

Traitement et représentation temps-fréquence des sons avec résolution adaptative

Sonia HAMNANE

Juin 2006

Rapport de stage

Master ATIAM (Acoustique Traitement du signal et Informatique Appliqués à la Musique)

Université Pierre et Marie Curie/Ircam

¹Adresse mail : hamnane@laposte.net

Remerciements

Je remercie Axel Roebel d'avoir accepté de m'encadrer sur ce stage. Je voudrais également exprimer mes remerciements à Rémi Gribonval d'avoir accepté de m'encadrer sur ce sujet, ainsi que pour ses nombreux conseils et son aide. Mes remerciements vont également à tous les membres permanents, doctorants et stagiaires des équipes Analyse-Synthèse à l'Ircam ainsi qu'à ceux de l'équipe Metiss à l'Irisa (Rennes) pour leur accueil chaleureux. En particulier, je tiens à remercier pour leur aide et leur soutien Geoffroy Peeters, Gilles Gonon et Ahmad Karfoul.

Table des matières

Résumé	7
1 Introduction générale	8
1.1 Contexte historique	8
1.2 Présentation du sujet	8
1.2.1 Propriétés acoustiques des signaux sonores et traitement du son.	8
1.2.2 Applications	9
1.2.3 Les objectifs du stage	9
1.3 Présentation du travail	9
1.3.1 Lieu et encadrement	9
1.3.2 Démarche adoptée, difficultés rencontrées	9
1.3.3 Plan du document	10
2 Représentations non adaptatives	12
2.1 Représentation de type Fourier (à court terme)	12
2.1.1 Rappel de la définition et les propriétés de la transformée de Fourier classique	12
2.1.2 Définition de la transformée de Fourier à court terme (TFCT)	12
2.1.3 Principe d'Heisenberg	13
2.1.4 Pavage temps-fréquence	15
2.1.5 Représentation graphique et ses limites	15
2.1.6 Conclusion	16
2.2 Représentation de type ondelettes	17
2.2.1 Principe des ondelettes/Interprétation	17
2.2.2 Définition et caractéristiques principales	17
2.2.3 Pavage temps-fréquence.	19
2.2.4 Conclusion sur cette représentation.	19
3 Critère de sélection	20
3.1 Introduction	20
3.2 Fonction d'ambiguïté	20
3.3 Les critères entropiques	22
3.3.1 Origine de l'entropie et explications	22
3.3.2 L'entropie de Shannon	22
3.3.3 L'entropie d'un spectre	23
3.3.4 L'entropie de Rényi et ses propriétés	24
3.3.5 Autres critères entropiques	25
3.3.6 L'entropie : un critère de parcimonie	26
3.3.7 Conclusion	27
4 Représentations adaptées et adaptatives	28
4.1 Recherche de meilleure base, 'best basis'	28
4.1.1 Principe	28
4.1.2 Algorithme plus détaillé	28
4.1.3 Conclusion et bilan	29
4.2 Les représentations adaptées en temps ou en fréquence.	30
4.2.1 La transformée de Gabor instationnaire : Généralités	30
4.2.2 La transformée covariante par translation temporelle	30
4.2.3 Conclusion et bilan	34

4.3	La représentation multicouche	35
4.3.1	Le partitionnement du plan	35
4.3.2	L'algorithme puzzle temps-fréquence	35
4.3.3	La représentation multicouche	36
4.4	Le Matching Pursuit	39
4.4.1	Matching Pursuit et représentation avec résolution adaptative	39
4.4.2	Principe du Matching Pursuit	40
4.4.3	La description de l'algorithme	40
4.4.4	Le choix du dictionnaire	40
4.4.5	Conclusion et bilan	42
5	Proposition alternative aux méthodes existantes.	43
5.1	Introduction	43
5.2	La théorie	43
5.2.1	Représentation parcimonieuse	43
5.2.2	Les approximations parcimonieuses	43
5.3	Les étapes de l'algorithme	43
5.3.1	Données d'entrée	44
5.3.2	Seuillage	44
5.3.3	La condition sur le dictionnaire	44
5.3.4	Itération et convergence	44
5.3.5	Les paramètres	45
5.3.6	Principe/Rapport avec une représentation multicouche.	45
5.4	Implémentation	45
5.4.1	Du dictionnaire à la tfct	45
5.4.2	La constante du dictionnaire	46
5.4.3	Itération et critère d'arrêt	47
5.4.4	Représentation des couches de l'algorithme	47
5.5	Evaluation	48
5.5.1	Paramètres et critères des expériences	48
5.5.2	Evaluation de la convergence	49
5.5.3	Evaluation de la parcimonie	51
5.5.4	Application à la détection de transitoire	53
5.5.5	Conclusion sur les résultats	56
5.5.6	Amélioration et recherches envisagées	56
6	Conclusion et perspectives	57
6.1	Conclusion	57
6.2	Perspectives	57

Résumé

L'objet du traitement du signal est d'analyser avec soin, de coder efficacement, de transmettre rapidement et enfin de reconstruire aussi fidèlement que possible le signal à la réception. Mais traiter un signal est avant tout une affaire de représentation puisque toute l'information contenue dans le signal est effectivement présente et cachée dans sa représentation graphique. On comprend l'importance de la représentation et la nécessité qu'elle dévoile "l'information contenue dans le signal".

La première étape de ce stage fut de faire le point sur les différentes méthodes d'adaptation d'une représentation temps-fréquence au signal. Mais qu'est-ce qu'une "bonne" représentation, une représentation "adaptée"? Le problème du critère de sélection de la taille de fenêtre est en effet un problème clef du sujet, nous lui avons donc consacré une partie importante du travail.

Par ailleurs, nous nous sommes penchés sur les avantages et inconvénients des différentes propositions dans l'état de l'art concernant les représentations temps-fréquence. Nous avons étudié comment ces représentations pouvaient être liées à un critère de sélection ainsi que d'autres propositions de représentations adaptatives.

Ensuite, à l'aide de cette étude nous avons été en mesure de sélectionner puis d'approfondir un algorithme afin d'en tirer une méthode alternative. En effet, celui-ci permet de choisir pour chaque lieu du plan temps-fréquence une taille de fenêtre parmi deux. Enfin, nous avons fait une évaluation de cette proposition en étudiant la convergence de l'algorithme et la parcimonie des représentations obtenues. Notons que si le sujet a été traité de telle sorte qu'il puisse avoir un intérêt pour tout type d'applications, il fut intéressant de se restreindre à la détection de transitoire pour évaluer le travail. Ceci nous permettra de conclure et d'ouvrir sur certaines perspectives de recherche pour la fin de ce stage.

1 Introduction générale

1.1 Contexte historique

La musique définie simplement comme l'art d'organiser les sons, est un objet d'étude pour les sciences, en particulier pour les mathématiques et le traitement du signal, dans la mesure où ceux-ci permettent de comprendre la diversité et la complexité des sons musicaux. Il est alors important de connaître et de comprendre les différents outils qui permettent d'organiser, d'analyser, de synthétiser et de modifier les sons musicaux ou tout simplement de les comprendre intimement grâce à leurs représentations graphiques. Nous nous plaçons donc ici dans le domaine du traitement du signal, dans laquelle il s'agit, entre autre, d'extraire d'un signal les informations pertinentes. Les applications industrielles de l'analyse et du traitement des signaux sont nombreuses et variées allant de la géophysique à l'imagerie médicale. Mais il s'agit ici d'aborder le sujet du point de vue de l'analyse sonore.

Faisons un bref parcours historique de ces outils depuis le 18ème siècle jusqu'à nos jours. Pendant longtemps les fonctions de base de l'analyse ont été le cosinus, le sinus et l'exponentielle imaginaire. Ainsi les fonctions $\frac{1}{\sqrt{2\pi}}e^{ikx}$ constituent une base orthonormée de l'espace de référence $L^2(0, 2\pi)$ et les séries de Fourier sont les combinaisons linéaires $\sum a_k e^{ikx}$. Mais l'analyse de Fourier présente des inconvénients majeurs qui ne permettent pas une analyse satisfaisante de toute sorte de signaux. En effet, on souhaiterait que les paramètres réalisent une analyse à la fois en temps et en fréquence, exactement comme à la manière d'une portée musicale où sont indiquées à la fois la fréquence et la durée des notes. C'est pourquoi l'analogie avec la notation musicale est communément utilisée pour introduire l'idée de représentation temps-fréquence, chaque note d'un morceau étant en effet associée à une certaine localisation aussi bien temporelle (son instant d'occurrence et sa durée) que fréquentielle (sa hauteur). L'analyse de Fourier classique est la plus ancienne de ces techniques, elle est performante en particulier grâce à la transformée de Fourier rapide, ceci ne permettant pas de donner commodément sens à la notion, pourtant intuitive, de fréquence instantanée. Cette difficulté, nous le verrons, est liée de façon fondamentale au fait que les variables de temps et de fréquence sont canoniquement conjuguées et leur représentation associée, sujette au principe d'Heisenberg.

Par la suite, une approche intuitive est utilisée sur la base de transformée de Fourier glissante ou à fenêtre. Apparaît ainsi la transformée de Gabor, le physicien Denis Gabor l'a proposée en 1940 (et on lui doit la formule de reconstruction.) Mais cette transformée présente encore des inconvénients et il sera nécessaire d'affiner à nouveau cet outil. Au début des années 80, de nombreux scientifiques utilisaient déjà les ondelettes comme une alternative à l'analyse de Fourier traditionnelle. Cet avancée scientifique apporte beaucoup au traitement du signal en général et aux mathématiques. S'il est évident qu'elles permettent un nouveau point de vue, elles ne résolvent pas tous les problèmes. Mais ceci encourage, tout récemment, de nombreux scientifiques à se porter sur l'adaptativité de l'analyse et des représentations temps-fréquence pour différents traitements.

1.2 Présentation du sujet

1.2.1 Propriétés acoustiques des signaux sonores et traitement du son.

La classe des signaux audio allant des signaux quasi-stationnaires aux très courtes transitoires, des notes de violon aux percussions, comporte une très large variété de structure temporelles et

fréquentielles. La diversité des timbres résulte d'une très grande variété de structure du signal en temps et en fréquence.

En particulier, l'attaque d'une note est une phase siège de très nombreuses fréquences. Bien qu'étant très brève, cette phase est caractéristique de l'instrument et de l'interprétation. Elle est donc tout aussi importante que la phase de maintien de la note qui contient en général des fréquences moins nombreuses et peut l'être pour certaines applications. De la même façon, les composantes provoquées par des instruments percussifs sont tout aussi importantes que celles provoquées par des instruments plus harmoniques comme le violon.

Si les traiteurs de signaux sonores savent depuis longtemps faire une analyse et une représentation temps-fréquence des sons, ils ont toujours été confrontés au problème de compromis entre l'adaptation aux différentes composantes du signal. En effet, les signaux sonores peuvent comporter des composantes aux caractéristiques temps-fréquence très différentes et qui nécessiteraient chacune une taille de fenêtre différente.

En conséquent, toutes ces propriétés physiques justifient la nécessité d'une représentation temps-fréquence avec une résolution adaptative.

1.2.2 Applications

De nombreux traitements du son pourraient gagner à utiliser une analyse/représentation adaptative. On peut penser par exemple à des applications comme le vocoder de phase, à la séparation de source, au codage, ou à la détection de transitoires ou de fondamentales. En effet, tous ces traitements temps-fréquence utilisent une représentation temps-fréquence à un moment donné. Par exemple, la référence [15] démontre l'utilité d'une représentation adaptative pour le codage.

Elle permettrait une représentation "plus souple" que celles proposées par les outils traditionnels. On peut d'ores et déjà remarquer que se posera le problème de la reconstruction du signal original, opération indispensable dans de nombreux traitements et ainsi comprendre la nécessité de développer ou d'utiliser des outils théoriques permettant une reconstruction relative.

1.2.3 Les objectifs du stage

On peut distinguer deux objectifs. D'une part de faire le bilan sur les différentes méthodes qui permettent d'adapter la taille de la fenêtre aux caractéristiques locales du signal, de confronter les différents critères d'optimalité ou de sélection de la taille de fenêtre, critères qui permettent de définir ce qu'est une "bonne représentation".

D'autre part, il s'agit de faire une proposition permettant une amélioration des méthodes présentes actuellement dans l'état de l'art.

1.3 Présentation du travail

1.3.1 Lieu et encadrement

Le sujet, proposé par Axel Roebel a été traité à l'Ircam dans l'équipe Analyse-Synthèse sous l'encadrement d'Axel Roebel du 20 Mars au 20 Avril, puis du 02 Mai au 28 Juin à l'Irisa dans l'équipe Metiss à Rennes sous l'encadrement de Rémi Gribonval.

1.3.2 Démarche adoptée, difficultés rencontrées

Le travail présenté ici a, en premier lieu, fait l'objet d'une étude théorique. En effet, les différents outils utilisés en traitement du signal pour traiter des représentations temps-fréquence font appel à des notions théoriques, en particulier mathématiques. Comprendre toutes ces notions

a donc été essentiel pour pouvoir juger des qualités des différentes approches et algorithmes proposés dans la littérature, ainsi que pour être en mesure de proposer de nouveaux outils. Ainsi en vue d'une évaluation des propositions de l'état de l'art et d'un apport à celui-ci, on a accordé une large partie à la recherche bibliographique, à la synthèse et à l'analyse de ces informations.

Nous avons donc dans un premier temps effectué une étude des différents critères de sélection de la représentation à savoir ceux donnés par les critères entropiques, les critères de parcimonie, la fonction d'ambiguïté, les fonctions de coût.

Par la suite, nous avons étudié différents algorithmes à relier à ces critères de sélection, et approfondi des outils de représentation pouvant être reliés à ces critères. Ceci a permis de définir des lieux d'études et de sélectionner les algorithmes pertinents d'un point de vue scientifique pour être décrit dans la partie théorique.

Une autre difficulté rencontrée pendant le stage a été de définir un cadre, une application, permettant de définir des critères objectifs d'évaluation des différents algorithmes aboutissant à une représentation temps-fréquence.

1.3.3 Plan du document

Comme on pourra le voir au long de ce rapport, la principale base d'étude de ce travail a été les documents bibliographiques réunis autour du sujet de représentation avec une résolution adaptative. Ainsi *l'effort majeur se situe au niveau de la synthèse dans un premier temps puis de l'analyse dans un second* de ces informations vis à vis du sujet traité. Une attention particulière a pu être portée sur les particularités, avantages et inconvénients ainsi qu'aux confrontations des méthodes ou algorithmes aussi bien du point de vu pratique que théorique.

La section 1 est l'introduction générale.Elle présente le sujet et le travail effectué pendant le stage.

La section 2 traite des représentations non adaptatives. Ce sont ces outils qui sont la base de construction des représentations adaptatives. Il est donc essentiel de les assimiler pour comprendre ceux développés dans la suite. Ceci justifiera le sujet tel qu'il est posé et permettra d'en percevoir les difficultés. Il s'agira plus précisément de traiter des représentations de type Fourier ou de type ondelettes. Les représentations de type Fourier sont depuis toujours une représentation incontournable. Les représentations de type ondelettes tiennent ici une importance non négligeable dans le sens où elles constituent un tournant fondamental dans le problème de représentation temps-fréquence.

La section 3 dresse un bilan des critères de sélection présents dans la littérature. En effet, pour traiter des représentations adaptatives, il est nécessaire de définir un critère permettant d'adapter la taille de la fenêtre aux caractéristiques locales du signal.

La section 4 détaille les propositions existantes pour adapter les outils d'analyse et de traitement des signaux en les étayant ou en les augmentant quand cela fut nécessaire. On y distinguera les représentations adaptées des représentations adaptatives. On entend par représentations adaptées celles dont le pavage temps-fréquence est "flexible" mais dont la taille des fenêtres ou

des atomes n'est pas relié à un critère qui permet de donner une représentation adaptée à chaque signal, et par représentation adaptative celles qui s'adaptent automatiquement aux caractéristiques du signal. On pourra alors étudier ces représentations, les avantages et inconvénients de chacune d'entre elles.

La section 5 présente un algorithme qui améliore les propositions existantes. On justifiera le choix de cette méthode au vue des propositions déjà existantes présentées plus haut. Ainsi, on détaillera la théorie nécessaire à la compréhension de l'algorithme, les étapes importantes de celui-ci ainsi que la façon dont nous avons procédé pour mettre en oeuvre son implémentation. Enfin on tentera de l'évaluer selon des critères adaptés que nous justifierons. Ceci afin d'en comprendre les limites et les améliorations à envisager.

Enfin nous serons en mesure de conclure et de réfléchir aux perspectives offertes par ce travail dans la section 6.

2 Représentations non adaptatives

On regroupe sous l'appellation "non adaptatif" la transformée de Fourier à court terme qu'on notera TFCT et la représentation en ondelettes. En effet, ces outils de représentation temps-fréquence donnent des pavages temps-fréquence fixés et non adaptés à un signal ni même à la classe de signaux.

Pour plus d'informations sur ce qui est présenté dans cette section, on pourra se référer à [14].

2.1 Représentation de type Fourier (à court terme)

On présente ici les outils de représentation de type Fourier. A savoir les définitions et propriétés mathématiques qui présentent un intérêt en traitement du signal.

