

Représentations pour la détection d'anomalies

Application aux données vibratoires des
moteurs d'avions

Thèse Cifre de doctorat de l'Université Paris-Saclay
préparée à l'Ecole CentraleSupélec et l'Université Paris Diderot
en partenariat avec SAFRAN

Ecole doctorale n° 000 XXXXX
Spécialité de doctorat: voir spécialités reconnues par l'ED

Thèse présentée et soutenue à l'Université Paris Diderot,
le 3 juillet 2018,
par

M. Mina ABDEL SAYED

Composition du jury:

M. Nicolas VAYATIS Professeur, ENS Cachan — CMLA	Président
Mme. Nadine MARTIN Directeur de Recherche, CNRS — GIPSA-lab	Rapporteur
M. Younès Bennani Professeur, Université Paris13 — LIPN	Rapporteur
M. Erwan Le Pennec Professeur, Ecole Polytechnique — CMAP	Examineur
M. Jérôme Lacaille Expert Emérite, SAFRAN	Examineur
M. Mohamed El Badaoui Professeur, Université Jean Monnet — LAPSI & SAFRAN	Examineur
M. Giles FAY Professeur, Ecole CentraleSupélec — MICS	Directeur de thèse
Mme. Mathilde Mougeot Professeur, Université Paris Diderot — LPMA	Co-Directrice de thèse

Remerciements

Ce manuscrit commence donc par les derniers ajouts en date mais non moins importants remerciements. La fin d'un long parcours qui aura été des plus instructifs, des plus intéressants, mais également des plus éreintants. Je tiens à remercier toutes les personnes qui ont contribué, de près comme de loin, à l'achèvement de cette thèse qui a permis d'aboutir à ce manuscrit et me faire grandir.

Mes premiers remerciements vont à mes directeurs de thèse, Mathilde Mougeot et Gilles Faÿ pour leurs soutiens, leurs patiences et les discussions toujours intéressantes que nous avons pu avoir au cours de ces quatre dernières années. Gilles, bien que tu sois parti dans le milieu industriel bien trop tôt, tu as continué à prendre de ton temps afin de relire mes papiers et de voir pour discuter de la thèse. Mathilde, tu as pris le relais derrière Gilles et as su m'encadrer, me guider et me conseiller tout au long de la thèse. Je ne te remercierais jamais assez de m'avoir fait découvrir, à travers tes cours, le monde passionnant qu'est celui du machine learning.

Je remercie chaleureusement Nadine Martin et Younès Bennani d'avoir accepté de rapporter ma thèse, ainsi que pour leur relecture minutieuse et leurs commentaires pertinents. Je leurs en suis très reconnaissant.

Je remercie Nicolas Vayatis, Erwan Le Pennec, Jérôme Lacaille et Mohamed El Badaoui d'avoir accepté de faire partie de mon jury. Leur présence dans mon jury est un honneur. Je remercie plus particulièrement Jérôme Lacaille et Mohamed El Badaoui avec qui j'ai eu le plaisir de pouvoir travailler au sein de SAFRAN. Leurs disponibilités, leurs soutiens ainsi que leurs volontés de faire avancer la science ont été un réel plus au cours de ma thèse et restent un avantage pour les doctorants et futurs doctorants de SAFRAN.

Je suis arrivé dans les débuts de SafranTech, le centre de recherche de SAFRAN. Les trois années au sein cette entité m'ont permis de voir se construire et se développer ce pôle de compétence, cela a été pour moi une expérience très enrichissante. Je remerci donc Daniel Duclos, le responsable du pôle TSI (Technologies du Signal et de l'Information) de SafranTech qui, malgré une disponibilité limitée due à la création de l'équipe, a pris du temps afin que nous ayons des discussions intéressantes. Je remercie également Dohy Hong, responsable de l'équipe analyse de données, qui, bien que arrivé dans la dernière année de ma thèse, m'a supervisé au cours de cette dernière et m'a beaucoup apporté au cours de nos discussions. Je remercie également Julien Grifaton qui m'a permis de bien mieux comprendre le fonctionnement des moteurs d'avions ainsi que de leurs vibrations.

