Ecole Centrale de Nantes **Université de Nantes**

Ecole Doctorale

Sciences et Technologies
de l'Information et des Matériaux

Année 2006

Thèse de Doctorat

Diplôme délivré conjointement par
l'Ecole Centrale de Nantes et l'Université de Nantes

Spécialité : Automatique et Informatique Appliquée

Présentée et soutenue publiquement par :

Corentin Dubois

le ? novembre 2006
à l'Ecole Centrale de Nantes

Titre

Méthodes de Monte Carlo séquentielles pour l'estimation de fréquences fondamentales
~
Application à la caractérisation et à la reconnaissance de l'effet de métabole, en environnement urbain

Jury

Président	Jean-Pierre Guédon	Professeur Université de Nantes
Rapporteurs	Bruno Torrésani	Professeur Université de Provence
	Olivier Cappé	Chargé de recherche CNRS, ENST, Paris
Examinateurs	Xavier Rodet	Professeur Université Paris VI
	Gérard Hégron	Professeur ENSA, Nantes
	Manuel Davy	Chargé de recherche CNRS, LAGIS, Lille
Membres invités	André Gilloire	France Télécom R&D, Lannion
	Nicolas Rémy	Maître assistant ENSA, Grenoble

Directeur de thèse : Gérard Hégron
Co-encadrants : Manuel Davy et Nicolas Rémy N° ED 0366xxx

Résumé

Le développement de systèmes pouvant répondre à une requête concernant le contenu informationnel d'un signal sonore, comme les moteurs de recherche multimédia, est aujourd'hui en plein essor. L'extraction de l'information nécessaire à l'ordinateur pour prendre une décision, se fait classiquement en caractérisant le signal sonore par une série de descripteurs audio. Si leur choix est souvent empirique, le triplet fréquence fondamentale/énergie/structure fréquentielle se démarque néanmoins. Il permet de reconstruire un signal semblable à l'original, sur le plan perceptif, captant ainsi l'information contenue dans le signal.

Une synthèse bibliographique montre que l'estimation de ce triplet de descripteurs est particulièrement difficile dans le cas polyphonique et nécessite souvent de fortes hypothèses structurelles sur le signal traité. Les signaux considérés dans cette thèse, sont issus de la superposition et de l'interaction d'un nombre inconnu et variable dans le temps de sources sonores, ces dernières pouvant être de nature différente : parole, musique, ambiance urbaine... Nous avons construit un algorithme séquentiel (filtrage particulaire) pour estimer le nombre de sources et leurs caractéristiques, au cours du temps. Les méthodes de Monte Carlo séquentielles permettent de s'affranchir d'hypothèses simplificatrices et de prendre en compte un large panel de signaux sonores. L'estimation est effectuée dans un cadre bayésien, à la fois rigoureux et flexible, avec lequel bon nombre des algorithmes existants peuvent être combinés. Après une application à la transcription automatique de la musique, les descripteurs sont utilisés pour la caractérisation de l'effet de métabole.

Mots-clés : Descripteurs audio, fréquences fondamentales multiples, estimation bayésienne, méthodes de Monte Carlo séquentielles, effets sonores.

Title : Sequential Monte Carlo Methods for Multipitch estimation - Application to the characterization and the recognition of the "métabole" effect in urban atmosphere.

Abstract

Systems aimed at performing content-based audio signal retrieval, such as multimedia search engine, are currently under intense development. The audio information that the computer needs to make a decision is usually characterized by audio features. Even if their choice often results from heuristics, the triplet fundamental frequency/energy/frequency structure is usualy preferred. Indeed, a signal can be reconstructed from it and its similarity to the original signal on a perceptive level, shows that the useful information is kept.

A bibliographical study emphasizes that the estimation of this triplet of features is quite hard to perform in the polyphonic case and often needs some strong structural assumptions about the processed signal. We address here signals that are composed of an unknown, time-varying number of interfering sources, which are of different kinds : speech, music, urban sound atmosphere... We propose a sequential algorithm (particle filtering) to estimate the number of sounds together with their characteristics at each time instant. Sequential Monte Carlo methods enable to decrease the number of simplifying assumptions and to take into account a large set of audio signals. Moreover, the estimation is performed within a both rigourous and flexible Bayesian framework with which most of existing method can be combined. Our algorithm is first applied to the automatic music transcription problem and then to characterize one particular sound effect named "métabole".

Key-words : Audio features, multipitch, Bayesian estimation, sequential Monte Carlo methods, sound effects.

Table des matières

Table des figures

Liste des tableaux

Introduction

L'imagination est plus importante
que la connaissance.
A. Einstein

Cet air de musique que je fredonne sans cesse, comment pourrais-je le retrouver dans une bibliothèque multimédia, si je ne connais ni son titre ni son auteur ? Cette bande sonore que je reçois, est-elle caractéristique d'une ambiance sereine ou décrit-elle une situation de détresse ? Je rentre de vacances et je trouve, sur mon répondeur téléphonique, un nombre considérable de messages. Ne serait-il pas possible de les trier automatiquement, pour faire ressortir ceux qui sont urgents ? Cet espace publique que je conçois, comment puis-je connaître l'ambiance sonore qui y régnera, je ne peux m'y promener ? Toutes ces interrogations renvoient, d'une manière ou d'une autre, à une même problématique générale : j'ai une requête concernant le contenu d'un signal sonore, comment puis-je en extraire l'information utile pour y répondre ?

Le développement d'un moteur de recherche multimédia, la mise en place d'un système de surveillance automatique, l'utilisation d'un outil informatique d'aide à la prise de décision ou la conception assistée par ordinateur, sont autant d'applications en plein essor aujourd'hui. Pourtant, pas forcément simple à la base, cette problématique, qui leur est sous-jacente, se complique considérablement lorsqu'il s'agit de la résoudre avec un ordinateur. En effet, l'analyse par ordinateur de signaux sonores rencontre plusieurs problèmes qui la rendent difficile. Si l'oreille humaine est un système remarquable qui est à la fois capable de distinguer des détails extrêmement fins et d'être insensible aux variations, l'ordinateur, lui, a besoin d'une définition stricte de ce à quoi ressemble l'information recherchée dans le signal et est généralement inflexible face aux variations. Et même, en imaginant qu'un algorithme de traitement du signal capable d'effectuer une analyse aussi précise que celle qui a lieu dans le cerveau humain, puisse être mis au point, apprendre à un ordinateur à prendre une décision n'est pas un problème trivial.

Ceci met en avant les deux grandes étapes que l'on peut distinguer dans le processus de prise de décision, par un ordinateur, concernant le contenu informationnel d'un signal audio. En effet, le cœur des techniques utilisées est la comparaison entre la requête et la base de données sur laquelle l'ordinateur s'appuie pour prendre sa décision. Ces deux éléments pouvant ne pas être du même ordre, il est nécessaire de les transformer pour les rendre comparables, c'est la première étape. Le signal sonore est placé dans un espace de représentation dans lequel, d'une part, l'information discriminante, celle qui nous intéresse, est mise en valeur et, d'autre part, le reste du signal, pas utile pour la tâche que l'on veut effectuer, est mis en arrière plan. La seconde

étape consiste alors en l'élaboration d'un outil de comparaison et d'une règle de décision dans cet espace.

Le travail effectué durant cette thèse, et présenté dans ce manuscrit, se focalise essentiellement sur la première étape.

.1 Espace de représentation du signal sonore

A l'instar de plusieurs solutions proposées dans le passé, pour résoudre différents problèmes, comme la segmentation et la classification de segments vidéos [Liu98b], la classification automatique du genre musical [Tza02b], la classification de signaux musicaux [Foo99, Tza02a, Her03] ou la reconnaissance d'instruments [Liv04], le signal traité est souvent caractérisé, à chaque instant d'analyse, par une collection de descripteurs audio [Dav02a]. A l'opposé des méthodes de représentation à proprement parler, comme les représentations temps-fréquence, en ondelettes, *etc*, les descripteurs audio fournissent un résumé synthétique caractérisant le signal, autour du moment d'analyse. Comme se sont eux qui définissent l'espace de représentation, leur choix est critique. En effet, ils doivent être discriminants par rapport à la requête à laquelle on cherche à répondre. C'est là que se situe la principale difficulté de l'approche par descripteurs car, face à la diversité des problèmes à résoudre, il est difficile de les relier à la requête et leur choix est souvent empirique.

A l'intermédiaire entre les représentations de type temps-fréquence et les descripteurs audio les plus courants, le triplet fréquence fondamentale/énergie/structure fréquentielle est un bon compromis. En effet, le contenu fréquentiel des signaux sonores occupe une place importante dans leur interprétation, comme peut le montrer une étude de la perception sonore par l'oreille humaine. De plus, ces descripteurs sont complémentaires et suffisent pour reconstruire un signal audible. Une écoute comparative entre ce signal et l'original, permet de voir que toute l'information contenue dans le signal original, est restituée. Cette validation perceptive du choix de ces descripteurs est justifiée par l'application à l'origine de cette thèse, qu'il est nécessaire d'expliciter dans cette partie introductive.

.2 Les effets sonores

La qualité de l'ambiance sonore en environnement urbain, est prise en compte avec une importance croissante dans la gestion de la ville. La description des situations sonores en intérieur, à l'aide de critères physiques [Pel91], est aujourd'hui quelquechose de possible et d'utilisable pour simuler les caractéristiques sonores d'une salle, c'est-à-dire ses qualités et ses défauts pour l'auditeur, avant même qu'elle ne soit construite. En urbanisme, et en extérieur en général, une telle description de l'ambiance sonore est beaucoup plus difficile et les mêmes descripteurs qu'en intérieur ne sont généralement pas utilisables.

Depuis les années 80, des recherches, menées principalement au CRESSON, à Grenoble, ont permis l'émergence d'un nouveau concept particulièrement bien adapté à la description des ambiances sonores extérieures : l'effet sonore [Aug82, Che94]. L'effet sonore prend en compte à la fois l'aspect quantitatif lié au signal sonore et l'aspect qualitatif lié à la perception que l'on a de ce signal sonore. Il se veut être l'intermédiaire entre l'objet sonore, introduit par

Schaeffer [Sch66], mais qui est souvent jugé trop objectif, et le concept de paysage sonore, introduit par Schafer [Sch80], mais qui lui, est jugé trop subjectif. L'effet sonore est à la fois quantitatif et qualitatif, ce qui le rend nécessairement multidisciplinaire. Il inclut, dans sa description, trois composantes :

- l'aspect physique : le signal sonore en lui-même, tel qu'il peut être enregistré
- l'aspect spatial : influence du contexte architectural dans lequel on se trouve lorsque l'on écoute le signal
- l'aspect socio-psychologique : influence de l'état d'esprit, du passé, de la culture de la personne qui écoute.

Sur les quelques quatre-vingts effets sonores définis [Aug95], l'importance relative de chacun de ces trois aspects peut varier considérablement. Ainsi, l'aspect socio-psychologique dans les effets dits mnémo-perceptifs ou psychomoteurs est prépondérant par rapport aux autres. A l'inverse, les effets élémentaires ou de composition sont beaucoup plus basés sur la physique du signal.

Plusieurs études [Odi96, Lav98] ont été menées, au sein du CRESSON, sur les effets sonores. Elles font ressortir que, si la détection des effets sonores est possible, elle reste très subjective, ce qui complique d'autant leur détection automatique. Dans cette thèse, nous avons essayé de caractériser, en utilisant l'approche par descripteurs, un effet sonore en particulier : celui de métabole. Une ambiance sonore métabolique peut être rapidement décrite par la difficulté, voire l'impossibilité, de distinguer ce qui fait partie du fond et ce qui peut en émerger. En fait, les sources sonores qui interagissent pour créer l'ambiance perçue, oscillent constamment entre le fond sonore et le premier plan, de sorte qu'il est difficile, pour l'auditeur, de fixer son attention sur l'une d'entre elles.

.3 Suivi de trajectoires spectrales

Si la fréquence fondamentale, la structure fréquentielle et l'énergie présentent des propriétés intéressantes du point de vue de la caractérisation du contenu informationnel du signal, leur estimation, au cours du temps, est particulièrement difficile. En effet, les sons qui nous parviennent et que nous souhaitons traiter automatiquement, sont le résultat du mélange de plusieurs sources sonores, chacune d'entre elles voyant sa fréquence fondamentale, sa structure fréquentielle ou son énergie avoir sa propre évolution temporelle. De plus, ces différentes sources interfèrent les unes sur les autres, ce qui vient perturber leur séparation et l'estimation de leurs caractéristiques individuelles. Dans ce manuscrit, nous avons développé un algorithme de filtrage particulaire capable, au cours du temps, de détecter l'apparition d'une source et de la suivre, dans l'espace de représentation, afin d'en estimer les caractéristiques, jusqu'à ce qu'elle disparaisse. La possibilité de traiter simultanément plusieurs sources est aussi une contrainte à laquelle l'algorithme répond.

.4 Organisation du manuscrit

Comme nous l'avons dit, le présent manuscrit est principalement focalisé sur l'estimation de fréquences fondamentales multiples. Le premier chapitre est consacré à l'étude de la perception sonore et plus particulièrement, des quantités physiques qui la caractérisent. Nous verrons que

la fréquence fondamentale ne doit pas être dissociée de la structure fréquentielle du signal et de son énergie au cours du temps. Ces trois descripteurs étant, en quelque sorte, complémentaires, ils suffisent à définir un espace de représentation du signal. Ce choix est aussi mis en relation avec les autres descripteurs audio souvent utilisés.

Dans le second chapitre, une revue de la littérature consacrée à l'estimation de fréquences fondamentales multiples. Les méthodes existantes peuvent être séparées en trois grandes catégories. D'abord, celles se basant sur le fonctionnement de la perception sonore chez l'être humain. Ensuite, celles utilisant des architectures permettant de combiner des processus *bottom-up* (typiquement, des méthodes de traitement du signal) avec des connaissances *top-down* (informations *a priori* que l'on cherche à inclure dans le processus d'estimation). Enfin, celles s'appuyant sur un modèle génératif du signal. Cette étude de l'existant permet aussi de mettre en évidence les difficultés auxquelles sont confrontés les algorithmes d'estimation et de fixer le contexte dans lequel nous nous situerons. En particulier, nous voulons poser le moins possible d'hypothèses restrictives, afin de pouvoir traiter un large panel de signaux.

Le troisième chapitre présente le cadre théorique et algorithmique à partir duquel nous avons développé notre méthode. Plus précisément, nous nous plaçons dans la troisième catégorie, c'est-à-dire qu'un modèle génératif est construit et que le problème devient alors l'estimation de ses paramètres. Cette estimation se fait par inférence bayésienne, dans un contexte séquentiel, en utilisant les méthodes de Monte Carlo séquentielles.

C'est dans le quatrième chapitre, que sont détaillés le modèle séquentiel et l'algorithme développé. Nous verrons, en particulier, que les méthodes de Monte Carlo séquentielles sont suffisamment générales, pour permettre de faire différents choix, à divers niveaux du modèle ou de l'algorithme, tous en restant dans un cadre rigoureux. Des exemples d'applications sont ensuite proposés sur des signaux synthétiques ou réels.

Enfin, dans le cinquième chapitre, l'application à la caractérisation de l'effet de métabole est développée. Pour cela, deux critères physiques sont définis et un processus de détection de la métabole est proposé. Une évaluation des performances est alors effectuée sur un corpus de signaux métaboliques et de signaux non métaboliques.

Chapitre I

Caractéristiques de la perception sonore

Nature gave us limbs for fight or flight,
and we invented athletics.
Nature gave us pitch to sort out the world,
and we invented music.
W. H. Hartmann

La perception d'un signal sonore et son interprétation par celui qui l'écoute, se basent sur des caractéristiques précises de ce signal. L'une des plus importantes d'entre elles est la *hauteur* (encore appelée *pitch*). L'auditeur se base sur la notion de hauteur perçue pour interpréter et séparer les sources sonores provenant de son environnement. Plusieurs évènements de la vie quotidienne peuvent produire des ondes acoustiques et ces sons individuels présentent une structure complexe qui varie au cours du temps. De plus, ils se recouvrent, voire se masquent dans les domaines temporel et fréquentiel et sont mélangés quand ils atteignent l'oreille. Malgré ce phénomène de fusion, le système perceptif humain est capable de retrouver les caractéristiques des évènements à l'origine des sons qui lui parviennent, en s'appuyant largement sur la hauteur [Ros98]. Cependant, la hauteur ne permet pas, à elle seule, une description complète des signaux sonores, il faut quelle soit complétée par les notions de *volume*, de *timbre* et même de *durée*. Si la hauteur, le volume et le timbre sont des descripteurs plutôt subjectifs, ils peuvent néanmoins être reliés à des quantités physiques comme la fréquence fondamentale, l'intensité (ou l'énergie) sonore et la structure fréquentielle du signal. Afin de mieux appréhender les liens qui existent entre ces quantités physiques et ces descripteurs subjectifs, il convient de donner une définition précise à chacun de ces termes, c'est le but de ce chapitre.

La première partie permet tout d'abord d'expliciter la structure des signaux plus ou moins complexes qui nous entourent et que nous voulons caractériser et traiter. La seconde partie est consacrée à la hauteur et à sa relation avec la fréquence fondamentale. Ensuite, en se basant sur le fonctionnement de l'oreille humaine, est expliqué le rôle prépondérant de la hauteur dans la perception sonore. Puis, les trois autres attributs sont étudiés ainsi que leur influence sur la perception de la hauteur. Enfin, dans la dernière partie, une discussion est proposée sur la pertinence d'une telle description.

I.1 Sons purs - Sons complexes

Parmi les sons les plus simples qui existent, ceux décrits par les ondes monochromatiques présentent un intérêt particulier. Ils peuvent s'écrire, à une phase pure près, sous la forme :

$$x(t) = a\cos(2\pi f t) \tag{I.1}$$

Dans ce cas précis, l'amplitude a et la fréquence f sont des constantes. Cependant, l'expérience auditive nous pousse à considérer des signaux légèrement plus compliqués en autorisant une possible évolution temporelle du contenu fréquentiel du signal et une description plus locale ou instantanée semble alors plus adaptée [Fla93]. En référence au cas monochromatique, les notions d'amplitude et de fréquence instantanées peuvent être définies, pour un signal x, par :

$$a(t) = |z_x(t)| \tag{I.2}$$

$$f(t) = \frac{1}{2\pi}\frac{\mathrm{d}\arg(z_x(t))}{\mathrm{d}t} \tag{I.3}$$

où z_x est le signal analytique calculé à partir de x par :

$$z_x(t) = x(t) + iH(x(t)) \tag{I.4}$$

avec H la transformée de Hilbert. Même si ces sons purs sont peu courants dans notre environnement sonore, ils présentent un intérêt particulier, tout d'abord parce qu'ils constituent une entité simple pouvant être à l'origine d'une réponse du système auditif et ensuite, parce que l'oreille se base sur leurs caractéristiques pour déterminer celles des sons plus complexes.

D'une manière générale, les signaux sonores peuvent être considérés comme un ensemble de sons générés, simultanément ou non, par une ou plusieurs sources distinctes. Ces sources peuvent être séparées en deux catégories : les sources périodiques ou quasi périodiques et celles non périodiques (ou percutantes).

Le spectre des sources périodiques, voir figure I.1, est constitué d'une série de composantes fréquentielles à peu près régulièrement espacées. Ces composantes fréquentielles, généralement appelées *partiels*, sont des sinusoïdes idéalement périodiques ou quasi périodiques, qui correspondent aux sons purs décrits ci-dessus. Lorsque la fréquence des partiels de rang supérieur ou égal à 2 sont des multiples entiers de celle du premier partiel, on parle de source harmonique, le premier partiel correspondant au fondamental, le second partiel au premier harmonique et ainsi de suite. Il existe des sources pour lesquelles la fréquence des partiels de rang h, $h \geq 2$, n'est pas exactement un multiple entier du fondamental mais s'en écarte de plus en plus quand h augmente. On parle alors de sources inharmoniques. En pratique, les notions d'harmonicité et d'inharmonicité peuvent se rejoindre. Par exemple, la plupart des instruments de musique présentent une inharmonicité relativement faible qui peut parfois être négligée sans pour autant être nulle. Fletcher et Rossing [Fle98] proposent un modèle d'inharmonicité pour le piano avec une relation non linéaire entre la fréquence f_1 du premier partiel et celle du partiel de rang h :

$$f_h = hf_1\sqrt{\frac{1+h^2\gamma}{1+\gamma}} \tag{I.5}$$

avec γ le coefficient d'inharmonicité. Une valeur de 0.0004 suffit pour décaler le partiel de rang 17 à la place de celui qui serait au rang 18 s'il n'y avait pas d'inharmonicité.

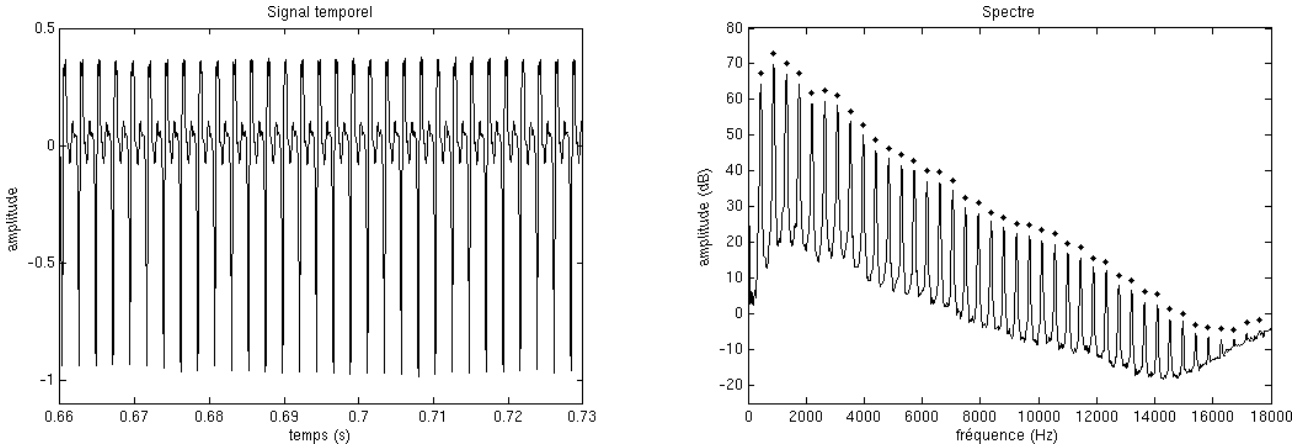


FIG. I.1 – Exemple de signal périodique : trompette jouant un la 3 (440 Hz). A droite, dans le spectre, les points correspondent aux multiples entiers de la fréquence fondamentale.

En revanche, les sources non périodiques n'ont souvent aucune structure explicite, voir l'exemple donné dans la figure I.2. Le son peut être modélisé par un signal stochastique et est généralement caractérisé par une enveloppe énergétique à large bande. Les sources de la première catégorie, qu'elles soient harmoniques ou inharmoniques, ont un spectre dit « de raies », c'est-à-dire dans lequel des partiels à des fréquences bien déterminées peuvent être mis en évidence. A l'opposé, les sources non périodiques présentent un spectre plus continu. Enfin, certaines de ces sources peuvent être réglées afin de donner l'illusion d'une note jouée, comme par exemple les tam-tams.

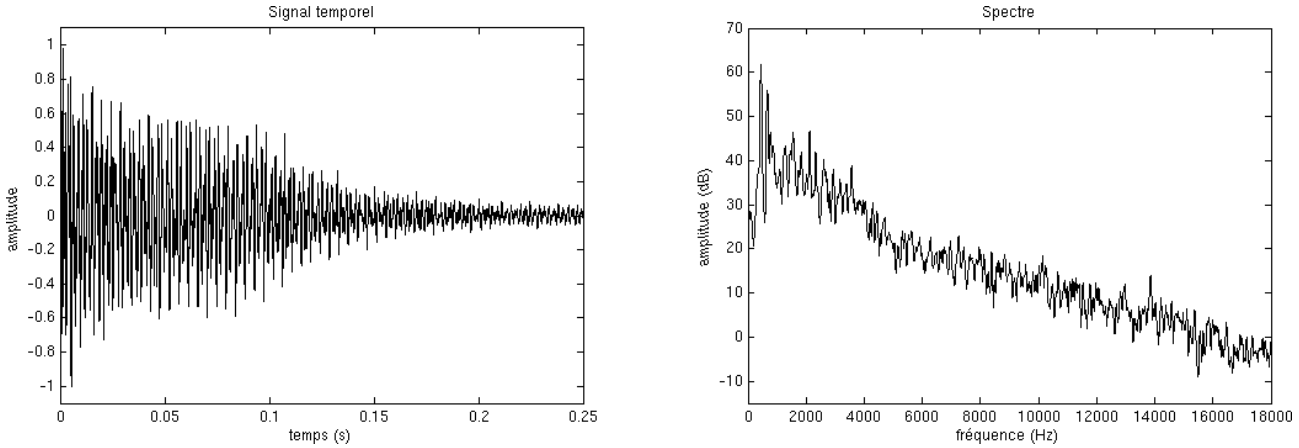


FIG. I.2 – Exemple de signal non périodique avec un instrument à percussion (timbales). Le spectre présente une structure continue, dans laquelle des pics régulièrement espacés, ne peuvent être trouvés.

Un peu à l'extrême, les cloches sont un bon exemple de mélange entre les sons quasi périodiques et les sons percutants. Après une étape transitoire plutôt non périodique, correspondant

au coup de balancier, elles présentent une structure essentiellement inharmonique avec aucune résonance proche de la fréquence fondamentale et une déviation par rapport au modèle harmonique telle qu'elle peut donner l'illusion que plusieurs notes sont sonnées. Une étude sur les cloches Hemony[1] a montré que la fréquence fondamentale peut être déterminée par les partiels 5 et 6, voire 7 [Fle98].

I.2 Hauteur et fréquence fondamentale

Des études [Wie77] portant sur l'estimation du plus petit changement de fréquence qu'un auditeur est capable de détecter, ont mis en évidence que l'oreille humaine est particulièrement sensible aux fréquences. Les expériences considérées consistaient en l'audition d'un signal parfaitement sinusoïdal, à une fréquence donnée. Dans ce cas précis, les notions de hauteur et de fréquence fondamentale peuvent être rigoureusement confondues. Cependant, s'il est incontestable que la fréquence fondamentale d'un son et sa hauteur sont deux notions intimement liées et, de fait, souvent confondues, pour les signaux réels de la vie quotidienne (qui sont plus complexes que les sinusoïdes pures), le lien qui les unit est plus subtil, il convient donc de leur donner une définition plus précise.

La hauteur est définie comme un attribut perceptif, ou lié à la sensation d'audition, permettant à l'auditeur d'ordonner les sons écoutés sur une échelle allant de « bas »à « élevé ». Elle peut être reliée à une fréquence dans la mesure où un auditeur humain est capable d'établir une correspondance, sur le plan perceptif, entre le son écouté et une sinusoïde pure à cette fréquence. Si une telle connexion ne peut pas être effectuée, on dit que le son n'a pas de hauteur. Cette définition autorise une certaine variabilité dans la détermination de la hauteur. En effet, à partir d'un certain nombre de tests d'écoute, un histogramme peut être établi et la hauteur d'un son peut, par exemple, être estimée à 440 ± 2 Hz. Cependant, cet histogramme peut présenter des ambiguïtés (en étant bi-modal, par exemple) et donner lieu à différentes conclusions (comme celle de la perception de plusieurs hauteurs). Une hauteur n'est attribuée à un son que si une large majorité d'auditeurs donne des estimations similaires et cohérentes.

La fréquence fondamentale est le pendant physique de la hauteur et correspond à l'inverse de la période du signal [Che02]. Elle est habituellement notée $F0$ dans la littérature. Cependant, cette définition ne s'applique qu'aux signaux parfaitement périodiques et les sons comme la parole, la musique ou les évènements sonores de la vie quotidienne s'écartent de la stricte périodicité. Pour la catégorie des signaux ayant un spectre de raies, la hauteur perçue par un auditeur est estimée à partir des fréquences des partiels du signal entendu. Plus précisément, la hauteur attribuée à un signal harmonique ou inharmonique, correspond à la fréquence du premier partiel[2], c'est-à-dire du fondamental [Kla04]. Les sons non périodiques peuvent aussi être classés selon une hiérarchie basée sur une pseudo hauteur qui est un reflet de la fréquence centrale du spectre continu [Hai03].

[1] L'art du fondeur de cloche consiste à réaliser une cloche dont les partiels sont accordés. Les frères Pieter et François Hemony furent parmi les premiers à trouver comment réaliser un tel réglage, au XVIIème siècle.

[2] Dans la suite de ce manuscrit, la fréquence du premier partiel d'une source quasi harmonique sera notée f_1 afin de la distinguer de $F0$. Cependant, conformément à un abus de langage très répandu dans la littérature, il y sera souvent fait référence par l'expression *fréquence fondamentale de la source*.

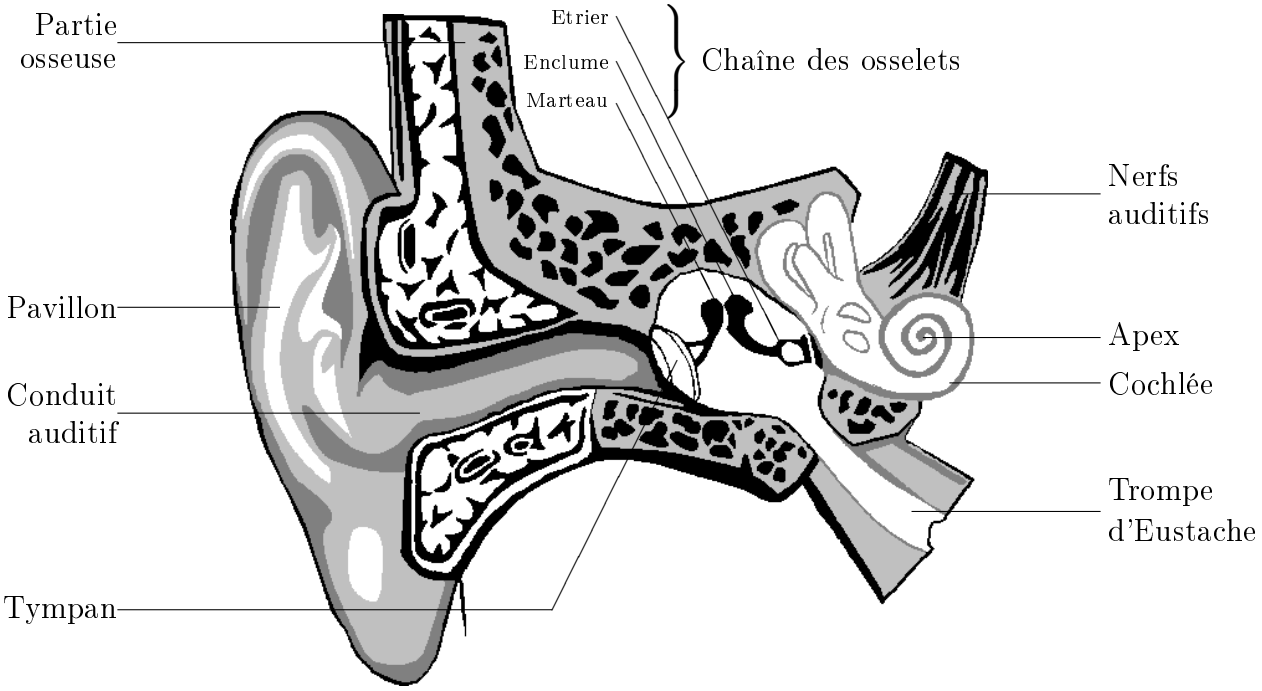


FIG. I.3 – Coupe de l'oreille.

I.3 La perception sonore et le système auditif

Le rôle essentiel que joue la hauteur dans le processus d'audition peut s'expliquer par une étude approfondie du fonctionnement de l'oreille humaine. L'organe clé dans la perception d'un son est la cochlée, dans l'oreille interne. Elle permet la transformation de l'onde de pression acoustique acheminée par la chaîne constituée du tympan, du marteau, de l'enclume et de l'étrier, en impulsions nerveuses codant convenablement l'information sonore. La cochlée, voir figure I.3, est un long tube caractérisé par sa forme hélicoïdale. Elle est divisée en deux sur toute sa longueur par la membrane basilaire. C'est cette membrane qui vibre sous l'excitation de l'onde acoustique et sur laquelle se trouvent des cellules ciliées. Ces cellules peuvent être classées en deux catégories, en fonction de l'analyse qu'elles effectuent. Les résultats expérimentaux résumés[3] dans [Har96] mettent en avant ces deux types d'analyse effectuée dans la cochlée, liés à ces deux catégories.

Le premier fait référence à l'organisation spatiale des cellules ciliées externes. Plus précisément, ces cellules ont une fréquence de résonance propre, définie par leur point d'innervation le long de la membrane basilaire. Les résonateurs des hautes fréquences se trouveraient au début de la cochlée et ceux des basses fréquences se situeraient plutôt vers l'apex de la cochlée. On parle alors d'une organisation tonotopique du système auditif dans lequel la fréquence d'un signal sinusoïdal pur est encodée par l'oreille en fonction de l'emplacement des cils qu'elle fait entrer en résonance. Le second type d'analyse, venant s'ajouter à ce système d'encodage tonotopique, permet la caractérisation d'une fréquence grâce à la structure temporelle des im-

[3]Une description plus détaillée de ces expériences est donnée dans [Har97]

pulsions dans les nerfs auditifs. Il a en effet été montré qu'il existait une corrélation entre les instants d'occurrences des impulsions nerveuses dans les nerfs auditifs et la fréquence du signal sinusoïdal pur écouté. Un des principaux arguments de l'approche temporelle vient des limites de l'approche tonotopique. En effet, cette analyse spectrale est comparable à une banque de filtres passe-bande. Comme dans une telle architecture, un principe d'incertitude comparable, bien qu'inexprimable dans les mêmes termes, au principe d'incertitude d'Heisenberg-Gabor en analyse temps-fréquence [Fla93], peut être mis en évidence. Il a été trouvé expérimentalement que la capacité d'un auditeur à distinguer deux fréquences proches diminuait avec la durée des signaux. Les psychoacousticiens s'accordent pour dire que les deux processus, spatial et temporel, sont impliqués dans la perception de la hauteur, avec néanmoins une dominance probable du processus temporel pour les basses fréquences et du processus spatial pour les hautes fréquences.

Dans le cas des signaux à spectre de raies, le modèle spatial du fonctionnement de l'oreille pour la perception de la hauteur semble être mis en défaut. En effet, les composantes du signal font entrer en résonance plusieurs cellules ciliées et il n'est pas évident de savoir quel endroit de la cochlée encodera la hauteur. Cette ambiguïté du modèle peut néanmoins être levée en considérant la règle arbitraire selon laquelle la plus basse fréquence encodée correspond à la hauteur. En revanche, le modèle temporel reste valable car, même s'il y a plusieurs partiels, la période du signal (ou la fréquence fondamentale) correspond toujours à la hauteur [Med91]. Il a aussi était montré que l'oreille était capable de retrouver la hauteur d'un son, même si le ou les premiers partiels ne sont pas physiquement présents dans le signal [Hou90]. Ce phénomène peut, par exemple, se produire lors de l'écoute, à travers un canal à bande limitée, d'une mélodie avec des passages à basses hauteurs. Même si l'énergie acoustique de la fréquence fondamentale des notes jouées est faible voire nulle, l'auditeur est toujours capable de suivre la mélodie sans ambiguïté. D'une manière générale, pour les signaux harmoniques, l'estimation de la hauteur reste précise quand le rang du plus petit harmonique présent est inférieur à 8. Au delà, l'acuité diminue progressivement jusqu'à atteindre un plateau pour un rang du plus petit harmonique égal à 12 ou 13 [Moo06].

Le cas des signaux purs ou périodiques est assez simple et les modèles perceptifs décrits précédemment ont déjà mis en évidence l'importance de la hauteur dans le codage des signaux sonores. Le cas des signaux non périodiques n'a pas été traité alors que l'oreille leur attribue, dans la majorité des cas, une hauteur. Des expérimentations [Gol73] ont montré que la connexion entre la hauteur et ces signaux complexes ressemble à un processus d'adéquation réalisé par la partie centrale du système auditif. Plus précisément, à partir des informations fournies par les capteurs que sont les oreilles, provenant à la fois de leurs analyses temporelle et spectrale, ce système central chercherait à adapter un schéma harmonique, caractérisé par une fréquence fondamentale, aux différentes composantes fréquentielles du signal sonore qu'il reçoit. Pour un signal donné, plusieurs fréquences fondamentales sont proposées et la hauteur finalement attribuée au signal correspond à celle qui s'adapte le mieux. Dans le cas où deux fréquences fondamentales différentes donnent le même résultat, on peut supposer que le signal d'entrée comporte deux hauteurs.

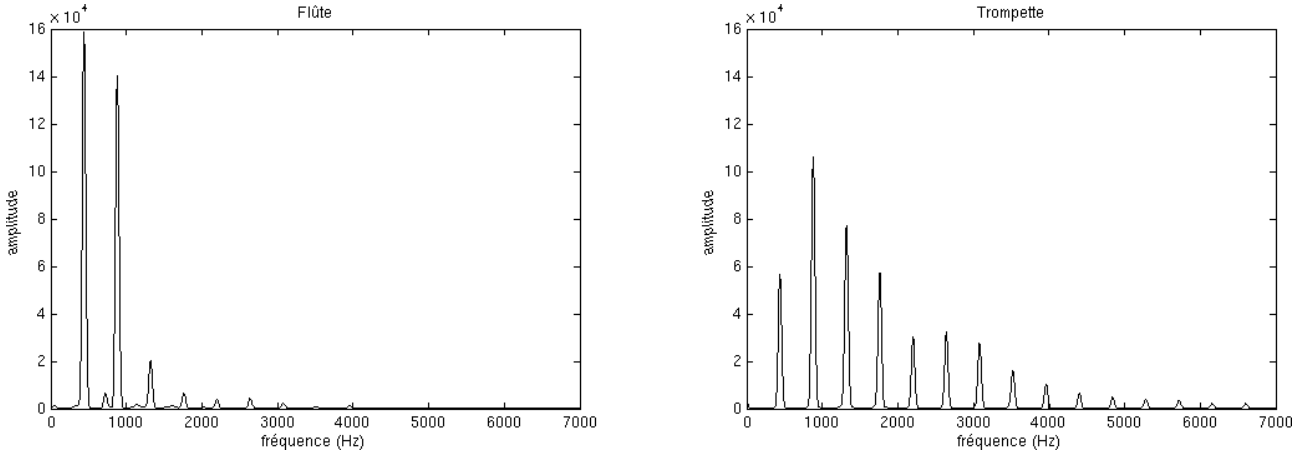


FIG. I.4 – Exemple de la distinction de deux instruments par le timbre. La flûte et la trompette jouent un la3 (440 Hz) à la même intensité.

I.4 Les autres attributs essentiels

Le volume et le timbre sont, comme la hauteur, des attributs subjectifs visant à décrire la sensation liée à la perception de différents types de signaux sonores. De même qu'il existe un lien fort entre la hauteur et la fréquence fondamentale, une corrélation directe peut être établie entre le volume et le timbre, d'une part, et les quantités physiques que sont l'intensité (ou l'énergie) sonore et la structure fréquentielle, d'autre part. Cependant, des relations exclusives hauteur/fréquence fondamentale, volume/intensité sonore ou timbre/structure fréquentielle n'existent pas et ces trois caractéristiques subjectives peuvent dépendre à la fois des trois caractéristiques physiques.

Le volume d'un son est la caractéristique permettant à l'auditeur de distinguer un son fort d'un son faible. Cette caractéristique subjective permet de décrire la force de la sensation liée à la perception sonore. Le volume d'une source sonore peut être directement relié à l'amplitude des partiels qui la composent. Des interactions fortes entre la perception de la hauteur et celle du volume peuvent être relevées. Les études menées par Wier *et al.* [Wie77] ont mis en évidence une dépendance entre l'intensité d'un signal pur et la capacité de l'auditeur à distinguer deux fréquences. L'influence du volume est plus importante dans les basses fréquences que dans les hautes fréquences, jusqu'à être pratiquement nulle aux environs de 8000 Hz. D'autres expérimentations portant sur l'évaluation de la hauteur [Fle34], ont montré qu'un son pur, ayant une fréquence constante, paraît baisser, pour des fréquences inférieures à 2500 Hz et, au contraire, s'élever légèrement si sa fréquence est supérieure à 2500 Hz, quand le volume augmente.

Le timbre, ou « couleur »d'un son, est un attribut perceptif qui est lié à la reconnaissance des sources sonores. Le timbre ne correspond pas à une simple propriété acoustique et le concept est traditionnellement défini par la négative : le timbre est la qualité d'un son par laquelle un auditeur peut distinguer deux sons de même volume et de même hauteur. Pour un signal musical, les psychoacousticiens associent volontiers la dimension psychologique de la hauteur à celle de la mélodie, afin de différencier la hauteur d'un signal sonore de son contenu fréquentiel,

qui est plutôt décrit par le timbre. La sensation de timbre est intimement liée au profil spectral du son, c'est-à-dire à la répartition de l'énergie entre les différents partiels à un instant donné, ainsi qu'à l'évolution de cette répartition au cours du temps. La figure I.4 donne un exemple, emprunté à la musique, de l'importance de l'intensité respective de chaque partiel dans la caractérisation du timbre et donc de la source sonore.

Encore une fois, la hauteur et le volume peuvent avoir une influence sur la perception du timbre [Jen61, Mar03]. En effet, il est bien connu qu'un changement dans la structure fréquentielle du signal peut induire une modification directe du timbre mais il est peut-être moins évident qu'un changement de hauteur ou de volume, sans aucune modification de la structure fréquentielle, puisse avoir une quelconque influence sur le timbre.

Le dernier attribut participant à la description de la perception sonore est la durée, qui peut simplement être définie comme l'intervalle de temps durant lequel l'oreille perçoit les vibrations du son. Les trois attributs étudiés précédemment ne sont définis que si la durée du signal dépasse un certain seuil (de l'ordre du centième de seconde). En deçà, le son est qualifié de claquement.

I.5 Discussion

Les attributs qui ont été étudié dans les parties précédentes, abordent le problème de la description de la perception sonore d'un point de vue plutôt psychoacoustique, c'est-à-dire en ayant le souci d'utiliser des descripteurs mettant en relation des grandeurs physiques et des réactions plus qualitatives, propres à chaque individu. C'est le cas de la fréquence fondamentale, de l'intensité, *etc.* De plus, ces grandeurs physiques caractérisent la forme d'onde du signal au travers d'un modèle implicite, pouvant éventuellement mener à sa reconstruction. Une autre approche pourrait être envisagée : décrire la source sonore plutôt du point de vue du signal physique, en considérant des descripteurs audio n'ayant pas forcément de signification perceptive. Quelques exemples, parmi les plus classiques, sont donnés dans le tableau I.1. Une grande partie d'entre eux peuvent être construits à partir d'une représentation temps-fréquence particulière : le spectrogramme. Pour un signal discret $\mathbf{x}$ donné, il est défini par [Hla92] :

$$\mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f] = \left|\mathrm{STFT}_{\mathbf{x}}^{\mathbf{w}}[t,f]\right|^2 \tag{I.6}$$

où STFT désigne la transformée de Fourier court terme, calculée avec la fenêtre $\mathbf{w}_t$, de longueur $L_{\mathbf{w}}$ et centrée sur l'instant t, par

$$\mathrm{STFT}_{\mathbf{x}}^{\mathbf{w}}[t,f] = \mathrm{DFT}(\mathbf{x}\cdot\mathbf{w}_t) \tag{I.7}$$

avec DFT la transformée de Fourier discrète.

Ces descripteurs présentent l'avantage de provenir d'une approche non paramétrique, ce qui permet de ne faire aucune hypothèse sur le comportement du signal étudié. Cependant, cela a pour conséquence, d'une part, que certains d'entre eux donnent une description redondante du signal et, d'autre part, que le choix des descripteurs utilisés est souvent arbitraire [Dav02a]. Le principe est alors d'essayer de trouver quelle combinaison permet de rendre compte de la caractéristique du son que nous voulons extraire et, souvent, ce choix est fait par apprentissage, en fonction de la tâche à effectuer [Foo99, Tza02a, Her03].

Nom	Définition	Description
Energie de bande	$\mathrm{BE}_{f_0,f_1}[t] = \frac{\sum_{f=f_0}^{f_1} \mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]}{\sum_{\tau=1}^{L_{\mathbf{w}}} \mathbf{w}[\tau]}$	L'énergie de bande est l'énergie contenue dans la bande $[f_0, f_1]$.
Taux d'énergie de bande	$\mathrm{BER}_{f_0,f_1}[t] = \frac{\sum_{f=f_0}^{f_1} \mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]}{\sum_{f} \mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]}$	Le taux d'énergie de bande est la proportion d'énergie contenue dans la bande $[f_0, f_1]$.
Fréquence moyenne	$\mathrm{MF}[t] = \frac{\sum_f f \mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]}{\sum_f \mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]}$	La fréquence moyenne est le moment du premier ordre du spectrogramme, à l'instant t.
Largeur de bande	$\mathrm{BW}[t] = \frac{\sum_f (f - \mathrm{MF}[t])^2 \mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]}{\sum_f \mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]}$	La largeur de bande est le moment du second ordre du spectrogramme, à l'instant t.
Coefficients Cepstraux	$\mathrm{C}_t[\tau] = \mathrm{IDFT}\Big(\frac{1}{2}\log\big(\mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]\big)\Big)$	Le cepstre est la transformée de Fourier inverse du logarithme du spectre. Pour la parole, les premiers coefficients cepstraux caractérisent les formants.
Taux de passages à zéro	$\mathrm{R0}[t] = \sum_{\tau=1}^{L_{\mathbf{w}}-1} \frac{\big\lvert \mathrm{sgn}(\mathbf{x}_t[\tau+1]) - \mathrm{sgn}(\mathbf{x}_t[\tau]) \big\rvert}{L_{\mathbf{w}}}$	C'est le nombre de changements de signe dans le signal fenêtré $\mathbf{x}_t$, défini par $\mathbf{x}_t = \mathbf{x} \cdot \mathbf{w}_t$.

TAB. I.1 – Quelques descripteurs audio parmi les plus classiques. La définition du spectrogramme $\mathrm{SP}_{\mathbf{x}}^{\mathbf{w}}[t,f]$ d'un signal temporel discret $\mathbf{x}$ est donnée par l'équation (I.6). IDFT désigne l'inverse de la transformée de Fourier discrète.

Dans les chapitres suivants, l'estimation des grandeurs physiques reliées aux attributs perceptifs décrits précédemment, a été privilégiée. Ce choix est avant tout justifié par le but ultime de ce travail : la détection d'un effet sonore ayant à la fois une composante physique (liée au signal) et perceptive (liée à la sensation induite par l'audition du signal sonore). Ce chapitre a mis en évidence le rôle déterminant que jouent la fréquence fondamentale, l'intensité et la structure fréquentielle dans le processus auditif. De plus, ces quantités donnent une description du signal parcimonieuse, complète et sans redondance dans la mesure où elles sont quasiment nécessaires et suffisantes pour reconstruire un signal produisant une sensation auditive comparable (il manque une information de phase mais elle influe peu sur la perception sonore). Enfin, même si une relation d'équivalence stricte ne peut pas être établie, ces trois quantités peuvent donner séparément une description du signal sonore ayant une signification propre. En poussant le parallèle à l'extrême, on pourrait considérer qu'elles définissent une base, au sens mathématique du terme, de représentation des signaux sonores.

Chapitre II

Estimation de fréquences fondamentales

Le travail de l'esprit ressemble à celui des abeilles.
Elles pillotent deçà delà les fleurs,
mais elles en font après le miel, qui est tout leur ;
ce n'est plus thym ni marjolaine.
M. de Montaigne

La hauteur joue un rôle prépondérant dans la perception sonore, comme il l'a été montré dans les études citées au chapitre précédent. L'oreille humaine est capable d'attribuer une hauteur à quasiment toutes sortes de sources sonores et de séparer différents sons en se basant sur cette hauteur. Une telle analyse est cependant très difficile à effectuer automatiquement avec un ordinateur et le développement de méthodes susceptibles de donner des résultats comparables reste un défi pour la recherche.

La matière première que les algorithmes traitent est le signal physique, c'est-à-dire la forme d'onde décrivant le mouvement oscillatoire d'un support (fluide ou solide) et générée par une perturbation mécanique de la pression régnant dans le milieu. Ce signal peut avoir différentes caractéristiques parmi lesquelles nous avons vu que la fréquence fondamentale, l'intensité sonore, la structure fréquentielle et la durée permettent d'obtenir une caractérisation de la perception sonore. On parle souvent d'algorithme d'estimation de fréquences fondamentales car cette quantité physique, de part son lien à la hauteur, joue un rôle important. Mais, de fait, ces algorithmes forment un tout dans lequel les quantités complémentaires doivent être prises en compte, que ce soit comme faisant aussi partie des sorties de l'algorithme ou alors, plus indirectement, comme intermédiaires pour atteindre la fréquence fondamentale. Elles peuvent même, dans certains cas, être considérées comme des paramètres de nuisance. Il est important de noter que le champs d'investigation de tels algorithmes est restreint aux sons périodiques ou quasi périodiques (sans forcément présenter une structure harmonique). Dans le premier chapitre, nous avons désigné cette catégorie de signaux comme ceux ayant un spectre de raies, c'est-à-dire comme étant constitué de composantes fréquentielles.

L'estimation de plusieurs fréquences fondamentales signifie que l'on cherche à estimer les fréquences fondamentales de plusieurs sons présents simultanément. Une donnée importante que les algorithmes doivent prendre en compte est précisément le nombre de sources sonores et son

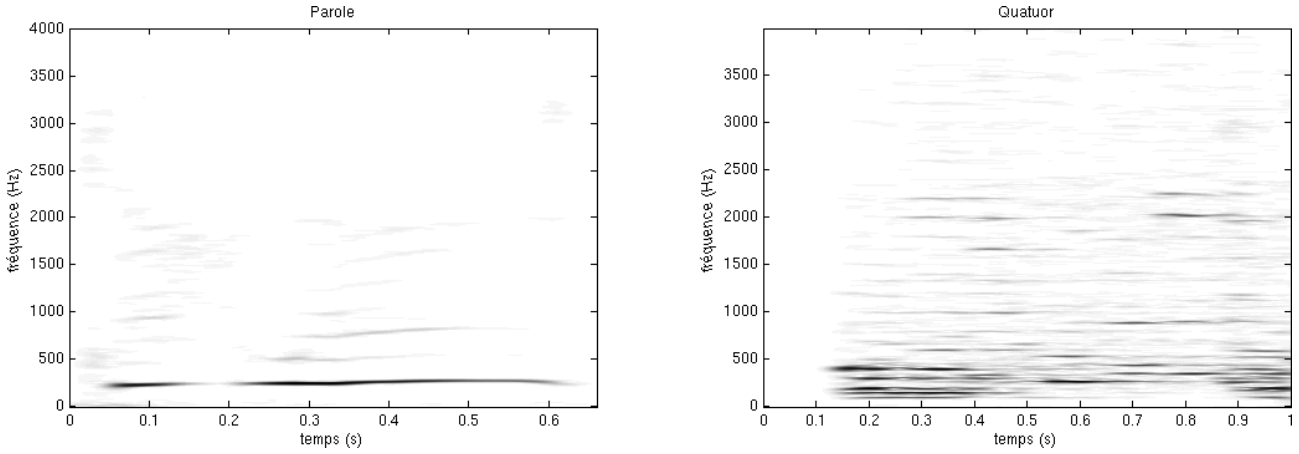


FIG. II.1 – Exemple de signaux monophonique et polyphonique. Le spectrogramme du signal de parole (mot *carrelage*) présente une structure plus simple que celle du spectrogramme du signal de musique (quatuor à cordes).

intégration parmi les paramètres à estimer fait partie des difficultés les plus reconnues [Yeh05]. Ici, une première catégorisation importante des méthodes peut être mise en avant : celles traitant les signaux monophoniques[1] et celles traitant les signaux polyphoniques. Dans la première classe, les signaux ne sont constitués au maximum que d'une seule source sonore tandis que dans la deuxième, peuvent coexister plusieurs sons. La figure II.1 donne un exemple de signal monophonique et de signal polyphonique. Le domaine de prédilection de l'estimation de fréquences fondamentales dans le cas monophonique est la parole. Dans le cas polyphonique, c'est la musique avec le problème particulier de la transcription automatique de la musique [Kla06]. Ceci explique pourquoi la grande majorité des méthodes existantes ont été développées dans l'un de ces deux domaines. Même si le cas monophonique peut généralement être considéré comme un cas particulier de polyphonie, cette subdivision garde tout de même sa pertinence car les principes fondamentaux et la difficulté de mise en œuvre de la procédure d'estimation ne sont pas les mêmes dans ces deux cas. Le fait de savoir qu'il n'y a qu'une seule source dans le signal étudié, permet de développer des méthodes à la fois plus simples et plus robustes [Plu02]. Dans le cas polyphonique, plusieurs hypothèses peuvent être faites concernant le nombre de sources. Il peut, par exemple, être supposé connu, ou encore être considéré inconnu mais fixé. Au delà de ces hypothèses, se pose le problème de son estimation, qui est intimement lié à celui de l'estimation des instants de début et de fin d'occurrence de chaque source. Là encore, deux types d'approches peuvent être distingués : les méthodes qui estiment le nombre de sources au cours du temps conjointement aux différents paramètres de chacune de ces sources et celles travaillant sur des signaux ou des portions de signal dont le nombre de sources est constant (connu ou non). Dans cette dernière catégorie, une segmentation des données est souvent un pré-traitement indispensable. Rossignol [Ros00] définit cette étape comme la détection des variations brusques

[1] Le terme monophonique peut être ambigu dans la mesure où il peut signifier le contraire de stéréophonique, c'est-à-dire désigner une technique de prise de son par un seul canal, ne donnant pas l'impression d'un relief sonore. Comme de tels procédés ne sont pas pris en compte dans ce manuscrit, cette signification ne sera pas retenue.

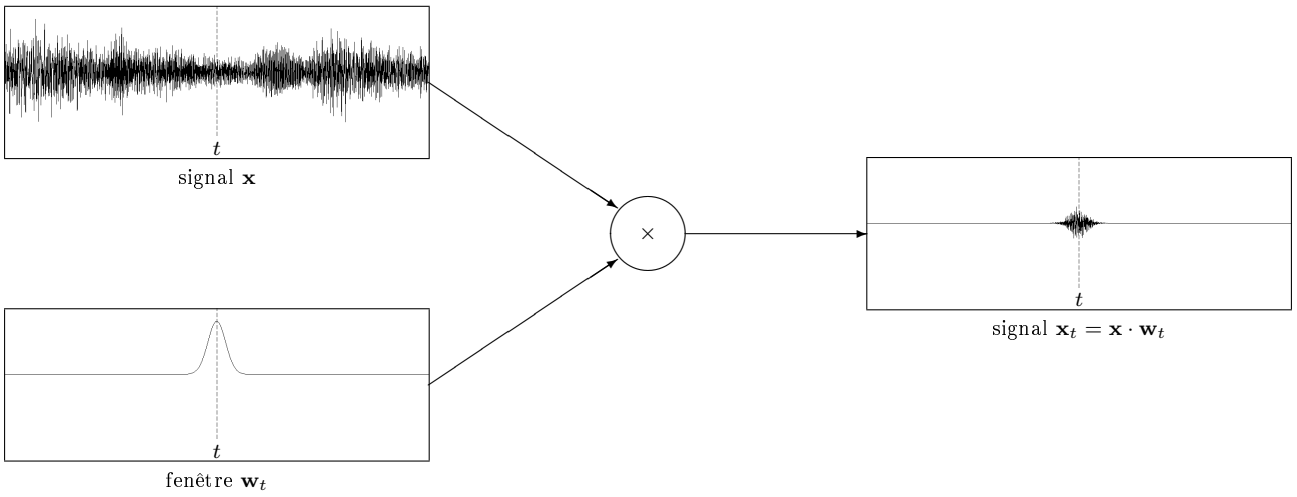


FIG. II.2 – Principe de calcul d'un segment de signal centré sur l'instant t. En pratique, le signal fenêtré $\mathbf{x}_t$ a la même longueur $L_\mathbf{w}$ que la fenêtre.

du signal, c'est-à-dire la détection des transitions entre deux zones stables successives. Beaucoup de travaux traitent de ce problème en se basant sur plusieurs outils comme les machines à vecteurs supports (*Support Vector Machine*) [Des05], les méthodes bayésiennes pour la régression [Pun02], la combinaison des différences d'énergie et des déviations de phase [Dux03] ou les modèles de Markov cachés (*Hidden Markov Model*) [Rap99]. Une revue étendue des travaux pour la segmentation dépasse le cadre de ce manuscrit, le lecteur intéressé pourra se reporter, par exemple, à [Bel05].

