

N° d'ordre : D.U :
EDSPIC :

UNIVERSITE BLAISE PASCAL - CLERMONT II
ECOLE DOCTORALE
SCIENCES POUR L'INGENIEUR DE CLERMONT-FERRAND

Thèse

Présentée par

Zahra HAMOU MAMAR

Pour obtenir le grade de

DOCTEUR D'UNIVERSITE

Spécialité : Informatique

Analyse Temps-Échelle et Reconnaissance des Formes pour le Diagnostic du Système de Guidage d'un Tramway sur Pneumatiques

Soutenue le 18/07/2008 devant le jury composé de :

M. A. QUILLIOT	Professeur Clermont Ferrand II
M. S. CANU	Professeur INSA Rouen
M. P. AKNIN	Directeur de recherche INRETS-LTN Paris
M. A. AUSSEM	Professeur Université Lyon 1
M. P. CHAINAIS	Maitre de Conférence Clermont Ferrand II
M. P. CHARLAT	Responsable du projet Tramway SMTC

À mes parents.

N'ABANDONNE JAMAIS !!!

***Quand les choses vont vraiment mal comme elles savent si bien le faire
quelquefois,***

*Quand la route sur laquelle tu chemines péniblement semble s'achever au sommet
d'une colline,
Quand les fonds sont bas et que les dettes culminent,
Tu voudrais sourire, et tu dois pousser des soupirs, quand le souci te pousse dans
la déprime,*

Repose-toi si tu veux... mais n'abandonne pas.

*La vie est si étrange avec ses revers et ses détours comme chacun de nous a pu
l'apprendre un jour, et beaucoup qui ont été abattus par un échec auraient pu
réussir s'ils avaient persévéré...*

***N'abandonne pas même si tout semble aller lentement, car un autre
souffle peut apporter la réussite.***

*Le succès n'est que l'envers de l'échec et tu ne peux jamais savoir à quelle distance
se trouve le but, qui peut être très proche alors qu'il te semble si lointain.*

***Aussi, continue la lutte au plus fort du combat car c'est quand tout te
semble perdu que tu ne dois pas abandonner.***

Auteur inconnu (Traduction d'un poème, anglais).

Remerciements

Je tiens tout d'abord à remercier mes directeurs de thèse, Alexandre Aussem et Pierre Chainais, pour avoir partagé avec moi leurs multiples compétences scientifiques et pour leur disponibilité, leur écoute et leur patience tout au long de ce travail de thèse.

Je remercie Mr. François Moulin, Directeur du SMTC, et Mr. Patrice Charlat, chef du projet Tramway, de m'avoir donné ma chance et d'avoir cru en mes compétences. Je remercie également tous mes collègues du SMTC sans exception pour leurs encouragements et pour tout le plaisir que j'ai eu à travailler auprès d'eux.

Mes remerciements s'adressent aussi à Monsieur Alain Quilliot, Directeur de l'Institut Supérieur d'Informatique de Modélisation et de leurs Applications, qui a eu la gentillesse de m'accueillir au sein du laboratoire LIMOS et me fait l'honneur de présider ce jury.

Je remercie les membres de mon jury de thèse, Stéphane Canu et Patrice Anin, pour avoir accepté d'être les rapporteurs de ce mémoire, pour l'intérêt qu'ils ont porté à mon travail et pour l'avoir enrichi de leurs remarques.

*Je n'aurais pas pu rêver meilleur environnement et meilleure compagnie que celle de mes amis pendant ces années de dur labeur. Je ne remercierai jamais assez Catherine, Michaël et Odile pour leur aide, leur soutien chaleureux et bienveillant et d'avoir été comme une deuxième famille pour moi. Il faudrait des pages et des pages pour le remercier tout le monde convenablement, je vous dis donc simplement "**Merci du fond du Coeur**".*

Je tiens à remercier particulièrement Nicolai Tchernev pour son soutien moral, son aide inestimable, et pour m'avoir permis d'exercer le métier qui me tenait tant à coeur, à savoir l'enseignement. Un grand merci également à mes collègues de l'Université d'Auvergne pour m'avoir adoptée et considérée comme l'une des leurs.

Je remercie Jonas pour ces nombreuses relectures fructueuses et pour m'avoir soutenu tout au long de cette thèse.

Je tiens aussi à remercier mes collègues de Logica et du service Supply Chain de Michelin, où je me sens si bien du fait de la qualité des rapports humains et de la diversité des âges, des origines et des compétences, qui s'avèrent très enrichissantes.

Et merci également à mes médecins, A. Schmit, N. Delage et M.T. Danyce, du

centre Anti Douleurs de l'hôpital Gabriel Montpied, pour le temps qu'ils m'ont consacré, pour leur soutien et leur investissement.

*Je dois beaucoup à Paulo Coelho qui, à travers son livre **L'Alchimiste**, m'a appris qu'on peut adopter face au monde deux postures fondamentalement différentes : soit on regarde le monde comme les malheureuses victimes d'un destin hasardeux, soit on se considère comme des aventuriers en quête d'un trésor. L'alchimiste a été pour moi une étape très importante dans ma vie, et j'aimerais partager avec vous l'extrait qui, à mon avis, résume l'objectif de ce sublime conte philosophique :*

"Une quête commence toujours par la Chance du Débutant. Et s'achève toujours par l'Épreuve du Conquérant." [Paulo Coelho] - Extrait de *L'Alchimiste*.

Enfin, un énorme merci à mes soeurs, mon frère, mes parents et mon mari, qui sont ma joie, ma vie, ma foi et ma raison d'être. Merci d'être présent dans mes joies et mes peines et tout simplement d'être ma famille.

Table des matières

Remerciements	i
Table des matières	ii
Introduction	1
1 Introduction générale	1
1.1 Modernisation des transports en commun	1
1.2 Transport en commun hybride	2
1.2.1 Les avantages	2
1.2.2 Les modes de guidage	2
1.2.3 La naissance du Translohr	2
1.3 Le Translohr, <i>bus ou tramway</i> ?	4
1.3.1 Historique	4
1.3.2 Transport sur voies réservées (TVR)	4
1.4 Système de guidage	6
1.4.1 Le principe de fonctionnement	6
1.4.2 La problématique industrielle	7
1.5 Le projet SYSCOGA : <i>SYStème de COntôle des GAlets</i>	7
1.5.1 La sûreté de fonctionnement	7
1.5.2 Le détecteur SysCoGa	8
1.5.3 Classes de défauts	9
1.6 Organisation de la thèse	13
2 Diagnostic par reconnaissance des formes	17
2.1 L'architecture d'un système de surveillance	17
2.1.1 Systèmes de surveillance	17
2.1.2 La perception	19
2.1.3 Le diagnostic	19
2.1.4 Quelques approches pour le diagnostic	20
2.2 Système de diagnostic par reconnaissance des formes	22
2.2.1 Les approches de reconnaissance de formes	22
2.2.2 Principe et formulation	23
2.3 Construction d'un système de reconnaissance de formes	23
2.3.1 Phase d'analyse	24
2.3.2 Phase de choix d'une méthode de décision	26
2.3.3 Phase d'exploitation	28
2.4 Approches de discrimination non-paramétrique	30
2.4.1 Principe de l'estimation de densité non paramétrique	31

2.4.2	Les arbres de décision	33
2.4.3	L'estimation des fonctions de classement	34
2.5	Outils pour la classification	35
2.5.1	Les réseaux de neurones, concepts et définitions	35
2.5.2	Les techniques de surveillance par réseaux de neurones	37
2.5.3	Le Perceptron Multi Couches (MLP)	38
2.5.4	Le réseau de neurones à fonctions de base radiales (RBF)	40
2.5.5	Le boosting probabiliste : Adaboost	43
2.5.6	Régression logistique	45
2.6	Les Machines à Vecteurs Supports (SVM)	46
2.6.1	Motivation	46
2.6.2	Principe des SVM linéaires	47
2.6.3	Principe des SVM non linéaires	50
2.6.4	L'optimisation quadratique sous contraintes pour l'apprentissage des SVM	51
	Conclusion	54
3	Acquisition des signaux vibratoires	57
3.1	Système d'acquisition	57
3.1.1	Configuration matérielle de base	57
3.1.2	Intégration sur site	58
3.2	Version finale du détecteur SYSCOGA	59
3.2.1	Installation sur la voie	59
3.2.2	Installation sur l'appareil de voie (l'ADV)	59
3.3	Étude de faisabilité	62
3.3.1	Premiers essais, premières difficultés	62
3.3.2	Configuration technique du véhicule et des défauts	65
3.3.3	Problèmes rencontrés	65
3.3.4	Premières analyses	67
3.3.5	Redirection de la thèse	68
3.4	Première analyse des signaux	68
3.4.1	Descriptif des signaux SYSCOGA-I à IV (sur la voie)	69
3.4.2	Descriptif des signaux SYSCOGA-V (sur l'ADV)	74
	Conclusion	84
4	Analyse temps-échelle	85
4.1	Les représentations temps-fréquence et temps-échelle	86
4.1.1	La transformée de Fourier	86
4.1.2	La transformée de Fourier à fenêtre glissante	87
4.1.3	Distribution de Wigner-Ville	90
4.1.4	La transformée en ondelettes continue	91
4.2	Choix de la représentation temps-fréquence	94
4.3	Étude des scalogrammes	95
4.3.1	Description qualitative des scalogrammes	95
4.3.2	Découpage temps-échelle des scalogrammes	96
4.4	Vers la classification	98
4.4.1	La dimension de l'espace de représentation	98
4.4.2	Interprétation des représentations temps-fréquence	99
4.5	Tentative de segmentation des images	100

4.5.1	Segmentation des scalogrammes par binarisation	100
4.5.2	Inconvénients et avantages de la binarisation	104
	Conclusion	107
5	Réduction de la dimension	109
5.1	Introduction	109
5.2	Sélection de variables	111
5.2.1	Principe de la sélection	111
5.2.2	Les étapes de la sélection de variables	111
5.3	Extraction des variables	115
5.3.1	Décomposition en valeurs singulières	115
5.4	La décomposition en valeurs singulières des matrices des coefficients en ondelettes CWTK	117
5.4.1	Formulation du problème	117
5.4.2	Étude de la décomposition en valeurs singulières d'un passage complet	118
5.5	Définition de signatures basées sur les vecteurs singuliers	125
5.5.1	Introduction	125
5.5.2	La pseudo signature de puissance (PPS)	126
5.5.3	Coefficient de corrélation et indicateur de distance basée sur la PPS	130
5.5.4	Fonction de distribution de probabilité des vecteurs singuliers	135
	Conclusion	141
6	Classification supervisée	143
6.1	Introduction	143
6.2	Démarche utilisée pour le diagnostic	145
6.3	Analyses et résultats	147
6.3.1	Performance des classifieurs	147
6.3.2	La classification multi-classe avec l'approche 1-vs-1	148
6.4	Spécifications de la T2C	158
6.4.1	Types de défauts retenus	158
6.4.2	Nouvelle démarche suivie et données utilisées	158
6.5	Détection de la présence d'un défaut	160
6.5.1	Classification avec les valeurs singulières	160
6.5.2	Classification avec les pseudo-signatures de puissances	160
6.6	Classification par SVM	165
6.6.1	Paramètres d'optimisation du calcul des SVM	166
6.6.2	Classification SVM avec les valeurs singulières	167
6.6.3	Classification SVM avec les pseudo-signatures de puissances .	172
	Conclusion	176
	Conclusions et perspectives	177
6.6.4	Perspectives	182
	Annexes	183
A	Planning prévisionnel du projet Syscoga	185

B Campagnes d'essais	187
B.1 Essais Syscoga-II	187
B.2 Essais Syscoga-IIbis	187
B.3 Essais Syscoga-IV	188
C Les mesures de l'erreur estimée pour un classifieur	191
C.1 Les mesures générales	191
C.2 Les mesures d'exactitude par classe	192
Bibliographie	195

INTRODUCTION GÉNÉRALE

1.1 Modernisation des transports en commun, *design et confort*

Depuis quelques années plusieurs collectivités locales ont adopté une nouvelle politique de transport, en incitant les citoyens à emprunter les moyens de transport en commun pour se déplacer en ville et reconquérir les centres historiques. Ainsi est réapparu le tramway synonyme de transport moderne, efficace, confortable, rapide et silencieux. Cependant, ce nouveau mode de vie, imposé par le mouvement de sensibilisation vis-à-vis de l'environnement et la nécessité de réduire les émissions polluantes en ville, entraîne aussi des inconvénients que d'avantages. Malheureusement, cet accroissement du trafic ferroviaire dans les villes induit une nuisance sonore qui nuit à la santé psychique et à la tranquillité des habitants. Par conséquent, le bruit revient en tête des préoccupations des autorités.

Le bruit généré se compose de plusieurs types dont le plus prépondérant est le bruit de roulement causé principalement par le contact roue/rail. Plusieurs recherches ont été consacrées, notamment à l'utilisation de nouveaux modes de transports innovant par leurs systèmes de roulement et/ou de guidage, afin de diminuer cette source de nuisance. Avec cette ère de renouveau, on observe la naissance de nouveaux moyens de transport dits "**guidés**". On entend en général par guidage, les moyens qui permettent d'imposer une trajectoire sans action des conducteurs. Le moyen de guidage le plus répandu est la combinaison de la voie ferrée et de la roue en acier. Cette solution est utilisée sur le tramway classique qui avait presque disparu en France dans les années 1960 et qui est revenu sur la scène urbaine au milieu des années 80.

Plus récemment, on assiste à une innovation technique importante orientée vers des systèmes hybrides, en tirant parti des avantages du tramway sans en subir les contraintes (nuisances sonores, etc.) ; ainsi est né le tramway sur pneumatiques. Ce nouveau mode de transport utilise des roues ou galets de guidage, distincts des roues porteuses, et coopérant avec des barres de guidage latérales ou un rail de guidage central.

1.2 Transport en commun hybride

1.2.1 Les avantages

Le tramway sur pneumatiques est un mode de transport en commun hybride (BHNS) situé entre le tramway, avec lequel il partage le guidage et l'alimentation électrique, et l'autobus et le trolleybus, avec lesquels il partage le roulement sur pneumatiques. Le véhicule ne roule pas sur des rails mais sur une chaussée ; le rail lui sert de guidage. Devant le succès du tramway et l'effet de mode, les concepteurs ont diversifié les modèles de tramway sur pneumatiques pour satisfaire les utilisateurs des transports en commun (design, confort et discrétion) sans pour autant porter atteinte aux normes de sécurité.

Il dispose de deux principaux atouts : le coût d'investissement qui est inférieur à celui d'un tramway classique et pour certains types de véhicules la possibilité de débrayage. Étant équipée d'un groupe électrogène, la rame peut quitter ponctuellement sa trajectoire en mode trolleybus ou autobus en cas d'accident, pour éviter un encombrement ou pour circuler en mode routier. Les rames peuvent s'affranchir alors de guidage et de ligne de contact aérienne (LAC). Circulant sur pneus, ces rames sont capables de franchir des pentes inaccessibles au tramway classique sans surcoûts et sans adaptations techniques particulières (infrastructure). Un autre avantage annoncé du tramway sur pneus est sa meilleure adhérence au sol qui lui procure une meilleure accélération et un freinage plus efficace. Enfin, son petit gabarit lui permet de s'insérer plus aisément dans les villes aux rues étroites.

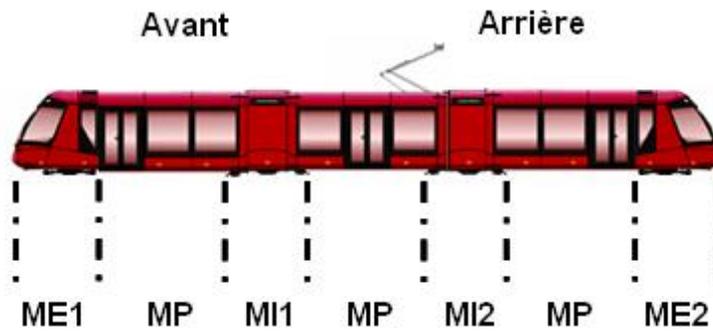
1.2.2 Les modes de guidage

Un tramway sur pneumatiques peut être guidé soit par un galet suivant un rail central, soit par une cellule optique suivant un trait tracé sur son chemin (guidage optique). En général on peut distinguer trois catégories :

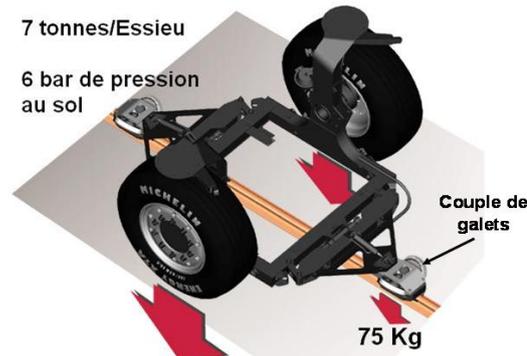
- guidage optique à l'aide d'une caméra installée à l'avant de la rame qui repère un marquage peint sur la chaussée ;
- guidage par rail central débrayable (exemple : le système de Nancy ou Caen) ;
- guidage par rail central non débrayable (exemple : la nouvelle ligne de Clermont-Ferrand).

1.2.3 La naissance du Translohr

Un système différent est exploité à Clermont-Ferrand, le **Translohr**, développé par la société Lohr Industrie. C'est un tramway sur pneumatiques à propulsion électrique alimenté par ligne aérienne de contact (LAC) mais il peut également recevoir un système de batteries pour circuler dans des zones où les câbles d'alimentation sont indésirables. Le véhicule consiste en un ensemble articulé de 2 (STE2) à 5 éléments



(a)



(b)

FIG. 1.1 – (a) Les différents modules d'un Translohr : ME, MP et MI représentent respectivement les modules extrêmes, passagers et intermédiaires. (b) Poids à vide par essieu et par couple de galets.

(STE5), avec une cabine de conduite à chaque extrémité, figure 1.1(a). Le confort est garanti grâce à la discrétion des systèmes de roulement et de guidage par rail central, figure 1.1(b). Pour limiter les nuisances sonores, les galets sont recouverts d'un bandage composite sensible à l'apparition d'usures. Les essieux, au nombre de 3 à 6, sont situés entre les modules (système de boggie-pneu). Sa longueur peut alors varier de $18m$, pour un STE2, jusqu'à $50m$ soit deux STE3 couplés.

Un Translohr STE4, par exemple, figure 1.1(a) est composé de :

- 2 modules d'extrémité en acier pour mieux absorber et amortir les chocs. Ils sont portés par les essieux moteurs avec les guidages ;
- 3 modules passagers (MP) en aluminium et matériaux composites, ce qui permet une réduction du poids de l'ensemble. Ils sont soit boulonnés aux modules d'extrémités d'un côté et articulés au module d'intercirculation (MI), soit articulés des deux côtés aux modules d'intercirculations. Ils comprennent un essieu porteur et son guidage ;
- 2 modules d'intercirculation (MI) ;
- 4 essieux dont 2 essieux extrêmes moteurs (EEM). Le poids à vide d'un essieu moteur est de 6.2 tonnes alors qu'un essieu non moteur (EI) pèse 4.2 tonnes. Il est plus léger que son équivalent sur rail (7 tonnes par essieu), figure 1.1(b).

1.3 Le Translohr, *bus ou tramway* ?

1.3.1 Historique

Le Translohr s'inscrit dans la nouvelle génération des tramways sur pneumatiques. Le confort est garanti grâce à la discrétion de son système de roulement : seuls les pneus absorbent la charge exercée sur le rail, qui devient ainsi un rail de guidage non porteur. Chaque essieu supporte deux roues pneumatiques et un système de guidage par rail central non débrayable, figure 1.2(a). Ce système est directement inspiré du *monorail Larmanjat*, dont une ligne fut mise en service entre Le Raincy et Montfermeil en 1868. Son moteur principal est électrique et l'alimentation se fait par lignes aériennes (LAC) de traction électrique (caténaire et pantographe). L'utilisation de pneus procure une traction supérieure au rail, le Translohr peut ainsi franchir des pentes de 13% (8% testé et confirmé à Clermont-Ferrand). Bien que le Translohr puisse être considéré comme un bus guidé, il ressemble beaucoup à un tramway classique.

La première ligne utilisant le Translohr est à Clermont-Ferrand. Le parc est composé de 20 rames STE4, de couleur "fleur de lave". D'autres villes utilisent également le Translohr comme Padoue, L'Aquila et la ligne pour relier Venise à Mestre, sur le continent. Dernièrement, le 6 décembre 2006, Lohr Industrie a inauguré l'infrastructure de la ligne chinoise de 8 km de la zone de développement économique de Tianjin. Elle utilisera 8 rames identiques à celles de Clermont-Ferrand.