On revient sur la transformée de Fourier classique pour aborder ensuite la transformée de Fourier à court terme dénommée couramment TFCT. Celle-ci donne une représentation sous forme de carte d'énergie temps-fréquence. On se penchera en particulier sur les limites de cette représentation (principe d'Heisenberg) qui justifient en partie le sujet traité.

2.1.1 Rappel de la définition et les propriétés de la transformée de Fourier classique

On rappelle brièvement ici les quelques définitions et propriétés de la transformée de Fourier classique.

On part d'un signal $x(t)$ défini sur toute la droite réelle, intégrable au sens de Lebesgue sur \mathbb{R} , et on pose

$$\hat{x}(\xi) = \int_{\mathbb{R}} x(t) e^{-2i\pi t \xi} dt$$

pour $\xi \in \mathbb{R}$. Cette fonction \hat{x} est bien définie sur \mathbb{R} et on l'appelle transformée de Fourier de x . (Une de ses propriétés basiques est qu'elle tend vers 0 à l'infini et est continue sur \mathbb{R} .)

Une autre propriété utile est la formule d'inversion de Fourier. Supposons que x et \hat{x} soient toutes les deux dans $L^1(\mathbb{R})$. Alors on a, pour presque tout $t \in \mathbb{R}$

$$x(t) = \int_{\mathbb{R}} \hat{x}(\xi) e^{i2\pi \xi t} d\xi.$$

On a par ailleurs l'importante conservation de l'énergie :

$$\int_{\mathbb{R}} |x(t)|^2 dt = \int_{\mathbb{R}} |\hat{x}(\xi)|^2 d\xi,$$

ce qui signifie que l'opérateur $\mathcal{F} : x \rightarrow \hat{x}$ de $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ dans $L^2(\mathbb{R})$ est uniformément continu et donc se prolonge de manière unique, grâce à la densité de $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ dans $L^2(\mathbb{R})$.

2.1.2 Définition de la transformée de Fourier à court terme (TFCT)

La transformée de Fourier à court terme d'une fonction $x(t) \in L^2(\mathbb{R})$ (ou plus généralement d'une distribution) est définie à l'aide d'une fenêtre w .

Par la suite, on est amené à faire glisser cette fenêtre (centrée en 0 et d'énergie 1) utilisée pour localiser en temps l'analyse :

(On notera bien que la fenêtre que l'on fait glisser est toujours la même et a donc la même résolution en temps et en fréquence sur tout le signal.)

On obtient alors une famille de coefficients $W_x(\lambda, b)$ (remplaçant les valeurs de $\hat{x}(\lambda)$ pour la transformée de Fourier classique) et où λ représente la fréquence et b localise l'analyse

$$W_x(\lambda, b) = \int_{-\infty}^{+\infty} x(t) \bar{\omega}(t-b) e^{-2i\pi\lambda t} dt.$$

L'application $x \rightarrow W_x$ s'appelle la transformée de Fourier à fenêtre glissante ou encore la transformée de Gabor continue. On note alors $\omega_{\lambda,b}(t) = \omega(t-b)e^{-2i\pi\lambda t}$ et on peut l'écrire sous la forme d'un produit scalaire dans L^2

$$W_x(\lambda, b) = (x, \omega_{\lambda,b}).$$

Comme pour l'analyse de Fourier, la connaissance des $W_x(\lambda, b)$ pour toutes les valeurs réelles de λ et de b déterminera complètement et de manière unique le signal x .

Remarque 2.1 *On notera bien que la fenêtre que l'on fait glisser est toujours la même et a donc la même résolution en temps et en fréquence sur tout le signal.*

On a, grâce à Gabor, la formule inverse donnant x à partir des $W_x(\lambda, b)$ dite formule de reconstruction ou de synthèse et la conservation de l'énergie.

Théorème 2.2 *Soit $\omega \in L^1 \cap L^2$ centrée en 0, telle que $|\hat{\omega}|$ soit une fonction paire et $\|\omega\|_2 = 1$. C'est ce qu'on appelle la fenêtre. On a posé $\omega_{\lambda,b}(t) = \omega(t-b)e^{-2i\pi\lambda t}$, λ et b étant des réels, et on considère pour tout signal $x \in L^2$ les*

$$W_x(\lambda, b) = \int_{-\infty}^{+\infty} x(t) \bar{\omega}(t-b) e^{-2i\pi\lambda t} dt.$$

On a la conservation de l'énergie

$$\int \int_{\mathbb{R}^2} |W_x(\lambda, b)|^2 d\lambda db = \int_{-\infty}^{+\infty} |x(t)|^2 dt,$$

et la formule de reconstruction

$$x(t) = \int \int_{\mathbb{R}^2} W_x(\lambda, b) \omega_{\lambda,b}(t) d\lambda db,$$

au sens où si $g_A(t) = \int_{b \in \mathbb{R}} \int_{|\lambda| \leq A} W_x(\lambda, b) \omega_{\lambda,b}(t) d\lambda db$ alors $g_A \rightarrow x$ dans L^2 .

2.1.3 Principe d'Heisenberg

On rappelle ici le principe d'Heisenberg. Il s'agit d'une propriété déterminante quand on veut traiter de résolution à la fois en temps et en fréquence.

Par souci de clarté pour la suite, on introduit dès à présent la notion d'atome temps-fréquence :

Définition 2.3 *On appelle atome un élément de la famille $(g_{t,f})_{(t,f)}$ où*

$$g_{t,f}(t') = g(t' - t) e^{i2\pi f t'}.$$

Avec (t,f) respectivement les paramètres de temps et de fréquence

On notera dans la suite $t_{g_{t,f}}$ et $f_{g_{t,f}}$ les centres de gravité temporel et fréquentiel de l'atome $g_{t,f}$, à savoir

$$t_{g_{t,f}} = \int_{\mathbb{R}^2} t' |g_{t,f}(t')|^2 dt'$$

$$f_{g_{t,f}} = \int_{\mathbb{R}^2} f' |\hat{g}_{t,f}(f')|^2 df'$$

Définition 2.4 On appelle dispersion d'énergie de $g_{t,f}$ en temps

$$\sigma_{g_{t,f}}^2 = \int_{\mathbb{R}^2} (t' - t_{g_{t,f}})^2 |g_{t,f}(t')|^2 dt'.$$

On appelle dispersion d'énergie de $g_{t,f}$ en fréquence

$$\sigma_{\hat{g}_{t,f}}^2 = \int_{\mathbb{R}^2} (f' - f_{g_{t,f}})^2 |\hat{g}_{t,f}(f')|^2 df'$$

On pose $E_{g_{t,f}} = \int_{\mathbb{R}} |g_{t,f}(t')|^2 dt$ l'énergie de f ; on introduit aussi

$$(\Delta t)^2 = \frac{(\sigma_{g_{t,f}})^2}{E_{g_{t,f}}},$$

et

$$(\Delta \lambda)^2 = \frac{(\sigma_{\hat{g}_{t,f}})^2}{E_{g_{t,f}}},$$

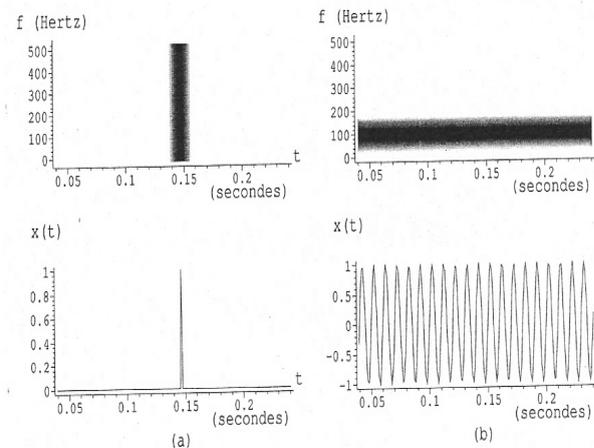
Le principe d'incertitude d'Heisenberg précise le lien entre la localisation d'un atome et celle de sa transformée, autrement dit entre les dispersions $\sigma_{g_{t,f}}$ et $\sigma_{\hat{g}_{t,f}}$, c'est-à-dire :

$$\sigma_{g_{t,f}} \times \sigma_{\hat{g}_{t,f}} \geq \frac{E_{g_{t,f}}}{4\pi} \implies \Delta t \times \Delta \lambda \geq \frac{1}{4\pi}.$$

Ceci pour un atome $g : \mathbb{R} \rightarrow \mathbb{C}$ de classe C^1 telle que g , g' et tg sont dans $L^2(\mathbb{R})$.

Remarquons qu'on obtient l'égalité pour g fonction gaussienne normalisée dans $L^2(\mathbb{R})$. Ce qui signifie que c'est la fonction qui réalise le meilleur compromis (mais pas forcément le plus adapté) à la localisation en temps et en fréquence de l'atome.

Ceci se comprend dans le sens où plus le support de g est "petit", plus le support de \hat{g} est "grand" et réciproquement.



Spectrogramme (en haut) et représentation temporelle (en bas) d'un Dirac en $t_0 = 147ms$ (a) et d'un sinusoïde de fréquence $f_0 = 100Hz$ (b). L'énergie, en décibels, est codée en niveaux de gris, du blanc (les -20dB) au noir (0dB). La fenêtre d'analyse est Gaussienne $g_a(t)$ avec $a = 31ms$.

Cette figure est extraite du livre [26] dans la partie écrite par Rémi Gribonval.

• **Conséquences du principe d'incertitude pour les représentations temps-fréquence :**

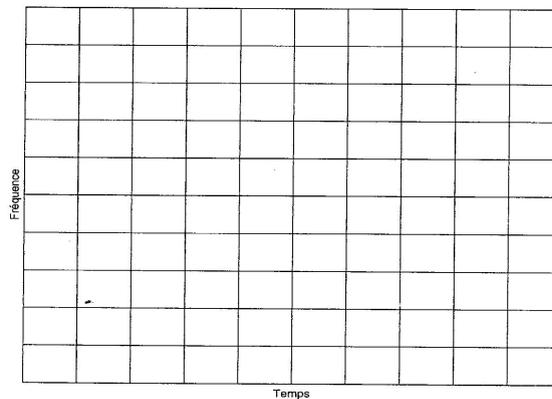
1-Le principe d'incertitude d'Heisenberg impose une valeur minimale pour la surface du "rectangle" représentant la localisation temps-fréquence d'un atome. En effet, on peut considérer que les valeurs de $\Delta\lambda$ et de Δt définies plus haut définissent des rectangles ;

2-Le choix d'une taille de fenêtre unique pour représenter un signal par un spectrogramme impose par conséquent un compromis parfois difficile entre la précision de la localisation temporelle et celle de la localisation fréquentielle.

2.1.4 Pavage temps-fréquence

On représente ci-dessous un pavage temps-fréquence pour une TFCT avec une taille de fenêtre constante

Figure : Pavage temps-fréquence d'une TFCT.



2.1.5 Représentation graphique et ses limites

Figure : Son de glockenspiel. Durée de 6 secondes. Fréquence d'échantillonnage de 44100 Hz.

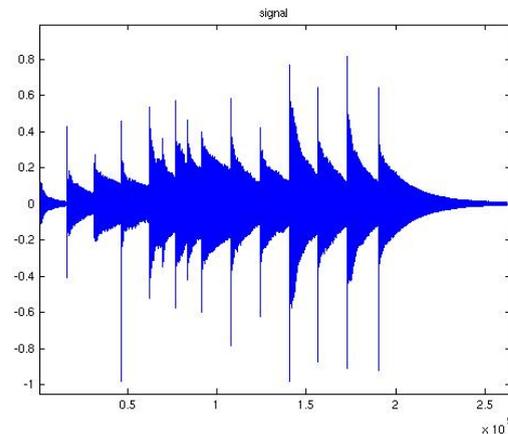
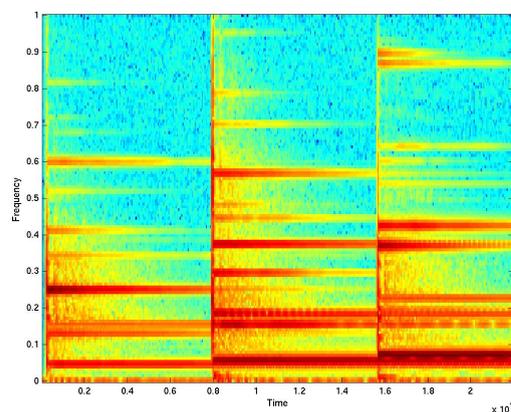


Figure : Exemple de représentation temps-fréquence. Le spectrogramme d'un son de glockenspiel.



On peut remarquer que, si on peut visualiser les harmoniques, leurs fréquences n'apparaît pas de façon suffisamment précise. D'autre part, l'énergie présente normalement au niveau de l'attaque se retrouve mal localisée en temps. Cette représentation est cependant un bon compromis entre les phases d'attaque et de maintien de la note. Si on voulait avoir plus de précisions en temps, on perdrait dans le même temps la précision en fréquence et vice-versa.

2.1.6 Conclusion

On peut noter un inconvénient majeur de la TFCT. En effet, la fenêtre est de longueur fixe ce qui est un handicap important lorsqu'on veut traiter des signaux dont les variations peuvent avoir des ordres de grandeur très variable comme c'est le cas pour les signaux musicaux.

Par ailleurs, on remarque qu'il pourrait être intéressant de faire varier la taille de la fenêtre non seulement par translation mais également par dilatation et contraction en fonction des caractéristiques du signal et de ce que l'on veut mettre en évidence. C'est précisément ce qui nous amène à la représentation de type ondelettes.

2.2 Représentation de type ondelettes

La représentation temps-fréquence grâce à l'outil d'analyse des ondelettes n'est pas, certes, une solution au problème de l'adaptativité mais a constitué surtout un tournant important pour le problème de représentation temps-fréquence, d'adéquation de la taille et de la localisation des atomes aux propriétés locales des signaux à traiter. Pour cette raison, on mentionne rapidement ici définitions et propriétés permettant de comprendre l'apport de cette théorie aux représentations temps-fréquence.

2.2.1 Principe des ondelettes/Interprétation

Nous avons vu précédemment que la TFCT présente un inconvénient majeur. Elle est de résolution temporelle et fréquentielle fixe.

La transformée en ondelettes va nous permettre de pallier cet inconvénient. On peut distinguer deux intérêts principaux des ondelettes en tant qu'outil mathématique d'analyse et de représentation des signaux.

Le premier est, comme dit précédemment, centré autour de l'idée de construire un outil d'analyse local en temps. Nous introduirons alors la transformée continue en ondelettes et nous verrons comment elle résout les problèmes posés par les outils classiques. La transformée en ondelettes est non seulement une analyse locale mais, de plus, sa résolution temporelle est variable. Elle permet donc de décrire le comportement local des signaux à différentes *échelles* de temps.

Le second est celui du problème de reconstruction et de la recherche de représentations pratiques. En effet, cette transformée nous permet bien de reconstruire le signal de départ et présente des avantages d'un point de vue numérique.

2.2.2 Définition et caractéristiques principales

La transformation en ondelettes continue consiste à créer, à partir d'une ondelette mère notée ψ (qui ressemble à une petite onde) ou encore appelée ondelette analysante, une famille d'ondelettes

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$$

où $a \in \mathbb{R}^{+*}$ est l'échelle et $b \in \mathbb{R}$ est la position (a sert à dilater et b à traduire). On dit également que tout couple $(a, b) \in \mathbb{R}^{+*} \times \mathbb{R}$ définit un atome de la transformée.

• Caractéristiques de l'ondelette mère et coefficients du signal

Les caractéristiques de l'ondelette mère sont nettement différentes de celles d'une fenêtre. En effet, une fenêtre avait plutôt l'allure d'un créneau, alors que la définition de ψ lui imposera d'être d'intégrale nulle. Par ailleurs, on imposera à ψ une bonne localisation autour de 0 et pour $\hat{\psi}$ une localisation autour d'une certaine fréquence $\omega_0 > 0$. On obtient ainsi des fonctions à décroissance rapide : elles oscillent puis s'amortissent ; ce qui leur donne leur nom.

L'analyse par ondelettes associe à la fonction analysée les coefficients

$$c_x(a, b) = (x, \Psi_{ab}) = \int_{-\infty}^{+\infty} x(t) \bar{\Psi}_{ab}(t) dt$$

qui comme les coefficients de Gabor décrivent le contenu de la fonction au voisinage de $(b, \frac{\omega_0}{a})$ si $\hat{\Psi}$ est localisée autour de ω_0 .

• **Définition d'une ondelette mère et d'une ondelette de base.**

Définition 2.5 Soit $\psi \in L^1 \cap L^2$ telle que

$$(i) \quad \int_{-\infty}^{+\infty} \frac{|\hat{\Psi}(\lambda)|}{|\lambda|} d\lambda < +\infty$$

et

$$(ii) \quad \|\psi\|_2 = 1.$$

On dit alors que ψ est une ondelette mère.

Remarque 2.6 1- La condition $\|\psi\|_2 = 1$ signifie que ψ a une énergie finie. Elle est de plus normalisée.

