

Spectrogramme d'Amplitude et de Fréquence Instantanées (SAFI)

Valentin EMIYA

DEA d'Acoustique,
Traitement du signal et Informatique Appliqués à la Musique
Université de la Méditerranée, Aix-Marseille II

Mémoire de stage
Mars-Juillet 2004



Responsable de stage : Vincent GIBIAT
Laboratoire d'Acoustique, de Métrologie et d'Instrumentation (LAMI)



Remerciements

Je tiens à remercier M. Vincent GIBIAT pour m'avoir donné l'opportunité de réaliser ce stage et avoir contribué à son bon déroulement. Merci également à Lionel Cros, élève du DEA SIA, avec qui j'ai étroitement travaillé sur le SAFI, pour le bon esprit dans lequel s'est déroulée notre collaboration.

Je remercie par ailleurs très chaleureusement pour leur accueil et leur sympathie l'ensemble des personnes que j'ai pu côtoyer au sein du LAMI.

Enfin, pour le temps qu'ils m'ont accordé, leur aide et leur soutien, un grand merci à Jeanne Clerc-Renaud, Abril Padilla, Solange Barrachina et Emmanuel Favreau (INA-GRM), Bertrand David (ENST), Pascal Gaillard (UTM), Cyrille Defaye, Gérard Assayag et Daniel Pressnitzer (IRCAM).

Résumé

Le présent rapport décrit l'élaboration et les propriétés du Spectrogramme d'Amplitude et de Fréquence Instantanées (SAFI). Cet outil de représentation temps-fréquence utilise le concept de fréquence instantanée pour mettre à profit les informations bien souvent négligées contenues dans la phase du signal. L'un des objectifs est d'apporter un renouveau dans le domaine des représentations temps-fréquence en améliorant notablement l'analyse en fonction des moyens théoriques et techniques disponibles aujourd'hui. D'autre part, ce nouvel outil ouvre la voie à d'autres formes d'analyse et d'interprétation comme l'analyse sur des bandes fréquentielles larges. Les premiers résultats intéressants obtenus en analysant des sons synthétiques et réels illustreront les capacités du SAFI.

Mot-clés :

Représentation temps-fréquence, transformée de Fourier à court terme, spectrogramme, analyse spectrale différentielle, amplitude instantanée, fréquence instantanée, phase instantanée, modulation, perception.

Abstract

This report describes the design and properties of an instantaneous amplitude and frequency based spectrogram (IAFS). The IAFS is a time-frequency representation that relies on instantaneous frequency concept in order to extract information from phase. This new time-frequency representation tool aims at improving sound analysis thanks to current theoretical and technical means. It also enables to develop other principles of analysis and understanding like wide-band analysis. Some noticeable results obtained with synthesized or real sounds illustrate the properties of the IAFS.

Keywords :

Time-frequency representation, spectral differential analysis, short time Fourier transform, spectrogram, instantaneous amplitude, instantaneous frequency, instantaneous phase, modulation, perception.

Table des matières

Remerciements	i
Résumé	iii
Liste des figures	viii
Introduction	1
1 Théorie	3
1.1 Rappels	3
1.1.1 Transformée de Fourier	3
1.1.2 Transformée de Fourier à Court Terme	3
1.1.3 Signal analytique	4
1.1.4 Amplitude et fréquence instantanées	4
1.2 Analyse Spectrale Différentielle (ASD)	5
1.2.1 ASD à temps continu	5
1.2.2 ASD à temps discret	6
1.2.3 Intérêt de l'ASD	6
1.2.4 Expression des coefficients de l'ASD appliquée au signal analytique	7
1.3 Etude de l'ASD	8
1.3.1 Signal sinusoïdal	8
1.3.2 Signal à fréquence variable	9
1.3.3 Signal à deux composantes sinusoïdales : phénomènes inter-bandes	10
1.3.4 Signal à plusieurs composantes : phénomènes intra-bandes	11
1.3.5 Estimation de l'amplitude instantanée	14
1.3.6 Fenêtre de pondération	15
1.4 Échelle fréquentielle ajustable	16
2 Spectrogramme d'Amplitude et de Fréquence Instantanées (SAFI)	17
2.1 Principe	17

2.2	Construction du SAFI	17
2.3	Paramètres	19
2.4	Complexité	20
3	Analyse de sons	21
3.1	Analyses de sons synthétiques	21
3.1.1	Chirp de durée variable	21
3.1.2	Analyse en bande large : modulations	24
3.1.3	Analyse en bande large : perception de la rugosité	25
3.2	Analyses de sons réels	31
3.2.1	Sons de steel drum	31
3.2.2	Sons de trompette marine	34
3.2.3	Sons de RER	36
	Conclusion	39
	Annexes	III
A	Développements mathématiques	III
A.1	Passage aux coefficients de l'ASD en discret	III
A.2	Coefficients ASD d'un signal réel	VI
A.3	Signal à deux composantes sinusoïdales : phénomènes inter-bandes	VI
B	Comparatif des différents spectrogrammes	IX
B.1	Représentations de différents algorithmes temps-fréquence appliqués au signal test	IX
B.1.1	Robustesse au bruit	XVII
B.1.2	Etude comparative de la complexité de calcul	XVIII
C	Echelle fréquentielle logarithmique : principe de construction	XXV
C.1	Principe de l'échelle logarithmique	XXV
C.2	Construction de l'échelle logarithmique	XXV
D	Acronymes et notations mathématiques	XXVII
E	Description de l'algorithme basé sur l'ASD	XXIX

Table des figures

1.1	<i>Portion de signal vue à travers une fenêtre rectangulaire glissante.</i>	5
1.2	<i>Comportement de $X_n(t)$ pour différentes valeurs de la fréquence du signal analysée (f_0, f_1, f_2) et proches de $\frac{n}{T}$.</i>	9
1.3	<i>Signal à deux composantes à 440Hz et 550Hz et analyse monobande.</i>	12
2.1	<i>Initialisation du programme (en haut) et bloc d'estimation de l'amplitude et de la fréquence instantanées (en bas).</i>	18
2.2	<i>Aperçu de l'interface graphique</i>	19
3.1	<i>Chirp élémentaire (2s)</i>	21
3.2	<i>Spectrogramme du chirp de durée variable (FFT sur 8192 points avec une fenêtre de Blackman-Harris de 2048 points et un recouvrement de 1968 points), avec les détails de la partie critique (figure du bas).</i>	22
3.3	<i>SAFI du chirp accéléré : représentation en 2D (à gauche), détail des chirps les plus rapides (à droite), et représentation en 3D (en bas)</i>	23
3.4	<i>SAFI d'une somme de deux sinusoides</i>	24
3.5	<i>Etude de la rugosité : signaux temporels</i>	25
3.6	<i>Etude de la rugosité : amplitude et fréquence instantanées pour $\varphi_0 = 0$</i>	27
3.7	<i>Etude de la rugosité : amplitude et fréquence instantanées pour $\varphi_0 = \frac{\pi}{6}$</i>	28
3.8	<i>Etude de la rugosité : amplitude et fréquence instantanées pour $\varphi_0 = \frac{2\pi}{6}$</i>	29
3.9	<i>Etude de la rugosité : amplitude et fréquence instantanées pour $\varphi_0 = \frac{\pi}{2}$</i>	30
3.10	<i>Spectrogramme classique du son de steel drum</i>	31
3.11	<i>SAFI du son de steel drum</i>	33
3.12	<i>SAFI trompette marine</i>	35
3.13	<i>Spectrogramme classique</i>	36
3.14	<i>SAFI RER</i>	37
B.1	<i>Répartition idéale du signal test $x_t(k)$ dans le plan temps-fréquence pour \tilde{f} allant de 0 à 0.25</i>	XII
B.2	<i>SAFI de x_t.</i>	XIII
B.3	<i>Spectrogramme classique de x_t.</i>	XIV

B.4	<i>Spectrogramme classique de x_t réalloué.</i>	XIV
B.5	<i>RTF de la distribution de Wigner-Ville de x_t.</i>	XV
B.6	<i>RTF de la pseudo distribution de Wigner-Ville de x_t.</i>	XV
B.7	<i>RTF de la distribution pseudo Wigner-Ville lissée de x_t.</i>	XVI
B.8	<i>RTF de la distribution pseudo Wigner-Ville de x_t lissée et réallouée.</i>	XVII
B.9	<i>SAFI de x_t.</i>	XVIII
B.10	<i>Spectrogramme classique de x_t.</i>	XIX
B.11	<i>Spectrogramme classique de x_t réalloué.</i>	XIX
B.12	<i>RTF de la distribution de Wigner-Ville de x_t.</i>	XX
B.13	<i>RTF de la pseudo distribution de Wigner-Ville de x_t.</i>	XX
B.14	<i>RTF de la distribution pseudo Wigner-Ville lissée de x_t.</i>	XXI
B.15	<i>RTF de la distribution pseudo Wigner-Ville de x_t lissée et réallouée.</i>	XXI

Introduction

Les représentations d'un signal dans le plan temps-fréquence donnent une description intuitive des signaux non-stationnaires dont le contenu fréquentiel varie au cours du temps et comportant des régimes transitoires. Avec les outils et les moyens de calcul actuels, il est donc d'un intérêt certain de s'appuyer sur ce type de représentations.

La transformée de Fourier représente sous forme fréquentielle l'information contenue dans le signal. La dimension temporelle n'apparaît pas compte tenu de la dualité des espaces temps et fréquence. La transformée de Fourier à court terme permet de réintroduire une dimension temporelle en effectuant une succession d'analyses locales du signal observé à travers une fenêtre. Il est alors possible de représenter la distribution de l'énergie du signal dans le plan temps-fréquence via le carré du module de la transformée de Fourier à court terme. Cette représentation, appelée spectrogramme, est la plus ancienne et la plus répandue. D'autres représentations quadratiques ont vu le jour, comme la distribution de Wigner-Ville.

L'inconvénient de ces outils qui représentent la distribution de l'énergie du signal est que l'information de phase est perdue. Or, certaines expériences et études mettent en évidence que l'oreille humaine est extrêmement sensible aux informations portées par la phase. Cette affirmation va à l'encontre de l'idée encore très répandue selon laquelle l'information perceptive est essentiellement voire totalement portée par l'amplitude, ce qui explique historiquement que le développement des représentations temps-fréquence ne prend en compte que l'amplitude.

L'outil élaboré utilise l'information de phase pour estimer la fréquence instantanée dans chaque bande d'analyse par analyse spectrale différentielle [2]. La représentation graphique en trois dimensions résultante, de type spectrogramme, s'obtient à partir du calcul des amplitudes et des fréquences instantanées dans des bandes fréquentielles ajustables. L'objectif est de pouvoir mettre en évidence des aspects de la perception auditive parfois déjà étudiés, qui, à notre connaissance, n'apparaissent pas explicitement sur les représentations temps-fréquence courantes. Les domaines d'application englobent les diverses branches de l'acoustique, la recherche et la création musicales, le traitement des signaux audio et de parole.

Chapitre 1

Théorie

1.1 Rappels

1.1.1 Transformée de Fourier

La Transformée de Fourier (TF) [14] d'une fonction $x(t)$ (réelle ou complexe) est définie (lorsque cette intégrale existe) par :

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-2i\pi ft} dt \quad (1.1)$$

La Transformée de Fourier Discrète (TFD) [15] d'un signal discret $x(k)$ (réel ou complexe) observé sur une durée de N échantillons est périodique de période N et définie par :

$$X(n) = \frac{1}{N} \sum_{m=0}^{N-1} x(m)e^{-2i\pi \frac{mn}{N}} \quad (1.2)$$

1.1.2 Transformée de Fourier à Court Terme

La Transformée de Fourier à Court Terme (TFCT) du signal $x(t)$ est définie par :

$$X_h(t, f) = \int_{-\infty}^{+\infty} x(\tau)h^*(t - \tau)e^{-2i\pi f\tau} d\tau \quad (1.3)$$

$$\text{avec } \begin{cases} t \in \mathbb{R}, \text{ le temps} \\ f \in \mathbb{R}, \text{ la fréquence} \\ h(t) \in \mathbb{R} \text{ ou } \mathbb{C}, \text{ la fenêtre de pondération utilisée} \end{cases}$$

La relation avec la TF $X(f)$ du signal est donnée par :

$$X_h(t, f) = [X(f) \otimes H_\tau^*(-f)](f) \quad (1.4)$$

où $H\tau(f)$ est la TF de $h(t-\tau)$ et \otimes désigne le produit de convolution. La TFCT sous forme numérique est définie par :

$$X_h(k, n) = \sum_{m=0}^{N-1} x(k+m)h^*(m)e^{-2i\pi\frac{nm}{N}} \quad (1.5)$$

1.1.3 Signal analytique

Le signal analytique $x_a(t)$ associé au signal $x(t)$ est défini par :

$$x_a(t) = x(t) + i H\{x(t)\} \quad (1.6)$$

où H désigne la transformation de Hilbert :

$$H\{x(t)\} = \frac{1}{\pi t} \otimes x(t) = \frac{1}{\pi} \text{VP} \int_{-\infty}^{+\infty} x(\tau) \frac{1}{t-\tau} d\tau \quad (1.7)$$

On a alors $x(t) = \text{Re}\{x_a(t)\}$.

La relation (1.6) exprimée dans le domaine fréquentiel devient :

$$X_a(f) = \begin{cases} 2X(f) & \text{si } f \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (1.8)$$

Ainsi, le signal analytique x_a peut-être obtenue par la TF inverse de $2X$ auquel on a préalablement annulé les composantes correspondant aux fréquences négatives. De plus, cette transformation s'effectue sans aucune perte d'information puisque la TF d'un signal réel vérifie la relation de symétrie hermitienne, à savoir :

$$X^*(f) = X(-f) \quad (1.9)$$

A temps discret, le signal analytique $x_a(k)$ associé à $x(k)$ est tel que les TFD respectives sur N points vérifient :

$$X_a(n) = \begin{cases} X(n) & \text{si } n = 0 \text{ ou } n = \frac{N}{2} \\ 2X(n) & \text{si } n = 2, 3, \dots, \frac{N}{2} - 1 \\ 0 & \text{si } n = \frac{N}{2} + 1, \dots, N - 1 \end{cases} \quad (1.10)$$

1.1.4 Amplitude et fréquence instantanées

La définition de la fréquence instantanée n'est pas unique et les premières tentatives de définition remontent au début du XXe siècle [4]. La définition classique la plus utilisée aujourd'hui est celle de Ville développée dans son article [12] et repose sur le concept de signal analytique explicité plus haut.

1.2 - Analyse Spectrale Différentielle (ASD)

Ainsi, à partir du signal analytique $x_a(t)$, il est possible de définir une amplitude instantanée $A(t)$ et une fréquence instantanée $f_i(t)$ par :

$$\begin{cases} A(t) &= |x_a(t)| \\ f_i(t) &= \frac{1}{2\pi} \frac{d}{dt} \arg\{x_a(t)\} \end{cases} \quad (1.11)$$

1.2 Analyse Spectrale Différentielle (ASD)

L'ASD est une forme de la TFCT associée à une technique de calcul. Elle permet de décrire un signal sous la forme d'une somme de composantes, dont les amplitudes et les fréquences évoluent dans le temps. De plus, elle se révèle peu exigeante en puissance de calcul du fait de sa simplicité algorithmique.

1.2.1 ASD à temps continu

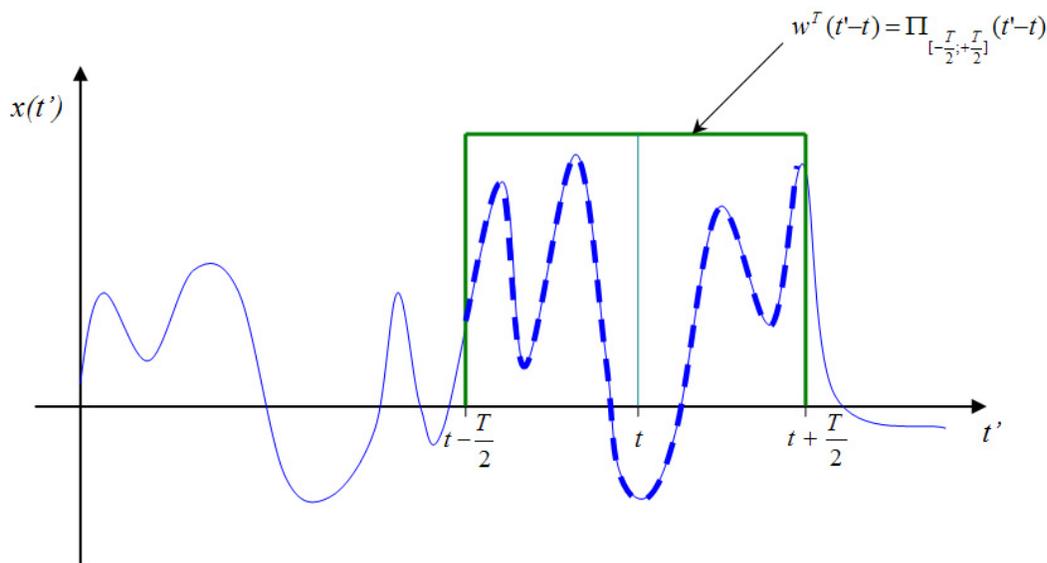


FIG. 1.1 – *Portion de signal vue à travers une fenêtre rectangulaire glissante.*

Considérons un signal réel $x(t)$. Les coefficients de l'ASD de ce signal sur une fenêtre de durée T sont une forme particulière de sa TFCT. On les obtient à partir de la portion du signal $x(t)$ vue à travers la fenêtre rectangulaire glissante $\Pi_{[-\frac{T}{2}, \frac{T}{2}]}(t' - t)$ (cf. figure

1.1). Ils sont définis par :

$$X(n, t) = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} x(u) e^{-2i\pi \frac{n}{T} u} du \quad (1.12)$$

En dérivant cette expression par rapport à t , on obtient :

$$\frac{dX(n, t)}{dt} = \frac{(-1)^n}{T} \left\{ x\left(t + \frac{T}{2}\right) - x\left(t - \frac{T}{2}\right) \right\} e^{-2i\pi \frac{n}{T} t} \quad (1.13)$$

Ce résultat est intéressant car le calcul de $\frac{dX(n, t)}{dt}$ ne dépend que des échantillons $x\left(t - \frac{T}{2}\right)$ et $x\left(t + \frac{T}{2}\right)$ situés aux extrémités de la fenêtre d'analyse . C'est sur cette propriété que repose la mise en oeuvre de l'ASD.

1.2.2 ASD à temps discret

La forme discrète des coefficients de l'ASD est donnée par :

$$X(n, k) = \frac{1}{N} \sum_{l=k-\frac{N-1}{2}}^{k+\frac{N-1}{2}} x(l) e^{-2i\pi \frac{n}{N} l} \quad (1.14)$$

Par différenciation, on retrouve une relation sous une forme très simple, qui constitue la base des algorithmes de calcul :

$$\boxed{\Delta X(n, k) = \frac{(-1)^n}{N} \left\{ x\left(k + \frac{N+1}{2}\right) - x\left(k - \frac{N-1}{2}\right) \right\} e^{-2i\pi \frac{n}{N} k}} \quad (1.15)$$

Des calculs plus détaillés sur l'obtention de cette équation de récurrence figurent dans l'annexe A.1.

1.2.3 Intérêt de l'ASD

On déduit de la relation (1.15) une loi de récurrence permettant de calculer facilement les coefficients :

$$\begin{aligned} X(n, k+1) &= X(n, k) + \Delta X(n, k) \\ &= X(n, k) + \frac{(-1)^n}{N} \left\{ x\left(k + \frac{N+1}{2}\right) - x\left(k - \frac{N-1}{2}\right) \right\} e^{-2i\pi \frac{n}{N} k} \end{aligned} \quad (1.16)$$

Le calcul de $\Delta X(n, k)$ se fait en un nombre constant d'opérations. La complexité du calcul du spectre entier à un instant donné, *ie* de $X(n, k)$ pour k fixé et $n \in [0; N-1]$ est donc en $O(N)$, alors que celle de l'algorithme de transformée de Fourier rapide (TFR) est en $O(N \log(N))$.

1.3 - Analyse Spectrale Différentielle (ASD)

D'autre part, l'estimation de la fréquence instantanée utilise l'égalité suivante (cf calculs exposés dans l'annexe A.1) :

$$f_i(n, k) = \frac{1}{2\pi} \frac{\operatorname{Re}\{X(n, k)\} \operatorname{Im}\{\Delta X(n, k)\} - \operatorname{Im}\{X(n, k)\} \operatorname{Re}\{\Delta X(n, k)\}}{\operatorname{Re}\{X(n, k)\}^2 + \operatorname{Im}\{X(n, k)\}^2} \quad (1.17)$$

L'estimation de la fréquence instantanée se fait donc facilement à partir des calculs de ΔX et de X issus de l'ASD. Par ailleurs, on peut remarquer qu'il n'y a pas de changement de phase à l'origine et que le déroulement de la phase ne pose pas de problème de continuité.

1.2.4 Expression des coefficients de l'ASD appliquée au signal analytique

Considérons un signal réel $x(t)$, et sa transformée de Fourier $X(f)$ écrite sous la forme :

$$X(f) = A(f) e^{i\Phi(f)} \quad (1.18)$$

On démontre (cf. annexe A.2) que les coefficients de l'ASD peuvent se mettre sous la forme :

$$X(n, t) = \int_{-\infty}^{+\infty} X(f') \operatorname{sinc} \left[\pi T \left(\frac{n}{T} - f' \right) \right] e^{-i2\pi \left(\frac{n}{T} - f' \right) t} df' \quad (1.19)$$

Cette dernière équation (1.19), développée en séparant les parties réelle et imaginaire de $X(f)$, fait apparaître un terme additif de fréquence $2f$ [1]. Celui-ci provient de la contamination de la partie positive du spectre de $X(f)$ par la partie négative. Une possibilité de contourner ce problème est de calculer les coefficients $X(n, t)$ à partir du signal analytique correspondant au signal réel considéré. En effet, la partie négative du spectre étant nulle, elle n'interfère plus sur la partie positive. Ceci permet donc l'élimination de ces oscillations et simplifie de manière significative l'expression analytique de $X(n, t)$.

Pour cela, il suffit de remplacer dans l'équation (1.19), $X(f)$ par $X_a(f)$, défini par (1.8), pour obtenir l'expression suivante des coefficients de l'ASD correspondant à un signal analytique :

$$\boxed{X(n, t) = e^{-i2\pi \frac{n}{T} t} \int_0^{+\infty} 2A(f') \operatorname{sinc} \left[\pi T \left(\frac{n}{T} - f' \right) \right] e^{i(2\pi f' t + \Phi(f'))} df'} \quad (1.20)$$

Désormais, on ne considérera plus que l'analyse des signaux analytiques correspondant aux signaux réels étudiés.

1.3 Etude de l'ASD

Afin de caractériser le comportement de l'ASD, nous nous intéressons ici à des cas simples où tous les paramètres sont connus et maîtrisés. Il est alors plus facile de conclure et d'établir des comparaisons.

