

Résumé

Les mesures de vibrations sont l'une des données les plus pertinentes pour détecter des anomalies sur les moteurs. Les vibrations sont acquises sur banc d'essai en phase d'accélération et de décélération pour assurer la fiabilité du moteur à la sortie de la chaîne de production. Ces données temporelles sont converties en spectrogrammes pour permettre aux experts d'effectuer une analyse visuelle de ces données et de détecter les différentes signatures atypiques. Les sources vibratoires correspondent à des raies sur les spectrogrammes. Dans cette thèse, nous avons mis en place un outil d'aide à la décision automatique pour analyser les spectrogrammes et détecter tout type de signatures atypiques, ces signatures ne proviennent pas nécessairement d'un endommagement du moteur. En premier lieu, nous avons construit une base de données numérique de spectrogrammes annotés. Il est important de noter que les signatures inusuelles sont variables en forme, intensité et position et se trouvent dans un faible nombre de données. Par conséquent, pour détecter ces signatures, nous caractérisons les comportements normaux des spectrogrammes, de manière analogue aux méthodes de détection de nouveautés, en représentant les patchs des spectrogrammes sur des dictionnaires comme les curvelets et la Non-negative matrix factorization (NMF), ainsi qu'en estimant la distribution de chaque point du spectrogramme à partir de données normales dépendamment ou non de leur voisinage. La détection des points atypiques est réalisée par comparaison des données tests au modèle de normalité estimé sur des données d'apprentissage normales. La détection des points atypiques permet la détection des signatures inusuelles composées par ces points.

mots clés : apprentissage de dictionnaire, curvelets, estimation de densité par noyau, détection de nouveautés, détection d'anomalies, vibrations

Abstract

Vibration measurements are one of the most relevant data for detecting anomalies in engines. Vibrations are recorded on a test bench during acceleration and deceleration phases to ensure the reliability of every flight engine at the end of the production line. These temporal signals are converted into spectrograms for experts to perform visual analysis of these data and detect any unusual signature. Vibratory signatures correspond to lines on the spectrograms. In this thesis, we have developed a decision support system to automatically analyze these spectrograms and detect any type of unusual signatures, these signatures are not necessarily originated from a damage in the engine. Firstly, we have built a numerical spectrograms database with annotated zones, it is important to note that data containing these unusual signatures are sparse and that these signatures are quite variable in shape, intensity and position. Consequently, to detect them, like in the novelty detection process, we characterize the normal behavior of the spectrograms by representing patches of the spectrograms in dictionaries such as the curvelets and the Non-negative matrix factorization (NMF) and by estimating the distribution of every points of the spectrograms with normal data depending or not of the neighborhood. The detection of the

unusual points is performed by comparing test data to the model of normality estimated on learning normal data. The detection of the unusual points allows the detection of the unusual signatures composed by these points.

keywords : dictionary learning, curvelets, kernel density estimation, novelty detection, anomaly detection, vibrations

Introduction

Le contexte industriel

Les moteurs manufacturés par Safran Aircraft Engines suivent une série de tests afin de vérifier leur viabilité avant d'être envoyés aux clients. Un de ces tests consiste en l'acquisition et l'analyse des signaux vibratoires sur banc d'essai durant différentes phases. Les moteurs d'avions sont des machines tournantes et subissent de nombreuses vibrations durant leurs services. Leurs mesures vibratoires contiennent des informations pouvant être liées au mauvais fonctionnement de différents éléments du moteur comme par exemple les roulements ou les engrenages (Figure 1). Les informations vibratoires inusuelles présentes sur les signaux ne sont pas nécessairement liées à un endommagement du moteur ou au mauvais fonctionnement de ce dernier, elles peuvent être issues de l'acquisition ou à l'excitation de certains modes non usuels du moteur. D'autres signatures présentes dans ces signaux sont normales et liées au fonctionnement nominal du moteur comme celles liées au fan (Figure 1). Ces dernières sont prévisibles et observables dans les signaux.

