

FIGURE 1.11 – Extraction du même patch des spectrogrammes de notre base de données avec encadrement des signatures inusuelles. Les types de signatures atypiques sur un même patch sont variés, certaines correspondent à des formes horizontales ou obliques, et d'autres ont des formes particulières. Les signatures inusuelles de même type (par exemple les 2 dernières lignes) sont également variables en intensité, en longueur et en position.

1.6 L'état de l'art de l'analyse vibratoire

1.6.1 L'état de l'art provenant de la littérature

Les vibrations sont étudiées depuis de nombreuses années, particulièrement dans le monde industriel, dans le but de détecter des anomalies ou de caractériser différentes pièces d'une machine [91]. Les différentes méthodes d'analyse vibratoire sont appliquées suivant la nature stationnaire ou non-stationnaire des signaux étudiés. Cette nature influe sur le domaine dans lequel les signaux sont représentés et étudiés [91, 117].

- Les signaux stationnaires
- Le domaine temporel
- Le domaine fréquentiel
- Les signaux non-stationnaires
- La cyclostationnarité
- le domaine temps-fréquence

Selon le domaine sélectionné, la résolution du signal dans ce dernier est définie à l'avance. D'après le principe d'incertitude d'Heisenberg, il est impossible d'obtenir de bonnes résolutions aussi bien dans le domaine fréquentiel que dans le domaine temporel. Il faut donc sélectionner le domaine de l'étude.

Le domaine temporel

Les vibrations correspondent à un signal échantillonné dans le temps, ce signal contient l'information vibratoire accessible depuis le capteur. Dans le cadre des vibrations des machines tournantes, ces signaux sont formés d'une composante périodique et d'une composante aléatoire. La moyenne synchrone est une technique couramment utilisée pour étudier ces signaux [65]. Elle permet d'extraire différentes sources du signal en considérant chaque cycle de la machine comme une réalisation. Dans [65], une moyenne synchrone angulaire est appliquée afin de soustraire du signal les sources déterministes liées au N_1 et au N_2 des moteurs d'avions. Ainsi seules les sources liées à aucun de ces régimes, telles les signatures atypiques, apparaissent dans le signal résiduel. Cependant cette approche ne peut être utilisée qu'en régime stationnaire.

Les modèles autorégressifs, tel que les modèles ARMA (autoregressive moving average), permettent également d'analyser les signaux temporels en détectant des fluctuations liées à un endommagement sur le signal. Ces méthodes peuvent être utilisées pour filtrer le signal vibratoire à partir d'un modèle calibré sur des signaux sans anomalie [115] ou pour donner une estimation sans anomalie de la suite du signal. Des filtres de Kalman ont également été utilisés dans [118] afin d'étudier des signaux vibratoires dans des conditions non stationnaires. Le RMS (Root Mean Square), le kurtosis, ainsi que les différents moments des signaux sont des indicateurs statistiques permettant de caractériser les signaux temporels.

L'Empirical Mode Decomposition (EMD) introduite dans [61] permet de séparer le signal en composantes presque orthogonales. Il s'agit d'une méthode itérative où chaque source correspond à la moyenne entre les enveloppes supérieure et inférieure du signal résiduel (les différentes sources découvertes retranchées du signal). Les sources constituent donc une base calibrée par le signal et non fixée à l'avance, ce qui est très avantageux pour un signal non-stationnaire [73]. Cette méthode possède cependant quelques défauts comme la superposition des modes et le manque de fondements théoriques.

Le domaine spectral

Le domaine spectral est fortement lié à la transformée de Fourier, qui permet de convertir un signal temporel en un signal fréquentiel. Cette transformée est applicable uniquement pour un signal stationnaire, ce qui n'est pas le cas de la majorité des signaux vibratoires issus du milieu industriel. D'autres outils ont donc été mis en place afin d'étudier ces signaux dans le domaine spectral. L'analyse cepstrale [45] permet également d'étudier les signaux vibratoires. Le cepstre consiste en la transformée de Fourier inverse du logarithme du spectre de puissance. Cependant, il est lui aussi applicable uniquement dans un cadre stationnaire.

