N° d'ordre : 2756

THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ BORDEAUX I

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

Par **Pierre HANNA**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

Modélisation statistique de sons bruités :
étude de la densité spectrale, analyse, transformation musicale et synthèse

---

Soutenue le : 12 Décembre 2003

Après avis des rapporteurs :

François Pachet	Maître de conférences habilité	
Daniel Arfib	Directeur de Recherche	

Devant la commission d'examen composée de :

Daniel Arfib	Directeur de Recherche	Rapporteur
Myriam Desainte-Catherine	Professeur	Examinateur
William M. Hartmann	Professeur	Examinateur
François Pachet	Maître de conférences habilité	Rapporteur
Xavier Rodet	Professeur	Examinateur
Robert Strandh	Professeur	Président

– 2003 –

Modélisation statistique de sons bruités :
étude de la densité spectrale, analyse, transformation musicale et synthèse

---

Résumé : L'ordinateur offre de nouvelles possibilités en terme de traitement du son. Les travaux de recherche actuels concernent essentiellement la modélisation des parties périodiques d'un signal. Les sons bruités sont pourtant de plus en plus employés lors de la composition musicale.

Les modèles existants considèrent ces sons, ou leurs parties bruitées, comme du bruit blanc filtré. Cette hypothèse est mise en défaut par les études psychoacoustiques sur la perception de la densité spectrale des fréquences. Nous présentons un nouveau modèle spectral et statistique, appelé CNSS, basé sur le modèle des bruits thermiques. Un son bruité est défini par une somme fixe de sinusoïdes. Le modèle CNSS propose de nouveaux contrôles liés aux distributions des fréquences et des phases des sinusoïdes, et autorise ainsi de nouvelles transformations musicales.

Une méthode d'analyse générale des sons bruités est envisagée. Nous décrivons une première méthode qui caractérise dans un spectre les partiels des composantes bruitées, en étudiant statistiquement les variations d'intensité. Une seconde méthode permet de définir, sur l'échelle temporelle, les parties composées de transitoires, en étudiant la distribution des échantillons du signal. Enfin, nous détaillons une nouvelle technique qui permet d'approximer la densité spectrale. Les paramètres du modèle CNSS peuvent ainsi être estimés.

Par la suite, les algorithmes de synthèse liés au modèle CNSS sont détaillés. De plus, les problèmes dus au recouvrement de fenêtres lors de la synthèse de sons aléatoires sont expliqués. Plusieurs solutions adéquates sont proposées.

Une implémentation du modèle CNSS a été effectuée, permettant la synthèse en temps-réel de sons bruités. Cette implémentation est destinée d'une part à être intégrée dans l'instrument d'éveil musical *Dolabip* développé au SCRIME, et d'autre part, à devenir une plateforme d'exploration sonore et un outil de synthèse pour des expériences psychoacoustiques.

Discipline : Informatique

Mots-clés : modélisation statistique, partie stochastique de son, densité spectrale, analyse et synthèse de sons, bruits, synthèse additive en temps-réel.

LaBRI,
Université Bordeaux 1,
351, cours de la libération
33405 Talence Cedex (FRANCE)

Statistical modeling of noisy sounds:
spectral density, analysis, musical transformations, and synthesis

Abstract: Computers offer new possibilities concerning sound processing. Existing research essentially concerns the modeling of the periodic parts of sounds. However, noisy sounds are more and more used in musical compositions.

Existing models consider sounds, or their noisy parts, as filtered white noise. Psychoacoustic studies about the perception of the spectral density of frequencies show that this assumption represents a restriction. We present a new spectral and statistical model, called CNSS, based on the thermal noise model. A noisy sound is defined by a fixed number of sinusoids. The CNSS model proposes new controls related to the distributions of the frequency and phase of the sinusoids, and thus provides new musical transformations.

A general analysis method for noisy sounds is proposed. We describe a first method which characterizes partials of the noisy components in the spectrum by statistically studying their intensity fluctuations. A second method defines parts of the temporal representation that contain transients by studying the distribution of the samples of the signal. Lastly, we detail a new technique which approximates the spectral density. The parameters of the CNSS model can then be estimated.

Then, the synthesis algorithms related to the CNSS model are then detailed. Moreover, problems due to the window overlap during the synthesis of random sounds are explained and some solutions are proposed.

The CNSS model has been implemented and provides real-time synthesis of noisy sounds. This implementation is planned to be integrated in the computer-assisted early-learning tool *Dolabip*, which is developed at the SCRIME. It is also a platform for exploring sounds and a sound synthesis for psychoacoustic experiences.

Discipline: Computer Science

Keywords: statistical modeling, stochastic part, spectral density, analysis and synthesis of sound, noises, real-time additive synthesis.

LaBRI,
Université Bordeaux 1,
351, cours de la libération
33405 Talence Cedex (FRANCE)

Remerciements

Je voudrais tout d'abord remercier chaleureusement :

- Daniel Arfib pour avoir accepté d'être un rapporteur de cette thèse, mais aussi d'avoir initié les conférences DAFX auxquelles j'ai participé ;
- François Pachet pour avoir également accepté d'être un rapporteur de cette thèse ;
- William Hartmann et Xavier Rodet pour m'avoir fait l'honneur de participer à mon jury.

Cette thèse s'est déroulée dans l'équipe Son du LaBRI. Je remercie donc tous les membres de cette équipe, en particulier Robert Strandh pour avoir présidé mon jury de thèse, mais également pour m'avoir aidé à corriger mes articles en anglais. J'ai également eu la chance de travailler dans le cadre du SCRIME, dont je remercie les membres, en particulier ceux qui m'ont permis d'obtenir des sons afin de pouvoir effectuer des expériences, comme Guillaume Largillier et Jean-Michel Rivet.

Je tiens aussi à remercier spécialement Anthony Beurivé qui a non seulement accepté d'adapter les algorithmes que je proposais pour pouvoir les tester en temps réel, mais avec qui j'ai eu de fructueuses discussions théoriques et techniques. Nul doute que sa présence a contribué à une progression plus rapide de mes travaux.

Je remercie également les collègues du LaBRI que j'ai cotoyés durant ma thèse.

Je voudrais profiter de cette page pour remercier aussi mes collègues thésards de bureau : Afif toujours prêt à faire tout et n'importe quoi, Aymeric, Benjamin, Bertrand, Florian, Pascal, Pascal, Swarup, Sylvie, q-Yvan, et enfin Yon, avec qui j'ai passé tout l'été, et que j'ai finalement réussi à devancer malgré toutes ses tentatives désespérées.

Au cours de ces trois ans, le football et la musique m'ont aidé à décompresser. Je voudrais donc remercier :

- mes coéquipiers de l'équipe de football de l'Afodib (non, je ne démissionnerai pas) ;
- mes compagnons de *PLAY*, en particulier les plus fidèles, Afif et Laurent.

Enfin, je voudrais terminer ces remerciements par les personnes les plus importantes :

- Ma "chef" Myriam Desainte-Catherine pour son enthousiasme, sa confiance, sa disponibilité et son implication dans mes travaux. J'apprécie vraiment beaucoup notre collaboration ;
- Ma famille et surtout mes parents qui m'ont toujours soutenu ;
- Mon amoureuse Christine car elle m'a toujours encouragé, a relu ce document et parce qu'elle est toujours là, tout simplement.

Table des matières

Liste des tableaux

Table des figures

Introduction

L'ordinateur fait partie depuis le milieu du 20ème siècle des instruments de création musicale. Il offre de nouvelles perspectives aux musiciens en matière de contrôle, d'effets ou de composition. En particulier, leur puissance, qui ne cesse de croître, permet d'envisager des traitements de plus en plus complexes sur les sons. L'essentiel des recherches sur ces traitements porte toutefois sur les sons instrumentaux. Ce choix provient sûrement de la culture musicale qui donne une priorité à la mélodie et au rythme, en ne retenant alors qu'une faible partie des propriétés perceptives des sons.

Dans le passé, le bruit n'a pas été considéré comme un élément musical. Ce n'est qu'au début du 20ème siècle que des compositeurs ont commencé à intégrer ces sons riches et complexes dans des pièces musicales. Des compositeurs comme Pierre Schaeffer ont ainsi créé de la musique à partir de sons bruités naturels, et ont été amenés à définir un nouveau vocabulaire adapté à ce type de sons [Sch66].

L'essentiel des recherches a été consacré à la modélisation des parties périodiques du signal. Un son périodique de base, comme par exemple une sinusoïde pure, n'est pourtant pas un son naturel. En effet, les sons naturels sont composés d'une partie pouvant être assimilée à du bruit. L'oreille est particulièrement sensible aux informations liées à cette partie. C'est ce qui explique pourquoi les représentations de sons, qui ne prennent pas en compte cette composante bruitée, ne donnent pas une sensation de sons réels.

Ce n'est que depuis quelques années que la partie bruitée du son est prise en considération. La qualité de rendu des sons n'en est que meilleure. Toutefois, la représentation de cette partie repose sur des hypothèses peu restrictives et approximatives, car les sons étudiés restent majoritairement des sons instrumentaux peu bruités. Les travaux que nous présentons dans ce document ont été menés dans le but de développer un modèle de représentation mathématique des sons dont la partie bruitée est importante, que nous appellerons sons bruités, et de proposer ainsi des contrôles musicaux propres à leurs caractéristiques stochastiques.

Les modèles existants considèrent les parties bruitées des sons comme des bruits blancs filtrés, sans apporter de justification théorique ou expérimentale. La motivation première de nos travaux a donc été la vérification de cette affirmation sur le plan mathématique et sur le plan de la perception. La représentation par du bruit blanc filtré apparaît limitée, par rapport à la richesse et la complexité des sons bruités naturels. Les travaux psychoacoustiques menés par William Hartmann représentent les fondations de notre étude, en mettant en avant la faculté du système auditif à discerner des bandes de bruit de densités spectrales différentes, et en proposant donc une nouvelle dimension perceptive.

Nous nous sommes donc attachés à essayer de comprendre quelques éléments de la perception des sons bruités, parmi lesquels la densité spectrale, de façon à les intégrer dans un modèle destiné à la création musicale, et à proposer ainsi aux compositeurs des contrôles originaux sur des sons complexes.

Chapitre 1 : Signal numérique, musique et perception

Le traitement informatique du son nécessite des connaissances dans différents domaines. Tout d'abord, la représentation numérique implique l'application d'opérations mathématiques adaptées à la discrétisation des données. La complexité des sons bruités que nous souhaitons étudier nous impose l'utilisation d'outils statistiques. Les buts des traitements étant la création et la composition, le vocabulaire musical doit être pris en compte de façon à pouvoir établir des liens avec les paramètres physiques du son. Enfin, la compréhension du système auditif permet de retenir les informations relatives à la perception.

Nous introduisons dans le premier chapitre les éléments de traitement du signal numérique que nous appliquons dans tout ce document, avec notamment les problèmes relatifs à la numérisation du son, le passage de la représentation temporelle à la représentation fréquentielle par la transformée de Fourier, le filtrage des signaux, et la notion d'enveloppe temporelle.

Nous posons ensuite les définitions des outils statistiques que nous employons dans la suite, avec notamment les fonctions de densité de probabilité classiques, et quelques calculs de base relatifs aux fonctions sinusoïdales.

Par la suite, quelques notions musicales sont décrites comme le volume, le timbre, les partiels et les transitoires. En particulier, nous donnons les différentes définitions du bruit, et nous détaillons les propriétés des bruits les plus classiques.

Enfin, nous présentons des éléments de la perception des sons pris en considération dans le reste du document, tels que l'intensité, la hauteur, l'enveloppe temporelle, et le masquage.

Chapitre 2 : Modèles pour les sons

De nombreux modèles ont été développés pour établir un lien entre l'onde sonore et les éléments de perception, en définissant des paramètres mathématiques relatifs à des propriétés musicales. Nous présentons dans ce chapitre plusieurs types de modèles, pouvant notamment être appliqués pour représenter les sons bruités.

Tout d'abord, nous présentons les modèles physiques, basés sur la définition mécanique du son, et les modèles abstraits, basés sur des formules mathématiques. Nous discutons de leurs limites dans le cadre de transformations musicales.

Ensuite, nous abordons les modèles relatifs à la représentation temporelle des sons. L'étude des liens entre les paramètres statistiques et les propriétés musicales montre les limites de cette approche.

Les modèles spectraux établissent un lien plus fort entre la perception et les paramètres du modèle, en se basant sur la représentation fréquentielle. Les approches classiques sont présentées, en insistant sur la modélisation des parties bruitées des sons, qui repose sur une hypothèse principale : ces parties stochastiques sont complètement représentées par du bruit blanc filtré.

Nous présentons des travaux menés en psychoacoustique qui mettent en avant la faculté du système auditif à discerner des bandes de bruit de densités spectrales différentes. L'étude de ce nouveau paramètre met en évidence les limites de l'hypothèse précédente. Ainsi, nous proposons un nouveau modèle spectral et statistique, appelé CNSS, basé sur le modèle du bruit thermique décrivant un bruit comme une somme fixe de sinusoïdes, permettant de prendre en compte de nouvelles propriétés perceptives telles que la densité spectrale. Les transformations musicales permises par ce modèle CNSS sont détaillées et expliquées.

Chapitre 3 : Analyse de sons bruités

Les modèles de sons induisent des méthodes d'analyse précises de façon à extraire les paramètres mathématiques définis par le modèle de sons naturels, et permettre ainsi des transformations de sons réels. Nous proposons, dans ce chapitre, un principe d'analyse générale que nous envisageons pour étudier les sons bruités. Cette approche comporte trois étapes. Pour chacune de ces étapes, nous présentons les alternatives existantes et les méthodes originales que nous avons développées.

A partir de la représentation fréquentielle, il est d'abord nécessaire d'établir quelles parties correspondent à des partiels, de façon à pouvoir les traiter par des techniques de haute précision. Nous présentons donc les méthodes existantes de détection de partiels, ainsi qu'une nouvelle approche que nous avons mise au point, destinée particulièrement aux sons bruités, reposant sur l'étude statistique des fluctuations d'intensité.

A partir de la représentation temporelle, les fortes variations d'énergie, correspondant à des attaques, doivent être localisées de façon à être représentées par des modèles adaptés et prises en compte lors de manipulations d'échelle temporelle. Les méthodes existantes sont présentées, ainsi qu'une nouvelle approche statistique que nous proposons.

Enfin, le modèle CNSS que nous avons développé pour représenter des propriétés perceptives comme la densité spectrale, nécessite des méthodes d'analyse adaptées, qui, à notre connaissance, n'existent pas. La transformée de Fourier, de par ses limites liées au compromis entre la précision temporelle et fréquentielle, ne permet pas d'extraire les informations relatives aux paramètres du modèle CNSS, notamment celles liées à la densité spectrale.

Ainsi, nous présentons une méthode originale, basée également sur l'étude statistique des fluctuations d'intensité, qui permet d'approximer la densité spectrale de bandes de bruit. Nous détaillons les algorithmes de calcul, et nous expliquons les limites de cette approche.

Enfin, nous proposons, à partir de cette méthode d'approximation de la densité spectrale, des algorithmes d'estimation des paramètres du modèle CNSS, permettant d'envisager l'analyse des valeurs mathématiques de sons naturels, et donc l'application de ce modèle aux sons bruités réels.

Chapitre 4 : Synthèse de sons bruités

A partir de paramètres mathématiques d'un modèle, la représentation temporelle du son modélisé doit être synthétisée, pour, grâce à la carte son de l'ordinateur et des haut-parleurs, produire une onde sonore audible. Cette synthèse doit être effectuée en temps-réel de façon à ce que les manipulations musicales puissent être directement perçues par le musicien.

Nous présentons dans ce chapitre les techniques de synthèse de sons bruités. Nous débutons par les algorithmes de synthèse de bruits classiques, dont les enveloppes spectrales sont des courbes mathématiques simples, en prenant en considération les propriétés statistiques de leur représentation temporelle.

Ensuite, nous détaillons les méthodes de synthèse spectrale de sons bruités, en discutant particulièrement leurs liens avec le modèle théorique du bruit thermique présenté dans le chapitre 2. La méthode de synthèse propre au modèle CNSS est ensuite décrite, en détaillant notamment les algorithmes de détermination aléatoire des caractéristiques des sinusoïdes composant le son synthétisé, à partir des paramètres mathématiques du modèle.

Les algorithmes performants existants de synthèse rapide de sinusoïdes qui peuvent être appliqués à la synthèse du modèle CNSS sont ensuite présentés et discutés.

Enfin, nous expliquons de nouveaux problèmes apparus lors de nos expérimentations, relatifs au caractère stochastique des sons étudiés, concernant notamment la technique de recouvrement de fenêtres de synthèse et le contrôle de la saturation du son. Pour le premier cas, le problème est détaillé du point de vue théorique, et plusieurs algorithmes permettant d'éviter les artefacts audibles induits sont proposés et discutés. Concernant l'amplitude du son, les différentes alternatives sont données.

Chapitre 5 : Applications et perspectives

Dans le dernier chapitre, nous présentons tout d'abord les applications possibles du modèle CNSS. L'implémentation du modèle, sous forme de bibliothèque et d'objets jMax, a permis d'effectuer des tests de synthèse et de confirmer son application en temps-réel. Le modèle CNSS, qui propose de nombreux paramètres de contrôle statistique, va ainsi être intégré dans l'outil pédagogique d'éveil musical Dolabip, développé au SCRIME.

Enfin, pour conclure, nous proposons quelques idées pouvant permettre d'améliorer le modèle CNSS en vue de la création musicale, en affinant notamment les méthodes d'estimation des paramètres à partir de sons naturels. Le modèle CNSS doit également permettre de servir de plateforme de tests psychoacoustiques, pour évaluer plus en détails l'influence des paramètres du modèle sur la perception. Une nouvelle distribution des fréquences est ainsi proposée, qui pourrait aboutir rapidement à une diminution du nombre de sinusoïdes nécessaires à la synthèse de sons bruités. Le modèle CNSS est surtout destiné à compléter d'autres modèles de sons non bruités, de façon à étendre leurs possibilités.

Chapitre 1

Signal numérique, musique et perception

L'onde sonore est un objet physique formé par des dépressions d'air se déplaçant dans toutes les directions de l'espace. L'étude de cette onde peut s'effectuer de plusieurs manières, selon qu'elle soit par exemple considérée par son aspect physique ou par les sensations qu'elle laisse percevoir. Parmi ces données, un modèle de représentation des sons doit prendre en compte celles qui sont pertinentes concernant les objectifs de la modélisation. La mise en oeuvre d'un modèle nécessite donc des connaissances pluridisciplinaires, comme l'illustre la figure 1.1. Dans ce chapitre, nous proposons quelques rappels et définitions de base que nous utilisons dans la suite du document. Dans la section 1.1, nous présentons quelques éléments de traitement du signal numérique. Comme nous allons particulièrement nous intéresser aux signaux aléatoires, nous donnons quelques définitions de mathématiques et de statistiques dans la section 1.2. Le vocabulaire d'acoustique musicale propre aux compositeurs et aux musiciens est expliqué dans la section acoustique musicale 1.3. Les signaux que nous considérons sont essentiellement des bruits, nous définissons ces signaux dans la section 1.4. Enfin, nous abordons dans la section 1.5 les notions essentielles de psychoacoustique.

1.1 Éléments de traitement du signal numérique

Le son est avant tout un signal. L'étude de la représentation des sons s'appuie sur des outils mathématiques du traitement du signal. Nous présentons dans cette section quelques uns de ces outils, et quelques notations que nous utiliserons dans toute la suite.

1.1.1 Signal numérique et échantillonnage

Représentation continue du son

Un signal audio peut être représenté par les variations d'amplitude de l'onde sonore en fonction du temps. Cette sorte de représentation est la représentation temporelle. Le signal

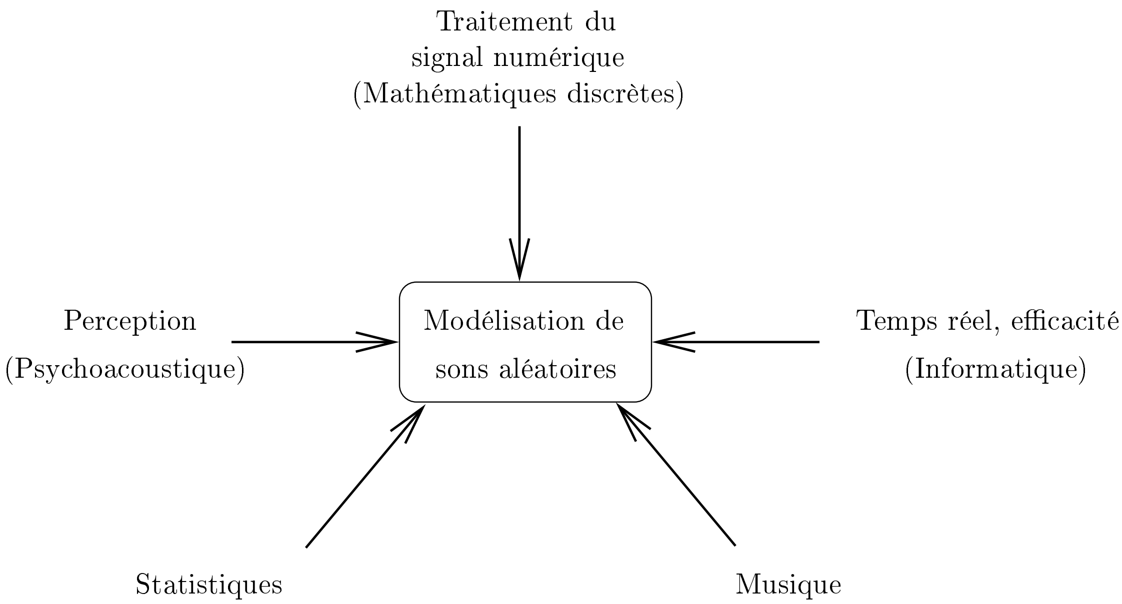

FIG. 1.1 – La modélisation des signaux aléatoires nécessite des connaissances en différents domaines.

est modélisé comme des variations continues de l'amplitude en fonction du temps t mesuré en secondes. Le signal est donc noté $x(t)$, où $t \in \mathbb{R}$.

Signal numérique

La représentation numérique d'un son est permise grâce à des convertisseurs analogique-numérique. Ceux-ci effectuent une quantification de la pression acoustique que crée l'onde sonore (mesurées sous forme de tension électrique à l'aide d'un microphone) à des intervalles de temps réguliers. Ces intervalles définissent la fréquence d'échantillonnage, notée F_s (par exemple 44, 1kHz pour une qualité CD, c'est-à-dire 44100 échantillons par seconde).

Les valeurs numériques ainsi obtenues permettent un traitement mathématique du signal. On parle alors de signal discret qui sera noté $x[n]$, où $n \in \mathbb{N}$.

Échantillonnage

L'une des règles de base de l'échantillonnage est la nécessité d'avoir au moins deux échantillons par période du signal pour pouvoir avoir les informations suffisantes concernant les fréquences qui le composent. C'est pourquoi il faut s'assurer que la fréquence d'échantillonage soit au moins le double de la fréquence du signal échantillonné. La fréquence limite est appelée *fréquence de Nyquist* [Roa98]. Comme l'oreille humaine n'est pas sensible aux fréquences supérieures à 20kHz, la fréquence d'échantillonnage doit être supérieure à 40kHz.

1.1.2 Convolution de signaux

La technique de convolution s"applique à deux signaux et permet des transformations sonores [Roa97]. L'opération de convolution (opération noté *) de deux signaux notés x et y est définie par les équations (1.1) et (1.2), respectivement dans le cas de signaux continus et de signaux discrets :

$$\forall t, \quad x(t) * y(t) = \int_{-\infty}^{\infty} x(k)y(k-t)dk \tag{1.1}$$

$$\forall n, \quad x[n] * y[n] = \sum_{-\infty}^{\infty} x[k]y[k-n]\Delta k \tag{1.2}$$

1.1.3 Fonction d'autocorrélation

L'étude des périodicités d'un signal peut s'effectuer à partir de l'observation des similitudes du signal avec ce même signal décalé temporellement. La fonction d'autocorrélation, notée Γ, permet d'étudier ces similitudes. Il existe deux définitions selon que le signal est de durée finie ou infinie, ainsi qu'une définition statistique.

Définition dans le cas fini

Dans le cas d'un signal fini, la définition de la fonction d'autocorrélation est donnée par :

$$\forall \tau, \quad \Gamma(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt \tag{1.3}$$

Cette fonction est donc la corrélation d'un signal avec lui-même décalé. Le temps de décalage est noté τ.

Définition dans le cas infini

Dans le cas infini, la définition de la fonction d'autocorrélation est donnée par :

$$\Gamma(\tau) = \lim_{T_0 \to \infty} \frac{1}{2T_0} \int_{-T_0}^{T_0} x(t)x(t+\tau)dt \tag{1.4}$$

Elle est définie par la convergence de l'intégrale définie ci-dessus, par rapport à une durée notée T_0, qui tend vers l'infini.

Définition statistique

Une autre définition statistique de la fonction d'autocorrélation existe, qui utilise la définition de l'espérance mathématique (voir section 1.2) :

$$\Gamma(\tau) = \mathcal{E}[X(t)X(t+\tau)] \tag{1.5}$$

La valeur de la fonction d'autocorrélation en 0 est la variance du signal aléatoire.

$$\Gamma(0) = \mathcal{E}[X^2(t)] = \sigma^2 = V(X) \tag{1.6}$$

Application et limites

Cette fonction d'autocorrélation est en général appliquée à des signaux discrets de longueur finie W_s échantillons, pour la détection de périodicités. L'algorithme de calcul se traduit par l'équation :

$$\forall \tau \in [0; W_s - k], \Gamma(\tau) = \frac{1}{W_s} \sum_{i=0}^{k} x[i]x[i+\tau] \tag{1.7}$$

Les choix de W_s et k sont prépondérants. La périodicité recherchée doit notamment être inférieure à la durée correspondant au nombre d'échantillons $W_s - k$.

Cas d'une sinusoïde

Dans cette sous-section, la fonction d'autocorrélation est exprimée dans le cas d'un signal sinusoïdal pur. Notons x ce signal :

$$\forall t, \quad x(t) = a_0 \sin(2\pi f_0 t + \phi_0) \tag{1.8}$$

La fonction d'autocorrélation pour ce signal x est :

$$\forall \tau, \quad \Gamma_x(\tau) = \lim_{T_0 \to \infty} \frac{a_0^2}{2T_0} \int_{-T_0}^{T_0} \sin(2\pi f_0 t + \phi_0) \sin(2\pi f_0 (t+\tau) + \phi_0) dt \tag{1.9}$$

Les relations de trigonométrie classiques permettent de simplifier cette équation, puisque :

$$\sin(2\pi f_0 (t+\tau) + \phi_0) = \sin(2\pi f_0 t + \phi_0) \cos(2\pi f_0 \tau) + \cos(2\pi f_0 t + \phi_0) \tag{1.10}$$

La fonction d'autocorrélation s'écrit donc :

$$\begin{aligned} \forall \tau, \quad \Gamma_x(\tau) &= \lim_{T_0 \to \infty} \frac{a_0^2}{2T_0} \int_{-T_0}^{T_0} \sin^2(2\pi f_0 t + \phi_0) \cos(2\pi f_0 \tau) \\ &+ \cos(2\pi f_0 t + \phi_0) \sin(2\pi f_0 t + \phi_0) \sin(2\pi f_0 \tau) dt \\ \Gamma_x(\tau) &= \lim_{T_0 \to \infty} \frac{a_0^2}{2T_0} \int_{-T_0}^{T_0} \sin^2(2\pi f_0 t + \phi_0) \cos(2\pi f_0 \tau) \\ &+ \frac{1}{2} \sin(4\pi f_0 t + \phi_0) \sin(2\pi f_0 \tau) dt \end{aligned} \tag{1.11}$$

Sachant que les intégrales de sinus sont nulles pour une fréquence non nulle (voir section 1.1.4) et que les intégrales de sinus au carré valent $\frac{1}{2}$, il vient :

$$\begin{aligned} \forall \tau, \quad \Gamma_x(\tau) &= \lim_{T_0 \to \infty} \frac{a_0^2}{2T_0} \int_{-T_0}^{T_0} \sin^2(2\pi f_0 t + \phi_0) \cos(2\pi f_0 \tau) dt \\ &= \lim_{T_0 \to \infty} \frac{a_0^2}{4T_0} 2T_0 \cos(2\pi f_0 \tau) \\ &= \frac{a_0^2}{2} \cos(2\pi f_0 \tau) \end{aligned} \tag{1.12}$$

Tout d'abord, cette équation montre que la fonction d'autocorrélation d'un signal sinusoïdal pur est indépendante de la valeur de la phase ϕ_0. En effet, quelle que soit la phase du signal, la similitude entre le signal et le signal retardé reste la même. D'autre part la fonction d'autocorrélation est une sinusoïde dont le maximum est en $\tau = 0$. Cette fonction d'autocorrélation Γ est périodique de période $\frac{1}{f_0}$. Cela montre bien que la similitude est maximale périodiquement, chaque fois que le retard τ est un multiple de la période du signal. La figure 1.2 illustre ces propriétés en montrant un signal sinusoïdal et son autocorrélation.

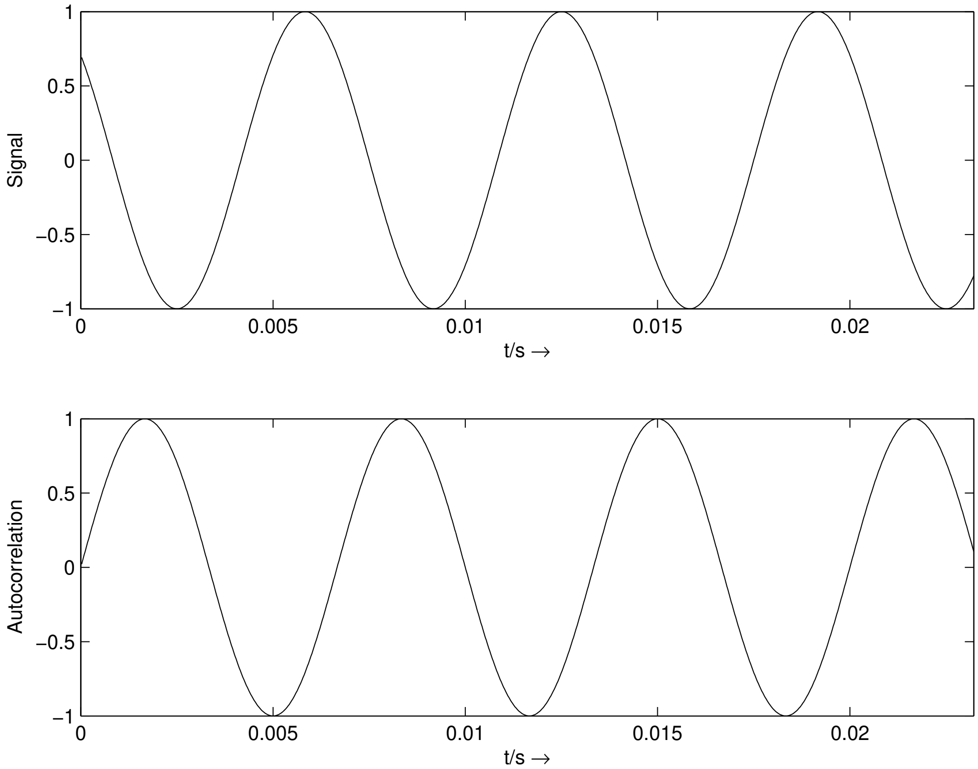

FIG. 1.2 – Signal sinusoïdal pur (en haut) et son autocorrélation (en bas).

Propriétés

– La fonction d'autocorrélation est une fonction paire :

$$\forall \tau, \quad \Gamma(\tau) = \Gamma(-\tau) \tag{1.13}$$

– Comme la fonction d'autocorrélation traduit la similitude entre le signal et une version retardée de lui-même, il est évident qu'elle prend une valeur maximale lorsque le retard est nul ($\tau = 0$) :

$$\forall \tau, \quad \Gamma(0) \geq |\Gamma(\tau)| \tag{1.14}$$

Seul le cas de signaux périodiques permet de définir des valeurs de la fonction égales aux valeurs de l'origine ($\Gamma(\tau) = \Gamma(0), \forall \tau \neq 0$). Dans ce cas-là, la fonction d'autocorrélation est périodique (voir un exemple pour un signal sinusoïdal pur dans l'équation (1.12). Ce maximum est l'énergie totale du signal, puisqu'il est défini par l'intégrale de la puissance du signal :

$$E_0 = \Gamma(0) = \int_{-\infty}^{\infty} x^2(t)dt \tag{1.15}$$

– Dans le cas de signaux non infinis, qui exclut le cas de signaux périodiques, la fonction d'autocorrélation converge vers 0 à l'infini :

$$\lim_{\tau \to \infty} |\Gamma(\tau)| = 0 \tag{1.16}$$

Théorème de Wiener-Khintchine

Le théorème de Wiener-Khintchine donne un résultat particulièrement important pour la compréhension et l'utilisation de la fonction d'autocorrélation lors de l'étude d'un signal. Il établit que la transformée de Fourier du résultat de l'application de la fonction d'autocorrélation à un signal est tout simplement le spectre de puissance du signal, c'est-à-dire le spectre composé des normes au carré des composantes du spectre du signal. La démonstration de ce théorème peut être trouvée, par exemple, dans [Har97].

Soit $X(k)$ la transformée de Fourier du signal $x(n)$ et $A(k)$ celle de son autocorrélation $\Gamma(\tau)$. Le théorème de Wiener-Khintchine s'écrit alors :

$$\forall k, \quad A(k) = |X(k)|^2 \tag{1.17}$$

1.1.4 Rappel de trigonométrie et nombres complexes

Nous proposons dans cette partie quelques rappels de trigonométrie pour introduire les notations utilisées pour tout ce document, ainsi que quelques calculs d'intégrales de fonctions sinusoïdales classiques. Ces calculs seront utiles dans les chapitres et sections suivantes. Nous notons T une durée fixée.

Les nombres complexes sont notés $c = a\exp(i\phi)$, où $i^2 = -1$, a est la norme (ou amplitude) du nombre complexe c et ϕ est sa phase. De plus, nous notons $\mathrm{Re}(c)$ et $\mathrm{Im}(c)$ respectivement la partie réelle et la partie imaginaire de c. Ainsi, le nombre complexe c s'écrit aussi $c = \mathrm{Re}(c) + i\mathrm{Im}(c)$. La partie réelle vaut $\mathrm{Re}(c) = a\cos(i\phi)$ et la partie imaginaire vaut $\mathrm{Im}(c) = a\sin(i\phi)$.

Nous notons T_k les périodes (en secondes), $f_k = \frac{1}{T_k}$ et $\omega_k = 2\pi f_k$, respectivement les fréquences (en secondes) et les pulsations (en radians par secondes) associées à ces périodes.

Intégrales de fonctions sinus et cosinus

L'intégrale d'une sinusoïde de fréquence non nulle ($\omega \neq 0$) est nulle :

$$\forall a, \forall \omega \neq 0, \forall \phi, \quad \frac{1}{T}\int_0^T a\sin(\omega t + \phi)dt = \frac{-a}{2\pi f T}[\cos(\omega T) - 1] = 0 \tag{1.18}$$

Il est important de noter que ce calcul est le même pour une sinusoïde, à une différence de phase près.

Intégrales de carrés de fonctions sinus et cosinus

L'intégrale d'une sinusoïde de fréquence non nulle ($\omega \neq 0$) est donnée par :

$$\begin{aligned} \forall a, \forall \omega \neq 0, \forall \phi, \quad \frac{1}{T}\int_0^T a\sin^2(\omega t+\phi)dt &= \frac{1}{T}\int_0^T \frac{a}{2}(1-\cos(2\omega t+2\phi))dt \\ &= \frac{a}{2} \end{aligned} \tag{1.19}$$

Intégrales de produits de fonctions sinus et cosinus

Les intégrales de produits de sinusoïdes donnent des résultats différents, selon deux cas possibles.

– Dans le cas $f_1 \neq f_2$:

$$\begin{aligned} I_{f_1,f_2} &= \frac{1}{T}\int_0^T a_1\cos(\omega_1 t+\phi_1)a_2\cos(\omega_2 t+\phi_2)dt \\ &= \frac{a_1a_2}{2T}\int_0^T \cos((\omega_1-\omega_2)t+\phi_1-\phi_2)+\cos((\omega_1+\omega_2)t+\phi_1+\phi_2)dt \\ &= 0 \end{aligned} \tag{1.20}$$

– Dans le cas $f_1 = f_2 = f$:

$$\begin{aligned} I_{f_1,f_2} &= \frac{a_1a_2}{2T}\int_0^T \cos(\phi_1-\phi_2)+\cos(2\omega t+(\phi_1+\phi_2))dt \\ &= \frac{a_1a_2}{2T}\int_0^T \cos(\phi_1-\phi_2)dt \\ &= \frac{a_1a_2}{2}\cos(\phi_1-\phi_2) \end{aligned} \tag{1.21}$$

1.1.5 Transformée de Fourier

La transformée de Fourier doit son nom au mathématicien français J.B. Fourier (1768-1830). Elle définit une série d'opérations mathématiques qui permet d'associer à une onde (un signal), une série de sinusoïdes de fréquences, d'amplitudes et de phases déterminées. Elle permet donc de passer de la représentation temporelle (amplitude en fonction du temps) à la représentation spectrale (amplitude en fonction de la fréquence).

Définition

La transformée de Fourier permet de définir le spectre complexe $X(f)$ du signal $x(t)$ par la relation :

$$X(f) = \int_{-\infty}^{\infty} x(t)\exp^{2\pi jft}\,dt \tag{1.22}$$

Propriétés

La transformée de Fourier possède quelques propriétés essentielles qui sont mises en application dans les chapitres suivants. Parmi ces propriétés, il est important de noter que :

- La transformée de Fourier est une application linéaire. Ainsi, pour deux signaux x et y, et pour deux réels α et β :

$$\mathrm{FT}(\alpha x + \beta y) = \alpha X(f) + \beta Y(f) \tag{1.23}$$

- La transformée de Fourier d'un produit de deux signaux est la convolution des transformées de Fourier des signaux. Ainsi, pour deux signaux x et y :

$$\mathrm{FT}(xy) = X(f) * Y(f) \tag{1.24}$$

$$\mathrm{FT}(x * y) = X(f)Y(f) \tag{1.25}$$

1.1.6 Transformée de Fourier inverse

La transformée de Fourier est une application réversible. La transformée inverse de Fourier est définie par la relation :

$$x(t) = \int_{-\infty}^{\infty} X(f) \exp^{2\pi j f t} df \tag{1.26}$$

1.1.7 Transformée discrète de Fourier

La transformée de Fourier est adaptée aux signaux numériques (échantillonnés) de durée limité W_s, par la transformée de Fourier discrète (DFT pour *discrete Fourier Transform*) :

$$X(k) = \frac{1}{W_s} \sum_{n=0}^{W_s-1} x(n) e^{-j\frac{2\pi}{W_s}nk} \tag{1.27}$$

où x représente le signal de longueur W_s échantillons et $X(k)$ l'estimation du spectre.

Discrétisation des fréquences

Comme le signal analysé est discret (suite d'échantillons), la transformée discrète de Fourier définit des amplitudes correspondant à des fréquences discrètes, notées F_k. Cette discrétisation est évidemment reliée à la fréquence d'échantillonnage F_s, mais aussi à la taille de la fenêtre de signal W_s pour laquelle le calcul est effectué. Les fréquences f_k valent donc :

$$\forall k \in \{0; W_s\}, \quad f_k = k\frac{F_s}{W_s} \tag{1.28}$$

Amplitudes et phases

Les amplitudes et les phases sont directement accessibles à partir des complexes $X(k)$ donnés par la transformée de Fourier :

$$A_k = |X(k)| = \sqrt{[\mathrm{Re}(X(k))]^2 + [\mathrm{Im}(X(k))]^2} \tag{1.29}$$

$$\phi_k = \arg(X(k)) = \arctan \frac{\mathrm{Im}(X(k))}{\mathrm{Re}(X(k))} \tag{1.30}$$

Transformée rapide de Fourier

L'algorithme de transformée rapide de Fourier (*Fast Fourier Transform* ou FFT) permet de calculer la transformée de Fourier d'un signal de longueur W_s avec une complexité de $O(W_s \log_2(W_s))$, avec la seule contrainte que W_s soit une puissance de 2. Le principe de cet algorithme peut être facilement trouvé dans de nombreux ouvrages ([Moo90] par exemple) et nous ne le détaillons pas à nouveau ici.

Une implémentation de cet algorithme est proposée par le MIT sous le nom de FFTW (*Fastest Fourier Transform in the West*) [FJ98].

1.1.8 Filtrage de signaux

Le filtrage est une des opérations les plus classiques en traitement du signal. D'ailleurs, de nombreux traitements comme la transformée de Fourier peuvent être décrits comme un ou des filtrages. Le filtrage est défini comme l'opération qui multiplie le spectre complexe d'un son en entrée par la fonction de transfert d'un filtre. Comme cette opération est effectuée sur des nombres complexes, elle consiste à multiplier les amplitudes et à ajouter les phases. Plus concrètement, elle consiste à amplifier ou préserver certaines fréquences tout en supprimant ou atténuant d'autres fréquences du spectre d'un son. En notant $X(f)$ la transformée de Fourier du signal en entrée $x(t)$ et $Y(f)$ la transformée de Fourier du signal en sortie, le filtrage peut être formalisé par :

$$\forall f, \quad Y(f) = H(f)X(f) \tag{1.31}$$

où $H(f)$ est la fonction de transfert du filtre.

Plusieurs familles de filtres sont particulièrement utilisées :

- Les filtres passe-bas (et passe-haut) consistent à atténuer (amplifier) les hautes (basses) fréquences tout en amplifiant (atténuant) les basses (hautes) fréquences du signal d'entrée. Ainsi, le son de sortie sera plus grave (aigu).
- Les filtres passe bande consistent à amplifier une bande de fréquences tout en atténuant toutes les fréquences se trouvant autour de cette bande.
- Les filtres variant avec le temps sont particulièrement utiles pour l'analyse et la synthèse de spectres. En effet, le spectre d'un son peut par exemple être modélisé par un bruit blanc filtré. Dans ce cas, le son est décomposé sur l'échelle du temps en morceaux, qui seront chacun représentés par un filtre différent. Cette modélisation sera abordée plus en détail dans la suite.

Comme cela a été précisé précédemment, l'opération de multiplication dans le domaine fréquentiel revient à une opération de convolution dans le domaine temporel. C'est à partir de cette propriété que peut être effectuée l'opération de filtrage dans le domaine temporel, en définissant la réponse impulsionnelle (notée $h(t)$) du filtre comme le résultat de la transformée inverse de Fourier de la fonction de transfert $H(f)$.

1.1.9 Enveloppe temporelle

L'enveloppe temporelle est un paramètre perceptif primordial (voir la section 1.5.5). Dans le langage courant, le terme enveloppe fait généralement référence à l'enveloppe spectrale. Il est important de noter la différence entre une enveloppe spectrale et une enveloppe temporelle.

L'enveloppe temporelle est définie comme la courbe de plus faibles variations qui *enveloppe* le signal sur sa représentation temporelle. Un exemple de signal composé de sinusoïdes, et son enveloppe temporelle sont représentés par la figure 1.3. Les formulations présentées ci-dessous sont tirées de [Har97] et permettent de définir mathématiquement ce paramètre qui sera utilisé dans toute la suite. C'est pourquoi nous avons choisi de détailler ici ces calculs.

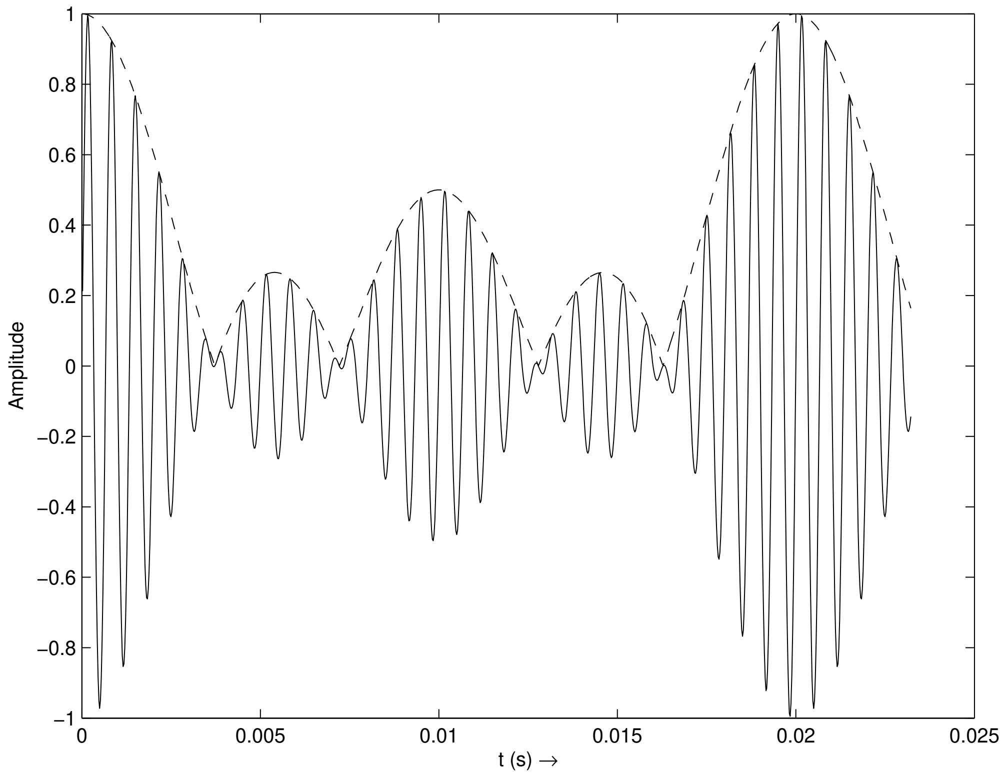

FIG. 1.3 – Représentation temporelle d'un signal composé de sinusoïdes (en ligne pleine) avec son enveloppe temporelle (en hachurés).

Définition mathématique

Pour pouvoir formaliser mathématiquement l'enveloppe d'un signal, l'hypothèse est faite que ce signal étudié, noté $x(t)$, peut être décomposé en une somme de N sinusoïdes. Nous verrons plus loin que cette hypothèse n'est pas restrictive puisqu'elle est valable pour des signaux complexes comme les bruits. Ainsi, nous pouvons écrire :

$$\forall t, \quad x(t) = \sum_{n=1}^{N} a_n \cos(\omega_n t + \phi_n) \tag{1.32}$$

Les pulsations composant le signal peuvent être exprimées comme une pulsation caractéristique notée ϖ à laquelle est ajoutée une pulsation de faible valeur notée δ_n :

$$\forall n \in \{1; N\}, \quad \omega_n = \varpi + \delta_n \tag{1.33}$$

Les formules classiques de trigonométrie permettent d'écrire :

$$\begin{aligned} \forall t, \quad x(t) &= \sum_{n=1}^{N} a_n \cos(\delta_n t + \phi_n) \cos(\varpi t) - \sum_{n=1}^{N} a_n \sin(\delta_n t + \phi_n) \sin(\varpi t) \\ &= R(t) \cos(\varpi t) - I(t) \cos(\varpi t) \end{aligned} \tag{1.34}$$

avec

$$\forall t, \quad \begin{cases} R(t) = E(t) \cos(\Phi(t)) = \sum_{n=1}^{N} a_n \cos(\delta_n t + \phi_n) \\ I(t) = E(t) \sin(\Phi(t)) = \sum_{n=1}^{N} a_n \sin(\delta_n t + \phi_n) \end{cases} \tag{1.35}$$

ce qui peut être exprimé sous forme complexe par :

$$R(t) + iI(t) = \sum_{n=1}^{N} a_n \exp[i(\delta_n t + \phi_n)] \tag{1.36}$$

Ainsi, l'enveloppe $E(t)$ s'écrit :

$$E(t) = |\sum_{n=1}^{N} a_n \exp[i(\omega_n + \phi_n)]| \tag{1.37}$$

L'enveloppe temporelle n'est donc juste qu'un élément d'une représentation mathématique possible d'un signal.

Propriétés

Dans cette section, nous présentons quelques propriétés découlant de la définition de l'enveloppe temporelle :

- L'enveloppe temporelle d'un signal est invariante en cas de translation spectrale du signal. En effet, dans ce cas, seule la valeur ϖ de l'équation (1.34) est modifiée. Or l'enveloppe temporelle est définie par les écarts des fréquences composant le signal, qui, eux, restent constants.
- L'enveloppe temporelle est invariante si les phases des sinusoïdes du signal sont incrémentées de la même valeur, notée $\Delta\phi$. Cette propriété est montrée par la définition donnée par l'équation (1.37) puisque le facteur $\exp(i\Delta\phi)$, de norme 1, peut y être mis en facteur. Cette propriété peut également être étendue aux incrémentations de phases proportionnelles aux fréquences de chaque sinusoïde composant le signal. Dans ce cas, cette modification revient à un changement d'échelle temporelle.

Puissance de l'enveloppe temporelle

L'enveloppe temporelle est donc définie par une norme. Il est alors plus aisé de travailler avec la puissance de cette fonction :

$$\forall t, \quad E^2(t) = \sum_{n=1}^{N} a_n^2 + 2\sum_{n=1}^{N}\sum_{n'=1}^{n-1} a_n a_{n'} \cos[(\omega_n - \omega_{n'})t + \phi_n - \phi_{n'}] \tag{1.38}$$

Cette formule met bien en évidence une propriété qui est utile dans la suite : la puissance de l'enveloppe temporelle d'un signal ne dépend que des différences des fréquences et des différences de phases des sinusoïdes composant ce signal.

A partir de cette puissance, la puissance moyenne de cette enveloppe peut être explicitée. En effet, en tenant compte des intégrales données en section 1.1.4, cette puissance moyenne s'écrit :

$$\begin{aligned}\forall t, \quad \overline{E^2(t)} &= \overline{\sum_{n=1}^{N} a_n^2} + 2\overline{\sum_{n=1}^{N}\sum_{n'=1}^{n-1} a_n a_{n'} \cos[(\omega_n - \omega_{n'})t + \phi_n - \phi_{n'}]} \\ &= \sum_{n=1}^{N} a_n^2 \end{aligned} \tag{1.39}$$

Par ailleurs, la puissance moyenne de l'enveloppe temporelle d'un signal peut être aisément reliée à la puissance moyenne du signal. Toujours dans le cas général d'un signal composé de sinusoïdes, sa puissance s'écrit :

$$\begin{aligned}\overline{x^2(t)} &= \overline{(\sum_{n=1}^{N} a_n \cos(\omega_n t + \phi_n))^2} \\ &= \overline{(\sum_{n=1}^{N} a_n^2 \cos^2(\omega_n t + \phi_n))} + \overline{(\sum_{n=1}^{N}\sum_{n'=1, n\neq n'}^{N} a_n \cos(\omega_n t + \phi_n))^2} \qquad (1.40) \\ &= \frac{1}{2}\sum_{n=1}^{N} a_n^2 \qquad (1.41)\end{aligned}$$

A partir de ces deux équations (1.39) et (1.41), nous pouvons établir la relation entre la puissance moyenne du signal et la puissance moyenne de son enveloppe temporelle :

$$\overline{E^2(t)} = 2\overline{x^2(t)} \tag{1.42}$$

1.2 Éléments de mathématiques et statistiques

Dans la suite, nous nous intéressons aux signaux aléatoires. La modélisation de ces signaux requiert l'utilisation des outils mathématiques et statistiques adaptés. Dans cette section, nous introduisons les notations employées dans ce document et nous rappelons rapidement quelques concepts de base.

1.2.1 Processus aléatoire

Un processus aléatoire $X[n, \Omega]$ (ou signal aléatoire discret) est une fonction à deux dimensions [DQ96]. La première dimension est le temps, la deuxième représente la réalisation (notée Ω). Nous notons $\mathcal{E}(X)$ l'espérance mathématique du processus aléatoire.

En considérant n constant, le processus aléatoire X est une variable aléatoire caractérisée par une loi de probabilité. Dans ce cas, notre étude est placée dans le domaine statistique. Nous noterons alors l'espérance $\mathcal{E}(X) = \langle X \rangle$.

Si Ω est constant, le processus aléatoire est une fonction temporelle. Notre étude est alors placée dans le domaine temporel. Nous noterons dans ce cas l'espérance $\mathcal{E}(X) = \overline{X}$. Ces deux approches ne sont équivalentes que si le processus est dit ergodique (voir ci-dessous dans la sous-section 1.2.3).

1.2.2 Moments

Dans cette sous-section, nous définissons les moments statistiques et temporels, et nous présentons les quatre premiers moments.

Définition

Un signal aléatoire est caractérisé par la connaissance de ses moments statistiques et/ou temporels. Cependant, pour certains types de signaux, la connaissance des premiers moments suffisent.

Les moments statistiques sont calculés en moyennant sur les réalisations alors que les moments temporels sont calculés en moyennant sur le temps. Les moments d'ordre k, notés m_k, et les moments centrés associés, notés σ_k, sont définis par la relation :

$$m_k = \mathcal{E}(X^k) \tag{1.43}$$

$$\sigma_k = \mathcal{E}(X - m_k)^k \tag{1.44}$$

Moment d'ordre 1 et moyenne

D'après la définition donnée par l'équation (1.44), le moment d'ordre 1 est $m_1 = \mathcal{E}(X)$, c'est-à-dire la moyenne du signal. En pratique, les signaux ont tendance à avoir une moyenne nulle. C'est pourquoi nous considérons cette hypothèse dans la suite.

Moment d'ordre 2 et variance

Le moment d'ordre 2 est $m_2 = \mathcal{E}(X^2)$, aussi appelé la variance. Ce moment caractérise la dispersion des valeurs du signal autour de la moyenne. Nous notons V la variance. L'écart-type, noté σ, est défini comme :

$$\sigma_X = \sqrt{m_2} = \sqrt{V(X)} \tag{1.45}$$

Moment d'ordre 3 et skewness

Une grandeur appelée skewness et notée S_X est définie à partir du moment d'ordre 3 :

$$S_X = \frac{\mathcal{E}(X^3)}{\sigma^3} \tag{1.46}$$

Cette mesure caractérise la symétrie de la densité de probabilité. Une variable aléatoire de densité de probabilité symétrique a un skewness nul. Plus la densité de probabilité est asymétrique, plus le skewness a une grande valeur. Quelques valeurs pour des distributions classiques sont données dans le tableau 1.1.

Moment d'ordre 4 et kurtosis

Une grandeur appelée kurtosis et notée K_X est définie à partir du moment d'ordre 4 :

$$K_X = \frac{\mathcal{E}(X^4)}{\sigma^4} \tag{1.47}$$

Cette mesure caractérise la forme de la densité de probabilité. Plus le kurtosis est important, plus la densité de probabilité est *pointue*. A l'opposé, un kurtosis faible signifie que la densité de probabilité est plate. Quelques valeurs pour des distributions classiques sont données dans le tableau 1.1.

Loi de Distribution	skewness	kurtosis
Uniforme	0	1.8
Gaussienne	0	3.0
Rayleigh	0.631	3.245

TAB. 1.1 – Valeurs de skewness et kurtosis pour quelques distributions classiques.

1.2.3 Ergodicité

Un processus aléatoire est dit ergodique si les résultats des études de son évolution au cours du temps sont les mêmes que les résultats des études sur les réalisations successives. Plus concrètement, un signal aléatoire discret est analysé sur une durée de temps limitée. Or une autre réalisation de ce signal aléatoire peut donner des résultats d'analyse différents. Il faudrait donc effectuer un grand nombre d'analyse sur un grand nombre de réalisations du signal pour avoir une information statistique de qualité. L'hypothèse d'ergodicité nous permet de supposer que les résultats sont les mêmes si une seule réalisation du signal est exploitée, et que l'étude est effectuée sur le temps. Dans ce cas, nous pouvons écrire :

$$\mathcal{E}(X) = \langle X \rangle = \overline{X} \tag{1.48}$$

Ainsi, une conséquence directe de cette hypothèse est la méthode de calcul des moments. Une différence existe entre les moments statistiques (calculés sur plusieurs réalisations) et les moments temporels (calculés par rapport au temps). L'hypothèse d'ergodicité, en les considérant comme égaux, permet de ne prendre en compte qu'un seul type de moments.

$$\mathcal{E}(X) = \lim_{T\to\infty} \frac{1}{T} \int_0^T X[n,\Omega]dn \tag{1.49}$$

1.2.4 Stationnarité

Un processus aléatoire est stationnaire si et seulement si ses propriétés statistiques sont indépendantes de l'intervalle de temps utilisé lors de leur définition. En notant $t = \{t_1, t_2, \ldots, t_n\}$ et $X(t) = \{X(t_1), X(t_2), \ldots, X(t_n)\}$, l'hypothèse de stationnarité peut s'écrire :

$$\forall k \in \mathbb{N}, \forall t, m_k(t) = m_k \tag{1.50}$$

On parle de stationnarité stricte si tous les moments statistiques vérifient cette indépendance. On parle de stationnarité au sens large si seulement les deux premiers moments la vérifient ($k \leq 2$).

En supposant un signal aléatoire stationnaire, nous nous autorisons à l'étudier sur des intervalles de temps dont l'origine peut varier tout en préservant les propriétés statistiques.

1.2.5 Fonctions de densité de probabilité

Définition

La fonction de densité de probabilité (PDF) d'une variable aléatoire X, notée ρ_X est la probabilité associée à chaque valeur possible de la variable. Ainsi la probabilité d'obtenir une valeur comprise entre x et $x + dx$ est $\rho(x)dx$. Cette fonction est normalisée puisque la somme des probabilités vaut 1, ce qui se traduit par la relation :

$$\int_{-\infty}^{\infty} \rho_X(x)dx = 1 \tag{1.51}$$

Estimation

A partir d'un signal, il est possible d'évaluer cette fonction de densité de probabilité en divisant l'ensemble des valeurs possibles en intervalles constants de taille δx. Si la taille des valeurs possibles de X est ΔV, alors le nombre d'intervalles est $\frac{\Delta V}{\delta x}$. Pour chaque valeur de X, l'intervalle correspondant $N(i)$ est incrémenté. Ainsi, un histogramme est obtenu, de la forme de la densité de probabilité. Un exemple d'histogramme pour une variable aléatoire suivant une loi gaussienne est donnée sur la figure 1.7.

L'espérance, et donc les moments, peuvent être déduits de la fonction de densité de probabilité et de la relation définissant l'espérance :

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} x\rho_X(x)dx \tag{1.52}$$

En appliquant cette relation à un signal aléatoire discret de longueur W_s échantillons, l'espérance est donnée par :

$$\mathcal{E}(X) = \sum_{0}^{W_s} x[n]\rho_X(x[n]) \tag{1.53}$$

Dans la suite, nous présentons quelques distributions classiques que nous serons amenés à utiliser. Pour chacune, les propriétés et les algorithmes de génération sont précisés.

1.2.6 Somme et différence de fonctions de densité de probabilité

Nous considérons ici une variable aléatoire Z définie comme la somme de deux variables aléatoires indépendantes X et Y. Il est possible de relier la fonction de densité de probabilité ρ_Z de Z à partir des fonctions de densité de probabilité ρ_X et ρ_Y (respectivement de X et Y). Nous définissons $\rho_Z(x_0)$ la probabilité que la somme fasse x_0. Pour obtenir cette valeur x_0, si la variable aléatoire X prend la valeur x' (probabilité $\rho_X(x')$), la variable aléatoire Y doit donc prendre la valeur $(x_0 - x')$ (probabilité $\rho_Y(x_0 - x')$). Ainsi, en sommant sur toutes les valeurs de x', il vient :

$$\rho_{X+Y}(x_0) = \int_{-\infty}^{\infty} \rho_X(x')\rho_Y(x_0 - x')dx' \tag{1.54}$$

Cette relation est en fait la convolution des deux fonctions de densité de probabilité. Nous pouvons donc écrire :

$$\rho_{X+Y}(x_0) = \rho_X(x_0) * \rho_Y(x_0) \tag{1.55}$$

De la même manière, il est aussi possible de définir la fonction de densité de probabilité de la différence entre deux variables aléatoires indépendantes :

$$\rho_{X-Y}(x_0) = \int_{-\infty}^{\infty} \rho_X(x_0 + x')\rho_Y(x')dx' \tag{1.56}$$

1.2.7 Loi de distribution uniforme

Définition

La loi de distribution uniforme est définie par une probabilité constante sur l'intervalle des valeurs possibles, noté $[a; b]$. Toutes les valeurs ont exactement la même probabilité d'être choisie dans cet intervalle, et ont une probabilité nulle en dehors de celui-ci. Cette loi est décrite par l'équation :

$$\forall x \in [a; b], \quad \rho(x) = \frac{1}{b - a} \tag{1.57}$$

La figure 1.4 représente la fonction de densité de probabilité ρ.

Algorithme

Il existe plusieurs algorithmes pour définir des valeurs pseudo-aléatoires qui suivent une loi uniforme. Quelques exemples peuvent être trouvés dans [Moo90] ou [Orf96]. En programmation, les bibliothèques proposent souvent une implémentation d'une telle fonction. En `C`, la

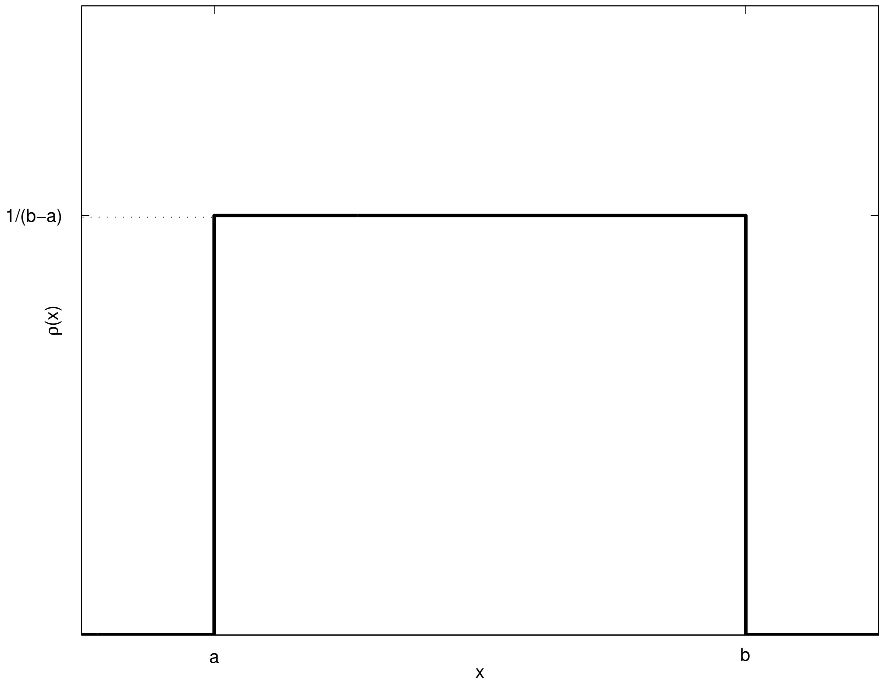

FIG. 1.4 – Représentation de la fonction de densité de probabilité relative à une loi de distribution uniforme.

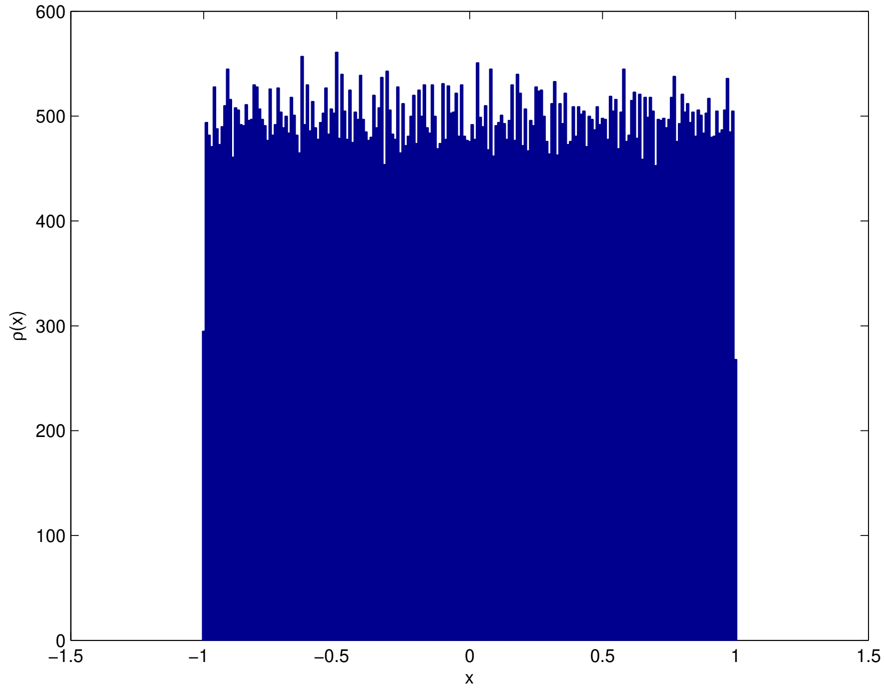

FIG. 1.5 – Fonction de densité de probabilité expérimentale obtenue avec la fonction `rand` de `Matlab` (tirage de 100000 valeurs).

fonction `rand` retourne un réel qui suit une telle loi. Il est calculé à partir d'une suite numérique définie pour que la répartition de ses valeurs successives suivent approximativement une distribution uniforme. La figure 1.5 montre un exemple de répartition des valeurs obtenues par l'utilisation de la fonction `rand` du logiciel de traitement du signal `Matlab`.

1.2.8 Loi de distribution de Gauss

Définition

La loi de Gauss (aussi appelée normale) est une fonction de densité de probabilité très importante à cause du théorème central limite (voir section 1.2.11). L'équation de cette fonction ρ est :

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(x-x_0)^2}{(2\sigma^2)}] \tag{1.58}$$

où $x \in \mathbb{R}$, x_0 représente la valeur moyenne et σ représente l'écart-type. La figure 1.6 représente cette fonction de densité de probabilité ρ.

Algorithme

La principale méthode de génération de valeurs aléatoires distribuées selon une loi de Gauss repose sur le théorème central limite. Les librairies existantes proposent des fonctions pseudo-aléatoires pour donner des valeurs uniformément distribuées. En additionnant un grand nombre de ces valeurs (notées x^u), le nombre total obtenu, noté X_i^g suit une loi de Gauss :

$$X^g = \sum_{j=1}^{N} x_j^u \tag{1.59}$$

La figure 1.7 montre un exemple de répartition des valeurs obtenues par l'utilisation de la fonction `randn` du logiciel de traitement du signal `Matlab`.

Propriétés

La fonction de densité de probabilité gaussienne est symétrique (paire). La probabilité maximale est la valeur moyenne, correspondant à x_0. Cette symétrie indique que le skewness est nul ($S_X = 0$). L'écart-type est donné par la valeur de σ. Le kurtosis vaut 3. On peut remarquer que dans certaines références la valeur de kurtosis d'une telle densité de probabilité sert de référence et vaut donc 0. Le kurtosis d'une distribution uniforme vaut alors -1.2 ($= 1.8 - 3$).

1.2.9 Loi de distribution de Rayleigh

Définition

La loi de Rayleigh est définie comme la fonction de densité de probabilité des valeurs x définies à partir de deux autres valeurs aléatoires a et b, distribuées selon une loi gaussienne :

$$x = \sqrt{a^2 + b^2} \tag{1.60}$$

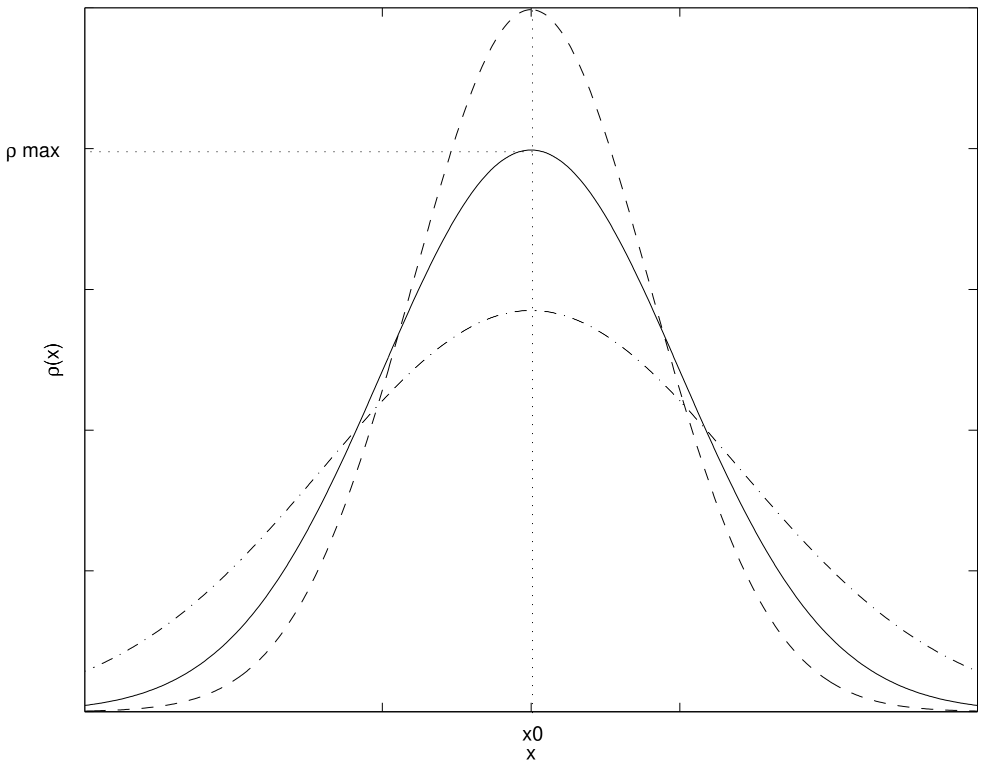

Fig. 1.6 – Représentation de la fonction de densité de probabilité relative à une loi de distribution gaussienne dans les cas $\sigma = 1$ (traits pleins), $\sigma = 0.8$ (traits hachurés) et $\sigma = 1.4$ (traits mixtes).

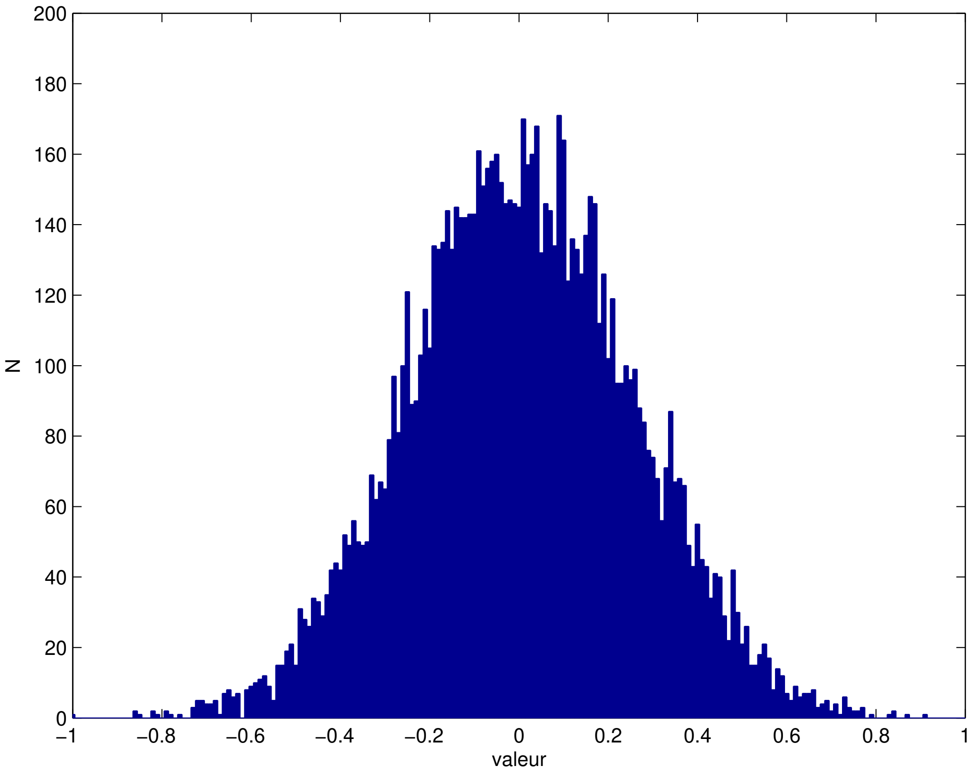

Fig. 1.7 – Estimation d'un fonction de densité de probabilité, ici un signal aléatoire dont les échantillons suivent une loi de Gauss (fonction `randn` de `Matlab`, tirage de 100000 valeurs).

Cette fonction de densité de probabilité ρ est définie par :

$$\rho(x) = \frac{x}{\sigma^2} \exp -(\frac{y^2}{2\sigma^2}) \tag{1.61}$$

où σ est l'écart-type de a et b. Cette fonction est normalisée :

$$\int_0^\infty \rho(x)dx = 1 \tag{1.62}$$

Algorithme

Un algorithme de génération de valeurs distribuées selon la loi de Rayleigh se base sur l'inverse de la fonction cumulative ($\int_0^x \rho(z)dz$) [Har97]. Les valeurs notées y_i peuvent être obtenues à partir de valeurs aléatoires uniformément distribuées x_i ($0 \leq x_i < 1$) :

$$y = \sigma\sqrt{-2\ln(1-x_i)} \tag{1.63}$$

Propriétés

La valeur moyenne de cette fonction de densité de probabilité ($\sigma\sqrt{\frac{\pi}{2}}$) n'est pas la valeur de probabilité maximale, qui correspond à $x = \sigma$: la courbe de représentation de la fonction de distribution n'est pas symétrique.

Les valeurs de la variance, du skewness et du kurtosis, dont les détails des calculs ne sont pas présentés ici, sont données dans le tableau 1.2.

Loi	moyenne	variance	skewness	kurtosis
Uniforme	0	$\frac{(b-a)^2}{12}$	0	1.8
Gauss	0	σ^2	0	3
Rayleigh	$\sigma\sqrt{\frac{\pi}{2}}$	$(2-\frac{\pi}{2})\sigma^2$	$(\pi-3)\sqrt{\frac{4\pi}{(4-\pi)^3}}$	$\frac{80-24\pi}{(4-\pi)^2}$

TAB. 1.2 – Valeurs des propriétés statistiques liées aux quatre premiers moments pour quelques distributions classiques.

1.2.10 Moyenne de signaux sinusoïdaux

Dans cette sous-section, nous nous intéressons à un signal sinusoïdal pur. Nous notons $x(t)$ ce signal. La moyenne de ce signal s'écrit :

$$\forall t, \quad \overline{x(t)} = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)dt \tag{1.64}$$

$$= \lim_{T\to\infty} \frac{1}{T}\int_0^T a\sin(\omega t+\phi)dt \tag{1.65}$$

Comme le signal est borné sur $[-a; a]$, la limite tend vers 0. Ainsi, la moyenne de ce signal est nulle.

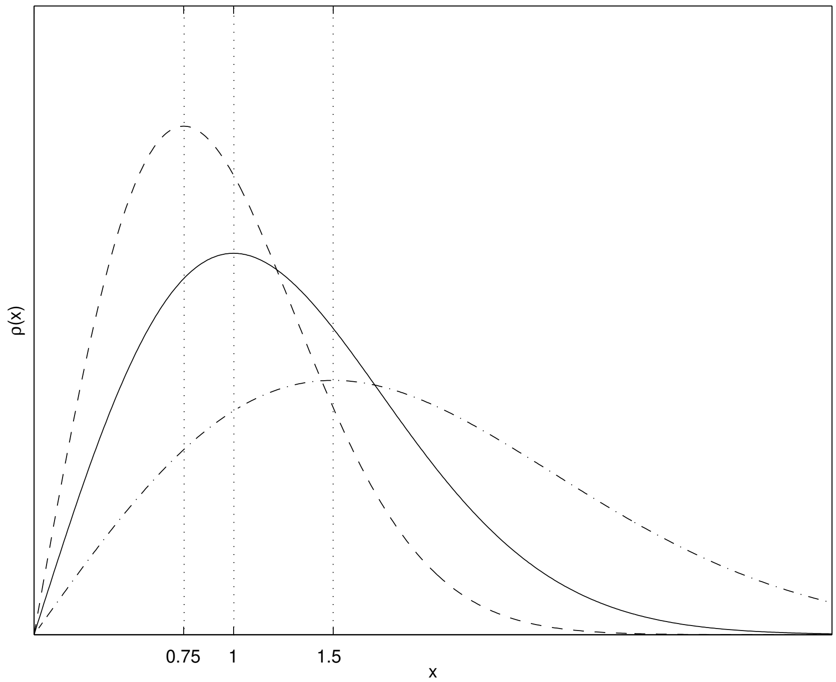

Fig. 1.8 – Représentation de la fonction de densité de probabilité relative à une loi de distribution de Rayleigh dans les cas $\sigma = 1$ (traits pleins), $\sigma = 0.75$ (traits hachurés) et $\sigma = 1.5$ (traits mixtes).

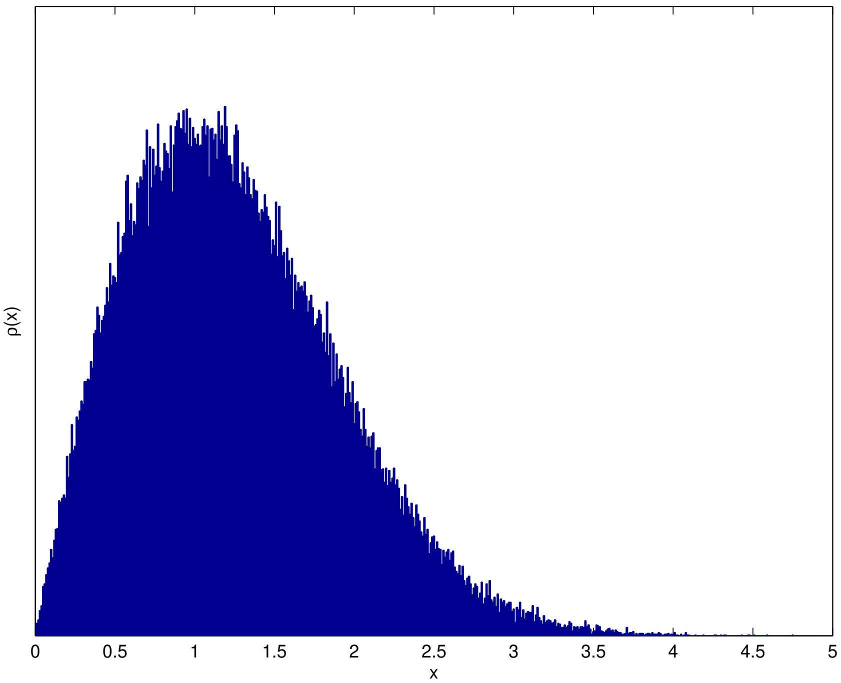

Fig. 1.9 – Fonction de densité de probabilité expérimentale obtenue en appliquant avec l'inverse de la fonction cumulative aux valeurs données par la fonction `rand` de `Matlab` (tirage de 100000 valeurs).

1.2.11 Théorème central limite

Le théorème central limite indique que la somme à l'infini de variables aléatoires indépendantes, de même loi, moyenne et variance (finie), est une variable aléatoire qui suit une loi de Gauss. Si X_n sont des variables aléatoires de variance σ, de moyenne x_0, et :

$$Y_n = \frac{\sum_{i=1}^{N} X_i - N x_0}{\sigma \sqrt{N}} \tag{1.66}$$

alors $\lim_{N \to \infty} Y_n$ suit une loi de Gauss.

Ce théorème peut être vérifié expérimentalement. La convergence est rapide, et le nombre N de variables nécessaire pour avoir une distribution quasiment gaussienne est très faible. Ce théorème est d'ailleurs à l'origine de l'algorithme de génération de valeurs de distribution gaussienne. De plus, il permet de valider l'hypothèse souvent émise sur le caractère gaussien des phénomènes aléatoires étudiés.

1.3 Éléments d'acoustique musicale

Le traitement du son a souvent un but artistique, et est donc généralement effectué par des musiciens ou des compositeurs. Ainsi, il est nécessaire d'avoir un lien très fort entre des paramètres mathématiques décrivant le son numérique et les termes musicaux utilisés par les artistes. Nous présentons dans cette section des notions communes faisant référence à la fois au traitement du signal et à la perception.

1.3.1 Fréquence

La hauteur est un attribut du son définissant un ordre sur les sons [Har97]. En musique, il existe deux dimensions liées à la fréquence : la note et l'octave. Une même note peut être jouée sur différentes octaves. C'est pourquoi certaines normes présentent les notes sous la forme d'une lettre et d'un chiffre. En considérant une sinusoïde pure de 440Hz, la note MIDI correspondante La3 (A3) indique non seulement que la note est un La, mais aussi qu'il s'agit de la troisième octave. Une représentation de cette approche a été proposée par Shepard sous forme d'un cylindre et d'une hélice, qui peut être aisément trouvée dans la littérature (par exemple [Har97, Mar00]). Il existe donc un lien direct entre la fréquence f fondamentale du son et la note de musique p (*pitch*) :

$$f = f_0 2^{\frac{p}{p_0}} \tag{1.67}$$

$$p = p_0 \log_2 \left(\frac{f}{f_0}\right) \tag{1.68}$$

où p_0 et f_0 sont respectivement les hauteurs et les fréquences de référence.

A l'échelle macroscopique, les variations de fréquences constituent la mélodie de la musique. Des variations périodiques à l'échelle microscopique provoquent un vibrato. La période de ces variations est dans ce cas-là faible (moins de 20Hz) pour pouvoir être perçue comme une variation de hauteur et pour ne pas rajouter des composants spectraux. Une variation de hauteur linéaire est associée à un glissando ou portamento.

La hauteur est un paramètre musical lié à la perception, mais pas un paramètre physique. Elle est fortement liée à la fréquence fondamentale (même si elle est *virtuelle*) dans le cas d'un son harmonique ou quasi-harmonique, mais cette notion de hauteur devient très complexe dans le cas de signaux tels que les bruits.

1.3.2 Amplitude

Le volume d'un son est lié à l'amplitude A du signal s. Cette relation est détaillée dans la section 1.5. L'amplitude en décibels du signal s'écrit :

$$A_{dB}(s) = 20 \log \left(\frac{A(s)}{A_0}\right) \tag{1.69}$$

où log représente le logarithme à base 10 et A_0 l'amplitude de référence (0dB).

La dynamique d'un morceau de musique est déterminée par les variations d'amplitude d'un son ou d'un ensemble de sons constituant ce morceau. Les variations peuvent être linéairement croissantes ou décroissantes (*fade-in, fade-out*). Elles peuvent aussi périodiques et le musicien utilise le terme tremolo. La période de ces variations est dans ce cas-là faible (moins de 20Hz) pour pouvoir être perçue comme une variation d'amplitude, et non comme un composant spectral supplémentaire.

1.3.3 Timbre

Le timbre est un terme général musical, étudié depuis des années [Sch66], et utilisé souvent pour définir un ensemble de propriétés du son. Par exemple, nous sommes capables de différencier une guitare et un piano qui jouent une même note avec le même volume parce qu'ils ont un timbre différent. Le timbre englobe donc par cette approche tout ce qui nous permet de distinguer deux sons de même hauteur et de même volume. Une autre définition en psychoacoustique présente le timbre comme les mécanismes perceptuels classifiant les sons en familles [Roa98].

En fait il apparaît que le timbre est intimement lié au spectre du son, mais que le timbre et le spectre ne sont pas équivalents. Il regroupe de nombreux facteurs tels que l'enveloppe temporelle, les vibrato et tremolo, la brillance, ..., et donc l'enveloppe spectrale et ses variations dans le temps.

Du point de vue de la perception, des expériences psychoacoustiques montrent que le timbre est dissocié de la hauteur dans notre mémoire auditive à court terme [SD91].

1.3.4 Partiels

Le terme partiel est fortement lié à la synthèse additive (voir section 2.3). Un son harmonique est en général composé d'une fréquence fondamentale (notée F_0) et d'harmoniques (notées F_k), multiples de cette fréquence.

$$\forall k, \quad F_k = kF_0 \tag{1.70}$$

Des exemples de sons harmoniques sont les voix chantées ou parlées, instruments à vent (trompette, clarinette,...) ou à corde (violon, guitare, ...). Un son peut aussi être composé d'une fréquence fondamentale et d'harmoniques qui ne sont pas forcément des multiples de la fondamentale. Les écarts peuvent être très faibles. On parle dans ce cas de sons quasi-harmoniques. Des exemples de sons quasi-harmoniques sont des sons issus d'instruments à corde ou certains instruments percussifs (xylophone, marimba, ...). Ces écarts de fréquences peuvent aussi être plus importants comme dans le cas de certains autres instruments percussifs (cloche, ...). On parle dans ce cas de sons inharmoniques.

Un partiel est défini comme une sinusoïde dont la fréquence et l'amplitude varient lentement par rapport au temps. Dans le cas de sons harmoniques et quasi-harmoniques, la fréquence fondamentale et ses harmoniques sont des partiels. De même pour les sons inharmoniques, les composants fréquentiels du spectre sont considérés commes des partiels. La figure 1.10 montre un spectre d'un son harmonique (la fondamentale avec quatre harmoniques). Ainsi, le son dont le spectre est représenté par cette figure est à priori composé de cinq partiels.

Un partiel correspond souvent à un pic (maximum local) dans le spectre. Pourtant cette propriété n'est pas toujours avérée. Ces problèmes d'analyse sont abordés dans le chapitre 3.

La définition du partiel est fortement liée à l'adverbe *lentement*. Il est possible de préciser la vitesse des variations. En effet, un son x composé d'un partiel dont l'amplitude est modulée par une sinusoïde, est équivalent à deux sinusoïdes :

$$\begin{aligned}
x(t) &= \sin(2\pi ft)\sin(2\pi Ft+\phi) \\
&= \frac{1}{2}\cos(2\pi(F+f)t+\phi) - \frac{1}{2}\cos(2\pi(F-f)t+\phi) \\
&= \frac{1}{2}\sin(2\pi(F+f)t+\phi+\frac{\pi}{2}) + \frac{1}{2}\sin(2\pi(F-f)t+\phi-\frac{\pi}{2}) \qquad (1.71)
\end{aligned}$$

Ainsi, si f est très faible (au moins inférieur à 20Hz), le son est perçu comme une sinusoïde dont l'amplitude varie, et peut donc être assimilée à un partiel. Par contre, si f est important, le son est perçu comme la somme de deux sinusoïdes et est assimilé à deux partiels.

1.3.5 Transitoires

Les transitoires sont assimilées à tous les phénomènes de brusques variations temporelles d'amplitude dans un signal. Elles se rapportent notamment aux attaques qui désignent la très courte durée durant laquelle le son monte en amplitude avant d'atteindre sa valeur maximale. Cette période correspond à une hausse rapide d'énergie dans tout le spectre, notamment dans les fréquences aiguës. La figure 1.11 illustre cette phase d'attaque en montrant une représentation théorique de l'enveloppe temporelle dynamique d'une note de musique, divisée en quatre parties : l'attaque, le déclin, le soutien et le relâchement (ADSR pour *attack, decay, sustain, release*). L'attaque est donc la première partie. Le déclin est la durée du son du haut de l'attaque jusqu'au point où le son commence à garder une amplitude constante et soutenue. Le soutien est la partie durant laquelle le son est conservé à un niveau constant jusqu'à son relâchement. Enfin, le relâchement est égal à la durée de la baisse d'intensité jusqu'au silence.

Des transitoires sont également présentes dans des sons complexes, notamment des sons bruiteux, sous forme de fortes variations d'énergie très localisées. Par exemple, le bruit des impacts de gouttes de pluie peut être assimilé à une succession de transitoires. Un autre exemple est donné par la figure 1.12 qui montre la représentation du son issu d'un bruit de machine et d'un impact bref.

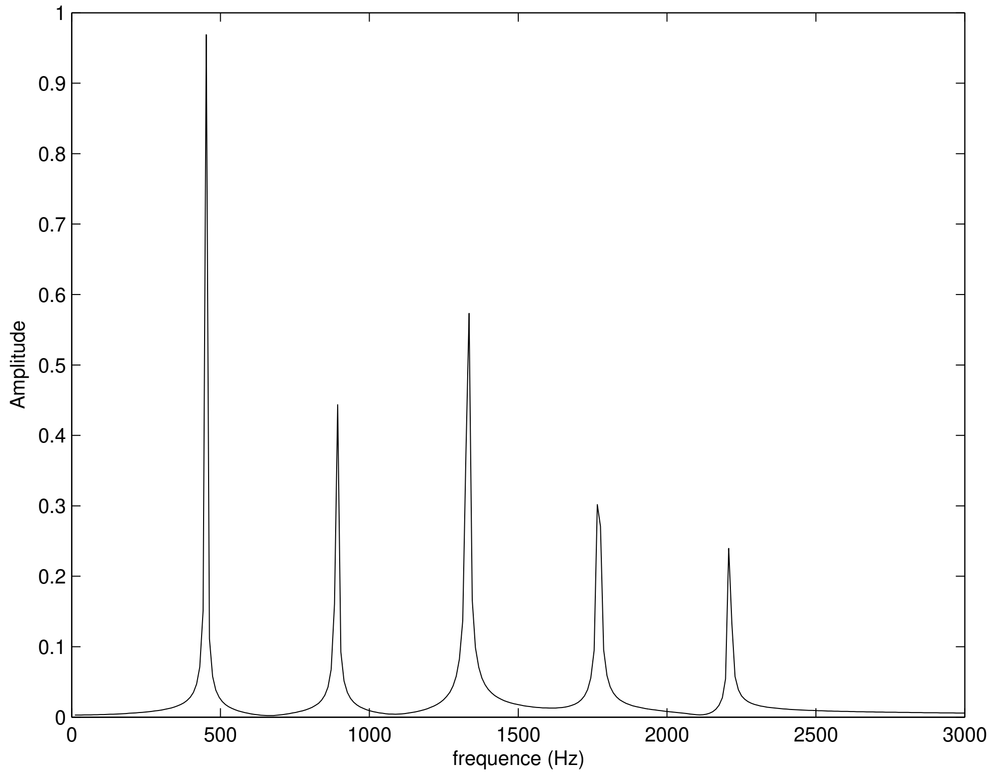

Fig. 1.10 – Exemple de spectre d'un son harmonique de fréquence fondamentale $F_0 = 440$Hz et de ses 4 premières harmoniques.

Les attaques sont perçues comme des *clics* dans le son et sont d'une grande importance dans le réalisme de sons instrumentaux notamment. Elles jouent également un rôle prépondérant dans la perception du timbre.

1.3.6 Bruits

Les bruits sont difficiles à définir. Dans la section suivante 1.4, nous détaillons les différentes définitions possibles. Mais il est important ici de noter que ce sont des sons dont le spectre est dense. Ainsi, le terme partiel n'est pas adapté à de tels sons, puisque les variations d'énergie dans le spectre sont rapides et nombreuses. Nous parlons plutôt de composants fréquentiels.

La notion de bruit est utilisée (parfois abusivement) pour décrire de nombreux sons provenant notamment d'instruments de musique. Ainsi, cette notion est associée aux sons denses comme les cymbales, aux turbulences d'air provoquées par le souffle pour les instruments à vent ou pour la parole, aux multiples chocs induits par les maracas, ...

Pour les modèles d'analyse et de synthèse de sons, le bruit est une partie prépondérante du point de vue du réalisme, en particulier en ce qui concerne les instruments de musique. Un instrument de synthèse est entendu comme *artificiel* si un bruit adapté n'est pas ajouté au son synthétisé. Par exemple, des synthétiseurs reposant sur la synthèse par modulation de fréquence (FM) sont très limités dès qu'il s'agit de sons d'instruments percussifs bruités.

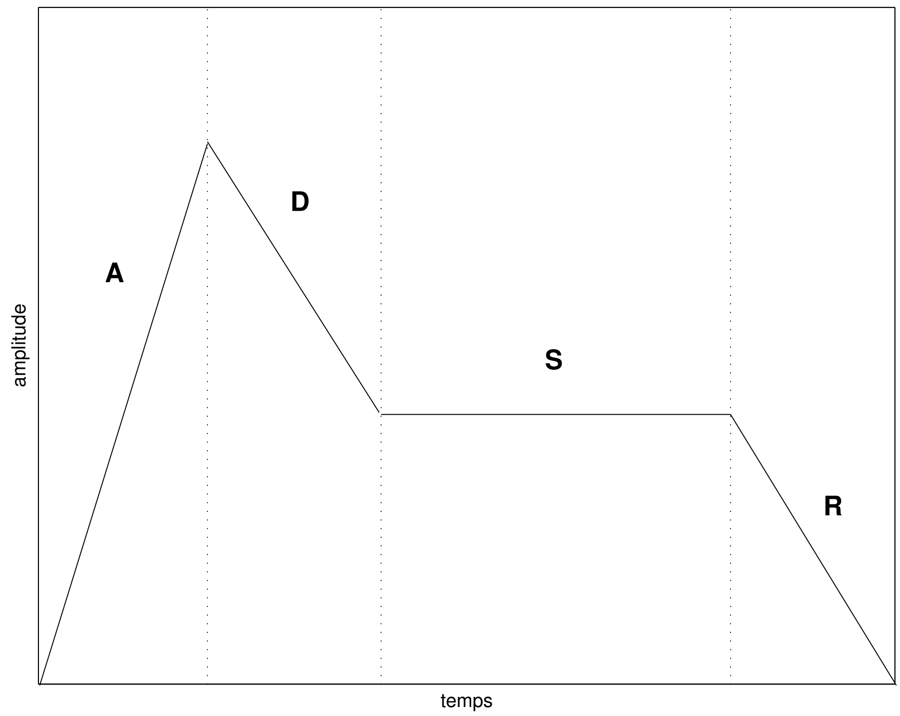

FIG. 1.11 – Illustration de l'enveloppe temporelle d'une note de musique. Quatre intervalles (ADSR) sont définis : l'attaque (*attack*), le déclin (*decay*), le soutien (*sustain*) et le relâchement (*release*).

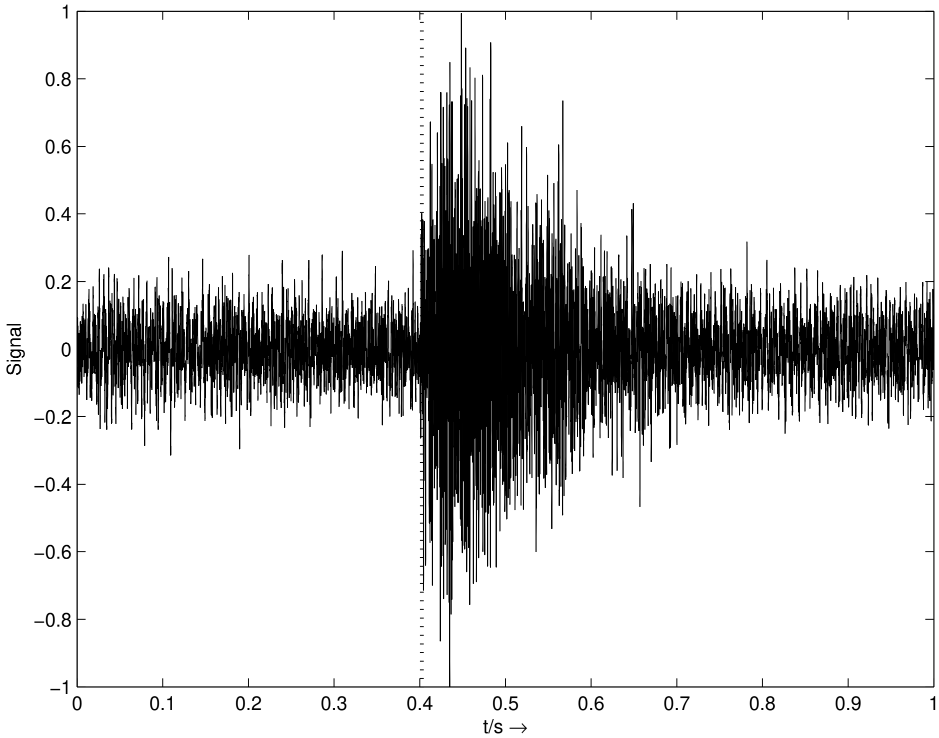

FIG. 1.12 – Exemple d'un bruit naturel issu d'une machine contenant une transitoire, dont l'origine est matérialisée par la ligne pointillée (à peu près 0.4s).

1.4 Définitions des bruits

Comme indiqué ci-dessus, il est très délicat de donner une définition précise des bruits. Il n'existe même aucune véritable définition indiscutable. Dans le langage courant, un bruit désigne un son non voulu et désagréable qui vient perturber l'écoute. Leur origine est souvent incomprise. Cette définition ne peut être satisfaisante car subjective, puisque certains types de musique agréables pour certaines personnes sont insupportables pour d'autres, ou un même morceau joué à un très fort volume est défini comme du bruit par certains. Nous présentons ici plusieurs autres approches possibles dépendantes du domaine abordé. Ensuite quelques types de bruits classiques sont définis et leurs propriétés sont expliquées.

1.4.1 Bruits en électronique

La notion de bruit a été introduite en électronique au début du 20ème siècle par Johnson et Nyquist. Le bruit est expliqué par les mouvements désordonnés (aléatoires) des porteurs de charges (électrons) d'un matériau sous l'influence de la température (non nulle sur l'échelle Kelvin). Ce bruit peut par exemple être facilement visualisé sur un écran de télévision dont l'antenne est débranché : on parle de "neige". Ce type de bruit est appelé bruit de Johnson ou de Nyquist, ou encore bruit thermique (*thermal noise*).

Le terme bruit est également utilisé pour qualifier les erreurs introduites lors d'un calcul ou lors d'une expérience de mesure.

1.4.2 Bruits en acoustique

Le bruit est présenté en acoustique comme un signal complexe dont le spectre est dense, et qui ne peut être défini qu'en faisant appel aux outils statistiques. Cette approche semble plus rigoureuse scientifiquement par rapport au langage commun, car un signal tel qu'une sinusoïde de haute fréquence (2000Hz) est désagréable et peut perturber, sans toutefois pouvoir être assimilé à du bruit. De plus, le besoin d'avoir recours à des outils statistiques semble d'autant plus juste, qu'elle s'appuie sur la perception. En effet, il est possible de générer plusieurs sons dont les représentations temporelles et fréquentielles sont complètement différentes au sens déterministe, qui ne peuvent pas être différenciés à l'écoute. La figure 1.13 propose deux représentations fréquentielles de deux sons complexes d'apparence différente, mais qui pourtant ont les mêmes qualités perceptives. Ainsi, pour de tels sons, la perception semble essentiellement tenir compte des propriétés statistiques (valeurs moyennes).

1.4.3 Bruit blanc

Définitions

Le bruit blanc doit son appelation à l'analogie avec la lumière blanche. Cette lumière blanche est dûe à la présence de photons de toutes les valeurs d'énergie, et est donc composée de toutes les couleurs. De même, un bruit blanc possède un spectre de fréquence continu et d'amplitude moyenne constante indépendante des fréquences.

$$\forall f, \quad \langle X(f) \rangle = X_0 \tag{1.72}$$

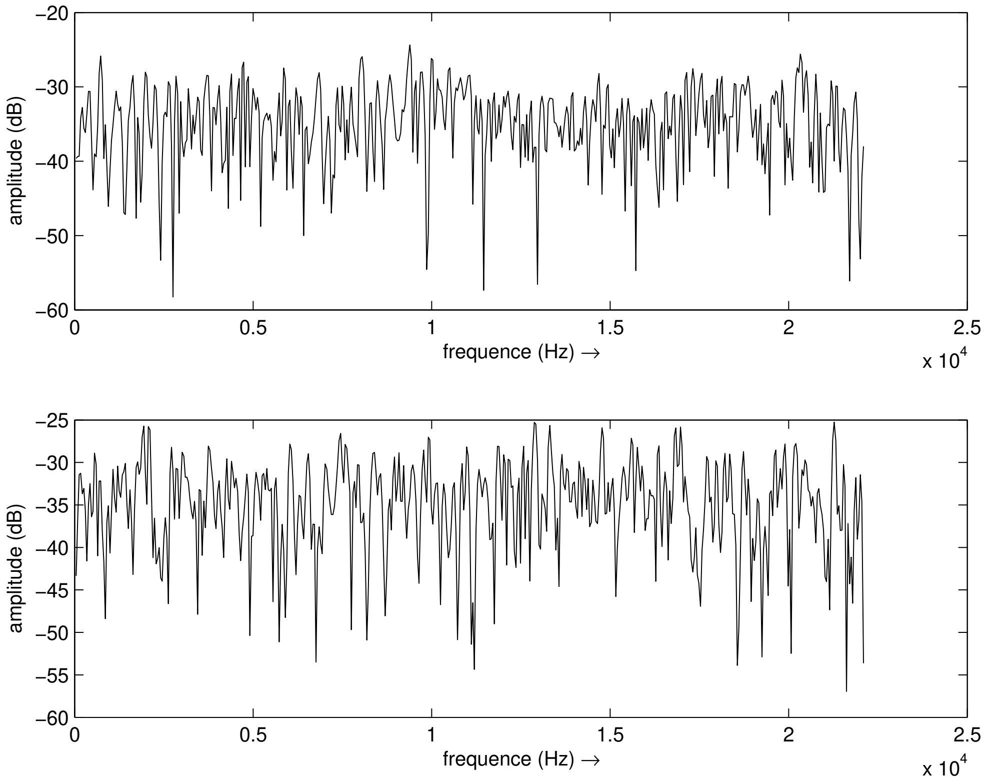

FIG. 1.13 – Deux spectres à court terme d'apparence différente de deux bruits qui sont pourtant impossibles à discerner sur le plan de la perception.

Le bruit blanc reste un modèle théorique et un tel bruit ne peut pas exister naturellement car il aurait sinon une puissance infinie :

$$P = \int_0^\infty X^2(f)df \tag{1.73}$$

A cause des limites de la perception, la définition du bruit blanc est restreinte à la bande de fréquence comprise entre 20Hz et 20000Hz. La figure 1.14 montre une représentation théorique du spectre d'un bruit blanc. En réalité, ce spectre reste uniquement une vision théorique puisqu'il représente un spectre moyen. Ainsi, l'analyse spectrale (quelle que soit la méthode utilisée) résulte en une courbe beaucoup plus chaotique que la courbe théorique. La figure 1.15 présente un exemple de spectre d'une portion de bruit blanc de synthèse.

Du point de vue de la représentation temporelle du bruit blanc, les échantillons successifs n'ont aucun lien entre eux. La figure 1.15 montre un exemple de courbe d'amplitude d'un court intervalle de temps pour un bruit blanc de synthèse.

De nombreux sons naturels se rapprochent de ce modèle de bruit blanc, comme par exemple des écoulements d'eau, d'air, etc...

