

PARCOURS MASTER 2

**ATiAM**

Parcours multi-mentions du Master Sciences et Technologies  
Université Pierre et Marie Curie - Paris 6  
en collaboration avec TELECOM ParisTech et l'Ircam



**ircam**  
**Centre**  
**Pompidou**

# Transformation des modulations et des gestes ornementaux dans les sons musicaux

Javier Contreras  
Université Pierre et Marie Curie, Paris VI

Sous la direction de : Axel Röbel  
Laboratoire IRCAM – Équipe analyse/synthèse

Mars 2009 – Juin 2009

# Remerciements

Arrivé à la fin de la rédaction du rapport de stage, je peux me poser, et réfléchir tranquillement qui je dois remercier. En faite, cela n'est pas facile ... il y a assez de monde, qu'il faudra compter avec une méthodologie appropriée. Bon, allons-y.

Je remercie beaucoup Axel Robël, mon maître de stage, pour sa patience et son aide dans le sujet que nous avons traité.

Je remercie aussi Lionel Feugere, sans qui je n'aurais pas pu écrire une seule page ! Et maintenant j'ai réussi ... ou pas ? Je te remercie aussi pour la musique qu'on a partagée !

Ensuite, je remercie beaucoup Camila y Jorge. Ils m'ont donné leur amitié et leur confiance depuis, et avant même, mon arrivée en France. Je les remercie beaucoup pour leur confiance en moi.

Je remercie aussi Batiste Bohelay, pour son aide lors de ce rapport, mais aussi pour les bons moments qu'on a partagés, pour les sujets de conversation, et pour son amitié. Pour les soirées, les cafés à l'IRCAM, et autant d'autres choses, merci beaucoup !

À Marion Courjaud, je la remercie pour sa confiance en moi, et son amitié. Merci beaucoup pour la maison ... et pour le vélo que j'utiliserai bientôt pour rentrer chez ... toi ! ( et merci pour le voyage à Poitiers ! )

Il est important pour moi de remercier mes amis Antonio Cardenas et Pierre Machart, pour leur amitié et leur confiance. Vous avez bougé, mais ne croyez pas vous avoir libéré de moi, je peux bouger moi aussi ...

À Gilles Chardon pour l'aide apportée dans ce rapport de stage, et pour avoir résolu pas mal de mes questions autour du traitement du signal.

Je remercie mes camarades de stage : Arnaud Dessen, Julien Moummé et Philippe Esling. Je n'oublie pas non plus John Mandereau, à qui je dois de bon moments en vélo et un sincère remerciement pour l'aide dans l'obscur domaine des mathématiques.

Je dois aussi remercier Amanda Sepulveda. Sans elle, je n'aurais jamais réussi une entreprise aussi dure et difficile comme partir à l'étranger, apprendre une langue, prendre des décisions, et réussir des examens, tout en ne comprenant rien de ce qui se passe autour de moi. Je t'embrasse depuis la France !

Je remercie aussi mes grands amis chiliens, sans lesquels je n'aurais pas trouvé repos ( ni fêtes ... ). J'envoie une dédicace spéciale à mes amis de la Cité Universitaire. On était plus ou moins dans la même situation. Je ne me rappelle pas quand est-ce que j'ai regardé dans les yeux de quelqu'un et que j'ai trouvé autant de compréhension comme chez mes voisins. Particulièrement, je tiens à remercier Aline pour son appui inconditionnel. J'embrasse aussi Paulette !

Je dois remercier le conseil régional d'Ile-de-France, pour le financement de mon projet de formation.

Enfin, je remercie enfin, toute la génération ATIAM 0809 pour les bons moments qu'on a passés ensemble.

# Résumé

Supposons que la transposition des fréquences d'un son avec vibrato est faite de manière à retirer le vibrato : à la différence du son original, le son transformé ne présente pas d'évolution temporelle des fréquences de ses partiels. Cependant, les amplitudes de ces partiels révèlent des minimas et maximas dans le spectre du signal dûs aux caractéristiques physiques de l'instrument et à cause du vibrato lui-même. À partir d'un modèle *source/filtre*, le signal est alors transposé en gardant son enveloppe spectrale afin d'éliminer ces variations d'amplitude. Suivant cette méthode, les variations d'amplitude ne sont pas complètement lissées, ce qui indique la présence d'une modulation de l'enveloppe spectrale. Durant un stage réalisé à l'IRCAM, une étape de filtrage a été proposée consistant à moyenniser les enveloppes spectrales à partir des enveloppes passées et en prenant en compte l'évolution instantanée de la fréquence du vibrato. Nous avons dirigé notre attention dans certains artefacts audibles qui ont été rapportés lors de ce filtrage.

Deux hypothèses ont été proposées lors de ce stage de recherche, afin d'expliquer ces artefacts. D'une part, la transposition du son a transposé des composantes fréquentielles qui n'ont pas été modulées avec le vibrato alors qu'ils sont modulés lors de la transposition. D'autre part, le filtre développé à partir de l'enveloppe spectrale a produit des variations d'amplitude des fréquences, qui n'appartiennent pas aux composantes harmoniques du son, mais qui sont proches. Dans ce cas, il pourrait se produire une modification de l'évolution temporelle de fréquence des partiels, ceci ayant pour conséquence la modification de la trajectoire des partiels.

Une de nos hypothèses est que la modulation d'amplitude dans le son avec vibrato peut être différente entre le bruit et la composante harmonique. Ainsi, on propose un modèle dit *sinusoïde plus bruit* pour aborder d'un côté la composante harmonique et de l'autre le bruit. Pour la composante harmonique, on a mis en œuvre un filtre à partir des amplitudes de partiels. Après avoir fait la transposition qui modifie l'étendue du vibrato, les amplitudes des fréquences associées aux partiels sont prises en compte pour obtenir un filtre qui varie au cours du temps. Le filtre, étant destiné aux composantes harmoniques, n'agit pas sur la composante du bruit. Les résultats montrent une diminution de variation d'amplitude des partiels, et il est tout de même mis en évidence l'existence d'une modulation d'amplitude dans les composantes non harmoniques du son transposé et filtré. On moyenne l'enveloppe spectrale de la composante du bruit à partir d'un filtre réalisé avec des coefficients LPCs. En additionnant les parties filtrées harmoniques et non harmoniques, on obtient le résultat d'un premier modèle de resynthèse.

D'autre part, puisqu'il peut y avoir des composantes du signal qui ne sont pas modulées lors du vibrato, on s'est interrogé sur la validité de la méthode consistant à transposer toutes les composantes du son vibrato pour modifier l'amplitude de la modulation en fréquence. Est-ce qu'on peut améliorer le modèle en appliquant la transposition uniquement aux fréquences des partiels, c'est à dire aux composantes qui sont modulées lors du vibrato ? Une deuxième méthode de resynthèse est alors proposée. À partir d'une analyse des partiels du signal, on reconstruit les composantes harmoniques et on transpose ces composantes. Les amplitudes de chaque partiel sont ensuite filtrées, ainsi que la composante bruitée. Finalement, on mélange les deux résultats.

À la fin du stage, on a commencé à aborder la question de la reconnaissance automatique du vibrato et de l'estimation de ses paramètres. On n'a cependant pas eu le temps d'arriver à des résultats.

Les résultats obtenus pour les filtres de lissage d'amplitude lors de la modification de l'étendue du vibrato ont été comparés. Le filtrage des deux composantes harmoniques et bruité, montrent une minoration des artefacts audibles par rapport à la méthode du filtrage du son complet. D'autre part, d'après des propos informels de l'auteur sur des exemples sonores, la séparation en composantes harmonique et bruité pour réaliser la transposition ne s'avère pas comme nécessaire.

# Table des matières

<b>Remerciements</b> .....	<b>2</b>
<b>Résumé</b> .....	<b>4</b>
<b>1. Introduction</b> .....	<b>7</b>
<b>2. Objectifs</b> .....	<b>9</b>
<b>3. État de l'art</b> .....	<b>10</b>
3.1. La morphologie du vibrato.....	10
3.2. Le modèle source/filtre.....	13
3.3. Détection de la fréquence fondamentale $f_0$ .....	14
3.4. Obtention des composantes lente et rapide de la trajectoire de.....	15
3.5. Modification de l'étendue du vibrato.....	17
3.6. Enveloppe spectrale.....	18
3.6.1. Estimation de l'enveloppe spectrale : « True envelope ».....	19
3.6.2. Filtre d'enveloppe spectrale.....	20
<b>4. Démarche réalisé</b> .....	<b>24</b>
4.1. Modèle sinusoïde plus bruit.....	25
4.1.1. Analyse de la composante sinusoïdal.....	26
4.1.2. Analyse du signal résiduel.....	27
4.1.3. Enveloppe spectrale du signal résiduel.....	28
4.2. Filtre des composantes harmoniques.....	28
4.5. Reconnaissance automatique du vibrato.....	30
4.4.1. Fonction d'observation.....	32
<b>5. Résultats</b> .....	<b>34</b>
<b>6. Conclusion et perspectives</b> .....	<b>39</b>
<b>Bibliographie</b> .....	<b>41</b>

# 1. Introduction

Aujourd'hui les transformations des sons musicaux permettent des résultats très naturels. Les algorithmes de transformation existants négligent par contre les gestes musicaux et les modulations. Celles-ci sont modifiées et elles perdent alors leur sens dans le son traité. Le but du présent stage a consisté à participer au développement des méthodes de transformation qui permettent de préserver le sens musical des modulations. On s'intéressera au cas du vibrato.

Différentes modifications d'un son avec vibrato sont envisageables. Par exemple, allonger ou raccourcir la durée du son sans changer la fréquence du vibrato. On peut aussi souhaiter de modifier la fréquence du vibrato sans modifier la durée du son, ou de modifier l'amplitude de la modulation de la fréquence du vibrato.

Dans le cadre d'un stage récemment achevé à l'IRCAM, Maller [11] a étudié différentes transformations du son contenant du vibrato. Son travail est basé sur le vocoder de phase et le modèle du signal *source/filtre*. Son attention s'est centrée sur le vibrato. Notamment, il a développé des algorithmes permettant la transformation indépendante de la durée du son et de la fréquence du vibrato. Un des problèmes centraux pour ces transformations est le fait que les modulations induisent des variations du timbre et que celles-ci doivent être prises en compte pour que les sons transformés restent convaincants.

Rappelons-nous que la dilatation/contraction temporelle d'un fichier audio dans le vocoder de phase est réalisée en changeant l'espacement entre les fenêtres de resynthèse pendant l'étape de overlap/add [16]. Ceci ne produit pas un changement de la hauteur du son, et puisque le vibrato est une modulation de fréquence au cours du temps, la dilatation/contraction du son modifiera la fréquence de cette modulation et non son amplitude. Maller propose d'utiliser cette dilatation/contraction pour modifier la fréquence du vibrato.

D'autre part, Maller propose l'allongement de la zone contenant le vibrato ( sans changement de sa fréquence ) en insérant les échantillons compris dans une période du vibrato déjà passée dans le fichier d'origine. C'est-à-dire, on fait une copie d'une période de vibrato et on la colle où on a besoin d'allonger le son. Inversement, il est possible de diminuer la longueur des zones du vibrato en enlevant des échantillons correspondant à la période du vibrato.

Finalement, la méthodologie utilisée par Maller pour effectuer une modification de l'amplitude de la modulation de fréquence du vibrato consiste à appliquer une transposition de hauteur sur chaque fenêtre d'analyse du son. Parce qu'on peut décrire l'instrument comme une source harmonique filtrée par des modes de résonance de son corps, donc rattaché au modèle *source/filtre*, une transposition en gardant l'enveloppe spectrale est choisie. Malgré que, de cette façon, les modulations d'amplitude des partiels du son sont diminuées, il en reste encore après le traitement. Maller propose donc, une étape de filtrage après avoir fait la modification de l'amplitude de modulation de fréquence du son avec vibrato. Cette étape consiste à obtenir un filtre qui varie au cours du temps à partir de l'enveloppe spectrale du son original. Le but est de moyenner l'évolution dans le temps des amplitudes du son dont on a modifié l'amplitude de la modulation de fréquence du vibrato.

La présente recherche est encadrée dans la continuité des recherches amenées dans l'équipe

d'analyse/synthèse de l'IRCAM, et notamment dans la continuité et la amélioration des résultats obtenus par Maller. Dans ce sens, nous croyons qu'il est possible qu'une généralisation de la méthode utilisée pour modifier l'amplitude de la modulation de fréquence du vibrato à la méthode de modification de la modulation d'amplitude. Cette généralisation simplifierait l'étape de filtrage post-modification de l'amplitude du fréquence du vibrato. D'autre part, la modification de l'amplitude de la modulation de fréquence et le facteur d'amplification/atténuation de la modulation d'amplitude seraient liés de façon naturelle.

L'étude de l'évolution de l'amplitude des partiels nous permettrait de modéliser le vibrato, ou plus précisément le vibrato produit par un couple instrument/instrumentiste. Il nous semble que ce modèle permettrait la reproduction et la modification totale ou partielle du vibrato dans un son. Dans le présent travail, nous ne chercherons pas à modéliser le vibrato, mais à introduire de nouvelles études et propositions autour de ce sujet.

La généralisation qu'on proposera ici prend en compte l'évolution dans le temps de l'amplitude de chaque partiel du son. À partir du modèle *sinusoïde plus bruit* nous pouvons estimer la fréquence et l'amplitude pour chaque partiel du son au cours du temps.

Après cette brève introduction nous présentons les objectifs du stage. Ensuite nous parcourons, dans le chapitre 3 « État de l'art », les concepts qu'on a besoin d'introduire pour le développement de nos idées. Ce chapitre commence par une description du vibrato, où on définit aussi les paramètres qui caractérisent le vibrato. On présente brièvement le modèle *source/filtre* sur lequel est basée la méthode existante pour la modification du vibrato. L'extraction de la fréquence fondamentale et la façon dont on modifie l'amplitude de la modulation de fréquence du vibrato est suivi par une explication de l'estimation de l'enveloppe spectrale et de la manière dont il est utilisé pour estimer les coefficients du filtre pour enlever la modulation d'amplitude. Dans le chapitre 4 « Démarche réalisée », nous exposons les idées et la façon par laquelle on les a réalisées pendant le stage. Il commence par une description du modèle *sinusoïde plus bruit*. Ensuite nous donnons la généralisation de la méthode de modification de l'amplitude de la modulation de fréquence à la modification de l'amplitude des partiels. Pour compléter ce traitement nous présentons une étape du filtrage du signal résiduel. Ce chapitre s'achève en décrivant ce qu'on a pu implémenter dans la détection du vibrato et de l'estimation automatique de ses paramètres.

Finalement, les résultats obtenus sont exposés dans le chapitre 5 « Résultats ». Nous proposons dans le chapitre 6 « Conclusion et perspectives » une discussion des résultats et donnons des perspectives pour continuer la recherche dans le traitement des sons avec vibrato.

## 2. Objectifs

Les objectifs du stage peuvent être résumés par les points suivants:

- S'inscrire dans la continuité de la recherche sur le vibrato menée par l'équipe d'analyse/synthèse. On s'intéressera, notamment, au perfectionnement et à la généralisation des recherches menées par Maller [11] dans la cadre d'un stage précédent au sein de l'équipe analyse/synthèse.
- Chercher des modèles qui permettent la modification indépendante des caractéristiques du vibrato et du son qui le contient.
- Étudier le signal d'un son contenant du vibrato et différents modèles pour le représenter. On étudiera notamment le modèle *source/filtre* et le modèle *sinusoïde plus bruit*.
- Comprendre, à partir d'un point de vue acoustique mais non exhaustif, le comportement du signal du son avec vibrato et la façon dont les différents facteurs qui le composent interagissent entre eux.
- Chercher un modèle de segmentation automatique qui permette de trouver les endroits où le vibrato existe et qui puisse estimer ses paramètres.

Pour accomplir les objectifs de notre recherche, nous devons délimiter notre démarche d'un part en fonction des attentes de l'équipe accueillant et de l'autre en limitant notre approche dans de conditions particulière. Dans ce sens nous considérons un signal contenant du vibrato comme un signal enregistré dans un cas idéal où cet enregistrement ne contient pas de réverbération et où l'enregistrement a un bon rapport signal sur bruit. De plus, nous étudierons le cas du vibrato dans des signaux instrumentaux considérés comme harmoniques.

