



HAL
open science

Deux approches de segmentation temps-fréquence : détection par modèle statistique et extraction de contour par le champ de vecteurs de réallocation

Fabien Millioz

► **To cite this version:**

Fabien Millioz. Deux approches de segmentation temps-fréquence : détection par modèle statistique et extraction de contour par le champ de vecteurs de réallocation. Traitement du signal et de l'image [eess.SP]. Institut National Polytechnique de Grenoble - INPG, 2009. Français. NNT : . tel-00421599

HAL Id: tel-00421599

<https://theses.hal.science/tel-00421599>

Submitted on 2 Oct 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT POLYTECHNIQUE DE GRENOBLE

N° attribué par la bibliothèque

THÈSE

pour obtenir le grade de

Docteur de L'Institut Polytechnique de Grenoble

Spécialité : « Signal, Image, Parole, Télécommunications »

préparée au laboratoire Grenoble Image Parole Signal et Communication

dans le cadre de l'**École Doctorale EEATS**

« Électronique, Électrotechnique, Automatique, Télécommunications, Signal »

présentée et soutenue publiquement

par

Fabien Millioz

le 16 septembre 2009

Titre :

Deux approches de segmentation temps-fréquence : détection par modèle statistique et extraction de contour par le champ de vecteurs de réallocation

Directrice de thèse : Nadine Martin

JURY

Monsieur	Jocelyn CHANUSSOT,	Président
Monsieur	Petar DJURIĆ,	Rapporteur
Monsieur	Patrick FLANDRIN,	Rapporteur
Madame	Nadine MARTIN,	Directrice de thèse
Monsieur	Pascal LARZABAL,	Examineur
Monsieur	Jean-Philippe OVARLEZ,	Examineur

*With magic, you can turn a frog into a prince.
With science, you can turn a frog into a Ph.D
and you still have the frog you started with.*

Terry PRATCHETT

Remerciements

Voici l'instant de mettre la touche finale à ce manuscrit et à quelques années de thèse, par la traditionnelle page de remerciements.

Un premier merci à Petar Djurić et à Patrick Flandrin, qui m'ont fait l'honneur d'accepter de relire ce présent document, ainsi qu'à Pascal Larzabal et à Jean-Philippe Ovarlez, pour avoir participé au jury en tant qu'examineurs. Un autre merci à Jocelyn Chanussot pour avoir présidé à la soutenance, avec toujours la classe et le chic qui lui sont bien particuliers.

Ces travaux de thèse ont eu lieu dans un premier temps au LIS, qui fusionna avec le LAG et l'ICP pour devenir le GIPSA-lab. De nombreuses personnes ont contribué directement ou indirectement à ce document, pardon d'avance à ceux que je vais hélas oublier.

Pour le côté scientifique, je me dois de remercier Julien Huillery, Cédric Gérot, Jean-Louis Lacoume, Christine Servière et Pierre Granjon. Merci aux stagiaires que j'ai encadré, à savoir Simon Thomas et Guillaume Linet.

Pour les autres côtés, un merci tout particulier à mes cobureaux présents et passés, à savoir Hussein, Lionel, Ludo, Mohammed, Bertrand. Merci pour toutes ces discussions, ces idioties, cette sympathie, mais merci aussi pour les moments de calme.

Merci à tous les autres, permanents, thésards présents ou déjà partis, pour leur contribution à l'ambiance de ce laboratoire. En vrac : Jérôme, Jocelyn, François (C.), Nicolas (LB.), Pierre-Olivier, Daniel, Éric, Christines (F. et S.), Steeve, Barbara, Cléo, Cyrille, Geneviève, Laurents (R. et H.), Jean-Marc (B., S. et C.), Glenn, Isabelles (M. et C.), Michels (G. et D.), Gabriel, Christian, Joël, Stéphane, Florent, Francis, Olivier, Lucia et Dinu d'une part ; et merci à Manuel, Moussa, Maxime, Geoffrey, Cédric (D.), Zaher, Mathieu et Matthieu, Tarek, Grégoire, Jérémy, Meryem, Frédéric, Sebastian, Gaël, Benoît, Nuttapol, Sébastien, Caroline, Julien (B.), Loïc, Cédrics (G.-P. et R.), Olivier, Ion, Nicolas (J.), Ali, Murtaza, Jérôme, Salem, Yulia, Silvia, Alberto, Ion, Christian, Rosilde, François (D.), Sophie, Cyrille, Lionel, Bastien, Benjamin, Jordi. Pour que personne ne soit jaloux sans raison, je tiens à préciser que l'ordre des noms ne porte aucun sens.

En dehors du département image et signal, je remercie également Hayate, Ahmad, Pascal, Christian, Sylvie, Nicolas (A.), Alexis, Alexandre, Sylvain, Florent, Antoine, Anne.

Merci également aux étudiants de l'ENSIEG puis de l'ENSE³ qui m'ont subi en tant qu'enseignant, vous avez indirectement contribué à ces travaux, en me distrayant régulièrement.

Merci à Lahire, Argine, Pallas, Lancelot, le 10 de carreau et les 27 autres cartes du jeu de coinche pour les pauses très appréciées du midi. Et merci aux très nombreux joueurs avec qui j'ai eu le plaisir de taper le carton.

Merci à mes relecteurs, Solène, Loïc et Bérenger, même s'ils n'ont pas osé relire les équations.

tions. S'il reste des fautes, ce sont les personnes à blâmer. Merci à Yannick d'avoir relu mon introduction, dommage que ton courage se soit épuisé juste après.

Merci à mes soutiens moraux et psychologiques, qu'ils soient de Savoie (Lionel, Jean-Jacques, Alex, Guillaume, et tous les autres dont la liste est bien trop longue – veuillez m'en excuser), B4 (Manuel, Yves-Laurent, Yohan et assimilés), ma famille, et tous ceux hors catégorie, qui se reconnaîtront.

Merci également à Linus Torvalds, Richard Stallman, et toutes les personnes bien moins médiatiques qui contribuent au logiciel libre. J'en profite pour remercier mon éternel voisin de bureau, Jean-Marc Sache, ainsi qu'Hervé Coluasonno.

Et un dernier merci tout particulier à Nadine Martin, à qui cette thèse doit évidemment beaucoup, que ce soit par ses conseils avisés ou son indéfectible optimisme quand il est pourtant évident que rien ne fonctionne comme espéré. Merci.

Il paraît que les remerciements sont la partie la plus lue d'une thèse. Je t'invite, ami(e) lecteur(ice), à lire malgré tout quelques passages qui suivent.

Table des matières

Introduction générale	1
Objectif de ces travaux	2
Organisation du document	3
I Segmentation statistique du plan temps-fréquence	7
Introduction	11
1 Transformée de Fourier à Court Terme et caractérisation statistique	13
1.1 Loi de la TFCT d'un bruit blanc stationnaire gaussien	15
1.1.1 Transformée de Fourier à Court Terme « rectifiée »	20
1.2 Cas d'un signal quelconque noyé dans un bruit blanc stationnaire gaussien	23
2 Détection des points temps-fréquence contenant du signal	25
2.1 Test binaire d'hypothèses	25
2.1.1 L'approche bayésienne	26
2.1.2 L'approche de Neyman-Pearson	28
2.2 Détection dans la TFCT rectifiée	29
3 Estimation itérative du bruit	33
3.1 Estimation : principe et définitions	34
3.1.1 Estimation par maximum de vraisemblance dans le cas du bruit seul	35
3.1.2 Estimation par maximum de vraisemblance dans le cas signal et bruit	37
3.2 Critère d'arrêt de l'estimation itérative : le kurtosis	38
3.2.1 Statistiques d'ordre supérieur	38
3.2.2 Kurtosis de la TFCT rectifiée	42
3.2.2.1 Cas particuliers	44
3.2.2.2 Influence de $\alpha[n, k]$	45
3.2.2.3 Choix d'un seuil d'arrêt	46
3.3 Conclusion	48
4 Algorithme de segmentation	49
4.1 Croissance de région	49
4.2 Description de l'algorithme de segmentation	50
4.3 Influence des paramètres	52
4.3.1 Probabilité de fausse alarme p_{fa}	53
4.3.2 Proportion de candidats segmentés à chaque sous-boucle p_{cand}	54

4.3.3	Seuil sur le kurtosis des points non segmentés s_k	55
4.3.4	Influence du Rapport Signal à Bruit	56
4.4	Extensions du principe de segmentation	56
4.4.1	Bruit non blanc ou non stationnaire	56
4.4.2	Transformée en ondelettes	57
4.4.3	Capogramme	59
4.5	Applications	61
4.5.1	Débruitage	61
4.5.2	Séparation de sources	62
4.5.3	Caractérisation de signaux	64
Conclusion		67
II Segmentation du plan temps-fréquence par les vecteurs de réallocation		69
Introduction		73
5 Réallocation		75
5.1	Réallocation du spectrogramme, interprétation du centre de gravité	75
5.2	Réallocation du spectrogramme, interprétation de la fréquence instantanée et du retard de groupe	78
5.2.1	Fréquence instantanée et retard de groupe	78
5.2.2	Mise en œuvre de la réallocation	80
5.3	Champ de vecteurs de réallocation	80
6 Détection de contour		83
6.1	Discrétisation du champ de vecteurs de réallocation	84
6.2	Détection de zone frontière	85
6.2.1	Application sur un signal vibratoire	87
6.3	Construction du contour	87
6.3.1	Degré de parallélisme local	87
6.3.2	Algorithme de construction de contour	89
6.4	Cas de signaux bruités	91
6.5	Influence des paramètres de la TFCT	94
6.6	Pour améliorer l'algorithme...	95
6.6.1	Réallocation multi-fenêtres	96
6.6.2	Réallocation pseudo multi-fenêtres	97
6.6.3	Lissage du résultat obtenu par contour actif	98
Conclusion		101
Conclusions générales		103
	Discussion	105
	Perspectives	106

Introduction générale

En traitement du signal, la notion de temps est duale de celle de fréquence. Ainsi, un signal représenté dans le domaine temporel, c'est-à-dire dont l'évolution est présentée en fonction du temps, ne présente directement aucune de ses caractéristiques fréquentielles. Sa transformée de Fourier, décrit le contenu fréquentiel du signal, mais le contenu temporel devient alors sous-jacent. Le passage du temporel au fréquentiel ne change pas l'information contenue dans le signal, mais uniquement sa présentation. Classiquement, un signal composé d'une fréquence pure de durée infinie a besoin d'une durée infinie pour être décrit temporellement, tandis qu'un seul point de sa transformée de Fourier suffit pour le décrire dans le domaine fréquentiel.

Dans le cas des signaux stationnaires, dont les propriétés n'évoluent pas dans le temps, ce dualisme ne pose aucun problème : l'information fréquentielle est la même à chaque instant temporel. Dans le cas des signaux non-stationnaires, cette opposition n'est plus acceptable. Le domaine fréquentiel présente l'information contenue globalement dans le signal, sans le rattacher à un instant donné, tandis que deux instants différents du signal peuvent présenter des caractéristiques fréquentielles différentes.

Pour pallier ce problème, les traiteurs de signaux ont développé depuis plus de cinquante ans les représentations *temps-fréquence* [Coh89, Fla93], qui cherchent à représenter le signal conjointement en temps et en fréquence. L'exemple historique est le spectrogramme, défini comme le module carré de la transformée de Fourier à court terme, ou transformée de Fourier glissante. Le signal est temporellement tronqué en plusieurs segments successifs, qui subissent une transformée de Fourier. Ainsi, on dispose du contenu fréquentiel du signal correspondant à chaque segment temporel. Le spectrogramme étant le module carré d'une transformée de Fourier, il représente l'énergie du signal dans le temps, selon le segment analysé, et dans la fréquence, selon le coefficient de Fourier.

Plusieurs autres représentations existent. De manière équivalente au spectrogramme, le caponogramme estime le spectre du signal par la méthode de Capon sur un segment du signal non stationnaire.

Dans un autre registre, il existe des représentations dites quadratiques, telle la distribution de Wigner-Ville. Cette distribution présente de bonnes propriétés de localisation de l'énergie dans le plan temps-fréquence, mais présente des interférences se manifestant par des valeurs alternativement positives et négatives. La représentation de Wigner-Ville engendre par son lissage toute une classe de représentations temps-fréquence, dont le spectrogramme fait partie, appelée classe de Cohen.

Dans toutes ces méthodes, une fois le type de représentation choisi, tout signal peut être étudié. Une autre approche est de modéliser le signal analysé par un nombre restreint de paramètres, par exemple en définissant un modèle auto-régressif ou un modèle de Prony pour décrire le signal. La représentation temps-fréquence associée décrit alors uniquement le modèle estimé, et non le signal lui-même.

Toutes ces méthodes n'ont qu'un but : présenter le signal sous une forme plus facilement exploitable que sa forme temporelle ou fréquentielle. Une modulation linéaire de fréquence, c'est-à-dire un signal sinusoïdal dont la fréquence augmente linéairement avec le temps, n'est que difficilement déchiffable dans le domaine temporel, et ressemble à un signal large bande dans le domaine fréquentiel. Toutefois, sa signification devient immédiate lorsque l'on observe une droite dans le plan temps-fréquence.

Objectif de ces travaux

Ces travaux, comme le titre l'indique, cherchent à segmenter le plan temps-fréquence. Le but est de déterminer quelles sont les zones temps-fréquence où le signal analysé apparaît. Pour reprendre l'exemple de la modulation linéaire de fréquence, une segmentation fournit automatiquement l'ensemble des points décrivant la droite dans le plan temps-fréquence. L'analyse des signaux est alors facilitée, en mettant en avant la forme spectrale du signal, jusqu'à permettre d'automatiser une tâche de caractérisation de ce signal. Bien d'autres applications d'une telle segmentation sont possibles, et seront décrites dans ce manuscrit.

Vient la question de la définition d'une composante spectrale. Pour une modulation linéaire de fréquence, c'est théoriquement une droite dans le plan temps-fréquence. Toutefois, la forme de cette droite dépend grandement du choix de la représentation, allant d'une droite au sens mathématique du terme dans la distribution de Wigner-Ville, à une forme rectiligne relativement large dans le spectrogramme.

Une autre question arrive alors, quelle représentation choisir ? Nous cherchons dans ces travaux à garder un large cadre d'applications, autrement dit imposer le minimum d'information *a priori* sur le signal à analyser. Naturellement, les représentations paramétriques ne conviennent pas. La distribution de Wigner-Ville possède de bonnes propriétés de localisation, mais ses interférences la rendent difficile à interpréter lorsque des signaux riches sont analysés. Le choix le plus simple est fait, à savoir le spectrogramme, qui est le module carré d'une transformation linéaire, la transformée de Fourier à court terme.

Les propriétés de la transformée de Fourier font qu'un signal borné en temps est infini en fréquence, et inversement. Par conséquent, une composante spectrale dans le spectrogramme a nécessairement un support infini. Il semble pourtant évident que l'énergie est concentrée sur quelques points uniquement : le long d'une droite dans le cas de la modulation linéaire de fréquence. Intuitivement, l'énergie devient négligeable à mesure que l'on s'éloigne de la droite. De manière plus générale, définir une composante spectrale revient à déterminer un support temps-fréquence de cette composante tel que l'énergie du motif hors de ce support est négligeable. La question reste alors entière.

Les deux approches de la segmentation temps-fréquence proposées dans ce manuscrit s'appuient chacune sur des définitions différentes d'un motif temps-fréquence, bien que ce motif soit dans les deux cas celui provenant du spectrogramme.

Approche statistique

Cette première approche se situe dans la continuité des travaux initiés au Laboratoire des Images et des Signaux (LIS) par Cyrille Hory pendant sa thèse [Hor02], et a été développée en

parallèle des travaux de Julien Huillery [Hui08]. Le principe est de définir un modèle statistique pour le signal analysé, tout en gardant le plus large champ d'application possible : le modèle le plus large consiste en un signal à segmenter perturbé par un bruit, supposé blanc, gaussien et stationnaire. Il en découle un modèle statistique dans le plan temps-fréquence, permettant de détecter et segmenter les points temps-fréquence contenant de l'énergie du signal.

Dans ce cas, la définition d'une composante spectrale est liée à l'énergie du bruit, plus précisément au rapport signal à bruit sur un point temps-fréquence. Si l'énergie de la composante est détectable sur ce point, alors ce point appartient à la composante spectrale, sinon il est considéré comme étant du bruit. De plus, un critère de connexité est retenu : une composante spectrale est un ensemble de points connexes. Un signal unique peut avoir plusieurs composantes.

Approche basée sur les vecteurs de réallocation

La définition d'une composante spectrale dans la première approche est directement liée au bruit. Que se passe-t-il dans le cas où il n'y a strictement aucun bruit ? De plus, la définition est basée sur la perturbation de la composante, plutôt que la composante elle-même.

L'idée de cette deuxième approche est de définir d'une manière originale une composante spectrale, en utilisant le champ de vecteurs de réallocation du spectrogramme, comme source d'information permettant de caractériser le spectrogramme associé. La réallocation, qui est une représentation temps-fréquence non-linéaire déplaçant l'énergie du spectrogramme d'un point à un autre, n'est pas dans ce cas utilisée en tant que représentation temps-fréquence. Plus que le motif lui-même, c'est sa frontière qui est définie à l'aide de la réallocation : de là découle une proposition de définition de son contour. Le support de la composante spectrale est alors la zone temps-fréquence située à l'intérieur du contour.

Cette définition *a priori* indépendante du bruit permet de ne pas faire l'hypothèse du modèle de signal de la première approche, étendant par conséquent le champ d'application potentiel de cette méthode.

Organisation du document

Deux approches différentes pour résoudre le problème de la segmentation temps-fréquence nécessitent deux parties séparées : le premier contient quatre chapitres, la seconde deux chapitres.

La première partie est consacrée à la segmentation basée sur un modèle statistique du signal temporel. Le premier chapitre présente le choix du modèle fait, à savoir un signal à segmenter perturbé par un bruit blanc gaussien, qui permet de limiter l'ajout d'information *a priori* sur le signal à segmenter. Ainsi, nous garantissons un large champ d'application à cette segmentation. Dans ce chapitre, nous nous restreignons à l'étude de la Transformée de Fourier à Court Terme (TFCT). Dans un premier temps, les statistiques des coefficients de la TFCT d'un bruit blanc gaussien sont étudiées, conduisant à la proposition d'une TFCT « rectifiée », dont les coefficients temps-fréquence suivent tous la même loi, une loi gaussienne circulaire, de variance dépendant de la variance du bruit additif. Dans un second temps sont étudiés les coefficients de la TFCT d'un signal quelconque à segmenter additionné d'un bruit blanc gaussien. Par l'absence de modèle sur le signal à segmenter, seules les caractéristiques permettant de différencier un point temps-fréquence contenant du signal d'un point temps-fréquence n'en contenant pas seront données.

Détecter l'ensemble des points temps-fréquence contenant du signal à segmenter revient à

effectuer sur chaque point un test binaire d'hypothèse. Le chapitre 2 présente alors les bases de la théorie de la détection, notamment les approches bayésienne et de Neyman-Pearson. L'application dans le cadre de la TFCT « rectifiée » est ensuite donnée.

Le test binaire d'hypothèse s'appuie sur l'hypothèse que la variance du bruit additif du modèle de signal est connue. Le chapitre 3 propose une méthode itérative afin d'estimer cette variance. Après un rappel des principes et définitions de base de l'estimation de paramètres, nous proposons une méthode d'estimation de la variance d'un bruit blanc gaussien dans le plan temps-fréquence. Cette estimation est ensuite appliquée dans deux cas : le premier, quand l'ensemble du plan temps-fréquence ne contient que du bruit. Sans surprise, l'estimateur est efficace. Le second cas est lorsque des points temps-fréquence contenant du signal à segmenter sont présents. La position de ces points est inconnue, ils sont utilisés par l'estimateur comme des points ne contenant que du bruit, ce qui biaise l'estimation et donne une valeur surestimée de la variance du bruit.

L'estimation de la variance du bruit se fait alors de manière itérative. Dans une première estimation, le niveau de bruit est surestimé, mais permet de détecter quelques points contenant du signal à segmenter. L'itération suivante se fait sur les points non segmentés, diminuant ainsi la surestimation, *et caetera*. Le principal problème de ce type d'estimation est le choix d'un critère pour déterminer quand arrêter ces itérations, autrement dit quand décider que l'estimation est satisfaisante.

Pour ce faire, le critère du kurtosis est retenu. Le kurtosis fait partie de la famille des statistiques d'ordre supérieur. Plus précisément, c'est le cumulatif d'ordre 4, qui présente la particularité d'être nul pour des variables aléatoires gaussiennes. Ainsi, le kurtosis est estimé sur les points non segmentés, jusqu'à ce qu'il soit nul, impliquant que tous les points non segmentés suivent une même loi gaussienne, autrement dit qu'ils soient tous issus du bruit uniquement. L'influence sur le kurtosis de la TFCT « rectifiée » par rapport à la TFCT classique est étudiée, ainsi que le choix d'un seuil en dessous duquel le kurtosis est considéré comme nul.

Le chapitre 4 décrit l'algorithme complet de segmentation, consistant en l'assemblage de chacune des briques présentées dans les chapitres précédents. Un algorithme de croissance de région est appliqué à chaque détection, permettant ainsi de créer les motifs spectraux correspondant aux points détectés. Trois paramètres en plus des paramètres de la TFCT contrôlent cet algorithme. Leur influence est successivement étudiée, ainsi que l'influence du rapport signal à bruit du signal à segmenter.

Des extensions à ce principe général de segmentation sont ensuite proposées, étendant le champ d'application de cet algorithme aux perturbations gaussiennes non blanches et non stationnaires, ainsi qu'à d'autres transformées, la transformée en ondelettes et le caponogramme. Des applications sur signaux simulés et signaux réels sont finalement présentées.

La seconde partie est consacrée à la segmentation basée sur les vecteurs de réallocation. Le chapitre 5 présente le principe de la réallocation du spectrogramme. La réallocation a pour but de déplacer l'énergie d'un point temps-fréquence sur un point qui lui donne plus de sens. Deux interprétations équivalentes peuvent en être faites. La première, la réallocation déplace l'énergie d'un point donné au point correspondant à son retard de groupe et à sa fréquence instantanée. La seconde, l'énergie est déplacée sur un point correspondant au centre de gravité local de la distribution de l'énergie du spectrogramme. Ainsi, l'énergie est focalisée par la réallocation. Le champ de vecteurs de réallocation est ensuite présenté, correspondant aux vecteurs de déplacement de l'énergie par la réallocation.

L'énergie étant déplacée sur son centre de gravité local par la réallocation, les vecteurs de réallocation sur la frontière d'un motif vont tous être dirigés vers l'intérieur du motif. Ainsi, ils seront localement parallèles. Nous avons défini la frontière d'un motif comme celle correspondant aux zones temps-fréquence où les vecteurs de réallocation sont localement parallèles. Cette approche se démarque de la proposition de Chassande-Mottin [CM98], qui partitionne le plan temps-fréquence en différents *bassins d'attraction*, déterminés par le champ de vecteurs de réallocation.

Le chapitre 6 commence par présenter le principal problème lié à cette définition, à savoir la discrétisation du plan temps-fréquence, qui empêche un strict parallélisme entre deux vecteurs voisins. Les solutions pour limiter les problèmes sont également présentées. Un premier algorithme de détection de zone frontière est proposé, et mis en application en tant qu'ajout d'information à la segmentation statistique. Ce principe est ensuite illustré sur un signal réel.

Par la suite, nous définissons un *degré de parallélisme* associé à chaque vecteur de réallocation, donnant le nombre de vecteurs qui lui sont parallèles sur un voisinage donné. Nous proposons un algorithme de création de contour, qui suit dans le plan temps-fréquence les maxima de ce degré de parallélisme. À partir de deux signaux synthétiques de test, une modulation linéaire de fréquence et un signal large bande provenant d'un bruit filtré, l'influence d'une perturbation stationnaire gaussienne et blanche sur la construction du contour est étudiée, ainsi que l'influence des paramètres de la TFCT.

Enfin, des pistes pour améliorer cet algorithme sont proposées.

Les conclusions et les perspectives de ces travaux concluront ce manuscrit.

Première partie

**Segmentation statistique du plan
temps-fréquence**

Table des matières

Introduction	11
1 Transformée de Fourier à Court Terme et caractérisation statistique	13
1.1 Loi de la TFCT d'un bruit blanc stationnaire gaussien	15
1.1.1 Transformée de Fourier à Court Terme « rectifiée »	20
1.2 Cas d'un signal quelconque noyé dans un bruit blanc stationnaire gaussien	23
2 Détection des points temps-fréquence contenant du signal	25
2.1 Test binaire d'hypothèses	25
2.1.1 L'approche bayésienne	26
2.1.2 L'approche de Neyman-Pearson	28
2.2 Détection dans la TFCT rectifiée	29
3 Estimation itérative du bruit	33
3.1 Estimation : principe et définitions	34
3.1.1 Estimation par maximum de vraisemblance dans le cas du bruit seul	35
3.1.2 Estimation par maximum de vraisemblance dans le cas signal et bruit	37
3.2 Critère d'arrêt de l'estimation itérative : le kurtosis	38
3.2.1 Statistiques d'ordre supérieur	38
3.2.2 Kurtosis de la TFCT rectifiée	42
3.3 Conclusion	48
4 Algorithme de segmentation	49
4.1 Croissance de région	49
4.2 Description de l'algorithme de segmentation	50
4.3 Influence des paramètres	52
4.3.1 Probabilité de fausse alarme p_{fa}	53
4.3.2 Proportion de candidats segmentés à chaque sous-boucle p_{cand}	54
4.3.3 Seuil sur le kurtosis des points non segmentés s_k	55
4.3.4 Influence du Rapport Signal à Bruit	56
4.4 Extensions du principe de segmentation	56
4.4.1 Bruit non blanc ou non stationnaire	56

TABLE DES MATIÈRES

4.4.2	Transformée en ondelettes	57
4.4.3	Capongramme	59
4.5	Applications	61
4.5.1	Débruitage	61
4.5.2	Séparation de sources	62
4.5.3	Caractérisation de signaux	64
Conclusion		67

Introduction

Cette première partie a pour objectif de présenter une méthode générique de segmentation de motifs issus d'une représentation temps-fréquence, s'appuyant sur une description statistique de la représentation temps-fréquence. Par « motif », nous entendons ici « zones connexes du plan temps-fréquence contenant du signal d'intérêt », en opposition aux zones où seul du bruit est présent. Nous appellerons support temps-fréquence du ou des signaux à segmenter l'ensemble de ces motifs.

Le but est ici de proposer une méthode automatique nécessitant le moins d'information *a priori* sur le signal, afin de proposer un algorithme susceptible de s'appliquer à une large gamme de signaux.

Pour ce faire, nous considérons un signal observé $x[m]$, composé de la somme d'un signal d'intérêt à segmenter $s[m]$ et d'un bruit blanc gaussien stationnaire $b[m]$ de variance inconnue σ_b^2 .

$$x[m] = s[m] + b[m] \tag{1}$$

Ce modèle de signal est une extension d'un modèle utilisé précédemment dans un travail similaire par Hory [Hor02], qui considérait un signal déterministe dans un bruit additif gaussien. Dans notre contexte, le signal d'intérêt $s[m]$ peut être un signal stochastique tout comme un signal déterministe. La seule condition restante sur $s[m]$ est de ne pas être un signal blanc gaussien et stationnaire, ce qui l'assimilerait à la perturbation $b[m]$.

Ce modèle de signal permet d'englober une large gamme de signaux, avec un nombre quelconque de composantes à bande étroite ou à large bande, avec modulation de fréquence ou d'amplitude.

Une fois un modèle de signal choisi, il faut ensuite choisir une transformée temps-fréquence. De nombreuses transformées existent pour représenter l'énergie du signal conjointement en temps et en fréquence; le lecteur pourra se reporter à [Fla93] ou à [Coh89] pour en faire un tour d'horizon du domaine.

Pour cette segmentation, nous nous intéresserons à la transformée de Fourier à Court Terme (TFCT), dont le module élevé au carré est appelé spectrogramme, et qui possède des caractéristiques intéressantes. Elle est exempte d'interférences lorsqu'il y a plusieurs composantes présentes, le spectrogramme est strictement positif et possède ainsi directement une notion de répartition d'énergie. De plus, le caractère linéaire de la transformée de Fourier rend facile sa mise en œuvre et son interprétation.

Une fois ce contexte défini, il ne reste plus qu'à déterminer comment segmenter le spectrogramme du signal modélisé par (1). Dans le chapitre 1, nous nous intéresserons aux propriétés

statistiques des coefficients temps-fréquence, et proposerons une normalisation de la TFCT telle que tous les coefficients issus d'un bruit blanc soient circulaires.

Le chapitre 2 présentera brièvement les processus de décision dans un contexte général afin de l'appliquer à la TFCT, et proposera une méthode pour déterminer quels sont les points temps-fréquence contenant du signal et quels sont ceux ne contenant que du bruit. Cette décision s'appuie sur l'hypothèse de variance de bruit connue : le chapitre 3 proposera donc une estimation de cette variance, ce qui nous conduira à un algorithme itératif. Le kurtosis, qui est le cumulatif d'ordre quatre issu des statistiques d'ordre supérieur, sera décrit et illustré dans le cadre de variables réelles et complexes, et permettra de contrôler la convergence de cet algorithme.

Tous ces éléments mis en place, l'algorithme dans son ensemble sera décrit et caractérisé dans le chapitre 4. Le principe de l'algorithme de segmentation étant assez général, des extensions à d'autres transformées que la TFCT, et au cas où le bruit additif $b[m]$ n'est ni blanc ni gaussien, seront décrites. Enfin, différentes applications de cette segmentation seront illustrées.

Chapitre 1

Transformée de Fourier à Court Terme Terme et caractérisation statistique

Le but de ce chapitre est de rappeler la Transformée de Fourier à Court Terme (TFCT), puis de caractériser les coefficients de cette transformée appliquée au modèle « signal + bruit » (équation (1) de l'introduction) utilisé dans ces travaux.

La TFCT d'un signal discret $x[m]$, ou Transformée de Fourier Glissante, notée $X_\phi[n, k]$, est l'ensemble des transformées de Fourier de N segments du signal $x[m]$, chaque segment étant apodisé par une fenêtre d'analyse $\phi[m]$ centrée à différents instants n :

$$X_\phi[n, k] = \sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} x[m]\phi[m-n]e^{-2j\pi k \frac{m}{M_\phi+Z}} \quad (1.1)$$

k est la fréquence réduite, M_ϕ la longueur impaire de la fenêtre d'analyse $\phi[m]$, Z le bourrage de zéros (en anglais *zero-padding*) afin de calculer $K = M_\phi + Z$ fréquences discrètes.

Le module carré de la TFCT, nommé spectrogramme et noté $S_{\phi,x}[n, k]$, représente l'évolution du spectre du signal $x[m]$ dans le temps aux différents instants n :

$$S_{\phi,x}[n, k] = |X_\phi[n, k]|^2 = (X^r[n, k])^2 + (X^i[n, k])^2 \quad (1.2)$$

où $X^r[n, k]$ et $X^i[n, k]$ sont respectivement les parties réelle et imaginaire de $X_\phi[n, k]$:

$$\begin{aligned} X_\phi^r[n, k] = \Re(X_\phi[n, k]) &= \sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} x[m]\phi[m-n] \cos\left(-2\pi k \frac{m}{M_\phi+Z}\right) \\ X_\phi^i[n, k] = \Im(X_\phi[n, k]) &= \sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} x[m]\phi[m-n] \sin\left(-2\pi k \frac{m}{M_\phi+Z}\right) \end{aligned} \quad (1.3)$$

La figure 1.1 donne un exemple de spectrogramme d'un signal synthétique. Dans ce cas d'une modulation linéaire de fréquence, le spectre du signal n'informe aucunement sur l'évolution de la fréquence du signal, contrairement au spectrogramme.

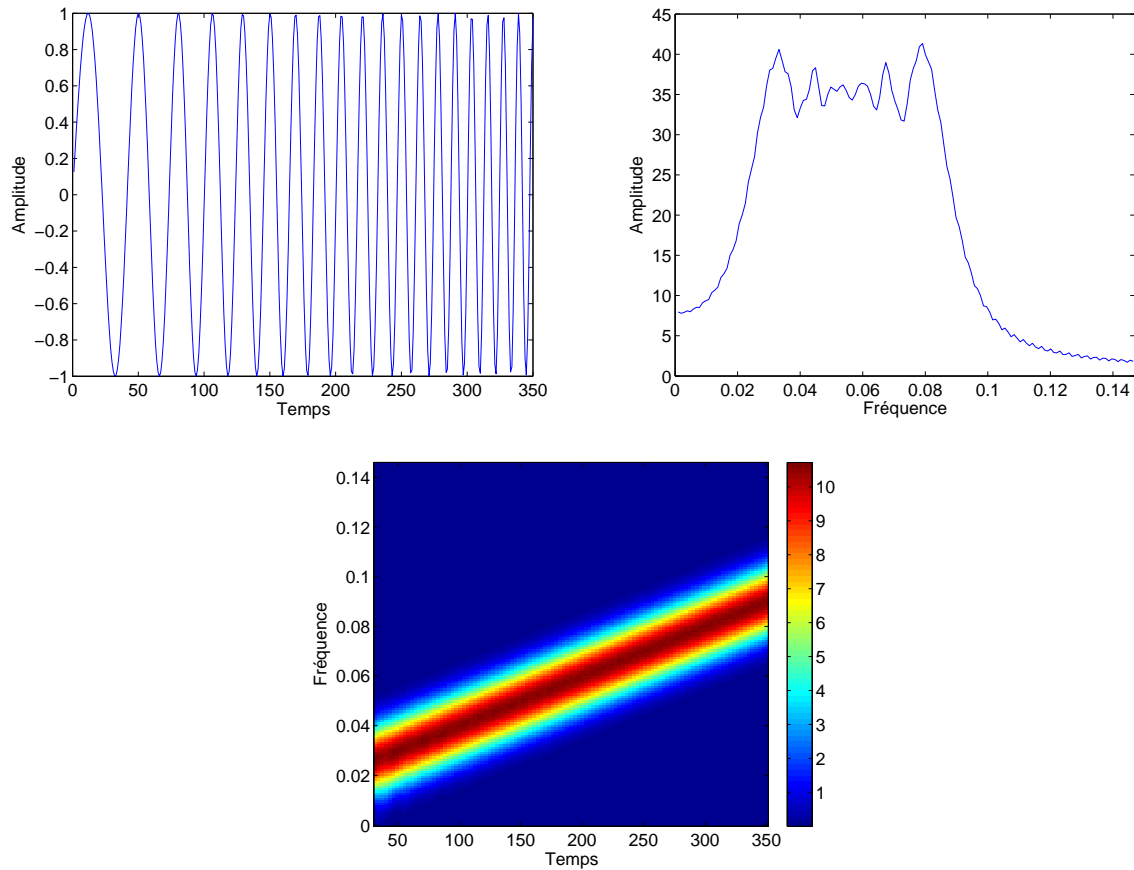


FIG. 1.1 : Illustration de la TFCT sur une modulation linéaire de fréquence. À gauche le signal temporel, à droite son spectre, en bas son spectrogramme. Dans le cas d'un signal non stationnaire, le spectre ne renseigne pas sur l'évolution du contenu spectral, contrairement au spectrogramme.

Nous considérerons par la suite uniquement des fenêtres normées en énergie :

$$\sum_{m=-\frac{M_\phi-1}{2}}^{+\frac{M_\phi-1}{2}} \phi[m]^2 = 1. \quad (1.4)$$

Pour alléger les notations, les indices des sommes seront ignorés.

Les paragraphes 1 et 2 étudient respectivement les propriétés statistiques des coefficients de la TFCT d'un bruit blanc gaussien, puis du modèle « signal + bruit », dans le but de déterminer un critère permettant de discriminer si un coefficient temps-fréquence contient uniquement l'énergie du bruit ou également une contribution du signal à segmenter.

1.1 Loi de la TFCT d'un bruit blanc stationnaire gaussien

Nous nous intéressons dans cette section à un signal constitué d'un bruit blanc gaussien stationnaire de variance σ_b^2 , $x[m] = b[m]$. L'équation (1.1) devient :

$$B_\phi[n, k] = \sum_m b[m] \phi[m-n] e^{\left(-2j\pi k \frac{m}{M_\phi + Z}\right)} = B^r[n, k] + jB^i[n, k] \quad (1.5)$$

$B^r[n, k]$ et $B^i[n, k]$ sont des sommes de M_ϕ variables gaussiennes indépendantes : elles suivent donc des lois gaussiennes de moyenne et de variance respectivement égales à la somme des moyennes et à la somme des variances de ces M_ϕ variables.

Ainsi, la loi de la variable $B_\phi[n, k]$ est une gaussienne complexe. Le bruit $b[m]$ étant centré, les moyennes de $B_\phi^r[n, k]$ et $B_\phi^i[n, k]$ sont nulles, par conséquent $B_\phi[n, k]$ est également centrée.

Le moment d'ordre deux de la variable complexe $B_\phi[n, k]$ est défini par [LAP97] :

$$\begin{aligned} \text{MO2}(B_\phi[n, k]) &= \text{E}(B_\phi[n, k] B_\phi^*[n, k]) \\ &= \text{E}(|B_\phi[n, k]|^2) = \text{E}\left((B_\phi^r[n, k])^2\right) + \text{E}\left((B_\phi^i[n, k])^2\right) \end{aligned} \quad (1.6)$$

Calculons les moments d'ordre deux de $B_\phi^r[n, k]$ et $B_\phi^i[n, k]$:

$$\text{E}(B_\phi^r[n, k]^2) = \sigma_b^2 \sum_m \phi[m-n]^2 \cos^2\left(-2\pi k \frac{m}{M_\phi + Z}\right) \quad (1.7)$$

$$\text{E}(B_\phi^i[n, k]^2) = \sigma_b^2 \sum_m \phi[m-n]^2 \sin^2\left(-2\pi k \frac{m}{M_\phi + Z}\right) \quad (1.8)$$

Étant donné que la fenêtre d'analyse est normalisée (1.4), les équations (1.7) et (1.8) conduisent à la simplification du moment d'ordre 2 de $B_\phi[n, k]$. L'équation (1.6) devient :

$$\text{MO2}(B_\phi[n, k]) = \text{E}(B_\phi^r[n, k]^2) + \text{E}(B_\phi^i[n, k]^2) = \sigma_b^2 \quad (1.9)$$

Le bruit gaussien b étant blanc et stationnaire, tous les points $[n, k]$ de la TFCT ont la même variance.

Le spectrogramme d'un bruit blanc étant la somme de deux gaussiennes élevées au carré (1.2), ses coefficients semblent suivre naturellement une loi du χ^2 . Ceci impliquerait que $B_\phi[n, k]$ est circulaire, c'est-à-dire que les parties réelle et imaginaire de la TFCT sont indépendantes et de même variance [Pic94].

Afin de caractériser plus finement la loi de $B_\phi[n, k]$, notamment vérifier sa circularité, nous nous intéressons à la répartition de la variance totale du bruit σ_b^2 entre les parties réelle et imaginaire de la TFCT. Pour ce faire, nous définissons le rapport $\alpha[n, k]$ entre le moment d'ordre deux de la partie réelle de la TFCT (1.7) et le moment d'ordre deux de la variable complexe donné en (1.9) :

$$\begin{aligned} \alpha[n, k] &= \frac{1}{\sigma_b^2} \text{E}(B_\phi^r[n, k]^2) = \sum_m \phi[m-n]^2 \cos^2\left(-2\pi k \frac{m}{M_\phi + Z}\right) \\ &= \frac{1}{2} + \frac{1}{2} \sum_m \phi[m]^2 \cos\left(-4\pi k \frac{m+n}{M_\phi + Z}\right) \end{aligned} \quad (1.10)$$

Les moments d'ordre deux (1.7) et (1.8) s'écrivent alors :

$$\mathbb{E}(B_\phi^r[n, k]^2) = \alpha[n, k]\sigma_b^2 \quad (1.11)$$

$$\mathbb{E}(B_\phi^i[n, k]^2) = (1 - \alpha[n, k])\sigma_b^2 \quad (1.12)$$

Le paramètre $\alpha[n, k]$ détermine totalement les lois des variables aléatoires $B_\phi^r[n, k]$ et $B_\phi^i[n, k]$, à la variance du bruit près.

Pour compléter la caractérisation de la loi de $B_\phi[n, k]$, nous nous intéressons à la corrélation entre les parties réelle et imaginaire, par l'intermédiaire du coefficient de corrélation $\rho[n, k]$:

$$\rho[n, k] = \frac{\mathbb{E}(B_\phi^r[n, k]B_\phi^i[n, k])}{\sqrt{\mathbb{E}(B_\phi^r[n, k]^2)\mathbb{E}(B_\phi^i[n, k]^2)}} \quad (1.13)$$

$$= \frac{\sigma_b^2 \sum_m \phi[m]^2 \cos(-2\pi k \frac{m+n}{M_\phi+Z}) \sin(-2\pi k \frac{m+n}{M_\phi+Z})}{\sigma_b^2 \sqrt{\alpha[n, k](1 - \alpha[n, k])}} \quad (1.14)$$

$$= \frac{1}{2\sqrt{\alpha[n, k](1 - \alpha[n, k])}} \sum_m \phi[m]^2 \sin(-4\pi k \frac{m+n}{M_\phi+Z}) \quad (1.15)$$

Cette définition implique que $\alpha[n, k] \neq 1$ et $\alpha[n, k] \neq 0$. Dans ces cas, la partie réelle ou imaginaire de la TFCT est strictement nulle, la corrélation entre les deux est nulle mais le coefficient de corrélation n'est pas défini.

La figure 1.2 montre l'évolution du paramètre α pour une fenêtre de Blackman de 31 points et 128 fréquences calculées. $\alpha[n, k]$ vaut $\frac{1}{2}$ pour un grand nombre de points : pour tous ces points, $\alpha[n, k] = 1 - \alpha[n, k]$, c'est-à-dire que les variances de $B_\phi^r[n, k]$ et $B_\phi^i[n, k]$ sont égales.

La figure 1.3 montre l'évolution du coefficient de corrélation pour une fenêtre de Blackman de 31 points et 128 fréquences calculées. Une corrélation entre les parties imaginaire et réelle de la TFCT d'un bruit blanc est présente sur les mêmes bandes de fréquence que les points pour lesquels $\alpha[n, k]$ ne vaut pas $\frac{1}{2}$. La largeur de cette bande de fréquence dépend des paramètres choisis pour calculer la TFCT : en effet les équations (1.10) et (1.15) dépendent du type de fenêtre d'analyse $\phi[m]$, de sa longueur M_ϕ et du bourrage de zéros Z .

On voit ainsi qu'une différence de variance entre les parties réelle et imaginaire de la TFCT apparaît pour les basses et hautes fréquences, ainsi qu'une corrélation.

Pour étudier ces phénomènes, nous définissons la fonction $\Phi[n, k]$, telle que :

$$\Phi[n, k] = e^{(-2j\pi 2k \frac{n}{M_\phi+Z})} \sum_m \phi[m]^2 e^{(-2j\pi 2k \frac{m}{M_\phi+Z})} \quad (1.16)$$

Les équations (1.10) et (1.15) s'écrivent en fonction de $\Phi[n, k]$:

$$\alpha[n, k] = \frac{1}{2} + \frac{1}{2} \Re(\Phi[n, k]), \quad (1.17)$$

$$\rho[n, k] = \frac{1}{2\sqrt{\alpha[n, k](1 - \alpha[n, k])}} \Im(\Phi[n, k]) \quad (1.18)$$

$\rho[n, k]$ n'est défini que pour $\alpha[n, k] \neq 1$ et $\alpha[n, k] \neq 0$, la partie réelle ou imaginaire étant nulle dans ces cas.