Le kurtosis spectral [7] est un outil puissant permettant d'indiquer les composantes non-gaussiennes dans le signal et de les localiser dans le domaine fréquentiel. Le kurtosis est calculé pour chaque fréquence afin de déceler la présence de non-stationnarité cachée. Cette méthode

est assez similaire au spectre de puissance avec le moment d'ordre 2 remplacé par le moment d'ordre 4. Le kurtosis spectral est calculé à partir du STFT. Il est donc dépendant des mêmes paramètres que cette dernière. Le kurtogramme [9] permet de pallier ce problème en donnant une représentation du kurtosis spectrale en 2 dimensions de manière analogue au spectrogramme.

La cyclostationnarité

La cyclostationnarité est une méthode courante de caractérisation des signaux étudiée en particulier dans [8]. La cyclostationnarité concerne les signaux ayant une composante déterministe ainsi qu'une composante aléatoire qui est elle-même périodique comme les signaux vibratoires des machines tournantes. Les signaux cyclostationnaires ont donc leurs moments qui sont périodiques, un signal cyclostationnaire d'ordre k signifie que tous ses moments d'ordre 1 à k sont périodiques. La transformée de Fourier est alors appliquée sur les moments du signal afin de mettre en évidence des anomalies dans les systèmes étudiés.

L'enveloppe spectrale [92], qui correspond à une cyclostationnarité d'ordre 2, permet de suivre les réponses hautes fréquences des machines tournantes comme les engrenages et les roulements. Les chocs émis par ces pièces sont très brefs et l'énergie produite est également très faible et répartie sur une grande plage de fréquences. L'analyse spectrale traditionnelle ne permet pas de détecter les possibles défauts dans ces composants.

Le domaine temps-fréquence

La majorité des signaux vibratoires dans les systèmes mécaniques ne sont pas stationnaires par nature. Les représentations temps-fréquence sont donc destinées à étudier ces signaux et permettent une représentation des signaux en 3 dimensions : temps-fréquence-amplitude.

La STFT permet d'obtenir ce type de représentation ; l'hypothèse de stationnarité du signal est faite sur des temps très courts. La transformée de Fourier peut être appliquée sur ces fenêtres de signal. La concaténation des spectres ainsi obtenus produit un spectrogramme. Notre base de données est constituée de spectrogrammes sur lesquels nous cherchons à détecter tout type de signatures atypiques présentes. Cette représentation a déjà été utilisée afin de détecter des défauts à partir des vibrations des systèmes mécaniques comme les engrenages [113, 114, 52].

L'utilisation de la STFT implique une décomposition du signal dans la base de Fourier, donc dans une base sinusoïdale. Les bases d'ondelettes [80] peuvent apporter une représentation parcimonieuse d'un signal dans une base appropriée. Le signal est représenté en fonction du temps et d'un paramètre d'échelle. Les signaux vibratoires peuvent ainsi être représentés dans différentes bases d'ondelettes. Les ondelettes ont déjà été utilisées à de nombreuses reprises pour caractériser et mettre en évidence des anomalies sur les signaux vibratoires [116]. Il est également possible de représenter les signaux vibratoires dans des fonctions définies par des

ondelettes comme les bandelettes [81], curvelets [22], wedgelets [42]. Cela permet de caractériser des formes plus spécifiques et en plus grandes dimensions. La représentation du signal dans une base de Fourier ou d'ondelettes correspond à une représentation dans un dictionnaire dont les atomes sont fixés à l'avance et définis par des fonctions.

La transformation de Wigner-Ville [12] constituent un autre moyen de caractériser les signaux dans un domaine temps-fréquence. Il s'agit d'une opération bilinéaire où la représentation de la somme de signaux ne correspond pas à la somme des représentations mais fait intervenir des termes d'interférence entre le temps et les fréquences.