Ce chapitre donne un aperçu des méthodes existantes ainsi que des principes fondamentaux et outils utilisés. Une présentation taxinomique des travaux est une tâche difficile au vu de la grande diversité des approches proposées dans le passé. De plus, une énumération exhaustive (si tant est qu'elle soit possible) dépasse le cadre de ce manuscrit (voir pour cela [Ste99, Hai01, Kla06]). La première partie est dédiée aux outils et concepts de base, utilisés dans beaucoup de méthodes. Puis, dans les trois parties suivantes, un certain nombre d'approches seront décrites. Elles sont classées en trois catégories, correspondant chacune à une partie : les méthodes s'appuyant sur le fonctionnement de la perception sonore par l'oreille humaine, celles utilisant des architectures permettant de combiner des processus *bottom-up* avec des connaissances *top-down* et enfin, celles se basant sur un modèle génératif du signal. La dernière partie fait une synthèse de cette revue et pose les bases de la méthode proposée dans ce manuscrit.

II.1 Quelques outils et notions de base

La plupart des algorithmes estiment la ou les fréquences fondamentales présentes à l'instant d'analyse t, à partir d'un fragment de signal (ou trame), centré sur l'instant t. La figure II.2 précise comment ces segments sont calculés à partir d'un signal discret $\mathbf{x}$, à l'aide d'une fenêtre d'analyse $\mathbf{w}_t$, de longueur $L_\mathbf{w}$ et centrée sur l'instant t. Parmi les fenêtres les plus classiques, on peut citer : rectangulaire, Hamming, Blackman, Gauss, ... Chacune de ces fenêtres a ses propres

caractéristiques, avantages et inconvénients, voir [Har78] pour une discussion plus complète.

II.1.1 La fonction d'auto-corrélation

Les signaux périodiques ou quasi périodiques présentent des périodicités à la fois dans les domaines temporel et spectral. Un outil particulièrement adapté à l'estimation de périodicités dans un signal est la fonction d'auto-corrélation et, de fait, bon nombre d'algorithmes se basent dessus pour estimer la fréquence fondamentale. Pour un segment de signal discret $\mathbf{x}_t$, défini à l'aide d'une fenêtre (voir la figure II.2), la fonction d'auto-corrélation est définie, dans sa version locale, par :

$$\mathrm{R}_t[\tau] = \sum_{u=-\frac{L_{\mathbf{W}}}{2}}^{\frac{L_{\mathbf{W}}}{2}} \mathbf{x}_t[u]\mathbf{x}_t[u-\tau] \tag{II.1}$$

Le caractère local fait référence au fait que la fonction d'auto-corrélation est rapportée à la date courante t et n'est pas contrainte à ne dépendre que du retard τ (comme ce serait le cas si le signal était stationnaire) [Fla93]. Pour des raisons de rapidité[2] [Sla90], elle peut être calculée à partir de la transformée de Fourier discrète :

$$\mathrm{R}_t[\tau] = \mathrm{IDFT}\left(\left|\mathrm{DFT}(\mathbf{x}_t)\right|^2\right) \tag{II.2}$$

Cette expression permet aussi de mettre en évidence le lien conceptuel qui peut exister entre la fonction d'auto-corrélation R_t et le cepstre C_t (voir tableau I.1, page 13). L'une se base sur le module au carré de la transformée de Fourier discrète et l'autre sur son logarithme. Pour l'estimation de la fréquence fondamentale, les performances des algorithmes utilisant l'une ou l'autre des approches peuvent être antagonistes. Par exemple, dans les signaux de parole, la fonction d'auto-corrélation met en évidence les pics correspondant aux composantes spectrales, en les faisant ressortir, mais est sensible à la structure des formants, tandis que le cepstre est capable de séparer les caractéristiques périodiques du signal mais est sensible au bruit [Nol64, Rab76].

Dans l'équation (II.1), la fonction d'auto-corrélation est calculée sur le signal brut afin de mesurer les périodicités temporelles. Pour reprendre la classification de Klapuri [Kla04], cela équivaut à rechercher des composantes fréquentielles à des emplacements précis dans le spectre de puissance. Une autre approche, pour estimer la fréquence fondamentale, consiste en la recherche d'intervalles spectraux. En effet, pour un signal quasi périodique ayant un spectre de raies, la distance entre deux pics consécutifs correspond à la fréquence fondamentale. Dans ce cas, la fonction d'auto-corrélation est calculée sur le spectre afin de mesurer les périodicités spectralles :

$$\tilde{\mathrm{R}}_t[\nu] = \sum_{v=1}^{L_{\mathbf{DFT}}} \left|\mathbf{X}_t[v]\right|\left|\mathbf{X}_t[v-\nu]\right| \tag{II.3}$$

avec $\mathbf{X}_t = \mathrm{DFT}(\mathbf{x}_t)$ et $L_{\mathbf{DFT}}$ le nombre de points utilisés pour calculer la transformée de Fourier discrète. Une limite de l'approche par auto-corrélation temporelle est qu'elle perd en

[2] Grâce, notament, à l'algorithme FFT (*Fast Fourier Transform*) de calcul rapide de la transformée de Fourier discrète [Coo65].

performance quand le signal n'est pas exactement périodique alors que l'approche spectrale s'avère plus à même de manipuler les signaux inharmoniques. Un autre point intéressant est que les méthodes basées sur R_t sont enclines à diviser la fréquence fondamentale par deux tandis que celles basées sur $\tilde{\mathrm{R}}_t$ la multiplieront par deux.

Plusieurs raffinements peuvent être apportés à la fonction d'auto-corrélation. Brown *et al.* [Bro89] proposent une méthode pour obtenir une fonction avec des pics plus étroits. Kunieda *et al.* [Kun96] proposent de prendre le logarithme du spectre et de lui appliquer un traitement pour en supprimer l'enveloppe spectrale avant de calculer la fonction d'auto-corrélation. Enfin, de Cheveigné et Kawahara [Che02] utilisent une série de cinq étapes de post-traitement de la fonction d'auto-corrélation afin de prévenir les erreurs d'estimation de la périodicité du signal.

II.1.2 Les représentations temps-fréquence

Une autre approche peut généraliser l'utilisation du concept de corrélation, c'est celle utilisant les représentations temps-fréquence [Hla92]. Ces représentations sont des transformations bilinéaires du signal dont un exemple très répandu est le spectrogramme (voir équation (I.6) page 12). Le spectrogramme appartient à la classe de Cohen et, comme toutes les représentations de cette classe, il peut être calculé à partir de la distribution de Wigner-Ville. Pour un signal continu x, elle est définie par :

$$\mathrm{WV}_x(t,f) = \int_\tau x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) \mathrm{e}^{-j2\pi f\tau}\mathrm{d}\tau \tag{II.4}$$

$$= \int_\nu X\left(f+\frac{\nu}{2}\right) X^*\left(f-\frac{\nu}{2}\right) \mathrm{e}^{j2\pi t\nu}\mathrm{d}\nu \tag{II.5}$$

où * désigne le complexe conjugué. La classe de Cohen peut aussi être construite à partir de la fonction d'ambiguïté, qui est définie par :

$$\mathrm{A}_x(\tau,\nu) = \int_t x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) \mathrm{e}^{-j2\pi\nu t}\mathrm{d}t \tag{II.6}$$

$$= \int_f X\left(f+\frac{\nu}{2}\right) X^*\left(f-\frac{\nu}{2}\right) \mathrm{e}^{j2\pi tf}\mathrm{d}f \tag{II.7}$$

grâce à la relation bijective qui lie WV_x et A_x :

$$\mathrm{A}_x(\tau,\nu) = \int_t\int_f \mathrm{WV}_x(t,f)\mathrm{e}^{-j2\pi(\nu t-\tau f)}\mathrm{d}t\mathrm{d}f \tag{II.8}$$

La fonction d'ambiguïté mesure essentiellement une corrélation temps-fréquence, c'est-à-dire le degré de ressemblance qu'un signal partage avec ses différentes translatées dans le plan [Fla93]. Pour construire une représentation appartenant à la classe de Cohen, il suffit de faire une double convolution de la distribution de Wigner Ville par un noyau, qui conférera ses propriétés à la représentation ainsi obtenue [Dav00].

La distribution de Wigner-Ville, même si elle est optimale pour le traitement des signaux n'ayant qu'une composante, fait apparaître des termes d'interférence si le signal en contient plusieurs, ce qui rend difficile sa lecture. Pielemeier *et al.* [Pie96] proposent un noyau pour lisser la

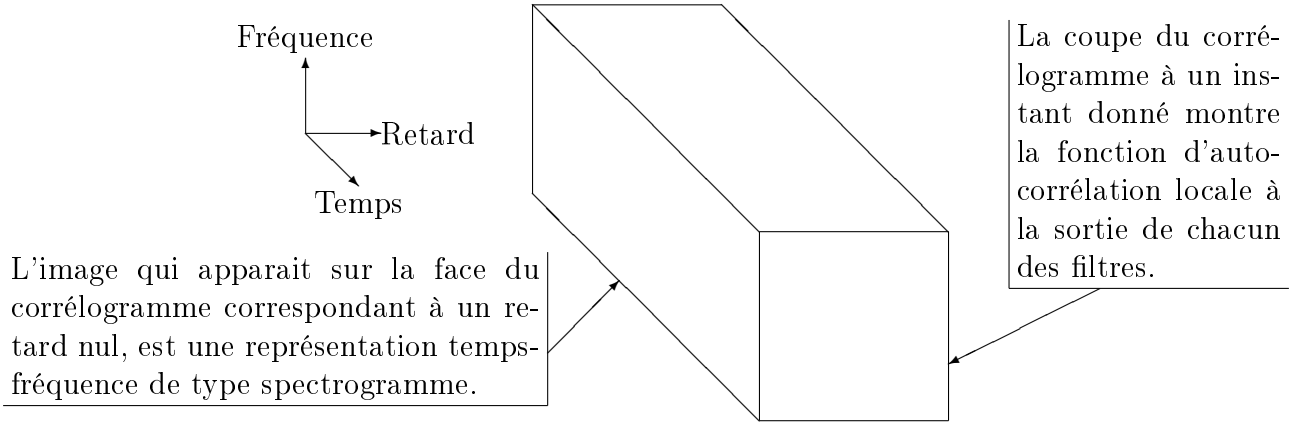


FIG. II.3 – Structure d'un corrélogramme.

distribution de Wigner Ville et obtenir une « distribution modale »optimisée pour le traitement des signaux musicaux. En effet, les lissages temporel et fréquentiel correspondent aux modulations possibles en amplitude et en fréquence. Sterian [Ste99] utilise cette distribution pour sa bonne résolution, à la fois en temps et en fréquence, par rapport à d'autres représentations plus classiques comme le spectrogramme ou la transformée en ondelettes.

II.1.3 Le corrélogramme

Le corrélogramme [Sla93b] permet de mieux appréhender la structure temporelle d'un signal. Il représente un son comme une fonction de trois variables : le temps, la fréquence et la périodicité. La première étape du calcul est le passage du signal acoustique au travers d'une banque de filtres passe-bandes, sensée modéliser l'analyse fréquentielle effectuée par la cochlée (voir section II.2.1), afin d'obtenir une carte en deux dimensions (temps et fréquence) du signal. Cette image dans le plan temps-fréquence est une distribution d'énergie et peut être assimilée à un spectrogramme. Puis, la périodicité de chacun des signaux en sortie des filtres est mesurée par la fonction d'auto-corrélation locale, ce qui donne la troisième dimension (retard) de la représentation (voir figure II.3).

II.2 Les méthodes basées sur un modèle auditif

Le fonctionnement de l'oreille humaine et l'estimation de fréquences fondamentales par un ordinateur sont deux taches différentes. Cependant, l'étude de la première a largement influencé le développement de la seconde et beaucoup de méthodes se basent sur un modèle auditif [Kla06]. La première partie de cette section présente les modèles auditifs qui peuvent être utilisés et la seconde donne quelques exemples d'application dans le cas polyphonique.

II.2.1 Modèles informatiques de la perception

Le fonctionnement de l'oreille, et plus généralement de la perception sonore, présenté rapidement dans le chapitre I peut être modélisé en vue de son imitation par un ordinateur en utilisant des outils de traitement du signal. Une telle modélisation est parfaitement justifiée dans le cadre du développement d'un algorithme d'estimation de fréquences fondamentales, dans la mesure où nous avons vu que l'oreille humaine utilise largement cette notion.

Les deux organes importants dans la transformation du signal sonore et son interprétation par un auditeur sont la cochlée et le cortex auditif, dans le cerveau. La cochlée transforme la vibration mécanique de la membrane basilaire en impulsions nerveuses. La modélisation de ce processus se fait par les deux étapes suivantes.

Analyse fréquentielle

Une analyse fréquentielle du signal acoustique permet de rendre compte des différents modes d'oscillation de la membrane basilaire. Elle consiste dans le passage du signal au travers d'une banque de plusieurs dizaines de filtres passe-bandes (*filtres auditifs*), dont la fréquence centrale est uniformément répartie sur une échelle logarithmique, pour obtenir une représentation multi-canal. Les composantes fréquentielles du signal étudié sont donc séparées par ces filtres et sont codées indépendamment à condition qu'elles ne soient pas trop rapprochées. Se pose ici la question de la largeur et de la forme de ces filtres, afin d'éviter le phénomène de masquage qui survient lorsque deux composantes se retrouvent dans la même bande de fréquences, c'est-à-dire lorsque l'écart de fréquence qui les sépare est inférieur à la largeur critique [Swe62, Pat76]. Un des filtres auditifs les plus répandus est celui basé sur la fonction *gammatone*. Ce choix peut être justifié par trois arguments [Pat96] : la réponse impulsionnelle de ce filtre correspond parfaitement aux données expérimentales obtenues par de Boer et de Jongh [Boe78], les propriétés de résolution fréquentielle de ce filtre s'accordent à celles mesurées lors d'expériences psychophysiques sur des auditeurs humains et, enfin, sa mise en œuvre technique est facilitée en utilisant une cascade de filtres du second ordre [Sla93a].

Génération des impulsions nerveuses

Le signal issu de chacune des bandes (ou *canaux auditifs*) est traité afin de modéliser la génération des impulsions nerveuses. Le modèle recherché prend en entrée les signaux filtrés issus de chacun des canaux et donne en sortie le train d'impulsions dans les nerfs auditifs. En pratique, étant donné le nombre élevé de cellules ciliées, on recherche plutôt le fonction de probabilité d'occurrence des impulsions nerveuses dans le canal considéré. Une suite de trois opérations permet d'obtenir cette fonction :

- **compression** : le but est de mettre les signaux filtrés à l'échelle en les rendant inversement proportionnel à leur variance.
- **redressement** : la mise à zéro de la moitié négative de l'onde (*half-wave rectification*) est un processus non linéaire permettant de faire la synthèse des périodicités temporelle et fréquentielle. En effet, il fait apparaître des pics dans le spectre correspondants aux écarts fréquentiels entre partiels. Dès qu'un signal contient plus d'une composante fréquentielle, il apparaît un phénomène de battement de la forme d'onde temporelle, qui se retrouve

dans l'enveloppe du signal. Le spectre de cette enveloppe est précisément composé des pics que cette transformation fait apparaître.
- **filtrage passe-bas** : il permet d'équilibrer le poids entre les pics spectraux des composantes originelles du signal et ceux de l'enveloppe. La fréquence de coupure est généralement de l'ordre de 1 kHz.

Un modèle plus proche du fonctionnement réel de l'oreille a été proposé par Meddis [Med86]. Il est basé sur un système de trois équations différentielles simulant l'activité des neurotransmetteurs au niveau des synapses des nerfs auditifs. Cependant, ce modèle réaliste, comme d'autres qui ont été avancés (voir [Hew91]), présente des limites notamment en ce qui concerne le niveau d'énergie maximum acceptable du signal d'entrée. Il reste néanmoins un modèle qui a marqué les esprits et sur lequel beaucoup d'algorithmes se basent.

Cette modélisation en deux étapes du fonctionnement de la cochlée est assez largement répandue et utilisée. En revanche, le processus de traitement effectué par le cerveau dans le cortex auditif, est moins bien connu. Cependant, dans les différentes théories avancées, deux étapes peuvent être distinguées : d'abord une analyse de la périodicité du signal dans chaque canal auditif est effectuée, puis une intégration transversale des données portées par les canaux est réalisée, afin d'obtenir une information sur la hauteur perçue. Les modèles existants de ce processus font souvent intervenir la fonction d'auto-corrélation, pour la première étape et une version sommée sur tous les canaux, pour la seconde.

Un exemple typique d'algorithme d'estimation de la fréquence fondamentale dans ce contexte, est donné par la méthode proposée par Meddis et Hewitt [Med91]. Après un pré-traitement permettant de reproduire les transformations que subit le signal dans les oreilles externe et moyenne, cette méthode opère en trois grandes étapes :
- une analyse fréquentielle est d'abord effectuée par une banque de 128 filtres gammatone
- la sortie de chaque filtre passe ensuite par le modèle de fonctionnement des cellules ciliées de Meddis
- enfin, la fonction d'auto-corrélation est calculée au sein de chaque canal sur la sortie de ce modèle

La fréquence fondamentale est estimée en prenant le maximum de la moyenne des résultats obtenus à la dernière étape.

II.2.2 Extension au cas polyphonique

Les algorithmes qui se basent sur un modèle auditif, comme celui de Meddis et Hewitt, s'ils donnent de bons résultats dans le cas monophonique, se généralisent souvent mal au cas polyphonique. En effet, les fréquences fondamentales qui se mélangent dans un signal sonore polyphonique, ne peuvent pas simplement être estimées à partir des pics de la fonction d'auto-corrélation sommée, car les différentes sources présentes peuvent avoir beaucoup d'informations en commun que la fonction d'auto-corrélation fusionne de telle sorte qu'il est extrêmement difficile de les séparer [Kla98a]. Différentes approches permettent d'étendre de manière robuste les modèles auditifs dans ce contexte multi-source. De Cheveigné [Che93] dégage deux grandes catégories de stratégies de séparation de sources : l'amélioration et l'annulation. D'une manière générale, l'amélioration permet de mettre en valeur telle ou telle source par rapport au reste du signal, en utilisant sa périodicité. A l'opposé, l'annulation permet de retirer une source en

utilisant sa structure harmonique. Il est important de noter que, préalablement à ces deux stratégies, une étapes d'estimation des fréquences fondamentales est nécessaire. Là encore, certaines méthodes les estiment de manière itérative et d'autres, de manière simultanée. Cette partie utilise cette distinction pour présenter différentes méthodes d'extension de modèles auditifs pour l'estimation dans le cadre polyphonique.

II.2.2.a Estimation itérative

Le principe de base est assez simple : une fois qu'une fréquence fondamentale est détectée (d'une manière ou d'une autre), les partiels lui correspondant sont retirés du mélange et le processus est itéré sur le résidu.

Meddis et Hewitt [Med92] ont proposé une méthode de traitement de la parole pour détecter et reconnaître deux voyelles prononcées simultanément avec, ou non, la même fréquence fondamentale. En amont du processus de reconnaissance des voyelles, un module de leur algorithme permet d'estimer la ou les deux fréquences fondamentales présentes. Il fonctionne de la manière suivante : après avoir estimé la période correspondant au plus grand pic dans la fonction d'auto-corrélation sommée, il retire les canaux présentant un pic à cette période dans leur fonction d'auto-corrélation. Avec les canaux restant, une nouvelle moyenne des fonctions d'auto-corrélation est calculée afin d'estimer la deuxième fréquence fondamentale. La décision sur le nombre de fréquences fondamentales présentes (1 ou 2) est prise en fonction du pourcentage de canaux retirés à la première étape (le seuil est de 80 %).

Toujours dans le cadre de l'expérience des deux voyelles, de Cheveigné [Che97b] propose l'utilisation de filtres d'annulation dans le domaine temporel. Ces filtres, quand ils sont réglés sur une certaine période, enlèvent du signal total le signal périodique dont l'intervalle entre les composantes est égal à cette période. Ainsi, comme dans la méthode de Meddis et Hewitt, la voyelle dominante est retirée du signal et le résidu est utilisé pour reconnaître la seconde voyelle. La différence entre les deux méthodes réside dans le fait que la séparation entre les deux voyelles se fait à l'intérieur de chacun des canaux et non pas de manière transversale. On peut noter que, pour ces deux méthodes, l'opération peut être répétée pour estimer d'autres périodes ou pour affiner les estimations initiales de façon récursive.

La méthode de Klapuri [Kla05] ne se limite pas à deux sources. L'originalité de la méthode réside dans la manière dont sont effectuées les étapes d'analyse de la périodicité et de combinaison entre tous les canaux. Le résultat est une fonction qui est un indicateur robuste d'une des fréquences fondamentales présentes dans le signal, d'où la mise en place du processus itératif pour estimer les autres fréquences fondamentales. On peut noter que la soustraction des partiels correspondant à la fréquence fondamentale détectée se fait dans le domaine fréquentiel. De plus, ce processus ne retire totalement que les partiels de bas rang. Dans les résultats présentés, le nombre de sources (1, 2, 4 ou 6 selon les tests) est donné en entrée de l'algorithme. A titre de comparaison, la méthode présentée dans [Kla03] ne fait pas d'hypothèse sur le nombre de sources, le processus d'itération est arrêté lorsque le poids correspondant à l'itération courante devient inférieur à un certain seuil fixé à l'avance. On peut aussi noter que cette méthode ne se base pas sur un modèle auditif mais sur des techniques spectrales.

II.2.2.b Estimation jointe

Une méthode d'estimation simultanée des fréquences fondamentales a été proposée par de Cheveigné et Kawahara [Che99]. L'espace des paramètres d'une cascade d'autant de filtres d'annulation que de sources considérées, est parcouru de manière exhaustive. Pour chaque jeu de paramètres, la puissance du signal de sortie est calculée. Une fois que tout l'espace est parcouru, les estimations des périodes sont données par les paramètres permettant de minimiser cette puissance.

Karjalainen et Tolonen [Kar99, Tol00] ont proposé une amélioration de la fonction d'auto-corrélation afin que toutes les fréquences fondamentales puissent être extraites directement du résultat. Leur méthode se démarque de l'approche plus classique de Meddis et Hewitt en plusieurs points. Tous d'abord, un pré-traitement est appliqué au signal pour supprimer les corrélations à court terme. Cette étape est fonctionnellement équivalente à celle de compression. Ensuite, le signal est étudié au travers de deux filtres : un passe-bas et un passe-haut avec une fréquence de coupure de 1 kHz. Le signal du canal passe-haut est ensuite traité par la série redressement/filtrage passe-bas afin d'en estimer son enveloppe. La suite correspond au schéma habituel où la fonction d'auto-corrélation est calculée dans chaque canal, puis les résultats sont additionnés. Une deuxième démarcation se trouve dans la façon de calculer la fonction d'auto-corrélation. L'équation (II.2) est utilisée mais le module de la transformée de Fourier est élevé à une puissance inférieure à 2. Enfin, un ultime bloc dans la chaîne de traitement permet de supprimer la redondance de la fonction d'auto-corrélation sommée. Cette dernière est d'abord écrêtée pour ne laisser que les valeurs positives puis allongée dans le temps d'un facteur 2 et soustraite à la fonction originale. Le résultat est de nouveau écrêté. Le processus est répété pour les facteurs 3, 4, *etc.* le but étant de supprimer les pics de la fonction d'auto-corrélation sommée correspondant aux partiels d'ordre supérieur à 2. A la fin, les pics de la fonction obtenue correspondent aux fréquences fondamentales présentes dans le signal.

Dans la méthode de Wu *et al.* [Wu03], l'estimation des périodes des K sources de parole présentes, avec $K = 0, 1$ ou 2, est effectuée avec le souci d'obtenir une méthode robuste au bruit. Après une analyse fréquentielle avec 128 filtres gammatone, les canaux sont séparés en deux groupes : ceux dont la fréquence centrale est inférieure à 800 Hz et ceux où elle est supérieure. L'enveloppe des signaux filtrés de la dernière catégorie est ensuite calculée. La fonction d'auto-corrélation normalisée est calculée dans tous les canaux. Une deuxième étape permet de sélectionner les canaux non corrompus par le bruit puis de sélectionner les pics de la fonction d'auto-corrélation de ces canaux, susceptibles de correspondre à des fréquences fondamentales. La troisième étape de l'algorithme permet une intégration des informations de périodicité au travers de tous les canaux. Cette intégration ne se fait pas par la manière classique de sommation de la fonction d'auto-corrélation, mais plutôt par le calcul d'une fonction de probabilité conditionnelle d'observation de l'ensemble des pics sélectionnés, étant donné les K périodes proposées. Enfin, pour obtenir des trajectoires de fréquences fondamentales continues, les auteurs proposent d'utiliser un modèle de Markov caché (*Hidden Markov Model*) à saut sur un espace qui est l'union de trois sous-espaces correspondants aux cas $K = 0, 1$ ou 2. Ces espaces sont discrétisés et la trajectoire est estimée avec l'algorithme de Viterbi [Vit67].

II.3 Insertion de connaissances - Groupement de partiels

La nécessité de tenir compte des autres sons présents, pour l'estimation de la fréquence fondamentale d'une source sonore noyée dans un flux polyphonique, pousse Bregman [Bre90] à lier le processus d'estimation de fréquences fondamentales à celui de la séparation de sources. Chaque source est composée de partiels dont la fréquence peut être reliée, d'une manière ou d'une autre, à la fréquence fondamentale de la source. Si chaque partiel était assigné à la source qui l'a généré, l'estimation des fréquences fondamentales serait plus aisée. C'est là que réside un des problèmes majeurs du cas polyphonique, car plusieurs partiels peuvent appartenir à la même source. Par exemple, dans un signal musical, tout ou partie de la structure spectrale d'une note peut être contenue dans celle d'une autre note[3].

Il est donc nécessaire d'avoir recours à des connaissances *a priori* (externes au signal sonore) afin de regrouper les partiels issus d'une même source et d'être capable de gérer le cas où un même partiel appartient à plusieurs sources. Différents types d'informations peuvent être utilisés, en fonction du problème posé. En se basant sur les principes gestalt[4] développés par les psychologues, Bregman propose quelques indices à prendre en compte pour mener à bien le processus de regroupement. On peut citer, parmi d'autres : proximité temporelle et spectrale, début et fin synchronisés, harmonicité, modulation en fréquence et en amplitude cohérentes, similarité de la localisation spatiale, timbre, *etc.* (voir [Bre90] pour une revue plus complète). Il existe différentes manières d'inclure ces informations extérieures. La première partie de cette section est consacrée aux architectures de type « tableau noir ». Dans la seconde partie, quelques méthodes importantes dans le contexte du regroupement de partiels sont présentées.

II.3.1 Architectures de type « tableau noir »

Les systèmes présentant une architecture de type « tableau noir »(*blackboard architecture*) ont été développés à l'origine dans le domaine de l'intelligence artificielle [Cor91] et permettent l'intégration de connaissances ou d'algorithmes divers dans un même processus d'estimation ou de décision. L'idée est de reproduire le fonctionnement du travail effectué par une équipe d'experts réunis autour d'un tableau, pour résoudre un problème donné. Les trois composantes principales d'une telle architecture sont :

- le tableau noir
- les sources de connaissances
- l'administrateur

Les sources de connaissance (ou experts) correspondent à de multiples modules de traitement qui échangent des informations au sein d'un espace commun (le tableau noir) sous la direction d'un administrateur. Le tableau peut être vu comme un espace hiérarchique dans lequel le plus bas niveau correspond au signal brut et le plus haut niveau à la solution du problème. Les experts regardent le problème évoluer sur le tableau et interviennent quand l'administrateur les y invite. Leur contribution consiste en la formulation d'hypothèses en se basant sur les données

[3]Dans la gamme tempérée utilisée dans la musique occidentale, le rapport des fréquences fondamentales des notes est de $\frac{2}{1}$ pour une octave, $\frac{3}{2}$ pour une quinte et $\frac{4}{3}$ pour une tierce.

[4]Le mot *gestalt* vient de l'allemand et signifie *schéma* (*pattern* en anglais). Initialement développés dans le domaine de la vision, ces principes décrivent le processus de création par le cerveau de schémas mentaux, permettant d'établir des connexions entre les données venant des sens.

à l'entrée du tableau et sur les hypothèses des autres sources de connaissance. Le problème est résolu lorsque tous les experts sont satisfaits par toutes les hypothèses présentes sur le tableau, avec une marge d'erreur donnée. L'étape d'entrée du processus est un traitement du signal étudié. Le but est d'obtenir plusieurs propositions de solution au problème traité et de les présenter sur le tableau noir. Dans notre cas, cela correspond souvent à extraire les différents partiels du signal et à les proposer comme candidats éventuels de fréquences fondamentales. Ici, plusieurs outils peuvent être utilisés (comme la fonction d'auto-corrélation, la transformée de Fourier, les représentations temps-fréquence, *etc.*). Ensuite, les experts, ou les sources de connaissance externes, interviennent individuellement quand leur domaine de compétence est requis.

Martin a développé une telle architecture, dédiée à la transcription de morceaux de piano à quatre voix. Dans la première version de son travail [Mar96a], le traitement initial du signal se fait par une analyse temps-fréquence (transformée de Fourier court terme) pour extraire les partiels présents dans chaque segment. La segmentation est faite en se basant sur les pics d'énergie du signal acoustique. Dans une seconde version [Mar96b], il préconise l'utilisation du corrélogramme proposé par Ellis [Ell96] afin d'améliorer la détection de notes en relation d'octave. L'entrée du tableau est une liste de composantes fréquentielles caractérisées par une fréquence, une amplitude et l'instant de début. Les sources d'information ont une structure « si/alors »et sont classées par ordre décroissant de bénéfice attendu. Martin utilise treize sources d'information pouvant se regrouper en trois grandes catégories : celles se basant sur la physique, celles se basant sur la théorie de la musique et celles permettant de gérer les conflits entre hypothèses. Les niveaux hiérarchiques qui composent le tableau sont au nombre de cinq : composante fréquentielle, partiel, note, écart entre notes et accord. La sortie du processus est la reconstruction de la partition.

Bello et Sandler [Bel00] ont proposé un système basé sur celui de Martin. La principale amélioration réside dans l'ajout d'une source de connaissance permettant la reconnaissance des accords. Elle consiste en un réseau de neurones dont les paramètres ont été appris à partir de plusieurs enregistrements de piano jouant un accord.

Dans la méthode de Godsmark et Brown [God99], une première analyse du signal se fait par une banque de filtres gammatone. A la sortie des filtres, des groupes de fréquences instantanées sont constitués en rassemblant les canaux adjacents excités par la même composante spectrale. Puis des chaînes synchronisées (*synchrony strands*) sont formées en mettant bout à bout ces fréquences en se basant sur trois principes : continuité temporelle, proximité fréquentielle et cohérence entre amplitudes. Ces chaînes correspondent aux composantes spectrales dominantes. Ce processus constitue le premier niveau du tableau et est largement basé sur la méthode proposée par Cooke [Coo91]. Dans un deuxième niveau, des caractéristiques sont extraites de ces chaînes afin de regrouper celles qui sont probablement issues de la même source. Ce processus confronte plusieurs sources de connaissance comme la synchronisation temporelle, l'harmonicité ou des mouvements fréquentiels similaires. A partir de ces regroupements, de nouvelles caractéristiques émergent, comme la fréquence fondamentale ou le timbre. Le troisième niveau se base sur elles pour proposer des rapprochements afin de constituer un flux sonore. Enfin, les auteurs suggèrent un dernier niveau permettant un traitement en vue d'obtenir des informations de plus haut niveau comme la récurrence d'une phrase mélodique dans un morceau de musique.

On pourrait aussi citer l'architecture IPUS (*Integrated Processing and Understanding of Signals*) qui offre un cadre général pour gérer les interactions entre les algorithmes de traitement du signal et les processus de compréhension du signal. De plus, la méthode a un fonctionnement itératif qui permet d'adapter les paramètres des différents algorithmes en fonction de l'environnement [Les95]. Ce cadre général a été appliqué à l'analyse de signaux audio [Kla98b] sans néanmoins se focaliser sur l'estimation de fréquences fondamentales. Ellis [Ell96] propose une approche qui inclut des modèles de sources afin d'introduire des prévisions d'évènements sonores dans le processus d'interprétation. Cela permet d'avoir une information complémentaire des caractéristiques acoustiques observées sur le signal.

II.3.2 Regroupement de partiels

Le système de Kashino *et al.* [Kas95] est semblable, dans l'esprit, à ceux présentant une architecture de type « tableau noir ». La méthode est organisée sur trois niveaux hiérarchiques : composantes fréquentielles, notes et accords. Des lignes fréquentielles sont d'abord extraites en se basant sur les pics du spectrogramme[5]. Ces composantes spectrales, qui constituent les données de bas niveau, sont ensuite intégrées dans la structure hiérarchique en utilisant des sources de connaissance. Les informations utilisées sont principalement de trois ordres. Des analyses statistiques sur différents morceaux de musiques permettant de déduire des probabilités d'occurrence d'une note au sein d'un accord, ainsi que des règles de transition entre accords. Des informations basées sur des modèles de timbre et des mémoires de sons permettent d'identifier les sources, ou de résoudre le problème des partiels coïncidents. Enfin, quelques-uns des indices proposés par Bregman, harmonicité ou instants d'attaque, sont utilisés pour séparer les sons. Pour gérer l'intégration de connaissances, Kashino utilise un réseau bayésien de probabilité. Ce type d'approche donne de bons résultats en présence de bruit et permet un compromis entre les croyances *a priori* et l'adéquation aux données.

Le méthode de Sterian [Ste99] commence par l'estimation des composantes fréquentielles au cours du temps, en reliant entre eux les pics de la distribution modale de Pielemeier [Pie96] avec un filtre de Kalman. Il partitionne ensuite cet ensemble de composantes afin de les répartir en notes ou d'en retirer les fausses alarmes. Cela se fait en représentant les règles de regroupement de partiels sous forme de vraisemblances, c'est-à-dire de probabilités d'avoir l'ensemble de composantes détecté étant donnée une certaine note. Des lois *a priori* sont aussi définies en se basant sur la probabilité d'occurrence d'une note (un peu comme dans la méthode de Kashino). La recherche de la partition optimale se fait donc dans un cadre bayésien en calculant le maximum *a posteriori*. Un algorithme de suivi multi-hypothèse (*Multiple Hypothesis Tracking*), comparable à ceux utilisés dans le domaine du suivi de cibles par radar, est utilisé afin de gérer l'apparition d'une nouvelle composante, ou pour retirer les partitions ayant une faible probabilité.

Dans son travail, Mellinger [Mel91] s'est intéressé aux différents signes ou indices pouvant être utilisés pour regrouper les partiels par source : attaques simultanées, harmonicité, variations fréquentielles cohérentes, localisation spatiale commune, *etc.* Le point de départ de son système est une analyse fréquentielle à base de corrélogramme. A partir de ce type de représentation, des filtres extraient les différents indices. Il s'est particulièrement focalisé sur les instants d'attaque

[5] L'article [Kas95] n'est pas très clair sur ce point.

et les variations fréquentielles. Le modèle utilise ces indices pour regrouper les caractéristiques locales en évènements sonores, puis les évènements sonores au cours du temps en sources sonores. Il a appliqué ce système à des signaux musicaux composés de plusieurs notes.

L'approche de Brown et Cooke [Bro94] est tout à fait comparable. Le traitement initial du signal en entrée du système correspond à celui effectué dans la méthode de Meddis et Hewitt [Med91]. A partir de la représentation multi-canal obtenue, les auteurs calculent plusieurs cartes de caractéristiques. La carte des variations de fréquence sert à l'élaboration de lignes fréquentielles, en faisant des interpolations linéaires sur les pics. Des hypothèses de fréquences fondamentales sont proposées à partir de la carte des auto-corrélations (similaire au corrélogramme). Un algorithme de programmation dynamique permet de relier chaque ligne fréquentielle à une hypothèse de fréquence fondamentale. Enfin, les lignes reliées à une même hypothèse sont regroupées pour former une note. Il est aussi proposé de compléter ce processus de regroupement par une information de timbre.

II.4 Méthodes utilisant un modèle génératif

Le problème d'estimation de fréquences fondamentales peut être entièrement reformulé en considérant un modèle génératif du signal. Génératif signifie qu'un signal semblable à l'original, voire le signal original lui-même, peut être généré à partir du modèle. Le problème devient donc la construction d'un tel modèle et l'estimation de ses paramètres. Dans le cas de l'estimation de fréquences fondamentales de sources quasi harmoniques, un modèle classique est la décomposition entre une partie déterministe et une autre stochastique [Ser90]. La fraction déterministe est souvent modélisée comme une somme de composantes spectrales de la forme

$$\begin{aligned} s(t) &= A(t)\cos(\phi(t)+\varphi) && \text{(II.9)} \\ &= a(t)\cos(\phi(t)) + b(t)\sin(\phi(t)) && \text{(II.10)} \end{aligned}$$

avec

$$f(t) = \frac{1}{2\pi}\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} \quad , \quad a(t) = A(t)\cos(\varphi) \quad \text{et} \quad b(t) = -A(t)\sin(\varphi) \qquad \text{(II.11)}$$

où $f(t)$ est la fréquence instantanée, $A(t)$ l'amplitude instantanée et φ la phase initiale de cette composante. En fonction du problème traité et de la connaissance des données, cette somme de composantes peut être structurée en K sources avec chacune H_k partiels, certaines méthodes prenant en compte l'estimation de ces deux quantités. Dans ce cas, la fréquence de la $h^{\text{ème}}$ composante de la $k^{\text{ème}}$ source peut être estimée en relation à la fréquence fondamentale $f_{k,1}$ de la source par $f_{k,h} = hf_{k,1}$ ou en considérant une possible inharmonicité (voir section I.1, page 6). La différence entre le signal et la partie déterministe est appelée le résidu, et est souvent modélisée par un processus stochastique à large bande. Cette fraction du modèle correspond à tout ce qui n'est pas périodique dans le signal étudié ou, plus précisément, à tout ce qui n'est pas à bande étroite, ainsi, les évolutions lentes au cours du temps de l'amplitude et de la phase sont prises en compte dans la partie déterministe [Goo96]. Le modèle global du signal devient donc

$$x(t) = \underbrace{\sum_{k=1}^{K}\sum_{h=1}^{H_k} s_{k,h}(t)}_{\text{partie déterministe}} + \underbrace{v(t)}_{\text{partie stochastique}} \qquad \text{(II.12)}$$

L'estimation des paramètres qui permettent au modèle d'expliquer au mieux les données, se fait classiquement en construisant une fonction de vraisemblance et en cherchant son maximum. Cependant, comme dans les méthodes présentées précédemment, le processus d'estimation peut être rendu plus robuste en incluant des informations *a priori*. L'approche paramétrique permet une combinaison optimale de l'information issue des données et de ces connaissances antérieures au signal, par le formalisme bayésien (ce formalisme sera étudié plus en détail dans le chapitre III). L'utilisation, pour l'estimation de fréquences fondamentales, de modèles paramétriques en général et du cadre bayésien en particulier, est assez récente. Cette partie propose une revue de quelques méthodes utilisant la maximisation de la fonction de vraisemblance, puis une présentation des principales approches bayésiennes qui définissent l'état de l'art.

II.4.1 Méthodes par maximum de vraisemblance

La méthode de Doval et Rodet [Dov91] permet l'estimation de la fréquence fondamentale d'un signal musical pseudo périodique. La première étape consiste à calculer le module de la transformée de Fourier court terme et d'en extraire les maxima. L'ensemble des composantes obtenu est une approximation de l'ensemble des partiels du signal. L'idée est alors de trouver la fréquence fondamentale dont la structure harmonique explique le mieux cet ensemble de partiels. Cela se fait en construisant une vraisemblance, c'est-à-dire la fonction de probabilité d'avoir l'ensemble de composantes étant donnée une fréquence fondamentale. L'estimation est obtenue en maximisant cette vraisemblance par rapport à la fréquence fondamentale. Afin de ne pas parcourir l'espace des fréquences en entier, à la recherche de la solution, les auteurs proposent une estimation en deux temps : la recherche de l'intervalle contenant la solution puis l'estimation précise de cette solution.

Yeh *et al.* [Yeh05] ont étendu cette approche au cas polyphonique en considérant connu le nombre de fréquences fondamentales. Le fondement est le même que la méthode précédente : les pics dans le domaine spectral correspondent soit à un partiel d'une (ou plusieurs) source soit à du bruit. A partir du spectre des données observées et d'un modèle spectral générateur quasi harmonique, la méthode génère une liste de candidats et affecte à chacune des fréquences fondamentales, une séquence de partiels potentielle, en prenant en compte une faible inharmonicité. Ensuite, chaque séquence est évaluée par une fonction score construite sur trois principes : harmonicité, lissage spectral et une évolution synchronisée des amplitudes au sein d'une même source. Cette fonction score est la somme pondérée de quatre critères formulant ces principes : harmonicité, largeur de bande moyenne, longueur effective de la séquence et l'écart-type de la durée moyenne de chaque composante de la séquence.

La méthode d'Irizarry [Iri98, Iri01] se base sur une analyse par fenêtre glissante. A chaque instant t, une petite portion du signal, centrée sur t, est extraite. Sur ce segment de signal supposé localement stationnaire est collé un modèle strictement harmonique. Ses paramètres (fréquence fondamentale et amplitudes des harmoniques) sont estimés en minimisant un critère des moindres carrés pondérés par la fenêtre utilisée. Le nombre de note au cours du temps est connu et fixé à un. Les valeurs estimées au cours du temps sont mises bout à bout afin d'obtenir une représentation paramétrique du signal.

II.4.2 Méthodes bayésiennes

Goto [Got01, Got04] a développé une méthode permettant de détecter la mélodie et la voix grave dans des signaux musicaux réels (issus de CDs). Le modèle qu'il propose ne porte pas sur le signal temporel mais sur son spectre à court terme. Un premier traitement permet d'extraire les composantes spectrales et de les séparer en deux groupes : l'un correspondant aux basses fréquences pour l'estimation de la voix grave et l'autre, correspondant aux fréquences intermédiaires et aiguës pour estimer la mélodie. On peut noter que les fréquences fondamentales recherchées n'appartiennent pas forcément à ces mêmes régions. Ensuite, une densité de probabilité de la fréquence fondamentale est construite. Elle représente l'importance relative de toutes les structures harmoniques possibles. Pour la former, l'algorithme regarde chaque groupe de composantes comme un mélange pondéré de tous les modèles de structure harmonique possibles et estime leur poids. Chaque modèle consiste en un nombre fixé de partiels qui sont représentés par des distributions normales centrées sur des multiples entiers de la fréquence fondamentale. Le modèle ayant le poids le plus élevé correspond à la fréquence fondamentale prédominante. L'estimation est effectuée par maximum *a posteriori* et avec l'algorithme EM (*Expectation-Maximization*). Enfin, comme cette valeur estimée n'est pas stable au cours du temps, une étape de suivi séquentiel avec une architecture multi-agent est ajoutée. L'estimation finale est donnée par la trajectoire temporelle la plus prégnante et la plus stable.

Cemgil *et al.* [Cem03, Cem06] ont développé un modèle génératif pour la transcription musicale. Le signal est modélisé à partir d'une représentation similaire aux rouleaux à musique des pianos mécaniques (*piano-roll*). L'idée est de considérer que les fréquences des notes appartiennent à une grille discrète de fréquences. Chaque fréquence de cette grille peut être soit active (1) soit inactive (0). Une variable indicatrice pour chaque fréquence donne son état en fonction du temps. Une attaque, c'est-à-dire le début d'une note, est caractérisée par le passage de l'état 0 à l'état 1. A ce moment là, une onde acoustique est générée avec un certain profil d'amortissement de l'amplitude qui reste valable tant que la fréquence est active. Lors du passage de 1 à 0, le profil change afin de s'annuler plus vite. Le signal total est modélisé par la somme des signaux correspondant à chaque fréquence et d'un bruit gaussien, ce qui permet d'avoir une fonction de vraisemblance au cours du temps. Des distributions séquentielles *a priori* sont obtenues en considérant des modèles d'évolution markoviens sur les variables indicatrices et sur les paramètres du modèle sinusoïdal permettant la génération du signal. Les auteurs ont donc construit un modèle bayésien permettant l'estimation des paramètres à partir de la distribution séquentielle *a posteriori*, en cherchant son maximum. Cependant, cette maximisation est la plupart du temps impossible à faire et les auteurs proposent des méthodes approximatives pour le cas polyphonique en se basant sur le cas monophonique qui est plus facilement traité.

Davy *et al.* [Dav02b, Dav06] proposent une approche hors-ligne pour analyser des segments audio dans lesquels il n'y a pas de changement de notes. La méthode permet l'estimation des paramètres caractéristiques d'un nombre inconnu (mais fixe) de sources, en s'appuyant sur un modèle harmonique polyphonique à base d'atomes de Gabor. En plus du nombre de sources, l'algorithme donne, pour chacune d'entre elles, la fréquence fondamentale, la structure fréquentielle (nombre de partiels pour chaque note et fréquence des partiels) qui peut éventuellement être inharmonique et les amplitudes de chacun des partiels. A partir de ce modèle, une fonction de vraisemblance est construite en supposant que les données sont égales à la somme du modèle déterministe et d'un bruit blanc gaussien de variance inconnue, également estimée par l'algo-

rithme. Afin d'effectuer l'estimation dans un cadre entièrement bayésien, des distributions *a priori* sur les paramètres sont définies. Enfin, le calcul des estimations des paramètres se basant sur la distribution *a posteriori* construite et impliquant des intégrales qui ne peuvent pas se faire analytiquement, les auteurs proposent d'avoir recours aux méthodes d'intégration numérique de Monte Carlo, parfaitement adaptées dans le cadre choisi. Le contexte étant hors-ligne et les données étant toutes disponibles au moment du traitement, l'algorithme d'estimation choisi est celui de MCMC (*Markov Chain Monte Carlo*), dont une version rapide est proposée. Dans les résultats donnés, différents cas de polyphonie ($K = 2, 3$ ou 4) impliquant différents instruments (flûte, clarinette, violon, trompette, *etc.*) sont traités, le nombre de notes jouées étant donné et non estimé.

Vincent et Plumbley [Vin05] ont adopté une approche bayésienne pour estimer le nombre d'*objets harmoniques* (*pitched objects*) ainsi que leurs paramètres caractéristiques. A partir d'une vraisemblance et d'*a priori* locaux, ils construisent une densité *a posteriori* et calculent des estimations par maximum *a posteriori* des paramètres inconnus en utilisant une méthode itérative à sauts déterministes. Ces estimations sont effectuées localement sur chaque fragment de signal puis affinées en prenant en compte des *a priori* de durée et de continuité.

Le but de l'approche de Raphael [Rap02] est d'affecter un label à un vecteur de caractéristiques issu d'un segment (ou *frame*) de signal, avec une application à des morceaux de piano. L'information la plus importante portée par les labels est l'accord joué (au sens d'une combinaison de notes issues d'un ensemble de notes possibles). Les labels contiennent aussi une information permettant de distinguer les moments d'attaque, de soutien et de relâchement de l'accord. L'auteur propose un modèle de Markov caché dont la sortie est le vecteur de caractéristiques observé et dont les variables cachées sont les labels. L'avantage de cette approche est de décrire l'évolution au cours du temps de la séquence de labels et de relier les observations courantes au label correspondant, d'une manière probabiliste. Après une étape permettant d'apprendre les différentes distributions impliquées dans le modèle, la séquence de labels maximisant la distribution *a posteriori* est recherchée. La taille de l'espace des labels étant beaucoup trop grande, l'utilisation directe de l'algorithme de Viterbi n'est pas possible. L'auteur propose alors une méthode permettant de limiter la taille de l'espace en cherchant, pour chaque segment, l'ensemble des labels les plus probables.

II.5 Discussion

Avant de tirer quelques conclusions de cette revue des méthodes d'estimation de fréquences fondamentales, il est important de noter que la catégorisation utilisée dans les parties précédentes n'est pas exclusive et que d'autres algorithmes auraient pu être présentés. En particulier, un pan non négligeable est celui utilisant des représentations issues de méthodes adaptatives [Jai05]. Dans ce contexte, il n'y a pas de modèle paramétrique ni de connaissance spécifique sur les sources, le but étant de les séparer en se basant uniquement sur les données (*data-driven methods*). Les principaux algorithmes utilisent l'analyse en composantes indépendantes et non négatives [Plu03], le *Matching Pursuit* [Mal93, Krs05], le codage parcimonieux (*sparse coding*) [Vir03, Abd06], *etc.* Cependant, les méthodes présentées dans ce chapitre représentent les principaux points de vue adoptés ces dernières années pour résoudre le problème d'estimation que nous étudions.

Nous avons vu au début de ce chapitre que c'est le contexte polyphonique qui nous intéresse plus spécifiquement. Si des approches aussi différentes ont été proposées pour répondre à la question de l'estimation des fréquences fondamentales dans un cadre polyphonique, c'est qu'elle soulève beaucoup de difficultés auxquelles chaque méthode apporte une solution plus ou moins efficace. Une liste des principaux obstacles à surmonter peut ici être dressée :

- Le mélange des sources. Les sources considérées présentent un spectre de raies, c'est-à-dire qu'elles contiennent plusieurs composantes. Dans le signal traité, ces composantes sont mélangées aussi bien dans le domaine temporel (présentes simultanément) que dans le domaine spectral (la structure fréquentielle de chaque source peut s'étendre sur un large intervalle et ces intervalles se superposent). Le problème est donc de pouvoir attribuer une source à chaque composante, autrement dit, de regrouper les composantes issues d'une même source.
- La pseudo périodicité des sources. Elle peut prendre deux aspects. Le premier est l'inharmonicité qui empêche de rechercher les composantes des sources à des positions fréquentielles prédéfinies. Le second est la non stationarité de l'amplitude des composantes. Il est reconnu que, d'une manière générale, l'amplitude des partiels de haut rang décroît plus vite que celle des partiels de bas rang.
- La résistance au bruit. La présence de sources non périodiques, modélisées par une partie stochastique dans le signal, ne doit pas trop perturber l'estimation des paramètres.
- Les partiels coïncidents. Il peut arriver que des sources aient des composantes situées aux mêmes fréquences et qui interfèrent (pouvant même donner l'illusion d'absence de contenu spectral à cette fréquence). Il est donc insuffisant de simplement répartir les partiels entre les sources.
- Stabilité au cours du temps. Les estimations effectuées à un instant donné doivent être cohérentes avec celles effectuées aux instants précédents ou suivants.

A cette liste, il convient d'ajouter le problème inhérent au cadre polyphonique qu'est l'estimation du nombre de sources.

Pour éviter ces difficultés, des hypothèses simplificatrices peuvent être posées : limiter la polyphonie, interdire la présence de sons interférant, considérer un type de source particulier (par exemple en se focalisant sur un instrument donné, dans le cas de la transcription automatique de la musique), faire une segmentation du signal en zones stables et traiter ces zones séparement, *etc.* L'utilisation d'informations *a priori* permet de résoudre certains problèmes : utiliser un modèle de timbre pour gérer les partiels coïncidents, considérer différents principes, comme ceux proposés par Bregman, pour associer les partiels, *etc.*

Au delà de ces considérations, il apparaît que le problème d'estimation de fréquences fondamentales est un problème difficile. Il n'existe pas de méthode suffisamment générale permettant de le résoudre sans poser trop d'hypothèses au préalable et s'appliquant à un large panel de signaux. La méthode proposée dans ce manuscrit peut être appliquée dans des situations dans lesquelles les difficultés citées ci-dessus peuvent être présentes. Le cadre choisi est le formalisme bayésien et l'approche est paramétrique. Les méthodes présentées dans ce chapitre mettent en avant la nécessité de combiner des connaissances *a priori* avec des informations issues des données, pour mener à bien l'estimation des fréquences fondamentales. C'est une première raison du choix du formalisme bayésien car cette alliance peut y être faite de manière conjointe et optimale (c'est-à-dire faisant partie intégrante du processus d'estimation). Une deuxième rai-

son est que cette approche offre un cadre rigoureux avec des algorithmes performants et des résultats de convergence établis. Enfin, la méthode proposée peut être combiner avec un bon nombre des méthodes présentées dans ce chapitre, en les incluant dans le formalisme bayésien mis en place.

Chapitre III

Filtrage bayésien

Un scientifique ne peut et ne doit jamais répondre
qu'en fonction d'un savoir donné, des éléments dont il dispose,
en traduisant sa réponse en termes de probabilités.
J. P. Petit

En reprenant la métaphore d'Anderson et Moore [And79], la notion de filtrage contient l'idée du passage d'une barrière, où ce qui nous parvient correspond à ce que nous voulons, tandis que la barrière retient ce qui est indésirable. Dans le domaine du traitement du signal, cela se transpose sous la forme de la question : comment séparer l'information qui nous intéresse de ce qui la contamine et nous empêche de l'observer directement ? L'information recherchée concerne l'état d'un système dynamique dont des observations sont disponibles au cours du temps. Par exemple, considérons un avion en plein vol dont nous souhaiterions connaître l'altitude à partir de l'écho radar qu'il renvoie. Le système est alors l'avion dont l'état recherché, l'altitude, varie au cours du temps et pour lequel des observations, l'écho radar, peuvent être obtenues. Utilisé dans ce sens, le terme filtrage désigne donc le processus d'estimation d'un vecteur $\boldsymbol{\theta}$ décrivant l'état d'un système dynamique, à partir d'observations $\mathbf{y}$ de ce système.