1.3.1 Signal sinusoïdal

Étudions le comportement de $X(n, t)$ pour $x(t) = A \cos(2\pi f_0 t + \varphi)$. La TF de $x(t)$ est :

$$X(f) = \frac{A}{2} (\delta(f - f_0) e^{i\varphi} + \delta(f + f_0) e^{-i\varphi}) \quad (1.21)$$

D'où, d'après (1.8), le spectre du signal analytique :

$$X_a(f) = A \delta(f - f_0) e^{i\varphi} \quad \forall f \geq 0 \quad (1.22)$$

D'après (1.20), on obtient :

$$\begin{aligned} X(n, t) &= e^{-i2\pi \frac{n}{T} t} \int_0^{+\infty} A \delta(f - f_0) \operatorname{sinc} \left[\pi T \left(f - \frac{n}{T} \right) \right] e^{i(2\pi f t + \varphi)} df \\ &= A \operatorname{sinc} \left[\pi T \left(f_0 - \frac{n}{T} \right) \right] e^{i(2\pi (f_0 - \frac{n}{T}) t + \varphi)} \end{aligned} \quad (1.23)$$

En posant $X(n, t) = X_n(t) e^{i\theta_n(t)}$, les coefficients de l'ASD s'écrivent :

$$\begin{cases} X_n(t) = A \operatorname{sinc} \left[\pi T \left(f_0 - \frac{n}{T} \right) \right] \\ \theta_n(t) = 2\pi \left(f_0 - \frac{n}{T} \right) t + \varphi \end{cases} \quad (1.24)$$

L'amplitude $X_n(t)$ reste constante et correspond à l'amplitude du signal réel multiplié par un sinus cardinal fonction de la différence de fréquences $(\frac{n}{T} - f_0)$. De plus, on remarque que $X_n(t) = A$ pour $f_0 = \frac{n}{T}$ et s'annule pour $f_0 = \frac{p}{T}$ avec $p \neq n$. Le schéma 1.2 illustre le comportement de $X_n(t)$ pour différentes valeurs de la fréquence du signal analysé.

Quant à la phase, elle reste linéaire et il est donc possible de définir sa pente qui donne l'écart entre la pulsation du signal et la pulsation centrale de la bande d'analyse.

Ainsi, on déduit de l'expression (1.24) :

$$\theta'_n(t) = 2\pi \left(f_0 - \frac{n}{T} \right) \quad (1.25)$$

On retrouve alors simplement la fréquence contenue dans le signal $x(t)$:

$$\boxed{f_0 = \frac{1}{2\pi} \theta'_n(t) + \frac{n}{T}} \quad (1.26)$$

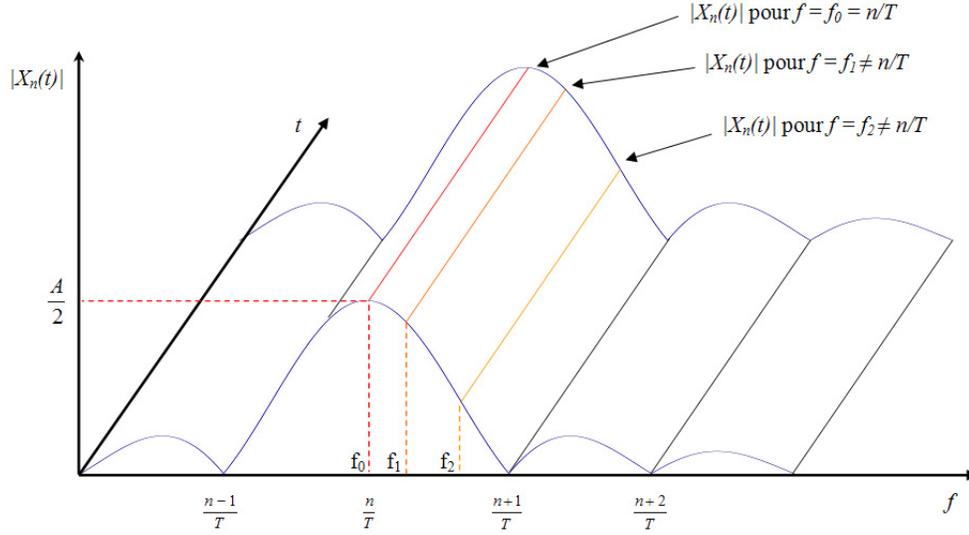


FIG. 1.2 – Comportement de $X_n(t)$ pour différentes valeurs de la fréquence du signal analysée (f_0, f_1, f_2) et proches de $\frac{n}{T}$.

1.3.2 Signal à fréquence variable

Considérons le signal non stationnaire $x(t) = A \cos(\phi(t))$ avec une phase $\phi(t)$ qui varie continûment. Si $\phi(u)$ est de plus développable au 1^{er} ordre en t sur l'intervalle $[t - \frac{T}{2}; t + \frac{T}{2}]$, on obtient l'approximation suivante :

$$\phi(u) \approx \phi(t) + (u - t) \phi'(t) \quad (1.27)$$

Cette hypothèse reste valable si la durée T de la fenêtre est suffisamment faible pour que la phase puisse être considérée comme une fonction linéaire du temps sur tout l'intervalle $[t - \frac{T}{2}; t + \frac{T}{2}]$.

Or,

$$\cos(\phi(t) + (u - t) \phi'(t)) = \cos[\phi'(t) u + \{\phi(t) - t \phi'(t)\}] \quad (1.28)$$

D'où,

$$X_a(f) \approx A \delta\left(f - \frac{1}{2\pi} \phi'(t)\right) e^{i(\phi(t) - t \phi'(t))} \quad \forall f \geq 0 \quad (1.29)$$

On en déduit donc, d'après (1.20) :

$$\begin{aligned} X(n, t) &= e^{i2\pi \frac{n}{T} t} \int_0^{+\infty} A \delta\left(f - \frac{1}{2\pi} \phi'(t)\right) \text{sinc}\left[\pi T \left(\frac{n}{T} - f\right)\right] e^{i(2\pi f t + \phi(t) - t \phi'(t))} df \\ &= e^{i2\pi \frac{n}{T} t} A \text{sinc}\left[\pi T \left(\frac{n}{T} - \frac{1}{2\pi} \phi'(t)\right)\right] e^{i\phi(t)} \\ &= A \text{sinc}\left[\pi T \left(\frac{n}{T} - \frac{1}{2\pi} \phi'(t)\right)\right] e^{i(\phi(t) - 2\pi \frac{n}{T} t)} \end{aligned} \quad (1.30)$$

D'où pour une composante non-stationnaire $x(t) = A(t) \cos(\phi(t))$, dans le cas où $A(t)$ n'est pas constant mais présente des variations lentes devant celles de $\cos(\phi(t))$, les coefficients de l'ASD s'écrivent :

$$X(n, t) = \overline{A(t)} \operatorname{sinc} \left[\pi T \left(\frac{1}{2\pi} \phi'(t) - \frac{n}{T} \right) \right] e^{i(\phi(t) - 2\pi \frac{n}{T} t)} \quad (1.31)$$

$\overline{A(t)}$ représentant la moyenne de $A(t)$ sur T .

Ou sous une autre forme, en posant $X(n, t) = X_n(t) e^{i\theta_n(t)}$:

$$\begin{cases} X_n(t) &= \overline{A(t)} \operatorname{sinc} \left[\pi T \left(\frac{1}{2\pi} \phi'(t) - \frac{n}{T} \right) \right] \\ \theta_n(t) &= \phi(t) - 2\pi \frac{n}{T} t \end{cases} \quad (1.32)$$

Une estimation de la fréquence instantanée s'en déduit donc facilement :

$$\boxed{f_n(t) \approx \frac{1}{2\pi} \theta'_n(t) + \frac{n}{T}} \quad (1.33)$$

On retrouve ainsi, en première approximation, une expression de la fréquence instantanée similaire au cas du signal sinusoïdal (1.26).

1.3.3 Signal à deux composantes sinusoïdales : phénomènes inter-bandes

Considérons maintenant le signal : $x(t) = A_0 \cos(2\pi f_0 t + \varphi_0) + A_1 \cos(2\pi f_1 t + \varphi_1)$.

On démontre (cf. annexe A.3) que le module et la fréquence instantanée des coefficients de l'ASD correspondant à $x(t)$ ont pour expressions :

$$\begin{cases} |X(n, t)| &= \sqrt{B_0^2 + B_1^2 + 2B_0 B_1 \cos(2\pi(f_0 - f_1)t + (\varphi_0 - \varphi_1))} \\ f_n(t) &= \frac{B_0^2 f_0 + B_1^2 f_1 + B_0 B_1 (f_0 + f_1) \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)}{B_0^2 + B_1^2 + B_0 B_1 \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)} \end{cases} \quad (1.34)$$

$$\text{avec } \begin{cases} B_0 = A_0 \operatorname{sinc} \left[\pi T \left(f_0 - \frac{n}{T} \right) \right] \\ B_1 = A_1 \operatorname{sinc} \left[\pi T \left(f_1 - \frac{n}{T} \right) \right] \end{cases}$$

Nous pouvons étudier l'influence d'une composante fréquentielle sur l'autre en supposant $|B_0| \gg |B_1|$. En utilisant le développement limité de $\frac{1}{1+x}$ à l'ordre 2 en 0 pour $x = \left(\frac{B_1}{B_0}\right)^2 + \frac{B_1}{B_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)$, ce qui reste justifié dans le cas où $|B_0| \gg |B_1|$, on simplifie (cf. annexe A.3) les expressions précédentes sous la forme :

$$\begin{aligned}
 |X(n, t)| &= \sqrt{B_0^2 \left(1 + \left(\frac{B_1}{B_0} \right)^2 + 2 \frac{B_1}{B_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right)} \\
 f_n(t) &\approx f_0 \left(1 + \left(\frac{B_1}{B_0} \right)^2 \left(\frac{f_1}{2f_0} - 1 \right) + \frac{B_1 f_1}{B_0 f_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right. \\
 &\quad \left. - \left(\frac{B_1}{B_0} \right)^2 \frac{f_1}{2f_0} \cos(4\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right) \quad (1.35)
 \end{aligned}$$

Ces expressions montrent une déviation de la partie constante de l'amplitude et de la fréquence instantanées ainsi que des oscillations aux fréquences multiples de $(f_1 - f_2)$. L'ampleur de ces phénomènes parasites dépend du rapport d'amplitudes $\frac{B_1}{B_0}$. Il est donc possible de les réduire en minimisant la valeur de B_1 . En pratique, B_1 correspond à l'amplitude A_1 de la composante gênante, modifiée par le terme en sinus cardinal dû à l'influence de la fenêtre qui peut-être vue comme un filtre passe-bas. Ce filtre n'est pas très sélectif (le premier lobe secondaire est à -13dB environ) et rend souvent l'influence de B_1 non négligeable. Une solution consiste à filtrer les coefficients obtenus par l'ASD de manière plus efficace avant d'estimer l'amplitude et la fréquence instantanées. Les termes d'amplitudes B_0 et B_1 dans les expressions (1.35) sont alors atténués par ce filtre. En pratique, on choisira un filtre passe-bas dont la fréquence de coupure est égale à la largeur de la bande d'analyse afin d'éliminer les composantes provenant d'autres bandes. L'ordre du filtre est ajustable par l'utilisateur.

1.3.4 Signal à plusieurs composantes : phénomènes intra-bandes

Que se passe-t-il lorsqu'une bande contient plus d'une composante d'amplitude non-négligeable ?

L'analyse par la TFCT donne pour ce canal une amplitude estimée $B(t)$ et une phase estimée $\Phi(t)$. Contrairement à une analyse traditionnelle où l'on suppose que chaque bande est assez étroite pour ne contenir au plus qu'une composante, l'analyse en bande large regroupe plusieurs composantes au sein de l'unique signal complexe $B(t)e^{i\Phi(t)}$. Comme le montre la figure 1.3, les estimations $B(t)$ de l'amplitude et $f_i(t) = \frac{1}{2\pi} \frac{d\Phi(t)}{dt}$ de la fréquence instantanée ont des allures plus complexes que dans le cas de composantes séparées.

Les signaux d'amplitude et de fréquence instantanées obtenus sont périodiques. Si les périodes respectives étaient relativement élevées (1s par exemple), la modulation d'amplitude serait perçue comme des battements et la modulation de phase comme un changement de hauteur. Ce cas rendrait la lecture des courbes intuitive. Dans le cas de l'exemple de la figure 1.3, la période étant de l'ordre de 10ms, la perception ne correspond pas à une lecture intuitive et simpliste des courbes.

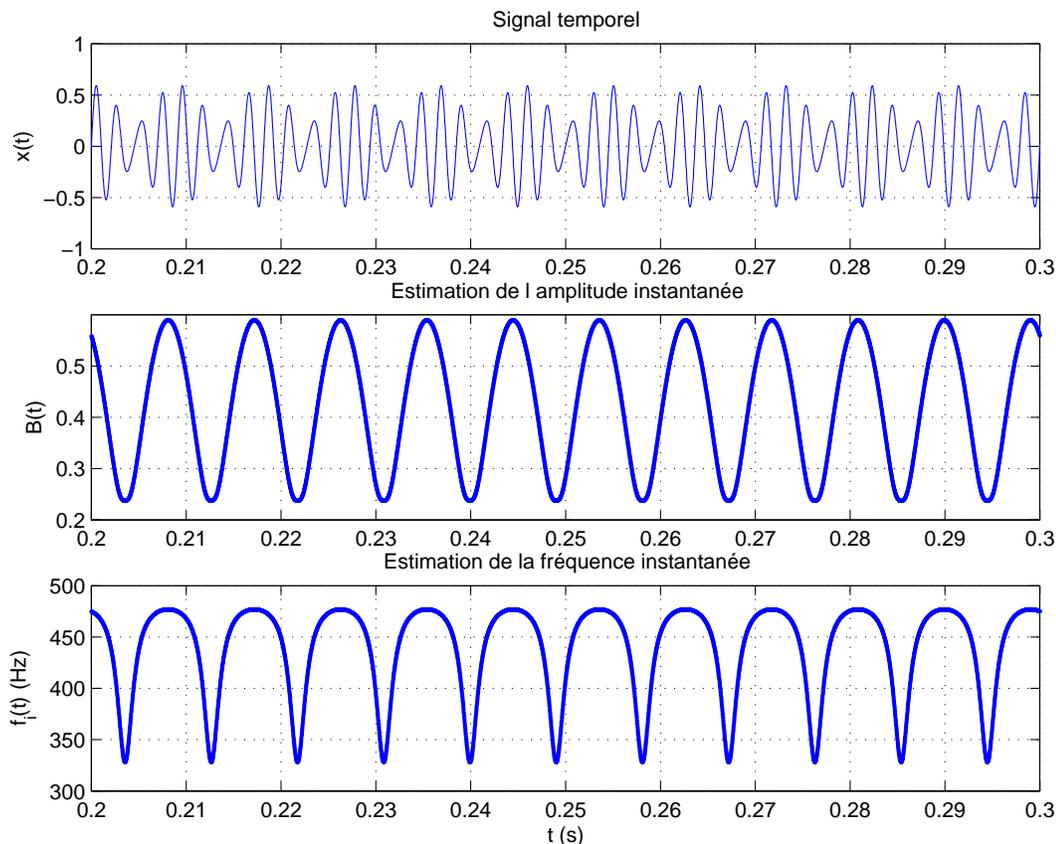


FIG. 1.3 – Signal à deux composantes à 440Hz et 550Hz et analyse monobande.

Nous essaierons de répondre à deux questions essentielles :

- Les estimations de l'amplitude et de la fréquence instantanées correspondent-elles toujours à des signaux ayant un sens physique, directement utilisables, ou bien les variations observées sont-elles inexploitable ? Dans le premier cas, y a-t-il une erreur d'estimation et de quel ordre est-elle ?
- Peut-on interpréter les représentations graphiques de cette amplitude et de cette fréquence instantanées ? Comment "lire" une telle représentation ?

Décomposition canonique en modulation d'amplitude et de phase

Nous allons utiliser une analogie afin de bien situer les éléments en jeu et notamment la notion de fréquence instantanée. La synthèse FM [18] consiste à moduler une fréquence porteuse f_c par une ou plusieurs fréquences modulantes f_m . Dans le cas simple d'une seule fréquence modulante, le signal synthétisé est du type $x(t) = A(t) \sin(2\pi f_c t + I \sin(2\pi f_m t))$ où $A(t)$ est l'amplitude et I un paramètre appelé index de modulation. Si

l'on pose $\Phi(t) = 2\pi f_c t + I \sin(2\pi f_m t)$, on a $x(t) = A(t) \sin(\phi(t))$ et l'on définit naturellement la fréquence instantanée du signal par $f_i(t) = \frac{1}{2\pi} \frac{d\Phi}{dt} = f_c + I f_m \cos(2\pi f_m t)$.

On peut généraliser le concept en considérant tout signal $x(t)$ comme une modulation canonique en amplitude et en phase. On a $x(t) = \text{Re}\{x_a(t)\} = \text{Re}\{a(t)e^{i\Phi(t)}\} = a(t) \cos(\Phi(t))$ où $x_a(t)$ désigne le signal analytique associé à $x(t)$ et où l'amplitude $a(t)$ et la phase $\Phi(t)$ sont définies de manière unique par :

$$\begin{cases} a(t) = |x_a(t)| \\ \Phi(t) = \arctan\left(\frac{\text{Im}\{x_a(t)\}}{\text{Re}\{x_a(t)\}}\right) \end{cases} \quad (1.36)$$

En appliquant cette décomposition à un signal à deux composantes, on obtient :

$$\begin{aligned} x(t) &= a_0 \cos(2\pi f_0 t) + a_1 \cos(2\pi f_1 t) & (1.37) \\ x_a(t) &= a_0 e^{2i\pi f_0 t} + a_1 e^{2i\pi f_1 t} \\ a(t) &= |x_a(t)| \\ &= \sqrt{a_0^2 + a_1^2 + 2a_0 a_1 \cos(2\pi(f_0 - f_1)t)} \\ \Phi(t) &= \arctan\left(\frac{a_0 \sin(2\pi f_0 t) + a_1 \sin(2\pi f_1 t)}{a_0 \cos(2\pi f_0 t) + a_1 \cos(2\pi f_1 t)}\right) \\ f_i(t) &= \frac{1}{2\pi} \frac{d\Phi}{dt} \\ &= \frac{a_0^2 f_0 + a_1^2 f_1 + a_0 a_1 (f_0 + f_1) \cos(2\pi(f_0 - f_1)t)}{a_0^2 + a_1^2 + 2a_0 a_1 \cos(2\pi(f_0 - f_1)t)} \end{aligned}$$

Dans le cas particulier où les deux composantes ont la même amplitude ($a_0 = a_1$), les résultats se simplifient :

$$\begin{aligned} x(t) &= a_0 (\cos(2\pi f_0 t) + \cos(2\pi f_1 t)) \\ x_a(t) &= a_0 (e^{2i\pi f_0 t} + e^{2i\pi f_1 t}) \\ a(t) &= 2 \left| a_0 \cos\left(2\pi \frac{f_0 - f_1}{2} t\right) \right| \\ f_i(t) &= \frac{f_0 + f_1}{2} \end{aligned}$$

Ces résultats nous permettent de préciser les concepts de fréquence et d'amplitude instantanées. Ces deux données varient ici périodiquement : les associations {amplitude instantanée - intensité sonore} et {fréquence instantanée - hauteur} qui permettent habituellement de faire correspondre les paramètres physiques et les paramètres perceptifs ne sont plus valables.

Estimations de l'amplitude et de la fréquence instantanées

L'analyse par la TFCT du signal à deux composantes donne d'après (1.26) et (1.34) :

$$\begin{aligned}
 x(t) &= a_0 \cos(2\pi f_0 t + \varphi_0) + a_1 \cos(2\pi f_1 t + \varphi_1) \\
 x_a(t) &= a_0 e^{2i\pi f_0 t + \varphi_0} + a_1 e^{2i\pi f_1 t + \varphi_1} \\
 X_n(t) &= \sqrt{B_0^2 + B_1^2 + 2B_0 B_1 \cos(2\pi(f_0 - f_1)t + (\varphi_0 - \varphi_1))} \\
 \theta_n(t) &= \arctan \left[\frac{B_0 \sin(2\pi f_0 t + \varphi_0) + B_1 \sin(2\pi f_1 t + \varphi_1)}{B_0 \cos(2\pi f_0 t + \varphi_0) + B_1 \cos(2\pi f_1 t + \varphi_1)} \right] - 2\pi \frac{n}{T} t \\
 f_n(t) &= \frac{B_0^2 f_0 + B_1^2 f_1 + B_0 B_1 (f_0 + f_1) \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)}{B_0^2 + B_1^2 + B_0 B_1 \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)} \\
 \text{avec } B_0 &= a_0 \operatorname{sinc} \left[\pi T \left(f_0 - \frac{n}{T} \right) \right] \\
 \text{et } B_1 &= a_1 \operatorname{sinc} \left[\pi T \left(f_1 - \frac{n}{T} \right) \right]
 \end{aligned}$$

Les résultats obtenus sont similaires à ceux provenant de la décomposition canonique en modulation d'amplitude et de fréquence. Les expressions sont identiques à l'exception des amplitudes des composantes qui sont multipliées par un terme en sinc $[\pi T (f - \frac{n}{T})]$. Ce terme est proche de 1 et peut être négligé si l'on considère une bande d'analyse ayant une largeur inférieure à celle du sinus cardinal.

Interprétation des résultats

Les estimations d'amplitude et de fréquence instantanées ont bien une signification physique dans le cas d'une analyse large bande. Elles constituent une approximation des mêmes paramètres obtenus en décomposant canoniquement le signal en une modulation d'amplitude et de fréquence. On peut donc étendre naturellement les concepts de fréquence et d'amplitude instantanées aux signaux de plusieurs composantes et l'analyse en TFCT permet de représenter ces paramètres moyennant l'approximation évoquée. En revanche, la compréhension de ces concepts est beaucoup moins aisée car les paramètres physiques et perceptifs ne sont plus directement reliés. Il convient alors de rapprocher l'analyse et la perception en bande large, lorsque cela est justifié, et d'étudier, comme dans le cas de la synthèse FM, la perception des différentes modulations d'amplitude et de fréquence dans des cas particuliers : (in)harmonicités, battements, perception de hauteur(s), couleurs sonores, etc. Ce travail de grande ampleur est illustré par la suite à travers quelques analyses de sons (cf. paragraphes 3.1.2 p.24 et 3.1.3 p.25).

1.3.5 Estimation de l'amplitude instantanée

Dans beaucoup d'applications, il est utile de connaître la valeur de l'amplitude instantanée $A(t)$ évoluant au cours du temps.

Or d'après l'étude faite précédemment, on s'aperçoit que l'amplitude des coefficients de l'ASD $X_n(t)$, dans le cas où il n'y a qu'une seule composante par bande, reste en première approximation proche de celle du signal réel. En effet, on a d'après (1.32), dans le cas d'un signal à fréquence variable :

$$X_n(t) \approx \overline{A(t)} \operatorname{sinc} \left[\pi T \left(\frac{1}{2\pi} \phi'(t) - \frac{n}{T} \right) \right]$$

$\overline{A(t)}$ représentant la moyenne de $A(t)$ sur T .

On en déduit donc très facilement une estimation de l'amplitude instantanée $A_{inst}(t)$:

$$A_{inst}(t) = \overline{A(t)} = \frac{X_n(t)}{\operatorname{sinc} \left[\pi T \left(\frac{1}{2\pi} \phi'(t) - \frac{n}{T} \right) \right]} \quad (1.38)$$

avec $\frac{1}{2\pi} \phi'(t) = \frac{p-n\pi}{\pi T}$ où $p \in \mathbb{Z}^*$

Cette expression reste valable dans le cas où le signal ne possède qu'une seule fréquence dans la bande d'analyse, et dans le cas où son amplitude présente des variations lentes devant la phase. Le cas du signal sinusoïdal est bien sûr un cas particulier.

