

Synthèse et transformation sonore par descripteurs de haut-niveau

Damien Tardieu

DEA

Acoustique, traitement du signal et informatique appliqués à la musique

Université Aix Marseille II

Mémoire réalisé à
IRCAM
75004 Paris
1 place Igor Stravinsky

Responsables : Xavier Rodet, Geoffroy Peeters

Septembre 2004

Remerciements

Je tiens à remercier :

Xavier Rodet de m'avoir accepté dans son équipe pour effectuer ce stage et pour toutes les conversations que nous avons pu avoir sur mon travail.

Geoffroy Peeters dont les conseils, orientations et enseignements m'ont permis de mener à bien mes travaux.

Steve McAdams pour avoir pris le temps d'écouter et de répondre à toutes mes questions sur la perception du timbre.

Alice Daquet, Romain Kronenberg, Matthias Demoucron et Grégoire Carpentier qui m'ont aidé et soutenu durant mes journées et mes nuits à l'IRCAM.

Tous les membres de l'équipe Analyse-Synthèse pour leur accueil et leur disponibilité.

Table des matières

1	Introduction	5
2	Etat de l'art	7
2.1	Description du son	7
2.1.1	L'approche perceptive et cognitive	7
2.1.2	L'approche signal	8
2.2	La transformation sonore de haut-niveau	9
2.2.1	Transformation basée sur le modèle SMS	9
2.2.2	La psychoacoustique	9
2.2.3	Transformations et synthèses basées sur les descripteurs obtenus par des méthodes statistiques	10
2.2.4	Le modèle de Jensen	10
2.3	Conclusion	11
3	Les descripteurs retenus pour le modèle de transformation	12
3.1	Description de l'enveloppe spectrale	12
3.1.1	L'enveloppe spectrale	12
3.1.2	Les partiels	14
3.2	Etude de la valeur des moments sur un ensemble de sons	14
4	Modélisation de l'enveloppe spectrale par splines	18
4.1	Introduction	18
4.2	Les splines	18
4.2.1	Description générale et définitions	18
4.2.2	Multiplicité des noeuds et continuité	20
4.2.3	Approximation de fonction a l'aide de splines	21
4.2.4	Produit scalaire dans $S_{k,t}$	21
4.3	Application à l'approximation et à la transformation d'enveloppe spectrale	22
4.3.1	Approximation de l'enveloppe spectrale	22
4.3.2	Transformation par changement de la valeur des moments	22
4.4	Modélisation de l'évolution temporelle des descripteurs	25

5	Description du système	28
5.1	L'analyse	28
5.1.1	L'analyse Additive	28
5.1.2	Estimation de l'enveloppe spectrale : Le cepstre discret	28
5.1.3	Approximation et transformation de l'enveloppe par la méthode des B-splines	30
5.2	La synthèse	30
5.2.1	Reconstruction de l'enveloppe à partir des nouveaux coefficients	30
5.2.2	La synthèse additive	30
6	Résultats	32
6.1	Présentation	32
6.1.1	Analyse et synthèse sans transformation	32
6.1.2	Changement de la valeur d'un des moments	32
6.1.3	Lissage des descripteurs au cours du temps	34
7	Conclusion	38

Chapitre 1

Introduction

De nombreuses méthodes de synthèse sonore existent, chacune fondée sur un paradigme différent : superposition de sinusoides, synthèse par formant, synthèse soustractive, synthèse par forme d'onde élémentaire, synthèse par modulation fréquentielle, etc. Ces méthodes permettent une modélisation très précise du signal sonore. Cependant il en résulte un nombre important de paramètres qu'il devient difficile de contrôler lorsqu'on décide de manipuler le timbre.

En effet, le timbre est un attribut multidimensionnel du son et ses dimensions sont mal connues. La difficulté de le manipuler de manière scientifique provient alors de la mauvaise compréhension que nous en avons. Mais inversement, nous pourrions aussi dire que la mauvaise compréhension que nous en avons provient de la difficulté de le manipuler [HKCH97]. Les questions de la compréhension et de la manipulation du timbre se posent alors en miroir. Elles sont sinon identiques, du moins intimement liées. Quelles axes de recherche nous permettrait alors de sortir de cette boucle ?

Se reposer sur une définition du timbre pose le problème du choix, car il n'en existe pas qui soit unanimement admise, et le timbre est en général défini par ce qu'il n'est pas. Pour notre étude, nous avons choisi la définition suivante : "Le timbre est l'attribut de la sensation auditive suivant lequel un auditeur peut différencier deux sons en utilisant tous critères autres que la hauteur, la sonie et la durée". Cette définition présente, en effet, l'avantage de proposer un axe de recherche exploitable : la perception, en se basant sur une sensation auditive pour qualifier le timbre. De plus, la supposition que des comparaisons de timbre peuvent être effectuées par un jugement de dissemblance perceptive entre deux sons de hauteurs différentes a récemment été confirmée [MdCMW03]. Nous nous proposons donc de prendre en compte les dimensions qui sous tendent la perception du timbre comme paramètres pour optimiser son contrôle. Les résultats des études de perception et cognition musicale, étant expérimentalement validés il semble pertinent de les privilégier. Cependant ces expériences fournissent un nombre limité de descripteurs. Nous devons en chercher d'autres, nous nous basons pour cela sur les travaux de description du son issus du traitement du signal.

Une fois identifié notre ensemble de descripteurs, il est question de les manipuler et donc de faire des choix. En effet, suite à l'analyse d'un son, plusieurs transformations sont

possibles, et il faut décider de quelle manière passer des valeurs obtenues de nos descripteurs aux valeurs souhaitées. Pour cela, nous devons créer un modèle ou une méthode de transformation.

Après avoir présenté les travaux effectués concernant les descripteurs du son et les modèles de synthèse déjà existants dans le chapitre 2, nous définissons les descripteurs que nous avons choisis (chapitre 3).

Ensuite, dans le chapitre 4, nous proposons une modélisation par splines de l'enveloppe spectrale ainsi qu'une méthode permettant d'obtenir une nouvelle enveloppe à partir de la valeur désirée des descripteurs. Enfin le résultat des modélisations et des transformations est étudié et critiqué dans le chapitre 6.

Les résultats de ce travail étant principalement des sons, nous en avons mis quelques uns sur la page web suivante :

www.ircam.fr/anasyn/dtardieu/exemples.html

Chapitre 2

Etat de l'art

2.1 Description du son

Dans ce chapitre, nous montrons l'existant en terme de description du son et de synthèse par descripteurs de haut-niveaux.

2.1.1 L'approche perceptive et cognitive

L'approche principalement adoptée dans l'étude du timbre en psychoacoustique consiste à décrire les dimensions perceptives du timbre en terme de propriétés abstraites du son [Don97], [MWD⁺95]. D'après cette approche, les sons seraient perçus selon certains descripteurs acoustiques (spectraux, temporels ou spectraux-temporels) créant un "espace de timbre".

La détermination de cette espace se fait généralement en utilisant des méthodes d'analyses multidimensionnelles sur les réponses de sujets à des tâches de jugement de (dis)similarités entre différents sons. L'analyse fournit des axes expliquant au mieux les réponses des sujets. Il convient ensuite de déterminer les corrélats acoustiques de ces axes. Les descripteurs acoustiques généralement mis en évidence sont le centroïde et l'étendue spectrale [PMH00], le temps d'attaque et le flux spectral. Ces études, réalisées sur des sons instrumentaux ou des synthèse croisées entre instruments, nous permettent de comprendre les descripteurs principaux permettant de différencier les sons. Cependant, toutes ces expériences sont réalisées sur des sons de même fréquence fondamentale. Mais qu'en est-il des comparaisons entre des sons de fréquences différentes ? [MdCMW03] démontre qu'il est possible d'effectuer des comparaisons de timbre pour des sons de hauteurs différentes (les intervalles étudiés sont de 2 et 11 demi-tons). De plus ses expériences semblent montrer que les positions relatives mais aussi les positions absolues des sons dans l'espace de timbre restent à peu près constantes malgré les changements de hauteur. Ceci suggère que la valeur d'un descripteur acoustique sous-tendant un axe perceptif doit être indépendante de la hauteur du son.

2.1.2 L'approche signal

La description du son a aussi été largement étudiée dans le domaine du traitement du signal. Deux approches différentes sont utilisées. La première consiste à utiliser des descripteurs acoustiques déjà connus, issus de divers domaines scientifiques, la seconde utilise des méthodes statistiques pour déterminer automatiquement les descripteurs.

Description utilisant des descripteurs connus

Pour des applications de classification automatique, un ensemble de descripteurs issus du signal sonore est créé. Ces descripteurs sont empruntés à différents domaines comme la perception (voir paragraphe 2.1.1), la physiologie, le traitement de la parole ou le traitement du signal. Ensuite, par apprentissage sur une base de données de sons, des valeurs de descripteurs sont associées à des classes de sons. Ainsi [Pee03b] a établi un ensemble d'une cinquantaine de descripteurs dans le cadre du projet européen CUIDADO. Ces descripteurs sont instantanés c'est à dire calculés sur chaque trame de signal ou globaux c'est à dire calculés pour la totalité du signal, ils décrivent la forme temporelle, la forme spectrale, le contenu harmonique et, grâce à un modèle simplifié du processus d'audition humain, ils décrivent aussi les sons en termes perceptifs. La valeur des descripteurs instantanés est calculée pour chaque trame de signal, ensuite leur évolution dans le temps est caractérisée par une moyenne, une variance, une approximation polynomiale ou une modulation. En terme de description de la forme spectrale, nous retiendrons les descripteurs suivants : le centroïde spectral, l'étendue spectrale issus de la psychoacoustique mais aussi les moments d'ordre supérieur, la skewness et la kurtosis.

Analyse en composantes principales sur des bases de données de sons

La seconde approche consiste à effectuer une analyse statistique de type PCA sur une base de donnée de son.

Ainsi Hourdin, Charbonneau et Moussa [HCM97a] ont créé un "espace de timbre" en effectuant une analyse proche de l'analyse en composante principale (Factorial Analysis of Correspondance) sur un ensemble de sons. Cet ensemble est constitué de 40 sons de différents instruments tous joués à la même hauteur (E4). Les sons sont analysés par la méthode des filtres hétérodynes qui donne l'amplitude et la fréquence des partiels d'un son quasi harmonique en fonction du temps. Le nombre de partiels choisi est de 40, ainsi un son est représenté par une trajectoire dans un espace à 80 dimensions (40 de fréquence et 40 d'amplitude). Le but est donc de réduire le nombre de dimensions tout en conservant les qualités perceptives du son. L'analyse des résultats indique que six facteurs contiennent 85 % de l'information et 3 en contiennent 68%. Une interprétation physique des trois premiers axes est aussi proposée. Le premier axe est associé à l'énergie du signal, le second à la largeur du spectre et le troisième à la proportion d'énergie concentrée dans les basses fréquence.

2.2 La transformation sonore de haut-niveau

Les descripteurs de haut niveau qui ont le plus fait l'objet d'étude sont la hauteur (transposition) et la durée des sons (dilatation/compression temporelle). Les méthodes classiques d'analyse-synthèse permettent des transformations très précises de ces deux descripteurs.

En revanche peu de méthodes ont été proposées pour le timbre du fait du manque de connaissance à son sujet, ce champ étant plutôt traité de manière empirique grâce à des effets tels que ceux utilisés par les ingénieurs du son (flanger, chorus, equalisation ...).

Les techniques proposées sont en général basées sur l'analyse/synthèse par addition de sinusoides. Permettant une analyse très fine du signal et une resynthèse presque parfaite mais au prix d'une multiplication du nombre de paramètres (fréquence, amplitude et phase des partiels pour chaque trame de signal), cette méthode pose le problème du contrôle de manière évidente. Ses paramètres sont tout à fait compréhensibles intuitivement mais leur manipulation individuelle est fastidieuse et les résultats peu prévisibles.