Ici, il est important de préciser la nature des systèmes considérés. Tout d'abord, l'étude se fait à temps discret, c'est-à-dire que le système est non pas décrit par des équations différentielles mais par des équations aux différences. Les observations sont obtenues à des instants précis, un peu comme des photos du système, et l'estimation du vecteur d'état est effectuée à ces instants. Ensuite, le système est supposé bruité. Ce bruit porte sur les entrées du système, par exemple si elles sont inconnues et que seules leurs propriétés statistiques peuvent être estimées. Plus précisément, l'évolution du vecteur d'état est modélisée par un processus de Markov, c'est-à-dire que la distribution de l'état $\boldsymbol{\theta}_t$ à l'instant t, $t \in \mathbb{N}^+$, ne dépend que de l'état à l'instant précédent :

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_{t-1}) = p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) \tag{III.1}$$

Il est complètement déterminé par la connaissance de sa distribution initiale $p(\boldsymbol{\theta}_0)$ et de son équation de transition $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$. Les observations recueillies au cours du temps peuvent ne pas porter directement sur l'état, mais sur une quantité qui lui est reliée par une équation éventuellement non linéaire. D'autre part, le bruit peut aussi corrompre les observations délivrées,

par exemple à cause de l'imperfection des capteurs utilisés ou du canal de transmission du signal. Cela peut se formaliser en supposant que les observations $\mathbf{y}_t$, $t \in \mathbb{N}^{+*}$, sont des variables aléatoires indépendantes conditionnellement aux états $\boldsymbol{\theta}_t$, $t \in \mathbb{N}^{+}$ et que la distribution marginale $p(\mathbf{y}_t|\boldsymbol{\theta}_t)$ est connue [Dou01]. On parle de système de Markov à espace d'état caché et une description résumée peut en être donnée par

$$\begin{cases} p(\boldsymbol{\theta}_0) & \\ p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) & \text{pour } t \geq 1 \\ p(\mathbf{y}_t|\boldsymbol{\theta}_t) & \text{pour } t \geq 1 \end{cases} \tag{III.2}$$

Différentes manières d'utiliser les observations dont on dispose sont possibles et le terme filtrage est aussi utilisé pour distinguer une approche particulière de deux autres qui lui sont proches : le lissage et la prédiction. En effet, l'estimation de l'état du système $\boldsymbol{\theta}_t$, à l'instant t, peut se faire en utilisant toutes les observations $\mathbf{y}_{1:t}$[1] jusqu'à l'instant t, c'est le filtrage. Il peut aussi y avoir un décalage entre l'instant d'analyse t et l'instant d'occurrence de la dernière observation. Si quelques observations futures sont utilisées, $\mathbf{y}_{1:t+T}$, il s'agit du lissage. Au contraire, si l'instant d'analyse est futur par rapport aux observations prises en compte, $\mathbf{y}_{1:t-T}$, on parle de prédiction. Le contexte qui nous intéresse ici est celui où les observations arrivent séquentiellement, et dans lequel l'estimation du vecteur d'état est mise à jour, instant après instant, grâce à la nouvelle information disponible. L'approche la plus immédiate est donc celle par filtrage (au deuxième sens du terme) et c'est donc elle qui sera choisie. Il est cependant important de noter qu'un même algorithme de filtrage peut être utilisé pour résoudre les problèmes de prédiction et de lissage [Fra69].

Ce chapitre fixe le cadre d'étude choisi dans ce manuscrit et présente les principaux algorithmes utilisés, l'application à la détection et à l'estimation des fréquences fondamentales au cours du temps étant l'objet du chapitre IV, page 65. Il est structuré en quatre parties. Dans la première, le formalisme bayésien est présenté et mis en relation avec l'approche par maximum de vraisemblance, afin d'en justifier le choix. Le problème de filtrage est alors reformulé en terme d'estimation de la distribution *a posteriori*. Les différentes approches possibles pour résoudre ce problème peuvent être regroupées en deux catégories. D'abord, celles qui font une approximation gaussienne de la distribution *a posteriori*, leur objectif étant alors d'estimer les deux premiers moments de cette gaussienne. Ces méthodes sont décrites dans la seconde partie de ce chapitre. Puis, les méthodes cherchant à représenter directement la distribution *a posteriori* de manière adéquate pour effectuer l'estimation qui nous intéresse. Elles sont introduites dans la troisième partie, avant de terminer ce chapitre par une synthèse.

III.1 Inférence statistique

Le principal but de la théorie statistique est d'effectuer, à partir d'observations d'un phénomène aléatoire, une inférence sur la distribution de probabilité sous-jacente à ce phénomène, c'est-à-dire que l'on suppose que les observations ont été générées par une distribution inconnue P. Cette description proposée par C. P. Robert [Rob01], met en avant deux choix qui doivent être faits dans un tel processus d'estimation : le modèle statistique et la méthode employée

[1] Pour un vecteur $\mathbf{x}$, on adopte la notation $\mathbf{x}_{a:b} \triangleq \{\mathbf{x}_a, \mathbf{x}_{a+1}, \ldots, \mathbf{x}_b\}$.

pour effectuer l'inférence statistique. En ce qui concerne le modèle, la procédure classique est de considérer un ensemble de distributions possibles

$$\{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\} \tag{III.3}$$

en conjecturant que la distribution $P \equiv P_{\boldsymbol{\theta}_{theo}}$ appartient à cet ensemble. Ici, deux grandes catégories d'approches se dégagent pour faire face à la complexité des observations :

- les méthodes paramétriques $\Rightarrow \Theta$ est un espace vectoriel de dimension finie
- les méthodes non paramétriques $\Rightarrow \Theta$ est un espace fonctionnel

Pour ces dernières, l'idée est de réduire le moins possible la complexité prise en compte et d'estimer la distribution en posant le minimum d'hypothèses. L'estimation se fait alors souvent par des méthodes de régression [Bir97]. A l'opposé, l'approche paramétrique représente la distribution recherchée sous la forme d'une densité $p(\mathbf{y}|\boldsymbol{\theta})$ où seul le vecteur de paramètres $\boldsymbol{\theta}$ est inconnu. C'est cette approche qui a été choisie dans ce manuscrit[2]. Une fois le modèle statistique choisi, les deux principales méthodes couramment utilisées pour effectuer l'inférence sur le vecteur de paramètres $\boldsymbol{\theta}$ (ou sur toutes estimations s'y rapportant) sont les méthodes par maximum de vraisemblance et les méthodes bayésiennes, décrites dans les deux sections suivantes.

III.1.1 Méthodes par maximum de vraisemblance

C'est entre 1912 et 1922 que Fisher [Ald97, Edw97] a ouvert la voie de l'estimation par maximum de vraisemblance (*Maximum Likelihood*, ML). Depuis, cette méthode statistique s'est répandue pour faire de l'inférence sur les paramètres de la distribution de probabilité sous-jacente à la génération de l'ensemble des observations. Etant donnée la famille de distributions de probabilité de l'équation (III.3), associée à une famille de densités de probabilité[3] $p_{\boldsymbol{\theta}}$, paramétrées par le vecteur $\boldsymbol{\theta} \in \Theta$ (avec, par exemple, $\Theta = \mathbb{R}^p$), il est possible de générer des échantillons $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t$ et de calculer la densité de probabilité associée à cet ensemble de données, appelée vraisemblance :

$$p_{\boldsymbol{\theta}}(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t|\boldsymbol{\theta}) \tag{III.4}$$

Soit L la fonction de $\boldsymbol{\theta}$, les échantillons $\mathbf{y}_{1:t}$ étant considérés fixes, définie par :

$$L_{\mathbf{y}_{1:t}}(\boldsymbol{\theta}) = p_{\boldsymbol{\theta}}(\mathbf{y}_{1:t}|\boldsymbol{\theta}) \tag{III.5}$$

La manipulation de la fonction de vraisemblance, au lieu de la vraisemblance elle-même, sert principalement à mettre en avant le fait que l'inférence porte sur les paramètres inconnus $\boldsymbol{\theta}$ et que dans ce processus d'estimation, les observations sont considérées fixes. La valeur de $\boldsymbol{\theta}$, notée $\hat{\boldsymbol{\theta}}_{ML}$, qui maximise la fonction de vraisemblance $L_{\mathbf{y}_{1:t}}(\boldsymbol{\theta})$ est appelée estimateur par maximum de vraisemblance :

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\ L_{\mathbf{y}_{1:t}}(\boldsymbol{\theta}) \tag{III.6}$$

[2]Les deux approches paramétrique et non paramétrique ont leurs avantages et leurs inconvénients et l'utilisation de l'une ou l'autre est un choix *a priori*. Quelques arguments en faveur du cas paramétrique peuvent être trouvés dans [Rob01].

[3]Dans ce manuscrit, nous utiliserons le terme densité de probabilité, qu'il s'agisse de variables aléatoires discrètes ou continues, en notant que nous faisons ici un abus de langage dans le cas discret.

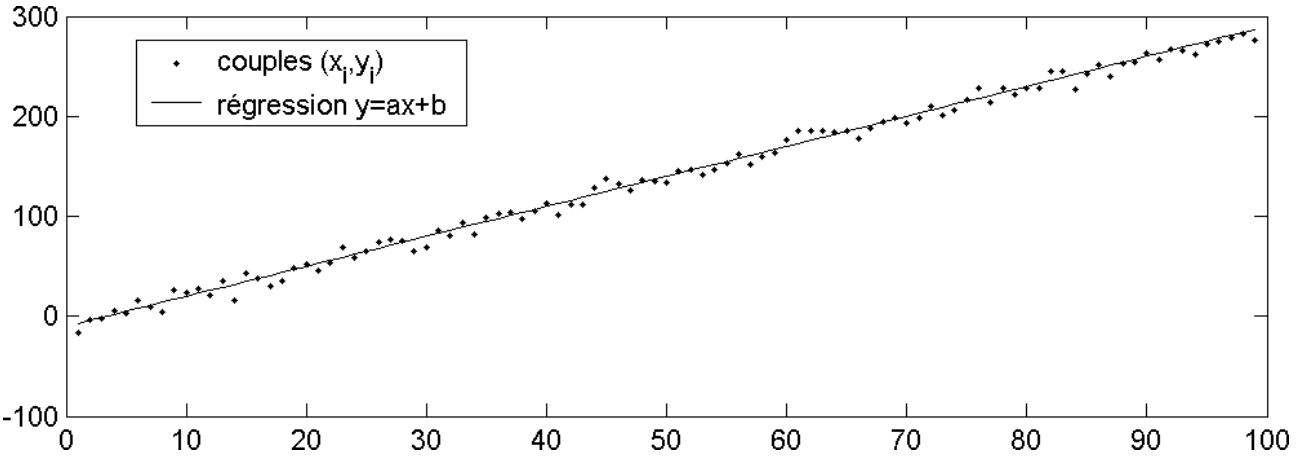


FIG. III.1 – Exemple d'estimation par maximum de vraisemblance : régression linéaire par les moindres carrés.

Chercher à maximiser la fonction de vraisemblance est une approche assez naturelle, car la vraisemblance peut être assimilée à une mesure de similarité entre les observations et le modèle. Autrement dit, c'est le terme qui évalue l'adéquation de ce dernier aux observations. Comme les observations sont pratiquement toujours supposées i.i.d. c'est-à-dire indépendantes et identiquement distribuées (générées par la même distribution), la densité (III.4) peut s'écrire

$$p_{\boldsymbol{\theta}}(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t|\boldsymbol{\theta}) = \prod_{\tau=1}^{t} p_{\boldsymbol{\theta}}(\mathbf{y}_\tau|\boldsymbol{\theta}) \tag{III.7}$$

et le problème de maximisation (III.6) se résout classiquement en considérant le logarithme de la fonction de vraisemblance

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{ML} &= \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\ L_{\mathbf{y}_{1:t}}(\boldsymbol{\theta}) && \text{(III.8)} \\ &= \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\ \log\left(L_{\mathbf{y}_{1:t}}(\boldsymbol{\theta})\right) && \text{(III.9)} \\ &= \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\ \sum_{\tau=1}^{t} \log\left(p_{\boldsymbol{\theta}}(\mathbf{y}_\tau|\boldsymbol{\theta})\right) && \text{(III.10)} \end{aligned}$$

Un exemple très répandue (bien que souvent insoupçonné) de l'utilisation du maximum de vraisemblance est l'estimation par les moindres carrés. En effet, considérons le cas simple de la régression linéaire sur des couples abscisses/ordonnées $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, voir figure III.1. Les paramètres à estimer sont le coefficient directeur de la droite a et l'ordonnée à l'origine b. Les moindres carrés cherchent les valeurs $\hat{a}$ et $\hat{b}$ qui minimisent la distance entre les données et la droite d'équation $y = ax + b$:

$$(\hat{a}, \hat{b}) = \underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - ax_i - b)^2 \tag{III.11}$$

Si maintenant on pose un modèle sur les données du type, pour $i = 1, \ldots, n$:

$$y_i = ax_i + b + \varepsilon_i \tag{III.12}$$

avec, par exemple, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ et les ε_i indépendants, alors le logarithme de la fonction de vraisemblance est proportionnel à :

$$\log(\sigma^{-n}) - \sum_{i=1}^{n} \frac{(y_i - ax_i - b)^2}{2\sigma^2} \tag{III.13}$$

et il devient évident que le couple $(\hat{a}, \hat{b})$ qui maximise l'équation (III.13) est égal à celui définit par l'équation (III.11), ceci quelque soit la valeur de σ^2 (plus de détails sur cet exemple particulier et sur d'autres sont donnés dans [Rob99]).

III.1.2 Méthodes bayésiennes

Le problème que cherche à résoudre l'analyse statistique est principalement un problème inverse, dans la mesure où l'objectif est de retrouver les causes (résumées dans le vecteur $\boldsymbol{\theta}$) à partir des conséquences (matérialisées par les observations $\mathbf{y}$). Ce point de vue est assez facile à retrouver dans la vraisemblance car le lien de cause à effet qui existe entre les paramètres et les observations y est explicité. Une description plus générale [Rob01] est cependant donnée par le théorème de Bayes. Considérons deux vecteurs de variables aléatoires $\boldsymbol{\theta} \in \Theta$ et $\mathbf{y}$, la distribution[4] conditionnelle de $\boldsymbol{\theta}$ étant donné $\mathbf{y}$ est

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_\Theta p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \tag{III.14}$$

Dans cette expression, on retrouve naturellement la vraisemblance $p(\mathbf{y}|\boldsymbol{\theta})$ mais elle est complétée par une distribution $p(\boldsymbol{\theta})$, appelée distribution *a priori*, qui vient modéliser aussi bien les connaissances disponibles (ou le manque de connaissances) que l'incertitude sur le vecteur de paramètres $\boldsymbol{\theta}$. L'inférence est alors effectuée sur la nouvelle distribution $p(\boldsymbol{\theta}|\mathbf{y})$ ainsi construite, appelée la distribution *a posteriori*. L'inclusion, dans le processus d'estimation, de connaissances sur les paramètres, peut aussi être faite dans les méthodes basées sur la vraisemblance, avec notament des solutions comme les approches par vraisemblance pénalisée [Gre99]. Cependant, sous certaines conditions, un lien formel entre ces approches et le contexte bayésien peut être établie. Le théorème de Bayes peut être résumé par

$$a\ posteriori \propto \text{vraisemblance} \times a\ priori \tag{III.15}$$

et signifie que, d'après Steve Gull :

> *ce que nous savons sur le vecteur de paramètres $\boldsymbol{\theta}$ après avoir obtenu les observations $\mathbf{y}$ est la combinaison de ce que nous savions sur $\boldsymbol{\theta}$ auparavant avec ce que les observations nous apprennent sur $\boldsymbol{\theta}$.*

[4] Les termes *distribution* et *densité* seront souvent confondus car nous nous situons dans un cadre dans lequel le contexte permet d'identifier clairement la quantité utilisée.

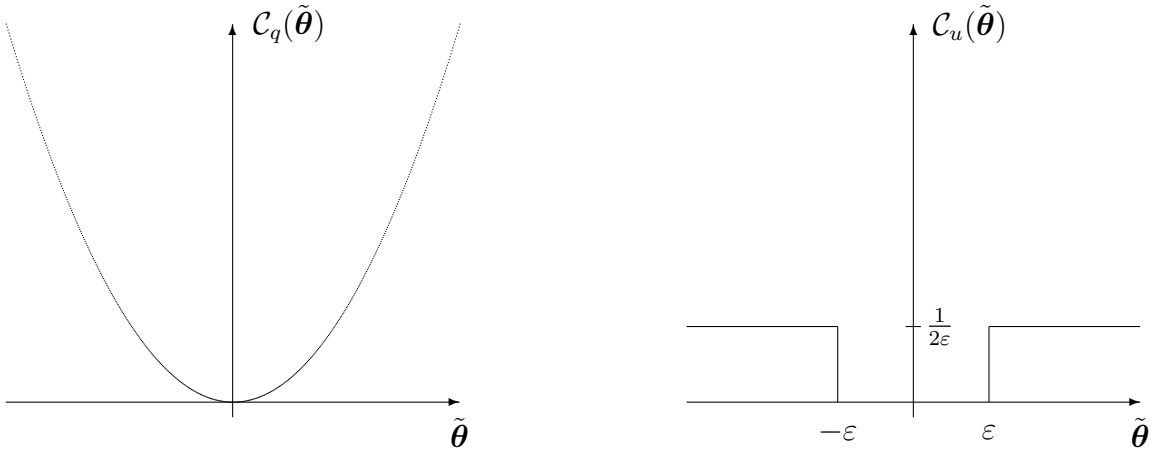


FIG. III.2 – Exemples de fonctions de coût : quadratique (à gauche) et uniforme (à droite).

Dans l'équation (III.14), le dénominateur, souvent appelé l'évidence, ne dépend pas de $\boldsymbol{\theta}$ et reste constant lorsque l'on fait de l'inférence sur $\boldsymbol{\theta}$, à partir de la distribution *a posteriori*. Ce qui explique le signe $\propto$ (« proportionnel à ») dans la formule (III.15).

L'inférence sur la distribution *a posteriori* peut se faire de différentes manières et plusieurs estimateurs peuvent être construits. Posons quelques notations. La valeur estimée du vecteur recherché $\boldsymbol{\theta}$ sera notée $\hat{\boldsymbol{\theta}}$ et l'erreur d'estimation commise avec $\hat{\boldsymbol{\theta}}$ sera notée $\tilde{\boldsymbol{\theta}}$ et est définie par $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$. Dans la théorie bayésienne de l'estimation, on introduit une fonction $\mathcal{R}$ qui mesure le risque pris en estimant $\boldsymbol{\theta}$ par $\hat{\boldsymbol{\theta}}$. Elle est définie par [Tol99] :

$$\mathcal{R}(\hat{\boldsymbol{\theta}}) = \int\int \mathcal{C}(\tilde{\boldsymbol{\theta}}) p(\boldsymbol{\theta}, \mathbf{y}) \mathrm{d}\mathbf{y} \mathrm{d}\boldsymbol{\theta} \tag{III.16}$$

où $p(\boldsymbol{\theta}, \mathbf{y})$ est la densité jointe de $\boldsymbol{\theta}$ et de $\mathbf{y}$. La fonction $\mathcal{C}$ est le critère d'évaluation du risque et est souvent appelée fonction de coût. Le problème de l'estimation du vecteur $\boldsymbol{\theta}$ se transforme en celui de la minimisation du risque $\mathcal{R}$, étant donnée une certaine fonction de coût $\mathcal{C}$. La valeur estimée minimisant le risque est dite optimale. Deux exemples courants de fonction de coût, voir figure III.2, sont la fonction de coût quadratique :

$$\mathcal{C}_q(\tilde{\boldsymbol{\theta}}) = ||\tilde{\boldsymbol{\theta}}||^2 \tag{III.17}$$

où la norme utilisée peut être du type Mahalanobis, qui est plus générale que la simple norme euclidienne ; et la fonction de coût uniforme :

$$\mathcal{C}_u(\tilde{\boldsymbol{\theta}}) = \begin{cases} 0 & ||\tilde{\boldsymbol{\theta}}|| < \varepsilon \\ \frac{1}{2\varepsilon} & ||\tilde{\boldsymbol{\theta}}|| \geq \varepsilon \end{cases} \tag{III.18}$$

Même si, théoriquement, la valeur optimale dépend de la fonction de coût, on peut noter que dans une large majorité des problèmes d'estimation, le choix de la fonction de coût n'a pas d'influence sur le résultat. Les deux estimateurs les plus importants dans l'approche bayésienne

sont l'estimateur par maximum *a posteriori* (MAP) et l'estimateur par minimum de variance, encore appelé estimateur à erreur quadratique moyenne minimum (*Minimum Mean Square Error*, MMSE). Ces deux estimateurs correspondent respectivement à la fonction de coût uniforme et à la fonction de coût quadratique, et peuvent être calculés [Tol99] à partir de la distribution *a posteriori* par :

$$\hat{\boldsymbol{\theta}}_{MAP} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\ p(\boldsymbol{\theta}|\mathbf{y}) \tag{III.19}$$

et

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{MMSE} &= \mathbb{E}_{p(\boldsymbol{\theta}|\mathbf{y})}[\boldsymbol{\theta}] && \text{(III.20)} \\ &= \int_{\Theta} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} && \text{(III.21)} \end{aligned}$$

De même, un estimateur optimal de tout paramètre particulier $\ell(\boldsymbol{\theta})$, ℓ étant une fonction donnée, peut être obtenue en calculant son espérance $\mathbb{E}_{p(\boldsymbol{\theta}|\mathbf{y})}[\ell(\boldsymbol{\theta})]$ par rapport à la distribution *a posteriori*.

III.1.3 Discussion

Dans le chapitre précédent, quelques justifications du choix de l'approche bayésienne avaient été données en invoquant des raisons pratiques liées à l'application qui nous préoccupe dans ce manuscrit. Ici, c'est d'un point de vue plus « philosophique »que sont données quelques justifications de ce choix, en renvoyant cependant le lecteur intéressé à des ouvrages plus complets sur la question [Rob01, Mac03].

L'hypothèse faite, en particulier par les méthodes par maximum de vraisemblance et par la théorie statistique classique (ou fréquentiste) en général, est que les observations ont été échantillonnées par un modèle génératif décrivant le système à l'origine de ces observations. De plus, il est supposé qu'il existe un jeu de paramètres de ce modèle permettant de rendre compte fidèlement du système (*ground truth parameter value*) et vouloir l'estimer est alors fondé. Dans ce cas, le sens des probabilités renvoie à une fréquence de réalisations d'un processus aléatoire, et l'enjeu est de connaître la distribution de probabilité sous-jacente à la génération des observations. Cependant, dans le cas où le modèle utilisé est incorrect ou n'est qu'une approximation du système (ce qui est pratiquement toujours le cas dans la réalité), la recherche de ce jeu de paramètres optimal perd de son sens. Dans le cadre de l'approche bayésienne, les probabilités sont plutôt utilisées pour décrire un degré de croyance en une proposition, qui n'est pas nécessairement reliée à des variables aléatoires (quelle est la probabilité qu'un dé soit truqué ?). Ainsi, l'analyse statistique bayésienne est utilisée pour décrire des hypothèses et les inférences effectuées étant données ces hypothèses. De part cette dépendance aux hypothèses, cette approche est souvent critiquée car qualifiée de subjective. Cependant, est-il possible de faire de l'inférence en ne posant aucune hypothèse ? La réponse à cette question est pratiquement toujours négative et c'est là que réside toute la puissance du paradigme bayésien, car cette incertitude tacite sur les informations dont on dispose est formellement prise en compte dans le processus d'estimation, au travers de la distribution *a priori*.

En revenant au problème de filtrage exposé au début de ce chapitre, le théorème de Bayes stipule que la probabilité qu'un vecteur $\boldsymbol{\theta}$ donné décrive l'état du système considéré, est égale au

produit de la vraisemblance (c'est-à-dire de la probabilité que les observations aient été générées par le système en l'état décrit par $\boldsymbol{\theta}$) par l'*a priori* (c'est-à-dire la probabilité que cet état du système apparaisse). Il s'avère donc être un outils performant pour l'estimation de l'état d'un système. En reprenant l'approche par modèle de Markov à espace d'état caché, nous pourrions récrire les équations du système (III.2) sous la forme

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) \quad \longleftrightarrow \quad \boldsymbol{\theta}_t = g_t(\boldsymbol{\theta}_{t-1}, \mathbf{v}^{\boldsymbol{\theta}}_{t-1}) \tag{III.22}$$

$$p(\mathbf{y}_t|\boldsymbol{\theta}_t) \quad \longleftrightarrow \quad \mathbf{y}_t = h_t(\boldsymbol{\theta}_t, \mathbf{v}^{\mathbf{y}}_{t-1}) \tag{III.23}$$

avec g_t et h_t deux fonctions déterministes connues (éventuellement non linéaires) et $\mathbf{v}^{\boldsymbol{\theta}}_{t-1}$ et $\mathbf{v}^{\mathbf{y}}_{t-1}$ deux bruits blancs centrés et indépendants. L'équation (III.22), dite équation de transition, modélise l'évolution de l'état au cours du temps et détermine l'*a priori* séquentiel $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$. L'équation (III.23), dite équation d'observation, relie l'état courant aux observations obtenues au même instant et détermine la vraisemblance $p(\mathbf{y}_t|\boldsymbol{\theta}_t)$.

Notre objectif est donc d'estimer, de manière séquentielle, la distribution *a posteriori* :

$$p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \tag{III.24}$$

et, en particulier, la densité de filtrage qui lui est associée :

$$p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \int p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})\mathrm{d}\boldsymbol{\theta}_{0:t-1} \tag{III.25}$$

Le reste de ce chapitre est consacré à la présentation de différentes méthodes permettant le calcul récursif de $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ ou de sa marginale $p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t})$. La densité de filtrage, lorsqu'elle est calculée de manière séquentielle, présente un intérêt particulier car il n'est pas nécessaire de garder en mémoire la trajectoire complète de l'état, depuis l'instant initial. De plus, plusieurs types d'estimation de l'état à l'instant courant peuvent être obtenus à partir de la densité de filtrage.

III.2 Filtre de Kalman et approximations

En supposant que les bruits d'état $\mathbf{v}^{\boldsymbol{\theta}}$ et de mesure $\mathbf{v}^{\mathbf{y}}$, dans les équations (III.22) et (III.23), suivent une loi normale et que les fonctions g et h sont linéaires par rapport au vecteur d'état et aux bruits $\mathbf{v}^{\boldsymbol{\theta}}$ et $\mathbf{v}^{\mathbf{y}}$ respectivement, le problème de l'estimation séquentielle de la densité de filtrage[5] $p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t})$ peut être résolu de manière analytique par le filtre de Kalman. L'hypothèse linéaire est néanmoins assez forte et, dans le cas non linéaire, on a souvent recours à un filtre construit sur le filtre de Kalman : le filtre de Kalman étendu. Cependant, si les non linéarités sont significatives, cette solution ne donne pas toujours de très bon résultats. Le filtre de Kalman sans parfum a été introduit récemment et il a été montré qu'il permettait de calculer des estimations plus précises que le filtre de Kalman étendu. Ces trois filtres sont décrits dans cette partie.

[5] Ou plus précisément, de ses deux premiers moments, ce qui revient au même car dans ce cas, la densité de filtrage est une gaussienne.

A l'instant $t = 0$

Initialisation

$$\boldsymbol{\mu}_{0|0} = \boldsymbol{\mu}_0$$
$$\boldsymbol{\Sigma}_{0|0} = \boldsymbol{\Sigma}_0$$

A l'instant $t \geq 1$

Prédiction

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{A}\boldsymbol{\mu}_{t-1|t-1}$$
$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{A}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}}$$

Correction

$$\mathbf{y}_{t|t-1} = \mathbf{C}\boldsymbol{\mu}_{t|t-1}$$
$$\mathbf{S}_{t|t-1} = \mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^{\mathrm{T}} + \mathbf{D}\mathbf{D}^{\mathrm{T}}$$
$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^{\mathrm{T}}\mathbf{S}_{t|t-1}^{-1}(\mathbf{y}_t - \mathbf{y}_{t|t-1})$$
$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^{\mathrm{T}}\mathbf{S}_{t|t-1}^{-1}\mathbf{C}\boldsymbol{\Sigma}_{t|t-1}$$

TAB. III.1 – Algorithme du filtre de Kalman.

III.2.1 Cas linéaire/gaussien

Dans le cas où les équations de transition et d'observation sont linéaires et où les bruits d'état et de mesure sont gaussiens, on peut récrire le système (III.22)-(III.23) sous la forme :

$$\boldsymbol{\theta}_t = \mathbf{A}\boldsymbol{\theta}_{t-1} + \mathbf{B}\mathbf{v}'_{t-1} \quad \text{(III.26)}$$
$$\mathbf{y}_t = \mathbf{C}\boldsymbol{\theta}_t + \mathbf{D}\mathbf{v}''_{t-1} \quad \text{(III.27)}$$

où les matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ et $\mathbf{D}$ sont de dimensions convenables et les bruits $\mathbf{v}'$ et $\mathbf{v}''$ ont une densité normale de moyenne nulle et de variance unité. L'état initial $\boldsymbol{\theta}_0$ est supposé être un vecteur aléatoire gaussien, de moyenne et de matrice de covariance connues :

$$\boldsymbol{\theta}_0 \sim \mathcal{N}(\boldsymbol{\theta}_0; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \quad \text{(III.28)}$$

Enfin, il est fait l'hypothèse que $\boldsymbol{\theta}_0$ est indépendant des bruits $\mathbf{v}'$ et $\mathbf{v}''$. Pour faire le lien avec les bruits d'état et de mesure introduits précédemment, on peut noter que :

$$\mathbf{v}^{\boldsymbol{\theta}} \sim \mathcal{N}(\mathbf{0}, \mathbf{B}\mathbf{B}^{\mathrm{T}}) \quad \text{(III.29)}$$
$$\mathbf{v}^{\mathbf{y}} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}\mathbf{D}^{\mathrm{T}}) \quad \text{(III.30)}$$

Le filtre de Kalman [Kal60] procède en deux étapes : prédiction et correction. Ces deux étapes peuvent être explicitées en écrivant la densité de filtrage à l'instant t, en fonction de celle à l'instant $t-1$:

$$\text{Prédiction} \Rightarrow p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t-1}) = \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})p(\boldsymbol{\theta}_{t-1}|\mathbf{y}_{1:t-1})\mathrm{d}\boldsymbol{\theta}_{t-1} \quad \text{(III.31)}$$

$$\text{Correction} \Rightarrow p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t-1})\mathrm{d}\boldsymbol{\theta}_t} \quad \text{(III.32)}$$

Ainsi, la densité de filtrage peut être mise à jour de manière récursive. De plus, grâce aux hypothèses posées, les deux densités (III.31) et (III.32) sont des densités normales

$$p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}) \quad \text{(III.33)}$$
$$p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \quad \text{(III.34)}$$

dont les deux premiers moments peuvent être calculés analytiquement en fonction des deux premiers moments à l'instant précédent [Bar94] (les calculs sont détaillés dans l'annexe A). L'algorithme complet du filtre de Kalman est donné dans le tableau III.1. Le vecteur $\boldsymbol{\mu}_{t|t-1}$ peut être interprété comme l'estimation de $\boldsymbol{\theta}_t$, avec la variance $\boldsymbol{\Sigma}_{t|t-1}$, étant données toutes les observations jusqu'à l'instant $t-1$. On parle alors de valeur prédite en fonction de l'information dont on dispose. Cette valeur estimée est ensuite corrigée, ou mise à jour, avec l'arrivée des dernières observations à l'instant t. On obtient alors le vecteur $\boldsymbol{\mu}_{t|t}$, avec la variance $\boldsymbol{\Sigma}_{t|t}$. Cette estimation de $\boldsymbol{\theta}_t$ est optimale au sens du minimum de variance comme au sens du maximum *a posteriori* [Che03, Dré03].

III.2.2 Filtre de Kalman étendu

Dans le filtre de Kalman étendu, les fonctions g_t et h_t peuvent être non linéaires et, en reprenant les mêmes notations que dans (III.26)-(III.27), le système considéré est :

$$\boldsymbol{\theta}_t = g_t(\boldsymbol{\theta}_{t-1}) + \mathbf{B}\mathbf{v}'_{t-1} \quad \text{(III.35)}$$
$$\mathbf{y}_t = h_t(\boldsymbol{\theta}_t) + \mathbf{D}\mathbf{v}''_{t-1} \quad \text{(III.36)}$$

Les deux densités $p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t-1})$ et $p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t})$, définies par les équations III.31 et III.32, sont alors approchées par des densités normales :

$$p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t-1}) \approx \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}) \quad \text{(III.37)}$$
$$p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t}) \approx \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \quad \text{(III.38)}$$

dont les deux premiers moments sont mis à jour récursivement en utilisant une linéarisation locale des fonctions g_t et h_t. Plus précisément, elles peuvent être décomposées en série de Taylor au voisinnage des points $\boldsymbol{\mu}_{t-1|t-1}$ et $\boldsymbol{\mu}_{t|t-1}$:

$$g_t(\boldsymbol{\theta}_{t-1}) = g_t(\boldsymbol{\mu}_{t-1|t-1}) + (\boldsymbol{\theta}_{t-1} - \boldsymbol{\mu}_{t-1|t-1}) \left.\frac{\partial g_t(\boldsymbol{\theta}_{t-1})}{\partial \boldsymbol{\theta}_{t-1}}\right|_{\boldsymbol{\theta}_{t-1}=\boldsymbol{\mu}_{t-1|t-1}} + \dots \quad \text{(III.39)}$$

$$h_t(\boldsymbol{\theta}_t) = h_t(\boldsymbol{\mu}_{t|t-1}) + (\boldsymbol{\theta}_t - \boldsymbol{\mu}_{t|t-1}) \left.\frac{\partial h_t(\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}\right|_{\boldsymbol{\theta}_t=\boldsymbol{\mu}_{t|t-1}} + \dots \quad \text{(III.40)}$$

On peut construire les vecteurs $\mathbf{a}_t = g_t(\boldsymbol{\mu}_{t-1|t-1}) - \mathbf{A}_t\boldsymbol{\mu}_{t-1|t-1}$ et $\mathbf{c}_t = h_t(\boldsymbol{\mu}_{t|t-1}) - \mathbf{C}_t\boldsymbol{\mu}_{t|t-1}$, où $\mathbf{A}_t$ et $\mathbf{C}_t$ sont les matrices jacobiennes des fonctions g_t et h_t :

$$\mathbf{A}_t = \left.\frac{\partial g_t(\boldsymbol{\theta}_{t-1})}{\partial \boldsymbol{\theta}_{t-1}}\right|_{\boldsymbol{\theta}_{t-1}=\boldsymbol{\mu}_{t-1|t-1}} \quad \text{et} \quad \mathbf{C}_t = \left.\frac{\partial h_t(\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}\right|_{\boldsymbol{\theta}_t=\boldsymbol{\mu}_{t|t-1}} \quad \text{(III.41)}$$

A l'instant $t = 0$

Initialisation

$\boldsymbol{\mu}_{0|0} = \boldsymbol{\mu}_0$

$\boldsymbol{\Sigma}_{0|0} = \boldsymbol{\Sigma}_0$

A l'instant $t \geq 1$

Prédiction

$\boldsymbol{\mu}_{t|t-1} = g_t(\boldsymbol{\mu}_{t-1|t-1})$

$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{A}_t \boldsymbol{\Sigma}_{t-1|t-1} \mathbf{A}_t^{\mathsf{T}} + \mathbf{B}\mathbf{B}^{\mathsf{T}}$

Correction

$\mathbf{y}_{t|t-1} = h_t(\boldsymbol{\mu}_{t|t-1})$

$\mathbf{S}_{t|t-1} = \mathbf{C}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{C}_t^{\mathsf{T}} + \mathbf{D}\mathbf{D}^{\mathsf{T}}$

$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1} \mathbf{C}_t^{\mathsf{T}} \mathbf{S}_{t|t-1}^{-1} (\mathbf{y}_t - \mathbf{y}_{t|t-1})$

$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1} \mathbf{C}_t^{\mathsf{T}} \mathbf{S}_{t|t-1}^{-1} \mathbf{C}_t \boldsymbol{\Sigma}_{t|t-1}$

TAB. III.2 – Algorithme du filtre de Kalman étendu.

Ainsi, en ne gardant que les termes jusqu'au premier ordre, un modèle linéarisé peut être construit :

$$\boldsymbol{\theta}_t = \mathbf{A}_t \boldsymbol{\theta}_{t-1} + \mathbf{a}_t + \mathbf{B}\mathbf{v}'_{t-1} \quad \text{(III.42)}$$

$$\mathbf{y}_t = \mathbf{C}_t \boldsymbol{\theta}_t + \mathbf{c}_t + \mathbf{D}\mathbf{v}''_{t-1} \quad \text{(III.43)}$$

Les équations du filtre de Kalman étendu sont obtenues en utilisant pour le modèle (III.42)-(III.43), les résultats du filtre de Kalman classique [And79]. L'algorithme est donné dans le tableau III.2.

III.2.3 Filtre de Kalman sans parfum

Le filtre de Kalman sans parfum (*unscented Kalman filter*, aussi appelé filtre de Kalman sans biais) se base sur la transformée sans parfum [Jul97] (voir Annexe B) pour estimer et mettre à jour au cours du temps, la moyenne $\boldsymbol{\mu}_{t|t}$ et la matrice de covariance $\boldsymbol{\Sigma}_{t|t}$ de l'approximation gaussienne de la densité de filtrage. Le principe sous-jacent à cette méthode est qu'il semble plus facile de faire l'approximation d'une gaussienne que d'une fonction non linéaire arbitraire. Elle a donc une approche radicalement opposée à celle du filtre de Kalman étendu qui linéarise les fonctions non linéaires. La figure III.3 compare les performances de ces deux méthodes sur un exemple donné. D'une manière générale, ce filtre donne de meilleurs résultats que le filtre de Kalman étendu car les non linéarités sont mieux approchées (à un ordre plus élevé). Les paramètres de l'approximation gaussienne sont alors plus précis.

L'algorithme complet du filtre de Kalman sans parfum est donné dans le tableau III.3. Le modèle considéré est celui décrit par les équations (III.35)-(III.36). Une nouvelle version du filtre de Kalman sans parfum a été proposée [Mer01] dans laquelle l'algorithme présente une meilleure stabilité numérique et assure le caractère défini positif de la matrice de covariance (ce qui n'est pas nécessairement le cas dans la version standard).

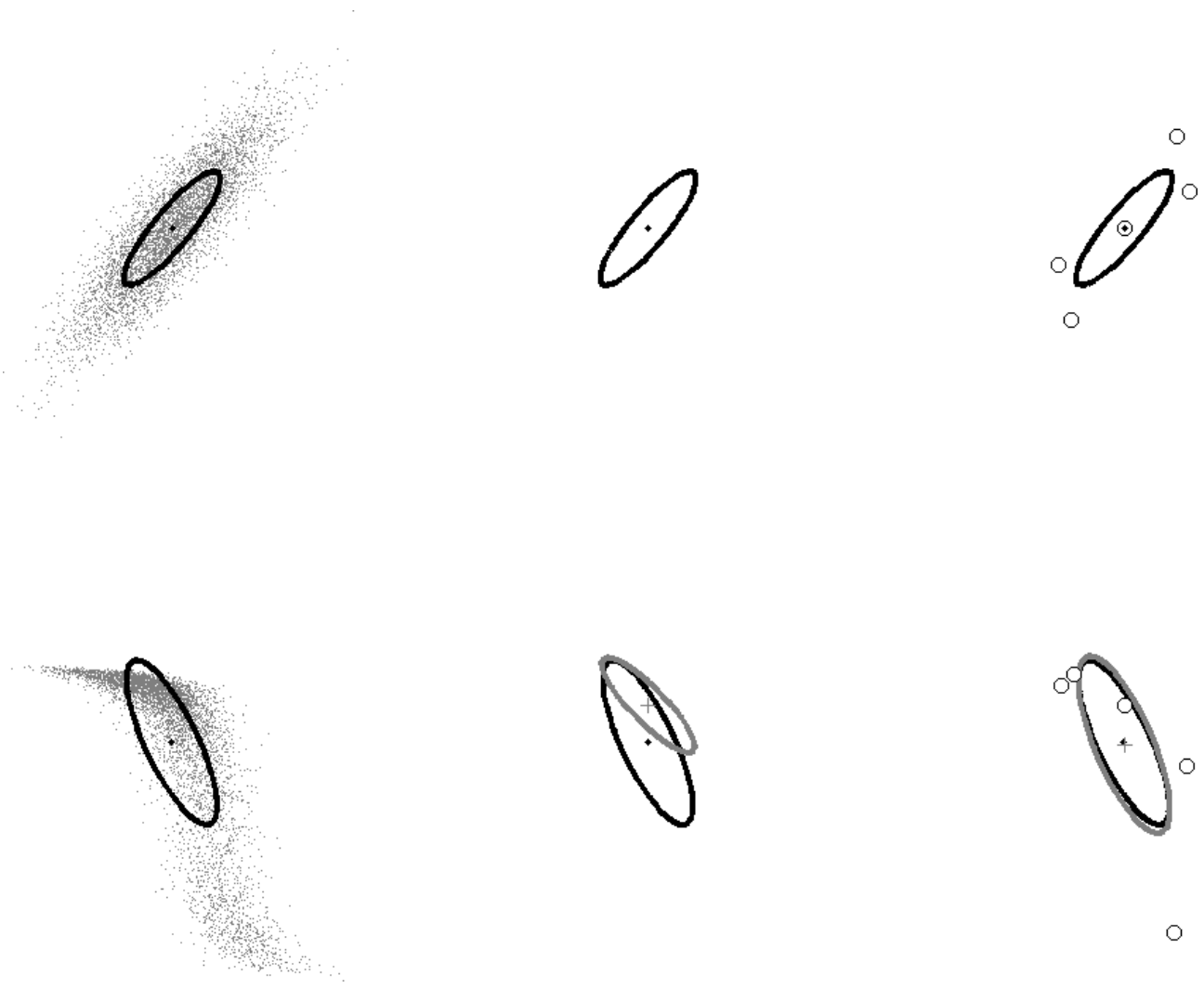


FIG. III.3 – Comparaison des estimations fournies par le filtre de Kalman étendu et la transformée sans parfum. Sur la colonne de gauche, un nuage d'échantillons gaussiens (en haut) est passé au travers d'une fonction non linéaire (en bas). La colonne du milieu montre l'estimation de la moyenne et de la matrice de covariance des échantillons transformés, par le filtre de Kalman étendu. L'estimation par la transformée sans parfum est faite sur la colonne de droite où les ronds correspondent aux *sigma points* (voir annexe B).

A l'instant $t = 0$

Initialisation

$\boldsymbol{\mu}_{0|0} = \boldsymbol{\mu}_0$

$\boldsymbol{\Sigma}_{0|0} = \boldsymbol{\Sigma}_0$

A l'instant $t \geq 1$

Calcul des échantillons

Calcul de $\mathbf{L}_{t-1|t-1}$ tel que $\boldsymbol{\Sigma}_{t-1|t-1} = \mathbf{L}_{t-1|t-1}\mathbf{L}_{t-1|t-1}$

$\mathcal{O}_{t-1|t-1} = \left\{\boldsymbol{\mu}_{t-1|t-1} \quad \boldsymbol{\mu}_{t-1|t-1} \pm \sqrt{n_{\boldsymbol{\theta}} + \kappa}(\mathbf{L}_{t-1|t-1})_i\right\}_{i=1,\ldots,n_{\boldsymbol{\theta}}}$

Prédiction

$\mathcal{O}_{t|t-1} = g_t(\mathcal{O}_{t-1|t-1})$

$\boldsymbol{\mu}_{t|t-1} = \sum_{i=0}^{2n_{\boldsymbol{\theta}}} W_i \mathcal{O}_{i,t|t-1}$

$\boldsymbol{\Sigma}_{t|t-1} = \sum_{i=0}^{2n_{\boldsymbol{\theta}}} W_i (\mathcal{O}_{i,t|t-1} - \boldsymbol{\mu}_{t|t-1})(\mathcal{O}_{i,t|t-1} - \boldsymbol{\mu}_{t|t-1})^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}}$

Correction

$\mathcal{Y}_{t|t-1} = h_t(\mathcal{O}_{t|t-1})$

$\mathbf{y}_{t|t-1} = \sum_{i=0}^{2n_{\boldsymbol{\theta}}} W_i \mathcal{Y}_{i,t|t-1}$

$\mathbf{P}_{\mathbf{yy}} = \sum_{i=0}^{2n_{\boldsymbol{\theta}}} W_i (\mathcal{Y}_{i,t|t-1} - \mathbf{y}_{t|t-1})(\mathcal{Y}_{i,t|t-1} - \mathbf{y}_{t|t-1})^{\mathrm{T}} + \mathbf{D}\mathbf{D}^{\mathrm{T}}$

$\mathbf{P}_{\boldsymbol{\theta}\mathbf{y}} = \sum_{i=0}^{2n_{\boldsymbol{\theta}}} W_i (\mathcal{O}_{i,t|t-1} - \boldsymbol{\mu}_{t|t-1})(\mathcal{Y}_{i,t|t-1} - \mathbf{y}_{t|t-1})^{\mathrm{T}}$

$\mathbf{S}_{t|t-1} = \mathbf{P}_{\boldsymbol{\theta}\mathbf{y}}\mathbf{P}_{\mathbf{yy}}^{-1}$

$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \mathbf{S}_{t|t-1}(\mathbf{y}_t - \mathbf{y}_{t|t-1})$

$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{S}_{t|t-1}\mathbf{P}_{\mathbf{yy}}\mathbf{S}_{t|t-1}^{\mathrm{T}}$

TAB. III.3 – Algorithme du filtre de Kalman sans parfum.

III.2.4 Discussion

Le filtre de Kalman apporte une solution optimale au problème de filtrage bayésien, sous les hypothèses linéaire et gaussienne. Cependant, beaucoup de systèmes réels sont non linéaires et des solutions comme le filtre de Kalman étendu ou sans parfum deviennent nécessaires. Quelques commentaires peuvent être faits sur ces deux extensions au cas non linéaire. D'un point de vue pratique, ils nécessitent le stockage en mémoire de matrices de covariance, ce qui peut devenir problématique lorsque la dimension $n_{\boldsymbol{\theta}}$ de l'état devient trop importante. De même, le filtre de Kalman sans parfum requiert $2n_{\boldsymbol{\theta}} + 1$ évaluations du modèle, ce qui, dans certains cas, peut devenir un facteur limitant. D'un point de vue plus théorique, ces deux filtres font une approximation gaussienne de la densité de filtrage. Or, du fait des non linéarités, cette densité peut présenter un caractère très piqué ou multimodal. Une généralisation possible est de la modéliser par un mélange de gaussiennes :

$$p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \sum_{m=1}^{M} c_m \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}_{t|t}^{(m)}, \boldsymbol{\Sigma}_{t|t}^{(m)}) \tag{III.44}$$

où les coefficients de mélange c_m sont strictement positifs et de somme unité. La structure du filtre ainsi construit [Als72] est alors une banque de filtres de Kalman étendu. Cependant, pour les raisons indiquées précédemment, d'autres schémas d'approximation de la distribution

a posteriori, comme les méthodes de Monte Carlo, deviennent nécessaires lorsque la dimension de l'espace d'état augmente.

III.3 Méthodes de Monte Carlo séquentielles

Le principe des trois filtres présentés dans la partie précédente est de faire une approximation gaussienne de la densité de filtrage. Elle est alors représentée par ses deux premiers moments et leur estimation récursive est basée sur les équations (III.31) et (III.32). Le principe des méthodes de Monte Carlo séquentielles est différent : elles cherchent à faire une approximation de la séquence de distributions *a posteriori* sans faire d'hypothèse sur leur forme. En particulier, les contraintes linéaires et/ou gaussiennes sur le modèle ne sont pas nécessaires à l'application de ces méthodes. Cette approximation séquentielle est basée sur la mise à jour, au cours du temps, d'un ensemble d'échantillons aléatoires pondérés, représentant une valeur possible du vecteur d'état. La représentation de la distribution *a posteriori* ainsi obtenue est spécialement adaptée au calcul d'estimateurs minimisant l'erreur quadratique moyenne. Si les méthodes de Monte Carlo séquentielles tirent leurs origines des recherches menées au cours de la Seconde Guerre Mondiale, c'est au début des années 90 qu'elles ont véritablement pris leur essor avec l'algorithme proposé par Gordon, Salmond et Smith [Gor93].

Cette partie donne une présentation générale des méthodes de Monte Carlo séquentielles, en se basant principalement sur [Dou01, Che03] ainsi que sur [Liu98a, Dou00, Aru02, Dou05]. Elle est structurée de la manière suivante : d'abord, le principe des méthodes de Monte Carlo et leur mise en œuvre pratique sont présentés et discutés. Puis, elles sont adaptées pour être utilisables dans un contexte séquentiel et l'algorithme de filtrage particulaire de base est construit. Enfin, plusieurs améliorations de cet algorithme sont présentées.

III.3.1 Echantillonnage de Monte Carlo

D'une manière générale, « Monte Carlo »fait référence à toute méthode visant à calculer une valeur numérique en utilisant des procédés stochastiques. Ce nom fait allusion aux jeux de hasard pratiqués dans les casinos et plus particulièrement dans celui situé à Monte Carlo, dans la principauté de Monaco.

Considérons un ensemble de N échantillons, aussi appelés particules, indépendants et identiquement distribués selon $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$:

$$\left\{\boldsymbol{\theta}_{0:t}^{(i)}\right\}_{i=1,\ldots,N} \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \tag{III.45}$$

Une approximation empirique de la distribution *a posteriori* est alors donnée par :

$$\hat{p}_N(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\boldsymbol{\theta}_{0:t}^{(i)}}(\boldsymbol{\theta}_{0:t}) \tag{III.46}$$

où $\delta_{\boldsymbol{\theta}_{0:t}^{(i)}}(\boldsymbol{\theta}_{0:t})$ est la fonction de Dirac localisée au point $\boldsymbol{\theta}_{0:t}^{(i)}$. Ainsi, tout estimateur optimal $I(\ell)$

d'une fonction ℓ de $\boldsymbol{\theta}_{0:t}$, défini par :

$$I(\ell) = \mathbb{E}_{p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})}[\ell(\boldsymbol{\theta}_{0:t})] \tag{III.47}$$

$$= \int \ell(\boldsymbol{\theta}_{0:t}) p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t} \tag{III.48}$$

peut être approché par :

$$\hat{I}_N(\ell) = \frac{1}{N} \sum_{i=1}^{N} \ell(\boldsymbol{\theta}_{0:t}^{(i)}) \tag{III.49}$$

Il est facile de voir que cet estimateur est non biaisé, $\mathbb{E}[\hat{I}_N(\ell)] = I(\ell)$. En supposant que la matrice de covariance de la fonction ℓ, définie par :

$$\mathrm{cov}[\ell(\boldsymbol{\theta}_{0:t})] = \int \big(\ell(\boldsymbol{\theta}_{0:t}) - I(\ell)\big)^2 p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t} \tag{III.50}$$

est finie, c'est-à-dire que ℓ est de carré intégrable sous $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$, la matrice de covariance $\mathrm{cov}[\hat{I}_N(\ell)]$ est donnée par :

$$\begin{aligned}
\mathrm{cov}[\hat{I}_N(\ell)] &= \mathbb{E}[\hat{I}_N(\ell)\hat{I}_N(\ell)^\intercal] - I(\ell)I(\ell)^\intercal \\
&= \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} \ell(\boldsymbol{\theta}_{0:t}^{(i)})\right)\left(\frac{1}{N}\sum_{i=1}^{N} \ell(\boldsymbol{\theta}_{0:t}^{(i)})\right)^\intercal\right] - I(\ell)I(\ell)^\intercal \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{E}[\ell(\boldsymbol{\theta}_{0:t}^{(i)})\ell(\boldsymbol{\theta}_{0:t}^{(j)})^\intercal] - I(\ell)I(\ell)^\intercal \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \mathbb{E}[\ell(\boldsymbol{\theta}_{0:t}^{(i)})\ell(\boldsymbol{\theta}_{0:t}^{(i)})^\intercal] + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} \mathbb{E}[\ell(\boldsymbol{\theta}_{0:t}^{(i)})]\mathbb{E}[\ell(\boldsymbol{\theta}_{0:t}^{(j)})]^\intercal - I(\ell)I(\ell)^\intercal \\
&= \frac{1}{N}\mathbb{E}[\ell(\boldsymbol{\theta}_{0:t})\ell(\boldsymbol{\theta}_{0:t})^\intercal] + \frac{N^2-N}{N^2} I(\ell)I(\ell)^\intercal - I(\ell)I(\ell)^\intercal \\
&= \frac{1}{N}\big(\mathbb{E}[\ell(\boldsymbol{\theta}_{0:t})\ell(\boldsymbol{\theta}_{0:t})^\intercal] - I(\ell)I(\ell)^\intercal\big) \\
&= \frac{1}{N}\mathrm{cov}[\ell(\boldsymbol{\theta}_{0:t})]
\end{aligned} \tag{III.51}$$

Ainsi, d'après la loi forte des grands nombres, $\hat{I}_N(\ell)$ converge presque sûrement vers $I(\ell)$ quand N tend vers l'infini et le taux de convergence est déterminé par le théorème de la limite centrale :

$$\sqrt{N}\big(\hat{I}_N(\ell) - I(\ell)\big) \underset{N\to\infty}{\Longrightarrow} \mathcal{N}(0, \mathrm{cov}[\ell(\boldsymbol{\theta}_{0:t})]) \tag{III.52}$$

où $\Longrightarrow$ désigne la convergence en distribution. Ce résultat met en avant un point important de la méthode d'approximation de Monte Carlo : la précision de l'estimateur $\hat{I}_N(\ell)$ ne dépend pas de la dimension de l'espace d'état et est inversement proportionnelle au nombre de particules utilisées. Le principal inconvénient de cette approche est qu'il est souvent impossible d'échantillonner selon la distribution *a posteriori*. Autrement dit, l'obtention des particules $\{\boldsymbol{\theta}_{0:t}^{(i)}\}$, $i = 1, \ldots, N$, n'est pas un problème trivial et le recours à des algorithmes spécifiques est nécessaire. Dans la suite, deux méthodes, l'échantillonnage d'importance et par acceptation/rejet, sont présentées.

III.3.1.a Echantillonnage d'importance

L'objectif de l'échantillonnage d'importance (*importance sampling*) est de générer des particules préférentiellement[6] dans les régions d'importance de la distribution *a posteriori*, c'est-à-dire dans les régions de l'espace d'état dans lesquelles elle prend des valeurs élevées. Il repose sur une densité de probabilité $\pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$, appelée densité d'importance, dont le support[7] contient celui de $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ et à partir de laquelle il est possible de générer des échantillons. D'autre part, cette approche est applicable dans le cas général où la distribution *a posteriori* n'est connue qu'à une constante près (ce qui se produit, par exemple, lorsqu'il est impossible d'évaluer la constante de normalisation de la densité). Pour fixer les idées, considérons que $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = c\tilde{p}(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ avec c la constante inconnue et $\tilde{p}(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ la partie évaluable, en un point donné, de la distribution *a posteriori*. Sous ces hypothèses, l'intégrale $I(\ell)$, définie par l'équation (III.48), peut alors être reformulée de la manière suivante :

$$\begin{aligned} I(\ell) &= \int \ell(\boldsymbol{\theta}_{0:t}) p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t} \\ &= c \int \ell(\boldsymbol{\theta}_{0:t}) \frac{\tilde{p}(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})}{\pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})} \pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t} \\ &= \frac{\int \ell(\boldsymbol{\theta}_{0:t}) \omega(\boldsymbol{\theta}_{0:t}) \pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t}}{\frac{1}{c}} \\ &= \frac{\int \ell(\boldsymbol{\theta}_{0:t}) \omega(\boldsymbol{\theta}_{0:t}) \pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t}}{\int \tilde{p}(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t}} \\ &= \frac{\int \ell(\boldsymbol{\theta}_{0:t}) \omega(\boldsymbol{\theta}_{0:t}) \pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t}}{\int \omega(\boldsymbol{\theta}_{0:t}) \pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\theta}_{0:t}} \end{aligned} \tag{III.53}$$

où $\omega(\boldsymbol{\theta}_{0:t})$ est le poids d'importance non normalisé, défini par :

$$\omega(\boldsymbol{\theta}_{0:t}) = \frac{\tilde{p}(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})}{\pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})} \tag{III.54}$$

En appliquant le principe de la méthode de Monte Carlo au numérateur et au dénominateur de l'équation (III.53), avec un ensemble de particules générées à partir de la densité d'importance, $I(\ell)$ peut être estimée par :

$$\hat{I}_N(\ell) = \frac{\frac{1}{N}\sum_{i=1}^{N} \omega(\boldsymbol{\theta}_{0:t}^{(i)}) \ell(\boldsymbol{\theta}_{0:t}^{(i)})}{\frac{1}{N}\sum_{i=1}^{N} \omega(\boldsymbol{\theta}_{0:t}^{(i)})} \tag{III.55}$$

$$= \sum_{i=1}^{N} \tilde{\omega}_t^{(i)} \ell(\boldsymbol{\theta}_{0:t}^{(i)}) \tag{III.56}$$

en définissant les poids d'importance normalisés $\tilde{\omega}_t^{(i)}$ par :

$$\tilde{\omega}_t^{(i)} = \frac{\omega(\boldsymbol{\theta}_{0:t}^{(i)})}{\sum_{i=1}^{N} \omega(\boldsymbol{\theta}_{0:t}^{(i)})} \tag{III.57}$$

[6] L'échantillonnage d'importance est parfois aussi appelé échantillonnage préférentiel.