1.3.6 Fenêtre de pondération

La fenêtre de pondération implicitement utilisée jusqu'ici est une fenêtre rectangulaire. Sa transformée de Fourier présente des lobes secondaires importants, défaut qui entraîne une forte influence des composantes entre elles. Le choix de mise en oeuvre via l'analyse spectrale différentielle ne permet cependant pas d'éviter l'utilisation d'une fenêtre rectangulaire : le "glissement" de la transformée de Fourier, illustré par la formule de récurrence suivante, implique une pondération uniforme des échantillons.

$$X(n, k+1) = X(n, k) + x \left(k + \frac{N+1}{2} \right) e^{-2i\pi \frac{n}{N} \left(k + \frac{N+1}{2} \right)}$$

$$- x \left(k - \frac{N-1}{2} \right) e^{-2i\pi \frac{n}{N} \left(k - \frac{N-1}{2} \right)}$$

Ce glissement s'effectue en soustrayant à la transformée de Fourier courante le point le plus ancien ($x \left(k - \frac{N-1}{2} \right)$) et en ajoutant le point suivant ($x \left(k + \frac{N+1}{2} \right)$). Si l'on voulait réaliser le même glissement avec une fenêtre de pondération non uniforme, il faudrait réévaluer la pondération de chaque point dans la somme donnant la transformée de Fourier ($m \in \left[k - \frac{N-1}{2}; k + \frac{N+1}{2} \right]$). Une telle modification n'est pas facilement envisageable et n'a pas été étudiée dans le cadre de ce projet. Elle n'est pour autant pas exclue et a déjà été abordée dans une forme récursive proche de l'ASD [16].

Dans notre cas, il est possible d'éviter les effets des lobes secondaires en appliquant un pré-traitement du type banc de filtres pour isoler chaque bande d'analyse ou un post-traitement consistant en un filtrage passe-bas (cf. paragraphe 1.3.3 p.10).

1.4 Échelle fréquentielle ajustable

La méthode décrite jusqu'ici utilise une répartition homogène des bandes d'analyse : la bande d'indice n a une fréquence centrale $f_n = \frac{n}{T}$ et une largeur $\Delta f_n = \frac{2}{T}$. Cette largeur est indépendante de n . Dans certains contextes comme l'étude de la perception, il est préférable de faire une analyse en utilisant des bandes de largeur variable. Cette partie montre comment abstraire le concept d'échelle, réduit jusqu'ici à l'emploi d'une échelle linéaire.

Une bande fréquentielle de l'échelle d'analyse était jusqu'ici caractérisée par deux paramètres : le numéro de la bande n et la longueur T de la fenêtre d'analyse. On en déduisait alors sa fréquence centrale $f_n = \frac{n}{T}$ et sa largeur $\Delta f_n = \frac{2}{T}$. Ce sont ces deux dernières données que l'on souhaite manipuler de façon à pouvoir choisir l'une indépendamment de l'autre.

La solution consiste à choisir la fréquence centrale f et la largeur Δf de chaque bande, puis à en déduire la valeurs de n et de T correspondantes. On obtient immédiatement :

$$\begin{aligned} n &= \frac{2f}{\Delta f} \\ T &= \frac{2}{\Delta f} \end{aligned} \tag{1.39}$$

Si l'on numérote par $m \in \mathbb{N}$ les bandes de l'échelle, on remarque alors deux différences par rapport à la situation précédente :

- la valeur de n est *a priori* différente du numéro m de la bande.
- la valeur de T varie en fonction m .

On notera donc n_m et T_m les deux paramètres caractéristiques de la bande m et l'on caractérisera les bandes de l'échelle d'analyse par le triplet (m, n_m, T_m) . Le cas particulier de l'échelle linéaire se retrouve en posant $\forall m, n_m = m$ et $T_m = T$. On peut de même obtenir facilement la distribution logarithmique des bandes fréquentielles en posant $\forall m, n_m = n$: seul le paramètre T_m varie. La construction de cette échelle est détaillée dans l'annexe C.

Chapitre 2

Spectrogramme d'Amplitude et de Fréquence Instantanées (SAFI)

2.1 Principe

Les estimations issues de l'ASD sont représentées sur un spectrogramme analogue au spectrogramme classique dans la disposition des axes temps-fréquence-amplitude. La différence essentielle réside dans la représentation des fréquences. En effet, le spectrogramme classique représente une estimation de la puissance instantanée du signal, censée décrire son contenu fréquentiel sur un nombre fini de bandes. Dans le cas du SAFI, on utilise la même notion d'amplitude instantanée et on introduit une donnée supplémentaire, la fréquence instantanée, qui permet de suivre l'évolution de la fréquence du signal à l'intérieur de la bande considérée.

2.2 Construction du SAFI

Le programme se divise en trois parties : un bloc d'initialisation, une boucle de calculs et un bloc de génération du graphique. Le bloc d'initialisation, représenté dans la partie supérieure de la figure 2.1, permet de lire les échantillons et les différents paramètres, de calculer le signal analytique et les conditions initiales des filtres.

Le second bloc, représenté dans la partie inférieure de la figure 2.1, analyse le signal après l'avoir segmenté, compte tenu de la taille des données à manipuler. Pour chaque segment de signal, on effectue une ASD suivie éventuellement d'un filtrage passe-bas, puis l'on estime les amplitudes et les fréquences instantanées. Un seuillage en fréquence permet de s'assurer que les bandes ne se recouvrent pas. On effectue ensuite un seuillage en amplitude paramétré par l'utilisateur afin de négliger l'information non pertinente.

Chapitre 2 - Spectrogramme d'Amplitude et de Fréquence Instantanées (SAFI)

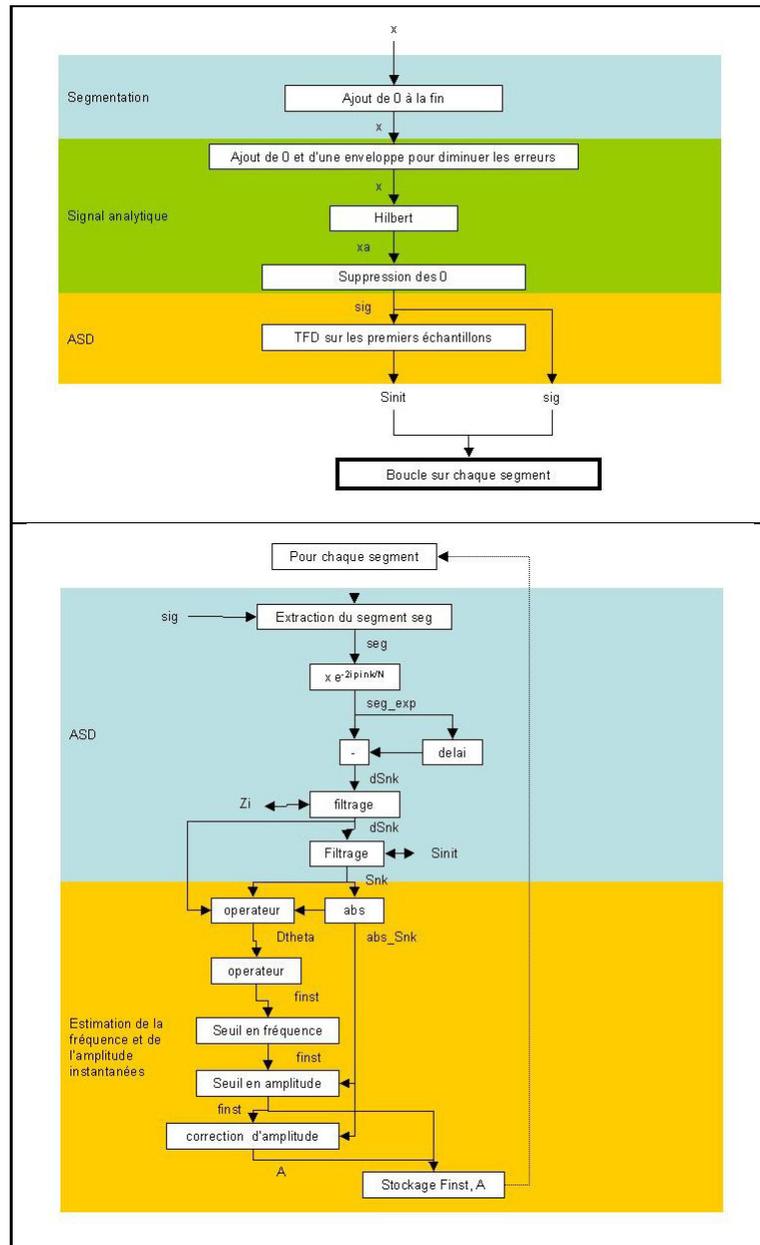


FIG. 2.1 – Initialisation du programme (en haut) et bloc d'estimation de l'amplitude et de la fréquence instantanées (en bas)

2.3 - Paramètres

Enfin une correction d'amplitude permet de minimiser l'erreur due au sinus cardinal. Les données sont ensuite sous-échantillonnées et stockées.

Le dernier bloc trace l'ensemble des courbes correspondant à chaque bande.

2.3 Paramètres

L'outil, conçu en Matlab, se présente sous la forme d'une interface de contrôle mettant à la disposition de l'utilisateur un certain nombre de paramètres (cf. figure 2.2). Il peut ensuite lancer le calcul et l'affichage des résultats.

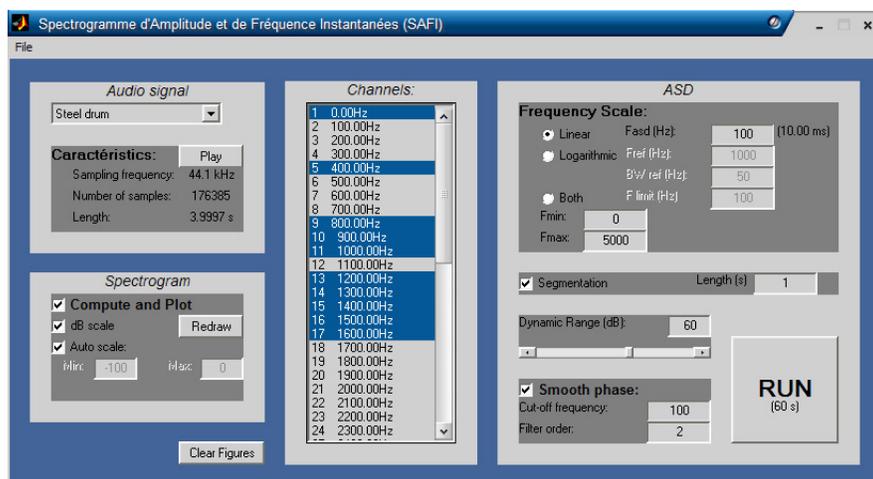


FIG. 2.2 – Aperçu de l'interface graphique

La partie «Audio signal» permet de sélectionner un fichier son à analyser, de l'écouter et de voir ses principales caractéristiques. A l'intérieur de la partie «ASD», le bloc «Frequency Scale» donne le choix entre une échelle fréquentielle linéaire, logarithmique ou mixte (une partie linéaire en basses fréquences et logarithmique au-delà d'une limite ajustable). Dans chaque cas, la taille des bandes est ajustable par une fréquence de référence et les bornes F_{min} et F_{max} limitent le nombre de bandes disponibles. L'utilisateur choisit ensuite de segmenter le signal s'il est trop long. Le seuil en amplitude est contrôlé par le paramètre «Dynamic Range». Enfin les paramètres du filtre passe-bas de post-traitement de l'ASD sont ajustables dans le bloc «Smooth Filter». Les bandes d'analyse déduites des paramètres de l'échelle fréquentielle sont affichées et sélectionnables dans la partie «Channels». La partie «Spectrogram» donne accès aux paramètres de représentation : bornes de l'échelle d'amplitude qui peut-être linéaire ou logarithmique, bouton

de génération du graphique. Une fois les paramètres ajustés, l'utilisateur peut lancer l'analyse en appuyant sur «RUN».

2.4 Complexité

La complexité de l'ASD est en $O(N)$ (cf. paragraphe 1.2.3 p.6). Chacun des mécanismes ajoutés dans le développement de l'outil a également un coût linéaire ou constant. L'ordre de la complexité reste donc en $O(N)$. Cette rapidité de calcul est très appréciable et rare parmi les représentations temps-fréquence.

Sur un micro-ordinateur de type Pentium III 500MHz avec 512Mo de mémoire vive, l'analyse complète par le SAFI d'une seconde de signal échantillonné à 44 100Hz prend moins d'une seconde par bande. Plusieurs façons sont envisageables pour améliorer les performances de l'outil : mise en oeuvre dans un langage plus rapide que Matlab, sous-échantillonnage en entrée de l'ASD (moyennant certaines précautions), optimisation algorithmique.

Chapitre 3

Analyse de sons

3.1 Analyses de sons synthétiques

3.1.1 Chirp de durée variable

Le chirp est un signal comportant une seule composante de fréquence instantanée variable. Nous utilisons dans cette partie un son construit à partir d'un chirp élémentaire. La fréquence instantanée de ce chirp parcourt linéairement l'intervalle $[20\text{Hz}; 1\text{kHz}]$ et est modulé en amplitude par une fenêtre de Hanning (cf. Figure 3.1).

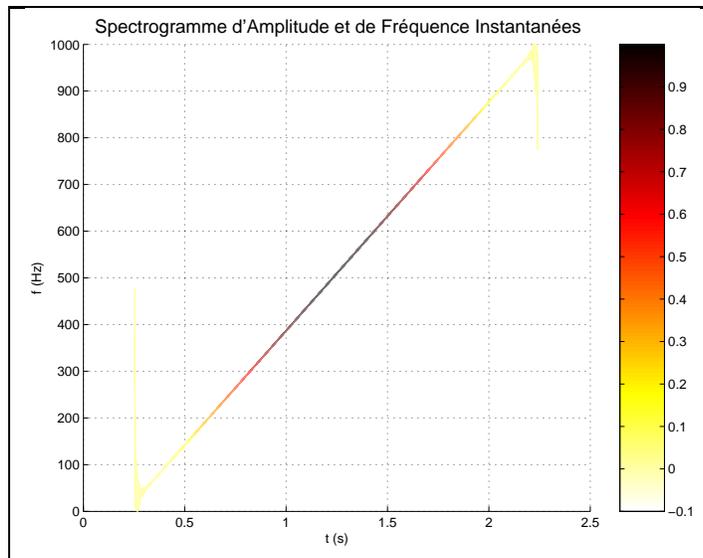


FIG. 3.1 – *Chirp élémentaire (2s)*

Le son analysé est composé d'une alternance de chirps de ce type et de silences. La durée des chirps est initialement égale à 0,5s et diminue progressivement jusqu'à 10ms.

On change ainsi le niveau de quasi-stationnarité du signal afin d'évaluer la robustesse de l'analyse.

La figure 3.2 montre le résultat obtenu en analysant le son à travers un spectrogramme classique.

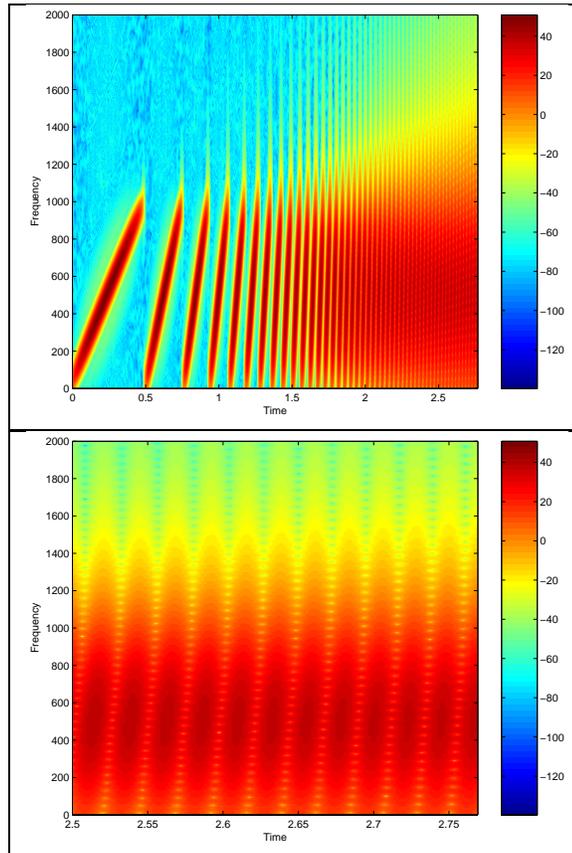


FIG. 3.2 – Spectrogramme du chirp de durée variable (FFT sur 8192 points avec une fenêtre de Blackman-Harris de 2048 points et un recouvrement de 1968 points), avec les détails de la partie critique (figure du bas)

On constate que l'étalement spectral introduit une incertitude importante sur la fréquence. La longueur de la fenêtre d'analyse est restreinte et accentue le phénomène. Cette longueur est pourtant optimale si l'on veut pouvoir voir les variations fréquentielles lorsque la durée du chirp diminue. Le compromis temps/fréquence est ici problématique.

Un SAFI de ce son est représentée sur la figure 3.3. L'analyse est faite sur une bande centrée en 500Hz, de largeur 1kHz. On constate que le caractère linéaire de la variation fréquentielle s'altère lorsque la durée du chirp diminue. On garde cependant une très bonne précision. L'analyse sur une fenêtre large fait reculer les limites du compromis

3.1 - Analyses de sons synthétiques

temps/fréquence évoqué précédemment.

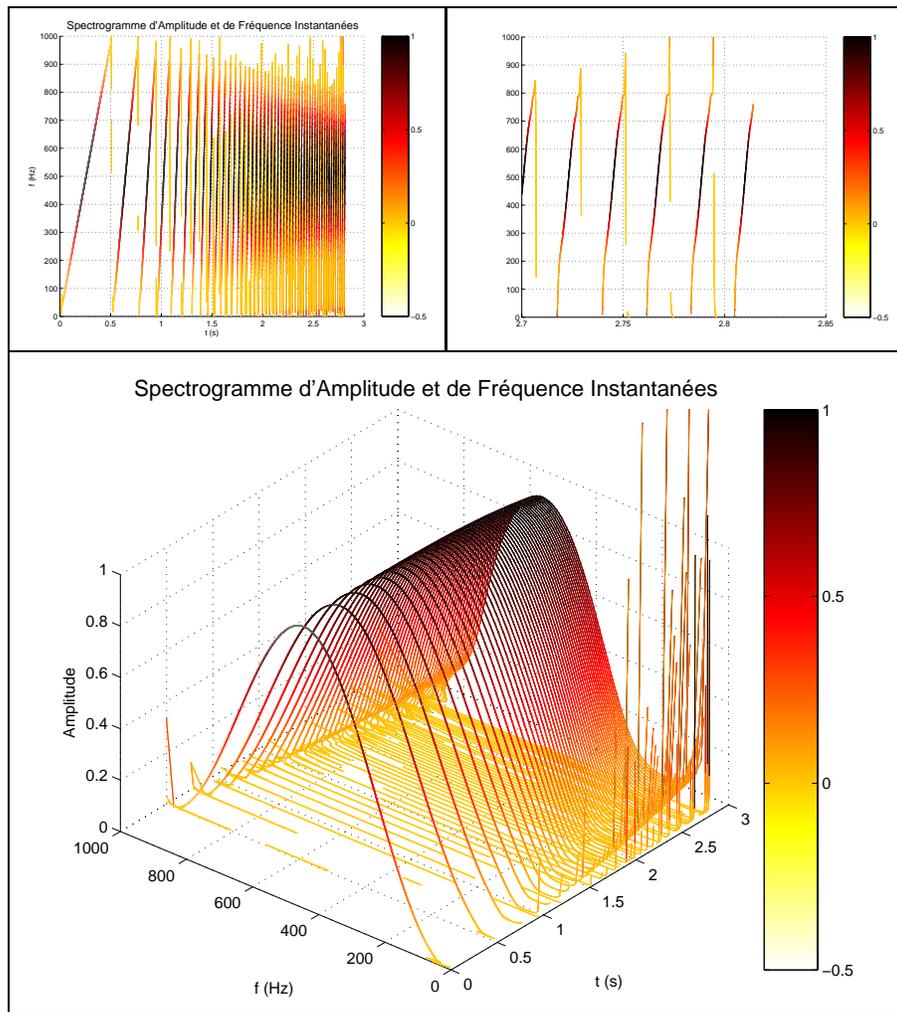


FIG. 3.3 – SAFI du chirp accéléré : représentation en 2D (à gauche), détail des chirps les plus rapides (à droite), et représentation en 3D (en bas)

3.1.2 Analyse en bande large : modulations

Nous analysons ici un son composé de deux sinusoïdes de fréquences voisines. Le signal temporel est du type :

$$x(t) = \sin(2\pi f_0 t) + a \sin(2\pi f_1 t) \quad (3.1)$$

Nous nous intéressons au cas où les amplitudes des sinusoïdes sont différentes, ie $a \neq 1$. Nous prendrons comme exemple $f_0 = 440\text{Hz}$, $f_1 = 442.5\text{Hz}$ et $a = 0.7$. Ce type de son est au premier abord perçu comme une sinusoïde modulée en amplitude. Une écoute plus fine permet de détecter une variation de hauteur.

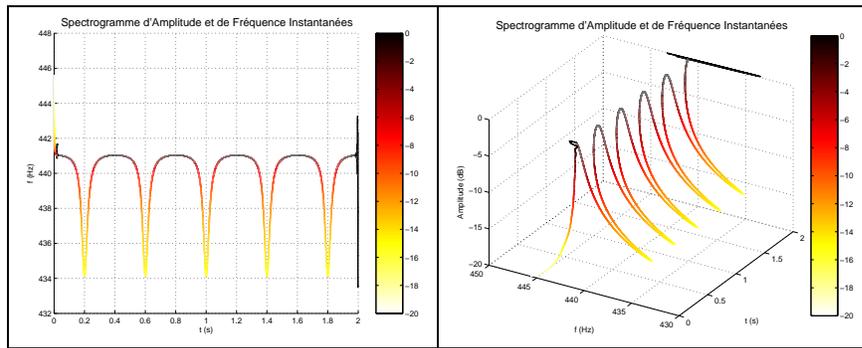


FIG. 3.4 – SAFI d'une somme de deux sinusoïdes

L'analyse sur une bande centrée en 500Hz de largeur 500Hz (cf. Figure 3.4) montre des oscillations à $f_1 - f_0 = 2.5\text{Hz}$ entre 0 et -15dB pour l'amplitude et entre 434 et 441Hz pour la fréquence instantanée (environ $\frac{1}{8}$ ton). Les oscillations des deux paramètres sont en phase. Lorsque a se rapproche de 1, l'amplitude des battements augmente en amplitude et en fréquence. A la limite ($a = 1$), la fréquence instantanée est constante, égale à $\frac{f_1 + f_0}{2}$ et l'amplitude instantanée reste oscillante.

Les modulations d'amplitude et de fréquence sont ici suffisamment lentes (2.5Hz) pour être perçues comme des battements et des changements de hauteur. Le SAFI permet de les visualiser clairement, sans être gêné par les effets dus à l'étalement spectral.

3.1.3 Analyse en bande large : perception de la rugosité

Nous approfondissons ici l'étude de l'analyse en bande large en nous intéressant à des sons étudiés dans [10]. Nous reprenons la première expérience de cet article, qui établit un lien entre la rugosité perçue dans un son et la différence entre les phases initiales des composantes.