## 3. État de l'art

Dans un premier temps, on s'intéressera à la définition du vibrato et à la reconnaissance de ses caractéristiques. À partir d'une étude bibliographique on décrira les caractéristiques du vibrato et les valeurs typiques des estimateurs de ses caractéristiques. Une fois décrite la morphologie du vibrato, on verra comment on encadre un signal contenant du vibrato dans le modèle *source/filtre* et de quelle façon on tire des avantages en utilisant ce modèle dans les transformations de sons.

Étant donné un son avec vibrato, l'amplitude de la variation en fréquences au cours du temps peut être estimée en regardant l'évolution de la fréquence fondamentale. La deuxième partie de ce chapitre expose les enjeux d'une recherche de fréquence fondamentale dans un son afin de déterminer sa meilleure représentation. On va appeler par la suite *trajectoire de la fréquence fondamentale* l'évolution de la fréquence fondamentale au cours du temps.

Ensuite, nous serons intéressés à discerner la partie de la trajectoire de la fréquence fondamentale qui est la note – la hauteur jouée – de la partie propre au mode de jeu qui est le vibrato. On exposera une manière de décomposer cette trajectoire en deux composantes : une composante lente qui est associée à la note, la hauteur du son, où encore à la mélodie jouée ; et une composante rapide qui est associée au vibrato.

Après, nous expliquerons en détail comment à partir des composantes lente et rapide on obtient un ensemble de coefficients qui nous permettront de modifier, à l'aide du vocoder de phase, l'amplitude de la modulation de fréquence du vibrato. On choisira un facteur de transposition  $\alpha$  qui permettra de régler la quantité de modification introduite pendant cette modification. Cette modification est faite à partir d'une transposition dans le vocoder de phase. À chaque fenêtre d'analyse est associé un coefficient de transposition de la hauteur du son. Ce coefficient est trouvé, et on expliquera par la suite comment on va procéder, à partir de la composante lente et de la composante rapide de la trajectoire de la fréquence fondamentale.

À partir de ce traitement on verra qu'il existe une modulation d'amplitude dans le signal produit par le vibrato. Nous allons étudier une étape de filtrage qui peut être utilisée pour réduire cette modulation. On commencera par introduire la méthode utilisée pour connaître l'enveloppe spectrale du son car c'est le cœur de cette étape de filtrage. Le filtre qui a été développé par Maller sera exposé par la suite.

### 3.1. La morphologie du vibrato

Le vibrato est un mode de jeu ornemental utilisé par les musiciens et les compositeurs depuis longtemps. Le mot *vibrato* ne désigne pas la façon dont cet effet est produit mais désigne la façon dont la note jouée est perçue. D'un point de vue perceptif, le vibrato est une petite variation de la hauteur d'une note réalisée de façon rapide et périodique. Malgré le changement de hauteur, le vibrato n'est pas perçu comme un changement effectif de la note mais comme un effet ornemental de la note jouée.

Dépendant de l'instrument, le vibrato est produit de différentes façons. Chez le violon, le vibrato est fait à partir d'un mouvement rythmique de doigt posé sur la corde. Le mouvement, étant réalisé de l'avant vers l'arrière suivant la largeur de la corde, produira un changement de la largeur effective. En conséquence, la hauteur de la note jouée oscillera autour d'une note centrale [12]. Chez le saxophone, il existe deux façons de produire le vibrato [7]. La première est réalisée par un mouvement de la mâchoire, et la deuxième est produite en faisant varier la pression de l'air contenu dans la cavité de la bouche à l'aide de la langue. Chez le hautbois, le vibrato peut être effectué de différentes façons : on peut utiliser la gorge, le diaphragme, les lèvres ou la mâchoire [23].

D'un point de vue du signal produit par le vibrato, l'effet correspond à une variation quasi périodique de la fréquence fondamentale autour d'une valeur centrale [19]. Le mouvement de la fréquence fondamentale peut être accompagné de fluctuations de l'amplitude et du timbre du son [23].

On définit par la suite l'étendue du vibrato comme l'amplitude de la modulation en fréquence ( en anglais « vibrato extent » ) , et la fréquence du vibrato comme la fréquence de modulation de la fréquence fondamentale du son ( en anglais « vibrato rate » ) lors de l'existence du vibrato [7].

En ce qui concerne la fréquence du vibrato, il existe des différences parmi les instruments mais on observe une valeur moyenne autour de 6 Hz. La fréquence du vibrato a été mesurée entre 5.6 et 6.7 Hz [22] sur un échantillon de plusieurs chanteurs. La fréquence moyenne pour le vibrato d'un violon a été mesurée entre 5.9 Hz [12] et 6.3 Hz [22], et entre 5.5 et 8 Hz pour la famille des instruments à cordes [23].

D'autre part, en comparant les instruments, la différence de l'étendue du vibrato pour chaque instrument est plus importante que celle de la fréquence du vibrato. Pour la voix, Seashore ( 1936, cité d'après [22] ) a trouvé comme moyenne 100 cents d'amplitude de modulation, alors que pour les cordes on trouve des étendues comprises entre 20 et 25 cents [23].

L'étendue et la fréquence du vibrato peuvent être accompagnées d'une variation d'amplitude du signal. Dans la perspective du modèle *source/filtre*, cette modulation d'amplitude est expliquée par le passage des composantes harmoniques de la source à travers le filtre. Cependant, il existe d'autres facteurs qui peuvent produire cette modulation, citons par exemple la variation de pression du souffle dans la flute [5].

Dans certains cas, l'étendue du vibrato est plus importante que la modulation d'amplitude du signal ( par exemple dans le cas de la voix ) . Inversement, dans d'autre cas la variation d'étendue est moins importante que la modulation d'amplitude ( par exemple chez les saxophones ) [7]. Une étude perceptive avec simulations des sons de violon avec vibrato [12] a permis de conclure que la modulation de fréquence n'avait pas beaucoup d'importance quant à la discrimination d'un son avec vibrato, alors que l'absence de modulation d'amplitude cause la détérioration de la qualité du son.

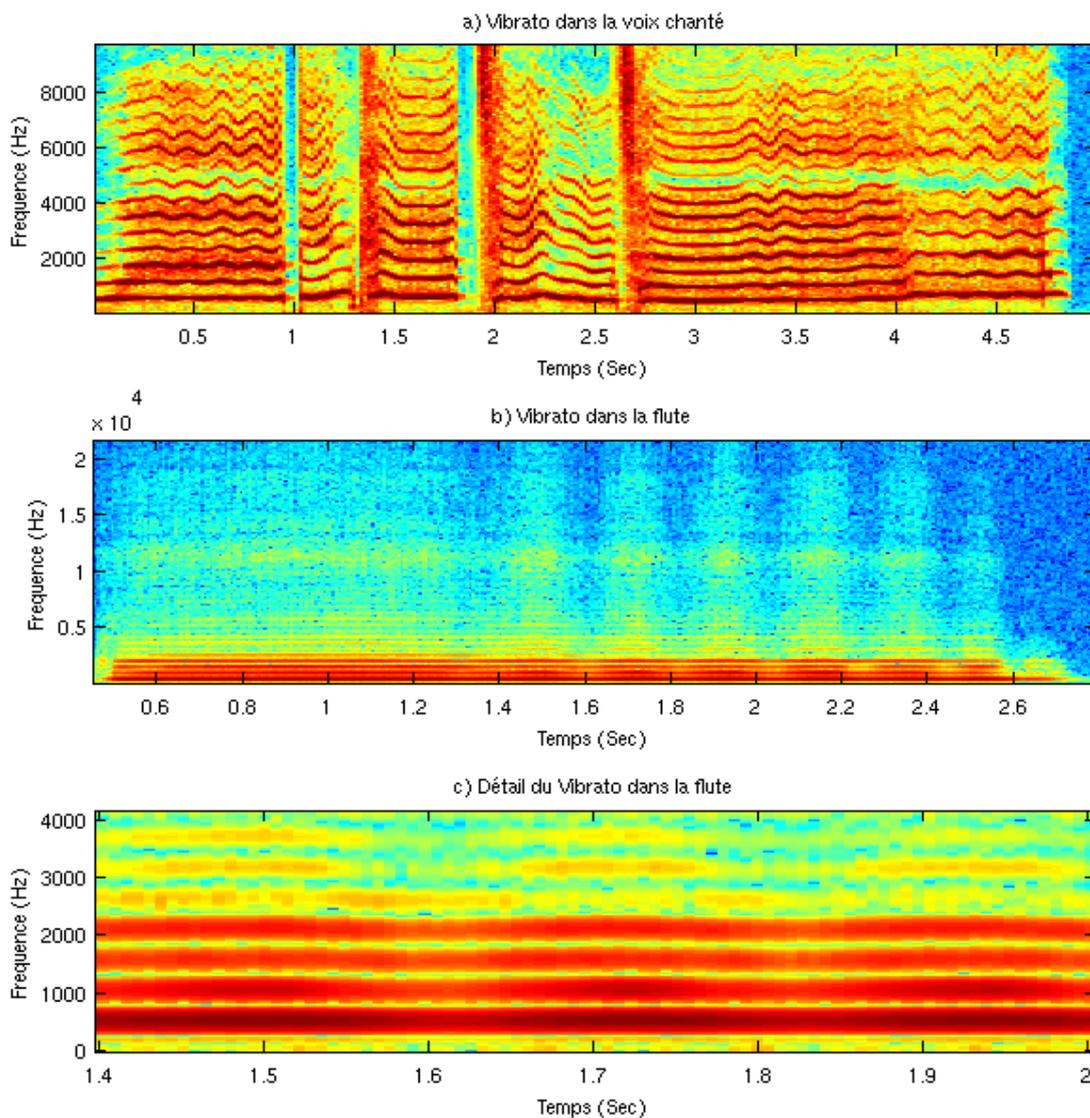
Finalement, la modulation de fréquence et la modulation d'amplitude du signal peuvent être accompagnées d'une modulation de l'enveloppe spectrale [24]. Par exemple, dans les bois et les cuivres, la modulation du flux d'air produit un enrichissement du spectre dû aux non-linéarités dans le tuyau. Quand le souffle est plus intense l'enrichissement spectral apparaît, et il disparaît quand le souffle est moins fort.

Finalement, les auteurs de [24] ont proposé que les variations d'amplitudes des partiels lors du vibrato peuvent être décrites sous l'angle d'une liaison de hystérésis. L'évolution de l'amplitude de

chaque partiel suivrait une forme cyclique où la symétrie entre la montée et la descente serait un cas particulier.

Si on se situe au niveau de la mélodie, on peut voir plusieurs notes, chacune pouvant contenir des gestes ornementaux. De ce point de vue, le vibrato peut être vu comme une composante du signal qui varie rapidement par rapport à la mélodie. Dans ce cas-là, il serait possible d'extraire d'une part une variation évoluant lentement qui contiendrait la mélodie, et d'autre part une variation rapide qui contiendrait le vibrato.

Pour trouver ces deux composantes on a besoin d'un modèle du signal qui nous permet de prendre en compte l'évolution de la hauteur d'un son instrumental. Par la suite, on verra comment nous pouvons utiliser une analyse de fréquence fondamentale  $f_0(t)$  pour trouver la variation lente  $f_{dc}(t)$ , et la variation rapide  $f_{ac}(t)$ .



*Illustration 1: Le sonogramme du vibrato dans la voix chanté montre une large étendue du vibrato (a), à la différence du vibrato dans la flute (c). Dans le cas de la flute, on distingue clairement une modulation d'amplitude des partiels (c) et du bruit (b).*

### 3.2. Le modèle source/filtre

Dans le modèle de *synthèse additive*, la source est décrite comme une addition de sinusoïdes dont chacune est multipliée par un coefficient. Une extension de ce modèle est le modèle *source/filtre*. Ceci définit une source comme un signal harmonique de spectre plat qui est ensuite modifiée par un filtre. Ce filtre amplifie ou atténue les amplitudes des partiels de la source pour reconstituer les amplitudes du son d'origine.

Ce modèle a une importante corrélation physique avec les instruments de musique. En tant que source, on modélise les systèmes d'excitation des instruments, par exemple : des cordes vocales, une corde, une anche, ou des lèvres. En ce qui concerne le filtre, on décrira leurs résonateurs, par exemple : le conduit vocal pour la voix, la caisse de résonance pour la guitare, et le tuyau dans le cas du saxophone et de la trompète.

Dans ce modèle, les variations d'amplitude trouvées pour les fréquences du partiels du son avec vibrato, par exemple dans le violon [12], peuvent être expliquées par l'interaction des modes de résonance de la caisse de l'instrument et du passage des partiels par ces résonances lors de la modulation en fréquence. De ce fait, on voit bien que cette variation n'est pas la même pour tous les partiels du son : d'un côté les modulations n'arrivent pas forcément en même temps, et de l'autre, la forme de la modulation diffère en fonction de la quantité et la distance entre les fréquences de résonances.

Pour utiliser ce modèle, on a besoin d'estimer les paramètres caractérisant la source et le filtre. La source peut être reconstruite à partir d'une analyse harmonique réalisée dans le domaine spectral. Pour le modèle du filtre, différentes approches sont proposées dans la littérature, parmi lesquels on trouve la « Linear Predictive Coding » (LPC) et la « true envelope » (une évaluation de ces deux méthodes est donnée en [15]).

L'intérêt de ce modèle est qu'on peut transposer dans le domaine fréquentiel uniquement la source. Après cette transposition, l'enveloppe spectrale est rétabli en garantissant une importante ressemblance entre l'enveloppe du son d'origine et celle du son transposé. Le but est de conserver le timbre d'un son après la transposition.

En termes pratiques, ce modèle est estimé à l'aide du vocoder de phase. À partir d'une analyse de fréquence fondamentale on trouvera les harmoniques et l'enveloppe spectrale du son pour chaque fenêtre d'analyse. Pour la modification de l'étendue du vibrato, une transposition est opérée en gardant l'enveloppe spectrale de la fenêtre originale. Pendant la resynthèse, la méthode de overlap/add avec recouvrement entre les fenêtres vont garantir la reconstruction du signal traité.

Le problème d'opérer une transposition en gardant l'enveloppe spectrale du son est que, comme on a déjà parlé, le vibrato peut produire des modulations d'enveloppe spectrale que seront remises lors de la transposition. Ainsi, pour enlever le vibrato, si on se contente de transposer la source dans le modèle qu'on vient de décrire, la variation du spectre sera toujours présente. C'est pour cela que Maller [11], a proposé une étape de filtrage du signal transposé avec un filtre variant dans le temps. La façon dont les coefficients du filtre sont obtenus est expliquée à la fin de ce chapitre.

Dans ce qui suit, on décrira les différentes analyses et processus qui font partie de la modification de l'étendue du vibrato dans le modèle *source/filtre* et le vocoder de phase. On commencera par le cœur du traitement qui est la détection de la fréquence fondamentale et de sa décomposition en une composante rapide et une composante lente. Ensuite, on parlera de la façon dont on obtient les

coefficients de transposition par fenêtre d'analyse. Finalement, on décrira le fonctionnement du filtre pour enlever la modulation d'amplitude et la modulation d'enveloppe spectrale.

### **3.3. Détection de la fréquence fondamentale $f_0$**

La fréquence fondamentale  $f_0$  peut être définie comme le plus grand diviseur commun des fréquences des harmoniques du son [21]. D'un point de vue musicale, la perception de la hauteur d'un son harmonique est liée à la fréquence fondamentale, et elle est généralement la fréquence la plus basse générée par l'instrument.

La variation de la fréquence fondamentale produit par le musicien pendant l'exécution d'une note avec vibrato, peut être observée à partir d'une analyse de fréquence fondamentale. Dans [6] on retrouve une perspective historique des méthodes et les modèles actuellement utilisées pour la détection de la  $f_0$ .

Dans notre cas, on va trouver la fréquence fondamentale à l'aide du logiciel SuperVP. L'algorithme de détection de  $f_0$  dans ce logiciel est fait dans le domaine fréquentiel. Après d'une étape de fenêtrage du signal d'entrée la TFD est obtenue pour chaque fenêtre. Ensuite, une détection de pics est réalisée et deux critères sont utilisés pour discerner quelle est leur fréquence fondamentale commune. Ces critères sont expliqués par la suite :

a ) « Spectral Match Weight »

Pour chaque fréquence fondamentale hypothétique est faite une somme de l'énergie contenue dans les pics qui appartiennent aux composantes harmoniques supérieures qui correspondent à cette fréquence. La fréquence fondamentale hypothétique qui explique la plus grande quantité d'énergie est alors choisie comme la fréquence fondamentale.

