Chloé Clavel
Stage DEA ATIAM Avril-Juillet 2003
(Paris 6, ENST, Aix Marseille II, UJF Grenoble I)

Séparation des sons musicaux.

Approche bayésienne et méthode de Monte-Carlo

-Sous la direction d'Yves Grenier-

Laboratoire de Traitement du Signal et des Images - ENST

Table des matières

Introduction

La séparation de sources sonores apparaît comme un maillon crucial pour répondre à un certain nombre de problèmes de traitement du signal tels que le débruitage, la compression, la reconnaissance automatique ou l'indexation de bases de données multimédia.

La séparation de sources est, dans ces contextes, une approche particulièrement pertinente car elle permet de décomposer un signal en un ensemble de sous-signaux élémentaires, privilégiant les composantes recherchées.

Dans le cadre de l'indexation, ces sous-signaux, versions approchées du mélange de signaux initiale peuvent être alors analysés de façon plus spécifique et plus efficace. En particulier dans le cas des contenus musicaux, détecter la présence de voix, de composantes rythmiques, d'instruments donnés voire de motifs sonores est une phase esssentielle pour caractériser automatiquement le style d'une oeuvre.

Une autre retombée s'est également avérée dans le domaine du codage : des codeurs de musique génériques basés sur l'analyse par modèles paramétriques atteignent désormais une haute qualité, même à très bas débit.

Il serait également possible de se diriger avers la transciption automatique de morceau de musique en séparant chacun des instrumentistes avant de détecter la partition qu'ils ont jouée.

Les enjeux sont également multiples dans le cadre du remixage d'un morceau de musique. La séparation de sources apparait alors comme l'étape préalable de demixage (cf paragraphe 1.1). Disposant des différentes sources séparément, le remixeur pourra par exemple atténuer ou accentuer une des sources selon l'esthétique qu'il veut donner au morceau.

Le problème de la séparation de sources peut se formuler de deux manières différentes selon que l'enregistrement du signal à traiter a été fait avec un ou plusieurs capteurs.

Dans une représentation multivoies, chaque instrument obtient en plus une signature spatiale unique, indice qui va pouvoir être exploité pour leur séparation. Les techniques de séparation sont alors largement inspirées des techniques utilisées dans le contexte des communications numériques.

Dans le cas de l'enregistrement monophonique, on doit recourir à des techniques qui identifient le contenu spectral du signal. Une étape indispensable sera donc l'estimation multipitch.

Présentation du département TSI.

Les recherches menées à l'ENST, au département TSI, concernent les différents domaines du traitement du signal et des images et leurs applications dans divers contextes des technologies de l'information. Dirigé par Henri Maître, ce département mobilise actuellement une cinquantaine de permanents, aussi bien chercheurs que membres administratifs et une soixantaine de thésards. Il se divise en cinq groupes de recherches :

· Le groupe Traitement et Interprétation des images. Son objectif est la mise en oeuvre de schémas complets de traitement, d'analyse et d'interprétation d'images, pour des applications en imagerie médicale (notamment l'imagerie cérébrale anatomique et fonctionnelle), en imagerie aérienne et satellitaire (avec un intérêt particulier pour l'imagerie radar et l'imagerie aérienne à très haute résolution des milieux urbains) et en imagerie des objets complexes tridimensionnels fixes ou animés.

· Le groupe Traitements statistiques et applications aux communications. Ce groupe travaille dans le domaine des communications, de la séparation de sources, de la reconstruction et la restauration d'images et du télétraffic.

· Le groupe Perception, apprentissage et modélisation. Il étudie le rôle des facteurs humains dans l'accès aux divers types d'information : la reconnaissance et l'identification de locuteurs pour la parole, la psychovision et l'imagerie de très haute qualité pour l'image, la structuration des documents pour l'écrit, les modalités perceptives dans l'appréhension de l'environnement et les interfaces multimodales.

· Le groupe Audio, acoustique et ondes. Il a pour vocation d'étudier la physique des ondes dans les deux domaines de l'optique et de l'acoustique, dans le but de modéliser la production des sons et leur perception (psychoacoustique, antennes acoustiques et prothèses auditives) et de stocker l'information dans les milieux optiques réinscriptibles.

· Le groupe Codage. Ce groupe travaille sur les techniques de compression de sources en vue de leur adaptation aux applications de l'audiovisuel et du multimédia. Les recherches qui y sont menées s'intéressent donc aussi bien aux domaines de l'image que de l'audio.

C'est Yves Grenier du groupe Audio, acoustique et ondes qui a assuré l'encadrement de mon stage.

Chapitre 1

Problématique.

1.1 Le projet DEMIX-REMIX

Ce projet a pour but de faire se rejoindre deux axes de recherche dans le domaine de l'audio :

- la spatialisation des sons qui donne à l'auditeur la sensation du relief sonore par le placement des sources dans l'espace, en utilisant plusieurs enceintes pour la restitution et en filtrant les signaux par les réponses spatiales de l'oreille (HRTF, Head Related Transfer Functions),
- la séparation des sources qui part d'un signal dans lequel plusieurs sources sont mélangées, et extrait les différentes sources.

En jargon d'ingénieur du son, " remix " est le raccourci pour signifier une opération de post-production qui vise à reconstituer un mixage différent de celui précédemment réalisé à partir des bandes originales.

Notre objectif serait de proposer à l'auditeur d'effectuer lui-même cette opération de mixage en lui laissant la possibilité de placer les sons à sa convenance.

Nous devons donc élaborer un dispositif permettant de combiner la spatialisation par HRTF et une restitution sur plusieurs enceintes (par exemple avec les 5+1 enceintes du son surround, voire avec plus d'enceintes). L'étape préparatoire au " remix ", est la partie " demix " elle consiste à isoler à partir des enregistrements originaux, les éléments de base à " remixer ", par exemple les instruments. Elle s'appuiera sur les techniques de séparation de source dans le domaine fréquentiel en mono-capteur, et les méthodes spatiales ou spatio-temporelles dans le cadre multivoies.

" Demix " et " Remix " peuvent se faire conjointement, au niveau de l'équipement de l'auditeur. On peut aussi envisager qu'ils se fassent aux deux bouts d'un canal véhiculant des signaux MPEG-4, car ce codage permet de coder un signal en le décomposant en plusieurs composantes, chacune pouvant d'ailleurs être codée avec une technique différente des autres composantes.

1.2 Problématique de la séparation de sources.

La séparation de sources, telles que les instruments de musique, dans le cas d'une prise de son monocapteur est un travail délicat, car de nombreux problèmes se posent rapidement. Une étape nécessaire et non moins fastidieuse pour espèrer pouvoir séparer des instruments de type harmonique est l'estimation multipitch. Une fois les notes présentes dans le signal de musique déterminées, ainsi que leurs caractéristiques, on peut alors songer à attribuer à chaque instrument les notes qui lui correspondent et ainsi resynthétiser chacune des sources séparément.

1.2.1 La détermination du nombre de sources.

La première difficulté réside en la détermination du nombre d'instruments présents dans le signal. Puisque l'on ne dispose que d'un seul capteur, tous les intruments de l'enregistrement se retrouvent mélangés, les harmoniques des différentes notes se retrouvent par conséquents mélangés, et peuvent même se superposer. Ce qui arrive fréquemment pour des notes jouées à la quinte juste ou à la tierce majeure.

Prenons l'exemple suivant : les trois premiers harmoniques d'un $Do4$ sont $f_1 = 261.6Hz$, $f_2 = 523.2Hz$ et $f_3 = 784.8Hz$ et ceux du $Sol5$ sont $f_1 = 784.3Hz$, $f_2 = 1568.6Hz$ et $f_3 = 2352.9Hz$.

Le troisième harmonique de Do4 et la fondamentale de Sol5 ne diffèrent que de 0.5 Hz.

On imagine bien le nombre d'instruments croissant, les harmoniques présents dans le signal pourront provenir simultanément de plusieurs instruments. La déterminaton du nombre d'instruments du mélange sera donc difficile à mettre en oeuvre, d'autant plus que certains instruments sont monophoniques tandis que d'autres sont polyphoniques. Plus encore tout instrument ne joue pas du début à la fin du morceau : il faut donc détecter les moments auxquels tel ou tel instrument est présent.

1.2.2 Les rapports d'intensité entre les différentes sources.

Lorsqu'un instrument joue plus fort que les autres (un soliste par exemple), un phénomèe de masquage apparaît : les harmoniques des instruments de niveau d'enregistrement plus faible risque d'être considérés comme du bruit et non comme de l'information utile, par comparaison avec celles de l'instrument de niveau plus fort. Pour notre problème qui est de séparer les différentes sources, ce cas, qui arrive fréquemment dans les enregistrements, pose un réel problème : une source non détectée ne pourra pas alors être isolée.

1.2.3 Le problème de l'octave.

Ce problème se pose lorsqu'on connait une suite de partiels et qu'on veut déterminer la note jouée, c'est à dire la fréquence fondamentale correspondante. Pour certains instruments, la fréquence fondamentale est d'amplitude négligeable par rapport aux partiels suivants, elle n'est pas forcément présente (ou mesurable) dans le spectre : on parle alors du phénomène de le fondamentale cachée. On considère le plus souvent que la fréquence fondamentale correspond au plus grand diviseur commun des fréquences des différents partiels. Il faut cependant être en mesure d'attribuer à chaque instrument ses partiels avant de s'autoriser l'utilisation de ce critère.

Un autre problème se pose lorsque plusieurs instruments jouent en même temps la même note décalée d'une ou plusieurs octave. Seule la note de fondamentale la plus grave risque d'être considérée, puisque tous les partiels de la note la plus haute seront confondus avec ceux de la note la plus grave, et pourront être considérées comme partiels de cette dernière. Seules, peut-être, des informations a priori sur les caractéristiques des instruments présents dans l'enregistrement, pourront permettre de séparer deux notes décalées d'une ou plusieurs octaves.

1.2.4 Problèmes de l'inharmonicité des instruments.

Il faut également être conscient du fait que si la plupart des instruments à sons entretenus sont harmoniques (les harmoniques correspondent aux multiples de la fréquences fondamentales), certains instruments, comme par exemple les instruments à corde tels que le piano ou la guitare, sont inharmoniques. Il faudra donc prendre en compte cette inharmonicité lorsque l'on cherchera à modéliser un instrument, en vue de l'estimation féquentielle.

1.2.5 Séparation des sources et reconnaissance des instruments.

Dans l'état de l'art actuel, les problèmes de séparation de source et d'identification des instruments sont traités indépendamment. Cependant, il est facile d'imaginer à quel point le travail de séparation pourrait être facilité, si on était capable de reconnaitre les différents instruments d'un enregistrement, si on disposait d'informations sur leurs caractéristiques temporelles et fréquentielles.

1.2.6 Les conditions d'enregistrement.

Les conditions d'enregistrement ont une influence certaine sur la détection des différentes sources. Par exemple, la réverbération présente dans les enregistrements pose un problème, en ceci qu'un long temps de réverbération agira de même qu'une source sonore supplémentaire, c'est à dire comme une nouvelle

voix, qui plus est bruitée.

1.3 Etats des travaux de l'ENST département TSI dans le domaine.

Plusieurs approches ont été développées pour traiter de la séparation des signaux, en distinguant le cas où le mélange de signaux est connu par enregistrement de plusieurs capteurs, et celui où on ne dispose que d'enregistrement monophonique.
Dans le cas multi-capteurs ([4]) les techniques reposent sur le concept de la formation de voies adaptative : il s'agit de construire plusieurs filtres agissant temporellement comme spatialement, chacun d'entre eux fabriquant un lobe de directivité pointant vers l'une des sources. Dans cette approche, la nature de chaque composante individuelle, que ce soit son enveloppe temporelle ou son contenu spectrale, n'intervient pratiquement pas. Seules interviennent les caractéristiques de la propagation : positions des capteurs et des sources, fonction de Green décrivant l'acoustique du local.

A partir d'un enregistrement monophonique, les approches choisies pour identifier le contenu spectral ont été des approches par maximum de vraisemblance ([18]) que nous développerons dans la suite du rapport.

1.4 Objectifs.

L'objectif est de fournir une vision globale des différentes méthodes utilisées dans le cadre de l'estimation multipitch en vue de la séparation des instruments.

Le travail de Paul Joseph Wamsley dans sa thèse [24] a particulièrement retenu notre attention pour des raisons que nous développerons dans la partie suivante. Son travail utilise la structure hiérarchique des modèles bayésiens afin de modéliser le lien entre la description haut niveau et bas niveau du son.

Chapitre 2

Etat de l'art : les différents algorithmes de séparation des sources.

Cet état de l'art est destiné à donner une vue d'ensemble des méthodes utilisées pour la séparation de sources dans le cas monocapteur et pour l'estimation multipitchs. Mais nous nous intéresserons également aux articles des modèles perceptifs mis en oeuvre lors de notre écoute, ainsi que des différents modèles de signal utilisés pour représenter le son musical.

2.1 Modèles perceptifs.

Que ce soit dans le cadre de la reconnaissance d'instruments de musique [7], ou celui de la séparation des instruments, il peut être intéressant, avant toute chose, de mettre en évidence les caractéristiques du signal audio qui entrainent notre perception des évènements séparés en un seul, un regroupement perceptif.

La problématique de la séparation de sources implique une réflexion préalable sur le comportement perceptif de notre oreille et sa capacité à séparer les sons et ainsi choisir les sons à séparer en fonction.

Par ailleurs, nous avons vu, dans le chapitre précédent, que la séparation d'instruments harmoniques passe nécessairement par une étape d'estimation multipitch. Nous nous intéresserons donc à comment notre oreille perçoit la hauteur d'une note d'un instrument, ainsi qu'à nos capacités à reconnaitre les différents instruments.

2.1.1 Sensation de hauteur.

La perception que l'on a de la hauteur d'une note, dépend de nombreux paramètres comme la période du son, la tessiture, l'intensité, la durée et la composition spectrale.

Tous ces paramètres sont perçus par notre oreille selon des processus complexes(tonotopie cochléaire ...). En effet, la première étape pour notre oreille consiste en la transmission d'une vibration acoustique à la cochlée, organe de l'oreille interne. La cochlée réalise alors une analyse de la fréquence et une compression dynamique. La vibration acoustique est alors transmise à la membrane basilaire (membrane située à l'intérieur de la cochlée), chaque fréquence du signal d'entrée excite une partie différente de la membrane. Les mouvements en ces différents points de la membrane permettent la transmission d'impulsions nerveuses qui sont envoyées par le nerf auditif vers le cerveau. C'est ensuite le cerveau qui permet d'interprêter ce flot d'impulsions comme la hauteur du son.

C'est cette interprêtation qui est difficile à modéliser, c'est à dire quels sont les indices prédominants dans la sensation que l'on a de la hauteur d'un son ?

2.1.2 Sensation de timbre.

Saisir les aspects physiques du timbre du son est un problème délicat, auquel les acousticiens se heurtent encore. Le timbre d'un son est associé en effet par notre oreille soit au matériau de l'objet ayant produit le son, soit à sa forme ou encore à son mode d'excitation. Notre première écoute d'un son est causale : nous cherchons, avant tout à identifier la source. Puis si les deux sons sont produits par des sources semblables, nous cherchons à distinguer des variations dans la durée de l'attaque ou dans le contenu spectral, afin de caractériser les différences de couleurs sonores entre les deux sons.

