

ces dernières rendent cette analyse délicate. Caractériser la normalité reste donc plus pertinent. Nous cherchons également à mettre en place des méthodes non dépendantes de la subdivision en patchs établie. Il est donc indispensable que les approches étudiées fonctionnent sur l'intégralité des patchs sans tenir compte de la répartition des moteurs dans ces derniers.

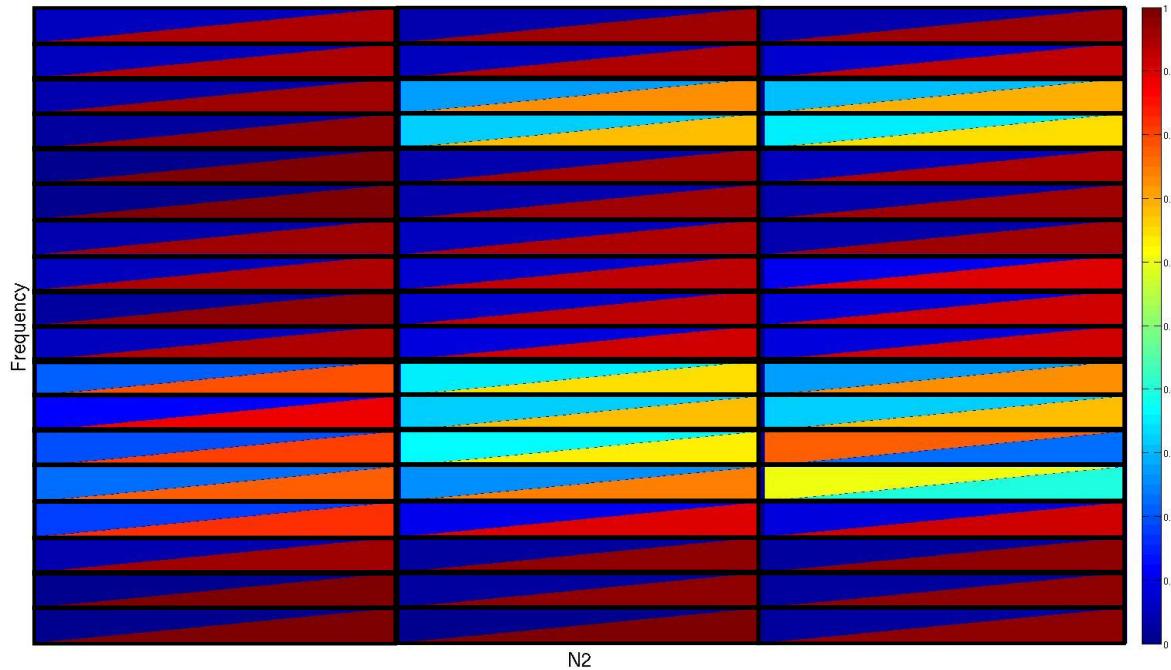


FIGURE 2.2 – Ratio de données normales et atypiques sur les différents patchs allant du bleu (faible proportion) au rouge (forte proportion). Chaque sous-rectangle correspond à un patch avec la partie inférieure droite correspondant au taux de données normales sur ce patch et la partie supérieure gauche aux données atypiques. La grande majorité des patchs possède peu de données atypiques, quelques patchs se distinguent avec une forte proportion de données possédant des signatures inusuelles.

2.4.2 Visualisation des résultats

Dans la thèse, nous présentons des résultats visuels de détection des signatures inusuelles sur un patch spécifique contenant différents types de signatures atypiques. Il s'agit du patch sur lequel quelques points de certains spectrogrammes ont été annotés. Nous avons sélectionné 5 données de ce patch spécifique dans la base de test pour présenter visuellement les résultats, une donnée ne contenant pas de signatures inusuelles et les 4 autres possédant différents formes de signatures atypiques (Figure 2.3). Les approches développées ne sont pas spécifiques à un type de signatures mais caractérisent la normalité présente sur les patchs afin de détecter tout type de signatures inusuelles. Nous représentons sur la figure 2.3 les différents patchs utilisés pour la représentation visuelle des résultats dans ce manuscrit avec un encadrement des signatures inusuelles présentes. Le patch sans encadrement correspond au patch normal.