[7] Le support de $\pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ est défini par $\{\boldsymbol{\theta}_{0:t} \in \mathbb{R}^{n_{\boldsymbol{\theta}} \times (t+1)} | \pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) > 0\}$.

L'estimateur défini par l'équation (III.56) est biaisé (ratio de deux estimateurs) mais consistant, c'est-à-dire que la loi forte des grands nombres peut s'appliquer et que le biais tend presque sûrement vers zéro quand N tend vers l'infini [Gew89].

III.3.1.b Echantillonnage par acceptation/rejet

Comme pour l'échantillonnage d'importance, l'échantillonnage par acceptation/rejet se base sur une densité intermédiaire $\pi(\boldsymbol{\theta}_{0:t})$[8] à partir de laquelle il est possible d'échantillonner, pour générer un ensemble de particules distribuées selon la densité cible $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$. En plus des hypothèses posées précédemment (même support que $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$), il est supposé que $\pi(\boldsymbol{\theta}_{0:t})$ est telle que :

$$p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) \leq C\pi(\boldsymbol{\theta}_{0:t}) \tag{III.58}$$

sur tout le support de $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$, avec $C < \infty$ une constante connue. Là encore, cette relation permet de considérer le cas où la distribution *a posteriori* n'est connue qu'à une constante près, aucune hypothèse n'étant faite sur la valeur de la constante C, ce qui lui permet d'incorporer la constante de normalisation. La méthode de génération des particules est alors décrite en trois points :

1. générer une variable aléatoire u selon la loi uniforme $\mathcal{U}([0,1])$
2. générer une particule $\boldsymbol{\theta}_{0:t}$ selon $\pi(\boldsymbol{\theta}_{0:t})$
3. si $\boldsymbol{\theta}_{0:t} \leq \frac{p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})}{C\pi(\boldsymbol{\theta}_{0:t})}$ l'accepter, sinon retourner à l'étape 1

Cette procédure est itérée jusqu'à obtenir le nombre voulu de particules. Deux remarques importantes peuvent être faite : les échantillons ainsi générés sont distribués selon $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ et n'ont pas besoin d'être pondérés. Cependant, pour obtenir un ensemble de N particules, il est nécessaire d'itérer la procédure un nombre de fois supérieur à N, le taux d'acceptation étant exactement égal à $\frac{1}{C}$ (il est nécessaire de faire environ C essais pour obtenir une particule correctement distribuée).

III.3.1.c Discussion

La nécessité de la connaissance au préalable de la constante C ou le nombre d'essais requis pour obtenir une particule convenablement distribuée sont des facteurs limitant pour l'utilisation de la méthode par acceptation/rejet. De la comparaison plus formelle des deux approches donnée par Robert et Casella [Rob99], il ressort que, sous certaines hypothèses, un estimateur construit à partir des échantillons acceptés par la méthode d'acceptation/rejet est dominé par l'estimateur construit avec l'ensemble des échantillons générés par cette méthode (ceux acceptés et ceux rejetés) utilisés dans un contexte d'échantillonnage d'importance. Enfin, ces deux méthodes d'échantillonnage, telle qu'elles sont présentées ici, sont inadaptées pour faire une estimation séquentielle de la distribution *a posteriori*. En effet, lorsqu'un nouveau vecteur d'observation $\mathbf{y}_t$ arrive, il faut recommencer les calculs pour la nouvelle trajectoire $\boldsymbol{\theta}_{0:t}$,

[8] La dépendance conditionnelle aux observations n'est pas explicitement indiquée ici, contrairement au cas précédent. En fait, pour les deux méthodes d'échantillonnage, cette dépendance n'est pas une nécessité. Le choix de l'exprimer dans la densité d'importance correspond à un souci de clarté et de cohérence par rapport à la suite de ce manuscrit.

étant données les observations $\mathbf{y}_{1:t}$. Cependant, l'échantillonnage d'importance peut être modifié afin de calculer une approximation de $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ en utilisant les particules passées $\boldsymbol{\theta}_{0:t-1}^{(i)}$, $i = 1, \ldots, N$, sans les changer. Cette modification est la base des méthodes de Monte Carlo séquentielles, comme le filtrage particulaire, présentées dans la prochaine section.

III.3.2 Filtre particulaire de base

L'aspect séquentiel du problème d'estimation de la distribution *a posteriori* $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ se résume dans l'utilisation du résultat à l'instant $t-1$ pour estimer la distribution à l'instant t. Les méthodes de Monte Carlo séquentielles permettent d'y répondre en proposant un certain nombre d'outils ayant pour point de départ l'échantillonnage d'importance. Plus précisément, supposons qu'à l'instant $t-1$, un ensemble de N particules $\boldsymbol{\theta}_{0:t-1}^{(i)}$, $i = 1, \ldots, N$, distribuées selon $p(\boldsymbol{\theta}_{0:t-1}|\mathbf{y}_{1:t-1})$ est disponible. L'objectif des méthodes de filtrage particulaire est alors de mettre à jour cet ensemble afin que les nouvelles particules soient distribuées selon $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ et qu'une approximation de la distribution *a posteriori* puisse être obtenue par :

$$\hat{p}_N(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} \tilde{\omega}_t^{(i)} \delta_{\boldsymbol{\theta}_{0:t}^{(i)}}(\boldsymbol{\theta}_{0:t}) \tag{III.59}$$

L'algorithme de base consiste en deux étapes qui sont maintenant présentées.

III.3.2.a Echantillonnage d'importance séquentiel

A chaque instant t, on ne souhaite pas échantillonner toute la trajectoire depuis l'instant 0, mais plutôt mettre à jour les particules de l'instant $t-1$ à l'instant t, sans changer leur passé. Cela revient à choisir une densité d'importance admettant la forme récursive suivante :

$$\pi_t(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = \pi_{t-1}(\boldsymbol{\theta}_{0:t-1}|\mathbf{y}_{1:t-1}) q_t(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t}) \tag{III.60}$$

A l'instant $t = 0$, les particules sont générées selon $\pi(\boldsymbol{\theta}_0)$. En pratique, la trajectoire $\boldsymbol{\theta}_{0:t-1}^{(i)}$ est concaténée avec un vecteur $\boldsymbol{\theta}_t^{(i)} \sim q_t(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$ pour construire la particule $\boldsymbol{\theta}_{0:t}^{(i)}$, $i = 1, \ldots, N$. On peut noter que la dimension des particules augmente, à chaque instant, de la dimension de l'état. Construites de cette manière, les particules à l'instant t sont distribuées selon $\pi_t(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$. Pour corriger cette différence avec la distribution *a posteriori*, il est nécessaire de pondérer les particules en calculant les poids d'importance. Là encore, il est nécessaire d'avoir recours à une formule de calcul récursive. A l'instar de la densité de filtrage, la distribution *a posteriori* vérifie une formulation récursive :

$$\begin{aligned}
p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) &= p(\boldsymbol{\theta}_{0:t-1}, \boldsymbol{\theta}_t|\mathbf{y}_{1:t-1}, \mathbf{y}_t) \\
&= \frac{p(\boldsymbol{\theta}_{0:t-1}, \boldsymbol{\theta}_t, \mathbf{y}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \\
&= p(\boldsymbol{\theta}_{0:t-1}|\mathbf{y}_{1:t-1}) \frac{p(\boldsymbol{\theta}_t, \mathbf{y}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \\
&= p(\boldsymbol{\theta}_{0:t-1}|\mathbf{y}_{1:t-1}) \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t, \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t-1}) p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \\
&= p(\boldsymbol{\theta}_{0:t-1}|\mathbf{y}_{1:t-1}) \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}
\end{aligned} \tag{III.61}$$

avec la convention $p(\boldsymbol{\theta}_0|\mathbf{y}_{1:0}) = p(\boldsymbol{\theta}_0)$. On retrouve, dans l'équation (III.61), l'*a priori* séquentiel $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$ et la vraisemblance $p(\mathbf{y}_t|\boldsymbol{\theta}_t)$, qui sont, rappelons-le, définis dans le cas général par le système d'équation (III.22)-(III.23). Il est en outre supposé que ces deux densités peuvent être évaluées, éventuellement à une constante près, en un point donné. Le calcul précédent s'effectue en utilisant plusieurs fois la relation $p(a,b) = p(a|b)p(b)$ et les différentes densités sont simplifiées grâce aux hypothèses posées, que nous rappelons ici :

- les observations sont supposées indépendantes conditionnellement à l'état, ainsi l'information $\mathbf{y}_t|\boldsymbol{\theta}_t, \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t-1}$ est équivalente à $\mathbf{y}_t|\boldsymbol{\theta}_t$
- l'état est supposé markovien, ainsi l'information $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}$ est équivalente à $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}$

En reprenant les notations de la section III.3.1.a, $\tilde{p}(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ désigne la partie connue, à une constante près, et évaluable de la distribution *a posteriori*. L'équation (III.61) peut alors se mettre sous la forme :

$$\tilde{p}(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = \tilde{p}(\boldsymbol{\theta}_{0:t-1}|\mathbf{y}_{1:t-1})p(\mathbf{y}_t|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) \tag{III.62}$$

En combinant les équations (III.60) et (III.62), on obtient une formule récursive permettant de calculer le poids d'importance non normalisé à l'instant t en mettant à jour celui normalisé à l'instant $t-1$:

$$\begin{aligned}
\omega_t^{(i)} &= \frac{\tilde{p}(\boldsymbol{\theta}_{0:t}^{(i)}|\mathbf{y}_{1:t})}{\pi_t(\boldsymbol{\theta}_{0:t}^{(i)}|\mathbf{y}_{1:t})} \\
&= \frac{\tilde{p}(\boldsymbol{\theta}_{0:t-1}^{(i)}|\mathbf{y}_{1:t-1})}{\pi_{t-1}(\boldsymbol{\theta}_{0:t-1}^{(i)}|\mathbf{y}_{1:t-1})} \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)})p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)})}{q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} \\
&= \omega_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)})p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)})}{q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} \\
&\propto \tilde{\omega}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)})p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)})}{q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})}
\end{aligned} \tag{III.63}$$

L'utilisation d'une formule définie à une constante près ne pose pas de problème, les poids étant ensuite normalisés par :

$$\tilde{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{i=1}^N \omega_t^{(i)}} \tag{III.64}$$

L'algorithme d'échantillonnage d'importance séquentiel est donné dans le tableau III.4. Ainsi, une approximation de la vraie distribution *a posteriori* à l'instant t, est obtenue en mettant à jour les particules et les poids à l'instant précédent, c'est-à-dire à partir de l'approximation obtenue à l'instant $t-1$.

III.3.2.b Rééchantillonnage

Il peut être montré [Kon94] que, avec une densité d'importance ayant la forme décrite dans l'équation (III.60), la variance des poids augmente avec le temps. Ce phénomène est communément appelé la dégénérescence des poids et il se traduit par une diminution du nombre de particules utiles. En effet, après quelques itérations de l'échantillonnage séquentiel, tous les

A l'instant $t = 0$
Initialisation
Pour $i = 1, \ldots, N$
Générer $\boldsymbol{\theta}_0^{(i)} \sim \pi(\boldsymbol{\theta}_0)$
$\tilde{\omega}_0^{(i)} = \frac{1}{N}$
A l'instant $t \geq 1$
Pour $i = 1, \ldots, N$
Mise à jour des particules
$\boldsymbol{\theta}_t^{(i)} \sim q_t(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$
$\boldsymbol{\theta}_{0:t}^{(i)} = \{\boldsymbol{\theta}_{0:t-1}^{(i)}\ ,\ \boldsymbol{\theta}_t^{(i)}\}$
Calcul des poids non normalisés
$\omega_t^{(i)} \propto \tilde{\omega}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)}) p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)})}{q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})}$
Normalisation des poids
Pour $i = 1, \ldots, N$
$\tilde{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{i=1}^N \omega_t^{(i)}}$

TAB. III.4 – Algorithme d'échantillonnage d'importance séquentiel.

poids deviennent nuls à l'exception d'un dont la valeur tend vers 1. La conséquence directe est que l'approximation de l'équation (III.59) est de plus en plus fausse. De plus, d'un point de vue pratique, cela signifie que beaucoup de temps de calcul sera nécessaire pour mettre à jour des particules qui, au final, auront un poids quasiment nul. Une solution à ce problème est alors de calculer une approximation équi-pondérée de la distribution *a posteriori*, ce qui se fait par l'ajout d'une étape de rééchantillonnage (encore appelée étape de sélection) à l'algorithme présenté dans le tableau III.4. L'introduction de cette étape dans un contexte de filtrage bayésien, a permis de développer le premier filtre particulaire véritablement opérationnel [Gor93]. Le principe du rééchantillonnage est de supprimer les particules ayant un poids faible et de multiplier celles ayant un poids élevé. Plus précisément, chaque particule $\boldsymbol{\theta}_{0:t}^{(i)}$, $i = 1, \ldots, N$, est dupliquée $N_t^{(i)}$ fois avec $\mathbb{E}[N_t^{(i)}] = N\tilde{\omega}_t^{(i)}$ et $\sum_{i=1}^N N_t^{(i)} = N$ (le nombre de particules reste donc constant au cours du temps) et un nouvel ensemble de particules $\{\boldsymbol{\theta}_{0:t}^{(i)*}\}_{i=1,\ldots,N}$ est obtenu. Les particules pour lesquelles $N_t^{(i)} = 0$ sont donc supprimées. La valeur de $N_t^{(i)}$ est choisie de telle sorte que l'approximation équi-pondérée

$$\hat{p}_N^*(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\theta}_{0:t}^{(i)*}}(\boldsymbol{\theta}_{0:t}) \tag{III.65}$$

est proche de l'approximation pondérée, équation (III.59), au sens où, pour toute fonction ℓ, on a :

$$\sum_{i=1}^N \tilde{\omega}_t^{(i)} \ell(\boldsymbol{\theta}_{0:t}^{(i)}) \approx \frac{1}{N}\sum_{i=1}^N \ell(\boldsymbol{\theta}_{0:t}^{(i)*}) \tag{III.66}$$

Classiquement, le calcul des $N_t^{(i)}$ se fait indirectement en générant les N nouvelles particules

Initialisation
Construction de la somme cumulative des poids normalisés
$c_1 = \tilde{\omega}_t^{(1)}$
Pour $k = 2, \ldots, N$
$c_k = c_{k-1} + \tilde{\omega}_t^{(k)}$
Choix d'un point de départ
$u \sim \mathcal{U}([0, \frac{1}{N}])$
$i = 1$
Rééchantillonnage
Pour $j = 1, \ldots, N$
$u_j = u + \frac{j-1}{N}$
Tant que $u_j > c_i$
$i = i + 1$
$\boldsymbol{\theta}_{0:t}^{(j)*} = \boldsymbol{\theta}_{0:t}^{(i)}$
Mettre les poids à $\frac{1}{N}$

TAB. III.5 – Algorithme d'échantillonnage systématique (méthode de rééchantillonnage proposée par [Kit96]).

selon l'approximation $\hat{p}_N(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ de la distribution *a posteriori*, obtenue par échantillonnage d'importance séquentiel à l'instant t (c'est la méthode du *bootstrap*). Ce processus peut être assimilé à une approximation de Monte Carlo, d'où la redéfinition des poids en $\frac{1}{N}$. Cette approche équivaut à générer les $N_t^{(i)}$ selon une loi multinomiale de paramètres les poids normalisés $\tilde{\omega}_t^{(i)}$. On a alors $\mathbb{E}[N_t^{(i)}] = N\tilde{\omega}_t^{(i)}$ et $\text{var}[N_t^{(i)}] = N\tilde{\omega}_t^{(i)}(1 - \tilde{\omega}_t^{(i)})$. D'autres méthodes de calcul existent [Car99] et elles assurent toutes que $\mathbb{E}[N_t^{(i)}] = N\tilde{\omega}_t^{(i)}$, la différence venant de la valeur de la variance. Une méthode particulièrement intéressante est celle de l'échantillonnage systématique [Kit96], dans la mesure où elle permet de minimiser la variance de $N_t^{(i)}$. L'algorithme est donné dans le tableau III.5.

Il est important de noter que, quel que soit la méthode utilisée pour calculer les $N_t^{(i)}$, l'étape de rééchantillonnage dégrade l'approximation de Monte Carlo. En effet, si dans l'équation (III.66), les deux estimateurs ont la même espérance, la variance du second est généralement supérieure à celle du premier. Il est donc parfois préférable de ne pas systématiquement passer par l'étape de rééchantillonnage. Ici encore, plusieurs stratégies sont possibles : rééchantillonner tout les T instants ou adopter une solution dynamique basée sur une mesure de la taille de l'ensemble des particules effectives (c'est-à-dire ayant un poids non négligeable). Une telle mesure [Kon94, Liu95] peut être définie par :

$$\begin{aligned} N_{\text{eff}} &= \frac{N}{1 + \text{var}[\tilde{\omega}_t]} \\ &= \frac{N}{1 + \mathbb{E}[\tilde{\omega}_t^2]} \end{aligned} \tag{III.67}$$

A l'instant $t = 0$
Initialisation
Pour $i = 1, \ldots, N$
Générer $\boldsymbol{\theta}_0^{(i)} \sim \pi(\boldsymbol{\theta}_0)$
$\tilde{\omega}_0^{(i)} = \frac{1}{N}$
A l'instant $t \geq 1$
Echantillonnage d'importance séquentiel
Pour $i = 1, \ldots, N$
Mise à jour des particules
$\boldsymbol{\theta}_t^{(i)} \sim q_t(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$
$\boldsymbol{\theta}_{0:t}^{(i)} = \{\boldsymbol{\theta}_{0:t-1}^{(i)}\ ,\ \boldsymbol{\theta}_t^{(i)}\}$
Calcul des poids non normalisés
$\omega_t^{(i)} \propto \tilde{\omega}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)}) p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)})}{q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})}$
Normalisation des poids
Pour $i = 1, \ldots, N$
$\tilde{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{i=1}^N \omega_t^{(i)}}$
Rééchantillonnage
$\hat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^2}$
Si $\hat{N}_{\text{eff}} < N_s$
Appliquer l'algorithme III.5
Sinon
$\boldsymbol{\theta}_{0:t}^{(i)*} = \boldsymbol{\theta}_{0:t}^{(i)}, \forall i$

TAB. III.6 – Algorithme de filtrage particulaire de base.

et peut être estimée par :

$$\hat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^2} \tag{III.68}$$

Le principe est alors de rééchantillonner les particules lorsque $\hat{N}_{\text{eff}}$ devient inférieur à un certain seuil N_s.

L'algorithme de filtrage particulaire de base, donné dans le tableau III.6, est ainsi construit à partir des deux étapes d'échantillonnage d'importance séquentiel et de rééchantillonnage.

III.3.3 Améliorations de l'algorithme de base

Dans les sections précédentes, l'influence de la densité d'importance $q_t(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t})$ a été peu abordé. Pourtant, elle joue un rôle primordial dans l'efficacité des algorithmes de filtrage particulaire et son choix est critique. En effet, combinée avec l'étape de rééchantillonnage, elle permet de limiter l'augmentation de la variance des poids. Il est alors naturel de chercher quelle est la densité d'importance qui serait la « meilleure », au sens d'un certain critère. Il peut

être montré [Dou98] que la densité d'importance optimale, c'est-à-dire celle qui minimise la variance du poids $\tilde{\omega}_t^{(i)}$, conditionnellement aux états passés $\boldsymbol{\theta}_{0:t-1}^{(i)}$ et aux observations $\mathbf{y}_{1:t}$, est $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}^{(i)}, \mathbf{y}_t)$. Dans ce cas, les poids non normalisés se calculent par :

$$\begin{aligned} \omega_t^{(i)} &\propto \tilde{\omega}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)}) p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)})}{p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)}, \mathbf{y}_t)} \\ &\propto \tilde{\omega}_{t-1}^{(i)} p(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)}) \end{aligned} \tag{III.69}$$

avec

$$p(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)}) = \int p(\mathbf{y}_t|\boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}^{(i)}) \mathrm{d}\boldsymbol{\theta}_t \tag{III.70}$$

L'utilisation de la densité d'importance optimale ne peut se faire que si :
- il est possible de générer des échantillons selon $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}^{(i)}, \mathbf{y}_t)$
- il est possible de calculer l'intégrale (III.70), à une constante près

En pratique, les deux conditions précédentes ne sont remplies que pour certains types de modèles (en particulier si les bruits sont supposés gaussiens et additifs et si l'équation d'observation est linéaire). Dans le cas général, il est nécessaire d'avoir recours à des solutions sous-optimales. Plusieurs approches existent dans la littérature, on peut donner quelques exemples :

- choisir l'*a priori* séquentiel $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$ comme densité d'importance. C'est la solution adoptée dans [Gor93]. L'algorithme ainsi construit, parfois appelé CONDENSATION, est simple et rapide. Les poids se calculent simplement par $\omega_t^{(i)} \propto p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)})$. Le principal inconvénient de cette solution est que l'information apportée à l'instant t par $\mathbf{y}_t$, n'est pas prise en compte pour mettre à jour les particules, ce qui se traduit par une grande variance.
- faire une approximation gaussienne de la densité optimale en faisant un pas de filtre de Kalman étendu [Che00] ou en utilisant la transformée sans parfum [Mer00]. En pratique, la seconde solution donne souvent de meilleurs résultats principalement grâce à l'approximation de la queue de la distribution qui est plus précise par la transformée sans parfum.

D'une manière générale, si elle n'est pas directement utilisable, la densité d'importance optimale permet néanmoins de définir des critères premettant de guider le choix de la densité d'importance. En particulier, on va chercher des densités de la forme $q_t(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}^{(i)}, \mathbf{y}_t)$, le plus proche possible de celle optimale, qui prennent en compte à la fois l'état passé et le vecteur d'observations courant et dont le support contient celui de la distribution *a posteriori*. En particulier, il est nécessaire que la densité soit suffisamment large pour prendre en compte les points marginaux. Enfin, la densité d'importance doit permettre d'avoir une variance la plus petite possible.

En plus du choix de la densité d'importance, d'autres stratégies peuvent être utilisées pour limiter la dégénérescence des poids. Dans les deux sections suivantes sont présentées le rééchantillonnage par variables auxiliaires, qui permet de réduire la perte de diversité due à l'étape de sélection et le filtre particulaire rao-blackwellisé qui permet, sous certaines conditions, de réduire la dimension de l'espace à explorer.

A l'instant $t = 0$

Initialisation

Pour $i = 1, \ldots, N$

Générer $\boldsymbol{\theta}_0^{(i)} \sim \pi(\boldsymbol{\theta}_0)$

A l'instant $t \geq 1$

Rééchantillonnage par variables auxiliaires

Pour $i = 1, \ldots, N$

Calculer $\hat{p}(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)})$

$\lambda_t^{(i)} \propto \omega_{t-1}^{(i)} \hat{p}(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)})$

$\tilde{\lambda}_t^{(i)} = \frac{\lambda_t^{(i)}}{\sum_{i=1}^N \lambda_t^{(i)}}$

Dupliquer/supprimer les particules $\boldsymbol{\theta}_{0:t-1}^{(i)}$ en fonction de $\tilde{\lambda}_t^{(i)}$ pour obtenir les particules $\boldsymbol{\theta}_{0:t-1}^{(i)*}$

Echantillonnage d'importance séquentiel

Pour $i = 1, \ldots, N$

Mise à jour des particules

$\boldsymbol{\theta}_t^{(i)} \sim q_t(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}^{(i)*}, \mathbf{y}_{1:t})$

$\boldsymbol{\theta}_{0:t}^{(i)} = \{\boldsymbol{\theta}_{0:t-1}^{(i)*}\ ,\ \boldsymbol{\theta}_t^{(i)}\}$

Calcul des poids non normalisés

$\omega_t^{(i)} \propto \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)}) p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)*})}{\hat{p}(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)}) q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{0:t-1}^{(i)*}, \mathbf{y}_{1:t})}$

TAB. III.7 – Algorithme de filtrage particulaire avec rééchantillonnage par variables auxiliaires.

III.3.3.a Rééchantillonnage par variables auxiliaires

L'étape de rééchantillonnage diminue la diversité au sein de l'ensemble de particules. En effet, les particules ayant un poids élevé sont souvent dupliquées tandis que celles ayant un poids faible, c'est-à-dire dans la queue de la distribution, finissent toujours par disparaître. Pour diminuer ce problème, Pitt et Shephard [Pit99] ont introduit le filtre particulaire avec une étape de rééchantillonnage par variables auxiliaires (*auxiliary particle filter*). Le principe est assez intuitif : au lieu d'étendre des particules qui seront par la suite supprimées, l'algorithme duplique d'abord les particules ayant un fort potentiel puis les nouvelles particules sont étendues de l'instant $t-1$ à l'instant t. La mesure du potentiel de la particule i, $i = 1, \ldots, N$, c'est-à-dire de sa capacité à avoir un poids élevé après l'étape de mise à jour, est donnée par la vraisemblance prédictive $p(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)})$. Comme nous l'avons vu précédemment, cette densité est définie par une intégrale qui est généralement incalculable et une approximation $\hat{p}(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)})$ est nécessaire. Dans l'article original [Pit99], les auteurs proposent de construire $\hat{p}(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)})$ en faisant une approximation de l'intégrale (III.70) en considérant $p(\mathbf{y}_t|\boldsymbol{\mu}(\boldsymbol{\theta}_{t-1}^{(i)}))$ au lieu de $p(\mathbf{y}_t|\boldsymbol{\theta}_t)$, avec $\boldsymbol{\mu}(\boldsymbol{\theta}_{t-1}^{(i)})$ la moyenne, le mode ou toute autre quantité caractérisant $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}^{(i)})$. Plus récemment, Andrieu *et al.* [And01] ont proposé une version améliorée de cet algorithme, basée sur une approximation gaussienne de la vraisemblance prédictive. Le calcul de ses deux premiers moments est effectué avec la transformée sans parfum (voir annexe B).

L'algorithme complet est donné dans le tableau III.7. Il est inversé par rapport au filtre particulaire classique. Les particules à l'instant $t-1$ sont d'abord rééchantillonnées en générant les nouvelles particules selon une loi multinomiale, comme dans l'étape de sélection préconisée par [Gor93]. La différence est que les paramètres de cette loi sont les variables auxiliaires $\tilde{\lambda}_t^{(i)}$, définies par :

$$\lambda_t^{(i)} \propto \omega_{t-1}^{(i)} \hat{p}(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)}) \tag{III.71}$$

$$\tilde{\lambda}_t^{(i)} = \frac{\lambda_t^{(i)}}{\sum_{i=1}^N \lambda_t^{(i)}} \tag{III.72}$$

On peut aussi noter que les poids non normalisés sont calculés par :

$$\omega_t^{(i)} \propto \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)}) p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)*})}{\hat{p}(\mathbf{y}_t|\boldsymbol{\theta}_{t-1}^{(i)}) q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{0:t-1}^{(i)*}, \mathbf{y}_{1:t})} \tag{III.73}$$

Une première remarque est qu'il n'est plus nécessaire de normaliser ces poids, la normalisation se faisant sur les variables auxiliaires. Ensuite, l'ajout de l'approximation de la vraisemblance prédictive au dénominateur est assez naturelle car dans cet algorithme, on peut considérer que l'étape de rééchantillonnage fait partie du processus de génération de la nouvelle particule à l'instant t.

III.3.3.b Rao-blackwellisation

Le principe de base de la rao-blackwellisation est d'exploiter au maximum la structure du modèle afin d'améliorer l'efficacité de l'inférence et, par conséquent, d'en réduire la variance. Le système dynamique considéré dans toute cette partie est décrit par les équations (III.22) et (III.23), que nous rappelons ici :

$$\begin{aligned} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) &\longleftrightarrow \boldsymbol{\theta}_t = g_t(\boldsymbol{\theta}_{t-1}, \mathbf{v}_{t-1}^{\boldsymbol{\theta}}) \\ p(\mathbf{y}_t|\boldsymbol{\theta}_t) &\longleftrightarrow \mathbf{y}_t = h_t(\boldsymbol{\theta}_t, \mathbf{v}_{t-1}^{\mathbf{y}}) \end{aligned}$$

Il n'est pas rare de pouvoir partitionner le vecteur d'état en deux parties, $\boldsymbol{\theta} = \{\boldsymbol{\theta}^a, \boldsymbol{\theta}^b\}$, et de se trouver dans la situation où, conditionnellement à $\boldsymbol{\theta}^a$, le modèle est linéaire/gaussien. Cela peut s'écrire sous la forme :

$$\boldsymbol{\theta}_t^a \sim p(\boldsymbol{\theta}_t^a|\boldsymbol{\theta}_{t-1}^a) \tag{III.74}$$

$$\boldsymbol{\theta}_t^b = \mathbf{A}(\boldsymbol{\theta}_t^a)\boldsymbol{\theta}_{t-1}^b + \mathbf{B}(\boldsymbol{\theta}_t^a)\mathbf{v}_{t-1}' \tag{III.75}$$

$$\mathbf{y}_t = \mathbf{C}(\boldsymbol{\theta}_t^a)\boldsymbol{\theta}_t^b + \mathbf{D}(\boldsymbol{\theta}_t^a)\mathbf{v}_{t-1}'' \tag{III.76}$$

où la dépendance à $\boldsymbol{\theta}_t^a$ des matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ et $\mathbf{D}$ est non linéaire. Dans ce cas précis, la distribution *a posteriori* peut se factoriser de la manière suivante :

$$p(\boldsymbol{\theta}_{0:t}^a, \boldsymbol{\theta}_{0:t}^b|\mathbf{y}_{1:t}) = \underbrace{p(\boldsymbol{\theta}_{0:t}^b|\mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}^a)}_{\text{(III.77)}'} \underbrace{p(\boldsymbol{\theta}_{0:t}^a|\mathbf{y}_{1:t})}_{\text{(III.77)}''} \tag{III.77}$$

A l'instant $t = 0$

Initialisation

Pour $i = 1, \ldots, N$

Générer $\boldsymbol{\theta}_0^{a(i)} \sim \pi(\boldsymbol{\theta}_0^a)$

$\boldsymbol{\mu}_{0|0}^{(i)} = \boldsymbol{\mu}_0$

$\boldsymbol{\Sigma}_{0|0}^{(i)} = \boldsymbol{\Sigma}_0$

$\tilde{\omega}_0^{(i)} = \frac{1}{N}$

A l'instant $t \geq 1$

Echantillonnage d'importance séquentiel

Pour $i = 1, \ldots, N$

Mise à jour des particules

$\boldsymbol{\theta}_t^{a(i)} \sim q_t(\boldsymbol{\theta}_t^a | \boldsymbol{\theta}_{0:t-1}^{a(i)}, \mathbf{y}_{1:t})$

$\boldsymbol{\theta}_{0:t}^{a(i)} = \{\boldsymbol{\theta}_{0:t-1}^{a(i)} \ , \ \boldsymbol{\theta}_t^{a(i)}\}$

Calcul de $\boldsymbol{\mu}_{t|t}^{(i)}$ et $\boldsymbol{\Sigma}_{t|t}^{(i)}$ en fonction de $\mathbf{y}_t$, $\boldsymbol{\theta}_t^{a(i)}$, $\boldsymbol{\mu}_{t-1|t-1}^{(i)}$ et $\boldsymbol{\Sigma}_{t-1|t-1}^{(i)}$ avec les équations du filtre de Kalman (voir annexe A)

Calcul des poids non normalisés

$\omega_t^{(i)} \propto p(\mathbf{y}_t | \boldsymbol{\theta}_{0:t}^{a(i)}, \mathbf{y}_{1:t-1}) \frac{p(\boldsymbol{\theta}_t^{a(i)} | \boldsymbol{\theta}_{t-1}^{a(i)})}{q_t(\boldsymbol{\theta}_t^{a(i)} | \boldsymbol{\theta}_{0:t-1}^{a(i)}, \mathbf{y}_{1:t})}$

Normalisation des poids

Pour $i = 1, \ldots, N$

$\tilde{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{i=1}^N \omega_t^{(i)}}$

Rééchantillonnage

Dupliquer/supprimer les particules en fonction de $\tilde{\omega}_t^{(i)}$

TAB. III.8 – Algorithme de filtrage particulaire rao-blackwellisé. On peut noter que dans la version présentée ici, les particules sont rééchantillonnées à chaque itération. De ce fait, il n'est pas nécessaire de tenir compte, pour le calcul des poids non normalisés, de la valeur du poids à l'itération précédente.

La densité conditionnelle (III.77)′ est une gaussienne dont les deux premiers moments peuvent être calculés analytiquement lorsque $\boldsymbol{\theta}_{0:t}^a$ est connu. Ainsi, l'estimation de la séquence de distribution *a posteriori* est effectuée en combinant un filtre particulaire, pour faire une approximation de (III.77)″, avec une banque de filtres de Kalman, pour calculer analytiquement (III.77)′. Cette combinaison est appelée filtre particulaire rao-blackwellisé [Fre02, And02].

L'approximation particulaire de (III.77)″ est donnée par :

$$\hat{p}(\boldsymbol{\theta}_{0:t}^a | \mathbf{y}_{1:t}) = \sum_{i=1}^{N} \tilde{\omega}_t^{(i)} \delta_{\boldsymbol{\theta}_{0:t}^{a(i)}}(\boldsymbol{\theta}_{0:t}^a) \tag{III.78}$$

et la densité conditionnelle (III.77)′ est un mélange de gaussiennes défini par :

$$\hat{p}(\boldsymbol{\theta}^b_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} \tilde{\omega}_t^{(i)} p(\boldsymbol{\theta}^b_{0:t}|\mathbf{y}_{1:t}, \boldsymbol{\theta}^{a(i)}_{0:t}) \tag{III.79}$$

$$= \sum_{i=1}^{N} \tilde{\omega}_t^{(i)} \mathcal{N}(\boldsymbol{\theta}^b_{0:t}; \boldsymbol{\mu}^{(i)}_{t|t}, \boldsymbol{\Sigma}^{(i)}_{t|t}) \tag{III.80}$$

L'algorithme, donné dans le tableau III.8, est essentiellement le même que celui du filtre particulaire classique. La principale différence réside dans le calcul des poids non normalisés :

$$\omega_t^{(i)} \propto \omega_{t-1}^{(i)} p(\mathbf{y}_t|\boldsymbol{\theta}^{a(i)}_{0:t}, \mathbf{y}_{1:t-1}) \frac{p(\boldsymbol{\theta}^{a(i)}_t|\boldsymbol{\theta}^{a(i)}_{t-1})}{q_t(\boldsymbol{\theta}^{a(i)}_t|\boldsymbol{\theta}^{a(i)}_{0:t-1}, \mathbf{y}_{1:t})} \tag{III.81}$$

La vraisemblance est une gaussienne définie par :

$$p(\mathbf{y}_t|\boldsymbol{\theta}^{a(i)}_{0:t}, \mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{y}_t; \mathbf{y}_{t|t-1}, \mathbf{S}_{t|t-1}) \tag{III.82}$$

où la moyenne $\mathbf{y}_{t|t-1}$ et la matrice de covariance $\mathbf{S}_{t|t-1}$ sont calculées par les équations du filtre de Kalman (voir annexe A). On peut noter qu'elle ne se simplifie pas en $p(\mathbf{y}_t|\boldsymbol{\theta}^{a(i)}_t)$, du fait d'une dépendance aux valeurs passées à cause de $\boldsymbol{\theta}^b_{0:t}$.

III.4 Synthèse

Dans ce long chapitre, a été introduit le problème du filtrage statistique, que nous avons mis en relation avec la théorie bayésienne et les méthodes de Monte Carlo. La réunion de ces trois champs des sciences dans une même partie, ne doit pas surprendre. En effet, le filtrage bayésien utilise les principes de l'inférence bayésienne pour résoudre le problème du filtrage. Cette solution est largement reconnue comme étant optimale, dans la mesure où toute l'information nécessaire à l'estimation du vecteur d'état, est contenue dans la distribution *a posteriori*. Cependant, dériver un estimateur à chaque instant, à partir de la séquence de distributions, n'est pas un problème trivial car généralement insoluble analytiquement. D'autant plus que le contexte qui nous intéresse est celui où les équations régissant l'évolution du vecteur d'état ou son lien avec les observations, ne sont pas linéaires. Même si les bruits d'état et de mesure sont supposés gaussiens et additifs, les non linéarités du modèle peuvent faire que la distribution *a posteriori* présente un caractère très piqué ou multimodal. C'est ici que les approximations gaussiennes proposées par les extensions du filtre de Kalman trouvent leurs limites. Par contre, les méthodes de Monte Carlo, et plus particulièrement leur formulation séquentielle, prennent toute leur importance, car elles apportent une solution efficace et des outils flexibles sous des hypothèses minimales. Plusieurs résultats de convergence des filtres particulaires ont été établis (voir [Cri01, Cri02] pour une approche accessible). En particulier, il peut être montré que si le noyau de transition de Markov de $\boldsymbol{\theta}_t$, dont la densité est définie par $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$, a un oubli exponentiel du passé, autrement dit que le système n'est pas à mémoire longue, alors on a une

convergence uniforme, dans le temps, de l'approximation particulaire de la densité de filtrage vers la vraie densité de filtrage :

$$\mathbb{E}\left[\left(\int \ell(\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t})\mathrm{d}\boldsymbol{\theta}_t - \int \ell(\boldsymbol{\theta}_t)\hat{p}_N(\boldsymbol{\theta}_t|\mathbf{y}_{1:t})\mathrm{d}\boldsymbol{\theta}_t\right)^2\right] \leq \frac{C||\ell||^2}{N} \tag{III.83}$$

avec C une constante indépendante de N. Le taux de convergence est indépendant de la dimension de l'état et est en $\frac{1}{N}$. On peut noter que la convergence est prise au sens de l'erreur quadratique moyenne et qu'elle est établie que si la fonction ℓ est bornée. Cependant, l'utilisation des méthodes de Monte Carlo pour résoudre le filtrage bayésien est assez récente et il est raisonnable de penser que des théorèmes de convergence moins restrictifs seront établis dans un futur proche. De plus, les bons résultats obtenus par les algorithmes de filtrage particulaire, dans de nombreux domaines, montrent le potentiel de ces méthodes pour résoudre le problème optimal du filtrage bayésien. A ce sujet, en suivant la pensée de J. W. Tukey, on peut dire que :

> *Il est préférable d'avoir une solution approximative d'un problème exact qu'une solution exacte d'un problème approximatif.*

Nous avons vu que les algorithmes construits sur le filtre de Kalman, même s'ils donnaient des résultats limités dans le cas non linéaire général, pouvaient être utilisés au sein des filtres particulaires. Ici se trouve un point crucial des méthodes de Monte Carlo séquentielles. En effet, l'efficacité du filtre particulaire repose en grande partie sur la densité d'importance, ce qui, à première vue, peut sembler une faiblesse de ce type d'approche. Cependant, nous avons aussi vu qu'il était possible de définir une densité d'importance optimale, selon le critère du minimum de variance. Bien que souvent inexploitable, ce résultat met en avant que certains choix peuvent donner de meilleurs résultats que d'autres. Le rôle de la densité d'importance est essentiellement de proposer un nouvel état, à l'instant t, pour mettre à jour les particules. Assez naturellement, pour que le filtre soit efficace, il faut proposer des états dans les régions de l'espace où la distribution *a posteriori* prend des valeurs élevées. Les conditions sur la densité d'importance sont assez légères et on peut arguer que n'importe qu'elle méthode, même *ad hoc* peut être utilisée. C'est ainsi que le choix de la densité d'importance peut se transformer en atout des méthodes de Monte Carlo séquentielles, car elles permettent de combiner une méthode d'estimation donnée avec le cadre bayésien que nous avons défini.

Un dernier point sur lequel il est nécessaire de revenir est le choix de l'approche par filtrage, ou plus précisément, d'une approche en-ligne, où seulement les observations jusqu'à l'instant courant sont utilisées pour estimer le vecteur d'état. Comme précisé au début de ce chapitre, le filtrage est à distinguer de la prédiction et du lissage. Dans le contexte qui nous intéresse, les observations sont supposées arriver instant après instant et le choix du filtrage s'impose de lui-même. Cependant, il est reconnu que les méthodes hors-ligne donnent des résultats plus précis et plus lisses. Ceci parce qu'elles disposent de toutes les observations et peuvent donc utiliser plus d'information pour effectuer l'estimation à l'instant t, que les approches en-ligne. Les méthodes de Monte Carlo peuvent d'ailleurs être utilisées dans ce contexte, notamment avec la simulation de Monte Carlo par chaîne de Markov (*Monte Carlo Markov Chain*, MCMC) [Rob99] et plusieurs applications montrent leur efficacité [And99, Dav06]. Le principal inconvénient de ces méthodes est d'ordre pratique. En effet, elles nécessitent souvent beaucoup de place mémoire (toutes les données sont stockées) et des temps de calculs importants. Il est cependant possible

d'opter pour une solution intermédiaire consistant à faire du lissage de manière séquentiel, c'est-à-dire d'estimer l'état $\boldsymbol{\theta}_t$ en utilisant un certain nombre d'observations futures, ce nombre étant constant au cours du temps. Ce qui revient à estimer la distribution $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t+T})$, avec $T > 0$ [Cla99, Dou04]. Une autre approche est de raffiner, à l'instant t, les estimations passées jusqu'à un horizon fixe $t - T$ [Dou06]. Cependant, toutes ces approches peuvent être mises en œuvre sur la base d'un algorithme de filtrage particulaire. C'est pourquoi, le choix initial de se placer dans un contexte de filtrage n'est pas limitatif.

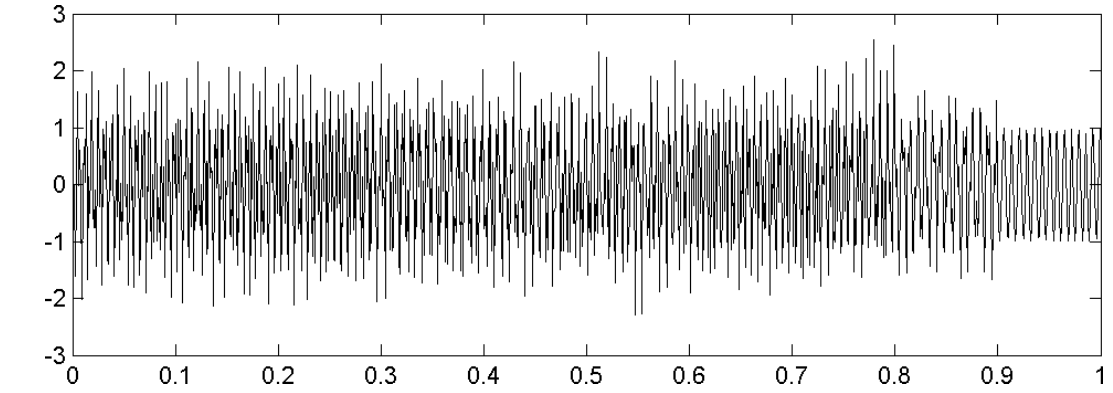
Chapitre IV

Algorithme séquentiel d'estimation

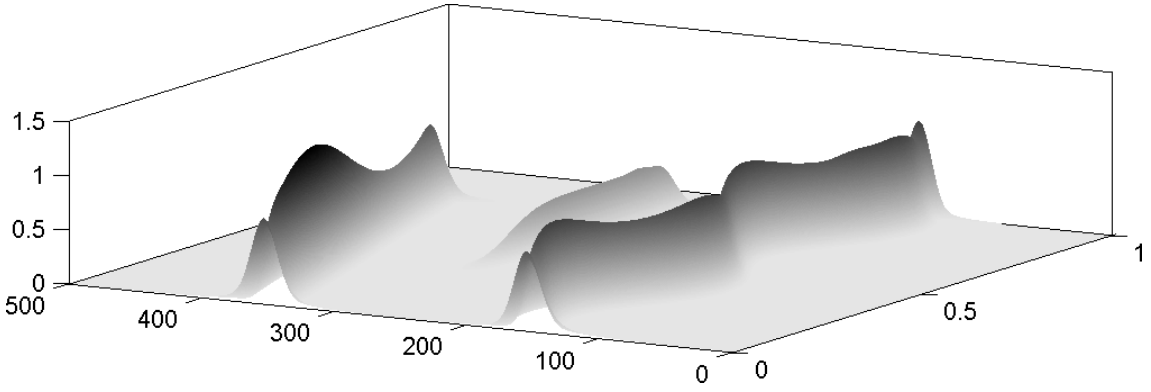
La théorie sans la pratique
est comme un vaisseau sans voile ni gouvernail,
la pratique sans théorie
est comme un vaisseau sans compas.
J.-B. A. Lebas

Les précédents chapitres ont posé les bases nécessaires pour aborder l'exposé de la méthode proposée dans ce manuscrit, pour résoudre le problème de la détection et de l'estimation séquentielles de fréquences fondamentales. Rappelons d'abord le contexte et l'objectif que l'on cherche à atteindre. Les signaux considérés résultent du mélange de plusieurs sources sonores, appartenant à deux catégories différentes : soit elles sont périodiques ou quasi périodiques, auquel cas elles peuvent être caractérisées par une certaine structure fréquentielle, soit elles sont non périodiques et sont modélisées par des bruits à large bande spectrale. Le nombre de sources interférant est inconnu et variable au cours du temps. Il est important de noter que ce nombre ne comptabilise que les sources de la première catégorie, les autres étant toutes regroupées dans le bruit. La structure fréquentielle des sources périodiques ou quasi périodiques est décrite par un ensemble de partiels dont les fréquences sont reliées, de manière harmonique ou inharmonique, à une fréquence fondamentale. L'amplitude et la fréquence de chacun des partiels peut évoluer au cours du temps. Enfin, la distribution de l'énergie sur l'ensemble des partiels, caractérisant plus ou moins le timbre de la source, est inconnue. On veut estimer au cours du temps les caractéristiques fréquentielles et énergétiques de chacun des partiels des sources détectées dans le signal.

Il est clair que ce problème est assez général et nous avons vu que sa résolution est loin d'être évidente. Le choix de s'y attaquer sans poser plus d'hypothèses restrictives résulte de la volonté de pouvoir traiter un large spectre de signaux. Il va aussi de soi que le champ d'action de la méthode proposée doit pouvoir être plus ciblé en fonction de la connaissance dont on dispose sur les sources. Par exemple, si l'algorithme est utilisé pour faire la transcription automatique d'un morceaux de piano, l'instrument peut être modélisé précisément : son inharmonicité est assez bien connue ainsi que le profil de décroissance de l'énergie des notes jouées. D'une manière générale, le principe essentiel du formalisme bayésien s'applique parfaitement à



Calcul du spectrogramme : passage du domaine temporel au plan temps-fréquence.



Détection et suivi des trajectoires temps-fréquence, à l'image du suivi de cibles dans un espace à plusieurs dimensions.

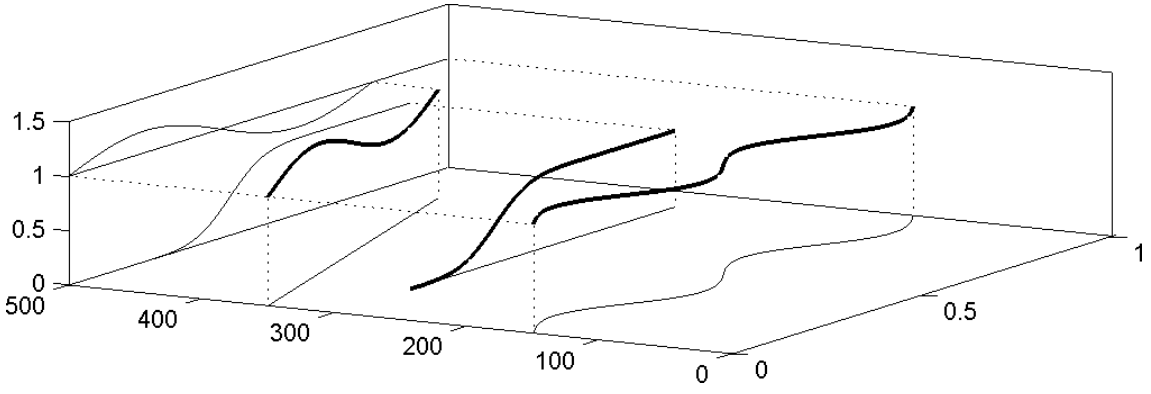


FIG. IV.1 – Illustration du principe d'extraction de trajectoires temps-fréquence.

notre problème : plus on a de connaissances *a priori* sur les sources et plus l'estimation pourra être robuste et précise. La durée des signaux traités étant inconnue (à terme, on veut pouvoir traiter en temps réel un flux audio continu), on souhaite effectuer une estimation séquentielle de tous les paramètres caractéristiques du signal. C'est donc d'eux-mêmes que les principes et les algorithmes du filtrage bayésien s'imposent pour apporter une solution à notre problème. L'idée est de considérer chaque source comme une cible en mouvement dans un espace à plusieurs dimensions. Chacune d'entre elles est décrite par un vecteur de paramètres dont il est possible de donner une équation d'évolution au cours du temps. Comme les seules données acquises dont on dispose sont le signal temporel, un vecteur d'observations de nos paramètres en est extrait à chaque instant. L'analogie avec l'approche décrite dans le chapitre précédent est alors flagrante et peut être illustrée par la figure IV.1. Une équation non linéaire (spectrogramme) relie les observations (signal temporel) aux paramètres (fréquences et amplitudes). Dans cet espace prenant en compte l'axe temporel, les paramètres évoluent selon des trajectoires qui peuvent être suivies au cours du temps. L'apparition ou la disparition d'une source dans le signal audio est assimilée à l'apparition ou à la disparition d'une trajectoire dans l'espace. Comme nous le verrons dans la suite, les non linéarités impliquées dans les équations de transition et d'observation, justifient pleinement l'utilisation des méthodes de Monte Carlo et, plus particulièrement, des algorithmes de filtrage particulaire, présentées dans le chapitre précédent.

Ce chapitre est organisé de la manière suivante. Le modèle bayésien séquentiel est défini dans la première partie. Les équations de transition et d'observation y sont notamment introduites sous leur forme la plus générale. Dans la partie suivante, trois mises en œuvre possibles de l'algorithme sont détaillées. Les différences résultent de la diversité des choix offerte par le domaine de validité des méthodes de Monte Carlo. Ces trois versions sont ensuite étudiées sur un exemple jouet avant de finir par l'étude de signaux réels.

IV.1 Notations et modèle

Le modèle adopté suit l'approche classique par fenêtre glissante. Plus précisément, soit $\mathbf{x} = [x_1, x_2, \ldots, x_t, \ldots]$, $t \in \mathbb{N}^{+*}$, un signal à temps discret. L'étude est rendue locale autour de l'instant t, en considérant une fenêtre d'analyse $\mathbf{w}_t$, de longueur $L_{\mathbf{w}}$ et centrée sur l'instant t (voir figure II.2, page 17). Le signal fenêtré $\mathbf{x}_t$ est supposé stationnaire par rapport à la longueur de la fenêtre, ce qui signifie que les évolutions fréquentielles et énergétiques de $\mathbf{x}$ se font sur une durée dont l'ordre de grandeur est supérieur à $L_{\mathbf{w}}$. Cette hypothèse est assez classique et est sous-jacente à toute méthode basée sur une analyse à fenêtre glissante. Il est modélisé par la somme de K_t composantes harmoniques stationnaires. On désigne par composante harmonique l'ensemble des composantes sinusoïdales (partiels) appartenant à une même structure fréquentielle. Ici, le terme harmonique est employé dans une acception générale, les fréquences n'étant pas forcément des multiples entiers de la fréquence fondamentale. La composante harmonique k, $k = 1, \ldots, K_t$, est constituée de H partiels dont la fréquence est notée $f_{t,k,h}$, $h = 1, \ldots, H$. Son fondamental a donc une fréquence notée $f_{t,k,1}$. La valeur de H n'est pas explicitement estimée, elle est fixée *a priori*. Ceci répond à un aspect pratique de l'implémentation de l'algorithme et n'est pas limitatif dans la mesure où H sera toujours choisi grand (l'algorithme cherchera donc à estimer plus de partiels que les sources en contiennent réellement). Deux modèles sinusoïdaux sont

envisageables pour les partiels. Le premier est de la forme :

$$A\cos(2\pi ft+\varphi) = a\cos(2\pi ft) + b\sin(2\pi ft) \tag{IV.1}$$

où A est une amplitude strictement positive et φ est la phase initiale. L'expression de droite est néanmoins celle qui est utilisée car elle n'est non linéaire que par rapport à la fréquence. Dans le second modèle, la phase initiale est négligée :

$$A\cos(2\pi ft) \tag{IV.2}$$

Ainsi, en fonction du modèle choisi, il faudra estimer, en plus des fréquences $f_{t,k,h}$, les amplitudes $a_{t,k,h}$ et $b_{t,k,h}$ ou les amplitudes $A_{t,k,h}$, $k = 1,\dots,K_t$ et $h = 1,\dots,H$. On désigne par $\mathbf{f}_t$ le vecteur contenant toutes les fréquences :

$$\mathbf{f}_t = \left[f_{t,1,1}\,,\dots,\,f_{t,k,h}\,,\dots,\,f_{t,K_t,H}\right]^{\mathrm{T}} \tag{IV.3}$$

Les vecteurs $\mathbf{a}_t$ et $\mathbf{b}_t$ ou $\mathbf{A}_t$ sont construits sur le même principe. Le problème de filtrage porte donc sur l'estimation séquentielle du paramètre discret K_t et du vecteur continu $\boldsymbol{\theta}_t = [\mathbf{f}_t, \mathbf{a}_t, \mathbf{b}_t]$ (ou $\boldsymbol{\theta}_t = [\mathbf{f}_t, \mathbf{A}_t]$, en fonction du modèle choisi). La dimension de $\boldsymbol{\theta}_t$ dépend du temps et est égale, au plus, à $3K_tH$ (ou $2K_tH$).