Ainsi, toutes les variations de $\alpha[n, k]$ et de $\rho[n, k]$ sont portées par la fonction $\Phi[n, k]$. Cette dernière (1.16) est composée de deux termes, un terme de phase $e^{(-2j\pi 2k \frac{n}{M_\phi+Z})}$ dépendant des

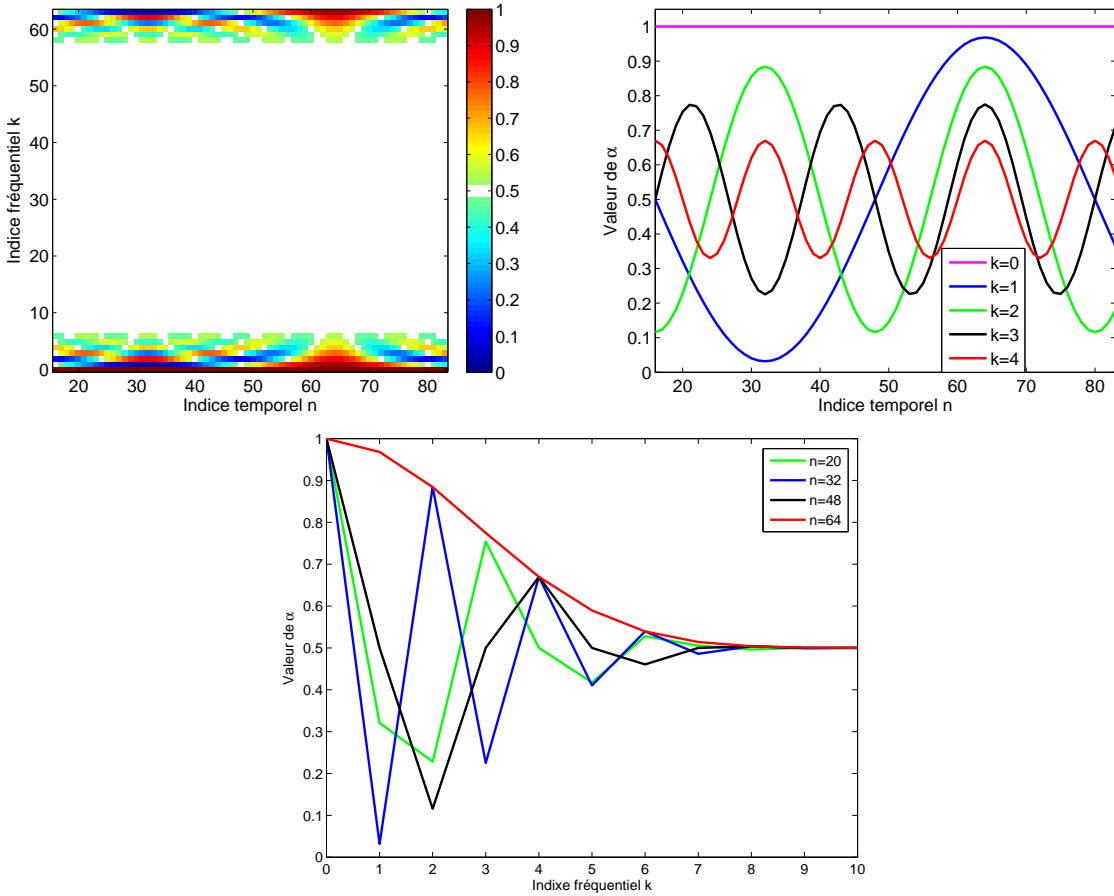


FIG. 1.2 : En haut à gauche, variations de $\alpha[n, k]$ pour une TFCT avec une fenêtre de Blackman de 31 points, un recouvrement de 30 points et 128 fréquences calculées. Les points blancs sont ceux pour lesquels $\alpha[n, k]$ est compris entre 0.485 et 0.515. En haut à droite, variations de α pour les fréquences $k = 0$ à 4. En bas, variations de α pour les instants $n = 20, 32, 48$ et 64 pour les premières valeurs de k .

indices temps-fréquence $[n, k]$ et une Transformée de Fourier Discrète (TFD) du carré de la fenêtre de pondération $\phi[m]$ et comprimée en fréquence d'un facteur 2.

De ce fait, $\alpha[n, k]$ diffère de $\frac{1}{2}$ et $\rho[n, k]$ est non nul sur une bande de fréquence dépendant de la TFD d'une fenêtre de pondération, c'est-à-dire autour de la fréquence nulle et des multiples de la fréquence d'échantillonnage. Cette bande de fréquence va logiquement dépendre de la taille de la fenêtre M_ϕ , du nombre de fréquences calculées $M_\phi + Z$ et du type de fenêtre. Plus la fenêtre temporelle est grande, plus faible est sa largeur spectrale ; plus on utilise de bourrage de zéros, plus on a de points pour décrire le spectre, et plus on a des valeurs de $\alpha[n, k]$ et de $\rho[n, k]$ variables.

La figure 1.4 illustre ce phénomène en montrant les variations de la valeur absolue de $\Phi[n, k]$ selon la fréquence k en fonction des paramètres de la TFCT.

En pratique, la TFCT est rarement calculée par l'équation décrite en (1.1), la translation temporelle étant portée par le signal $x[m]$ plutôt que par la fenêtre $\phi[m]$. Par exemple, la TFCT

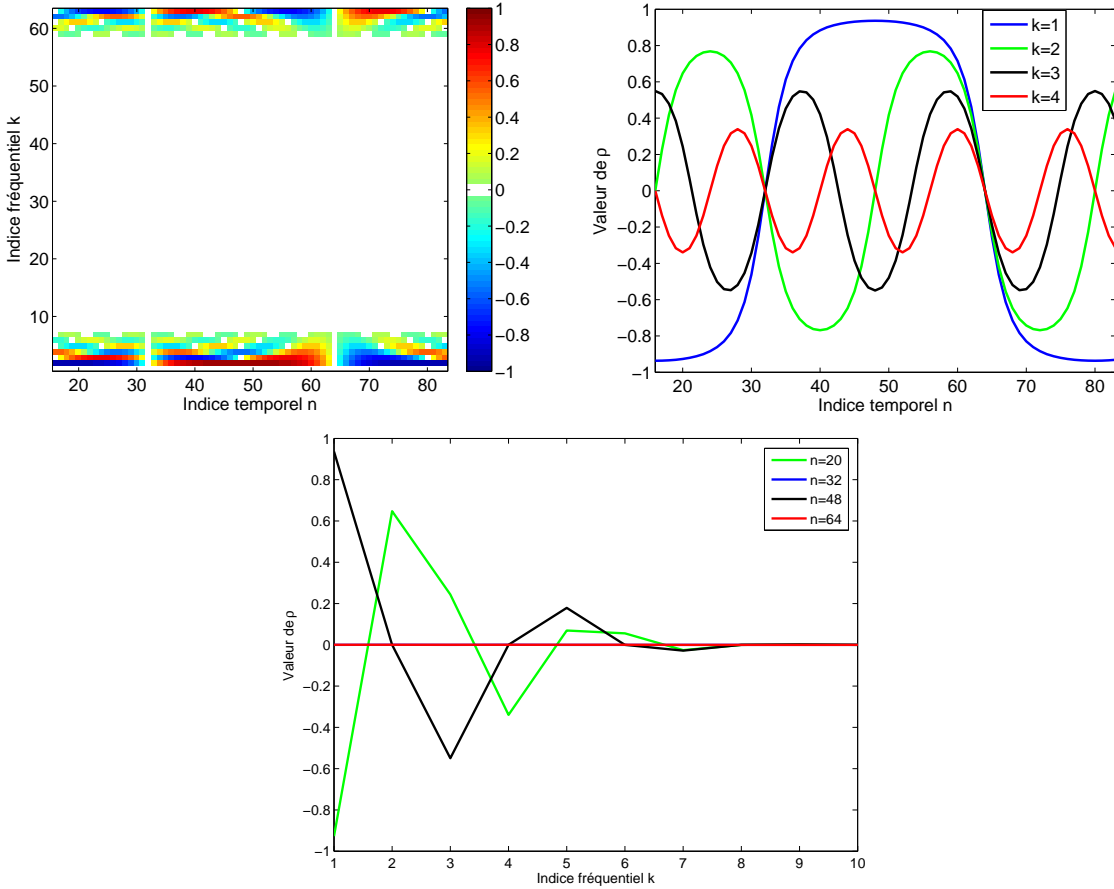


FIG. 1.3 : En haut à gauche, variations du coefficient de corrélation $\rho[n, k]$ pour une TFCT avec une fenêtre de Blackman de 31 points, un recouvrement de 30 points et 128 fréquences calculées. $\rho[n, k]$ n'étant pas défini pour $k = 0$, ces valeurs ne sont pas représentées. Les points présentant de la corrélation se situent aux mêmes fréquences que les valeurs de $\alpha[n, k]$ différentes de $\frac{1}{2}$ (voir figure 1.2). En haut à droite, variations de ρ pour les fréquences $k = 1$ à 4. En bas, variations de ρ pour les instants $n = 20, 32, 48$ et 64 pour les premières valeurs de k .

donnée par la Boîte à Outils Temps-Fréquence¹ est calculée selon l'équation :

$$X_\phi[n, k] = \sum_{m=0}^{\frac{M_\phi-1}{2}} x[n+m] \phi[m] e^{\left(-2j\pi k \frac{m}{M_\phi+Z}\right)} + \dots$$

$$\sum_{m=M_\phi+Z-\frac{M_\phi-1}{2}}^{M_\phi+Z-1} x[n+m-(M_\phi+Z)] \phi[m-(M_\phi+Z)] e^{\left(-2j\pi k \frac{m}{M_\phi+Z}\right)} \quad (1.19)$$

¹Cette Boîte à Outils est une collection de scripts pour GNU Octave et Matlab[®] concernant l'analyse de signaux non stationnaires, distribuée sous la licence publique générale GNU et disponible sur http://tftb.nongnu.org/index_fr.html

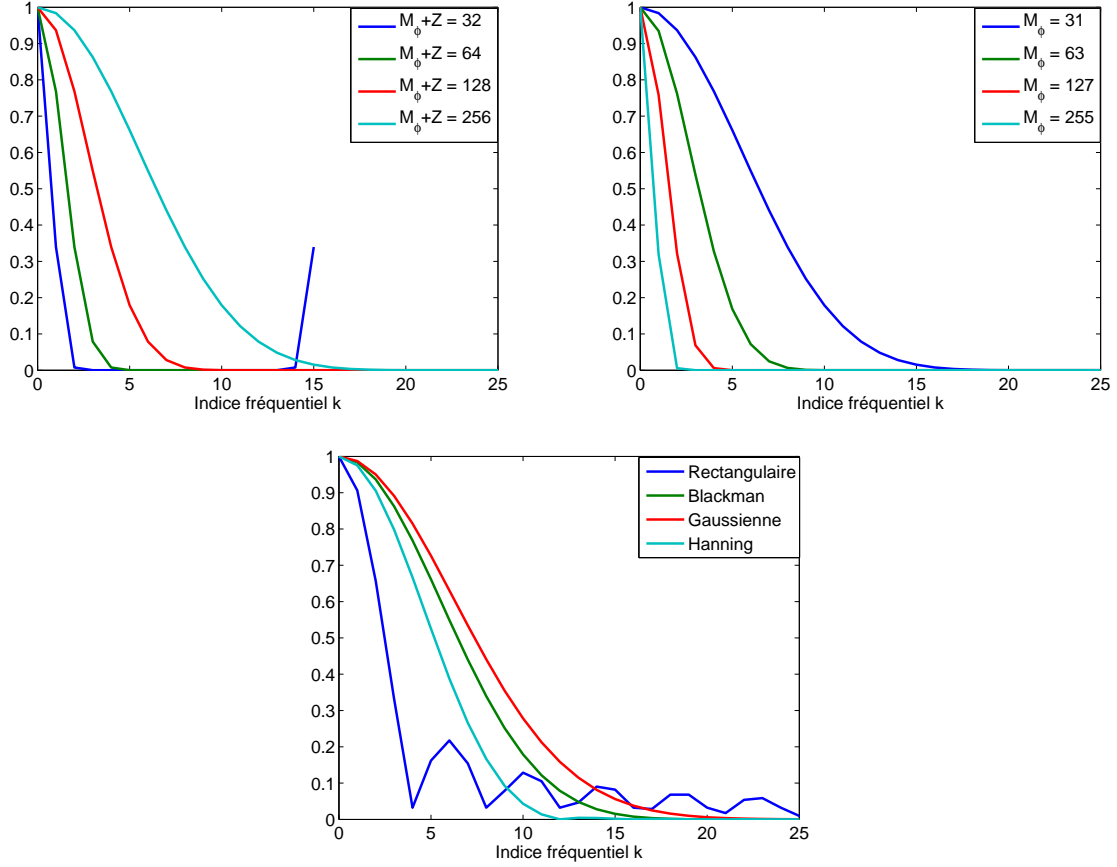


FIG. 1.4 : Variation de la valeur absolue de $\Phi[n, k]$ (1.16) selon les paramètres de la TFCT, pour les premières valeurs de la fréquence réduite k . En haut à gauche, l'influence du nombre de fréquences calculées $M_\phi + Z$, pour une fenêtre de Blackman de 31 points ; en haut à droite, l'influence de la taille de la fenêtre de Blackman, pour 256 fréquences calculées ; en bas, l'influence du type de fenêtre, de 31 points pour 256 fréquences calculées. La bande de fréquence où $\phi[n, k]$ n'est pas nul correspond aux points temps-fréquence où $\alpha[n, k]$ est différent de $\frac{1}{2}$ et $\rho[n, k]$ différent de 0.

où la fenêtre d'analyse $\phi[m]$ garde la même définition que précédemment (1.4). En utilisant les propriétés de périodicité de l'exponentielle complexe, l'équation (1.19) s'écrit :

$$X_\phi[n, k] = \sum_{m=-\frac{M_\phi-1}{2}}^{+\frac{M_\phi-1}{2}} x[n+m] \phi[m] e^{-2j\pi k \frac{m}{M_\phi+Z}} \quad (1.20)$$

En utilisant cette définition, les moments d'ordre deux des parties réelle et imaginaire de la TFCT d'un bruit blanc (1.7,1.8) ne dépendent plus de l'instant n . La fonction $\Phi[n, k]$ est alors égale à :

$$\Phi[n, k] = \sum_{m=-\frac{M_\phi-1}{2}}^{+\frac{M_\phi-1}{2}} \phi[m]^2 e^{-2j\pi 2k \frac{m}{M_\phi+Z}} = \Phi[0, k] \quad (1.21)$$

$\Phi[n, k]$ est dans ce cas stationnaire, par conséquent $\alpha[n, k]$ et $\rho[n, k]$ également (équations (1.17) et (1.18)).

De plus, les fenêtres de pondérations utilisées couramment sont paires. En utilisant la formule d'Euler, l'équation (1.21) s'écrit :

$$\begin{aligned} \Phi[n, k] &= \phi[0]^2 + \sum_{m=1}^{+\frac{M_\phi-1}{2}} \phi[m]^2 \left(\exp(-2j\pi 2k \frac{m}{M_\phi + Z}) + \exp(2j\pi 2k \frac{m}{M_\phi + Z}) \right) \quad (1.22) \\ &= \phi[0]^2 + 2 \sum_{m=1}^{+\frac{M_\phi-1}{2}} \phi[m]^2 \cos(4\pi k \frac{m}{M_\phi + Z}) \quad (1.23) \end{aligned}$$

Dans le cas d'une fenêtre d'analyse $\phi[m]$ paire, la fonction $\Phi[n, k]$ est réelle. D'après l'équation (1.18), la corrélation entre la partie réelle et imaginaire de tous les points temps-fréquence est nulle.

Les deux paramètres $\alpha[n, k]$ et $\rho[n, k]$ caractérisent précisément la loi de $X_\phi[n, k]$, et permettent d'en déduire la loi du spectrogramme $S_\phi[n, k]$. Les points $[n, k]$ pour lesquels $\alpha[n, k]$ est différent de $\frac{1}{2}$ ou $\rho[n, k]$ est non nul ne suivent pas une loi du χ^2 de degré deux, contrairement aux autres.

Dans le cadre de spectres discrets, Koopmans [Koo74] et Johnson et Long [JL99] avaient déjà déterminé que les coefficients du spectre d'un bruit blanc ne suivent pas une loi du χ^2 de degré deux aux fréquences $k = 0$ et $k = \frac{M_\phi}{2}$ pour une fenêtre rectangulaire sans bourrage de zéros ($Z=0$), et aux fréquences $k = 0$, $k = 1$, $k = \frac{M_\phi}{2} - 1$ et $k = \frac{M_\phi}{2}$ pour une fenêtre de Hanning sans bourrage de zéros. La fonction $\Phi[n, k]$ généralise et définit tous les points pour lesquels la loi du χ^2 de degré deux n'est pas applicable.

D'un autre point de vue, Durrani [DN73] caractérise la loi de chaque point du spectre par une loi du χ^2 de degré variable, inférieur ou égal à deux, et détermine par là quels sont les points qui ne suivent pas une loi de degré deux.

1.1.1 Transformée de Fourier à Court Terme « rectifiée »

Pour les fréquences réduites proches de 0 et de 0.5, les coefficients de la TFCT d'un bruit blanc ne sont pas circulaires et présentent une dissymétrie entre les variances de leurs parties réelle et imaginaire. Toutefois, cette dissymétrie est connue et caractérisée par le paramètre $\alpha[n, k]$, qui dépend uniquement des paramètres de la TFCT. Il est possible de proposer une rectification des valeurs de la TFCT et de définir $X_\phi[n, k]' = X_\phi^r[n, k]' + jX_\phi^i[n, k]'$, tel que :

$$\begin{aligned} X_\phi^r[n, k]' &= X_\phi^r[n, k], \\ X_\phi^i[n, k]' &= X_\phi^i[n, k] \quad \text{si } \alpha[n, k] = 0 \text{ ou } \alpha[n, k] = 1 \quad (1.24) \end{aligned}$$

$$\begin{aligned} X_\phi^r[n, k]' &= X_\phi^r[n, k] / \sqrt{2\alpha[n, k]}, \\ X_\phi^i[n, k]' &= X_\phi^i[n, k] / \sqrt{2(1 - \alpha[n, k])} \quad \text{sinon.} \quad (1.25) \end{aligned}$$

Ainsi, pour les points où $\alpha[n, k]$ est différent de 0 et 1, c'est-à-dire les coefficients de la TFCT qui ne sont pas strictement réels ou imaginaires, tous les coefficients des parties réelle et imaginaire suivent une loi gaussienne de même variance $\frac{\sigma_b^2}{2}$.

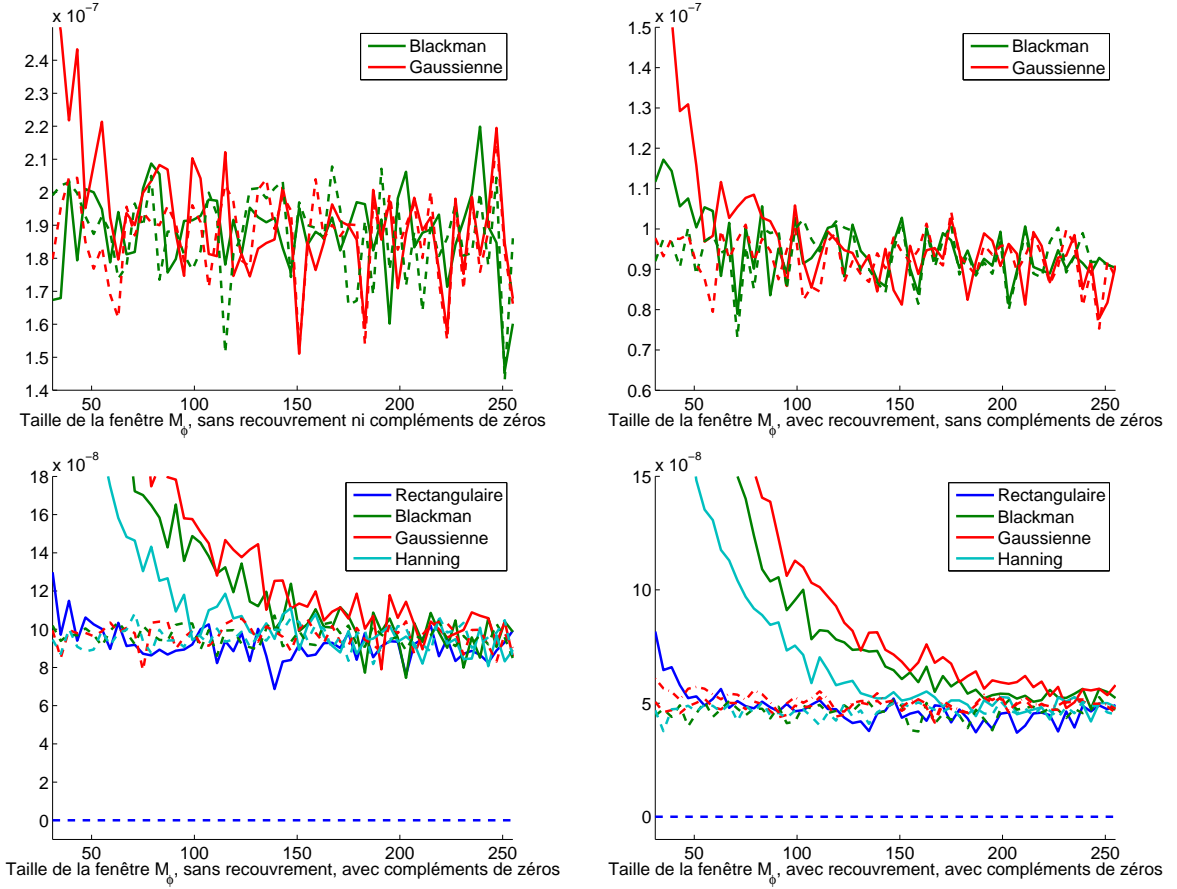


FIG. 1.5 : Divergence de Kullback-Leibler entre une loi du χ^2 de degré deux et les coefficients du spectrogramme d'un bruit blanc de 2000 points (trait plein) et les coefficients du spectrogramme « rectifié » (équation (1.26), trait pointillé), pour différentes fenêtres de taille variant de 31 à 255. En haut à gauche, dans le cas d'une TFCT sans recouvrement ni bourrage de zéros ; en haut à droite, une TFCT avec 75% de recouvrement, sans bourrage de zéros ; en bas à gauche, une TFCT sans recouvrement et quatre fois plus de fréquences calculées que de points dans la fenêtre ; en bas à droite, une TFCT avec 75% de recouvrement et quatre fois plus de fréquences calculées que de points dans la fenêtre. Pour des raisons de lisibilité, toutes les fenêtres ne sont pas affichées.

De plus, en choisissant une construction de la TFCT telle que la corrélation entre parties réelle et imaginaire est nulle (équation (1.20) avec une fenêtre $\phi[m]$ paire), tous les coefficients temps-fréquence complexes sont circulaires. Par conséquent tous les coefficients du spectrogramme « rectifié », définis par :

$$S_{\phi,x}[n, k]' = X_{\phi}^r[n, k]^2 + X_{\phi}^i[n, k]^2 \quad (1.26)$$

suivent une loi du χ^2 , de degré un si $\alpha[n, k] = 0$ ou 1, de degré deux sinon.

Huillery [HMM08] a montré qu'une loi du χ^2 de degré deux reste une bonne approximation de la loi d'un coefficient du spectrogramme même en présence de bruit corrélé, pour peu que la fenêtre d'analyse choisie s'annule à ses bords. Dans notre cas, nous cherchons à vérifier l'approximation de la loi de l'ensemble des coefficients du spectrogramme par une même loi du χ^2

de degré deux, et de la comparer avec la loi du spectrogramme « rectifié ».

Pour apprécier la ressemblance de l'ensemble des coefficients du spectrogramme avec une loi du χ^2 de degré 2, nous utilisons la divergence de Kullback-Leibler entre la loi p et la loi q , notée $D_{KL}(p, q)$, et définie par [CT91] :

$$D_{KL}(p, q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \quad (1.27)$$

La divergence de Kullback-Leibler mesure l'adéquation ou la dissimilarité entre deux lois. C'est une valeur positive, égale à zéro quand les lois p et q sont identiques. Dans la définition (1.27), le logarithme népérien est utilisé, l'unité est par conséquent le « nat ».

La figure 1.5 affiche les divergences de Kullback-Leibler entre respectivement les lois du spectrogramme $S_{\phi, x}$ (1.2) et du spectrogramme « rectifié » $S'_{\phi, x}$ (1.26) et la loi du χ^2 de degré deux qui nous sert de référence. Les divergences sont calculées à partir de 1000 réalisations du spectrogramme du signal $x[m] = b[m]$ de 2000 points.

Comme on le voit sur cette figure, la rectification du spectrogramme permet de mieux approximer la loi des coefficients du plan temps-fréquence par une loi du χ^2 de degré deux. Plus la taille de la fenêtre est grande, plus l'approximation est bonne, et moins les effets de la rectification se font sentir.

En conclusion, nous avons montré que les parties réelle et imaginaire de la TFCT d'un bruit blanc gaussien b suivent toutes deux une loi gaussienne centrée, de variance respectivement $\alpha[n, k]\sigma_b^2$ et $(1 - \alpha[n, k])\sigma_b^2$. En dehors des basses et hautes fréquences, ce paramètre $\alpha[n, k]$ vaut $\frac{1}{2}$, ce qui implique que les variances des parties réelle et imaginaire sont égales. Dans les hautes et basses fréquences, sur une bande de fréquence dépendant du choix des paramètres de la TFCT, ce paramètre diffère de $\frac{1}{2}$, une différence de variance entre les parties réelle et imaginaire apparaissent, ainsi qu'une corrélation. Cette corrélation et cette différence de variance impliquent que la loi des coefficients du spectrogramme à ces fréquences devient différente d'une loi du χ^2 de degré deux.

La corrélation entre coefficients peut disparaître avec une construction judicieuse de la TFCT et en utilisant une fenêtre d'analyse paire, tandis que la « rectification » des coefficients du spectrogramme pour égaliser la variance entre parties réelle et imaginaire permet d'améliorer l'approximation de la loi des coefficients du spectrogramme par une loi du χ^2 de degré deux.

1.2 Cas d'un signal quelconque noyé dans un bruit blanc stationnaire gaussien

Dans cette section, le signal étudié est celui de l'équation (1), c'est-à-dire $x[m] = s[m] + b[m]$. Le bruit b est blanc et gaussien de variance σ_b^2 , le signal d'intérêt s est inconnu et de propriétés statistiques quelconques, si ce n'est le fait qu'il ne doit pas être un bruit blanc gaussien stationnaire. En effet, il serait *de facto* intégré dans le bruit b , augmentant sa variance. L'équation (1.1) devient :

$$X_\phi[n, k] = \sum_m x[m] \phi[m - n] e^{\left(-2j\pi k \frac{m}{M_\phi + Z}\right)} = S_\phi[n, k] + B_\phi[n, k] \quad (1.28)$$

où $S_\phi[n, k]$ est la TFCT du signal s uniquement.

Le but de cette section est de déterminer, à partir des statistiques de $X_\phi[n, k]$, si un point temps-fréquence contient uniquement du bruit (1.5) ou un mélange de signal et de bruit (1.28). Le signal s est inconnu, et peut être un signal aléatoire centré. Dans ce cas l'espérance de $X_\phi[n, k] = S_\phi[n, k] + B_\phi[n, k]$ est nulle, et ne permet donc pas de savoir si ce point contient ou non du signal. Nous nous intéressons alors au moment d'ordre deux :

$$\begin{aligned} \text{MO2}(X_\phi[n, k]) &= \text{E}(X_\phi[n, k] X_\phi^*[n, k]) \\ &= \sum_m \sum_{m'} \text{E}(x[m] x[m']) \phi[m] \phi[m'] e^{\left(-2j\pi k \frac{m-m'}{M_\phi + Z}\right)} \\ &= \gamma_s[n, k]^2 + \sigma_b^2 \end{aligned} \quad (1.29)$$

Où $\gamma_s[n, k]^2$ est le moment d'ordre deux de la TFCT du signal inconnu s uniquement :

$$\begin{aligned} \gamma_s[n, k]^2 &= \text{E}(S_\phi[n, k] S_\phi^*[n, k]) \\ &= \sum_m \sum_{m'} \text{E}(s[m] s[m']) \phi[m - n] \phi[m' - n] e^{\left(2j\pi k \frac{m-m'}{M_\phi + Z}\right)} \end{aligned} \quad (1.30)$$

Dans le cas particulier où $s[m]$ est un signal déterministe $d[m]$, nous avons $\text{E}(s[m] s[m']) = d[m] d[m']$, l'équation (1.30) devient :

$$\gamma_s[n, k]^2 = |D_\phi[n, k]|^2 \quad (1.31)$$

Avec $D_\phi[n, k]$ la TFCT du signal déterministe uniquement.

La loi du bruit seul avec la TFCT rectifiée $X_\phi[n, k]'$ étant une loi du χ^2 , il peut être intéressant de s'intéresser au moment d'ordre deux de cette variable dans le cas « signal + bruit » :

$$\begin{aligned} \text{MO2}(X_\phi[n, k]') &= \text{E}(X_\phi^r[n, k]'^2 + X_\phi^i[n, k]'^2) \\ &= \gamma_s[n, k]'^2 + \sigma_b^2 \end{aligned} \quad (1.32)$$

Où $\gamma_s[n, k]'^2$ est le moment d'ordre deux de la TFCT normalisée du signal s uniquement :

$$\gamma_s[n, k]'^2 = \text{E}(S_\phi^r[n, k]'^2 + S_\phi^i[n, k]'^2) \quad (1.33)$$

Quelle que soit la nature du signal d'intérêt s , nous avons montré que le moment d'ordre deux de la TFCT normalisée du signal $X_\phi[n, k]'$ décrit par l'équation (1.32) est la somme de deux termes positifs, le premier dépendant uniquement du signal s , le second étant la variance du bruit σ_b . Par conséquent, un point temps-fréquence $[n, k]$ contenant de l'énergie du signal à segmenter aura un moment d'ordre deux plus grand qu'un point contenant du bruit uniquement.

Dans le cas où le signal à segmenter s est déterministe, la loi du spectrogramme au point $[n, k]$ devient une loi du χ^2 décentrée, où le paramètre de décentrage est la valeur du spectrogramme du signal s seul au point considéré [Hor02, Hui08]. Dans notre cas, la loi des points contenant du signal et du bruit est inconnue.

La section suivante propose à partir de ce constat une méthode permettant de détecter automatiquement les points temps-fréquence contenant du signal, en supposant le niveau de bruit connu.

Chapitre 2

Détection des points temps-fréquence contenant du signal

Dans le chapitre précédent, nous avons déterminé qu'un point temps-fréquence contenant uniquement du bruit a un moment d'ordre deux plus faible que s'il contient également une part du signal à segmenter. Il s'agit à présent de définir une méthode pour décider à partir d'une réalisation de la variable $X_\phi[n, k]$ si ce point $[n, k]$ est constitué de bruit uniquement, ce que nous nommerons par la suite l'hypothèse nulle \mathcal{H}_0 , ou s'il y a du signal en plus du bruit, l'hypothèse alternative \mathcal{H}_1 .

Dans une première section, nous allons présenter le cas classique du test binaire d'hypothèses, et diverses méthodes de résoudre ce problème. Ce problème a été abordé par les communautés mathématiques [LR05] et de traitement du signal [Kay98b] avec leur vocabulaire propre, nous utiliserons dans ce document celui utilisé par les traiteurs de signaux.

Une seconde partie présentera le critère de détection retenu, compte tenu des spécificités liées au modèle de signal choisi dans le cadre de ce travail.

2.1 Test binaire d'hypothèses

Un test binaire d'hypothèses consiste à construire une fonction de décision $d(x)$ permettant d'associer une observation de dimension finie x à une hypothèse parmi deux disponibles, notées \mathcal{H}_1 et \mathcal{H}_0 . Usuellement, l'hypothèse \mathcal{H}_0 est appelée l'hypothèse nulle, et correspond au cas où x est issue d'une réalisation d'un bruit. L'hypothèse alternative \mathcal{H}_1 correspond au cas où x est originaire d'un signal perturbé par le bruit.

Si l'observation x évolue dans un espace noté \mathcal{R} , le but du test est de partitionner cet espace en deux régions \mathcal{R}_0 et \mathcal{R}_1 , en définissant la fonction de décision $d(x)$ par :

$$d(x) = \begin{cases} 0 & \text{si } x \in \mathcal{R}_0 \\ 1 & \text{si } x \in \mathcal{R}_1. \end{cases} \quad (2.1)$$

L'hypothèse \mathcal{H}_0 sera décidée si $d(x) = 0$, l'hypothèse \mathcal{H}_1 sera retenue sinon.

Suite à cette prise de décision, quatre cas de figures peuvent apparaître. Les deux premiers correspondent aux cas de bonne détection :

- une bonne détection de signal, l'hypothèse \mathcal{H}_1 est correctement décidée ;
- une bonne détection de bruit seul, l'hypothèse \mathcal{H}_0 est correctement décidée ;

les deux autres correspondent aux erreurs possibles :

- une non détection de signal, l'hypothèse \mathcal{H}_0 est décidée alors que l'hypothèse \mathcal{H}_1 est vraie ;
- une fausse alarme, l'hypothèse \mathcal{H}_1 est décidée alors que seul du bruit (hypothèse \mathcal{H}_0) était présent.

En notant $p_{\mathcal{H}_0}(x)$ et $p_{\mathcal{H}_1}(x)$ les densités de probabilité des observations x conditionnelles aux hypothèses \mathcal{H}_0 et \mathcal{H}_1 respectivement, on peut définir les probabilités associées aux quatre situations suivant la détection :

- probabilité de bonne détection de signal P_{ds}

$$P_{ds} = \int_{\mathcal{R}_1} p_{\mathcal{H}_1}(x) dx, \quad (2.2)$$

- probabilité de bonne détection de bruit seul P_{db}

$$P_{db} = \int_{\mathcal{R}_0} p_{\mathcal{H}_0}(x) dx, \quad (2.3)$$

- probabilité de non détection de signal P_{nd}

$$P_{nd} = \int_{\mathcal{R}_0} p_{\mathcal{H}_1}(x) dx, \quad (2.4)$$

- probabilité de fausse alarme P_{fa}

$$P_{fa} = \int_{\mathcal{R}_1} p_{\mathcal{H}_0}(x) dx. \quad (2.5)$$

La figure 2.1 illustre ces différentes probabilités dans un cas à une dimension. Dans le cas où les courbes $p_{\mathcal{H}_0}(x)$ et $p_{\mathcal{H}_1}(x)$ se chevauchent, il est évident qu'il n'existe pas de seuil de détection tel que les probabilités d'erreur soient conjointement nulles. Des compromis sont à faire afin de choisir le seuil de détection.

Dans un cas général, le choix d'une fonction de décision consiste à donner une importance relative à chacune de ces quatre issues. Pour ce faire, nous allons présenter deux approches, l'approche bayésienne et celle de Neyman-Pearson.

2.1.1 L'approche bayésienne

L'approche bayésienne consiste à minimiser le risque du test, défini comme l'espérance du coût de la décision. Ce risque bayésien est basé sur un coût attribué à chaque issue de la décision. Des coûts C_{00} , C_{11} , C_{10} et C_{01} sont respectivement associés à une bonne détection de bruit, à une bonne détection de signal, à une non détection de signal et à une fausse alarme.

Le risque bayésien r est alors défini par [Tre68] :

$$r = p_0 C_{00} P_{db} + p_1 C_{11} P_{ds} + p_1 C_{10} P_{nd} + p_0 C_{01} P_{fa} \quad (2.6)$$

Où p_0 et p_1 sont respectivement les probabilités *a priori* d'occurrence des hypothèses \mathcal{H}_0 et \mathcal{H}_1 . En remarquant que d'après les définitions (2.2, 2.4) et (2.3, 2.5), nous avons $P_{ds} = 1 - P_{nd}$ et $P_{db} = 1 - P_{fa}$, le risque bayésien (2.6) s'écrit sous la forme :

$$r = p_1 C_{11} + p_0 C_{10} + \int_{\mathcal{R}_0} (p_0 (C_{00} - C_{10}) p_{\mathcal{H}_0}(x) + p_1 (C_{01} - C_{11}) p_{\mathcal{H}_1}(x)) dx \quad (2.7)$$

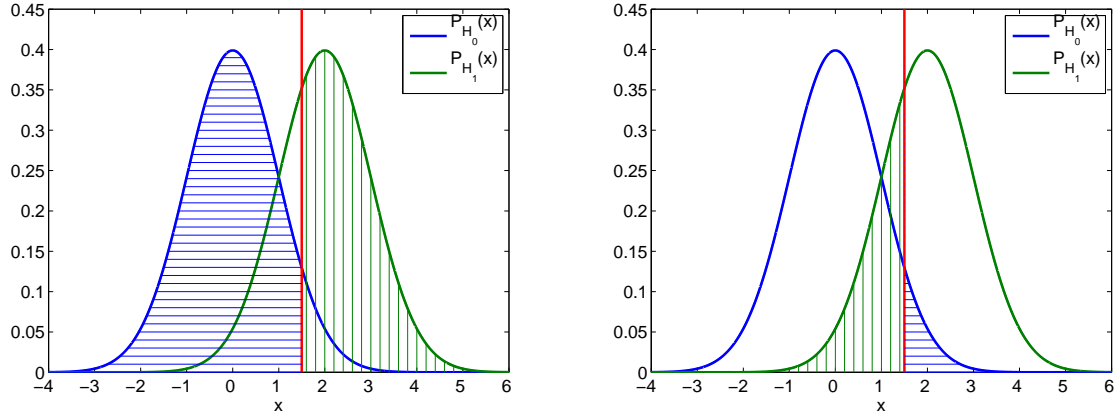


FIG. 2.1 : Illustration des probabilités de bonne et de mauvaise détection. L'observation x suit une loi gaussienne centrée de variance 1 sous l'hypothèse \mathcal{H}_0 (courbes bleues), une loi gaussienne de moyenne 2 et de variance 1 sous l'hypothèse \mathcal{H}_1 (courbes vertes). L'espace des réalisations de x est \mathbb{R} , partitionné en deux sous régions délimité par un seuil arbitraire de 1.5 (ligne rouge) : la région \mathcal{R}_0 est l'intervalle $]-\infty, 1.5]$, la région \mathcal{R}_1 l'intervalle $]1.5, +\infty[$. Le test d'hypothèse associera une observation inférieure ou égale à 1.5 à l'hypothèse \mathcal{H}_0 , à l'hypothèse \mathcal{H}_1 sinon. À gauche, les probabilités de bonne décision : en hachures verticales la probabilité de bonne détection de signal P_{ds} (2.2), en hachures horizontales, la probabilité de bonne détection du bruit P_{db} (2.3). À droite, les probabilités de mauvaise décision : en hachures verticales la probabilité de non détection P_{nd} (2.4), en hachures horizontales la probabilité de fausse alarme P_{fa} (2.5).

Déterminer les régions \mathcal{R}_0 et \mathcal{R}_1 se fait en minimisant le risque bayésien r , ce qui revient à ne garder sous l'intégrale de l'équation (2.7) que les termes négatifs. En supposant que les coûts liés aux erreurs C_{01} et C_{10} sont supérieurs aux coûts liés aux bonnes décisions C_{00} et C_{11} , la région \mathcal{R}_0 est l'ensemble des observations x telles que :

$$p_1(C_{01} - C_{11})p_{\mathcal{H}_1}(x) - p_0(C_{10} - C_{00})p_{\mathcal{H}_0}(x) < 0 \quad (2.8)$$

La fonction de décision se met alors sous la forme d'un test sur le rapport de vraisemblance des hypothèses :

$$\mathcal{L}(x) = \frac{p_{\mathcal{H}_1}(x)}{p_{\mathcal{H}_0}(x)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{p_0(C_{10} - C_{00})}{p_1(C_{01} - C_{11})} \quad (2.9)$$

La fonction de décision consiste à comparer le rapport de vraisemblance des hypothèses \mathcal{H}_0 et \mathcal{H}_1 avec un seuil défini par les connaissances *a priori* des probabilités de ces hypothèses, et des coûts associés à chacune des issues de la décision.

Cette solution générale (2.9) se décline en plusieurs cas particuliers.

Lorsqu'aucune contrainte n'existe sur les coûts, il est classique de mettre les coûts des bonnes détections à zéro, c'est-à-dire $C_{00} = C_{11} = 0$, et un coût égal aux deux mauvaises décisions, $C_{10} = C_{01}$. Le risque bayésien n'est alors constitué que des probabilités de fausse alarme et de non détection, minimiser ce risque revient à minimiser la probabilité d'erreur.

Le test de décision s'écrit alors :

$$\frac{p_{\mathcal{H}_1}(x)}{p_{\mathcal{H}_0}(x)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \frac{p_0}{p_1} \quad (2.10)$$

Ce qui se réécrit en utilisant la règle de Bayes :

$$p(\mathcal{H}_1(x)|x) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} p(\mathcal{H}_0(x)|x) \quad (2.11)$$

Où $p(\mathcal{H}_i(x)|x)$ désigne la probabilité de \mathcal{H}_i conditionnellement à x . Autrement dit, la fonction de décision consiste dans ce contexte à garder l'hypothèse la plus probable compte tenu de l'observation x . De ce fait, ce critère est appelé « maximum *a posteriori* ».

Lorsque les probabilités *a priori* p_0 et p_1 ne sont pas connues, il n'y a pas de raison de ne pas les considérer égales, c'est-à-dire qu'on suppose que les deux hypothèses \mathcal{H}_0 et \mathcal{H}_1 sont équiprobables. Dans ce nouveau cas, le test de décision s'écrit :

$$\frac{p_{\mathcal{H}_1}(x)}{p_{\mathcal{H}_0}(x)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} 1 \quad (2.12)$$

On décide l'hypothèse dont la vraisemblance est la plus grande pour l'observation x . Ce critère est alors nommé « maximum de vraisemblance ».

Cette approche nécessite dans tous les cas la connaissance des densités de probabilités de l'observation x sous les deux hypothèses \mathcal{H}_0 et \mathcal{H}_1 .

2.1.2 L'approche de Neyman-Pearson

L'approche bayésienne consiste à minimiser un risque relatif à la décision, tenant compte de toutes les issues possibles de la décision. L'approche de Neyman-Pearson cherche quant à elle à minimiser la probabilité d'erreur P_e , définie par :

$$P_e = p_0 P_{fa} + p_1 P_{nd} \quad (2.13)$$

Minimiser les probabilités de fausse alarme et de non détection à la fois est impossible : diminuer P_{fa} en réduisant la région \mathcal{R}_0 (2.5) diminue également la probabilité de détection du signal (2.2), et donc augmente P_{nd} . La stratégie de Neyman-Pearson consiste à déterminer les régions de décisions \mathcal{R}_0 et \mathcal{R}_1 sous la contrainte de garder la probabilité de fausse alarme égale à une valeur α , tout en minimisant la probabilité d'erreur [Lan02].

Ce problème se résout en utilisant le multiplicateur de Lagrange, la fonction à minimiser est :

$$F = \lambda(P_{fa} - \alpha) + P_{nd}, \quad \text{où } \lambda > 0 \text{ est un multiplicateur de Lagrange.} \quad (2.14)$$

F s'écrit en fonction des équations (2.4) et (2.5) :

$$F = \lambda(1 - \alpha) + \int_{\mathcal{R}_0} (p_{\mathcal{H}_1}(x) - \lambda p_{\mathcal{H}_0}(x)) dx \quad (2.15)$$

Comme précédemment, minimiser la fonction F revient à définir la région \mathcal{R}_0 telle qu'elle regroupe les observations de x où le contenu de l'intégrale est négatif. La région \mathcal{R}_1 regroupe alors les autres points. Le test de décision revient alors à un test du rapport de vraisemblance :

$$\frac{p_{\mathcal{H}_1}(x)}{p_{\mathcal{H}_0}(x)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda \quad (2.16)$$

Le seuil λ correspond détermine la région \mathcal{R}_1 , suivant la contrainte :

$$P_{fa} = \int_{\mathcal{R}_1} p_{\mathcal{H}_0}(x) dx = \alpha \quad (2.17)$$

Ainsi, le seuil λ n'est défini qu'en fonction de la probabilité de l'hypothèse nulle \mathcal{H}_0 , contrairement à l'approche bayésienne qui se base sur le rapport de vraisemblance lui-même, et donc des probabilités des deux hypothèses. De plus, une fois les régions \mathcal{R}_0 et \mathcal{R}_1 déterminées, l'utilisation du rapport de vraisemblance n'est plus nécessaire : il suffit de regarder à quelle région appartient l'observation x .

Cette méthode a toutefois un inconvénient : il n'y a aucun contrôle quant à la probabilité de bonne détection de signal.

2.2 Détection dans la TFCT rectifiée

Dans le cadre de la segmentation du plan temps-fréquence, nous cherchons à déterminer si un point $[n, k]$ contient uniquement du bruit b ou également une part de signal à segmenter s . Nous avons donc deux hypothèses :

$$\begin{cases} \text{l'hypothèse nulle } \mathcal{H}_0 : & X_\phi[n, k]' = B_\phi[n, k]' \\ \text{l'hypothèse alternative } \mathcal{H}_1 : & X_\phi[n, k]' = S_\phi[n, k]' + B_\phi[n, k]' \end{cases} \quad (2.18)$$

Où $X_\phi[n, k]'$ désigne la TFCT rectifiée définie en (1.24,1.25).

Dans notre cas, la loi sous l'hypothèse \mathcal{H}_1 est inconnue, tandis que nous connaissons parfaitement la loi du bruit uniquement, qui est une loi du χ^2 de degré deux¹. Nous adoptons alors comme méthode de détection l'approche de Neyman-Pearson, qui consiste à maximiser la probabilité de détection en fixant la probabilité de fausse alarme à une valeur arbitraire p_{fa} .

Le chapitre précédent nous a montré que le moment d'ordre deux d'un coefficient de la TFCT rectifiée est plus élevé en présence de bruit : en effet les équations (1.6) et (1.32) nous indiquent qu'un point temps-fréquence aura un plus grand moment d'ordre deux s'il contient du bruit que s'il n'en contient pas. Le moment d'ordre deux étant une variable positive, l'espace des réalisations \mathcal{R} de la variable aléatoire $|X_\phi[n, k]'|^2$ sera \mathbb{R}^+ .

Le moment d'ordre deux d'un point contenant du signal, autrement dit sous l'hypothèse \mathcal{H}_1 , est plus grand que celui d'un point sous l'hypothèse \mathcal{H}_0 : l'espace des réalisations sera divisé en deux sous espaces, $\mathcal{R}_0 = [0, s_{\sigma_b^2}]$ et $\mathcal{R}_1 =]s_{\sigma_b^2}, \infty[$. Le problème consiste à définir le seuil de détection $s_{\sigma_b^2}$.