L'analyse spectrographique de toutes les notes d'un même instrument montre d'importantes variations spectrales alors que notre oreille est capable de les percevoir comme provenant d'un même instrument, d'extraire donc un invariant : le timbre de l'instrument.

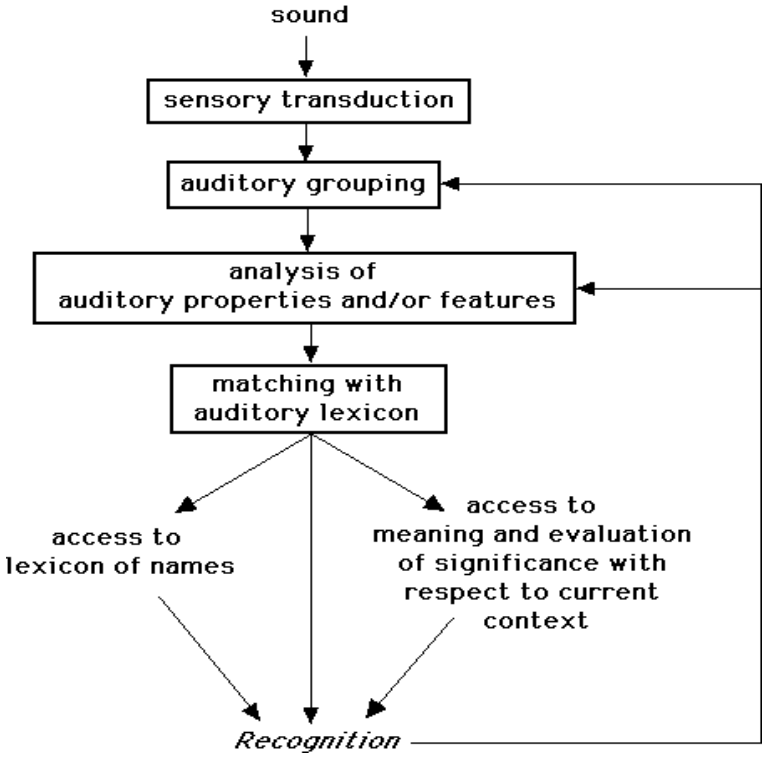
Caractériser le timbre d'un instrument pour les séparer est donc un problème qu'il va falloir surmonter au mieux en cherchant dans le signal les caractéristiques qui nous font percevoir un invariant dans les différentes notes d'un instrument.

2.1.3 Catégorisation auditive des sources sonores.

Nous nous réfèrerons à [7], et à [5], en ce qui concerne les modèles perceptifs de catégorisation des sources sonores.

Selon Mac Adams [2], le processus de reconnaissance des différentes sources sonores se décompose en pluseurs étapes. Il suppose que le lien entre les qualités perceptuelles des différentes sources, leur représentation abstraite dans notre mémoire, leur identité, est le résultat d'un processus séquentiel rétroactif :

FIG. 2.1 – *Les différentes étapes du processus de reconnaissance des différentes sources sonores selon Mac Adams.*



Dans la phase de regroupement auditif, le flot d'informations arrivant au cerveau, après l'étape de transduction réalisée par l'oreille et décrite dans le paragraphe 2.1.1, est transformé en représentations auditives séparées, correspondant à chacune des sources sonores. Ce qui signifie que les composantes appartenant à chacune des sources sont intégrées dans un même groupe.

2.2 Le modèle de signal.

Dans les applications qui nous concernent, le signal de musique est le plus souvent modélisé par des sinusoïdes, les méthodes d'analyse pouvant être paramétriques ou non paramétriques.

Le modèle sinusoïdal est en effet particulièrement adapté pour représenter la parole et les instruments de musique harmoniques. Ces modèles permettent d'obtenir les composantes fréquentielles du signal et sa reconstruction est alors assez satisfaisante. Les modèles sinusoïdes plus bruit permettent d'introduire la composante bruit qui est également essentielle pour la plupart des instruments de musique.

2.2.1 Modèles sinusoïdaux non paramétriques.

Dans le cas d'une analyse non paramétrique, le modèle sous-jacent n'est pas forcément explicité. Les inférences faites pour extraire des caractéristiques au signal sont faites à partir de transformations, de manipulations (ex : filtrage). Dans [20], le signal est modélisé sous la forme de sinusoïdes et son analyse se fait à l'aide de la TFCT (Transformée de Fourrier à Court Terme). Cette analyse permet de déterminer pour chaque trame les fréquences et les amplitudes de chaque pic du spectre.

Ce modèle nécessite des fenêtres assez longues (en fonction la longueur d'ondes de la différence des fréquences des deux notes, soit ici 300 ms) pour pouvoir estimer deux fréquences proches. En effet, la largeur du lobe principal du sinus cardinal est inversement proportionnel à la longueur de la fenêtre et donc la résolution fréquentielle est de l'ordre de grandeur de l'inverse du temps total d'analyse.

Le pas entre les fenêtres d'analyse doit être petit afin de de détecter de manière précise les harmoniques les plus élevés, témoins de la modulation de fréquence ou d'amplitude.

2.2.2 Les modèles sinusoïdaux paramétriques.

Le modèle est alors explicite, construit et paramétré par un ensemble de caractéristiques physiquement significatives et qu'il va falloir estimer.

Modèle de signal utilisé par la méthode ESPRIT [15].

Le signal de musique, compte tenu de son aspect périodique et de son amortissement au cours du temps est modélisé par une somme de M sinusoïdes exponentiellement amorties, ce qui de façon formelle s'écrit :

$$x(t) = \sum_{m=1}^{M} a_m \exp(-d_m t) \cos(2\pi f_m t + \phi_m)$$

où $x(t)$ est le signal discret observé aux instants $0, .., N-1$. M est l'ordre du modèle, a_m sont les amplitudes des sinusoïdes, d_m sont les coefficients d'amortissement réels, $f_m \in [-\frac{1}{2}, \frac{1}{2}[$ sont les fréquences normalisées, et $\phi_m \in [-\pi, \pi[$ sont les phases à l'origine. L'équation peut être réécrite en utilisant les notations complexes :

$$x(t) = \sum_{m=1}^{M} (\alpha_m z_m^t + \alpha_m^* z_m^{t*})$$

avec α_m les amplitudes complexes et z_m les pôles complexes.

L'estimation des paramètres du modèle consistera alors en l'estimation des coefficients d'amortissement et des fréquences des sinusoïdes.

L'avantage d'un tel modèle est que l'estimation des fréquences fondamentales n'est alors plus contrainte en théorie par la largeur des lobes principaux. Elle permet donc de travailler avec des fenêtres d'analyse beaucoup plus courtes (environ 30 ms).

Modèle de signal utilisé par les méthodes du maximum de vraisemblance [18], [21].

Le modèle utilisé dans [18] est une somme de sinusoïdes à laquelle on ajoute un bruit auto-régressif dont l'ordre sera à estimer.

$$x(t) = \sum_{m=1}^{M} a_m \cos(2\pi f_m t) + b_m \sin(2\pi f_m t) + v(t)$$

où $v(t)$ est un bruit AR gaussien d'ordre P.

2.3 Les méthodes de séparation.

L'objectif de ces différentes parties est d'expliquer différentes méthodes et de voir comment celles-ci résolvent les différents problèmes rencontrés lors de l'estimation des paramètres du modèle.

2.3.1 Une méthode par classification des composantes harmoniques.

Dans l'article [20], Tuomas Virtanen et Anssi Klapuri proposent une classification des différentes composantes sinusoïdales en paquets quasi-harmoniques. Cette classification se fait à l'intérieur même d'une trame mais peut être également aidée par les composantes des blocs voisins, dans un schéma de poursuite de paquets en passant de blocs en blocs au cours du temps (voir paragraphe 2.3.5).

Afin de simplifier le problème de l'estimation multipitch, cet article traite le cas où les instruments présents dans le mélange sont harmoniques et n'ont pas la même fréquence fondamentale.

Son algorithme de séparation des sources musicales se décompose en différentes étapes :

Analyse du signal.

-Analyse en TFCT pour déterminer l'amplitude et la fréquence correspondant aux pics des différentes trames suivie de la méthode F-test (cf [22]) permettant de sélectionner les pics appartenant à la partie harmonique du signal.
-extraction des trajets de partiels.
-interpolation des trajectoires en supprimant les ruptures causées par les modulations d'amplitude, les coupures et les bruits.

Mesure de distances perceptuelles entre les différentes trajectoires sinusoïdales.

Les distances prennent en compte un certain nombre d'indices perceptuels la proximité spectrale, la concordance harmonique, les changements synchrones tels que la modulation d'amplitude, de fréquence et les mouvements équidirectionnels du spectre.

Plus de détails sur ces indices se retrouvent dans [5].

Classification des composantes sinusoïdales par sources sonores à l'aide de la distance perceptuelle.

A partir d'un calcul de distance entre les différentes trajectoires, cette étape consiste à regrouper les trajectoires les plus proches en différente classes.

Détection des trajectoires qui se heurtent entre sources.

Les harmoniques des deux sources sonores peuvent se recouvrir, notamment lorsque les deux notes à séparer sont à la quinte juste ou à la tierce majeure. Pour détecter les trajectoires qui se rencontrent, Virtanen et Klapuri proposent de calculer la distance harmonique entre les trajectoires .

L'utilisation de trajectoires dans cet article correspond en réalité à faire de la poursuite des paramètres sur plusieurs fenêtres. Les fenêtres qui correspondent au même harmonique d'un même son sont regroupées et permettent de déterminer la trajectoire d'un partiel. La classification, ici, ne se fait pas fenêtre par fenêtre mais trajectoire par trajectoire.

2.3.2 Méthodes par maximum de vraisemblance.

Parmi les différents estimateurs, l'estimateur dit du maximum de vraisemblance joue un rôle essentiel pour l'estimation des fréquences des sinusoïdes. Il s'agit à l'origine d'une généralisation d'une idée due à Gauss qui postule que le meilleur modèle est celui qui donne à l'évènement observé la plus grande probabilité.

Nous avons décrit précédemment le modèle sinusoïdal utilisé pour cette méthode : le signal observé $s(n)$est modélisé comme la somme d'un signal purement périodique $x(n)$, de période T et d'un bruit blanc gaussien $e(n)$ de variance σ^2. L'estimation des paramètres du modèle se fait en maximisant la vraisemblance ou log-vraisemblance des observations :

$$p(s/T, x, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} exp(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(s(n) - x(n))^2)$$

Le maximum de vraisemblance consiste alors à maximiser par rapport à T, x et σ^2 la log-vraisemblance définie par :

$$L(T, x, \sigma^2) = ln(p(s/T, x, \sigma^2))$$

L'estimateur du maximum de vraisemblance possède l'avantage d'être robuste et permet de traiter le cas d'observations bruitées.

C'est pourquoi, dans la problématique de séparation de sources, chaque source d'un mélange peut être vue comme noyée dans un environnement bruité,

et la méthode du maximum de vraisemblance intervient comme une solution efficace d'estimation des différents paramètres de chacune des sources.

C'est cette approche qui est choisie dans [18], avec un critère de pénalité pour le nombre d'harmoniques.

Choix de l'ordre du modèle.

Un problème qui se pose lors de la mise en oeuvre des méthodes précédentes est le choix de l'ordre du modèle, en d'autres termes combien de sinusoïdes faut-il choisir pour estimer au mieux le signal ? L'intuition pousse à croire qu'une solution efficace à ce problème serait de surestimer la dimension de l'espace signal. Malheureusement, cette solution n'est pas forcément la meilleure. En effet, surestimer la dimension de l'espace signal consisterait en fait à considérer que le bruit appartient en partie à cet espace, ce qui est en contradiction avec le modèle imposé. Choisir une valeur de M trop grande perturberait l'analyse spectrale. De nombreux estimateurs existent et parmi les plus connus on compte le critère AIC (Akaike Information Criterion) et le critère MDL (Minimum Length Description).

Le critère d'ordre AIC.

Le critère AIC développé par Akaike est défini par :

$$AIC = -2\ln(f(d,\hat{\phi})) + 2k$$

où $\hat{\phi}$ est l'estimé du maximum de vraisemblance du vecteur de paramètres ϕ du modèle et k le nombre de paramètres libres de ϕ. Le critère AIC correspond à la somme de la log-vraisemblance de l'estimateur du maximum de vraisemblance des paramètres et d'un terme de correction, correspondant au critère d'ordre.

Le critère d'ordre MDL - [9], [17].

L'approche de Rissanen a quant à elle été développée à partir de l'idée que chaque modèle peut-être utilisé pour encoder les données. Le critère MDL sélectionne alors le modèle qui va fournir la longueur de code minimale :

$$MDL = -\ln(f(d,\hat{\phi})) + \frac{1}{2}k\ln(N)$$

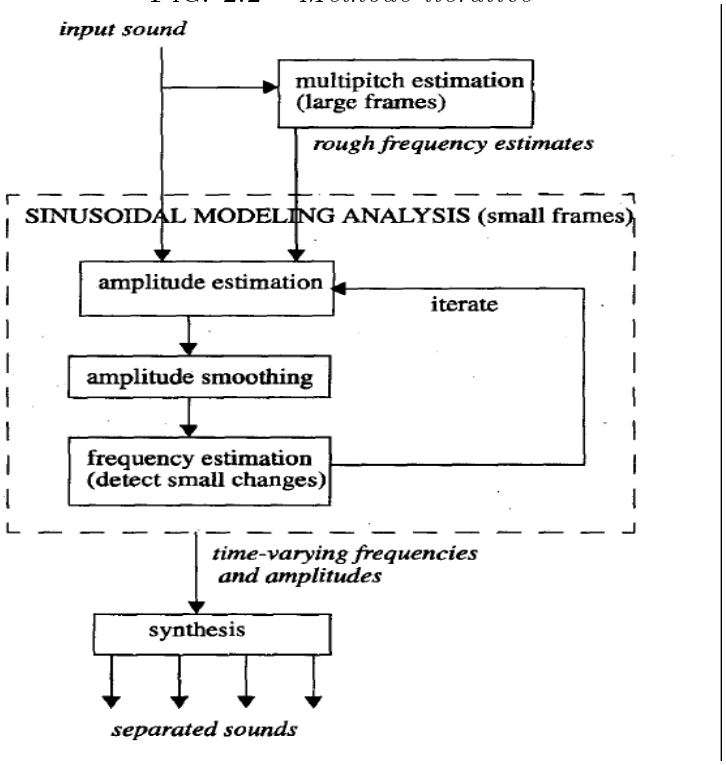
Méthode itérative.

L'algorithme décrit dans [20] ne fonctionne pas si on doit séparer plus de deux sons. Dans [21], Klapuri et Virtanen proposent une méthode itérative, utilisant le maximum de vraisemblance qui se décomposent en différentes étapes.

- Estimation multipitch sur des fenêtres longues estimation du nombre de sons présents, la fréquence fondamentale de chaque son, et les fréquences de chaque composante harmonique de chaque son (méthode décrite dans [1])

Cette étape consiste en l'initialisation du sytème d'analyse sinusoïdale qui estime les fréquences et amplitudes du modèle de manière itérative en utilisant des modèles d'enveloppes spectrales.
- Estimation itérative des paramètres sur des fenêtres plus courtes. Cette étape estime de manière plus précise les paramètres des différentes composantes du signal selon le schéma suivant :

FIG. 2.2 – *Méthode itérative*



L'estimation des amplitudes, des phases et des fréquences se fait par moindres carrés en utilisant le modèle sinusoïdal.