Parmi les méthodes que nous allons présenter, nous insisterons particulièrement sur le "Timbre Model" de Kristofer Jensen qui est le seul à notre connaissance à proposer un modèle complet basé sur des descripteurs perceptifs.

2.2.1 Transformation basée sur le modèle SMS

Xavier Serra et Jordi Amatrian [SB98], [ABLS02] proposent un ensemble de transformations basées sur le modèle SMS (Spectral Modeling System). Ce modèle est une variante du modèle additif. Les transformations sont en général des applications directes des possibilités de SMS comme la translation ou la distorsion du spectre harmonique. Trois transformations donnent accès à des descripteurs de plus haut niveau : le vibrato, le tremolo (où l'on postule que la fréquence ou l'amplitude des partiels varient sinusoidalement) et le changement du genre d'un locuteur.

Les auteurs suggèrent la possibilité d'effectuer d'autres transformations mais ne précisent pas de méthodes.

2.2.2 La psychoacoustique

Pour vérifier les hypothèses relatives aux dimensions perceptives du timbre obtenues sur des sons instrumentaux, la psychoacoustique a recours à des modèles spectraux permettant de contrôler les paramètres physiques sous tendant les dimensions perceptives. Le modèle généralement adopté est que le spectre est une droite. Le centroïde spectral est alors relié à la pente de la droite. Ensuite des ajustements sont effectués sur l'amplitude des harmoniques pour contrôler d'autres descripteurs comme la proportion d'harmonique pairs et impairs.

Ce modèle, parfaitement adapté à son objectif (qui est de vérifier la possibilité de prédire la position d'un son dans un espace de timbre), ne semble pas adapté à un usage musical, du fait de la trop grande approximation de la forme du spectre.

2.2.3 Transformations et synthèses basées sur les descripteurs obtenus par des méthodes statistiques

L'analyse décrite au paragraphe 2.1.2 a donné lieu à une méthode de transformation et de synthèse [HCM97b].

L'espace réduit obtenu permet d'effectuer des interpolations entre instruments ou des synthèses en déplaçant les coordonnées du sons. En effet, du fait de la méthode d'analyse utilisée, il est possible de recréer un son à partir d'une trajectoire dans l'espace.

2.2.4 Le modèle de Jensen

Kristofer Jensen propose dans [Jen99] un modèle de timbre que nous résumons ici.

Le modèle d'enveloppe

Jensen définit l'enveloppe spectrale comme le maximum au cours du temps de l'amplitude $a_k(t)$ des partiels.

$$a_k = \max_t a_k(t)$$

Ainsi une seule enveloppe spectrale est définie pour tout le son. Celle-ci est modélisée par quatre descripteurs proches de la perception :

La brillance

$$T_b = \frac{\sum_{k=1}^N k a_k}{\sum_{k=1}^N a_k} \quad (2.1)$$

Le tristimulus 1

$$T_1 = \frac{a_1}{\sum_{k=1}^N a_k} \quad (2.2)$$

Le tristimulus 2

$$T_2 = \frac{a_2 + a_3 + a_4}{\sum_{k=1}^N a_k} \quad (2.3)$$

La proportion d'harmoniques impairs

$$T_o = \frac{\sum_{k=2}^{N/2} a_{2k-1}}{\sum_{k=1}^N a_k} \quad (2.4)$$

Bien sur, il existe une infinité d'enveloppe pour une valeur de ces descripteurs. Pour limiter le nombre de possibilité, Jensen impose une modélisation linéaire de l'enveloppe spectrale en log-amplitude pour les partiels de rang supérieur à 4 :

$$a_k = \begin{cases} B^{-k}, & \text{si } k \text{ impair} \\ k_0 B^{-k}, & \text{si } k \text{ pair} \end{cases} \quad (2.5)$$

où B et k_0 dépendent de la brillance T_b et de la proportion d'harmoniques impairs T_0 selon les expressions suivantes :

$$B = \frac{T_b}{T_b - 1}$$

$$k_0 = (B + B^2 - 1) \frac{T_0}{1 - T_0}$$

Ces expressions correspondent à une approximation où toute l'enveloppe serait décrite par (2.5) et où le nombre de partiels serait infini. Une fois B et k_0 obtenus, les coefficients a_1 jusqu'à a_4 peuvent être calculés grâce aux équations (2.1), (2.2), (2.3), (2.4).

Ce système donne parfois des résultats négatifs, dans ce cas une procédure itérative permet de déterminer les valeurs des harmoniques basses en ajustant B et k_0 et en tenant compte de l'irrégularité du spectre définie comme $\sum_{k=1}^N (a_k - a_{k+1})^2 / \sum_{k=1}^N a_k^2$.

2.3 Conclusion

Parmi tous les descripteurs acoustiques décrivant la timbre nous avons choisi d'en garder quatre, les quatre premiers moments du spectre. Les deux premiers, le centroïde et l'étendue spectral, sont souvent identifiés comme sous-tendant le deuxième et le troisième axe de l'espace perceptif du timbre. Le rôle de la skewness et de la kurtosis n'a pas été réellement étudié, cependant elles sont utilisées en classification et la skewness peut être associée au troisième axe de [HCM97a]. De plus ces deux descripteurs ont une cohérence mathématique avec le centroïde et l'étendue spectrale ce qui facilite la modélisation.

Nous allons donc modéliser l'enveloppe spectrale d'un son principalement par ses quatre premiers moments. Nous verrons dans la suite que d'autres descripteurs se sont révélés nécessaires pour obtenir un modèle intéressant.

Chapitre 3

Les descripteurs retenus pour le modèle de transformation

Ce chapitre présente l'ensemble des paramètres que nous utiliserons dans notre modèle d'enveloppe spectrale.

3.1 Description de l'enveloppe spectrale

Généralement, les descripteurs spectraux sont calculés soit sur le spectre d'une trame de son, soit sur les amplitudes des harmoniques de cette trame, soit sur un banc de filtre. Nous avons choisi calculer ces descripteurs sur l'enveloppe spectrale. Les valeurs calculées sur l'enveloppe spectrale ne sont pas les mêmes que celles calculées sur les harmoniques mais elles sont proportionnelles entre elles tant que le nombre d'harmoniques est suffisamment élevé. Or, dans notre contexte, se sont bien les valeurs relatives des descripteurs qui nous intéressent, de plus nous cherchons à contrôler des sons qui contiennent un grand nombre d'harmoniques, puisque, si un son en contient peu, il est plus aisé de manipuler les harmoniques individuellement que de manipuler l'enveloppe spectrale.

3.1.1 L'enveloppe spectrale

Nous avons choisi de décrire l'enveloppe spectrale par sa forme et sa largeur (que nous appellerons bande passante). Nous travaillons sur une enveloppe de bande passante normalisée à 1. La forme de l'enveloppe est alors obtenue grâce à ses quatre premiers moments. Ensuite la véritable enveloppe est obtenue par changement d'échelle de facteur égal à la bande passante.

Tous les moments sont calculés sur l'enveloppe spectrale $a(\nu)$ en fréquence linéaire normalisée $\nu \in [0, 1]$ et amplitude logarithmique normalisée. La fréquence normalisée 0 est associée à la fréquence du premier harmonique f_1 , tandis que la fréquence 1 est associée à la fréquence de coupure voisée/non-voisée f_c , c'est à dire la fréquence au-delà de laquelle

le spectre est mieux expliqué par du bruit que par des partiels quasi-harmoniques [Pee01]. Les amplitudes logarithmiques sont calculées comme suit :
 Soit e l'amplitude linéaire et a l'amplitude logarithmique de l'enveloppe spectrale.

$$a = \log_{10}\left(\frac{e}{e_0}\right)$$

où e_0 est un seuil choisi arbitrairement pour que les amplitudes en logarithmes soit toujours positives. Dans nos expériences nous prenons $e_0 = 10^{-8}$. Ensuite les amplitudes sont normalisées :

$$a_{norm}(\nu) = \frac{a(\nu)}{\int_0^1 a(\nu) d\nu}$$

Dans la suite nous appelons $a\nu$ l'enveloppe normalisée.

Le centroïde spectral

Le centroïde spectral ou centre de gravité à été identifié comme étant l'un des paramètres les plus importants pour la description du timbre dès les premières expériences sur la perception du timbre. Plusieurs définitions existent, nous adopterons la suivante :

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

L'étendue spectrale

Nous définissons l'étendue spectrale comme l'étendue du spectre autour de sa valeur moyenne.

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

La skewness

La skewness donne une mesure de la symétrie du spectre autour de sa moyenne. Elle est calculée à partir du moment d'ordre 3.

$$\gamma_1 = \frac{1}{\sigma^3} \int_0^1 (\nu - \mu)^3 a(\nu) d\nu$$

Une skewness nulle correspond à une distribution symétrique, une skewness négative indique une plus grande énergie à droite et une skewness positive une plus grande énergie à gauche. Le spectre d'un son étant en général décroissant la skewness est positive.

La kurtosis

la kurtosis donne une mesure de l'aplatissement du spectre autour de sa moyenne. Il est calculé à partir du moment d'ordre 4.

$$\gamma_2 = \frac{1}{\sigma^4} \int_0^1 (\nu - \mu)^4 a(\nu) d\nu$$

Une distribution normale a une kurtosis de 3. Plus la distribution est plate plus la kurtosis est faible, plus elle est pointue, plus la kurtosis est élevée.

La bande passante

Elle est définie par la différence entre la fréquence de coupure voisée/non-voisée et celle du premier partiel. Cette valeur est directement reliée par la fréquence fondamentale au nombre d'harmonique du son.

Dépendance des moments avec la bande passante

La kurtosis et la skewness sont indépendante de l'intervalle de fréquence utilisé pour les calculer. En revanche, le centroïde et l'étendue spectrale en dépendent. Ainsi si au lieu de les calculer sur des valeurs normalisées comprises entre 0 et 1, nous les calculons sur des fréquences réelles $f \in [0, F]$, ils se trouvent tous les deux multipliés par F .

Les équations suivantes indiquent comment calculer les moments en valeurs fréquentielles à partir des moments en valeurs normalisées, de la bande passante et de la fréquence f_1 du premier harmonique (qui est généralement égale à la fréquence fondamentale f_0).

$$\mu_f = \mu B_w + f_1$$

$$\sigma_f = \sigma B_w$$

$$\gamma_{1f} = \gamma_1$$

$$\gamma_{2f} = \gamma_2$$

3.1.2 Les partiels

Nous n'avons pas modélisé les fréquences des partiels, elles sont donc conservées, ainsi que leur phase et leur index.

3.2 Etude de la valeur des moments sur un ensemble de sons

Pour tenter de comprendre un peu mieux nos descripteurs spectraux, nous les avons calculés pour tous les éléments d'une base de données d'extraits sonores (Studio En Ligne

- IRCAM). Les extraits sont des notes seules jouées sur des instruments harmoniques soutenus ou non à différentes hauteurs. La base de données contient 1003 sons de 16 instruments.

Les moments sont calculés sur la fréquence et l'amplitude des partiels. L'amplitude et la fréquence des partiels est obtenues par analyse additive, les partiels dont la fréquence est supérieure à la fréquence de coupure voisé/non-voisé sont supprimés.

La figure 3.1 montre les corrélations entre les moments en valeurs fréquentielles, la figure 3.2 montre celles des moments normalisés.

La légende de chaque figure est la corrélation entre les deux moments représentés calculée par la méthode de Pearson.

Tout d'abord, nous remarquons que le centroïde et l'étendue spectrale sont fortement corrélés (.98) lorsqu'ils sont calculés en fréquences réelle. En effet, du fait de la forme dissymétrique du spectre des instruments de musique, les deux ont tendance à augmenter lorsque la largeur du spectre augmente. D'ailleurs, le fait de normaliser les fréquences (ce qui annule l'effet de la largeur du spectre) rend la corrélation très faible. En revanche, la corrélation est déplacée à la skewness. En effet, dans ce cas, un son dont le centroïde est élevé aura plus d'énergie à droite.