Les sons utilisés sont définis par :

$$x_{\varphi_0}(t) = \frac{1}{2}\sin(2\pi(f_c - f_m)t) + \sin(2\pi f_c t + \varphi_0) + \frac{1}{2}\sin(2\pi(f_c + f_m)t) \quad (3.2)$$

Ils sont facilement synthétisables et se composent donc d'une fréquence centrale et de deux fréquences adjacentes. Les phases initiales des fréquences adjacentes sont nulles. L'étude porte sur la variation de la sensation de rugosité en fonction des valeurs de la phase initiale φ_0 de la fréquence centrale.

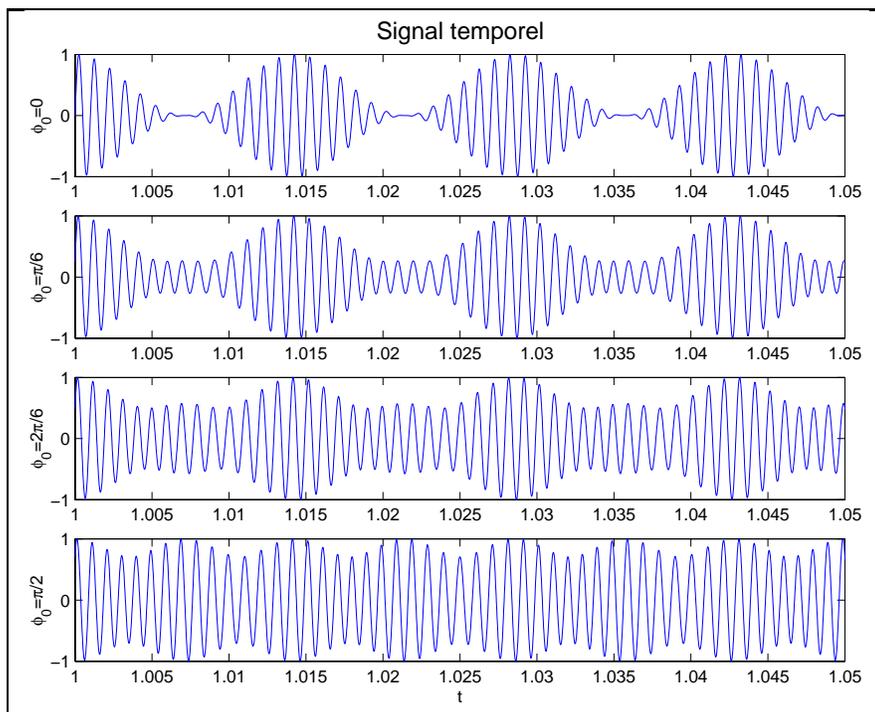


FIG. 3.5 – Etude de la rugosité : signaux temporels

Les résultats de l'article font apparaître un lien entre le caractère rugueux du son et la valeur de la phase initiale : les sujets de l'expérience perçoivent le son comme de plus en plus rugueux lorsque φ_0 passe de $\frac{\pi}{2}$ à 0. Les auteurs expliquent ce phénomène en remarquant qu'un son est jugé d'autant plus rugueux que son enveloppe temporelle

fluctue de manière importante (cf. figure 3.5). On peut ainsi observer des passages à zéros de l'enveloppe dans le cas du son le plus rugueux ($\varphi_0 = 0$) alors que pour le son le moins rugueux ($\varphi_0 = \frac{\pi}{2}$), l'enveloppe reste comprise entre 0.7 et 1.

L'analyse a été faite sur une seule bande de largeur $1000Hz$ centrée en $1000Hz$. Les SAFI obtenus pour $\varphi_0 \in \{0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{\pi}{2}\}$ avec $f_c = 1000Hz$ et $f_m = 70Hz$ sont représentés respectivement sur les figures 3.6, 3.7, 3.8 et 3.9. On peut reprendre l'hypothèse émise sur les enveloppes temporelles et l'étendre à l'amplitude et à la fréquence instantanées. En effet, les variations de ces signaux pour un son très rugueux sont caractéristiques : l'amplitude des oscillations est forte (50 dB pour l'amplitude instantanée) et la raideur des pentes élevée. A l'inverse, pour les sons moins rugueux, les oscillations sont beaucoup plus régulières et moins importantes. On peut également remarquer que la fréquence instantanée pour $\varphi_0 = 0$ diverge et que la représentation graphique donne des pics erronés aux points de divergence. Cette divergence peut-être mise en évidence graphiquement en donnant à φ_0 des valeurs proches de 0.

Cette analyse en bande large permet de représenter la cohérence relative des composantes entre elles. La phase initiale est une constante qui semble n'être présente ni dans le terme d'amplitude ni dans celui de fréquence instantanée du fait de la dérivation opérée. Cette intuition se révèle être fautive dans le cas présent et il est remarquable de pouvoir représenter simplement un tel phénomène. Le lien entre modulation d'amplitude, modulation de fréquence et enveloppe temporelle du signal est très étroit et repose ici sur leurs variations temporelles, qui semblent être responsables de la sensation de rugosité.

3.1 - Analyses de sons synthétiques

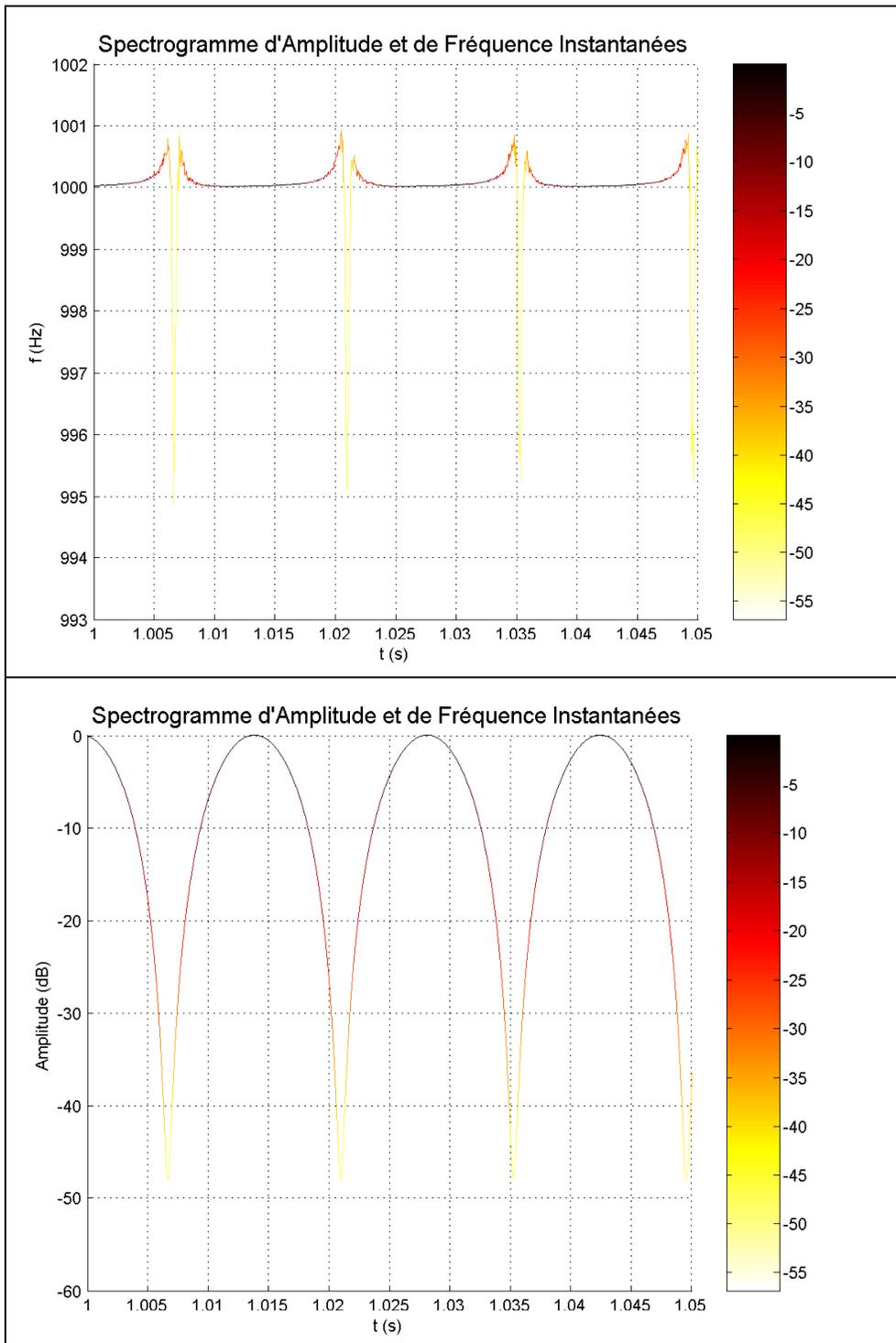


FIG. 3.6 – Etude de la rugosité : amplitude et fréquence instantanées pour $\varphi_0 = 0$

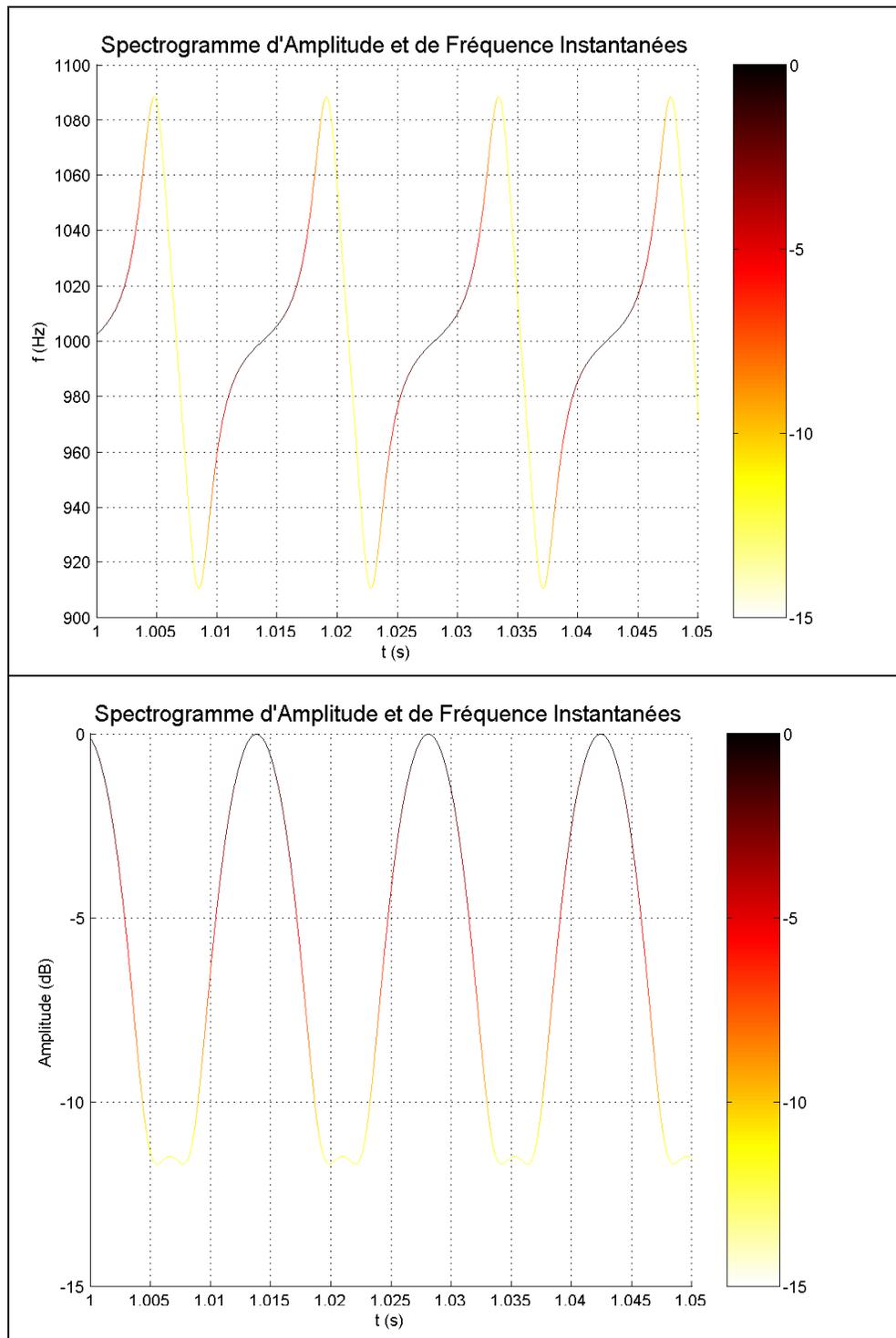


FIG. 3.7 – Etude de la rugosité : amplitude et fréquence instantanées pour $\varphi_0 = \frac{\pi}{6}$

3.1 - Analyses de sons synthétiques

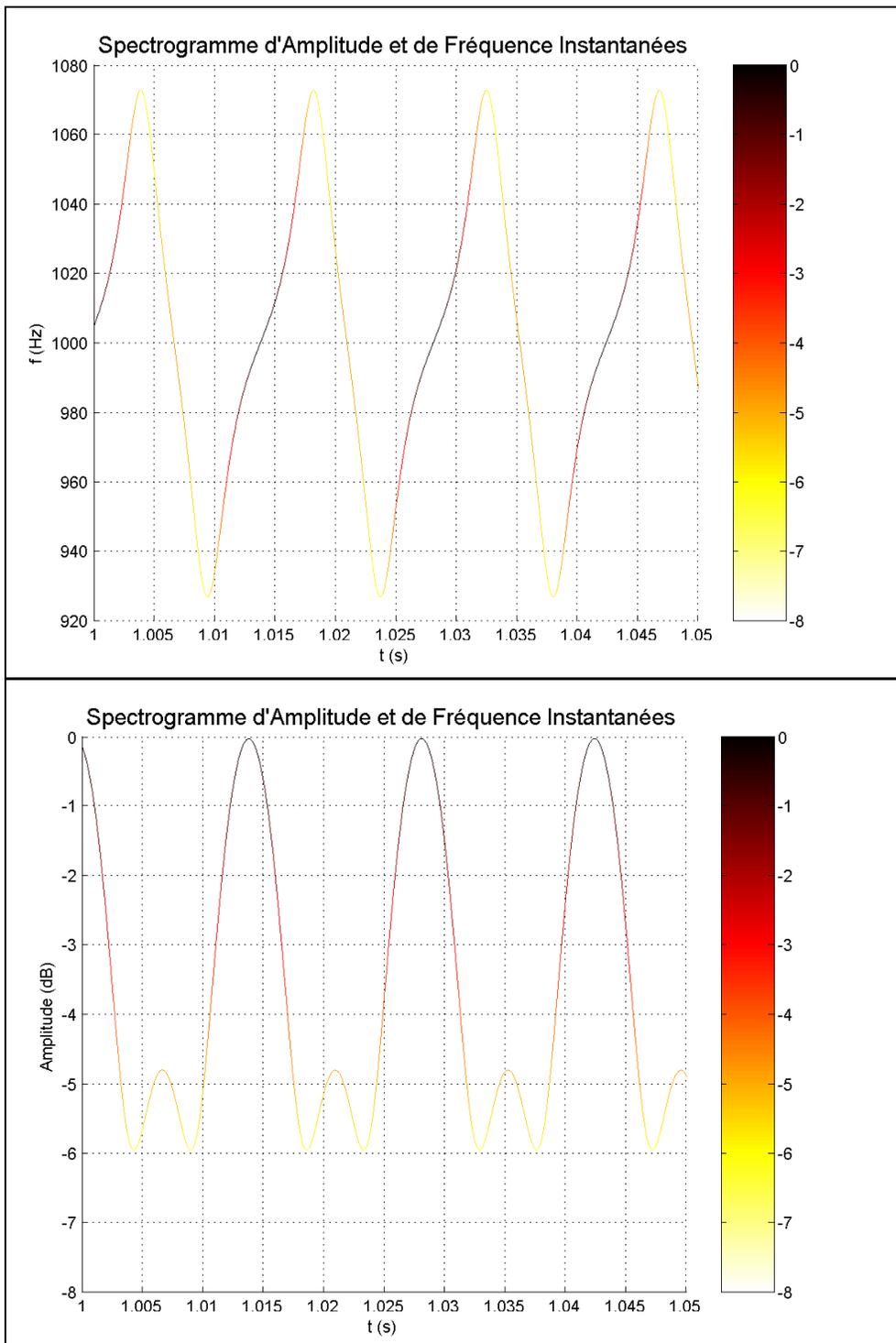


FIG. 3.8 – Etude de la rugosité : amplitude et fréquence instantanées pour $\varphi_0 = \frac{2\pi}{6}$

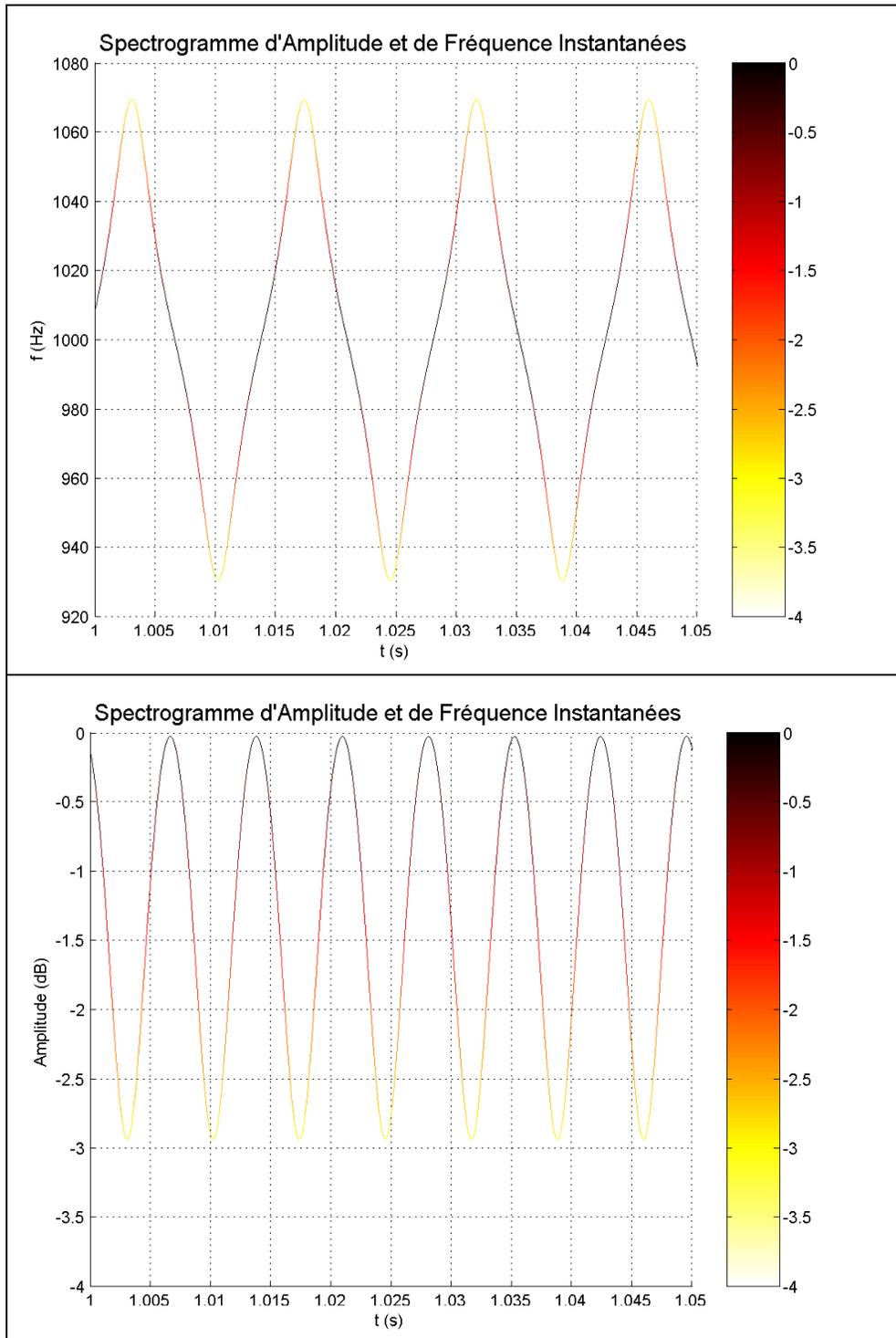


FIG. 3.9 – Etude de la rugosité : amplitude et fréquence instantanées pour $\varphi_0 = \frac{\pi}{2}$

3.2 Analyses de sons réels

3.2.1 Sons de steel drum

Nous analysons ici un son issu d'une collection d'enregistrements de notes de steel drum que nous a aimablement fourni Pascal Gaillard du Laboratoire Jacques Lordat (Université de Toulouse Le Mirail). Le steel drum est un instrument à percussion caribéen obtenu à partir de bidons en métal. Le facteur sculpte la partie supérieure du bidon pour isoler des zones correspondant aux différentes notes. Chaque zone est accordée à l'aide d'un marteau afin d'obtenir une note précise lorsqu'on la frappe. Il s'agit donc d'oscillations non-entretenues sur un matériau dont les partiels ont été accordés pour avoir une sensation de hauteur lors du jeu.

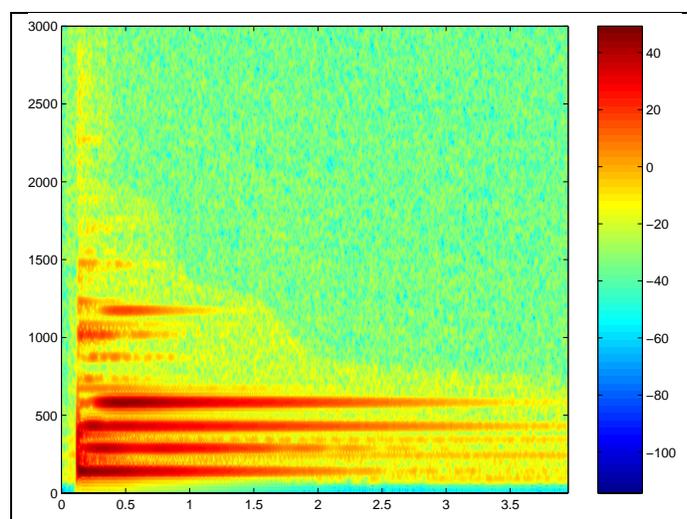


FIG. 3.10 – *Spectrogramme classique du son de steel drum*

La figure 3.10 montre l'analyse obtenue à partir d'un spectrogramme classique. Quatre composantes principales se détachent par leur amplitude et leur durée. On distingue ensuite une grande quantité de composantes secondaires.

L'analyse de ce son par un SAFI est représentée sur la figure 3.11. Elle a été réalisée sur 150 bandes de largeur $15Hz$ avec un post-filtrage d'ordre 3 et de fréquence de coupure $15Hz$.

On retrouve l'ensemble des informations visibles auparavant. Certains détails apparaissent ou gagnent en précision, comme le nombre de partiels visibles et le détail de leur localisation. Le temps de montée des partiels est très bien défini : on remarque en particulier que le quatrième partiel principal arrive avec un retard important, tout

comme un autre partiel autour de $1200Hz$. Ces montées retardées, déjà visibles sur le spectrogramme classique, sont maintenant analysables en détail. Il en est de même pour la décroissance de chaque partiel. Elle est relativement linéaire (sur une échelle en dB) mais apparaissent parfois des oscillations très localisées.

