Temps-fréquence et décision — Une introduction

Patrick Flandrin

1 Introduction

C'est une évidence et une banalité de dire que le monde qui nous entoure est "non stationnaire". Si s'offrent bien sûr à notre observation des phénomènes empreints d'une grande régularité (du moins rapportés à une échelle temporelle humaine, comme par exemple le mouvement des astres), ce par quoi le monde naît à notre perception passe avant tout par le changement, la variation, la différence. D'une façon tout à fait fondamentale, l'information (dans son sens aussi bien commun que technique, "à la Shannon") est indissociable de l'imprévu et donc de la "non-stationnarité", en tant qu'idée générale de non-homogénéité spatio-temporelle. De plus, si l'on admet que la nature est en soi non stationnaire, il apparaît aussi qu'une capacité de suivre ses non-stationnarités est attachée à nos sens — outils naturels permettant de l'appréhender —, et de cette conjonction résulte la possibilité même (étymologique) d'une connaissance. Observateurs impassibles d'un univers immuable, nous serions confrontés, non seulement à un indicible ennui, mais encore à l'incapacité même d'en dire quelque chose, voire d'en percevoir l'existence.

Au vu de ce préambule très général, il semblerait donc raisonnable que les signaux, supports physiques de toute information, soient traités par des procédures reposant à la base sur des pré-supposés de non-stationnarités, c'est-à-dire à l'aide d'outils offrant un langage naturel pour la description de ces dernières. Or, force est de constater que, si l'expérience quotidienne va bien dans ce sens (quoique d'une façon essentiellement perceptuelle), ce n'est que depuis un passé assez récent que de tels outils ont été forgés pour appréhender des classes très larges de non-stationnarités, dans lesquelles l'évolution temporelle de propriétés spectrales joue un rôle prédominant. Nul besoin pourtant de longues explications pour justifier du bien-fondé de la notation musicale, de l'évidence de la reconnaissance d'un instrument de musique (ou d'un locuteur) à son timbre, ou encore de l'à-propos du diagnostic acoustique d'un moteur par l'oreille experte d'un garagiste. Il a en fait fallu attendre les années quatre-vingts pour voir émerger (et se faire accepter) un paradigme nouveau, selon lequel c'est dans un plan temps-fréquence qu'il convient de décrire et de manipuler des signaux non stationnaires. Si un des éléments décisifs de ce changement a certainement été la "révolution des ondelettes", le véritable basculement paradigmatique va au-delà d'une technique particulière, et les approches qui ont pu être développées s'appuient sur tout un arsenal de techniques introduites précédemment à des fins essentielles d'analyse, sans nécessaire souci de traitement proprement dit.

Comme cela peut être le cas dans des tâches d'analyse ou de compression, changer d'espace de représentation pour prendre une décision quant à un signal (détection, estimation, classification, reconnaissance) répond au souci de disposer de l'information que le signal recèle, sous la forme la plus propice à son extraction. Lorsqu'il s'agit d'analyse, la question est de trouver une représentation "adaptée" au signal, en accord avec son interprétation physique et sa description phénoménologique. Dans le cas de la compression, l'objectif est davantage de "concentrer" l'information sur un petit nombre de coefficients, mais sans que ceux-ci se prêtent nécessairement à une interprétation particulière. Si c'est une décision qui est visée, l'idée est souvent double, reposant pour part sur l'existence d'une "signature" dans le plan (et donc sur la nécessité de choisir une représentation qui soit à même de mettre en évidence de façon simple une telle signature), et pour part sur une procédure de comparaison avec une (ou des) référence(s), opérant dans le plan.

La théorie de la décision étant un sujet classique et largement couvert dans la littérature [54, 64, 66], il faut bien sûr s'interroger sur la nécessité qu'il peut y avoir de recourir à des approches temps-fréquence dans un tel contexte. D'une manière très générale, la mise en comparaison de stratégies temps-fréquence avec des approches usuelles (opérant en temps et/ou en fréquence) peut se symboliser de la façon suivante. Soit $x(t;\theta)$ un signal connu (à un vecteur de paramètres θ près) et $r(t) := x(t) + b(t)$ une observation bruitée. Sur la base de l'observation disponible $r(t)$ et de connaissances statistiques a priori $\mathcal{B}$ quant au bruit $b(t)$, divers critères (maximum de vraisemblance, maximum de contraste, stratégie bayésienne, de Neyman-Pearson. . .) permettent alors de construire un récepteur optimal dont la sortie $\Lambda_\theta(r|x,\mathcal{B})$ est une statistique permettant la meilleure prise de décision possible au sens du critère d'optimalité choisi. Élargir la démarche à un cadre temps-fréquence consiste à considérer le diagramme

$$\begin{tikzcd}
\{r(t), x(t;\theta)\} \arrow[r, mapsto] \arrow[dd, "?"] & \Lambda_\theta(r|x,\mathcal{B}) \arrow[dd, leftrightarrow, "?"] & \\
 & & \mathcal{B} \arrow[ul] \arrow[dl] \\
\{\rho_r(t,f), \rho_x(t,f;\theta)\} \arrow[r, mapsto, "?"] & \tilde{\Lambda}_\theta(\rho_r|\rho_x,\mathcal{B}) &
\end{tikzcd}$$

dans lequel $\rho_x(t,f;\theta)$ et $\rho_r(t,f)$ sont les images temps-fréquence respectives de $x(t;\theta)$ et $r(t)$, et où $\tilde{\Lambda}_\theta(\rho_r|\rho_x,\mathcal{B})$ est la statistique de sortie d'un récepteur opérant directement dans le plan.

Un tel diagramme pose évidemment plusieurs questions relatives aux nombreux degrés de liberté mis en jeu : quelle représentation temps-fréquence ρ choisir ? quel critère retenir pour la construction de la statistique $\tilde{\Lambda}$? quel lien entre $\tilde{\Lambda}$ et Λ ? Il met aussi en évidence qu'au moins deux approches sont possibles pour aborder ces questions : la première consiste à prendre le chemin $r \mapsto \Lambda \to \tilde{\Lambda}$, c'est-à -dire à partir de stratégies usuelles et à en donner des formulations temps-fréquence associées ; la deuxième, qui emprunte cette fois la route $r \to \rho_r \mapsto \tilde{\Lambda}$, est plus radicale dans la mesure où elle prend son point de départ dans le plan et cherche à y ancrer sa stratégie. Dans le premier cas, il est clair que la recherche de la commutativité du diagramme n'est pas une fin en soi. C'est bien davantage un point de départ, destiné avant tout à garantir l'existence de formulations alternatives de stratégies optimales et à fournir une base pour des modifications éventuelles. Dans le deuxième cas, on peut espérer qu'une approche opérant directement dans le plan permette, grâce à une adéquation entre la nature des données et l'espace choisi pour leur représentation, d'aborder des problèmes difficiles à formuler dans un espace monodimensionnel, et d'atteindre des propriétés d'optimalité de façon simple. La situation est en fait assez analogue à celle qui prévaut en analyse temps-fréquence. En effet, dans le cas de représentations "continues", l'idée même de passer d'une description monodimensionnelle à une représentation bidimensionnelle peut paraître créatrice d'une augmentation de redondance inutile. Il n'en est cependant pas nécessairement ainsi, comme le montre l'exemple simple d'un "chirp" linéaire, dont la paramétrisation temps-fréquence peut s'avérer plus économique que la donnée des échantillons du signal de départ. La raison en est que le passage au plan permet une opération en deux temps : dans un premier temps, l'augmentation potentielle de redondance offre en quelque sorte au signal la possibilité de se déployer de manière structurée ; dans un deuxième temps, cette structuration — lorsqu'elle est correctement identifiée et mise à profit — se traduit alors par une diminution de la redondance effective.