Cette étude nous permet aussi d'avoir une idée sur l'intervalle que couvre chaque moment (cf. tableau 3.1). Ces intervalles pourront être comparés aux résultats des transformations que nous allons proposer.

Moment	Minimum	Maximum
centroïde	0.3	0.5
étendue	0.26	0.34
skewness	0	0.8
kurtosis	1.5	2.5

TAB. 3.1 – Intervalles de valeurs des moments normalisés du Studio On Line

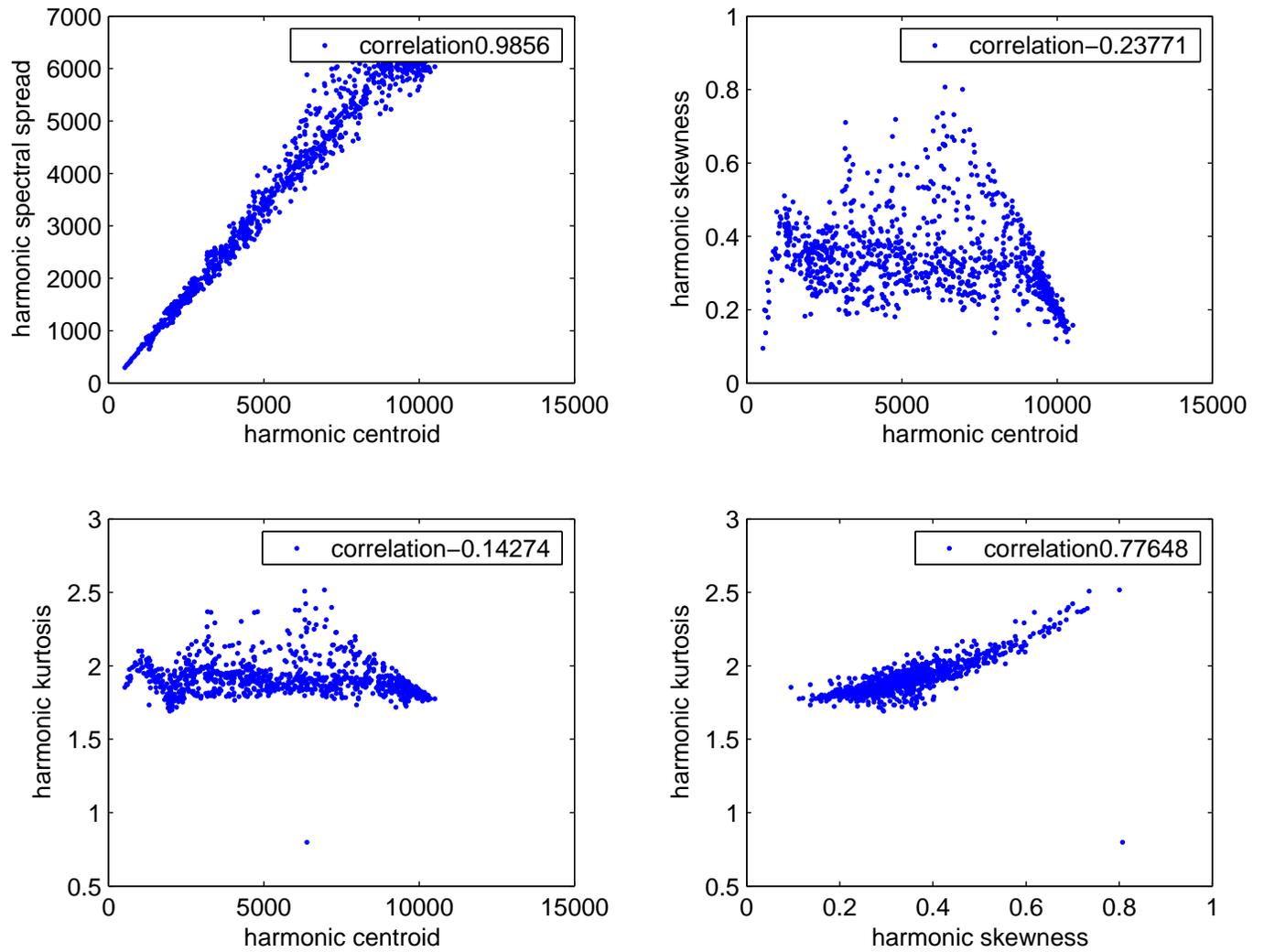


FIG. 3.1 – Etude de la corrélation entre les moments des sons du Studio En Ligne

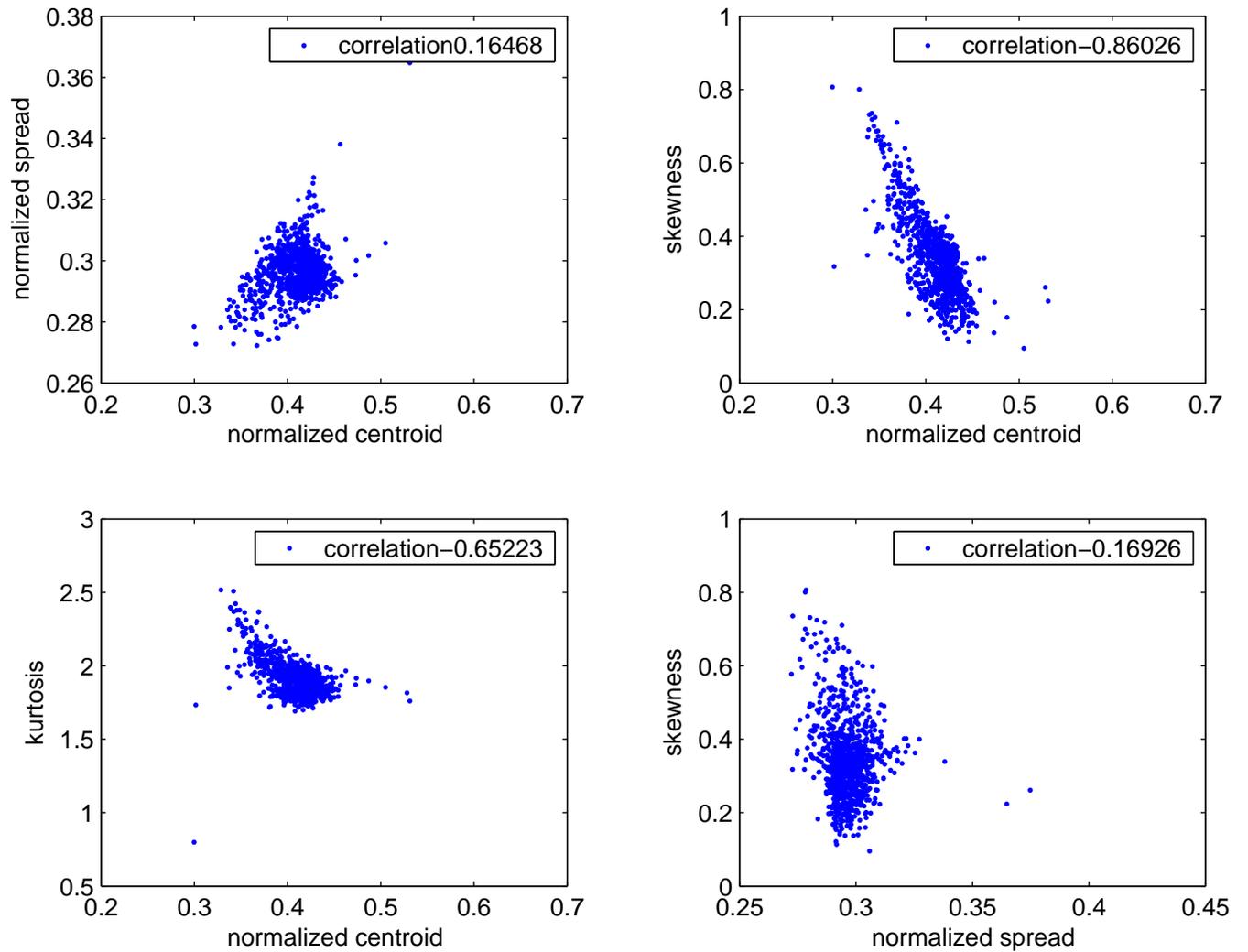


FIG. 3.2 – Etude de la corrélation des moments normalisés des sons du Studio En Ligne

Chapitre 4

Modélisation de l'enveloppe spectrale par splines

4.1 Introduction

Ce chapitre propose un modèle d'enveloppe spectrale permettant de transformer un son en changeant la valeur des descripteurs proposés au chapitre 3. Ce modèle est basé sur l'approximation en courbes splines. Ces courbes ont été choisies car elles permettent une approximation précise, lisse et cohérente de l'enveloppe, d'autre part nous pouvons lier analytiquement la valeur des descripteurs aux coefficients de la courbe. Enfin les propriétés des splines permettent que, quelle que soit la transformation opérée, l'enveloppe résultante reste toujours continue.

Nous commençons par définir ces courbes et en donnons quelques propriétés, ensuite nous montrons comment retrouver une enveloppe à partir de la valeur des descripteurs désirés.

4.2 Les splines

Les définitions sont empruntées à [dB93] et [UAE93]

4.2.1 Description générale et définitions

Les splines sont des polynômes par morceaux qui peuvent satisfaire certaines conditions de continuité. Les morceaux sont tous de même degré, ainsi l'ordre d'une spline est défini comme étant égale au degré des polynômes plus un. Les extrémités de chaque morceau sont appelées noeuds. L'ensemble des noeuds est appelé séquence de noeuds. Cette dernière peut contenir plusieurs fois le même noeuds, dans ce cas nous parlerons de noeuds multiples (double, triple ...).

Nous appelons $S_{k,t}$ l'ensemble des splines de degré k et de séquence de noeuds t .

Toutes splines peut-être définie comme une somme pondérée de fonctions de bases appelées B-splines. A chaque séquence de noeuds est associée un ensemble unique de B-splines dans lequel chaque B-spline est associé à un noeud. Ainsi nous pouvons définir $S_{k,t}$ comme suit :

Definition 4.2.1 Une spline d'ordre k et de séquence de noeuds t est une combinaison linéaire de B-splines $B_{i,k}$ associés à la séquence de noeuds. L'ensemble de ces splines est défini par

$$S_{k,t} = \left\{ \sum_i B_{i,k} a_i : a_i \in \mathbb{R} \right\}$$

Où les a_i sont appelés coefficients de la spline et les fonctions de base, ou B-splines $B_{i,k}$ sont définies par récurrence,

Definition 4.2.2 Soit une partition ou séquence de noeuds, i.e., une séquence croissante $t = (t_i)$. La B-spline $B_{i,k}(t)$ d'ordre $k > 1$ associée au noeud t_i est défini par :

$$B_{i,k}(t) = \omega_{i,k} B_{i,k-1} + (1 - \omega_{i+1,k}) B_{i+1,k-1}$$

avec

$$B_{i,1}(t) = X(t) = \begin{cases} 1, & \text{si } t_i \leq t < t_{i+1} \\ 0, & \text{sinon.} \end{cases}$$

et

$$\omega_{i,k}(t) = \begin{cases} \frac{t-t_i}{t_{i+k-1}-t_i}, & \text{si } t_i \neq t_{i+k-1} \\ 0, & \text{sinon.} \end{cases}$$

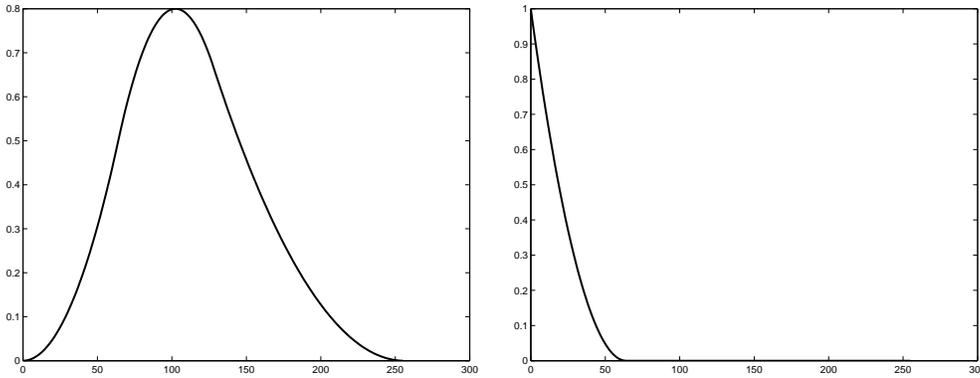


FIG. 4.1 – B-splines d'ordre 3 associées à (a) un noeud simple (b) un noeud triple