2.3.3 Méthode ESPRIT : [15]

L'estimation des fréquences et des coefficients d'amortissement des sinusoïdes se fait par le biais de la méthode ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) exploitant des propriétés d'invariance rotationnelle du sous-espace signal. L'estimation des amplitudes et des phases à l'origine est faite par un calcul du type moindres carrés.

Ce type de méthode permet d'avoir une meilleure résolution que celle de Fourier et est ainsi mieux adapté en théorie au problème de séparation où la présence de plusieurs signaux implique de faibles écarts entre les pics spectraux.

La discrimination entre deux harmoniques très proches de deux notes possédant une harmonicité élevée devient alors théoriquement possible.

2.3.4 Estimation bayésienne.

Nous ne détaillerons pas ici cette méthode qui sera amplement développée dans le chapitre 3. En résumé, cette méthode utilise, comme le maximum de vraisemblance, un modèle harmonique dont les paramètres sont à estimer. Pour cela, au lieu de maximiser la vraisemblance des observations connaissant le modèle, cette méthode choisit de considérer la loi a posteriori, c'est à dire cherche à maximiser la probabilité du modèle connaisant les observations.

2.3.5 Poursuite sur plusieurs fenêtres

Les méthodes précédentes peuvent être améliorées par la poursuite des paramètres sur plusieurs fenêtres (tracking en anglais). La poursuite étudie, en effet, l'évolution temporelle des fréquences. En utilisant par exemple les informations extraites des fenêtres voisines de notre fenêtre on peut ainsi améliorer la qualité de son analyse.

L'algorithme SINTRACK

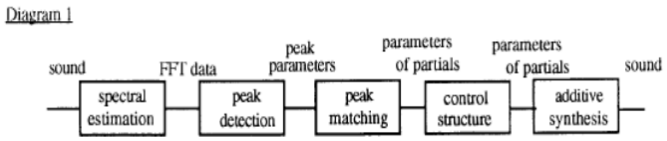
Dans [16], l'algorithme de poursuite utilisé est l'algorithme SINTRACK, décrit dans [13]. C'est un algorithme de poursuite de sinusoïdes amorties. La méthode SINTRACK permet de détecter les changements dans le modèle de signal et de déterminer la variation temporelle des différents paramètres. Sintrack est envisagé selon une problématique de générale de détection et d'estimation conjointes. La procédure de détection repose sur l'observation permanente de la valeur absolue de l'erreur de prédiction rétrograde, laquelle renseigne, par son allure caractéristique sur les éventuelles ruptures de modèle et la nature des non-stationnarités du signal.

Poursuite par HMM.

Dans [12], X. Rodet, P. Depalle et C. Garcia utilisent les Modèles de Markov Cachés (HMM) pour faire de la poursuite sur plusieurs fenêtres pour la synthèse additive de sons. La structure d'ensemble de leur algorithme est décrite dans la figure suivante :

L'article s'intéresse donc à la partie "peak matching" de l'algorithme d'analyse/synthèse qui consiste au classement des différents pics détectés dans la

FIG. 2.3 – *Processus d'analyse-synthèse additive.*



transformée de Fourier du signal à l'aide d'une poursuite par HMM.

La poursuite consiste ici en une identification des trajectoires des pics spectraux successifs. Une trajectoire est considérée comme une séquence temporelle de pics spectraux qui satisfait certaines contraintes de continuité.

La méthode alors utilisée est d'identifier les trajectoires dont les pentes d'amplitudes et de fréquences évoluent régulièrement au cours du temps.
Voici le modèle de Markov décrit dans cet article :

Soit h_k le nombre de pics contenus dans la fenêtre k $P_k(j)$ avec $0 \leq j \leq h_k$ ordonnés en fréquences et étiquetés par un indice $I_k(j)$. Lorsqu'un pic est considéré comme non valable (c'est à dire dû à un bruit ou aux lobes de la frenêtre d'analyse) on pose $I_k(j) = 0$

- Les données d'observations discrètes sont définies par une paire formée par le nombre de pics comptés dans la fenêtre $k-1$ et dans la fenêtre k,

(h_{k-1}, h_k)
- Au temps k un vecteur d'état S_k est défini par une paire ordonnée de vecteurs

$$S_k = (I_{k-1}, I_k)$$

où

$$I_k = \begin{pmatrix} I_k(0) \\ I_k(1) \\ . \\ . \\ I_k(h_k)) \end{pmatrix}$$

- Les amplitudes et les fréquences des pics correspondant à chaque $P_k(j)$ sont considérés comme les paramètres des HMM et sont utilisés pour calculer la probabilité de transition entre deux états S_{k-1} et S_k. Pour chaque pic, on évalue un critère d'assortiment ("matching criterion") qui dépend des pics des fenêtres voisines ayant le même index.

A partir de ce modèle, déterminer les trajectoires des partiels revient à trouver la séquence optimale d'états de laquelle dérive la séquence d'observation. Cette étape se réalise en pratique par l'algorithme de Viterbi.

2.4 Séparation des composantes bruitées.

L'isolation des composantes de bruits à chaque instrument nécessite un traitement distinct. Dans le cas de notes simultanées provenant d'un même instrument (accords du piano) le problème se pose autrement : un simple remixage de la composante bruit entre les différentes notes pourraient suffire.

En revanche, dans le cas où chaque note correspond à un instrument différent, chaque instrument produit alors une composante bruit (composante sans structure harmonique) qui lui est propre. Ainsi les composantes bruit du violon (frottement de l'archet), de la clarinette (excitation de l'anche), ou du piano (choc du marteau contre la corde) vont présenter des caractéristiques différentes.

La séparation de chacune de ces composantes, consiste donc à identifier les mélanges de modèles stochastiques.

2.5 La synthèse séparée des sources.

Une fois que toutes les trajectoires appartenant à deux sources différentes sont séparées, on peut représenter le signal synthétique et le resynthétiser. Dans [21] et [20], les fréquences, les amplitudes et les phases sont interpolées

fenêtre par fenêtre et le signal temporel est obtenu en sommant les composantes harmoniques de chaque son.

Chapitre 3

L'estimation multipitch par les modèles Bayésiens et la méthode de Monte-Carlo.

Le but de cette partie est d'étudier cette méthode d'estimation multipitch et de voir son intérêt dans le cadre de notre objectif final qui est la séparation d'instruments. Pour cela nous choisirons de traiter le cas où les instruments présents dans l'enregistrement et que l'on veut séparer sont connus.

Notre travail se fera à partir de différents articles traitant des modèles Bayésiens et des MCMCs pour l'estimation des sinusoïdes, qu'ils soient appliqués directement à notre problématique ([14],[24],[19]) ou à une problématique plus générale ([3]).

Nous allons, dans un premier temps, décrire la méthode sur le plan théorique, afin de bien comprendre ses avantages, et ses limites. Puis, nous illustrerons cette méthode en l'implantant étape par étape pour nous diriger peu à peu vers une méthode de séparation des instruments.

3.1 L'apport d'une approche Bayésienne utilisée conjointement à la méthode de Monte-Carlo.

La méthode du maximum de vraisemblance permet d'estimer avec des calculs simples les paramètres du modèle. Elle recquiert cependant une connaissance préalable du nombre de sinusoïdes présentes.

L'utilisation des méthodes AIC (Akaike's Information Criterion) et MDL (Minimum Description Length) exige l'estimation des paramètres du modèle pour chaque modèle possible et l'évaluation du critère.

Cependant leur estimation du nombre de composants est en général fausse dans le cas d'échantillons de faible taille et d'un faible rapport signal sur bruit.

3.1.1 Principe de l'approche bayésienne.

Règle de Bayes.

Dans notre état de l'art, nous avons rencontré des méthodes axées sur le maximum de vraisemblance afin de modéliser au mieux le signal observé dans une fenêtre d'analyse.

Ce que propose l'approche bayésienne repose sur le principe suivant : au lieu de considérer le maximum de vraisemblance (p(observations/modele)) pour l'estimation des paramètres, on considère le maximum a posteriori (probabilité du modèle connaissant les observations) qui fait intervenir des informations a priori (p(modele)) sur les paramètres du modèle.

L'expression de la loi a posteriori se fait grâce à la règle de Bayes :

$$\boxed{p(\text{modele/observations})p(\text{observations}) = p(\text{observations/modele})p(\text{modele})}$$

Selection de modèle.

L'analyse bayésienne permet de sélectionner un modèle parmi plusieurs, et ainsi de comparer plusieurs modèles entre eux. C'est la capacité du modèle à représenter les données qui est évaluée ici. Au cours de notre travail, cet aspect de l'analyse bayésienne interviendra dans le cadre de la sélection de l'ordre du modèle harmonique, c'est à dire l'estimation du nombre de sinusoïdes présentes dans le signal.

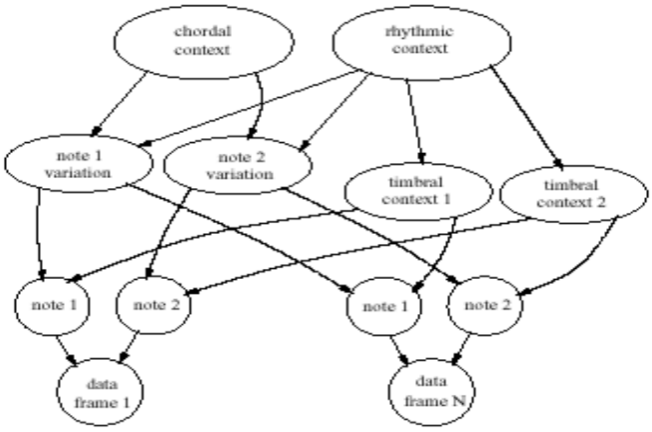
Le modèle bayésien permet de définir une probabilité postérieure sur l'espace des structures possibles du signal. Ici nous ne connaissons pas le nombre de sinusoïdes, la distribution postérieure est alors définie comme une union finie de sous-espaces de dimension variable.

Intérêt de l'approche bayésienne.

Les modèles de Bayes permettent une structure de modèles hiérarchiques qui peut être utilisée pour représenter la structure du signal à différents niveaux, du bas niveau (en termes de sinusoïdes) à des niveaux plus hauts (voir figure 3.1)

Ainsi la structure bas niveau est représentée par des notes dont les paramètres sont dépendants des différentes formes du contexte musical(contexte harmoniques accords, contexte rythmique.) Par exemple, les informations sur le timbre décrivent la variation des amplitudes des harmoniques au cours du

FIG. 3.1 – *Modèle graphique pour la représentation de la structure hiérarchique du signal.*



temps et peuvent permettre l'identification des attaques et d'atténuation des notes.

Les modèles bayésiens permettent ainsi de contrôler les informations a priori que l'on injecte dans le modèle.

Dans le cadre que l'on s'est fixé, où les instruments présents sont connus, le contexte timbral pourrait nous fournir des informations facilitant l'estimation multipitch en vue de la séparation des différents sons présents dans le signal.

L'estimation bayésienne consiste à maximiser la probabilité a posteriori, en fonction du modèle, permettant ainsi le contrôle des informations a priori que l'on a choisies d'injecter dans le modèle.

3.1.2 La méthode de Monte Carlo pour résoudre des problèmes d'optimisation.

L'approche bayésienne a pour inconvénient de devoir maximiser une probabilité a posteriori longue à calculer. La méthode de Monte-Carlo intervient alors pour chercher le maximum d'une distribution en explorant les états où le maximum est susceptible de se trouver. Les états qui nous intéressent pour la distribution postérieure sont ainsi atteints plus rapidement.

Deux algorithmes existent pour simuler la distribution postérieure par une chaîne de Markov :

L'échantillonneur de Gibbs.

L'échantillonneur de Gibbs est une méthode qui permet de construire une chaîne de Markov pour simuler la distribution postérieure.

Pour passer de l'état X^{k-1} à X^k on simule composante par composante, si X est le vecteur d'état de la chaîne de Markov de dimension n selon :

$$X_i^k \sim p(x_i \setminus x_1^k, .., x_{i-1}^k, x_{i+1}^{k-1}, .., x_n^{k-1})$$

pour $i = 1..n$ et où p est la loi a posteriori à une constante près. Chaque transition de l'échantillonneur de Gibbs ne change qu'un paramètre à la fois.

L'algorithme de Metropolis-Hastings.

L'algorithme de Metropolis-Hastings est une autre méthode pour construire la chaîne de Markov constituée des états susceptibles de maximiser la probabilité postérieure. Les états sont donc sous la forme d'un vecteur des paramètres du modèle.

Dans cet algorithme, un noyau de transition $q(X^*/X))$ propose un état X^* à partir de l'état courant X en utilisant une densité de probabilité qui est généralement dépendante de l'état X. Ce nouvel état est alors accepté avec une probabilité $Q(X/X^*))$ déterminée par la fonction d'acceptation de Metropolis-Hastings :

$$Q(X/X^*) = \min(1, PR(X)TR(X))$$

où

$$\begin{cases} PR(X) = \frac{p(X^*/d)}{p(X/d)} \\ TR(X) = \frac{q(X^*/X^*)}{q(X^*/X)} \end{cases}$$

Autrement dit, on accepte l'état X^* avec la probabilité Q, c'est à dire :

- si $PR(X)TR(X) > 1$, l'état X^* est conservé.

- sinon, l'état X^* est conservé selon un tirage de pile ou face avec la probabilité $PR(X)TR(X)$.

Choisir des noyaux de transition efficaces.

La désignation de noyaux de transition, qui soient à la fois efficaces au niveau calcul et qui explorent les régions de haute probabilité, peut accélérer la convergence de la chaîne de Markov. Combinée avec un choix judicieux de valeurs initiales (par exemple les résultats d'une analyse de données, proches temporellement et spatialement de l'observation, ou des estimateurs obtenus par des méthodes non-paramétriques), une bonne performance de cette méthode peut être obtenue, bien qu'en théorie, la chaîne de Markov converge indépendamment des valeurs initiales.

Le choix des noyaux de transision, ici, se fait de manière à exploiter les caractéristiques a posteriori du modèle considéré.

Nous allons décrire, ici, différents types génériques de noyaux de transition qui permettent différents types de mouvements dans l'espace des paramètres. Nous décrirons les noyaux effectivement utilisés pour notre problème dans la partie suivante.

Noyau d'exploration globale.
Ce noyau permet de proposer des candidats dans les régions de haute-probabilités de l'espace des paramètres. Le but de ce noyau est de construire un état indépendant dont la distribution approche la distribution posterieure. Celui-ci permet d'accélérer la convergence en permettant de traverser l'espace d'état rapidement.

Noyau d'exploration locale.
Il est avantageux d'explorer localement les régions intéressantes de notre distribution a posteriori. Les noyaux de transition alors utilisés sont des perturbations aléatoires de l'état courant, des marches aléatoires. Ces formes de transitions ont cependant tendance à explorer l'espace des paramètres lentement.

De petites perturbations sont susceptibles de fournir un haut taux d'acceptation mais une exploration lente alors que de grosses pertubations peuvent explorer plus largement pour un taux d'acceptation plus bas.

Noyau d'exploration des modes.
Certains modèles peuvent avoir une distribution posterieure multimodale,

et dans ce cas il faut éviter de se laisser emprisonner dans la région d'un seul de ces modes. Si le noyau d'exploration global ne permet pas forcément une exploration de tous les modes, il peut être utile de proposer des candidats dans les autres modes.