De nombreuses méthodes d'analyse des signaux vibratoires pour détecter des anomalies existent déjà [91]. Ces dernières se basent principalement sur les signaux vibratoires temporels sur lesquels sont appliquées diverses méthodes de traitement du signal comme la transformée de Fourier, la transformée en ondelettes [80], des moyennes synchrones, la cyclostationnarité [8]. Des représentations de type temps-fréquence sont également utilisées pour étudier les comportements vibratoires non-stationnaires, tels que les spectrogrammes ou les scalogrammes [117].

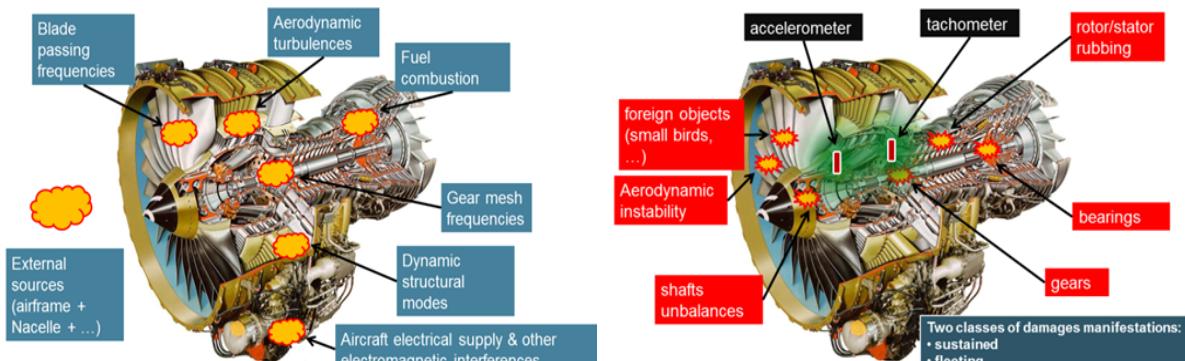


FIGURE 1 – Sources de vibrations normales des moteurs (gauche) et de vibrations anormales (droite)