¹Excepté aux points où $\alpha[n, k] = 0$ ou 1. Dans ces cas, la loi du bruit seul est également une loi connue, un χ^2 de degré 1. Ce cas sera ignoré dans la suite du manuscrit, pour alléger le texte.

En utilisant la stratégie de détection de Neyman-Pearson qui fixe une probabilité de fausse alarme à p_{fa} , le seuil de détection se définit par :

$$\begin{aligned} s_{\sigma_b^2} & / \text{Prob}(|X_\phi[n, k]'|^2 > s_{\sigma_b^2}) = p_{fa} \\ s_{\sigma_b^2} & / \int_{s_{\sigma_b^2}}^{\infty} p_{\mathcal{H}_0}(x) dx = p_{fa} \end{aligned} \quad (2.19)$$

En considérant les variables de la TFCT rectifiée du bruit seul, chaque coefficient $X_\phi[n, k]'$ est circulaire, c'est-à-dire que ses parties réelle et imaginaire suivent une même loi gaussienne de variance $\frac{\sigma_b^2}{2}$ et sont décorrélées.

La loi du χ^2 de degré de liberté δ est définie comme la densité de probabilité d'une variable aléatoire constituée de la somme des carrés de δ variables indépendantes de même loi gaussienne de moyenne nulle et de variance 1, et s'exprime :

$$f(x) = \begin{cases} \frac{1}{2^{\delta/2}\Gamma(\delta/2)} x^{\delta/2-1} e^{-x/2} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases} \quad (2.20)$$

Le moment d'ordre deux d'un coefficient de la TFCT rectifiée d'un bruit seul est la somme de ses parties réelle et imaginaire élevées au carré, qui sont indépendantes et de même variance $\frac{\sigma_b^2}{2}$. Le degré de liberté est par définition $\delta = 2$. La densité de probabilité associée $p_{\mathcal{H}_0}$ s'écrit alors :

$$p_{\mathcal{H}_0} = \begin{cases} \frac{1}{\sigma_b^2} e^{-x/\sigma_b^2} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases} \quad (2.21)$$

Le seuil de détection $s_{\sigma_b^2}$ (2.19) s'exprime directement en fonction du niveau de bruit σ_b^2 :

$$s_{\sigma_b^2} / \int_{s_{\sigma_b^2}}^{\infty} \frac{1}{\sigma_b^2} e^{-x/\sigma_b^2} dx = p_{fa} \quad (2.22)$$

$$s_{\sigma_b^2} = \sigma_b^2 \ln(p_{fa}^{-1}) \quad (2.23)$$

Ce seuil est directement proportionnel à la variance du bruit du modèle « signal + bruit » (équation (1) de l'introduction), et au logarithme de l'inverse de la probabilité de fausse alarme p_{fa} choisie. La figure 2.2 illustre l'influence de ces deux paramètres sur la valeur du seuil $s_{\sigma_b^2}$. Ce seuil s'appuie sur le moment d'ordre deux de la TFCT rectifiée, autrement dit sur la valeur du spectrogramme. Un point temps-fréquence $[n, k]$ sera considéré comme contenant une part du signal à segmenter si la valeur du spectrogramme rectifié $S_{\phi, x}[n, k]'$ est supérieure au seuil $s_{\sigma_b^2}$.

Dans le cas où le signal s est déterministe, l'hypothèse \mathcal{H}_1 est connue pour chaque point temps-fréquence $[n, k]$. Le moment d'ordre deux de $X_\phi[n, k]' = B_\phi[n, k]' + S_\phi[n, k]'$ suit une loi du χ^2 de degré deux, de coefficient de proportionnalité $\sigma_b^2/2$ et de paramètre de décentrage de $S_{\phi, s}[n, k]^2$, qui s'exprime [JKB94] :

$$p_{\mathcal{H}_1}(x) = \begin{cases} \frac{1}{\sigma_b^2} \exp\left(-\frac{x+S_{\phi, s}[n, k]^2}{\sigma_b^2}\right) I_0\left(\frac{2\sqrt{xS_{\phi, s}[n, k]^2}}{\sigma_b^2}\right) & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases} \quad (2.24)$$

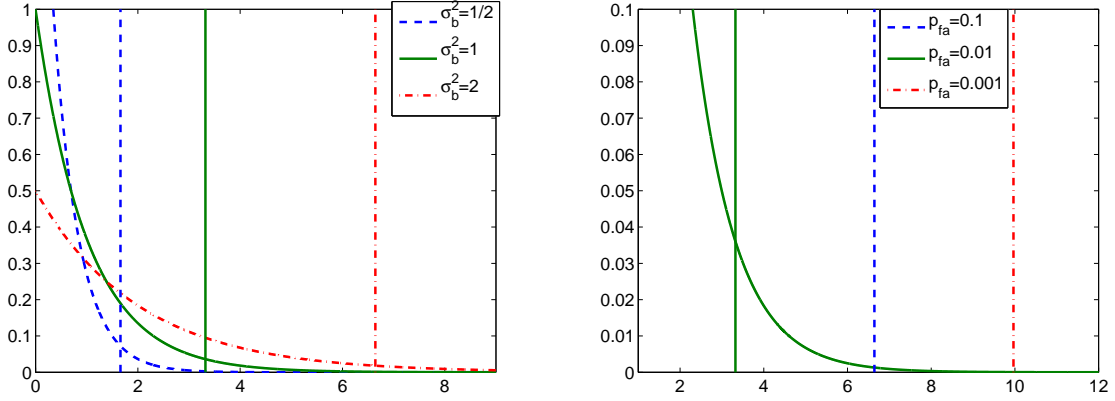


FIG. 2.2 : Illustration du seuil selon la loi du bruit seul et la probabilité de fausse alarme p_{fa} choisie. À gauche, influence du niveau de bruit σ_b^2 , pour $p_{fa} = 0.1$. À droite, influence de p_{fa} , pour $\sigma_b^2 = 1$.

Huillery [Hui08] montre que le seuil de détection, qui varie alors pour chaque point temps-fréquence, s'écrit en fonction du Rapport Signal sur Bruit (RSB) local $\xi[n, k] = S_{\phi, s}[n, k]/\sigma_b^2$:

$$s_{\sigma_b^2, MV}[n, k] = \sigma_b^2(0.26\xi[n, k] + 1) \quad (2.25)$$

Toutefois, le RSB local est une inconnue. Choisir un seuil de détection revient à choisir un RSB local *de focus* ξ_0 . Les points temps-fréquence dont le RSB local sera supérieur à ce RSB de focus seront considérés comme contenant du signal. Le lien entre le choix du RSB de focus et probabilité de fausse alarme pour déterminer un seuil de détection est immédiat (2.23, 2.25) :

$$\ln(p_{fa}^{-1}) = 0.26\xi_0 + 1 \quad (2.26)$$

De par son côté intuitif, ce critère de RSB de focus peut aider à choisir une probabilité de fausse alarme. Dans ces travaux, étant donné le caractère potentiellement stochastique du signal à segmenter s , nous nous contenterons du critère de Neyman-Pearson et de la probabilité de fausse alarme qui y est liée.

En conclusion, compte tenu du modèle général choisi pour décrire le signal temporel (équation (1) de l'introduction) le choix d'une fonction de décision se fait par un critère de Neyman-Pearson, à probabilité de fausse alarme choisie. Il en résulte un seuil de détection portant sur les moments d'ordre deux de la TFCT rectifiée, c'est-à-dire sur les coefficients du spectrogramme rectifié. Ce seuil est directement proportionnel au niveau de bruit σ_b^2 du bruit du modèle, et au logarithme népérien de l'inverse de la probabilité de fausse alarme p_{fa} choisie. Ce dernier paramètre résulte d'un choix de segmentation, tandis que la puissance du bruit est intrinsèque au signal à analyser : le chapitre suivant propose une méthode itérative afin de l'estimer.

Chapitre 3

Estimation itérative du bruit

Le chapitre précédent a fourni une méthode de détection des points temps-fréquence contenant du signal, à condition de connaître la puissance du bruit blanc additif b . Par hypothèse, la variance du bruit b du modèle du signal (1) est inconnue. Il faut donc l'estimer afin de déterminer le seuil σ_b^2 *via* l'équation (2.19). Diverses méthodes ont déjà été utilisées pour ce type de problème. On peut notamment citer l'utilisation des statistiques minimales [Mar06, Hui08], considérant que les Z plus faibles valeurs du spectrogramme ne proviennent que du bruit et les méthodes itératives [Hor02, YR06], qui à chaque itération éliminent des points contenant du signal et estiment le niveau de bruit sur l'ensemble des points restants. Comme l'indique le titre de ce chapitre, ce deuxième choix a été retenu.

À l'instar de la détection, la théorie de l'estimation est un domaine très classique du traitement du signal. Dans une première section, nous présenterons les bases de l'estimation d'un paramètre déterministe quelconque. Une estimation de paramètre aléatoire peut être trouvée par exemple dans [Kay98a]. Ce principe sera ensuite appliqué dans le cas de l'estimation de la variance du bruit, considérée comme un paramètre déterministe, dans le plan temps-fréquence en présence de signal.

Cette estimation nous conduira à une estimation itérative du niveau de bruit similaire à celle développée par Hory [Hor02], permettant successivement de détecter des points contenant du signal. Hory prenait comme critère d'arrêt la convergence de la vraisemblance des caractéristiques des points non détectés comme contenant du signal, c'est-à-dire les points supposés n'être issus que du bruit. Dans ce travail, nous remplaçons ce critère par une estimation du kurtosis de l'ensemble des points supposés ne contenir que du bruit. Le bruit étant gaussien, le kurtosis des coefficients temps-fréquence ne contenant que du bruit est censé être nul, ce qui permet d'avoir un critère absolu indépendant d'une convergence variable selon les itérations.

Ainsi, une seconde section présentera brièvement les statistiques d'ordre supérieur dans le cas d'une variable aléatoire réelle et complexe ainsi que leurs estimations. L'utilisation du kurtosis dans le cadre de la segmentation du plan temps-fréquence sera ensuite illustrée, afin de déterminer un critère d'arrêt aux itérations d'estimations et de détections. L'influence de la non circularité de la TFCT sera discutée. Le critère d'arrêt étant un seuil sur la valeur estimée du kurtosis des points non segmentés, une estimation automatique de ce seuil sera proposée.

3.1 Estimation : principe et définitions

Le problème de l'estimation se pose dès que l'on cherche un paramètre θ , caractérisant la densité de probabilité $p_X(x; \theta)$ d'une variable aléatoire X , à partir d'un ensemble de N réalisations de cette variable aléatoire, notés $\{x_1, \dots, x_N\}$. Le paramètre θ peut être une grandeur à une ou plusieurs dimensions, et ainsi représenter un ensemble de plusieurs paramètres de la densité de probabilité de X .

On appelle estimation du paramètre θ la fonction $\hat{\theta}$ des observations x_i :

$$\hat{\theta} = f(x_1, \dots, x_N) \quad (3.1)$$

Cette fonction d'estimation est issue d'un ensemble de réalisations d'une variable aléatoire, la valeur estimée est par conséquent elle aussi une variable aléatoire. Plusieurs critères de qualité caractérisent un estimateur.

Le biais $b(\hat{\theta})$ est défini comme la différence entre l'espérance de l'estimation et la valeur recherchée.

$$b(\hat{\theta}) = E(\hat{\theta}) - \theta \quad (3.2)$$

Un estimateur sera dit sans biais si $b(\hat{\theta}) = 0$, c'est à dire que l'espérance de l'estimateur est la valeur recherchée. Il sera *asymptotiquement non biaisé* si le biais tend vers zéro quand le nombre d'échantillon N tend vers l'infini.

L'erreur quadratique moyenne $EQM(\hat{\theta})$ est l'espérance de la différence au carré entre la valeur estimée et la valeur recherchée.

$$EQM(\hat{\theta}) = E\left((\hat{\theta} - \theta)^2\right) \quad (3.3)$$

Cette erreur, qui caractérise la dispersion de la valeur estimée autour de la valeur recherchée, est intimement liée à la variance de l'estimateur :

$$\begin{aligned} \text{Var}(\hat{\theta}) &= E\left((\hat{\theta} - E(\hat{\theta}))^2\right) \\ EQM(\hat{\theta}) &= \text{Var}(\hat{\theta}) + b(\hat{\theta})^2 \end{aligned} \quad (3.4)$$

Un estimateur sera dit plus *efficace* qu'un autre si son erreur quadratique moyenne est inférieure.

Un estimateur est dit consistant s'il tend vers la valeur à estimer θ lorsque le nombre N d'observations tend vers l'infini. Autrement dit, un estimateur est consistant s'il est non biaisé ou asymptotiquement non biaisé, et si son erreur quadratique moyenne tend vers zéro quand le N tend vers l'infini.

La borne de Cramér-Rao, notée $BCR(\hat{\theta})$, dans sa formulation concernant les estimateurs non biaisés, est la variance minimale de tout estimateur non biaisé de θ [Kay98a]. La borne de Cramér-Rao est définie par l'inverse de l'information de Fisher $I(\theta)$, soit :

$$BCR(\hat{\theta}) = \frac{1}{I(\theta)} = - \left(E \left(\frac{\partial^2 \ln p_X(x; \theta)}{\partial \theta^2} \right) \right)^{-1} \quad (3.5)$$

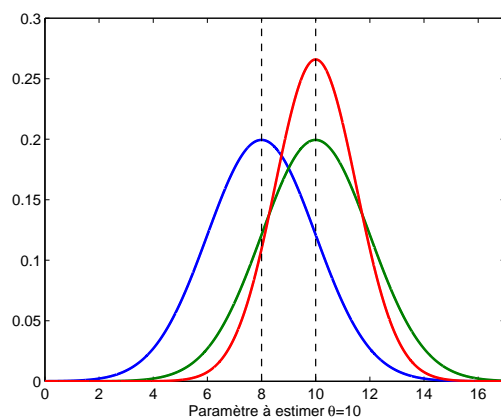


FIG. 3.1 : Illustrations des densités de probabilité de trois exemples d'estimateur d'un paramètre $\theta = 10$. La courbe bleue présente un estimateur biaisé, la verte un estimateur non biaisé et la rouge un estimateur non biaisé plus efficace que le précédent.

Où \ln est le logarithme népérien, $p_X(x; \theta)$ la densité de probabilité de X , en fonction du paramètre θ .

Ainsi, tout estimateur non biaisé de θ a une variance supérieure ou égale à cette borne. L'estimateur dont la variance est égale à la borne de Cramér-Rao est dit *efficace*.

La figure 3.1 illustre ces différentes caractéristiques pour trois exemples d'estimateur d'un même paramètre scalaire.

De nombreuses méthodes existent pour construire un estimateur : en recherchant un estimateur de variance minimale, en se limitant à une combinaison linéaire des observations, etc. [Kay98a]. La section suivante propose un estimateur au sens du maximum de vraisemblance de la variance d'une variable aléatoire complexe suivant une loi gaussienne centrée, ce qui correspond à l'hypothèse « bruit seul » dans le plan temps-fréquence. Quand ce type d'estimateur existe, son biais tend vers zéro et sa variance tend vers la borne de Cramér-Rao à mesure que le nombre d'échantillons augmente.

3.1.1 Estimation par maximum de vraisemblance dans le cas du bruit seul

Dans cette section, nous cherchons à estimer la variance d'une variable aléatoire issue d'une loi gaussienne complexe centrée circulaire. En effet, dans le cas de la TFCT « rectifiée » d'un bruit blanc¹ (1.25), les coefficients complexes $B_\phi[n, k]'$ suivent tous une même loi gaussienne circulaire. Le paramètre inconnu à estimer est la variance du bruit $\theta = \sigma_b^2$.

$B_\phi[n, k]'$ est une variable complexe, que l'on peut représenter comme une variable réelle à deux dimensions, sa partie réelle et sa partie imaginaire. Ces deux dimensions étant indépendantes, la densité de probabilité de $B_\phi[n, k]'$ est le produit des densités de probabilité de ses parties réelle et imaginaire.

¹Comme dans le chapitre précédent, les points $[n, k]$ strictement réels ou imaginaires, correspondant aux cas $\alpha[n, k] = 0$ ou 1, seront ignorés.

$$p_{B'_\phi}(B_\phi[n, k]'; \sigma_b^2) = p_{B'_\phi}(B_\phi^r[n, k]'; \sigma_b^2) \times p_{B'_\phi}(B_\phi^i[n, k]'; \sigma_b^2) \quad (3.6)$$

$$= \frac{1}{\sigma_b/\sqrt{2}\sqrt{2\pi}} e^{-\frac{(B_\phi^r[n, k]')^2}{\sigma_b^2}} \times \frac{1}{\sigma_b/\sqrt{2}\sqrt{2\pi}} e^{-\frac{B_\phi^i[n, k]'^2}{\sigma_b^2}} \quad (3.7)$$

$$= \frac{1}{\pi\sigma_b^2} e^{-\frac{B_\phi^r[n, k]'^2 + B_\phi^i[n, k]'^2}{\sigma_b^2}} \quad (3.8)$$

Cette loi est la même pour tout point $[n, k]$ du plan temps-fréquence qui contient $N \times K$ coefficients. Nous cherchons à estimer le paramètre σ_b^2 à partir de NK observations de B'_ϕ .

La vraisemblance du paramètre σ_b^2 à estimer en fonction des NK observations s'écrit :

$$\mathcal{L}(\sigma_b^2) = \prod_{n=1, k=1}^{n=N, k=K} p_{B'_\phi}(B_\phi[n, k]'; \sigma_b^2) \quad (3.9)$$

Maximiser cette fonction de vraisemblance donne une estimation du paramètre recherché qui possède des propriétés intéressantes [Kay98a] : les estimateurs ainsi construits sont asymptotiquement non biaisés, et atteignent asymptotiquement la borne de Cramér-Rao.

Puisque le logarithme est une fonction monotone, maximiser la vraisemblance est identique à maximiser son logarithme. Il est ainsi plus aisé de maximiser ce dernier, aussi appelé *log-vraisemblance*, qui s'écrit :

$$\ln \mathcal{L}(\sigma_b^2) = \sum_{n=1, k=1}^{n=N, k=K} \ln p_{B'_\phi}(B_\phi[n, k]'; \sigma_b^2) \quad (3.10)$$

$$= -NK \ln(\pi) - NK \ln(\sigma_b^2) - \frac{1}{\sigma_b^2} \sum_{n=1, k=1}^{n=N, k=K} (B_\phi^r[n, k]'^2 + B_\phi^i[n, k]'^2) \quad (3.11)$$

L'estimation $\hat{\sigma}_b^2$ au sens du maximum de vraisemblance est la valeur de σ_b^2 qui maximise l'équation (3.11), qu'on détermine en calculant les valeurs nulles de la dérivée de la log-vraisemblance par rapport à σ_b^2 :

$$\frac{\partial \ln \mathcal{L}(\sigma_b^2)}{\partial \sigma_b^2} = -\frac{NK}{\sigma_b^2} + \frac{1}{(\sigma_b^2)^2} \sum_{i=1}^{NK} (B_\phi^r[n, k]'^2 + B_\phi^i[n, k]'^2) \quad (3.12)$$

L'estimateur au sens du maximum de vraisemblance du niveau de bruit, sous réserve qu'il est non nul, s'écrit :

$$\hat{\sigma}_b^2 = \frac{1}{NK} \sum_{n, k} (B_\phi^r[n, k]'^2 + B_\phi^i[n, k]'^2) \quad (3.13)$$

Cet estimateur a un biais nul :

$$b(\hat{\sigma}_b^2) = E\left(\frac{1}{NK} \sum_{n, k} (B_\phi^r[n, k]'^2 + B_\phi^i[n, k]'^2) - \sigma_b^2\right) \quad (3.14)$$

$$= \frac{1}{NK} \sum_{n, k} (E(B_\phi^r[n, k]'^2 + B_\phi^i[n, k]'^2)) - \sigma_b^2 \quad (3.15)$$

$$= \frac{1}{NK} \sum_{n, k} (\sigma_b^2) - \sigma_b^2 = 0 \quad (3.16)$$

L'étude de sa variance est plus délicate : en effet, selon les paramètres de la TFCT, notamment le complément de zéro et le recouvrement entre deux fenêtres consécutives, les observations $B_\phi[n, k]$ présenteront une corrélation entre les différentes réalisations $[n, k]$. Nous nous contenterons ici de regarder le cas où les échantillons sont indépendants. Dans le cas d'échantillons corrélés, l'estimateur verra sa variance augmenter.

Dans ce cas, on peut remarquer d'après l'équation (3.13) que $\hat{\sigma}_b^2$ suit à un facteur $\frac{1}{NK}$ près une loi du χ^2 à $2NK$ degrés de liberté, provenant de NK réalisations d'une loi gaussienne de variance σ_b^2 . La variance de l'estimateur est donc [JKB94] :

$$\text{Var}(\hat{\sigma}_b^2) = \frac{1}{(NK)^2} \times \frac{\sigma_b^4}{4} 4NK \quad (3.17)$$

$$= \frac{\sigma_b^4}{NK} \quad (3.18)$$

L'estimateur ainsi créé est consistant, puisque sa variance tend vers zéro quand le nombre d'observations NK tend vers l'infini.

Pour vérifier son efficacité, on compare sa variance à sa borne de Cramér-Rao (3.5) :

$$BCR(\hat{\sigma}_b^2) = \left(\mathbb{E} \left(\frac{\partial^2 \ln \mathcal{L}(\sigma_b^2)}{\partial (\sigma_b^2)^2} \right) \right)^{-1} \quad (3.19)$$

$$= - \left(\frac{NK}{\sigma_b^4} - \frac{2}{\sigma_b^6} \mathbb{E} \left(\sum_{n=1, k=1}^{n=N, k=K} (B_\phi^r[n, k]^2 + B_\phi^i[n, k]^2) \right) \right)^{-1} \quad (3.20)$$

$$= - \left(\frac{-NK}{\sigma_b^4} \right)^{-1} = \frac{\sigma_b^4}{NK} \quad (3.21)$$

Dans le cadre d'une estimation du niveau de bruit à partir de la TFCT du bruit seul, l'estimateur MV (3.13) est non biaisé et a une variance égale à sa borne de Cramér-Rao, il est donc efficace. La section suivante applique cet estimateur dans le cadre du modèle « signal + bruit ».

3.1.2 Estimation par maximum de vraisemblance dans le cas signal et bruit

Dans le cas général, la TFCT est un mélange de points contenant uniquement du bruit et de points contenant du signal en plus du bruit. Le moment d'ordre deux d'un point temps-fréquence contenant du signal est supérieur à celui d'un point ne contenant que du bruit. En utilisant les statistiques minimales du plan temps-fréquence, il est possible d'estimer le niveau de bruit à partir d'échantillons ne contenant que du bruit, en supposant que les coefficients de plus faibles moments d'ordre deux ne contiennent aucune contribution énergétique du signal à segmenter [Mar06, Hui08].

Dans ce travail, nous considérons une autre approche. L'algorithme proposé est initié à partir de l'ensemble des points temps-fréquence, en prenant en considération le fait que certains points

contiennent du signal. À l'initialisation, l'estimateur (3.13) devient dans ce cas biaisé :

$$\begin{aligned}
 E(\widehat{\sigma_b^2}) &= E\left(\frac{1}{NK} \sum (X_\phi^r[n, k]^2 + X_\phi^i[n, k]^2)\right) \\
 &= \frac{1}{NK} \left(E(\sum B_\phi^r[n, k]^2 + B_\phi^i[n, k]^2) + E(\sum S_\phi^r[n, k]^2 + S_\phi^i[n, k]^2) \dots \right. \\
 &\quad \left. + 2E(\sum B_\phi^r[n, k]' S_\phi^r[n, k]' + B_\phi^i[n, k]' S_\phi^i[n, k]') \right) \\
 &= \sigma_b^2 + \frac{1}{NK} \sum \gamma_s[n, k]^2
 \end{aligned}$$

En considérant l'ensemble des points temps-fréquence, l'estimateur (3.24) surestime le niveau du bruit σ_b^2 , proportionnellement à l'énergie du signal s .

Pour pallier ce problème, nous proposons une méthode itérative. L'idée est de se servir d'un premier niveau de bruit surestimé afin de déterminer un seuil de détection (2.19), qui permet de détecter une première série de points temps-fréquence contenant de l'énergie du signal s .

Sachant que ces points biaisent l'estimation de σ_b^2 , le niveau de bruit est réestimé dans une nouvelle itération sans tenir compte de ces points contenant du signal. Le niveau de bruit est ainsi moins surestimé, ce qui donne un seuil de détection plus bas, et permet de segmenter de nouveaux points contenant du signal. Le processus se poursuit jusqu'à ce que tous les points contenant du signal soient segmentés.

Reste donc à déterminer un critère permettant de déterminer lorsque l'ensemble des points non segmentés ne contiennent que du bruit. Le kurtosis semble être une bonne piste, puisque sa valeur est nulle lorsqu'il est calculé à partir d'échantillons issus d'une variable aléatoire gaussienne. La prochaine section présente les statistiques d'ordre supérieur dont est issu le kurtosis, afin de l'utiliser comme critère d'arrêt.

3.2 Critère d'arrêt de l'estimation itérative : le kurtosis

3.2.1 Statistiques d'ordre supérieur

Les statistiques d'ordre supérieur sont utilisées essentiellement en complément aux statistiques usuelles d'ordre un et deux, en permettant de compléter la caractérisation d'une variable aléatoire X . En notant $p_X(u)$ la densité de probabilité de la variable X , les moments généralisés sont définis pour toute fonction réelle $g(u)$ par [LAP97] :

$$E(g(x)) = \int_{-\infty}^{+\infty} g(u) p_X(u) du \quad (3.25)$$

En prenant des fonctions polynomiales pour $g(u)$, on obtient les moments classiques. En utilisant des fonctions exponentielles, on obtient les fonctions caractéristiques associées aux variables aléatoires.

La première fonction caractéristique est :

$$\Phi_x(v) = \int_{-\infty}^{+\infty} e^{jvu} p_X(u) du = E(e^{jvx}) \quad (3.26)$$

On peut remarquer que la première fonction caractéristique correspond à la transformée de Fourier de sa densité de probabilité. La seconde fonction caractéristique est définie comme le logarithme népérien de la première :

$$\Psi_x(v) = \ln(\Phi_x(v)) \quad (3.27)$$

En développant la première fonction caractéristique en série de Taylor au voisinage de l'origine, on obtient les moments d'ordre r :

$$\mu_{x,(r)} = (-j)^r \frac{d^r \Phi_x(v)}{dv^r} \Big|_{v=0} = E(x^r) \quad (3.28)$$

En faisant de même avec la seconde fonction caractéristique, on définit les *cumulants* d'ordre r :

$$\kappa_{x,(r)} = (-j)^r \frac{d^r \Psi_x(v)}{dv^r} \Big|_{v=0} = \text{Cum}(x, x, \dots, x) \quad (3.29)$$

Les cumulants d'ordre r peuvent être calculés à partir des moments d'ordre inférieur ou égal à r [KS94]. Pour les ordres de 1 à 4, ces relations sont :

$$\kappa_{x,(1)} = \mu_{x,(1)} \quad (3.30)$$

$$\kappa_{x,(2)} = \mu_{x,(2)} - \mu_{x,(1)}^2 \quad (3.31)$$

$$\kappa_{x,(3)} = \mu_{x,(3)} - 3\mu_{x,(1)}\mu_{x,(2)} + 2\mu_{x,(1)}^3 \quad (3.32)$$

$$\kappa_{x,(4)} = \mu_{x,(4)} - 4\mu_{x,(3)}\mu_{x,(1)} - 3\mu_{x,(2)}^2 + 12\mu_{x,(2)}\mu_{x,(1)}^2 - 6\mu_{x,(1)}^4 \quad (3.33)$$

Les cumulants d'ordre un et deux sont respectivement le moment d'ordre un et le moment centré d'ordre deux. Nous considérerons des variables centrées ($\kappa_{x,(1)} = \mu_{x,(1)} = 0$) pour alléger les expressions de la suite de ce chapitre.

Si X suit une loi gaussienne, sa seconde fonction caractéristique s'écrit [LAP97] :

$$\Psi_x(v) = j\mu_{x,(1)}v - \frac{1}{2}\mu_{x,(2)}v^2 \quad (3.34)$$

Ainsi, tous les cumulants d'ordre supérieur à deux d'une variable aléatoire gaussienne sont nuls.

Les cumulants d'ordre 3 et 4 sont très utilisés sous leurs formes normalisées, où ils prennent le nom respectivement d'asymétrie et de kurtosis. En considérant une variable aléatoire centrée, l'asymétrie et le kurtosis s'écrivent :

$$\text{Asym}(x) = \frac{\kappa_{x,(3)}}{\kappa_{x,(2)}^{3/2}} \quad (3.35)$$

$$\mathcal{K}(x) = \frac{\kappa_{x,(4)}}{\kappa_{x,(2)}^2} = \frac{\mu_{x,(4)}}{\mu_{x,(2)}^2} - 3 \quad (3.36)$$

L'asymétrie est nulle si la densité de probabilité de x possède un axe de symétrie.

Le kurtosis caractérise le comportement à l'infini des densités de probabilités [LAP97] : une densité qui tend vers zéro à l'infini plus rapidement qu'une loi gaussienne aura un kurtosis négatif et sera dite « sous-gaussienne », tandis qu'une densité qui tend vers zéro plus lentement qu'une loi gaussienne aura un kurtosis positif, et sera dite « sur-gaussienne ». Le kurtosis peut ainsi servir de mesure d'écart à la gaussianité.

Pour estimer le kurtosis, il faut tout d'abord estimer les cumulants qui le composent. La première idée est de tirer partie des équations liant cumulants et moments (3.30 à 3.33), en

utilisant les estimations au sens du maximum de vraisemblance des moments. L'estimation du cumulants d'ordre 4 à partir de N échantillons d'une variable aléatoire X centrée s'écrit alors :

$$\hat{\kappa}_{x,(4)} = \hat{\mu}_{x,(4)} - 3\hat{\mu}_{x,(2)}^2 = \frac{1}{N} \sum_{i=1}^N x_i^4 - 3 \left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^2 \quad (3.37)$$

Toutefois, ce genre d'estimation conduit à un biais [LAP97]. Une estimation non biaisée des cumulants, appelé *k-statistique* s'écrit pour l'ordre 4 [KS94] :

$$\hat{\kappa}_{x,(4)} = \frac{N+2}{N(N-1)} \sum_{i=1}^N x_i^4 - \frac{3N}{N-1} \left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^2 \quad (3.38)$$

L'estimation du kurtosis est ensuite réalisée grâce aux estimateurs non biaisés des cumulants d'ordre 4 et 2. Elle s'écrit sous la forme :

$$\hat{\mathcal{K}}(x) = \frac{\hat{\kappa}_{x,(4)}}{\hat{\kappa}_{x,(2)}^2} = \frac{N(N+2)}{N-1} \frac{\sum_{i=1}^N x_i^4}{(\sum_{i=1}^N x_i^2)^2} - \frac{3N}{N-1} \quad (3.39)$$

En raison de sa complexité, les propriétés de cet estimateur ont été étudiées seulement de façon approchée, et pour de grandes valeurs de N [KS94]. On notera qu'il est biaisé. Toutefois, dans le cas gaussien correspondant à l'hypothèse où seul du bruit est présent sur l'ensemble du plan temps-fréquence, des résultats exacts existent [KS94]. L'espérance et la variance de l'estimateur (3.39) sont :

$$\mathbb{E}(\hat{\mathcal{K}}(x)) = 0, \quad (3.40)$$

$$\text{Var}(\hat{\mathcal{K}}(x)) = \frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)} \approx \frac{24}{N} \quad (3.41)$$

Le kurtosis d'une variable gaussienne est théoriquement nul, l'estimateur (3.39) est donc non biaisé dans le cas gaussien. De plus, sa variance tendant vers 0 quand N tend vers l'infini, il est consistant.

Dans le cas des variables complexes, notées par la suite $z = a + jb$, la même définition des cumulants (3.29) est applicable [AGL96, LAP97], en associant à la variable aléatoire z son conjugué z^* . Sa densité de probabilité, au lieu d'être définie en fonction de ses parties réelle et imaginaire, est alors définie en fonction de z et z^* . Ses fonctions caractéristiques s'écrivent alors :

$$\Phi_{z,z^*}(u, u^*) = \mathbb{E} \left(e^{\frac{j(z^*u + u^*z)}{2}} \right) \quad (3.42)$$

$$\Psi_{z,z^*}(u, u^*) = \ln(\Phi_{z,z^*}(u, u^*)) \quad (3.43)$$

Cette approche montre que la description statistique d'une variable complexe fait intervenir des moments et des cumulants croisés entre la variable complexe z et sa conjuguée z^* . Il y a ainsi $r+1$ les moments et les cumulants d'ordre $r = p + q$, où p correspond à l'ordre des termes non conjugués, et q à l'ordre des termes conjugués. Les moments et cumulants sont comme précédemment issus respectivement des dérivées des première et seconde fonctions caractéristiques :

$$\mu_z \begin{matrix} (q) \\ (p) \end{matrix} = \mathbb{E}(z^p z^{*q}) = (-2j)^r \frac{\partial^r \Phi_{z,z^*}(u, u^*)}{\partial u^p \partial u^{*q}} \Big|_{u=0} \quad (3.44)$$

$$\kappa_z \begin{matrix} (q) \\ (p) \end{matrix} = \text{Cum} \left(\underbrace{z, \dots, z}_p \text{ fois}, \underbrace{z^*, \dots, z^*}_q \text{ fois} \right) = (-2j)^r \frac{\partial^r \Psi_{z,z^*}(u, u^*)}{\partial u^p \partial u^{*q}} \Big|_{u=0} \quad (3.45)$$

Pour des variables circulaires, les cumulants d'ordre r sont nuls si z et z^* ne sont pas au même ordre $p = q$ [AGL96]. Dans le cas où nous n'avons que du bruit dans le plan temps-fréquence, chaque point temps-fréquence est distribué selon une gaussienne complexe circulaire. Par conséquent, nous nous intéressons seulement au cumulants d'ordre 4 $\kappa_z^{(2)}$, qui nous conduit à la définition du kurtosis spectral [VGS03] :

$$\mathcal{KS}(z) = \frac{\kappa_z^{(2)}}{\kappa_z^{(1)}} = \frac{\mathbb{E}(z^2 z^{*2})}{\mathbb{E}(z z^*)^2} - 2 \quad (3.46)$$

À noter, en considérant une variable complexe $z = a + jb$ circulaire [Pic94], le kurtosis spectral peut s'exprimer en fonction du kurtosis de sa partie réelle (ou de manière équivalente, du kurtosis de sa partie imaginaire). Dans le cas circulaire, la loi de z est invariante par multiplication d'un facteur $e^{j\theta}$, impliquant que les parties réelle a et imaginaire b suivent la même loi. On peut alors exprimer la variance de z en fonction de celle de a :

$$\mathbb{E}(z z^*) = \mathbb{E}(a^2 + b^2) = \mathbb{E}(a^2) + \mathbb{E}(b^2) = 2\mathbb{E}(a^2) \quad (3.47)$$

En exprimant le moment d'ordre 4 de la même manière, on obtient :

$$\mathbb{E}(z^2 z^{*2}) = \mathbb{E}(a^4 + b^4 + 2a^2 b^2) = 2\mathbb{E}(a^4) + 2\mathbb{E}(a^2 b^2) \quad (3.48)$$

En écrivant a et b en fonction de z et z^* :

$$a = \frac{z + z^*}{2}, \quad b = \frac{z - z^*}{2i} \quad (3.49)$$

L'équation (3.48) devient :

$$\mathbb{E}(z^2 z^{*2}) = 2\mathbb{E}(a^4) - \frac{1}{8}\mathbb{E}(z^4 + z^{*4} - 2z^2 z^{*2}) \quad (3.50)$$

Du fait de la circularité de z , les espérances de z^4 et de z^{*4} sont toutes deux nulles. Le moment d'ordre 4 de z s'exprime en fonction de celui de a :

$$\mathbb{E}(z^2 z^{*2}) = \frac{8}{3}\mathbb{E}(a^4) \quad (3.51)$$

En reprenant les expressions du kurtosis spectral de z (3.46) et du kurtosis de a (3.36), leur rapport est :

$$\frac{\mathcal{KS}(z)}{\mathcal{K}(a)} = \frac{\mathbb{E}(z^2 z^{*2}) - 2\mathbb{E}(z z^*)^2}{\mathbb{E}(z z^*)^2} \times \frac{\mathbb{E}(a^2)^2}{\mathbb{E}(a^4) - 3\mathbb{E}(a^2)^2} \quad (3.52)$$

$$= \frac{\frac{8}{3}\mathbb{E}(a^4) - 8\mathbb{E}(a^2)^2}{4(\mathbb{E}(a^4) - 3\mathbb{E}(a^2)^2)} \quad (3.53)$$

$$= \frac{2}{3} \quad (3.54)$$

Le kurtosis spectral d'une variable complexe circulaire est égal à $\frac{2}{3}$ du kurtosis de sa partie réelle ou de sa partie imaginaire.

L'estimateur du kurtosis spectral s'écrit, en utilisant les *k-statistiques* [VGS03] :

$$\widehat{\mathcal{KS}}(z) = \frac{N}{N-1} \left(\frac{(M+1) \sum_{i=1}^N |z_i|^4}{(\sum_{i=1}^N |z_i|^2)^2} - 2 \right) \quad (3.55)$$

Cette définition ne fait intervenir que la norme de z , le kurtosis spectral est par conséquent une valeur réelle. À l'instar de l'estimateur du kurtosis réel (3.39), cet estimateur est non biaisé dans le cas gaussien, et sa variance s'écrit :

$$\text{Var}(\widehat{\mathcal{KS}}(z)) = \frac{4N^2}{(N-1)(N+2)(N+3)} \approx \frac{4}{N} \quad (3.56)$$

Cet estimateur est consistant.

3.2.2 Kurtosis de la TFCT rectifiée

Dans ces travaux, le plan temps-fréquence est constitué d'un nombre inconnu de points contenant un mélange « signal + bruit », les autres contenant du bruit uniquement. En supposant le niveau de bruit connu, le chapitre 2 nous permet de détecter les points contenant du signal. Toutefois, l'estimation du niveau de bruit à partir de tous les points temps-fréquence conduit à une surestimation (3.24). Cette surestimation permet de détecter une première vague de points contenant du signal, et une nouvelle estimation sans ces points détectés permet de diminuer la surestimation du niveau de bruit.

Il faut alors contrôler les points non segmentés, pour savoir quand arrêter ces itérations. C'est là qu'intervient le kurtosis, qui est nul dans le cas d'une variable aléatoire gaussienne : les itérations d'estimation/détection seront arrêtées lorsque le kurtosis des points non segmentés atteindra zéro dans le cas gaussien.

Le kurtosis spectral a quant à lui déjà été utilisé dans des buts de caractérisation et de classification de signaux, en l'utilisant séparément pour chaque fréquence [CSL96, Ser98, VGML04]. Dans le cadre de ce travail, il va être utilisé de la même manière que le kurtosis unidimensionnel, en tant que mesure de blancheur d'un ensemble de points temps-fréquence.

Nous utilisons comme représentation temps-fréquence la TFCT « rectifiée » (1.25), qui donne une variable complexe $X_\phi[n, k]'$. Deux approches relatives au kurtosis sont possibles : soit on considère le kurtosis réel des parties réelle et imaginaire de $X_\phi[n, k]'$, soit on considère directement le kurtosis spectral de la variable complexe. Les deux méthodes seront étudiées en parallèle, pour déterminer quelle est la plus efficace. Dans la suite du texte, la mention du kurtosis sans plus de précisions sous-entendra qu'il s'agit indifféremment du kurtosis spectral ou des kurtosis des parties réelle et imaginaires de la TFCT rectifiée.

Dans le cas « bruit seul », le chapitre 1 a montré que tous les échantillons de la TFCT rectifiée suivent une loi gaussienne complexe circulaire², c'est-à-dire que leurs parties réelle et imaginaire sont indépendantes et de même variance.

Dans le cas « signal + bruit », le nombre de coefficients temps-fréquence composés de la somme de la TFCT du bruit seul et de la TFCT rectifiée du signal seul est inconnu. En considérant les parties réelle et imaginaire séparément, le fait d'avoir des points contenant du signal va rajouter des valeurs extrêmes comparé aux points de bruit seul. La distribution globale des coefficients aura tendance à être sur-gaussienne, et son kurtosis à être positif.

²Encore une fois, les cas $\alpha[n, k] = 0$ ou 1 seront ignorés.

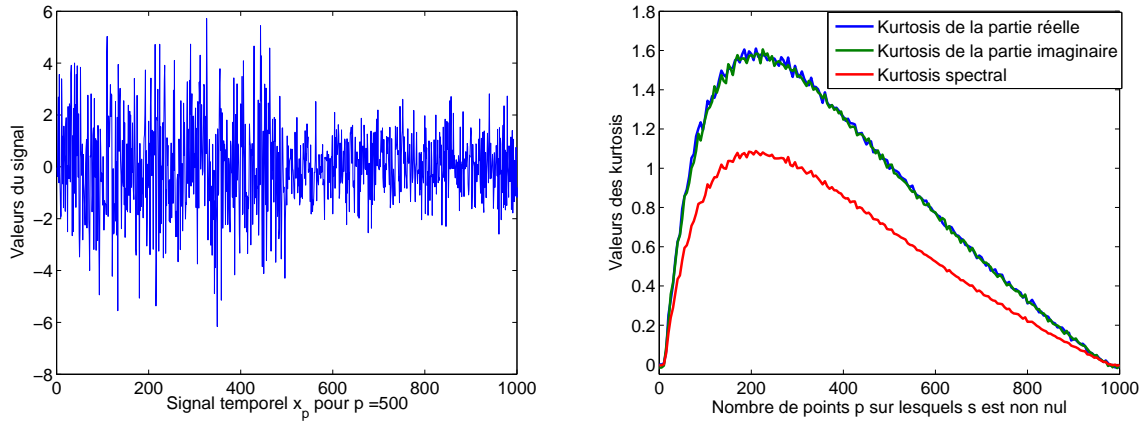


FIG. 3.2 : À gauche, illustration du signal temporel $x_p[m] = s[m] + b[m]$ décrit par l'équation (3.57). Le bruit b est gaussien blanc de variance 1, tandis que le signal s est gaussien blanc de variance 3 sur les p premiers points du signal, puis s'annule. À droite, le kurtosis spectral et les kurtosis des parties réelle et imaginaire de la TFCT rectifiée sont représentés, en fonction du nombre de points sur lesquels le signal s est non nul. Lorsqu'il n'y a que du bruit, ou lorsque le mélange est composé de deux gaussiennes indépendantes stationnaires, le kurtosis est nul. Dans les autres cas, c'est-à-dire en présence d'un signal non stationnaire, le kurtosis est positif. Le kurtosis spectral a la même allure que les kurtosis des parties réelle et imaginaire, diminuée d'un facteur $\frac{2}{3}$. Les kurtosis ont été obtenus en moyennant mille réalisations de b et de s .

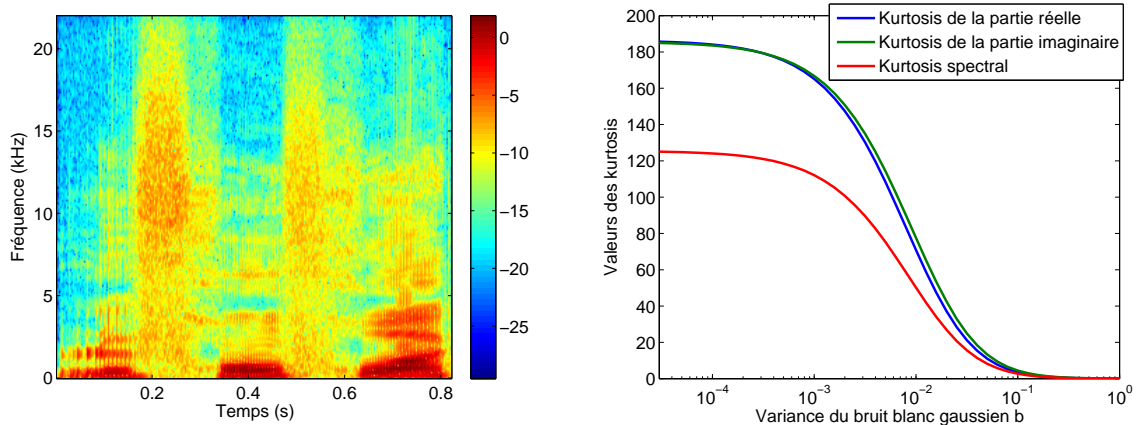


FIG. 3.3 : À gauche, spectrogramme d'un signal de parole. À droite, évolution du kurtosis spectral et des kurtosis des parties réelle et imaginaire de la TFCT rectifiée du signal auquel on ajoute un bruit blanc gaussien, en fonction de la variance de ce bruit. Lorsque le signal n'est pas négligeable par rapport au bruit, le kurtosis de l'ensemble du plan temps-fréquence est strictement positif. Comme dans la figure 3.2, le kurtosis spectral a la même allure que les kurtosis des parties réelle et imaginaire diminuée d'un facteur $\frac{2}{3}$. Les kurtosis ont été obtenus en moyennant mille réalisations de bruit.