Ces trois années au sein de SAFRAN furent également l'occasion de faire des rencontres agréables et enrichissantes. Je remercie toute l'équipe TSI : Dany Abboud, Axel Barrau, Haithem Boussaid, Veronique Brion, Andrei Bursuc, Jorge Hernandez, Alexis Huck, Frédéric Jenson, Yann Le Guilloux, Marie Lienou, Vincent Morad, Estelle Parra, Sylvaine Picard, Edward Romero, David Tourais, Sebastien Razakarivony, Hélène Vorobieva. Je remercie également les membres des autres équipes pour les bons moments passés ensemble.

Je porte une attention particulière à mes compagnons thésard de SAFRAN : Elisa Hubert, Michel Moukari, Paul Chauchat, Edouard Pineau, Arturo Mendoza, Camille Chapdelaine qui ont rendu les moments difficiles plus supportables et moins solitaires. Je les remercie pour la bonne ambiance et leur sens de l'humour.

Bien que je n'ai pas passé beaucoup de temps dans les laboratoires, je tiens également à

remercier leurs membres pour m’avoir accueilli, ainsi que pour les discussions intéressantes que nous avons pu avoir. Je remercie mes compagnons de salle, doctorants du LPMA : Anna Benhamou, Cyril Benezet, Lucas Benigni, Laure Marêché et Assaf Shapira. Je tiens à remercier tout particulièrement l’assistante de l’école doctorale Paris Saclay Suzanne Thuron, pour sa patience et son aide indispensable pour les différentes procédures administratives.

Je remercie également mes nouveaux collègues au sein de la Société Générale : Lionnel Massoulard, David Duong, Thomas Seleck, Marc Van Ousdedhen, Pierre Titeux, Marion Cabrol, Fabien Vavrand, Guillaume , Alexandre Miot. **peut être ajouté Gilles, Marwen et Pasha.**

Les années de thèse auraient été beaucoup plus dures qu’elles ne l’ont été sans le soutien, la bonne humeur de mes amis proches : Tristan Delcourt, Michael Richard, Richard Muller, Christophe Calduch, Elias Artuso, Frederic Khatami, en particulier Jennifer Lamalle et Fabien Lamalle.

Cette thèse n’aurait pas été la même sans le soutien de ma famille, je vous remercie donc maman, papa, Dina et John. Je remercie également ceux que je considère comme ma seconde famille, Yasmina Ben Aoun (que je considère comme ma seconde mère), Mokrane Kadri, Charlie, ainsi que tout particulièrement mes amis de très longues dates Cédric Ben Aoun et Samedani. Je remercie également ma belle-famille : Philippe Malherbe, Caroline Malherbe, Arnaud Messé pour leur soutien. Une dernière pensée me vient pour Brigitte Cauchy.

Enfin, cette thèse n’aurait en aucun cas été possible sans le soutien inconditionnel, la patience extraordinaire et les conseils de ma compagne Véronique Malherbe au quotidien. Tous les mots du monde ne suffiraient à lui exprimer ma gratitude.

Table des matières

Remerciements	III
Table des matières	VIII
Résumé	1
Abstract	1
Introduction	3
Le contexte industriel	3
Problématique	4
Les approches préconisées	6
Contributions	8
Structure du mémoire	8
I L’analyse vibratoire et détection de nouveautés	9
1 L’analyse et les données vibratoires	13
1.1 Introduction	13
1.2 Les moteurs d’avions	13
1.2.1 Les caractéristiques des moteurs	13
1.2.2 Les vibrations du moteur	15
1.2.3 L’acquisition des mesures vibratoires et des vitesses de rotation	16
1.3 Conversion des signaux temporels en spectrogrammes	16
1.3.1 Intérêt de cette conversion	16
1.3.2 La transformation du signal temporel en spectrogramme	18
1.3.3 Gains et limites de cette représentation	19
1.4 Construction de la base de données	21
1.4.1 L’annotation manuelle des experts	21
1.4.2 Extraction automatique des zones anormales sur les données textuelles . .	23
1.4.3 La base de données enrichie	24
1.5 Étude des spectrogrammes par patch	25
1.5.1 Localisation des signatures inusuelles sur le spectrogramme	25
1.5.2 Subdivision du spectrogramme en patches	26
1.5.3 Labélisation ponctuelle du patch - enrichissement de la base de données .	29
1.5.4 La grande variabilité des signatures inusuelles	30
1.6 L’état de l’art de l’analyse vibratoire	31
1.6.1 L’état de l’art provenant de la littérature	31
1.6.2 Les algorithmes d’analyse vibratoire de Safran Aircraft Engines	34
1.7 Une première approche de détection	36
1.7.1 Représentation des patches par leurs histogrammes d’intensités vibratoires	37

