

exactement la même localisation sur les différents spectrogrammes.

Cette représentation possède également quelques inconvénients compliquant la mise en place d'algorithmes de détection automatique. La figure 1.5 montre une grande variabilité sur les relations N_1/N_2 des différents moteurs pouvant dépendre de la température ambiante. Ainsi bien que les signatures vibratoires de même ordre fréquentiel issues de l'arbre HP se trouvent toujours à la même localisation, cette affirmation est erronée en ce qui concerne les signatures vibratoires de même ordre issues de l'arbre BP. Une seconde contrainte liée à cette représentation est sa grande dimension, chaque spectrogramme possède 1.5 million de mesures vibratoires à différents régimes et fréquences. Cependant comme on l'observe sur la figure 1.4, les signatures non usuelles consistent en une infime partie du spectrogramme caractérisée par quelques centaines de points sur ce dernier. Les informations pertinentes que sont les signatures inusuelles risquent fortement d'être noyées par des informations non pertinentes, comme les raies normales ou du bruit, en cas d'étude du spectrogramme dans sa globalité. Nous verrons par la suite comment nous pallions cette disproportion. Les informations non pertinentes que sont les signatures normales et le bruit de mesures peuvent être assimilées à du bruit par rapport à notre étude. Nous pouvons alors dire que le rapport signal sur bruit du spectrogramme est très faible.

Les spectrogrammes consistent en une concaténation de spectres eux même bruités. Les intensités des signatures inusuelles peuvent être proches du niveau du bruit sur les spectrogrammes. Nous cherchons à détecter toute trace de ces signatures atypiques sur les spectrogrammes et non pas uniquement celles synonymes d'endommagement du moteur, donc également celles ayant des intensités proches du bruit.

La dernière contrainte présente dans nos données est l'absence d'étiquetage numérique des différentes vibrations inusuelles présentes sur les spectrogrammes. Cette contrainte n'est pas liée aux spectrogrammes en eux-même, à l'acquisition ou la transformation des données. La base de données numérique dont nous disposons contient uniquement les spectrogrammes des moteurs, mais manque d'informations sur les signatures atypiques présentes dans ces derniers. Dans la section suivante, nous discutons de la création d'une base de données annotées.

1.4 Construction de la base de données

1.4.1 L'annotation manuelle des experts

L'analyse par les experts des spectrogrammes s'effectue visuellement et manuellement. Safran Aircraft Engines dénomme cette opération le "screening visuel". Pour chaque nouveau moteur, les spectrogrammes en phase d'accélération et de décélération sont observés et les experts encadrent (en jaune) les zones suspectées de contenir des signatures atypiques. Cette annotation est effectuée sur des captures d'écran de plages de fréquences des spectrogrammes (Figure 1.6). Les plages de fréquences observées sont les mêmes pour chaque spectrogramme analysé. Ces informations