Le problème de cet estimation est qu'il peut exister un pic maximal plus bas que la fréquence fondamentale pouvant expliquer les mêmes partiels que la vrai  $f_0$  plus des faux partiels appartenant au bruit du signal. Pour résoudre ce problème, ce critère est accompagné d'un deuxième critère qu'on explique ci-dessous.

b ) « Envelope Weight » :

On assume l'enveloppe spectrale des instruments de musique comme lissée, c'est-à-dire sans changement abrupt d'amplitude entre les fréquences. Dans le cas où la fréquence fondamentale estimée serait plus basse que la vrai fréquence fondamentale, l'estimation expliquera aussi des faux partiels qui seront entre les fréquences des partiels réels. L'enveloppe spectrale de cette estimation sera alors coupée par des chutes d'amplitudes appartenant aux pics du bruit entre les amplitudes des vrais partiels. Les candidates pour la fréquence fondamentale générant des variations rapides sont évaluées moins bonnes et sont alors rejetées.

D'après l'analyse de  $f_0$  dans SuperVP, on caractérise la fréquence fondamentale avec trois paramètres: la fréquence, l'amplitude réelle de la fréquence et un indicateur de la périodicité de la fréquence estimée. Cet analyse de périodicité nous donne un aperçu de la fiabilité de l'estimation de la fréquence fondamentale. Si la valeur de cet indicateur est petite, la fréquence fondamentale peut indiquer une erreur d'estimation, alors que si la valeur est grande l'estimation est certaine.

Pendant l'étape d'obtention de la TFD, la taille de la fenêtre doit être choisie par rapport à la fréquence fondamentale et à la fréquence du vibrato. Une taille de fenêtre trop longue aura une bonne résolution spectrale mais une mauvaise résolution temporelle, et dans ce cas là, on risque de ne pas voir le changement de fréquence lors du vibrato. D'autre part, une taille de fenêtre trop petite ne permet pas de résoudre les partiels car les lobes deviennent plus larges et on va avoir une évaluation de la fréquence fondamentale moins bonne.

On va prendre une taille de fenêtre  $M_{f_0}$  échantillons pour l'analyse de fréquence fondamentale ( 25 millisecondes ) . Aussi, on va prendre une fenêtre d'analyse chaque  $M_{f_0}/O_{f_0}$  échantillons,  $O_{f_0}$  étant le chevauchement entre fenêtres. La fréquence d'échantillonnage de la trajectoire  $f_0$  sera donc :

$$F_{ef_0} = O_{f_0} / M_{f_0} \quad (1)$$

À partir de la trajectoire de la fréquence fondamentale, on va trouver ses composantes lente et rapide. Ci-dessous, nous allons expliquer de quelle façon ceci est fait.

### **3.4. Obtention des composantes lente et rapide de la trajectoire de $f_0$**

D'après le point de vue du signal, le vibrato peut être vu comme une variation quasi-périodique de la fréquence fondamentale dont sa fréquence de variation se trouve autour de 6 Hz. À partir de cette affirmation, la trajectoire de la fréquence fondamentale peut être décrite comme l'addition d'une composante lente  $f_{dc}$  et d'une composante rapide  $f_{ac}$  . La première serait estimée comme une moyenne locale de la fréquence fondamentale, alors que la deuxième comme la différence entre la fréquence fondamentale et la moyenne locale estimée. D'après un point de vue musical, la composante lente  $f_{dc}$  pourrait être interprétée comme l'intonation [1].

D'une façon générale on peut décrire la trajectoire de la fréquence fondamentale comme ce qui suit :

$$f_0(n) = f_{dc}(n) + f_{ac}(n) \quad (2)$$

où, pour un vibrato sinusoïdal, on a :

$$f_{ac}(n) = A_{vib} \times \cos(2\pi f_{vib} n) \quad (3)$$

$f_{vib}$  est la fréquence du vibrato en Hz,  $A_{vib}$  est l'amplitude de la modulation de la fréquence du vibrato en Hz, et  $n$  représente le temps en échantillons. Il faut remarquer que la période d'échantillonnage de  $n$  est déterminée par  $F_{ef_0}$  ( voir équation (1) ) et non par la fréquences d'échantillonnage du son  $F_e$  .

Pour l'estimation de la fréquence moyenne locale nous pouvons utiliser une transformée de Fourier à court terme de la trajectoire de la fréquence fondamentale. En sachant que la composante continue se situe dans la fréquence 0 de la transformée de Fourier, et en prenant en compte une fenêtre d'analyse, on a :

$$f_{dc}(II) = \frac{\sum_{n=II}^{II+Nvib-1} f0(n)w(n-II)}{\sum w(n)} \quad (4)$$

où  $l \in \{0,1,2,\dots\}$  et  $I$  est la différence ( en échantillons ) entre les positions des fenêtres d'analyse  $w(t)$  ( habituellement appelées en anglais « hop size » ).  $f0(n)$  est la trajectoire de la fréquence fondamentale dans un fenêtre de taille  $Nvib$  échantillons. La division par la somme des points appartenant à la fenêtre sert à normaliser cette équation.

Sauf si la différence temporelle  $I$  , prise pour l'estimation de  $f_{dc}(n)$  , est d'un échantillon, le signal  $f_{dc}(II)$  sera un signal de fréquence d'échantillonnage inférieure à celle du  $f0(n)$  . Une interpolation linéaire doit être réalisée pour trouver un signal  $f_{dc}(n)$  .

Un bon compromis entre la résolution temporelle et fréquentielle est déterminant pour le choix de la taille de fenêtre  $Nvib$  . On la détermine à partir de l'équation suivante:

$$Nvib = CLF \times T_{vib} \times F_{ef0} \quad (5)$$

Où CLF est une valeur prise empiriquement ( nous utilisons une valeur de  $CLF=4$  ).  $T_{vib}$  est la période du vibrato qu'on attend trouver. Un valeur typique de la période du vibrato est 160 millisecondes, ce qui correspond à une fréquence de modulation approximée de 6 Hz. Mais la période  $T_{vib}$  peut être comprise entre 125 et 330 millisecondes, ce qui correspond à une fréquence de 8 Hz et 3 Hz respectivement.

Finalement, on obtient la composante rapide en soustrayant la valeur de la fréquence moyenne locale à la trajectoire du  $f0(n)$  qu'on vient d'estimer:

$$f_{ac}(n) = f0(n) - f_{dc}(n) \quad (6)$$

Ci-dessous, on verra comment à partir des composantes lente et rapide, on peut trouver les coefficients de transposition pour modifier l'étendue du vibrato d'un son.

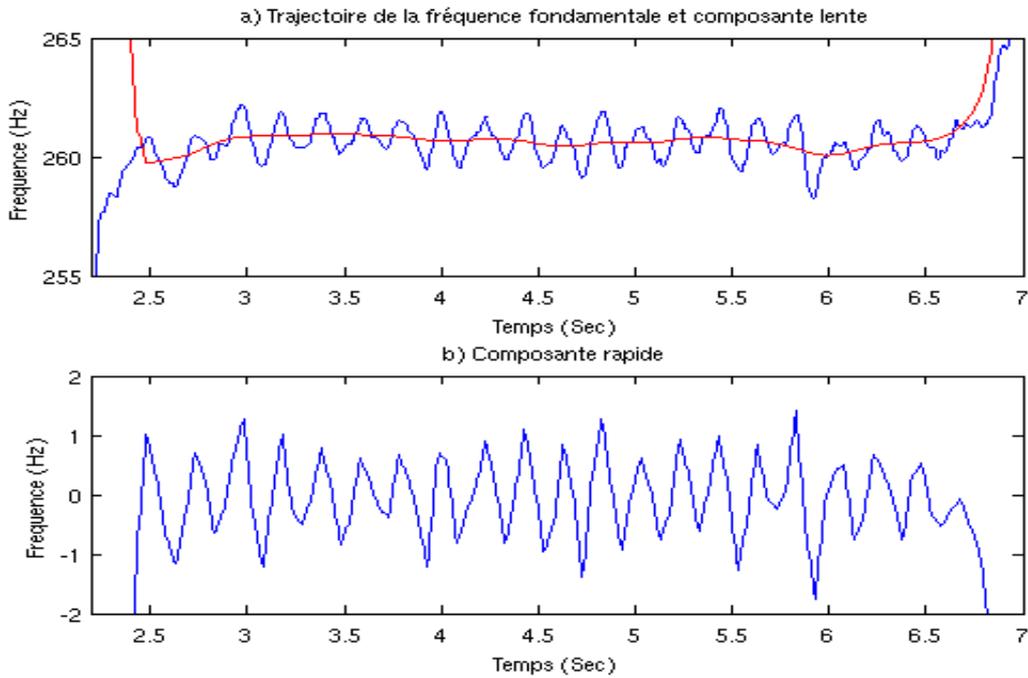


Illustration 2: D'après la trajectoire de la fréquence fondamentale d'un son de trompette avec vibrato (a, courbe en bleue) la composante lente est obtenue (a, courbe en rouge). La soustraction de la composante lente de la trajectoire  $f_0$  donne la composante rapide (b).

### 3.5. Modification de l'étendue du vibrato

Comme on l'a déjà dit, la modification de l'étendue du vibrato sera faite à partir d'un modèle *source/filtre* dans le vocoder de phase. Pour obtenir une modification de l'étendue du vibrato, une transposition du spectre d'amplitude du son avec vibrato sera faite à chaque trame d'analyse. Un facteur de transposition  $\alpha$ , donné par l'utilisateur, nous permettra d'indiquer la quantité de modification souhaitée.

Les coefficients de transposition sont obtenus à partir du rapport entre la  $f_{dc}$  et la  $f_0$  et le facteur de transposition  $\alpha$  comme il suit :

$$T(n) = (1 - \alpha) \times 1200 \frac{\log(f_{dc}(n)/f_0(n))}{\log(2)} \quad (7)$$

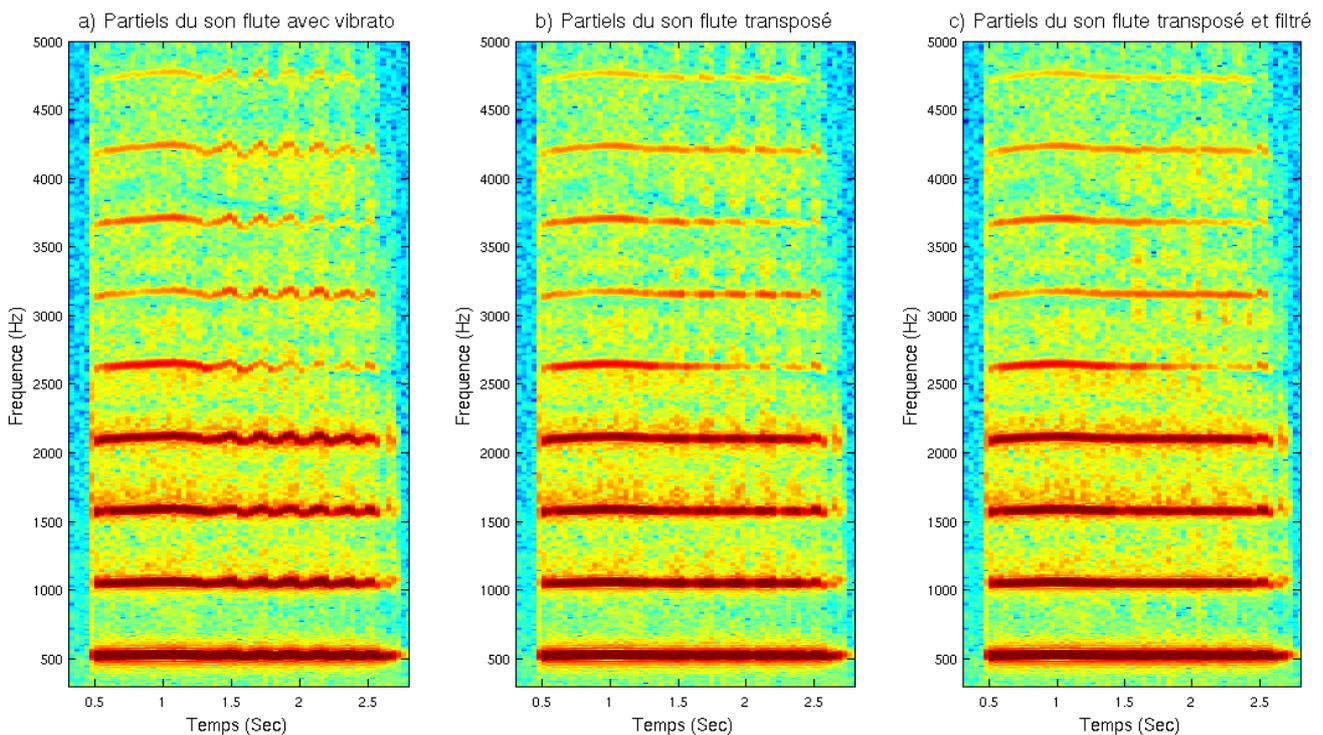
où  $T(n)$  est un vecteur de transposition en cents à appliquer à chaque instant  $n$ . Le facteur de transposition est utilisé de la façon suivante :

$$\begin{cases} \alpha < 0, & \text{inversion de phase de l'étendue du vibrato} \\ \alpha \in [0, 1], & \text{atténuation complète de l'étendue '0' jusqu'au vibrato sans alteration '1'} \\ \alpha > 1, & \text{amplification de l'étendue du vibrato} \end{cases} \quad (8)$$

La transposition d'un son dans le modèle *source/filtre* est faite en séparant la source du filtre et en transposant seulement la source. Après cette opération, le filtre est appliqué sur la source transposée pour rétablir les caractéristiques spectrales du son d'origine [15]. Dans le cas où le son a une modulation d'enveloppe spectrale au cours du temps, cette modulation sera remise dans le son traité. Ceci est le cas des instruments de musique joués avec vibrato.

D'après la figure 3.a, on voit que chaque amplitude des partiels est modulée d'une façon différente. En fait, chaque fréquence du partiel passe par différents modes de résonance de l'instrument lors de la modulation de fréquence produit par le vibrato, donc il est logique que chaque amplitude de partiel ait sa propre modulation. De plus, puisqu'il existe une modulation d'enveloppe spectrale, la forme de cette modulation est différente selon que la trajectoire de la fréquence fondamentale monte ou descende lors du vibrato .

D'après ces résultats, quand on veut modifier l'étendue du vibrato, il s'avère nécessaire d'avoir une étape de filtrage en cascade à l'étape de transposition du son. Dans le sous-chapitre suivant, nous exposons la façon dont on obtient l'enveloppe spectrale d'un son et comment on peut s'en servir pour l'obtention des coefficients d'un filtre qui nous permettra de minorer ces modulations d'amplitude.



*Illustration 3: La fréquences fondamentale et les huit premiers partiels du son de flute avec vibrato (a). Après la transposition, (b) la modulation en fréquence a disparu, mais on observe des modulations d'amplitudes dans les partiels du son. Dans (c) est présenté le signal après l'étape de filtrage par la méthode développée par Maller. Notons en (c) quelques pics isolés qui ont été créés lors de ce filtrage.*

### 3.6. Enveloppe spectrale

Les applications de l'enveloppe spectrale dans le domaine de la modification du son sont variées : la transposition de façon indépendante de la hauteur et du timbre, l'utilisation de l'enveloppe spectrale d'un son pour filtrer un autre son ( hybridation ), le modification de l'enveloppe spectrale du son pour accentuer ou diminuer le timbre d'un son, etc [15]. L'importance de l'estimation de l'enveloppe spectrale dans notre travail, réside d'une part dans le fait qu'on va obtenir, à partir d'elle, la source dans le modèle *source/filtre* du son avec vibrato, et d'autre part qui va nous permettre d'obtenir les coefficients d'un filtre pour enlever la modulation d'enveloppe spectrale du son vibrato dont on a modifié son étendue.

Malgré le fait d'avoir un nom qui exprime bien ce qu'elle essaye d'estimer, et le fait que nous pouvons avoir une idée de ce qu'on estime avec elle, l'enveloppe spectrale manque d'une définition mathématique unique. Il existe différentes façons de l'estimer, notons par exemple le cepstre réel, le cepstre discret, la LPC et la « True envelope » [20], [15]. Dans notre travail de recherche on a utilisé les deux derniers estimateurs. La méthode « True envelope » sera décrite dans ce chapitre, et une vision générale de la méthode LPC sera donnée dans le chapitre suivant.