IV.1.1 Equation d'observation

Dans la figure IV.1, le vecteur d'observations à l'instant t est construit à partir du signal fenêtré par :

$$\mathbf{y}_t = \left|\mathrm{DFT}(\mathbf{x}_t)\right|^2 \tag{IV.4}$$

$$= \left|\mathrm{DFT}(\mathbf{x}\cdot\mathbf{w}_t)\right|^2 \tag{IV.5}$$

où DFT désigne la transformée de Fourier discrète. Il est clair que le spectrogramme est insensible à la phase initiale des composantes et qu'un modèle sinusoïdal du type de l'équation (IV.1) n'est pas envisageable. On adopte alors le second modèle et on définit la fonction h_t par :

$$\begin{array}{rccc} h_t: & \mathbb{N}\times\mathbb{R}^{2K_tH} & \longrightarrow & \mathbb{R}^{L_{\mathbf{w}}} \\ & (K_t,\boldsymbol{\theta}_t) & \longmapsto & \left|\mathrm{DFT}(\mathbf{s}_t\cdot\mathbf{w}_t)\right|^2 \end{array} \tag{IV.6}$$

avec, pour $\tau = t+1-\frac{L_{\mathbf{w}}}{2},\dots,t+\frac{L_{\mathbf{w}}}{2}$:

$$\mathbf{s}_t[\tau] = \sum_{k=1}^{K_t}\sum_{h=1}^{H} A_{t,k,h}\cos(2\pi f_{t,k,h}\tau) \tag{IV.7}$$

Le passage par le plan temps-fréquence, comme espace de représentation du signal, permet d'obtenir un premier algorithme d'estimation séquentielle de la fréquence fondamentale [Dub05c, Dub05a]. Cependant, si le spectrogramme a permis l'émergence de l'idée de suivi de trajectoires

temps-fréquence (voir figure IV.1), son utilisation n'est pas indispensable et la perte de l'information de phase peut être évitée en restant dans le domaine temporel [Dub05b, Dub]. Le vecteur d'observations à l'instant t est alors pris égal au signal fenêtré :

$$\begin{aligned} \mathbf{y}_t &= \mathbf{x}_t & \text{(IV.8)} \\ &= \mathbf{x} \cdot \mathbf{w}_t & \text{(IV.9)} \end{aligned}$$

et la fonction h_t devient, en adoptant, cette fois-ci, le modèle sinusoïdal de l'équation (IV.1) :

$$\begin{array}{rccc} h_t : & \mathbb{N} \times \mathbb{R}^{3K_tH} & \longrightarrow & \mathbb{R}^{L_\mathbf{w}} \\ & (K_t, \boldsymbol{\theta}_t) & \longmapsto & \mathbf{s}_t \cdot \mathbf{w}_t \end{array} \qquad \text{(IV.10)}$$

avec, pour $\tau = t + 1 - \frac{L_\mathbf{w}}{2}, \ldots, t + \frac{L_\mathbf{w}}{2}$:

$$\mathbf{s}_t[\tau] = \sum_{k=1}^{K_t} \sum_{h=1}^{H} a_{t,k,h} \cos(2\pi f_{t,k,h}\tau) + b_{t,k,h} \sin(2\pi f_{t,k,h}\tau) \qquad \text{(IV.11)}$$

Quels que soient la définition de la fonction h_t et le choix du modèle sinusoïdal pour les partiels, l'équation d'observation est de la forme :

$$\mathbf{y}_t = h_t(K_t, \boldsymbol{\theta}_t) + \mathbf{v}_t^\mathbf{y} \qquad \text{(IV.12)}$$

où $\mathbf{v}_t^\mathbf{y}$ est un vecteur aléatoire gaussien de moyenne nulle et de matrice de covariance $r^\mathbf{y}\mathbf{I}_{L_\mathbf{w}}$, $\mathbf{I}_{L_\mathbf{w}}$ étant la matrice identité de dimensions $L_\mathbf{w} \times L_\mathbf{w}$.

IV.1.2 Equations de transition

Comme nous l'avons dit au début de ce chapitre, le suivi des composantes harmoniques est comparable au suivi de cibles, dont le vecteur d'état est classiquement composé de leurs coordonnées spatiales et de la vitesse. Dans ce cas, les équations d'évolution entre deux instants consécutifs sont données par les lois de la physique. Elles sont faciles à obtenir lorsque les mouvements possibles de la cible, c'est-à-dire en conformité avec ses propriétés dynamiques, sont connus. En musique, de telles équations pourraient être obtenues pour un instrument donné, utilisé dans des conditions maîtrisées. Par exemple, en considérant la guitare, la physique ondulatoire permet de décrire le mécanisme de vibration de la corde. Dans le cas général, il est difficile de considérer un modèle applicable à tout type de signaux, pour caractériser l'évolution de K_t et $\boldsymbol{\theta}_t$ au cours du temps. Les équations de transition adoptées se basent alors sur l'hypothèse heuristique que les fréquences et les amplitudes ont des trajectoires assez lisses et régulières. En d'autres termes, on considère qu'il n'y a pas de changement brusque au sein de l'évolution des paramètres d'une composante harmonique et que de tels changements dans le signal, reflètent l'apparition ou la disparition d'une ou plusieurs composantes harmoniques. Les trois sections suivantes détaillent l'application de cette hypothèse pour définir la forme générale des équations de transition du nombre de composantes, des fréquences et des amplitudes.

	$K_{t-1} = K_{min}$	$K_{min} < K_{t-1} < K_{max}$	$K_{t-1} = K_{max}$
b_t (%)	80	10	0
e_t (%)	80	80	80
d_t (%)	0	10	20

TAB. IV.1 – Exemple de valeurs, pour les probabilités de transition du nombre de composantes harmoniques. Les probabilités b_t, e_t et d_t sont définies dans l'équation (IV.13).

IV.1.2.a Nombre de composantes

L'évolution du paramètre discret K_t est modélisée par une chaîne de Markov dont les probabilités de transition sont connues :

$$K_t = K_{t-1} + \begin{cases} 1 & \text{avec la probabilité } b_t \\ 0 & \text{avec la probabilité } e_t \\ \text{-1} & \text{avec la probabilité } d_t \end{cases} \tag{IV.13}$$

La valeur précise des probabilités b_t, e_t et d_t n'a pas une grande importance. Le principe de cette modélisation est de permettre une augmentation ou une diminution du nombre de composantes harmoniques, tout en privilégiant le cas où il reste constant. Cela se traduit par une supériorité relative de la probabilité e_t par rapport à b_t et d_t. Ce choix reprend l'hypothèse heuristique que le nombre de composantes harmoniques est, la plupart du temps, supposé constant. Cela ne veut pas dire que l'évolution de K_t est limitée, mais signifie simplement que l'on considère que la durée de vie d'une composante harmonique est d'un ordre de grandeur supérieur à l'écart entre deux instants d'analyse. De même, la probabilité b_t peut être prise inférieure à d_t, afin de favoriser un faible nombre de composantes harmoniques. D'autre part, pour des raisons pratiques, il est nécessaire de borner les valeurs possibles pour K_t. Dans la suite, elles sont supposées comprises entre une valeur minimale K_{min} (qui peut être nulle) et une valeur maximale K_{max}. Un exemple de l'ordre de grandeur des probabilités b_t, e_t et d_t est donné dans le tableau IV.1.

Dans l'équation (IV.13), l'amplitude de l'évolution de K_t est fixée à 1. Il est important de noter que ce choix n'est pas limitatif. En particulier, il n'est pas nécessaire d'autoriser des évolutions d'amplitude supérieure, pour gérer le cas de l'apparition ou de la disparition simultanée de plusieurs composantes. Afin de mettre ceci en évidence, prenons un exemple. Supposons que les histogrammes de la distribution *a posteriori* du nombre de composantes harmoniques, aux instants $t-1$ et t, sont ceux donnés dans la figure IV.2. Des évolutions d'amplitude 1 peuvent parfaitement expliquer l'histogramme à l'instant t, étant donné celui à l'instant $t-1$, alors que les estimations par maximum *a posteriori* du nombre de composantes harmoniques, aux instants $t-1$ et t, sont respectivement $\hat{K}_{t-1} = 5$ et $\hat{K}_t = 2$. En effet, pour obtenir l'histogramme de droite, dans la figure IV.2, un ensemble de 10 000 échantillons a été généré à partir de l'histogramme de gauche. Chaque échantillon a ensuite été mis à jour en utilisant l'équation (IV.13) et en favorisant une diminution du nombre de composantes harmoniques, information qui peut être apportée par le nouveau vecteur d'observations $\mathbf{y}_t$, à l'instant t.

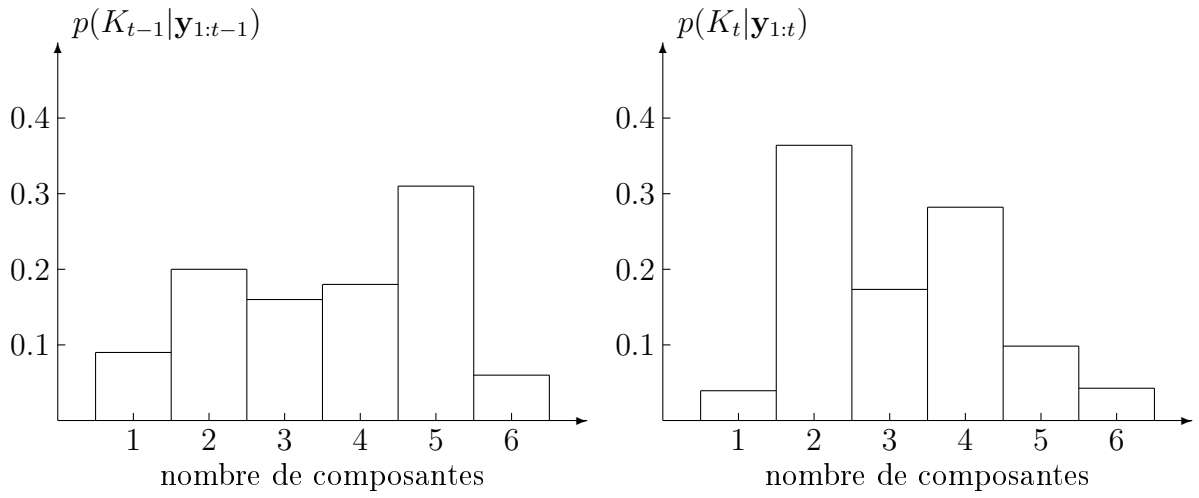


FIG. IV.2 – Exemple d'évolution de la distribution *a posteriori* du nombre de composantes harmoniques.

IV.1.2.b Fréquences

Avant de donner l'équation de transition des fréquences, il est nécessaire de préciser quel peut être le contenu du vecteur $\mathbf{f}_t$, définit par l'équation (IV.3). En effet, plusieurs cas de figure peuvent être envisagés, en fonction de la connaissance disponible sur le signal traité.

Pour les signaux parfaitement harmoniques, la fréquence des partiels de rang $h \geq 2$ est donnée par $f_{t,k,h} = hf_{t,k,1}$. Dans ce cas précis, il n'est donc pas nécessaire d'estimer ces fréquences et le vecteur $\mathbf{f}_t$ n'est composé que des K_t fréquences fondamentales $f_{t,k,1}$, $k = 1, \ldots, K_t$. En pratique, ce modèle n'est utilisable que dans un nombre restreint de situations (par exemple, la partie voisée des signaux de parole est reconnue comme étant harmonique). D'une manière générale, les sources considérées sont supposées inharmoniques. Ici, plusieurs solutions sont possibles pour prendre en compte cette inharmonicité, toujours en fonction de la connaissance que nous avons du signal. Si un modèle d'inharmonicité est disponible, comme par exemple pour le piano, équation (I.5), ou pour les instruments à cordes pincées [Kla03], pour lesquels il existe une relation du type $f_{t,k,h} = hf_{t,k,1}\sqrt{1 + (h^2 - 1)\gamma_{t,k}}$, alors, il n'est toujours pas nécessaire d'estimer toutes les fréquences des partiels de rang $h \geq 2$. Par rapport au cas précédent, il suffit d'ajouter dans le vecteur $\mathbf{f}_t$, les K_t paramètres d'inharmonicité $\gamma_{t,k}$. Cependant, dans un contexte moins particulier, il sera nécessaire d'estimer la fréquence de chacun des partiels. L'approche la plus simple est certainement d'effectuer cette estimation indépendamment et le vecteur $\mathbf{f}_t$ correspond alors à l'équation (IV.3). Une solution néanmoins plus robuste est de modéliser la fréquence des partiels de rang $h \geq 2$ par $f_{t,k,h} = c_{t,k,h}hf_{t,k,1}$. Dans ce cas, il faut estimer la fréquence du fondamental et les coefficients correcteurs $c_{t,k,h}$, $h \geq 2$ ($c_{t,k,1} = 1$) qui permettent de s'écarter de la fréquence harmonique. Il faut cependant noter que si ces coefficients n'appartiennent pas à un intervalle borné, cette modélisation est équivalente au cas simple précédent.

La méthode exposée dans ce chapitre peut prendre en compte facilement toutes ces solu-

tions et, pour des commodités de notation, les composantes du vecteur $\mathbf{f}_t$, qui peuvent être les fréquences des partiels, les paramètres d'inharmonicité ou les coefficients correcteurs, seront dans la suite désignées par $f_{t,k,h}$, en accord avec l'équation (IV.3), le contexte permettant de savoir à quoi cela renvoie. La dimension de $\mathbf{f}_t$ n'est pas fixe et peut être, dans le cas où toutes les fréquences sont estimées, égale à $K_t H$.

La forme générale de l'équation de transition des fréquences, entre les instants $t-1$ et t, est, pour $k = 1, \ldots, K_t$ et $h = 1, \ldots, H$:

$$f_{t,k,h} = g_T^{\mathbf{f}}(f_{t-T:t-1,k,h}) + v_{t-1,k,h}^{\mathbf{f}} \tag{IV.14}$$

avec $g_T^{\mathbf{f}}$ une fonction de lissage qui calcule une valeur pour la fréquence à l'instant t, en prenant en compte les $T > 0$ valeurs passées. Afin de conserver le caractère markovien, dans le cas où $T \neq 1$, il est nécessaire de redéfinir le vecteur d'état, en y incluant les T valeurs passées. La fonction $g_T^{\mathbf{f}}$ peut être définie par un modèle auto-régréssif [Dav98], encore appelé modèle de prédiction linéaire, d'ordre T. Elle peut aussi prendre en compte des caractéristiques comme la dérivée des paramètres, pour utiliser, par exemple, la formulation bayésienne séquentielle des fonctions splines [Wah78, Wah83]. A l'opposé, prendre $T = 1$ et $g_T^{\mathbf{f}} = \mathbf{1}$ (la fonction identité), revient à considérer une simple marche aléatoire comme équation de transition. Là encore, la flexibilité de l'algorithme permet l'utilisation d'un large panel de fonction. Dans l'équation (IV.14), $v_{t-1,k,h}^{\mathbf{f}}$ est un bruit gaussien de moyenne nulle et de variance $r_{t-1,k,h}^{\mathbf{f}}$. Dans le modèle, la variance $r_{t-1,k,h}^{\mathbf{f}}$ peut évoluer au cours du temps, selon l'équation :

$$\log\left(r_{t,k,h}^{\mathbf{f}}\right) = \log\left(r_{t-1,k,h}^{\mathbf{f}}\right) + \varphi_{t-1,k,h} \tag{IV.15}$$

avec

$$\varphi_{t-1,k,h} \sim \mathcal{N}(0, \sigma_\varphi^2) \tag{IV.16}$$

Autoriser les variances à évoluer au cours du temps permet au modèle de s'adapter aux différentes situations, l'évolution des paramètres pouvant être stationnaire ou plus dynamique (comme lorsque l'on a une modulation). Dans ce dernier cas, la rapidité des changements est considérée à l'échelle de l'analyse par fenêtre glissante, c'est-à-dire qu'elle reste d'un ordre de grandeur supérieur à $L_{\mathbf{w}}$. D'un point de vue théorique, la diminution de la variance en cas de stabilisation de l'évolution est assurée par la distribution *a posteriori*. En effet, nous avons vu que les algorithmes de filtrage particulaire favorisent les régions dans lesquelles cette distribution prend des valeurs élevées, ce qui se traduit par une diminution de la valeur de la variance. Or, cela n'est possible que lorsque la vraisemblance et la distribution *a priori* coïncident et ceci de manière durable dans le temps. Cette situation correspond bien à une stabilisation de l'évolution du paramètre considéré. De plus, la diminution de la variance permet aussi de rendre l'estimation des paramètres plus précise au fur et à mesure de l'établissement de la stationnarité. Cette étape de stabilisation apparaît très souvent dans l'évolution des paramètres, car elle correspond au régime libre de la source, qui arrive après le régime forcé, à l'origine de l'excitation sonore. Par exemple, lorsqu'un musicien joue de la guitare, après l'étape de pincement de la corde, cette dernière oscille librement à une fréquence qui n'évolue pas au cours du temps. La loi d'évolution des variances porte sur leur logarithme, afin de garantir la positivité de $r_{t,k,h}^{\mathbf{f}}$. Enfin, la valeur de σ_φ n'a pas besoin d'être ajustée finement. En choisissant $\sigma_\varphi = 0.35 \approx \ln(2)/2$, 95% des valeurs de $\exp(\varphi_{t-1,k,h})$ appartiendront à l'intervalle $[0.5, 2]$, ce qui signifie que $r_{t-1,k,h}^{\mathbf{f}}$ peut

être doublée ou divisée par deux avec une probabilité non négligeable, ce qui offre une large possibilité d'évolution pour la variance.

IV.1.2.c Amplitudes

Comme il l'a été dit au début du chapitre II, un algorithme d'estimation de fréquences fondamentales ne se cantonne pas à l'estimation de ces dernières mais il doit aussi considérer, d'une manière ou d'une autre, les autres paramètres caractérisant la source : la structure fréquentielle et les amplitudes des partiels. Dans la partie précédente, plusieurs solutions ont été exposées pour prendre en compte facilement la structure fréquentielle. De même, en ce qui concerne les amplitudes, plusieurs manières de procéder peuvent être distinguées, que ce soit pour le choix des paramètres à estimer ou la définition de l'équation de transition. A l'instar de la fréquence, il n'est pas forcément nécessaire d'estimer les amplitudes de chacun des partiels. En effet, on pourrait considérer un modèle paramétrique du profil de décroissance de l'énergie en fonction du rang du partiel, il suffirait alors d'estimer les paramètres de ce modèle. Cependant, si un tel modèle peut être utilisé dans certains cas précis, il est difficile d'en choisir un assez général pouvant s'appliquer à différents types de signaux. Nous avons donc opté pour l'estimation de l'amplitude de tous les partiels, avec néanmoins la possibilité d'introduire, par exemple, une dépendance entre les amplitudes de deux partiels consécutifs.

Pour ce qui est de l'équation de transition, nous pouvons lui donner la même forme générale que celle des fréquences. Pour $k = 1, \ldots, K_t$ et $h = 1, \ldots, H$, on a :

$$a_{t,k,h} = g_T^{\mathbf{a}}(a_{t-T:t-1,k,h}) + v_{t-1,k,h}^{\mathbf{a}} \quad \text{(IV.17)}$$
$$b_{t,k,h} = g_T^{\mathbf{b}}(b_{t-T:t-1,k,h}) + v_{t-1,k,h}^{\mathbf{b}} \quad \text{(IV.18)}$$

$$\log\left(r_{t,k,h}^{\mathbf{a}}\right) = \log\left(r_{t-1,k,h}^{\mathbf{a}}\right) + \alpha_{t-1,k,h} \quad \text{(IV.19)}$$
$$\log\left(r_{t,k,h}^{\mathbf{b}}\right) = \log\left(r_{t-1,k,h}^{\mathbf{b}}\right) + \beta_{t-1,k,h} \quad \text{(IV.20)}$$

$$\alpha_{t-1,k,h} \sim \mathcal{N}(0, \sigma_\alpha^2) \quad \text{(IV.21)}$$
$$\beta_{t-1,k,h} \sim \mathcal{N}(0, \sigma_\beta^2) \quad \text{(IV.22)}$$

ou, si le second modèle sinusoïdal est choisi

$$A_{t,k,h} = g_T^{\mathbf{A}}(A_{t-T:t-1,k,h}) + v_{t-1,k,h}^{\mathbf{A}} \quad \text{(IV.23)}$$

$$\log\left(r_{t,k,h}^{\mathbf{A}}\right) = \log\left(r_{t-1,k,h}^{\mathbf{A}}\right) + \Lambda_{t-1,k,h} \quad \text{(IV.24)}$$

$$\Lambda_{t-1,k,h} \sim \mathcal{N}(0, \sigma_\Lambda^2) \quad \text{(IV.25)}$$

Comme pour σ_φ, la valeur précise de σ_α, σ_β ou σ_Λ n'a pas tellement d'importance et nous pouvons la prendre égale à 0.35, pour les mêmes raisons que précédemment.

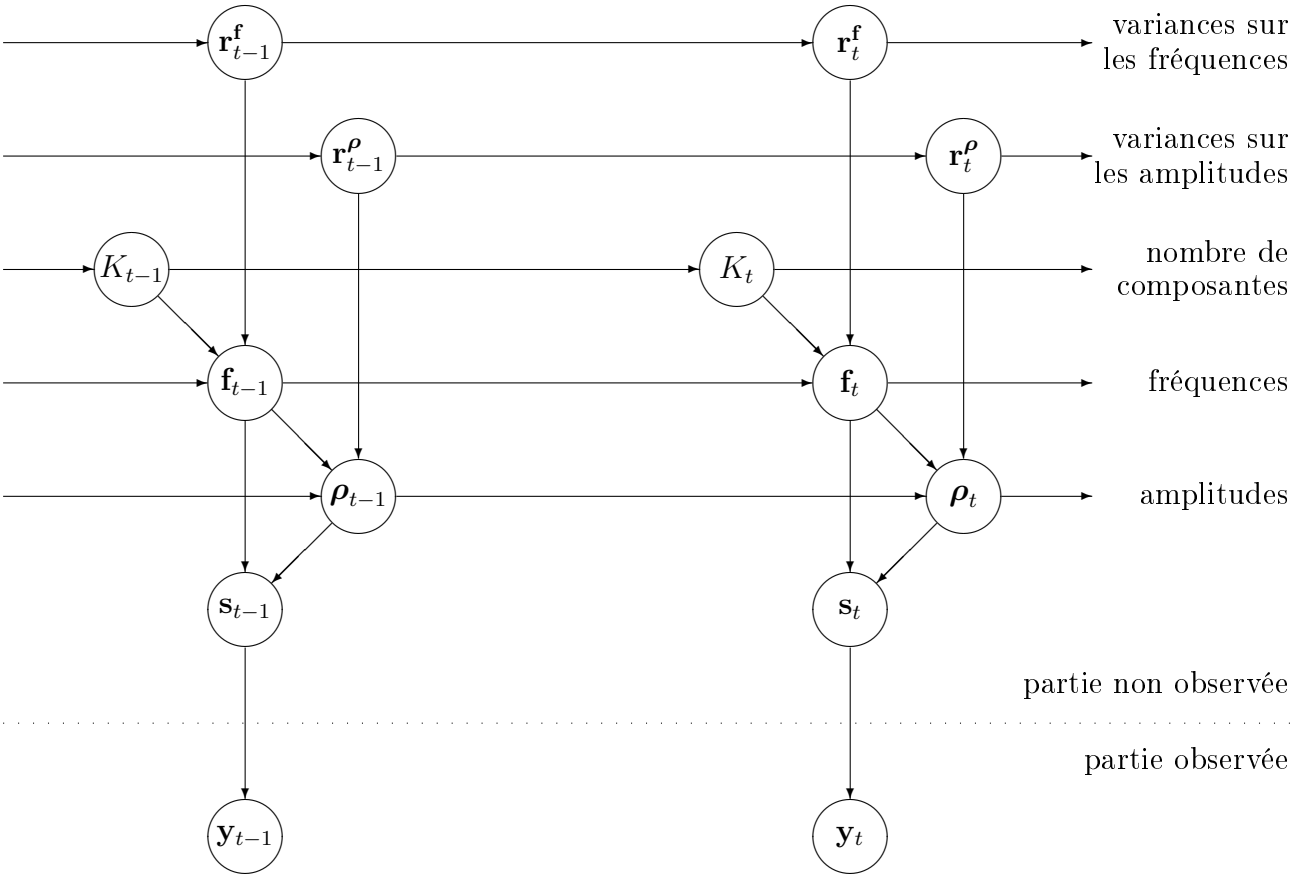


FIG. IV.3 – Représentation du modèle séquentiel global sous forme de graphe. La notation $\boldsymbol{\rho}$ désigne les amplitudes $[\mathbf{a}, \mathbf{b}]$ ou $\mathbf{A}$, en fonction du modèle sinusoïdal choisi.

IV.2 Trois implémentations de l'algorithme

Les équations de transition et d'observation qui viennent d'être présentées, l'ont été sous leur forme la plus générale. Il a aussi été mentionné que différents choix pouvaient être faits quant aux paramètres à estimer. L'intention derrière cet exposé généraliste, est de mettre en avant la flexibilité offerte par les algorithmes de filtrage particulaire. Le modèle séquentiel global est résumé, sous forme de graphe, dans la figure IV.3. Chaque flèche reprend une équation définie dans la partie précédente. La représentation graphique permet aussi de mettre en évidence la relation hiérarchique qui existe, d'une part, entre le nombre de composantes harmoniques et les fréquences et, d'autre part, entre les fréquences et les amplitudes. En effet, avant d'estimer les fréquences, il faut savoir combien il y en a. De même, il faut savoir quels sont les partiels dont il faut estimer les amplitudes.

Depuis le début de ce chapitre, nous construisons petit à petit le modèle et l'algorithme utilisé pour estimer ses paramètres. Nous avons aussi évoqué les multiples choix possibles, à divers niveaux. En fait, ces choix ont une influence les uns sur les autres et il devient maintenant nécessaire de quitter un peu la généralité pour en expliciter quelques-uns. Dans cette partie, nous allons présenter trois mises en œuvre différentes de l'algorithme de filtrage particulaire :

une dans laquelle le vecteur d'observations est dans le plan temps-fréquence et deux dans lesquelles il est dans le domaine temporel. Cette première différence, portant sur le choix du vecteur d'observations, à une conséquence directe sur le modèle sinusoïdal utilisable. Elle induit aussi un changement dans la manière d'estimer les amplitudes, étant donné les fréquences et le nombre de composantes. En effet, en choisissant le domaine temporel et en considérant certaines équations de transition sur les amplitudes, l'estimation de ces dernières peut être plus ou moins découplée de celle des fréquences. Or, nous avons vu dans le chapitre précédent, qu'un tel découplage dans le vecteur d'état, permettait d'opter pour des solutions algorithmiques plus performantes.

Dans la suite, après avoir présentée la structure générale de l'algorithme, chacune des versions est détaillée.

IV.2.1 Points communs

Les trois algorithmes présentés dans la suite ont la même structure générale, donnée dans le tableau IV.2. La mention *voir version* signifie que cette étape particulière de l'algorithme dépend de la version choisie. Elle sera donc décrite dans les sections suivantes. Ici, nous nous intéressons d'abord aux points communs aux trois versions, et plus particulièrement à la fonction de proposition des fréquences (notée fpf dans l'algorithme), à l'étape de mise à jour des hyperparamètres portant sur les fréquences et à la sortie de l'algorithme à chaque itération.

Fonction de proposition des fréquences

Cette fonction, notée fpf, sert à calculer, à un instant t donné, un vecteur de fréquences[1] $\mathbf{f}_t^{\mathbf{y}}$ à partir du vecteur d'observations $\mathbf{y}_t$. Elle peut être définie par n'importe qu'elle procédure *ad hoc* ou encore être l'implémentation d'une des nombreuses méthodes exposées dans le chapitre II, le principe étant d'essayer d'extraire les fréquences fondamentales présentes dans $\mathbf{y}_t$. Cet emboîtement d'une méthode d'estimation de fréquences fondamentales au sein d'un algorithme qui a le même objectif, peut sembler paradoxal. En fait, le vecteur $\mathbf{f}_t^{\mathbf{y}}$ est utilisé pour initialiser les particules et pour proposer une nouvelle composante harmonique, lorsque K_t augmente. Il peut aussi servir à construire (nous verrons comment plus loin) la densité d'importance nécessaire à tout algorithme de filtrage particulaire. De ce fait, les hypothèses nécessaires à son calcul, ne se transmettent pas à l'algorithme général et les erreurs commises seront corrigées par l'estimation séquentielle.

Nous présentons ici, la construction de la fonction fpf utilisée dans la suite de ce manuscrit. Elle cherche à estimer K_{max} composantes harmoniques, dans le signal fenêtré $\mathbf{x}_t$, servant à la construction de $\mathbf{y}_t$. Elle part du principe que chaque pic de la transformée de Fourier de $\mathbf{x}_t$, correspond potentiellement à un partiel. Elle commence donc par extraire la fréquence de tous les pics. Vient ensuite une étape itérative, où toutes les fréquences de la liste obtenue, étant à peu près en relation harmonique avec une autre fréquence de cette liste, sont supprimées. Plus précisément, notons ν_1 la plus petite fréquence de la liste. Cette fréquence est mise de côté et supprimée de la liste. Pour $h = 2, 3, \ldots$, la fréquence de la liste la plus proche de $h\nu_1$, et

[1] Il faut rappeler que les composantes du vecteur $\mathbf{f}$ ne sont pas forcément des fréquences mais peuvent être des coefficients correcteurs ou des paramètres d'inharmonicité (voir section IV.1.2.b).

A l'instant $t = 0$
- **Initialisation**
  - Pour $i = 1, \ldots, N$
    - Générer $K_0^{(i)} \sim \mathcal{U}([K_{min}, K_{max}])$
    - Initialiser $\boldsymbol{\theta}_0^{(i)}$ en utilisant fpf
    - Poser $(\mathbf{r}_0^{\boldsymbol{\theta}})^{(i)} = \mathbf{r}_{ini}^{\boldsymbol{\theta}}$

A l'instant $t \geq 1$
- Calculer le vecteur d'observations (voir version)
- **Echantillonnage d'importance séquentiel**
  - Pour $i = 1, \ldots, N$
    - Générer $K_t^{(i)}$ avec l'équation de transition : $K_t^{(i)} = K_{t-1}^{(i)} + v_t^K$
    - Si $v_t^K = 1$ alors ajouter une composante avec fpf
    - Si $v_t^K = -1$ alors enlever une composante choisie aléatoirement
    - Mise à jour des hyper-paramètres
      - sur les fréquences : générer $(\mathbf{r}_t^{\mathbf{f}})^{(i)} \sim q_t(\mathbf{r}_t^{\mathbf{f}} | (\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)}, \mathbf{y}_t)$
      - sur les amplitudes (voir version)
    - Mise à jour des paramètres
      - fréquences (voir version)
      - amplitudes (voir version)
    - Calcul des poids non normalisés (voir version)
  - Normalisation des poids
    - Pour $i = 1, \ldots, N$
      - $\tilde{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{i=1}^N \omega_t^{(i)}}$
- **Rééchantillonnage**
  - Dupliquer/supprimer les particules en fonction de $\tilde{\omega}_t^{(i)}$
- **Sortie**
  - Nombre de composantes : calcul de $\hat{K}_t$ par l'équation (IV.29)
  - Fréquences : calcul de $\hat{\mathbf{f}}_t$ par l'équation (IV.30)
  - Amplitudes (voir version)

TAB. IV.2 – Structure générale des algorithmes d'estimation de fréquences fondamentales.

étant dans l'intervalle $[h\nu_1 - h\frac{\delta}{2}, h\nu_1 + h\frac{\delta}{2}]$, est aussi supprimée. La largeur de l'intervalle autour de la fréquence $h\nu_1$ grandit avec h, afin de prendre en compte une éventuelle inharmonicité dans la structure fréquentielle. Elle est néanmoins bornée, afin de rester inférieure à $\frac{\nu_1}{2}$. Ce processus est itéré, en considérant la nouvelle fréquence la plus petite restant dans la liste, jusqu'à ce qu'il n'y ait plus de fréquence dans la liste. Les fréquences ainsi successivement mises de côté, sont de bons candidats de fréquences fondamentales. Un inconvénient de cette procédure, est qu'elle est incapable de proposer des fréquences fondamentales multiples l'une de l'autre (par exemple, en relation d'octave). Pour palier à ce problème, il suffit d'agrandir la liste des fréquences candidates en y ajoutant des multiples entiers de ces fréquences. Cet ajout peut être systématique ou basé sur l'amplitude relative des partiels d'une composante harmonique candidate. Enfin, le nombre de fréquences fondamentales ainsi obtenues est souvent, sinon toujours, supérieur à K_{max}. Il faut donc faire un choix pour construire le vecteur $\mathbf{f}_t^{\mathbf{y}}$. Il se fait simplement en pondérant chaque fréquence fondamentale par la somme des amplitudes de chacun des partiels lui correspondant. L'amplitude est calculée en faisant le produit scalaire entre $\mathbf{x}_t$ et le partiel considéré. Les fréquences finalement choisies sont les K_{max} premières, après les avoir ordonnées par ordre décroissant de poids. Cette procédure peut être légèrement modifiée si on veut, en sortie, non pas la structure fréquentielle, mais les coefficients correcteurs ou les paramètres d'inharmonicité. Un vecteur d'amplitudes peut aussi être construit, sur la base des fréquences proposées.

Mise à jour des hyper-paramètres

Dans le modèle général présenté au début du chapitre, nous avons vu qu'en plus des fréquences et des amplitudes, nous voulons estimer la variance de leurs équations de transition. Ces hyper-paramètres font partie intégrante du vecteur d'état et sont estimés par filtrage particulaire. Leur estimation est prise en compte dans le calcul du poids, par un terme de la forme :

$$\frac{p((\mathbf{r}_t^{\mathbf{f}})^{(i)}|(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)})}{q_t((\mathbf{r}_t^{\mathbf{f}})^{(i)}|(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)}, \mathbf{y}_t)} \tag{IV.26}$$

où le numérateur est la densité *a priori*, donnée par l'équation (IV.15). Reste à définir la densité d'importance q_t. Dans le chapitre précédent, il a été expliqué que cette densité doit prendre en compte la valeur des paramètres à l'instant précédent ainsi que la nouvelle observation obtenue à l'instant courant. Une solution simple pour combiner ces deux informations est de prendre, en posant $\mathbf{R}_t = \log(\mathbf{r}_t)$:

$$q_t(\mathbf{r}_t^{\mathbf{f}}|(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)}, \mathbf{y}_t) = \mathcal{N}(\mathbf{R}_t^{\mathbf{f}}; |\mathbf{f}_{t-1}^{(i)} - \mathbf{f}_t^{\mathbf{y}}|, \sigma_\varphi^2 \mathbf{I}) \tag{IV.27}$$

avec $\mathbf{I}$ est la matrice identité. Une autre possibilité est d'utiliser, en inversant les étapes de mise à jour des paramètres et des hyper-paramètres, dans le tableau IV.2 :

$$q_t(\mathbf{r}_t^{\mathbf{f}}|(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)}, \mathbf{y}_t) = \mathcal{N}(\mathbf{R}_t^{\mathbf{f}}; |\mathbf{f}_{t-1}^{(i)} - \mathbf{f}_t^{(i)}|, \sigma_\varphi^2 \mathbf{I}) \tag{IV.28}$$

Afin de ne pas avoir des valeurs trop petites ou trop grandes, il est nécessaire que les composantes du vecteur $\mathbf{r}_t^{\mathbf{f}}$ soient comprises entre une valeur minimale et une valeur maximale. Ces valeurs ne sont pas forcement égales pour toutes les composantes, la variance pouvant porter sur différentes quantités (fréquences, coefficients correcteurs ou paramètres d'inharmonicité).

Sortie de l'algorithme

Les estimations $\hat{K}_t$ et $\hat{\mathbf{f}}_t$ du nombre de composantes et des fréquences sont calculées à partir de l'approximation particulaire de la distribution *a posteriori* donnée par l'équation (III.59). En particulier, $\hat{K}_t$ est calculé par maximum *a posteriori*, de la manière suivante :

$$\hat{K}_t = \underset{K_{min} \leq k \leq K_{max}}{\operatorname{argmax}} \; N_{t,k} \tag{IV.29}$$

où $N_{t,k}$ est le nombre de particules ayant k composantes harmoniques. L'estimation $\hat{\mathbf{f}}_t$ minimise l'erreur quadratique moyenne et est calculée par, pour $k = 1, \ldots, \hat{K}_t$:

$$\hat{\mathbf{f}}_{t,k} = \frac{1}{N'_{t,k}} \sum_{i=1}^{N'_{t,k}} \mathbf{f}_{t,k}^{(i)} \tag{IV.30}$$

Il convient d'apporter quelques précisions sur le calcul effectué dans l'équation (IV.30). Une fois $\hat{K}_t$ obtenu, on pourrait utiliser les $N_{t,k}$ particules avec $k = \hat{K}_t$ composantes harmoniques, pour calculer $\hat{\mathbf{f}}_t$. Dans ces particules, les fréquences sont d'abord triées et ordonnées selon une référence commune, par exemple $\mathbf{f}_t^{\mathbf{y}}$ ou $\hat{\mathbf{f}}_{t-1}$, afin d'effectuer des moyennes qui ont un sens. En effet, avant l'étape de classement, les composantes harmoniques n'ont aucune raison d'être dans le même ordre, d'une particule à l'autre. Il peut aussi être noté que les particules contenant plus de $\hat{K}_t$ composantes harmoniques peuvent être utilisées pour calculer $\hat{\mathbf{f}}_t$, ce qui permet de prendre en compte plus de particules et donc, d'obtenir une estimation plus précise. Ceci explique le nombre $N'_{t,k}$ dans l'équation (IV.30), au lieu de $N_{t,k}$.

IV.2.2 Dans le plan temps-fréquence

Dans cette première version de l'algorithme, le vecteur d'observations est la colonne du spectrogramme correspondant à l'instant courant. Il est défini dans l'équation (IV.4). La conséquence de ce choix est qu'il n'est pas possible d'estimer la phase initiale. Le modèle sinusoïdal est donc celui de l'équation (IV.2) et les paramètres à estimer sont donc K_t et $\boldsymbol{\theta}_t = [\mathbf{f}_t, \mathbf{A}_t]$. La dimension du vecteur $\mathbf{A}_t$ est $K_t H$, l'amplitude de tous les partiels étant estimée. Les équations de transition choisies sont de simples marches aléatoires, données par :

$$\mathbf{f}_t = \mathbf{f}_{t-1} + \mathbf{v}_t^{\mathbf{f}} \tag{IV.31}$$

$$\mathbf{A}_t = \mathbf{A}_{t-1} + \mathbf{v}_t^{\mathbf{A}} \tag{IV.32}$$

et l'équation d'observation est, rappelons-le :

$$\mathbf{y}_t = h_t(K_t, \boldsymbol{\theta}_t) + \mathbf{v}_t^{\mathbf{y}} \tag{IV.33}$$

La fonction h_t, définie par l'équation (IV.6), est non linéaire par rapport aux fréquences et aux amplitudes. Cela implique que le vecteur $\boldsymbol{\theta}_t$ en entier doit être estimé par filtrage particulaire. C'est la caractéristique principale de cette version de l'algorithme. Il nous faut donc définir une densité d'importance pour $\mathbf{f}_t$ et $\mathbf{A}_t$. En reprenant les algorithmes utilisés dans [Dub05c, Dub05a], nous avons opté, dans cette première version, pour l'utilisation de la transformée sans

parfum et des équations du filtre de Kalman, afin de construire la densité d'importance pour les fréquences. Plus précisément, pour chaque particule i, $i = 1, \ldots, N$, un vecteur $\bar{\mathbf{f}}_t^{(i)}$ et une matrice $\mathbf{P}_t^{(i)}$ sont construits à partir de $\bar{\mathbf{f}}_{t-1}^{(i)}$, $(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)}$ et $\mathbf{y}_t$, en utilisant les équations du filtre de Kalman sans parfum données dans le tableau III.3. Le vecteur $\bar{\mathbf{f}}_0^{(i)}$ est initialisé en étant pris égal à $\mathbf{f}_0^{(i)}$. On obtient donc :

$$q_t(\mathbf{f}_t|\mathbf{f}_{t-1}^{(i)}, \mathbf{y}_t) = \mathcal{N}(\mathbf{f}_t; \bar{\mathbf{f}}_{t-1}^{(i)}, \mathbf{P}_t^{(i)}) \tag{IV.34}$$

En ce qui concerne l'amplitude, la construction de la densité d'importance se base sur le fait que, lorsque deux composantes ne sont pas trop près l'une de l'autre, le spectrogramme peut être considéré comme pratiquement linéaire, ce qui nous permet d'opter pour une solution plus simple que le filtre de Kalman sans parfum. Après avoir généré $\mathbf{f}_t^{(i)}$ selon l'équation (IV.34), un vecteur $\mathbf{A}_t^{\mathbf{y}(i)}$ est construit en calculant les amplitudes correspondant aux fréquences, dans le signal fenêtré. La densité d'importance est alors simplement définie par :

$$q_t(\mathbf{A}_t|\mathbf{A}_{t-1}^{(i)}, \mathbf{y}_t) = \mathcal{N}\left(\mathbf{A}_t; \frac{1}{2}(\mathbf{A}_{t-1}^{(i)} + \mathbf{A}_t^{\mathbf{y}(i)}), (\mathbf{r}_t^{\mathbf{A}})^{(i)}\right) \tag{IV.35}$$

On peut noter que ces deux densités d'importance répondent bien à la nécessité de prendre en compte l'information apportée par le nouveau vecteur d'observations et celle apportée par les paramètres à l'instant précédent. Les hyper-paramètres $(\mathbf{r}_t^{\mathbf{A}})^{(i)}$ sont mis à jour de la même manière que pour les fréquences. Les poids non normalisés se calculent par la formule :

$$\omega_t^{(i)} \propto p(\mathbf{y}_t|K_t^{(i)}, \boldsymbol{\theta}_t^{(i)}) \frac{p(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)})}{q_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(i)}, \mathbf{y}_t)} \frac{p((\mathbf{r}_t^{\boldsymbol{\theta}})^{(i)}|(\mathbf{r}_{t-1}^{\boldsymbol{\theta}})^{(i)})}{q_t((\mathbf{r}_t^{\boldsymbol{\theta}})^{(i)}|(\mathbf{r}_{t-1}^{\boldsymbol{\theta}})^{(i)}, \mathbf{y}_t)} \tag{IV.36}$$

La vraisemblance $p(\mathbf{y}_t|K_t^{(i)}, \boldsymbol{\theta}_t^{(i)})$ est calculée par :

$$p(\mathbf{y}_t|K_t^{(i)}, \boldsymbol{\theta}_t^{(i)}) = \mathcal{N}(\mathbf{y}_t; h_t(K_t^{(i)}, \boldsymbol{\theta}_t^{(i)}), r^{\mathbf{y}}\mathbf{I}_{L_{\mathbf{w}}}) \tag{IV.37}$$

$$= \frac{1}{(2\pi r^{\mathbf{y}})^{\frac{L_{\mathbf{w}}}{2}}} \exp\left(-\frac{\left\|\mathbf{y}_t - h_t(K_t^{(i)}, \boldsymbol{\theta}_t^{(i)})\right\|^2}{2r^{\mathbf{y}}}\right) \tag{IV.38}$$

Si le nombre de composantes est nul, le poids devient alors proportionnel à la vraisemblance, qui est définie par :

$$p(\mathbf{y}_t|K_t^{(i)} = 0) = \frac{1}{(2\pi r^{\mathbf{y}})^{\frac{L_{\mathbf{w}}}{2}}} \exp\left(-\frac{\left\|\mathbf{y}_t\right\|^2}{2r^{\mathbf{y}}}\right) \tag{IV.39}$$

Enfin, le calcul de $\hat{\mathbf{A}}_t$ est effectué de la même manière que $\hat{\mathbf{f}}_t$, dans l'équation (IV.30).

IV.2.3 Dans le domaine temporel

Dans les deux autres versions, le vecteur d'observations est directement égal au signal fenêtré, et reste donc dans le domaine temporel. A l'inverse du spectrogramme, il n'y a donc pas de perte d'information, ce qui nous autorise à considérer le modèle sinusoïdal de l'équation(IV.1), qui est le plus complet des deux. Les paramètres à estimer sont donc K_t et $\boldsymbol{\theta}_t = [\mathbf{f}_t, \mathbf{a}_t, \mathbf{b}_t]$. De même

que précédemment, toutes les amplitudes sont estimées et les vecteurs $\mathbf{a}_t$ et $\mathbf{b}_t$ sont, au plus, de dimension $K_t H$. L'équation de transition des fréquences est aussi la même que précédemment et l'équation d'observation a la même forme, la fonction h_t étant, cette fois-ci, définie par l'équation (IV.10). Comme mentionné dans [Dub05b, Dub], un autre avantage de la modélisation dans le domaine temporel, est la possibilité de découpler l'estimation des fréquences de celle des amplitudes, ce qui améliore les résultats de l'algorithme face à certaines difficultés (l'étude comparative de la partie suivante permettra de mieux mettre cela en évidence). Ce découplage peut avoir lieu grâce à la linéarité, conditionnellement à K_t et $\mathbf{f}_t$, de l'équation d'observation, par rapport aux amplitudes. Plus précisément, elle peut se mettre sous la forme :

$$\mathbf{y}_t = \mathbf{C}(K_t, \mathbf{f}_t) \begin{bmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{bmatrix} + \mathbf{v}_t^{\mathbf{y}} \tag{IV.40}$$

où $\mathbf{C}$ est la matrice des fonctions de base, de dimensions $L_{\mathbf{w}} \times 2n$ avec n le nombre total de fréquences[2]. Les n premières colonnes correspondent à la fonction cosinus, les n suivantes, à la fonction sinus. En faisant en plus certains choix d'équation de transition sur les amplitudes, seuls le nombre de composantes harmoniques et les fréquences devront être estimées par filtrage particulaire, l'estimation des amplitudes étant prises en compte par une méthode extérieure ou indirectement. Avant de présenter les deux équations de transition permettant de dériver ces deux versions de l'algorithme, la construction de la densité d'importance $q_t(\mathbf{f}_t|\mathbf{f}_{t-1}^{(i)}, \mathbf{y}_t)$, qui leur est commune, utilisée pour proposer un nouveau vecteur de fréquences à chaque itération, est expliquée.

En fait, cette densité est construite en appliquant simplement le principe, que nous avons exprimé plusieurs fois, stipulant que la densité d'importance doit prendre en compte l'information apportée par le nouveau vecteur d'observations, à l'instant courant et le vecteur de paramètres, à l'instant précédent. Elle utilise le vecteur $\mathbf{f}_t^{\mathbf{y}}$, donné par la fonction de proposition des fréquences fpf, présentée un peu plus tôt, qu'elle combine avec le vecteur $\mathbf{f}_{t-1}^{(i)}$ pour construire une densité gaussienne. On note avec un indice j, la $j^{\text{ème}}$ composante du vecteur correspondant. La densité d'importance pour cette composante est définie par :

$$q_t(f_{t,j}|f_{t-1,j}^{(i)}, \mathbf{y}_t) = \mathcal{N}(f_{t,j}; \lambda_{t,j}^{(i)} f_{t,j}^{\mathbf{y}} + (1 - \lambda_{t,j}^{(i)}) f_{t-1,j}^{(i)}, (r_{t,j}^{\mathbf{f}})^{(i)}) \tag{IV.41}$$

Le paramètre $\lambda_{t,j}^{(i)}$, compris entre 0 et 0.5, règle le compromis entre l'information apportée par les observations et celle apportée par la valeur des paramètres à l'instant précédent. Dans la mesure du possible, il est pris égal à 0.5. Cependant, lorsque $f_{t,j}^{\mathbf{y}}$ est trop éloigné de $f_{t-1,j}^{(i)}$, par rapport à la variance $(r_{t,j}^{\mathbf{f}})^{(i)}$, sa valeur tend vers 0, afin de favoriser une évolution régulière de la trajectoire du paramètre.

IV.2.3.a Premier modèle sur les amplitudes

En combinant l'équation d'observation (IV.40), avec une équation de transition des amplitudes définie par une simple marche aléatoire, de la forme :

$$\begin{bmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{t-1} \\ \mathbf{b}_{t-1} \end{bmatrix} + \mathbf{v}_{t-1}^{\mathbf{a},\mathbf{b}} \tag{IV.42}$$

[2]Il faut noter que, quelquesoit le contenu du vecteur $\mathbf{f}_t$, et donc sa dimension, n représente le nombre total de partiels détectés dans le signal fenêtré et dont on veut estimer les amplitudes.

le filtre particulaire obtenu peut être rao-blackwellisé. C'est la caractéristique principale de cette seconde version de l'algorithme. En reprenant les résultats exposés dans le chapitre précédent, le nombre de composantes et les fréquences sont estimés par filtrage particulaire et les amplitudes par une banque de filtres de Kalman. L'approximation particulaire de la densité conditionnelle de la trajectoire des amplitudes, étant données toutes les observations, est un mélange de gaussiennes défini par :

$$\hat{p}(\mathbf{a}_{0:t}, \mathbf{b}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} \tilde{\omega}_t^{(i)} p(\mathbf{a}_{0:t}, \mathbf{b}_{0:t}|\mathbf{y}_{1:t}, K_{0:t}^{(i)}, \mathbf{f}_{0:t}^{(i)}) \tag{IV.43}$$

$$= \sum_{i=1}^{N} \tilde{\omega}_t^{(i)} \mathcal{N}(\mathbf{a}_{0:t}, \mathbf{b}_{0:t}; \boldsymbol{\mu}_{t|t}^{(i)}, \boldsymbol{\Sigma}_{t|t}^{(i)}) \tag{IV.44}$$

le vecteur $\boldsymbol{\mu}_{t|t}^{(i)}$ et la matrice $\boldsymbol{\Sigma}_{t|t}^{(i)}$ étant calculés, pour chaque particule, à partir de $\mathbf{y}_t$, $k_t^{(i)}$, $\mathbf{f}_t^{(i)}$, $\boldsymbol{\mu}_{t-1|t-1}^{(i)}$ et $\boldsymbol{\Sigma}_{t-1|t-1}^{(i)}$, avec les équations du filtre de Kalman (voir l'annexe A). Cette estimation séquentielle des deux premiers moments, nous permet, pour simplifier, de ne pas mettre à jour les hyper-paramètres concernant les amplitudes. Les poids non normalisés se calculent alors par la formule :

$$\omega_t^{(i)} \propto p(\mathbf{y}_t|K_{0:t}^{(i)}, \mathbf{f}_{0:t}^{(i)}, \mathbf{y}_{1:t-1}) \frac{p(\mathbf{f}_t^{(i)}|\mathbf{f}_{t-1}^{(i)})}{q_t(\mathbf{f}_t^{(i)}|\mathbf{f}_{t-1}^{(i)}, \mathbf{y}_t)} \frac{p((\mathbf{r}_t^{\mathbf{f}})^{(i)}|(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)})}{q_t((\mathbf{r}_t^{\mathbf{f}})^{(i)}|(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)}, \mathbf{y}_t)} \tag{IV.45}$$

où la vraisemblance $p(\mathbf{y}_t|K_{0:t}^{(i)}, \mathbf{f}_{0:t}^{(i)}, \mathbf{y}_{1:t-1})$ est une gaussienne définie par :

$$p(\mathbf{y}_t|K_{0:t}^{(i)}, \mathbf{f}_{0:t}^{(i)}, \mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{y}_t; \mathbf{y}_{t|t-1}^{(i)}, \mathbf{S}_{t|t-1}^{(i)}) \tag{IV.46}$$

dont les deux premiers moments sont aussi donnés par les équations du filtre de Kalman (voir l'annexe A). Enfin, les estimations $\hat{\mathbf{a}}_t$ et $\hat{\mathbf{b}}_t$ des amplitudes, se calculent sur le même principe que les fréquences dans l'équation (IV.30). Pour $k = 1, \ldots, \hat{K}_t$, on a :

$$\begin{bmatrix} \hat{\mathbf{a}}_{t,k} \\ \hat{\mathbf{b}}_{t,k} \end{bmatrix} = \frac{1}{N'_{t,k}} \sum_{i=1}^{N'_{t,k}} \boldsymbol{\mu}_{t|t,k}^{(i)} \tag{IV.47}$$

IV.2.3.b Second modèle sur les amplitudes

Dans la seconde version de l'algorithme qui vient d'être présentée, on peut noter que, si l'estimation de K_t et $\mathbf{f}_t$ est séparée de l'estimation de $\mathbf{a}_t$ et $\mathbf{b}_t$, l'une ne peut pas se faire sans l'autre. L'estimation des amplitudes intervient dans le filtre particulaire, au niveau du calcul du poids, et plus particulièrement, de la vraisemblance. Nous avons vu, dans le chapitre précédent, qu'elle ne se simplifiait pas en $p(\mathbf{y}_t|K_t^{(i)}, \mathbf{f}_t^{(i)})$, du fait d'une dépendance aux valeurs passées des amplitudes. Afin de pouvoir découpler totalement les deux estimations, il est nécessaire de supprimer cette dépendance au passé, en remplaçant la marche aléatoire de l'équation (IV.42), par une équation du type :

$$\begin{bmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{bmatrix}; \boldsymbol{\mu}_t(K_t, \mathbf{f}_t), r^{\mathbf{y}} \boldsymbol{\Sigma}_t(K_t, \mathbf{f}_t) \right) \tag{IV.48}$$

où le vecteur $\boldsymbol{\mu}_t$ et la matrice $\boldsymbol{\Sigma}_t$ sont construits à partir de K_t et de $\mathbf{f}_t$. La formule du calcul des poids non normalisés peut alors se simplifier, pour donner la formule du filtre particulaire classique :

$$\omega_t^{(i)} \propto p(\mathbf{y}_t|K_t^{(i)}, \mathbf{f}_t^{(i)}) \frac{p(\mathbf{f}_t^{(i)}|\mathbf{f}_{t-1}^{(i)})}{q_t(\mathbf{f}_t^{(i)}|\mathbf{f}_{t-1}^{(i)}, \mathbf{y}_t)} \frac{p((\mathbf{r}_t^{\mathbf{f}})^{(i)}|(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)})}{q_t((\mathbf{r}_t^{\mathbf{f}})^{(i)}|(\mathbf{r}_{t-1}^{\mathbf{f}})^{(i)}, \mathbf{y}_t)} \tag{IV.49}$$

Il reste que le terme de vraisemblance $p(\mathbf{y}_t|K_t^{(i)}, \mathbf{f}_t^{(i)})$ ne peut pas se calculer directement. En effet, à un moment ou un autre, il est tout de même nécessaire de faire intervenir des amplitudes. Le calcul peut néanmoins se faire en notant que $p(\mathbf{y}_t|K_t^{(i)}, \mathbf{f}_t^{(i)})$ est la marginale d'une densité plus globale, prenant en compte les amplitudes. En choisissant $\boldsymbol{\mu}_t = \mathbf{0}$, dans l'équation (IV.48), la vraisemblance peut être calculée, à une constante près, par :

$$p(\mathbf{y}_t|K_t^{(i)}, \mathbf{f}_t^{(i)}) \propto \frac{|\mathbf{S}_t|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_t|^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_t - \mathbf{y}_t^{\mathrm{T}}\mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C} + \boldsymbol{\Sigma}_t^{-1})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y}_t}{2r^{\mathbf{y}}}\right) \tag{IV.50}$$

avec $\mathbf{S}_t^{-1} = \mathbf{C}^{\mathrm{T}}\mathbf{C} + \boldsymbol{\Sigma}_t^{-1}$ et $\mathbf{C}$ une notation allégée de $\mathbf{C}(K_t^{(i)}, \mathbf{f}_t^{(i)})$. Les détails du calcul sont donnés dans l'annexe C. Ainsi, le nombre de composantes et les fréquences sont estimés au cours du temps, sans jamais estimer directement les amplitudes. La vraisemblance est calculée en intégrant sur tout l'espace des amplitudes et c'est là que réside un autre avantage de cette approche, car elle revient à essayer toutes les amplitudes possibles, pour un partiel donné. Ceci nous permet aussi, à l'instar de la version précédente et pour simplifier, de ne pas mettre à jour les hyper-paramètres sur les amplitudes. Enfin, à l'instant t, une fois $\hat{K}_t$ et $\hat{\mathbf{f}}_t$ calculés, une estimation des amplitudes peut être obtenue par [Dav04] :

$$\begin{bmatrix} \hat{\mathbf{a}}_t \\ \hat{\mathbf{b}}_t \end{bmatrix} = (\mathbf{C}^{\mathrm{T}}\mathbf{C} + \boldsymbol{\Sigma}_t^{-1})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y}_t \tag{IV.51}$$

où $\mathbf{C}$ est une notation allégée de $\mathbf{C}(\hat{K}_t, \hat{\mathbf{f}}_t)$.

IV.3 Etude comparative sur un exemple jouet

Les trois algorithmes qui viennent d'être présentés, sont basés sur des choix précis effectués à différents niveaux. Afin d'évaluer leurs performances et de les comparer, ils ont été utilisés pour estimer le nombre de composantes harmoniques existantes dans un signal synthétique, ainsi que leurs caractéristiques. Le signal utilisé et le cadre de l'étude sont d'abord présentés. Puis, les résultats obtenus par les algorithmes sont donnés et évalués. Enfin, une synthèse est faite sur ces trois versions.