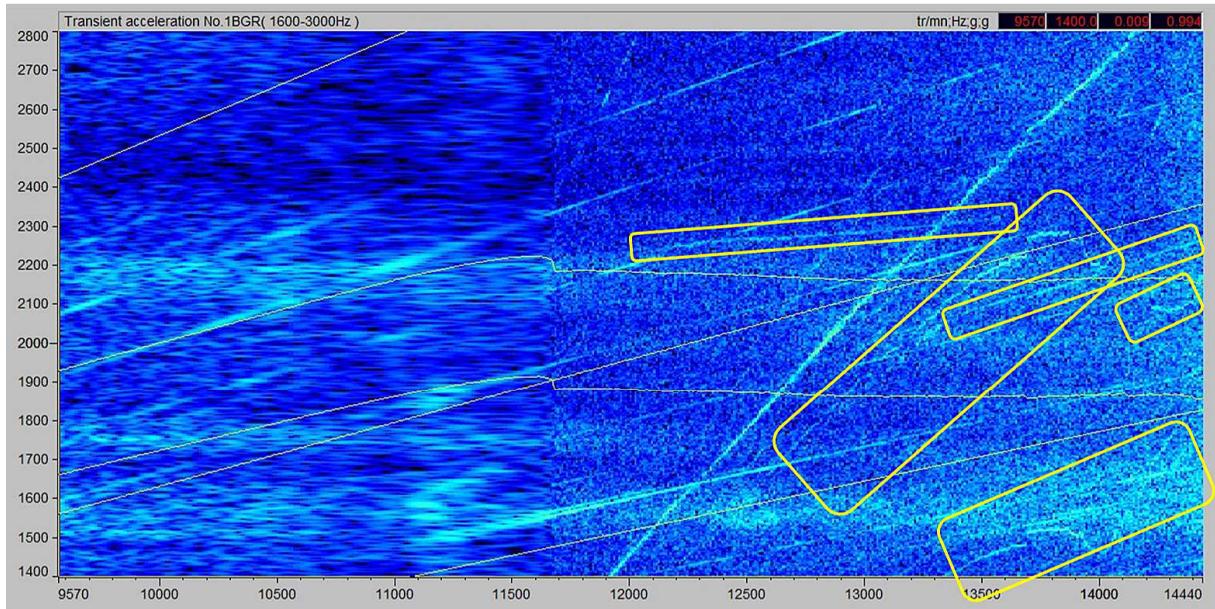


FIGURE 1.6 – Annotation manuelle des spectrogrammes par des experts. Chaque cadre jaune correspond à une zone suspectée par un expert de contenir une signature inusuelle. Les plages de fréquences observées peuvent contenir plusieurs signatures atypiques, ces dernières peuvent également se superposer.

ne permettent pas de donner la localisation précise des signatures inusuelles sur les différents spectrogrammes. Il n'existe pas de traces numériques de ces signatures, les seules informations disponibles étant :

- les données numériques des spectrogrammes moteurs sans annotations des signatures inusuelles,
- les données textuelles des captures d'écran des annotations manuelles des experts encadrant les zones contenant des signatures inusuelles.

Plusieurs signatures atypiques peuvent être présentes sur le spectrogramme au niveau de la plage de fréquences étudiées.

La figure 1.6 montre une capture d'écran annotée par un expert sur la plage de fréquences 1400–2800Hz. Les éléments caractérisés sont des zones du spectrogramme contenant la signature décrétée comme atypique. Cela ne signifie pas que tous les points de la zone encadrée sont à détecter, mais qu'une signature inusuelle est présente dans la zone. Même au sein de la zone encadrée, une grande partie des points restent normaux. La figure montre également une grande diversité des signatures inusuelles au sein d'une même zone, cette diversité et leurs faibles nombres empêchent de mettre en place des modèles de données atypiques. Notre avons construit une base de données numérique associant les données numériques des spectrogrammes aux données textuelles des captures d'écran d'annotations manuelles (Figure 1.6) des experts.

La visualisation, par les experts de Safran Aircraft Engines, de ces spectrogrammes permet de vérifier que les signatures atypiques présentes ne sont pas liées à un endommagement du moteur. Les travaux réalisés durant cette thèse ont pour but de mettre en évidence ces signatures atypiques, et ainsi de pouvoir acquérir automatiquement ces différentes signatures inusuelles.

1.4.2 Extraction automatique des zones anormales sur les données textuelles

Nous cherchons à mettre en place des algorithmes de détection d'anomalies sur les spectrogrammes. Cependant pour se faire, il est nécessaire de connaître les zones contenant des signatures inusuelles afin de sélectionner les données sur lesquelles les modèles de normalité (caractérisant les informations normales sur les spectrogrammes) sont calibrés. Les captures d'écran des annotations manuelles des experts contiennent cette information. Nous avons extrait les cadres (zones) définis par les experts pour convertir cette information textuelle en données numériques pour la construction d'une base de données numérique de spectrogrammes annotés. L'algorithme 3 d'extraction analyse les captures d'écran sur lesquelles nous cherchons à détecter les pixels de couleur jaune (Figure 1.6). Nous récupérons une information de position des pixels sur l'image que nous projetons sur les axes des régimes et fréquences connus afin d'avoir une estimation des fréquences et régimes des pixels concernés. La plage de fréquences est connue car elle est identique sur tous les spectrogrammes analysés. La plage des régimes N_2 est également connue car les spectrogrammes sont analysés sur la plage totale des régimes dont nous disposons dans notre base numérique de spectrogrammes. Une fois l'ensemble des pixels déterminés, nous cherchons un chemin fermé de ces points. Ce chemin constitue le (ou les cadres) annoté(s) par les experts. Plusieurs cas de figures peuvent apparaître :

