse du geste V, normalisée entre 0 et 1, fait varier linéairement la fréquence de

TABLE III.2 – Valeurs extrêmes pour le mapping linéaire entre le contrôle de rugosité de surface ρ et les paramètres de synthèse liés à la série d'impacts (amplitude A et intervalle entre impacts Δ_T).

		$\rho_{min} = 0$	$\rho_{max} = 1$
A	σ_A	0.04	0.04
	a_1	-0.97	-0.93
	b_1	0.07	0.32
	μ_A	0.43	0.27
$\mathbf{\Delta_T}$	σ_{Δ_T}	0.19 (ms)	0.85 (ms)
	a_1	-0.97	-0.93
	b_1	-0.34	0.35
	μ_{Δ_T}	3.1 (ms)	6.4 (ms)

coupure du filtre passe-bas entre 0 et 4000 Hz. La taille perçue S (normalisée entre 0.1 et 1) de l'objet interagissant sur la surface est contrôlée par ζ' par le mapping suivant :

$$t_0 = 7.88 \times 10^{-4}\, S \tag{III.5}$$

t_0 étant la durée d'un impact en secondes. Cela laisse donc à l'utilisateur la liberté de raffiner le prototype avant d'effectuer des transitions continues entre interactions.

Prototype de l'interaction "rouler" Pour l'interaction rouler, les contrôles possibles sont la rugosité de la surface ρ (normalisée entre 0 et 1), la taille de bille et sa vitesse, ainsi que son asymétrie. La rugosité de la surface définit les paramètres bas-niveau liée à la statistique des impacts. Les valeurs des paramètres, interpolés linéairement par rapport à la valeur de la rugosité, sont rappelés dans la table III.2. La vitesse de la bille contrôle la fréquence de coupure f_c du filtre comme pour les interaction frotter et gratter, ainsi que la fréquence de la modulation d'amplitude ν_m dont le mapping est :

$$\nu_m = 3\,\frac{V}{S} \tag{III.6}$$

La profondeur de modulation m, par défaut fixée à 0.3, permet de passer d'un objet parfaitement symétrique (m proche de 0) à un objet roulant très déformé (m proche de 1), comme une ellipse par exemple. Enfin, la taille perçue contrôle la durée d'impact par le mapping :

$$t_0 = 7.88 \times 10^{-4}\, S\, A^{-0.29} \tag{III.7}$$

t_0 étant la durée d'un impact en secondes, dépendant donc également de l'amplitude de l'impact A.

D.3.2 Espace sonore des interactions

On peut maintenant proposer une stratégie de navigation entre ces différentes interactions. L'espace de contrôle proposé pour le contrôle de l'interaction perçue est présenté sur la figure III.10.

Les prototypes des interactions frotter, gratter et rouler sont placés respectivement aux angles 0, $\frac{2\pi}{3}$ et $\frac{4\pi}{3}$ sur la circonférence d'un disque de rayon 1. Autour de la circonférence, le son de synthèse $\mathcal{S}$, dont la position est caractérisée par son angle θ, est défini

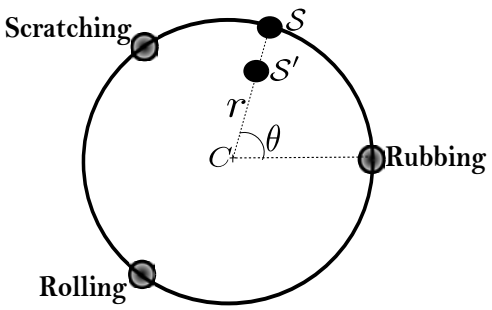

FIGURE III.10 – Schématisation de l'espace sonore des interactions du synthétiseur.

par le jeu de paramètres de synthèse bas-niveau $\mathfrak{P}_{\mathcal{S}}(\theta)$ comme :

$$\mathfrak{P}_{\mathcal{S}}(\theta) = T(\theta)\,\mathfrak{P}_{\text{rub}} + T\left(\theta - \frac{2\pi}{3}\right)\mathfrak{P}_{\text{scratch}} + T\left(\theta - \frac{4\pi}{3}\right)\mathfrak{P}_{\text{roll}} \tag{III.8}$$

où la fonction $T(\theta)$ est définie par :

$$T(\theta) = \begin{cases} -\frac{3}{2\pi}\theta + 1 & , \quad \theta \in \left[0; \frac{2\pi}{3}\right[\\ 0 & , \quad \theta \in \left[\frac{2\pi}{3}; \frac{4\pi}{3}\right[\\ \frac{3}{2\pi}\theta - 2 & , \quad \theta \in \left[\frac{4\pi}{3}; 2\pi\right[\end{cases} \tag{III.9}$$

A l'intérieur du disque, un son $\mathcal{S}'$, caractérisé par son angle θ et son rayon r, est défini par le jeu de paramètres $\mathfrak{P}_{\mathcal{S}'}(\theta, r)$:

$$\mathfrak{P}_{\mathcal{S}'}(\theta, r) = (1 - r)\mathfrak{P}_{C} + r\mathfrak{P}_{\mathcal{S}}(\theta) \tag{III.10}$$

où

$$\mathfrak{P}_{C} = \frac{1}{3}\left(\mathfrak{P}_{\text{rub}} + \mathfrak{P}_{\text{scratch}} + \mathfrak{P}_{\text{roll}}\right) \tag{III.11}$$

et $\mathfrak{P}_{\mathcal{S}}(\theta)$ est défini dans l'équation (III.8).

En plus de la navigation dans cet "espace d'interaction", le geste est pris en compte dans la stratégie de contrôle. En effet, pour de telles interactions continues, le geste sous-jacent à la production du son est un attribut perceptif fondamental qui peut être transmis par la dynamique du son (Merer *et al.*, 2013; Thoret *et al.*, 2014). Le geste est pris en compte par le filtrage passe-bas relié à la vitesse relative entre l'objet qui interagit (la main par exemple) et la surface comme proposé par Van Den Doel *et al.* (2001). Si on associe cet effet à une loi biologique du mouvement humain, alors cette calibration spécifique du profil de vitesse permet l'évocation d'un geste humain (Thoret *et al.*, 2014). La figure III.11 représente l'interface du synthétiseur ($\sharp$).

E Perspectives d'élargissement de l'espace sonore des interactions : vers la friction non-linéaire

Dans cette section, on présentera des travaux préliminaires ayant pour but d'élargir l'espace sonore des interactions continues afin d'y inclure la friction non-linéaire. On présentera tout d'abord un modèle source-filtre de friction non-linéaire développé dans l'équipe (Thoret *et al.*, 2013), adapté au paradigme action-objet. Puis on proposera une reformulation du modèle de synthèse des interactions "frotter", "rouler" et "gratter" afin de rendre compatible les deux modèles.

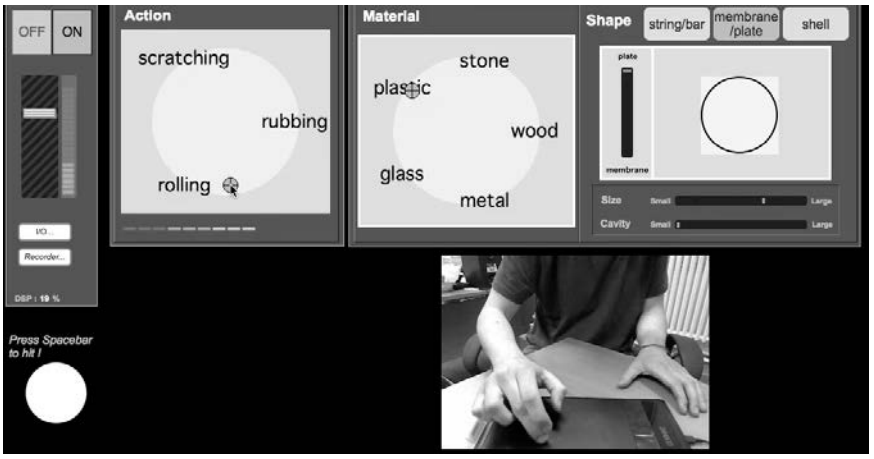

FIGURE III.11 – Interface du synthétiseur. On peut voir l'utilisateur contrôler le synthétiseur grâce à une tablette graphique qui permet de capter la vitesse de son geste.

E.1 Modèle de synthèse source-filtre de sons de friction non-linéaire

Les sons de friction tels que les couinements (doigts mouillés sur une assiette, bruit d'essuie-glace, verre de table qui chante, auto-oscillations d'une corde de violon...) sont produits par des phénomènes physiques non-linéaires, i.e. un fort couplage a lieu entre les deux objets en contact. Afin de permettre la génération de ce type de sons grâce à un modèle source-filtre, Thoret *et al.* (2013) proposent de prendre en compte ces non-linéarités directement dans le signal source. Pour dériver ce modèle, un premier phénomène physique est étudié : la friction de Coulomb. Ce phénomène peut être décrit par un modèle phénoménologique simple qui est présenté sur la figure III.12 (haut). On considère une masse posée sur un tapis roulant et attachée par un ressort à un point fixe. En fonction de la vitesse du tapis, la raideur du ressort et le poids de la masse, celle-ci alterne des phases où elle adhère au tapis et d'autres où elle glisse sur le tapis, qu'on appelle phénomène de "collé-glissé" (ou "stick-slip" en anglais). Ce modèle décrit le comportement le plus simple de la friction non-linéaire, qu'on peut voir sur la figure III.12 (bas). Le déplacement résultant $x(t)$ (en négligeant la durée de la phase de glissement) correspond à un signal en dent de scie dont la décomposition en série de Fourier est :

$$x(t) = \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \sin(2\pi k f_0 t) \tag{III.12}$$

où f_0 correspond à la fréquence du mouvement de collé-glissé. Ici, ce modèle d'excitation est harmonique. Les modèles prenant en compte la durée de la phase de glissement ont également un spectre harmonique dont l'amplitude des composantes diminue avec l'ordre de l'harmonique. Ce modèle permet donc de donner un a priori sur le comportement du signal source à considérer afin de pouvoir obtenir un modèle de synthèse plus élaboré de friction non-linéaire, c'est-à-dire qu'on cherchera un spectre harmonique.

Les auteurs se sont ensuite basés sur des observations empiriques de sons de friction non-linéaire enregistrés (cf figures III.13 et III.14) :

— Une porte qui grince
— Les couinements produits par un doigt humide sur une assiette
— Un doigt humide faisant couiner et chanter (auto-oscillation) un verre.

Une étude d'harmonicité grâce à une méthode de suivi de partiels (Ellis, 2003) a tout d'abord permis de vérifier l'hypothèse faite grâce au modèle simple du tapis roulant.

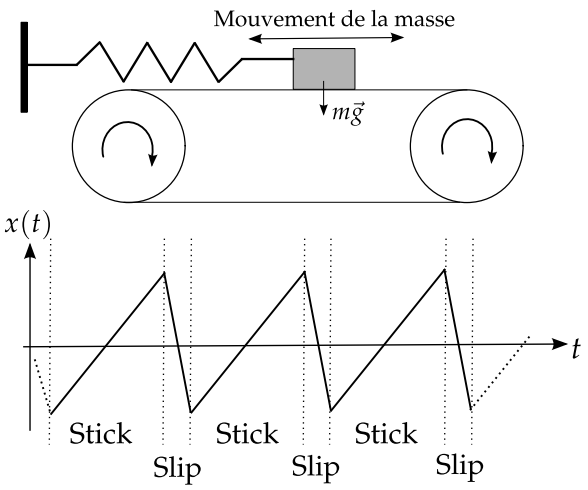

FIGURE III.12 – Modèle phénoménologique du "collé-glissé" ("stick-slip" en anglais)

Ensuite, l'étude empirique des représentations temps-fréquence a permis de révéler les morphologies suivantes :

— Dans le cas de la porte qui grince, on observe des transitions soudaines entre différents régimes de vibration, qui se traduisent par une montée en fréquence progressive du spectre harmonique, et une modulation de fréquence aléatoire.
— Dans le cas des couinements de l'assiette humide, on observe une moins grande plage de variation du spectre harmonique que pour la porte, mais une modulation de fréquence aléatoire plus importante autour d'une valeur centrale de f_0. Lorsqu'un partiel est proche d'un mode de l'assiette, celle-ci se met à résonner (ce qui renforce l'idée de séparer excitateur et résonateur).
— Dans le cas du verre frotté sur son bord, on observe deux phases différentes. La première correspond à des couinements et a un comportement identique à celui de l'assiette humide. La deuxième a lieu quand f_0 se bloque sur un mode du verre. On observe alors des modulations d'amplitudes des partiels. Les transitions entre ces différents régimes semblent difficilement prédictibles.

Ainsi, la morphologie acoustique liée à l'évocation de différents sons de friction non-linéaire est liée à l'évolution de la fréquence fondamentale d'un spectre harmonique. Basés sur ces observations, Thoret *et al.* (2013) ont proposé un modèle contrôlable de source d'un modèle source-filtre permettant de synthétiser ces différents sons. Le terme source peut être décrit ainsi :

$$f(t) = \sum_{k=1}^{N} \mathrm{AM}_k(t) \sin\left(\varphi_k(t)\right) \tag{III.13}$$

tel que :

$$\varphi_k(t) = \int_0^t 2\pi k f_0(\tau) \mathrm{d}\tau \tag{III.14}$$

$f_0(t)$ étant la fréquence instantanée du fondamental du peigne harmonique, et $\mathrm{AM}_k(t)$ étant la loi de modulation en amplitude du $k^{\text{ième}}$ harmonique. La fréquence fondamentale $f_0(t)$ est contrôlée par la loi suivante :

$$f_0(t) = (1 + \gamma\eta(t))\Gamma(v(t), p(t)) \tag{III.15}$$

où $\eta(t)$ est un bruit blanc filtré passe-bas ($<$ 20 Hz) permettant de simuler les variations aléatoires des partiels et γ pondère son importance. $\Gamma(v(t), p(t))$ est une fonction

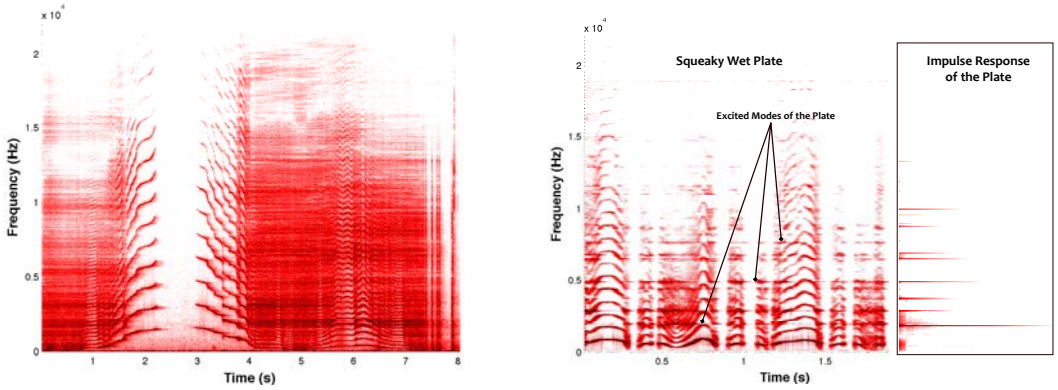

FIGURE III.13 – Représentation temps-fréquence d'un son de porte grinçante (gauche) et des sons de couinements produits par un doigt humide sur une assiette (droite). Figures extraites de (Thoret *et al.*, 2013).

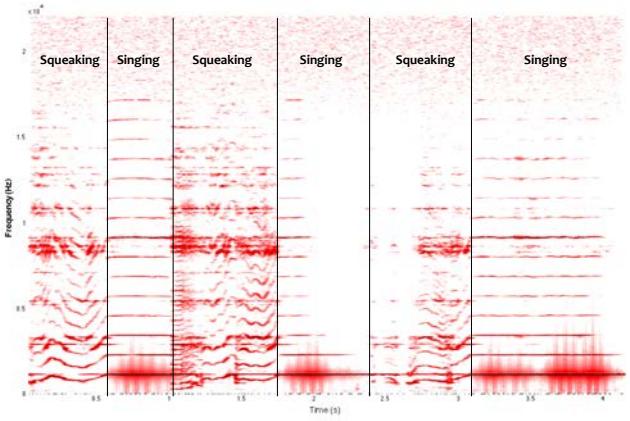

FIGURE III.14 – Représentation temps-fréquence du son d'un doigt humide faisant couiner et chanter (auto-oscillation) un verre. Figure extraite de (Thoret *et al.*, 2013).

permettant de définir un mapping entre la fréquence fondamentale et les paramètres de vitesse $v(t)$ et de pression $p(t)$ du geste[1]. Ainsi pour les sons de couinements ou grincements, on choisira $\gamma > 0$, tandis que pour les auto-oscillations (verre qui chante, corde frottée par un archet...) on choisira $\gamma = 0$ et $\Gamma(v(t), p(t)) = \tilde{f}_{(n,0)}$, le $n^{\text{ième}}$ mode de résonance du résonateur. La loi de modulation d'amplitudes $\mathrm{AM}_k(t)$ est fixée à $\frac{1}{k}$ pour les sons de couinements, grincements ou de corde frottée par un archet. Pour les auto-oscillations du verre, la loi de modulation d'amplitude suit :

$$\mathrm{AM}_k(t) = \frac{1}{k} \sin\left(2\pi \int_0^t \frac{v(\tau)}{\pi D} \mathrm{d}\tau\right) \tag{III.16}$$

v étant la vitesse du doigt frottant le rebord du verre et D le diamètre du verre, d'après une étude de Rossing (1994).

1. Ce paramètre de mapping permet donc de contrôler finement le comportement du modèle, i.e. de définir des zones de pression/vitesse (différentes de ce que la physique du phénomène impose) menant à tel ou tel autre comportement (auto-oscillation ou couinement par exemple), et on peut ainsi comprendre l'intérêt d'un tel modèle pour des applications comme le guidage du geste ($\sharp$).

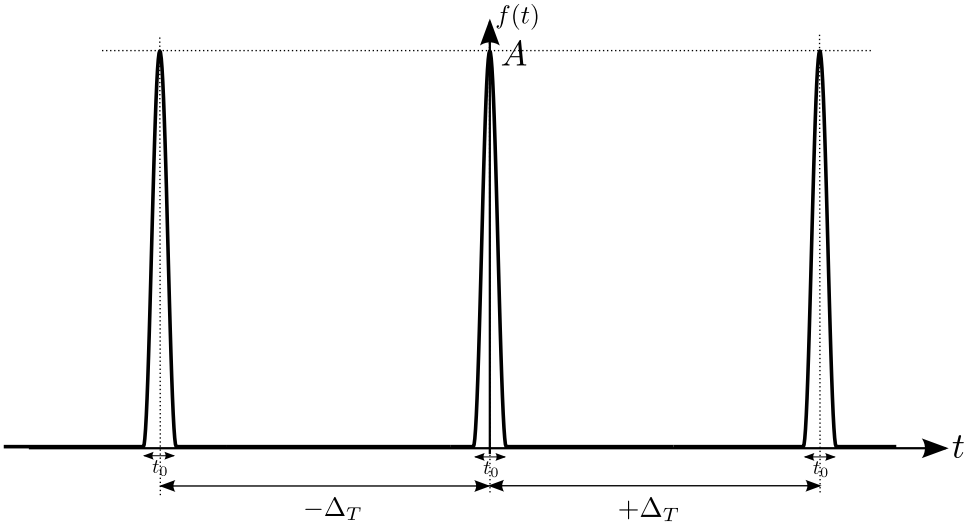

FIGURE III.15 – Excitation $f(t)$ pour le modèle simplifié.

Ce modèle de synthèse de sons de friction non-linéaire en source-filtre est tout à fait compatible avec le paradigme action-objet que l'on considère dans cette thèse. Cependant, l'implémentation actuelle du modèle de signal pour les interactions rouler, frotter et gratter ne permet pas d'effectuer des transitions continues entre ces interactions et la friction non-linéaire. Dans la section qui suit, on décrira une implémentation alternative des excitations rouler, frotter et gratter en synthèse additive, qui laisse envisager l'intégration de la friction non-linéaire à l'espace sonore des interactions. Ces travaux sont actuellement en cours.

E.2 Construction du modèle d'excitation pour les actions "rouler", "frotter" et "gratter" en synthèse additive

Dans cette partie on va proposer de construire l'excitation f (équation III.4) en synthèse additive, afin de la rendre compatible avec le modèle générique pour la friction non-linéaire proposé par Thoret *et al.* (2013) (cf équation III.13). On étudiera tout d'abord un modèle simplifié dont on viendra ensuite faire varier les paramètres pour obtenir les excitations pour le rouler, frotter et gratter.

E.2.1 Modèle simplifié

Supposons tout d'abord une interaction très simple $f(t)$ qui est constituée d'une suite d'impacts ϕ (eq. II.9) identiques (mêmes durées t_0, mêmes amplitudes A) et régulièrement espacés de Δ_T dans le temps. Cette fonction, représentée sur la figure III.15, peut se définir comme périodisation de la fonction $\phi(t)$ par un peigne de Dirac :

$$f(t) = \left[\text{Ш}_{\Delta_T} * \phi\right](t) \tag{III.17}$$

où Ш_{Δ_T} est le peigne de Dirac :

$$\text{Ш}_{\Delta_T} = \sum_{p=-\infty}^{+\infty} \delta_{p\Delta_T}(t) = \sum_{p=-\infty}^{+\infty} \delta(t - p\Delta_T) \tag{III.18}$$

$\delta(t)$ étant la distribution de Dirac qui vaut ∞ en $t = 0$ et 0 partout ailleurs, et dont l'intégrale sur $\mathbb{R}$ vaut 1.

Série de Fourier d'un signal périodisé La fonction f étant périodique de période Δ_T, elle admet une décomposition en série de Fourier telle que :

$$f(t) = \sum_{k=-\infty}^{+\infty} c_k e^{j\frac{2\pi k}{\Delta_T}t} \tag{III.19}$$

$$= a_0 + \sum_{k=1}^{+\infty} \left(a_k \cos\left(2\pi \frac{kt}{\Delta_T}\right) + b_k \sin\left(2\pi \frac{kt}{\Delta_T}\right)\right) \tag{III.20}$$

Les coefficients a_k peuvent être déduits des coefficients complexes c_k par :

$$\begin{cases} a_k = c_k + c_{-k} \quad , \quad k > 0 \\ a_0 = c_0 \end{cases} \tag{III.21}$$

et la fonction f étant paire, les coefficients b_k sont nuls. Pour la suite, il sera plus simple de considérer que chaque impact "commence" en $k\Delta_T$ ($k \in \mathbb{Z}$). On pourra alors écrire notre signal f de la sorte :

$$f(t) = a_0 + \sum_{k=1}^{+\infty} a_k \cos\left(2\pi \frac{k}{\Delta_T}\left(t - \frac{t_0}{2}\right)\right) \tag{III.22}$$

Mais pour le calcul des coefficients il est plus simple de considérer le signal comme décrit sur la figure III.15, et de calculer d'abord les coefficients complexes c_k pour en déduire les coefficients réels. Il est plus simple de calculer les coefficients complexes de la décomposition en série de Fourier de f car ceux-ci peuvent être obtenus par échantillonnage de la transformée de Fourier de ϕ aux points $k\frac{1}{\Delta_T}$, $k \in \mathbb{Z}$. En effet, la fonction f étant de produit de convolution d'un peigne de Dirac de période Δ_T et de la fonction ϕ (cf. équation III.17), sa transformée de Fourier $\hat{f}(\nu)$ (on choisit de noter ν la fréquence au lieu de l'habituel f afin de ne pas prêter à confusion avec la fonction qui s'appelle également f) est le produit des transformées de Fourier respectives $\frac{1}{\Delta_T}\text{Ш}_{\frac{1}{\Delta_T}}(\nu)$ et $\hat{\phi}(\nu)$ du peigne et de ϕ :

$$\hat{f}(\nu) = \hat{\phi}(\nu) \times \left(\frac{1}{\Delta_T}\text{Ш}_{\frac{1}{\Delta_T}}(\nu)\right) = \hat{\phi}(\nu) \times \left(\frac{1}{\Delta_T}\sum_{k=-\infty}^{+\infty} \delta(\nu - k\frac{1}{\Delta_T})\right) \tag{III.23}$$

Les coefficients complexes de la série de Fourier de f sont donc :

$$c_k = \hat{f}\left(k\frac{1}{\Delta_T}\right) = \frac{1}{\Delta_T}\hat{\phi}\left(k\frac{1}{\Delta_T}\right) \quad \text{pour} \quad k \in \mathbb{Z} \tag{III.24}$$

Transformée de Fourier de la fonction ϕ Comme vu précédemment, afin d'obtenir les coefficients de la décomposition en série de Fourier de f, il nous faut calculer la transformée de Fourier $\hat{\phi}$ de ϕ. Cette fonction est définie comme (cf. équation II.9) :

$$\phi(t) = \frac{A}{2}\left[1 + \cos\left(\frac{2\pi t}{t_0}\right)\right]\Pi_{t_0}(t) \tag{III.25}$$

où $\Pi_{t_0}(t)$ est la fonction porte qui vaut 1 pour $t \in [-t_0, +t_0]$ et 0 ailleurs. Sa transformée de Fourier s'exprime donc comme la somme de trois contributions :

$$\hat{\phi}(\nu) = \frac{A}{2}\left[\left(\delta_0 + \frac{1}{2}\delta_{\frac{1}{t_0}} + \frac{1}{2}\delta_{-\frac{1}{t_0}}\right) * \hat{\Pi}_{t_0}\right](\nu) \tag{III.26}$$

$$= \frac{A}{2}\left[\hat{\Pi}_{t_0}(\nu) + \underbrace{\frac{1}{2}\hat{\Pi}_{t_0}\left(\nu - \frac{1}{t_0}\right) + \frac{1}{2}\hat{\Pi}_{t_0}\left(\nu + \frac{1}{t_0}\right)}_{\alpha}\right] \tag{III.27}$$

où la transformée de Fourier de la fonction porte est :

$$\hat{\Pi}_{t_0}(\nu) = t_0 \frac{\sin(\pi t_0 \nu)}{\pi t_0 \nu} = t_0 \operatorname{sinc}(\pi t_0 \nu) \tag{III.28}$$

La contribution du terme α se réduit à :

$$\alpha = \frac{t_0}{2}\left[\operatorname{sinc}(\pi t_0 \nu - \pi) + \operatorname{sinc}(\pi t_0 \nu + \pi)\right] \tag{III.29}$$

$$= -\frac{(\nu t_0)^2}{(\nu t_0)^2 - 1}\hat{\Pi}_{t_0}(\nu) \tag{III.30}$$

et finalement :

$$\hat{\phi}(\nu) = \frac{A t_0}{2}\frac{1}{1-(\nu t_0)^2}\operatorname{sinc}(\pi t_0 \nu) \tag{III.31}$$

Décomposition en série de Fourier de la fonction f En reprenant l'équation III.24 qui permet d'obtenir les coefficients complexes c_k de la décomposition en série de Fourier de la fonction f à partir de la transformée de Fourier de ϕ (eq. III.31), on obtient finalement :

$$c_k = \frac{1}{\Delta_T}\hat{\phi}\left(k\frac{1}{\Delta_T}\right) \quad \text{pour} \quad k \in \mathbb{Z} \tag{III.32}$$

$$= \frac{1}{\Delta_T}\frac{A t_0}{2}\frac{1}{1-(\frac{k t_0}{\Delta_T})^2}\operatorname{sinc}\left(\pi\frac{k t_0}{\Delta_T}\right) \tag{III.33}$$

Finalement, la décomposition en série de Fourier de f est :

$$\begin{cases} f(t) = a_0 + \sum_{k=1}^{+\infty} a_k \cos\left(2\pi\frac{k}{\Delta_T}\left(t - \frac{t_0}{2}\right)\right) \\ a_0 = \frac{A t_0}{2\Delta_T} \\ a_k = \frac{A}{2}\frac{1}{1-k^2\left(\frac{t_0}{\Delta_T}\right)}\left(\frac{t_0}{\Delta_T}\right)\operatorname{sinc}\left(k\pi\left(\frac{t_0}{\Delta_T}\right)\right) \end{cases} \tag{III.34}$$

E.2.2 Application aux interactions rouler, frotter et gratter

Afin d'appliquer notre représentation des signaux d'excitation en additif à des interactions "réelles", on considère des nouvelles séries temporelles $A(t)$ et $\Delta_T(t)$ obtenue d'après A^n et Δ_T^n comme représentées sur la figure III.16 (la nouvelle variable $t_0(t)$ est toujours déduite d'après $A(t)$ par l'équation II.12).