La méthode de Monte-Carlo permet d'accélérer l'étape de maximisation de la probabilité a posteriori en simulant une chaîne de Markov dont les états sont les régions des paramètres à haute probabilité. Elle consiste en le choix de noyaux de transition qui proposent les nouveaux états, acceptés ou non selon l'algorithme de Metropolis-Hastings.

3.2 Pour bien comprendre : le cas monophonique.

Le choix que nous avons fait concernant l'implémentation de cette méthode est de procéder pas à pas afin de bien saisir toutes ses subtilités. La première étape va consister en l'implémentation sous Matlab du cas monophonique afin de bien choisir les noyaux de transition les plus adaptés à notre situation.

3.2.1 Une modélisation du signal de musique.

Le modèle choisi dans cet article est le modèle **somme de sinusoïdes plus bruit gaussien**.
Pour chaque fenêtre d'analyse, on décrit le i^{me} échantillon sous la forme suivante :

$$d(t_i) = \sum_{k=1}^{H} a_k \cos(k\omega t_i) + b_k \sin(k\omega t_i) + e(t_i)$$

On part de l'hypothèse où le signal varie lentement à l'intérieur d'une fenêtre d'analyse, les amplitudes et la fréquence fondamentale sont considérées constantes sur une fenêtre. L'écriture matricielle du modèle a donc la forme suivante :

$$\begin{pmatrix} d(t_1) \\ d(t_2) \\ . \\ . \\ d(t_N) \end{pmatrix} = \begin{pmatrix} sin(\omega t_1) & .. & sin(H\omega t_1) & cos(\omega t_1) & .. & cos(H\omega t_1) \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ sin(\omega t_N) & .. & sin(H\omega t_N) & cos(\omega t_N) & .. & cos(H\omega t_N) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ bH \\ a_1 \\ a_2 \\ . \\ . \\ aH \end{pmatrix} + e$$

soit :

$$\boxed{d = Gb + e}$$

où G est la matrice des sinus et des cosinus de taille $N * 2H$, où N est la longueur de la fenêtre, et H le nombre d'harmoniques associés à une note.

Les paramètres du modèle (ϕ) sont :

- les amplitudes des sinusoïdes (vecteur b)
- la fréquence fondamentale (ω)
- le nombre d'harmoniques (H)
- la variance du bruit (σ_e)

3.2.2 Expression de la vraisemblance.

Calcul de l'estimée $\hat{d}$ du signal.

On considère que le bruit gaussien est IID pour donner une estimation des amplitudes des sinusoïdes. On utilise pour cela **l'estimateur des moindres carrés** :

$$\hat{d} = G\hat{b} = G(G^tG)^{-1}G^td$$

Première approximation.

Pour des tailles en nombre d'échantillons par fenêtre suffisament grande (pour N assez grand), on peut réaliser une première approximation qui consiste à approcher les sommes de Riemann $\frac{1}{N}\sum_{n=1}^{N}\sin(k\omega t_n)\sin(k'\omega t_n)$ et $\frac{1}{N}\sum_{n=1}^{N}\cos(k\omega t_n)\cos(k'\omega t_n)$ qui forment les coefficients de la matrice G^tG par les intégrales associées.

Ce qui nous ammène à la relation suivante :

$$(GG^t)^{-1} \simeq \frac{2}{N}I_N$$

Et :

$$\boxed{\hat{b} = \frac{2}{N}G^td}$$

Cette approximation se révèle en pratique tout à fait pertinente pour N supérieur à quelques centaines d'échantillons, donc pour des fenêtres de temps utilisées de l'ordre de 20ms.

Le maximum de vraisemblance.

Trouver les paramètres qui maximisent la vraisemblance des observations connaissant le modèle revient à minimiser l'erreur $d - \hat{d}$ c'est à dire maximiser la probabilité gaussienne du bruit :

$$\boxed{(\omega^*, H^*, \sigma_e^*, b^*) = \max_{(\omega,H,\sigma_e,b)} p(d/\Phi) = \max_{(\omega,H,\sigma_e,b)} \frac{1}{(2\pi\sigma_e^2)^{N/2}}\exp(-\frac{\parallel d - \hat{d} \parallel^2}{2\sigma_e^2})}$$

où $\hat{d}$ est l'estimateur des moindres carrés qui s'exprime en fonction de (ω, H, b).

Etape de Marginalisation dans la vraisemblance : réduire l'espace des paramètres.

La distribution à maximiser est une fonction des quatre paramètres du modèle (fréquence fondamentale, nombre d'harmoniques, amplitude, bruit).

Afin de simplifier le problème, on s'intéresse dans un premier temps uniquement à l'estimation de la fréquence fondamentale et du nombre d'harmoniques en se "débarrassant" des deux autres paramètres : c'est l'étape de marginalisation.

L'étape de marginalisation consiste à se débarrasser des paramètres qui forment inévitablement une part du modèle mais dont les valeurs ont peu d'intérêt.

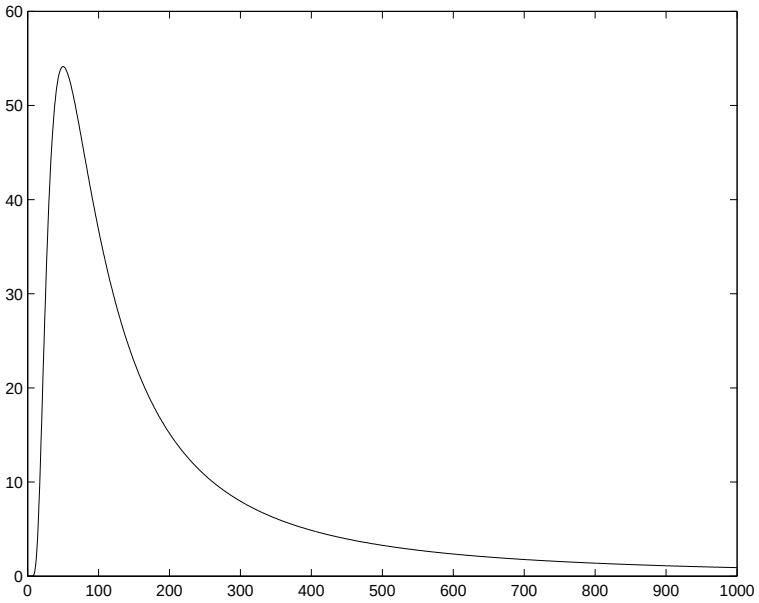
Si on dispose de l'espace des paramètres $\Theta = (\theta_1, \theta_2)$, la marginalisation du paramètre θ_2 dans la vraisemblance donne :

$$\boxed{p(\theta_1/d) = \int_{\Theta_2} p(\theta_1, \theta_2/d) d\theta_2}$$

La marginalisation nécessite de définir une loi pour les variables que l'on veut marginaliser. Cette loi doit être assez vague pour ne pas injecter de fausses informations.

1. *Cas des amplitudes*
   Paul Walmsley ([24]) préconise une loi multivariable gaussienne pour les amplitudes :
   $$p(b/d, \phi) = N(0, \sigma_e^2 * \Sigma)$$
   où $\Sigma = \delta^2 (G^t G)^{-1}$ et où δ est le rapport signal sur bruit, $\delta^2 = E(\frac{\|Gb\|^2}{\|e\|^2})$
   Le calcul de la marginalisation des amplitudes dans l'expression de la vraisemblance se fait de la manière suivante :
   $$p(d/\Phi - b) = \int p(d/\Phi) p(b/\Phi - b) db$$
   Ce qui nous permet d'obtenir l'expression de la nouvelle loi a posteriori :
   $$p(d/\Phi - b) = (2\pi\sigma_e^2)^{\frac{N}{2}} (1+\delta^2)^{-\frac{M}{2}} exp[-\frac{d^t d - \frac{\delta^2}{1+\delta^2} \hat{d}^t \hat{d}}{2\sigma_e^2}]$$
   Cette dernière équation nous donne la probabilité marginale par rapport à b. Si un nouveau paramètre a été introduit, le rapport signal sur bruit, il s'agit en réalité d'une variable dont la valeur numérique sera fixée une fois pour toute en fonction du niveau de bruit escompté dans le signal. Ici on prendra $\delta^2 = 100$.
2. *Cas de la variance du bruit.*
   On choisit pour la variance du bruit la loi inverse Gamma :
   $$p(\sigma_e^2/d, \phi) = IG(\alpha; \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma_e^{-2\alpha-2} e^{-\frac{\beta}{\sigma_e^2}}$$
   Cette loi possède une moyenne de $\frac{\beta}{\alpha-1}$ et une variance de $\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$. Elle possède un mode (point ayant la plus grande probabilité) en $\frac{\beta}{\alpha+1}$. La justification de ce choix vient du fait qu'elle est d'une part peu informative

FIG. 3.2 – *La loi inverse Gamma, IG(α, β) avec $\alpha = 1$ et $\beta = 0.01$*



et d'autre part qu'elle est intégrable. C'est aussi une généralisation d'une loi qui joue un grand rôle dans ce type de problèmes, la loi de Jeffrey qui est $p(\sigma_e) = \frac{1}{\sigma_e^2}$, obtenue pour la limite $\alpha = 0$ et $\beta = 0$, mais qui a l'inconvénient de ne pas être intégrable en 0, et donc de ne pas être normalisable sur $R+$ En introduisant les deux paramètres associés à la loi inverse-gamma α_e et β_e, on obtient pour la vraisemblance l'expression suivante :

$$p(d/\omega^q, H^q) \propto \frac{(1+\delta^2)^{-\frac{M}{2}}}{(d^t d - \frac{\delta^2}{1+\delta^2}\hat{d}^t\hat{d} + 2\beta_e)^{\epsilon}}$$

3.2.3 Expression de la loi a posteriori à maximiser

D'après la règle de Bayes énoncée dans la section précédente, l'estimation des paramètres se résume, alors, à résoudre le problème suivant :

$$(\hat{\omega}, \hat{H}) = \max_{\omega,H} p(\omega, H/d) = \max_{\omega,H} p(d/\omega, H)p(\omega, H)$$

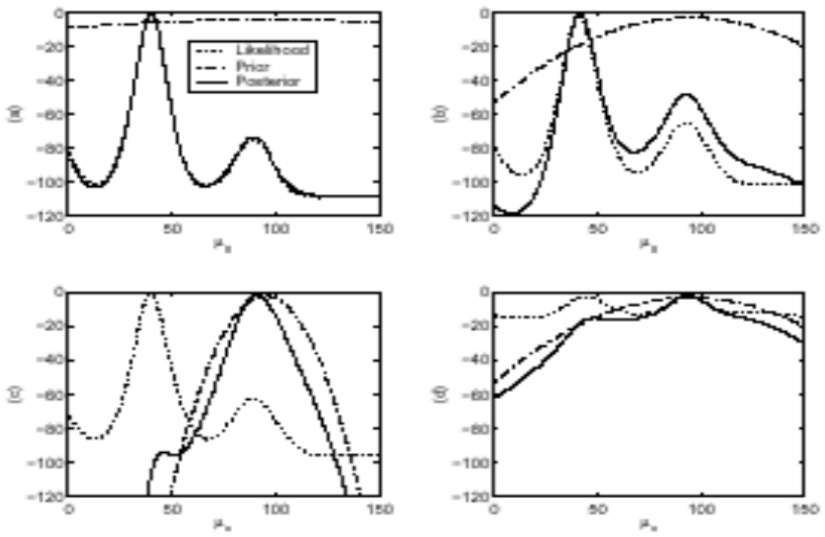
La probabilité a posteriori à maximiser s'exprime alors sous la forme suivante :

$$\boxed{p(\omega, H/d) \propto (1+\delta^2)^{-H}[\| d \|^2 - \frac{\delta^2}{1+\delta^2} \| \hat{d} \|^2 + 2\beta_e]^{-\varepsilon} p(\omega)p(H)}$$

Choix des a priori sur les paramètres (ω, H)**.**

La figure ci-dessous met en évidence l'influence de l'injection d'information a priori sur notre estimation.

FIG. 3.3 – *Impact des a priori sur la distribution a posteriori. Chaque figure montre la log vraisemblance, l'a priori et la probabilité a posteriori pour deux pics gaussien dans un bruit gaussien additif. La dernière figure montre l'effet d'une augmentation du bruit dans la donnée : la connaissance a priori compense le fait que l'information soit masquée dans le bruit.*



Le choix des informations a priori doit être fait prudemment. Quand nous connaissons peu de choses sur les valeurs probables des paramètres, il vaut mieux ne leur attribuer qu'une probabilité a priori qui est non-informative. L'a priori doit être diffus comparé à la vraisemblance et ne pas porter plus d'information qu'il n'en dispose.

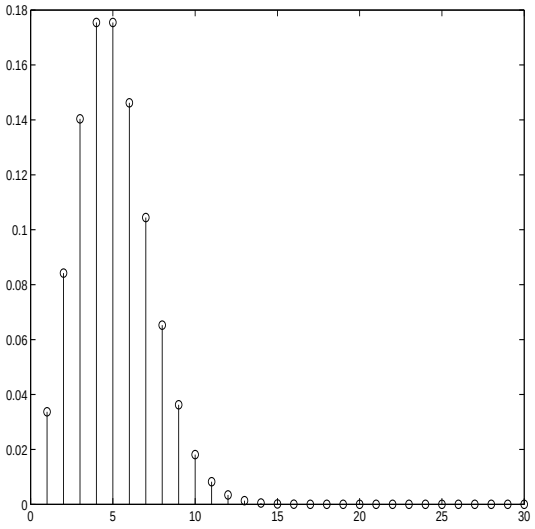
Nous choisissons dans un premier temps de n'injecter que des a priori simples sur les paramètres. Ce sont les a priori que Paul Walmsley a utilisé dans [24].

1. Loi a priori sur les fréquences : loi uniforme sur un intervalle de fréquences donné et qui correspond à la tessiture de l'instrument.
2. Loi a priori sur le nombre d'harmoniques : loi de poisson.

$$p(H) = \frac{\mu^h}{h! \exp(-\mu)}$$

Le paramètre μ dépend fortement de l'instrument dont on veut estimer la fréquence. Il serait intéressant de justifier le choix de cette loi en la testant sur une grande base de données.

FIG. 3.4 – *La loi de poisson centrée en 5*

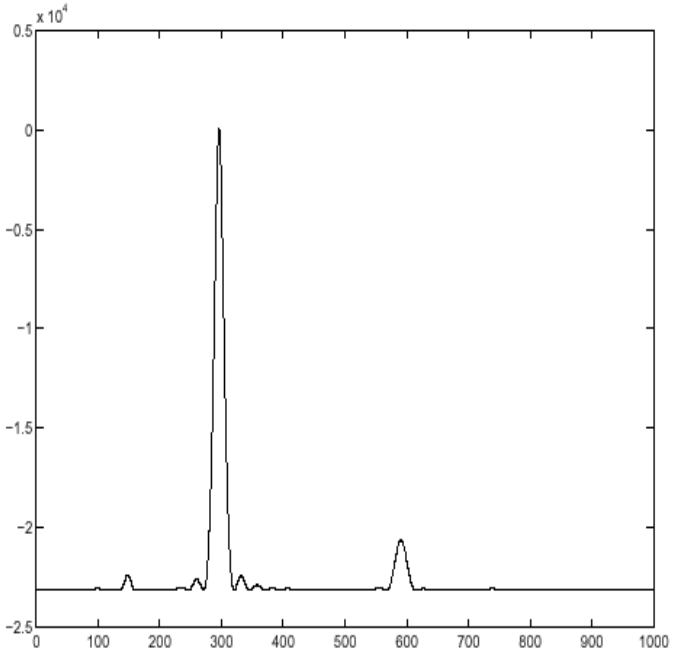


Voici donc l'allure de la loi a posteriori calculée sur une fenêtre d'analyse d'une note de saxophone D4 et marginalisée par rapoort au nombre d'harmoniques.