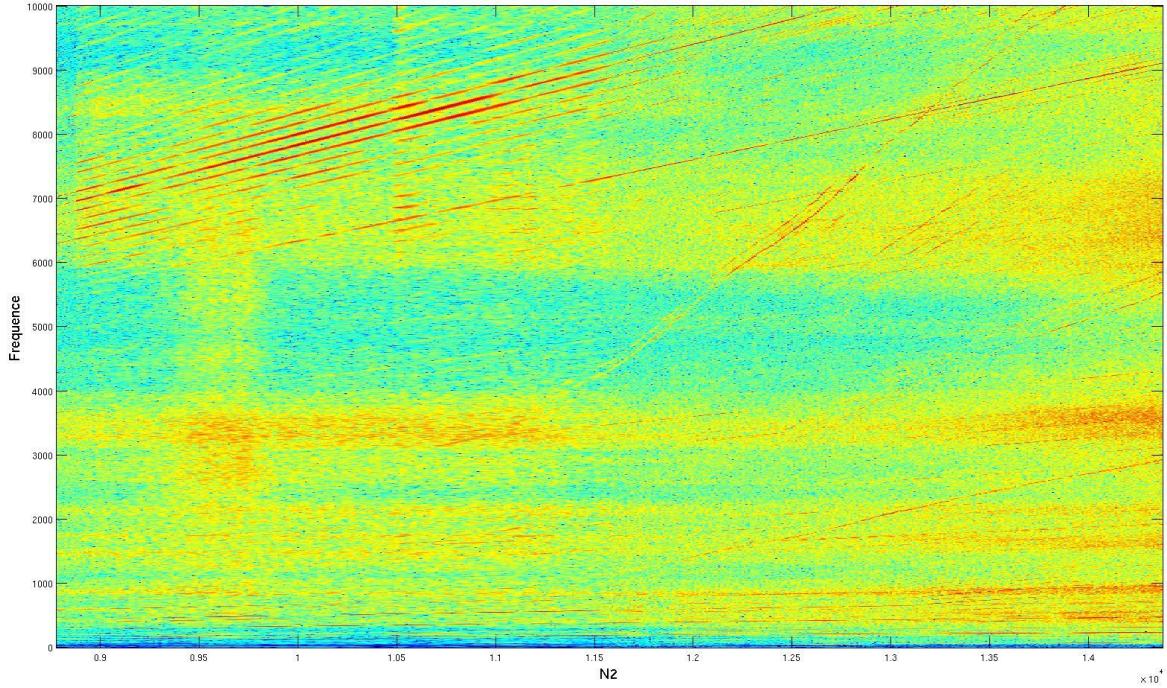


FIGURE 2 – Spectrogramme indexé en ordre (fréquence de vibration du moteur dépendant de la vitesse de rotation) durant une décélération lente. L'axe des abscisses correspond à une vitesse de rotation. Pour chaque vitesse, un spectre est calculé et est représenté sur l'axe des ordonnées, les couleurs allant du bleu au rouge donnent l'intensité de l'énergie pour la vitesse de rotation et la fréquence observée.

Problématique

Nous nous intéressons principalement à un type particulier de spectrogramme acquis pendant des transitoires lents (accélération et décélération) dont l'abscisse horizontale qui initialement représente le temps est remplacée par la vitesse de rotation d'un arbre du moteur, le régime moteur (Figure 2). Ces spectrogrammes sont représentatifs des signaux vibratoires sur banc d'essai de réception (avant livraison au client) pour lesquels nous cherchons des représentations pertinentes pour la détection automatique de signatures non usuelles potentielles. Les spectrogrammes vibratoires représentent l'intensité vibratoire à différentes fréquences f et différents régimes N_2 (une des vitesses de rotation du moteur). Ces mesures sont fortement bruitées, chaque point du spectrogramme peut être considéré comme une variable aléatoire. La construction des spectrogrammes à partir des mesures vibratoires est donnée dans le chapitre 1. A l'heure actuelle, les spectrogrammes (enregistrés dans un format numérique) sont analysés visuellement et annotés manuellement dans un format textuel par des experts. Il faut souligner qu'aucune trace numérique des signatures détectées n'est conservée. Par ailleurs, certains algorithmes sont utilisés pour rechercher des signatures spécifiques sur les spectrogrammes.

Nous disposons d'une base de données numérique des spectrogrammes issus des bancs d'essais, cette base ne contient qu'un unique moteur connu comme étant endommagé. Ces moteurs proviennent de la chaîne de production et ne contiennent que très rarement des signatures anor-

males liées à des endommagements. Endommager volontairement un moteur afin d'obtenir des données contenant des signatures anormales n'est pas envisageable car les moteurs ainsi que ces campagnes d'essai sont très coûteux. De plus, le nombre de possibles endommagements du moteur, bien qu'excessivement rares, est très large étant donnée la complexité du système. Un moteur est déclaré comme endommagé en cas de signature anormale trop importante et non pas uniquement en cas de présence d'un artéfact sur l'image du spectrogramme. Ainsi les moteurs déclarés comme normaux par les experts peuvent contenir des signatures inusuelles sans risque annotées et analysées par ces mêmes experts. Nous cherchons à détecter et mettre en évidence toutes les signatures inusuelles de manière automatique pour apporter une aide à la décision aux experts. Ces signatures atypiques correspondent aux artefacts des spectrogrammes devant être contrôlés par les experts. Ces signatures particulières, n'étant pas relatives à des endommagements du moteur, sont d'intensités faibles à des niveaux proches du bruit et donc peuvent être difficilement détectables visuellement sur les spectrogrammes. La mise en place de méthodes permettant de mettre en évidence ces anomalies automatiquement est pertinente pour l'efficacité de l'analyse vibratoire.