IV.3.1 Cadre de l'étude

Le signal utilisé ne cherche pas à ressembler à un signal réel mais doit plutôt être considéré comme un concentré des différentes difficultés auxquelles les algorithmes d'estimation de fréquences fondamentales sont souvent confronté. Son spectrogramme ainsi que les représentations temps-fréquence théoriques de ses composantes, sont donnés dans la figure IV.4. Il dure 1 s et est échantillonné à une fréquence de 10 000 Hz. Le nombre de composantes harmoniques présentes simultanéement varie de 0 à 3. Voici les détails concernant chacune d'entre elles :

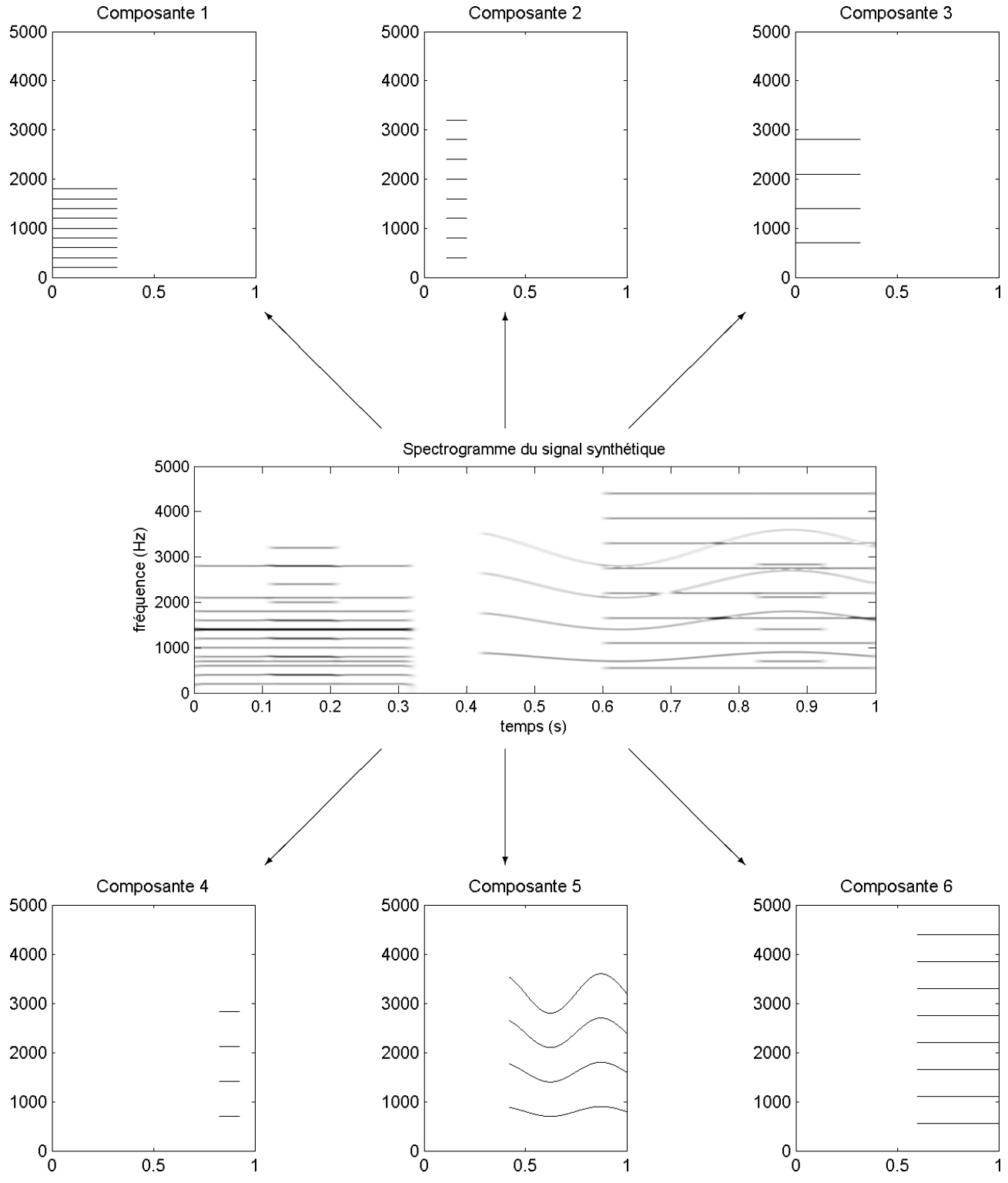


FIG. IV.4 – Décomposition du signal synthétique.

Composante 1 : elle est composée de 9 partiels. La fréquence fondamentale est 200 Hz et il n'y a pas d'inharmonicité. Tous les partiels ont la même amplitude.

Composante 2 : elle est composée de 8 partiels. La fréquence fondamentale est 400 Hz et il n'y a pas d'inharmonicité. Tous les partiels ont la même amplitude. Cette composante est en relation d'octave avec la première, ce qui représente une difficulté pour l'algorithme.

Composante 3 : elle est composée de 4 partiels. La fréquence fondamentale est 700 Hz et il n'y a pas d'inharmonicité. Tous les partiels ont la même amplitude. Le second partiel, de fréquence 1400 Hz, est confondu avec le septième partiel de la première composante, ce qui constitue une autre difficulté pour l'algorithme.

Composante 4 : elle est composée de 4 partiels. La fréquence fondamentale est 700 Hz. Tous les partiels ont la même amplitude. Cette composante est inharmonique : le quatrième partiel est à la fréquence 2833.4 Hz au lieu de 2800 Hz. La fréquence des partiels s'éloigne de plus en plus de la fréquence harmonique, ce qui constitue une troisième difficulté pour l'algorithme.

Composante 5 : elle est composée de 4 partiels. L'amplitude des partiels décroît avec le rang. La fréquence fondamentale est modulée sinusoïdalement autour de la fréquence 800 Hz, ce qui constitue une quatrième difficulté pour l'algorithme.

Composante 6 : elle est composée de 8 partiels. La fréquence fondamentale est 550 Hz. Tous les partiels ont la même amplitude. Les partiels de cette composante se croisent avec ceux de la cinquième composante, ce qui constitue une dernière difficulté pour l'algorithme.

Pour cette étude comparative, nous avons choisi de considérer un modèle d'inharmonicité de la forme, pour $h = 2, \ldots, H$:

$$f_h = f_1 h \sqrt{1 + \gamma h^2} \tag{IV.52}$$

Le vecteur $\mathbf{f}_t$ contient donc les K_t fréquences fondamentales $f_{t,k}$ et les K_t coefficients d'inharmonicité $\gamma_{t,k}$. Il est donc de dimension $2K_t$. En ce qui concerne les amplitudes des partiels, si elles sont toutes estimées, nous avons choisi de mettre une dépendance entre les différents partiels, au niveau de la variance du modèle d'évolution. Plus précisément, on pose la relation :

$$r^{\boldsymbol{\rho}}_{t,k,h} = \left(\frac{-0.9}{H-1} h + \frac{H - 0.1}{H-1} \right) r^{\text{ampl}} \tag{IV.53}$$

où $\boldsymbol{\rho}$ désigne $\mathbf{a}$, $\mathbf{b}$ ou $\mathbf{A}$, suivant la version, et H le nombre maximum de partiels estimés. Cette relation permet de faire décroître la variance, quand le rang h du partiel augmente. Ce choix est motivé par le fait que, dans une grande majorité des cas, l'énergie des partiels diminue lorsque l'on s'éloigne de la fréquence fondamentale. Dans l'étude comparative, les hyper-paramètres des amplitudes ne sont donc pas estimés et la variance r^{ampl} est fixe au cours du temps. Les autres choix concernant le modèle sont ceux présentés avec les trois versions de l'algorithme. Les différents paramètres des algorithmes sont donnés dans le tableau IV.3.

Au lieu de fixer la variance de l'équation d'observation, nous avons choisi de la faire varier au cours du temps, en la rendant proportionnelle à la variance du signal fenêtré.

	Version 1 (section IV.2.2)	Version 2 (section IV.2.3.a)	Version 3 (section IV.2.3.b)
Nombre de particules	$N = 100$		
Nombre minimum de composantes harmoniques	$K_{min} = 0$		
Nombre maximum de composantes harmoniques	$K_{max} = 4$		
Nombre maximum de partiels	$H = 10$		
Longueur de la fenêtre	$L_{\mathbf{w}} = 51.2$ ms (soit 512 points)		
Recouvrement des fenêtres entre deux instants consécutifs	98 %		
Variances particulières pour les fréquences fondamentales	$r^f_{ini} = 2^2 \quad , \quad r^f_{min} = 0.5^2 \quad , \quad r^f_{max} = 5^2$		
Variances particulières pour les paramètres d'inharmonicité	$r^\gamma_{ini} = 0.00005^2 \quad , \quad r^\gamma_{min} = 0.00001^2 \quad , \quad r^\gamma_{max} = 0.0001^2$		
Variance particulière sur les amplitudes	$r^{\text{ampl}} = 0.03^2$	$r^{\text{ampl}} = 0.5^2$	
Variance de l'équation d'observation	$r^{\mathbf{y}} = \text{var}(\mathbf{x}_t)/2$	$r^{\mathbf{y}} = \text{var}(\mathbf{x}_t)$	$r^{\mathbf{y}} = 2\,\text{var}(\mathbf{x}_t)$

TAB. IV.3 – Valeur des différents paramètres pour les trois versions de l'algorithme.

IV.3.2 Présentation des résultats

Les différents résultats d'estimation des trois versions de l'algorithme, sont donnés dans les figures IV.5, IV.6, IV.7 et IV.8. A chaque fois, les quantités théoriques, c'est-à-dire simulées, sont rappelées. Dans la figure IV.7, les estimations des fréquences des partiels, au cours du temps, sont données. Cela permet de vérifier l'estimation du nombre de partiels ainsi que de l'inharmonicité éventuelle des composantes. Une évaluation quantitative est effectuée par le calcul de l'erreur RMS (*Root Mean Square*) au cours du temps, voir figure IV.8. Le calcul se fait selon la formule :

$$\text{rms}_t = \sqrt{\frac{\sum_{i=1}^{N} \tilde{\omega}_t^{(i)} \|\mathbf{y}_t - h_t(K_t^{(i)}, \boldsymbol{\theta}_t^{(i)})\|^2}{L_{\mathbf{w}}}} \tag{IV.54}$$

Il est important de remarquer que, selon la version de l'algorithme, $\mathbf{y}_t$ et h_t ne sont pas définis de la même manière. C'est pourquoi, dans la figure IV.8, deux erreurs RMS sont données pour la première version de l'algorithme : la première est effectivement calculée avec l'équation (IV.54), la seconde est calculée sur le même principe mais en remplaçant $\mathbf{y}_t$ par le signal fenêtré $\mathbf{x}_t$, afin que la comparaison avec les deux autres versions se fasse sur la même base.

IV.3.3 Discussion

La principale remarque que nous pouvons faire en regardant les résultats des différentes versions, est que seule le troisième algorithme semble capable de détecter la présence d'une

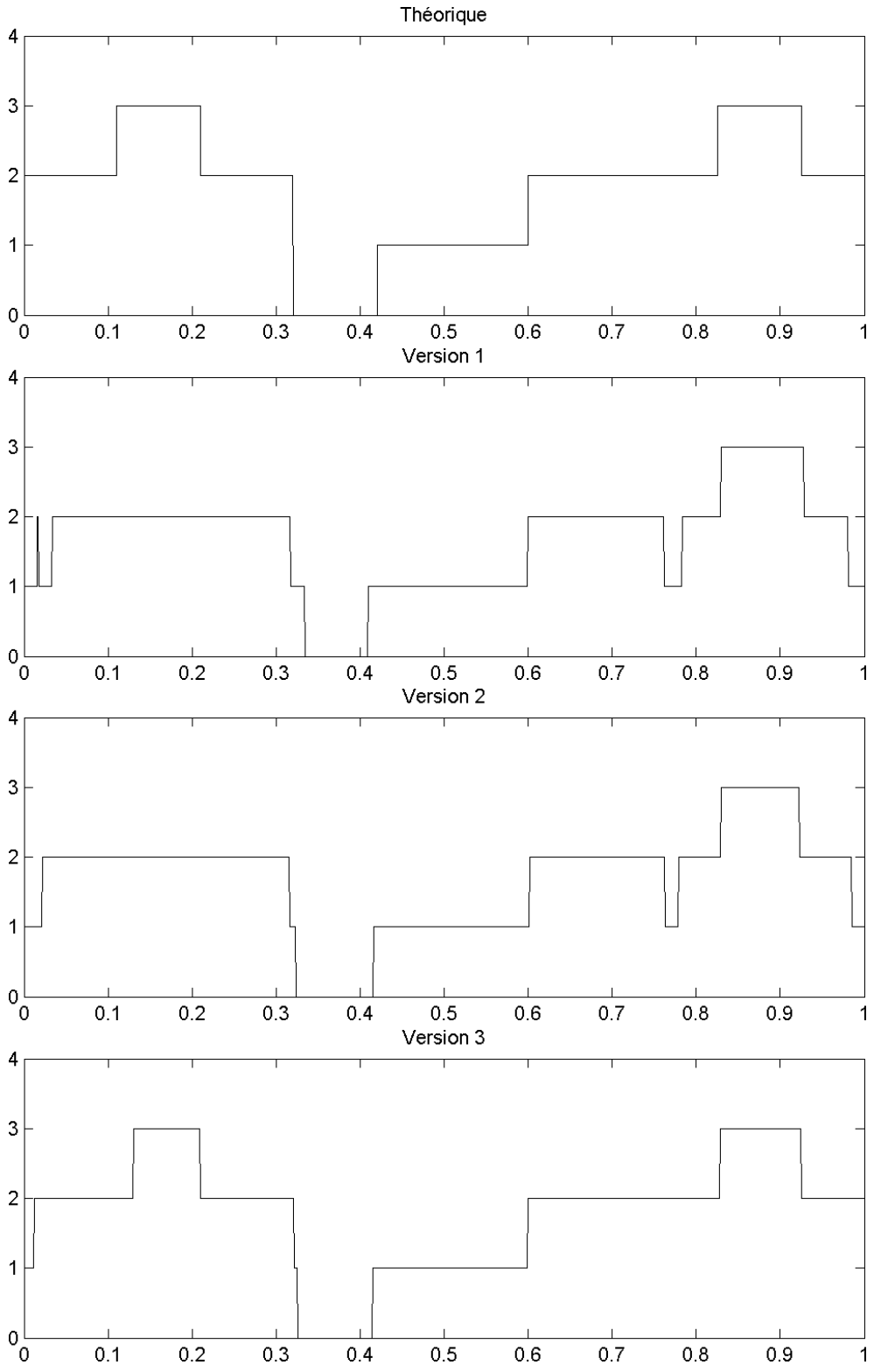


FIG. IV.5 – Estimation du nombre de composantes, au cours du temps, pour le signal synthétique.

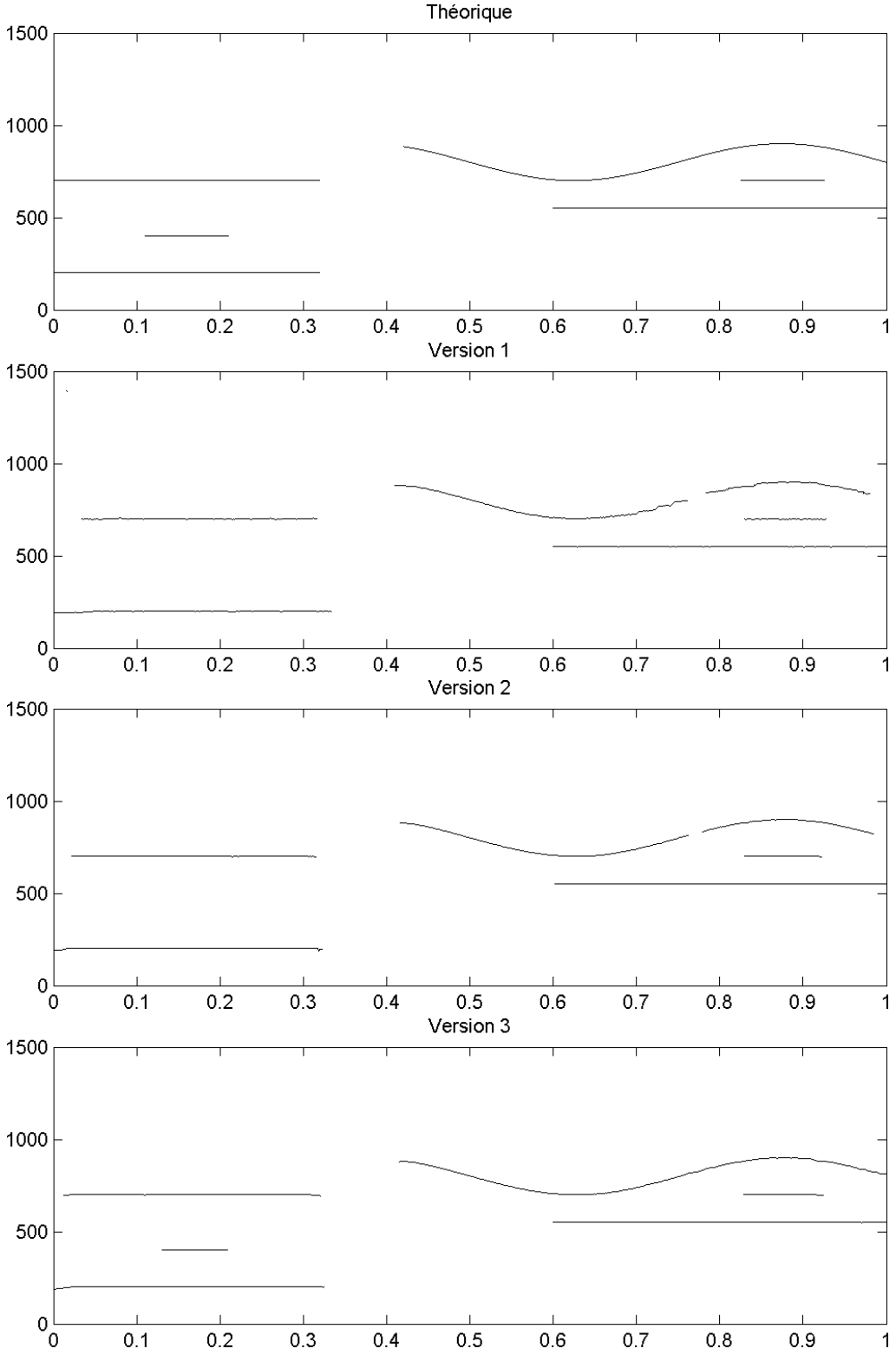


FIG. IV.6 – Estimation de fréquences fondamentales, au cours du temps, pour le signal synthétique.

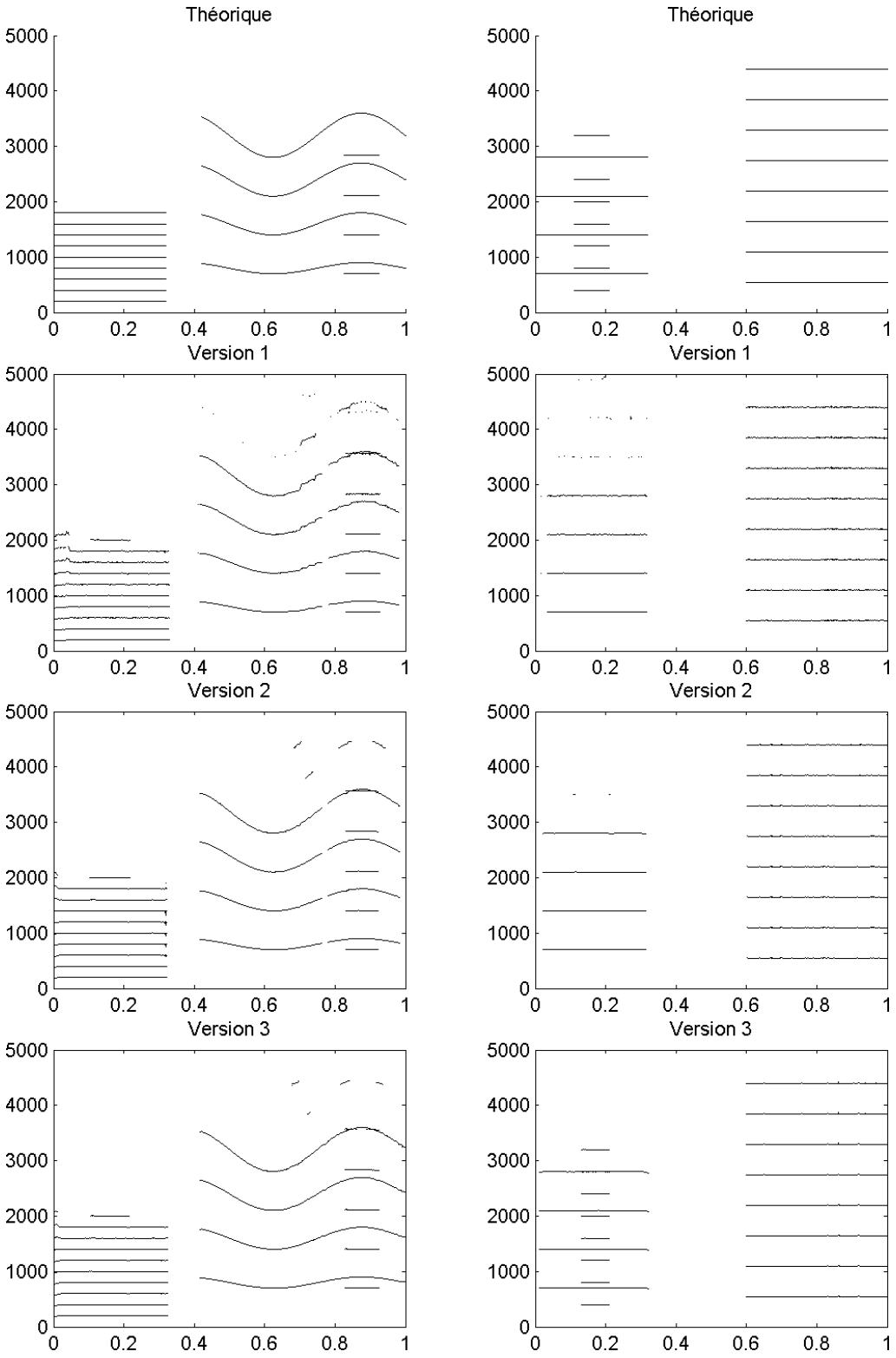


FIG. IV.7 – Estimation des fréquences des partiels, au cours du temps, pour le signal synthétique. Les différentes composantes ont été réparties entre deux graphes, pour plus de lisibilité.

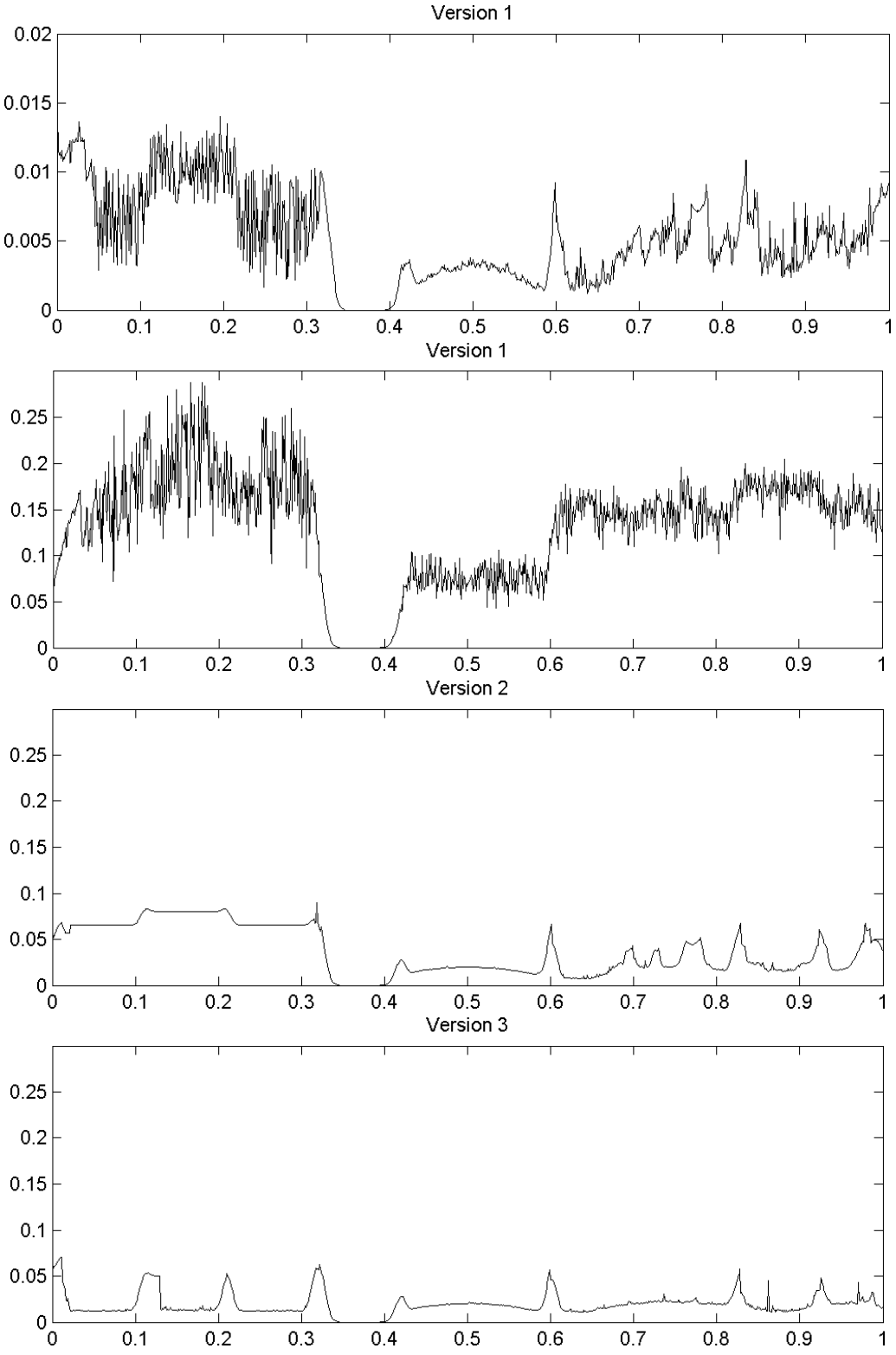


FIG. IV.8 – Erreur RMS, au cours du temps, pour le signal synthétique.

composante, dont la fréquence fondamentale est en relation d'octave avec celle d'une autre. Avant d'essayer de donner quelques explications à ce fait, il faut signaler que la relation d'octave simulée dans le signal synthétique, est particulièrement pathologique. En effet, les deux composantes concernées ne sont pas inharmoniques et leurs partiels ont tous la même amplitude. L'algorithme ne peut donc pas avoir recours à une éventuelle différence de structure fréquentielle, pour les distinguer. Une première explication à cette non détection, pourrait trouver son origine dans le fenêtrage. En effet, le recours à une fenêtre de pondération à la conséquence bien connue de créer un compromis entre les résolutions fréquentielle et temporelle. En particulier, la fenêtre vient, en quelque sorte, étaler les évènements brusques sur une certaine durée. Or, les deux premières versions de l'algorithme estiment l'amplitude au cours du temps, en considérant un modèle d'évolution reliant les valeurs estimées d'un instant à l'autre. Avec l'effet de fenêtre, cela leur laisse le temps de mettre à jour leurs amplitudes et ainsi de confondre l'apparition de la composnate en relation d'octave, avec une brusque augmentation de l'amplitude. Dans la troisième version, à chaque instant, toutes les amplitudes possibles sont testées, en imposant une répartition de l'amplitude entre les partiels telle que, l'amplitude d'un partiel de rang h a une forte probabilité d'être inférieure à celle d'un partiel de rang $h' < h$. Cette répartition est modélisée par le profil de décroissance de la variance que nous avons choisi. Au delà de cette explication, il est important de noter que la détection de fréquences fondamentales multiples l'une de l'autre, est un problème particulièrement difficile. Nous avons vu, dans le chapitre II, que les méthodes qui veulent prendre en compte une telle situation, nécessitent pratiquement toujours l'insertion de connaissances *a priori*, concernant la structure fréquentielle des sources. La solution mise en œuvre dans la troisième version est plus générale et nous verrons dans la section suivante, qu'elle donne de bons résultats sur des signaux réels.

Dans la figure IV.8, on peut noter que l'erreur RMS présente parfois des pics. Cela correspond aux instants où il y a un changement du nombre de composantes. L'apparition ou la disparition des composantes, dans notre signal synthétique, étant brusque (il n'y a pas de transition), il est normal de voir apparaître ces pics, ces étapes étant non harmoniques.

Enfin, l'ordre de grandeur du temps de calcul, pour chacune des versions, est d'environ 8 heures pour la première, 13 heures pour la seconde et 45 minutes pour la troisième.

Les trois versions présentées précédemment et évaluées sur cet exemple jouet, reprennent les différentes étapes de l'élaboration de notre algorithme et d'autres évaluations ont été effectuées au fur et à mesure de l'avancement [Dub05c, Dub05a, Dub05b, Dub]. La version finalement retenue est, assez logiquement, la dernière.

IV.4 Application à la transcription automatique de la musique

Nous avons vu, dans le chapitre II, que la musique est le domaine de prédilection des algorithmes d'estimation de fréquences fondamentales, notamment pour résoudre le problème de la transcription automatique. Il est donc assez naturel de vouloir étudier le comportement de notre algorithme sur des signaux musicaux. La transcription automatique de la musique est un problème comportant de multiples facettes comme l'estimation des notes jouées, l'analyse du rythme, la séparation de sources, la reconnaissance des instruments ou l'étude de la struc-

ture musicale [Kla06]. Notre algorithme peut apporter une réponse à certaines d'entre elles. En effet, l'entité de base qu'il cherche à extraire, est ce que nous avons appelé une composante harmonique. Pour un morceau de musique, cela correspond à une note. De plus, comme toute la structure fréquentielle de la note est estimée, l'algorithme peut aussi séparer les différents instruments. Dans cette partie, nous avons appliqué l'algorithme à deux types de signaux : le premier est généré à partir d'un fichier MIDI[3] et le second est un signal réel, issu d'un CD. La représentation que l'on cherche à obtenir est comparable aux rouleaux à musique des pianos mécaniques (*piano-roll*), c'est-à-dire donnant la succession des fréquences fondamentales correspondant aux notes jouées dans le signal. D'après les résultats obtenus dans la partie précédente, nous avons choisi d'utiliser la troisième version de l'algorithme pour estimer les notes successives de ces deux signaux. Les paramètres estimés sont les mêmes que précédemment, en particulier, nous avons considéré le même modèle d'inharmonicité ainsi qu'un profil de décroissance de la variance sur les amplitudes.

IV.4.1 Signal MIDI

Le signal est un duo flûte/piano, jouant la célèbre chanson française « Le roi Dagobert ». La partition à partir de laquelle le signal MIDI a été généré, est donnée dans la figure IV.9. Il est échantillonné à une fréquence de 11 025 Hz et dure 25 s. Le nombre de notes jouées simultanément varie entre 0 et 2 et les notes peuvent être en relation d'octave.

Les paramètres de l'algorithme sont les suivants :

- longueur de la fenêtre : $L_{\mathbf{w}} = 1024$ points (≈ 93ms)
- écart entre deux instants d'analyse : 50 points (≈ 4.5ms) le recouvrement est de 95.1%
- nombre maximum de notes : $K_{max} = 5$
- nombre minimum de notes : $K_{min} = 0$
- nombre maximum de partiels : $H = 40$
- nombre de particules : $N = 500$

Les résultats sont donnés dans les figures IV.10 et IV.11, l'ordre de grandeur du temps de calcul, étant d'environ 12 heures. Le premier commentaire concerne le nombre de notes estimé. Même si, à première vue, il semble très différent du nombre théorique, il faut noter que le résultat obtenu est normal. En effet, le nombre de notes théorique, au cours du temps, a été calculé à partir de la partition, en considérant la durée des différents types de note (noire, croche, ...). Le nombre estimé, lui, prend aussi en compte l'écho de la note : on entend la note jouée plus longtemps (ou moins longtemps) que sa durée théorique. Cela se vérifie dans la figure IV.11, où les surestimations du nombre de composantes sont souvent consécutives à un changement de la note jouée par l'instrument. A l'inverse, à la fin du morceau, par exemple, le piano joue successivement deux blanches pointées et on peut remarquer que, dans le graphe des fréquences, la durée correspondante est inférieure. En fait, cela est dû au piano, pour lequel il est difficile de faire durer la note aussi longtemps que sa durée théorique. En ce qui concerne les fréquences, on peut constater qu'elles sont assez bien estimées. On remarque, cependant, qu'il y a pratiquement toujours une fluctuation de la fréquence au début ou à la fin des notes.

[3]MIDI est la contraction de *Musical Instrument Digital Interface*. Un fichier MIDI contient toutes les informations (instruments utilisés, partition jouée, paramètres du jeu musical, ...) nécessaires à la synthèse du morceau de musique par un synthétiseur électronique.

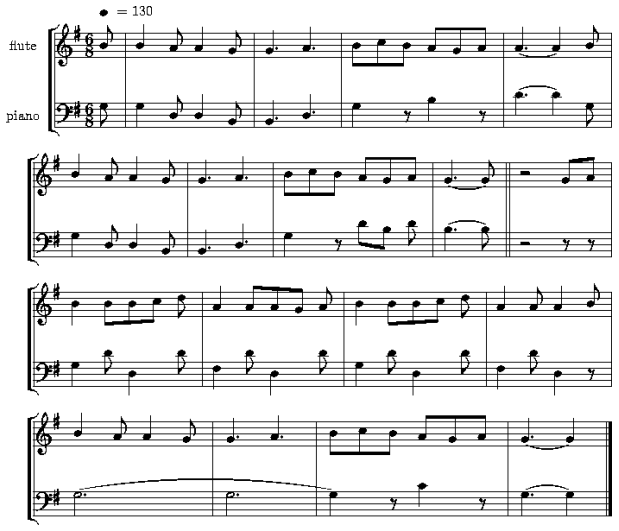


FIG. IV.9 – Partition de la chanson « Le roi Dagobert », à partir de laquelle le signal MIDI a été généré.

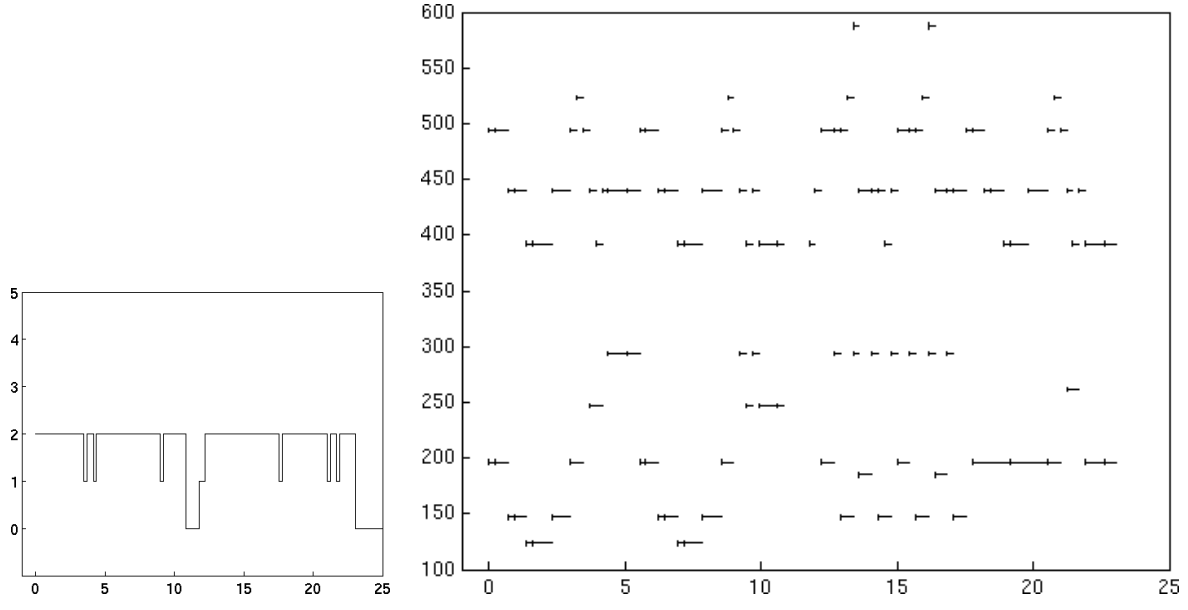


FIG. IV.10 – Nombre de notes (à gauche) et fréquences fondamentales (à droite) théoriques, au cours du temps, pour le signal MIDI. Les traits verticaux, dans le graphe de droite, indiquent les instants d'attaque des notes.

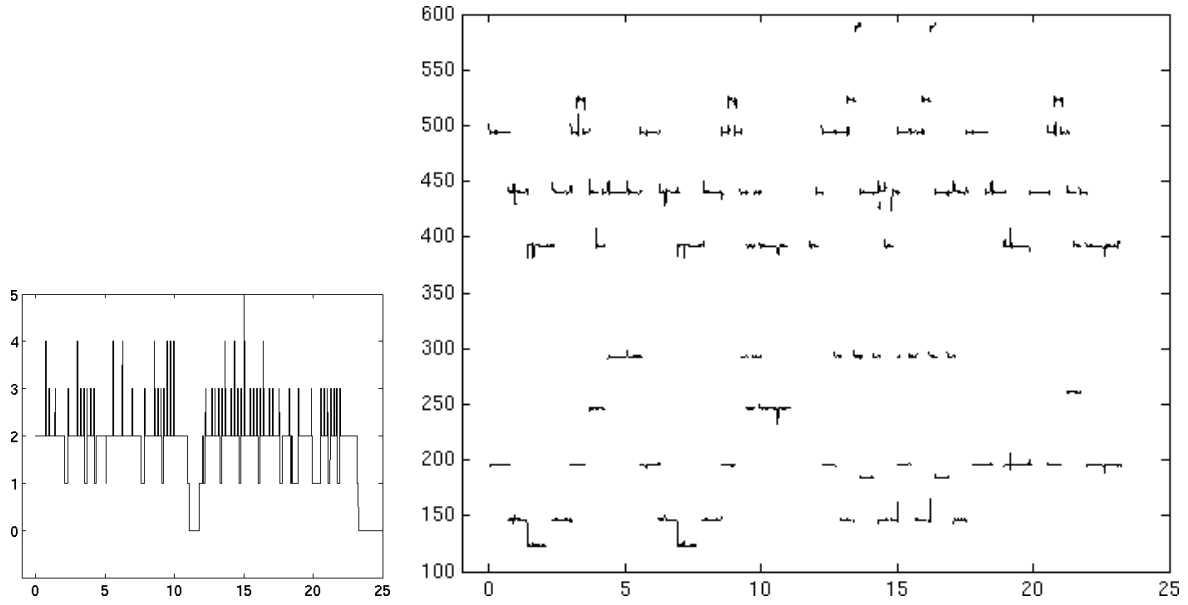


FIG. IV.11 – Nombre de notes (à gauche) et fréquences fondamentales (à droites) estimés, au cours du temps, pour le signal MIDI.

Elle est causée par l'attaque. En effet, cette courte étape transitoire est non harmonique et l'algorithme ne trouve pas forcement tout de suite la bonne fréquence. Du reste, une amplitude faible est souvent associée à ces fluctuations transitoires et elles ne perturbent pas l'écoute du signal reconstruit à partir des notes estimées.

IV.4.2 Signal réel

Le signal réel utilisé est un extrait du Canon en Ré de Johann Pachelbel. La fréquence d'échantillonnage est de 22 050 Hz et le signal dure 34 s. Les paramètres de l'algorithme sont les mêmes que précédemment, on peut simplement noter que la fréquence d'échantillonnage étant double, la longueur de la fenêtre devient égale à 2048 points. Même si sa durée reste identique (≈ 93 ms), cela a une répercussion sur le temps de calcul, les matrices considérées étant deux fois plus grandes. L'ordre de grandeur du temps de calcul est alors d'environ 61 heures. Les résultats obtenus sont donnés dans les figures IV.12, IV.13 et IV.14.

Les commentaires concernant ces résultats sont du même ordre que pour le signal MIDI. N'ayant pas à disposition, la partition jouée par les musiciens, il est difficile d'évaluer l'estimation du nombre de composantes et des fréquences. Une écoute comparative du signal d'origine et du signal reconstruit, permet néanmoins de se rendre compte des résultats obtenus : les deux instruments sont nettement restitués ainsi que la mélodie jouée. Une évaluation plus quantitative est effectuée par le calcul de l'erreur RMS au cours du temps, voir figure IV.14. Le calcul se fait selon l'équation (IV.54).

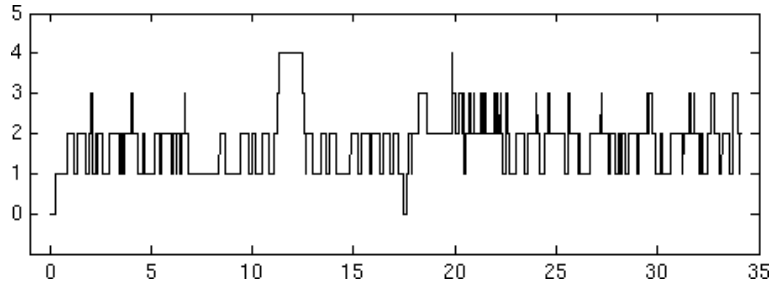


FIG. IV.12 – Estimation du nombre de notes, au cours du temps, pour le signal réel.

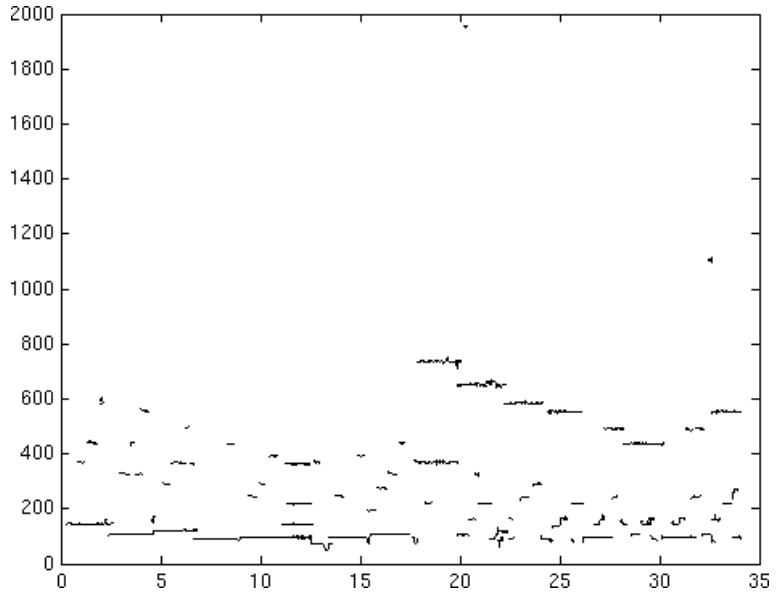


FIG. IV.13 – Fréquences fondamentales estimées, au cours du temps, pour le signal réel.

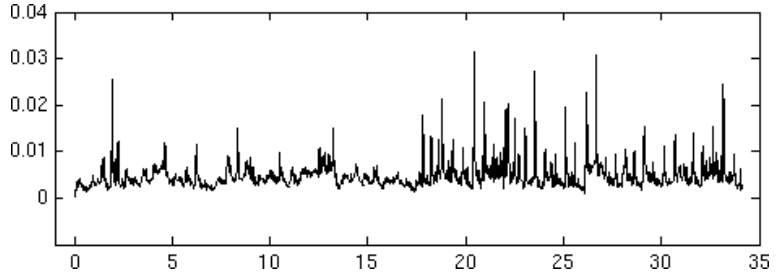


FIG. IV.14 – Erreur RMS au cours du temps, pour le signal réel.

Chapitre V

Application à la métabole

C'est par l'expérience que la science et l'art
font leur progrès chez les hommes.
Aristote

La notion d'effet sonore, telle qu'elle a été définie par le CRESSON, à Grenoble, a était brièvement introduite au début de ce manuscrit. Avant de se focaliser sur l'un d'entre eux, l'effet de métabole, il convient de rappeler les fondements de cet outils, ce qui est fait dans la première partie de ce chapitre. L'effet de métabole est ensuite étudié en détail et la décomposition en composantes harmoniques, effectuée par l'algorithme élaboré tout au long de ce manuscrit, est utilisée afin de le caractériser et éventuellement de le reconnaître, avant de terminer par une partie conclusive.

V.1 Les effets sonores

Le sonore est une dimension trop souvent négligée dans l'étude des modes de vie et d'habitat, sauf quand il s'agit d'en étudier les aspects négatifs (lutte contre le bruit et les perturbations qu'il engendre). Pourtant, il constitue toujours un révélateur privilégié de la diversité des typologies architecturales, des pratiques sociales ou des cultures locales [Bar99]. L'enjeu est donc la description et l'analyse des ambiances sonores, et ceci de manière générique, c'est-à-dire sans *a priori* sur la nature des sources à l'origine de cette ambiance perçue. Si la caractérisation sonore en intérieur est possible, grâce à des paramètres objectifs [Pel91], cette tâche s'avère difficile en extérieur. En effet, les critères retenus pour l'acoustique des salles ne sont guères adaptés aux sites extérieurs et il n'existe pas, au moins jusqu'à un passé récent, d'outils capables de mesurer les performances sonores, sur le plan de l'impact perceptif, des espaces ouverts et des petits espaces clos, avec toute la finesse souhaitable [Odi96]. De plus, l'influence de la dimension humaine dans la perception sonore, n'est que partiellement prise en compte dans une évaluation quantitative et le recours à une composante qualitative dans la manière de décrire les ambiances sonores devient alors nécessaire. Cependant, il ne faut pas oublier qu'un son est, avant tout, un phénomène physique. Les outils utilisés pour la description de l'environnement

sonore doivent donc résulter d'une juste complémentarité entre le quantitatif et le qualitatif, et non d'une simple juxtaposition des deux. C'est sur ce constat essentiel que les outils utilisant les objets sonores [Sch66] ou le paysage sonore [Sch80], comme descripteurs, sont mis à défaut. En effet, le premier est trop élémentaire, c'est-à-dire qu'il s'attache à la caractérisation son par son, et se situe à une échelle d'analyse trop basique. Le second, *a contrario*, s'avère trop large et trop flou et renvoie plus à ce qui est perceptible comme une unité esthétique, dans le milieu sonore [Aug95].

Il faut aussi rappeler que le contexte à l'origine de l'émergence des effets sonores, est celui de la ville. S'il y a une déformation perceptive du signal sonore, liée au sujet qui écoute, l'architecture et l'environnement physique dans lequel l'auditeur se trouve, ont aussi une influence sur l'apparition et l'appréciation de telle ou telle ambiance sonore. Avec l'interdisciplinarité inhérente à la prise en compte conjointe du quantitatif et du qualitatif et la capacité à s'écarter du purement esthétique, la contrainte de l'adéquation de l'échelle d'analyse avec celle du milieu urbain forme le troisième pilier sur lesquels est élaboré le concept d'effet sonore. Un effet sonore donné se manifeste et trouve une description dans trois champs complémentaires, correspondant à ces trois piliers.

Sciences humaines

L'environnement sonore est considéré comme un réservoir de possibilités sonores. Chacune d'entre elles pourra être à l'origine d'une déformation perceptive, d'une sélection d'informations et d'une attribution de significations, qui dépendront du contexte (psychologique, culturel, social, ...) dans lequel évolue d'auditeur.

Architecture et urbanisme

Ici, c'est la ville, comme origine et champ de propagation des sources sonores, qui est prise en compte. L'aménagement urbain et l'espace construit peuvent, en effet, directement façonner certains effets sonores. D'autre part, une analyse architecturale de l'endroit dans lequel règne l'ambiance sonore étudiée, permet d'objectiver les conclusions et remarques récoltées *in situ* auprès des habitants.

Acoustique appliquée

C'est le signal physique et toutes les mesures quantitatives qui s'y rapportent, qui sont ici considérés. Il ne doit pas cependant être pris indépendamment des circonstances et des conditions d'audition, l'expérience architecturale, par exemple, pouvant permettre de prévoir certaines performances physiques du signal.

Sur les quelques quatre-vingts effets sonores répertoriés [Aug95], plusieurs taxinomies complémentaires peuvent être envisagées. Une première règle de classification est la séparation entre les effets dits « majeurs » et ceux dits « mineurs ». Trois critères sont utilisés pour déclarer qu'un effet est majeur : le fait qu'il puisse être défini comme un effet de base auquel plusieurs variantes peuvent être rattachées, son caractère omniprésent dans le processus d'écoute (la réverbération est présente dans toute propagation) et, enfin, son caractère substantiel à l'environnement et aux processus étudiés (la métabole est caractéristique du milieu urbain). De même, il est clair que l'importance relative des champs, ou domaines de repérage, cités auparavant, varie d'un ef-

fet à l'autre. Ainsi, cinq catégories d'effets sonores peuvent être définies : les effets élémentaires, les effets de composition, les effets liés à l'organisation perceptive, les effets psychomoteurs et les effets sémantiques. Si l'acousticien est plus familier avec la première catégorie, l'urbaniste ou l'architecte le sera plus avec la seconde, tandis que le sociologue se tournera plus naturellement vers la quatrième ou la cinquième catégorie. C'est parce que les effets sonores traitent de manière interdisciplinaire les interactions entre les sources acoustiques, le milieu aménagé et la perception, qu'ils peuvent prétendre être une approche possible des ambiances sonores. En fait, on peut considérer que les effets sonores se proposent d'être des descripteurs des situations de la vie quotidienne.

Le principe des études exploratoires qui ont déjà été menées sur quelques ambiances sonores particulières [Bar93, Che97a, Odi96, Bar99], s'articule autour des trois axes caractéristiques des effets sonores. Le dénominateur commun est que tous les résultats s'appuient sur des investigations menées *in situ*. Parmi ces dernières, on peut citer : entretiens directs ou semi-directifs, comptes-rendus de perception en mouvement, mesures physiques des ambiances, analyse architecturale et prise de son. Une part importante du corpus de données est obtenue à l'aide de la méthode des parcours commentés ou des parcours d'écoute qualifiée. Celle-ci a pour objectif d'obtenir des comptes-rendus de perception en mouvement en demandant à des passants d'effectuer un cheminement et de décrire ce qu'ils perçoivent et ressentent au fur et à mesure du trajet. Ici, c'est évidemment l'aspect sociologique de l'effet sonore qui prédomine. En ce qui concerne l'approche architecturale, l'idée est d'évaluer la prédictibilité de tel ou tel effet sonore, à partir de l'analyse des formes construites (architecturales et urbaines) et des éléments fonctionnels (usages et activités). En effet, les effets sonores sont susceptibles de se produire si certaines conditions, notamment spatiales, sont réunies. La disposition des éléments construits affecte les conditions d'émission, de propagation et de réception sonore. Pour l'aspect physique de l'effet sonore, les mesures acoustiques effectuées permettent de tirer les renseignements suivants : temps de réverbération, niveau sonore équivalent, atténuation avec la distance en niveau global ou par bandes de fréquence et analyse fréquentielle (essentiellement basée sur des représentations de type spectrogramme). Il faut enfin noter que d'autres paramètres physiques, tels que l'hygrométrie, la température, le vent, *etc*, peuvent rentrer en ligne de compte et doivent aussi être mesurés.

De toutes ces études, il ressort, d'une manière générale, que la caractérisation des effets sonores par des critères physiques, est difficile et que l'objectif de reconnaissance de l'effet sur un enregistrement audio reste un problème posé. C'est à travers cette problématique et en se focalisant sur l'effet de métabole, que nous nous sommes intéressés aux effets sonores.

V.2 L'effet de métabole

Cette partie est le cœur de ce chapitre. L'effet de métabole y est étudié en détail et quelques pistes de caractérisation physique sont données.

V.2.1 Définition

Caractéristique du milieu urbain, l'effet de métabole appartient aux effets liés à l'organisation perceptive. Dans le répertoire des effets sonores [Aug95], il est ainsi défini :

> *Effet perceptif sonore décrivant les relations instables et métamorphiques entre les éléments composant un ensemble sonore. Figure classique de la rhétorique, la métabole caractérise l'instabilité dans le rapport structural qui lie les parties d'un ensemble, et donc, la possibilité de commuter dans n'importe quel ordre les composants élémentaires d'une totalité, la faisant percevoir comme étant en perpétuelle transition. En grec ancien, le mot* metabolos *signifie ce qui est changeant, quelque chose qui est en métamorphose. Ici, le changement considéré affecte le rapport des éléments qui composent l'environnement sonore, celui-ci pouvant se définir comme l'addition et la superposition de sources multiples entendues simultanément.*

Transdisciplinaires par essence, les effets sonores existent et peuvent être expérimentés dans divers champs du savoir et de la pratique. Cependant, certains domaines de repérage sont plus adaptés que d'autres, pour l'identification de tel ou tel effet. Pour l'effet de métabole, la dominante de repérage est le domaine de la psychologie et de la physiologie de la perception [Aug95]. Extraits du répertoire, nous pouvons citer ces quelques points, appartenant à ce domaine :

- *L'effet de métabole comporte deux critères fondamentaux : l'instabilité de la structure perçue dans le temps et la distinctibilité des parties ou de l'ensemble dans une composition donnée.*
- *L'oreille a la capacité de percevoir des sons multiples comme une entité, mais en même temps, son pouvoir discriminatoire lui permet d'écouter plus particulièrement ou de sélectionner certains d'entre eux.* [...] *Certains éléments se détachent, formant les « figures » alors que les autres restent en « fond ».* [...] *Il existe des situations sonores créatrices d'effet de métabole, lorsque tout se fond, lorsque, d'un ensemble composite, n'émerge pas plus un son qu'un autre. De telles situations engendrent l'instabilité perceptive entre figure et fond.*
- *Outre cette relativité existant entre figure et fond sonores, on peut comprendre l'effet de métabole par le phénomène qui consiste à ne pouvoir distinguer clairement les sons les uns des autres, à les percevoir plutôt comme un tout.*

Cette description de l'effet de métabole permet de mieux l'appréhender et, peut être, de le reconnaître dans notre environnement quotidien. Un exemple typique d'endroit et d'ambiance métabolique sont les marchés[1] ou encore les cocktails. Cependant, elle ne permet pas de définir des critères physiques et mesurables, pouvant amener à une caractérisation de l'effet de métabole, dans un but de détection automatique dans un signal enregistré. Le second domaine dans lequel la nature de cet effet est facilement identifiable, est celui de l'acoustique physique et appliquée. Là encore, on peut tirer du répertoire des effets sonores, la description suivante :

> *L'effet de métabole est favorisé par toutes les similitudes pouvant exister au niveau de chaque critère du son : celles des timbres, des hauteurs, des intensités, des*

[1] Dans [Tix01], on trouve ce témoignage, lors d'un enregistrement *in situ*, à l'approche d'un marché : « *la rumeur du marché augmente. Ce n'est plus une rumeur continue, un brouhaha uniforme, mais une dynamique bien vivante et signifiante. On est dedans. Ici, tout est proximité et tout est changeant, tout disparaît et réapparaît, rien ne reste et tout est là. On ne distingue plus rien de continu, il n'y a plus des éléments de fond et d'autres qui restent au premier plan. On passe d'un son à un autre, sans forcément s'en rendre compte. C'est à la fois tout le temps différent et c'est à la fois tout le temps identique* ».

rythmes, des localisations... Lorsque ces paramètres sont clairement distincts pour des évènements sonores différents et simultanés, il y a moins d'ambiguïté perceptive.