1.6.2 Les algorithmes d'analyse vibratoire de Safran Aircraft Engines

Algorithme de détection du palier#4

Safran Aircraft Engines a développé des algorithmes spécifiques pour la détection d'anomalies sur les spectrogrammes qui utilisent l'information physique des signatures vibratoires. Les signatures liées à des endommagements de roulements moteur sont les plus problématiques et les plus fréquentes. Ces signatures sont représentées dans les spectrogrammes par des raies dont les équations sont connues. Les raies inusuelles liées au palier#4 (roulement corrotatif) sont représentées par l'équation suivante dans le spectrogramme :

$$f = k_1 N_1 + k_2 N_2 \text{ avec } k_1 = -k_2 \quad (1.3)$$

L'algorithme 4 cherche à détecter spécifiquement les raies correspondant à ce type de signature. La méthode de détection consiste en la suppression des signatures prévisibles et de toutes les intensités faibles par un seuillage. Les coefficients de l'équation 1.3 sont calculés pour chaque paire de points restants. La détection est établie dans cet espace des coefficients en considérant les taux d'apparition des couples (k_1, k_2) . La raie inusuelle liée au palier#4 n'est pas supprimée par le filtrage, ainsi tous les points la composant restent présents et donnent les mêmes valeurs pour le couple (k_1, k_2) . Si les valeurs du couple correspondent à une signature de palier#4, que le taux d'apparition du couple et l'intensité maximale le long de la raie sont suffisamment grands, une alarme est déclenchée et le moteur est suspecté d'être anormal. Cet algorithme permet de bien détecter des anomalies de palier#4, mais aucun autre type d'anomalie. De plus si les anomalies sont d'intensités trop faibles, ces dernières sont supprimées lors de l'étape de prétraitement.

Algorithme 4 : DéTECTeur de signature palier#4

Données : Le spectrogramme S, le seuil s_1 d'apparition du couple, le seuil s_2 d'intensité maximale

Résultat : Indicateur de présence d'une signature atypique de type palier#4
Suppression des intensités faibles, des harmoniques liées aux arbres HP, BP et du RDS.

pour chaque paire de points non supprimés **faire**

- | Calcul du couple (k_1, k_2) de l'équation 1.3.
- | Incrémentation du *compteur* correspondant au couple (k_1, k_2) calculé

fin

si *compteur* $\geq s_1$ et $S(N_2, k_1(N_2 - N_1(N_2))) \geq s_2$ **alors**

- | Déclenchement d'une alarme

fin

Algorithme de détecteur de raies

Cet algorithme [52] estime un masque sur l'intégralité du spectrogramme qui est considéré comme l'ensemble de l'information normale présente sur le spectrogramme. Il est défini à partir d'un ensemble de spectrogrammes supposés normaux. Afin d'obtenir ce masque, les spectrogrammes sont dans un premier temps normalisés à partir de la méthode de Clifton et Tarassenko [33]. Deux masques sont déterminés et soustrait intégralement du spectrogramme. Le premier masque, paramétrique, correspond aux différentes harmoniques entières du N_1 et du N_2 présentes sur le spectrogramme test. Ces harmoniques sont calculées à partir de la transformée de Hough H [10]. Il s'agit d'un opérateur s'appliquant à des fonctions de $\mathbb{R}^2 \mapsto \mathbb{R}$ et permettant de détecter des formes en renseignant la famille des courbes paramétriques auxquelles elles appartiennent. Les harmoniques du N_1 et du N_2 appartiennent à la famille $\{\mathcal{C}_\alpha^j, j \in \{1, 2\}, \alpha \in \mathbb{N}\}$ des courbes linéaires en $N_j, j \in \{1, 2\}$ paramétrées par α

$$\mathcal{C}_\alpha^j = \{(x, \alpha N_j(x)), x \in [\min N_2, \max N_2]\}.$$

La transformée de Hough correspond à l'intégrale le long de la courbe. Dans le cas des spectrogrammes, la somme des intensités des points le long des courbes est prise en compte.

$$H : \mathcal{C}_\alpha^j \rightarrow \int_{\mathcal{C}_\alpha^j} S^i(x, \alpha N_j(x)) dx.$$

Pour toute valeur de α dont la transformée de Hough est supérieure à un seuil fixé arbitrairement, l'intégralité de la courbe correspondante est supprimée du spectrogramme.