Les figures 3.2 et 3.3 illustrent ce fait de deux manières différentes. La figure 3.2 donne l'évolution du kurtosis des parties réelle et imaginaire de la TFCT d'un mélange x « signal +

bruit » selon la durée sur laquelle le signal est présent. Le signal s est un bruit blanc gaussien stationnaire de variance $\sigma_s^2 = 3$ sur les p premiers points, puis est nul sur les points suivants. Le mélange $x_p[m]$ de paramètre p s'écrit alors :

$$x_p[m] = \begin{cases} s[m] + b[m] & \text{si } m \geq p \\ b[m] & \text{sinon.} \end{cases} \quad (3.57)$$

Ici, le bruit b est un bruit blanc gaussien stationnaire de variance $\sigma_b^2 = 1$ de 1000 points. Le signal s étant supposé indépendant du bruit additif b , le mélange $x_p[m]$ suit une loi gaussienne de variance $\sigma_s^2 + \sigma_b^2 = 4$ sur les p premiers points, puis une loi gaussienne identique à celle du bruit seul, c'est-à-dire de variance $\sigma_b^2 = 1$.

La figure 3.2 montre que lorsque le signal s est absent ($p=0$), c'est-à-dire $x_p[m] = b[m]$, le kurtosis de la TFCT rectifiée est nul. De même, quand s est présent sur la totalité du signal ($p=1000$), tous les points du mélange x_p suivent une loi gaussienne de variance $\sigma_b^2 + \sigma_s^2 = 4$: le kurtosis est à nouveau nul.

Entre ces deux extrêmes, le kurtosis est toujours positif. Ses plus grandes valeurs sont situées quand 10 à 20% de signal est présent dans le mélange. En effet, rajouter quelques points de grande amplitude augmente beaucoup le cumuland d'ordre 4 et affecte peu la variance. Au contraire, quand peu de bruit b est présent seul, la variance totale du plan temps-fréquence tend à s'approcher de celle du mélange, le caractère sur-gaussien du plan complet tend à s'effacer.

La figure 3.3 montre l'évolution du kurtosis de la TFCT rectifiée d'un signal composé de la somme d'un signal de parole et d'un bruit blanc gaussien b , en fonction de la variance du bruit additif. Lorsque la variance du bruit est élevé, supérieure ou égale à 1, le signal devient négligeable et obtient un kurtosis pratiquement nul. Dans le cas contraire, le kurtosis augmente avec la diminution de la variance du bruit jusqu'à une valeur limite, correspondant à une variance du bruit négligeable. Dans ces deux cas, le kurtosis spectral de la TFCT rectifiée est égal au facteur $\frac{2}{3}$ près au kurtosis des parties réelle et imaginaire.

De manière générale, le kurtosis du plan temps-fréquence d'un mélange « signal + bruit » sera positif. Par les itérations d'estimation de bruit/détection du signal, les points non segmentés ne seront finalement composés que de bruit, avec un kurtosis nul. Un critère d'arrêt de ces itérations peut donc se porter sur le kurtosis. Toutefois, les variations du kurtosis au cours de ces itérations ne sont pas monotones : comme l'illustre la figure 3.2, peu de points de signal mélangés à beaucoup de points de bruit conduisent à un kurtosis plus grand que dans le cas d'un plus grand nombre de points.

3.2.2.1 Cas particuliers

Dans l'utilisation du kurtosis dans le cadre de la segmentation, on fait l'hypothèse que le kurtosis de l'ensemble du plan temps-fréquence est positif. Cette hypothèse est vérifiée dans la plupart des cas, notamment quand le signal est concentré dans certaines zones temps-fréquence. Il y aura un grand nombre de points contenant uniquement du bruit, mélangés avec quelques points de grande amplitude, ce qui résultera en une grande valeur de kurtosis.

Cependant, des contre-exemples existent, par exemple un signal s composé de 10 fréquences pures régulièrement réparties entre 0.02 et 0.47 en fréquences réduites, d'équation :

$$s[m] = \sum_f \sin(2\pi f m + \phi_f), \quad (3.58)$$

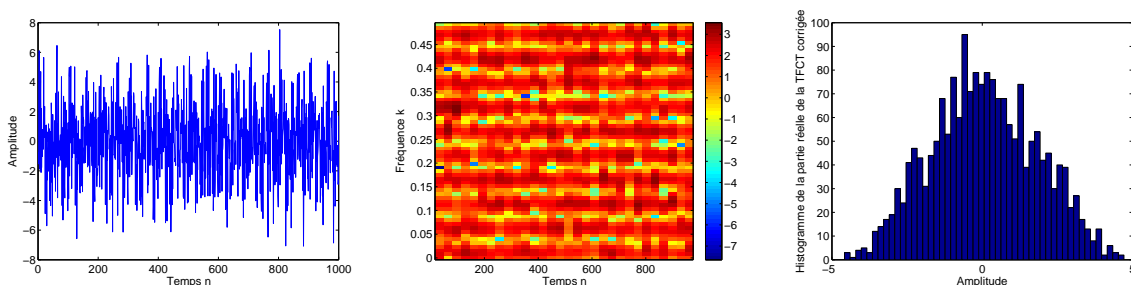


FIG. 3.4 : Cas particulier d'un signal composé de plusieurs fréquences pures régulières réparties en fréquences, avec une phase aléatoire. À gauche, le signal temporel, additionné d'un bruit blanc de variance 1. Au centre, son spectrogramme évalué avec une fenêtre de Hanning de 63 points, un recouvrement de 31 points et 128 fréquences calculées. À droite, l'histogramme de la partie réelle de la TFCT rectifiée.

avec $f = [0.02, 0.07 \dots 0.47]$ et $\phi_f \in [0, 2\pi[$ une phase aléatoire.

Comme illustré sur la figure 3.4, l'ensemble du plan temps-fréquence contient de l'énergie du signal. Pour une TFCT rectifiée évaluée avec une fenêtre de Hanning de 63 points, un recouvrement de 31 points et 128 fréquences calculées, la distribution globale des points devient sous-gaussienne, résultant en un kurtosis négatif. En rajoutant un bruit gaussien de variance un, et en moyennant les kurtosis sur 1000 réalisations du signal et du bruit, le kurtosis moyen des parties réelle et imaginaires de la TFCT rectifiée est approximativement de -0.65, le kurtosis spectral approximativement de -0.4.

Ce type de signal sera considéré comme suffisamment rare pour être ignoré. De plus, par son absence de points temps-fréquence contenant uniquement du bruit, la méthode de segmentation décrite par ces travaux est inapplicable, et une fenêtre d'analyse plus résolutive en fréquence permettra de retrouver des points de bruit seul entre les fréquences des sinusoides.

Un autre cas particulier peut apparaître, quand la phase de la TFCT rectifiée du signal seul est constante et conduit à ce que sa partie réelle ou imaginaire est nulle. Par exemple, un signal constitué d'une sinusoïde, avec une fenêtre de TFCT d'une longueur égale à un nombre entier de période de la sinusoïde, et aucun recouvrement entre fenêtres, et la phase de la sinusoïde telle que la partie imaginaire de la TFCT rectifiée est nulle. Dans la TFCT du mélange « signal + bruit », la partie imaginaire est constituée de bruit uniquement, et a donc un kurtosis nul, contrairement au kurtosis de la partie réelle. À noter, le kurtosis spectral n'est pas affecté, puisqu'il est estimé (3.55) par le module de la TFCT rectifiée.

3.2.2.2 Influence de $\alpha[n, k]$

Dans ce chapitre, nous n'avons traité jusqu'ici que de la TFCT rectifiée, dont chaque coefficient temps-fréquence issu d'un bruit blanc $B_\phi[n, k]'$ suit une loi gaussienne circulaire, les parties réelle et imaginaire étant décorréelées et de même variance. Que devient le kurtosis d'un bruit blanc dans le cas où l'on ne rectifie pas la TFCT par le paramètre $\alpha[n, k]$, décrit par l'équation (1.10) ?

Ce paramètre $\alpha[n, k]$ indique que la variance de la partie réelle des coefficients de la TFCT n'est pas égale à celle de leurs parties imaginaires sur toutes les bandes de fréquence. Les lois des parties réelle et imaginaire de la TFCT sont un mélange de gaussiennes, de variances respectivement $\alpha[n, k]$ et $1 - \alpha[n, k]$. Comme illustré sur la figure 3.2, le kurtosis d'un mélange

de gaussiennes de variances différentes est positif. Ainsi, dans le cas de la TFCT non rectifiée d'un bruit blanc, un biais positif apparaît sur le kurtosis des parties réelle et imaginaire des coefficients. Dans le cas d'un bruit blanc de 1000 points et d'une TFCT évaluée avec une fenêtre de Hanning de 63 points, un recouvrement de 31 points et 128 fréquences calculées, le kurtosis moyenné sur 1000 réalisations de bruit est égal à 0.12 pour la partie réelle et 0.17 pour la partie imaginaire, tandis qu'il est de -0.016 dans le cas de la TFCT rectifiée.

Un biais d'un facteur dix sur la valeur estimée du kurtosis est ainsi évité par l'utilisation de la rectification de la TFCT.

Dans le cas du kurtosis spectral, le paramètre $\alpha[n, k]$ ne devrait pas avoir d'influence. En effet, seule la norme de $B_\phi[n, k]$ est utilisée dans son estimation. D'après l'équation (1.9), l'espérance de la norme d'un coefficient de la TFCT est :

$$E(B_\phi[n, k]B_\phi[n, k]^*) = \alpha[n, k]\sigma_b^2 + (1 - \alpha[n, k])\sigma_b^2 = \sigma_b^2 \quad (3.59)$$

Tandis que dans le cas rectifié, l'équation précédente devient :

$$E(B_\phi[n, k]'B_\phi[n, k]'^*) = \frac{1}{2}\sigma_b^2 + \frac{1}{2}\sigma_b^2 = \sigma_b^2 \quad (3.60)$$

En pratique, le même phénomène que pour les parties réelle et imaginaire apparaît : le kurtosis spectral moyen, dans les mêmes conditions que précédemment, est de 0.055 dans le cas non rectifié, et de -0.006 dans le cas rectifié.

Quel que soit le choix du kurtosis, spectral ou non, la prise en compte de la non circularité de la TFCT d'un bruit blanc et sa correction permet de limiter le biais dans l'estimation du kurtosis d'un facteur 10.

3.2.2.3 Choix d'un seuil d'arrêt

Lors de la segmentation, le kurtosis des points non segmentés va tendre vers zéro, à mesure que l'on détecte des points contenant du signal. L'estimateur du kurtosis est lui-même une variable aléatoire : selon les réalisations, la valeur estimée du kurtosis sera variable. Il convient alors de définir un seuil au-dessous duquel le kurtosis sera considéré comme suffisamment faible pour être issu d'une variable gaussienne uniquement.

Une première solution peut être de donner à l'utilisateur de l'algorithme de segmentation la responsabilité de choisir un seuil. Cette solution apparemment inintéressante permet à l'utilisateur d'intégrer des informations *a priori* s'il le souhaite. Une autre solution consiste à définir automatiquement un seuil, basé sur la variance connue de l'estimateur du kurtosis.

Cette variance est approximativement de $\frac{24}{N}$ (3.41) dans le cas à une dimension, et de $\frac{4}{N}$ (3.56) dans le cas complexe. Deux phénomènes expliquent principalement la variance bien plus faible du kurtosis spectral : d'une part, les parties réelle et imaginaire sont indépendantes dans le cas gaussien, il y a alors deux fois plus d'information dans un coefficient complexe utilisé par le kurtosis spectral que dans la partie réelle ou imaginaire de ce coefficient ; d'autre part, le rapport entre le kurtosis spectral d'une variable complexe circulaire et le kurtosis de sa partie réelle ou imaginaire est de $\frac{2}{3}$ (3.54).

La « variance équivalente » du kurtosis estimé à partir des parties réelle et imaginaire de la TFCT d'un bruit blanc, pour la comparer à celle du kurtosis spectral, est :

$$\text{Var}_{\text{eq}}(\hat{\mathcal{K}}(x)) \approx \left(\frac{2}{3}\right)^2 \frac{1}{2} \frac{24}{N} \approx \frac{5.33}{N} \quad (3.61)$$

nombre de fréquences calculées recouvrement	256	511	1024	2048
31	9.4%	9.4%	9.3%	9.4%
63	8.3%	7.6%	8.4%	9.6%
127	4.4%	4.6%	4.6%	4.9%

TAB. 3.1 : Pourcentage de kurtosis spectraux dépassant le seuil décrit par l'équation (3.65). 1000 estimations de ces kurtosis spectraux sont calculées sur l'ensemble d'une TFCT rectifiée d'une réalisation d'un bruit blanc de variance 1 et de longueur 10000, en fonction du recouvrement et du nombre de fréquences calculées avec une fenêtre de Hanning de 255 points.

Cette variance reste néanmoins toujours supérieure à celle du kurtosis spectral.

Dans ces calculs de variance, les N réalisations sont supposées indépendantes, ce qui n'est évidemment pas le cas des coefficients $[n, k]$ du plan temps-fréquence, de par le recouvrement entre fenêtres et le complément de zéros. Nous définissons alors un *facteur de redondance* noté f_r , défini comme le rapport entre le nombre NK de coefficients de la TFCT à segmenter, et le nombre de coefficients de la TFCT sans redondance, c'est-à-dire sans recouvrement ni bourrage de zéros. Cette dernière valeur est égale à la longueur M du signal à segmenter $s[m]$. Le facteur de redondance s'écrit alors :

$$f_r = \frac{NK}{M} \quad (3.62)$$

Pour l'estimation du kurtosis de $N_i K_i$ points non segmentés, le nombre de réalisations indépendantes est approché par N_a , tel que :

$$N_a = \frac{N_i K_i}{f_r} \quad (3.63)$$

En supposant que l'estimateur suit une loi gaussienne, 97.5% des réalisations de cette variable aléatoire sont inférieures à deux fois son écart-type. Nous choisissons ainsi de définir un seuil sur la valeur du kurtosis s_k dans le cas à une dimension en dessous duquel le kurtosis sera considéré nul par :

$$s_k = 2\sqrt{\text{Var}(\hat{\mathcal{K}}(x))} = 2\sqrt{\frac{24}{N_a}} \quad (3.64)$$

Et dans le cas du kurtosis spectral par :

$$s_{ks} = 2\sqrt{\text{Var}(\widehat{\mathcal{KS}}(x))} = 2\sqrt{\frac{4}{N_a}} \quad (3.65)$$

Pour valider ce seuil, nous calculons 1000 réalisations de TFCT rectifiée d'un bruit blanc de variance 1 de longueur 10000 points, pour une fenêtre Hanning de 255 points pour diverses valeurs de recouvrements et de nombre de fréquences calculées. Chacune de ces réalisations conduit à une estimation du kurtosis spectral d'un ensemble de coefficients gaussiens circulaires de variance 1.

Le tableau 3.1 indique le pourcentage de ces 1000 estimations de kurtosis spectraux qui sont supérieures au seuil décrit par l'équation (3.65), avec $N_a = M = 10000$. Les valeurs obtenues sont très nettement supérieures aux 2.5% supposés : l'hypothèse de la gaussiannité de l'estimateur est fautive, ou la redondance de la TFCT est mal prise en compte. Toutefois, ces valeurs sont suffisamment faibles pour considérer le seuil (3.65) comme acceptable.

3.3 Conclusion

Dans ce chapitre, nous avons déterminé un estimateur efficace pour déterminer le niveau de bruit quand la TFCT est issue d'un bruit seul. Cette valeur est surestimée dès que des points contenant du signal sont présents. Une méthode itérative surestimant volontairement le niveau de bruit à l'initialisation a été utilisée, détectant successivement les points contenant du signal et permettant ainsi de les retirer de l'estimation du bruit.

Ce principe avait déjà été utilisé [Hor02], la contribution de ces travaux a été l'utilisation du kurtosis des points non segmentés afin de contrôler l'arrêt des itérations d'estimation du bruit/détection de signal. Le kurtosis spectral, défini à partir des coefficients complexes de la TFCT rectifiée, a été démontré $\frac{2}{3}$ plus faible que le kurtosis des parties réelle ou imaginaire. Cette différence n'influe pas sur la capacité du kurtosis, spectral ou non, à caractériser correctement si les points sur lesquels le kurtosis est calculé proviennent d'une variable aléatoire gaussienne ou non.

De plus, l'influence de la rectification de la TFCT telle que définie dans le chapitre 1 a été illustrée, conduisant à une surestimation du kurtosis et du kurtosis spectral dans le cas d'une variable aléatoire gaussienne.

En tant que critère d'arrêt des itérations estimation du bruit/détection de signal, un seuil pour l'estimation du kurtosis a été proposé, seuil en dessous duquel les itérations seront arrêtées. Ce seuil est basé sur le nombre de réalisations approximativement indépendantes des points non segmentés du plan temps-fréquence, en fonction des variances théoriques des estimateurs du kurtosis et du kurtosis spectral.

Enfin, la variance de l'estimateur du kurtosis spectral a été montrée plus faible que celle du kurtosis des parties réelle et imaginaire. Le kurtosis spectral permet en plus de traiter directement les coefficients de la TFCT complexe. Par conséquent, c'est ce dernier qui sera retenu comme critère d'arrêt de l'algorithme complet de segmentation, qui sera présenté et caractérisé dans le prochain chapitre.

Chapitre 4

Algorithme de segmentation

Le cœur de l'algorithme présenté dans cette première partie de thèse a été présenté dans le chapitre précédent. Des itérations successives d'estimation du niveau de bruit permettent de détecter des points temps-fréquence contenant du signal. Toutefois, cette détection ne donne que le support temps-fréquence du signal d'origine, sans permettre de relier ce support à des notions de composantes temps-fréquence. De plus, le résultat final comprend des points de fausse alarme, dont la quantité dépendra de la probabilité de fausse alarme choisie.

Une première section présente un algorithme de croissance de région, appliqué à chaque itération sur les points détectés, qui a pour but de résoudre ces deux problèmes conjointement, au prix d'un nouveau paramètre contrôlant la segmentation. Une seconde section décrira l'algorithme complet de segmentation, et caractérisera son comportement en fonction des paramètres d'entrée, en s'appuyant sur l'exemple d'un signal synthétique.

Une troisième section présentera les extensions de ce principe de segmentation, dans le cas d'un bruit non blanc ou non stationnaire d'une part, et dans le cas d'autres représentations temps-fréquence d'autre part. Enfin, des exemples d'applications seront donnés.

4.1 Croissance de région

La croissance de région ([AB94]) est un des algorithmes les plus simples de segmentation d'image, ici le plan temps-fréquence. L'idée est de partir d'un point, appelé germe, et de chercher dans ses plus proches voisins lesquels partagent une même caractéristique pour les affecter à la même classe que le germe. Ces points nouvellement segmentés deviennent à leur tour des germes de cette même classe, qui contient au final tous les points connexes partageant la même caractéristique que le germe initial.

En considérant qu'un motif temps-fréquence est un ensemble connexe de points contenant du signal, l'idée est d'appliquer cette croissance de région à chaque itération, après détection des points contenant du signal. La caractéristique à suivre est simple : est-ce que le point temps-fréquence a été détecté comme contenant du signal ou non ?

Le choix du germe est dans ce cas peu discriminant, tous les points détectés connexes finissant dans une même classe. Selon l'équation (1.30), la présence de signal sur un point temps-fréquence augmente la valeur de son moment d'ordre deux. Le germe le plus logique est donc celui de plus grand moment d'ordre deux. Les points détectés comme contenant du signal, connexes au germe initial, sont ainsi tous segmentés en une même classe.

Ainsi, la croissance de région va permettre d'agrèger les points détectés connexes en un seul motif... s'ils sont tous connexes. Dans le cas où tous les points détectés ne sont pas connexes, un nouveau germe est à choisir parmi les points non segmentés, qui crée une nouvelle classe. À l'itération suivante, les classes déjà créées servent de germes afin de les agrandir avant de créer une nouvelle classe si besoin est.

L'application de la croissance de région après la phase de détection permet de segmenter facilement des composantes proches, lors des premières itérations, qui seraient fusionnées à la fin.

Il est possible de ne segmenter à chaque itération qu'une proportion donnée de points détectés, appelés par la suite *points candidats* à la segmentation. Se faisant, on peut éviter de segmenter des motifs isolés correspondant à des fausses alarmes. En notant N_b le nombre de points non segmentés ni détectés, c'est-à-dire le nombre de points qui contiennent uniquement du bruit, et en supposant que le niveau de bruit ait été correctement estimé, il y aura approximativement $N_b \times p_{fa}$ points détectés à tort comme contenant du signal. En notant N_{cand} le nombre de points détectés, la proportion de candidats à segmenter pour éviter les motifs de fausses alarmes notée p_{cand} est :

$$p_{cand} = \frac{N_{cand} - p_{fa}N_b}{N_{cand}} \quad (4.1)$$

À noter, selon le nombre de points candidats à la segmentation N_{cand} et le nombre de points de bruit uniquement N_b , une grande valeur de p_{fa} résultera en une valeur négative de p_{cand} , empêchant alors de segmenter de nouveaux points indépendamment du critère d'arrêt de la segmentation. Pour éviter ce genre de cas, une valeur minimale strictement positive est attribuée à ce paramètre.

En arrêtant de créer des nouveaux germes lorsque l'on a segmenté une proportion de points candidats au moins supérieure à p_{cand} , on a peu de chance de créer un nouveau germe alors que les seuls candidats restants sont issus de fausses alarmes. Toutefois, les points candidats issus de fausses alarmes connexes à des points contenant du signal seront segmentés par la croissance du motif lié à ces points de signal.

Ce paramètre p_{cand} peut également être fixé arbitrairement à une valeur choisie, entre 0 et 1. L'utilité d'un tel choix sera illustrée dans la section suivante.

4.2 Description de l'algorithme de segmentation

L'algorithme de segmentation proposé consiste principalement en deux boucles imbriquées, illustrées sur la figure 4.1. La représentation temps-fréquence du signal est ici la TFCT rectifiée (1.25).

On entre ensuite dans la boucle principale de l'algorithme. À chaque itération, la variance du bruit est estimée sur les points non segmentés, ce qui conduit à une surestimation de cette valeur s'il y a encore des points contenant du signal non segmentés (3.24). Une stratégie de détection suivant l'approche de Neyman-Pearson, avec une probabilité de fausse alarme p_{fa} fixée, permet avec la valeur estimée du niveau de bruit de déterminer un seuil de détection $s_{\sigma_b^2}$ (2.19). Ce seuil détermine les *candidats* à la segmentation.

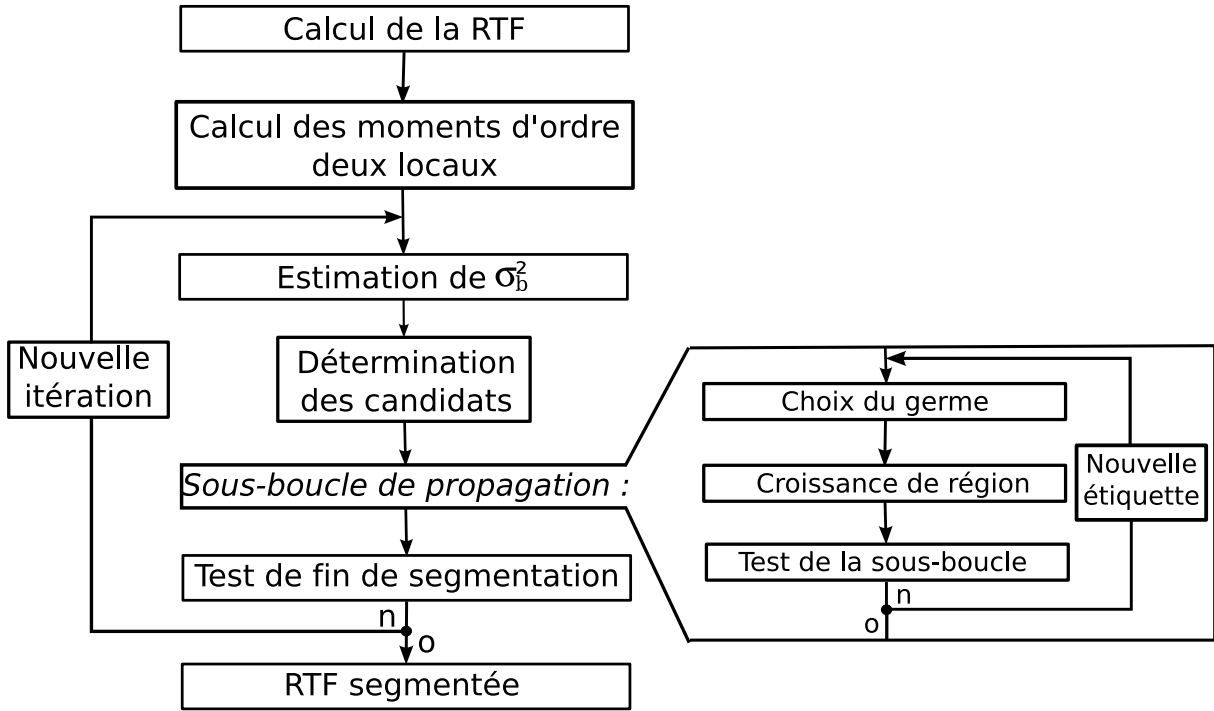


FIG. 4.1 : Schéma de l'algorithme de segmentation décrit dans la section 4.2.

La seconde boucle, imbriquée dans la première, consiste à créer les classes de la segmentation. Un algorithme de croissance de région est itérativement appliqué sur les points candidats, tant que la proportion de candidats segmentés est inférieure à une proportion donnée p_{cand} , fixée arbitrairement ou calculée selon (4.1).

À chacune de ces itérations un ou des germes sont choisis, qui contaminent tous les points qui leurs sont connexes dans une même classe. Deux cas de figure apparaissent pour choisir un germe :

- Des motifs existent déjà, qui n'ont pas encore été agrandis à cette itération. La classe la plus récente est choisie comme germe, permettant ainsi de l'étendre. Le choix de la classe la plus récente permet de contrebalancer les effets de l'algorithme de croissance de région, qui tend à privilégier la propagation des motifs extraits en premier.
- Il n'y a encore aucun motif segmenté, ou tous les motifs ont déjà été agrandis. Dans ce cas, il faut choisir un germe parmi les candidats, en lui attribuant une nouvelle étiquette, distincte de celles classes déjà créées. Puisque la présence de signal sur un point temps-fréquence augmente son moment d'ordre deux, c'est le point temps-fréquence de plus grand moment d'ordre deux qui est choisi comme germe. Il est ainsi peu probable de sélectionner un point issu d'une fausse alarme comme candidat plutôt qu'un point contenant du signal.

À chaque fois qu'un germe est choisi et agrandi, on compare la proportion de candidats nouvellement segmentés avec la valeur minimale p_{cand} . Si elle n'est pas atteinte, un nouveau germe est à choisir ; sinon, on retourne dans la boucle principale.

Le critère d'arrêt de la segmentation est le kurtosis spectral des points non segmentés. Lorsque ce kurtosis devient inférieur à un seuil s_k , fixé manuellement ou déterminé selon l'équation (3.64), les points non segmentés sont supposés ne provenir que du bruit seul : l'algorithme s'arrête. Ce

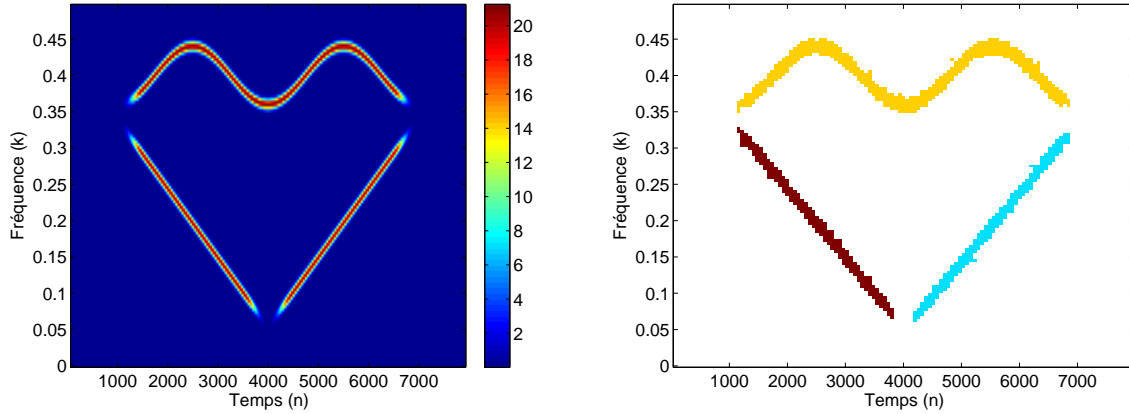


FIG. 4.2 : Spectrogramme du signal synthétique utilisé dans le test de l'influence des paramètres. Il est composé de trois modulations de fréquence d'amplitude 1, deux de ces modulations sont linéaires tandis que la troisième est sinusoïdale. Les paramètres du spectrogramme sont ceux utilisés dans le test, c'est-à-dire une fenêtre de Hanning de 127 points, 256 fréquences calculées et un recouvrement de 63 points. À droite, un exemple de bonne segmentation du spectrogramme du signal, chaque composante étant segmentée dans un motif.

critère sert d'arrêt aux deux boucles, et non seulement à la boucle principale. En effet, nul n'est besoin de créer de nouvelles classes ou d'agrandir des motifs existants si tous les points non segmentés sont supposés ne contenir que du bruit.

Ainsi, outre les paramètres de la TFCT, à savoir le type de fenêtre de pondération, sa taille, le recouvrement entre deux fenêtres successives et le nombre de fréquences calculées, trois paramètres contrôlent la segmentation : la probabilité de fausse alarme p_{fa} servant à la détection, la proportion de candidats p_{cand} segmentés à chaque itération et le seuil sur la valeur du kurtosis spectral. Ces deux derniers possèdent des valeurs déterminables automatiquement. L'influence de tous ces paramètres est décrit dans le paragraphe suivant.

4.3 Influence des paramètres

Trois paramètres interviennent dans l'algorithme de segmentation décrit :

- la probabilité de fausse alarme p_{fa} , qui détermine le seuil de détection à chaque itération (2.23) ;
- la proportion de points candidats p_{cand} qui seront segmentés par croissance de région parmi l'ensemble des points détectés ;
- le seuil du kurtosis spectral s_k , qui détermine la fin de l'algorithme.

Pour caractériser leurs influences respectives, nous utilisons un signal synthétique de 8000 points, composé de trois signaux non stationnaires différents : deux modulations linéaires de fréquence, et une modulation sinusoïdale de fréquence, toutes les trois d'amplitude 1. La figure 4.2 montre le spectrogramme de ce signal, calculé pour une fenêtre de Hanning de 127 points, 256 fréquences calculées et un recouvrement de 63 points.

D'après la définition du signal de test, la segmentation parfaite extrait un motif temps-fréquence pour chacune de ces composantes. La caractérisation des paramètres va se faire en étudiant le nombre de classes segmentées en fonction des choix des paramètres et selon le rapport

Signal sur Bruit (RSB). Le RSB est défini ici en fonction de l'amplitude des modulations de fréquence :

$$\text{RSB} = 10 \log_{10} \frac{A^2/2}{\sigma_b^2} \quad (4.2)$$

Avec A l'amplitude de la modulation de fréquence égale à 1.

Pour tester l'influence de chaque paramètre, trois tests séparés sont faits. À chacun d'eux, deux des trois paramètres sont fixés, afin de regarder l'influence du troisième en fonction du RSB, contrôlé par l'ajout d'un bruit blanc gaussien de variance connue. Pour chaque RSB, 1000 réalisations du mélange « signal + bruit » sont segmentées.

Pour caractériser l'efficacité de la segmentation, on regarde le nombre de classes segmentées au final. Idéalement, trois et seulement trois classes sont créées, une par composante du signal. Nous appellerons « sous-segmentation » le fait d'en obtenir deux ou moins, et « sur-segmentation » le fait d'en obtenir plus de trois. Le critère retenu pour caractériser les paramètres sera les pourcentages de sous-segmentation et de sur-segmentation obtenus sur les 1000 réalisations du bruit. Plus ces pourcentages seront faibles et plus la segmentation sera considérée comme correcte.

Une sur-segmentation apparaît lorsque des motifs issus de fausses alarmes détectées sont segmentés, ou lorsqu'une même composante est segmentée en plusieurs classes. Une sous-segmentation provient d'une absence de segmentation de l'une des composantes. Par conséquent, une segmentation ratant une composante mais qui segmente une autre composante en deux classes conduit à un résultat supposé bon, le nombre de classe final étant correct. Ce cas de figure sera supposé suffisamment rare pour être ignoré.

Toutefois, ce critère permet de donner un sens intuitif à ce que doit être la segmentation, sans avoir de définition précise de ce que doit être le support d'un motif temps-fréquence. On ne raisonne ainsi pas en fonction d'une définition donnée d'un motif temps-fréquence.

4.3.1 Probabilité de fausse alarme p_{fa}

Comme montré par l'équation (2.23), la probabilité de fausse alarme p_{fa} influe sur le seuil de détection selon le niveau de bruit estimé à chaque itération, et par conséquent contrôle le nombre de points candidats. Ainsi, dans le cas de faibles p_{fa} , peu de candidats seront détectés. Les points de plus grands moments d'ordre deux d'un même motif peuvent être détectés de manière disjointe, et segmentés en classes différentes, conduisant à une sur-segmentation d'un motif.

À l'inverse, plus p_{fa} sera élevée, plus il y aura de candidats à une itération donnée. Il sera ainsi plus probable qu'un grand nombre de candidats soient connexes et segmentés ensemble, ce qui est susceptible de créer des sous-segmentations. Le choix de la valeur de p_{fa} est un compromis à faire entre la sur-segmentation et la sous-segmentation.

Enfin, lorsque le niveau de bruit devient correctement estimé, des motifs de fausse alarme seront détectés comme candidats et pourront être segmentés à tort, créant une sur-segmentation.

La figure 4.3 illustre ces effets en montrant les pourcentages de sous-segmentations et de sur-segmentations du signal de test, avec la proportion de candidats p_{cand} fixée à 50%, le seuil sur le kurtosis s_k fixé à 0.1, et pour différents RSB. Pour un RSB de -15dB, le signal n'est jamais totalement segmenté, conduisant à un taux de 100% de sous-segmentation, par conséquent 0% de sur-segmentations.

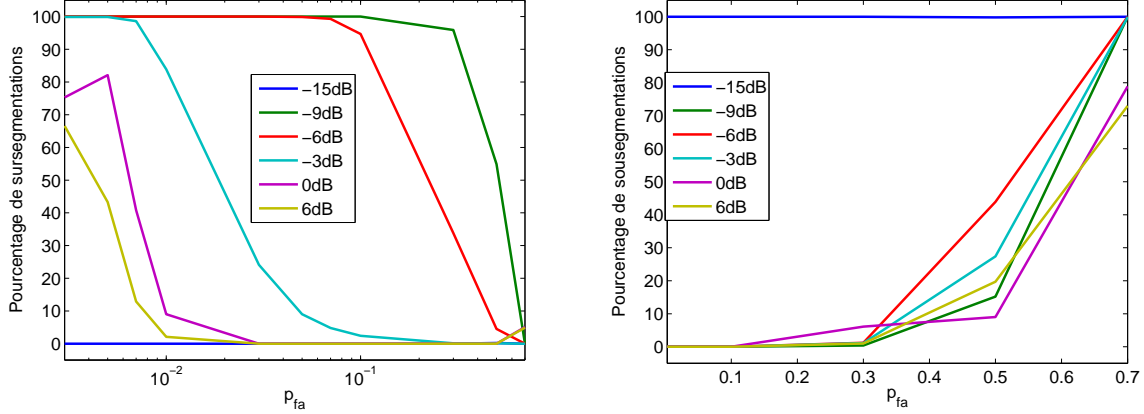


FIG. 4.3 : Pourcentages de sur-segmentations à gauche et de sous-segmentations à droite en fonction de la probabilité de fausse alarme p_{fa} et du RSB, lors de 1000 segmentations d'un mélange « signal de test+bruit ». La proportion de candidats est fixée à $p_{cand} = 0.5$ et le seuil sur le kurtosis à $s_k = 0.1$. Pour des raisons de lisibilité, la courbe de sur-segmentation est présentée en échelle logarithme, celle de sous-segmentation en échelle linéaire.

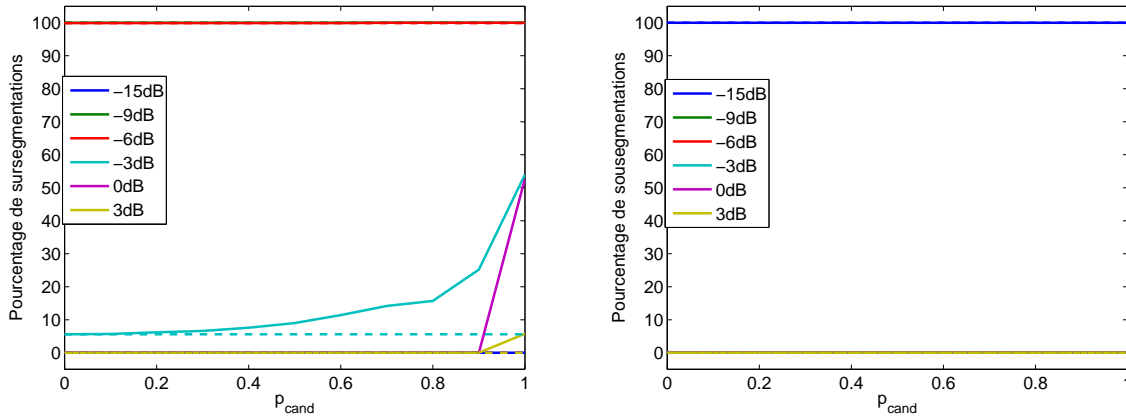


FIG. 4.4 : Pourcentages de sur-segmentations à gauche et de sous-segmentations à droite en fonction de la proportion de candidats p_{cand} segmentés à chaque itération et du RSB, lors de 1000 segmentations d'un mélange « signal de test + bruit ». La probabilité de fausse alarme est fixée à $p_{fa} = 0.1$ et le seuil sur le kurtosis à $s_k = 0.1$. Les traits pointillés indiquent le résultat pour p_{cand} déterminé à chaque itération par l'équation (4.1). Le pourcentage de sur-segmentations et de sous-segmentations est dans ce cas égal aux valeurs obtenues pour $p_{cand} = 10^{-3}$.

Une valeur de p_{fa} inférieure à 0.3 évite toute sous-segmentation, tandis qu'une valeur supérieure à 10^{-2} limite les sur-segmentations.

4.3.2 Proportion de candidats segmentés à chaque sous-boucle p_{cand}

Le paramètre p_{cand} fixe la quantité minimale de candidats qui doivent être segmentés par l'algorithme de croissance de région avant d'estimer à nouveau le bruit. Une valeur proche de 0

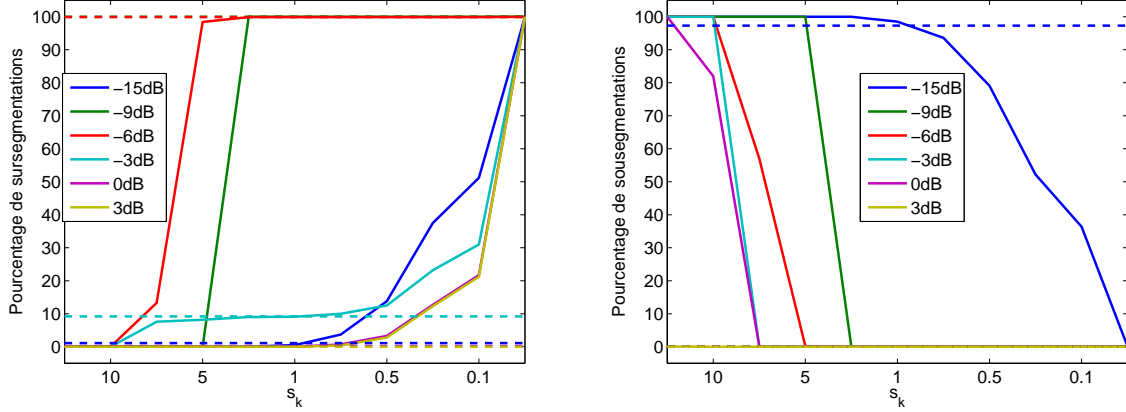


FIG. 4.5 : Pourcentages de sur-segmentations à gauche et de sous-segmentations à droite en fonction du seuil sur le kurtosis s_k et du RSB, lors de 1000 segmentations d'un mélange « signal de test+bruit ». La probabilité de fausse alarme est fixée à $p_{fa} = 0.1$ et la proportion de candidats à $p_{cand} = 0.5$. Les traits pointillés indiquent le résultat pour s_k déterminé à chaque itération par l'équation (3.64). Le pourcentage de sur-segmentations et de sous-segmentations est dans ce cas approximativement égal aux valeurs obtenues pour $s_k = 0.1$.

signifie qu'on va privilégier à chaque itération l'agrandissement des motifs existants, ce qui peut conduire à la segmentation de deux composantes ensemble et donc à des sous-segmentations. À l'opposé, une valeur proche de 1 implique que chaque point détecté, même isolé, sera segmenté dans un motif, ce qui risque de conduire à une sur-segmentation.

Comme montré sur la figure 4.4, une grande valeur de p_{cand} entraîne plus de sur-segmentations qu'une faible valeur. Ce paramètre n'a toutefois aucune influence sur la sous-segmentation du signal de test. À cause de la faible valeur de p_{fa} choisie, la valeur de p_{cand} calculée selon l'équation (4.1) donne les mêmes résultats que pour $p_{cand} = 10^{-3}$.

La proportion de candidats p_{cand} n'a qu'une faible influence sur les mauvaises segmentations, tant qu'elle reste inférieure à 0.9.

4.3.3 Seuil sur le kurtosis des points non segmentés s_k

Le seuil sur le kurtosis détermine la valeur minimale que doit atteindre le kurtosis des points non segmentés afin de les considérer comme issus d'une même loi gaussienne, et par conséquent que tous les points du signal ont été segmentés. Plus ce seuil sera haut, plus l'algorithme s'arrêtera tôt, au risque de ne pas segmenter des motifs complets. Inversement, une valeur trop proche de zéro peut continuer à segmenter des points alors que toutes les composantes ont déjà été segmentées, induisant une sur-segmentation.

La figure 4.5 confirme ces phénomènes. Plus le seuil s_k est bas, plus on sur-segmente, jusqu'à segmenter tout le temps dans le cas d'un seuil négatif. Le seuil calculé automatiquement est dans tous les cas un bon compromis entre sur-segmentations et sous-segmentations.

4.3.4 Influence du Rapport Signal à Bruit

Les figures 4.3, 4.4 et 4.5 montrent que le nombre de mauvaises segmentations, qu'elles proviennent d'une sous-segmentation ou d'une sur-segmentation, diminue lorsque le RSB augmente. Toutefois, le choix optimal des paramètres de la segmentation dépend du RSB, bien que les choix de laisser la proportion de candidats p_{cand} et le seuil du kurtosis s_k calculés par l'algorithme à chaque itération donnent de bons résultats.

4.4 Extensions du principe de segmentation

Pour résumer l'algorithme, l'idée est d'estimer itérativement le bruit et de segmenter des points de signal, jusqu'à ce que les points non segmentés suivent une loi conforme à celle du bruit seul. Ce principe a été appliqué dans le cas de la TFCT et d'un bruit blanc additif gaussien, mais peut facilement être étendu à d'autres cas, traités dans ce paragraphe.

4.4.1 Bruit non blanc ou non stationnaire

L'hypothèse de bruit blanc permet de considérer en un même ensemble tous les points non segmentés. Dans le cas d'un bruit non blanc ou non stationnaire, la loi de l'ensemble des points ne contenant plus de signal ne suit plus une loi gaussienne centrée de variance donnée.

Toutefois, on peut considérer que le bruit est localement blanc et stationnaire, c'est-à-dire que tous les échantillons ne contenant que du bruit sur un voisinage centré sur le point $[n, k]$ suivent une même loi gaussienne de variance $\sigma_b^2[n, k]$. En rajoutant un *a priori* sur le modèle de signal originel, à savoir la taille de ce voisinage, la variance de chaque point temps-fréquence peut être estimée de la même manière que dans le cas du bruit blanc stationnaire, en ne considérant que le voisinage du point. En notant \mathcal{V} le voisinage de V_n points temporels et V_k points fréquentiels, l'estimation du bruit au sens du maximum de vraisemblance (3.13) devient :

$$\widehat{\sigma}_b^2[n, k] = \frac{1}{V_n V_k} \sum (X_\phi^r[n, k]^2 + X_\phi^i[n, k]^2) \quad \forall [n, k] \in \mathcal{V} \quad (4.3)$$

De même, le seuil de détection (2.23) dépend du point temps-fréquence considéré, et s'écrit :

$$s_{\sigma_b^2}[n, k] = \widehat{\sigma}_b^2[n, k] \ln(p_{fa}^{-1}) \quad (4.4)$$

Le test d'arrêt de segmentation nécessite que tous les points non segmentés suivent une même loi gaussienne. Or, si chaque point suit une loi gaussienne de variance différente, la loi de l'ensemble devient un mélange de gaussiennes, et non pas une loi gaussienne. Pour pallier ce problème, l'idée est de blanchir la TFCT uniquement pour calculer le kurtosis des points non segmentés. Le critère d'arrêt devient dans ce cas (3.55) :

$$\widehat{\mathcal{KS}} = \frac{M}{M-1} \left(\frac{(M+1) \sum_{n,k} |X'_\phi[n, k]/\sigma_b[n, k]|^4}{(\sum_{n,k} |X'_\phi[n, k]/\sigma_b[n, k]|^2)^2} - 2 \right) \quad (4.5)$$

Seuls sont pris en compte les coefficients $X'_\phi[n, k]$ non segmentés, avec M leur nombre. Par la normalisation de la variance du bruit à 1, le kurtosis permet de caractériser si tous les points non segmentés proviennent bien d'une loi gaussienne de variance égale à celle estimée localement.