L'excitation de roulement peut alors se réécrire comme une somme de sinusoïdes modulées en phase et en amplitude :

$$f(t) = a_0(t) + \sum_{k=1}^{+\infty} a_k(t)\cos(\varphi_k(t)) \tag{III.35}$$

où $a_0(t)$ et $a_k(t)$ (définis dans l'équation III.34) sont maintenant des fonctions dépendant des nouvelles séries temporelles $A(t)$, $\Delta_T(t)$ et $t_0(t)$, et $\varphi_k(t)$ est la phase instantanée dépendant de la pulsation instantanée $\Omega_k(t)$ telle que :

$$\varphi_k(t) = \int_0^t \underbrace{\frac{2\pi k}{\Delta_T(\tau)}\left[1 - \frac{t_0(\tau)}{2}\right]}_{\Omega_k(\tau)} \mathrm{d}\tau \tag{III.36}$$

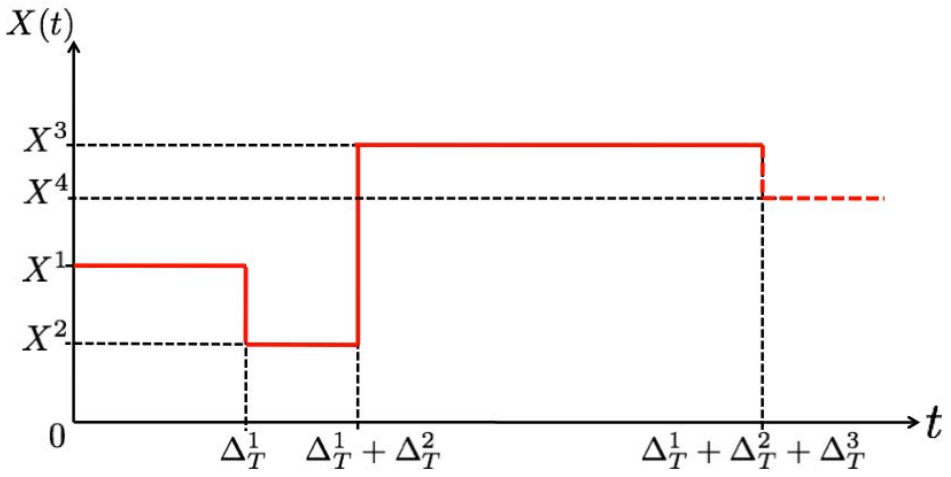

FIGURE III.16 – Nouvelles séries temporelles (ici X représente A ou Δ_T).

Si on reconstruit $f(t)$ directement à partir des nouvelles séries temporelles comme représentées sur la figure III.16, on a des "clics" à la reconstruction car on a des changements brusques de la pulsation instantanée. L'idée ici étant simplement de montrer la faisabilité du modèle en additif, pour éviter ces clics on lisse les séries en les convoluant par une fenêtre normalisée (une gaussienne de 40-50 échantillons à 44100 Hz fonctionne bien) ($\sharp$). Afin d'aller plus loin sur ce modèle, on pourrait plutôt envisager un modèle polynomiale de phase qui permettrait un meilleur contrôle de l'approximation de ces séries $A(t)$ et Δ_T en contrôlant l'ordre du polynôme. De plus, la convolution par une fenêtre a pour effet de faire "traîner" les variations rapides de ces séries, tandis qu'une approximation polynomiale permet de contrôler parfaitement le début et la fin de l'approximation.

En modifiant les statistiques des séries temporelles $A(t)$ et $\Delta_T(t)$ comme proposé précédemment, on peut ainsi synthétiser les excitations pour le modèle source-filtre pour faire du "rouler", du "frotter" et du "gratter". Une des perspectives de ces travaux serait de proposer une modélisation statistique unique des fréquences et amplitudes instantanées, ce qui permettrait de passer continûment entre l'ensemble des interactions.

F Discussion générale

Dans ce chapitre, on s'est intéressé à étendre le modèle développé pour la synthèse des sons de roulement (chapitre II) pour synthétiser d'autres sons d'interactions : "frotter" et "gratter". A notre connaissance, il n'existe pas d'étude sur les différences perceptives entre ces deux interactions dans la littérature. Dans le domaine de la synthèse sonore, ces deux interactions sont nommées sans jamais être différenciées. Afin d'étudier s'il existe effectivement des différences perceptives entre ces deux interactions, un test d'écoute et de catégorisation de sons enregistrés a été effectué. Ce test révèle qu'effectivement certains sons sont associés unanimement à l'une ou l'autre de ces interactions. Une étude qualitative des signaux catégorisés nous a permis de supposer qu'une différence majeure entre ces deux interactions est due à la densité temporelle d'impacts entre la surface et l'objet interagissant sur cette surface : les sons associés à gratter contiennent un moins grand nombre d'impacts que les sons associés à frotter. Cette hypothèse a été par la suite testée et validée par la synthèse : un modèle phéno-

ménologique de synthèse sonore permettant de contrôler cette densité d'impact a été proposé puis validé grâce à un test perceptif. Un modèle générique et une stratégie de contrôle associée permettant de naviguer continûment dans l'espace sonore des interactions a finalement été proposée. Enfin, des travaux préliminaires ayant pour but d'élargir cet espace sonore des interactions à la friction non-linéaire ont été présentés.

Une première perspective évidente de ce travail est de proposer l'espace sonore complet des interactions continues incluant la friction non-linéaire. Un test perceptif de validation de cet espace complet serait également intéressant et pourrait permettre de raffiner le mapping des transitions continues. En effet, les transitions perceptives continues entre rouler et frotter, et entre rouler et gratter n'ont pas été validées formellement. Les modèles des interactions frotter et gratter pourraient également être perfectionnés, grâce à des méthodes d'analyse/synthèse (e.g. (Lagrange *et al.*, 2010; Lee *et al.*, 2010)) ou à l'étude de modèles physiques (e.g. (Ben Abdelounis *et al.*, 2011; Ye, 2004)).

D'un point de vue applications, les modèles de synthèse sonore permettant de créer des continua entre différentes évocations sont intéressants pour étudier la perception auditive (Aramaki *et al.*, 2009a; Micoulaud-Franchi *et al.*, 2011). Le modèle proposé, du fait de sa contrôlabilité, est également intéressant pour des applications de guidage de geste ou d'étude de l'influence du son sur l'équilibre par exemple (Avanzini *et al.*, 2004; Danna *et al.*, 2013; Serafin *et al.*, 2013; Gandemer *et al.*, 2014). De même que pour le modèle de synthèse de sons de roulements, l'apport de la plateforme de navigation dans l'espace sonore des interactions pour la synthèse sonore dans les jeux vidéos est actuellement évalué dans le cadre du projet ANR PHYSIS [2].

Dans le chapitre suivant, on s'intéressera à l'utilisation des modèles de synthèse sonore d'interactions continues pour créer des métaphores sonores, en interagissant sur des textures sonores. Effectivement, comme on le verra, le paradigme action-objet proposé se prête tout particulièrement à cet effet, en laissant la possibilité de représenter les caractéristiques de la texture sonore dans la partie objet, i.e. l'idée n'est plus d'interagir (par exemple rouler) sur un objet "réel" (comme une plaque en bois par exemple), mais sur une texture "inouïe" (comme une nappe de synthétiseur par exemple).

2. `http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2[CODE]=ANR-12-CORD-0006`

Chapitre IV

Métaphores Sonores

Sommaire

A	Les textures sonores : définition	79
B	Synthèse de textures sonores : état de l'art	80
C	Proposition d'un modèle d'analyse/synthèse de textures sonores	83
D	Création de métaphores sonores	91
E	Construction du corpus sonore pour les tests perceptifs	94
F	Expérience 1 : métaphores sonores vs. mélange des sons	97
G	Expérience 2 : reconnaissance des interactions dans la métaphore	100
H	Expérience 3 : reconnaissance des textures originales dans la métaphore	103
I	Discussion générale	105

Dans les chapitres précédents, on a proposé un modèle source-filtre générique et contrôlable à haut-niveau qui permet de synthétiser des sons évoquant différentes interactions : frotter, gratter et rouler. Ces travaux se sont basés sur un paradigme *action-objet*, qui stipule que les sons résultent d'une interaction sur un objet et peuvent ainsi être découplés en deux parties indépendantes. Ceci a été possible en mettant en avant des morphologies du signal sonore particulières, et plus précisément des invariants transformationnels responsables de l'évocation de ces interactions. L'idée de ce chapitre est d'étudier la possibilité de modifier grâce aux invariants structurels mis en évidence une "texture sonore" afin de lui faire évoquer une interaction particulière (par exemple le roulement), tout en restant reconnaissable (i.e. en conservant une partie de son timbre original), et ainsi créer une "métaphore sonore". De tels outils peuvent être intéressant pour le design sonore, fournissant ainsi au designer tout un paradigme permettant de modifier des textures sonores "inertes" en textures informants également sur une interaction, éventuellement une taille et une vitesse. Ces outils sont donc d'intérêt tant pour l'art que pour l'industrie, par exemple pour la sonification informative des véhicules électriques qui du fait de leur nature silencieuse peuvent présenter un danger pour les piétons. En effet, le projet ANR MétaSon fait clairement apparaître une demande de l'industrie automobile de pouvoir créer des sons avec un contenu sémiotique spécifique. Notamment, pour la voiture, l'évocation du "rouler" est cruciale, mais il est également important de laisser une liberté créative au niveau du timbre des sonorités, qui est une marque de l'identité du véhicule. Il est donc important de proposer une méthode valable sur un grand nombre de textures sonores.

Pour créer ces métaphores sonores, on utilisera le principe de la synthèse croisée (Zölzer, 2002), qui consiste à hybrider deux sons différents afin de conserver certaines

caractéristiques de chacun au sein d'un mélange de ces 2 sons. Cette façon de faire fusionner deux sources sonores est bien connue des musiciens avec par exemple la *talkbox*, technique de synthèse croisée mécanique, où le son d'un instrument (e.g. une guitare électrique ou un synthétiseur) est conduit dans la bouche du musicien et est ainsi filtré par les formants du conduit vocal, permettant ainsi de faire "parler son instrument" [1]. Différentes techniques classiques existent, comme celles basées sur le vocoder à canaux (Dudley, 1939; Gold et Rader, 1967) ou le vocoder de phase (Flanagan et Golden, 1966; Moorer, 1978), la prédiction linéaire (Moorer, 1979; Keiler *et al.*, 2000) ou encore sur le cepstre du signal (technique appelée traitement homomorphique du signal, cf. Oppenheim et Schafer (1975)). Polansky et Erbe (1996) proposent quant à eux des "mutations spectrales", qu'ils définissent par différentes interpolations entre les modules des transformées de Fourier à court-terme de 2 signaux sonores. Enfin on peut également citer les travaux de Olivero *et al.* (2012) qui proposent une approche mathématique rigoureuse de la synthèse croisée basée sur l'altération de représentations temps-fréquence par des multiplicateurs de Gabor. Nous choisirons une approche prédiction linéaire par la suite.

Dans ce chapitre, on définira tout d'abord la notion de textures sonores, puis on passera en revue l'état de l'art sur la synthèse de ce type de sons. On proposera ensuite un modèle d'analyse/synthèse de textures sonores compatible avec notre paradigme action-objet, en représentant les caractéristiques de la texture au niveau de la partie objet (i.e. la partie filtre du modèle source-filtre), puis on détaillera le principe de synthèse croisée choisi. Une série de tests perceptifs permettra finalement de valider le concept de métaphores sonores proposé.

A Les textures sonores : définition

Bien que d'après Saint-Arnaud et Popat (1995) il ne semble pas y avoir de consensus réel sur ce qu'est exactement une texture sonore, ces derniers en donnent une définition intéressante :

> *A sound texture is like wallpaper : it can have local structure and randomness, but the characteristics of the fine structure must remain constant on the large scale.*

Pour être considéré comme une texture sonore, un son doit avoir des propriétés constantes au cours du temps. Ainsi, un extrait de quelques secondes d'une texture donnée ne doit pas différer significativement (au moins d'un point de vue perceptif) d'un autre extrait de cette même texture. Par exemple, la hauteur tonale d'une texture sonore (si définie, ou bien à défaut une grandeur telle que son barycentre spectral) n'augmentera pas continument au cours du temps comme celui d'une voiture qui accélère, son "rythme" n'accélèrera pas, etc : les textures sonores sont stationnaires.

Une manière intéressante d'isoler les textures sonores représentée sur la figure IV.1, consiste à considérer l'évolution au cours du temps de "l'information potentielle" apportée par la texture. Ici, "information" est à prendre au sens cognitif du terme : une information est une *"donnée pertinente que le système nerveux central est capable d'interpréter pour se construire une représentation du monde et pour interagir correctement avec lui"* [2]. La parole et la musique peuvent potentiellement apporter de nouvelles informations à chaque instant. Par exemple, je vais entendre quelqu'un parler. Puis me rendre compte

1. Voir par exemple Roger Troutman montrant le principe de la talkbox sur un synthétiseur : `https://www.youtube.com/watch?v=L_CBZkd2tGE`
2. `http://fr.wikipedia.org/wiki/Information#Perception`

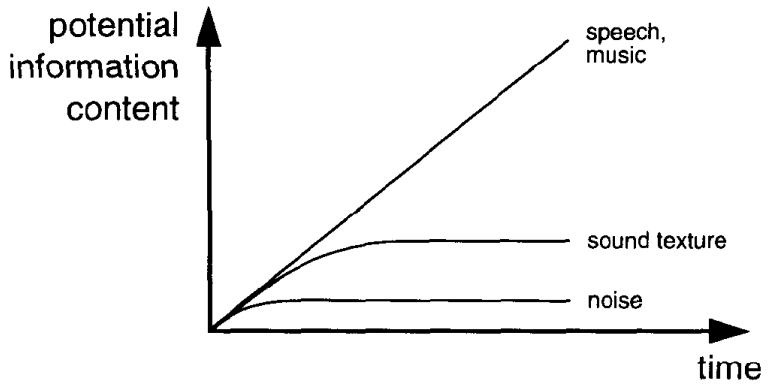

FIGURE IV.1 – Information potentielle contenue par différentes catégories de signaux sonores au cours du temps (figure extraite de (Saint-Arnaud et Popat, 1995)).

qu'il parle de moi. Je vais l'entendre dire que je suis gros, puis que je suis moche, puis que je suis idiot. L'information contenue dans son discours pourrait continuer d'augmenter si je n'allais pas lui mettre une baffe. A l'inverse, j'entends de la pluie (stationnaire). L'information que j'en tirerai se limitera sûrement au fait qu'il pleuve beaucoup ou non, et j'en déduirai si je dois mettre un manteau ou non. Mais je ne pourrai guère obtenir plus d'informations. Cette quantité limitée d'information est due au fait que les propriétés de la texture sont similaires au cours du temps. Un bruit n'évoquant pas d'événement sonore particulier apportera quant à lui encore moins d'information qu'une "texture sonore" (il sera peut-être "ennuyeux", "agréable", mais guère plus).

Saint-Arnaud et Popat (1995) résument tout ceci par la définition suivante : **(1)** Les textures sonores sont formées d'éléments sonores simples, appelés atomes ; **(2)** Ces atomes surviennent selon un motif haut-niveau régissant la texture, qui peut être périodique, aléatoire, ou les deux ; **(3)** Les caractéristiques haut-niveau doivent rester identiques sur une longue période temporelle ; **(4)** Le motif haut-niveau doit être délivré totalement en quelques secondes (qu'on appelle la "durée d'attention" ; les textures se démarquent ainsi des "paysages sonores" (*sound scapes*), terme employé pour caractériser des extraits sonores de durée généralement assez longue qui sont composés d'une multitude d'événements sonores, y compris des textures ; voir par exemple Schafer (1993)) ; **(5)** Les textures sonores peuvent être fortement aléatoires, pourvu qu'on puisse avoir un bon exemple de ces propriétés aléatoires pendant la "durée d'attention".

Ainsi, les applaudissements d'une foule, le son de la pluie, du vent, ou bien le final soutenu d'un orchestre classique peuvent être considérés comme des textures sonores.

B Synthèse de textures sonores : état de l'art

De multiples auteurs se sont penchés sur l'analyse des textures sonores, que ce soit à des fins de synthèse sonore ou de classification. Comme en témoigne l'excellent état de l'art sur la synthèse de textures sonores de Schwarz (2011), il n'est pas évident d'ériger une taxonomie des différentes méthodes de synthèse (i.e. additive, soustractive, granulaire, par apprentissage sur des bases d'ondelettes...) : certains modèles se retrouvent dans plusieurs catégories, une partie se base sur des méthodes d'analyse/synthèse complète tandis que d'autres modèles sont purement génératifs et se basent sur des considérations physiquement informées...

Ici, on prend le parti de séparer en deux les approches existantes de la littérature :

— Les approches potentiellement compatibles avec notre paradigme action-objet. En effet, le but de ce chapitre est de proposer un modèle cohérent avec le paradigme action-objet, afin de représenter la texture comme un objet sur lequel on pourra interagir. Ces approches seront donc plutôt soustractives.
— Les autres approches, comme la synthèse additive ou par concaténation par exemple.

B.1 Synthèse soustractive de textures sonores

Dans les modèles de synthèse de textures sonores de type source-filtre, on trouve essentiellement deux approches. La première consiste en un schéma classique analyse/(transformation)/synthèse. Ce type de méthode est souvent relativement général et peut s'appliquer à un grand nombre de textures différentes. La seconde approche se base sur des considérations physiquement informées. Ce type d'approche est souvent dédié à une sous-catégorie de textures particulières, mais a l'avantage de permettre en général de dériver plus facilement des contrôles haut-niveau, la morphologie typique de la texture étant prise en compte dès la conception du modèle de synthèse.

Méthodes d'analyse/synthèse Athineos et Ellis (2003) ont proposé une méthode d'analyse-synthèse basée sur la prédiction linéaire (LPC) sur des fenêtres à court-terme du signal, en temps, puis en fréquence sur la transformée en cosinus discret du signal résiduel. De la même façon que la LPC en temps permet d'estimer l'enveloppe spectrale du signal (comme on le verra plus bas), la LPC en fréquence permet d'estimer l'enveloppe temporelle du signal (Herre et Johnston, 1996). Pour effectuer la synthèse, le schéma inverse est appliqué sur un bruit blanc gaussien. Une variété de textures sonores plutôt bruitées (applaudissements, pluie, feu...) peut ainsi être synthétisée de manière plus fidèle qu'une seule LPC en temps avec le même nombre de coefficients [3]. Zhu et Wyse (2004) ont proposé une modification de ce modèle en considérant que les textures sonores sont composées d'un bruit de fond et d'événements singuliers transitoires. Ces événements singuliers sont détectés via un critère énergétique puis retirés du signal et stockés. La resynthèse du bruit de fond est ensuite similaire à la méthode de Athineos et Ellis (2003), et les événements singuliers sont déclenchés selon une loi de Poisson dont le paramètre est estimé sur le signal analysé. On peut constater qu'on est déjà ici à la limite de la synthèse soustractive et qu'on a plutôt à faire à une synthèse hybride, les événements discrets étant synthétisé par un processus granulaire.

McDermott *et al.* (2009) et McDermott et Simoncelli (2011) ont également proposé une méthode d'analyse/synthèse soustractive de textures sonores par analogie avec le domaine de la vision (Julesz, 1962; Portilla et Simoncelli, 2000). Cette méthode est basée sur des considérations neurophysiologiques permettant d'hypothétiser des calculs probablement effectués par le système nerveux à la sortie système auditif périphérique. Ainsi, lors de la phase d'analyse, la texture est filtrée par un banc de filtres cochléaires, et les enveloppes du signal en sortie de chaque filtre sont calculées. Ensuite, différentes statistiques sont calculées sur ces enveloppes ainsi que des corrélations entre ces enveloppes. La synthèse est ensuite effectuée en imposant les statistiques estimées à un bruit blanc gaussien via une descente de gradient itérative, jusqu'à ce qu'un critère d'erreur soit en dessous d'un seuil fixé par l'utilisateur. Cette méthode produit des resynthèses très satisfaisantes [4] mais est cependant très coûteuse en temps de calcul.

3. `http://www.ee.columbia.edu/~marios/ctflp/ctflp.html`
4. `http://mcdermottlab.mit.edu/texture_examples/index.html`

Enfin, Bruna et Mallat (2013) ont proposé plus récemment une méthode d'analyse/synthèse de textures sonores à partir d'une transformée nommée "scattering", qui consiste à décomposer un signal par un banc de filtres en ondelettes et à calculer les modules de chaque sous-bande, puis à réitérer l'opération au sein de chaque sous-bande (Anden et Mallat, 2014). La phase de synthèse des textures sonores est similaire à la méthode utilisée par McDermott *et al.* (2009); McDermott et Simoncelli (2011), puisque les coefficients obtenus par la transformée, appelés moments de "scattering", sont ensuite imposés progressivement à un bruit blanc gaussien grâce à une méthode de descente de gradient. Les resynthèses sont là aussi très satisfaisantes [5], au prix d'un coup de calcul relativement élevé.

Méthodes purement génératives D'autres auteurs se sont plutôt basés sur des considérations phénoménologiques et physiques (observation de représentations temps-fréquence, questionnements de la nature du processus physique ayant produit le son [6], ...) afin de dériver des modèles de synthèse de textures sonores bien particulières. Verron (2010) et Verron *et al.* (2010) ont notamment proposé, dans le cadre du développement d'un synthétiseur de sons d'environnement spatialisé et contrôlable à haut-niveau, des modèles soustractifs pour les sons de vent et de vagues [7], en se basant sur les travaux de Farnell (2010). Enfin, Peltola *et al.* (2007) ont proposé un modèle source-filtre d'applaudissements de foule. Les paramètres des filtres de "claquements" sont réglés grâce à des enregistrements de claquements uniques, et diverses lois statistiques sont proposées, permettant ainsi la synthèse d'applaudissements de foule et le contrôle de différents niveaux d'enthousiasme et de synchronisation de la foule.

B.2 Autres méthodes de synthèse de textures sonores

Parmi les méthodes de synthèse de textures sonores non basées sur un modèle soustractif, une grande partie considère des méthodes granulaires (Roads, 1988). La méthode granulaire est en effet assez naturelle si l'on considère le point de la définition de Saint-Arnaud et Popat (1995) qui propose les textures sonores sont composés d'éléments sonores simples, des "atomes" [8] .

Parmi ces approches, certains auteurs proposent de modéliser directement ces atomes, comme par exemple pour les sons de liquides (van den Doel, 2004; Farnell, 2010; Verron, 2010; Verron *et al.*, 2010), puis de les déclencher selon des lois statistiques empiriques, provenant de l'observation des signaux. A noter que les approches proposées dans (Verron, 2010; Verron *et al.*, 2010) sont totalement hybrides, combinant à la fois synthèse granulaire d'atomes modélisés (somme de sinusoïdes amorties, chirps, etc)

5. `http://cims.nyu.edu/~bruna/Audio_Texture_Synthesis.html`

6. Voici un exemple de questionnement typique, d'après Andy Farnell (`http://www.obiwannabe.co.uk/tutorials/html/tutorial_rain.html`) : *"What is the nature of rain ? What does it do ?" According to the lyrics of certain shoegazing philosophies it's "Always falling on me", but that is quite unhelpful. Instead consider that it is nearly spherical particles of water of approximately 1-3mm in diameter moving at constant velocity impacting with materials unknown at a typical flux of 200 per second per meter squared. All raindrops have already attained terminal velocity, so there are no fast or slow ones. All raindrops are roughly the same size, a factor determined by their formation at precipitation under nominally uniform conditions, so there are no big or small raindrops to speak of. Finally raindrops are not "tear" shaped as is commonly held, they are in fact near perfect spheres. The factor which prevents rain being a uniform sound and gives rain its diverse range of pitches and impact noises is what it hits. Sometimes it falls on leaves, sometimes on the pavement, or on the tin roof, or into a puddle of rainwater.*

7. `http://www.charlesverron.com/thesis.html`

8. Cette vision du son comme composé "d'atomes" acoustiques a été proposée pour la première fois par Gabor (1947) pour représenter les signaux acoustiques de manière plus cohérente en temps et en fréquence simultanément, contrairement à la représentation adoptée jusque là qui séparait le domaine temporel du domaine de Fourier.

et synthèse soustractive (bruit large bande ou bande étroite). Une approche plus extrême consiste en la simulation physique préalable de sons de liquides, extrêmement coûteuse en temps de calcul, mais permettant des résultats impressionnants en terme de réalisme (Moss *et al.*, 2010) [9].