L'objectif de ce Chapitre n'est pas de fournir un panorama exhaustif et détaillé de l'ensemble des solutions possibles au problème de la décision dans le plan temps-fréquence, mais davantage de fournir quelques clés introductives à quelques familles d'approches détaillées plus avant dans ce Volume. La question essentielle qui est abordée ici est celle de la détection (élargie le cas échéant à l'estimation ou la classification) de signaux de forme plus ou moins connue, corrompus par un bruit d'observation additif, le cadre d'étude choisi étant celui des distributions temps-fréquence "à la Cohen", dont on supposera que le lecteur connaît les bases [9, 15, 24, 51]. On passera ainsi sous silence au moins deux grandes catégories de problématiques complémentaires à celles retenues. La première concerne les *objectifs* visés, qui pourraient inclure la détection de ruptures (temporelles et/ou spectrales), telle qu'elle peut être revisitée dans le plan temps-fréquence, point abordé en détail ailleurs [40, 44, 45, 48] et en particulier dans la suite de cet ouvrage [19, 41]. La deuxième catégorie concerne les *méthodes*, qui pourraient inclure les techniques reposant sur des bases de décompositions linéaires (Gabor, ondelettes, paquets d'ondelettes), en particulier pour la détection de transitoires, interprétés comme une forme d'introduction locale de non-stationnarité : là encore, on pourra se reporter à [28, 44, 45, 55] ou [5, 33] pour un traitement explicite de ces questions.

Ce que contient finalement ce Chapitre est organisé de la façon suivante. Dans un premier temps, on discute pourquoi et comment les approches temps-fréquence (au sens large) peuvent offrir un cadre naturel de *ré-écriture*, dans le plan, de stratégies optimales de décision connues par ailleurs. Dans un deuxième temps, l'approche temps-fréquence est justifiée par la capacité qu'elle offre d'aller au-delà de cette simple ré-écriture, en permettant d'*adapter* les schémas définis précédemment à des situations non nominales, ou d'offrir un espace de représentation adéquat à certains problèmes difficiles à traiter en temps ou en fréquence. De manière plus précise, cette partie concerne la *détection à hypothèses composites*, en particulier dans le cas important de la détection de *chirps*, ainsi que des questions de type *robustesse*. Dans un troisième temps, on renverse la perspective en considérant directement le plan temps-fréquence comme l'espace d'observation, l'objectif étant de donner un sens précis à l'idée intuitive de détection par reconnaissance d'une *signature* temps-fréquence. On discute ainsi l'idée de *filtrage adapté temps-fréquence*, la question de comment faire usage d'une base d'*apprentissage* pour piloter le choix (et l'usage qui en est fait) d'une distribution, avec quelques considérations sur les possibilités offertes par les techniques de *reconnaissance de formes* et/ou d'*analyse d'images*.

2 Réécrire

Soit le problème initial de détection binaire :

$$\begin{cases} H_0 & : \quad r(t) = b(t) \\ H_1 & : \quad r(t) = x(t) + b(t) \end{cases}$$

dans lequel il s'agit de détecter un signal d'énergie finie $x(t) \in L^2(\mathbb{R})$, déterministe, parfaitement connu, à partir d'une observation $\{r(t); t \in T\}$ corrompue par un bruit additif $b(t)$.

Dans le cas le plus simple où le bruit $b(t)$ est supposé centré et blanc, de densité spectrale de puissance N_0, le filtre *linéaire* dont la sortie (au temps $t = 0$) maximise le contraste (ou rapport signal-sur-bruit de sortie)

$$D(\Lambda) := \frac{|\mathbb{E}\{\Lambda|H_1\} - \mathbb{E}\{\Lambda|H_0\}|}{(\mathrm{var}\{\Lambda|H_0\})^{1/2}}$$

n'est autre que le filtre adapté, de réponse impulsionnelle $h(t) = x_-(t) := x(-t)$:

$$\Lambda(r) := \langle x_-, r \rangle.$$

C'est là le prototype de l'intuition selon laquelle détecter un signal revient à trouver dans l'observation qui en est faite un degré de ressemblance jugé suffisant avec la référence dont on dispose. Au sens du produit scalaire considéré (celui des fonctions de carré sommable), maximiser le degré de ressemblance se fait en utilisant comme mesure la corrélation, ce qui est équivalent à minimiser la distance quadratique (donc la norme associée) entre observation et référence. La situation qui vient d'être évoquée repose sur une structure linéaire *imposée* mais, sous l'hypothèse plus restrictive de gaussiannité du bruit, le même résultat aurait pu être obtenu par un argument de maximum de vraisemblance. Dans les deux cas, le caractère *linéaire* du détecteur est intimement lié à la nature *déterministe* du signal à détecter.

Les deux approches mentionnées (maximum de vraisemblance et contraste) offrent plusieurs niveaux de généralisations si l'on suppose désormais que le signal à détecter $x(t)$ est lui-même aléatoire [54]. Si l'on considère ainsi que $x(t)$ est gaussien, centré, et noyé dans un bruit $b(t)$ coloré et centré, le détecteur à maximum de vraisemblance prend la forme générale :

$$\Lambda(r) = \langle \mathbf{R}_b^{-1} \mathbf{R}_x (\mathbf{R}_x + \mathbf{R}_b)^{-1} r, r \rangle,$$

où $\mathbf{R}_x$ et $\mathbf{R}_b$ sont les opérateurs de covariance associés au signal et au bruit, respectivement : le détecteur obtenu est dans ce cas une fonction *quadratique* de l'observation. Remarquons que l'on aurait là aussi pu imposer ce caractère quadratique sans recourir à l'hypothèse de gaussiannité, en utilisant le critère de contraste évoqué précédemment. On aurait alors obtenu comme solution la quantité :

$$\Lambda(r) = \langle \mathbf{R}_b^{-1} \mathbf{R}_x \mathbf{R}_b^{-1} r, r \rangle.$$

Là encore, c'est dans les deux cas la nature *aléatoire* du signal à détecter qui fixe le caractère *quadratique* des détecteurs.

De manière peut-être plus explicite, la solution du problème de détection d'un bruit gaussien coloré dans un bruit gaussien blanc peut s'écrire en développant les opérateurs de covariance mis en jeu sur la base de Kahrunen-Loève du processus d'intérêt [64]. Si l'on considère ainsi le signal $x(t)$ à détecter comme étant gaussien de moyenne $m_x(t) \neq 0$, tout en supposant le bruit additif $b(t)$ centré et blanc, la statistique de décision issue du principe de maximum de vraisemblance s'exprime alors comme somme de deux contributions :

$$\Lambda(r) = 2 \sum_n \frac{1}{\lambda_n + N_0} \operatorname{Re}\{\langle m_x, \varphi_n \rangle \langle \varphi_n, r \rangle\} + \frac{1}{N_0} \sum_n \frac{\lambda_n}{\lambda_n + N_0} \, |\langle \varphi_n, r \rangle|^2,$$

expression dans laquelle les λ_n et les $\varphi_n(t)$ sont, respectivement, les valeurs propres et fonctions propres de $\mathbf{R}_x$. La première de ces contributions est une fonction *linéaire* de l'observation : on peut l'interpréter comme relative à la détection de la moyenne $m_x(t)$, considérée comme partie déterministe de $x(t)$. La deuxième contribution, *quadratique*, s'attache quant à elle à la détection de la partie aléatoire de $x(t)$ (fluctuations autour de la valeur moyenne). On obtient ainsi une structure de récepteur linéaire-quadratique, en accord avec les exemples précédents. Plus généralement, le problème gaussien-gaussien complet admet une solution de même nature, que l'on n'explicitera pas par souci de simplicité.