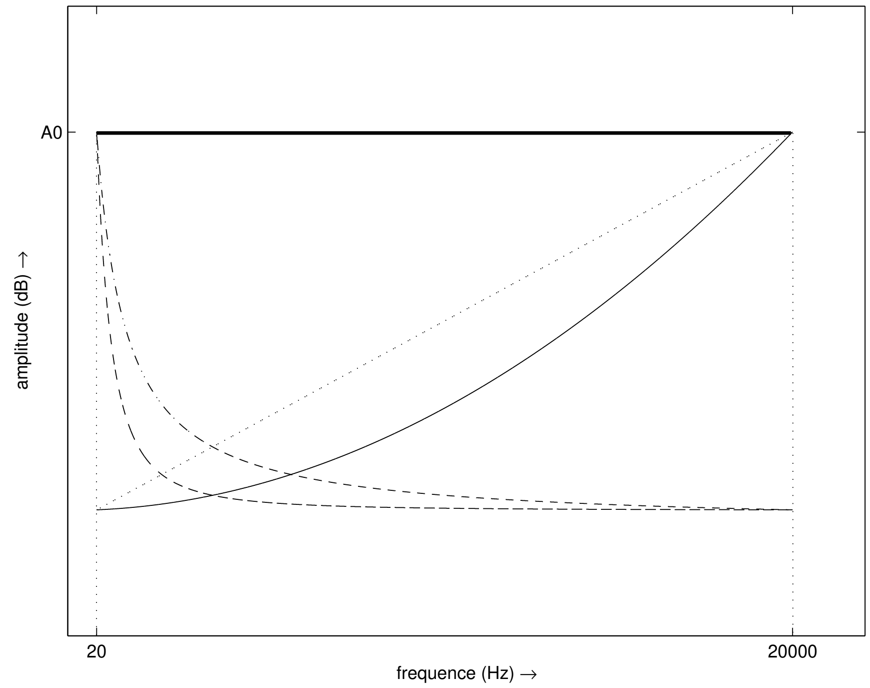

FIG. 1.14 – Illustration des spectres théoriques des bruits classiques : le bruit blanc (gras), le bruit rose (mixtes), le bruit brownien (tirets), le bruit bleu (pointillés) et le bruit violet (plein). La largeur de bande est fixée par les limites de la perception ($20 - 20000$Hz).

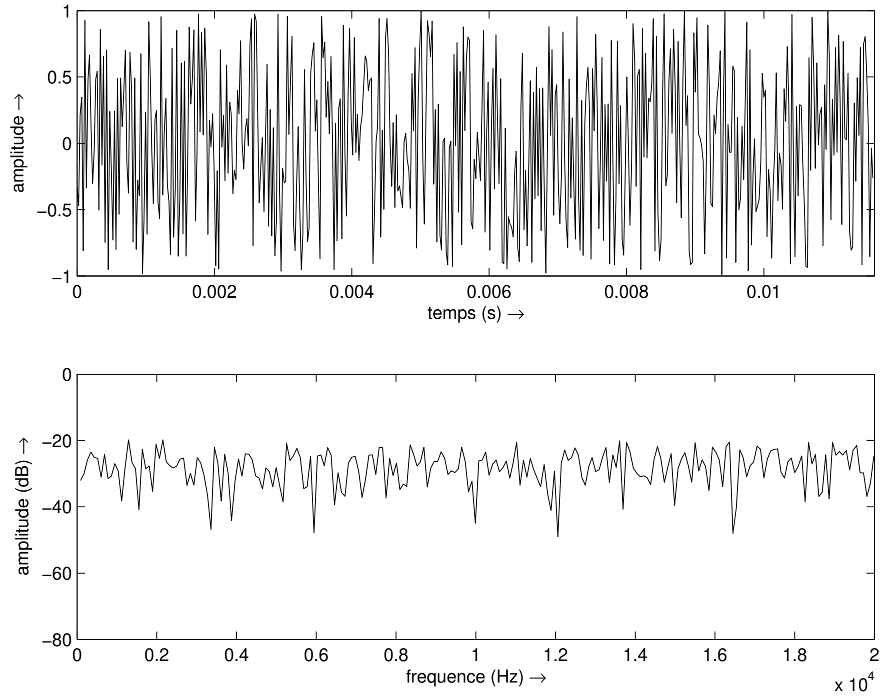

FIG. 1.15 – Courbe d'amplitude (en haut) et spectre obtenu par transformée de Fourier (en bas) d'un bruit blanc de durée 11.6ms. Le spectre apparaît perturbé par rapport au spectre théorique représenté par la figure 1.14.

Propriétés

Les propriétés auditives essentielles des bruits blancs sont l'absence de perception de hauteur, dûe à la supposée constance du spectre par rapport aux fréquences, et l'absence de perception de rythme, dûe à la supposée absence de périodicités de l'enveloppe temporelle. C'est notamment à cause de ces propriétés que ces sons sont fréquemment utilisés lors d'expériences psychoacoustiques pour mettre au point des modèles perceptifs.

La définition du bruit blanc se traduit par des propriétés statistiques. Ce signal aléatoire, noté $B[n]$, est représenté par des échantillons, de moyenne nulle, indépendants (donc décorrélés) entre eux, mais qui suivent la même loi de distribution. Ainsi, lorsque le bruit est supposé stationnaire, nous pouvons écrire :

$$\forall n, \quad \mathcal{E}(B[n]) = 0 \tag{1.74}$$

$$\forall n, \forall \tau \neq 0, \quad \Gamma(\tau) = \mathcal{E}(B[n]B[n+\tau]) = 0 \tag{1.75}$$

1.4.4 Bruits colorés

Les bruits doivent aussi leur dénomination à l'analogie avec la lumière. Ce bruits n'ont pas un spectre d'amplitude constant. Au sens strict du terme, le terme bruit coloré est affecté à un bruit dont le spectre est caractérisé par une enveloppe très lisse avec une bande large d'amplitude plus importante. Ainsi, la couleur correspondante au spectre de la lumière associée est affectée au bruit. Par exemple, les bruits roses, bleus, violets, gris, ... ont ainsi été définis.

Bruit rose

Les bruits peuvent être considérés comme rose (*pink noise*) si leur distribution de puissance moyenne (ou énergie sur une longue période de temps) suit une échelle logarithmique, ce qui revient à ce que leur spectre d'amplitude soit proportionnel à l'inverse de la fréquence. Une représentation du spectre théorique d'un bruit rose est montrée par la figure 1.14. Le spectre d'un bruit peut donc se mettre sous la forme :

$$\forall f \geq f_{\min}, \quad S(f) = \frac{S_0}{f} \tag{1.76}$$

Ainsi, la puissance du bruit est aussi importante dans l'octave $[200; 400]$Hz que dans l'octave $[2000 - 4000]$Hz. Autrement dit, le spectre possède la même quantité d'énergie sur des intervalles de fréquence de largeur égale :

$$\forall f \geq f_{\min}, \quad \int_f^{2f} S(f)df = S_0 \ln \frac{2f}{f} = S_0 \ln(2) \tag{1.77}$$

De par sa définition, le bruit rose est souvent référencé comme le bruit en $\frac{1}{f}$. L'amplitude d'une fréquence f est divisée par 2 par rapport à l'amplitude correspondant au double de la fréquence $S(2f) = \frac{1}{2}S(f)$. Le spectre est donc aussi caractérisé par une baisse de -3dB par octave (à peu près -10dB par décade) :

$$\forall f \geq f_{\min}, \quad 10 \log \frac{S(f)}{S(2f)} = -10 \log 2 = -3 \tag{1.78}$$

Sur le plan de la perception, les hautes fréquences sont atténuées. Ce bruit apparaît être finalement un bruit blanc plus justifié, car le système auditif est plus sensible à l'échelle logarithmique qu'à l'échelle linéaire. Les applications de ce bruit sont nombreuses et dépassent largement le cadre du traitement du son : biologie, astronomie, géologie, ... Concernant l'analyse de musique, plusieurs études ont montré l'utilité de ce bruit pour la composition aléatoire automatique [BG97, Moo90].

Bruit brownien

Avant tout, il est important de noter que le bruit brownien (*brownian noise*) n'a aucun rapport avec le bruit marron (*brown*), comme pourrait le laisser penser le nom en anglais. Son nom vient des mouvements dits browniens des molécules dans un fluide. On parle aussi de marche aléatoire (*random walk*). Les bruits dits browniens sont les bruits dont les spectres d'amplitude sont proportionnels à l'inverse de la fréquence au carré. On parle aussi de bruit en $\frac{1}{f^2}$. Une représentation du spectre théorique d'un bruit brownien est montrée par la figure 1.14. Le spectre peut donc s'écrire sous la forme :

$$\forall f \geq f_{\min}, \quad S(f) = \frac{S_0}{f^2} \tag{1.79}$$

L'amplitude d'une fréquence f est divisée par 4 par rapport à l'amplitude correspondant au double de la fréquence $S(2f) = \frac{1}{4}S(f)$. Le spectre est donc aussi caractérisé par une baisse de 6dB par octave.

Bruit bleu et violet

Les bruits bleu et violet sont respectivement définis par un spectre d'amplitude proportionnel à la fréquence et au carré de la fréquence. Une représentation du spectre théorique d'un bruit bleu et d'un bruit violet est montrée par la figure 1.14. Leur spectre peut donc se mettre sous la forme :

$$\forall f, \quad S(f) = S_0 f^\alpha \tag{1.80}$$

où $\alpha = 1$ dans le cas du bruit bleu, et $\alpha = 2$ dans le cas du bruit violet.

Leur spectre est donc caractérisé par une hausse respective de +3 et +6 dB par octave. Ces bruits ont leurs basses fréquences atténuées.

Bruit blanc filtré

La défintion au sens large du bruit coloré est rapporté à la définition du bruit blanc filtré. En effet si l'appelation bruit coloré n'implique aucune hypothèse sur la forme de l'enveloppe spectrale, alors tous les bruits blanc filtrés peuvent être appelés bruit colorés. Ainsi, par cette définition, un bruit coloré peut donner, au contraire des bruits précédemment présentés, une impression de hauteur (voire plusieurs hauteurs) si le filtre définit une (ou plusieurs) bande spectrale de forte amplitude (voir la section 1.5.4).

1.4.5 Bruits ondulants

Les bruits ondulants (*rippled noise*) sont particulièrement utilisés lors d'expériences sur la modélisation et la compréhension de la perception de la hauteur (voir la section 1.5.4) car ils donnent une sensation assez précise de hauteur. Ce type de bruits n'est pas naturel (à notre connaissance) et résulte de l'addition à un bruit blanc de lui-même retardé. Ainsi, l'enveloppe spectrale de ces bruits est constituée de pics périodiques : on parle de pics spectraux.

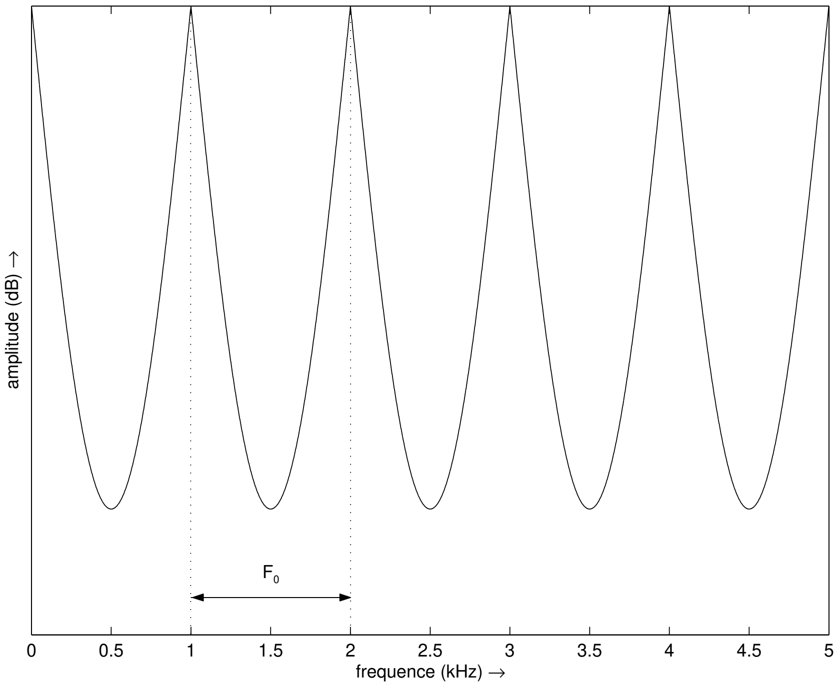

FIG. 1.16 – Représentation spectrale théorique d'un bruit ondulant (*rippled noise*) : la fréquence F_0 correspond à la hauteur perçue.

1.4.6 Low-noise Noise

Le *low-noise noise* définit le bruit dont les fluctuations d'intensité sont les plus faibles possibles. Il est principalement utilisé pour les communications en permettant de limiter les distortions et d'avoir un meilleur rapport signal-bruit [HP91], mais aussi lors d'expériences psychoacoustiques sur les propriétés de masquage des bruits (voir section 1.5.6), mais aussi pour étudier les réponses spectrales d'un système [Pum85]. En fait, ce bruit reste théorique et sont qualifiés de *low-noise noise* les bruits dont les fluctuations d'intensité sont bien plus faibles que les fluctuations d'un bruit blanc. Plusieurs algorithmes de synthèse de tels bruits ont été proposés et sont présentés dans le chapitre 4.

1.5 Éléments de psychoacoustique

Dans cette section nous introduisons quelques concepts et résultats liés à la perception. Le mode de fonctionnement et les propriétés de l'oreille humaine étant particulièrement complexes, les modèles de perception présentés (pour l'intensité, la hauteur et la densité spectrale notamment) sont simplifiés, comme dans quasiment tous les travaux de traitement du signal, sujets par ailleurs à de nombreuses approximations mathématiques. En effet, pour de tels travaux, la complexité qu'imposeraient des modèles de la perception plus rigoureux, ne serait certainement pas compensée par le gain en qualité obtenu. Quelques ouvrages de référence ([ZF99], [Har97]) proposent des résultats d'expériences donnant lieu à des modèles plus complexes.

1.5.1 Perception de l'intensité

Au sens strict du terme, l'intensité d'un son est un paramètre physique, lié à l'amplitude de l'onde sonore, c'est-à-dire à l'amplitude du signal. Il existe une forte corrélation avec l'intensité perçue, paramètre perceptif appelé sonie (*loudness*), et cette amplitude. En effet, l'augmentation de l'amplitude d'un signal implique dans la très grande majorité des cas une augmentation de la sonie.

Cette différence est expliquée par le système auditif qui tient compte de nombreux facteurs comme les informations temporelles (enveloppe, durée, ...) et fréquentielles (spectre, répartition de l'énergie, ...). Par exemple, l'intensité perçue d'une sinusoïde pure dépend de sa fréquence car l'oreille est plus sensible sur certaines régions du spectre. La figure 1.17 montrent les courbes de même sonie en fonction de la fréquence (courbes de Fletcher-Munson). Nous pouvons remarquer que le maximum de sensibilité se situe à peu près sur l'intervalle $2700 - 3200$Hz (le seuil de la douleur est lui évalué vers 120dB). Une sinusoïde de fréquence 3000Hz est perçue plus fortement qu'une sinusoïde de même amplitude mais de fréquence 5000Hz ou 500Hz.

Pour les sons complexes, chaque composant fréquentiel contribue à la sonie selon ces mêmes courbes, mais de nombreux effets supplémentaires interviennent (masquage, bandes critiques, ...), qui compliquent fortement le phénomène.

Il est également établi que l'oreille perçoit l'intensité sur une échelle logarithmique. C'est pour cela que l'unité décibels (dB) est généralement utilisée. Sachant que la pression est définie comme le carré de l'intensité, nous pouvons écrire le niveau sonore L du signal x sous la forme :

$$L(x) = 20log(\frac{P(x)}{P_0}) \tag{1.81}$$

où P_0 est la pression acoustique de référence (10^{-5} Pa), correspondant au seuil de l'audition à 1000Hz. Sachant que l'amplitude du signal est proportionnelle à la pression acoustique, nous définissons le niveau sonore L par l'équation :

$$L(x) = 20log(\frac{A(x)}{A_0}) \tag{1.82}$$

Le volume V d'un son est une mesure subjective, qui est fortement corrélée au niveau sonore L. L'hypothèse couramment admise suppose que ces deux valeurs sont égales.

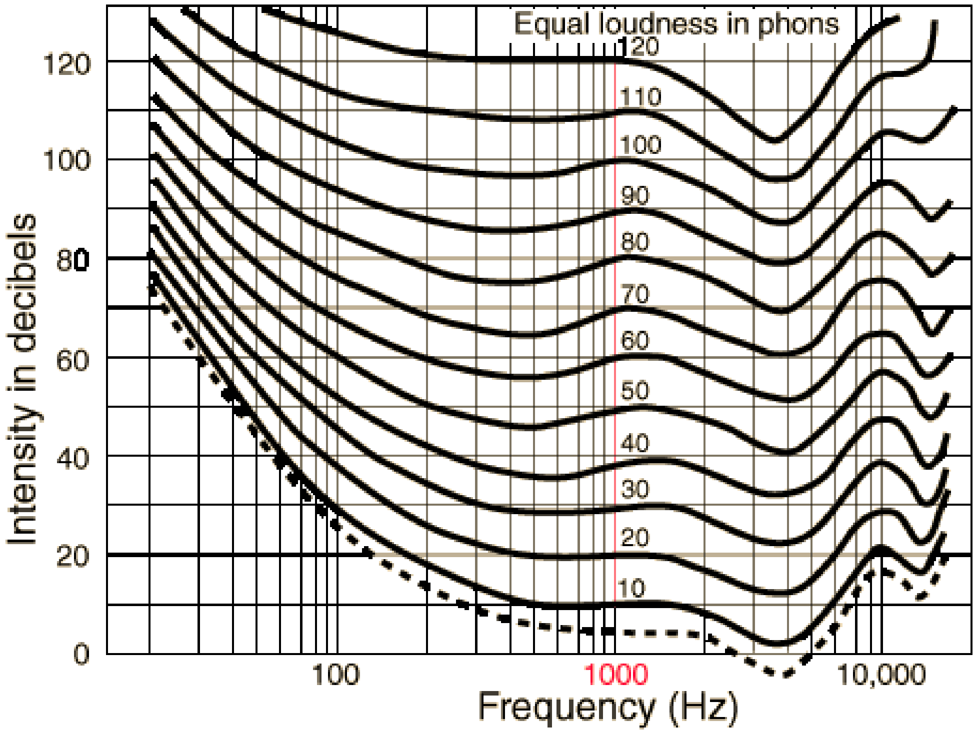

FIG. 1.17 – Courbes d'intensité perçue (sonie) constante (Fletcher et Munson 1933) : pour chaque courbe, le niveau en phones (unité correspondant à l'intensité perçue pour une sinusoïde de 1000Hz) est indiqué.

Amplitude RMS

L'amplitude peut à priori être définie à partir de la représentation temporelle en choisissant l'amplitude de valeur absolue maximale. Cette approche devient très discutable dans le cas de sons complexes, et en particulier dans le cas de signaux aléatoires. En effet, un seul pic d'amplitude ne peut définir l'amplitude du signal complet. L'amplitude RMS (*Root Mean Square*) est une mesure de l'amplitude d'un signal x, dans les deux cas continus et discrets (W_s échantillons) :

$$A_{\mathrm{RMS}}(x) = \lim_{T\to\infty} \sqrt{\frac{1}{T}\int_0^T x^2(t)dt} \tag{1.83}$$

$$A_{\mathrm{RMS}}(x) = \sqrt{\frac{1}{W_s}\sum_{n=1}^{W_s} x^2[n]} \tag{1.84}$$

Cette mesure d'amplitude peut aussi bien s'appliquer aux sinusoïdes pures qu'aux signaux aléatoires.

Intensité perçue d'un partiel

A partir des équations (1.84) et (1.82), nous pouvons exprimer le volume associé à une sinusoïde pure d'amplitude a_0 :

$$\begin{aligned} A_{\mathrm{RMS}}(x) &= \lim_{T\to\infty}\sqrt{\frac{1}{T}\int_0^T a_0\sin^2(\omega\frac{n}{F_s}+\phi)dt} \\ &= \frac{a_0}{\sqrt{2}} \end{aligned} \tag{1.85}$$

Intensité perçue d'un bruit

Dans le cas d'un signal aléatoire X, l'amplitude RMS peut se mettre sous la forme :

$$A_{\mathrm{RMS}}(X) = \sqrt{\langle X^2\rangle} \tag{1.86}$$

Dans le cas d'un signal de moyenne nulle, l'amplitude RMS est donc équivalente à la racine carrée de la variance, c'est-à-dire l'écart-type.

$$A_{\mathrm{RMS}}(X) = \sigma \tag{1.87}$$

1.5.2 Perception de la hauteur

La hauteur est le paramètre perceptif (et musical comme vu dans la section 1.3.1) relatif au paramètre physique qu'est la fréquence. Comme pour le volume, le fonctionnement de l'audition pour la perception de la hauteur des sons est complexe. En effet, cette perception est basée à la fois sur les déplacements d'une partie spécifique de l'oreille (la membrane basilaire), mais aussi des mécanismes temporels de détection de périodicités.

Un son est bien entendu composé de nombeuses fréquences, parfois liées entre elles (sons harmoniques ou quasi-harmoniques), parfois sans aucun lien (sons inharmoniques). Dans le premier cas, le système auditif permet d'associer une seule hauteur à chaque son, qui correspond en général à la fréquence fondamentale. Néammoins, il peut arriver que la hauteur perçue ne corresponde à aucune fréquence présente. Par exemple, un son composé d'un certain nombre d'harmoniques (par exemple 880Hz, 1320Hz, 1760Hz, ...) laisse percevoir une hauteur équivalente à la fréquence fondamentale (440Hz) : on parle dans ce cas-là de fondamentale manquante.

L'appareil auditif arrive aussi à détecter parfois des hauteurs pour des sons inharmoniques et même des bruits (voir ci dessous). Il est également à noter que, même si nous pouvons entendre des fréquences allant de 20Hz à 20000Hz, notre sensibilité aux hauteurs semble être plus réduite [Roa98].

1.5.3 Perception des fréquences

Discrimination fréquentielle

La perception de plusieurs fréquences n'est pas non plus un sujet trivial. Lorsque deux fréquences sont éloignées, le système auditif arrive à les discerner. Si elles sont trop proches, le système les *fusionne* pour donner une sensation de battements (voir section 1.3.4) et/ou de rugosité [Pre00] : une seule fréquence qui varie est alors perçue. La différence de fréquences (intervalle) pour passer d'une sensation à une autre dépend également de la valeur des fréquences et leur relation n'est pas linéaire. Pour simplifier, en dessous de 500Hz, deux fréquences seront discernées si elles sont séparées de 1.8Hz. Au dessus de 500Hz, l'écart autour d'une fréquence ne doit pas dépasser 0.35%.

Bandes critiques

De plus, concernant l'intensité, lorsque les deux fréquences sont suffisamment proches, l'intensité perçue de la somme des deux fréquences est moins importante que la somme des deux intensités. Par contre, si elles sont éloignées, les sensations d'intensité s'ajoutent. Ici aussi, des intervalles de fréquence sont définis. Ces intervalles sont liés aux propriétés physiques de la membrane basilaire et sont appelés bandes critiques. L'existence de ces bandes expliquent aussi certaines propriétés de masquage de sinusoïdes pures par du bruit [ZF99]. Plusieurs modélisations de ces bandes critiques ont été proposées, dont une est particulièrement utilisée : l'échelle des Barks.

Bandes Barks

L'échelle des Barks[1] a été présentée par Zwicker en 1961. Elle est disponible dans la littérature (par exemple [ZF99]). En dessous de 500Hz, la largeur des bandes critiques est constante. Au dessus de 500Hz, cette largeur augmente avec la fréquence. La figure 1.18 illustre la définition des bandes Barks. Il n'existe pas de formule exacte pour passer des Hertz aux Barks, mais seulement des formules d'approximation pour déterminer les frontières de ces bandes. En notant b la bande bark, et f une fréquence en Hz, on a :

$$b(f) = \begin{cases} \frac{f}{100} & \text{pour } f \leq 500 \\ 9 + 4\log_2(\frac{f}{1000}) & \text{pour } f > 500 \end{cases} \tag{1.88}$$

$$f(b) = \begin{cases} 100b & \text{pour } b \leq 5 \\ 1000 \times 2^{\frac{b-9}{4}} & \text{pour } b > 5 \end{cases} \tag{1.89}$$

Une formule peut également donner la largeur d'une bande critique, notée Δf (Hz), en fonction de la fréquence centrale f (Hz) :

$$\Delta f = 25 + 75[1 + 1.4(\frac{f}{1000})^2]^{0.69} \tag{1.90}$$

Bandes ERB

Une échelle différente a été proposée par Moore et Glasberg, proche de celle des Barks : l'échelle ERB *Equivalent Rectangular Band*. Les détails à propos de cette échelle, et de la comparaison avec l'échelle Bark sont présentés dans [SA99]. Ces bandes ont été utilisées pour la modélisation de résidus [Goo96]. Nous revenons sur cette approche dans la section 2.3.5.

[1] du nom du chercheur Allemand Barkhausen (1881-1956)

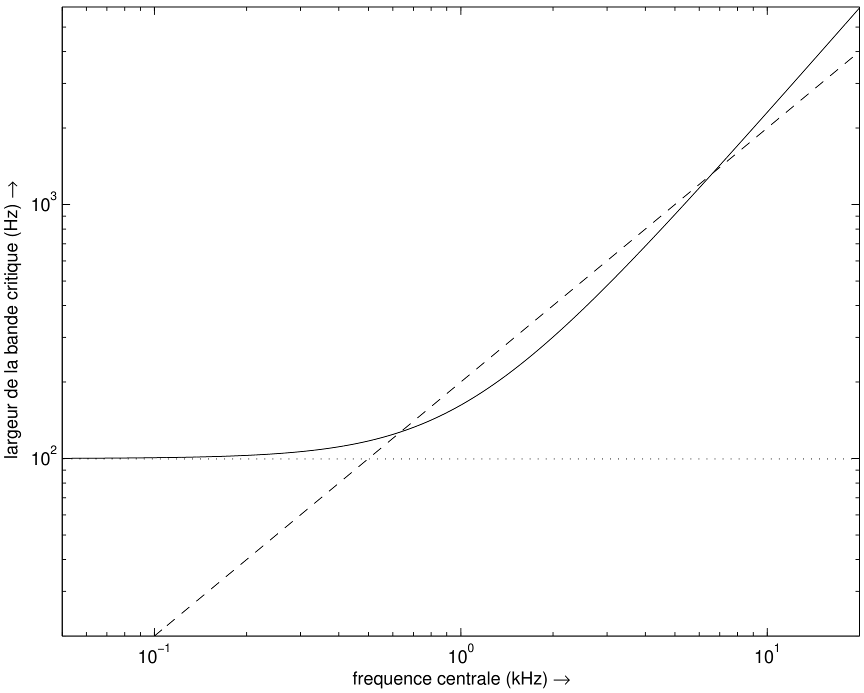

Fig. 1.18 – Représentation de la largeur de la bande critique en fonction de la fréquence. Les fonctions d'approximation sont aussi indiquées, avec, pour les basses fréquences une largeur constante à 100Hz (pointillés), et avec, pour les hautes fréquences, une largeur proportionnelle à la fréquence $0.2f$ (tirets).

Bande Bark	f_{inf}	f_{sup}	f_c	Δf	Bark	f_{inf}	f_{sup}	f_c	Δf
$0-1$	0	100	50	100	$12-13$	1720	2000	1850	280
$1-2$	100	200	150	100	$13-14$	2000	2320	2150	320
$2-3$	200	300	250	100	$14-15$	2320	2700	2500	380
$3-4$	300	400	350	100	$15-16$	2700	3150	2900	450
$4-5$	400	510	450	110	$16-17$	3150	3700	3400	550
$5-6$	510	630	570	120	$17-18$	3700	4400	4000	700
$6-7$	630	770	700	140	$18-19$	4400	5300	4800	900
$7-8$	770	920	840	150	$19-20$	5300	6400	5800	1100
$8-9$	920	1080	1000	160	$20-21$	6400	7700	7000	1600
$9-10$	1080	1270	1170	190	$21-22$	7700	9500	8500	1800
$10-11$	1270	1480	1370	210	$22-23$	9500	12000	10500	2500
$11-12$	1480	1720	1600	240	$23-24$	12000	15500	13500	3500

Tab. 1.3 – Tableau définissant les 24 bandes Barks avec, pour chaque bande, la fréquence limite inférieure f_{inf}, la fréquence limite supérieure f_{sup}, la fréquence centrale f_c et la largeur de la bande Δf.

1.5.4 Perception de la hauteur pour les bruits

Des bruits caractérisés par des spectres composés de variations importantes localisées peuvent entraîner des sensations de hauteur. Ainsi, un bruit blanc modifié par un filtre passe-bas ou passe-haut de pente aiguë (120dB par décade) laissera percevoir une hauteur correspondant à la fréquence de coupure de ce filtre. Cela peut aussi être le cas pour un filtre passe-bande qui génère une bande de bruit. Dans ce cas, la hauteur perçue dépend également de la largeur de la bande de bruit, mais aussi de la fréquence centrale. En simplifiant un peu, nous indiquons ici les deux cas.

Cas de bandes larges

Lorsque la bande de bruit est large (supérieure à 800Hz [ZF99], ou supérieure à un cinquième d'une octave [Roa98]), l'auditeur perçoit deux hauteurs correspondant à peu près aux deux fréquences extrêmes de la bande (les deux fréquences de coupure du filtre passe-bande). La figure 1.5.4 illustre ce cas.

Cas de bandes étroites

Lorsque la bande est étroite, le système auditif *fusionne* les deux fréquences aux extrémités de la bande pour produire une sensation de hauteur correspondant à la fréquence centrale de la bande. La figure 1.5.4 illustre ce cas.

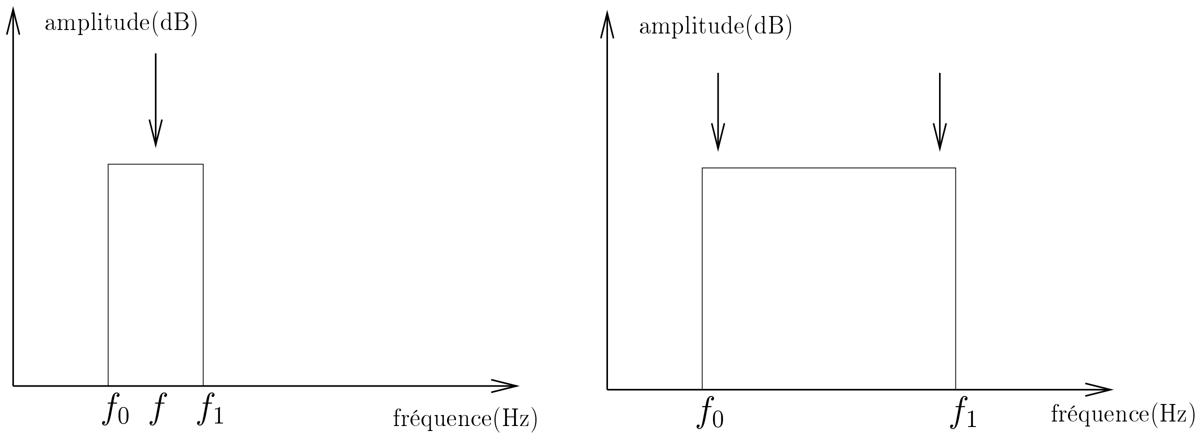

FIG. 1.19 – Perception de hauteur pour un bruit : pour une bande de bruit étroite, la hauteur perçue correspond à la fréquence centrale (figure de gauche), alors que pour une bande de bruit large, deux hauteurs sont perçues correspondant à peu près (4%) aux deux extrémités de la bande (figure de droite).

Bruit ondulant

Le bruit ondulant est défini dans la section 1.4.5. Ces bruits donnent une sensation de hauteur assez précise. Cette sensation peut s'expliquer par le système de détection des fréquences, lié à la membrane basilaire, ou par le système de détection des périodes, voire les deux.

1.5.5 Perception de l'enveloppe temporelle

Une sinusoïde de fréquence et d'amplitude fixes est perçue comme un son parfaitement stable, alors que l'amplitude de l'onde varie beaucoup au cours du temps. Cette stabilité peut être expliquée par le fait que l'enveloppe temporelle reste constante au cours du temps. A l'opposé, un signal qui laisse une impression de mouvements est caractérisé par une enveloppe temporelle plus chahutée. La figure 1.20 illustre ce phénomène.

La caractérisation mathématique de l'enveloppe temporelle a été détaillée dans la section 1.1.9. Des expériences psychoacoustiques lui attribuent un rôle essentiel lors de la détection d'une sinusoïde dans une bande de bruit : le système auditif utilise en plus de l'information liée à la différence d'énergie, la stabilité de l'enveloppe temporelle [Har97]. Une autre application de ce modèle d'enveloppe temporelle est l'explication de la faculté à discerner des bandes de bruit de densité spectrale différente. Nous revenons sur ce point dans la prochaine section, mais aussi plus loin.

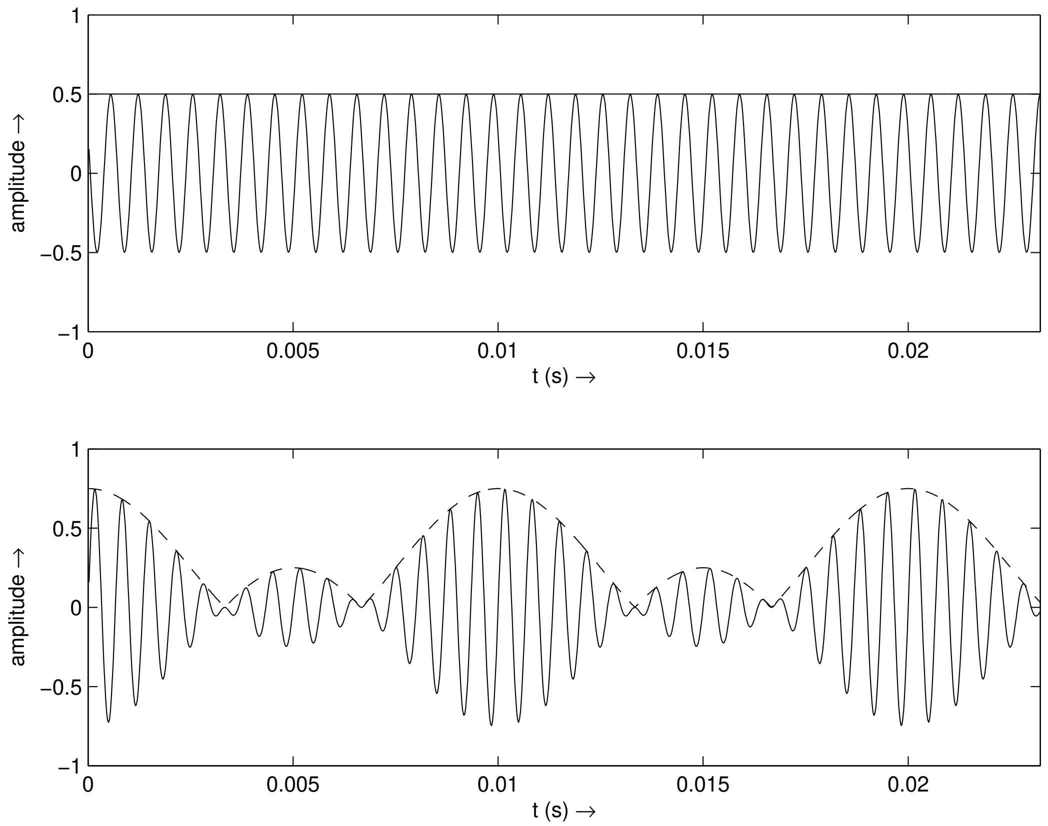

Fig. 1.20 – Exemple de deux enveloppes temporelles de deux signaux simples : sur la figure du haut, une sinusoïde laisse une impression de stabilité car son enveloppe temporelle est constante, alors que sur la figure du bas, le signal composé de plusieurs sinusoïdes donne une impression de mouvement, qui s'explique en partie par les variations de cette même enveloppe.

1.5.6 Masquage par le bruit

Le masquage est un phénomène psychoacoustique auquel tout le monde est soumis : un son peut en cacher un autre partiellement ou entièrement. Ainsi, une sinusoïde plongée dans du bruit n'est pas toujours perçue. En musique, un son d'un instrument de volume constant peut être masqué par un son provenant d'autre instrument dont le volume est plus fort.

En général, de nombreuses expériences sont menées pour tester la perception de sinusoïdes avec du bruit, des bandes de bruit et d'autres sinusoïdes. Le masquage est un phénomène qui peut intervenir sous deux formes. Le masquage fréquentiel apparaît lorsque deux fréquences proches ont une grande différence d'amplitude. Dans ce cas, la fréquence d'amplitude importante peut masquer la seconde fréquence et la rendre inaudible. Ces propriétés du système auditif sont mis en application pour la compression de données [GP99], la synthèse accélérée [LM01], ...

Des expériences mettent également en évidence deux phénomènes de masquage temporel [ZF99]. Le post-masquage est mis en évidence par la persistance de l'effet masquant d'un son sur un autre. Même lorsque le son masquant s'arrête, le second son continue à être masqué pendant une petite durée de temps : on parle de post-masquage (*post-masking*). De plus, il peut arriver qu'un son soit masqué avant que le son masquant ne commence à être perçu. Cet effet est moins important que le précédent. On parle alors de pré-masquage (*pre-masking*).

D'autres études portent sur le masquage de sons purs par des bandes de bruit et sur l'efficacité du masquage en fonction des propriétés de ces bandes de bruit. Des expériences ont donc mis en évidence le lien entre la perception des fluctuations d'intensité et le phénomène de masquage, puisqu'une bande de bruit étroite dont les fluctuations d'intensité sont faibles (*low-noise noise* décrit dans la section 1.4.6) est moins efficace pour masquer des sinusoïdes [HP88].

Chapitre 2

Modèles pour les sons

La manipulation des sons à l'aide de l'ordinateur nécessite le développement d'une représentation mathématique adaptée de l'onde physique. Un modèle sonore associe donc des valeurs numériques aux phénomènes réels, et définit ainsi un ensemble de paramètres qui, pour permettre des transformations musicales intuitives, doivent être liés de la manière la plus proche possible à la perception des sons.

Plusieurs sortes de modèles peuvent être proposées, selon la fonctionnalité voulue. Les modèles de compression s'efforceront de définir des paramètres peu coûteux en mémoire pour représenter de façon fidèle des sons naturels. Des modèles de perception s'efforceront de modéliser de manière rigoureuse le fonctionnement du système auditif. Certains modèles de synthèse proposeront de générer des sons originaux, sans disposer de son de base. Même si des réflexions sont menées afin de définir des critères de qualité entre ces différentes approches [Jaf95], il est impossible de les comparer puisque leur développement a souvent été effectué dans des buts complètement différents.

Dans ce travail, nous nous intéressons en particulier aux modèles d'analyse et de synthèse, destinés à la composition musicale. Ces modèles sont définis par un ensemble de paramètres indépendants, qui sont extraits d'un son réel lors de l'étape d'analyse. Ils peuvent être ensuite modifiés, pour générer un nouveau son, résultant de la transformation du son original. La figure 2.1 illustre le principe de cette famille de modèles. Les recherches concernant le développement et l'amélioration de ces modèles sont très nombreuses actuellement, et mènent à des résultats de très haute qualité. Cependant, il apparaît que les travaux existants concernent essentiellement les sons instrumentaux très peu bruités, et que leur application à ce type de sons pose de nouveaux problèmes. Dans la première section, nous présentons les modèles physiques et abstraits en montrant leurs intérêts et leurs limites concernant les transformations musicales. Ensuite, nous détaillons les approches possibles pour définir un modèle temporel adapté aux signaux aléatoires ou bruits. Enfin, après avoir introduit les modèles spectraux existants et leurs limites, nous proposons un nouveau modèle pour représenter les sons bruités, à l'aide de paramètres statistiques : le modèle CNSS. Les transformations musicales permises par ce modèle sont expliquées et détaillées.

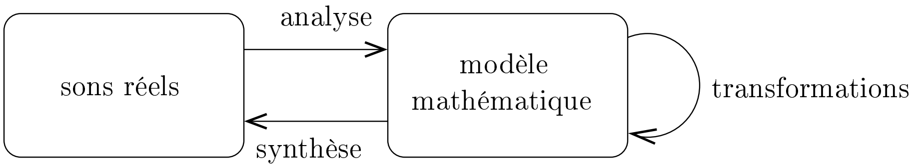

FIG. 2.1 – Schéma général de l'analyse, la transformation et la synthèse d'un son.

2.1 Types de modélisation

Dans cette section nous présentons brièvement deux grandes familles de modèles de sons : les modèles physiques et les modèles abstraits. Pour chaque type de modèle, nous essayons de dégager leurs principaux intérêts et limites, dans le cadre que nous nous sommes fixé.

2.1.1 Modèles physiques

Les modèles physiques sont définis à partir d'une représentation physique de la source sonore, par exemple un instrument de musique. A partir de ce modèle de source sonore, il est possible de générer, à l'aide des connaissances en mécanique (fluide ou solide) et acoustique, les ondes sonores correspondant à chaque source, et donc le signal émis. Ces modèles peuvent à la fois servir à la composition sonore [Cad02], mais aussi à la simulation de phénomènes physiques.

Principe

Pour simplifier, la modélisation physique est basée sur l'interaction entre un excitateur et un résonateur. L'excitateur va déclencher des vibrations, comme par exemple le pincement d'une corde, un coup sur une membrane, ... Le résonateur va, lui, définir la réponse du corps de l'instrument à cette excitation, et se représente donc par un filtre variant dans le temps.

Un des outils mathématiques le plus couramment employé en modélisation physique est le paradigme masse-ressort [Roa98]. Il est notamment utilisé pour la modélisation de cordes vibrantes (cordes d'instrument) en reliant successivement ces objets mathématiques sur une dimension, mais aussi de surfaces vibrantes (peaux de tambours), en les reliant sur deux dimensions. Le déplacement de ces ondes vibrantes est aussi régi par les lois physiques selon un système de filtres : le modèle des guides d'ondes (*waveguides*) [Smi97].

La qualité des sons produits par ce type de modèle est reconnue. Cette qualité est due à la rigueur des modèles développés par les physiciens. Ainsi, de nombreux synthétiseurs commerciaux les mettent en application.

Transformations

Les transformations autorisées par les représentations de sons liées à des modèles physiques sont très nombreuses, voire même illimitées car ces modèles permettent de développer des instruments qui seraient impossibles à construire en réalité (tambours gigantesques, corps d'instruments qui varient dans le temps, etc. . .).

Limites

En dépit de la qualité des sons produits et des transformations envisageables, la modélisation physique présente quelques défauts. Nous présentons ici les principales limites. La première concerne l'efficacité du modèle de synthèse. En effet, les modèles physiques imposent un temps de calcul très important qui rend quasiment impossible le rendu en temps-réel, même en tenant compte de la puissance des machines actuelles [CC00]. Les seules possibilités de diminuer les temps de calcul ne peuvent être envisagées qu'en simplifiant les modèles physiques, ce qui n'est envisageable que pour certains sons [Coo97].

Une autre limite de cette approche physique provient du manque de contrôle musical. Bien que certains musiciens tentent de mettre en avant les possibilités de ces techniques [Cad02], il est difficile pour un musicien de modifier la perception d'un son en agissant sur des caractéristiques physiques souvent obscures pour des novices en physique (coefficients d'élasticité, viscosité, etc...).

Enfin, il est parfois difficile de trouver le modèle physique adéquat pour représenter un son. Il est d'abord nécessaire de connaître l'origine physique d'un son à synthétiser, ce qui n'est pas toujours le cas, puisqu'un son complexe, notamment les bruits, trouve son origine dans une multitude de phénomènes physiques indépendants qui peuvent avoir des natures complètement différentes. Un modèle physique peut s'avérer, dans de tels cas, très complexe à développer, et le temps de calcul requis rendrait son application impossible.

2.1.2 Modèles abstraits

Les modèles dits abstraits sont les premiers modèles développés pour la synthèse de sons, car ils sont rapidement apparus comme facilement applicables au traitement informatique des sons.

Principe

Les modèles abstraits représentent les sons à partir de formules mathématiques. L'exemple de modèle abstrait le plus employé est la synthèse par modulation de fréquence (FM) [Roa98]. Le son synthétisé est produit à partir de sinusoïdes dont les fréquences varient périodiquement dans le temps. Cette variation rapide densifie le spectre du son en ajoutant de nombreux composants spectraux, et modifie donc le timbre du son [Cho73]. Quasiment toutes les sortes de sons peuvent être ainsi synthétisées en n'ayant recours qu'à un nombre très limité d'oscillateurs. Cette technique ne requiert donc pas une puissance de calcul très importante, ce qui a permis son application dans le domaine public dès le début des années 80, avec notamment la commercialisation du synthétiseur *Yamaha DX7*.

Limites

Les paramètres des modèles abstraits sont donc des paramètres purement mathématiques. Certains musiciens ont appris par la force des choses à manipuler ces paramètres, mais ils demeurent éloignés des paramètres musicaux ou perceptifs.

De plus, les modèles abstraits ne sont que des modèles de synthèse (synthèse FM, par modulation d'amplitude, par modulation en anneau, ...). Il n'existe pas de méthode d'extraction des paramètres mathématiques pour ces modèles. Ainsi il n'est pas possible d'effectuer des transformations musicales à partir de sons naturels, mais seulement sur des sons de synthèse existants, imitant des sons naturels.

Enfin, ces modèles sont souvent liés aux techniques de modulation qui produisent des sons présentant des caractéristiques du timbre souvent ressemblantes, qui sont reconnues par un musicien expérimenté. C'est pourquoi les recherches concernant ces modèles stagnent à présent, et se dirigent vers des méthodes réclamant certes plus de puissance de calcul, mais permettant des manipulations musicales plus variées.

2.2 Modèles temporels

Comme nous l'avons vu précédemment, un signal sonore peut être représenté sur l'échelle temporelle ou sur l'échelle fréquentielle. La première approche est la plus courante. La quasi-majorité des formats des fichiers audio (`wav`, `aiff`, ...) sauvegarde un signal sous sa forme temporelle. Pourtant, comme nous allons le voir dans cette partie, cette approche n'est pas forcément celle qui est le plus naturellement liée à la perception.

Sous sa forme temporelle, un signal sonore numérique est représenté par une suite de valeurs obtenues par quantification du signal électrique résultant d'un transducteur (microphone), provoqué par les dépressions d'air dues à l'onde physique sonore. Ces valeurs sont quantifiées à des intervalles de temps réguliers. La représentation temporelle d'un son est donc une fonction discrète de l'amplitude en fonction du temps.

Les problèmes relatifs à l'échantillonnage ne seront pas abordés dans ce document puisque ils sont beaucoup traités dans la littérature sur le traitement du signal numérique (par exemple [Moo90], [Orf96]). Nous présentons les paramètres qui peuvent être obtenus directement à partir de la représentation temporelle d'un son, leurs liens avec la perception et leur dépendance entre eux. Ensuite, nous présentons les transformations possibles ainsi que les principales limites.

2.2.1 Principe

La courbe de l'amplitude du son en fonction du temps contient toutes les informations du signal. Parmi ces informations, seules celles liées à la perception nous intéressent. Le premier paramètre classique est le volume. Celui-ci est lié aux valeurs d'amplitude du signal, comme cela a été vu dans la section 1.5.1. Il est utilisé dans tous les modèles d'analyse et synthèse de sons. En ce qui concerne les signaux aléatoires, nous étudions les propriétés de la fonction de distribution des échantillons, en essayant de voir son lien avec la perception dans le cas général.

Répartition des échantillons

Il apparaît assez rapidement qu'il n'existe que très peu de lien entre la distribution des échantillons et la perception fréquentielle. En effet, quelques exemples très simples peuvent en attester. Tout d'abord, un son pur tel qu'une sinusoïde de fréquence et d'amplitude constantes présente une fonction de distribution à peu près constante quelles que soient sa fréquence et sa phase initiale. Pourtant la valeur de la fréquence est évidemment un élément perceptif prépondérant. La figure 2.2 montre l'équivalence entre la fonction de distribution théorique d'une sinusoïde d'amplitude 1.0 et une fonction de distribution calculée expérimentalement à partir d'une sinusoïde de fréquence 440Hz.

Par ailleurs, un autre exemple peut être donné en considérant un signal composé de plusieurs sinusoïdes, dont les phases sont choisies selon une loi uniforme. Nous verrons que cette approche est conforme à la définition théorique de bruits (section 2.4.2). L'hypothèse d'ergodicité nous permet de comparer la fonction de densité de probabilité des échantillons produits à celle d'un bruit blanc gaussien. Il apparaît que cette fonction présente, sauf lors d'exceptions liées aux choix particuliers de fréquences et/ou de phases, les mêmes caractéristiques. Ainsi, un signal généré avec seulement six sinusoïdes de fréquences aléatoires dans une bande de largeur 5000Hz présente des moments statistiques très proches de ceux d'une distribution gaussienne (voir tableau 2.1). La sensation produite par ces deux sons est pourtant très différente. Cette tendance pour un signal à avoir une distribution gaussienne est liée au théorème central limite (ou loi des grands nombres) présenté dans la section 1.2.11.

De même, le lien entre les propriétés du timbre et la fonction de densité de probabilité dans le cas d'un son aléatoire apparaît faible. Si nous considérons un bruit blanc gaussien et un bruit blanc uniforme, les fonctions de densité de probabilité sont différentes, et pourtant la sensation auditive est la même. Une fonction de densité de probabilité d'un signal aléatoire différente n'implique pas forcément une perception différente du son. Des exemples concernant également des signaux aléatoires peuvent être décrits pour mettre en avant le manque d'informations contenues dans la fonction de distribution des échantillons. Ainsi, la fonction de densité de probabilité d'un bruit rose est identique à celle d'un bruit blanc gaussien. La figure 2.3 montre les deux fonctions dans ces deux cas, et le tableau 2.2 présente quelques valeurs de moments statistiques calculées et illustre quantitativement cette similitude. Une fonction de densité de probabilité équivalente n'implique donc pas forcément des propriétés auditives équivalentes. Cette observation peut d'ailleurs être étendue aux cas de bruits blancs filtrés (bruits colorés). Ces bruits auront tendance à avoir une fonction de densité de probabilité proche d'une fonction caractéristique d'une loi de Gauss. La fonction de densité de probabilité ne contient pas, dans ce cas, l'information relative à l'enveloppe spectrale du son.

Un autre exemple peut être fourni par un son produit par répétitions successives d'une même fenêtre temporelle d'un bruit blanc gaussien. D'un point de vue statistique, ce son est identique à un bruit blanc gaussien de durée infinie. Pourtant, à l'écoute, ce son laisse percevoir une hauteur, liée à sa périodicité. La fonction de densité de probabilité n'est pas suffisante dans ce cas, pour définir cette propriété auditive primordiale.

La fonction de densité de probabilité d'un signal sonore ne contient donc pas forcément les informations perceptives liées à la hauteur ou à l'enveloppe spectrale du son. Néanmoins, cette fonction peut s'avérer utile pour décrire d'autres propriétés auditives liées aux fluctuations temporelles du signal. Ainsi, les sons harmoniques de hautes fréquences, correspondant à des notes aiguës, contenant très peu de bruits, pourraient être caractérisés par des fonctions de densité de probabilité proches de la forme de celles des sinusoïdes pures. A l'opposé, les sons contenant une attaque, comme les sons percussifs, présenteront un saut d'amplitude très localisé et seront donc caractérisés par une fonction de densité de probabilité plus "fine" et

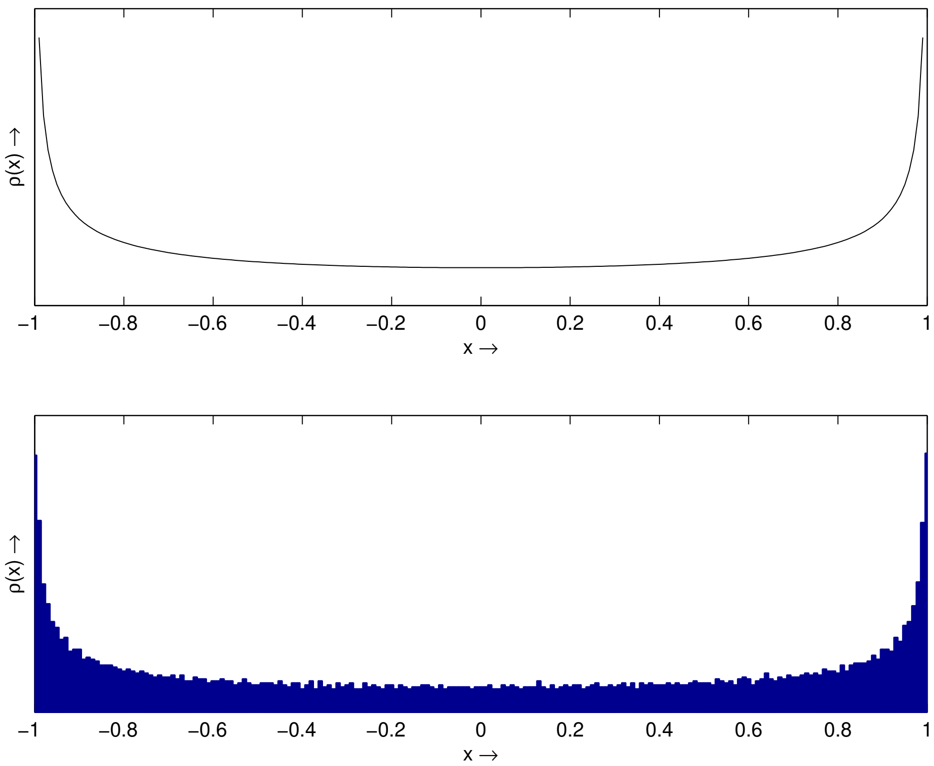

FIG. 2.2 – Représentation théorique de la fonction de distribution d'une sinusoïde pure : la fonction théorique est montrée sur la figure du haut alors que le résultat mesuré à partir de 4096 échantillons (93ms) d'une sinusoïde de 440Hz, de phase nulle, d'amplitude 1.0 est présenté sur la figure du bas. Quelle que soit la fréquence de cette sinusoïde, la fonction de distribution a la même forme et les échantillons du signal ont les mêmes propriétés statistiques.

Signal	moyenne	minimum	maximum	skewness	kurtosis
gaussienne théorique	0	-1	1	0	3
6 sinusoïdes	0.0018	-0.9739	0.9093	0.0072	2.9947
bruit blanc gaussien	-0.0034	-0.8677	0.9075	0.0257	3.0246

TAB. 2.1 – Exemples de valeurs de quelques caractéristiques statistiques (mesurées sur 4096 échantillons) d'un bruit blanc gaussien et d'un signal constitué de 6 sinusoïdes dont les fréquences sont aléatoirement tirées selon une loi uniforme entre 0 et 5000Hz et dont les phases sont uniformément tirées entre 0 et 2π.

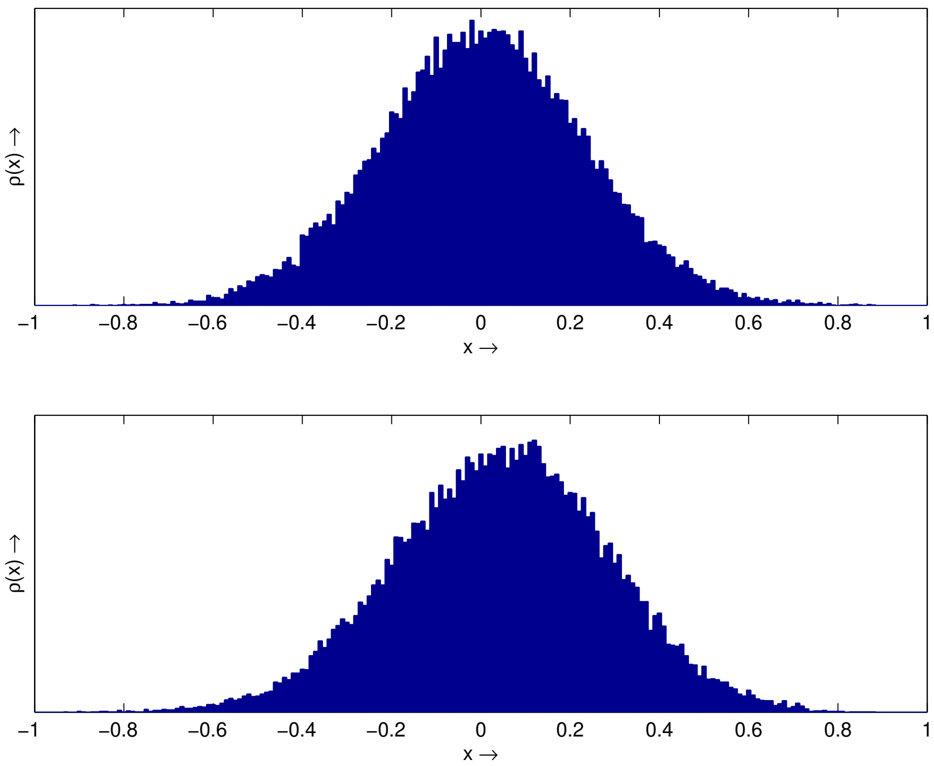

Fig. 2.3 – Représentation de la répartition de 4096 échantillons (93ms) d'un bruit blanc gaussien (figure du haut) et d'un bruit rose (figure du bas) : les deux approximations des densités de probabilité sont comparables.

Signal	moyenne	minimum	maximum	skewness	kurtosis
gaussienne théorique	0	-1	1	0	3
bruit rose	0.0465	-0.9333	0.8797	-0.0896	3.0098
bruit blanc gaussien	0.0035	-0.9081	0.8802	0.0194	3.0112

Tab. 2.2 – Exemples de valeurs de quelques caractéristiques statistiques (mesurées sur 4096 échantillons) d'un bruit blanc gaussien et d'un bruit rose.

plus “pointue”. La figure 2.4 montre les fonctions de densité de probabilité (calculées sur 4096 échantillons) d'un son harmonique faiblement bruité issu d'un hautbois, et d'un son percussif issu d'un xylophone. Les données quantitatives relatives à ces différences de forme sont en particulier les quatrièmes moments statistiques (kurtosis). Ce moment est lié à l'aplatissement de la loi de distribution. Dans le cas d'un son harmonique peu bruité, ce moment prend une valeur proche de 1.5 (kurtosis théorique d'une sinusoïde), inférieure à 3 (3 étant la valeur associée à la loi de Gauss), alors que dans le cas d'un son contenant une attaque, ce moment prend une valeur bien supérieure à 3. Le tableau 2.3 montre que le kurtosis associé au son percussif est plus important que celui associé au son harmonique, mais aussi plus généralement à celui des bruits gaussiens. Le kurtosis est donc un paramètre qui peut s'avérer utile pour caractériser des sons ou des segments de sons. Nous étudions les applications de cet outil statistique à la segmentation dans la section 3.3. Il est également à noter que W. Hartmann utilise aussi le quatrième moment statistique de la courbe d'amplitude du signal sonore pour mesurer les fluctuations des bruits dans le cadre d'expériences psychoacoustiques sur le masquage [HP88]. Nous y revenons plus en détails dans la section 2.4.4.

La fonction de densité de probabilité ne contient donc pas les données du son relatives au spectre et, en particulier son enveloppe. Mais l'étude de sa forme donne des informations sur les variations d'amplitude au cours du temps.

Influence des écarts entre les échantillons sur les propriétés spectrales

Après avoir vu quelles informations nous pouvions retirer de l'approximation de la fonction de répartition des échantillons, nous nous intéressons aux approches liées à la représentation temporelle permettant d'extraire des informations spectrales. Les périodicités d'un signal sont reliées aux écarts entre les échantillons successifs d'un même signal sonore. L'exemple le plus connu est le bruit brownien : l'écart entre les échantillons est aléatoirement choisi selon une loi uniforme. Cette propriété est mise en application pour la synthèse de ce type de bruits (voir la section 4.1.3). Un changement de la loi de distribution entre deux échantillons consécutifs produit un changement de la forme de l'enveloppe spectrale. Cependant, la loi définissant l'écart ne caractérise pas toujours cette forme. Par exemple, un bruit blanc gaussien est caractérisé par une distribution gaussienne des écarts (la différence de deux variables aléatoires gaussienne suit une loi gaussienne), alors qu'un bruit blanc uniforme est caractérisé par une distribution des écarts de type triangulaire (voir la figure 2.5 et le tableau 2.4). Pourtant ces sons présentent le même spectre et les mêmes propriétés auditives.

Le contrôle statistique des écarts entre deux échantillons consécutifs ne permet de définir que la forme grossière de l'enveloppe spectrale et, par exemple, peut être utile pour colorer basiquement un bruit. Pour pouvoir effectuer une transformation plus complexe de l'enveloppe spectrale, il faut pouvoir définir statistiquement les écarts entre plusieurs échantillons consécutifs. Cette approche revient à une opération de filtrage dont les coefficients du filtre sont des variables aléatoires. Elle est donc vite rendue très complexe par les problèmes classiques liés à la stabilité des filtres.

L'écart entre les échantillons ne peut donc être contrôlé qu'à partir de valeurs déterministes fixées. Une enveloppe spectrale plus complexe peut être représentée sous la forme de coefficients représentant le filtre à appliquer à un bruit blanc pour obtenir le son souhaité. La technique la plus classique (surtout en traitement de la parole) pouvant être utilisée dans ce cadre

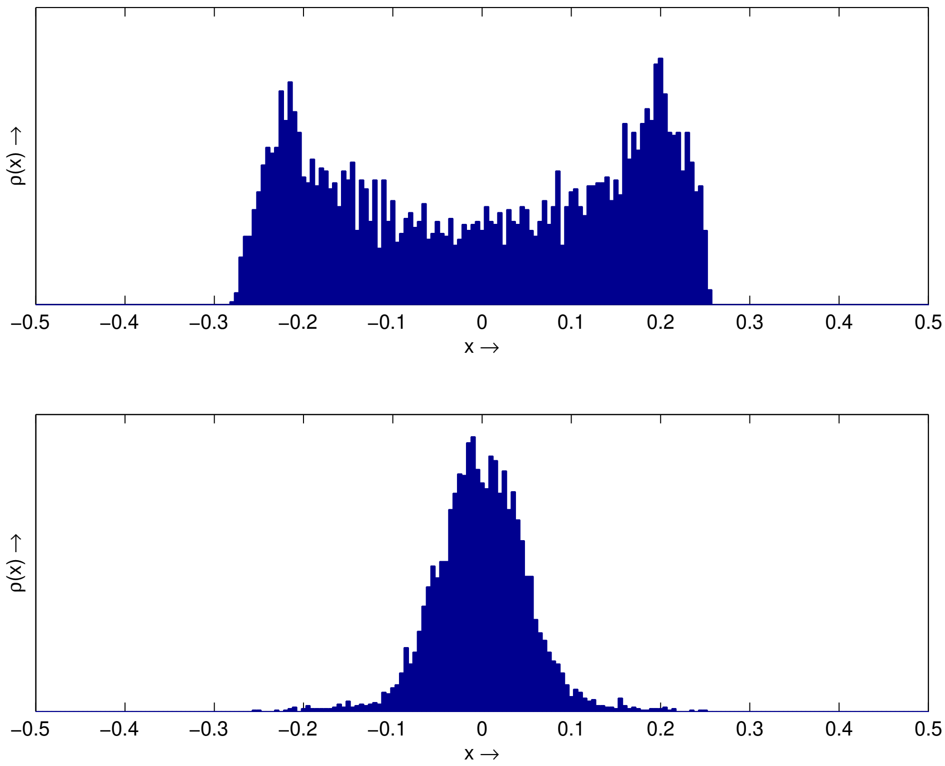

Fig. 2.4 – Représentation de la répartition de 4096 échantillons (93ms) provenant d'un son harmonique de hautbois (figure du haut) et d'un son percussif de xylophone (figure du bas) : les deux approximations des densités de probabilité ont des caractéristiques bien distinctes.

Signal	moyenne	minimum	maximum	skewness	kurtosis
Son harmonique	0.0000	−0.2778	0.2561	−0.0755	1.5287
Son percussif	−0.0008	−0.2531	0.2520	−0.0241	4.9965

Tab. 2.3 – Exemples de valeurs de quelques caractéristiques statistiques (mesurées sur 4096 échantillons) d'un son harmonique issu d'un hautbois et d'un son percussif.

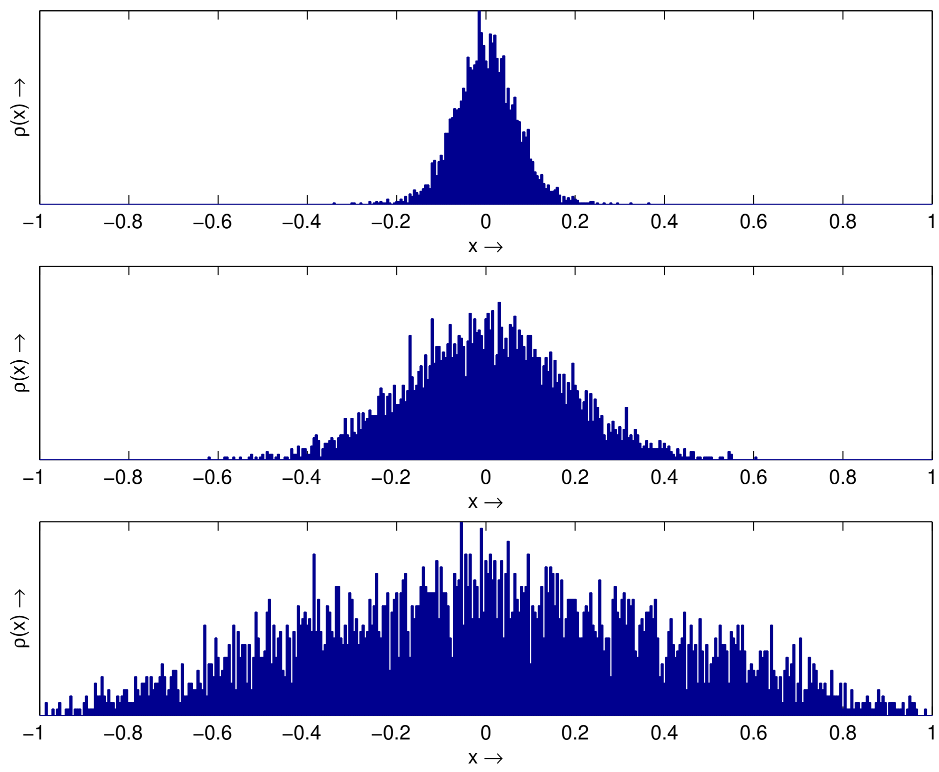

FIG. 2.5 – Représentation de la répartition de 4096 écarts entre deux échantillons consécutifs provenant d'un bruit rose (en haut), d'un bruit blanc gaussien (au milieu) et d'un bruit blanc uniforme (en bas).

Signal	moyenne	minimum	maximum	skewness	kurtosis
Bruit blanc gaussien	0.0001	−0.6210	0.6034	−0.0099	2.9751
Bruit blanc uniforme	0.0001	−0.9861	0.9855	0.0158	2.4143
Bruit rose	−0.0000	−0.6580	0.6210	0.0108	2.9287

TAB. 2.4 – Exemples de valeurs de quelques caractéristiques statistiques (mesurées sur 4096 échantillons) des écarts entre les échantillons d'un bruit blanc gaussien, un bruit blanc uniforme et d'un bruit rose.

est la prédiction linéaire (LPC *Linear Prediction Coding*)[Mak75] [Moo90], qui approxime chaque échantillon du signal à une combinaison linéaire d'un certain nombre p d'échantillons précédents. Le filtre correspondant peut s'écrire sous la forme :

$$Y[n] = a_0 X[n] + \sum_{i=1}^{p} b_i Y[n-i] \tag{2.1}$$

où a_i et b_i sont les coefficients du filtre, $X[n]$ le signal source ou l'erreur, et $Y[n]$ le signal modélisé. Cette technique reste néanmoins limitée pour les variations de l'enveloppe spectrale. Celle-ci est obligatoirement modélisée sous une forme lissée, pour éviter des problèmes relatifs à la stabilité et à la complexité des calculs. En effet, plus l'enveloppe est complexe, c'est-à-dire plus elle contient de creux et de bosses, plus le nombre de coefficients à déterminer est important.

Cette approche de prédiction linéaire revient à un système excitation-filtre, vu dans la section 2.1.1. Dans le cas de sons bruités (en parole les sons non voisés), le signal source est un bruit blanc gaussien. D'autres types de signaux peuvent être donnés en entrée du filtre [KAU00]. Des travaux ont été menés sur les propriétés statistiques de ce signal source, en particulier les moments statistiques [DT97]. Ces propriétés statistiques permettent d'établir un classement du timbre des instruments de musique lors de l'étude de leur partie bruitée (résidu), mais modéliser ces parties bruitées par ces valeurs semble difficile [Dub96]. Une autre approche très récente consiste à appliquer cette même technique de prédiction linéaire à la source, mais cette fois-ci dans le domaine fréquentiel, dans le but de modéliser les variations de l'enveloppe temporelle [AE03].

Enveloppe temporelle

L'enveloppe temporelle (ou enveloppe dynamique) est également une donnée importante sur le plan de la perception, comme nous l'avons vu dans la section 1.5.5. A partir de la représentation temporelle d'un signal, il est possible de modifier cette enveloppe temporelle en multipliant le signal original par un signal basse fréquence, ce qui revient à une approche de type modèle abstrait par modulation. Le problème principal provient du fait qu'une modification de l'enveloppe temporelle par modulation implique une modification de l'enveloppe spectrale si la fréquence de modulation se situe dans le domaine audible (au dessus de 20Hz). Cette propriété est même la base de la synthèse FM [Cho73]. Et, à l'inverse, le filtrage d'un son pour modifier l'enveloppe spectrale peut impliquer un changement des propriétés de l'enveloppe dynamique [HP88]. Ces deux opérations ne sont pas indépendantes, alors que ces deux paramètres le sont.

Conclusion

Nous pouvons donc conclure de cette étude sur les modèles temporels, que, pour l'instant, seulement deux paramètres sont envisageables pour un modèle : l'enveloppe spectrale et l'enveloppe temporelle. Sur le plan de la perception, ce paramètre est très important, car il est lié à la notion de timbre. Cela semble toutefois limité pour représenter des sons complexes. Les autres paramètres, liés aux propriétés statistiques ne paraissent pas être directement significatifs des propriétés auditives, à part les variations de l'enveloppe temporelle. Mais l'influence de l'opération de filtrage sur cette enveloppe rend impossible les manipulations de ces paramètres indépendamment.

2.2.2 Transformations

Aucune hypothèse n'est nécessaire pour représenter un son sous sa forme temporelle (variations d'amplitude en fonction du temps). Cette représentation permet de reproduire sans distortion audible tous les sons, à condition que la fréquence d'échantillonnage soit adaptée. Cette représentation permet également d'effectuer à priori toutes les transformations possibles, concernant la dynamique, la hauteur, l'espace (spatialisation, déplacements), la durée, le timbre.

Amplification

L'amplification d'un son s'effectue en multipliant tous les échantillons de sa représentation temporelle par une valeur, pouvant éventuellement varier au cours du temps : si cette variation est trop rapide, le spectre du son est modifié (modulation d'amplitude), et donc son timbre. Si elle est lente, l'enveloppe dynamique est modifiée.

Modification de l'enveloppe spectrale (couleur)

Comme nous l'avons déjà abordé juste avant, il est possible de modifier l'enveloppe spectrale (la couleur) d'un son à partir de sa représentation temporelle, en filtrant le signal. La technique la plus couramment utilisée dans ce cadre est la prédiction linéaire.

En effet, la technique de prédiction linéaire est particulièrement adaptée à l'analyse et la synthèse de l'enveloppe spectrale lissée de signaux aléatoires, car les techniques de filtrage sont très efficaces et réclament un temps de calcul très raisonnable. Un problème peut néanmoins être soulevé quant à l'intuitivité musicale des paramètres mathématiques que sont les coefficients du filtre. Un musicien se retrouvant face à ces valeurs mathématiques ne saura certainement pas quels coefficients modifier pour amplifier ou atténuer telles bandes de fréquences. Or les musiciens doivent être les principaux utilisateurs de ces modèles de sons. C'est pourquoi il nous semble qu'une approche spectrale est plus adaptée, car plus proche de la perception.

La définition d'un filtre peut être effectuée dans le domaine spectral (voir section 1.1.8). Il est toutefois important de rappeler ici que la complexité du calcul de l'enveloppe spectrale voulue dépend du nombre de bosses et de creux. Par cette méthode, l'enveloppe spectrale est forcément lissée.

Effets musicaux

D'autres effets utilisés fréquemment pour la musique sont applicables à partir de cette représentation temporelle. Les exemples sont nombreux, et nous pouvons citer les effets chorus, flanger, etc. Cette représentation est également particulièrement adaptée pour des effets tels que les réverbérations ou les échos, qui se traduisent par l'application de filtres correspondant aux propriétés physiques d'une salle de concert par exemple.

Modification de la durée

Un effet particulièrement intéressant et dont l'intérêt reste indiscutable est le changement d'échelle temporelle. Il est facilement possible d'agrandir ou de réduire la durée d'un son à partir de sa représentation temporelle. Il suffit pour cela de modifier la fréquence d'échantillonnage du signal, ce qui reviendrait à modifier la vitesse de défilement d'une bande sonore [Smi00]. Par exemple, si cette fréquence d'échantillonnage F_s est divisée par 2, alors W_s échantillons représenteront une durée deux fois plus importante. Le problème est que cette méthode modifie également les fréquences du signal, et donc son spectre. Ainsi, par exemple, une sinusoïde de fréquence f_0 verrait sa fréquence divisée par 2 dans le cas d'une fréquence d'échantillonnage multipliée par 2. Cette propriété se traduit en pratique : si un son est accéléré, les fréquences deviennent plus aiguës (effet *Mickey* lorsque cela concerne les sons provenant de la parole).

Une solution a été proposée pour éviter cette transformation spectrale indésirable dans le cas de sons harmoniques [Mou90]. Elle provient de techniques employées pour la synthèse de la voix. Ces méthodes appelées PSOLA (*Pitch Synchronized OverLap Add*) considèrent le signal comme une répétition périodique de formes d'onde. Cette périodicité est associée à la hauteur fondamentale du son. Ainsi, les opérations dans le temps sont effectuées en enlevant ou en répétant ces segments temporels. Le spectre du son est ainsi préservé [PR98a], avec notamment ses formants. La partie bruitée du signal peut aussi être traitée de manière synchrone à la périodicité fondamentale [Pee01]. Bien que de nombreuses améliorations aient été apportées, cette approche reste limitée aux sons harmoniques comprenant une partie bruitée de faible énergie.

Modification de la hauteur

Comme nous l'avons vu dans la section précédente, les modifications de la hauteur sont liées aux modifications de durée. Ainsi, la méthode de base pour augmenter la hauteur d'un son consiste à modifier la fréquence d'échantillonnage. Cette méthode implique forcément un changement de durée du son.

Pour les sons harmoniques, la méthode PSOLA permet de faire ces modifications de hauteur sans modifier la durée. Pour cela, il faut en fait commencer par modifier la durée du son selon le principe vu juste avant. Par exemple, si on veut augmenter la hauteur d'un son, on commence par augmenter la durée d'un même facteur. Ensuite, un ré-échantillonnage est effectué qui entraîne à la fois le changement de hauteur, et le changement de durée. Ce dernier compense alors le premier changement induit par la technique PSOLA.

En ce qui concerne les translations de bandes de bruit, qui peuvent dans certains cas s'assimiler à une transposition de hauteur, elles ne peuvent être effectuées qu'en modifiant la fréquence d'échantillonnage. Cette transformation entraîne donc forcément aussi une modification de la bande de bruit originale.

2.2.3 Limites et conclusions

La représentation temporelle est la plus couramment utilisée. Cependant, même si elle permet de reproduire fidèlement n'importe quel son ainsi que d'effectuer avec efficacité beaucoup de transformations, elle présente plusieurs inconvénients majeurs. Tout d'abord, les transformations liées à la durée et au timbre au sens général (l'enveloppe spectrale notamment) ne peuvent pas être effectuées indépendamment, sans émettre d'hypothèses sur les propriétés du son. Dans le cas des signaux aléatoires comme les bruits, il n'existe pas, à notre connaissance, de méthodes envisageables pour effectuer ces transformations en ne modifiant qu'un seul paramètre perceptif.

De plus, il apparaît souvent difficile d'établir des liens entre les paramètres mathématiques extraits et des paramètres perceptifs et musicaux, sauf dans le cas du volume. L'exemple le plus significatif est certainement le contrôle de l'enveloppe spectrale par les coefficients du filtre qui ne semblent pas intuitifs pour un musicien. Pour représenter cette enveloppe spectrale, le meilleur moyen reste de visualiser la courbe lissée correspondant à la notion de couleur. Plus généralement, de nombreux paramètres perceptifs sont liés à des données spectrales. C'est pourquoi les approches spectrales prédominent l'essentiel des avancées en matière de modèles de ces dernières années.

2.3 Modèles spectraux

La représentation spectrale des sons provient, à l'origine, de la synthèse additive. Cette méthode consiste à décomposer le signal en formes d'ondes élémentaires. L'étude de plusieurs formes d'ondes a mené à des modèles comme par exemple les ondes carrées pour la synthèse dite de Walsh [Wal23]. D'autres approches utilisent des petites ondes limitées dans le temps. En les étirant dans le domaine du temps et des fréquences, une famille d'ondes est définie, servant de base à la transformée par ondelettes [KM88]. Même si les applications de ces modèles par ondelettes montrent des résultats probants sur le plan de la qualité [Min98], les transformations permises sont effectuées empiriquement en modifiant des coefficients mathématiques, dont le lien avec des paramètres perceptifs n'a pas été formellement établi.

La forme d'onde la plus couramment utilisée reste cependant les sinusoïdes. En effet, elles respectent des règles d'orthogonalité et peuvent être reliées à la perception de hauteur. Les modèles spectraux sont liés aux périodicités du signal sonore et se placent au niveau de la membrane basilaire du système auditif. Ils paraissent donc plus proches de la perception que les modèles temporels. La transformée de Fourier est l'outil mathématique adéquat pour se placer dans le domaine fréquentiel.

2.3.1 Transformée de Fourier à court terme et vocodeur de phase

La transformée de Fourier (voir la section 1.1.5) permet de décomposer un signal en une somme de sinusoïdes. Les fréquences, les phases et les amplitudes de ces sinusoïdes sont les résultats de cette transformation. Celle-ci permet donc de passer de la représentation temporelle d'un son à sa représentation spectrale, en indiquant les variations d'amplitude en fonction des fréquences, pour définir le spectre d'amplitude. Pour les signaux numériques, la transformée de Fourier discrète (DFT) est employée. Un algorithme rapide (*Fast Fourier Transform*) permet son application dans un temps limité. Comme la transformée de Fourier repose sur une hypothèse de stationnarité du signal, plusieurs petits segments temporels (appelés fenêtres) de sons sont successivement étudiés : c'est le principe de la transformée de Fourier à court terme.

Le vocodeur de phase est un modèle développé à partir de la transformée de Fourier à court terme [FG66]. Il représente le son par une succession de spectres (d'amplitude et de phase) calculés en appliquant la transformée discrète de Fourier à des fenêtres temporelles du son étudié, pouvant éventuellement se chevaucher. De nombreuses implémentations basées sur la transformée rapide de Fourier sont disponibles [Por76]. Ce modèle a de nombreuses applications [Dol86], notamment du point de vue musical [Moo78]. Les plus connues sont les étirements et compressions dans le temps et les changements de hauteur [Ser97a], qui peuvent être théoriquement effectués indépendamment. Ces changements sont faits en modifiant le taux de chevauchement (le temps d'avancement entre les fenêtres temporelles successives de signal) avant d'éventuellement ré-échantillonner le signal. Il est nécessaire toutefois de recalculer les phases des sinusoïdes pour chaque fenêtre, de façon à ne pas créer d'artefact audible dû à la rupture d'un partiel. Ce calcul complique les transformations, notamment dans le cas de variations de l'enveloppe spectrale. Des améliorations ont été envisagées [LD99], qui proposent de ne recalculer les phases que pour les fréquences correspondant aux maxima locaux du spectre d'amplitude, et d'appliquer ces modifications de phase aux voisins de ces maxima de la même façon.

Les limites du vocodeur de phase sont donc liées aux transformations sur l'échelle du temps et de la hauteur. Le fait d'appliquer directement la transformée de Fourier implique l'utilisation de paramètres mathématiques (taille de la fenêtre, forme de la fenêtre, temps d'avancement, etc., voir section 3.4) qui sont essentiels à la qualité de la représentation, mais qui ne sont pas intuitifs sur le plan musical. De plus, bien que des améliorations existent, le vocodeur de phase n'est destiné qu'à un certain type de sons notamment ceux dont les partiels ne varient pas trop au cours du temps. Ainsi, un son harmonique contenant un vibrato important risque d'être resynthétisé avec des artefacts, puisque les fréquences variant dans le temps risquent de dépasser la résolution fréquentielle imposée par la transformée discrète de Fourier. Si cette technique est appliquée aux sons bruités, la resynthèse fait perdre au son original toutes ses caractéristiques de *mouvements spectraux* pour donner un son *métallique*.

Cette technique est donc relativement simple et assez rapide, ce qui a rendu son utilisation très populaire. A présent, compte tenu des nombreuses améliorations, le vocodeur de phase est surtout utilisé pour illustrer les limites de la transformée discrète de Fourier.

2.3.2 Synthèse additive

La synthèse additive consiste à additionner plusieurs signaux simples afin de générer un signal plus complexe. Chaque signal simple composant le son résultant donne une contribution à la qualité de rendu. Bien évidemment, cela reste un terme général qui regroupe un grand nombre de techniques différentes. La plus classique reste l'addition de sinusoïdes pures, de fréquence, d'amplitude et de phase constantes (partiels). Mais nous allons voir dans la suite que des signaux plus complexes, comme par exemple des bruits, peuvent être ajoutés pour rendre un son synthétisé plus réaliste.

2.3.3 Modèles hybrides

Principe

A l'origine, les modèles spectraux ont été conçus pour l'analyse et la synthèse de la parole. Or, la parole peut être correctement perçue même si le bruit contenu dans le signal n'est pas bien représenté et synthétisé. C'est pourquoi ces modèles reposent sur des hypothèses de stationnarité et d'absence de bruit qui ne peuvent être vérifiées dans le cas de certains sons instrumentaux ou d'autres sons réels. L'application de ces méthodes d'analyse et de synthèse mène dans ces cas à des sons *artificiels*, qui demeurent éloignés sur le plan de la perception des sons originaux. C'est pour résoudre ces limites que les modèles hybrides sont apparus à la fin des années 80. Ces modèles représentent le signal analysé par plusieurs composantes, dont les méthodes d'analyse, de transformations et de synthèse sont différenciées et adaptées à la spécificité de chacune de ces composantes.

Le premier modèle du genre fût le modèle SMS [SS90, Ser89, Ser97b] (*Spectral Modeling Synthesis*) qui décompose le signal, noté x, en une partie déterministe, notée d, et une partie stochastique, notée s :

$$\forall n, \quad x[n] = d[n] + s[n] \tag{2.2}$$

La partie déterministe correspond aux sinusoïdes du signal dont les amplitudes et les fréquences varient lentement au cours du temps, c'est-à-dire aux partiels. Elle correspond donc aux approches classiques utilisées dans la modélisation de la parole. L'originalité de ce modèle SMS réside dans le fait d'ajouter une partie stochastique qui correspond au bruit contenu dans le signal. En pratique, il correspond au résidu du signal après avoir retiré l'énergie correspondant à la partie déterministe.

Une approche semblable a, depuis, été mise en application pour l'amélioration du traitement de la parole [Sty98, SDS98]. La comparaison avec une méthode classique comme PSOLA permet de mettre en avant l'intérêt de l'ajout d'une composante bruitée sur la qualité sonore [SSG+98]. Les méthodes d'analyse et de synthèse sont différentes de celles de la méthode SMS, puisqu'elles sont adaptées aux propriétés particulières de la parole.

Le modèle SMS a été amélioré en décomposant la partie stochastique de façon à prendre en compte les transitoires (attaques), notées t. En effet, les transitoires ne peuvent être modélisées ni par des sinusoïdes, sans prendre en compte les informations liées aux phases, ni par du bruit. Ainsi, le signal original est partagé en trois parties :

$$\forall n, \quad x[n] = d[n] + t[n] + s[n] \tag{2.3}$$

Ces approches hybrides permettent donc d'enrichir le son en proposant des traitements spécifiques à chacune de ses parties. Les résultats sonores sont sans conteste probants et montrent l'amélioration de qualité. Toutefois cette décomposition implique un traitement indépendant des parties. En effet, du fait des modèles différents pour chaque partie, il est nécessaire pour le musicien de prendre en compte ces différences et d'être capable de manipuler le son en différentes étapes.

Certaines approches reprennent l'idée de l'ajout de bruit au son composé de partiels, mais en essayant de développer un modèle dit homogène, car il ne comporte qu'un seul type de paramètre [FHC00]. Nous abordons également ce modèle dans la suite.

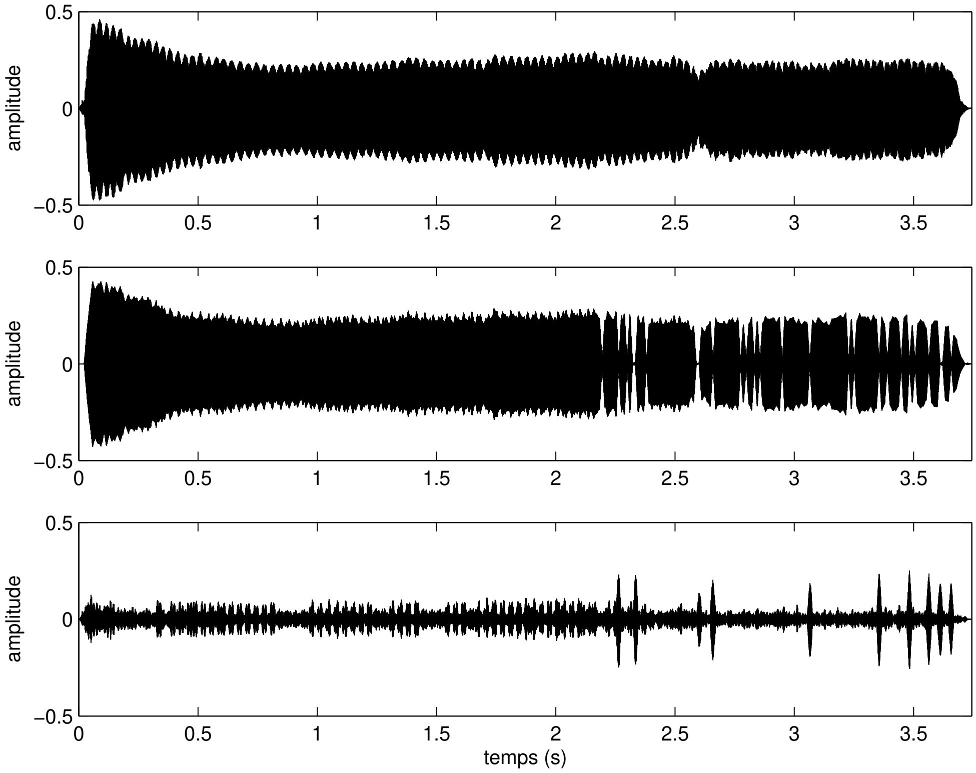

Fig. 2.6 – Exemple de la décomposition d'un son de flûte (figure du haut) en une partie déterministe définie par des partiels (figure du milieu) et une partie stochastique (résidu), définie par soustraction spectrale (figure du bas).

Transformations possibles

En comparant aavec les modèles sinusoïdaux, les transformations d'échelle temporelle et de hauteur sont de meilleure qualité par cette approche hybride, puisqu'une méthode de synthèse adéquate est appliquée à chaque composante du signal original. Ainsi, les distortions pouvant être introduites par une méthode de synthèse générale peuvent être évitées (par exemple l'utilisation de partiels pour synthétiser du bruit).

Les autres transformations possibles, propres aux modèles hybrides, sont liées à la différence de traitement possible entre les différentes parties du signal. Ainsi, il est permis de modifier la partie stochastique, tout en préservant la partie déterministe, ce qui revient à bruiter le signal original (l'inverse est également envisageable, ce qui reviendrait à enlever les bruits). Des synthèses croisées sont aussi permises, en synthétisant un son dont les parties déterministes et stochastiques proviennent par exemple de deux sons différents.

2.3.4 Modélisation de la partie déterministe

Les modèles hybrides considèrent une partie déterministe comme une partie constituée des partiels détectés dans le signal analysé. Ce modèle sinusoïdal, devenu maintenant très classique et à la base de nombreux modèles spectraux, a été amélioré ces dernières années. Nous présentons ici le modèle de base et une amélioration récente de ce modèle, avec les transformations musicales qu'ils autorisent.

Modèles sinusoïdaux

Le modèle sinusoïdal est une amélioration du vocodeur de phase. Il a été introduit au milieu des années 80 par plusieurs travaux effectués au même moment. Les références restent néanmoins le modèle de McAulay et Quatieri [MQ86], destiné à la modélisation de la parole, et les améliorations apportées par Serra et Smith [SS87]. La base de ces modèles est la transformée de Fourier à court terme. Le signal est donc considéré comme une somme de N partiels, d'amplitude a_n, de fréquence f_n et de phase ϕ_n :

$$x(t) = \sum_{n=1}^{N} a_n \sin(\phi_n(t)) \tag{2.4}$$

$$\frac{d\phi_n(t)}{dt}(t) = 2\pi f_n(t) \tag{2.5}$$

Mais, à la différence du vocodeur de phase, toutes les valeurs de spectre discret résultant ne sont pas prises en compte. Seuls les maxima locaux ou pics (valeur de fréquence dont les amplitudes sont supérieures aux deux valeurs voisines) sont pris en considération et supposés être représentatifs des partiels présents dans le signal.

En se déplaçant dans le temps, d'un spectre à court terme à un autre, les partiels présents sont reliés entre eux pour considérer des partiels de durée plus importante que la durée correspondant à la longueur W_s de la fenêtre de la transformée de Fourier. Les fréquences et les amplitudes sont sauvegardées lors de l'analyse, mais les phases ne sont pas prises en compte. C'est seulement lors de la phase de synthèse que les phases des partiels sont calculées de façon à lier les sinusoïdes synthétisées dans deux fenêtres consécutives :

$$\phi_n(t_0 + t) = \phi_n(t_0) + 2\pi \int_{t_0}^{t} f_n(u)du \tag{2.6}$$

Dans le cas d'une fréquence fixe, cette relation devient dans le cadre discret (fenêtre i) :

$$\phi_n^{i+1} = \phi_n^i + 2\pi W_s f_n \tag{2.7}$$

En pratique, comme les fréquences et les amplitudes évoluent d'une fenêtre sur l'autre (même si elles peuvent se chevaucher), des heuristiques différentes sont employées pour décider si deux pics correspondent au même partiel : ce sont les problèmes classiques de suivi de partiels (*partial tracking*). La méthode originale de McAulay et Quatieri propose de comparer les fréquences d'une fenêtre sur l'autre. Si la différence est inférieure à un seuil préalablement fixé, les deux fréquences sont rattachées à un même partiel. D'autres approches ont été proposées depuis pour améliorer cette technique en utilisant les informations des amplitudes, des fréquences et des phases. Nous ne les détaillerons pas ici, mais nous pouvons mentionner par exemple les approches statistiques par modèle de Markov [DGR93], une approche par prédiction linéaire [LMRR03], ou enfin une analyse de la distortion du spectre de phase [PR99]. La définition d'états “zombies” permet également d'éviter des coupures de partiels lorsque la phase de détection des pics n'a pas été correcte [Ser97b].

Cette méthode de modélisation sinusoïdale a été implémentée dans de nombreux programmes d'analyse/synthèse de sons. Nous pouvons citer par exemple `Lemur` [FH96] du *CERL Sound Group* disponible en *shareware* [FHH+], les logiciels `InSpect`, `ProSpect` et `ReSpect` du *SCRIME* [MS99] téléchargeables librement [Mar], et enfin un programme de test `Matlab` développé par Dan Ellis et téléchargeable librement [Ell]. C'est ce dernier programme que nous avons utilisé pour effectuer quelques tests, illustrés par la figure 2.7.

Modèle additif structuré (SAS)

Le modèle additif structuré appelé SAS (*Structured Additive Synthesis*) est un modèle spectral, basé sur la modélisation sinusoïdale [Mar00]. Il définit quatre paramètres physiques proches de la perception, d'après les travaux de description du timbre effectués par des musiciens [DCM99a] :

l'amplitude est une fonction du temps et est relative au volume du son modélisé. Mathématiquement, il est relié à l'amplitude RMS du signal, comme nous l'avons vu dans la section 1.5.1.

la fréquence représente la hauteur du signal sonore.

la couleur est une interpolation (version lissée) de l'enveloppe spectrale.

l'inharmonicité (*warping*) est une fonction qui définit l'écart des fréquences des harmoniques à leurs valeurs idéales d'harmonicité.

Le principe de ce modèle reste proche de la modélisation sinusoïdale dans le sens qu'il représente le signal comme une somme de partiels. Les paramètres du modèle sont définis à partir de ces partiels. L'intérêt du modèle réside dans le fait que l'utilisateur (musicien) n'a pas besoin de manipuler un par un des partiels pour modifier les propriétés du son, mais peut manipuler directement les quatre paramètres perceptifs à un plus haut niveau.

Limites

Ces modèles sinusoïdaux, qu'ils soient ou non structurés présentent des limites liées à la définition même du modèle. Définir un son par une somme de sinusoïdes dont les amplitudes et les fréquences varient lentement dans le temps implique une restriction sur la catégorie de sons à considérer. Les hypothèses de base sont l'absence de réverbération et l'unicité de la source dans le cas du modèle SAS. En effet, la présence de bruits ou de transitoires perturbe non seulement la phase d'analyse qui peut considérer des pics dans le spectre qui ne correspondent pas à un partiel correct (voir le chapitre 3) et ainsi entraîner des mauvaises connections, et peut aussi représenter l'énergie correspondant au bruit ou aux transitoires par des partiels. Le résultat de synthèse d'un son bruité par un modèle sinusoïdal n'est souvent pas satisfaisant, donnant une impression "métallique" et les attaques sont très atténuées mais multiples (sensation d'écho). De plus les transformations telles que le changement d'échelle temporelle risque d'amplifier ces défauts sonores. La figure 2.7 montre qu'un son de hautbois est bien représenté par des fréquences stables dans le temps, alors que des trajectoires de partiels semblent plus chaotiques dans le cas du son de trompette, car ce son est composé des perturbations d'airs propres à certains instruments à vent. De nouveaux partiels dûs au bruit sont définis et pris en compte dans le modèle.

Néanmoins, comme les résultats sont très satisfaisants pour les sons peu bruités sans transitoires, cette représentation sinusoïdale a été améliorée en traitant différemment le bruit, puis les transitoires dans les modèles hybrides.

Transformations possibles

Lorsque ce modèle est appliqué à des sons peu bruités, sans transitoire, de nombreuses transformations musicales sont permises comme le filtrage à partir du spectre, les changements d'échelle d'amplitude ou de temps. Ces transformations temporelles (allongement ou réduction de la durée du son) sont de meilleure qualité, en comparaison avec le modèle du vocodeur de phase. En effet, il suffit pour cela de synthétiser les partiels sur une plus longue période de temps. Cet avantage est du à l'absence de prise en compte de la phase à l'analyse et au calcul de leurs valeurs seulement à la synthèse.

Concernant le changement de hauteur, le modèle SAS permet d'effectuer des transpositions indépendamment du timbre, ou du moins de l'enveloppe spectrale et donc des formants [Mar99]. Avec le modèle sinusoïdal, le changement de hauteur est effectué en modifiant l'échelle des fréquences du spectre. L'enveloppe spectrale est alors également modifiée, et les formants sont notamment translatés. En passant par un modèle structuré comme le modèle SAS, les valeurs des fréquences des partiels sont certes modifiées, mais leurs amplitudes sont affectées à partir du paramètre couleur qui, lui, peut rester constant.

2.3.5 Modélisation de la partie stochastique

Les modèles sinusoïdaux sont restreints aux sons peu bruités et sans transitoire. Les premiers modèles hybrides ont vu leur origine dans l'intérêt de définir un modèle de bruit pour pouvoir améliorer le réalisme des sons produits.

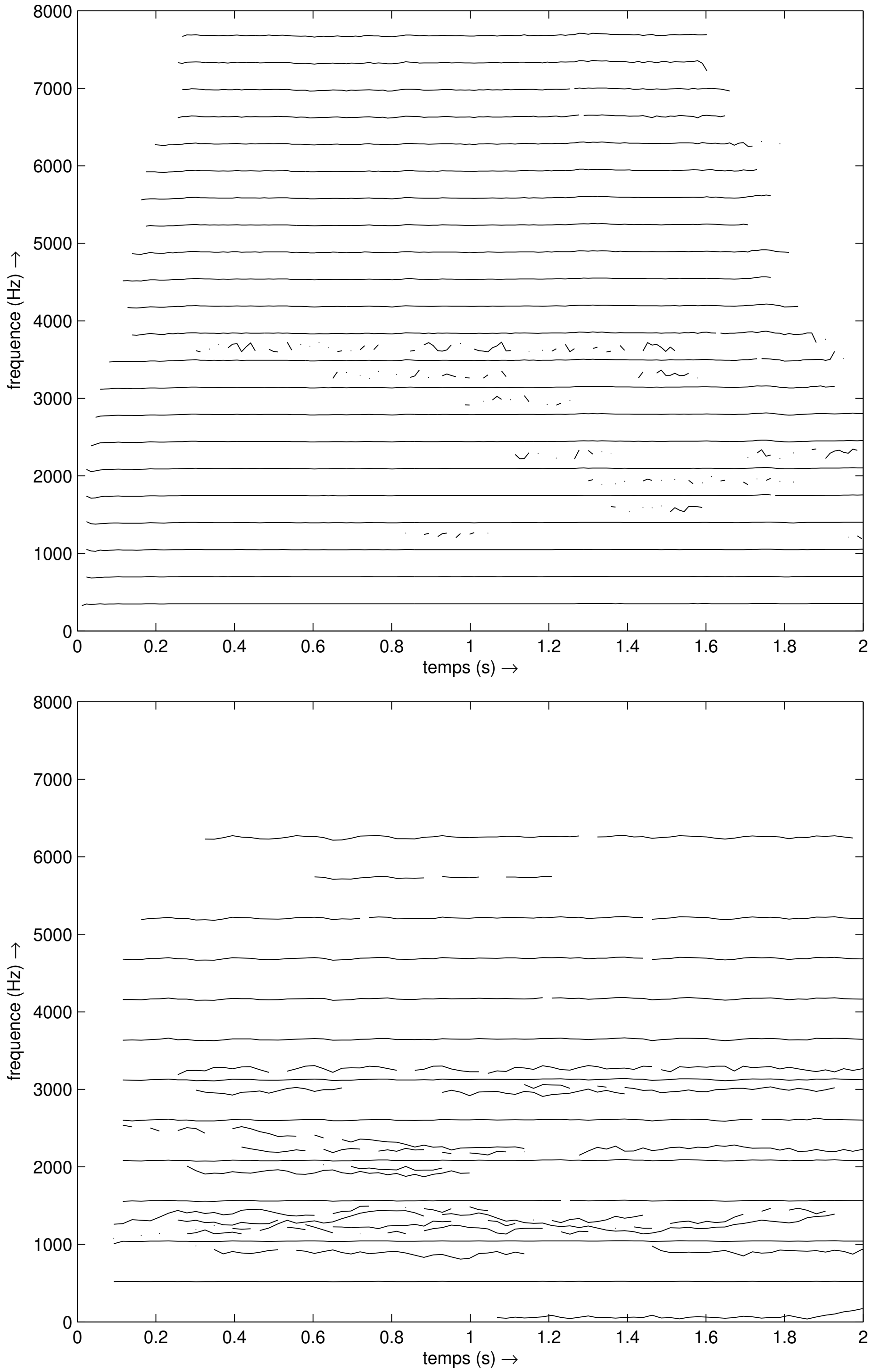

Fig. 2.7 – Évolutions des partiels pour des sons de hautbois (figure du haut) et de trompette (figure du bas) selon le modèle sinusoïdal de type McAulay et Quatieri (fenêtres de 1024 échantillons) : le bruit contenu dans le son de la trompette est représenté par quelques partiels apparaissant entre les premières harmoniques.

Modèle SMS (bruit blanc filtré)

Le modèle SMS décompose le signal x en une partie déterministe d et une partie stochastique s. La partie déterministe suit le modèle sinusoïdal décrit ci-dessus. La partie stochastique est obtenue par soustraction entre le signal original et les partiels analysés. La figure 2.8 illustre le principe général de la méthode d'analyse pour ce modèle. Ce principe explique pourquoi le terme résidu est souvent employé. Cette soustraction est spectrale pour rendre cette opération indépendante des phases des partiels [SS90]. La précision de l'analyse des partiels définit donc l'exactitude de la partie stochastique. En notant X, D et S respectivement les spectres du signal, de la partie déterministe et de la partie stochastique, nous pouvons écrire :

$$\forall f, \quad |S(f)| = ||X(f)| - |D(f)|| \tag{2.8}$$

L'hypothèse de base concernant la partie stochastique dans le modèle SMS est qu'elle est entièrement définie par sa couleur, c'est-à-dire son enveloppe spectrale lissée. Cette hypothèse n'est d'ailleurs pas vraiment justifiée dans la littérature, si ce n'est par le caractère supposé stochastique de cette partie. Plusieurs méthodes peuvent être employées pour représenter l'enveloppe spectrale de la partie stochastique. Nous les présentons dans la section 3.5.

Des expériences menées à partir des logiciels implémentant cette représentation [Ser97c] montrent les limites de ce modèle pour les sons bruités. La soustraction spectrale s'avère notamment une source d'erreur assez importante. Par exemple, l'analyse d'une sinusoïde pure mélangée à du bruit blanc fournit une partie stochastique laissant entendre par moment la hauteur correspondant à cette sinusoïde. Ce phénomène peut s'expliquer par les erreurs produites par les méthodes d'analyse des paramètres des partiels, qui semblent inévitables même si de nombreuses améliorations ont été apportées notamment en ce qui concerne l'estimation des fréquences et des amplitudes [KZ01, AF95, DCM00], la sélection de pics [KM02, PR98b, HDC03d] et le suivi des partiels [DGR93, LMRR03, PR99]. Le fait de définir la partie stochastique à partir d'une estimation de la partie déterministe implique que toutes les erreurs induites lors de la définition des partiels seront ajoutées dans la partie stochastique.

Modélisation par bandes de Barks

Plusieurs modèles de bruits ont été proposés dans le domaine de la compression audio [Lev98]. Ces modèles ont la même approche que le modèle SMS dans le sens que le bruit est assimilé au résidu du signal. Il est donc défini par soustraction entre le signal original et l'énergie représentée par les partiels du signal. La différence entre ces modèles provient de la représentation de l'énergie contenue dans le spectre. Le modèle SMS retient et extrait l'enveloppe spectrale lissée, alors que dans les modèles proposés par Goodwin [Goo96, Goo97], le résidu est représenté par la quantité d'énergie contenue dans chaque bande Bark. Cette approche repose sur l'hypothèse que l'oreille n'est pas sensible aux variations d'énergie à l'intérieur de chaque bande Bark, mais seulement à leur quantité totale. Les bandes Barks sont définies dans la section 1.5.3. La compression pour une telle approche est particulièrement intéressante puisque seulement 25 bandes Barks sont définies sur l'intervalle des fréquences audibles (jusqu'à 20000Hz).