D'autres auteurs s'attachent à une décomposition plus "mathématique" du signal, sur des bases plus abstraites. Ainsi plusieurs auteurs proposent de décomposer les textures sonores sur une base d'ondelettes (Bar-Joseph *et al.*, 1999; Dubnov *et al.*, 2002; Kersten et Purwins, 2010, 2012) ou sur un banc de filtres en octaves (Saint-Arnaud et Popat, 1995), puis d'effectuer un apprentissage statistique sur les coefficients afin de capturer les relations haut-niveau régissant une texture. Toujours basés sur une décomposition en ondelettes, Miner et Caudell (1997) ont plutôt recherché des règles de paramétrisation des coefficients de la décomposition permettant de générer et contrôler des textures, avec la possibilité de passer par exemple progressivement d'une pluie fine à une pluie forte, ou bien de synthétiser différents types de sons de moteurs.

Une autre approche de la synthèse granulaire, couramment appelée synthèse concaténative, consiste à segmenter en courts fragments une texture ou un corpus de sons puis déclencher ces grains aléatoirement par la suite selon des règles prédéfinies. Ces règles sont généralement définies par des statistiques particulières (voir par exemple les travaux de Bascou et Pottier (2005)), éventuellement modélisées à partir de descripteurs (par exemple à partir du calcul des "Mel-frequency cepstral coefficients" (Lu *et al.*, 2004)). Une approche intéressante consiste à positionner ces grains dans un espace où les dimensions représentent des descripteurs acoustiques (barycentre spectral, intensité, etc) et permettre à l'utilisateur de générer dynamiquement le son en se déplaçant dans cet espace (Schwarz, 2004, 2013; Bernardes *et al.*, 2012).

Enfin, d'autres méthodes de synthèse non-standards ont été proposées, basées sur une approche "système dynamique" du son. Ainsi, Di Scipio (1999) propose de générer des textures sonores inédites et inattendues par itération de fonctions non-linéaires. Mackenzie (1994) propose lui de reconstruire un attracteur à partir de la série temporelle qui représente la texture, permettant ainsi de générer de nouvelles séries temporelles, qu'on peut voir comme des simulations d'autres observations de la même texture.

C Proposition d'un modèle d'analyse/synthèse de textures sonores

Dans cette partie, on proposera un modèle de synthèse de textures sonores à partir de l'analyse d'une texture sonore donnée. Le but est d'avoir un modèle compatible avec notre paradigme action-objet, afin de pouvoir considérer la texture comme un objet sur lequel on pourra interagir (rouler par exemple), et créer ainsi une "métaphore sonore". Plus loin dans ce chapitre, on voudra évaluer perceptivement ce concept de métaphore sonore. Afin de bien isoler le problème, on propose de se limiter à des textures sonores relativement "simples". On choisit donc de se limiter à une représentation de la partie "objet" (i.e. de la texture) fixe, c'est-à-dire que les caractéristiques de la texture seront reproduites par des filtres invariants dans le temps. Cela exclut également les textures constituées en partie d'atomes discrets, comme la pluie (atomes de gouttes) ou le feu (atomes de crépitements) par exemple (voir (Verron, 2010; Verron *et al.*, 2010)).

On suppose donc que les textures sonores sont le résultat d'un processus stochastique (on considèrera un bruit blanc gaussien) filtré par un filtre linéaire invariant dans

9. `http://gamma.cs.unc.edu/SoundingLiquids/`

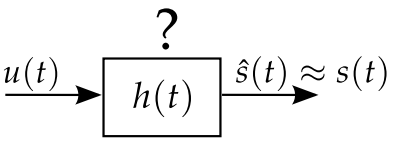

FIGURE IV.2 – Position du problème : étant donnée une texture sonore $s(t)$ connue, comment estimer le filtre $h(t)$ tel que $s(t) \approx \hat{s}(t) = [h * u](t)$, où u est un bruit blanc gaussien de moyenne nulle et où $\hat{s}$ représente la texture estimée ?

le temps. Ainsi, étant donné une texture sonore $s(t)$ connue, le but est d'estimer le filtre $h(t)$ tel que $s(t) \approx \hat{s}(t) = [h * u](t)$, où u est un bruit blanc gaussien de moyenne nulle et où $\hat{s}$ représente la texture estimée (cf figure IV.2).

C.1 Définition du modèle

Une méthode classique et relativement efficace pour modéliser un signal discret $s(n)$ consiste à considérer qu'il est la sortie d'un système prenant en entrée un signal $u(n)$ tel que :

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k) + G \sum_{l=0}^{q} b_l u(n-l) \ , \quad b_0 = 1 \tag{IV.1}$$

où $(a_k)_{1 \leq k \leq p}$, $(b_l)_{1 \leq l \leq q}$ et G sont les paramètres supposés du modèle, i.e. que $s(n)$ est une combinaison linéaire de ses propres valeurs passées ainsi que de la valeur présente et des valeurs passées de l'entrée $u(n)$ (Makhoul, 1975). L'équation (IV.1) peut également s'écrire dans le domaine fréquentiel (plus précisément dans le domaine en z), en considérant la transformée en z :

$$H(z) = \frac{S(z)}{U(z)} = G \frac{1 + \sum_{l=1}^{q} b_l z^{-l}}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{IV.2}$$

où

$$S(z) = \sum_{n=-\infty}^{\infty} s(n) z^{-n} \tag{IV.3}$$

est la transformée en z de $s(n)$ et $U(z)$ la transformée en z de $u(n)$. Dans le cas où les coefficients $(a_k)_{1 \leq k \leq p}$ et $(b_l)_{1 \leq l \leq q}$ sont non-nuls, on a à faire à un système pôle-zéro, aussi appelé modèle Auto-Régressif à Moyenne Ajustée (ARMA). Si les coefficients $(a_k)_{1 \leq k \leq p}$ sont tous nuls, le système ne comporte que des zéros (qui dans le domaine fréquentiel se traduisent par des creux dans le spectre de puissance) et est appelé modèle tout-zéro ou à Moyenne Ajustée (MA). Enfin, si les coefficients $(b_l)_{1 \leq l \leq q}$ sont tous nuls, le système ne comporte que des pôles (qui dans le domaine fréquentiel se traduisent par des pics dans le spectre de puissance) et est appelé modèle tout-pôle ou Auto-Régressif (AR).

C.2 Estimation d'un modèle AR

Dans le cas d'un modèle AR, où l'entrée $u(n)$ est inconnue, le signal $s(n)$ est approché par $\hat{s}(n)$, une combinaison linéaire de ses p valeurs passées, où :

$$\hat{s}(n) = -\sum_{k=1}^{p} a_k s(n-k) \tag{IV.4}$$

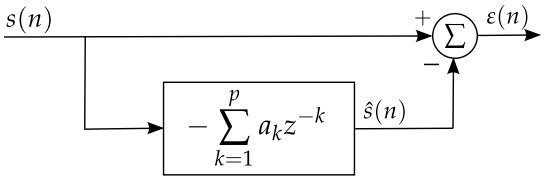

FIGURE IV.3 – Erreur commise lors de la prédiction linéaire.

L'erreur commise, également appelée résiduel, entre le signal mesuré et le signal prédit est donc :

$$\varepsilon(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^{p} a_k s(n-k) \tag{IV.5}$$

et on peut ainsi estimer les coefficients $(a_k)_{1\leq k\leq p}$ en minimisant $\|\varepsilon\|^2$ (cf figure IV.3). Minimiser l'erreur quadratique moyenne revient à résoudre le système suivant, connu sous le nom d'équations de Yule-Walker ou estimation LPC (Makhoul, 1975) :

$$\sum_{k=1}^{p} a_k R_{ss}(i-k) = -R_{ss}(i) \ , \quad 1 \leq i \leq p \tag{IV.6}$$

où (dans le cas des signaux à valeurs réelles) :

$$R_{ss}(i) = \mathbb{E}\left[s(n)s(n-i)\right] \tag{IV.7}$$

est la fonction d'auto-corrélation de s dont un estimateur temporel est :

$$R_{ss}(i) = \sum_{n=-\infty}^{\infty} s(n)s(n-i) \tag{IV.8}$$

L'équation (IV.6) peut s'écrire sous forme matricielle et les coefficients $(a_k)_{1\leq k\leq p}$ sont alors :

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = -\underbrace{\begin{bmatrix} R_{ss}(0) & R_{ss}(1) & \dots & R_{ss}(p-1) \\ R_{ss}(-1) & R_{ss}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_{ss}(1) \\ R_{ss}(1-p) & \dots & R_{ss}(-1) & R_{ss}(0) \end{bmatrix}^{-1}}_{\mathbf{R}^{-1}} \begin{bmatrix} R_{ss}(1) \\ R_{ss}(2) \\ \vdots \\ R_{ss}(p) \end{bmatrix} \tag{IV.9}$$

et le gain G est égal à la valeur RMS du résiduel.

Si le signal s est bien un signal AR, c'est-à-dire s'il est le résultat du filtrage d'un bruit blanc gaussien moyenne nulle et de variance 1 par le filtre :

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{IV.10}$$

on peut alors synthétiser un signal aléatoire ayant les mêmes propriétés statistiques. On voit que comme l'estimation LPC approxime la fonction d'auto-corrélation du signal $s(n)$, elle approxime aussi sa densité spectrale de puissance (qui est définie comme la transformée de Fourier de la fonction d'auto-corrélation du signal). Sur la figure IV.4, on a représenté le spectre d'une voyelle, ainsi que les spectres obtenus par approximation LPC pour différents ordres de modélisation. On peut constater que le degré de

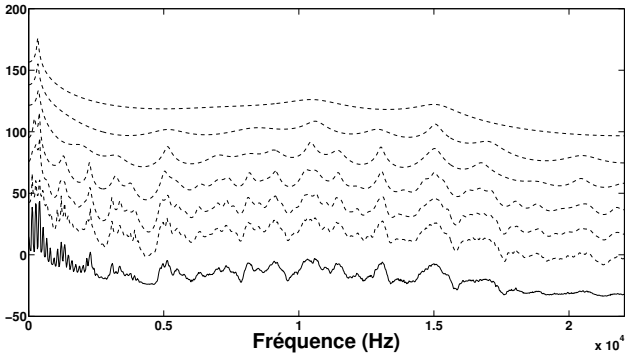

FIGURE IV.4 – Comparaison du spectre obtenu pour différents ordres de modélisation LPC (lignes pointillées) sur une voyelle - "e" français de fenêtre - (trait plein). Les différents ordre de modélisation sont $p = 10, 20, 40, 80, 160, 320$, de haut en bas (décalage vertical ajouté pour la lisibilité). Echelle logarithmique arbitraire en ordonnée.

raffinement augmente avec l'ordre de modélisation [10]. En théorie on peut approximer de manière arbitrairement proche le spectre du signal en augmentant le nombre de pôles de la modélisation. On voit également grâce à cet exemple que la méthode d'estimation ne favorise aucune bande de fréquence, ainsi en augmentant le nombre de pôles, beaucoup d'entre eux sont "gaspillés" pour approximer finement des parties en haute fréquence qui ne sont sûrement pas perceptivement importantes, et afin de bien approximer des portions du spectre sûrement plus importantes d'un point de vue perceptif (comme les pics spectraux entre 0 et 2500 Hz environ), il faudra d'autant plus augmenter l'ordre de modélisation.

L'expérience nous montre que pour avoir des bonnes resynthèses de textures sonores, il nous faut monter à des ordres importants, de l'ordre de 800 pôles. Quand plus loin dans le chapitre on voudra "interagir" avec les textures (rouler, frotter...) tout en conservant leur timbre, on verra qu'il faudra encore augmenter le nombre de pôles afin d'avoir une "réponse impulsionnelle" de la texture plus résonante, ce qui amène à un très grand nombre de pôles.

Afin de limiter le nombre de pôles, on pourrait également considérer des zéros dans un modèle ARMA. Comme on l'a vu précédemment, on peut obtenir une approximation arbitrairement proche du spectre de puissance du signal par un modèle AR en augmentant le nombre de pôles, et si le spectre contient des zéros (creux dans le spectre), alors un grand nombre de pôles sera utilisé pour modéliser ces zéros. Cependant, les méthodes permettant d'estimer des modèles ARMA ne produisent pas toujours une solution stable, en particulier lorsque l'ordre de modélisation commence à être important. On peut toujours se ramener à une solution stable en réfléchissant à l'intérieur du cercle unité les pôles qui sont à l'extérieur (Smith, 2007). Cependant, l'ordre élevé de modélisation pose ici encore un problème. En effet, les solutions fournies par les algorithmes d'estimation pour le dénominateur sont sous la forme d'un polynôme développé, or pour pouvoir réfléchir les pôles instables à l'intérieur du cercle unité, il est nécessaire de factoriser ce polynôme afin d'obtenir les pôles, ce qui est numériquement dangereux lorsque l'on a des polynômes d'ordre élevé. En revanche dans le cas des systèmes AR,

10. La modélisation LPC est très courante en codage de la parole, où l'on considère un petit nombre de pôles pour modéliser l'enveloppe spectrale due au filtrage effectué par le conduit vocal (filtrage formantique). Pour la resynthèse on vient ensuite estimer si on a un signal non-voisé (le résiduel est blanc) ou voisé (le résiduel est un train d'impulsions dont on peut estimer la fréquence).

le système prédit sera plus stable à mesure que la séquence observée $s(n)$ est grande [11] (sauf si le nombre de points utilisé pour l'estimation des coefficients d'auto-corrélation devient proche du nombre de pôles demandé). Comme nous nous limitons à modéliser les textures sonores par un filtrage linéaire invariant dans le temps, on peut utiliser toute la séquence, ce qui nous garantira un résultat stable. Dans la suite, on proposera une estimation LPC en sous-bandes, ce qui aura pour avantage de réduire le nombre de pôles à utiliser, et permettra également une meilleure répartition des pôles sur l'ensemble du spectre.

C.3 Estimation d'un modèle AR en sous-bandes

On choisit de construire un banc de filtres multi-résolution afin de séparer la texture sonore en plusieurs bandes de fréquence de largeur de bande différentes, banc composé de filtres d'analyse et de filtres de synthèse. Deux filtres miroir en quadrature (*Quadrature Mirror Filters*, QMF) sont utilisés comme base pour construire ce banc de filtres d'analyse/synthèse, qui permet d'obtenir en sortie une reconstruction parfaite (RP) du signal d'entrée (Vaidyanathan, 1993). Ces deux filtres, un passe-bas et un passe-haut, sont le miroir l'un de l'autre car ils respectent la relation de complémentarité en puissance :

$$A_1(z) = A_0(-z) \tag{IV.11}$$

où $A_0(z)$ est le filtre passe-bas d'analyse et $A_1(z)$ le filtre passe-haut d'analyse (voir figure IV.5). Afin d'assurer la RP, les filtres de synthèse respectent également la condition d'annulation du repliement spectral :

$$S_0(z) = A_1(-z), S_1(z) = -A_0(-z) \tag{IV.12}$$

où $S_0(z)$ et $S1(z)$ sont les filtres de synthèse. Ainsi, on peut construire les 4 filtres QMF en ne concevant qu'un seul filtre passe-bas de bande passante $[0, \frac{1}{4}]$ (en fréquence normalisée, i.e. $\frac{1}{2}$ est la fréquence de Nyquist). $\downarrow M$ est l'opération de décimation, i.e. ne prendre qu'un échantillon tout les M dans le signal. $\uparrow M$ est l'opération d'interpolation, i.e. insérer $M-1$ zéros entre deux échantillons successifs du signal. L'opération de décimation a pour effet de créer une version du signal qui est la somme de ce signal dont le spectre est dilaté d'un facteur M et de $M-1$ versions dont le spectre est également dilaté d'un facteur M et est décalé de $2\pi k$, $1 \leq k \leq M-1$, i.e. :

$$X(e^{j2\pi f})|_{\downarrow M} = \frac{1}{M}\sum_{k=0}^{M-1} X(e^{j2\pi(f-k)/M}) \tag{IV.13}$$

où f est la fréquence. L'opération d'interpolation quant à elle contracte le spectre du signal d'un facteur M, i.e. :

$$X(e^{j2\pi f})|_{\uparrow M} = X(e^{j2\pi M f}) \tag{IV.14}$$

On voit donc que ces deux opérations successives introduisent du repliement spectral sur le signal. Les conditions (IV.11) et (IV.12) permettent d'annuler ce repliement à la resynthèse. L'opération de décimation est donc intéressante, car elle permet au sein de chaque canal de réassigner la bande de fréquence d'intérêt sur $[0, \frac{1}{2}]$. Ainsi, pour le codage, on peut faire en sorte que l'algorithme ne se concentre que sur certaines portions du spectre, ce qui est particulièrement intéressant quand on utilise des algorithmes qui minimisent une erreur au sens des moindres carrés comme le LPC.

11. La matrice de covariance **R** dans le système d'équations (IV.9) tend alors vers vers la matrice d'auto-corrélation, qui est définie positive, condition pour obtenir un filtre stable (Makhoul, 1975).

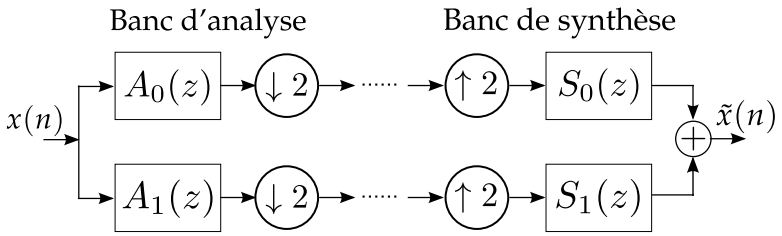

FIGURE IV.5 – Banc de filtres QMF de base.

A partir de ce banc de filtres QMF de base qui sépare le signal en 2 bandes fréquentielles, on peut en créer un qui sépare le signal en plus de bandes fréquentielles en réappliquant le même procédé à l'intérieur d'une branche (entre les opérations de décimation et d'interpolation, pointillés sur la figure IV.5). Pour ce qui suit, on utilisera le banc de filtre utilisé par Lee *et al.* (2010), qui proposent de créer un banc de filtres ayant 4 sous-bandes, de fréquences de coupures $\frac{1}{16}$, $\frac{1}{8}$ et $\frac{1}{4}$, qu'on peut représenter comme un banc de filtres classique (cf figure IV.6) en utilisant les identités nobles (cf figure IV.7) (Vaidyanathan, 1993). Les filtres équivalents d'analyse/synthèse sont finalement :

$$\begin{cases} D_1(z) = A_0(z)A_0(z^2)A_0(z^4) & , \quad R_1(z) = S_0(z)S_0(z^2)S_0(z^4) \\ D_2(z) = A_0(z)A_0(z^2)A_1(z^4) & , \quad R_2(z) = S_0(z)S_0(z^2)S_1(z^4) \\ D_3(z) = A_0(z)A_1(z^2) & , \quad R_3(z) = S_0(z)S_1(z^2) \\ D_4(z) = A_1(z) & , \quad R_4(z) = S_1(z) \end{cases} \tag{IV.15}$$

Afin d'éviter les distorsions de phase introduites par les filtres, on souhaite de plus que les filtres de base du banc QMF soient à phase linéaire, afin que la seule distorsion introduite soit un retard qu'on peut compenser par translation temporelle (retard de groupe constant). Nguyen et Vaidyanathan (1989) ont proposé une paire de filtres à réponse impulsionnelle finie (RIF) permettant la RP tout en ayant une phase linéaire. Ceci est possible en relaxant la condition de complémentarité en puissance (IV.11), i.e. $A_1(z) \neq A_0(-z)$ (Vaidyanathan, 1985). En pratique, on a directement implémenté les filtres RIF de longueur 64 coefficients listés dans (Nguyen et Vaidyanathan, 1989), sans réimplémenter la méthode d'optimisation proposée dans ce papier. Ces filtres ont une atténuation de 42.5 dB dans leur bande de réjection et sont représentés sur la figure IV.8. La réponse du banc de filtres multi-résolution (figure IV.6), est représentée sur la figure IV.9.

La texture sonore finalement synthétisée après l'estimation est donc :

$$\hat{s}(n) = \sum_{k=1}^{4} \hat{s}_k(n + \tau_k) \tag{IV.16}$$

où τ_k est le retard de groupe introduit par les filtres d'analyse et de synthèse dans la sous-bande k et :

$$\hat{s}_k(n) = \left[\left[[u * d_k]_{\downarrow M_k} * h_k \right]_{\uparrow M_k} * r_k \right](n) \quad , \quad 1 \leq k \leq 4 \tag{IV.17}$$

est le signal synthétisé dans la sous-bande k, u est le bruit blanc gaussien en entrée du système pour la resynthèse, d_k et r_k les réponses impulsionnelles filtres d'analyse et de synthèse de la sous-bande k, M_k le facteur de décimation de la sous-bande k, et h_k la réponse impulsionnelle du système AR estimé dans la sous-bande k.

Afin d'apprécier l'avantage de la méthode en sous-bandes, on a tracé sur la figure IV.10 le spectre obtenu avec la méthode en sous-bandes présentée dans cette section

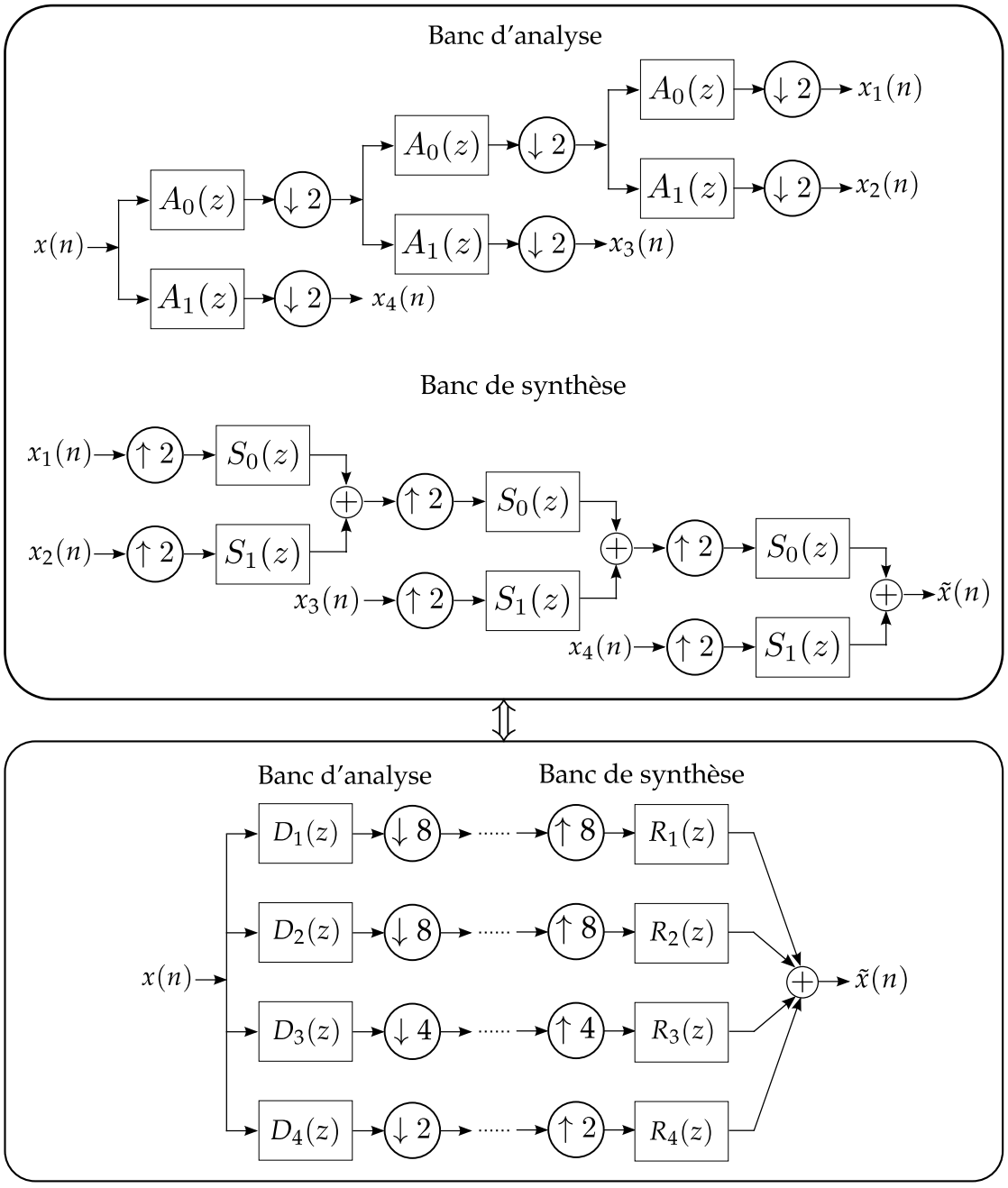

FIGURE IV.6 – Banc de filtres multi-résolution. Haut : Cascade de banc de filtres. Bas : Banc de filtres équivalent.

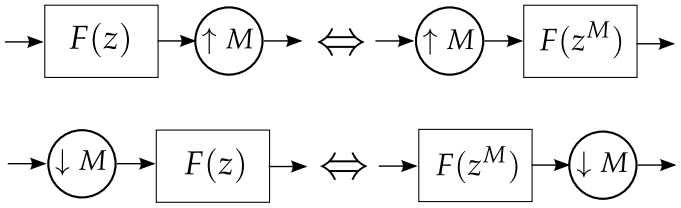

FIGURE IV.7 – Identités nobles permettant de remonter au banc de filtres équivalent à la cascade de filtres.

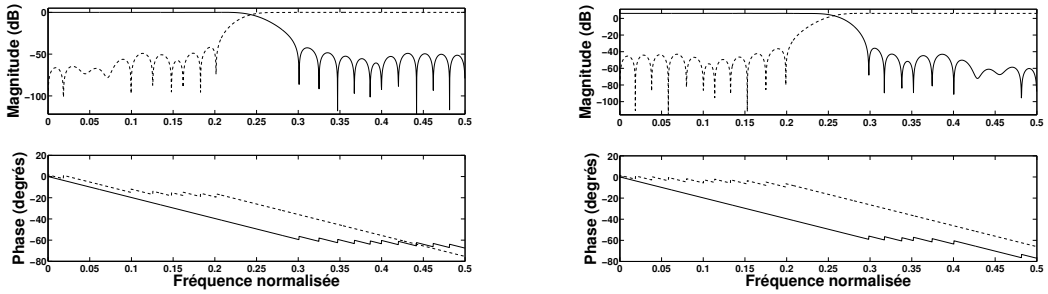

FIGURE IV.8 – Réponse en fréquence des filtres d'analyse (gauche, $A_0(z)$ en trait plein, $A_1(z)$ en pointillés) et de synthèse (droite, $S_0(z)$ en trait plein, $S_1(z)$ en pointillés) du banc QMF à 2 canaux.