Proposition 4.2.1 Une B-spline d'ordre k associé au noeud t_i est nulle hors de l'intervalle $[t_i, t_{i+k})$, i.e.,

$$\forall k \forall t \notin [t_i, t_{i+k}) B_{i,k}(t) = 0$$

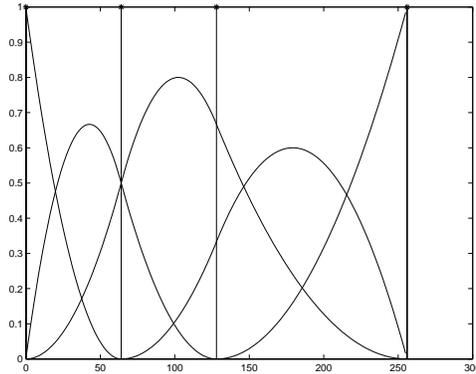


FIG. 4.2 – Ensemble des B-splines d’ordre associés à la séquence $\{0, 0, 0, 64, 128, 256, 256, 256\}$

4.2.2 Multiplicité des noeuds et continuité

La multiplicité d’un noeud permet de changer la condition de continuité à ce noeud, i.e. plus la multiplicité est élevée, moins la spline sera lisse.

Proposition 4.2.2 *soit $s \in S_{k,t}$ une spline d’ordre k et de séquence de noeuds t avec $t_i \in t$ un noeud de multiplicité m . Alors s est $k - m$ fois dérivable en t_i et les $k - m - 1$ premières dérivées sont continues.*

La figure 4.3 montre deux splines d’ordre 3 dont la position des noeuds est identique. Les noeuds aux extrémités de la courbe sont triples pour permettre à la courbe de ne pas partir de zéro, le troisième noeud est double dans (a) et triple dans (b), tous les autres sont simples. Notons que lorsque la multiplicité d’un noeud est égale à l’ordre de la spline, alors la courbe est discontinue.

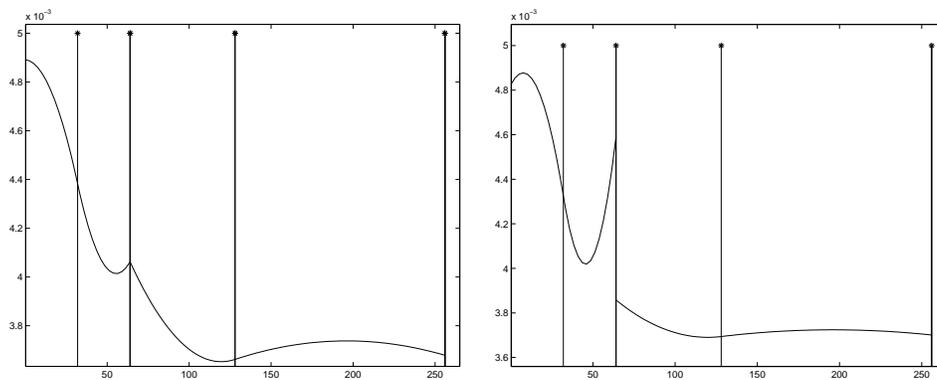


FIG. 4.3 – Splines de degré 3 avec (a) un noeud double (b) un noeud triple

4.2.3 Approximation de fonction a l'aide de splines

Les splines permettent d'approximer précisément une grande variété de courbes. L'approximation se fait en agissant sur les coefficients a_i afin de minimiser l'erreur quadratique. Deux paramètres doivent être déterminés à priori, ce sont la position des noeuds et l'ordre de la spline.

Dans notre contexte l'intérêt des splines se trouve, d'une part dans la qualité des approximations, d'autre part dans le contrôle qu'elles permettent. En effet, chaque B-spline est non nulle sur un intervalle restreint, ainsi en intervenant sur un coefficient nous ne modifions qu'une partie du spectre tout en conservant la continuité aux noeuds. Nous obtenons donc un nombre réduit, mais cohérent fréquemment, de paramètres de contrôle.

4.2.4 Produit scalaire dans $S_{k,t}$

Nous allons voir comment calculer un produit scalaire dans $S_{k,t}$. Comme $S_{k,t}$ n'est pas un espace orthogonal, il nous faut effectuer quelques manipulations pour simplifier le calcul.

Soit f et g deux fonctions de \mathbb{R} . Elles peuvent se décomposer sur $S_{k,t}$: $f(x) = \sum_i k_i B_i(x)$, $g(x) = \sum_i l_i B_i(x)$.

Nous avons,

$$\begin{aligned} \langle f, g \rangle &= \int f(x) g(x) dx \\ &= \int \sum_i k_i B_i(x) \sum_j l_j B_j(x) dx \\ &= \sum_i \sum_j k_i l_j \int B_i(x) B_j(x) dx \end{aligned}$$

Posons, $A_{ij} = \int B_i(x) B_j(x) dx$, et $\alpha_i = \sum_j l_j A_{ij}$

Nous obtenons l'expression suivante :

$$\langle f, g \rangle = \sum_i l_i \alpha_i \tag{4.1}$$

Ceci nous sera utile lorsque nous modifierons les moments de l'enveloppe spectrale pour déterminer les coefficients de l'enveloppe transformée.

4.3 Application à l'approximation et à la transformation d'enveloppe spectrale

Nous allons maintenant voir comment utiliser les fonctions splines pour changer la forme de l'enveloppe spectrale en modifiant la valeur de ses quatre premiers moments.

4.3.1 Approximation de l'enveloppe spectrale

Choix des noeuds et de l'ordre

Nous avons choisi des splines d'ordre 3, ce qui correspond à des polynômes de degré 2. Nous faisons ce choix car c'est l'ordre minimum qui donne une approximation précise de l'enveloppe.

Comme nous avons cinq descripteurs (les moments d'ordre 0 à 4) à notre disposition, nous voulons une spline caractérisée par cinq coefficients. D'après les propriétés de ces courbes et pour obtenir les bonnes conditions de continuités, cela permet de diviser l'axe des fréquences en trois parties. Cette partition est non uniforme pour donner un contrôle plus précis dans les basses fréquences. Nous avons donc choisi la séquence de noeuds suivante :

$$t = \{0, 1/4, 1/2, 1\}$$

Résiduel de modélisation

Les splines que nous utilisons donnent la forme globale de l'enveloppe mais n'en suivent pas tous les détails. Nous extrayons donc un résiduel de modélisation en soustrayant l'approximation à l'enveloppe originale. Ce résiduel est conservé tout au long du traitement puis rajouté après transformation. Ceci permet de conserver la "structure fine du spectre". Le calcul des descripteurs spectraux, réalisé sur l'approximation, n'est influencé par cette opération que de façon négligeable.

La figure 4.4 représente une enveloppe spectrale de guitare. La courbe lisse représente l'approximation en spline et les courbes en pointillés sont les B-splines. Cette figure permet aussi de voir quel sera le résiduel de modélisation.

4.3.2 Transformation par changement de la valeur des moments

Dans ce paragraphe, nous cherchons à déterminer un système reliant la valeur des moments de l'enveloppe spectrale aux coefficients de son approximation en splines.

Soit $e(\nu)$ la fonction décrivant l'enveloppe spectrale et $a(\nu)$ son approximation dans $S_{k,t}$ ie. $a(\nu) = \sum_i k_i B_i(x)$.

La valeur d'un moment est déterminée par une expression de la forme $m = k \int a(\nu) f(\nu) d\nu$ où f est égale à $\nu, (\nu - \mu)^2, (\nu - \mu)^3 \dots$

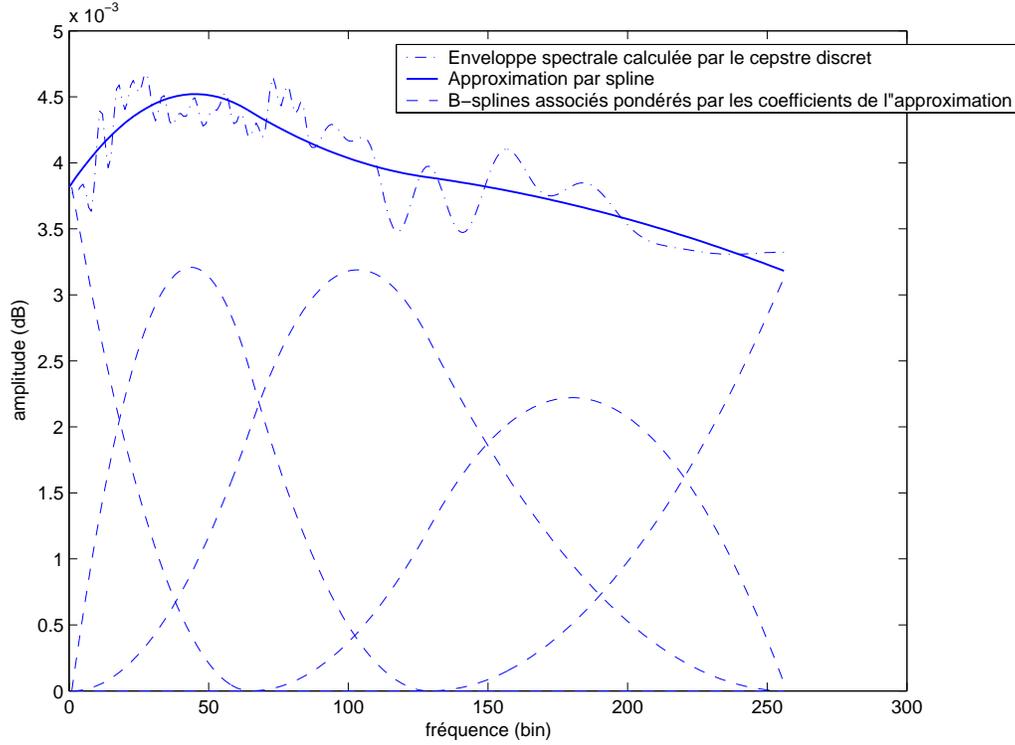


FIG. 4.4 – Enveloppe spectrale d’une guitare et son approximation par spline

Ceci peut-être interprété comme le produit scalaire entre a et f . Ainsi, si nous approximons f par splines, ie. $f(\nu) = \sum_i l_i B_i(\nu)$ et si nous posons :
 $\alpha_i = \sum_j l_j A_{ij}$ avec $A_{ij} = \int B_i(x) B_j(x) dx$
 on peut s’écrire d’après 4.1 de la manière suivante :

$$m = k \langle f, a \rangle = k \sum_i l_i \alpha_i$$

Ceci nous donne une équation reliant les l_i à la valeur des moments.

Cependant, les coefficients de l’équation α_i dépendent des l_i et donc de f . Or f est dépendant de μ pour les moments d’ordre supérieur à 2, elle est donc susceptible de varier au cours du temps et pendant les transformations.

Pour des questions de simplicité algorithmique nous prendrons donc $f = \nu, \nu^2, \nu^3 \dots$ ce qui nous permettra d’obtenir des α_i indépendants de μ . Pour cela il nous faut, dans un premier temps déterminer la valeur du produit scalaire entre $a(\nu)$ et les ν^i en fonction des moments.