Pour chaque fenêtre d'analyse, nous considérons B_k casiers (*bins*), ou valeurs de fréquence du spectre discret S, dans une bande Barks k et nous notons $\mathcal{F}_k$ l'ensemble des fréquences

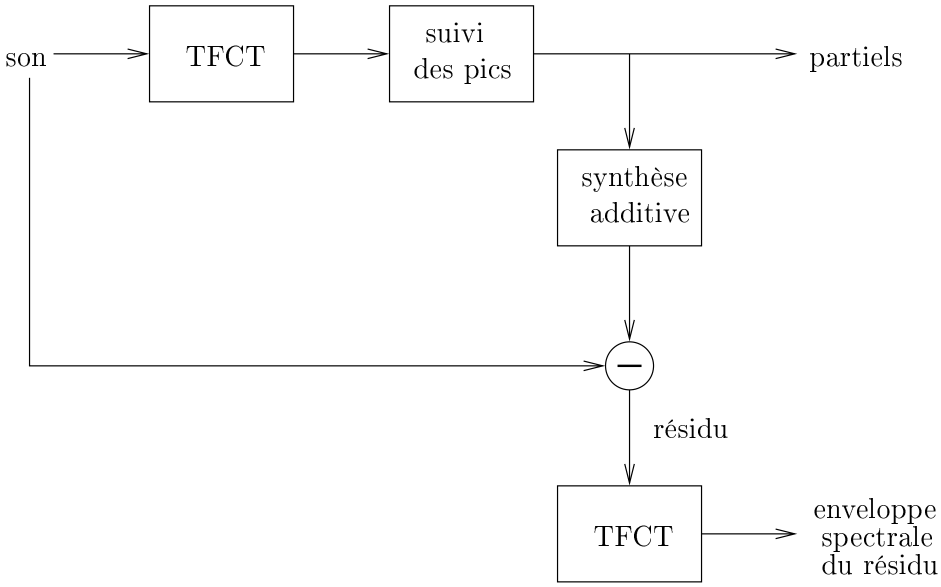

FIG. 2.8 – Schéma général de l'analyse pour le modèle hybride SMS (sinusoïdes + bruit).

appartenant à cet ensemble. Ainsi, nous pouvons écrire la relation entre la valeur de l'amplitude S' pour chaque fréquence d'une même bande en fonction des valeurs de fréquences initiales f :

$$\forall f \in \mathcal{F}_k, \quad S'(f) = \sqrt{\frac{1}{B_k} \sum_{f \in \mathcal{F}_k} S^2(f)} \tag{2.9}$$

Cette méthode est décrite comme réaliste pour la resynthèse des sons, notamment avec l'impression que le résidu fusionne mieux avec la partie sinusoïdale [Lev98]. Ce sentiment n'est certainement plus le même lorsque des transformations, comme un allongement dans le temps par exemple, sont effectuées.

Malgré les différences avec le modèle SMS, cette approche reste basée sur l'approximation de l'enveloppe spectrale. Elle effectue un échantillonnage non uniforme de l'énergie en se basant sur une échelle des fréquences non linéaire. Ces approximations ne peuvent être vérifiées que dans le cas de sons dont l'énergie liée au bruit reste très faible.

Élargissement des pics spectraux

Certains travaux se sont efforcés de représenter les sons d'une façon plus homogène que les modèles hybrides ne le proposent, en argumentant un manque de fusion entre les différentes parties lors des transformations, surtout dans le cas de sons instrumentaux [Fre99]. Une autre idée envisage donc de se baser sur le modèle sinusoïdal, mais en ajoutant du bruit directement à chaque partiel. On parle de modèle additif avec largeur de bande adaptée (*bandwidth enhanced additive model*) [FHC00]. Un seul type de composant est considéré, défini comme un partiel (de pulsation ω, d'amplitude a) autour duquel de l'énergie est ajoutée aléatoirement dans le spectre. Cet oscillateur particulier peut s'écrire [Fit99] :

$$y[n] = a(\sqrt{1-\mathcal{K}} + \sqrt{2\mathcal{K}}(X[n] * h[n])) \exp(j\omega n) \tag{2.10}$$

où $\mathcal{K}$ représente la proportion d'énergie du partiel associée au bruit, $X[n]$ est un bruit blanc et $h[n]$ est la réponse impulsionnelle d'un filtre passe-bas. Lorsque $\mathcal{K}$ est nul, l'oscillateur représente une sinusoïde pure, alors que lorsque $\mathcal{K}$ vaut 1, il représente un bruit. Lors de la phase d'analyse, l'énergie est donc répartie autour du partiel en modifiant cette valeur $\mathcal{K}$.

Paul Masri propose également une ébauche d'un nouveau modèle à partir d'un partiel bruité (*noisy partial*), défini comme une sinusoïde dont la fréquence est modulée par du bruit [Mas96].

Ce modèle est original car homogène puisqu'il ne propose qu'un seul type de composants spectraux. Toutefois, concernant la partie bruit d'un signal, il propose la même approche que pour le modèle SMS, c'est-à-dire le filtrage d'un bruit blanc. Une version logicielle libre est disponible [FHC].

Transformations possibles

Les modèles que nous venons de présenter ne proposent finalement qu'un seul paramètre pour représenter la partie stochastique du signal : l'enveloppe spectrale lissée. Ainsi, tous ces modèles représentent un bruit naturel par un bruit blanc filtré et les seules transformations musicales permises seront donc liées à la modification de la couleur et de l'échelle de temps. La complexité et la multitude de sons naturels existants laissent penser qu'un seul paramètre de représentation est insuffisant pour décrire toutes les propriétés musicales des bruits naturels.

A l'aide de ces représentations, un son peut être allongé et raccourci temporellement tout en préservant, pour leurs parties stochastiques, les caractéristiques propres aux bruits. Ces opérations sont effectuées en modifiant le taux de recouvrement des fenêtres de synthèse dans le cas du modèle SMS [SS90], et en augmentant la durée des oscillateurs bruités dans le cas du modèle des pics élargis, suivant le principe de la modélisation sinusoïdale. Toutefois, en ce qui concerne le modèle SMS, nous avons mis en valeur un problème théorique dans l'application de cette méthode d'allongement, lié au chevauchement des fenêtres de synthèse dans le cas d'un signal aléatoire (comme la partie stochastique). Nous évoquons ce problème et les solutions proposées, dans la section 4.5.

2.3.6 Modélisation de transitoires

Les attaques ou transitoires contenues dans les sons sont reconnues comme étant mal représentées par des partiels, car les variations d'énergies sont trop importantes sur une durée très courte, pour pouvoir être correctement représentées par des sinusoïdes. De même, le fait d'analyser la partie stochastique à l'aide d'une transformée de Fourier effectuée à partir de fenêtres successives se chevauchant, mène fréquemment à une atténuation des attaques, mais aussi à une multiplication de leurs reproduction lors de la synthèse (on parle de *pré-écho*) car plusieurs fenêtres d'analyse et de synthèse contiennent ce saut d'énergie. De plus, lors de transformations telles qu'un étirement dans le temps, les transitoires ne peuvent être traitées de la même manière que les partiels ou le bruit. C'est à cause de ces particularités que des travaux ont porté et portent toujours sur la modélisation de ces transitoires localisées dans le temps, dans le but d'être incorporées dans des modèles hybrides.

Les techniques de détection de transitoires sont décrites dans la section 3.3. Concernant les modèles, l'exemple le plus abouti est certainement le modèle TMS (*Transient Modeling*

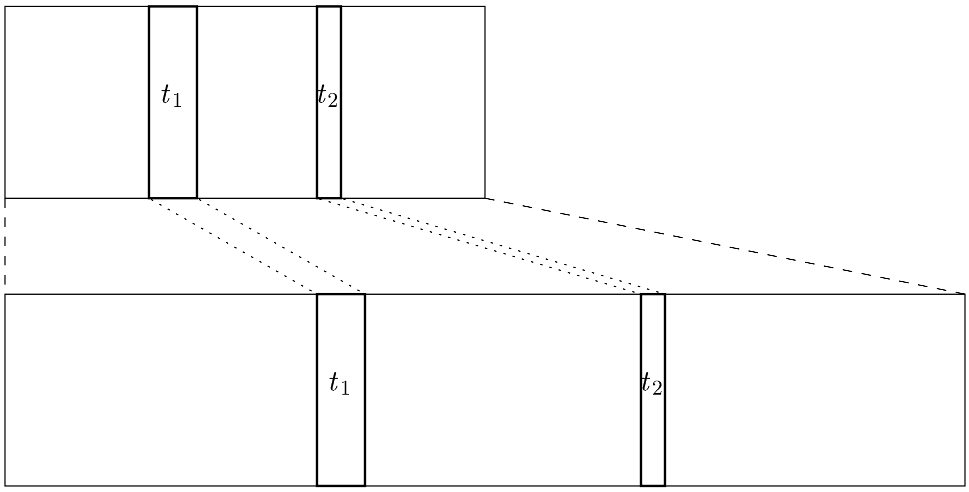

FIG. 2.9 – Principe de l'allongement d'un son contenant des transitoires : les segments temporels, notés t_1 et t_2, contenant ces transitoires sont préservés alors que le reste du signal (sinusoïdes et bruit) est dilaté.

Synthesis) [VLM99]. Il repose sur la propriété qu'une transitoire (une impulsion) dans le domaine temporel correspond à une sinusoïde variant lentement dans le domaine spectral. Ainsi, une analyse sinusoïdale est effectuée à partir d'une représentation fréquentielle et permet à la fois de repérer et de modéliser les transitoires d'un signal.

Ce modèle de transitoire a ensuite été incorporé dans le modèle classique SMS de type sinusoïde et bruit pour donner le modèle STN (*Sines+Transients+Noise*) [VM00]. La figure 2.10 illustre le principe général de la méthode d'analyse pour ce modèle. Il permet d'effectuer des modifications d'échelle temporelle en traitant différemment les transitoires. Les fréquences des sinusoïdes dans le domaine fréquentiel sont modifiées, ce qui déplace l'instant de la transitoire à l'intérieur d'une fenêtre d'analyse, et la durée de ces sinusoïdes, toujours dans le domaine fréquentiel, est modifiée pour augmenter la taille de la fenêtre d'analyse temporelle contenant les transitoires [VM98]. A notre connaissance, il n'existe toutefois pas d'implémentation libre de ce modèle.

D'autres approches pour représenter mathématiquement les transitoires ont été proposées. Nous pouvons citer entre autres l'utilisation des fonctions de Prony (sinusoïde dont l'enveloppe décroît de façon exponentielle) [TG01], la détection de zones dans les spectres à court terme [RJ01] et la détection dans le temps de façon à synchroniser les fenêtres d'analyse et de synthèse pour préserver les transitoires [MB96].

Même si nous n'avons pas pu tester tous les modèles de transitoires que nous citons, il apparaît que ces approches restent essentiellement destinées à compléter les modèles hybrides de type SMS, pour prendre en compte les attaques de sons harmoniques. Ainsi, ces modèles n'introduisent pas de paramètre de contrôle musical. Aucune transformation, spécifique aux transitoires, n'a été proposée. Ces modèles sont essentiellement utiles à la prise en compte des spécificités des transitoires lors des opérations d'étirement ou de compression temporelle. En effet, les attaques ne peuvent être, par définition, allongées. La figure 2.9 montre le principe de la prise en compte des transitoires lors de l'allongement d'un son.

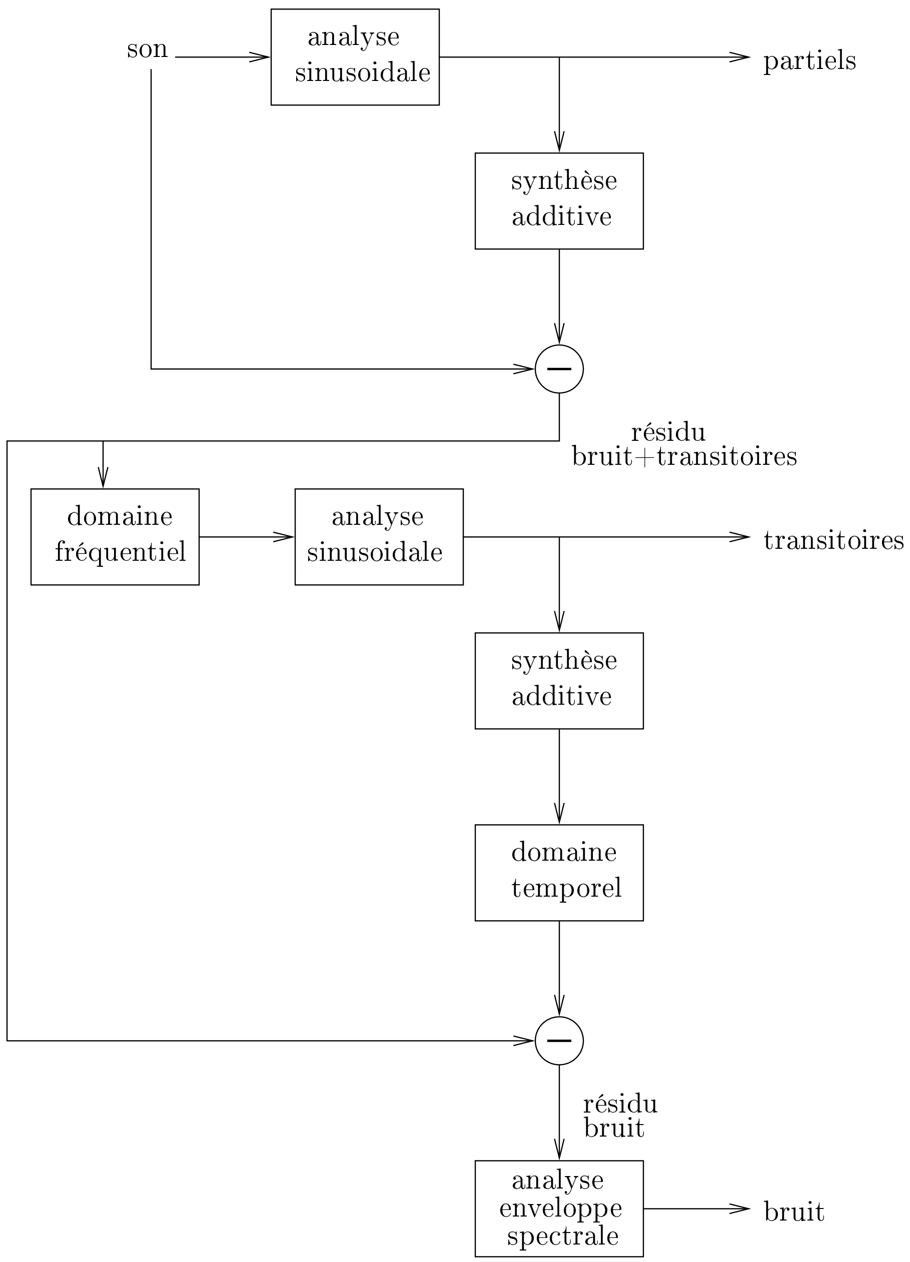

FIG. 2.10 – Schéma général de l'analyse du modèle hybride STN (sinusoïdes + transitoires + bruit).

2.4 Modèle CNSS

Nous avons vu dans les sections précédentes les modèles envisageables pour la synthèse de sons. Parmi ceux-ci, les modèles spectraux semblent les plus adaptés à la manipulation musicale. Cependant, les approches existantes restent essentiellement destinées à la représentation de sons peu bruités. L'application de ces techniques d'analyse et synthèse à des bruits laisse apparaître des distortions dues à la faillite des hypothèses de stationnarité et préconise une adaptation spécifique pour pouvoir envisager la modélisation de sons complexes aléatoires.

Nous proposons dans cette section un modèle original pour les sons bruités (ou les parties stochastiques des modèles hybrides de type SMS), basé sur une représentation spectrale et statistique, et permettant d'effectuer de nouvelles transformations musicales propres à ce type particulier de sons. Dans un premier temps, nous justifions le choix d'un modèle spectral et statistique. Nous présentons également les travaux menés sur la perception de la densité spectrale, paramètre auditif propre aux signaux aléatoires complexes. Enfin, après avoir rappelé une première approche expérimentale, nous proposons les paramètres de ce nouveau modèle CNSS.

2.4.1 Modèle spectral

Les modèles sinusoïdaux sont à présent bien avancés et très performants pour des sons peu bruités. Ils proposent des contrôles musicaux intuitifs. Il paraît donc intéressant de partir de cette base de modèles pour pouvoir représenter des sons plus complexes. Cette démarche a donc conduit aux modèles hybrides. L'étude de la représentation de la partie bruitée des sons par ces modèles nous amène cependant à supposer qu'il est possible d'envisager un modèle proposant plus de paramètres de contrôle musical, puisque les approches existantes n'en proposent pour l'instant qu'un seul : la couleur, sous la forme de l'enveloppe spectrale lissée. Pour cela, nous nous sommes intéressés aux paramètres perceptifs étudiés par les psychoacousticiens et concernant les bruits. La densité spectrale paraît ainsi utile pour décrire les propriétés auditives d'un bruit. Cette densité spectrale, définie dans la suite, peut être contrôlée en utilisant une approche spectrale. Et comme la couleur peut également être intuitivement représentée à l'aide d'une approche spectrale, il nous semble opportun de nous concentrer sur le domaine fréquentiel et sur la possibilité de modéliser un bruit par une somme finie de sinusoïdes. L'avantage de cette approche est donc double. Il nous permet aussi d'envisager une extension d'un modèle sinusoïdal performant pour définir un modèle spectral hybride. Mais il doit permettre de contrôler indépendamment l'enveloppe spectrale (la couleur) et la densité spectrale, et ainsi proposer aux musiciens de nouvelles transformations originales.

Toutefois, il n'apparaît pas évident, de prime abord, de pouvoir justifier la représentation mathématique d'un signal aléatoire par une somme finie de sinusoïdes de fréquence, de phase et d'amplitude fixées. Heureusement la physique des particules a déjà envisagé dans le passé une telle approche.

2.4.2 Modèle du bruit thermique

A température non nulle, les électrons sont soumis à des mouvements aléatoires dans un conducteur. Ces mouvements produisent une tension, par conséquence aléatoire elle aussi, aux bornes de ce conducteur, et définissent ainsi le bruit thermique. Ce bruit, noté X, a été décrit par Einstein et Hopf en 1910 [Har97], par la relation :

$$X(t) = \sum_{n=1}^{N} C_n \sin(\omega_n t + \Phi_n) \tag{2.11}$$

où N est le nombre de fréquences, n est un entier, ω_n sont les pulsations régulièrement espacées, C_n sont des amplitudes aléatoires, distribuées selon une loi de Rayleigh, et Φ_n sont des phases aléatoires, uniformément distribuées entre 0 et 2π. Dans ce cas, le bruit X voit ses échantillons suivre une distribution de Gauss.

Cette définition représente donc un bruit par une somme finie de sinusoïdes. Elle permet de conforter l'idée d'un modèle spectral pour représenter les bruits naturels et constitue donc la base du modèle de sons aléatoires développé dans la suite. Toutefois, ce modèle ne précise ni le nombre de sinusoïdes nécessaires, ni l'écart entre les fréquences du signal. Pourtant choisir $N = 2$ sinusoïdes sur une bande de fréquence de largeur 20kHz ne permet pas de synthétiser un bruit équivalent du point de vue de la perception à un bruit blanc synthétisé par un tirage aléatoire des échantillons selon une loi gaussienne. Même si la distribution des échantillons a rapidement tendance à suivre une loi de Gauss à cause du théorème central limite, il semble que la perception soit sensible (jusqu'à un certain point) à ce nombre N, par rapport à la largeur de la bande de bruit considérée. L'étude de cette perception est présentée dans la section suivante.

Il est important de noter que la représentation de la partie stochastique dans les modèles hybrides, tels que le modèle SMS, est basée implicitement sur le modèle du bruit thermique. La seule différence provient du caractère déterministe. En effet, la resynthèse (détaillée en section 4.2) est effectuée à partir de l'enveloppe spectrale en définissant un spectre, par la distribution aléatoire des valeurs des phases selon une loi uniforme, avant d'effectuer une transformée inverse de Fourier. Or cette opération mathématique consiste à sommer un nombre fixe de sinusoïdes dont les fréquences sont également espacées, et dont les amplitudes sont fixées. En notant F_s la fréquence d'échantillonnage et W_s la taille de la fenêtre de synthèse, l'écart entre les fréquences est $\frac{F_s}{W_s}$, et le nombre de sinusoïdes vaut $\frac{W_s}{2}$ (nombre de fréquences dans le spectre audible). Ce modèle de synthèse peut donc s'écrire :

$$X(t) = \sum_{n=1}^{\frac{W_s}{2}} c_n \sin(2\pi n \frac{F_s}{W_s} t + \Phi_n) \tag{2.12}$$

Nous pouvons néanmoins légitimement nous poser la question du nombre requis de sinusoïdes. Même si les expériences confirment que le nombre implicitement utilisé en effectuant une transformée inverse de Fourier s'avère suffisant, nous pouvons nous demander s'il est nécessaire. Nous pouvons aussi nous interroger sur la nécessité de définir aléatoirement les amplitudes. Là aussi, les expériences semblent nous indiquer que fixer les amplitudes des sinusoïdes lors de la synthèse n'implique pas de différence audible.

2.4.3 Première version du modèle

Une première ébauche de modèle spectral et statistique a été mise au point de façon à prendre en compte les propriétés des sons liées à la densité et à l'enveloppe spectrale [HDC00].

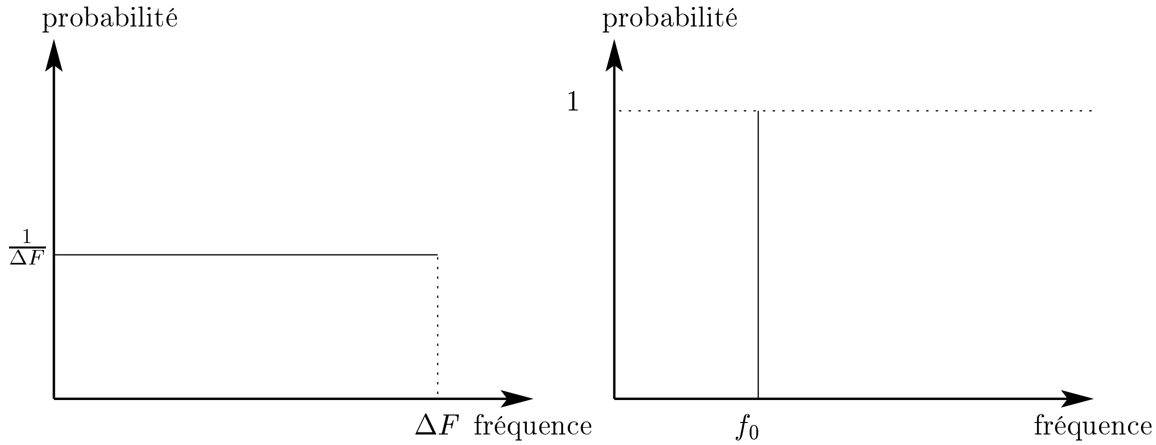

Fig. 2.11 – Fonction théorique de densité de probabilité (FDP) des fréquences dans le cas d'un bruit blanc (à gauche) et d'une sinusoïde pure de fréquence f_0 (à droite).

Ce modèle est général puisqu'il permet de représenter des sons allant d'une simple somme de partiels jusqu'au bruit blanc. Il définit pour cela quatre paramètres statistiques qui peuvent être reliés à la perception. Ces paramètres sont extraits des spectres à court terme du signal, et notamment les pics de ces spectres.

Fonction de densité de probabilité des fréquences

FDP* : *fréquence* $\times$ *temps* $\rightarrow$ *réel

Pour chaque fenêtre d'analyse spectrale, les composants prédominants sont extraits, c'est-à-dire ici les maxima locaux (pics), comme dans les cas de la modélisation sinusoïdale. Le premier paramètre du modèle est donc la probabilité pour chaque fréquence d'être présente dans le signal. En pratique, un ensemble de fenêtres consécutives est étudié, définissant la *fenêtre FDP*. La densité de probabilité des fréquences est liée à la notion musicale d'harmonicité. En effet, cette fonction dans le cas, par exemple, d'un son harmonique sans bruit est théoriquement une somme de fonctions de Dirac. Alors que si cette même fonction est constituée de bandes étroites autour des valeurs de fréquences harmoniques, elle représente alors l'inharmonicité (*warping*) telle qu'on la retrouve dans le modèle SAS [Mar99]. La différence est qu'ici l'inharmonicité est représentée par une densité de probabilité qui implique un caractère aléatoire à la synthèse, pouvant rendre le son plus réaliste. Deux exemples théoriques de fonctions de densité de probabilité des fréquences sont donnés par la figure 2.4.3.

Nombre de fréquences

FC* : *temps* $\rightarrow$ *entier

Le nombre de pics contenus dans chaque fenêtre spectrale est également un paramètre de ce modèle. Son rôle est évidemment fortement lié à la fonction de densité de probabilité des fréquences. Ainsi, un son harmonique d'instrument à vent est souvent considéré comme possédant un spectre composé d'étroites bandes autour de la valeur de fréquence de chaque harmonique [Fre99]. Ces bandes peuvent être représentées par des fonctions de densité de probabilité de fréquences élargies, au lieu de fonctions de Dirac, et par un nombre de fréquences plus important. Ce paramètre est bien sûr aussi lié à la notion de densité spectrale.

Probabilité de continuité

CP* : *fréquence* $\times$ *temps* $\rightarrow$ *réel

Une probabilité est attribuée à chaque valeur de fréquence pour déterminer aléatoirement si une sinusoïde peut avoir une durée supérieure à celle correspondant à la longueur de la fenêtre de synthèse. Ainsi, à l'analyse, comme dans certaines approches de la modélisation sinusoïdale, les phases associées aux pics sont testées pour déterminer cette probabilité.

Couleur

C* : *fréquence* $\times$ *temps* $\rightarrow$ *amplitude

La couleur est un des paramètres du modèle SAS [DCM99a]. Il représente donc une version interpolée et lissée de l'enveloppe spectrale. Son rôle dans la perception du timbre est reconnu.

Ce modèle peut être appliqué théoriquement à tous les sons à partir du moment où ils ne présentent pas de transitoire. Il repose en partie sur le modèle sinusoïdal puisqu'il permet de considérer des sinusoïdes de longue durée qui varient peu en amplitude et en fréquence, tout en permettant de représenter également le bruit par des sinusoïdes de courte durée suivant le modèle du bruit thermique. De plus, il ne décompose pas le signal en plusieurs parties comme les modèles hybrides le font. Des expériences ont été menées à partir de sons naturels, mettant en évidence les limites de cette approche.

Ces limites de ce modèle expérimental sont essentiellement liées à la technique d'extraction des paramètres, notamment l'estimation de la fonction de densité de probabilité des fréquences et le nombre des fréquences. En effet, à notre connaissance, de telles techniques n'existent pas et celles que nous avons développées pour ce modèle ne sont pas assez performantes pour pouvoir être appliquées sur des sons naturels.

Ce modèle expérimental général a été développé il y a quelques temps déjà, et nous ne bénéficions pas à l'époque des connaissances suffisantes à propos de la modélisation par les sinusoïdes du bruit. C'est pourquoi nous nous sommes attachés, depuis, à restreindre notre étude aux sons bruités, que nous pouvons considérer comme stochastiques. Nous nous sommes donc concentrés sur les signaux aléatoires, en sachant que de nombreux travaux de recherche continuent d'être menés, concernant les modèles sinusoïdaux. Ces études nous ont conduit à développer le modèle CNSS que nous présentons ici.

2.4.4 Perception de la densité spectrale

Expériences psychoacoustiques

Pour les expériences psychoacoustiques, les bruits peuvent être générés de deux manières différentes. La première consiste à filtrer du bruit blanc produit par distribution aléatoire des échantillons selon une loi uniforme ou gaussienne. La seconde possibilité propose de définir le spectre du bruit désiré, et de synthétiser le son en sommant des sinusoïdes. Ces méthodes de synthèse seront détaillées dans le chapitre 4. La méthode spectrale, qui repose sur le modèle des bruits thermiques, est cependant souvent préférée car elle présente l'avantage d'être plus explicite, en permettant un contrôle direct du spectre [Har87].

Cette approche pose donc des questions sur le nombre de sinusoïdes nécessaires pour générer un bruit qui ne pourra pas être différencié d'un bruit naturel ou d'un bruit synthétisé par production aléatoire d'échantillons. Les premières expériences ont ainsi été menées par A. Gerzso dès 1978 [Ger78]. Elles étaient aussi justifiées par des études concernant la mise au point des synthétiseurs pour connaître le nombre d'oscillateurs nécessaires pour recouvrir entièrement l'échelle des fréquences perçues par le système auditif.

Les premières expériences consistaient à comparer une bande de bruit synthétisé par filtrage passe-bande à un bruit produit par addition de sinusoïdes, d'amplitudes fixées de façon à représenter la même énergie, de fréquences aléatoirement choisies à l'intérieur de la même bande selon une loi uniforme, et de phases uniformément distribuées entre 0 et 2π. Ces expériences ont montré que le nombre de sinusoïdes admis comme nécessaire par des expériences plus anciennes, basées sur l'étude des propriétés de masquage (1 sinusoïde par Hz), était trop important. Elles ont aussi mis en évidence le rôle de la réverbération, puisque le nombre de sinusoïdes nécessaires est inférieur dans un environnement réverbérant que dans un environnement *sec*.

Ces expériences ont été reprises et poussées par W. Hartmann et al. [HMGB86] pour étudier notre capacité à discerner des bandes de bruit dont le nombre de sinusoïdes est différent. Un des bruits est composé d'un nombre de sinusoïdes volontairement très important, pouvant être considéré comme dépassant la valeur maximale. Il est comparé par des cobayes à un bruit généré à partir d'un nombre variable de sinusoïdes, plus faible. La largeur de la bande de bruit étudiée ainsi que la valeur de fréquence centrale sont également des variables de ces expériences.

Les conclusions de ces expériences sont nombreuses. Le nombre d'oscillateurs requis pour pouvoir synthétiser un bruit dont le spectre peut être considéré comme de densité maximale s'avère être assez faible dans le cas de bande étroite. Par exemple, les expériences montrent que seulement quatre sinusoïdes suffisent pour une bande de bruit de largeur 50Hz. De plus le mécanisme de discrimination de la densité de bandes étroites est indépendant de la valeur de la fréquence centrale, et donc de la résolution spectrale du système auditif. Ainsi, une première conclusion indique que ce mécanisme est principalement basé sur la perception des fluctuations d'intensité [Har87].

En ce qui concerne les bandes larges de bruit, ce lien est moins évident. La résolution spectrale apparaît être le facteur majeur utilisé par le système auditif pour discerner les bandes de bruit de densités différentes [HMGB86]. Dans ce cas, l'oreille perçoit les creux d'énergie du spectre correspondant à un trop grand écart entre deux sinusoïdes voisines.

Pour conclure sur ces expériences, nous pouvons utiliser, lors de l'élaboration d'un modèle de bruits, le fait que l'oreille est naturellement (sans apprentissage particulier) sensible au nombre de sinusoïdes utilisées pour synthétiser spectralement des bandes de bruit. Ainsi, ce nombre peut être appréhendé par la suite comme un paramètre perceptif d'une bande de bruit. Il est donc lié au contenu spectral (présence de creux d'énergie) ou aux fluctuations d'intensité. Le contrôle de ce nombre de sinusoïdes permet donc de maîtriser ces deux paramètres physiques. Des modèles mathématiques présentés dans la section suivante formalisent ces liens.

Par ailleurs, ces expériences confirment qu'une approche spectrale pour synthétiser du bruit semble matériellement possible. En effet, un nombre raisonnable d'oscillateurs au regard de la puissance des machines actuelles est nécessaire pour pouvoir synthétiser des bruits blancs dont la densité et la largeur de bande sont maximales, et qui correspondent donc au cas le plus critique du point de vue du temps de calcul.

Définition de la densité spectrale

Les expériences psychoacoustiques mettent donc en évidence la densité spectrale comme un paramètre perceptif permettant de différencier des bandes de bruit. Cependant donner une définition exacte et complète de cette densité spectrale ne semble pas être si facile. A. Gerzso relie la densité spectrale au rapport du nombre de sinusoïdes par la largeur de la bande de fréquences concernée [Ger78]. En notant ρ la densité spectrale d'une bande de bruit de largeur ΔF formée à partir de N sinusoïdes, cette définition basique peut donc s'écrire :

$$\rho = \frac{N}{\Delta F} \tag{2.13}$$

Néanmoins, en étudiant les résultats des expériences de W. Hartmann et lors de nos propres expériences, nous avons été amenés à mettre en défaut cette définition, sans toutefois parvenir à en formaliser une plus correcte. Nous pouvons quand même mettre en avant les limites de cette définition. En effet, elle ne prend pas en compte la distribution des fréquences. Ainsi, si nous considérons une bande de bruit formée à partir de N sinusoïdes réparties non linéairement sur l'intervalle des fréquences, plusieurs bruits provoquant des sensations de densités différentes peuvent être générés. Par exemple, nous pouvons envisager, pour une bande de largeur 200Hz, le cas de $N-1$ sinusoïdes très proches (sur un intervalle de 1Hz) à la limite inférieure de la bande et une sinusoïde dont la fréquence correspond à la limite supérieure. Dans ce cas, la densité spectrale serait certainement proche de $\rho = \frac{2}{200}$, car la bande serait équivalente du point de vue de la perception à une bande de même largeur composée de seulement deux sinusoïdes. Cet exemple volontairement extrême, illustré par la figure 2.12, montre la nécessité de prendre en compte dans la définition de la densité spectrale, la distribution des fréquences des sinusoïdes, ou alors de restreindre cette définition aux bandes de bruits formées de sinusoïdes dont les fréquences sont également espacées, comme pour la définition des bruits thermiques.

De plus, cette définition doit prendre en compte la durée de la bande de bruit. En effet, il apparaît lors de nos propres expériences que le système auditif n'est pas capable de différencier deux bruits d'une durée courte, alors qu'il est capable de le faire si la durée est allongée. Un exemple simple peut être donné. Un bruit blanc formé à partir de 512 sinusoïdes linéairement réparties entre 0 et 22050Hz (correspondant à une transformée inverse de Fourier pour une fenêtre de 1024 échantillons), ne pourra pas être différencié d'un bruit blanc naturel sur une durée de 1024 échantillons, alors qu'il le sera si la durée est doublée. Il est d'ailleurs étonnant que ce paramètre de durée n'ait pas été intégré à la liste des paramètres testés lors des expériences en psychoacoustique, puisqu'il nous a semblé prépondérant et qu'il est souvent présenté comme majeur dans le système auditif (intégration temporelle). Dans la suite, nous traiterons indépendamment le nombre de sinusoïdes utilisé et la durée du signal puisqu'ils seront deux paramètres distincts du modèle de bruit proposé.

Par ailleurs, les expériences menées par W. Hartmann sur la discrimination de la densité spectrale, étudiaient des bandes de bruit générées en utilisant des sinusoïdes dont les fréquences étaient linéairement réparties sur la bande de fréquence fixée. Nous pouvons penser, vue la tendance du système auditif à considérer des échelles logarithmiques (bandes critiques) que des écarts non linéaires modifieraient les résultats obtenus concernant le nombre de sinusoïdes nécessaires. Toutefois, à notre connaissance, aucune étude n'a été menée sur ce sujet. Dans la section 5.2, nous indiquons quelques idées qui pourraient permettre de définir une répartition non linéaire des fréquences et qui pourraient conduire à étudier l'influence de telles répartitions sur la perception de la densité spectrale.

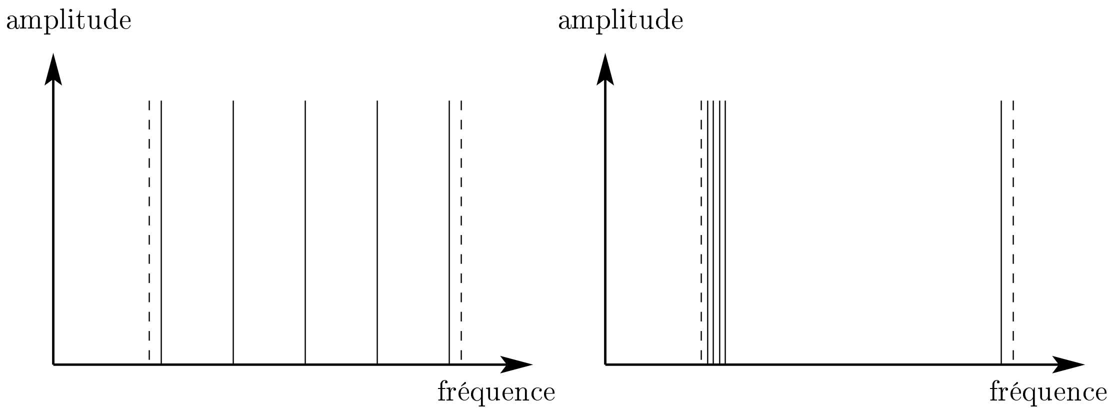

Fig. 2.12 – Illustration d'un exemple mettant en exergue les limites de la définition de la densité spectrale donnée par l'équation (2.13) : suivant cette définition, ces deux spectres de deux sons ont une même densité spectrale (ils sont constitués du même nombre de sinusoïdes pour une même largeur de bande), alors qu'ils ne doivent pas produire les mêmes sensations. La distribution des fréquences des sinusoïdes doit être prise en compte dans la définition.

2.4.5 Modèles de perception de la densité spectrale

Des expériences montrent que la perception de la densité spectrale des bandes étroites de bruit est principalement liée aux fluctuations d'intensité. Ce paramètre physique apparaît donc être directement lié à un paramètre perceptif. Il existe plusieurs approches mathématiques pour expliciter ces fluctuations d'intensité. Elles sont détaillées dans deux articles [HP91, HP88]. Nous les présentons rapidement ici, en nous basant sur l'hypothèse que le signal considéré est de moyenne nulle ($\mathcal{E}(X) = 0$).

Kurtosis du signal

La première méthode pour décrire les fluctuations d'intensité consiste à exprimer le quatrième moment du signal X, c'est-à-dire le kurtosis (voir équation (1.47)). La puissance instantanée, notée $P(t)$, s'exprime en fonction du carré du signal :

$$P(t) = X^2(t) \tag{2.14}$$

La variation de puissance, notée ΔP, s'écrit donc comme la différence entre la puissance et la puissance moyenne $\overline{P(t)}$, c'est-à-dire la variance $V(X)$:

$$\begin{aligned} \Delta P &= P(t) - \overline{P(t)} && (2.15) \\ \Delta P &= P(t) - V(X) && (2.16) \end{aligned}$$

Comme la moyenne de ΔP est nulle, la quantité étudiée est la moyenne de la puissance élevée au carré :

$$\begin{aligned} \overline{\Delta P^2} &= m_4(X) - 2V^2(X) + V^2(X) \\ \overline{\Delta P^2} &= m_4(X) - V^2(X) && (2.17) \end{aligned}$$

En normalisant cette valeur par le carré de la variance, nous obtenons :

$$\rho_1 = \frac{\overline{\Delta P^2}}{V^2(X)} = \frac{m_4(X)}{V^2(X)} - 1 = K_X - 1 \tag{2.18}$$

Ainsi, le kurtosis K_X permet de quantifier les variations d'intensité d'un bruit. Toutefois, il présente une incohérence par rapport à la perception. En effet, si nous considérons une sinusoïde pure, de fréquence et d'amplitude constantes, la valeur du kurtosis K_X vaut 1.5, et donc $\rho_1 = 0.5$. Pourtant ces fluctuations ne sont pas perçues par notre système auditif, puisqu'une sinusoïde pure donne une impression de stabilité [Har87].

Cette incohérence provient du fait que cette mesure ρ_1 prend en compte des variations que le système auditif ne perçoit pas, car ces variations sont trop rapides. Ainsi, en considérant un son x formé de N sinusoïdes d'amplitudes constantes ($\forall n, a_n = 1$) et de fréquences f_n régulièrement espacées, nous pouvons expliciter, par exemple, sa puissance par :

$$\begin{aligned} x^2(t) &= [\sum_{n=1}^{N} a_n \sin(2\pi f_n t + \phi_n)]^2 \\ &= \sum_{n=1}^{N} \sin^2(2\pi f_n t + \phi_n) + 2\sum_{n=1}^{N}\sum_{n'=1}^{n-1} \sin(2\pi f_n t + \phi_n)\sin(2\pi f_{n'} t + \phi_{n'}) \\ &= \sum_{n=1}^{N} \frac{1}{2} + \sum_{n=1}^{N}\sum_{n'=1}^{n-1} \sin[2\pi(f_n - f_{n'})t + (\phi_n - \phi_{n'})] \\ &+ \sum_{n=1}^{N}\sum_{n'=1}^{n-1} \sin[2\pi(f_n + f_{n'})t + (\phi_n + \phi_{n'})] && (2.19) \end{aligned}$$

La puissance moyenne est donc $\frac{N}{2}$, les moyennes des sinusoïdes étant nulles. Les contributions apportées par les deux autres parties peuvent être distinguées. Celle due aux sinusoïdes dont les fréquences sont des sommes de fréquences, qui varient donc rapidement, ne contribue que très rarement à la perception des fluctuations. Pourtant elles sont prises en compte par cette mesure qui surestime donc les fluctuations d'intensité par rapport à ce que l'oreille perçoit.

Fluctuations d'enveloppe temporelle

Une autre mesure mathématique permet de modéliser les fluctuations d'intensité, sans présenter les défauts liés à la puissance du signal. Cette mesure est basée sur la puissance de l'enveloppe temporelle, définie dans la section 1.1.9. En considérant un bruit suivant le modèle du bruit thermique (équation (2.11)), la puissance de l'enveloppe temporelle est obtenue en mettant au carré l'équation (1.37) :

$$E^2(t) = \sum_{n=1}^{N} C_n^2 + 2\sum_{n=1}^{N}\sum_{n'=1}^{n-1} C_n C_{n'} \cos[(\omega_n - \omega_{n'})t + \phi_n - \phi_{n'}] \tag{2.20}$$

Nous pouvons remarquer que, contrairement à la puissance du signal, seules les différences de fréquences interviennent dans la définition de la puissance de l'enveloppe.

Des études psychoacoustiques [Har97] montrent que seules les fluctuations d'enveloppe de basses fréquences sont perçues. Cette propriété est caractérisée par un filtre passe-bas et sa fonction de transfert TMTF (*temporal modulation transfer function*). En effet, le système auditif ne répond pas immédiatement aux variations d'intensité. La puissance de l'enveloppe est donc filtrée pour accentuer les basses fréquences. La fonction de transfert s'écrit :

$$h(t) = \frac{1}{\tau}\exp(-\frac{t}{\tau}) \tag{2.21}$$

où τ est de l'ordre de 3ms. L'opération de filtrage s'écrit donc :

$$E_\mathrm{f}^2(t) = \int_0^\infty E^2(t-t')h(t')dt' \tag{2.22}$$

Pour simplifier les équations, nous notons :

$$\Delta_{nn'} = \omega_n - \omega_{n'} \tag{2.23}$$

$$y_{nn'} = (\omega_n - \omega_{n'})t + \phi_n - \phi_{n'} \tag{2.24}$$

La puissance de l'enveloppe filtrée devient donc :

$$E_\mathrm{f}^2(t) = \sum_{n=1}^{N} C_n^2 + 2\sum_{n=1}^{N}\sum_{n'=1}^{n-1} C_n C_{n'} \frac{\cos(y_{nn'}) + \tau\Delta_{nn'}\sin(y_{nn'})}{1+(\tau\Delta_{nn'})^2} \tag{2.25}$$

D'après la définition de la variance donnée par l'équation (1.45), la variance de la puissance de l'enveloppe est donnée par la relation :

$$V(E_\mathrm{f}^2) = \left\langle (E_\mathrm{f}^2(t) - \left\langle E_\mathrm{f}^2(t)\right\rangle)^2 \right\rangle \tag{2.26}$$

Or, sachant que les moyennes de sinusoïdes, de fréquences différentes de 0, sont nulles, il vient :

$$\left\langle E_\mathrm{f}(t)^2 \right\rangle = \sum_{n=1}^{N} C_n^2 \tag{2.27}$$

Ainsi, nous pouvons expliciter la variance, en utilisant la propriété que les moyennes des double produits($\langle \sin\cos\rangle$) sont nulles dans le cas d'une distribution de phases uniforme (voir la section 1.1.4) :

$$V(E_\mathrm{f}^2) = \left\langle 4\sum_{n=1}^{N}\sum_{n'=1}^{n-1} {C_n}^2 {C_{n'}}^2 \frac{\cos^2(y_{nn'}) + (\tau\Delta_{nn'})^2\sin^2(y_{nn'})}{(1+(\tau\Delta_{nn'})^2)^2} \right\rangle \tag{2.28}$$

De plus, comme $\left\langle cos^2(y_{nn'})\right\rangle = \frac{1}{2}$ et $\langle sin(y_{nn'})\rangle = \frac{1}{2}$, il vient donc :

$$V(E_{\mathrm{f}}^2) = 2\left\langle \sum_{n=1}^{N}\sum_{n'=1}^{n-1} {C_n}^2 {C_{n'}}^2 \frac{1}{1+(\tau\Delta_{nn'})^2}\right\rangle \tag{2.29}$$

En normalisant cette valeur par le carré de la puissance de l'enveloppe, nous obtenons donc dans le cas d'un bruit blanc :

$$\frac{V(E_{\mathrm{f}}^2)}{\langle E^2\rangle^2} = \frac{2\sum_{n=1}^{N}\sum_{n'=1}^{n-1} C_n^2 C_{n'}^2 \left\langle\frac{1}{1+(\tau\Delta_{nn'})^2}\right\rangle}{\sum_{n=1}^{N}\sum_{n'=1}^{N} C_n^2 C_{n'}^2} = (1-\frac{1}{N})\left\langle\frac{1}{1+(\tau\Delta_{nn'})^2}\right\rangle \tag{2.30}$$

Nous notons cette mesure $\mathcal{V}(E_{\mathrm{f}}^2)$. Elle dépend donc de la distribution des fréquences composant le bruit. W. Hartmann a proposé plusieurs distributions. La distribution la plus générale divise la bande des fréquences ΔF en M casiers, à l'intérieur desquels sont choisies aléatoirement N fréquences [Har97]. Dans ce cas, l'intégration donne la relation :

$$\begin{aligned}\mathcal{V}(E_{\mathrm{f}}^2) = \frac{(1-\frac{1}{N})}{2\pi\Delta F\tau}\frac{2}{1-\frac{1}{M}}\Bigg\{&\tan^{-1}(2\pi\Delta F\tau) - \tan^{-1}(\frac{2\pi\Delta F\tau}{M})\\ &+\frac{1}{4\pi\Delta F\tau}\left[M\ln\left[1+(\frac{2\pi\Delta F\tau}{M})^2\right] - \ln[1+(2\pi\Delta F\tau)^2]\right]\Bigg\}\end{aligned} \tag{2.31}$$

Nous pouvons faire ici plusieurs remarques qui confirment la qualité du lien de ce modèle avec la perception pour représenter les fluctuations d'intensité. Tout d'abord, au contraire de la puissance du signal, la valeur $\mathcal{V}(E_{\mathrm{f}}^2)$ est nulle lorsque N vaut 1, c'est-à-dire dans le cas d'un signal composé d'une seule sinusoïde. De plus, ces fluctuations augmentent lorsque le nombre de sinusoïdes N augmente, et lorsque la largeur de la bande de fréquence ΔF diminue. Pour autant, cette mesure n'est pas une fonction simple du rapport $\frac{N}{\Delta F}$ [HMGB86]. Ce modèle des fluctuations d'intensité correspond donc aux résultats des expériences psychoacoustiques dans le cas de bandes de bruit étroites.

2.4.6 Modèle statistique

Le modèle CNSS que nous avons développé est un modèle spectral basé sur le modèle du bruit thermique, définissant donc un son comme une somme finie de sinusoïdes. Les expériences psychoacoustiques révèlent la capacité du système auditif à discerner des sons de densités spectrales différentes. Pour expliquer cette capacité, un modèle mathématique des fluctuations d'intensité a été proposé 2.4.5. Celui-ci montre que ces fluctuations, perçues par l'oreille, dépendent de la distribution des fréquences des sinusoïdes composant le signal. En changeant cette distribution, nous obtenons un contrôle d'un paramètre perceptif des bruits, en modifiant ces fluctuations dans certaines bandes du signal, ou en imposant des creux d'énergie dans le spectre par un écart trop important entre deux fréquences voisines. Or le modèle du bruit thermique définit les fréquences des sinusoïdes comme régulièrement espacées. Le modèle CNSS se différencie donc de ce modèle en définissant les fréquences comme des variables aléatoires, dont la distribution est paramétrable par des valeurs mathématiques que nous détaillons dans la section 2.4.8. L'approche statistique se justifie par le fait que l'oreille ne paraît pas être sensible de manière très précise aux fluctuations d'intensité ou à l'emplacement exact de creux. Il semble donc inutile lors de la représentation d'un bruit de retenir ces informations de répartition des fréquences de manière déterministe.

Par ailleurs, l'étude des fluctuations d'intensité fait également apparaître leur dépendance à la distribution des phases (passage de l'équation (2.26) à l'équation (2.28)). Cette dépendance peut être illustrée de façon simple en prenant deux cas extrêmes. Le cas des phases uniformément distribuées est traité par l'étude précédente. Si ces phases prennent une même valeur fixée $\frac{\pi}{2}$, les fluctuations augmentent beaucoup. En effet, chaque sinusoïde composant le signal présente une valeur maximale à $t = 0$, quelle que soit sa fréquence. La somme de ces sinusoïdes a par conséquence une valeur égale à la somme des amplitudes de toutes les sinusoïdes. Un pic d'intensité a alors lieu à la date $t = 0$. L'oreille est sensible à cette irrégularité et perçoit une impulsion (transitoire). C'est pourquoi le modèle CNSS propose également de contrôler la distribution des valeurs des phases des sinusoïdes composant le signal.

Concernant les amplitudes des sinusoïdes, le modèle thermique indique que ce sont des variables aléatoires distribuées selon la loi de Rayleigh. En pratique, il apparaît toutefois que le choix d'amplitudes fixées de manière non aléatoire permet de générer des bandes de bruits qui ne peuvent être différenciées de bandes de bruits produites à l'aide de sinusoïdes dont les amplitudes sont des variables aléatoires suivant la loi de Rayleigh. De plus, nous n'avons pas réussi théoriquement à justifier l'intérêt de définir ces amplitudes de manière aléatoire, et d'établir un lien entre une éventuelle distribution et une propriété perceptive. C'est pourquoi nous nous sommes restreints, pour l'instant, à définir des amplitudes de façon déterministe pour le modèle CNSS.

2.4.7 Base du modèle CNSS

Le modèle CNSS (*Colored Noise by Sum of Sinusoids*) définit un son en tant que processus aléatoire X_k, comme une somme finie de N sinusoïdes, dont les amplitudes a_n sont fixées, les phases Φ_n, appartenant à l'intervalle $[0; 2\pi[$, et les fréquences F_n, appartenant à une bande de fréquence de largeur ΔF, sont des variables aléatoires, suivant des lois distinctes :

$$X_k = \sum_{n=1}^{N} a_n \sin(2\pi F_n \frac{k}{F_s} + \Phi_n) \tag{2.32}$$

Nous pouvons vérifier mathématiquement que cette approche spectrale et statistique, proche de celle du modèle du bruit thermique, peut permettre de définir un bruit blanc. Un bruit blanc doit respecter la condition mathématique donnée dans la section 1.4.3 par l'équation (1.75). En écrivant, pour tout (p, q), le produit de l'espérance de X_p par celle de X_{p+q}, il vient :

$$
\begin{aligned}
\mathcal{E}(X_pX_{p+q}) &= \mathcal{E}(\sum_{n=1}^{N} a_n \sin(2\pi F_n \frac{p}{F_s} + \Phi_n) \sum_{l=1}^{N} a_l \sin(2\pi F_l \frac{p+q}{F_s} + \Phi_l)) \\
\mathcal{E}(X_pX_{p+q}) &= \sum_n \sum_l a_n a_l \mathcal{E}(\sin(2\pi F_n \frac{p}{F_s} + \Phi_n) \sin(2\pi F_l \frac{p+q}{F_s} + \Phi_l)) \\
\mathcal{E}(X_pX_{p+q}) &= \frac{1}{2} \sum_n \sum_l a_n a_l [\mathcal{E}(\cos(2\pi F_n \frac{p}{F_s} + 2\pi F_l \frac{p+q}{F_s}) + \Phi_n + \Phi_l) \\
&+ \mathcal{E}(\cos(2\pi F_n \frac{p}{F_s} - 2\pi F_l \frac{p+q}{F_s}) + \Phi_n - \Phi_l)]
\end{aligned} \quad (2.33)
$$

Comme cette espérance est définie en effectuant une intégration sur les phases, supposées ici distribuées uniformément entre 0 et 2π, nous pouvons donc simplifier par :

$$
\mathcal{E}(\cos(2\pi F_n \frac{p}{F_s} + 2\pi F_l \frac{p+q}{F_s}) + \Phi_n + \Phi_l) = 0 \quad (2.34)
$$

et, pour $l \neq n$:

$$
\mathcal{E}(\cos(2\pi F_n \frac{p}{F_s} - 2\pi F_l \frac{p+q}{F_s}) + \Phi_n - \Phi_l) = 0 \quad (2.35)
$$

Il vient donc :

$$
\mathcal{E}(X_pX_{p+q}) = \frac{1}{2} \sum_n a_n^2 \mathcal{E}(\cos(2\pi F_n \frac{q}{F_s})) \quad (2.36)
$$

Nous pouvons donc conclure que pour tout couple d'entiers (p, q) :

- $\mathcal{E}(X_pX_{p+q}) = 0$, si $q \neq 0$
- $\mathcal{E}(X_pX_{p+q}) = \frac{1}{2}\sum_n a_n^2$, si $q = 0$

ce qui correspond à la définition du bruit blanc.

2.4.8 Paramètres du modèle CNSS

Le modèle statistique et spectral CNSS que nous avons développé propose donc de représenter un son par une somme finie de sinusoïdes, dont les fréquences et les phases sont des variables aléatoires. En pratique, comme dans les autres modèles spectraux (SMS, STN, SAS, ...), le son est analysé et synthétisé par ajout successif, avec ou sans chevauchement, de plusieurs segments temporels, des fenêtres (*frames*). Cette considération n'apparaît ni dans la définition théorique du bruit thermique, ni dans celle du modèle CNSS. Toutefois, elle se justifie de plusieurs manières. La première est le fruit des expériences que nous avons été amenés à réaliser pour tester le modèle. Comme nous l'avons déjà fait remarquer, le nombre de sinusoïdes utiles pour synthétiser un bruit blanc dépend de la durée du son à produire. Par contre, si plusieurs réalisations d'un même signal de durée réduite se succèdent dans le temps, la perception du son est similaire à celle d'un bruit blanc. Le fait d'utiliser plusieurs fenêtres temporelles successives nous permet de restreindre le nombre de sinusoïdes utilisées pour la modélisation et surtout pour la synthèse, ce qui permet de réduire le temps de calcul, très intimement lié au nombre d'oscillateurs. La seconde justification provient du déroulement de la méthode d'analyse pour extraire les paramètres mathématiques du modèle. Des approches possibles pour cette analyse seront détaillées dans le chapitre 3. Néanmoins, l'utilisation des statistiques pose le problème de l'hypothèse de stationnarité et d'ergodicité. Or la phase d'analyse ne doit s'effectuer qu'à partir d'une seule réalisation d'un signal supposé aléatoire. La seule façon de traiter ce signal de manière statistique consiste à se baser sur l'hypothèse d'ergodicité, en analysant plusieurs fenêtres temporelles supposées indépendantes. En contrepartie, l'hypothèse de stationnarité nous oblige à ne pas employer une taille de fenêtre trop large. Ces problèmes de choix de taille de fenêtres sont aussi abordés dans le chapitre consacré à l'analyse.

Le modèle CNSS représente donc une succession de fenêtres temporelles de signal aléatoire (bruit). Chaque fenêtre sera modélisée par plusieurs paramètres mathématiques que nous allons détailler ici et qui sont illustrés par la figure 2.13.

Largeur de la bande de fréquences

Le signal représenté est supposé être une bande de bruit. Un des paramètres est la largeur de cette bande, qui définit aussi l'intervalle de la fonction de densité de probabilité des fréquences. Elle sera notée ΔF, et sera en fait définie par une borne minimale $F_{\min}$ et une borne maximale $F_{\max}$:

$$\Delta F = F_{\max} - F_{\min} \tag{2.37}$$

Comme nous supposons $\Delta F > 0$, nous supposons aussi $F_{\min} < F_{\max}$. L'unité de ce paramètre est le Hertz.

Évidemment ce paramètre est limité par les propriétés de la perception fréquentielle. Comme le système auditif ne perçoit que les fréquences comprises entre à peu près 20 et 20000Hz, nous pourrions attribuer ces mêmes limites. Cependant, pour des aspects pratiques, nous fixons généralement les bornes de cet intervalle à 0 et $\frac{F_s}{2} = 22050$Hz, puisque la fréquence d'échantillonnage employée est souvent 44100Hz. Cet intervalle correspond à la bande de fréquences utilisée lors d'une transformée (inverse) de Fourier. C'est donc celle qui est implicitement employée par les modèles spectraux comme la partie stochastique du modèle SMS.

Nombre de casiers

Pour décrire la fonction de densité de probabilité des fréquences, nous définissons des casiers (*bins*), de taille constante, recouvrant toute la bande de fréquence choisie ΔF. Nous noterons ce paramètre M, qui est sans unité. Le choix du nom est évidemment le même que pour décrire la résolution fréquentielle de la transformée de Fourier, qui, elle aussi, divise l'intervalle des fréquences en un certain nombre de casiers de taille égale.

Chaque casier, noté B_i avec i entier compris entre 0 et $M-1$, décrit un intervalle de fréquence à partir de largeur de la bande ΔF. La largeur de tous les casiers, notée ΔB, est constante et vaut donc :

$$\forall i \in [0; M-1[, \quad \Delta B = B_{i+1} - B_i = \frac{\Delta F}{M} \tag{2.38}$$

L'intervalle I_{B_i} décrit par chaque casier B_i est :

$$\forall i \in [0; M-1], \quad I_{B_i} =]F_{\min} + i\frac{F_{\max} - F_{\min}}{M}; F_{\min} + (i+1)\frac{F_{\max} - F_{\min}}{M}] \tag{2.39}$$

Les figures 2.13 et 2.15 illustrent deux valeurs différentes du nombre de casiers, en choisissant à l'intérieur de chacun une fréquence.

Le nombre de casiers M a la particularité de ne pas être borné. Ainsi, il peut prendre toutes les valeurs positives non nulles, et peut aller jusqu'à l'infini. Ce cas du nombre de casiers infini est illustré par la figure 2.17. Nous verrons plus loin la signification d'un nombre de casiers infini.

$$M \in [1; \infty] \tag{2.40}$$

A l'intérieur de chaque casier recouvrant ainsi toute la bande de fréquence choisie, une fréquence au plus sera aléatoirement choisie selon une loi de distribution définie par les autres paramètres du modèle.

Nombre de sinusoïdes

Le nombre de sinusoïdes, noté N et apparaissant dans l'équation (2.32) définissant le modèle CNSS, permet de contrôler le nombre de composants fréquentiels qui vont représenter le spectre du son. Ce paramètre est sans unité, mais présente une contrainte liée au nombre de casiers M. En effet, il n'est pas possible de choisir (aléatoirement) plusieurs fréquences dans un même casier. Le nombre maximum de sinusoïdes correspond donc au nombre de casiers M, sauf dans le cas où celui-ci est infini. La figure 2.13 illustre le cas où $N = M$, et la figure 2.16 illustre le cas où $N < M$. Par ailleurs, un son doit être composé au minimum d'une sinusoïde :

$$1 \leq N \leq M \tag{2.41}$$

En pratique, l'influence du nombre de sinusoïdes est très importante sur l'efficacité de la synthèse. Nous revenons en détails sur ce point dans le chapitre 4. L'intérêt est de minimiser cette valeur, indépendamment des propriétés perceptives du son produit. Or l'influence de ce nombre de sinusoïdes N est primordiale sur la perception de la densité spectrale. Nous proposons une approximation du nombre maximum de sinusoïdes. Nos convictions reposent sur les expériences que nous avons menées, mais que nous n'avons pas quantifiées contrairement aux études psychoacoustiques de A. Gerzso [Ger78] et W. Hartmann [HMGB86]. Le nombre de sinusoïdes maximum nous paraît être lié à la taille de la fenêtre W_s du signal. Cette taille n'apparaît pas dans les expériences psychoacoustiques. Pourtant, le modèle du bruit thermique définit le bruit comme une somme de sinusoïdes dont les fréquences sont également espacées, ce qui revient à la définition mathématique d'un son harmonique. La différence entre un bruit et un son harmonique provient de la valeur de l'écart entre ces fréquences, qui correspond à la fréquence fondamentale. Or le son doit avoir au minimum une durée équivalente à la période, pour que sa hauteur (fréquence fondamentale) puisse être perçue. Cette considération nous indique donc que si N fréquences sont définies sur une bande de largeur $\frac{F_s}{2}$Hz (à peu près l'intervalle des fréquences audibles), l'écart entre les fréquences est donc $\frac{F_s}{2N}$Hz. La durée minimale du signal pour que cette fréquence soit audible vaut $\frac{2N}{F_s}$s, ce qui correspond à $2N$ échantillons. Ainsi, pour définir un bruit de densité spectrale maximale de W_s échantillons, il faut donc utiliser $\frac{W_s}{2}$ sinusoïdes. Il est intéressant de noter que cette limite correspond à celle implicitement employée lors de la transformée inverse de Fourier (partie stochastique du modèle SMS par exemple). Dans le cas d'un bruit blanc, c'est-à-dire d'une bande de fréquence de largeur $\frac{F_s}{2}$, ce nombre maximum correspond donc à la moitié de la taille de la fenêtre W_s :

$$N \leq \frac{W_s}{2} \tag{2.42}$$

De même, nous faisons également l'hypothèse que le nombre de sinusoïdes maximum est relié de façon linéaire à la largeur de la bande étudiée. Dans la section 5.2, nous discutons cette hypothèse. Ainsi, une bande de bruit de largeur ΔF quelconque, de durée W_s et de densité spectrale maximum, sera représentée par :

$$N_{\max} = \frac{\Delta F.W_s}{F_s} \tag{2.43}$$

Le nombre de sinusoïdes, utilisées pour représenter les sons dans le modèle CNSS, est donc défini dans l'intervalle $[0; \frac{\Delta F.W_s}{F_s}]$. Il est toutefois important de rappeler que ce seuil n'a pas été validé pour le moment par des expériences psychoacoustiques adaptées. Cela fait partie des travaux futurs (voir chapitre 5).

Largeur de la fonction de densité de probabilité

La valeur de la fréquence à l'intérieur d'un casier est décidée aléatoirement suivant une loi uniforme. Un paramètre noté L permet toutefois de contrôler la largeur de cette loi (le domaine de définition de la fonction de densité de probabilité). Ce paramètre est une valeur relative, réelle, sans unité, comprise entre 0 et 1 ($0 \leq L \leq 1$). Lorsque cette largeur L est nulle, la fonction de densité de probabilité associée à tous les casiers est une fonction de Dirac, correspondant à la valeur de l'extrémité supérieure (borne maximale) de l'intervalle défini par le casier, c'est-à-dire pour le casier B_i à la valeur $F_{\min} + (i+1)\frac{F_{\max}-F_{\min}}{M}$. Ce cas est illustré par la figure 2.14. Au contraire, si cette largeur L prend sa valeur maximale 1, la fonction de densité de probabilité associée au casier B_i est une fonction rectangulaire (constante) définie sur l'intervalle B_i. Ce cas est illustré par la figure 2.13. Ainsi, dans le cas où le nombre de fréquences est égal au nombre de casiers, nous pouvons écrire, pour le casier B_i :

$$\begin{aligned} \text{si} \quad L = 0, \quad F_i &= F_{\min} + (i+1)\frac{F_{\max}-F_{\min}}{M} \\ \forall L \in]0;1], \quad F_i &\in \left]F_{\min} + (i+1-L)\frac{F_{\max}-F_{\min}}{M}; F_{\min} + (i+1)\frac{F_{\max}-F_{\min}}{M}\right] \end{aligned} \tag{2.44}$$

La fonction de densité de probabilité uniforme des fréquences, notée ρ, associée à un casier B_i peut donc s'écrire, pour tout entier i compris entre 0 et $N-1$ et pour $L \neq 0$:

$$\forall F_i \in]F_{\min} + (i+1-L)\frac{\Delta F}{M}; F_{\min} + (i+1)\frac{\Delta F}{M}], \quad \rho(F_i) = \frac{1}{L\frac{\Delta F}{M}}$$

Ce paramètre permet ainsi de définir la régularité des écarts entre les fréquences composant le son modélisé. Ainsi, dans le cas d'une largeur nulle ($L = 0$) et d'un nombre de fréquences égal au nombre de casiers ($N = M$), toutes les fréquences sont régulièrement espacées :

$$\forall i \in [0; N-2], \quad F_{i+1} - F_i = \frac{\Delta F}{M} \tag{2.45}$$

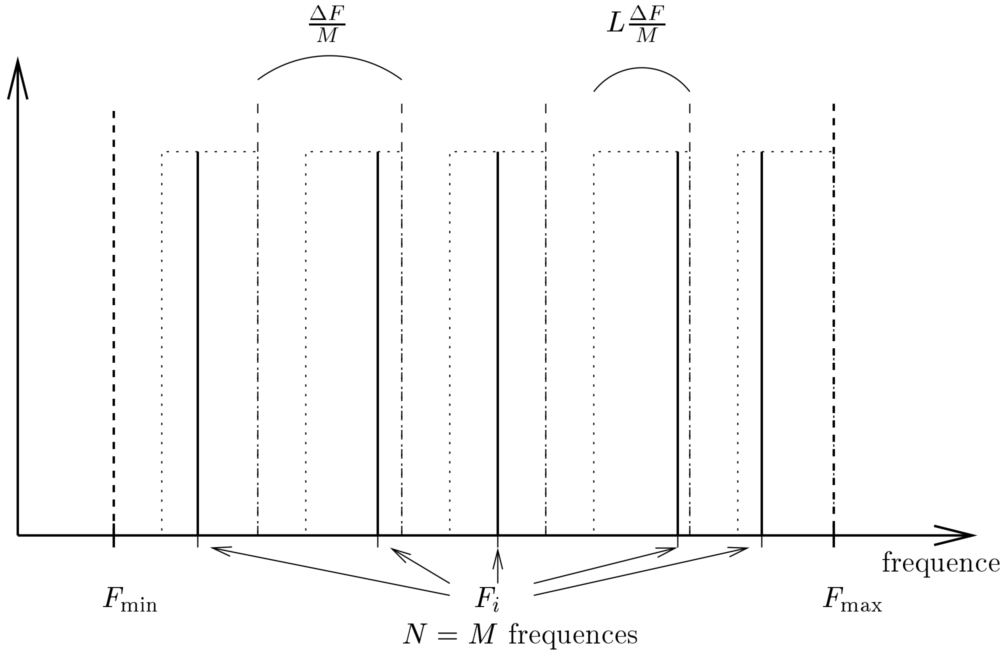

FIG. 2.13 – Illustration des paramètres du modèle CNSS concernant la distribution des fréquences : N fréquences notées F_i sont aléatoirement choisies dans M casiers réguliers sur une bande de largeur $\Delta F = F_{\max} - F_{\min}$, selon une loi de distribution uniforme de largeur $L\frac{\Delta F}{M}$. Dans cet exemple le nombre de casiers correspond au nombre de fréquences.

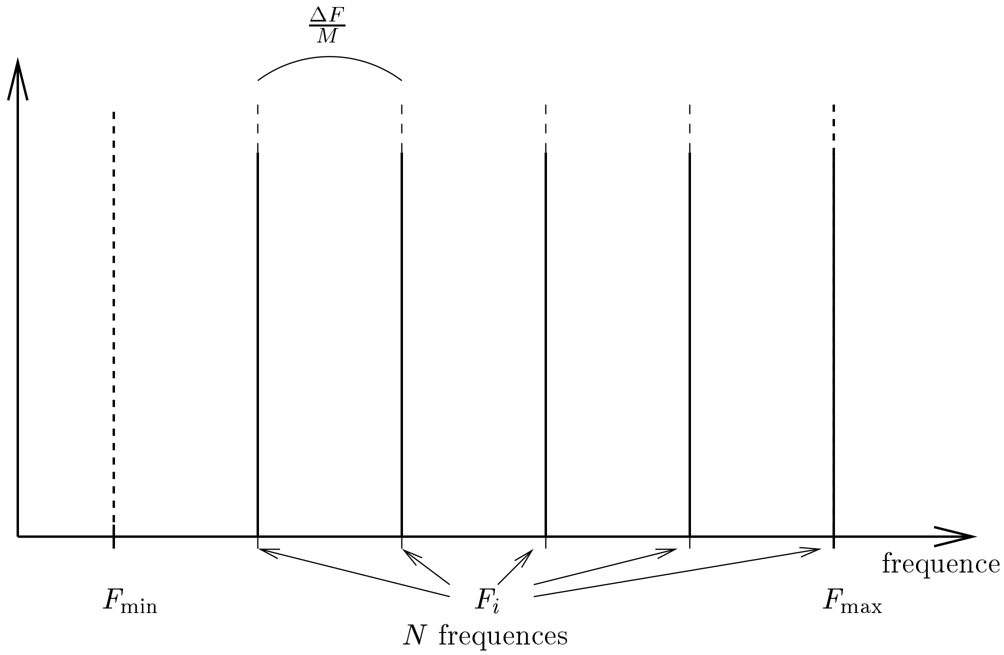

FIG. 2.14 – Exemple de tirage de fréquences contrôlé par les paramètres du modèle CNSS : le paramètre L étant nul, les fréquences sont régulièrement espacées conformément à une transformée inverse de Fourier ou au modèle du bruit thermique.

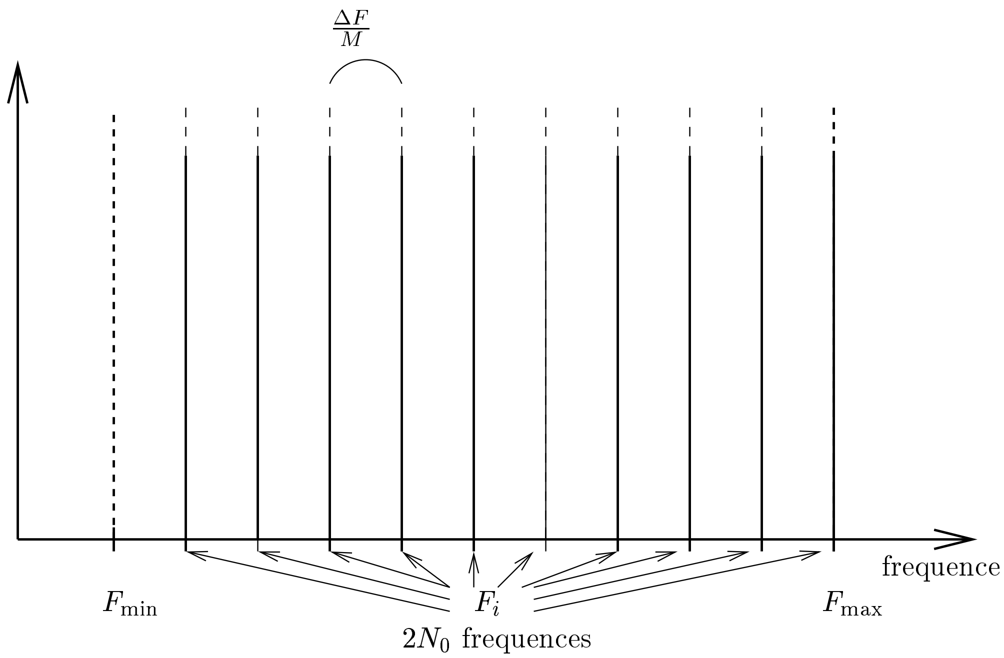

FIG. 2.15 – Exemple de tirage de fréquences contrôlé par les paramètres du modèle CNSS : ici, le nombre de casiers (et de fréquences) est doublé par rapport à la figure 2.14, mais les fréquences sont toujours régulièrement réparties.

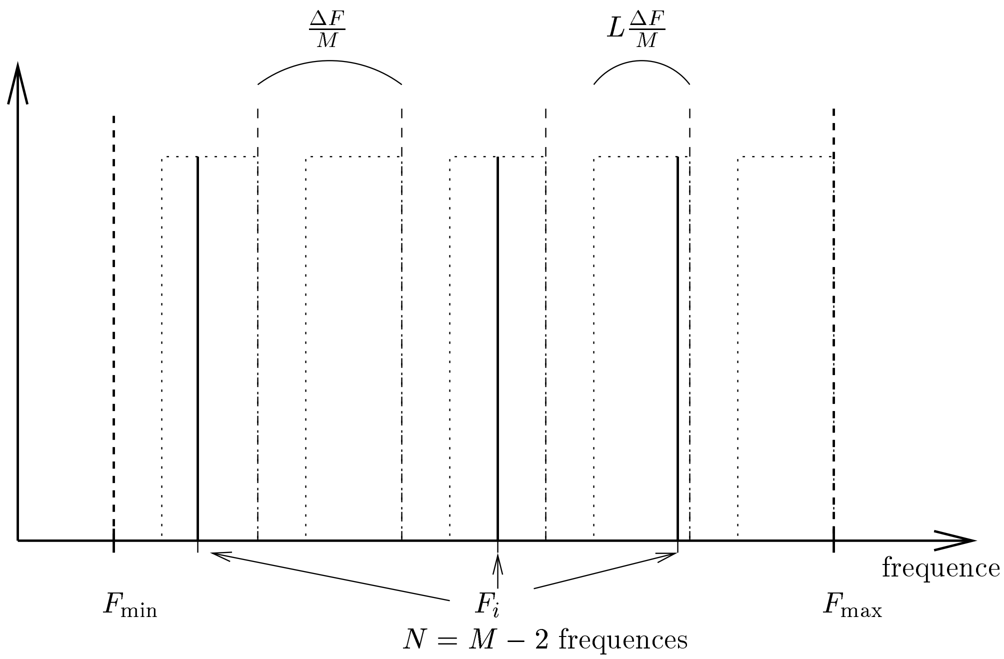

FIG. 2.16 – Exemple de tirage de fréquences contrôlé par les paramètres du modèle CNSS : ici, le nombre de fréquences N est inférieur au nombre de casiers M recouvrant la bande de fréquence, ce qui peut provoquer des écarts de fréquence plus importants.

Longueur de la fenêtre d'analyse/synthèse

Le modèle CNSS représente le son en considérant des fenêtres temporelles successives qui peuvent éventuellement se chevaucher (*OverLap and Add*). La figure 2.19 illustre cette méthode. Nous notons cette taille de fenêtre W_s (*window size*). Elle est entière, positive et est exprimée en nombre d'échantillons, c'est-à-dire en unité correspondant à $\frac{1}{F_s}$s, où F_s est la fréquence d'échantillonnage.

Largeur de la fonction de densité de probabilité des phases

Le modèle du bruit thermique décrit les phases de chaque sinusoïde composant le signal comme des variables aléatoires distribuées selon une loi uniforme entre 0 et 2π. Le modèle CNSS permet de modifier cette loi de distribution en limitant l'intervalle $[0;2\pi[$. Ainsi, il définit la largeur de la fonction de densité de probabilité des phases par une valeur relative, notée P, qui est un réel compris entre 0 et 1 ($0 \leq P \leq 1$) . Ce paramètre est sans unité. Lorsqu'il est nul, la fonction de densité de probabilité associée aux phases est une fonction de Dirac, les phases des sinusoïdes représentant le signal sont des valeurs constantes fixées, telles que leur valeur soit exactement $\frac{\pi}{2}$ à l'instant t_0 correspondant à la moitié de la durée de la fenêtre W_s :

$$t_0 = \frac{W_s}{2} \tag{2.46}$$

Lorsque ce paramètre P prend la valeur 1, les phases sont représentées par des variables aléatoires de loi uniforme entre 0 et 2π, conformément au modèle du bruit thermique. L'instant de référence pour le contrôle des phases correspond à la moitié de la fenêtre W_s, à cause d'un éventuel chevauchement des fenêtres qui s'accompagne généralement d'une multiplication par une fenêtre de pondération nulle aux extrémités (voir le principe du chevauchement dans la section 4.5). L'intérêt de ce paramètre résidant dans le fait de pouvoir contrôler les impulsions d'intensité éventuelles du signal, il est nécessaire de ne pas annuler ces impulsions par ces fenêtres de pondération. Pour cela, les phases sont synchronisées au milieu de la fenêtre.

En considérant les phases $\Phi_i^{t_0}$ à l'instant t_0 des sinusoïdes du signal, nous pouvons écrire le lien entre le paramètre P et ces phases, pour tout entier i compris entre 0 et $N-1$:

$$\forall P \in [0;1], \quad \Phi_i \in [\frac{\pi}{2} - P\pi; \frac{\pi}{2} + P\pi]; \tag{2.47}$$

Nous pouvons alors expliciter les phases des sinusoïdes à l'instant initial (début de la fenêtre $t=0$), en fonction de leur fréquence F_i :

$$\forall P \in [0;1], \quad \Phi_i \in [\frac{\pi}{2} - \frac{\pi F_i W_s}{F_s}; \frac{\pi}{2} + 2P\pi - \frac{\pi F_i W_s}{F_s}]; \tag{2.48}$$

Couleur

Le paramètre couleur est un paramètre déjà utilisé par d'autres modèles classiques (SMS, STN) sous la forme d'enveloppe spectrale lissée. Le modèle SAS [Mar00] définit ce même paramètre. Son nom est dû à l'analogie effectuée avec la lumière, et le spectre visible. Ce terme est déjà utilisé pour définir des bruits, comme nous l'avons vu dans la section 1.4.4. Il est aussi couramment utilisé en musique pour qualifier le timbre d'instruments.

Pour le modèle CNSS, la couleur est l'enveloppe spectrale lissée d'un son. Ce paramètre sera noté C et représente donc les variations de l'amplitude en fonction des fréquences. C'est une fonction positive, et théoriquement une fonction continue, mais elle sera modélisée sous la forme d'un nombre fini de points, pas forcément régulièrement espacés (échantillonnage non uniforme). L'importance du lissage de l'enveloppe spectrale est relative à la nature du son (voir la figure 2.18). Concernant les bruits, le lissage paraît pouvoir être important, puisqu'il ne devrait apparaître dans le spectre ni pic ni variation brusque d'énergie, comme c'est le cas pour les sons instrumentaux non bruités par exemple.

Il est également possible de représenter la couleur sous la forme de fonctions mathématiques [Sch98] comme les splines par exemple, ou sous la forme de coefficients du filtre associé (prédiction linéaire) [Rod97a]. La littérature propose des comparaisons de ces différentes approches [SR99], et nous discutons des possibilités de l'analyse de cette enveloppe dans la section 3.5. La représentation à l'aide de fonctions géométriques permet de réduire le nombre de valeurs nécessaires pour modéliser l'enveloppe spectrale. Leur utilisation n'est justifiée que dans le cadre d'une volonté de compression des données relatives au son.

Nous avons choisi de représenter l'enveloppe spectrale par une série de points, car cette représentation nous semble être la plus intuitive pour le musicien. En effet, l'utilisation des coefficients de filtre impose des contraintes. La modification d'un seul coefficient entraîne la modification de l'enveloppe entière pour toutes les fréquences. De plus ces coefficients n'ont à priori aucune signification sur le plan musical ou sur le plan de la perception.

Le point important à noter concernant le modèle CNSS et l'enveloppe spectrale est son indépendance vis à vis de la densité spectrale. En effet, l'information relative à la densité spectrale peut certainement être donnée et sauvegardée par les variations très localisées d'énergie dans le spectre. Les modèles représentant des sons fortement bruités seulement par une approximation de l'enveloppe spectrale, ne prennent pas en compte les propriétés du son relatives à la densité spectrale. A l'opposé, les modèles qui décrivent ces sons par le spectre complet non approximé et non lissé, retiennent ces deux informations à la fois et ne permettent pas de modifier l'une sans modifier l'autre. Le modèle CNSS, en définissant la couleur comme l'enveloppe spectrale lissée, et indépendamment la distribution des fréquences, permet le traitement séparé de ces deux propriétés.

2.4.9 Liens entre le modèle CNSS et les modèles de résidu

Les paramètres mathématiques de contrôle de la distribution des fréquences utilisés dans le modèle CNSS permettent non seulement de proposer d'infinies possibilités de distributions différentes, mais aussi de définir les fréquences des sinusoïdes composant le signal comme des valeurs non aléatoires, et se ramener ainsi aux modèles existants.

En considérant une bande de fréquence de largeur équivalente à la moitié de la fréquence d'échantillonnage ($\Delta F = \frac{F_s}{2}$, c'est-à-dire 22050Hz dans la plupart des cas), le nombre de fréquences N égal au nombre de casiers M et la largeur relative L nulle, les fréquences ne sont plus des variables aléatoires. Leur valeur fixée est telle que l'écart entre deux fréquences successives est constant. Dans le cas où ce nombre de fréquences vaut la moitié de la fenêtre d'analyse, nous nous ramenons exactement au cas de la transformée inverse de Fourier, employée par des

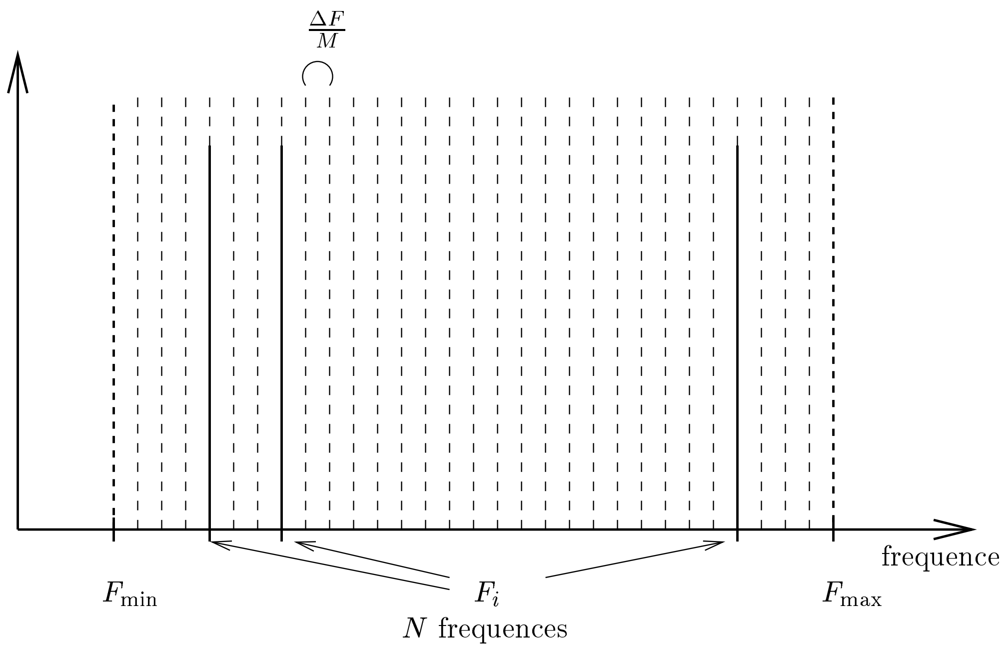

FIG. 2.17 – Exemple de tirage de fréquences contrôlé par les paramètres du modèle CNSS : ici, le nombre de casiers est très large ($M \to \infty$), ce qui revient à effectuer un tirage uniforme sur toute la largeur de la bande ΔF.

Nom	Notation	Type	Domaine de définition
Nombre de casiers	M	entier	$[1; \infty]$
Largeur de la bande de fréquences	ΔF	entier	$]0; \frac{F_s}{2}]$
Borne inférieure de la bande de fréquences	$F_{\min}$	entier	$]0; \frac{F_s}{2} - \Delta F]$
Nombre de sinusoïdes	N	entier	$[1; M]$
Largeur de la PDF dans chaque casier	L	réel	$[0; 1]$
Longueur de la fenêtre	W_s	entier	$[1; \infty[$
Largeur de la PDF des phases	P	réel	$[0; 1]$
Enveloppe spectrale	C	ensemble de points	$[0; 1] \times [0; \frac{F_s}{2}]$

TAB. 2.5 – Récapitulatif des paramètres du modèle CNSS.

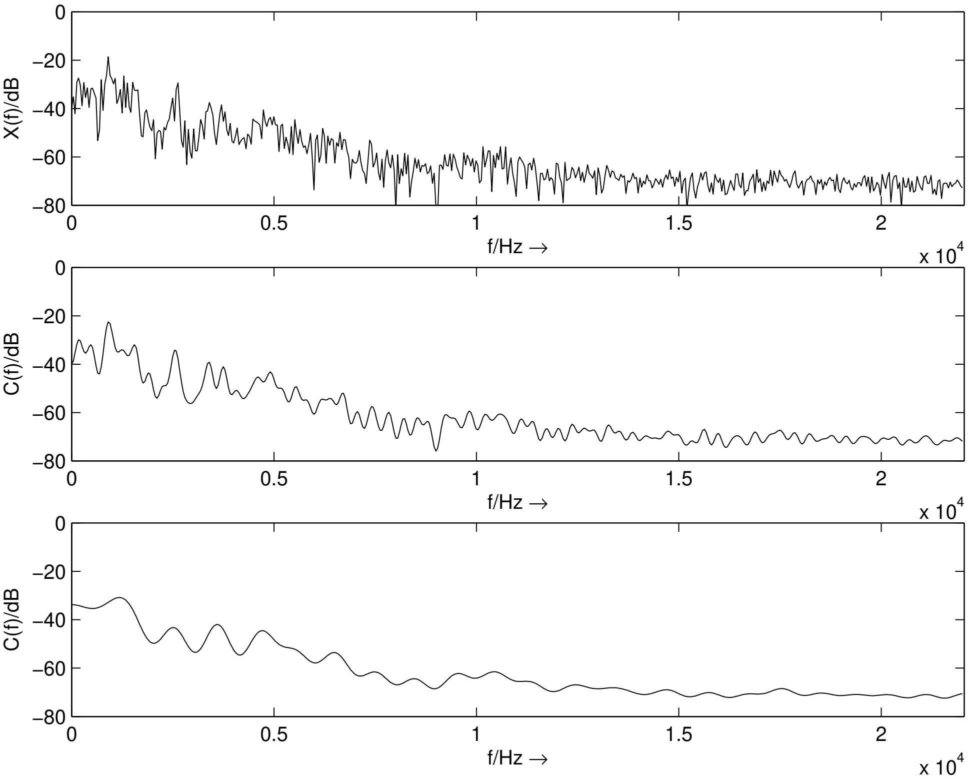

Fig. 2.18 – Illustration de la force du lissage de l'enveloppe spectrale : la figure du haut représente le spectre à court terme d'un son de voix chuchotée, qui est légèrement lissée dans la figure du milieu, et plus fortement dans le figure du bas.

modèles de résidu, comme par exemple le modèle SMS :

$$N = M = \frac{W_s}{2}, \quad L = 0, \quad \Delta F = \frac{F_s}{2} \tag{2.49}$$

Contrairement aux modèles de résidu basés sur la transformée inverse de Fourier, le modèle CNSS permet de modifier le nombre de sinusoïdes N nécessaires tout en préservant les caractéristiques liées à la distribution de l'énergie dans le spectre. En effet, si le nombre de sinusoïdes N et le nombre de casiers M sont égaux, et si la largeur de la distribution des casiers L est fixée à 1, alors les fréquences sont régulièrement réparties sur la bande de fréquence. Il est donc possible en modifiant le nombre de sinusoïdes (et le nombre de casiers) de changer l'écart régulier entre les fréquences, tout en respectant le modèle des bruits thermiques :

$$N = M, \quad L = 0 \tag{2.50}$$

Les figures 2.14 et 2.15 illustrent cette propriété. L'intérêt de ce contrôle se situe non seulement sur le plan musical, puisqu'il reste lié à la densité spectrale, mais aussi sur le plan de l'efficacité de la synthèse.

Une autre possibilité offerte par le modèle CNSS est la distribution uniforme des fréquences sur toute la bande considérée. Dans ce cas, il suffit de définir M très grand ($M \to \infty$). Le nombre de casiers est alors très important, comparé au nombre de sinusoïdes N. La définition des casiers impose la contrainte que deux fréquences ne peuvent être choisies trop proches, cette résolution fréquentielle étant imposée par la taille du casier. Or celle-ci est définie par le rapport entre la largeur de la bande et le nombre de casiers. Si ce dernier nombre est considéré comme infini, cette taille tend vers une valeur très petite (théoriquement vers 0). La contrainte imposée par les casiers s'avère donc inutile. Ainsi, quelle que soit la valeur de la largeur de la distribution L, les fréquences seront aléatoirement choisies uniformément. La figure 2.17 illustre une telle distribution.

En dehors de ces trois possibilités de configuration qui permettent de se rapporter à des approches classiques (partie stochastique du modèle SMS, bruit thermique et distribution uniforme des fréquences), le modèle CNSS permet par les paramètres statistiques proposés de contrôler directement la distribution des fréquences et ne fait donc que donner la possibilité d'envisager des sons composés de sinusoïdes dont les fréquences sont aléatoirement distribuées. Le modèle CNSS étend les modèles existants. Il est donc également possible d'appliquer ce modèle pour représenter les sons de manières similaires à celles envisagées par certains modèles existants, comme la partie stochastique du modèle SMS.

2.5 Transformations musicales possibles avec le modèle CNSS

Le modèle CNSS, que nous avons développé, définit plusieurs paramètres mathématiques. Dans cette section, nous détaillons les transformations permises par ce modèle et donc les propriétés perceptives, voire musicales, engendrées par des changements de valeur de ces paramètres. Il est utile de rappeler que les modèles existants, destinés à la représentation de bruits, ne proposent comme paramètre de contrôle que la couleur, c'est-à-dire l'enveloppe spectrale lissée, et la dynamique (l'intensité perçue). Nous allons voir que le modèle CNSS en envisage de nombreux autres, même si ils restent encore expérimentaux.

2.5.1 Densité spectrale

Les expériences psychoacoustiques établissent que l'oreille humaine est sensible au nombre de sinusoïdes utilisées pour synthétiser une bande de bruit. Cette sensation est expliquée par la perception des fluctuations d'intensité dans le cas de bandes étroites, et par la perception de différences trop importantes entre les fréquences, liée à la résolution fréquentielle du système auditif, pour les bandes plus larges. Nous étudions ici l'influence des paramètres du modèle CNSS sur ces deux propriétés, d'une part les fluctuations d'intensité, et d'autre part les distributions de différence entre les fréquences voisines.

Contrôle des fluctuations d'intensité

Nous avons proposé une étude expérimentale de l'influence des paramètres de contrôle de la distribution des fréquences dans [HDC01]. L'étude théorique associée repose sur les relations établies par W. Hartmann, et détaillées dans la section 2.4.4, en particulier l'équation (2.31). Celle-ci décrit les fluctuations d'intensité en fonction du nombre de sinusoïdes N, du nombre de casiers M et de la largeur de bande ΔF. Cette fonction $\mathcal{V}$ est croissante par rapport au nombre de sinusoïdes N, puisque les variations sont en $1-\frac{1}{N}$ (voir l'équation (2.31)). De même, cette fonction est aussi croissante par rapport au nombre de casiers M. La démonstration de ces variations est moins évidente. Nous ne proposons ici que la courbe des variations de cette fonction sur la figure 2.20, pour des valeurs de N, τ et ΔF fixées.

Ainsi, en considérant une bande de bruit étroite, les fluctuations d'intensité perçues par le système auditif peuvent être augmentées ou diminuées en modifiant le nombre de sinusoïdes et/ou le nombre de casiers. Pour confirmer cela expérimentalement, nous avons produit des bandes de bruits et nous avons mesuré les fluctuations d'intensité sur un grand nombre de réalisations de signaux. Ces expériences, dont les résultats sont représentés par les figures 2.21 et 2.22, appuient la théorie donnée par l'équation (2.31) et confirment les modifications possibles d'intensité à partir des paramètres N et M du modèle CNSS.

Contrôle des différences entre les fréquences

La perception de la densité spectrale est liée à la différence entre deux fréquences voisines composant le signal. Comme dans le modèle CNSS, les fréquences sont aléatoirement distribuées selon des fonctions de densité de probabilité contrôlées par les paramètres N, M et L, nous pouvons étudier les fonctions de densité de probabilité des différences de fréquences par rapport à ces mêmes paramètres. Ces fonctions sont établies en utilisant les propriétés données dans la section 1.2.5.

Le premier cas considéré est le cas où le nombre de sinusoïdes N est égal au nombre de casiers M, et où la largeur de distribution est fixée à 1. La fonction de densité de probabilité des différences des fréquences, représentée par la figure 2.23 est donnée par l'équation :

$$\begin{aligned}
\Delta \in [0; \frac{\Delta F}{M}], \qquad \rho(x) = & \ \frac{2M}{\Delta F^2}\Delta \\
\Delta \in [\frac{\Delta F}{M}; \Delta F], \qquad \rho(x) = & \ \frac{2M}{(M-1)\Delta F} - \frac{2M}{(M-1)\Delta F^2}\Delta
\end{aligned} \tag{2.51}$$

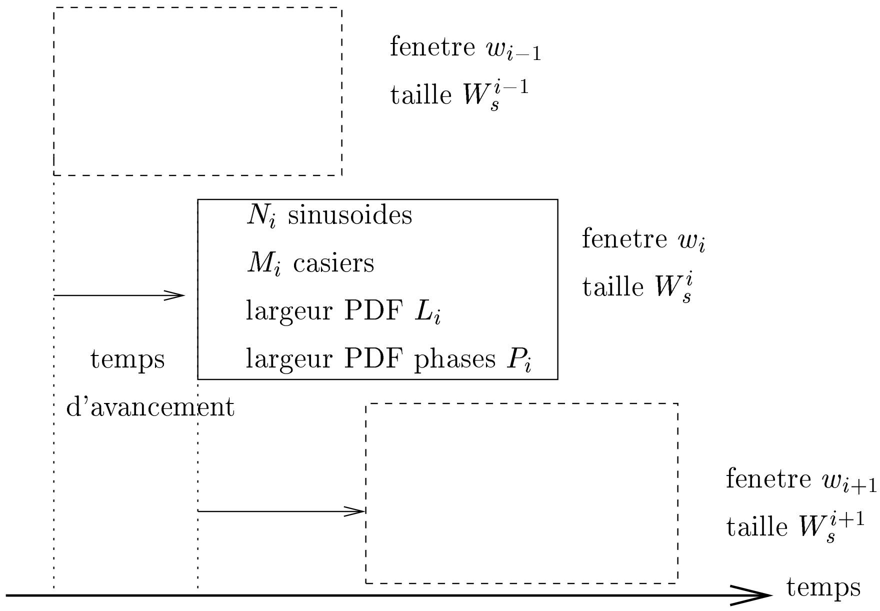

FIG. 2.19 – Illustration du principe du modèle CNSS : le signal est représenté par des fenêtres successives définies par des paramètres mathématiques. La taille W_s de ces fenêtres est un paramètre du modèle.

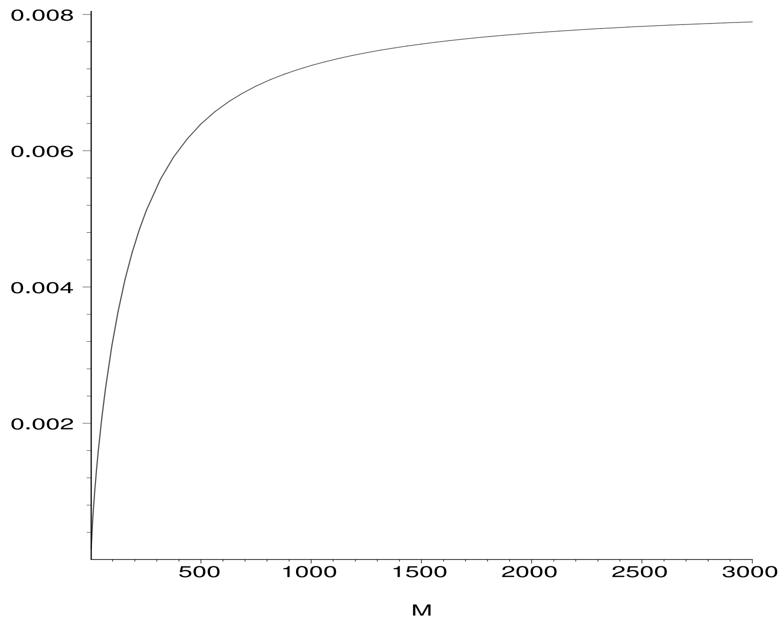

FIG. 2.20 – Évolutions des fluctuations d'intensité en fonction du nombre de casiers M, en fixant $N = 500$, $\Delta F = 20000$Hz et $\tau = 0.003$s : les variations sont croissantes.

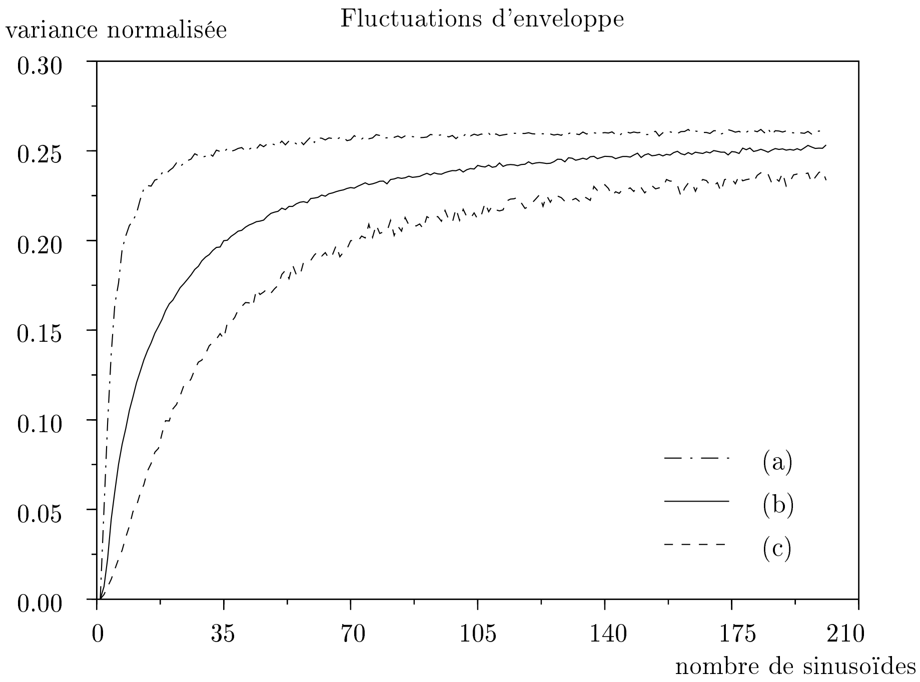

FIG. 2.21 – Variations des fluctuations d'intensité en fonction du nombre de sinusoïdes N (1000 réalisations de 65536 échantillons du signal, de largeur $\Delta F = 2000$ Hz), lorsque (a) M est infini, (b) $M = N$ et $L = 1$ (c) $M = N$ et $L = 0$.

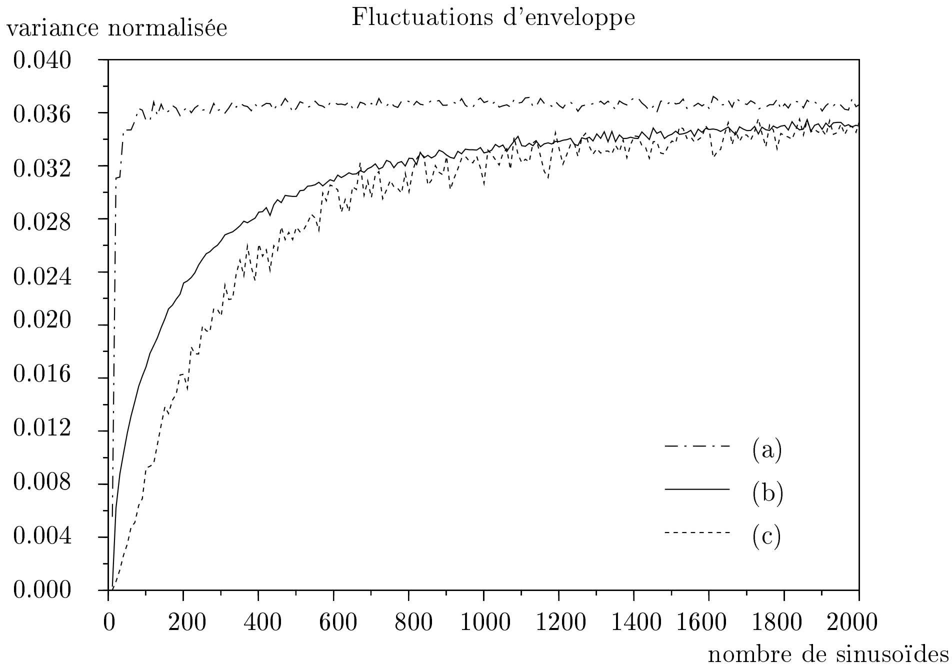

FIG. 2.22 – Variations des fluctuations d'intensité en fonction du nombre de sinusoïdes N (1000 réalisations de 65536 échantillons du signal, de largeur $\Delta F = 20000$ Hz), lorsque (a) M est infini, (b) $M = N$ et $L = 1$ (c) $M = N$ et $L = 0$.

Cette fonction de densité est associée aux différences entre *toutes* les fréquences, et pas seulement les fréquences voisines. C'est pour cela que leur domaine de définition s'étend sur toute la bande ΔF. L'écart le plus probable est $\frac{\Delta F}{M}$, ce qui ramène à l'écart le plus probable entre deux fréquences consécutives. Si nous considérons à présent la fonction des différences entre deux fréquences voisines, nous pouvons dire que l'écart minimum sera quasiment nul, et l'écart maximum sera $2\frac{\Delta F}{M}$. En effet, ces deux cas correspondent à deux fréquences sur deux casiers voisins, respectivement choisies aux deux extrémités de leur casier. Ainsi, une diminution du nombre de casiers M (ou de sinusoïdes N ici) aura pour conséquence une augmentation de l'écart le plus probable, mais également une augmentation de l'écart maximum possible. Sur le plan de l'écoute, ces deux augmentations rendent plus probable l'existence de *trous* d'énergie dans le spectre, dont la taille est supérieure à la résolution fréquentielle du système auditif, et qui sont donc perceptibles.

Il est possible de généraliser la fonction de densité de probabilité des différences entre deux fréquences voisines au cas d'une largeur de distribution L inférieure à 1. Dans ce cas, la fonction ρ', définie sur l'intervalle $[[\frac{(1-L)\Delta F}{M};\frac{(1+L)\Delta F}{M}]$, dépend donc de L et s'écrit :

$$\begin{array}{ll} \forall x \in [\dfrac{(1-L)\Delta F}{M};\dfrac{\Delta F}{M}], & \rho'(x) = \dfrac{(L-1)}{L^2}\dfrac{M}{\Delta F} + \dfrac{M^2}{L^2\Delta F^2}x \\ \forall x \in [\dfrac{\Delta F}{M};\dfrac{(1+L)\Delta F}{M}], & \rho'(x) = \dfrac{(L+1)}{L^2}\dfrac{M}{\Delta F} - \dfrac{M^2}{L^2\Delta F^2}x \end{array} \tag{2.52}$$

Le cas limite de cette distribution est le cas où $L = 0$. Dans ce cas, la fonction ρ' est une fonction de Dirac.