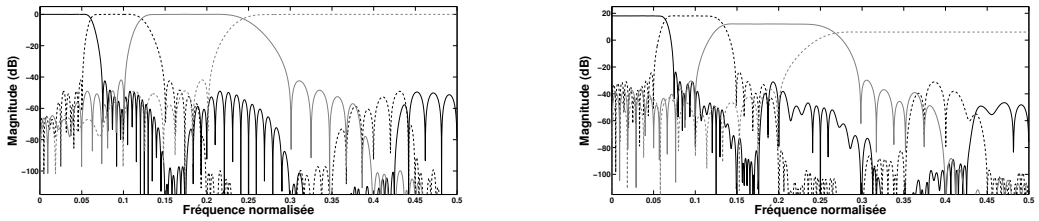

FIGURE IV.9 – Réponse en fréquence des filtres d'analyse (gauche, $D_1(z)$ en trait plein noir, $D_2(z)$ en pointillés noirs, $D_3(z)$ en trait plein gris, $D_4(z)$ en pointillés gris) et de synthèse (droite, $R_1(z)$ en trait plein noir, $R_2(z)$ en pointillés noirs, $R_3(z)$ en trait plein gris, $R_4(z)$ en pointillés gris) du banc QMF à 4 canaux.

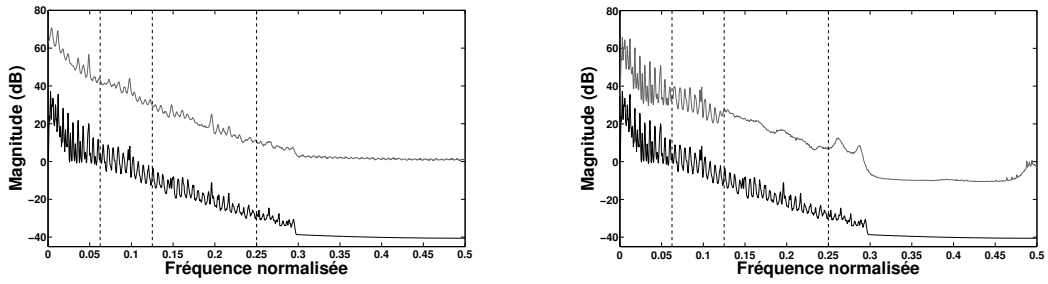

FIGURE IV.10 – Gauche : Spectre d'une texture sonore (noir) et de son estimation LPC en pleine bande (gris) avec 160 pôles. Droite : Spectre d'une texture sonore (noir) et de son estimation LPC en sous-bandes (gris) avec respectivement 80, 60, 10, 10 pôles dans les bandes $[0, \frac{1}{16}]$, $[\frac{1}{16}, \frac{1}{8}]$, $[\frac{1}{8}, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$. Sur chaque figure, un décalage vertical de $+40$ dB a été ajouté à l'estimation pour la lisibilité.

ainsi que celui obtenu par une estimation en pleine bande (section C.2) d'une texture sonore. Le même nombre total de pôles (160) a été utilisé dans les 2 cas, en répartissant les pôles de manière non-uniforme dans le cas des sous bandes (respectivement 80, 60, 10, 10 pôles dans les bandes $[0, \frac{1}{16}]$, $[\frac{1}{16}, \frac{1}{8}]$, $[\frac{1}{8}, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$). Ainsi on a permis que l'algorithme estime finement les parties importantes du spectre en imposant beaucoup de pôles dans les deux premières bandes (entre 0 et environ 5500 Hz) et n'estime que grossièrement les deux sous-bandes haute-fréquence. L'estimation en pleine bande s'est quant à elle attachée à modéliser "au mieux" l'ensemble du spectre au sens des moindres carrés, ce qui donne au final un résultat perceptivement peu satisfaisant, contrairement à la méthode en sous-bandes ($\sharp$).

On a donc proposé un modèle de synthèse de textures sonores cohérent avec notre paradigme action-objet de départ, en considérant que la texture est le résultat d'un bruit blanc gaussien filtré par un banc de filtre linéaire invariant dans le temps. Ainsi on va pouvoir "interagir" avec cette texture en remplaçant le bruit blanc gaussien en entrée par une excitation autre, roulement ou rebond par exemple. C'est l'objet de la section suivante.

D Création de métaphores sonores

Afin de créer des métaphores sonores, on utilisera le principe de la synthèse croisée (Zölzer, 2002). L'hypothèse de base du modèle d'analyse/synthèse de textures sonores est que la texture est le résultat d'un bruit blanc gaussien filtré par filtrage linéaire invariant dans le temps (partie C). On voit donc qu'on peut maintenant "interagir" avec la texture, i.e. la faire rouler, la frotter, la gratter ou bien la faire couiner. En effet, toute l'information retenue sur la texture par notre modèle d'analyse/synthèse est contenue dans la partie filtre (objet) du modèle source-filtre (action-objet). Ainsi, en remplaçant le bruit blanc gaussien en entrée du filtre par un des signaux issus des modèles d'interaction précédemment proposés, on peut "faire rouler" (ou "rouler sur" selon le point de vue) une texture sonore par exemple. Pour la suite de ce chapitre, on introduira quelques notations :

— La texture originale est nommée $\mathbb{T}_j$, et la réponse impulsionnelle du filtre estimé $T_j(t)$.
— Les signaux source évoquant les interactions seront nommés $I^i(t)$, avec les indices $i = \{c, f, r\}$ respectivement pour couiner, frotter et rouler.
— La métaphore obtenue en filtrant le signal évoquant une interaction par le filtre estimé est nommée $M_j^i(t) = \left[I^i * T_j\right](t)$

Ce principe est illustré sur la figure IV.11. Une limite de cette méthode est qu'elle nécessite que les deux signaux aient un support spectral commun, du moins qu'une bonne partie de l'énergie de chacune 2 sources soit contenue dans la même bande fréquentielle, et qu'ainsi la convolution des deux signaux (et donc la multiplication de leurs densités spectrales de puissances respectives) ne résulte pas en un signal n'ayant retenue que trop peu d'informations sur ces deux signaux.

D.1 Contribution de l'interaction

On choisira dans le reste de ce chapitre de se focaliser sur trois interactions qui sont "couiner" (cf section E.1), "frotter" (cf section C) et "rouler" (cf chapitre II). Pour la synthèse croisée, certains auteurs préconisent de blanchir l'excitation avant l'entrée dans les filtres estimés par LPC, i.e. utiliser le résiduel d'une estimation LPC d'ordre faible (typiquement 4 à 6 pôles) sur l'excitation (Moorer, 1979; Keiler *et al.*, 2000). Néanmoins

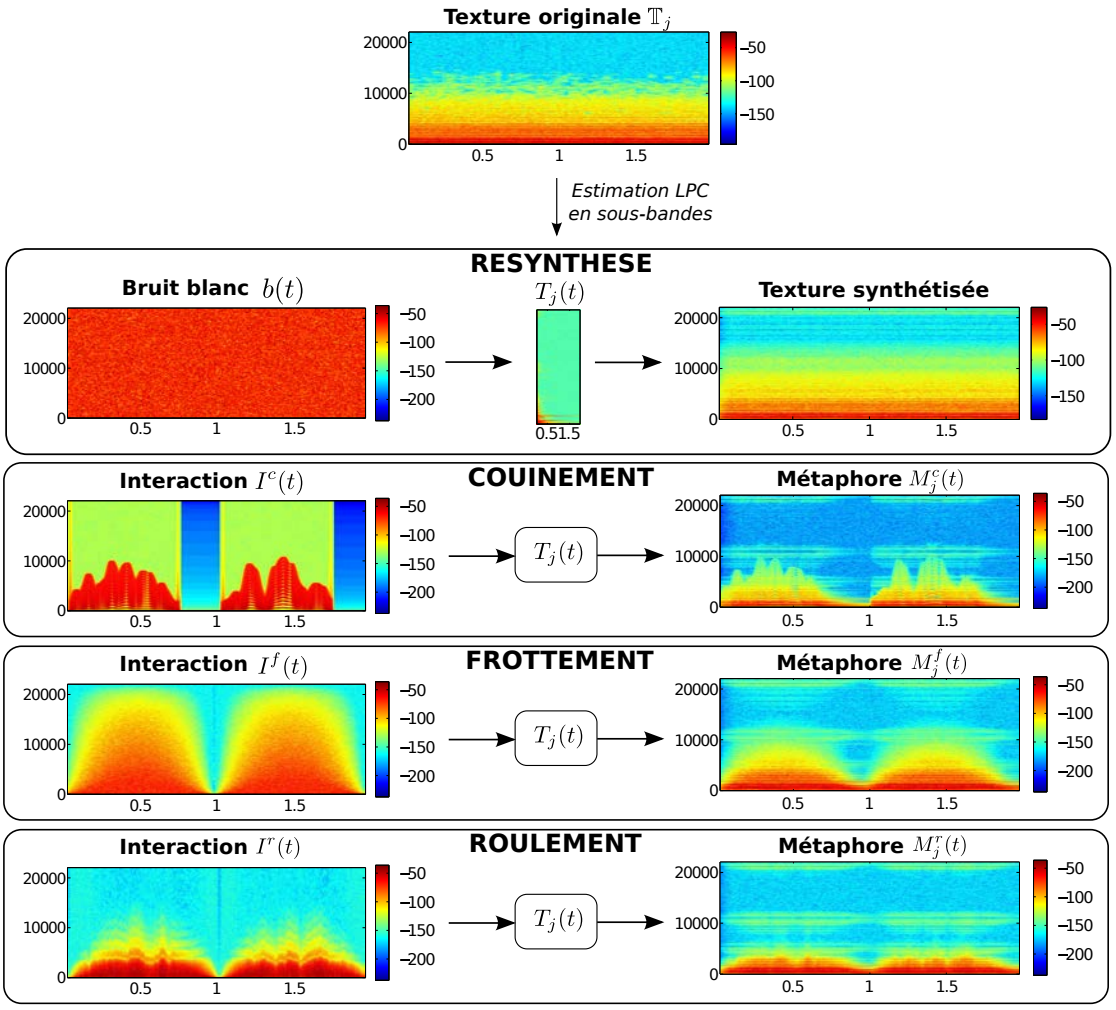

FIGURE IV.11 – Illustration du principe de génération des métaphores sonores. Pour l'estimation on a considéré 320, 240, 160 et 80 pôles respectivement dans l'ordre des bandes de fréquences croissantes. Représentations temps-fréquence avec temps en abscisse (en secondes) et fréquence en ordonnée (en Hz), échelle de couleurs en dB. Les temps-fréquence ont été calculés par transformée de Fourier à court-terme sur une fenêtre de Blackman-Harris de 2048 points avec un recouvrement de 90%.

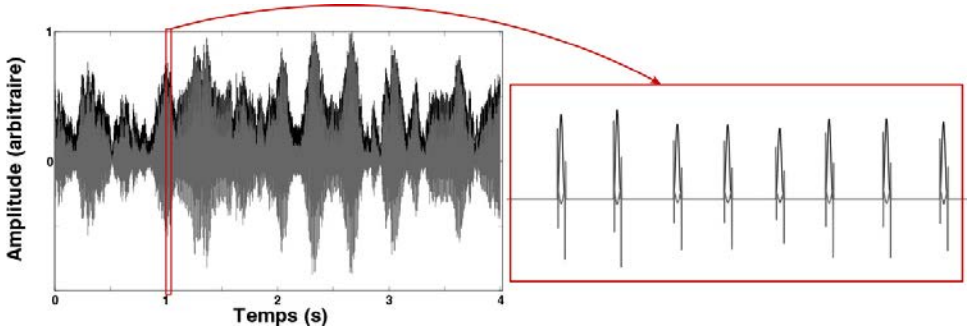

FIGURE IV.12 – Excitation de roulement (noir) et résiduel de la LPC avec 4 pôles sur ce même signal (gris). Les 2 signaux ont été redimensionnés en amplitude par rapport à leur valeur maximale.

cette solution ne sera pas retenue car pouvant nuire à la reconnaissance de l'interaction. Par exemple, dans le cas du roulement, le blanchiment a pour effet global de déconvoluer la forme des impacts ϕ^n (cf équation (II.5)). Un exemple est présenté sur la figure IV.12, où l'excitation liée au roulement (en noir) a été blanchie (en gris) par LPC (4 pôles). En blanchissant on perd donc toute l'information liée à la forme des impacts, et l'interaction est bien moins reconnaissable (♯). Cette possibilité ne sera donc pas retenue.

D.2 Exemples de métaphores sonores

Sur la figure IV.11 on a représenté les spectrogrammes de la texture originale et de sa resynthèse. La texture sonore choisie est un extrait du requiem de Mozart : un choeur accompagné d'un orchestre tenant un long accord. Les représentations temps-fréquence des excitations de roulement, frottement et couinement, et celles des métaphores associées sont également sur la figure IV.11 (♯). Pour chaque interaction, on a imposé un profil de vitesse sinusoïdal de période 1 s, évoquant le geste sous-jacent à l'interaction. On peut voir que pour la métaphore de frottement, à mesure que la vitesse augmente, on retrouve la quasi intégralité de la texture (en comparant par rapport à la représentation temps-fréquence de la texture resynthétisée). Pour la métaphore de couinement, les "modes" de la texture sont sélectivement excités à mesure que les partiels de l'excitation s'en approchent. Le motif temps-fréquence de l'excitation est toujours clairement visible dans le spectrogramme de la métaphore, comme observé par Thoret *et al.* (2013) sur des sons de couinements enregistrés. Enfin, pour le roulement, le motif temps-fréquence de l'excitation seule est toujours légèrement visible dans la métaphore (bien que nettement moins clair que pour le couinement).

Dans la suite du chapitre, on validera le concept de métaphores sonores proposé par une série de trois tests perceptifs. L'expérience 1 a pour but de comparer qui de la méthode proposée ou d'une approche plus "simpliste", qui consiste à mélanger les 2 flux par "addition" (l'excitation liée à l'interaction et la texture sonore), permet une meilleure fusion des 2 sons. Les deux tests suivants permettent d'évaluer dans quelle mesure les caractéristiques respectives de l'interaction et de la texture sont conservées lors de la fusion par la méthode de synthèse croisée proposée. L'expérience 2 évalue donc si les sujets sont capables de reconnaître dans la métaphore l'interaction effectuée (i.e. "couiner", "frotter" ou "rouler"). L'expérience 3 quant à elle évalue si la texture originale est toujours reconnaissable dans la métaphore. Les mêmes sujets ont participé aux trois expériences. L'expérience 1 a toujours été effectuée en premier. L'ordre de

passage des expériences 2 et 3 a été contre-balancé entre les sujets afin d'éviter un effet d'ordre de présentation. Dans la section qui suit on présentera le corpus sonore choisi pour les tests perceptifs.

E Construction du corpus sonore pour les tests perceptifs

E.1 Choix des textures sonores et resynthèse

Douze textures différentes ont été sélectionnées pour les tests perceptifs de validation, provenant de diverses sources (voir annexe C). Ces textures ont été choisies en fonction de leur caractéristiques acoustiques et se déclinent en 3 classes de 4 textures chacunes :

— Les **textures tonales**, i.e. présentant des pics spectraux marqués et piqués. Les 4 textures sont : une note tenue d'accordéon (texture numéro 9), la voyelle française "a" (texture numéro 10), une note tenue de clarinette (texture numéro 11) et un accord de synthétiseur (texture numéro 12).

— Les **textures bruitées** plutôt large bande, i.e. ne présentant pas de pics spectraux clairs. Les 4 textures sont : une nappe de guitare électrique avec beaucoup de distorsion (texture numéro 1), les 3 autres (numérotées de 2 à 4) sont difficilement associables à des sources sonores particulières.

— Les textures entre ces deux catégories, présentant généralement une très grande densité de pics spectraux. On nommera cette catégorie les **textures intermédiaires**. Les 4 textures sont : un son au timbre très rugueux difficilement associable à un événement sonore particulier (texture numéro 5), un accord tenu par un choeur et un orchestre (texture numéro 6), un final d'orchestre (texture numéro 7) et un autre son difficilement associable à une source sonore particulière (texture numéro 8).

Cette catégorisation a été effectuée à la fois en se basant sur les représentations spectrales des textures et sur des tests d'écoute informels. Les représentations temps-fréquence des textures sonores du corpus et des resynthèses sont présentées sur les figures IV.13 (textures bruitées), IV.14 (textures intermédiaires) et IV.15 (textures tonales) (♯). On peut notamment constater que, du fait de la représentation des caractéristiques de la texture sur un banc de filtres invariant dans le temps, la synthèse a pour effet de rendre plus stationnaire la texture originale. Le nombre de pôles par bande est précisé dans les légendes des figures et a été ajusté selon 3 critères : **(1)** permettre une resynthèse perceptivement satisfaisante de la texture sonore (i.e. que la texture sonore soit facilement reconnaissable) ; **(2)** permettre que la texture sonore soit toujours reconnaissable lors de la métaphore ; **(3)** permettre que l'interaction (i.e. l'excitation en entrée du filtre) soit toujours reconnaissable lors de la métaphore. Si on se base uniquement sur le critère **(1)**, la resynthèse peut être bonne mais la texture est parfois difficilement reconnaissable dans le cas de la métaphore car on ne l'entend pas assez "résonner". Il faut donc augmenter le nombre de pôles par bande afin d'avoir une réponse impulsionnelle plus longue. Cependant, si le nombre de pôles est trop élevé, la réponse impulsionnelle est trop longue et a ainsi tendance à masquer la dynamique de l'excitation. Basé sur ces critères, le nombre de pôles a été raffiné par essai-erreur.

D'une manière générale, on peut noter que le nombre de pôles nécessaires est nettement supérieur pour les textures bruitées que pour les textures tonales. On suppose que ceci vient du fait que les textures tonales présentent des pics spectraux très marqués, qui sont de fait très résonants. En effet, si on considère un pic spectral comme modélisé par une sinusoïde exponentiellement amortie de constante de temps τ et de fréquence

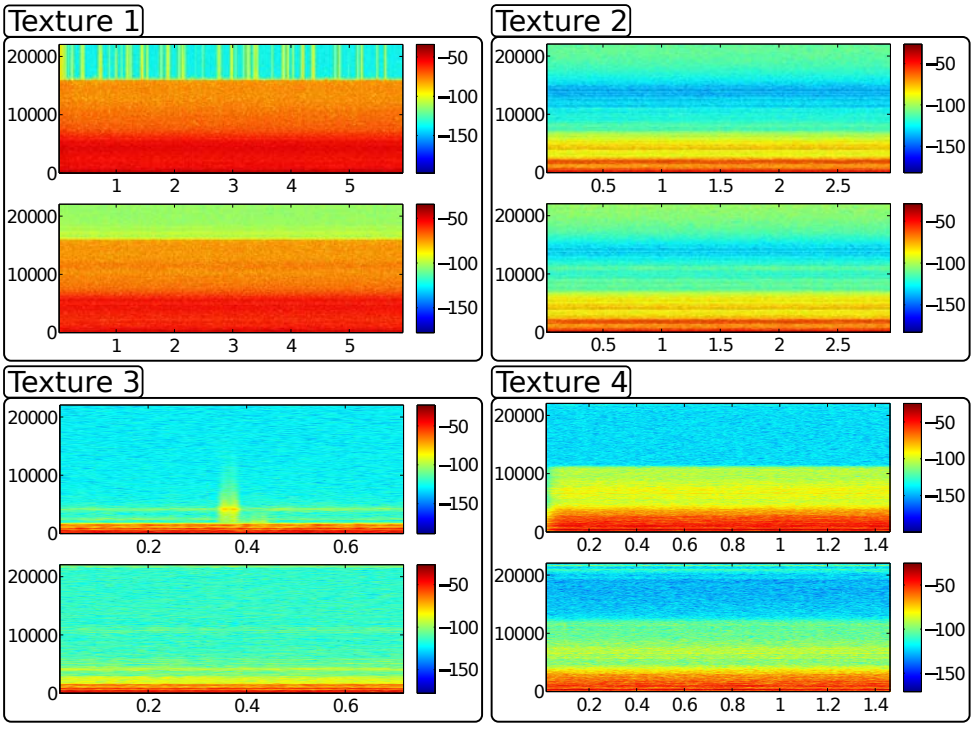

FIGURE IV.13 – Représentations temps-fréquence des 4 textures sonores bruitées, dans chaque cadre on a l'original en haut et la resynthèse en bas. Temps en abscisse et fréquence en ordonnée, échelle de couleurs en dB. Les représentations temps-fréquence ont été calculées par transformée de Fourier à court-terme sur une fenêtre de Blackman-Harris de 2048 points avec un recouvrement de 90%. Nombre de pôles par bande, de la bande de fréquence la plus basse à la plus élevée : $[320, 240, 160, 80]$ pour toutes sauf pour la texture 4 $[240, 180, 120, 60]$.

f, alors sa transformée de Fourier est une fonction Lorentzienne centrée en f et de largeur à mi-hauteur $\frac{1}{\pi\tau}$. Par conséquent, plus la sinusoïde met de temps à s'amortir, plus le pic spectral est étroit et marqué, et inversement. On comprend donc pourquoi il est nécessaire d'avoir un grand nombre de pôles pour que les textures bruitées résonnent plus longtemps, si on voit ces textures comme une somme de sinusoïdes très amorties. Pour les textures 5 à 8, entre ces deux catégories, le nombre de pôles nécessaires est intermédiaire.

E.2 Choix des interactions

Pour les stimuli des tests de validation des métaphores sonores, les interactions "couiner" (cf section E.1), "frotter" (cf section C) et "rouler" (cf chapitre II) ont été choisies. Pour ces 3 interactions, un profil de vitesse sinusoïdal normalisé entre 0 et 1 et de période 1 s a été imposé. 5 périodes ont été considérées pour l'ensemble des stimuli.

Interaction de roulement Pour cette interaction, on a considéré pour tous les stimuli une taille $S = 0.5$ (cf équation (II.17)), une rugosité de la surface $\rho = 0.5$ (cf section D.3), une profondeur de modulation $m = 0.3$ (cf équation (II.7)) et le même mapping pour la fréquence de modulation que dans l'équation II.8. Le même filtrage passe-bas en fonction de la vitesse que pour l'excitation de frottement (voir ci-après) est également appliqué.

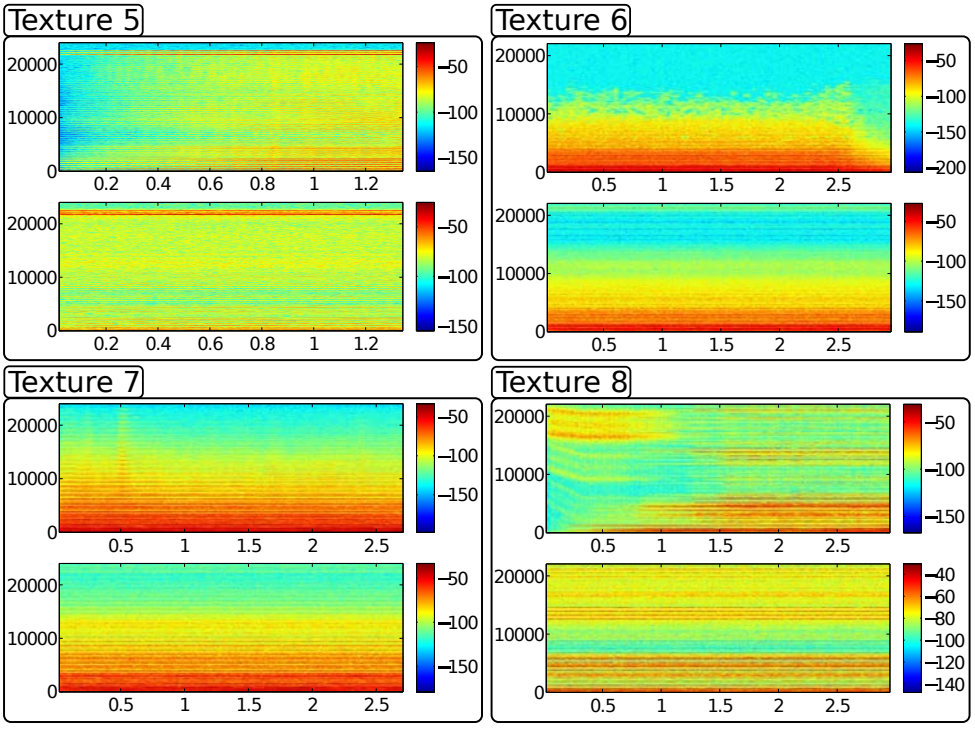

FIGURE IV.14 – Représentations temps-fréquence des 4 textures sonores intermédiaires, légende identique à la figure IV.14. Nombre de pôles par bande, de la bande de fréquence la plus basse à la plus élevée : $[160, 120, 80, 40]$ pour les textures 5 et 8, et $[320, 240, 160, 80]$ pour les textures 6 et 7.

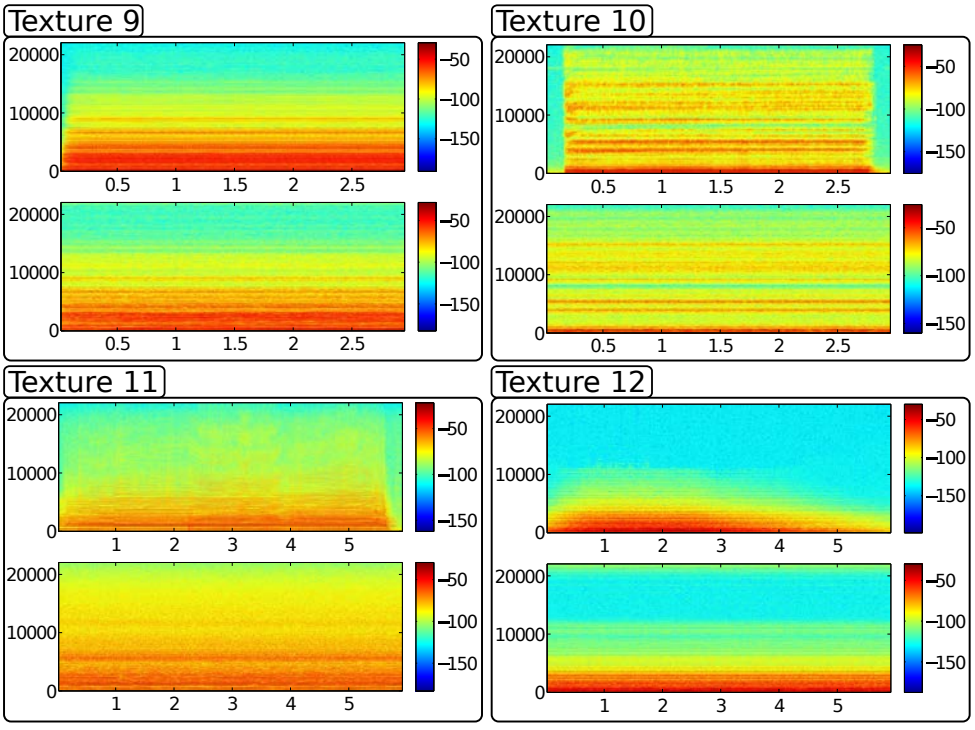

FIGURE IV.15 – Représentations temps-fréquence des 4 textures sonores tonales, légende identique à la figure IV.14. Nombre de pôles par bande, de la bande de fréquence la plus basse à la plus élevée : $[160, 120, 80, 40]$ pour toutes sauf pour la texture 11 $[80, 60, 40, 20]$.