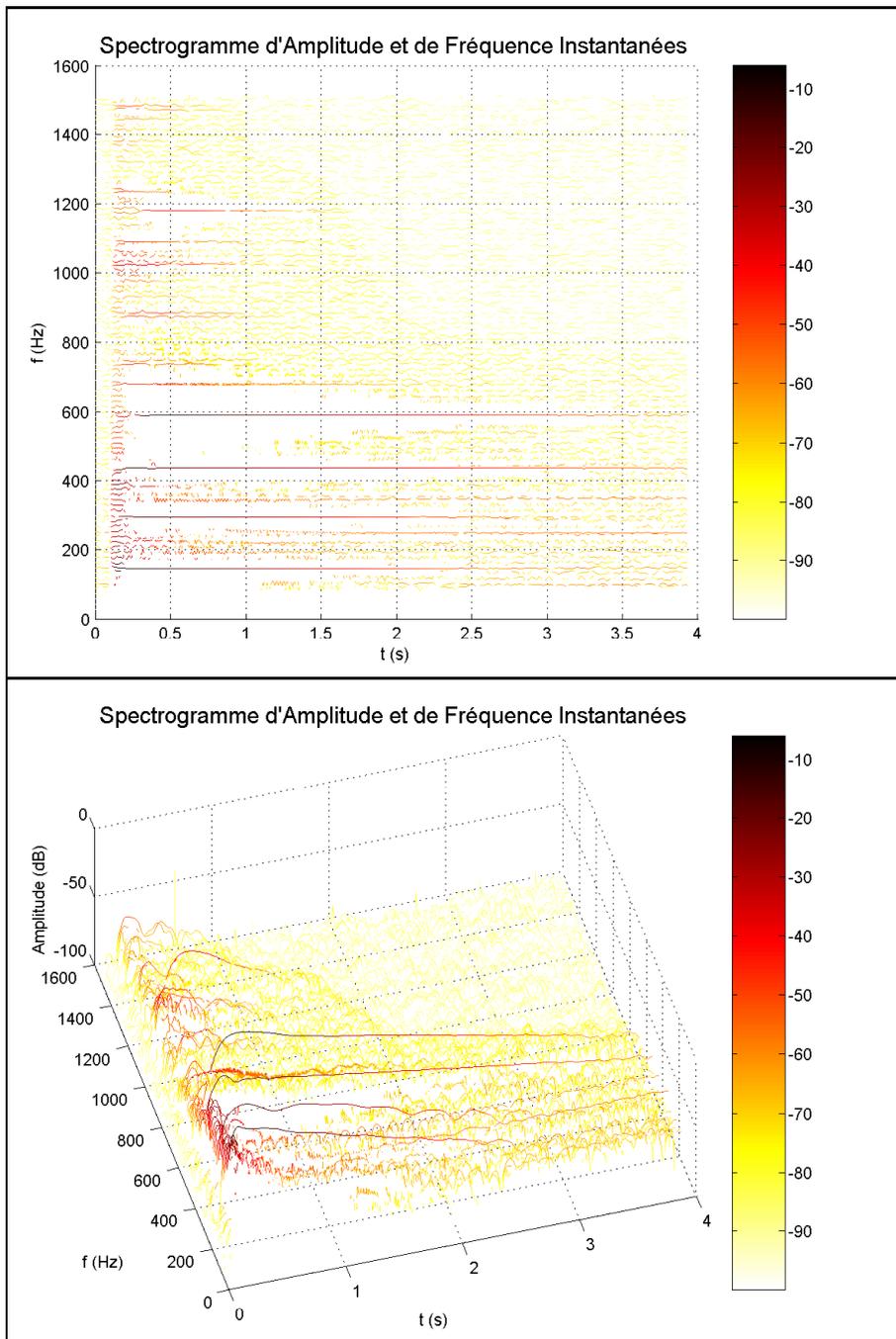


FIG. 3.11 – SAFI du son de steel drum

3.2.2 Sons de trompette marine

Le son est obtenu en jouant l'harmonique 4 sur une trompette marine. La séquence se décompose en deux parties : le son entretenu lorsque l'instrumentiste joue avec l'archet puis les oscillations libres quand l'archet n'est plus en contact avec la corde.

Dans la partie entretenue, on distingue la fréquence fondamentale du son et les harmoniques. On observe également beaucoup de bruit dans certaines tranches fréquentielles, dû probablement aux chocs du chevalet contre la table.

Lors des oscillations libres, le bruit se dissipe très rapidement car le système chevalet/table est très amorti. On voit apparaître des composantes entre les harmoniques précédemment évoqués : il s'agit des partiels de la corde qui vibre librement. Ces partiels sont répartis de manière quasi-harmonique, d'où l'impression d'en voir trois régulièrement espacés entre deux harmoniques du son joué. La corde étant beaucoup moins amortie que le chevalet, ces composantes se maintiennent à des niveaux relativement élevés par rapport au bruit du chevalet très vite dissipé.

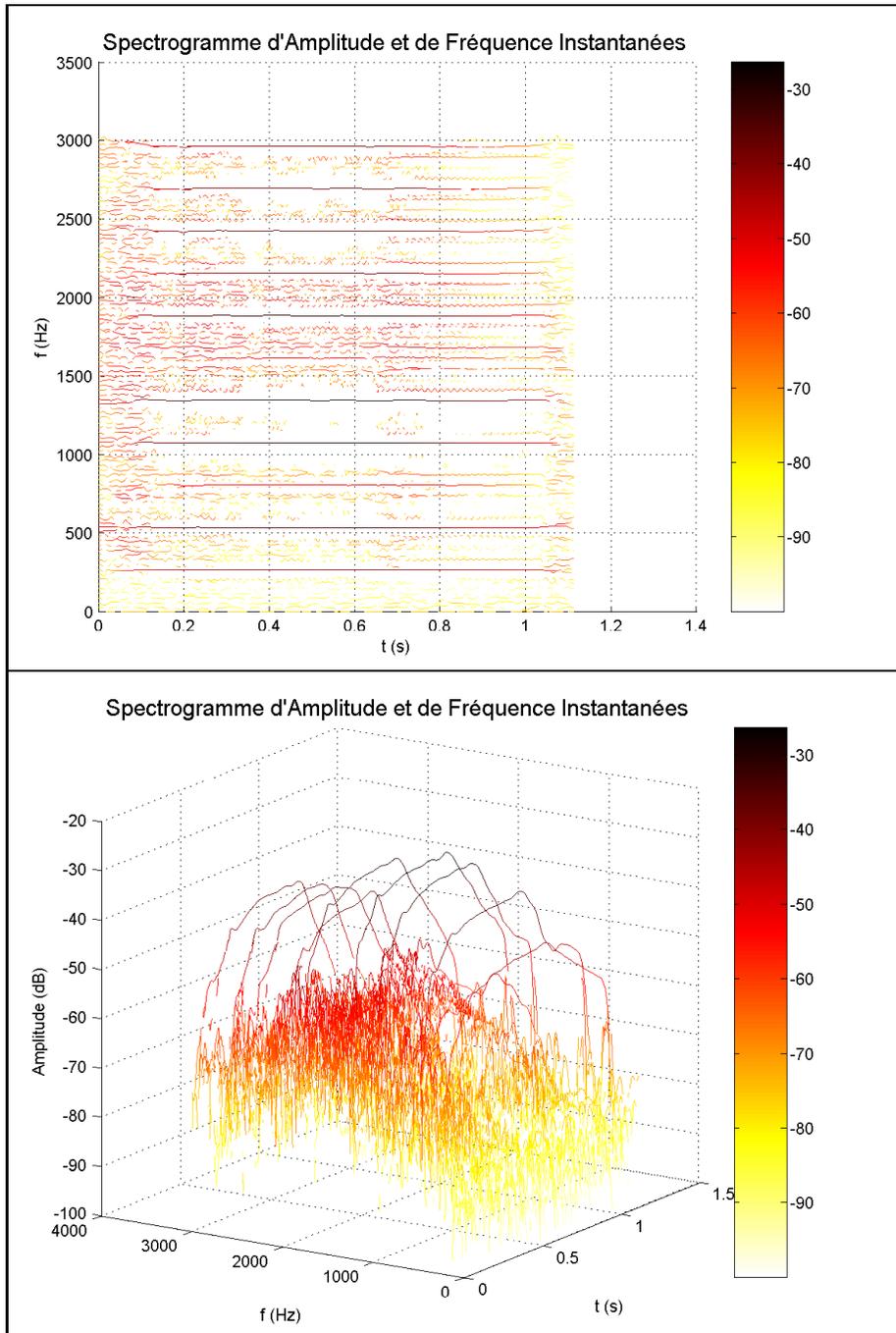


FIG. 3.12 – SAFI trompette marine

3.2.3 Sons de RER

Ce son est issu d'enregistrements effectués par Abril Padilla dans le RER parisien, dans le cadre de ses travaux de composition électroacoustique. L'enregistrement est extrêmement bruité à l'écoute. L'extrait sonore a été choisi pour les sifflements, crissements et autres détails qu'il contient.

L'analyse est effectuée sur des bandes de $400Hz$ (cf. figure 3.14). On observe trois composantes fréquentielles étroites : la première se situe un peu au-dessus de $7000Hz$ au début du son. Les deux suivantes (une fondamentale et la deuxième harmonique) occupent le dernier tiers. On remarque également des «chocs» caractérisés par de l'énergie répartie sur toutes les bandes, sur une durée courte, et irrégulièrement espacés. Enfin, une bande de bruit assez large occupe la plage $[5000Hz; 7500Hz]$. Il est intéressant de noter, outre la finesse de la représentation, la distinction entre les composantes fréquentielles étroites et les bruits de bande assez large. Une composante étroite se détache à la fois par sa finesse et par une zone blanche périphérique que l'on pourrait qualifier de zone de masquage par le sinus cardinal. Dans ce cas de figure, les bandes adjacentes détectent la composante énergétique dominante qui masque le signal présent à l'intérieur de la bande. En revanche, les zones bruitées (chocs ou bande centrale dans notre exemple) conservent leur largeur sur la représentation.

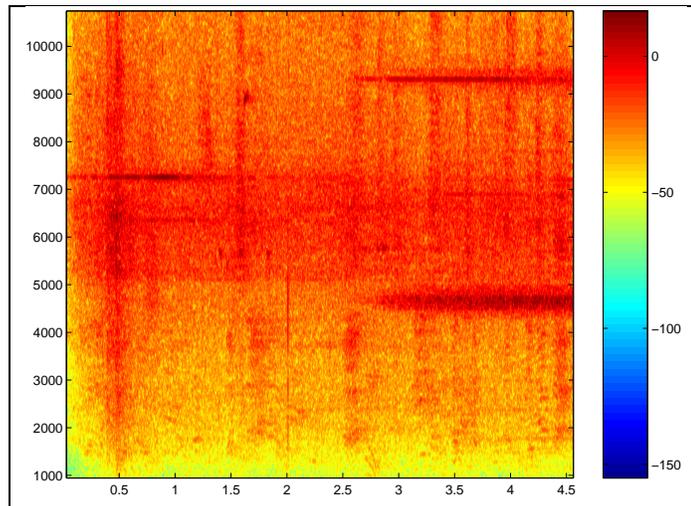


FIG. 3.13 – *Spectrogramme classique*

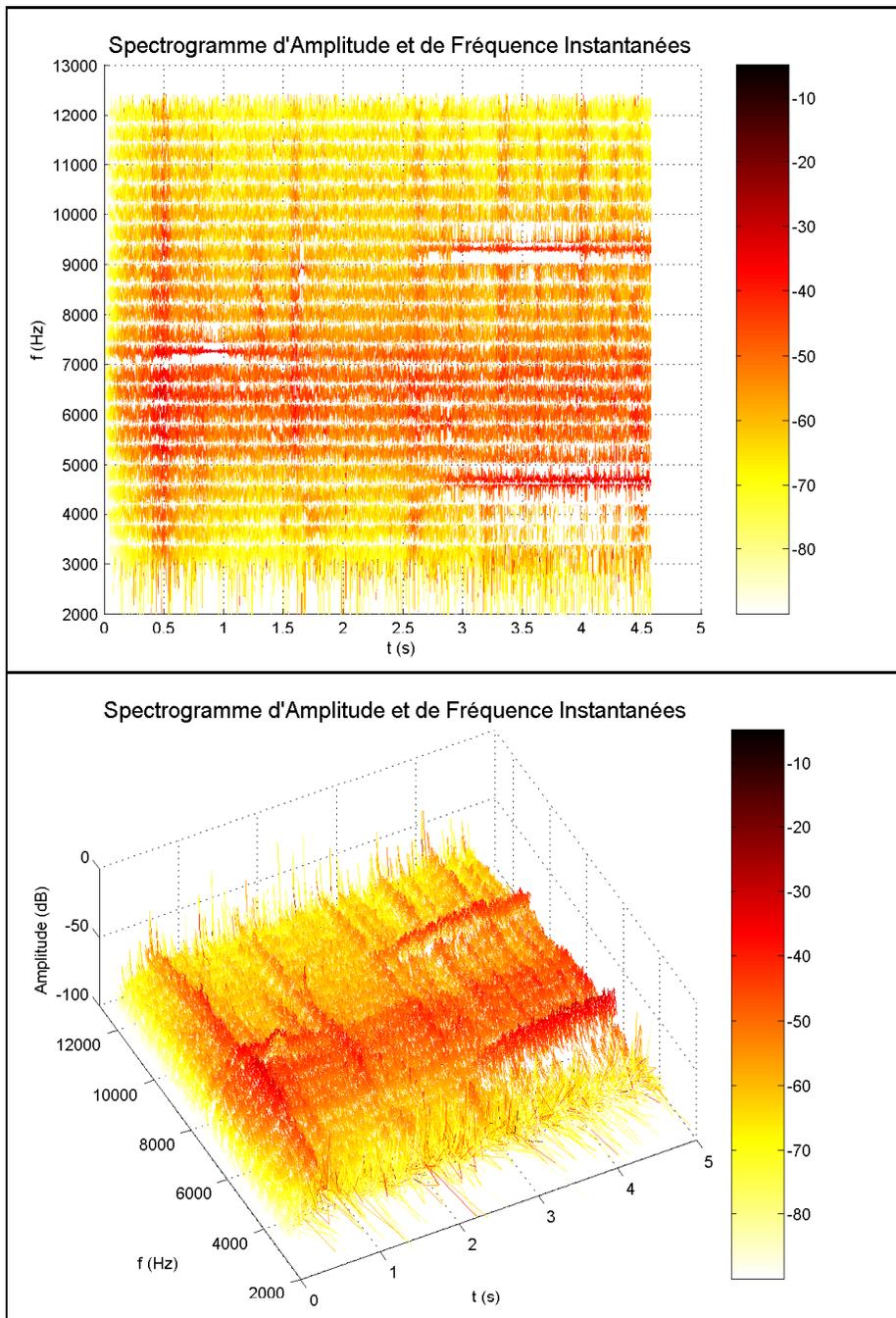


FIG. 3.14 – SAFI RER

Conclusion

Intérêts de la méthode

Le SAFI est avant tout un outil simple dans sa réalisation, bénéficiant ainsi d'un domaine d'application très étendu. Il permet de retrouver et bien souvent d'améliorer les résultats obtenus via un spectrogramme classique. Cette amélioration s'observe tout d'abord dans l'approche élémentaire du spectrogramme : la largeur des lobes due à l'étalement spectral n'est plus visible sur la représentation, remplacée par un tracé fin permettant une précision fréquentielle d'autant plus importante qu'elle résulte de l'estimation de la fréquence instantanée et non de l'approximation des fréquences de Fourier. L'efficacité de la séparation des composantes provient d'une part de cette précision fréquentielle, d'autre part de la possibilité de filtrer les coefficients de l'ASD. Par ailleurs, si l'amplitude est calculée sur le même principe que pour le spectrogramme classique, elle semble cependant gagner en détail grâce à son association avec la fréquence instantanée sous forme de courbe ainsi qu'à la possibilité de perspectives ajustables en trois dimensions. Cet aspect s'observe en particulier sur les transitoires d'attaque des instruments et les différentes formes de décroissance ou d'entretien des composantes (décroissances oscillantes ou linéaires, tremolo,...). Enfin, le SAFI met en évidence la présence de bruit ainsi que la largeur spectrale associée, et accentue la différenciation entre composantes fréquentielles pures et zones bruitées.

Outre ces améliorations qualitatives par rapport au spectrogramme classique, l'étude des capacités du SAFI permet d'envisager de nouveaux principes d'interprétation de la représentation. Ainsi, l'analyse en bandes larges fait apparaître les modulations d'amplitude et de fréquence présentes. On peut ensuite chercher à donner du sens à ces modulations. Il ne s'agit plus de séparer et d'extraire des composantes mais de considérer le signal analytique résultant de la présence de plusieurs d'entre elles. Un des intérêts de cette démarche réside dans le lien possible avec la perception des sons, comme l'illustre l'exemple de la perception de la rugosité. L'outil peut alors accompagner les études qui considèrent que le système auditif perçoit certaines modulations en tant que telles ou qu'il extrait des paramètres issus soit de ces modulations, soit plus généralement de données pertinentes présentes sur une représentation temps-fréquence.

D'un point de vue technique, l'ASD consiste en un calcul récursif sur chaque bande qui se traduit dans la mise en oeuvre par une grande efficacité en termes de temps de calcul. Elle offre également de la souplesse dans le choix des bandes fréquentielles ainsi que dans leur largeur qui peut théoriquement être ajustée individuellement. Enfin, l'aspect modulaire permet d'envisager une amélioration de chaque bloc ou l'utilisation d'autres blocs plus spécifiques à l'application choisie.

Critiques et limites

Les limitations constatées sur le SAFI sont généralement liées à celles de la transformée de Fourier à court terme. Le sinus cardinal conséquent à l'utilisation d'une fenêtre rectangulaire induit des interférences entre les bandes adjacentes et un phénomène de masquage qui peut être gênant. De plus, taille de la fenêtre d'analyse et largeur de bande restent directement liées. Par ailleurs, il n'a pas été envisagé pour l'instant de sous-échantillonner le signal de chaque bande avant l'analyse comme c'est le cas dans beaucoup d'usages de la TFCT afin d'optimiser le coût calculatoire. On ne pourrait cependant pas atteindre la limite théorique du taux de sous-échantillonnage car la représentation qui en découlerait manquerait de résolution temporelle. Elle nécessiterait alors un sur-échantillonnage pour que l'on puisse voir confortablement les variations temporelles des données.

Perspectives

Au cours de cette étude, les bases du SAFI ont été posées tant en ce qui concerne sa construction que sa justification théorique et son utilité. Le cadre ainsi établi, le SAFI a été appliqué à des sons d'origines diverses, dont l'analyse donne des résultats intéressants et originaux. La conception du SAFI se situe donc dans un état d'avancement intermédiaire, dans la mesure où l'outil en lui-même peut encore être largement développé et où ses possibilités d'application n'en sont qu'à leurs débuts.

Les principaux points de développement technique à explorer sont les possibilités de sous-échantillonnage avant analyse, l'aspect «banc de filtres» qui peut paraître simpliste dans sa version actuelle, la correction d'amplitude qui ne tient pas compte du post-filtrage, l'utilisation du signal analytique conjointement au post-filtrage, les possibilités de zéro-padding dans le but d'avoir un découpage fréquentiel plus fin en gardant la même longueur de fenêtre d'analyse. Il serait également intéressant de garder les mêmes objectifs d'analyse et de représentation tout en changeant le bloc «ASD», pour utiliser par exemple des filtres auditifs et obtenir ainsi une représentation plus proche des signaux analysés par l'oreille.

Parmi les applications du SAFI, les exemples étudiés ici seraient à reprendre et à

approfondir. L'analyse des sons de steel drums étant concluante, on pourrait de la même manière valider l'étude des transitoires pour d'autres instruments de musique. Mais il reste surtout à explorer le potentiel applicatif du SAFI, de l'étude des bruits impulsionsnels à la recherche à l'interprétation des modulations en passant par la validation des résultats de travaux de psychoacoustique passés ou à venir.

Plus généralement, l'étude du SAFI et de ses applications s'inscrit dans une vision moderne des problématiques de l'acoustique et vise à proposer des outils de représentation plus proches des phénomènes à étudier. Remettre en question l'héritage d'outils de référence tels le spectrogramme s'avère utile pour améliorer les représentations. Si suite à leur conception datant de plusieurs décennies, ces outils se sont révélés d'une grande efficacité, ils n'en ont pas moins des lacunes qui les rendent aujourd'hui dépassés. Dans le cas du SAFI, la prise en compte de la phase du signal montre une possibilité pour étendre le concept de spectrogramme et faire apparaître des éléments déterminants et absents jusqu'alors.

Annexes

Annexe A

Développements mathématiques

A.1 Passage aux coefficients de l'ASD en discret

Considérons un signal réel $x(t)$. Les coefficients de l'ASD de ce signal sur une fenêtre de durée T sont définis par :

$$X(n, t) = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} x(u) e^{-2i\pi \frac{n}{T} u} du \quad (\text{A.1})$$

D'où, l'expression précédente exprimée sous forme numérique :

$$X(n, k) = \frac{1}{N} \sum_{l=k-\frac{N-1}{2}}^{k+\frac{N-1}{2}} x(l) e^{-2i\pi \frac{n}{N} l} \quad (\text{A.2})$$

D'où,

$$\begin{aligned} X(n, k+1) &= \frac{1}{N} \sum_{l=(k+1)-\frac{N-1}{2}}^{(k+1)+\frac{N-1}{2}} x(l) e^{-2i\pi \frac{n}{N} l} \\ &= X(n, k) + \frac{1}{N} \left\{ x\left(k - \frac{N-1}{2}\right) e^{-2i\pi \frac{n}{N} \left(k - \frac{N-1}{2}\right)} \right. \\ &\quad \left. - x\left(k + \frac{N-1}{2} + 1\right) e^{-2i\pi \frac{n}{N} \left(k + \frac{N-1}{2} + 1\right)} \right\} \end{aligned}$$

Donc, par différenciation :

$$\begin{aligned} \Delta X(n, k) &= X(n, k+1) - X(n, k) \\ &= \frac{1}{N} \left\{ x\left(k + \frac{N+1}{2}\right) e^{-2i\pi \frac{n}{N} \left(\frac{N+1}{2}\right)} \right. \\ &\quad \left. - x\left(k - \frac{N-1}{2}\right) e^{2i\pi \frac{n}{N} \left(\frac{N-1}{2}\right)} \right\} e^{-2i\pi \frac{n}{N} k} \end{aligned}$$

$$\text{Or } \begin{cases} 2\pi \frac{n}{N} \frac{N-1}{2} = n\pi \frac{N-1}{N} \approx n\pi & \text{si } N \gg 1 \\ 2\pi \frac{n}{N} \frac{N+1}{2} = n\pi \frac{N+1}{N} \approx n\pi & \text{si } N \gg 1 \end{cases} \quad \text{et } e^{in\pi} = e^{-in\pi} = (-1)^n$$

D'où la relation suivante exprimée sous une forme très simple, qui constitue la base des algorithmes de calcul de l'ASD :

$$\boxed{\Delta X(n, k) = \frac{(-1)^n}{N} \left\{ x\left(k + \frac{N+1}{2}\right) - x\left(k - \frac{N-1}{2}\right) \right\} e^{-2i\pi \frac{n}{N}k}} \quad (\text{A.3})$$

Ainsi, cette équation donne la loi de réactualisation des coefficients de l'ASD au cours du temps, sous une forme rappelons le, fort simple à mettre en oeuvre et très peu coûteuse en temps de calcul.

Physiquement, $\Delta X(n, k)$ mesure l'écart à la périodicité en N du signal $x(k)$ au voisinage du point k .