Cette équation montre que cette largeur L ne modifie pas la valeur de la différence la plus probable qui reste $\frac{\Delta F}{M}$, mais modifie l'écart maximum entre deux fréquences voisines. Elle est illustrée par la différence entre les figures 2.23 et 2.25, représentant les fonctions de densité de probabilité des différences entre toutes les fréquences, avec respectivement $L = 1$ et $L = 0.5$. L'influence de ce paramètre sur les écarts est cependant moins forte que celle du nombre de casiers M, puisque ce paramètre permet certes de modifier en moyenne les écarts entre les fréquences, mais ces écarts restent de toute façon bornés :

$$(1-L)\frac{\Delta F}{M} < F_{i+1} - F_i \leq (1+L)\frac{\Delta F}{M} \tag{2.53}$$

Nous pouvons illustrer cette influence par les figures 2.21 et 2.22. De plus, si le nombre de casiers M est supérieur au nombre de sinusoïdes N, la valeur de M a une plus grande incidence sur les écarts entre les fréquences. Et si ces deux paramètres sont égaux, le paramètre L a un autre effet qui sera décrit dans la section suivante.

Le cas où le nombre de sinusoïdes N est inférieur au nombre de casiers M est plus complexe à exprimer mathématiquement, car il doit non seulement prendre en compte la distribution des fréquences à l'intérieur de ces casiers, mais aussi les probabilités de choisir des casiers consécutifs, ou de laisser un certain nombre de casiers vides. Nous pouvons toutefois parvenir à justifier l'influence de la différence entre ces deux paramètres. Elle provient essentiellement de la possibilité de ne pas choisir de fréquence à l'intérieur de i casiers consécutifs. La probabilité n'est nulle que pour $i > M - N$. Évidemment, plus cette différence est importante, plus la probabilité de ne pas choisir i casiers consécutifs est importante. Comme par ailleurs, l'écart le plus probable entre deux fréquences consécutives reste $\frac{\Delta F}{M}$, une augmentation du nombre de casiers M entraîne donc d'une part une diminution de la valeur de l'écart le plus probable, mais également une augmentation de la probabilité pour des écarts plus importants :

$$0 < F_{i+1} - F_i \leq (2 + M - N)\frac{\Delta F}{M} \tag{2.54}$$

La figure 2.24 propose les répartitions des écarts des fréquences consécutives, calculées à partir de 100000 réalisations, pour plusieurs nombres de casiers.

Le cas limite de cette augmentation du nombre de casiers revient au cas où il peut être considéré comme infini. Comme nous l'avons déjà vu, ce cas est équivalent à une distribution uniforme des fréquences sur toute la bande ΔF. Ainsi, les écarts entre les fréquences sont représentés par la figure 2.23, et, en particulier, les écarts entre deux fréquences successives ont une probabilité maximale pour une valeur nulle, et ne sont pas bornés, pouvant théoriquement aller jusqu'à la largeur de la bande :

$$\forall(i,j), \quad 0 < F_{i+1} - F_i \leq \Delta F \tag{2.55}$$

En conclusion de cette étude, nous avons donc montré que les variations des paramètres N et de M du modèle CNSS permettent de modifier les fluctuations d'intensité et d'introduire des écarts d'énergie dans le spectre, et donc de contrôler la densité spectrale d'une bande de bruit. Dans le cas d'une bande étroite (au sens des résultats des expériences psychoacoustiques de W. Hartmann [HMGB86]), les effets induit par les deux paramètres sont quasiment équivalents, puisqu'une augmentation des deux paramètres, simultanée ou séparée, provoque une augmentation des fluctuations d'intensité et donc de la densité spectrale. Par contre, dans le cas d'une bande plus large, l'augmentation du nombre de casiers M permet d'augmenter les probabilités d'obtenir des écarts plus importants entre les fréquences, et donc de donner la sensation d'une densité spectrale plus faible. L'effet de ce paramètre semble donc contradictoire, selon la largeur de la bande de fréquence considérée.

Toutefois, le nombre de casiers M est directement lié à l'écart des fréquences le plus probable. Si le nombre de sinusoïdes N est égal ou très proche de ce paramètre, il décide donc directement de la densité spectrale, puisque les écarts entre les fréquences sont réguliers. Par contre, si ces deux valeurs sont distinctes, la différence $M - N$ apparaît finalement être la valeur qui décide de la probabilité de trous importants dans le spectre, puisque nous avons vu que les écarts sont distribués entre 0 et $(2 + M - N)\frac{\Delta F}{M}$. Nous pouvons donc dire pour résumer que, dans le cas de bandes larges, le nombre de casiers M définit la densité spectrale en fixant l'écart entre les fréquences le plus probable, alors que la différence entre M et le nombre de sinusoïdes N, en fixant l'écart maximum possible, définit la proportion d'écarts supérieurs à cet écart le plus probable. Les paramètres M et N permettent donc de contrôler la densité spectrale moyenne, alors que la différence entre ces deux paramètres $M - N$, permet de contrôler les variations de la densité spectrale le long du spectre.

Taille des écarts perceptibles

Les transformations liées à la densité spectrale posent le problème de la taille des écarts entre deux fréquences, perçus par l'oreille humaine. Des résultats existent lorsque deux sinusoïdes pures sont mises en présence. Mais nous nous plaçons ici dans le cas d'un son dense, dont le spectre présente une forte variation d'énergie très localisée. A notre connaissance, aucune expérience n'a été menée dans de telles conditions. Il n'est donc pas possible à priori

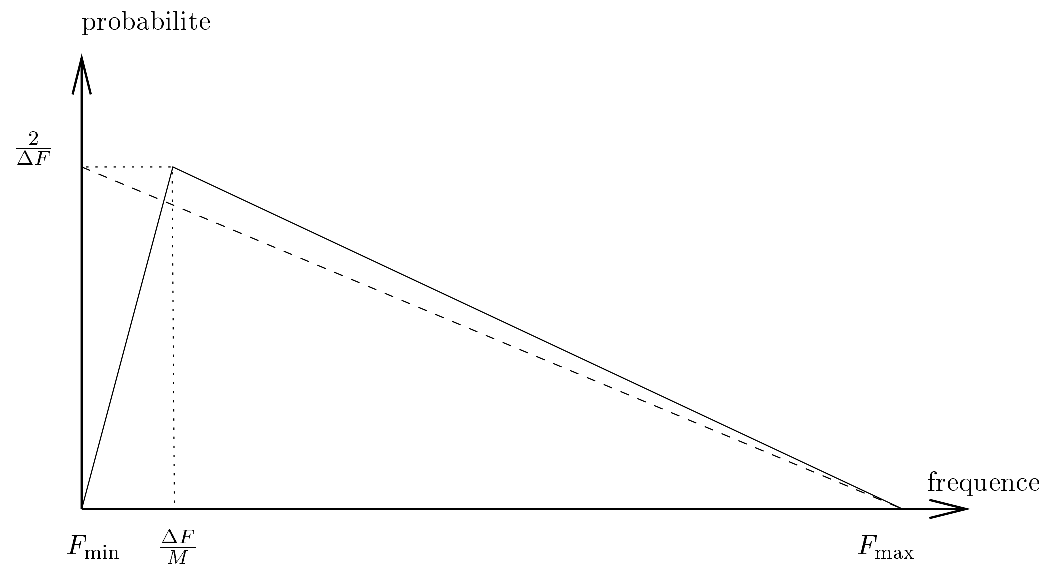

FIG. 2.23 – Fonction de densité de probabilité des différences de N fréquences lorsqu'elles sont aléatoirement choisies uniformément ($L = 1$) dans M casiers. En hachuré, la courbe correspond au cas limite où le nombre de casiers est infini $M \to \infty$.

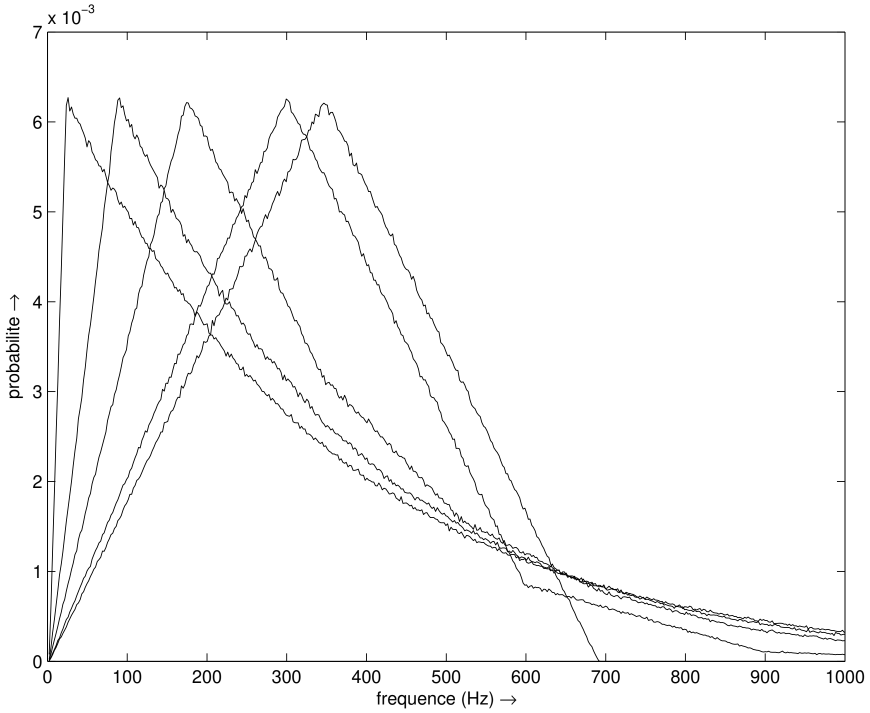

FIG. 2.24 – Fonction de densité de probabilité expérimentale (calculée sur 100000 réalisations) des différences de N fréquences lorsqu'elles sont uniformément choisies dans M casiers ($N = 64$, $\Delta F = 22000$Hz, $L = 1$), respectivement de gauche à droite, pour des valeurs de M de 1024, 256, 128, 74, 64.

de connaître la taille minimale des écarts qui sont perçus. Ce problème se ramène à celui du nombre de sinusoïdes nécessaires pour représenter un son dont la densité spectrale est maximale, c'est-à-dire équivalente à celle d'une bande de bruit produite par tirage aléatoire d'échantillons et filtrage. Nous avons vu dans la section 2.4.8 que le nombre de sinusoïdes supposé nécessaire pour représenter W_s échantillons d'une bande de bruit de largeur ΔF, de densité maximum est $\frac{\Delta F W_s}{F_s}$. Pour un tel bruit, un écart supérieur à $\frac{F_s}{W_s}$ est susceptible d'être perçu par l'oreille.

Conclusion pour les transformations musicales liées à la densité spectrale

Le modèle CNSS permet donc d'effectuer deux transformations originales sur une bande de bruit [HDC02c]. La première consiste à contrôler la densité spectrale moyenne en modifiant de la même façon le nombre de casiers M et le nombre de sinusoïdes N, ce qui a pour conséquence soit de varier les fluctuations d'intensité perçues dans le cas de bandes étroites, soit de créer ou de supprimer des *trous* réguliers d'énergie perceptibles dans le spectre dont la taille est supérieure à la résolution fréquentielle du système auditif.

La seconde transformation consiste à modifier la répartition des sinusoïdes en faisant varier la différence entre le nombre de casiers M et le nombre de sinusoïdes N, ce qui aura tendance à générer aléatoirement et irrégulièrement des *trous* dans le spectre. Ces distributions d'écarts, si ceux-ci sont suffisamment importants, traduisent les impressions de mouvement dans le spectre du bruit.

2.5.2 Harmonicité

Le paramètre noté L du modèle CNSS définit la largeur relative de la fonction de densité de probabilité uniforme des fréquences à l'intérieur de chaque casier. Son influence, même si elle existe, est faible en ce qui concerne la densité spectrale. Par contre il définit indirectement la régularité des écarts entre les fréquences des sinusoïdes composant le signal. L'équation (2.52) donne les différences entre deux fréquences consécutives en fonction de L et du nombre de casiers M (le nombre de sinusoïdes N étant égal à M). Cette fonction est représentée par la figure 2.26. Elle montre que lorsque ce paramètre vaut 0, les écarts entre les fréquences sont constants (la fonction de densité de probabilité ρ' est une fonction de Dirac). Plus ce paramètre prend une valeur proche de 1, plus les écarts sont distribués. Dans le cas d'un son de densité spectrale moyenne faible, c'est-à-dire représenté par un nombre de sinusoïdes et un nombre de casiers inférieurs au seuil correspondant à la densité spectrale maximale ($N_{\max}$), l'écart entre les fréquences peut correspondre à une fréquence audible et donc à une hauteur qui sera perçue. Or cette hauteur sera d'autant plus nette que ce paramètre L prendra une valeur proche de 0. C'est pourquoi nous pouvons définir L comme un coefficient d'harmonicité. Sur le plan musical, ce paramètre permet donc de passer, pour un bruit de densité spectrale moyenne non maximale, d'un son de liquide ($L \approx 1$) à un son d'insecte ou de machine ($L \approx 0$).

L'influence de ce paramètre n'est perceptible que lorsque le son n'est pas de densité spectrale moyenne importante ($N \approx M$ et $M << N_{\max}$). Si la densité spectrale moyenne est trop proche du maximum, l'écart entre les fréquences est trop faible pour pouvoir être perçu puisqu'il n'entre pas dans l'intervalle des fréquences audibles. De même, si des zones du spectre sont de densité spectrale importante $N << M$, les fréquences ne seront plus régulièrement espacées, et le son ne sera plus harmonique.

2.5.3 Impulsivité

Dans les calculs de fluctuations d'intensité effectués pour modéliser la perception de la densité spectrale (voir la section 2.4.4), les phases sont considérées comme des variables aléatoires uniformément distribuées entre 0 et 2π, conformément au modèle des bruits thermiques. Ces fluctuations ne suivent pas les mêmes lois si cette distribution des phases est modifiée. Pour comprendre ces variations, il suffit de se placer dans le cas limite où les phases prennent exactement la même valeur 0. Toutes les sinusoïdes composant le signal s'ajoutent alors au même instant, dans l'exemple choisi à $t = 0$:

$$x[0] = \sum_{n=1}^{N} a_n \tag{2.56}$$

Ce choix implique donc des pics d'intensité dans le domaine temporel qui seront perçus comme des transitoires de très courtes durées. La figure 2.27 montre la représentation temporelle d'un tel signal. Ces impulsions provoquent une discontinuité dans l'enveloppe et sont donc perceptibles. Le modèle CNSS introduit comme paramètre la largeur P de la fonction de densité de probabilité des phases, qui permet de contrôler les probabilités que plusieurs phases prennent la même valeur et donc la force des impulsions. En effet, l'intervalle des phases aléatoires est de $P2\pi$ et la probabilité qu'une phase soit de valeur ϕ_0 est donc de $\frac{1}{P2\pi}$. Plus la largeur P est proche de 0, plus le nombre de phases égales sera important, et donc plus le nombre d'amplitudes des sinusoïdes a_n qui s'ajouteront à un instant pour former une impulsion sera grand. Le paramètre P définit donc la force de l'impulsion. Le lien avec les fluctuations d'intensité peut être fait puisque, plus les impulsions sont importantes, plus les mesures des fluctuations d'intensité donnent des valeurs grandes [Har97].

Lorsque la largeur P est maximale ($P = 1$), les phases sont uniformément choisies entre 0 et 2π, conformément au modèle du bruit thermique. Comme cet intervalle de valeur correspond au plus large possible, les fluctuations d'intensité devraient être les plus faibles. Pourtant des expériences montrent qu'il est possible, pour certaines valeurs de phases, de produire des bandes de bruit dont les variations d'intensité sont encore plus faibles, menant au *low-noise noise* . Nous avons essayé de définir les distributions de phase qui pourraient permettre de pouvoir réduire les fluctuations d'intensité jusqu'à ce seuil. Nous n'y sommes pas encore parvenus car les calculs sont complexes. Les seules alternatives se confirment donc être des formules empiriques comme celles de l'algorithme de Schroeder, décrit dans la section 4.1, ou l'algorithme itératif de Pumplin [Pum85]. Pour le modèle CNSS, les bruits formés par une distribution uniforme des phases entre 0 et 2π seront les bruits de plus faibles fluctuations d'intensité qu'il sera possible de représenter. Les transformations liées aux fluctuations seront donc limitées de la même manière. Toutefois, l'intérêt de cette baisse des fluctuations réside essentiellement sur les propriétés de masquage de la bande de bruit [HP91]. Il reste donc limité sur le plan musical, mais est utile en psychoacoustique.

Lorsque la largeur P de la fonction de densité de probabilité des phases est inférieure à 1, une impulsion est localisée dans chaque fenêtre du signal. Si le son synthétisé présente une impulsion régulière (une par fenêtre par exemple), et si la période est relative à une fréquence audible, cette impulsion va donner une sensation de hauteur. La figure 2.27 montre un tel exemple : les impulsions y sont régulièrement espacées.

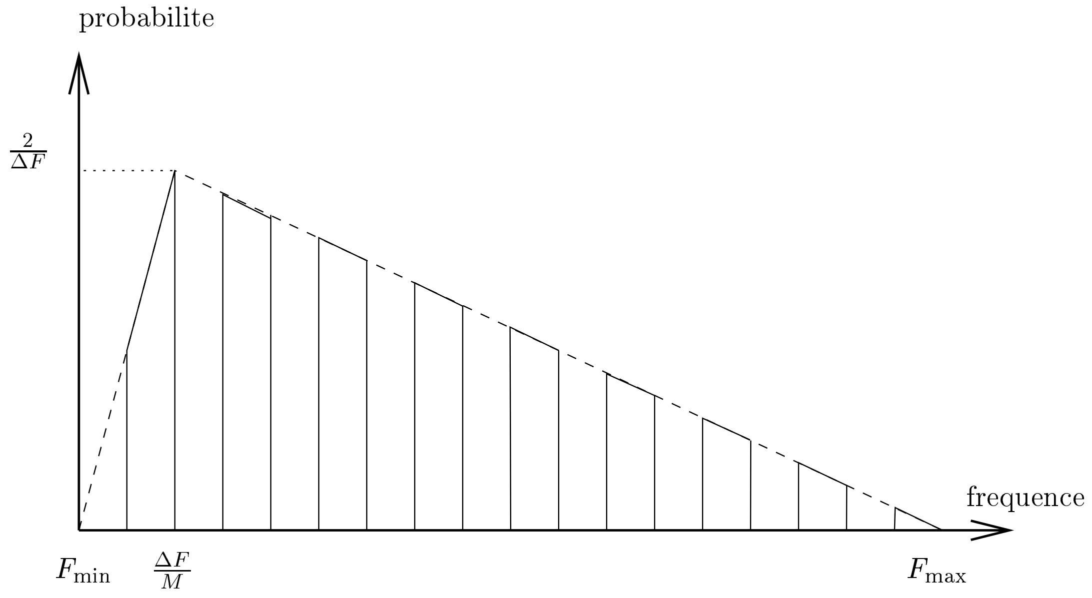

FIG. 2.25 – Fonction de densité de probabilité des différences de N fréquences lorsqu'elles sont aléatoirement choisies dans M casiers. Par rapport à la figure 2.23, la largeur de la distribution L à l'intérieur de chaque casier vaut 0.5.

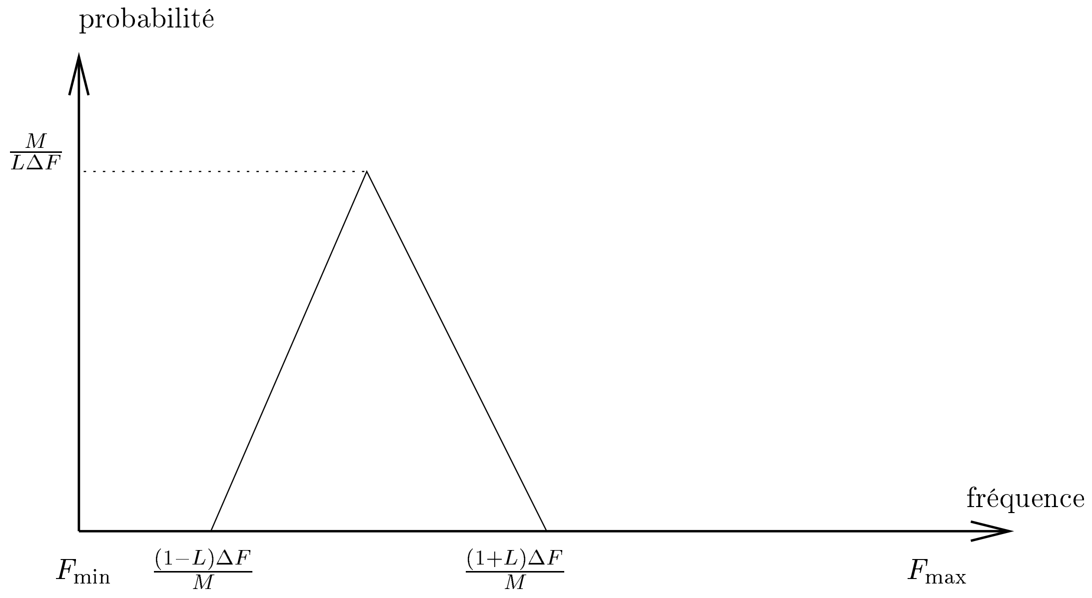

FIG. 2.26 – Fonction de densité de probabilité des différences de N fréquences lorsqu'elles sont aléatoirement choisies dans M casiers ($N = M$) avec une largeur de distribution uniforme L à l'intérieur de chaque casier.

2.5.4 Modification de la bande de fréquence

Le modèle CNSS définit comme paramètre la largeur de la bande ΔF du son considéré. Les transformations liées à ce paramètre sont nombreuses. Il est ainsi possible de translater la bande considérée, indépendamment des autres paramètres, notamment par rapport à la couleur. La figure 2.28 illustre ce principe. L'intérêt de cette transformation concerne les bandes de bruit qui provoquent une sensation de hauteur liée à leur limite de fréquence inférieure ou supérieure dans le cas de bandes larges, ou à leur fréquence centrale dans le cas de bandes étroites (voir la section 1.5.4). Une telle translation permet dans ce cas de modifier la(les) hauteur(s) perçue(s).

La modification de la largeur de la bande impose quelques contraintes supplémentaires. En effet, un changement de largeur, en laissant le nombre de sinusoïdes et de casiers constant, implique une variation de la densité spectrale moyenne, puisque, d'une part les fluctuations d'intensité seront plus faibles, et d'autre part l'écart le plus probable entre les fréquences est modifié. Pour préserver les propriétés auditives du son liées à la densité spectrale, il est donc nécessaire de préserver le rapport $\frac{\Delta F}{M}$ constant. Modifier cette largeur ΔF impose donc de modifier le nombre de casier M. Mais comme la différence $M - N$ est aussi reliée à un élément perceptif (par la répartition des écarts des fréquences), le nombre de sinusoïdes doit également être changé.

Ainsi, si la largeur de la bande est variée d'une quantité δF, les nombres de casiers et de sinusoïdes, notés respectivement M' et N' deviennent :

$$\Delta F' = \Delta F + \delta f \tag{2.57}$$

$$M' = M + \frac{M(\Delta F + \delta f)}{\Delta F} \tag{2.58}$$

$$N' = N + \frac{M(\Delta F + \delta f)}{\Delta F} \tag{2.59}$$

Cette transformation peut donc être rendue indépendante des autres paramètres perceptifs, contrairement aux autres modèles qui ne permettent cette modification que par un changement d'enveloppe spectrale.

2.5.5 Changement d'échelle temporelle

Le changement d'échelle temporelle est une transformation très utile mais qui est difficile à effectuer parfaitement. Ce processus doit impérativement préserver les caractéristiques propres aux bruits du signal original. Le modèle CNSS permet de faire ces transformations [HDC03c]. Les modèles existants, tels que la représentation de la partie stochastique du modèle SMS [SS90], permettent l'allongement ou la réduction de la durée des sons en modifiant le temps d'avancement des fenêtres de synthèse (*hop size*), ou en modifiant la taille de ces fenêtres.

La première solution impose une limite dans le cas de l'allongement. Si l'analyse est effectuée avec un taux d'avancement H ($0 \leq H \leq 1$), l'allongement de la durée ne peut pas être supérieur à un certain pourcentage de la durée initiale d, qui dépend de H :

$$d' \leq \frac{d}{H} \tag{2.60}$$

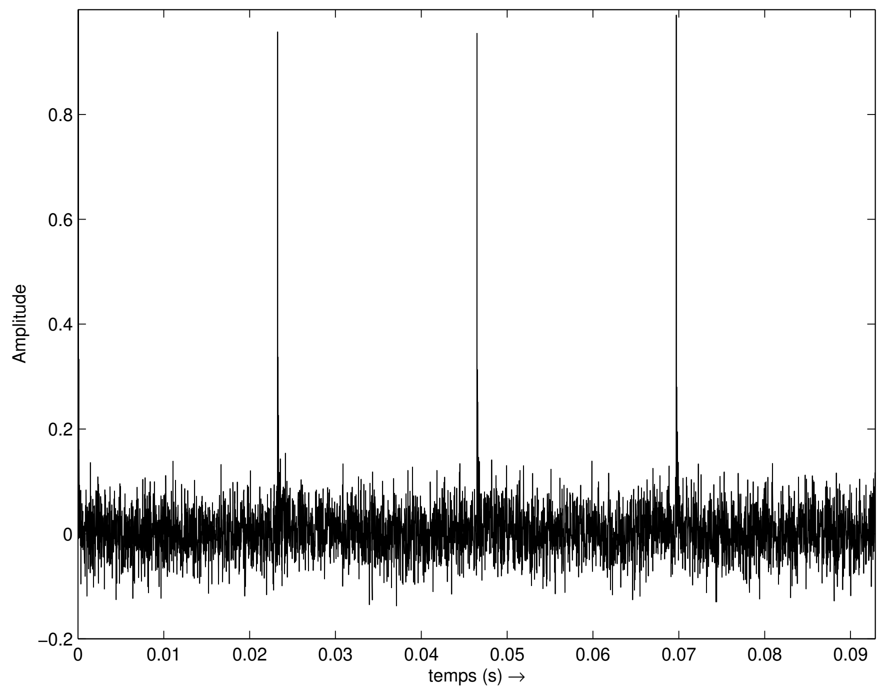

Fig. 2.27 – Bruit généré à partir de 4 fenêtres de 1024 échantillons produits par addition de 512 sinusoïdes dont les fréquences sont aléatoirement choisies sur 4096 casiers définis sur l'intervalle $[0; 22050]$Hz, et dont les phases sont nulles.

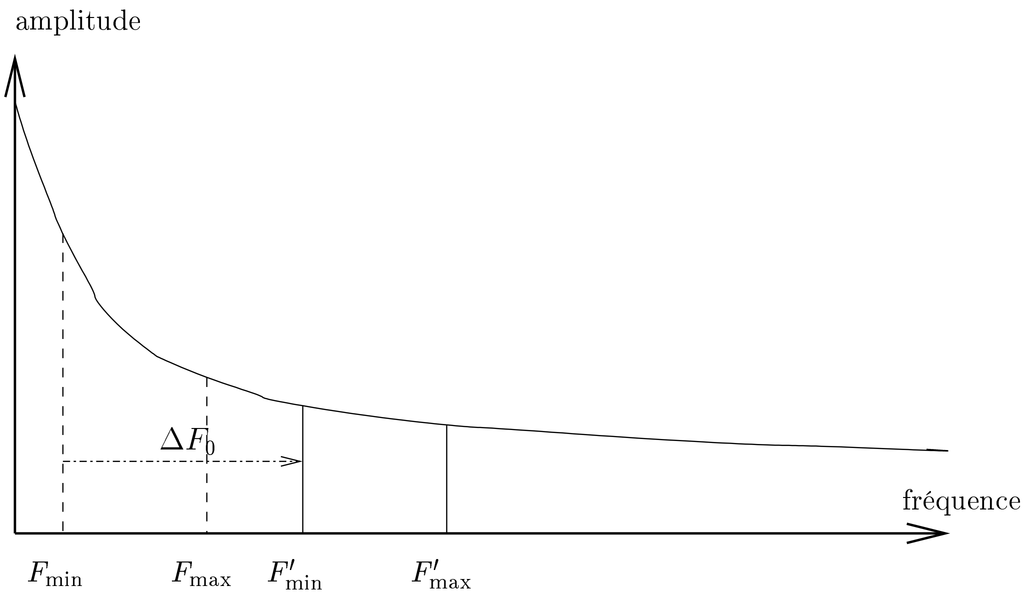

Fig. 2.28 – Illustration de la translation d'une bande de bruit par le modèle CNSS : la couleur du son n'est pas modifiée.

Par exemple, si le taux d'avancement à l'analyse est de 50% ($H = 0.5$), la durée ne peut être au maximum que multipliée par 2 ($d' \leq 2d$). Quelle que soit la valeur de ce taux d'avancement, l'allongement est forcément limité.

La seconde solution, qui consiste à agrandir la fenêtre de synthèse, impose une interpolation de l'enveloppe spectrale. Or, comme ce paramètre est le seul paramètre du modèle (avec parfois le volume), il contient toutes les informations du spectre et doit donc être précis. Si l'interpolation est trop importante, l'enveloppe spectrale est dégradée à un tel point qu'elle provoque des artefacts audibles et fait perdre au son synthétisé certaines de ses propriétés perceptives. Par ailleurs, cette augmentation de fenêtre implique une augmentation du temps de synthèse puisque cette taille impose directement le nombre de sinusoïdes utilisées pour la génération du son.

Par ailleurs, nous montrons dans la section 4.5 que le chevauchement des fenêtres de synthèse implique des variations des propriétés statistiques du son, et provoque donc des artefacts audibles qui le distordent. Les expériences que nous avons menées ont d'ailleurs confirmé les études théoriques proposées. Nous proposons plusieurs solutions qui peuvent être mises en application avec le modèle CNSS [HDC02a] et qui sont également détaillées dans la section 4.5.

La méthode de synthèse pour le modèle CNSS autorise plusieurs approches pour modifier la durée d'un son. Cependant, les propriétés perceptives liées à la densité spectrale dépendent en partie du nombre de sinusoïdes. Ce nombre doit donc rester constant au cours du temps. Or si les fenêtres se recouvraient partiellement, il serait différent pour les segments de sons d'une fenêtre et ceux composés par le chevauchement de plusieurs fenêtres. La densité spectrale varierait donc au cours du temps. C'est pourquoi nous choisissons un taux d'avancement constant, soit qui est vaut 1 (le recouvrement est nul), soit qui représente la moitié de la taille des fenêtres ($H = 0.5$). Les transformations temporelles seront donc effectuées sans modifier ce taux.

Pour développer le principe des manipulations temporelles avec le modèle CNSS, nous nous restreignons ici aux étirements (la durée du son est augmentée), sachant que les contractions sont basées sur le même principe. Nous notons α le coefficient de dilatation, défini par :

$$d' = \alpha d \tag{2.61}$$

où d représente la durée du signal initial et d' la durée du signal après étirement temporel. Concernant le modèle CNSS, la technique la plus simple et la plus efficace concerne les durées de son qui sont des multiples de la taille de la fenêtre d'analyse, notée W_s^a. Dans ce cas, le nombre de fenêtres de synthèse, noté $\mathcal{N}_f$, est simplement fixé par ce rapport et augmenté dans le cas d'un étirement (diminué dans le cas d'une contraction) :

$$\exists k \in \mathbb{N} \backslash d' = kW_s^a \Longleftrightarrow \mathcal{N}_f = k \tag{2.62}$$

Cette opération présente deux avantages : elle n'impose pas de temps de calcul supplémentaire, puisque le même nombre de sinusoïdes est synthétisé, et elle ne porte pas préjudice à la perception, du fait du tirage aléatoire des fréquences et des phases pour définir de nouvelles sinusoïdes composant chaque fenêtre de synthèse. Cette technique n'est par contre pas conseillée pour des modèles du type de la partie stochastique du modèle SMS, dont la synthèse est basée sur la transformée inverse de Fourier, pour lesquels seul le spectre de phase est aléatoirement défini. Pour ces modèles, reproduire plusieurs fenêtres de synthèse provoque des spectres d'amplitude strictement équivalents, et a tendance à générer des sons de type "métalliques".

Si la durée du son désirée ne vérifie pas cette condition, elle peut s'écrire sous la forme :

$$\exists k \in \mathbb{N} \backslash d' = kW_s^a + k_0 \tag{2.63}$$

où $k_0 \in \mathbb{R}$ et $0 < k_0 < 1$. Dans ce cas, le nombre de fenêtre est toujours donné par k, mais les tailles W_s^s des fenêtres de synthèse sont également modifiées :

$$W_s^s = (1 + \frac{k_0}{k})W_s^a \tag{2.64}$$

Toutefois, pour le modèle CNSS, une augmentation de la taille de la fenêtre s'accompagne forcément d'une augmentation du nombre de casiers et du nombre de sinusoïdes, et donc d'une augmentation du temps de calcul à la synthèse. En effet, comme nous l'avons déjà dit, cette taille de fenêtre est liée à la densité spectrale et donc au nombre de sinusoïdes et de casiers composant le signal. La définition basique de la densité spectrale établit un rapport linéaire entre cette taille W_s et le nombre de sinusoïdes N. Cela implique que, si la taille de la fenêtre W_s est multipliée par un facteur α, alors les nombres de sinusoïdes N et de casiers M doivent être multipliés par ce même facteur α pour pouvoir conserver les propriétés perceptives liées à la densité spectrale.

L'augmentation de la taille de la fenêtre impose aussi une interpolation de l'enveloppe spectrale, mais qui a peu d'incidence sur la qualité du son vu qu'elle est supposée très lissée, contrairement au paramètre des autres modèles. De plus, ces inconvénients restent mineurs puisque l'augmentation de la taille de la fenêtre sera toujours limitée, le facteur $\frac{k_0}{k}$ étant généralement faible vu que k_0 est défini inférieur à 1 et le nombre de fenêtres k est important dès que la durée du son original est raisonnable.

L'opération de changement d'échelle temporelle avec le modèle CNSS est donc de meilleure qualité que celle effectuée sous les modèles existants pouvant être appliqués aux bruits, tout en nécessitant un temps de calcul pour la synthèse à peu près constant. En effet, la génération de nouvelles fenêtres de signal est effectuée à partir du même nombre de sinusoïdes, donc ce nombre par unité de temps ne varie pas. La seule consommation de ressources supplémentaires provient de l'agrandissement des tailles de fenêtres. Mais comme l'incrémentation du nombre de fenêtres est privilégiée par rapport à la taille des fenêtres, le temps de calcul ne varie que très peu.

Les limites de cette transformation proviennent du traitement des transitoires présentes dans le signal. Plusieurs modèles ont été proposés autorisant des transformations dans le domaine temporel en traitant séparément ces transitoires, comme le modèle STN [VM98]. Le modèle CNSS ne les modélisent pas véritablement pour l'instant. La seule approche encore expérimentale est liée à la distribution des phases. Pourtant, l'allongement ou le raccourcissement de la durée d'un son doit tenir compte de ces variations d'énergie, en les préservant. L'allongement réaliste d'un choc reste un choc. Pour cela, des méthodes adaptées sont proposées dans le chapitre 3 pour pouvoir localiser ces brusques variations d'énergie. Ainsi, une fois que les transitoires sont détectées, il est possible de segmenter le signal à l'analyse, en définissant les fenêtres comportant une transitoire. A la synthèse, si une transformation d'échelle temporelle est effectuée, cette fenêtre n'est ni démultipliée (allongement) ni supprimée (contraction). La figure 2.9 illustre cette approche. En appliquant cette technique souvent répertoriée [Mas96], le modèle CNSS permet d'envisager les changements d'échelle temporelle pour les sons bruités contenant des transitoires.

2.5.6 Échelle microscopique et macroscopique

Les tailles des fenêtres d'analyse et de synthèse ne sont à priori pas des paramètres qui ont un intérêt musical. Leur rôle est lié à la qualité et la précision de l'analyse, et, dans de nombreux cas, à l'efficacité de la synthèse. Les expériences menées pour tester la synthèse CNSS, notamment avec des musiciens, nous ont convaincus de son intérêt. Lorsque cette taille est réduite (inférieure à quelques centaines de millisecondes), les variations sont si rapides qu'elle permet de manipuler le son à l'échelle microscopique, en définissant un son bruité. C'est évidemment la base du modèle qui permet de représenter et de manipuler ainsi ces sons complexes. Mais lorsque cette taille de fenêtre devient plus importante, et que le nombre de sinusoïdes est restreint, le système auditif peut arriver à percevoir chaque fréquence de chaque sinusoïde qui compose le signal sous la forme de hauteurs. L'échelle de manipulation devient macroscopique, puisque ce n'est plus vraiment le son qui est entendu, mais plutôt la musique créée par la suite de hauteurs induites. Ces mélodies, ou suite d'accords, sont formées aléatoirement selon les paramètres de distribution des fréquences, ou plutôt, dans ce cas, de notes.

2.5.7 Couleur

Dans le modèle CNSS, la couleur est donc l'enveloppe spectrale lissée. Elle fait partie des paramètres de nombreux modèles spectraux [DCM99a, Ser97b], même si ce n'est pas toujours exactement sous cette forme. Les transformations permises par le modèle CNSS sont les mêmes que celles décrites par l'application de ces modèles [Mar99]. La plus classique consiste à modifier l'enveloppe spectrale de façon à changer la coloration d'un bruit, permettant par exemple de passer d'un bruit blanc à un bruit rose. Plus généralement, cette modification d'enveloppe revient à filtrer le signal. Dans le cas du modèle CNSS, cette transformation peut donc s'effectuer indépendamment des autres modifications possibles telles que le changement de densité spectrale, d'harmonicité, d'impulsivité, etc....

2.5.8 Volume

Comme dans tous les modèles spectraux, le volume perçu peut être aisément modifié en multipliant les amplitudes de toutes les sinusoïdes par la même constante réelle γ. Si cette constante est inférieure à 1, le son est atténué, alors, que si elle est supérieure à 1, le son est amplifié.

$$\forall n \in [1; N], \quad {a'}_n = a_n \times \gamma \tag{2.65}$$

2.6 Conclusion, limites actuelles et perspectives

Le modèle CNSS que nous avons développé est un modèle spectral et statistique, entièrement original, même s'il repose sur le modèle du bruit thermique. Il représente les sons bruités par des successions de fenêtres, chacune étant générée par une somme de sinusoïdes dont les fréquences et phases sont aléatoirement définies à partir de paramètres mathématiques décrivant la distribution, et dont les amplitudes sont fixées à partir de l'enveloppe spectrale choisie. Les modèles existants, souvent destinés à la représentation de parties bruités de faible énergie pour compléter des modèles de sons harmoniques, ne proposent qu'un seul paramètre de contrôle : l'enveloppe spectrale. le modèle CNSS, lui, propose, au travers des nombreux paramètres qu'il définit, des contrôles musicaux originaux. Le tableau 2.5 récapitule toutes ces transformations. Ainsi, des propriétés démontrées expérimentalement comme éléments perceptifs, par exemple la densité spectrale, peuvent être modifiées indépendamment de la couleur. Les transformations d'échelle temporelle sont également possibles, avec une qualité accrue et de manière infinie.

Néanmoins, ce modèle reste expérimental et doit encore être amélioré, puisque tous les paramètres proposés ne sont pas encore totalement indépendants, ce qui ne rend pas toujours les transformations intuitives. Il est nécessaire dans le futur de structurer les paramètres, de la même façon que le modèle SAS structure les modèles sinusoïdaux [DCM99b]. Le lien entre le nombre de casiers et le nombre de sinusoïdes doit être notamment rendu plus facilement manipulable, sous la forme d'un seul paramètre : la densité spectrale. De plus, le modèle CNSS nécessite encore de nombreux tests psychoacoustiques pour définir quantitativement les limites des intervalles de définition des paramètres. L'exemple le plus marquant est le nombre de sinusoïdes nécessaires pour synthétiser un bruit blanc, indiscernable d'un bruit blanc synthétisé par des méthodes plus classiques.

Nom	Notation	Effets
Nombre de casiers	M	densité spectrale
Largeur de la bande de fréquences	ΔF	bande de fréquences
Borne inférieure de la bande de fréquences	$F_{\min}$	bande de fréquences
Nombre de sinusoïdes	N	densité spectrale
Largeur de la PDF dans chaque casier	L	harmonicité
Longueur de la fenêtre	W_s	échelle micro/macro
Largeur de la PDF des phases	P	impulsivité
Enveloppe spectrale	C	couleur

TAB. 2.6 – Transformations musicales permises par les paramètres du modèle CNSS.

Chapitre 3

Analyse de sons bruités

Les modèles présentés dans le chapitre précédent sont basés sur des paramètres mathématiques, liés à la perception, permettant de définir des sons du monde réel. Pour pouvoir effectuer des transformations de ces sons naturels, il est nécessaire de développer des méthodes d'analyse dans le but d'extraire les paramètres mathématiques associés au modèle à partir de la représentation temporelle des sons. La qualité de ces méthodes dépend essentiellement de la précision des paramètres extraits des sons.

Comme beaucoup de modèles spectraux ont été proposés pour le traitement des sons instrumentaux, considérés comme des sons peu bruités, des méthodes d'analyse adaptées ont été développées, basées sur des approximations liées notamment à l'absence de bruit. Il existe peu de travaux sur la représentation musicale des sons bruités. Le modèle statistique et spectral CNSS que nous présentons dans cette thèse modélise ce type de sons par une somme de sinusoïdes aléatoirement définies. La méthode d'analyse relative à ce modèle est d'autant plus complexe à mettre en oeuvre qu'elle ne peut appliquer les méthodes employées lors de l'analyse des sons non bruités. Nous proposons quelques approches complètement originales, basées sur des études statistiques des propriétés des sons.

Dans la section 3.1, nous proposons un schéma d'analyse général des sons bruités, mettant en oeuvre des méthodes de détection de sinusoïdes, présentées dans la section 3.2, ainsi que des méthodes de détection de transitoires, décrites dans la section 3.3. Ensuite, nous nous intéressons à l'analyse spécifique des parties bruitées des sons. Nous commençons par détailler dans la section 3.4 l'analyse par transformée de Fourier, base des méthodes d'analyse des modèles sinusoïdaux, et les limites qui apparaissent dans le cas de sons bruités. Dans la section 3.5, nous répertorions les méthodes d'analyse existantes de l'enveloppe spectrale, que nous pouvons appliquer pour toutes les sortes de sons. Une méthode originale d'estimation de la densité spectrale de bandes de bruits est ensuite proposée dans la section 3.6. Enfin, une application de cette technique à l'approximation des paramètres du modèle CNSS est décrite dans la section 3.7.

3.1 Principe général

D'un point de vue général, les sons naturels peuvent être considérés comme des compositions de partiels, de bruit et de transitoires, conformément à l'approche des modèles hybrides, vus dans la sous-section 2.3.3. Chacune de ces parties est caractérisée par des propriétés distinctes et impose l'utilisation de méthodes d'analyse spécifiques.

En particulier, le traitement de la partie bruitée des sons peut s'effectuer en se basant sur la représentation mathématique que nous avons proposée, et qui définit le modèle spectral et statistique CNSS. Les méthodes statistiques d'estimation des paramètres du modèle qui doivent être employées imposent des hypothèses sur le caractère aléatoire des propriétés spectrales. Ces hypothèses sont mises en défaut par les partiels et les transitoires présents dans le signal. C'est pourquoi la première étape d'une méthode d'analyse générale doit consister à différencier les parties spectrales et/ou temporelles associées aux partiels et aux transitoires.

3.1.1 Segmentation

L'analyse des parties déterministes des sons s'effectue sur le plan spectral, à partir de la transformée discrète de Fourier. De nombreux travaux ont porté sur les améliorations de cette technique, en posant des hypothèses sur le nombre de partiels représentant cette partie, leur écart, ainsi que leur variation d'amplitude et/ou de fréquence. Ainsi des méthodes d'analyse de hautes qualités ont été mises au point et validées pour ce type de sons, harmoniques ou quasi-harmoniques, et non bruités. Ces méthodes étudient les pics des spectres discrets à court terme successifs obtenus par transformée de Fourier. Il est donc nécessaire de déterminer les pics correspondant à des partiels, qui pourront être traitées par ces analyses, et ceux correspondant au bruit, qui seront traités statistiquement. Pour cela, nous présentons dans la section 3.2 les techniques existantes, ainsi qu'une nouvelle approche que nous avons développée. Une segmentation sur le domaine spectral est ainsi effectuée, puisque des pics sont associés à des partiels, alors que les parties restantes du spectre sont associées au bruit.

Les transitoires (attaques) correspondent aux fortes variations d'amplitude sur le domaine temporel, et sont généralement détectées en étudiant les évolutions de l'énergie du signal. Quelques méthodes de détection sont présentées dans la section 3.3, ainsi qu'une nouvelle technique que nous proposons. Cette segmentation sur le domaine temporel permet de traiter de façon adaptée les fenêtres de signal contenant une transitoire, de celles composées uniquement de bruit.

3.1.2 Analyse

L'analyse d'un son naturel peut s'envisager sur trois étapes consécutives, montrées par la figure 3.1. Dans le domaine spectral, les partiels sont détectés et sont traités par les méthodes d'analyse haute résolution adaptées aux suivi de partiels, qui sont à la base des modèles sinusoïdaux. Parallèlement à ce processus, une méthode de détection définit dans le domaine temporel les segments de son contenant une transitoire. Ces segments doivent être traités de manière particulière avec des techniques adaptées. Deux approches sont possibles : soit un modèle de représentation des transitoires est employé pour pouvoir effectuer des transformations propres aux transitoires, soit les segments temporels sont préservés lors de la resynthèse, permettant les transformations d'échelle temporelle (voir la figure 2.9).

Le principe majeur de ce système d'analyse est de proscrire toute opération de soustraction, contrairement aux méthodes habituelles associées aux modèles hybrides. En effet, cette

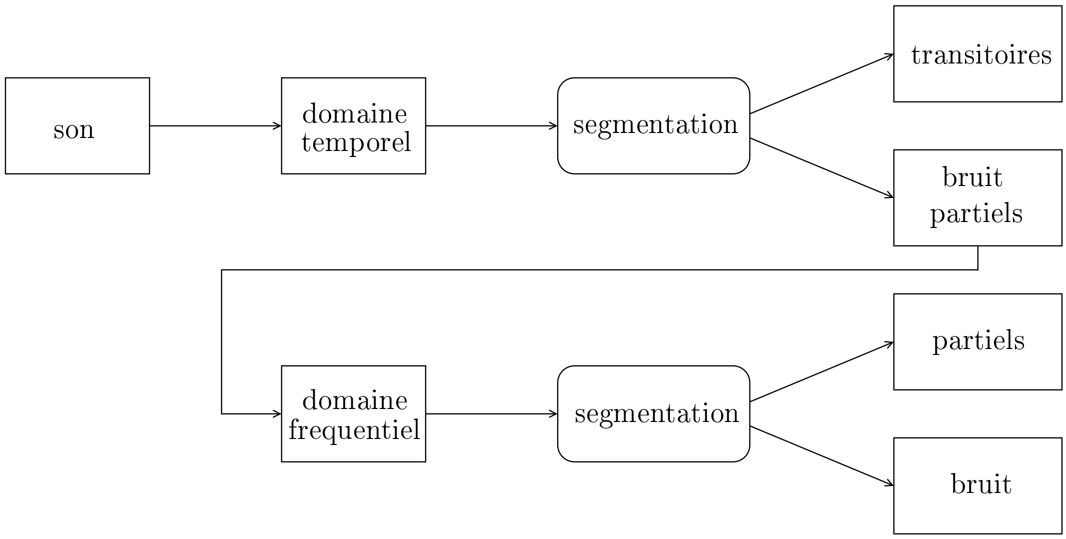

FIG. 3.1 – Illustration du principe général envisagé pour l'analyse de sons bruités.

opération impose une très haute précision pour les techniques d'analyse des paramètres mathématiques, car chaque erreur se retrouve introduite dans les signaux résiduels. Or il semble que la précision des méthodes actuelles, malgré leur très bonne performance, ne peut empêcher ces problèmes d'approximation. Nous souhaitons donc mettre en place une analyse capable d'extraire les informations propres à chaque composante du signal, sans modifier le son original.

Pour l'instant, nous nous sommes attachés à référencer les méthodes de détection de partiels et de transitoires, ainsi qu'à en développer de nouvelles, plus adaptées aux sons bruités. Nous nous sommes également concentré sur l'analyse de la composante bruit de ces signaux en nous restreignant lors de nos expériences à l'étude de sons qui ne contiennent ni partiel, ni transitoire. Dans le chapitre 5, nous introduisons quelques idées permettant de mettre en oeuvre le schéma d'analyse complet de sons bruités sur lequel nous envisageons de nous baser, et que nous avons présenté ici.

3.2 Détection de partiels

Les modèles sinusoïdaux sont destinés aux sons harmoniques ou quasi-harmoniques avec très peu de bruit. Mais les sons réels sont souvent composés d'une grand part d'énergie assimilée à du bruit. Les méthodes d'analyse associées à ces modèles sont basées sur la transformée de Fourier à court terme et consistent à connecter sur des fenêtres temporelles successives des partiels. La figure 3.2 illustre ce processus. Pour cela, tous les pics (maxima locaux) des spectres à court terme sont supposés représenter un partiel. En pratique, certains correspondent réellement à une sinusoïde, alors que d'autres sont induits par l'énergie du bruit. Un exemple est donné par la figure 3.3 qui montre un spectre à court terme d'un son de hautbois, en mettant en évidence les pics pris en considération pour le suivi de partiel. Les limites de cette approche sont imposées par la transformée de Fourier et son traditionnel compromis entre la précision temporelle et la précision fréquentielle. Les améliorations apportées à la transformée de Fourier concernent essentiellement la précision des mesures des fréquences, amplitudes et phases (voir la section 3.4), mais peu souvent la sélection de pics et la détection de partiels. Il est pourtant utile d'être capable d'estimer quels pics peuvent être considérés pour améliorer les performances du suivi de partiels, et ainsi éviter de définir des partiels incorrects.

Dans cette section, nous présentons les méthodes envisagées pour améliorer la sélection de pics dans les spectres à court terme, en mettant en évidence pour chacune les limites par rapport à la trop forte présence de bruit. Nous décrivons également une méthode que nous avons proposée [HDC02b] et améliorée [HDC03d], qui prend en compte l'information statistique contenue dans les fluctuations d'intensité et qui reste indépendante des limites de résolution de la transformée de Fourier.

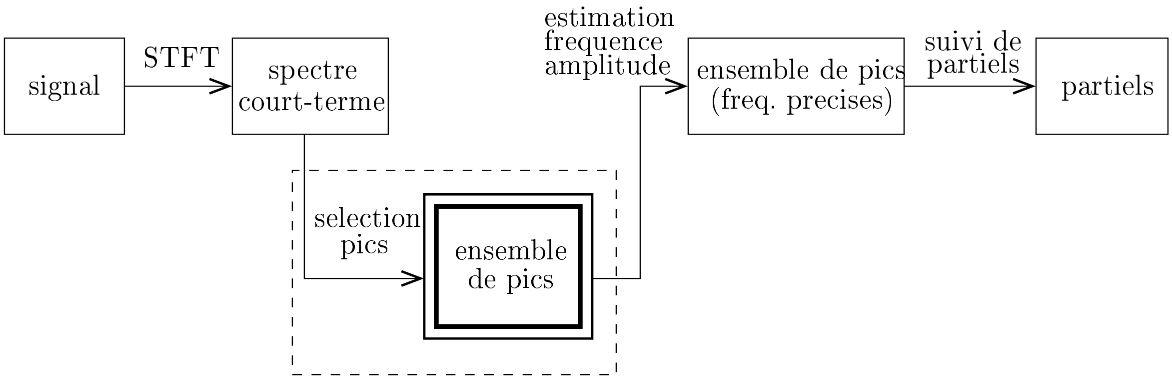

FIG. 3.2 – Illustration de l'algorithme général pour l'analyse spectrale de sons selon une modélisation sinusoïdale.

3.2.1 Méthodes existantes

Maximum local

La technique la plus classique, utilisée dans de nombreuses méthodes d'analyse, consiste à choisir tous les pics, notés $S(k)$, c'est-à-dire les maxima locaux du spectre d'amplitude à court terme :

$$\forall k \in \mathbb{N}, \quad \left\{ \begin{array}{lcl} |S(k-1)| & < & |S(k)| \\ |S(k+1)| & \leq & |S(k)| \end{array} \right. \tag{3.1}$$

Un exemple de l'application de cette méthode est proposé par la figure 3.3.

Cette technique permet d'effectuer la sélection de pics de manière rapide et efficace, mais si les partiels sont mélangés à du bruit d'énergie importante (par exemple un rapport signal sur bruit négatif, SNR < 0), les amplitudes correspondant aux partiels ne sont pas toujours supérieures aux amplitudes correspondant à l'énergie du bruit. Un local maximum n'apparaît pas toujours dans ce cas. De plus, de nombreux pics artificiels, dûs au bruit, risquent d'être détectés. Le nombre de pics peut donc être très supérieur au nombre réel de partiels présents dans le signal.

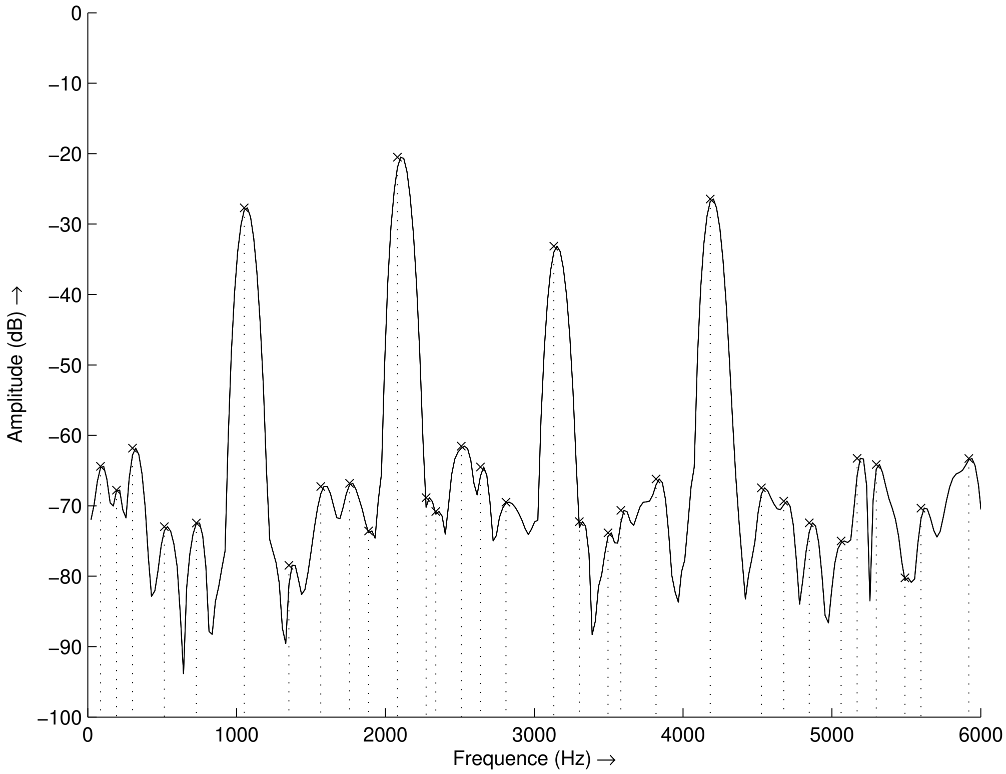

FIG. 3.3 – Exemple de la détection de pics (maxima locaux) sur une fenêtre d'un son de hautbois (1024 échantillons).

Deux améliorations basiques sont en général appliquées lors de l'utilisation de cette technique, en particulier dans le cas d'analyse de sons harmoniques. La première consiste à fixer le nombre précis de partiels que la méthode de sélection doit choisir. Cette approche peut toutefois donner de mauvais résultats car elle suppose la connaissance préalable du nombre des partiels du signal. Si ce nombre est surestimé, cette méthode choisit des pics correspondant à du bruit, alors que si ce nombre est sous-estimé, comme par exemple dans le cas de sons polyphoniques, elle ne sélectionnera pas assez de pics.

La seconde amélioration couramment employée préconise l'utilisation d'un seuil d'amplitude. Tous les pics dont l'amplitude est supérieure à cette valeur constante sont sélectionnés. Contrairement à l'amélioration précédente, le nombre de pics à sélectionner n'est pas fixé. Néanmoins, cette approche ne permet pas de résoudre les problèmes liés aux pics incorrects dûs au bruit présent dans le signal, car le seuil d'amplitude est fixe pour toutes les valeurs de fréquence. En effet, les sons naturels ont tendance à avoir plus d'énergie dans les basses fréquences. Les pics correspondant à du bruit, présents dans les basses fréquences, ont ainsi tendance à être associés à des partiels, alors que les pics dûs aux partiels et présents dans les hautes fréquences, ont tendance à être associés à du bruit.

Il est évident que cette méthode ne peut pas être utilisée avant d'avoir préalablement ajusté les paramètres tels que le seuil d'amplitude ou le nombre de partiels, qui dépendent essentiellement de la nature du son considéré. Cette méthode d'analyse ne peut donc pas être automatisée.

Nous pouvons également noter que cette technique de sélection des pics peut être aussi améliorée en utilisant la technique de rajout de zéros lors de l'application de la transformée de Fourier. Mais, de toute façon, ces améliorations ne peuvent pas permettre de réduire toutes les erreurs dues à la présence de bruit et de variations d'amplitude des pics.

Seuil d'amplitude

Une autre méthode de sélection consiste à distinguer les pics corrects correspondant aux partiels des pics dûs au bruit, en comparant la valeur $\Gamma(k)$ liée à l'importance relative de l'amplitude du pic, calculée pour tous les maxima locaux du spectre d'amplitude. Cette valeur représente la somme des valeurs absolues des différences entre le pic et ses deux minima locaux adjacents :

$$\Gamma(k) = |S(k)|_{dB} - \frac{|S(k^+)|_{dB} + |S(k^-)|_{dB}}{2} \tag{3.2}$$

où $S(k^+)$ et $S(k^-)$ représentent l'amplitude des minima locaux adjacents de chaque côté du pic.

Cette méthode suppose évidemment la définition d'un seuil pour ces valeurs Γ. Celles qui sont inférieures à ce seuil prédéfini ne sont pas sélectionnées, alors que les pics dont les valeurs Γ sont supérieures au seuil sont supposés correspondre à des partiels du signal. Ce choix arbitraire du seuil détermine donc la précision de cette méthode. Bien sûr, il peut entraîner le rejet de pics corrects, dont les amplitudes ne sont pas assez importantes comparées à celles des creux voisins. De plus, la valeur calculée $\Gamma(k)$ peut ne pas être significative dans des environnements bruités. En effet, les minima locaux adjacents peuvent avoir des amplitudes importantes si elles correspondent à de l'énergie due au bruit.

Corrélation

La caractérisation proposée dans [Rod97b] calcule la corrélation complexe entre la transformée de Fourier de la fenêtre d'analyse H centrée sur la fréquence étudiée et chaque pic S de la transformée discrète de Fourier du signal fenêtré par H :

$$\forall \omega, \quad \Gamma(\omega) = \left| \sum_{k,|\omega-\omega_k|<W} H(\omega - \omega_k) S(\omega_k) \right| \tag{3.3}$$

Le coefficient Γ est une valeur comprise entre 0 et 1. La valeur 0 indique la présence de bruit, alors que la valeur 1 indique un partiel. Théoriquement, cette méthode peut être appliquée à n'importe quel type de sons. Toutefois, elle présente quelques défauts qui sont détaillés dans [PR98b]. Nous les résumons brièvement ici :

Stationnarité du signal : La précision de cette mesure repose sur l'hypothèse selon laquelle toutes les propriétés du signal (fréquences, amplitudes) sont supposées constantes pendant la durée correspondant à la taille de la fenêtre d'analyse. Avec des sons naturels, cette hypothèse est rarement vérifiée, menant à des erreurs de détection.

Détections de composants non sinusoïdaux : La mesure appliquée aux pics correspondant aux lobes secondaires de sinusoïdes indique une sinusoïde pure.

Résolution temporelle et fréquentielle : L'utilisation de la transformée de Fourier pose le traditionnel problème de la résolution fréquentielle et temporelle. Pour pouvoir calculer Γ, une bonne précision fréquentielle, induite par une large fenêtre d'analyse, est nécessaire alors que l'hypothèse de stationnarité requiert une bonne précision temporelle, induite par une fenêtre d'analyse réduite.

Dans [PR98b], Peeters propose quelques améliorations possibles pour cette mesure :

Normalisation de la fréquence fondamentale : En posant l'hypothèse que la fréquence de modulation des signaux est corrélée à la fréquence fondamentale, le signal est rééchantillonné pour minimiser l'influence de ces modulations de fréquence.

Mesure des variations de phase : Cette autre mesure est basée sur la différence de phase à deux instants (deux fenêtres d'analyse). Ce calcul mène à des estimations différentes et est utilisé pour compléter l'approche par corrélation. Ainsi, c'est l'apport de cette technique qui permet de ne pas prendre en compte les pics incorrects correspondant aux lobes secondaires.

Soustraction de pics voisins : Chaque pic préalablement supposé comme correspondant à un partiel du signal est soustrait du spectre d'analyse pour annuler son influence sur les autres pics éventuels.

Ces améliorations sont efficaces mais restent liées aux hypothèses que les sons étudiés sont harmoniques ou pseudo-harmoniques, et qu'ils contiennent peu de bruit. De plus, elles nécessitent une connaissance préalable à propos de la présence de partiels ou à propos des variations de phase. Cette dernière condition impose l'application de l'analyse à partir de plusieurs fenêtres successives de signal.

Améliorations de la FFT

Plusieurs méthodes ont été proposées pour améliorer l'estimation des paramètres des partiels d'un signal (voir la section 2.3.5), notamment l'estimation des fréquences d'après le spectre obtenu par transformée de Fourier. La plupart d'entre elles sont présentées et testées dans [KM02]. La détection des pics peut être effectuée à partir de ces méthodes. Durant la première phase d'analyse, tous les pics du spectre à court terme sont considérés. Ensuite, une méthode haute précision d'estimation de la fréquence est appliquée à chacun de ces pics. En comparant ce résultat à la localisation fréquentielle du pic, il est possible de décider s'il est considéré comme associé à un partiel du signal. La méthode propose donc de conserver tous les pics dont la différence entre la fréquence de référence du pic et celle estimée est inférieure à la largeur d'un casier de la transformée de Fourier.

En notant k le casier choisi, $S(k)$ étant donc le maximum local associé, f la fréquence estimée par une méthode haute résolution, F_s la fréquence d'échantillonnage, et W_s la taille de la fenêtre d'analyse, ce critère de sélection peut s'écrire :

$$|f - k\frac{F_s}{W_s}| \leq \frac{F_s}{W_s} \tag{3.4}$$

Cette méthode est entièrement basée sur la transformée de Fourier. Dans des environnements bruités, le nombre de pics sélectionnés peut être assez important. Des exemples sont donnés dans la section 3.2.6, qui montrent les limites de cette approche.

3.2.2 Analyse des fluctuations d'enveloppe temporelle

A la différence des autres techniques proposées, nous nous intéressons aux sons bruités, qui ne sont évidemment pas harmoniques. La méthode que nous présentons dans cette section est basée sur la définition des bruits thermiques, rappelée dans la section 2.4.2, et plus particulièrement sur l'étude des fluctuations d 'enveloppe temporelle, abordée dans la section 2.4.5.

Nous considérons dans la suite une bande de fréquences étroite, composée d'un partiel et de bruit. Le modèle des bruits thermiques nous permet de définir ce bruit comme une somme finie de sinusoïdes dont les amplitudes, fréquences et phases sont aléatoirement fixées. Les amplitudes de ces sinusoïdes peuvent être supposées faibles par rapport à celle de la sinusoïde principale (partiel), pour que celle-ci puisse être perçue. La méthode de sélection et de détection de pics correspondant aux partiels repose sur les fluctuations d'intensité du signal étudié sur une bande de fréquences (le signal filtré). Théoriquement, ces fluctuations sont négligeables si la bande de fréquences étudiée contient un partiel, alors qu'elles sont importantes si cette même bande est seulement composée de bruit. L'estimation de ces fluctuations d'intensité peut donc permettre de détecter la présence de partiels.

Nous avons vu dans la section 2.4.5 un modèle pour représenter ces fluctuations d'intensité à partir de la variance normalisée de la puissance de l'enveloppe temporelle filtrée. L'enveloppe temporelle est filtrée pour atténuer les hautes fréquences que le système auditif ne perçoit pas [Har97]. Nous ne nous intéressons pas ici à la perception. Cette opération de filtrage n'est pas effectuée puisque nous allons être amené à ne considérer que des bandes de fréquences très étroites. Les différences entre les fréquences composant le signal, qui sont à la base de la définition de la puissance de l'enveloppe temporelle (voir l'équation (2.20)) sont donc très faibles. Le filtrage des hautes fréquences n'apparaît pas nécessaire et peut peut-être même nous priver d'informations utiles. C'est pourquoi les équations et relations établies dans cette section sont différentes de celles établies par Hartmann [Har97] et présentées dans la section 2.4.5.

Variance de la puissance de l'enveloppe

Une mesure possible des fluctuations d'enveloppe est l'expression de la variance de la puissance de l'enveloppe :

$$V(E^2) = \left\langle (E^2 - \left\langle E^2 \right\rangle)^2 \right\rangle \tag{3.5}$$

La puissance de l'enveloppe est donnée par l'équation (2.20). Comme $\langle \cos[(\omega_n - \omega_{n'})t + \phi_n - \phi_{n'}] \rangle = 0$, et cela pour tout $n \neq n'$, la moyenne de la puissance de l'enveloppe se simplifie :

$$\left\langle E^2 \right\rangle = \sum_{n=1}^{N} C_n^2 \tag{3.6}$$

Ainsi, la variance devient :

$$V(E^2) = \left\langle 4(\sum_{n=1}^{N} \sum_{n'=1}^{n-1} C_n C_{n'} \cos[(\omega_n - \omega_{n'})t + \phi_n - \phi_{n'}])^2 \right\rangle \tag{3.7}$$

En appliquant la propriété de linéarité de la moyenne, c'est-à-dire $\langle \alpha X + \beta Y \rangle = \alpha \langle X \rangle + \beta \langle Y \rangle$ avec $\{\alpha, \beta\} \in \mathbb{R}^2$, et en reprenant les notations introduites par les équations (2.23) et (2.24), l'équation précédente s'écrit :

$$\begin{aligned} V(E^2) &= 4\left\langle \sum_{n=1}^{N}\sum_{n'=1}^{n-1} C_n^2 C_{n'}^2 \cos^2[y_{nn'}] \right\rangle \\ &+ 4\left\langle \sum_{n=1}^{N}\sum_{n'=1}^{n-1}\sum_{m=1}^{N}\sum_{m'=1}^{m-1} [cos(y_{nn'} - y_{mm'}) + cos(y_{nn'} + y_{mm'})] \right\rangle \end{aligned} \tag{3.8}$$

Comme $n \neq n'$, nous pouvons simplifier par :

$$\left\langle \cos^2[y_{nn'}] \right\rangle = \frac{1}{2}$$

et

$$\left\langle \sum_{n=1}^{N}\sum_{n'=1}^{n-1}\sum_{m=1}^{N}\sum_{m'=1}^{m-1} [cos(y_{nn'} - y_{mm'})] \right\rangle = \left\langle \sum_{n=1}^{N}\sum_{n'=1}^{n-1}\sum_{m=1}^{N}\sum_{m'=1}^{m-1} [cos(y_{nn'} + y_{mm'})] \right\rangle = 0$$

La variance de la puissance de l'enveloppe vaut donc :

$$V(E^2) = 2\sum_{n=1}^{N}\sum_{n'=1}^{n-1} C_n^2 C_{n'}^2 \tag{3.9}$$

Or, le carré de cette puissance moyenne est :

$$\left\langle E^2 \right\rangle^2 = \sum_{n=1}^{N}\sum_{n'=1}^{N} C_n^2 C_{n'}^2 \tag{3.10}$$

donc la variance normalisée de la puissance de l'enveloppe, notée V_{NEP}, est :

$$V_{\text{NEP}} = \frac{V(E^2)}{\langle E^2 \rangle^2} = \frac{2\sum_{n=1}^{N}\sum_{n'=1}^{n-1} C_n^2 C_{n'}^2}{\sum_{n=1}^{N}\sum_{n'=1}^{N} C_n^2 C_{n'}^2} \tag{3.11}$$

Dans le cas d'un bruit blanc ($\forall n \in \{1, \ldots, N\}, C_n = C_0$), cette équation devient :

$$V_{\text{NEP}} = \frac{V(E^2)}{\langle E^2 \rangle^2} = 1 - \frac{1}{N} \tag{3.12}$$

Maximisation de la variance de la puissance de l'enveloppe

Nous étudions ici la variance de la puissance de l'enveloppe dans le cas de phases nulles ($\forall k, \phi_k = 0$). Ainsi, les valeurs $y_{nn'}$ peuvent s'écrire :

$$y_{nn'} = (\omega_n - \omega_{n'})t \tag{3.13}$$

L'équation (3.8) devient :

$$\begin{aligned} V(E^2) &= 2\sum_{n=1}^{N}\sum_{n'=1}^{n-1} C_n^2 C_{n'}^2 + 4\left\langle \sum_{n=1}^{N}\sum_{n'=1}^{n-1}\sum_{m=1}^{N}\sum_{m'=1}^{m-1} cos(y_{nn'} - y_{mm'}) \right\rangle \\ &+ 4\left\langle \sum_{n=1}^{N}\sum_{n'=1}^{n-1}\sum_{m=1}^{N}\sum_{m'=1}^{m-1} cos(y_{nn'} + y_{mm'}) \right\rangle \end{aligned} \tag{3.14}$$

Les deux moyennes de sommes de sinusoïdes dépendent du nombre de quadruplets (n, n', m, m') qui vérifient $y_{nn'} + y_{mm'} = 0$ et $y_{nn'} - y_{mm'} = 0$. Pour ces quadruplets, la moyenne de la somme est :

$$\langle cos(y_{nn'} + y_{mm'}) \rangle = 1 \tag{3.15}$$

Il est donc évident que plus le nombre de sinusoïdes N est important, plus le nombre de ces quadruplets est aussi important, et donc plus la valeur $V(E^2)$ est grande (V_{NEP} de même).

3.2.3 Application à la sélection de pics

L'équation précédente (3.12) montre que pour des amplitudes constantes, la mesure V_{NEP} que nous avons définie dépend uniquement de la distribution des fréquences, et en particulier du nombre de sinusoïdes N. En considérant le son analysé comme un mélange de bruit et de partiels, et en appliquant la définition du modèle du bruit thermique, le signal peut donc être complètement représenté par une somme de sinusoïdes.

Soit x le signal aléatoire correspondant au bruit blanc, suivant le modèle du bruit thermique. En considérant un mélange, noté m, composé d'une sinusoïde (partiel), notée s, dont l'amplitude a est égale à 1 et la fréquence est constante, et de ce bruit blanc x :

$$\forall t, \quad m(t) = x(t) + s(t) \tag{3.16}$$

Nous faisons l'hypothèse que le rapport signal sur bruit est nul. L'amplitude RMS $A_{\text{RMS}}^{\text{noise}}$ est :

$$A_{\text{RMS}}^{\text{noise}} = \sqrt{\langle x^2 \rangle} = \sqrt{\sum_n C_n^2} \tag{3.17}$$

où C_n est l'amplitude de la $n^{\text{ième}}$ sinusoïde associée au bruit. Comme l'amplitude RMS du bruit x est égale à l'amplitude RMS du partiel s (SNR= 0dB), nous pouvons écrire :

$$\frac{a}{\sqrt{2}} = \frac{1}{\sqrt{2}} = \sqrt{\sum_n C_n^2} \tag{3.18}$$

Lorsque N est important, les amplitudes C_n sont inférieures à a, même avec des rapports signal/bruit négatifs. Donc, en considérant tous les sons comme un mélange de sinusoïdes et de bruit blanc, nous pouvons approximer, dans une bande de fréquence étroite contenant un partiel, N à 1. Cela implique que la variance de la puissance de l'enveloppe est plus petite pour cette bande, en comparaison aux bandes de largeur équivalente, contenant seulement du bruit.

3.2.4 Méthodes d'approximation

Dans cette sous-section, nous présentons trois techniques permettant de calculer la mesure V_{NEP} décrite dans la section précédente. Dans l'équation (4.52), la variance de la puissance de l'enveloppe dynamique est calculée en moyennant sur le temps (sur les échantillons) et sur les phases aléatoires. Toutefois, la variance que nous mesurons n'est rien d'autre que la variance d'une seule réalisation du signal. Les expériences que nous avons menées montrent que les valeurs des variances peuvent beaucoup varier, selon le nombre d'échantillons utilisés pour le calcul. Les résultats peuvent ainsi être faussés et peuvent aussi faire perdre de la précision. Nous proposons trois solutions à ce problème. La première effectue une moyenne de la variance en distribuant aléatoirement les valeurs du spectre de phase du signal. La seconde calcule la variance maximale. Enfin, la troisième consiste à calculer la variance en itérant sur le nombre d'échantillons utilisés. Chacune de ces améliorations a ses avantages et ses inconvénients. Nous étudions dans la suite leur précision et leur efficacité respective.

méthode des phases aléatoires : De nombreuses réalisations peuvent être obtenues à partir du signal original en modifiant le spectre de phase. D'après la définition du modèle du bruit thermique, les phases sont supposées être des variables aléatoires uniformément distribuées entre 0 et 2π. Cette méthode consiste donc à générer de multiples spectres de phase, en donnant des valeurs aléatoires à chaque casier du spectre de phase obtenu en appliquant la transformée discrète de Fourier.

méthode des phases constantes : La variance est maximisée en modifiant le spectre de phase en mettant à zéro toutes ses valeurs ($\forall k, \phi_k = 0$).

méthode de la variation de taille : La variance est moyennée en utilisant un nombre variable d'échantillons du signal analysé lors du calcul. Contrairement aux deux autres méthodes, celle-ci ne requiert aucune modification du spectre de phase.

3.2.5 Algorithmes

Dans cette section, nous détaillons les trois algorithmes pour générer la mesure de sinusoïdalité, en se basant sur la théorie de la variance de la puissance de l'enveloppe développée dans les sections précédentes. Pour chaque description de ces algorithmes, nous utilisons trois fenêtres de signal de tailles différentes. La première fenêtre, notée w_1, est utilisée pour filtrer le signal, puisque des bandes étroites de fréquence sont analysées successivement. La seconde fenêtre, notée w_2, est utilisée pour calculer le spectre de phase du signal, qui sera modifié pendant l'analyse. La troisième fenêtre, notée w_3, sert à calculer la variance.

Filtrage fréquentiel

La théorie indique que la variance de la puissance de l'enveloppe varie de façon significative entre un signal composé d'une sinusoïde principale (partiel), et un signal composé uniquement de bruit (somme de plusieurs sinusoïdes au sens du bruit thermique). Pour pouvoir analyser un signal composé seulement d'une seule sinusoïde ou composé seulement de bruit (sans sinusoïde), seule une partie étroite du spectre du signal doit être étudiée. Ainsi, une petite fenêtre temporelle, mais aussi une petite fenêtre fréquentielle sont analysées. Les algorithmes peuvent prendre en compte la possibilité de filtrer le signal à partir de sa représentation spectrale. Une fois que la transformée de Fourier est produite, le spectre est modifié pour extraire la bande fréquentielle à analyser. Ensuite, la puissance de l'enveloppe de cette bande est calculée en appliquant la transformée de Fourier inverse. Néanmoins, il est possible d'employer d'autres algorithmes de filtrage puisque cette étape est indépendante des autres.

Méthode des phases aléatoires

Une fois que la transformée de Fourier est effectuée, il est possible de remplacer les valeurs du spectre de phase par des valeurs aléatoires, uniformément distribuées entre 0 et 2π. Ensuite, une réalisation du signal est synthétisée en appliquant la transformée inverse de Fourier. Cette phase est répétée autant de fois que nécessaire. Le nombre de réalisations est d'ailleurs un paramètre de cette méthode qui sera discuté dans la section 3.2.7.

Un algorithme plus efficace est envisageable en remarquant que le spectre d'un signal est le même que celui de son enveloppe temporelle, en ce qui concerne sa partie positive [Har97]. Ainsi, une seule transformée de Fourier peut être effectuée : l'enveloppe temporelle peut être calculée et le spectre de phase modifié durant la même étape. La figure 3.4 présente une description d'un tel algorithme. Comme pour les autres algorithmes, trois fenêtres sont définies (w_1, w_2 et w_3), avec chacune une taille éventuellement différente. Dans cet algorithme, la taille de la fenêtre w_2 utilisée pour effectuer la transformée de Fourier et pour modifier le spectre de phase, est prépondérante. Il est implicitement supposé que le signal analysé est stationnaire le temps durant cette fenêtre. Les fenêtres w_1 et w_3 ont moins d'influence sur la précision de cette méthode. En effet, les erreurs dues au nombre d'échantillons utilisés pour calculer la variance sont moyennées par le nombre de réalisations du signal, générées par les modifications du spectre de phase. L'influence de ces paramètres est expérimentée dans la section 3.2.7.

Méthode des phases constantes

De la même manière que dans les sous-sections précédentes, les valeurs des spectres de phase sont remplacées par des valeurs nulles, après avoir effectué la transformée de Fourier. Ici, la différence principale vient du fait que les spectres de phase ne sont modifiés qu'une seule fois. La figure 3.4 présente une illustration de cet algorithme. Comme précédemment, trois fenêtres différentes sont utilisées. La taille de la fenêtre w_2 est la plus importante pour la précision de la méthode, à cause de l'hypothèse de stationnarité. La fenêtre w_3 a moins d'influence car les fluctuations sont maximisées en produisant un pic d'intensité au début de la fenêtre d'analyse. La valeur de la variance est complètement liée à l'amplitude de ce pic. L'influence de ces paramètres est testée et discutée dans la section 3.2.7.

Méthode de la variation de taille

Cette méthode requiert la modification des spectres de phase du son analysé. Elle peut être implémentée sans appliquer une transformée de Fourier. D'abord, le signal est filtré et la puissance de l'enveloppe est calculée. Ensuite, la variance est définie pour différentes parties du signal, supposées correspondre à différentes réalisations du même signal. Le moyen de sélectionner différentes parties du signal vient de la possibilité de réduire, pas à pas, le nombre d'échantillons pour calculer V_{NEP}. Le nombre de réalisations est un paramètre de cette méthode et est discuté dans la section 3.2.7. La figure 3.5 propose une illustration de cet algorithme. Au contraire de deux autres méthodes, la fenêtre w_2 n'a quasiment aucune influence sur la précision. Mais la fenêtre w_3 est importante car elle définit une partie du signal qui est supposée stationnaire. La taille de la fenêtre w_3 détermine donc la précision de la mesure.

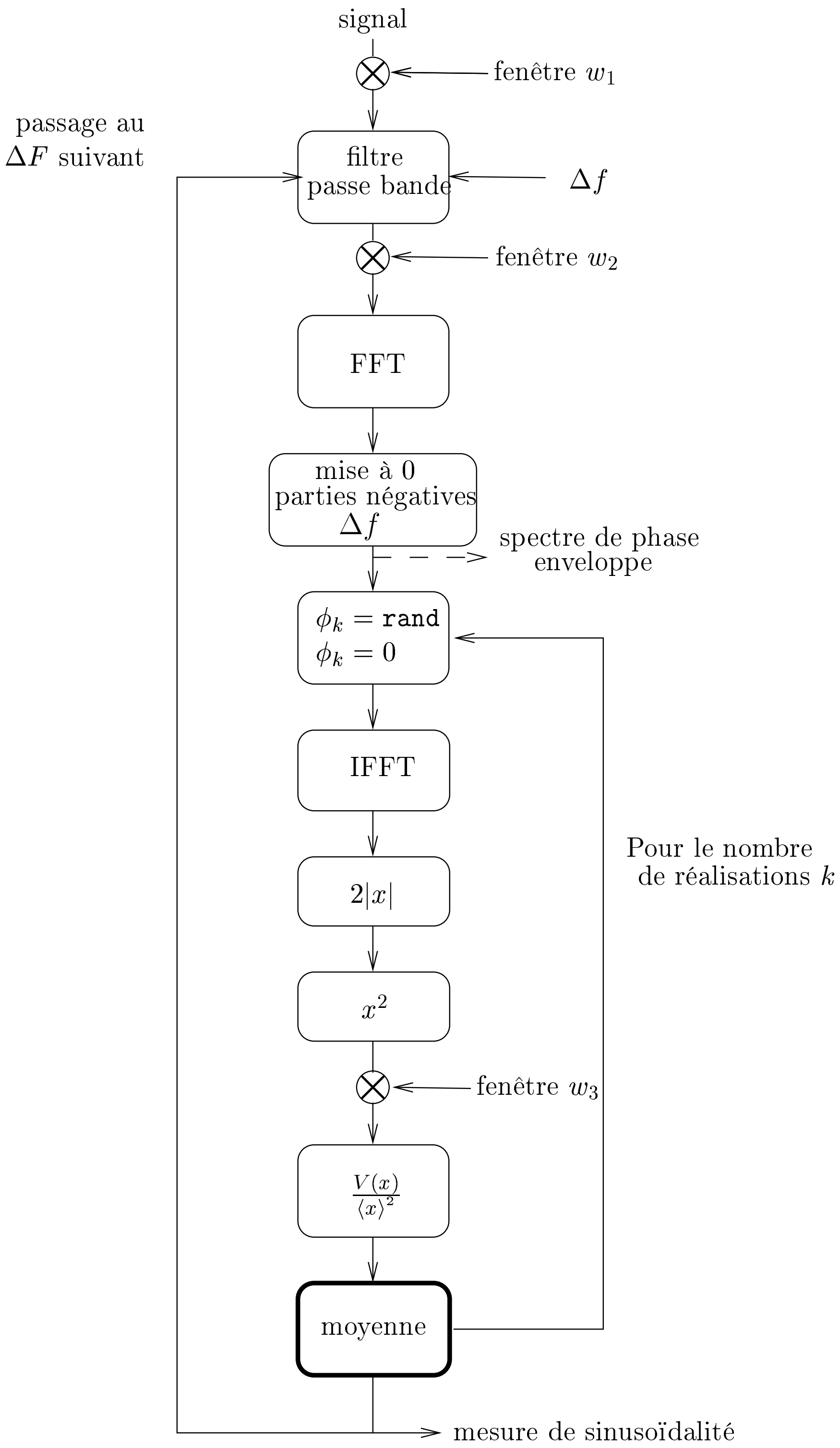

FIG. 3.4 – Algorithmes pour la méthode des phases constantes et des phases aléatoires.

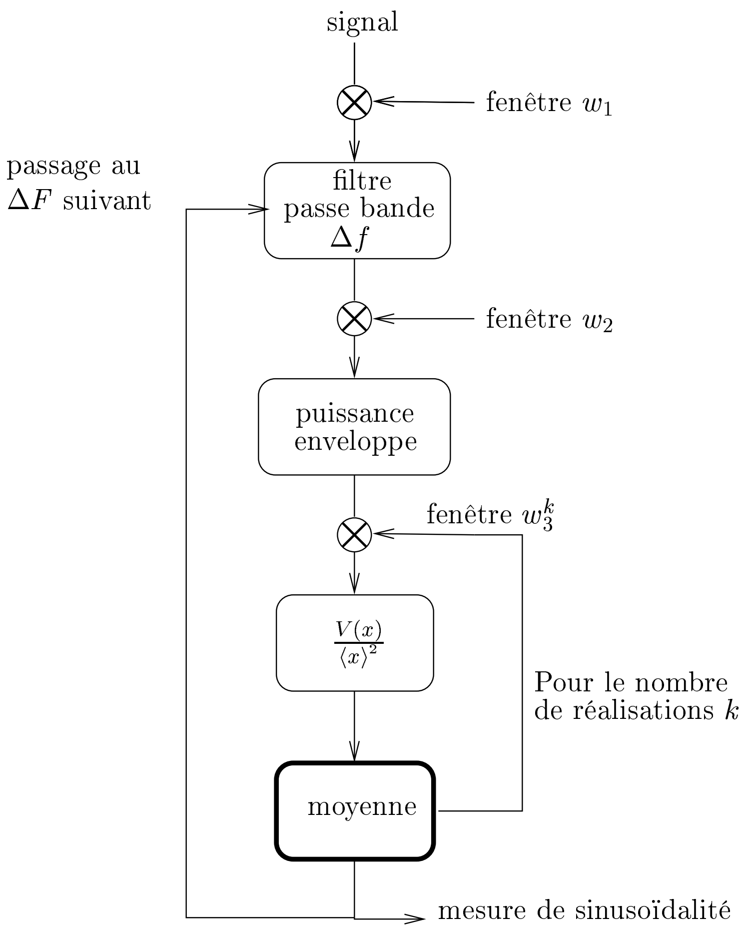

FIG. 3.5 – Algorithme pour la méthode de la variation de taille.

Algorithme de sélection de pics

Les trois algorithmes sont utilisés par un algorithme général de sélection de pics, que nous appellerons PS-algorithme. Cet algorithme consiste à générer un spectre à court terme en appliquant la transformée de Fourier sur des fenêtres de signal. Tous les maxima locaux (pics) des spectres d'amplitude successifs sont sélectionnés. Ensuite, nous appliquons un des trois algorithmes introduits précédemment pour calculer la mesure V_{NEP} à partir d'une bande étroite de fréquences autour de chaque pic considéré. Tous les pics dont la mesure associée V_{NEP} est inférieure à un seuil fixé sont finalement sélectionnés.

La valeur de ce seuil est évidemment importante. Les études théoriques (voir la section 3.2.2) montrent qu'elle est indépendante de l'amplitude générale du son analysé. Plus le seuil est important, plus le nombre de pics sélectionnés sera important. Cette remarque est confirmée par les expériences que nous avons menées dans la section suivante.

Un autre moyen d'appliquer les trois algorithmes de détection de pics présentés est de calculer V_{NEP} pour tous les points du spectre. Bien sûr, le nombre total de points sélectionnés sera dans la plupart des cas plus important de cette manière, pouvant donc entraîner la détection de partiels erronés.

3.2.6 Expériences

Nous présentons dans cette section quelques résultats des nombreuses expériences que nous avons effectuées pour tester la qualité, les avantages mais aussi les limites de notre approche originale. Nous commençons par une ou plusieurs sinusoïdes d'amplitudes et de fréquences constantes, plongées dans du bruit blanc. Ensuite, des sinusoïdes dont les amplitudes et les fréquences sont modulées sont utilisées pour préciser les possibilités des algorithmes que nous avons développés dans des cas non stationnaires. Enfin, nous terminons par un son instrumental naturel.

Tous les sons testés sont définis à partir d'une fréquence d'échantillonnage $F_s = 44100$Hz. Dans toutes les expériences menées, le bruit blanc est synthétisé en générant aléatoirement des échantillons, suivant une distribution de Gauss. L'écart-type de cette distribution σ est relié à l'amplitude RMS du son synthétisé. L'amplitude RMS d'une sinusoïde, d'amplitude a_0, est $\frac{a_0}{\sqrt{2}}$. Ainsi, le rapport signal sur bruit (SNR) vaut :

$$\mathrm{SNR} = 20 \log \frac{a_0}{\sigma\sqrt{2}} \tag{3.19}$$

Pour pouvoir présenter des résultats significatifs, nous avons procédé à un grand nombre d'expériences, pour lesquels nous avons calculé les résultats moyens. Les mesures effectuées sont répertoriées dans la table 3.1. Nous supposons ici qu'une sinusoïde détectée est correcte si sa fréquence f varie au maximum d'un casier de la transformée discrète de Fourier correspondante à la taille d'analyse :

$$f - f_0 \leq \frac{F_s}{W_s} \tag{3.20}$$

où W_s est la taille de la fenêtre (en échantillons) et f_0 est la fréquence de la sinusoïde du signal analysé.

$\overline{N}_{\text{sinus}}$	Nombre moyen de sinusoïdes détectées par fenêtre.
$\overline{N}_{\text{exact}}$	Nombre moyen de sinusoïdes détectées par fenêtre, qui sont effectivement dans le signal.
$\overline{N}_{\text{faux}}$	Nombre moyen de sinusoïdes détectées par fenêtre, qui ne sont pas présents dans le signal.
$\overline{N}_{\text{manque}}$	Nombre moyen de sinusoïdes par fenêtre qui sont présentes dans le signal, mais qui n'ont pas été détectées.

TAB. 3.1 – Détails des mesures effectuées pour les tests de détection de sinusoïdes.

Bien entendu, la précision des méthodes proposées dépend non seulement des valeurs des seuils choisis, mais également des autres paramètres de la méthode d'analyse. Une discussion est proposée dans la section 3.2.7. Nous avons essayé de montrer les résultats les plus significatifs. Mais, bien entendu, ces mesures, pour pouvoir être reproduites, doivent être effectuées avec les mêmes paramètres : d'autres valeurs peuvent donner d'autres résultats, et peut-être même dans certains cas, des meilleurs.

Tous les résultats peuvent être comparés aux autres techniques basées sur la transformée de Fourier à court terme, en ce qui concerne la précision et l'efficacité. Nous comparons d'ailleurs les résultats que nous avons obtenus à ceux présentés dans [KM02], car nous avons utilisé les mêmes paramètres d'analyse.

Partiel seul dans le bruit

Les sons analysés sont synthétisés en utilisant une seule sinusoïde dont la fréquence f_0 et l'amplitude a_0 sont constantes, mélangée avec du bruit blanc (voir l'équation (3.21)). Le bruit blanc est généré par tirage aléatoire des échantillons suivant une distribution de Gauss. Nous appliquons les PS-algorithmes sur 1500 fenêtres de signal de longueur W_s échantillons, chaque fenêtre étant donc composé de bruit blanc et d'un partiel, de fréquence différente pour chaque fenêtre. Cette fréquence varie entre 300Hz et 1600Hz et est aléatoirement choisie dans des casiers consécutifs de largeur $\frac{F_s}{4W_s}$, selon une loi uniforme. Ce processus assure que les fréquences prennent à peu près toutes les valeurs à l'intérieur de chaque casier associé à la transformée discrète de Fourier. Par ailleurs, les phases sont aussi des valeurs aléatoires, uniformément distribuées entre 0 et 2π.

$$s(n) = a_0 \sin(2\pi f_0 \frac{n}{f_S} + \phi) \tag{3.21}$$

Nous testons les trois méthodes introduites précédemment, chacune avec trois rapports signal sur bruit différents (SNR=6dB, 0dB et -6dB). La taille de la fenêtre d'analyse, supposant la stationnarité du signal, est de 1024 échantillons.

Méthode de la variation de taille

Ces expériences ont été effectuées avec une fenêtre de filtrage w_1 de 4096 échantillons. Cette taille peut être évidemment plus importante, car, pour cette méthode, elle n'impose pas de contrainte relative à la stationnarité du signal. La mesure V_{NEP} est calculée, en moyennant 100 signaux, à partir de fenêtre w_3 de taille variant donc de 1024 à $1024 - 99$ échantillons. La largeur des bandes de fréquences successivement étudiées est de 8 casiers, c'est-à-dire 345Hz. Les fréquences des partiels sont comprises entre 323Hz et 16473Hz. Les résultats sont présentés dans le tableau 3.2. Pour chaque niveau de bruit, nous avons utilisé différentes valeurs de seuil de détection.

Méthode des phases constantes

Pour cette méthode, la fenêtre de filtrage w_1 a la même taille que la fenêtre w_2 utilisée pour modifier le spectre de phase. Elle a une taille de 1024 échantillons. La largeur de la bande de fréquence étudiée est de 12 casiers, soit 517Hz. La variance est calculée à partir de 256 échantillons. La fréquence de chaque partiel est comprise entre 237Hz et 16387Hz. Les

Méthode de la variation de taille						
SNR (dB)	6		0		-6	
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$
0.35	6.919	1.000	7.601	1.000	8.137	0.999
0.3	3.827	1.000	4.379	1.000	4.823	0.995
0.25	2.294	1.000	2.711	1.000	2.991	0.975
0.2	1.687	1.000	2.041	1.000	2.144	0.927
0.15	1.503	1.000	1.829	1.000	1.515	0.751

TAB. 3.2 – Résultats de détection de pics pour une sinusoïde stationnaire, en appliquant la méthode de la variation de taille.

résultats sont présentés dans le tableau 3.3. Pour chaque niveau de bruit, nous avons utilisé différentes valeurs de seuil de détection.

Méthode des phases constantes						
SNR (dB)	6		0		-6	
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$
2.0	9.316	1.000	9.553	1.000	9.332	0.990
1.8	5.623	1.000	6.063	1.000	5.923	0.953
1.6	4.189	1.000	4.691	0.999	4.299	0.863
1.4	3.697	1.000	4.000	0.993	3.487	0.725
1.2	3.666	0.999	3.683	0.937	2.835	0.551

TAB. 3.3 – Résultats de détection de pics pour une sinusoïde stationnaire, en appliquant la méthode des phases constantes.

Méthode des phases aléatoires

Ici, le nombre de réalisations du signal obtenues par modification du spectre de phase est fixé à 100. La fenêtre de filtrage w_1 a la même taille que la fenêtre w_2 utilisée pour modifier le spectre de phase, c'est-à-dire 1024 échantillons. La largeur de la bande de fréquence étudiée est de 12 casiers, soit 517Hz. La variance est calculée à partir de 256 échantillons. La fréquence de chaque sinusoïde est comprise entre 237Hz et 16387Hz. Les résultats sont présentés dans le tableau 3.4. Pour chaque niveau de bruit, nous avons utilisé différentes valeurs de seuil de détection.

Conclusion

Les résultats présentés dans les tableaux de cette sous-section montrent qu'une sinusoïde stationnaire peut être parfaitement détectée pour des rapports signal sur bruit de +6dB et 0dB (pour chaque méthode, quelques valeurs de seuil permettent d'obtenir 100% des sinusoïdes détectées). Pour un rapport signal sur bruit de -6dB, quelques sinusoïdes ne sont pas détectées, mais cette probabilité reste faible (moins de 5% d'erreurs).

Méthode des phases aléatoires						
SNR (dB)	6		0		-6	
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$
0.5	4.829	1.000	5.033	1.000	5.321	0.998
0.475	3.653	1.000	3.931	1.000	4.239	0.994
0.45	3.259	1.000	3.600	1.000	3.882	0.975
0.425	3.093	0.996	3.471	0.993	3.687	0.948
0.4	3.013	0.987	3.393	0.975	3.507	0.908

TAB. 3.4 – Résultats de détection de pics pour une sinusoïde stationnaire, en appliquant la méthode des phases aléatoires.

méthode	FFT	parabolique	triangle	reassignement	dérivée	vocodeur de phase
$\overline{N}_{\text{sinus}}$	140.1	40.8	52.2	139.4	118.9	134.9

TAB. 3.5 – Résultats de détection de pics pour une sinusoïde stationnaire dans du bruit (SNR= 12dB), présentés dans [KM02].

La précision, et donc l'intérêt, de ces méthodes sont déterminés par le nombre de sinusoïdes qui sont détectées mais qui ne sont pas présentes dans le signal étudié. Les nombres moyens de sinusoïdes détectés par fenêtre pour des méthodes classiques sont répertoriés dans le tableau 3.5. Ces nombres sont bien plus importants par rapport au nombre moyen de pics sélectionnés en appliquant les PS-algorithmes que nous présentons. Les meilleurs résultats sont obtenus par la méthode de variation de taille, en particulier dans les cas de rapport signal sur bruit de +6dB et 0dB. En effet, dans ces conditions, moins de 2 sinusoïdes (en incluant la sinusoïde correcte) sont détectées pour chaque fenêtre. Pour les deux autres méthodes, ce nombre augmente légèrement : la méthode des phases aléatoires permet de détecter entre 3 et 4 sinusoïdes, alors que le plus mauvais résultat est obtenu avec la méthode des phases constantes avec 5 à 6 sinusoïdes. Il est toutefois important de remarquer que, même dans ce cas, le PS-algorithme obtient de bien meilleurs résultats que les méthodes classiques, en détectant au total à peu près dix fois moins de sinusoïdes.

Pour ce qui concerne les rapports signal sur bruit faibles (−6dB), le nombre de sinusoïdes manquées accroît. Les erreurs peuvent même approcher 10%, alors que le nombre total de sinusoïdes détectées reste le même. Un autre seuil peut alors être choisi pour empêcher l'absence de pics corrects, ce qui peut entraîner selon les méthodes une hausse du nombre total de sinusoïdes. Pour la méthode des phases constantes, ce nombre peut atteindre 10. Pour la méthode des phases aléatoires, il reste entre 2 et 3. Pour la méthode de la variation de taille, il s'approche à peu près de 2. Néanmoins, les PS-algorithmes restent très précis en comparaison des méthodes basées sur la transformée de Fourier.

Sons harmoniques

Dans cette partie, nous expérimentons les PS-algorithmes en analysant les sons synthétisés avec 20 sinusoïdes de fréquences f_k et d'amplitudes a_0 constantes, mélangées avec du bruit blanc gaussien (voir l'équation (3.22)). Comme le but de ces tests est de simuler des sons harmoniques, les fréquences f_k ($1 \leq k \leq 20$) de chaque sinusoïde valent kf_o, où f_0 est une valeur aléatoire uniformément distribuée dans l'intervalle $[860 - 990]$Hz. Ces valeurs assurent que les fréquences sont bien réparties tout le long du spectre audible. Les amplitudes de chaque sinusoïde sont égales et déterminées de telle manière que le rapport signal sur bruit vaut 0dB. Les phases ϕ_k sont aussi aléatoirement choisies selon une loi uniforme entre 0 et 2π.

$$s(n) = \sum_{k=0}^{N-1} a_0 \sin(2\pi k f_0 \frac{n}{F_s} + \phi_k) \tag{3.22}$$

Les PS-algorithmes ont été appliqués avec les mêmes paramètres que ceux employés lors des expériences décrites dans les paragraphes précédents. Pour chaque méthode, 1000 signaux composés chacun de 20 sinusoïdes ont été aléatoirement synthétisés et testés. Nous montrons dans le tableau 3.6 les moyennes de ces résultats.

Méthode de la variation de taille				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.25	37.664	20.000	17.664	0.000
0.2	37.129	20.000	17.129	0.000
0.15	35.967	19.997	15.970	0.003
Méthode des phases constantes				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
1.8	77.462	20.000	57.462	0.000
1.6	75.754	19.982	55.772	0.018
1.4	74.505	19.824	54.681	0.176
Méthode des phases aléatoires				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.5	72.517	20.000	52.517	0.000
0.4	68.405	19.427	48.978	0.573

TAB. 3.6 – Résultats de détection de pics pour des sons synthétiques harmoniques (20 sinusoïdes) mélangés à du bruit blanc gaussien (SNR = 0dB).

Ces résultats confirment la qualité de ces méthodes et sont parfaitement comparables aux données des tableaux de la section précédente 3.2.6. Le nombre de sinusoïdes qui ne sont pas détectées mais qui sont néanmoins présentes dans le signal, est approximativement le même. Ainsi, nous pouvons conclure que le nombre de sinusoïdes n'altère pas la précision des méthodes dans ces conditions.

De plus, la proportion du nombre total de pics sélectionnés par rapport au nombre de sinusoïdes présentes décroît légèrement dans le cas de la méthode de la variation de taille, plus significativement dans le cas des phases aléatoires et constantes. Par exemple, avec la méthode des phases aléatoires (seuil de 0.5), seulement 72.5 pics sont sélectionnés pour 20 composant réellement le signal analysé (3.6 sinusoïdes détectées pour 1 vraiment présente), alors que 5 sinusoïdes sont détectées pour 1 sinusoïde correcte dans le tableau 3.4 (même seuil de 0.5). Cette différence s'explique par la répartition régulière des fréquences sur tout le spectre étudié.

D'autres expériences ont été effectuées avec des sons synthétiques inharmoniques. Les fréquences f_k de chaque sinusoïde ne sont plus de la forme kf_0, mais $(k+\delta)f_0$ où δ est une valeur aléatoire. Pour ces sons, les résultats sont très similaires à ceux présentés dans cette section.

Modulation d'amplitude

Dans cette sous-section, nous testons les PS-algorithmes sur des signaux composés de bruit blanc gaussien et d'une sinusoïde de fréquence constante mais dont l'amplitude varie périodiquement avec le temps, conformément à l'équation (3.23). La profondeur de la modulation d'amplitude a_1 est 1.0 lorsque l'amplitude moyenne de la sinusoïde a_0 vaut 2.0 ($a_1 = \frac{a_0}{2}$). Cette amplitude a_0 est déterminée de façon à ce que le rapport signal sur bruit sans modulation d'amplitude vaille 6dB. La fréquence de modulation f_1 est fixée à 20Hz, ce qui représente la limite reconnue d'une modulation audible. Les phases ϕ_1 et ϕ sont aléatoirement choisies selon une distribution uniforme entre 0 et 2π. Comme dans la section 3.2.6, la fréquence f_0 est aléatoirement choisie dans des casiers consécutifs, de largeur $\frac{f_S}{4W_s}$, selon une distribution uniforme.

$$s(n) = (a_0 + a_1 \sin(2\pi f_1 \frac{n}{f_S} + \phi_1)) \sin(2\pi f_0 \frac{n}{f_S} + \phi) \tag{3.23}$$

Les résultats de cette sélection de pics sont montrés dans le tableau 3.7. Ils peuvent être comparés à ceux de la section 3.2.6 avec des rapports signal sur bruit de 0dB. Pour des rapports plus faibles, les erreurs deviennent assez nombreuses. Concernant les méthodes des phases aléatoires et des phases constantes, les résultats sont très similaires. Le nombre moyen de pics sélectionnés est très proche du nombre détecté dans le cas des sinusoïdes stationnaires (tableaux 3.3 et 3.4).

Pour la méthode de la variation de taille, les résultats montrent que le nombre moyen de pics non détectés est plus important que celui du tableau 3.2. Cela peut être expliqué par les fluctuations d'intensité introduites par la modulation d'amplitude, qui augmentent la variance de l'enveloppe temporelle et se répercute donc sur la mesure V_{NEP} que nous utilisons. Néanmoins, les erreurs introduites restent négligeables, et le nombre total de pics sélectionnés demeure très bas.