Interaction de frottement Pour cette interaction, on a considéré pour tous les stimuli le filtre à 2 pôles et 2 zéros proposé équations (III.1) et (III.2), avec un facteur de qualité $Q = \frac{1}{\sqrt{2}}$ et une fréquence de coupure variant linéairement avec la vitesse de 0 Hz (vitesse nulle) à 4000 Hz (vitesse normalisée égale à 1).

Interaction de couinement Pour cette interaction, on a considéré pour tous les stimuli une fréquence fondamentale du peigne harmonique variant linéairement avec la vitesse de 0 Hz (vitesse nulle) à 600 Hz (vitesse normalisée égale à 1) et 15 harmoniques en tout. La composante aléatoire de la fréquence fondamentale est un bruit blanc filtré à 20 Hz (filtre Butterworth du second ordre), recentré autour de 0 et redimensionné entre -50 et 50. Enfin, un seuil a été fixé sur la vitesse, de telle sorte que si la vitesse normalisée est inférieure à 0.2, aucun son n'est émis (on paramètre ici la fonction de mapping $\Gamma(v(t), p(t))$ où l'on ne considère que le paramètre de vitesse, cf équation (III.15)).

F Expérience 1 : métaphores sonores vs. mélange des sons

Le mélange "additif" de sons est une approche classique largement utilisée pour créer des effets sonores. Bien qu'un designer sonore expert aille sûrement plus loin qu'un simple mélange des deux sources sonores, en appliquant par exemple la même réverbération aux deux sons pour les faire mieux fusionner ou en leur appliquant un effet Doppler permettant de simuler une source en mouvement (Kronland-Martinet et Voinier, 2008), cela a pour effet de dénaturer d'autant plus la texture sonore originale, c'est-à-dire qu'on n'a déjà plus affaire au même son, car l'ajout d'effet type réverbération ou Doppler apporte des "informations" supplémentaires. L'approche proposée ici est une alternative pertinente à ces techniques dans la mesure où la synthèse croisée permet une fusion à un niveau plus intrinsèque des morphologies sonores. Nous proposons ici d'évaluer si l'approche proposée permet une meilleure fusion de l'interaction et de la texture qu'une approche plus classique qui consiste à mélanger les 2 sons.

F.1 Sujets

16 sujets ont participé à l'expérience : 4 femmes et 12 hommes, 26 ans en moyenne (écart-type : 4 ans). Aucun d'entre eux n'avait eu connaissance des stimuli avant l'expérience ni ne présentait de troubles auditifs.

F.2 Stimuli

36 métaphores M_j^i ont été générées par toutes les combinaisons possibles des textures du corpus ($j = \{1, 2, ..., 12\}$ le numéro de la texture) et des interactions choisies ($i = \{r, c, f\}$ respectivement pour roulement, couinement et frottement), i.e. 3 interactions $\times$ 12 textures (cf section E). 36 mélanges des interactions et textures ont également été générés. Les mélanges ont été réalisés de manière minutieuse afin de pouvoir percevoir les 2 sons sans pour autant que le son d'interaction ne soit prépondérant, ce qui a pour effet de rendre le mélange trop grossier en faisant automatiquement percevoir 2 flux sonores séparés du fait des morphologies différentes de la texture et de l'interaction. Les textures ont été resynthétisées en filtrant un bruit blanc gaussien par les filtres estimés et leur enveloppe a été modulée par le profil de vitesse sinusoïdal, de manière synchrone par rapport aux excitations de roulement, frottement et couinement. Enfin, pour chacune des textures, une nouvelle réalisation de chaque interaction est effectuée

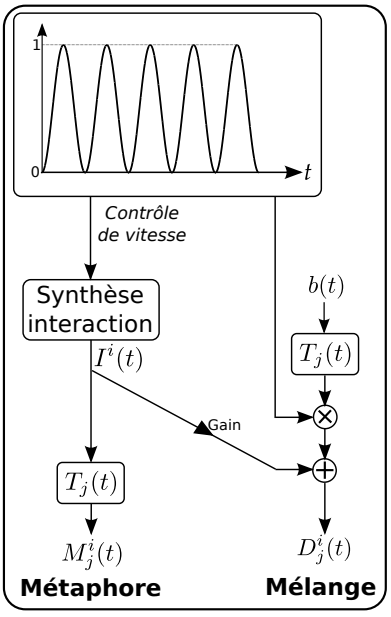

FIGURE IV.16 – Génération de la métaphore $M_j^i(t)$ et du mélange $D_j^i(t)$, obtenus à partir de l'interaction $I^i(t)$ et du filtre $T_j(t)$ estimé sur la texture $\mathbb{T}_j$. $b(t)$ est un bruit blanc gaussien.

(utilisée pour la métaphore et le mélange). Ainsi lors des tests perceptifs les sujets ne peuvent pas se baser sur un motif particulier mais plutôt sur une morphologie sonore particulière. Le schéma présenté sur la figure IV.16 récapitule la génération des stimuli.

F.3 Protocole

Le test s'est déroulé dans une pièce calme, sur un ordinateur portable muni d'un casque Sennheiser HD-650. Un protocole de comparaison par paire a été choisi, chaque paire (M_j^i, D_j^i) étant comparée une seule fois par chaque sujet, et deux textures (ou interactions) différentes n'étaient jamais comparées, soit un total de 36 comparaisons à effectuer. A chaque essai les sujets devaient répondre à la question suivante : "Dans lequel de ces 2 sons entend-on le moins 2 sons différents ?" en sélectionnant leur réponse sur une interface dédiée développée sous MAX/MSP. Avant le test, la consigne a été expliquée aux sujets, i.e. que les sons à comparer étaient tous deux issus d'un mélange de 2 sons différents mais que la manière de les mélanger était différente. Il était également précisé que les 2 sons avant la fusion étaient identiques. Les sons avant mélange et métaphore (i.e. les interactions seules et les textures seules) n'ont pas été présentés aux sujets avant le début du test. Lors de la comparaison, les sujets pouvaient écouter les sons autant de fois que souhaité. Afin d'éviter une influence de l'ordre de présentation des stimuli, 2 séries aléatoires ont été testées, chacune sur la moitié des sujets (i.e. la moitié des sujets se voyait présenter les sons dans un ordre aléatoire, l'autre moitié dans un autre ordre aléatoire).

F.4 Résultats

Afin d'évaluer si globalement les métaphores ont été perçues avec une meilleure fusion des 2 sons que les mélanges, un test de Student unilatéral à un échantillon par rapport au hasard (50%) avec un risque de 5% a été effectué sur l'ensemble des 36 condi-

tions moyennées (moyenne de 85.8% pour les métaphores ; $t(35) = 20.82$; $p < .001$). La métaphore est toujours significativement jugée comme meilleure dans la fusion des 2 sons pour chaque classe de texture (moyennes de 84.4%, 84.4% et 88.5% respectivement pour les textures bruitées, intermédiaires, et tonales, $p < .001$), et chaque type d'interaction (moyennes de 83.3%, 84.4% et 89.6% respectivement pour les interactions couinement, frottement et roulement, $p < .001$). Des tests de Student bilatéraux sur échantillons deux à deux ne montrent pas de différences significatives entre les catégories de textures ($p = 0.31$ au minimum). Des tests de Student bilatéraux sur échantillons deux à deux montrent une différence significative uniquement entre les interactions "couiner" et "rouler" ($t(22) = -2.38$;$p < .05$).

F.5 Discussion

Comme les résultats du test perceptif l'attestent, la méthode proposée de "métaphores sonores" permet une meilleure fusion de l'interaction et de la texture qu'une méthode plus classique qui consiste à mélanger les deux sources sonores. Dans l'analyse de scènes auditives, il a été montré que plusieurs composantes spectrales fusionnent plus aisément si elles ont un "destin commun", e.g. que ces composantes présentent des modulations de fréquence ou d'amplitude similaires (Bregman, 1994). Bien que les sources sonores utilisées dans notre cas soient plus complexes que celles utilisées habituellement dans les tests ayant pour but de mettre en évidence des schémas permettant de fusionner ou ségréguer différentes composantes, on voit bien que dans le cas du mélange des deux sources, le seul "destin commun" est une modulation d'amplitude. Au contraire, dans le cas de la synthèse croisée, la source sonore "interaction", en fonction de l'évolution de son énergie dans le plan temps-fréquence, va directement exciter sélectivement les "résonances" estimées de la texture, ce qui aide à lier les deux sources sonores.

Les comparaisons par type de texture montrent de plus qu'il ne semble pas y avoir de textures parmi celles utilisées pour lesquelles la métaphore permet une moins bonne fusion. Les comparaisons par type d'interaction montrent qu'il existe une différence significative dans le jugement de la fusion texture-interaction entre les interactions "couiner" et "rouler", et la métaphore est moins préférée pour le couinement que pour le roulement, même si les moyennes restent élevées et valident dans tous les cas la préférence des sujets pour les métaphores. On peut avancer une explication probable. Comme détaillé dans la partie E.2, on peut voir l'excitation de roulement comme une somme de sinusoïdes modulées en fréquence et en amplitude. Ces modulations sont indépendantes de la vitesse de l'objet roulant, et n'ont donc pas de lien avec la modulation d'amplitude appliquée à la texture. Dans le cas d'un mélange, le "destin commun" de ces deux sources sonores est donc faible. Basé sur ces morphologies très différentes, le système auditif sépare peut-être ces deux sources sonores. En comparaison, le mélange de la texture modulée et de l'interaction "couiner ", dont la fréquence fondamentale du peigne harmonique augmente avec la vitesse du geste (cf sections E.1 et E.2) et est donc directement proportionnelle à la modulation d'amplitude appliquée à la texture, présente des variations plus cohérentes qui favorisent peut-être l'intégration simultanée des deux composantes du mélange.

Le résultat principal à retenir de ce test perceptif est cependant le fait que la méthode de synthèse croisée proposée pour les métaphores sonores permet une fusion nettement meilleure de l'interaction et de la texture sonore qu'un simple mélange des deux sources.

G Expérience 2 : reconnaissance des interactions dans la métaphore

Le test perceptif précédent a permis de montrer que l'approche proposée pour les métaphores sonores permet une meilleure fusion du son d'interaction (i.e. "rouler", "frotter" ou "couiner") et de la texture sonore. L'objectif de ce test est d'évaluer si la métaphore porte bien l'information quant à l'interaction, i.e. si les sujets sont capables de reconnaître l'interaction dans la métaphore. Comme expliqué précédemment, l'ordre de ce test et du suivant est contrebalancé entre les sujets afin d'éviter un effet d'ordre de présentation.

G.1 Sujets

Les mêmes sujets que dans l'expérience 1 ont passé ce test.

G.2 Stimuli

Les 36 métaphores utilisées dans l'expérience 1 ont été réutilisées dans ce test.

G.3 Protocole

Pour évaluer la capacité de reconnaissance des interactions dans les métaphores, on a opté pour un protocole de test à choix forcé. Pour chaque texture $j = \{1, 2, ..., 12\}$, les 3 métaphores M_j^i générées pour les 3 interactions étaient disposées de manière aléatoire sur l'interface à chaque essai. Les sujets devaient ensuite réassocier chaque métaphore à l'une des trois interactions. Ainsi à chaque essai, seule la nature de l'interaction variait entre les 3 métaphores. Une interface de "glisser-déposer" a été développée spécialement sous le logiciel MAX/MSP (cf figure IV.17). L'association avec les 3 interactions a été effectuée par chaque sujet une fois pour chaque texture, soit un total de 12 essais. Ni les textures seules ni les interactions seules n'ont été présentées aux sujets avant le début du test. Les sujets pouvaient écouter les sons autant de fois que souhaité. Afin d'éviter une influence de l'ordre de présentation des textures, 2 séries aléatoires ont été testées, chacune sur la moitié des sujets.

G.4 Résultats

Afin d'évaluer les résultats du test, on s'intéresse aux matrices de confusion, qu'on nommera respectivement $\mathbb{M}_B$, $\mathbb{M}_T$ et $\mathbb{M}_I$ pour les catégories de textures bruitées, tonales et intermédiaires, ainsi que $\mathbb{M}_{tot}$, la matrice de confusion globale. Ces matrices sont définies dans le tableau IV.1. Les coefficients sont définis de la manière suivante : chaque association est notée 1 et 0 sinon ; on obtient ainsi une matrice binaire pour chaque texture et chaque sujet ; les matrices sommées par catégorie de textures pour chaque sujet ; les matrices sont enfin moyennées sur tous les sujets puis normalisées (de manière à avoir un total de 100% pour chaque ligne et colonne). Les matrices de confusion par catégorie de texture obtenues sont :

$$\mathbb{M}_B = \begin{bmatrix} \mathbf{100} & 0 & 0 \\ 0 & \mathbf{92.19} & 7.81 \\ 0 & 7.81 & \mathbf{92.19} \end{bmatrix}, \mathbb{M}_I = \begin{bmatrix} \mathbf{100} & 0 & 0 \\ 0 & \mathbf{95.31} & 4.69 \\ 0 & 4.69 & \mathbf{95.31} \end{bmatrix}, \mathbb{M}_T = \begin{bmatrix} \mathbf{96.88} & 3.12 & 0 \\ 1.56 & \mathbf{96.88} & 1.56 \\ 1.56 & 0 & \mathbf{98.44} \end{bmatrix} \quad \text{(IV.18)}$$

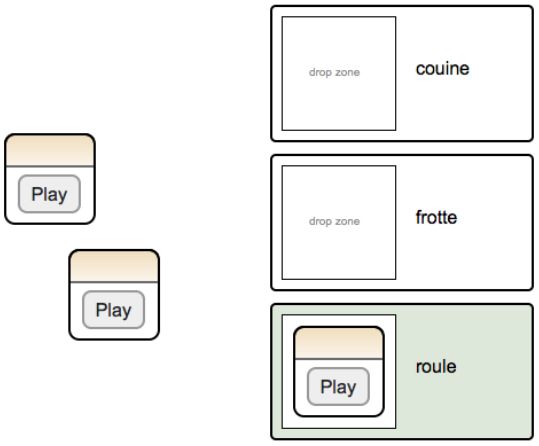

FIGURE IV.17 – Interface de test de l'expérience 2. Les métaphores M_j^i (boîtes avec l'icône "Play" qui permet de jouer le son) doivent être réassociées aux interactions de référence.

TABLE IV.1 – Construction des matrices de confusion dans l'association d'une métaphore $M_j^{c,f,r}$ avec une interaction $I^{c,f,r}$ pour l'expérience 2. Une matrice $\mathbb{M}_B$, $\mathbb{M}_I$ et $\mathbb{M}_T$ est construite pour chaque catégorie de texture, avec j respectivement de 1 à 4, de 5 à 8 et de 9 à 12, et une matrice de confusion globale $\mathbb{M}_{tot}$ avec j de 1 à 12.

	I^c	I^f	I^r
M_j^c	.	.	.
M_j^f	.	.	.
M_j^r	.	.	.

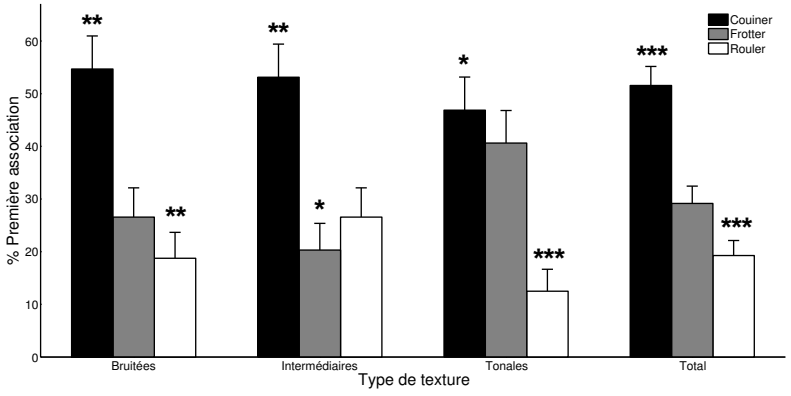

FIGURE IV.18 – Résultats du test sur le nombre moyen de fois où chaque interaction a été classifiée en premier (moyenne et erreur type). Les étoiles montrent si le son a significativement été en moyenne plus ou moins associé en premier avec $p < 0.05$ (1 étoile), $p < 0.01$ (2 étoiles) ou $p < 0.001$ (3 étoiles).

et la matrice de confusion globale obtenue est :

$$\mathbb{M}_{tot} = \begin{bmatrix} \mathbf{98.96} & 1.04 & 0 \\ 0.52 & \mathbf{94.79} & 4.69 \\ 0.52 & 4.17 & \mathbf{95.31} \end{bmatrix} \tag{IV.19}$$

où toutes les valeurs en gras sont les associations significativement au-dessus du hasard (33%) à $p < .001$ (t-test unilatéral à un échantillon avec un risque de 5%).

Une analyse statistique de l'ordre de classification des interactions, et en particulier du nombre moyen de fois où chaque interaction (i.e. couiner, frotter ou rouler) a été associée en premier, a également été effectuée. Pour chaque interaction (dans chaque type de textures et sur le total), un t-test bilatéral à un échantillon avec un risque de 5% par rapport à 33% (i.e. chacune des 3 interactions a été associée autant de fois en premier) est calculé. Les résultats sont présentés sur la figure IV.18.

G.5 Discussion

Les résultats du test nous permettent de conclure que les trois interactions sont très bien reconnues quelque soit la catégorie de texture utilisée (taux d'association >90% quelque soit l'interaction). Ces résultats montrent que, malgré le fait que les métaphores produites évoquent des percepts "irréels" (e.g. croisement d'une interaction de roulement et d'une nappe du requiem de Mozart) par rapport à des situations du quotidien (e.g. une bille qui roule sur une plaque en bois), les morphologies sonores proposées précédemment dans cette thèse contiennent l'information nécessaire aux auditeurs pour reconnaître l'interaction dans une tâche de catégorisation forcé.

Ces scores élevés révèlent probablement un effet plafond. En revanche, cela ne préjuge pas forcément de la facilité de la tâche étant donné le protocole de test choisi. En effet, le protocole de test d'association implique une interdépendance des évaluations, et il peut y avoir par exemple une interaction qui est toujours associée par défaut (toujours en dernier, et donc jamais en premier). Les résultats de l'analyse sur le nombre moyen de fois où chaque interaction a été classifiée en premier montrent que l'interaction "couiner" a été plus souvent associée en premier, ce qui peut signifier qu'elle a été la première interaction reconnue. Il est probable que ceci soit dû à sa morphologie

sonore très différente des deux autres interactions. En effet, le roulement et le frottement présentent une similitude qui est la variation de la fréquence de coupure du filtre passe-bas en fonction de la vitesse, et ont ainsi sensiblement la même variation de leur support fréquentiel au cours du temps. Globalement, l'interaction "rouler" a été moins associée en premier que les 2 autres. Cela n'implique pas qu'elle n'a pas été reconnue (elle a été associée en premier dans 20% du nombre total d'essai), mais laisse supposer qu'elle est plus difficile à reconnaître dans cette tâche. Afin d'évaluer la difficulté, un protocole de test à choix forcés (i.e. on ne présente qu'un son à la fois et le sujet doit choisir quelle interaction le son lui évoque parmi une liste) serait intéressant.

H Expérience 3 : reconnaissance des textures originales dans la métaphore

Le test précédent a permis de montrer que les interactions "couiner", "frotter" et "rouler" étaient très bien reconnues dans les métaphores créées par la méthode de synthèse croisée proposée. L'objectif de ce test est d'évaluer si la texture modifiée par la méthode de synthèse croisée est toujours reconnaissable, i.e. si elle a conservé une partie suffisante du timbre original pour pouvoir être réassociée à la texture originale. Comme expliqué précédemment, l'ordre de ce test et du précédent est contrebalancé entre les sujets afin d'éviter un ordre de présentation des interactions.

H.1 Sujets

Les mêmes sujets que dans les expériences 1 et 2 ont passé ce test.

H.2 Stimuli

Les 36 métaphores des expériences 1 et 2 ont été évaluées dans ce test. Les 12 textures ont également été resynthétisées avec les paramètres proposés sur les figures IV.13, IV.14 et IV.15 (i.e. résultat d'un bruit blanc gaussien filtré par les filtres estimés), toutes d'une durée de 3 secondes.

H.3 Protocole

La capacité des sujets à reconnaître les textures sonores contenues dans les métaphores a été évaluée par un protocole de catégorisation forcée. Pour chaque interaction et chaque catégorie de texture, les 4 stimuli M_j^i générés pour les 4 textures de la catégorie (i.e. $j = \{1,2,3,4\}$ ou $j = \{5,6,7,8\}$ ou $j = \{9,10,11,12\}$) étaient disposés de manière aléatoire sur la gauche de l'interface de test (voir figure IV.19). Les textures "statiques" étaient disposées dans les cases de droite. Les sujets devaient réassocier les métaphores M_j^i aux textures statiques grâce à l'interface de "glisser-déposer". Chaque ensemble de 4 stimuli M_j^i a été jugé une seule fois par sujet, soit un total de 9 essais. Ainsi, à chaque essai, seule la nature de la texture sonore variait entre les 4 métaphores. Ni les textures seules ni les interactions seules n'ont été présentées aux sujets avant le début du test. Les sujets pouvaient écouter les sons autant de fois que souhaité. Afin d'éviter une influence de l'ordre de présentation des stimuli, 2 séries aléatoires ont été testées, chacune sur la moitié des sujets.

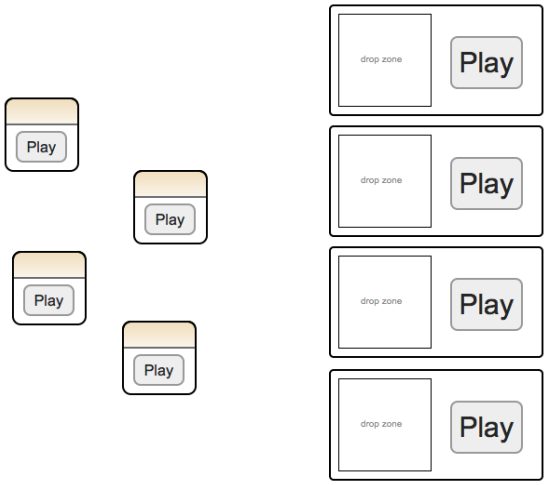

FIGURE IV.19 – Interface de test de catégorisation forcée pour évaluer la reconnaissance des textures. Les métaphores M^i_j (à gauche) doivent être réassociées aux textures "statiques" de référence (à droite).

TABLE IV.2 – Construction des matrices de confusion pour l'expérience 3. Une matrice $\mathbb{M}^i_B$, $\mathbb{M}^i_I$ $\mathbb{M}^i_T$ est construite pour chaque catégorie de texture et interaction i, avec respectivement $\{j_1, j_2, j_3, j_4\} = \{1, 2, 3, 4\}$ ou $\{5, 6, 7, 8\}$ ou $\{9, 10, 11, 12\}$ les numéros de texture.

	T_{j1}	T_{j2}	T_{j3}	T_{j4}
M^i_{j1}	.	.	.	.
M^i_{j2}	.	.	.	.
M^i_{j3}	.	.	.	.
M^i_{j4}	.	.	.	.

H.4 Résultats

Comme pour le test précédent, afin d'évaluer la capacité de reconnaissance des textures sonores, on s'intéresse aux matrices de confusion. Pour chaque interaction et chaque catégorie de texture (i.e. pour chaque essai du test), ces matrices sont construites comme dans le test précédent, et on nomme de même ces matrices $\mathbb{M}^i_B$, $\mathbb{M}^i_T$ et $\mathbb{M}^i_I$ pour les catégories de textures bruitées, tonales et intermédiaires, et l'exposant i représente l'interaction. Ces matrices sont définies dans le tableau IV.2. Les matrices sont remplies de la même manière que dans l'analyse des résultats précédents (cf section G.4). Les matrices de confusion obtenues les catégories de textures bruitées et tonales (i.e. les résultats sont sommés sur les 3 interactions différentes) sont :

$$\mathbb{M}_B = \begin{bmatrix} \mathbf{97.92} & 0 & 2.08 & 0 \\ 0 & \mathbf{100} & 0 & 0 \\ 2.08 & 0 & \mathbf{95.84} & 2.08 \\ 0 & 0 & 2.08 & \mathbf{97.92} \end{bmatrix}, \mathbb{M}_T = \begin{bmatrix} \mathbf{97.92} & 0 & 2.08 & 0 \\ 0 & \mathbf{100} & 0 & 0 \\ 2.08 & 0 & \mathbf{97.92} & 0 \\ 0 & 0 & 0 & \mathbf{100} \end{bmatrix} \tag{IV.20}$$

et la matrice de confusion $\mathbb{M}_I$ pour la catégorie de textures intermédiaires est l'identité (100% de reconnaissance pour toutes les textures). Les valeurs en gras sont les associations significativement au-dessus du hasard (25%) à $p < .001$ (t-test unilatéral à un

échantillon avec un risque de 5%).

H.5 Discussion

Les résultats de ce test montrent que la méthode proposée pour les métaphores sonores permet de conserver suffisamment le timbre des textures analysées pour que celles-ci soient clairement reconnaissables, et ce bien que le spectre des excitations de roulement, de frottement ou de couinement ne soient pas large bande et donc n'excitent pas l'intégralité du spectre des textures sonores initiales. On observe dans ce test également un effet plafond. En effet, plusieurs sujets ont rapporté que ce test leur avait semblé assez facile. Cette facilité peut s'expliquer notamment du fait que les sujets pouvaient se baser sur d'autres indices que le timbre de la texture, comme par exemple la hauteur tonale pour les textures tonales ou bien la bande passante des textures. Afin d'étudier plus profondément la capacité de reconnaissance, il serait nécessaire d'étudier des textures sonores choisies de telle sorte que les sujets ne puissent se baser sur ce genre d'indices (e.g. hauteur tonale, bande passante), ou de modifier les textures utilisées ici dans ce but.