1.3.2 Transport sur voies réservées (TVR)

De par l'usage de pneumatiques et l'appellation "tram", le concept du Translohr s'approche du transports sur voies réservées (TVR) du constructeur Bombardier. En réalité, en plus du guidage, figure 1.3, de nombreuses différences les opposent. Les TVR de Nancy (à perches) ou Caen (à pantographe) sont des véhicules routiers, immatriculés comme des autobus, et dont le guidage est optionnel. Ils sont bimode, munis de moteurs diesel et d'un système de direction. Il en résulte qu'il leur est impossible, contrairement au Translohr :

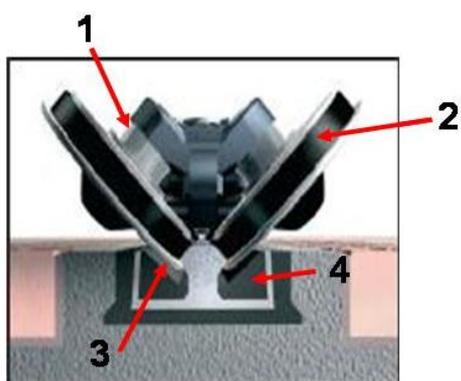
- de circuler en unités multiples (plusieurs rames accouplées) comme des tramways ;
- d'être rallongés, leur longueur étant tributaire de la législation routière ;
- d'être réversibles : ils doivent disposer de boucles de retournement à chaque terminus.

À l'inverse, le Translohr est un véhicule ferroviaire :

- il est réversible et peut ainsi circuler dans les deux sens ; bidirectionnel car il possède une cabine de conduite à chaque extrémité ;
 - il n'est pas immatriculé ;
-



(a)



(b)



(c)

FIG. 1.2 – Système de roulement composé de deux pneumatiques et un système de guidage. (b) Un couple de galets en V : 1– la fusée, 2– la bande de roulement, 3– le mentonnet et 4– caoutchouc du rail. (c) Le patin de retour de courant : essieu intermédiaire.

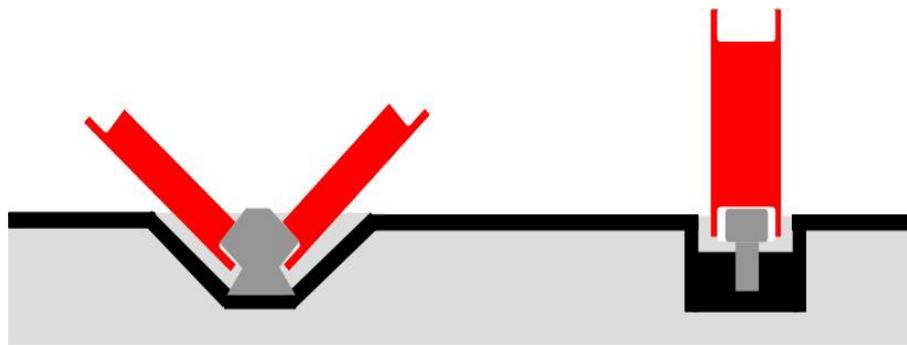


FIG. 1.3 – Guidage Translohr, à gauche, et TVR à droite.

- son guidage est permanent et n'est pas bimode (si ce n'est les batteries) ;
- il peut circuler en unités multiples ou être rallongé (gamme évolutive) ;
- le Translohr n'est pas enterré. Il roule sur la chaussée de la voirie urbaine et non sur des pistes de roulements spécifiques. Cela contraint à le confondre avec des véhicules routiers.

1.4 Système de guidage

1.4.1 Le principe de fonctionnement

Le Translohr est guidé par un rail central. Chaque essieu est orienté par un bras et deux galets directeurs (GD), qui enserrant le rail en V, situés vers l'avant dans le sens de la marche. Un deuxième jeu de galets suiveurs (GS) est fixé à l'arrière de chaque essieu, figure 1.2(a). Le dispositif de guidage prévoit l'inversion des galets directeurs et suiveurs pour permettre au véhicule, bidirectionnel, de circuler dans les deux sens. Ainsi, le vérin limiteur de débattement permet d'activer le couple de galets situé en avant de l'essieu (GD) dans le sens de la marche, en solidarissant dispositif de guidage et essieu, et de laisser un jeu latéral à l'organe de guidage situé en arrière (GS). Ce système de vérins est piloté par le réseau de conduite. La seule fonction surveillée par le système est l'état des vérins des limiteurs de débattement. Un dysfonctionnement des vérins ne met pas en cause la sécurité du système. Tout au plus, ce dysfonctionnement engendre une usure plus rapide des bandages. Dans ce cas de figure, le chauffeur est informé du dysfonctionnement.

Les galets sont inclinés à 45° ce qui permet de réduire la charge verticale sur les galets et de transmettre les efforts sur le rail par roulement des galets et non par frottement des mentonnets, figure 1.2(b). En fonctionnement normal, il existe toujours un jeu entre les mentonnets, en acier traité, et le rail. Tous les efforts latéraux (vent, force centrifuge, etc.) sont donc repris par les pneus. Seuls les efforts nécessaires à l'orientation des essieux sont repris par le guidage.

Les galets sont portés par une fusée et par des roulements. Les supports des galets, des essieux intermédiaires, portent les patins de retour de courant, figure 1.2(c). Cependant, les essieux extrêmes sont équipés de patins de mise qui servent égale-

ment de chasse-pierres. Chaque galet (en fer) est recouvert d'un bandage composite, figure 1.2(b), jouant le rôle de séparateur entre ce dernier et le rail. Positionné ainsi, le bandage permet de minimiser les nuisances sonores qui peuvent être provoquées naturellement par le contact fer/fer (galet sans bandage et le rail).

1.4.2 La problématique industrielle

Comme le guidage se fait par roulement du bandage sur le rail. Le galet devient une pièce d'usure. L'apparition de défauts (trous, éclatement, etc.) peut être causée par la présence de pièces métalliques, cailloux et autres objets sur le rail, ainsi que par la fréquence de roulement et de freinage normal ou brusque. Une usure anormale du bandage ne met pas en cause la sécurité du système puisque les galets gardent le contact avec le rail mais elle détériore le confort des usagers. Dans un état de fonctionnement normal, galets avec bandage sans défaut, le niveau acoustique engendré est faible. Cependant, dans le cas d'usure du bandage le mentonnet roule dans le fond de l'ornièrre et le guidage induit ainsi des nuisances sonores indésirables (contact acier/fer). Ce changement du niveau acoustique engendré sera ressenti tout de suite par le chauffeur et les usagers car il est supérieur à la normale. L'effort dynamique résultant de la circulation du tramway, dans ce fonctionnement extrême, peut éventuellement créer la fatigue et l'usure de l'organe lui même, le galet. Le paragraphe suivant présente les différents types d'usures pouvant surgir. L'usure sera dans un premier temps contrôlée visuellement tous les jours au dépôt. Automatiser ce processus de surveillance garantira à la fois la disponibilité du tramway et le confort des usagers à moindre coûts. C'est l'objectif du projet SYSCOGA.

1.5 Le projet SYSCOGA : *SYStème de COntrôle des GAlets*

1.5.1 La sûreté de fonctionnement

Le système de guidage est un système dynamique pouvant évoluer sous deux modes de fonctionnement, normal (sans défaut) et anormal (avec défaut). Ce dernier mode de fonctionnement correspond, dans notre cas, à l'existence de défauts sur le bandage des galets de guidage et l'état sous lequel évolue le système devient générateur de nuisances ou de défaillances. Cela induit un changement dans le comportement, niveau acoustique, du système de guidage et voire même une panne générale du tramway. Écouter les bruits lors du passage du tramway pour détecter des signes révélateurs de défauts ou contrôler visuellement l'état des galets de guidage serait une tâche coûteuse en temps et en argent. Le parc de Clermont Ferrand se compose de 20 rames STE5 dont chacune à 10 couples de galets. Au total 400 galets donc 400 bandages à contrôler, au minimum, 3 fois par jours, donc 3 équipes de techniciens en permanence.

L'opérateur humain intervient souvent dans la surveillance de ce type de systèmes dynamiques. Cependant, du fait de la complexité et de la qualité de service requises dans ces systèmes industriels, l'intervention de l'opérateur humain est susceptible d'engendrer des erreurs. Il est alors indispensable d'automatiser le processus de surveillance et de mettre en oeuvre des techniques plus fiables pour détecter de façon précise l'apparition d'un signe de défaillance. Ainsi leur continuité de fonctionnement est assurée par la mise en place d'un processus de maintenance préventive et corrective. En particulier la maintenance préventive des installations industrielles a été la source, pendant la dernière décennie, du développement de procédures de surveillance et diagnostic permettant d'alerter l'exploitant, à un stade plus ou moins précoce, sur l'existence d'une anomalie source potentielle d'une rupture ou d'un accident grave. L'intégration d'un tel processus préserve hautement la sûreté de fonctionnement : la disponibilité, fiabilité et maintenabilité (DFM). En effet, la fiabilité et la sûreté de leur fonctionnement permettent en partie d'assurer la sécurité des personnes côtoyant le système, la qualité du service et la rentabilité des installations. Malheureusement, l'utilisation de nouveaux procédés industriels de plus en plus complexes rend cette tâche de plus en plus difficile.

Pour atteindre cette finalité et assurer la sûreté de fonctionnement du nouveau tramway Translohr, un problème particulier nous a été soumis par le Syndicat Mixte des Transports en Commun de l'agglomération clermontoise (SMTC).

Le comportement du système de guidage repose partiellement¹ sur l'état des galets de guidage. L'existence d'un défaut sur les galets (bandage) n'implique pas nécessairement une défaillance, cependant leur maintenance permet d'identifier les défauts dès leur apparition avant de causer des dommages irréversibles, sinon l'état du système de guidage peut s'aggraver et paralyser tout le système (arrêt de fonctionnement). La disponibilité du tramway et le confort des usagers reposent sur l'état du système de guidage. Ainsi, sa maintenabilité et la réduction du bruit nécessitent la surveillance du matériel roulant, pneus et galets de guidage, qui doit être effectuée régulièrement. Le sujet de thèse est ainsi défini à partir de ces besoins industriels. Il s'agit d'automatiser le processus de surveillance de l'état des galets (bandage) de guidage pour garantir la disponibilité du tramway en minimisant les coûts de maintenance à savoir, le nombre d'interventions pour le changement des galets défectueux et par conséquent la réduction des périodes d'immobilisation pour la maintenance.

1.5.2 Le détecteur SysCoGa

Le système de surveillance automatique, nommé détecteur SYSCOGA : *Système de Contrôle des Galets*, est développé dans le cadre d'un partenariat industriel entre le Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS², UMR CNRS 6158), le Syndicat Mixte des Transport en Commun de l'agglomération clermontoise (SMTC), LOHR Industrie³ et la société Signal Dévelop-

¹les galets représentent des composantes de l'ensemble de système de guidage

²<http://www.isima.fr/limos/>

³<http://www.lohr.fr/transport-public-tech.htm>

pement⁴(SD). Un planning prévisionnel du déroulement du projet est présenté dans l'annexe A.

Chaque type de défaut se manifeste dans une bande de fréquence bien définie, basses fréquences pour un défaut structurel et hautes fréquences pour un défaut local. La société SD a proposé d'enregistrer pour chaque passage du tramway l'accélération du rail en espérant que cette différence du comportement vibratoire d'un galet, avec défaut local, structurel ou sans défaut, sera ressentie au niveau de l'accélération. Pour chaque défaut (trous, usure patatoïde, etc.) une base de données est construite et composée à partir d'un ensemble d'observations recueillies lors des campagnes d'essais effectuées sur la piste de la société Lohr Industrie (chapitre 3). Les bases de données sont utilisées par la suite dans l'étape de diagnostic par reconnaissance de formes (RdF). Toutefois, l'expérience acquise par la société SD montre que la principale difficulté pour élaborer une méthode par apprentissage réside dans la définition des paramètres pris en compte, qui doivent être pertinents et stables, et en s'intéressant notamment à plusieurs gammes de fréquences. Pertinents, car ils doivent réagir à la présence d'un défaut et stables, car ils doivent peu évoluer pour un même défaut vu plusieurs fois dans des contextes plus ou moins différents (emplacements sur le tramway).

Le projet SYSCOGA, tel qu'il est défini dans le premier cahier des charges (Juin 2003), se divise en une étape d'acquisition et de paramétrisation des signaux suivie d'une étape de diagnostic d'état des galets. La première partie devait être assurée complètement par la société SD et la deuxième partie a fait l'objet de ce travail de thèse. Le détecteur SYSCOGA sera lié à une sous station de maintenance de manière à envoyer en continu et pour chaque passage du tramway les signaux enregistrés, la date et l'heure. L'envoi du numéro de la rame est assuré par un système d'identification présenté en annexe, figure B.3. L'application finale doit fournir un rapport d'état des couples de galets. La décision - arrêt total du tramway ou changement des galets usés - sera prise par un expert ou un technicien. Ce premier cahier des charges a été ensuite complété et enrichi en 2006.

1.5.3 Classes de défauts

Un ensemble de campagnes d'essais sur site nous a permis, dans un premier temps, d'étudier la faisabilité du projet SYSCOGA, de trouver la bonne configuration matérielle pour l'acquisition des signaux et enfin de définir les classes de défauts à considérer. Vu la nouveauté du système de guidage et la méconnaissance des types d'usures pouvant surgir, un ensemble de défauts ont été élaborés manuellement (tournevis, chalumeau, ...) par les techniciens de la société Lohr. Au fur et à mesure de l'avancement du projet SYSCOGA et des campagnes d'essais, leur nombre a diminué et leur forme⁵ s'est progressivement définie. Six campagnes d'essais au total ont été effectuées. Dans chacune d'elle une nouvelle installation matérielle a été testée : accéléromètres, cellules photoélectriques (type et nombre), carte d'ac-

⁴<http://www.signal-developpement.fr/>

⁵La forme signifie ici la nature de l'usure et la manière dont elle est générée manuellement ou mécaniquement.

quisition, etc. Le chapitre 3 résume les différents composants matériels utilisés pour l'enregistrement des signaux vibratoires.

Deux types principaux de défauts peuvent être distingués. Le premier type correspond aux défauts structuraux des galets qui produisent des vibrations à basses fréquences (ou à grandes échelles). Ces structures basses fréquences additionnent les effets des défauts liés à une déformation de la structure du galet, une erreur de fixation du galet, un mauvais calibrage et les effets d'instabilité dynamique du chariot de guidage. Le second type produit des vibrations à hautes fréquences (petites échelles) et est caractérisé par les défauts surfaciques locaux. Parmi l'ensemble des défauts testés durant les campagnes d'essais on peut citer :

(i) **Les défauts locaux :**

- les trous fig. 1.5(a) ;
- l'usure aléatoire : déchirure du bandage fig. 1.5(f), 50% d'usures + trous, corps étrangers + impacts et usure patatoïde (usure aléatoires) fig. 1.5(c) ;
- l'usure parallèle : 25% d'usures, 50% d'usures, 75% d'usures figure 1.5(g), 100% d'usures avec un peu de bandage et 100% d'usures sans bandage, figure 1.5(h).

(ii) **Les défauts structuraux :**

- l'absence du patin, figure 1.4(a) ;
- la pression nulle sur le vérin d'appui de guidage ;
- l'ajout de 20 g au niveau de l'équilibrage du galet, figure 1.4(b) ;
- etc.

D'autres types de défauts peuvent être considérés. Ils peuvent être dus par exemple à une mauvaise fixation du galet, aux fissures internes du galet⁶ ou même liés à la dynamique du tramway. Cependant, seuls les défauts sur le bandage, défauts locaux, sont considérés et étudiés ici.

⁶Ce défaut n'est pas testé expérimentalement et peut être lié par exemple à un défaut de fabrication.

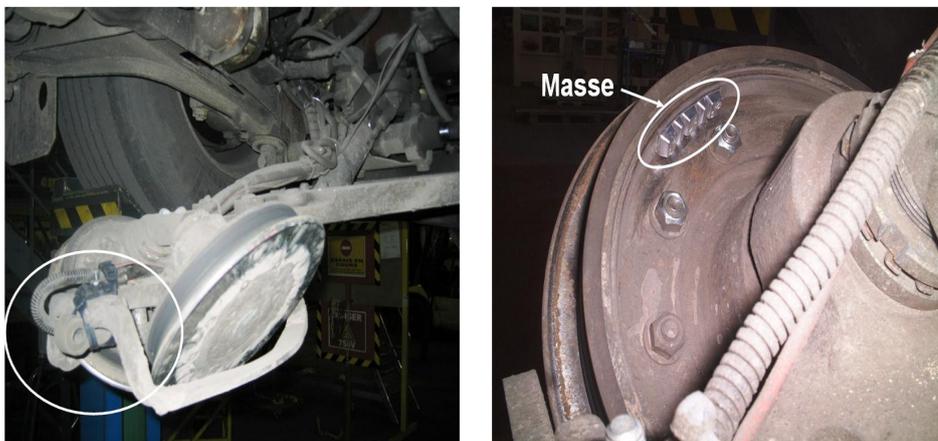


FIG. 1.4 – (a) Absence du patin avant. (b) Masse sur le galet.

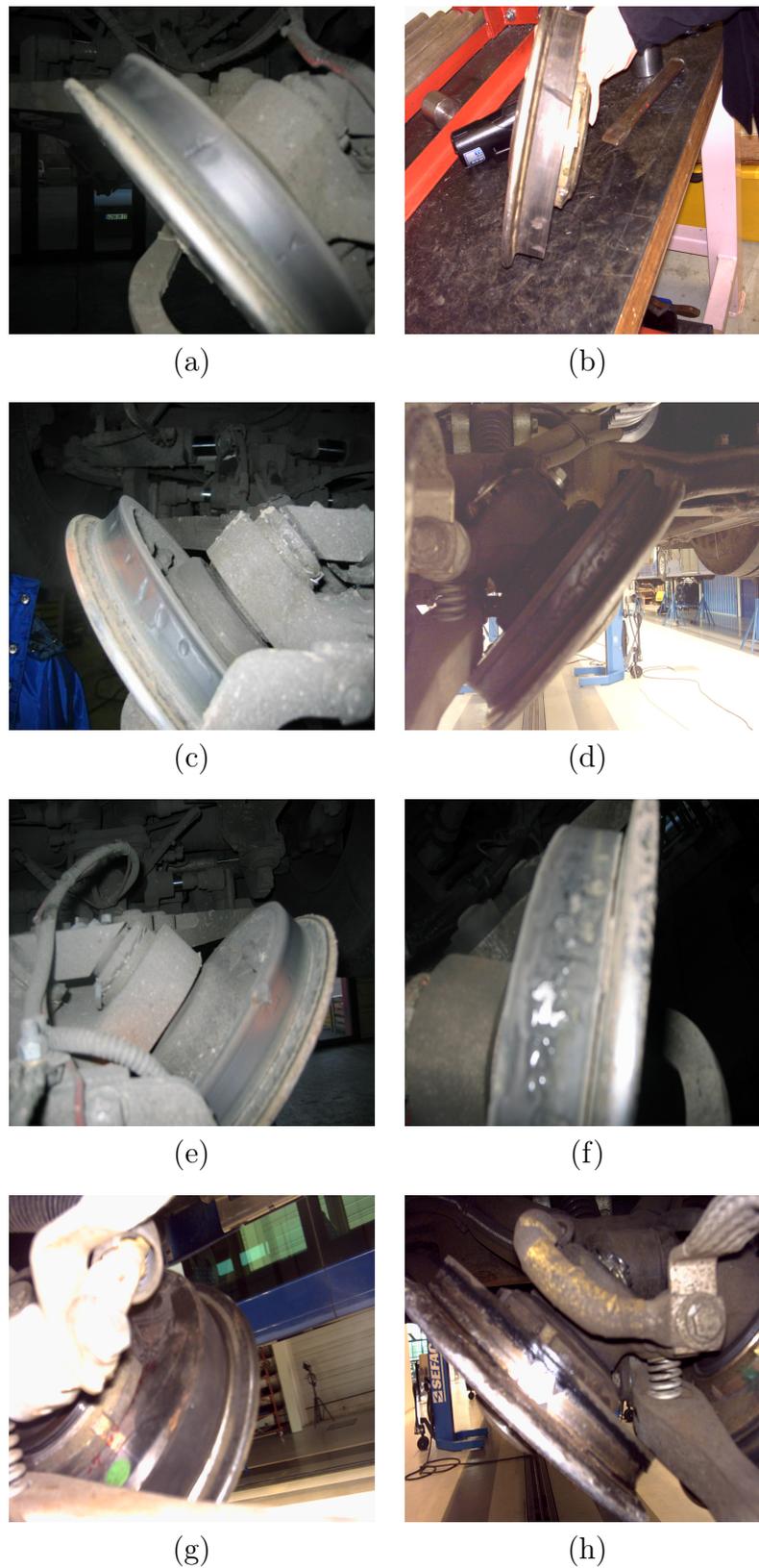


FIG. 1.5 – (a) Bandage avec trous (SYSCOGA-III). (b) Bandage avec trous (SYSCOGA-V). (c) Bandage avec usure patatoïde (SYSCOGA-III). (d) Bandage avec usure patatoïde (SYSCOGA-V). (e) Arrachement interne du bandage. (f) Usure maxi aléatoire. (g) Usure parallèle du bandage à 75%. (h) Usure parallèle à 100% sans bandage.