Le second masque, statistique, correspond à l'apprentissage de la présence ou non d'information vibratoire sur des patchs du spectrogramme de petites tailles. Il s'agit d'imagettes de taille 5×3 à partir de méthodes d'analyse discriminante ou de SVM [49]. Le classifieur appris est appliqué sur l'ensemble des imagettes des spectrogrammes utilisés pour la calibration du masque. Ce dernier est binaire et alloue la valeur 1 à l'ensemble des points dont les imagettes possédant

ce point en leurs centres ont été classifiées plusieurs fois comme contenant de l'information vibratoire pertinente dans la base de données. Ce masque est également soustrait du spectrogramme. L'anomalie est déclarée par rapport au nombre de points restants après la suppression par les 2 masques.

Cette méthode présente les avantages de tenir compte du décalage des raies N_1 par le masque paramétrique, de pouvoir considérer toutes les raies présentes dans le spectrogramme et d'être établie comme une approche de détection de nouveautés ne nécessitant que des spectrogrammes normaux pour la calibration du modèle. Cependant, elle ne tient pas du tout compte de l'intensité vibratoire. Les raies vibratoires sont considérées comme des variables binaires de présence. De plus, cet algorithme est appliqué sur le spectrogramme complet. Or comme nous l'avons énoncé précédemment, plusieurs signatures inusuelles sont présentes sur les spectrogrammes normaux également. Ainsi, plus la base d'apprentissage est grande, plus le masque statistique risque d'apprendre de l'information inusuelle également. Si cette base est trop petite, une grande quantité d'information normale risque de ne pas être prise en compte.

L'application de la transformée de Hough pour détecter les raies paramétriques est intéressante. Cependant il serait plus pertinent d'apprendre le masque statistique sur de patchs sur lesquels nous pouvons donner la présence ou non de signatures inusuelles. Une autre amélioration serait de tenir compte de l'intensité et de la continuité des points vibratoires restants. Les différentes méthodes énoncées cherchent à détecter des signatures liées à des endommagements, donc des signatures d'intensités importantes et possédant plusieurs harmoniques dans le signal. Dans cette thèse, nous cherchons à détecter toute signature inusuelle présente sur les spectrogrammes, indépendamment du fait qu'elles correspondent à un endommagement. Les signatures à détecter correspondent donc à des signatures en faible dimension et souvent d'intensités faibles.

1.7 Une première approche de détection

Cette partie contient différentes approches afin de détecter les moteurs atypiques sur un patch particulier de la subdivision \mathcal{K}^{128} . Nous considérons donc notre jeu de données comme l'ensemble du même patch sur les différents spectrogrammes.

$$\{Z_{\mathcal{K}_j^{128}}^i\}_{i=1 \dots n}$$

Pour chacun des patchs, nous disposons du label du patch $Y_{Z_{\mathcal{K}_j}}^i$:

$$Y_{Z_{\mathcal{K}_j}}^i = \begin{cases} 1 & \text{si } \exists k \text{ tel que } \text{card}(\text{zones}_{ano}^i(k) \cap \mathcal{K}_j^{128}) \neq 0 \\ 0 & \text{sinon} \end{cases} \quad (1.4)$$

$\text{zones}_{ano}^i(k)$ correspond à la k^{ieme} zone atypique extraite dans la base de données sur le spectrogramme du moteur i . Nous cherchons, avec les méthodes présentées ci-dessous, à discriminer les patchs inusuels (labélisés 1) par rapport aux patchs normaux (labélisés 0).

