

Etude croisée geste-son des articulations du violon

Florian Kaiser

Mémoire de Master 2 Recherche ATIAM année 2005-2006
Université Pierre et Marie Curie, Paris VI

Laboratoire d'accueil : Ircam - Applications Temps Réel
Responsables : Frédéric Bevilacqua, Nicolas Rasamimanana

22 juin 2006

Remerciements

Je tiens à remercier :

Frédéric Bevilacqua et Nicolas Rasamimanana pour m'avoir permis de faire ce stage, et pour leur disponibilité tout au long de ces trois mois.

Norbert Schnell et toute l'équipe d'Applications Temps Réel pour leur accueil chaleureux.

Alice Dacquet, Patrice Tisserand et Arshia Cont pour la place qu'ils ont su me faire dans leur bureau.

Table des matières

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 2 | Le violon augmenté | 5 |
| 2.1 | Capture du geste | 5 |
| 2.2 | Analyse du geste | 6 |
| 2.3 | <i>BogenLied</i> de Florence Baschet | 8 |
| 3 | Éléments d’acoustique du violon | 11 |
| 3.1 | Régimes de vibration de la corde | 11 |
| 3.1.1 | Régime stationnaire des oscillations | 11 |
| 3.1.2 | Diagramme de Schelleng | 13 |
| 3.1.3 | Transition entre deux régimes stationnaires | 14 |
| 3.2 | Conclusions pour notre étude | 16 |
| 4 | Description sonore des modes de jeu | 19 |
| 4.1 | La classification audio | 19 |
| 4.1.1 | Définition | 19 |
| 4.1.2 | Extraction des descripteurs | 20 |
| 4.1.3 | Sélection de descripteurs | 22 |
| 4.2 | Application aux modes de jeu <i>Détaché</i> , <i>Martelé</i> et <i>Spiccato</i> | 23 |
| 4.2.1 | Création d’une base de données | 23 |
| 4.2.2 | Calcul des descripteurs | 23 |
| 4.2.3 | Sélection des descripteurs | 24 |
| 4.3 | Résultats | 25 |
| 4.3.1 | Description des modes de jeu sur des notes de même hauteur | 25 |
| 4.3.2 | Variation de dynamique | 27 |
| 4.3.3 | Variations de la hauteur de la note et du joueur | 28 |
| 4.4 | Conclusion | 31 |
| 5 | Caractérisation des transitoires du violon | 33 |
| 5.1 | Retour | 33 |
| 5.2 | Extraction du résiduel | 34 |
| 5.2.1 | Extraction avec additive | 34 |
| 5.2.2 | Les méthodes à haute résolution | 37 |

| | | |
|----------|---|-----------|
| 5.2.3 | Application au son du violon | 40 |
| 5.2.4 | Confrontation méthode ESPRIT/Synthèse Additive . . . | 45 |
| 6 | Etude de la relation entre le résiduel et l'accélération | 49 |
| 6.1 | Mesures et protocole | 49 |
| 6.1.1 | Base d'étude | 49 |
| 6.1.2 | Prétraitements sur les signaux | 49 |
| 6.1.3 | Modes de jeu considérés | 50 |
| 6.2 | Résultats | 50 |
| 6.2.1 | Relation temporelle | 50 |
| 6.2.2 | Relation entre les amplitudes | 57 |
| 6.3 | Discussion | 57 |
| 7 | Conclusion et Perspectives | 59 |

Chapitre 1

Introduction

Le travail présenté dans ce rapport s'inscrit dans le cadre des recherches sur la modélisation du geste instrumental menées à l'Ircam. Ces recherches ont entre autre pour but le développement de nouveaux contrôleurs gestuels de processus sonores pour la composition. Autour de ce projet, les scientifiques sont donc en étroite collaboration avec un groupe de compositeurs, incluant Florence Baschet, Philippe Manoury, Frank Bedrossian et Jérôme Combier, qui participe activement au développement de ces nouvelles interfaces homme/machine. L'un des enjeux de cette recherche est de définir une relation entre geste et son qui procure à l'instrumentiste un contrôle des processus de synthèse avec lequel il se sente à l'aise et qui lui permette de s'exprimer pleinement.

Dans ce contexte, un violon augmenté, i.e. un violon acoustique équipé d'un système de captation du geste, a été développé à l'Ircam. Le geste est analysé via un système de capteurs montés sur l'archet. On peut alors mesurer une série de signaux allant de la position de l'archet sur la corde à l'accélération de l'archet. Ces signaux peuvent ensuite servir à la fois de base d'étude pour la description du geste, et comme paramètres d'entrée pour le contrôle de processus électroniques. L'utilisation d'instruments augmentés pour le contrôle de synthèse présente un intérêt majeur : l'utilisateur a déjà acquis une technique de jeu, de laquelle il sait tirer profit afin de donner de l'expressivité à son jeu.

Se pose alors la question de la relation qui existe entre les signaux gestuels mesurés par le violon augmenté, et le son produit par le violon. D'après les études acoustiques, une partie de la réponse à cette question se trouve dans l'étude des transitoires du violon, i.e. les articulations entre notes. C'est la voie que nous empruntons dans ce rapport pour caractériser la l'interaction geste-son sur le violon.

Méthodologie

L'analyse du geste avec le violon augmenté s'est articulée autour de la caractérisation de trois techniques de jeu, i.e. détaché, martelé, et spiccato [6]. L'étude gestuelle de ces techniques de jeu, chacune associée à un mouvement

d'archet précis, a permis de déterminer un espace de descripteurs gestuels dans lequel chaque mode de jeu est distinct. Cette modélisation est intéressante car elle permet de caractériser à la fois les différences et les similarités entre les modes de jeu. En effet, on observe des régions dans l'espace de descripteurs où les modes de jeu forment des sortes d'hybrides, qui sont tout à fait cohérents avec la façon de jouer de l'instrumentiste.

Une propriété intéressante de cette modélisation du geste est qu'elle est fortement invariante aux variations de dynamiques et de tempo, et dans une certaine mesure au niveau du joueur.

Notre étude croisée geste-son s'articulera également autour de l'analyse de ces trois modes de jeu. Il y a cela deux raisons majeures. Premièrement on dispose déjà d'une modélisation gestuelle de chacune de ces techniques de jeu. On pourra donc essayer de déterminer une modélisation des modes de jeu qui soit basée cette fois sur une description sonore, et qui satisfasse la même propriété d'invariance aux variations de dynamiques et de joueurs que le modèle gestuel. On pourra alors tenter de définir une relation entre un modèle gestuel et un modèle sonore des modes de jeu (*cf* fig. 1.1).

Deuxièmement, l'articulation entre ces techniques de jeu dans un contexte musical donne à l'instrumentiste un moyen riche d'expressivité. Il sera intéressant de traduire dans la relation geste-son ce moyen d'expressivité, d'abord en caractérisant la relation geste-son qui existe pour chacun des modes de jeu, puis en essayant de définir un modèle qui puisse traduire une continuité entre les modes de jeu.

L'étude des parties transitoires constitue un des points clés de notre approche. En effet, les études acoustiques menées par Guettler [4] ont montré que le couple Force/Accélération donné à l'archet pendant les transitoires conditionne la mise en vibration de la corde (*cf* chapitre 3). Guettler émet l'hypothèse que les intentions d'expression de l'instrumentiste peuvent être traduites par ce couple Force/Accélération. Pour montrer cela, Guettler a utilisé des simulations d'attaques, avec des profils d'accélération constante. Nous proposons d'aller plus loin dans ce champ en étudiant maintenant des profils d'accélération des transitoires correspondant à une situation réelle de jeu, et en les confrontant au signal sonore.

Nous disposons pour l'instant uniquement d'une modélisation gestuelle. La première étape dans cette étude sera donc de donner une description fine des sons associés à chaque mode de jeu, afin de déterminer un espace dans lequel chaque mode de jeu peut être représenté.

Dans un premier temps, nous utiliserons les nombreux outils développés pour cela dans le cadre de la classification audio des instruments (*cf* chapitre 4).

On montrera alors que le modèle obtenu avec une telle description n'est pas satisfaisant, premièrement parce qu'il ne permet pas de donner une description simple des modes de jeu qui soit valable pour plusieurs joueurs et des notes

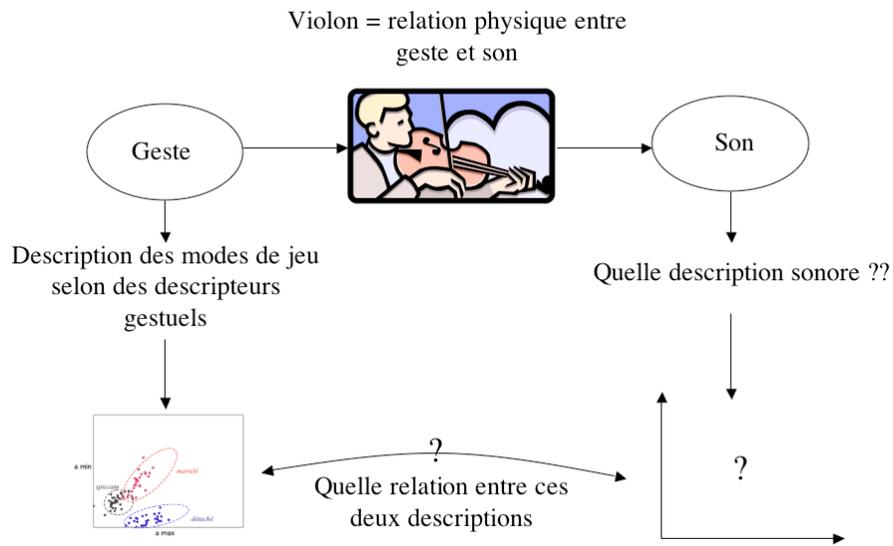


FIG. 1.1 – Dans l'étude croisée geste-son sur le violon, une approche consiste à mettre en parallèle une description du geste avec une description du son

de hauteurs différentes, et deuxièmement parce qu'elle ne permet pas, de par les outils d'analyse sur lesquels elle repose, une caractérisation des transitions entre notes.

Il conviendra donc de se pencher ensuite sur l'étude des transitoires en séparant les contributions harmoniques et non-harmoniques dans le signal sonore. Nous montrerons que cette séparation constitue un outil puissant pour la caractérisation des transitoires et qu'il existe de plus une forte relation de temporalité, variable selon les modes de jeu, entre le bruit émis par le violon au moment des transitoires et le profil d'accélération de l'archet (*cf* chapitre 5 et 6). On donnera ainsi un moyen efficace pour caractériser les parties transitoires correspondant à chaque mode de jeu. A terme, cette description pourrait nous permettre de définir une continuité entre les mouvements d'archet, et ainsi de pouvoir modéliser des articulations entre les modes de jeu à des fins expressives.

Chapitre 2

Le violon augmenté

Dans le cadre des recherches sur la modélisation du geste instrumental et le contrôle gestuel de processus électroniques pour la composition, un violon augmenté, i.e. un violon acoustique équipé d'un système de captation du geste, a été développé à l'Ircam.

Dans ce chapitre, nous présentons ce violon augmenté et les données du geste qu'il permet de mesurer. Nous introduirons ensuite les résultats de l'étude menée sur la modélisation du geste de l'archet, pour enfin présenter brièvement la pièce *Bogenlied* de Florence Baschet, écrite pour le violon augmenté.

2.1 Capture du geste

Le violon augmenté de l'Ircam est équipé d'un système de captation du geste développé par E.Flety en 2004. Ce système sans fil est composé d'un microcontrôleur, de deux accéléromètres et d'un émetteur radio numérique. Les principales données mesurables avec un tel système sont l'accélération de l'archet, la position de l'archet et l'angle d'inclinaison de l'archet par rapport à la corde. Ces données sont transmises à un ordinateur via *EtherSense*, i.e. un système d'acquisition développé par E.Flety [2], pour être traitées en temps réel ou différé.

Mesure de l'accélération

Deux accéléromètres placés sous le talon de l'archet permettent de mesurer l'accélération de l'archet selon trois axes : dans la direction de l'archet, dans la direction de la corde, et perpendiculairement à l'archet dans la direction verticale.

La vitesse de l'archet peut ensuite être calculée par intégration des signaux d'accélération.

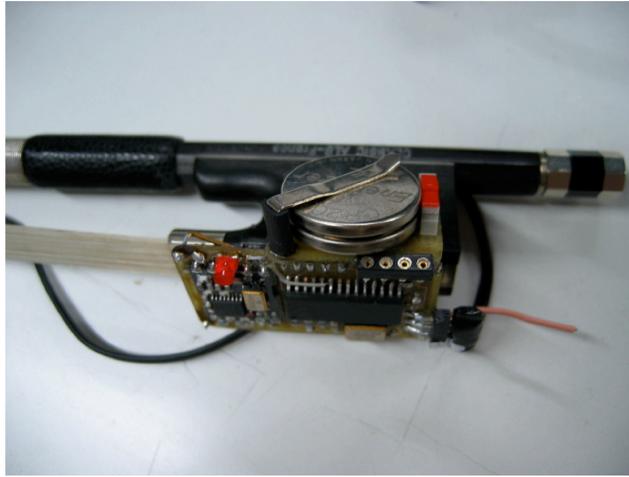


FIG. 2.1 – Archet augmenté

Un système non-intrusif

La contrainte principale imposée pour la conception d'un instrument augmenté est la non-intrusivité du système de captation du geste. En effet, il est important que l'instrumentiste ne soit pas perturbé en situation de jeu. L'archet de l'Ircam est représenté sur la figure 2.1. Le poids total du système est de 17g. Tous les professionnels ayant testé l'archet ont jugé les conditions de jeu acceptables, à la fois par rapport à la position des doigts sur l'archet et à la surcharge de poids apportée par les capteurs.

2.2 Analyse du geste

Autour de tout instrument acoustique se construit un vocabulaire gestuel, exploité et modifié au cours de l'histoire par l'instrumentiste. Ce vocabulaire gestuel donne au joueur une source riche d'expressivité.

En donnant accès à la mesure du geste de l'instrumentiste en situation de jeu, les instruments augmentés permettent d'étudier ce vocabulaire gestuel associé à l'instrument.

Dans le cadre du projet "Violon Augmenté" de l'Ircam, l'approche adoptée par N.Rasamimanana et F.Bevilacqua a donc été de caractériser et de modéliser au moyen du violon augmenté les différentes techniques de jeu du violon. Une description détaillée des travaux menés dans ce champ pourra être trouvée dans [6]. Dans cette partie, nous résumons les principaux résultats obtenus.

Trois modes de jeu sont considérés :

- *Détaché* : la course de l’archet est inversée à chaque note (tiré, poussé). La vitesse de l’archet est quasi constante. La note est attaquée de façon plus ou moins molle.
- *Martelé* : accélération très forte de l’archet au début de la note. L’archet est ensuite brusquement stoppé. Ce geste assez violent donne une attaque presque percussive.
- *Spiccato* : L’archet commence sa course au dessus de la corde. Il décrit une sorte de parabole et frotte la corde une fois arrivé au point le plus bas de la parabole. Le son est léger et précis.

A chacun de ces modes de jeu correspond un mouvement d’archet précis.

Les coups d’archet correspondants aux trois modes de jeu ont été enregistrés par des violonistes professionnels et amateurs, effectuant des gammes dans deux tempos, 60 bpm et 120 bpm, et trois dynamiques *pianissimo*, *mezzoforte* et *fortissimo*.

Le signal d’accélération dans la direction de l’archet est représenté pour les trois modes de jeu sur la figure 2.2. Il présente généralement deux pics, désignés sur la figure par a_{max} et a_{min} . La figure 2.3 représente les courbes d’accélération et de vitesse pour un coup d’archet en détaché.