Quoique n'épuisant bien sûr pas tous les cas intéressants en pratique, ces premiers exemples illustrent néanmoins le fait que de larges classes de problèmes de détection puissent admettre comme structure de détection optimale une statistique de décision de la forme :

$$\Lambda(r) \propto \langle h(x, \mathcal{B}), r \rangle + \langle \mathbf{L}(x, \mathcal{B}) r, r \rangle,$$

où h est une fonction et $\mathbf{L}$ un opérateur linéaire, tous deux dépendants du signal à détecter x et de la connaissance a priori $\mathcal{B}$ que l'on peut avoir du bruit additif.

En donner une formulation temps-fréquence revient donc essentiellement à exprimer les produits scalaires mis en jeu sous une forme équivalente opérant dans le plan. On peut remarquer qu'à ce niveau, il n'y a évidemment de gain à attendre d'un tel point de vue alternatif que d'intelligibilité, et non d'information. On y reviendra plus loin.

2.1 Classes générales

Soit $\rho_{x,y}(t, f)$ une distribution temps-fréquence (croisée) [9, 15, 24]. On dira qu'elle est *unitaire* (ou encore qu'elle vérifie la *formule de Moyal*) si elle "conserve" le produit scalaire au sens où l'égalité :

$$\langle x_1, x_2 \rangle_{L^2(\mathbb{R})} \langle x_4, x_3 \rangle_{L^2(\mathbb{R})} = \langle \langle \rho_{x_1,x_3}, \rho_{x_2,x_4} \rangle \rangle_{L^2(\mathbb{R}^2)}$$

est satisfaite pour tous signaux $\{x_i(t) \in L^2(\mathbb{R}); i = 1, \dots 4\}$. Une telle propriété est vérifiée en particulier par la distribution de Wigner-Ville $W_{x,y}(t, f)$, avec la conséquence que l'on a alors (avec la notation simplifiée $W_x := W_{x,x}$) :

$$\langle h, r \rangle = \langle \langle W_h, W_{h,r} \rangle \rangle / \|h\|_2^2$$

et

$$\langle \mathbf{L} r, r \rangle = \langle \langle W_{\mathbf{L}}, W_r \rangle \rangle$$

pour tout opérateur linéaire $\mathbf{L}$ de symbole de Weyl [37] associé $W_{\mathbf{L}}(t, f)$.

Quoique le choix de la distribution de Wigner-Ville ne soit pas unique (dans la classe de Cohen, par exemple, toute distribution de noyau unimodulaire (dans sa représentation dans le plan des ambiguïtés) convient également [24], tout comme la distribution unitaire de Bertrand dans la classe affine [8, 60]), on pourra convenir de le retenir pour des raisons de simplicité. Ce faisant, il devient dès lors possible de reformuler les statistiques de détection linéaires-quadratiques évoquées plus haut sous la forme :

$$\tilde{\Lambda} \propto \langle \langle W_{h(x,\mathcal{B})}, W_{h(x,\mathcal{B}),r} \rangle \rangle / \|h(x, \mathcal{B})\|_2^2 + \langle \langle W_{\mathbf{L}(x,\mathcal{B})}, W_r \rangle \rangle,$$

mettant en évidence le fait que détecter un signal (non stationnaire) peut se faire en comparant la "signature" temps-fréquence de son observation (telle qu'elle est donnée par la distribution de Wigner-Ville) avec une signature de référence associée au signal à détecter et au bruit qui le corrompt.

Deux exemples — Deux exemples simples permettent d'illustrer ce point de vue alternatif : celui du *canal de Rayleigh* et celui de la détection *localement optimale* [22].

Dans le premier cas, on suppose que $x(t) := a\, x_d(t)$, où $x_d(t)$ est une forme d'onde déterministe et a une variable aléatoire gaussienne, centrée. Le détecteur se réduit alors à sa partie quadratique et s'écrit simplement $\tilde{\Lambda}(W_r) \propto \langle\langle W_{x_d}, W_r \rangle\rangle$.

Dans le deuxième cas, on suppose que le signal à détecter est aléatoire, centré, et que l'observation est à faible rapport signal-sur-bruit, c'est-à-dire que les valeurs propres de la covariance sont telles que $\lambda_n \ll N_0$ pour tout n. On en déduit que

$$\sum_n \frac{\lambda_n}{\lambda_n + N_0} \, |\langle \varphi_n, r \rangle|^2 \approx \frac{1}{N_0} \sum_n \lambda_n \, |\langle \varphi_n, r \rangle|^2$$

d'où, par unitarité, $\tilde{\Lambda}(W_r) \propto \langle\langle \mathbb{E}W_x, W_r \rangle\rangle$, car $\sum_n \lambda_n \, W_{\varphi_n}(t,f) = \mathbb{E}W_x(t,f)$.

Dans les deux cas, la statistique de détection est construite comme produit scalaire entre la signature de l'observation (donnée par sa distribution de Wigner-Ville, ou par toute distribution unitaire) et une signature de référence (distribution analogue pour un signal déterministe ou valeur moyenne des distributions de ses réalisations pour un signal aléatoire).

Une parenthèse historique — L'idée de détecter ou reconnaître un signal par comparaison de signatures temps-fréquence est bien sûr ancienne. Tout comme pour l'analyse temps-fréquence — qui peine à trouver une paternité à la transformation de Fourier à court-terme — des approches de cette nature ont pu être utilisées de façon intuitive bien avant qu'une formulation précise en soit donnée. Parmi les premiers travaux "explicites" sur le thème, on peut citer [26] et [2], qui tous deux identifiaient distributions temps-fréquence et spectrogrammes, conduisant nécessairement à des solutions sous-optimales. Un point très important de ces travaux était néanmoins d'être guidés par le souci d'offrir une description "naturelle", qui soit proche des mécanismes d'audition chez les mammifères (et en particulier de la chauve-souris dans les travaux de R.A. Altes). Le lien avec les stratégies optimales usuelles, via l'unitarité et le recours essentiellement à la distribution de Wigner-Ville, suit des travaux fondateurs [36, 39], poursuivis dans [22], qui sont à la base de la plupart des développements ultérieurs [20, 27, 46, 49, 52, 61, 62].

3 Adapter

Les situations évoquées dans la Section précédente permettent de justifier qu'une approche temps-fréquence puisse offrir une formulation différente, mais toujours *équivalente*, à des structures optimales connues par ailleurs. Le premier intérêt de cette re-formulation est de se prêter à une interprétation physique simple et sans doute plus naturelle, car mettant explicitement en jeu la signature temps-fréquence des non-stationnarités à détecter. Le deuxième intérêt est que ce nouveau point de vue offre en fait un point de départ adéquat pour des variations relativement aux situations nominales connues. Là encore, c'est de l'association entre structure mathématique et interprétation physique que peut naître la possibilité de telles variations.

3.1 Hypothèses composites

La première façon d'élargir le cadre précédent — dont on pourra trouver une présentation complète dans [61, 62] — consiste à considérer le cas d'hypothèses composites :

$$\begin{cases} H_0 & : \quad r(t) = b(t) \\ H_1 & : \quad r(t) = x(t;\theta) + b(t) \end{cases}$$

dans lequel le signal à détecter dépend d'un vecteur de paramètres $\theta \in \mathbf{\Theta}$, inconnu ou aléatoire de densité de probabilité $p(\theta)$.