Les a priori sur les paramètres choisis sont pour l'instant peu informatifs, l'impact qu'ils ont sur la loi a posteriori par rapport au maximum de vraisemblance est donc moindre.

FIG. 3.5 – *La loi a posteriori marginalisée par rapport au nombre d'harmoniques pour une note de saxophone alto D4*

FIG. 1.3 – *La loi a posteriori pour une note de saxophone alto D4*



3.2.4 L'estimation des paramètres.

L'estimation des paramètres comme maximum de notre distribution a posteriori se fait par une simulation de chaîne de Markov par la méthode de Monte-Carlo, et plus précisément par l'algorithme de Metropolis-Hastings décrit précédemment (cf paragraphe 3.1.2)

Dans le cas monophonique, cet algorithme se décrit de la manière suivante :

1. Initialisation : (ω^0, H^0) par le noyau indépendant.
2. Pour chaque itération i de la chaîne de Markov :
   (a) proposition d'un nouveau candidat pour la fréquence fondamentale et le nombre d'harmoniques : (ω^*, H^*) par le noyau de transition $q(\omega^*, H^* \setminus \omega^{i-1}, H^{i-1})$
   (b) calculer la probabilité d'acceptation associée à ce nouvel état.
   (c) selon la fonction d'acceptation de Metropolis-Hasting :
   - $\omega^i = \omega^*$ et $H^i = H^*$ ou,
   - $\omega^i = \omega^{i-1}$ et $H^i = H^{i-1}$
   (d) $i \rightarrow i + 1$

Les trois noyaux de transition :

Le choix des noyaux de transition se fait de la façon suivante :

Noyau indépendant de l'état courant.

Ce noyau propose des candidats très probables de fréquences car elles sont présentes dans le spectre du signal. Ce noyau génère donc des candidats sans prendre en compte l'état précédent de la chaîne de Markov.
Pour les harmoniques, une valeur de H est engendrée à partir d'une distribution $q(H^*)$, qui est ici l'a priori.
Pour ω^*, plusieurs possibilités se présentent. Une première idée est de tirer aléatoirement parmi les pics de la fft une fréquence qui servira de candidat. Cette méthode est bien adaptée dans le cas monophonique mais dans le cas polyphonique celle-ci risque de poser un certain nombre de problèmes (à détailler).

1. *Rappels sur la transformée harmonique ou somme spectrale.*
   La transformée harmonique d'ordre P du signal d(t) s'exprime sous la forme suivante :

$$H_P(l) = \sum_{p=1}^{P} |D(pl)|^2$$

   où l est le numéro de l'échantillon fréquentiel $l = 1..L, L = \lfloor \frac{Nfft}{P} \rfloor$. On suppose qu'on a procédé a un zéro-padding sur Nfft sur le signal d.

La transformée calcule l'énergie des composantes fréquentielles aux multiples $1..P$ fois la fréquence de chaque échantillon fréquentiel. Le critère de Nyquist impose que L soit réduit pour les plus hautes fréquences afin que $pL < \frac{Nfft}{2}$

L'énergie de la projection pour ω, H peut être ensuite approchée par :

$$||f||^2 \approx \frac{2}{N} H_P(\lfloor \frac{\omega N f f t}{2\pi f s} \rfloor)$$

FIG. 3.6 – *Transformée de Fourier (module au carré) pour une note de saxophone alto D4*

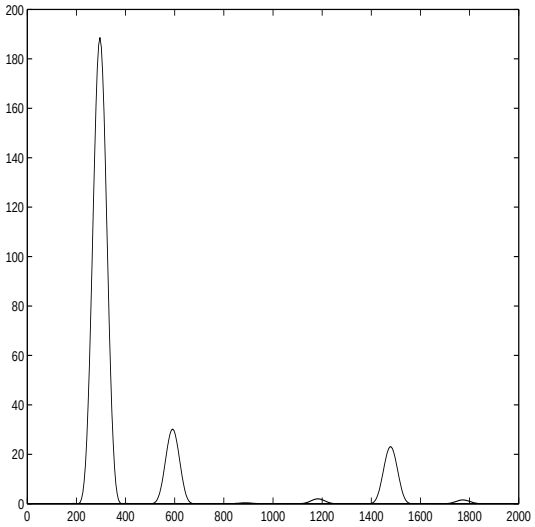
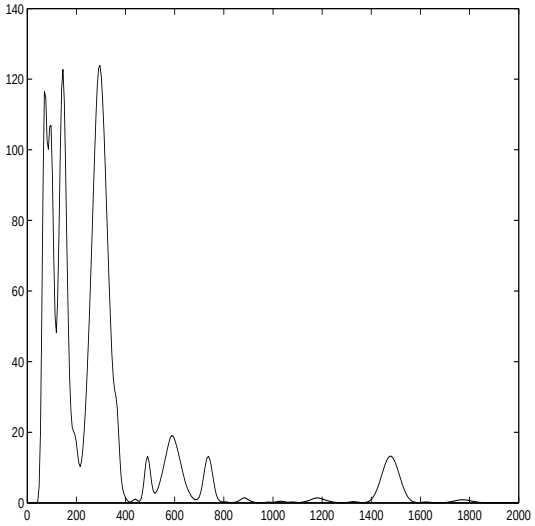


FIG. 3.7 – *Somme spectrale pour une note de saxophone alto D4*



2. *Noyau basé sur la transformée harmonique*
   La transformée harmonique et un outil utile pour le générateur aléatoire.

   Il s'agit ici de décrire la grille fréquentielle possible des fondamentales par pas égal à la résolution sur la transformée de Fourier et à pondérer chacun des points de cette grille par la probabilité postérieure approchée, enfin à sommer tous ces points en les ayant chacun légèrement décentrés au moyen d'une loi gaussienne de variance bien choisie. Le nouveau candidat est alors engendré de la manière suivante :

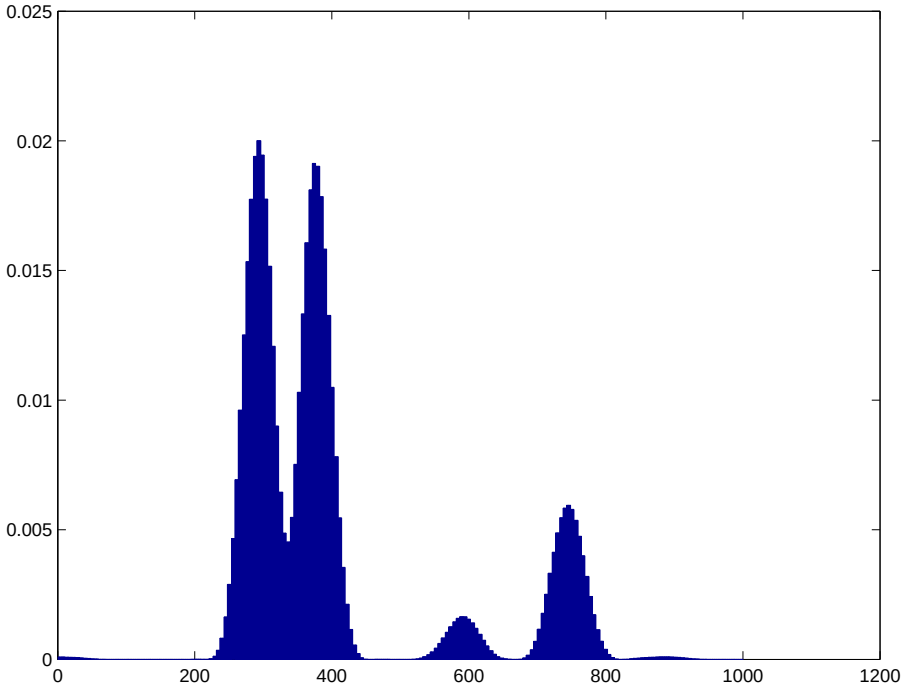
$$\omega_n^* = \sum_{l=1}^{L} N(\frac{2\pi fsl}{Nfft}, \sigma_\omega^2)(1+\delta^2)^{-H^*}[\| d \|^2 - \frac{2\delta^2}{N(1+\delta^2)} H_{H^*}(l) + 2\beta_e]^{-\varepsilon} p(\omega^*) p(H^*)$$

3. *Noyau d'exploration des régions autour des pics de la transformée de Fourier*
   Dans [3] les candidats de fréquences sont tirés selon la loi :

$$q(\omega^*) \propto \sum_{l=0}^{Nfft-1} |D(l)|^2 1_{[\frac{l*fs}{Nfft}, \frac{(l+1)*fs}{Nfft}]}(\omega^*)$$

   L'idée de base est de proposer un candidat fréquentiel indépendant de l'état précédent dans les régions autour de pics de la transformée de Fourier $D(l)$, régions déterminées en fonction du pas d'échantillonnage :

FIG. 3.8 – *La loi pour la proposition d'un nouveau candidat de fréquence pour deux notes de saxophone alto D4 (293.6) et Gb4(369.7*



Noyau d'exploration des multiples fréquentiels.

Ce noyau indépendant de l'état courant, permet de résoudre des erreurs d'octave, que l'on rencontre dans la détermination de la fréquence fondamentale. En effet, le couple de paramètres (ω, H) et celui $(2\omega, H/2)$ maximise tous les deux la probabilité postérieure, car ils possèdent tous les deux la même estimée $\hat{d}$.

C'est le $(1+\delta^2)^{-H}$ de la distribution postérieure qui privilégie le couple de paramètres qui présente le moins d'harmoniques.

Le nouveau candidat est alors engendré de la manière suivante :

$$\begin{cases} \omega_n^* = r\omega_n \\ H_n^* = poisson(\frac{H_n}{r}) \end{cases}$$

où r est tiré uniformément dans l'ensemble : $\{\frac{1}{3}, \frac{1}{3}, \frac{1}{2}, 2, \frac{3}{2}, 3\}$

$$q(\omega_n^*, H_n^*/\omega_n, H_n) = q(\omega_n, H_n/\omega_n^*, H_n^*)$$

Marche aléatoire.

Ce troisième noyau explore les régions autour des états probables pour déterminer avec une plus grande précision la position du maximum.

Les deux paramètres vont être engendrés séparément.

Le nouveau candidat est engendré par une loi gaussienne.

$$\begin{cases} \omega_n^* = N(\omega_n, \sigma_\omega^2) \\ H_n^* = poisson(H_n) \end{cases}$$

Le choix de modifier la fréquence ou le nombre d'harmoniques se fait avec la probabilité $\frac{1}{2}$.

La variance de la loi est ajustée en fonction de la précision recherchée.

Combinaison des trois noyaux.

Tirage aléatoire à chaque étape n parmi un des trois noyaux de transition avec un poids plus grand pour la marche aléatoire.

Méthode MCMC à sauts réversibles.

L'estimation du nombre de partiels n'est pas très satisfaisante avec la méthode utilisée précédemment. Les noyaux de transitions qui proposent de nouveaux candidats pour le nombre d'harmoniques n'explorent pas forcément toutes les possibilités.
cf [19]

L'estimation du nombre d'harmoniques nécessite de pouvoir augmenter ou diminuer le nombre de partiels de la façon suivante :

- augmenter le nombre de partiels avec la probabilité $b^n(H) : H \to H + n$
- diminuer le nombre de partiels avec la probabilité $d^n(H) : H \to H - n$
- mettre à jour les paramètres avec la probabilité $u(H)$

avec

$$b^n(H) = c \min\{1, \frac{p(H+n)}{p(H)}\}$$

pour $H < H_{max} - n + 1$ et

$$d^n(H) = c \min\{1, \frac{p(H)}{p(H-n)}\}$$

pour $H > n$ o ù $c = 0.15$.

Dans le cas monophonique, on peut prendre $n = 1$, dans le cas polyphonique, il sera nécessaire d'augmenter ou de diminuer le nombre de partiels avec n éventuellemnt plus grand que 1 afin de se débarrasser du problème d'un même harmonique partagé par plusieurs notes. En effet dans le cas de deux notes, l'augmentation du nombre de partiels pour la première note qui placerait un partiel à l'endroit d'un partiel existant pour la note2 ne serait jamais accepté, et par conséquent les partiels de fréquences plus hautes relatifs à la note 1 ne seront jamais atteints. En proposant d'ajouter plus d'un partiel, on a la possibilité de "sauter" le partiel commun.

C'est pour cette raison que l'algorithme est appelé algorithme MCMC à sauts réversibles.

Choix du nombre d'itérations et analyse des taux d'acceptation.

Un premier problème qui se pose est de savoir si les noyaux de transition ont été bien choisis et si le taux d'acceptation est suffisant.

Convergence de la chaîne de Markov.

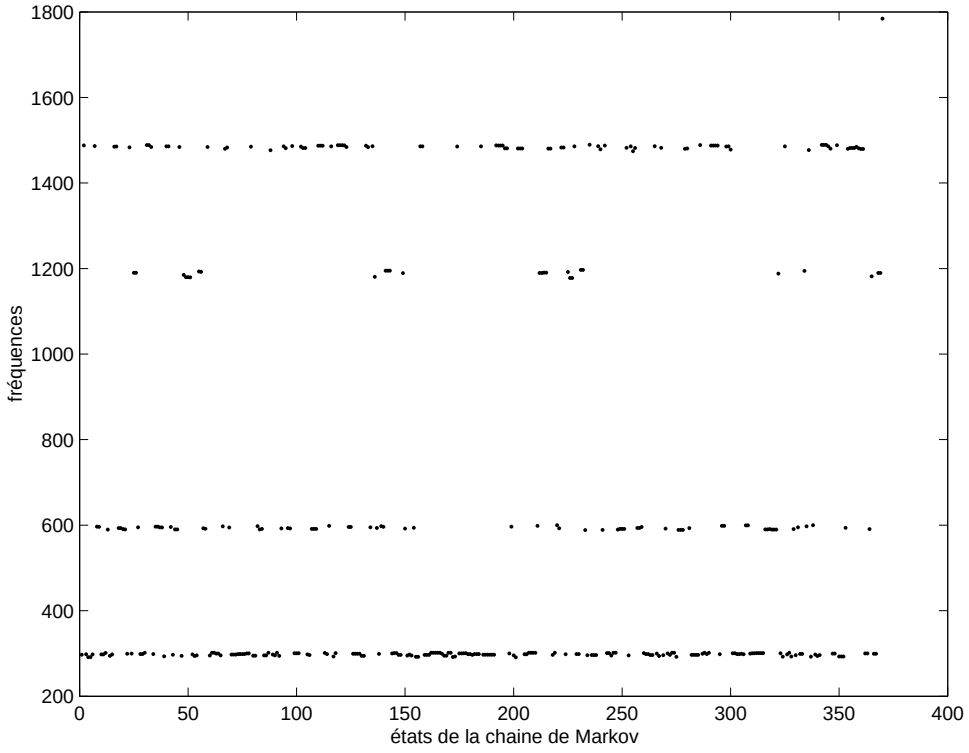
La séquence markovienne engendrée par la méthode Monte-Carlo converge en loi vers la distribution a posteriori $p(x/d)$ où $x = (\omega, H)$. On a en effet avec $(X^{(k)})_{k\in N} = (\omega^k, H^k)_{k\in N}$:

1. $lim_{k\to\infty} P(X^k = x) = p(x)$
2. $lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} h(X^(j)) = \sum_{x\in\mathcal{X}} h(x)p(x)$(presque sûrement)

En implémentant les MCMC, il est important de déterminer combien de temps la simulation doit tourner et de jeter un certain nombre d'itérations. Garder toutes les simulations d'une MCMC peut consommer beaucoup de mémoire, particulièrement quand les itérations consécutives sont fortement corrélées. La question suivante se pose alors : comment déterminer à l'avance le nombre d'itérations nécessaires pour un niveau de précision donné dans un algorithme MCMC.
Ce problème est traité en détail dans [25] et [11] Il est recommandé d'avoir un taux d'acceptation calibré à environ 0.25 pour un modèle haute dimension et 0.50 pour des modèles de dimension 1 ou 2.