Modulation de fréquence

Dans cette sous-section, nous utilisons des signaux de bruit blanc gaussien et d'une sinusoïde dont l'amplitude est constante mais dont la fréquence varie périodiquement au cours du temps, conformément à l'équation (3.24). La fréquence du vibrato introduit est fixé à 20Hz, ce qui représente la limite de l'audition. Les phases ϕ_1 et ϕ sont aléatoirement choisies selon une distribution uniforme entre 0 et 2π. Comme dans les tests précédents, la fréquence f_0 est aléatoirement choisie dans des casiers consécutifs, de largeur $\frac{F_s}{4W_s}$, selon une distribution uniforme. L'amplitude a de la sinusoïde est déterminée de façon à ce que le rapport signal sur bruit vaille 0dB. Pour des rapports plus faibles, les erreurs deviennent assez nombreuses.

$$s(n) = a \sin(2\pi(f_0 + f_2 \sin(2\pi f_1 \frac{n}{F_s} + \phi_1)) \frac{n}{F_s} + \phi) \tag{3.24}$$

Les résultats de la sélection de pics sont montrés dans le tableau 3.8. Ils peuvent être comparés à ceux obtenus dans la section 3.2.6, dans le cas d'un rapport signal sur bruit nul

Méthode des phases aléatoires				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.525	8.350	1.000	7.350	0.000
0.5	4.783	0.999	3.783	0.001
0.45	3.119	0.995	2.124	0.005
Méthode de la variation de taille				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.25	2.277	0.999	1.278	0.001
0.2	1.608	0.992	0.616	0.008
0.15	1.335	0.917	0.418	0.083
Méthode des phases constantes				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
1.8	5.419	0.996	4.423	0.004
1.7	4.546	0.992	3.554	0.008
1.6	3.953	0.983	2.970	0.017
1.5	3.597	0.975	2.621	0.025

TAB. 3.7 – Résultats de détection de pics pour une sinusoïde dont l'amplitude est modulée, mélangé à du bruit blanc gaussien (SNR = 6dB).

(SNR= 0dB).

Les méthodes que nous utilisons ici diffèrent des précédentes. En effet, dans le cas d'une sinusoïde dont la fréquence est modulée, le maximum local du spectre discret obtenu par transformée discrète de Fourier ne correspond pas toujours au casier associé à la fréquence de la sinusoïde analysée. C'est pour cette raison que nous choisissons d'appliquer les PS-algorithmes à toutes les valeurs du spectre discret. La principale conséquence de ce choix est l'augmentation du nombre moyen de pics sélectionnés, puisque, par exemple, deux casiers consécutifs peuvent être choisis. Ainsi, pour la méthode de la variation de taille, le nombre moyen de casiers sélectionnés est très important ($\approx 4,5$) en comparaison avec celui obtenu dans le tableau 3.2 (≈ 2).

A l'exception du nombre total de sinusoïdes détectées qui accroît, le nombre moyen d'erreurs reste inchangé lorsque la fréquence des sinusoïdes est modulée. Toutefois, l'analyse de sinusoïdes dont les fréquences sont modulées requiert l'utilisation de bandes de fréquences plus larges pour pouvoir prendre en compte les variations de fréquence des sinusoïdes induites par la modulation, pendant le temps correspondant à la taille de la fenêtre d'analyse.

L'utilisation d'une valeur plus faible pour la profondeur du vibrato f_2 peut améliorer la qualité des résultats et peut nous permettre d'appliquer les PS-algorithmes aux maxima locaux du spectre discret uniquement : cela peut donc réduire le nombre total de casiers sélectionnés. Néanmoins, même dans des conditions critiques, les méthodes proposées arrivent à choisir les casiers corrects du spectre, qui correspondent aux sinusoïdes présentes dans le signal analysé.

Méthode de la variation de taille				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.2	5.792	1.000	4.792	0.000
0.15	4.753	0.994	3.759	0.006
0.1	3.607	0.909	2.698	0.091
Méthode des phases aléatoires				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.425	6.737	0.991	5.746	0.009
0.4	6.593	0.979	5.614	0.021
Méthode des phases constantes				
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
1.5	8.572	1.000	7.572	0.000
1.45	8.065	1.000	7.065	0.000
1.4	7.691	1.000	6.691	0.000

TAB. 3.8 – Résultats de détection de pics pour une sinusoïde dont la fréquence est modulée, mélangé à du bruit blanc gaussien (SNR= 0dB).

Sons naturels

Nous avons effectué de nombreuses expériences avec des sons naturels. Toutefois, il est difficile de connaître les valeurs exactes des fréquences et des amplitudes des sinusoïdes contenues dans le signal analysé, et donc de vérifier la justesse des résultats obtenus. L'expérience présentée ici concerne un son issu d'un hautbois de fréquence fondamentale $\approx$ 522Hz (C5). Nous montrons dans la figure 3.6 les maxima locaux du spectre à court terme de ce signal. Nous avons généré la mesure V_{NEP} présentée, sur chacun de ces maxima, en utilisant la méthode de la variation de taille ($w_3 = 1024$, $w_1 = 4096$, $\Delta f = 43Hz$, seuil $= 0.5$) et nous proposons sur la figure 3.6 correspondante pour illustrer l'utilité de cette méthode de sélection. Le même son auquel a été ajouté du bruit blanc gaussien (SNR= 12dB) a ensuite été analysé en appliquant les deux mêmes processus pour illustrer les possibilités des PS-algorithmes dans des conditions bruitées.

La figure 3.7 montre les pics sélectionnés en appliquant les PS-algorithmes à chaque maximum local du spectre à court terme. Nous pouvons observer que le nombre total de sinusoïdes détectées décroît considérablement. Toutefois, la majorité des pics corrects est préservée. Même les harmoniques avec une faible amplitude (dans les hautes fréquences) peuvent être extraites. Il est important de noter que chaque fenêtre temporelle successive est traitée sans aucune information sur le futur, contrairement aux méthodes du type du vocodeur de phase qui comparent les informations de phase pour chaque pic sur deux fenêtres consécutives. De plus, bien que le volume général du son analysé diminue avec le temps et que les amplitudes des harmoniques

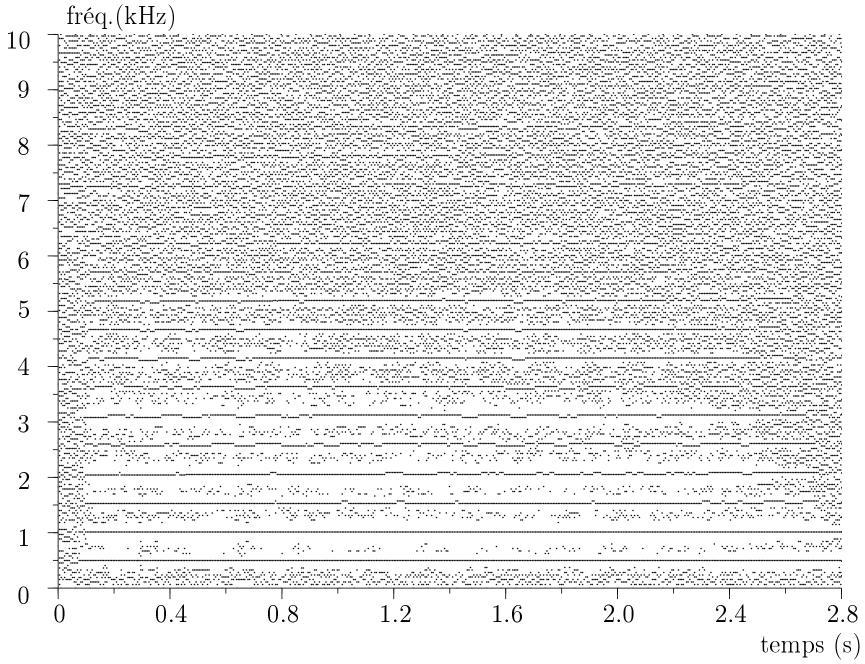

FIG. 3.6 – Localisation des maxima locaux des spectres à court terme d'un son de hautbois.

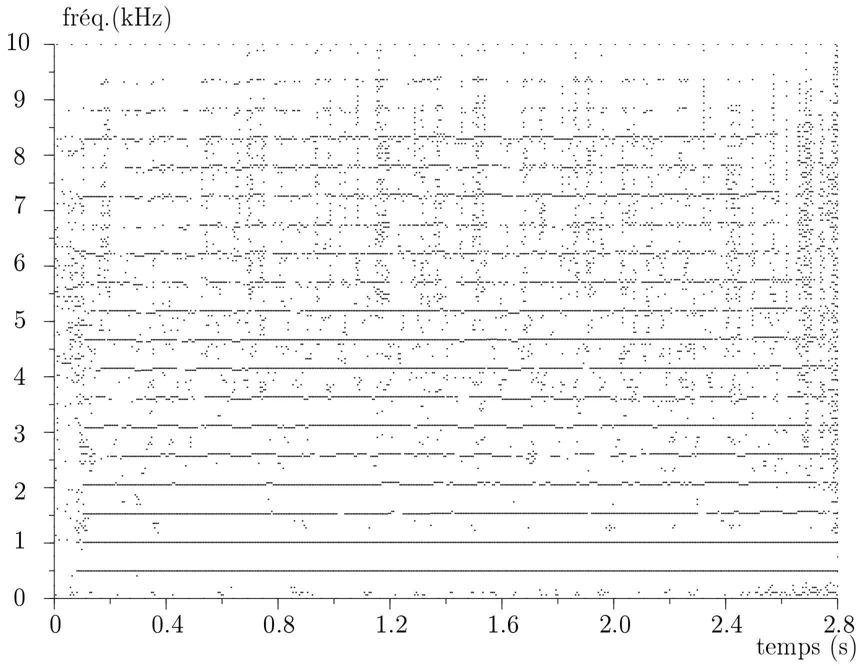

FIG. 3.7 – Localisation des pics sélectionnés en appliquant la méthode de la variation de taille à un son de hautbois.

de hautes fréquences soient plus faibles que les premières, aucun ajustement du seuil n'est nécessaire. Néanmoins, les limites des PS-algorithmes apparaissent dans ce cas (9900Hz).

Un des principaux intérêts de ces algorithmes est leur capacité à détecter des sinusoïdes dont l'amplitude est assez faible. Bien qu'elles ne soient pas toujours audibles (masquage), il est important de les prendre en compte dans le cas de certaines transformations musicales. L'utilisation des PS-algorithmes n'empêche pas l'application des méthodes de suivi de partiels, mais améliorent juste leur qualité en évitant des mauvais liens entre des pics de fenêtres consécutives. De plus, des algorithmes de suivi de partiels ont été développés pour reconstruire les trajectoires de partiels (voir la section 2.3.3), même si des pics corrects ont été manqués (présence de trous dans les trajectoires).

Un second exemple de détection de pics concerne le même son original, mais dans un environnement bruité. Nous avons choisi de ne montrer que la partie du spectre correspondante aux basses fréquences, car elle est composée d'harmoniques dont les amplitudes restent importantes comparées à l'amplitude du bruit. La figure 3.8 représente les maxima locaux du spectre à court terme alors que la figure 3.9 montre les pics sélectionnés par les PS-algorithmes (méthode de la variation de taille avec $w_3 = 1024$, $w_1 = 4096$, $\Delta f = 172Hz$ et un seuil de $= 0.15$). Comme précédemment, l'utilisation d'un PS-algorithme diminue considérablement le nombre total de pics sélectionnés.

3.2.7 Discussion des paramètres de l'analyse

Nous étudions et discutons dans cette section l'influence des paramètres de cette méthode d'analyse. Les expériences sont récapitulées dans des tableaux proposés en annexe A.

Taille de la fenêtre pour le calcul de la variance

Les résultats montrent tout d'abord que la modification de cette taille de fenêtre w_3 entraîne une variation de la mesure V_{NEP}. Lorsque cette taille augmente, les seuils doivent aussi être augmentés pour pouvoir trouver le même nombre moyen de sinusoïdes. Ces expériences confirment que ce paramètre n'a pas une influence primordiale sur la précision des méthodes des phases constantes et aléatoires. Par exemple, dans le tableau A.1, le même nombre moyen de sinusoïdes exactes $\overline{N}_{\text{exact}} = 0.931$ peut être obtenu avec un seuil 0.375 (taille $w_3 = 256$) ou avec un seuil 0.55 (taille $w_3 = 512$). En comparaison, le nombre moyen de sinusoïdes sélectionnées est légèrement plus grand pour $w_3 = 512$ que pour $w_3 = 256$. Mais cette différence n'est pas réellement significative.

Pour la méthode de la variation de taille, l'influence de la taille w_3 est confirmée être beaucoup plus importante pour la précision (voir les explications dans la section 3.2.4). Par exemple, dans le tableau A.1, un seuil de 0.15 implique approximativement la sélection du même nombre moyen de pics (1.943 et 1.829). Mais le nombre moyen de sinusoïdes qui sont dans le signal mais qui ne sont pas détectées est différent : aucune erreur n'est introduite avec la taille $w_3 = 1024$ alors que 4% des pics ne sont pas détectés. Ce paramètre détermine la précision de la méthode de la variation de taille.

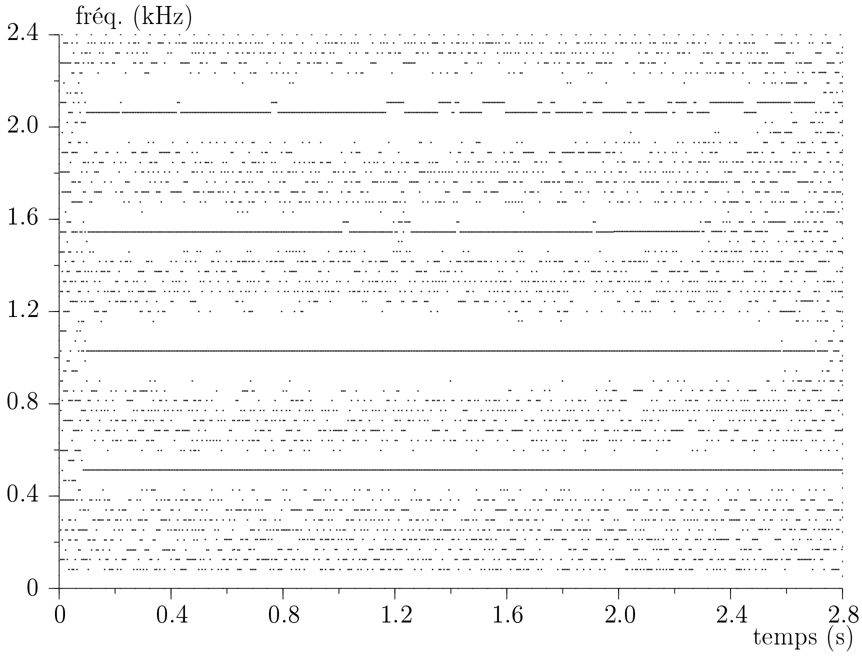

Fig. 3.8 – Localisation des maxima locaux des spectres à court terme d'un son de hautbois auquel a été rajouté du bruit blanc gaussien (SNR= 12dB).

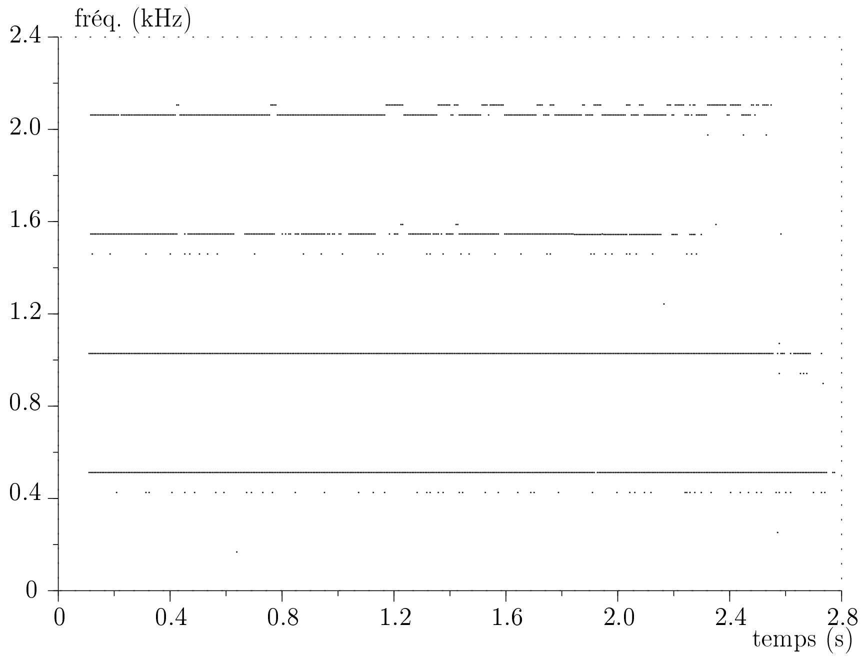

Fig. 3.9 – Localisation des pics sélectionnés en appliquant la méthode de la variation de taille à un son de hautbois auquel a été rajouté du bruit blanc gaussien (SNR= 12dB).

Taille de la fenêtre pour les modifications spectrales

A l'opposé de la taille de la fenêtre w_3, la taille de la fenêtre w_2 concerne principalement la méthode des phases constantes et aléatoires, comme cela a été expliqué dans la section 3.2.4. Les expériences avec la méthode de la variation de taille montrent clairement que l'influence de cette taille est mineure (voir le tableau A.4) : Les différences pour le même seuil sont négligeables.

Pour les deux autres méthodes, l'augmentation de cette taille implique une meilleure précision. Par exemple, dans le tableau A.3, pour le même nombre moyen d'erreurs ($\overline{N}_{\text{exact}} = 0.921$ et $\overline{N}_{\text{exact}} = 0.931$), le nombre moyen de pics sélectionnés est bien plus important avec la taille $w_3 = 512$ ($\overline{N}_{\text{sinus}} = 5.857$) qu'avec une taille $w_3 = 1024$ ($\overline{N}_{\text{sinus}} = 3.292$). De la même manière, pour la méthode des phases constantes (voir la tableau A.3), le même nombre moyen de sinusoïdes détectées ($\overline{N}_{\text{sinus}} = 4.327$ et $\overline{N}_{\text{sinus}} = 4.245$) est perturbé par plus d'erreurs avec la taille $w_3 = 512$ ($\overline{N}_{\text{exact}} = 0.686$) qu'avec la taille $w_3 = 1024$ ($\overline{N}_{\text{exact}} = 0.998$).

Nombre de réalisations

Ces deux séries de résultats montrent clairement que leur qualité est améliorée quand le nombre d'itérations augmente. Par exemple, le nombre moyen de sinusoïdes détectées qui sont présentes dans le signal ($\overline{N}_{\text{exact}}$) est approximativement le même que dans le tableau A.6, quel que soit le nombre d'itérations. Il est légèrement amélioré dans le tableau A.5. Mais le point important ici est que le nombre total moyen de sinusoïdes diminue avec le nombre d'itérations, même si le nombre moyen d'erreurs reste constant (ou diminue).

Le problème est que le temps de calcul augmente proportionnellement avec ce nombre d'itérations. On peut observer que la multiplication de ce nombre par deux ne multiplie pas la qualité de l'analyse par deux. L'utilisation de cette méthode requiert le choix d'un nombre de réalisations le plus important possible, la principale limite étant liée au temps de calcul nécessaire.

Largeur de la bande de fréquence analysée

La modification de la bande de fréquence requiert un réajustement du seuil de détection. En effet, l'augmentation de la largeur permet d'augmenter le nombre de pics donnés par la mesure V_{NEP}. C'est pourquoi les seuils pourraient être légèrement augmentés dans le but de trouver la même nombre moyen de pics détectés.

Pour la méthode des phases constantes, ce paramètre a une influence mineure, à part pour le niveau général de la mesure V_{NEP}. Par exemple, le même nombre moyen de pics sélectionnés ($\overline{N}_{\text{sinus}} = 4.985$ et $\overline{N}_{\text{sinus}} = 4.962$) implique approximativement le même nombre moyen d'erreurs ($\overline{N}_{\text{manque}} = 0.999$ and $\overline{N}_{\text{manque}} = 0.989$).

Concernant la méthode des phases aléatoires, l'augmentation de la largeur de la bande de fréquence provoque la sélection d'un plus grand nombre de pics. Cela implique évidemment que le nombre de sinusoïdes manquées diminue. Par exemple, le seuil minimum pour lequel toutes les sinusoïdes présentes sont détectées ($\overline{N}_{\text{exact}} = 1.0$), induit la sélection d'un plus grand nombre de pics non corrects avec une largeur de 16 casiers ($\overline{N}_{\text{faux}} = 3.936$), qu'avec une largeur de 12 casiers ($\overline{N}_{\text{faux}} = 2.601$). Les résultats opposés sont observés avec la méthode de la variation de taille. L'élargissement de la bande de fréquence étudiée permet dans une certaine mesure de sélectionner moins de pics sans pour autant augmenter le taux d'erreurs.

Ces remarques sont valables seulement lorsque les PS-algorithmes sont appliqués aux maxima locaux des spectres à court terme. En effet, plus la bande de fréquences est large, plus le nombre de fenêtres fréquentielles contenant une éventuelle sinusoïde est important. L'application des PS-algorithmes aux seuls maxima locaux des spectres à court terme empêche la sélection de deux casiers consécutifs, puisque deux sinusoïdes peuvent dans ce cas être détectées dans deux fenêtres fréquentielles voisines. Si les algorithmes sont directement appliqués à tous les casiers des spectres, le nombre moyen de pics sélectionnés augmente considérablement. Dans ce cas, l'utilisation d'une fenêtre fréquentielle trop large aura pour conséquence une diminution de la précision dans le processus de sélection de pics. Il est alors conseillé de calculer les mesures V_{NEP} à partir de bandes de fréquences étroites.

De plus, cette largeur de bande de fréquence impose une restriction sur la différence entre deux sinusoïdes consécutives. En effet, si une bande étudiée contient deux (ou plus) sinusoïdes, la mesure V_{NEP} associée donne des valeurs plus larges qu'avec une seule sinusoïde. Dans ce cas, la probabilité de choisir le casier du centre de cette fenêtre est plus faible, puisque la mesure a plus de chances d'être supérieure au seuil choisi. C'est pourquoi il est aussi recommandé de choisir une largeur de bande de fréquence étroite pour analyser des sons harmoniques dont la fréquence fondamentale est faible, ou des sons composés de sinusoïdes proches. La limite de cette différence est donnée trivialement par la moitié de la largeur Δf de la bande de fréquences choisie.

Le choix de ce paramètre dépend aussi de la possibilité pour les sinusoïdes recherchées composant le signal analysé, d'être modulée en fréquence. Dans ce cas, une bande plus large améliore assez sensiblement la qualité des résultats.

3.2.8 Application

Nous avons donc proposé une méthode originale pour améliorer l'analyse spectrale des sons en sélectionnant les pics significatifs de spectres à court terme. Cette méthode est basée sur l'étude statistique des fluctuations d'intensité, alors que les méthodes existantes reposent essentiellement sur la transformée discrète de Fourier et donc dépendent des hypothèses de stationnarité.

Nous avons proposé trois algorithmes pour appliquer cette méthode. Ces algorithmes ont été testés avec un certain succès sur des sons synthétiques bruités et sur des sons naturels. Du point de vue de la complexité, deux algorithmes sont particulièrement intéressants car ils nécessitent seulement le calcul de la puissance de l'enveloppe filtrée du son analysé. A l'opposé, la méthode des phases aléatoires qui nécessite la synthèse de nombreuses réalisations du signal original, consomme beaucoup de ressources.

Ces méthodes ont quelques avantages notables par rapport aux techniques existantes :

Haute précision dans des conditions bruitées : Les comparaisons avec les méthodes existantes montrent que les algorithmes que nous proposons sont particulièrement efficaces lorsque le son analysé comporte beaucoup de bruit. Des sinusoïdes peuvent ainsi être parfaitement détectées dans le cas de rapports signal sur bruit négatifs. Ces méthodes peuvent donc s'avérer particulièrement utiles pour l'analyse de sons naturels bruités et/ou de bruits naturels.

méthode	seuil très bruité	seuil peu bruité	largeur Δf de la bande de fréquence (Hz)	nombre de réalisations
variation de taille	0.3	0.1	517	1
phases aléatoires	0.5	0.45	517	25
phases constantes	1.8	1.2	517	1

TAB. 3.9 – Exemples de valeurs de paramètres envisageables pour l'application des PS-algorithmes.

Indépendance du seuil de détection par rapport à l'amplitude : Nous pouvons montrer que la mesure V_{NEP} est indépendante de l'amplitude du signal analysé. Quel que soit le volume général du son étudié, les mêmes valeurs de seuil conduisent au même résultat. Cette propriété est particulièrement intéressante lorsque la dynamique des sons varient. Ces méthodes ne nécessitent pas de changement de seuil, et peuvent donc être complètement automatisées. Néanmoins, les paramètres utilisés par les PS-algorithmes peuvent être précisément définis pour obtenir une meilleure précision. Des exemples de paramètres sont proposés dans le tableau 3.9. Ces paramètres dépendent évidemment de la nature du son analysé (bruité, harmonique, ...), et éventuellement de sa hauteur.

Indépendance par rapport à la transformée de Fourier : Les PS-algorithmes présentés sont très différents des algorithmes classiques employés pour l'analyse de sons, puisque les informations prises en compte sont essentiellement extraites de la variance de l'enveloppe temporelle, et non de la transformée de Fourier. Les résultats donnés par les méthodes que nous avons développées peuvent parfaitement être combinés avec des résultats provenant d'autres approches, dans le but d'améliorer l'analyse.

Après avoir appliqué une des techniques que nous proposons au son analysé, il est possible d'effectuer les liens entre les pics considérés sur des fenêtres temporelles successives, pour reconstruire les trajectoires des partiels. Les méthodes que nous proposons permettent d'éviter d'extraire des pics erronés et donc de considérer des partiels faux dus à l'énergie du bruit présent.

Nous envisageons surtout leur utilisation dans le cadre d'un schéma général d'analyse de sons bruités, destiné au modèle CNSS que nous avons développé pour la représentation de sons bruités.

3.3 Détection de transitoires

De nombreuses méthodes de détection de transitoires ont été développées, que ce soit au niveau microscopique (à l'échelle du son) pour relever avec précision l'instant d'une transitoire, ou au niveau macroscopique (à l'échelle musicale) pour analyser le début d'une note. Nous nous intéressons ici aux techniques liées à l'analyse de son. Nous ne présentons que les plus connues. Bien sûr, la méthode de détection idéale n'existe pas, et il est envisageable de combiner simultanément plusieurs techniques pour essayer d'améliorer la précision. En général, ces méthodes de détection reposent sur l'étude de l'énergie. L'énergie, notée E, peut être obtenue à partir de la représentation temporelle ou spectrale :

$$\mathsf{E} = \sum_k |X_f(k)|^2 = \sum_k x[k]^2 \tag{3.25}$$

où X_f représente le spectre discret d'amplitude du signal x.

3.3.1 Méthodes existantes

Dans toutes les méthodes, le signal est analysé par des fenêtres successives, indexées par leur date de début. Par exemple, nous notons E_r l'énergie dans une fenêtre indexée par r.

Variations d'énergie temporelle

La méthode la plus classique consiste à étudier les variations d'énergie locale, en comparant l'énergie de la fenêtre courante et l'énergie moyenne du signal. Pour cela, l'énergie relative à la fenêtre courante est calculée. La détection des transitoires est effectuée à partir du rapport entre cette énergie et la moyenne des énergies précédentes, établies en sommant les énergies pondérées relatives aux T fenêtres précédentes [Lev98]. Si ce rapport devient bien supérieur à 1, cela signifie qu'une forte variation d'énergie est localisée sur la fenêtre courante, et qu'une transitoire peut être supposée présente :

$$\frac{\mathsf{E}_r}{\frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{r-t}} >> 1 \tag{3.26}$$

où r représente l'indice de la fenêtre courante de signal.

D'un point de vue pratique, cette méthode peut être appliquée en utilisant des fenêtres de taille 512 échantillons, multipliées par la fenêtre de Hamming, et se chevauchant avec un temps d'avancement correspondant à la moitié de cette taille.

Variations d'énergie des hautes fréquences

Des améliorations de la méthode précédente existent, pour prendre en compte les propriétés des transitoires dans les signaux musicaux. En effet, les expériences montrent que les variations d'énergie dans les basses fréquences sont parfois dues aux variations d'amplitude des harmoniques du signal. Par contre, les variations dans les hautes fréquences ne peuvent être la conséquence que de la présence d'une transitoire. C'est à cause de cette observation qu'une première amélioration possible consiste à ne considérer que les parties du spectre supérieures à 1000Hz pour calculer l'énergie [Pee01].

Toujours en s'appuyant sur cette propriété, Masri décrit une méthode de détection basée sur l'énergie et sur le contenu des hautes fréquences (*high frequency content*), noté HFC [MB96]. Cette dernière quantité est définie par la somme des amplitudes du spectre, élevées au carré, chacune étant pondérées par les fréquences correspondantes :

$$\mathrm{HFC}_r = \sum_k [|X_f(k)|^2.k] \tag{3.27}$$

où r représente l'indice de la fenêtre courante de signal.

La mesure introduite, notée $\mathcal{F}_i$, est définie par la multiplication du rapport de variation du contenu des hautes fréquences par le rapport entre le contenu des hautes fréquences et l'énergie :

$$\mathcal{F}_r = \frac{\mathrm{HFC}_r^2}{\mathrm{HFC}_{r-1}\mathsf{E}_r} \tag{3.28}$$

Un seuil est défini et toutes les valeurs de $\mathcal{F}_r$ supérieures à ce seuil indiquent la présence d'une transitoire dans la fenêtre r correspondante. La figure 3.10 montre un exemple de l'évolution de cette mesure au cours du temps, pour un son produit par plusieurs notes successives de guitare acoustique. L'analyse est effectuée sur 512 échantillons et un taux de chevauchement des fenêtres de 50%.

Les limites de cette méthode apparaissent rapidement lorsque le son analysé comporte un fond important, que ce soit du bruit ou des sons harmoniques. En effet, dans ce cas, les variations d'énergie deviennent insuffisantes pour pouvoir impliquer une hausse de valeur pour la mesure considérée $\mathcal{F}\cdot$ Les figures 3.15 et 3.16 illustrent ces limites pour des sons percussifs mélangés à du bruit.

Similarités spectrales

Une autre méthode portant sur l'étude des variations du spectre a été proposée par Masri [Mas96]. Celle-ci définit une mesure de dissimilitude, notée $\mathcal{D}$, qui quantifie les différences entre deux spectres consécutifs, sans s'attacher à la localisation fréquentielle de ces différences :

$$\mathcal{D}_r = \sum_k \frac{||X_f^r(k)|^2 - |X_f^{r-2}(k)|^2|}{|X_f^{r-2}(k)|^2} \tag{3.29}$$

où X_f^r représente le spectre à court terme associée à la fenêtre r. Elle se rapproche en cela de la première méthode de variation d'énergie que nous avons montrée. Un seuil doit être choisi pour pouvoir automatiquement définir comme fenêtre comportant une transitoire, les fenêtres pour lesquelles la dissimilitude $\mathcal{D}$ est supérieure à ce seuil.

Distance de spectre

Une autre façon de mesurer la différence entre deux spectres consécutifs est proposée par Peeters [PR99]. Le spectre de la fenêtre étudiée est considéré comme la densité de probabilité du spectre. Les recherches de théorie de l'information utilisent dans ce cas la distance dite de Kullback-Leibler [Bas89]. Ainsi, une mesure dérivée de cette approche statistique, notée $\mathcal{D}_{\mathrm{KL}}$, se calcule à partir du spectre X_f^r de la fenêtre courante r et celui X_f^{r-1} de la fenêtre précédente $r-1$:

$$\mathcal{D}_{\mathrm{KL}} = \sum_k \mathcal{F}\left(\frac{X_f^{r-1}(k)}{X_f^r(k)}\right) \tag{3.30}$$

où la fonction $\mathcal{F}$ est définie par :

$$\forall x, \quad \mathcal{F}(x) = x - \log(x) - 1 \tag{3.31}$$

Un seuil est également défini pour cette méthode et toutes les valeurs de distances $\mathcal{D}_{\mathrm{KL}}$ supérieures à ce seuil indiquent la présence d'une transitoire dans la fenêtre correspondante.

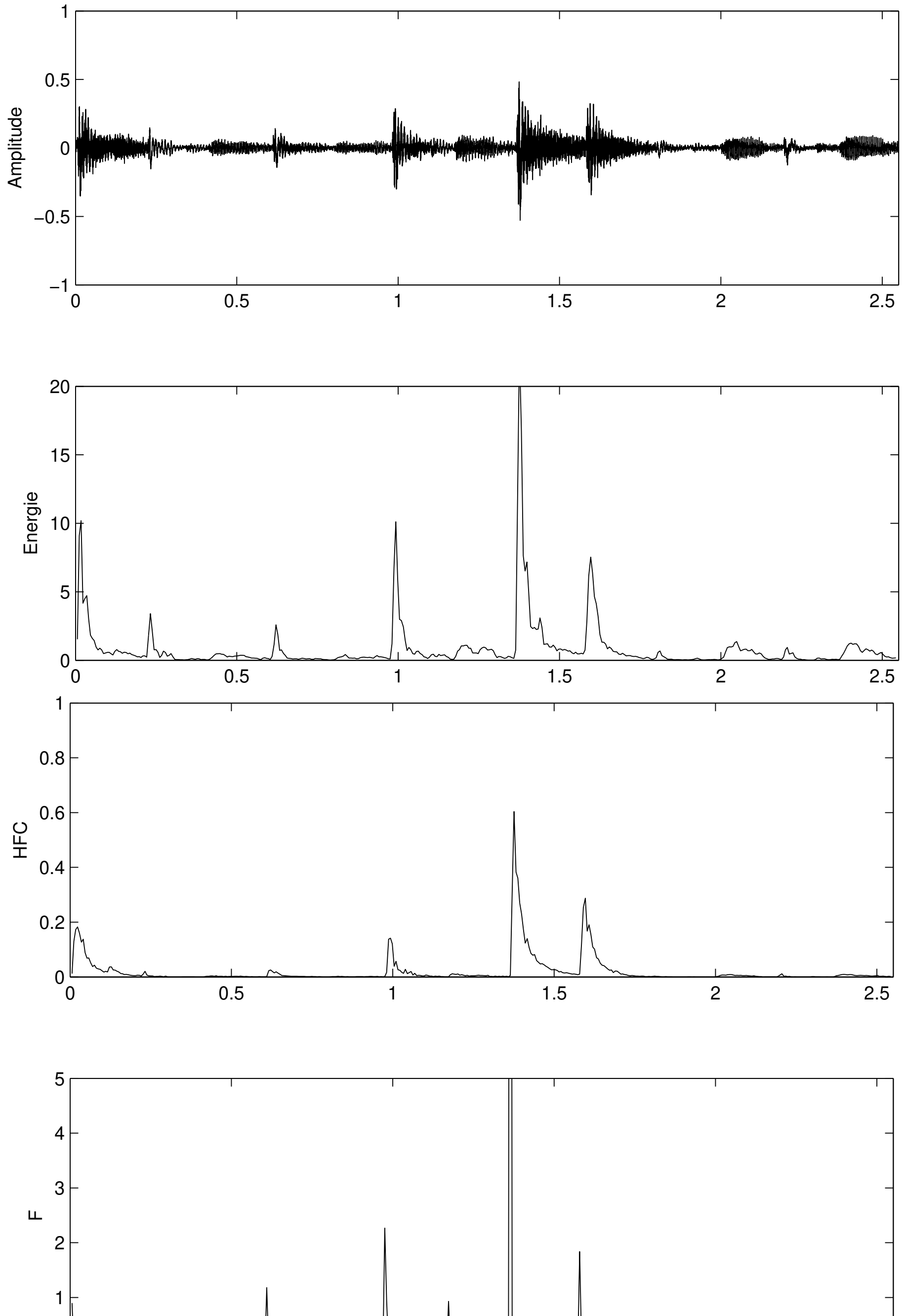

Fig. 3.10 – Exemple de détection de transitoires dans un son par la méthode des variations d'énergie dans les hautes fréquences (512 échantillons, 50% de recouvrement) : la figure du haut représente le signal issu d'une mélodie de guitare acoustique, la seconde figure représente l'énergie, la troisième le contenu hautes fréquences (HFC) et enfin la dernière figure montre la mesure des transitoires $\mathcal{F}$ résultante.

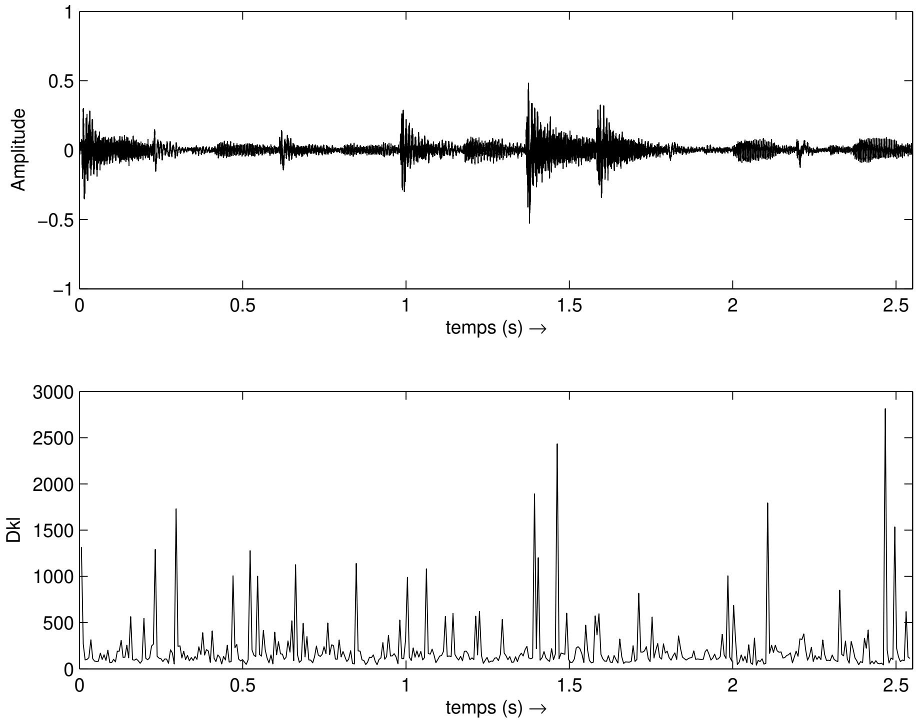

FIG. 3.11 – Exemple de détection de transitoires dans un son par la méthode de la distance de Kullback-Leibler (512 échantillons, 50% de recouvrement) : la figure du haut représente le signal issu d'une voix alors que celle du bas montre la mesure $\mathcal{D}_{\mathrm{KL}}$ associée.

La figure 3.10 montre un exemple de l'évolution de cette mesure au cours du temps, pour un son produit par plusieurs notes successives de guitare acoustique. L'analyse est effectuée sur 512 échantillons et un taux de chevauchement des fenêtres de 50%.

Cette méthode est finalement très similaire aux méthodes précédentes d'étude des variations d'énergie et présente donc les mêmes limites. La détection peut être faussée si le son étudié comporte du bruit ou un signal harmonique important. Ainsi une amélioration est proposée dans [Pee01] qui préconise l'utilisation d'une version approximée et lissée du spectre (plus précisément de son enveloppe). La version que nous avons testée ne prend pas en compte cette amélioration.

Comparaisons d'énergie signal/résidu

Les méthodes de détection de transitoires sont souvent développées dans le but d'être intégrées à des méthodes d'analyse et de synthèse de sons reposant sur le principe des modèles hybrides sinusoïdes/bruit. Le principe de l'analyse associée à ces méthodes repose d'abord sur la détection et le suivi de partiels, qui sont ensuite soustraits du signal. Il est évident que si le signal original comporte des transitoires, celles-ci devraient se trouver dans la partie résiduelle, résultant de cette soustraction. Ainsi, Levine propose dans [Lev98] de comparer l'énergie de cette partie résiduelle et de la partie sinusoïdale pour estimer la présence de transitoires. La mesure notée $\mathcal{F}$ est définie par le rapport de ces deux parties du signal à analyser :

$$\mathcal{F} = \frac{\mathsf{E}_{\text{résidu}}}{\mathsf{E}_{\text{sinus}}} \tag{3.32}$$

où $\mathsf{E}_{\text{résidu}}$ représente l'énergie de la partie résiduelle, et $\mathsf{E}_{\text{sinus}}$ celle de la partie sinusoïdale. Un rapport proche de 0 indique l'absence présumée de transitoires. Un rapport important, supérieur à un seuil préalablement défini, indique la présence de transitoires.

Les limites de cette méthode sont bien sûr liées aux limites déjà présentées de l'analyse des partiels et de la définition du résidu par soustraction spectrale. De plus, un son bruité peut présenter une énergie relative au résidu très importante. La grande valeur du rapport $\mathcal{F}$ n'est alors pas due à la présence de transitoire, mais à la forte présence de bruit.

3.3.2 Analyse de la distribution des échantillons

La méthode de détection des transitoires, destinée à être utilisée lors de la première étape de l'analyse pour le modèle CNSS, doit à priori respecter quelques contraintes, dont certaines ont été proposées dans [RJ01] :

a. elle ne doit reposer sur aucune hypothèse à propos du contenu spectral des transitoires détectées.
b. elle ne doit pas utiliser des résultats provenant d'une méthode d'analyse sinusoïdale.
c. elle doit induire des résultats très précis du point de vue temporel.
d. elle doit pouvoir être utilisable en milieu fortement bruité.
e. elle doit être simple et intuitive d'utilisation, proposant un nombre réduit de paramètre (voire aucun), et pouvant être entièrement automatisable.

Les deux premières conditions sont contraires à l'application de la méthode du rapport résidu-signal, nécessitant une analyse sinusoïdale, et de la méthode du contenu hautes fréquences, reposant sur l'hypothèse selon laquelle les transitoires sont essentiellement constituées de hautes fréquences, contrairement au reste du signal. Plus généralement, les méthodes existantes, essentiellement basées sur l'étude des variations d'énergie et/ou de l'enveloppe spectrale, présentent des limites importantes lorsque le son analysé est fortement bruité.

Par ailleurs, la dernière condition, même si elle n'est pas la plus importante, pose également un problème par rapport aux méthodes présentées. En effet, chacune nécessite le choix d'un seuil qui dépend beaucoup de la nature du son étudié. Une valeur de seuil commune à tous les sons ne peut être raisonnablement envisagée.

Nous proposons ici une nouvelle méthode destinée à la détection de transitoires dans les signaux bruités, basée sur la distribution des échantillons du signal sonore, et qui, à notre connaissance, n'a pas encore été développée.

Principe

Ici, nous nous intéressons à la répartition des échantillons. Comme nous l'avons vu dans la section 2.2, celle-ci, même si elle ne donne pas d'information très précise sur les propriétés perceptives du son, peut s'avérer être un bon indicateur de la nature général d'un son (voir par exemple la figure 2.4). Des propriétés de la distribution des échantillons peuvent être quantifiées par les premiers moments. Parmi ceux-ci, le quatrième, aussi appelé kurtosis, est significatif de l'aplatissement de la fonction de densité de probabilité. Par exemple, une loi gaussienne peut être caractérisée par une valeur de kurtosis de 3 (c'est la valeur de référence), alors que la valeur associée à une loi uniforme vaut 1.8. Ainsi, une impulsion temporelle sera caractérisée par une fonction de densité de probabilité très aiguë, ce qui induira une hausse de kurtosis. Cette observation est illustrée par la figure 3.12. Cette approche a d'ailleurs été mise en application pour définir et caractériser les bruits impulsifs se produisant sur les lieux de travail à des fins médicales [Erd86].

Par ailleurs, il est possible d'établir un lien théorique entre le kurtosis et les fluctuations d'intensité (voir la section 2.4.5). Plus les fluctuations d'intensité sont importantes, plus le kurtosis prend une grande valeur. Or il est évident qu'une transitoire va augmenter fortement ces fluctuations. Ceci est montré par l'étude des bruits générés par somme de N sinusoïdes, de fréquences régulièrement réparties, d'amplitudes et de phases ϕ constantes. L'amplitude de la transitoire (impulsion) générée par ces sons dépend du nombre de sinusoïdes N. La relation entre le kurtosis et ce nombre N est donnée dans [HP91] :

$$K = \frac{1}{N}(N^2 + \frac{1}{2}) + \frac{(N-1)(N-2)}{3N}\cos(2\phi) \tag{3.33}$$

Quelle que soit la valeur des phases ϕ, le kurtosis augmente lorsque N augmente, et donc lorsque l'amplitude relative à la transitoire croît.

Algorithme

La méthode que nous proposons est donc basée sur le quatrième moment statistique du signal, celui-ci étant considéré comme aléatoire. En fait une transitoire sera localisée par une forte variation de kurtosis. Pour définir l'évolution du kurtosis au cours du temps, le signal est découpé en fenêtres temporelles de taille fixée W_s, pouvant éventuellement se chevaucher (taux de recouvrement H). Chaque fenêtre définit un signal dont le kurtosis peut être exprimé, à partir de la définition donnée par l'équation (1.47) :

$$m_r = \frac{1}{W_s}\sum_{n=n_r}^{n_r+W_s} x[n] \tag{3.34}$$

$$K_r = \frac{\frac{1}{W_s}\sum_{n=n_r}^{n_r+W_s}(x[n]-m_r)^4}{(\frac{1}{W_s}\sum_{n=n_r}^{n_r+W_s}(x[n]-m_r)^2)^2} \tag{3.35}$$

où r représente l'indice de la fenêtre courante. Sachant que nous pouvons considérer dans le cas de signaux sonores la moyenne des échantillons nuls, cette équation devient :

$$K_r = \frac{\sum_{n=n_r}^{n_r+W_s} x[n]^4}{\frac{1}{W_s}(\sum_{n=n_r}^{n_r+W_s} x[n]^2)^2} \tag{3.36}$$

Contrairement aux méthodes basées sur la transformée discrète de Fourier, l'ensemble d'échantillons W_s n'est pas multiplié par une fenêtre de type hamming ou hann. En effet, si c'était le cas, les transitoires seraient déformées, ce qui pourrait donc entraîner une distortion du kurtosis associé.

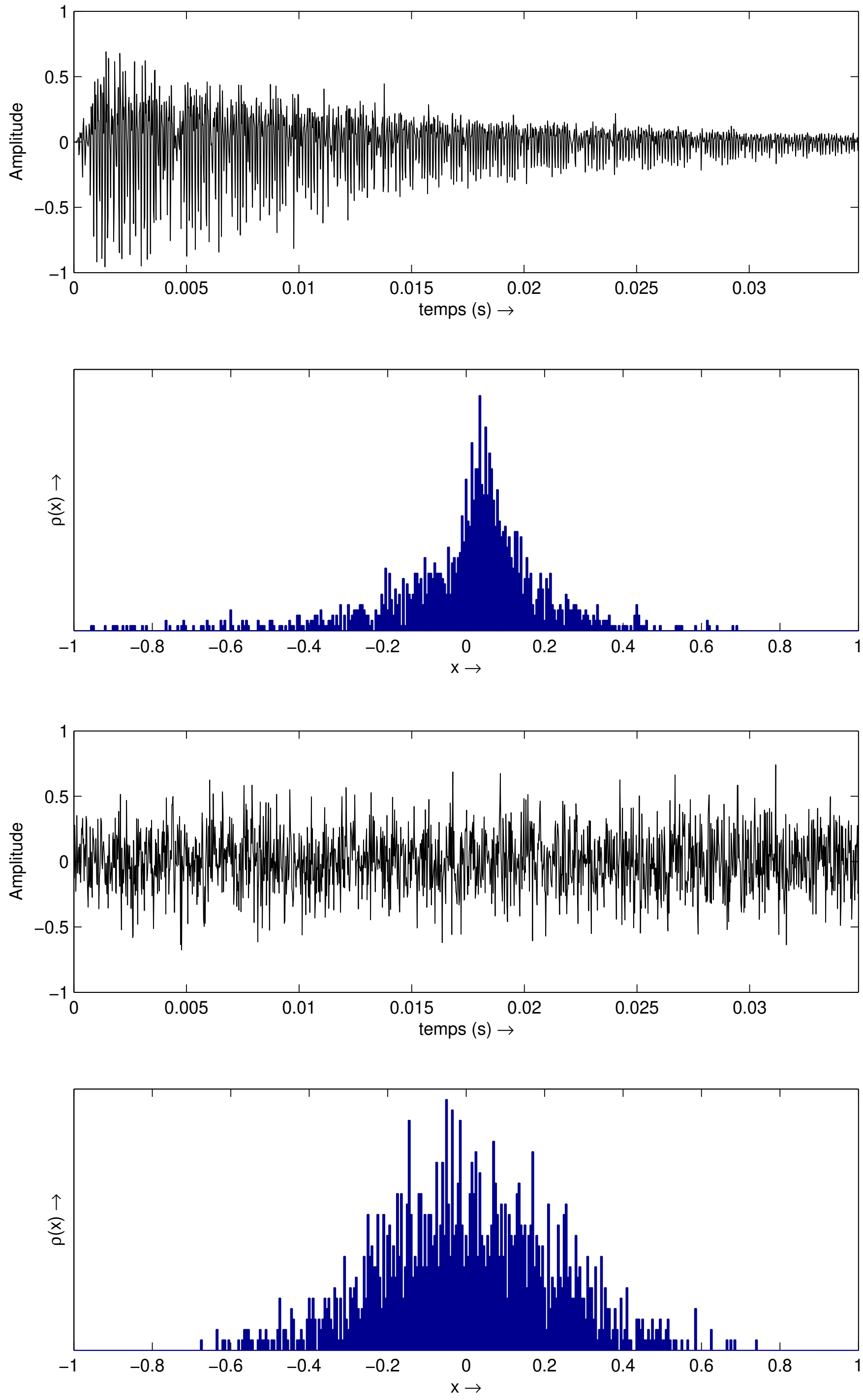

Fig. 3.12 – Illustration de la caractérisation de signaux par le kurtosis : la figure du haut montre l'évolution d'une transitoire (issue d'une percussion) dans le temps (durée de 1536 échantillons) et la deuxième figure montre la répartition des échantillons associée qui apparaît pointue (kurtosis 6.45). La troisième figure montre la représentation temporelle d'un bruit blanc gaussien (durée de 1536 échantillons) et la quatrième figure présente la distribution associée : celle-ci est plus plate (kurtosis 3).

Le fait que le kurtosis soit calculé à partir d'un nombre réduit d'échantillons induit souvent quelques erreurs de précision. Cette observation est appuyée par les expériences simples que nous avons menées, récapitulées dans le tableau 3.10. Par exemple, si le kurtosis d'un bruit blanc Gaussien (théoriquement de 3) est mesuré à partir de 512 échantillons, les valeurs sont comprises entre 2.35 et 4.12. C'est cette dispersion des valeurs qui pose les limites de cette méthode.

Une valeur de kurtosis est donc associée à chaque fenêtre, c'est-à-dire à plusieurs instants régulièrement répartis. Un seuil est défini de façon à proscrire les valeurs de kurtosis dues aux imprécisions liées aux moyennes. Ainsi, chaque valeur de kurtosis supérieure à ce seuil sera synonyme de présence de transitoire. Plus précisément, une valeur très forte indique le début de la transitoire. Nous avons aussi pu observer qu'une variation très légère apparaît parfois à la fin de la transitoire. Ces indicateurs permettent d'établir avec une bonne précision, liée à la taille de la fenêtre et au taux de recouvrement, l'instant de début et parfois celui de fin de la transitoire. Un exemple d'un signal bruité contenant trois attaques de percussions est donné par la figure 3.13 qui montre les résultats obtenus par l'application du calcul du kurtosis. Dans ce cas, ce calcul permet de définir très précisément l'instant de début de chaque transitoire, et pour deux attaques, l'instant de fin.

La présence d'une transitoire induit une variation de kurtosis. Celle-ci peut être mieux mise en valeur par le calcul de la différence $\mathcal{D}K$ avec la moyenne des kurtosis correspondant aux T fenêtres précédentes :

$$\mathcal{D}K = K_r - \frac{1}{T}\sum_{t=1}^{T} K_t \tag{3.37}$$

Ce calcul, qui revient à un filtre passe-bas, permet d'éviter les pics correspondant à des variations légères dues aux approximations, pour mieux mettre en relief les pics correspondant aux transitoires. La figure 3.13 montre l'évolution de cette différence au cours du temps.

Expériences

La méthode que nous venons de présenter a été expérimentée sur des sons naturels et synthétiques. La figure 3.14 montre l'exemple de la détection de transitoires pour un son naturel issu de castagnettes. Ce son n'est quasiment pas bruité. D'autres tests ont été effectués avec des sons de percussion (cymbales, caisses) synthétiques, provenant des sons utilisés par Masri [MB96]. Nous avons rajouté du bruit blanc gaussien à ces sons et nous leur avons appliqué deux autres méthodes (celle du contenu hautes fréquences et celle basée sur la distance de Kullback Leibler), pour essayer d'effectuer des comparaisons. Il nous semble que l'approche par la mesure du kurtosis est plus résistante au bruit. Les figures 3.15 et 3.16 montrent ces résultats, en faisant apparaître un seuil (kurtosis de 5). Un tel seuil, bien supérieur au kurtosis maximum donné par le tableau 3.10 pour une taille de fenêtre de 512 échantillons, permet de détecter quasiment toutes les transitoires du signal analysé.

Discussion

Cette méthode de détection de transitoires, basée sur le calcul du kurtosis, impose le choix de trois paramètres. La taille de la fenêtre d'analyse impose une hypothèse de stationnarité du signal, puisque le calcul suppose que les propriétés statistiques, notamment le kurtosis,

W_s	$\langle K \rangle$	$\min(K)$	$\max(K)$
∞	3.00	3.00	3.00
2048	2.9946	2.6284	3.6202
1024	2.9902	2.5119	3.7909
512	2.9837	2.3539	4.1169
256	2.9664	2.1317	5.1209
128	2.9309	1.9269	6.1102

TAB. 3.10 – Variations du calcul expérimental du kurtosis d'un bruit blanc Gaussien (fonction `randn` de `Matlab`) en fonction du nombre d'échantillons considérés.

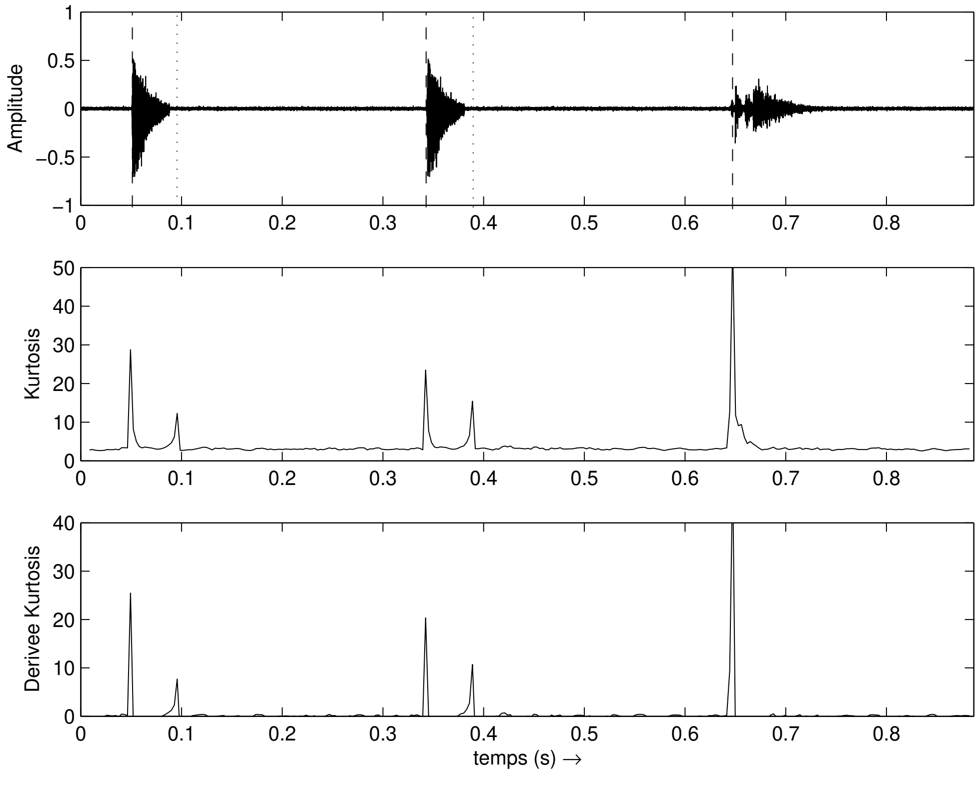

FIG. 3.13 – Illustration de la détection de transitoires par le calcul du kurtosis : la première figure montre le signal analysé composé de trois transitoires (sons de percussion) auquel un faible bruit a été ajouté, la deuxième figure présente l'évolution du kurtosis au cours du temps (fenêtre de 512 échantillons, taux de chevauchement de 25%) et la troisième figure les variations du kurtosis (différence par rapport à la moyenne des 4 valeurs précédentes). Les trois transitoires sont détectées, leur temps de départ analysé étant représenté sur la première figure en traits hachurés, et leur temps de fin en pointillés. Seul l'instant de fin de la troisième transitoire n'a pu être analysé.

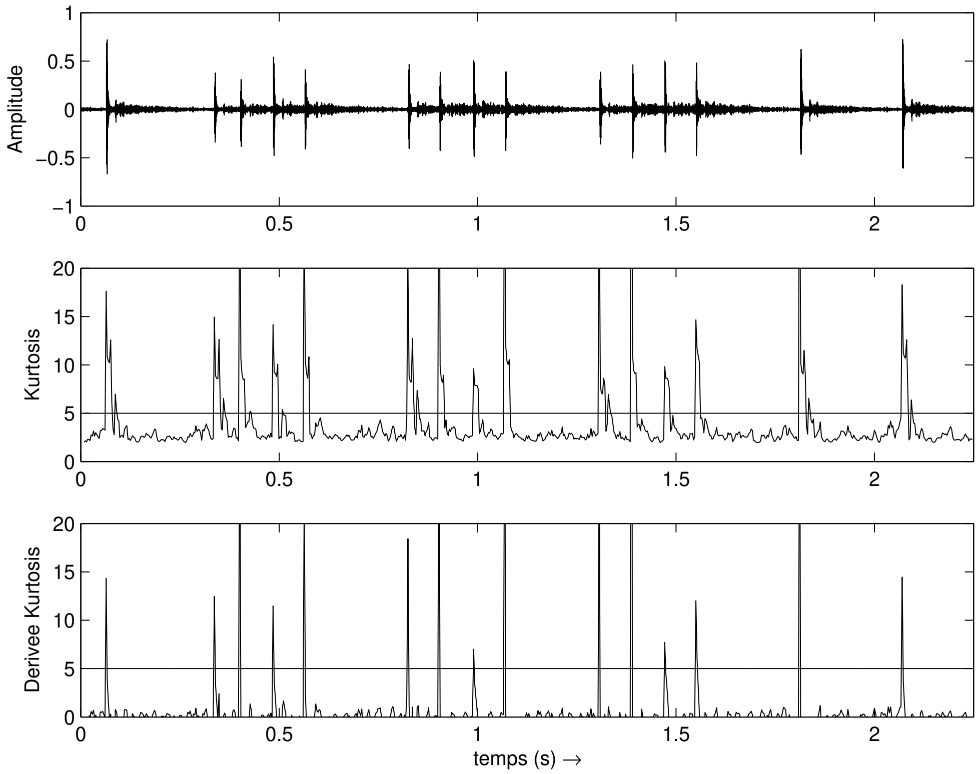

FIG. 3.14 – Illustration de la détection de transitoires par le calcul du kurtosis : la première figure montre le son naturel analysé (son de castagnettes), la deuxième figure présente l'évolution du kurtosis au cours du temps (fenêtre de 512 échantillons, taux de chevauchement de 25%) et la troisième figure les variations du kurtosis (différence par rapport à la moyenne des 4 valeurs précédentes).

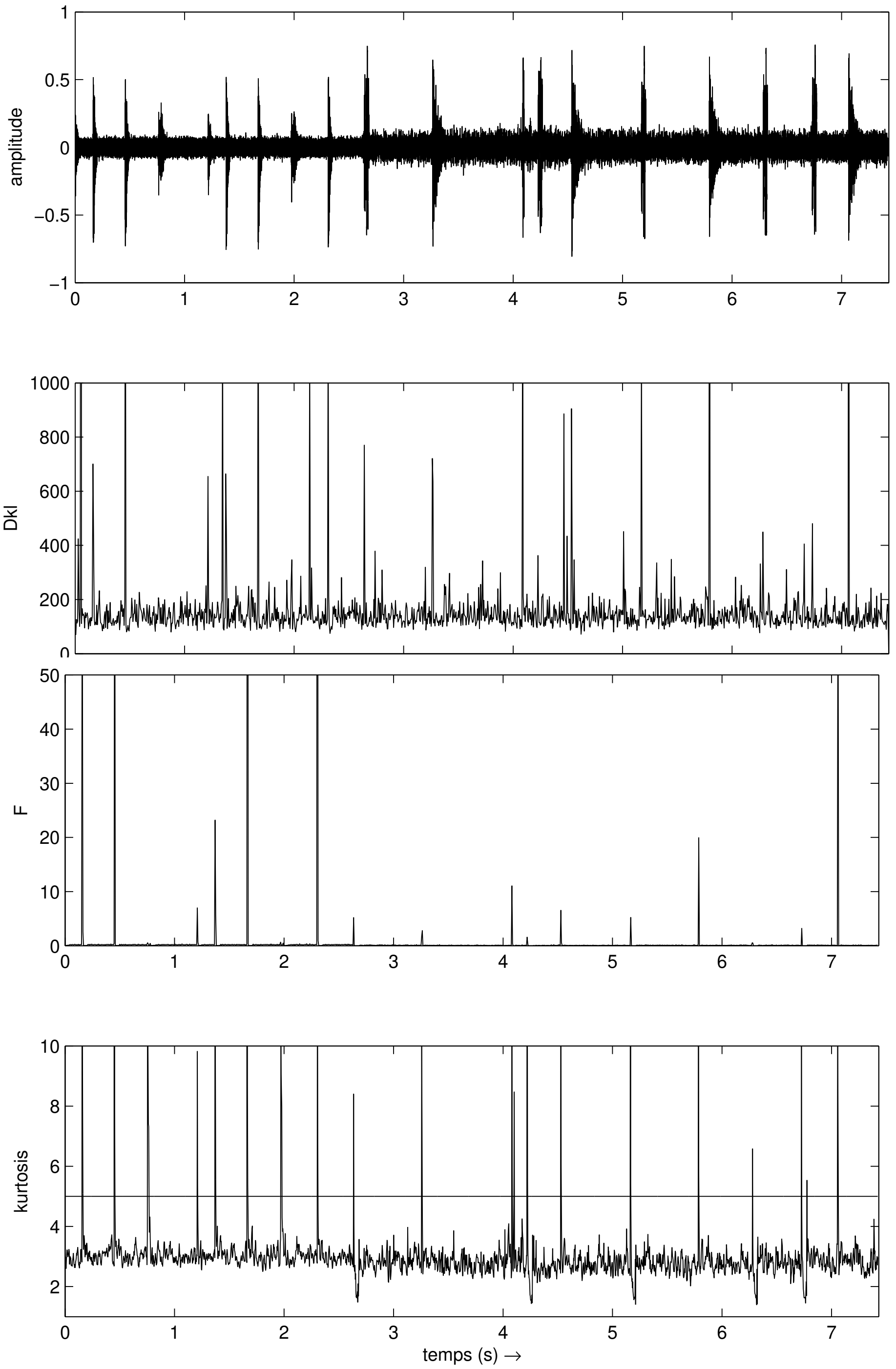

Fig. 3.15 – Détection par différentes méthodes de transitoires dans un son de percussions plongé dans du bruit blanc gaussien : le signal étudié est représenté temporellement sur la première figure alors que les trois autres montrent les résultats obtenus respectivement, par le calcul de la distance de Kullback-Leibler (512 échantillons, 50% de recouvrement), par la méthode du contenu hautes fréquences (512 échantillons, 50% de recouvrement), et enfin par le calcul du kurtosis (512 échantillons, 25% de recouvrement, seuil de 5).

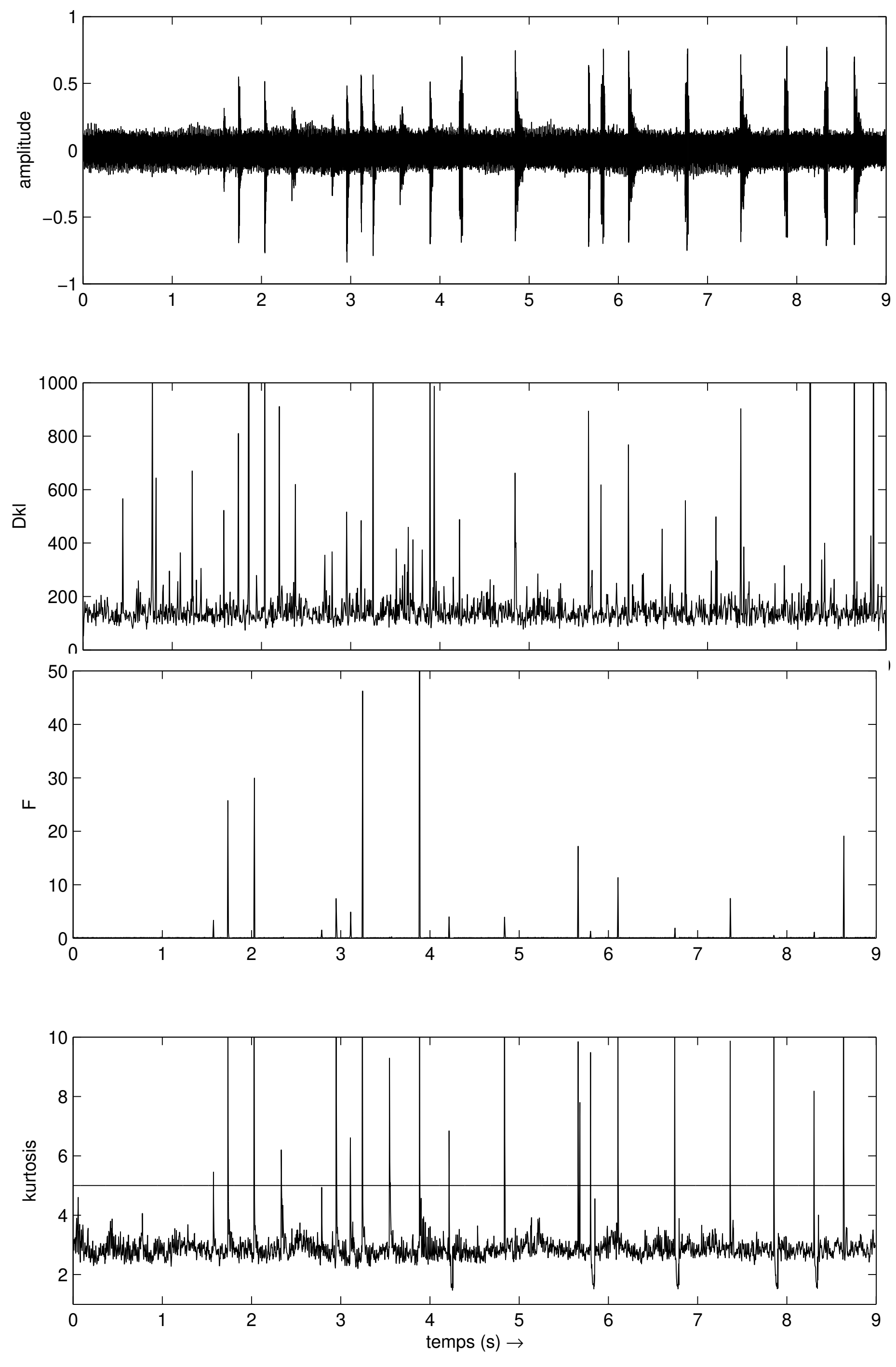

FIG. 3.16 – Détection par différentes méthodes de transitoires dans un son de percussions plongé dans du bruit blanc gaussien : le signal étudié est représentée temporellement sur la première figure alors que les trois autres montrent respectivement les résultats obtenus par le calcul de la distance de Kullback-Leibler (512 échantillons, 50% de recouvrement), par la méthode du contenu hautes fréquences (512 échantillons, 50% de recouvrement), et enfin par le calcul du kurtosis (512 échantillons, 25% de recouvrement, seuil de 5).

sont constantes pendant la durée correspondant à cette taille. Dans le cas idéal, elle ne doit contenir que du bruit ou un signal harmonique, ou alors une transitoire. Une fenêtre contenant une partie d'une transitoire et un bruit sera quantifiée par un kurtosis dont la valeur sera moyennée. Plus cette fenêtre est petite, plus la probabilité de définir exactement une transitoire existe, impliquant un kurtosis très important. Par contre, cette valeur risque de varier plus fortement, même en présence de bruit, comme le montre le tableau 3.10. A l'opposé, une grande fenêtre aura tendance à produire des variations de kurtosis moins importantes et donc à rendre l'influence de la transitoire moins significative.

Le taux de recouvrement est directement relié à la précision de la localisation de la transitoire. Ce taux définit l'avancement de la fenêtre utilisée pour le calcul du kurtosis et donc l'échantillonnage de la courbe de mesure. Plus ce taux est faible, plus le nombre de points de mesure est grand, plus l'influence des transitoires risque d'impliquer des fortes variations de kurtosis. Plus il est fort, plus le nombre de points à calculer est important, et plus la méthode d'analyse est coûteuse en ressources et donc en temps. Nous pouvons d'ailleurs remarquer que l'algorithme de calcul du kurtosis peut être assez facilement amélioré, si le taux de chevauchement est assez faible (inférieur à 50%). Dans ce cas, il est moins coûteux de soustraire et de rajouter les contributions des échantillons à la variance et au moment d'ordre 4.

A la différence des autres méthodes, le seuil peut être fixé, quelle que soit la nature du son étudié. Nous avons ainsi appliqué une valeur de seuil de 5 lors de nos expériences. Cependant, il peut être sujet à une étude plus complète, pour essayer de l'affiner. Plus le son est bruité, plus le seuil doit être proche de 3 pour pouvoir détecter les transitoires les plus faibles. Forcément, cela implique un risque supplémentaire de considérer des transitoires correspondant en réalité à des variations de kurtosis dues à des approximations de calcul.

Conclusion

Nous proposons donc une méthode efficace et précise, qui semble plus robuste à la présence de bruit que les méthodes existantes basées sur l'étude des variations d'énergie. Comme cette méthode s'appuie sur les propriétés statistiques du signal, elle peut être appliquée en complément d'autres techniques, de la même façon que dans [Lev98]. Son autre principal intérêt réside dans la possibilité de fixer le seuil de sélection des transitoires indépendamment de la nature du son étudié, puisque le kurtosis est normalisé et prend la même valeur quel que soit le volume du son. Il est donc possible d'envisager une méthode entièrement automatisée, à condition que le seuil soit précisément défini par un très grand nombre d'expériences sur des bases de données de sons variés. Ce long travail reste à accomplir. Des études restent également à mener pour tenter d'améliorer encore la précision.

3.4 Analyse par la transformée de Fourier

Les méthodes d'analyse des paramètres de modèles spectraux sont basées essentiellement sur les informations de la représentation fréquentielle du signal étudié. Cette représentation fréquentielle est obtenue par la transformée discrète de Fourier. Cette opération est effectuée par un algorithme efficace, appelé la transformée rapide de Fourier (FFT). Nous présentons ici brièvement le principe de cette méthode, ses améliorations possibles, et ses limites, en particulier dans le cadre de l'analyse de sons bruités.

3.4.1 Principe

La transformée de Fourier impose une hypothèse de stationnarité du signal sur son contenu fréquentiel. Comme un son naturel est composé de sinusoïdes de fréquences et d'amplitudes variant dans le temps, et que son spectre évolue donc parfois rapidement dans le temps, il est nécessaire de réduire la durée du signal étudié, pour rendre cette hypothèse de stationnarité la plus valide possible. C'est pourquoi les méthodes d'analyse basées sur la transformée de Fourier considèrent des fenêtres temporelles successives du son étudié, pouvant éventuellement se chevaucher, et les spectres dits à court terme associés. Chacun de ces spectres peut donc s'écrire :

$$X(k) = \frac{1}{W_s} \sum_{n=0}^{W_s-1} x(n)w(n)e^{-j\frac{2\pi}{W_s}nk} \tag{3.38}$$

où x représente le signal de longueur W_s échantillons et $X(k)$ l'estimation du spectre à court terme. La fenêtre rectangulaire w qui définit la portion de signal considéré s'écrit :

$$\begin{cases} w(n) & = 1 \quad ,\forall n \in [\![0, W_s - 1]\!] \\ w(n) & = 0 \quad ,\text{sinon} \end{cases} \tag{3.39}$$

L'intervalle de temps défini par la longueur W_s de la fenêtre détermine l'échelle discrète des fréquences, et en particulier l'écart entre les valeurs obtenues de fréquences f_k :

$$\forall k \in [\![0; W_s]\!], \quad f_k = k\frac{F_s}{W_s} \tag{3.40}$$

Ainsi, plus la taille W_s de la fenêtre est importante, plus l'écart entre les fréquences sera faible, et inversement. Il faut donc effectuer un compromis entre la résolution temporelle et la résolution fréquentielle.

3.4.2 Améliorations

Le spectre continu associé à une sinusoïde pure est théoriquement une fonction de Dirac. Dans le cas discret, ce n'est pas le cas. En effet, d'après les propriétés de la transformée de Fourier présentées dans la section 1.1.5, le spectre obtenu X est en fait le résultat de la convolution entre le spectre S du signal s et le spectre W de la fenêtre w :

$$x = s \times w \Longleftrightarrow X = S * W \tag{3.41}$$

Ainsi, le choix de la fenêtre multiplicative w est prépondérante pour la précision fournie par l'approximation du spectre. Il est donc nécessaire de définir une fenêtre dont les effets seront limités. Par exemple, le spectre des fenêtres est composé de lobes consécutifs qui peuvent donner lieu à plusieurs raies pour un signal composé d'une seule sinusoïde pure. La fenêtre idéale n'existe pas. De nombreuses fenêtres ont été définies avec des caractéristiques différentes permettant de choisir l'une ou l'autre en fonction de l'analyse effectuée. Le choix des fenêtres ne sera pas détaillé ici, car de nombreuses discussions existent à ce sujet dans la littérature [Mar00, Moo90, Roa98]. Nous illustrons juste cette propriété à l'aide de la figure 3.17 qui montre plusieurs fenêtres classiques et pour chacune le spectre d'amplitude associé.

Les spectres obtenus à partir d'un signal numérique par la transformée de Fourier discrétise l'échelle des fréquences. La méthode la plus classique permettant d'interpoler ces valeurs

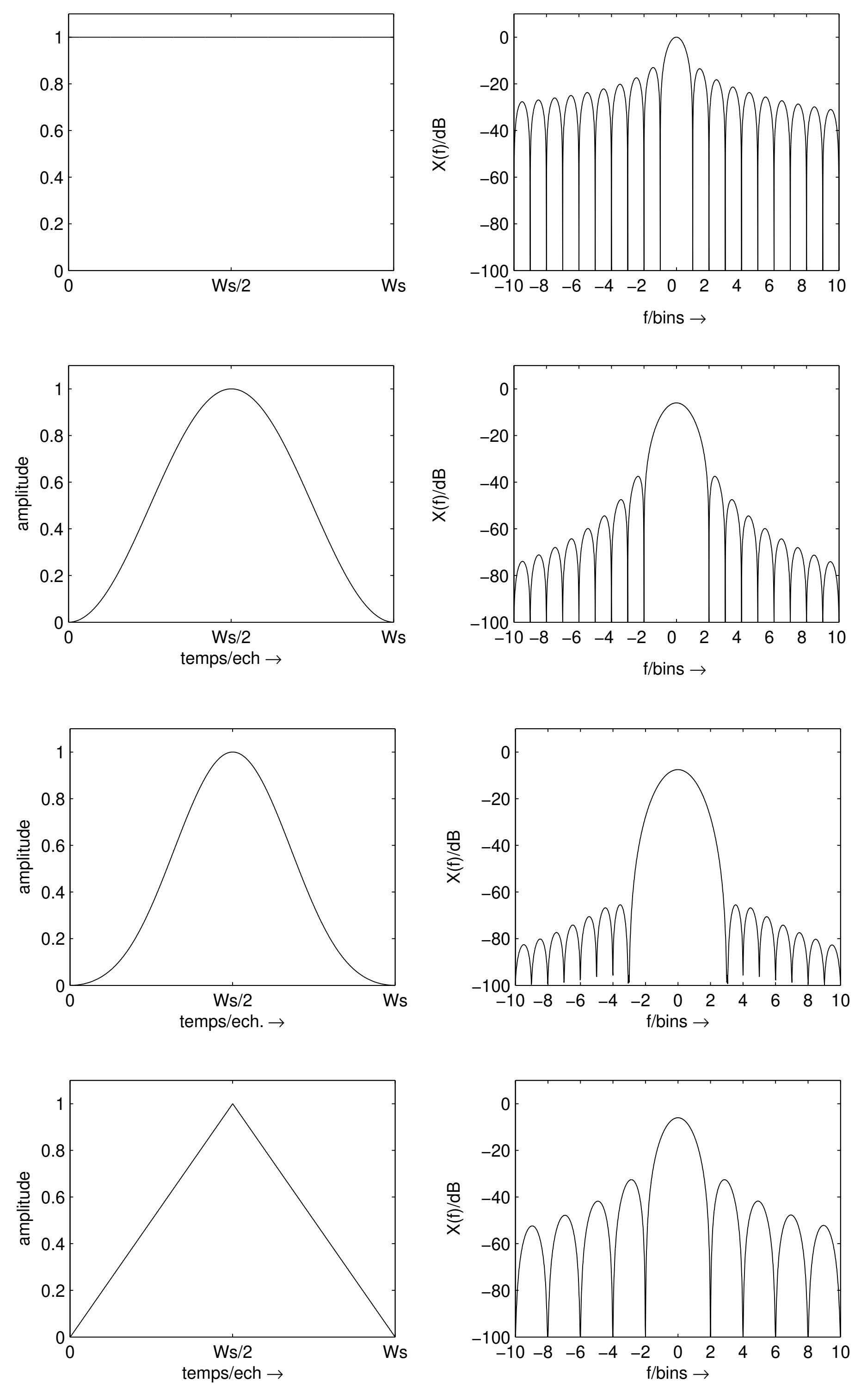

FIG. 3.17 – Fenêtres (respectivement fenêtre rectangulaire, de Hann, de Blackman et de Bartlett employées pour l'analyse par la transformée discrète de Fourier, avec pour chacune à gauche leur représentation temporelle et à droite leur spectre d'amplitude associé.

consiste à augmenter artificiellement la taille W_s de la fenêtre d'analyse en ajoutant des valeurs nulles : c'est le bourrage par zéro (*zero-padding*). Evidemment cette technique a un coût important, puisque la complexité de la transformée rapide de Fourier dépend du nombre d'échantillons. Cette interpolation a toutefois des limites, puisque, par exemple, elle ne permet pas toujours de pouvoir produire deux pics distincts sur le spectre pour deux sinusoïdes de fréquences différentes. La figure 3.18 illustre cette limite, en montrant deux cas de signaux composés de deux fréquences proches et leur spectre associé. Dans le deuxième cas, même si l'interpolation est très importante, un seul pic est visible. Le bourrage par zeros reste une opération d'interpolation et ne rajoute aucune information sur le signal.

D'autres techniques pour améliorer l'estimation des paramètres des partiels ont été mises au point, notamment pour pouvoir augmenter la qualité des modèles sinusoïdaux :

L'interpolation parabolique : Cette technique repose sur la ressemblance entre le lobe principal du spectre d'une fenêtre et une parabole. Par la méthode de Brent, le maximum de la parabole associée aux valeurs discrètes du spectre à court terme est estimé.

L'algorithme du triangle : Keiler et Zölzer proposent d'appliquer une fenêtre correspondant à une fenêtre de forme simple (triangulaire) dans le domaine fréquentiel [KZ01], de façon à pouvoir estimer facilement le maximum, comme pour le cas précédent de la parabole.

Le réassignement spectral : Par la connaissance de la dérivée de la fenêtre d'analyse employée, la fréquence est ajustée à l'intérieur du casier du spectre concerné [AF95].

La dérivée du signal : Le rapport entre le signal et sa dérivée est mathématiquement liée à la fréquence exacte du signal. C'est donc par le calcul de cette dérivée que les paramètres sont estimés [DCM00].

La distortion de phase : Lorsque l'hypothèse de stationnarité des sinusoïdes du signal n'est pas respectée, des distortions significatives apparaissent dans le spectre de phase, permettant d'extraire des informations sur les variations de fréquence et d'amplitude [MC98].

3.4.3 Limites

Même si de nombreuses améliorations ont été apportées à la transformée discrète de Fourier, les limites concernant la résolution fréquentielle et temporelle, et la présence de bruit rendent à priori très difficile son application à l'étude des sons bruités. En effet, le bruit perturbe énormément le spectre obtenu et provoque des erreurs d'estimation des paramètres. Mais surtout, l'approche du modèle CNSS, basée sur le modèle du bruit thermique, semble incompatible avec une analyse basée sur la transformée de Fourier, puisqu'un bruit sera défini comme une somme finie et importante de sinusoïdes, et que donc les fréquences composant le signal seront trop proches et dépasseront la résolution fréquentielle imposée, pour pouvoir être estimées.

3.5 Analyse de l'enveloppe spectrale

La donnée de base de l'analyse spectrale est le spectre discret de Fourier à court terme. Le passage du domaine temporel au domaine spectral préserve toutes les informations du signal. Le modèle CNSS, mais aussi d'autres modèles spectraux, définissent le paramètre couleur,

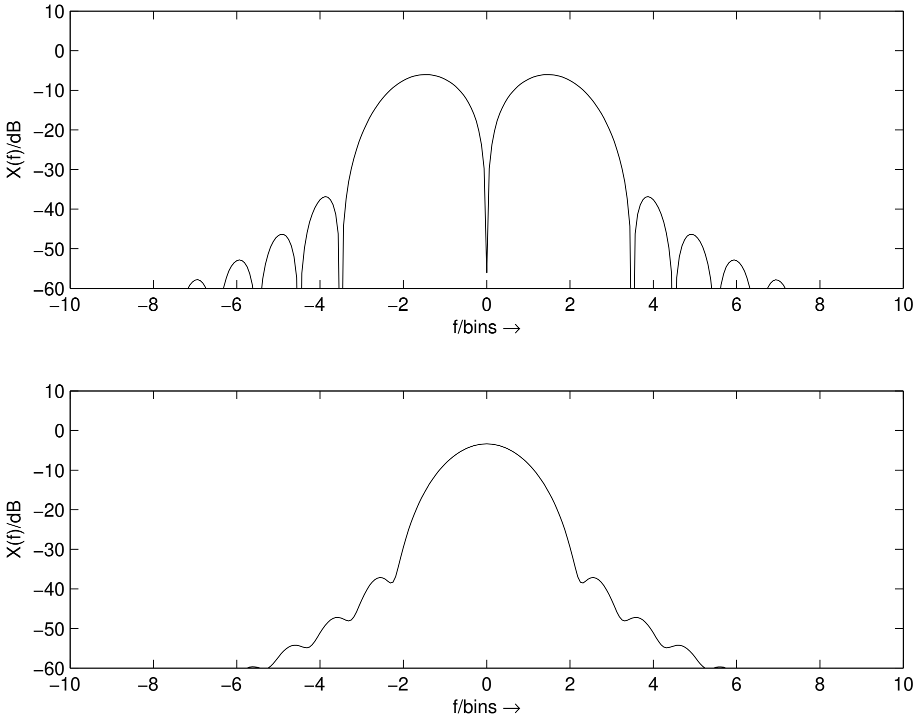

Fig. 3.18 – Les limites de la résolution fréquentielle de la transformée discrète de Fourier apparaît lorsque deux fréquences sont trop proches ($W_s = 1024$ échantillons). Dans la figure du haut, l'interpolation par bourrage de zéro permet de visualiser la présence de deux sinusoïdes de fréquences proches (différence de 2.5 casiers, soit 108Hz), alors que dans la figure du bas, malgré le bourrage de zéro, ces deux fréquences ne peuvent être distinguées (différence de 0.5 casier, soit 21.5Hz).

relatif à la forme générale de l'enveloppe spectrale. Il est donc nécessaire de développer une technique permettant d'extraire la forme de cette enveloppe, à partir du spectre discret à court terme. Les travaux existants à ce sujet sont nombreux. Nous en présentons quelques-uns et montrons quelques résultats sur un même son bruité pour ces différentes techniques.

3.5.1 Prédiction linéaire

La méthode de prédiction linéaire (LPC *Linear Prediction Coding*)[Mak75], déjà introduite dans la section 2.2, est une des plus employées. Elle définit un filtre linéaire, appelé filtre LPC, qui, appliqué à un signal source (par exemple un bruit blanc), permet d'approximer un signal original. Le filtre peut donc s'écrire :

$$y[n] = a_0 x[n] + \sum_{i=1}^{p} b_i y[n-i] \tag{3.42}$$

où a_0 et b_i sont les coefficients du filtre, $x[n]$ le signal source ou l'erreur, et $y[n]$ le signal modélisé. L'indice p, appelé ordre de prédiction, définit donc le nombre de coefficients du filtre. Il existe plusieurs algorithmes de calcul des coefficients b_i, mais certains peuvent mener à la définition de filtres non stables (méthode de covariance). Une méthode classique est la méthode dite d'autocorrélation qui minimise l'énergie de l'erreur de prédiction $\mathsf{E}(e)$:

$$\mathsf{E}(e) = \frac{1}{W_s} \sum_{n=1}^{W_s} x[n]^2 \tag{3.43}$$

Cette opération mène à un système d'équations, qui peut se résoudre par l'algorithme de Levinson et Durbin [Mak75].

La fonction de transfert $\mathcal{H}$ associée à ce filtre est une approximation de l'enveloppe spectrale du signal. Elle peut être calculée à partir des coefficients b_i par une transformée de Fourier avec un bourrage de zéros, puisque le nombre de coefficients p est inférieur au nombre d'échantillons qu'il représente. Ce nombre p définit donc le nombre de points utilisés pour définir l'enveloppe spectrale et donc la qualité de son lissage. Entre ces points, l'enveloppe spectrale est interpolée par le complément à zéro de la transformée de Fourier. Plus l'ordre de prédiction p est important, plus l'enveloppe spectrale est composée de pics et de creux, et plus elle est proche de l'enveloppe originale. Plus il est faible, plus l'enveloppe spectrale est lisse. Un exemple de modélisation de l'enveloppe spectrale d'un son bruité (une consonne "s" prononcée) est représentée par la figure 3.19, pour plusieurs valeurs d'ordre de prédiction p.

Comme le nombre de points p utiles pour représenter le signal original est moins important que le nombre d'échantillons définissant celui-ci, cette technique LPC peut être employée pour le codage de sons et/ou la compression. Néanmoins, pour ce qui concerne le modèle CNSS, ce n'est pas la priorité. Du point de vue du contrôle musical, la manipulation de coefficients mathématiques pour effectuer des transformations est peu intuitive, et semble poser une limite importante, même si il est possible d'envisager des interfaces entre les coefficients du filtre et la fonction de transfert (c'est-à-dire l'enveloppe spectrale) par la transformée de Fourier.

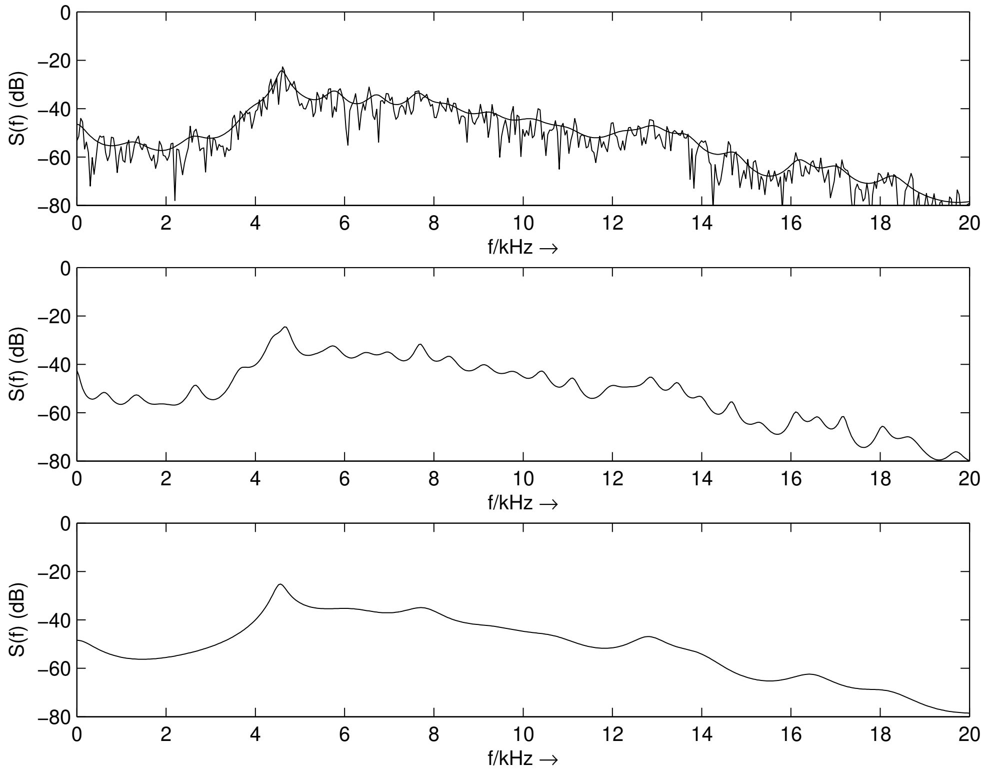

Fig. 3.19 – Illustration de la représentation mathématique de l'enveloppe spectrale d'un son (voix prononçant un "s") par la prédiction linéaire (LPC) : la figure du haut montre l'estimation du spectre par la transformée de Fourier ainsi que l'approximation par la prédiction linéaire par 50 coefficients en traits hachurés, la figure du milieu montre l'approximation par 75 coefficients, et la figure du bas par 25 coefficients.

3.5.2 Filtrage du cepstre

Le cepstre (mot défini par l'inversion des 4 premières lettres de "spectre") est un outil mathématique développé dans le cadre du traitement de la voix. Il est noté c et est défini comme la transformée inverse de Fourier du logarithme de la transformée de Fourier du signal étudié, noté x :

$$\mathsf{c} = \mathcal{F}^{-1}(\log(\mathcal{F}(x))) \tag{3.44}$$

où $\mathcal{F}$ représente la transformée de Fourier.

L'application du cepstre peut être décrite de deux manières différentes [Sch98]. La première définit l'enveloppe spectrale comme une version lissée de la courbe d'amplitude (en décibels) obtenue par la transformée discrète de Fourier, ce qui peut se traduire par un filtrage passe-bas du spectre à court terme. Ainsi, cette opération de filtrage peut s'effectuer en représentant le spectre dans le domaine spectral par une transformée de Fourier (inverse), à donner des valeurs nulles aux fréquences supérieures à un seuil fixé (l'ordre du cepstre), avant de repasser dans le domaine temporel par une transformée de Fourier.

L'autre façon d'appréhender cette méthode est de considérer un signal x comme une source s modifiée par un filtre h :

$$x = s * h \tag{3.45}$$

ce qui revient, en appliquant une transformée de Fourier et en prenant le logarithme du résultat $(\forall(a,b), \log(ab) = \log(a) + \log(b))$, à :

$$\log(X) = \log(S) + \log(H) \tag{3.46}$$

où X, S et H sont respectivement les transformées de Fourier de x, s,et h. Le signal source est supposé n'être constitué que de hautes fréquences. L'enveloppe spectrale du signal original, définie par la contribution de H, est alors donnée par les faibles valeurs de $\log(X)$.

D'un point de vue pratique, l'enveloppe spectrale peut être lissée en suivant l'algorithme suivant. Dans un premier temps, le cepstre d'une fenêtre du signal est évalué selon l'équation (3.44). Ensuite, les p premières valeurs du cepstre sont préservées, et les autres sont mises à zéro. L'ordre p permet donc de définir la force de l'effet de lissage de l'enveloppe spectrale finalement obtenue. Un exemple de modélisation de l'enveloppe spectrale d'un son bruité (une consonne "s" prononcée) obtenue par filtrage du cepstre, est représentée par la figure 3.19, pour plusieurs valeurs d'ordre p.

Le lissage obtenu par cette méthode de filtrage du cepstre ne modifie pas le nombre de points de l'enveloppe lissée. Ce n'est donc pas une méthode envisageable dans le cadre d'un modèle de compression. Elle est peu employée pour des modèles sinusoïdaux car l'enveloppe définie ne passe pas forcément par les pics (maxima locaux) du spectre. Par contre, en ce qui concerne le paramètre couleur du modèle CNSS, cette méthode est parfaitement applicable, puisque cette condition n'est pas nécessaire.

3.5.3 Approximation par des segments

Les techniques de prédiction linéaire et de filtrage du cepstre définissent des enveloppes spectrales qui ne passent pas forcément par les pics des spectres à court terme. Serra propose une autre méthode pour le modèle SMS [SS90]. Une suite de p fenêtres fréquentielles de taille

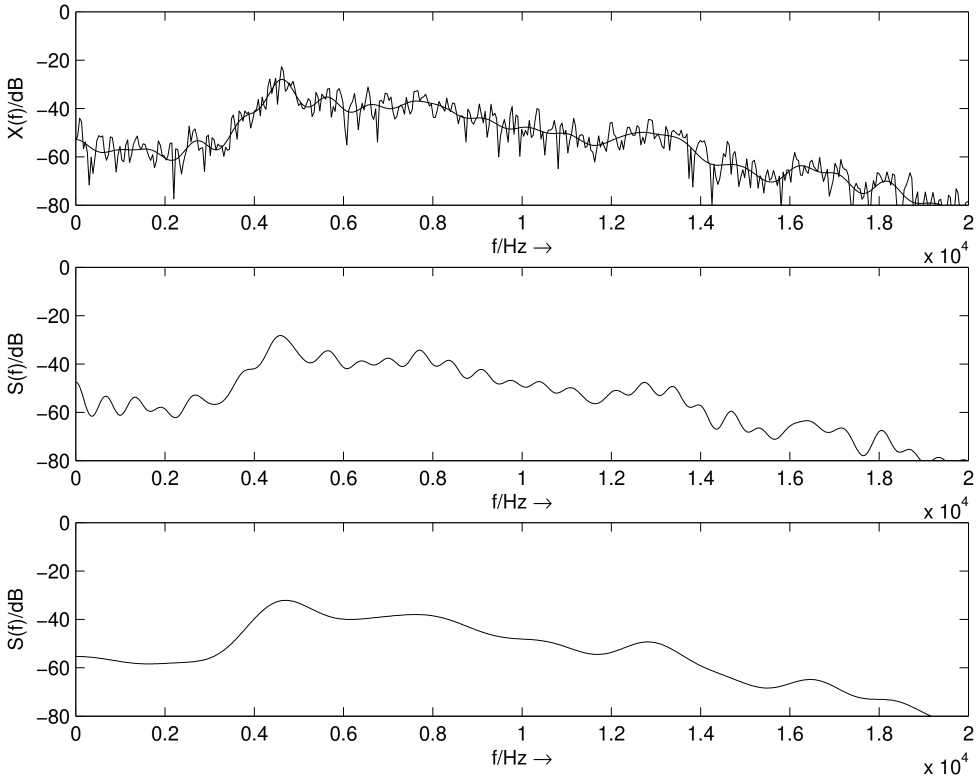

Fig. 3.20 – Illustration de la représentation mathématique de l'enveloppe spectrale d'un son (voix prononçant un "s") par le cepstre : la figure du haut montre l'estimation du spectre par la transformée de Fourier ainsi que l'approximation par le cepstre par 50 coefficients en traits hachurés, la figure du milieu montre l'approximation par 75 coefficients, et la figure du bas par 25 coefficients.

Q échantillons, recouvrant entièrement le spectre de taille W_s (taille de la fenêtre temporelle de signal), est définie :

$$p = \frac{W_s}{2Q} \tag{3.47}$$

$$\forall k \in [\![0; p-1]\!], \quad I_k = [kQ; (k+1)Q[\tag{3.48}$$

où I_k sont les intervalles de fréquence définis par les p fenêtres fréquentielles de taille Q. Pour chacun de ces p intervalles I_k, la valeur du maximum d'amplitude du spectre est cherchée. Elle est affectée à l'intervalle des fréquences correspondant, et plus particulièrement à la fréquence du centre f_c^k de cet intervalle :

$$f_c^k = \frac{k+1}{2}Q \tag{3.49}$$

Ainsi, un nouveau spectre noté $\widetilde{S}$ est défini à partir de p points :

$$\widetilde{S}(k) = \max_{i \in [\![\frac{-Q}{2}; \frac{Q}{2}-1]\!]} |S(i+kQ)| \tag{3.50}$$

où k est un entier compris entre 0 et $p-1$. Les fenêtres fréquentielles peuvent éventuellement se chevaucher, ce qui augmente le nombre de points p définis. Ce nombre p doit évidemment être inférieur au nombre de casiers du spectre discret obtenu par transformée de Fourier :

$$p < \frac{W_s}{2} \tag{3.51}$$

Pour les autres $\frac{W_s}{2} - p$ valeurs de $\widetilde{S}$ non définies, l'amplitude est linéairement interpolée à partir des deux valeurs d'amplitude voisines. Un exemple de modélisation de l'enveloppe spectrale d'un son bruité (une consonne "s" prononcée) obtenu par segments, est représenté par la figure 3.21, pour plusieurs valeurs d'ordre p.

Le paramètre p définit donc le nombre de points composant l'enveloppe spectrale lissée. Plus ce paramètre est important, moins l'enveloppe spectrale est lissée. Il est possible également de modifier le taux de chevauchement des fenêtres pour définir plus de points. Par ailleurs, dans l'approche que nous avons présentée, les points sont régulièrement répartis. Il est possible de sélectionner des points particuliers, de façon à imposer à l'enveloppe spectrale de passer par ces points. Cela peut par exemple s'avérer utile dans la modélisation sinusoïdale [SR99].

Cette méthode représente l'enveloppe spectrale d'un son à partir d'un nombre réduit de points, ce qui rend son application possible pour la compression. Elle a également été choisie dans le modèle SMS en raison de sa flexibilité. De plus, elle permet d'assurer, avec un nombre raisonnable de points, le passage de l'enveloppe spectrale par les pics du spectre. Néanmoins, pour une analyse destinée au modèles de bruits, en particulier le modèle CNSS, ces propriétés sont soit non requises (compression), soit non désirées (passage par le pics).

3.5.4 Autres méthodes

Les méthodes de prédiction linéaire et de filtrage du cepstre peuvent être appliquées à tout type de son alors que la méthode de segmentation est plutôt destinée aux sons harmoniques. D'autres méthodes spécifiques à des natures de sons particulières ont été proposées.

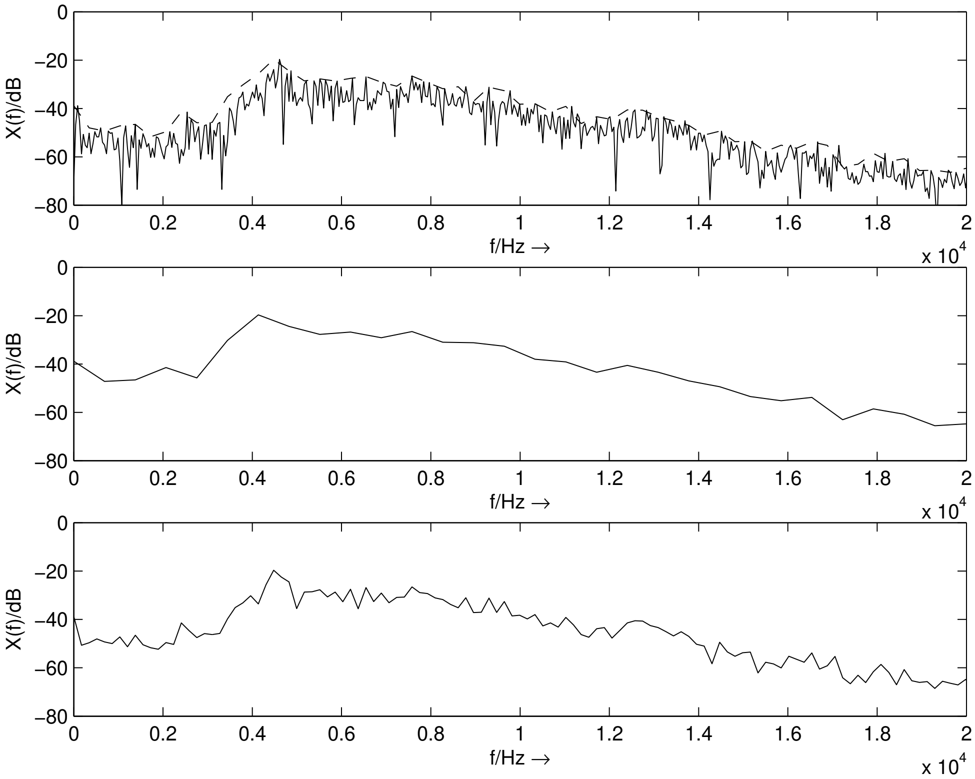

Fig. 3.21 – Illustration de la représentation mathématique de l'enveloppe spectrale d'un son (voix prononçant un "s") par segmentation : la figure du haut montre l'estimation du spectre par la transformée de Fourier ainsi que l'approximation par 64 segments linéaires en traits hachurés, la figure du milieu montre l'approximation par 32 segments, et la figure du bas par 128 segments.

Par exemple, le modèle SAS propose un paramètre couleur liée à l'enveloppe spectrale, représenté par l'interpolation de la courbe définie par les pics correspondant à des partiels du spectre discret à court terme [Mar00].

Laurenti et DePoli proposent une méthode d'estimation de l'enveloppe spectrale du résidu [LG00].En considérant une courte fenêtre de signal (taille W_s), l'inverse du spectre R est moyennée par un filtre médian sur N_f points :

$$\overline{R}(k) = \frac{1}{N_f} \sum_{h=-\frac{N_f}{2}}^{\frac{N_f}{2}-1} R_{(k-h)\mathrm{mod}W_s} \tag{3.52}$$

L'inverse de $R(k)$ constitue une approximation du spectre du résidu. Celle-ci est alors ensuite lissée, par des techniques classiques. Toutefois, cette approche donne des résultats très similaires aux techniques LPC et de filtrage du cepstre, pour des ordres d'approximation faibles.

3.5.5 Conclusion

Les expériences que nous avons effectuées montrent que les méthodes de prédiction linéaire et de filtrage du cepstre sont applicables à l'estimation de l'enveloppe spectrale pour des sons bruités. En ce qui concerne le modèle CNSS, l'ordre d'approximation doit être relativement faible de façon à ce que la couleur soit définie par une courbe très lisse.

3.6 Estimation de la densité spectrale

Le modèle CNSS que nous avons développé représente à l'aide de paramètres mathématiques et statistiques la densité spectrale d'un bruit. Il envisage des transformations musicales liées à cette propriété montrée comme importante sur le plan perceptif. Il est donc nécessaire de proposer une méthode d'estimation de la densité spectrale d'un son. Ce problème s'avère être très difficile du fait des limites dues à la discrétisation du signal étudié et à sa complexité. A notre connaissance, aucun travail de recherche n'a été mené sur ce domaine. Néanmoins, nous avons réussi à développer une méthode permettant d'estimer la densité spectrale de bandes de bruits [HDC03a]. Nous la présentons et la discutons dans cette section.