Dans le premier cas, on peut recourir à une approche de maximum de vraisemblance généralisé :

$$\Lambda(r) \to \max_{\theta \in \mathbf{\Theta}} \Lambda_\theta(r)$$

et, dans le second, à une statistique de décision pondérée par la connaissance a priori relative au paramètre perturbateur :

$$\Lambda(r) \to \int_{\mathbf{\Theta}} \Lambda_\theta(r)\, p(\theta)\, d\theta.$$

Il en est ainsi si l'on considère par exemple que $x(t;\theta) := x_d(t)\ \exp\{i\theta\}$, avec $\{p(\theta) = 1/2\pi; \mathbf{\Theta} = [-\pi, \pi)\}$, c'est-à-dire que le signal recherché est déterministe, mais connu à une phase uniforme près : on montre alors [66] que le détecteur devient $\Lambda(r) \propto |\langle x_d, r\rangle|^2$, ce qui ramène à la classe des détecteurs quadratiques.

Dans des cas plus généraux de paramètres inconnus, il est souvent difficile d'exprimer et d'interpréter le détecteur résultant d'un lissage par la densité de probabilité, alors que l'approche temps-fréquence peut se révêler bien davantage informative. En effet, la “signature” temps-fréquence se transforme dans une telle situation selon :

$$W_{\mathbf{L}(x,\mathcal{B})}(t,f) \to \int_{\mathbf{\Theta}} W_{\mathbf{L}(x(\theta),\mathcal{B})}(t,f)\, p(\theta)\, d\theta,$$

soit essentiellement un “épaississement” de la signature nominale. De façon exacte [22], si l'incertitude observée est induite par une *gigue* en temps et en fréquence, c'est-à-dire si $\theta := (\tau, \xi)$, le détecteur devient $\langle\langle C_{\mathbf{L}(x,\mathcal{B})}(p), W_r\rangle\rangle$, expression dans laquelle $C_{\mathbf{L}}(t,f;p)$ est, mutatis mutandis, l'analogue du symbole de Weyl, mais construit sur la distribution de la classe de Cohen de noyau de lissage $p(t, f)$ (dans le plan temps-fréquence), en lieu et place de la distribution de Wigner-Ville de noyau $\delta(t)\,\delta(f)$.

Il est intéressant de noter que, suivant la façon dont les incertitudes en temps et fréquence peuvent varier, ce cadre formel englobe une variété de détecteurs a priori très différents. En effet, si l'on suppose pour simplifier la présentation que $C_{\mathbf{L}(x,\mathcal{B})}(t, f; p) \propto C_{x_d}(t, f; p)$, il est facile de se convaincre (par raison de convolution, et éventuellement à une symétrie près) que l'on peut permuter les rôles dévolus à la référence et à l'observation, et que la structure générale de décision peut en fait s'écrire :

$$\tilde{\Lambda}(C_r) \propto \langle\langle W_{x_d}, C_r(p)\rangle\rangle,$$

soit la stratégie initiale de comparaison avec la signature nominale, mais appliquée à une version *lissée* de la distribution de Wigner-Ville de l'observation, le lissage s'identifiant exactement au degré d'incertitude sur la localisation de la référence [22, 24]. Prendre en compte cette incertitude permet donc d'obtenir un détecteur qui se situe "entre" les cas limites répertoriés dans le tableau suivant :

$p(t,f)$	$\Lambda(r)$	type de détecteur
$\delta(t)\,\delta(f)$	$\lvert\langle x_d, r\rangle\rvert^2$	filtre adapté + détecteur d'enveloppe
$\delta(t)$	$\langle \lvert x_d\rvert^2, \lvert r\rvert^2\rangle$	corrélateur d'intensité
$\delta(f)$	$\langle \lvert X_d\rvert^2, \lvert R\rvert^2\rangle$	corrélateur de densité spectrale
1	$\lVert x_d\rVert_2^2 . \lVert r\rVert_2^2$	détecteur d'énergie

On voit ainsi que la formulation temps-fréquence offre une grande flexibilité pour passer *continûment* d'un récepteur semi-cohérent (filtre adapté suivi d'un détecteur d'enveloppe) à des récepteurs totalement incohérents (détecteur d'énergie, corrélateurs d'intensités), tout en restant à l'intérieur d'une seule et même structure unifiée.

Que ces variations soient construites in fine autour de la classe de Cohen trouve son origine dans le fait que cette dernière repose sur un principe de covariance par les translations en temps et en fréquence. Si le vecteur de paramètres θ avait été choisi différemment, des résulats analogues auraient été néanmoins obtenus, moyennant que l'on transpose à des classes de distributions covariantes par rapport au groupe de transformations induit par le nouveau paramètre [62]. C'est en particulier le cas du couple "translation en temps + changement d'échelle", qui donne naissance à des détecteurs temps-échelle basés sur la classe affine [61].

3.2 Chirps

Une catégorie particulièrement importante de signaux non stationnaires pour lesquels l'approche temps-fréquence est pertinente est celle des "chirps", c'est-à-dire des signaux de la forme $x(t) = a_x(t)\,\exp\{i\varphi_x(t)\}$, pour lesquels on admet que la variation de l'amplitude $a_x(t) \geq 0$ est suffisamment lente à l'échelle des oscillations induites par la phase $\varphi_x(t)$ pour que la dérivée temporelle de cette dernière puisse s'interpréter comme une fréquence instantanée $f_x(t)$. L'idée de base est que, dans ce cas, la signature temps-fréquence du chirp doit être "filaire", c'est-à-dire se réduire essentiellement à une contribution non nulle dont le support dans le plan s'identifie à la trajectoire de fréquence instantanée. Détecter un chirp revient alors à rechercher une contribution cohérente et localisée le long de cette trajectoire.

Pour que cette intuition prenne tout son sens, il faut bien sûr qu'il y ait adéquation entre le type de chirp envisagé et la distribution temps-fréquence susceptible d'en assurer la localisation. Si tel est le cas, c'est-à-dire s'il existe une distribution unitaire ρ telle que $\rho_x(t,f) = A(t)\,\delta(f - f_x(t))$, on peut écrire :

$$\langle\langle \rho_x, \rho_r \rangle\rangle = \int A(t)\,\rho_r(t, f_x(t))\,dt,$$

d'où un détecteur qui se réduit à une intégration de chemin le long de la trajectoire de fréquence instantanée.

Il en est ainsi pour les chirps unimodulaires ($|a_x(t)| = 1$) de phase quadratique $\varphi(t) := 2\pi(\alpha t^2/2 + f_0 t + \psi)$, lorsqu'on les analyse par la distribution de Wigner-Ville, puisque l'on sait qu'alors $W_x(t,f) = \delta(f - (f_0 + \alpha t))$. Par raison d'équivalence unitaire, il est facile d'obtenir un résultat identique pour les chirps hyperboliques [52] en faisant usage de la distribution d'Altes-Marinovic [9, 24, 53] en lieu et place de celle de Wigner-Ville. Moyennant quelques aménagements, le même argument s'applique au couple "chirps en lois de puissance + distribution unitaire de Bertrand" [11, 12] et peut, suivant les besoins, être étendu à d'autres distributions (comme les spectrogrammes réalloués [11, 12] ou les arêtes de scalogrammes [10]) ou à d'autres formes spécifiques de lois de modulation. Ce problème de la détection temps-fréquence de chirps est détaillé plus avant dans [13].