I Discussion générale

Dans ce chapitre, on a proposé une méthode générique permettant de faire des "métaphores sonores", c'est-à-dire d'effectuer des interactions comme le roulement ou le frottement sur des textures sonores, et ainsi créer des sons inouïs. Une méthode d'analyse-synthèse, basée sur la prédiction linéaire en sous-bandes, a été développée. Cette méthode permet de resynthétiser un certain nombre de textures, en considérant qu'elles sont le résultat d'un bruit blanc filtré par un filtre linéaire invariant dans le temps. Ainsi, les caractéristiques de la texture sonore analysée sont contenues dans la partie filtre du modèle source-filtre. Dans le cadre du paradigme action-objet sur lequel se sont basés les travaux de cette thèse, la texture s'inscrit donc dans la partie objet. En remplaçant le bruit blanc en entrée du filtre estimé (i.e. en modifiant la partie action du paradigme action-objet) par un signal évoquant une interaction particulière, on crée ainsi la métaphore sonore. La méthode proposée permet en outre d'étendre à une durée quelconque les textures sonores analysées.

Afin d'évaluer la méthode proposée, une série de trois tests perceptifs a été effectuée. Un corpus de textures sonores présentant des caractéristiques acoustiques variées et pour la plupart issues "du commerce" (i.e. obtenues à partir de divers sources, cf annexe C et non préparées par nos soins) a été construit et trois interactions ont été choisies : "rouler", "frotter" et "couiner". Le premier test perceptif a permis de montrer que la méthode de synthèse croisée proposée permet une meilleure fusion de l'interaction et de la texture qu'un simple mélange des deux sources sonores par "addition". Le second test perceptif a permis de montrer que, dans une tâche d'association, les 3 interactions (frottement, couinement et roulement) sont bien reconnues dans la métaphore sonore, et ce malgré le côté "irréel" des sons, i.e. les interactions ne sont pas associées avec des objets du quotidien (e.g. plaque en métal ou assiette en céramique) mais sur des textures sonores (e.g. nappe d'orchestre, de synthétiseur, textures sonores abstraites...). Pour aller plus loin dans la validation de la reconnaissance des interactions, on pourrait proposer un test de catégorisation (choix entre plusieurs interactions) en ne présentant aux sujets qu'un seul son à la fois. Enfin le dernier test a permis de montrer que les textures sélectionnées étaient toujours reconnaissables dans les métaphores sonores.

D'un point de vue des applications, il est prévu que le partenaire industriel du projet ANR MétaSon se base sur les travaux proposés afin de construire des sons informatifs évoquant notamment le roulement à partir de textures sonores générées par ailleurs pour la sonification des véhicules électriques. Afin d'étendre la méthode pour qu'elle soit applicable à un plus grand nombre de textures sonores, par exemple à des sons évoluant dans le temps (variation du barycentre spectral au cours du temps par exemple), la méthode proposée peut aisément se décliner en une version où les paramètres sont estimés par trame temporelle. Pour étendre les possibilités à des sons plus "granulaires" contenant une multitude d'événements discrets (pluie par exemple), une méthode comme celle proposée par Athineos et Ellis (2003) où est appliquée par trame temporelle une prédiction linéaire en temps (estimation de la densité spectrale de puissance du signal) puis en fréquence (estimation de l'enveloppe temporelle du signal) serait intéressante à envisager.

Conclusion et Perspectives

Les travaux présentés dans cette thèse ont fait appel à plusieurs domaines scientifiques : de la modélisation physique, du traitement du signal, de la psychologie expérimentale (tests d'écoute) et des statistiques (analyse de données). Afin de conclure cette thèse, on récapitulera tout d'abord les apports de cette thèse, puis on élargira en proposant des perspectives sur ces travaux.

Conclusion

Dans cette thèse, on s'est donc tout d'abord intéressé au développement de modèles de synthèse de sons d'interactions solidiennes et du contrôle intuitif associé. Ces travaux se sont inscrits dans le cadre du paradigme *action-objet* de description sémantique des sons, basé sur l'approche écologique de la perception, qui stipule l'existence d'*invariants perceptifs* : les *invariants structurels*, liés aux objets résonants, et les *invariants transformationnels*, liés aux actions effectuées par/sur ces objets. Cette description du son en terme d'actions et d'objets est particulièrement propice à une implémentation *source-filtre* des modèles de synthèse, en décrivant les invariants structurels, liés aux objets résonants, dans la partie filtre et les invariants transformationnels, liés aux actions, dans la partie source.

Les travaux de cette thèse se sont dans un premier temps particulièrement focalisés sur la partie action du modèle, i.e. à la modélisation de signaux pour la partie source du modèle de synthèse. Ainsi, dans les chapitres II et III, les morphologies des signaux permettant l'évocation des interactions "rouler", "frotter" et "gratter" ont été étudiées. Ces études ont permis de proposer un synthétiseur temps-réel, contrôlable à haut-niveau, et permettant notamment des transitions continues entre ces trois interactions. La figure IV.20 schématise le synthétiseur final, et les contrôles haut-niveau et bas-niveau associés. Ces transitions ont notamment été rendues possible en observant dans un premier temps que ces interactions sont toutes trois des suites d'impacts ayant des propriétés statistiques particulières. Basé sur ces observations, un modèle génératif unique permettant de synthétiser des suites d'impacts tout en contrôlant leurs statistiques a été proposé et validé par des tests perceptifs.

Dans la dernière partie de cette thèse (chapitre IV), on s'est plutôt intéressé à la partie objet du modèle pour la proposition de métaphores sonores. Les métaphores sonores ont été développées du fait de la demande du partenaire de l'industrie automobile du projet ANR MétaSon, qui fait clairement apparaître une demande de pouvoir créer des textures sonores avec un contenu sémiotique spécifique. Notamment dans le cas de la sonification des voitures électriques, l'évocation du "rouler" est cruciale, tout en laissant une liberté créative au designer sonore au niveau du timbre des sonorités, marque de l'identité du véhicule. On a ainsi proposé une méthode d'analyse synthèse de textures sonores où les caractéristiques de la texture sont représentées dans la partie filtre du paradigme action-objet. Ceci permet alors d'interagir sur la texture avec les diffé-

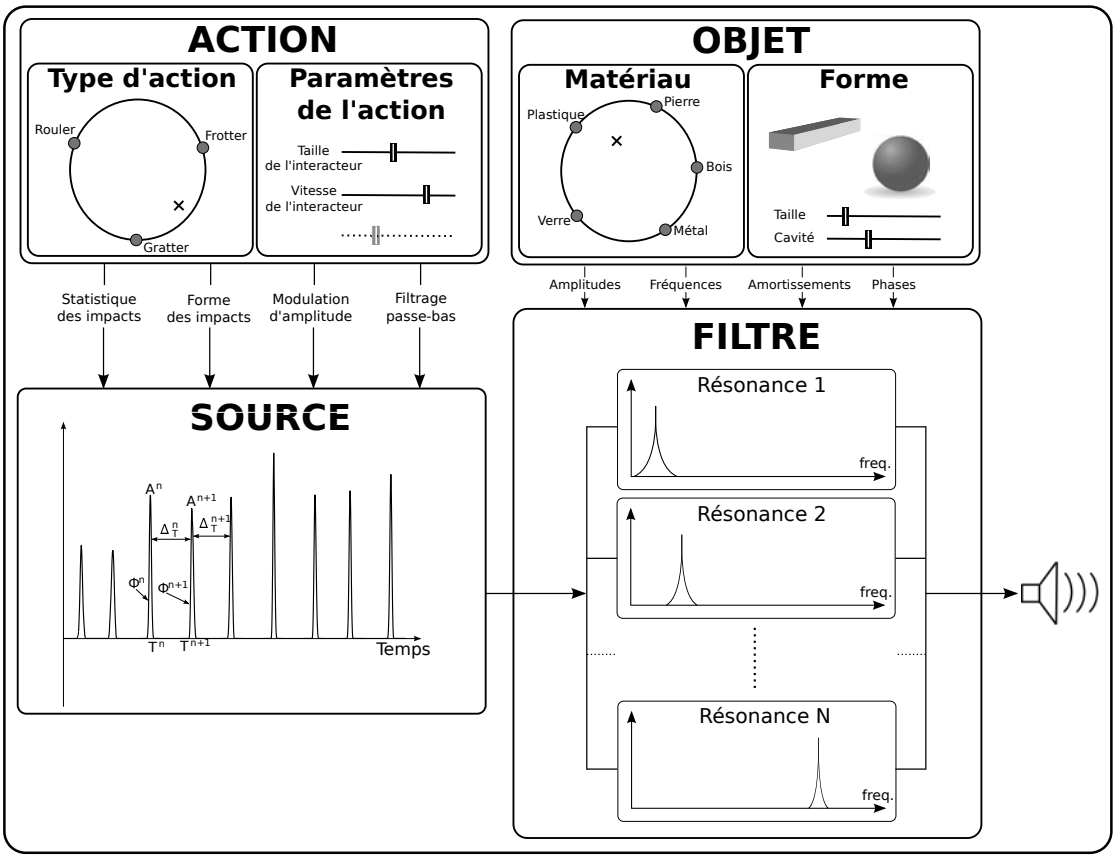

FIGURE IV.20 – Schématisation du synthétiseur final et des contrôles haut-niveau et bas-niveau associés.

rentes interaction proposées, “rouler” ou “frotter” par exemple. Le modèle proposé a par la suite été validé par une série de tests perceptifs montrant : **1)** que la méthode proposée permet une meilleure fusion de la texture et de l’interaction qu’un mélange additif des deux sources sonores ; **2)** que l’interaction spécifique (ici “frotter”, “rouler” ou “couiner”) est toujours reconnaissable dans la métaphore sonore ; **3)** que la texture est toujours reconnaissable dans la métaphore sonore.

Perspectives

Les modèles développés dans cette thèse se sont basés sur l’analyse d’un modèle physique, ainsi que sur une analyse qualitative de sons d’interactions enregistrés. Il pourrait être intéressant de pousser ces travaux en développant une méthode d’analyse/synthèse sur des sons enregistrés, comme proposé par Lagrange *et al.* (2010) et Lee *et al.* (2010), ou en partant sur des méthodes d’analyse n’ayant pas encore été envisagées pour ce type de sons, comme le “matching-pursuit” (Mallat et Zhang, 1993; Daudet, 2006), ce qui pourrait permettre d’extraire l’invariant transformationnel du signal. Ces méthodes pourraient être pré-informées grâce aux invariants transformationnels mis en avant dans cette thèse. De plus, une méthode fournissant une bonne extraction de la source, combinée avec les morphologies des invariants transformationnels, serait d’intérêt pour la reconnaissance automatique de l’interaction effectuée, par exemple dans le domaine de l’analyse automatique de scènes sonores (*Computational Auditory Scene Analysis*, cf. (Wang et Brown, 2006) par exemple).

Une autre perspective intéressante serait de comparer les morphologies des invariants transformationnels à des imitations vocales produites par des sujets. En effet, comme montré par Gygi *et al.* (2004), la bande fréquentielle $[1200, 2400]$ Hz est la plus importante pour la reconnaissance des sons du quotidien, ce qui suggère que les informations pertinentes pour la reconnaissance de ces sons peuvent être imitées vocalement. L'efficacité de ces imitations vocales pour la reconnaissance de sons a été par la suite montrée par Lemaitre et Rocchesso (2014). Ces résultats suggèrent donc que les imitations vocales contiennent de l'information pertinente pour reconnaître les sons du quotidien (on s'imagine par exemple aisément qu'on retrouvera des similarités acoustiques entre des sons de frottement ou de grattement enregistrés, des imitations vocales de ces mêmes sons, et les morphologies sonores mises en avant dans cette thèse).

Suied *et al.* (2014) ont montré que quelques millisecondes seulement étaient nécessaires pour reconnaître la catégorie d'un son (e.g. voix, instrument, percussion...). Ces résultats impliquent que les indices de timbre pour la reconnaissance des sons sont disponibles à différentes échelles temporelles, et en particulier des échelles de temps très courtes. Un tel protocole sur des sons d'interactions continues permettrait probablement de mettre en avant le temps minimum nécessaire pour percevoir l'invariant transformationnel lié à l'interaction continue spécifique, et donner ainsi une indication sur l'échelle temporelle à considérer pour rechercher ces invariants transformationnels. L'approche "esquisse auditive" (Suied *et al.*, 2013) pourrait également apporter à la recherche des invariants. Cette approche consiste à construire une représentation parcimonieuse du signal, en sélectionnant un petit nombre d'atomes à partir d'une représentation obtenue par un modèle auditif, de manière à ce que le son obtenu soit très appauvri mais toujours reconnaissable. Ainsi, si "l'esquisse auditive" obtenue est toujours reconnue, on peut raisonnablement supposer que le nécessaire pour retrouver l'invariant (éventuellement dégradé) s'y trouve.

Une étude intéressante, qui permettrait d'avancer sur la notion d'invariant transformationnel, serait d'étudier l'influence du profil de vitesse sur la perception de l'interaction. En effet, parmi les interactions proposées, "frotter" et "gratter" sont des interactions faisant intervenir l'humain (son d'écriture par exemple) ou non (cas d'un solide qui glisse sur une surface inclinée par exemple) tout au long de la production du son, tandis que "rouler" est une interaction qui peut être initiée par une action humaine mais le mouvement qui s'en suit n'est pas forcément régi par un geste. En effet, l'invariant de la loi en "1/3", qui relie la courbure de la trajectoire et la vitesse (Lacquaniti *et al.*, 1983), contraint la perception auditive des mouvements biologiques humains (Thoret *et al.*, 2014). Il serait intéressant d'étudier si par exemple cet invariant influe sur la perception de l'interaction, e.g. si l'interaction "rouler" va être plus jugé comme "frotter" si cet invariant du mouvement biologique humain lui est appliqué. Inversement, si l'on applique un profil de vitesse type oscillateur harmonique (profil de vitesse d'une bille dans un bol) amorti à l'interaction "frotter", il est possible que cette interaction soit plus jugée comme "rouler". Cela impliquerait probablement de diviser les invariants transformationnels en deux niveaux, l'un plus lié à l'interaction, l'autre au mouvement.

Sur les métaphores sonores, des perspectives ont été présentées, notamment sur la nécessité de proposer un algorithme toujours cohérent avec le paradigme action-objet mais permettant une resynthèse d'un plus large type de textures (notamment les textures de type "granulaire", i.e. constituées d'atomes sonores discrets, comme la pluie par exemple). Quelques pistes de recherches ont été proposées. Une autre approche qui pourrait permettre des possibilités sonores supplémentaires, serait d'exprimer le filtre (dans le cas présenté dans le chapitre IV, i.e. on considère un filtre linéaire invariant

dans le temps) en un banc de résonateurs du second ordre en parallèle, comme l'on exprime la réponse impulsionnelle des objets résonants dans la partie objet du synthétiseur. Ceci permettrait d'appliquer des lois d'amortissement en fonction de la fréquence de matériaux connus (comme proposé par Aramaki *et al.* (2011) pour le contrôle du matériau perçu de sons d'impacts synthétiques), et ainsi modifier aisément les textures : un designer sonore pourrait ainsi facilement modifier des textures sonores en les métallisant ou en les boisant, par exemple. Cependant, des travaux préliminaires ont été effectués en ce sens au cours de cette thèse et laissent apparaître que ce problème est loin d'être trivial. En effet, comme on l'a vu dans le chapitre IV, un nombre important de pôles dans le filtre estimé est nécessaire pour obtenir une resynthèse correcte des textures sonores. Or pour exprimer le filtre estimé en résonateurs du second ordre en parallèle, une décomposition en éléments simples, et donc une factorisation de polynôme, est nécessaire. Cette opération est particulièrement instable numériquement s'il y a beaucoup de pôles, et en particulier lorsque beaucoup de pôles sont regroupés (et ont donc des racines proches) (Smith, 2007). Afin d'éviter ces problèmes, une méthode considérant un grand nombre de sous-bandes peut être utilisée, ou bien une méthode comme l'estimation de modèles AR ou ARMA avec zoom en fréquence (FZ-ARMA, cf Karjalainen *et al.* (2002)). Cependant, comme noté par Karjalainen *et al.* (2002), si on veut exprimer le banc de filtres résultant en pleine bande sans considérer de banc de filtres de synthèse et de décimations/interpolations (Vaidyanathan, 1993), la tâche est loin d'être évidente. Une recherche approfondie sur l'estimation de ce qu'on pourrait appeler les "modes de résonance" de la texture serait donc nécessaire à cet effet. Ainsi, les textures pourraient être directement intégrées et contrôlées à haut-niveau en temps-réel dans la partie objet du synthétiseur.

Bibliographie

J.-M. ADRIEN : *The missing link : Modal synthesis*, chapitre 8, pages 269–298. MIT Press, 1991.

A. AKAY : Acoustics of friction. *The Journal of the Acoustical Society of America*, 111:1525–1548, 2002.

J.B. ALLEN et D.A. BERKLEY : Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am*, 65(4):943–950, 1978.

J. ANDEN et S. MALLAT : Deep scattering spectrum. *Signal Processing, IEEE Transactions on*, 62(16):4114–4128, Aug 2014. ISSN 1053-587X.

M. ARAMAKI, M. BESSON, R. KRONLAND-MARTINET et S. YSTAD : *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, chapitre Timbre perception of sounds from impacted materials : behavioral, electrophysiological and acoustic approaches, pages 1–17. Springer, 2009a.

M. ARAMAKI, M. BESSON, R. KRONLAND-MARTINET et S. YSTAD : Controlling the perceived material in an impact sound synthesizer. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(2):301–314, 2011.

M. ARAMAKI, C. GONDRE, R. KRONLAND-MARTINET, T. VOINIER et S. YSTAD : Thinking the sounds : an intuitive control of an impact sound synthesizer. *In Proceedings of the 15th International Conference on Auditory Display*. Aalborg Universitet, Copenhagen, Denmark, May 18 - 22 2009b.

M. ARAMAKI et R. KRONLAND-MARTINET : Analysis-synthesis of impact sounds by real-time dynamic filtering. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(2):695–705, 2006.

U. M. ASCHER et L. R. PETZOLD : *Computer methods for ordinary differential equations and differential-algebraic equations*, volume 61. Siam, 1998.

M. ATHINEOS et D. P. W. ELLIS : Sound texture modelling with linear prediction in both time and frequency domains. *In IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings*, volume 5, pages 648–651. IEEE, 2003.

H. ATTIAS et C. E. SCHREINER : Temporal low-order statistics of natural sounds. *Advances in neural information processing systems*, pages 27–33, 1997.

F. AVANZINI, M. RATH et D. ROCCHESSO : Physically-based audio rendering of contact. *In Proceedings of the IEEE International Conference on Multimedia and Expo*, volume 2, pages 445–448. IEEE, 2002.

F. AVANZINI et D. ROCCHESSO : Controlling material properties in physical models of sounding objects. *In Proceedings of the International Computer Music Conference*, pages 91–94, 2001a.

F. AVANZINI et D. ROCCHESSO : Modeling collision sounds : Non-linear contact force. *In Proceedings of the Conference on Digital Audio Effects*, pages 61–66. Citeseer, 2001b.

F. AVANZINI et D. ROCCHESSO : Physical modeling of impacts : theory and experiments on contact time and spectral centroid. *In Proceedings of the Conference on Sound and Music Computing*, pages 287–293, 2004.

F. AVANZINI, D. ROCCHESSO et S. SERAFIN : Friction sounds for sensorial substitution. *In Proceedings of the International Conference on Auditory Display*, 2004.

F. AVANZINI, S. SERAFIN et D. ROCCHESSO : Interactive simulation of rigid body interaction with friction-induced sound generation. *IEEE Transactions on Speech and Audio Processing*, 13(5):1073–1081, 2005.

R. BADEAU, R. BOYER et B. DAVID : EDS parametric modeling and tracking of audio signals. *In Proceedings International Conference on Digital Audio Effects*, 2002.

Z. BAR-JOSEPH, D. LISCHINSKI, M. WERMAN, S. DUBNOV et R. EL-YANIV : Granular synthesis of sound textures using statistical learning. *In Proceedings of the International Computer Music Conference*, pages 178–181, 1999.

C. BASCOU et L. POTTIER : GMU, a flexible granular synthesis environment in max/msp. *In Proceedings of the Sound and Music Computing Conference*. Salerno, Italy, 2005.

J.P. BELLO, C. DUXBURY, M. DAVIES et M. SANDLER : On the use of phase and energy for musical onset detection in the complex domain. *IEEE Signal Processing Letters*, 11 (6):553–556, 2004.

H. BEN ABDELOUNIS, A. LE BOT, J. PERRET-LIAUDET et H. ZAHOUANI : An experimental study on roughness noise of dry rough flat surfaces. *Wear*, 268(1):335–345, 2010.

H. BEN ABDELOUNIS, H. ZAHOUANI, A. LE BOT, J. PERRET-LIAUDET et M. B. TKAYA : Numerical simulation of friction noise. *Wear*, 271(3):621–624, 2011.

J. BENSA, K. JENSEN et R. KRONLAND-MARTINET : A hybrid resynthesis model for hammer-string interaction of piano tones. *EURASIP Journal on Applied Signal Processing*, 2004:1021–1035, 2004.

G. BERNARDES, C. GUEDES et B. PENNYCOOK : Eargram : an application for interactive exploration of large databases of audio snippets for creative purposes. *In Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval*, pages 265–277, 2012.

S. BILBAO : *Numerical Sound Synthesis : Finite Difference Schemes and Simulation in Musical Acoustics*. John Wiley & Sons, 2009.

N. BÖTTCHER : Current problems and future possibilities of procedural audio in computer games. *Journal of Gaming & Virtual Worlds*, 5(3):215–234, 2013.

A. S. BREGMAN : *Auditory scene analysis : The perceptual organization of sound*. MIT press, 1994.

J. BRUNA et S. MALLAT : Audio texture synthesis with scattering moments. *arXiv preprint arXiv :1311.0407*, 2013.

P. A. CABE et J. B. PITTENGER : Human sensitivity to acoustic information from vessel filling. *Journal of experimental psychology : human perception and performance*, 26(1):313, 2000.

C. CADOZ, A. LUCIANI, J.-L. FLORENS, C. ROADS et F. CHADABE : Responsive input devices and sound synthesis by stimulation of instrumental mechanisms : The cordis system. *Computer music journal*, pages 60–73, 1984.

C. CARELLO, K.L. ANDERSON et A.J. KUNKLER-PECK : Perception of object length by sound. *Psychological Science*, 9(3):211, 1998.

U. CASTIELLO, B. L. GIORDANO, C. BEGLIOMINI, C. ANSUINI et M. GRASSI : When ears drive hands : the influence of contact sound on reaching to grasp. *PloS one*, 5 (8):e12240, 2010.

A. CHAIGNE et V. DOUTAUT : Numerical simulations of xylophones. i. time-domain modeling of the vibrating bars. *Journal of the Acoustical Society of America*, 101(1):539–557, 1997.

A. CHAIGNE et C. LAMBOURG : Time-domain simulation of damped impacted plates. i. theory and experiments. *The Journal of the Acoustical Society of America*, 109:1422–1432, 2001.

J. CHOWNING : The synthesis of complex audio spectra by means of frequency modulation. *J. Audio Eng. Soc*, 21, 1973.

S. CONAN, M. ARAMAKI, R. KRONLAND-MARTINET, E. THORET et S. YSTAD : Perceptual differences between sounds produced by different continuous interactions. *In Acoustics 2012, Nantes*, 23-27 april 2012.

S. CONAN, O. DERRIEN, M. ARAMAKI, S. YSTAD et R. KRONLAND-MARTINET : A synthesis model with intuitive control capabilities for rolling sounds. *IEEE/ACM Transactions on Speech, Audio and Language Processing*, 22(8):1260–1273, 2014a.

S. CONAN, E. THORET, M. ARAMAKI, O. DERRIEN, C. GONDRE, R. KRONLAND-MARTINET et S. YSTAD : Navigating in a space of synthesized interaction-sounds : Rubbing, scratching and rolling sounds. *In Proceedings of the 16th International Conference on Digital Audio Effects*, Maynooth, Ireland, September 2013.

S. CONAN, E. THORET, M. ARAMAKI, O. DERRIEN, C. GONDRE, S. YSTAD et R. KRONLAND-MARTINET : An intuitive synthesizer of continuous interaction sounds : Rubbing, scratching and rolling. *Computer Music Journal*, 38(4):24–37, 2014b.

P. R. COOK : Modeling bill's gait : Analysis and parametric synthesis of walking sounds. *In Audio Engineering Society Conference : 22nd International Conference : Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, 2002.

J. DANNA, J.L. VELAY, V. PAZ-VILLAGRÁN, A. CAPEL, C. PETROZ, C. GONDRE, E. THORET, M. ARAMAKI, S. YSTAD et R. KRONLAND-MARTINET : Handwriting movement sonification for the rehabilitation of dysgraphia. *In Sound, Music & Motion-Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research*, pages 200–208, 2013.

L. DAUDET : Sparse and structured decompositions of signals with the molecular matching pursuit. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1808–1816, 2006.

H. A. DAVID : *The method of paired comparisons*. Oxford University Press, 1988.

F. DE SAUSSURE : *Cours de linguistique générale*. Payot, Paris, 1916.

N. DELPRAT, P. GUILLEMAIN et R. KRONLAND-MARTINET : Parameter estimation for non-linear resynthesis methods with the help of a time-frequency analysis of natural sounds. *In Proceedings of th International Computer Music Conference*, pages 88–90, 1990.

S. DENJEAN, V. ROUSSARIE, R. KRONLAND-MARTINET, S. YSTAD et J.-L. VELAY : How does interior car noise alter driver's perception of motion ? multisensory integration in speed perception. *Acoustics 2012 Nantes*, 2012.

S. DENJEAN, J.-L. VELAY, R. KRONLAND-MARTINET, V. ROUSSARIE, J.-F. SCIABICA et S. YSTAD : Are electric and hybrid vehicles too quiet for drivers ? *In Internoise 2013*, 2013.

M. DESAINTE-CATHERINE et S. MARCHAND : Structured additive synthesis : Towards a model of sound timbre and electroacoustic music forms. *In Proceedings of the International Computer Music Conference*, pages 260–263, 1999.

A. DI SCIPIO : Synthesis of environmental sound textures by iterated nonlinear functions. *In Proceedings of the Workshop on Digital Audio Effects*, pages 109–117, 1999.

P. DOORNBUSCH : Computer sound synthesis in 1951 : The music of csirac. *Computer Music Journal*, 28(1):10–25, 2004.

S. DUBNOV, Z. BAR-JOSEPH, R. EL-YANIV, D. LISCHINSKI et M. WERMAN : Synthesizing sound textures through wavelet tree learning. *IEEE Computer Graphics and Applications*, 22(4):38–48, 2002.