A partir de nos différentes approches, nous cherchons à détecter les points composant ces

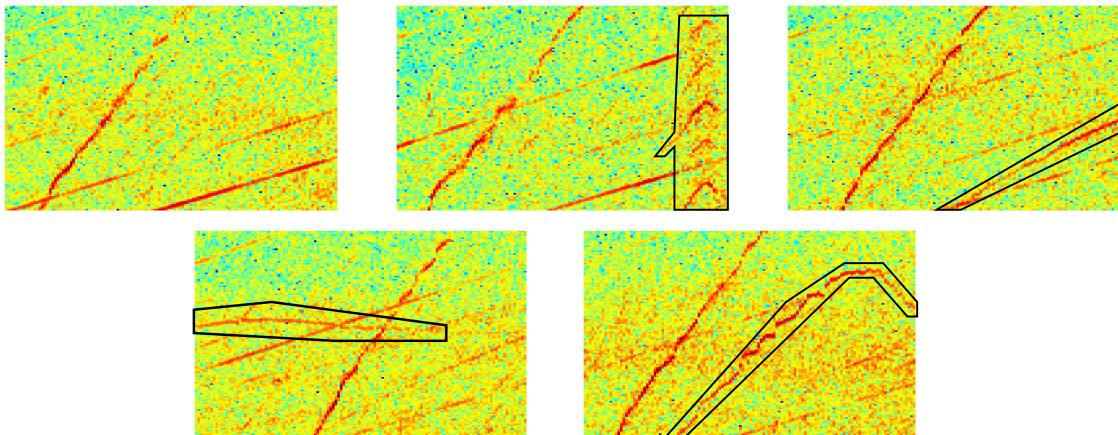


FIGURE 2.3 – Patches de la base de test utilisés pour présenter les résultats visuels des différentes approches. Le premier patch ne contient pas de signature inusuelle, les suivants possèdent tous différentes signatures inusuelles encadrées.

signatures atypiques et ainsi les mettre en évidence pour des détections visuelles. Pour cela, nous modélisons les comportements normaux de chaque patch dans leur globalité et ponctuellement.

Deuxième partie

Les approches de représentation globale par dictionnaire

Introduction

Les approches par dictionnaires permettent de représenter les données en grande dimension dans un nouvel espace. Dans le cadre de la détection de nouveautés, ce dictionnaire doit permettre de caractériser le comportement normal de nos données, c'est-à-dire l'ensemble des raies présentes sur la très grande majorité des spectrogrammes. Cette approche est appliquée patch par patch de manière indépendante. Les dictionnaires sont définis sur chaque patch et n'ont aucune relation avec les dictionnaires définis sur d'autres patchs. Les patchs sont définis à partir de la subdivision \mathcal{K}^{128} dont chaque patch est de dimension 128×128 . Afin de caractériser la normalité, ces dictionnaires sont calibrés sur les patchs normaux sans signature inusuelle. Nous nous servons donc des zones atypiques extraites de la base de données et de la classification des patchs établis afin de sélectionner les patchs d'apprentissage.

Nous cherchons à définir un espace, défini par les atomes du dictionnaire, caractérisant les éléments normaux sur les patchs afin d'y projeter nos données. Nous obtenons ainsi une représentation normale des données à partir de laquelle nous reconstruisons les patchs. Dans ces reconstructions issus du dictionnaire de normalité, les signatures atypiques présentes dans les données se retrouvent absentes ou réduites. Les reconstructions correspondent alors à des estimations normales des patchs. Nous utilisons des dictionnaires dont la reconstruction s'écrit comme une combinaison linéaire des atomes de ces derniers. Ce choix est volontaire et vient de notre interprétation des spectrogrammes comme une superposition de différentes raies pouvant être interprétées comme des sources. Les atomes du dictionnaire sont appris sur les données ou sélectionnés dans un dictionnaire plus large à partir d'un seuillage des coefficients de la décomposition. Nous avons étudié deux de ces dictionnaires :