Tout d'abord, on va définir l'enveloppe spectrale d'un son comme une courbe reliant les maxima du spectre de magnitude obtenu lors d'une transformation de Fourier d'un son [20]. Nous verrons par la suite la méthode « True envelope » et ensuite on discutera son utilisation dans la transposition du son avec vibrato.

Malgré que les résultats obtenus lors de la transposition en gardant l'enveloppe spectrale indiquent une diminution des amplitudes des composantes harmoniques, ils montrent aussi qu'il existe une modulation dans l'enveloppe spectrale qu'on va remettre après la transposition, induisant ainsi une modulation d'amplitude dans le son.

En fait, la préservation de l'enveloppe spectrale pendant la transposition d'un son dans le vocoder de phase ne garantit pas la crédibilité du son transformé [14]. La comparaison de deux échantillons sonores provenant d'un même instrument, mais en changeant seulement la hauteur jouée, montre des différences entre les enveloppes spectrales provenant des ces deux sons. De plus, pour le même instrument mais différentes hauteurs, les formants pourraient changer leurs niveaux. En effet, ceci dépend de l'instrument mais aussi du style de musique joué.

L'analyse d'enveloppe spectrale est réalisé dans le vocoder de phase SuperVP et enregistrée dans le format SDIF<sup>1</sup>. Dans Matlab, la lecture de ce fichier est assurée par la bibliothèque EASDIF<sup>2</sup> [2].

#### 3.6.1. Estimation de l'enveloppe spectrale : « True envelope »

Une version de l'enveloppe spectrale peut être obtenue à partir d'un filtrage du cepstre réel. L'atténuation des composantes de haute fréquence dans le cepstre ( équation (9) ) produit une représentation approximée, et donc lissée, du spectre d'amplitude. Le choix des composantes à prendre en compte est donné par le nombre des fréquences des harmoniques  $P_c$  ( ou ordre du filtre cepstral ) . Dû à la perte d'énergie impliquée par le filtrage du cepstre, le résultat de cette

---

1 SDIF ( « Sound description interchange format for audio features » [2] ) est un format de fichier ouvert et extensible développé par l'IRCAM, le CNMAT et le MTG pour stoker des descripteurs audio.

2 Pour plus d'information s'adresser à la page web : <http://www.ircam.fr/sdif/download/Easdif>

opération ne va pas créer une enveloppe suivant les maxima du spectre, mais suivant leurs valeurs moyennes.

Le cepstre réel  $C(l)$  est donné par :

$$C(l) = \sum_{k=0}^{K-1} \log(|X(k)|) e^{i \frac{2\pi kl}{K}} \quad (9)$$

Où  $X(k)$  est la DFT de K point d'un signal  $x(n)$ .

L'analyse appelée « true envelope » propose une solution itérative pour trouver les valeurs des maxima à partir de cette moyenne.

Soit  $V_i(k)$  la représentation de l'enveloppe spectrale d'ordre  $P_c$  dans la  $i$ -ème itération, et les valeurs initiales de  $A_0(k) = \log(|X(k)|)$  et  $V_0(k) = -\infty$ , l'algorithme garde les valeurs maximales entre les deux vecteurs  $A_0(k)$  et  $V_0(k)$  à l'itération  $i$  :

$$A_i(k) = \max(A_{i-1}(k), V_{i-1}(k)) \quad (10)$$

Le filtre cepstral est appliqué alors sur l'actuel valeur de  $A(k)$ , et une nouvelle itération est commencée. De cette façon, à chaque itération, les vallées entre les pics seront remplis et l'enveloppe spectrale augmentera son énergie jusqu'à qu'elle passe par les pics de  $A_0(k)$ . Comme règle d'arrêt de l'algorithme, on spécifie une valeur de seuil pour la différence de niveau entre les maxima de  $A_0(k)$  et l'enveloppe spectrale estimée. Des valeurs typiques pour ce seuil sont compris entre 3 et 2 dB [14].

### 3.6.2. Filtre d'enveloppe spectrale

On s'intéresse maintenant à trouver un filtre qui permette de moyennner l'évolution temporelle de l'enveloppe spectrale. En moyennant cette évolution au cours du temps, on s'attend à minorer la variation d'amplitude qu'elle produit. Dans la bibliographie, on trouve différentes manières d'arriver au même but. Notamment, les auteurs de [9] proposent d'estimer l'enveloppe spectrale dans un point de la trajectoire de  $f_0$  qui appartient à la composante lente  $f_{dc}$  et à partir de cette enveloppe, une interpolation des amplitudes peut être appliquée pour chaque fréquence pour toutes les enveloppes spectrales.

L'algorithme proposé par Maller applique une pondération sur des enveloppes spectrales passées. La différence entre les modulations d'amplitude des différents partiels justifierait une séparation entre les enveloppes du vibrato lors de la montée en fréquence et lors de la descente [11].

L'algorithme fait une recherche des endroits où il existe un croisement entre la trajectoire de la fréquence fondamentale  $f_0$  et de la composante continue  $f_{dc}$ . Cette recherche est faite par fenêtres, et un pas d'avancement est incrémenté jusqu'à avoir parcouru tout la trajectoire de la  $f_0$ .

À chaque fenêtre d'analyse, dans le point où la condition de croisement est accomplie, les enveloppes spectrales du temps présent  $env(n, k)$ <sup>3</sup> et celle du temps précédent  $env(n-1, k)$  sont pondérées en fonction de l'étendue du vibrato comme ci-dessous :

$$R_{es}(n, k) = \frac{env(n, k) \times |f_{ac}(n-1)| + env(n-1, k) \times |f_{ac}(n)|}{|f_{ac}(n-1) - f_{ac}(n)|} \quad (11)$$

Pour chaque fenêtre d'analyse, l'enveloppe  $R_{es}(n, k)$  est gardée selon le sens de la trajectoire  $f_0$ . Si cette trajectoire monte, l'enveloppe résultante est enregistrée dans une matrice  $E_m(j, k)$ , et dans le cas où la trajectoire  $f_0$  descend, l'enveloppe est gardée dans une matrice  $E_d(i, k)$ . Une fois finie l'analyse sur la fenêtre actuelle, on fait une moyenne entre les enveloppes montantes  $\bar{E}_m(k)$  et une moyenne pour les enveloppes descendantes  $\bar{E}_d(k)$ .

Finalement, pour obtenir l'enveloppe spectrale moyennée  $\hat{E}(k)$  pour le temps  $n$ , on prend en compte les enveloppes de montée et de descente. Si  $\bar{E}_m(k)$  et  $\bar{E}_d(k)$  ne sont pas des vecteurs vides, et le sens de la trajectoire de  $f_0$  est montant :

$$\hat{E}(k) = \frac{1+\alpha}{2} \times \bar{E}_m(k) + \frac{1-\alpha}{2} \times \bar{E}_d(k) \quad (12.a)$$

et dans le cas où la trajectoire de  $f_0$  descende :

$$\hat{E}(k) = \frac{1-\alpha}{2} \times \bar{E}_m(k) + \frac{1+\alpha}{2} \times \bar{E}_d(k) \quad (12.b)$$

Par contre, s'il existe qu'une des deux moyennes  $\bar{E}_m(k)$  ou  $\bar{E}_d(k)$ , l'enveloppe  $\hat{E}(k)$  prend la valeur de l'enveloppe existante.

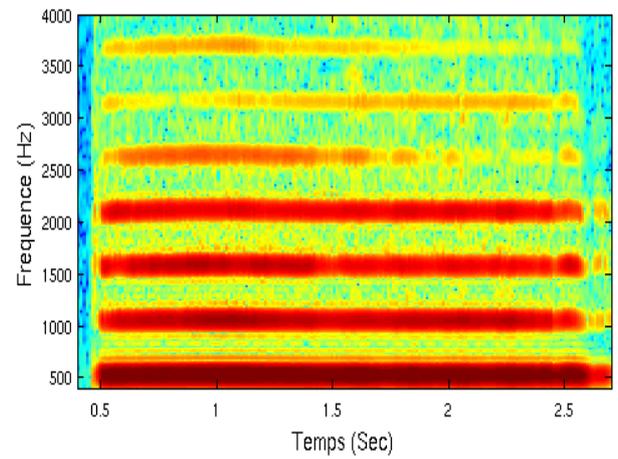
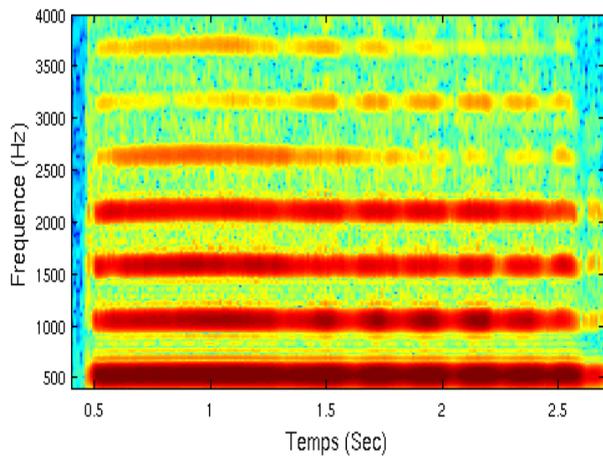
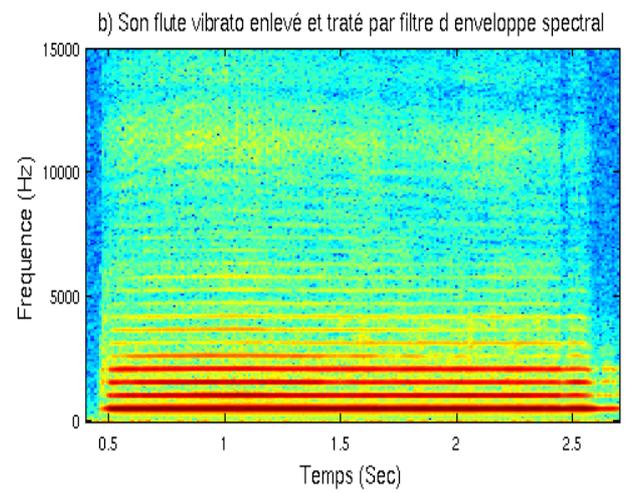
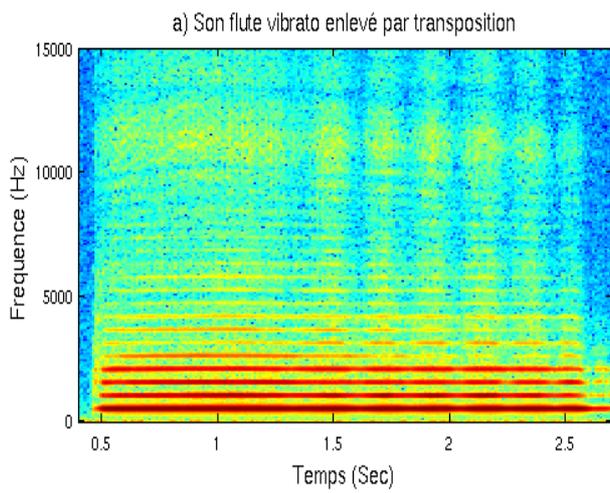
Finalement, les coefficients du filtre  $H_{spec}(n, k)$  qui varie au cours du temps  $n$  sont déterminés dans le domaine fréquentiel à partir de l'équation suivante:

$$H_{spec}(n, k) = \frac{\hat{E}(n, k)}{env(n, k)} \quad (13)$$

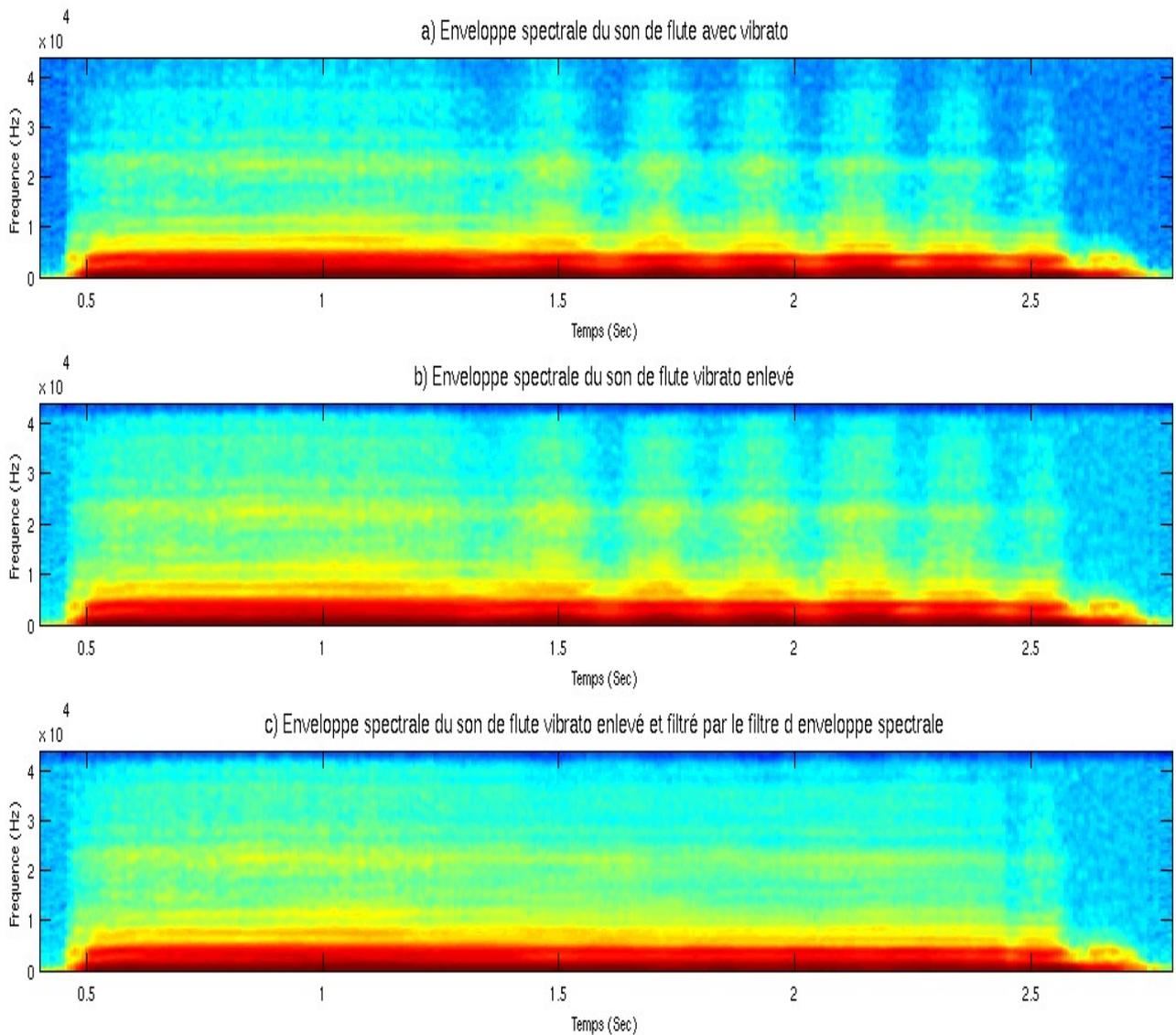
Ces coefficients sont remis à une échelle linéaire avant d'être enregistrés dans un fichier SDIF. L'étape de filtrage est fait dans le logiciel SuperVP.

---

3 Les amplitudes des enveloppes spectrales sont ici exprimées en log-amplitude.



*Illustration 4: Comparaison entre les enveloppes spectrales d'un son de flute avec vibrato (a), celles du même son après avoir modifié l'étendue du vibrato, et (c) celle d'un son avec vibrato modifié et filtré par le filtre d'enveloppe spectrale. Les amplitudes des enveloppes sont en log-magnitud (dB).*



*Illustration 5: Dans la colonne (a) on montre le spectrogramme d'un son de flute avec vibrato qui a été transposé pour annuler le vibrato, on présent en bas un zoom du spectrogramme. Dans la colonne (b) on peut voir le spectrogramme du même son après l'avoir filtré par le filtre d'enveloppe spectrale décrite ci-dessus. On observe une diminution de modulation d'amplitude dans tout le spectre. En bas de la colonne (b) on observe autour de 3000 Hz, entre les secondes 2 à 3, des amplifications de fréquences dans le bruit du signal. Dans le troisième partiel on n'observe pas beaucoup d'atténuation de la modulation d'amplitude.*

## 4. Démarche réalisé

Dans le chapitre précédent, on a vu comment à partir du modèle *source/filtre* et du vocoder de phase on a traité le son vibrato pour modifier son étendue. Dans un première temps, on a fait une étape d'analyse basée sur la détection de la fréquence fondamentale et l'extraction de ses composantes lente et rapide. Dans un deuxième temps, cette analyse nous a permis de trouver les coefficients de transposition pour modifier l'étendue du vibrato lors de la resynthèse dans le vocoder de phase. Finalement, après avoir mis en évidence la modulation d'amplitude dans les partiels du son avec vibrato, on a présenté brièvement une étape de filtrage basée sur l'enveloppe spectrale qui permet de diminuer cette modulation.