2- L'hypothèse (i) implique que $\hat{\Psi}(0) = \int_{-\infty}^{+\infty} \psi(t) dt = 0$ car $\hat{\Psi}$ est continue. Dans les cas pratiques, cette condition est suffisante pour obtenir les résultats ci-dessus. En effet, si on suppose que ψ et $x\psi$ sont intégrables, on aura $\hat{\Psi} \in C^1$ et $\lambda \rightarrow \frac{|\hat{\Psi}(\lambda)|}{|\lambda|}$ continue en 0. On aura donc la convergence en l'infini puisque $\psi \in L^2$.

3- Dans le cas du traitement du signal musical, il est commode d'extraire de l'analyse des informations à propos de la distribution d'énergie et du comportement en phase dans la représentation de la transformée en ondelettes. Ceci peut être fait facilement en utilisant une ondelette à valeur complexe. Il est donc nécessaire de faire l'hypothèse supplémentaire suivante :

(iii) $g(t)$ ne contient que des composantes à fréquence positive ($\omega < 0 \implies \hat{g}(\omega) = 0$).

Définition 2.7 On appelle ondelette de base

$$\Psi_{ab} = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right)$$

où a et b sont dans \mathbb{R}^* et \mathbb{R} , et on considère pour tout signal $x \in L^2$ les coefficients d'ondelettes

$$c_x(a, b) = \int_{-\infty}^{+\infty} x(t) \bar{\Psi}_{ab}(t) dt.$$

Remarque 2.8 1- On peut donner pour les coefficients c_x une définition équivalente

$$c_x(a, b) = \sqrt{a} \int_{-\infty}^{+\infty} \hat{x}(t) \bar{\hat{\Psi}}(at) e^{2i\pi tb} dt$$

ou

$$c_x(a, b) = (x * \check{\Psi}_a)(b)$$

où $\check{\Psi}_a(t) = \frac{1}{\sqrt{a}} \Psi(-\frac{t}{a})$.

2- Vu du côté temporel (ou spatial) t , $c_x(a, b)$ renseigne sur le signal x autour du point b dans un voisinage de taille a . Vu du côté fréquentiel ν , $c_x(a, b)$ renseigne sur le signal \hat{x} autour de la fréquence $\frac{1}{a}$.

• **Théorème fondamental**

On pose $K = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\lambda)|}{|\lambda|} d\lambda < +\infty$, nous avons le théorème suivant :

Théorème 2.9 *On a avec les notations précédentes :*

a) *La conservation de l'énergie*

$$\frac{1}{K} \int_{\mathbb{R}^2} |c_x(a, b)|^2 \frac{da db}{2} = \int_{-\infty}^{+\infty} |x(t)|^2 dt.$$

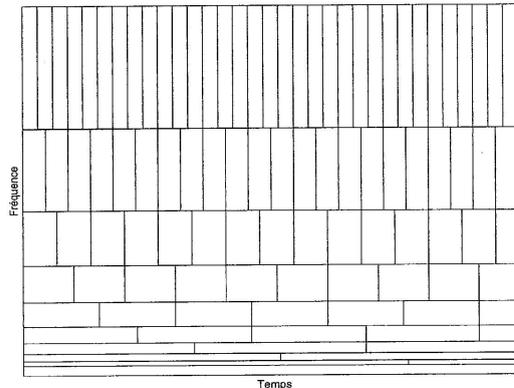
b) *La formule de reconstruction*

$$x(t) = \frac{1}{K} \int_{\mathbb{R}^2} c_x(a, b) \psi_a b(t) \frac{da db}{a^2}$$

au sens où si $x_\varepsilon(t) = \frac{1}{K} \int_{|a| \geq \varepsilon, b \in \mathbb{R}} c_x(a, b) \psi_a b(t) \frac{da db}{a^2}$ alors $x_\varepsilon \rightarrow x$ dans L^2 lorsque ε tend vers 0.

2.2.3 Pavage temps-fréquence.

On représente ici le pavage temps-fréquence donné par la représentation de type ondelettes.



Remarque 2.10 *On aura pu remarquer sur la définition que la transformée en ondelettes est continue est infiniment redondante. Cette représentation du pavage temps-fréquence sous-entend une discrétisation des atomes sur le plan temps-fréquence.*

2.2.4 Conclusion sur cette représentation.

Cet outil de représentation a clairement l'avantage sur la TFCT de posséder des atomes de différentes tailles dont la résolution en temps et en fréquence varie selon la localisation de l'atome dans le plan temps-fréquence. Cependant, comme nous avons pu le voir précédemment sur la représentation du pavage temps-fréquence, la localisation des différents atomes selon leur taille est fixée à l'avance. Cet outil malgré les différentes résolutions qu'il autorise, ne permet pas d'adapter cette résolution à un signal donné.

Il va donc s'agir à présent de comprendre les différents problèmes posés par le concept "d'adaptativité" de la représentation à un signal donné, les différents critères permettant de définir cette adaptativité, ainsi que les outils de représentation permettant de les mettre en oeuvre.

3 Critère de sélection

3.1 Introduction

Il s'agira dans cette partie de traiter des différents critères de sélection ou d'optimalité de la représentation des "bonnes" tailles de fenêtres proposés dans la littérature. En effet, comme nous l'avons déjà fait remarqué, la possibilité d'une représentation temps-fréquence avec résolution adaptative suppose qu'il existe une "bonne représentation", une représentation "adaptée" à un signal que ce soit de façon locale ou globale. Cela implique non seulement de savoir ce qu'est une représentation adaptée mais en plus, de savoir comment la sélectionner de manière systématique grâce à ce qu'on appellera *un critère de sélection*.

Plus précisément, à quoi sert un critère de sélection ? Ou plutôt, comment s'en sert-on ? On peut distinguer plusieurs usages possibles :

Premièrement, on peut penser à une optimisation de la représentation globale. Par exemple : une taille de fenêtre de 256 ou de 1024 échantillons est-elle plus adaptée pour ce signal ? Mais ceci n'est pas satisfaisant étant donné les différences entre les caractéristiques temps-fréquence des composantes d'un signal. Par exemple les phases d'attaques d'un note produite par un instrument de musique requiert une petite fenêtre alors que la phase de maintien aurait tendance à encourager plutôt le choix d'une grande fenêtre.

Deuxièmement, on peut penser à une optimisation locale avec intervention d'un utilisateur : on cherche quelle taille de fenêtre est le mieux adaptée à cette composante du plan temps-fréquence qu'on sélectionne. Mais cette approche ne nous intéresse pas puisqu'il s'agit dans ce stage d'établir une adaptation *automatique* de la taille de la fenêtre.

Troisièmement, il est possible d'établir une optimisation locale sans intervention de l'utilisateur. On souhaite donc établir un algorithme qui trouve pour chaque partie du plan temps-fréquence quelle est la taille de fenêtre, en temps et en fréquence, la mieux adaptée et ceci automatiquement.

Par ailleurs, nous cherchons un critère de sélection qui permette de faire un calcul dans un délai raisonnable. Une attention particulière sera donc apportée au temps de calcul qu'imposeront les différents critères. Enfin, on peut soulever dès à présent le problème du lien avec la représentation temps-fréquence. En effet, il n'est pas immédiat d'associer un critère à une représentation temps-fréquence. Si on trouve bien le critère qui sélectionne, il faut encore savoir ce qu'il sélectionne : des bases différentes, une taille de fenêtre parmi deux ou plus, des atomes de tailles ou de formes différentes. Mais nous verrons ces différentes possibilités dans la section 4 qui traite explicitement des représentations adaptées et adaptatives.

Nous essaierons donc dans cette section de comprendre les avantages, inconvénients et différences entre les différents critères utilisés dans l'état de l'art pour traiter le problème.

3.2 Fonction d'ambiguïté

On note ici une approche proposée dans [33] et dans [22]

Définition 3.1 On définit la fonction d'ambiguïté d'un signal x par

$$A_x(\tau, \omega) = \int_{\mathbb{R}} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi\omega t} dt$$

Remarquons que la fonction d'ambiguïté peut être facilement comprise comme une généralisation de la fonction de corrélation au temps et à la fréquence.

Définition 3.2 On définit la fenêtre optimale par $g_{opt} = \arg \max_{\{g \in L^2(\mathbb{R}) / \|g\|=1\}} (|A_x|^2, |A_g|^2)$,

La remarque précédente permet donc de comprendre qu'il s'agit tout simplement de définir la fenêtre optimale comme étant celle dont la corrélation en temps et en fréquence de la fenêtre d'analyse avec le signal est la plus importante.

Par ailleurs, on peut remarquer comme dans [23] que cette approche imposera un coût de calcul assez élevé, c'est en partie pour cette raison que nous ne retiendrons pas cette approche pour traiter le sujet.

• **Remarques importantes :**

1-Propriété globale du signal :

Il est assez clair que la fonction d'ambiguïté ne donne qu'une propriété *globale* du signal. Par conséquent, une question se pose. Comment appliquer localement ce critère ?

2-Vers une optimisation locale et supervisée.[23]

Afin de résoudre le problème du point précédent, on peut penser à une optimisation locale mais supervisée. C'est-à-dire nécessitant l'intervention de l'utilisateur. Plus précisément, on réalise une première analyse avec un outil comme la TFCT par exemple, l'utilisateur sélectionne une zone du plan temps-fréquence auquel il accorde un intérêt particulier et pour laquelle il désire avoir une représentation temps-fréquence adaptée. C'est à ce moment qu'intervient le critère de sélection de la "bonne" taille de fenêtre avec g_{opt} . On pourra ensuite refaire une analyse du signal avec la taille de fenêtre choisie

3-En conclusion, la procédure faisant intervenir un utilisateur n'est pas satisfaisante car nous voudrions que l'adaptation se fasse de manière automatique d'une part, et d'autre part qu'elle soit adaptée à toutes les composantes présentes dans le signal comme nous l'avons expliqué dans la section 3.1.

Remarque 3.3 Pour un signal aléatoire, cette définition peut être adaptée en prenant $\mathbb{E}\{A_x\}$ au lieu de A_x .

3.3 Les critères entropiques

De nombreuses informations dans cette partie ont été tirées de [15]

3.3.1 Origine de l'entropie et explications

L'entropie est une notion qui trouve son origine dans la thermodynamique. Elle a aujourd'hui des applications dans des domaines très différents comme l'économie ou la théorie de l'information. Mais c'est son utilisation dans la théorie de l'information à laquelle nous nous intéresserons.

D'une manière générale, c'est une mesure quantitative du désordre, et dans notre cadre, elle définit une mesure de "quantité d'information" contenue dans le signal. Dans le cas de la représentation du signal, il s'agit de mesurer la finesse et la précision de la localisation de l'information, localisation à la fois en temps et en fréquence.

L'entropie nous permettra donc de voir où l'énergie du signal se localise dans le plan temps-fréquence. Plus précisément, elle nous permettra de savoir quelle représentation temps-fréquence permet la meilleure localisation de l'information dans le plan.

Nous rappellerons ici les définitions de l'entropie dans le cadre de la théorie de l'information et donnerons quelques unes de ses propriétés avant de détailler son utilisation dans l'adaptation des représentations temps-fréquence au signal étudié. Nous verrons ensuite en quoi l'entropie peut être considérée comme un critère de parcimonie et comment on pourra utiliser cette remarque importante pour traiter le sujet.

3.3.2 L'entropie de Shannon

Shannon définit dans les années 40 l'entropie :

Définition 3.4 *L'entropie de Shannon d'une variable aléatoire discrète \mathbf{x} prenant les valeurs $(x_k)_{k \in [1, N]}$ et de probabilités respectives $(p_k)_{k \in [1, N]}$ (où évidemment $\sum_{k=1}^{k=N} p_k = 1$) est notée $H(\mathbf{x})$ et est définie par*

$$H(\mathbf{x}) = \sum_{k=1}^{k=N} -p_k \log p_k$$

Remarque 3.5 *$H(\mathbf{x})$ est donc définie comme espérance mathématique, ou valeur attendue, de la quantité d'information. La quantité d'information étant définie par $\log p_k$*

Cette définition de l'entropie comme mesure de la quantité d'information produite par un processus discret aléatoire vérifie les trois propriétés suivantes :

Proposition 3.6 *(On écrira ici $H(\mathbf{x})$ comme $H_n(p_1, \dots, p_n)$)*

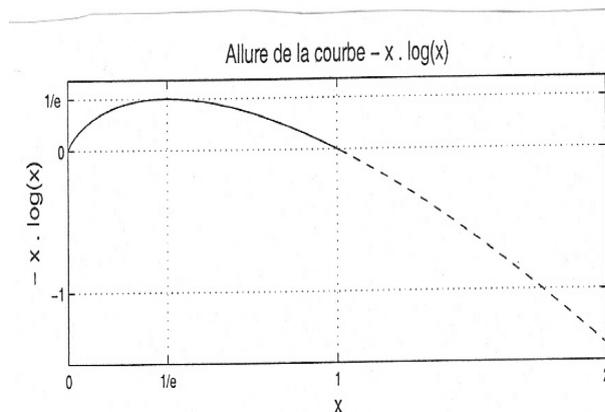
1. *L'entropie est une fonction continue des probabilités de chaque occurrence. C'est-à-dire que pour tout n , $H_n(p_1, \dots, p_n)$ est une fonction continue de chaque occurrence.*

2. *L'entropie d'une variable aléatoire uniforme augmente avec le nombre des occurrences. C'est-à-dire que pour tout n , $H_n(p_1, \dots, p_n) < H_{n+1}(p_1, \dots, p_n, p_{n+1})$.*

3. L'entropie d'une variable aléatoire est égale à la somme pondérée des entropies des sous-choix possibles le cas échéant. L'entropie doit donc être une somme pondérée des probabilités de ses occurrences. En particulier, une probabilité nulle ne contribue pas à l'entropie. Pour tout n , on a $H_{n+1}(p_1, \dots, p_n, 0) = H_n(p_1, \dots, p_n)$.

4. Remarquons que les conditions d'existence de l'entropie (à savoir que $\sum_{k=1}^{k=N} p_k = 1$ et $0 \leq p_k \leq 1, \forall k \in [1, N]$) donnent l'encadrement : $0 \leq H(\mathbf{x}) \leq \log(N)$

On représente ici l'allure de la courbe $-x \log(x)$ tirée de [15]



3.3.3 L'entropie d'un spectre

Nous avons jusqu'à présent défini l'entropie d'une variable aléatoire. Comment adapter cette notion pour en déduire des propriétés de la représentation temps-fréquence ? C'est de la thèse [15] que nous avons trouvé ces informations.

Pour un signal numérique la densité spectrale de puissance (DSP) pourra être assimilée à une densité de probabilité, même si pour cela, il faudra au préalable la normaliser.

En effet, on peut considérer que les valeurs possibles données par les fréquences peuvent être considérées comme une probabilité proportionnelle à leur énergie. On utilise dans ce cas le module au carré de la Transformée de Fourier discrète (TFD) en tant qu'estimateur de la DSP des tranches du signal.

Le "désordre" mesuré par l'entropie représente donc dans ce contexte :

- La richesse du contenu fréquentiel du signal en terme de richesse harmonique pour des signaux stationnaires ;

- et la richesse du contenu fréquentiel en terme d'imprévisibilité temporelle pour les signaux non stationnaires et brefs.

Définition 3.7 La TFD d'un signal numérique $x[n]$ pour $n \in [0, N-1]$ est définie pour $k \in [0, N-1]$ par :

$$X[k] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x[k] \cdot e^{-2i\pi k \frac{n}{N}}.$$

Il est important de remarquer que l'entropie est une propriété globale d'un signal. Rappelons que nous voulons l'utiliser pour donner un critère d'optimalité d'une représentation temps-fréquence localement. Une des difficultés est donc de définir une entropie de telle sorte qu'elle puisse nous donner des informations sur le signal de façon locale de telle sorte que ces informations puissent faire sens pour comparer des représentations temps-fréquence.

Soit un signal numérique quelconque noté $(x[n])_{n \in [0, N-1]}$, l'entropie au sens de la répartition énergétique des fréquences est bien définie avec la définition suivante :

Définition 3.8 L'entropie d'un signal numérique $(x[n])_{n \in [0, N-1]}$ est définie, pour $v = \sqrt{\sum_{k=0}^{N-1} x[n]^2}$ par :

$$H_{[0, N-1]}(x) = \sum_{k=0}^{N-1} \left(\frac{|X[k]|}{v} \right)^2 \cdot \log \left(\frac{|X[k]|}{v} \right)^2$$

On pourra remarquer que dans le cas d'un signal réel, il faut modifier un peu la définition à cause du fait que la répartition d'énergie est symétrique par rapport à $\frac{F_e}{2}$, où F_e est la fréquence d'échantillonnage. Il est donc nécessaire de n'effectuer la somme que pour les valeurs prises par un signal réel, soit sur $[0, \frac{F_e}{2}]$. La définition devient donc (définie dans [15]) :

Définition 3.9 L'entropie d'un signal numérique à valeurs réelles $(x[n])_{n \in [0, N-1]}$ est définie, pour $v = \sqrt{\frac{1}{2} (v^2 - |X[0]|^2 - |X[\frac{N}{2}]|^2)}$ par :

$$H_{[0, N-1]}(x) = \sum_{k=0}^{N/2} \left(\frac{|X[k]|}{v} \right)^2 \cdot \log \left(\frac{|X[k]|}{v} \right)^2$$

3.3.4 L'entropie de Rényi et ses propriétés

L'entropie de Rényi est très tôt proposée pour résoudre le problème d'adaptivité. On pourra se référer aux articles et thèses [23], [5]. L'entropie de Rényi est une généralisation de l'entropie de Shannon, elle a l'inconvénient de ne plus remplir les quatre propriétés (3.6) mais ces derniers auteurs considèrent qu'elle a des avantages d'un point de vue de l'implémentation.