1.6 Organisation de la thèse

L'étude concerne la détection de défauts ou l'usure des galets de guidage (bandage). Plusieurs techniques de détection de défauts peuvent être utilisées, on peut citer la thermographie infrarouge, la détection ultrasonore et l'analyse des vibrations mécaniques. La maintenance par analyse vibratoire est largement utilisée pour le diagnostic des installations industrielles, notamment le diagnostic des roulements des machines tournantes. Elle fait appel à de nombreux outils de traitement du signal tels que la transformée de Fourier (l'analyse du spectre et du cepstre), la transformée temps-fréquence, la transformation de Hilbert (pour la recherche de modulations d'amplitude et/ou de fréquence), ainsi que le calcul d'indicateurs scalaires tels que la valeur efficace, le facteur de crête et le Kurtosis. Pour notre application, la détection de défaut du bandage repose sur l'analyse des vibrations enregistrées avec des accéléromètres mis en place pour le diagnostic. Ces vibrations sont des processus non déterministes mais leurs propriétés statistiques varient dans le temps de façon périodique (ou quasi-périodique). Cependant, la propriété de cyclostationnarité (ou quasi-cyclostationnarité) des vibrations est négligée dans la majorité des démarches de diagnostic. De même, dans notre approche, le caractère cyclostationnaire des galets n'est pas pris en compte. Il est également envisageable d'élargir nos investigations à d'autres sources d'informations telles que acoustiques. Sans connaissance a priori sur le système de guidage, nous nous plaçons dans une approche sans modèle utilisant des méthodes de reconnaissance de formes (RdF).

Dans ce premier chapitre nous avons défini, le cadre de notre travail ainsi que la formulation du problème posé. Après avoir introduit la notion des transports en commun hybride, nous avons passé une grosse partie de ce chapitre à présenter le tramway sur pneumatiques (le Translohr), détailler son système de guidage, ses composantes et son fonctionnement. Nous avons décrit, dans la troisième partie, la problématique posée par les industriels. Les différents types d'usure pris en compte dans ce travail ont été aussi exposés. Nous avons défini deux types d'usure dont chacun regroupe un ensemble de classes d'usures. Chaque type d'usure a un comportement fréquentiel qui le caractérise. L'exploitation de cette différence pourra nous fournir des données que nous pourrions utiliser dans l'étape de diagnostic par reconnaissance de forme. Par convention, nous substituons le terme "défaut" au terme "usure" dans la suite du document.

Après avoir introduit les aspects techniques et théoriques de notre travail, la suite du mémoire comporte cinq chapitres :

- Le deuxième chapitre est consacré à la démarche du diagnostic par reconnaissance des formes (RdF). Sa définition, sa formulation et ses principes sont décrits. Les différentes phases pour construire un système par RdF sont aussi illustrées. La dernière phase du processus de diagnostic par RdF consiste en une étape de classification. Pour notre application, nous avons adopté la classification supervisée étant donné que les classes de défauts sont connues préalablement (classes étiquetées). Nous exposons un ensemble d'approches de discrimination non-paramétrique les plus utilisées dans la littérature. Nous
-

comparons les résultats obtenus à l'aide des méthodes testées sur nos signaux dans le chapitre 6 ;

- Le troisième chapitre décrit le système d'acquisition utilisé pour l'enregistrement des signaux vibratoires. La première partie du chapitre détaille la configuration matérielle de base utilisée. Les conditions d'intégration sur site sont aussi exposées. Deux installations de la version finale du détecteur SYSCOGA ont été testées et sont présentées dans la deuxième partie. La troisième partie du chapitre est consacrée à l'étude de faisabilité du projet. Les premiers tests d'essai, les premières analyses et les problèmes rencontrés durant le déroulement du projet sont alors évoqués. La quatrième partie donne un descriptif détaillé des signaux vibratoires acquis avec les deux installations matérielles. Comme mentionné dans le deuxième chapitre, le système de guidage du Translohr se compose de 8 couples de galets. Le signal issu du système d'acquisition présente alors l'information superposée des huit couples de galets. Pour obtenir la position exacte de chaque couple de galets sur le signal complet, une segmentation temporelle s'impose et fera l'objet de cette dernière partie du chapitre ;
 - Le chapitre 4 présente un état de l'art des méthodes temps-fréquences les plus répandues dans la littérature et les plus adaptées pour l'analyse des signaux non stationnaires. Nous proposons ensuite une présentation détaillée de l'approche choisie, à savoir la transformée en ondelettes continue (TOC), et ses propriétés. Un ensemble d'arguments sont présentés afin de valider notre choix pour la TOC. Une étude du scalogramme obtenu à l'aide de la TOC permet de spécifier la bande d'échelles (fréquences) utiles pour notre étude. En conséquence, un découpage fréquentiel et temporel est utilisé. La dernière partie du chapitre propose une première tentative de classification qui repose sur la segmentation des scalogrammes à l'aide de techniques de segmentation d'images. Pour la segmentation, nous avons utilisé la méthode de binarisation d'images à niveau de gris par la méthode d'OTSU. La dernière partie du chapitre présente les résultats obtenus, ses avantages et ses inconvénients ;
 - Le chapitre 5 aborde le problème de la réduction de dimension en classification supervisée. Deux approches peuvent être utilisées. La sélection de variables et l'extraction de caractéristiques. La première comporte principalement deux étapes : un algorithme de recherche et un critère d'évaluation mesurant la pertinence des sous-ensembles de variables choisies parmi l'ensemble des variables d'origine. L'extraction de caractéristiques consiste en une projection des variables initiales dans un espace de dimension plus réduite. Nous nous intéressons plus particulièrement aux méthodes d'extraction de caractéristiques à partir des représentations temps-échelle. Nous présentons, dans les deux premières parties du chapitre, ces deux méthodes. La dernière partie présente l'application de la décomposition en valeurs singulières sur les matrices temps-échelle. Nous exposons par la suite, quelques résultats obtenus sur des signaux issus de l'installation sur la voie et de l'installation sur l'appareil de voie (ADV) ;
-

- Nous commençons, dans le dernier chapitre, par introduire les méthodes utilisées pour la classification supervisée. Nous présentons la démarche générale du projet SYSCOGA. Nous décrivons, par la suite, les différents modules intervenant dans la conception de notre système de diagnostic automatique de l'état des galets de guidage. Ce système constitue un cadre d'application des différentes approches choisies et exposées dans les chapitres 3, 4 et 5. Pour mieux répondre aux besoins de l'exploitant, plusieurs techniques de classification et plusieurs voies ont été explorées. Une comparaison de leurs performances est donnée et les résultats obtenus sont largement interprétés et discutés. Nous avons essayé d'étudier le comportement de chaque défaut afin de définir les vraies classes de défauts. Ce travail peut être considéré comme une ouverture vers d'éventuelles travaux sur ce nouveau système de guidage et, pourquoi pas, vers d'autres types de tramways ;

À la fin de ce mémoire, nous concluons ces travaux de recherche, en donnant quelques pistes pour améliorer notre méthode et nous dressons quelques perspectives scientifiques.

DIAGNOSTIC PAR RECONNAISSANCE DES FORMES

2

Si les faits sont correctement observés, il doit y avoir des moyens de les expliquer et de les coordonner entre eux. [Bullard 1965]

Ce présent chapitre est consacré à l'élaboration d'un système de diagnostic par reconnaissance de forme correspondant à la nature de nos données. Néanmoins, avant de discuter plus particulièrement du système de diagnostic de l'état des galets de guidage il semble intéressant, dans un premier temps, de décrire l'architecture globale d'un système de surveillance. Quel que soit le système considéré, dans une démarche de surveillance, trois principales problématiques s'imposent :

- quelles sont les informations pertinentes pour juger l'état du système ?
- comment expliquer la situation à partir de ces informations a priori ?
- si nécessaire, comment remédier à cette situation ?

La première partie du chapitre illustre la stratégie de surveillance qui peut être mise en place pour répondre aux besoins industriels. La deuxième partie du chapitre introduit la notion de diagnostic et passe en revue les différentes approches de diagnostic afin de se positionner par rapport à ce qu'on trouve dans la littérature scientifique. Celle qui s'applique le mieux à notre problématique repose sur l'approche sans modèle et sera amplement détaillée dans la partie 2.2. Les différentes étapes de diagnostic ainsi que l'ensemble des approches qui peuvent être adoptées dans chacune d'elles sont aussi exposées.

2.1 L'architecture globale d'un système de surveillance

2.1.1 Systèmes de surveillance

La surveillance automatique s'inscrit dans un processus global de conduite et de supervision. La figure 2.1 présente l'architecture générale d'un système de sur-

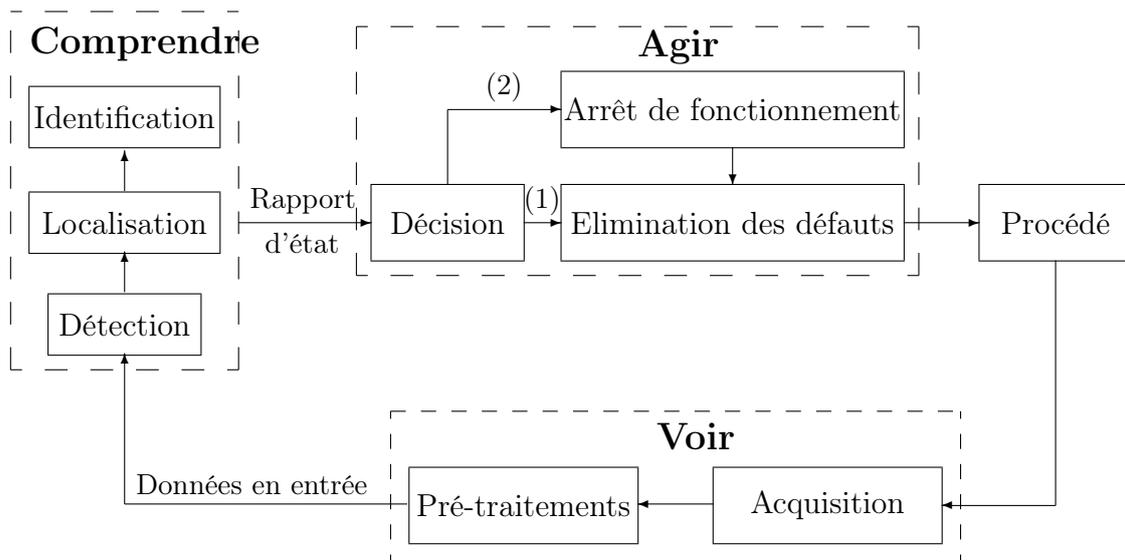


FIG. 2.1 – L'architecture globale d'un système de surveillance. L'état (1) représente un arrêt momentané du tramway avec une reprise presque immédiate. Par contre dans l'état (2), la situation est plus grave et nécessite un retrait du tramway de la circulation.

veillance. On y retrouve les trois grandes fonctions - "Voir", "Comprendre" et "Agir" - nécessaires à une bonne surveillance telle que résumée dans [83]. Les deux tâches essentielles en diagnostic sont donc l'observation des symptômes de défaillance ainsi que l'identification de la cause de défaillance à l'aide d'un raisonnement logique fondé sur des observations acquises sur le système. À partir des informations disponibles sur son état de fonctionnement, l'objectif est de détecter, localiser et identifier les défaillances qui peuvent affecter sa sûreté de fonctionnement. Généralement, sur un procédé global, on distingue trois familles principales de défauts : sur le procédé, sur les actionneurs ou sur les capteurs. Une bonne procédure de surveillance doit être capable de traiter un défaut affectant aussi bien le système physique proprement dit, que l'un de ses organes de conduite (actionneurs, capteurs).

Dans la plupart des cas, la surveillance se base sur des systèmes de traitement d'alarmes. Des valeurs de seuil sont définies préalablement par des experts du procédé selon des critères de sécurité fixés par l'exploitant. Ce système de traitement d'alarmes est l'outil de base pour aider l'opérateur dans sa tâche de surveillance. Il reste cependant à l'opérateur à analyser la situation et à prendre une décision adaptée. Son efficacité et sa performance à répondre à une alarme dépendent de son expérience. Cependant, un tel système de surveillance ne permet pas de gérer le caractère évolutif d'un état normal ou anormal du système. L'utilisation d'un système de surveillance basé sur des techniques plus évoluées permettra de pallier ces limites. Dernièrement, la surveillance et la maintenance des équipements industriels ont fait appel à des techniques de diagnostic de plus en plus sophistiquées. Il existe plusieurs procédures de diagnostic [280]. Le choix d'une approche est lié à la connaissance que l'on souhaite acquérir sur le système, mais aussi à la complexité de ce système.

Dans ce travail de thèse nous traitons la surveillance des composants, les galets de guidage, et nous nous focalisons dans la suite sur les composantes "Voir" et "Comprendre". On s'intéresse donc aux deux étapes élémentaires : la perception et le diagnostic, qui seront décrites dans les paragraphes suivants. Le système de diagnostic et d'aide à la décision conçu dans le cadre du projet SYSCOGA fournit à l'utilisateur les informations sur l'état des galets de guidage, à chaque instant, avec une possibilité d'interagir par une interface Homme-Machine (IHM). Cet outil sera utilisé par la suite comme support pour la maintenance et la prise de décisions. En présence de défaillance et en se basant sur le rapport d'état des galets, le technicien peut intervenir et réaliser une maintenance corrective hors-ligne pour le remplacement des galets défectueux s'il juge que l'état des galets est dégradé. Les différents degrés d'usure réelle seront fixés par l'exploitant, la société de Transport en Commun de l'agglomération Clermontoise (la T2C), lors des prochaines campagnes d'essais planifiées sur site à Clermont-Ferrand. Le suivi et la maintenance de l'installation matérielle - défauts sur les capteurs - seront assurés par la société SD. Toutefois, les défauts sur les actionneurs ne sont pas considérés dans la suite de cette étude.

2.1.2 La perception

L'étape de perception constitue la source principale des informations sur le système. Elle se compose de deux phases. Une phase d'acquisition des données qui consiste à déterminer la configuration matérielle (le type, le nombre de capteurs à utiliser et la période d'échantillonnage, etc.) nécessaire pour le recueil des signaux sur le système étudié. Les signaux acquis doivent fournir l'information utile afin de juger de l'état de fonctionnement dans lequel se trouve le système. Cette première phase est suivie d'une phase de conditionnement des signaux (filtrage, débruitage, ...).

Cette première étape constitue l'étape d'acquisition des signaux vibratoires. Elle est réalisée par la société SD et sera détaillée dans le chapitre 3. Les résultats de cette étape de perception représentent l'information a priori sur le fonctionnement du système étudié et seront utilisés par le module de diagnostic. Ces informations représentent la matière première, leur pertinence conditionne la pertinence du diagnostic.

2.1.3 Le diagnostic

La définition donnée par les instances nationales et internationales de normalisation [AFNOR. Maintenance - Concepts et définitions des activités de maintenance, Norme NF X 60-010. 1994] du terme diagnostic en automatique est "*l'identification de la cause probable de la (ou des) défaillance(s) à l'aide d'un raisonnement logique fondé sur un ensemble d'informations provenant d'une inspection, d'un contrôle ou d'un test*". À l'origine, le diagnostic se limitait aux applications industrielles à haut niveau de risque comme le nucléaire ou l'aéronautique [59, 67], ainsi qu'aux secteurs d'activité de pointe tels que l'industrie de l'armement ou l'aérospatial [123]. Les pre-

miers travaux concernant le thème du diagnostic datent du début des années 1970, résumés notamment dans les articles de synthèse [182, 122, 207], ainsi que dans des ouvrages synthétiques tels que [38, 280].

Dans un système de surveillance, le diagnostic représente la composante principale et consiste à localiser et à identifier les éléments défectueux ou à déterminer si l'état courant du système est normal ou anormal à partir des descripteurs générés par le module de perception. Le diagnostic est constitué, dans la plupart des cas, de plusieurs étapes, figure 2.1 [90] :

- la détection consiste à identifier le mode de fonctionnement du système physique. En se basant sur une connaissance préalable des modes de fonctionnement normaux et anormaux du système, elle permet de décider s'il est en bon ou en mauvais état [46] ;
- la localisation consiste à déterminer de manière plus approfondie les composants défaillants ou la cause probable de la défaillance (fault isolation). Les principaux travaux dans ce domaine sont ceux de Reiter [226] et de De Kleer et Williams [144] ;
- l'identification des défauts a pour but de déterminer le type (classe, taille, ...) d'un défaut (classification).

D'une manière générale, fournir un diagnostic sur système consiste à faire une comparaison entre l'information mesurée, lors du fonctionnement réel du système, et la connaissance a priori de ses modes de fonctionnement. En se basant sur les résultats obtenus de cette comparaison, l'utilisateur peut intervenir et mettre en place les actions correctrices nécessaires à un retour à la normale. Une bonne approche de diagnostic doit être capable de détecter les défauts de faible amplitude avant que ceux-ci ne se propagent dans tout le système et n'entraînent des pannes. Elle doit, en outre, éviter ou diminuer la fréquence des fausses alarmes qui occasionneraient des arrêts inutiles du système.

2.1.4 Quelques approches pour le diagnostic

Deux grandes familles de procédures sont définies dans la littérature. Les méthodes de diagnostic à base de modèles analytiques reposent sur le suivi des paramètres et des grandeurs de la machine, au moyen d'algorithmes d'observation. Le principal avantage de ces méthodes réside dans l'intégration d'une connaissance a priori du système et donc un filtrage de l'information. Les méthodes sans modèle, basées sur l'extraction d'informations, exploitent le traitement des signaux mesurés. Les signaux mesurables tels que la vitesse, les vibrations, les émissions sonores, etc. peuvent fournir des informations significatives sur les défauts. À partir de ces grandeurs caractéristiques du fonctionnement, la mise en oeuvre de méthodes décisionnelles permet de concevoir des systèmes de surveillance et/ou de diagnostic. Les performances de ces méthodes sont étroitement liées à la pertinence des indicateurs de défauts retenus et à la finesse d'analyse des mesures (la perception).

Dans ce qui suit, nous passons en revue quelques approches utilisées pour le

diagnostic. Le domaine est très vaste, le but n'est pas de faire une synthèse exhaustive mais de montrer les différentes possibilités qui s'offrent au concepteur de système de diagnostic. Plusieurs méthodes sont proposées dans la littérature pour établir un diagnostic sur l'état de fonctionnement d'un système donné. Le choix d'une méthode par rapport à une autre dépend fortement de la connaissance a priori du système ainsi que du type de la représentation retenue pour ses modes de fonctionnement. Elles peuvent être classées comme suit :

1. les méthodes basées sur des modèles analytiques ou mathématiques qui peuvent être issus de différents domaines :
 - traitement du signal : séries temporelles, analyse spectrale, analyse temps-fréquence, etc.
 - automatique : observateurs, filtrage de Kalman, etc.
2. les méthodes symboliques : les systèmes experts, graphes causaux, etc.
3. les méthodes basées sur la reconnaissance des formes : approches statistiques, approches floues.

Quand les observations issues du système sont de type numérique et que l'on dispose d'un modèle mathématique du système, l'approche par modèle mathématique est privilégiée pour le diagnostic. La modélisation se base sur une bonne connaissance du comportement entrée/sortie du procédé. Cette connaissance est généralement exprimée sous forme de lois physiques, c'est ce qu'on appelle un modèle de connaissance ou modèle interne. Différentes approches de diagnostic ont été développées par la communauté de l'automatique [98, 122]. Le point commun entre toutes ces approches est la nécessité d'avoir une bonne connaissance du système pour construire un modèle et pour trouver des relations de redondance analytique. Dans ce type de modèle, un diagnostic peut être défini littéralement comme un état plausible du système physique et habituellement décomposé en deux étapes. L'étape de détection produit des symptômes. L'étape de localisation des défauts collecte les symptômes des tests de détection et fournit les diagnostics possibles. Parmi les différentes méthodes utilisant des modèles mathématiques, on trouve principalement l'espace de parité, les observateurs et l'estimation paramétrique. Pour plus de détails sur ces approches, le lecteur peut consulter les références suivantes [20, 74, 86].

Lorsque le mode de représentation de la connaissance est de type symbolique, l'approche par système expert est particulièrement adaptée [46]. Ce type d'approche est adapté aux problèmes nécessitant la manipulation d'une quantité d'information importante de données non homogènes et/ou contextuelles. Pour des applications industrielles on pourra se référer à [9]. En revanche, lorsqu'on ne peut construire un modèle mathématique correspondant aux divers modes de fonctionnement du procédé à surveiller ou si sa complexité le rend inexploitable, les méthodes basées sur la reconnaissance des formes sont les plus adaptées. Par conséquent, l'approche par reconnaissance de formes sert à la conception d'un système de diagnostic en lui procurant les moyens d'appréhender la connaissance a priori incomplète des modes de fonctionnement.

L'ensemble des méthodes citées ci-dessus permettent de générer l'information pertinente (paramètres, vecteur formes, règles, ...) pour l'élaboration des signatures

de chaque état de fonctionnement (avec ou sans défaut). Cependant, la complexité, la méconnaissance, la nouveauté du système étudié dans ce travail ainsi que les contraintes imposées par les industriels ne permettent pas la modélisation du système physique. Nous pensons par conséquent qu'une approche de surveillance et de diagnostic par reconnaissance de formes (RdF) [28, 74] est la plus adaptée. Nous utiliserons simplement une base de données composée à partir d'un ensemble d'observations recueillies lors des campagnes d'essais effectuées sur la piste d'essais de la société Lohr Industrie à Duppigheim (Strasbourg) et décrite dans le chapitre 3.