1.7.2	Représentation des histogrammes dans un espace réduit	39
1.8	Conclusions	39
2	La détection de nouveautés	41
2.1	Définition	41
2.2	État de l'art de la détection de nouveautés	42
2.2.1	Les approches probabilistes	43
2.2.2	Les approches basées sur les distances	44
2.2.3	Les approches basées sur la reconstruction des données	44
2.2.4	Les approches basées sur la caractérisation des limites des données normales	45
2.2.5	Les approches basées sur la théorie de l'information	46
2.3	La détection de nouveautés appliquée aux données vibratoires	46
2.3.1	Application aux données temporelles et fréquentielles	46
2.3.2	Application sur les harmoniques du signal	47
2.3.3	Application aux spectrogrammes	48
2.4	Caractérisation de la base de données construite	49
2.4.1	Répartition des données en sous-ensembles	49
2.4.2	Visualisation des résultats	51
II	Les approches de représentation globale par dictionnaire	53
3	Représentation par dictionnaire fixe - les curvelets	57
3.1	Introduction	57
3.1.1	La représentation par dictionnaire	57
3.1.2	Les dictionnaires non-adaptatifs	59
3.2	La transformée en curvelet	60
3.2.1	La transformée en ridgelet	60
3.2.2	Les ridgelets multi-échelles	61
3.2.3	La construction de la transformée en curvelet	62
3.2.4	Le dictionnaire des curvelets	64
3.3	Application des curvelets aux spectrogrammes vibratoires	64
3.3.1	Caractérisation des raies vibratoires à partir des curvelets	64
3.3.2	Comparaison des représentations en curvelet	66
3.4	Normalité définie dans le dictionnaire des curvelets	68
3.4.1	Le modèle de normalité	68
3.4.2	Normalité définie par optimisation avec contraintes de parcimonie	72
3.4.3	Comparaison des supports normaux	73
3.5	Le score de détection d'anomalies sur le patch	75
3.5.1	Les scores de détection	75
3.5.2	Le score de normalité	76
3.5.3	Résultats sur la base de test	78
3.6	L'étude des résidus ponctuels	80
3.6.1	Les résidus du modèle de normalité $\mathcal{D}_{Supp^*}^c$	80
3.6.2	La détection d'anomalies	81
3.6.3	Calibration des paramètres des modèles sur Ω_{Val}	83
3.6.4	Résultats sur la base de test Ω_{Test}	89
3.7	Conclusions	93