V.2.2 Premières études

Différents travaux [Lav98, Tix01, Rém01] ont déjà été effectué plus spécifiquement en rapport à l'effet de métabole. Nous avons vu, dans sa définition, que cet effet est principalement caractérisé par sa composante perceptive et sociologique. C'est pourquoi les entretiens d'auditeurs plus ou moins avertis, permet de faire ressortir la perception d'un effet de métabole, lors d'un trajet donné en ville. D'un point de vue plus acoustique, dans [Tix01], l'auteur tente de développer un modèle physique permettant la synthèse de sons métaboliques, en essayant de reproduire les phénomènes sonores engendrés par les grands mouvements de dunes dans le désert. Les résultats obtenus ne permettent cependant pas de généraliser à la détection d'un fragment métabolique dans un signal sonore. Il ressort aussi de ces études, que l'effet de métabole semble surtout se caractériser par sa grande difficulté à être abordé par les outils classiques de l'acoustique physique. A ce sujet, il convient de revenir plus en détails sur le travail effectué par Laveaud [Lav98].

Les objectifs de ce travail étaient, dans un premier temps, d'établir une base de données sonores d'espaces métaboliques urbains. Le résultat de cette étude est la mise en évidence des causes circonstancielles à l'apparition de l'effet de métabole :

- D'un point de vue architectural : des espaces plutôt fermés (par du bâti ou par des dispositifs mobiles)
- D'un point de vue physique : la présence de réverbérations (non synchronisées) qui peuvent s'entendre d'un même point d'écoute
- D'un point de vue social : des activités humaines ayant un caractère collectif

Dans un second temps, il fallait tester les critères classiques de la métrologie acoustique pour voir s'ils permettaient de quantifier une structure sonore métabolique. En partant de la définition de la métabole, l'auteur s'est focalisé sur deux points susceptibles d'être caractérisés par des quantités physiques : l'instabilité de la structure sonore perçue au cours du temps et la distinctibilité des parties ou de l'ensemble. Les critères quantitatifs retenus sont :

- l'évolution temporelle des niveaux sonores en utilisant le niveau de bruit équivalent[2] L_{eq}
- l'analyse fréquentielle des évènements sonores. Elle repose sur une représentation spectrale par 1/3 d'octave entre 100 Hz et 20 kHz. L'idée est de comparer les spectres des différents évènements sonores qui constitue le signal étudié, dans une logique évolutive, de la séquence totale vers le détail sonore. Cela se fait en considérant des intervalles temporels d'analyse de plus en plus petits.
- le temps de réverbération[3] T_R

Il faut enfin noter que ces quantités servaient de représentation visuelle du signal, aucune procédure de reconnaissance automatique n'étant envisagée. A l'issue de ces analyses, principalement comparatives, sur l'ensemble des terrains d'étude, il est apparu difficile de cerner des critères

[2]Le niveau de bruit équivalent est le niveau de pression acoustique d'un bruit stable qui donnerait la même énergie acoustique que le signal étudié, pendant une durée T donnée.

[3]Le temps de réverbération est, pour un point donné éloigné de la source, la durée que met un bruit stable pour décroître de 60 dB lorsqu'il est brutalement coupé. Il est exprimé en secondes. Cette mesure peut également se faire sur une octave donnée.

métrologiques capables de caractériser toute structure métabolique à partir uniquement du signal. Un critère qui s'avère pertinent pour une situation, ne l'étant plus pour une autre.

V.2.3 Critères de caractérisation

En reprenant la seconde définition de l'effet de métabole, le lien avec l'algorithme qui a été élaboré tout au long de ce manuscrit devient évident. Il faut toutefois noter que des critères comme le rythme ou les localisations, ne peuvent pas être pris en compte. En effet, l'étude du rythme est un problème à part entière, qui sort du champ d'application de notre algorithme. En fait, avec l'estimation de fréquences fondamentales, la mesure du rythme est un pan complet de la transcription automatique de la musique [Kla06] et suscite encore beaucoup de recherche. D'autre part, nous nous sommes limités, dans cette étude, aux signaux monophoniques[4], il n'y a donc aucune spatialisation du son qui est prise en compte.

En partant de cette seconde définition, nous avons choisi de nous focaliser sur la quantification de deux similitudes : celle sur l'intensité et celle sur le timbre. A chaque instant d'analyse, l'algorithme de filtrage particulaire extrait les différentes composantes harmoniques qui constituent le signal fenêtré. La fréquence, en relation avec la fréquence fondamentale, et les amplitudes de chaque partiel de ces composantes, sont estimées. Pour $k = 1 \dots K_t$ et $h = 1 \dots H$, on note $A_{t,k,h}$ l'énergie du partiel de fréquence $f_{t,k,h}$. A partir des amplitudes estimées $a_{t,k,h}$ et $b_{t,k,h}$, elle est calculée par :

$$A_{t,k,h} = \sqrt{a_{t,k,h}^2 + b_{t,k,h}^2} \tag{V.1}$$

Les effets sonores, et donc la métabole, se situent au niveau de la perception auditive. Il paraît alors logique de chercher à reproduire la sensibilité de l'oreille humaine, dans la gamme de fréquences 20 Hz - 20 kHz, en filtrant ces énergies avec un filtre acoustique. Au vue des niveaux d'énergie considérés, la pondération A s'impose d'elle-même [IEC93]. La fonction de transfert de ce filtre pondérateur est :

$$H_A(j\omega) = \frac{7397050000\ (j\omega)^4}{(j\omega + 129.4)^2(j\omega + 676.7)(j\omega + 4636)(j\omega + 76655)^2} \tag{V.2}$$

La courbe de pondération A, en fonction de la fréquence, est donnée dans la figure V.1.

Similitude sur les amplitudes

La première quantité définie, que nous appellerons ampl_t, mesure la similitude relative entre les énergies pondérées A de chacune des composantes harmoniques présentes à l'instant t. Elle est calculée de la manière suivante :

$$\text{ampl}_t = \frac{1}{K_t} \sum_{k=1}^{K_t} (A_{t,k} - \bar{A}_t)^2 \tag{V.3}$$

avec

$$\bar{A}_t = \frac{1}{K_t} \sum_{k=1}^{K_t} A_{t,k} \quad \text{et} \quad A_{t,k} = \sum_{h=1}^{H} A_{t,k,h} \tag{V.4}$$

[4]Ici, le terme monophonique est pris comme le contraire de stéréophonique.

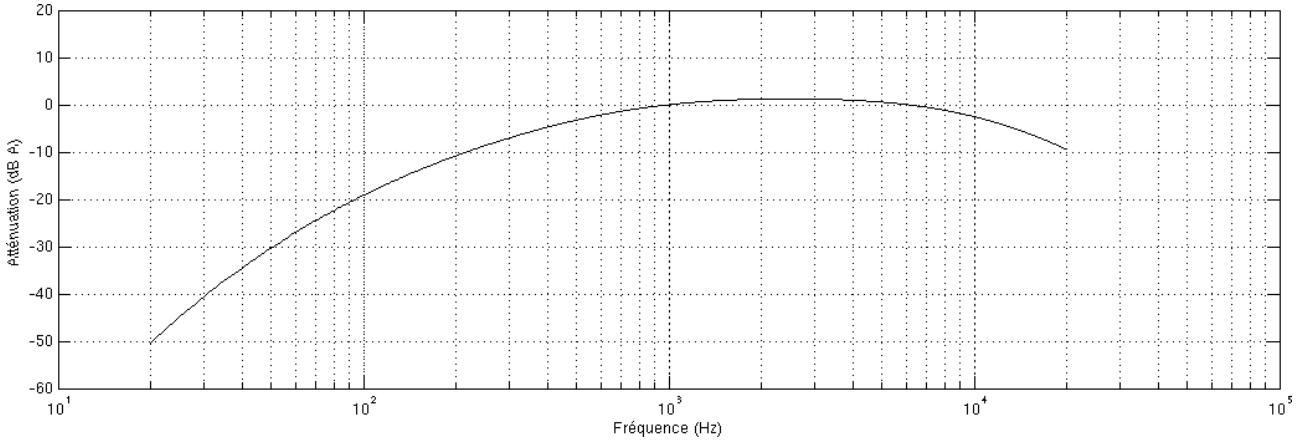


FIG. V.1 – Atténuation, en dB (A), de l'énergie en fonction de la fréquence.

Si on considère que l'énergie $A_{t,k}$ de la composante k, est une variable aléatoire, alors la quantité ampl_t est une estimation de sa variance. On peut alors dire que plus elle sera faible et plus les énergies des composantes harmoniques seront similaires.

Similitude sur les timbres

La deuxième quantité définie, que nous appellerons timb_t, mesure la similitude relative entre les profils de décroissance de l'énergie des partiels de chacune des composantes harmoniques. Elle est calculée sur le même principe que ampl_t. Plus précisément, on a :

$$\text{timb}_t = \text{trace}\left(\frac{1}{K_t}\sum_{k=1}^{K_t}(\mathbf{p}_{t,k} - \bar{\mathbf{p}}_t)(\mathbf{p}_{t,k} - \bar{\mathbf{p}}_t)^{\mathsf{T}}\right) \tag{V.5}$$

avec

$$\bar{\mathbf{p}}_t = \frac{1}{K_t}\sum_{k=1}^{K_t}\mathbf{p}_{t,k} \quad \text{et} \quad \mathbf{p}_{t,k} = \left[A_{t,k,1}\,,\ldots,\,A_{t,k,h}\,,\ldots,\,A_{t,k,H}\right]^{\mathsf{T}} \tag{V.6}$$

Là encore, si on considère que le profil de décroissance de l'énergie des partiels $\mathbf{p}_{t,k}$ de la composante k, est un vecteur aléatoire, alors la quantité timb_t est la trace de l'estimation de sa matrice de covariance. De même, plus elle sera faible et plus on pourra considérer que les profils de décroissance des différentes composantes sont similaires.

V.2.4 Application et résultats

Les critères définis dans la partie précédente, vont maintenant être testés sur un ensemble de signaux, dont certains sont métaboliques et d'autres ne le sont pas, afin d'évaluer leur capacité à rendre compte du caractère métabolique d'un signal.

Numéro	Classe	Numéro	Classe	Numéro	Classe
1	métabole	11	métabole	24	métabole
2	métabole	12	métabole	25	métabole
3	non métabole	13	métabole	26	métabole
4	non métabole	14	non métabole	27	métabole
5	non métabole	15	non métabole	28	non métabole
6	non métabole	16	non métabole	29	non métabole
8	métabole	17	non métabole	30	non métabole
9	métabole	18	non métabole	39	non métabole
10	métabole	19	non métabole	40	non métabole

TAB. V.1 – Ensemble des signaux métaboliques et non métaboliques, utilisé pour l'étape de validation. Les numéros des signaux ne se suivent pas toujours car la base de données utilisée est plus riche : elle contient aussi des signaux dits « indéterminés », qui ne sont pas considérés ici.

Base de données

Les signaux utilisés pour cette étape de validation, sont issus de la base de données du CRESSON, dont une bonne partie sont des enregistrements effectués *in situ*. Cet ensemble est constitué de 12 signaux métaboliques et de 15 signaux non métaboliques, le détail est donné dans le tableau V.1, l'identification ayant été faite par les membres du CRESSON. La troisième version de l'algorithme, décrite dans la section IV.2.3.b, est utilisée pour décomposer chacun de ces signaux en composantes harmoniques. Les choix des différents modèles sont les mêmes que pour l'application à la transcription automatique de la musique, effectuée dans la partie IV.4. Les signaux sont échantillonnés à une fréquence de 44 100 Hz et durent en moyenne 10 s. Les paramètres de l'algorithme sont les suivants :

- longueur de la fenêtre : $L_{\mathbf{w}} = 3000$ points ($\approx$ 68ms)
- écart entre deux instants d'analyse : 100 points ($\approx$ 2.2ms) le recouvrement est de 96.7%
- nombre maximum de notes : $K_{max} = 5$
- nombre minimum de notes : $K_{min} = 0$
- nombre maximum de partiels : $H = 40$
- nombre de particules : $N = 500$

Le temps de calcul moyen est d'environ 2 heures par seconde de signal.

Procédure de détection

Nous avons vu, au début de ce chapitre, que les effets sonores, et plus particulièrement l'effet de métabole, ne pouvaient pas se limiter à un des trois aspects fondamentaux (sciences humaines, acoustique appliquée et architecture et urbanisme), même si leur importance relative peut varier d'un effet à l'autre. Par conséquent, plutôt que de chercher une méthode qui permettrait de prendre la décision binaire métabole/non métabole, uniquement à partir de critères physiques, nous avons choisi d'opter pour une solution intermédiaire qui consiste à attribuer à chaque signal, à partir de critères physiques, un taux de « métabolicité », compris entre 0,

pour un signal non métabolique, et 1, pour un signal métabolique. Cette approche est justifiée par les études passées qui ont été effectuées sur l'effet de métabole, et qui mettent en avant le caractère subjectif de sa perception.

Les deux quantités physiques ampl_t et timb_t, calculées à chaque instant t, mesurent la similitude des énergies et des timbres des sources présentes simultanément. D'après la définition de l'effet de métabole, ces deux critères doivent être faibles pour un signal présentant cette caractéristique. L'idée est donc de définir deux seuils, S_{ampl} et S_{timb}, en deçà desquels le fragment de signal auquel se rapportent les deux quantités, est considéré comme métabolique. Il suffit alors de calculer la proportion de fragments métaboliques par rapport au signal complet, pour obtenir les deux taux de « métabolicité » λ_{ampl} et λ_{timb}.

Dans le but ultime de décider si un signal est métabolique ou non, il faut choisir deux autres seuils, $S_{\lambda_{\text{ampl}}}$ et $S_{\lambda_{\text{timb}}}$, au delà desquels les taux λ_{ampl} et λ_{timb} correspondent à un signal métabolique.

Evaluation des performances

Deux systèmes de reconnaissance de l'effet de métabole peuvent être définis : un qui utilise l'information de similitude des énergies (ampl_t) et l'autre qui utilise l'information de similitude des timbres (timb_t). Ces deux systèmes se basent sur des seuils pour prendre leur décision. Les courbes ROC (*Receiver Operating Characteristic*) [Dud01], permettent d'évaluer les performances de ces détecteurs, en étudiant les variations de leur spécificité et de leur sensibilité, pour différentes valeurs des seuils de discrimination.

Les courbes ROC sont construites en portant, sur l'axe des abscisses, la probabilité qu'un signal non métabolique soit considéré comme métabolique (taux de fausses alarmes), notée P_{fal}, et sur l'axe des ordonnées, la sensibilité du test, c'est-à-dire la probabilité qu'un signal métabolique soit effectivement reconnu comme tel (taux de biens détectés), notée P_{hit}. La spécificité du test est égale à $1 - P_{fal}$. La construction de la courbe se fait de manière empirique en calculant des estimées de P_{fal} et P_{hit}, pour différentes valeurs des seuils de discrimination. Dans notre cas, nous avons deux seuils à considérer : S_{ampl} et $S_{\lambda_{\text{ampl}}}$ (resp. S_{timb} et $S_{\lambda_{\text{timb}}}$) pour le détecteur se basant sur ampl_t (resp. timb_t), ce qui induit que chaque détecteur est caractérisé par une série de courbes (voir figures V.2 et V.3).

Un seuil est optimal s'il permet de séparer clairement les signaux métaboliques des signaux non métaboliques, sans commettre d'erreur. Il est donc défini par une sensibilité et une spécificité égales à 1, ce qui correspond au coin en haut à gauche des graphiques des figures V.2 et V.3. Au contraire, si le détecteur n'est pas informatif, c'est-à-dire si le critère sur lequel il se base n'est pas discriminant, il ne sera pas possible de trouver un seuil permettant de faire la détection et la courbe ROC sera confondue avec la diagonale représentée en pointillés dans les figures V.2 et V.3. Un détecteur sera donc d'autant meilleur que sa courbe ROC sera proche du coin en haut à gauche et sera éloignée de la diagonale. L'aire sous la courbe ROC, notée AUC (*Area Under Curve*), est un estimateur de l'efficacité globale du détecteur. Si elle est de 1/2, cela signifie que le détecteur fait sa discrimination de manière aléatoire, au contraire, si elle est égale à 1, cela signifie qu'il est parfaitement discriminant.

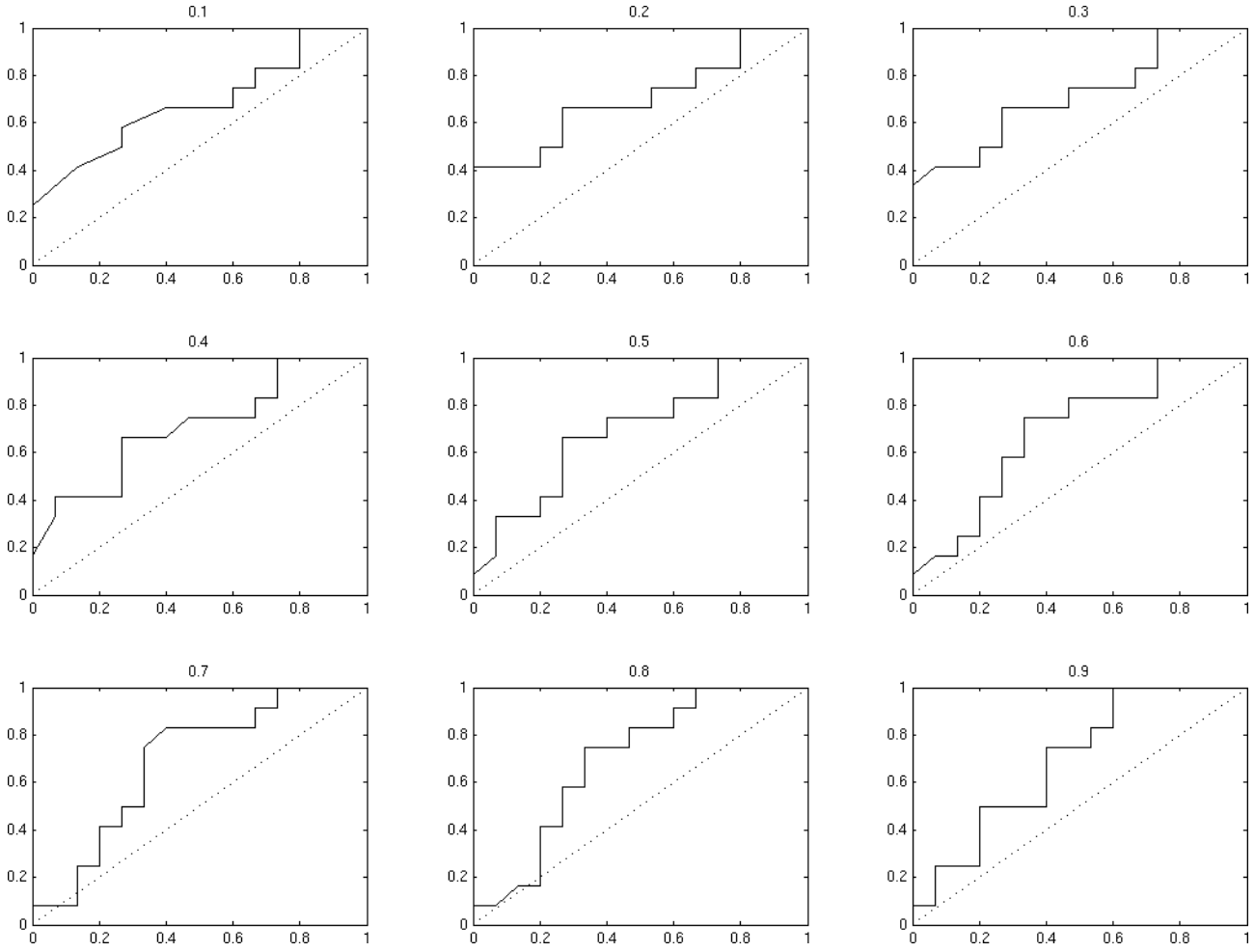


FIG. V.2 – Courbes ROC quand la détection est effectuée sur la base de la similitude des énergies. Pour chaque courbe, le seuil S_{ampl} varie entre 0 et 0.05. Le seuil $S_{\lambda_{\text{ampl}}}$ est indiqué au-dessus de chaque graphique.

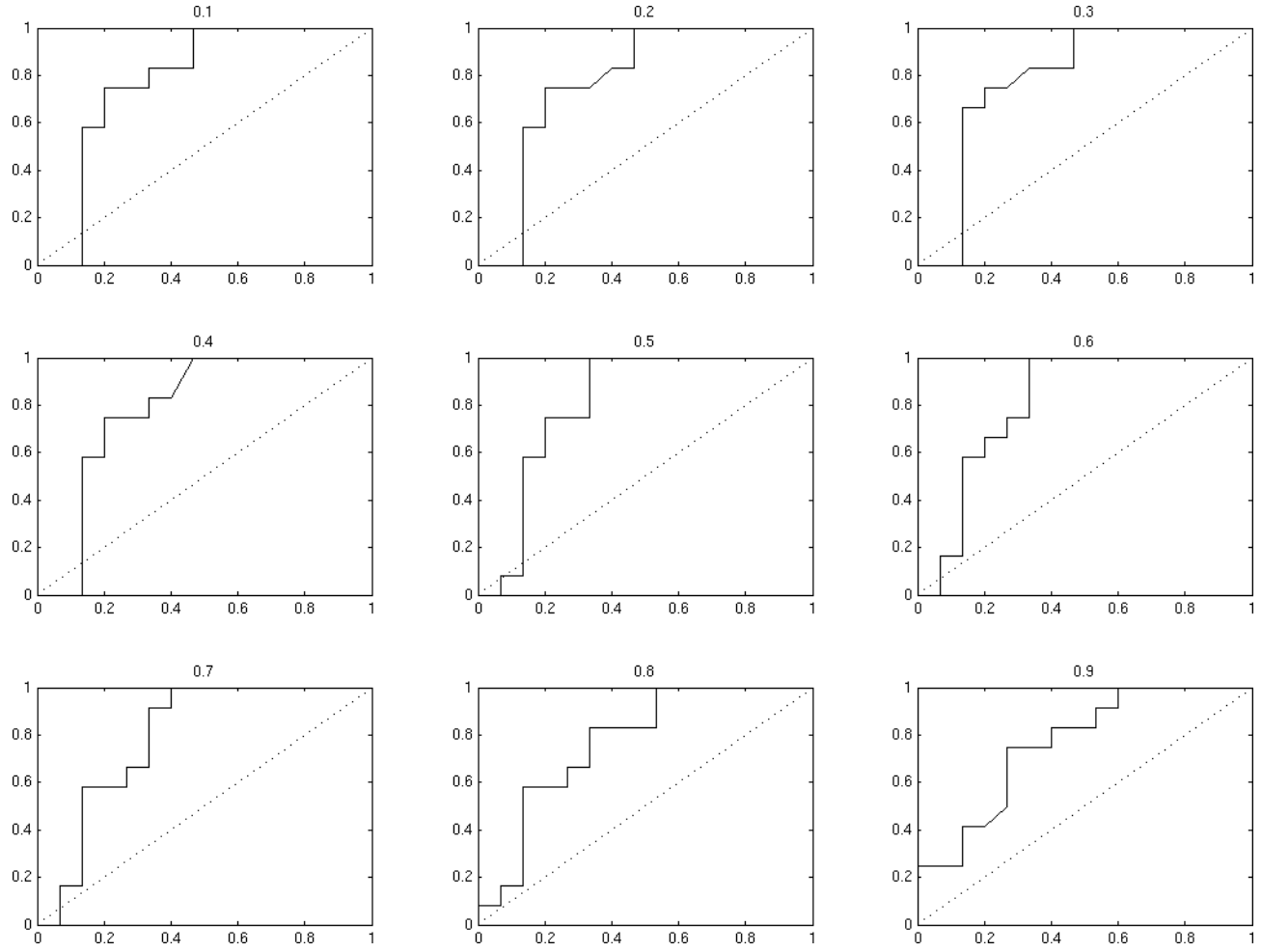


FIG. V.3 – Courbes ROC quand la détection est effectuée sur la base de la similitude des timbres. Pour chaque courbe, le seuil S_{timb} varie entre 0 et 0.04. Le seuil $S_{\lambda_{\text{timb}}}$ est indiqué au-dessus de chaque graphique.

Détecteur sur ampl

$S_{\lambda_{\text{ampl}}}$	AUC
0.1	0.6833
0.2	0.7055
0.3	0.7194
0.4	0.7083
0.5	0.6971
0.6	0.6917
0.7	0.6918
0.8	0.6972
0.9	0.6944

Détecteur sur timb

$S_{\lambda_{\text{timb}}}$	AUC
0.1	0.7835
0.2	0.7807
0.3	0.7918
0.4	0.7891
0.5	0.8113
0.6	0.8113
0.7	0.7946
0.8	0.7724
0.9	0.7639

TAB. V.2 – Aires sous les courbes ROC correspondant aux figures V.2 et V.3.

La méthode de calcul précise de cette aire est encore un thème de recherche, mais, afin d'avoir une première idée de l'efficacité des deux détecteurs mis en place, nous avons choisi d'en faire une estimation empirique en utilisant les sommes de Riemann. Les résultats sont donnés dans le tableau V.2.

Les courbes ROC et les valeurs AUC mettent bien en avant les potentialités des deux détecteurs mis en place. Le principe de courbes ROC est de voir comment évolue le compromis entre le nombre de biens détectés et le nombre de fausses alarmes, pour un système de détection donné, en fonction de la valeur du seuil de discrimination. Le choix des seuils ne se fait donc pas de manière automatique. En effet, il dépend du coût que l'utilisateur attribue aux erreurs potentielles du détecteur. S'il veut à tout prix détecter le caractère métabolique d'un signal, il choisira pour S_{ampl} (resp. S_{timb}) une valeur élevée et pour $S_{\lambda_{\text{ampl}}}$ (resp. $S_{\lambda_{\text{timb}}}$) une valeur plutôt faible, même si cela fait augmenter le taux de fausses alarmes. Et inversement.

Afin d'aller un peu plus loin dans l'étude des résultats de discrimination des 27 signaux traités, nous avons choisi deux valeurs de seuil, $S_{\text{ampl}} = 0.0052$ et $S_{\text{timb}} = 0.001$, afin de voir où se situaient les signaux dans le plan des taux de « métabolicité » $(S_{\lambda_{\text{ampl}}}, S_{\lambda_{\text{timb}}})$. Les résultats sont donnés dans la figure V.4.

Avant toute interprétation de ces résultats, il est important d'avoir à l'esprit que le choix des seuils est fait de manière arbitraire, ou plutôt, qu'il ne répond à aucune « politique »de détection donnée. On remarque néanmoins que les signaux métaboliques sont principalement concentrés autour du point $(1, 1)$, c'est-à-dire qu'ils correspondent bien aux signaux ayant les proportions les plus élevés de quantités ampl_t et timb_t faibles (en-dessous des seuils choisis). D'autres remarques peuvent être faites quant à la région du plan occupée par certains signaux. Par exemple, les signaux 8, 9 et 10 sont métaboliques et assez proches les uns des autres dans le plan. Ceci semble normal car ils sont issus du même signal et se suivent chronologiquement. Les signaux 17, 18 et 19 ne sont pas métaboliques : les deux plans sonores que l'on peut y distinguer ne se mélangent pas. Cette propriété se retrouve dans le descripteur qui mesure la similitude des timbres : elle est faible et quasi identique pour les trois signaux. Les signaux 3, 4, 5 et 6 ne sont pas métaboliques du fait de la présence de réverbération, qui lisse les

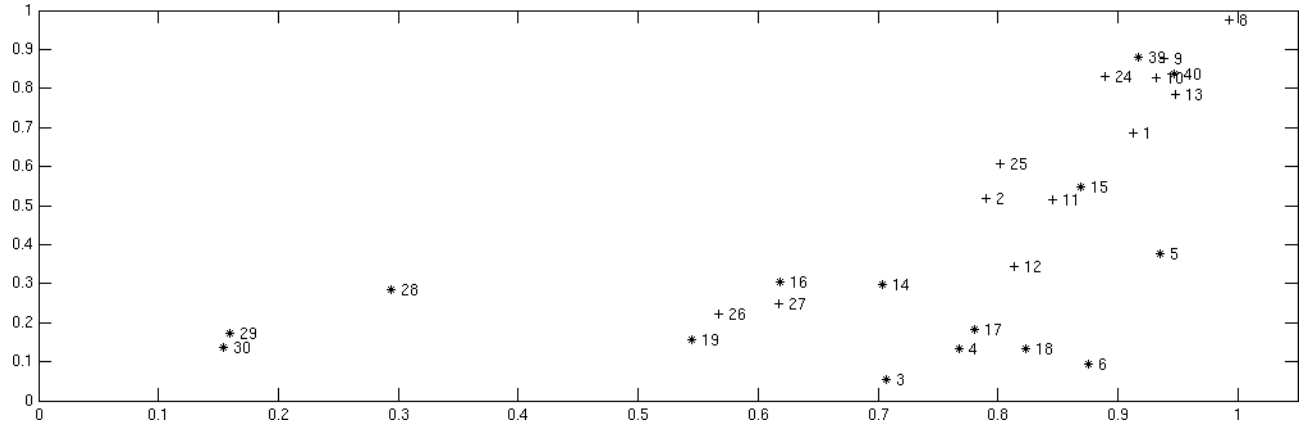


FIG. V.4 – Répartition des signaux pour un choix de seuils donné : $S_{\text{ampl}} = 0.0052$ et $S_{\text{timb}} = 0.001$. Les signaux représentés par un « + » correspondent aux métaboles et ceux représentés par un « ∗ » correspondent aux signaux non métaboliques.

interactions entre sources sonores et crée un sorte de brouhaha. Ces quatre signaux sont dans une même zone du plan. On peut noter aussi que le signal 5 semble plus près des métaboles que les autres. Peut-être la réverbération y est moins présente ? On retrouve peut-être, ici, le caractère subjectif de l'appréciation de la métabole. Cela est plus frappant pour les signaux 39 et 40. Même s'ils sont issus du même signal et se retrouvent dans la même région du plan, ils ne sont pas considérés comme métaboliques alors qu'ils se situent près du point $(1, 1)$. Au vue des résultats, on pourrait croire qu'il s'agit de cas pathologiques. En recommençant l'étude précédente mais en les enlevant de la base, les courbes ROC, figures V.5 et V.6, et les estimations des aires sous la courbe, tableau V.3, obtenues sont bien meilleures.

Détecteur sur ampl

$S_{\lambda_{\text{ampl}}}$	AUC
0.1	0.7116
0.2	0.7501
0.3	0.7692
0.4	0.7629
0.5	0.7693
0.6	0.7756
0.7	0.7852
0.8	0.7852
0.9	0.7756

Détecteur sur timb

$S_{\lambda_{\text{timb}}}$	AUC
0.1	0.9070
0.2	0.9006
0.3	0.9166
0.4	0.9134
0.5	0.9294
0.6	0.9230
0.7	0.9038
0.8	0.8749
0.9	0.8429

TAB. V.3 – Aires sous les courbes ROC correspondant aux figures V.5 et V.6.

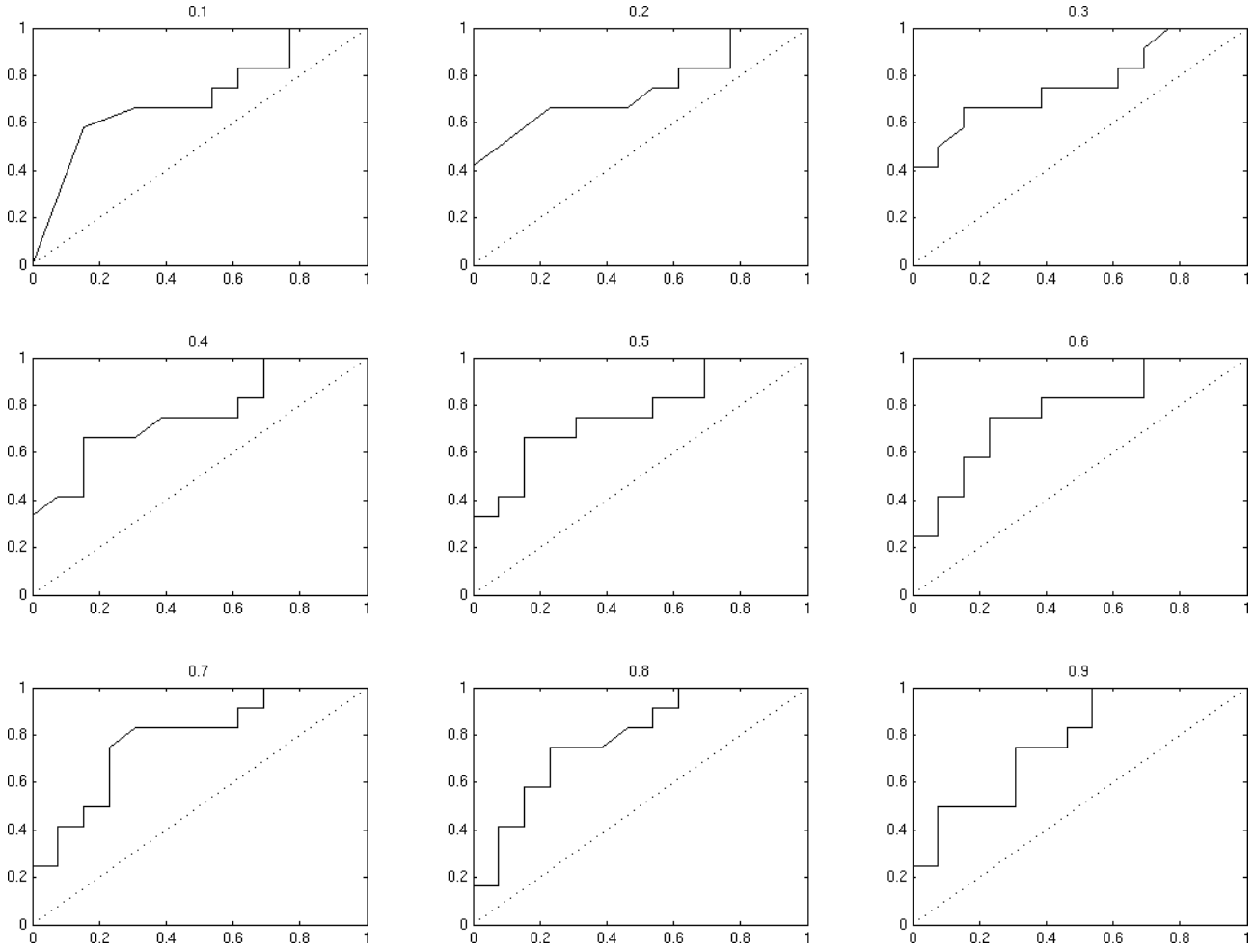


FIG. V.5 – Courbes ROC quand la détection est effectuée sur la base de la similitude des énergies. Pour chaque courbe, le seuil S_{ampl} varie entre 0 et 0.05. Le seuil $S_{\lambda_{\text{ampl}}}$ est indiqué au-dessus de chaque graphique. Les signaux 39 et 40 ont été enlevés de la base.

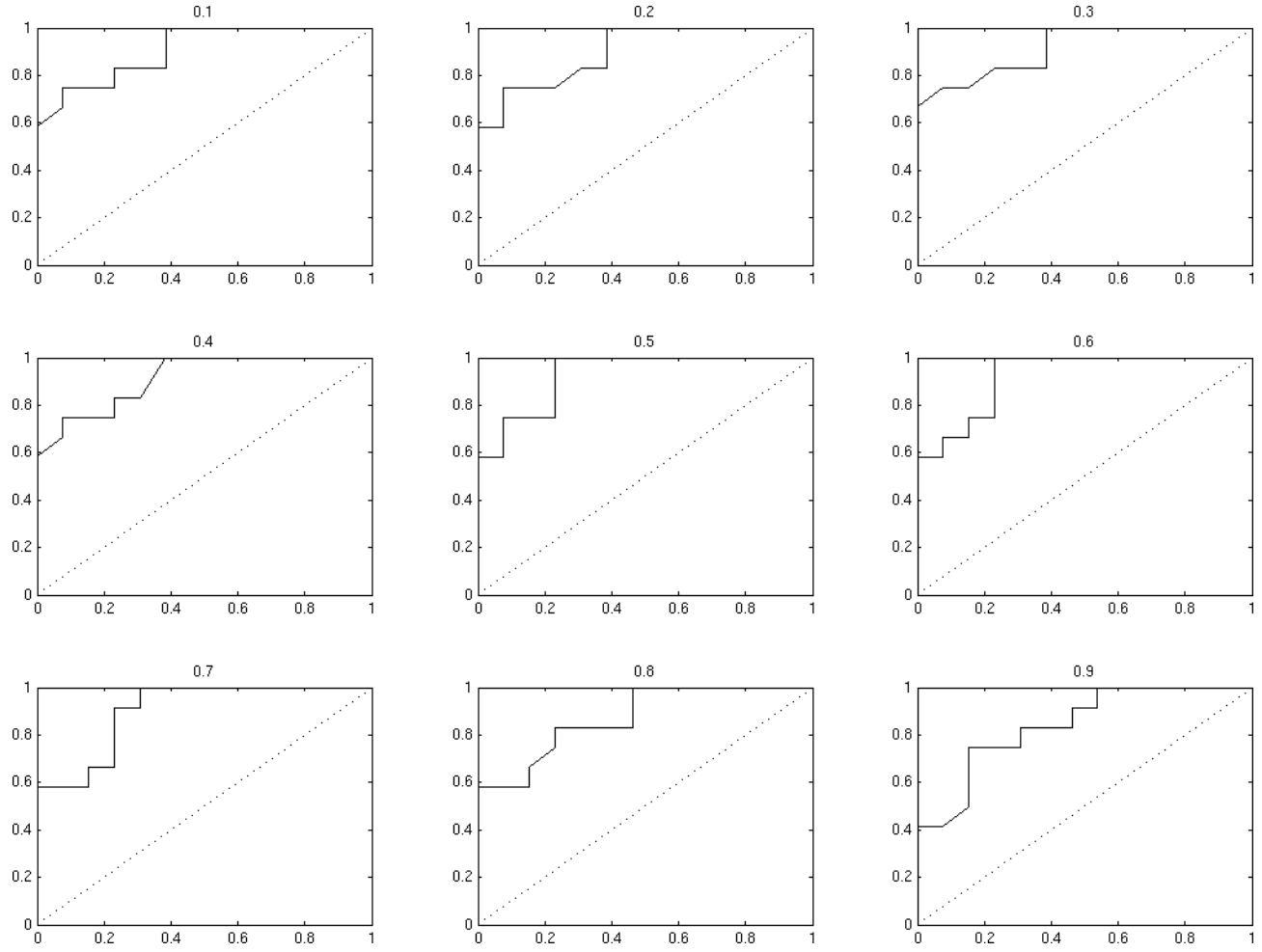


FIG. V.6 – Courbes ROC quand la détection est effectuée sur la base de la similitude des timbres. Pour chaque courbe, le seuil S_{timb} varie entre 0 et 0.04. Le seuil $S_{\lambda_{\text{timb}}}$ est indiqué au-dessus de chaque graphique. Les signaux 39 et 40 ont été enlevés de la base.

V.3 Bilan

La caractérisation de l'effet de métabole que nous proposons, se base sur l'évaluation de deux critères physiques, l'un mesurant la similitude des énergies et l'autre celle des timbres, à un instant donné, des sources présentes simultanément dans le signal. A partir de ces deux quantités, un système de détection du caractère métabolique est mis en place, sur la base de deux seuils. Il est important de noter que la décision prise avec le seuil S_{ampl} ou S_{timb}, porte sur le caractère métabolique d'une portion de signal tandis que celle prise avec le seuil λ_{ampl} ou λ_{timb}, porte sur le signal dans son intégralité. Ce processus de détection, incluant deux échelles d'analyse différentes, a deux avantages. Tout d'abord, il permet de s'affranchir de divers phénomènes ponctuels, qui pourraient perturber la reconnaissance de l'effet de métabole. Par exemple, dans le signal 9, qui est métabolique, il y a une émergence qui survient à un moment donné, c'est-à-dire que, temporairement, une des sources présentes dans le signal se détache plut nettement du fond sonore. Cette courte émergence ne remet pas en cause le caractère métabolique du signal, mais se pose alors la question de la durée minimale que doit avoir un signal sonore, pour pouvoir raisonnablement y détecter un effet de métabole. Cette question est fondamentale mais n'est pas du ressort du processus de détection, et c'est là que réside le second avantage de l'approche proposée. En effet, elle peut être considérée comme une méthode objective, car basée sur des critères physiques, d'aide à la reconnaissance de l'effet de métabole. Or, la réponse à la question de la durée minimale nécessaire à l'instauration de cet effet, relève peut-être plus de la subjectivité. Cette relativité, liée à l'auditeur, peut néanmoins être introduite dans le processus de détection, au travers du choix des seuils, et peut-être plus spécifiquement du second.

Les performances encourageantes obtenues, invitent à creuser plus profondement dans la direction proposée, afin d'arriver, peut être, à une meilleure caractérisation de l'effet de métabole. En effet, nous avons vu que le label « métabole » peut correspondre à une grande diversité de signaux, le caractère métabolique pouvant se retrouver à différents niveaux. Plusieurs paramètres physiques supplémentaires peuvent alors être pris en compte. Par exemple, la spatialisation de la perception auditive peut être considérée en traitant des signaux stéréophoniques et en comparant les résultats obtenus sur chacun des canaux. Le système de détection peut aussi être amélioré. En effet, les deux détecteurs définis dans la partie précédente, sont utilisés en parallèle. La conception d'un détecteur conjoint, dans lequel la frontière entre les deux catégories de signaux pourrait être plus compliquée qu'une simple frontière linéaire, permettrait sans doute d'optimiser les résultats. Enfin, le traitement amont, l'extraction des composantes harmoniques, qui est effectué par l'algorithme élaboré durant cette thèse, peut certainement être amélioré, notamment en ce qui concerne le temps de calcul. Nous avons mentionner qu'il fallait, en moyenne, deux heures pour traiter une seconde de signal[5], ce qui nous amène à environ 22 jours de calcul pour les 27 signaux considérés. Même si, à cette étape du développement de l'algorithme, des optimisations peuvent être envisagées, cette contrainte a obligatoirement une conséquence sur le nombre de signaux, métaboliques ou non, traités et donc sur le design du détecteur. En effet,

[5]Il est important de rappeler que le temps de calcul dépend de plusieurs paramètres : la fréquence d'échantillonnage du signal traité, la longueur de la fenêtre d'analyse, le nombre de particules. Compte tenu de cela, on peut considérer que deux heures pour une seconde de signal est le temps de calcul maximum. De plus, à ces paramètres, il faut ajouter la « complexité du signal ». En effet, s'il n'est composé que d'une seule source, à chaque instant, le temps de calcul est plus faible.

face à la diversité qui peut exister entre deux signaux métaboliques, une évaluation des critères de caractérisation doit se faire sur un nombre plus élevé d'exemples. On peut néanmoins considérer que, s'ils ne sont pas capables de rendre compte de toute la diversité que peut recouvrir la métabole sonore, les deux critères proposés dans ce manuscrit peuvent peut être en décrire un certain type.

Conclusion

On fait la science avec des faits,
comme on fait une maison avec des pierres,
mais une accumulation de faits n'est pas plus une science
qu'un tas de pierres n'est une maison.
H. Poincaré

A l'origine de ce travail de thèse, se trouve la question de la reconnaissance automatique d'un effet sonore, la métabole, dans des signaux audio issus du milieu urbain. A partir de ce problème concret, nous avons voulu nous placer à un niveau plus fondamental pour y répondre, afin d'élargir le champs applicatif de la solution proposée et de pouvoir la transposer à d'autres problèmes. Le fil rouge que nous avons essayé de suivre tout au long de notre travail, est la généralité, c'est-à-dire la volonté de développer une méthode qui ne se restreint pas à une application donnée ou à un type de signal prédéfini. En effet, la thématique dans laquelle nous nous situons porte sur l'extraction du contenu informationnel d'un signal audio, à des fins de décision ou de classification, et trouve des applications qui sortent du seul domaine des effets sonores. Le dénominateur commun est la nécessité de caractériser la perception que l'on peut avoir d'un signal sonore, à l'aide de quantités physiques, les ordinateurs n'étant pas capables de manipuler des descripteurs qualitatifs.

A l'issue du premier chapitre, nous avons mis en évidence l'importance du triplet de quantités physiques constitué de la fréquence fondamentale, de l'énergie et de la structure fréquentielle, pour décrire le signal issu d'une source sonore harmonique ou pseudo harmonique. L'étude bibliographique menée dans le second chapitre montre que l'estimation de ce triplet, au cours du temps, n'est pas un problème simple surtout si on considère des signaux polyphoniques quelconques, c'est-à-dire constitués de plusieurs sources dont les propriétés ne sont pas connues *a priori*. En fait, les auteurs ont souvent recours à des hypothèses simplificatrices[1], qui éliminent une partie de l'intérêt du problème initial. Le constat que l'on peut néanmoins faire est qu'il est nécessaire d'inclure certaines connaissances *a priori* pour mener à bien ce problème d'estimation séquentielle. C'est là que se situe une difficulté majeure, car ces connaissances et la manière de les prendre en compte, peuvent être très spécifique à un problème donné.

[1] Comme considérer que le nombre de sources est connu ou alors inconnu mais fixe au cours du temps. Une autre hypothèse est de choisir de travailler sur un type de source donné (un instrument particulier, par exemple).

Contributions

La nécessité d'estimer conjointement le nombre de sources et le triplet qui les caractérise, à chaque instant, et la recherche d'une solution efficace qui ne pose pas ou peu d'hypothèses simplificatrices, nous ont conduit au choix de l'approche paramétrique dans le paradigme bayésien. La construction du modèle s'est faite en plusieurs étapes, pour aboutir à une solution dans laquelle sont prises en compte toutes les caractéristiques temporelles et fréquentielles des sources. Le modèle retenu est un modèle de Markov à sauts dont les paramètres, non directement observés, doivent être estimés à partir du signal temporel. Le nombre de ces paramètres, à chaque instant, est assez élevé[2] et les équations qui permettent de les relier entre eux ou au signal temporel, sont non linéaires.

Ces propriétés du modèle statistique, associées à la contrainte d'une estimation séquentielle, nous ont assez naturellement amené aux méthodes de Monte Carlo séquentielles. C'est dans ce cadre que nous avons développé un algorithme de filtrage particulaire, capable de s'affranchir des principales difficultés inhérentes au contexte polyphonique et suffisamment flexible pour pouvoir s'adapter au niveau de connaissance, plus ou moins élevé, que l'on peut avoir sur le signal traité. Le cœur de cette approche réside dans le design d'une fonction de proposition et, à cette fin, nous avons proposé une solution simple, répondant au même souci de généralité que celui qui nous a animé tout au long de l'élaboration de l'algorithme. Une application typique de l'estimation de fréquences fondamentales dans des signaux polyphoniques, est la transcription automatique de la musique. Il était donc naturel de valider notre méthode en essayant de traiter ce problème. Même si l'objectif initial n'était pas de le résoudre, les résultats obtenus sont intéressants et encourageants.

Le dernier chapitre de ce manuscrit est consacré à l'application de notre algorithme à la reconnaissance de l'effet de métabole. A cet effet, nous avons défini deux critères capables de décrire deux caractéristiques précises des métaboles. Ces quantités sont estimées à partir des résultats que donne l'algorithme de filtrage particulaire et servent d'entrée à un système de détection de la métabole. Son principe repose sur deux seuils qui permettent d'effectuer une analyse à deux échelles différentes. Les résultats obtenus sur le corpus étudié offrent des perspectives d'approfondissement, qui apporte une première validation à la méthode proposée.

Retour sur nos choix

Une modélisation paramétrique du signal sonore, est un choix *a priori* qui peut être discuté. Il est vrai qu'en faisant ce choix, nous nous interdisons le traitement des signaux ne comportant pas de partie harmonique ou essentiellement caractérisés par leur rythmique. Cependant, au sein de l'ensemble des signaux concernés, c'est-à-dire contenant une partie harmonique ou quasi harmonique, ce qui est quand même le cas de la majorité des signaux qui nous entourent, l'approche paramétrique présente l'avantage de fournir une description précise et synthétique du signal. Cette information peut être aisément manipulée en fonction de l'application en vue, comme nous avons pu le montrer dans ce manuscrit. Par contre, une difficulté inhérente à utilisation de modèles peut être la généralité. Or, dès le début, nous avons voulu avoir une

[2]Si on considère que, à un instant donné, le signal est composé de 3 sources et que chacune de ces sources a 40 partiels, le modèle est alors caractérisé par 360 paramètres.

approche généraliste afin d'être capable de traiter un large panel de signaux. Apparemment paradoxale, cette situation n'est cependant pas insoluble. En fait, les modèles choisis et surtout la méthode utilisée pour estimer leurs paramètres, sont suffisemment flexibles pour pouvoir s'adapter à différentes situations. On peut alors se demander s'il est possible de concevoir une méthode permettant l'extraction et la caractérisation du contenu informationnel d'un signal, applicable dans tous les cas. La réponse est bien évidemment négative mais l'idée derrière notre démarche est de proposer une méthode de traitement qui fonctionne relativement bien lorsque l'on n'a pas beaucoup d'informations sur le signal étudié et qui peut être de plus en plus raffiner, en fonction des connaissances *a priori*.

Un autre aspect lié à la modélisation paramétrique est la méthode d'estimation des paramètres du modèle. Dans notre cas, les méthodes de Monte Carlo séquentielles se sont imposées assez naturellement. Le cœur des algorithmes de filtrage particulaire repose sur la fonction de proposition utilisée pour mettre à jour les particules. En effet, elle a un rôle important dans les performances générales de l'algorithme. Dans ce manuscrit, nous avons développé une telle fonction et les résultats obtenus sont satisfaisants. Comme nous l'avons déjà dit, il est possible d'utiliser une méthode existante, même si elle nécessite plus d'hypothèses simplificatrices, l'idée étant de se dire que le paradigme bayésien que notre algorithme rajoute à cette méthode, a de forte chance d'améliorer les résultats que la méthode seule aurait donnés. Il faut alors se poser la question : quand faut-il mettre en œuvre l'algorithme de filtrage particulaire et quand faut-il se contenter d'une solution plus simple ? La réponse est liée à la nature des signaux traités et à l'information que l'on veut en extraire. Il est clair que, malgré la flexibilité et la capacité d'adaptation de notre algorithme, si on cherche à estimer la fréquence fondamentale d'un signal musical dans lequel, à chaque instant, un seul instrument est présent et si, en plus, on connait les caractéristiques de cet instrument, notre approche, de par sa lourdeur, n'est pas adaptée.

Perspectives

Les axes de développement que nous pouvons envisager, suite au travail présenté dans ce manuscrit, portent sur trois points.

D'abord, au niveau de l'algorithme en lui-même, plusieurs améliorations peuvent être apportées, notament en ce qui concerne le temps de calcul. Ici, se trouve certainement une des principales limites de notre approche. Cependant, au vue des progrès réalisés, ne serait-ce qu'au niveau des trois versions de l'algorithme, nous avons bon espoir d'arriver à une version plus rapide, à plus ou moins court terme. Les pistes envisageables peuvent bien sûr être du côté de la modélisation elle-même, afin de réduire ou de supprimer les calculs lourds mais le principal travail reste du côté de l'optimisation du codage de la méthode. La recherche d'approximations moins coûteuses en temps de calcul et qui ne dégraderaient pas les performances générales de l'algorithme, est aussi une piste à explorer. Une autre amélioration qui suscite notre intérêt est le lien qui existe entre la fonction de proposition et l'algorithme particulaire. Aujourd'hui, si la fonction de proposition ne détecte pas une composante harmonique, l'algorithme particulaire met un certain temps avant de palier à ce problème. Afin d'augmenter se réactivité, une solution serait d'inclure une étape de MCMC (*Monte Carlo Markov Chain*), ce qui permettrait de faire des sauts plus grands dans l'espace d'état, et donc de l'explorer plus rapidement.

Le deuxième point sur lequel il serait intéressant de travailler, concerne l'application de l'al-

gorithme au problème de la transcription automatique de la musique. En particulier, il semble parfaitement envisageable d'inclure une étape de reconnaissance de l'instrument, basé sur un modèle de timbre. La conséquence de cette prise en compte se traduirait par une amélioration de la robustesse de l'estimation du nombre de composantes harmoniques et par une meilleure capacité à détecter les notes en relation d'octave. Enfin, un modèle de timbre faciliterait grandement le suivi des trajectoires. Par exemple, supposons qu'un violon et qu'une trompette jouent ensemble et que la première note estimée corresponde au violon et la seconde à la trompette. Tant que les instruments jouent, il n'y a pas de confusion des trajectoires mais, aujourd'hui, après un temps de silence, il n'y a aucune raison que la première note estimée corresponde toujours au violon.

Le dernier point porte sur la reconnaissance de la métabole. Dans ce problème, si un premier pas à déjà été franchi avec l'approche proposée dans ce manuscrit, il reste encore de nombreuses étapes à franchir. En particulier en ce qui concerne l'information prise en compte par le système de détection. Dans les études préliminaires qui ont été menées par le CRESSON, il est mentionné que des paramètres physiques indépendants du signal sonore doivent être pris en compte. Il serait aussi intéressant de quantifier l'influence de l'architecture afin qu'elle puisse aussi être prise en compte dans un processus de détection automatique de l'effet de métabole.

Annexe A

Equations du filtre de Kalman

Commençons par quelques notations. L'expression $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ signifie que le vecteur aléatoire $\mathbf{x}$, de dimension n, est distribué selon une distribution normale de moyenne $\boldsymbol{\mu}$ et de matrice de covariance $\boldsymbol{\Sigma}$. Dans la suite de cette annexe, la notation $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ sera aussi utilisée pour désigner la densité de probabilité :

$$\frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{A.1}$$

en particulier, lors de l'usage de l'abus de notation $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{\mu}; \mathbf{x}, \boldsymbol{\Sigma})$.