2.2 Système de diagnostic par reconnaissance des formes

Rappelons que sur des problèmes complexes, la modélisation du procédé est souvent difficile à mettre en oeuvre. On privilégie alors l'approche par reconnaissance de formes. Celle-ci sera présentée comme une solution alternative à l'approche de diagnostic par modèle puisque les modes de fonctionnement sont modélisés mais pas de manière analytique. On ne dispose pas de modèle issu de considérations sur les comportements du système ou de ses composantes, mais d'une base de données composée à partir d'un ensemble de mesures (individus) de ses modes de fonctionnement [74, 28, 35, 34]. En fonction de l'application considérée, un individu est un ensemble de mesures réalisées sur un système physique ou un ensemble d'informations collectées lors de l'observation d'un phénomène. Dans la suite, un individu est représenté par le terme "forme".

2.2.1 Les approches de reconnaissance de formes

La reconnaissance de formes se base sur la définition d'algorithmes permettant de classer des formes en les comparant à des formes-types. Elle intervient dans de nombreux domaines tels que la reconnaissance vocale, la reconnaissance de caractères, l'automatisation industrielle, le diagnostic médical, la classification de documents, etc. De manière générale on distingue deux types de reconnaissance des formes :

- les approches syntaxiques, structurelles et la mise en correspondance des formes (Template Matching) utilisent les hypothèses sur les distributions des données à l'intérieur des classes. Les procédures de classification, dans ce cas, sont construites à l'aide d'hypothèses probabilistes (Ex. classement Bayésien). Elles se basent sur une représentation des formes à l'aide de grammaires et nécessitent des moyens de calculs importants. De ce fait, ce type d'approche n'est pas approprié aux applications en temps réel [2];
 - la reconnaissance de formes statistique, contrairement à la précédente, s'appuie sur une représentation numérique des formes en se basant sur des méthodes paramétriques ou non paramétriques. Les méthodes non paramétriques ne posent pratiquement aucune hypothèse sur la forme des distributions (Ex.
-

K-plus proches voisins [93]). Parmi les méthodes les plus utilisées, on cite les méthodes connexionnistes qui constituent un sous-ensemble des méthodes statistiques [116, 181]. Un état de l'art concernant les méthodes statistiques a été proposé dans [2]. C'est l'approche statistique qui sera utilisée et développée dans la suite du document.

2.2.2 Principe et formulation

On considère N formes, chacune définie par un ensemble de p paramètres regroupées dans un vecteur appelé vecteur forme x . L'espace \mathbb{R}^p défini par ces paramètres est appelé l'espace de représentation. Les formes-types ou prototypes constituent des points représentatifs de l'espace de représentation. Dans le cas réel, les observations sont bruitées et sont rarement confondues avec l'un des prototypes. Cela signifie que chaque forme-type est représentée par une zone géométrique désignée par le terme "classe" en reconnaissance de formes.

Le principe est qu'on observe des formes de M classes différentes, $\Omega = \{\omega_i, i = 1, \dots, M\}$. L'ensemble Ω définit "l'espace de décision". L'objectif est de construire des frontières réalisant une partition de façon à affecter un nouveau vecteur forme à l'une des classes $\omega_1, \dots, \omega_M$. Cette association désigne l'opération de classement ou discrimination.

En diagnostic, l'ensemble des données issues du système de perception sert à construire le vecteur forme. La notion d'état de fonctionnement du procédé permet de construire l'ensemble d'apprentissage, celui-ci est composé de vecteurs formes associés à chaque mode de fonctionnement ou classe.

De nombreux travaux [26, 103, 247], ont permis de montrer l'intérêt de ce type d'approche dans plusieurs domaines d'applications. Le domaine industriel est devenu l'un des premiers champs d'application pour le diagnostic par reconnaissance des formes [277, 35, 194].

2.3 Construction d'un système de reconnaissance de formes

Un système de reconnaissance de formes se décompose généralement en cinq étapes séquentielles [218], figure 2.2. La qualité de chaque étape du processus dépend de la qualité des étapes précédentes. Néanmoins, pour faciliter la compréhension, nous les regroupons en trois phases principales : phase d'analyse, phase de choix de la méthode de décision et enfin la phase d'exploitation.

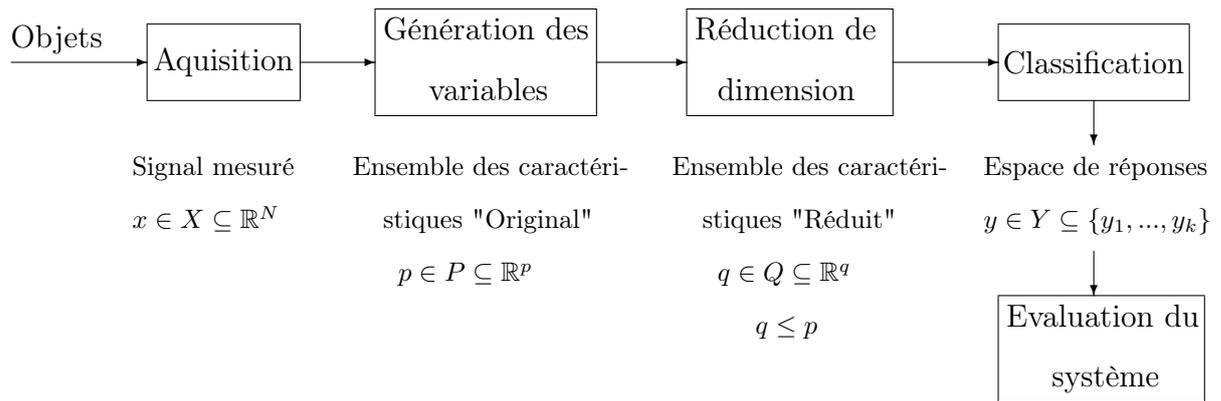


FIG. 2.2 – Système de reconnaissance de formes.

2.3.1 Phase d'analyse

Cette première phase consiste à déterminer les paramètres et les techniques de prétraitement nécessaires pour permettre une bonne classification. La collecte d'informations sur le procédé est effectuée grâce au module d'acquisition, durant l'étape de perception. Suivant le domaine d'application, ces données peuvent être de nature qualitative, quantitative ou les deux à la fois. L'exploitation de ces connaissances permet d'élaborer le module de diagnostic.

Détermination de l'espace de représentation

Il s'agit de construire le vecteur forme x . Les données collectées par le module de perception sont des vecteurs réalisations de variables qui correspondent à des mesures réalisées sur un système physique ou à des informations collectées lors d'une observation d'un phénomène. Ces variables ne sont pas toutes aussi informatives. Elles peuvent correspondre à du bruit, peuvent être peu significatives, corrélées ou redondantes, aberrantes ou simplement inexploitable. De ce fait, une étape de génération de variable s'impose.

L'étape de génération est définie comme la transformation initiale qui vient après la mesure du signal, produisant un ensemble original de paramètres dénoté $p \in P \subseteq \mathbb{R}^p$. On utilise en général des méthodes d'analyse de données ou de traitement du signal pour effectuer cette transformation. Chaque observation effectuée à un instant donné sera caractérisée par l'ensemble de p variables, $x = (x_1, x_2, \dots, x_p)$. L'objectif fondamental est d'accentuer les informations importantes du signal acquis. Cela implique une transformation du vecteur représentant le signal du domaine temporel vers un domaine où les informations contenues dans le signal seront mieux représentées. Cependant, suivant l'information recherchée plusieurs possibilités s'offrent à nous. Les différentes techniques d'analyse du signal pouvant répondre à nos besoins sont détaillées dans le chapitre 4.

En diagnostic, l'étape de caractérisation des états de fonctionnement du système consiste à traiter les mesures recueillies lors de la phase d'acquisition pour extraire une représentation stable et concise, nécessaire à la mise en forme des caractéristiques du fonctionnement normal ou anormal du système. Elle représente le fruit de l'expérience acquise sur le processus et conditionne le succès du système de diagnostic par reconnaissance de formes.

Les données résultantes sont ensuite regroupées dans un tableau numérique de dimension $(N \times p)$ qui correspond à un ensemble de N formes pour lesquelles on connaît la valeur de p variables quantitatives. Les N formes (x_1, \dots, x_N) recueillies sur le système constituent l'ensemble d'apprentissage. La détermination de l'espace de représentation constitue l'étape clé dans la construction du module de diagnostic. La dimension p des variables disponibles à la fin de cette étape peut être très importante. Une forme peut se distinguer mieux des autres en éliminant des variables qui ne sont pas significatives ou en changeant de représentation. Ainsi, travailler dans un espace de représentation réduit, facilitera largement la tâche de la classification. En mode d'apprentissage, l'étape de réduction de variables permet de définir l'espace de représentation le plus approprié pour la classification. C'est dans cet espace de représentation que sont définis les formes à reconnaître.

Réduction de la dimension de l'espace de représentation

Il s'agit de réduire la dimension de l'espace de représentation afin, d'une part, de diminuer le temps de classification d'une nouvelle observation et, d'autre part, d'éviter la dégradation des performances à cause d'une dimension trop élevée de l'espace de représentation [60, 30]. Cette opération de réduction de dimension peut être obtenue par une sélection ou une extraction de variables.

La sélection de variables consiste à ne conserver, parmi l'ensemble des p variables originales, que les plus pertinentes ou les plus informatives, au sens d'un critère de séparabilité entre les classes et d'un critère de condensation des points d'une même classe, comme par exemple le rapport de Fisher. Tandis que l'approche d'extraction, contrairement à la première, utilise toute l'information des variables initiales en les projetant le mieux possible dans un espace de dimension plus petite. En combinant entre elles les variables initiales, on définit de nouvelles variables par transformations linéaires ou non linéaires. Les méthodes d'extraction linéaires les plus utilisées sont l'Analyse en Composante Principale (ACP) [131] et l'Analyse Factorielle Discriminante (AFD). En revanche si cette corrélation est non linéaire, des méthodes d'extraction non linéaire telle que le réseau Vector Quantization and Projection (VQP) [66] sont alors utilisées. D'autres méthodes de projection non linéaires particulièrement adaptées à la visualisation des données, comme la projection de Sammon, sont également utilisées pour l'extraction des variables.

Les différentes méthodes de sélection et d'extraction des variables les plus utilisées sont exposées dans le chapitre 5. Leur application est présentée à travers quelques expérimentations.

Détermination de l'espace de décision

Le diagnostic par reconnaissance de formes est réalisé en utilisant une méthode de classification qui associe une nouvelle observation à une classe correspondant à un mode de fonctionnement. Les méthodes de classification se divisent généralement en deux familles : les méthodes supervisées, et les méthodes non supervisées [134, 218].

L'espace de décision est défini par l'ensemble des classes possibles. Si l'on dispose d'un ensemble de formes étiquetées dont la classe d'appartenance est connue, la classification est dite supervisée. L'étiquetage permet de définir (ou d'apprendre) un ensemble fini de M classes d'objets en général mutuellement exclusives. L'objectif est alors de définir des règles d'affectation pour une forme inconnue x à l'une des M classes apprises. Ainsi, on parle de prédiction en statistique et de classement en reconnaissance de formes. Nous utiliserons, par la suite, les termes discriminateur ou classifieur pour désigner une règle d'affectation.

Dans le cas contraire, si la classe d'origine de chaque observation de l'ensemble d'apprentissage - l'espace de décision - n'est pas parfaitement connue, l'apprentissage des classes se fait en mode non supervisé. Il existe de nombreuses méthodes de classification non supervisée, nommées les méthodes de coalescence telles que les méthodes hiérarchiques [152] et les méthodes des nuées dynamiques [70]. Le principe général de ces méthodes est de détecter dans l'ensemble d'apprentissage une structure de classes disjointes au sens d'un critère de similarité, de telle façon que les observations d'une même classe soient semblables et que celles de classes différentes soient dissemblables.

Étant donné que les deux modes de fonctionnement, sans défaut et avec défaut, ainsi que l'ensemble des classes de défauts sont connus préalablement dans notre application, nous ne nous intéressons, dans le cadre de notre travail, qu'aux méthodes de classification supervisée.

2.3.2 Phase de choix d'une méthode de décision

Une fois l'espace de décision défini, il s'agit de construire des frontières de décision entre les classes. Les méthodes de décision utilisées définissent une règle de décision pour la classification des nouvelles observations aux différentes classes de l'ensemble d'apprentissage. La performance de ces méthodes est estimée en utilisant un indice de performance. En général on choisit la probabilité de mauvaise classification. La règle de décision peut être construite en utilisant une approche statistique ou une approche analytique [74].

Dans une approche statistique, on associe à chaque vecteur forme x une densité de probabilité conditionnelle $f(x|\omega_i)$ par rapport à chaque classe de ω . Les concepts issus de la théorie statistique de la décision [87] et de l'analyse discriminante [175] peuvent être utilisés pour établir les frontières de décision entre les classes. En fonction de la connaissance, complète ou incomplète, dont on dispose sur la distribution de $f(x|w_i)$ on trouve différentes approches.

- lorsque la connaissance est complète, modèle et paramètres de $f(x|\omega_i)$ étant supposés connus, la règle de décision optimale est basée sur la théorie Bayésienne de la décision [93, 74],
- lorsque la connaissance est incomplète c'est-à-dire dans la plupart des cas, le modèle de $f(x|\omega_i)$ inconnu, différentes stratégies peuvent être utilisées :
 - si la forme de $f(x|\omega_i)$ est supposée connue, on estime les paramètres de $f(x|\omega_i)$,
 - si la forme de $f(x|\omega_i)$ est inconnue, on utilise des méthodes non paramétriques pour l'estimation de $f(x|\omega_i)$ telles que les noyaux de Perzen [74]. Sinon on estime les probabilités a posteriori en utilisant par exemple la méthode des k plus proches voisins (k-PPV) et ses extensions [55], ou les réseaux neuromimétiques, etc.

Que ce soit pour les méthodes paramétriques ou non paramétriques, l'estimation de la densité conditionnelle $f(x|\omega_i)$ se fait sur un ensemble d'apprentissage.

Il existe d'autres méthodes de classification qui divisent directement l'espace de représentation en des zones de décision en utilisant des indicateurs, des sorties binaires, telles que les réseaux de neurones basés sur la notion de distance [57] ou les réseaux Learning Vector Quantization [146] ou encore les classifieurs linéaires tels que les réseaux perceptron.

Dans l'approche analytique, la fonction discriminante est déterminée en estimant les paramètres d'une fonction mathématique de manière à séparer au mieux les classes c'est-à-dire à minimiser la probabilité de mauvaise classification. Le choix du modèle de la fonction mathématique dépend de la complexité de la frontière de décision. En pratique, on commence par des fonctions linéaires puis si les performances de classification ne sont pas satisfaisantes, on passe à des modèles plus complexes, quadratiques, réseaux neuromimétiques, etc.

C'est à l'issue de ces deux premières phases qu'est élaboré le module de diagnostic qui décidera à quelle classe ω_i sera affectée une nouvelle observation x . Nous nous sommes intéressés aux méthodes de classification non paramétrique, plus adaptées aux cas industriels, pour lesquelles la densité de probabilité est complètement inconnue. Nous comptons environ 500 types de classifieur [222] qui peuvent être utilisés lors de l'étape de la classification. Il est difficile de proposer un classement général des méthodes de classification supervisées et une même méthode peut être présentée sous divers points de vue. Toutefois, nous avons vu dans la partie 2.3.2, que dans un problème réel de RdF, deux familles d'approches peuvent être utilisées pour définir le modèle probabiliste que suivent les données. Celles retenues pour nos traitements sont les méthodes non-paramétriques qui ne posent pratiquement aucune hypothèse sur la forme des distributions. Le paragraphe 2.4 expose celles les plus citées dans la littérature.

2.3.3 Phase d'exploitation

Évaluation du système de diagnostic

Le système de reconnaissance de formes permet de classer chaque nouvelle observation recueillie sur le système dans une des classes connues, en appliquant la règle de décision élaborée dans la phase précédente. La détermination de cette classe permet de connaître le mode de fonctionnement actuel du système en élaborant un diagnostic.

L'élaboration du système de diagnostic suppose la connaissance exhaustive de tous les modes de fonctionnement du système, hypothèse rarement vérifiée, ce qui nécessite une adaptabilité du système de décision lors de l'apparition de nouveaux modes de fonctionnement.

De nombreux travaux se sont intéressés au problème d'adaptabilité du système de décision et ont apporté des solutions [103, 26]. Dans la plupart des cas, le système est adapté hors ligne. Ces solutions sont peu adaptées lorsque la surveillance porte sur un système évolutif (dynamique) dérivant lentement vers un mode inconnu où il est nécessaire de détecter, de manière précoce, l'apparition de ce mode.

Une fois l'apprentissage effectué, il est indispensable d'évaluer la qualité de la règle de classement induite. L'idée est d'appliquer une phase de test pour estimer l'erreur réelle de classement (taux de mauvais classement). Dans la pratique, l'échantillon S de taille N est divisé en deux ensembles : un ensemble d'apprentissage et un ensemble test. Selon la répartition entre les deux ensembles, il existe de nombreuses méthodes pour estimer l'erreur de classement :

- la re-substitution : les deux ensembles d'apprentissage et de test sont identiques et correspondent à l'ensemble de l'échantillon. Si l'on note n_e le nombre d'erreurs commises, on aura $Erreur = \frac{n_e}{N}$;
- la méthode Holdout [153] : on partitionne l'échantillon S en un ensemble d'apprentissage et un ensemble test. La répartition entre les deux ensembles doit être aléatoire ; en général dans des proportions 1/2, 1/2 pour chacun des deux ensembles ou 2/3 pour l'ensemble d'apprentissage et 1/3 pour l'ensemble test. Si m est la taille de l'ensemble test, on aura $Erreur = \frac{n_e}{m}$;
- la D-validation croisée [40] : l'échantillon est partitionné en D parties de tailles (approximativement) égales. L'apprentissage se fait alors sur les $(D - 1)$ sous-ensembles et le test sur le D^{me} sous-ensemble restant. On réalise alors D apprentissages en laissant à chaque fois une des parties de côté pour le test. Si l'on note $n_e(d)$ le nombre d'erreurs de classement commises sur le d -ème sous-ensemble, l'estimation de l'erreur est la moyenne arithmétique des $n_e(d)$, $Erreur = \frac{1}{N} \sum_{d=1}^D n_e(d)$;
- la méthode du Leave-one-out : cette technique est un cas particulier de la validation croisée pour lequel $D = N$. La procédure de classement est répétée N fois sur $(N - 1)$ instances ce qui la rend très coûteuse en temps de calcul, mais recommandée si l'ensemble E est de petite taille. En effet, elle conduit à

- un minimum de biais dans l'estimation de l'erreur ;
- la méthode de ré-échantillonnage (Bootstrap) [78, 1] : étant donné un échantillon E de taille N , on tire avec remise un ensemble d'apprentissage de même taille N (un élément de N peut ne pas appartenir à l'ensemble d'apprentissage, ou y figurer plusieurs fois), l'ensemble test est N . L'erreur est alors la moyenne des erreurs obtenues pour un certain nombre d'itérations de l'algorithme d'apprentissage.

Dans le cas supervisé, la méthode de re-substitution est connue pour être facile et rapide à mettre en oeuvre. Cependant, l'ensemble de test étant identique à l'ensemble d'apprentissage, l'estimation risque de souffrir d'un biais optimiste quasi systématique. On ne teste alors pas la capacité en généralisation mais plutôt la faculté d'apprendre éventuellement par coeur. La méthode "Holdout" est connue pour être pessimiste et biaisée surtout lorsque la taille des ensembles d'apprentissage et/ou de test est faible. Un partitionnement différent des ensembles d'apprentissage et test donne un résultat d'estimation différent, ce qui se traduit par une variance importante. A l'inverse, la technique du "Leave-One-Out (LOO)" est sans biais mais souffre de l'influence du choix de la base d'apprentissage. Les méthodes de "Bootstrap", quant à elles, fournissent un bon estimateur de l'erreur de variance moins importante que la méthode du "Leave-one-out", mais sont encore plus coûteuses en temps de calcul. Elles sont très utiles pour les échantillons de petite taille [1]. La technique de validation croisée réalise un bon compromis entre le biais et la variance [40].

Ces méthodes d'évaluation peuvent être interprétées différemment selon que l'on effectue un apprentissage supervisé ou non supervisé. Notons que toutes ces méthodes sont des méthodes d'estimation de l'erreur théorique. Une évaluation de la qualité de ces méthodes nécessite le rappel de deux propriétés importantes d'un estimateur : le biais et la précision. Soit T l'estimateur de l'erreur réelle e à partir de l'échantillon N de taille n . On notera $\hat{e} = T_n(x_1, \dots, x_n)$, la valeur estimée de e par T_n .