1. le cadre (rectangle) extrait par le chemin fermé est droit et n'est pas superposé avec un autre cadre (Figure 1.7a) ; dans ce cas, le chemin des pixels jaunes et le cadre récupéré correspond à celui annoté par les experts
2. le cadre est oblique et n'est pas superposé avec un autre cadre (Figure 1.7b) ; pour plus de simplicité, nous cherchons à caractériser les zones atypiques uniquement par des rectangles droits. Ainsi, pour caractériser ces cadres obliques, nous récupérons le plus petit rectangle droit englobant le rectangle oblique.
3. Différents cadres se superposent (qu'ils soient droits ou obliques) (Figure 1.7c). Nous n'avons pas cherché à discriminer un cadre par rapport à un autre dans ce cas de figure, bien que des techniques plus avancées de reconnaissance de forme le permettraient. Le croisement des différents cadres entraînent des bifurcations pour les chemins fermés. Pour caractériser ces zones atypiques, nous prenons encore une fois le plus petit rectangle droit englobant l'ensemble des différents cadres superposés.

Les zones atypiques extraites sont définies par une plage de fréquences et de régimes définissant un rectangle droit dans le spectrogramme, même si la zone annotée initiale correspondait à un rectangle oblique. Ces cadres englobent donc plus d'informations normales et donc non significatives que les cadres créés par les experts mais contiennent bien les signatures inusuelles. Cette extraction a permis la mise en place d'une base de données numérique annotée contenant les spectrogrammes ainsi qu'une information sur les positions des signatures inusuelles dans ces derniers. Nous considérons donc ces zones comme contenant des signatures inusuelles mais nous ne considérons pas tous les points de ces zones comme atypiques.

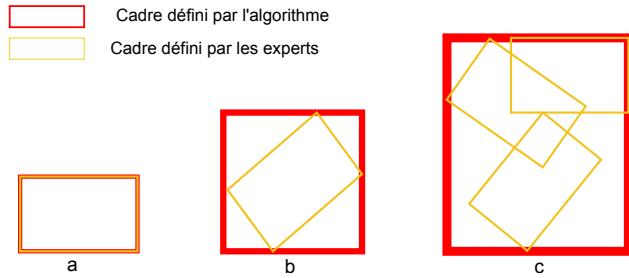


FIGURE 1.7 – Les différents cas de figure d'extraction des zones atypiques (rouge) sur les annotations manuelles (jaune).

Algorithme 3 : Extraction des annotations manuelles

Données : Les annotations manuelles, le spectrogramme vibratoire, l'axe des régimes et l'axe des fréquences

Résultat : L'ensemble des zones contenant des signatures inusuelles

Conversion de l'annotation manuelle en image.

Détection de tous les pixels associés à la couleur jaune sur l'image.

Récupération des formes connexes parmi les pixels détectés.

pour chaque composante connexe **faire**

Détermination des coordonnées de pixels minimales et maximales.

Projection de ces coordonnées sur l'axe des régimes et des fréquences.

Définition du rectangle englobant.

fin

1.4.3 La base de données enrichie

Nous disposons dans notre base de données de $n = 493$ moteurs (492 déclarés sans endommagement et 1 avec endommagement) annotés par les experts. Les labels des signatures inusuelles ne sont pas indiqués, ces zones sont donc indexées par un simple numéro. Chaque spectrogramme ne possède pas le même nombre de zones atypiques. Chaque donnée contient les différentes informations nécessaires à l'étude des spectrogrammes (Figure 1.8) :

- le spectrogramme du moteur i , $S^i \in \mathbb{R}^{\text{card}(f) \times \text{card}(N_2)}$ représentées comme une matrice de dimension la taille $\text{card}(f)$ de l'échantillonnage fréquentiel calculé par la transformée de Fourier et la taille $\text{card}(N_2)$ de la discréttisation des vitesses de rotation N_2 de l'arbre HP atteint par le moteur sur le banc d'essai. Les fréquences sont les mêmes pour tous les spectrogrammes de la base de données, l'échelle des régimes N_2 possède une plage commune de régimes sur l'ensemble de la base,
- les vitesses de rotation de l'arbre HP du moteur i , N_2^i correspondant aux régimes sur lesquels le spectrogramme a été calculé,
- les vitesses de rotation de l'arbre BP du moteur i , N_1^i qui correspondent aux régimes BP associés aux régimes HP,
- les zones atypiques (Zone_Ano) détectées par l'algorithme 3 paramétrées par des fréquences minimale et maximale et des régimes N_2 minimal et maximal,