Regardons à présent comment on la définit précisément.

Définition 3.10 On définit le spectrogramme normalisé par

$$S_x^g(t, f) = \frac{|G_x^g(t, f)|^2}{\|x\|^2}$$

où G_x^g est la transformée de Gabor continue du signal x avec la fenêtre g .

Remarque 3.11 On pourrait aussi utiliser la transformée de Wigner-Ville pour définir le spectrogramme normalisé.

Définition 3.12 *L'entropie de Rényi d'une représentation temps-fréquence est définie par*

$$R_\alpha(g) = \frac{1}{1-\alpha} \log_2 \int \int (S_x^g(t, f))^\alpha dt df$$

pour $\alpha \in]0, 1[$.

On pourrait ensuite utiliser ce critère en appliquant la fonction suivante :

$$\hat{s} = \arg \min_s R_\alpha(g_s)$$

Remarque 3.13 *Le paramètre α influencera évidemment la valeur de l'entropie : Lors de son utilisation pour le choix de la représentation temps-fréquence, elle aura tendance à favoriser davantage une représentation avec une petite fenêtre ou bien avec une grande fenêtre. La référence [22] précise que "de faibles valeurs de α ont tendance à privilégier la localisation temporelle, alors que les grandes valeurs de α privilégient la localisation fréquentielle". Ainsi il est clair que le choix, totalement arbitraire d'un paramètre α pour tout le plan temps-fréquence est d'intérêt limité. L'idéal serait de pouvoir choisir le critère selon le type de composante que l'on souhaite a priori mettre en évidence.*

L'entropie de Rényi est une généralisation de l'entropie de Shannon au sens où $H(x) = \lim_{\alpha \rightarrow 1} H_\alpha(x)$

Pour plus d'informations sur les entropies de Rényi et les informations qu'elles donnent sur les propriétés temps-fréquence on pourra se référer à l'article [5]. On y trouvera en particulier les problèmes de définition de l'entropie pour certains signaux, l'étude des propriétés et le potentiel des entropies de Rényi pour certaines applications.

3.3.5 Autres critères entropiques

Il existe de nombreux autres critères entropiques, nous citons ici les plus importants ainsi que ceux qui permettent une généralisation de la notion.

Les informations sur les critères dont nous parlons ci-dessous sont présents dans [15]

Il existe trois critères développés pour permettre une segmentation des signaux audio basés sur une mesure entropique :

- la courbe de dimension théorique (CDT), obtenue en prenant l'exponentielle de l'entropie.
- le critère entropique local (CEL) qui est une fonction similaire au test entropique utilisé dans les algorithmes de meilleure base.
- le contraste entropique local (CTEL) est une fonction de contraste calculée à partir de l'entropie de deux fenêtres adjacentes.

Par ailleurs, nous verrons dans la section suivante qu'on peut choisir une base de manière adaptative en fonction du signal. Les algorithmes de meilleure base que nous verrons en détail (dans 4.1) sélectionnent une base dans "une bibliothèque" de base, par minimisation d'une fonction de coût concave. Précisons quelles peuvent être ces fonctions de coût.

On pense en particulier aux fonctions de coût l^p pour $p < 2$, définies par $\phi(x) = x^{p/2}$. ϕ est une fonction concave pour $x > 0$. Le coût correspondant à cette fonction s'écrit :

$$C(f, B) = \sum_{m=1}^N \frac{|\langle f, g_m \rangle|^p}{\|f\|^p}$$

où B est la base dont on calcule le coût.

Remarque 3.14 *On peut remarquer ici le lien avec la définition de l'entropie de Rényi dans le cas où $p=2$.*

• Limites de l'utilisation globale et locale du critère

Si la question importante est de savoir quel critère choisir, il est aussi important de comprendre comment l'utiliser et quelles sont les limites de son utilisation.

Il est évident, comme nous l'avons déjà expliqué, que l'utilisation du critère comme étant un critère *global* du signal n'est pas satisfaisant. En effet, les différentes composantes d'un signal peuvent avoir des caractéristiques temps-fréquence tellement différent que cela ne fait pas forcément sens. Par ailleurs, il s'agit dans ce sujet de proposer une méthode permettant une *résolution variable*.

Nous devons donc nous tourner vers une utilisation locale. C'est-à-dire qu'après sélection et identification d'une région particulière du plan temps-fréquence aux caractéristiques temps-fréquence homogènes, nous pouvons être en mesure de dire quelle taille de fenêtre est la mieux adaptée.

Cependant, si cette approche peut être dans certains cas pratique, elle nécessite l'intervention de l'utilisateur et n'est donc pas satisfaisante.

Enfin, faisons une brève remarque sur une utilisation du critère entropique : on pourrait penser utiliser le critère trame par trame et faire le calcul localement de cette façon. C'est-à-dire : on calcule l'entropie d'une trame avec une petite fenêtre, et l'entropie d'une trame avec une plus grande fenêtre. Cependant, il n'est pas sensé de comparer les deux entropies puisqu'elles ne sont pas évaluées sur le même extrait de signal, la taille de la trame ayant été modifiée.

3.3.6 L'entropie : un critère de parcimonie

Comme nous l'avons vu, les critères de sélection utilisant l'entropie considèrent qu'une représentation temps-fréquence est de bonne qualité quand elle concentre les informations (à savoir l'énergie) sur une faible quantité de points, ou du moins sur la quantité la plus faible possible. En effet, nous avons déjà expliqué que l'entropie favorisait les représentations avec des coefficients non nuls sur le plus faible nombre de points. C'est donc tout simplement un critère de parcimonie de la représentation.

D'une manière plus générale, on peut par conséquent considérer tous les critères de parcimonie comme critère de sélection. C'est ce que nous avons choisi pour la suite du travail et pour proposer une méthode alternative aux propositions présentées ici.

3.3.7 Conclusion

Nous avons eu l'occasion dans cette section de dresser un tableau quasi-exhaustif de tous les critères de sélection d'une bonne représentation temps-fréquence. Par ailleurs, nous avons pu comprendre la difficulté d'utiliser ces outils pour faire un choix de "bonne" représentation *localement*. Nous avons également soulevé le fait qu'un critère, pour être effectif, doit nécessairement être "relié" à une certaine représentation. En effet, il faut que le critère puisse être un outil de sélection ou de comparaison entre des représentations, mais cette opération ne fait pas forcément sens comme nous l'avons vu dans les limites de l'utilisation de tels critères. Nous nous penchons dans la section suivante sur le problème de la représentation à associer au critère de sélection. En effet, comme nous l'avons déjà soulevé, si le critère sélectionne il n'est pas évident de savoir ce qu'il sélectionne : une base avec une taille de fenêtre, une base avec des tailles de fenêtres différentes, des atomes ... C'est ce que nous étudierons dans la section 4.

4 Représentations adaptées et adaptatives

On traite dans cette section, toutes les représentations "adaptées ou adaptatives". On considère dans la suite qu'une représentation est adaptée ou adaptative si elle permet plusieurs types de pavage temps-fréquence et ainsi différentes représentations et résolutions temps-fréquence.

4.1 Recherche de meilleure base, 'best basis'

4.1.1 Principe

Les algorithmes de meilleure base ([7]) permettent d'adapter la base utilisée au signal x . Le critère d'optimalité est la minimisation d'une fonction de coût additive ce qui revient à un critère entropique en ce qu'elle concentre autant que possible son énergie sur peu de coefficients. Les algorithmes de meilleure base choisissent une base dans une bibliothèque de base orthonormale.

Ces décompositions conduisent à des segmentations temporelles ou fréquentielles, elles permettront le choix de meilleure base. Il s'agit de trouver la meilleure combinaison des atomes générés par les bases des différentes itérations.

4.1.2 Algorithme plus détaillé

Nous présentons ici de façon détaillée l'algorithme de recherche de meilleure base dans l'article [7]. Il est à noter que cet algorithme fut créé dans le but de la compression. Il s'appuie sur une structure dyadique, où chaque sous-bande est le père de deux fils. En voici les différentes étapes :

1. La décomposition est effectuée jusqu'à la profondeur équivalente à la résolution la plus fine souhaitée.

2. Une fonction de coût est calculée pour les paquets obtenus aux itérations successives. On appelle poids la valeur de la fonction de coût d'un paquet. Les critères fréquemment utilisés sont l'entropie de Shannon, l'énergie et toutes les normes de L^2 .

3. En partant de la profondeur maximale, à savoir le bas de la décomposition, la somme des poids de deux fils est comparée au poids de leur père. Deux cas sont possibles :

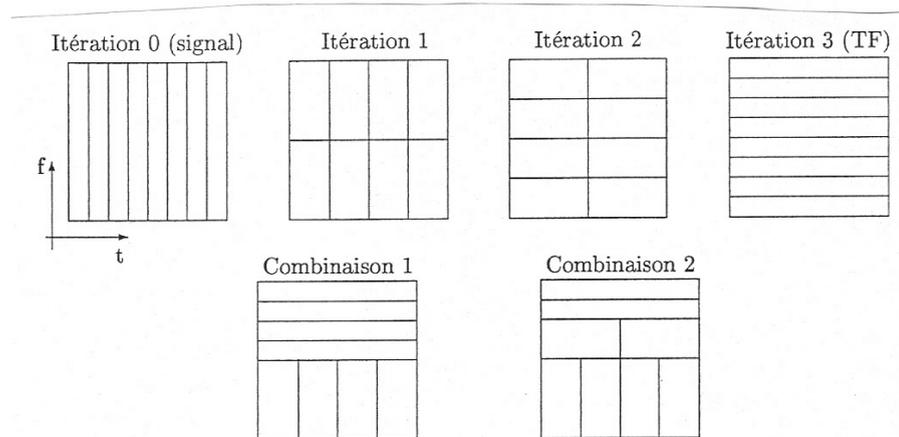
–Si la somme des poids des fils est supérieure au poids du père, alors le père est conservé. Autrement dit, l'atome temps-fréquence représenté par le père est plus adapté au signal que les deux fils, il est alors logique de le garder.

–Si la somme des poids des fils est inférieure au poids du père, alors les deux fils sont conservés. Le poids du père est remplacé par la somme des poids des fils pour la suite du test à la profondeur supérieure.

4. Le test reprend à l'étape 1 à la profondeur supérieure.

En résumé, l'algorithme conserve le père lorsque la sous-bande est vide ou de faible énergie et conserve le fils lorsqu'au moins un des deux contient du signal.

Nous représentons ici les bases obtenues aux différentes itérations de l'algorithme décrit ci-dessus sur la première ligne. Sur la seconde, nous représentons deux combinaison linéaire de ces bases.



Cette figure est tirée de[15]

4.1.3 Conclusion et bilan

Nous pouvons voir que l'algorithme tel qu'il est décrit et construit précédemment ne permet pas une adaptativité totale de la représentation au signal. En effet, certaines configurations, qui pourraient être les plus avantageuses, ne sont pas possible. Par exemple, il est impossible d'avoir au début du signal la succession de fenêtres suivante : 'petite, grande, petite'. Nous allons donc essayer de remédier à cette "rigidité" de l'adaptation.

4.2 Les représentations adaptées en temps ou en fréquence.

Nous présentons ici la transformée de Gabor instationnaire présentée dans [23] et appelée également transformée covariante par translation temporelle ou par translation fréquentielle. En particulier, nous (re)démontrons dans la section 4.2.1 le cas discret nécessaire à une implémentation. Cette partie n'étant pas présente dans les références à savoir dans [23], il a fallu la démontrer pour valider le sens d'une implémentation vis à vis de la théorie.

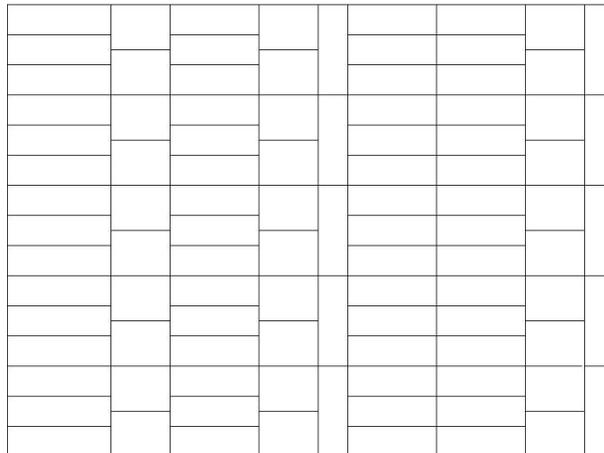
4.2.1 La transformée de Gabor instationnaire : Généralités

Le principe de la transformée de Gabor instationnaire est grossièrement une analyse de Gabor avec une fenêtre pouvant être de taille variable en temps ou en fréquence.

Comme nous l'avons dit précédemment, le terme transformée de Gabor instationnaire regroupe la transformée covariante par translation temporelle et la transformée covariante par translation fréquentielle. Comme nous le verrons, le temps et la fréquence étant duaux dans ces définitions, il n'y a pas de différences fondamentales entre les deux.

• Pavage de la transformée covariante par translation temporelle.

On présente ci-dessous le pavage d'une transformée covariante par translation temporelle.



4.2.2 La transformée covariante par translation temporelle

On définit :

Définition 4.1

– **La famille de fenêtre** : Soit $(g_\tau)_{\tau \in \mathbb{R}}$ une famille de fenêtre (τ est un paramètre temporel) définie dans $L^2(\mathbb{C})$ telle que $\forall t \in \mathbb{R}, \int_{\mathbb{R}} |g_\tau(t)|^2 d\tau = 1$ (condition d'admissibilité)

– **Les atomes** : $g_{\tau\nu}(t) = g_\tau(t)e^{i2\pi\nu(t-\tau)}$

Définition 4.2 On définit la transformée de Gabor instationnaire $H_x^g(\tau, \nu)$ d'un signal $x \in L^2(\mathbb{C})$ avec une famille $g = (g_\tau)_{\tau \in \mathbb{R}}$ par le produit scalaire sur les atomes de la décomposition :

$$H_x^g(\tau, \nu) = \int_{\mathbb{R}} x(t) \overline{g_\tau(t)} e^{-i2\pi\nu(t-\tau)} dt$$

Remarque 4.3 sur cette transformée*1-Principe de la décomposition atomique*

Cette transformée peut être évidemment vue comme une décomposition atomique :

On choisit pour chaque valeur de τ , une fenêtre d'analyse g_τ bien localisée temporellement autour de τ . On a maintenant la possibilité de faire varier la taille et/ou la forme de g_τ en fonction de τ . Ainsi les caractéristiques temps-fréquence varient avec le temps. Notons bien que la résolution fréquentielle ne varie pas ici à un instant τ donné.

2-Rapport avec la TFCT

La transformée de Gabor instationnaire peut, comme la transformée de Gabor classique vue, à un instant τ donné, comme un spectre à court terme du signal.

• Propriétés fondamentales de la transformée

On mentionne ici les propriétés fondamentales pour son utilisation pratique.

Proposition 4.4 On a les propriétés suivantes :

- *Linéarité* : $\forall (\lambda, \mu) \in \mathbb{C}^2, \forall (x, y) \in (L^2(\mathbb{R}))^2, H_{\lambda x + \mu y}^g = \lambda H_x^g + \mu H_y^g$
- *Covariance* : Soit $v_0 \in \mathbb{R}, x_{v_0} = x(t)e^{i2\pi v_0 t}, \forall (\tau, v) \in \mathbb{R}^2, H_{x_{v_0}}^g(\tau, v) = e^{i2\pi v_0 \tau} H_x^g(\tau, v - v_0)$
- *Isométrie* : La transformée conserve le produit scalaire $(H_x^g, H_y^g) = (x, y)$
- *Inversion* : $\forall x \in L^2(\mathbb{R}), x = \int \int H_x^g(\tau, v) d\tau dv$.
- *Conservation de l'énergie* $\int |x(t)|^2 dt = \int \int |H_x^g(\tau v)|^2 d\tau dv$

• Version discrète de la transformée utile pour l'implémentation

Enfin, on définit la transformée dans le cas discret et redémontre l'isométrie dans ce cas permettant de retrouver les propriétés importantes dans ce cas-ci (inversion et conservation de l'énergie).