Définition 2.3.1

- (i) *le biais* : On dira qu'un estimateur est non biaisé si son espérance est égale à la valeur réelle e , $E(T_n) = e$. L'estimateur ne sera qu'asymptotiquement sans biais si $\lim_{n \rightarrow \infty} E(T_n) = e$. Un estimateur biaisé approche la valeur vraie avec un décalage systématique $E(T_n) = e + \epsilon$.
- (ii) *la précision* : Un estimateur est précis ou efficace s'il donne des résultats proches d'un échantillon à l'autre. Cette précision est déterminée par la variance de l'estimateur, $E((T - E(T))^2)$.

Les résultats obtenus lors de la phase d'exploitation et d'évaluation sont exposés dans le chapitre 6 en comparant les différentes méthodes utilisées dans l'étape de la classification.

Notion de rejet

On peut penser qu'un problème de diagnostic n'est autre qu'un problème de reconnaissance de formes. Cependant, il existe une différence significative entre les deux dans la connaissance a priori du nombre de classe : en RdF cette connaissance est complète tandis qu'en diagnostic cette information est souvent incomplète. Dans ce cas, le plus judicieux est d'intégrer cette méconnaissance dans la phase de décision de telle sorte qu'on peut dire, "on ne sait pas dans quel état se trouve le système" ou encore "cette observation correspond peut être à un nouvel état de fonctionnement". La décision finale est prise après une étape de vérification. Cette possibilité de doute et de ne pas prendre une décision exclusive est offerte par la notion dite de "rejet". Son intégration au sein d'un système de diagnostic par reconnaissance de formes permet de diminuer l'erreur de classification.

Dans le cas où tous les modes de fonctionnement (classes) ne sont pas connus, on utilise donc une classe α appelée classe de rejet pour combler le manque d'informations. Deux types de rejets sont possibles :

- Il se peut qu'une observation soit affectée à plusieurs classes avec des certitudes identiques. Dans ce cas, au lieu de pencher pour un état ou un autre en forçant la décision, il est intéressant de rejeter l'observation sans être obligé de prendre la décision de l'affecter à l'une des classes en conflits ni de créer un nouvel état de fonctionnement. Ce rejet dit en "ambiguïté", introduit par C. Chow [49], affecte une observation, se situant approximativement à égale distance entre plusieurs classes, dans une nouvelle classe appelée la classe de rejet en ambiguïté ω_o , figure 2.3. Cette classe rassemble les observations correspondant au cas d'indécision.
- Si l'observation à classer se trouve dans une zone de l'espace éloignée de toute classe, l'observation située loin des classes connues est affectée à la classe de rejet en distance ω_d [74], figure 2.3. Cela veut dire que le système se trouve dans un mode de fonctionnement inconnu. Cette notion de rejet est la base de toute procédure de décision adaptative [34]. Les rejets en ambiguïté peuvent amener à fusionner deux classes et les rejets en distance peuvent engendrer la création de nouvelles classes sous le contrôle du module d'adaptation.

2.4 Approches de discrimination non-paramétrique

Les approches non-paramétriques conviennent particulièrement au cas où les distributions des données sont inconnues ou multimodales. Trois approches sont communément suivies :

- l'estimation des densités conditionnelles $P(x|w_i)$;
 - l'estimation des probabilités a posteriori $P(w_i|x)$;
 - l'estimation des fonctions de classement ou la construction directe des frontières de décision.
-

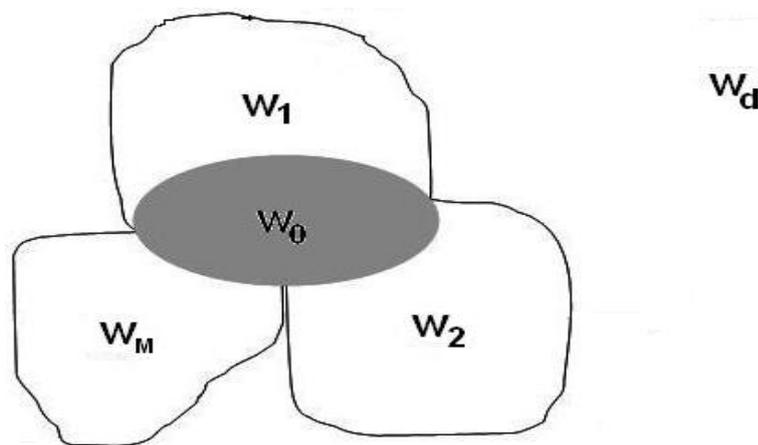


FIG. 2.3 – Partition de l’espace de représentation en $M + 2$ classes où M représente l’ensemble des classes ou modes de fonctionnement.

Les méthodes correspondant à la première approche, telles que les k plus proches voisins ($k - ppv$) et les méthodes d’estimation par noyaux de Perzen [200], estiment les paramètres de la surface séparatrice, d’une forme connue, et optimisent la décision. Nous présentons dans ce cadre la méthode des $k - ppv$. La seconde approche inclut les règles de classement par arbres de décision. Ces deux premières approches sont plus simples que la dernière, dans le sens où elles permettent d’ajuster facilement le discriminateur constitué lorsque de nouvelles données s’ajoutent à l’ensemble d’apprentissage.

Dans la dernière approche, nous trouvons par exemple la méthode des moindres carrés, la méthode de Fisher, les réseaux de neurones et les machines à vecteurs support (SVM) [260]. Ce sont les méthodes appartenant à cette dernière approche qui seront utilisés dans l’étape de classification. La partie 2.5 détaille les plus utilisées dans la littérature.

2.4.1 Principe de l’estimation de densité non paramétrique

Nous supposons ici que l’espace des observations est muni d’une distribution de probabilité définie par une densité f quantifiable suivant la mesure de Lebesgue. Soit R une région de l’espace des observations, on peut définir la probabilité d’observer x dans R et son volume par

$$P(R) = \int_R f(x) dx$$

$$V(R) = \int_R dx$$

La densité moyenne de la région R s’écrit $f(R) = \frac{P(R)}{V(R)}$. Si l’on souhaite estimer $f(R)$ à partir d’un échantillon de taille n , la probabilité $P(R)$ est estimée à partir

du nombre n_r de points observés dans R

$$\hat{P}(R) = \frac{n_r}{n}$$

Il est intéressant de souligner que $\hat{P}(R)$ tend vers la vraie valeur de $P(R)$ quand n tend vers l'infini. En remplaçant donc $P(R)$ par son estimation, on obtient un estimateur non biaisé $\hat{f}(R) = \frac{\hat{P}(R)}{V(R)}$ de $f(R)$. Dans la mesure où l'on souhaite obtenir la densité ponctuelle d'un point particulier x , on effectue un passage à la limite

$$\hat{f}(\mathbf{x}) = \lim_{x \in R, V(R) \rightarrow \infty} \hat{f}(R)$$

Dans l'approche des k plus proches voisins (k -ppv) [93], le volume de la région \mathbf{R} autour du point x est choisi de manière à contenir exactement K points de l'échantillon observé, k étant un paramètre de la méthode. Ainsi, plus le volume $V_x(\mathbf{R})$ est important, plus la densité estimée sera faible

$$\hat{f}(\mathbf{x}) = \frac{1}{V_x(\mathbf{R})} \frac{k}{n}$$

L'estimateur obtenu est sans biais [93]. En pratique, on choisit généralement des volumes réguliers centrés en x (Ex. hypercubes, hypersphères). L'espace de représentation \mathbb{R}^p est muni d'une mesure de distance (Ex. distance euclidienne, métrique adaptative [41]) qui détermine le type de voisinage. Dans ce cas, la règle de décision des k -ppv consiste à affecter un individu x à la classe majoritairement représentée parmi les k plus proches voisins dans l'ensemble d'apprentissage.

Les performances de la méthode des k -ppv dépendent de la valeur de k . En pratique, pour une taille finie de l'ensemble d'apprentissage, la valeur optimale de k est souvent déterminée expérimentalement par validation croisée. Un cas particulier et simple de cette méthode consiste à ne considérer que le plus proche voisin ($k = 1$). L'avantage est qu'aucun paramètre n'est nécessaire à fixer, cependant les frontières de décision construites à partir de cette règle sont souvent non linéaires.

La méthode des k -ppv est très coûteuse en temps de calcul puisqu'il est nécessaire de calculer, à chaque nouvelle affectation, toutes les distances de x à chacun des exemples d'apprentissage. On estime alors la complexité de l'algorithme des k -ppv à $O(kn)$. Une amélioration possible des performances de recherche se trouve dans l'exploitation judicieuse des propriétés géométriques de l'espace des distributions [176]. D'autres approches visent à réduire au préalable le nombre d'individus d'apprentissage en cherchant à ne conserver que les exemples pertinents pour la discrimination [234].

2.4.2 Les arbres de décision

Les arbres de décision ont été largement étudiés et développés dans les années 80¹, notamment grâce aux travaux de Breiman, donnant le modèle CART [149] et ceux de Quinlan, donnant le modèle Id3 [212] puis C4.5 ou J48 [213]. Les arbres de décision sont des algorithmes d'apprentissage symbolique. Ils utilisent des mesures probabilistes pour apprendre des classifications en construisant des partitions arborescentes qui peuvent s'interpréter comme des arbres de décision. Nous présenterons ici uniquement le modèle C4.5 (J48) de Quinlan. Il offre l'avantage d'être assez simple et générique, dans la mesure où la plupart des arbres de décision actuels lui sont proches. Les résultats obtenus par les deux modèles Id3 et J48 sont comparés dans le chapitre 6.

Un arbre de décision est un arbre dont les noeuds sont :

- soit des feuilles contenant des individus appartenant tous à une même classe. Les feuilles représentent donc des sous-ensembles de la même classe w_i ;
- soit des noeuds de décision qui partitionnent les données suivant plusieurs sous-ensembles, chaque sous-ensemble correspond à un résultat d'une fonction de test caractérisant le noeud de décision.

Algorithme de construction

Un arbre de décision est un arbre hiérarchique, chaque noeud est un test d'un ou quelques attributs, chaque branche correspondant à un résultat de ce test. Le principe de construction consiste à diviser récursivement, et le plus efficacement possible, l'ensemble d'apprentissage (échantillon N). Chaque division est faite en appliquant des tests définis à l'aide des variables (caractéristiques) jusqu'à ce que l'on obtienne des sous-ensembles ne contenant que des individus appartenant tous à la même classe.

Choix des tests

On suppose que les variables peuvent être binaires, qualitatives (à valeurs dans un ensemble fini de modalités) ou continues (à valeurs réelles). Pour les variables continues, on utilise généralement des heuristiques qui permettent de les discrétiser [124, 278]. Dans C4.5, les tests portent uniquement sur une variable à la fois. Ils sont donc de la forme $x = a$ pour des données qualitatives, ou $x \leq a$ pour des données quantitatives, x désignant une variable et a une valeur.

Nous allons maintenant, introduire et rappeler quelques notations. Soit un échantillon N de taille n , un ensemble de classes $\Omega = \{w_1, \dots, w_i, \dots, w_c\}$ et un arbre de décision T . À chaque noeud de décision d de T correspond un sous-ensemble de

¹Signalons que certains des arbres de décision (Ex. Id3) étaient déjà connus et utilisées bien avant les années 80 [257]

N constitué des individus qui satisfont les tests de la racine jusqu'à ce noeud. Par conséquent, on peut définir pour tout noeud d de T , les quantités suivantes :

- E_d : l'ensemble des individus de N associés à d ;
- $n(d)$: leur nombre ($\text{card}(E_d)$) ;
- $n(i|d)$: le cardinal de l'ensemble des individus associés à d appartenant à la classe w_i ;
- $\frac{n(i|d)}{n(d)}$: la proportion d'individus de la classe w_i associée au noeud d ;
- d_1, d_2, \dots, d_R : les noeuds fils de d ;
- $E_{d_1}, E_{d_2}, \dots, E_{d_R}$: la partition de l'ensemble E_d , chaque sous-ensemble E_{d_r} étant associé à un fils de d .

La construction de l'arbre se ramène à la détermination des tests à appliquer et l'ordre dans lequel on effectue les tests, de manière à trouver l'arbre le plus petit possible. Pour ce faire, on utilise des heuristiques car ce problème est NP-complet [96]. La plupart des méthodes construisent des fonctions d'évaluation et choisissent le test courant qui donne la meilleure valeur. Le but est de permettre de mesurer le degré de mélange des individus entre les différentes classes. Une telle fonction doit prendre son minimum lorsque tous les individus appartiennent à la même classe (le noeud est pur) et son maximum lorsque les individus sont répartis équitablement. Il existe différentes fonctions qui satisfont ces propriétés, citons parmi celles-ci :

- l'entropie [212] ;
- la fonction de Gini [149].

Ces fonctions mesurent la quantité d'information (ou le degré de mélange) contenue dans un noeud d , celle apportée par le partitionnement du noeud d en R fils est

$$\text{Gain}(x) = H(d) - \sum_{r=1}^R \frac{n(d_r)}{n} H(d_r)$$

L'objectif est alors de choisir le test (ou variable) qui maximise cette fonction gain. Cependant, cette fonction a l'inconvénient de favoriser les tests sur les variables ayant un grand nombre de valeurs possibles. Pour pallier ce problème, C4.5 utilise une version normalisée du gain, appelée le Gain-Ratio

$$\text{GainRatio}(d, x) = \frac{\text{Gain}(d, x)}{\sum_{r=1}^R \left(\frac{n(d_r)}{n} \right) \log \left(\frac{n(d_r)}{n} \right)}$$

2.4.3 L'estimation des fonctions de classement

Les techniques de surveillance par l'estimation des fonctions de classement sont fondées sur l'existence d'une base de données d'apprentissage constituée d'un ensemble de mesures de différents modes de fonctionnement normal ou anormal du système à surveillé. Le principe est de trouver une relation entre les variables d'entrée et les variables de sortie. Les variables d'entrée peuvent être qualitatives ou quantitatives. Pour ce dernier cas, la sortie du réseau de neurones représente l'état

du fonctionnement du système considéré. Le problème de la surveillance est donc considéré comme un problème de reconnaissance des formes. La forme à reconnaître est caractérisée par l'ensemble des données (quantifiables et/ou qualifiables) et les classes d'appartenance représentant les différents modes de fonctionnements. Ce type de méthode peut ainsi nous renseigner sur l'état de fonctionnement tout en assurant à la fois la fonction de détection (fonctionnement normal ou pas) et la fonction de diagnostic, puisqu'il précise le mode de défaillance. Nous introduisons, dans la section suivante, les méthodes les plus utilisées dans le domaine du diagnostic et la surveillance.

2.5 Outils pour la classification

2.5.1 Les réseaux de neurones, concepts et définitions

Les réseaux de neurones sont des outils d'intelligence artificielle (IA), capables d'effectuer des opérations de classification par apprentissage. Leur principal avantage par rapport aux autres outils est leur capacité d'apprentissage et de généralisation de leurs connaissances à des entrées inconnues. Une des qualités de ce type d'outils de classification est son adéquation pour la mise au point de systèmes de surveillance capables de s'adapter à un système complexe avec reconfigurations multiples. Durant une étape d'apprentissage le réseau de neurones essaye d'apprendre la relation "ensembles de données en entrée - modes de fonctionnement", ceci va lui permettre d'associer les classes représentant les modes de fonctionnement aux formes, mesurées sur le système à surveiller, en entrées. Cette étape d'apprentissage permet de déterminer ou modifier les paramètres du réseau de neurones afin d'adopter un comportement désiré. Plusieurs algorithmes d'apprentissage ont été développés depuis la première règle d'apprentissage de Hebb en 1949. Ils sont classés en deux catégories : supervisé et non supervisé.

Dans l'apprentissage supervisé, la valeur de sortie (la classe désirée) que le réseau de neurones doit associer au vecteur d'entrée x est connue. Cette valeur ou vecteur de sortie (appelé cible ou sortie désirée) est fournie par un superviseur (ou expert humain). L'apprentissage consiste à modifier les paramètres du réseau de neurones afin de minimiser l'erreur entre la sortie désirée et la sortie réelle du réseau de neurones. Par contre, dans l'apprentissage non supervisé, aucune connaissance a priori n'est fournie sur la sortie désirée, il s'agit dans ce cas de déterminer les paramètres du réseau de neurones suivant un critère à définir.

La capacité de généralisation est la capacité du réseau de neurones à élargir les connaissances acquises après l'étape d'apprentissage à des données nouvelles et d'associer donc un vecteur d'entrée x , qui n'a pas fait l'objet d'un apprentissage, à une classe donnée. En d'autres termes, c'est l'aptitude du réseau de neurones à approximer une fonction uniquement à partir d'une partie des données. C'est l'une des raisons qui motivent l'étude et le développement des réseaux de neurones artificiels. On peut distinguer les généralisations locale et globale. Dans l'approche locale, chaque neurone est associé à une région d'activation (région d'influence) localisée

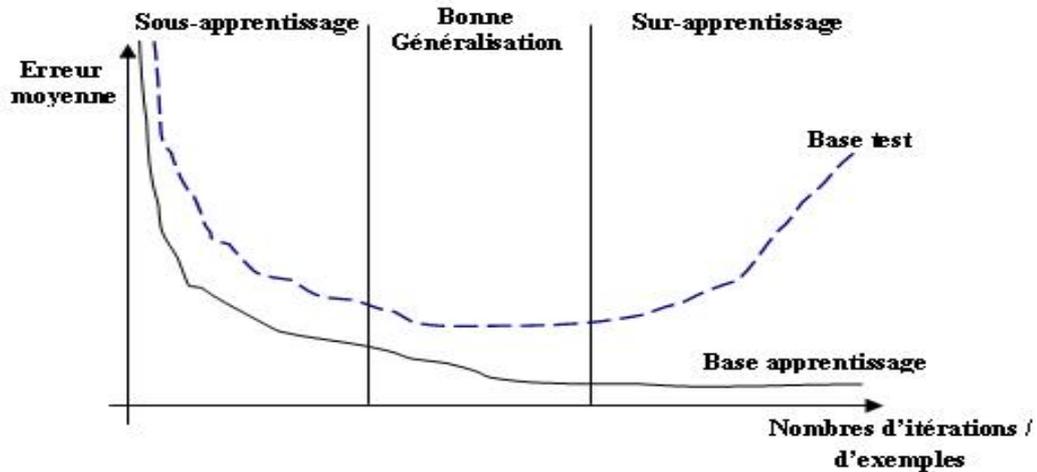


FIG. 2.4 – La complexité d'un réseau de neurones : pour trouver un compromis entre l'erreur de la base d'apprentissage et l'erreur de la base de test, on compare pour différentes complexités k du modèle, l'erreur moyenne sur la base d'apprentissage et sur une base de test indépendante. La complexité optimale du modèle correspond à la valeur de k pour laquelle l'erreur sur la base de test est minimale.

dans l'espace des données. Seule une partie des neurones participe donc à la réponse du réseau. Deux types d'architectures neuronales possèdent cette particularité : les Réseaux à base de Fonctions Radiales (RBF) et la carte de Kohonen.

Contrairement à la première approche, dans l'approche globale, l'ensemble des neurones du réseau participe à l'élaboration de la sortie du réseau. L'information est donc distribuée dans le réseau tout entier. C'est le cas des réseaux de neurones de type Perceptron Multi Couches (MLP) ou le modèle de Hopfield. L'approche globale est supposée plus robuste aux pannes éventuelles de quelques neurones isolés. Par ailleurs, lors d'un apprentissage incrémental, des problèmes d'interférences peuvent apparaître : la modification des paramètres d'un neurone a des répercussions sur l'ensemble de la fonction modélisée par le réseau. Tandis que le réseau apprend dans une région de l'espace des données, le modèle peut oublier ce qu'il a appris dans d'autres régions.

D'une manière générale, la capacité de généralisation du réseau de neurones est très liée à la notion de sous-apprentissage et de sur-apprentissage (Over-Fitting). On parle de sur-apprentissage quand le réseau a trop parfaitement appris les exemples proposés et de sous-apprentissage dans le cas contraire. Ces deux caractéristiques sont complètement antagonistes et dans les deux cas le réseau de neurones sera incapable de généraliser. La figure 2.4 illustre clairement ce compromis entre sur-apprentissage, sous-apprentissage et une bonne généralisation.

Il existe plusieurs types de réseaux de neurones et plusieurs algorithmes d'apprentissage. Les architectures neuronales les plus utilisées dans le domaine de la surveillance et le diagnostic industriel sont le Perceptron Multi Couche, le Réseau à Fonctions de base Radiales et la carte de Kohonen (apprentissage non supervisé) [146]. Seules les deux architectures, le Perceptron Multi Couches (MLP) et les Réseaux à Fonctions de base Radiales (RBF), sont décrites dans la suite.

2.5.2 Les techniques de surveillance par réseaux de neurones

Ces techniques de surveillance sont fondées sur l'existence d'une base de données d'apprentissage constituée d'un ensemble de mesures de différents modes de fonctionnement, normal ou anormal, du système à surveiller. Les variables d'entrée peuvent être quantitatives ou qualitatives. Comme dans notre cas la sortie du réseau de neurones représente alors l'état du système considéré. Ainsi, le problème de la surveillance est donc considéré comme un problème de reconnaissance des formes. La forme à reconnaître est caractérisée par l'ensemble des données (quantifiables et/ou qualifiables) et les classes d'appartenance représentant les différents modes de fonctionnement. Le réseau de neurones peut alors nous renseigner sur l'état de fonctionnement tout en assurant à la fois la fonction de détection (fonctionnement normal ou pas) et la fonction de diagnostic, puisqu'il précise le mode de défaillance.