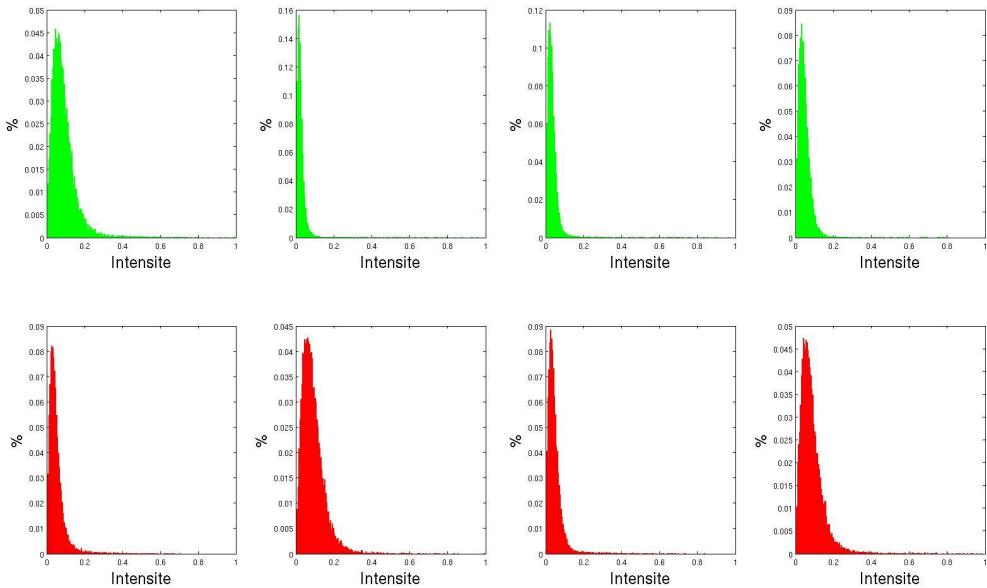


FIGURE 1.12 – Histogrammes des intensités vibratoires d'un patch spécifique pour différents moteurs ayant des spectrogrammes normaux (partie supérieure) et atypiques (partie inférieure) sur ce patch.

1.7.1 Représentation des patchs par leurs histogrammes d'intensités vibratoires

Les patchs $\{Z_{\mathcal{K}_j^{128}}^i\}_{i=1 \dots n}$ peuvent être considérés comme différentes réalisations d'une même donnée appartenant à $\mathbb{R}_+^{128^2}$. Les histogrammes des patchs normaux doivent donc être très proches les uns des autres. Sous la forme d'histogramme d'intensités vibratoires, la localisation des raies est perdue et ainsi seule la présence ou non d'une raie est déterminante. Les raies inusuelles entraînent un nombre de points à intensité élevé plus important que pour les spectrogrammes normaux, les histogrammes devraient donc avoir des queues de distribution différentes.

Afin de pallier les différences d'intensités vibratoires sur les mêmes patchs des spectrogrammes, les intensités vibratoires du patch considéré ont été normalisées par l'intensité maximale sur ce même patch. Ainsi, tous les points des patchs appartiennent à l'intervalle $[0, 1]$, nous notons $\tilde{Z}_{\mathcal{K}_j^{128}}^i$ les patchs normalisés.

$$\tilde{Z}_{\mathcal{K}_j^{128}}^i = \frac{Z_{\mathcal{K}_j^{128}}^i}{\max Z_{\mathcal{K}_j^{128}}^i} \quad (1.5)$$

Les histogrammes correspondent à une subdivision de l'intervalle $[0, 1]$ en sous-intervalles de taille 0.005, les histogrammes sont donc alors des représentations du patch en dimension 200. La figure 1.12 illustre les différents histogrammes obtenus sur des patchs normaux (ligne supérieure) et atypiques (ligne inférieure). Nous pouvons remarquer une variabilité aussi bien pour