Il est à remarquer, de même, que l'on peut reconstituer le signal original à partir des coefficients de l'ASD sur une seule bande n . En effet, d'après la relation précédente :

$$x\left(k + \frac{N+1}{2}\right) - x\left(k - \frac{N-1}{2}\right) = \frac{(-1)^n}{N} \Delta X(n, k) e^{-2i\pi \frac{n}{N}k} \quad \forall k, n \in \mathbb{N}$$

De plus, il est possible de calculer toutes les bandes de l'ASD à partir d'une seule :

$$\Delta X(p, k) = \Delta X(n, k) e^{-2i\pi \frac{n-p}{N}k}$$

Ainsi, chaque bande individuelle contient toute l'information contenue dans le signal, mais en pratique, il reste intéressant d'en calculer plusieurs, notamment quand le signal est riche en composantes fréquentielles.

Algorithme de calcul

Introduisons les notations suivantes :

$$\begin{cases} X(n, k) = X_{n,k} e^{i\theta_{n,k}} = A + iB \\ \Delta X(n, k) = X(n, k+1) - X(n, k) = \Delta A + i\Delta B \\ \Delta x = x\left(k + \frac{N+1}{2}\right) - x\left(k - \frac{N-1}{2}\right) \end{cases}$$

Ainsi, d'après (A.3) :

$$\begin{aligned} \Delta X(n, k) &= \frac{(-1)^n}{N} \Delta x e^{-2i\pi \frac{n}{N}k} \\ &= \frac{(-1)^n}{N} \Delta x \left\{ \cos\left(2\pi \frac{n}{N}k\right) - i \sin\left(2\pi \frac{n}{N}k\right) \right\} \end{aligned}$$

D'où,

$$\begin{cases} \Delta A = \frac{(-1)^n}{N} \Delta x \cos\left(2\pi \frac{n}{N} k\right) \\ \Delta B = \frac{(-1)^{n+1}}{N} \Delta x \sin\left(2\pi \frac{n}{N} k\right) \end{cases}$$

De même,

$$\begin{cases} A = X_{n,k} \cos(\theta_{n,k}) \\ B = X_{n,k} \sin(\theta_{n,k}) \end{cases}$$

De plus,

$$\begin{aligned} X(n, k+1) &= X(n, k) + \Delta X(n, k) \\ &= (A + \Delta A) + i(B + \Delta B) \\ &= X_{n,k+1} e^{i\theta_{n,k+1}} \end{aligned}$$

avec
$$\begin{cases} X_{n,k+1} = \sqrt{(A + \Delta A)^2 + (B + \Delta B)^2} \\ \theta_{n,k+1} = \arctan\left(\frac{B + \Delta B}{A + \Delta A}\right) \end{cases}$$

de même
$$\begin{cases} X_{n,k} = \sqrt{A^2 + B^2} \\ \theta_{n,k} = \arctan\left(\frac{B}{A}\right) \end{cases}$$

Ainsi,

$$\begin{aligned} \Delta\theta_{n,k} &= \theta_{n,k+1} - \theta_{n,k} \\ &= \arctan\left(\frac{B + \Delta B}{A + \Delta A}\right) - \arctan\left(\frac{B}{A}\right) \end{aligned}$$

Ce qui peut, en 1^{re} approximation, s'approcher par :

$$\Delta\theta_{n,k} \approx \Delta\left(\arctan\left(\frac{B}{A}\right)\right) = \frac{1}{1 + \left(\frac{B}{A}\right)^2} \times \frac{\Delta BA - B\Delta A}{A^2}$$

c'est à dire :

$$\boxed{\Delta\theta_{n,k} \approx \frac{A\Delta B - B\Delta A}{A^2 + B^2} = \frac{(-1)^n}{N} \frac{\Delta x}{X_{n,k}} \sin\left(\theta_{n,k} - 2\pi \frac{n}{N} k\right)} \quad (\text{A.4})$$

L'utilisation de cette relation reste justifiée pour $\frac{1}{N} \left| \frac{\Delta x}{X_{n,k}} \right| \ll 1$ c'est à dire, tant que :

$$\frac{|\Delta x|}{N} \ll |X_{n,k}|$$

Cette relation est d'autant plus vérifiée que le signal est sur-échantillonné.

A.2 Coefficients ASD d'un signal réel

Considérons un signal réel $x(t)$, et sa Transformée de Fourier $X(f)$ définie comme (1.1) :

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-i2\pi ft} dt = A(f) e^{i\Phi(f)} \quad (\text{A.5})$$

Les coefficients d'ASD de ce signal sur une fenêtre de durée T sont définis par :

$$X(n, t) = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} x(u) e^{-i2\pi \frac{n}{T} u} du \quad (\text{A.6})$$

$$= \int_{-\infty}^{+\infty} \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} x(u) e^{-i2\pi fu} du \delta\left(f - \frac{n}{T}\right) df \quad (\text{A.7})$$

Soit :

$$X(n, t) = \int_{-\infty}^{+\infty} \left(X(f) * \text{sinc}(\pi f T) e^{-i2\pi ft} \right) \delta\left(f - \frac{n}{T}\right) df$$

$$\text{avec } \text{sinc}(\pi f T) = \frac{\sin(\pi f T)}{\pi f T}$$

d'où

$$X(n, t) = \int_{-\infty}^{+\infty} X(f') \text{sinc}\left[\pi T \left(\frac{n}{T} - f'\right)\right] e^{-i2\pi \left(\frac{n}{T} - f'\right)t} df' \quad (\text{A.8})$$

A.3 Signal à deux composantes sinusoïdales : phénomènes inter-bandes

Considérons le signal : $x(t) = A_0 \cos(2\pi f_0 t + \varphi_0) + A_1 \cos(2\pi f_1 t + \varphi_1)$.

On en déduit, d'après 1.22, la TF du signal analytique :

$$X_a(f) = A_0 \delta(f - f_0) e^{i\varphi_0} + A_1 \delta(f - f_1) e^{i\varphi_1} \quad \forall f \geq 0 \quad (\text{A.9})$$

or $X(n, t)$ est linéaire, donc $X(n, t) = X_0(n, t) + X_1(n, t)$, ce qui implique d'après (1.24) :

$$X(n, t) = B_0 e^{i[2\pi(f_0 - \frac{n}{T})t + \varphi_0]} + B_1 e^{i[2\pi(f_1 - \frac{n}{T})t + \varphi_1]} \quad (\text{A.10})$$

$$\text{avec } \begin{cases} B_0 = A_0 \text{sinc}\left[\pi T \left(f_0 - \frac{n}{T}\right)\right] \\ B_1 = A_1 \text{sinc}\left[\pi T \left(f_1 - \frac{n}{T}\right)\right] \end{cases}$$

ce qui s'écrit encore sous la forme : $X(n, t) = X_n(t) e^{i\theta_n(t)}$

$$\text{avec } \begin{cases} X_n(t) = \sqrt{B_0^2 + B_1^2 + 2B_0B_1 \cos(2\pi(f_0 - f_1)t + (\varphi_0 - \varphi_1))} \\ \theta_n(t) = \arctan\left[\frac{B_0 \sin(2\pi f_0 t + \varphi_0) + B_1 \sin(2\pi f_1 t + \varphi_1)}{B_0 \cos(2\pi f_0 t + \varphi_0) + B_1 \cos(2\pi f_1 t + \varphi_1)}\right] - 2\pi \frac{n}{T} t \end{cases} \quad (\text{A.11})$$

Comme précédemment, en dérivant la phase $\theta_n(t)$ par rapport au temps, on peut en déduire une estimation de la fréquence instantanée constituée des fréquences f_0 et f_1 , ainsi que des amplitudes B_0 et B_1 :

$$f_n(t) = \frac{1}{2\pi} \theta'_n(t) + \frac{n}{T}$$

Soit :

$$f_n(t) = \frac{B_0^2 f_0 + B_1^2 f_1 + B_0 B_1 (f_0 + f_1) \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)}{B_0^2 + B_1^2 + B_0 B_1 \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)} \quad (\text{A.12})$$

Nous pouvons étudier l'influence d'une composante fréquentielle sur l'autre en supposant $|B_0| \gg |B_1|$.

D'après (A.11),

$$|X(n, t)|^2 = B_0^2 \left(1 + \left(\frac{B_1}{B_0} \right)^2 + 2 \frac{B_1}{B_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right) \quad (\text{A.13})$$

L'expression de la fréquence instantanée se met sous la forme :

$$f_n(t) = f_0 \frac{1 + \left(\frac{B_1}{B_0} \right)^2 \frac{f_1}{f_0} + \frac{B_1}{B_0} \left(1 + \frac{f_1}{f_0} \right) \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)}{1 + \left(\frac{B_1}{B_0} \right)^2 + \frac{B_1}{B_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)} \quad (\text{A.14})$$

En utilisant le développement limité de $\frac{1}{1+x}$ à l'ordre 2 en 0 pour $x = \left(\frac{B_1}{B_0} \right)^2 + \frac{B_1}{B_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1)$, ce qui reste justifié dans le cas où $|B_0| \gg |B_1|$, on a :

$$f_n(t) \approx f_0 \left\{ 1 + \left(\frac{B_1}{B_0} \right)^2 \frac{f_1}{f_0} + \frac{B_1}{B_0} \left(1 + \frac{f_1}{f_0} \right) \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right\} \\ \times \left\{ 1 - \left[\left(\frac{B_1}{B_0} \right)^2 + \frac{B_1}{B_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right] \right. \\ \left. + \left[\left(\frac{B_1}{B_0} \right)^2 + \frac{B_1}{B_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right]^2 \right\}$$

On développe et on ne garde que les puissances de $\frac{B_1}{B_0}$ inférieures à 2 :

$$f_n(t) \approx f_1 \left(1 + \left(\frac{B_1}{B_0} \right)^2 \left(\frac{f_1}{2f_0} - 1 \right) + \frac{B_1}{B_0} \frac{f_1}{f_0} \cos(2\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right. \\ \left. - \left(\frac{B_1}{B_0} \right)^2 \frac{f_1}{2f_0} \cos(4\pi(f_0 - f_1)t + \varphi_0 - \varphi_1) \right) \quad (\text{A.15})$$

Annexe B

Comparatif des différents spectrogrammes

Cette étude a été réalisée par Lionel Cros (DEA SIA, 2004).

B.1 Représentations de différents algorithmes temps-fréquence appliqués au signal test

Ces signaux simples sont de même amplitude et créés en raisonnant sur des fréquences normalisées. Le signal test x_t de 2048 échantillons est ensuite normalisé, puis échantillonné à 2kHz, ce qui le rend audible sur une durée de 1,024s.

Les paragraphes 1.2 et 1.3 ont mis en évidence certaines propriétés de l'ASD afin de caractériser le comportement de notre spectrogramme. L'application du SAFI à des signaux synthétiques de référence va nous permettre d'étudier les différents phénomènes remarquables que l'on peut mettre en évidence. Il faut se baser sur des résultats connus et par conséquent établir une étude comparative avec différents spectrogrammes de référence, largement utilisés et dont les caractéristiques et les propriétés sont souvent exposées [4] [8].

Tout d'abord, rappelons brièvement la signification et la définition de quelques propriétés souhaitables qui vont nous servir d'éléments de comparaisons :

– *Covariance en translation* :

$$y(t) = x(t - t_0)e^{-i2\pi f_0 t} \iff \text{RTF}_y(t, f) = \text{RTF}_x(t - t_0, f - f_0) \quad (\text{B.1})$$

– *Conservation de l'énergie* :

$$\iint_{-\infty}^{+\infty} \text{RTF}_x(t', f') dt' df' = \int_{-\infty}^{+\infty} |x(t)|^2 dt \quad (\text{B.2})$$

sous réserve que la fonction d'ambiguïté de la RTF soit unitaire.

– *Positivité* :

$$\text{RTF}_x(t, f) \geq 0 \quad \forall t, f \in \mathbb{R} \quad (\text{B.3})$$

Le spectrogramme de référence par rapport au SAFI reste bien sûr le spectrogramme de Fourier défini comme le module au carré de la TFCT du signal :

$$S(t, f) = \left| \int_{-\infty}^{+\infty} x(\tau) h^*(t - \tau) e^{-2i\pi f\tau} d\tau \right|^2 \quad (\text{B.4})$$

On peut montrer que les propriétés que possède cette représentation quadratique sont [4] :

- la covariance par translation dans le plan temps-fréquence
- la conservation de l'énergie sous réserve que h soit d'énergie unité
- la positivité

D'autres représentations bilinéaires appartenant aussi à la classe de Cohen possèdent d'autres propriétés intéressantes comme la localisation parfaite des chirps. On s'intéressera en particulier à la Distribution de Wigner-Ville (WV) et à ses dérivées.

Rappelons leur définition respective [8] :

– *Distribution de Wigner-Ville* :

$$\text{WV}(t, f) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi f\tau} d\tau \quad (\text{B.5})$$

– *Distribution de pseudo Wigner-Ville* :

$$\text{PWV}(t, f) = \int_{-\infty}^{+\infty} h(\tau) x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi f\tau} d\tau \quad (\text{B.6})$$

– *Distribution de pseudo Wigner-Ville lissée* :

$$\text{SPWV}(t, f) = \int_{-\infty}^{+\infty} h(\tau) \int_{-\infty}^{+\infty} g(s - t) x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) ds e^{-i2\pi f\tau} d\tau \quad (\text{B.7})$$

– *Distribution de pseudo Wigner-Ville lissée et réallouée (en temps et fréquence)* : [7]

$$\text{RSPWV}(t, f) = \iint_{-\infty}^{+\infty} \text{SPWV}(t, f; g, h) \delta(t' - \hat{t}(x; t, f)) \delta(f' - \hat{f}(x; t, f)) dt df \quad (\text{B.8})$$

$$\text{avec} \begin{cases} \hat{t}(x; t, f) &= t - \frac{\text{SPWV}(t, f; \tau_g, h)}{2\pi \text{SPWV}(t, f; g, h)} \\ \hat{f}(x; t, f) &= f - i \frac{\text{SPWV}(t, f; g, \frac{dh(t)}{dt})}{2\pi \text{SPWV}(t, f; g, h)} \end{cases}$$

Description du principe de la réallocation

Les distributions présentées précédemment, pseudo WV et pseudo WV lissée, contribuent à l'amélioration de la lisibilité des représentations de la DWV. Cependant, quelle que soit l'approche retenue, l'opération de filtrage destinée à atténuer les composantes interférentielles se traduit en contrepartie par une diminution de la résolution dans le plan temps-fréquence. Une façon d'améliorer la lisibilité des représentations quadratiques est de remarquer que l'information de phase relative à la distribution sous-jacente est perdue. L'amélioration proposée en 1976 par Kodera, Gendrin et de Villedary, puis reformulée récemment par Auger et Flandrin [4][7], consiste alors à prendre en compte cette information tout en gardant à l'esprit une distribution d'énergie. Cette opération présente l'avantage de recentrer l'énergie du spectrogramme au voisinage de la loi de fréquence instantanée.

De plus, on peut montrer que le processus de réallocation préserve quelques propriétés intéressantes de la classe de Cohen. En effet, les représentations obtenues sont covariantes par translation dans le plan temps-fréquence et conservent l'énergie. Elles permettent aussi une localisation parfaite des modulations de fréquence linéaires et des impulsions, ce qui reste une propriété assez rare et recherchée. Avant d'être étendue à l'ensemble des représentations de la classe de Cohen, la méthode de réallocation a été proposée dans le but d'améliorer les propriétés de localisation du spectrogramme.

Dans notre étude comparative, nous avons choisi de considérer la réallocation du spectrogramme et de la distribution de pseudo WV lissée. Afin d'obtenir des renseignements complémentaires sur la réallocation temps-fréquence, on pourra se référer à [7] et [17].

Etude comparative sur un signal test

L'étude comparative s'est faite grâce à la toolbox "Temps-Fréquence" sur Matlab, développée par François Auger, Patrick Flandrin, Paulo Gonçalves et Olivier Lemoine. Elle met en oeuvre des algorithmes temps-fréquence efficaces et permet notamment l'application du principe de la réallocation.

Pour simplifier et homogénéiser l'étude, nous avons créé un signal réel qui est compatible avec les différents algorithmes temps-fréquence mis en oeuvre. Ce dernier est constitué d'un ensemble de signaux simples de référence, bien répartis sur le domaine temps-fréquence, et dont les caractéristiques sont parfaitement connues :

- une *sinusoïde* définie par : $x_1(t) = \cos(2\pi 0.8 t)$.
- un *chirp linéaire* décroissant défini par : $x_2(t) = \text{Re} \left[e^{i2\pi(a_0 t + \frac{a_1 t^2}{2})} \right]$ avec a_0 et a_1 choisis pour que le signal commence à la fréquence réduite $\tilde{f} = 0.2$ et finisse à $\tilde{f} = 0.05$.

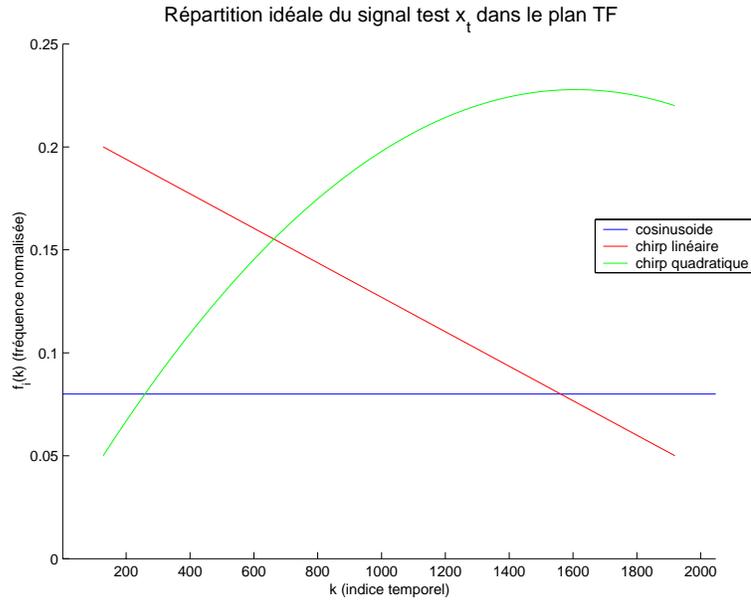


FIG. B.1 – Répartition idéale du signal test $x_t(k)$ dans le plan temps-fréquence pour \tilde{f} allant de 0 à 0.25

- un *chirp quadratique* croissant défini par : $x_3(t) = \text{Re} \left[e^{i2\pi(a_0t + \frac{a_1t^2}{2} + \frac{a_2t^3}{3})} \right]$ avec a_0 et a_1 choisis pour que le signal passe par les fréquences normalisées 0.05, 0.2, 0.22 en des temps particuliers.

Ces signaux simples sont de même amplitude et créés en raisonnant sur des fréquences normalisées. Le signal test x_t de 2048 échantillons est ensuite normalisé, ce qui le rend audible sur une durée de 1,024s avec une fréquence d'échantillonnage de 2kHz.

La figure B.1 montre la répartition de x_t dans le plan temps-fréquence. Afin de comparer au mieux les RTF entre elles, et d'atténuer l'amplitude des termes interférentiels provenant des interactions entre les composantes portées par les fréquences positives et celles portées par les fréquences négatives, nous avons choisi d'effectuer les calculs des spectrogrammes à partir du signal analytique de x_t . L'amplitude est représentée en décibel (dB). Les courbes en pointillé noir (quand elles sont présentes) représentent la RTF idéale du signal x_t .

L'analyse visuelle de ces différentes RTF permet de cerner rapidement l'intérêt de chacune d'elles. On constate que les diverses zones de croisements des signaux élémentaires mettent en difficulté l'ensemble des RTF, souvent caractérisées par des oscillations locales.

Tout d'abord le SAFI (figure B.2) donne une représentation simple et très proche

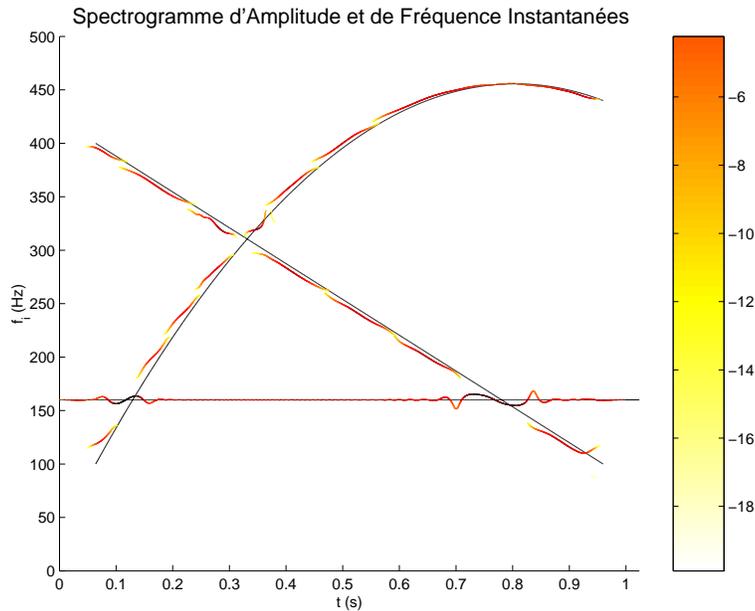


FIG. B.2 – SAFI de x_t .

de la représentation idéale de x_t (figure B.1). Il présente, parfois, un léger écart avec la répartition théorique. Le confort visuel est notable.

Le spectrogramme (figure B.3) donne une représentation approximative de x_t dans le plan temps-fréquence. Il a notamment beaucoup de mal à détecter la partie basse fréquence du chirp quadratique.

La distribution de Wigner-Ville (figure B.5) met en évidence une multitude de termes d'interférence qui gênent considérablement la lisibilité de la RTF. Par contre, elle permet une excellente localisation des chirps sous réserve de connaître *a priori* la constitution du signal étudié et les propriétés de symétrie des interférences. La distribution de pseudo WV (figure B.6) atténue l'amplitude des termes interférentiels et le lissage effectué ensuite (figure B.7) permet de les faire disparaître en quasi-totalité. La lisibilité est donc améliorée au prix d'une perte de localisation. La précision est, tout de même, approximativement le double de celle du spectrogramme. Le temps de calcul augmente logiquement avec le nombre de corrections appliquées à la distribution de WV.

La réallocation en temps et en fréquence du spectrogramme (figure B.4 XIV) et de la distribution de pseudo WV lissée (figure B.8) augmente la lisibilité des RTF de manière cruciale. La différence visuelle entre les deux représentations n'est, par contre, pas flagrante alors que le coût calculatoire varie d'un facteur 2 entre le spectrogramme et la distribution de WV (cf. tableau B.2). Cette dernière notion est d'ailleurs la plus importante, car on arrive facilement à plusieurs minutes de calcul sur un P4 2.4GHz, ce

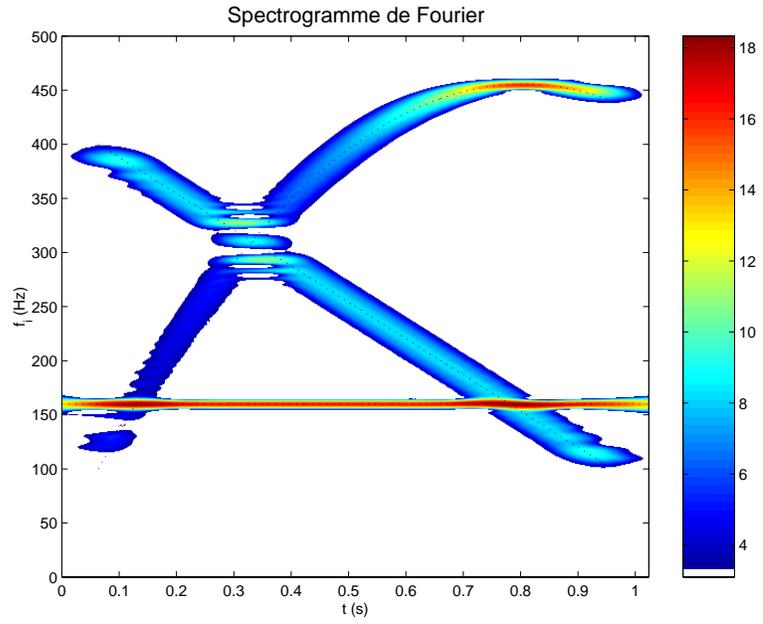


FIG. B.3 – Spectrogramme classique de x_t .