Ce voisinage \mathcal{V} permet de considérer des bruits gaussiens stationnaires mais non blancs, en prenant une taille temporelle V_n égale à la longueur temporelle de la TFCT ; inversement, une taille fréquentielle V_k égale au nombre de fréquences calculées permet de considérer des bruits

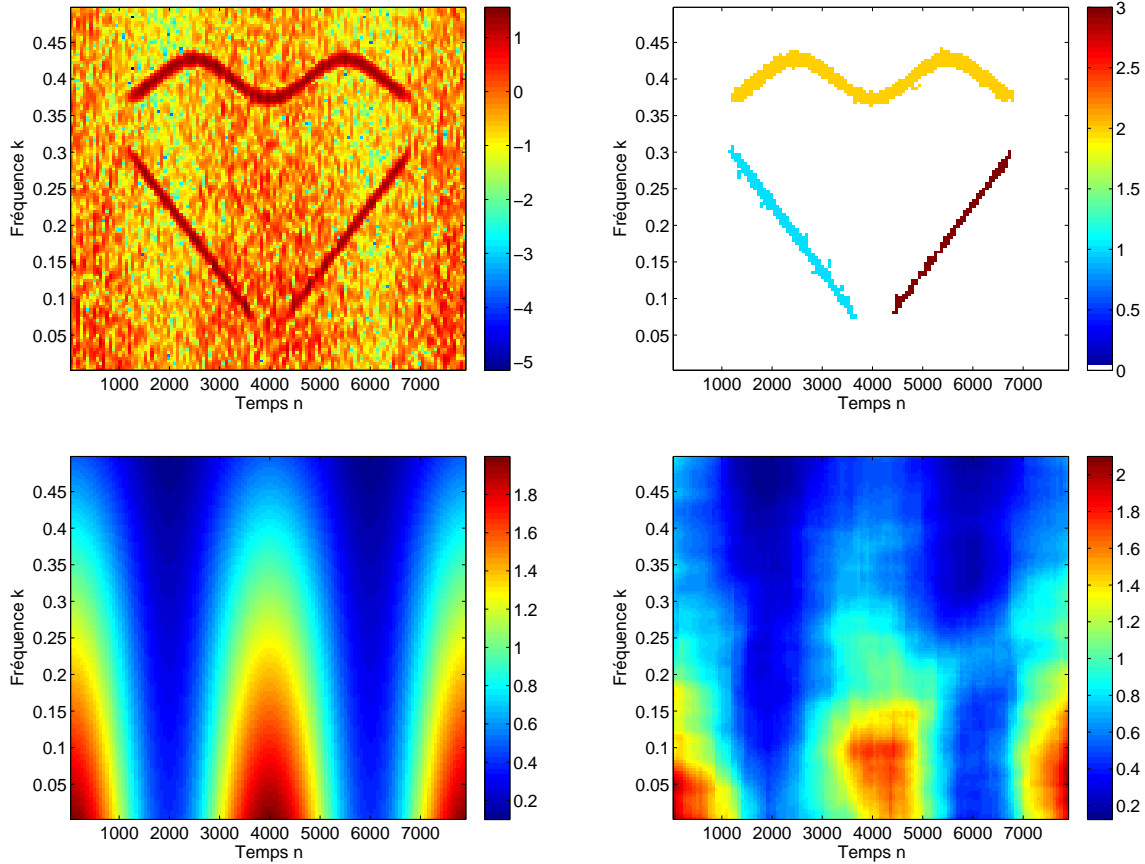


FIG. 4.6 : Illustration de la segmentation avec un bruit non blanc et non stationnaire. En haut à gauche, le spectrogramme bruité du signal de test (figure 4.2) auquel un bruit non blanc et non stationnaire a été ajouté. En haut à droite, la segmentation obtenue, pour une probabilité de fausse alarme $p_{fa} = 0.1$, une proportion de candidats et un seuil sur le kurtosis calculés automatiquement. En bas à gauche, la variance théorique du bruit ajouté, en bas à droite l'estimation du bruit faite par l'algorithme.

blancs non stationnaires. Et de manière évidente, le cas blanc et stationnaire se retrouve en prenant comme voisinage \mathcal{V} l'ensemble du plan temps-fréquence.

La figure 4.6 illustre la segmentation du signal synthétique de la figure 4.2 auquel un bruit non blanc et non stationnaire a été ajouté. La proportion de candidats p_{cand} et le seuil sur le kurtosis ont été calculés automatiquement, tandis que la p_{fa} était fixée à 0.1. Une cellule de 21 points temporels et 21 points fréquentiels a été utilisée pour estimer le bruit. On obtient d'une part la segmentation des composantes spectrales, comme dans le cas du bruit blanc stationnaire, et d'autre part l'estimation du niveau de bruit dans tout le plan temps-fréquence.

4.4.2 Transformée en ondelettes

L'algorithme de segmentation a été décrit pour une transformation linéaire, la TFCT, qui permet d'avoir des coefficients de bruit seul distribués selon une loi gaussienne. Une autre trans-

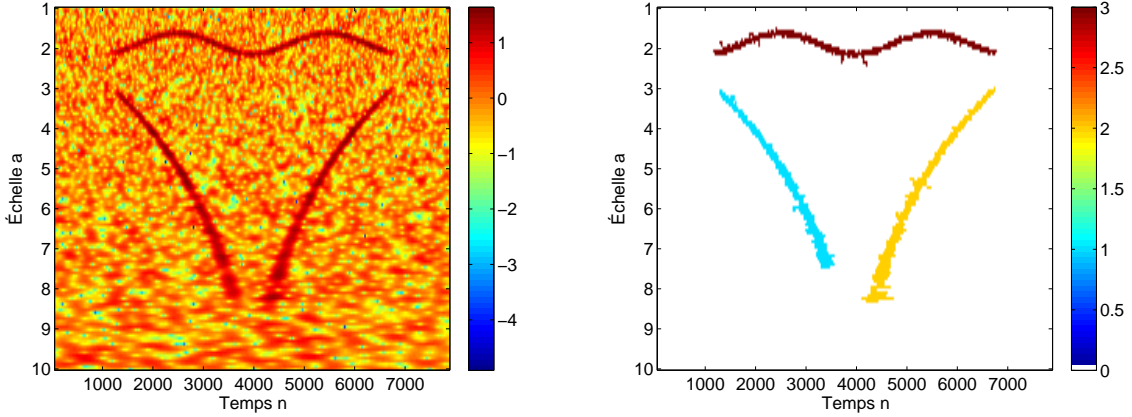


FIG. 4.7 : À gauche, le scalogramme du signal de test de la figure 4.2 avec un bruit blanc gaussien de variance 1, calculé avec une ondelette mère de Morlet de 63 points. À gauche, le résultat de la segmentation pour une probabilité de fausse alarme $p_{fa} = 0.1$, une proportion de candidats p_{cand} estimée par l'algorithme, et un seuil sur le kurtosis $s_k = 0.1$.

formation linéaire est la transformée en ondelettes (TO). Alors que la TFCT a une résolution temps-fréquence constante sur tout le plan temps-fréquence, fixée par le choix de la fenêtre d'analyse ϕ , la TO permet une résolution variable selon l'échelle d'analyse.

La TO projette le signal analysé $x[m]$ sur une base d'ondelettes $\psi_{n,a}[m]$ définies en fonction de l'ondelette mère $\psi_{0,1}[m]$ par :

$$\psi_{n,a}[m] = \frac{1}{\sqrt{a}} \psi_{1,0} \left[\frac{m-n}{a} \right] \quad (4.6)$$

La TO du signal $x[m]$ s'écrit alors :

$$X_\psi[n, a] = \sum_{m=1}^M x[m] \psi_{n,a}^*[m] \quad (4.7)$$

Le choix de la fenêtre d'analyse de la TFCT revient ici à choisir une ondelette mère. Le module carré de la TO, nommé scalogramme, représente l'énergie du signal analysé selon l'échelle d'analyse.

La forme de l'équation de la TO est très similaire à celle de la TFCT (1.1). Par le même raisonnement, et en considérant une ondelette normée en énergie, c'est-à-dire $\sum_m |\psi_{0,1}[m]|^2 = \sum_m |\psi_{n,a}[m]|^2 = 1$, chaque coefficient temps-échelle issu de l'analyse d'un bruit blanc gaussien stationnaire de variance σ_b^2 suit une loi gaussienne de même variance. La différence principale entre les deux transformées vient du choix de l'ondelette, qui peut être réelle ou complexe.

Ainsi, le même algorithme de segmentation peut être appliqué, en utilisant comme critère d'arrêt le kurtosis spectral dans le cas d'une ondelette complexe, et donc d'une TO complexe, ou le kurtosis classique dans le cas d'une ondelette réelle. La figure 4.7 montre la segmentation de la TO du signal de test de la figure 4.2 auquel un bruit blanc gaussien de variance $\sigma_b^2 = 1$ a été ajouté. La TO a été calculée avec une ondelette mère de Morlet de 63 points, et a été segmentée avec une probabilité de fausse alarme $p_{fa} = 0.1$, une proportion de candidats p_{cand} automatiquement choisie, et un seuil sur le kurtosis $s_k = 0.1$.

Les trois composantes spectrales sont correctement segmentées.

4.4.3 Capogramme

Une autre transformée linéaire est le *capogramme*, issu d'une estimation spectrale à court terme par la méthode de Capon, aussi appelée minimum de variance [Cas03, Cap69]. Le principe de l'estimation spectrale par cette méthode est de construire un banc de filtres à réponse impulsionnelle finie de fréquence de coupure f_c selon deux contraintes :

- à la fréquence f_c , le filtre transfère le signal d'entrée avec un gain unité ;
- la puissance de sortie du filtre aux fréquences différentes de f_c est minimisée.

En considérant un signal $x[m]$ échantillonné à la fréquence $F_e = T_e^{-1}$, ces deux contraintes se résolvent à l'aide des multiplicateurs de Lagrange et donnent comme coefficients complexes \mathbf{A}_{f_c} du filtre à la fréquence f_c :

$$\mathbf{A}_{f_c} = \frac{\mathbf{R}_x^{-1} \mathbf{E}_{f_c}}{\mathbf{E}_{f_c}^H \mathbf{R}_x^{-1} \mathbf{E}_{f_c}} \quad (4.8)$$

Avec $(\cdot)^H$ la transposée hermitienne, \mathbf{R}_x la matrice de covariance du signal x tel que $\mathbf{R}_x = \mathbf{E}(\mathbf{X} \cdot \mathbf{X}^H)$ et $\mathbf{E}_{f_c}^T = (1, e^{2\pi j f_c T_e}, \dots, e^{2\pi j (M-1) f_c T_e})$. Le paramètre M est l'ordre de la méthode, qui est à choisir lors de l'estimation. Il correspond au nombre de coefficients du filtre. Plus M est grand, plus la méthode est résolutive, au prix de l'inversion d'une matrice $M \times M$.

Le spectre du signal x est alors estimé par la puissance de sortie P notée $P(f_c)$:

$$P_x(f_c) = \frac{1}{\mathbf{E}_{f_c}^H \mathbf{R}_x^{-1} \mathbf{E}_{f_c}} \quad (4.9)$$

Lagunas [LG84, LSGM86] a proposé une normalisation de ce spectre, pour que l'estimation du spectre soit homogène à une densité spectrale plutôt qu'à une puissance. Cet estimateur noté $S_x(f_c)$ s'écrit :

$$S_x(f_c) = \frac{P_x(f_c)}{1/T_e \mathbf{A}_{f_c}^H \mathbf{A}_{f_c}} \quad (4.10)$$

Dans le cadre d'une représentation temps-fréquence, le capogramme consiste à l'instar de la TFCT à fenêtrer le signal $x[m]$ à différents instants n , et à estimer le spectre par la méthode de Capon normalisée sur ces différentes portions de signal. Les coefficients du capogramme d'ordre M sont alors définis par :

$$C_M[n, k] = S_{x_n}(k) \quad (4.11)$$

Avec $x_n[1..N] = x[n - \frac{N-1}{2} .. n + \frac{N-1}{2}]$ le signal tronqué à N points centré à l'instant n , et k les différentes fréquences de coupures utilisées pour déterminer le banc de filtres $\mathbf{A}_{n,k}$.

Le filtrage étant linéaire, l'accès à la variable filtrée par le filtre \mathbf{A}_{f_c} permet d'utiliser le même algorithme de segmentation que dans le cas d'une TFCT ou d'une TO. Toutefois, l'estimation spectrale par la méthode de Capon se basant sur la matrice de corrélation du signal analysé, le signal filtré n'est pas directement accessible. Plutôt que d'utiliser l'équation (4.10), nous définissons le *capogramme linéaire* :

$$C_{lin,M}[n, k, i] = \frac{1}{\sqrt{1/T_e \mathbf{A}_{f_c}^H \mathbf{A}_{f_c}}} \mathbf{x}_{n,i}^H \mathbf{A}_{n,k} \quad (4.12)$$

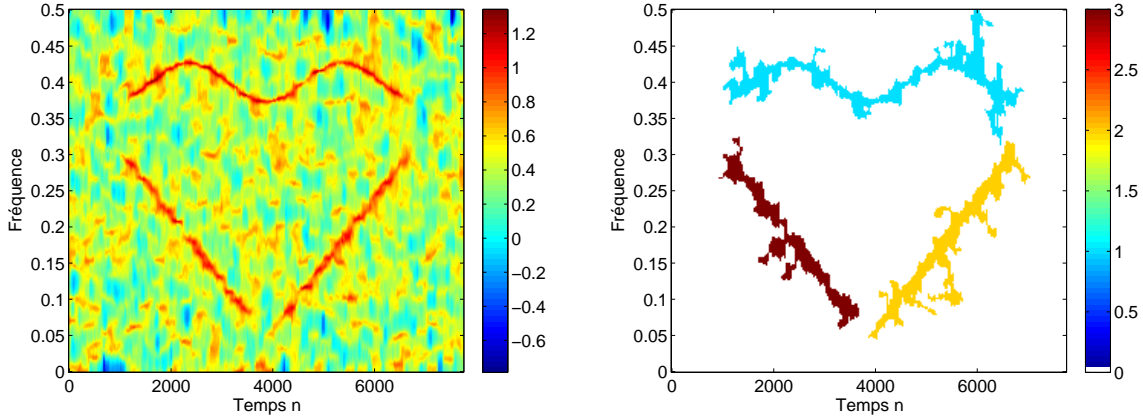


FIG. 4.8 : À gauche, caponogramme normalisé du signal de test de la figure 4.2 auquel un bruit blanc gaussien de variance 2 est ajouté. Le caponogramme est calculé pour un ordre $M = 50$, la fenêtre est de $N = 255$ points, chaque fenêtre séparée de la suivante par 240 points, et 250 fréquences k sont calculées entre les fréquences réduites 0 et 0.5. À droite, le résultat de la segmentation pour une probabilité de fausse alarme $p_{fa} = 0.1$, une proportion de candidats déterminée par l'algorithme, et un seuil sur le kurtosis $s_k = 0.1$

Avec $\mathbf{x}_{n,i}^T = (x_n[i] \dots x_n[M])$, $i \in [1, N - M]$. La redondance d'information dans chaque coefficients $C_{lin,M}[n, k, i]$ permet d'obtenir une bonne estimation de la puissance du signal filtré, et évite l'estimation de la matrice de corrélation du signal.

La relation entre le caponogramme classique et le caponogramme linéaire est :

$$C_M[n, k] = \frac{1}{N - M} \sum_{i=1}^{N-M} C_{lin,M}^H[n, k, i] C_{lin,M}[n, k, i] \quad (4.13)$$

Comme dans le cas de la TFCT et de la TO, chaque coefficient du caponogramme linéaire calculé à partir d'un bruit blanc gaussien de variance σ_b^2 suit une loi gaussienne complexe de variance σ_b^2 . Toutefois, étant donné les $N - M$ coefficients $C_{lin,M}[n, k, i]$ pour chaque point temps-fréquence, l'énergie du point $[n, k]$, autrement dit le coefficient $C_M[n, k]$, ne suit pas une simple loi du χ^2 de degré deux. De plus, la grande corrélation présente entre ces coefficients implique que la loi de $C_M[n, k]$ ne suit pas non plus une loi du χ^2 de degré $M - N$, mais de degré inférieur.

Cette loi est approximée par une loi du χ^2 de degré $\delta = N/M$, correspondant au nombre de variables indépendantes parmi les coefficients $C_{lin,M}[n, k, i]$.

Une fois le degré de liberté de la loi du χ^2 estimé, l'algorithme de segmentation se déroule de la même manière que pour les autres transformées. La variance du bruit est estimée sur les points non segmentés, qui avec le degré de liberté δ caractérise entièrement la loi du bruit seul. La stratégie de Neyman-Pearson donne un seuil de détection, en fonction de la probabilité de fausse alarme p_{fa} choisie. L'algorithme s'arrête lorsque le kurtosis de l'ensemble des coefficients $C_{lin,M}[n, k, i]$ devient inférieur au seuil s_k .

La figure 4.8 montre la segmentation du signal de test de la figure 4.2 additionné d'un bruit blanc gaussien de variance 2. Le caponogramme est calculé pour un ordre $M = 50$, la fenêtre est de $N = 255$ points, chaque fenêtre séparée de la suivante par 240 points. 250 fréquences k sont calculées entre les fréquences réduites 0 et 0.5. La segmentation se fait avec une probabilité de

fausse alarme $p_{fa} = 0.1$, une proportion de candidats déterminée par l'algorithme, et un seuil sur le kurtosis $s_k = 0.1$.

Encore une fois, les composantes spectrales sont bien détectés, le caponogramme permettant de les détecter avec un bruit de variance 2.

4.5 Applications

Cette section propose quelques exemples d'application d'un algorithme de segmentation. De manière évidente, la segmentation de motifs temps-fréquence à partir d'une représentation bruitée permet de mettre en avant les zones temps-fréquence d'intérêt, et ainsi de permettre une première lecture simplifiée du signal analysé.

Quelques applications plus précises peuvent être envisagées.

4.5.1 Débruitage

Le débruitage est l'opération de filtrage du bruit, afin de rendre le signal analysé plus lisible et plus facilement exploitable. Une segmentation temps-fréquence permet d'effectuer un filtrage non-stationnaire, en mettant à zéro les coefficients de la TFCT qui n'ont pas été détectés comme contenant du signal. Le signal temporel filtré s'obtient alors par inversion de la TFCT filtrée. Cette méthode est alors équivalente dans le cas stationnaire à un filtrage spectral à l'aide d'une fenêtre rectangulaire. La figure 4.9 montre le spectrogramme issu de la TFCT du signal de test (figure 4.2) filtrée par sa segmentation.

Malgré le choix simpliste de la fenêtre de lissage, le débruitage se révèle très bon avec de faibles erreurs de reconstruction. Pour caractériser l'efficacité de ce filtrage, nous définissons le RSB avant filtrage RSB_{av} , le RSB après filtrage RSB_{ap} et l'erreur de reconstruction ϵ , respectivement par :

$$RSB_{av} = 10 \log_{10} \left(\frac{\sum s[m]^2}{\sum b[m]^2} \right) \quad (4.14)$$

$$RSB_{ap} = 10 \log_{10} \left(\frac{\sum s_{rec}[m]^2}{\sum (s[m] - s_{rec}[m])^2} \right) \quad (4.15)$$

$$\epsilon = \frac{\sum (s[m] - s_{rec}[m])^2}{\sum s[m]^2} \quad (4.16)$$

Où $s[m]$ est le signal seul de longueur M , $b[m]$ le bruit seul et $s_{rec}[m]$ le signal reconstruit par inversion de la TFCT filtrée.

En utilisant le signal de test de la figure 4.9 auquel on ajoute des bruits blancs gaussiens de variance connue σ_b^2 , le tableau 4.1 donne les valeurs de l'erreur de reconstruction ϵ , du RSB avant et après filtrage, en fonction de la variance du bruit. Les filtrages ont été effectués à partir d'une segmentation avec une probabilité de fausse alarme $p_{fa} = 0.1$, les paramètres p_{cand} et s_k étant calculés par l'algorithme.

Pour des valeurs trop grandes de σ_b^2 (16 et 8), le signal est entièrement noyé dans le bruit et la segmentation ne retrouve que des motifs de fausse alarme liés au bruit. Cela conduit à des erreurs de reconstruction proches de 100%. Pour $\sigma_b^2 = 4$, seules quelques portions du support temps-fréquence complet du signal sont segmentées, l'erreur finale reste grande. Pour des valeurs de σ_b^2 plus faible, le signal est correctement segmenté, l'erreur est alors faible. Lorsque la segmentation se déroule correctement, le gain après filtrage est de l'ordre de 10dB.

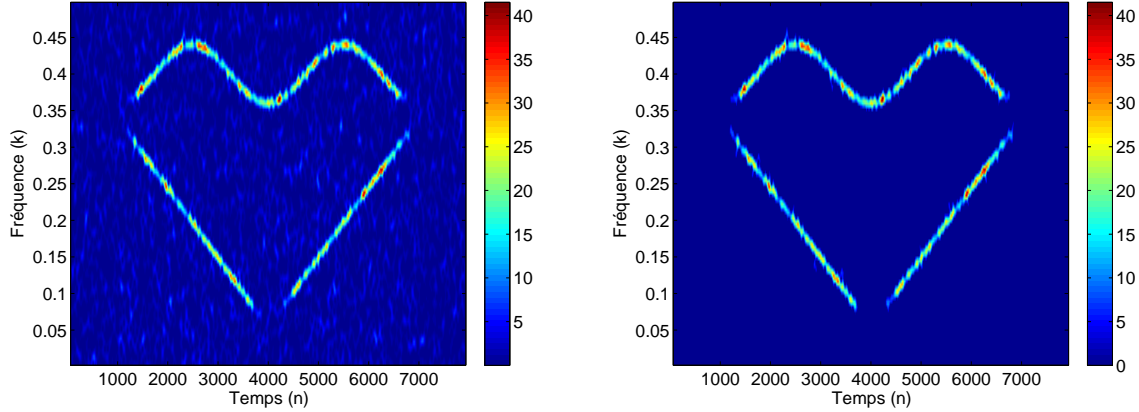


FIG. 4.9 : À gauche, spectrogramme du signal de test bruité, avec un RSB de 0dB. À droite, spectrogramme issu de la TFCT filtrée par la segmentation. Tous les coefficients supposés ne contenir que du bruit ont été mis à zéro, les autres étant inchangés.

σ_b^2	16	8	4	2	1	1/2	1/4	1/8	1/16	1/32	1/64
$RSB_{av}(dB)$	-14	-11	-8	-5	-1.9	1	4.1	7.1	10.1	13.1	16.1
$RSB_{ap}(dB)$	—	-8.1	-0.8	5.1	8.5	11.6	14.7	17.4	20.1	22.9	25.8
ϵ (%)	100	99,4	60,7	23,5	11,7	5,9	3,1	1,8	0,9	0,5	0,3

TAB. 4.1 : RSB avant filtrage RSB_{av} , RSB après filtrage RSB_{ap} et erreur de reconstruction ϵ lors du filtrage du signal de test, pour différentes valeurs de variance du bruit additif σ_b^2 .

Cette méthode de filtrage a été utilisée avec succès dans le cadre de la reconnaissance de parole par [RLEM03, LBR⁺04].

4.5.2 Séparation de sources

Le but de la séparation de sources aveugle est de déterminer les signaux sources à partir d'un mélange de ces signaux. Quand le nombre de sources est supérieur au nombre de mélanges, le problème est dégénéré puisque la technique classique d'inversion matricielle ne peut être appliquée. Toutefois, il est possible de séparer les signaux à condition de trouver une représentation dans laquelle les supports temps-fréquence des sources sont disjoints, puis en segmentant ces supports.

La séparation de sources par des méthodes temps-fréquence a déjà été étudiée [Row00, YR04] dans le cadre de signaux de parole, en utilisant un masque temps-fréquence binaire.

L'algorithme de segmentation proposé dans ces travaux peut être utilisé pour déterminer les masques binaires nécessaires à la séparation de sources, dans le cas où les motifs dans le plan temps-fréquence sont disjoints. Cette condition est appelée *W-disjoint orthogonality* [YR04]. Ainsi, une séparation de sources mono-capteur, c'est-à-dire avec un seul mélange de sources, est réalisable.

Contrairement au cas du débruitage où l'ensemble des points segmentés sert de masque de filtrage, seuls les points segmentés comme appartenant à une classe donnée servent à créer le

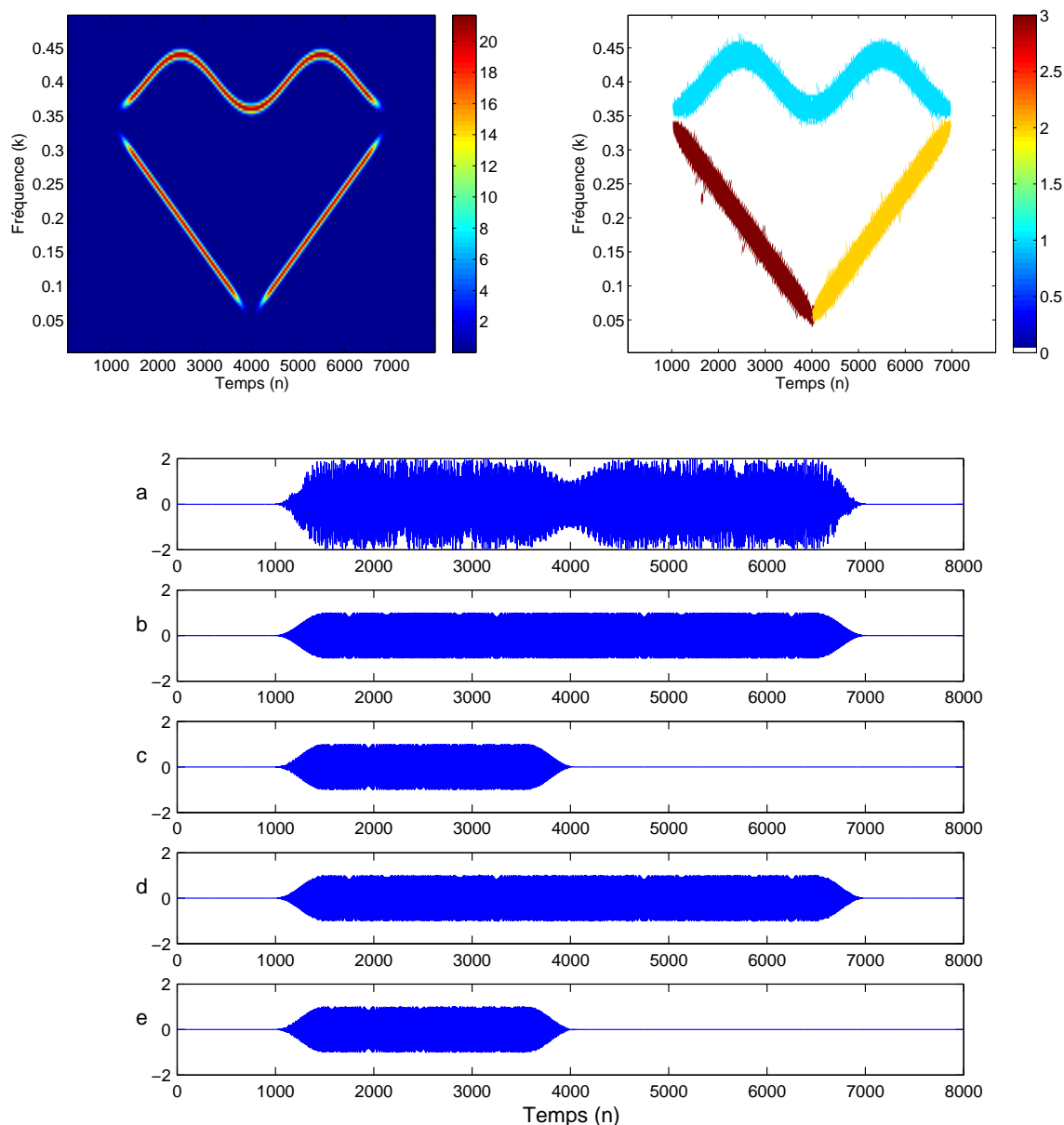


FIG. 4.10 : En haut à droite, spectrogramme du signal de test, mélange de trois composantes. Pour les besoins de convergence de l'algorithme, un bruit blanc de variance $\sigma_b^2 = 1/1000$ a été ajouté. En haut à gauche, segmentation de la TFCT du signal, chaque composante est affectée à une et une seule classe. La courbe a représente l'allure temporelle du signal segmenté, la courbe b celle de la modulation sinusoïdale de fréquence, la courbe c celle de la première modulation linéaire de fréquence. Les courbes d et e sont respectivement les signaux reconstruits de la modulation sinusoïdale de fréquence et de la première modulation linéaire de fréquence à partir du masque de la segmentation des classes 1 et 3.

masque temps-fréquence.

La figure 4.10 illustre une séparation de sources, à partir du signal de test de la figure 4.2.

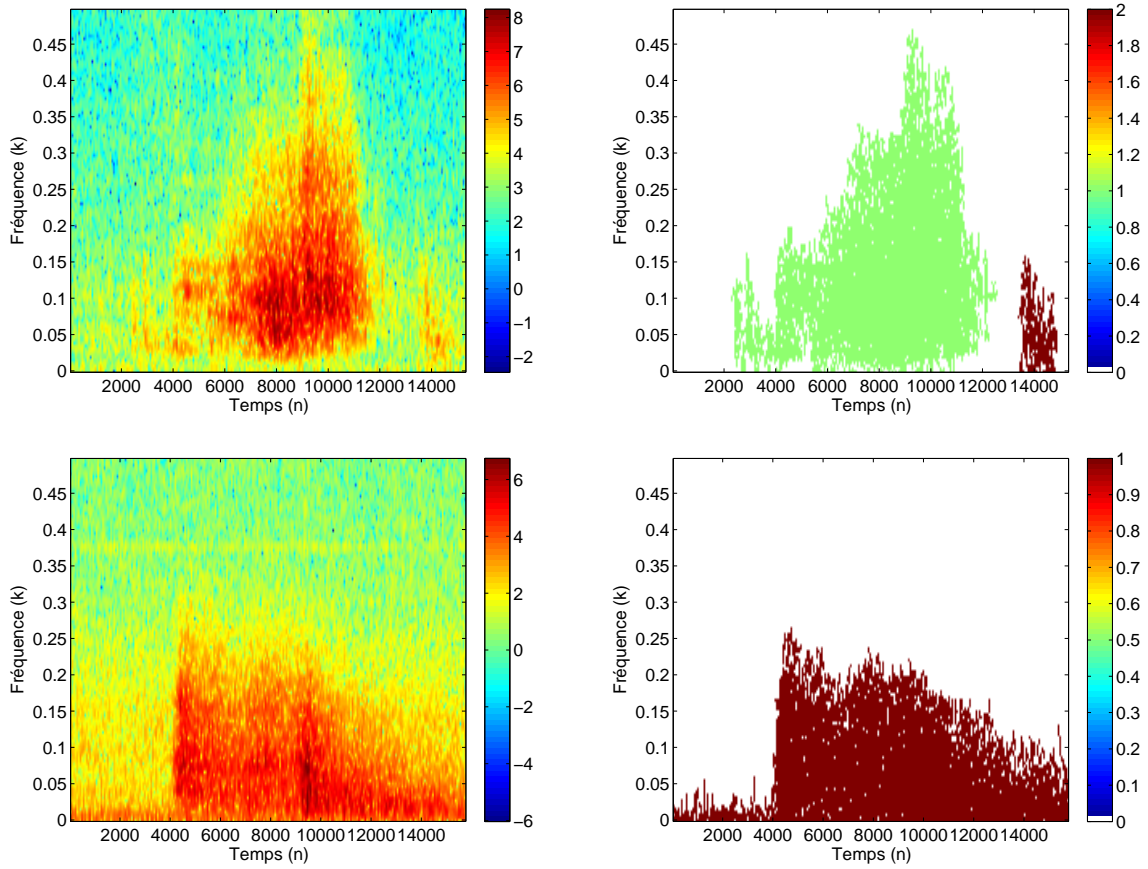


FIG. 4.11 : À gauche, spectrogrammes d'un signal d'avalanche (en haut) et d'un séisme (en bas). À droite, les segmentations associées.

Ce signal est construit à partir de trois composantes dont les supports temps-fréquence sont disjoints. L'algorithme de segmentation se basant sur les caractéristiques du bruit du modèle « signal + bruit », un bruit blanc gaussien de variance négligeable, $\sigma_b^2 = 1/1000$, a été ajouté pour la segmentation.

La segmentation attribue une classe unique à chacun des supports des composantes du signal. En considérant successivement comme masque binaire temps-fréquence les ensembles de points segmentés dans les classes 1, 2, et 3, la modulation sinusoïdale de fréquence, la deuxième modulation linéaire de fréquence et la première modulation linéaire de fréquence sont reconstruites.

4.5.3 Caractérisation de signaux

Une autre application possible d'une segmentation temps-fréquence se trouve dans la caractérisation et la classification de signaux. La classification de signaux est équivalente à un test d'hypothèse, tel que décrit dans le chapitre 2. Dans ce cas, les hypothèses à discriminer ne sont plus l'hypothèse bruit seul et l'hypothèse signal, mais des hypothèses liées à des types de signaux recherchés.

Dans le cas où ces signaux possèdent des caractéristiques temps-fréquence distinctes les unes des autres, la segmentation permet de les extraire automatiquement et ainsi sert de base au test

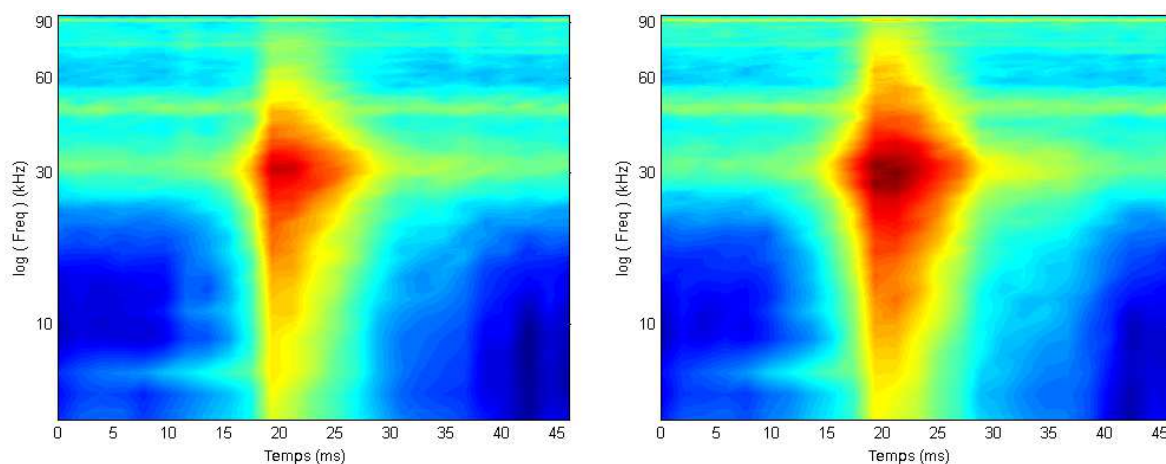


FIG. 4.12 : Caponogrammes d'un bruit de Barkhausen provenant d'un acier de dureté 53.6 HRC à gauche, de dureté 37.6 HRC à droite. Le bruit le plus énergétique correspond à la dureté la plus faible.

d'hypothèse effectuant la classification.

Une telle classification a été utilisée dans le cadre d'une collaboration avec Météo France en 2004 [Sax04]. À partir d'un capteur placé en montagne, le but était de compter les avalanches, en évitant de comptabiliser d'autres signaux tels les séismes ou les hélicoptères.

La figure 4.11 montre les spectrogrammes d'un signal d'avalanche et d'un signal de séisme. Les répartitions de l'énergie dans le plan temps-fréquence de ces deux signaux sont bien différentes, et sert de critère afin de les discriminer. Les segmentations sont calculées en prenant comme paramètre de segmentation $p_{fa} = 0.1$, une proportion de candidats p_{cand} automatique et un seuil sur le kurtosis $s_k = 6$. Le seuil sur le kurtosis est choisi élevé afin de ne segmenter que les plus grandes valeurs du signal.

En définissant deux gabarits, retranscrivant les formes typiques des motifs temps-fréquence des avalanches et des séismes respectivement, il a ainsi été possible de mettre en place un test d'hypothèse, basé sur une distance entre le motif segmenté et les gabarits.

Une seconde application de caractérisation de signaux est présentée dans [PMM09], pour caractériser la dureté d'aciers par analyse de leur bruit de Barkhausen. Ce bruit est un phénomène magnétique produit quand un champ magnétique variable induit des mouvements de paroi dans les matériaux ferromagnétiques. Ces mouvements, non pas continus mais discrets, sont causés par des défauts dans la microstructure du matériau, et génèrent des impulsions magnétiques qui peuvent être mesurées par une bobine placée sur la surface du matériau.

Ce bruit renseignant sur la structure du matériau, dépend ainsi de la dureté de l'acier analysé [LNPW00, JPH⁺99]. Toutefois, les techniques usuelles de caractérisation du bruit de Barkhausen ne considèrent que des grandeurs à une dimension, comme l'énergie du signal ou sa valeur maximale, ou des grandeurs à deux dimensions, comme son enveloppe. Les représentations temps-fréquence peuvent se montrer utiles, comme la figure 4.12 l'illustre avec les caponogrammes de deux aciers SAE 4140, de dureté de Rockwell de 53.6 HRC et 37.6 HRC. Comme prévu, plus la dureté est grande, moins le bruit est énergétique.

Pour caractériser ces duretés, une première méthode s'intéresse aux courbes iso-niveaux des

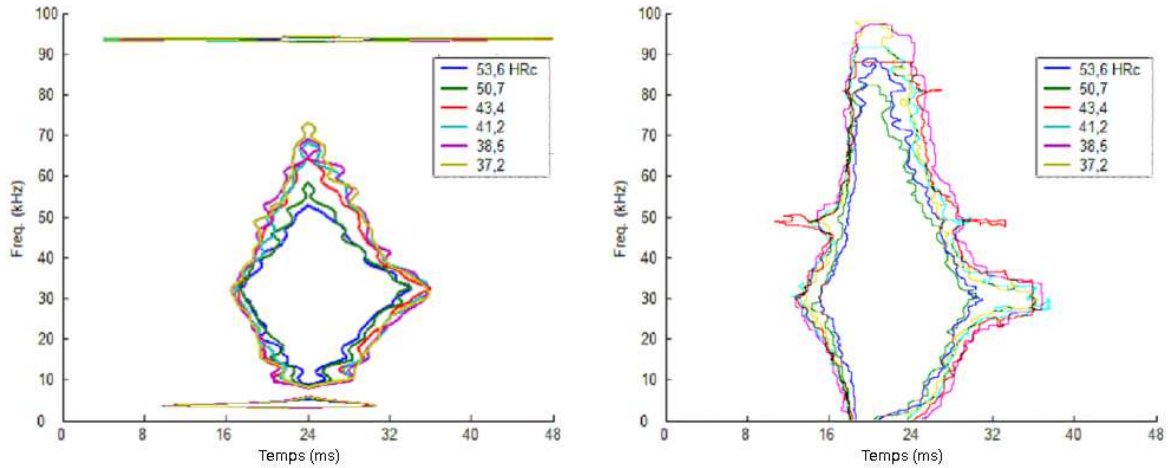


FIG. 4.13 : À gauche, le tracé des courbes iso-niveaux des caponogrammes des bruits de Barkhausen de 6 aciers de dureté différente. L'iso-niveau correspond à 40 de la valeur maximale. À droite, contours obtenus par l'algorithme de segmentation, pour un probabilité de fausse alarme $p_{fa} = 10^{-2}$ et un seuil sur le kurtosis $s_k = 3.5$.

caponogrammes, à 40 % du niveau maximum de leur représentation temps-fréquence. La figure 4.13 à gauche montre les six courbes correspondant à autant d'aciers de dureté différente.

Pour automatiser l'analyse, et ne plus dépendre du choix *ad hoc* de la courbe iso-niveau à 40 % de la valeur maximale, l'algorithme de segmentation a été appliqué sur chacun des six caponogrammes avec une probabilité de fausse alarme $p_{fa} = 10^{-2}$, une proportion de candidats calculée selon (4.1). Dans ce cas, le bruit additif n'est pas blanc : il y a notamment des fréquences pures à environ 30 et 50kHz. Le seuil du kurtosis est par conséquent fixé à $s_k = 3.5$, correspondant au kurtosis des points qui sont considérés comme ne contenant que du bruit, c'est-à-dire ceux compris entre 0 et 5 secondes d'une part, et entre 40 et 45 secondes d'autre part. La figure 4.13 à droite en donne le résultat pour les six aciers considérés. Pour des soucis de lisibilité, seuls les points extrêmes des motifs, autrement dit les contours, sont affichés.

De manière similaire aux courbes iso-niveaux, les segmentations sont d'autant plus fines que la dureté de l'acier analysé est grande.

Conclusion

Dans cette première partie, nous avons présenté une méthode automatique de segmentation de la Transformée de Fourier à Court Terme (TFCT), basée sur les propriétés statistiques d'un signal analysé. Le signal analysé est modélisé par un signal d'intérêt, non-stationnaire et inconnu, auquel s'ajoute un bruit blanc gaussien de variance inconnue.

De ce modèle de signal résultent les propriétés statistiques des coefficients temps-fréquence : dans le cas où un coefficient ne provient que du bruit, il suit une loi gaussienne ; sinon, sa loi possède un moment d'ordre deux supérieur au cas du bruit seul. Il apparaît que tous les coefficients de la TFCT d'un bruit blanc gaussien ne suivent pas une même loi gaussienne, dont nous avons caractérisé la variance selon son emplacement sur le plan temps-fréquence.

Toutefois, ces différences de variance sont connues et ne dépendent que des paramètres de la TFCT. Il est possible de définir une TFCT rectifiée telle que chaque coefficient issu d'un bruit blanc suive une même loi gaussienne. Chacun de ces coefficients est alors circulaire.

La segmentation s'effectue à partir de cette TFCT rectifiée, par une succession d'itération. À chaque itération, le niveau de bruit est estimé sur les points non segmentés, ce qui conduit à sa surestimation si des points contenant du signal ne sont pas encore segmentés. Grâce à cette estimation et par une détection de type Neyman-Pearson, les points temps-fréquence ayant les plus grands moments d'ordre deux sont détectés comme contenant du signal, et segmentés en classe par un algorithme de croissance de région.

Le critère d'arrêt de l'algorithme est basé sur le kurtosis spectral des coefficients non segmentés. Dans le cas d'une variable complexe circulaire, le kurtosis spectral, défini directement à partir de la variable complexe, a été démontré égal aux deux tiers du kurtosis de ses parties réelle et imaginaire. Il est par conséquent nul s'il est estimé sur un ensemble de points provenant d'une loi gaussienne complexe circulaire, comme la loi de la TFCT rectifiée d'un bruit blanc gaussien. Un seuil sur le kurtosis spectral, que l'algorithme peut estimer automatiquement, sert ainsi de critère d'arrêt des itérations d'estimation et de détection.

Outre les paramètres de la TFCT, la segmentation possède trois paramètres intrinsèques paramétrant son comportement. Ces paramètres sont la probabilité de fausse alarme utilisée pour la détection des points contenant du signal, la proportion de candidats segmentés à chaque itération par la croissance de région, et le seuil sur le kurtosis qui définit la fin de la segmentation. L'influence de chacun de ses paramètres a été étudiée sur un signal de test, ainsi que l'intérêt de la détermination automatique par l'algorithme de la proportion de candidats et du seuil sur le kurtosis.

Le principe de segmentation se base sur la propriété de linéarité de la TFCT, qui préserve le caractère gaussien du signal dans le plan temps-fréquence. Ainsi, la transformée en ondelettes, qui est au scalogramme ce que la TFCT est au spectrogramme, et le caponogramme, qui applique

un filtre à réponse impulsionnelle finie pour l'obtention de chaque coefficient temps-fréquence, ont été utilisés comme représentation temps-fréquence à segmenter.

La contrainte de stationnarité et de blancheur du bruit blanc gaussien additif dans tout le plan temps-fréquence est très forte, et peut être remplacée par une stationnarité et une blancheur locale. En estimant le bruit sur un voisinage local pour chaque point temps-fréquence, l'algorithme de segmentation est adaptable aux bruits non blancs et non stationnaires. Une détection est faite sur chaque point en fonction de son niveau de bruit estimé, tandis que le kurtosis spectral est calculé sur l'ensemble des points non segmentés, mais normalisés par la variance de leur bruit associé afin de retomber dans le cas blanc et stationnaire.