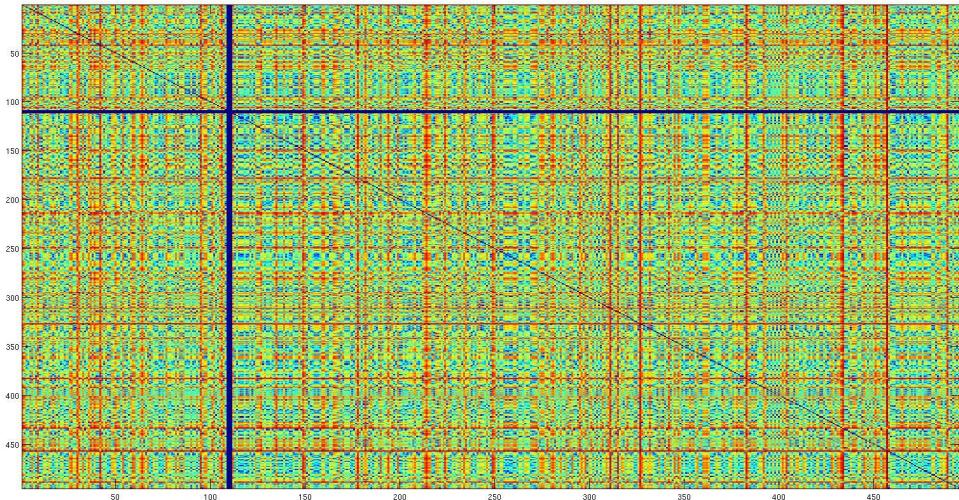


FIGURE 1.13 – Distance entre les histogrammes des mêmes patchs pour différents spectrogrammes. Le premier quadrant (partie supérieure gauche) correspond à la distance entre les patchs normaux, le dernier quadrant (partie inférieure droite) à la distance entre les patchs atypiques, les deux autres à la distance entre les patchs normaux et atypiques. La matrice obtenue n'est pas diagonale par bloc. Cela montre l'inefficacité de la méthode pour discriminer les patchs atypiques des patchs normaux.

les histogrammes des patchs normaux et inusuels, mais également une certaine proximité entre ces mêmes histogrammes. La détection d'anomalies s'effectue à partir de la distance L_2 de ces représentations par histogramme. La figure 1.13 représente la matrice des distance entre les histogrammes des même patchs des différents moteurs de la base de données. Les premiers éléments de la matrice (premier quadrant : premières lignes et colonnes) correspondent aux patchs normaux, les suivants (quatrième quadrant : les lignes et colonnes suivantes) aux patchs inusuels, chaque point de la matrice donne la distance entre les patchs des colonnes et lignes correspondantes. Nous n'observons pas de structure diagonale par bloc dans la matrice des distances permettant de discriminer les patchs normaux des patchs inusuels. La représentation des patchs par leurs histogrammes d'intensités vibrations n'est donc pas un indicateur pertinent pour la détection d'anomalies sur tout un patch.

Cette première étude illustre les problématiques rencontrées dans l'étude des spectrogrammes. Les signatures inusuelles se retrouvent noyées par l'information normale. Les points inusuels consistent en un faible nombre de points sur le patch n'influençant pas suffisamment la distance L_2 entre les histogrammes. Les points atypiques sur le patch ne sont pas les points de plus fortes intensités, les vibrations normales étant de plus forte amplitudes. Les points atypiques possèdent de fortes amplitudes vibratoires par rapport aux points de mêmes coordonnées en fréquence et régime sur les spectrogrammes normaux. Cependant cette représentation en histogramme ne tient pas compte de l'information de position noyant les points inusuels. De plus, les signatures normales des spectrogrammes possèdent également une grande variabilité en présence des raies, position et intensité.