La méthode proposée jusqu'ici, prend implicitement les composantes harmonique et bruit du son avec vibrato. Mais étant donné qu'une de nos hypothèses est que la modulation d'amplitude dans le bruit et dans les harmoniques peut être différente, nous allons travailler, dans ce qui suit, à partir du modèle *sinusoïde plus bruit*.

Nous prenons ce modèle aussi dans le but d'accomplir l'objectif de généraliser les recherches déjà menées par l'équipe analyse/synthèse. Dans ce sens nous croyons que le fait de pouvoir travailler avec l'amplitude de chaque partiel du son nous donne deux avantages importants par rapport au modèle antérieur.

D'un part, il permettrait de développer une étape de filtrage plus simple que celle qui a été implémentée pour enlever la modulation d'amplitude. Pour faire ce filtre, il faudrait prendre en compte un filtrage en parallèle de la composante harmonique et du bruit. La méthodologie du filtrage des composantes harmoniques serait une extension de la méthodologie qu'on a utilisée pour le changement de l'étendue du vibrato. Le filtrage du bruit serait fait avec un filtre obtenu à partir d'une enveloppe spectrale de bas ordre.

D'autre part, à partir de la composante rapide de la fréquence fondamentale, de l'évolution de l'amplitude de chaque partiel, et de l'évolution de l'enveloppe spectrale du bruit au cours du temps on pourrait modéliser le vibrato d'un son [9]. Cette idée est intéressante parce qu'elle nous permettrait de modifier complètement un son avec vibrato mais aussi de mettre du vibrato dans un son qui n'en a pas. Des effets audios intéressants sont envisageables avec ce modèle, par exemple l'utilisation d'un modèle de vibrato de n'importe quel son dans un autre ( Hybridation ) .

Ainsi, la méthodologie que nous proposons est la suivante. D'abord, on va extraire les composantes lente et rapide de la fréquence fondamentale et des amplitudes des partiels. Pour la modification de l'étendue du vibrato la transposition est opérée sur chaque fenêtre d'analyse dans toute la partie du signal qui contient l'effet du vibrato. Le filtrage d'enveloppe spectrale est alors remplacé par une étape de filtrage parallèle : d'un côté les composantes sinusoïdales et de l'autre le bruit. Le filtrage de la partie harmonique est fait à partir de la composante rapide des amplitudes des partiels. Pour le filtrage du bruit, on utilise le filtre d'enveloppe spectrale expliqué dans le chapitre précédent, avec la différence que cette fois, nous utilisons une estimation d'enveloppe spectrale à partir d'un modèle LPCs de bas ordre. Une fois filtrées, les deux composantes sont additionnées et ainsi, le son résultant est le son vibrato modifié. Finalement un modèle du vibrato peut être établi à partir des estimations antérieures : la fréquence du vibrato est modélisée comme la composante rapide de la

$f_0$  , et toutes les modulations d'amplitudes sont modélisées comme les composantes rapides des amplitudes de partiels plus la modulation d'enveloppe spectrale du bruit.

Ce chapitre est organisé de la façon suivante. Tout d'abord, on commence par un rappel du modèle *sinusoïde plus bruit* dont on explique comment on obtient les composantes harmonique et du bruit. Dans le même sous-chapitre, on explique l'estimation d'enveloppe spectrale à partir des coefficients LPCs. On continue par une exposition du filtrage des amplitudes des partiels. On expliquera notamment, comment à partir de l'extraction des composantes lente et rapide des amplitudes de partiels on trouve ce filtre. Enfin nous exposons les démarches réalisées en ce qui concerne la détection automatique du vibrato.

#### 4.1. Modèle *sinusoïde plus bruit*

Le modèle *sinusoïde plus bruit* est une extension du modèle *additif*. On cherche à représenter un signal comme une somme de sinusoïdes et un bruit résiduel. Cette représentation est utile lors de l'analyse d'un signal provenant des instruments de musique du son tenu dans le temps. Par exemple, dans le cas des instruments à cordes frottées, la composante sinusoïdale est produite par les modes de vibration de la corde, alors que la composante du bruit est due au frottement de la corde par l'archet et des autres non-linéarités dues au couplage archet-corde-résonateur [21]. Les partiels d'un son harmonique sont alors représentés par des sinusoïdes du modèle et toutes les composantes du son qui n'appartiennent pas aux partiels sont représentées comme le bruit ou le composant résiduel.

Le signal sonore  $s(t)$  est exprimé comme la somme des composantes harmoniques plus un signal de bruit  $e(t)$  :

$$s(t) = e(t) + \sum_{r=1}^R A_r(t) \cos(\theta_r(t)) \quad (14)$$

Où  $A_r(t)$  et  $\theta_r(t)$  sont respectivement les amplitudes et phases instantanées pour la sinusoïde  $r$  .  $R$  est le nombre de sinusoïdes représentant des partiels du son.

La phase instantanée est prise comme l'intégrale de la fréquence instantanée  $\omega_r(t)$  en radians.

$$\theta_r(t) = \int_0^t \omega_r(\tau) d\tau \quad (15)$$

Par signal déterministe on définira la somme de composantes sinusoïdales dans l'équation (14), celles-ci variant lentement en amplitude et fréquences dans une fenêtre d'analyse.

Malgré que la composante de bruit soit normalement décrite en fonction de sa densité spectrale de puissance, en termes pratiques elle sera obtenue en soustrayant du signal la composante sinusoïdale.

La composante résiduelle  $e(t)$  peut être vue comme un bruit blanc filtré. Soit  $h(t, \tau)$  un filtre variant dans le temps  $t$  , et  $u(t)$  un bruit blanc, on a :

$$e(t) = \int_0^t h(t, \tau) u(\tau) d\tau \quad (16)$$

Le signal  $s(t)$  est analysé par trames ( ou fenêtres ). Chaque trame est multipliée par une fonction de fenêtrage ayant comme but l'élimination des discontinuités dans les bords des trames. Ensuite, le spectre d'amplitude est obtenu pour chaque fenêtre d'analyse en utilisant la transformée de Fourier discrète. Après, une détection de fréquence fondamentale est réalisée et suivie par une analyse des partiels du son. En ayant trouvé les partiels, on a obtenu la partie sinusoïdale du modèle ou composante harmonique du son.

Le signal résiduel est obtenu à partir d'une synthèse soustractive. D'abord la composante harmonique est resynthétisée ( synthèse additive ) en utilisant l'information de fréquence, d'amplitude et de phase obtenues d'après l'analyse harmonique du signal. Ensuite, cette composante harmonique est soustraite du son original dans le domaine temporel.

#### 4.1.1. Analyse de la composante sinusoïdal

On s'intéresse maintenant à trouver les valeurs des paramètres d'amplitudes  $A_r(n)$  , de fréquences  $\omega_r(n)$  et de phase  $\theta_r(t)$  pour chaque sinusoïde  $r$  des expressions (14) et (15).

La première sinusoïde correspond à la fréquence fondamentale et sa reconnaissance a été discutée dans le chapitre précédant. C'est à partir de celle-ci qu'on va chercher la fréquence et l'amplitude associée à chaque partiel du son.

Pendant l'étape d'analyse, les paramètres à contrôler sont associés à la TFD : type et taille de la fenêtre d'analyse, pas d'avancement entre fenêtres et taille  $N_a$  de la TFD. Pour bien distinguer entre deux partiels voisins, nous utilisons une fenêtre de Blackman de longueur  $M_a = M_{f0}$  échantillons. L'atténuation des premiers lobes latéraux de cette fenêtre nous aide à diminuer l'influence de l'amplitude des partiels proches.

D'autre part, les contraintes à donner pour trouver les paramètres des sinusoïdes associées aux partiels du son sont le nombre de partiels à trouver  $R$  et un seuil de magnitude. Ce seuil indique la différence maximale qui peut exister entre l'amplitude du partiel la plus petite et la valeur de l'amplitude maximale dans le spectre.

La précision spectrale ( ou définition spectrale ) est l'aptitude à mesurer une valeur de fréquence, et elle est liée à la fréquence d'échantillonnage  $F_e$  et le numéro de points de calcul du spectre du signal ( taille de la TFD ), comme  $F_e/N_a$  [3]. L'utilisation du *zero-padding* augmente le nombre de points dans le spectre par Hz de fréquence, et avec ça, on améliore la précision de l'estimation des paramètres des sinusoïdes. Pour optimiser l'algorithme de la TFD, la taille  $N_a$  est choisie d'être la puissance de deux immédiatement plus grande que quatre fois la taille de la fenêtre d'analyse harmonique  $M_a$  .

Puisqu'on veut bien distinguer deux pics de partiels qui sont proches, il est important d'avoir une bonne résolution spectrale. Les contraintes pour déterminer la taille de la fenêtre d'analyse sont les mêmes que celles pour trouver la fréquence fondamentale. Si la taille de la fenêtre d'analyse harmonique  $M_a$  est grande, on gagne de la résolution fréquentielle mais on perd de la résolution temporelle et on risque de ne pas mesurer le vibrato. D'autre part si on prend une taille de fenêtre petite, on améliore la résolution temporelle mais on risque de confondre les partiels.

Dans la figure 6, on montre une estimation des paramètres d'amplitude et de fréquence pour un son violon avec vibrato. On peut facilement apercevoir l'influence de l'enveloppe spectrale associée à l'instrument dans les amplitudes de partiels.

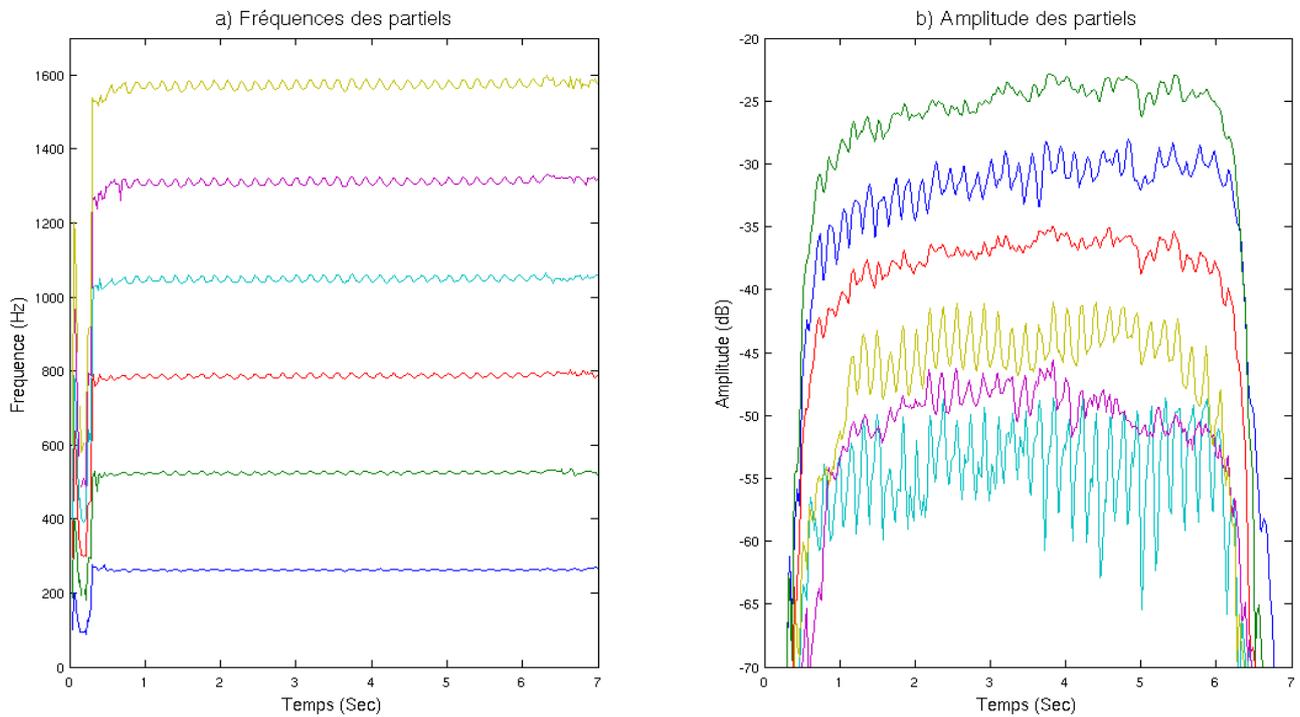


Illustration 6: Estimation des paramètres des sinusoides d'un son violon avec vibrato.

#### 4.1.2. Analyse du signal résiduel

Une fois trouvée la composante sinusoidale, le signal résiduel sera obtenu à partir d'une soustraction de celle-ci au son original. Cette soustraction sera faite dans le domaine temporel. Pour faire cette soustraction, la phase des sinusoides doit être préservée [21].

On partait de l'hypothèse que l'enregistrement a été fait dans des conditions idéales, de façon à trouver dans le bruit du signal seulement l'influence de l'instrument enregistré. Dans le cas contraire, la composante du bruit aura d'autres signaux parasites, comme de la réverbération ou du bruit de fond.

A partir de l'estimation des paramètres des sinusoides, un signal contenant la partie harmonique est synthétisé par frames en considérant que la fréquence  $y$  est constante, comme suit :

$$d(m) = \sum_{r=1}^R A_r \cos(\omega_r m + \varphi_r) \quad (17)$$

$$m = 0, 1, 2, \dots, S - 1$$

Où  $R$  est le nombre de sinusoides à reconstruire dans la frame, et  $S$  est la taille de la frame.

Le signal résiduel est enfin obtenu:

$$e(n) = (s(n) - d(n)) \quad (18)$$

Dans notre cas idéal, le bruit ne présente pas de composantes sinusoïdales. Mais puisque le vibrato peut produire une modulation d'enveloppe spectrale, le bruit pourrait avoir une modulation d'amplitude. Dans ce cas-là, le signal résiduel peut être filtré en utilisant le filtre d'enveloppe spectrale décrit dans le chapitre précédant. La différence cette fois-ci, est qu'on ne cherche pas à trouver le comportement entre les pics d'amplitude des partiels d'un signal mais à trouver une moyenne entre ses composantes. Cette différence se traduit en une réduction de la complexité dans le calcul de l'enveloppe. Nous allons utiliser par la suite le modèle LPC d'ordre bas, pour l'estimation de l'enveloppe spectrale du bruit.

### 4.1.3. Enveloppe spectrale du signal résiduel

La méthode LPC ( « Linear predictive coding » ) a été développée pour la transmission et compression de la parole. Elle est utilisée aussi pour estimer l'enveloppe spectrale du son [20]. La méthode consiste à représenter une valeur du signal  $s(n)$  à partir de la somme des  $j$  valeurs passés dans le domaine temporel :

$$\hat{s}(n) = \sum_{i=1}^j a_i s(n-i) \quad (19)$$

Où  $j$  est l'ordre de la prédiction. Les coefficients  $a_i$  sont cherchés de façon à minorer l'erreur de prédiction  $e(n) = \hat{s}(n) - s(n)$ . La récursion de Levinson-Durbin est utilisée pour résoudre l'équation provenant de la minoration de cette erreur. Cette méthode est réalisée sur des fenêtres du signal  $s(n)$  avec un pas d'avancement déterminé. Un fois que le modèle  $\hat{s}(n)$  du signal  $s(n)$  est obtenu, le spectre d'amplitude du premier montre un comportement moyenné par rapport au deuxième.

## 4.2. Filtre des composantes harmoniques

Comme on a déjà vu, l'amplitude de chaque partiel est modifiée au cours du temps par le vibrato. Il peut y avoir seulement une modulation d'amplitude produite par le balayage des fréquences autour des modes de résonances de l'instrument, ou il peut y avoir une modulation d'enveloppe spectrale produite en fonction de la façon de réaliser le vibrato. Quelque soit le cas, la modulation d'amplitude de chaque partiel peut être décomposée dans une série de Fourier où la fréquence fondamentale sera déterminée par la fréquence du vibrato.