Théorème 1
Soient

$\boldsymbol{\mu}_1$	*un vecteur de dimension* n_1
$\boldsymbol{\Sigma}_1$	*une matrice symétrique définie positive, de dimension* $n_1 \times n_1$
$\boldsymbol{\mu}_2$	*un vecteur de dimension* n_2
$\boldsymbol{\Sigma}_2$	*une matrice symétrique définie positive, de dimension* $n_2 \times n_2$
$\mathbf{Q}$	*une matrice rectangulaire, de dimension* $n_1 \times n_2$
$\mathbf{x}$	*un vecteur aléatoire, de dimension* n_2

Alors

$$\mathcal{N}(\mathbf{Q}\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\mathcal{N}(\mathbf{Q}\boldsymbol{\mu}_2; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1 + \mathbf{Q}\boldsymbol{\Sigma}_2\mathbf{Q}^{\mathrm{T}})$$

avec

$$\boldsymbol{\Sigma} = (\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\mathbf{Q} + \boldsymbol{\Sigma}_2^{-1})^{-1} \quad \textit{et} \quad \boldsymbol{\mu} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1 - \mathbf{Q}\boldsymbol{\mu}_2)$$

La démonstration de ce théorème nécessite quelques développements calculatoires et ne présente pas de difficulté particulière. Rappelons la formule récursive de la densité de filtrage à l'instant $t+1$, en fonction de celle à l'instant t :

$$p(\boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t}) = \int p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t})\mathrm{d}\boldsymbol{\theta}_t \tag{A.2}$$

$$p(\boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t+1}) = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\theta}_{t+1})p(\boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t})}{\int p(\mathbf{y}_{t+1}|\boldsymbol{\theta}_{t+1})p(\boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t})\mathrm{d}\boldsymbol{\theta}_{t+1}} \tag{A.3}$$

A l'instant $t = 0$, il n'y a pas d'observation. On pose alors $\mathbf{y}_{1:0} = \emptyset$ et $p(\boldsymbol{\theta}_0|\mathbf{y}_{1:0}) = p(\boldsymbol{\theta}_0)$. En reprenant les notations des équations (III.26) et (III.27) et sous les hypothèses gaussienne et linéaire, on a :

$$\begin{aligned} p(\boldsymbol{\theta}_0) &= \mathcal{N}(\boldsymbol{\theta}_0; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) && \text{(A.4)} \\ p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t) &= \mathcal{N}(\boldsymbol{\theta}_{t+1}; \mathbf{A}\boldsymbol{\theta}_t, \mathbf{B}\mathbf{B}^{\mathrm{T}}) && \text{(A.5)} \\ p(\mathbf{y}_{t+1}|\boldsymbol{\theta}_{t+1}) &= \mathcal{N}(\mathbf{y}_{t+1}; \mathbf{C}\boldsymbol{\theta}_{t+1}, \mathbf{D}\mathbf{D}^{\mathrm{T}}) && \text{(A.6)} \end{aligned}$$

On suppose que la densité de filtrage à l'instant t est de la forme :

$$p(\boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \tag{A.7}$$

Cette hypothèse est vraie à l'instant $t = 0$, en posant $\boldsymbol{\mu}_{0|0} \equiv \boldsymbol{\mu}_0$ et $\boldsymbol{\Sigma}_{0|0} \equiv \boldsymbol{\Sigma}_0$. Il s'agit maintenant de trouver une formule récursive permettant d'exprimer les deux premiers moments de la densité de filtrage à l'instant $t+1$ en fonction de ceux de la densité de filtrage à l'instant t. En appliquant le théorème 1 à l'équation (A.2) et en utilisant (A.5) et (A.7), on obtient :

$$\begin{aligned} p(\boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t}) &= \int \mathcal{N}(\boldsymbol{\theta}_{t+1}; \mathbf{A}\boldsymbol{\theta}_t, \mathbf{B}\mathbf{B}^{\mathrm{T}}) \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \mathrm{d}\boldsymbol{\theta}_t \\ &= \int \mathcal{N}(\mathbf{A}\boldsymbol{\theta}_t; \boldsymbol{\theta}_{t+1}, \mathbf{B}\mathbf{B}^{\mathrm{T}}) \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \mathrm{d}\boldsymbol{\theta}_t \\ &= \mathcal{N}(\mathbf{A}\boldsymbol{\mu}_{t|t}; \boldsymbol{\theta}_{t+1}, \mathbf{A}\boldsymbol{\Sigma}_{t|t}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}}) \int \mathcal{N}(\boldsymbol{\theta}_t; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathrm{d}\boldsymbol{\theta}_t \\ &= \mathcal{N}(\boldsymbol{\theta}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t}) && \text{(A.8)} \end{aligned}$$

avec

$$\begin{aligned} \boldsymbol{\mu}_{t+1|t} &= \mathbf{A}\boldsymbol{\mu}_{t|t} && \text{(A.9)} \\ \boldsymbol{\Sigma}_{t+1|t} &= \mathbf{A}\boldsymbol{\Sigma}_{t|t}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}} && \text{(A.10)} \end{aligned}$$

Les valeurs exactes de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ sont données par le théorème 1. Elles présentent cependant peu d'intérêt car l'intégrale d'une loi gaussienne normalisée est égale à 1. On a ainsi pu mettre en évidence que $p(\boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t})$ est une gaussienne dont les deux premiers moments peuvent être calculés analytiquement. De même, en appliquant le théorème 1 à l'équation (A.3) et en utilisant (A.6) et (A.8), on obtient :

$$\begin{aligned} p(\boldsymbol{\theta}_{t+1}|\mathbf{y}_{1:t+1}) &= \frac{\mathcal{N}(\mathbf{y}_{t+1}; \mathbf{C}\boldsymbol{\theta}_{t+1}, \mathbf{D}\mathbf{D}^{\mathrm{T}}) \mathcal{N}(\boldsymbol{\theta}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t})}{\int \mathcal{N}(\mathbf{y}_{t+1}; \mathbf{C}\boldsymbol{\theta}_{t+1}, \mathbf{D}\mathbf{D}^{\mathrm{T}}) \mathcal{N}(\boldsymbol{\theta}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t}) \mathrm{d}\boldsymbol{\theta}_{t+1}} \\ &= \frac{\mathcal{N}(\mathbf{C}\boldsymbol{\theta}_{t+1}; \mathbf{y}_{t+1}, \mathbf{D}\mathbf{D}^{\mathrm{T}}) \mathcal{N}(\boldsymbol{\theta}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t})}{\int \mathcal{N}(\mathbf{C}\boldsymbol{\theta}_{t+1}; \mathbf{y}_{t+1}, \mathbf{D}\mathbf{D}^{\mathrm{T}}) \mathcal{N}(\boldsymbol{\theta}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t}) \mathrm{d}\boldsymbol{\theta}_{t+1}} \\ &= \frac{\mathcal{N}(\boldsymbol{\theta}_{t+1}; \boldsymbol{\mu}_{t+1|t+1}, \boldsymbol{\Sigma}_{t+1|t+1}) \mathcal{N}(\mathbf{C}\boldsymbol{\mu}_{t+1|t}; \mathbf{y}_{t+1}, \mathbf{D}\mathbf{D}^{\mathrm{T}} + \mathbf{C}\boldsymbol{\Sigma}_{t+1|t}\mathbf{C}^{\mathrm{T}})}{\int \mathcal{N}(\boldsymbol{\theta}_{t+1}; \boldsymbol{\mu}_{t+1|t+1}, \boldsymbol{\Sigma}_{t+1|t+1}) \mathcal{N}(\mathbf{C}\boldsymbol{\mu}_{t+1|t}; \mathbf{y}_{t+1}, \mathbf{D}\mathbf{D}^{\mathrm{T}} + \mathbf{C}\boldsymbol{\Sigma}_{t+1|t}\mathbf{C}^{\mathrm{T}}) \mathrm{d}\boldsymbol{\theta}_{t+1}} \\ &= \mathcal{N}(\boldsymbol{\theta}_{t+1}; \boldsymbol{\mu}_{t+1|t+1}, \boldsymbol{\Sigma}_{t+1|t+1}) && \text{(A.11)} \end{aligned}$$

avec

$$\boldsymbol{\Sigma}_{t+1|t+1} = (\boldsymbol{\Sigma}_{t+1|t}^{-1} + \mathbf{C}^{\mathrm{T}}(\mathbf{D}\mathbf{D}^{\mathrm{T}})^{-1}\mathbf{C})^{-1} \tag{A.12}$$

$$\boldsymbol{\mu}_{t+1|t+1} = \boldsymbol{\mu}_{t+1|t} + \boldsymbol{\Sigma}_{t+1|t+1}\mathbf{C}^{\mathrm{T}}(\mathbf{D}\mathbf{D}^{\mathrm{T}})^{-1}(\mathbf{y}_{t+1} - \mathbf{C}\boldsymbol{\mu}_{t+1|t}) \tag{A.13}$$

En utilisant la formule de Woodbury :

$$(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^{\mathrm{T}})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{A}^{-1} \tag{A.14}$$

et sa variante, si $\mathbf{P}$ et $\mathbf{R}$ sont définies positives :

$$(\mathbf{P}^{-1} + \mathbf{B}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{P}\mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{P}\mathbf{B}^{\mathrm{T}} + \mathbf{R})^{-1} \tag{A.15}$$

et en posant

$$\text{Gain} \Rightarrow \mathbf{S}_{t+1|t} = \mathbf{C}\boldsymbol{\Sigma}_{t+1|t}\mathbf{C}^{\mathrm{T}} + \mathbf{D}\mathbf{D}^{\mathrm{T}} \tag{A.16}$$

$$\text{Innovation} \Rightarrow \mathbf{y}_{t+1|t} = \mathbf{C}\boldsymbol{\mu}_{t+1|t} \tag{A.17}$$

les équations (A.12) et (A.13) se simplifient pour donner :

$$\boldsymbol{\mu}_{t+1|t+1} = \boldsymbol{\mu}_{t+1|t} + \boldsymbol{\Sigma}_{t+1|t}\mathbf{C}^{\mathrm{T}}\mathbf{S}_{t+1|t}^{-1}(\mathbf{y}_{t+1} - \mathbf{y}_{t+1|t}) \tag{A.18}$$

$$\boldsymbol{\Sigma}_{t+1|t+1} = \boldsymbol{\Sigma}_{t+1|t} - \boldsymbol{\Sigma}_{t+1|t}\mathbf{C}^{\mathrm{T}}\mathbf{S}_{t+1|t}^{-1}\mathbf{C}\boldsymbol{\Sigma}_{t+1|t} \tag{A.19}$$

Ainsi, nous avons montré que le caractère gaussien de la densité de filtrage demeurait au cours des itérations et que ses deux premiers moments pouvaient être calculés analytiquement.

Annexe B

Transformée sans parfum

Soient $\mathbf{x}$ un vecteur aléatoire de dimension n_x, de moyenne $\boldsymbol{\mu}_x = \mathbb{E}\{\mathbf{x}\}$ et de matrice de covariance $\boldsymbol{\Sigma}_x = \mathbb{E}\{(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^\mathsf{T}\}$ et $g : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_y}$ une fonction non linéaire.

La transformée sans parfum tente de résoudre le problème de la propagation d'une distribution de probabilité au travers d'une fonction non linéaire. Plus précisément, elle calcule une estimation des deux premiers moments de la distribution du vecteur aléatoire $\mathbf{y}$, de dimension n_y, définit par $\mathbf{y} = g(\mathbf{x})$. La méthode opère en trois étapes :

- Etape 1 : construction d'un ensemble de $2n_x + 1$ échantillons capturant totalement l'information de moyenne et de covariance du vecteur $\mathbf{x}$
- Etape 2 : propagation de cet ensemble au travers de la fonction g
- Etape 3 : estimation de la moyenne et de la covariance du vecteur $\mathbf{y}$, à partir de l'ensemble transformé

Ces trois étapes sont maintenant détaillées.

Etape 1

L'information portée par la moyenne $\boldsymbol{\mu}_x$ et la matrice de covariance $\boldsymbol{\Sigma}_x$ du vecteur $\mathbf{x}$ est résumée par un ensemble de $2n_x + 1$ échantillons pondérés, appelés *sigma points*, construit de manière déterministe par :

$$\mathcal{X}_0 = \boldsymbol{\mu}_x \tag{B.1}$$

$$\mathcal{X}_i = \boldsymbol{\mu}_x + \sqrt{n_x + \kappa}(\mathbf{L}_x)_i \quad \text{pour } i = 1, \ldots, n_x \tag{B.2}$$

$$\mathcal{X}_i = \boldsymbol{\mu}_x - \sqrt{n_x + \kappa}(\mathbf{L}_x)_i \quad \text{pour } i = n_x + 1, \ldots, 2n_x \tag{B.3}$$

où $\kappa \in \mathbb{R}$ est un paramètre de mise à l'échelle et $(\mathbf{L}_x)_i$ est la $i^{\text{ème}}$ colonne de la matrice $\mathbf{L}_x$ définie par $\boldsymbol{\Sigma}_x = \mathbf{L}_x\mathbf{L}_x$ ($\mathbf{L}_x$ est la racine carrée matricielle de $\boldsymbol{\Sigma}_x$. Elle existe et est unique parce que $\boldsymbol{\Sigma}_x$ est définie positive). Les poids associés à chacun des échantillons sont définis par :

$$W_0 = \frac{\kappa}{n_x + \kappa} \tag{B.4}$$

$$W_i = \frac{1}{2(n_x + \kappa)} \quad \text{pour } i = 1, \ldots, n_x \tag{B.5}$$

$$W_i = \frac{1}{2(n_x + \kappa)} \quad \text{pour } i = n_x + 1, \ldots, 2n_x \tag{B.6}$$

Etape 2

Chaque échantillon construit précédemment est propagé au travers de la fonction g :

$$\mathcal{Y}_i = g(\mathcal{X}_i) \quad i = 0, \dots 2n_x \tag{B.7}$$

Etape 3

Les estimations $\hat{\boldsymbol{\mu}}_y$ et $\hat{\boldsymbol{\Sigma}}_y$ de $\boldsymbol{\mu}_y$ et de $\boldsymbol{\Sigma}_y$ sont calculées par

$$\hat{\boldsymbol{\mu}}_y = \sum_{i=0}^{2n_x} W_i \mathcal{Y}_i \tag{B.8}$$

$$\hat{\boldsymbol{\Sigma}}_y = \sum_{i=0}^{2n_x} W_i (\mathcal{Y}_i - \hat{\boldsymbol{\mu}}_y)(\mathcal{Y}_i - \hat{\boldsymbol{\mu}}_y)^{\mathtt{T}} \tag{B.9}$$

Il peut être montré [Jul96] que la valeur prédite $\hat{\boldsymbol{\mu}}_y$ est précise jusqu'au troisième ordre. En effet, en faisant un développement en série de Taylor de la fonction g, on peut obtenir une expression, sous la forme d'une somme infinie, de la vraie moyenne $\boldsymbol{\mu}_y$ et de la moyenne estimée $\hat{\boldsymbol{\mu}}_y$. Ces expressions sont égales jusqu'au troisième ordre, les erreurs d'estimation n'apparaissant qu'à partir du quatrième ordre. De même, la valeur prédite $\hat{\boldsymbol{\Sigma}}_y$ est précise jusqu'au deuxième ordre. A titre de comparaison, la valeur prédite de $\boldsymbol{\mu}_y$ par le filtre de Kalman étendu, c'est-à-dire en passant par une linéarisation de la fonction g, n'est précise qu'au premier ordre.

Une généralisation de la transformée sans parfum [Jul02] a été proposée afin de modifier la manière dont les échantillons $\mathcal{X}$ sont mis à l'échelle. Le but de cette mise à l'échelle, réglée par le paramètre κ dans la méthode présentée dans cette annexe, est de placer les échantillons plus ou moins loin de la moyenne du vecteur $\mathbf{x}$, en fonction de la non linéarité globale de la fonction g. Dans la nouvelle version de la méthode, la mise à l'échelle peut être effectuée dans n'importe qu'elle direction.

Annexe C

Calcul de la vraisemblance

Cette annexe donne les détails du calcul de la vraisemblance, intervenant dans le poids d'importance de la troisième version de l'algorithme de filtrage particulaire, présentée à la section IV.2.3.b, page 81. Rappelons d'abord le contexte. On souhaite calculer la vraisemblance $p(\mathbf{y}_t|K_t,\mathbf{f}_t)$ (l'indice de la particule est ici omis). Pour alléger les notations, on pose $\mathcal{F}_t=\{K_t,\mathbf{f}_t\}$ et $\mathcal{A}_t=\{\mathbf{a}_t,\mathbf{b}_t\}$. Les deux équations qui nous intéressent sont :

$$\begin{aligned}\mathcal{A}_t &\sim \mathcal{N}(\mathcal{A}_t;\boldsymbol{\mu}_t(\mathcal{F}_t),r^{\mathbf{y}}\boldsymbol{\Sigma}_t(\mathcal{F})) &\text{(C.1)}\\ \mathbf{y}_t &= \mathbf{C}(\mathcal{F}_t)\mathcal{A}_t+\mathbf{v}_t^{\mathbf{y}} &\text{(C.2)}\end{aligned}$$

où $\mathbf{v}_t^{\mathbf{y}}$ est un bruit gaussien de moyenne nulle et de variance $r^{\mathbf{y}}$. Toujours pour alléger les notations, la dépendance à $\mathcal{F}_t$ de $\mathbf{C}$, $\boldsymbol{\mu}_t$ et $\boldsymbol{\Sigma}_t$ sera omise dans la suite. Enfin, soit n la dimension de $\mathbf{a}_t$, $\mathcal{A}_t$ est alors de dimension $2n$ et $\mathbf{C}$ de dimensions $L_{\mathbf{w}}\times 2n$.

La vraisemblance $p(\mathbf{y}_t|\mathcal{F}_t)$ peut être calculée par :

$$\begin{aligned}p(\mathbf{y}_t|\mathcal{F}_t) &= \int_{\mathcal{A}_t} p(\mathbf{y}_t,\mathcal{A}_t|\mathcal{F}_t)\mathrm{d}\mathcal{A}_t\\ &= \int_{\mathcal{A}_t} p(\mathbf{y}_t|\mathcal{A}_t,\mathcal{F}_t)p(\mathcal{A}_t|\mathcal{F}_t)\mathrm{d}\mathcal{A}_t &\text{(C.3)}\end{aligned}$$

avec

$$\begin{aligned}p(\mathbf{y}_t|\mathcal{A}_t,\mathcal{F}_t) &= \frac{1}{(2\pi r^{\mathbf{y}})^{\frac{L_{\mathbf{w}}}{2}}}\exp\left(-\frac{1}{2r^{\mathbf{y}}}\big(\mathbf{y}_t-\mathbf{C}\mathcal{A}_t\big)^{\mathsf{T}}\big(\mathbf{y}_t-\mathbf{C}\mathcal{A}_t\big)\right) &\text{(C.4)}\\ p(\mathcal{A}_t|\mathcal{F}_t) &= \frac{1}{(2\pi r^{\mathbf{y}})^n|\boldsymbol{\Sigma}_t|^{\frac{1}{2}}}\exp\left(-\frac{1}{2r^{\mathbf{y}}}\big(\mathcal{A}_t-\boldsymbol{\mu}_t\big)^{\mathsf{T}}\boldsymbol{\Sigma}_t^{-1}\big(\mathcal{A}_t-\boldsymbol{\mu}_t\big)\right) &\text{(C.5)}\end{aligned}$$

L'équation (C.3) s'écrit donc

$$p(\mathbf{y}_t|\mathcal{F}_t)=\frac{1}{(2\pi)^{\frac{L_{\mathbf{w}}}{2}}}\frac{1}{(2\pi)^n|\boldsymbol{\Sigma}_t|^{\frac{1}{2}}}(r^{\mathbf{y}})^{-n-\frac{L_{\mathbf{w}}}{2}}\int_{\mathcal{A}_t}\exp\left(-\frac{1}{2r^{\mathbf{y}}}E\right)\mathrm{d}\mathcal{A}_t \qquad \text{(C.6)}$$

où E est défini par

$$E=\big(\mathbf{y}_t-\mathbf{C}\mathcal{A}_t\big)^{\mathsf{T}}\big(\mathbf{y}_t-\mathbf{C}\mathcal{A}_t\big)+\big(\mathcal{A}_t-\boldsymbol{\mu}_t\big)^{\mathsf{T}}\boldsymbol{\Sigma}_t^{-1}\big(\mathcal{A}_t-\boldsymbol{\mu}_t\big) \qquad \text{(C.7)}$$

On pose

$$\begin{aligned}
\mathbf{S}_t^{-1} &= \mathbf{C}^{\mathrm{T}}\mathbf{C} + \boldsymbol{\Sigma}_t^{-1} && \text{(C.8)}\\
\mathbf{m}_t &= \mathbf{S}_t\mathbf{C}^{\mathrm{T}}\mathbf{y}_t + \mathbf{S}_t\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t && \text{(C.9)}\\
\mathbf{P}_t &= \mathbf{I}_{L_{\mathbf{w}}} - \mathbf{C}\mathbf{S}_t\mathbf{C}^{\mathrm{T}} && \text{(C.10)}\\
\mathbf{Q}_t &= \boldsymbol{\Sigma}_t^{-1} - \boldsymbol{\Sigma}_t^{-1}\mathbf{S}_t\boldsymbol{\Sigma}_t^{-1} && \text{(C.11)}
\end{aligned}$$

l'équation (C.7) peut alors s'écrire

$$E = \big(\mathcal{A}_t - \mathbf{m}_t\big)^{\mathrm{T}}\mathbf{S}_t^{-1}\big(\mathcal{A}_t - \mathbf{m}_t\big) + \mathbf{y}_t^{\mathrm{T}}\mathbf{P}_t\mathbf{y}_t + \boldsymbol{\mu}_t^{\mathrm{T}}\mathbf{Q}_t\boldsymbol{\mu}_t - \mathbf{y}_t^{\mathrm{T}}\mathbf{C}\mathbf{S}_t\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t - \boldsymbol{\mu}_t^{\mathrm{T}}\boldsymbol{\Sigma}_t^{-1}\mathbf{S}_t\mathbf{C}^{\mathrm{T}}\mathbf{y}_t \quad \text{(C.12)}$$

L'intégrale dans (C.6) peut alors être calculée et on obtient, après simplifications :

$$\begin{aligned}
p(\mathbf{y}_t|\mathcal{F}_t) &= \frac{(r^{\mathbf{y}})^{-\frac{L_{\mathbf{w}}}{2}}}{(2\pi)^{\frac{L_{\mathbf{w}}}{2}}}\frac{|\mathbf{S}_t|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_t|^{\frac{1}{2}}}\exp\left(-\frac{\mathbf{y}_t^{\mathrm{T}}\mathbf{P}_t\mathbf{y}_t + \boldsymbol{\mu}_t^{\mathrm{T}}\mathbf{Q}_t\boldsymbol{\mu}_t - \mathbf{y}_t^{\mathrm{T}}\mathbf{C}\mathbf{S}_t\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t - \boldsymbol{\mu}_t^{\mathrm{T}}\boldsymbol{\Sigma}_t^{-1}\mathbf{S}_t\mathbf{C}^{\mathrm{T}}\mathbf{y}_t}{2r^{\mathbf{y}}}\right)\\
&\propto \frac{|\mathbf{S}_t|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_t|^{\frac{1}{2}}}\exp\left(-\frac{\mathbf{y}_t^{\mathrm{T}}\mathbf{P}_t\mathbf{y}_t + \boldsymbol{\mu}_t^{\mathrm{T}}\mathbf{Q}_t\boldsymbol{\mu}_t - \mathbf{y}_t^{\mathrm{T}}\mathbf{C}\mathbf{S}_t\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t - \boldsymbol{\mu}_t^{\mathrm{T}}\boldsymbol{\Sigma}_t^{-1}\mathbf{S}_t\mathbf{C}^{\mathrm{T}}\mathbf{y}_t}{2r^{\mathbf{y}}}\right) \qquad \text{(C.13)}
\end{aligned}$$

Les paramètres "libres" dans l'équation (C.13) sont $\boldsymbol{\mu}_t$ et $\boldsymbol{\Sigma}_t$. Plusieurs choix peuvent être faits, chacun pouvant simplifier le calcul de la vraisemblance.

Choix 1

On utilise le *g-prior* introduit par Zellner [Zel86] pour définir la matrice $\boldsymbol{\Sigma}_t$:

$$\boldsymbol{\Sigma}_t^{-1} = \frac{1}{\sigma_0^2}\mathbf{C}^{\mathrm{T}}\mathbf{C} \qquad \text{(C.14)}$$

Les matrices $\mathbf{S}_t$, $\mathbf{P}_t$ et $\mathbf{Q}_t$ deviennent alors

$$\begin{aligned}
\mathbf{S}_t &= \frac{\sigma_0^2}{\sigma_0^2+1}(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1} && \text{(C.15)}\\
\mathbf{P}_t &= \mathbf{I}_{L_{\mathbf{w}}} - \frac{\sigma_0^2}{\sigma_0^2+1}\mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}} && \text{(C.16)}\\
\mathbf{Q}_t &= \frac{1}{\sigma_0^2+1}\mathbf{C}^{\mathrm{T}}\mathbf{C} && \text{(C.17)}
\end{aligned}$$

et l'équation (C.13) se simplifie en :

$$p(\mathbf{y}_t|\mathcal{F}_t) \propto (\sigma_0^2+1)^{-n}\exp\left(-\frac{\big(\mathbf{C}\boldsymbol{\mu}_t - \mathbf{y}_t\big)^{\mathrm{T}}\big(\mathbf{C}\boldsymbol{\mu}_t - \mathbf{y}_t\big) + \sigma_0^2\big(\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_t - \mathbf{y}_t^{\mathrm{T}}\mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y}_t\big)}{2(\sigma_0^2+1)r^{\mathbf{y}}}\right) \qquad \text{(C.18)}$$

Choix 2

Le vecteur $\boldsymbol{\mu}_t$ est considéré nul :

$$\boldsymbol{\mu}_t = \mathbf{0} \tag{C.19}$$

L'équation (C.13) se simplifie alors en :

$$p(\mathbf{y}_t|\mathcal{F}_t) \propto \frac{|\mathbf{S}_t|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_t|^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_t - \mathbf{y}_t^{\mathrm{T}}\mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C} + \boldsymbol{\Sigma}_t^{-1})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y}_t}{2r^{\mathbf{y}}}\right) \tag{C.20}$$

Choix 3

Combinaison des choix 1 et 2. L'équation (C.13) se simplifie en :

$$p(\mathbf{y}_t|\mathcal{F}_t) \propto (\sigma_0^2 + 1)^{-n} \exp\left(-\frac{\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_t - \frac{\sigma_0^2}{\sigma_0^2+1}\mathbf{y}_t^{\mathrm{T}}\mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y}_t}{2r^{\mathbf{y}}}\right) \tag{C.21}$$

Bibliographie

[Abd06] S. A. Abdallah et M. D. Plumbley. Unsupervised Analysis of Polyphonic Music by Sparse Coding. *IEEE transactions on Neural Networks*, volume 17, pages 179–196, janvier 2006.

[Ald97] J. Aldrich. R. A. Fisher and the Making of Maximum Likelihood 1912-1922. *Statistical Science*, volume 12, pages 162–176, août 1997.

[Als72] D. L. Alspach et H. W. Sorenson. Nonlinear Bayesian Estimation using Gaussian Sum Approximations. *IEEE transactions on Automatic Control*, volume 20, pages 439–447, 1972.

[And79] B. D. O. Anderson et J. B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.

[And99] C. Andrieu et A. Doucet. Joint Bayesian Model Selection and Estimation of Noisy Sinusoids via Reversible Jump MCMC. *IEEE transactions on Signal Processing*, volume 47, pages 2667–2676, octobre 1999.

[And01] C. Andrieu, M. Davy et A. Doucet. Improved Auxiliary Particle Filtering : Applications to Time-Varying Spectral Analysis. *Dans les actes de IEEE SSP*. 2001.

[And02] C. Andrieu et A. Doucet. Particle Filtering for Partially Observed Gaussian State Space Models. *Journal of the Royal Statistical Society B*, volume 64, pages 827–836, 2002.

[Aru02] M. S. Arulampalam, S. Maskell, N. Gordon et T. Clapp. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE transactions on Signal Processing*, volume 50, pages 174–188, février 2002.

[Aug82] J. F. Augoyard, O. Balay, G. Chelkoff et O. Belle. Sonorité, sociabilité, urbanité. Rapport technique, Ecole d'Architecture de Grenoble, 1982.

[Aug95] J. F. Augoyard et H. Torgue. *A l'écoute de l'environnement - Répertoire des effets sonores*. Ed. Parenthèses, Marseille, 1995.

[Bar93] J.-L. Bardyn. L'appel du port, recherche exploratoire pluridisciplinaire sur l'ambiance sonore de cind ports européens. Rapport technique, CRESSON - ARCHIMEDA, Grenoble, octobre 1993.

[Bar94] A. L. Barker, D. E. Brown et W. N. Martin. Bayesian Estimation and the Kalman Filter. Rapport technique IPC-TR-94-002, Institute of Parallel Computation, School of Engineering and Applied Science, septembre 1994.

[Bar99] J.-L. Bardyn. La Portée ferroviaire, ambiances sonores des gares européennes. Rapport technique, CRESSON - ARCHIMEDA, Grenoble, avril 1999.

[Bel00] J. P. Bello et M. Sandler. Blackboard System and Top-Down Processing for the Transcription of Simple Polyphonic Music. *Dans les actes de COST G-6 Conference on Digital Audio Effects*. Verona, Italy, décembre 2000.

[Bel05] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies et M. B. Sandler. A Tutorial on Onset Detection in Music Signals. *IEEE transactions on Speech and Audio Processing*, volume 13, pages 1035–1047, septembre 2005.

[Bir97] L. Birgé et P. Massart. From Model Selection to Adaptative Estimation. *Dans les actes de* D. Pollard, E. Torgersen et G. L. Yang, rédacteurs, *Festschrift for Lucien Le Cam. Research Papers in Probability and Statistics*, Springer, pages 55–87. 1997.

[Boe78] E. de Boer et H. R. de Jongh. On Cochlear Encoding : Potentialities and Limitations of the Reverse-Correlation Technique. *Journal of the Acoustical Society of America (JASA)*, volume 63, pages 115–135, janvier 1978.

[Bre90] A. S. Bregman. *Auditory Scene Analysis : the Perceptual Organization of Sound*. MIT Press, 1990.

[Bro89] J. C. Brown et M. S. Puckette. Calculation of a narrowed autocorrelation function. *Journal of the Acoustical Society of America (JASA)*, volume 85, pages 1595–1601, avril 1989.

[Bro94] G. J. Brown et M. Cooke. Perceptual Grouping of Musical Sounds : A Computational Model. *Journal of the New Music Research*, volume 23, pages 107–132, 1994.

[Car99] J. Carpenter, P. Clifford et P. Fearnhead. Building Robust Simulation-based Filters for Evolving Data Sets. Rapport technique, University of Oxford, 1999.

[Cem03] A. T. Cemgil, B. Kappen et D. Barber. Generative Model Based Polyphonic Music Transcription. *Dans les actes de IEEE WASPAA*. New Paltz, NY, USA, octobre 2003, pages 181–184.

[Cem06] A. T. Cemgil, H. J. Kappen et D. Barber. A Generative Model for Music Transcription. *IEEE transactions on Audio, Speech and Language Processing*, volume 14, pages 679–694, mars 2006.

[Che93] A. de Cheveigné. Separation of Concurrent Harmonic Sounds : Fundamental Frequency Estimation and a Time-Domain Cancellation Model of Auditory Processing. *Journal of the Acoustical Society of America (JASA)*, volume 93, pages 3271–3290, juin 1993.

[Che94] G. Chelkoff et J. P. Odion. Les effets sonores, un outils pour l'évaluation des espaces en extérieur. *Journal de Physique*, volume 4, pages C5–101–C5–104, 1994.

[Che97a] G. Chelkoff, J.-P. Thibaud, J.-L. Bardyn, B. Belchun et M. Leroux. Ambiances sous la ville, une approche écologique des espaces publics souterains. Rapport technique, CRESSON, Grenoble, septembre 1997.

[Che97b] A. de Cheveigné. Concurrent Vowel Identification. III. A Neural Model of Harmonic Interference Cancellation. *Journal of the Acoustical Society of America (JASA)*, volume 101, pages 2857–2865, mai 1997.

[Che99] A. de Cheveigné et H. Kawahara. Multiple Period Estimation and Pitch Perception Model. *Speech Communication*, volume 27, pages 175–185, avril 1999.

[Che00] R. Chen et J. S. Liu. Mixture Kalman Filters. *Journal of the Royal Statistical Society B*, volume 62, pages 493–508, 2000.

[Che02] A. de Cheveigné et H. Kawahara. YIN, a Fundamental Frequency Estimator for Speech and Music. *Journal of the Acoustical Society of America (JASA)*, volume 111, pages 1917–1930, avril 2002.

[Che03] Z. Chen. Bayesian Filtering : From Kalman Filters to Particle Filters, and Beyond. *Internet publication*, 2003.

[Cla99] T. Clapp et S. J. Godsill. Fixed-lag Smoothing using Sequential Importance Sampling. *Dans les actes de Bayesian Statistics 6*. 1999, pages 743–752.

[Coo65] J. W. Cooley et J. W. Tukey. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, volume 19, pages 297–301, avril 1965.

[Coo91] M. P. Cooke. *Modelling Auditory Processing and Organisation*. Thèse de doctorat, Department of Computer Science, university of Sheffield, mai 1991.

[Cor91] D. D. Corkill. Blackboard Systems. *AI Expert*, volume 6, pages 40–47, septembre 1991.

[Cri01] D. Crisan. Particle Filters - A Theoretical Perspective. *Dans les actes de* A. Doucet, N. de Freitas et N. Gordon, rédacteurs, *Sequential Monte Carlo Methods in Practice*, Springer, pages 17–41. 2001.

[Cri02] D. Crisan et A. Doucet. A Survey of Convergence Results on Particle Filtering Methods for Practitioners. *IEEE transactions on Signal Processing*, volume 50, pages 736–746, mars 2002.

[Dav98] M. Davy, B. Leprettre, C. Doncarli et N. Martin. Tracking of Spectral Lines in an ARCAP Time-Frequency Representation. *Dans les actes de EUSIPCO*. Island of Rhodes, Greece, 1998, pages 633–636.

[Dav00] M. Davy. *Noyaux Optimisés pour la Classification dans le Plan Temps-Fréquence*. Thèse de doctorat, Université de Nantes et Ecole Centrale de Nantes, septembre 2000.

[Dav02a] M. Davy et S. J. Godsill. Audio Information Retrieval : a Bibliographical Study. Rapport technique CUED/F-INFENG/TR.429, Signal Processing Group, Cambridge University Engineering Department, février 2002.

[Dav02b] M. Davy et S. J. Godsill. Bayesian Harmonic Models for Musical Signal Analysis. *Dans les actes de Bayesian Statistics 7*. Valencia, Spain, juin 2002.

[Dav04] M. Davy et J. Idier. Fast MCMC Computations for the Estimations of Sparse Processes from Noisy Observations. *Dans les actes de IEEE ICASSP*. Montréal, Québec, Canada, mai 2004.

[Dav06] M. Davy, S. J. Godsill et J. Idier. Bayesian Analysis of Polyphonic Western Tonal Music. *Journal of the Acoustical Society of America (JASA)*, volume 119, pages 2498–2517, avril 2006.

[Des05] F. Desobry, M. Davy et C. Doncarli. An Online Kernel Change Detection Algorithm. *IEEE transactions on Signal Processing*, volume 53, pages 2961–2974, août 2005.

[Dou98] A. Doucet. On Sequential Simulation-Based Methods for Bayesian Filtering. Rapport technique, Signal Processing Group, Department of engineering, University of Cambridge CB2 1PZ Cambridge, 1998.

[Dou00] A. Doucet, S. Godsill et C. Andrieu. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *Statistics and Computing*, volume 10, pages 197–208, juillet 2000.

[Dou01] A. Doucet, N. de Freitas et N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[Dou04] A. Doucet et S. Sénécal. Fixed-lag Sequential Monte Carlo. *Dans les actes de EUSIPCO*. Vienna, Austria, septembre 2004, pages 861–864.

[Dou05] A. Doucet et X. Wang. Monte Carlo Methods for Signal Processing. *IEEE Signal Processing Magazine*, pages 152–170, novembre 2005.

[Dou06] A. Doucet, M. Briers et S. Sénécal. Effcient Block Sampling Strategies for Sequential Monte Carlo Methods. *Journal of Computational & Graphical Statistics*, 2006. To appear.

[Dov91] B. Doval et X. Rodet. Estimation of Fundamental Frequency of Musical Sound Signals. *Dans les actes de ICASSP*. Toronto, Ontario, Canada, avril 1991, volume 5, pages 3657–3660.

[Dré03] J. P. Drécourt. Kalman Filtering in Hydrological Modelling. Rapport technique DAIHM 2003-1, DHI Water & Environment, mai 2003.

[Dub] C. Dubois et M. Davy. Joint Detection and Tracking of Time-Varying Harmonic Components : a Flexible Bayesian Approach. Submitted at IEEE transactions on Audio, Speech and Language Processing. Under second review.

[Dub05a] C. Dubois et M. Davy. Harmonic Tracking using Sequential Monte Carlo. *Dans les actes de SSP*. Bordeaux, France, juillet 2005.

[Dub05b] C. Dubois et M. Davy. Suivi de Trajectoires Temps-Fréquence par Filtrage Particulaire. *Dans les actes de GRETSI*. Louvain-la-Neuve, Belgium, septembre 2005.

[Dub05c] C. Dubois, M. Davy et J. Idier. Tracking of Time-Frequency Components using Particle Filtering. *Dans les actes de IEEE ICASSP*. Philadelphia, PA, USA, mars 2005, volume 4, pages 9–12.

[Dud01] R. O. Duda, P. E. Hart et D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2001. Second edition.

[Dux03] C. Duxbury, J. P. Bello, M. Davies et M. Sandler. A combined Phase and Amplitude Based Approach to Onset Detection for Audio Segmentation. *Dans les actes de European Workshop on Image Analysis for Multimedia Interactive Services*. London, UK, avril 2003, pages 275–280.

[Edw97] A. W. F. Edwards. What Did Fisher Mean by "Inverse Probability" in 1912-1922 ? *Statistical Science*, volume 12, pages 177–184, août 1997.

[Ell96] P. W. Ellis. *Prediction-Driven Computational Auditory Scene Analysis*. Thèse de doctorat, M. I. T. Departement of Electrical Engineering and Computer Science, juin 1996.

[Fla93] P. Flandrin. *Temps-fréquence*. Hermès, 1993.

[Fle34] H. Fletcher. Loudness, Pitch and the Timbre of Musical Tones and their Relation to the Intensity, the Frequency and the Overtone Structure. *Journal of the Acoustical Society of America (JASA)*, volume 6, pages 59–69, octobre 1934.

[Fle98] N. H. Fletcher et T. D. Rossing. *The Physics of Musical Instruments*. Springer, 1998.

[Foo99] J. Foote. An Overview of Audio Information Retrieval. *Multimedia Systems*, volume 7, pages 2–10, janvier 1999.

[Fra69] D. C. Fraser et J. E. Potter. The Optimum Linear Smoother as a Combination of Two Optimum Linear Filters. *IEEE transactions on Automatic Control*, volume 14, pages 387–390, août 1969.

[Fre02] N. de Freitas. Rao-Blackwellised Particle Filtering for Fault Diagnosis. *Dans les actes de IEEE Aerospace Conference*. Big Sky, MT, USA, mars 2002, volume 4, pages 1767–1772.

[Gew89] J. Geweke. Bayesian Inference in Econometric Models using Monte Carlo Integration. *Econometrica*, volume 57, pages 1317–1340, 1989.

[God99] D. Godsmark et G. J. Brown. A Blackboard Architecture for Computational Auditory Scene Analysis. *Speech Communication*, volume 27, pages 351–366, avril 1999.

[Gol73] J. L. Goldstein. An Optimum Processor Theory for the Central Formation of the Pitch of Complex Tones. *Journal of the Acoustical Society of America (JASA)*, volume 54, pages 1496–1516, décembre 1973.

[Goo96] M. Goodwin. Residual Modeling in Music Analysis-Synthesis. *Dans les actes de ICASSP*. Atlanta, GA, USA, mai 1996, volume 2, pages 1005–1008.

[Gor93] N. J. Gordon, D. J. Salmond et A. F. M. smith. Novel Approach to Nonlinear/non-Gaussian Bayesian State Estimation. *IEE Proceedings-F*, volume 140, pages 107–113, 1993.

[Got01] M. Goto. A Predominent-F0 Estimation Method for CD Recordings : MAP Estimation using EM Algorithm for Adaptative Tone Models. *Dans les actes de IEEE ICASSP*. Salt Lake City, UT, USA, mai 2001, volume 5, pages 3365–3368.

[Got04] M. Goto. A Real-Time Music-Scene-Description System : Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-World Audio Signals. *Speech Communication*, volume 43, pages 311–329, septembre 2004.

[Gre99] P. J. Green. Penalized Likelihood. *Encyclopaedia of Statistical Sciences*, volume 3, pages 578–586, 1999.

[Hai01] S. W. Hainsworth. Analysis of Musical Audio for Polyphonic Transcription. Rapport de première année, septembre 2001.

[Hai03] S. W. Hainsworth. *Techniques for the Automated Analysis of Musical Audio*. Thèse de doctorat, Signal Processing Group, Department of Engineering, University of Cambridge, décembre 2003.

[Har78] F. J. Harris. On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. *Proceedings of the IEEE*, volume 66, pages 51–83, janvier 1978.

[Har96] W. M. Hartmann. Pitch, Periodicity and Auditory Organization. *Journal of the Acoustical Society of America (JASA)*, volume 100, pages 3491–3502, décembre 1996.

[Har97] W. M. Hartmann. *Signals, Sound, and Sensation*. Springer-Verlag, 1997.

[Her03] P. Herrera, G. Peeters et S. Dubnov. Automatic Classification of Musical Instrument Sounds. *Journal of the New Music Research*, volume 32, pages 3–21, mars 2003.

[Hew91] M. J. Hewitt et R. Meddis. An Evaluation of Eight Computer Models of Mammalian Inner Hair-Cell Function. *Journal of the Acoustical Society of America (JASA)*, volume 90, pages 904–917, août 1991.

[Hla92] F. Hlawatsch et G. F. Boudreaux-Bartels. Linear and Quadratic Time-Frequency Signal Representations. *IEEE Signal Processing Magazine*, volume 9, pages 21–67, avril 1992.

[Hou90] A. J. M. Houtsma et J. Smurzynski. Pitch Identification and Discrimination for Complex Tones with many Harmonics. *Journal of the Acoustical Society of America (JASA)*, volume 87, pages 304–310, janvier 1990.

[IEC93] IEC. Electroacoustics - Intruments for the measurement of sound intensity - Measurement with pairs of pressure sensing microphones. *International Electrotechnical Commission*, décembre 1993. First edition.

[Iri98] R. A. Irizarry. *Statistics and Music : Fitting a Local Harmonic Model to Musical Sound Signals*. Thèse de doctorat, University of California, Berkeley, 1998.

[Iri01] R. A. Irizarry. Local Harmonic Estimation in Musical Sound Signals. *Journal of the American Statistical Association*, volume 96, pages 357–367, juin 2001.

[Jai05] F. Jaillet. *Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore*. Thèse de doctorat, Université de la Méditerranée - Aix-Marseille II, juin 2005.

[Jen61] R. A. Jenkins. Perception of Pitch, Timbre and Loudness. *Journal of the Acoustical Society of America (JASA)*, volume 33, pages 1550–1557, novembre 1961.

[Jul96] S. J. Julier et J. K. Uhlmann. A General Method for Approximating Nonlinear Transformations of Probability Distributions. *Internet publication*, novembre 1996.

[Jul97] S. J. Julier et J. K. Uhlmann. A New Extension of the Kalman Filter to Nonlinear Systems. *Dans les actes de SPIE Aerosense*. Orlando, FL, USA, avril 1997, volume 3068, pages 182–193.

[Jul02] S. J. Julier. The Scaled Unscented Transform. *Dans les actes de American Control Conference*. Anchorage, AK, USA, mai 2002, volume 6, pages 4555–4559.

[Kal60] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of the Basic Engineering*, volume 82, pages 35–45, mars 1960.

[Kar99] M. Karjalainen et T. Tolonen. Multi-Pitch and Periodicity Analysis Model for Sound Separation and Auditory Scene Analysis. *Dans les actes de IEEE ICASSP*. Phoenix, AZ, USA, mars 1999, volume 2, pages 929–932.

[Kas95] K. Kashino, K. Nakadia, T. Kinoshita et H. Tanaka. Application of Bayesian Probability Network to Music Scene Analysis. *Dans les actes de IJCAI, CASA workshop*. Montréal, Québec, Canada, août 1995, pages 52–59.

[Kit96] G. Kitagawa. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, volume 5, pages 1–25, 1996.

[Kla98a] A. P. Klapuri. *Automatic Transcription Music*. Mémoire de master, Tampere University of Technology, avril 1998.

[Kla98b] F. I. Klassner, V. R. Lesser et S. H. Nawab. The IPUS Blackboard Architecture as a Framework for Computational Auditory Scene Analysis. *Dans les actes de* D. F. Rosenthal et H. G. Okuno, rédacteurs, *Computational Auditory Scene Analysis*, Lawrence Erlbaum Associates, NJ, pages 105–114. janvier 1998.

[Kla03] A. P. Klapuri. Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness. *IEEE transactions on Speech and Audio Processing*, volume 11, pages 804–816, novembre 2003.

[Kla04] A. P. Klapuri. *Signal Processing Methods for the Automatic Transcription of Music*. Thèse de doctorat, Tampere University of Technology, mars 2004.

[Kla05] A. P. Klapuri. A Perceptually Motivated Multiple-F0 Estimation Method. *Dans les actes de IEEE WASPAA*. New Paltz, NY, USA, octobre 2005, pages 291–294.

[Kla06] A. P. Klapuri et M. Davy. *Signal Processing Techniques for Music Transcription*. Springer, 2006.

[Kon94] A. Kong, J. S. Liu et W. H. Wong. Sequential Imputations and Bayesian Missing Data Problems. *Journal of the American Statistical Association*, volume 89, pages 278–288, 1994.

[Krs05] S. Krstulović, R. Gribonval, P. Leveau et L. Daudet. A Comparison of Two Extensions of the Matching Pursuit Algorithm for the Harmonic Decomposition of Sounds. *Dans les actes de WASPAA*. New Paltz, NY, USA, octobre 2005, pages 259–262.

[Kun96] N. Kunieda, T. Shimamura et J. Suzuki. Robust Method of Measurement of Fundamental Frequency by ACLOS : Autocorrelation of Log-Spectrum. *Dans les actes de IEEE ICASSP*. Atlanta, GA, USA, mai 1996, volume 1, pages 232–235.

[Lav98] S. Laveaud. *L'effet de métabole, possibilité d'une caractérisation acoustique ?*. Mémoire de master, 1998.

[Les95] V. R. Lesser, S. H. Nawab et F. I. Klassner. IPUS : an Architecture for the Integrated Processing and Understanding of Signals. *Artificial Intelligence*, volume 77, pages 129–171, janvier 1995.

[Liu95] J. S. Liu et R. Chen. Blind Deconvolution via Sequential Imputations. *Journal of the American Statistical Association*, volume 90, pages 567–576, 1995.

[Liu98a] J. S. Liu et R. Chen. Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association*, volume 93, pages 1032–1044, septembre 1998.

[Liu98b] Z. Liu, Y. Wang et T. Chen. Audio Feature Extraction and Analysis for Scene Segmentation and Classification. *Journal of VLSI Signal Processing*, volume 20, pages 61–79, juin 1998.

[Liv04] A. A. Livshin et X. Rodet. Intrument Recognition Beyond Separate Notes - Indexing Continuous Recordings. *Dans les actes de ICMC*. Coral Gables, FL, USA, novembre 2004.

[Mac03] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[Mal93] S. G. Mallat et Z. Zhang. Matching Pursuit with Time-Frequency Dictionaries. *IEEE transactions on Signal Processing*, volume 41, pages 3397–3415, décembre 1993.

[Mar96a] K. D. Martin. A Blackboard System for Automatic Transcription of Simple Polyphonic Music. Rapport technique 385, M. I. T. Media Laboratory Perceptual Computing Section, 1996.

[Mar96b] K. D. Martin. Automatic Transcription of Simple Polyphonic Music : Robust Front End Processing. Rapport technique 399, M. I. T. Media Laboratory Perceptual Computing Section, 1996.

[Mar03] J. Marozeau, A. de Cheveigné, S. McAdams et S. Winsberg. The Dependency of Timbre on Fundamental Frequency. *Journal of the Acoustical Society of America (JASA)*, volume 114, pages 2946–2957, novembre 2003.

[Med86] R. Meddis. Simulation of Mechanical to Neural Transduction in the Auditory Receptor. *Journal of the Acoustical Society of America (JASA)*, volume 79, pages 702–711, mars 1986.

[Med91] R. Meddis et M. J. Hewitt. Virtual Pitch and Phase Sensitivity of a Computer Model of the auditory Periphery. I : Pitch Identification. *Journal of the Acoustical Society of America (JASA)*, volume 89, pages 2866–2882, juin 1991.

[Med92] R. Meddis et M. J. Hewitt. Modeling the Identification of Concurrent Vowels with Different Fundamental Frequencies. *Journal of the Acoustical Society of America (JASA)*, volume 91, pages 233–245, janvier 1992.

[Mel91] D. K. Mellinger. *Event Formation and Separation in Musical Sound*. Thèse de doctorat, Departement of Computer Science, University of Standford, décembre 1991.

[Mer00] R. van der Merwe, A. Doucet, N. de Freitas et E. Wan. The Unscented Particle Filter. Rapport technique, Department of engineering, University of Cambridge CB2 1PZ Cambridge, 2000.

[Mer01] R. van der Merwe et E. A. Wan. The Square-Root Unscented Kalman Filter for State and Parameter Estimation. *Dans les actes de IEEE ICASSP*. Salt Lake City, UT, USA, mai 2001, volume 6, pages 3461–3464.

[Moo06] B. C. J. Moore, B. R. Glasberg et H. J. Flanagan. Frequency Discrimination of complex Tones ; Assessing the Role of Component Resolvability and Temporal Fine

Structure. *Journal of the Acoustical Society of America (JASA)*, volume 119, pages 480–490, janvier 2006.

[Nol64] A. M. Noll. Short-Time Spectrum and Cepstrum Techniques for Vocal-Pitch Detection. *Journal of the Acoustical Society of America (JASA)*, volume 36, pages 296–302, février 1964.

[Odi96] J. Odion, J. Augoyard, G. Chelkoff et J. Bardyn. Testologie architecturale des effets sonores, prédictabilité de la qualité sonore. Rapport technique, CRESSON, 1996.

[Pat76] R. D. Patterson. Auditory Filter Shapes Derived with noise Stimuli. *Journal of the Acoustical Society of America (JASA)*, volume 59, pages 640–654, mars 1976.

[Pat96] R. D. Patterson et J. Holdsworth. A Functional Model of Neural Activity Patterns and Auditory Images. *Dans les actes de* W. A. Ainsworth, rédacteur, *Advances in Speech, Hearing and Language Processing*, JAI Press inc., volume 3, pages 547–563. 1996.

[Pel91] X. Pelorson. *Pertinence des paramètres objectifs utilisés pour caractériser la qualité acoustique d'une salle.* Thèse de doctorat, Université du Maine, 1991.

[Pie96] W. J. Pielemeier et G. H. Wakefield. A High-Resolution Time-Frequency Representation for Musical Instrument Signals. *Journal of the Acoustical Society of America (JASA)*, volume 99, pages 2382–2396, avril 1996.

[Pit99] M. K. Pitt et N. Shephard. Filtering via Simulation : Auxiliary Particle Filter. *Journal of the American Statistical Association*, volume 94, pages 590–599, 1999.

[Plu02] M. D. Plumbley, S. A. Abdallah, J. P. Bello, M. E. Davies, G. Monti et M. B. Sandler. Automatic Music Transcription and Audio Source Separation. *Cybernetics and Systems*, volume 33, pages 603–627, septembre 2002.

[Plu03] M. D. Plumbley. Algorithms for Nonnegative Independent Component Analysis. *IEEE transactions on Neural Networks*, volume 14, pages 534–543, mai 2003.

[Pun02] E. Punskaya, C. Andrieu, A. Doucet et W. J. Fitzgerald. Bayesian Curve Fitting using MCMC with Applications to Signal Segmentation. *IEEE transactions on Signal Processing*, volume 50, pages 747–758, mars 2002.

[Rab76] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg et C. A. McGonegal. A Comparative Performance Study of Several Pitch Detection Algorithms. *IEEE transactions on Acoustics, Speech and Signal Processing*, volume 24, pages 434–442, octobre 1976.

[Rap99] C. Raphael. Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models. *IEEE transactions on Pattern Analysis and Machine Intelligence*, volume 21, pages 360–370, avril 1999.

[Rap02] C. Raphael. Automatic Transcription of Piano Music. *Dans les actes de ISMIR.* Paris, France, octobre 2002.

[Rém01] N. Rémy. *Maîtrise et prédictibilité de la qualité sonore du projet architectural : application aux espaces publics en gare.* Thèse de doctorat, Université de Nantes, 2001.

[Rob99] C. P. Robert et G. Casella. *Monte Carlo Statistical Methods*. Springer, 1999.

[Rob01] C. P. Robert. *The Bayesian Choice*. Springer, 2001.

[Ros98] D. F. Rosenthal et H. G. Okuno. *Computational Auditory Scene Analysis*. Lawrence Erlbaum Associates, 1998.

[Ros00] S. Rossignol. *Ségmentation et Indexation des Signaux Sonores Musicaux*. Thèse de doctorat, Université Paris 6, juillet 2000.

[Sch66] P. Schaeffer. *Traité des objets musicaux*. Ed. du Seuil, Paris, 1966.

[Sch80] R. M. Schafer. *Le paysage sonore*. Ed. J. C. Lattès, Paris, 1980.

[Ser90] X. Serra et J. Smith. Spectral Modeling Synthesis : A Sound Analysis/Synthesis System based on a Deterministic plsu Stochastic Decomposition. *Computer Music Journal*, volume 14, pages 12–24, 1990.

[Sla90] M. Slaney et R. F. Lyon. A Perceptual Pitch Detector. *Dans les actes de IEEE ICASSP*. Albuquerque, NM, USA, avril 1990, volume 1, pages 357–360.

[Sla93a] M. Slaney. An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank. Rapport technique 35, Perception Group, Advanced Technology Group, Apple Computer, 1993.

[Sla93b] M. Slaney et R. F. Lyon. On the Importance of Time - A Temporal Representation of Sound. *Dans les actes de* M. Cooke, S. Beet et M. Crawford, rédacteurs, *Visual Representations of Speech Signals*, John Wiley & Sons Ltd, pages 95–116. 1993.

[Ste99] A. D. Sterian. *Model-Based Segmentation of Time-Frequency Images for Musical Transcription*. Thèse de doctorat, University of Michigan, 1999.

[Swe62] J. A. Swets, D. M. Green et W. P. Tanner(Jr). On the Witdh of Critical Bands. *Journal of the Acoustical Society of America (JASA)*, volume 34, pages 108–113, janvier 1962.

[Tix01] N. Tixier. *Morphodynamique des ambiances construites*. Thèse de doctorat, Université de Nantes, novembre 2001.

[Tol99] R. Tolding. Estimation Theory and Foundations of Atmospheric Data Assimilation. Rapport technique DAO Office Note 1999-01, Data Assimilation Office, Goddard Space Flight Center, juin 1999.

[Tol00] T. Tolonen et M. Karjalainen. A Computationally Efficient Multipitch Analysis Model. *IEEE transactions on Speech and Audio Processing*, volume 8, pages 708–716, novembre 2000.

[Tza02a] G. Tzanetakis. *Manipulation, Analysis and Retrieval Systems for Audio Signals*. Thèse de doctorat, Princeton University, juin 2002.

[Tza02b] G. Tzanetakis et P. Cook. Musical Genre Classification of Audio Signals. *IEEE transactions on Speech and Audio Processing*, volume 10, pages 293–302, juillet 2002.

[Vin05] E. Vincent et M. D. Plumbley. A Prototype System for Object Coding of Musical Audio. *Dans les actes de IEEE WASPAA*. New Paltz, NY, USA, octobre 2005, volume 15, pages 239–242.

[Vir03] T. VIRTANEN. Sound Source Separation Using Sparse Coding with Temporal Continuity Objective. *Dans les actes de ICMC*. Singapore, octobre 2003.

[Vit67] A. VITERBI. Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *IEEE transactions on Information Theory*, volume 13, pages 260–269, avril 1967.

[Wah78] G. WAHBA. Improper Priors, Spline Smoothing and the Problem of Guarding against Model Errors in Regression. *Journal of the Royal Statistical Society B*, volume 40, pages 364–372, 1978.

[Wah83] G. WAHBA. Bayesian Confidence Intervals for the Cross-Validated Smoothing Spline. *Journal of the Royal Statistical Society B*, volume 45, pages 133–150, 1983.

[Wie77] C. C. WIER, W. JESTEADT et D. M. GREEN. Frequency Discrimination as a Function of Frequency and Sensation Level. *Journal of the Acoustical Society of America (JASA)*, volume 61, pages 178–184, janvier 1977.

[Wu03] M. WU, D. WANG et G. J. BROWN. A Multipitch Tracking Algorithm for Noisy Speech. *IEEE transactions on Speech and Audio Processing*, volume 11, pages 229–241, mai 2003.

[Yeh05] C. YEH, A. RÖBEL et X. RODET. Multiple Fundamental Frequency Estimation of Polyphonic Music Signals. *Dans les actes de IEEE ICASSP*. Philadelphia, PA, USA, mars 2005, volume 3, pages 225–228.

[Zel86] A. ZELLNER. On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions. *Dans les actes de* P. GOEL et A. ZELLNER, rédacteurs, *Bayesian Inference and Decision Techniques*, Elsevier. 1986.