Enfin, la segmentation est illustrée sur différentes applications, dans des contextes de débruitage, de séparation de sources et de caractérisation et classification de signaux. Le modèle temporel de signal étant peu contraignant, ce principe de segmentation est susceptible de s'employer sur une large gamme de signaux.

Deuxième partie

Segmentation du plan temps-fréquence par les vecteurs de réallocation

Table des matières

Introduction	73
5 Réallocation	75
5.1 Réallocation du spectrogramme, interprétation du centre de gravité	75
5.2 Réallocation du spectrogramme, interprétation de la fréquence instantanée et du retard de groupe	78
5.2.1 Fréquence instantanée et retard de groupe	78
5.2.2 Mise en œuvre de la réallocation	80
5.3 Champ de vecteurs de réallocation	80
6 Détection de contour	83
6.1 Discrétisation du champ de vecteurs de réallocation	84
6.2 Détection de zone frontière	85
6.2.1 Application sur un signal vibratoire	87
6.3 Construction du contour	87
6.3.1 Degré de parallélisme local	87
6.3.2 Algorithme de construction de contour	89
6.4 Cas de signaux bruités	91
6.5 Influence des paramètres de la TFCT	94
6.6 Pour améliorer l'algorithme...	95
6.6.1 Réallocation multi-fenêtres	96
6.6.2 Réallocation pseudo multi-fenêtres	97
6.6.3 Lissage du résultat obtenu par contour actif	98
Conclusion	101
Conclusions générales	103
Discussion	105
Perspectives	106

Introduction

Dans la première partie de ce document, nous avons présenté une méthode de segmentation de représentations temps-fréquence basée sur un modèle temporel du signal, à savoir un signal quelconque à segmenter auquel s'ajoute un bruit gaussien. Les propriétés statistiques des coefficients temps-fréquence ne contenant que du bruit permettent de détecter quels sont les points contenant du signal. La segmentation crée des motifs spectraux à l'aide d'un critère de connexité dans le plan temps-fréquence, par l'intermédiaire d'un algorithme de croissance de région.

Outre la connexité, le critère déterminant la forme et la taille d'un motif temps-fréquence est lié à l'ensemble des points détectés comme contenant du signal, et donc du niveau de bruit réel et estimé. Plus fort sera le bruit, plus faible sera le nombre de points détectés, à probabilité de fausse alarme constante. Ainsi, le motif temps-fréquence s'amenuise à mesure que le niveau de bruit augmente. La définition du support du motif est donc directement liée au choix d'un seuil dépendant de la puissance du bruit additif.

Dans cette seconde partie, nous proposons une toute autre méthode de segmentation, visant à reconstruire le contour d'un motif temps-fréquence. De telles détections de contours existent déjà, par exemple [LM02], basée sur la morphologie mathématique. Cependant, ce type de méthodes basées sur le gradient de la représentation temps-fréquence ne donne pas des résultats satisfaisants pour de faibles contrastes.

L'idée est d'utiliser la méthode de la réallocation du spectrogramme, non pas en tant que représentation temps-fréquence, mais comme source d'information sur la structure temps-fréquence du signal afin de définir un contour. Cette approche permet de se passer de la définition d'un bruit additif, et fournit ainsi un contour basé sur les propriétés du signal et de la transformée temps-fréquence.

Le chapitre 5 présente le principe de la méthode de réallocation appliquée au spectrogramme, conduisant à la création d'un champ de vecteurs de réallocation. Ce champ de vecteurs est à la base de la définition du contour d'un motif temps-fréquence proposée dans le chapitre 6. Une méthode pour détecter les zones temps-fréquence contenant le contour sera décrite, pour construire le contour en lui-même par un algorithme de détection de crêtes. Enfin, quelques pistes d'amélioration de cette segmentation seront proposées.

Chapitre 5

Réallocation

La réallocation est une méthode originellement introduite par Kodera, Gendrin et Villedary [KVG76, KVG78] en 1976. L'idée est de refocaliser une distribution naturellement lissée, le spectrogramme. Cette méthode a été étendue par Auger et Flandrin [AF94, AF95] plusieurs années plus tard à la classe de Cohen, l'ensemble des distributions d'énergie en temps-fréquence bilinéaires covariantes par les translations en temps et en fréquence, et à la classe affine, qui regroupe les distributions d'énergie temps-échelle covariantes par les translations en temps et en échelle.

Seule la réallocation du spectrogramme sera présentée dans ce chapitre.

5.1 Réallocation du spectrogramme, interprétation du centre de gravité

Comme décrit dans la première partie (1.1,1.2), le spectrogramme $S_{\phi,x}(t, f)$ d'un signal $x(t)$ est le module carré de la Transformée de Fourier à Court Terme. En considérant des variables continues, il s'écrit :

$$S_{\phi,x}(t, f) = X_{\phi}(t, f)X_{\phi}(t, f)^* = \left| \int_{-\infty}^{+\infty} x(s)\phi^*(s-t)e^{-2\pi jfs} ds \right|^2 \quad (5.1)$$

Le spectrogramme peut également être exprimé à partir de la distribution de Wigner-Ville $W_x(t, f)$, définie pour un signal $x(t)$ par :

$$W_x(t, f) = \int_{-\infty}^{+\infty} x(t + \tau/2)x^*(t - \tau/2)e^{-2\pi jf\tau} d\tau \quad (5.2)$$

Cette expression peut s'interpréter comme la transformée de Fourier de la fonction d'intercorrélation instantanée $C_x(t, s) = x(t + s/2)x^*(t - s/2)$, de la même manière que la densité spectrale de puissance d'un signal stationnaire s'obtient par la transformée de Fourier de sa fonction de corrélation (théorème de Wiener-Khintchine).

Toutefois, la nature bilinéaire de la distribution de Wigner-Ville fait qu'il y a une apparition d'artefacts, nommés interférences [Fla93], lorsque le signal analysé est constitué de plusieurs composantes. Ces interférences sont constituées de valeurs positives et négatives en des lieux temps-fréquence où aucune énergie n'est censée être présente, constituant ainsi un frein à une interprétation facile de la distribution de Wigner-Ville.

Étant donné que les interférences présentent des valeurs positives et négatives, un lissage temps-fréquence, autrement dit un filtre passe-bas, permet de les atténuer. Il est possible de

définir ainsi toutes les distributions $C_x(t, f)$ temps-fréquence de la classe de Cohen, selon le noyau de lissage $\Pi(t, f)$:

$$C_x(t, f) = \iint W_x(t, f) \Pi(\tau - t, \nu - f) d\tau d\nu = W_x(t, f) * \Pi(\tau - t, \nu - f) \quad (5.3)$$

avec $*$ le produit de convolution.

Le spectrogramme défini par l'équation (5.1) appartient à la classe de Cohen ; il peut également s'interpréter comme la représentation de Wigner-Ville lissée par un noyau $W_\phi(t, f)$ égal à la distribution de Wigner-Ville de la fenêtre $\phi(t)$:

$$S_x^h(t, f) = \iint W_x(\tau, \nu) W_\phi(\tau - t, \nu - f) d\tau d\nu \quad (5.4)$$

Dans le cas du spectrogramme, le lissage permet de supprimer toutes les interférences issues de la distribution de Wigner-Ville du signal $x(t)$, au prix d'une perte de résolution temps-fréquence. Ces phénomènes sont illustrés par la figure 5.1 : les interférences entre les deux modulations de fréquence sont parfaitement supprimées, tandis que les supports temps-fréquence des modulations sont élargis.

Partant de ces constats, l'idée de la réallocation du spectrogramme est de « refocaliser » la répartition temps-fréquence de l'énergie, étant donné que l'on connaît par le noyau de lissage $W_h(t, f)$ la manière dont elle a été étalée. L'énergie d'un point du spectrogramme (t, f) est alors déplacée au point (t', f') , défini par le barycentre de la distribution de Wigner-Ville du signal dans un voisinage fonction de la fenêtre d'analyse $W_\phi(t, f)$:

$$t' = t'_{\phi,x}(t, f) = \frac{1}{S_{\phi,x}(t, f)} \iint \tau W_x(\tau, \nu) W_\phi(\tau - t, \nu - f) d\tau d\nu \quad (5.5)$$

$$f' = f'_{\phi,x}(t, f) = \frac{1}{S_{\phi,x}(t, f)} \iint \nu W_x(\tau, \nu) W_\phi(\tau - t, \nu - f) d\tau d\nu \quad (5.6)$$

Le spectrogramme réalloué est obtenu par le déplacement des valeurs du spectrogramme des points (t, f) aux points $(t'_{\phi,x}(t, f), f'_{\phi,x}(t, f))$, et s'écrit :

$$S'_{\phi,x}(t, f) = \iint S_{\phi,x}(t, f) \delta(t - t'_{\phi,x}(\tau, \nu), f - f'_{\phi,x}(\tau, \nu)) d\tau d\nu \quad (5.7)$$

La figure 5.1 illustre la construction d'un spectrogramme réalloué, composé de deux modulations linéaires de fréquence. Dans la représentation de Wigner-Ville, les modulations linéaires de fréquence sont représentées avec une bonne localisation temps-fréquence, mais des termes oscillants apparaissent entre les deux droites correspondant au signal. Le spectrogramme lisse la représentation de Wigner-Ville par un noyau de lissage correspondant à la distribution de Wigner-Ville de la fenêtre d'analyse du spectrogramme. La réallocation du spectrogramme (5.7) refocalise l'énergie étalée par ce lissage, en déplaçant l'énergie de chaque point du spectrogramme sur son centre de gravité local. La distribution ainsi obtenue se rapproche alors de la représentation idéale du signal, conjuguant la bonne localisation temps-fréquence de Wigner-Ville à l'absence de termes d'interférences du spectrogramme.

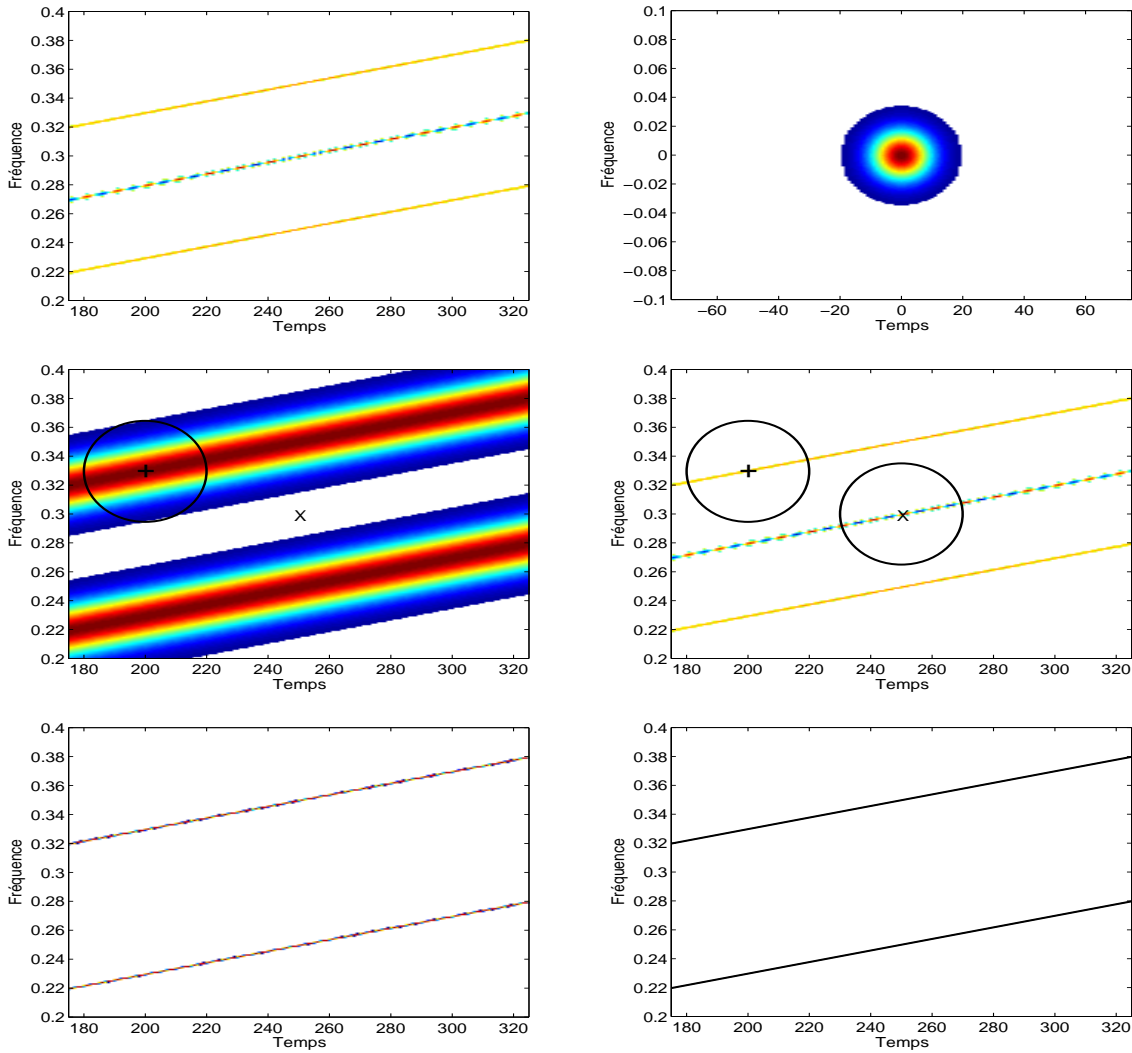


FIG. 5.1 : Illustration de la réallocation du spectrogramme. Le signal $x(t)$ est constitué de deux modulations linéaires de fréquence. Les représentations ont été choisies de manière à ne pas faire apparaître d'effets de bords. En haut à gauche, la distribution de Wigner-Ville (5.2) $W_x(t, f)$ du signal. La fenêtre de lissage du spectrogramme $W_\phi(t, f)$, représentée en haut à droite, est choisie gaussienne. La convolution de $W_x(t, f)$ par $W_\phi(t, f)$ conduit au spectrogramme (5.1,5.4), au milieu à gauche. Au milieu à droite sont représentés deux points caractéristiques de l'effet du lissage : au point X , le noyau de lissage est appliqué sur des valeurs alternativement positives et négatives, conduisant à un point sur le spectrogramme de valeur nulle. Au contraire, sur le point $+$, le lissage va étaler l'énergie concentrée dans la distribution de Wigner-Ville, conduisant aux larges motifs du spectrogramme. La réallocation de ce spectrogramme (5.7) est donnée en bas à gauche. En refocalisant l'énergie étalée par le spectrogramme, la réallocation conjugue la bonne localisation temps-fréquence de Wigner-Ville à l'absence de termes d'interférences du spectrogramme. La représentation obtenue s'approche alors de la représentation idéale du signal, illustrée en bas à droite.

5.2 Réallocation du spectrogramme, interprétation de la fréquence instantanée et du retard de groupe

5.2.1 Fréquence instantanée et retard de groupe

Le point temps-fréquence (t', f') sur lequel l'énergie d'un point du spectrogramme (t, f) est déplacée est défini selon le barycentre local de la répartition de l'énergie du spectrogramme. Une autre interprétation est possible, basée sur la fréquence instantanée et le retard de groupe. En effet, en remplaçant la distribution de Wigner-Ville par son expression (5.2) dans l'équation (5.5) :

$$t' = \frac{1}{S_{\phi,x}(t, f)} \iiint \tau x\left(\tau + \frac{u}{2}\right) x^*\left(\tau - \frac{u}{2}\right) e^{-2\pi j \nu u} \phi\left(\tau - t + \frac{v}{2}\right) \phi^*\left(\tau - t - \frac{v}{2}\right) e^{-2\pi j(\nu-f)v} du dv d\tau \quad (5.8)$$

avec $\phi(t)$ la fenêtre d'analyse réelle. En remarquant que l'intégrale selon ν ne fait intervenir que l'expression $e^{-2\pi j(u+v)\nu}$, la distribution de Dirac en $(u+v)$, notée $\delta(u+v)$ apparaît. L'expression (5.8) devient :

$$t' = \frac{1}{S_{\phi,x}(t, f)} \iint \tau x\left(\tau + \frac{u}{2}\right) x^*\left(\tau - \frac{u}{2}\right) \phi\left(\tau - t - \frac{u}{2}\right) \phi^*\left(\tau - t + \frac{u}{2}\right) e^{-2\pi j f u} du d\tau \quad (5.9)$$

Avec un changement de variables tel que $u_1 = \tau + \frac{u}{2}$ et $u_2 = \tau - \frac{u}{2}$, on obtient :

$$\begin{aligned} t' &= \frac{1}{S_{\phi,x}(t, f)} \iint \frac{u_1 + u_2}{2} x(u_1) x^*(u_2) \phi(u_2 - t) \phi^*(u_1 - t) e^{-2\pi j f (u_1 - u_2)} du_1 du_2 \\ &= \frac{1}{S_{\phi,x}(t, f)} \frac{1}{2} \left(\int u_1 x(u_1) \phi^*(u_1 - t) e^{-2\pi j f u_1} du_1 \int x^*(u_2) \phi(u_2 - t) e^{2\pi j f u_2} du_2 \right. \\ &\quad \left. + \int x(u_1) \phi^*(u_1 - t) e^{-2\pi j f u_1} du_1 \int u_2 x^*(u_2) \phi^*(u_2 - t) e^{2\pi j f u_2} du_2 \right) \end{aligned} \quad (5.10)$$

On reconnaît alors la partie réelle d'un nombre complexe z exprimée sous la forme $\frac{1}{2}(z + z^*)$, conduisant à l'expression :

$$t' = \frac{1}{S_{\phi,x}(t, f)} \Re \left(X_{\phi}^*(t, f) \int u x(u) \phi^*(u - t) e^{-2\pi j f u} du \right) \quad (5.11)$$

avec $\Re(\cdot)$ la partie réelle.

En procédant de même pour la fréquence f' décrite par l'équation (5.6), on obtient en détaillant les expressions des distributions de Wigner-Ville :

$$f' = \frac{1}{S_{\phi,x}(t, f)} \iiint \nu x\left(\tau + \frac{u}{2}\right) x^*\left(\tau - \frac{u}{2}\right) e^{-2\pi j \nu u} \phi\left(\tau - t + \frac{v}{2}\right) \phi^*\left(\tau - t - \frac{v}{2}\right) e^{-2\pi j(\nu-f)v} du dv d\tau \quad (5.12)$$

Dans ce cas, l'intégration selon ν fait intervenir l'expression $\int \nu e^{-2\pi j(u+v)\nu} d\nu$. On reconnaît la TF (au sens des distributions) de ν , conduisant à l'expression $\frac{-1}{2j\pi} \delta'(u+v)$, avec $\delta'(t)$ la dérivée de la distribution de Dirac. L'expression (5.12) s'écrit alors :

$$\begin{aligned} f' &= \frac{-j}{2\pi S_{\phi,x}(t, f)} \iint x\left(\tau + \frac{u}{2}\right) x^*\left(\tau - \frac{u}{2}\right) \left(\frac{1}{2} \phi'\left(\tau - t - \frac{u}{2}\right) \phi^*\left(\tau - t + \frac{u}{2}\right) e^{-2\pi j f u} \right. \\ &\quad \left. - \frac{1}{2} \phi\left(\tau - t - \frac{u}{2}\right) \phi'^*\left(\tau - t + \frac{u}{2}\right) e^{-2\pi j f u} + \phi\left(\tau - t - \frac{u}{2}\right) \phi^*\left(\tau - t + \frac{u}{2}\right) j 2\pi f e^{-2\pi j f u} \right) du d\tau \end{aligned} \quad (5.13)$$

En effectuant le même changement de variables que précédemment, $u_1 = \tau + \frac{u}{2}$ et $u_2 = \tau - \frac{u}{2}$, on obtient :

$$\begin{aligned} f' &= \frac{-j}{2\pi S_{\phi,x}(t,f)} \left(j2\pi f \int x(u_1)\phi^*(u_1-t)e^{-2\pi j f u_1} du_1 \int x^*(u_2)\phi(u_2-t)e^{2\pi j f u_2} du_2 \right. \\ &\quad + \frac{1}{2} \int x(u_1)\phi^*(u_1-t)e^{-2\pi j f u_1} du_1 \int x^*(u_2)\phi'(u_2-t)e^{2\pi j f u_2} du_2 \\ &\quad \left. - \frac{1}{2} \int x(u_1)\phi^{*'}(u_1-t)e^{-2\pi j f u_1} du_1 \int x^*(u_2)\phi(u_2-t)e^{2\pi j f u_2} du_2 \right) \end{aligned} \quad (5.14)$$

Deux termes apparaissent, l'un sous la forme zz^* correspondant à la valeur absolue du nombre complexe z , l'autre sous la forme $\frac{1}{2j}(z - z^*)$, correspondant à la partie imaginaire de z . L'expression ainsi obtenue est :

$$f' = f - \frac{1}{2\pi S_{\phi,x}(t,f)} \Im \left(X_{\phi}^*(t,f) \int x(u)\phi^{*'}(u-t)e^{2\pi j f u} du \right) \quad (5.15)$$

avec $\Im(\cdot)$ la partie imaginaire.

Les expressions de t' (5.11) et f' (5.15) font apparaître des termes correspondant à la dérivée partielle selon f et t respectivement de la TFCT. t' et f' se réécrivent alors :

$$\begin{aligned} t' &= \frac{1}{X_{\phi}(t,f)X_{\phi}^*(t,f)} \Re \left(X_{\phi}^*(t,f) \frac{1}{-2\pi j} \frac{\partial}{\partial f} X_{\phi}(t,f) \right) \\ &= \frac{1}{2\pi} \Im \left(\frac{1}{X_{\phi}(t,f)} \frac{\partial}{\partial f} X_{\phi}(t,f) \right) \end{aligned} \quad (5.16)$$

$$\begin{aligned} f' &= f - \frac{1}{2\pi X_{\phi}(t,f)X_{\phi}^*(t,f)} \Im \left(-X_{\phi}^*(t,f) \frac{\partial}{\partial t} X_{\phi}(t,f) \right) \\ &= f + \frac{1}{2\pi} \Im \left(\frac{1}{X_{\phi}(t,f)} \frac{\partial}{\partial t} X_{\phi}(t,f) \right) \end{aligned} \quad (5.17)$$

En exprimant la TFCT par son module et son argument $X_{\phi}(t,f) = M(t,f)e^{j\Phi(t,f)}$, les dérivées partielles de la TFCT s'écrivent :

$$\begin{aligned} \frac{\partial}{\partial f} X_{\phi}(t,f) &= jM(t,f)e^{j\Phi(t,f)} \frac{\partial}{\partial f} \Phi(t,f) + e^{j\Phi(t,f)} \frac{\partial}{\partial f} M(t,f) \\ &= X_{\phi}(t,f) \left(j \frac{\partial}{\partial f} \Phi(t,f) + \frac{1}{M(t,f)} \frac{\partial}{\partial f} M(t,f) \right) \end{aligned} \quad (5.18)$$

$$\begin{aligned} \frac{\partial}{\partial t} X_{\phi}(t,f) &= jM(t,f)e^{j\Phi(t,f)} \frac{\partial}{\partial t} \Phi(t,f) + e^{j\Phi(t,f)} \frac{\partial}{\partial t} M(t,f) \\ &= X_{\phi}(t,f) \left(j \frac{\partial}{\partial t} \Phi(t,f) + \frac{1}{M(t,f)} \frac{\partial}{\partial t} M(t,f) \right) \end{aligned} \quad (5.19)$$

$$(5.20)$$

En injectant les équations (5.18) et (5.19) dans les équations (5.16) et (5.17) respectivement, on obtient finalement :

$$t' = \frac{1}{2\pi} \frac{\partial}{\partial f} \Phi(t,f) \quad (5.21)$$

$$f' = f + \frac{1}{2\pi} \frac{\partial}{\partial t} \Phi(t,f) \quad (5.22)$$

Ces expressions sont intéressantes pour deux raisons. Elles montrent que la réallocation du spectrogramme se fait en intégrant une information liée à la phase de la TFCT, qui est ignorée lors du calcul du spectrogramme.

De plus, les expressions (5.21,5.22) montrent que l'énergie du spectrogramme au point (t, f) est déplacée au point (t_g, f_i) correspondant au retard de groupe t_g et à la fréquence instantanée f_i du signal. De manière duale, on peut considérer que l'énergie d'un point (t, f) est déplacée sur un point qui est censé mieux la représenter, tandis que chaque point temps-fréquence regroupe les portions d'énergies de même retard de groupe et de même fréquence instantanée.

5.2.2 Mise en œuvre de la réallocation

Ainsi, le paragraphe précédent a mis en évidence deux points de vue existants pour interpréter la réallocation, le premier basé sur la notion de centre de gravité local de la répartition de l'énergie dans le plan temps-fréquence, le second basé sur la notion de fréquence instantanée et de retard de groupe du signal. Toutefois, ces deux cas posent des problèmes pratiques de calcul : le calcul d'un barycentre est très coûteux, tandis que dériver une phase cause des problèmes de stabilité, notamment liés au déroulement de la phase.

Une mise en œuvre plus efficace est réalisable en remarquant dans l'équation de t' (5.11) l'expression d'une TFCT calculée avec une fenêtre d'analyse égale à $t\phi(t)$, notée $X_{t\phi}(t, f)$. Ce qui nous conduit à :

$$t' = \frac{1}{S_{\phi,x}(t, f)} \Re (X_{\phi}^*(t, f)X_{t\phi}(t, f) + tX_{\phi}(t, f)) \quad (5.23)$$

De même, la TFCT calculée avec la dérivée de la fenêtre $\phi(t)$ se retrouve dans l'équation de f' (5.15) :

$$f' = f - \frac{1}{2\pi S_{\phi,x}(t, f)} \Im (X_{\phi}^*(t, f)X_{(d\phi(t,f)/dt)}) \quad (5.24)$$

Le nouveau point (t', f') auquel l'énergie du point (t, f) est déplacée s'exprime à l'aide de quotients de TFCT :

$$t' = t + \Re \left(\frac{X_{t\phi}(t, f)}{X_{\phi}(t, f)} \right) \quad (5.25)$$

$$f' = f - \frac{1}{2\pi} \Im \left(\frac{X_{(d\phi(t,f)/dt)}}{X_{\phi}(t, f)} \right) \quad (5.26)$$

Ainsi, le calcul de la réallocation du spectrogramme s'effectue en calculant trois TFCT, avec pour fenêtres d'analyse : la fenêtre $\phi(t)$, paramètre du spectrogramme à réallouer, la fenêtre $t\phi(t)$ et la fenêtre dérivée $d\phi()/dt$, puis en les combinant selon les équations (5.25,5.26). La réallocation proprement dite peut alors se faire à partir des coefficients du spectrogramme obtenus par le module carré de la première TFCT, déplacés selon l'équation (5.7).

5.3 Champ de vecteurs de réallocation

Dans les chapitres suivants, nous allons nous intéresser plus particulièrement à la manière dont l'énergie du spectrogramme est déplacée dans le plan temps-fréquence, autrement dit nous allons nous intéresser au champ de vecteurs de réallocation, noté $\mathbf{r}_x(t, f)$, défini par :

$$\mathbf{r}_x(t, f) = (t'(t, f) - t, f'(t, f) - f)^t \quad (5.27)$$

Ce champ de vecteurs représente l'ensemble des vecteurs de déplacement de l'énergie des points (t, f) aux nouveaux points (t', f') .

D'après les équations précédentes, il se calcule aisément à l'aide des quotients de TFCT (5.25,5.26) :

$$\mathbf{r}_x(t, f) = \left(\Re \left(\frac{X_{t\phi}(t, f)}{X_\phi(t, f)} \right), \frac{-1}{2\pi} \Im \left(\frac{X_{(d\phi(t,f)/dt)}(t, f)}{X_\phi(t, f)} \right) \right)^t \quad (5.28)$$

Ce champ de vecteurs de réallocation dépend grandement de la fenêtre d'analyse $\phi(t)$. Ainsi, une fenêtre $\phi(t)$ choisie telle que sa distribution de Wigner-Ville $W_\phi(t, f)$ soit très étendue en temps et réduite en fréquence tendra à créer des vecteurs de réallocation dirigés suivant l'axe temporel. Pour s'affranchir de cette influence, il peut être utile de s'intéresser au champ de vecteurs de réallocation normalisé, défini par :

$$\mathbf{R}_x(t, f) = \left(\frac{t' - t}{\Delta t}, \frac{f' - f}{\Delta f} \right)^t \quad (5.29)$$

avec Δt et Δf les largeurs temporelle et fréquentielle de la fenêtre de pondération, définis par :

$$\Delta t = \left(\int t^2 W_\phi(t, f) dt df \right)^{\frac{1}{2}} \quad (5.30)$$

$$\Delta f = \left(\int f^2 W_\phi(t, f) dt df \right)^{\frac{1}{2}} \quad (5.31)$$

La figure 5.2 illustre ce champ de vecteurs pour deux signaux différents, avec des fenêtres d'analyse gaussiennes. Le premier est une modulation linéaire, d'équation $x(t) = e^{i*2*\pi*t^2/2}$. La réallocation déplace l'énergie du spectrogramme sur la droite $t = f$ [CM98], chaque vecteur de réallocation pointant alors exactement sur un point de cette droite.

L'autre signal d'illustration est un logon gaussien, d'équation $x(t) = \pi^{-\frac{1}{4}} e^{-\frac{t^2}{2}}$. La réallocation fait subir à son spectrogramme une homothétie de rapport $\frac{1}{2}$, centrée sur le centre du logon [CM98]. Dans ce cas, les vecteurs de réallocation ne pointent pas tous vers un même lieu, mais restent dirigés vers le centre du logon.

D'après la formulation du spectrogramme donnée par l'équation (5.4), le spectrogramme va étaler une énergie issue d'une représentation de Wigner-Ville. D'après les équations (5.5,5.6), l'énergie d'un point (t, f) va être déplacée sur le barycentre local de l'énergie. Ainsi, sur les bords d'un motif temps-fréquence, l'énergie issue de la distribution de Wigner-Ville étalée par le spectrogramme va être déplacée vers le centre de gravité local de l'énergie, autrement dit vers l'intérieur du motif. Ainsi, les vecteurs de réallocation situés aux bords du motifs pointeront vers le motif.

Le chapitre suivant utilise cette idée afin de proposer une détection des zones temps-fréquence susceptibles d'appartenir aux bords d'un motif temps-fréquence, pour ensuite reconstruire le contour à partir de cette première détection.

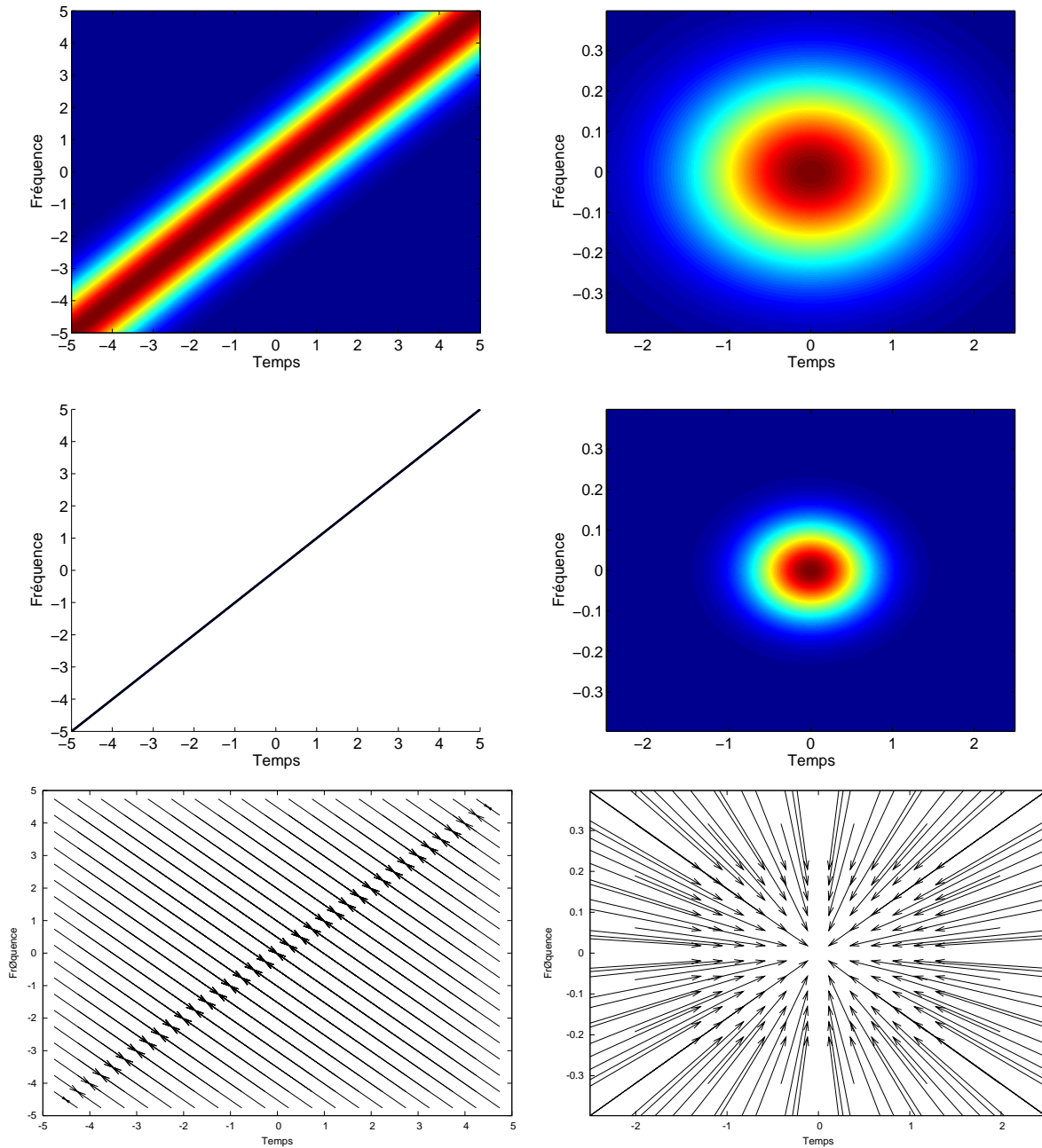


FIG. 5.2 : Illustrations de deux champs de vecteurs de réallocation : en haut, spectrogrammes des signaux calculés avec une fenêtre gaussienne, une modulation linéaire de fréquence d'équation $e^{i*2*\pi t^2/2}$ à gauche, un logon gaussien d'équation $\pi^{-\frac{1}{4}} e^{-\frac{t^2}{2}}$ à droite. Au milieu, les réallocations des spectrogrammes. En bas, les champs de vecteurs associés aux réallocations. Tous les vecteurs de réallocation sont dirigés vers l'intérieur des motifs spectraux.

Chapitre 6

Détection de contour

Dans ces travaux, nous n'utilisons pas la réallocation en tant que représentation temps-fréquence, mais uniquement comme source d'information sur la structure du signal. L'idée est ici de définir comme *frontière* d'un motif temps-fréquence l'ensemble des points temps-fréquence contenant de l'énergie du spectrogramme, mais ne contenant pas d'énergie de la distribution de Wigner-Ville. Autrement dit, la zone frontière d'un motif est celle où l'énergie issue de la distribution de Wigner-Ville a été étalée par la convolution avec la fenêtre d'analyse par le spectrogramme (5.4). La figure 6.1 montre cette zone frontière pour une modulation linéaire de fréquence fenêtrée dans le temps.

Sur cette zone frontière, tous les vecteurs de réallocation vont pointer vers l'intérieur du motif temps-fréquence. On suppose de plus que les variations du contour sont suffisamment douces pour considérer que les vecteurs de réallocation sont localement parallèles sur cette zone frontière. Détecter cette zone revient alors à détecter les zones temps-fréquence où les vecteurs de réallocation sont parallèles.

La construction du contour se fait ainsi en deux étapes : tout d'abord, nous proposons une méthode pour détecter les points susceptibles d'appartenir à cette zone frontière, dont l'utilité de cette méthode simple sera illustrée en tant que complément à la méthode de segmentation présentée dans la première partie de cette thèse. Ensuite, nous proposerons une construction du contour d'un motif temps-fréquence à l'aide d'un algorithme de détection de crête. L'influence des paramètres du spectrogramme et du rapport signal à bruit sur la détermination de ce contour sera discutée.

Pour illustrer ces méthodes, deux signaux synthétiques analytiques seront utilisés. Le premier, utilisé dans la figure 6.1, est une modulation linéaire de fréquence limitée en temps. Le second est un bruit blanc, fenêtré en temps et en fréquence afin de la rendre non blanc et non stationnaire, dont le spectrogramme est montré sur la figure 6.2 à gauche.

La réallocation est connue pour être particulièrement efficace sur les modulations linéaires de fréquence. Toutefois, elle n'est que peu utile sur les signaux large bande : en effet, la réallocation a pour effet de déplacer l'énergie issue du spectrogramme sur les barycentres locaux de l'énergie. Dans le cas d'un signal issu d'une réalisation d'un bruit blanc, son spectrogramme va présenter des fluctuations, chaque valeur maximale locale servant alors de point de focalisation pour l'énergie déplacée par la réallocation. Il en résulte en une représentation nervurée de l'énergie dans le plan temps-fréquence (6.2, figure de droite).

L'approche proposée, en s'appliquant à la fois aux modulations de fréquence et aux signaux large bande, permet ainsi d'étendre indirectement le champ d'application de la réallocation.

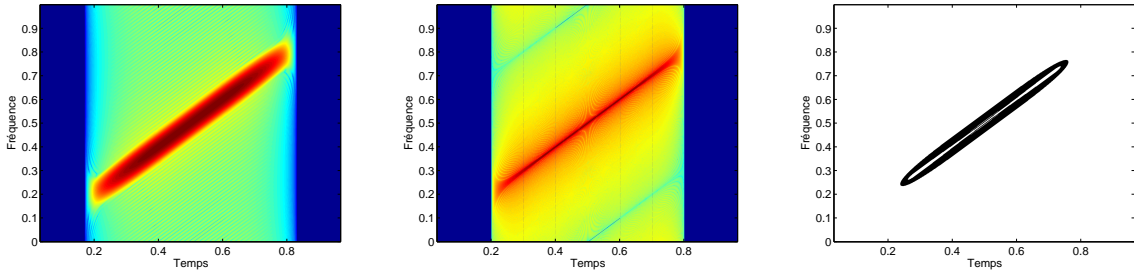


FIG. 6.1 : Illustration de la zone frontière d'un motif temps-fréquence. À gauche, spectrogramme d'une modulation linéaire de fréquence, fenêtrée dans le temps. Au centre, la distribution de Wigner-Ville associée. À droite, la zone frontière du motif, définie comme la différence de support entre le motif tiré du spectrogramme et le motif issu de Wigner-Ville. Les supports temps-fréquence ont été seuillés à 1% de l'amplitude maximale.

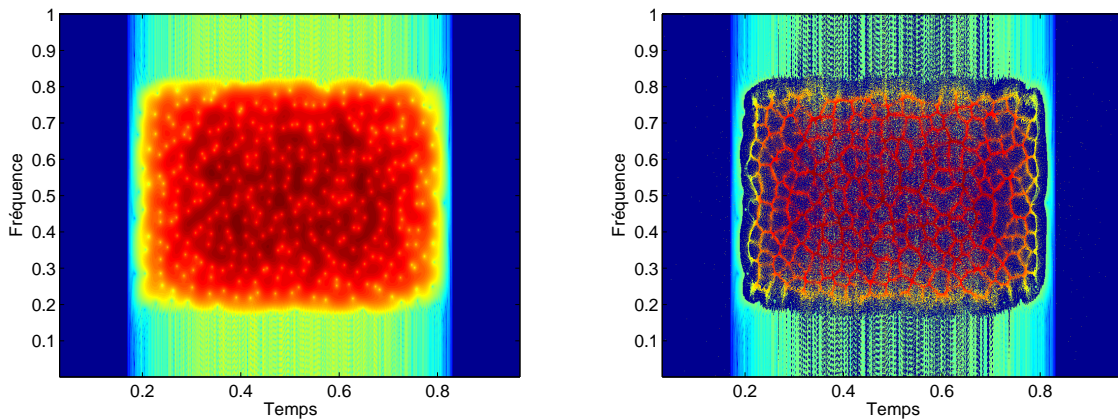


FIG. 6.2 : Illustration du signal large bande synthétique utilisé dans ce chapitre. À gauche, son spectrogramme, à droite son spectrogramme réalloué. L'énergie du spectrogramme est déplacée sur les maxima locaux du spectrogramme, autrement dit sur les fluctuations liées à la réalisation du signal aléatoire.

6.1 Discrétisation du champ de vecteurs de réallocation

Une mise en œuvre de la construction du contour passe par des moyens informatiques, et impose par conséquent d'utiliser des grandeurs discrètes. L'hypothèse de parallélisme local des vecteurs de réallocation sur la zone frontière se révèle alors fautive, comme l'illustre la figure 6.3. Deux points voisins dans le plan temps-fréquence peuvent cibler deux points différents et avoir alors deux angles différents. Pour pallier ce problème, nous considérerons le parallélisme des vecteurs à un angle θ_0 près, autrement dit deux vecteurs seront considérés parallèles si la différence entre leurs angles est inférieure à θ_0 .

Cet angle permet d'une part d'éviter le problème lié à la discrétisation du plan temps-fréquence, d'autre part, il laisse une flexibilité sur la contrainte du parallélisme recherché, et

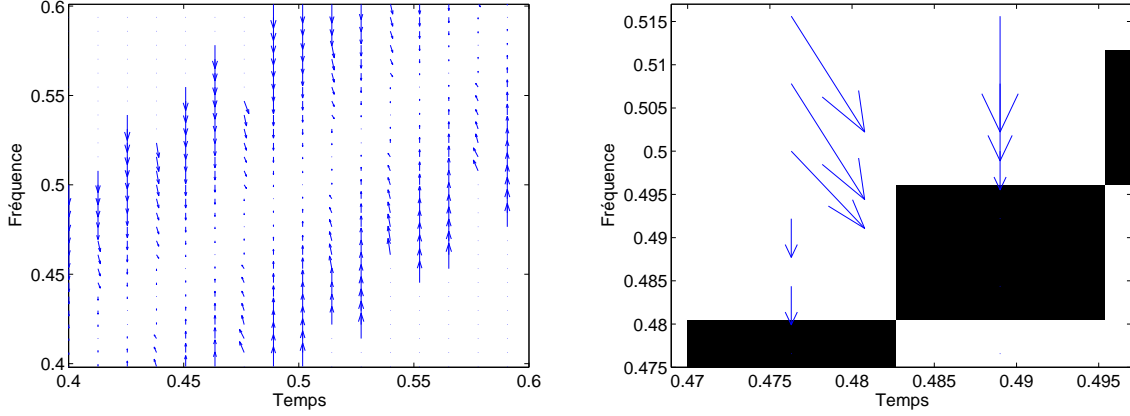


FIG. 6.3 : Illustration du problème de la discrétisation du champ de vecteurs de réallocation. À gauche, un premier zoom du champ de vecteurs de réallocation calculé en pratique pour la modulation linéaire de fréquence de la figure 5.2. Pour une meilleure lisibilité, la norme des vecteurs a été réduite. À droite, un second zoom de ce champ de vecteurs. Les pixels noirs correspondent aux points temps-fréquence où l'énergie du spectrogramme est réallouée. Deux vecteurs voisins dans le plan temps-fréquence peuvent avoir des angles différents, selon leur position et celle du point qu'ils visent.

permet alors de suivre les variations du contour. Le choix de θ_0 est fixé à :

$$\theta_0 = \frac{\pi}{4} \quad (6.1)$$

Cet angle correspond à la borne supérieure de la différence entre les angles de deux vecteurs de réallocation voisins sur une zone frontière, et constitue une valeur suffisamment grande pour pouvoir suivre de fortes variations de contour.

Dans le cas continu, un vecteur de réallocation s'écrit $\mathbf{r}(t, f) = (t'(t, f) - t, f'(t, f) - f)^t$ (5.27), ses dimensions sont ainsi homogènes à un temps et à une fréquence respectivement. Un choix est à faire sur le champ de vecteurs de réallocation discrétisé : doit-on garder des grandeurs physiques ou ne considérer que des déplacement de pixels ?

Comme illustrée par la figure 6.3, la discrétisation pose problème quant au parallélisme local des vecteurs. En considérant des grandeurs physiques, dépendant de la fréquence d'échantillonnage du signal, un décalage d'un point discret peut correspondre à des millihertz en fréquence et à des heures en temps : le problème peut ainsi grandement s'amplifier. Le choix d'un champ de vecteurs de réallocation sans lien physique, uniquement en tant que déplacement de points dans un espace discret est donc à privilégier.

Pour cette raison, la normalisation du champ de vecteurs (5.29) est inutile dans notre cas.