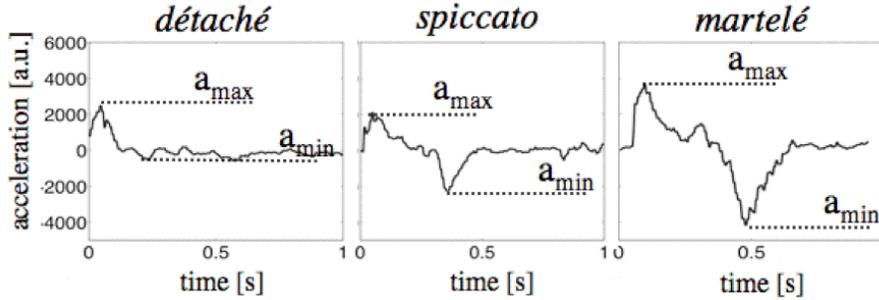


FIG. 2.2 – Signal d’accélération de l’archet pour un coup d’archet dans les trois modes de jeu (*mf*, *60bpm*)

Un ensemble de descripteurs du mouvement de l’archet a été développé sur la base des signaux mesurés par le violon augmenté. Cet ensemble contient notamment a_{max} , a_{min} , v_{min} et v_{max} .

Une Analyse Linéaire Discriminante (LDA [9]) sur l’ensemble de ces descripteurs montre que a_{max} et a_{min} sont les descripteurs les plus pertinents pour la modélisation des coups d’archet. La figure 2.4 représente différents coups d’archet donnés dans les trois modes de jeu selon les axes a_{max} et a_{min} . On observe que les trois modes de jeu occupent chacun une zone distincte et relativement compacte de l’espace. Ces zones sont appelées *clusters*.

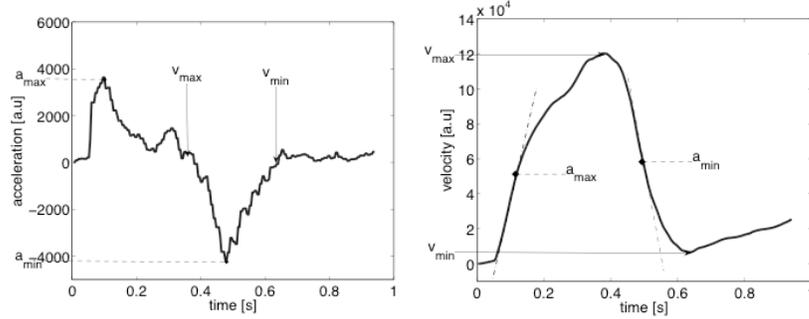


FIG. 2.3 – Paramètres a_{max} , a_{min} , v_{min} et v_{max} extraits des courbes d'accélération et de vitesse d'une note jouée en martelé

La variabilité des clusters formés par chaque mode de jeu a été étudiée en fonction de changements de nuances, de tempo, et de joueurs. Les clusters montrent une forte propriété d'invariance pour ces trois variables. Un système a donc été implémenté sous Max/MSP pour la reconnaissance en temps réel du mode de jeu dans lequel un coup d'archet est joué.

2.3 *BogenLied* de Florence Baschet

Le résultat de l'analyse gestuelle des trois modes de jeu détaché, martelé et spiccato, a été utilisé pour la composition de *BogenLied*, une pièce mixte de Florence Baschet pour violon augmenté. La pièce se divise en neuf sections (fig. 2.5), alternativement avec et sans processus électroniques. Dans les sections *II*, *IV* et *VI*, la partition est centrée sur l'articulation entre les trois modes de jeu. A chacun de ces modes de jeu est associé un processus de synthèse spécifique. Dans la section *VIII*, les trois modes de jeu apparaissent successivement. Le choix du processus de synthèse est alors déterminé par le système de reconnaissance des modes de jeu.

Dans la dernière section, l'instrumentiste joue des sortes d'hybrides entre les modes de jeu. L'électronique est alors construite par une combinaison des processus des sections *II*, *IV* et *VI*.

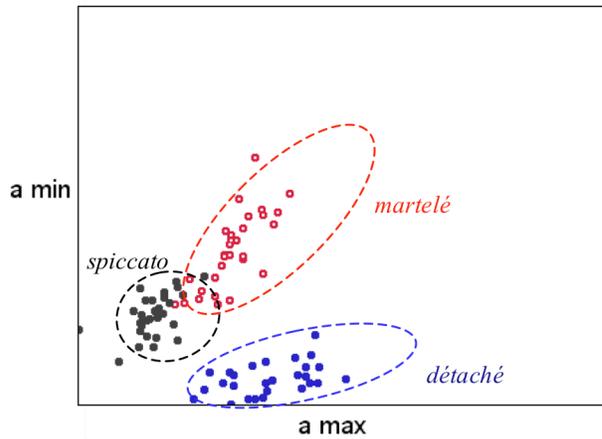


FIG. 2.4 – Représentation des coups d’archet selon a_{max} et a_{min} . Chaque point correspond à un coup d’archet l’un des trois modes de jeu : détaché en bleu, martelé en rouge et spiccato en noir. Dans cet espace, les modes de jeu sont bien distincts et peuvent être reconnus.

| Gesture analysis | sound | bowing style | section |
|--|----------|----------------------------|---------|
| | acoustic | | I |
| gesture intensity | mixed | <i>détaché</i> | II |
| | acoustic | | III |
| gesture intensity | mixed | <i>martelé</i> | IV |
| | acoustic | | V |
| gesture intensity | mixed | <i>spiccato</i> | VI |
| | acoustic | | VII |
| gesture intensity + bow stroke recognition | mixed | <i>all 3 bowing styles</i> | VIII |
| gesture intensity + bow stroke recognition | mixed | <i>hybrid</i> | IX |

FIG. 2.5 – Structure de BogenLied

Chapitre 3

Eléments d'acoustique du violon

A la base de l'interaction geste-son sur le violon se trouve le frottement de l'archet sur la corde. Dans ce chapitre, seront introduits quelques éléments sur les différents régimes de vibration des cordes frottées.

On présentera dans un premier temps le mouvement de Helmholtz, qui donne une modélisation du mouvement de la corde lorsque celle-ci oscille dans un régime stationnaire. On verra ensuite avec le diagramme de Schelleng les paramètres du mouvement de l'archet qui conditionnent les oscillations de la corde dans un tel régime.

Le dernier point abordé concerne spécifiquement le régime transitoire de vibration de la corde. Nous verrons avec les travaux de Guettler quels paramètres du mouvement conditionnent la perception d'une attaque.

On exploitera ensuite cette description physique de l'interaction corde-archet pour orienter notre étude.

3.1 Régimes de vibration de la corde

3.1.1 Régime stationnaire des oscillations

A l'oeil nu, une corde frottée semble vibrer comme une corde libre dans son mode fondamental. Mais la réalité est différente. Helmholtz a montré en 1877 qu'une fois un son soutenu obtenu, la corde prend la forme d'un V, dont la pointe se déplace selon une trajectoire de type parabolique (fig. 3.1) . Cette trajectoire est parcourue une fois à chaque période de la fondamentale de la note jouée. La pointe du "V" formé est appelée "Coin de Helmholtz". Cette description du mouvement de la corde frottée a depuis été vérifiée et complétée, notamment par Schelleng [10].

Le frottement entre l'archet et la corde est un frottement sec. Le système corde-archet alterne donc deux états :

- Adhérence : l'archet et la corde sont solidaires et ont donc la même vitesse.
- Glissement : rupture d'adhérence déclenchée par le passage du coin de Helmholtz sous l'archet. Le mouvement de la corde s'oppose alors à celui de l'archet.

Le mouvement de la corde sur une période est décomposé sur la figure 3.2. En 1, la corde et l'archet sont solidaires. Le passage du coin de Helmholtz sous l'archet entre 1 et 2 provoque une rupture d'adhérence, et la corde glisse. L'archet et la corde retrouvent un état d'adhérence une fois que la corde a repris spontanément la vitesse de l'archet.

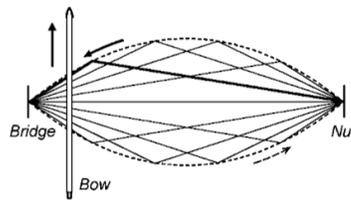


FIG. 3.1 – Mouvement de Helmholtz

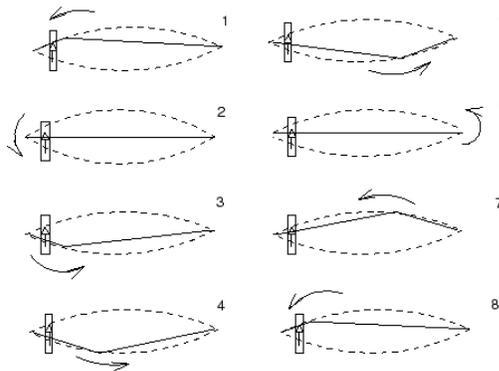


FIG. 3.2 – Décomposition du mouvement de Helmholtz

A cette description idéale du mouvement de Helmholtz, il faut ajouter certains éléments. Premièrement, la rigidité de la corde provoque un arrondissement du coin de Helmholtz. D'après Cremer [1], l'arrondi du coin de Helmholtz a une forte influence sur le contenu spectral du son. Plus le coin est aiguisé, plus le son est brillant. C'est pour cela que beaucoup d'études acoustiques sur l'interaction

geste-son ont pour objet le contrôle des paramètres gestuels sur l'arrondi du coin de Helmholtz. Deuxièmement, la vibration réelle de la corde de violon est une superposition de plusieurs mouvements de Helmholtz.

3.1.2 Diagramme de Schelleng

Pour que la corde entre dans un régime de Helmholtz et n'en sorte pas, l'archet doit fournir suffisamment d'énergie à la corde pour que celle-ci puisse s'opposer à son mouvement et déclencher les phases de glissement. Les conditions limites pour le maintien du mouvement de Helmholtz en fonction de la pression d'archet, de la position de l'archet et de la vitesse de l'archet sont résumées dans le diagramme de Schelleng (fig. 3.3) [10]. Schelleng observe des changements de timbre causés par une variation de la pression d'archet. Il délimite ainsi sur son diagramme des régions auxquelles sont associées des caractéristiques de timbre. Par exemple, pour des pressions d'archet élevées, il note que le son est plus brillant. Même s'il ne définit pas une relation explicite entre geste et son, ce diagramme est l'une des premières études visant à donner une interprétation sonore au geste de l'instrumentiste.

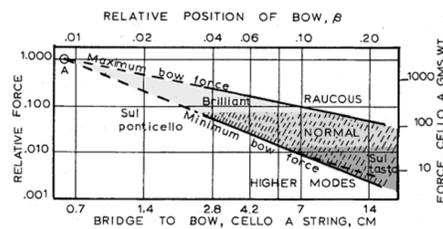


FIG. 3.3 – Diagramme de Schelleng pour une vitesse donnée

Vers une étude croisée geste-son...

On peut alors préciser les caractéristiques du geste que l'on considère pour l'étude croisée geste-son. Sur le diagramme de Schelleng, le geste est représenté selon trois paramètres :

- Pression de l'archet sur la corde
- Vitesse de l'archet
- Position de l'archet sur la corde

K.Guettler, E.Schoonderwaldt et A.Askenfelt [3] ont étudié l'influence de la vitesse et de la position de l'archet sur le spectre du son. A l'aide d'un archet mécanique, ils sont capables de ne faire varier qu'un seul paramètre gestuel à la fois, et d'étudier l'influence de cette variation sur le spectre du son. Ils concluent que la position de l'archet n'a pas d'influence majeure sur le spectre, tandis qu'une variation seule de la pression ou de la vitesse de l'archet modifie fortement le contenu hautes-fréquences du spectre, et donc le timbre. Ils ajoutent

aux paramètres gestuels l'inclinaison de l'archet par rapport à la corde. Dans [3], K.Guettler résume les principaux résultats obtenus dans les études menées sur l'influence des paramètres gestuels sur le timbre (tableau 3.1)

TAB. 3.1 – Influence des paramètres gestuels sur le timbre :

| <i>Paramètre gestuel</i> | <i>Timbre</i> |
|--------------------------------------|--|
| augmentation de la pression d'archet | Les partiels hautes fréquences sont accentués, le son est donc plus brillant |
| Augmentation de la vitesse d'archet | Les partiels hautes fréquences sont atténués, le son est donc moins brillant |
| position | influence moindre |
| Inclinaison | son sensiblement plus brillant |

3.1.3 Transition entre deux régimes stationnaires

D'une façon générale, le régime transitoire est défini comme faisant le lien entre deux états d'un système. Dans le cas d'une corde frottée, il s'agit de l'établissement du régime de Helmholtz, ou parfois de la transition d'un régime de Helmholtz à un autre régime de Helmholtz (ex : transition entre deux notes jouées en détaché).

Le mouvement de la corde pendant l'attaque est assez chaotique. Sur la figure 3.4, on désigne par Pre-Helmholtz l'établissement du mouvement de Helmholtz. On voit sur cette figure que la vibration de la corde (et donc le son) est fortement apériodique pendant le régime de Pre-Helmholtz. On considère que le régime de Helmholtz est établi une fois qu'il y a déclenchement périodique (à la fréquence de la fondamentale) des phases de glissement et d'adhérence.

K.Guettler a étudié les transitoires d'attaque des cordes frottées [5] et [4]. Il s'est notamment intéressé à la perception des bonnes et des mauvaises attaques. Son parti-pris est que plus le temps d'établissement du mouvement de Helmholtz est court, meilleure est l'attaque.

A l'aide d'un archet mécanique, Guettler simule des attaques pour des pressions et accélérations d'archet variables. Pour chaque combinaison de ces deux paramètres, il mesure le nombre de périodes avant création du mouvement de Helmholtz sur le signal de vitesse de la corde (à chaque glissement, la vitesse de la corde chute par rapport à la vitesse de l'archet).

Le résultat est représenté sur la figure 3.5.

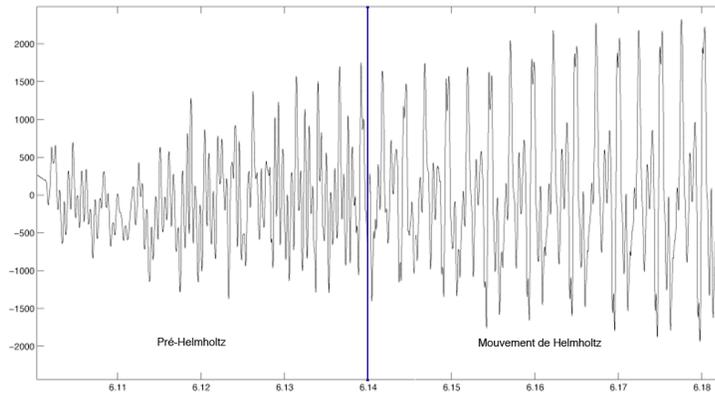


FIG. 3.4 – Enregistrement avec un microphone du transitoire d’attaque d’une note de violon. On ne visualise pas à proprement parler la vibration de la corde. Les résonances de la caisse du violon et de la salle viennent s’ajouter à la vibration de la corde.

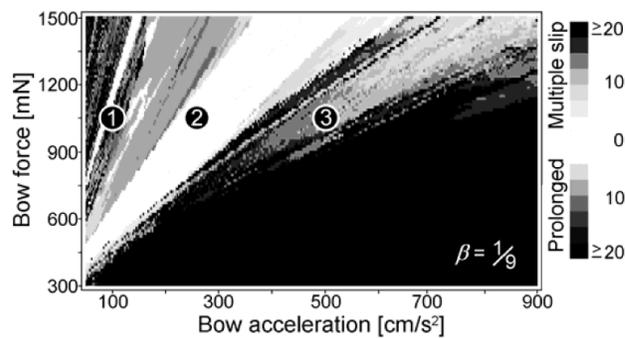


FIG. 3.5 – Simulation d’attaques. Chaque pixel représente le nombre de période avant le déclenchement du mouvement de Helmholtz. Trois catégories d’attaque : (1) :”Choked” ou ”Creaky”, (2) :”Perfect”, (3) :”Loose” ou ”Scratchy”

En fonction du couple Pression/Accélération, plusieurs scénarios sont alors possibles. Dans le cas (1), on a une forte pression pour une accélération faible. Pendant la création du coin de Helmholtz, l'intervalle entre deux glissements est alors plus long que la période de la fondamentale (*Prolonged* sur le diagramme). Guettler caractérise alors le son de "creaky" (Grinçant). Sur le diagramme de Schelleng, cette situation correspond à la zone appelée "Raucous".