3.3 Robustesse

La détection de chirps offre un exemple simple de situation dans laquelle une approche temps-fréquence offre un cadre conceptuel particulièrement bien adapté pour modifier un récepteur optimal lorsque l'on s'écarte des conditions nominales pour lesquelles il a été conçu.

Tolérance à l'effet Doppler — Considérons dans un premier temps un chirp de temps d'arrivée inconnu, $x(t) := x_0(t - t_0)$, et une distribution temps-fréquence ρ, covariante par les translations temporelles. La stratégie d'intégration de chemin peut dans ce cas permettre d'aborder le problème conjoint de la détection de $x(t)$ et de l'estimation de t_0, selon :

$$\hat{t_0} = \arg\max_{\tau} \int A(t)\, \rho_r(t+\tau, f_{x_0}(t))\, dt.$$

Dans cette perspective, la formulation temps-fréquence de l'estimation optimale d'un retard permet une interprétation particulièrement simple lorsque l'observation est affectée d'effet Doppler (large bande). On sait en effet que, dans le cas général, les estimations de taux Doppler et de retard ne sont pas découplées, ce qui conduit à un biais sur l'estimation de ce dernier si le Doppler est inconnu [64, 66]. Par suite, la question se pose naturellement de concevoir des signaux qui soient tolérants à l'effet Doppler, au sens où un taux quelconque de Doppler ne biaise pas l'estimation du retard. De tels signaux existent et sont caractérisés par une fréquence instantanée (ou un retard de groupe) hyperbolique [3, 59] : l'approche temps-fréquence permet d'en donner une justification géométrique très simple [21, 24, 25]. En effet, un effet Doppler de taux η correspond à la transformation du plan définie par $(t, f) \to (\eta t, f/\eta)$. Il s'ensuit qu'une estimation non biaisée du retard est alors possible en utilisant la stratégie mentionnée plus haut (intégration de chemin) à condition que la distribution utilisée soit (i) covariante par translations, (ii) unitaire (pour assurer l'optimalité) et (iii) localisée sur la courbe laissée invariante par la transformation Doppler. Cette courbe étant une hyperbole, les trois conditions requises sont satisfaites univoquement par la distribution unitaire de Bertrand, grâce laquelle le problème de la tolérance à l'effet Doppler reçoit une solution temps-fréquence purement géométrique.

Gabarit — L'intégration de chemin est une façon intuitive de mettre en œuvre un filtre adapté, le degré de ressemblance entre observation et référence étant mesuré par le plus ou moins grand niveau de recouvrement des trajectoires associées. Dans le cas idéalisé d'une localisation parfaite, on voit que la réponse d'un tel détecteur est "en tout ou rien", le support d'intersection de deux courbes étant de mesure nulle dès que les courbes ne sont pas exactement superposées, et infini lorsqu'elles le sont. Sans aller jusqu'à cette situation idéalisée, le point de vue temps-fréquence illustre de façon claire la chute de performance du filtre adapté lorsque la référence est mal connue ou lorsque le modèle d'observation s'écarte des conditions nominales prévues. Rendre *robuste* un détecteur à de tels écarts est donc souhaitable, ce qui n'est pas facilement formalisable au sens des stratégies usuelles (opérant en temps ou en fréquence) mais s'envisage de façon très naturelle dans le plan.

Considèrons en effet, pour fixer les idées, le cas d'une phase quadratique $\varphi(t) = 2\pi(\alpha(t-\tau)^2/2 + \xi(t-\tau) + \psi)$, avec $\alpha = \alpha_m \pm \delta\alpha$, $\tau = \tau_m \pm \delta\tau$, ou encore $\xi = \xi_m \pm \delta\xi$, où $\delta\alpha \in \Delta\alpha$, $\delta\tau \in \Delta\tau$ et $\delta\xi \in \Delta\xi$ mesurent des écarts possibles au modèle nominal $(\alpha_m, \tau_m, \xi_m)$. Balayant l'ensemble des valeurs possibles de $(\delta\alpha, \delta\tau, \delta\xi) \in \Delta\alpha \times \Delta\tau \times \Delta\xi$, on obtient en fait pour la fréquence instantanée $\dot{\varphi}(t)/2\pi$ un faisceau de droites possibles pour la référence, en lieu et place de la droite nominale unique $f = \alpha t + \xi$. Assurer la détection revient donc à remplacer le chemin d'intégration nominal par le domaine défini par l'ensemble des chemins possibles dans le plan. Ce faisant, on "élargit" en fait la référence d'une quantité représentative du degré d'incertitude que l'on a sur les paramètres mis en jeu, ce qui ramène, au moins conceptuellement, au cas des hypothèses composites et aux détecteurs reposant sur des distributions lissées.

Minimax — D'une manière plus générale, il est raisonnable de penser que l'approche temps-fréquence puisse offrir un cadre adéquat pour rendre des détecteurs véritablement *robustes*, au sens par exemple d'un critère minimax [35].

Quoique la théorie n'en soit encore qu'ébauchée, on peut, pour en soutenir l'intuition, reprendre le cas simple étudié en [23]. Le problème considéré est celui de la détection (dans un bruit blanc, gaussien, centré, de densité spectrale de puissance N_0) d'un transitoire $x(t) := a \, \exp\{-\lambda t + i2\pi f_0 t\} \, \mathbf{1}_{[0,+\infty)}(t)$, d'amplitude $a \in \mathcal{N}(0, \sigma^2)$ aléatoire, de fréquence centrale f_0 connue mais de facteur d'amortissement $\alpha > 0$ inconnu, celui-ci étant distribué selon une loi de densité $p(\alpha)$ au voisinage d'une valeur nominale α_0. Utilisant pour la détection une statistique simplifiée d'intégration sur la droite $f = f_0$ (stratégie extensible aux chirps linéaires), on obtient :

$$\tilde{\Lambda}(W_r) = \int_0^{+\infty} P(t) \, W_r(t, f_0) \, dt; \qquad P(t) := \int_0^{+\infty} p(\alpha) \, \exp\{-2\alpha t\} \, d\alpha.$$

Evaluant le contraste $D(\tilde{\Lambda})$ en sortie de ce détecteur, on peut alors rattacher le problème posé à celui considéré dans [38] et montrer que :

$$P_*(t) := \arg \max_P \min_{\alpha \in A} D(\tilde{\Lambda}) = \frac{4\alpha_0 t}{\sqrt{4\alpha_0 \Delta + 4\alpha_0 t}} \, \exp\{-2\alpha_0 t\}$$

si le domaine de variation possible du facteur d'amortissement α est défini par :

$$A := \left\{ \alpha \, \middle| \int_0^{+\infty} t^2 \, \left[\exp\{-2\alpha t\} - \exp\{-2\alpha_0 t\}\right]^2 \, dt \leq \Delta \right\}.$$

Ce faisant, on voit que la prise en compte d'un domaine maximal de variation du paramètre fluctuant définit implicitement la fonction de poids $P_*(t)$ à utiliser de façon optimale dans le détecteur temps-fréquence, celui-ci héritant de fait des propriétés de robustesse de l'approche minimax.

4 Partir du plan

Comme cela a été dit dans l'Introduction, on peut imaginer, plutôt que de construire des détecteurs optimaux opérant sur le signal pour en donner ensuite seulement une formulation temps-fréquence équivalente, de prendre comme espace de représentation de départ le plan temps-fréquence lui-même.