Ces spectrogrammes sont de grande dimension, chaque spectrogramme consiste en 1.5 million de points environ, chaque point étant lié à une fréquence f et un régime N_2 . Les signatures inusuelles présentes sur les spectrogrammes consistent en général en une infime partie de ces derniers (une signature inusuelle peut ne consister qu'en une centaine de points). Etudier le spectrogramme dans sa globalité compliquerait la tâche de représentation et de détection car les signatures atypiques seraient noyées par de l'information vibratoire normale et non pertinente. Nous avons donc décidé de travailler sur une décomposition des spectrogrammes en patchs (sous-zones du spectrogramme) définis par des plages de fréquences et de régimes. Les signatures inusuelles restent largement minoritaires sur les patchs et possèdent différentes localisations sur les spectrogrammes. Une forte disproportion entre les patchs contenant des signatures atypiques (nettement moins nombreux) et ceux pouvant être considérés comme totalement normaux est présente dans la base de données. La plupart des éléments de la subdivision contiennent aucune ou très peu de données possédant une signature inusuelle sur cet élément. Une autre caractéristique des spectrogrammes est la grande variabilité aussi bien au niveau de la nature des signatures atypiques au sein d'un même patch que de l'intensité vibratoire.

Nous notons $Y_{Z_{\mathcal{K}_j}}^i \in \{0, 1\}$ et $Y_{f, N_2}^i \in \{0, 1\}$ les variables aléatoires représentant la classe normale (valeur 0) ou atypique (valeur 1) pour respectivement le patch j de la subdivision \mathcal{K} (section 1.5.2) et le point de fréquence f et de régime N_2 du spectrogramme issu du moteur i . La classe atypique correspond à la présence d'une signature inusuelle sur le patch ou le point considéré. Nous cherchons dans notre étude à estimer ces variables aléatoires pour chaque moteur. Nous verrons par la suite qu'il est possible de disposer d'une vérité terrain pour $Y_{Z_{\mathcal{K}_j}}^i$ et Y_{f, N_2}^i . La vérité terrain des patchs a été obtenue à partir de l'extraction des annotations manuelles des experts, la vérité terrain des différents points du patch a été établie manuellement pendant cette thèse pour quelques points uniquement.

Les approches préconisées

Le sujet de cette thèse correspond à une problématique de modélisation des patchs des spectrogrammes à des fins de classification. Pour répondre à la problématique, nous utilisons des méthodes de machine learning [49, 15], consistant à apprendre un modèle sur les données et à calibrer un classifieur. Les méthodes de classification sont généralement supervisées avec un apprentissage des modèles pour les différentes classes présentes dans la base de données entièrement labélisée et équilibrée (la base de données contient une proportion suffisante de données de chaque classe).

Dans notre cas d'étude, les approches supervisées ne sont pas adaptées. La base de données contient un unique spectrogramme issu d'un moteur endommagé et les patchs contenant des signatures inusuelles ne sont pas en quantité suffisante. De plus, la grande variabilité de ces signatures atypiques ainsi que le manque d'exemples des différentes signatures possibles sont des obstacles à la création d'un modèle d'anomalie (Figure 3). Nous cherchons à détecter les différentes signatures inusuelles présentes sur les spectrogrammes vibratoires, cependant notre base de données ne contient pas toutes les anomalies possibles étant donnée la complexité du système. Il est important que la méthodologie mise en place permette la détection de signatures inusuelles jamais observées dans la base de données.