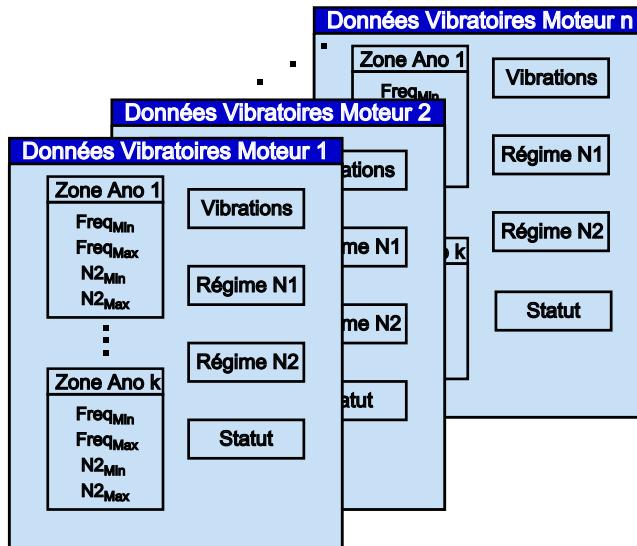


FIGURE 1.8 – Illustration de la base de données et de ses éléments. Chaque observation de la base de données contient le spectrogramme (Vibrations) avec ses différentes zones atypiques définies par une plage de fréquences et de régimes N_2 . Les régimes N_1 , N_2 et le statut normal ou endommagé du moteur sont également des attributs de l’observation.

- le statut du moteur correspondant à l’état normal ou endommagé du moteur défini par les experts ; notre base contient presque intégralement des moteurs classés comme normaux.

Nous notons cette base de données \mathcal{B}_0 . A partir de cette base, il est possible de manière simple de sélectionner une partie spécifique du spectrogramme et de déterminer les moteurs dont les spectrogrammes sont normaux ou atypiques sur cette partie. Il n’est cependant pas possible pour le moment de sélectionner un structure de signature et d’en déterminer l’ensemble des spectrogrammes où la signature est présente.

1.5 Étude des spectrogrammes par patch

1.5.1 Localisation des signatures inusuelles sur le spectrogramme

Les moteurs déclarés comme normaux par les experts peuvent posséder des signatures inusuelles sur leurs spectrogrammes ne correspondant pas nécessairement à un endommagement du moteur. Ces signatures atypiques restent des informations très localisées sur ces derniers. Une même anomalie apparaît en général à des fréquences bien spécifiques. La figure 1.9 présente une carte des proportions des zones atypiques présentes dans notre base de données. Cette carte donne point par point sur le spectrogramme la proportion de moteurs de notre base de données dont le point étudié sur le spectrogramme appartient à une zone atypique. Les éléments de ce résultat sont à nuancer car notre algorithme d’extraction défini au-dessus inclut de nombreux points normaux dans les zones extraites. Cette carte nous apporte tout de même des informations concernant les différentes zones où les irrégularités sont présentes. Les zones atypiques sont localisées, la majorité des zones ne sont pas du tout (ou très peu) touchées par des signatures