Définition 4.5 On définit la condition d'admissibilité comme :

$$\forall n \in \mathbb{N}, \sum_{m \in \mathbb{Z}} |h_m[n]|^2 = 1$$

Définition 4.6 On définit la famille d'atomes discrète par

$$\forall n \in \mathbb{N}, h_{m,k}[n] = h_m[n] e^{i2\pi \frac{k}{N}(n - \tau_m)}$$

Définition 4.7 On définit la transformée dans l'espace suivant :

$$\mathcal{E} = \left\{ X : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{C} / \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{m \in \mathbb{Z}} \sum_{k=-N}^{k=+N} |X(m, k)|^2 < \infty \right\}$$

Nous donnons ici les propriétés importantes dans le cas discret :

Proposition 4.8 *Nous avons les propriétés et définitions suivantes :*

– *Produit scalaire : \mathcal{E} muni du produit scalaire $\forall (X, Y) \in \mathcal{E}$,*

$$(X, Y) = \lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{m \in \mathbb{Z}} \sum_{k \in [-N, N]} X(m, k) \bar{Y}(m, k)$$

(espace de Hilbert)

– *Définition : Transformée du signal x définie sur \mathcal{E}*

$$I_x^h(m, k) = \sum_{n \in \mathbb{N}} x[n] \bar{h}_m[n] e^{-i2\pi \frac{k}{N}(n-\tau_m)}$$

– *Propriétés conservées :*

Propriété d'isométrie : $\forall (x, y) \in (L^2(\mathbb{R}))^2$, $(I_x^h, I_y^h) = (x, y)$.

Formule d'inversion $x = \lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{m \in \mathbb{Z}} \sum_{k \in [-N, N]} I_x^h(m, k) h_{m,k}$

Toutes ces propriétés sont données facilement par la formule d'isométrie. Pour cette raison, on ne les démontrera pas. Par contre nous démontrons ci-dessous la propriété d'isométrie. Précisons que cette preuve ne figure pas dans les références et qu'il a donc fallu la démontrer pour valider une implémentation dans le cas discret.

Preuve 4.9 *Nous vérifions ici la propriété d'isométrie avec les définitions posées plus haut.*

La définition de l'ensemble sur lequel on travaille est la suivante :

$$\mathcal{E} = \left\{ X : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{C} \text{ telle que } \lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{k \in [-N, +N]} \sum_{m \in \mathbb{Z}} |X(m, k)|^2 < \infty \right\}.$$

On pose pour un signal x

$$I_x^h(m, k) = \sum_{n \in \mathbb{N}} x[n] \bar{h}_m[n] e^{-i2\pi \frac{k}{N}(n-\tau_m)}$$

La première chose à montrer est que si x est dans $\ell^2(\mathbb{Z}^2)$ alors I_x^h est dans \mathcal{E} . Pour cela on montre que cette transformation est une isométrie, c'est-à-dire que pour tous signaux x et y on a

$$\langle I_x^h, I_y^h \rangle = \langle x, y \rangle$$

avec

$$\langle I_x^h, I_y^h \rangle := \lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{k \in [-N, +N]} \sum_{m \in \mathbb{Z}} \left(\sum_{n \in \mathbb{Z}} x[n] h_m[n] e^{2i\pi \frac{k}{N}(n-\tau_m)} \right) \overline{\left(\sum_{n \in \mathbb{Z}} y[n] h_m[n] e^{2i\pi \frac{k}{N}(n-\tau_m)} \right)}$$

On pose pour cela P la longueur de y , K la longueur de x et M le maximum des longueurs des fenêtres h_m . On pose alors $L = \max(K, P)$.

On a

$$\begin{aligned}
& \frac{1}{2N+1} \sum_{k \in [-N, +N]} \sum_{m \in [-M, +M]} \left(\sum_{l \in [0, L]} x[l] h_m[l] e^{2i\pi \frac{k}{N} (l - \tau_m)} \right) \overline{\left(\sum_{p \in [0, L]} y[p] h_m[p] e^{2i\pi \frac{k}{N} (p - \tau_m)} \right)} = \\
& \frac{1}{2N+1} \sum_{m \in [-M, +M]} \sum_{k \in [-N, +N]} \left(\sum_{l \in [0, L]} x[l] h_m[l] e^{2i\pi \frac{k}{N} (l - \tau_m)} \right) \overline{\left(\sum_{p \in [0, L]} y[p] h_m[p] e^{2i\pi \frac{k}{N} (p - \tau_m)} \right)} = \quad (1) \\
& \frac{1}{2N+1} \sum_{m \in [-M, +M]} \sum_{l \in [0, L]} \sum_{p \in [0, L]} x[l] h_m[l] \overline{y[p] h_m[p]} \sum_{k \in [-N, +N]} e^{2i\pi \frac{k}{N} (l - \tau_m)} e^{-2i\pi \frac{k}{N} (p - \tau_m)} = \\
& \frac{1}{2N+1} \sum_{m \in [-M, +M]} \sum_{l \in [0, L]} \sum_{p \in [0, L]} x[l] h_m[l] \overline{y[p] h_m[p]} \sum_{k \in [-N, +N]} e^{2i\pi \frac{k}{N} (l - p)}
\end{aligned}$$

Remarquons que

$$\sum_{k \in [-N, +N]} e^{2i\pi \frac{k}{N} (l - p)} = \frac{\sin\left(\pi \frac{2N+1}{N} (l - p)\right)}{\sin\left(\frac{\pi}{N} (l - p)\right)}$$

et on en déduit que si l est différent de p on a

$$\lim_{N \rightarrow +\infty} \sum_{k \in [-N, +N]} e^{2i\pi \frac{k}{N} (l - p)} = 0$$

et par contre si $p = l$

$$\lim_{N \rightarrow +\infty} \sum_{k \in [-N, +N]} e^{2i\pi \frac{k}{N} (l - p)} = 2N + 1$$

Dès lors

$$\begin{aligned}
& \lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{m \in [-M, +M]} \sum_{l \in [0, L]} \sum_{p \in [0, L]} x[l] h_m[l] \overline{y[p] h_m[p]} \sum_{k \in [-N, +N]} e^{2i\pi \frac{k}{N} (l - p)} = \\
& \sum_{m \in [-M, +M]} \sum_{l \in [0, L]} |h_m[l]|^2 x[l] \overline{y[l]} = \sum_{l \in [0, L]} x[l] \overline{y[l]} \text{ avec la condition d'admissibilité} \quad (2)
\end{aligned}$$

et en passant à la limite sur M et L on a bien

$$\langle I_x^h, I_y^h \rangle = \sum_l x[l] \overline{y[l]} = \langle x, y \rangle$$

• Quelques remarques sur la construction des atomes

On cherche donc une famille $(h_m)_{m \in \mathbb{Z}}$ telle que $\forall n \in \mathbb{N}, \sum_{m \in \mathbb{Z}} |h_m[n]|^2 = 1$ (1). Nous savons que pour cela il suffit de trouver une famille η_k telle que il existe (A, B) vérifiant $0 < A \leq B < \infty$ et $\forall n \in \mathbb{N}, A \leq \sum_k |\eta_k[n]|^2 \leq B$ qu'on normalisera pour vérifier (1).

Regardons de plus près comment choisir les η_k . On les choisit tels que :

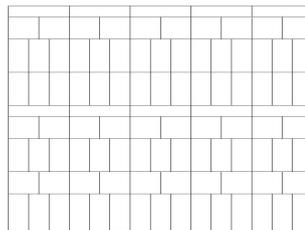
- $\text{Card}\{\text{Supp}(\eta_k)\} < \infty$ (1)
- $\eta_k > 0, \forall n \in \text{Supp}(\eta_k)$ (2)
- $\bigcup_k \text{Supp}(\eta_k) = \mathbb{N}$ (3)
- $\exists N \in \mathbb{N}^* / \forall n \in \mathbb{N}, \text{Card}\{k/n \in \text{Supp}(\eta_k)\} < N$ (4)

Remarque 4.10 On peut remarquer que la propriété (1) impose que la fenêtre soit de taille finie, que la propriété (2) impose qu'elle soit positive, la (3) que toutes les fenêtres recouvrent l'ensemble du signal et la propriété (4) impose que chaque échantillon ne soit recouvert que par un nombre fini de fenêtres. Seule la propriété (4) est nécessaire à une bonne construction, les autres correspondent à la façon dont on les construit dans la pratique.

• **La transformée covariante par translation temporelle.**

Dans le cas de la transformée covariante par translation temporelle, nous avons des définitions et propriétés complètement analogues. On indique ici à titre indicatif quelques définitions ainsi que le pavage temps-fréquence correspondant à la transformée.

- Condition d’admissibilité $\forall \xi \in \mathbb{R}, \int |\hat{g}_k(\xi)|^2 d\nu = 1$
- On définit l’atome par $g_{\tau\nu}(t) = g_\nu(t - \tau)$
- et la transformée par $K_x^g(\tau, \nu) = \int x(t) \bar{g}_\nu(t - \tau) dt$
- Pavage temps-fréquence



4.2.3 Conclusion et bilan

On peut à présent analyser les avantages et inconvénients de cette transformée.

Avantages

Elle présente tout d’abord l’avantage d’avoir une résolution temporelle ou une résolution fréquentielle variable. Tout ceci en assurant la formule d’inversibilité avec une charge de calcul limitée. Par ailleurs, il est possible avec cette transformée et associée à un critère de sélection de la taille de réaliser une optimisation locale en sélectionnant à l’avance la zone du plan temps-fréquence pour laquelle on veut choisir la ”meilleure” taille de fenêtre.

Inconvénient

Par ailleurs, on peut voir qu’elle présente quelques inconvénients. Comme nous avons pu le remarquer, la résolution s’effectue en temps ou en fréquence mais ne peut pas s’effectuer à la fois en temps et en fréquence. Ce qui limite d’ores et déjà les possibilités d’adaptativité de cette transformée. Par ailleurs, elle impose un condition d’admissibilité sur les fenêtres même si de petites variations au niveau local de la taille permet de respecter cette condition. Enfin, même si l’implémentation est simple, il n’est pas possible de la relier à un critère de sélection automatique sans intervention de l’utilisateur.

4.3 La représentation multicouche

On présente dans cette section la représentation multicouche issue de [23]. On présentera successivement, le partitionnement en fréquence qui est un concept important de la représentation, l'algorithme puzzle temps-fréquence qui réalise la représentation multicouche et la reconstruction. Enfin, on tentera de comprendre plus précisément la nature d'une représentation multicouche et les qualités de l'algorithme décrit.

4.3.1 Le partitionnement du plan

Le partitionnement du plan est la première étape de l'algorithme.

On partitionne le plan temps-fréquence en rectangles $(R_{m,n})_{(m,n) \in \mathbb{Z}^2}$ de largeur temporelle $\Delta_\tau > 0$ et de largeur fréquentielle $\Delta_\nu > 0$:

$$R_{m,n} = [(m - \frac{1}{2})\Delta_\tau, (m + \frac{1}{2})\Delta_\tau] \times [(n - \frac{1}{2})\Delta_\nu, (n + \frac{1}{2})\Delta_\nu].$$

Ils constituent bien une partition du plan temps-fréquence car $\mathbb{R}^2 = \bigcup_{(m,n) \in \mathbb{Z}^2} R_{m,n}$

On choisit ensuite dans chaque rectangle $(R_{(m,n)})_{(m,n) \in \mathbb{Z}^2}$ la meilleure fenêtre au sens du critère de sélection.

Un problème déterminant pour l'algorithme est qu'il n'y a pas de reconstruction parfaite. En effet, de trop nombreuses erreurs dues aux zones de transition entre les représentations empêchent cette reconstruction. Pour y remédier, il va donc falloir construire l'algorithme itératif de "puzzle temps-fréquence" que nous décrivons ci-dessous.

4.3.2 L'algorithme puzzle temps-fréquence

• Principe de l'algorithme

Nous décrivons étape par étape le principe de l'algorithme :

- Sélection de la fenêtre : On teste pour chaque fenêtre et pour chaque rectangle le critère entropique, on choisit la fenêtre qui minimise le critère pour un rectangle donné. Ceci pour chacun des rectangles.
- On inverse le résultat même s'il n'y a pas de reconstruction exacte.
- On calcule le résidu, c'est-à-dire la différence entre le signal initial et le signal reconstruit.
- On itère le processus en effectuant exactement le même traitement sur le résidu.
- On constate que l'algorithme converge même s'il n'existe pas de preuve mathématique de la convergence à ce jour.

• Détails de l'algorithme

- On définit un partitionnement du plan temps-fréquence comme décrit plus haut, c'est-à-dire en rectangles :

$(R_{m,n})_{(m,n) \in \mathbb{Z}^2}$ de largeur temporelle $\Delta_\tau > 0$ et de largeur fréquentielle $\Delta_\nu > 0$:

$$R_{m,n} = [(m - \frac{1}{2})\Delta_\tau, (m + \frac{1}{2})\Delta_\tau] \times [(n - \frac{1}{2})\Delta_\nu, (n + \frac{1}{2})\Delta_\nu]$$

On notera dans la suite chaque rectangle R_s où $(s = (m, n))$

- On définit la grille \mathcal{L} (*Lattice*) qui représente la grille de discrétisation due à *un certain repère de Gabor* r par :

$$\mathcal{L}^r = \tau_r \mathbb{Z} \times \nu_r \mathbb{Z}$$

avec τ_r et ν_r pas de discrétisation posés a priori arbitrairement.

- Ensuite, pour chaque repère r , on définit $\mathcal{L}^{r,s} = \mathcal{L}^r \cap R_s$ où $(s = (m, n))$ qui identifie les atomes qui se trouvent dans un rectangle donné. $\mathcal{L}^{r,s}$ représentent ainsi les éléments de la grille qui sont dans R_s .

- On choisit ensuite le fenêtre optimale au sens du critère défini (on considère ici l'entropie de Rényi) : $r(x, s) = \operatorname{argmin}_{r=0 \dots R-1} H_\alpha(v(x, r, s))$ où $v(x, r, s)$ est $\langle x, g_{mn}^s \rangle$, $(m, n) \in \mathcal{L}^{r,s}$. De cette façon, on définit pour le rectangle s du signal x un repère de Gabor adapté et noté $r(x, s)$.

- On obtient ainsi un premier résultat (où on note (\cdot, \cdot) le produit scalaire) :

$$x^{(1)} = \sum_s \sum_{\lambda \in \mathcal{L}^{r(x,s),s}} (x, g_\lambda^{r(x,s)}) g_\lambda^{r(x,s)}$$

et le résidu $\mathcal{R}^1(x) = x - x^{(1)}$

- Et on peut enfin itérer le processus pour obtenir une bonne reconstruction du signal en plusieurs couches. On a ainsi l'équation suivante à la k ème itération

$$x^{(k)} = \sum_s \sum_{\lambda \in \mathcal{L}^{r(\mathcal{R}^{k-1}(x),s),s}} (\mathcal{R}^{k-1}(x), g_\lambda^{r(\mathcal{R}^{k-1}(x),s)}) g_\lambda^{r(\mathcal{R}^{k-1}(x),s)}$$

qui représente le "bout" du signal reconstruit à la k ème itération.

- et le résultat final $x = \sum_{k=1}^K x^{(k)} + \mathcal{R}^k(x)$ qui représente le signal reconstruit à la k ème itération.

4.3.3 La représentation multicouche

• Principe

Nous essayons ici de comprendre plus précisément en quoi l'algorithme décrit ci-dessus réalise une représentation multicouche. Pour un repère r fixé, on isole les composantes du signal qui sont "bien représentées" avec le repère r .

On appelle ainsi une couche les composantes qui sont bien représentées selon un repère r donné. En effet, si on effectue une reconstruction partielle pour chacun des repères r , on obtient une représentation multicouche du signal.

Cette représentation est particulièrement intéressante avec deux types de repères de Gabor différents. C'est-à-dire avec un repère avec des petites fenêtres en temps et un autre avec de petites fenêtres en fréquence. Considérer cette représentation avec uniquement deux repères de Gabor différents permet d'isoler les composantes avec des caractéristiques temps-fréquences différentes : on isolera les composantes localisées en temps à savoir les transitoires des composantes tonales qui sont quasi-stationnaires.

• Etude de l'algorithme

On présente ici quelques uns des résultats présentés dans [23] lors de tests permettant de comprendre sa dépendance à certains paramètres :

On constate une forte dépendance de l'algorithme au type de signal ainsi qu'à la taille des rectangles. Dans le cas avec deux couches :

- Type de signal : On calcule le rapport signal à bruit en 20 itérations
 - pour un bruit blanc : 140dB
 - pour un glockenspiel : 160dB
 - pour des sinusoides+impulsions : 170dB
- Taille des rectangles : La vitesse de convergence augmente avec la taille des rectangles.
 - L'énergie présente dans le résidu provient des erreurs de reconstruction aux interfaces entre les rectangles temps-fréquence.
 - Il y a donc un compromis à faire car une grande taille de rectangles permet une convergence rapide mais ne permet pas une localisation concentrée dans le plan temps-fréquence.

• Puzzle et seuillage entropique

Un inconvénient de cet algorithme est que pour certaines composantes, la "meilleure" représentation par l'algorithme n'est pas forcément "bonne", c'est-à-dire suffisamment parcimonieuse. Indépendamment de la fenêtre utilisée, certaines composantes comme un bruit blanc par exemple, n'auront pas de représentation parcimonieuses.

Ceci peut être problématique pour interpréter l'algorithme multicouche. Par exemple, on peut se demander comment seront représentées les composantes bruitées ? Elles les seront par les deux tailles de fenêtres ; si on utilise un algorithme avec deux couches, à savoir, deux repères de Gabor. Ainsi des composantes peuvent être partagées entre deux couches.