Sur 1000 itérations de notre algorithme, environ 370 états ont été acceptés et sont représentés sur la figure suivante :

FIG. 3.9 – *états engendrés par les MCMC pour les fréquences pour une note de saxophone alto D4 (293.6)-noyau utilisé : nombre d'harmoniques déterministe et pour les fréquences pics de la fft*



Les différents états explorés correspondent bien à la fréquence désirée et à ses harmoniques (noyau d'exploration des multiples fréquentiels) et les états correspondant au voisinage de ces fréquences ont également bien été explorés (noyau de la marche aléatoire).

L'exploitation de la chaîne de Markov.

Les états de la chaîne de Markov ayant été déterminés, il s'agit de trouver le couple (ω, H) ayant la plus haute densité de probabilité. On obtient alors l'estimateur du maximum a posteriori.

Après 1000 itérations, notre chaîne comporte trop peu de points pour espèrer tracer un histogramme à deux dimensions . Pour résoudre ce problème, nous déterminons d'abord la fréquence sur la distribution marginalisé par rapport à H, nous déterminons d'abord la fréquence sur la distribution marginalisée par rapport à H, puis le nombre H le plus probable connaissant cette fréquence..

Voici les histogrammes grossier et fin obtenus sur une fenêtre pour une note de saxo alto D4 (293.6Hz)

Le plus grand pic de l'histogramme se situe bien au voisinage de la fréquence recherchée (293.6 Hz), les autres pics correspondant aux multiples de

FIG. 3.10 – *histogramme grossier pour une note de saxophone alto D4*

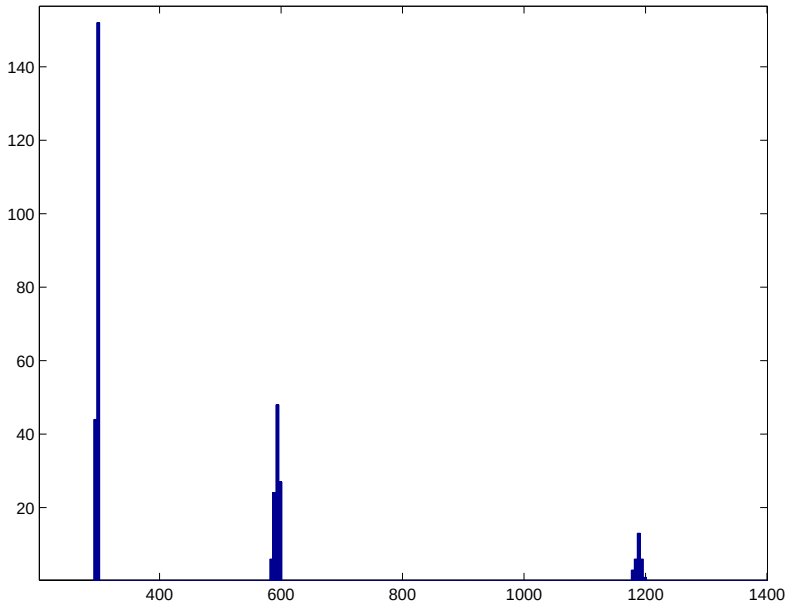
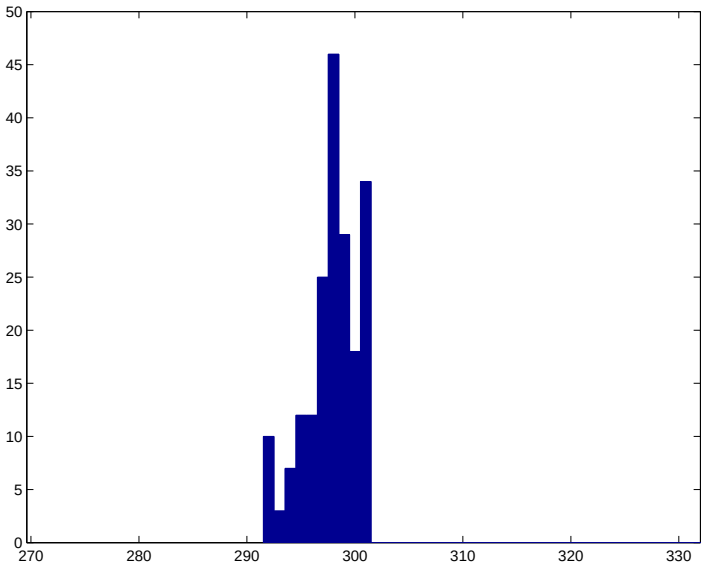


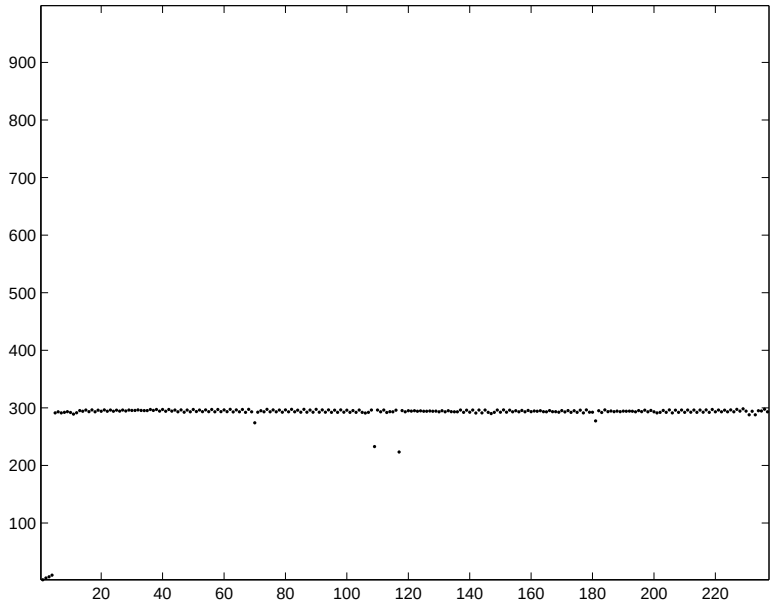
FIG. 3.11 – *histogramme fin pour une note de saxophone alto D4*



cette fréquence.

Voici les résultats obtenus pour une note tenue.

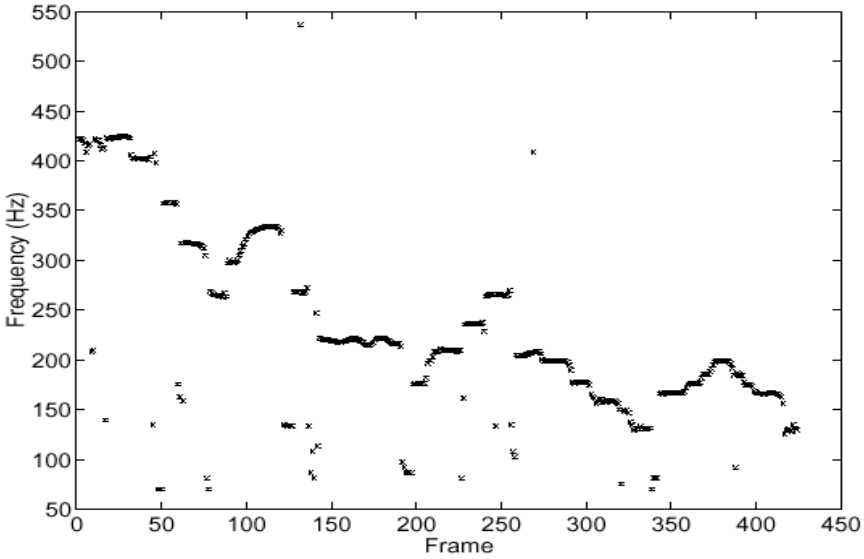
FIG. 3.12 – *Estimation de la fréquence sur plus de 200 fenêtres pour une note de saxophone alto D4 (293.6Hz)*



On voit sur cette figure que quelques erreurs apparaissent sur quatre fenêtres, erreurs qui pourront être facilement corrigées par une méthode de poursuite sur plusieurs fenêtres (voir plus loin).

L'algorithme d'estimation monopitch a également été testé sur un solo de saxophone, les notes changent à un tempo rapide, d'où l'intérêt d'une analyse sur des fenêtres de 20 ms.

FIG. 3.13 – *Contour fréquentiel obtenu pour un solo de saxophone*



Les erreurs (points isolés) tombent souvent au niveau des transitions brusques entre les notes et correspondent, la plupart du temps à des sous-harmoniques de la fréquence réelle.

3.3 Extension au signal polyphonique.

Le nombre de notes jouées simultanément n'est au départ pas une variable aléatoire . Pour chaque note q on rajoute une variable aléatoire booléenne :

$$\gamma^q = 1$$

si la note est présente dans le signal et 0 sinon.

Le modèle s'écrit alors sous la forme suivante :

$$\boxed{d = \sum_{q=1}^{q=Q} \gamma^q G_q b_q + e}$$

On cherche alors à maximiser la probabilité postérieure suivante :

$$\boxed{p(\omega^q, H^q, \gamma^q) \propto (1+\delta^2)^{-\frac{M}{2}} [\| d \|^2 - \frac{\delta^2}{1+\delta^2} \| \hat{d} \|^2 + 2\beta_e]^{-\varepsilon} p(\omega^q, H^q))}$$

où le paramètre $M \leq \sum_{q=1}^{q=Q} H^q$

3.3.1 Méthode résiduelle.

La méthode résiduelle est celle utilisée dans la thèse de Walmsley [24].

L'espace des paramètres possèdent maintenant 3Q paramètres à estimer. Afin de réduire le coût de calcul, on se ramène au cas d'estimation monophonique en soustrayant Q-1 notes au signal à chaque étape et en essayant d'estimer la dernière.

L' estimation de chaque note se fait itérativement à partir de l'estimation déjà obtenue des autres . Les états de la chaînes de Markov sont mis à jour les uns après les autres.

Ce choix de méthode n'est pas implanté ici car nous pensons que la soustraction dans le cas où le nombre de notes à séparer est élevé est une perte d'information importante.

3.3.2 Méthode itérative.

Afin de pallier au problème de la soustraction sans multiplier le coût de l'algorithme, une méthode est d'estimer une première fréquence (celle de la note d'intensité maximale) puis réinjecter dans le modèle l'information donnée par cette fréquence afin de réestimer les autres fréquences et les amplitudes conjointement.

3.3.3 Optimisation conjointe.

Le choix que nous avons fait est d'éviter toute soustraction de notes en cherchant à optimiser la probabilité a posteriori exprimée ci-dessus, conjointement pour toutes les notes présentes simultanément dans le signal.

3.4 Cas de deux notes présentes simultanément.

3.4.1 Simulation de la chaîne de Markov.

Il s'agit de maximiser la probabilité a posteriori en fonction de ses 4 paramètres $(\omega_1, \omega_2, H_1, H_2)$.

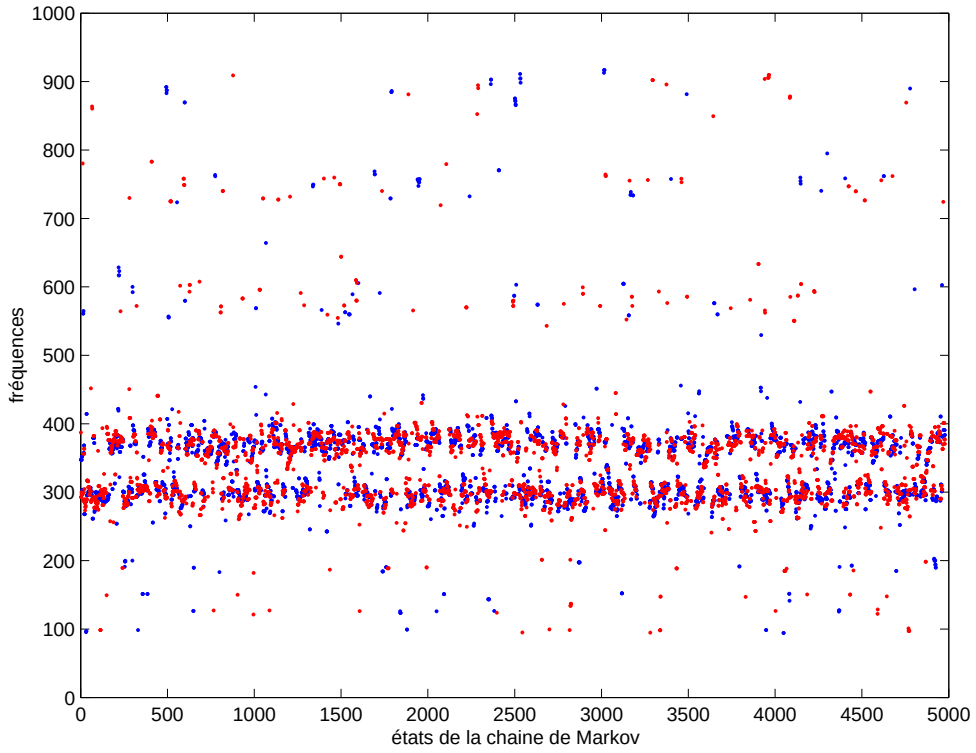
Pour cela, on va simuler une nouvelle chaîne de Markov dont les états seront représentés par un vecteur de ces 4 paramètres. L'algorithme utilisé est alors un mélange entre l'algorithme de Gibbs et l'algorithme de Metropolis-Hastings :

```
Pour la note 1 :
  1. choix d'un noyau de transition par tirage
     aléatoire parmi les trois noyaux
  2. proposer un nouveau candidat de fréquence
     pour la note i selon le générateur
     aléatoire choisi à l'étape précedente.
  3. accepter ou non le nouvel état composé
     de ce nouveau candidat pour la note1 et
     du résultat de l'état précedent pour la
     note2.
  4. inverser les rôles note1/note2.
```

Pour chaque nouvel état simulé, on ne change donc qu'une note à la fois regardant la probabilité d'acceptation en fonction de ce nouveau candidat pour cette note et du candidat proposé et accepté à l'étape précédente pour les autres notes.