- les curvelets [22] forment un dictionnaire non-adaptatif (du fait de la non modification des atomes en fonction des données) défini à partir de fonctions. Les atomes de ce dictionnaire s'apparentent à des raies sur de petites échelles et dans différentes orientations. Ce dictionnaire permet donc la caractérisation des signatures vibratoires en combinant différents atomes pour reconstruire la raie.
- La Non-Negative Matrix Factorization (NMF) [70], un dictionnaire adaptatif dont les atomes sont appris à partir des données. Les atomes de ce dictionnaire prennent en compte la structure globale du patch par la procédure d'apprentissage. Ce dictionnaire impose des combinaisons additives de ces atomes uniquement. Les signatures atypiques non présentes dans les données d'apprentissage ne sont pas bien caractérisées par le dictionnaire et ne peuvent pas être reconstruites.

Ces dictionnaires ne sont généralement pas utilisés dans le cadre de la détection d'anomalies. Nous les avons donc adaptés pour qu'ils permettent de répondre à notre problématique. Dans cette partie, nous montrons que ces dictionnaires sont pertinents et complémentaires pour finement détecter les points atypiques sur les patchs.

Chapitre 3

Représentation par dictionnaire fixe - les curvelets

3.1 Introduction

3.1.1 La représentation par dictionnaire

La représentation des données a un rôle primordial dans les performances des méthodes de machine learning [13]. Cette représentation doit avoir un sens aussi bien mathématique qu'explicatif vis-à-vis du problème étudié. Elle doit permettre de réaliser l'étude et de mettre en valeur les éléments pertinents sur les données au niveau de la représentation. Pour la détection dans un cadre supervisé, il est important que la représentation des données soit discriminante, tandis que pour la détection de nouveautés, il est important que la représentation généralise la normalité.

Dans cette partie, nous avons opté pour des dictionnaires reconstruisant les données de manière linéaire. Ce choix a été réalisé car nous considérons les spectrogrammes comme une superposition linéaire de différentes raies pouvant être considérées comme des sources. Les dictionnaires correspondent à un ensemble d'éléments sur lesquels les données sont représentées ; chaque élément du dictionnaire est un atome. Les dictionnaires étudiés permettent de représenter les données comme une combinaison linéaire des atomes de ce dictionnaire. Dans ce cadre, il existe deux moyens de représenter une donnée $x \in \mathbb{R}^p$ à partir d'un dictionnaire $\mathcal{D} = [d_1, \dots, d_r] \in \mathbb{R}^{p \times r}$ [98, 93] :

- la méthode d'analyse où la donnée x est représentée par sa décomposition $C_{\mathcal{D}}^A$ dans le dictionnaire \mathcal{D} , c'est à dire par le produit scalaire de \mathcal{D} avec x ,

$$C_{\mathcal{D}}^A = \mathcal{D}^T x,$$

- la méthode de synthèse où la donnée x est considérée comme une combinaison linéaire des atomes de \mathcal{D} pondérés par la représentation $C_{\mathcal{D}}^S$, la décomposition est obtenue par des

méthodes d'optimisation.

$$x = \mathcal{D}C_{\mathcal{D}}^{\mathcal{S}}$$

Ces deux approches sont identiques dans le cas où le dictionnaire \mathcal{D} est orthogonal. Lorsque $r < p$, la donnée est représentée dans un espace de plus petite dimension, on parle alors de réduction de dimension [35]. Lorsque $r > p$ la représentation est effectuée dans un espace de plus grande dimension, il s'agit alors d'un dictionnaire sur-complet [4]. Ce second cas de figure est utilisé plus particulièrement pour obtenir des représentations parcimonieuses des données.