Le principe de la méthode du filtrage des composantes harmoniques, développé pendant le stage, est analogue à la méthode qu'on a utilisée pour trouver les coefficients de transposition pour la modification de l'étendue du vibrato. Tout d'abord, on va trouver les composantes lente et rapide de l'amplitude estimée pour chaque partiel. Ensuite, on va déterminer les facteurs pour lequel l'amplitude de la fréquence de chaque partiel doit être multipliée pour subir le changement souhaité. Dans la détermination de ces facteurs intervient le facteur de transposition  $\alpha$  qu'on a utilisé pendant la modification de l'étendue du vibrato.

Une fois la modification de l'étendue du vibrato a été opérée dans le son d'origine ( son avec vibrato ), une analyse de fréquence fondamentale doit être refaite sur le son modifié. À partir de cette analyse, on cherche les amplitudes et les fréquences au cours du temps des partiels du son. L'analyse harmonique est faite à l'aide du logiciel pm2 développé à l'IRCAM.

Les composantes lente et rapide de l'amplitude de chaque partiel peuvent être obtenues à partir d'une généralisation de l'équation (4) , de la façon suivante:

$$Adc_p(II) = \frac{\sum_{n=II}^{II+Nvib-1} A_p(n)w(n-II)}{\sum w(n)} \quad (20)$$

Où  $A_p(n)$  et  $Adc_p(II)$  sont des vecteurs contenant l'amplitude associée au partiel  $p$  et la composante lente de l'amplitude de la fréquence du partiel  $p$  respectivement. La fonction du fenêtrage  $w(n)$  qu'on a choisie est celle de Hann, et la taille de la fenêtre est déterminée par l'équation (5) .  $Adc_p(n)$  est obtenu en interpolant linéairement les valeurs de  $Adc_p(II)$  .

Maintenant, nous sommes intéressés à modifier l'amplitude des fréquences des partiels  $A_p(n)$  du son dont on a modifié le vibrato. Parce que nous devons maintenir une cohérence entre la modification de l'étendue du vibrato et la modification de l'amplitude des fréquences de partiels, nous utilisons le facteur de transposition  $\alpha$  . Ceci nous indique le degré de modification que nous devons appliquer à ces amplitudes. Alors, pour chaque trajectoire du partiel, on peut trouver un facteur  $h_p(n)$  qui, multiplié par l'amplitude de la fréquence du partiel  $A_p(n)$  , nous donne la valeur de l'amplitude souhaitée.

Pour l'amplitude de la fréquence du partiel  $p$  , le facteur  $h_p(n)$  est obtenu à partir de l'équation suivante :

$$h_p(n) = \frac{(1-\alpha) \times Adc_p(n) + \alpha \times A_p(n)}{A_p(n)} \quad (21)$$

Bien évidemment, quand l'amplitude  $A_p(n)=0$  , le facteur est défini comme  $h_p(n)=1$  , de façon à ne pas modifier l'amplitude d'une fréquence qui n'a pas été trouvé ou qui n'existe pas.

À partir de cette équation on va implémenter un filtre, dans le domaine fréquentiel, qui varie au cours du temps. Ce filtre sera ensuite enregistré et utilisé, dans le domaine fréquentiel, pour le multiplier avec le spectre d'amplitude du son dont on a modifié l'étendue du vibrato. Une fois la multiplication réalisée, la TFD inverse et une étape de overlap/add nous garantissent une resynthèse sans perte du son filtré.

D'après (21), on voit que quand on veut éliminer complètement le vibrato, cette-à-dire en utilisant un facteur de transposition  $\alpha=0$  , le filtre  $h_p(n)$  sera déterminé comme le rapport entre la composante lente de l'amplitude de chaque partiel  $Adc_p$  sur l'amplitude de chaque partiel  $A_p$  . Par conséquence, dans la multiplication des deux spectres d'amplitude, celle du signal à traiter et celle du filtre, l'amplitude résultante sera la composante lente  $Adc_p(n)$  .

Pour la création du filtre, on fait un vecteur  $v_{df}(k)$  à chaque trame  $n$ , avec  $k \in \{0, 1, 2 \dots N_a/2 + 1\}$ .  $N_a$  est la taille de la TFD de l'analyse harmonique ( la même taille sera utilisée pendant l'étape du filtrage ). Ce vecteur peut être vu comme la première moitié d'un spectre d'amplitude, celle-ci obtenue après une TFD sur la réponse impulsionnelle de notre filtre. Dans cette perspective, pour chaque fenêtre d'analyse hypothétique, on va créer un vecteur  $v_{df}(k)$ . Ce vecteur contiendra les facteurs pour lesquels chaque amplitude de fréquence, représentant des partiels du son à modifier, doit être multipliée. Dans ce cas, on peut écrire ce vecteur, qui évolue au cours du temps, de la façon suivante:

$$v_{df}(k) = 1 + \sum_{p=1}^P (h_p(n) - 1) g_p(k) \quad (22)$$

Où grand  $P$  est le nombre de partiels à filtrer. Et  $g_p(k)$  un filtre passe-bande carré centré autour de la fréquence du partiel  $p$  et de largeur de bande comme on la déterminera par la suite et d'amplitude 1 à la fréquence du partiel.

Pour trouver le filtre  $g_p(k)$ , il nous reste à trouver le rapport entre les bins fréquentiels  $k$  et les fréquences des partiels  $p$ . À partir de l'analyse harmonique on connaît la fréquence associée à chaque partiel  $p$  au cours du temps. Ces fréquences doivent être approchées par la valeur de  $k \times F_e / N_a$  la plus proche. On appellera par la suite  $f_k^p$  ces nouvelles valeurs de fréquence. On choisit la largeur de bande du filtre comme la largeur du bande du lobe principal de la TFD de la fenêtre qu'on utilisera sur le signal à filtrer. Cette largeur est déterminée par le type et la taille de la fenêtre. Nous avons choisi une fenêtre de Hann de taille 0.025 secondes.

D'après Oppenheim et Schafer [13], la largeur du lobe central de la fenêtre de Hann peut être calculée en radians par:

$$ML_{rad} = \frac{4\pi}{M \times F_e + 1} \quad (23)$$

où  $M$  est la taille de la fenêtre en secondes,  $F_e$  est la fréquence d'échantillonnage. Pour le calcul de la largeur en bins du lobe principal on a:

$$ML_{bins} \approx \frac{2N}{M \times F_e} \quad (24)$$

Où  $N$  est la taille de la TFD.

La largeur de bande du filtre  $g_p(k)$  est alors comprise entre  $f_k^p - ML_{bins}$  et  $f_k^p + ML_{bins}$ . Cette largeur nous permette la modification de l'amplitude seulement dans le lobe principal des fréquences de partiels trouvés.

Une matrice  $V_{df}(n, k)$  constituée à partir des valeurs trouvées pour chacune des trames  $v_{df}(k)$  est enregistrée en utilisant le format de fichier SDIF. L'étape de filtrage du signal vibrato modifié est réalisée, dans le domaine fréquentiel, par le logiciel SuperVP.

## 4.5. Reconnaissance automatique du vibrato

À la fin de notre stage, nous nous sommes intéressés à la détection automatique des segments contenant l'effet vibrato. Il existe plusieurs motivations pour s'intéresser à la segmentation automatique. La modification du vibrato appliquée sur des transitoires d'un son, produit un changement de fréquence qui n'est pas souhaité. On pourrait connaître à l'avance les segments des transitoires du son, pour n'appliquer pas les traitements associés à la modification du vibrato. Lors de la détection automatique du vibrato on pourrait l'enlever sans devoir indiquer à la main les points du début et du fin du segment à traiter. D'autre part, l'estimation des paramètres pourraient servir à classer des sons, par exemple dans une base de données. Dans ce contexte, les auteurs du [7] proposent, que les différentes façons de produire le vibrato dans les saxophones pourraient être reconnues à l'aide des coefficients trouvés lors de sa détection. D'autre part, pour la segmentation automatique dans des zones stables, le vibrato détériore les performances des algorithmes [19] : quand le vibrato est présent dans un son, la détection de la transition entre deux notes successives devient problématique, le vibrato pourrait être supprimé avant d'estimer cette transition.

Dans le but de la segmentation en zones stables, Rossignol [19] s'attache au problème de la détection du vibrato, de l'estimation de ses paramètres, et de sa suppression. L'auteur distingue deux groupes de méthodologie pour sa détection. D'une part, les méthodes qui travaillent directement dans le domaine amplitude/temps du signal ( signal analytique ) , et d'autre part, les méthodes qui utilisent une analyse de fréquence fondamentale.

Pour commencer notre travail, nous nous sommes appuyés sur les recherches de Rossignol, en particulier sur une des méthodologies basée dans la trajectoire de la fréquence fondamentale. La méthode proposée consiste à observer cette trajectoire comme une somme de sinusoides et à estimer les paramètres de ces sinusoides. La TFD de la trajectoire de la fréquence fondamentale est calculée par fenêtres avec un pas d'avancement. Le principe de détection peut alors se poser en termes d'évaluation de la position du maxima dans le spectre d'amplitude ou d'une comparaison entre l'énergie contenue dans la bande de fréquences possibles du vibrato et l'énergie qui n'appartient pas à cette bande.

Nous avons choisi de commencer par la première de ces méthodes car, en même temps qu'il permet de réparer les endroits avec vibrato, il permet de trouver aussi ses paramètres : la fréquence du vibrato, son amplitude et sa phase [19]. Selon notre approche *sinusoïdale plus bruit*, l'évolution des amplitudes de partiels et l'évolution du bruit au cours du temps seraient eux aussi des paramètres estimables du vibrato, mais on n'a pas abordé cet question ici.

Notre approche est expliquée par la suite. La variation quasi-périodique de la fréquence fondamentale autour de sa valeur centrale peut être vue comme une somme de sinusoides dont les amplitudes et les fréquences instantanées varient dans le temps [19]. Notre signal à analyser est donc la composante rapide de la trajectoire de la fréquence fondamentale  $f_{ac}(n)$  . À partir de la TFD opérée sur la  $f_{ac}(n)$  on obtient le spectre d'amplitude du vibrato. On utilisera la même taille de fenêtre  $N_{vib}$  que pendant l'extraction des composantes lente et rapide dans la  $f_0$  . On utilisera un chevauchement de  $O_{vib}$  échantillons et un pas d'avancement  $P_{av} = N_{vib} / O_{vib}$  échantillons.

L'équation de la TFD s'exprime :

$$F_{AC}(lP_{av}, k) = \frac{\sum_{n=1}^{P_{av} + N_{vib} - 1} f_{ac}(n) \times w(n - lP_{av}) e^{-j\pi \frac{kn}{K}}}{\sum w(n)} \quad (25)$$

où  $l$  est l'actuelle frame d'analyse.  $w(n)$  est une fenêtre de Hann. Le spectre des amplitudes est ensuite calculé comme :

$$F_{AC}(lP_{av}, k) = |F_{AC}(lP_{av}, k)| \quad (26)$$

et puis l'énergie peut être calculée comme :

$$EF_{AC}(lP_{av}, k) = |F_{AC}(lP_{av}, k)|^2 \quad (27)$$

Le maxima de chaque fenêtre est ensuite gardée dans un vecteur  $Max(lP_{av})$ . Pour avoir la même fréquence d'échantillonnage que  $f_{ac}(n)$  une interpolation linéaire entre les valeurs de pics de chaque fenêtre est faite, on obtient  $Max(n)$ . Pour déterminer l'existence ou pas du vibrato, on va réunir un ensemble d'indicateurs dans une fonction d'observation.

#### 4.4.1. Fonction d'observation

Pour prendre la décision de l'existence ou non du vibrato dans un trajectoire de  $f0$ , on définit une fonction d'observation. Cette fonction prend comme entrée des données correspondant à différents critères, l'utilisateur étant chargé de choisir ses seuils. Après avoir évalué les différentes données d'entrée et choisi celles qui restent valables après les seuils de discrimination, elles seront pondérées pour produire une réponse. La fonction donne comme sortie un ensemble de valeurs de temps correspondant au début et à la fin des morceaux où l'on a trouvé l'effet du vibrato.

Plusieurs façons de détecter le vibrato ont été proposées dans la bibliographie [19], mais aucune d'entre elles s'avère parfaite ou meilleur qu'une autre. Souvent, le problème de la détection du vibrato est déplacé vers une problématique de pondération d'indicateurs de présence du vibrato.

L'approche que nous avons choisi prend en compte trois indicateurs de présence du vibrato.

- 1) Le premier indicateur vient de la détection de fréquence fondamentale et représente la fiabilité de l'estimation du  $f0$ . Cette valeur est conservée lors de l'obtention de la composante rapide  $f0_{ac}$ , et nous allons dire que cet indicateur est partagé entre l'analyse  $f0$  et sa composante rapide. Le seuil déterminera la fiabilité minimale qui peut avoir une certaine valeur de la  $f0_{ac}$ .

- 2) À partir de la décomposition du trajet de la fréquence fondamentale comme une somme de sinusoides, une détection du maxima dans le spectre d'amplitude nous donne une estimation de la fréquence associée au vibrato. Quand la valeur de cette fréquence est comprise entre une valeur minimale  $Fv_{min}$  et une valeur maximale  $Fv_{max}$ , on va dire qu'il existe du vibrato. Des valeurs typiques peuvent être comprises entre  $[3,5]$  Hz pour la première et  $[7,9]$  Hz pour la deuxième.
- 3) On considérera le vibrato comme inexistant quand l'amplitude ou la fréquence n'a pas pu être obtenue à travers l'analyse de fréquence fondamentale.

D'autres indicateurs peuvent être implémentés, nous avons déjà commenté le rapport entre l'énergie contenue dans la bande du vibrato  $[Fv_{min}, Fv_{max}]$  sur l'énergie dans le reste du spectre d'amplitude de la  $f_{ac}$ . Un autre indicateur à prendre en compte est un seuil de l'amplitude de l'étendue du vibrato [23].

Comme on utilise l'analyse  $f_0$ , c'est à dire une représentation du signal sous-échantillonné, les valeurs obtenues dans notre fonction d'observation peuvent subir un léger décalage [19].

## 5. Résultats

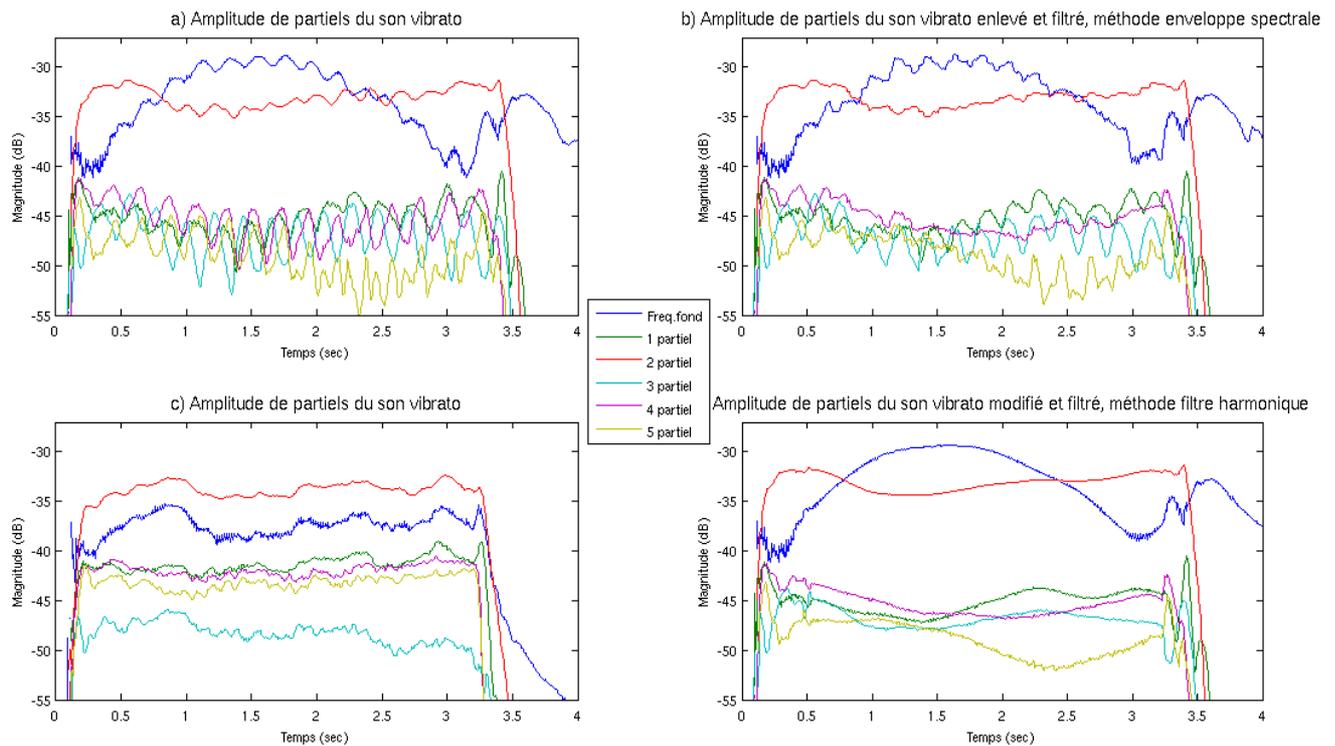
Pour l'évaluation des méthodes proposées dans ce rapport, ainsi que pour l'étude en général, nous avons utilisé des sons de différents instruments de musique obtenus à partir de deux bases de données. La première est la base de données « IRCAM's Studio Online Instrumental sounds ». Les enregistrements ont été faits avec une résolution de 24 bits, à une fréquence d'échantillonnage de 48 Khz. Trois techniques de disposition du microphones ont été utilisées : microphone en contact avec l'instrument, en champs proche et lointain. La salle d'enregistrement était une salle à acoustique variable ( salle ESPRO ) . La deuxième base de données est « The University of Iowa Musical Instrument Samples ». Les sons ont été enregistrés dans une chambre anéchoïque en utilisant un microphone Neumann KM 84, une table de mixage Mackie 1402-VLZ, et ils ont été enregistrés en DAT avec un enregistreur Panasonic SV-3800. Les sons sont enregistrés sur 16 bits, avec une fréquence d'échantillonnage de 44.1 Khz. À la différence de celle de l'IRCAM, cette base de données est disponible gratuitement sur Internet<sup>4</sup>.