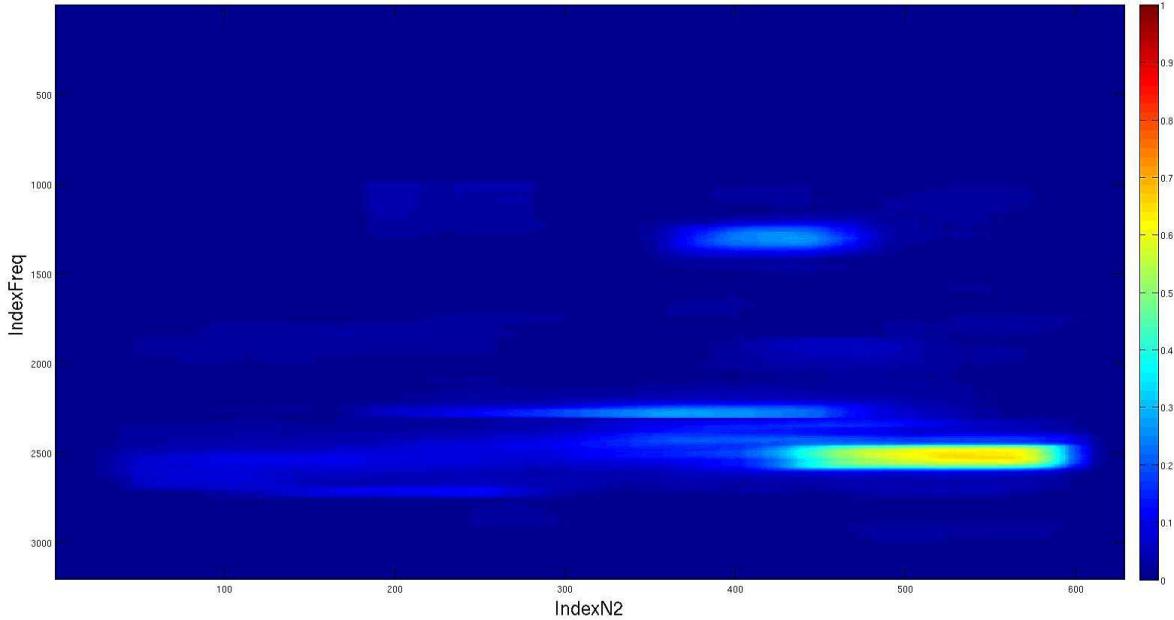


FIGURE 1.9 – Proportion (allant du bleu au rouge) pour chaque point du nombre de moteurs dans la base de données dont le point sur le spectrogramme appartient à une zone extraite. La plupart des points appartiennent rarement à des zones atypiques, une zone particulière attire notre attention avec plus de la moitié des moteurs de la base de données possédant une signature atypique sur cette zone dans leurs spectrogrammes.

inusuelles alors que d'autres possèdent une signature inusuelle dans plus de la moitié de notre base de données. Les zones atypiques présentes sur les spectrogrammes sont concentrées sur des positions spécifiques du spectrogramme, il est donc intéressant d'étudier chacune de ces zones séparément et de ne pas définir de modèle sur l'intégralité du spectrogramme. Nous présentons dans la partie suivante les bénéfices apportés par une approche de ce type.

1.5.2 Subdivision du spectrogramme en patchs

Les signatures inusuelles présentes sur les spectrogrammes sont de dimensions infimes comparées à la taille des spectrogrammes, la dimension correspondant au nombre de pixels. Il est donc fort probable que la signature inusuelle soit noyée par l'information normale présente sur le spectrogramme en effectuant une analyse statistique directement sur l'intégralité de ce dernier. Deux points d'attention empêchent l'étude du spectrogramme dans sa globalité comme une unique donnée :

- à l'échelle du spectrogramme, le rapport signal à bruit est très faible, si on considère le bruit comme l'information non pertinente constituée des raies normales et le bruit de mesure et le signal comme les raies atypiques ;
- bien que les moteurs dont sont issus les spectrogrammes sont considérés comme non endommagés, les spectrogrammes peuvent contenir des signatures inusuelles qu'il est important de détecter. En considérant le spectrogramme dans sa globalité, peu de données pourraient être considérées comme exemptes de signature atypique.

Une méthode pour pallier ces problèmes est de subdiviser les spectrogrammes en patchs sur une plage commune de régimes. Chaque patch correspond à une même plage de fréquences et de régimes sur l'ensemble de la base de données sur lequel une analyse est effectuée indépendamment des autres patchs. Nous définissons alors une subdivision \mathcal{K} comme un ensemble d'intervalles de régimes et de fréquences décomposant le spectrogramme en patchs Z correspondant à une extraction d'un rectangle sur le spectrogramme complet :

$$\mathcal{K} = \{\mathcal{K}_j = [f_{jini}, f_{jend}] \times [N_{2,jini}, N_{2,jend}], j \in \{1, \dots, \text{card}(\mathcal{K})\}\} \quad (1.2)$$

$$Z_{\mathcal{K}_j}^i = S^i(\mathcal{K}_j)$$

avec i le moteur considéré et S^i le spectrogramme issu de ce moteur, \mathcal{K}_j correspond au patch j de la subdivision \mathcal{K} .