Les atomes des dictionnaires peuvent être définis de manière analytique par des fonctions, ou sont appris à partir des données. Un état de l'art de ces différentes approches est réalisé dans [97]. Les dictionnaires analytiques possèdent leurs atomes définis par des fonctions prédéfinies. Ce type de dictionnaire est non-adaptatif, car les atomes ne sont pas modifiés en fonction des données, mais possède des propriétés mathématiques intéressantes. Ces dictionnaires donnent une décomposition en général unique et rapide à calculer. Ils sont basés sur la représentation d'un signal complexe à partir de classes de fonctions mathématiques plus simples. Ils sont généralement sélectionnés car leurs atomes correspondent à la structure des données étudiées et sont susceptibles de donner une représentation parcimonieuse de ces dernières. Les dictionnaires *data-driven* apprennent directement leurs atomes sur les données. Les représentations acquises à partir de ces dictionnaires ne sont généralement pas uniques. Afin de rendre ces dictionnaires et les décompositions dans ces derniers plus robustes et d'éviter le surapprentissage, des contraintes sont ajoutées lors de l'apprentissage telles que la parcimonie, la positivité, l'invariance par translation ou par rotation... Ce type de dictionnaire s'est rapidement développé dans la fin du 20ème siècle et le début du 21ème grâce à son adaptabilité aux données et aux méthodes d'optimisation comme les « basis pursuit » [28].

Nous nous intéressons à ces approches de dictionnaire dans un cadre de la détection de nouveautés sur nos spectrogrammes. Nous définissons le modèle de normalité de nos données à partir de ces dictionnaires. Nous ne cherchons pas à extraire des caractéristiques sur les données pour la détection d'anomalies, mais à représenter les éléments normaux des données afin de pouvoir donner une reconstruction normale (sans signature atypique) de ces dernières. Nous définissons la représentation des données par la combinaison du dictionnaire et la décomposition des données dans ce dernier. Dans le cadre de la détection de nouveautés, nous définissons la représentation normale par la reconstruction des données à partir du dictionnaire caractérisant la normalité, il s'agit donc d'une estimation normale des données. Dans ce chapitre, nous étudions les dictionnaires non-adaptatifs à travers les curvelets [22], les dictionnaires adaptatifs sont étudiés dans le chapitre suivant.

3.1.2 Les dictionnaires non-adaptatifs

Ces dictionnaires permettent de représenter les données à partir d'ensemble de fonctions simples. L'exemple le plus connu de ce type de dictionnaire est la transformée de Fourier caractérisant les signaux à partir de fonctions sinusoïdales. Cependant, ne bénéficiant pas de fonction de localisation, cette transformée ne permet pas de caractériser efficacement des signaux discontinus. Des approches multi-échelles tenant compte de la localisation ont permis d'obtenir de meilleurs résultats de caractérisation pour certains types de données. Il s'agit de la transformée en ondelettes [80], fondée sur une famille de fonctions paramétrées par un facteur d'échelle et de position. Cette transformée est inversible et permet de décomposer efficacement des signaux discontinus sans aucun apriori avec peu de coefficients. Autrement dit, la représentation dans ce dictionnaire est généralement parcimonieuse. Dans le cas de la transformée de Fourier, les signaux discontinus influent sur toutes les fréquences et donc sont représentés par un grand nombre de coefficients. Il y a un choix à effectuer par rapport à la nature des fonctions utilisées pour les atomes du dictionnaire. Il existe différentes catégories d'ondelettes comme les ondelettes de Haar, de Daubechies, de Morlet,...[80]. Ces dictionnaires constituent généralement une base de L^2 . La décomposition dans ces dictionnaires se calcule donc généralement à partir d'un produit scalaire entre les données et les atomes, donc par analyse.