4.1 Filtrage adapté temps-fréquence

En se plaçant dans le cas où il s'agit de détecter un signal déterministe $x(t)$ dans un bruit $b(t)$, formuler le problème dans le plan temps-fréquence revient à écrire [23] :

$$\begin{cases} H_0 & : \quad \rho_r(t,f) = \rho_b(t,f) \\ H_1 & : \quad \rho_r(t,f) = \rho_{x+b}(t,f) \end{cases}$$

Par analogie avec la théorie classique du filtrage adapté, l'idée est alors de construire une statistique de décision basée sur un filtre temps-fréquence $G(t,f)$, selon :

$$\tilde{\Lambda}(\rho_r; G) := \langle\langle G, \rho_r \rangle\rangle,$$

de sorte qu'en soit maximisé le contraste (ou rapport signal-sur-bruit de sortie).

Le problème ainsi posé se heurte à plusieurs degrés de difficultés et d'arbitraire. Il faut d'une part choisir, a priori, quelle (classe de) distribution(s) utiliser, dont il serait souhaitable d'autre part de maîtriser les propriétés statistiques.

Bien que quelques résultats existent en ce qui concerne le deuxième point [**?**], l'état de l'art actuel ne permet pas de faire une théorie aussi aboutie que dans le domaine temporel (ou fréquentiel). On se contentera ici de signaler quelques résultats et quelques pistes, dans le cas particulier de la classe de Cohen et d'un bruit d'observation blanc. Sur la base de telles hypothèses, on peut en fait montrer [23] que

$$\arg\max_G \tilde{\Lambda}(C_r; G) = C_x,$$

ce qui s'accorde avec l'intuition selon laquelle le contraste est maximisé lorsque la réponse du filtre temps-fréquence s'identifie à la distribution du signal à détecter (concept de *filtre adapté temps-fréquence*).

Ayant trouvé la réponse optimale pour une distribution donnée à l'intérieur de la classe de Cohen, il est possible d'envisager un deuxième niveau d'optimisation relativement au noyau $\varphi(\xi,\tau)$. Explicitant le contraste maximum (à φ fixé), on montre alors que

$$\tilde{\Lambda}(C_r(\varphi); C_x(\varphi)) \leq \frac{\|x\|_2^2}{N_0},$$

avec égalité pour toutes les distributions (unitaires) caractérisées par une fonction de paramétrisation telles que $|\varphi(\xi,\tau)| = 1$. Dans de tels cas, le contraste maximal est exactement égal à celui que l'on obtiendrait par un filtrage adapté (classique) suivi d'une détection d'enveloppe. On peut noter que les spectrogrammes se trouvent de fait exclus de cette situation d'optimalité, ce qui rejoint des remarques faites précédemment et explique qu'une détection temps-fréquence à base de spectrogrammes nécessite des procédures auxiliaires de déconvolution[2].

4.2 Apprentissage

Lorsque le signal à détecter ne peut être considéré comme déterministe, on a dit précédemment que divers points de vue (contraste, maximum de vraisemblance) pouvaient conduire à des statistiques de décision de la forme $\tilde{\Lambda}(\rho_r) \propto \langle\langle \mathbb{E}\rho_x, \rho_r \rangle\rangle$, c'est-à-dire à la comparaison d'une distribution relative à l'observation avec la moyenne d'ensemble de celles associées au signal. Les difficultés soulevées par ce point de vue sont alors d'ordre théorique (dispose-t-on d'un modèle pour la statistique des données, et de leurs distributions ?) et pratique (comment estimer la moyenne d'ensemble sur la base d'un nombre souvent réduit de situations pouvant servir à l'apprentissage ?).

Dans le cas de la distribution de Wigner-Ville, une solution à ce problème est de partir d'une structure linéaire imposée, semblable à celle prévalant dans le cas du filtrage adapté temps-fréquence. On montre alors [56] que la maximisation de tout critère (contraste, Fisher. . .) ne faisant usage que de propriétés de premier et second ordres de $\tilde{\Lambda}$ conduit à une solution optimale $G(\alpha)$, qui est paramétrée par un nombre $\alpha \in [0, 1]$ ne dépendant que du critère choisi. L'approche retenue est alors de choisir la valeur α_* qui minimise la probabilité d'erreur du détecteur. Cette optimisation n'est cependant pas satisfaisante à elle seule, car on sait que les performances d'un détecteur construit par apprentissage imposent d'adapter sa complexité à la taille des données disponibles (faible capacité d'apprentissage si la complexité n'est pas assez grande, faible capacité de généralisation si elle l'est trop). En d'autres termes, l'objectif véritable est d'optimiser conjointement le critère de détection et la complexité du détecteur, de telle sorte que la probabilité d'erreur soit minimisée, à taille fixée de la base d'apprentissage. Plusieurs stratégies sont possibles pour atteindre cet objectif [57, 58], dont une consiste, conceptuellement, à tronquer un développement propre de $G(\alpha)$ de telle sorte que les composantes mises à zéro soient préférentiellement celles qui induisent les plus faibles écarts par rapport à $G(\alpha_*)$.

L'idée de piloter une statistique de décision dans le plan par apprentissage est récente et peut se décliner de différentes manières [4]. On peut par exemple, plutôt que fixer le choix de la représentation et optimiser la mesure de performances, préférer fixer cette dernière tout en faisant porter l'optimisation sur le choix de la représentation. Il en est ainsi dans le problème de classification (à deux classes) considéré en [16], où choix est fait d'utiliser un critère de contraste de type Fisher et des distributions de la classe de Cohen. De manière plus précise, le critère choisi consiste à minimiser une distance intra-classes moyenne tout en maximisant la distance inter-classes correspondante, la distance d étant de type Kolmogorov et l'optimisation opérant sur la fonction de paramétrisation $\varphi(\xi, \tau)$ des distributions de Cohen (normalisées) $C_x(t, f; \varphi)$ d'un ensemble d'apprentissage $X_1 \times X_2$:

$$\varphi_* := \arg\max_{\varphi} \frac{\langle d(C_x, \langle C_x \rangle_{x \in X_2}) \rangle_{x \in X_1} + \langle d(C_x, \langle C_x \rangle_{x \in X_1}) \rangle_{x \in X_2}}{\langle d(C_x, \langle C_x \rangle_{x \in X_1}) \rangle_{x \in X_1} + \langle d(C_x, \langle C_x \rangle_{x \in X_2}) \rangle_{x \in X_2}}.$$

Une façon efficace d'assurer cette optimisation est de choisir pour $\varphi(\xi, \tau)$ un modèle gaussien radialement symétrique.

4.3 Reconnaissance de formes

Le dénominateur commun de la plupart des stratégies temps-fréquence de détection étant de trouver dans la représentation d'une observation un degré de ressemblance avec une "signature" du signal à détecter, utilisée comme référence, il est tentant d'aborder le problème en termes de reconnaissance de formes.

D'une manière générale, il s'agit de réduire une représentation temps-fréquence de référence à un certain nombre d'attributs qui puissent ensuite être comparés à ceux extraits de la représentation d'une observation. Il y a nécessairement un assez grand arbitraire dans la sélection de ces attributs et dans la façon de les mettre en relation, rendant difficile une méthodologie unifiée et conditionnant dans une large mesure les solutions proposées aux applications particulières auxquelles elles sont dédiées.