Dans le cas (2), on a tout de suite un glissement par période. Pour Guettler, l'attaque est alors parfaite.

Dans le dernier cas, on donne beaucoup d'accélération à l'archet, en appliquant une pression faible. On a alors la situation (1) inversée, à savoir plusieurs glissements par période de la fondamentale (*multiple slip* sur le diagramme). Le son donne alors l'impression d'être flûté (Guettler appelle ce type de son "scratchy").

Ces simulations d'attaques ont fait l'objet d'une étude perceptive. Un groupe de violonistes professionnels a donné son jugement sur la qualité des attaques simulées. Sur 1694 attaques simulées avec des couples Force/Accélération également répartis dans le diagramme 3.5, 50 % sont jugées "parfaites" et utilisables dans un contexte musical. Pour Guettler, le résultat de cette étude montre que chaque violoniste est capable de naviguer dans le diagramme 3.5 à des fins d'expressivité.

Considérer l'accélération de l'archet dans l'étude des transitions paraît donc être une piste très prometteuse.

3.2 Conclusions pour notre étude

L'étude de Guettler sur le régime transitoire de vibration des cordes frottées soulève un point fondamental pour notre étude : en donnant plus ou moins d'accélération à l'archet, le violoniste peut modifier la perception du transitoire d'attaque. Dans un contexte musical, cette possibilité de variation du transitoire donne au violoniste un moyen riche d'expression. Le diagramme établi par Guettler (3.5) nous conforte dans l'idée que la compréhension de la relation geste-son à des fins expressives doit se faire sur la transition entre les notes, c'est à dire quand la corde oscille entre deux régimes de vibration.

En appréhendant la relation geste-son avec le violon augmenté, on considère le geste dans un contexte de jeu. Il y a là une différence fondamentale avec les travaux présentés dans cette partie. En effet, l'utilisation d'un archet mécanique dans [3] et [4] permet d'étudier indépendamment l'influence de chaque paramètre gestuel sur le son. Or dans une situation de jeu, le geste de l'instrumentiste est une combinaison complexe de tous ces paramètres.

L'intérêt d'étudier la relation geste-son avec un violon augmenté est donc que le résultat sera directement applicable dans un contexte musical. Appliquer les résultats obtenus avec un archet mécanique à une situation de jeu paraît être une tâche plus ardue.

Dans les études acoustiques, l'influence du geste sur le son est essentiellement définie en terme de variation de timbre. Or le timbre est un concept assez flou et très mal défini. Sans description fine du son, il paraît difficile de quantifier précisément l'interaction geste-son. Dans la suite de cette étude, nous allons donc nous attacher à donner une paramétrisation plus fine du son.

Chapitre 4

Description sonore des modes de jeu

A chaque mode de jeu du violon est associée une sonorité particulière. Pour mettre en relation cette sonorité avec la modélisation gestuelle des modes de jeu donnée par le violon augmenté, il nous faut déterminer les propriétés sonores qui sont spécifiques à chacun de ces modes, et qui permettent de les différencier voire de les modéliser.

Dans ce chapitre, on se propose de donner une description de chaque mode de jeu. Nous utilisons pour cela les nombreux outils de description des sons musicaux développés dans le cadre de la classification audio des sons instrumentaux et de la parole.

Avec cette étude, nous mettrons en avant que la description obtenue avec ces outils permet une caractérisation de la partie entretenue des sons, mais pas la modélisation des parties transitoires.

4.1 La classification audio

4.1.1 Définition

Le but de la classification audio est d'organiser une base de données (en l'occurrence des sons) selon un ensemble de classes. Chaque classe est associée à une ou plusieurs propriétés particulières du son. Prenons un exemple : pour une tâche de classification de timbres, on veut définir les classes "son brillant" et "son sourd". Le but est de déterminer selon quelle propriété du signal sonore on peut différencier ces deux types de son. Pour ce faire, on prend un échantillon de chaque classe, et on utilise le traitement de signal pour les décrire. On pourra par exemple observer la transformée de Fourier des deux échantillons, et remarquer que dans les hautes fréquences, l'énergie du son brillant est plus élevée que celle du son sourd. En représentant notre base de sons selon l'axe "*Energie dans*

les hautes fréquences”, on peut alors distinguer les sons brillants des sons sourds.

A travers cet exemple simple, on voit se dessiner un protocole de classification : extraction avec le traitement de signal de descripteurs des données de chaque classe, sélection des descripteurs qui permettent de distinguer les classes, et extraction des descripteurs sur une base de données plus large que l’on peut ainsi classifier.

Le but de notre étude n’étant pas de classifier les sons du violon, mais de décrire et de reconnaître les sons appartenant à chacun des modes de jeu, nous nous arrêterons à l’étape *sélection des descripteurs*.

4.1.2 Extraction des descripteurs

L’ensemble des descripteurs développés pour les diverses applications de la classification étant très large, il est difficile d’établir une hiérarchie limpide entre les descripteurs. Néanmoins, les descripteurs appliqués aux sons musicaux et à la parole peuvent être distingués selon leur mode d’extraction ;

Descripteurs globaux et instantanés On distingue deux grandes familles parmi les descripteurs. La première regroupe les descripteurs qui sont calculés sur une fenêtre d’analyse, comprenant tout le signal. Ces descripteurs donnent donc une description globale du son. Par exemple, le temps d’attaque est un descripteur global.

La seconde famille est constituée des descripteurs instantanés. Chaque descripteur est calculé sur une fenêtre d’analyse translatée sur tout le signal.

Voici maintenant les cinq groupes de descripteurs que nous avons considérés :

Descripteurs temporels Ils sont extraits de chaque trame de signal sans pré-traitement. Exemple : coefficients d’auto-corrélation, Zero-Crossing-Rate (taux de passage par zero).

Descripteurs spectraux Calculés à partir du spectre du signal sonore. Exemple : barycentre spectral, taux de décroissance du spectre , etc...

Descripteurs harmoniques A partir d’une modélisation harmonique du signal, on peut calculer par exemple le rapport harmoniques/bruit, tristimulus (rapport entre l’amplitude des premières harmoniques et la somme de toutes les harmoniques) , etc...

Descripteurs perceptifs On applique au spectre un modèle perceptif de l’oreille.

Descripteurs du contenu énergétique Exemples : puissance du signal, puissance de la modélisation harmonique et du résiduel.

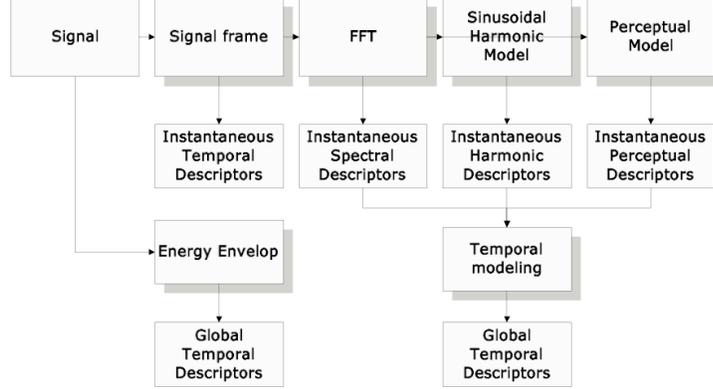


FIG. 4.1 – Extraction des descripteurs [G.Peeters]

Dans le cadre du projet CUIDADO, G.Peeters a créé un ensemble de descripteurs, dont le détail pourra être trouvé dans [8].

Quelques descripteurs importants...

Le centroïde spectral Le centroïde spectral, ou centre de gravité du spectre. Les études perceptives montrent qu'il est un des paramètres les plus pertinents dans la description du timbre. Dans [8] il est défini par :

$$\mu = \int_0^1 \nu a(\nu) d\nu \quad (4.1)$$

avec ν la fréquence réduite et $a(\nu)$ le spectre d'amplitude normalisé.

L'étendue spectrale : L'étendue spectrale (ou spectral spread) est une mesure de l'étendue du spectre autour de sa valeur moyenne. Elle calculée avec du moment d'ordre 2 :

$$\sigma^2 = \int_0^1 (\nu - \mu)^2 a(\nu) d\nu \quad (4.2)$$

avec μ la valeur moyenne de $a(\nu)$.

Spectral slope : Mesure le taux de décroissance du spectre d'amplitude par une régression linéaire.

Tristimulus : Les tristimuli donnent une description des premiers harmoniques. Ils sont de trois types :

$$Tristimulus1 = \frac{a(1)}{\sum_h a(h)} \quad (4.3)$$

$$Tristimulus2 = \frac{a(1) + a(2) + a(3)}{\sum_h a(h)} \quad (4.4)$$

$$Tristimulus3 = \frac{\sum_{h=5:H} a(h)}{\sum_h a(h)} \quad (4.5)$$

avec $a(k)$ l'amplitude de la k_{ieme} harmonique.

4.1.3 Sélection de descripteurs

De manière générale, à chaque tâche de classification correspond un ensemble de descripteurs. Le champ d'application de la classification s'élargissant, le nombre de descripteurs a considérablement augmenté ces dernières années. Décrire un signal avec l'ensemble de ces descripteurs introduit forcément des informations redondantes (descripteurs colinéaires). Les prendre tous en compte dans un algorithme de classification affecterait donc les performances de reconnaissance des classes. Parallèlement à la croissance ininterrompue de la population des descripteurs se sont donc développés des algorithmes de sélection de descripteurs.

On présente maintenant un critère de sélection de descripteurs *Inertia Ratio Maximization*.

Inertia Ratio Maximization Pour éviter la confusion entre les classes, les valeurs d'un descripteur pour une classe spécifique doivent être distinctes des valeurs du même descripteur pour toutes les autres classes. On peut exprimer cela de façon mathématique par le ratio de la variance du descripteur à l'intérieur de la classe spécifiée sur la variance du descripteur pour l'ensemble des classes . On considère K classes. On définit N_k le nombre de sons représentant la classe k , et N le nombre total de sons. Pour le i_{eme} descripteur d_i , on définit le ratio r :

$$r = \frac{B}{T} = \frac{\sum_{k=1}^K \frac{N}{N_k} (m_{i,k} - m_i)(m_{i,k} - m_i)'}{\frac{1}{N} \sum_{n=1}^N (d_{i,n} - m_i)(d_{i,n} - m_i)'} \quad (4.6)$$

avec m_i la moyenne du descripteur d_i sur l'ensemble des classes, et $m_{i,k}$ sa moyenne sur les données de la classe k . B est appelée *Between-class inertia Matrix*, T est appelée *Total class inertia Matrix*.

Si pour le descripteur d_i le ratio r est grand, alors ce descripteur discrimine bien les classes. Pour sélectionner nos descripteurs, on calcule donc r pour chacun des descripteurs, et on sélectionne ceux pour lesquels r est le plus grand.

4.2. APPLICATION AUX MODES DE JEU DÉTACHÉ, MARTELÉ ET SPICCATO 23

Ce critère est appelé *Inertia Ratio Maximization* .

On peut également utiliser un second critère pour s'assurer que l'ensemble des descripteurs sélectionnés ne contiennent pas d'informations redondantes. Un moyen simple est de projeter orthogonalement le dernier descripteur sélectionné sur l'ensemble des descripteurs déjà sélectionnés.

Les résultats obtenus avec ces deux critères de sélection pour la classification d'instruments acoustiques sont illustrés dans [7].

4.2 Application aux modes de jeu *Détaché, Martelé et Spiccato*

On définit trois classes de sons : *note jouée en détaché*, *note jouée en martelé*, et *note jouée en spiccato*. Le but de cette étude est de déterminer un ensemble de descripteurs sonores dans lequel on puisse distinguer ces trois classes.

4.2.1 Création d'une base de données

Pour représenter chaque classe dans l'algorithme de sélection des descripteurs, nous avons besoin d'enregistrements sonores de notes jouées dans les trois modes de jeu. Pour cela, nous avons utilisé les enregistrements réalisés dans le cadre de l'analyse gestuelle des coups d'archet avec le violon augmenté (*cf* chapitre 1). Douze élèves d'une classe de violon, ainsi que leur professeur et son assistant ont participé à ces enregistrements. Le niveau technique des élèves va de moyen (DEM 1ère année) à bon (Prix supérieur).

Chaque violoniste joue une gamme dans les trois modes de jeu, *détaché*, *martelé* et *spiccato*, et dans trois dynamiques (*pp* et *ff*).

Le signal audio est enregistré avec un micro cardioïde KM-140 placé au dessus des écouteilles du violon. Le signal est ensuite numérisé via une carte son M-Audio 410. Dans le même temps, on enregistre les paramètres gestuels transmis par le violon augmenté.

Notre base de sons a été constituée en découpant les notes de chacune des gammes. Chaque mode de jeu est donc représenté par des notes de hauteurs différentes, jouées dans plusieurs dynamiques, et par des instrumentistes de niveau variable.

Le signal audio d'une note jouée dans les trois modes de jeu est représenté sur la figure 4.2. A chaque mode de jeu est associée une forme d'onde particulière.

4.2.2 Calcul des descripteurs

Les descripteurs sont calculés pour chaque son de la base de données. Nous avons utilisé pour cela l'algorithme GDclass développé par G.Peeters dans le

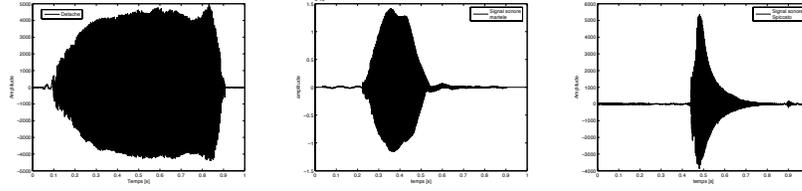


FIG. 4.2 – De gauche à droite, forme d’onde d’une note jouée en détaché, martelé et spiccato dans la nuance *mf*.

cadre du projet européen CUIDADO. Le set de descripteurs calculés dans GD-class est détaillé dans [8]. On a ainsi un ensemble de 60 descripteurs instantanés pour décrire chacun des sons de la base de données.

Les paramètres définis pour l’extraction sont les suivants :

- Les descripteurs sont calculés toutes les 10 ms sur une fenêtre de 30 ms
- Pour la modélisation harmonique du signal, la fréquence fondamentale est cherchée dans l’intervalle [50Hz 150Hz] et les partiels dans l’intervalle [150Hz 10kHz].

4.2.3 Sélection des descripteurs

Les descripteurs calculés pour une note peuvent s’écrire sous la forme matricielle suivante :

$$\mathbf{note} = \begin{pmatrix} & \text{descripteur1} & \text{descripteur2} & \dots & \text{descripteurK} \\ \text{observation} - 1 : & d_{1,1} & d_{2,1} & \dots & d_{K,1} \\ \text{observation} - 2 : & d_{1,2} & d_{2,2} & \dots & d_{K,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{observation} - n : & d_{1,n} & d_{2,n} & \dots & d_{K,n} \end{pmatrix}$$

Avec $d_{i,k}$ la valeur du i_{eme} descripteur à l’instant d’observation k , K le nombre de descripteurs, et n la durée de la note.

On applique à l’ensemble de nos descripteurs le critère de sélection défini dans la première partie de ce chapitre, i.e. *Inertia Ratio Maximization*. La variance et la moyenne de chaque descripteur sont donc calculées à la fois sur les notes représentant chacune des classes et sur l’ensemble des notes représentant les trois classes confondues. On calcule ensuite le ratio r (eq. 4.4) pour chaque descripteur, et on choisit ceux pour lesquels il est le plus grand.