6.2 Détection de zone frontière

Un algorithme très simple de détection des zone frontière a déjà été proposé ([MM06],[MM07a]). Il consiste à parcourir tous les points (t, f) du plan temps-fréquence afin d'évaluer le parallélisme

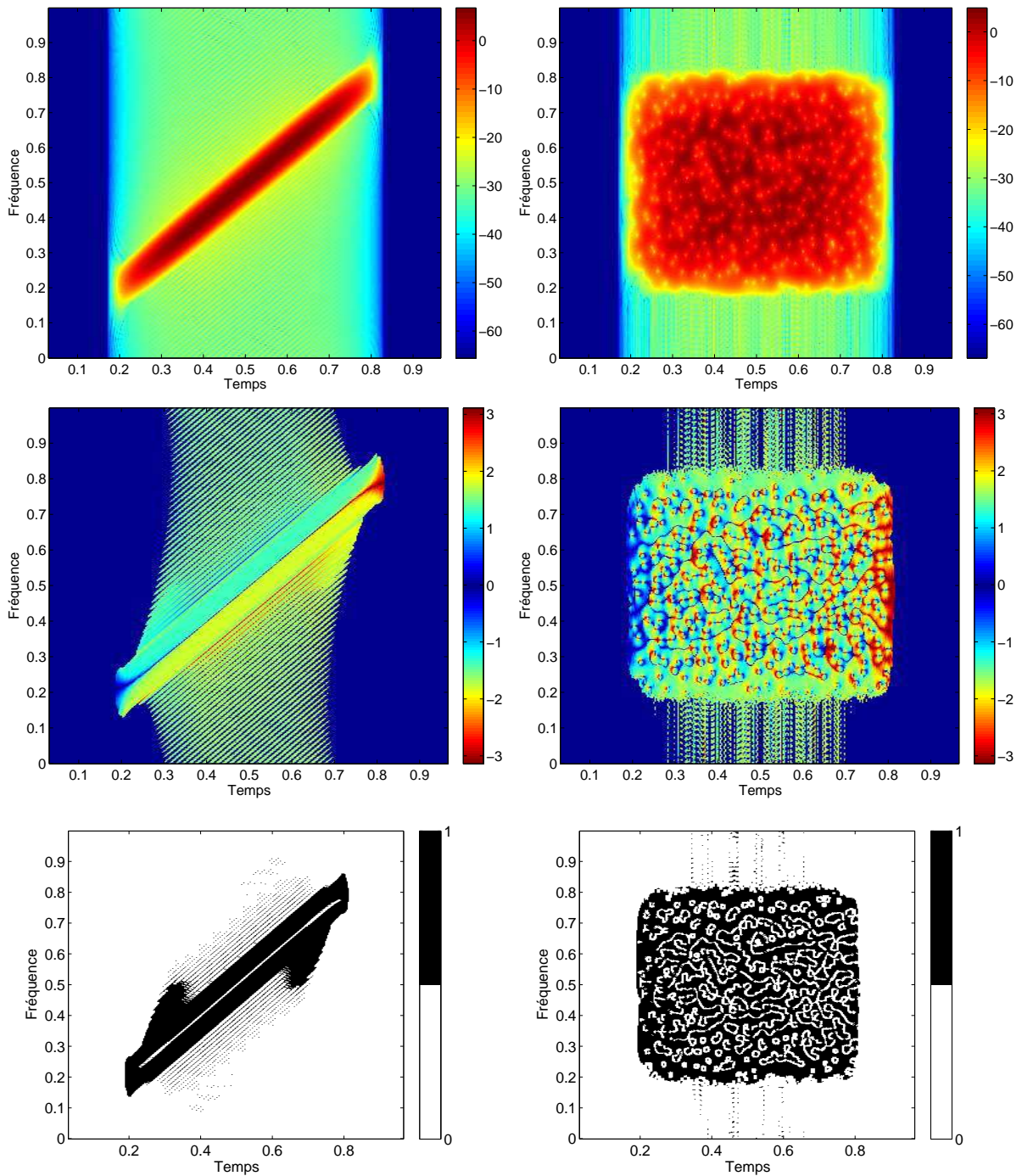


FIG. 6.4 : Illustration de la détection de zone frontière : en haut, les spectrogrammes de la modulation linéaire de fréquence et du signal large bande, calculés pour une fenêtre gaussienne de 63 points, un recouvrement de 60 points et 512 fréquences calculées. Au milieu, les angles des vecteurs de réallocation. L'échelle est telle que π et $-\pi$ sont de la même couleur. En bas, la détection de zone frontière obtenue pour un voisinage $l_1 \times l_2 = 3 \times 3$.

des vecteurs de réallocation sur un voisinage donné, renvoyant un indicateur $\text{ind}(t, f)$ défini par :

$$\text{ind}(t, f) = \begin{cases} 1 & \text{si } |\text{angle}(\mathbf{r}(t, f)) - \text{angle}(\mathbf{r}(\tau, \nu))| \leq \theta_0, \quad \forall (\tau, \nu) \in N_{t,f} \\ 0 & \text{sinon.} \end{cases} \quad (6.2)$$

où $\text{angle}(\mathbf{r}(t, f))$ est l'angle du vecteur de réallocation \mathbf{r} et $N_{t,f}$ un voisinage rectangulaire de taille de $l_1 \times l_2$.

La figure 6.4 illustre le résultat de cet algorithme sur les deux signaux de test, pour un spectrogramme calculé avec une fenêtre gaussienne de 63 points, un recouvrement de 60 points et 512 fréquences calculées. Le voisinage est fixé arbitrairement à $l_1 \times l_2 = 3 \times 3$. Les vecteurs de réallocation sur les bords des motifs sont bien dirigés vers l'intérieur des motifs. L'algorithme détecte correctement toute la zone frontière, mais génère beaucoup de fausses alarmes, notamment liées aux parallélismes locaux des vecteurs de réallocation autour de maxima ponctuels de l'énergie du spectrogramme.

6.2.1 Application sur un signal vibratoire

Cet algorithme, malgré ses nombreuses fausses alarmes, est déjà utile en tant que complément d'information à une autre segmentation, telle que présentée dans la première partie de ces travaux. Dans ce cas, nous cherchons à surveiller le moteur du banc de test GOTIX¹ [MM07b]. Le signal, enregistré par des accéléromètres sur l'axe du moteur, est composé d'harmoniques, dont un zoom du spectrogramme est présenté en haut à gauche de la figure 6.5. Le but est de segmenter l'harmonique principale, sans les harmoniques secondaires. Dans le cadre d'une segmentation aveugle, on ne sait pas combien d'harmoniques ont été fusionnées en un seul motif spectral.

Une mauvaise segmentation, obtenue avec un mauvais choix de paramètres et englobant plusieurs harmoniques, est montrée en haut à droite de la figure 6.5. Grâce à l'utilisation des vecteurs de réallocation, on détecte la présence d'une zone frontière à l'intérieure du motif segmenté (figure 6.5, en bas), indiquant ainsi que la segmentation ne comprend pas qu'un seul motif.

Dans le cas d'une bonne segmentation, ne comprenant qu'un motif, la zone frontière est correctement détectée au bord du motif uniquement.

Afin d'améliorer la détection de zone frontière et supprimer les nombreuses fausses alarmes, nous cherchons à construire un contour au motif, autrement dit à déterminer une ligne entourant le motif.

6.3 Construction du contour

6.3.1 Degré de parallélisme local

Le détecteur de frontière $\text{ind}(t, f)$ (6.2) indique de manière binaire si tous les vecteurs sont parallèles (à un angle θ_0 près) sur une cellule $l_1 \times l_2$. Plutôt que de considérer cette information de manière binaire, on s'intéresse au *degré de parallélisme* sur la cellule $l_1 \times l_2$, que l'on notera $\delta_{\parallel}(t, f)$. Nous définissons ce degré de parallélisme $\delta_{\parallel}(t, f)$ comme le nombre de vecteurs de

¹Pour plus de détail, voir <http://www.gipsa-lab.inpg.fr/gotix>

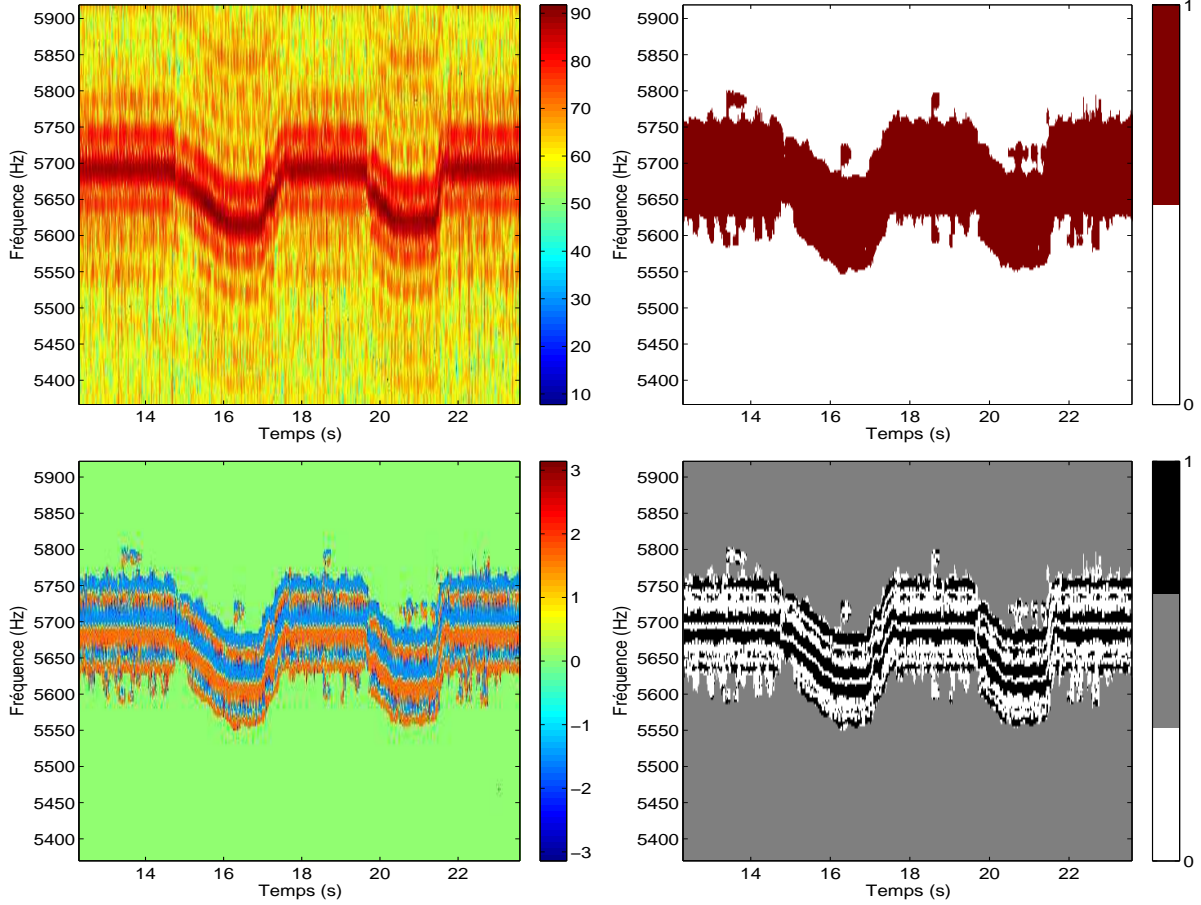


FIG. 6.5 : Illustration de l'intérêt des vecteurs de réallocation comme apport d'information complémentaire à la segmentation du spectrogramme. En haut à gauche, spectrogramme d'un signal vibratoire d'une machine tournante. En haut à droite, un cas de mauvaise segmentation : plusieurs harmoniques sont segmentées ensemble. En bas à gauche, l'angle des vecteurs de réallocation segmentés. En bas à droite, le résultat de la détection de zone frontière appliquée sur les points segmentés. La zone grise correspond aux points non segmentés. Une zone frontière est détectée à l'intérieur du motif, distincte des bords : il y a plus d'une composante segmentée.

réallocation $\mathbf{r}(\tau, \nu)$ parallèles au vecteur $\mathbf{r}(t, f)$ sur le voisinage $l_1 \times l_2$.

$$\delta_{\parallel}(t, f) = \sum_{(\tau, \nu)} \|\left(\mathbf{r}(\tau, \nu), \mathbf{r}(t, f)\right), (\tau, \nu) \in N_{t, f} \quad (6.3)$$

avec $\|(\mathbf{a}, \mathbf{b}) = \begin{cases} 1 & \text{si } |\text{angle}(\mathbf{a}) - \text{angle}(\mathbf{b})| \leq \theta_0, \\ 0 & \text{sinon.} \end{cases}$

Le voisinage $l_1 \times l_2 = 3 \times 3$ utilisé pour la détection de zone frontière de la figure 6.4 ne peut convenir dans le cas du degré de parallélisme δ_{\parallel} . En effet, de nombreux vecteurs de réallocation sont parallèles à tous leurs huit voisins, donnant un indicateur $\text{ind}(t, f)$ égal à 1, correspondant à un degré de parallélisme maximal. Les fausses alarmes à l'intérieur du motif large bande ne sont pas différenciées des points appartenant réellement au contour.

Par conséquent, il faut un voisinage plus grand. La définition retenue pour la zone frontière

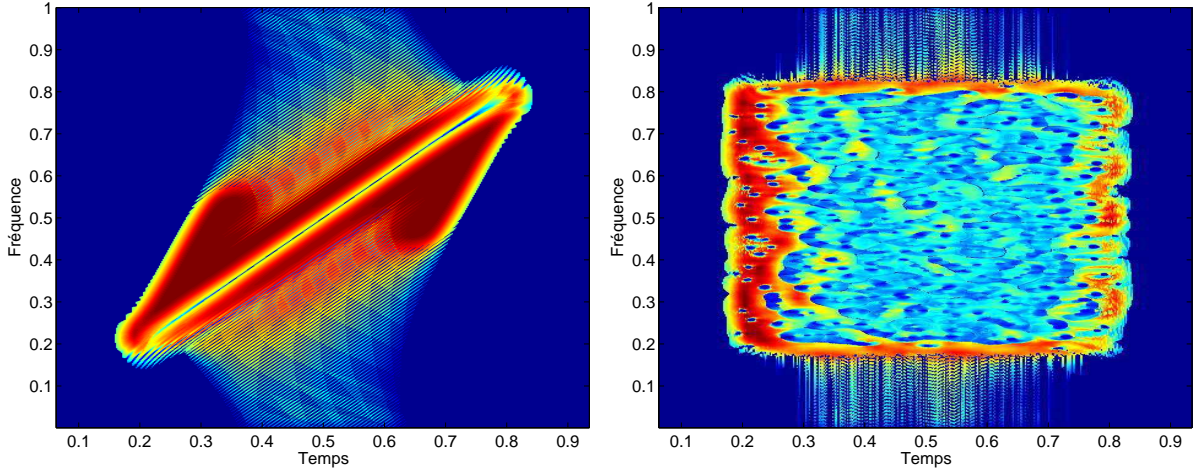


FIG. 6.6 : Illustration du degré de parallélisme δ_{\parallel} sur les deux signaux de test. Les spectrogrammes sont calculés avec une fenêtre gaussienne de 127 points, un recouvrement de 125 points, et 1024 fréquences calculées. Le voisinage $l_1 \times l_2$ est de 31×53 , correspondant au support du noyau de lissage $W_{\phi}(t, f)$. Le degré de parallélisme est localement maximal aux frontières des motifs temps-fréquence

d'un motif temps-fréquence est la zone où l'énergie de la distribution de Wigner-Ville est étalée par le noyau de lissage créant le spectrogramme 5.4. Nous pouvons considérer comme voisinage $N_{(t,f)}$ le support temps-fréquence du noyau de lissage $W_{\phi}(t, f)$, autrement dit l'ensemble des points tels que $W_{\phi}(t, f)$ est positif. Toutefois, ce support est infini, que ce soit en temps ou en fréquence : nous prendrons comme support temps-fréquence du noyau uniquement les points temps-fréquence contenant au moins 10% de la valeur maximale du noyau, en supposant le reste négligeable.

Ainsi, le voisinage retenu $N(t, f)$ s'écrit :

$$N(t, f) = \{ (t, f) \mid W_{\phi}(t, f) > \frac{1}{10} \max(W_{\phi}(t, f)) \} \quad (6.4)$$

Pour des raisons pratiques, la forme de $N(t, f)$ est choisie rectangulaire $l_1 \times l_2$. $N(t, f)$ est la plus petite cellule rectangulaire englobant le support temps-fréquence du noyau (6.4).

La figure 6.6 montre les résultats de degré de parallélisme δ_{\parallel} (6.3) de la modulation linéaire de fréquence et du signal large bande définis en 6, pour des spectrogrammes calculés avec une fenêtre gaussienne de 127 points, un recouvrement de 125 points, et 1024 fréquences calculées. Le parallélisme entre deux vecteurs est toujours considéré à une valeur θ_0 près, le voisinage choisi correspond au support temps-fréquence du noyau de lissage (6.4). Les plus grandes valeurs de δ_{\parallel} se situent effectivement sur les bords des motifs temps-fréquence. De nombreuses fausses alarmes dans la détection de zone frontière (figure 6.4) ont un degré de parallélisme bien plus faible que les points situés aux frontières, ce qui permet désormais de les discriminer.

6.3.2 Algorithme de construction de contour

L'idée présentée ici est de s'appuyer sur le degré de parallélisme $\delta_{\parallel}(t, f)$ pour construire un contour fermé entourant le motif temps-fréquence. L'algorithme retenu consiste à parcourir le

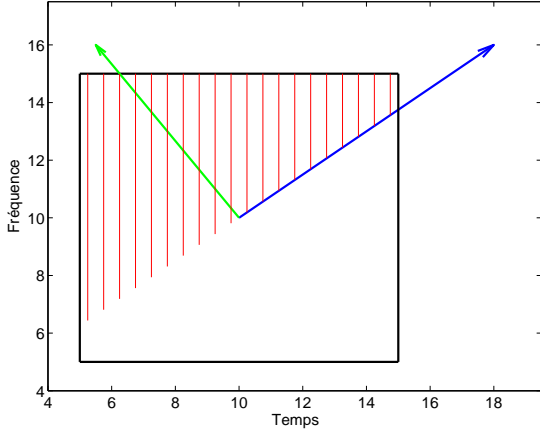


FIG. 6.7 : Illustration du demi-voisinage $\mathcal{V}(t, f)$ correspondant au vecteur de réallocation (en bleu) $\mathbf{r}(t, f)$, sur lequel le point suivant du contour est recherché. Ce voisinage, hachuré en rouge, correspond à la moitié du support du noyau de lissage $W_\phi(t, f)$ (indiqué par le carré noir) située à gauche du vecteur de réallocation. En vert, le vecteur perpendiculaire servant à calculer le demi-voisinage (6.5).

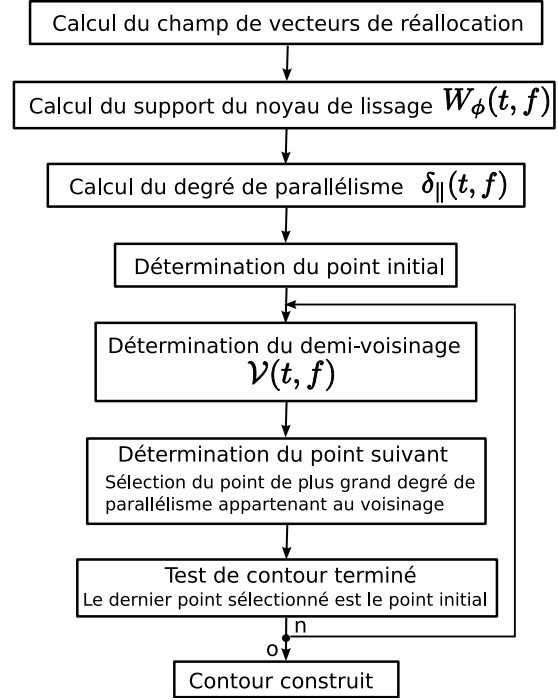


FIG. 6.8 : Algorithme de construction de contour d'un motif temps-fréquence.

plan temps-fréquence en suivant les maxima locaux du degré de parallélisme $\delta_{\parallel}(t, f)$, semblable à un algorithme de suivi de ligne de crête.

On suppose dans un premier temps connu un point du contour, noté $C(l) = (t_l, f_l)$. Le vecteur de réallocation associé à ce point $\mathbf{r}(t_l, f_l)$ pointe vers l'intérieur du motif temps-fréquence. Le point suivant $C_{l+1} = (t_{l+1}, f_{l+1})$ est déterminé selon deux critères :

- il appartient à la zone frontière du motif, c'est-à-dire que le degré de parallélisme associé $\delta_{\parallel}(t_{l+1}, f_{l+1})$ est grand ;
- il est voisin du point précédent, afin de garder une continuité dans la construction du contour.

Le voisinage retenu ici est le support temps-fréquence du noyau de lissage $N(t, f)$ (6.4), correspondant à la largeur temps-fréquence de la zone frontière.

Il reste à déterminer un sens de parcours : de manière arbitraire, le contour sera créé à partir de la gauche du point initial, par rapport à la direction pointée par le vecteur de réallocation associé. Afin d'assurer que les points sélectionnés pour construire le contour vont dans la bonne direction par rapport au contour déjà créé, le voisinage sur lequel le point suivant est choisi, noté $\mathcal{V}(t, f)$, est limité aux points à gauche du vecteur de réallocation courant. Son expression mathématique est :

$$\mathcal{V}(t, f) = (\tau, \nu) \in N(t, f) / (\tau - t, \nu - f)^t \cdot (-(f' - f), (t' - t))^t \geq 0 \quad (6.5)$$

avec \cdot dénotant le produit scalaire, et $(-(f' - f), (t' - t))^t$ le vecteur perpendiculaire au vecteur $\mathbf{r}(t, f) = (t' - t, f' - f)^t$, pointant à sa gauche.

La figure 6.7 montre un exemple de ce demi-voisinage. En bleu, le vecteur de réallocation associé au point (t, f) , en vert le vecteur perpendiculaire servant à la détermination du demi-

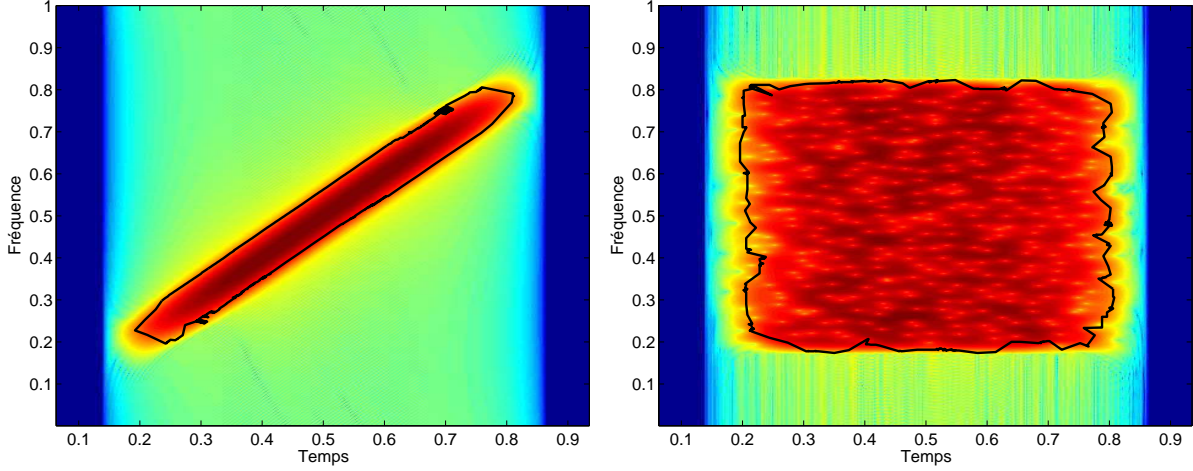


FIG. 6.9 : Illustration du contour reconstruit

voisinage $\mathcal{V}(t, f)$. En noir, la cellule $N(t, f) = l_1 \times l_2 = 11 \times 11$. La zone hachurée en rouge correspond au demi-voisinage $\mathcal{V}(t, f)$, l'ensemble des points de la cellule $N(t, f)$, situés à gauche du vecteur de réallocation.

Une fois ce demi-voisinage $\mathcal{V}(t, f)$ obtenu, il ne reste qu'à sélectionner un point pour continuer le contour. Le critère principal étant un grand degré de parallélisme $\delta_{\parallel}(t, f)$, le point suivant C_{l+1} choisi est celui ayant le plus grand degré de parallélisme sur le demi-voisinage :

$$C_{l+1} = (t_{l+1}, f_{l+1}) / \delta_{\parallel}(t_{l+1}, f_{l+1}) = \max(\delta_{\parallel}(\tau, \nu)), \forall (\tau, \nu) \in \mathcal{V}(t, f) \quad (6.6)$$

Enfin, il reste à déterminer un point initial C_0 . Pour la même raison que précédemment, il est choisi comme celui ayant le plus grand degré de parallélisme, cette fois-ci en considérant l'ensemble du plan temps-fréquence, et s'écrit :

$$C_0 = (t_0, f_0) / \delta_{\parallel}(t_0, f_0) = \max(\delta_{\parallel}(\tau, \nu)), \forall (\tau, \nu) \in \mathcal{R}^2 \quad (6.7)$$

Le schéma de l'algorithme est résumé sur la figure 6.8.

6.4 Cas de signaux bruités

Contrairement à la première partie de ces travaux de thèse, nous ne nous intéressons pas ici aux statistiques des coefficients temps-fréquence d'un signal bruité, ni aux statistiques des vecteurs de réallocation associés. Le lecteur curieux pourra toutefois se référer aux travaux de Chassande-Mottin [CM98, CMFA98], donnant la densité de probabilité du vecteur de réallocation dans le cas d'un bruit seul et d'un mélange « signal + bruit ».

Dans le cadre de la construction de contour, le bruit additif n'est intéressant que dans la mesure où il perturbe les vecteurs de réallocation présents sur la zone frontière des motifs temps-fréquence.

Pour des niveaux de bruit faibles, la construction du contour se déroule de manière similaire au cas sans bruit, comme l'illustre la figure 6.10. Les contours sont ceux des signaux de test auxquels est ajouté un bruit blanc gaussien de variance 10^{-2} .

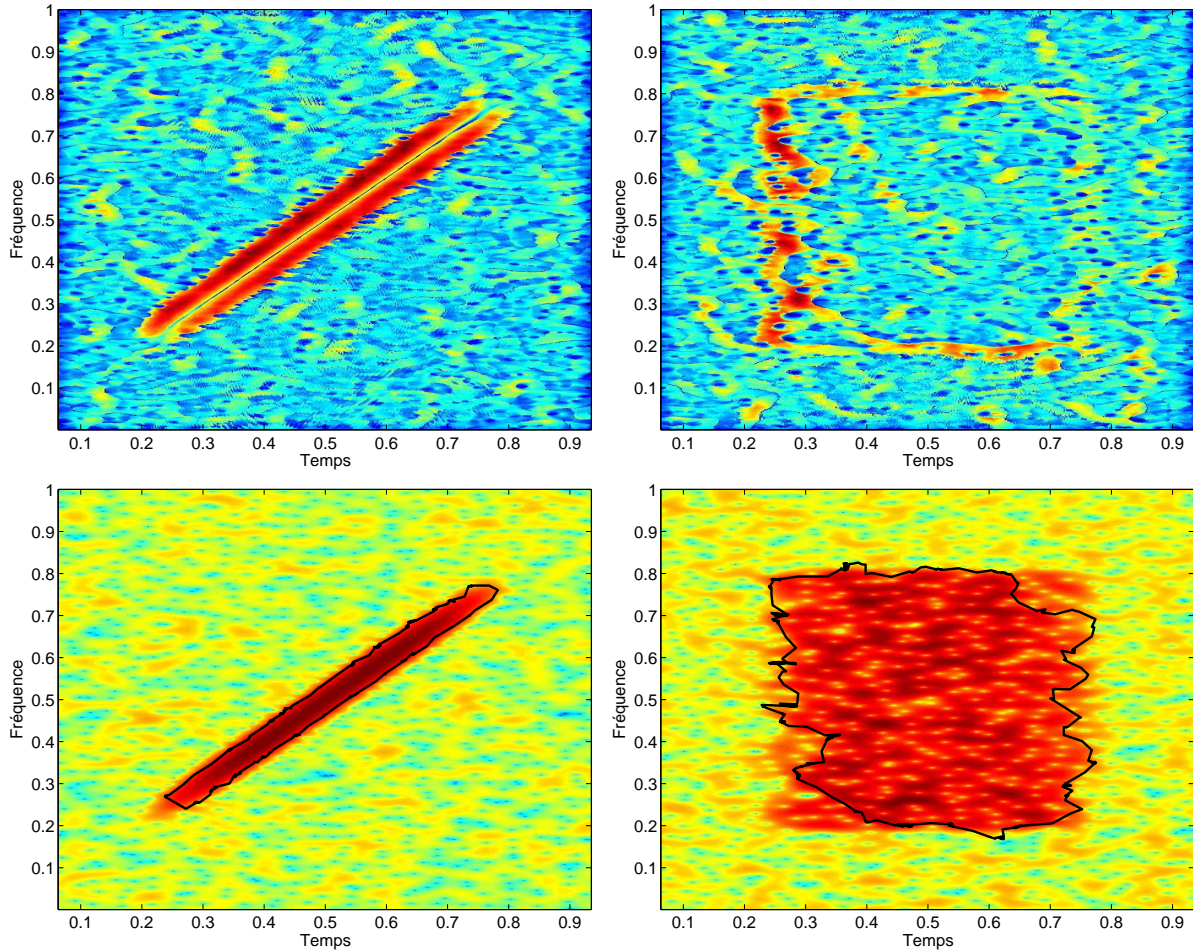


FIG. 6.10 : En haut, degré de parallélisme δ_{\parallel} et en bas, contours reconstruits, pour les deux signaux de test auxquels est ajouté un bruit blanc gaussien de variance 10^{-2}

Toutefois, les zones temps-fréquence ne contenant que du bruit présentent des valeurs de degré de parallélisme $\delta_{\parallel}(t, f)$ variables, qui peuvent localement être supérieures au degré de parallélisme des points appartenant à la zone frontière du motif. Ainsi, lorsque le niveau de bruit devient trop élevé, la poursuite du contour peut sélectionner des points de bruit comme appartenant au contour plutôt que des points appartenant à la zone frontière du motif. L'algorithme de création de contour quitte alors la zone frontière du motif, et ne peut plus construire correctement le contour du motif. Cet effet est illustré par la figure 6.11. Sur ces exemples, l'algorithme construit correctement le début du contour, mais poursuit des fluctuations du degré de parallélisme liées au bruit uniquement.

Dans le spectrogramme sans bruit additif, les points temps-fréquence sur la zone frontière d'un motif contiennent moins d'énergie à mesure qu'ils s'éloignent du motif, selon la fenêtre de lissage utilisée. Ainsi, ils sont plus susceptibles d'être perturbés lorsque le bruit blanc est ajouté au signal. Les vecteurs de réallocation associés à ces points ont alors un comportement correspondant aux points de bruit, et ne présentent plus le parallélisme attendu sur la zone frontière. Par conséquent, cette zone frontière se réduit : il en va de même pour le contour

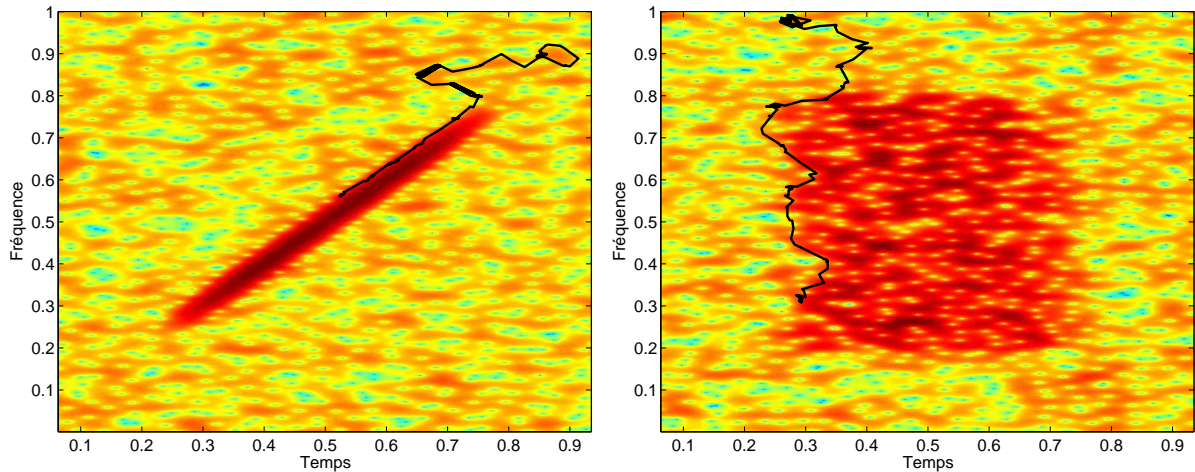


FIG. 6.11 : Exemples de contours non reconstruits. Pour des bruit trop grands, ici de variance 10^{-1} , l'algorithme de construction de contour peut suivre des points de haut degré de parallélisme δ_{\parallel} liés au bruit uniquement. Le contour créé quitte alors la zone frontière du motif, et devient incapable d'entourer correctement le motif temps-fréquence.

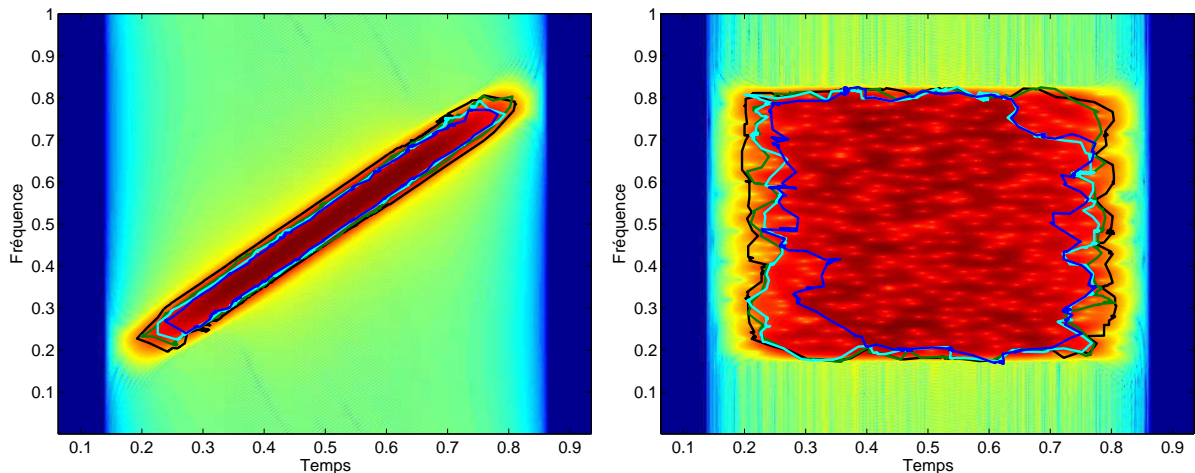


FIG. 6.12 : Contours reconstruits pour différentes variances de bruit blanc gaussien, ajouté aux signaux de test. En noir, le contour déterminé pour une variance du bruit additif nulle ; en vert, pour une variance de 10^{-4} ; en bleu clair, pour une variance de 10^{-3} ; en bleu foncé, pour une variance de 10^{-2} . Plus la variance du bruit additif est grande, plus la surface entourée par le contour reconstruit est petite.

reconstruit, qui s'appuie sur la zone frontière.

Ce phénomène est illustré sur la figure 6.12. Sur le spectrogramme des signaux non bruités sont représentés les contours déterminés avec des bruits additifs de variances 0, 10^{-4} , 10^{-3} et 10^{-2} . Quel que soit le type de signal, modulation de fréquence ou large bande, le contour reconstruit décrit un motif de surface plus petit dans le plan temps-fréquence à mesure que le niveau de bruit augmente.

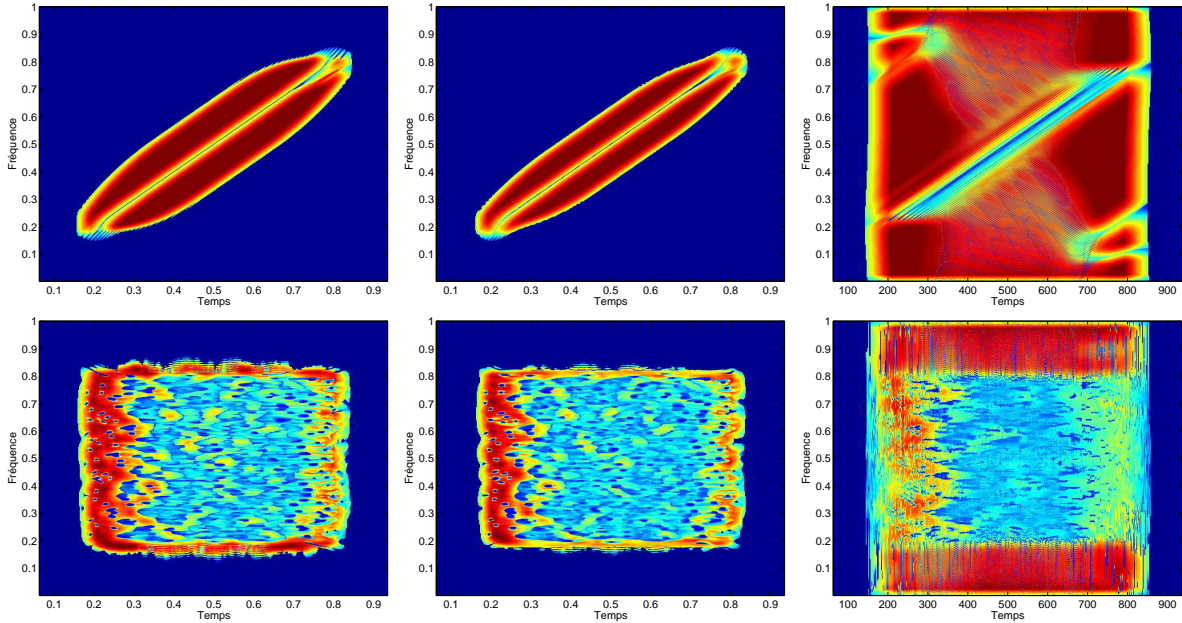


FIG. 6.13 : De gauche à droite, degré de parallélisme δ_{\parallel} des signaux de test sans bruit pour une fenêtre d'analyse de Hanning, de Blackman et rectangulaire respectivement. Ce même paramètre est donné pour une fenêtre gaussienne sur la figure 6.6. Le choix du noyau de lissage influe grandement sur la qualité de la description de la zone frontière par le degré de parallélisme.

La construction de contour proposée ici est influencée de la même manière que les motifs temps-fréquence segmentés dans la première partie de ces travaux de thèse : dans le cas de la segmentation par motif, les points temps-fréquence contenant du signal de faible énergie sont noyés par l'énergie du bruit, et ne sont alors plus détectés comme contenant du signal. Le support temps-fréquence du motif dépend alors du niveau de bruit additif. Dans le cas de la segmentation par contour décrite dans cette partie, ce sont les vecteurs de réallocation associés aux points de faible énergie sur la zone frontière du motif qui sont masqués par l'influence du bruit, réduisant de manière similaire la surface décrite par le contour.

Ce phénomène semble inhérent à l'analyse spectrale : les points de faible énergie sont indiscernables du bruit, quelle que soit la définition d'un motif temps-fréquence retenue.

6.5 Influence des paramètres de la TFCT

Le paramètre principal d'une TFCT – respectivement d'un spectrogramme – est le choix de sa fenêtre d'analyse $\phi(t)$ – respectivement de son noyau de lissage $W_{\phi}(t, f)$. Une fenêtre temporelle courte impliquera une grande fenêtre fréquentielle, autrement dit un choix est à faire sur les largeurs temporelle Δt (5.30) et fréquentielle Δf (5.31) du noyau de lissage. Les vecteurs de réallocation $\mathbf{r}(t, f)$ étant définis relativement à la taille du noyau de lissage, une faible largeur temporelle associée à une grande largeur fréquentielle créeront majoritairement des vecteurs de réallocation ayant une faible composante temporelle et une grande composante fréquentielle, c'est-à-dire dirigés verticalement.

Comme signalé dans la section 6.1, nous ne considérons que des vecteurs de réallocation discrétisés, sans lien direct avec les notions de temps et de fréquence. La discrétisation du plan

temps-fréquence se fait selon le nombre de fréquences calculées d'une part, et selon l'écart temporel entre deux fenêtres d'analyse successives d'autre part. Les pas de quantification en temps et en fréquence sont indépendants des largeurs temporelle et fréquentielle du noyau de lissage. Par conséquent, le voisinage $N(t, f)$, exprimé en points, sur lequel se calcule le degré de liberté δ_{\parallel} est indépendant du choix de la fenêtre d'analyse $\phi(t, f)$. Il n'y a *a priori* aucune raison de privilégier la dimension temporelle à la fréquentielle, et inversement. Ainsi, un choix de voisinage $N(t, f)$ approximativement de même taille en temporel et en fréquentiel semble naturel, choix qui a été retenu en pratique. De plus, la taille du voisinage retenue $N(t, f)$ a été choisie assez grande (31×53) pour garantir une grande plage de variation pour le degré de parallélisme δ_{\parallel} , impliquant une grande redondance dans le plan temps-fréquence, et donc des temps de calculs relativement longs.

Outre les largeurs temporelle et fréquentielle du noyau de lissage, considérées en continu ou de manière discrète, la forme de la fenêtre d'analyse, et donc celle du noyau de lissage dans le plan temps-fréquence joue grandement sur la qualité de la description de la zone frontière par le degré de parallélisme $\delta_{\parallel}(t, f)$. La figure 6.13 montre le degré de parallélisme calculé pour les deux signaux de test sans bruit, avec pour fenêtre d'analyse des fenêtres de Hanning, de Blackman et rectangulaire. Le degré de parallélisme pour une fenêtre gaussienne a déjà été illustré sur la figure 6.6. Les fenêtres de Hanning et de Blackman, de variations douces en temps et en fréquences, sans discontinuités temporelles présentent des degrés de parallélisme homogènes sur la zone frontière des motifs. Au contraire, la fenêtre rectangulaire, de par les forts lobes secondaires du sinus cardinal présentent des résultats totalement incohérents avec la zone frontière attendue. La fenêtre rectangulaire est donc particulièrement à proscrire dans le cadre de l'algorithme proposé dans ce chapitre, ne permettant pas de reconstruire le contour voulu.

Dans cette section, l'influence des paramètres de la TFCT a plus été illustrée que réellement étudiée. La forme de la fenêtre, la dissymétrie entre taille temporelle et taille fréquentielle du voisinage $N(t, f)$, la surface de ce voisinage et l'influence des largeurs temporelle Δt et fréquentielle Δf à voisinage discret $N(t, f)$ constant demandent encore à être caractérisées plus précisément.

6.6 Pour améliorer l'algorithme...

L'algorithme proposé souffre de plusieurs problèmes, notamment relatifs à la construction du contour lui-même : lors des itérations de construction, le point suivant est déterminé uniquement à partir du dernier point de contour, ce qui constitue un horizon très local. Si le dernier point sélectionné est hors de la zone frontière, ou si son angle ne pointe pas correctement vers le motif, le point suivant ne pourra que difficilement rectifier le problème. De plus, cette vision locale empêche de considérer le contour globalement, et peut produire un contour très chahuté, alors que les motifs dans le spectrogramme sont issus d'un lissage dans le plan temps-fréquence. Enfin, aucune information relative à l'énergie du motif n'est prise en compte.

L'autre problème majeur vient de la discrétisation du plan temps-fréquence, qui peut produire de grandes variances entre les angles de vecteurs de réallocation voisins sur une zone frontière du motif (*cf.* paragraphe 6.1). Pour pallier ce problème, nous cherchons à obtenir une version moyennée des vecteurs de réallocation, en supposant que la différence entre les angles de vecteurs voisins diminue, améliorant ainsi le degré de parallélisme δ_{\parallel} . Dans un premier temps, nous nous intéressons à la réallocation multi-fenêtres, récemment proposée par Xiao et Flandrin [XF06, XF07].

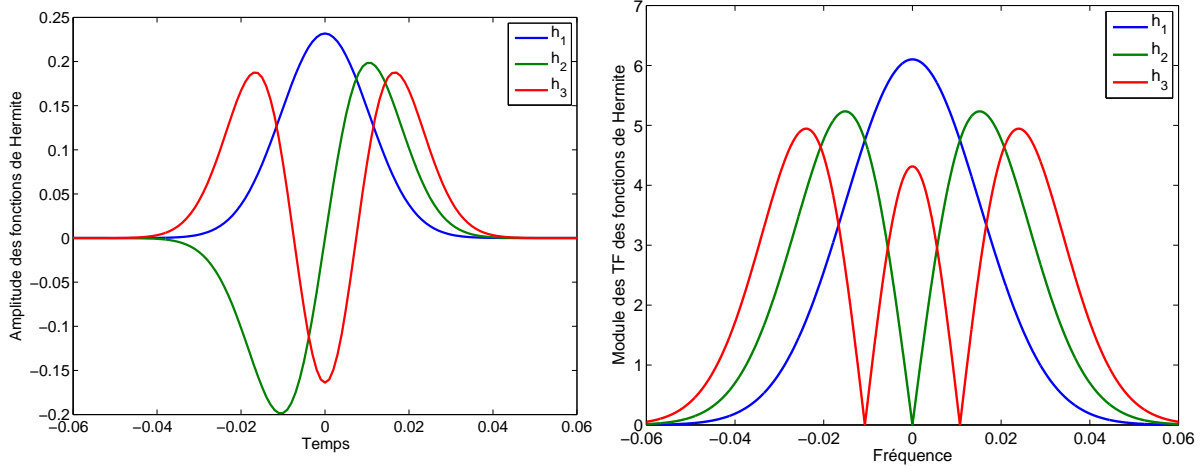


FIG. 6.14 : À gauche allure temporelle des trois premières fonctions de Hermite utilisées comme fenêtres d'analyse $\phi(t)$ du spectrogramme. À droite, le module des transformées de Fourier de ces fenêtres. De par les oscillations temporelles de ces fenêtres, le module du spectre présente des fluctuations dans sa bande passante.