Les états conservés lors de la simulation des deux chaînes (une couleur pour chaque chaîne) sont représentés sur la figure suivante :

FIG. 3.14 – *Etats conservés pour les fréquences lors de la simulation pour deux notes de saxophone alto D4 (293.6 Hz) et Gb4 (369.7)- 16000 itérations*



Les états s'accumulent autour de deux fréquences correspondant bien aux fréquences des deux notes présentes dans le mélange. Les deux chaînes sautent d'une fréquence à l'autre.

Plaçons nous maintenant dans le plan des deux fréquences pour voir comment les états se répartissent en fonction des deux fréquences.

FIG. 3.15 – *répartition des états en fonction des deux fréquences à estimer pour deux notes de saxophone alto D4 (293.6 Hz) et Gb4 (369.7)- 16000 itérations*

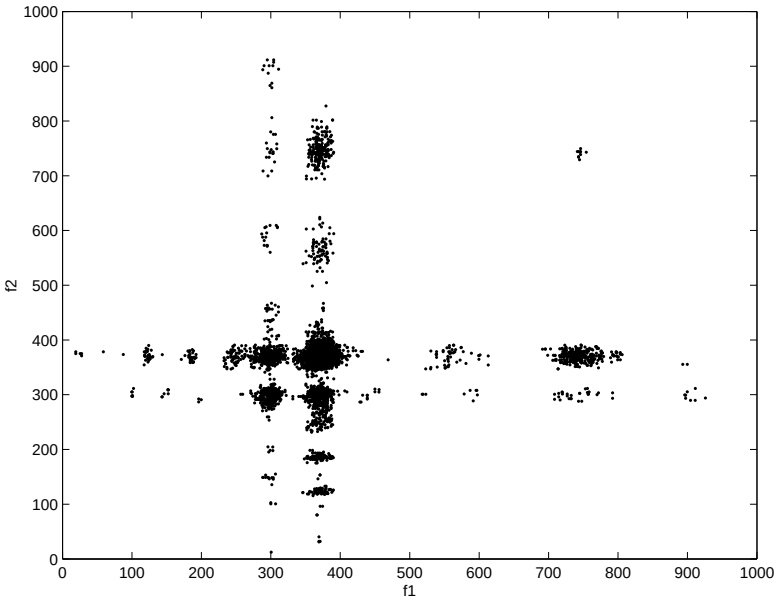
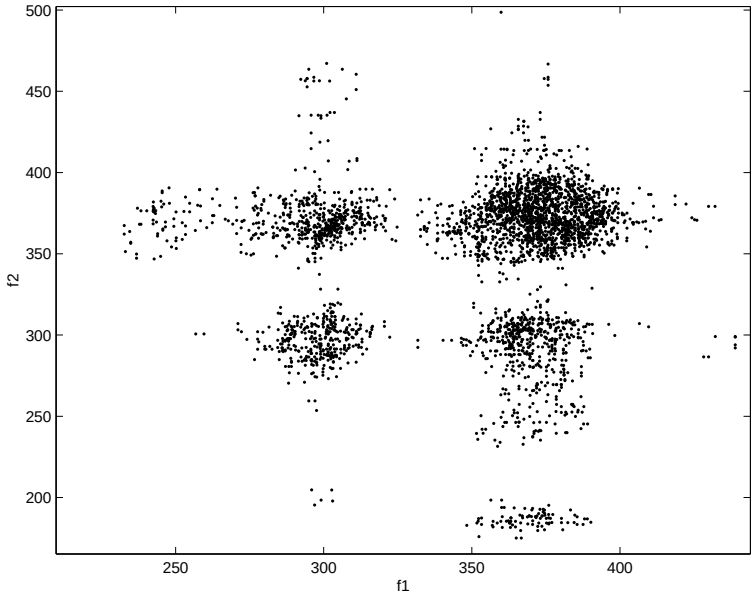


FIG. 3.16 – *répartition des états en fonction des deux fréquences à estimer pour deux notes de saxophone alto D4 (293.6 Hz) et Gb4 (369.7)- zoom - 16000 itérations*



Le plus grand nombre d'états semblent être constitués par deux fréquences identiques (autour du Sol4). Pour exploiter la chaîne de Markov, il faudra donc introduire la contrainte d'inégalité entre les deux fréquences.

3.4.2 Exploitation de la chaîne de Markov.

La méthode choisie est de concaténer les deux chaînes et de considérer la répartition des fréquences des états conservés par la simulation de la chaîne de Markov.

Pour cela, on trace l'histogramme des fréquences marginalisé par rapport au nombre d'harmoniques, les fréquences désirées correspondent aux deux plus grands pics de l'histogramme.

FIG. 3.17 – *histogramme pour deux notes de saxophone alto D4 (293.6 Hz) et Gb4 (369.7)-16000 itérations*

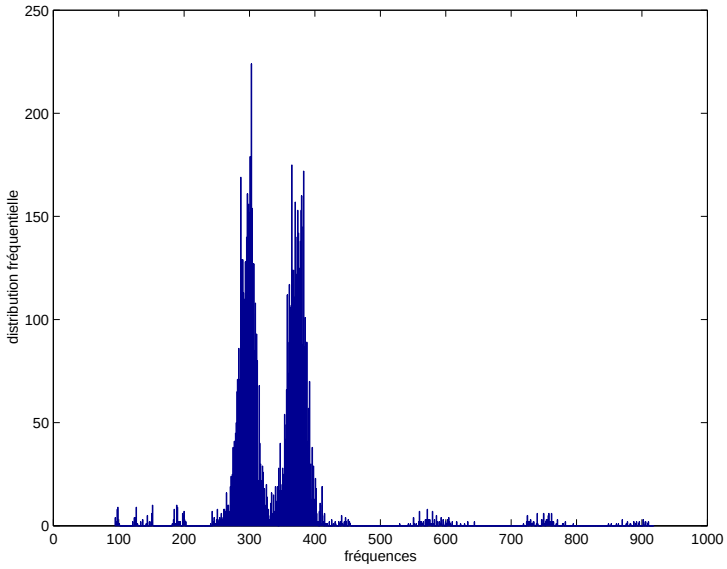
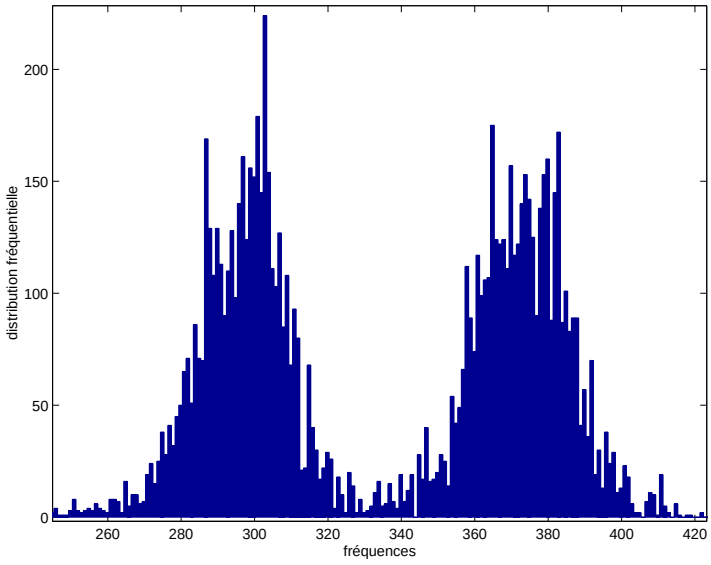


FIG. 3.18 – *histogramme pour deux notes de saxophone alto D4 (293.6 Hz) et Gb4 (369.7)-16000 itérations - zoom*



3.4.3 Commentaires et limites de la méthode.

Cette méthode implique que les signaux aient approximativement la même puissance et des fréquences différentes et suffisament espacées.

Il semblerait que la marginalisation par rapport aux amplitudes ne nous permettent pas d'estimer correctement les fréquences dans le cas où les deux signaux présents ne sont pas d'égale intensité. Il faudrait réimplémenter la méthode en introduisant un nouveau paramètre : le paramètre des amplitudes (voir paragraphe 3.6)

3.5 Tests.

3.5.1 Synthèse de signaux tests

Les signaux sont synthétisés suivant le modèle harmonique avec une phase aléatoire suivant une loi gaussienne et avec un taux d'amortissement pour les amplitudes des harmoniques de 0.3.

FIG. 3.19 – *Signal synthétisé de fréquence fondamentale 378 Hz représenté en temps*

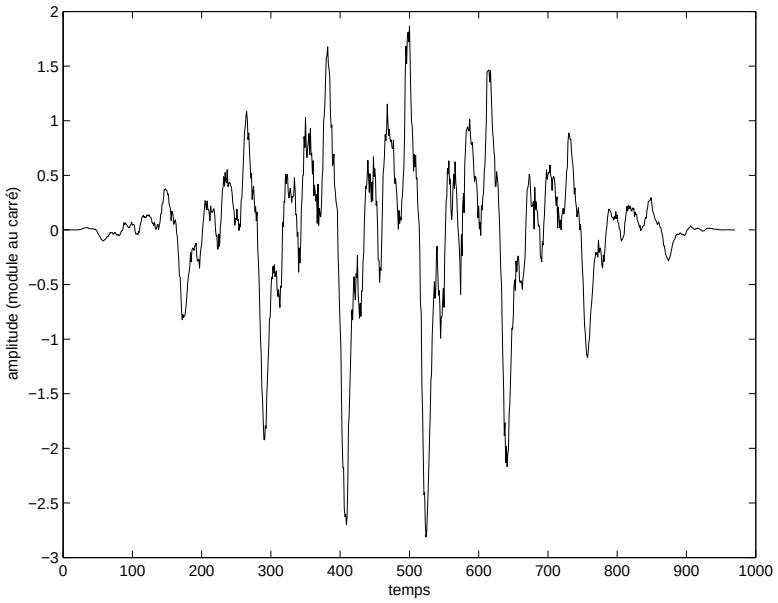
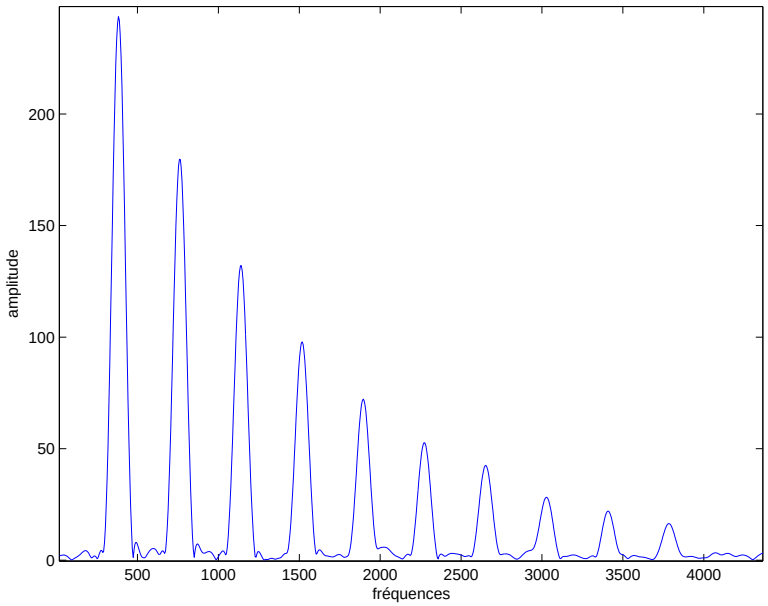


FIG. 3.20 – *Signal synthétisé de fréquence fondamentale 378 Hz représenté en fréquences*



3.5.2 Cas monophonique.

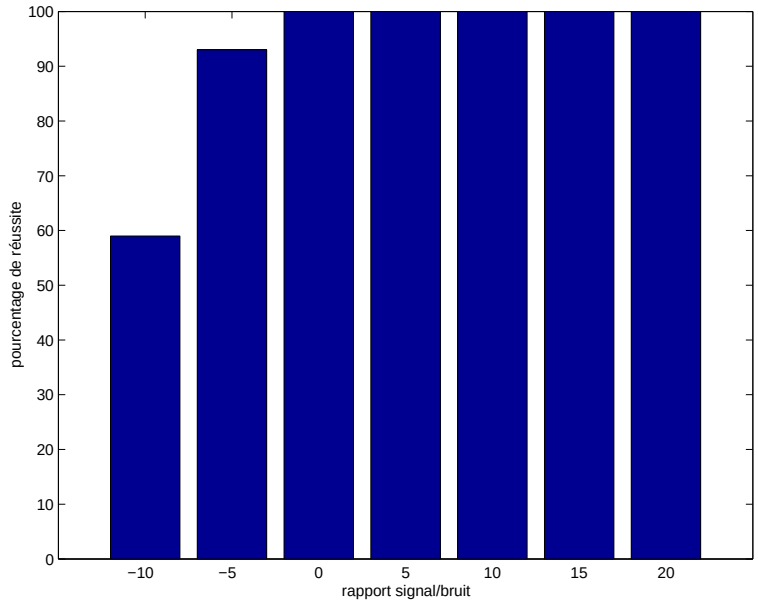
Après avoir testé notre algorithme sur une note de saxophone, nous cherchons ici à tester la robustesse de notre méthode au bruit.

Pour cela, nous allons développer un programme permettant de tester notre méthode sur des signaux synthétiques avec des rapports signal/bruit variant de 30 dB à 0 (ou -10 dB)

C'est notre estimation de la fréquence et (du nombre d'harmoniques ?) qui va être testée ici sur 100 (ou 200) signaux.

La précision fréquentielle exigée pour ces tests est de 3 pour cent de la fréquence.

FIG. 3.21 – *Pourcentage de réussite dans l'estimation de la fréquence d'une note pour différents rapports signal/bruit*

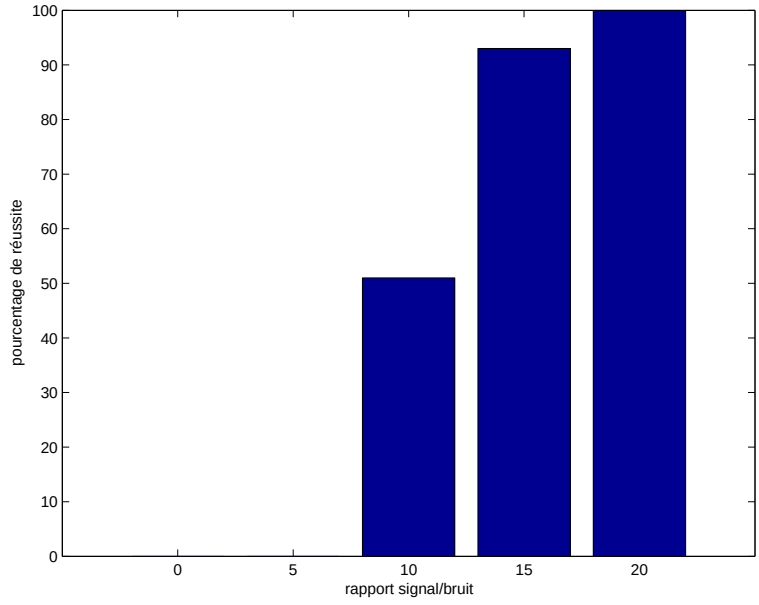


3.5.3 Estimation monopitch de mélange de signaux.

On teste ici la capacité de notre méthode à estimer la fréquence du signal dominant, et par la même occasion sa robustesse au bruit.

Les signaux du mélange ont une puissance qui décroit par pallier de 5 dB.