4.3 Résultats

4.3.1 Description des modes de jeu sur des notes de même hauteur

Une large partie des descripteurs de GDclass est calculée à partir du contenu spectral du son, et on peut s’attendre à ce que la valeur de ces descripteurs varie beaucoup avec la hauteur de la note. Nous avons donc étudié dans un premier temps la description des modes de jeu en considérant des notes de même hauteur, jouées dans les trois modes de jeu.

Observation des modes de jeu et sélection des descripteurs :

On extrait la valeur des descripteurs à partir de trois sons de notre base de données : un La4, joué dans les trois modes de jeu. Les trois notes sont jouées dans la dynamique *ff* par un seul et même joueur.

La fenêtre d’analyse des signaux sonores étant translatée avec un pas de 10 ms, on dispose d’environ 60 observations de chaque mode de jeu, pour lesquelles les descripteurs de GDclass sont calculés. Chacun des trois sons peut donc être représenté sous la forme matricielle définie dans la section 4.2.3.

Le résultat de la sélection de descripteurs est donné dans le tableau 4.3.1. Le classement entre les descripteurs est établi en fonction de la valeur du ratio r défini dans la section 4.1.3. Plus celui-ci est grand, plus le descripteur considéré discrimine les classes entre elles.

| Classement | Descripteur | r |
|------------|-------------------------------|------|
| 1 | Spectral Slope | 1535 |
| 2 | Spectral Centroid | 1456 |
| 3 | AudioFilterbank Centroid | 808 |
| 4 | AudioFilterbank Slope | 482 |
| 5 | AudioFilterbank Tristimulus 3 | 335 |
| ⋮ | ⋮ | ⋮ |
| 14 | AudioFilterbank Decrease | 15 |

TAB. 4.1 – Descripteurs sélectionnés pour la distinction entre les modes de jeu. Chaque mode de jeu est représenté par les descripteurs extraits au cours d’une note (La4), jouée dans la dynamique *ff* par un même joueur.

En étudiant la liste des descripteurs sélectionnés, il apparaît que les descripteurs de poids le plus fort semblent décrire les mêmes propriétés du spectre. Le *Spectral Centroid* est une mesure du centre de gravité du spectre. Le descripteur *AudioFilterbank Centroid* est également une mesure du centre de gravité du spectre, pris cette fois dans une échelle de Bark. En projetant orthogonalement les descripteurs sélectionnés entre eux, on constate alors qu’une large partie des

descripteurs sélectionnés sont colinéaires entre eux.

Le tableau 4.3.1 donne cette fois la liste des descripteurs sélectionnés non-colinéaires entre eux.

| | Descripteur | Inertia Ratio |
|----|-------------------------------|---------------|
| 1 | Spectral Slope | 1535 |
| 5 | AudioFilterbank Tristimulus 3 | 300 |
| 14 | AudioFilterbank Decrease | 30 |

TAB. 4.2 – Descripteurs sélectionnés et non colinéaires entre eux :

Les deux premiers descripteurs sélectionnés sont le *Spectral Slope* et le *AudioFilterbank Tristimulus 3*. Le premier est une mesure du taux de décroissance du spectre et le second un calcul du ratio entre l'énergie des partiels haute fréquence sur l'énergie de l'ensemble des partiels (cf section 4.1.2). Ces deux descripteurs décrivent essentiellement le timbre du son. Le poids du troisième descripteur sélectionné étant nettement inférieur à celui des deux premiers, nous ne l'avons pas considéré par la suite.

Espace de descripteurs... On définit un espace pour la représentation des modes de jeu, dont les axes sont le *Spectral Slope* et le *Tristimulus 3*. Sur la figure 4.3, les classes sont représentées dans cet espace. Chaque point représente la valeur du *Spectral Slope* et du *Tristimulus 3* au cours des observations de chaque mode de jeu : détaché en bleu, martelé en rouge et spiccato en noir.

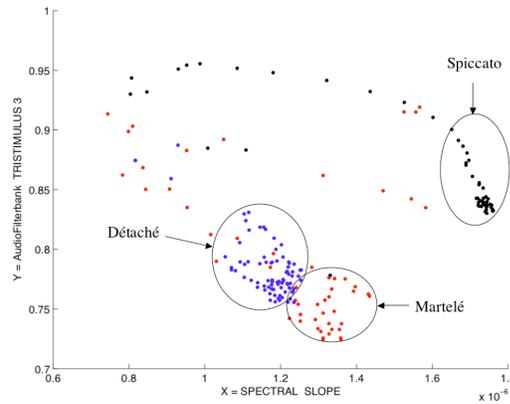


FIG. 4.3 – Représentation des modes de jeu dans un espace 2D(X = Spectral Slope, Y= Tristimulus 3). Détaché en bleu, martelé en rouge, et spiccato en noir.

On peut presque identifier trois clusters distincts, correspondant chacun aux trois modes de jeu. Quelques points sont néanmoins éparpillés dans l'espace de façon incohérente. Ces points correspondent aux valeurs des descripteurs pour les observations de la partie transitoire des sons. Lorsqu'il dispose uniquement des observations des parties transitoires de chacune des notes, l'algorithme de sélection de descripteur est impuissant. En effet, nous sommes incapables de déterminer un seul descripteur qui puisse séparer les trois modes de jeu. Ce résultat n'est pas tellement surprenant. En effet, la plupart des descripteurs calculés dans GDclass sont extraits à partir d'une transformée de Fourier du signal. Or la brièveté des attaques rend difficile une analyse de Fourier (*cf* chapitre suivant). Pour bien décrire les transitoires, les descripteurs définis dans GDclass sont donc insuffisants.

Dans la suite de ce chapitre, on ne considérera plus que les observations correspondant à la partie entretenue du son.

On sélectionne maintenant les descripteurs sur la base des observations de la partie entretenue des sons. Le résultat est donné dans le tableau suivant ;

| Descripteur | Inertia Ratio |
|-------------------------------|---------------|
| Spectral Slope | 2520 |
| AudioFilterbank Tristimulus 3 | 636 |

Nos deux descripteurs *Spectral slope* et *Tristimulus 3* sont toujours les premiers sélectionnés, cette fois avec un poids beaucoup plus fort que lorsque l'ensemble des observations est considéré.

Sur la figure 4.4, ces observations sont représentées, toujours dans le même espace défini par le *Spectral Slope* et le *Tristimulus 3*.

Les clusters correspondant à chaque mode de jeu sont cette fois beaucoup mieux distincts

Il est intéressant de noter que les clusters correspondant aux modes de jeu détaché et martelé sont proches, alors que le mode de jeu spiccato est très éloigné. Lorsqu'une note est jouée en spiccato, l'archet ne reste pas en contact avec la corde pendant toute la note, alors que c'est le cas pour les notes martelées et les notes jouées en détaché.

4.3.2 Variation de dynamique

On étudie maintenant les propriétés de l'espace défini par les descripteurs *Spectral Slope* et *Tristimulus 3*, en considérant un changement de dynamique. Pour cela, chaque mode de jeu est maintenant représenté par deux notes de même hauteur, jouées dans les dynamiques *ff* et *pp* par un seul et même joueur.

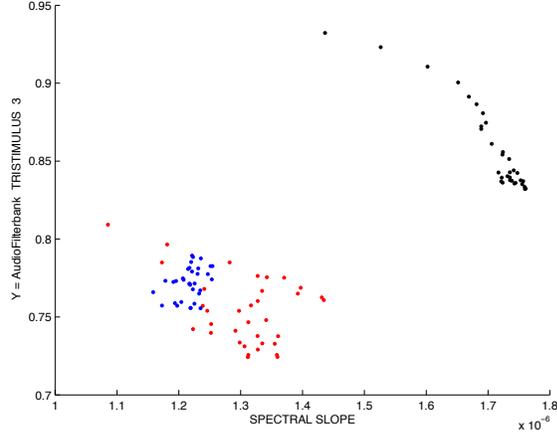


FIG. 4.4 – Représentation des modes de jeu dans un espace 2D ($X = \text{Spectral Slope}$, $Y = \text{Tristimulus 3}$). Les points correspondent aux observations de la partie entretenue des notes prises dans les trois modes de jeu : Détaché en bleu, martelé en rouge, et spiccato en noir.

Les descripteurs sont évalués au cours de la partie entretenue des sons correspondant à ces trois notes.

Chaque observation est représentée sur la figure 4.5 dans notre espace de descripteurs. Les notes jouées *pianissimo* sont représentées par le symbole (+) et les notes jouées *fortissimo* par le symbole (·).

Les clusters correspondants aux modes de jeu détaché et martelé sont invariants, tandis que les notes jouées *pp* et *ff* en Spiccato forment des clusters distincts. Néanmoins on conserve des clusters distincts pour chaque mode de jeu.

4.3.3 Variations de la hauteur de la note et du joueur

Nous avons défini un espace de descripteurs (*Slope et Tristimulus 3*) dans lequel nous sommes capables de reconnaître les modes de jeu dans lesquels sont jouées (par un seul joueur) des notes de même hauteur. On veut maintenant savoir si notre espace conserve cette propriété lorsque les notes évaluées sont jouées à des hauteurs différentes et par des instrumentistes différents.

Chaque mode de jeu est maintenant représenté par les observations des parties entretenues de deux notes de hauteurs différentes (un sol4 et un la4). La figure 4.6 montre la valeur des descripteurs au cours de ces observations dans notre espace de descripteurs.

| <i>Descripteur</i> | poids dans la combinaison |
|--------------------|----------------------------------|
| Slope | 0.46 |
| Centroïd | 0.18 |
| Skewness | 0.17 |
| Tristimulus 3 | 0.17 |
| Autocorrelation | 0.15 |

TAB. 4.3 – Combinaison linéaire de descripteurs qui maximise la séparation entre les modes de jeu (déterminée avec une LDA sur l’ensemble des descripteurs). Le poids affecté au reste des descripteurs est quasi nul.

Si on considère les deux notes séparément, on distingue les clusters de chaque mode de jeu. Par contre, il n’apparaît aucun cluster qui regroupe les modes de jeu sur l’ensemble des observations. Notre espace de description des modes de jeu n’est donc pas invariant selon la note et les clusters associés à chaque mode de jeu se déplacent dans l’espace en fonction de la hauteur de la note. On tente donc de déterminer un ensemble de descripteurs qui puisse caractériser les modes de jeu sur des notes de hauteurs différentes. En utilisant uniquement le critère de sélection *Inertia Ratio Maximization*, cela n’est pas possible. En effet, aucun descripteur n’est capable de discriminer à lui seul les modes de jeu évalués sur des notes de hauteur différente.

On peut alors utiliser une Analyse Linéaire Discriminante (LDA), qui évalue les combinaisons linéaires de descripteurs qui maximisent la distance entre les classes [9]. Nous avons procédé à cette analyse sur un ensemble d’observations contenant une gamme de notes jouées dans les trois modes de jeu. La combinaison linéaire de descripteurs qui maximise la séparation entre les modes de jeu est donnée dans le tableau 4.3

Finalement, cette combinaison de descripteurs nous permet de créer un nouveau descripteur. En représentant les trois modes de jeu selon ce descripteur, on peut effectivement distinguer relativement bien les modes de jeu pour des notes de hauteurs différentes, jouées par un même joueur. Néanmoins, ce descripteur ne permet pas de représenter tous les joueurs dans un même espace.

L’inconvénient avec ce type d’analyse est que le descripteur déterminé par LDA n’a pas nécessairement de rapport immédiat avec le geste.

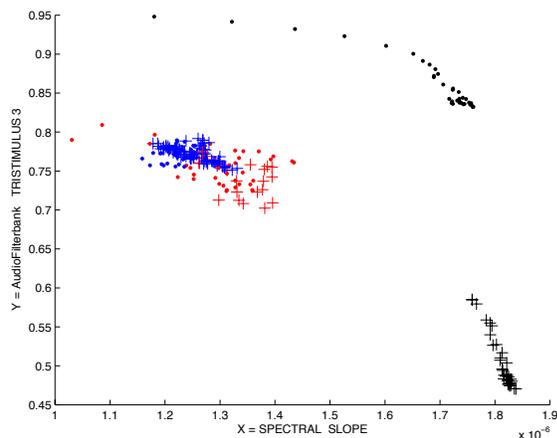


FIG. 4.5 – Variation de la dynamique. Symbole (+) pour les notes jouées *pp*, symbole (•) pour les notes jouées *ff*.

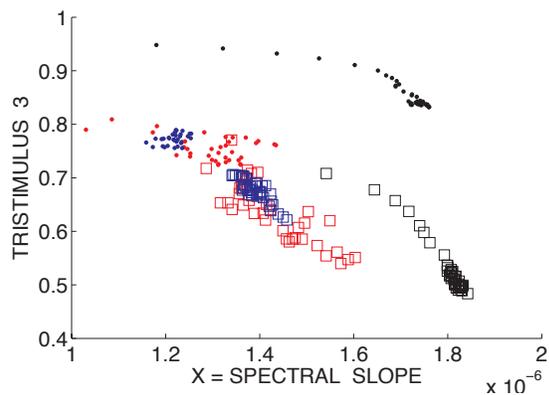


FIG. 4.6 – Deux notes jouées dans les trois modes de jeu sont représentées dans l'espace de edscripteurs (*Slope et Tristimulus 3*) (détaché en bleu, martelé en rouge, spiccato en noir). (\square : Sol4, \bullet : La4)

4.4 Conclusion

La description des modes de jeu avec les outils de classification audio nous a permis de déterminer un espace de descripteurs (*Slope et Tristimulus β*) dans lequel nous sommes capables de reconnaître les modes de jeu dans lesquels sont jouées des notes de mêmes hauteurs, dans deux dynamiques (*pp* et *ff*) et jouées par un seul et même joueur.

Malheureusement, la propriété n'est pas conservée lorsque des notes de hauteurs différentes sont évaluées, et encore moins quand des joueurs différents jouent les notes.

Nous avons étudié la possibilité d'utiliser une Analyse Discriminante Linéaire sur l'ensemble des descripteurs afin de déterminer des combinaisons de descripteurs qui permettraient de distinguer les classes indépendamment de la note jouée. Nous apportons néanmoins deux critiques à cette solution. Premièrement le descripteur créé à partir d'une combinaison de l'ensemble des descripteurs n'aura pas forcément de rapport direct avec le geste. Interpréter le comportement des modes de jeu représentés selon ce descripteur est donc difficile, surtout si on veut décrire dans le son un phénomène physique (comme le frottement de la corde avec l'archet). Deuxièmement, la sélection de descripteurs avec LDA est très sévère, et maximise coûte que coûte la séparation entre les classes. Cette maximisation shunte toute continuité entre les modes de jeu dans une description. Or, l'analyse gestuelle permet de caractériser ce lien entre les modes de jeu et il est donc important de le conserver dans notre description sonore.

Nous avons également soulevé l'insuffisance des descripteurs de GDclass pour la modélisation des parties transitoires du son. Or d'après les études acoustiques, c'est dans ces parties transitoires que la relation entre geste et son est très forte.

Ces éléments nous amènent donc dans les deux chapitres suivants à créer nous-mêmes un descripteur des parties transitoires.

Chapitre 5

Caractérisation des transitoires du violon

Nous avons soulevé dans le chapitre précédent différents problèmes liés à la description des modes de jeu avec les outils de classification. Parmi eux, le fait que les descripteurs audio classiquement utilisés ne permettent pas de caractériser les transitoires d'attaque. Or, il a été montré dans le chapitre 3 qu'une étroite relation entre le geste et le son se met en place au moment des transitoires. Nous nous penchons donc maintenant sur la création d'un nouveau descripteur qui puisse modéliser un transitoire d'attaque et l'articulation entre deux notes.

L'étude que l'on mène ici part d'un constat : les transitoires d'attaques des notes de violon sont fortement bruités. De plus, ce "bruit" semble être simultané aux pics d'accélération de l'archet (*cf* fig. 5.3).