Le premier exemple que l'on peut citer est celui (déjà rencontré) des chirps, pour lesquels la "signature" naturelle est celle d'une trajectoire temps-fréquence, et les attributs associés les paramètres permettant d'en décrire la loi. En termes de filtrage adapté, on a vu précédemment que détecter un chirp pouvait se réduire à une intégration de chemin le loi de la loi supposée du chirp. En termes de reconnaissance de formes, on peut adopter une perspective inversée, qui consiste à considérer chaque point du plan temps-fréquence comme lieu potentiel de passage pour toutes les lois de chirp possibles, puis à attribuer, dans le plan des paramètres (par exemple la pente et l'ordonnée à l'origine de la droite de fréquence instantanée, dans le cas d'un chirp linéaire), un poids correspondant à la valeur locale considérée de la distribution : c'est la transformation de Hough. Appliquer la transformation de Hough dans le plan temps-fréquence a été proposé pour différents types de distributions [6], avec un avantage tout particulier pour les spectrogrammes réalloués [7], lié bien sûr à leur grande capacité de localisation sur des structures filaires.

Dans les cas où les structures filaires ne sont pas adaptées, d'autres paramétrisations de surfaces temps-fréquence sont possibles, via par exemple des énergies, des dates et/ou des fréquences centrales, évaluées dans des fenêtres temps-fréquence locales. Etant donné une base d'apprentissage, la sélection du jeu des paramètres les plus discriminants (au sens d'un critère prescrit) repose alors sur une optimisation du choix de la fenêtre temps-fréquence : on trouvera dans [29] un exemple d'une telle approche associant distribution de Wigner-Ville et critère d'information mutuelle.

On peut enfin remarquer que l'approche par reconnaissance de formes est étroitement liée à l'adoption d'un point de vue "image" relatif aux distributions temps-fréquence [1]. Ce point de vue appelle au moins deux remarques. La première est que, si une distribution temps-fréquence se présente en effet souvent à l'utilisateur sous l'aspect d'une image, celle-ci comporte, par opposition à des images du monde physique, un degré important de structuration, portant trace de la transformation sous-jacente utilisée, qu'il est souvent difficile d'intégrer dans les post-traitements et qui, dans le cas où on l'ignore, limite nécessairement l'efficacité des outils standard du traitement d'images. La deuxième remarque est que, si l'aspect “image” est qualitativement parlant pour une analyse humaine, la lisibilité demandée par cette dernière ne va pas nécessairement de pair avec l'optimalité d'une décision. D'un point de vue élémentaire, il n'est que de songer aux termes interférentiels des distributions de la classe de Cohen (comme par exemple la distribution de Wigner-Ville), qui sont à la fois réputés altérer la lisibilité et connus pour être nécessaires à l'unitarité garantissant l'optimalité. Moyennant la prise en compte des spécificités du problème, considérer une distribution temps-fréquence comme une image est néanmoins possible et ouvre la voie à des procédures de décision basées sur des outils comme la détection de contours, l'extraction de lignes de crête, de partage des eaux ..., qui peuvent alors s'avérer très efficaces d'un point de vue algorithmique.

5 Conclusion

References

[1] Abeysekera, R., Boashash, B. (1991). “Methods of signal classification using the images produced by the Wigner-Ville distribution,” *Pattern Recogn. Lett.* **12**, 717–729.

[2] Altes, R.A. (1980). “Detection, estimation and classification with spectrograms,” *J. Acoust. Soc. Amer.* **67**(4), 1232–1246.

[3] Altes, R.A., Titlebaum, E.L. (1970). “Bat signals as optimally Doppler tolerant waveforms,” *J. Acoust. Soc. Amer.* **48**(2, Part 2), 1014–1020.

[4] Atlas, L., Droppo, J., McLaughlin, J. (1997). “Optimizing time-frequency distributions for automatic classification,” in Proc. SPIE **3162**.

[5] Cf. Chapitre 2.3.

[6] Barbarossa, S. (1995). “Analysis of nonlinear LFM signals by a combined Wigner-Hough transform,” *IEEE Trans. on Signal Proc.* **43**(6).

[7] Barbarossa, S., Lemoine, O. (1996). “Analysis of nonlinear FM signals by pettern recognition of their time-frequency representation,” *IEEE Signal Proc. Lett.* **3**(4), 112–115.

[8] Bertrand, J., Bertrand, P. (1992). “A class of affine Wigner functions with extended covariance properties,” *J. Math. Phys.* **33**(7), 2515–2527.

[9] Boudreaux-Bartels, G.F. (1996). "Mixed time-frequency signal transformations," in *The Transforms and Applications Handbook* (A.D. Poularikas, *ed.*), CRC Press, Boca Raton (FL), 829–885.

[10] Carmona, R., Hwang, H.L., Torrésani, B. (1998). *Practical Time-Frequency Analysis*, Academic Press, San Diego (CA).

[11] Chassande-Mottin, E., Flandrin, P. (1999). "On the time-frequency detection of chirps," *Appl. Comp. Harm. Anal.* **6**(2), 252–281.

[12] Chassande-Mottin, E., Flandrin, P. (1999). "On the time-frequency detection of chirps and its application to gravitational waves," in Proc. 2nd Workshop on Gravitational Wave Data Analysis, Editions Frontires, Paris, 47–52.

[13] Cf. Chapitre 2.4.

[14] Cohen, F.S., Kadambe, S., Boudreaux-Bartels, G.F. (1993). "Tracking of unkown nonstationary chirp signals using unsupervised clustering in the Wigner distribution space," *IEEE Trans. on Signal Proc.* **41**, 2085–3001.

[15] Cohen, L. (1995). *Time-Frequency Analysis*, Prentice-Hall, Englewood Cliffs (NJ).

[16] Davy, M., Doncarli, C. (1998). "Optimal kernels of time-frequency representations for signal classification," in Proc. IEEE-SP Int. Symp. on Time-Freq. and Time-Scale Anal. TFTS-98, Pittsburgh (PA), 585–588.

[17] Cf. Chapitre 4.

[18] Droppo, J., Atlas, L. (1998). "Application of classifier-optimal time-frequency distributions to speech signals," in Proc. IEEE-SP Int. Symp. on Time-Freq. and Time-Scale Anal. TFTS-98, Pittsburgh (PA), 581–584.

[19] Cf. Chapitre 2.1.

[20] Escudié, B. (1992). "Wavelet analysis of asymptotic signals — Part II : A tentative model for bat sonar receiver," in *Wavelets and Applications* (Y. Meyer, *ed.*) **RMA-20**. Masson, Paris (F), 28–37.

[21] Flandrin, P. (1986). "Time-frequency interpretation of matched filtering," in Proc. IEEE-Academia Sinica Workshop on Acoust., Speech and Signal Proc. WASSP-86, Pkin (RPC), 287–290.

[22] Flandrin, P. (1988). "A time-frequency formulation of optimum detection," *IEEE Trans. on Acoust., Speech and Signal Proc.* **36**(9), 1377–1384.

[23] Flandrin, P. (1989). “Signal detection in the time-frequency plane,” in Proc. IFAC Workshop on Adv. Info. Proc. in Auto. Control AIPAC-89, Nancy (F), 131–134.

[24] Flandrin, P. (1998). *Temps-Fréquence* (2ème éd.), Hermès, Paris. Traduction anglaise : Flandrin, P. (1999). *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego (CA).

[25] Flandrin, P., Gonçalvès, P. (1994). “From wavelets to time-scale energy distributions,” in *Recent Advances in Wavelet Analysis* (L.L. Schumaker and G. Webb, *eds.*), Academic Press, New York (NY), 309–334.

[26] Flaska, M.D. (1976). “Cross-correlation of short-time spectral histories,” *J. Acoust. Soc. Amer.* **59**(2), 381–388.