Cette approche est similaire au processus des experts en mécanique de Safran Aircraft Engines. Le screening visuel consiste à regarder le spectrogramme sur des plages de fréquences prédéfinies afin d'avoir une meilleure observation des signatures inusuelles. La subdivision est indépendante du spectrogramme considéré et apporte divers avantages. Elle répond aux deux points d'attention énoncés ci-dessus. Le rapport signal à bruit de chaque patch est nettement supérieur par rapport au même rapport sur le spectrogramme complet, les signatures atypiques sont bien plus visibles au niveau des patchs. A l'inverse de la donnée complète, où nous disposons de peu de données sans aucune signature inusuelle pour mettre en place les modèles caractérisant la normalité, à l'échelle du patch ce problème ne se pose pas. A partir de la base de données construite (section 1.4), il est possible de sélectionner l'ensemble des spectrogrammes avec ou sans anomalie sur le patch considéré. Pour cela, il suffit de vérifier sur le spectrogramme étudié l'intersection entre les zones atypiques extraites dans la base données construite et le patch considéré. Si cette intersection est vide, le patch du moteur étudié est considéré comme normal. Nous disposons donc, à partir de la base de données et pour chaque patch de la subdivision, de la vérité terrain $Y_{Z_{\mathcal{K}_j}}^i$ caractérisant le caractère normal ou atypique pour le spectrogramme i du patch j de la subdivision \mathcal{K} . Rappelons que cette information à l'échelle ponctuelle (c'est-à-dire au niveau de chaque point du spectrogramme) Y_{f,N_2}^i n'est pas accessible car les annotations d'experts s'effectuent sur des zones du spectrogramme et non pas au niveau des points de ce dernier. Nous ne pouvons pas considérer chaque point des zones atypiques comme inusuels.

L'étude des spectrogrammes à partir d'une subdivision apporte d'autres avantages qu'une réduction du rapport signal sur bruit ou la classification des patchs comme normaux ou atypiques :

- la variabilité au niveau du patch est nettement inférieure à celle du spectrogramme complet ce qui permet de mieux caractériser les comportements normaux à cette échelle ;
- les patchs sont de dimension inférieure, ce qui permet une mise en place plus rapide des modèles ;
- l'anomalie est détectée uniquement sur le ou les patchs contenant les signatures inusuelles. Nous disposons donc d'une localisation approximative de ces signatures ;
- l'analyse des différents patchs est parallélisable, car ils sont étudiés de manière indépen-

dante.

Trois paramètres définissent la subdivision des spectrogrammes en patchs :

- la taille des patchs en longueur et largeur,
- le recouvrement des patchs.

Les signatures inusuelles correspondent à des raies. Afin de pouvoir les distinguer, il est nécessaire que les patchs soient suffisamment larges afin d'avoir un maximum d'information les concernant. Cependant si le patch est trop grand, la signature inusuelle risque d'être noyée par de l'information normale. La taille du patch doit apporter un équilibre entre la quantité relative d'information inusuelle présente dans le patch par rapport aux zones atypiques et la quantité d'information normale. De plus certaines méthodes de représentation par dictionnaire exigent que la donnée d'entrée soit sous la forme de carré dyadique. Nous avons donc choisi cette structure de patch. Avoir la totalité de la signature inusuelle dans le patch permettrait de maximiser le taux d'information inusuelle. Nous avons donc déterminé la taille moyenne du carré permettant d'englober les zones atypiques de la base de données.

$$\hat{L}_{\mathcal{K}} = \frac{1}{\text{card}(\text{zones}_{ano})} \sum_{k \in \text{zones}_{ano}} \max(L_{\text{zones}_{ano}^k}, l_{\text{zones}_{ano}^k})$$

$L_{\text{zones}_{ano}^k}, l_{\text{zones}_{ano}^k}$ correspondent à la longueur et la largeur des zones atypiques zones_{ano}^k de la base de données.

Nous obtenons $\hat{L}_{\mathcal{K}} \approx 147$. Cette valeur est supérieure à la réalité, les zones extraites par l'algorithme 3 sont surdimensionnées (prise du rectangle droit uniquement, fusion des zones) par rapport aux zones annotées par les experts. Nous avons donc opté pour une subdivision en carré dyadique de taille 128 pixels, cela signifie une plage de fréquences de $400Hz$ et une plage de régimes de 1280rpm. Une taille de 256 pixels serait trop importante et noierait les signatures inusuelles. Des patchs de dimension 64 ne permettraient pas de contenir toute la signature inusuelle. Le recouvrement des patchs n'est pas indispensable pour les approches étudiées dans cette thèse. Nous utilisons donc une subdivision \mathcal{K}^{128} du spectrogramme en patchs carrés de taille 128×128 sans recouvrement (Figure 1.10). Dans la suite, nous étudions également les spectrogrammes de manière ponctuelle comme un ensemble de points d'intensités vibratoires. Nous pouvons considérer cela comme une subdivision du spectrogramme en patchs de taille 1. Cette subdivision est caractérisée par les coordonnées en fréquence f et en régime N_2 de chaque point.