Nous nous sommes orientés vers des approches non-supervisées de type *one-class* (Figure 3), un modèle est appris pour une unique classe de la base de données, les données sont comparées au modèle défini afin de tester leur appartenance à cette classe. Ces approches s'apparentent à la détection de nouveautés [90, 83, 84], d'anomalies [26] ou d'*outliers*¹ [25, 60]. Ces termes sont considérés comme des synonymes dans la littérature du fait de leur proximité. Ce type de méthode mesure la non similarité entre les données tests et les données d'apprentissage appartenant à une certaine classe, elle est utilisée lorsqu'il existe une grande disparité entre le nombre de données labélisées normales et le nombre de données labélisées inusuelles. Ce phénomène est assez fréquent dans le monde industriel où le produit manufacturé est trop complexe et/ou coûteux pour pouvoir être testé dans différentes conditions afin d'obtenir une base de données de cas atypiques. Ce type d'approche est également très présent dans le milieu médical. Ces méthodes s'appuient sur des données normales pour apprendre un modèle de normalité des données, et cherchent à détecter les différences entre ce modèle et les nouvelles données. Dans le cadre des spectrogrammes, il s'agit donc de représenter (à travers des modèles) leurs comportements normaux et usuels.

Nous avons utilisé deux approches différentes. Une première approche consiste à utiliser des dictionnaires [75, 13, 80] pour représenter les données, qui sont alors projetées sur un sous-espace engendré par les éléments du dictionnaire. Ce sous-espace est défini sur des données normales et permet une reconstruction des données sans signature inusuelle. Le dictionnaire est appris sur les patchs et non pas sur le spectrogramme entier. Les dictionnaires n'ayant pas été calibrés avec les signatures atypiques, ces dernières sont susceptibles d'être mal reconstruites, elles se trouveraient

1. donnée aberrante

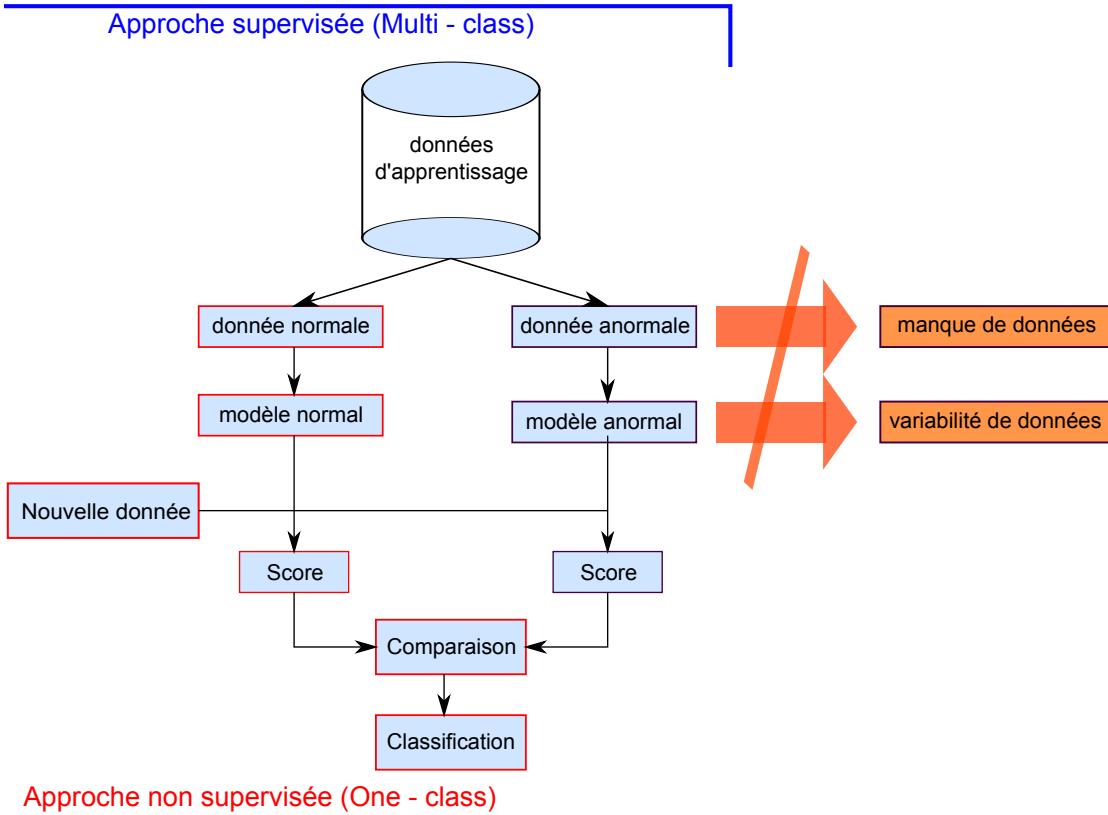


FIGURE 3 – Illustration des approches multi-class (en bleu) et one-class (en rouge). Les éléments sur le côté donnent les arguments dans nos données contre les approches multi-class.