Tout d’abord, les amplitudes sont normalisées :

$$\int a(\nu) d\nu = 1 \tag{4.2}$$

Puis d'après la définition de la moyenne :

$$\begin{aligned}\mu &= \int \nu a(\nu) d\nu \\ &= \langle \nu, a(\nu) \rangle\end{aligned}$$

Nous obtenons :

$$\boxed{\langle \nu, a(\nu) \rangle = \mu} \quad (4.3)$$

De même la variance est définie par :

$$\begin{aligned}\sigma^2 &= \int (\nu - \mu)^2 a(\nu) d\nu \\ &= \int (\nu^2 - 2\mu\nu + \mu^2) a(\nu) d\nu \\ &= \langle \nu^2, a(\nu) \rangle - 2\mu \langle \nu, a(\nu) \rangle + \mu^2 \langle 1, a(\nu) \rangle \\ &= \langle \nu^2, a(\nu) \rangle - \mu^2\end{aligned}$$

Donc,

$$\boxed{\langle \nu^2, a(\nu) \rangle = \sigma^2 + \mu^2} \quad (4.4)$$

La skewness est définie par :

$$\begin{aligned}\gamma_1 &= \frac{1}{\sigma^3} \int (\nu - \mu)^3 a(\nu) d\nu \\ &= \frac{1}{\sigma^3} \int (\nu^3 - 3\mu\nu^2 + 3\mu^2\nu - \mu^3) a(\nu) d\nu \\ &= \frac{1}{\sigma^3} (\langle \nu^3, a(\nu) \rangle - 3\mu \langle \nu^2, a(\nu) \rangle + 3\mu^2 \langle \nu, a(\nu) \rangle - \mu^3 \langle 1, a(\nu) \rangle) \\ &= \frac{1}{\sigma^3} (\langle \nu^3, a(\nu) \rangle - 3\mu(\sigma^2 + \mu^2) + 3\mu^3 - \mu^3) \\ &= \frac{1}{\sigma^3} (\langle \nu^3, a(\nu) \rangle - 3\mu\sigma^2 - \mu^3)\end{aligned}$$

Donc,

$$\boxed{\langle \nu^3, a(\nu) \rangle = \sigma^3\gamma_1 + 3\mu\sigma^2 + \mu^3} \quad (4.5)$$

Enfin, la kurtosis est définie par :

$$\begin{aligned}\gamma_2 &= \frac{1}{\sigma^4} \int (\nu - \mu)^4 a(\nu) d\nu \\ &= \frac{1}{\sigma^4} \int (\nu^4 - 4\mu\nu^3 + 6\nu^2\mu^2 - 4\mu^2\nu + \mu^4) a(\nu) d\nu \\ &= \frac{1}{\sigma^4} (\langle \nu^4, a(\nu) \rangle - 4\sigma^3\gamma_1\mu - 6\mu^2\sigma^2 - \mu^4)\end{aligned}$$

Donc,

$$\boxed{\langle \nu^4, a(\nu) \rangle = \gamma_2 \sigma^4 + 4\sigma^3 \gamma_1 \mu + 6\mu^2 \sigma^2 + \mu^4} \quad (4.6)$$

Maintenant, à partir des équations précédentes, nous pouvons déterminer les coefficients de la spline associée à l'enveloppe dont les moments ont les valeurs désirées.

Posons

$$\nu^n = \sum_i l_{n,i} B_i(x), \text{ pour } n \in \{0, 1, 2, 3, 4\}$$

$$A_{ij} = \int B_i(x) B_j(x) dx$$

$$\mathcal{M}_{n,i} = \sum_j l_{n,j} A_{ij}.$$

D'après (4.1)

$$\langle \nu^n, a(\nu) \rangle = \sum_i \mathcal{M}_{n,i} k_i$$

De plus En utilisant (4.2), (4.3), (4.4), (4.5), (4.6)

$$d = \begin{pmatrix} \langle 1, a(\nu) \rangle \\ \langle \nu, a(\nu) \rangle \\ \langle \nu^2, a(\nu) \rangle \\ \langle \nu^3, a(\nu) \rangle \\ \langle \nu^4, a(\nu) \rangle \end{pmatrix} = \begin{pmatrix} 1 \\ \mu \\ \sigma^2 + \mu^2 \\ \sigma^3 \gamma_1 + 3\mu \sigma^2 - \mu^3 \\ \gamma_2 \sigma^4 + 4\sigma^3 \gamma_1 \mu + 6\mu^2 \sigma^2 + \mu^4 \end{pmatrix}$$

Nous obtenons le système $\mathcal{M}k = d$ reliant les coefficients $k = (k_i)$ de l'approximation en spline de l'enveloppe spectrale et la valeurs des quatre premiers moments de cette enveloppe.

Pour obtenir les coefficients, il suffit d'inverser \mathcal{M} .

4.4 Modélisation de l'évolution temporelle des descripteurs

Nous avons essayé de modéliser l'évolution temporelle des descripteurs pour étudier l'influence d'un lissage et apporté un meilleur contrôle au cours du temps. Le signal est divisé en plusieurs parties selon une simplification du modèle ADSR (Attack, Sustain, Decay, Release). Ce modèle segmente l'enveloppe temporelle de l'énergie en quatre partie : l'attaque, une décroissance, une partie tenue et l'extinction. Dans notre cas l'algorithme découpe le son en deux ou trois partie selon ses caractéristique. Par exemple une note de guitare (voir figure 4.6) est découpée en trois (attaque, décroissance et extinction) et une note de contrebasse jouée à l'archet (voir figure 4.5) est découpée en une attaque et une extinction. Ensuite le trajet temporel des descripteurs est lissé par des splines d'ordre 2 pour conserver la continuité aux points de coupe [Pee03a]. Ainsi chaque zone est représentée par une droite dont nous pouvons contrôler les coefficients.

Il est donc possible de faire varier les moments indépendamment sur chaque partie en changeant les coefficients du polynôme correspondant ou, lors de synthèses croisées, conserver

la correspondance entre les parties des deux sons.

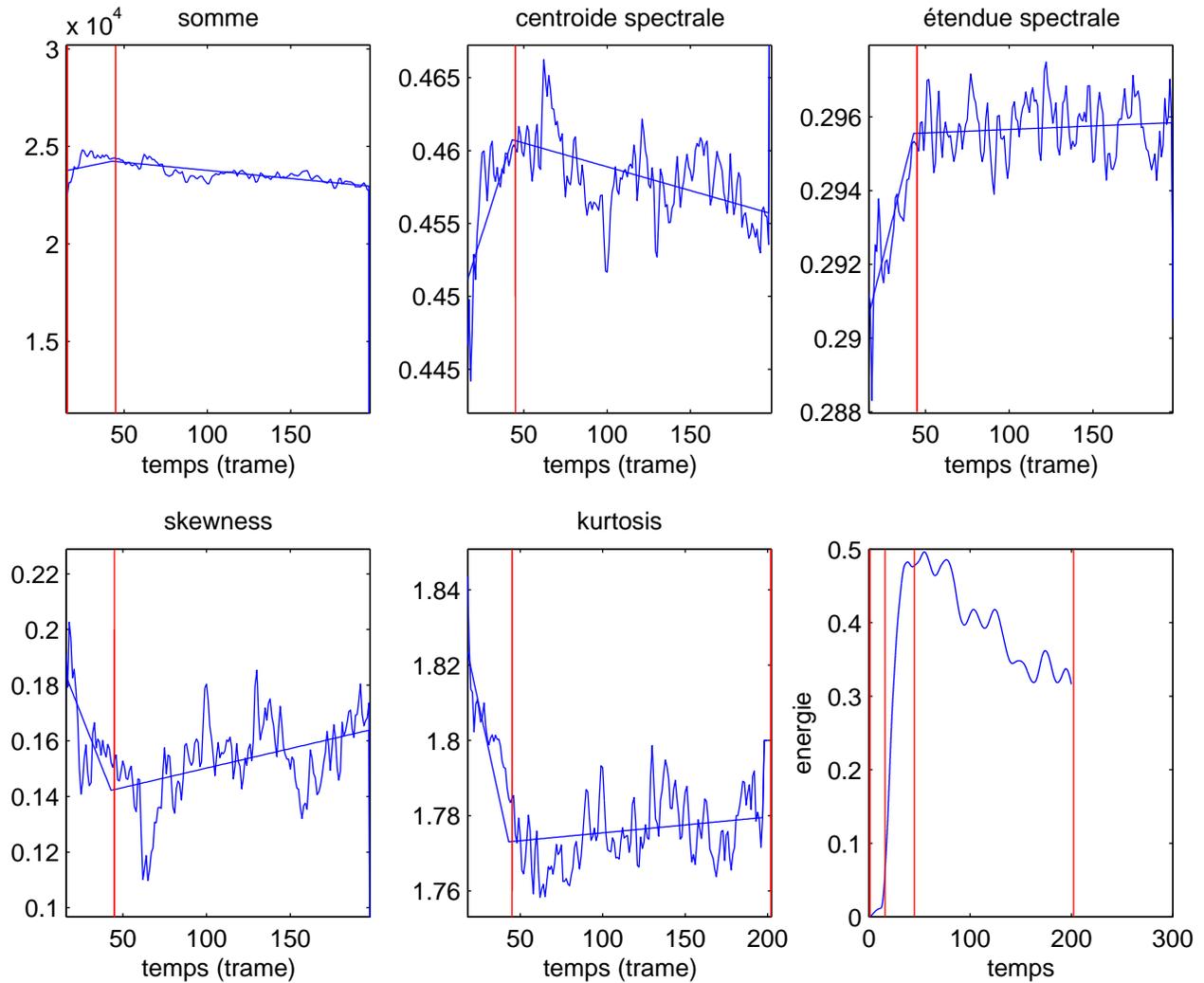


FIG. 4.5 – Evolution temporelle des quatre premiers moments d’une enveloppe spectrale de contrebasse et approximation par splines d’ordre 2 des parties correspondant à l’attaque et à la partie soutenue