FIG. 3.22 – *Pourcentage de réussite dans l'estimation de la fréquence d'une note en présence de deux signaux pour différents rapports de puissance.*



Ces résultats ne sont pas convaincants, la fréquence prédominante commence à être estimée our des rapports d'intensité supérieur à 10dB. De plus, la fréquence de l'autre signal n'est jamais celle estimée. Ces résultats montrent donc l'importance de considérer également les amplitudes comme paramètres de la probabilité a posteriori à maximiser.

3.5.4 Cas polyphonique.

Le programme test, qu'il faudrait implanter ici, permettrait de tester l'estimation multipitch dans un mélange de deux signaux synthétiques pour des rapports de puissances entre les signaux variant de 20 dB à -20 dB pour un rapport signal sur bruit constant et égal à 5 dB. On pourrait ainsi tester la robustesse de l'algorithme dans des situations où les différents instruments présents ne jouent pas à la même puissance (ce qui arrive fréquemment dans les enregistrements musicaux).

3.6 Vers un modèle plus précis.

C'est le modèle utilisé dans l'article [19].

3.6.1 Dans le cas monophonique.

On ne considère plus ici que les amplitudes des harmoniques restent constants sur une fenêtre d'analyse et on introduit un paramètre δ_m pour modéliser l'inharmonicité.

Le modèle sur chaque fenêtre s'écrit alors sous la forme suivante :

$$d(t) = \sum_{k=1}^{H} a_k(t)\cos[(k+\delta_k)\omega t] + b_k(t)\sin[(k+\delta_k)\omega t] + e(t)$$

Les amplitudes sont projetées sur un petit nombre de fonctions de base $\phi_i, i = 0, ...I$:

$$a_k(t) = \sum_{i=0}^{I} a_{k,i}\phi_i(t)$$

$$b_k(t) = \sum_{i=0}^{I} b_{k,i}\phi_i(t)$$

Les fonctions de base sont par exemple la fenêtre de Hamming ou de Hanning et vérifient :

$$\phi_i(t) = \phi(t - i\frac{N-1}{I})$$

où N est le nombre d'échantillons de la fenêtre d'analyse et $t \in [O; N-1]$

Il est important que le bruit $e(t)$ modélise autant que possible les différentes sources d'erreurs de modélisation (erreurs dues par exemple au phénomène d'écho ou de réverbération...). Une solution est de prendre pour $e(t)$ un bruit AR d'ordre p :

$$e(t) = \alpha_1 e(t-1) + .. + \alpha_p e(t-p) + \epsilon(t)$$

où ϵ est un bruit blanc gaussien.

3.6.2 Dans le cas polyphonique.

Le modèle ci-dessus de vient alors :

$$d(t) = \sum_{q=1}^{Q}\sum_{k=1}^{H}\sum_{i=0}^{I} \phi(t - i\frac{N-1}{I})\{a_{q,k,i}\cos[(k+\delta_k)\omega_q t] + b_{q,k,i}\sin[(k+\delta_k)\omega_q t]\} + e(t)$$

3.6.3 Avantage de l'amélioration.

L'amplitude n'étant plus considérée comme constante sur une fenêtre on peut ainsi modéliser l'attaque ou la décroissance d'une note.

D'autre part, l'inharmonicité est également modélisée et le bruit AR permet de modéliser les fréquences résiduelles dûes par exemple au bruit environnant.

L'espace des paramètres du modèle est donc augmenté des coefficients $\alpha = (\alpha_1, ..., \alpha_p)$, des amplitudes et du paramètre d'inharmonicité.

Nous nous réfèrerons à [19] pour le choix des a priori sur ces paramètres et les noyaux de transitions alors utilisés dans le cas de ce modèle.

3.7 Cas où le nombre de notes présentes par fenêtre n'est pas connu.

.

On introduit la variable booléenne avec le noyau d'apparition et de disparition de la note.

Noyau de naissance et de mort d'une note. Ce noyau de transition supplémentaire est celui qui contrôle le nombre de notes actives dans notre modèle, en le faisant varier d'une seule note à la fois. Pour ce noyau seulement, le tirage de la note q n'a pas lieu dans l'ensemble des notes actives (autrement dit telles que $\gamma^q = 1$ mais dans toutes les notes de 1 à Q. Deux possibilités de présentent alors :

- soit la note $q^{'}$ pour laquelle on propose un nouveau candidat dans la fenêtre considérée est active.
- soit la note $q^{'}$ pour laquelle on propose un nouveau candidat ne fait pas partie du modèle dans la fenêtre considérée. et là qu'est ce qu'on fait ?

3.8 L'utilisation d'hyperparamètres pour la poursuite.

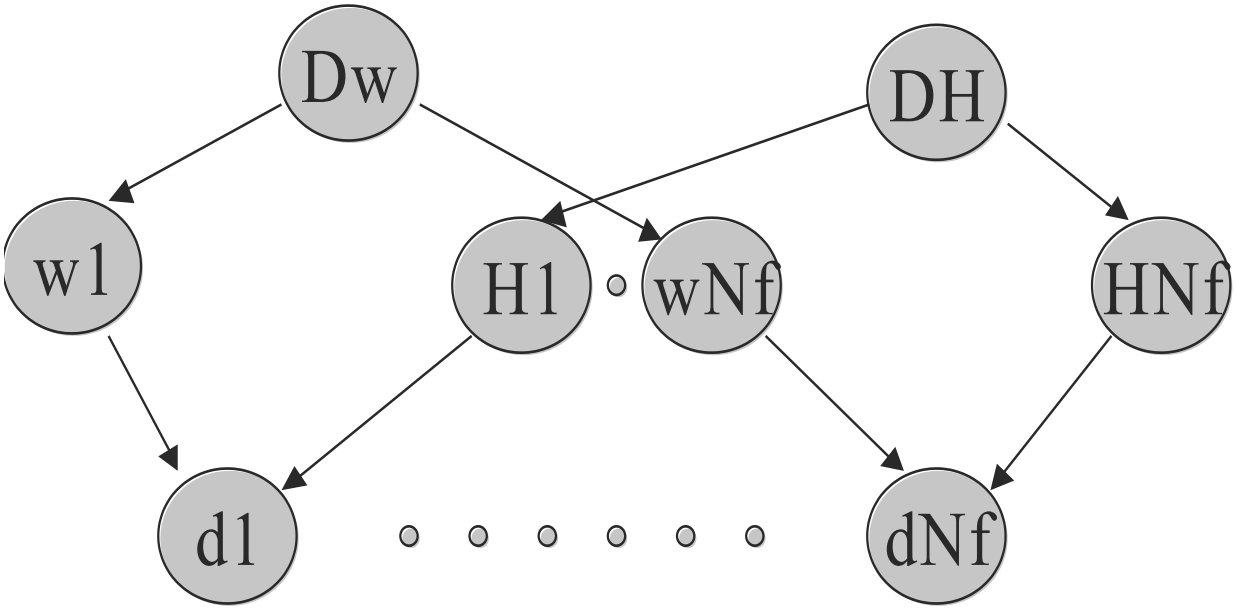
Jusqu'ici la méthode propose une analyse fenêtre par fenêtre, chaque fenêtre étant indépendante l'une de l'autre et sur chaque fenêtre la fréquence était supposée constante.

Nous allons montrer ici comment une information a priori sur la haute corrélation qui existe entre les paramètres des fenêtres adjacentes peut être exploitée dans la structure de modèle Bayésien.

3.8.1 Modèle graphique pour une analyse multi-fenêtres.

Les modèles graphiques permettent de représenter les dépendances entre les observations, les paramètres du modèle, et leurs hyperparamètres.

FIG. 3.23 – *Modèle graphique pour une analyse multifenêtres*



Une manière d'insérer des hypothèses de base sur la variation des paramètres au cours du temps est de d'imposer une dépendance markovienne entre les paramètres des fenêtres successives. L'hypothèse fondamentale est que la valeur des paramètres d'une fenêtre est proche de celle des fenêtres précédentes.

Les hyperparamètres gouvernent les variations de fréquences $(\Delta_\omega, \Sigma^2_\omega)$ et du nombre d'harmoniques (Δ_H) le long d'un bloc (ensemble de plusieurs fenêtres adjacentes). voir figure 6.8 p 175.

3.8.2 Connaissances a priori sur les hyperparamètres.

Dans cette section, nous allons présenter les connaissances a priori choisies dans la thèse de P.Wamsley [24] et dans différents articles de l'université de Cambridge ([14], [19]).

Dépendances statistiques entre les paramètres et les hyperparamètres.

On part dorénavant de l'hypothèse que les fréquences suivent la loi normale dépendant des hyperparamètres $\Delta_\omega, \Sigma_\omega^2$ de la manière suivante :

$$p(\omega_i/\Delta_\omega, \Sigma_\omega^2) = LN(\omega_i; \Delta_\omega, \Sigma_\omega^2)$$

Pour les harmoniques, on conserve la loi de poisson en faisant varier le paramètre de la loi en fonction de l'hyperparamètre Δ_H) :

$$p(H_i/\Delta_H) = Poisson(H_i, \Delta_H)$$

A priori sur les hyperparamètres.

Pour l'hyperparamètre Δ_ω régissant la variation des fréquences sur un bloc de fenêtre, deux possibilités se présentent. On peut choisir simplement une loi uniforme dont l'étendue dépend de la tessiture des instruments présents dans le mélange :

$$p(\Delta_\omega) = \frac{1}{\omega_{high} - \omega_{low}}$$

ou alors choisir un a priori markovien centré sur la valeur estimée du bloc précédent pour augmenter la dépendance des paramètres à des échelles temporelles plus longues.

Quant à la variance Σ_ω^2, elle suit la loi inverse Gamma :

$$p(\Sigma_\omega^2) = IG(\beta_\omega, \alpha_\omega)$$

Le nombre d'harmoniques suit encore la loi de poisson :

$$p(\Delta_H) = Poisson(\Delta_H, H_0)$$

3.9 Méthode de reconstruction.

Nous n'avons malheureusement pas eu le temps de traiter cette étape de la séparation de source. Nous allons cependant présenter grossièrement les différentes possibilités que nous envisageons pour la reconstruction séparée des différentes sources.

3.9.1 Synthèse des signaux séparés.

A partir des paramètres du modèle sinusoïdal estimés, il s'agit de reconstruire la partie harmonique du signal. Nous n'avons malheureusement pas pu traiter le résiduel pour synthétiser la partie bruit.

3.9.2 Soustraction.

L'idée est de soustraire les signaux synthétisés selon le modèle sinusoïdal des autres sources à l'ensemble du signal afin qu'il ne reste qu'une seule des sources. Selon l'application (restitution sur différents haut-parleurs), cette méthode pourrait être adaptée.

Conclusions et perspectives

Le travail réalisé dans le cadre de ce stage a été de bien comprendre la méthode Bayésienne et ces enjeux, en l'implémentantimplantant dans des situations simples, où tout son intérêt n'a pas pu être exploité. Il reste donc un travail important à réaliser pour enrichir et affiner la méthode. Nous avons répertorié, dans le chapitre précédent, les différentes améliorations qu'il reste à implanter.

Nous avons choisi de nous intéresser essentiellement à l'estimation multipitch, étape indispensable à la séparation d'instruments harmoniques, les étapes de traitement de la partie bruitée et de reconstruction ayant été laissées de côté.

Il reste encore, également à réfléchir sur le choix des informations a priori à injecter dans les modèles bayésiens. Dans le contexte où les instruments présents dans l'enregistrement sont connus, l'idée serait pour chaque instrument d'essayer de déterminer des statistiques permettant de modéliser ses caractéristiques propres dans les a priori sur les paramètres ou les hyperparamètres. Pour cela, il faudrait disposer d'une grande base de données pour chaque instrument permettant de tester les différentes lois choisies. Par exemple, pour la clarinette, on pourrait utiliser les informations dont on dispose sur ses harmoniques (harmoniques impaires).

Bibliographie

[1] J-M Holm A. Klapuri, T. Virtanen. Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals. In In Proc. COST-G6 Conference on Digital Audio Effects, Verona,Italie, 2000.

[2] S. Mac Adams. Recognition of auditory sound sources and events. thinking in sound : The cognitive psychology of human audition. Oxford University Press, 1993.

[3] C. Andrieu A.Doucet. Joint bayesian model selection and estimation of noisy sinusoids via reversible jump mcmc. IEEE Transaction of signal Processing, 47(10), October 1999.

[4] S. Affes and Y. Grenier. A signal subspace tracking algorithm for microphone array processing of speech. In IEEE Transactionon speech and Audio Proceedings, volume 5, Hamburg,Germany, September 1997.

[5] A.S. Bregman. Auditory scene analysis. MIT Press, 1990.

[6] K. Chen. Tracking of patials for additive sound synthesis using dynamic programming. ECE 497, 2000.

[7] A. Eronen. Musical Instrument Recognition. PhD thesis, Tampere University of Technology, 2001.

[8] P. Djurié H.T. Li. An iterative procedure for joint bayesian spectrum and parameter estimation of harmonic signals. IEEE-International Symposium on Circuits and ystems, Mai 1996.

[9] S. Kay and V. Nagesha. Extraction of periodic signals in colored noise. IEEE Transactions on Speech and Audio Processing, 9, October 1992.

[10] A. Klapuri. Number theorical means of resolving a mixture of several harmonic sounds. MIT Press, 1990.

[11] Jun S. Liu. Monte-Carlo strategises in Scientific Computing. Springer, 1996.

[12] X. Rodet P. Depalle, G. Garcia. Tracking of partials for additive sound synthesis using hidden markov models. Mineapolis, Minnesota, April 1992.

[13] P.Duvaut. Traitement du signal. Hermès, 1996.

[14] P.J.W. Rayner P.J. Walmsley, S.J. Godsill. Bayesian graphical models for polyphonic pitch tracking. In Diderot forum, Vienna, December 1999.

[15] B. David R. Badeau, R. Boyer. Eds parametric modeling and tracking of audio signals. In Proceeding of the 5th Conference on Digital Audio Effects (DAFx-02), Hamburg,Germany, September 2002.

[16] G. Richard R. Badeau, B. David. Sintrack analysis for tracking components of musical signals. Proceeding of Forum Acusticum Sevilla, 2002.

[17] J. Rissanen. Modelling by the shortest data description, 1978.

[18] J. Rosier and Y. Grenier. Pitch estimation for the separation of musical sounds. In 112th AES Convention, Münich,Germany, Mai 2002.

[19] M. Davy S.J. Godsill. Bayesian harmonic models for musical pitch estimation and analysis., 2002.

[20] A. Klapuri T. Virtanen. Separation of harmonic sound sources using sinusoïdal modeling. IEEE, 2000.

[21] A. Klapuri T. Virtanen. Separation of harmonic sound sources using multipitch analysis and iterative parameter estimation. In IEEE, New York, October 2001.

[22] D.J. Thomson. Spectrum estimation and harmonic analysis. Proceedings of the IEEE, 1982.

[23] T. Tolonen. Separation of harmonic sound sources using sinusoïdal modeling. In Seminar on the Content Analysis of Music and Audio, 1999.

[24] P.J. Walmsley. Signal Separation of musical instrument. PhD thesis, University of Cambridge, 1999.

[25] S. Richardson W.R. Gilks and D.J. Spiegelhalter. Markov Chain Monte-Carlo in Practice. Chapman and Hall, 2001.