Pour commencer notre analyse, nous comparons les résultats obtenus lors du filtrage d'un son qui a été transposé pour obtenir une modification de l'étendue du vibrato. Dans la figure 7 première colonne, on compare les amplitudes des partiels entre un son de contrebasse avec vibrato et un son de même contrebasse avec le même musicien mais la note est jouée sans vibrato. Dans la deuxième colonne, on montre le premier son de contrebasse avec vibrato mais qui a été traité pour enlever la modulation en fréquence. En haut, se trouve le son auquel nous avons appliqué le filtre obtenu avec l'enveloppe spectrale, et en bas le son auquel nous appliquons le filtre des harmoniques en parallèle avec le filtre de bruit.

D'après la figure 7.a on observe une oscillation des valeurs d'amplitude pour un son avec vibrato de 1 dB pour la fréquence fondamentale et le deuxième partiel, alors que pour le 3ième et 4ième partiels, les amplitudes arrivent à une différence de 8 dB. On observe aussi des variations d'amplitude dans le son sans vibrato 7.c, mais cette fois-ci, la différence qu'on pourrait associer à la modulation du vibrato ne dépasse pas 1.5 dB dans le 3ième partiel. Si on analyse les résultats obtenus lors du filtrage par le filtre obtenu à partir des enveloppe spectrale, on peut voir des oscillations de 1.5 dB pour l'amplitude de la fréquence fondamentale et de 4 dB pour l'amplitude du 3ième partiel. Dans le 4ième partiel, on observe une importante réduction de la modulation d'amplitude. Malgré qu'il existe une réduction des modulations, en ce qui concerne les composantes sinusoïdales, ce filtrage n'a pas été efficace. Finalement, on observe dans le filtrage des composantes harmoniques présentées dans 7.d, une importante réduction de modulation des amplitudes dans les six composantes sinusoïdales. En effet, l'amplitude des partiels à moins de variation d'amplitude que celle que produit l'instrumentiste quand il joue sans vibrato.

---

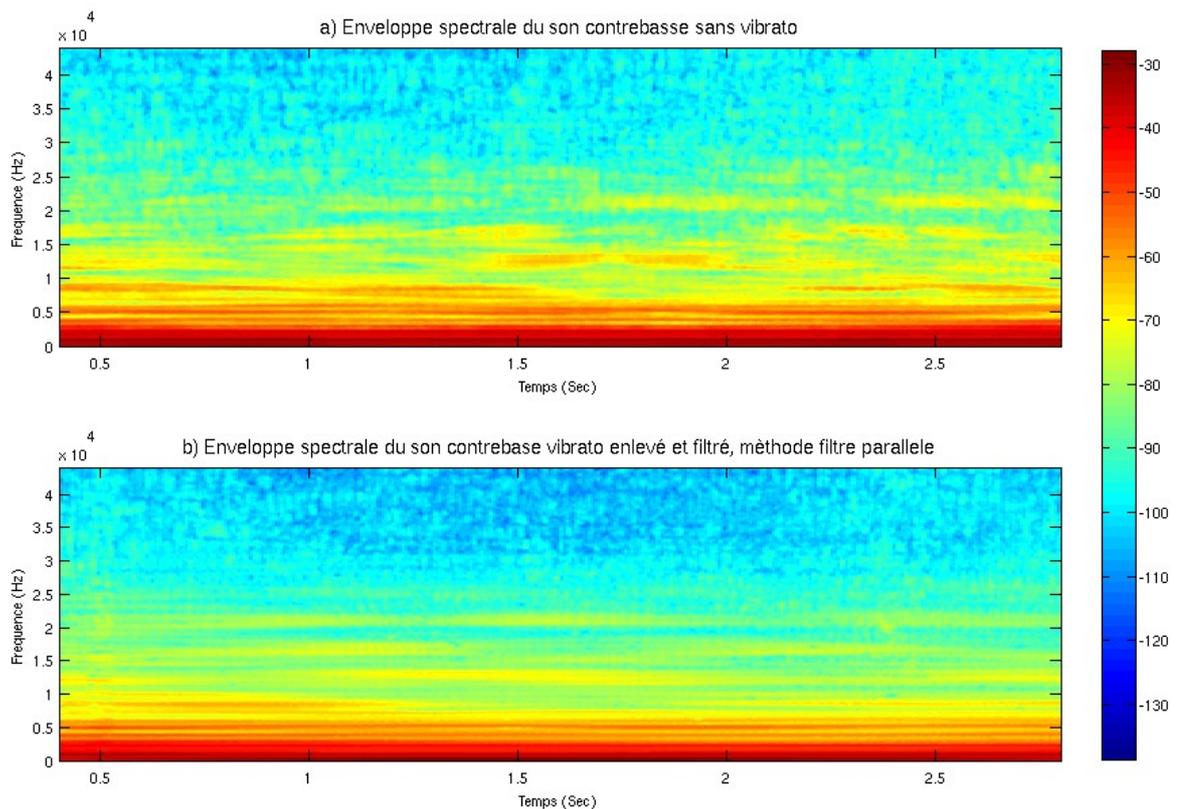
4 Basse de données disponible sur la page web <http://theremin.music.uiowa.edu/>



*Illustration 7: Comparaison des amplitudes pour la fréquence fondamentale et les 5 premiers partiels d'un son de contrebasse avec vibrato (a), de la même contrebasse mais sans vibrato (c), et les deux méthodes de filtrage après avoir enlevé le vibrato: dans (b) la méthode de filtrage par enveloppe spectrale, et en (d) la méthode de filtrage des composantes harmoniques en parallèle au filtrage du bruit.*

Autour de l'instant  $t=0.5$  seconde, on observe dans 7.d un pic d'amplitude dans les six composantes harmoniques. L'amplitude la plus grande d'entre elles est de 2.5 dB dans le 3<sup>ème</sup> partiel. Ces pics sont dus au changement abrupte entre les segments audio dont on ne filtre pas les amplitudes et les segments où on le fait. Pour choisir les limites du filtrage, afin d'éviter les discontinuités dans les amplitudes, il faut faire attention à la continuité des amplitudes entre le segment du signal qui ne sera pas traité et l'amplitude cible dans chaque partiel. On retrouvera cette problématique dans les limites du traitement pour modifier l'étendue du vibrato : pour obtenir une trajectoire de fréquence continue lors de la transposition, la fréquence cible doit être proche de celle de l'instant précédent. Dans la même figure, ce phénomène s'observe aussi près de l'instant  $t=3.3$  secondes, correspondant au moment où l'on finit le filtrage. Mais cette fois-ci, on l'observe dans les partiels 3 et 4.

Comme nous avons déjà vu dans la figure 4, la méthode du filtrage à partir du filtre fait avec les enveloppes spectrales, vu au chapitre 3, présente de bons résultats au niveau de l'enveloppe spectrale du son : on ne la voit plus modulée dans 4.c. Dans le étape de filtrage que nous proposons, nous utilisons le même filtre d'enveloppe pour la composante du bruit, donc nous nous attendons à avoir des résultats de performances égales à ceux de la méthode précédente. En fait, nous nous attendons à de meilleurs résultats car la partie harmonique a été traitée séparément ( les résultats ont été montrés ci-dessus ) . Cependant, l'utilisation des coefficient LPC et le bas ordre qu'on a choisi pourraient produire des différences. Nous n'avons pas mesuré ces différences car nous estimons qu'elles seraient négligeables au niveau perceptif. Le bruit se trouve suffisamment en dessous des amplitudes des principaux partiels pour que le bruit soit masqué par eux.



*Illustration 8: Comparaison de l'enveloppe spectrale d'un son de contrebasse joué sans vibrato (a) et de l'enveloppe spectrale d'un son vibrato qu'on a transposé pour éliminer le vibrato et qui a été ensuite filtré en parallèle par le filtre des composantes harmoniques et le filtre du bruit.*

Nous offrons dans la figure 8, une comparaison entre l'enveloppe spectrale d'un son de contrebasse joué sans vibrato ( figure 7.c) et l'enveloppe d'un son obtenu après avoir été transposé pour annuler le vibrato et filtré par le filtre parallèle ( figure 7.d ).

Les résultats obtenus lors du filtrage proposé dans ce stage nous semblent attractifs, mais la garantie d'un bon processus dans le traitement des sons musicaux ne peut se faire qu'à partir d'une évaluation perceptive. Ceci ne fait pas parti de nos objectifs et nécessiterait plus de temps. Nous considérons tout de même que des vibratos d'étendue moyenne ( 30 cents ) , comme ceux que l'on a vu dans ce rapport, peuvent être traités sans produire d'artefacts audibles. Mais, pendant que l'étendue du vibrato augmente ( par exemple à 100 cents ) , des artefacts produits par une transposition incorrecte du son lors de la modification de l'étendue seront perçus. En fait, lors que l'étendue augmente, la vitesse de changement de fréquence dans la trajectoire de la fréquence fondamentale augmente. Si cette vitesse est importante, la fréquence estimée ne sera pas stationnaire dans une fenêtre d'analyse, et son estimation peut donner des valeurs incorrectes ou valables seulement dans un point particulier de la fenêtre. Lors de la transposition, le changement de la fréquences fondamental et des partiels du son sera réalisé à partir de cette estimation, et la fréquence fondamentale résultante de cette transposition sera proche de la fréquence cible, mais il y aura une différence associée à cette erreur d'estimation. Cette erreur est propagée par tous les partiels du son, et étant donnée la périodicité du vibrato, l'erreur sera systématiquement reproduite à chaque période du vibrato.

Nous exposons par la suite les résultats de notre implémentation de la détection automatique du vibrato. Cet algorithme a été mis à preuve dans des segments audio ne contenant qu'une note. Puisque les résultats sur une note ne sont pas encore satisfaisants, nous n'avons pas évalué la détection sur des mélodies.

Dans la figure 9, nous présentons en rouge les segments sur lesquels on a trouvé l'effet vibrato. Comme référence on montre en bleu la trajectoire de la fréquence fondamentale. A partir de la figure, nous avons considéré que le vibrato commençait à 1.25 seconde et qu'il finissait vers 2.6 secondes. L'algorithme a trouvé un segment avec vibrato entre les secondes 1.4 et 2.5. Ce segment est plus petit que celui que nous avons estimé préalablement. D'autre part, l'algorithme a trouvé aussi du vibrato où il n'existe pas: vers 0.5 et 0.7 secondes. Entre ces deux segments, le fait de produire une transposition sur le premier rendrait des artefacts audibles. Ceci parce que la composante lente n'est pas encore arrivée à la valeur de la fréquence de la hauteur du son, et lors du calcul des coefficients de transposition, la valeur de la composante lente ne représentera pas la hauteur moyenne de la note jouée. La modification de la hauteur dans ce segment, sera incohérent avec la fréquence de la note et avec la modulation de fréquence du vibrato. Le résultat sera un artefact audible. Pour le deuxième segment, le problème ne sera pas aussi important. La composante lente est déjà arrivée à la valeur de la fréquence moyenne de la note, et alors la fréquence résultante lors de la transposition sera proche de celle d'origine. Dans la plus part des cas, le résultat de cette transposition peut être perçue comme un artefact.

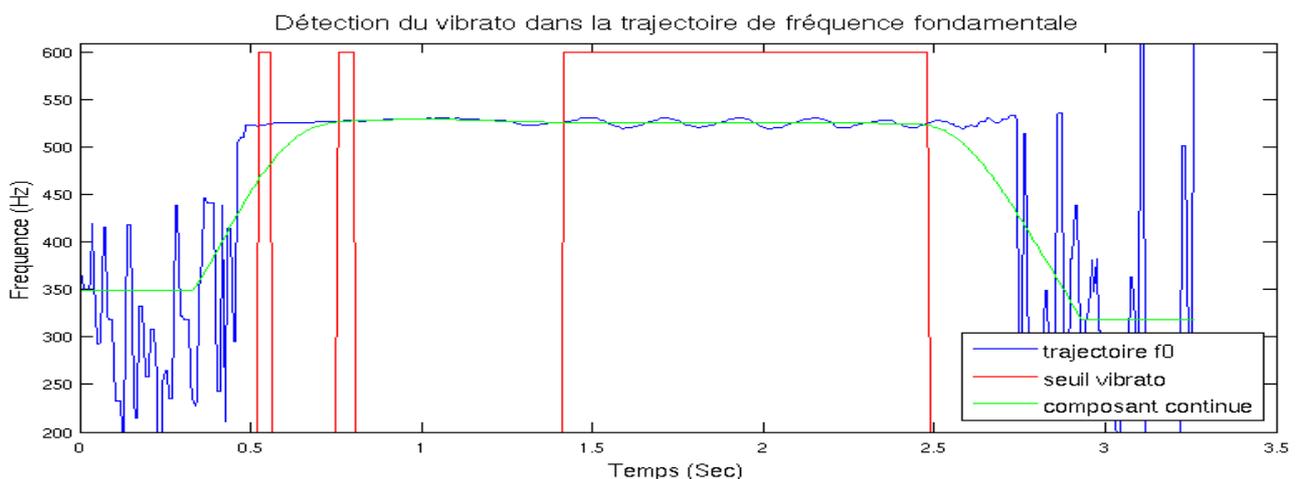


Illustration 9: Détection automatique du vibrato dans un son de flute.

Un autre problème associé à la modification de l'étendu du vibrato à partir de sa détection automatique sont les point de début et de fin du segment à traiter. Après avoir traité le segment contenant le vibrato, on attend que le son résultant redevienne continu. Cette-à-dire qu'il ne présente pas de changements abruptes de fréquence aux instants où les segments non traités et ceux qui ont été traité se suivent. Cette condition arrive pour les valeurs de la trajectoire de la fréquence fondamentale où  $f_0(n) = f_{dc}(n)$ . Si cette condition n'arrive pas, le changement abrupte dans cette trajectoire de fréquence fondamentale, et dans tous les partiels du son, pourraient être perçues comme des artefacts.

Dans la figure 11 nous montrons un zoom sur la détection automatique du vibrato de la figure 9. On observe dans la ligne rouge les points de transition entre le segment qui sera traité et celui qui ne le sera pas. Dans le début du segment à traiter ( première ligne rouge ) nous voyons la condition de

continuité s'accomplir. Par contre, à la fin de ce segment, nous voyons une différence entre la fréquence fondamentale et sa composante continue.