3.6.1 Limites des techniques habituelles

L'analyse de la distribution de l'énergie dans les fréquences requiert généralement l'utilisation de la transformée de Fourier à court terme. Deux limites essentielles nous amènent ici à considérer d'autres approches. Elles sont illustrées par la figure 3.29, qui montre un exemple de bruit dont la densité spectrale n'est pas constante, et dont l'estimation du spectre par la transformée de Fourier ne permet pas de mettre en avant ces variations. Le son dont le spectre à court terme est montré, est composé d'une bande de bruit de largeur 500Hz, dont la densité spectrale est faible (un quart des sinusoïdes théoriquement requises est synthétisé). La première limite de la transformée de Fourier à court terme concerne la résolution et le problème classique du compromis entre la précision fréquentielle et la précision temporelle (voir la section 3.4). La seconde limite est liée à l'analyse de la densité spectrale. Une idée basique serait de détecter les creux (minima locaux) de la même manière que les pics (maxima locaux) sont considérés dans les modèles sinusoïdaux. Une détermination des seuils serait alors nécessaire et ceux-ci seraient alors forcément liés à l'amplitude général du signal étudié. De plus, les expériences montrent que les approximations induites par la transformée discrète de Fourier mènent à des creux d'amplitude qui ne sont pas toujours dus à une diminution locale de densité spectrale. C'est pour ces raisons que nous avons été amenés à envisager une autre méthode basée sur l'étude statistique des fluctuations d'intensité, dérivée de la méthode employée pour la détection de pics dus à des sinusoïdes et présentée dans la section 3.2.

3.6.2 Principe de la méthode d'estimation de la densité spectrale

Analyse des fluctuations d'intensité

Les fluctuations d'intensité ont été étudiées et modélisées dans le but d'expliquer la faculté des humains à discerner des bruits de densités spectrales différentes [HMGB86]. Une autre étude théorique de ces fluctuations d'intensité, détaillée dans la section 3.2, mène à des résultats comparables [HDC03d]. La variance de la puissance de l'enveloppe temporelle de tout signal est liée au nombre de sinusoïdes le composant. Nous rappelons la définition de cette variance normalisée, notée V_{NEP} :

$$V_{\text{NEP}} = \frac{V(E^2)}{\langle E^2 \rangle^2} \tag{3.53}$$

Nous considérons dans cette approche des fenêtres fréquentielles étroites. En supposant que le spectre d'amplitude est constant dans cette bande définie, nous avons montré que la mesure V_{NEP} est directement reliée au nombre de sinusoïdes représentant le signal :

$$V_{\text{NEP}} = 1 - \frac{1}{N} \tag{3.54}$$

La méthode que nous proposons consiste à produire des valeurs V_{NEP} successives à partir de fenêtres fréquentielles, pouvant éventuellement se chevaucher, du signal analysé. Les valeurs obtenues consécutives représentent donc une approximation des fluctuations d'intensité, et sont donc relatives au nombre de sinusoïdes composant ce signal. En effet, si la bande de fréquences analysée est composée de bruit, ce qui revient à une somme de plusieurs sinusoïdes (modèle du bruit thermique, voir la section 2.4.2), la mesure V_{NEP} est proche de 1 ($N >> 1$). A l'opposé, si la bande n'est composée que d'une seule sinusoïde ($N \approx 1$), la mesure V_{NEP} est presque nulle. La figure 3.22 illustre le principe de cette méthode.

Pour tenter d'atténuer les limites de cette technique basée sur une mesure statistique, et donc d'améliorer la précision, nous déterminons un seuil en dessous duquel les valeurs de la mesure V_{NEP} seront considérées comme significatives de la présence d'une seule sinusoïde. Par ailleurs, il est important de rappeler que la mesure V_{NEP} est indépendante du volume général de la bande de signal étudiée. Le seuil employé par la méthode peut donc être défini sans avoir de connaissances préalables sur la nature du son étudié.

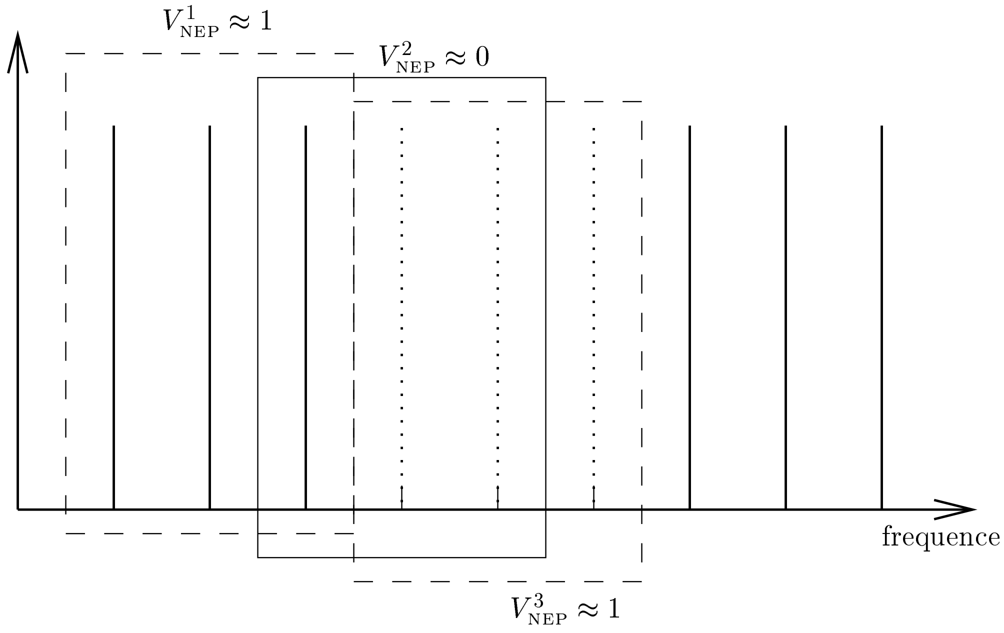

FIG. 3.22 – Application de l'étude des fluctuations d'intensité à l'approximation de la densité spectrale : la première fenêtre fréquentielle constituée de 3 sinusoïdes est associée à une valeur V^1_{NEP} grande (proche de 1) de même que la troisième fenêtre qui est constituée de faible bruit, donc de sinusoïdes ($V^3_{\text{NEP}} \approx 1$). A l'opposé, la deuxième fenêtre composée d'une seul sinusoïde, donnera une valeur V_{NEP} proche de 0.

Hypothèses

La méthode d'estimation de la densité spectrale que nous proposons repose sur deux hypothèses principales. La première concerne les variations de l'enveloppe spectrale. Cette enveloppe est un paramètre de tous les modèles spectraux hybrides. Parfois même c'est le seul paramètre de la partie stochastique (bruit) du modèle. L'étape d'analyse associée à cette enveloppe spectrale pour ces modèles consiste à définir des points de la courbe d'amplitude obtenue par la transformée de Fourier de façon à pouvoir interpoler à partir de ces points une version lissée de cette enveloppe spectrale. L'une des techniques possibles est d'atténuer les hautes fréquences de l'enveloppe spectrale (filtrage passe bas). Un exemple d'enveloppe spectrale ainsi lissée (méthode du cepstre) est présentée sur la figure 3.29. C'est à cause de cette interpolation et de ce lissage que nous pouvons considérer dans une bande de fréquence étroite que l'enveloppe spectrale reste constante. Nous supposons en tout cas qu'elle ne varie pas suffisamment pour introduire des variations de valeurs de V_{NEP} pour une telle bande étroite, les amplitudes des sinusoïdes représentant cette bande étant constantes. La justesse de cette hypothèse dépend évidemment de la largeur de la bande considérée. Si une bande trop large est choisie, cette hypothèse ne peut être vérifiée, et des erreurs de précision en résultent.

La seconde hypothèse concerne l'approximation de la densité spectrale des bandes de fréquence étudiées. Comme nous calculons le nombre de valeurs V_{NEP} qui sont inférieures à un seuil préalablement fixé, nous supposons que la densité spectrale est constante dans chaque bande de fréquence analysée. Bien sûr, là aussi, la validité de cette hypothèse dépend principalement de la largeur de la bande considérée. Plus la bande étudiée est large, plus la probabilité que la densité spectrale soit constante est faible.

3.6.3 Algorithme

Dans cette sous-section, nous détaillons les étapes successives de l'algorithme que nous proposons pour approximer la densité spectrale d'une bande de bruit.

Filtrage passe bande

La première étape de cet algorithme consiste à filtrer le signal original de façon à ne prendre en compte qu'une bande de fréquence (filtre passe bande) pour laquelle seront estimées ensuite les fluctuations d'intensité. Ce processus peut être effectué à partir d'un grand nombre d'échantillons du son original, puisque ce choix n'impose aucune supposition de stationnarité. La largeur de la bande du filtre, notée ΔF_0, est un paramètre de la méthode d'estimation. Ces bandes de fréquence peuvent bien entendu se chevaucher. Nous notons H_0 la différence relative entre deux bandes consécutives. Si $H_0 \geq 1$, les bandes ne se chevauchent pas.

Calcul de la mesure V_{NEP}

La mesure V_{NEP} est obtenue à partir de la puissance de l'enveloppe temporelle du signal et est associée à la valeur de fréquence F_0 située au centre de la bande étudiée ΔF_0. La figure 3.23 propose une illustration de cette étape de l'algorithme. Comme nous l'avons déjà vu, il existe deux moyens de générer la puissance de l'enveloppe : soit en sommant les carrés des échantillons puisque les puissances du signal et de l'enveloppe sont proportionnelles [Har97], soit en annulant les parties négatives du spectre obtenu par la transformée de Fourier. Comme le filtrage précédent peut également être effectué à partir du spectre discret, la seconde approche sera préféré. Cependant, le calcul à partir de la puissance du signal est plus efficace.

Seuillage

Une fois que les valeurs V_{NEP} ont été produites pour chaque valeur de fréquence (au pas choisi près), l'étape suivante consiste à compter le nombre de valeurs $V_{\text{NEP}}(f)$ pour la bande de fréquence étudiée $\Delta F'$ qui se situent sous le seuil choisi t_h. Elle s'effectue simplement par des comparaisons successives. Après cette étape, un nombre de valeurs, noté $\mathcal{N}$, est affecté à chaque bande de fréquence étudiée $\Delta F'$, et plus particulièrement à la fréquence associée, notée F, située au centre de cette bande. La figure 3.24 illustre ce processus.

A cette étape de l'algorithme, nous considérons une bande de fréquence de largeur $\Delta F'$ et le nombre $\mathcal{N}$ associé de valeurs de la mesure V_{NEP}. Or ce nombre n'est pas normalisé puisqu'il dépend des paramètres de la méthode d'approximation, tels que les largeurs de bande et les seuils employés. C'est pourquoi une comparaison avec des valeurs de référence préalablement définies par des expériences est nécessaire. Le tableau 3.11 propose quelques valeurs $\mathcal{N}$ calculées à partir de 100 signaux de densité spectrale théoriquement équivalente, pour deux valeurs de seuil différentes. La densité spectrale de la bande étudiée serait donc approximée par la valeur de référence la plus proche de celle obtenue. Pour éviter cette étape de calibration

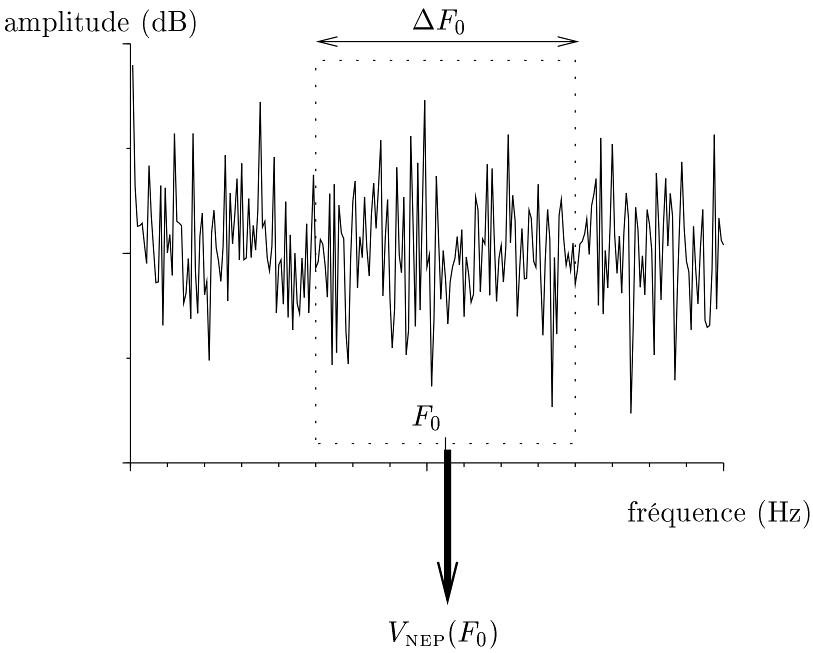

FIG. 3.23 – Illustration de la première étape de l'algorithme d'approximation de la densité spectrale : la mesure V_{NEP} relative aux fluctuations d'intensité est calculée pour une fenêtre fréquentielle de largeur ΔF_0 et est associée à la fréquence F_0, centre de cette fenêtre.

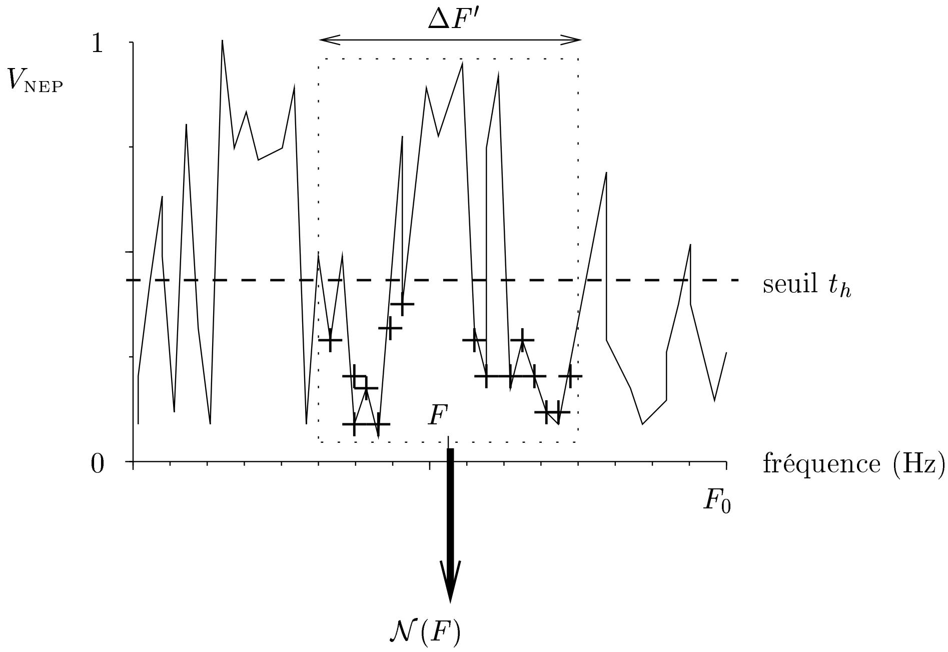

FIG. 3.24 – Illustration de la deuxième étape de l'algorithme d'approximation de la densité spectrale : un nombre $\mathcal{N}$ de valeurs V_{NEP} situées sous le seuil t_h est affecté à chaque bande de fréquence $\Delta F'$ centrée en F.

complexe, nous proposons une autre technique qui consiste à trouver la largeur qui implique le plus grand nombre $\mathcal{N}$.

taux de sinusoïdes	Seuils	
	0.225	0.325
bruit blanc	3.46	11.94
1.0	3.35	10.910
0.5	3.12	10.47
0.33	22.00	36.59
0.25	43.96	57.32

TAB. 3.11 – Résultats moyens (100 signaux) du nombre $\mathcal{N}$ de valeurs V_{NEP} calculés pour des bruits synthétiques de densité spectrale connue, en fonction du seuil t_h et du taux de sinusoïdes utilisées (un taux de 1 signifie que la densité spectrale est considérée maximale). Les paramètres d'analyse sont $\Delta F_0 = 172\text{Hz}$, $\Delta F' = \frac{F_s}{2}$, $H_0 = 0.125$ et V_{NEP} est calculé avec $w_3 = 1024$ échantillons.

Maximisation

Un nombre, noté $\mathcal{N}$, de valeurs de V_{NEP} inférieures à un seuil fixe est donc calculé pour chaque bande de fréquence $\Delta F'$, assimilée à la fréquence centrale, notée F. Nous supposons que le spectre est composé de N fréquences, notées f_i, telles que :

$$\forall i \in [\![0; N-2]\!], \quad f_i < f_{i+1} \tag{3.55}$$

Nous énonçons la propriété suivante que nous allons appliquer dans cette sous-section :

Propriété : Le nombre $\mathcal{N}$, qui dépend de la taille ΔF_0 de la fenêtre fréquentielle considérée pour la première étape, est maximum lorsque cette largeur ΔF_0 de bande correspond à la différence $f_{i+1} - f_i$ entre deux fréquences consécutives du spectre analysé.

Pour justifier cette propriété, nous considérons l'exemple d'un son de durée W_s échantillons, composé d'une bande de bruit de largeur ΔF, produite par $\frac{\Delta F W_s}{F_s}$ sinusoïdes de fréquences régulièrement réparties. Ce son est supposé être de densité maximale (voir la section 2.4.8). Nous considérons à présent un son, obtenu à partir du précédent en supprimant régulièrement 3 sinusoïdes sur 4. Des trous d'énergie perceptibles apparaissent, dus à des différences importantes entre les fréquences.

Nous envisageons les deux cas distincts suivants, à partir de deux sons différents :

- La première possibilité à envisager est une bande de largeur $\Delta F'$ suffisamment importante pour que plusieurs trous d'énergie consécutifs soient pris en compte. La figure 3.25 illustre une telle bande. Dans ce cas, la largeur ΔF_0 induisant le plus grand nombre de fenêtres composées d'une seule fréquence, est égale à la différence créée entre les fréquences (ici, 4 casiers, c'est-à-dire $\frac{4F_s}{W_s}$Hz). En effet, pour une valeur ΔF_0 plus grande, plusieurs fréquences sont dans la bande définie, ce qui entraîne une augmentation de la mesure V_{NEP}. Pour une valeur ΔF_0 plus petite, le nombre $\mathcal{N}$ de fenêtres fréquentielles induisant une mesure V_{NEP} nulle est moins important.

– La seconde possibilité concerne une bande de largeur $\Delta F'$ suffisamment réduite pour qu'un seul trou d'énergie soit présent. La figure 3.26 illustre une telle bande. Dans ce cas, pour toutes les largeurs ΔF_0 inférieures à la différence entre les fréquences (4 casiers), le nombre $\mathcal{N}$ de fenêtres fréquentielles induisant une mesure V_{NEP} nulle est constant. Pour des largeurs supérieures, comme la fenêtre fréquentielle contient toujours au moins 2 fréquences, le nombre $\mathcal{N}$ est plus faible. Ainsi, la plus grande largeur ΔF_0 correspondant au nombre $\mathcal{N}$ maximum est bien égale à la différence créée entre les fréquences.

Les expériences que nous avons menées confirment cette théorie, même si les premières limites de cette approche apparaissent pour le cas d'un seul trou d'énergie.

L'algorithme de maximisation de $\mathcal{N}$ consiste donc à effectuer une itération sur la valeur de ΔF_0 dans le but de trouver la valeur $\Delta F_0^{i_0}$ qui induit la valeur maximale de $\mathcal{N}$, notée $\mathcal{N}^{i_0}$:

$$\forall i, \quad \mathcal{N}^i(F) \leq \mathcal{N}^{i_0}(F) \tag{3.56}$$

La valeur de largeur de bande de fréquence $\Delta F_0^{i_0}$ est supposée être la différence entre deux fréquences consécutives de sinusoïdes, ou, pour le formuler de manière différente, la taille du *trou* d'énergie dans le spectre. A la fin de cette étape, une valeur $\Delta F_0^{i_0}$ est affectée à chaque fréquence F considérée.

Approximation de la densité spectrale d'une bande de bruit

L'algorithme présenté permet d'associer à chaque sous-bande de fréquence d'un bruit analysé une valeur de largeur de fréquence $\Delta F_0^{i_0}$ correspondant au nombre $\mathcal{N}^{i_0}$ de valeurs de la mesure V_{NEP} inférieures à un seuil t_h. Ainsi, une courbe discrète de valeurs de largeur $\Delta F_0^{i_0}$ est obtenue en fonction de la fréquence. Un exemple d'une telle courbe est donné par la figure 3.29. Cette courbe est liée à la densité spectrale du son puisqu'elle représente les différences entre les sinusoïdes du signal étudiée. Ainsi, un pic à une valeur de $\Delta F_0^{i_0} = 4$ indique un trou d'énergie dans le spectre de 4 casiers, ce qui peut être assimilé à l'absence de 3 sinusoïdes consécutives (1 seule présente au lieu de 4), et signifie donc une densité spectrale 4 fois moins importante. La courbe des variations de $\Delta F_0^{i_0}$ peut donc être vue comme la courbe de variations de l'inverse de la densité spectrale du son.

En pratique, les conclusions relatives à ces courbes calculées ne sont pas si simples à établir. Tout d'abord, il est important de remarquer que de nombreuses approximations sont effectuées au cours des différentes mesures, notamment pour les calculs de V_{NEP}. De plus, l'utilisation de fenêtres pour l'étude de la composition spectrale induit une perte de précision sur la localisation fréquentielle des trous d'énergie. Malgré ces hypothétiques, mais fort probables erreurs de mesure, les résultats demeurent exploitables. En effet, la qualité requise de précision d'une méthode d'analyse de propriétés auditives d'un son dépendent évidemment de la précision de la perception pour ces propriétés. En ce qui concerne la perception de la densité spectrale, le système auditif ne semble pas la percevoir avec une grande précision (voir la section 2.4.4). En effet, deux bandes de bruit de même densité spectrale moyenne peuvent comporter des trous d'énergie plus importants (dans une certaine mesure) sur certaines parties du spectre, et des bandes très denses sur d'autres parties. Pourtant l'oreille ne perçoit que la densité moyenne, sauf dans le cas de trous vraiment larges. De plus, pour reproduire un son équivalent sur le plan de la perception, il n'est pas nécessaire de retenir les informations exactes sur la

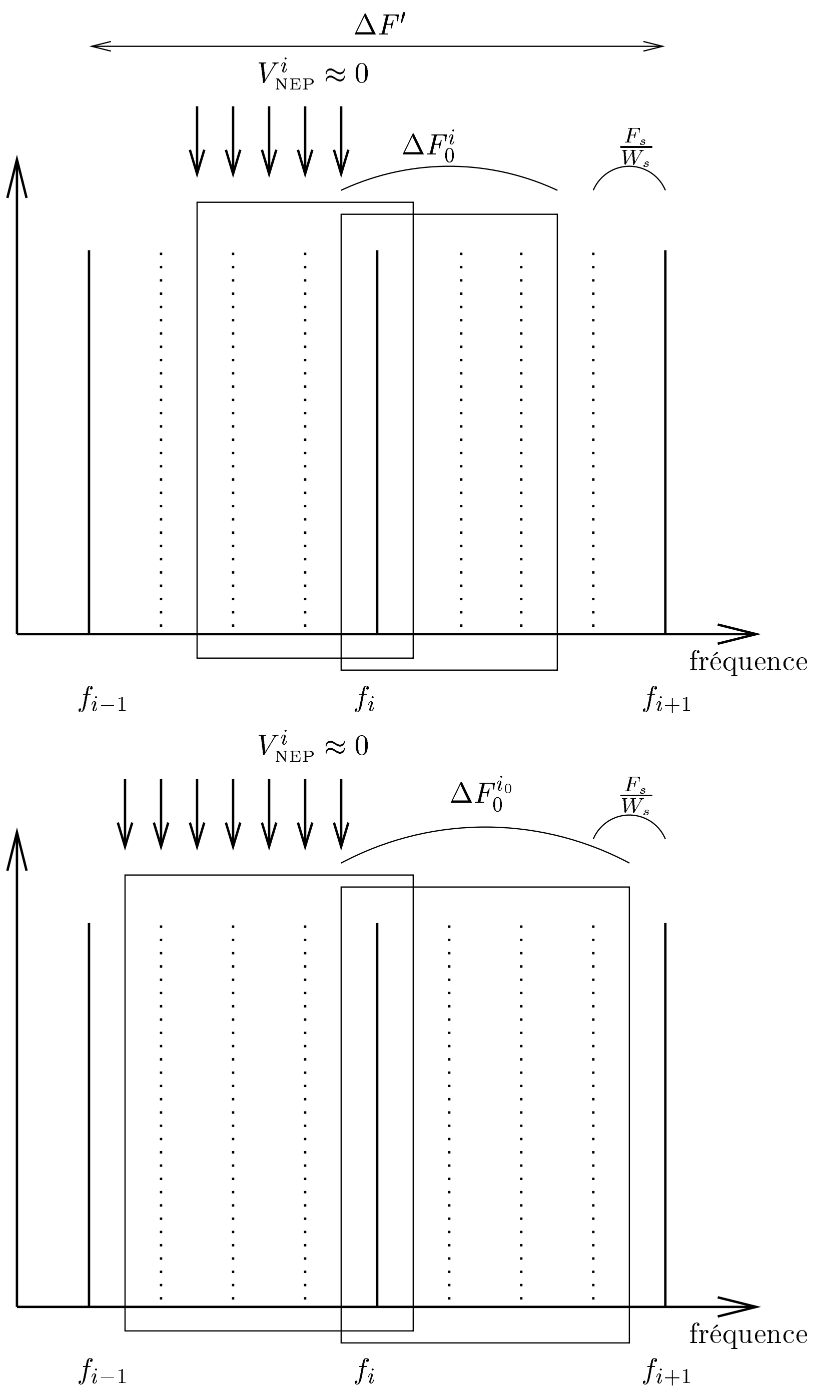

FIG. 3.25 – Illustration de la propriété énoncée concernant la maximisation de la largeur de la fenêtre fréquentielle ΔF_0, dans le cas d'une bande de bruit constituée de plusieurs trous réguliers d'énergie : le nombre de valeurs de la mesure V_{NEP} est moins important lorsque cette largeur ΔF^i_0 est inférieure (figure du haut) à la largeur correspondant à la différence entre deux fréquences consécutives $\Delta F^{i_0}_0 = f_{i+1} - f_i$ (figure du bas). Une flèche matérialise une possibilité de fenêtre donnant une mesure V_{NEP} nulle, dans le cas d'un chevauchement de 50%. Le nombre de flèches représente donc la valeur de $\mathcal{N}$ théorique.

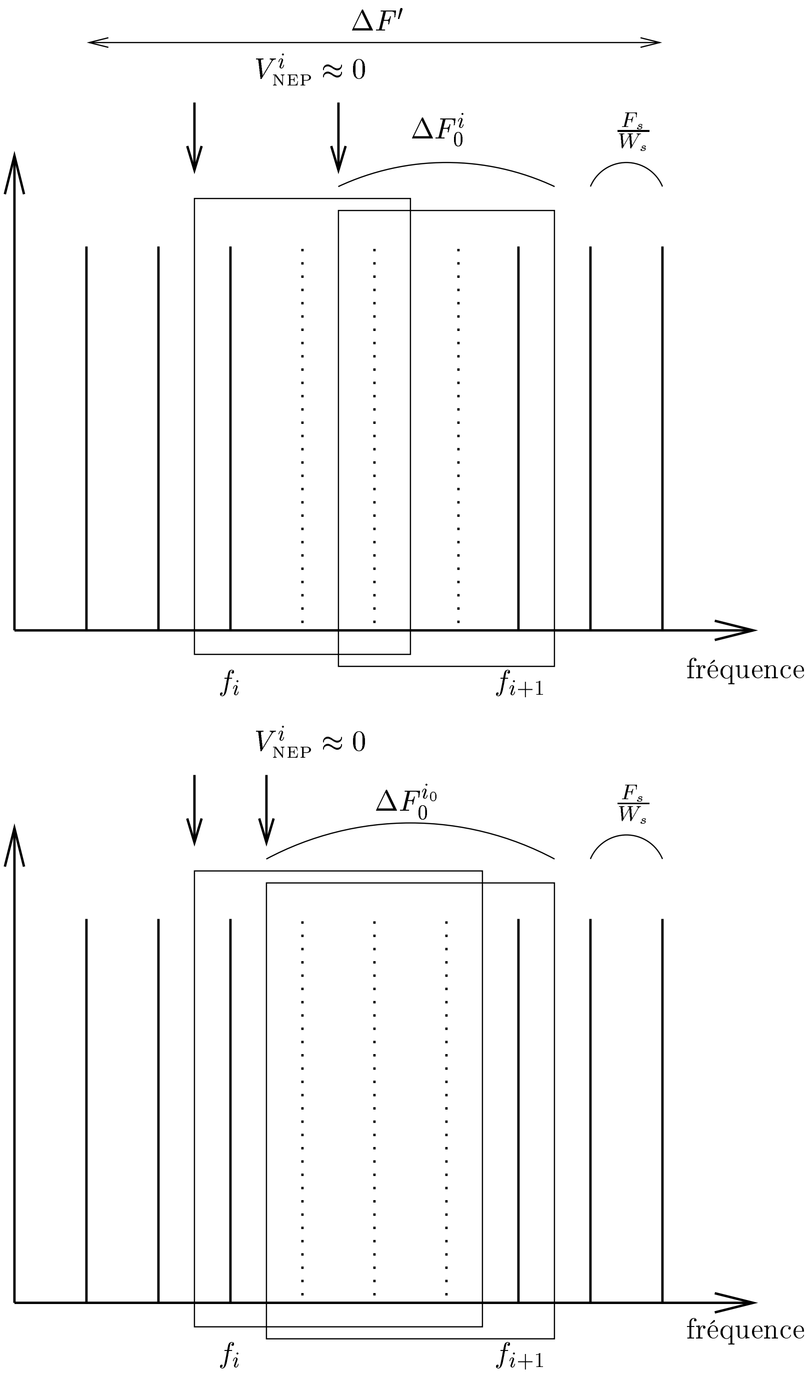

FIG. 3.26 – Illustration de la propriété énoncée concernant la maximisation de la largeur de la fenêtre fréquentielle ΔF_0, dans le cas d'une bande de bruit constituée d'un seul trou d'énergie : le nombre de valeurs de la mesure V_{NEP} est équivalent lorsque cette largeur ΔF^i_0 est inférieure (figure du haut) à la largeur correspondant à la différence entre deux fréquences consécutives $\Delta F^{i_0}_0 = f_{i+1} - f_i$ (figure du bas). Une flèche matérialise une possibilité de fenêtre donnant une mesure V_{NEP} nulle, dans le cas d'un chevauchement de 50%. Le nombre de flèches représente donc la valeur de $\mathcal{N}$ théorique.

répartition d'énergie dans le spectre. Le modèle CNSS repose sur cette approximation, dont la validité a été confirmée par l'écoute : savoir qu'un trou de largeur iHz située en f_0 n'est pas une information nécessaire à la représentation d'un son. Par contre, savoir qu'il y a une probabilité certaine que le spectre comporte un trou de largeur approximative iHz est une donnée importante.

Ainsi, une haute précision fréquentielle n'est pas nécessaire. L'intérêt de l'analyse de la densité spectrale réside surtout dans le fait de pouvoir détecter des bandes de bruit plus ou moins denses, sans obligatoirement savoir où elles se situent avec exactitude. Pour cela les expériences montrent que la courbe donnée par l'algorithme est exploitable, et permet d'estimer l'évolution de la densité spectrale de bandes de bruit.

3.6.4 Expériences

Dans cette sous-section, nous présentons quelques résultats des expériences que nous avons effectuées pour montrer la précision et l'efficacité de la méthode que nous avons développée pour l'estimation de la densité spectrale de bandes de bruit. Nous nous sommes restreints lors de ces tests à des sons synthétiques, car il n'est pas possible de connaître à priori la densité spectrale d'un son naturel, et donc de comparer les résultats donnés. La seule autre possibilité pour mener des expérimentations à partir de sons naturels serait de procéder à un processus complet d'analyse et de synthèse en estimant la densité spectrale et en synthétisant un son d'après les valeurs analysées afin d'effectuer des tests psychoacoustiques de comparaison du son obtenu par rapport à l'original. Ce procédé est lourd et complexe, et nous n'avons pas pu le mettre en place pour l'instant. Nous proposons toutefois quelques exemples sur des sons naturels, sans pouvoir mesurer l'exactitude de nos résultats. Les tailles de fenêtres considérées sont $W_s = 1024$ échantillons, soit 23ms.

Bruits de densités spectrales constantes

Les premières expériences ont été effectuées à partir de bruits de durée $W_s = 1024$ échantillons, synthétisés par somme de sinusoïdes, dont les amplitudes sont fixes et constantes, dont les phases sont des variables aléatoires distribuées uniformément entre 0 et 2π, et dont les fréquences sont régulièrement réparties sur la bande des fréquences audibles (ici, l'intervalle entre 0 et $\frac{F_s}{2}$Hz). La différence constante entre les fréquences décide ici de la densité spectrale. La densité maximale est associée à un nombre de sinusoïdes de $N = \frac{W_s}{2}$. Nous avons d'abord testé notre méthode d'approximation sur un bruit d'une telle densité. Ensuite, nous avons effectué d'autres expériences, en ne préservant qu'une certaine proportion de sinusoïdes par rapport à cette référence. Par exemple, un rapport de $\frac{1}{4}$ implique un nombre de sinusoïdes de $N = \frac{W_s}{8}$, et une différence entre les fréquences quatre fois plus importante.

La figure 3.27 montre l'exemple de l'estimation de la densité spectrale d'un bruit formé d'un nombre de sinusoïdes de $N = \frac{W_s}{2}$, et celui avec quatre fois moins de sinusoïdes $N = \frac{W_s}{8}$. Cette expérience montre la qualité de l'approximation dans ces deux cas, puisque les courbes obtenues varient très légèrement autour des valeurs 1 et 4 (lignes hachurées). Ces variations font apparaître cependant les premières limites de notre méthode d'approximation. En effet, lors de la maximisation du nombre $\mathcal{N}$, la largeur ΔF_0 de la bande de fréquence est itérativement testée par des pas successifs de 0.25 casiers ($0.25\frac{F_s}{W_s}$), et les variations, donc les

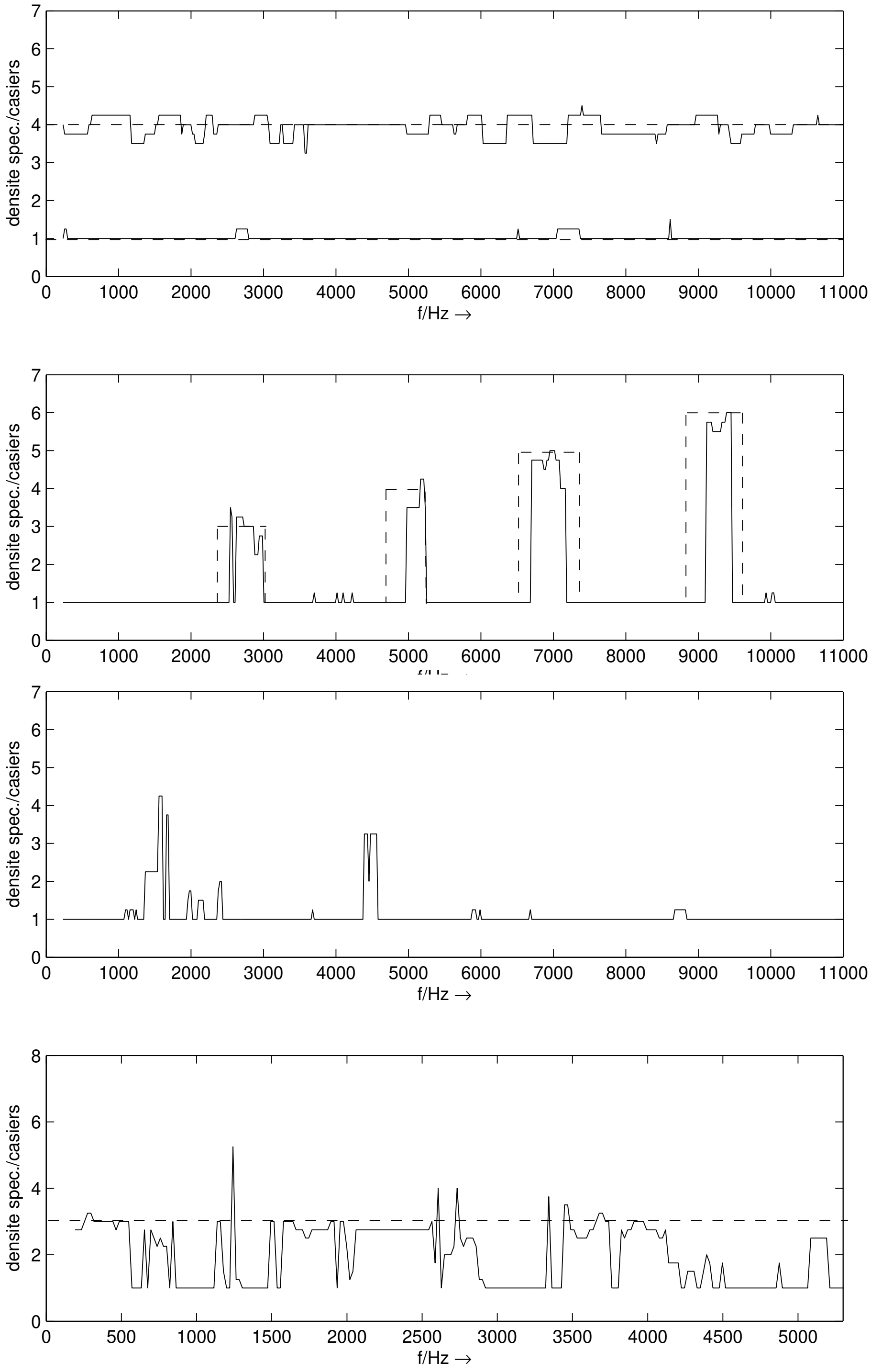

FIG. 3.27 – Illustration de l'approximation de la densité spectrale d'un bruit avec les exemples de bruit de densités spectrales fixées et constantes (figure du haut avec une densité maximale (1) et une 4 fois plus faible), un bruit composé de bandes de densité différentes $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$ et $\frac{1}{6}$ (deuxième figure), un bruit blanc synthétisé par tirage aléatoire des échantillons (troisième figure), et enfin un son naturel de rasoir électrique.

erreurs, sont parfois supérieures à ce pas.

Bruit blanc

D'autres expériences ont été effectuées à partir de bruit blanc gaussien, synthétisé par tirage aléatoire d'échantillons, selon une distribution de Gauss. Les bruits blancs peuvent, à priori, être considérés comme des bruits dont la densité spectrale est maximale et constante le long du spectre. Une première indication de résultat est donnée par le tableau 3.11, dans lequel le nombre moyen $\mathcal{N}$ de valeurs V_{NEP} situées sous un seuil t_h fixé ($\Delta F_0 = 172$Hz, $\Delta F' = \frac{F_s}{2}$Hz) est comparable au nombre moyen obtenu pour une densité spectrale maximale (taux de sinusoïdes de 1).

La figure 3.27 propose, sur sa troisième courbe, l'évolution de l'approximation de la densité spectrale obtenue par notre méthode. Nous devrions obtenir une courbe très similaire à celle d'un bruit de synthèse de densité maximale. Cependant, ici, nous observons quelques grandes variations très localisées. Nous avons donc été amené par ces résultats à considérer la possibilité qu'un tel bruit synthétisé par tirage aléatoire d'échantillons ne soit pas de densité constante. Pour vérifier cela, nous avons procédé à de brèves expériences de comparaison de bruits (le type de ces expériences est décrit dans la section 5.2.6). Elles ont confirmé dans une certaine mesure les résultats de notre estimation de densité, puisqu'elle nous permet non seulement de différentier un bruit blanc synthétisé par tirage aléatoire d'échantillons, d'un bruit blanc synthétisé par somme de $N = \frac{W_s}{2}$ sinusoïdes, mais ces expériences confirment que la densité n'est pas maximale pour le premier type de sons. Nous revenons sur cette observation dans la section 5.2.6.

Ainsi, les expériences psychoacoustiques, même si elles demandent à être confirmées, corroborent les résultats que montrent les courbes produites à l'aide de notre méthode d'estimation de la densité spectrale. Par ce fait, cette méthode s'avère donc à la fois utile et suffisamment précise, puisqu'elle permet par cette expérience de mettre en évidence une propriété perceptive du bruit étudié.

Bandes de bruit de densités spectrales différentes

Nous avons synthétisé, pour cette série d'expériences, des bandes de bruit dont la densité spectrale est maximale, à part pour une ou plusieurs bandes étroites, dont la densité spectrale est plus faible, la proportion de sinusoïdes utilisées pour la générer étant réduite. Le premier son synthétique étudié est un son contenant 4 bandes étroites, de largeurs respectives 860Hz, 947Hz, 947Hz et 860Hz. Pour chacune de ces bandes, les proportions de sinusoïdes sont respectivement $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$ et $\frac{1}{6}$. La figure 3.27 montre sur sa deuxième courbe l'évolution de la densité spectrale de ce son, estimée par l'application de la méthode que nous proposons, ainsi que la densité spectrale théorique (en lignes hachurées).

Nous pouvons observer sur ces résultats que la méthode, malgré certaines erreurs de précision, parvient à mettre en évidence les variations de densité spectrale, introduites par les quatre bandes de bruit. Elle permet d'approximer à peu près correctement la localisation fréquentielle des variations de densité spectrale (localisation de creux d'énergie) ainsi que leurs forces (amplitude des creux).

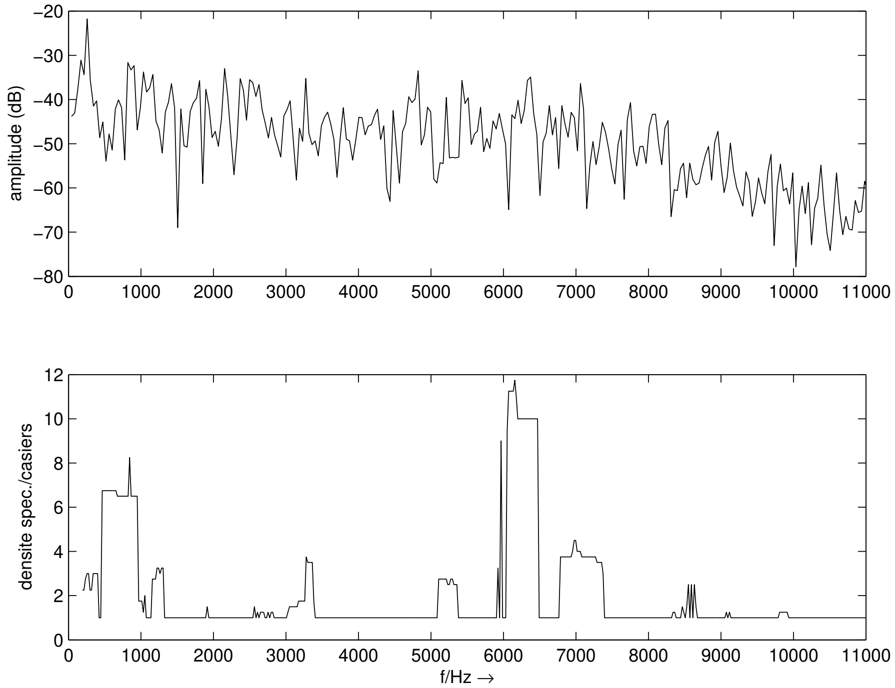

FIG. 3.28 – Illustration de l'approximation de la densité spectrale d'un bruit avec l'exemple d'un son de cymbale crash (fenêtre de 1024 échantillons, $H_0 = 0.125$, $\Delta F' = 344$Hz) : la densité spectrale n'est ni constante, ni maximale.

Sons naturels

Nous nous sommes également attachés à tester la méthode d'approximation de la densité spectrale sur des sons bruités naturels. Pour de tels sons, il est impossible de connaître la densité spectrale réelle, et donc d'effectuer une critique de la qualité de l'analyse. Le premier exemple concerne un son issu d'un rasoir électrique. La figure 3.27 montre sur sa quatrième courbe un exemple d'approximation de densité spectrale sur une fenêtre d'analyse (1024 échantillons). Nous remarquons que la densité est relativement faible par rapport au bruit blanc et se rapproche d'un son composé d'une sinusoïde sur 3 par rapport à la densité considérée comme maximale. Ces résultats apparaissent conforme à la perception : ce son se rapproche auditivement d'un son harmonique puisqu'il laisse percevoir une hauteur et indique donc par cette sensation une densité non maximale. Nous continuons l'étude de ce son dans la section 3.7.6.

Le deuxième exemple que nous présentons concerne un son issu de la frappe d'une cymbale crash. Nous ne considérons pas le tout début du signal puisqu'il est composé d'une transitoire qui pourrait fausser les résultats obtenus. La figure 3.28 montre l'approximation de la densité spectrale obtenue ainsi que le spectre associé d'un son de cymbale crash (1024 échantillons). Des variations de densité assez importantes peuvent être observées. Nous nous sommes demandés si elles sont dues à la présence de partiels, ce qui entraîne un pic d'amplitude dans la bande étudiée. Les analyses sinusoïdales que nous avons effectuées ne semblent toutefois pas confirmer cette possibilité.

3.6.5 Limites et conclusion

La méthode proposée reste encore expérimentale et sera très certainement améliorée prochainement, notamment en affinant les valeurs des paramètres d'analyse. Les résultats obtenus ne peuvent cependant pas être comparés car, à notre connaissance, il n'existe aucune technique permettant d'estimer le nombre de sinusoïdes nécessaire pour représenter un son bruité.

Il est également difficile de discuter de la précision de cette méthode, puisqu'il n'existe, toujours à notre connaissance, aucun travail concernant la résolution requise pour une telle méthode. Il faudrait pour cela posséder des informations plus précises sur la perception de bandes de bruit de densités spectrales différentes, les effets de masquage entre ces bandes, etc.... Néanmoins, les expériences que nous avons menées à partir de l'implémentation du modèle CNSS pour la synthèse de bruit (décrite dans la section 5.1) montrent que la précision imposée par la méthode que nous proposons semble suffisante pour analyser quelques propriétés relatives à la densité spectrale d'un son synthétique, et être ainsi capable de reproduire un son très ressemblant sur le plan de la perception. En effet, la différence entre deux valeurs de densité spectrale moyenne doit être importante pour que le système auditif perçoive cette différence. Les erreurs de précision de la mesure de la densité ne devraient donc avoir que rarement des répercussions sur la perception des sons synthétisés.

Cette technique d'estimation de la densité spectrale présente en l'état actuel un défaut majeur, que nous n'avons pas encore réussi à résoudre. Pour des densités spectrales comprises entre la valeur maximale et une proportion de sinusoïdes réduites de moitié, les erreurs d'analyse sont assez importantes. Il nous est donc difficile de différentier par cette méthode des sons de densité spectrale importante, mais non maximale. Toutefois, des expériences continuent d'être menées pour tenter d'améliorer cette précision, par un choix judicieux des paramètres d'analyse.

Influence des paramètres

Comme souvent, les paramètres jouent un rôle fondamental pour cette méthode d'analyse. Celui qui est apparu comme le plus important est la largeur $\Delta F'$ de la bande de bruit, considérée pour compter le nombre $\mathcal{N}$ de valeurs V_{NEP} inférieures à un seuil fixé. En effet, plus cette largeur est grande, plus important est le nombre de valeurs V_{NEP} utilisées pour calculer ce nombre $\mathcal{N}$, empêchant la détection de variations très localisées de densité. La conséquence directe d'un choix d'une largeur $\Delta F'$ trop grande est l'absence de détection de creux d'énergie trop étroits, et donc de densité spectrale faible. C'est ce qui explique par exemple, les difficultés d'approximation de bandes de bruit de densité deux fois moins importante que la densité maximale. A l'opposé, choisir une largeur trop faible impliquera une plus grande sensibilité aux changements d'énergie dans le spectre, entraînant une détection plus appuyée des variations de densité. Ce paramètre pose le problème du compromis de la précision de la mesure d'estimation $\mathcal{N}$ par rapport à la précision de la localisation fréquentielle.

Les autres paramètres ont une influence beaucoup moins importante. Le seuil t_h ne permet que de réduire ou d'augmenter le nombre de valeurs V_{NEP} considérées. Il faut juste s'assurer qu'il n'est ni trop grand, entraînant un nombre $\mathcal{N}$ important dont les variations seront moins significatives, ni trop petit, auquel cas ce nombre $\mathcal{N}$ risque d'être trop souvent nul, quelle que soit la largeur ΔF_0.

De toute façon, comme souvent pour les méthodes d'analyse, il est nécessaire d'effectuer de nombreux tests sur de grandes bases de données de sons, avant de parvenir à trouver les valeurs de paramètres impliquant les meilleurs résultats. Nous pensons que cette phase de test, toujours en cours, peut encore nous permettre d'affiner les résultats obtenus.

Conclusion et perspectives

La méthode que nous introduisons dans cette section est une méthode originale permettant d'estimer un paramètre perceptif de sons bruités mis en évidence par des études psychoacoustiques et intégré dans le modèle CNSS. Cette méthode ne peut être comparé ni à une méthode existante, ni aux méthodes d'analyse sinusoïdale, limitées lorsqu'elles sont appliquées à des sons bruités. Les quelques expériences que nous avons menées sur des sons synthétiques semblent indiquer que la précision donnée par la méthode est suffisante pour mettre en avant des propriétés du son, jusqu'ici inexploitées. C'est par exemple le cas d'un bruit blanc synthétisé par tirage aléatoire d'échantillons, dont la méthode nous a permis d'envisager la possibilité qu'il ne soit pas de densité spectrale maximale. La précision, même si elle apparaît rapidement limitée, peut s'avérer suffisante, puisque l'analyse de la densité spectrale ne requiert ni une importante précision sur la localisation fréquentielle, ni une grande précision temporelle. L'oreille n'est en effet pas sensible à ces informations sous leur forme exacte.

Peu d'expériences ont été effectuées avec des sons naturels car il est nécessaire pour mesurer la qualité de la méthode de connaître la densité spectrale du son étudié, ce qui est à priori impossible à savoir. De plus, les hypothèses imposées par la méthode introduite supposent que le son analysé ne doit contenir aucun partiel. Cette hypothèse est difficile à vérifier, car l'application des analyses liées aux modèles sinusoïdaux existants à des sons bruités mènent fréquemment à la définition artificielle de partiels, alors que le son n'en contient pas véritablement.

Cette méthode d'approximation de la densité spectrale peut donc être appliquée lors d'une méthode plus générale d'analyse et de synthèse de sons bruités, comme celle destinée au modèle CNSS. La densité spectrale est appréhendée pour ce modèle par une approche statistique. Les résultats obtenus par la technique que nous proposons nous permet d'estimer la densité spectrale moyenne de bandes de bruit successives. Dans ce cas, elle est donc d'autant plus précise que la bande étudiée est large. C'est donc l'approximation de la densité moyenne qui semble l'application possible la plus prometteuse et la plus précise.

Le modèle CNSS sépare la représentation spectrale d'un son étudié en plusieurs paramètres dont l'enveloppe spectrale et la densité spectrale. Un exemple de cette décomposition est donnée par la figure 3.29, qui présente le signal sous sa représentation temporelle et fréquentielle, et les deux paramètres extraits, c'est-à-dire l'enveloppe spectrale et la densité spectrale. Le son analysé est un bruit blanc de densité spectrale constante et maximale, excepté pour une bande de bruit de largeur 500Hz ($[3100; 3600]$Hz) pour laquelle seulement une sinusoïde sur quatre a été synthétisée. Cet exemple montre bien d'une part la qualité de la méthode d'estimation de la densité spectrale, d'autre part l'intérêt de cette décomposition en deux paramètres indépendants pour pouvoir effectuer des transformations.

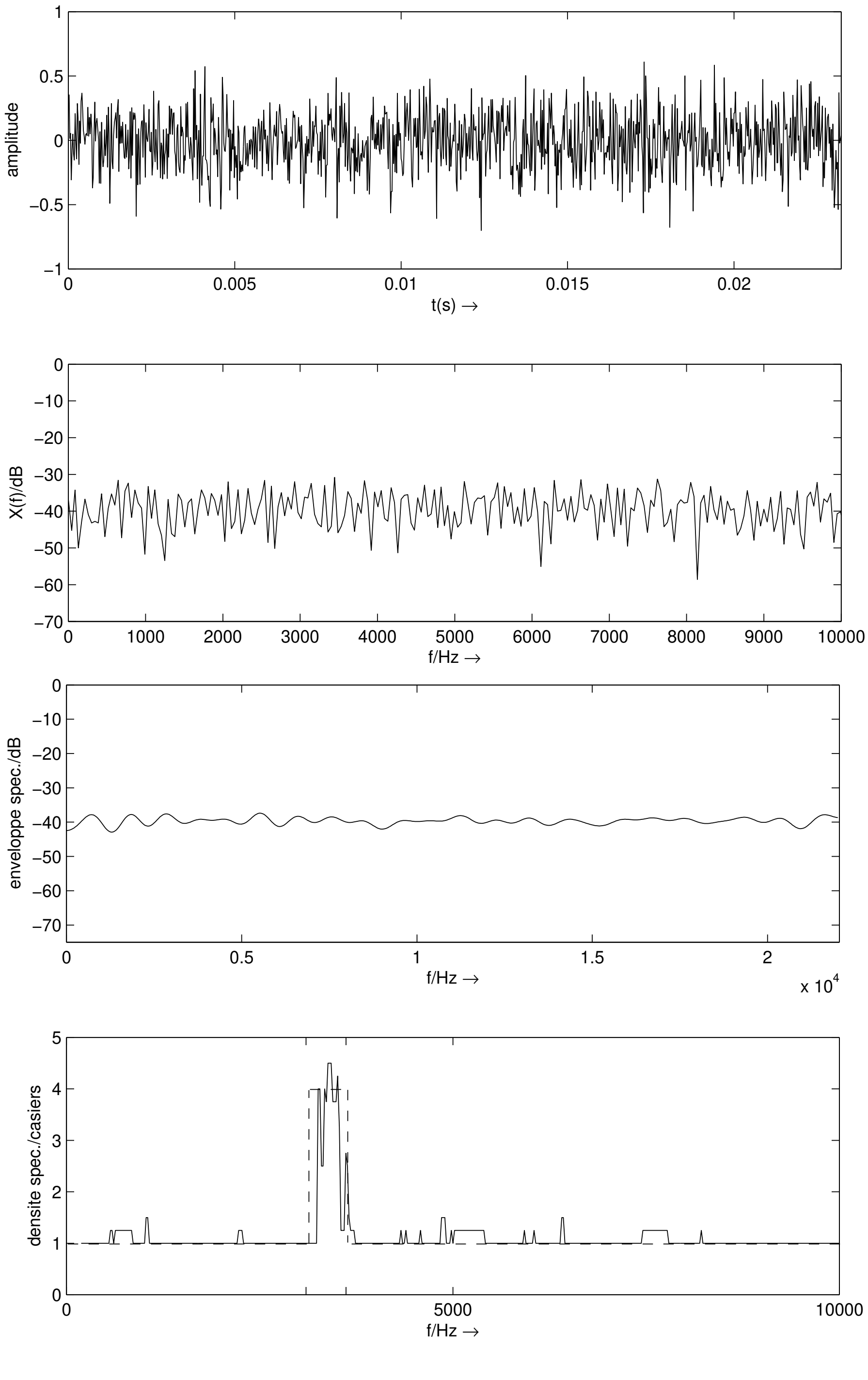

Fig. 3.29 – Illustration de l'approximation de la densité spectrale d'un bruit avec l'exemple d'un bruit blanc de densité spectrale maximale, excepté pour une bande de largeur 500Hz ([3100; 3600]Hz), pour laquelle elle est 4 fois moins importante (1 sinusoïde sur 4 utilisée à la synthèse). Les deux figures du haut montrent les représentations temporelles et spectrales, la troisième figure présente l'enveloppe spectrale lissée (couleur dans le cas du modèle CNSS) et la dernière figure montre l'estimation de la densité spectrale. Pour chacune, les traits hachurés indiquent la courbe théorique associée.

3.7 Extraction des paramètres du modèle CNSS

Le modèle CNSS présenté dans la section 2.4 représente la distribution des fréquences composant le spectre du signal étudié à l'aide de paramètres mathématiques : le nombre de casiers M, le nombre de sinusoïdes N et la largeur relative L de la fonction de densité de probabilité dans chaque casier. La méthode d'analyse associée à ce modèle doit donc donner les valeurs adéquates à ces paramètres à partir du signal original. Pour cela, nous avons vu dans la section précédente une méthode d'approximation de la densité spectrale d'une bande de bruit. La méthode d'analyse pour le modèle CNSS va donc consister à relier les résultats obtenus par cette méthode aux paramètres mathématiques liés à la distribution des fréquences.

3.7.1 Estimation des différences entre les fréquences

Les paramètres du modèle CNSS permettent de modéliser la distribution des fréquences des sinusoïdes représentant le signal. Ils ne peuvent donc être estimés qu'à partir de la fonction de densité de probabilité des fréquences. Il apparaît deux problèmes majeurs. Le premier repose sur la difficulté d'établir une fonction de densité de probabilité à partir d'un nombre réduit de signaux. En pratique, en supposant l'ergodicité du signal étudié, nous pouvons envisager une étude statistique à partir d'un certain nombre de fenêtres consécutives. Cependant, plus ce nombre est important, plus le signal doit être considéré comme stationnaire.

La seconde difficulté provient de l'analyse de la répartition des fréquences. Celle-ci, à cause des limites de résolution de la transformée de Fourier, est impossible. En fait, comme les études de la perception de la densité spectrale nous indiquent que, dans le cas de sons bruités, le système auditif n'est pas sensible à la localisation des fréquences, mais plutôt à l'existence de creux d'énergies dans le spectre, nous basons l'estimation des paramètres CNSS sur la répartition des différences entre les fréquences.

Pour approximer la fonction de densité de probabilité des différences entre les fréquences, la méthode décrite dans la section précédente peut être appliquée. En effet, dans les algorithmes que nous avons présentés, la valeur ΔF_0^{io} provoquant un nombre $\mathcal{N}$ maximum de valeurs V_{NEP} inférieures à un seuil fixé est supposée être la différence entre deux sinusoïdes voisines à l'intérieur de la bande étudiée, ou, pour le dire de façon différente, la taille du creux dans le spectre.

Nous notons q la fonction de densité de probabilité de la différence entre deux fréquences consécutives (ou de la taille des creux d'énergie). L'algorithme d'extraction des paramètres du modèle CNSS débute donc par la génération de la fonction cumulative des valeurs ΔF_0^{io} en fonction des fréquences correspondantes F. Cette fonction, même si elle n'est pas rigoureusement égale à la fonction de densité q, en présente toutefois les mêmes caractéristiques. L'étude théorique de la fonction de densité q nous a permis de dégager quelques propriétés définissant des liens avec les paramètres de contrôle de la répartition des fréquences du modèle CNSS.

En effet, nous ne pouvons définir ces paramètres que par la prise en compte de certaines caractéristiques des fonctions. D'une part, les méthodes d'estimation des différences entre les fréquences reposent sur quelques approximations, et nous avons vu dans la section précédente qu'elles ne sont pas très précises. D'autre part, l'étude théorique de la différence entre deux fréquences théoriques à partir des paramètres mathématiques est très complexe, et nous n'avons pas été en mesure d'arriver à formuler cette répartition dans le cas général. Cette complexité s'explique par la prise en compte du cas où le nombre de sinusoïdes et le nombre de casiers sont différents. Le tirage s'effectue alors sur deux niveaux : tout d'abord le tirage d'un certain nombre de casiers, ensuite le tirage de la fréquence dans le casier. La définition de la fonction de densité de probabilité des différences entre deux fréquences consécutives se ramène à un problème de mots aléatoires composés de M nombres (0 ou 1), dont N nombres 1. La question est de savoir trouver les probabilités de trouver des motifs 0^i, pour toutes les valeurs i, sachant qu'il est possible que le mot contienne plusieurs motifs. Dans le cas simple où le nombre de casiers et de sinusoïdes est le même, la fonction de densité de probabilité des différences de fréquences est définie par l'équation (2.52).

C'est donc pour ces deux principales raisons que nous nous avons décidé d'éviter cette étude précise pour nous restreindre aux quelques caractéristiques que nous avons exploitées et que nous allons détailler.

3.7.2 Principe général de l'analyse

Les différentes étapes de cette méthode d'extraction de paramètres sont décrites dans [HDC03b] et illustrées par la figure 3.30. La première partie consiste à estimer le nombre de casiers M car nous allons voir qu'il est directement relié au maximum de la fonction de densité de probabilité des différences de fréquences. La seconde partie effectue un test pour déterminer si le nombre de fréquences N est différent du nombre de casiers M. Si il est analysé comme différent, ce nombre est ensuite estimé. Enfin, la largeur relative de la fonction de densité de probabilité dans chaque casier est approximée indépendamment. Nous détaillons et justifions ces étapes successives.

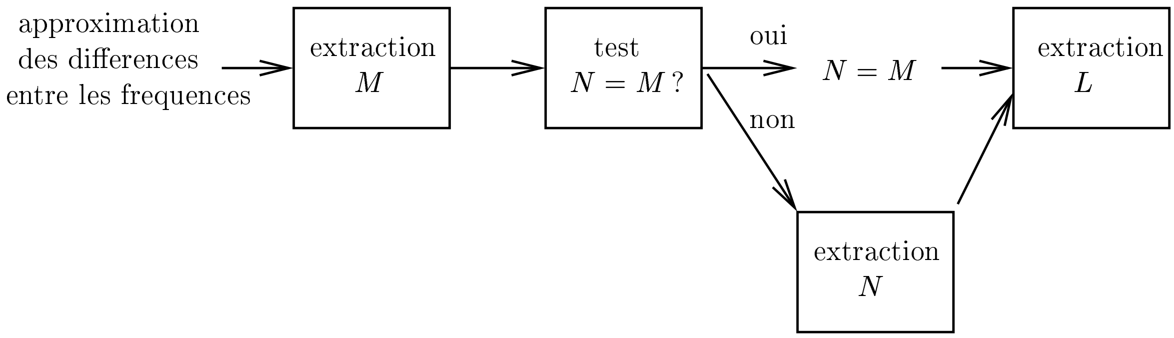

Fig. 3.30 – Principe général de la méthode d'extraction des paramètres du modèle CNSS.

3.7.3 Extraction du nombre de casiers

Le nombre de casiers, noté M est un paramètre mathématique du modèle CNSS, présenté dans la section 2.4.8. Il définit le nombre d'intervalles dans lesquels une valeur de fréquence au plus est choisie.

Principe

En étudiant l'influence du nombre de casiers sur la répartition des différences des fréquences, il apparaît un lien direct entre la fonction de densité de probabilité et ce nombre M. En effet, quelle que soit la valeur des autres paramètres, le maximum de la fonction de densité de probabilité, c'est-à-dire la valeur la plus probable, dépend directement du nombre de casiers M. La figure 2.24 illustre cette propriété en montrant la répartition expérimentale obtenue pour plusieurs valeurs de N. Pour chaque valeur de M, la différence de fréquence la plus probable correspond à $\frac{\Delta F}{M}$:

$$\delta q^{\max} = \frac{\Delta F}{M} \tag{3.57}$$

Le cas limite où le nombre de casiers M est infini vérifie également cette propriété. Dans ce cas, la différence des fréquences la plus probable est $\lim_{M\to\infty} \frac{\Delta F}{M} = 0$, avec $\Delta F \neq 0$.

En considérant une bande de bruit à analyser, l'algorithme d'estimation du nombre de casiers M nécessaire à la représentation par le modèle CNSS débute par un décompte du nombre d'occurrences de chaque valeur $\Delta F_0^{i_0}$ obtenues. Le paramètre M est approximé à partir de la valeur $\Delta F_0^{i_0}$ dont le nombre d'occurrences est le plus important :

$$M = \frac{\Delta F}{\Delta F_0^{\max}} \tag{3.58}$$

où $\Delta F_0^{\max}$ est exprimé en Hertz. Si les écarts entre les fréquences sont évalués en casiers, et notés $\Delta\mathsf{F}_0^{\max}$, le nombre de casiers M est obtenu par la relation :

$$M = \frac{\Delta F W_s}{F_s \Delta\mathsf{F}_0^{\max}} \tag{3.59}$$

Expériences

Nous avons testé l'algorithme d'estimation du nombre de casiers nécessaire à la représentation de bandes de bruit sur des sons synthétiques générés en sommant un nombre fixe de sinusoïdes, de fréquences déterminées à partir des paramètres N et L précis, dans une bande de largeur $\Delta F = \frac{F_s}{2}$, de phases uniformément distribuées entre 0 et 2π, et d'amplitude constante fixée.

Le tableau 3.12 présente quelques résultats de précision obtenus pour plusieurs valeurs des paramètres N et L. A chaque fois, le pourcentage de résultats justes est indiqué, ainsi que le pourcentage de résultats variant d'un certain nombre de pas. Une unité de pas, noté δf, est donc définie par :

$$\forall i, \quad \delta = \Delta F_0^i - \Delta F_0^{i+1} \tag{3.60}$$

Dans nos expériences, le pas δ de largeur de bandes ΔF_0 est de 0.25 casiers (soit $\delta = \frac{F_s}{4W_s}$). Ainsi, 10% d'erreur pour la colonne 1 précise que 10% des résultats ont été analysés avec un pas d'erreur, soit une estimation de la différence des fréquences la plus probable à un quart de casier de précision (soit pour une taille de fenêtre de 1024, de 10.75Hz).

Nous n'avons pas pris en compte les valeurs $\Delta F_0^{\max}$ correspondant à une densité spectrale maximale (différence entre les fréquence d'un casier), car elle représente très souvent (et donc fréquemment par erreur) le nombre maximum d'occurrences. Les expériences que nous avons menées avec des bandes de bruit confirment la validité de ce choix. En effet, l'estimation du nombre de casiers M pour un son de densité spectrale maximale conduit à un résultat correspondant à la valeur $\Delta F_0^{i_0}$ égale à un casier additionné au pas δ, ce qui correspond à l'approximation la plus proche de la réalité. La seule limite de ce choix concerne les sons de densités spectrales très importantes et très proches du maximum, qui peuvent alors être assimilées à cette valeur maximale. Néanmoins, pour un pas δ assez faible, la conséquence sur le plan de la perception est supposée négligeable.

Paramètres CNSS			Erreur δ					
N	M	L	0	1	2	3	4	> 4
128	128	0.0	77.78	22.22	0	0	0	0
128	128	1.0	12.22	31.11	27.78	17.78	8.89	2.22
146	146	0.0	91.11	8.89	0	0	0	0
146	146	1.0	21.11	27.78	18.89	17.78	10.00	4.44
171	171	0.0	52.22	47.78	0	0	0	0
171	171	1.0	20.00	35.56	25.56	12.22	2.22	4.44
128	146	0.0	87.78	12.22	0	0	0	0
128	146	1.0	21.11	27.78	18.89	17.78	10.00	4.44

TAB. 3.12 – Expérimentations de la méthode d'estimation du nombre de casiers pour le modèle CNSS (90 réalisations, $W_s = 1024$ échantillons, $\Delta F = \frac{F_s}{2}$Hz, $\Delta F' = 344$Hz). L'erreur est précisée en pas δ, paramètre de la méthode (ici $\delta = \frac{F_s}{4W_s}$).

Limites

La précision de la méthode est décidée par le choix du pas δ pour le calcul des valeurs ΔF_0^i. Pour les expériences avec des sons synthétiques, nous avons choisi $\delta = \frac{F_s}{4W_s}$ et $W_s = 1024$ échantillons. Quelles que soient les valeurs du nombre de sinusoïdes N et de la largeur relative L de la répartition des fréquences, la méthode permet d'approximer la valeur de M utilisée pour générer le son testé. Le tableau 3.12 montre que, pour des valeurs de L nulles, les erreurs sont limitées à la valeur du pas δ, même pour des nombres M de casiers supérieurs au nombre N de sinusoïdes. Bien évidemment, les erreurs sont d'autant plus nombreuses lorsque la largeur relative L est grande, puisque les fréquences sont plus fortement distribuées. Ainsi, l'écart le plus souvent obtenu n'est pas forcément celui correspondant à l'écart le plus probable, étant donné que nous ne considérons qu'une seule fenêtre de son, c'est-à-dire une seule réalisation du signal aléatoire.

Des erreurs de précision apparaissent surtout lorsque la densité étudiée est proche de la densité maximale, à cause des limites de la méthode d'approximation des différences entre les fréquences des sinusoïdes représentant le signal. Ces limites sont décrites dans la section 3.6.

3.7.4 Test d'égalité entre le nombre de sinusoïdes et le nombre de casiers

L'étape suivante de la méthode d'extraction des paramètres du modèle CNSS consiste à estimer le nombre de sinusoïdes N composant la bande de bruit étudiée. L'étude théorique montre certaines propriétés particulières dans le cas où le nombre de casiers M et le nombre de sinusoïdes N sont égaux. Nous avons donc décidé de prendre en considération cette possibilité par un algorithme de test relativement simple, de façon à nous éviter d'effectuer éventuellement une étape supplémentaire.

Principe

Cette méthode de test prend en compte l'estimation du nombre de casiers M donnée par l'algorithme précédent. Lorsque le nombre de sinusoïdes N est égal au nombre de casiers M, la fonction cumulative des différences de fréquence est théoriquement symétrique, quelle que soit la valeur de la largeur relative L. L'axe central de cette fonction symétrique est défini par la valeur la plus probable, c'est-à-dire $\frac{\Delta F}{M}$. La figure 3.31 illustre cette propriété, dans les cas où $N = M$ et $N < M$.

L'algorithme que nous proposons pour déterminer si le nombre de fréquences N et le nombre de casiers M sont les mêmes effectue un test sur la valeur maximale ΔF_0^i qui est non nulle. Nous notons cette valeur ΔF_0^m . Si le nombre de sinusoïdes N et le nombre de casiers M sont égaux, cette valeur est, dans le cas limite où $L = 1$, approximativement deux fois la valeur $\frac{\Delta F}{M}$ (voir la section 2.5). Dans le cas contraire, le nombre de fréquences N est plus petit que le nombre de casiers M. Cette condition peut donc s'écrire :

$$\Delta F_0^m > 2\frac{\Delta F}{M} \tag{3.61}$$

Il est toutefois nécessaire de tenir compte des approximations effectuées dans le calcul des différences entre les fréquences, et donc sur la fonction cumulative de ces différences. Pour cela, l'algorithme va compter le nombre de valeurs ΔF_0^i supérieure à $2\frac{\Delta F}{M}$. Si ce nombre est inférieur à un seuil préalablement fixé et noté $\mathcal{N}^{\text{seuil}}$, indépendant du nombre de casier M, les nombres de casiers et de sinusoïdes sont supposés égaux :

$$\mathcal{N}_{(\Delta F_0^i > 2\frac{\Delta F}{M})} > \mathcal{N}^{\text{seuil}} \tag{3.62}$$

où $\mathcal{N}_{(\Delta F_0^i > 2\frac{\Delta F}{M})}$ représente le nombre de valeurs de ΔF_0^i supérieures à $2\frac{\Delta F}{M}$.

Expériences

L'algorithme pour détecter une égalité entre le nombre de sinusoïdes N et le nombre de casiers M suppose la connaissance préalable de ce dernier nombre. Nous avons donc effectué nos expériences selon deux procédés différents. Dans les premiers tests, nous avons effectué les comparaisons exprimées par l'équation (3.61) à partir du nombre de casiers M expérimental, estimé par l'algorithme présenté précédemment. Évidemment cet algorithme est une source d'erreurs qui risquent de se répercuter sur la précision de l'algorithme d'estimation de l'égalité entre N et M. C'est pourquoi nous avons également choisi de procéder aussi à partir de la valeur théorique du nombre de casiers, pour montrer le lien entre les éventuelles erreurs et les erreurs d'estimation de M. Pour ces deux méthodes, un seuil différent a été choisi.

Nous avons donc effectué plusieurs expériences que nous récapitulons dans le tableau 3.14, qui présente le taux de réussite des tests, pour différents sons synthétiques analysés, selon les deux procédés expliqués précédemment (M expérimental avec un seuil de $\mathcal{N}^{\text{seuil}} = 55$ et M théorique avec un seuil de $\mathcal{N}^{\text{seuil}} = 35$).

Ces résultats indiquent que l'algorithme donne des résultats excellents quelles que soient les caractéristiques des sons analysés, à partir du moment où le nombre de casiers M est juste (cas du M théorique). Lorsque ce nombre n'est pas exact (cas du M expérimental), des erreurs apparaissent, surtout pour les expériences où les nombres de sinusoïdes N et de casiers M sont égaux. Par ailleurs, le paramètre a une certaine influence puisque les résultats semblent d'autant meilleurs que ce paramètre est proche de 0. Cette observation s'explique par le fait que plus le paramètre L est proche de 1, plus les différences entre les fréquences sont distribuées, et donc induisent de plus fortes variations statistiques.

Limites

Comme les résultats le montrent clairement, la précision de cette méthode dépend essentiellement du nombre de casiers M employé par l'algorithme. Si celui-ci est juste, l'estimation est très bonne. Dans le cas contraire, même si elle reste très correcte, elle est moins précise. Nous pouvons toutefois indiquer que les valeurs des seuils $\mathcal{N}^{\text{seuil}}$ utilisés doivent être affinés en menant de nombreuses autres expériences. De plus, les méthodes d'estimation du nombre de sinusoïdes N, appliquées si l'égalité avec le nombre de casiers M n'est pas mise en évidence, pourra éventuellement indiquer une valeur proche de celle du nombre de casiers et ainsi rattraper l'erreur introduite par cet algorithme.

3.7.5 Estimation du nombre de sinusoïdes

Si l'application de l'algorithme précédent conduit à estimer que le nombre de sinusoïdes N est inférieur au nombre de casiers M, il est alors nécessaire d'approximer ce nombre.

Principe

Le nombre de sinusoïdes N composant le son analysé n'a pas une influence sur la différence la plus probable entre les fréquences mais en a une sur le nombre de différences importantes pouvant être détectées. En effet, quel que soit le nombre de casiers, plus le nombre de sinusoïdes est faible, plus le nombre de larges différences est important. Cette propriété est illustrée par la figure 3.31 qui montre deux fonctions de densité de probabilité des différences des fréquences pour un même nombre de casiers, mais un nombre de sinusoïdes plus importants ($N = 171$) pour la courbe en traits hachurés que pour la courbe la plus à gauche ($N = 128$).

Cette remarque nous permet d'envisager plusieurs méthodes d'estimation du nombre de sinusoïdes, basées sur le nombre de différences supposées importantes. La première possibilité est de compter les différences ΔF_0^i supérieures à $\Delta F_0^{\text{ref}} = 2\frac{\Delta F}{M}$, sur le même principe que l'algorithme précédent :

$$\mathcal{N}^{\text{mesuré}} = \mathcal{N}_{(\Delta F_0^i > \Delta F_0^{\text{ref}})} \tag{3.63}$$

Le nombre ainsi obtenu est caractéristique du nombre de différences importantes entre les fréquences, et donc du nombre de sinusoïdes N.

Cet algorithme calcule une valeur, qui est d'autant plus grande que le nombre de sinusoïdes N est petit. En effet, plus le nombre de sinusoïdes est faible, plus d'importantes différences entre les fréquences sont détectées. La valeur calculée $\mathcal{N}^{\text{mesuré}}$ n'est pas normalisée. Pour pouvoir estimer ce nombre, il est donc nécessaire d'établir des valeurs de référence pour pouvoir

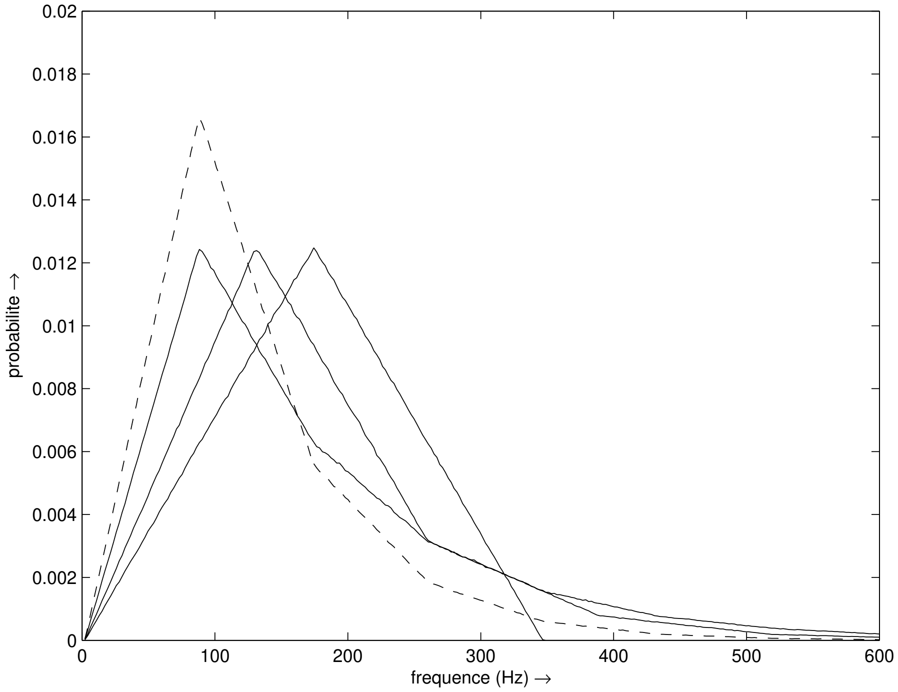

FIG. 3.31 – Illustration expérimentale ($\Delta F = \frac{F_s}{2}$Hz) de la fonction de densité de probabilité de la différence entre deux fréquences consécutives dans les cas où $N = 128$ et $M = 128$ (à droite), $M = 171$ (au centre) and $M = 256$ (à gauche), et le cas où $N = 171$ et $M = 256$ (en traits hachurés). La maximum correspond à $dF_{\max} = \frac{\Delta F}{M}$ (respectivement 2, 3 et 4 casiers, c'est-à-dire 86Hz, 129Hz et 172Hz), mais seulement la courbe correspondant au cas où $N = M$ est symétrique.

Paramètres CNSS			M expérimental		M théorique	
N	M	L	$N = M$	$N \neq M$	$N = M$	$N \neq M$
171	171	1.0	52.22	47.78	100.00	0.00
146	171	1.0	7.78	92.00	0.00	100.00
128	171	1.0	1.11	98.89	0.00	100.00
128	146	1.0	26.67	73.33	0.00	100.00
128	146	0.0	0.00	100.00	0.00	100.00
146	146	1.0	54.44	45.56	100.00	0.00
146	146	0.0	100.00	0.00	100.00	0.00

TAB. 3.13 – Résultats des expériences concernant l'algorithme de détection de l'égalité entre le nombre de casiers M et le nombre de sinusoïdes N pour différents sons synthétiques (moyennes sur 90 signaux, de taille $W_s = 1024$ échantillons, $\Delta F = \frac{F_s}{2}$Hz, $\Delta F' = 172$Hz).

comparer les résultats obtenus avec ces valeurs calculées à partir de sons de synthèse dont le nombre de sinusoïdes est connu. Cette méthode requiert donc une phase de calibration.

Expériences

Comme l'algorithme de test d'égalité précédent, l'algorithme d'estimation du nombre de sinusoïdes N prend en compte le nombre de casiers M. Les éventuelles erreurs d'estimation de M peuvent donc se répercuter sur la précision de cet algorithme. Nous avons procédé à des expériences, à partir du nombre estimé par l'algorithme présenté précédemment (M expérimental), mais également à partir du nombre de casiers M théorique. Les tableau 3.14 montre quelques exemples de résultats $\mathcal{N}^{\text{mesuré}}$, obtenus à partir de sons synthétiques. Ces résultats confirment que $\mathcal{N}^{\text{mesuré}}$ dépend essentiellement du nombre de sinusoïdes. Cette observation est vérifiée pour les résultats moyens, même pour des valeurs de N proches (128 et 146).

Paramètres CNSS			M expérimental			M théorique		
N	M	L	moyenne	max	min	moyenne	max	min
128	146	1.0	101.6	389.0	0.0	108.8	172	66
146	146	1.0	50.76	286.0	0.0	19.18	45	3
146	146	0.0	10.5	25.0	0.0	10.5	25.0	0.0
128	171	1.0	199.2	707	37	235.0	305	177
146	171	1.0	135.0	340.0	18.0	131.3	191.0	85.0
171	171	1.0	81.4	537.0	0.0	19.6	48.0	2.0

TAB. 3.14 – Exemples de nombres de différences ΔF_0^i supérieures à une valeur de référence ($\Delta F_0^{\text{ref}} = 2\frac{\Delta F}{M}$), obtenus pour plusieurs nombres de sinusoïdes N et de casiers M : pour des nombre de casiers équivalents, nous remarquons qu'ils sont caractéristiques des nombres de sinusoïdes N (les moyennes, minimum et maximum ont été calculés à partir de 50 signaux aléatoires de taille $W_s = 1024$ échantillons, $\Delta F = \frac{F_s}{2}$Hz).

Limites

Le tableau 3.14 montre quelques limites, puisqu'il présente des valeurs $\mathcal{N}^{\text{mesuré}}$ mesurées qui peuvent indiquer un nombre de sinusoïdes plus important ou plus faible qu'il ne l'est véritablement. C'est notamment le cas de certaines valeurs maximales associées à un certain nombre de sinusoïdes (par exemple $N = 146$ et $M = 171$) qui dépassent certaines valeurs minimales relatives à un nombre de sinusoïdes inférieur ($N = 128$ et $M = 171$). Il existe certaines probabilités que le nombre de sinusoïdes soient alors mal estimé.

Ainsi, cet algorithme permet de calculer des valeurs liées au nombre de sinusoïdes. Cependant, le calcul direct des nombres de sinusoïdes est pour l'instant impossible. Il est donc nécessaire de procéder à des comparaison avec des valeurs établies à partir de sons synthétiques. Cette contrainte rend cet algorithme plus complexe que ceux présentés dans cette section. D'autres algorithmes peuvent certainement être mis au point pour éviter ces comparaisons, et seul le temps nous a manqué pour y parvenir.

3.7.6 Analyse de l'harmonicité

L'algorithme général d'extraction des paramètres du modèle CNSS doit également effectuer une estimation de la largeur relative L de la fonction de densité de probabilité uniforme associée à chaque casier. Cette étape peut être effectuée avant ou après les précédentes puisqu'elle est complètement indépendante et ne requiert aucune information concernant les autres paramètres du modèle.

Principe

Le paramètre L du modèle CNSS décrivant la largeur relative de la fonction de densité de probabilité des fréquences dans chaque casier est lié à l'harmonicité du son et donc à sa périodicité. En effet, comme l'illustre la figure 3.32, une faible valeur de L (proche de 0) induit une différence à peu près constante entre les fréquences consécutives, alors qu'une valeur de L plus large (proche de 1) implique une distribution plus importante de ces différences. Comme nous l'avons vu dans la section 2.5, ce paramètre peut être présenté comme un coefficient d'harmonicité, puisque la valeur $p_{\max}$ la plus probable des écarts entre deux fréquences successives est inversement proportionnel à ce paramètre :

$$\forall L \in]0;1] \quad q_{\max} = q(\frac{\Delta F}{M}) = \frac{M}{L\Delta F} \tag{3.64}$$

La méthode d'estimation de ce paramètre peut donc se ramener à un problème classique de mesure de la force d'une hauteur perçue pour un son harmonique ou bruité. Nous pouvons citer plusieurs travaux sur des modèles psychoacoustiques sur ce domaine comme par exemple les études de Demany [KD98] ou ceux de Yost [Yos96] et de Wiegrebe [WPDC98] pour les bruits ondulants. Ces recherches proposent des méthodes de mesure basées sur la fonction d'autocorrélation, définie dans la section 1.1.3, et plus particulièrement le rapport entre le second pic maximum de cette fonction et sa première valeur, qui représente l'énergie totale du signal et le maximum de la fonction. Cette technique est également appliquée dans le traitement de la parole, notamment pour discriminer les parties voisées et non voisées.

Expériences

L'algorithme d'estimation consiste à calculer la fonction d'autocorrélation du signal, notée Γ, d'après l'équation (1.7). Sachant que la première valeur $\Gamma(0)$ de cette fonction est le maximum, nous cherchons le pic secondaire $\Gamma(m)$, qui vérifie :

$$\begin{cases} \forall x \neq 0, \quad \Gamma(m) & > \quad \Gamma(x) \\ \Gamma(m) \geq \Gamma(m-1) & \text{et} \quad \Gamma(m) \geq \Gamma(m+1) \end{cases} \tag{3.65}$$

Le rapport, notée r_L, entre la première valeur de la fonction et ce maximum secondaire est alors déduit :

$$r_L = \frac{\Gamma(m)}{\Gamma(0)} \tag{3.66}$$

La valeur obtenue est ensuite comparée à une courbe d'étalonnage expérimentale, comme par exemple les valeurs proposées dans le tableau 3.15, pour permettre d'en déduire la valeur du paramètre L associée. De plus, il est important de remarquer que la valeur m correspondant au maximum secondaire de la fonction d'autocorrélation est théoriquement liée à l'écart (plus ou moins régulier) entre les fréquences :

$$m = \frac{M}{F_s \Delta F} \tag{3.67}$$

où m est exprimé en échantillons. Cette valeur peut donc être également utile pour vérifier la valeur du nombre de casiers M estimée par un autre algorithme, surtout dans le cas d'une harmonicité forte (L proche de 0).

Les expériences que nous avons effectuées vérifient la précision de l'algorithme proposé, mais servent également de valeurs de référence pour déduire des rapports calculés les valeurs du paramètre L. Deux exemples sont répertoriés dans le tableau 3.15 et présentent les variations statistiques des mesures pour différentes valeurs du paramètre L et pour deux sons synthétiques de durée 1024 échantillons, générés à partir de 128 et 171 sinusoïdes ($N = M$). Plusieurs conclusions peuvent être tirées de ces expériences. Tout d'abord, le nombre de sinusoïdes n'influe quasiment pas sur les valeurs des rapports d'autocorrélation associées aux valeurs de L, et donc sur la précision de l'estimation. Ainsi, une seule courbe d'étalonnage est suffisante pour déduire une approximation de la valeur de L à partir de la valeur calculée du rapport r_L. De plus, ils montrent que les écarts-types de la mesure sont maximaux pour des valeurs de L proches de 0.5. Pour des valeurs extrêmes, proches de 0 ou de 1, la précision est plus importante.

Limites

Les limites de l'algorithme d'estimation du paramètre L du modèle CNSS sont tout d'abord liées aux paramètres employés pour le calcul de la fonction d'autocorrélation. Il est nécessaire de s'assurer qu'ils permettent de prendre en compte les périodicités éventuelles du signal (voir la section 1.1.3). D'autre part, cet algorithme nécessite le précalcul de valeurs de référence pour pouvoir déduire de la valeur du rapport d'autocorrélation la valeur du paramètre L. Le tableau montre que cette étape peut être effectuée préalablement, puisqu'elle ne requiert pas la connaissance du nombre de sinusoïdes.

La précision indiquée par les écart-types paraît suffisante au regard de la précision du système auditif. En effet, les expériences de synthèse que nous avons effectuées mettent en évidence notre sensibilité par rapport au paramètre L, mais la résolution requise ne paraît pas être très importante. En effet, il nous a semblé généralement difficile de différentier une bande de bruit produite avec un paramètre L de 0.6 de celle produite avec un paramètre L de 0.5.

Nous avons effectué également des tests à partir de sons fortement bruités naturels. Parmi les sons qui indiquent une grande harmonicité, les sons de machine sont particulièrement intéressants pour mettre en valeur le processus d'analyse du paramètre L. En effet, ils sont suffisamment bruités pour être difficilement analysés par des méthodes sinusoïdales, mais laissent percevoir une hauteur. La figure 3.34 présenté les résultats de l'algorithme d'estimation de l'harmonicité de ce son. De nombreuses valeurs de rapport r_L sont situées entre 0.8 et 1.0, ce qui indique des valeurs de L proche de 0, et confirme l'harmonicité du son. Par ailleurs nous indiquons sur cette figure les valeurs des périodes associées, proches de 340 échantillons, c'est-à-dire relatives à une hauteur perçue d'approximativement 120Hz. Cette périodicité permet également d'estimer le nombre de casiers M, qui s'approche ici de 180, et donc de confirmer les nombres de casiers M estimés par l'algorithme présenté précédemment.

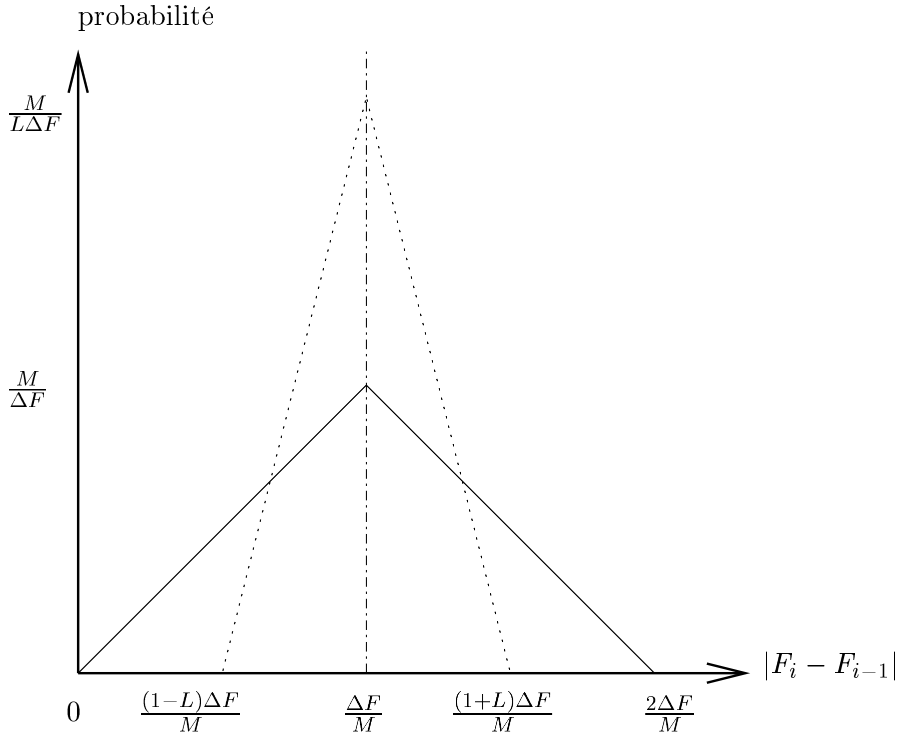

FIG. 3.32 – Représentation de la fonction de densité de probabilité de différence entre deux fréquences consécutives pour des valeurs différentes du paramètre L.

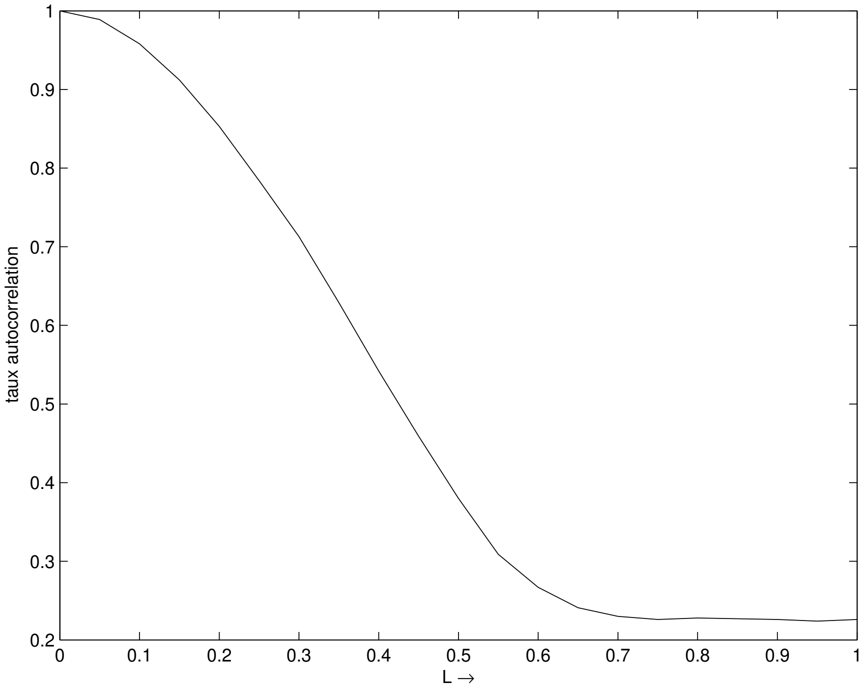

FIG. 3.33 – Variations mesurées sur un son synthétique du rapport entre le second pic maximum et la première valeur de la fonction d'autocorrélation par rapport à la valeur du paramètre L ($N = M = 128$).

	$N = 128$				$N = 171$			
L	moyenne	écart-type	max	min	moyenne	écart-type	max	min
0.0	1.0	0.0	1.0	1.0	0.999	0.004	1.009	0.975
0.05	0.989	0.009	1.017	0.960	0.988	0.010	1.016	0.946
0.1	0.958	0.018	1.012	0.889	0.960	0.019	1.015	0.880
0.15	0.912	0.024	0.987	0.824	0.914	0.027	0.990	0.828
0.2	0.853	0.032	0.956	0.755	0.853	0.034	0.953	0.698
0.25	0.784	0.040	0.901	0.644	0.783	0.043	0.916	0.664
0.3	0.713	0.046	0.874	0.548	0.707	0.048	0.853	0.548
0.35	0.629	0.054	0.814	0.448	0.625	0.056	0.808	0.447
0.4	0.542	0.061	0.735	0.314	0.546	0.061	0.777	0.331
0.45	0.459	0.069	0.668	0.260	0.461	0.069	0.673	0.241
0.5	0.380	0.070	0.584	0.188	0.382	0.070	0.613	0.204
0.55	0.309	0.063	0.514	0.182	0.310	0.064	0.549	0.180
0.6	0.267	0.051	0.485	0.161	0.262	0.050	0.475	0.171
0.65	0.241	0.039	0.441	0.150	0.237	0.035	0.377	0.151
0.7	0.230	0.031	0.381	0.148	0.230	0.032	0.366	0.165
0.75	0.226	0.029	0.375	0.161	0.226	0.030	0.355	0.157
0.8	0.228	0.028	0.363	0.167	0.227	0.030	0.350	0.157
0.85	0.227	0.030	0.350	0.162	0.226	0.030	0.392	0.146
0.9	0.226	0.028	0.330	0.165	0.225	0.029	0.359	0.158
0.95	0.224	0.029	0.330	0.161	0.225	0.029	0.343	0.155
1.0	0.226	0.030	0.346	0.155	0.224	0.028	0.350	0.159

TAB. 3.15 – Résultats moyens des rapports r_L d'autocorrélation pour des signaux synthétiques ($N = 128$, 1000 signaux de durée $W_s = 1024$ échantillons) en fonction de différentes valeurs pour le paramètre L.

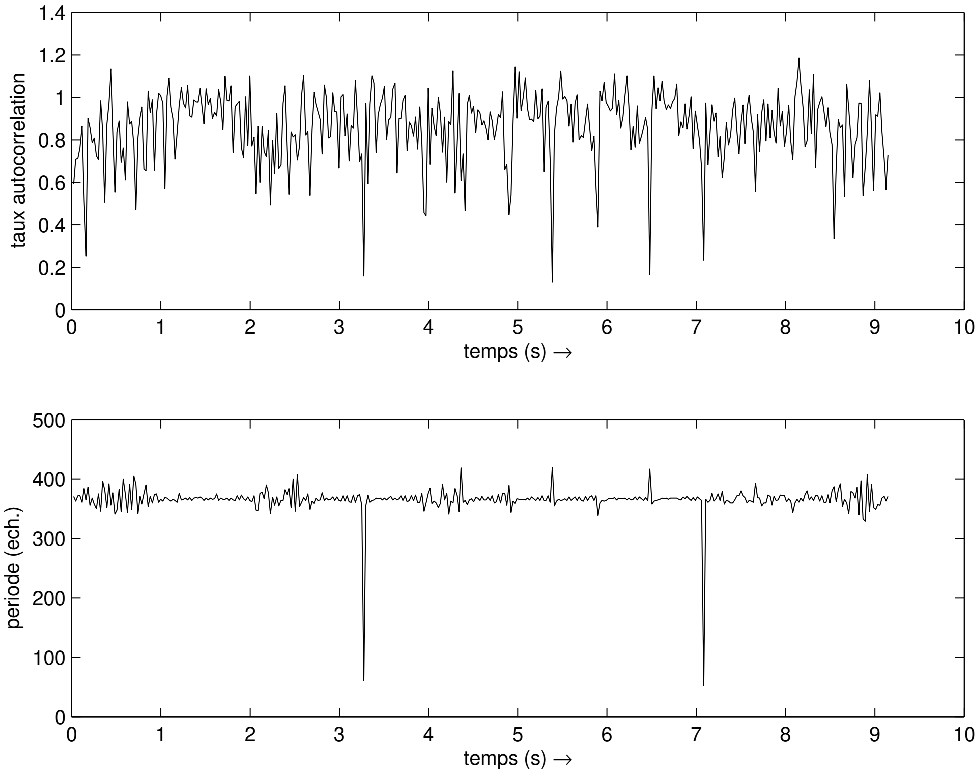

Fig. 3.34 – Exemple de l'estimation du paramètre L du modèle CNSS par la mesure du rapport entre la première valeur de la fonction d'autocorrélation et son pic secondaire, pour un son de rasoir électrique (1024 échantillons, taille de la fenêtre d'autocorrélation $k = 450$ échantillons) : la courbe des variations de ce rapport indique que l'harmonicité est importante pour ce son, et que la périodicité associée est relativement constante.

3.7.7 Discussion

Les limites des algorithmes d'estimation de paramètres sont nombreuses. Par exemple, les différences trop faibles entre les fréquences sont encore mal évaluées, ce qui implique des erreurs sur l'approximation des paramètres N et M, si ils sont trop proches de leur valeur maximale. De plus, les études que nous proposons concernent des bandes de bruit de largeurs importantes ($\frac{F_s}{2}$ par exemple). Nous supposons qu'une étude sur des bandes étroites risque de provoquer des erreurs plus importantes.

Toutefois, il est encore difficile d'évaluer ces méthodes d'analyse, puisque la résolution des paramètres du modèle CNSS n'a pas encore été établie avec précision. Pour cela il est nécessaire d'effectuer de nombreuses expériences psychoacoustiques. Néanmoins, les expériences de synthèse que nous avons menées ont semblé montrer qu'une résolution importante n'était pas nécessaire, car le modèle CNSS représente le signal par des paramètres moyens, et que le système auditif est surtout sensible à ces valeurs moyennes.

Par ailleurs, le fait d'étudier des sons supposés aléatoires implique forcément une étude statistique, et donc des calculs de moyenne reposant sur des hypothèses de stationnarité. Ces calculs induisent donc inévitablement des erreurs s'ils ne sont pas effectués sur un nombre important de signaux (ou de fenêtres temporelles de signaux). C'est pourquoi les résultats que nous montrons sont encourageants et prometteurs, car l'analyse de sons doit être effectuée sur plusieurs fenêtres d'analyse, ce qui laisse supposer une précision des estimations supérieures à celles présentées ici.

Les algorithmes d'estimation des paramètres du modèle CNSS sont encore expérimentaux, et nécessitent encore de nombreux tests, à la fois sur des sons de synthèse pour tenter d'améliorer la précision par des choix plus judicieux des paramètres des méthodes d'analyse, mais aussi sur des sons naturels pour montrer l'intérêt du modèle et de ces méthodes d'analyse. Nous pensons que les améliorations qui découleront de ces tests permettront d'améliorer encore les algorithmes que nous avons présentés.

Chapitre 4

Synthèse de sons bruités

La synthèse consiste à calculer les échantillons représentant un son dans le domaine temporel à partir de paramètres mathématiques d'un autre modèle, afin de pouvoir être convertis par une carte son en signal électrique analogique (convertisseur numérique/analogique). La synthèse est une étape du processus de traitement des sons qui, si elle est effectuée en temps réel, permet aux utilisateurs (compositeurs, musiciens, ...) d'entendre directement les modifications effectuées sur le son produit. Les méthodes de synthèse sont donc décrites sous la forme d'algorithmes de calcul des échantillons, et sont comparées non seulement sur le plan de la qualité du son généré, mais aussi sur le plan de l'efficacité.

Dans ce chapitre, nous présentons essentiellement les algorithmes propres aux sons bruités, même si certains d'entre eux ont été mis au point dans le cadre de la synthèse additive destinée aux sons harmoniques. Deux types de synthèse y sont abordés : la synthèse directe des échantillons, basée sur la représentation temporelle du son, ou la synthèse spectrale, basée sur la représentation fréquentielle.

La synthèse analogique du bruit peut être effectuée en amplifiant la tension aux bornes d'une résistance. En effet, à température non nulle, les électrons sont soumis à des mouvements aléatoires dans un conducteur (agitation thermique). Ces mouvements produisent une variation de potentiel (également aléatoire) aux bornes de ce conducteur, et donc une tension aléatoire, dite tension de bruit.

Dans la section 4.1, nous décrivons quelques algorithmes de synthèse de bruits classiques, à partir des propriétés de leur représentation sur le domaine temporel. Ensuite, après avoir détaillé les méthodes de synthèse spectrale de sons bruités dans la section 4.2, nous proposons dans la section 4.3 les algorithmes de synthèse propres au modèle CNSS que nous avons développé. Puis, dans la section 4.4, nous présentons les algorithmes rapides existants, employés pour le calcul des échantillons produits par des sinusoïdes. Enfin, nous proposons une discussion sur de nouveaux problèmes relatifs à la synthèse de sons bruités, ainsi que les solutions que nous envisageons, dans la section 4.5 pour ce qui concerne le chevauchement des fenêtres de synthèse et dans la section 4.6 l'amplitude et les risques de saturation.

4.1 Synthèse temporelle de bruit

Comme nous l'avons déjà vu, le mode de représentation le plus courant du signal sonore numérique est une suite de valeurs quantifiant les dépressions d'air à des intervalles de temps réguliers. La technique de synthèse adaptée à ce type de représentation est donc la génération directe d'échantillons, c'est-à-dire de valeurs numériques, définissant les variations d'amplitude au cours du temps. Concernant les sons bruités, pouvant être considérés comme des signaux aléatoires, cette méthode s'appuie sur les propriétés statistiques des échantillons, ou des différences entre plusieurs échantillons. Comme nous l'avons expliqué dans la section 2.2, cette approche ne permet seulement que la définition de sons bruités, dont l'enveloppe spectrale est de forme mathématique simple. Nous présentons donc dans cette section quelques algorithmes connus de synthèse de bruits classiques, à partir de leurs propriétés spécifiques.

4.1.1 Bruit blanc

Un bruit blanc est défini par un spectre d'amplitude continu et de moyenne constante (voir la section 1.4.3). En fait, la bande de fréquence de définition est restreinte à la bande de fréquence audible, c'est-à-dire à l'intervalle $[20; 20000]$Hz. En pratique, l'intervalle considéré sera plutôt $[0; \frac{F_s}{2}]$Hz, ce qui est pratiquement équivalent pour une fréquence d'échantillonnage de 44100Hz.

D'un point de vue statistique, les échantillons successifs sont indépendants, induisant une absence de périodicité. C'est cette propriété qui est utilisée pour l'algorithme de synthèse que nous proposons. Les échantillons sont ainsi considérés comme des variables aléatoires indépendantes, suivant une même loi de distribution. Plusieurs lois de distribution sont envisageables, pouvant donner lieu à des différences de perception liées à l'intensité perçue, mais surtout des différences du point de vue de l'efficacité. Nous abordons ici les deux lois les plus classiques, en précisant à chaque fois ces deux propriétés.