On peut y remédier avec l'ajout d'un seuil sur les valeurs d'entropies pour avoir la parcimonie satisfaisante. De cette façon, on fait en sorte que les composantes pour lesquelles la représentation n'est pas suffisamment parcimonieuse (suffisamment étant représenté par la valeur du seuil) ne soient pas représentées et restent dans le résidu.

• Conclusion sur la représentation multicouche

Avantages

Ajoutons tout d'abord qu'il existe une deuxième variante de cet algorithme présente dans [23]. Nous ne l'avons pas décrit ici car il semblait que la première était suffisante pour comprendre les concepts importants de cette représentation. Vis à vis de la transformée de Gabor instationnaire

dont nous avons parlé plus haut, elle représente l'avantage indéniable de permettre une adaptativité de la représentation en temps ET en fréquence. Et ceci avec une reconstruction aussi parfaite que souhaitée. On peut en effet, rendre l'erreur de reconstruction aussi petite que souhaitée grâce à la convergence (expérimentale) de l'algorithme.

On remarque en particulier que l'adaptativité proposée ici est particulièrement intéressante si les fenêtres sont très différentes et dans le cas de deux couches seulement. Pour l'instant, l'intérêt d'avoir un plus grand nombre de couches n'est pas évident.

Inconvénients

Enfin, si les résultats sont convaincants, la complexité de l'algorithme et la difficulté de l'implémenter nous a encouragé à chercher d'autres méthodes qui pourraient être traitées dans le temps cours et limité de ce stage.

4.4 Le Matching Pursuit

Le Matching Pursuit s'inscrit dans les algorithmes de poursuite dont il existe plusieurs sortes. Quelques mots sur les algorithmes de poursuite pour le situer dans son contexte :

Les bases de Fourier et d'ondelette donnent une approximation linéaire très précise pour des signaux particulièrement réguliers. La poursuite de base ou de vecteurs ajoute des degrés de liberté supplémentaires dans le choix de la base ou dans celui des vecteurs pour représenter le signal. Les vecteurs ainsi choisis mettent en avant les structures principales du signal et donnent par ce biais ses caractéristiques temps-fréquence.

Pour mieux comprendre les qualités des approximations par poursuite, donnons l'analogie trouvée dans [28] :

”Un enregistrement musical comprend souvent au même instant des notes de durées différentes [...]. La même note de musique peut aussi avoir des durées différentes quand elle est jouée à des instants différents [...]. Pour approximer efficacement un signal musical, il faut que la décomposition dispose de la même liberté que le compositeur, qui peut choisir librement les atomes temps-fréquence (les notes) les mieux adaptées à la représentation d'un son”

4.4.1 Matching Pursuit et représentation avec résolution adaptative

Nous allons décrire précisément l'algorithme de Matching Pursuit dans la suite mais tentons d'expliquer dès à présent en quoi cette méthode permet une représentation temps-fréquence avec résolution adaptative.

Le Matching Pursuit cherche à approcher un signal x grâce à une décomposition atomique où les atomes sont *choisis* dans un dictionnaire (ensemble d'atomes temps-fréquence) fixé au départ. La seule condition imposée au départ sur le dictionnaire est qu'il soit complet, c'est-à-dire que les combinaisons linéaires de ses atomes soient denses dans l'ensemble de définition du signal, $L^2(\mathbb{R})$. Par la suite, les atomes qui seront utilisés dans la décomposition linéaire sont choisis de manière adaptative par rapport au signal x donné au départ.

Le choix des atomes compris dans le dictionnaire est donc relativement libre et on peut le choisir de telle sorte qu'il contienne *plusieurs tailles* d'atomes temps-fréquence. L'algorithme Matching Pursuit choisira parmi les différents atomes ceux dont la taille est la mieux adaptée aux caractéristiques locales du signal.

Ainsi, il est clair que le Matching Pursuit définit des représentations adaptatives de signaux sonores en représentant le signal sous forme de combinaisons linéaires d'atomes extraits du dictionnaire. Le critère de sélection est ici la corrélation de l'atome avec le signal. Plusieurs méthodes sont proposées dans [13] permettant chacune de représenter certaines structures caractéristiques. En particulier, le Matching Pursuit Harmonique[13] décompose le signal en composantes harmoniques élémentaires alors que le Matching Pursuit Chirpé[13] extrait les variations de fréquence instantanée.

4.4.2 Principe du Matching Pursuit

Au début de l'algorithme, on sélectionne l'atome le plus corrélé avec le signal observé. On obtient ainsi une estimée à la première itération de l'algorithme. Celle-ci est le projeté orthogonal du signal sur l'atome sélectionné. Si on fait la différence de cette estimée avec le signal du départ, on obtient le résidu.

L'algorithme consiste à recommencer l'étape précédente sur le résidu obtenu : on cherche l'atome du dictionnaire le plus corrélé avec le résidu, puis on estime le résidu en le projetant sur l'atome sélectionné. Ensuite on renouvelle le processus jusqu'à ce que le critère d'arrêt soit atteint. Dans la description détaillée de l'algorithme, on notera $r^{(k)}$ le résidu de l'estimée $\hat{x}^{(k)}$ du signal x à l'étape k de l'algorithme, g_k l'atome choisi à l'étape k de l'algorithme.

4.4.3 La description de l'algorithme

Plus précisément, on distingue les étapes suivantes :

0. On initialise $\hat{x}^{(0)}$ à 0, $r^{(0)}$ à x puis k à 1.

A l'itération k ,

1. On calcule les corrélations $\{\langle g_i, r^{(k-1)} \rangle = g_i^T \cdot r^{(k-1)}\}_{i=1..n}$

2. On choisit la plus forte de ces corrélations :

$$i_k = \arg \max_i |\langle g_{i_k}, r^{(k-1)} \rangle|$$

3. On met à jour l'estimée ainsi que son résidu :

$$\begin{aligned}\hat{x}^{(k)} &= \hat{x}^{(k-1)} + \langle g_{i_k}, r^{(k-1)} \rangle g_{i_k} \\ r^{(k)} &= x - \hat{x}^{(k)} = r^{(k-1)} - \langle g_{i_k}, r^{(k-1)} \rangle g_{i_k}\end{aligned}$$

4. On incrémente k de 1 et on reprend les étapes 1, 2 et 3 jusqu'à ce que le critère d'arrêt soit atteint. Le critère d'arrêt peut être le nombre d'itérations tout comme un rapport signal à bruit minimum.

4.4.4 Le choix du dictionnaire

On donne ici quelques définitions de dictionnaires multi-échelles d'atomes temps-fréquences. Précisons que le terme multi-échelle signifie qu'il y a plusieurs tailles de fenêtres.

• Dictionnaire de Gabor

On définit le dictionnaire multi-échelle de Gabor D_{Gabor} par l'ensemble des atomes temps-fréquence de la forme :

$$g_{a,t,f}(t') = \frac{1}{\sqrt{a}} g\left(\frac{t' - t}{a}\right) e^{i2\pi f t'}$$

Ces atomes ont été obtenus par dilatation, translation temporelle et modulation fréquentielle d'une fenêtre g . On choisit pour g une fonction paire, positive, dont l'énergie est localisée autour du temps 0

• Dictionnaire de chirplets

Le dictionnaire de "chirplets" D_{chirp} généralise le dictionnaire de Gabor. En effet, comme nous le voyons dans la définition ci-dessous, le D_{chirp} comprend les atomes dont la fréquence peut varier avec le temps. Ils s'écrivent :

$$g_{a,t,f}(t') = \frac{1}{\sqrt{a}} g\left(\frac{t' - t}{a}\right) e^{i2\pi\left[f t' + \frac{c}{2}(t' - t)^2\right]}$$

On peut ainsi remarquer que chaque atome a une fréquence instantanée de

$$f_{g_{a,t,f},c}(t') = f + c(t' - t)$$

• Utilisation de la librairie MPTK

MPTK (Matching Pursuit Toolkit) est une implémentation efficace, développée à l'IRISA, de l'algorithme Matching Pursuit, pour la décomposition parcimonieuse de signaux multicanaux. MPTK permet de traiter des problèmes à grande échelle (ex : signaux de plus de 50 millions d'échantillons) en un temps raisonnable.

Elle permet de définir de manière très précise le dictionnaire que l'on veut utiliser, les atomes qui y sont et leurs caractéristiques temps-fréquence.

Pour le manuel d'utilisation, on pourra se référer à [25]

• Choix du dictionnaire

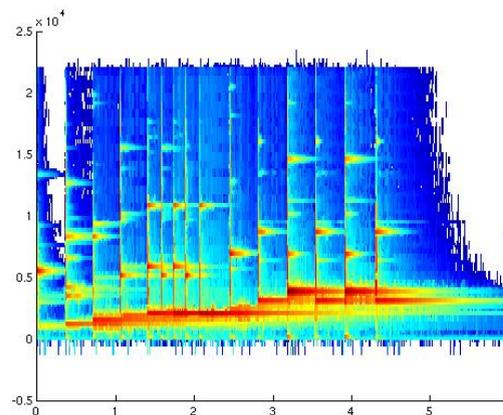
À présent que nous avons les outils théoriques et pratiques pour l'utilisation du Matching Pursuit, il ne reste qu'à choisir précisément le dictionnaire à utiliser.

Comme nous l'avons déjà noté, l'algorithme Matching Pursuit est intéressant pour traiter notre sujet car on peut utiliser plusieurs tailles et plusieurs types d'atomes. Dans un premier temps, il est intéressant d'utiliser un dictionnaire de Gabor composé de deux tailles de fenêtre différentes, une petite de 64 échantillons par exemple et une grande de 256 échantillons par exemple. Cela permettra de détecter les différentes composantes du signal analysé ainsi que ses caractéristiques.

• Représentation graphique

On représente la décomposition en atomes avec Matching Pursuit fait avec les paramètres suivants :

- Le son est un son de glockenspiel de 6 secondes
- Le critère d'arrêt était le nombre d'itération fixé à 100 000.
- Les atomes sont des atomes de Gabor de taille : 32,128,256,32768 échantillons représentant respectivement 0.7 ; 2.9 ; 5.8 ; et 743 ms.



Remarque 4.11 *Il est intéressant de noter ici que la MPTK nous permettrait de sélectionner les atomes de petites tailles qui ont été choisis par l'algorithme de Matching Pursuit et de les représenter. En effet, une fois une première approximation faite, on peut sélectionner les atomes qui ont été choisis pour la reconstruction.*

4.4.5 Conclusion et bilan

• Avantages.

Un des avantages est entre autre l'existence de la MPTK qui facilite considérablement l'utilisation de l'algorithme. On peut très facilement changer les atomes présents dans le dictionnaire et donc mettre en évidence certaines caractéristiques du signal selon ce que l'on sait a priori sur le signal utilisé.

• Inconvénient.

La représentation temps-fréquence du Matching Pursuit est pratique car elle fait apparaître les principales caractéristiques temps-fréquence du signal étudié. Cependant, comme nous pouvons le voir, tout le plan temps-fréquence n'est pas rempli par les atomes et nous restons sans informations sur les parties du plan qui n'ont pas été analysées par les atomes sélectionnés. Par ailleurs, il est donc impossible de juger de la parcimonie de la représentation dans ces conditions.

5 Proposition alternative aux méthodes existantes.

5.1 Introduction

Nous avons expliqué dans la section 3.2.6 en quoi le critère entropique pouvait être considéré comme un critère de parcimonie. Nous allons dans la suite détailler un algorithme utilisant ce critère.

On présente dans cette partie un algorithme décrit dans l'article d'Ingrid Daubechies, Michel Defrise et Christine De Mol "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint"[9]. En effet, on peut trouver de nombreux avantages d'une représentation intégrant ce critère de sélection et nous verrons cela en détail dans ce qu'il va suivre.

5.2 La théorie

Nous présentons brièvement ici des notions qui seront nécessaires pour comprendre l'algorithme décrit dans cette section.

5.2.1 Représentation parcimonieuse

On utilisera dans la suite les représentations parcimonieuses minimisant un critère de type l^τ . On les définit de la façon suivante :

Définition 5.1 On définit une représentation parcimonieuse pour un critère de type l^τ par

$$C_{\mathbf{x}}^\tau = \arg \min_{C_{\mathbf{x}} | C_{\mathbf{x}} \phi = \mathbf{x}} \sum_{k=1}^K |c_x(k)|^\tau$$

5.2.2 Les approximations parcimonieuses

On peut également définir ce qu'on appelle une approximation parcimonieuse en intégrant la représentation (à savoir le dictionnaire ϕ) à la minimisation du critère de parcimonie.

Définition 5.2 Une approximation parcimonieuse est donnée par

$$C_{\mathbf{x}}^{\tau, \lambda} = \arg \min \left\{ \|\mathbf{x} - C_{\mathbf{x}} \phi\|_F^2 + \lambda \sum_{k=1}^K |c_x(k)|^\tau \right\}$$

où $\|\mathbf{x}\|_F = \sum_p |x_p|^2$ désigne la norme de Fröbenius et où ϕ représente le dictionnaire.

Remarque 5.3 Le paramètre τ donne différents critères de parcimonie.

5.3 Les étapes de l'algorithme

On présente dans cette partie différentes étapes importantes de l'algorithme du point de vue théorique.

5.3.1 Données d'entrée

On se donne au départ, un signal x , un fonction de seuillage que l'on détaillera plus bas dans le cas $\tau = 1$, et un dictionnaire qui comprendra tous les atomes de la décomposition. Dans notre cas, il est intéressant de considérer des atomes de Gabor de plusieurs tailles, pour une première approche, deux tailles suffisent.

On se donne aussi une matrice c quelconque qui servira dans le déroulement de l'algorithme.

5.3.2 Seuillage

On effectuera une certaine opération sur les coefficients mais une étape particulièrement importante est celle du seuillage que nous détaillons ici.

Seuillage itératif pour $1 \leq \tau \leq 2$.

Pour $1 \leq \tau \leq 2$, des méthodes itératives efficaces ont été proposées. On s'est particulièrement penché sur celle développée dans [9]. Celle-ci repose sur une fonction de seuillage notée $S_{\mu,\tau}$. Pour $\tau = 1$, cette fonction de seuillage est définie de la manière suivante :

$$S_{\mu,\tau} = \begin{cases} t + \frac{\mu}{2} & \text{si } t \leq -\frac{\mu}{2} \\ 0 & \text{si } |t| < \frac{\mu}{2} \\ t - \frac{\mu}{2} & \text{si } t \geq \frac{\mu}{2} \end{cases}$$

Nous pourrions discuter plus tard de l'influence du paramètre μ dans la définition de la fonction de seuillage.

5.3.3 La condition sur le dictionnaire

L'algorithme assure la convergence (paragraphe suivante) sous la condition sur le dictionnaire suivante :

$$\|\phi c\|_2^2 \leq \|c\|_2^2, \forall c$$

Cependant, lorsque l'algorithme est constitué d'atomes unité, ce qui sera le cas ici, on ne pourra avoir dans ce cas que

$$\|\phi c\|_2^2 \leq M \|c\|_2^2, \forall c(*)$$

pour un certain M fini.

Nous donnons ci-dessous l'algorithme dans chacun de ces deux cas.

5.3.4 Itération et convergence

Sous la condition ci-dessus, le théorème 3.1 de [9] nous assure la convergence de l'algorithme de calcul de $C_x^{\lambda,\tau}$ à partir d'une initialisation arbitraire $c^{(0)}$ vers un minimiseur. Il s'agit d'itérer l'étape de seuillage suivante sur m :

$$c^{(m)} = S_{\lambda,\tau} \left(c^{(m-1)} + (x - c^{(m-1)}\phi)\phi^H \right)$$

où on a défini la fonction de seuillage $S_{\lambda,\tau}$ par $S_{\lambda,\tau}(c) = \left(S_{\lambda,\tau}(c(k)) \right)_{1 \leq k \leq K}$

Sous la condition (*), il sera nécessaire de modifier légèrement l'algorithme :

$$c^{(m)} = S_{\frac{\lambda}{M}, \tau} \left(c^{(m-1)} + \frac{(x - c^{(m-1)}\phi)\phi^H}{M} \right)$$

De plus, si ϕ est une base (ce qui sera le cas ici), nous savons qu'alors le minimiseur est unique.

5.3.5 Les paramètres

Il sera intéressant dans la suite d'étudier les résultats de l'algorithme en faisant varier certains paramètres. Nous étudierons :

- comment il est possible de trouver ou de fixer le paramètre M de manière correcte et quelle est son influence sur l'algorithme.
- comment fixer judicieusement la valeur du seuil μ définie par la variable λ
- comment choisir les tailles de fenêtre de manière à ce qu'elles détectent correctement les différentes composantes du signal.

5.3.6 Principe/Rapport avec une représentation multicouche.

On peut considérer que cet algorithme va réaliser une représentation multicouche comme celle présentée précédemment. En effet, on espère que l'étape de seuillage va permettre dans chaque représentation de garder les coefficients importants donc représentant bien le signal et de négliger ceux qui ont une faible valeur. Ainsi, cet algorithme tel qu'il est décrit sélectionne pour chaque couche du plan temps-fréquence les composantes qui sont bien localisées et dont la représentation met le mieux en évidence l'énergie présente dans cette composante.