La subdivision est établie sur les plages de fréquences et de régimes communes à l'ensemble des spectrogrammes. Les vibrations sous et au-dessus d'un certain régime ne sont pas considérées car il s'agit de régimes non communs à l'ensemble de la base. Les vibrations à des fréquences supérieures à un certain seuil ne sont pas prises en compte non plus du fait de la non-annotation par les experts (repliement spectral à des fréquences très élevées). Nous considérons donc les spectrogrammes comme l'ensemble des patchs les composant.

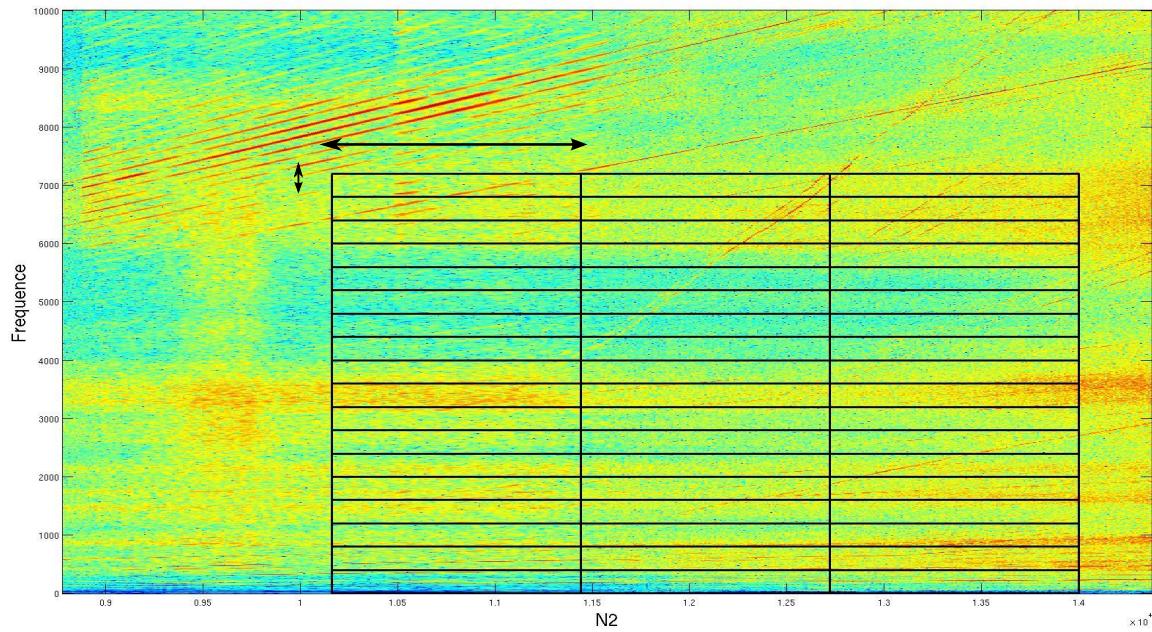


FIGURE 1.10 – Subdivision \mathcal{K}^{128} . Les flèches correspondent à la taille des patchs et chaque cadre à un des patchs. Les intensités vibratoires sous et au-dessus de certains régimes ne sont pas prises en compte, tout comme les intensités au-dessus d'une certaine fréquence.