4	Représentation par dictionnaire data-driven-NMF	95
4.1	Les dictionnaires adaptatifs/data-driven	95
4.1.1	Définition mathématique du problème	95
4.1.2	Les méthodes de résolution	96
4.1.3	Quelques exemples de dictionnaires data-driven (adaptatifs)	97
4.1.4	Les dictionnaires appris pour caractériser les spectrogrammes	98
4.2	Non-Negative Matrix Factorization (NMF)	99
4.2.1	Formulation mathématique	99
4.2.2	Résolution de la problématique	99
4.3	Le modèle de normalité défini à partir de la NMF	100
4.3.1	L'apprentissage du dictionnaire	100
4.3.2	Définition du rang du dictionnaire	101
4.3.3	Représentation de la normalité à partir du dictionnaire de la NMF	102
4.4	Détection d'anomalies sur les patches	105
4.4.1	Les scores de détection	105
4.4.2	Résultats sur la base de test	108
4.5	Les erreurs ponctuelles issues du dictionnaire de la NMF	109
4.5.1	Les résidus de la NMF	109
4.5.2	La détection des points inusuels	110
4.5.3	Calibration des paramètres	112
4.5.4	Résultats sur la base de test Ω_{Test}	115
4.5.5	Complémentarité des approches adaptatives et non-adaptatives	116
4.6	Conclusions	119
III	Analyse ponctuelle des spectrogrammes	121
5	Analyse ponctuelle indépendante	125
5.1	Introduction	125
5.1.1	Considération ponctuelle des points des spectrogrammes	125
5.1.2	La base de données	126
5.1.3	Les modèles de normalité	127
5.2	Modélisation paramétrique de la distribution de normalité	127
5.2.1	Le modèle de normalité	127
5.2.2	Le score de détection	128
5.2.3	Calibration des seuils de détection sur la base de validation Ω_{Val}	129
5.2.4	Résultats sur la base de test Ω_{Test}	130
5.3	Estimation non paramétrique de la densité par noyau	135
5.3.1	Formulation	135
5.3.2	L'estimation de la densité par noyau gamma	136
5.3.3	Estimation de l'échelle du noyau	137
5.4	Distribution de normalité estimée par les noyaux gaussiens	139
5.4.1	Le modèle de normalité	139
5.4.2	Le score de détection	140
5.4.3	Calibration des seuils	142
5.4.4	Résultats sur la base de test Ω_{Test}	143
5.5	Distribution de normalité estimée par les noyaux gamma	147
5.5.1	Le modèle de normalité	147
5.5.2	Le score de détection	148
5.5.3	Calibration des seuils de détection	150
5.5.4	Résultats sur la base de test Ω_{Test}	151
5.6	Conclusions	153

6	Analyse ponctuelle conditionnelle au voisinage	155
6.1	Considération des points voisins	155
6.1.1	Dépendance par rapport au voisinage des points	155
6.1.2	Modèle de normalité par rapport au voisinage	156
6.2	Estimation de densité conditionnelle par noyau	157
6.2.1	L'estimation de densité par noyau dans un cadre multidimensionnel	157
6.2.2	Calibration de la matrice d'échelle	158
6.2.3	L'estimation de la densité conditionnelle par noyau	159
6.3	Le modèle de normalité défini à partir du voisinage d'ordre 1	160
6.3.1	La structure du voisinage	160
6.3.2	Le modèle de normalité	161
6.3.3	Le score de détection	162
6.3.4	Calibration du modèle	165
6.3.5	Résultats sur la base de test Ω_{Test}	167
6.4	Normalité défini en fonction de la direction du voisinage	169
6.4.1	Le voisinage directionnel	169
6.4.2	Le modèle de normalité	170
6.4.3	Le score de détection	171
6.4.4	La caractérisation des signatures inusuelles	172
6.4.5	Calibration du seuil de détection	173
6.4.6	Résultats sur la base de test Ω_{Test}	173
6.5	Fusion des différentes approches	180
6.5.1	Comparaison des approches par dictionnaire et ponctuelles	180
6.5.2	Fusion des approches par dictionnaire et ponctuelles	180
6.6	Conclusions	184
	Conclusions et perspectives	187
	Conclusions	187
	Perspectives	191
	Bibliographie	193
	Table des figures	200
	Liste des tableaux	207
	Annexes	209
	A Les tests multiples	209