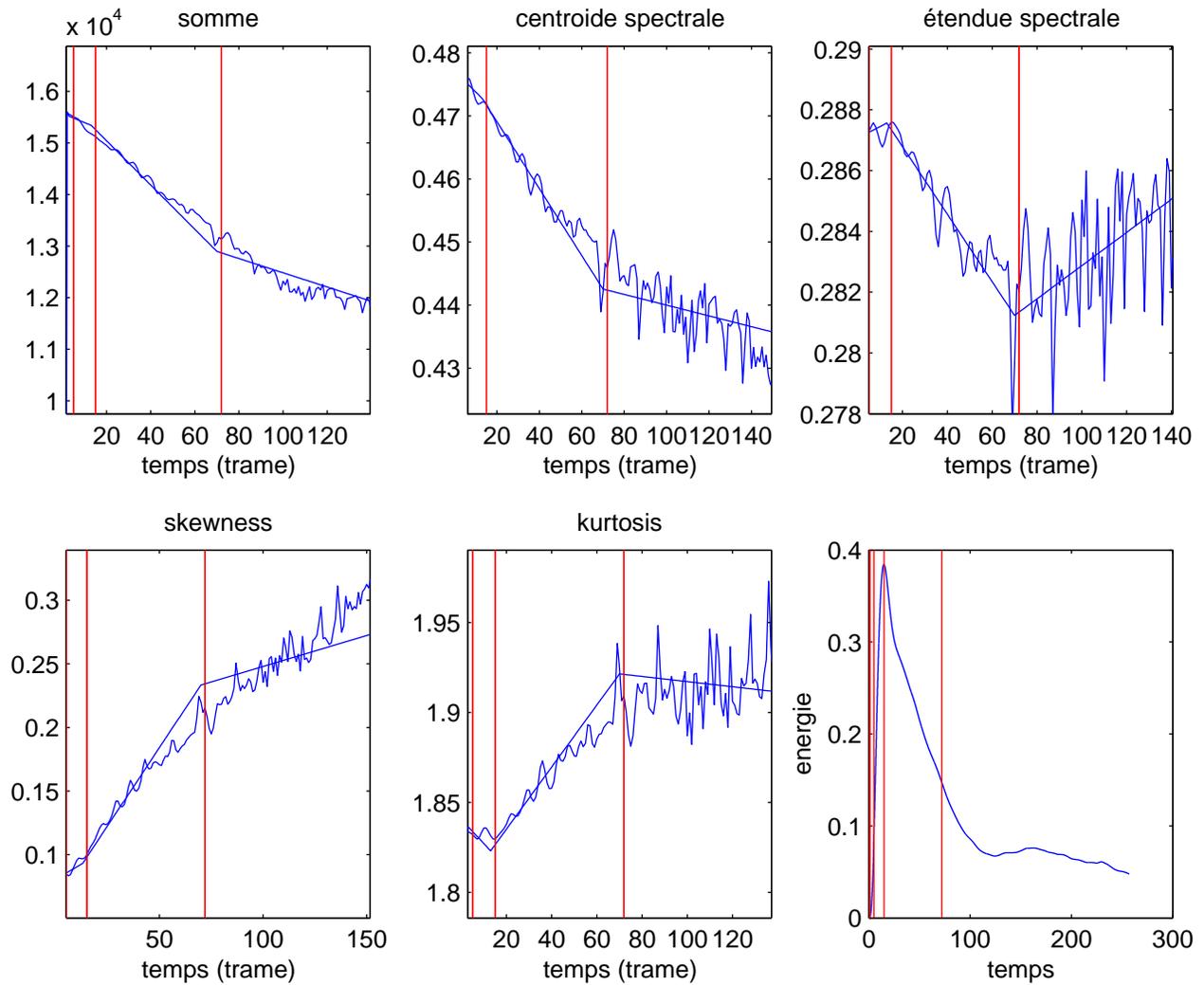


FIG. 4.6 – Evolution temporelle des quatre premiers moments d’une enveloppe spectrale de guitare et approximation par splines d’ordre 2 des parties correspondant à l’attaque, à la décroissance et à la décoissance

Chapitre 5

Description du système

L'ensemble du système est représenté sur la figure 5.1. Les paragraphes suivants détaillent chaque élément.

5.1 L'analyse

5.1.1 L'analyse Additive

La méthode d'analyse du signal utilisé est l'analyse additive [MQ86], [SS90]. Elle permet de décomposer le signal en une partie harmonique et un résiduel de modélisation. A chaque instant n le signal est modélisé de la manière suivante :

$$s(n) = \sum_{i=1}^{I(k)} a_i \cos(2\pi f_{i,k}n + \phi_{i,k}) + r(n) , \text{ pour chaque trame } k.$$

Où $I(k)$ représente le nombre de partiels (variable au cours du temps), $a_{i,k}$, $f_{i,k}$ et $\phi_{i,k}$ respectivement l'amplitude, la fréquence et la phase du partiel i , et $r(n)$ le résiduel de modélisation.

Outre les paramètres ci-dessus, l'analyse nous permet d'obtenir une fréquence de voisement, c'est à dire une fréquence au-delà de laquelle le signal est considéré comme plus bruité qu'harmonique. Ainsi, il est inutile de chercher des pics spectraux ou d'estimer l'enveloppe spectrale au-dessus de cette fréquence. La méthode utilisée est décrite dans [Pee01]. Celle-ci nous fourni une fréquence de voisement $f_v(k)$ pour chaque trame. Nous utilisons dans la suite une valeur unique de $f_v(k)$ pour tout le signal :

$$f_{vmax} = \max_k(f_v(k))$$

5.1.2 Estimation de l'enveloppe spectrale : Le cepstre discret

L'estimation de l'enveloppe spectrale se fait par la méthode du cepstre discret [GR90],[OCM95]. L'enveloppe spectrale est calculée directement à partir de points du plan fréquence/amplitude

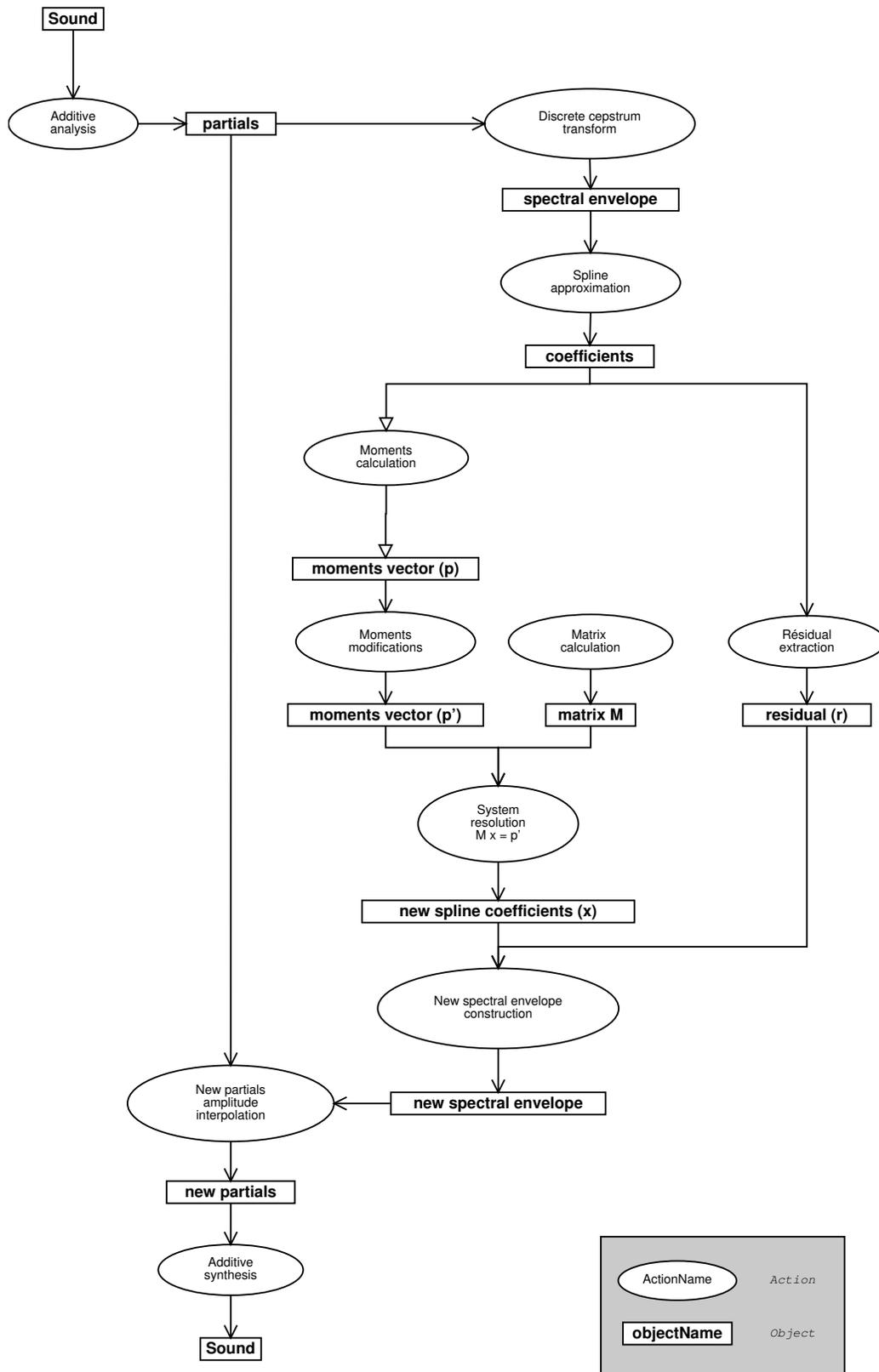


FIG. 5.1 – Les différentes étapes mises en jeu dans la transformation de haut niveau de l'enveloppe spectrale d'un son

des partiels. Typiquement, et dans le cas présent, les points sont les pics spectraux extraits par l'analyse additive.

L'enveloppe en log-amplitude est décrite par les coefficients cepstraux réel c_i sous la forme :

$$A_c(f) = c_0 + 2 \sum_{i=1}^p c_i \cos(2\pi f i)$$

Avec p l'ordre du cepstre.

La méthode consiste à déterminer les c_i qui minimisent l'erreur :

$$\epsilon = \sum_{k=1}^L w_k \| 20 \log_{10} a_k - A_c(f_k) \|^2$$

Certaines techniques sont utilisées pour améliorer les résultats comme la "Cloud method" ou l'utilisation de log-fréquence dans le haut du spectre [GR90].

Nous avons choisi d'utiliser le cepstre discret car il est très bien adapté à l'analyse additive. L'amplitude des partiels est bien approchée tandis que l'enveloppe reste relativement lisse dans la plupart des cas. Comme pour l'analyse additive le paramétrage se fait en grande partie à la main selon les caractéristiques de chaque signal.

5.1.3 Approximation et transformation de l'enveloppe par la méthode des B-splines

Voir chapitre 4

5.2 La synthèse

5.2.1 Reconstruction de l'enveloppe à partir des nouveaux coefficients

Le calcul de la nouvelle enveloppe spectrale se fait à partir des coefficients p'_i obtenus et du résiduel $\epsilon(f)$:

$$a'(f) = \sum_i p'_i B_i(f) + \epsilon(f)$$

5.2.2 La synthèse additive

Nous obtenons l'amplitude des partiels, dont la fréquence n'a pas changé au cours des transformations, à partir de l'enveloppe spectrale.

La synthèse se fait par addition de sinusoides en tenant compte des phases du signal d'origine.

Enfin, le résiduel de modélisation $r(n)$ issu de l'analyse additive peut-être ajouté au signal. Toutefois si les transformations effectuées sur le spectre harmonique sont importantes ou si de fortes modifications d'amplitudes au cours du temps ont été réalisées, la partie aléatoire peut avoir perdu son lien avec la partie harmonique. La superposition des deux peut amener à des sons peu cohérents. Un travail doit donc être effectué sur la partie aléatoire pour la transformer de manière cohérente à la partie harmonique.

Chapitre 6

Résultats

6.1 Présentation

6.1.1 Analyse et synthèse sans transformation

La modélisation utilisée permet de resynthétiser un son de manière parfaite si le résiduel est ajouté durant la reconstruction de l'enveloppe spectrale, si l'enveloppe spectrales est bien calculée par le cepstre discret et si la valeur des moments est conservée pour chaque trame. Nous entendons par parfaite que le résultat est le même que celui obtenu directement par analyse/resynthèse.

Lorsque le résiduel de modélisation n'est pas ajouté, les résultats dépendent de la régularité de l'enveloppe spectrale originale dans la zone des premiers harmoniques. En effet le modèle a tendance à lisser l'enveloppe, ce qui s'entend beaucoup sur les premiers harmoniques qui sont dans des bandes critiques différentes.

La figure 6.1 montre un spectre de contrebasse et un spectre de trompette. La trompette a une enveloppe très lisse, comme nous pouvons l'entendre (voir site ftp) les synthèses avec et sans résiduel sons quasiment indiscernables. En revanche, elles sont différentes dans le cas de la contrebasse dont les premiers harmoniques sont assez irréguliers.

Range de la transformation. Différence entre les descripteurs.

6.1.2 Changement de la valeur d'un des moments

Tout d'abord, nous devons vérifier que les modifications désirées (ou les non-modifications) sont bien celles que nous obtenons. En effet, les modifications sont appliquées à une enveloppe lissée, le fait d'ajouter le résiduel pourrait fausser les résultats et surtout, les valeurs de moments calculées sur l'enveloppe spectrale pourrait être différentes de celles calculées sur les harmoniques. Comme nous l'avions prévu, les valeurs des moments de l'enveloppe respectent parfaitement les valeurs cibles malgré l'ajout du résiduel, de plus les valeurs relatives des moments harmoniques sont presque justes.