Dans ce chapitre, nous présentons une technique efficace pour l'extraction de la partie bruitée d'un son de violon. On étudiera dans le chapitre suivant le lien entre ce bruit et le geste de l'archet pendant les transitoires.

5.1 Retour

Pendant le transitoire d'attaque, la corde est mise en vibration. Naturellement elle ne passe pas instantanément de l'état de repos à un mouvement de Helmholtz. La corde passe par une phase dite de "Pré-Helmholtz", où son mouvement est un mélange de vibrations périodiques et apériodiques (*cf* chapitre 3). C'est ce caractère apériodique des vibrations de la corde particulièrement fort lors du transitoire d'attaque qui explique le grand niveau de bruit que l'on observe.

Nous partons de l'hypothèse que le joueur est capable de maîtriser avec sa technique d'archet le niveau de bruit généré par la mise en vibration de la corde. On se propose d'étudier ce contrôle via les mesures du geste dont on dispose.

Le son du violon peut être considéré comme la somme d'un signal harmonique (ou pseudo-harmonique) et d'un bruit. Les outils de traitement de signal ne nous permettent pas de modéliser directement le bruit, qui est non sinusoïdal. On modélise donc la partie harmonique du son, et on soustrait ensuite cette modélisation au signal d'origine. On obtient ainsi un signal résiduel qui ne contient en théorie que la composante bruitée du son.

Il est difficile de modéliser exactement la partie harmonique d'un son, et on ne peut pas affirmer que le signal résiduel ne contient que le bruit additif. Par la suite, on préférera donc parler de résiduel plutôt que de bruit.

5.2 Extraction du résiduel

Pour la consistance de notre étude, il est important que le signal résiduel ne contienne plus de partiels. Pour cela, il faut synthétiser correctement la partie pseudo-harmonique du signal. Dans cette section nous présentons deux techniques de synthèse : la synthèse additive, qui utilise la transformée de Fourier, et les méthodes à haute résolution.

5.2.1 Extraction avec additive

Dans un premier temps nous avons synthétisé la partie harmonique du signal avec le logiciel *Additive* de l'IRCAM.

Additive est basé sur un modèle paramétrique du signal, représenté comme une somme de sinusoides perturbées par un bruit additif. Notre signal peut donc s'écrire pour tout $n \in \mathbb{Z}$:

$$x(n) = \sum_{k=1}^K a_k \cos(j2\pi f_k n + \phi_k) + b(n) \quad (5.1)$$

avec $K \in \mathbb{N}$ le nombre de partiels.

Pour suivre l'évolution temporelle des partiels, le signal observé est découpé en trames. Pour chaque trame, les paramètres a_k , f_k et ϕ_k sont estimés dans le spectre de Fourier du signal. La partie harmonique peut ainsi être resynthétisée avec une technique *Overlap and Add*.

La figure 5.1 montre le résiduel extrait de notes jouées en détaché. Nous avons utilisé des fenêtres de 30ms translatées avec un pas de 10ms. La fondamentale est cherchée dans l'intervalle [0Hz 150Hz], et les partiels dans l'intervalle [150Hz 10000Hz].

On voit avec cette confrontation entre les deux signaux l'intérêt d'extraire le

résiduel pour l'étude des parties transitoires. En effet, l'énergie du résiduel semble être maximale à chaque transition. On peut donc penser que nous avons là un bon outil pour l'étude de la mise en vibration de la corde.

Sur la figure 5.3, l'accélération et le résiduel sont représentés simultanément. Dans tous les modes de jeu, l'accélération et le résiduel sont quasi-simultanément maximums lors des transitions (attaque, transition entre deux notes, et extinction de la note). C'est pourquoi nous proposons dans ce chapitre d'étudier de la relation entre le signal résiduel et le signal d'accélération.

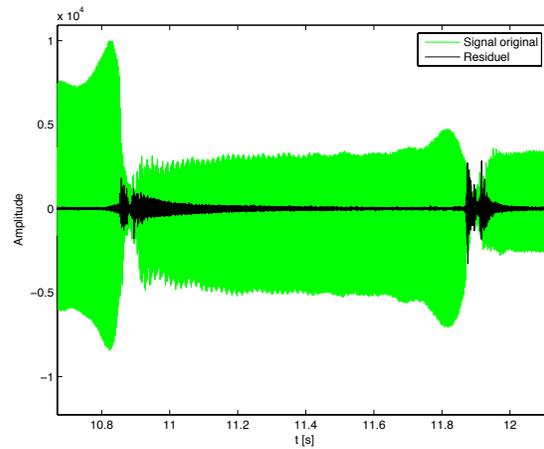


FIG. 5.1 – Résiduel et signal original, mode de jeu : Détaché

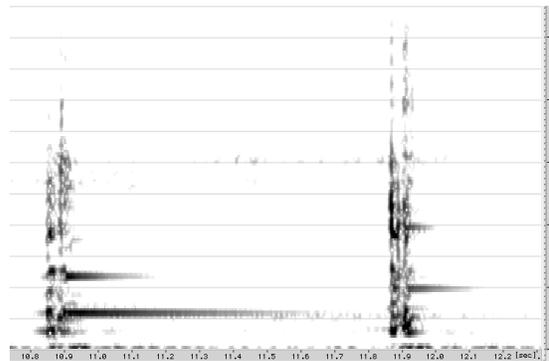


FIG. 5.2 – Spectrogramme du résiduel pour un mode de jeu Détaché

On voit sur la représentation temps-fréquence du résiduel (fig. 5.2) que cer-

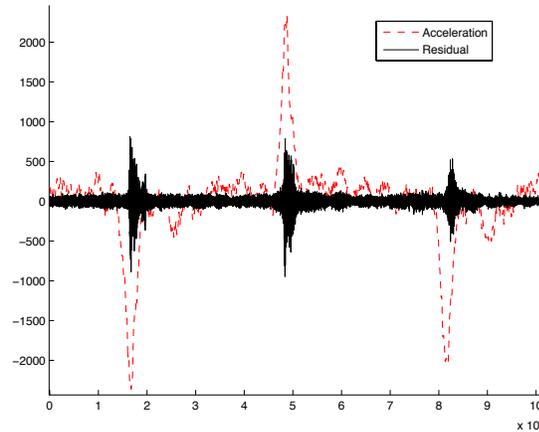


FIG. 5.3 – Représentation simultanée du résiduel et du signal d'accélération

tains partiels n'ont pas été retirés. Il y a plusieurs explications à cela.

Premièrement, le violon est un instrument inharmonique. Or, pour tracker les partiels, le logiciel utilise un a priori harmonique. En effet, une fois qu'il a détecté la fréquence fondamentale, des zones fréquentielles sont déterminées autour des fréquences harmoniques de la fondamentale, dans lesquelles les maxima d'amplitude du spectre sont recherchés. Avec un son légèrement inharmonique, on peut donc détecter convenablement les partiels. Cependant, dans notre cas, il faudrait introduire le modèle d'inharmonicité du violon.

Deuxièmement, on touche ici une limite de la transformée de Fourier. En effet, la brièveté des attaques (en moyenne 50ms pour le violon) nous incite à utiliser des fenêtres de taille très petite pour bien décrire l'évolution temporelle des partiels. Parallèlement, on a besoin d'une bonne résolution spectrale pour distinguer les partiels les uns des autres. Malheureusement il est difficile d'obtenir les deux. En effet, la taille de fenêtre conditionne la résolution fréquentielle de la transformée de Fourier (la capacité de distinguer deux proches). Pour une fenêtre de N points, la précision fréquentielle est en $\frac{2}{N}$, donc plus la fenêtre est grande, mieux on est capable de discriminer les fréquences. Les fenêtres que l'on a utilisées sont de 30ms, soit $N=1323$ points pour un son échantillonné à 44,1kHz. La résolution fréquentielle est donc $\Delta f = \frac{2}{N} F_S = 66.6\text{Hz}$. Pendant le régime entretenu, les partiels sont bien distincts et évoluent de façon quasi-stationnaire, il n'y a donc pas de problème pour les distinguer avec la transformée de Fourier. Dans le cas des transitoires d'attaques, l'évolution des partiels n'est plus du tout stationnaire, et leurs trajectoires peuvent se croiser. On a donc besoin pour les suivre à la fois de discriminer fortement les fréquences et d'avoir des fenêtres temporelles très petites. La résolution fréquentielle obtenue avec des fenêtres de

30ms est déjà insuffisante, et on ne peut donc plus la réduire. La transformée de Fourier n'est donc pas un outil approprié à la modélisation harmonique des transitoires d'attaque du violon.

5.2.2 Les méthodes à haute résolution

Nous avons soulevé les problèmes de résolution spectrale liés au fenêtrage lorsqu'on utilise la transformée de Fourier pour analyser notre signal. Les méthodes à haute résolution permettent en théorie de s'affranchir de cette limite, ou en tout cas dans la pratique de réduire considérablement la taille de la fenêtre d'observation, sans affecter l'estimation des fréquences.

Ces méthodes utilisent un modèle paramétrique du signal représenté comme une somme de sinusoides exponentiellement amorties et perturbées par un bruit blanc additif gaussien complexe. Pour l'estimation des paramètres, les méthodes à haute résolution développées récemment à l'ENST sont basées sur une décomposition en sous-espaces propres du signal observé. Nous allons maintenant introduire quelques éléments de théorie sur cette décomposition et présenter ensuite le système d'analyse/synthèse basé sur la *méthode ESPRIT* que nous avons utilisée pour extraire le résiduel.

Modèle de signal

Pour tout $t \in \mathbb{Z}$, le modèle $s(t)$ est tel que :

$$s(t) = \sum_{k=0}^{K-1} \alpha_k z_k^t \quad (5.2)$$

Avec $K \in \mathbb{N}^*$, $k \in \{0 \dots K-1\}$. $\alpha_k \in \mathbb{C}^*$ et $z_k \in \mathbb{C}^*$.

Les pôles complexes z_k s'écrivent :

$$z_k = e^{\delta_k} e^{i2\pi f_k} \quad (5.3)$$

où $f_k \in \mathbb{R}$ et $\delta_k \in \mathbb{R}$, sont respectivement la fréquence de la $k_{ième}$ sinusoïde et le taux d'amortissement associé.

Les amplitudes complexes α_k s'écrivent :

$$\alpha_k = a_k e^{i\phi_k} \quad (5.4)$$

où $a_k \in \mathbb{R}_+^*$ et $\phi_k \in \mathbb{R}$, sont respectivement l'amplitude et la phase associées à la fréquence f_k .

Les paramètres que l'on cherche à estimer sont donc a_k , δ_k , f_k et ϕ_k .

Le signal observé $x(t)$ est modélisé comme la somme du signal $s(t)$ et d'un bruit blanc gaussien complexe $b(t)$, centré et de variance σ^2 . On a ainsi :

$$x(t) = s(t) + b(t) \quad (5.5)$$

Structures des données

Considérons l'horizon d'observation $[0 \dots N-1]$ avec $N > 2K$. Le signal non bruité $s(t)$ peut s'écrire sous la forme d'une matrice de Haenkel sur cet horizon :

$$\mathbf{S} \begin{pmatrix} s(0) & s(1) & \dots & s(l-1) \\ s(1) & s(2) & \dots & s(l) \\ \vdots & \vdots & \vdots & \vdots \\ s(n-1) & s(n) & \dots & s(N-1) \end{pmatrix}$$

où $n > K$, $l > K$, et $N = n + l - 1$.

Si $s(t)$ satisfait le modèle défini en (1.1) alors la matrice \mathbf{S} peut être factorisée sous la forme :

$$\mathbf{S} = \mathbf{V}^n \mathbf{D} \mathbf{V}^{l^T} \quad (5.6)$$

où $\mathbf{D} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_{K-1})$, et \mathbf{V}^n et \mathbf{V}^l sont les matrices de Vandermonde de dimensions $n \times K$ et $l \times K$ respectivement.

On rappelle que la matrice de Vandermonde \mathbf{V}^N de dimensions $N \times K$ est définie par :

$$\mathbf{V}^N \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_K \\ \vdots & \vdots & \vdots & \vdots \\ z_0^N & z_1^N & \dots & z_K^N \end{pmatrix}$$

On définit la matrice de corrélation : $R_{ss} = \frac{1}{l} \mathbf{S} \mathbf{S}^H$. Si \mathbf{S} est factorisable selon (1.5), on montre que R_{ss} est également factorisable :

$$R_{ss} = \mathbf{V}^n \mathbf{P} \mathbf{V}^{l^H} \quad (5.7)$$

avec $\mathbf{P} = \frac{1}{l} \mathbf{V}^{l^T} \mathbf{V}^{l^*} \mathbf{D}^H$.

On fait maintenant l'hypothèse suivante : tous les pôles z_k sont distincts et, $\alpha_k \neq 0$ et $z_k \neq 0 \forall k \in \{0 \dots K-1\}$. Dans ce cas, la matrice R_{ss} est de rang K , et son espace image est engendré par la matrice \mathbf{V}^n . L'espace image de R_{ss} est appelé espace signal.

Une diagonalisation de la matrice R_{ss} nous en donne une base orthonormée composée des vecteurs propres $\{w_m\}_{m=0 \dots n-1}$, associés aux valeurs propres $\lambda_0 \geq \lambda_1 \geq \dots \lambda_{n-1} \geq 0$. La matrice R_{ss} étant de rang K , on peut écrire alors :

- $\lambda_i > 0, \forall i \in \{0 \dots K-1\}$
- $\lambda_i = 0, \forall i \in \{K \dots n-1\}$

On appelle alors W la matrice $[w_0 \dots w_{K-1}]$ et W_{\perp} la matrice $[w_K \dots w_{n-1}]$. On peut montrer que $\text{Im}(W) = \text{Im}(V^n)$. W est donc une autre base de l'espace signal.

Dans la pratique, on dispose seulement des observations du signal bruité $x(t) = s(t) + b(t)$. $x(t)$ peut également s'écrire sous la forme d'une matrice de Haenkel X sur l'horizon $[0 \dots N-1]$. Le bruit blanc additif $b(t)$ est centré et de variance σ^2 . A partir des observations de X , on peut estimer la matrice de corrélation \hat{R}_{xx} :

$$\hat{R}_{xx} = \frac{1}{l} X X^H = R_{ss} + \sigma^2 I_n \quad (5.8)$$

D'après la décomposition de R_{ss} , on peut montrer que $\{w_m\}_{m=0 \dots n-1}$ est également une base orthonormée de vecteurs propres de la matrice \hat{R}_{xx} , associés aux valeurs propres μ_m telles que :

- $\mu_m = \lambda_m + \sigma^2, \forall m \in \{0 \dots K-1\}$
- $\mu_m = \sigma^2, \forall m \in \{K \dots n-1\}$

On appelle alors $\text{Im}(W)$ l'espace signal et $\text{Im}(W_{\perp})$ l'espace bruit. On peut donc estimer l'espace signal en diagonalisant \hat{R}_{xx} (en triant les valeurs propres par ordre croissant), et en sélectionnant les K premiers vecteurs propres, dont les valeurs propres associées ont toutes une valeur strictement supérieure à σ^2 .

Il existe plusieurs méthodes à haute résolution basées sur une décomposition en sous-espaces propres. Parmi elles, la méthode ESPRIT que nous allons maintenant présenter.

La méthode ESPRIT

L'estimation des paramètres f_k et δ_k avec la méthode ESPRIT (*Estimation of Signal Parameters via Rotational Invariance Techniques*) est basée sur la propriété d'invariance rotationnelle de la matrice V^n .