Les réseaux de neurones peuvent fournir, dans certains cas, des solutions plus intéressantes que les autres outils de surveillance, à condition de choisir judicieusement le type de l'architecture neuronale et surtout de bien mener la phase d'apprentissage. Globalement, ils sont utilisés :

- soit comme outil secondaire pour la surveillance, c'est-à-dire comme approximateur de fonctions pour l'identification des systèmes dynamiques grâce à une boîte noire neuronale. La relation entrées-sorties d'un réseau de neurones peut avoir pour objet, non pas de donner un diagnostic direct, mais de reconstruire une quantité utile à une décision ultérieure. La variable de sortie est une variable réelle. Les réseaux de neurones sont dans ce cas utilisés comme un approximateur de fonctions non linéaires liant les variables de sortie aux variables d'entrée du système industriel. Ils représentent donc une boîte noire non linéaire modélisant le système. Nous pouvons citer les travaux présentés par Böhme dans [192] qui constituent une application très intéressante de détection et localisation des défauts de capteurs d'une centrale d'épuration hydraulique. Ainsi que Vemuri qui a utilisé, dans ses travaux [262] et [261], les réseaux de neurones comme outils complémentaires à la modélisation d'un manipulateur robot ;
- soit comme outil principal de détection et diagnostic, en l'occurrence tous les travaux de classification. Lors de la mise en place d'un système de surveillance par reconnaissance de formes, l'expert est censé connaître les modes de bon fonctionnement et certains des modes de défaillance. L'opération de diagnostic menée par l'expert est souvent très complexe et demande des connaissances ainsi qu'un raisonnement, généralement difficiles à formaliser. Les informations contenues dans l'historique de fonctionnement représentent la base d'apprentissage supervisé du réseau de neurones. La réussite d'une telle application est donc tributaire de la qualité des informations contenues dans l'historique de fonctionnement. La relation entrées/sorties représente dans ce cas directement l'opération de diagnostic [74].

L'application décrite dans [252] est caractéristique de cette démarche. Un réseau RBF est utilisé pour le diagnostic de défaillance d'un véhicule mobile

autonome. Le réseau de neurones associe cinq modes de défaillance à neuf variables mesurées en entrée. Les auteurs mettent en évidence les capacités du réseau RBF à reconnaître de nouvelles situations non apprises. Cette capacité de généralisation locale des réseaux RBF est souvent appréciée dans les applications de surveillance industrielle.

Un autre aspect important des réseaux de neurones a été exploité dans les techniques de surveillance dynamique. Ce type de traitement permet d'apprendre la dynamique d'un signal fournit par un capteur afin de prédire une défaillance. Les réseaux de neurones temporels offrent cette possibilité de prise en compte de l'aspect temporel des données. L'article présenté par [227] constitue une autre application très intéressante des réseaux de neurones pour la surveillance dynamique d'une unité de craquage catalytique fluide.

L'article de Koivo [147] présente une bonne synthèse sur l'application des réseaux de neurones en surveillance industrielle. Il présente les architectures neuronales les plus utilisées dans ce domaine avec des résultats pratiques dans des applications de diagnostic en statique et en dynamique. Trois types de réseaux de neurones ont été testés : le Perceptron Multi Couche, le Réseau à Fonctions de base Radiales et la carte de Kohonen. Les deux premiers réseaux donnent d'assez bons résultats en classification et en identification de processus. Les RBF peuvent se montrer plus performants que les MLP à condition de déterminer judicieusement leurs paramètres. Par contre, le réseau de Kohonen n'est pas aussi performant que les deux premiers mais ses capacités d'auto-adaptation (apprentissage non supervisé) sont très appréciées. L'auteur insiste dans sa conclusion sur le fait que chaque type de réseau a ses avantages et ses inconvénients, la pertinence d'un type par rapport aux autres est fortement liée à l'application considérée. D'autres références peuvent être trouvées dans [84].

2.5.3 Le Perceptron Multi Couches (MLP)

Un MLP est un réseau dont les neurones sont disposés en plusieurs couches successives et où chaque neurone d'une couche est connecté à tous les neurones de la couche suivante. En revanche il n'y a pas de connexion entre les neurones de la même couche. Un MLP est composé d'une couche entrée, une ou plusieurs couches cachées et une couche de sortie. L'addition de cette couche cachée permet au réseau de modéliser des fonctions de décision complexes non linéaires entre n'importe quel espace d'entrée et de sortie. Dans un problème de classification, le nombre de neurones en sortie est fonction du codage adopté. Chaque sortie est dédiée à une classe donnée.

Soit le réseau à deux couches décrit par la figure 2.5 dans lequel les unités de sortie sont notées y_i , les unités cachées v_j et les unités d'entrée x_k . Les connexions des unités d'entrée aux unités cachées sont notées w_{jk} et celles des unités cachées aux unités de sortie par w_{ij} . L'entrée k a pour valeur x_k^l lorsque la donnée l est présentée au réseau. Ces valeurs peuvent être binaires (0/1 ou +1/-1) ou continues.

Pour estimer les paramètres et adapter les poids du réseau de neurones une étape d'apprentissage est nécessaire. L'ajustement est obtenue en minimisant un critère par exemple la somme des erreurs quadratiques.

Pour la donnée d'entrée l , la valeur de sortie de l'unité cachée j est donnée par

$$v_j^l = f(a_j^l) = f\left(\sum_k w_{jk} x_k^l\right),$$

où w_{jk} est le poids de la connexion qui relie l'unité k à l'unité j , a_j est l'activation de l'unité j , f est la fonction d'activation des unités. Cette fonction d'activation est du type fonction à seuil.

Les unités de sortie ont comme valeur

$$y_i^l = f(a_i^l) = f\left(\sum_j w_{ij} v_j^l\right).$$

Les fonctions d'erreurs partielles et globales sont alors définies par

$$E^l = \frac{1}{2} \sum_i (y_i^l - S_i^l)^2 \text{ et } E = \sum_l E^l. \quad (2.1)$$

où S_i^l est la sortie désirée et y_i est la sortie réelle du réseau de neurones.

La minimisation de la fonction d'erreur globale va se faire par une descente de gradient. Par conséquent, après la présentation de tous les vecteurs d'entrée de la base d'apprentissage, la valeur de chaque connexion est modifiée par la règle de delta généralisée

$$\Delta w = -\eta \frac{\partial E}{\partial w} = -\eta \sum_l \frac{\partial E^l}{\partial w}. \quad (2.2)$$

Il existe plusieurs algorithmes d'apprentissage. L'algorithme de rétropropagation (back-propagation) est de loin le plus utilisé [28] et se divise en deux étapes :

- une phase de propagation, qui consiste à présenter les exemples à l'entrée du réseau, puis à propager cette entrée de proche en proche de la couche d'entrée à la couche de sortie en passant par les couches cachées ;
- une phase de rétropropagation, qui consiste à minimiser l'erreur commise sur la base d'apprentissage en modifiant les paramètres du réseau. Après avoir calculé la variation des poids des connexions pour tous les neurones de sortie on calcule alors la variation des poids des connexions de la couche cachée. On met ainsi à jour les poids des connexions de la couche de sortie jusqu'à la couche d'entrée : on rétropropage ainsi le signal d'erreur. C'est de là que vient le nom de cet algorithme : rétropropagation du gradient de l'erreur. Du fait de sommer les w_{ij} pour tous les vecteurs l de la base d'apprentissage puis de remettre à jour les poids avec la variation totale ainsi calculée, l'algorithme est appelé gradient total.

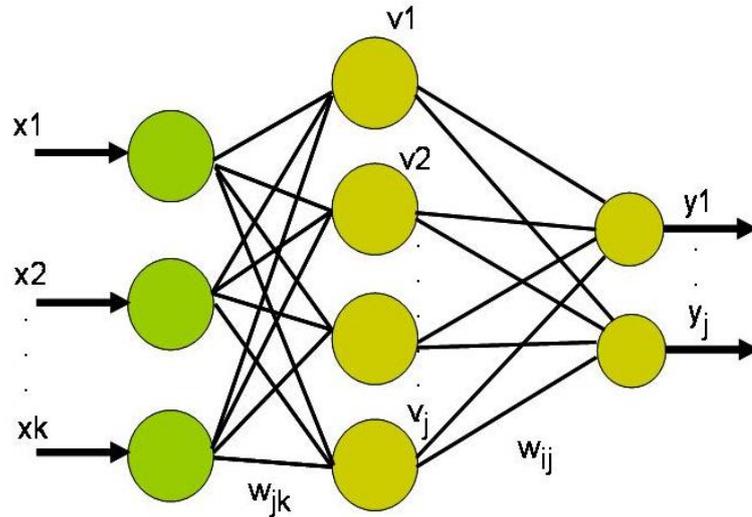


FIG. 2.5 – Perceptron Multi Couches.

L'algorithme de rétropropagation du gradient de l'erreur, utilisé pour l'apprentissage dans un MLP, a permis de dépasser les limites du Perceptron simple. Il s'avère capable de résoudre un grand nombre de problèmes de classification et de reconnaissance des formes. Il souffre néanmoins de nombreux défauts, parmi lesquels :

- le temps de calcul : l'apprentissage est très long ;
- Une grande sensibilité aux conditions initiales, c'est-à-dire l'initialisation des poids des connexions ;
- La présence éventuelle de minima locaux dans la fonction d'erreur. Ce problème est en partie résolu avec le gradient stochastique, mais il subsiste quand même ;
- Le problème de dimensionnement du réseau. La rétro-propagation apprend une base d'apprentissage sur un réseau dont la structure est fixée a priori. La structure est définie par le nombre de couches cachées, le nombre de neurones par couche et la topologie des connexions. Un mauvais choix de la structure peut dégrader considérablement les performances du réseau.

2.5.4 Le réseau de neurones à fonctions de base radiales (RBF)

L'utilisation des réseaux de neurones à fonctions de base radiales remonte aux années soixante-dix pour résoudre des problèmes d'interpolation multi variables [111, 237]. Les bases théoriques de ces réseaux ont été ensuite approfondies par M. J. D. Powell [209], T. Poggio et F. Girosi [208]. D'autres travaux se sont succédés où l'application des RBF a été élargie à d'autres domaines, à savoir la prédiction de l'évolution des systèmes dynamiques et la classification de phonèmes. La particularité de ces réseaux réside dans le fait qu'ils sont capables de fournir une représentation locale de l'espace grâce à des fonctions de base radiales $g(\|\cdot\|)$ dont l'influence est restreinte à une certaine zone de cet espace. $\|\cdot\|$ représente la norme euclidienne.

Un réseau de neurones à fonction de base radiale [99] est un réseau de type feedforward avec une couche d'entrée, une couche cachée qui contient les neurones

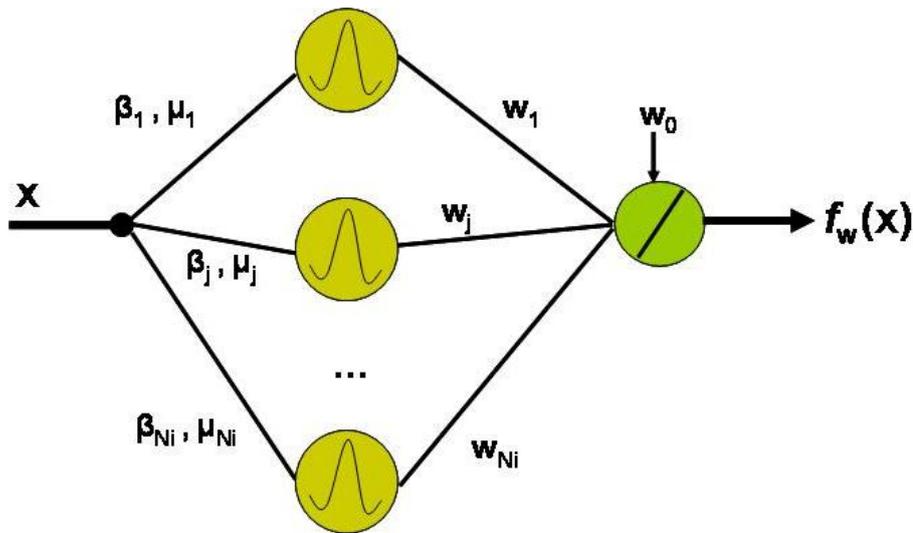


FIG. 2.6 – Réseau à fonctions de base radiales.

RBF et une couche de sortie qui contient une fonction linéaire, figure 2.6. Chaque couche est complètement connectée à la suivante. Chaque neurone RBF contient une gaussienne centrée sur un point de l'espace d'entrée. Pour une entrée x_i la sortie du neurone RBF est l'amplitude de la gaussienne en ce point.

La sortie du réseau est simplement une combinaison linéaire des sorties des neurones RBF pondérées par les poids $w_i, i \in 1, \dots, N_i$ de leur connexions respectives où N_i est le nombre de neurones dans la couche cachée (couche RBF).

Deux paramètres sont associés à cette fonction de base : un vecteur de référence μ_j appelé centre ou prototype et la dimension β_j du champ d'influence appelé rayon d'influence. La réponse de la fonction de base dépend de la distance du vecteur d'entrée x au vecteur prototype μ_j , et de la taille du champ d'influence

$$g_j(x) = g(\|x - \mu_j\|, \beta_j),$$

où la fonction $g(\|\cdot\|)$ est généralement maximale lorsque $x = \mu_j$ et décroît d'une façon monotone vers 0 quand $r = \|x - \mu_j\| \rightarrow \infty$. La fonction de base la plus utilisée est la gaussienne. Elle s'exprime, sous sa forme la plus générale, par

$$g_j(x) = \exp\left(-\frac{\|x - \mu_j\|^2}{2\beta_j^2}\right),$$

où β_j^2 désigne la variance associée à la cellule. La décroissance de la gaussienne est la même pour toutes les directions de l'espace. Les courbes d'iso-activation des cellules cachées sont alors des hypersphères. Un nombre restreint de fonctions de base participent au calcul de la sortie pour une entrée donnée.

Lorsque les RBF sont utilisés pour résoudre les problèmes de classification, cette dernière est considérée comme une classification probabiliste. La loi du vecteur d'entrée x quand on ne connaît pas sa classe d'appartenance est donnée par la loi mélange

$g(x)$

$$g_j(x) = \sum_{k=0}^N Pr(\alpha_k)g(x|\alpha_k),$$

où $g(x|\alpha_k)$ est supposée connue. Elle représente la loi conditionnelle d'appartenance du vecteur x à la classe α_k et $Pr(\alpha_k)$ représente la probabilité a priori des classes α_k , supposée connue. Cette expression ressemble à l'expression de sortie d'un réseau RBF [99]

$$f_w(x) = \sum_{k=0}^{N_i} w_{ik}g_k(x) + w_0,$$

avec w_{ik} représentant le poids de la connexion entre le k -ième neurone radial et le i -ième neurone de sortie. D'après cette représentation, les centres des gaussiennes peuvent être considérés comme étant des vecteurs représentatifs. La sortie des unités cachées représente la probabilité a posteriori d'appartenance du vecteur d'entrée x à la classe α_k . Les poids des connexions représentent la probabilité a priori des classes. La sortie de tout le réseau représente la probabilité a posteriori de la classe de sortie C_i .

La sortie du réseau est simplement une combinaison linéaire des sorties des neurones RBF pondérées par les poids $\{w_i\}_{i=1}^{N_i}$ de leurs connexions respectives et donnée par

$$f_w(x) = \sum_{j=1}^{N_i} w_j g_j(x) + w_0 \quad (2.3)$$

avec

$$g_j(x) = \exp\left(-\frac{\|x - \mu_j\|^2}{2\beta_j^2}\right) \quad (2.4)$$

L'apprentissage des RBF permet de déterminer les paramètres du réseau à partir de la base d'apprentissage $D = \{x_i, y_i\}_{i=1}^N$ en minimisant un critère quadratique régularisé en utilisant la descente de gradient sur E_i [125]

$$E_i = \frac{1}{2} \sum_k (\zeta_k^i - y_k(x^i))^2,$$

où ζ_k^i représente la sortie désirée du k -ième neurone de sortie pour l'exemple i et $y_k(x^i)$ représente sa sortie réelle par rapport à l'entrée x de l'exemple i .

Ces paramètres sont :

- le nombre N_i de neurones RBF ;
- les centres des gaussiennes (prototypes) μ_j ;
- la largeur β_j de ces gaussiennes ;
- le biais et les poids des connexions entre les neurones de la couche intermédiaire et ceux de la couche de sortie $[w_0, w_1, \dots, w_k]$.

Toute modification d'un de ces paramètres entraîne directement un changement de comportement du réseau. Le nombre de neurones RBF N_i et la position des centres sont intimement liés :

- le nombre de neurones RBF est choisi égal au nombre d'exemples dans la base d'apprentissage $N_i = N$, s'il n'est pas trop grand. Les centres sont alors placés sur toutes les observations de la base d'apprentissage ;
- si N , la taille de l'ensemble d'apprentissage, est grande le positionnement des centres et le choix de leur nombre devient un problème de clustering. Plusieurs techniques peuvent être utilisées pour trouver le nombre de neurones RBF. On peut citer la méthode des K-means [72] et la quantification vectorielle (Learning Vector Quantization, LVQ) [222].

Une fois les centres μ_j choisis, il faut déterminer aussi les largeurs des gaussiennes β_j . Une règle empirique consiste à prendre

$$\beta = \frac{d}{\sqrt{2N_i}}, \quad d = \max \|\mu_i - \mu_j\|, \quad 1 \leq i, j \leq N_i,$$

si on choisit la même largeur β_j pour toutes les gaussiennes.

Une autre règle consiste à déterminer une valeur β_j pour chaque centre μ_j [239]

$$\beta_j = \frac{1}{N_{\Omega_j}} \sum_{i/x_i \in \Omega_j} \|x_i - c_j\|$$

où Ω_j est la classe (ou cluster) des observations correspondant au centre μ_j et N_{Ω_j} est son effectif. Il existe d'autres méthodes de détermination des largeurs dans la littérature.

Au même titre que le SVM que nous verrons dans le prochain chapitre, sa fonction de décision peut être perçue comme une combinaison linéaire des mesures de similarité entre l'observation de test et un ensemble de prototypes (les centres des gaussiennes). La fonction de décision correspondant à la classe i s'écrit :

$$f_i(x) = \text{Sign} \left(\sum_{k=1}^r g_k(x) w_{ki} + w_{0i} \right) \quad (2.5)$$

g_k étant la fonction de probabilité de la k -ème gaussienne, r étant le nombre de gaussiennes et "Sign" étant une fonction sigmoïde d'activation. Les paramètres w_{ki} , $k = 1, \dots, r$ représentent les poids du perceptron de la classe i et w_{0i} représente le biais de la fonction de décision de la classe. Ce réseau estime les probabilités conditionnelles des classes. Pour éviter le sur-apprentissage, il requiert habituellement une quantité plus importante de données que le MLP.

2.5.5 Le boosting probabiliste : Adaboost

L'idée principale, dans l'algorithme standard adaptive boosting (AdaBoost), est de définir à chacune de ses étapes $1 \leq t \leq T$, une nouvelle distribution de probabilité

a priori sur les exemples d'apprentissages en fonction des résultats de l'algorithme à l'étape précédente. Le poids à l'étape t d'un exemple (x_i, y_i) d'indice i est noté $D_t(i)$.

Initialement, tous les exemples ont un poids identique, puis à chaque étape, les poids des exemples mal classés par le classifieur sont augmentés. Ceci le force à se concentrer sur les exemples difficiles de l'échantillon d'apprentissage.

À chaque étape t , le classifieur cherche une hypothèse $h_t : \chi \rightarrow \{-1, +1\}$ bonne pour la distribution D_t sur χ . La performance de l'apprenant est mesurée par l'erreur

$$\epsilon_t = P_{D_t} [h_t(x_i) \neq u_i] = \sum_{i: h_t(x_i) \neq u_i} D_t(i).$$

L'erreur est mesurée en fonction de la distribution D_t sur laquelle le classifieur est entraîné. En pratique :

- soit les poids des exemples sont effectivement modifiés ;
- soit c'est la probabilité du tirage des exemples qui est modifiée et l'on utilise un tirage avec remise.

Chaque hypothèse h_t apprise est affectée d'un poids α_t mesurant l'importance qui sera donnée à cette hypothèse dans la combinaison finale. Ce poids est positif si $\epsilon_t \leq 1/2$. Plus l'erreur associée à l'hypothèse h_t est faible, plus celle-ci est dotée d'un coefficient α_t important.