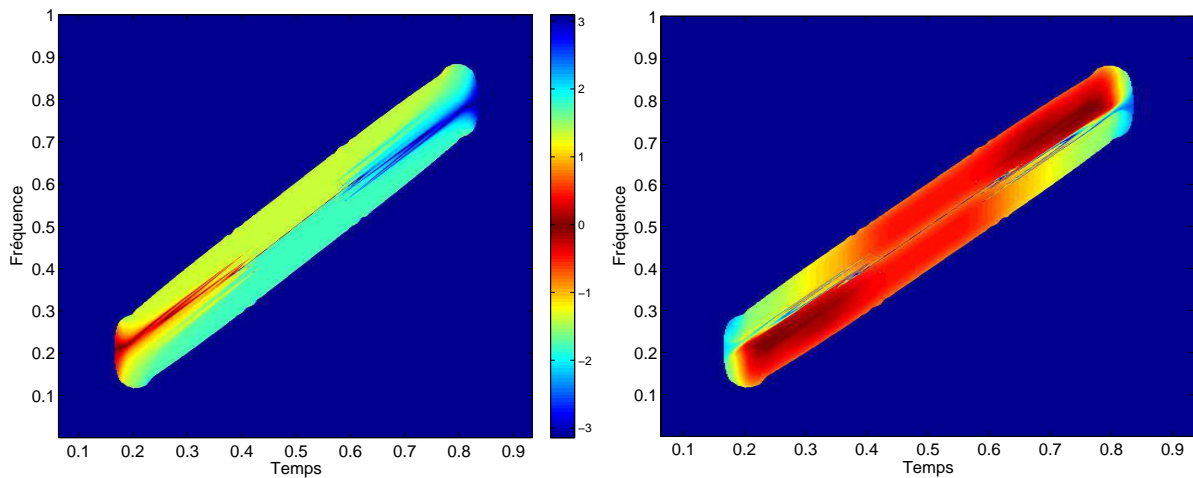


FIG. 6.15 : À gauche, angle des vecteurs de réallocation pour un spectrogramme calculé avec la cinquième fonction de Hermite. Les oscillations de la fenêtre d'analyse créent des oscillations dans l'angle des vecteurs de réallocation, perturbant le degré de parallélisme sur la zone frontière, qui est illustré à droite.

6.6.1 Réallocation multi-fenêtres

Cette réallocation part d'une proposition de Thomson [Tho82] pour estimer le spectre d'un signal x . Plutôt que d'utiliser la méthode de Welch, qui consiste à fenêtrer des durées limitées du signal et à moyenner les spectres obtenus sur chacune de ces durées², l'idée est de considérer le signal entier, mais en utilisant diverses fenêtres orthogonales. Cette méthode permet alors de diminuer la variance du spectre sans pour autant augmenter ce biais.

²Autrement dit, moyenner le spectrogramme d'un signal temporel dans le temps.

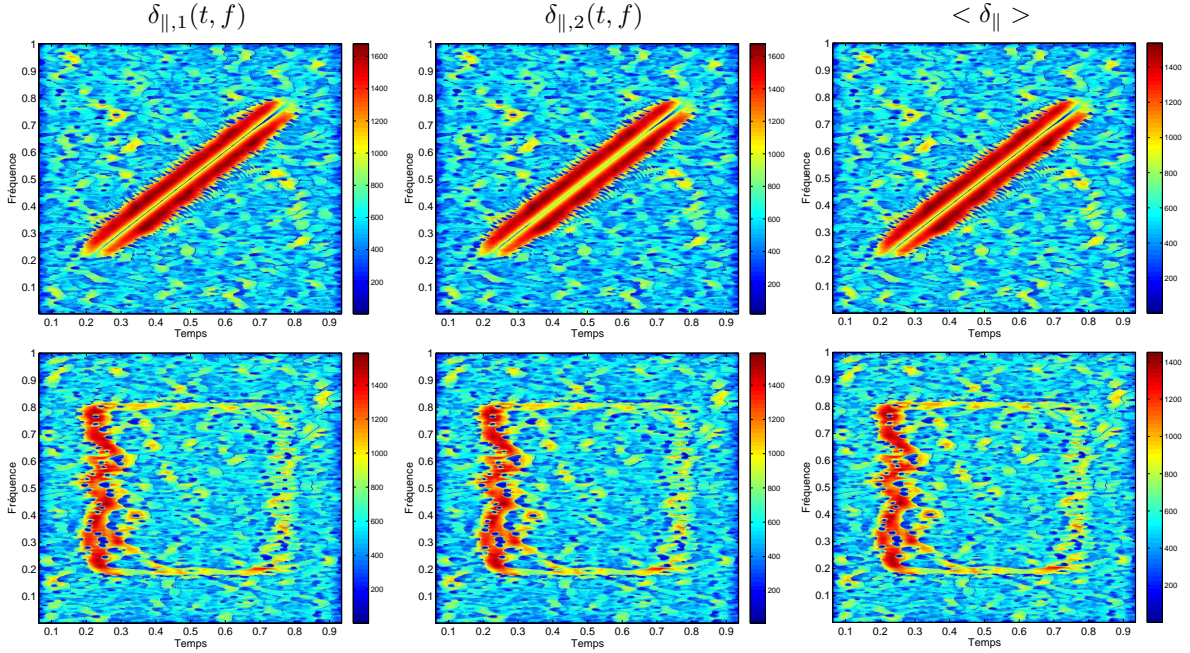


FIG. 6.16 : À gauche, degré de parallélisme $\delta_{||,1}(t, f)$ calculé à partir du champ de vecteurs de réallocation moyen de l'équation (6.8). Au centre, degré de parallélisme $\delta_{||,2}(t, f)$ calculé avec le champ de vecteurs de réallocation moyenné selon (6.9). À droite, degré de parallélisme moyen $\langle \delta_{||} \rangle$, issu de la moyenne entre les degrés de parallélisme calculés séparément pour la réallocation avec les fenêtres gaussienne, de Hanning et de Blackman. Les différences entre ces méthodes sont minimales.

Le même principe appliqué aux signaux non stationnaires conduit au spectrogramme multi-fenêtres [FB94, Tho01]. Dans ce cas, les fenêtres orthogonales sont les fonctions de Hermite [XF07]), illustrées sur la figure 6.14. Ces fonctions orthogonales présentent des oscillations temporelles et fréquentielles.

La réallocation multi-fenêtres réalloue chacun de ces spectrogrammes pour ensuite moyennner les spectrogrammes réalloués. Dans le cas de la construction de contour, nous ne regardons que les vecteurs de réallocation, plus précisément la manière dont ils pointent vers l'intérieur du motif. Toutefois, les fluctuations des fenêtres d'analyse utilisées ici perturbent les vecteurs de réallocation. Comme illustré sur la figure 6.15 pour la cinquième fonction de Hermite, certains vecteurs de réallocation présents sur la zone frontière ne sont pas correctement dirigés vers l'intérieur du motif. Par conséquent, le degré de parallélisme $\delta_{||}$ basé sur cet angle présente de grandes disparités le long de la zone frontière.

6.6.2 Réallocation pseudo multi-fenêtres

Plutôt que d'utiliser des fenêtres orthogonales, qui ont de bonnes propriétés pour diminuer la variance du spectrogramme, mais sont gênantes pour les vecteurs de réallocation, nous proposons ici une réallocation multi-fenêtres, avec des fenêtres quelconques. En l'occurrence, le but est d'obtenir plusieurs champs de vecteurs de réallocation afin de les moyennner, et limiter ainsi les problèmes liés à la discrétisation du plan temps-fréquence. Autrement dit, la diminution de la

variance du spectrogramme ne revêt que peu d'importance dans ce contexte.

Comme vu dans la section précédente, les fenêtres d'analyse douces en temps et en fréquence, à savoir les fenêtres gaussienne, de Hanning et de Blackman, conduisent à des vecteurs de réallocation sur la zone frontière du motif pointant correctement vers le motif, respectivement notés $\mathbf{r}_{x,g}(t, f)$, $\mathbf{r}_{x,h}(t, f)$ et $\mathbf{r}_{x,b}(t, f)$.

Deux moyennes de champs de vecteurs de réallocation sont envisagés. La première consiste à moyenner de manière classique les vecteurs de réallocation issus des réallocations selon les trois fenêtres d'analyse, ce qui s'écrit :

$$\mathbf{r}_1(t, f) = \frac{1}{3} (\mathbf{r}_{x,g}(t, f) + \mathbf{r}_{x,h}(t, f) + \mathbf{r}_{x,b}(t, f)) \quad (6.8)$$

La seconde ne considère que les angles de ces vecteurs de réallocation, qui constituent l'unique critère sur lequel est basée la construction de contour. Autrement dit, les vecteurs de réallocation sont normés à 1 avant d'être moyennés, conduisant à un champ de vecteurs moyenné de norme également égale à 1 :

$$\mathbf{r}_2(t, f) = \frac{1}{3} \left(\frac{\mathbf{r}_{x,g}(t, f)}{|\mathbf{r}_{x,g}(t, f)|} + \frac{\mathbf{r}_{x,h}(t, f)}{|\mathbf{r}_{x,h}(t, f)|} + \frac{\mathbf{r}_{x,b}(t, f)}{|\mathbf{r}_{x,b}(t, f)|} \right) \quad (6.9)$$

Les champs de vecteurs de réallocation $\mathbf{r}_1(t, f)$ et $\mathbf{r}_2(t, f)$ conduisant à deux degrés de parallélisme différents, notés $\delta_{\parallel,1}(t, f)$ et $\delta_{\parallel,2}(t, f)$ respectivement, calculés selon l'équation (6.3). Ces deux degrés de parallélisme sont comparés avec le degré de parallélisme moyen $\langle \delta_{\parallel} \rangle$, défini comme la moyenne des degrés de parallélisme calculés séparément pour chaque fenêtre d'analyse considérée. La figure 6.16 montre les degrés de parallélisme issus des moyennages des champs de vecteurs de réallocation et le degré de parallélisme moyen.

Les degrés de parallélisme entre ces différents moyennages étant sensiblement les mêmes, la construction du contour est similaire dans tous les cas, et ne présente par conséquent aucun avantage.

6.6.3 Lissage du résultat obtenu par contour actif

Les contours actifs ont été introduits par Kass, Witkin et Terzopoulos dans les années 80 [KWT88]. Le principe repose sur la minimisation d'une fonction appelée par analogie physique « énergie ». Le contour est ici un objet déformable, dont l'énergie est composée de deux termes : le premier est un terme lié à la régularité du contour, qui sera d'autant plus faible que la courbe est régulière ; le second est un terme d'attraction vers les données.

Outre le critère d'attraction vers les données, la difficulté des contours actifs est de définir un contour initial. Une extension de ce principe, la méthode du ballon [Coh91] propose de partir d'un petit contour situé à l'intérieur du contour à construire, et d'ajouter une force d'expansion dans l'énergie, faisant grandir ce contour jusqu'à ce qu'il tombe sur un minimum d'énergie lié à l'image.

Cette dernière méthode a été testée pour construire un contour de motif temps-fréquence à partir du degré de parallélisme $\delta_{\parallel}(t, f)$. Le critère lié à l'image consiste alors à maximiser la somme des degrés de parallélisme situés sur les points de contour. Ce critère s'est toutefois révélé trop peu efficace, le ballon représentant le contour actif continuant à croître lorsqu'il se situe sur la zone frontière du motif temps-fréquence.

Le contour actif semble toutefois rester une bonne solution pour lisser *a posteriori* le contour construit par l'algorithme 6.8.

Conclusion

Cette partie a présenté un algorithme de détection et de construction de contours de motifs du spectrogramme, basé sur des concepts uniquement « traitement du signal », et sans information *a priori* sur le signal. Cet algorithme s'appuie sur une définition d'un motif temps-fréquence – plus précisément sur une définition de la frontière d'un motif temps-fréquence – lié à la réallocation du spectrogramme.

La réallocation est une représentation temps-fréquence du signal non-linéaire, qui, sachant que le spectrogramme est une distribution de Wigner-Ville lissée par un noyau temps-fréquence lié à la fenêtre d'analyse du spectrogramme, déplace l'énergie du spectrogramme sur le centre de gravité local du signal dans le plan temps-fréquence. À la frontière d'un motif temps-fréquence, là où l'énergie du spectrogramme n'est présente que par le lissage par le noyau temps-fréquence, les vecteurs de réallocation pointent tous vers l'intérieur du motif, et sont par conséquent localement parallèles.

Cette définition permet de n'avoir aucune considération statistique sur le signal, tout en étendant l'intérêt de la réallocation aux signaux large bande.

Après avoir illustré les problèmes liés à la discrétisation du plan temps-fréquence, qui imposent de considérer deux vecteurs parallèles à un angle θ_0 près, un premier algorithme de détection de zones correspondant aux frontières d'un motif temps-fréquence a été proposé. Si tous les vecteurs de réallocation sur un voisinage restreint sont parallèles (θ_0 près), le point est considéré comme appartenant à la frontière du motif temps-fréquence. Cet algorithme a été utilisé comme source d'information supplémentaire à la segmentation présentée dans la première partie de cette thèse, dans le cadre d'une segmentation d'harmoniques provenant d'un accéléromètre placé sur un axe de machine tournante. Il permet ainsi de savoir si une seule ou plusieurs harmoniques ont été segmentées dans une même classe.

Un second algorithme a été proposé, se basant sur un degré de parallélisme des points temps-fréquence, défini comme le nombre de vecteurs parallèles sur un voisinage lié à la taille du noyau temps-fréquence de lissage du spectrogramme. La construction du contour, une ligne entourant le motif temps-fréquence, se fait en suivant la ligne de crête de ce degré de parallélisme.

L'influence d'un bruit blanc gaussien additif a été examinée. Un niveau de bruit peut créer des points temps-fréquence présentant un haut degré de parallélisme hors de la zone frontière des motifs. Dans ce cas, la construction d'un contour est impossible. Le bruit additif garde une influence pour des niveaux modérés : à l'instar de la segmentation statistique par motif, qui voit disparaître les points de plus faible énergie du signal quand le bruit augmente, les vecteurs de réallocation correspondant aux faibles énergies du signal sont noyés par le bruit. Ainsi, la surface décrite par le contour construit est plus faible à mesure que le niveau de bruit augmente.

L'influence des paramètres de la TFCT a également été illustrée. Le choix de la fenêtre

d'analyse est primordial pour avoir une construction de contour possible, et doit être régulière en temps et en fréquence. De plus, l'algorithme actuel nécessite une grande redondance dans le spectrogramme, afin d'avoir un voisinage de calcul du degré de parallélisme suffisamment grand pour que le degré de parallélisme ait une plage de variations suffisamment grande. Toutefois, l'influence exacte des paramètres reste à étudier.

Enfin, quelques pistes d'améliorations, par la réallocation multi-fenêtres et les contours actifs, ont été proposées et n'ont pas donné de résultats.

Afin de rendre cet algorithme utilisable en pratique, plusieurs points restent à régler. Les soucis soulevés par cette méthode sont de deux types.

Le premier est issu de la discrétisation du plan temps-fréquence, qui crée artificiellement une grande variabilité sur l'angle de deux vecteurs de réallocation voisins sur une zone frontière, empêchant ainsi leur strict parallélisme. Les essais pour limiter ce problème, à savoir le moyennage de champ de vecteurs de réallocation provenant de différentes réallocations du spectrogramme, n'ont donné jusqu'ici aucun résultat probant.

L'autre problème est lié à l'algorithme de création de contour en lui-même. Le contour se construit point à point en fonction du degré de parallélisme sur les points proches du dernier point de contour. Aucune vision globale du contour n'est prise en compte : le caractère lisse d'un contour de motif temps-fréquence n'est pas pris en compte, ce qui laisse le contour osciller localement autour de quelques points de grand degré de parallélisme. De plus, l'énergie dans le plan temps-fréquence est totalement ignorée. Ces informations supplémentaires pourraient améliorer la robustesse de l'algorithme, et éviter de quitter la zone frontière lorsqu'un niveau d'un bruit additif devient trop grand.

Enfin, le seul cas traité ici est un signal à segmenter ne comportant qu'une seule composante. Le cas multi-composantes reste à développer.

Conclusions générales

L'objectif de ces travaux de thèse est de présenter deux méthodes distinctes de segmentation de motifs temps-fréquence, en tâchant de rester dans un cadre le plus général possible, c'est-à-dire en limitant au maximum les hypothèses *a priori* sur le signal à segmenter.

La première de ces méthodes cherche à segmenter le *support temps-fréquence* d'un motif, autrement dit l'ensemble des points temps-fréquence sur lequel une contribution énergétique provenant du signal à segmenter est présente, dans la continuation des travaux de thèse de Cyrille Hory [Hor02], et en parallèle de ceux de Julien Huillery [Hui08]. Pour ce faire, on s'appuie sur un modèle statistique de signal, à savoir un signal d'intérêt que l'on cherche à segmenter, auquel un bruit blanc gaussien de variance inconnue est ajouté. L'intérêt de ce modèle est de n'imposer aucune condition spécifique sur le signal à segmenter, ce dernier pouvant être aléatoire comme déterministe, une modulation de fréquence comme un signal large bande.

Un chapitre 1 s'est attaché à caractériser en deux temps les lois que suivent les coefficients de la Transformée de Fourier à Court Terme (TFCT) d'un signal suivant un tel modèle. Tout d'abord, nous nous sommes penchés sur le cas « bruit seul », quand il n'y a pas de signal à segmenter. Nous avons montré que les coefficients temps-fréquence suivent chacun une loi gaussienne, et déterminé la variance de ces lois selon le point temps-fréquence considéré. De plus, selon la mise en œuvre de la TFCT retenue, une corrélation entre parties réelle et imaginaire des coefficients peut apparaître. Une proposition de TFCT « rectifiée » a été faite, de sorte que tous les points temps-fréquence suivent une même loi gaussienne de variance identique, sans corrélation entre parties réelle et imaginaire. Ainsi, une loi unique permet de décrire tous les coefficients de la TFCT d'un bruit blanc gaussien.

Le cas « signal inconnu+bruit » a également été étudié. En l'absence d'un modèle de loi pour le signal à segmenter, la loi d'un coefficient temps-fréquence contenant du signal et du bruit est inconnue. Toutefois, nous avons montré que le moment d'ordre deux d'un tel point reste supérieur au moment d'ordre deux d'un point ne contenant que du bruit.

Ce constat nous permet dans le chapitre 2 de définir un test binaire d'hypothèse, en supposant le niveau de bruit connu. Ainsi, nous définissons une manière de décider si un point de la TFCT « rectifiée » contient de l'énergie du signal à segmenter ou non. L'énergie du signal à un point temps-fréquence considéré étant inconnue, une stratégie de Neyman-Pearson, fixant une probabilité de fausse alarme, est retenue.

Le bruit reste toutefois inconnu, donc à estimer. Pour ce faire, une méthode itérative similaire à celle retenue par Hory [Hor02] a été étudiée dans le chapitre 3. À chaque itération, le niveau de bruit est estimé sur les points non détectés comme contenant du signal. Aux premières itérations, le niveau de bruit est surestimé par la présence du signal : cette surestimation permet toutefois de détecter des points contenant du signal à segmenter. L'itération suivante produit une estimation

moins biaisée, des points de signal ayant été détectés. L'originalité des travaux présentés dans ce document est de prendre comme critère d'arrêt un moment d'ordre 4 complexe, le kurtosis spectral.

D'un côté, il est connu que le kurtosis (spectral ou non) d'une variable gaussienne est nul. De l'autre, nous avons illustré le fait qu'un ensemble de points temps-fréquence ne contenant que du bruit et de points contenant du signal et du bruit possède en général un kurtosis bien supérieur à zéro. Ainsi, le critère d'arrêt des itérations consiste en un simple seuil sur le kurtosis spectral des points non segmentés, en dessous duquel tous les points non segmentés seront considérés comme issus d'une même loi gaussienne, autrement dit du bruit uniquement.

L'influence de la diversité des variances des lois des coefficients de la TFCT du bruit seul a été étudiée. L'estimation du kurtosis spectral de points ne contenant que du bruit s'est révélée moins biaisée dans le cas de la TFCT « rectifiée » que dans le cas de la TFCT usuelle.

Une fois l'estimation du bruit et la détection de points contenant du signal mises en place, l'algorithme en lui-même a été présenté dans le chapitre 4. À chaque itération d'estimation de bruit et de détection de point de signal, des motifs sont créés à l'aide d'un algorithme de croissance de région. Les points détectés sont alors regroupés en classes, chaque classe appartenant à un motif temps-fréquence distinct. L'algorithme proposé est contrôlé par trois paramètres, outre les paramètres inhérents à la TFCT. Le premier est la probabilité de fausse alarme, liée à la détection de Neyman-Pearson ; le second est la proportion de points détectés à chaque itération qui sont affectés à des classes par la croissance de région ; le dernier est le seuil sur le kurtosis spectral des points non segmentés, déterminant la fin de l'algorithme. Des propositions de choix automatiques ont été faites pour les deux derniers paramètres, le premier restant un choix de l'utilisateur de l'algorithme. L'influence de ces trois paramètres sur le résultat de la segmentation a été étudiée, montrant que les choix automatiques forment des compromis satisfaisants.

Le principe de cet algorithme a ensuite été étendu selon deux dimensions. Dans un premier temps, le caractère blanc et stationnaire du bruit additif gaussien a été supprimé. Une estimation locale dans le plan temps-fréquence du niveau de bruit permet de normaliser le plan temps-fréquence de telle sorte que le bruit équivalent soit blanc et stationnaire. Ainsi nous pouvons appliquer le même algorithme. Ensuite, nous nous sommes intéressés à deux autres transformations qui conservent le caractère gaussien complexe des points ne contenant que du bruit, à savoir la transformée en ondelettes et le caponogramme, issu de l'estimation à court terme du spectre par la méthode de Capon. Toutes ces extensions ont été illustrées sur un signal synthétique.

Enfin, des applications potentielles de cet algorithme ont été illustrées, sur signaux synthétiques et réels, à savoir le filtrage temps-fréquence, la séparation de source et la caractérisation de signaux.

La seconde méthode présentée dans cette thèse ne nécessite aucun modèle de signal. L'idée est de s'appuyer sur la réallocation du spectrogramme, non pas comme nouvelle représentation temps-fréquence, mais comme source d'information sur la structure du signal. Par conséquent, la méthode de réallocation a été brièvement présentée dans le chapitre 5. Considérant que le spectrogramme d'un signal peut être vu comme la distribution de Wigner-Ville de signal lissé en temps-fréquence par un noyau défini par la distribution de Wigner-Ville de la fenêtre d'analyse du spectrogramme, la réallocation déplace l'énergie du spectrogramme vers le centre de gravité local de l'énergie dans le plan temps-fréquence. Ainsi, le spectrogramme, version « étalée » de Wigner-Ville, est refocalisé.

La réallocation considérée ainsi nous inspire une définition de la zone frontière d'un motif :

sur la frontière d'un motif dans le spectrogramme, l'énergie va être déplacée vers l'intérieur du motif. Tous les vecteurs de réallocation décrivent comment l'énergie est déplacée, et vont pointer vers l'intérieur du motif. Ainsi, les vecteurs de réallocation seront localement parallèles sur cette zone frontière.

Dans le chapitre 6, un algorithme est proposé afin de détecter cette zone frontière, puis de construire un contour entourant le motif temps-fréquence. Dans un premier temps, le problème de la discrétisation du plan temps-fréquence est soulevé. Cette discrétisation empêche un strict parallélisme des vecteurs de réallocation, qui sont alors considérés parallèles à un angle de seuil près. En regardant sur un petit voisinage temps-fréquence si tous les vecteurs de réallocation sont parallèles, nous proposons un premier algorithme de détection de frontière, présentant de très nombreuses fausses alarmes. Toutefois, cet algorithme a pu être utilisé en pratique, en complément d'information à la segmentation présentée en première partie.

Dans un deuxième temps, nous avons défini un *degré de parallélisme* pour chaque vecteur de réallocation, égal au nombre de vecteurs parallèles à ce vecteur sur un voisinage dépendant de la taille du noyau de lissage du spectrogramme. Un algorithme de type suivi de ligne de crête a pu être proposé, construisant le contour du motif temps-fréquence.

Cet algorithme a été illustré pour deux signaux de test, une modulation linéaire de fréquence et un signal large bande, provenant d'un bruit blanc filtré. Le contour étant reconstruit pour ces deux types de signaux, nous avons montré l'intérêt que présente la réallocation dans le cas des signaux large bande. L'influence d'un bruit additif blanc gaussien a été illustrée, ainsi que les paramètres de la TFCT conduisant au spectrogramme. L'algorithme nécessite beaucoup de redondance dans le spectrogramme afin d'être efficace, ainsi qu'un rapport signal à bruit assez bon, rendant son utilisation pour des signaux réels difficile. Des pistes d'améliorations ont été proposées, qui ne se sont pas révélées fructueuses.

Discussion

Ces travaux ont cherché par deux méthodes différentes à détecter le support temps-fréquence d'une composante spectrale, que ce soit en déterminant directement ce support par la segmentation statistique, ou en décrivant un contour entourant ce support, par la segmentation basée sur le champ de vecteurs de réallocation. La question de définition d'un support temps-fréquence reste ouverte : en effet, un signal borné en temps présente de l'énergie sur toutes les fréquences, tandis qu'un signal borné en fréquence présentera une durée infinie. Un support temps-fréquence n'est pas par conséquent borné. Toutefois, l'énergie d'un motif reste localisée dans le plan temps-fréquence. Comment définir proprement son support ?

Dans le cas de la segmentation statistique, le support temps-fréquence d'un motif est directement lié au rapport signal à bruit local [Hui08], c'est-à-dire le rapport entre l'énergie du signal et l'énergie du bruit seul à un point temps-fréquence donné. Pour un rapport signal à bruit local très faible, l'énergie du signal est noyée dans le bruit, et devient ainsi indétectable. Quand le signal devient indiscernable du bruit, il devient *de facto* négligeable. Le support temps-fréquence du motif est alors l'ensemble des points détectables.

Dans le cas de la segmentation par la réallocation, le support du motif est limité par le seuil sur l'énergie du noyau de lissage, distribution de Wigner-Ville de la fenêtre d'analyse du spectrogramme. Sans cette troncature, le noyau de lissage aurait un support infini, conduisant à un voisinage infini pour déterminer le degré de parallélisme des vecteurs de réallocation. Toutefois, même en ignorant ce seuil, il est apparu que le contour reconstruit d'un motif temps-fréquence d'un signal perturbé par un bruit blanc gaussien dépend du niveau du bruit additif. Le contour construit entoure une surface d'autant plus faible que le bruit est fort.

Le support temps-fréquence varie selon le bruit de la même manière dans les deux segmentations. Y a-t-il un lien direct ? Par corollaire, est-ce la seule manière de définir le support d'un motif temps-fréquence dans le cas d'un signal perturbé par du bruit ?

Perspectives

Concernant l'algorithme de segmentation statistique, plusieurs pistes sont à envisager. Tout d'abord, le problème de l'estimation du bruit, qui est effectué de manière itérative : il en résulte en un temps de calcul très variable selon les signaux analysés, dépendant de la convergence du critère d'arrêt. Une méthode d'estimation de type « statistiques minimales » [Mar06], telle que présentée dans les travaux de Huillery [Hui08], peut pallier ce problème. Des premiers essais utilisant le kurtosis spectral pour déterminer quels sont les points temps-fréquence sur lesquels estimer la variance d'un bruit blanc additif semblent prometteurs. Une estimation du niveau de bruit avant l'algorithme de segmentation permet de faire disparaître les itérations, ainsi que l'intérêt de l'algorithme de croissance de région : le seul côté « traitement d'image » de cet algorithme est supprimé. Un critère de connexité permet de reconstruire des classes disjointes dans le plan temps-fréquence, sans rajout d'informations *a priori*.

La phase de la TFCT n'a été utilisée pour l'instant que pour caractériser les points ne contenant que du bruit, qui est gaussien complexe circulaire, c'est-à-dire que la répartition de la phase est uniforme. Un complément d'information peut être apporté par le spectrogramme de phase, proposé par Léonard [L00, L06]. Ce spectrogramme de phase consiste en une construction judicieuse de la TFCT, notamment en utilisant des fenêtres d'analyse présentant des phases constantes [RV70], puis en ne considérant que l'angle de cette TFCT. Le spectrogramme de phase résultant offre d'intéressantes caractéristiques, susceptibles d'aider à la détection d'un point temps-fréquence contenant du signal d'un part, et de caractériser les motifs segmentés d'autre part.

Une troisième perspective est d'étendre l'algorithme de segmentation statistique à d'autres transformées, comme cela a déjà été initié par les extensions à la transformée en ondelettes et au caponogramme présentées dans le chapitre 4. Toute transformée linéaire, gardant le caractère gaussien des points ne contenant que du bruit, est susceptible de servir de base à ce type d'algorithme. Toutefois, de nombreuses représentations temps-fréquence sont quadratiques, offrant directement une représentation en énergie. Une extension à l'ensemble de la classe de Cohen et à la classe affine nécessite alors de déterminer un critère équivalent au kurtosis dans ce cadre.

Pour l'algorithme basé sur la réallocation du spectrogramme, le principal objectif reste de créer un algorithme robuste et fiable utilisable sur signaux réels. Jusqu'à présent, les essais sur signaux réels avec cet algorithme se sont révélés décevants. Plusieurs pistes ont été évoquées dans le chapitre 6, à savoir l'amélioration de l'algorithme de création lui-même, le moyennage de plusieurs champs de vecteurs de réallocation afin de limiter le problème lié à la discrétisation du plan temps-fréquence de battement d'angle des vecteurs de réallocation, et un post-traitement par une méthode de contour actif.

De plus, cet algorithme n'a fait l'objet d'aucune évaluation. Une comparaison avec des méthodes classiques de détection de contour, telles que celle de Canny [Can86] basée sur le gradient de l'image est nécessaire, ainsi qu'une comparaison avec d'autres algorithmes de segmentation

de motifs temps-fréquence.

Enfin, les deux segmentations présentées ont été développées et testées séparément l'une de l'autre, et sont basées sur deux concepts bien différents, la première étant statistique, la deuxième basée sur les vecteurs de réallocation du spectrogramme. Un algorithme général mêlant les informations provenant de ces deux méthodes produirait probablement des résultats plus robustes, fusionnant les intérêts propres à chaque méthode.

Bibliographie

- [AB94] Rolf Adams et Leanne Biscof. Seeded Region Growing. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(6) :641–647, 1994.
- [AF94] François Auger et Patrick Flandrin. The why and how of time-frequency reassignment. *Proc. of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Octobre 1994.
- [AF95] François Auger et Patrick Flandrin. Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method. *IEEE Trans. Signal Proc.*, 43 :1068–1089, 1995.
- [AGL96] P.O. Amblard, M. Gaeta, et J.L. Lacoume. Statistics for complex variables and signals - Part I : Variables. *Signal Processing*, 53 :1–13, 1996.
- [Can86] J. Canny. A Computational Approach To Edge Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, (8) :679–714, 1986.
- [Cap69] J. Capon. High resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57 :1408–1418, August 1969.
- [Cas03] Francis Castanié. *Analyse spectrale*, chapter 7 (Minimum de variance), pages 171–203. Hermès, 2003.
- [CM98] Eric Chassande-Mottin. *Méthodes de réallocation dans le plan temps-fréquence pour l'analyse et le traitement de signaux non stationnaires*. Thèse de doctorat, Université de Cergy-Pontoise, Septembre 1998.
- [CMFA98] Eric Chassande-Mottin, Patrick Flandrin, et François Auger. On the statistics of spectrogram reassignment vectors. *Multidimensional Syst. Signal Process.*, 9(4) :355–362, 1998.
- [Coh89] L. Cohen. Time-Frequency Distributions - A Review. *Proceedings of the IEEE*, 77(7) :941–981, Juiller 1989.
- [Coh91] Laurent D. Cohen. On active contour models and balloons. *CVGIP : Image Understanding*, 53(2) :211–218, March 1991.
- [CSL96] C. Capdevielle, , C. Servièrè, et J-L. Lacoume. Blind separation of wide-band sources : Application to rotating machine signals. In *European Signal Processing Conf.*, Trieste, Italia, Septembre 1996.
- [CT91] Thomas M. Cover et Joy A. Thomas. *Elements of information theory*. Wiley and Sons, New York, NY, USA, 1991.

- [DN73] T.S. Durrani et J.M. Nightingale. Probability distributions for discrete Fourier spectra. *Proc. IEE*, 120(2) :299–311, Février 1973.
- [FB94] G. Fraser et B. Boashash. Multiple window spectrogram and time-frequency distributions. In *Proceedings of the Acoustics, Speech, and Signal Processing, 1994. on IEEE International Conference*, pages 293–296, Washington, DC, USA, 1994.
- [Fla93] P. Flandrin. *Temps-fréquence*. Hermès, 1993.
- [HMM08] Julien Huillery, Fabien Millioz, et Nadine Martin. On the Description of Spectrogram Probabilities With a Chi-Squared Law. *IEEE Transactions on Signal Processing*, 56(6) :2249–2258, Juin 2008.
- [Hor02] Cyrille Hory. *Mélanges de distributions du χ^2 pour l'interprétation d'une représentation temps-fréquence*. Thèse de doctorat, INP Grenoble, Décembre 2002.
- [Hui08] Julien Huillery. *Support temps-fréquence d'un signal inconnu en présence de bruit additif gaussien*. Thèse de doctorat, INP Grenoble, Juiller 2008.
- [JKB94] N.L. Johnson, S. Kotz, et N. Balakrishnan. *Continuous Univariate Distributions*, volume 2. Wiley and sons, 2nd édition, 1994.
- [JL99] P.E. Johnson et D.G. Long. The Probability Density of Spectral Estimates Based on Modified Periodogram Averages. *IEEE Transactions on Signal Processing*, 47(5) :1255–1261, Mai 1999.
- [JPH⁺99] H.T. Jeong, D.G. Park, J.H. Hong, Y.S. Ahn, et G.M. Kim. The effect of microstructural changes on magnetic barkhausen noise in Mn-Mo-Ni pressure vessel steel. *J. Korean Phys. Soc.*, 34(5) :429–433, 1999.
- [Kay98a] S.M. Kay. *Fundamentals of Statistical Signal Processing, Volume 1 : Estimation Theory*. Prentice Hall, 1998.
- [Kay98b] S.M. Kay. *Fundamentals of Statistical Signal Processing, Volume 2 : Detection Theory*. Prentice Hall, 1998.
- [Koo74] L.H. Koopmans. *The Spectral Analysis of Time Series*. Academic Press, 1974.
- [KS94] M. Kendall et A. Stuart. *The advanced theory of statistics*. Griffin, 1994.
- [KVG76] K. Kodera, C. De Villedary, et R. Gendrin. A new method for the numerical analysis of nonstationary signals. *Phys. Earth and Plan. Int.*, 12, 1976.
- [KVG78] K. Kodera, C. De Villedary, et R. Gendrin. Analysis of Time-Varying Signals with Small BT Values. *Acoustics, Speech and Signal Processing, IEEE Trans. on*, 26, 1978.
- [KWT88] Michael Kass, Andrew Witkin, et Demetri Terzopoulos. Snakes : Active contour models. *International Journal of Computer Vision*, V1(4) :321–331, January 1988.
- [L00] François Léonard. Spectrogramme de phase et spectrogramme de fréquence. *TS. Traitement du signal*, 17(4) :269–286, 2000.
- [L06] François Léonard. Phase and frequency spectrograms. In *Proceedings of International Congress on Sound and Vibration 13*, Vienna, Austria, Juiller 2006.

-
- [Lan02] R. Langelle. *Décision et reconnaissance de formes en signal*. Hermès, 2002.
- [LAP97] J.-L. Lacoume, P.-O. Amblard, et P. Comon. *Statistiques d'ordre supérieur pour le traitement du signal*. Masson, 1997.
- [LBR⁺04] W.M. Liu, V.J. Rivas Bastante, F. Romero Rodriguez, N.W.D. Evans, et J.S.D. Mason. Morphological Filtering of Spectrograms for Automatic Speech Recognition. In *Proceedings of Visualization, Imaging, and Image Processing*, Marbella, Spain, Septembre 2004.
- [LG84] M. A. Lagunas et A. Gasull. An improved maximum likelihood method for power spectral density estimation. *IEEE Transactions on Acoustics Speech Signal Processing*, ASSP-32(1) :170–173, Feb. 1984.
- [LM02] Benoit Leprettre et Nadine Martin. Extraction of pertinent subsets from time-frequency representations for detection and recognition purposes. *Signal Processing*, 82 :229–238, 2002.
- [LNPW00] C. Lachmann, Th. Nitschke-Pagel, et H. Wohlfahrt. Characterization of residual stress relaxation in fatigue loaded welded joints by x-ray diffraction and barkhausen noise method. Number 347–349, pages 374–379, 2000.
- [LR05] E.L. Lehmann et J. P. Romano. *Testing Statistical Hypothesis*. Springer, 3rd édition, 2005.
- [LSGM86] M. A. Lagunas, M. E. Santamaria, A. Gasull, et A. Moreno. Maximum likelihood filters in spectral estimation problems. *Signal Processing*, 10(1) :19–34, Jan. 1986.
- [Mar06] Rainer Martin. Bias compensation methods for minimum statistics noise power spectral density estimation. *Signal Process.*, 86(6) :1215–1229, 2006.
- [MM06] F. Millioz et N. Martin. Detection of the boundaries of time-frequency patterns with a reassignment-based method. In *Proceedings of the Seventh International Conference on Mathematics in Signal Processing*, Cirencester, England, Décembre 2006.
- [MM07a] F. Millioz et N. Martin. Réallocation du spectrogramme pour la détection de frontières de motifs temps-fréquence. In *21e colloque GRETSI sur le traitement du signal et des images*, Troyes, France, Septembre 2007.
- [MM07b] F. Millioz et N. Martin. Reassignment Vector Field for Time-Frequency Segmentation. In *Proceedings of International Congress on Sound and Vibration 14*, Cairns, Australia, Juiller 2007.
- [Pic94] B. Picinbono. On Circularity. *Proceedings of the IEEE*, 42(12) :3473–3482, Décembre 1994.
- [PMM09] L.R. Padovese, N. Martin, et F. Millioz. Time-Frequency and Time-Scale analysis of Barkhausen noise signals. *Proc. IMechE, Part G : J. Aerospace Engineering*, (223(G5)) :577–588, 2009.
- [RLEM03] F. Romero Rodriguez, W.M. Liu, N.W.D. Evans, et J.S.D. Mason. Morphological Filtering of Speech Spectrograms in the Context of Additive Noise. In *Proceedings of Eurospeech*, Geneva, Switzerland, Septembre 2003.

- [Row00] S. T. Roweis. One microphone source separation. In *Advances in Neural Information Processing Systems*, pages 793–799, 2000.
- [RV70] D.C. Rife et D.A. Vincent. Use of the Discrete Fourier Transform in the Measurements of Frequencies and Level of Tones. *Bell System Technical Journal*, 49(2) :197–228, 1970.
- [Sax04] Olivier Saxod. *Critères d'identification de signaux sismiques d'avalanches*. Rapport de DEA, INP Grenoble, Août 2004.
- [Ser98] C. Servière. Feasibility of source separation in frequency domain. In *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing*, Seattle, USA, Mai 1998.
- [Tho82] David J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9) :1055–1096, 1982.
- [Tho01] David J. Thomson. Multitaper analysis of nonstationary and nonlinear time series data. In R. L. Smith, W. J. Fitzgerald, A. T. Walden, et P. C. Young, editors, *Nonlinear and nonstationary signal processing*, chapter 11, pages 317–394. Cambridge Univ. Press, Cambridge, UK, 2001.
- [Tre68] H.L. Van Trees. *Detection, Estimation, and Modulation Theory*, volume Part 1. Wiley and sons, 1968.
- [VGML04] V. Vrabie, P. Granjon, C.-S. Maroni, et B. Leprettre. Application of spectral kurtosis to bearing fault detection in induction motors. In *5th International Conference on Acoustical and Vibratory Surveillance, Methods and Diagnostic Techniques*, Senlis, France, oct 2004.
- [VGS03] V. Vrabie, P. Granjon, et C. Servière. Spectral kurtosis : from definition to application. In *IEEE-EURASIP International Workshop on Nonlinear Signal and Image Processing*, Grado, Italie, jun 2003.
- [XF06] Jun Xiao et Patrick Flandrin. Multitaper Time-Frequency Reassignment. In *Proceedings of EUSIPCO*, Florence, Italy, 2006.
- [XF07] Jun Xiao et Patrick Flandrin. Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement. *IEEE Transactions on Signal Processing*, 55(6) :2851–2860, 2007.
- [YR04] Özgör Yilmaz et Scott Rickard. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, Vol. 52, No. 7, pages 1830–1847, July 2004.
- [YR06] Chunghsin Yeh et Axel Röbel. Adaptive noise level estimation. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 145–148, Montreal, Quebec, Canada, Sept. 18–20, 2006.

Résumé

Les représentations temps-fréquence montrent l'évolution spectrale d'un signal dans le temps. L'objectif de ces travaux de recherche est de proposer deux principes de segmentation du plan temps-fréquence, cherchant à déterminer quelles sont les zones temps-fréquence présentant un intérêt relatif au signal analysé.

La première méthode s'appuie sur une méthode statistique, modélisant le signal analysé par un signal d'intérêt à segmenter perturbé par un bruit blanc gaussien additif de variance inconnue. Le but est de déterminer le support temps-fréquence, ensemble des points sur lesquels l'énergie du signal à segmenter est répartie. Une détection de type Neyman-Pearson à probabilité de fausse alarme fixée permet de détecter les points temps-fréquence contenant du signal, à niveau de bruit connu. L'algorithme proposé est itératif, estimant le niveau de bruit à partir des points non segmentés, ce niveau de bruit servant à détecter de nouveaux points contenant du signal. Un critère basé sur le kurtosis spectral des points non segmentés permet de définir l'arrêt des itérations. Les applications de cette méthode sont illustrées sur des signaux synthétiques et réels, et pour différentes représentations temps-fréquence.

La seconde méthode est basée sur le principe de la réallocation du spectrogramme, non pas comme représentation temps-fréquence réallouée, mais uniquement en tant que source d'information sur le spectrogramme. La réallocation déplace l'énergie du spectrogramme sur le centre de gravité local de l'énergie. Aux frontières d'un motif temps-fréquence, l'énergie sera déplacée vers l'intérieur du motif. Ainsi, les vecteurs de réallocation, décrivant le déplacement de l'énergie du spectrogramme par la réallocation, sont localement parallèles sur la frontière d'un motif. Nous définissons alors un « degré de parallélisme » pour chaque vecteur, égal au nombre de ses vecteurs voisins qui lui sont parallèles. Un algorithme de type suivi de crête, parcourant le plan temps-fréquence en suivant les maxima de ce degré de parallélisme, construit alors un contour entourant le motif temps-fréquence.

Mots-clés : représentation temps-fréquence, segmentation statistique, kurtosis spectral, support temps-fréquence, réallocation, spectrogramme, frontière d'un motif temps-fréquence, contour.

Abstract

Time-frequency representations show the spectral evolution of a signal in time. The goals of this work is to propose two segmentation principles of time-frequency plane, trying to determine the time-frequency areas which present an interest in relation to the analyzed signal.

The first method is based on a statistical method, taking as a model of the analyzed signal the sum of a signal of interest to segment and a white Gaussian noise of unknown variance. The aim is to determine the time-frequency support, that is all the points on which the energy of the signal to segment is distributed. A Neyman-Pearson detection with a given probability of false alarm can detect the time-frequency points containing signal for a known noise level. The proposed algorithm is iterative, estimating the noise level from non-segmented points, this noise level permitting to detect new points containing signal. A criterion based on the spectral kurtosis of non-segmented points define the moment to stop the iterations. The application of this method are illustrated on synthetic and real signals, and for different time-frequency representations.

The second method is based on the principle of the reassignment of the spectrogram, not as a reassigned time-frequency representation, but only as a source of information on the spectrogram. The reassignment shifts the energy spectrogram to the local center of gravity of the energy. On the boundary of a time-frequency pattern, energy will be moved inside the pattern. Thus, the reassignment vectors describing the displacement of the energy spectrogram by the reassignment are locally parallel on the boundary of a pattern. We then define a "parallelism degree" for each vector, being the number of its neighboring vectors which are parallel to it. A tracking algorithm searching for the maxima of the parallelism degree along the time-frequency plane built finally a closed contour encircling the time-frequency pattern.

Keywords : time-frequency representation, statistical segmentation, spectral kurtosis, time-frequency support, reassignment, spectrogram, time-frequency pattern boundary, contour.