Propriété d'invariance rotationnelle : Soit A une matrice de dimensions $n \times K$. On définit A_{\downarrow} la matrice de dimensions $(n-1) \times K$ composée des $n-1$ premières lignes de A , et A_{\uparrow} la matrice de dimensions $(n-1) \times K$ composée des $n-1$ dernières lignes de A . La propriété d'invariance rotationnelle de la matrice de Vandermonde nous donne :

$$V_{\uparrow}^n = V_{\downarrow}^n J \quad (5.9)$$

où $J = \text{diag}(z_0, z_1, \dots, z_{K-1})$

On a vu que la matrice W était également une base de la matrice R_{xx} . On effectue donc le changement de base suivant : $V^n = W C$. Ainsi, la matrice W satisfait elle aussi la propriété d'invariance rotationnelle : $W_{\uparrow}^n = W_{\downarrow}^n \Phi$,

où $\Phi = C D C^{-1}$ est appelée matrice spectrale.

La propriété intéressante de Φ est que ses valeurs propres sont les pôles $\{z_k\}_{k \in \{0 \dots K-1\}}$. Si on peut déterminer ces pôles, alors on connaît les paramètres f_k et δ_k .

On calcule donc la matrice Φ : $\Phi = (W_{\downarrow}^H W_{\downarrow})^{-1} W_{\downarrow}^H W_{\uparrow}$

En diagonalisant Φ , on peut donc estimer les pôles $\{\hat{z}_k\}_{k \in \{0 \dots K-1\}}$, et on a :

- $\hat{f}_k = \frac{1}{2\pi} \text{angle}(\hat{z}_k)$
- $\hat{\delta}_k = \ln(|\hat{z}_k|)$

En résumé :

- On calcule \hat{R}_{xx}
- On diagonalise \hat{R}_{xx} et on en déduit W
- On calcule Φ
- On diagonalise Φ et on en déduit les paramètres f_k et δ_k

Estimation des amplitudes et des phases

On utilise l'estimateur des moindres carrés pour déterminer les amplitudes complexes α_k .

Si on a $x = V^n \alpha + b$, alors, l'estimateur des moindres carrés $\hat{\alpha}$ de α est :

$$\hat{\alpha} = V^{n\dagger} x \quad (5.10)$$

Avec $V^{n\dagger}$ la matrice pseudo-inverse de V^n . Les amplitudes réelles et les phases se calculent ensuite par :

- $\hat{a}_k = |\hat{\alpha}_k|$
- $\hat{\phi}_k = \arg(\hat{\alpha}_k)$

Remarque : si le bruit additif est blanc, alors l'estimateur des moindres carrés est sans biais et atteint la borne de Cramer-Rao.

5.2.3 Application au son du violon

La synthèse d'un son de violon avec les méthodes à haute résolution donne des résultats au moins aussi bons voire meilleurs que ceux obtenus avec une synthèse additive. Modéliser le signal comme une somme de sinusoides exponentiellement amorties paraît donc être adapté au son du violon.

Dans cette partie nous présentons l'extraction du signal résiduel d'un son de violon avec les méthodes à haute résolution. Nous avons utilisé pour cela le système d'analyse/synthèse développé par Roland Badeau à l'ENST. Ce système utilise la méthode ESPRIT pour estimer les pôles complexes du signal. On trouvera une présentation détaillée de ce système dans [1].

Choix de l'ordre de modélisation

L'inconvénient des méthodes à haute résolution est qu'elles supposent connu l'ordre K du modèle de signal, i.e. le nombre K de sinusoides présentes dans le signal. Or dans le cas du violon, nous ne connaissons pas K .

Si le rapport signal à bruit est suffisamment grand, les valeurs propres de la

matrice \widehat{R}_{xx} associées à l'espace signal sont strictement supérieures à σ^2 , la variance du bruit, et on distingue très nettement la limite entre l'espace signal et l'espace bruit. Dans ce cas, on peut déduire l'ordre de modélisation.

Dans le cas du violon, le rapport signal à bruit n'est pas très bon, en particulier au moment des attaques. La limite entre les deux sous-espaces est donc beaucoup plus floue. (cf fig. 5.4)

Le processus d'analyse décompose le signal en sous-bandes dans lesquelles il est sous échantillonné avec un facteur $M=8$. Les pôles sont estimés dans chaque sous bande. Sur la figure 5.4 sont tracées dans l'ordre décroissant les valeurs propres de la matrice \widehat{R}_{xx} pour une fenêtre d'analyse de 64 points prise dans la première sous-bande. Le signal est une note de violon jouée en martelé. La fenêtre est prise pendant l'attaque de la note. On voit que seulement quelques valeurs propres se distinguent. A partir de la cinquième valeur propre, commence une zone de confusion entre les valeurs propres associées à l'espace signal et à l'espace bruit.

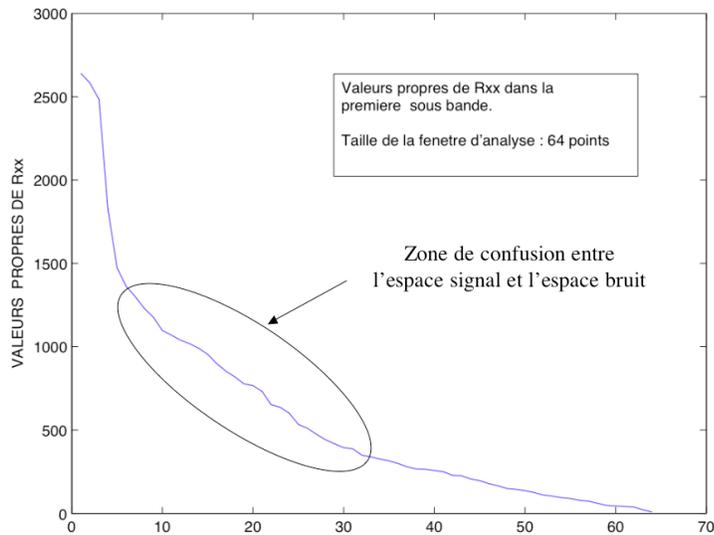


FIG. 5.4 – Valeurs propres de la matrice \widehat{R}_{xx} calculées sur la première fenêtre d'analyse de la première sous bande. Le signal sonore est une note jouée en martelé.

En choisissant un ordre de modélisation trop grand, on introduit dans le signal synthétisé une partie du bruit, et inversement si on sous-modélise le signal. Notre but étant de récupérer exactement le bruit, nous avons pensé dans un premier temps qu'il valait mieux sous-estimer l'ordre de modélisation afin de ne perdre aucune contribution du bruit, quitte à lisser ensuite le résiduel obtenu pour effacer les partiels restants. Malheureusement, une sous-modélisation de

l'espace signal affecte l'estimation des amplitudes. En effet, la variance de l'estimateur des amplitudes $(\hat{\alpha} V^n \alpha)$ vaut : $\text{var}(\hat{\alpha}) \sigma^2 (V^n V^{n*})^{-1}$. Or la matrice V^n est de rang K , donc plus K est grand, plus la variance de l'estimateur est faible.

Pour mesurer les effets d'une surmodélisation, nous avons synthétisé un signal composé de trois sinusoïdes, noyées dans un bruit blanc. Afin de récupérer exactement le bruit additif avec les méthodes à haute résolution, il nous faut modéliser le signal avec un ordre d'au moins 10 sinusoïdes. Pour notre application, nous avons donc surestimé l'ordre du modèle pour ne pas affecter l'estimation des amplitudes, qui est critique dans le cas d'une extraction de résiduel (cf paragraphe suivant). Nous avons également étudié le comportement des valeurs propres de \hat{R}_{xx} en fonction de la hauteur de la note et des modes de jeu. Notre conclusion est qu'il est possible de garder le même ordre de modélisation quels que soient la hauteur et le mode de jeu dans lequel la note est jouée. Dans le cas du violon on peut fixer l'ordre K à 30 sinusoïdes par sous bande.

Artefacts dus à la soustraction

L'extraction d'un résiduel n'est pas la première application visée dans la conception d'un processus de synthèse, et il y a quelques précautions à prendre lors de la soustraction du signal synthétique au signal original.

On considère les deux paramètres du modèle que sont la fréquence et l'amplitude réelle des sinusoïdes. Les méthodes à haute résolution nous ont permis d'améliorer la précision de l'estimation des fréquences. Les besoins de précision en fréquence sont d'ailleurs les mêmes que ce soit pour une application de synthèse ou pour l'extraction d'un résiduel.

Pour l'estimation des amplitudes, il y a un point sur lequel il est important de s'attarder. Notre signal est une somme de sinusoïdes noyées dans un bruit blanc centré. L'amplitude de chaque sinusoïde s'ajoute donc à la variance du bruit additif. Les amplitudes estimées sont donc la somme des amplitudes réelles du signal et de la variance du bruit. On comprend alors en quoi la soustraction du signal synthétique au signal original peut poser un problème. En effet, si l'estimation des amplitudes est exacte, on supprime à la fois les harmoniques et la contribution du bruit.

Pour illustrer notre propos, prenons un signal $s(t)$ composé d'une sinusoïde s_1 de fréquence f_1 et d'amplitude a_1 , noyée dans un bruit blanc $b(t)$, centré et de variance σ^2 . A la fréquence f_1 , le spectre $A(f)$ de notre signal vérifie donc : $A(f_1) = a_1 + \sigma^2$. On veut maintenant retrouver le bruit $b(t)$ à partir du signal $s(t)$. Pour cela on estime la fréquence et l'amplitude de la sinusoïde s_1 pour la resynthétiser. On soustrait ensuite ce signal synthétique au signal $s(t)$.

Il y a alors plusieurs scénarios possibles lors de la soustraction. Dans le premier cas, l'estimateur des amplitudes est centré et de variance nul. On estime donc exactement l'amplitude de la sinusoïde dans le signal. Le spectre du résiduel présente alors un trou à la fréquence f_1 , et on a donc un signal résiduel erroné.

Deuxième cas, l'estimateur des amplitudes est centré et de variance σ_1^2 . Si

σ_1^2 et σ^2 sont du même ordre de grandeur. L'amplitude du spectre du résiduel à la fréquence f_1 est alors comparable à celle du spectre de $b(t)$ à cette même fréquence, et on peut dire que l'estimation du résiduel est satisfaisante. Si maintenant la variance de l'estimateur σ_1^2 est plus élevée que la variance du bruit σ^2 , on obtient alors dans le signal résiduel une sinusoïde à la fréquence f_1 et d'amplitude σ_1^2 .

Synthèse et extraction de résiduel d'un son de violon avec les méthodes HR

On synthétise une note de violon jouée en martelé avec les méthodes à haute résolution. L'ordre de modélisation dans chaque sous-bande est égal à 33. Les figures 5.5 et 5.6 montrent les périodogrammes du signal original, du signal synthétisé avec les méthodes à haute résolution et du résiduel obtenu. Sur la figure 5.5, la fenêtre d'analyse est placée sur la partie entretenue du son. Le résiduel semble quasi-exempt de partiels. Sur la figure 5.6, la fenêtre d'analyse est maintenant prise pendant l'attaque du son. Là encore, le résiduel est propre. Le signal a été relativement bien synthétisé.

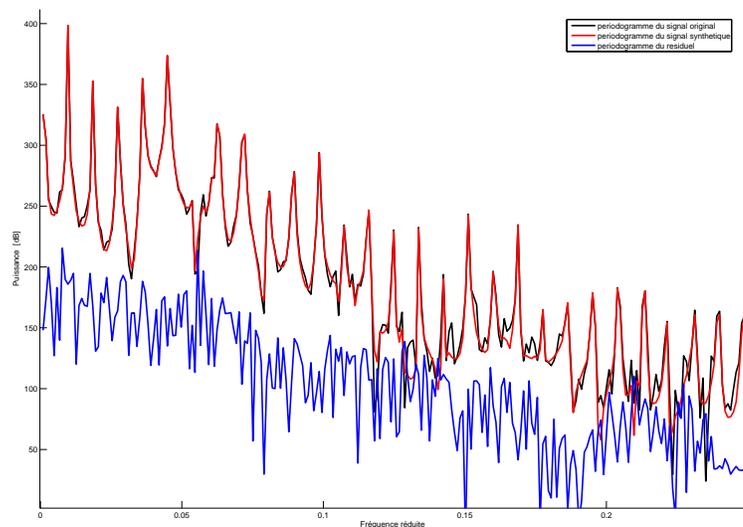


FIG. 5.5 – Périodogrammes du signal original, du signal synthétisé avec les méthodes à haute résolution et du signal résiduel. Le signal original est une note jouée en martelé. La trame d'analyse est prise sur **la partie entretenue du son**.

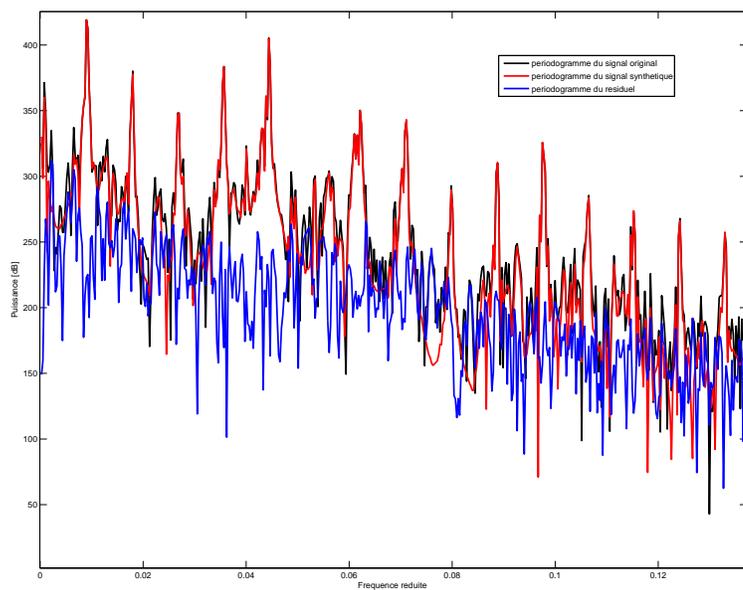


FIG. 5.6 – Périodogrammes du signal original, du signal synthétisé avec les méthodes à haute résolution et du signal résiduel. Le signal original est une note jouée en martelé. La trame d'analyse est prise **pendant l'attaque de la note**.

5.2.4 Confrontation méthode ESPRIT/Synthèse Additive

Afin de valider l'amélioration apportée avec les méthodes HR, nous comparons maintenant les résiduels obtenus avec *Additive* et les méthodes à haute résolution.

On trace une représentation temps-fréquence des signaux résiduels obtenus avec les deux méthodes pour une gamme jouée en *Détaché* et en *Martelé* (fig. 5.7 et 5.8).

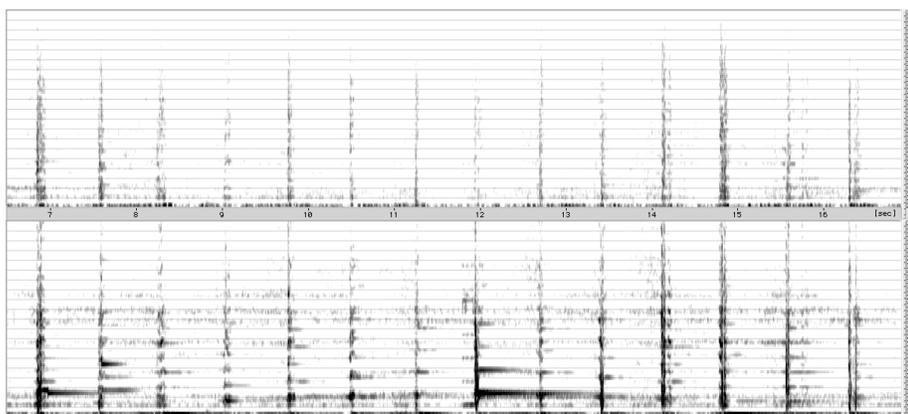


FIG. 5.7 – TFCT du résiduel gamme en *Détaché* (Méthodes HR en haut, Synthèse additive en bas)

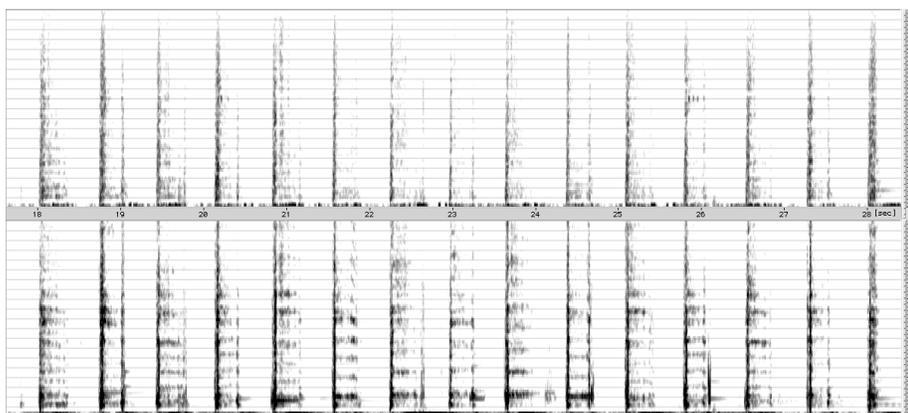


FIG. 5.8 – TFCT du résiduel gamme en *Martelé* (Méthodes HR en haut, Synthèse additive en bas)

La figure 5.9 montre les périodogrammes des signaux résiduels obtenus avec les méthodes à haute résolution et avec additive pendant la partie entretenue du son et pendant l'attaque.