Liste des symboles

f	Fréquence
N_2	Régime de l'arbre haute pression du moteur
N_1	Régime de l'arbre basse pression du moteur
S^i	Spectrogramme vibratoire du moteur i
S_{f,N_2}^i	Intensité vibratoire à la fréquence f et au régime N_2 du moteur i
\mathcal{K}	Subdivision du spectrogramme en patches
\mathcal{K}_j	Élément j de la subdivision \mathcal{K} du spectrogramme en patches
$Z_{\mathcal{K}_j}^i$	Patch correspondant à la l'élément j de la subdivision \mathcal{K} extrait du spectrogramme i
$\hat{Z}_{\mathcal{K}_j}^i$	Estimation normale du patch correspondant à la l'élément j de la subdivision \mathcal{K} extrait du spectrogramme i
Ω_{App}^j	Base d'apprentissage du patch j de la subdivision \mathcal{K}
Ω_{Val}^j	Base de validation du patch j de la subdivision \mathcal{K}
Ω_{Test}^j	Base de test du patch j de la subdivision \mathcal{K}
Ω_{App}^{f,N_2}	Base d'apprentissage des points aux coordonnées (f, N_2)
Ω_{Val}^{f,N_2}	Base de validation des points aux coordonnées (f, N_2)
Ω_{Test}^{f,N_2}	Base de test des points aux coordonnées (f, N_2)
$Y_{Z_{\mathcal{K}_j}}^i$	Vérité terrain de la classe normale ou atypique pour le spectrogramme i du patch Z_j provenant de l'élément j de la subdivision \mathcal{K}
$\hat{Y}_{Z_{\mathcal{K}_j}}^i$	La classe normale ou atypique estimée pour le spectrogramme i du patch Z_j provenant de l'élément j de la subdivision \mathcal{K}
Y_{f,N_2}^i	Vérité terrain sur la classe normale ou atypique du point de coordonnées (f, N_2) pour le spectrogramme i
\tilde{Y}_{f,N_2}^i	La classe, normale ou atypique, estimée sans filtrage par voisinage du point de coordonnées (f, N_2) pour le spectrogramme i
\hat{Y}_{f,N_2}^i	La classe, normale ou atypique, estimée avec filtrage par voisinage du point de coordonnées (f, N_2) pour le spectrogramme i
\mathcal{V}_{f,N_2}	Voisinage du point de coordonnées (f, N_2)
\mathcal{V}_{f,N_2}^k	Voisinage du point de coordonnées (f, N_2) dans la direction k
\mathcal{D}^C	Dictionnaire des curvelets
\mathcal{D}^{NMF}	Dictionnaire de la NMF

$SuppZ_{\mathcal{K}_j}^i$	Atomes des curvelets activés pour le patch $Z_{\mathcal{K}_j}^i$
$Supp^*$	Atomes des curvelets activés pour un pourcentage des données d'un patch de la base d'apprentissage
$R(Z_{\mathcal{K}_j}^i)$	Résidus de reconstruction du patch $Z_{\mathcal{K}_j}^i$
$R^+(Z_{\mathcal{K}_j}^i)$	Résidus positifs de reconstruction du patch $Z_{\mathcal{K}_j}^i$
$R^-(Z_{\mathcal{K}_j}^i)$	Résidus négatifs de reconstruction du patch $Z_{\mathcal{K}_j}^i$
$\mathcal{N}(\mu, \sigma)$	Loi et densité gaussienne de paramètre (μ, σ)
f_X	densité de la variable aléatoire X
$f_{\mathcal{N}}$	Densité de la loi gaussienne
$F_{\mathcal{N}}$	Fonction de répartition de la loi gaussienne
$K^{\mathcal{N}}$	Noyau gaussien
K^{Γ}	Noyau gamma
h	Echelle du noyau en une dimension
H	Matrice d'échelle du noyau multidimensionnel
\mathcal{F}	Transformée de Fourier
$\mathbf{1}$	Fonction indicatrice
$\ x\ _0$	norme 0, $\ x\ _0 = \text{card}(i : x_i \neq 0)$ avec $x = (x_1, \dots, x_n)$
$\ x\ _1$	norme 1, $\ x\ _1 = \sum x_i $ avec $x = (x_1, \dots, x_n)$
$\ x\ _2$	norme 2, $\ x\ _2 = \sqrt{\sum x_i ^2}$ avec $x = (x_1, \dots, x_n)$