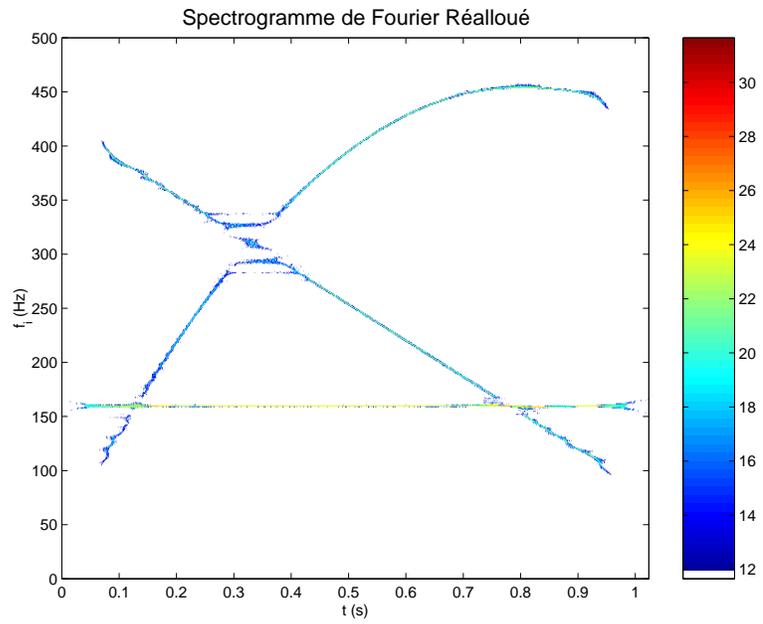


FIG. B.4 – Spectrogramme classique de x_t réalloué.

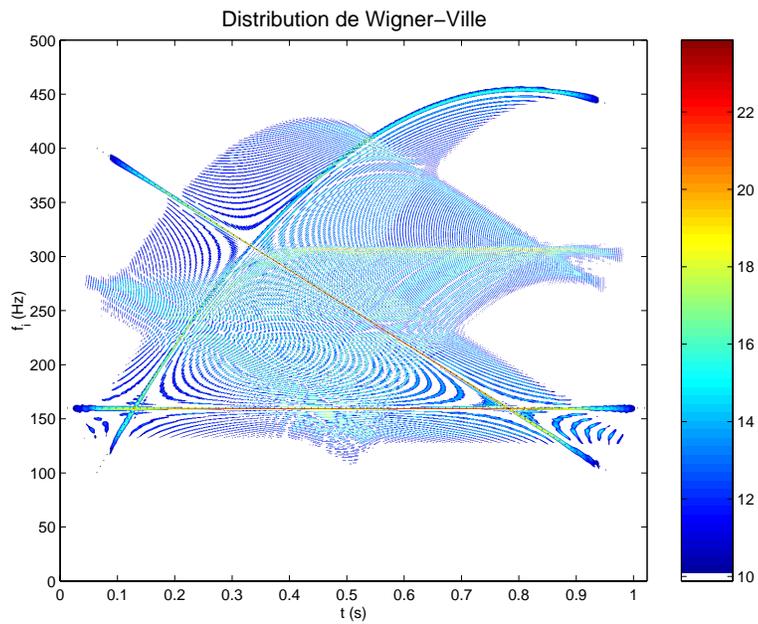


FIG. B.5 – RTF de la distribution de Wigner-Ville de x_t .

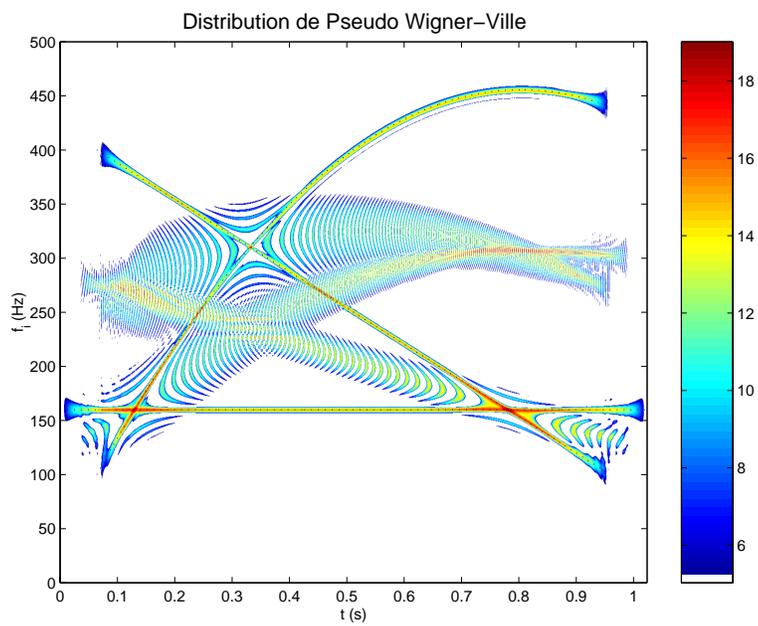


FIG. B.6 – RTF de la pseudo distribution de Wigner-Ville de x_t .

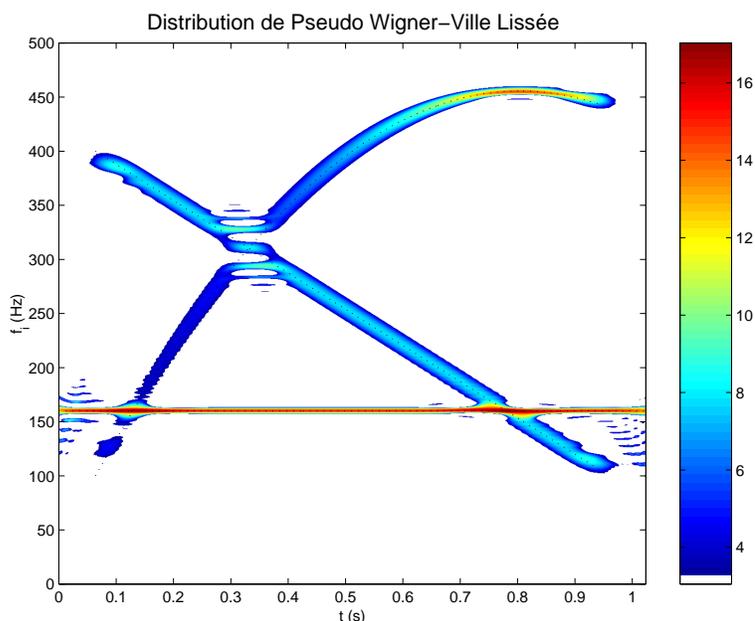


FIG. B.7 – RTF de la distribution pseudo Wigner-Ville lissée de x_t .

	Spectro-gramme	DWV	PWV	SPWV	SPWV réalloué	SAFI
<i>Covariance en translation</i>		✓	✓	✓	✓	✓
<i>Conservation de l'énergie</i>	✓	✓	✓	✓	✓	✓
<i>Positivité</i>	✓					✓
<i>Excellente localisation des chirps</i>		✓			✓	✓
<i>Localisation parfaite des sinusoides</i>		✓			✓	✓

TAB. B.1 – Tableau récapitulatif de diverses propriétés mathématiques de base des différentes RTF étudiées

qui diminue considérablement son domaine d'application. En effet, dans ce cas, il n'est pas réalisable, par exemple, d'utiliser la méthode de réallocation sur des signaux réels comportant des millions d'échantillons.

Ainsi, on remarque que le SAFI possède une excellente lisibilité même si l'on peut observer de légers écarts avec la représentation théorique, et cela avec un coût calculatoire faible, sans commune mesure avec la méthode de réallocation. L'inconvénient toutefois, reste le choix délicat et souvent empirique, des paramètres du SAFI. Un avantage certain, est par contre, la possibilité de visualisations et d'interprétations avec précision qu'offre la représentation en trois dimensions du SAFI mais aussi la possibilité de calculer en une durée raisonnable, la RTF d'un signal sonore sur un très grand nombre d'échantillons.

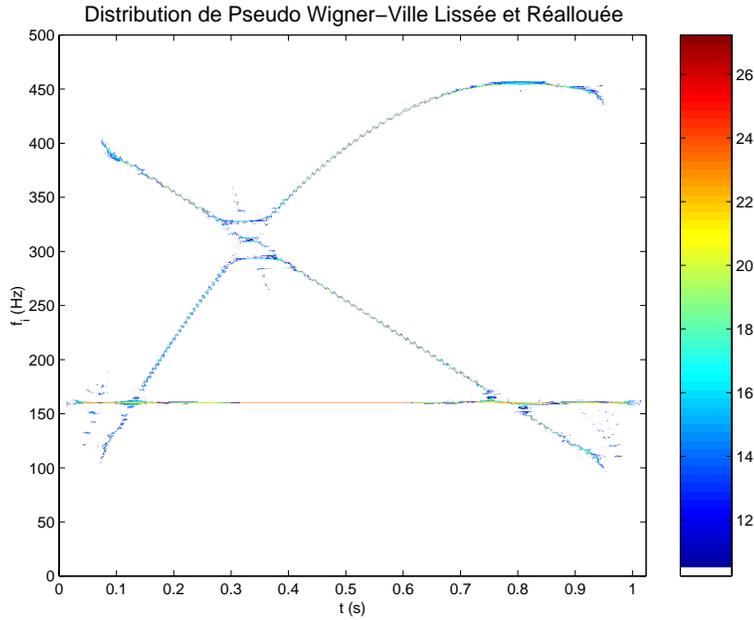


FIG. B.8 – RTF de la distribution pseudo Wigner-Ville de x_t lissée et réallouée.

B.1.1 Robustesse au bruit

Dans le but de tester la robustesse au bruit des différentes RTF présentées ci-dessus, un même bruit blanc gaussien additif a été ajouté à chaque nouvelle analyse du signal x_t . Le rapport signal sur bruit (RSB_{dB}) est estimé à partir de la formule empirique suivante :

$$RSB_{dB} = 10 \log_{10} \left(\frac{\frac{1}{N} \sum_{k=0}^{N-1} x_t^2(k)}{\sigma_b^2} \right) \quad (\text{B.9})$$

σ_b^2 représentant la puissance du bruit blanc.

Le SAFI (figure B.9) détecte distinctement les trois composantes contenues dans le signal x_t et ce, malgré le bruit qui génère de légères oscillations autour de la valeur théorique. Le spectrogramme (figure B.10) par contre, donne une RTF très bruitée qui rend la lecture de l'image très difficile et très incertaine. Les termes d'interférence ajoutés au bruit gaussien, rendent la représentation de la DWV (figure B.12) et de la distribution de pseudo WV (figure B.13) quasiment ininterprétables, surtout si l'on ne dispose pas d'informations *a priori* sur la composition du signal. Par contre, le lissage de la distribution de pseudo WV (figure B.14) permet une lecture beaucoup plus fiable que le spectrogramme et plus agréable que la DWV. Enfin la réallocation en temps et en fréquence du spectrogramme (figure B.11) améliore la lisibilité et l'interprétabilité de celui-ci. Ceci se

vérifie sur la distribution de pseudo WV (figure B.15) lissée et réallouée. Cette dernière représentation est assez proche de la répartition théorique dans le plan temps-fréquence (figure B.1).

Le SAFI se révèle donc être assez robuste au bruit blanc comparé aux autres RTF. La lisibilité de la distribution de WV réallouée et du SAFI présentent des similarités, ce qui peut amener à effectuer des comparaisons directes entre ces deux représentations.

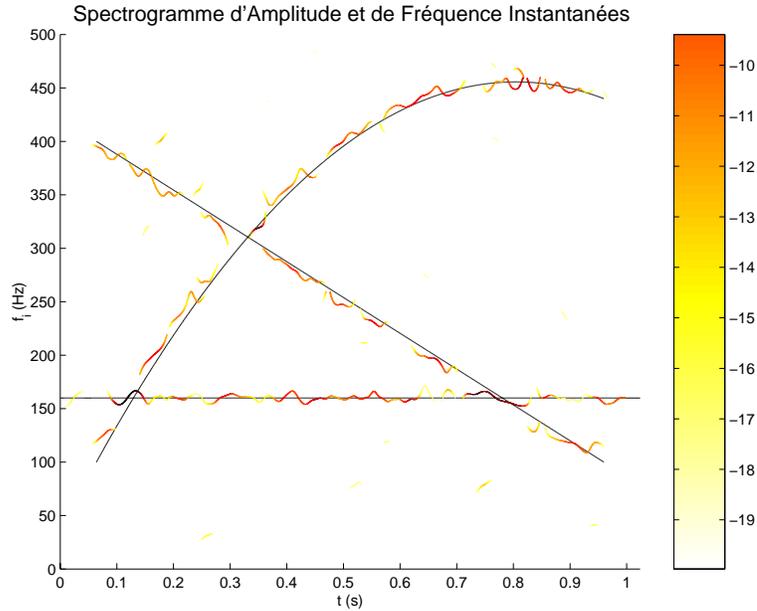


FIG. B.9 – SAFI de x_t .

B.1.2 Etude comparative de la complexité de calcul

Un problème essentiel souvent occulté dans les écrits traitant des RTF reste les difficultés pratiques liées à l'évaluation de ces représentations, et en particulier au volume de calcul souvent prohibitif qu'induit l'analyse temps-fréquence d'un signal dit "long".

Pour tenter de palier à ce problème, on recense deux approches. Une première vise à limiter la redondance des calculs, et une seconde consiste en l'évaluation récursive des RTF. Nous pouvons remarquer dès à présent que le SAFI englobe ces deux notions. En effet, le calcul des coefficients de l'ASD à l'aide d'une fenêtre rectangulaire, du fait de sa symétrie particulière, permet d'en déduire une relation de récurrence simple (1.15) permettant d'estimer la fréquence instantanée avec une faible complexité de calcul.

Un aperçu de la remarquable efficacité des algorithmes récursifs pour le calcul des RTF est disponible dans [16].

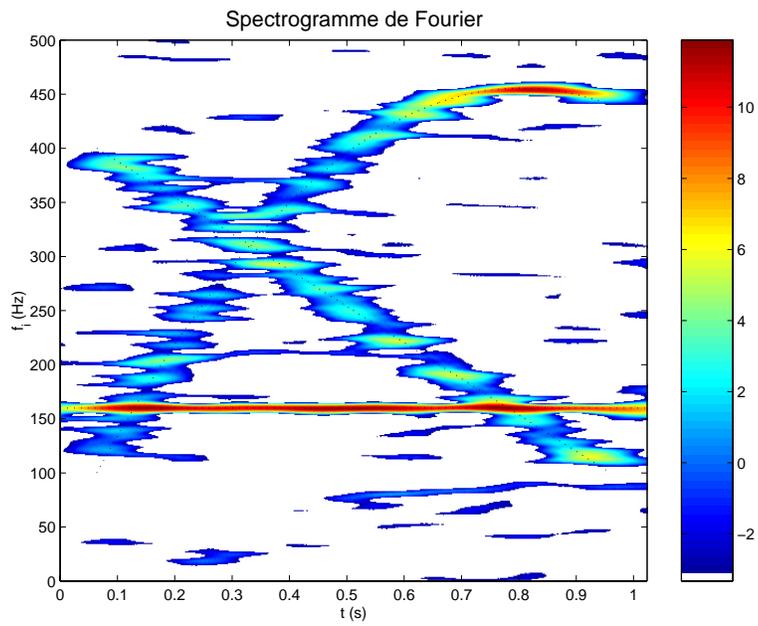


FIG. B.10 – Spectrogramme classique de x_t .

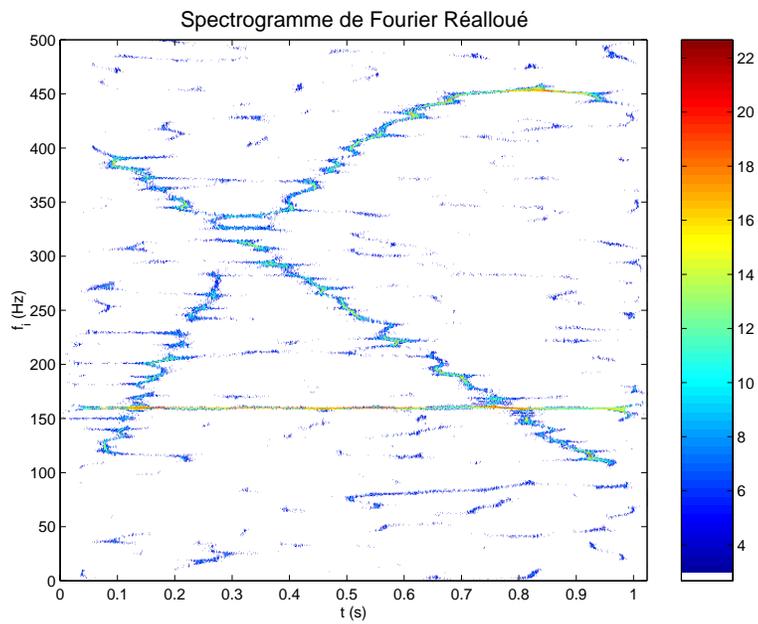


FIG. B.11 – Spectrogramme classique de x_t réalloué.

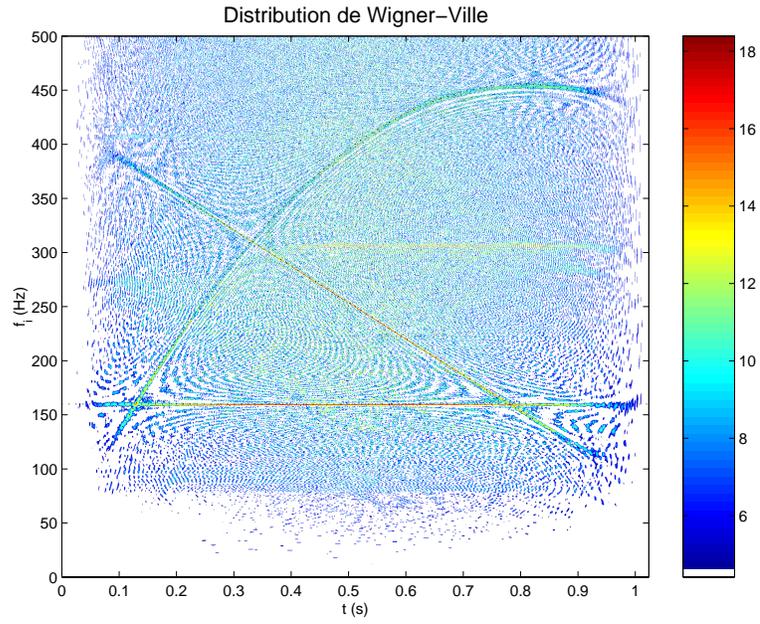


FIG. B.12 – RTF de la distribution de Wigner-Ville de x_t .

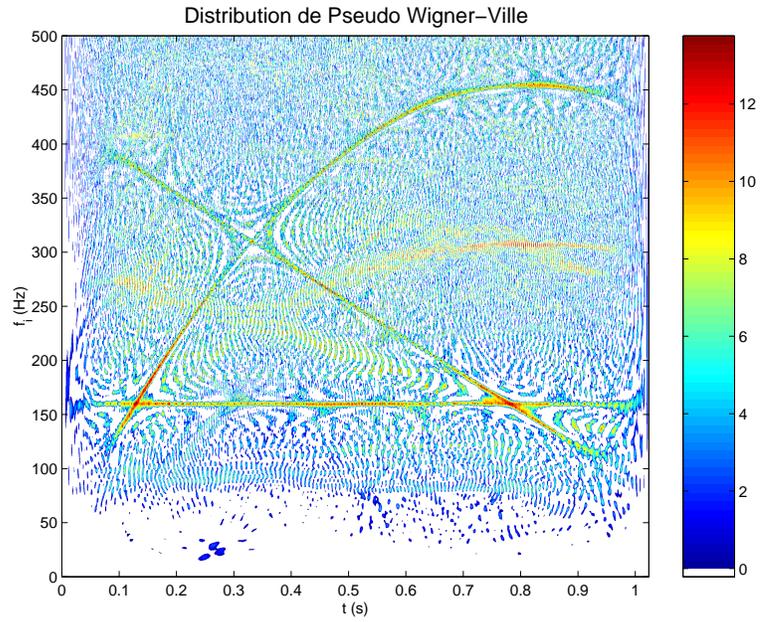


FIG. B.13 – RTF de la pseudo distribution de Wigner-Ville de x_t .

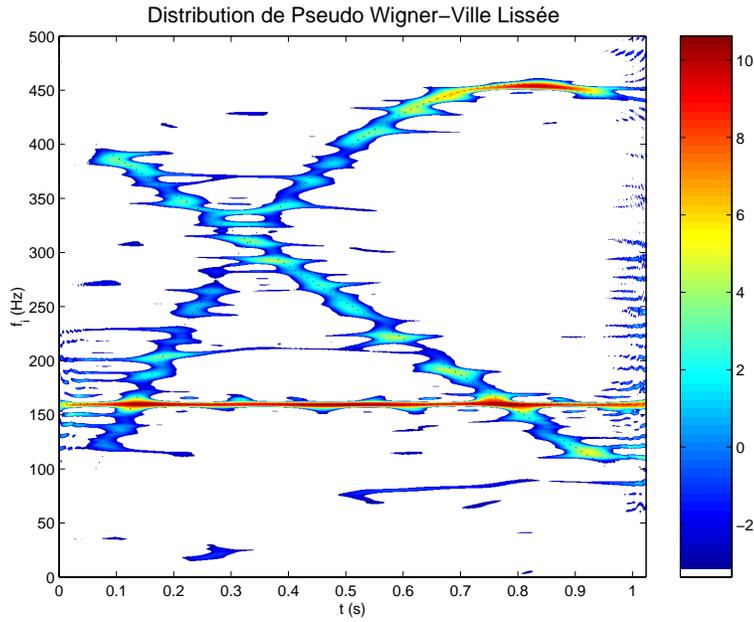


FIG. B.14 – *RTF de la distribution pseudo Wigner-Ville lissée de x_t .*

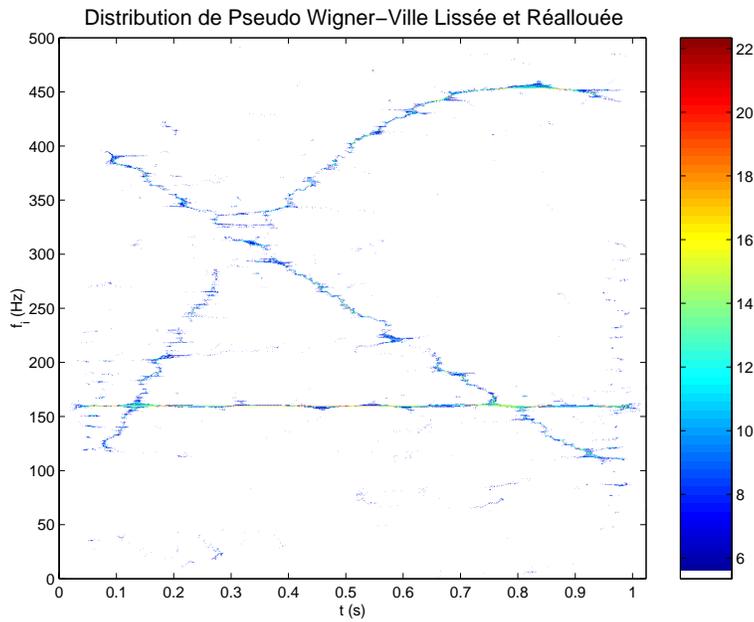


FIG. B.15 – *RTF de la distribution pseudo Wigner-Ville de x_t lissée et réallouée.*

L'efficacité de chacun des algorithmes temps-fréquence est évaluée au moyen de calculs de complexité¹. Par souci d'homogénéité, on choisira dans chacun des cas qui suivent, une fenêtre rectangulaire dont la longueur correspond à une puissance de 2, afin de permettre l'usage optimal de la Transformée de Fourier Rapide (TFR) lorsque cela est possible.