[27] Fowler, M.L., Sibul, L.H. (1991). “A unified formulation for detection using time-frequency and time-scale methods,” in Proc. Asilomar Conf. on Sig., Syst. and Comp., Pacific Grove (CA), 637–642.

[28] Friedlander, B., Porat, B. (1991). “Detection of transient signals by the Gabor representation,” *IEEE Trans. on Acoust., Speech and Signal Proc.* **37**(2), 169–180.

[29] Grall-Maës, E., Beauseroy, P. (1998). “Features extraction for signal classification based on Wigner-Ville distribution and mutual information criterion,” in Proc. IEEE-SP Int. Symp. on Time-Freq. and Time-Scale Anal. TFTS-98, Pittsburgh (PA), 589–592.

[30] Hlawatsch, F. (1998). *Time-Frequency Analysis and Synthesis of Linear Spaces : Time-Frequency Filters, Signal Detection and Estimation and range-Doppler Estimation*. Kluwer, Boston (MA).

[31] Hlawatsch, F., Boudreaux-Bartels, G.F. (1992). “Linear and quadratic time-frequency signal representations,” *IEEE Signal Proc. Mag.* **9**, 21–27.

[32] Hlawatsch, F., Flandrin, P. (1998). “The interference structure of the Wigner distribution and related time-frequency signal representations,” in [51, pp. 59–133].

[33] cf. Chapitre 6.1.

[34] Jones, D.L., Sayeed, A.M. (1995). “Blind quadratic and time-frequency based detectors from training data,” in Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-95, detroit (MI), 1767–1770.

[35] Kassam, S.A., Poor, H.V. (1985). “Robust techniques for signal processing : A survey,” *Proc. IEEE* **73**, 433–481.

[36] Kay, S., Boudreaux-Bartels, G.F. (1985). "On the optimality of the Wigner distribution for detection," in Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-85, Tampa (FL), 27.2.1–27.2.4.

[37] Kozek, W. (1992). "Time-frequency signal processing based on the Wigner-Weyl framework," *Signal Proc.* **29**(1), 77–92.

[38] Kuznetsov, V.P. (1976). "Stable detection when the signal and spectrum of normal noise are inaccurately known," *Telecomm. Radio Eng.* **30:31**, 58–64.

[39] Kumar, B.V.K.V., Carroll, C.W. (1984). "Performance of Wigner distribution function based detection methods," *Opt. Eng.* **23**(6), 732–737.

[40] Laurent, H., Doncarli, C. (1998). "Stationarity index for abrupt changes detection in the time-frequency plane," *IEEE Signal Proc. Lett.* **5**(2), 43–45.

[41] Cf. Chapitre 2.2.

[42] Lemoine, O. (1995). *Détection de signaux non stationnaires par représentation temps-fréquence*, Thèse de Doctorat, Univ. de Nice.

[43] Li, W. (1987). "Wigner distribution method equivalent to dechirp method for detecting a chirp signal," *IEEE Trans. on Acoust., Speech and Signal Proc.* **35**.

[44] Mallat, S. (1997). *A Wavelet Tour of Signal Processing*, Academic Press, San Diego (CA).

[45] Mallat, S., Papanicolaou, G., Zhang, Z. (1998). "Adaptive covariance estimation of locally stationary processes," *Ann. Stat.* **26**.

[46] Marinovich, N.M., Roytman, L.M. (1991). "Signal/noise subspace decomposition for random transient detection," in *SVD and Signal Processing — II : Algorithms, Analysis and Applications* (R.J. Vaccaro, *ed.*). Elsevier, Amsterdam (NL), 391–401.

[47] Cf. Chapitre 5.

[48] Martin, W., Flandrin, P. (1985). "Detection of changes of signal structure by using the Wigner-Ville spectrum," *Signal Proc.* **8**(2), 215–233.

[49] Matz, G., Hlawatsch, F. (1996). "Time-frequency formulation and design of optimal detectors," in Proc. IEEE-SP Int. Symp. on Time-Freq. and Time-Scale Anal. TFTS-96, Paris (F), 213–216.

[50] Matz, G., Hlawatsch, F. (1999). "Minimax robust time-frequency processing," in Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-99, Phœnix (AZ), .

[51] Mecklenbräuker, W.F.G., Hlawatsch, F., *eds.* (1998). *The Wigner Distribution — Theory and Applications in Signal Processing*, Elsevier, Amsterdam (NL).

[52] Papandreou, A., Kay, S.M., Boudreaux-Bartels, G.F. (1994). "The use of hyperbolic time-frequency representations for optimum detection and parameter estimation of hyperbolic chirps," in Proc. IEEE-SP Int. Symp. on Time-Freq. and Time-Scale Anal. TFTS-94, Philadelphia (PA), 369–372.

[53] Papandreou-Suppappola, A., Hlawatsch, F., Boudreaux-Bartels, G.F. (1998). "Quadratic time-frequency representations with scale covariance and generalized time-shift covariance : a unified framework for the affine, hyperbolic, and power classes," *Digital Signal Proc.* **8**, 3–48.

[54] Poor, H.V. (1988). An Introduction to Signal Detection and Estimation. Springer-Verlag, New York (NY).

[55] Ravier, Ph., Amblard, P.-O. (1998). "Denoising using wavelet packets and the kurtosis : application to transient detection," in Proc. IEEE-SP Int. Symp. on Time-Freq. and Time-Scale Anal. TFTS-98, Pittsburgh (PA), 625–628.

[56] Richard, C., Lengellé, R. (1997). "Une nouvelle approche pour la détection linéaire optimale dans le plan temps-fréquence," in Proc. 16ème Coll. GRETSI, Grenoble, 659–662.

[57] Richard, C., Lengellé, R. (1998). "Structural risk minimization for reduced-bias time-frequency based detectors design," in Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-98, Phœnix (AZ), 2397–2400.

[58] Richard, C., Lengellé, R. (1998). "Two algorithms for designing optimal reduced-bias data-driven time-frequency detectors," in Proc. IEEE-SP Int. Symp. on Time-Freq. and Time-Scale Anal. TFTS-98, Pittsburgh (PA), 601–604.

[59] Rihaczek, A.W. (1968). *Principles of High-Resolution Radar*. McGraw-Hill, New York (NY).

[60] Rioul, O., Flandrin, P. (1992). "Time-scale energy distributions: a general class extending wavelet transforms," *IEEE Trans. on Signal Proc.* **40**(7), 1746–1757.

[61] Sayeed, A.M., Jones, D.L. (1995). "Optimal detection using bilinear time-frequency and time-scale representations," *IEEE Trans. on Signal Proc.* **43**(12), 2872–2883. (erratum : *IEEE Trans. on Signal Proc.* **45**(3), 761–762, 1997.)

[62] Sayeed, A.M., Jones, D.L. (1996). "Optimal quadratic detection and estimation using generalized joint signal representations," *IEEE Trans. on Signal Proc.* **44**, 2959–2970.

[63] Senhadji, L., Shamsollahi, M. (2000). "Une application en biomédical : signaux SEEG," ce volume.

[64] Van Trees, H.L. (1968). *Detection, Estimation and Modulation Theory — Part I.* Wiley, New York (NY).

[65] Wang, M., Chan, A.K., Chui, C.K. (1998). "Linear frequency-modulated signal detection using Radon-ambiguity transform," *IEEE Trans. on Signal Proc.* **46**(3), 571–586.

[66] Whalen, A.D. (1971). *Signal detection in Noise.* Academic Press, New York (NY).