Vers la fin de l'apprentissage, le poids des exemples difficiles à apprendre devient largement dominant. Si on peut trouver une hypothèse performante sur ces exemples c'est-à-dire avec $\epsilon_t \approx 0$, elle sera alors dotée d'un coefficient α_t considérable. L'une des conséquences possibles est que les exemples bruités, sur lesquels finit par se concentrer l'algorithme, perturbent l'apprentissage par boosting. À la fin de cet algorithme, chaque règle de classification h_t est pondérée par une valeur α_t calculée en cours de route. La classification d'un nouvel exemple se fait en utilisant la règle

$$H(x) = \text{sign} \left(\sum_{t=1}^{t=T} \alpha_t h_t(x) \right)$$

Nous remarquons que le boosting construit l'hypothèse finale comme une série additive dans une base de fonctions, dont les éléments sont les hypothèses h_t . C'est la même démarche utilisée dans les méthode d'apprentissage comme par exemple les SVM, les méthodes d'approximation bayésiennes, etc. Pour nos tests nous avons adopté la méthode AdaboostM1 de Yoav Freund et Robert E. Schapire proposée dans l'article [91] utilisée dans le logiciel data mining Weka.

2.5.6 Régression logistique

Dans la régression logistique il s'agit d'analyser une relation causale entre une variable dépendante Y et une ou plusieurs variables indépendantes X_i . En d'autres termes, la régression permet de déterminer les effets des variables X_i sur la variable dépendante Y . Cette analyse permet de mieux comprendre et d'expliquer les phénomènes étudiés, voire même de les prédire, si le modèle est suffisamment solide. En régression logistique, ce qui nous intéresse, c'est de savoir si la ou les variables indépendantes X_i peuvent prédire l'appartenance de Y à l'une ou l'autre des catégories ou classes (avec défauts et sans défaut, 0 ou 1). Pour cela, on calcule la probabilité que la variable Y appartienne à l'une ou l'autre des catégories. Quand la variable indépendante Y est binaire (0 et 1, etc.), on parle de régression logistique binomiale. Quand la variable indépendante Y a plus de deux catégories (0, 1, 2 et 3), on parle de régression logistique multinomiale. Il est essentiel que les catégories soient exhaustives et mutuellement exclusives. La procédure utilisée dans Weka est celle décrite dans l'article de leCessie et van Houwelingen [156], avec quelques modifications. S'il y a k classes pour N exemples (vecteurs formes) avec m variables, la probabilité pour la classe j à l'exception de la dernière classe est donnée par

$$P_j(X_i) = \frac{\exp(X_i B_j)}{(\sum_{j=1}^{k-1} \exp(X_i B_j) + 1)}.$$

La dernière classe a la probabilité

$$1 - \sum_{j=1}^{k-1} P_j(X_i) = \frac{1}{(\sum_{j=1}^{k-1} \exp(X_i B_j) + 1)}.$$

La probabilité de vraisemblance multinomiale est

$$L = - \sum_{i=1}^n \left(\sum_{j=1}^{k-1} (Y_{ij} \ln(P_j(X_i))) + (1 - (\sum_{j=1}^{k-1} Y_{ij})) \ln(1 - \sum_{j=1}^{k-1} P_j(X_i)) \right) + \text{ridge}(B^2).$$

B représente la matrice paramètre de dimension $(m \times (k - 1))$. Afin de trouver la matrice B pour laquelle L est minimisé, la méthode "Quasi-Newton" est utilisée pour trouver les valeurs optimisées à partir de l'ensembles des $(m \times (k - 1))$ variables initiales. Pour plus d'informations sur l'algorithme modifié voir [156].

Bien que performantes sur beaucoup de problèmes, la principale lacune des méthodes discriminantes telles que le MLP et le réseau RBF est la difficulté d'adapter correctement l'architecture du classifieur (réseau de poids) au problème de classification. L'autre maillon faible de ces méthodes est leur sensibilité à la dimensionalité des données.

Comme nous allons le voir dans le paragraphe 2.6, le SVM procure un cadre théorique qui résout le problème du choix du nombre de paramètres puisque celui

se fait automatiquement. L'efficacité du modèle pour la classification de données de grande dimension telles que du texte sur un très grand dictionnaire de mots a été établie par des travaux comme ceux de Joachims et al. [136, 137]. La robustesse et l'efficacité du SVM à traiter des données de nature et de complexité arbitraire a été établie sur beaucoup d'applications de reconnaissance de formes [157].

2.6 Les Machines à Vecteurs Supports (SVM)

2.6.1 Motivation

L'origine des machines à vecteurs de support (SVM) remonte à 1975 lorsque Vapnik et Chervonenkis proposèrent le principe du risque structurel et la dimension VC pour caractériser la capacité d'une machine d'apprentissage. Depuis, les SVMs n'ont cessé de susciter l'intérêt de plusieurs communautés de chercheurs de différents domaines d'expertise. Par exemple, l'identification/vérification du locuteur, la catégorisation de textes, la reconnaissance des visages et récemment la reconnaissance des instruments de musique [174, 80]. Schmidt et al. [238] ont exploré la tâche de reconnaissance d'orateur. Osuna et al. ont traité de la reconnaissance d'images de visages [77].

Le SVM est un modèle discriminant qui tente de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les données des classes. La maximisation de la marge est une méthode de régularisation qui réduit la complexité du classifieur. Elle sert à pénaliser les paramètres du modèles de la même façon que la méthode du "weight decay" qui altère les poids de grande amplitude dans un MLP. La pénalisation des paramètres du SVM dans le but de maximiser la marge est une méthode de sélection de modèle implicite à l'apprentissage. Ce processus produit un ensemble réduit de prototypes faisant partie de l'ensemble d'apprentissage qu'on appelle communément vecteurs de support. Son comportement est, par ailleurs, conditionné par le type du noyau utilisé et par les valeurs octroyées à ses paramètres. Le noyau d'un SVM est une fonction symétrique défini-positive qui permet de projeter les données dans un espace transformé de grande dimension dans lequel s'opère plus facilement la séparation des classes, de la même façon que les neurones cachés d'un MLP permettent de projeter les données entrées dans un espace de représentation dans lequel les poids de la couche de sortie définissent des fonctions discriminantes linéaires des classes. Les valeurs des paramètres de noyaux affectent directement la complexité de la frontière de décision du classifieur. Aussi, ses valeurs influencent le nombre de vecteurs de support du classifieur.

De nombreux travaux ont démontré la supériorité du SVM sur les méthodes discriminantes classiques telles que les MLP, le discriminant de Fisher, les réseaux RBF, etc. Des versions modifiées du SVM permettent les meilleures performances sur plusieurs bases de données standards. Elles présentent l'avantage d'être discriminantes par opposition aux approches génératives (telles que les approches GMM, etc.) qui présupposent une structure particulière, souvent mal justifiée, des formes de densité des données, et exhibent en pratique sa robustesse vis à vis de la dimen-

sionalité des données et une très bonne capacité de généralisation. Ses propriétés font que le SVM est nettement plus avantageux. Seules les notions de base de la classification par Machines à Vecteurs Supports sont ici introduites. Pour plus de détails, le lecteur peut se référer aux références [260, 39, 56, 250, 16, 80].

Les Machines à Vecteurs Supports sont de puissants classifieurs qui font appel aux notions de Minimisation du Risque Empirique (MRE) et de Minimisation du Risque Structurel (MRS). Ils sont par essence des classifieurs bi-classes qui visent à séparer les exemples de chaque classe ω_q , $1 \leq q \leq 2$, au moyen d'un hyperplan H_{w_0, b_0} choisi de manière à garder un maximum de marge de séparation entre n'importe quels exemples d'apprentissage et H_{w_0, b_0} .

2.6.2 Principe des SVM linéaires

De façon plus formelle, en se donnant un ensemble, $D = (x_1, y_1), \dots, (x_N, y_N)$ dans $\mathbb{R}^d \times \{-1, +1\}$, linéairement séparables en deux classes M_1 et M_2

$$D = \{(x_1, y_1), \dots, (x_N, y_N)\}, x \in \mathbb{R}^p, y_i \in \{-1, +1\}$$

$$M_1 = \{x_i \text{ tel que } y_i = +1\}, M_2 = \{x_i \text{ tel que } y_i = -1\},$$

et en munissant \mathbb{R}^d d'un produit scalaire, on cherche un hyperplan qui sépare les données, c'est-à-dire une paire $(w_0 \in \mathbb{R}^p, b_0 \in \mathbb{R})$ telle que :

$$f_w(x_i) = \langle w_0, x_i \rangle + b_0 = \sum_{j=1}^p w_j x_{ij} + b_0 > 0, \forall x_i \in M_1, \quad (2.6)$$

$$f_w(x_i) = \langle w_0, x_i \rangle + b_0 = \sum_{j=1}^p w_j x_{ij} + b_0 < 0, \forall x_i \in M_2 \quad (2.7)$$

Notons (\cdot) le produit scalaire, il s'agit donc de déterminer l'hyperplan optimal

$$H_{w_0, b_0} : w_0 \cdot x + b_0 = 0; w_0 \in \mathbb{R}^d, b_0 \in \mathbb{R}, \quad (2.8)$$

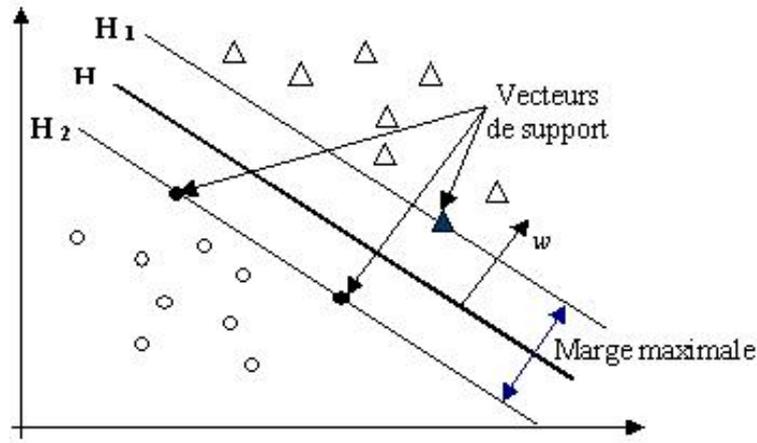
solution de

$$\max_{w \in \mathbb{R}^d, b \in \mathbb{R}} \min_{x, i} \{\|x - x_i\|; x \in \mathbb{R}^d, w \cdot x + b = 0, i = 1, \dots, l\} \quad (2.9)$$

Nous avons supposé, dans un premier temps, que les données sont linéairement séparables. En utilisant une mise à l'échelle appropriée de w et b , il est possible de contraindre les exemples de chaque classe à satisfaire les conditions

$$w \cdot x_i + b \geq +1, \text{ pour } y_i = +1, \quad (2.10)$$

$$w \cdot x_i + b \leq -1, \text{ pour } y_i = -1, \quad (2.11)$$

FIG. 2.7 – Hyperplan linéaire dans R^2 .

qui peuvent être combinées en une même inégalité

$$y_i(w \cdot x_i + b) - 1 \geq 0, \forall i. \quad (2.12)$$

Les deux hyperplans

$$H_1 : w \cdot x_j + b = +1, x_j \in M_1, \quad (2.13)$$

$$H_2 : w \cdot x_k + b = -1, x_k \in M_2, \quad (2.14)$$

permettent de définir la marge. H_1 et H_2 sont parallèles et sont appelés hyperplans canoniques. Grâce aux équations (2.10) et (2.11) il n'existe aucun point entre les deux. La distance qui sépare H_1 et H_2 est appelée la marge et notée Δ . L'hyperplan optimal est celui qui assure une marge maximale, c'est-à-dire qui rend minimale la quantité $\|w\|$. En effet, il existe plusieurs hyperplans linéaires séparateurs, mais il y en a un seul qui maximise la marge Δ (qui maximise la distance entre lui et les points les plus proches des deux classes). Les points qui se trouvent sur les hyperplans H_1 et H_2 sont appelés les vecteurs supports (Support Vectors). La figure 2.7 en donne une illustration.

Le problème posé ne dépend en fait que de ces points particuliers en ce sens que si tous les autres points sont éliminés, la solution du problème reste la même. Ainsi, l'hyperplan optimal est solution du problème d'optimisation

$$\text{minimiser } \tau(w) = \frac{1}{2} \|w\|^2; w \in \mathbb{R}^d, b \in \mathbb{R}, \quad (2.15)$$

$$\text{sous les contraintes } y_i(w \cdot x_i + b) - 1 \geq 0, \forall i = 1, \dots, l.$$

Un premier remède consiste à rendre moins rigides les contraintes (2.12) en introduisant des variables d'écart positives ξ_i pour que les contraintes deviennent

$$w \cdot x_i + b \geq 1 + \xi_i, \text{ si } y_i = +1, \quad (2.16)$$

$$w \cdot x_i + b \leq -1 - \xi_i, \text{ si } y_i = -1, \quad (2.17)$$

$$\xi_i \geq 0, \forall i. \quad (2.18)$$

Pour qu'un exemple d'apprentissage x_i soit mal classé, il faut que le ξ_i correspondant soit supérieur à 1. Par suite $\sum_i \xi_i$ est une borne supérieure sur le nombre d'erreurs de classification qui peuvent être pénalisées en modifiant la fonction objectif $\tau(w)$ par

$$r(w, \xi_i) = \frac{1}{2} \|w\|^2 + C \left(\sum_i \xi_i \right), \quad (2.19)$$

où $\xi = [\xi_1, \dots, \xi_l]^T$ et $C > 0$ est un paramètre permettant de contrôler le compromis entre le fait de maximiser la marge et minimiser les erreurs de classification commises sur l'ensemble d'apprentissage. On parle alors de classifieur à marge souple. Notons qu'il est souvent préférable de tolérer certaines erreurs, au bénéfice d'une marge plus grande car ces erreurs peuvent être dues à des outliers, observations aberrantes, non-significatives de la classe qui leur est associée. Il existe une autre réponse au problème de données non linéairement séparables qui mène à l'obtention de surfaces de décisions non-linéaires. Nous exposerons cela dans la partie 2.6.3.

Le problème (2.15) est un problème d'optimisation sous contraintes qui est résolu en introduisant des multiplicateurs de Lagrange $(\alpha_i)_{1 \leq i \leq l}$ et un Lagrangien

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i (w \cdot x_i + b) - 1], \quad (2.20)$$

où $\alpha = [\alpha_1, \dots, \alpha_l]^T$. Le Lagrangien L doit être minimisé par rapport aux variables dites primales w et b , et maximisé par rapport aux variables duales α_i : ce sont les conditions de Karush-Kuhn-Tucker (KKT) [243].

Dans le cas où la fonction objectif, $\tau(w)$, et les contraintes, $c_i(x_i) = y_i(w \cdot x_i + b) - 1$, sont convexes, les conditions KKT sont nécessaires et suffisantes, et la solution du problème est telle que

$$\sum_{i=1}^l \alpha_i y_i = 0, \quad (2.21)$$

$$\sum_{i=1}^l \alpha_i y_i x_i = w, \quad (2.22)$$

$$\alpha_i [y_i (w \cdot x_i + b) - 1] = 0, \quad (2.23)$$

$$\alpha_i \geq 0. \quad (2.24)$$

De plus, l'expression (2.23) implique que tous les points x_i qui ne sont pas vecteurs supports, c'est à dire ceux qui ne vérifient pas l'égalité $y_i(w \cdot x_i + b) - 1 = 0$ ($0 < \alpha_i < C$), sont associés à des α_i nuls. Ainsi, on retrouve que l'hyperplan optimal ne dépend que des n_s vecteurs supports du problème ($n_s \leq l$)

$$w = \sum_{i=1}^{n_s} \alpha_i y_i x_i, \quad (2.25)$$

En remplaçant w par son expression de l'équation (2.20), la fonction objective $L(w, b, \alpha)$ prend la forme d'un Lagrangien dual à maximiser, fonction des α_i seulement. Il s'écrit

$$L(w, b, \alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j}^{n_s} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j), \quad (2.26)$$

avec les contraintes $\alpha_i \geq 0$, $i = 1, \dots, n_s$ et $\sum_{i=1}^{n_s} \alpha_i y_i = 0$.

La fonction de décision est alors définie par le "Sign" de

$$f(x) = w \cdot x + b, \quad (2.27)$$

$$= \sum_{i=1}^{n_s} \alpha_i y_i (x \cdot x_i) + b. \quad (2.28)$$

Le paramètre b peut être déterminé au travers de la condition (2.23) en choisissant un indice i tel que $\alpha_i \neq 0$, ou encore en moyennant les valeurs obtenues en utilisant tous les points x_i associés à des α_i non nuls (pour une meilleure robustesse numérique). Pour plus de détails concernant le calcul des SVM nous invitons le lecteur à consulter [243, 39].

2.6.3 Principe des SVM non linéaires

Boser et al. [33], ont su produire des frontières de décision non-linéaire avec le SVM. L'idée est d'utiliser un noyau de Mercer qui permet de projeter les données dans un espace éventuellement plus grand dans lequel une séparation linéaire des classes est possible. Il est alors important que la formulation du produit scalaire dans l'équation (2.26) reste intacte après l'introduction du noyau. Boser et al. démontrent qu'un noyau de Mercer ayant la forme (2.29) ne change pas la nature de la fonction objective à optimiser qui équivaut toujours à un problème de maximisation quadratique.

Il s'agit donc de doter les SVM d'un mécanisme permettant de produire des surfaces de décision non-planes. L'idée est de transformer les données de l'espace de départ \mathbb{R}^d dans un espace de Hilbert E de dimension supérieure dans lequel les données transformées deviennent linéairement séparables. Ainsi, en exploitant une application l'algorithme SVM linéaire appliqué aux données $\Phi(x_i)$ dans l'espace E produit des surfaces de décision non-planes dans l'espace \mathbb{R}^d (mieux appropriées aux données de départ pour un choix judicieux de $\Phi(x_i)$). Cette procédure peut être rendue très efficace en utilisant une astuce appelée "kernel trick" permettant d'effectuer les calculs nécessaires à l'algorithme dans l'espace de départ \mathbb{R}^d sans passer explicitement dans E .

Du fait que les données apparaissent dans tous les calculs uniquement sous forme de produits scalaires $(x \cdot x_i)$, il suffit de trouver une façon efficace de calculer $\Phi(x) \cdot \Phi(x_i)$. Cela est réalisé en faisant appel à une fonction noyau $k(x, x_i)$, définie

par

$$k(x, x_i) = \langle \Phi(x), \Phi(x_i) \rangle. \quad (2.29)$$

Le Lagrangien dual de la fonction objective à maximiser est alors

$$L(w, b, \alpha) = \sum_{i=1}^{n_s} \alpha_i - \frac{1}{2} \sum_{i,j}^{n_s} \alpha_i \alpha_j y_i y_j k(x_i, x_j), \quad (2.30)$$

Tout le développement présenté dans le paragraphe 2.6.2 reste valable en remplaçant simplement les termes $x \cdot x_i$ par $k(x, x_i)$. La fonction de décision est définie par le "Sign" de

$$f(x) = \sum_{i=1}^{n_s} \alpha_i y_i k(x, x_i) + b, \quad (2.31)$$

Notons que $f(x)$ est une combinaison linéaire de termes non-linéaires $k(x, x_i)$ qui représente la similarité entre les images des points x et x_i . Pour une observation de test x quelconque, la valeur de $f(x)$ indique la classe d'appartenance inférée par le SVM. L'avantage d'une telle approche réside dans le fait qu'il n'est pas nécessaire de connaître Φ explicitement. Il suffit d'obtenir des noyaux convenables.

• Noyaux

Différentes formes de noyaux, vérifiant les conditions de Mercer, ont été proposées. Nous examinerons

- le noyau linéaire

$$k(x, y) = x \cdot y, \quad (2.32)$$

- le noyau polynômial de degré δ

$$k(x, y) = (x \cdot y)^\delta, \quad (2.33)$$

- le noyau radial (RBF) exponentiel

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right). \quad (2.34)$$

2.6.4 L'optimisation quadratique sous contraintes pour l'apprentissage des SVM

Afin de trouver les paramètres du SVM, il est nécessaire de résoudre le problème d'optimisation quadratique convexe donné par l'équation (2.30) dont la formulation matricielle est donnée par

$$\begin{aligned} \max_{\alpha \in \mathbb{R}^n} & -\frac{1}{2} \alpha^t \widehat{K} \alpha + U^t \alpha, \\ & y^t \alpha = 0, \\ & 0 \leq \alpha_i \leq C_i \in [1, \dots, n] \end{aligned} \quad (2.35)$$

où \widehat{K} est une matrice semi-définie positive. Elle désigne la matrice noyau pondérée par les étiquettes des point, $\widehat{K}_{ij} = y_i y_j k(x_i, x_j)$ et U est le vecteur unité de même taille que α . Ce système dual est un programme quadratique convexe avec n inconnues et $2n + 1$ contraintes. Comme la fonction objective est convexe, tout maximum local est aussi un maximum global. Toutefois, il peut y avoir des solutions optimales différentes en terme de α_i donnant lieu à des performances différentes.