La décomposition en ondelettes est pertinente pour caractériser des singularités ponctuelles sur des signaux unidimensionnels ou multidimensionnels. Cependant, dans les dimensions supérieures, les singularités peuvent également correspondre à des hyperplans qui sont alors caractérisés par un grand nombre de coefficients. La transformée en ondelettes ne permet plus une représentation parcimonieuse en grande dimension. Bien qu'elle tienne compte de la position, la transformation en ondelettes classique ne prend pas en compte les orientations et donc la géométrie des singularités. Ceci entraîne la nécessité d'un grand nombre de coefficients afin de caractériser une discontinuité linéaire en 2 dimensions. De nouvelles approches basées sur les ondelettes et la représentation multi-échelle ont vu le jour pour pallier ce problème et principalement caractériser des singularités courbes dans des images, donc en dimension 2. Dans [42], un dictionnaire appelé les wedgelets est mis en place, il consiste en la division de l'image en 4 carrés dyadiques et la définition d'une droite séparatrice pour chacun de ces carrés. Cette droite permet de représenter la partie supérieure du carré par une valeur et la partie inférieure par une autre. Si la droite séparatrice ne permet pas une assez bonne caractérisation du carré dyadique, c'est-à-dire que ce carré est traversé par une forme non linéaire, ce dernier est à nouveau décomposer en 4 sous-carrés dyadiques et chacun d'entre eux est une nouvelle fois divisé par une droite séparatrice. Le processus est itéré jusqu'à avoir une bonne caractérisation des données. Le dictionnaire correspond alors à l'ensemble des carrés dyadiques et des droites séparatrices correspondantes. Il permet de caractériser efficacement des formes courbes en les décomposant en sous-formes linéaires à travers des carrés dyadiques d'échelle de plus en plus fine. Les ridgelets [21] permettent de caractériser des discontinuités linéaires, elles sont basées sur une application des ondelettes le long de droites, les curvelets [22] caractérisent les discontinuités courbes. Plusieurs autres approches permettent de caractériser des discontinuités sur des images comme les

contourlets [39], les bandelets [67]... Le scattering network [18] permet de représenter les données dans un réseau convolutionnel où chaque noeud correspond au module des convolutions successives de transformées d'ondelettes de Morlet. Cette représentation a l'avantage d'être invariante par translation et rotation. Une étape non-linéaire de seuillage est généralement ajoutée à ces représentations non-adaptatives afin de donner une représentation parcimonieuse car un nombre limité de coefficients est généralement suffisant pour caractériser et/ou reconstruire la donnée d'entrée.

Le dictionnaire des curvelets permet de caractériser les formes courbes, nous utilisons donc ce dictionnaire afin de caractériser les raies normales présentes sur les patchs sans signature atypique. La caractérisation des raies normales permet de définir un dictionnaire de normalité dans lequel nous pouvons projeter les patchs pour obtenir une reconstruction normale de ces derniers et étudier les résidus associés afin de vérifier la présence potentielle de signatures inusuelles sur le patch.

3.2 La transformée en curvelet

La transformée en curvelet [22] permet la caractérisation de singularités courbes. Elle est basée sur la transformée en ridgelet [21] et la caractérisation d'une forme courbe à partir d'une succession de formes linéaires à petite échelle (Figure 3.1).

3.2.1 La transformée en ridgelet

Les ridgelets ont été introduites dans [19] et sont définies sous la forme d'une composition de fonctions *ridges* avec une ondelette ψ (3.1). Elles sont caractérisées par un paramètre d'échelle $a > 0$, un paramètre de position $b \in \mathbb{R}$ et un paramètre d'orientation $\theta \in [0, 2\pi[$. Ce nouveau paramètre diffère des ondelettes et permet la caractérisation de toutes formes linéaires sous différentes directions. On note $\psi_{a,b,\theta}$ la ridgelet de paramètre (a, b, θ) .

$$\psi_{a,b,\theta}(t = (t_1, t_2)) = a^{-\frac{1}{2}}\psi\left(\frac{t_1 \cos \theta + t_2 \sin \theta - b}{a}\right) \quad (3.1)$$

Cette fonction est constante le long des droites $t_1 \cos \theta + t_2 \sin \theta = \text{constante}$ et se comporte comme une ondelette perpendiculairement. Dans le domaine des fonctions ridges, une singularité linéaire est définie par un point (la constante à laquelle est associée la droite la caractérisant). Dans cet espace, la singularité linéaire devient donc une singularité ponctuelle pouvant être caractérisée efficacement par les ondelettes avec peu de coefficients. Les ridgelets permettent donc une représentation efficace des formes linéaires.

La transformée en ridgelet \mathcal{R}_x d'un signal x correspond alors au produit scalaire entre la