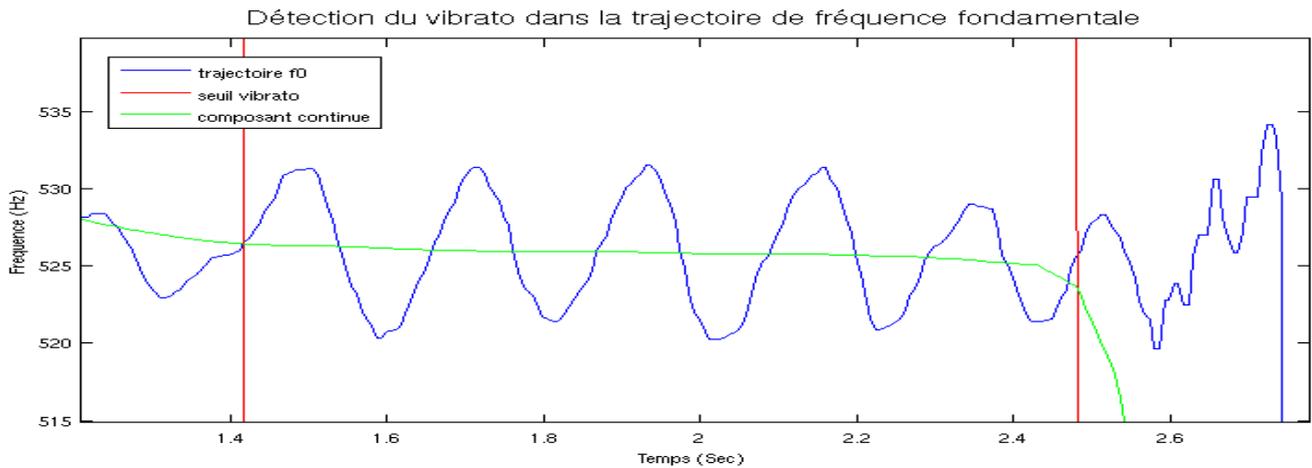


Illustration 10: à la fin du segment dans le 2.5 secondes, la trajectoire de la fréquence fondamentale après la modification de l'étendue montrera une discontinuité.

Finalement, l'estimation des paramètres du vibrato pour le son de flute des figures 3, 4 et 9, est présentée dans l'illustration 11. La fréquence du vibrato estimée a été pris en compte pour la détection du vibrato.

L'implémentation de la détection automatique du vibrato tel quelle a été implémenté ici ne est pas encore utilisable et doit être améliorée.

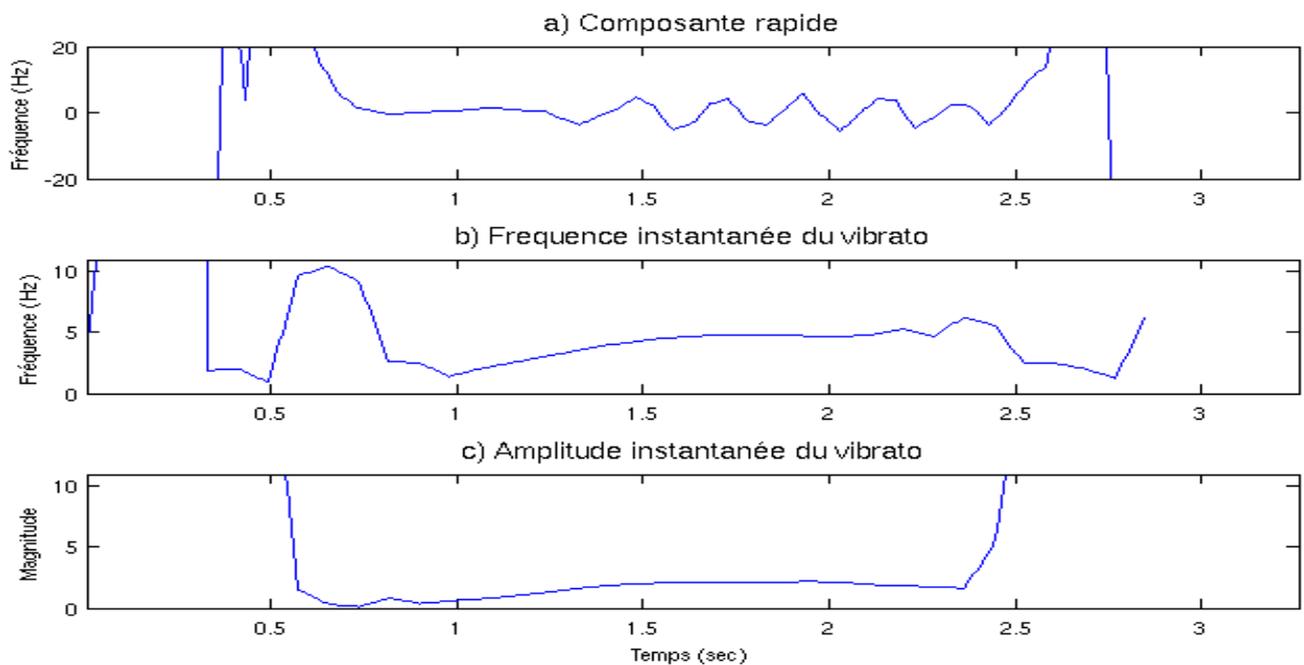


Illustration 11: à partir de la composante rapide  $f_{ac}$  l'amplitude et la fréquence du vibrato sont estimées.

## 6. Conclusion et perspectives

En filtrant les composantes harmoniques d'un côté et le bruit d'un autre, les modulations d'amplitudes lors du vibrato sont diminuées effectivement. Cette méthode présente de meilleurs résultats que la méthode du filtrage de tout le signal à partir de l'enveloppe spectrale. En effet cette dernière méthode présente de bons résultats pour le filtrage de l'enveloppe spectrale du bruit mais des résultats moins satisfaisants pour les composantes harmoniques.

Les erreurs associés au filtrage de l'enveloppe spectrales peuvent être associés au fait que chaque période du vibrato a différentes paramètres ( l'étendue, la modulation de amplitude et modulation d'enveloppe spectrale ).

Avec ces résultats on peut dire qu'un traitement des composantes harmoniques doit être réalisé indépendamment du traitement du bruit lors de la modification de la modulation d'amplitude vis à vis le changement d'étendu du vibrato.

Une des causes des artefacts qui ont émergé lors de la modification de l'étendue du vibrato dans le modèle *source/filtre* a été attribuée à un erreur provenant de l'analyse de la fréquence fondamentale. Quand l'étendu du vibrato est grande, la fréquence n'est pas invariante dans une fenêtre d'analyse et la fréquence estimé sera alors une moyenne de son évolution. Lors de la transposition, cet évolution sera transposée et cette trajectoire de la fréquence fondamentale présentera une discontinuité. Vu l'importance d'avoir une taille de fenêtre correcte, nous pourrions déterminer cette taille à partir de l'analyse de fréquence fondamentale ( « pitch-synchronous analyse » ) pour chaque fenêtre d'analyse.

Pour des sons enregistrés dans des conditions idéales, la transposition de la composante du bruit indépendamment de la composante harmonique ne s'avère pas nécessaire. Nous n'avons pas trouvé de composante non harmonique qui ne soit pas modulée lors du vibrato et qui soit modulée lors de la transposition du son. Nous croyons, que dans le cas de son enregistré dans un environnement bruyant la méthode de transposer séparément le bruit et les composantes harmoniques peut donner de bons résultats.

En ce qui concerne la détection automatique du vibrato, la méthode propose ici n'est pas encore utilisable. De notre expérience, nous pouvons tirer certaines conclusions :

Premièrement, le traitement du vibrato ne doit pas être réalisé dans les transitoires des notes. Afin d'éviter le traitement des transitoires, on pourrait les détecter et les signaler dans la fonction d'observation comme des endroits où le vibrato n'existe pas. On trouve dans [18] un modèle *sinusoidal plus bruit plus transitoire* qui on pourrait éventuellement utiliser.

Pour éviter qu'une discontinuité entre la fréquence fondamentale d'un segment non traité et la fréquence d'un segment transposé soit produite pendant l'étape de recherche du vibrato, on pourrait choisir le point où la  $f_0$  croise la composante  $f_{dc}$  la plus proche des extrémités du segment avec vibrato.

L'estimation de la fréquence du vibrato à partir du spectre d'amplitude de  $f_{ac}$  se présente comme une estimation peu robuste sur l'existence ou non du vibrato. Ceci parce que dans les segments où il n'y a pas de vibrato, il peut exister un pic spectral dans la bande du vibrato. Pour améliorer la détection, on pourrait faire un test d'harmonicité entre les pics du spectre. À la différence du spectre du vibrato, nous attendrions que dans le bruit on ne trouve pas de composante harmonique. D'autre part, le critère du rapport entre l'énergie contenue dans la bande de fréquence possible du vibrato sur l'énergie contenue ailleurs dans le spectre, peut présenter de meilleurs résultats ( voir équation (27) ).

Pour détecter les sinusoides d'un signal périodique à partir du spectre d'amplitude calculé avec la FFT de la trajectoire de  $f_0$ , la taille de la fenêtre d'analyse doit être d'environ 4 fois la période fondamentale de ce signal. Dans le cas d'une fréquence du vibrato de 6 Hz, la taille de la fenêtre doit être d'environ 0.6 secondes. Cette taille est importante car le signal du vibrato n'est pas stationnaire sur 0.6 secondes. Dans [19] il est proposé d'utiliser une modélisation AR (Auto Regressif) pour prédire les valeurs passées et futures d'une portion de signal de  $f_0(n)$  à partir de laquelle les coefficients AR sont calculés. Ainsi, il serait possible de réduire la taille de la fenêtre d'analyse et d'augmenter la précision fréquentielle.

Finalement, un seuil de l'amplitude du vibrato pourrait aussi servir comme variable pour la fonction d'observation du vibrato [10].

D'autre part, pour modéliser le vibrato, on devrait prendre un nombre limité de partiels et les composantes rapides de leurs amplitudes. On aurait aussi besoin de la composante rapide de la trajectoire de la fréquence fondamentale. Finalement, pour modéliser le bruit, on pourrait prendre les coefficients LPC trouvés lors de l'analyse d'enveloppe spectrale du bruit. Dans cette perspective, une étude pour déterminer quel est le meilleur ordre pour représenter le bruit, et combien de partiels sont nécessaires, doit être menée. Pour cela, on devrait réaliser une recherche à partir des échantillons des sons instrumentaux contenant du vibrato. Pour un travail de modélisation, on pourrait aussi, prendre en compte les transitions entre les notes. En effet, les musiciens anticipent les transitions et attaquent la prochaine note de façon cohérente avec le vibrato [24].

En termes créatives, il serait intéressant d'utiliser le modèle de vibrato d'un instrument pour un autre instrument. Ce processus est appelé « sound morphs » ou « hybrids » [21]. En effet l'auteur fait une différence entre la synthèse croisée ( « cross-synthesis » ) et la « hybridation ».

Pour obtenir des modifications du son avec vibrato à partir du modèle comme on le propose ici, il serait nécessaire d'éliminer le vibrato en opérant une transposition suivie d'un filtrage des amplitudes de partiels, comme nous l'avons fait dans le sous-chapitre 3.5 et 4.2 respectivement. Ensuite, une dilatation ou une contraction du temps peut être faite à volonté sur le signal sans vibrato. Dans le cas où on fait une contraction dans le temps, on peut simplement tronquer la composante  $f_{ac}$ , et appliquer la transposition correspondant pour avoir une trajectoire  $f_0$  avec vibrato. La modulation des amplitudes des partiels du vibrato peut être remise d'une façon analogue. Pour la dilatation temporelle, un modèle AR peut être obtenu d'après les composantes rapides des trajectoires de  $f_0$  et des amplitudes de partiels. L'obtention des échantillons manquants peuvent être obtenus à partir d'une prédiction linéaire. Pour le changement de la fréquence du vibrato sans changer la durée de la note, on devrait faire un changement de fréquence d'échantillonnage des paramètres du modèle du vibrato.

Finalement, une extension de ce travail pourrait être faite, afin d'appliquer les mêmes principes étudiés ici pour d'autres gestes ornementaux, comme par exemple le trémolo et le trille.

# Bibliographie

- [1] I. Arroabarren, X. Rodet, A. Carlosena ; *On the measurement of the instantaneous frequency and amplitude of partials in vocal vibrato*; dans « IEEE transaction on audio, speech, and language processing »; vol. 14, N° 4, juillet 2006.
- [2] J. Burred, C. Cella, G. Peeters, A. Röbel, D. Schwarz; *Using the SDIF sound Description interchange format for audio features*; dans « ISMIR 2008 – Session 4a – Data Exchange, archiving and evaluation »; 2008.
- [3] M. Charbit, G. Blanchet; *Traitement du signal audio-numérique*; notes du cours ATIAM, 2008.
- [4] B. Doval, X. Rodet; *Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMMs*; dans « Proceedings of ICASSP »; avril 1993.
- [5] N. H. Fletcher, T. D. Rossing; *The physics of musical instruments*; seconde édition; Ed. Springer; EEUU, 1998.
- [6] D. Gerhard; *Pitch extraction and fundamental frequency : History and current techniques*; Rapport technique, « Department of Computer Science » de l'université de Regina; Canada, novembre 2003.
- [7] J. Gilbert, L. Simon, J. Terroir; *Vibrato of saxophones*; dans « Journal of Acoustic society of america »; octobre 2005.
- [8] F. Harris; *On the use of windows for harmonic analysis with the discrete fourier transform*; dans « Proceedings of the IEEE », vol 66, No 1; EEUU, janvier 1978.
- [9] P. Herrera, J. Bonada; *Vibrato extraction and parameterization in the spectral modeling synthesis framework*; « Workshop on Digital Audio Effects ( DAFx'98 ) », 1998.
- [10] H. Järveläinen ; *Perception-based control of vibrato parameters in string instrument synthesis*; dans « International Computer Music Conference »; Sweden, 2002.
- [11] S. Maller; *Transformation des sons musicaux avec préservation des articulations musicales*; Rapport de stage sous la direction de A. Röbel et G. Laurent; IRCAM, France, juin 2008.
- [12] M. Mellody, G. H. Wakefield; *The time-frequency characteristic of violin vibrato : Modal distribution analysis and synthesis*; dans « Journal of Acoustic society of america »; janvier 2000.
- [13] A. V. Oppenheim, R. W. Schafer; *Discrete-time signal processing*; Ed. Prentice–Hall; EEUU, 1989.

- [14] A.Röbel, X.Rodet; *Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation*; dans « Proc. of the 8 conference on Digital Audio Effects »; Espagne, 2005.
- [15] A.Röbel; *Source filter modeling and spectral envelope estimation*; « Lecture on analysis, modeling and transformation of audio signals »; Institute of Communication Science TU-Berlin; août 2006.
- [16] A. Röbel; *Signal modifications using the STFT*; « Lecture on analysis, modeling and transformation of audio signals »; Institute of communication science TU-Berlin; août 2006.
- [17] A. Röbel; *Analysis/resynthesis with the short time fourier transform*; « Lecture on analysis, modeling and transformation of audio signals »; Institute of communication science TU-Berlin; août 2006.
- [18] D. Rocchesso; *Introduction to sound processing*; Université de Verona; 2003.
- [19] M. Rossignol; *Segmentation et indexation des signaux sonores musicaux*; thèse de doctorat de l'université Paris 6; dirigé par X. Rodet; France, 2000.
- [20] D. Schwarz; *Spectral envelopes in sound analysis and synthesis*; Rapport de « Master in computer science » à l'« Institute of computer science of the university of Stuttgart »; Dirigé par Xavier Rodet; l'IRCAM, 1998; Téléchargement disponible depuis le site internet: <http://recherche.ircam.fr/equipes/analyse-synthese/schwarz/da/>
- [21] X. Serra; *Musical sound modeling with sinusoids plus noise*; dans « Musical signal processing »; Ed. Swets & Zeitlinger Publisher.
- [22] J. Shonle, K. Horan; *The pitch of vibrato tones*; dans « Journal of Acoustic society of america »; janvier 1980.
- [23] R. Timmers, P. Desain; *Vibrato : questions and answers from musicians and science*; dans « Proceedings of the sixth ICMPC », Keele, 2000.
- [24] V. Verfaille, C. Gustavino, P. Depalle; *Perceptual evaluation of vibrato models*; dans « Proceedings of the conference on interdisciplinaire musicology ( CIM05 ) »; Canada, 2005.
- [25] C. Yeh, A. Röbel; *A new score function for joint evaluation of multiple  $f_0$  hypotheses*; dans « Proc. Of the 7<sup>th</sup> conference on digital audio effects »; Italie, octobre 2004.