A la fois pendant la partie entretenue et l'attaque, les méthodes à haute résolution nous ont permis d'effacer dans le signal beaucoup plus de partiels que la synthèse Additive ne le fait. On le voit par exemple sur la figure 5.9 (fenêtre du bas), aux fréquences réduites 0.06 et 0.07 par exemple, le périodogramme du résiduel obtenu avec Additive se colle quasiment à celui du signal original.

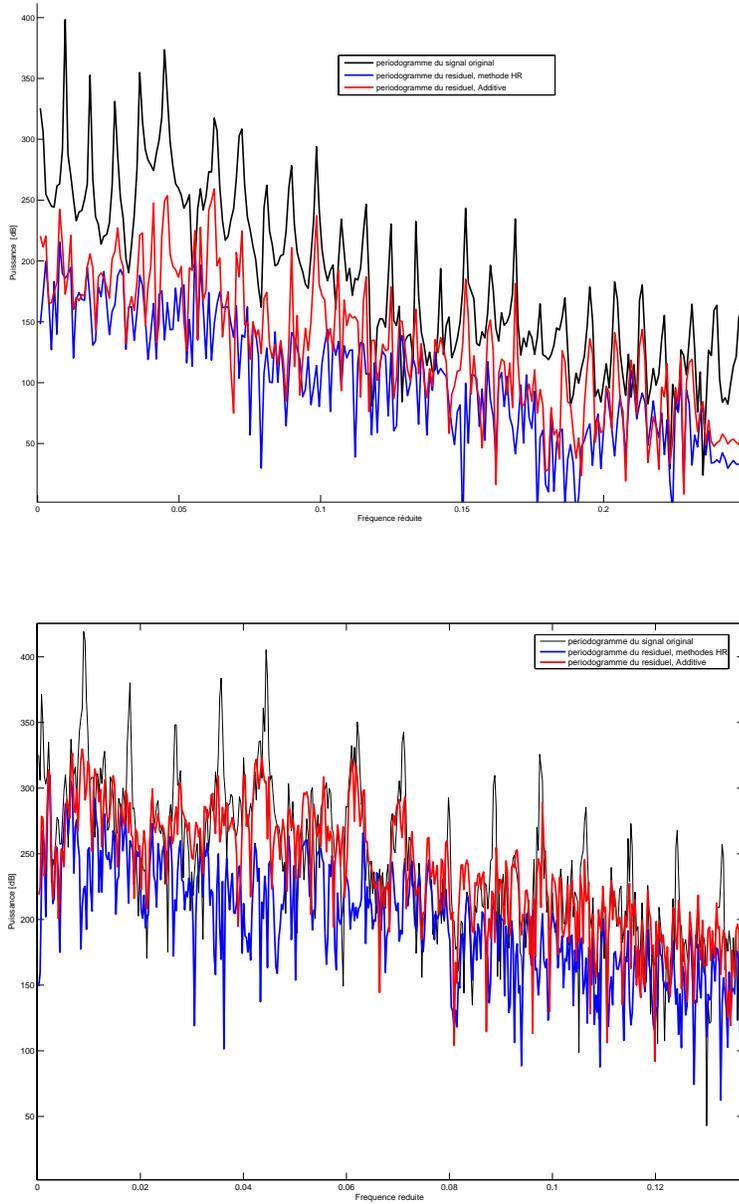


FIG. 5.9 – Périodogrammes du signal résiduel obtenu avec les méthodes HR (en bleu) et avec Additive (en rouge). Le périodogramme du signal original est représenté en noir. La première figure correspond à une fenêtre d'analyse prise sur la partie entretenue du son. Sur la figure du dessous, la fenêtre d'analyse est prise pendant l'attaque.

Chapitre 6

Etude de la relation entre le résiduel et l'accélération

Nous avons montré dans la partie précédente comment extraire proprement la partie non-harmonique d'un son de violon avec les méthodes à haute résolution. Dans ce chapitre nous confrontons le signal résiduel au signal d'accélération.

Nous montrerons qu'une description croisée geste-son des modes de jeu permet sur quelques dizaines de ms le mode de jeu dans lequel est attaqué une note.

6.1 Mesures et protocole

6.1.1 Base d'étude

Nous disposons d'enregistrements de gammes, jouées dans les trois modes de jeu, dans différentes nuances et par différents instrumentistes. Pour chacun de ces enregistrements, les signaux résiduels sont extraits avec les méthodes à haute résolution (cf chapitre 5).

Notre étude portant sur la caractérisation des transitoires, nous avons isolé les transitions entre notes sur chacune des gammes. On pourra ainsi étudier le comportement des signaux résiduels et d'accélération pour diverses configurations de jeu et divers joueurs.

6.1.2 Prétraitements sur les signaux

Extraction de la puissance du résiduel Dans ce rapport nous avons confronté uniquement la puissance du résiduel au signal d'accélération. Ceci constitue une première étape dans la confrontation entre bruit et accélération et ne remet pas en cause le fait d'avoir utilisé une méthode aussi sophistiquée que les méthodes à haute résolution pour l'extraction du résiduel.

L'enveloppe du résiduel est extraite en filtrant le signal avec une demi fenêtre de hanning de 1024 points (env. 23ms).

Seuillage du signal d'accélération Nous considérons le signal d'accélération mesuré dans la direction de l'archet. L'angle d'inclinaison de l'archet par rapport à la corde crée un offset dans le signal d'accélération. L'offset étant variable au cours du temps (le joueur est malheureusement physiquement incapable de maintenir le même angle), il ne suffit pas de centrer le signal pour l'annuler. Pour parer à cela, nous appliquons un seuil au signal d'accélération. Ce seuil est fixé à 10 % du maximum d'amplitude du signal. Pour être comparé au résiduel, le signal d'accélération est suréchantillonné d'un facteur $M = \frac{F_s}{F_{acc}}$, avec $F_{acc} = 333\text{Hz}$, la fréquence d'échantillonnage de l'accélération, et $F_s = 44,1\text{kHz}$ la fréquence d'échantillonnage du son.

6.1.3 Modes de jeu considérés

Pour amorcer l'étude croisée entre l'accélération et la puissance du résiduel, seuls les modes de jeu détaché et martelé sont considérés. Le mode de jeu spiccato a été temporairement écarté pour une simple raison : lorsqu'une note est jouée en spiccato, l'archet est très peu en contact avec la corde, et le signal d'accélération de l'archet est difficilement interprétable dans une étude croisée geste-son.

Restriction sur le signal d'accélération martelé La figure 6.1 représente le résiduel et le signal d'accélération d'une note martelée. On observe deux pics de puissance du signal résiduel, au début et à la fin de la note. Le second pic de bruit est généré lorsque l'instrumentiste stoppe l'archet. Pour l'instant, l'étude du transitoire d'extinction a été écartée. Nous considérons ici uniquement le transitoire d'attaque. Nous avons donc centré l'étude de la relation entre résiduel et accélération sur le premier lobe du signal d'accélération.

6.2 Résultats

Nous avons constaté dans le chapitre 5 que le signal résiduel et le signal d'accélération présentaient tous deux un pic à chaque transitoire. Nous étudions maintenant de manière plus approfondie la relation qui peut exister entre ces deux signaux.

6.2.1 Relation temporelle

La figure 6.2 donne une représentation simultanée de la puissance du résiduel et du signal d'accélération pour une transition dans les deux modes de jeu détaché et martelé.

Nous avons étudié ce genre de graphique pour une large partie de notre base

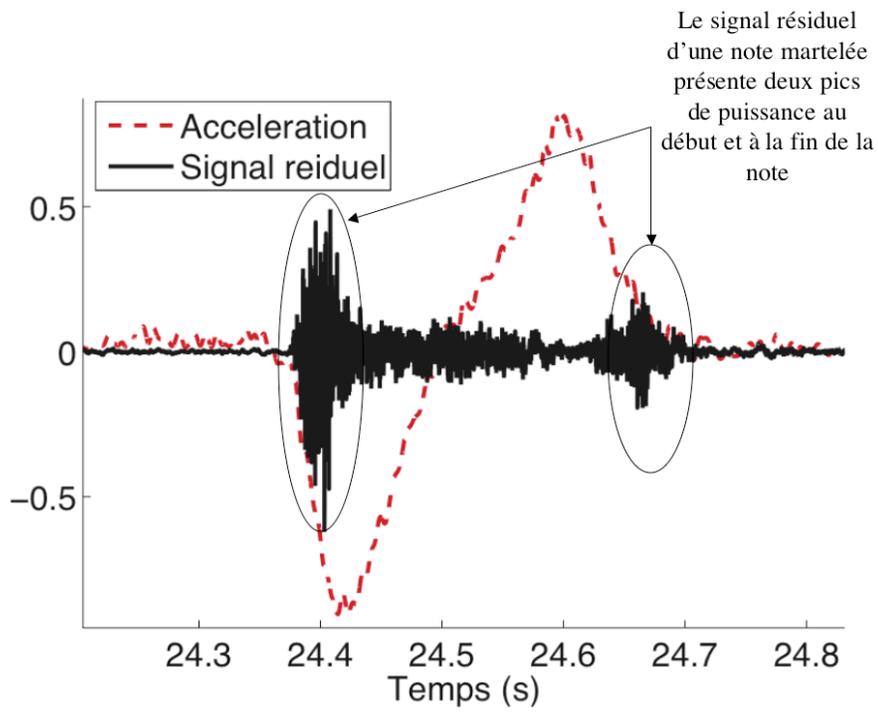


FIG. 6.1 – default

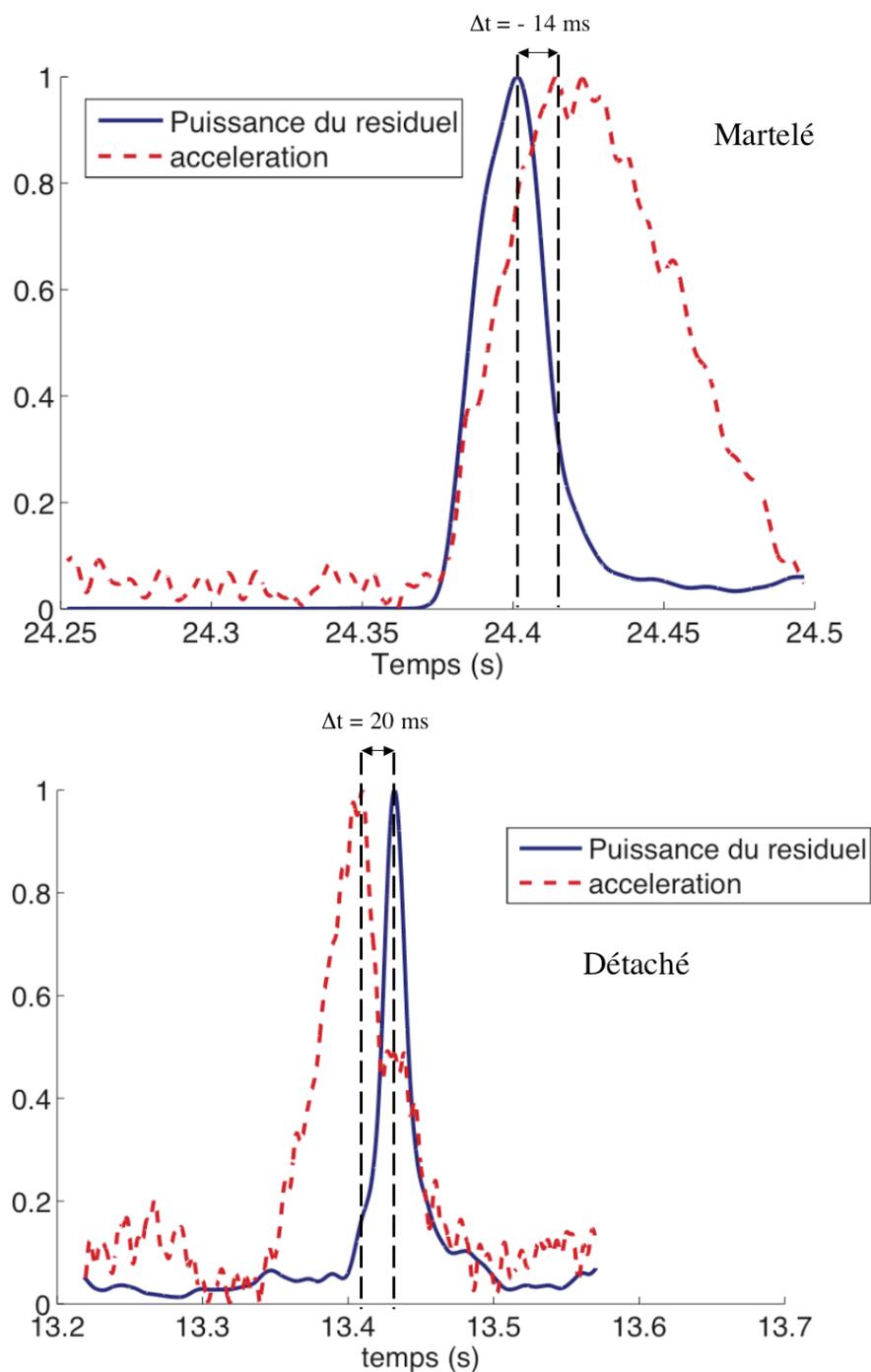


FIG. 6.2 – Comparaison du signal d'accélération à la puissance du résiduel dans deux modes de jeu : en haut Martelé (signaux extraits pendant l'attaque), en bas Détaché (signaux extraits pendant la transition entre deux notes). Les signaux sont normalisés.

de son. Il apparaît de façon très claire que dans la majorité des cas étudiés, le pic de puissance du résiduel précède le pic d'accélération quand la transition est effectuée dans un mode martelé, et inversement quand la transition est effectuée en détaché. Ceci, apparemment quels que soient la hauteur de la note, la dynamique, et le joueur.

On appelle ΔT , le temps de décalage entre les deux pics. Nous étudions maintenant la pertinence de la mesure Δt pour caractériser les transitoires de chaque mode de jeu.

On considère dans un premier temps une gamme jouée par un seul joueur (le professeur de la classe de violon) dans les deux modes de jeu (avec la même nuance). La mesure de ΔT pour les transitions correspondant à ces deux gammes est représentée sur la figure 6.3. L'axe Y donne la valeur de ΔT . Chacun des points représente un transitoire (\bullet : Martelé ; X : détaché). Sur l'axe X, on indique simplement la position de la note dans la gamme.

Ce premier résultat est très encourageant. En effet, pour chaque transitoire, on peut parfaitement identifier dans quel mode de jeu la note va être jouée. Il est intéressant d'étudier maintenant comment réagit ΔT à une variation de nuance.