Les figures 6.2 et 6.3, montrent un exemple de modification de la kurtosis sur un son de contrebasse. Nous avons changé la kurtosis de -1%, -5%, -10% et +10%. Nous consta-

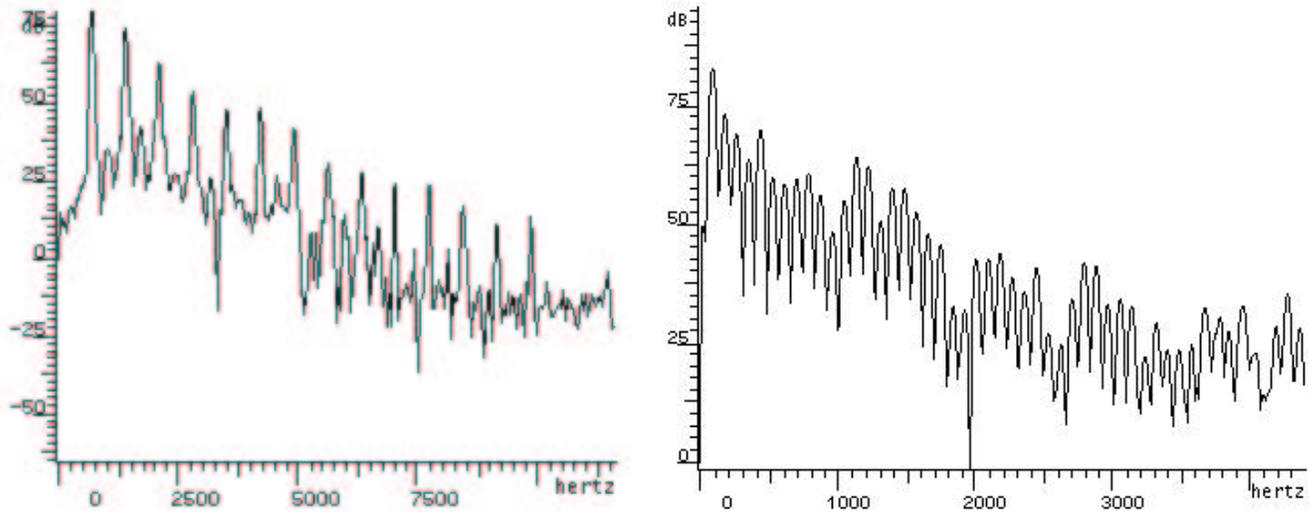


FIG. 6.1 – (a) Spectre de trompette (fa4) (b) Spectre de contrebasse (fa1)

tons que les autres moments harmoniques restent quasiment inchangés (Les courbes sont confondues).

D’autres part les intervalles possibles pour les descripteurs sont proches de ceux observés dans le paragraphe 3.2.

Etudions maintenant les types de modifications qu’opèrent sur l’enveloppe les changements de valeurs des moments.

Les figure 6.4 et 6.5 présente le résultat de la modification indépendante de chaque moment. Agir ainsi sur les moments entraîne de très grandes variations de l’enveloppe pour des petites variations de leurs valeurs. Ainsi, nous voyons sur la figure 6.4(a) qu’une modification de l’ordre de 2% de la valeur du centroïde entraîne déjà de fortes modifications en particulier en basse et haute fréquence. Le changement perceptif est frappant (voir exemples sonores).

Nous avons fait écouter à quelques personnes de manière informelle un petit ensemble de sons. Les observations recueillies sont résumées et commentées ci-dessous. Les modifications le long des axes définis par les quatre moments sont différentes perceptivement. Tout d’abord, l’axe un est l’axe le mieux défini. Cela n’a rien d’étonnant dans la mesure où il modifie le centroïde spectrale, comme attendu sa modification change la brillance. Ensuite l’axe 3 semble correctement définie. En effet, la brillance n’est pas modifiée lorsque l’on fait varier le descripteur lui correspondant. Ce qui semble changer est alors ce que nous pourrions qualifier de “résonateur”. En augmentant la valeurs de la skewness, le “résonateur” donne l’impression de se resserrer. Enfin les résultats des axes 2 et 4 sont plus mitigés. Des variations le long de ces axes ne changent pas d’attributs perceptifs précis. Les déplacements sur l’axe 2 donne une impression de passer d’un son de gorge à un son de nez, et change aussi la brillance du son. L’axe 4 est plus complexe, et les résultats

sonores varient considérablement selon le son de départ. Il est important de souligner que ce problème est valable pour tout les axes, de manière plus ou moins importante. Comme nous l'avons déjà dit, l'axe 1 est le mieux définie, c'est donc celui qui subit le moins les changements de son. Les axes 2 et 4 sont ceux sur lesquels une différence de son au départ donnera les résultats les moins cohérents, c'est à dire les sons dont les variations seront les plus difficiles à qualifier sous un même terme, comme nous pouvons qualifier les changements le long de l'axe 1 de variations de brillance.

Le problème d'interprétations des résultats qui se pose, peut s'illustrer en prenant des exemples plus précis de sons présentés en exemples. On constate, en effet, que certaines modifications sont moins perceptibles pour le son de trompette que pour le son de guitare. Nous voyons là les limites de notre modèle. Nous ne pouvons nous contenter que de décrire empiriquement les résultats des synthèses effectués et ce presque son par son. Les solutions envisageables pour rendre ce modèle consistant serait de définir des catégories de sons qui après synthèse donne le même type de variations perceptives. Plus précisément, nous pourrions regarder par exemple si les sons frottés donnent après transformation des résultats sonores cohérents et interprétables, et différents des sons soufflés qui donneraient alors d'autres résultats sonores cohérents entre eux. Pour valider ce travail scientifiquement et ne pas se contenter d'observations empiriques, il serait souhaitable de mettre en place des expériences de perception et cognition musicale. Cela permettrait de valider scientifiquement nos recherches et peut être d'aider à la définition et à la compréhension du timbre.

6.1.3 Lissage des descripteurs au cours du temps

Le lissage des descripteurs par splines est en général peut perceptible si le son ne contient pas de modulation (vibrato, trémolo). Les microvariations observée sur les valeurs des moments au cours du temps ne semble pas avoir de réelle influence perceptive. De plus dans le cas d'instruments non entretenus (cordes frappées ou pincées) il ne parait pas nécessaire de couper l'attaque et la décroissance en deux, une seule droite suffit.

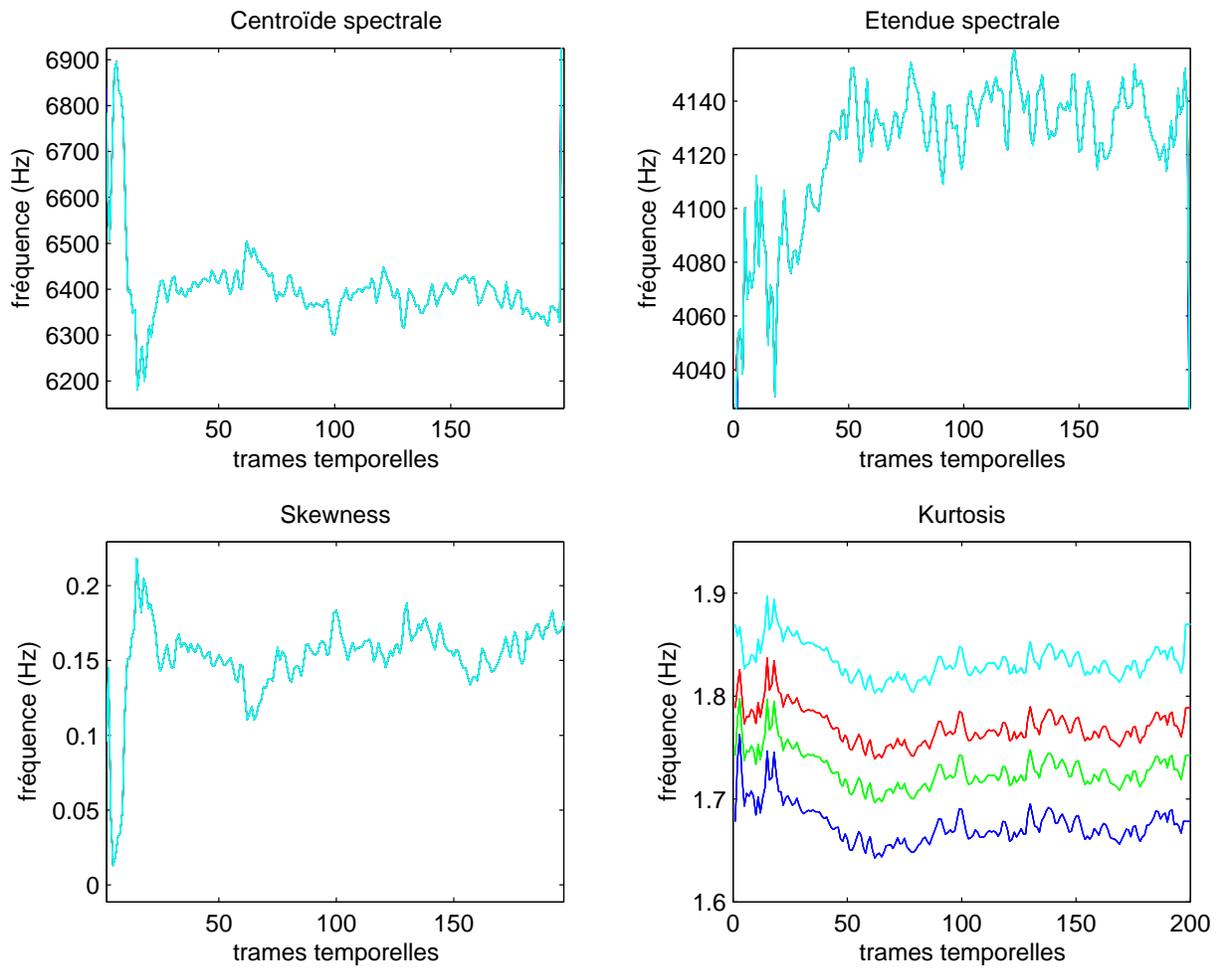


FIG. 6.2 – Evolution temporelle de la valeur des quatre premiers moments de l’enveloppe de la contrebasse pour quatre valeurs de la kurtosis