donc dans les résidus de la reconstruction. Différents dictionnaires ont été étudiés :

- La Non-Negative Matrix Factorization (NMF) [70] : il s'agit d'un dictionnaire appris à partir des données (data-driven), ce dictionnaire s'adapte donc aux données,
- Les curvelets [22] : ce dictionnaire est défini par des fonctions et est donc indépendant du type de données.

Une seconde approche, localisée cette fois, reconside^re le spectrogramme comme une mesure physique en chaque point à une fréquence f et un régime N_2 . Ainsi la distribution de chaque point est apprise de manière non paramétrique [49] sur les spectrogrammes considérés comme normaux et des tests statistiques permettent de déterminer la conformité ou non des différents points du spectrogramme.

Nous avons ainsi conçu un outil d'aide à la décision qui permet d'analyser les spectrogrammes automatiquement et de guider les experts vers les signatures inusuelles présentes. Cet outil permet de donner à l'expert les points composant les signatures inusuelles, ce dernier devra alors analyser ces signatures afin de définir si la signature correspond à un endommagement et de juger de la sévérité de l'endommagement si celui-ci est présent.

Contributions

Cette thèse a donné lieu à trois articles présentés dans des conférences internationales :

- [3] *NMF-based Decomposition for Anomaly Detection applied to Vibration Analysis* (International Conference on Condition Monitoring and Machinery Failure Prevention Technologies CM-MFPT2015, Oxford, UK),
- [1] *Anomaly detection on spectrograms using data-driven and fixed dictionary representations* (European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning ESANN2016, Bruges, Belgique),
- [2] *Dictionary Comparison for Anomaly Detection on Aircraft Engine Spectrograms* (International Conference on Machine Learning and Data Mining MLDM2016, New-York, Etats-Unis).

Ces articles portent principalement sur les approches par dictionnaire, nous sommes en cours de rédaction d'un article de journal sur les différentes approches par dictionnaire et ponctuelles mises en place au cours de cette thèse pour la détection d'anomalies sur les spectrogrammes.

Nous avons mis en place dans cette thèse des outils d'aide à la décision pour l'analyse des spectrogrammes vibratoires permettant la mise en évidence des signatures atypiques. Les codes en langage MATLAB liées à cet outil ont été livrés à Safran Aircraft Engines.

Structure du mémoire

Ce mémoire de thèse est divisé en trois parties, chacune portant sur une échelle d'analyse différente de nos données.

La première partie correspond à une présentation globale des spectrogrammes et du type d'approche proposée. Le chapitre 1 contient une description de nos données et de l'analyse vibratoire. Le chapitre 2 donne un aperçu de la détection de nouveautés.

Nous présentons dans la seconde partie les approches par dictionnaire avec lesquelles nous étudions les spectrogrammes à l'échelle du patch. Le dictionnaire fixe des curvelets est étudié dans le chapitre 3. Le chapitre 4 porte sur le dictionnaire adaptatif déduit de la NMF.

Les spectrogrammes sont analysés ponctuellement dans la troisième partie à partir de la définition d'un modèle de normalité pour chaque point paramétré par une fréquence et un régime. Les différents points des spectrogrammes sont considérés comme indépendants dans le chapitre 5. La dépendance entre les points et leurs voisinages est prise en compte dans le chapitre 6.

Première partie

L'analyse vibratoire et détection de nouveautés