G. DUBUS et R. BRESIN : A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE*, 8(12):e82491, 12 2013. URL `http://dx.doi.org/10.1371%2Fjournal.pone.0082491`.

H. DUDLEY : Remaking speech. *The Journal of the Acoustical Society of America*, 11(2):169–177, 1939.

D. P. W. ELLIS : Sinewave and sinusoid+noise analysis/synthesis in Matlab, 2003. URL `http://www.ee.columbia.edu/~dpwe/resources/matlab/sinemodel/`. online web resource.

A. ELMAIAN : *Méthodologies de simulation des bruits automobiles induits par le frottement*. Thèse de doctorat, Université du Maine, 2013.

A. ELMAIAN, F. GAUTIER, C. PEZERAT et J.-M. DUFFAL : How can automotive friction-induced noises be related to physical mechanisms ? *Applied Acoustics*, 76:391–401, 2014.

E. FALCON, C. LAROCHE, S. FAUVE et C. COSTE : Behavior of one inelastic ball bouncing repeatedly off the ground. *The European Physical Journal B-Condensed Matter and Complex Systems*, 3(1):45–57, 1998.

A. FARNELL : *Designing sound*. MIT Press Cambridge, 2010.

A. FIELD : *Music, electronic media, and culture*, chapitre Simulation and reality : the new sonic objects. Ashgate Publishing, Ltd., 2000.

J. L. FLANAGAN et R. M. GOLDEN : Phase vocoder. *Bell System Technical Journal*, 45 (9):1493–1509, 1966.

K. FRANINOVIĆ et S. SERAFIN : *Sonic Interaction Design*. Mit Press, 2013.

D. J. FREED : Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, 87:311–322, 1990.

D. GABOR : Acoustical quanta and the theory of hearing. *Nature*, 159:591–594, 1947.

L. GANDEMER, G. PARSEIHIAN, R. KRONLAND-MARTINET et C. BOURDIN : The influence of horizontally rotating sound on standing balance. *Experimental brain research*, pages 1–8, 2014.

W. W. GAVER : *Everyday listening and auditory icons*. Thèse de doctorat, UNIVERSITY OF CALIFORNIA, SAN DIEGO, 1988.

W. W. GAVER : How do we hear in the world ? explorations in ecological acoustics. *Ecological psychology*, 5(4):285–313, 1993a.

W. W. GAVER : What in the world do we hear ? : An ecological approach to auditory event perception. *Ecological psychology*, 5(1):1–29, 1993b.

M. N. GEFFEN, J. GERVAIN, J. F. WERKER et M. O. MAGNASCO : Auditory perception of self-similarity in water sounds. *Frontiers in Integrative Neuroscience*, 5, 2011.

J. J. GIBSON : *The senses considered as perceptual systems.* Houghton Mifflin, 1966.

J. J. GIBSON : *The ecological approach to visual perception.* Boston : Houghton Miflin Co, 1979.

B. L. GIORDANO et S. MCADAMS : Material identification of real impact sounds : Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America*, 119:1171–1181, 2006.

B. L. GIORDANO et K. PETRINI : Hardness recognition in synthetic sounds. *In Proceedings of the Stockholm Music Acoustics Conference, Stockholm, Sweden*, 2003.

B. GOLD et C. M. RADER : The channel vocoder. *Audio and Electroacoustics, IEEE Transactions on*, 15(4):148–161, 1967.

A. GOUNAROPOULOS et C. JOHNSON : Synthesising timbres and timbre-changes from adjectives/adverbs. *Applications of Evolutionary Computing*, pages 664–675, 2006.

M. GRASSI : Do we hear size or sound ? balls dropped on plates. *Attention, Perception, & Psychophysics*, 67(2):274–284, 2005.

J. M. GREY : Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5):1270–1277, 1977.

B. GYGI, G. R. KIDD et C. S. WATSON : Spectral-temporal factors in the identification of environmental sounds. *The Journal of the Acoustical Society of America*, 115:1252, 2004.

L. HAMSTRA-BLETZ et A. W. BLÖTE : A longitudinal study on dysgraphic handwriting in primary school. *Journal of Learning Disabilities*, 26(10):689–699, 1993.

C. HEINRICHS et A. MCPHERSON : Mapping and interaction strategies for performing environmental sound. *In 1st Workshop on Sonic Interactions for Virtual Environments at IEEE VR 2014*, 2014.

T. HERMANN, J. NEUHOFF et A. HUNT : *The Sonification Handbook*. Logos Verlag, Berlin, Germany, 2011.

D. J. HERMES : Synthesis of the sounds produced by rolling balls. Internal IPO report no. 1226, IPO, Center for User-System Interaction, Eindhoven, The Netherlands, September 1998.

J. HERRE et J. D. JOHNSTON : Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS). *In Audio Engineering Society Convention 101*. Audio Engineering Society, 1996.

G. W. HILL : Programming for high-speed computers. Mémoire de D.E.A., University of Sidney, Australia, 1954.

M. HOFFMAN et P. R. COOK : Real-time feature-based synthesis for live musical performance. *In Proceedings of the 7th international conference on New interfaces for musical expression*, pages 309–312. ACM, 2007.

M. HOFFMAN et P.R. COOK : Feature-based synthesis : Mapping acoustic and perceptual features onto synthesis parameters. *In Proceedings of the 2006 International Computer Music Conference*, volume 33, pages 536–539. Citeseer, 2006.

A. HORNER, J. BEAUCHAMP et L. HAKEN : Machine tongues xvi : Genetic algorithms and their application to fm matching synthesis. *Computer Music Journal*, pages 17–29, 1993.

M. M. M. HOUBEN : *The Sound of Rolling Objects, Perception of size and speed*. Thèse de doctorat, Technische Universiteteit, Eindhoven, 2002.

M.M.J. HOUBEN, A. KOHLRAUSCH et D.J. HERMES : Auditory cues determining the perception of size and speed of rolling balls. *In Proc. Int. Conf. Auditory Display (ICAD-01)*, pages 105–110, 2001.

M.M.J. HOUBEN, A. KOHLRAUSCH et D.J. HERMES : Perception of the size and speed of rolling balls by sound. *Speech communication*, 43(4):331–345, 2004.

M.M.J. HOUBEN, A. KOHLRAUSCH et D.J. HERMES : The contribution of spectral and temporal information to the auditory perception of the size and speed of rolling balls. *Acta acustica united with acustica*, 91(6):1007–1015, 2005.

O. HOUIX, G. LEMAITRE, N. MISDARIIS, P. SUSINI et I. URDAPILLETA : A lexical analysis of environmental sound categories. *Journal of Experimental Psychology : Applied*, 18 (1):52, 2012.

K.H. HUNT et F.R.E. CROSSLEY : Coefficient of restitution interpreted as damping in vibroimpact. *J. Appl. Mech.*, 42(2):440–445, 1975.

B. JULESZ : Visual pattern discrimination. *IRE Transactions on Information Theory*, 8 (2):84–92, 1962.

J. H. JUSTICE : Analytic signal processing in music computation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(6):670–684, 1979.

M. KARJALAINEN, P. A. A. ESQUEF, P. ANTSALO, A. MÄKIVIRTA et V. VÄLIMÄKI : Frequency-zooming arma modeling of resonant and reverberant systems. *Journal of the Audio Engineering Society*, 50(12):1012–1029, 2002.

K. KARPLUS et A. STRONG : Digital synthesis of plucked-string and drum timbres. *Computer Music Journal*, 7(2):43–55, 1983.

A.G. KATSIAMIS, E.M. DRAKAKIS et R.F. LYON : Practical gammatone-like filters for auditory processing. *EURASIP Journal on Audio, Speech, and Music Processing*, 2007, 2007.

F. KEILER, D. ARFIB et U. ZÖLZER : Efficient linear prediction for digital audio effects. *In Proceedings of the International Conference on Digital Audio Effects*, 2000.

John L KELLY et Carol C LOCHBAUM : Speech synthesis. *In Proceedings of the 4th International Congress on Acoustics Proceedings of the 4th International Congress on Acoustics*, pages 1–4, 1962.

S. KERSTEN et H. PURWINS : Sound texture synthesis with hidden markov tree models in the wavelet domain. *In Proceedings of Sound and Music Computing Conference*, 2010.

S. KERSTEN et H. PURWINS : Fire texture sound re-synthesis using sparse decomposition and noise modelling. *In Proc. Int. Conf. on Digital Audio Effects DAFx*, 2012.

R.L. KLATZKY, D.K. PAI et E.P. KROTKOV : Perception of material from contact sounds. *Presence : Teleoperators & Virtual Environments*, 9(4):399–410, 2000.

R. KRONLAND-MARTINET et T. VOINIER : Real-time perceptual simulation of moving sources : Application to the leslie cabinet and 3d sound immersion. *EURASIP Journal on Audio, Speech, and Music Processing*, 2008:7, 2008.

R. KRONLAND-MARTINET, S. YSTAD et M. ARAMAKI : High-level control of sound synthesis for sonification processes. *AI & society*, 27(2):245–255, 2012.

A. J. KUNKLER-PECK et M. T. TURVEY : Hearing shape. *Journal of Experimental Psychology : Human Perception and Performance*, 26(1):279, 2000.

F. LACQUANITI, C. TERZUOLO et P. VIVIANI : The law relating the kinematic and figural aspects of drawing movements. *Acta psychologica*, 54(1):115–130, 1983.

M. LAGRANGE, G. SCAVONE et P. DEPALLE : Analysis/synthesis of sounds generated by sustained contact between rigid objects. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3):509–518, 2010.

J. LAIRD : *The Physical Modelling Of Drums Using Digital The Physical Modelling Of Drums Using Digital Waveguides*. Thèse de doctorat, University of Bristol, 2001.

S. LAKATOS, S. MCADAMS et R. CAUSSÉ : The representation of auditory source characteristics : Simple geometric form. *Attention, Perception, & Psychophysics*, 59(8):1180–1190, 1997.

C. LAMBOURG, A. CHAIGNE et D. MATIGNON : Time-domain simulation of damped impacted plates. ii. numerical model and results. *The Journal of the Acoustical Society of America*, 109:1433–1447, 2001.

S. LE GROUX et P. F. VERSCHURE : Perceptsynth : mapping perceptual musical features to sound synthesis parameters. *In Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 125–128. IEEE, 2008.

J.S. LEE, P. DEPALLE et G. SCAVONE : Analysis/synthesis of rolling sounds using a source-filter approach. *In Proceedings of the International Conference on Digital Audio Effects*, 2010.

G. LEMAITRE, A. DESSEIN, P. SUSINI et K. AURA : Vocal imitations and the identification of sound events. *Ecological Psychology*, 23(4):267–307, 2011.

G. LEMAITRE et L. M. HELLER : Evidence for a basic level in a taxonomy of everyday action sounds. *Experimental brain research*, 226(2):253–264, 2013.

G. LEMAITRE et L.M. HELLER : Auditory perception of material is fragile while action is strikingly robust. *Journal of the Acoustical Society of America*, 131:1337–1348, February 2012.

G. LEMAITRE, O. HOUIX, N. MISDARIIS et P. SUSINI : Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology : Applied*, 16(1):16, 2010.

G. LEMAITRE et D. ROCCHESSO : On the effectiveness of vocal imitations and verbal descriptions of sounds. *The Journal of the Acoustical Society of America*, 135(2):862–873, 2014. URL `http://scitation.aip.org/content/asa/journal/jasa/135/2/10.1121/1.4861245`.

J. LEONARD, C. CADOZ, N. CASTAGNÉ et A. LUCIANI : A virtual reality platform for musical creation. *In International Symposium on Computer Music Modeling and Retrieval*, 2013.

X. LI, R. LOGAN et R. PASTORE : Perception of acoustic source characteristics : Walking sounds. *Journal of the Acoustical Society of America*, 90(6):3036–3049, 1991.

D. B. LLOYD, N. RAGHUVANSHI et N. K. GOVINDARAJU : Sound synthesis for impact sounds in video games. *In Symposium on Interactive 3D Graphics and Games*, pages 55–62. ACM, 2011.

L. LU, L. WENYIN et H.-J. ZHANG : Audio textures : Theory and applications. *Speech and Audio Processing, IEEE Transactions on*, 12(2):156–167, 2004.

J. M. LUCK et A. MEHTA : Bouncing ball with a finite restitution : chattering, locking, and chaos. *Physical Review E*, 48(5):3988, 1993.

R. A. LUTFI et E. L. OH : Auditory discrimination of material changes in a struck-clamped bar. *The Journal of the Acoustical Society of America*, 102(6):3647–3656, 1997.

J. MACKENZIE : *Using strange attractors to model sound*. Thèse de doctorat, King's College, 1994.

J. MAKHOUL : Linear prediction : A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, 1975.

S. G. MALLAT et Z. ZHANG : Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

M. MATHEWS : The digital computer as a musical instrument. *Science*, 142(11):553–557, 1963.

M. MATHEWS et J.O. SMITH : Methods for synthesizing very high q parametrically well behaved two pole filters. *In Proceedings of the Stockholm Musical Acoustics Conference (SMAC 2003)(Stockholm), Royal Swedish Academy of Music (August 2003)*, 2003.

M. V. MATHEWS et N. GUTTMAN : Generation of music by a digital computer. *In Proceedings of the International Congress on Acoustics*, 1959.

S. E. MCADAMS : *Thinking in sound : The cognitive psychology of human audition.*, chapitre Recognition of sound sources and events. Oxford Science Publications, 1993.

R. MCAULAY et T. QUATIERI : Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(4):744–754, 1986.

J. H. MCDERMOTT, M. SCHEMITSCH et E. P. SIMONCELLI : Summary statistics in auditory perception. *Nature Neuroscience*, 2013.

J.H. MCDERMOTT, A.J. OXENHAM et E.P. SIMONCELLI : Sound texture synthesis via filter statistics. *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 297–300. IEEE, 2009.

J.H. MCDERMOTT et E.P. SIMONCELLI : Sound texture perception via statistics of the auditory periphery : evidence from sound synthesis. *Neuron*, 71(5):926–940, 2011.

A. MERER, M. ARAMAKI, S. YSTAD et R. KRONLAND-MARTINET : Perceptual characterization of motion evoked by sounds for synthesis control purposes. *ACM Transactions on Applied Perception (TAP)*, 10(1):1, 2013.

C.F. MICHAELS et C. CARELLO : *Direct perception*. Prentice-Hall Englewood Cliffs, NJ, 1981.

J.A. MICOULAUD-FRANCHI, M. ARAMAKI, A. MERER, M. CERMOLACCE, S. YSTAD, R. KRONLAND-MARTINET et J. VION-DURY : Categorization and timbre perception of environmental sounds in schizophrenia. *Psychiatry research*, 189(1):149–152, 2011.

N.E. MINER et T.P. CAUDELL : Using wavelets to synthesize stochastic-based sounds for immersive virtual environments. *In International Conference on Auditory Display, Palo Alto, CA*, 1997.

J. A. MOORER : The use of the phase vocoder in computer music applications. *Journal of the Audio Engineering Society*, 26(1/2):42–45, 1978.

J. A. MOORER : The use of linear prediction of speech in computer music applications. *Journal of the Audio Engineering Society*, 27(3):134–140, 1979.

V. MORICEAU : Un modèle de représentation sémantique de la métaphore : le cas des métaphores d'orientation. Mémoire de D.E.A., Université Paul Sabatier, Toulouse III, 2003.

P.M. MORSE et K.U. INGARD : *Theoretical acoustics*. Mc Graw-Hill Book Company, 1968.

W. MOSS, H. YEH, J.-M. HONG, M. C. LIN et D. MANOCHA : Sounding liquids : Automatic sound synthesis from fluid simulation. *ACM Transactions on Graphics (TOG)*, 29(3):21, 2010.

D. MURPHY, A. KELLONIEMI, J. MULLEN et S. SHELLEY : Acoustic modeling using the digital waveguide mesh. *IEEE Signal Processing Magazine*, 24(2):55–66, 2007.

E. MURPHY, M. LAGRANGE, G. SCAVONE, P. DEPALLE et C. GUASTAVINO : Perceptual evaluation of rolling sound synthesis. *Acta Acustica united with Acustica*, 97(5):840–851, 2011.

T. Q. NGUYEN et P. P. VAIDYANATHAN : Two-channel perfect-reconstruction FIR QMF structures which yield linear-phase analysis and synthesis filters. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(5):676–690, 1989.

B. NINNESS, A. WILLS et S. GIBSON : The university of newcastle identification toolbox (unit). *In Proceedings of theWorld Congress of the International Federation of Automatic Control*, 2005.

R. NORDAHL, S. SERAFIN, L. TURCHET et N.C. NILSSON : A multimodal architecture for simulating natural interactive walking in virtual environments. *PsychNology Journal*, 9(3):245–268, 2011a.

R. NORDAHL, L. TURCHET et S. SERAFIN : Sound synthesis and evaluation of interactive footsteps and environmental sounds rendering for virtual reality applications. *IEEE Transactions on Visualization and Computer Graphics*, 17(9):1234–1244, 2011b.

A. OLIVERO, P. DEPALLE, B. TORRÉSANI et R. KRONLAND-MARTINET : Sound morphing strategies based on alterations of time-frequency representations by Gabor multipliers. *In Audio Engineering Society Conference : 45th International Conference : Applications of Time-Frequency Processing in Audio*. Audio Engineering Society, 2012.

H. OMER et B. TORRÉSANI : Estimation of frequency modulations on wideband signals ; applications to audio signal analysis. *In* Goetz PFANDER, éditeur : *Proceedings of the 10th International Conference on Sampling Theory and Applications*, pages 29–32. Eurasip Open Library, 2013. URL `http://hal.archives-ouvertes.fr/hal-00822186`.

A. V. OPPENHEIM et R. W. SCHAFER : *Digital signal processing*. Prentice-Hall, 1975.

S. PAPETTI, F. AVANZINI et D. ROCCHESSO : Numerical methods for a nonlinear impact model : a comparative study with closed-form corrections. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2146–2158, 2011.

C. V. PARISE, K. KNORRE et M. O. ERNST : Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, page 201322705, 2014.

J. PARKER et T. BOVERMANN : Dynamic FM synthesis using a network of complex resonator filters. *In Proceedings of the Sound and Music Computing Conference*, 2013.

R. G. PAYNE : A microcomputer-based analysis/resynthesis scheme for processing sampled sounds using FM. *In Proceedings of the International Computer Music Conference*, pages 282–289, 1987.

L. PELTOLA, C. ERKUT, P. R. COOK et V. VALIMAKI : Synthesis of hand clapping sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1021–1029, 2007.

G. E. PETERSON et H. L. BARNEY : Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2):175–184, 1952.

C. PICARD, N. TSINGOS et F. FAURE : Retargetting example sounds to interactive physics-driven animations. *In Audio Engineering Society Conference : 35th International Conference : Audio for Games*. Audio Engineering Society, 2009.

L. PIZZAMIGLIO, T. APRILE, G. SPITONI, S. PITZALIS, E. BATES, S. D'AMICO et F. DI RUSSO : Separate neural systems for processing action- or non-action-related sounds. *Neuroimage*, 24(3):852–861, 2005.

L. POLANSKY et T. ERBE : Spectral mutation in soundhack. *Computer Music Journal*, pages 92–101, 1996.

J. PORTILLA et E. P. SIMONCELLI : A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–70, 2000.

C.-E. RAKOVEC, M. ARAMAKI et R. KRONLAND-MARTINET : Perception of material and shape of impacted everyday objects. *In International Symposium on Computer Music Modeling and Retrieval*, 2013.

M. RATH : *Interactive Realtime Sound Models for Human–Computer Interaction—A Sound Design Concept and Applications*. Thèse de doctorat, Department of Computer Science, University of Verona, 2004.

M. RATH, F. AVANZINI, N. BERNARDINI, G. BORIN, F. FONTANA, L. OTTAVIANI et D. ROCCHESSO : An introductory catalog of computer-synthesized contact sounds, in real-time. *In Proc. Colloquium of Musical Informatics*, pages 103–108, 2003.

M. RATH et D. ROCCHESSO : Continuous sonic feedback from a rolling ball. *IEEE Multimedia*, 12(2):60–69, 2005.

R. E. REMEZ, P. E. RUBIN, D. B. PISONI et T. D. CARRELL : Speech perception without traditional speech cues. *Science*, 212(4497):947–949, 1981.

Z. REN, H. YEH et M. C. LIN : Synthesizing contact sounds between textured models. *In IEEE Virtual Reality Conference*, pages 139–146. IEEE, 2010.

B. H. REPP : The sound of two hands clapping : An exploratory study. *The Journal of the Acoustical Society of America*, 81(4):1100–1109, 1987.

J.-C. RISSET et M. V. V MATHEWS : Analysis of musical-instrument tones. *Physics today*, 22:23–30, 1969.

C. ROADS : *Foundations of computer music*, chapitre Granular Synthesis of Sound. MIT Press, 1985.

C. ROADS : Introduction to granular synthesis. *Computer Music Journal*, pages 11–13, 1988.

C. ROADS : *L'audionumérique*, chapitre 8. Paris, Dunod, 1998.

D. ROCCHESSO et F. FONTANA : *The sounding object*. Mondo estremo, 2003.

Xavier RODET, Yves POTARD et Jean-Baptiste BARRIERE : The chant project : from the synthesis of the singing voice to synthesis in general. *Computer Music Journal*, pages 15–31, 1984.

M. W. M. RODGER, W. R. YOUNG et C. M. CRAIG : Synthesis of walking sounds for alleviating gait disturbances in parkinson's disease. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 22(3):543 – 548, May 2014.

T.D. ROSSING : Acoustics of the glass harmonica. *The Journal of the Acoustical Society of America*, 95:1106, 1994.

R. ROY et T. KAILATH : Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984–995, 1989.

P. M. RUIZ : A technique for simulating the vibration of strings with a digital computer. Mémoire de D.E.A., University of Illinois at Urbana-Champaign, 1969.

N. SAINT-ARNAUD et K. POPAT : Analysis and synthesis of sound textures. *In Readings in Computational Auditory Scene Analysis*. Citeseer, 1995.

R.S. SAYLES et T.R. THOMAS : Surface topography as a nonstationary random process. *Nature*, 271:431–434, 1978.

R. M. SCHAFER : *The soundscape : Our sonic environment and the tuning of the world*. Inner Traditions/Bear & Co, 1993.

K. W. SCHINDLER : Dynamic timbre control for real-time digital synthesis. *Computer Music Journal*, pages 28–42, 1984.

E. SCHUIJERS, W. OOMEN, B. den BRINKER et J. BREEBAART : Advances in parametric coding for high-quality audio. *In Audio Engineering Society Convention 114*, 2003.

D. SCHWARZ : *Data-driven concatenative sound synthesis*. Thèse de doctorat, Université Paris 6 - Pierre et Marie Curie, 2004.

D. SCHWARZ : State of the art in sound texture synthesis. *In Proceedings of th International Conference on Digital Audio Effects*, 2011.

D. SCHWARZ : Retexture – towards interactive environmental sound texture synthesis through inversion of annotations. *In Proc. of the 10th International Symposium on Computer Music Multidisciplinary Research*, 2013.

J.-F. SCIABICA : *Caractérisation acoustique et perceptive du bruit moteur dans un habitacle automobile*. Thèse de doctorat, Aix Marseille 1, 2011.

A. SEDDA, S. MONACO, G. BOTTINI et M. A. GOODALE : Integration of visual and auditory information for hand actions : preliminary evidence for the contribution of natural sounds to grasping. *Experimental brain research*, 209(3):365–374, 2011.

S. SERAFIN : *The sound of friction : real-time models, playability and musical applications*. Thèse de doctorat, Stanford University, 2004.

S. SERAFIN, P. HUANG, S. YSTAD, C. CHAFE et J. O. SMITH : Analysis and synthesis unusual friction-driven musical instruments. *In Proceedings of International Computer Music Conference*, 2002.

S. SERAFIN, N. C. NILSSON et D. SKAARUP : Influence of auditory and haptic feedback on a balancing task. *International Journal of Autonomous and Adaptive Communications Systems*, 6(4):366–376, 2013.

X. SERRA : *A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition*. Thèse de doctorat, Stanford University, 1989.

J. O. SMITH : Physical modeling using digital waveguides. *Computer music journal*, pages 74–91, 1992.

J. O. SMITH : *Introduction to digital filters : with audio applications*. W3K Publishing, 2007.

J. O. SMITH : *Physical Audio Signal Processing*. `http://ccrma.stanford.edu/~jos/-pasp/`, 2014. online book, 2010 edition.

B. C. M. SMITS-ENGELSMAN, A. S. NIEMEIJER et G. P. van GALEN : Fine motor deficiencies in children diagnosed as DCD based on poor grapho-motor ability. *Human movement science*, 20(1):161–182, 2001.

C. SPENCE et M. ZAMPINI : Auditory contributions to multisensory product perception. *Acta Acustica united with Acustica*, 92(6):1009–1025, 2006.

C. STOELINGA et A. CHAIGNE : Time-domain modeling and simulation of rolling objects. *Acta Acustica united with Acustica*, 93(2):290–304, 2007.

C. N. J. STOELINGA, D. J. HERMES, A. HIRSCHBERG et A. J. M. HOUTSMA : Temporal aspects of rolling sounds : A smooth ball approaching the edge of a plate. *Acta Acustica united with Acustica*, 89(5):809–817, 2003.

C.N.J. STOELINGA : *A psychomechanical study of rolling sounds*. Thèse de doctorat, ENSTA ParisTech, 2007.

C. SUIED, T. R. AGUS, S. J. THORPE, N. MESGARANI et D. PRESSNITZER : Auditory gist : Recognition of very short sounds from timbre cues. *The Journal of the Acoustical Society of America*, 2014. ISSN 00014966.

C. SUIED, A. DRÉMEAU, D. PRESSNITZER et L. DAUDET : Auditory sketches : Sparse representations of sounds based on perceptual models. *In From Sounds to Music and Emotions*, pages 154–170. Springer, 2013.

P. SUSINI, O. HOUIX et N. MISDARIIS : Sound design : an applied, experimental framework to study the perception of everyday sounds. *The New Soundtrack*, 4(2):103–121, 2014.

T. TAKEUCHI : Auditory information in playing tennis. *Perceptual and motor skills*, 76 (3c):1323–1328, 1993.

W. F. D. THERON : *Analysis of the rolling motion of loaded hoops*. Thèse de doctorat, University of Stellenbosch, 2008.

E. THORET : *Caractérisation acoustique et perceptive des relations sensori-motrices entre les mouvements biologiques et la perception sonore*. Thèse de doctorat, Aix-Marseille Université, 2014.