Sur la figure 6.4, chaque point représente la valeur de ΔT pour une transition dans un mode de jeu (\bullet : Martelé ; X : Détaché). Les mesures sont extraites des enregistrements d'une gamme dans les quatre configurations suivantes : détaché joué *pp* (en rouge), détaché joué *mf* (en noir), martelé joué *pp* (en rouge), martelé joué *mf* (en noir). Chacune des quatre gammes est jouée par un seul instrumentiste (le professeur de la classe de violon de Dijon).

Mis à part une légère confusion pour les notes 3 et 4, on distingue encore une fois parfaitement la limite entre les deux modes de jeu. Il apparaît que ΔT augmente quand on passe de la nuance *mf* à la nuance *pp*. Néanmoins, dans la mesure où on ne dispose pas de l'information de pression d'archet, il est difficile de tirer une conclusion de ce graphique sur l'influence de la nuance sur le décalage entre les deux pics. Il faut pour cela considérer la puissance sonore.

On étudie enfin la valeur de ΔT pour des transitoires toujours effectués dans les deux modes de jeu détaché et martelé, mais cette fois par des joueurs de niveaux différents et dans trois nuances. Cette évaluation est représentée sur la figure 6.5 Les niveaux des quatre joueurs sont très variables. Les élèves 1 et 2 sont en première année de DEM (Diplôme d'Etude Musicale). Ils ne sont pas débutants, mais n'ont pas encore une technique de jeu totalement éprouvée. L'élève 7 est d'un niveau confirmé. Le quatrième joueur considéré est la professeur. Il est intéressant de constater que la distinction entre les modes de jeu est plus ou moins nette selon le niveau des joueurs. Pour les élèves 1 et 2, on observe parfois des inversions (en $X=5$, 6 et 7), pour les transitoires martelés.

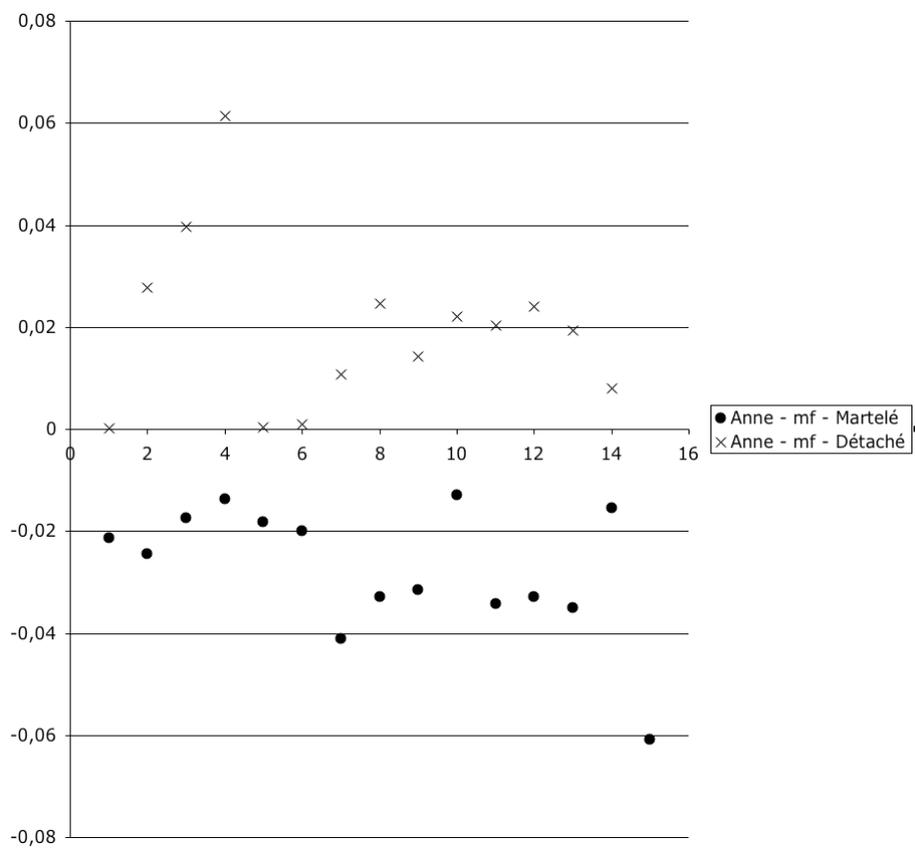


FIG. 6.3 – Mesure de ΔT sur une gamme. ● : Martelé; X : Détaché. Chaque transitoire est effectué dans la nuance *mf* par un seul de joueur de niveau confirmé

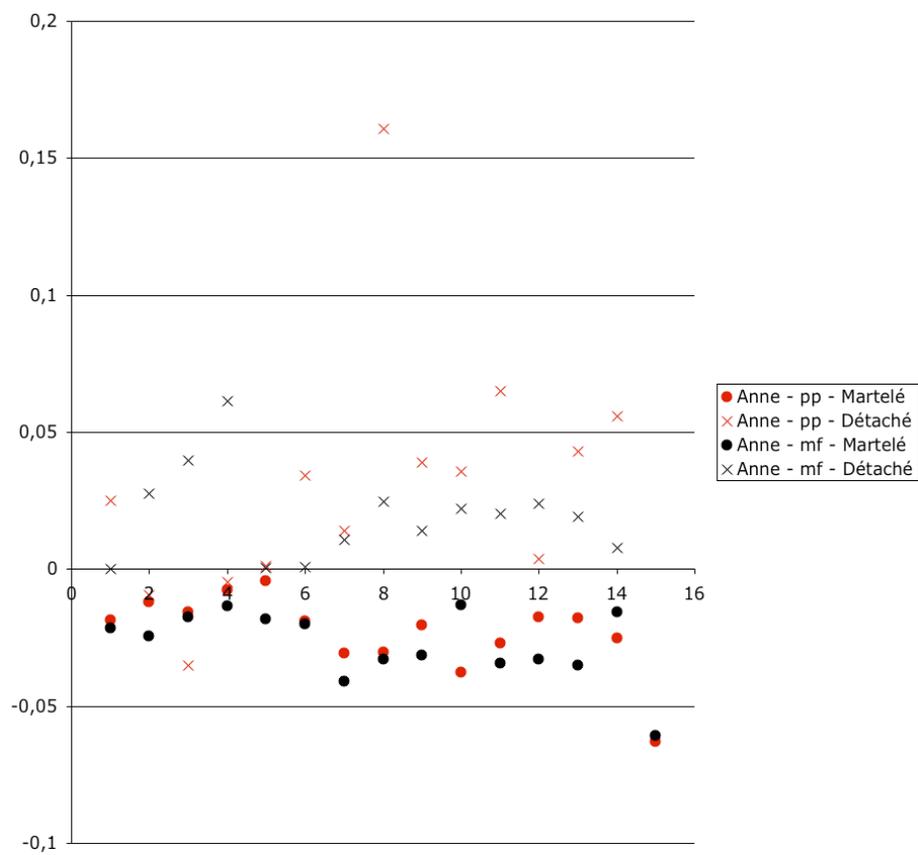


FIG. 6.4 – Mesure de ΔT sur une gamme. \bullet : Martelé; X : détaché; mf en noire et pp en rouge. La position de la note dans la gamme est donnée par l'axe X.

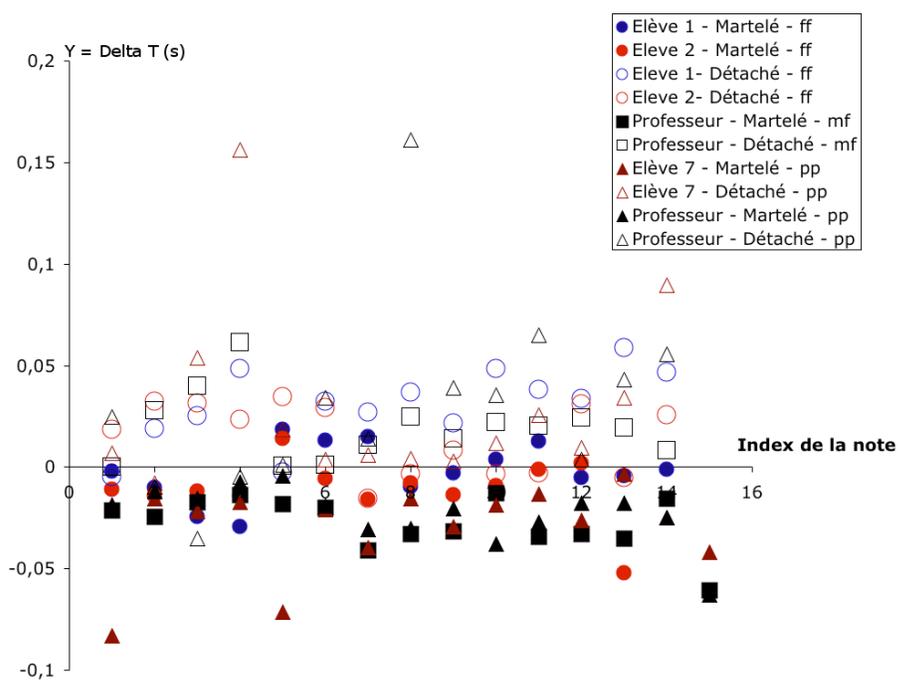


FIG. 6.5 – Chaque symbole représente une nuance : ● : *ff*, □ : *mf*, Triangle : *pp*. Les symboles pleins correspondent au mode de jeu détaché, les symboles creux au mode de jeu martelé. Chaque joueur est représenté par une couleur.

6.2.2 Relation entre les amplitudes

De l'observation des deux signaux sur une gamme, le premier constat que l'on peut faire est que leurs amplitudes n'ont pas un lien direct. En effet, le geste de l'instrumentiste est relativement constant sur la gamme, et l'amplitude des pics d'accélération varie faiblement entre chaque coup d'archet. Par contre, la puissance du résiduel varie fortement avec la gamme. On peut partiellement expliquer cela par le fait que la longueur vibrante de la corde diminue avec la hauteur de la note (les gammes sont toutes jouées sur une seule corde). En résulte une baisse de la dynamique du son. Pour quantifier la relation entre l'amplitude du résiduel et le geste, il nous manque donc la pression d'archet. Les recherches en cours à l'Ircam devraient bientôt aboutir à la création d'un capteur de pression sur l'archet. Ce nouvel élément dans la modélisation du geste devrait nous permettre d'approfondir le lien entre résiduel et geste. Prendre en compte la pression pour l'étude des transitoires est d'ailleurs un des résultats des études de Guettler présentées dans le chapitre 3.

6.3 Discussion

Mesurer le temps de décalage entre le maximum de puissance du résiduel et le pic d'accélération sur un transitoire semble permettre de caractériser efficacement nos deux modes de jeu. On peut apporter quelques éléments d'explication. Pour une note martelée, le violoniste comprime dans l'archet de l'énergie qu'il libère d'un coup dans la corde. L'accélération violente provoque un fort niveau de bruit (*cf* Diagramme de Schelleng, Chapitre 3). Pour une note détachée cet effet est beaucoup moins présent, et on se situe plus dans la zone (1) du diagramme de Guettler.

En plus d'être efficace dans la caractérisation des modes de jeu, cette description croisée entre geste et son est extrêmement locale en temps. La modélisation des modes de jeu par le geste donne de bons résultats de reconnaissance, mais elle nécessite que chaque note soit effectuée entièrement. Avec la description que nous apportons, il suffit d'une cinquantaine de millisecondes pour reconnaître le mode de jeu dans lequel va être jouée la note. Une autre propriété intéressante de ce descripteur est qu'il montre une forte propriété d'invariance au niveau du joueur. On peut donc imaginer donner une modélisation universelle des transitoires en affinant l'étude de la relation entre le bruit et le signal d'accélération.

Chapitre 7

Conclusion et Perspectives

Nous avons introduit dans ce rapport les résultats d'une analyse croisée geste-son des articulations du violon. En nous basant sur une étude physique du système corde-archet, nous avons orienté notre étude vers la description de la relation geste-son pendant les transitoires du violon. Afin de caractériser ces transitoires, nous avons cherché à en donner une description sonore. Nous sommes partis pour cela du constat simple que chaque transitoire est toujours accompagné d'un bruit. Par l'extraction de ce bruit avec une méthode fiable, nous avons pu vérifier l'étroite relation entre geste et son pendant les transitoires en remarquant la simultanéité des maxima du signal d'accélération et du bruit.

L'étude croisée du signal d'accélération de l'archet et de la puissance de ce bruit a permis d'introduire un descripteur puissant des transitoires du violon. Simplement basé sur le temps de latence entre l'arrivée des pics d'accélération et de puissance du résiduel dans le temps, ce descripteur permet de déterminer efficacement en quelques millisecondes le mode de jeu dans lequel sera jouée la note. Il présente de plus une forte propriété d'invariance aux changements de joueur.

Ces propriétés de notre descripteur trouveront un domaine d'application privilégié dans le cadre de la composition musicale. En effet, la reconnaissance en temps-réel des différents modes de jeu du violon constitue un matériau riche pour la composition (*cf* chapitre 2). Une analyse seule du geste a déjà permis de créer un système de reconnaissance temps-réel des modes de jeu. Néanmoins, ce système a besoin d'analyser la quasi totalité de la note pour identifier le mode de jeu. Calculé au début de la note sur une durée très brève, le descripteur croisé geste-son peut permettre d'améliorer considérablement les performances de ce type de systèmes, à la fois sur un plan technologique (amélioration du coût de calcul) et sur un plan musical (système plus réactif).

Pour bien modéliser l'expression musicale sur le violon, cette description pourra être approfondie. Nous avons quelques pistes pour son amélioration. Premièrement, nous n'avons pas encore pu intégrer à notre étude la pression de l'archet sur la corde. Les récentes recherches effectuées à l'Ircam permettront d'ici peu On peut encore une fois s'appuyer sur l'étude physique du système

corde-archet pour penser que c'est une piste justifiée dans l'étude des transitoires du violon. Deuxièmement, nous n'avons considéré ici que la puissance du signal résiduel. Une des voies d'amélioration serait d'utiliser le signal résiduel plus finement.

On pourra par exemple appliquer ces résultats au contrôle d'un modèle physique d'une corde frottée avec les signaux gestuels du violon augmenté.

On peut également imaginer construire à partir de notre description un modèle qui permette de passer d'un mode de jeu à un autre. Piloter en temps réel un système de synthèse sonore avec un tel modèle permettrait de traduire une partie des possibilités d'expressivité donnée par le violon acoustique.

Bibliographie

- [1] Roland Badeau. *Méthodes à haute résolution pour l'estimation et le suivi de sinusoïdes modulées. Application aux signaux de musique*. PhD thesis, ENST, 2005.
- [2] Emmanuel Fléty, Nicolas Leroy, Jean-Christophe Ravarini, and Frédéric Bevilacqua. Versatile sensor acquisition system utilizing network technology. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2004.
- [3] K Guettler, E Schoonderwaldt, and A Askenfelt. Bow speed or bowing position - which one influences spectrum the most? In *Proceedings of the Stockholm Music Acoustics Conference, SMAC*, 2003.
- [4] Knut Guettler. Looking at starting transients and tone coloring of the bowed string.
- [5] Knut Guettler. *Applications of the bluestein filter in bowed-string analysis*. PhD thesis, KTH, 2003.
- [6] Frédéric Bevilacqua Nicolas Rasamimanana, Emmanuel Flety. Gesture analysis of violoin bow strokes.
- [7] Geoffroy Petters. Automatic classification of large musical instrument databases using hierarchical classifiers with inertia ratio maximization. In *AES Convention Paper*, 2003.
- [8] Geoffroy Petters. A large set of audio features for sound description (similarity and classification) in the cuidado project. 2004.
- [9] S.Balakrishnama and A.Ganapathiraju. Linear discriminant analysis - a brief tutorial.
- [10] J.C. Schelleng. The bowed string and the player. *J. Acoust. Soc. Am.*, 53(1), 1973.