Distribution uniforme

La distribution la plus courante est la distribution uniforme, définie par une probabilité égale pour toutes les valeurs du domaine de définition :

$$\forall (x_1, x_2) \in]-1; 1]^2, p(x = x_1) = p(x = x_2) \tag{4.1}$$

Pour un signal numérique, le domaine de définition est l'intervalle réel $]-1; 1]$. La fonction de densité de probabilité ρ employée est donc définie par :

$$\forall x \in]-1; 1], \quad \rho(x) = \frac{1}{2} \tag{4.2}$$

En pratique, les langages de programmation proposent souvent une implémentation de fonctions de génération de nombres aléatoires (par exemple la fonction `rand` du langage `C`). Nous ne détaillons pas ici leurs algorithmes, mais quelques exemples peuvent être trouvés dans [Orf96] ou [Moo90]. Les valeurs fournies par ces fonctions sont généralement des nombres, notés Y_n, compris entre 0 et 1. Pour en déduire des valeurs d'échantillons, nous effectuons une simple opération :

$$\forall n, \quad X_n = 1 - 2Y_n \tag{4.3}$$

Le tableau 4.1 propose les valeurs des propriétés statistiques des premiers moments, obtenues à partir de 100 signaux de 44100 échantillons, générés en appliquant la fonction `rand` du logiciel `Matlab`. Ces propriétés montrent que les algorithmes semblent suffisamment performants pour synthétiser des bruits blancs. La figure 4.1 propose le spectre d'un tel signal.

moment	moyenne	écart-type	minimum	maximum
moyenne	-0.0003	0.0028	-0.0072	0.0066
variance	0.3333	0.0014	0.3301	0.3376
skewness	0.0006	0.0069	-0.0163	0.019
kurtosis	1.8000	0.0055	1.7855	1.812

TAB. 4.1 – Exemples de valeurs des propriétés statistiques (100 signaux) des quatre premiers moments pour un bruit blanc (durée 44100 échantillons, soit 1s), généré par tirage uniforme des échantillons (fonction `rand` du logiciel `Matlab`).

L'intensité perçue des bruits blancs synthétisés par distribution uniforme des échantillons peut être explicitée à partir des équations (4.2) et (1.45), puisqu'elle est reliée à l'amplitude RMS du signal, qui est égale à sa variance dans le cas d'un signal aléatoire. Sachant que la moyenne des échantillons produits est nulle ($\langle X \rangle = 0$), nous pouvons écrire :

$$\begin{aligned} A_{\mathrm{RMS}} &= \sqrt{\langle X^2 \rangle} \\ &= \sqrt{\int_{-1}^{1} x^2 \rho(x) dx} \\ &= \sqrt{\int_{-1}^{1} x^2 \frac{1}{2} dx} \\ &= \sqrt{\frac{1}{2} . \frac{2}{3}} \\ &= \frac{1}{\sqrt{3}} \end{aligned} \tag{4.4}$$

Par ailleurs, en appliquant cet algorithme de synthèse, les amplitudes définies par les échantillons sont toujours bornées (inférieures en valeur absolue à 1), ce qui permet d'assurer que le son produit ne présente aucune saturation :

$$\forall n, \quad |X_n| \leq 1 \tag{4.5}$$

Distribution gaussienne ou normale

Une autre distribution classique est la distribution de Gauss, aussi appelée distribution normale. Elle est définie dans la section 1.2.8. Elle est particulièrement intéressante, puisque les échantillons des sons réels ont une tendance naturelle à avoir une telle distribution.

L'algorithme de génération de valeurs distribuées selon cette loi est basé sur le théorème central limite (présenté en section 1.2.11). Ainsi, une addition d'un certain nombre N de variables aléatoires uniformément distribuées est définie par une loi de Gauss. En pratique, la somme de $N = 12$ variables uniformes X_i, choisies entre 0 et 1, est suffisante pour avoir une bonne approximation [Orf96] :

$$\forall n, \quad X_n = \sum_{i=1}^{N} X_n^i \tag{4.6}$$

Comme les variables X_n^i sont indépendantes, les variances et les espérances s'additionnent. Nous pouvons donc écrire la moyenne :

$$\begin{aligned} \forall n, \quad \mathcal{E}(X_n) &= \sum_{i=1}^{N} \mathcal{E}(X_n^i) \\ &= \frac{N}{2} \end{aligned} \tag{4.7}$$

ainsi que la variance :

$$\begin{aligned} \forall n, \quad V(X_n) &= \sum_{i=1}^{N} V(X_n^i) \\ &= \sum_{i=1}^{N} \langle X_n^2 \rangle - \langle X_n \rangle^2 \\ &= \sum_{i=1}^{N} \int_0^1 x_i^2 dx - (\frac{1}{2})^2 \\ &= \sum_{i=1}^{N} \frac{1}{3} - \frac{1}{4} \\ &= \frac{N}{12} \end{aligned} \tag{4.8}$$

Ainsi, l'algorithme suivant peut être déduit, permettant de définir N variables aléatoires distribuées selon une loi de Gauss de moyenne m et de variance σ^2 :

Entrée : W_s, N, m, σ
Sortie : tableau samples de taille W_s
début
  pour $i = 0$ *à* $W_s - 1$ **faire**
    sum = 0;
    pour $j = 0$ *à* $N - 1$ **faire**
      sum = sum + `rand`(0,1);
    fin
    samples[i] = σ.(sum-$\frac{N}{2}$)+m;
  fin
fin

Le tableau montre des exemples de propriétés statistiques des moments associés à 100 bruits blancs de $W_s = 44100$ échantillons, synthétisés à partir de l'algorithme précédent, avec une valeur N de 12. La figure 4.1 propose le spectre d'un tel signal. Nous pouvons vérifier que les échantillons vérifient convenablement les propriétés d'une loi de Gauss. Nous pensons

moment	moyenne	écart-type	minimum	maximum
moyenne	0.0001	0.0047	-0.0115	0.0111
variance	0.9994	0.0066	0.9800	1.0121
skewness	0.0009	0.0103	-0.0217	0.0224
kurtosis	2.8974	0.0233	2.8436	2.9761

TAB. 4.2 – Exemples de valeurs des propriétés statistiques (100 signaux) des quatre premiers moments pour un bruit blanc (durée $W_s = 44100$ échantillons, soit 1s) généré par tirage selon une loi de Gauss des échantillons ($N = 12$ avec la fonction `rand` du logiciel `Matlab`).

cependant que le choix de N à 12 est un peu limité, comme le montrent les valeurs du kurtosis du tableau de résultats. Une valeur de N supérieure semble nécessaire ($N \approx 25$).

Nous pouvons détailler l'intensité perçue d'un tel bruit. L'amplitude RMS d'un bruit blanc synthétisé par une distribution gaussienne, d'écart-type σ, des échantillons est :

$$A_{\text{RMS}} = \sigma \tag{4.9}$$

L'intensité perçue est plus faible comparée à celle d'un bruit blanc uniforme. La figure 4.1 met d'ailleurs en valeur cette différence.

Du fait de la définition de la fonction de densité de probabilité d'une loi de Gauss, donnant des probabilités non nulles pour des valeurs importantes, les valeurs des échantillons ne sont pas bornées, contrairement au cas de la distribution uniforme. La probabilité de tirer une valeur supérieure en valeur absolue à 1 n'est jamais nulle. Il existe donc une certaine possibilité que le son synthétisé sature. Nous étudions plus en détails ces probabilités dans la section 4.6. Nous pouvons toutefois dire que la détermination du maximum d'amplitude permet de déterminer une valeur pour σ. Un compromis est à trouver pour cette valeur : plus σ est important, plus la probabilité de saturer est importante, alors que plus σ est faible, plus l'intensité perçue est faible, mais plus la probabilité de saturer est faible.

4.1.2 Bruit rose

Les bruits roses , ou bruits en $1/f$, sont des bruits classiques, couramment utilisés, notamment en composition musicale aléatoire. Ils sont définis dans la section 1.4.4. Il existe des algorithmes simples pour générer de tels bruits, en contrôlant grossièrement les différences entre les échantillons consécutifs, qui ne sont plus indépendants.

Algorithme de Voss

L'algorithme que nous décrivons ici est le plus connu, et a été mis au point par Voss [VC78]. Il est relativement simple et efficace [Bol83], même s'il existe des versions encore plus rapides. Il est présenté dans quelques ouvrages parmi lesquels [Moo90]. Il repose sur deux principes de base. Tout d'abord, chaque échantillon est défini par la somme de N variables aléatoires, uniformément distribuées. Ensuite, chaque échantillon est obtenu à partir du précédent en ne tirant à nouveau qu'un certain nombre de ces variables. Ainsi, deux variables successives auront plus de probabilités d'être proches. Les hautes fréquences sont donc atténuées, conformément à la définition du spectre de tels bruits.

indice i	0	1	2	3	4	5	6	7
bits	000	001	010	011	100	101	110	111
retirages	3	1	2	1	3	1	2	1

TAB. 4.3 – Illustration du principe de l'algorithme de Voss pour la synthèse d'un bruit rose. Dans cet exemple, seulement 3 bits sont utilisés pour décider le nombre de retirages effectués pour déterminer successivement chaque échantillon.

En pratique, le nombre de tirages est décidé par le nombre de changements de bits représentant l'indice de l'échantillon courant. Ainsi, pour synthétiser $65536 = 2^{16}$ échantillons, le nombre de bits sera de 16. Au début, tous les bits sont à 0 et le premier échantillon d'indice 0 est produit par l'addition de 16 nombres aléatoires uniformément distribués. Le second échantillon d'indice 1 sera défini par ces mêmes valeurs, exceptée une seule (un seul bit différent entre la représentation de 0 et de 1) qui est tirée à nouveau. L'itération continue jusqu'à obtenir 65536 échantillons pour cet exemple. Le tableau 4.3 montre un autre exemple plus simple avec seulement 8 échantillons à produire (3 bits utilisés) et indique, pour chaque itération, le nombre de tirages nécessaires, le premier échantillon étant défini par 3 valeurs aléatoires. L'intervalle entre deux échantillons (variables aléatoires) successifs est donc une valeur aléatoire qui dépend du nombre de valeurs uniformément distribuées redéfinies, c'est-à-dire du nombre de bits de l'indice entier i modifiés. L'algorithme général suivant détaille les étapes successives de la synthèse de 2^N échantillons d'un bruit rose dont la figure 4.1 propose le spectre :

```
Entrée : N
Sortie : tableau samples de taille 2^N
début
  pour i = 0 à N − 1 faire
    oldbit[i] = 1;
  fin
  pour K = 0 à 2^N − 1 faire
    K = bit[0], bit[1], . . . , bit[N-1];
    sum = 0;
    pour i = 0 à N − 1 faire
      si bit[i] ≠ oldbit[i] alors
        d[i] = rand(0.1);
      fin
      sum = sum + d[i];
      oldbit[i] = bit[i];
    fin
    samples[K] =sum;
  fin
fin
```

Puisqu'un certain nombre de valeurs aléatoires uniformément distribuées sont ajoutées, pouvant entraîner une somme totale supérieure en valeur absolue à 1, des saturations sont probables. Pour éviter cela, nous avons choisi de définir un échantillon comme une somme de 16 nombres compris entre -0.1 et 0.1. La littérature indique de choisir les valeurs dans l'intervalle compris entre $-\frac{1}{N}$ et $\frac{1}{N}$ [Moo90].

4.1.3 Bruit brownien

Le bruit brownien , ou bruit en $1/f^2$ est défini, dans la section 1.4.4, par rapport aux caractéristiques du spectre. Une autre définition repose sur les liens entre les échantillons de tels bruits. L'écart entre deux échantillons X est une variable aléatoire uniformément distribuée, notée Y :

$$\forall n, \quad X_{n+1} = X_n + Y_n \tag{4.10}$$

C'est à partir de cette définition qu'un algorithme de synthèse simple et efficace peut être proposé. Il est d'ailleurs intéressant de noter que cette définition permet de montrer qu'un bruit brownien est synthétisé à partir d'un bruit blanc et d'un filtre à réponse impulsionnelle infinie.

Algorithme

L'algorithme de génération des bruits browniens consiste à définir aléatoirement un premier échantillon. Les autres échantillons sont produits en ajoutant à cette valeur, un autre réel aléatoire uniformément distribuée entre $-k$ et k, avec $k < 1$. La figure 4.1 propose le spectre d'un signal synthétisé par cette méthode, décrite par l'algorithme suivant :

Entrée : W_s
Sortie : tableau samples de taille W_s
début
    sum = `rand`(−1,1);
    oldsum = sum;
    pour $i = 0$ *à* $W_s - 1$ **faire**
        sum = oldsum + `rand`(−0.1,0.1);
        tant que *sum* > 1.0 *OU sum* < -1.0 **faire**
            sum = oldsum + `rand`(−0.1,0.1);
        fin
        samples[i] =sum;
        oldsum = sum;
    fin
fin

Les bornes, définies par le paramètre k, ont une importance en ce qui concerne les bornes des échantillons produits, et donc une éventuelle saturation. Comme dans le cas du bruit blanc gaussien, il est probable qu'un échantillon prenne une valeur supérieure en valeur absolue à 1. Cette probabilité est d'autant plus grande que la valeur k est importante. Lors de nos expériences, nous avons été amené à utiliser la valeur $k = 0.1$, mais aussi à effectuer une vérification pour éviter ces débordements. Ces choix, proposés dans l'algorithme présenté, sont également conseillés dans [Moo90].

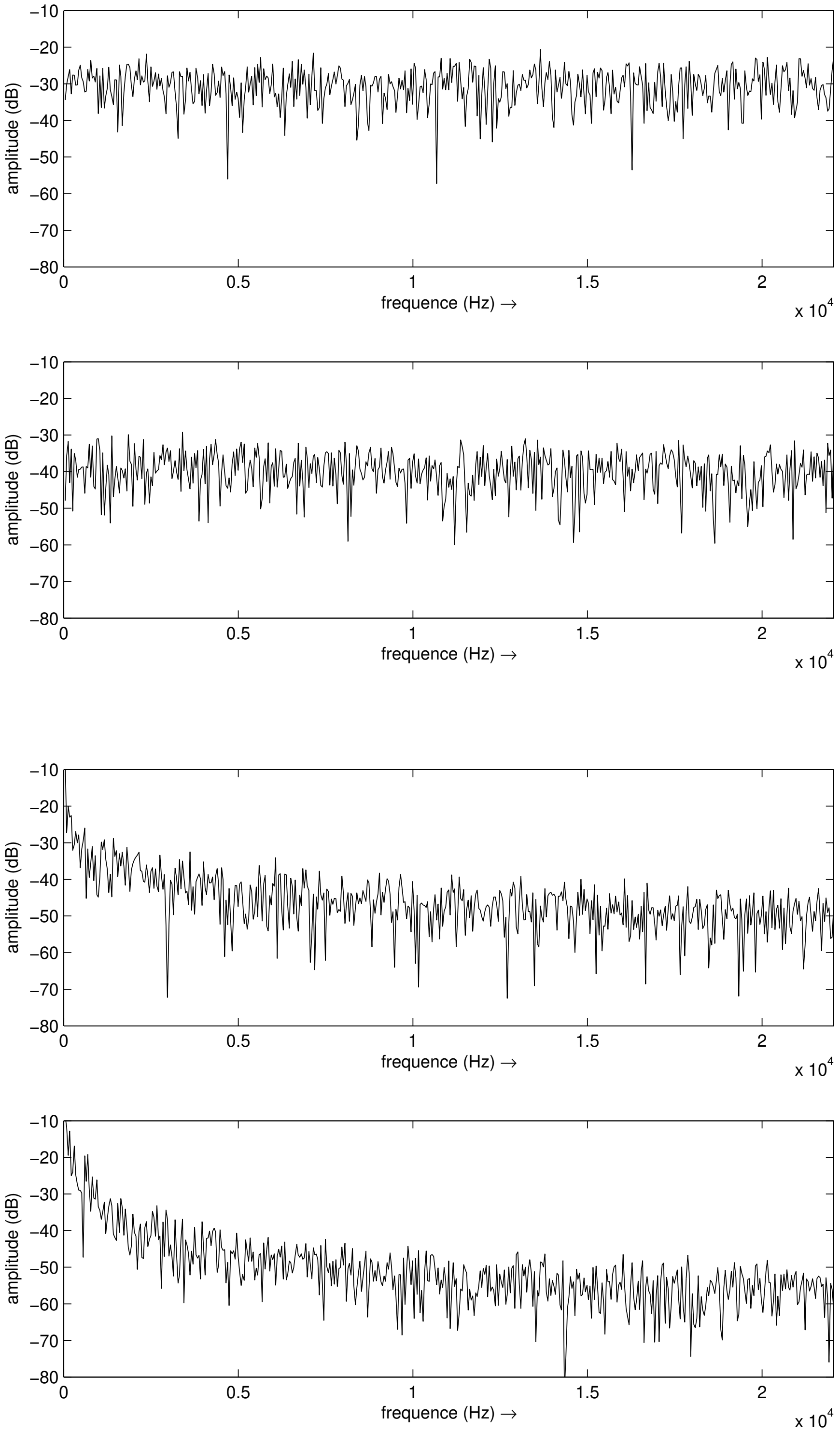

FIG. 4.1 – Représentations spectrales sur 1024 échantillons (43ms) de bruits synthétisés selon les algorithmes présentés, respectivement de haut en bas d'un bruit blanc uniforme, d'un bruit blanc gaussien, d'un bruit rose et d'un bruit brownien.

4.1.4 Bruit ondulant

Le bruit ondulant (*rippled noise*) est défini dans la section 1.4.5. Il présente la particularité de laisser percevoir une hauteur [Yos96]. Il est fréquemment utilisé lors d'expériences psychoacoustiques, et les méthodes de synthèse employées ont été développées dans le cadre de ces expériences. Elles consistent à ajouter à un bruit blanc une version retardée et atténuée de lui-même, comme l'illustre la figure 4.2.

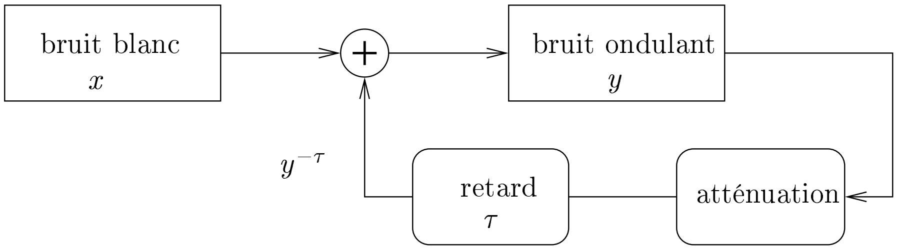

Fig. 4.2 – Illustration de la méthode de synthèse du bruit ondulant.

Algorithme

L'algorithme de synthèse d'un bruit ondulant définit comme paramètre un taux d'atténuation (réel a inférieur à 1), et un temps de retard τ (en échantillons) qui est directement lié à la hauteur que laisse percevoir le son produit. Les τ premiers échantillons sont aléatoirement choisis selon une distribution uniforme (ou une distribution gaussienne). Les échantillons suivants d'indice i sont calculés par la somme pondérée d'un échantillon aléatoire et de l'échantillon précédent d'indice $i - \tau$:

Entrée : a, τ
début
  pour $i = 0$ *à* $n - 1$ **faire**
    $x[i] =$ `rand`$(-1, 1)$;
  fin
  pour $i = 0$ *à* $\tau - 1$ **faire**
    samples[i] $= x[i]$;
  fin
  pour $i = \tau$ *à* $n - 1$ **faire**
    samples[i] $= a\times$ samples[$i - \tau$]$+\ x[i]$;
  fin
  Résultat : samples
fin

Applications

La hauteur f_0 que laisse percevoir ce son est directement liée à la valeur τ, qui définit la périodicité :

$$f_0 = \frac{F_s}{\tau} \tag{4.11}$$

Le taux d'atténuation a détermine, lui, la force de cette périodicité. Le spectre d'un bruit ondulant, tel que celui présenté par la figure 4.3, laisse apparaître des pics régulièrement espacés, d'une distance f_0Hz. Des études liées à la perception de la hauteur ont porté sur ces bruits, et notamment sur le fonctionnement du système auditif pour mesurer la force de la sensation de hauteur [WPDC98]. La fonction d'autocorrélation permet de mettre en valeur cette périodicité, comme le montre la figure 4.3.

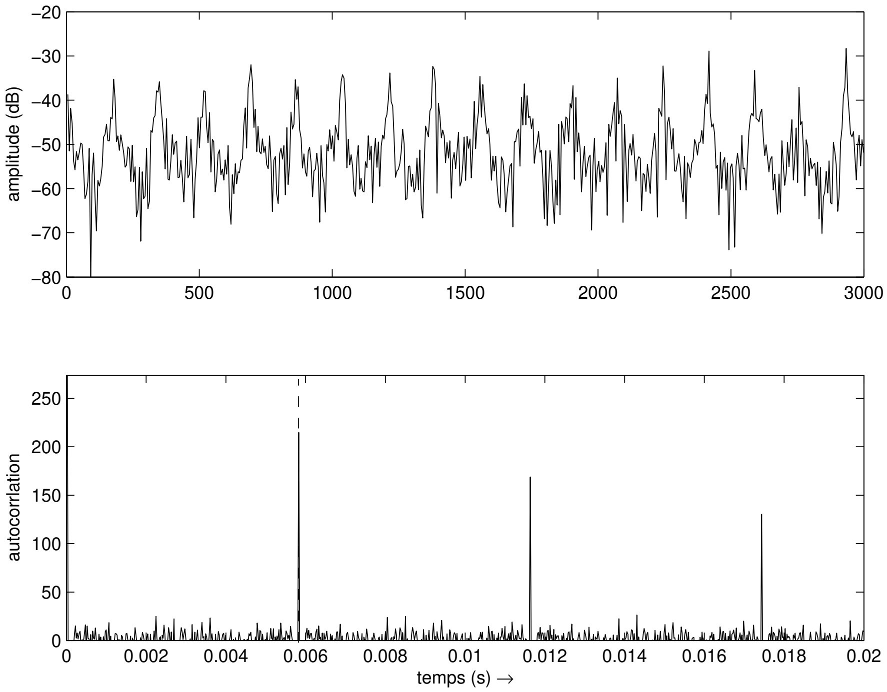

FIG. 4.3 – Représentation spectrale sur 8192 échantillons (0.185s) et fonction d'autocorrélation d'un bruit ondulant de synthèse (temps de retard de 256 échantillons, soit 5.8ms) : l'écart entre les pics réguliers de l'enveloppe spectrale est de 172Hz, fréquence qui doit correspondre à la hauteur perçue et à la périodicité mise en valeur par la fonction d'autocorrélation (en traits hachurés, la valeur de 5.8ms).

4.1.5 Bruit blanc filtré

Les algorithmes précédents permettent de générer un bruit dont l'enveloppe spectrale a une forme particulière, mais qui reste simple (en $\frac{1}{f}$, en $\frac{1}{f^2}$, etc...). Des méthodes de synthèse temporelles autorisent la production d'échantillons définissant un bruit dont l'enveloppe spectrale est plus complexe. La technique principale est la prédiction linéaire présentée dans les sections 2.2 et 3.5, ou plus généralement le filtrage d'un bruit blanc. C'est le principe de la synthèse soustractive.

Principe

L'équation générale d'un filtre peut s'écrire sous la forme suivante :

$$y[n] = \sum_{i=1}^{N} a_i x[n] + \sum_{i=1}^{M} b_i y[n-i] \tag{4.12}$$

où a_i et b_i sont les coefficients du filtre, x le signal source ou l'erreur, et y le signal synthétisé. A partir de la connaissance des coefficients et du signal d'entrée (généralement un bruit blanc), les échantillons sont produits successivement. La mise au point d'un filtre impose une contrainte sur les coefficients pour s'assurer que le filtre est bien stable, c'est-à-dire que les échantillons générés restent bornés et ne convergent pas vers l'infini.

Cette méthode est bien sûr très intéressante du point de vue de l'efficacité puisque le nombre d'opérations nécessaires pour produire un échantillon est très réduit, ne dépendant que de la complexité du filtre (valeurs N et M).

Limites

La principale limite de cette approche a déjà été abordée dans la section 2.2, et est liée à la représentation non intuitive de l'enveloppe spectrale du son par des coefficients mathématiques. Les transformations musicales sont complexes, car la modification d'un coefficient implique un changement de l'enveloppe spectrale pour toutes les valeurs de fréquences.

Par ailleurs, la complexité de l'enveloppe spectrale est liée au nombre de coefficients, qui dépend ainsi du nombre de pics et de creux de l'enveloppe. Plus l'enveloppe est perturbée, plus il est difficile de définir un filtre stable. Les enveloppes spectrales envisageables restent donc de forme relativement basique, permettant juste de définir des formants, utiles par exemple pour la synthèse de la parole.

4.2 Synthèse spectrale de son bruité

La synthèse additive permet de synthétiser un son en sommant les contributions aux sons de formes d'ondes élémentaires. Le cas le plus classique, et le plus proche de la perception est la somme de sinusoïdes. Cette synthèse est particulièrement adaptée à la modélisation sinusoïdale, puisque les sons concernés par ces modèles sont représentés par un nombre réduit de sinusoïdes, qui dépend de la hauteur fondamentale.

Concernant les bruits, il a été difficile jusqu'à présent d'envisager cette méthode. Tout d'abord la représentation de bruits par des sinusoïdes paraissait forcément impliquer des artefacts audibles, du fait de leurs variations trop lentes. Ensuite, la théorie indique que cette représentation n'est possible que si un nombre important de sinusoïdes est considéré, limitant les applications en temps-réel. Les performances des machines actuelles autorisent depuis quelques temps une telle approche. Nous abordons dans cette section le principe général de la synthèse spectrale de sons bruités, basée sur le modèle du bruit thermique.

4.2.1 Principe de la synthèse spectrale

Comme l'indique la définition des bruits thermiques (voir la section 2.4.2), un son bruité peut être défini comme une somme fixe de sinusoïdes. Les caractéristiques (fréquences, amplitudes, phases) de ces sinusoïdes permettent de contrôler les propriétés auditives du son bruité associé. Une fois que ces caractéristiques ont été décidées, les échantillons induits par chaque sinusoïde sont calculés et sommés pour former le son résultant.

Il est évident que le son synthétisé est défini par des fenêtres successives. Cela permet d'une part de réduire le nombre de sinusoïdes à calculer, puisque, comme nous l'avons vu dans la section 2.4.8, ce nombre est relatif à la taille de la durée de la fenêtre. D'autre part, cela permet de pouvoir prendre en compte les variations du son au cours du temps.

4.2.2 Définition des sinusoïdes

Chaque sinusoïde composant le son est définie par une fréquence, une amplitude et une phase. Nous détaillons dans cette sous-section les possibilités pour définir ces caractéristiques, les liens avec les modèles présentés et l'influence sur la perception.

Amplitudes de Rayleigh et phases aléatoires

La définition du bruit thermique impose des fréquences régulièrement espacées, des phases uniformément distribuées entre 0 et 2π, et des amplitudes aléatoirement choisies selon une loi de Rayleigh (définie dans la section 1.2.9) :

$$\forall n \in [\![0; W_s - 1]\!], \quad X_n = \sum_{k=1}^{N} A_k \sin(2\pi f_k \frac{n}{F_s} + \Phi_k) \tag{4.13}$$

Ainsi, pour définir une bande de bruit de largeur $\Delta F = F_{\max} - F_{\min}$, de densité maximale, et de durée W_s échantillons, il est nécessaire de produire $N = \frac{W_s.\Delta F}{F_s}$ fréquences f_k régulièrement espacées :

$$\forall k \in [\![1; \frac{W_s.\Delta F}{F_s}]\!], \quad f_k = F_{\min} + k\frac{F_s}{W_s} \tag{4.14}$$

Les phases sont donc aléatoirement choisies entre 0 et 2π, selon une loi uniforme avec une fonction du type `rand` du langage `C` :

$$\forall k \in [\![1; \frac{W_s.\Delta F}{F_s}]\!], \quad \Phi_k = 2\pi.X_k \tag{4.15}$$

où X est une variable aléatoire uniformément distribuée entre 0 et 1.

Les amplitudes sont générées aléatoirement à partir de l'enveloppe spectrale voulue, qui induit une amplitude a_k relative à la fréquence f_k. La valeur a_k précise la loi de Rayleigh utilisée pour la génération de l'amplitude A_k aléatoire, en définissant la valeur la plus probable de la loi, notée σ dans l'équation (1.61). Ainsi, la valeur aléatoire A_k est obtenue à partir de l'équation (1.63) :

$$A_k = a_k\sqrt{-2\ln(1 - X_i)} \tag{4.16}$$

où X_i sont des variables aléatoires uniformément distribuées entre 0 et 1.

L'avantage de cette technique est que le bruit produit voit ses échantillons suivre une loi de Gauss, ressemblant ainsi aux sons naturels, dès que le nombre de sinusoïdes est supérieur ou égal à 1 [Har97]. Mais l'inconvénient associé provient du fait que l'enveloppe spectrale est aléatoirement définie, ce qui ne permet pas de décider exactement de son spectre. En dehors de la définition mathématique des bruits thermiques, nous pouvons également nous interroger sur la nécessité de définir aléatoirement les amplitudes, puisque l'oreille ne semble pas particulièrement sensible aux faibles variations ainsi introduites. A notre connaissance, il n'existe aucun travail de recherche qui justifie ce choix par rapport à la perception. C'est pourquoi les méthodes habituelles employées font abstraction de cette contrainte pour les amplitudes. Des expériences restent à mener pour justifier ou non ce choix.

Amplitudes fixes et phases aléatoires

L'autre possibilité de technique de synthèse de sons bruités consiste à fixer les amplitudes a_k associées aux sinusoïdes, qui ne sont alors plus des variables aléatoires. Elles sont complètement définies par l'enveloppe spectrale voulue. Dans ce cas, seules les valeurs de phases Φ_k des sinusoïdes restent des variables aléatoires uniformément distribuées entre 0 et 2π :

$$\forall n \in [\![0; W_s - 1]\!], \quad X_n = \sum_{k=1}^{N} a_k \sin(2\pi f_k \frac{n}{F_s} + \Phi_k) \tag{4.17}$$

Les signaux produits ont alors exactement le même spectre d'amplitude.

Cette méthode n'est pas fidèle à la définition des bruits thermiques. Toutefois, le théorème central limite (voir la section 1.2.11) nous assure que, pour un nombre raisonnable de sinusoïdes (dès $N = 5$ [Har97]), le signal produit est de distribution gaussienne.

C'est cette approche qui est implicitement appliquée pour les modèles spectraux existants, notamment les parties bruitées des modèles hybrides telles que la partie stochastique du modèle SMS et ses dérivés (STN, ...), présentées dans la section 2.3.5. En effet, La synthèse de la partie stochastique du modèle SMS s'effectue à partir d'une approximation de l'enveloppe spectrale, qui permet par interpolation de définir les amplitudes a_k des sinusoïdes. La technique de synthèse consiste d'abord à définir ainsi le spectre d'amplitude. Ensuite, le spectre de phases est aléatoirement généré par tirage uniforme des phases entre 0 et 2π, permettant de déterminer le spectre complexe symétrique. Enfin, une transformée inverse de Fourier permet de produire les W_s échantillons du son synthétisé. La figure 4.4 illustre cette méthode de synthèse.

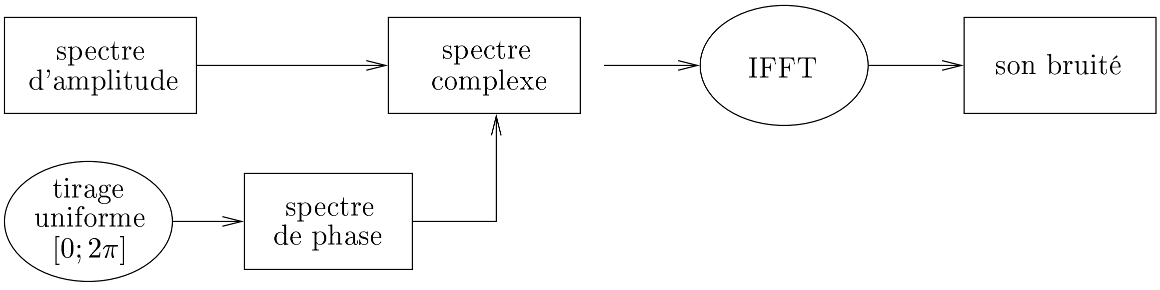

FIG. 4.4 – Illustration de la méthode de synthèse de la partie stochastique du modèle hybride SMS.

Ainsi, les valeurs des fréquences f_k sont définies par la discrétisation de l'échelle fréquentielle de la transformée de Fourier. Elles sont séparées de la valeur $\frac{F_s}{W_s}$ et donc implicitement espacées de façon régulière. Cette opération est équivalente à la synthèse additive de $\frac{W_s}{2}$ sinusoïdes :

$$\forall n \in [\![0; W_s - 1]\!], \quad X_n = \sum_{k=1}^{\frac{W_s}{2}} a_k \sin(2\pi k \frac{F_s}{W_s} \frac{n}{F_s} + \Phi_k) \tag{4.18}$$

Il est important de rappeler que cette méthode de synthèse considère donc le nombre de sinusoïdes nécessaires fixé et directement lié à la durée du son (ou de la fenêtre de son) produit. Il ne peut d'ailleurs être modifié, à moins de définir des spectres d'amplitudes dont certaines valeurs sont nulles.

4.2.3 Low-noise Noise

Dans les deux algorithmes précédents, basés sur la définition du bruit thermique, les phases sont supposées être uniformément distribuées entre 0 et 2π. La conséquence d'une réduction de cet intervalle est l'augmentation des fluctuations d'intensité, comme cela est montré dans la sous-section 2.5.3. De même, une augmentation de cet intervalle implique également une hausse de ces fluctuations, à cause des propriétés de périodicités des sinusoïdes. Par exemple, si l'intervalle est doublé, il n'y aura aucune conséquence. S'il est augmenté de moitié, la probabilité de choisir des valeurs entre 0 et π sera doublée, ce qui entraîne une augmentation des valeurs de phases égales, et donc une probabilité plus importante de pics d'intensité.

Des travaux de recherche en psychoacoustique portant sur les propriétés de masquage des bandes de bruit ont étudié la possibilité de modifier cette distribution, dans le but de réduire leurs fluctuations d'intensité. Nous présentons ici deux algorithmes permettant de générer un bruit dont les fluctuations d'intensité sont inférieures à celles produites par une distribution uniforme des phases entre 0 et 2π.

Algorithme de Schroeder

L'algorithme de Schroeder [Sch70] présenté en 1970, décrit dans [HP88], propose une méthode de détermination des phases conduisant à la synthèse d'un bruit dont les fluctuations d'intensité, décrites par le facteur de crête (la rapport de l'amplitude maximale sur l'amplitude RMS), sont faibles. Les phases ne sont pas dans ce cas des variables aléatoires et sont déterminées par la relation :

$$\phi_n = \phi_0 - \frac{\pi n^2}{N} \tag{4.19}$$

où n représente l'indice des sinusoïdes et ϕ_0 est une valeur de phase constante. L'algorithme de synthèse qui en est déduit est évidemment trivial, puisque les amplitudes sont fixes, les fréquences sont également réparties, et les phases sont calculées de façon simple.

Toutefois, cette méthode de synthèse mène certes à un bruit dont les fluctuations d'intensité sont moindres que dans le cas de phases aléatoires uniformément distribuées, mais les sensations qu'il provoque sont éloignées de celles d'un bruit, car le son produit se comporte comme une sinusoïde dont la fréquence augmente linéairement dans le temps [Pum85]. Il ne peut donc être appliqué à la synthèse de sons bruités.

Algorithme de Pumplin

Un autre algorithme de synthèse a été proposé par Pumplin [Pum85], permettant de synthétiser un bruit de faibles fluctuations d'intensité, tout en préservant les qualités perceptives caractéristiques des bruits. Il consiste à minimiser les fluctuations d'intensité, en minimisant le kurtosis par itérations successives, à partir d'un bruit produit avec des valeurs de phases uniformément distribuées entre 0 et 2π. Une valeur de phase est déterminée et est modifiée par petits pas successifs, de façon à s'approcher de la valeur qui induit le kurtosis minimum. Cette opération est effectuée autant de fois que voulu, sur plusieurs valeurs de phase. Bien évidemment l'algorithme est très coûteux en temps de calcul, d'autant plus que la précision souhaitée est importante.

Un algorithme dérivé de cette méthode est également proposé dans [Pum85], utilisant la transformée rapide de Fourier. Au début, le signal est généré à partir de valeurs de phases uniformément distribuées. Un seuil t_h est arbitrairement fixé, et les valeurs d'amplitudes supérieures à ce seuil sont coupées :

$$\forall t, \forall |x(t)| > t_h, \quad \begin{cases} x(t) = -t_h & \text{si} \quad x(t) < 0 \\ x(t) = t_h & \text{si} \quad x(t) > 0 \end{cases} \tag{4.20}$$

Le spectre d'amplitude obtenu par transformée rapide de Fourier est modifié de façon à restaurer les valeurs d'amplitude originale, tout en préservant le spectre de phase. En repassant sur la représentation temporelle par transformée inverse de Fourier, le signal est à nouveau coupé par rapport au seuil t_h, etc.... Les itérations successives convergent vers un spectre d'amplitude qui diminue les fluctuations d'intensité. Cet algorithme est plus rapide, mais le son synthétisé ne parvient pas au niveau de fluctuations aussi faible que ne le permet l'algorithme proposé par Pumplin. Ils peuvent ainsi être combinés.

Applications

Les applications de ces types de bruits sont essentiellement liées au phénomène de masquage. Il a été montré expérimentalement dans [HP88] que les bandes étroites de tel bruit sont moins efficaces dans le masquage d'une sinusoïde de fréquence constante que les bandes de même largeur, mais générées à partir de phases uniformément distribuées. Du point de vue musical, leur intérêt est moins évident. Toutefois, dans le cas d'un modèle hybride, ces propriétés de masquage pourraient conduire à des sons dont la partie déterministe (sinusoïdale) serait plus ou moins masquée par la partie bruitée.

4.3 Synthèse CNSS

Dans cette section, nous détaillons les algorithmes de définition des sinusoïdes à partir des paramètres mathématiques du modèle CNSS, détaillé dans la section 2.4. Le modèle CNSS est un modèle spectral, basé sur le modèle du bruit thermique. L'algorithme général de synthèse produit les échantillons du signal en déterminant les caractéristiques des sinusoïdes le représentant, c'est-à-dire leur fréquence, amplitude et phase. Nous présentons successivement les algorithmes liés à chaque caractéristique.

4.3.1 Détermination des casiers

La définition des casiers s'effectue à partir du nombre de casier M et de la largeur de la bande de fréquence $\Delta F = F_{\max} - F_{\min}$. Nous notons ΔB la taille constante des casiers, calculée selon l'équation :

$$\Delta B = \frac{\Delta F}{M} \tag{4.21}$$

A partir de cette différence calculée, les M casiers consécutifs sont définis, à partir de leur valeur de fréquence supérieure, notée B_i :

$$\forall i \in [\![1; M]\!], \quad B_i = F_{\min} + i.\frac{\Delta F}{M} \tag{4.22}$$

C'est à partir de ces valeurs que l'algorithme définit, si besoin est, les fréquences dans chaque casier.

Dans le cas où le nombre de sinusoïdes N est strictement inférieur au nombre de casiers M, l'algorithme doit aléatoirement décider quels casiers contiennent une fréquence, et quels casiers sont vides. Le premier choix naïf consiste à effectuer un tirage uniforme sur tous les casiers, ce qui se ramène à choisir aléatoirement un nombre entre 1 et M, affecté à l'indice d'un casier. Une comparaison est effectuée avec les nombres déjà tirés. Lorsqu'un casier (un indice) a déjà été choisi, il est rejeté. Cette opération est répétée tant que N casiers n'ont pas été sélectionnés.

Cette façon de procéder n'est pas statistiquement correcte, puisque la distribution uniforme n'est dans ce cas plus respectée. De plus, elle impose pour chaque tirage une série de comparaisons, ce qui accroît la complexité. La deuxième solution que nous proposons et que nous avons appliquée provient des algorithmes de génération aléatoire de permutations. Un indice i (un casier B_i) est aléatoirement choisi sur l'intervalle $[\![1; M]\!]$ selon une loi uniforme. Après ce tirage, les indices M et i sont échangés. Un nouveau tirage uniforme est alors effectué sur les $M - 1$ casiers restants, cette opération se répétant jusqu'à ce que N casiers aient été choisis. La figure 4.5 illustre le principe de cet algorithme.

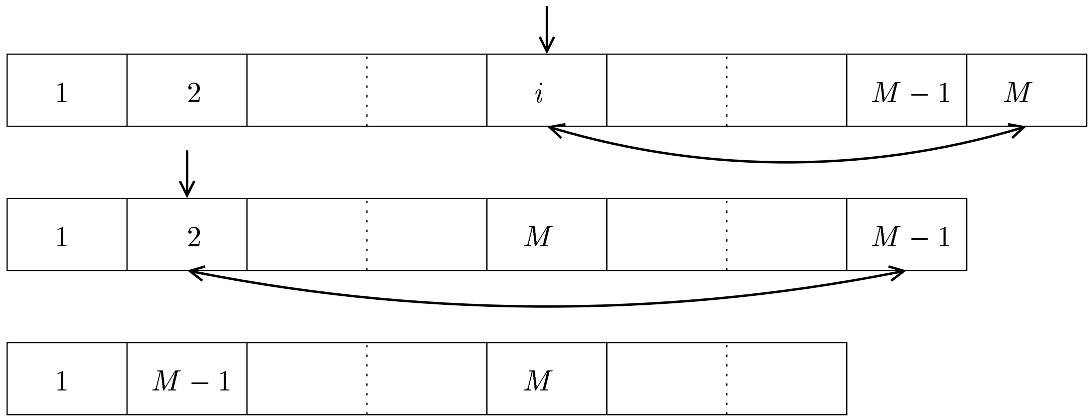

FIG. 4.5 – Illustration de l'algorithme de tirage de N casiers parmi M possibles : le premier tirage uniforme détermine la casier i, qui est alors échangé avec le dernier casier M (première figure), le deuxième tirage donne le deuxième casier qui est alors échangé avec le dernier casier $M - 1$, etc....

Cet algorithme a un coût non négligeable lorsque le nombre de casiers M est important comparé au nombre de sinusoïdes N, car un tableau de taille importante doit être initialisé et manipulé pour chaque fenêtre. Nous pouvons toutefois prendre en compte le fait que si le nombre de casiers M est vraiment très important par rapport au nombre de sinusoïdes N ($M >> 10N$), il peut alors être considéré comme infini. Le choix aléatoire des N fréquences est alors effectué par un tirage uniforme sur toute la bande de fréquence considéré ΔF, ce qui diminue le temps de calcul et l'espace mémoire utilisé.

Nous pouvons également considérer le cas particulier où le nombre de casier M est égal au nombre de sinusoïdes N. L'algorithme précédent de détermination des casiers est bien entendu valable dans ce cas, mais il implique des tirages successifs qui sont inutiles. Pour économiser du temps de calcul, une comparaison peut être effectuée et, en cas d'égalité, associer directement les sinusoïdes d'indice i aux casiers B_i. Néanmoins, nous voyons dans la section 4.5 que certaines techniques liées au chevauchement des fenêtres de synthèse dans le cas de signaux aléatoires imposent un ordre aléatoire des fréquences, c'est-à-dire que la $i$ème sinusoïde ne doit pas toujours correspondre au $i$ème casier. Plus généralement, les expériences montrent que la quasi-majorité du temps de calcul lors de la synthèse est utilisée lors du calcul des sinusoïdes. C'est pourquoi nous avons choisi, dans les implémentations que nous avons développées, d'appliquer l'algorithme des permutations aléatoires, même si le nombre de sinusoïdes N est égal au nombre de casiers M.

4.3.2 Détermination des fréquences

Une fois que les casiers sont déterminés, les valeurs de fréquences sont aléatoirement choisies à partir du paramètre L de largeur de la fonction de densité de probabilité uniforme pour chaque casier. Pour chaque fréquence, un tirage uniforme d'un réel r est effectué entre 0 et L. La fréquence F_i, contenue dans le casier défini par B_j, est déduite de la relation :

$$\forall i \in [\![1; M]\!], \quad F_i = B_j - r\Delta B \tag{4.23}$$

L'algorithme suivant résume les différentes étapes de la définition des fréquences des N sinusoïdes du signal :

Dans le cas particulier où le nombre de casiers M est considéré comme infini ($M >> 10N$), les fréquences ne sont calculées qu'à partir de la largeur de bande ΔF du signal, en effectuant un tirage uniforme :

$$F_i = F_{\min} + \texttt{rand}(\Delta F) \tag{4.24}$$

où `rand` représente une fonction de tirage de nombre aléatoire, selon une loi uniforme, compris entre 0 et l'argument de la fonction.

4.3.3 Détermination des amplitudes

Les amplitudes relatives aux fréquences préalablement définies sont déterminées à partir du paramètre couleur, c'est-à-dire d'une approximation de l'enveloppe spectrale. Celle-ci est représentée par une suite de points, plus ou moins espacés selon la méthode employée. Par exemple, dans le cas de la méthode du filtrage du cepstre, l'amplitude a_i correspondant à la fréquence F_i est calculée par interpolation linéaire entre les deux valeurs de points de l'enveloppe spectrale les plus proches.

Entrée : Paramètres M,N,L,ΔF
début
  pour *chaque fenêtre* **faire**
    $\Delta B = \frac{\Delta F}{M}$;
    pour $i \in 0, \ldots, M-1$ **faire**
      $B[i] = F_{\min} + (i+1)\frac{\Delta F}{M}$;
    fin
    Définir un tableau b d'entiers $[\![0; M-1]\!]$;
    pour $n \in 0, \ldots, N-1$ **faire**
      Tirer un entier k à partir des $M-n$ entiers restants;
      Tirer un réel r dans $[0; L]$;
      $F_i = B_j - r\Delta B$;
      $b[k] = b[M-n-1]$;
    fin
  fin
fin

Des problèmes apparaissent pour les sons bruités pour la détermination de l'amplitude générale du son et la possibilité de saturation. En effet, le nombre de sinusoïdes N est lié à l'intensité perçue, approximée par l'amplitude RMS. Ainsi, il est nécessaire de modifier les amplitudes des fréquences en fonction du nombre de sinusoïdes N. Nous étudions les différentes possibilités que nous avons envisagées pour cela dans la section 4.6.

4.3.4 Détermination des phases

Le modèle CNSS est basé sur le modèle du bruit thermique, qui impose aux valeurs de phases des sinusoïdes composant le signal d'être uniformément distribuées entre 0 et 2π. Le paramètre noté P permet de définir la largeur de la fonction de densité de probabilité de ces phases, l'intérêt étant de pouvoir contrôler un éventuel pic d'intensité se produisant lorsque beaucoup de phases sont égales. En ce qui concerne la synthèse, l'utilisation du chevauchement des fenêtres successives de signal induit une multiplication par une fenêtre de pondération, de valeur nulle à ses limites. Il est nécessaire de prendre en compte cette caractéristique en synchronisant à l'instant correspondant à chaque moitié des fenêtres de synthèse les phases aléatoirement choisies.

Ainsi, l'algorithme de synthèse de phases consiste à d'abord effectuer N tirages uniformes de réels, notés R_i, situés entre 0 et 1, puis à les utiliser pour en déduire les valeurs de phases Φ_i selon la relation :

$$\Phi_i = \frac{\pi}{2} + 2(R_i - \frac{1}{2})P\pi + \frac{\pi F_i W_s}{F_s} \tag{4.25}$$

Nous étudions actuellement une extension de cette approche pour la détermination des phases. En effet, pour l'instant, le tirage est effectué conformément à la même loi de distribution, sur toutes les fenêtres de synthèse. Une valeur faible du paramètre P conduit à un pic d'intensité (une impulsion), au milieu de chaque fenêtre. Ainsi, pour envisager la synthèse d'un bruit composé d'impulsions dont la période est supérieure à la taille de la fenêtre, il est obligatoire de modifier régulièrement et rapidement la valeur du paramètre P. Il est plutôt souhaitable de définir des paramètres permettant l'automatisation de ces changements. Pour cela, nous expérimentons l'utilisation de ces nouveaux paramètres pour le modèle de synthèse, en supplément du paramètre P :

- La longueur τ de la période moyenne des impulsions définit le temps moyen s'écoulant entre deux impulsions, c'est-à-dire la durée entre deux fenêtres pour lesquelles le paramètre P (différent de 1) sera pris en compte.
- L'intervalle des durées T entre deux impulsions permet de définir les bornes des durées envisageables entre deux impulsions. Un tirage aléatoire est ainsi effectué pour chaque fenêtre contenant une impulsion, déterminant l'instant de la prochaine.
- la largeur relative g ($0 \leq g \leq 1$) de la fonction de densité de probabilité de la valeur de la phase moyenne pour toutes les sinusoïdes. Pour l'instant, celle-ci était considérée comme toujours égale à $\frac{\pi}{2}$, conduisant à un pic d'intensité maximale dans le cas où le paramètre P vaut 0. Cette valeur peut être modifiée, permettant de diminuer la valeur maximale de ce pic. Le paramètre g définit les variations aléatoires de cette valeur maximale.
- La largeur relative P de la fonction de densité de probabilité des valeurs de phases est le même paramètre que celui employé précédemment. La différence ici provient du milieu de l'intervalle considéré, défini aléatoirement à partir du paramètre précédent.

Ces quatre nouveaux paramètres permettent de déterminer de manière plus complexe les phases, mais autorisent des contrôles plus nombreux. L'algorithme de calcul des phases procède de la manière suivante. Tout d'abord, pour chaque fenêtre de synthèse, un nombre aléatoire G est aléatoirement choisi entre $-\pi g$ et πg. Ensuite, pour chaque sinusoïde, un nombre aléatoire R_i est uniformément tiré entre 0 et 1. Les phases Φ_i des sinusoïdes sont donc déterminées à partir de la relation :

$$\Phi_i = \frac{\pi}{2} + G + 2(R_i - \frac{1}{2})P\pi + \frac{\pi F_i T_s}{F_s} \tag{4.26}$$

La durée T_s permet de spécifier le prochain instant de synchronisation des phases. Elle est aléatoirement définie par la relation :

$$T_s = \tau - T + \texttt{rand}(2\mathrm{T}) \tag{4.27}$$

A chaque fois que cette durée est dépassée, elle est à nouveau aléatoirement déterminée.

Cette nouvelle approche, qui reste encore à l'étude, permet non seulement de définir des écarts entre les impulsions plus importantes que la durée d'une fenêtre de synthèse, mais également de traiter statistiquement ces écarts. Elle pourrait trouver un grand intérêt dans le traitement de sons bruités impulsifs, composés de nombreux impacts aléatoirement répartis dans le temps (sons de gouttes de pluie, écoulement d'eau, bris de verre, etc. . .).

4.4 Calcul des échantillons

Les méthodes de synthèse s'effectuent en deux étapes distinctes, même si la première étape de détermination des caractéristiques est parfois implicite, comme dans le cas de la synthèse de la partie stochastique du modèle SMS. Une fois que cette première étape est finie, c'est-à-dire que toutes les caractéristiques des sinusoïdes composant le signal sont déterminées, un algorithme permet de calculer les échantillons synthétisés à partir de ces sinusoïdes. Il existe pour cela plusieurs possibilités, réparties en deux familles, de nombreuses études ayant porté sur ce domaine. Nous les présentons assez rapidement, car les détails peuvent être trouvés dans la littérature (par exemple [Smi03] ou [Mar00]).

4.4.1 Calcul de sinusoïdes

Les algorithmes de calcul des sinusoïdes ne peuvent raisonnablement pas faire appel aux fonctions implémentées dans les différents langages, essentiellement pour des raisons d'efficacité. En effet, le calcul d'une telle fonction est très coûteuse. C'est pour cela que plusieurs techniques ont été proposées, que nous présentons dans cette sous-section.

Une sinusoïde à calculer est notée s et est définie par :

$$s[n] = a\cos(n\frac{2\pi f}{F_s} + \phi_0) = a\cos(n\frac{\omega}{F_s} + \phi_0) \tag{4.28}$$

où f est sa fréquence, et ϕ_0 sa phase à l'origine. Nous notons $\Omega_s = \frac{\omega}{F_s}$.

Tables d'échantillons

La première méthode est une méthode de base calculant préalablement à l'exécution des algorithmes de synthèse de sinusoïdes, les échantillons relatifs aux différentes fréquences possibles, sauvegardés dans des tableaux en mémoire. L'accès à ces valeurs s'effectue par rapport aux indices des tableaux, qui peuvent être associés à une échelle de temps. Bien évidemment, cette technique impose une grande consommation de mémoire pour la sauvegarde de ces nombreuses valeurs. La précision de cette technique est liée à la mémoire utilisée (le nombre d'échantillons précalculés doit être important) et/ou à la rapidité d'exécution des méthodes d'interpolation entre les échantillons précalculés.

Cette approche était plus intéressante dans le passé, avec des machines dont le temps de calcul était limité, mais avec des accès mémoire suffisamment rapides. Dorénavant, avec les machines actuelles, la précision de la synthèse peut être meilleure en utilisant des méthodes de calcul des échantillons, ce qui privilégie la taille de la mémoire nécessaire.

Forme couplée

La première méthode appelée forme couplée (*coupled form*) repose sur la prise en considération d'un nombre complexe dont la partie réelle est $x[n]$, noté S_n :

$$S_n = x[n] + i.a\sin(n\Omega_s + \phi_0) = x[n] + i.y[n] = a\exp(in\Omega_s + \phi_0) \tag{4.29}$$

Au temps $n+1$, ce nombre complexe peut s'écrire sous la forme :

$$S_{n+1} = a\exp(in\Omega_s + \phi_0).\exp(i\Omega_s) = S_n.\exp(i\Omega_s) \tag{4.30}$$

Cette relation constitue la base de la technique présentée. Nous notons $c[n] = \cos(\Omega_s)$ et $s[n] = \sin(\Omega_s)$. La sinusoïde est définie alors par :

$$\forall n > 1, \quad \left\{ \begin{array}{lcl} x[n] & = & c[n].x[n-1] - s[n].y[n-1] \\ y[n] & = & s[n].x[n-1] + c[n].y[n-1] \end{array} \right. \tag{4.31}$$

avec $x[0] = \cos(\phi_0)$ et $y[0] = \sin(\phi_0)$.

Ainsi, le calcul d'un échantillon de la sinusoïde x nécessite 4 multiplications et 2 additions. Mais d'autres algorithmes permettent d'effectuer ce même calcul avec moins d'opérations.

Cercle magique

L'algorithme du cercle magique (*magic circle*) ne nécessite seulement que 2 multiplications et 2 additions pour chaque échantillon. Il repose sur la propriété selon laquelle $\sin(2x) = 2\sin(x)\cos(x)$. Nous définissons la valeur $\epsilon = 2\sin(\frac{\pi f}{F_s})$. La sinusoïde x que nous voulons calculer peut alors s'écrire :

$$\forall n > 1, \quad \left\{ \begin{array}{rcl} x[n] & = & x[n-1] + \epsilon.y[n-1] \\ y[n] & = & -\epsilon.x[n-1] + y[n-1] \end{array} \right. \tag{4.32}$$

Résonateur numérique

L'algorithme réclamant le moins d'opérations (1 multiplication et 1 addition) est celui du résonateur numérique (*digital resonator*) qui repose sur la propriété selon laquelle :

$$\forall n > 1, \quad x[n+1] = \varepsilon.x[n] - x[n-1] \tag{4.33}$$

où $\varepsilon = 2\cos(\omega_s)$.

Cet algorithme a été implémenté dans le logiciel Respect [MS99], disponible librement [Mar]. Les expériences ont montré qu'il était possible de synthétiser ainsi à peu près 2 sinusoïdes par MHz de vitesse du microprocesseur Pentium II [Mar00]. L'application de cet algorithme nous assure avec les machines actuelles la possibilité de pouvoir synthétiser suffisamment de sinusoïdes pour pouvoir générer un son bruité.

4.4.2 Synthèse par transformée de Fourier inverse

Une méthode de synthèse très différente des précédentes, présentée dans [RD92b, RD92a], consiste à synthétiser les échantillons du signal souhaité en définissant d'abord son spectre, puis en appliquant la transformée inverse de Fourier. Elle est appelée méthode FFT^{-1}. Cette approche est justifiée par la rapidité des algorithmes mis en oeuvre pour la transformée de Fourier (FFT), mais également par la définition des sons en prenant compte de leur enveloppe spectrale. Cette dernière remarque est particulièrement valable pour les parties bruitées rajoutées par les modèles hybrides.

Reconstruction du spectre

La première étape de cet algorithme consiste à définir le spectre complexe du signal à synthétiser. En fait, de par la définition de la transformée de Fourier, seulement la moitié du spectre doit être reconstruite, l'autre moitié symétrique étant définie par conjugaison. Chaque sinusoïde apporte une contribution au spectre. Cette contribution est idéalement une valeur d'amplitude à additionner à la fréquence associée. Toutefois, nous avons vu dans la section 3.4 qu'en réalité, le spectre d'une sinusoïde de fréquence et d'amplitude fixe est composé de plusieurs pics d'amplitude, dont la localisation et l'amplitude dépendent de la forme de la fenêtre d'analyse utilisée. Ainsi, la contribution d'une sinusoïde à un spectre peut être approximée, en ne tenant compte que des valeurs significatives, à 12 valeurs pour une fenêtre de Bartlett, et à 6 valeurs pour une fenêtre de Hann [Mar00].

Chevauchement des fenêtres

Les fenêtres de signal produites peuvent éventuellement se chevaucher. Les fenêtres de Hann et de Bartlett, qui peuvent être utilisées pour l'analyse, permettent l'application de cette technique de recouvrement (voir la sous-section 4.5.2). Néanmoins, il est également possible d'employer d'autres fenêtres pour l'analyse, qui garantissent un nombre peu élevé de valeurs du spectre significatives pour une sinusoïde, tout en effectuant un chevauchement des fenêtres de synthèse en utilisant un autre type de fenêtre de pondération [RD92b], sachant que les effets des deux fenêtres sont produits sur le domaine temporel pour l'une (fenêtre de pondération) et le domaine fréquentiel pour l'autre (fenêtre d'analyse).

Ajustement des phases

Les phases des sinusoïdes entre les différentes fenêtres successives de synthèse doivent être synchronisées de façon à ne pas provoquer de modulation d'amplitude (en cas de chevauchement) ou d'artefacts (sans chevauchement). Pour cela, les phases sont calculées le plus précisément possible et le spectre complexe du signal à synthétiser est généré à partir de ces valeurs.

Transformée inverse de Fourier

La transformée inverse de Fourier peut être effectuée à l'aide de l'algorithme rapide de la transformée de Fourier. Pour une fenêtre de synthèse de durée W_s échantillons, la complexité de cet algorithme est en $W_s \log(W_s)$. La bibliothèque FFTW (*Fastest Fourier Transform in the West*) [FJ98] est une implémentation libre de cet algorithme.

4.4.3 Conclusions

Les deux types de méthode (calcul direct des sinusoïdes ou calcul par transformée inverse de Fourier) peuvent être comparés sur le plan de la qualité et sur le plan de l'efficacité de la synthèse. Du point de vue de la qualité, les méthodes sont équivalentes lorsque les sinusoïdes sont d'amplitudes et de fréquences constantes durant le temps correspondant à la longueur de la fenêtre de synthèse. Par contre, dès qu'elles ne sont plus stables, la technique directe de calcul des sinusoïdes s'avère plus précise car elle peut être adaptée pour prendre en compte ces variations [MS99]. Pour les modèles sinusoïdaux, elle doit être préférée.

Ce choix est également confirmé lorsque nous étudions l'efficacité des méthodes. Une discussion pour les sons harmoniques ou quasi-harmoniques est donnée dans [Mar00]. Elle préconise l'utilisation du calcul direct des sinusoïdes, argumentant le fait que le nombre de sinusoïdes est constant et grand pour la méthode FFT^{-1}, alors que les sons harmoniques nécessitent rarement (fréquence très grave) autant de partiels.

En ce qui concerne les sons bruités, ce choix apparaît plus discutable, puisque le nombre de sinusoïdes requis est très important. Ainsi, les méthodes de synthèse des parties bruitées des modèles spectraux hybrides sont basées sur cette méthode de synthèse FFT^{-1}, en reconstituant le spectre à partir d'une approximation de l'enveloppe spectrale et d'un tirage aléatoire des phases. Néanmoins, nous montrons dans la section suivante que l'application de la technique de recouvrement à la synthèse de sons bruités supposés aléatoires impose des méthodes adaptées, dont la plus précise implique un décalage temporel. Cette opération semble complexe à réaliser dans le domaine spectral, même si des études restent à mener (voir la section 5.2.2).

4.5 Chevauchement des fenêtres de synthèse

Les méthodes de synthèse liées aux modèles spectraux, comme par exemple le modèle CNSS que nous avons développé, s'appliquent sur des successions de fenêtres de signal, qui peuvent éventuellement se chevaucher. Elles commencent par déterminer les caractéristiques des sinusoïdes composant le signal avant de calculer les échantillons produits par ces sinusoïdes. Nous étudions dans cette section les conséquences sur la qualité de la synthèse de cette considération du signal par des petites fenêtres consécutives, et les algorithmes envisageables pour corriger les défauts produits.

4.5.1 Succession de fenêtres

Les expériences que nous avons menées pour mettre en application la synthèse de sons à partir du modèle CNSS ont mis en avant des artefacts audibles, dus à la succession de fenêtres. Nous expliquons brièvement ces défauts et les solutions proposées.

Fenêtres rectangulaires

Dans le cas où les fenêtres de signaux se succédant sont rectangulaires (signal multiplié par une fenêtre de pondération w telle que $\forall n \in [\![0; W_s - 1]\!], w[n] = 1$, et $w[n] = 0$ sinon), quelques artefacts peuvent être perçus. La périodicité de la succession des fenêtres, ainsi que les discontinuités du signal sont perçues par le système auditif. De nombreux clics peuvent aussi être entendus dans le cas de sons synthétisés avec un très petit nombre de sinusoïdes (sons harmoniques ou quasi-harmoniques de hautes fréquences fondamentales).

Fenêtres de tailles aléatoires

Pour éliminer ces artefacts, nous avons proposé de rendre aléatoire la taille de la fenêtre de synthèse pour empêcher la périodicité due à leur succession régulière. Nos expériences utilisant cette approche permettent de synthétiser des sons sans la moindre fréquence audible indésirable. Mais les clics restent toujours présents à cause des transitions discontinues entre les fenêtres temporelles successives.

Technique de recouvrement

Nous avons alors envisagé l'application de la technique de recouvrement, définie ci-dessous : les fenêtres de signaux synthétisés se chevauchent et s'additionnent après avoir été multipliées par des fenêtres de pondération (Hann, Bartlett, ...). Les clics disparaissent alors. Mais, quel que soit le type de la fenêtre, de nouveaux artefacts sont entendus lorsque cette technique est appliquée à des sons bruités. Ces distortions sont décrites et expliquées dans les sections suivantes.

4.5.2 Principe de la technique de recouvrement

La technique du chevauchement des fenêtres de signal x, appelée OLA (*OverLap and Add*), consiste à synthétiser des fenêtres notées s_p de W_s échantillons (voir la figure 4.6) qui se recouvrent en partie. Chaque fenêtre temporelle s_p est multipliée par une fenêtre de pondération notée w :

$$\left\{ \begin{array}{lcl} x_p[n] = & s_p[n]w[n] & n \in [\![0; W_s - 1]\!] \\ x_p[n] = & 0 & \forall n < 0, \quad \forall n \geq W_s \end{array} \right. \tag{4.34}$$

Cette fenêtre doit satisfaire la condition selon laquelle la somme de tous les poids affectés aux différentes fenêtres se chevauchant, doit être égale à 1 :

$$\forall n, \quad \sum_p w[n - pH] = 1 \tag{4.35}$$

où H représente le nombre d'échantillons d'avancement et p est l'indice de la fenêtre. Par exemple, pour une valeur de $H = 0.25W_s$, les fenêtres successives se chevauchent de $\frac{3}{4}$, et donc un échantillon est défini par l'addition de 4 fenêtres. Lorsque la condition donnée par l'équation (4.35) n'est pas vérifiée, le signal de sortie voit son amplitude modulée par un signal périodique.En notant P le nombre de fenêtres, nous pouvons écrire :

$$\forall n, \quad x[n] = \sum_{p=0}^{P-1} x_p[n - pH] \tag{4.36}$$

Plusieurs types de fenêtres de pondération ont été définis. Toutefois, comme les méthodes d'analyse basées sur la transformée de Fourier multiplient déjà le signal original par une fenêtre particulière pour atténuer les distortions du spectre discret obtenu (voir la section 3.4), l'utilisation de ces mêmes fenêtres pour la synthèse facilite le processus d'analyse et de synthèse complet et permet d'envisager une reproduction plus fidèle du son original. Cependant, ces fenêtres de pondération doivent vérifier la condition donnée par l'équation (4.35).

Fenêtre triangulaire

La fenêtre la plus simple mathématiquement est la fenêtre dite de Bartlett, ou fenêtre triangulaire. Comme son nom l'indique, elle est constituée de deux parties linéaires, et présente un maximum en son centre. Elle peut s'assimiler à des fondus enchaînés (*fade-in* puis *fade-out*). La figure 4.7 présente cette fenêtre. Sa définition mathématique est donnée par les relations :

$$\left\{ \begin{array}{lll} w(n) & = \frac{2n}{W_s-1} & , \forall n \in [\![0; \frac{W_s}{2} - 1]\!] \\ w(n) & = 2 - \frac{2n}{W_s-1}0 & , \forall n \in [\![\frac{W_s}{2}; W_s - 1]\!] \end{array} \right. \tag{4.37}$$

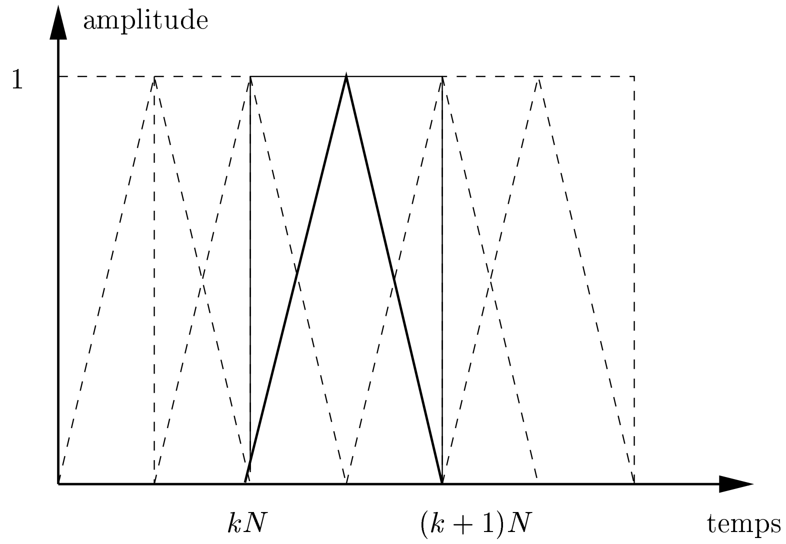

FIG. 4.6 – illustration de la méthode de recouvrement des fenêtres (OLA) avec ici la fenêtre de Bartlett ($H = \frac{N}{2}$).

En faisant abstraction du fenêtrage des signaux lors de l'étude par transformée de Fourier, cette fenêtre est la plus couramment employée dans le cadre de la technique de recouvrement puisqu'elle vérifie trivialement la condition imposée (équation (4.35)).

Fenêtre de Hann

La fenêtre de Hann est utilisée lors de l'analyse par la transformée de Fourier car elle présente des propriétés intéressantes par rapport à son spectre. La figure 4.7 montre cette fenêtre, définie par la relation :

$$\forall n \in [\![0; W_s - 1]\!], \quad w_{\text{hann}}(n) = \frac{1}{2}(1 - \cos(\frac{2\pi n}{W_s - 1})) \tag{4.38}$$

Il est intéressant de noter qu'elle peut parfaitement être employée pour des techniques de recouvrement, puisque la condition de l'équation (4.35) est vérifiée en effectuant une très légère approximation, pour $H = \frac{W_s}{2}$:

$$\begin{aligned}
\forall n, \quad w_{\text{hann}}(n) + w_{\text{hann}}(n + \frac{W_s}{2}) &= \frac{1}{2}(1 - \cos(\frac{2\pi n}{W_s - 1})) + \frac{1}{2}(1 - \cos(\frac{2\pi(n + \frac{W_s}{2})}{W_s - 1})) \\
w_{\text{hann}}(n) + w_{\text{hann}}(n + \frac{W_s}{2}) &= \frac{1}{2}(1 - \cos(\frac{2\pi n}{W_s - 1})) + \frac{1}{2}(1 - \cos(\frac{2\pi n}{W_s - 1} + \pi\frac{W_s}{W_s - 1}))
\end{aligned}$$

En approximant $\frac{W_s}{W_s - 1}$ à 1, nous pouvons écrire :

$$\begin{aligned}
w_{\text{hann}}(n) + w_{\text{hann}}(n + \frac{W_s}{2}) &= \frac{1}{2}(1 - \cos(\frac{2\pi n}{W_s - 1})) + \frac{1}{2}(1 - \cos(\frac{2\pi n}{W_s - 1} + \pi)) \\
w_{\text{hann}}(n) + w_{\text{hann}}(n + \frac{W_s}{2}) &= \frac{1}{2}(1 - \cos(\frac{2\pi n}{W_s - 1})) + \frac{1}{2}(1 + \cos(\frac{2\pi n}{W_s - 1})) \\
w_{\text{hann}}(n) + w_{\text{hann}}(n + \frac{W_s}{2}) &= 1 \qquad (4.39)
\end{aligned}$$

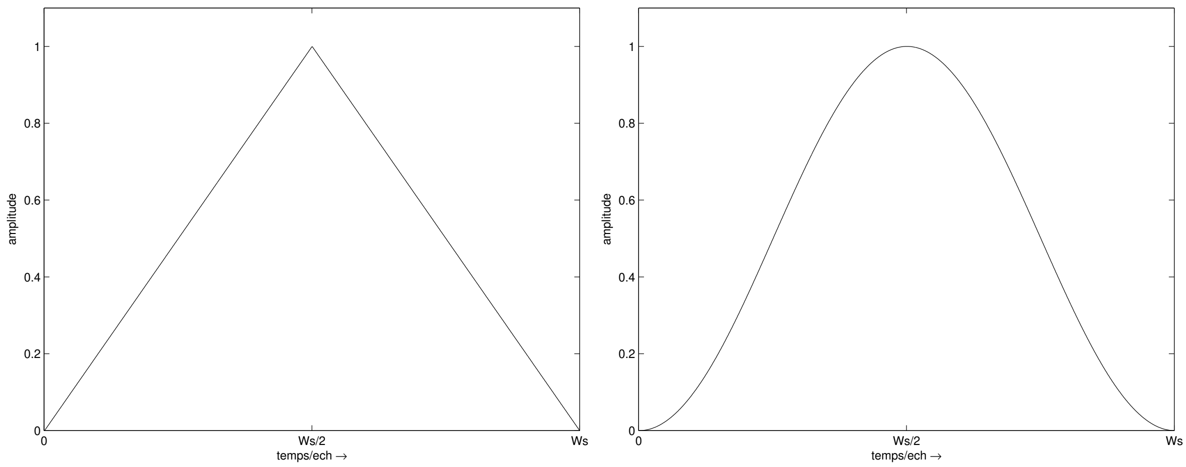

FIG. 4.7 – Représentation temporelle des fenêtres de Bartlett et de Hann utilisées pour la méthode du recouvrement (OLA).

Fenêtre de Hamming

La fenêtre de Hamming est également une fenêtre couramment utilisée lors d'analyse. Elle est définie par la relation :

$$\forall n \in [\![0; W_s - 1]\!], \quad w_{\text{hamming}}(n) = 0.54 - 0.46\cos(\frac{2\pi n}{W_s - 1})) \tag{4.40}$$

Cette définition montre une certaine ressemblance avec la fenêtre de Hann. Nous pouvons montrer, comme nous l'avons fait pour cette fenêtre, que pour un avancement de $H = \frac{W_s}{2}$, la somme des fenêtres de Hamming se chevauchant est constante :

$$w_{\text{hamming}}(n) + w_{\text{hamming}}(n + \frac{W_s}{2}) = 1.08 \tag{4.41}$$

Cette somme ne vaut pas 1, contrairement à la condition imposée pour une fenêtre de pondération. Néanmoins, cette fenêtre peut être utilisée. La définition donnée par l'équation (4.35) peut en effet être généralisée par l'équation :

$$\sum_{l=0}^{L-1} w(n - pH) = K_1 \tag{4.42}$$

où K_1 représente une valeur constante. La différence entre les deux définitions provient uniquement du fait que les signaux produits par recouvrement seront très légèrement amplifiés, car leur amplitude sera multipliée par la constante K_1.

4.5.3 Application au modèle CNSS

La technique de recouvrement peut être appliquée à la méthode de synthèse liée au modèle CNSS. Les spécificités de ce modèle imposent d'autres conditions à respecter. La notion de densité spectrale est liée au nombre de sinusoïdes composant le signal. Il est donc nécessaire de s'assurer qu'à chaque instant, le nombre de sinusoïdes n'est pas modifié par l'application de cette technique de recouvrement. Pour cela, la seule possibilité est de considérer un temps d'avancement constant, égal à la moitié du signal :

$$H_{\mathrm{CNSS}} = 0.5W_s \tag{4.43}$$

Bien évidemment une autre solution consisterait à ne considérer aucun recouvrement ($H = W_s$), mais cette approche conduirait à des artefacts dus aux transitions entre les fenêtres.

Lorsque la technique de recouvrement est employée, nous pouvons envisager deux cas de figures, selon que la synthèse soit effectuée ou non après une modification du son original.

Synthèse après analyse sans modification

Les méthodes d'analyse de sons extraient des paramètres mathématiques à partir desquels le son est synthétisé. Pour cela, elles utilisent la technique de recouvrement. Comme les sons bruités sont considérés par les modèles adaptés comme des signaux aléatoires, cela implique que les fenêtres successives de ces sons ne peuvent être considérées comme indépendantes, comme l'illustre la figure 4.8. Nous allons voir plus loin que c'est cette propriété qui explique que, lorsqu'aucune modification n'est effectuée, les expériences montrent que le signal synthétisé ne présente aucun artefact audible. Bien évidemment, une synthèse directe après une analyse, sans modification (changement d'échelle temporelle, transposition, ...), ne présente que très peu d'intérêt.

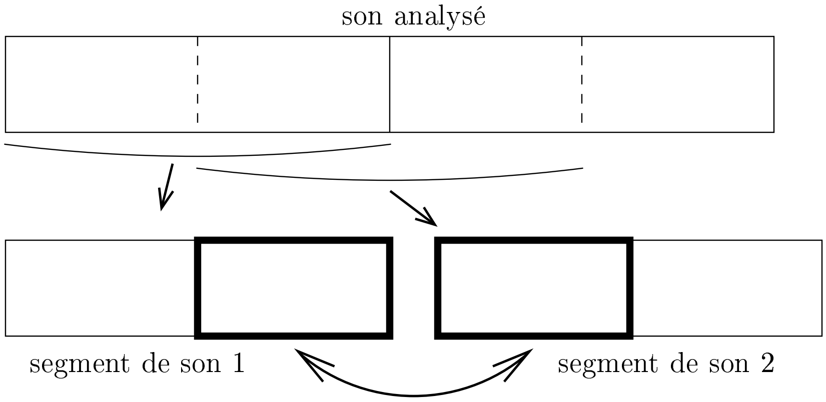

FIG. 4.8 – Étape d'analyse : deux fenêtres successives du son ne sont pas indépendantes.

.

Synthèse après transformation du signal

Par contre, lorsque la technique de recouvrement est appliquée, et qu'une transformation du son analysée est effectuée, les expériences que nous avons menées, mettent en évidence des artefacts audibles, liés à des variations périodiques d'intensité. Ces artefacts sont confirmés par des expériences effectuées à partir de deux bruits blancs synthétiques, générés par tirage aléatoire des échantillons selon une loi gaussienne. Ces deux bruits sont donc indépendants. Les bruits blancs résultant sont déduits des deux bruits précédents par addition, pour des pondérations différentes. Le tableau 4.4 met en évidence les variations de variances, selon les pondérations utilisées.

(α, β)	$V(s_1)$	$V(s_2)$	$V(\alpha s_1 + \beta s_2)$
$(0.5, 0.5)$	8.312	8.361	4.160
$(0.4, 0.6)$	8.297	8.349	4.342
$(0.3, 0.7)$	8.304	8.340	4.812
$(0.2, 0.8)$	8.350	8.330	5.677
$(0.1, 0.9)$	8.337	8.411	6.899
$(0.0, 1.0)$	8.323	8.341	8.341

TAB. 4.4 – Variations de la variance du signal résultant de l'addition de deux signaux aléatoires indépendants, produits par distribution gaussienne des échantillons (fenêtre de Bartlett, 100000 réalisations). Les coefficients (α, β) correspondent ici à la pondération imposée par une fenêtre de Bartlett.

Ainsi ces expériences font apparaître le fait que si la technique de recouvrement est appliquée après avoir effectué une transformation du signal analysé, les sons aléatoires synthétisés présentent, quelle que soit la fenêtre de pondération utilisée (Bartlett, Hann), des propriétés statistiques qui varient au cours du temps, avec notamment une variance non constante.

L'exemple de la représentation temporelle d'un bruit blanc synthétisé par recouvrement de plusieurs bruits blancs (fenêtre de Bartlett) est montré par la figure 4.9. Cet exemple montre les variations périodiques d'intensité. Ces effets indésirables sont parfaitement expliqués par les propriétés statistiques des sons.

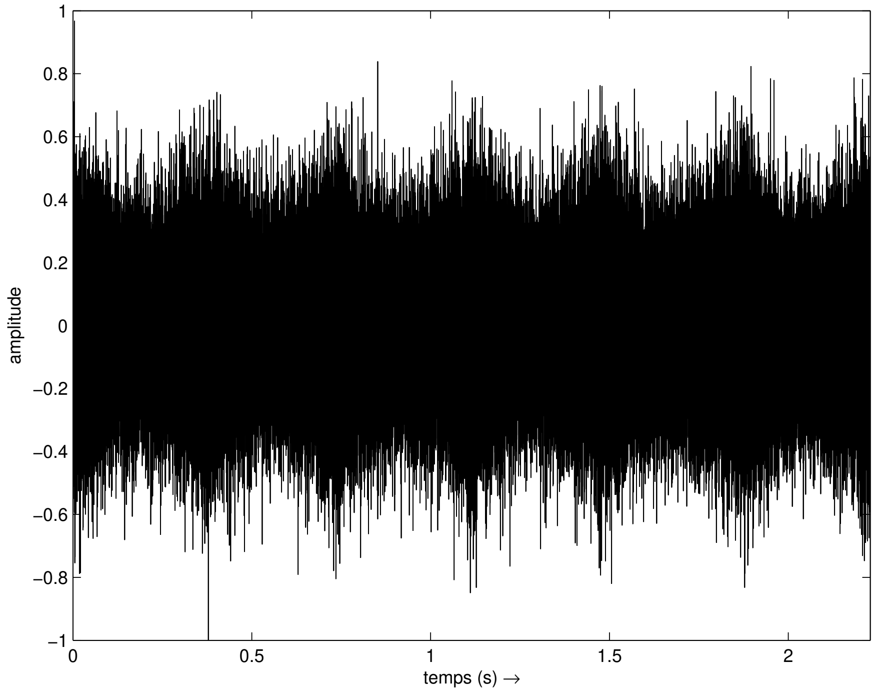

FIG. 4.9 – Exemple de bruit blanc synthétisé en appliquant la technique de recouvrement (fenêtre de Bartlett) : les variations d'intensité induites par cette technique sont mises en évidence.

4.5.4 Technique de recouvrement pour des signaux aléatoires

Nous avons vu dans la description de la technique de recouvrement que deux fenêtres temporelles étaient multipliées par une fenêtre de pondération avant d'être ajoutées. Quel que soit le type de la fenêtre, elle doit satisfaire l'équation (4.35). En notant x' le signal synthétisé, nous pouvons écrire à partir de l'équation (4.36) :

$$\forall n, \quad x'[n] = \sum_{p=0}^{P-1} x_p(n - pH) \tag{4.44}$$

et l'équation (4.34) nous amène à :

$$\forall n, \quad x'[n] = \sum_{p=0}^{P-1} s_p(n - pH)w(n - pH) \tag{4.45}$$

Dans le cas du bruit blanc, les signaux s_p ($p \in [\![0; P-1]\!]$) sont des réalisations de la même variable aléatoire X. Ils présentent les mêmes propriétés statistiques, et le signal résultant du processus de chevauchement et d'addition doit présenter les mêmes propriétés. Nous notons X' la variable aléatoire associée au signal $x'[n]$. La moyenne de X' reste constante mais le moment du second ordre varie. En notant $V(X')$ la variance de X', nous pouvons écrire :

$$V(X') = V(\sum_{p=0}^{P-1} X_p(n - pH)w(n - pH)) \tag{4.46}$$

Si le signal original a été analysé, mais qu'aucune transformation n'a été effectuée, les signaux X_p ne sont pas indépendants (voir la sous-section précédente) et nous vérifions $V(X') = V(X)$. Nous considérons ici le cas où les signaux s_p sont indépendants, c'est-à-dire qu'aucune analyse n'a été faite, ou qu'une modification du signal analysée a été effectuée. Il vient alors :

$$V(X') = \sum_{p=0}^{P-1} V(X_p(n - pH)w(n - pH)) \tag{4.47}$$

La définition de la fenêtre w implique que $w(n - pH)$ n'est pas une variable aléatoire. Nous pouvons donc écrire :

$$V(X') = \sum_{p=0}^{P-1} w^2(n - pH)V(X_p(n - pH)) \tag{4.48}$$

Comme les signaux s_p sont des réalisations de la même variable aléatoire X, nous pouvons écrire :

$$V(X') = V(X) \sum_{p=0}^{P-1} w^2(n - pH) \tag{4.49}$$

La définition de la fenêtre de pondération $w(n)$ est $\sum_p w(n - pH) = 1$, ce qui amène à $\sum_p w^2(n - pH) = f(n)$. Cette somme n'est pas constante. C'est pour cette raison que la variance $V(X')$ varie dans le temps. De plus, cette variation dépend du temps d'avancement et est donc périodique.

En supposant que $V(X')$ et $V(X)$ sont égales, ce calcul mène à la condition sur la fenêtre w :

$$\forall n, \quad \sum_{p=0}^{P-1} w^2(n - pH) = 1 \tag{4.50}$$

ou, plus généralement :

$$\forall n, \quad \sum_{p=0}^{P-1} w^2(n - pH) = K_2 \tag{4.51}$$

où K_2 est une constante.

Par exemple, dans le cas d'un taux de recouvrement de 50% ($H = \frac{W_s}{2}$) et lorsque la fenêtre de pondération w est triangulaire (Bartlett, équation (4.37)), l'équation (4.49) devient :

$$V(X') = V(X)(1 - \frac{4n}{W_s - 1} + \frac{8n^2}{(W_s - 1)^2}) \tag{4.52}$$

où $n \in [\![kW_s; kW_s + \frac{W_s-1}{2}]\!]$, pour tous les entiers k.

La figure 4.10 montre les variations de la variance en fonction du temps avec une fenêtre de longueur $W_s = 16384$ échantillons et $H = \frac{W_s}{2}$. Les variations dans le cas (b) correspondent parfaitement aux fluctuations d'intensité de la figure 4.9.

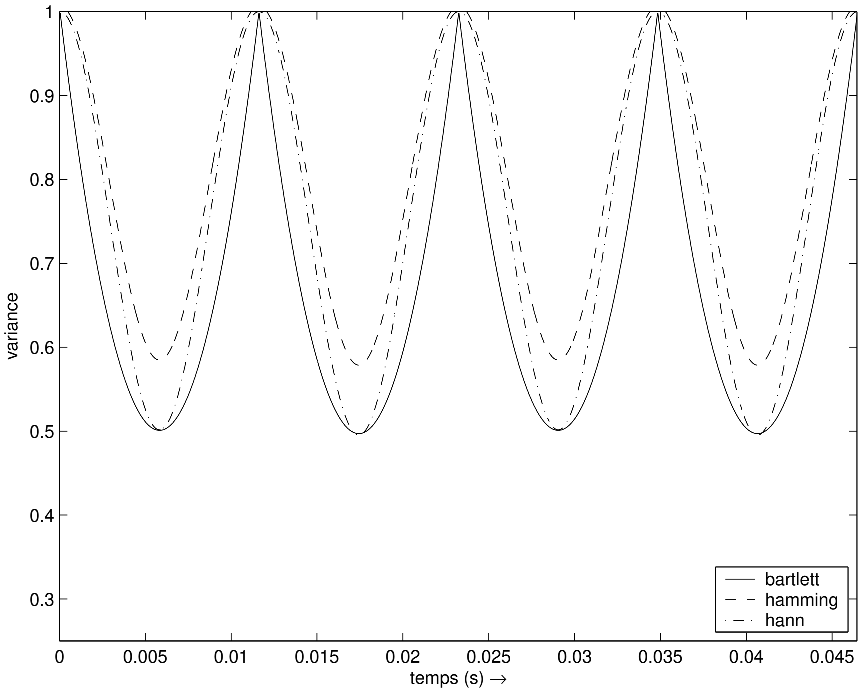

FIG. 4.10 – Variations de la variance du signal pour différentes fenêtres de pondération (fenêtres de 1024 échantillons, temps d'avancement de $H = \frac{W_s}{2}$).

Ces variations de variance expliquent les fluctuations régulières d'intensité perçues lors des expériences que nous avons menées, et mises en évidence sur la figure 4.9. En effet, comme nous l'avons montré dans la section 1.5.1, l'intensité perçue est liée à l'amplitude RMS. Or dans le cas d'un son bruité, supposé aléatoire, cette amplitude RMS vaut la racine carrée de la variance, c'est-à-dire l'écart-type (voir l'équation (1.87)). Ainsi, ces variations de variance induisent des variations d'intensité perçue.

4.5.5 Solutions envisagées

Nous avons montré que l'utilisation de la technique de recouvrement lors de la synthèse de sons transformés après avoir été analysés, implique des artefacts audibles dus à la variation périodique d'intensité. Nous avons proposé dans [HDC02a] des méthodes permettant d'appliquer un recouvrement des fenêtres de signaux synthétisés. Nous détaillons dans cette sous-section deux d'entre elles, qui présentent des limites.

Compensation de l'amplitude

La méthode la plus simple pour empêcher de produire les artefacts dus au variations périodiques d'intensité est de compenser ces variations en multipliant périodiquement le signal par de nouvelles fenêtres de pondération inversement proportionnelles aux variations de variance introduites par l'application de la technique de recouvrement. Pour cela, nous savons que pour tout réel k, $V(kX) = k^2V(X)$. Dans le cas d'une fenêtre de pondération w de Bartlett, la fenêtre de compensation d'amplitude à appliquer, notée w_2, est définie par :

$$\forall n, \quad w_2[n] = \frac{1}{\sqrt{1 - 2w[n] + 2w[n]^2}} \tag{4.53}$$

Les inconvénients de cette méthode apparaissent lors de la synthèse de sons composés d'un très petit nombre de sinusoïdes (sons bruités de faible densité spectrale). Dans ce cas, les variations d'intensité sont perçues par l'oreille comme des variations d'amplitude propres à chaque sinusoïde. Ces variations sont représentées par la figure 4.10. Ainsi, une méthode de synthèse homogène, représentant un son par une somme de sinusoïdes, ne peut être envisagée, puisqu'elle nécessite alors une fenêtre de pondération supplémentaire pour les parties du son correspondant au bruit.

Existence d'une fenêtre satisfaisante

L'une des premières idées pour empêcher les artefacts audibles introduits par l'application de la technique de recouvrement est de définir une fenêtre de pondération valable pour les sons non bruités et pour les sons bruités. Les deux conditions que doivent remplir une telle fenêtre de pondération sont décrites par les équations (4.35) et (4.50). Ces deux équations peuvent être généralisées à des égalités à des constantes, comme l'indiquent les équations (4.51) et (4.42). Nous pouvons montrer qu'il est impossible de définir une fenêtre pouvant vérifier ces deux équations, même si chaque constante de chaque équation (notée K_1 et K_2) est différente l'une de l'autre ($K_1 \neq K_2$).

Pour simplifier la démonstration, nous nous plaçons ici dans le cas où $H = \frac{W_s}{2}$ et où K_1, K_2 et W_s sont fixés.

$$\forall n \in [\![0; \frac{W_s}{2}]\!], \quad \begin{cases} w(n) + w(n + \frac{W_s}{2}) = K_1 \\ w^2(n) + w^2(n + \frac{W_s}{2}) = K_2 \end{cases} \tag{4.54}$$

Nous notons $\alpha_n = w(n)$ et $\beta_n = w(n + \frac{W_s}{2})$.

$$\begin{cases} \alpha_n + \beta_n = K_1 \\ \alpha_n^2 + \beta_n^2 = K_2 \end{cases} \tag{4.55}$$

Ce système est équivalent au système suivant :

$$\begin{cases} \alpha_n + \beta_n = K_1 \\ 2\beta_n^2 - 2K_1\beta_n + (K_1^2 - K_2) = 0 \end{cases} \tag{4.56}$$

Dans le meilleur des cas ($\Delta = -4K_1^2 + 8K_2 > 0$), la deuxième équation du système n'a que **deux** solutions. Cela implique que le système n'a, au maximum, que **deux** solutions, et donc cela montre bien qu'il est impossible de trouver une série de nombres (α_n, β_n) pour des valeurs de K_1 et K_2 fixées, de façon à définir une fenêtre.

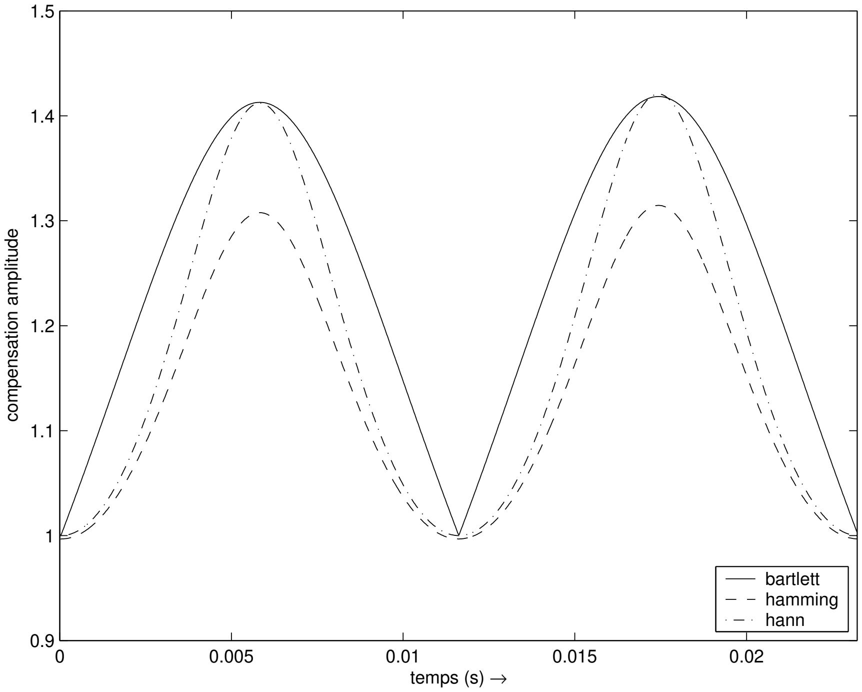

FIG. 4.11 – Compensation d'amplitude nécessaire pour différentes fenêtres de pondération (fenêtres de 1024 échantillons, temps d'avancement de $H = \frac{W_s}{2}$).

4.5.6 Fenêtre sinusoïdale

La première méthode, proposée également dans [HDC02a], permet l'application de la technique de recouvrement pour la synthèse des sons bruités, par l'utilisation d'une fenêtre de pondération spécifique.

Principe

La technique de recouvrement pour la synthèse de sons étant générés aléatoirement impose l'utilisation d'une fenêtre de pondération vérifiant l'équation (4.50). Les fenêtres couramment

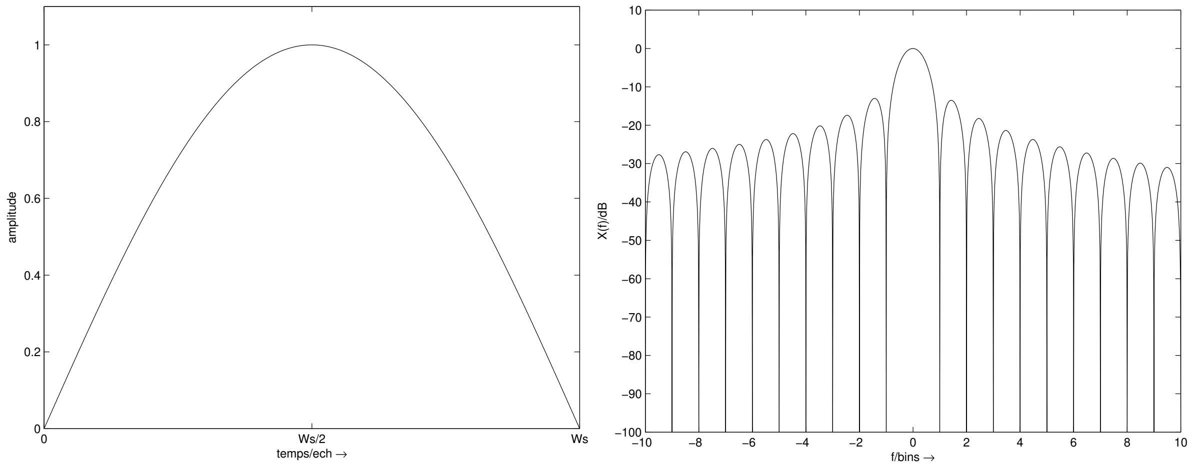

FIG. 4.12 – Représentation temporelle et spectrale de la fenêtre sinusoïdale proposée pour la technique de recouvrement des sons considérés comme aléatoire.

employées ne répondant pas à cette condition, nous avons envisagé la définition d'une fenêtre de pondération spécifique. Nous proposons une fenêtre de type sinusoïdale (de la famille de celle de Hann), notée w_{sinus}, définie par la relation :

$$\forall n \in [\![0; W_s - 1]\!], \quad w_{\text{sinus}}[n] = \sin(\frac{\pi n}{W_s - 1}) \tag{4.57}$$

où W_s est la taille de la fenêtre du signal synthétisé, strictement supérieure à 1. Cette fenêtre est représentée sur les domaines temporel et spectral, par la figure 4.12.

Nous pouvons vérifier que cette fenêtre vérifie la condition de l'équation (4.50), dans le cas où le temps d'avancement vaut $H = \frac{W_s}{2}$:

$$\begin{aligned} w^2_{\text{sinus}}[n] + w^2_{\text{sinus}}[n + \frac{W_s}{2}] &= \sin^2(\frac{\pi n}{W_s - 1}) + \sin^2(\frac{\pi n}{W_s - 1} + \frac{\pi}{2}) \\ &= \sin^2(\frac{\pi n}{W_s - 1}) + \cos^2(\frac{\pi n}{W_s - 1}) \\ &= 1 \end{aligned} \tag{4.58}$$

Algorithme

L'algorithme de synthèse en adaptant la technique de recouvrement est donc identique à l'algorithme employé habituellement avec les fenêtres de pondération classiques. Du point de vue de l'efficacité, ces algorithmes sont complètement équivalents.

Cette technique de recouvrement est de même parfaitement applicable aux méthodes de synthèse par calcul des sinusoïdes et par transformée inverse de Fourier. Il est juste nécessaire, dans ce dernier cas, de prendre en compte les caractéristiques spectrales de cette fenêtre dans le calcul des contributions d'une sinusoïde au spectre.

Limites

L'utilisation d'une fenêtre sinusoïdale permet d'empêcher l'apparition d'artefacts audibles pour la synthèse de sons bruités. Par contre, comme cette fenêtre de pondération ne vérifie pas l'équation (4.35), elle introduit des variations d'intensité lors de la synthèse de sons déterministes, comme par exemple des sons harmoniques issus d'instrument de musique. Nous pouvons exprimer ces variations périodiques d'amplitude :

$$\begin{aligned} \forall n, \quad w(n) + w(n + \frac{W_s}{2}) &= \sin(\frac{\pi n}{W_s - 1}) + \sin(\frac{\pi n}{W_s - 1} + \frac{\pi}{2}) \\ &= \sin(\frac{\pi n}{W_s - 1}) + \cos(\frac{\pi n}{W_s - 1}) \\ &= 2\sin(\frac{\pi n}{W_s - 1} + \frac{\pi}{4})\cos(\frac{\pi}{4}) \\ &= \sqrt{2}\sin(\frac{\pi n}{W_s - 1} + \frac{\pi}{4}) \end{aligned} \tag{4.59}$$

L'utilisation de cette fenêtre induit donc, dans le cas de la synthèse de sons composés d'un nombre plus faible de partiels, une modulation d'amplitude explicitée par l'équation (4.59) et représentée par la figure 4.13.

4.5.7 Décalage temporel des oscillateurs

La méthode précédente, consistant à utiliser comme fenêtre de pondération une fenêtre sinusoïdale, induit des artefacts audibles dans le cas de synthèse de sons peu bruités. Nous avons donc envisagé une autre technique pouvant être employée aussi bien avec les sons bruités qu'avec les sons déterministes (harmoniques par exemple). Celle-ci ne peut être utilisée que pour la synthèse par addition de sinusoïdes.

Principe

Les variations périodiques d'intensité sont dues à la fenêtre de pondération (de type Hann, Bartlett, Hamming, etc...). Cet effet ne peut pas être évité, mais, par contre, il peut être réparti régulièrement sur tout le signal. Pour cela, au lieu de multiplier le signal résultant de la somme des sinusoïdes par la fenêtre de pondération, nous considérons chaque sinusoïde indépendamment, avec, pour chacune, un instant de début différent, régulièrement réparti. Cette approche revient à considérer le signal à synthétiser comme une somme de signaux (des sinusoïdes), chacun étant synthétisé par recouvrement. Ainsi, chaque signal introduit une variation d'intensité du fait de la multiplication par la fenêtre de pondération. Mais comme les sinusoïdes sont décalées, les variations sont réparties, annulant la variation globale d'intensité du signal résultant. La figure 4.14 illustre cette méthode avec, pour cet exemple, un son composé de 3 sinusoïdes.

Nous pouvons formaliser cette approche. Le nombre de partiels est noté N. Les décalages introduits, notés d_m, sont compris entre 0 et H, soit la moitié de la taille de la fenêtre de synthèse pour $H = \frac{W_s}{2}$. Les sinusoïdes formant le signal, notées s_p^m, sont de phases constantes.

$$X'(n) = \sum_{p=0}^{P-1} \sum_{m=0}^{N-1} s_p^m(n - pH - d_m) w(n - pH - d_m) \tag{4.60}$$

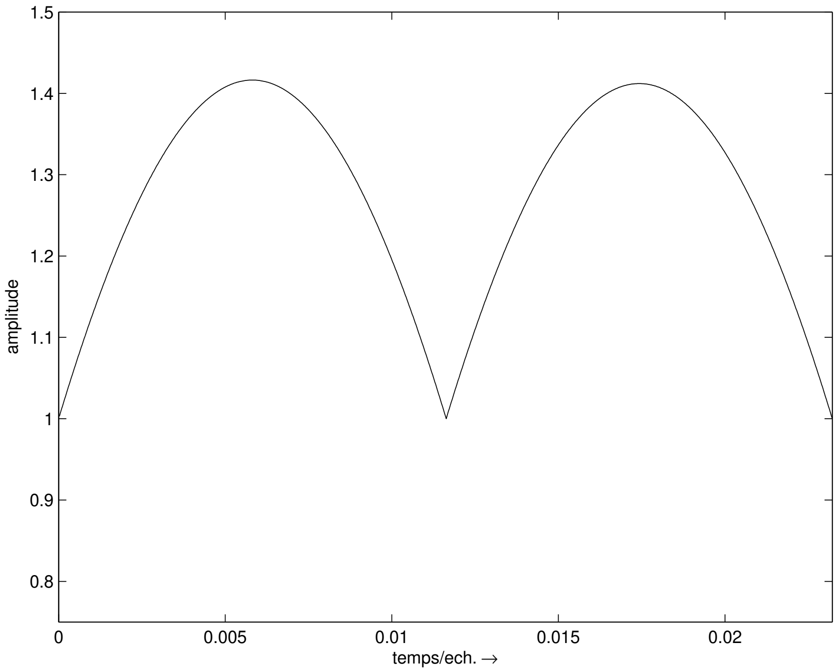

FIG. 4.13 – Modulation d'amplitude introduite par l'utilisation d'une fenêtre de pondération sinusoïdale dans le cas de sons non aléatoires (fenêtres de 1024 échantillons, temps d'avancement de $H = \frac{W_s}{2}$).

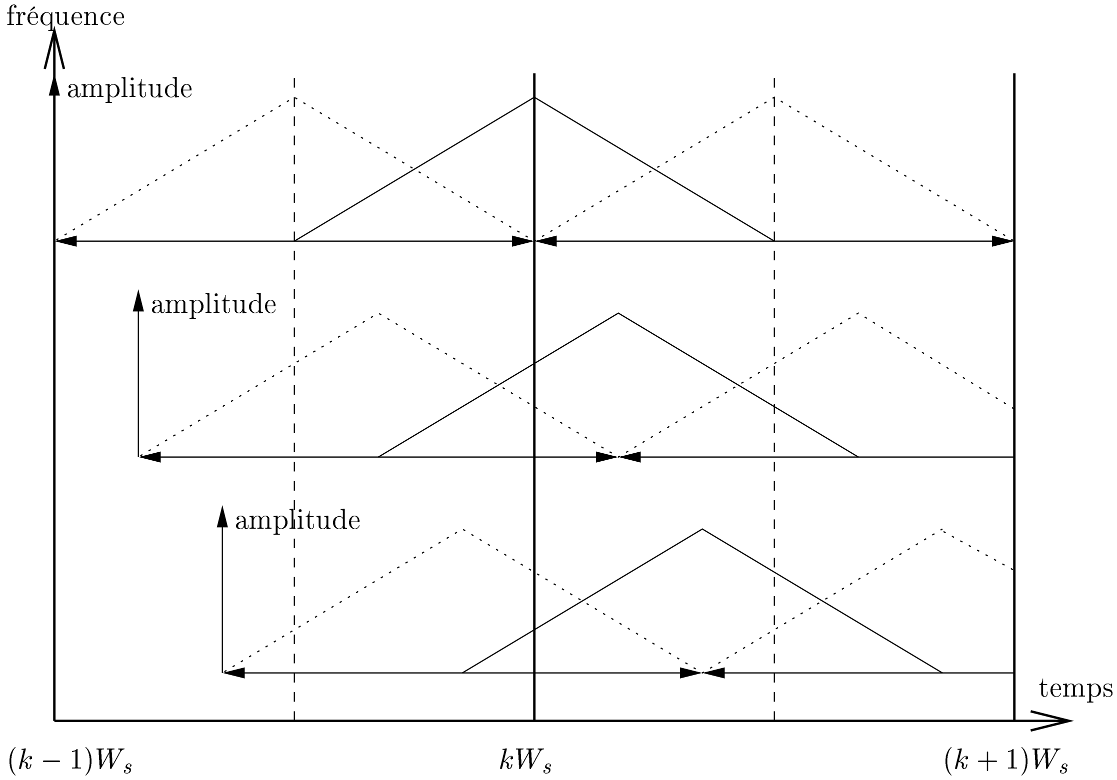

FIG. 4.14 – Exemple de décalage de 3 sinusoïdes pour l'algorithme de recouvrement.