5.4 Implémentation

Le théorème présenté ci-dessus assure la convergence de l'algorithme. Cependant, celui-ci présenté sous cette forme n'assure pas une charge de calcul raisonnable. En effet, la construction du dictionnaire implique de pouvoir représenter pour plusieurs tailles de fenêtres tous les atomes du plan temps fréquence. D'autre part, la taille d'une trame n'étant pas fixée le signal doit être stocké dans un vecteur et non dans une matrice ce qui allourdi davantage la charge de calcul. Il a donc fallu adapter et "traduire" l'algorithme pour permettre une implémentation rapide et efficace de celui-ci.

5.4.1 Du dictionnaire à la tfct

L'utilisation du dictionnaire pour faire l'analyse du signal n'étant pas totalement satisfaisant en terme de coût de calcul, il semblait intéressant d'utiliser la tfct (Transformée de Fourier à Court Terme) pour faire le calcul.

L'algorithme initial :

$$c^{(m)} = S_{\frac{\lambda}{M}, \tau} \left(c^{(m-1)} + \frac{(x - c^{(m-1)}\phi)\phi^H}{M} \right)$$

peut d'abord s'écrire sous forme matricielle avec ϕ_1 et ϕ_2 les matrices comprenant respectivement les atomes de l'analyse avec une "petite" fenêtre d'analyse W_1 et avec une "grande fenêtre" W_2 :

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

$$c = (c_1 c_2), \text{ alors } c\phi = (c_1\phi_1 + c_2\phi_2)$$

est traduit avec des tfct et des itfct (transformée de Fourier à court terme inverse) de la façon suivante :

A l'itération m , on a

$$- c_1^{m-1} \text{ et } c_2^{m-1}$$

- On note $r = x - \text{itfct}(c^{m-1}, W_1) - \text{itfct}(c^{m-1}, W_2)$ le résidu.

- On effectue l'opération

$$S(c_1^{m-1} + \text{itfct}(r, W_1)) = c_1^m$$

$$S(c_2^{m-1} + \text{itfct}(r, W_2)) = c_2^m$$

où S est la fonction de seuilage.

$$- \text{on obtient } c^{(m)} = [c_1^{(m)} c_2^{(m)}]$$

5.4.2 La constante du dictionnaire

Comme nous l'avons vu précédemment, pour que l'algorithme converge, il est nécessaire de respecter une condition sur le dictionnaire. Notons que nous n'utilisons pas de dictionnaire dans notre implémentation mais deux fenêtres. Il est donc nécessaire de traduire cette condition sur le dictionnaire dans le cadre de notre utilisation. Pour cela nous avons deux possibilités :

- La première est d'écrire le dictionnaire correspondant à l'analyse faite avec deux tfct de tailles différentes, puis de vérifier directement la condition sur le dictionnaire. En effet, on peut construire une matrice possédant tous les atomes de Gabor de deux tailles différentes pour tous les temps et toutes les fréquences présentent dans le signal. Il est alors possible en diagonalisant cette matrice de trouver la constante grâce aux valeurs propres de celle-ci.

- La seconde est de traduire la condition dans le cas d'une tfct et d'en déduire une méthode pour trouver la constante M à partir des deux tailles de fenêtres.

Si c'est en premier lieu sur la première possibilité que nous nous sommes penchés, face aux difficultés diverses lors de la construction de la matrice c'est la seconde qui a été retenue.

Rappelons qu'il s'agit de trouver M assez grand et le plus petit possible vérifiant $\|\phi c\|_2^2 \leq M \|c\|_2^2, \forall c$

Ainsi, en notant $(\phi_k)_t = (w(t)e^{2i\pi \frac{kt}{N}})_t$, (où la fonction w est évidemment la fenêtre), avec N le nombre d'échantillons de la taille de la fenêtre en fréquence, on a dans le cas d'une TFD :

$$\|\phi c\|_2^2 = \|\sum_{k=0}^{N-1} w(t) c_k e^{2i\pi \frac{kt}{N}}\|_2^2 \leq \|w\|_\infty^2 \|\sum_{k=0}^{N-1} c_k e^{2i\pi \frac{kt}{N}}\|_2^2 \leq \|w\|_\infty^2 N \|c\|_2^2$$

Etant donné que ϕ est en réalité composé de ϕ_1 et de ϕ_2 , nous devons nous placer dans le cas de deux stft. D'une part, on peut considérer que si $M = \|w\|_\infty^2 N$ est correct dans le cas d'une TFD, on aura la même valeur de M dans le cas d'un stft. D'autre part dans notre cas où il y a deux stft dans l'analyse on obtiendra une valeur tout aussi correcte en prenant le max pour chacune des stft. On obtient donc une bonne estimation du plus petit M avec $M = \max(N_1 \|w_1\|_\infty^2, N_2 \|w_2\|_\infty^2)$.

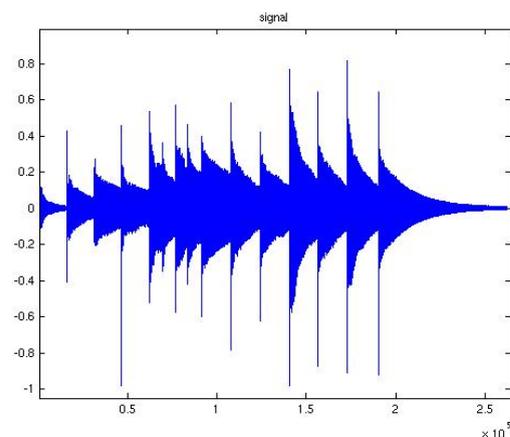
5.4.3 Itération et critère d'arrêt

Nous avons choisis dans les expériences qui vont suivre de mettre un critère d'arrêt sur le nombre d'itérations. On pourrait également mettre comme critère une erreur minimale.

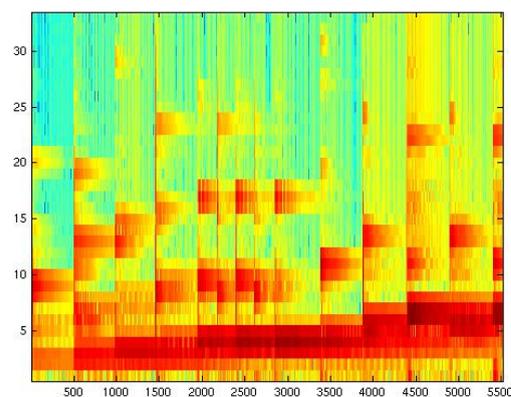
5.4.4 Représentation des couches de l'algorithme

On représente ci-dessous le signal et les représentation temps-fréquence à l'aide de deux fenêtres différentes. Nous attendons des différentes couches qu'elles mettent en évidence les transitoires visibles dans la première représentation dans une couche et dans une seconde couche les différentes harmoniques visibles dans la seconde.

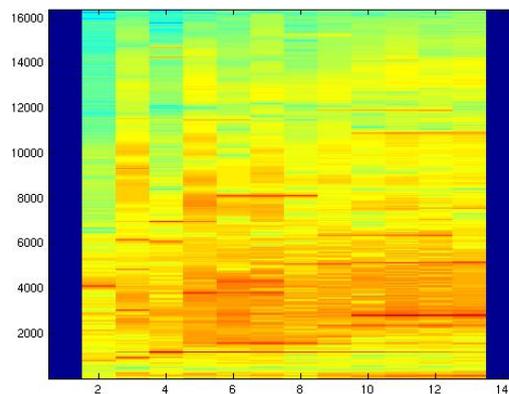
Représentation du son de glockenspiel.



Représentation de l'analyse avec une fenêtre de 64 échantillons.



Représentation de l'analyse avec une fenêtre de 32768 échantillons.



On peut ainsi voir sur les deux figures précédentes que ce sont des composantes différentes qui apparaissent clairement dans chacun des cas. On voit sur la première les transitoires alors que la seconde met davantage en évidence les harmoniques. Nous devons préciser ici que ces deux tailles de fenêtres ont été choisies à partir des caractéristiques du signal. On a pu en effet remarquer que les transitoires avaient une taille d'environ 60 échantillons alors que les composantes quasi-stationnaires avaient une durée d'environ 32 000 échantillons.

Il s'agit de mettre en évidence ces deux composantes dans les différentes couches de l'algorithme. On représente ci-dessous les représentations obtenues à l'aide de l'algorithme décrit plus haut.

5.5 Evaluation

Les critères possibles pour évaluer les qualités d'une représentation temps-fréquence peuvent être divers. Pour pouvoir évaluer correctement ce travail, nous avons défini une évaluation valide selon les objectifs fixés au départ.

Par ailleurs, il a semblé particulièrement important et intéressant de se restreindre à une application particulière pour être en mesure de juger de la fiabilité ainsi que de l'utilité d'un tel algorithme.

Enfin nous pourrions conclure sur les avantages et limites de l'algorithme ici développé et évalué.

5.5.1 Paramètres et critères des expériences

• Les paramètres

Rappelons quels sont les différents paramètres de l'algorithme :

- le seuil : Il s'agit du paramètre λ de la partie 5.4.1 qui définit le seuil (voir définition de la fonction seuil 5.3.2 où $\tau = \frac{\lambda}{M}$. Décrivons rapidement comment on le calcule dans l'algorithme :

- On regarde quel est la valeur absolue du coefficient maximum sur la représentation temps-fréquence (tfct) avec la petite fenêtre.
- On fait de même avec la représentation temps-fréquence (tfct) faite avec la grande fenêtre.
- On calcul le maximum de ces deux valeurs
- On la multiplie par un coefficient appelé "coefficient seuil". C'est ce coefficient qu'on fera varier entre 0.2 et 0.8 dans les expériences suivantes
- les tailles de fenêtres. Il s'agit de prendre deux tailles de fenêtre qui permettront de "capter" les composantes caractéristiques du signal. Comme on le verra, les tailles de fenêtre que l'on choisit pour cet algorithme peuvent avoir un impact non négligeable sur les qualités de l'algorithme. On étudiera donc leur influence.

•Les critères

Nous évaluerons l'erreur quadratique. (On pourra se référer à la section 5.4.1). Il s'agit de la quantité :

$$E_m = \frac{\|x - \text{itfct}(c^{m-1}, W_1) - \text{itfct}(c^{m-1}, W_2)\|^2}{\|x\|^2}$$

Nous proposons donc d'évaluer cet algorithme selon les critères suivants :

- Evaluation de la convergence de l'erreur selon les différents paramètres.
- Evaluation de la parcimonie de la représentation pour chacune des couches en fonction des différents paramètres de l'algorithme.
- Evaluation de l'algorithme pour une application à la détection de transitoire sur un signal réel.

Toutes les évaluations qui seront présentées ont été réalisées sur un son de glockenspiel, instrument possédant des différences très importantes entre les caractéristiques temps-fréquence de la phase de maintien et celles de la phase d'attaque, et qui nous permettra donc d'évaluer les qualités de l'algorithme pour "séparer" ces deux phases dans ce cas simple.

5.5.2 Evaluation de la convergence

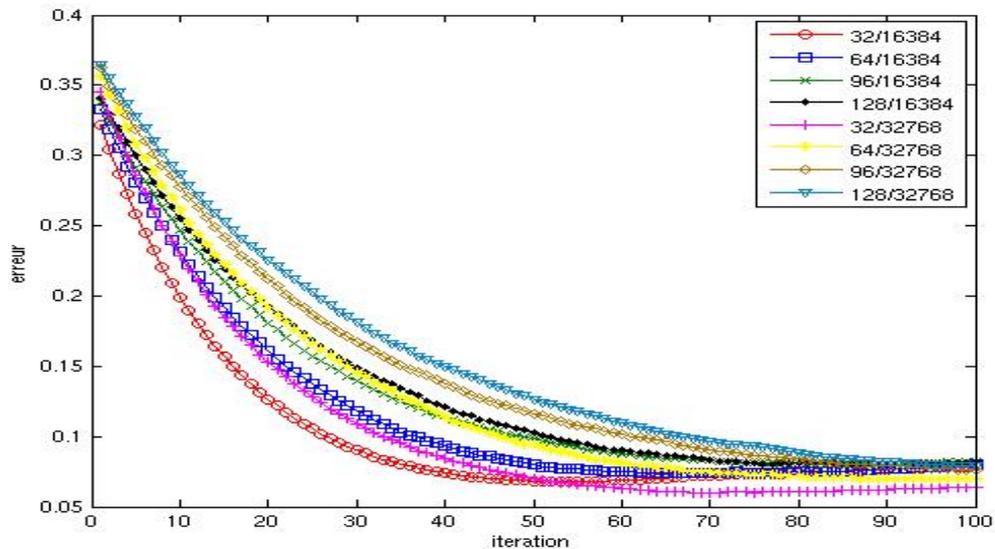
Présentation de l'expérience

On étudie les convergence de l'algorithme décrit en (5.3 et 5.4), à savoir sa vitesse de convergence en fonction des différentes valeurs du seuil et de la taille de fenêtres.

Nombre d'itérations nécessaires à la convergence selon la taille des fenêtres.

On regarde ensuite la rapidité de la convergence pour un seuil fixé et pour plusieurs tailles de fenêtres

Figure : Représentation de l'erreur en fonction du nombre d'itération pour différentes tailles de fenêtres.



Conclusion

On peut remarquer que deux tailles de fenêtres donnent une "bonne convergence". Il s'agit des fenêtres de taille 32/16384 et 32/32768. Comme nous l'avons déjà fait remarquer celle-ci correspond à la tailles des différentes composantes.

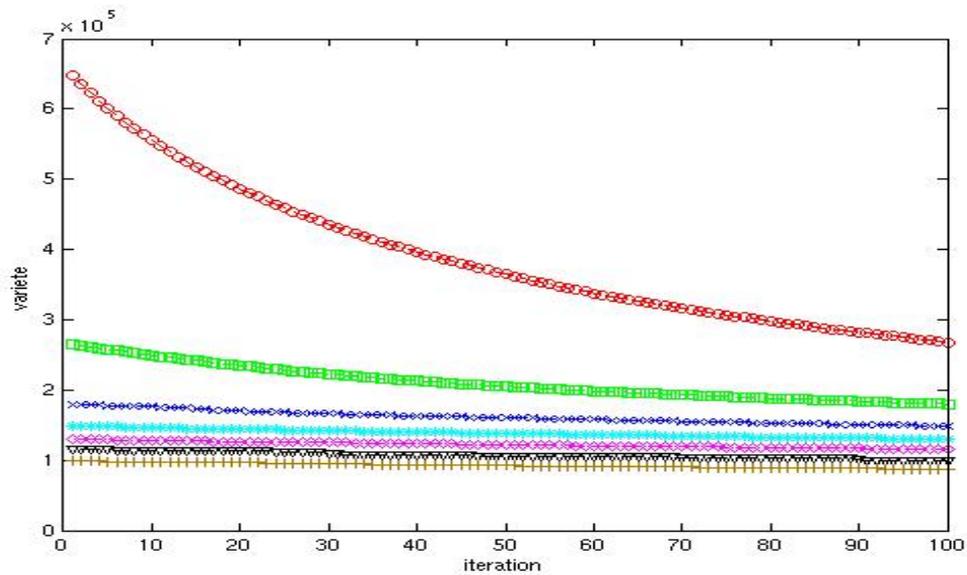
5.5.3 Evaluation de la parcimonie

Nous calculons ici la parcimonie au sens de la norme l^0 . Il s'agit donc de savoir si les représentations obtenues sont bien parcimonieuses comme on le souhaite.

Pour évaluer la parcimonie au sens de la norme l^0 , on calcule le nombre de coefficients non nuls dans la représentation temps-fréquence. Etant donné que nous avons deux représentations temps-fréquence, nous sommes les résultats obtenus pour chacune d'entre elle. Par ailleurs, remarquons que pour le calcul qui est ici fait, plus la valeur est importante (il y a beaucoup de coefficients non nuls) moins bonne sera la parcimonie, et inversement, plus le critère sera faible (il y a peu de coefficients non nuls) meilleure sera la parcimonie. En toute logique, nous n'avons donc pas nommé ce critère parcimonie mais plutôt utilisé le terme de "variété".

Nombre d'itérations nécessaire à la convergence de la parcimonie

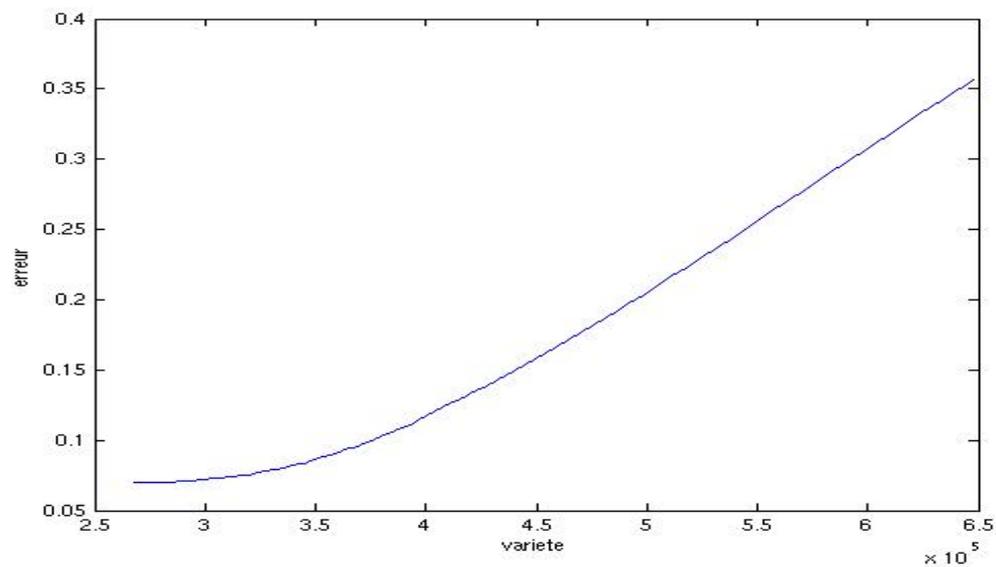
Nous évaluons ici la convergence de la parcimonie en fonction du nombre d'itération de l'algorithme. On espère que la parcimonie va augmenter (donc la variété diminuer) avec le nombre d'itération.