E. THORET, M. ARAMAKI, C. GONDRE, R. KRONLAND-MARTINET et S. YSTAD : Controlling a non linear friction model for evocative sound synthesis applications. *In Proceedings of the International Conference on Digital Audio Effects*, Maynooth, Ireland, September 2013.

E. THORET, M. ARAMAKI, R. KRONLAND-MARTINET, J.-L. VELAY et S. YSTAD : From sound to shape : Auditory perception of drawing movements. *Journal of Experimental Psychology : Human Perception and Performance, Advance online publication*, 2014.

S. TUCKER et G.J. BROWN : Investigating the perception of the size, shape and material of damped and free vibrating plates. *University of Sheffield, Department of Computer Science Technical Report CS-02-10*, 2002.

P. P. VAIDYANATHAN : On power-complementary FIR filters. *IEEE transactions on circuits and systems*, 32(12):1308–1310, 1985.

P. P. VAIDYANATHAN : *Multirate systems and filter banks*. Pearson Education India, 1993.

V. VALIMAKI, H.-M. LEHTONEN et M. TAKANEN : A perceptual study on velvet noise and its variants at different pulse densities. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(7):1481–1488, July 2013.

K. van den DOEL : Physically based models for liquid sounds. *In Proceedings of the International Conference on Auditory Display*, Sidney, Australia, July 6-9 2004.

K. VAN DEN DOEL, P.G. KRY et D.K. PAI : Foleyautomatic : physically-based sound effects for interactive simulation and animation. *In Proceedings of the conference on Computer graphics and interactive techniques*, pages 537–544. ACM, 2001.

K. van den DOEL et D.K. PAI : Modal synthesis for vibrating objects. *Audio Anectodes. AK Peter, Natick, MA*, 2003.

N. J. VANDERVEER : *Ecological acoustics : Human perception of environmental sounds (Unpublished)*. Thèse de doctorat, Cornell University, 1979.

C. VELASCO, R. JONES, S. KING et C. SPENCE : The sound of temperature : What information do pouring sounds convey concerning the temperature of a beverage. *Journal of Sensory Studies*, 28(5):335–345, 2013.

J. L. VERHEY : *The Oxford Handbook of Auditory Science : Hearing*, volume 3, chapitre 5. Oxford University Press, 2010.

C. VERRON : *Synthèse Immersive de Sons d'Environnement (Immersive Synthesis of Environmental Sounds)*. Thèse de doctorat, Université de Provence, 2010.

C. VERRON, M. ARAMAKI, R. KRONLAND-MARTINET et G. PALLONE : A 3-d immersive synthesizer for environmental sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1550–1561, 2010.

C. VERRON et G. DRETTAKIS : Procedural audio modeling for particle-based environmental effects. *In Audio Engineering Society Convention 133*. Audio Engineering Society, 2012.

J. VILLENEUVE et C. CADOZ : Inverse problem in sound synthesis and musical creation using mass-interaction networks. *In 9th Sound and Music Computing Conference*, pages 49–54, 2012.

P. VIVIANI, G. BAUD-BOVY et M. REDOLFI : Perceiving and tracking kinesthetic stimuli : Further evidence of motor–perceptual interactions. *Journal of Experimental Psychology : Human Perception and Performance*, 23(4):1232, 1997.

P. VIVIANI et N. STUCCHI : The effect of movement velocity on form perception : Geometric illusions in dynamic displays. *Perception & Psychophysics*, 46(3):266–274, 1989.

P. VIVIANI et N. STUCCHI : Biological movements look uniform : evidence of motor-perceptual interactions. *Journal of Experimental Psychology : Human Perception and Performance*, 18(3):603, 1992.

R. F. VOSS et J. CLARKE : 1/f noise in speech and music". *Nature*, 258:317–318, 1975.

M. M. WANDERLEY et P. DEPALLE : Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644, 2004.

M. M. WANDERLEY et J. MALLOCH, éditeurs. *Advances in the Design of Mapping for Computer Music*, volume 4, Fall 2014. Computer Music Journal, MIT Press.

D. WANG et G. J. BROWN, éditeurs. *Computational auditory scene analysis : Principles, algorithms, and applications*. Wiley-IEEE Press, 2006.

W.H. WARREN et R.R. VERBRUGGE : Auditory perception of breaking and bouncing events : A case study in ecological acoustics. *Journal of Experimental Psychology : Human Perception and Performance*, 10(5):704–712, 1984.

D. L. WESSEL : Timbre space as a musical control structure. *Computer music journal*, pages 45–52, 1979.

R. P. WILDES et W. A. RICHARDS : Recovering material properties from sound. *In* W. A. RICHARDS, éditeur : *Natural computation*, pages 356–363. Cambridge, Massachusetts, 1988.

Z. YE : Sound generated by rubbing objects. *Physics Letters A*, 327(2):91–94, 2004.

D. YOUNG et S. SERAFIN : Playability evaluation of a virtual bowed string instrument. *In Proceedings of the conference on New interfaces for musical expression*, pages 104–108. National University of Singapore, 2003.

S. YSTAD et T. VOINIER : A virtually real flute. *Computer Music Journal*, 25(2):13–24, 2001.

H. ZAHOUANI, R. VARGIOLU et J.-L. LOUBET : Fractal models of surface topography and contact mechanics. *Mathematical and Computer modelling*, 28(4):517–534, 1998.

M. ZAMPINI, S. GUEST et C. SPENCE : The role of auditory cues in modulating the perception of electric toothbrushes. *Journal of dental research*, 82(11):929–932, 2003.

M. ZAMPINI et C. SPENCE : The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *Journal of sensory studies*, 19(5):347–363, 2004.

C. ZHENG et D.L. JAMES : Rigid-body fracture sound with precomputed soundbanks. *ACM Transactions on Graphics*, 29(4):69, 2010.

C. ZHENG et D.L. JAMES : Toward high-quality modal contact sound. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2011)*, 30(4), août 2011. URL `http://www.cs.cornell.edu/projects/Sound/mc`.

X. ZHU et L. WYSE : Sound texture modeling and time-frequency LPC. *In Proceedings of the international conference on digital audio effects*, volume 4, 2004.

U. ZÖLZER, éditeur. *DAFX : digital audio effects*. Wiley Online Library, 2002.

E. ZWICKER et H. FASTL : *Psychoacoustics : Facts and Models.* Springer-Verlag, 1990.

Annexe A

Modélisation des densités de probabilité des séries d'impacts du roulement

Figures supplémentaires représentant les densités de probabilités estimées et modélisées de ${}^{c}\tilde{A}$ et ${}^{c}\tilde{\Delta}_T$. Quand ils ne varient pas, les paramètres sont fixés à $\kappa = \kappa_2$, $\mu = \mu_2$ et $\beta = -0.5$.

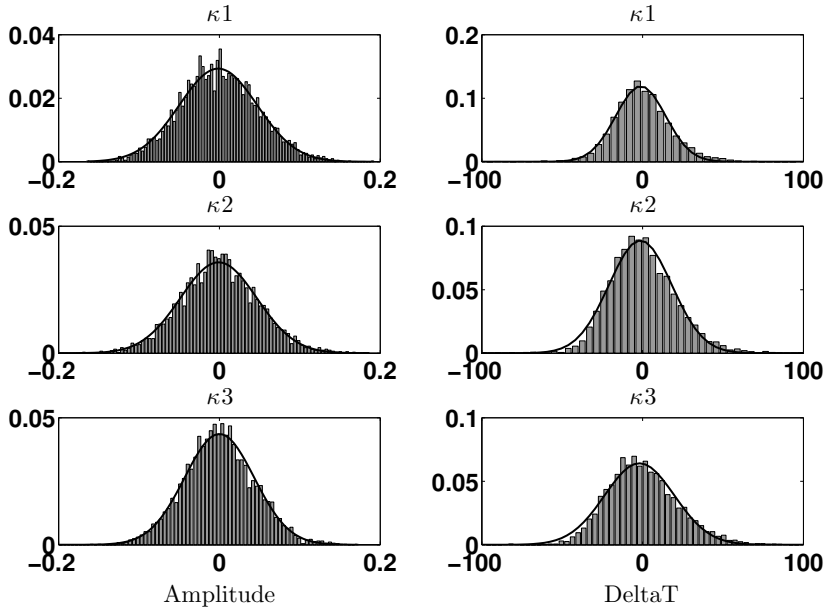

FIGURE A.1 – Variation du paramètre κ. Barres : densité de probabilité estimée de ${}^{c}\tilde{A}$ *(gauche)* et ${}^{c}\tilde{\Delta}_T$ *(droite)*. Courbe noire : estimation par les moindres carrés des lois gaussiennes.

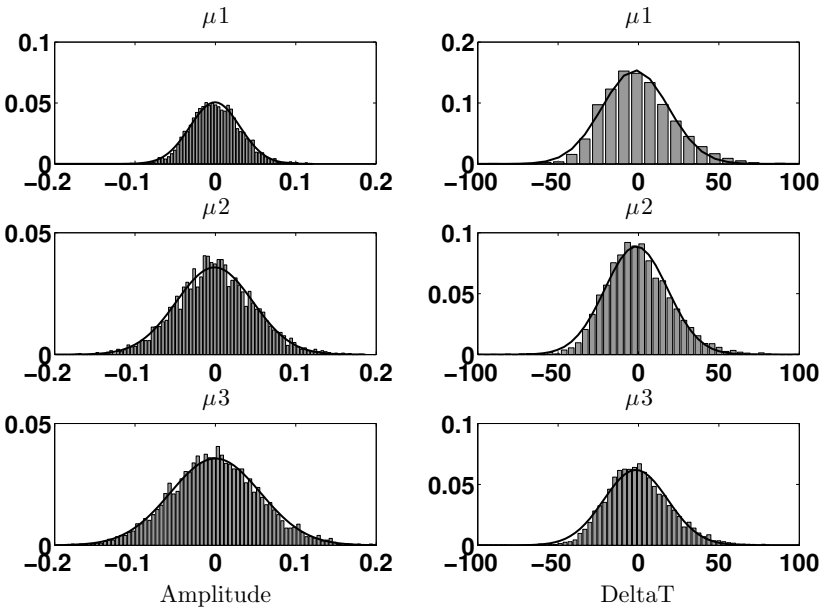

FIGURE A.2 – Variation du paramètre μ. Barres : densité de probabilité estimée de ${}^{c}\tilde{A}$ *(gauche)* et ${}^{c}\tilde{\Delta}_T$ *(droite)*. Courbe noire : estimation par les moindres carrés des lois gaussiennes.

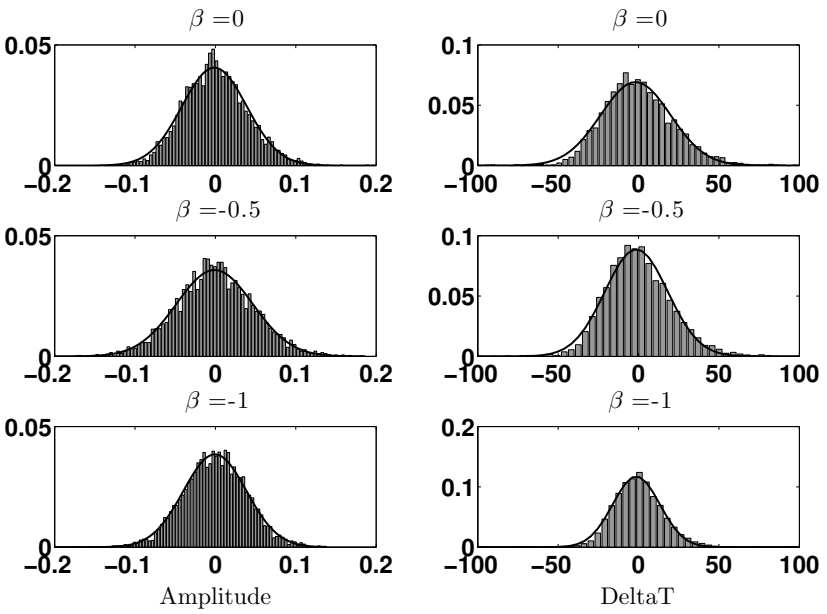

FIGURE A.3 – Variation du paramètre β. Barres : densité de probabilité estimée de ${}^{c}\tilde{A}$ *(gauche)* et ${}^{c}\tilde{\Delta}_T$ *(droite)*. Courbe noire : estimation par les moindres carrés des lois gaussiennes.

Annexe B

Fonction de détection d'attaques

Dans cette annexe, on décrit la fonction de détection d'attaques proposées par Bello *et al.* (2004). L'idée générale pour construire une fonction de détection d'attaque pour un signal $s(n)$ est d'observer les variations d'énergie et de phase entre des trames finies successives $s(m)$ de ce signal. Pour ce faire, on considère la transformée de Fourier à court terme de $s(m)$:

$$S_k(m) = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} s(mh+n)w(n)e^{-j\frac{2\pi nk}{N}} \ , \ k = 0, 1, ..., N-1 \tag{B.1}$$

où $w(n)$ est une fenêtre de taille N et $h \in [1, N]$ est la taille du pas. On peut réécrire cette transformée sous la forme polaire :

$$S_k(m) = R_k(m)e^{j\phi_k(m)} \ , \ k = 0, 1, ..., N-1 \tag{B.2}$$

où $R_k(m)$ et $\phi_k(m)$ sont le module et la phase de la trame m de la transformée de Fourier à court terme. On pose :

$$\hat{S}_k(m) = \hat{R}_k(m)e^{j\hat{\phi}_k(m)} \ , \ k = 0, 1, ..., N-1 \tag{B.3}$$

où $\hat{R}_k(m)$ et $\hat{\phi}_k(m)$ sont les valeurs du module et de la phase "attendues" à la trame m pour le $k^{\text{ième}}$ coefficient de la transformée s'il n'y a pas de variation d'énergie et si la déviation de phase est constante. Si l'énergie ne varie pas, alors on a simplement $\hat{R}_k(m) = R_k(m-1)$. Si la déviation de phase est constante, alors cette déviation $\Delta_\phi(m)$ entre les trames $m-1$ et m est identique à celle $\Delta_\phi(m-1)$ entre entre les trames $m-1$ et m. La phase du $k^{\text{ième}}$ coefficient de la transformée à la trame m est donc égale à celle du $k^{\text{ième}}$ coefficient de la transformée à la trame $m-1$ à laquelle on ajoute la variation de phase $\Delta_\phi(m-1)$ attendue, i.e. :

$$\hat{\phi}_k(m) = \phi_k(m-1) + \underbrace{(\phi_k(m-1) - \phi_k(m-2))}_{\Delta_\phi(m-1)} \tag{B.4}$$

$$= 2\phi_k(m-1) - \phi_k(m-2) \ , \ k = 0, 1, ..., N-1 \tag{B.5}$$

En se plaçant dans le plan complexe, on peut quantifier ces variations de phase et d'énergie en mesurant la distance euclidienne entre la valeur attendue et la valeur mesurée pour le $k^{\text{ième}}$ coefficient de la transformée à la trame m :

$$\Gamma_k(m) = \left\{ \left[\mathbf{Re}(\hat{S}_k(m)) - \mathbf{Re}(S_k(m))\right]^2 + \left[\mathbf{Im}(\hat{S}_k(m)) - \mathbf{Im}(S_k(m))\right]^2 \right\}^{\frac{1}{2}} \tag{B.6}$$

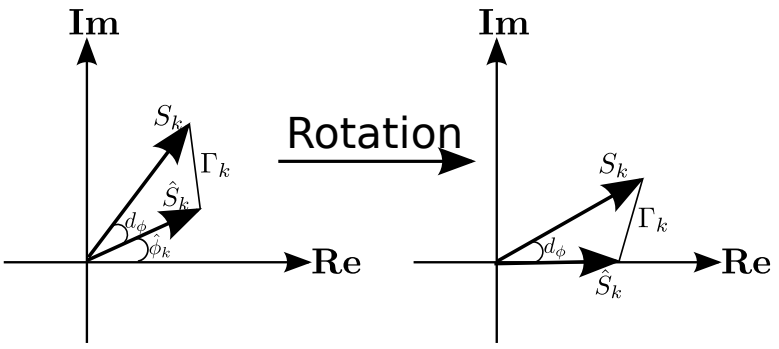

FIGURE B.1 – Rotation pour simplifier le calcul de la fonction de détection d'attaques

où **Re** et **Im** sont respectivement les parties réelles et imaginaires. Cette équation peut être simplifiée en effectuant une rotation de $\hat{S}_k(m)$ et $S_k(m)$ dans le plan complexe (cf figure B.1) de telle sorte que $\hat{S}_k(m)$ soit sur l'axe réel (i.e. $\hat{\phi}_k(m) = 0$), et on obtient alors :

$$\Gamma_k(m) = \left\{ R_k(m-1)^2 + R_k(m)^2 - 2R_k(m-1)R_k(m)\cos\left(d_{\phi_k}(m)\right) \right\}^{\frac{1}{2}} \tag{B.7}$$

avec

$$d_{\phi_k}(m) = \phi(m) - 2\phi(m-1) + \phi(m-2) \tag{B.8}$$

La fonction de détection d'attaques $\eta(m)$ est finalement obtenue à chaque trame m en sommant $\Gamma_k(m)$ sur tous les coefficients de la transformée :

$$\eta(m) = \sum_{k=0}^{N-1} \Gamma_k(m) \tag{B.9}$$

Annexe C

Crédits corpus de textures sonores

Provenance des textures sonores utilisées dans les tests du chapitre 4 :

- **Texture 1 :** Extrait de *Colours Move* par Fuck Buttons.
- **Texture 2 :** Extrait de `http://www.freesound.org/people/Jovica/sounds/51005/` (dernière visite 27 septembre 2014).
- **Texture 3 :** Banque de son interne du laboratoire, son conçu par Adrien Merer.
- **Texture 4 :** Banque de son interne du laboratoire, son conçu par Adrien Merer.
- **Texture 5 :** Banque de son interne du laboratoire, son conçu par Adrien Merer.
- **Texture 6 :** Extrait du Requiem en ré mineur de Mozart.
- **Texture 7 :** Extrait de `http://www.freesound.org/people/sandyrb/sounds/85894/` (dernière visite 27 septembre 2014).
- **Texture 8 :** Banque de son interne du laboratoire, son conçu par Adrien Merer.
- **Texture 9 :** Extrait de `http://www.freesound.org/people/hookhead/sounds/13275/` (dernière visite 27 septembre 2014).
- **Texture 10 :** Enregistrement par Simon Conan.
- **Texture 11 :** Extrait de `http://www.freesound.org/people/arightwizard/sounds/172117/` (dernière visite 27 septembre 2014).
- **Texture 12 :** Extrait de `http://www.soundsnap.com/node/18185` (dernière visite 27 septembre 2014).

Annexe D

Publications associées à la thèse

Articles de revues à comité de lecture (2)

2. **Conan, S.**, Thoret, E. , Aramaki, M., Derrien, O., Gondre, C., Ystad, S., Kronland-Martinet, R. (2014). An Intuitive Synthesizer of Continuous Interaction Sounds : Rubbing, Scratching and Rolling. *Computer Music Journal 38(4), 24–37 (Invited Paper)*

1. **Conan, S.**, Derrien, O., Aramaki, M., Ystad, S., Kronland-Martinet, R. (2014). A Synthesis Model With Intuitive Control Capabilities for Rolling Sounds. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(8), 1260–1273.* doi :10.1109/TASLP.2014.2327297

Chapitre d'ouvrage (1)

1. **Conan, S.**, Aramaki, M., Kronland-Martinet, R., Ystad, S. (2013). Intuitive Control of Rolling Sound Synthesis. In *From Sounds to Music and Emotions, LNCS vol 7900*, 99–109. Springer Berlin Heidelberg

Conférences internationales à comité de lecture (4)

4. **Conan, S.**, Thoret, E., Gondre, C., Aramaki, M., Kronland-Martinet, R., Ystad, S. (2013). An Intuitive Synthesizer of Sustained Interaction Sounds. *Sound, Music and Motion, Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, 15-18 Oct., Marseille – Demonstration

3. **Conan, S.**, Thoret, E., Aramaki, M., Derrien, O., Gondre, C., Kronland-Martinet, R., Ystad, S. (2013). Navigating in a Space of Synthesized Interactions-Sounds : Rubbing, Scratching and Rolling, *Proceedings of the 16th Conference on Digital Audio Effects (DAFx)*, 2-5 Sept. 2013, Maynooth, Ireland – Oral presentation, *Best paper award in sound synthesis*

2. **Conan, S.**, Aramaki, M., Kronland-Martinet, R., Ystad, S. (2012). Rolling Sound Synthesis : Work In Progress, *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval*, 19-22 June, London, UK – Oral Presentation

1. **Conan, S.**, Aramaki, M., Kronland-Martinet, R., Thoret, E., Ystad, S. (2012). Perceptual differences between sounds produced by different continuous interactions, *Proceedings of Acoustics 2012*, 23-27 April Nantes, France – Poster

Conférences nationales (2)

2. **Conan, S.** (2012) Recherche des invariants acoustiques liés à l'évocation d'événements sonores particuliers. Une application à la synthèse et au contrôle des sons de roulements. *2èmes Journées Perception Sonore - SFA*, 10–11 Dec. 2012, Marseille – Poster
1. Danna, J., Paz-Villagran, V., Velay, J.-L., Gondre, C., Kronland-Martinet, R., Ystad, S., Aramaki, M., Thoret, E., **Conan, S.**, Voinier, T., Omer, H., and Torrésani, B. (2012). Sonifier l'écriture : un outil pour le diagnostic et la remédiation de la dysgraphie, IIIèmes Journée Scientifique du Centre de Référence des Troubles d'Apprentissage (CERTA)/RESODYS, Marseille, May 25, 2012 – Oral Presentation

Autres publications (1)

1. Danna, J., Paz-Villagran, V., Velay, J.-L., Gondre, C., Kronland-Martinet, R., Ystad, S., Aramaki, M., Thoret, E., **Conan, S.**, Voinier, T., Omer, H., and Torrésani, B. (2012). Sonifier l'écriture : un outil pour le diagnostic et la remédiation de la dysgraphie, *Développements*, 12, 2012, 32-40

Titre – Contrôle intuitif de la synthèse sonore d'interactions solidiennes : vers les métaphores sonores

Résumé – Un des enjeux actuels de la synthèse sonore est le contrôle perceptif (i.e. à partir d'évocations) des processus de synthèse. En effet, les modèles de synthèse sonore dépendent généralement d'un grand nombre de paramètres de bas niveau dont la manipulation nécessite une expertise des processus génératifs. Disposer de contrôles perceptifs sur un synthétiseur offre cependant beaucoup d'avantages en permettant de générer les sons à partir d'une description du ressenti et en offrant à des utilisateurs non-experts la possibilité de créer et de contrôler des sons intuitivement. Un tel contrôle n'est pas immédiat et se base sur des hypothèses fortes liées à notre perception, notamment la présence de morphologies acoustiques, dénommées "invariants", responsables de l'identification d'un évènement sonore.

Cette thèse aborde cette problématique en se focalisant sur les invariants liés à l'action responsable de la génération des sons. Elle s'articule suivant deux parties. La première a pour but d'identifier des invariants responsables de la reconnaissance de certaines interactions continues : le frottement, le grattement et le roulement. Le but est de mettre en œuvre un modèle de synthèse temps-réel contrôlable intuitivement et permettant d'effectuer des transitions perceptives continues entre ces différents types d'interactions (e.g. transformer progressivement un son de frottement en un son de roulement). Ce modèle s'inscrira dans le cadre du paradigme "action-objet" qui stipule que chaque son résulte d'une action (e.g. gratter) sur un objet (e.g. une plaque en bois). Ce paradigme s'adapte naturellement à une approche de la synthèse par modèle source-filtre, où l'information sur l'objet est contenue dans le "filtre", et l'information sur l'action dans la "source". Pour ce faire, diverses approches sont abordées : études de modèles physiques, approches phénoménologiques et tests perceptifs sur des sons enregistrés et synthétisés.

La seconde partie de la thèse concerne le concept de "métaphores sonores" en élargissant la notion d'objet à des textures sonores variées. La question posée est la suivante : étant donnée une texture sonore quelconque, est-il possible de modifier ses propriétés intrinsèques pour qu'elle évoque une interaction particulière comme un frottement ou un roulement par exemple ? Pour créer ces métaphores, un processus de synthèse croisée est utilisé dans lequel la partie "source" est basée sur les morphologies sonores des actions précédemment identifiées et la partie "filtre" restitue les propriétés de la texture. L'ensemble de ces travaux ainsi que le paradigme choisi offre dès lors de nouvelles perspectives pour la constitution d'un véritable langage des sons.

Mots-clés – contrôle intuitif de la synthèse sonore, sons d'interactions, analyse-synthèse, perception des sons

---

Title – Intuitive control of solid-interaction sound synthesis : toward sonic metaphors

Abstract – Perceptual control (i.e. from evocations) of sound synthesis processes is a current challenge. Indeed, sound synthesis models generally involve a lot of low-level control parameters, whose manipulation requires a certain expertise with respect to the sound generation process. Thus, intuitive control of sound generation is interesting for users, and especially non-experts, because they can create and control sounds from evocations. Such a control is not immediate and is based on strong assumptions linked to our perception, and especially the existence of acoustic morphologies, so-called "invariants", responsible for the recognition of specific sound events.

This thesis tackles the problem by focusing on invariants linked to specific sound generating actions. If follows two main parts. The first is to identify invariants responsible for the recognition of three categories of continuous interactions : rubbing, scratching and rolling. The aim is to develop a real-time sound synthesizer with intuitive controls that enables users to morph continuously between the different interactions (e.g. progressively transform a rubbing sound into a rolling one). The synthesis model will be developed in the framework of the "action-object" paradigm which states that sounds can be described as the result of an action (e.g. scratching) on an object (e.g. a wood plate). This paradigm naturally fits the well-known source-filter approach for sound synthesis, where the perceptually relevant information linked to the object is described in the "filter" part, and the action-related information is described in the "source" part. To derive our generic synthesis model, several approaches are treated : physical models, phenomenological approaches and listening tests with recorded and synthesized sounds.

The second part of the thesis deals with the concept of "sonic metaphors" by expanding the object notion to various sound textures. The question raised is the following : given any sound texture, is it possible to modify its intrinsic properties such that it evokes a particular interaction, like rolling or rubbing for instance ? To create these sonic metaphors, a cross-synthesis process is used where the "source" part is based on the sound morphologies linked to the actions previously identified, and the "filter" part renders the sound texture properties. This work, together with the chosen paradigm offers new perspectives to build a sound language.

Keywords – intuitive control of sound synthesis, interaction sounds, analysis-synthesis, sound perception

---

CNRS - Laboratoire de Mécanique et d'Acoustique (UPR7051)
31, Chemin Joseph Aiguier, 13402 Marseille Cedex 20