1.5.3 Labélisation ponctuelle du patch - enrichissement de la base de données

Nous cherchons à donner un résultat de détection sur les patchs considérés. Il est également possible de mettre en place des algorithmes afin de détecter ponctuellement les signatures inusuelles, c'est-à-dire la détection des différents points composant cette signature et non pas du patch la contenant. Cependant, nous ne disposons pas (et nous n'avons pas la possibilité) d'obtenir une annotation de tous les points des différents spectrogrammes. Il s'agit d'une des raisons pour lesquelles nous avons opté pour une détection au niveau du patch. Afin de donner des résultats numériques de cette détection, combinés aux résultats visuels, nous avons annoté manuellement quelques points des spectrogrammes. Pour une soixantaine de spectrogrammes, une douzaine de points ont été récupérés et chaque point est affecté à une des 4 classes suivantes :

- classe "normal" : le point fait partie d'une signature normale liée à l'arbre HP,
- classe "inusual" : le point fait partie d'une signature atypique,
- classe "bruit" : bruit du spectrogramme, le point ne se trouve pas sur une raie et ne représente donc pas la vibration spécifique d'un arbre ou d'une irrégularité,
- classe "décalé" : le point fait partie d'une signature normale, mais cette dernière n'est pas exactement identique sur l'ensemble des spectrogrammes, il s'agit des signatures liées à l'arbre BP n'ayant pas la même position dans tous les spectrogrammes.

Seule la classe 2 doit être détectée comme atypique.

Trois points issus de chaque classe sont récupérés sur un patch spécifique d'un ensemble de 60 spectrogrammes. Le patch sélectionné correspond à celui contenant un nombre important de signatures inusuelles dans notre base de données (Figure 1.9). Cette extraction d'information

est possible uniquement manuellement et est coûteuse en temps. Il n'a donc pas été possible d'effectuer cette opération sur tous les points, patchs, ou spectrogrammes. Notons cette base de données \mathcal{B}_1 . Elle contient uniquement les données du patch des spectrogrammes desquels les informations ponctuelles ont été extraites. Cette extraction permet de donner des résultats de détection ponctuelle sur les spectrogrammes.

1.5.4 La grande variabilité des signatures inusuelles

Dans un cadre supervisé, une méthode pour classifier les patchs comme normaux ou atypiques serait de construire un classifieur pour chaque patch indépendamment. Considérons un patch Z_j issu d'une des subdivisions, l'approche serait de définir ou d'apprendre un modèle sur les patchs normaux Z_j^{normal} et sur les patchs inusuels Z_j^{ano} (Figure 3). Le classifieur correspondrait alors à l'appartenance la plus vraisemblable d'un nouveau patch Z_j^{test} à l'une des classes (les modèles sont définis pour un patch en particulier, les patchs tests correspondent exactement à la même zone sur les spectrogrammes tests). C'est de cette manière que des méthodes telles que la régression logistique ou les arbres de classification [49] sont construites. Cependant, pour effectuer une telle approche, il est indispensable que la base de données d'apprentissage contienne suffisamment d'échantillons des différentes classes. Il est également nécessaire que la variabilité au sein d'une même classe ne soit pas trop importante.

Notre base de données ne satisfait aucun de ces deux critères. Comme nous l'avons vu sur la figure 1.9, certaines zones contiennent très peu de données avec des signatures inusuelles. Mettre en place un modèle d'anomalie dans ce cadre n'est donc pas adapté. Certaines zones, à l'inverse, contiennent un nombre suffisant de données atypiques. Cependant ces derniers ne satisfont pas le second critère. Il existe une grande variabilité des signatures singulières au sein d'un même patch et pas suffisamment de données pour chacune d'entre elles (Figure 1.11).

Les différents types de signatures inusuelles du même patch (Figure 1.11) présentent de grandes variabilités en forme, position et intensité. Les signatures atypiques de même type possèdent également les mêmes variabilités. Une signature liée au palier #4 est dépendante du régime N_2 et N_1 . Ainsi du fait de la variabilité des relations N_1/N_2 (Figure 1.5), chacune de ces signatures est différente. De plus, il est possible que ce ne soit pas la fréquence fondamentale de la signature inusuelle qui soit observable sur le spectrogramme mais une de ses harmoniques. Nous ne disposons pas dans notre base de données de plusieurs échantillons de tous les types possibles de signatures inusuelles afin de calibrer un modèle pour chacun d'entre eux.

Néanmoins, nous disposons de suffisamment de données sans anomalie afin d'apprendre les caractéristiques normales des patchs des spectrogrammes. La détection d'anomalies sur les patchs s'effectue alors en mesurant l'écart entre la donnée test et ce modèle normal. Cette procédure est associée aux approches de type détection d'anomalies, d'outliers, de nouveautés qui permettent d'effectuer des détections de manière automatique sans (ou avec très peu) de données atypiques.