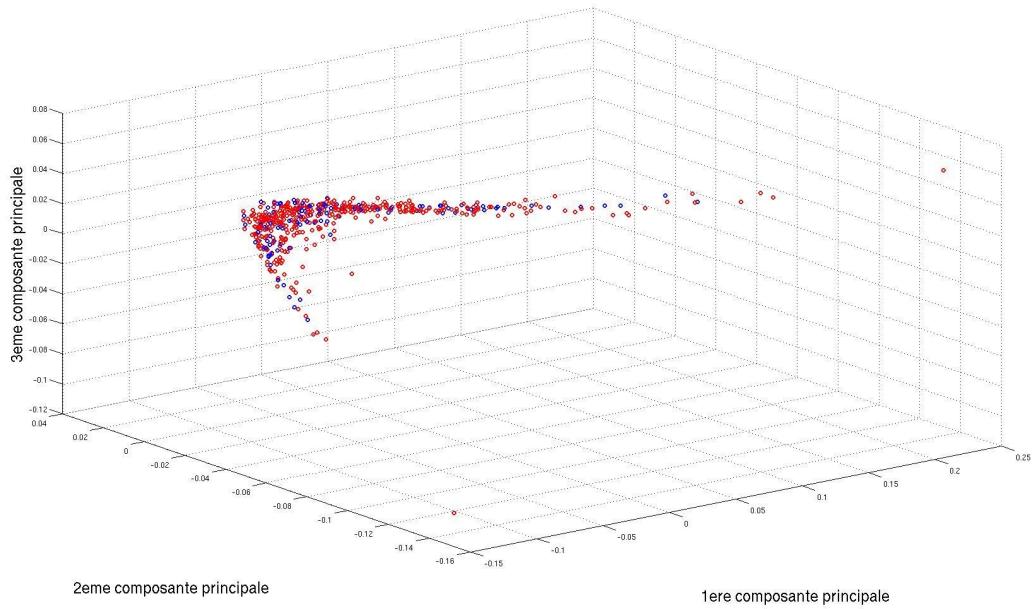


FIGURE 1.14 – Projection des histogrammes de représentation sur les 3 premières composantes principales de la ACP. Les points bleus et rouges correspondent respectivement aux patchs normaux et inusuels.

1.7.2 Représentation des histogrammes dans un espace réduit

La distance L_2 peut être assimilée à une moyenne sur les éléments de l'histogramme noyant les signatures inusuelles. Il est donc plus pertinent de considérer les différents éléments de l'histogramme comme une variable. Chaque patch est donc caractérisé par un vecteur de taille 200. Nous projetons ce vecteur dans un espace réduit maximisant la variance entre ces indicateurs à partir de l'Analyse en Composantes Principales (ACP). Les différentes données sont projetées dans l'espace engendré par les 3 premières composantes principales (Figure 1.14). Les patchs normaux sont représentés en bleu et les patchs atypiques en rouge. La figure 1.14 montre que les indicateurs issus de l'histogramme ne sont pas gaussiens. Les 3 premiers composantes principales ne permettent pas la discrimination des patchs normaux des patchs inusuels. Cela signifie que la variabilité de l'information normale des patchs est plus importante que la variabilité apporté par les signatures inusuelles. Cela montre une nouvelle fois que les signatures inusuelles en très faible dimension sont noyées par de l'information normale. Il est donc nécessaire de caractériser cette information normale afin de faire apparaître les signatures inusuelles.

1.8 Conclusions

Nous disposons d'une base de données de spectrogrammes vibratoires de moteurs d'avions échantillonnés en régime N_2 pour les études réalisées dans cette thèse. Cette base de données est fortement déséquilibrée avec un unique moteur déclaré comme contenant un endommagement.

Il est important de souligner qu'un moteur déclaré sans anomalie peut contenir des signatures inusuelles dans les données. Les spectrogrammes dont nous disposons contiennent des signatures inusuelles cataloguées comme sans gravité par les experts de Safran Aircraft Engines. Il s'agit de ces signatures, parfois d'intensités très faibles, que nous cherchons à détecter dans nos études. Les signatures inusuelles correspondent à des intensités vibratoires plus importantes sur des fréquences f et régimes N_2 non normales.

Dans un premier temps, nous avons récupéré les annotations d'experts afin de construire une base de données numérique de spectrogrammes contenant des zones labélisées comme normales ou atypiques. Nous disposons ainsi d'une base de données de 493 spectrogrammes annotés. Une analyse globale des spectrogrammes n'étant pas pertinente, nous avons donc décidé de subdiviser le spectrogramme en patchs carrés de taille 128 sans recouvrement et de les étudier indépendamment les uns des autres. Un modèle est donc défini pour chacun des patchs. Cette démarche permet d'avoir une meilleure mise en évidence des signatures inusuelles et une labélisation de chaque patch à partir de la base de données construite.

Les signatures inusuelles sont très variées en forme, en intensité et en position et le nombre de spectrogrammes ayant des signatures inusuelles est très faible au niveau de certains patchs. Ces deux effets entraînent l'impossibilité de mettre en place des modèles de données atypiques et donc d'utiliser des méthodes supervisées de classification. Il est donc indispensable que les algorithmes mis en place ainsi que les représentations tiennent compte de ce déséquilibre et de la structure normale des patchs. Les méthodes développées dans cette thèse sont donc principalement portées sur des approches de type one-class de détection de nouveautés.