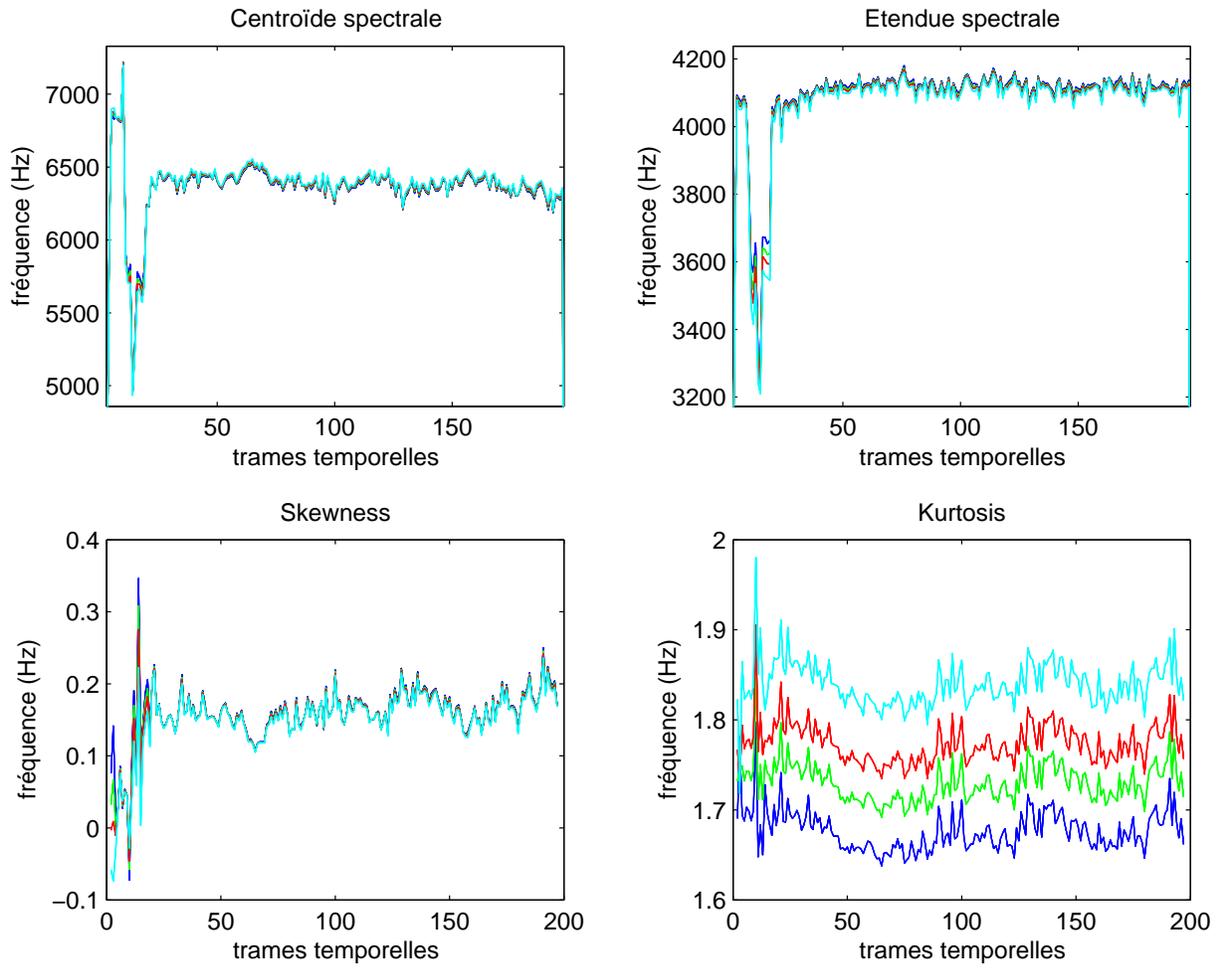


FIG. 6.3 – Evolution de la valeur des quatre premiers moments des partiels de la contrebasse pour quatre valeurs de la kurtosis

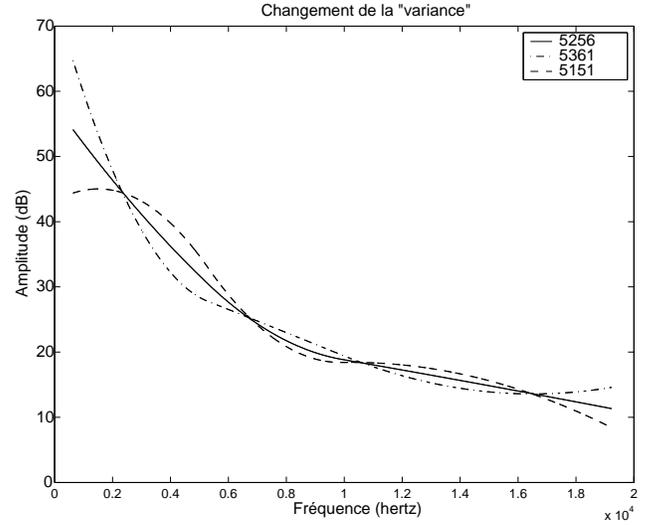
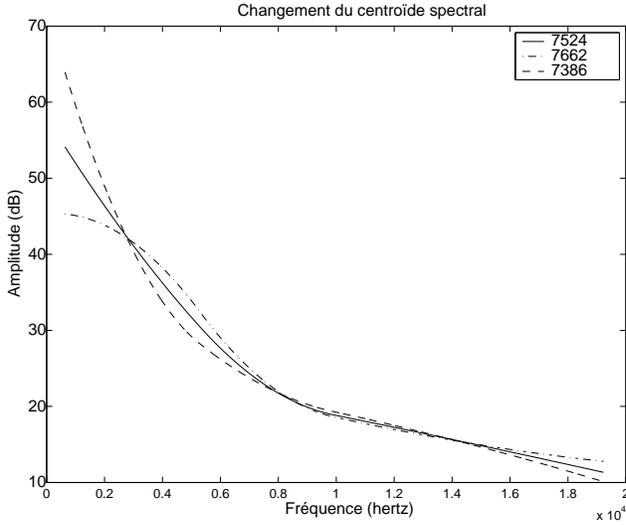


FIG. 6.4 – B-splines de degré 3 associés à (a) un noeud simple (b) un noeud triple

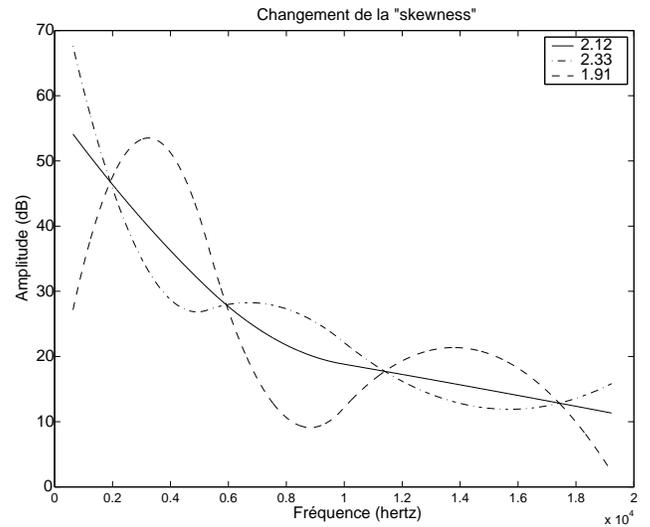
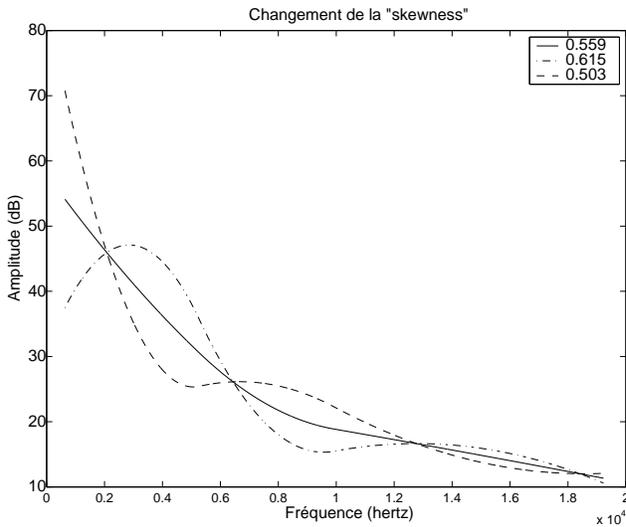


FIG. 6.5 – B-splines de degré 3 associés à (a) un noeud simple (b) un noeud triple

Chapitre 7

Conclusion

Dans cette étude, nous avons proposé un modèle permettant de manipuler un son à l'aide de paramètres de haut-niveau. Ces paramètres sont les quatre premiers moments et la largeur de l'enveloppe spectrale. La technique proposée permet une modification réellement indépendante de chaque paramètre. Un modèle simple de l'évolution temporelle des paramètres est aussi proposé.

Les transformations du son rendues possibles par ce modèle sont satisfaisantes. Néanmoins, elles sont parfois peu prévisibles. Nous avons observé que l'évolution du son le long des axes définis par les descripteurs n'est pas perceptivement constante et dépend du son traité, les résultats les plus probants étant obtenus pour des sons contenant beaucoup d'harmoniques. De plus l'évolution de la sonie au cours du temps peut-être fortement modifiée ce qui perturbe le jugement. Il semble donc nécessaire d'ajouter une contrainte sur la sonie au modèle. Ce dernier pourrait aussi être complété par l'ajout d'un modèle de comportement fréquentiel des harmonique ou en étendant la modélisation temporelles des descripteurs. Nous n'avons réalisé que des écoutes informelles sur un petit nombre de son. Nous ne pouvons alors pas apporter de validation scientifique au modèle proposé. Il serait intéressant de réaliser des expériences de perception et cognition musicale comme des tâches de catégorisation libre ou de jugement de dissimilarités sur un ensemble de sons issus de transformations, pour s'assurer d'une qualification unanime concernant les variations observés sur chaque axes.

Bibliographie

- [ABLS02] X. Amatriain, J. Bonada, A. Loscos, and X. Serra, *Spectral processing*, DAFX : Digital Audio Effects (Udo Zölzer, ed.), John Wiley and Sons, Ltd., 2002, pp. 373–438.
- [dB93] Carl de Boor, *B(asic)-spline basics*, Fundamental Developments of Computer-Aided Geometric Modeling (Les Piegl, ed.), Academic Press, 1993.
- [Don97] S. Donnadiou, *Représentation mentale du timbre des sons complexes et effets de contexte*, Thèse de doctorat, Université Paris V, Paris, 1997.
- [GR90] Thierry Galas and Xavier Rodet, *An Improved Cepstral Method for Deconvolution of Source-Filter Systems with Discrete Spectra : Application to Musical Sound Signals*, Proceedings of the International Computer Music Conference (ICMC) (Glasgow), September 1990.
- [HCM97a] C. Hourdin, G. Charbonneau, and T. Moussa, *A multidimensional scaling analysis of musical instruments' time-varying spectra*, Computer Music Journal **21,2** (1997), 56–68.
- [HCM97b] ———, *A sound-synthesis based on multidimensional scaling of spectra*, Computer Music Journal **21,2** (1997), 40–55.
- [HKCH97] J. Hajda, R. Kendall, E. Carterette, and M. Harshberger, *Methodological issues in timbre research*, Perception and cognition of music (I. Deliège and John Sloboda, eds.), Psychology Press, UK, 1997, pp. 253–306.
- [Jen99] K. Jensen, *Timbre models of musical sounds*, Ph.D. dissertation, Department of Computer Science, University of Copenhagen, 1999, Report no. 99/7.
- [MdCMW03] J. Marozeau, A. de Cheveigné, S. McAdams, and S. Winsberg, *The dependency of timbre on fundamental frequency*, Journal of the Acoustical Society of America **114** (2003), 2946–2957.
- [MQ86] R.J. McAulay and Th.F. Quatieri, *Speech analysis/synthesis based on a sinusoidal representation*, IEEE Trans. on Acoust., Speech and Signal Proc. **34** (1986), 744–754.
- [MWD⁺95] S. McAdams, S. Winsberg, S. Donnadiou, G. De Soete, and J. Krimphoff, *Perceptual scaling of synthesized musical timbres : Common dimensions, specificities, and latent subject classes*, Psychological Research **58** (1995), 177–192.

- [OCM95] J. Laroche O. Cappé and E. Moulines, *Regularized estimation of cepstrum envelope from discrete frequency points*, 1995.
- [Pee01] Geoffroy Peeters, *Modèles et modification du signal sonore adaptés à ses caractéristiques locales*, Ph.D. thesis, Université Paris 6, 2001.
- [Pee03a] Geoffroy Peeters, *Audio feature extractions*, Tech. report, Projet CUIDADO, 2003.
- [Pee03b] ———, *Automatic classification of large musical instrument databases using hierachical classifiers with inertia ratio maximization*, AES 115th Convention (New-York, USA), Octobre 2003.
- [PMH00] G. Peeters, S. McAdams, and P. Herrera, *Instrument description in the context of mpeg-7*, Proceedings of International Computer Music Conference 2000 (Berlin, Germany), 2000.
- [Rod97] Xavier Rodet, *Musical sound signal analysis/synthesis : Sinusoidal+residual and elementary waveform models*, 1997.
- [SB98] X. Serra and J. Bonada, *Sound transformations based on the sms high level attributes*, Proceedings of the 1st Digital Audio Effects Workshop (Barcelona, Spain), 1998.
- [SS90] X. Serra and J.O. Smith, *Spectral modeling synthesis : A sound analysis/synthesis system based on a deterministic plus stochastic decomposition.*, Computer Music Journal **14** (1990), no. 4, 12–24.
- [UAE93] M. Unser, A. Aldroubi, and M. Eden, *B-Spline signal processing : Part I—Theory*, IEEE Transactions on Signal Processing **41** (1993), no. 2, 821–833.