En calculant la variance du signal ainsi défini, nous pouvons écrire :

$$V(X') = V(\sum_{m=0}^{N-1}\sum_{p=0}^{P-1} s_p^m(n-pH-d_m)w(n-pH-d_m)) \tag{4.61}$$

Comme nous savons que $\sum_p w(n-pH-d_m) = 1$ (voir la définition d'une fenêtre de pondération) :

$$V(X') = V(\sum_{m=0}^{N-1} t^m(n-d_m)) \tag{4.62}$$

où $t^m(n)$ est une sinusoïde. Nous savons qu'un bruit, c'est-à-dire un signal aléatoire, peut être défini mathématiquement par une somme de sinusoïdes. Le signal obtenu est de propriétés statistiques fixes, si le nombre de sinusoïdes le composant est constant. Cette condition est remplie si les écarts d_m sont régulièrement répartis sur la moitié de la fenêtre de synthèse (pour le cas d'un recouvrement de moitié).

Algorithme

Cette méthode induit un algorithme légèrement plus compliqué que les algorithmes de synthèse par recouvrement. Tout d'abord il faut définir les valeurs des décalages. Ces valeurs dépendent du nombre de sinusoïdes que contient le signal, et doivent être modifiées si celui-ci varie. Nous avons expérimenté plusieurs possibilités, notamment l'intérêt de définir ces écarts aléatoirement. Nous sommes arrivé à la conclusion que plusieurs solutions sont envisageables, mais une technique particulière pose des problèmes.

La première possibilité est de définir aléatoirement les décalages d_m entre 0 et H. En théorie, les écarts sont censés recouvrir complètement l'intervalle $[0; H]$. Toutefois, en pratique, il est possible que plusieurs sinusoïdes aient le même instant de début, ce qui est incompatible avec la théorie de la méthode. Cela reste néanmoins suffisamment rare pour ne pas provoquer de variations d'intensité assez fortes pour être perçues.

L'autre possibilité est de définir les décalages régulièrement répartis sur l'intervalle $[0; H]$. Les écarts entre les décalages d_m sont alors constants et valent $\delta d = \frac{H}{N}$:

$$\forall m \in [\![0; N-1]\!], \quad d_m = m.\delta d \tag{4.63}$$

La seule contrainte par rapport à cette méthode est imposée par les résultats de nos expériences. Il est apparu que lorsque les fréquences des sinusoïdes sont triées, les sons synthétisés laissent percevoir des artefacts. Nous n'avons pas trouvé d'explications rigoureuses, mais nous en avons tenu compte dans nos implémentations, pour la méthode de synthèse du modèle CNSS. Pour cela, nous utilisons les algorithmes de détermination des fréquences (section 4.3) qui choisissent les fréquences dans un ordre non croissant, du fait de l'application de l'algorithme de tirage aléatoire des casiers. Sinon, si les fréquences sont triées, il est nécessaire de définir les N décalages de façon aléatoire, à partir des N valeurs possibles données par l'équation (4.63).

Une fois que les décalages ont été déterminées, les sinusoïdes doivent être calculées séparément et multipliées chacune par la même fenêtre de pondération. L'algorithme suivant récapitule les étapes de cette méthode :

Entrée : Liste de N sinusoïdes, tableau d'échantillons de sortie
début
  pour i *à* N **faire**
    choisir un décalage d_i;
    synthétiser la sinusoïde courante;
    multiplier les échantillons obtenus par la fenêtre de pondération w;
    décaler le tableau de sortie de la valeur d_i;
    ajouter les échantillons au tableau de sortie;
  fin
fin

Limites

Les tests que nous avons effectués montrent que cet algorithme permet de synthétiser des sons qui ne laissent percevoir aucune variation d'intensité. L'avantage principal de cette technique est qu'elle peut être utilisée pour les algorithmes de synthèse de tous les modèles sinusoïdaux. Il est donc possible avec cette approche de concevoir une méthode de synthèse générale pour les sons déterministes et les sons bruités stochastiques.

Cette méthode présente toutefois deux inconvénients. Le premier provient du fait qu'elle ne peut être utilisée qu'avec des modèles de sons représentant le signal par des sommes de sinusoïdes. Pour d'autres modèles, la solution de la fenêtre sinusoïdale doit être envisagée. Le second inconvénient de cette méthode est liée à son efficacité. En effet, elle impose pour chaque sinusoïde composant le signal une multiplication par la fenêtre de pondération, ce qui revient à une multiplication supplémentaire par sinusoïde et par échantillon. Néanmoins, nous avons testé et validé son application pour synthétiser des sons en temps-réel.

4.6 Amplitude du signal et saturation

Comme nous venons de la voir, l'application des modèles de synthèse destinés à la représentation et la synthèse de sons bruités, considérés comme des signaux stochastiques, comme par exemple le modèle CNSS, pose de nouveaux problèmes propres aux caractéristiques statistiques de tels sons. Un de ces problèmes, que nous avons été amené à appréhender, est lié à l'amplitude du signal synthétisé, et aux possibilités de saturation, menant à des artefacts audibles.

4.6.1 Problèmes de saturation

Les modèles sinusoïdaux sont particulièrement adaptés aux sons peu bruités. Ils représentent les sons par des partiels. Le nombre de partiels varie peu au cours du temps. Les phases de ces sinusoïdes ont très peu d'influence sur la perception. Elles déterminent toutefois les maxima d'amplitude de la représentation temporelle du son. Lorsqu'elles prennent une même valeur fixe, le maximum d'amplitude vaut la somme des amplitudes. Il est donc possible de prévoir les amplitudes des partiels pour s'assurer que le signal produit ne sature pas.

En ce qui concerne les sons bruités représentés par une somme de sinusoïdes, il est plus complexe de contrôler cette éventuelle saturation. En effet, dans le cas du modèle CNSS, les fréquences et les phases sont redéfinies aléatoirement pour chaque fenêtre de signal, entraînant donc également une redéfinition des amplitudes associées. De plus, le nombre de sinusoïdes peut varier d'une fenêtre à l'autre. Nous présentons ici les différentes méthodes envisageables, que nous avons toutes testées. Nous notons a_n les amplitudes associées aux fréquences f_n par rapport à l'approximation de l'enveloppe spectrale, et b_n les amplitudes définies par rapport aux valeurs de a_n,x selon les méthodes proposées.

4.6.2 Amplitudes en $\frac{1}{N}$

La première technique pour s'assurer que le signal produit ne sature pas, est de diviser les amplitudes par la somme des amplitudes des sinusoïdes :

$$\forall n \in [\![0; N-1]\!], \quad b_n = \frac{a_n}{\sum_n a_n} \tag{4.64}$$

en supposant que, pour toutes les sinusoïdes, l'amplitude a_n est inférieure à 1. En appliquant cette méthode, le cas limite où toutes les phases des sinusoïdes composant le signal sont égales, est pris en compte. Dans ce cas, l'amplitude maximale vaut $\sum_n b_n = 1$. Pour ce cas extrême, nous sommes assuré qu'une saturation ne peut se produire.

Nous pouvons également exprimer l'amplitude RMS en fonction des amplitudes, dans le cas où les phases sont uniformément distribuées :

$$A_{\text{RMS}} = \frac{1}{\sqrt{2}} \sqrt{\sum_n (b_n)^2} = \frac{1}{\sqrt{2} \sum_n a_n} \sqrt{\sum_n a_n^2} \tag{4.65}$$

Nous montrons donc que l'amplitude RMS, décrivant l'intensité perçue, décroît lorsque le nombre N de sinusoïdes augmente. Par exemple, en considérant le cas d'un bruit blanc où toutes les amplitudes a_n sont égales et valent 1, l'amplitude RMS vaut $\frac{1}{\sqrt{2N}}$. Cette méthode ne peut donc être employée que lorsque le nombre de sinusoïdes reste constant au cours du temps.

4.6.3 Amplitudes en $\frac{1}{\sqrt{N}}$

Une autre possibilité est de diviser les amplitudes des sinusoïdes par la racine carrée de la somme des carrés des amplitudes :

$$\forall n \in [\![0; N-1]\!], \quad b_n = \frac{a_n}{\sqrt{\sum_n a_n^2}} \tag{4.66}$$

Cette approche permet de conserver l'amplitude RMS du signal produit, quel que soit le nombre de sinusoïdes utilisés :

$$A_{\text{RMS}} = \frac{1}{\sqrt{2}} \sqrt{\sum_n b_n^2} = \frac{1}{\sqrt{2 \sum_n a_n^2}} \sqrt{\sum_n a_n^2} = \frac{1}{\sqrt{2}} \tag{4.67}$$

Il est toutefois important de remarquer qu'une possibilité de saturation subsiste toujours. C'est pourquoi nous avons cherché à établir une probabilité de saturation.

4.6.4 Amplitudes en $\frac{1}{\sqrt{KN}}$

Nous pouvons étudier les probabilités de saturation du signal généré par addition de N sinusoïdes, dont les phases sont uniformément distribuées entre 0 et 2π. Pour cela, leur amplitude est divisée par la racine carrée du nombre de sinusoïdes, ainsi que par une constante :

$$\forall n \in [\![0; N-1]\!], \quad a'_n = \frac{a_n}{\sqrt{K \sum_n a_n^2}} \tag{4.68}$$

où $K \in \mathbb{R}$.

La somme de sinusoïdes avec des phases uniformément distribuées est équivalente à un signal aléatoire x de distribution normale (gaussienne). La fonction de densité de probabilité $\rho_N(x)$ d'un tel signal est donnée par l'équation (1.58).

L'écart-type σ de la distribution gaussienne des échantillons étant égale à l'amplitude RMS du signal, nous pouvons écrire :

$$\sigma = \frac{1}{\sqrt{2}} \sqrt{\sum_n (b_n)^2} = \frac{1}{\sqrt{2K \sum_n a_n^2}} \sqrt{\sum_n a_n^2} = \frac{1}{\sqrt{2K}} \tag{4.69}$$

Des résultats connus en statistique établissent une probabilité de 99.7% d'obtenir des valeurs supérieures à 3σ. Une saturation du signal aura donc lieu si un échantillon prend une valeur (en valeur absolue) supérieure à 1.0. Cette valeur représente donc la limite $x_{\text{lim}} = 1.0$. L'écart-type σ vaut alors $\frac{1}{3}$ ce qui signifie que K peut être fixé à $K = \frac{9}{2}$.

Sinon, en faisant l'hypothèse que la probabilité associée à la valeur x_{lim} est 10^{-z}, avec $z \in \mathbb{N}$, il vient :

$$\ln(\sqrt{2\pi}10^{-z}) = \frac{-x_{\text{lim}}^2}{2\sigma^2} - \ln(\sigma) \tag{4.70}$$

Ainsi, par exemple, une valeur de $K = 10$ mène à une variance $V(X) = 0.05$, c'est-à-dire $\sigma = 0.25$, et implique une probabilité de saturation de $10^{-5} = 0.00001$.

Il est donc possible d'établir un lien direct entre le paramètre K et les probabilités de saturation du signal. Toutefois, nous pouvons remarquer que cette étude repose sur l'hypothèse que les phases sont uniformément distribuées entre 0 et 2π. Pour une autre loi de distribution, ou un autre intervalle de distribution, les probabilités augmentent fortement.

4.6.5 Conclusions

Comme nous n'avons pas pu établir de méthodes de calcul des amplitudes des sinusoïdes garantissant que le signal généré ne sature pas tout en préservant une amplitude RMS fixe quel que soit le nombre de sinusoïdes, nous sommes obligé d'envisager d'autres solutions dans le cas général où aucune hypothèse sur la distribution des phases n'est possible.

En pratique, lors du développement du modèle CNSS pour la synthèse temps-réel de sons bruités (voir la section 5.1), nous avons employé la dernière méthode décrite ci-dessus (amplitude en $\frac{1}{\sqrt{KN}}$). Mais pour empêcher d'éventuelles saturations, nous avons également développé un algorithme qui, pour chaque fenêtre de sons synthétisée, vérifie si une saturation n'a pas lieu, c'est-à-dire si un échantillon ne dépasse pas 1 en valeur absolue. Si c'est le cas, l'algorithme effectue alors une interpolation linéaire entre le maximum d'amplitude de la fenêtre courante, et le maximum d'amplitude de la prochaine fenêtre. Les expérimentations ont montré que cet algorithme est efficace, mais ajoute un léger coût supplémentaire en temps de calcul. C'est néanmoins le prix à payer pour empêcher le son de produire des clics désagréables, dus aux saturations.

Chapitre 5

Applications et perspectives

Le modèle CNSS a donné lieu à de nombreuses applications, au départ justifiées par la nécessité de conduire des expériences auditives pour vérifier les études théoriques présentées dans les chapitres précédents. Dans la section 5.1, nous présentons ces applications basées sur le modèle CNSS. Une bibliothèque a ainsi été implémentée, de même qu'un objet jMax permettant de synthétiser des sons en temps-réel, à partir des paramètres du modèle CNSS. Ces deux applications définissent un outil expérimental d'exploration et de composition sonore, qui a ensuite été utilisé dans le cadre du projet Dolabip, proposant un instrument d'éveil musical aux enfants.

Ce modèle reste encore expérimental et nous donnons dans la section 5.2 quelques idées qui pourraient être mises en oeuvre afin d'améliorer le modèle théorique et son application à la musique. De plus, les expériences menées à partir de applications implémentées ont permis de soulever un certain nombre de problématiques, essentiellement liées à la perception, que nous envisageons d'étudier lors d'expériences psychoacoustiques.

5.1 Applications

Le modèle CNSS a d'abord été testé, puis implémenté en tant que modèle de synthèse pour pouvoir expérimenter les paramètres de contrôle musical, et en particulier vérifier les liens entre les paramètres mathématiques du modèle et la perception. Cette implémentation permet ainsi de disposer d'un outil expérimental de composition et d'exploration, mais est également destinée à être intégrée dans un logiciel pédagogique d'éducation musicale.

5.1.1 Bibliothèque `libcnss`

Le modèle CNSS a d'abord été implémenté pour effectuer des expériences sur la fiabilité des fonctions de synthèse, et sur l'intérêt des liens entre les paramètres mathématiques et la perception. Ces premières implémentations ont été effectuées en langage `C`, interfacé en `Scheme` [Cha96] (`STK` [Gal98]).

Pour être capable de tester les capacités en temps-réel du modèle CNSS que nous avons développé, nous l'avons implémenté sur l'un des logiciels libres de synthèse audio en temps-réel. Les détails de cette implémentation sont disponibles dans [HBDC02]. Nous les rappelons ici. L'objectif premier était de contrôler tous les paramètres de synthèse aussi rapidement que possible, alors que le son est généré. La première cible a été jMax [jma], car le SCRIME possède déjà une certaine expérience de ce logiciel, le modèle SAS (voir la section 2.3), développé également au SCRIME, ayant déjà été adapté.

Tout d'abord, le modèle lui-même a été implémenté, sous la forme d'une bibliothèque, appelée `libcnss`. Celle-ci est écrite de telle manière qu'il est possible de la relier à de nombreux logiciels de traitement audio (*clients*), qu'ils soient ou non temps-réel. Du point de vue de la bibliothèque, la différence entre ces deux sortes de clients repose essentiellement sur le nombre d'échantillons demandés pour un certain intervalle de temps. Alors qu'un client non temps-réel typique demande un grand nombre d'échantillons (par exemple, 10 secondes d'un son) puis plus rien, un client temps-réel demande des petits nombres d'échantillons (par exemple 64 échantillons, c'est-à-dire 1.45ms pour une fréquence d'échantillonnage de $44,1$kHz), mais effectue cette requête régulièrement.

Pour tenir compte de ces deux possibilités de clients, il est nécessaire d'appliquer les contraintes imposées par les plus exigeants, c'est-à-dire ici les clients temps-réel. Ces clients réclament des algorithmes efficaces où l'essentiel du temps est consommé pour la génération d'échantillons. Pour cela, nous avons directement appliqué les algorithmes de synthèse rapide [SM99], mis au point lors des développements liés au modèle SAS. Ces algorithmes sont capables de produire des sinusoïdes sans accès mémoire et avec très peu d'instructions pour le processeur (voir la section 4.4).

Les aspects expérimentaux du modèle CNSS ont également été pris en considération. Le modèle est décrit dans cette implémentation par plusieurs petits composants de rôles différents, reliés entre eux pour produire des sons bruités. Certains composants concernent la taille de la fenêtre de synthèse, alors que d'autres contrôlent la distribution aléatoire des fréquences ou des phases des sinusoïdes composant le son produit. De temps en temps, de nouveaux composants doivent pouvoir être introduits pour vérifier les conséquences sur la perception de nouveaux paramètres mathématiques. Cette condition a rendu le développement plus difficile, car nous voulions préserver une architecture modulaire claire, facile à étendre.

Dans la bibliothèque actuelle, la solution suivante est adoptée. Tous les composants sont disponibles pour le client sous la forme de *méthodes* CNSS. Le tableau 5.1 énumère les méthodes actuellement disponibles, avec une courte description. Chaque méthode possède ses propres paramètres de synthèse qui peuvent être contrôlés en temps-réel par le client. Le client a également la possibilité de créer des sources de sons CNSS, déterminées à partir de séquences de telles méthodes. Ainsi, une source sonore peut être définie, par exemple, à partir de la séquence *[partials bins winsize nooffset ola]*. Les paramètres que le client peut contrôler sont alors ceux proposés par chaque méthode de la séquence. Par exemple, la méthode *bins* propose quatre paramètres de contrôle, parmi lesquels il est possible de modifier le nombre de casiers (paramètre noté M du modèle CNSS, voir la section 2.4.8), définissant les intervalles de la bande de bruits dans lesquels sont aléatoirement définies les fréquences. Cette méthode permet donc de modifier ce nombre de casiers pendant la synthèse temps-réel de la source sonore.

Méthode CNSS	Paramètres CNSS	Description
uniform	$F_{\min}$, $F_{\max}$	Génère un ensemble de sinusoïdes. Toutes les sinusoïdes ont leur fréquence uniformément distribuée (nombre de casiers infini) dans un intervalle défini.
bins	$F_{\min}$, $F_{\max}$, M, L	Équivalent à la méthode *uniform*, mais les fréquences sont aléatoirement choisies à partir de casiers régulièrement répartis.
winsize	W_s	Définit le nombre d'échantillons de la fenêtre de synthèse.
dl	W_s, δW_s	Définit aléatoirement le nombre d'échantillons de la fenêtre de synthèse.
nooffset		Remplit la fenêtre de synthèse avec des échantillons générés par la synthèse des sinusoïdes préalablement définies. Chaque sinusoïde démarre au début de la fenêtre de synthèse.
offset		Équivalent à la méthode *nooffset*, mais chaque sinusoïde commence à un instant aléatoirement choisi, et se voit appliquer une fenêtre de pondération triangulaire.
ola	H	Applique la technique du recouvrement des fenêtres consécutives de synthèse, avec une fenêtre de pondération de Bartlett.
olasinus		Équivalent à la méthode *ola*, mais la fenêtre de pondération est sinusoïdale.
phi	P	Définit aléatoirement les phases des sinusoïdes.
silence		Remplit les fenêtres de synthèse avec du silence (échantillons nuls).
synchro	P, G, τ, T (sous section 4.3)	Définit les phases des sinusoïdes de façon à ce qu'elles soient aléatoirement choisies à un certain pourcentage de la taille de la fenêtre de synthèse. Elles peuvent éventuellement être aléatoirement *désynchronisées*.
color	C	Définit les amplitudes des sinusoïdes par rapport au paramètre couleur (série de points de l'approximation de l'enveloppe spectrale).
declicker		Teste les éventuelles saturations de façon à atténuer le signal, le cas échéant, en le multipliant par une enveloppe dynamique et empêcher ainsi les clics (voir la section 4.6).
divide		Divise les amplitudes par le nombre de sinusoïdes N (cas du bruit blanc, voir la section 4.6).
normalize		Normalise chaque fenêtre de synthèse.
partials	N	Définit un ensemble de sinusoïdes dont les fréquences, amplitudes et phases sont nulles.
sqrt	K	Divise les amplitudes par $\sqrt{KN}$, où K est un paramètre (voir la section 4.6).

Tab. 5.1 – Liste et descriptions des méthodes CNSS actuellement implémentées.

5.1.2 Objet `cnss4jmax` pour jMax

Comme nous l'avons précisé précédemment, nous avons choisi jMax pour être le premier client de la bibliothèque. Une extension de jMax, appelée `cnss4jmax`, a ainsi été développée pour relier quelques types de données, disponibles dans la bibliothèque, aux types d'objets suivants définis dans jMax :

– `cnss-methods` : Les objets de ce type sont capables de produire une liste des méthodes disponibles dans la bibliothèque, ainsi que d'en fournir une description, en terme de fonctionnalité et de paramètres de contrôle.

– `cnss-frame` : Les objets de ce type permettent d'accéder et de modifier les paramètres de synthèse reliés aux séquences de méthodes CNSS.

– `cnss-source~` : Les objets de ce type sont les synthétiseurs de sons CNSS, ce qui signifie qu'ils produisent les échantillons conformément aux séquences de méthodes (`cnss-methods`) et aux suites de paramètres de synthèse (`cnss-frame`).

La figure 5.1 est une capture d'écran du *patch* jMax contenant des objets CNSS. Sur cette figure, nous pouvons voir qu'une `cnss-frame` reliée à une séquence de méthodes CNSS *[partials bins dl]* permet de contrôler huit paramètres. Elle est connectée à une seule `cnss-source~` avec la séquence de méthodes *[partials bins dl nooffset ola]*, contrôlant les algorithmes utilisés dans la synthèse de la source. La source est elle même connectée à un objet `dac~` qui représente la sortie gauche et droite de la carte son.

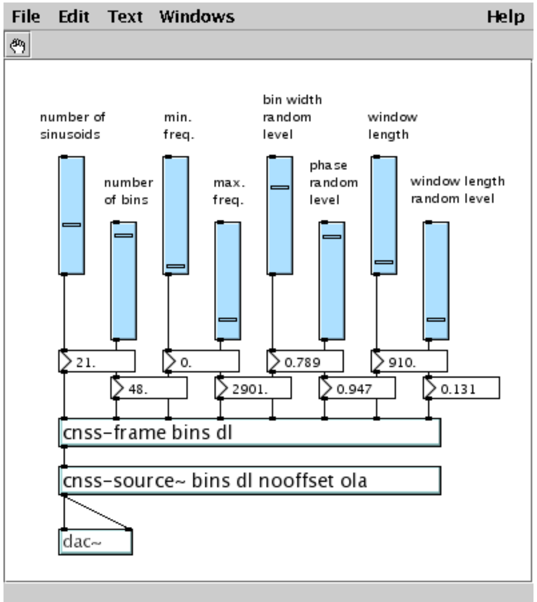

FIG. 5.1 – Capture d'écran d'un *patch* jMax composé d'objets CNSS.

L'ordre des séquences de méthodes est très important. Une permutation dans la séquence peut conduire à une annulation de quelques méthodes. Par exemple, la séquence *[partials bins winsize synchro nooffset]* synthétise du son avec des sinusoïdes dont les fréquences sont aléatoirement définies et dont les phases sont calculées de façon à produire un pic d'intensité plus ou moins important (voire nul). Mais si les méthodes *synchro* et *winsize* sont échangées, les phases ne sont plus synchronisées. En effet, la méthode *synchro* nécessite la connaissance de la taille de la fenêtre de synthèse.

Nous pouvons proposer quelques exemples de séquences de méthodes pour synthétiser des sons bruités :

- *[partials uniform phi winsize nooffset] :*

  Cet exemple est une des séquences de base les plus simples. Les sinusoïdes sont synthétisées dans des fenêtres successives qui ne se recouvrent pas. Les seuls paramètres sont le nombre de sinusoïdes N, la largeur relative L de la fonction de densité de probabilité des phases, et la taille W_s des fenêtres de synthèse.
- *[partials bins winsize dl synchro nooffset ola] :*

  Cet exemple montre que deux méthodes de synthèse peuvent être associées. Ici, les fenêtres de synthèse se chevauchent et leur taille est aléatoirement choisie. Toutefois, il est important de noter que toutes les méthodes ne peuvent pas être employées simultanément, comme par exemple les méthodes *offset* et *ola*.
- *[partials bins winsize dl synchro offset]* and *[partials bins winsize dl synchro nooffset olasinus] :*

  Ces deux séquences sont parmi les plus utilisées pour synthétiser des sons bruités. Elles permettent de contrôler neuf paramètres et envisagent la synthèse de sons dont les propriétés peuvent être complètement différentes.

En l'état actuel, l'implémentation est suffisamment performante pour expérimenter la synthèse en temps-réel de nombreux sons représentés par le modèle CNSS. Sur une machine un peu ancienne Pentium III 500Mhz exécutant GNU/Linux et jMax, nous ne rencontrons aucun problème à contrôler la synthèse en temps-réel de quelques centaines de sinusoïdes pour des fenêtres de synthèse relativement courtes. Nos prochaines expériences nous amèneront à tester plusieurs sources sonores sur différentes bandes de fréquence, nécessitant probablement plus de 500 sinusoïdes en temps-réel. Ce type de synthèse peut certainement déjà être possible avec des ordinateurs récents.

Néanmoins cette implémentation reste en développement constant. Bien que les algorithmes de génération des sinusoïdes soient d'ores et déjà suffisamment rapides, il est évident que les autres parties de la bibliothèque peuvent encore être améliorées pour être mieux adaptées aux contraintes de l'exécution temps-réel. Cette remarque est d'autant plus valable que nous souhaiterions, dans le futur, pouvoir contrôler plusieurs bandes de sons bruités définies à partir de différents paramètres. Le modèle CNSS est également destiné à être amélioré et étendu prochainement.

La bibliothèque `libcnss` [BHa] et l'extension jMax [BHb] associée sont des logiciels libres développés sur une plateforme GNU/Linux. Ils sont disponibles sur la page du SCRIME [scr].

5.1.3 Outil de composition et d'exploration

Même si le modèle reste encore principalement un outil de recherche, le développement d'un objet jMax permet d'appliquer ce modèle à l'exploration des bruits et la composition musicale. Ainsi, l'influence des paramètres existants du modèle sur les propriétés auditives des sons peut être expérimentée en temps-réel. De plus l'objet jMax permet d'être facilement étendu en rajoutant des méthodes pour modifier les algorithmes de détermination des fréquences, amplitudes et phases des sinusoïdes composant le signal. Il peut donc servir de plateforme de recherche pour mettre en évidence de nouveaux paramètres musicaux.

Le modèle CNSS a été développé pour la représentation d'une bande de bruit. Nous pouvons toutefois envisager l'extension de cette approche à un son composé de plusieurs bandes de bruit dont les propriétés sont différentes. D'ores et déjà, la polyphonie est permise, en définissant plusieurs objets CNSS avec des paramètres mathématiques indépendants. Ainsi, cet outil permet d'effectuer des expériences psychoacoustiques sur l'interaction entre différentes bandes de bruit.

5.1.4 Incorporation dans un outil pédagogique d'éducation musicale pour les enfants (Dolabip)

Une autre application du modèle CNSS est son intégration dans un projet d'outil pédagogique pour l'éducation musicale des enfants [DCKS+03].

Projet Dolabip

Le projet Dolabip a été mis en place au Studio de Création et de Recherche en Informatique et en Musique Electroacoustique (SCRIME) [scr]. Il a vu son origine dans le constat que les logiciels de composition musicale existants sont inadaptés à l'apprentissage et la manipulation des sons par des jeunes enfants. Un nouvel outil pédagogique d'éveil musical a été mis au point, basé au départ sur le modèle de sons SAS.

Ce projet a conduit au développement d'un logiciel proposant une interface entre une partie matérielle composée de potentiomètres et de boutons (joystick), et une partie logicielle de synthèse de sons reposant sur le modèle SAS, développée par la société Algory. Cette interface permet de relier chaque axe et/ou bouton du joystick à un paramètre du modèle et autorise ainsi un contrôle musical en temps-réel. La représentation de ces liens s'effectue à l'aide d'un arbre d'objets musicaux, comme le montre la figure 5.2. Une première version de Dolabip a ainsi été conçue, basée uniquement sur le modèle de sons SAS. Les premières expérimentations sur le plan pédagogique ont été effectuées dans des écoles, sous le contrôle d'enseignants, de musiciens et de psychologues. Leur objectif est double : d'une part, la conception avec l'enseignant d'ensemble d'objets musicaux sous Dolabip proposant un intérêt pédagogique, et, d'autre part, l'observation du comportement des enfants lors de compositions et d'improvisations musicales.

Logiciel jMax

Une seconde version du logiciel, développée par Anthony Beurivé, existe sous système GNU/Linux, et est disponible librement selon les termes de la licence GPL [Beu]. Elle est destinée au logiciel jMax. Son interface est certes moins intuitive, mais ce logiciel est portable et permet d'utiliser d'autres modèles de sons, notamment le modèle CNSS. La figure 5.3 présente un exemple de cette deuxième version composé d'objets jMax, parmi lesquels des objets CNSS.

Cette version est constituée d'un ensemble d'objets jMax, notamment ceux se rapportant aux modèles SAS et CNSS, ainsi qu'un objet permettant d'obtenir les informations provenant d'un joystick. Le modèle CNSS définissant un nombre important de paramètres, il est

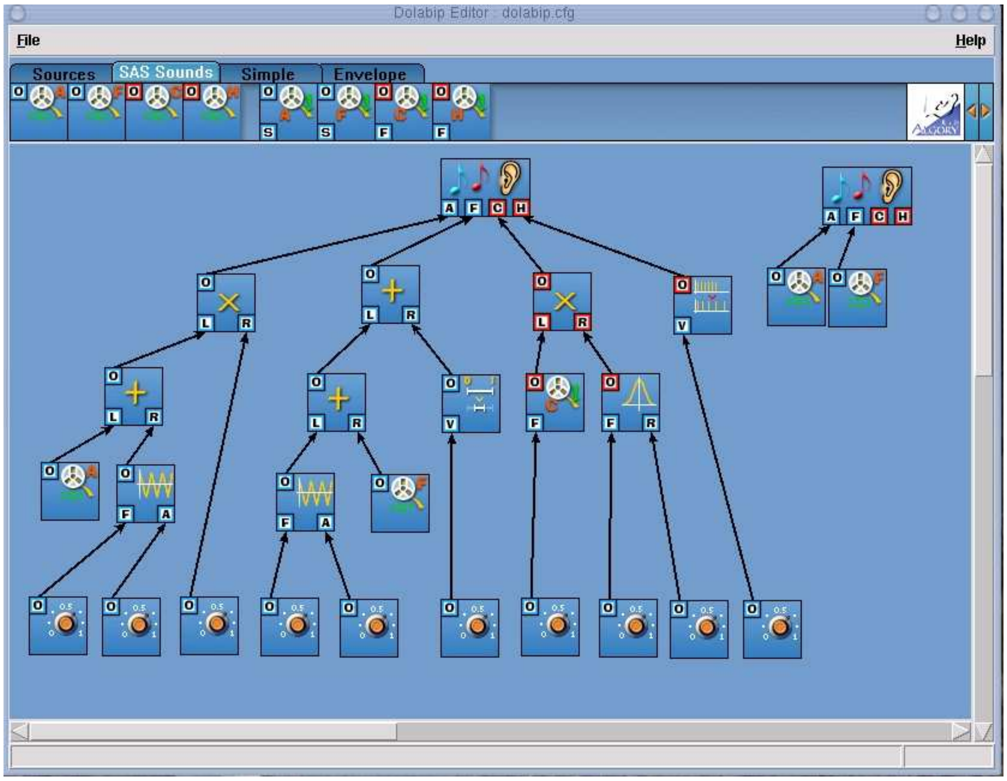

FIG. 5.2 – Capture d'écran de la première version de Dolabip.

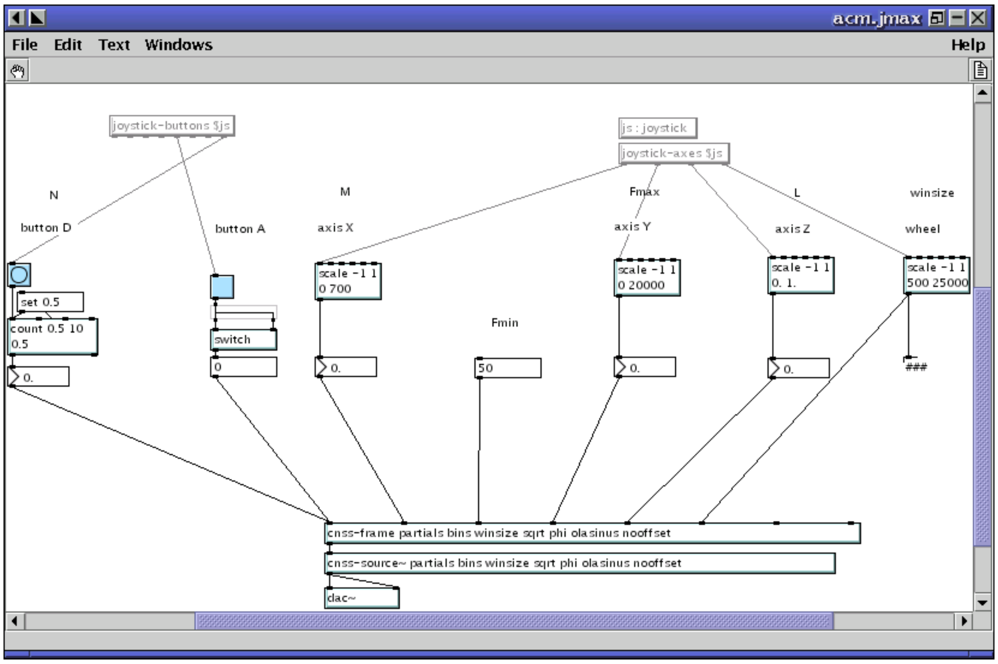

FIG. 5.3 – Capture d'écran de la deuxième version de Dolabip, qui utilise des objets CNSS.

nécessaire d'étudier et de mettre au point plusieurs configurations de méthodes autorisant des contrôles directs, évidents et précis, de propriétés perceptives des sons produits. Cette étude ne peut être validée que par de multiples expériences de synthèse, réunissant des enseignants et des musiciens. Une troisième version de Dolabip est en cours. Elle combine l'interface souple de la première version avec la possibilité d'appliquer plusieurs modèles de sons, proposée par la seconde version.

Les premières possibilités d'appliquer le modèle CNSS de sons bruités d'une part à la composition musicale et à l'exploration du son, et d'autre part à l'éveil musical par le jeu pour les jeunes enfants, doivent permettre de recueillir l'avis de musiciens et de compositeurs, afin de modifier et d'étendre ce modèle.

5.2 Perspectives

Le modèle CNSS, même s'il apparaît prometteur, reste encore expérimental. L'étude des sons bruités que nous avons effectuée pour mettre au point ce modèle nous a permis de mettre en lumière des problèmes propres aux propriétés particulières de ce type de sons, que ce soit sur le plan de la perception, du traitement statistique du signal, ou du contrôle musical, etc. . . .

Certains de ces problèmes ont été présentés en détails et nous avons pu proposer des solutions adaptées, comme, par exemple, les problèmes liés au chevauchement des fenêtres de sons aléatoires lors de la synthèse, ou l'importance de la densité spectrale menant à une méthode d'analyse spécifique. Toutefois, d'autres problèmes restent encore à l'étude. Nous essayons dans cette section de présenter quelques problématiques, et pour certaines, des esquisses de solutions et/ou des expériences à mener pour obtenir des informations supplémentaires sur le plan de la perception des bandes de bruit.

5.2.1 Distributions des fréquences

Le modèle CNSS représente des bandes de bruit par une somme fixe de sinusoïdes. Ces sinusoïdes sont définies par des amplitudes fixes, et par des phases et des fréquences aléatoires. La distribution des fréquences est déterminée par les paramètres du modèle, en divisant la bande de fréquence ΔF considérée en M intervalles réguliers (casiers). Ce choix est justifié mathématiquement par le modèle des bruits thermiques qui définit les fréquences des sinusoïdes comme régulièrement espacées.

Les recherches en psychoacoustique (voir la section 1.5.3) montrent l'intérêt d'une division logarithmique de l'échelle des fréquences par rapport à la perception de l'intensité et à la discrimination de fréquences. Nous pouvons donc nous interroger sur la possibilité de définir les fréquences des sinusoïdes composant une bande de bruit sur une échelle logarithmique.

Nous présentons dans cette sous-section les distributions que nous pouvons envisager, les algorithmes de synthèse associés, et les perspectives de recherche liées à ces nouvelles distributions.

Distribution linéaire des fréquences

Dans le modèle CNSS, les M casiers contenant au plus une fréquence sont régulièrement répartis sur la bande de fréquence $\Delta F = F_{\max} - F_{\min}$. Nous rappelons ici brièvement les définitions vues dans la section 2.4.8, où elles sont expliquées plus en détails. Chaque casier est défini par sa limite supérieure notée B_i :

$$\forall i \in [1; M], \quad B_i = F_{\min} + \frac{\Delta F}{M} i \tag{5.1}$$

Ces limites sont donc décrites par une relation linéaire de l'indice i du casier. Les intervalles, notés I_{B_i}, associés à chaque casier B_i, dans lesquels sont aléatoirement choisies les fréquences sont ainsi définis :

$$\forall i \in [0; M-1], \quad I_{B_i} =]F_{\min} + i\frac{F_{\max} - F_{\min}}{M}; F_{\min} + (i+1)\frac{F_{\max} - F_{\min}}{M}] \tag{5.2}$$

Les écarts entre les casiers sont donc constants :

$$\forall i \in [1; M[, \quad \Delta B_i = B_{i+1} - B_i = \frac{\Delta F}{M} \tag{5.3}$$

La figure 5.4 représente ces écarts constants en fonction des fréquences.

Distribution logarithmique des fréquences

La première distribution des casiers que nous avons envisagée permet non plus de conserver une largeur d'intervalle de fréquences constante, mais un rapport entre deux intervalles successifs fixé :

$$\forall i \in [1; M[, \quad \frac{B_{i+1}}{B_i} = K \tag{5.4}$$

Cette propriété induit une distribution logarithmique des casiers, et donc des fréquences :

$$\forall i \in [1; M[, \quad B_i = F_{\min}(\frac{F_{\max}}{F_{\min}})^{\frac{i}{M}} \tag{5.5}$$

où la basse fréquence limite $F_{\min}$ ne doit pas être nulle. Nous pouvons vérifier que les intervalles définis par chaque casier ne sont alors plus constants :

$$\forall i \in [1; M[, \quad \Delta B_i = F_{\min}(\frac{F_{\max}}{F_{\min}})^{\frac{i}{M}}[(\frac{F_{\max}}{F_{\min}})^{\frac{1}{M}} - 1] \tag{5.6}$$

La figure 5.4 représente ces intervalles en fonction des valeurs de fréquences définissant les limites de ces casiers, ce qui revient, dans le cas général, à montrer les variations des écarts entre les fréquences, en fonction des fréquences.

En notant Λ le rapport entre la fréquence maximale $F_{\max}$ et la fréquence minimale $F_{\min}$, nous pouvons exprimer l'indice i correspondant à l'égalité entre les écarts dans le cas d'une distribution linéaire et dans celui d'une distribution logarithmique :

$$i = M.\frac{\ln[\frac{\Lambda - 1}{M(\Lambda^{\frac{1}{M}} - 1)}]}{\ln(\Lambda)} \tag{5.7}$$

Dans le cas où $F_{\min} = 1$, $F_{\max} = \frac{F_s}{2}$ et $M = 512$, nous obtenons $i_0 = 393.61$. La fréquence correspondant à cet indice pour une distribution logarithmique est $f_{i_0} = 2183$Hz. Ce calcul indique que, pour 512 casiers définis sur une bande de $\frac{F_s}{2} = 22050$Hz de large, les différences entre les fréquences deviennent plus importantes pour une distribution logarithmique que pour une distribution linéaire, pour des fréquences supérieures à approximativement 2200Hz. Les expériences que nous avons effectuées à partir de cette nouvelle distribution montrent que, pour un même nombre de sinusoïdes, le bruit synthétisé donne une sensation de densité spectrale plus faible. Cette observation va à l'opposé de l'effet escompté, puisque nous souhaitons diminuer le nombre de sinusoïdes. Elle s'explique toutefois par le fait que les premiers casiers sont trop étroits, ce qui accroît inutilement le nombre de sinusoïdes nécessaires, et que les casiers supérieurs à 2200Hz (dans l'exemple du calcul) sont alors trop larges. Il faut donc définir une nouvelle distribution permettant d'augmenter la taille des intervalles pour les basses fréquences.

Distribution généralisée des fréquences

Nous proposons une nouvelle distribution des fréquences, permettant de contrôler les largeurs des intervalles des casiers pour les basses fréquences, et ainsi augmenter la valeur de fréquence pour laquelle les largeurs des casiers correspondant à cette nouvelle distribution deviennent plus importantes que les largeurs d'une distribution linéaire. Pour cela, nous introduisons un nouveau paramètre que nous notons K_0. Ce paramètre représente la taille minimale des intervalles associés aux casiers. Les casiers sont définis par la relation :

$$\forall i \in [1; M[, \quad B_i = F_{\min}\left(\frac{F_{\max} - MK_0}{F_{\min}}\right)^{\frac{i}{M}} + K_0 i \tag{5.8}$$

Les largeurs des intervalles ne sont pas constantes, de même que leur rapport. Nous pouvons exprimer ces largeurs :

$$\forall i \in [1; M[, \quad \Delta B_i = F_{\min}\left(\frac{F_{\max} - MK_0}{F_{\min}}\right)^{\frac{i+1}{M}} - F_{\min}\left(\frac{F_{\max} - MK_0}{F_{\min}}\right)^{\frac{i}{M}} + K_0 \tag{5.9}$$

Cette relation montre bien la différence avec la distribution logarithmique précédente, avec le paramètre K_0 qui précise la largeur minimale. La figure 5.4 représente ces intervalles en fonction des valeurs de fréquences définissant les limites de ces casiers, ce qui revient à montrer les variations des écarts entre les fréquences, en fonction des fréquences.

Le domaine de définition du paramètre K_0 dépend de la largeur de la bande de fréquences et du nombre de casiers :

$$0 \leq K_0 \leq \frac{\Delta F}{M} \tag{5.10}$$

Les deux bornes de cet intervalle de définition ramènent aux cas des précédentes distributions. En effet, dans le cas où $K_0 = \frac{\Delta F}{M}$, la distribution est linéaire, alors que dans le cas où $K_0 = 0$, la distribution est logarithmique. Entre ces deux bornes, une infinité de distributions différentes peut être déterminée.

Compensation des amplitudes

Le fait d'appliquer une distribution des fréquences non linéaire induit des variations au niveau de la répartition de l'énergie dans le spectre. En effet, l'augmentation ou la réduction

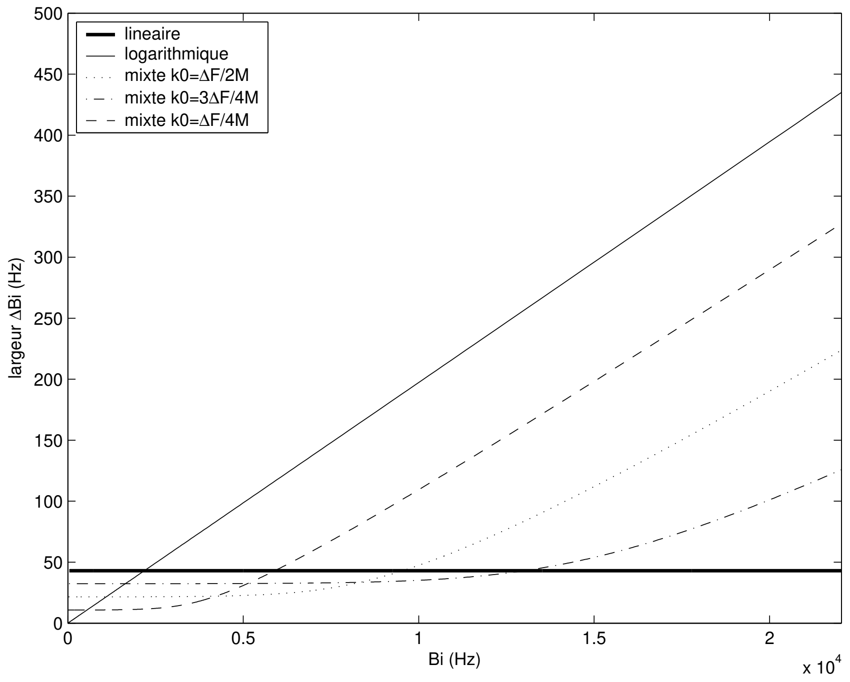

FIG. 5.4 – Illustration des distributions envisageables (linéaire, logarithmique, générale pour différentes valeurs de K_0) des fréquences pour la modélisation de bandes de bruit en montrant les variations des différences entre les fréquences en fonction de ces fréquences.

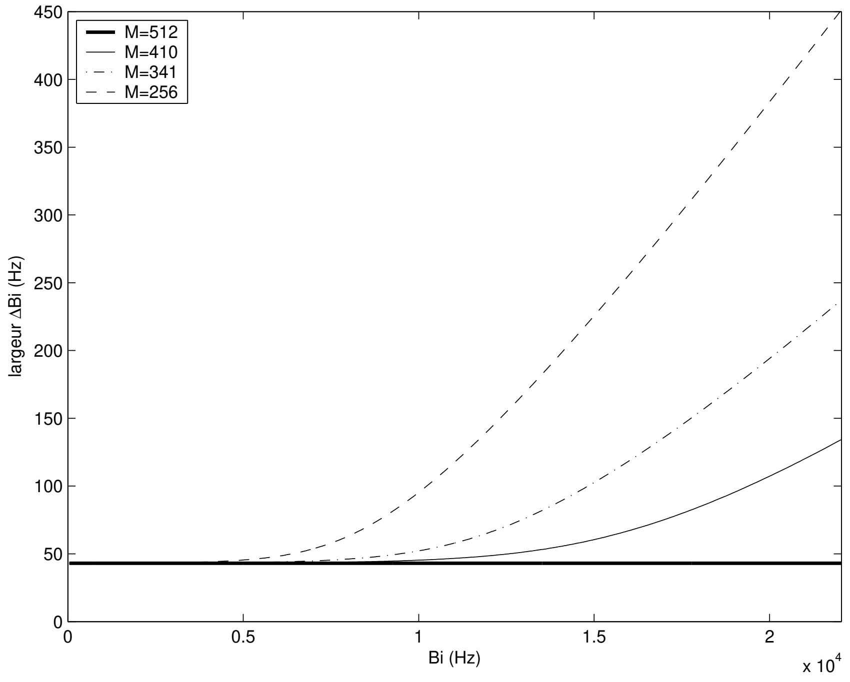

FIG. 5.5 – Illustration des distributions des fréquences pour la modélisation de bandes de bruit obtenues en fixant le paramètre K_0 ($K_0 = \frac{22049}{512}$), pour différentes valeurs du nombre de casiers M.

du nombre de sinusoïdes dans une certaine bande de fréquences résulte en une augmentation ou une réduction de l'énergie dans cette bande. Pour pouvoir comparer les distributions, nous devons mettre au point une technique de compensation des amplitudes. Ainsi, une bande contenant N sinusoïdes d'amplitudes a_0 a une énergie comparable à celle d'une bande de $\frac{N}{k}$ sinusoïdes d'amplitude $\frac{a_0}{\sqrt{k}}$.

Pour cela, il est nécessaire d'étudier les variations du nombre de sinusoïdes pour une distribution choisie en fonction du nombre correspondant à une distribution linéaire, qui reste la distribution de référence. Le nombre de sinusoïdes dans le cas linéaire correspond, à une constante multiplicative près, à l'indice i qui lui est associé. Dans le cas général, le nombre de sinusoïdes est décrit par la variation dB_i, où B_i est donné par l'équation (5.8).

L'expression théorique de ces variations s'est avérée trop complexe à définir, et nous n'y sommes pas parvenu pour l'instant. Toutefois, l'algorithme de synthèse lié au modèle CNSS, présenté dans la section 4.3, nous permet de contourner ce problème. En effet, dans cet algorithme, nous utilisons les indices des fréquences (des casiers). Chaque indice i correspond à un casier dont la répartition est linéaire. C'est à partir de ces indices que nous pouvons exprimer la variation dB_i.

L'équation (5.8) peut s'écrire :

$$\forall i \in [1; M[, \quad B_i = F_{\min} \exp(\frac{i}{M}. \ln(\frac{F_{\max} - MK_0}{F_{\min}})) + K_0 i \tag{5.11}$$

La dérivée de B_i en fonction de l'indice i est définie par :

$$\forall i \in [1; M[, \quad \frac{dB}{di} = F_{\min} \frac{\ln(\frac{F_{\max}-MK_0}{F_{\min}})}{M}. \exp(i \frac{\ln(\frac{F_{\max}-MK_0}{F_{\min}})}{M} + K_0 \tag{5.12}$$

Ainsi, les valeurs des amplitudes de chaque sinusoïde d'indice i valent donc :

$$\forall i \in [1; M[, \quad a'_i = a_i \sqrt{\frac{dB}{di}} \tag{5.13}$$

Algorithmes

L'algorithme que nous proposons permet de synthétiser des bandes de bruit à partir de sinusoïdes dont les fréquences sont aléatoirement choisies dans des casiers pouvant être répartis non linéairement. Il est basé sur les algorithmes de synthèse associés au modèle CNSS, présentés dans la section 4.3.

Il consiste tout d'abord à calculer une valeur k, définie à partir du nombre de casiers M et de la bande de fréquence ΔF par la relation :

$$k = \exp[\frac{\ln(\frac{F_{\max}-MK_0}{F_{\min}})}{M}] \tag{5.14}$$

où K_0 représente le paramètre de contrôle de la distribution des casiers.

L'algorithme détermine ensuite aléatoirement (algorithme de tirage des permutations) les indices i des casiers dans lesquels sont aléatoirement choisies les fréquences. Pour chaque indice, les amplitudes associées aux fréquences sont obtenues à partir du paramètre couleur C (enveloppe spectrale). Elles sont alors divisées par un facteur correspondant à $\sqrt{\frac{dB}{di}}$.

Données : Paramètres N,L,M,ΔF, K_0, couleur C, tableaux `freq` et `casier`
début
  pour *chaque fenêtre* **faire**
    Calculer k;
    Tirer les casiers $B[i]$;
    pour $i \in 0, \ldots, N-1$ **faire**
      Tirer un réel r dans $[0; L]$;
      Calculer $a = F_{\min}.k^{B[i]} + B[i].K_0$;
      Calculer $b = F_{\min}.k^{B[i]+1} + (B[i]+1).K_0$;
      $\text{freq}[i] = a + (1-r) * (b-a)$;
      $\text{ampli}[i] = C(\text{freq}[i])$;
      $\text{ampli}[i] = \text{ampli}[i] * \sqrt{F_{\min}.k.\exp(ki) + K_0}$;
    fin
  fin
fin

Cet algorithme est évidemment plus coûteux qu'un algorithme de détermination linéaire des fréquences. Mais ce coût n'est pas important par rapport à celui des algorithmes de synthèse additive des sinusoïdes.

Perspectives

Cet algorithme a été implémenté pour effectuer les expérimentations concernant l'exactitude de la compensation de l'amplitude des sinusoïdes. Notre algorithme nous a semblé valable en n'induisant aucune modification de couleur. Il va bientôt être implémenté sous la forme d'une méthode CNSS pour la bibliothèque `libcnss` et pour l'objet jMax `cnss4jmax`. Ainsi des expériences plus approfondies pourront être menées.

L'intérêt d'une distribution non linéaire des fréquences est la possible réduction du nombre de sinusoïdes nécessaires pour synthétiser une bande de bruit. La figure 5.5 montre les distributions des fréquences pour différents nombres de casiers et un paramètre K_0 constant. Elle confirme qu'il semble possible de diminuer le nombre de sinusoïdes tout en préservant des différences faibles entre les basses fréquences. Ainsi, l'efficacité des méthodes de synthèse spectrale de bruit semble pouvoir être améliorée.

Il est important de remarquer qu'une distribution non linéaire des fréquences annule l'intérêt du paramètre L (largeur de la fonction de densité de probabilité dans un casier) du modèle CNSS. En effet, un écart plus ou moins constant sur une échelle linéaire se traduira par un rapport plus ou moins constant, pas forcément entier entre les fréquences. Le fait de pouvoir introduire un rapport entier constant doit induire un effet sur le plan de la perception, et donne l'espoir de pouvoir définir un nouveau paramètre de contrôle musical.

Plus généralement, ces distributions différentes de fréquences doivent permettre de mener des études pour approfondir la compréhension du mode de fonctionnement de la perception de la densité spectrale de bandes de bruit. En effet, cette perception est basée sur les fluctuations d'intensité dans le cas d'une bande étroite. Le fait de définir les fréquences sur une échelle linéaire induit des périodicités sur la puissance de l'enveloppe temporelle. L'utilisation d'une échelle logarithmique risque de réduire cet effet et cette échelle pourra être utilisée lors d'expériences psychoacoustiques pour donner de nouvelles informations sur le fonctionnement du système auditif.

5.2.2 Modèle de sons bruités avec des paramètres de haut niveau

Le modèle CNSS reste pour l'instant un modèle expérimental, utile pour étudier la perception et pour valider les méthodes de définition statistique des caractéristiques des sinusoïdes composant le son bruité synthétisé. Il constitue également un outil de composition, mais en l'état actuel, reste limité par sa complexité, induite par le trop grand nombre de paramètres pas toujours indépendants, et par les limites des méthodes d'analyse. Nous proposons dans cette sous-section quelques idées pour améliorer ces différents points.

Paramètres haut niveau

Les paramètres que propose le modèle CNSS sont liés à des paramètres perceptifs, mais pas toujours de manière directe. De plus ils ne sont pas toujours indépendants entre eux. Un peu comme l'a proposé le modèle SAS pour les modèles sinusoïdaux, il est nécessaire de proposer des paramètres musicaux de plus haut niveau, permettant au compositeur un contrôle du son. Par exemple, les paramètres notés M et N (respectivement le nombre de casiers et le nombre de fréquences) sont tous deux liés à la perception de la densité spectrale, mais de manière différente. Si un utilisateur modifie le paramètre N, il modifie cette densité, et il en va de même pour le paramètre M. Il est donc nécessaire de proposer plutôt deux paramètres perceptifs indépendants : la densité spectrale moyenne et les variations de densité spectrale.

Paramètre L non linéaire

Le paramètre noté L, représentant la largeur de la fonction de densité de probabilité dans chaque casier, a une influence sur la force de la hauteur que laisse percevoir le son bruité modélisé. Les études menées lors du développement de la méthode d'analyse de ce paramètre (voir la section 3.7) laissent apparaître que cette relation ne semble pas linéaire. Les expériences de synthèse ont confirmé cette remarque. Approximativement nous pouvons dire que, pour des valeurs comprises entre 0.8 et 1.0, les variations sont plus faibles que pour des valeurs comprises entre 0.2 et 0.8. Ainsi, nous pouvons envisager à l'avenir de considérer ce paramètre d'harmonicité sur une échelle logarithmique pour proposer ainsi une précision supplémentaire dans la transformation des sons.

Variation aléatoire des amplitudes

Dans le modèle CNSS, les amplitudes des sinusoïdes composant le son bruité sont déterminées à partir du paramètre couleur, c'est-à-dire une approximation de l'enveloppe spectrale. Pourtant le modèle des bruits thermiques, sur lequel est basé le modèle CNSS, définit les amplitudes des sinusoïdes comme des variables aléatoires, distribuées selon une loi de Rayleigh. Le théorème central limite nous permet de justifier théoriquement le choix de donner aux amplitudes des valeurs déterminées.

Nous pouvons néanmoins nous interroger sur l'intérêt de cette définition statistique des amplitudes. Comme, à notre connaissance, il n'existe pas de résultat de recherche en psychoacoustique que nous pourrions relier au rôle des variations statistiques d'amplitude, nous pensons qu'il est nécessaire d'effectuer des tests pour établir des comparaisons sur le plan de la perception entre des sons dont l'enveloppe spectrale est fixée, et d'autres dont l'enveloppe spectrale moyenne est fixée, à partir de laquelle les amplitudes sont aléatoirement distribuées. Cette étude peut éventuellement aboutir à la définition de nouveaux paramètres de contrôle, liés à ces distributions, mais également à la possibilité de diminuer le nombre de sinusoïdes nécessaires pour synthétiser un son de densité maximale, en supposant que ces variations d'amplitude interviennent dans la perception de la densité spectrale. Pour simplifier, nous pouvons nous poser la question de savoir si le système auditif arrive à percevoir une régularité dans un signal généré par des fenêtres successives de signaux d'enveloppes spectrales strictement égales.

Distribution des phases

Le modèle des bruits thermiques considère les phases des sinusoïdes formant les bruits comme uniformément distribuées entre 0 et 2π. Nous avons proposé pour le modèle CNSS de modifier cet intervalle par l'intermédiaire de sa largeur relative notée P, ce qui permet de contrôler l'impulsivité du bruit synthétisé. Ce paramètre s'avère cependant limité par rapport à la perception. En effet, comme nous l'avons vu dans la section 4.2.3, une distribution uniforme des phases ne permet pas d'obtenir une bande de bruit dont les fluctuations d'intensité sont minimales. Pour cela, il est nécessaire de mettre en place un algorithme complexe, comme celui proposé par Pumplin, de façon à pouvoir réduire ces fluctuations et s'approcher ainsi d'un low-noise noise.

Amélioration de l'analyse de la densité spectrale

Les méthodes d'approximation de la densité spectrale que nous proposons dans la section 3.6 présentent quelques limites et doivent être encore améliorées. Tout d'abord, de nombreuses autres expériences doivent être réalisées à partir d'une grande variété de sons bruités naturels, pour adapter de la meilleure façon possible les paramètres complexes et nombreux de ces méthodes.

Ensuite les résultats des expériences que nous avons montrés ont été obtenus à partir d'une seule fenêtre de signal. Nous envisageons, concernant l'analyse de bandes de bruit, d'étudier le son sur des fenêtres temporelles w_2 (par exemple de taille $W_s^2 = 4096$ échantillons) composées elles-mêmes de fenêtres temporelles w_1 plus petites (par exemple de taille $W_s^1 = 1024$ échantillons). La figure 5.6 illustre ce principe. Chaque analyse permet de déterminer des valeurs des paramètres du modèle CNSS pour chaque fenêtre w_1. Ces valeurs sont ensuite moyennées pour définir les paramètres associés à chaque fenêtre w_2. L'intérêt de cette approche doit être validé par des expériences d'analyse et de synthèse de sons bruités.

Enfin, des améliorations récentes concernant l'analyse de sinusoïdes non stationnaires permettent d'extraire des sinusoïdes dont les fréquences et les amplitudes varient [LMR02], qu'il semble possible de corréler avec la densité spectrale d'une bande de bruit. Une étude mérite certainement d'être menée pour adapter ces techniques à l'approximation de la densité spectrale.

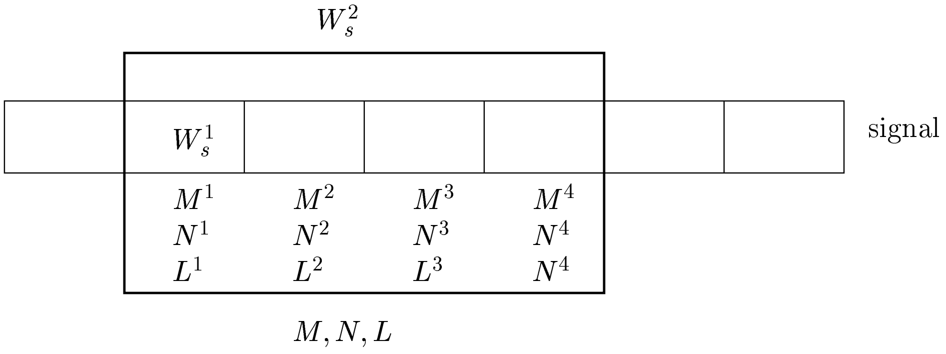

FIG. 5.6 – Illustration de la possibilité de définir les paramètres du modèle CNSS pour une fenêtre de taille W_s^2, à partir des paramètres obtenus pour des fenêtres plus petites de taille W_s^1.

Synthèse FFT^{-1} de sons bruités

Dans la section 4.5, nous proposons plusieurs algorithmes permettant d'effectuer une synthèse spectrale de bruit sans introduire d'artefact (variation d'intensité) dû au caractère aléatoire du son. Un de ces algorithmes propose de décaler les temps de début de chaque sinusoïde dans chaque fenêtre temporelle de synthèse, et de les multiplier indépendamment par une fenêtre de pondération. Cet algorithme rajoute un coût non négligeable à la synthèse de ces sons qui peut s'approximer à une multiplication supplémentaire par échantillon et par sinusoïde.

Cette technique a été développée pour être appliquée à la synthèse de bruits par somme de sinusoïdes, en calculant chacune d'elle, c'est-à-dire en rajoutant successivement chacune de leur contribution. Toutefois, nous pouvons aussi envisager son application à la synthèse par transformée inverse de Fourier (FFT^1). En effet, pour cette méthode de synthèse, la fenêtre de pondération est prise en considération pour chaque sinusoïde, et donc le fait d'appliquer cette fenêtre à chacune d'entre elle n'impliquera aucun coût supplémentaire. Par contre, contrairement à l'utilisation habituelle, les sinusoïdes ne sont pas définies sur la totalité de chaque fenêtre de synthèse, puisque certaines commencent en retard et d'autres terminent en avance. La contribution de ces sinusoïdes est plus complexe à calculer : il est nécessaire de prendre en compte les valeurs dans le spectre dues à la fenêtre de pondération, mais également l'interpolation induite par la durée de la sinusoïde dans la fenêtre de synthèse. Cette interpolation rajoute probablement des composants fréquentiels dans le spectre, donc du temps de calcul. Mais cette contrainte peut être compensée par le temps gagné en effectuant la synthèse par transformée inverse de Fourier. Cet algorithme doit être étudié, mis au point et implémenté, pour ensuite effectuer des tests de performance afin d'évaluer le gain éventuel en rapidité de calcul.

5.2.3 Synthèse de la parole

Les modèles de sons bruités ont de nombreuses applications pour le traitement de la parole. Ainsi, nous pouvons penser que le modèle CNSS peut être utilisé pour synthétiser et transformer des signaux issus de la voix, en particulier les parties de sons correspondant aux consonnes non voisées. Nous avons déjà effectué quelques expériences pour reproduire quelques consonnes, et pour vérifier la qualité de l'allongement dans le temps par le modèle CNSS. Les résultats obtenus sont prometteurs.

Par ailleurs, les voix chuchotées peuvent être appréhendées comme des sons bruités, et semblent donc pouvoir être représentées par les paramètres du modèle CNSS. Là encore, nous avons reproduit des phrases chuchotées, et modifié leur durée sans nuire à la qualité et à la compréhension des phrases.

Ces résultats encourageants nous ont amenés à envisager d'autres types de traitement. Ainsi, nous espérons, par le modèle CNSS, pouvoir analyser une voix normale, et être capable de synthétiser la même phrase, mais chuchotée. Ce type de transformation est déjà proposé par les modèles hybrides, mais également par les techniques classiques de vocodeur de phase, en modifiant les informations liées aux spectres de phase. Néanmoins, nous espérons pouvoir proposer une meilleure qualité.

5.2.4 Vers un modèle général

Le modèle CNSS que nous présentons dans cette thèse est destiné aux sons bruités, et peut parfaitement s'appliquer à la représentation de la partie dite stochastique d'un modèle hybride. En effet, les études théoriques que nous avons détaillées et que nous avons développées confirment la possibilité de représenter un son aléatoire par une somme de sinusoïdes de phases aléatoires. Ce processus est pour l'instant couramment employé sans avoir été rigoureusement justifié. De plus, les expériences que nous avons menées ont mis en lumière des limites dans l'application des techniques de synthèse utilisées pour des sons déterministes. Les solutions que nous proposons permettent d'adapter ces techniques aux types de son tels que les bruits. Enfin le choix de définir comme des variables aléatoires les fréquences de ces sinusoïdes autorise la manipulation des sons sur le plan musical par des contrôles originaux. L'application du modèle CNSS pour représenter la partie résiduelle d'un modèle hybride semble donc prometteuse, et permet une avancée qui se traduira certainement par une amélioration de la qualité de rendu des sons.

Par ailleurs nous proposons dans la section 3.1 un principe d'analyse général de sons. Cette approche n'a encore pas pu être développée sur sa totalité, mais devrait rapidement l'être pour offrir une alternative originale à l'approche classique des modèles hybrides qui effectuent une soustraction spectrale pour définir la composante stochastique du son. Cette opération reste à notre avis une source d'erreur trop importante. Le principe général proposé effectue une segmentation sur l'échelle temporelle définissant les fenêtres contenant une transitoire. Ces segments sont soit préservés, soit traités par des modèles spécifiques. Ensuite, une segmentation spectrale définit les pics du spectre qui seront traités indépendamment par une méthode d'analyse sinusoïdale de haute précision. Les parties restantes du spectre sont soumises à une analyse conforme au modèle CNSS, en tenant en compte des segments du spectre qui correspondent à des partiels.

Un développement complet de cette approche doit à présent être effectué de façon à pouvoir procéder à des tests sur une grande variété de sons, et à mettre en avant les éventuels avantages de cette façon d'aborder les problèmes de l'analyse.

5.2.5 Modélisation des transitoires

Comme nous l'avons déjà expliqué dans la section 2.3.6, les transitoires nécessitent un traitement particulier en ce qui concerne l'analyse, la synthèse et les transformations sur l'échelle temporelle. Dans le cadre du développement du modèle CNSS, nous avons mené une étude sur l'influence de la distribution des phases. Une distribution de largeur étroite, sur l'intervalle $[0; 2\pi[$ induit un pic d'intensité sur le signal de synthèse. Ce pic peut être assimilé à une transitoire. Nous pouvons donc penser qu'il est possible de modéliser certaines transitoires par une somme de sinusoïdes de fréquences et d'amplitudes aléatoires, et de phases constantes ou distribuées sur un intervalle très limité.

Des tests doivent être menés pour confirmer cette possibilité de représenter et de synthétiser toute sorte de transitoires, et montrer les limites de cette approche. De nouveaux paramètres de contrôle pourraient être définis par rapport à la distribution des fréquences, des amplitudes et des phases. L'avantage de cette approche est qu'elle permettrait de rendre homogène la représentation du bruit et des transitoires bruitées.

5.2.6 Expériences psychoacoustiques

Le modèle CNSS peut parfaitement servir de modèle de synthèse de bruits dans le cadre d'expériences psychoacoustiques. En effet, il repose sur le modèle des bruits thermiques, qui est couramment appliqué pour synthétiser les sons servant aux expériences, puisqu'il permet de contrôler avec exactitude l'enveloppe spectrale entre autres. Nous souhaitons prochainement mettre en place des expériences permettant de quantifier l'influence des paramètres du modèle sur la perception et de mesurer les propriétés de masquage entre des bandes de bruit de propriétés différentes formant un même son.

Outil pour les expériences psychoacoustiques

Les expériences psychoacoustiques mettant en jeu un bruit sont souvent effectuées par application du modèle du bruit thermique, pour pouvoir contrôler parfaitement l'enveloppe spectrale du son synthétisé. Ainsi le modèle CNSS peut servir d'outil de synthèse en temps-réel pour effectuer des expériences. En effet, il permet de synthétiser en temps réel des bruits classiques, mais, en plus, de par ses nouveaux paramètres de contrôle des distributions, il permet de générer des sons aux propriétés originales. Nous pouvons donner comme exemple un bruit généré à partir d'une densité spectrale non maximale, avec des fréquences régulières ($L = 0$). Le son produit ressemble au bruit ondulant en laissant percevoir une hauteur, alors que l'algorithme de synthèse est complètement différent. Un tel son peut présenter un intérêt dans la compréhension de la perception de la hauteur.

L'implémentation du modèle CNSS en tant que modèle de synthèse peut être étendue en rajoutant des méthodes de détermination des caractéristiques des sinusoïdes composant le son. L'exemple que nous avons déjà abordé est celui du low-noise noise, souvent employé lors d'expériences sur l'étude du masquage des bandes de bruit. Il est tout à fait possible d'implémenter l'algorithme proposé par Pumplin (voir la section 4.2.3) afin de synthétiser ce type de bruit.

Nombre de sinusoïdes utiles

La définition de la densité spectrale donnée dans la section 2.4.4 a été mise en défaut lors de nos études pratiques, de même que les premières expériences menées par W. Hartmann [HMGB86]. Il nous semble donc utile de procéder à de nouvelles expériences psychoacoustiques permettant de définir le nombre de sinusoïdes suffisantes et/ou nécessaires pour pouvoir synthétiser un bruit blanc qui ne pourra être différencié d'un bruit blanc généré par distribution gaussienne des échantillons, en étudiant notamment l'influence des paramètres du modèle CNSS, c'est-à-dire l'influence de la distribution des fréquences.

Par ailleurs, les tests concernant la méthode d'approximation de la densité spectrale ont montré que les bruits blancs générés par tirage aléatoire des échantillons ne sont pas de densité spectrale maximale. D'autres expériences doivent être effectuées pour confirmer cette remarque, et définir les paramètres du modèle CNSS permettant de synthétiser un bruit comparable sur le plan de la perception.

Masquage de bandes de bruit

La qualité des méthodes d'analyse concernant l'approximation de la densité spectrale ne peut actuellement être rigoureusement discutée, car nous ne possédons pas les résolutions du système auditif liées à la perception de la densité spectrale. En particulier, il est nécessaire d'étudier l'influence d'une bande de bruit de densité spectrale faible située entre des bandes de bruit de densité spectrale maximale. Ces expériences reviendraient à étudier les propriétés de masquage de bandes de bruit entre elles, en fonction de leur densité spectrale, de leur largeur et de leur fréquence centrale. Ces résultats auront une importance non seulement pour fixer les limites de résolution de la méthode d'approximation de la densité spectrale, mais aussi dans le cadre de la représentation et de la synthèse de sons formés de bandes de bruits dont les propriétés sont distinctes.

Conclusion

L'objectif principal de cette étude était de discuter l'hypothèse posée sur la partie bruitée des sons par les modèles existants, en considérant cette composante comme du bruit blanc filtré. Un autre enjeu était d'apporter une justification à la représentation (souvent implicite) de sons bruités par une somme de sinusoïdes, qui est couramment employée sans être complètement expliquée. Nous devions pour cela essayer de mieux comprendre les travaux psychoacoustiques existants sur la perception des bruits.

Dans le chapitre 2, nous avons donc apporté une justification mathématique à la modélisation de sons bruités par une somme fixe de sinusoïdes d'amplitude et de fréquence constantes, en nous basant sur le modèle des bruits thermiques. Les travaux menés en psychoacoustique sur la faculté à discerner des bandes de bruit, que nous avons rappelés en détails, mettent en avant un nouveau paramètre perceptif : la densité spectrale. Nous avons discuté la définition de ce paramètre, et le lien entre la répartition des fréquences des sinusoïdes composant le signal et les éléments du signal relatifs à cette densité spectrale, perçus par l'oreille.

Nous avons ainsi montré l'intérêt d'une approche spectrale, mais également statistique. Nous avons proposé un nouveau modèle, appelé CNSS, définissant des paramètres mathématiques de contrôle de la distribution des fréquences du signal. Les liens avec la perception ont été détaillés, pour prouver l'intérêt musical des nouveaux contrôles autorisés par ce nouveau modèle. Nous avons également montré que le modèle CNSS permet des transformations d'échelle temporelle moins restrictives et de meilleure qualité que les modèles existants.

Dans le chapitre 3, nous avons décrit un principe d'analyse général envisagé pour les sons bruités. Celui-ci effectue une segmentation fréquentielle pour détecter les pics des spectres à court terme correspondant à des partiels. Ces pics doivent être traités par des méthodes haute résolution adaptées. Nous avons décrit les méthodes existantes, et nous avons proposé une méthode originale basée sur les modèles mathématiques développés par W. Hartmann, qui étudie statistiquement les fluctuations d'intensité de bandes fréquentielles du signal. Nous avons détaillé les algorithmes relatifs à cette nouvelle méthode, ainsi que les expériences que nous avons effectuées. Les résultats de ces expériences mettent en valeur la qualité de cette approche pour des sons bruités.

La méthode d'analyse générale effectue également une segmentation temporelle pour prendre en compte les transitoires (attaques) présentes dans le signal étudié. Nous avons décrit les méthodes de détection de transitoires existantes, et nous avons présenté une nouvelle méthode basée sur l'étude statistique de la répartition des échantillons du signal.

Ces deux segmentations permettent de définir les parties correspondant à du bruit. La densité spectrale de ces bandes de bruit doit être estimée. Nous avons discuté les limites de la transformée de Fourier pour une telle analyse. Nous avons ainsi proposé une méthode originale d'approximation de la densité spectrale de bandes de bruit, basée elle aussi sur l'étude statistique des fluctuations d'intensité. Nous avons détaillé un algorithme, et nous avons présenté quelques résultats d'expériences. Cette méthode peut ensuite être appliquée à l'analyse des paramètres du modèle CNSS. Nous avons décrit une méthode générale d'estimation de ces paramètres ainsi que les algorithmes associés. Les résultats que nous avons présentés, même s'ils mettent en évidence quelques limites, sont encourageants dans la mise au point d'une méthode d'analyse complète.

Dans le chapitre 4, nous avons justifié les méthodes de synthèse spectrale couramment employées, qui reposent sur des amplitudes fixées, des fréquences régulièrement réparties et des phases aléatoires. Nous avons détaillé les méthodes de synthèse liées au modèle CNSS, en décrivant les algorithmes de calcul des caractéristiques des sinusoïdes composant le signal à partir des paramètres mathématiques du modèle. Ensuite, nous avons présenté les algorithmes efficaces existants de calcul des échantillons du signal à partir des sinusoïdes ainsi définies.

Les études menées ont mis en évidence quelques problèmes spécifiques à la nature stochastique des sons bruités. Ainsi, nous avons démontré pourquoi l'utilisation de la technique classique de recouvrement induit des artefacts audibles lorsqu'elle est appliquée à de tels sons. Nous avons proposé plusieurs solutions et les algorithmes associés, que nous avons discutés.

Enfin, nous avons proposé une étude des solutions envisageables pour éliminer les problèmes liés à la saturation des sons bruités synthétisés.

Dans le chapitre 5, nous avons décrit les applications existantes du modèle CNSS. Nous avons développé, en collaboration avec Anthony Beurivé, une bibliothèque et un objet jMax, pour pouvoir confirmer la synthèse en temps-réel de sons bruités à partir des paramètres du modèle CNSS. Un outil expérimental d'exploration sonore est ainsi disponible librement [BHb]. Cette implémentation va également être intégrée dans l'outil pédagogique d'éveil musical pour les enfants Dolabip, développé au SCRIME.

Ensuite, dans ce même chapitre, nous avons proposé plusieurs idées pour améliorer le modèle, et en particulier les méthodes d'analyse. Nous avons également commencé à définir une nouvelle approche statistique de distribution logarithmique des fréquences, qui devrait permettre de réduire le nombre de sinusoïdes nécessaires pour synthétiser des sons bruités.

Nous avons également indiqué plusieurs axes de recherche envisageables pour l'avenir. Nous pensons à terme pour associer le modèle CNSS à un modèle sinusoïdal de haute précision de façon à pouvoir représenter et transformer une grande variété de sons naturels.

Les études que nous avons effectuées et qui sont récapitulées dans ce document concernent un type de sons qui n'a fait l'objet que de très peu de travaux de recherche. Les théories que nous décrivons dans ce document servent avant tout à justifier les méthodes couramment employées. Ensuite, elles précisent les limites de ces méthodes, qui ne considèrent les sons bruités que comme des bruits blancs filtrés. Cette hypothèse apparaît trop restrictive et le modèle CNSS que nous avons développé constitue désormais une alternative possible.

Comme la qualité des modèles sinusoïdaux, pouvant être appliqués aux sons instrumentaux peu bruités, a atteint un niveau très intéressant, nous pensons qu'un des enjeux principaux des prochaines recherches pour la modélisation informatique des sons concernera la représentation de la composante bruitée de sons naturels. Nous espérons que ce document, et notamment le modèle CNSS qui y est décrit, pourra servir de base à ces futurs travaux.

Annexe A

Résultats d'expériences pour la détection de partiels

Cet annexe présente des résultats d'expériences concernant la méthode de détection spectrale de partiels, présentée dans la section 3.2.

Méthode des phases aléatoires								
fenêtre w_3	512				256			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.55	3.394	0.931	2.463	0.069	17.168	1.000	16.168	0.000
0.45	2.781	0.729	2.052	0.271	3.600	1.000	2.601	0.001
0.375	2.439	0.621	1.819	0.379	3.292	0.931	2.361	0.069
Méthode des phases constantes								
fenêtre w_3	512				256			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
1.8	2.969	0.631	2.339	0.369	6.063	1.000	5.063	0.000
1.6	2.786	0.584	2.202	0.416	4.691	0.999	3.691	0.001
1.4	2.563	0.541	2.021	0.459	4.000	0.993	3.007	0.007

TAB. A.1 – Résultats de détections pour une sinusoïde pure stationnaire, avec une fenêtre de variance w_3 de tailles 512 et 256 échantillons (SNR = 0dB).

fenêtre w_3	512				1024			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.3	9.843	1.000	8.843	0.000	4.379	1.000	3.379	0.000
0.2	3.478	0.997	2.481	0.003	2.041	1.000	1.041	0.000
0.15	1.943	0.959	0.984	0.041	1.829	1.000	0.829	0.000

TAB. A.2 – Résultats de détections d'une sinusoïde pure stationnaire par la méthode de variation de taille, avec une fenêtre de variance w_3 de tailles 512 et 1024 échantillons (SNR = 0dB).

méthode des phases aléatoires								
fenêtre w_2	512				1024			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.525	5.857	0.921	4.936	0.079	8.397	1.000	7.397	0.000
0.475	2.949	0.829	2.120	0.171	3.931	1.000	2.931	0.000
0.4	1.714	0.701	1.013	0.299	3.393	0.975	2.419	0.025
0.375	1.542	0.665	0.877	0.335	3.292	0.931	2.361	0.069
Méthode des phases constantes								
fenêtre w_2	512				1024			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
1.8	12.507	0.851	11.656	0.149	6.063	1.000	5.063	0.000
1.5	5.668	0.718	4.950	0.282	4.245	0.998	3.247	0.002
1.4	4.327	0.686	3.641	0.314	4.000	0.993	3.007	0.007
1.2	2.553	0.607	1.945	0.393	3.683	0.937	2.745	0.063

Tab. A.3 – Résultats de détection d'une sinusoïde pure stationnaire, avec une fenêtre de modifications spectrales w_2 de tailles 512 et 1024 échantillons (SNR = 0dB).

fenêtre w_2	2048				4096			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.3	4.164	1.000	3.164	0.000	4.379	1.000	3.379	0.000
0.2	2.059	1.000	1.059	0.000	2.041	1.000	1.041	0.000
0.15	1.841	0.999	0.842	0.001	1.829	1.000	0.829	0.000

Tab. A.4 – Résultats de détection d'une sinusoïde pure stationnaire par la méthode de variation de taille, avec une fenêtre de modifications spectrales w_2 de tailles 2048 et 4096 échantillons (SNR = 0dB).

nombre réal.	25			50			100		
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$
0.55	35.551	0.998	34.553	24.889	1.000	23.889	17.055	1.000	16.055
0.5	12.927	0.994	11.933	7.141	0.999	6.141	4.829	1.000	3.829
0.45	4.936	0.984	3.952	3.493	0.995	2.498	3.259	1.000	2.259
0.4	3.150	0.953	2.197	2.997	0.974	2.023	3.013	0.987	2.027

Tab. A.5 – Résultats de détection d'une sinusoïde pure stationnaire (SNR = 6dB) obtenus à partir de 25, 50 et 100 spectres de phase (nombre de réalisations).

nombre réal.	25			50			100		
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$
0.25	9.422	1.000	8.422	9.724	1.000	8.724	10.054	1.000	9.054
0.2	4.926	1.000	3.926	4.986	1.000	3.986	5.243	1.000	4.243
0.15	2.417	0.998	1.419	2.470	0.999	1.471	2.590	0.998	1.592
0.1	1.293	0.895	0.398	1.311	0.902	0.409	1.319	0.881	0.437

TAB. A.6 – Résultats de détection d'une sinusoïde pure stationnaire (SNR = 0dB) obtenus à partir de 25, 50 et 100 spectres de phase (nombre de réalisations).

Méthode des phases constantes								
fenêtre w_2	16 bins				12 bins			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
1.9	5.025	0.999	4.027	0.001	7.403	1.000	6.403	0.000
1.7	4.985	0.989	3.993	0.011	4.962	0.999	3.963	0.001
1.4	4.816	0.916	3.901	0.084	4.000	0.993	3.007	0.007
1.2	4.538	0.788	3.749	0.212	3.683	0.937	2.745	0.063
Random phase method								
bande fréq. Δf	16 bins				12 bins			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.55	5.283	1.000	4.283	0.000	17.168	1.000	16.168	0.000
0.525	4.936	1.000	3.936	0.000	8.397	1.000	7.397	0.000
0.5	4.845	0.999	3.845	0.001	5.033	1.000	4.033	0.000
0.45	4.774	0.996	3.778	0.004	3.600	1.000	2.601	0.001

TAB. A.7 – Résultats de détection d'une sinusoïde pure stationnaire avec des largeurs de bande de fréquence Δf de 16 et 12 casiers (SNR = 0dB).

bande fréq. Δf	4 bins				8 bins			
seuils	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$	$\overline{N}_{\text{sinus}}$	$\overline{N}_{\text{exact}}$	$\overline{N}_{\text{faux}}$	$\overline{N}_{\text{manque}}$
0.25	10.054	1.000	9.054	0.000	2.711	1.000	1.711	0.000
0.2	5.243	1.000	4.243	0.000	2.041	1.000	1.041	0.000
0.15	2.590	0.998	1.592	0.002	1.829	1.000	0.829	0.000

TAB. A.8 – Résultats de détection d'une sinusoïde pure stationnaire avec des largeurs de bande de fréquence Δf de 4 et 8 casiers (SNR = 0dB).

Bibliographie

[AE03] M. Athineos and D.P.W. Ellis. Sound texture modelling with linear prediction in both time and frequency domains. *Proceedings of the IEEE International Conference On Acoustics, Speech, and Signal Processing (ICASSP'03), Hong-Kong, China*, 2003.

[AF95] F Auger and P. Flandrin. Improving the readibility of time-frequency and time-scale representations by the reassignment mthod. *IEEE Transactions on Signal Processing*, 43(5) :1068–1089, 1995.

[Bas89] M. Basseville. Distance measures for signal processing and pattern recognition. *European Journal Signal Processing*, 18(4) :349–369, 1989.

[Beu] A. Beurivé. *Dolabip.* Scrime, University Bordeaux 1, France. `http ://scrime.labri.fr/logiciels/dolabip/`.

[BG97] G.P. Bernardi, A. Bugna and De Poli G. Musical analysis with chaos. In *Musical signal processing*, pages 187–220. RoadsSwets and Zeitlinger, 1997.

[BHa] A. Beurivé and P. Hanna. *CNSS Library.* Scrime, University Bordeaux 1, France. `http ://scrime.labri.fr/logiciels/`.

[BHb] A. Beurivé and P. Hanna. *CNSS Object for Jmax.* Scrime, University Bordeaux 1, France. `http ://scrime.labri.fr/logiciels/`.

[Bol83] T. Bolognesi. Automatic composition : Experiments with self-similar music. *Computer Music Journal*, 7(1) :25–36, 1983.

[Cad02] C. Cadoz. Le modèle physique, métaphore pour la création musicale. *Proceedings of the Journées d'Informatique Musicale (JIM'02), Marseille, France*, pages 59–69, 2002.

[CC00] N. Castagne and C. Cadoz. Physical modeling synthesis : balance between realism and computing speed. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'00), Verona, Italy*, pages 217–222, 2000.

[Cha96] J. Chazarain. *Programmer avec Scheme.* International Thomson Publishing, 1996.

[Cho73] J.M. Chowning. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7) :526–534, 1973.

[Coo97] P.R. Cook. Physically informed sonic modeling (phism) :synthesis of percussive sounds. *Computer Music Journal*, 21(3) :38–49, 1997.

[DCKS+03] Myriam Desainte-Catherine, György Kurtag, Catherine Semal, Sylvain Marchand, and Pierre Hanna. Playing electro-acoustic music like playing video games. *ACM Computers in Entertainment*, 2003.

[DCM99a] M. Desainte-Catherine and S. Marchand. Structured additive synthesis : Towards a model of sound timbre and electroacoustic music forms. *Proceedings of International Computer Music Conference (ICMC'99), Beijing, China*, pages 260–263, 1999.

[DCM99b] M. Desainte-Catherine and S. Marchand. Vers un modèle pour unifier musique et son dans une composition multiéchelle. *Proceedings of the Journées d'Informatique Musicale (JIM'99), Paris, France*, 1999.

[DCM00] M. Desainte-Catherine and S. Marchand. High precision fourier analysis of sounds using signal derivatives. *Journal of Audio Engineering Society*, 48(7/8) :654–667, 2000.

[DGR93] P. Depalle, G. Garcia, and X. Rodet. Tracking of partials for additive sound synthesis using hidden markov models. *Proceedings of the IEEE International Conference On Acoustics, Speech, and Signal Processing (ICASSP'93), Minneapolis, Minnesota, USA*, 1993.

[Dol86] M. Dolson. The phase vocoder : A tutorial. *Computer Music Journal*, 10(4) :14–27, 1986.

[DQ96] D. Declercq and A. Quinquis. *Le signal aléatoire*. Hermès, 1996.

[DT97] S. Dubnov and N. Tishby. Analysis of sound textures in musical and machine sounds by means of higher order statistical features. *Proceedings of the IEEE International Conference On Acoustics, Speech, and Signal Processing (ICASSP'97), Munich, Germany*, 1997.

[Dub96] S. Dubnov. *Polyspectral analysis of musical timbre*. PhD thesis, Hebrew University, 1996.

[Ell] D. Ellis. *Matlab codes for sinusoidal modeling*. Columbia University, New-York, USA. `http ://www.ee.columbia.edu/ dpwe/resources/matlab/sinemodel/`.

[Erd86] J. Erdreich. A distribution based definition of impulse noise. *Journal of Acoustical Society of America*, 79(4) :990–998, 1986.

[FG66] J.L. Flanagan and R.M. Golden. Phase vocoder. *Bell System Technical Journal*, pages 1493–1509, November 1966.

[FH96] K. Fitz and L. Haken. Sinusoidal modeling and manipulation using lemur. *Computer Music Journal*, 20(4) :44–59, 1996.

[FHC] K. Fitz, L. Haken, and P. Christensen. *Loris Software*. CERL Sound Group, University of Illinois, USA. `http ://www.cerlsoundgroup.org/Loris/`.

[FHC00] K. Fitz, L. Haken, and P. Christensen. A new algorithm for bandwidth association in bandwidth-enhanced additive sound modeling. *Proceedings of International Computer Music Conference (ICMC'00), Berlin, Germany*, pages 178–181, 2000.

[FHH$^+$] K. Fitz, L. Haken, B. Holloway, E. Tellman, and B. Walker. *Lemur Software*. CERL Sound Group, University of Illinois, USA. `http ://www.cerlsoundgroup.org/Lemur/`.

[Fit99] K. Fitz. *The Reassigned Bandwidth-Enhanced Method of Additive Synthesis*. PhD thesis, Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1999.

[FJ98] M. Frigo and S.G. Johnson. *FFTW User's Manual*, 1998.

[Fre99] A. Freed. Spectral line broadening with transform domain additive synthesis. *Proceedings of International Computer Music Conference (ICMC'99), Beijing, China*, pages 78–81, 1999.

[Gal98] E. Gallesio. *Stk Reference Manual*, 1998.

[Ger78] A. Gerzso. Density of spectral components : preliminary experiments. report, Ircam, 1978.

[Goo96] M. Goodwin. Residual modeling in music analysis-synthesis. *Proceedings of the IEEE International Conference On Acoustics, Speech, and Signal Processing (ICASSP'96), Atlanta, GA, USA*, pages pp. 1005–1008, 1996.

[Goo97] M. Goodwin. *Adaptive Signal Models : Theory, Algorithms and Audio Applications*. PhD thesis, Berkeley, University of California, 1997.

[GP99] G. Garcia and J. Pampin. Data compression of sinusoidal modeling parameters based on psychoacoustic masking. *Proceedings of International Computer Music Conference (ICMC'99), Beijing, China*, pages 40–43, 1999.

[Har87] W.M. Hartmann. Temporal fluctuations and the discrimination of spectrally dense signals by human listeners. In W.A. Yost and C.S. Watson, editors, *Auditory Processing of Complex Sounds*, pages 126–135. Erlbaum, Hillsdale, NJ, 1987.

[Har97] W.M. Hartmann. *Signals, Sound, and Sensation*. Modern Acoustics and Signal Processing AIP Press, 1997.

[HBDC02] P. Hanna, A. Beurivé, and M. Desainte-Catherine. Real-time noise synthesis with control of the spectral density. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'02), Hamburg, Germany*, pages 151–156, 2002.

[HDC00] P. Hanna and M. Desainte-Catherine. Statistical approach for sound modeling. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'00), Verona, Italy*, pages 91–96, 2000.

[HDC01] P. Hanna and M. Desainte-Catherine. Influence of frequency distribution on intensity fluctuations of noise. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX'01), Limerick, Ireland*, pages 125–129, 2001.

[HDC02a] P. Hanna and M. Desainte-Catherine. Adapting the overlap-add method to the synthesis of noise. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'02), Hamburg, Germany*, pages 101–104, 2002.

[HDC02b] P. Hanna and M. Desainte-Catherine. Detection of sinusoidal components in sounds using statistical analysis of intensity fluctuations. *Proceedings of International Computer Music Conference (ICMC'02), Göteborg, Sweden*, pages 100–103, 2002.

[HDC02c] P. Hanna and M. Desainte-Catherine. Influence de la densité spectrale sur la synthèse de sons bruités. *Proceedings of the Journées d'Informatique Musicale (JIM'02), Marseille, France*, pages 17–24, 2002.

[HDC03a] P. Hanna and M. Desainte-Catherine. Analysis method to approximate the spectral density of noises. *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03), Mohonk, New-Paltz, NY.*, 2003.

[HDC03b] P. Hanna and M. Desainte-Catherine. Cnss model : a statistical and spectral model for noise. *Internal report, LaBRI, University of Bordeaux 1*, 2003.

[HDC03c] P. Hanna and M. Desainte-Catherine. Time scale modification of noises using a spectral and statistical model. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), Hong Kong, China*, 2003.

[HDC03d] P. Hanna and M. Desainte-Catherine. Using statistical analysis of the intensity fluctuations to detect sinusoids in noisy signals. Technical report, LaBRI, University of Bordeaux 1, 2003.

[HMGB86] W.M. Hartmann, S. McAdams, A. Gerzso, and P. Boulez. Discrimination of spectral density. *Journal of Acoustical Society of America*, 79(6) :1915–1925, 1986.

[HP88] W.M. Hartmann and J. Pumplin. Noise power fluctuations and the masking of sine signals. *Journal of Acoustical Society of America*, 83(6) :2277–2289, 1988.

[HP91] W.M. Hartmann and J. Pumplin. Periodic signals with minimal power fluctuations. *Journal of Acoustical Society of America*, 90(4) :1986–1999, 1991.

[Jaf95] D.A. Jaffe. Ten criteria for evaluating synthesis techniques. *Computer Music Journal*, 19(1) :76–87, 1995.

[jma] jmax. `http ://www.ircam.fr/jmax`.

[KAU00] F. Keiler, D. Arfib, and Zölzer U. Efficient linear prediction for digital audio effects. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'00), Verona, Italy*, pages 19–24, 2000.

[KD98] C. Kaernbach and L. Demany. Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *J. Acoust. Soc. Am.*, 104(4) :2298–2306, 1998.

[KM88] R. Kronland-Martinet. The wavelet transform for analysis, synthesis and processing of speech and music sounds. *Computer Music Journal*, 12(4) :11–20, 1988.

[KM02] F. Keiler and S. Marchand. Survey on extraction of sinusoids in stationary sounds. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'02), Hamburg, Germany*, pages 51–58, 2002.

[KZ01] F. Keiler and U. Zölzer. Extracting sinusoids from harmonic signals. *Journal of New Music Research, Special Issue :"Musical Applications of Digital Signal Processing"*, 30(3) :243–258, September 2001.

[LD99] J. Laroche and M. Dolson. New phase vocoder technique for pitch-shifting, harmonizing and other exotic effects. *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'99), Mohonk, New-Paltz, NY.*, pages 91–94, 1999.

[Lev98] S. Levine. *Audio Representations for Data Compression and Compressed Domain Processing*. PhD thesis, CCRMA, Stanford University, 1998.

[LG00] N. Laurenti and De Poli G. A method for spectrum separation and envelope estimation of the residual in spectrum modeling of musical sound. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'00), Verona, Italy*, pages 233–236, 2000.

[LM01] M. Lagrange and S. Marchand. Real-time additive synthesis of sound by taking advantage of psychoacoustics. *Proceedings of the COST G-6 conference on Digital Audio Effects (DAFx'01), Limerick, Ireland*, pages 5–9, 2001.

[LMR02] M. Lagrange, S. Marchand, and J.B. Rault. Sinusoidal parameter extraction and component selection in a non-stationnary model. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'02), Hamburg, Germany*, pages 59–64, 2002.

[LMRR03] M. Lagrange, S. Marchand, M. Raspaud, and J.B. Rault. Enhanced partial tracking using linear prediction. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'03), London, UK*, 2003.

[Mak75] J. Makhoul. Linear prediction : a tutorial review. *Proceedings of the IEEE*, 63(4) :561–580, April 1975.

[Mar] S. Marchand. *InSpect, ReSpect and ProSpect Softwares.* Scrime, University Bordeaux 1, France. `http ://scrime.labri.fr/logiciels/`.

[Mar99] S. Marchand. Musical sound effects in the SAS model. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'99), Trondheim, Norway*, 1999.

[Mar00] S. Marchand. *Sound models for computer music : analysis, transformation, synthesis of musical sound.* PhD thesis, LaBRI, Université Bordeaux I, 2000.

[Mas96] P. Masri. *Computer Modelling of sound.* PhD thesis, University of Bristol, UK, 1996.

[MB96] P. Masri and A. Bateman. Improved modelling of attack transients in music analysis-resynthesis. *Proceedings of International Computer Music Conference (ICMC'96), Hong-Kong, China*, pages 100–103, 1996.

[MC98] P. Masri and C.N. Canagarajah. Extracting more detail from the spectrum with phase distortion analysis. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'98), Barcelona, Spain*, pages 119–122, 1998.

[Min98] N.E. Miner. *Creating Wavelet-based Models for Real-Time Synthesis of Perceptually Convincing Environmental Sounds.* PhD thesis, University of New Mexico, 1998.

[Moo78] J.A. Moorer. The use of the phase vocoder in computer music applications. *Journal of the Audio Engineering Society*, 26(1), 1978.

[Moo90] F.R. Moore. *Elements for Computer Music.* Prentice Hall, 1990.

[Mou90] E. Moulines. *Algorithmes de codage et de modification des paramètres prosodiques pour la synthèse de la parole à partir du texte.* PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France, 1990.

[MQ86] R.J. McAulay and T.F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(4) :744–754, 1986.

[MS99] S. Marchand and R. Strandh. Inspect and respect : Spectral modeling, analysis and real-time synthesis software tools for reseachers and composers. *Proceedings of International Computer Music Conference (ICMC'99), Beijing, China*, pages 341–344, 1999.

[Orf96] S.J. Orfanidis. *Introduction to Signal Processing.* Prentice Hall, 1996.

[Pee01] G. Peeters. *Modèles et Modification du signal sonore adaptés à ses caractéristiques locales.* PhD thesis, IRCAM, Université de Paris 6, 2001.

[Por76] M.R. Portnoff. Implementation of the digital phase vocoder using the fast fourier transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(3) :243–248, 1976.

[PR98a] G. Peeters and X. Rodet. Analyse et synthèse de sons musicaux par la méthode psola. *Proceedings of the Journées d'Informatique Musicale (JIM'98), La Londe-les-Maures, France)*, 1998.

[PR98b] G. Peeters and X. Rodet. Signal characterization in terms of sinusoidal and non-sinusoidal components. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'98), Barcelona, Spain*, 1998.

[PR99] Geoffroy Peeters and Xavier Rodet. Sinola : A new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum. *Proceedings of International Computer Music Conference (ICMC'99), Beijing, China*, pages 153–156, 1999.

[Pre00] D. Pressnitzer. Modèles psychoacoustiques et perception de hauteur. *JIM Proceedings*, 2000.

[Pum85] J. Pumplin. Low-noise noise. *Journal of Acoustical Society of America*, 78(1) :100–104, 1985.

[RD92a] X. Rodet and P. Depalle. A new additive synthesis method using inverse fourier transform and spectral envelopes. *Proceedings of International Computer Music Conference (ICMC'92), San Jose, California, USA*, pages 410–411, 1992.

[RD92b] X. Rodet and P. Depalle. Spectral envelopes and inverse fft synthesis. *AES'92, San Fransisco, USA*, 1992.

[RJ01] X. Rodet and F. Jaillet. Detection and modeling of fast attack transients. *Proceedings of International Computer Music Conference (ICMC'01), Havana, Cuba*, 2001.

[Roa97] C. Roads. Sound tranformation by convolution. In *Musical signal processing*, pages 411–438. RoadsSwets and Zeitlinger, 1997.

[Roa98] C. Roads. *L'Audionumérique (trad. The Computer Music Tutorial)*. Dunod, 1998.

[Rod97a] X. Rodet. Musical sound signals analysis/synthesis : Sinusoidal+residual and elementary waveform models. *Proceedings of the IEEE Time-Frequency and Time-Scale Workshop (TFTS'97), University of Warwick, Coventry, UK*, 1997.

[Rod97b] X. Rodet. Musical sound signals analysis/synthesis : Sinusoidal+residual and elementary waveform models. *Applied Signal Processing*, 4 :131–141, 1997.

[SA99] J.O. Smith and J.S. Abel. Bark and erb bilinear transforms. *IEEE Transactions on Speech and Audio Processing*, 7(6) :697–708, November 1999.

[Sch66] P. Schaeffer. *Traité des objets musicaux.* Seuil, 1966.

[Sch70] M.R. Schroeder. Synthesis of low peak factor signals and binary sequences with low autocorrelation. *IEEE Transactions on Information theory*, 16 :85–89, 1970.

[Sch98] D. Schwarz. Spectral envelopes in sound analysis and synthesis. Diplomarbeit nr. 1622, Universität Stuttgart, Fakultät Informatik, 1998.

[scr] Scrime. `http ://www.scrime.labri.fr`.

[SD91] C. Semal and L. Demany. Dissociation of pitch from timbre in auditory short-term memory. *J. Acoust. Soc. Am.*, 89 :2404–2410, 1991.

[SDS98] Y. Stylianou, T. Dutoit, and J. Schroeter. Diphone concatenation using a harmonic plus noise model of speech. *Proceedings of the Eurospeech Conference (EUROSPEECH'97), Rhodes, Greece*, pages 613–616, 1998.

[Ser89] X. Serra. *A System for Sound Analysis/Transformation/Synthesis Based on a DEterministic plus Stochastic Decomposition*. PhD thesis, CCRMA, Stanford University, 1989.

[Ser97a] M.H. Serra. Introducing the phase vocoder. In *Musical signal processing*, pages 31–90. RoadsSwets and Zeitlinger, 1997.

[Ser97b] X. Serra. Musical sound modeling with sinusoids plus noise. In *Musical signal processing*, pages 91–122. RoadsSwets and Zeitlinger, 1997.

[Ser97c] X.. Serra. *Spectral Modeling Synthesis Tools*, 1997.

[SM99] R. Strandh and S. Marchand. Real-time generation of sound from parameters of additive synthesis. *Proceedings of the Journées d'Informatique Musicale (JIM'99), Paris, France*, pages 83–88, 1999.

[Smi97] J.O. Smith. Acoustic modeling using digital waveguides. In *Musical signal processing*, pages 221–263. RoadsSwets and Zeitlinger, 1997.

[Smi00] J.O. Smith. Digital audio resampling home page. Technical report, CCRMA, Stanford University, `http ://ccrma-www.stanford.edu/~jos/resample/`, 2000.

[Smi03] *Digital Waveguide Modeling of Musical Instruments*. Department of Music, Stanford University, `http ://www-ccrma.stanford.edu/~jos/waveguide/`, 2003. Online Book.

[SR99] D. Schwarz and X. Rodet. Spectral envelope estimation and representation for sound analysis-synthesis. *Proceedings of International Computer Music Conference (ICMC'99), Beijing, China*, pages 351–354, 1999.

[SS87] J.O. Smith and X. Serra. PARSHL : an analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation. *Proceedings of International Computer Music Conference (ICMC'87), San Fransisco, USA*, 1987.

[SS90] X. Serra and J. Smith. Spectral modeling synthesis : a sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4) :12–24, 1990.

[SSG+98] A.K. Syrdal, Y.G. Stylianou, L.F. Garrison, A. Conkie, and J. Schroeter. TD-PSOLA versus harmonic plus noise model in diphone based speech synthesis. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'98), Seattle, USA*, pages 273–276, 1998.

[Sty98] Y. Stylianou. Concatenative speech synthesis using a harmonic plus noise model. *Proceedings of the 3rd ESCA/COCOSDA Workshop on Speech Synthesis, Jenolan Caves, Australia*, pages 613–616, 1998.

[TG01] H. Thornburg and F. Gouyon. A flexible analysis/synthesis method for transients. *Proceedings of International Computer Music Conference (ICMC'00), Berlin, Germany*, pages 400–403, 2001.

[VC78] R.F. Voss and J. Clarke. 1/f noise in music : music from 1/f noise. *Journal of Acoustical Society of America*, 63(1) :258–263, 1978.

[VLM99] T. Verma, S. Levine, and T. Meng. Transient modeling synthesis : a flexible analysis/synthesis tool for transient signals. *Proceedings of International Computer Music Conference (ICMC'97), Thessaloniky, Greece*, 1999.

[VM98] T.S. Verma and T.H.Y. Meng. Time scale modification using a sines+transient+noise signal model. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX'98), Barcelona, Spain*, pages 49–52, 1998.

[VM00] T.S. Verma and T.H.Y. Meng. Extending spectral modeling synthesis with transient modeling synthesis. *Computer Music Journal*, 24(2) :47–59, 2000.

[Wal23] J. Walsh. A closed set of normal orthogonal functions. *American Journal of Mathematics*, 45 :5–24, 1923.

[WPDC98] L. Wiegrebe, R.D. Patterson, L. Demany, and R.P. Carlyon. Temporal dynamics of pitch strength in regular interval noises. *Journal of Acoustical Society of America*, 104(4) :2307–2313, 1998.

[Yos96] W.A. Yost. Pitch of iterated rippled noise. *Journal of Acoustical Society of America*, 100(1) :3329–3335, 1996.

[ZF99] E Zwicker and H. Fastl. *Psychoacoustics : facts and models.* Springer, 1999.

Index

V