Ainsi l'évaluation de la TFCT $X(k, n/N)$ pour $n = 0, \dots, N - 1$, résulte de la TFR d'un signal de N échantillons. Par conséquent, le calcul de la TFCT nécessite $N \log_2(N)$ additions complexes A_c et $\frac{N}{2} \log_2 N$ multiplications complexes M_c à chaque instant. Afin de pouvoir comparer ces résultats entre eux, il est indispensable de convertir les opérations complexes en opérations réelles. Pour cela, on peut remarquer qu'une addition complexe A_c fait intervenir deux additions réelles A_r et qu'une addition complexe M_c requiert quatre multiplications réelles M_r et deux additions réelles. On a ainsi la correspondance suivante[15] :

$$\begin{cases} A_c &= 2A_r \\ M_c &= 4M_r + 2A_r \end{cases} \quad (\text{B.10})$$

Donc, le calcul direct de la TFCT nécessite, pour une fenêtre rectangulaire, à chaque instant, le nombre d'opérations réelles suivant :

$$\text{TFCT} \Rightarrow \begin{cases} A_r &= 3N \log_2(N) \\ M_c &= 2N \log_2(N) \end{cases} \quad (\text{B.11})$$

De plus, avec le même raisonnement que précédemment, on montre que la distribution de pseudo WV lissée et réallouée nécessite, pour une fenêtre rectangulaire, à chaque instant, le nombre d'opérations réelles suivant [16] :

$$\text{RSPWV} \Rightarrow \begin{cases} A_r &= 9L \log_2(L) + (12M + 11)L \\ M_c &= 6L \log_2(L) + (8M + 15)L \end{cases} \quad (\text{B.12})$$

où les demi-largeurs des fenêtres h et g de lissage temporel et fréquentiel sont notées respectivement M et L . La fenêtre g est supposée à valeurs complexes¹.

On peut remarquer que ces deux distributions basées sur des algorithmes rapides sont de complexité en $O(N \log_2(N))$. Une possibilité de diminuer la complexité de calcul est de considérer les formes récursives correspondantes aux expressions précédentes. On montre [16], que l'on peut atteindre une complexité linéaire, notamment dans le cas de l'expression récursive de la TFCT.

Si on s'intéresse maintenant à l'étude de la complexité de l'ASD, on remarque d'après

¹Pour plus de détail, on pourra se reporter à [16] où l'on peut trouver une étude plus complète de la complexité de calcul de différents algorithmes temps-fréquence

	Temps de calcul
SAFI	400 (31 ms/bande)
PDWV	1 800
DWV	2 300
Spectrogramme	2 700
SPDWV	90 000
Spectrogramme réalloué	100 000
SPDWV réalloué	200 000

TAB. B.2 – *Tableau comparatif sur le temps de calcul, en ms, des différents algorithmes temps-fréquence considérés.*

(1.15) que les coefficient de l'ASD peuvent s'écrire sous la forme récursive suivante :

$$\begin{aligned}
 X(n, k + 1) &= X(n, k) + \Delta X(n, k) \\
 &= \left\{ x\left(k + \frac{N + 1}{2}\right) - x\left(k - \frac{N - 1}{2}\right) \right\} \frac{(-1)^n}{N} e^{-2i\pi \frac{n}{N}k}
 \end{aligned}$$

Ainsi, on en déduit que la complexité de calcul de $X(n, k)$ est linéaire ($O(N)$).

Annexe C

Echelle fréquentielle logarithmique : principe de construction

C.1 Principe de l'échelle logarithmique

Afin que l'analyse et la représentation se rapprochent de la perception, une première approximation consiste à utiliser une échelle fréquentielle dont la répartition des fréquences centrales et des largeurs des bandes est logarithmique. Nous allons montrer que cela implique un système à T_m variable et $n_m = n$ constant. Nous verrons ensuite comment choisir la répartition des bandes.

Les données initiales du problème sont les suivantes :

- une analyse entre f_{min} et f_{max} .
- une précision de référence Δf_{ref} pour une fréquence de référence f_{ref} .
- une distribution logarithmique des fréquences.

Le principe de distribution logarithmique des fréquences se traduit par la relation suivante : pour une fréquence $f = \alpha f_{ref}$ donnée, la largeur de bande associée est $\Delta f = \alpha \Delta f_{ref}$.

D'où $\frac{n}{T} = \alpha \frac{n_{ref}}{T_{ref}}$ et $\frac{2}{T} = \alpha \frac{2}{T_{ref}}$ puis $n = n_{ref}$.

n est donc une constante quelle que soit la bande, avec $n = \frac{f}{\Delta f}$. La bande m est alors caractérisée par une fréquence centrale $f_m = \frac{n}{T_m}$ et une largeur de bande $\Delta f_m = \frac{2}{T_m}$, où T_m est la longueur de la fenêtre d'analyse.

C.2 Construction de l'échelle logarithmique

Il s'agit de construire une suite de bandes m . On fixe d'abord n : sa valeur théorique est $\frac{f_{ref}}{\Delta f_{ref}}$ mais n doit être entier, ce qui revient à prendre la partie supérieure de cette valeur. On a donc $n = \text{ceil}(\frac{f_{ref}}{\Delta f_{ref}})$. Il reste à construire la suite des T_m . En numérique,

Annexe C - Echelle fréquentielle logarithmique : principe de construction

cela revient à construire la suite des $N_m = T_m F_s$ qui étaient auparavant constants (N_{asd}).

Nous prenons en compte les deux contraintes :

- $N_m \in N$ (contrainte liée au calcul numérique)
- $f_{m+1} - f_m < \frac{\Delta f_{m+1} - \Delta f_m}{2}$ (contrainte de recouvrement des bandes)

Le choix pour $m = 0$ s'obtient en considérant que f_m doit appartenir à la première bande :

$$f_0 - \frac{\Delta f_0}{2} \leq f_{min} < f_0 + \frac{\Delta f_0}{2} \quad (C.1)$$

ie

$$n \frac{F_s}{N_0} - \frac{F_s}{N_0} \leq f_{min} < n \frac{F_s}{N_0} - \frac{F_s}{N_0} \quad (C.2)$$

ie

$$(n-1) \frac{F_s}{f_{min}} \leq N_0 < (n+1) \frac{F_s}{f_{min}} \quad (C.3)$$

$N_0 = \left\lfloor \frac{n F_s}{f_{min}} \right\rfloor$ convient, où $\lfloor x \rfloor$ désigne la partie entière de x .

Construisons N_{m+1} à partir de N_m . La contrainte de recouvrement impose

$$f_{m+1} - f_m < \frac{\Delta f_{m+1} + \Delta f_m}{2} \quad (C.4)$$

ie

$$\frac{n F_s}{N_{m+1}} - \frac{n F_s}{N_m} < \frac{F_s}{N_{m+1}} + \frac{F_s}{N_m} \quad (C.5)$$

puis

$$N_{m+1} > \frac{n-1}{n+1} N_m \quad (C.6)$$

On doit donc choisir N_{m+1} dans l'intervalle $\left[\left\lfloor \frac{n-1}{n+1} N_m \right\rfloor ; N_m \right]$ où $\lfloor x \rfloor$ désigne la partie supérieure de x . En première approximation, la valeur $N_{m+1} = \frac{n}{n+1} N_m$ convient. Elle se situe au milieu de l'intervalle précédent. Le choix de N_{m+1} pose problème lorsque l'intervalle $\left[\left\lfloor \frac{n-1}{n+1} N_m \right\rfloor ; N_m \right]$ ne contient pas d'entier. Ceci peut être le cas dès que $N_m - N_{m+1} < 1$:

$$m > \frac{\ln\left(\frac{n+1}{\left\lfloor \frac{n F_s}{f_{min}} \right\rfloor}\right)}{\ln\left(\frac{n}{n+1}\right)} \quad (C.7)$$

Une répartition possible des bandes est donc définie par $N_m = \left\lfloor \left(\frac{n}{n+1}\right)^m N_0 \right\rfloor$ et $N_0 = \left\lfloor \frac{n F_s}{f_{min}} \right\rfloor$. Les fréquences centrales associées sont $f_m = \frac{n F_s}{\left\lfloor \left(\frac{n}{n+1}\right)^m N_0 \right\rfloor}$. D'après ce que nous venons de voir, cette répartition est valable jusqu'à une certaine fréquence, il convient donc de vérifier que f_{max} est dans le domaine de validité.

Annexe D

Acronymes et notations mathématiques

Acronymes :

ASD	Analyse Spectrale Différentielle
DWV	Distribution de Wigner-Ville
RTF	Représentation Temps Fréquence
SAFI	Spectrogramme d'Amplitude et de Fréquence Instantanées
TF	Transformée de Fourier
TFCT	Transformée de Fourier à Court Terme
TFD	Transformée de Fourier Discrète
TFR	Transformée de Fourier Rapide
WV	Wigner-Ville

Notations mathématiques :

$\theta(t)$	phase instantanée
$A(t)$	amplitude instantanée
$f_i(n, t)$	fréquence instantanée calculée pour la bande n
$f_i(t)$	fréquence instantanée
T	durée de la fenêtre d'analyse
$x(*)$	signal (continu (t) ou discret (k))
$X(*)$	TF ou TFD de x
$X(*, *)$	coefficients de l'ASD de x
$x_a(*)$	signal analytique correspondant à x
$X_a(*)$	TF ou TFD de x_a
$X_h(*, *)$	TFCT de x avec la fenêtre de pondération h
$\arg\{*\}$	argument de $*$
$\text{Im}\{*\}$	partie imaginaire de $*$
$\text{Re}\{*\}$	partie réelle de $*$
$\text{RTF}_*(t, f)$	représentation temps-fréquence de $*$
\otimes	produit de convolution

Annexe E

Description de l'algorithme basé sur l'ASD

Cette annexe a pour but de donner un aperçu général du déroulement de l'algorithme principal de calcul du spectrogramme (SAFI).

Récupération des variables et des paramètres

- Caractéristiques du signal : x , FS , L_{sig} .
- Paramètres de l'analyse : N_{asd_m} , n_m , m .
- Paramètres de post-traitement : $THRESHOLD$, $PHASE_FILTER_...$

Initialisations et calculs de variables et de paramètres

- Sélection des numéros de bandes ($n_m(m)$) et de la largeur de la fenêtre associée en nombre d'échantillons ($N_{asd_m}(m)$)
- Calcul du nombre de bandes sélectionnées (Nb_{band})
- Calcul de la longueur ($L_{seg_{in}}$) et du nombre (Nb_{seg}) de segments
- Calcul des fréquences limites entre les bandes d'analyse ($freq_{lim}$)
- Calcul de la largeur maximale de la fenêtre d'analyse ($N_{asd_{max}}$)
- Calcul de la longueur ($L_{seg_{out}}$) des vecteurs contenant les incréments $\Delta S_{n,k}$
- Ajustement de la longueur du signal par ajout de 0 en fin de signal

Calcul de l'ASD

1. Calcul de la constante pour chaque bande m :

$$cst(m) = -i2\pi \frac{n_m(m)}{N_{asd_m}(m)} \quad (E.1)$$

2. Calcul des valeurs initiales de $S_{n,k}$:

$$\begin{aligned} S_{init_{n,k}}(m) &= \frac{1}{N_{asd_m}(m)} \sum_{k=1}^{N_{asd_m}(m)} x(k) e^{cst(m) \frac{(k-1)}{N_{asd_m}(m)}} \\ &= \frac{1}{N_{asd_m}(m)} \sum_{k=1}^{N_{asd_m}(m)} x(k) e^{-i2\pi \frac{(k-1) n_m(m)}{N_{asd_m}(m)}} \\ &= \frac{1}{N_{asd_m}(m)} TFD \left[x(1 : N_{asd_m}(m)) \right] \end{aligned} \quad (E.2)$$

3. Pour chaque segment :

- (a) Calcul des indices correspondants aux segments (n_{seg})
- (b) (facultatif) Calcul du signal analytique par la transformée d'Hilbert
- (c) Calcul de la stroboscopie :

$$seg_{exp} = \left(x_e(m, k) \right)_{Nb_{band} \times L_{seg_{in}}} \quad (E.3)$$

avec :

$$\begin{aligned} x_e(m, k) &= x(k) e^{cst(m)} \\ &= x(k) e^{-i2\pi \frac{n_m(m)}{N_{asd_m}(m)}} \end{aligned} \quad (E.4)$$

- (d) Calcul des incréments des coefficients ASD :

$$\Delta S_{n,k} = \left(\Delta S_{n,k}(m, k) \right)_{Nb_{band} \times L_{seg_{out}}} \quad (E.5)$$

avec :

$$\Delta S_{n,k}(m, k) = x_e(m, k + N_{asd_m}(m)) - x_e(m, k) \quad (\text{E.6})$$

(e) Calcul des coefficients ASD par filtrage, selon le filtre : $F(z) = \frac{1}{1-z^{-1}}$.

Ainsi,

$$S_{n,k} = \left(\left(S_{init_{n,k}} \right)_{Nb_{band} \times 1} \left(S_{n,k}(m, K) \right)_{Nb_{band} \times L_{segout} - 1} \right) \quad (\text{E.7})$$

avec :

$$S_{n,k}(m, K) = S_{init_{n,k}}(m) + \sum_{k=N_{asd_m}(m)+1}^K \Delta S_{n,k}(m, k) \quad (\text{E.8})$$

(f) Mise à jour de $S_{init_{n,k}}$:

$$S_{init_{n,k}} = S_{n,k}^T(end) \quad \forall m \quad (\text{E.9})$$

(g) Suppression du dernier terme de $S_{n,k}$ pour éviter la redondance :

$$S_{n,k} = \left(S_{n,k}(m, K) \right)_{Nb_{band} \times L_{segout}} \quad (\text{E.10})$$

(h) Calcul de $|S_{n,k}|$

(i) Calcul des incréments de phase :

$$\Delta \theta_{n,k} = \frac{A_{n,k} \Delta B_{n,k} - B_{n,k} \Delta A_{n,k}}{A_{n,k}^2 + B_{n,k}^2} \quad (\text{E.11})$$

avec $S_{n,k}$ réécrit sous la forme :

$$S_{n,k} = A_{n,k} + iB_{n,k} \quad (\text{E.12})$$

(j) (facultatif) Lissage des incréments de phase

(k) Estimation de la phase instantanée ($\theta_{n,k}$) telle que :

$$\theta_{n,k}(m, k) = \begin{cases} \Delta\theta_{n,k}(m, 1) & \text{si } k = 1 \\ \theta_{n,k}(m, k-1) & \text{si } k \in [2; L_{segout}] \end{cases} \quad (\text{E.13})$$

(somme cumulée)

(l) Estimation de la fréquence instantanée (f_{inst}) tel que :

$$f_{inst}(m, k) = \left(\frac{1}{2\pi} \Delta\theta_{n,k}(m, k) + \frac{n_m(m)}{N_{asdm}(m)} \right) \cdot FS \quad (\text{E.14})$$

(m) Suppression de l'information hors de la bande d'analyse

(n) Suppression de l'information en dessous du seuil (*THRESHOLD*)

(o) Estimation de l'amplitude instantanée (A) tel que :

$$A(m, k) = \frac{|S_{n,k}|}{\text{sinc} \left[\left(f_{inst}(m, k) - \frac{n_m(m)}{N_{asdm}(m)} FS \right) \times \frac{N_{asdm}(m)}{FS} \right]} \quad (\text{E.15})$$

(p) Sauvegarde des variables principales

4. Reconstitution des segments

<i>A</i>	$Nb_band * Lsig$	Flottant	Amplitude des différentes composantes
<i>dSnk</i>	$Nb_band * Lseg_out$	Flottant	Incréments des coefficients d'ASD
<i>Finst</i>	<i>Lsig</i>	Flottant	Estimation de la fréquence instantanée
<i>freq_lim</i>	$Nb_band + 1$	Flottant	Liste de fréquences qui délimitent les bandes les unes des autres
<i>FS</i>	1	Entier	Fréquence d'échantillonnage du signal
<i>HILBERT_IN_SEGMENT</i>	1	Binaire	Calcul de Hilbert par segment
<i>Lseg_in</i>	1	Entier	Nombre d'échantillons par segments
<i>Lseg_out</i>	1	Entier	Nombre d'échantillons en sortie
<i>Lsig</i>	1	Entier	Nombre total d'échantillons du signal ($Lsig = Nb_seg * Lseg_in$)
<i>m</i>	Nb_band	Entier	Indice de la bande d'analyse
<i>n_seg</i>	$Lseg_in$	Entier	Indices des échantillons du signal à sélectionner
<i>Nasd_m</i>	Nb_band	Entier	Largeur de la fenetre d'analyse en nombre de points
<i>Nasd_max</i>	1	Entier	Largeur de la fenetre d'analyse maximale en nombre de points
Nb_band	1	Entier	Nombre de bandes sélectionnées pour l'analyse
Nb_seg	1	Entier	Nombre de segment
<i>nm</i>	Nb_band	Entier	Numéro de la bande d'analyse de largeur $T : fm = nm/T$
<i>no_seg</i>	1	Entier	Indice du segment en cours
<i>Phase</i>	<i>Lsig</i>	Flottant	Estimation de la phase
<i>PHASE_FILTER_FC</i>	1	Entier	Fréquence de coupure du filtre de lissage de la phase
<i>PHASE_FILTER_ORDER</i>	1	Entier	Ordre du filtre de lissage de la phase
<i>seg</i>	$Nb_band * Lseg_in$	Flottant	Segment de signal
<i>seg_exp</i>	$Nb_band * Lseg_in$	Flottant	Segment de signal multiplié terme à terme par l'exponentielle
<i>segmentation</i>	1	Binaire	Effectuer un calcul en segmentant le signal
<i>sig</i>	<i>Lsig</i>	Flottant	Echantillons du signal
<i>Sinit</i>	Nb_band	Flottant	Coefficient de l'ASD initial $1/N * TFD(x)$ ou $Snk(end)$
<i>Snk</i>	$Nb_band * Lseg_out$	Flottant	Coefficient de l'ASD
<i>SPECTRO_AMPLITUDE</i>	1	Binaire	Calculer le spectre d'amplitude
<i>t</i>	<i>Lsig</i>	Flottant	Vecteur temps
<i>THRESHOLD</i>	1	Flottant	Seuil
<i>z1</i>	$Nasd_max$	Flottant	Valeurs initiales du 1er filtre de lissage de la phase
<i>z2</i>	$Nasd_max$	Flottant	Valeurs initiales du 2eme filtre de lissage de la phase

Bibliographie

- [1] Vincent Gibiat, Pierre Jardin, François Wu : *Analyse spectrale différentielle - Application aux signaux sonars de Myotis Mystacinus*, ACUSTICA Vol. 63 (1987), pp. 90-99.
- [2] Vincent Gibiat, François Wu, Pierre Perio, Sylviane Chaintreuil, André Blanc-Lapierre : *SIGNAUX ET SYSTEMES - Analyse spectrale différentielle (A.S.D.)* , C.R. Acad. Sc. Paris, t.294 (15 mars 1982), Série II pp. 633-636.
- [3] François Léonard : *Spectrogramme de phase et spectrogramme de fréquence*, Traitement du Signal, Vol. 17 no. 4 (2000), pp. 269-286.
- [4] Patrick Flandrin : *Temps-fréquence*, Springer Verlag, Berlin 1972.
- [5] J.L. Flanagan : *Speech Analysis Synthesis and Perception*, Hermes, Paris 1993.
- [6] Pascale Jardin : *Évaluation des performances d'une technique de Fourier glissant (Analyse Spectrale Différentielle) pour le traitement des signaux de parole*, thèse de l'Université Pierre et Marie Curie - Paris 6, soutenue le 30 novembre 1984.
- [7] François Auger, Patrick Flandrin : *Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method*, IEEE Transactions on signal processing, Vol. 43 no. 5 (mai 1995), pp. 1068-1089.
- [8] François Auger, Patrick Flandrin, Paulo Gonçalves, Olivier Lemoine : *Time-Frequency Toolbox : For use with MATLAB - Tutorial*, CNRS & Rice University (1995-1996).
- [9] Nathalie DELPRAT : *Analyse Temps-Fréquence de Sons Musicaux : Exploration d'une Nouvelle Méthode d'Extraction de Données Pertinente pour un Modèle de Synthèse*, thèse de l'Université d'Aix-Marseille II, soutenue le 2 avril 1992.
- [10] Daniel Pressnitzer, Stephen McAdams : *Two phase effects in roughness perception*, J. Acoust. Soc. Am., Vol. 105 no. 5 (mai 1999), pp. 2773-2782.
- [11] T.A.C.M Claasen, W.F.G. Mecklenbräuker : *Time-Frequency Signal Analysis by means of the Wigner Distribution*, Philips Research Laboratories, Eindhoven, The Netherlands, Proceedings of IEEE international conference on acoustics, speech and signal processing, 30,31 mars & 1 avril 1981.
- [12] J. Ville : *Théorie et applications de la notion de signal analytique*, C&T, 2 A., no.1 (1948).
- [13] Stéphane Chabrel, Gérard Charbonneau : *Un système d'analyse de sons par ordinateur fonctionnant en mode conversationnel*, L'onde électrique, Vol. 56 no. 8-9 (août-septembre 1976), pp.358-366.
- [14] Hervé Carfantan : *Traitement Numérique du Signal - Notes de cours*, cours de l'Université Paul Sabatier de Toulouse, France, 3 mars 2003.

Annexe E - BIBLIOGRAPHIE

- [15] Murat Kunt : *Traitement numérique des signaux*, Traité d'Électricité, 2^e édition, 1989.
- [16] Cédric Richard : *Une méthodologie pour la Détection à Structure Imposée - Applications au plan temps-fréquence*, thèse de l'Université de Technologie de Compiègne, soutenue le 22 décembre 1998.
- [17] Eric Chassande-Mottin : *Méthodes de réallocation dans le plans temps-fréquence pour l'analyse et le traitement de signaux non stationnaires*, thèse de l'Université de Cergy-Pontoise, soutenue le 28 septembre 1998.
- [18] John Chowning : *The synthesis of complex audio spectra by means of frequency modulation*, Journal of the Audio Engineering Society no. 21, pp.526-534, 1973.