Figure :Parcimonie en fonction des itérations.

Conclusion On constate que la parcimonie augmente quelque soit le seuil ce qui est bien le résultat attendu de l'algorithme.

Parcimonie au sens l^0 en fonction de l'erreur

On évalue ici la parcimonie en fonction de l'erreur. Rappelons que nous souhaitons que la parcimonie augmente (donc que la variété décroisse) avec l'erreur.

Figure :Parcimonie en fonction de l'erreur

Conclusion

On constate que plus l'erreur est faible, plus la parcimonie l'est également ce qui encourage à choisir des valeurs de seuils pour lesquelles l'erreur diminue.

5.5.4 Application à la détection de transitoire

On évalue dans cette partie la performance de l'algorithme à une application particulière : la détection de transitoire. En effet, on espère que la couche de la petite fenêtre va donner les composantes transitoires alors que la couche de la grande fenêtre donnera les composantes harmoniques. Comme depuis le début, on effectuera ces expériences sur un son de glockenspiel.

Nous représentons ci-dessous, le signal de glockenspiel, le signal reconstruit à partir de la grande fenêtre puis de la petite et enfin les représentations temps-fréquence de ces trois sons.

Figure :Représentation du son de glockenspiel.

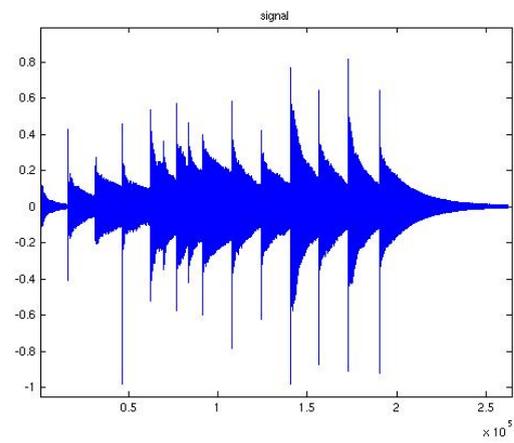


Figure :Représentation du son reconstitué de la grande fenêtre.

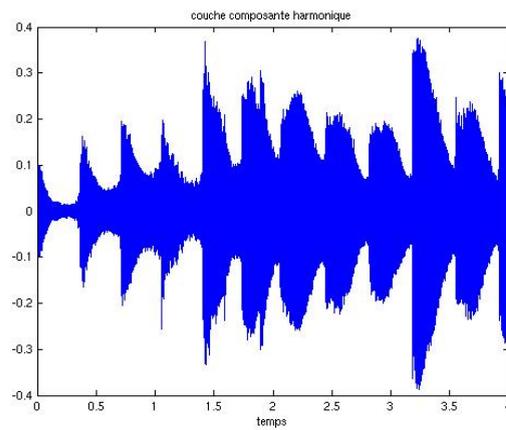


Figure :Représentation du son reconstitué de la petite fenêtre.

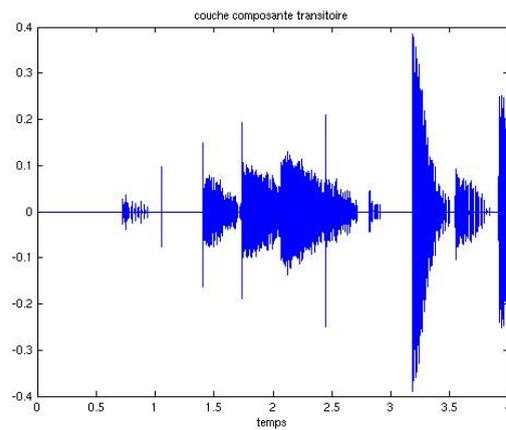


Figure :Représentation de la couche avec la grande fenêtre.

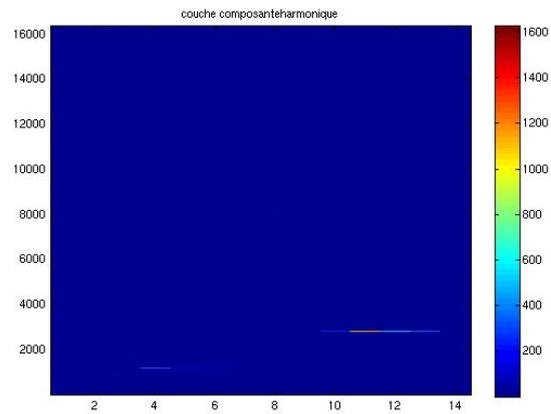


Figure :Représentation de la couche avec la grande fenêtre zoomée.

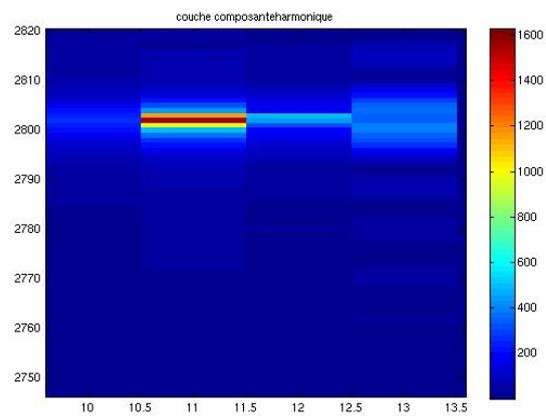
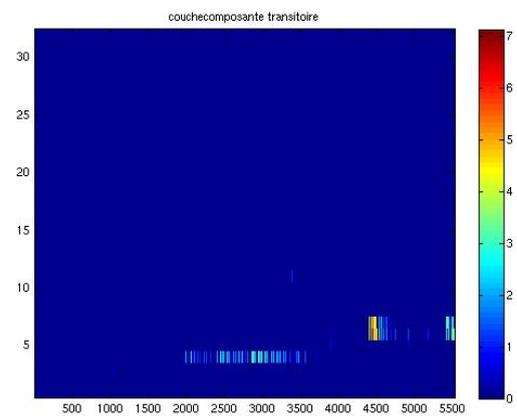


Figure :Représentation de la couche avec la petite fenêtre.



Remarque 5.4 *S'il est évident que les représentations obtenues sont bien parcimonieuses, il est n'est pas possible de bien localiser l'énergie sur la représentation sans zoomer la figure.*

- **Conclusion**

On peut bien distinguer que la couche de la petite fenêtre reconstruit essentiellement les phases d'attaque du signal et que la couche de la grande fenêtre reconstruit les harmoniques du signal. On présente ici deux représentations temps-fréquence dans le cas de la grande couche la seconde représentation étant le zoom sur une des composante de la première.

L'algorithme ici implémenté et décrit produit bien l'effet escompté sur ce signal et pour cette application à la détection d'attaque.

5.5.5 Conclusion sur les résultats

On déduit de ces évaluations que l'algorithme implémenté produit bien une analyse multi-couche dans le sens où il permet d'analyser les différentes composantes du signal et de les séparer.

Nous avons étudié l'évolution de la parcimonie, de l'erreur en fonction des différents seuils et des différentes tailles de fenêtres. Nous avons également étudié son efficacité en étudiant son utilisation dans le cas de la détection de transitoire sur un son de glockenspiel. Nous avons ainsi pu constater que les représentations étaient bien parcimonieuses et que chaque couche représente bien un type de composante.

5.5.6 Amélioration et recherches envisagées

Il serait intéressant pour la suite et fin du stage de prolonger cette étude :

Dans un premier temps, il sera intéressant de tester ce même algorithme sur d'autres instruments que le glockenspiel qui présentait des caractéristiques particulières pour cette étude.

En vue d'une amélioration de l'algorithme on pourra essayer d'utiliser des seuils différents selon les couches et étudier si cela produit une amélioration des performances de l'algorithme comme on pourrait s'y attendre.

Enfin, on peut également penser à utiliser cet algorithme à d'autres applications que la détection de transitoire, comme par exemple la séparation de source.

6 Conclusion et perspectives

6.1 Conclusion

Nous avons dans un premier temps présenté les définitions et propriétés les plus fondamentales sur les représentations temps-fréquences de type Fourier ou ondelettes. Nous avons ainsi compris en quoi elles n'étaient pas suffisamment satisfaisantes pour traiter des signaux sonores et pourquoi nous devons nous tourner vers des représentations avec résolution adaptatives.

Dans un deuxième temps, nous avons soulevé le problème du critère de sélection de la résolution selon les caractéristiques locales du signal. Nous avons fait le bilan des types de critères existant dans l'état de l'art actuellement, de la fonction d'ambiguïté aux critères entropiques et de parcimonie. Ainsi, il a été possible d'une part, de comprendre leur rôle dans les méthodes présentées dans la section suivante, d'autre part, d'évaluer les avantages et inconvénients de chacun d'entre eux pour être en mesure de faire un choix pour un apport à l'état de l'art.

Ensuite, nous avons fait la synthèse et l'analyse des représentations qui présentent une résolution variable. Nous avons ainsi présenté plusieurs transformées, représentations temps-fréquence ou algorithme reliés ou non à un critère de sélection. Plus précisément, nous avons parlé de l'algorithme de meilleure base, des transformées covariantes par translation temporelle ou fréquentielle, nous avons également présenté un algorithme effectuant une représentation multicouche à l'aide de plusieurs fenêtres. Enfin, nous avons présenté l'algorithme de Matching Pursuit, en quoi et comment il pouvait être une solution au problème que nous nous posons. Dans chacun de ces cas, après avoir effectué une synthèse des points importants de la représentation ou de l'algorithme, nous nous sommes attachés à faire l'analyse des qualités de celle(celui)-ci, pour ensuite faire le bilan des avantages, inconvénients de chaque méthode. Cette synthèse et analyse était en effet fondamentale pour être en mesure de proposer une méthode alternative à celles actuelles.

Enfin, grâce à l'étude faite précédemment, nous avons été en mesure de proposer une méthode présentant de nombreux avantages autant en terme de résolution du problème posé qu'en terme de coût de calcul ou de diversité des applications. Nous avons présenté l'algorithme du point de vue théorique en détaillant chacune des étapes importantes, adapté cet algorithme à une implémentation efficace et expliqué ses atouts et particularités. Nous en sommes venu à proposer une évaluation de l'algorithme selon les critères de convergence et de parcimonie en fonction des paramètres de l'algorithme (seuil, taille des fenêtres et constante du dictionnaire). Ceci nous permet d'expliquer les améliorations et recherches envisagées pour la fin de ce stage.

6.2 Perspectives

Nous avons déjà présenté les recherches envisagées pour la fin du stage à savoir l'étendue des utilisations possibles de cet algorithme à diverses applications et signaux sonores aux propriétés diverses. Il pourra être également intéressant de démontrer le théorème utilisé dans le cas multicanal car cela permettrait probablement d'un point de vue pratique un avancé importante des qualités de l'algorithme en terme par exemple de séparation de source.

D'autre part, à l'issue de ce travail les outils nécessaires à un bilan final sur les algorithmes de ce type, leurs apports et leurs limites seront constitués. A ce moment, il sera judicieux d'évaluer les possibilités en terme d'avancées scientifiques de cette approche et de réfléchir à d'autres possibilités.

Références

- [1] F. Auger, P. Flandrin. Improving the Readability of Time-Frequency Time-Scale Representations by Reassignment Method. *IEEE SP*, Vol.43, No.5, 1995, pp.1068-1089.
- [2] F.Auger ,P. Flandrin, P.Goncalves,O. Lemoine. *Time Frequency Toolbox*. 1995-1996.CNRS.Isis.
- [3] Cohen. *Time-Frequency analysis*. Prentice-Hall Signal Processing.1997.
- [4] M. Bayram and R. Baraniuk. Multiple Window Time-Varying Spectrum Estimation. *Conf. Info. Sci. and Sys. (CISS)*, vol. 30, Mar. 1996.
- [5] R. G. Baraniuk et Augustus G. Baraniuk J. E. M. Janssen. Measuring Time-Frequency Information Content Using the Rényi Entropies. *IEEE Trans. on Information Theory*, 1998.
- [6] E. Chassande-Mottin, I. Daubechies. Differential Reassignment. *IEEE Signal Proc. Lett.*, 4 :293-294, 1997
- [7] R.Coifman, Mladen Victor Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theory*, 38(2), March 1992.
- [8] Corson. 1999. Mémoire de DEA ATIAM. Etude de la décomposition des sons en sinusoides plus bruit.
- [9] I.Daubechies, M.Defrise, C.De Mol. An iterative thresholding algorithm for linear inverse problemes with a sparsity constraint. *Soumis*
- [10] M. Dörfler.*Time-frequency Analysis for Musical Signal. A Mathematical Approach*.
- [11] P.Flandrin. Some features time-frequency representations of multicomponents signals *ICASSP 84*.
- [12] Patrick Flandrin. *Temps-fréquence*. Editions Hermès.1993.
- [13] Rémi Gribonval. *Approximations non-linéaires pour l'analyse des signaux sonores*. Thèse de doctorat.2000.
- [14] Claude Gasquet et Patrivk Witomski.*Analyse de Fourier et applications*.Dunod.2004
- [15] Gilles Gonon. Thèse de doctorat.Proposition d'un schéma d'analyse/sytnèse adaptatif dans le plan temps-fréquence basé sur des critères entropiques. Application au codage audio par transformée. 13 juin 2002.

-
- [16] Michael Mark Goodwin. Adaptive signal models : theory, algorithms, and audio applications. Ph. D. dissertation, Memorandum no. UCB/ERL M97/91, Electronics Research Laboratory, College of Engineering,
- [17] F.Hlawatsch et G.F. Boudreaux-Bartels. Linear and Quadratic Time-Frequency Signal Representations. Avril 1992. IEEE SP Magazine
- [18] Oliver Johnson et Christophe Vignat. Some results concerning maximum Rényi entropy distributions. 19-07-2005.Soumis
- [19] Douglas L. Jones and Richard G. Baraniuk. A simple scheme for adapting time-frequency representations. IEEE Trans. Signal Processing, vol. 42, pp. 3530–3535, Dec. 1994.
- [20] Douglas L. Jones and Richard G. Baraniuk. An adaptive optimal-kernel time-frequency representation. IEEE Transactions on Signal Processing, October 1995.
- [21] Florent Jaillet et Bruno Torrèsani. Time-Frequency Jigsaw Puzzle : Adaptive multiwindow and multilayered Gabor expansions. Prépint.Soumis.
- [22] Florent Jaillet et Bruno Torrèsani. Remarques sur l’adaptativité des représentations temps-fréquence. Proceedings of the GRETSI’03, 19^{ème} colloque sur le traitement du signal et des images, Paris, 8-11 septembre, 2003.
- [23] Florent Jaillet. Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore. Thèse de doctorat.2005.
- [24] A. Jourjine, S. Richard, and O.Yilmaz, Blind separation of Disjoint Orthogonal Signals : Demixing N sources from 2 Mixtures, IEEE Conference on Acoustics, Speech and Signal Processing.
- [25] Sacha Krstulovic, Rémi Gribonval. The Matching Pursuit Tool Kit User Manual and Developer’s Notes.15 Mai 2006.
- [26] Franz Hlawatsch et François Auger (sous la direction de). Temps-fréquence, concepts et outils. Editions Hermès. 2005
- [27] Stéphane Mallat et Z.Zhang. Matching Pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing, 41(12), December 1993
- [28] Stéphane Mallat. Une exploration des signaux en ondelettes. Les éditions de l’école polytechnique.2000
- [29] Yves Meyer. Les ondelettes, algorithmes et applications. Editions Armand Colin.1992.

-
- [30] Axel Roebel, Miroslav Zivanovic, and Xavier Rodet. Signal decompositions by means of classifications of spectral peaks. International Computer Music Conference (ICMC), Miami, 2004
- [31] Eva Wesfreid et Mladen Victor Wickerhauser. Adapted local Trigonometric Transform and Speech Processing. IEEE Trans. SP, Vol. 41, No. 12, pp. 3596–3600, Dec. 1993.
- [32] K. Zyczkowski. Rényi extrapolation of Shannon Entropy. Open Systems and Information Dynamics, Volume 10, Number 3, September 2003, pp. 297-310(14)
- [33] Werner Kozek. Matched Weyl-Heisenberg Expansions of Nonstationary Environments. Phd an der technischen universität Wien Fakultät für Elektrotechnik. Septembre 1996.