Comme les contraintes sont linéaires et en nombre fini, la solution optimale α vérifie pour chacune de ses composantes α_i les conditions de Karuch Kuhn Tucker (KKT), c'est-à-dire :

$$\alpha_i = 0 \rightarrow (f(x_i) + b)y_i > 1 \quad (2.36)$$

$$0 < \alpha_i < C \rightarrow (f(x_i) + b)y_i = 1 \quad (2.37)$$

$$\alpha_i = C \rightarrow (f(x_i) + b)y_i < 1 \quad (2.38)$$

$$(2.39)$$

Cela implique que seuls les points non contraints dans la solution, ($0 < \alpha_i < C$), nécessitent le calcul de leur coefficient (la parcimonie). Cette propriété est le point clef de la résolution des grands problèmes quadratiques. Les autres points - contraints à 0 et contraints à C $\alpha_i = 0$ et $\alpha_i = C$ - ont pour coefficient une valeur fixe donnée par le problème : celle de la borne qui les contraint. À partir de là, il suffit simplement de connaître la répartition des points dans les trois groupes : I_0 , I_C et I_w correspondant respectivement aux points contraints à 0, contraints à C et non contraints pour la résoudre le problème. Par conséquent, la résolution du problème quadratique initial se traduit en une résolution de système linéaire sur les points non contraints.

Ce constat a amené à une grande variété de méthodes de décomposition et différents algorithmes traitant de la résolution de problèmes quadratiques sous contraintes. Cependant, quelques méthodes seulement sont capables de traiter un grand nombre d'exemples souvent sous hypothèse que la matrice de Gram Schmidt K soit creuse. Dans le cas contraire, l'apprentissage d'un SVM de quelques centaines d'exemples prendrait énormément de temps de calcul et assez de ressources memoire. Seulement, il est possible de dériver des algorithmes qui exploitent la forme particulière de la fonction objective duale du SVM.

Parmi les méthodes de décomposition les plus connues on peut citer la plus simple, connue sous le nom de "Chunking", et proposée en 1982 par V. Vapnik. Ou bien d'autres méthodes de décomposition plus sophistiquées comme celles de Osuna et al. proposée en 1997, Chang et Lin proposée en 2004, J. Platt [206] et Vishwanathan et al. proposée en 2003. Ces méthodes sont décrites brièvement dans le paragraphe suivant.

Méthodes de décomposition

Les méthodes de décomposition, comme leur nom l'indique, dévisent le problème initial en plusieurs sous-problèmes de manière à résoudre des problèmes de petite dimension. Les résultats de ces sous-problèmes sont ensuite combinés pour ensuite arriver à la solution optimale globale. On peut citer :

- *La méthode "Chunking" de V. Vapnik [260]* est la première méthode de décomposition utilisée dans l'apprentissage des SVM. Cette technique consiste à éliminer les points dont les $\alpha = 0$ au fur et à mesure de l'avancement de l'algorithme. L'idée est de résoudre le problème quadratique (QP) pour un sous ensemble de points $S1$. Le sous ensemble suivant est composé des vecteurs supports de la résolution précédente et des $S2$ points sélectionnés dans le reste de la base comme étant ceux qui violent le plus les contraintes KKT. La limite de cette méthode réside dans la taille de la solution finale. Comme elle utilise l'ensembles des points non nuls identifiés, la taille de matrice noyau résultante peut être trop grande pour être stocker intégralement en mémoire ;
- *La méthode de décomposition de Osuna et al.* est une technique de décomposition est similaire au chunking. Les auteurs ont suggéré de garder, à chaque étape, une taille de sous ensemble constante et de geler le reste des données d'apprentissage. Tout en laissant la possibilité de retirer du sous ensemble courant des points vecteurs supports ;
- *L'algorithme SMO (algorithme d'optimisation par minimisation séquentielle ou en anglais Sequential Minimal Optimization) proposé par J. Platt [206].* Méthode simple et rapide proposée pour résoudre le problème de programmation quadratique des SVMs sans la nécessité de stocker une grande matrice en mémoire et sans une routine numérique itérative pour chaque sous-problème. Elle peut être perçue comme le cas extrême des méthodes de décomposition successive et consiste à ne considérer que deux points à chaque étape. Chaque sous-problème quadratique peut être alors résolu analytiquement et donc nul besoin de recourir à une résolution numérique coûteuse. Encore faut-il choisir le bon couple (α_i, α_j) à optimiser durant chaque itération. L'auteur utilise des heuristiques qui sont basées sur les conditions de KKT et ne considère que les directions admissibles qui ne modifient que deux coefficients α_i et α_j par des valeurs opposées.

Plusieurs variantes de la méthode sont proposées dans la littérature [141, 246]. La plupart des implémentations de SMO font appel à la technique de shrinking, proposée par Joachims en 1999, de manière à restreindre l'espace de recherche des α . Cette heuristique vise à déterminer au cours de l'avancement de l'algorithme les points, sans avoir à calculer leur coefficient α , qui seront certainement exclus de la solution ou bornés. On pourra ainsi ne plus tenir compte de ces points et réduire mécaniquement la taille du problème à résoudre.

Pour nos traitements nous avons utilisé l'algorithme SMO implémenté dans le package java open source Weka. Cette implémentation remplace globalement

toutes les valeurs manquantes et transforme les valeurs nominales en binaire et normalise également toutes les valeurs par défaut. Dans ce cas, les valeurs en sortie sont basés sur les données normalisées, et non pas sur les données d'origine, ce qui est important pour l'interprétation de la classification. Le noyau utilisé est de type polynomial : $K(x, y) = \langle x, y \rangle^p$ où $K(x, y) = (\langle x, y \rangle + 1)^p$. Pour plus d'informations sur cette méthode, le lecteur peut se référer aux articles [206, 233].

L'utilisation de SMO permet de se dégager de l'utilisation de solveurs QP . Pour les autres méthodes de décomposition, il reste pour chaque sous problème à résoudre un problème quadratique.

Résolution du problème quadratique

Nous allons par la suite exposer deux méthodes de résolution :

- Les méthodes de *point intérieur*, proposées par Karmarkar en 1984, sont utilisées pour résoudre la programmation linéaire en temps polynomial. Ces méthodes se sert de tous les exemples de la base d'apprentissage pour construire la solution. Bien qu'elles soient très efficaces, elles sont limitées dans le cas des bases de données de grande taille ;
- Contrairement à la première, le principe des méthodes de *contraintes actives* est de répartir le plus efficacement possible les points dans les trois groupes I_0, I_C et I_w et de trouver les valeurs des $\alpha_i, i \in I_w$ par la résolution d'un système linéaire. L'algorithme consiste à résoudre pour le groupe I_w , à chaque étape, la minimisation sans contrainte. Si la solution trouvée ne se situe plus dans la boîte de contraintes ($0 < \alpha_i < C$), α est projetée dans l'espace admissible. Pour les SVM, l'algorithme est appelé "SimpleSVM". Développée au laboratoire PSI sous Matlab [164] est nommée à l'origine "monqp". En parallèle, publié par Vishwanathan [267] et est renommée en conséquence "SimpleSVM". Cet algorithme s'est révélé être très efficace en terme de temps de calcul et de précision.

Cet algorithme peut s'appliquer à tout problème posé sous la forme quadratique convexe sous contraintes de boîtes et permet de trouver itérativement la répartition en trois groupes. Pour plus de détails sur cette méthode, le lecteur peut se référer aux articles [165, 267].

Conclusion

Nous avons consacré ce présent chapitre à la démarche de diagnostic par RdF. Après avoir introduit les concepts de base de l'approche de reconnaissance de formes, (cf. 2.2), nous avons présenté les principales étapes de construction d'un système de diagnostic par RdF, figure 2.2, (cf. 2.3). Le module d'acquisition collecte l'ensemble des informations sur le procédé, le système de guidage. Une fois l'étape de perception

effectuée, le processus de diagnostic se déclenche en se basant sur l'ensemble des signaux vibratoires enregistrés. Seule l'approche de diagnostic par reconnaissance des formes (RdF) est prise en compte et utilisée dans nos travaux.

Dans les systèmes de diagnostic par RdF, la précision des informations, sur lesquelles nous basons nos traitements ainsi que leur fiabilité sont limitées. La justesse de la décision finale prise dépend de la prise en compte de l'incertitude et de l'imprécision des informations utilisées. Cela exige des méthodes très performantes permettant la modélisation et la manipulation des imperfections sans perdre d'information. Dans la dernière partie du chapitre, nous rappelons les principes de la théorie de la décision et analysons quelques unes des principales méthodes discriminantes utilisées pour la reconnaissance des formes telles que le MLP et le réseau RBF. Des différences considérables existent entre le MLP et les RBF. Le MLP est caractérisé par une approche globale tandis que les RBF ont une approche locale. Une première conséquence directe de cette différence, en classification, est que les RBF sont capables de dire "je ne sais pas" contrairement au MLP qui donne une réponse (même mauvaise) quel que soit le vecteur d'entrée. Cette caractéristique propre aux RBF est très importante en surveillance industrielle. Les bases de connaissances du fonctionnement des équipements industriels sont rarement exhaustives. Le système de surveillance se doit donc de pouvoir reconnaître les nouveaux modes susceptibles d'être rencontrés tout au long du fonctionnement de l'équipement. Les RBF attirent ainsi l'attention de l'expert humain sur un nouveau mode rencontré alors que le MLP a tendance à donner une mauvaise réponse et l'expert ne saura jamais qu'un nouveau mode a été rencontré. Par conséquent, les RBF laissent une part importante à l'expert humain lors de l'occurrence d'un nouveau mode. L'expert peut ainsi analyser le nouveau mode et le faire apprendre au réseau sans vraiment modifier les informations apprises au préalable. Toutefois, malgré ces arguments les MLP peuvent se montrer plus efficace que les RBF dans certains cas. Les résultats de classification obtenus avec les deux architectures sont comparés dans le chapitre 6.

Comme nous avons vu dans le paragraphe (cf. 2.6), le SVM procure un cadre théorique qui résout le problème du choix du nombre de paramètres puisque celui se fait automatiquement. La robustesse et l'efficacité du SVM à traiter des données de nature et de complexité arbitraire a été établie sur beaucoup d'applications de reconnaissance de formes. Les machines à noyaux en général et les SVM en particulier apportent un certain nombre de réponses aux problèmes évoqués dans les deux paragraphes précédents. C'est la conclusion que nous allons vérifier en pratique dans le chapitre 6.

Une fois la démarche de diagnostic par RdF définie. Nous allons consacrer les chapitres restants à chaque phase de diagnostic. Nous commençons par présenter dans le chapitre suivant, le système d'acquisition utilisé pour l'enregistrement des signaux. L'ensemble des tests d'essais sur site Lohr Industrie et le matériel choisi sont aussi détaillés.

ACQUISITION DES SIGNAUX VIBRATOIRES

Les objectifs fixés au début de la thèse étaient de définir un ensemble de défauts à la fois facilement maîtrisable et reflétant les cas réels d'usures, de trouver un bon compromis entre la sensibilité des capteurs à la présence de défauts et une influence plus faible au bruit (l'isolation). Pour répondre à ces exigences, une structure multi-capteur a été utilisée pour enregistrer les vibrations du rail causées par les galets de guidage au passage du Translohr, en marche avant et arrière. Le système d'acquisition (le détecteur SYSCOGA) a été placé, dans un premier temps, sur la voie et dans un deuxième temps sur l'appareil de voie (ADV). Ces deux installations sur site sont décrites dans les parties 3.2.1 et 3.2.2. La suite du chapitre est réservée à l'étude et l'analyse des signaux vibratoires acquis à l'aide de chacune des deux installations.

3.1 Système d'acquisition

3.1.1 Configuration matérielle de base

L'existence d'un défaut sur le bandage peut se manifester par un changement de comportement du système de guidage au contact du rail, à l'origine d'un bruit gênant. Pour détecter ce changement nous avons besoin d'une information vibratoire et/ou acoustique. Quelle que soit la technique utilisée, il s'agit de détecter ces changements. Ces deux méthodes, l'analyse vibratoire et acoustique, ont plus ou moins les mêmes limitations. Le choix dépend surtout de l'application et de l'environnement. Pour notre étude le choix s'est porté sur l'analyse vibratoire.

L'apparition d'usure, de fissures sur le bandage ou même un défaut sur le galet (calibrage, fixation, ...), se traduit par des chocs qui se caractérisent par un signal vibratoire particulier. Ces vibrations traduisent les mouvements relatifs d'une structure soumise à une sollicitation dynamique. Leur analyse fournit une mine d'informations dont on espère qu'elles permettront d'identifier les défauts dès leur apparition avant de causer des dommages irréversibles. Un galet avec bandage usé ou même neuf se caractérise par une signature vibratoire particulière. En se basant sur ces signatures vibratoires, le système de surveillance et d'aide à la décision élaboré devra juger si le couple de galets diagnostiqué est avec défauts ou sans défauts.

Pour l'enregistrement des signaux vibratoires, la société SD a choisi d'utiliser une méthode par redondance matérielle. C'est une approche pénalisante en termes de poids, volume, coût d'achat et de maintenance. Toutefois, elle est essentiellement réservée aux cas où la continuité du service est obligatoire. Dans notre cas, le confort des utilisateurs et la disponibilité du tramway sont les services qui doivent être assurés en continu par le système de surveillance.

Multiplier physiquement les accéléromètres nous permet d'avoir toute l'information utile et liée à la cyclostationnarité des galets. Nous disposons ainsi d'un ensemble d'informations complémentaires : le maximum de sensibilité de chaque accéléromètre correspond à un secteur différent de la circonférence des galets. Ainsi, le résultat du diagnostic n'en sera que plus fiable.

De plus, cette redondance matérielle assure la fiabilité du système et limite les interventions en cas de panne d'un accéléromètre. En effet, la manipulation des signaux issus des éléments redondants permet d'effectuer d'une part des comparaisons pour valider la décision finale et d'autre part de détecter une incohérence matérielle. Si un même accéléromètre contredit toujours le reste, ce dernier est jugé défaillant. Une fois la défaillance matérielle détectée et localisée le reste de l'équipement en bon état peut être utilisé en mode dégradé pour assurer le fonctionnement du système d'acquisition en attendant un retour à la normal. Cette tâche de détection d'erreurs sur les accéléromètres est assurée par la société SD.

3.1.2 Intégration sur site

Une fois cette configuration maintenue (le nombre d'accéléromètre), un premier problème se pose concernant le type et l'intégration des accéléromètres dans le rail : *"peut-on les fixer sur le rail ?, par soudage ou vissage ? quel est l'encombrement autorisé ?"*. Un deuxième problème se pose ici concernant la détection du passage des galets : *"comment peut-on avoir la position exacte d'un couple de galets donné ?"*.

Pour répondre à ces besoins, différentes configurations matérielles et composants matériels (type, nombre et position) ont été testées avec un ensemble de défauts pouvant être rencontrés dans le cas réel. Finalement, le système de mesure, le détecteur SYSCOGA, est constitué d'un ensemble d'accéléromètres permettant chacun d'enregistrer les vibrations horizontales et verticales du rail provoquées par les galets à chaque passage du tramway à vitesse normale (~ 35 km/h). Une ou plusieurs cellules photoélectriques, composées d'un émetteur et d'un réflecteur, fournissent les dates de référence permettant un repérage dans le temps du passage des galets et/ou des essieux.

Suite à des contraintes liées aux normes de sécurité ferroviaire, l'emplacement du détecteur SYSCOGA a été changé. Ainsi, deux emplacements ont été testés : sur la voie et sur l'appareil de voie (l'ADV : la plaque d'aiguillage). Ce changement d'emplacement est imposé par des contraintes de sécurité ferroviaire. En effet, le système de mesure :

- ne doit pas être encombrant ni toucher la sécurité des usagers ;
- doit être protégé et en aucun cas gêner la circulation routière (travaux, embulance, etc.) ;
- ne doit pas modifier la structure de l’installation et l’infrastructure (rail). Le vissage est interdit. Le trou propre au vis peut se propager et causer des fissures au niveau du rail. Ce qui va entraîner à terme l’usure du rail et donc mettre en péril la sécurité des usagers.

3.2 Version finale du détecteur SYSCOGA

3.2.1 Installation sur la voie

Le système de mesure figure 3.1(a), est un prototype d’essai installé sur une distance totale de 74.2 cm correspondant à la circonférence d’un galet. Il est constitué de trois accéléromètres permettant chacun d’enregistrer les vibrations horizontales et verticales du rail provoquées par le passage des galets sur la voie à chaque passage du tramway à vitesse normale (~ 35 km/h). Les accéléromètres sont vissés sur la semelle du rail, figure 3.1(b). Deux cellules photoélectriques sont positionnées avant et après les capteurs pour fournir les dates de référence permettant un repérage dans le temps. Chaque cellule est une barrière optique de taille réduite, dont la partie émergente ne dépasse pas 5 cm, composée d’un émetteur avec faisceau infrarouge invisible et d’un réflecteur placés de part et d’autre du tramway. Les cellules sont espacées d’environ 0.80 m.

L’acquisition dure 8 secondes et commence 2 s avant l’entrée du tramway dans la zone d’acquisition, le top de la première balise, et se termine 6 s après ce top. Seules les 3 secondes centrales sont utiles pour notre étude. La fréquence d’échantillonnage F_e est égale à 44100 Hz. Le choix de cette fréquence relativement élevée par la société SD durant la phase d’étude est dû à une méconnaissance a priori de la gamme de fréquences utile.

3.2.2 Installation sur l’appareil de voie (l’ADV)

L’installation du détecteur SYSCOGA sur la voie ne respectait pas certaines contraintes de sécurité. D’un côté, le trou conçu pour visser les accéléromètres sur la semelle du rail peut à long terme s’étaler sur une surface plus grande du rail et ainsi le fragiliser. D’un autre côté, l’existence d’objets encombrants sur le rail ou la voie est strictement interdite. Il nous a fallu donc trouver un autre emplacement, respectant les normes de sécurité, pour le détecteur SYSCOGA. D’où l’idée de tester la possibilité d’intégrer le détecteur SYSCOGA sur l’appareil de voie (l’ADV ou plaque d’aiguillage).

Les essais sur l’appareil de voie ont eu lieu au mois de Mai 2006. Le système de mesure utilisé durant cette campagne d’essai est pratiquement le même que celui

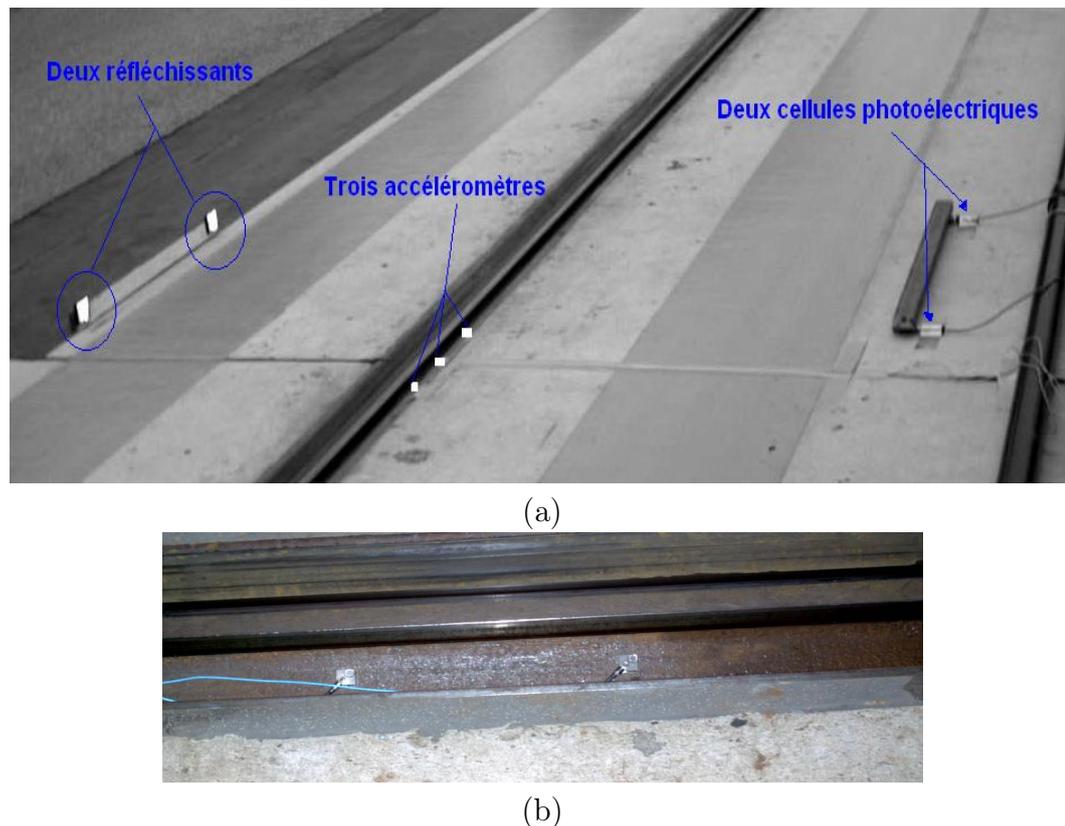


FIG. 3.1 – (a) Système d’acquisition sur site : Installation sur la voie. (b) Fixation des accéléromètres sur la semelle du rail.

utilisé dans l’installation sur la voie. La différence est que cette fois-ci le détecteur SYSCOGA est monté sur la plaque d’aiguillage. L’appareil de voie, mesurant 2.18 m, est instrumenté par trois accéléromètres collés dans des emplacements préparés préalablement, figure 3.2(b) et (c). Les capteurs mesurent les accélérations suivant les directions X, Y et Z. Le Translohr circule dans le sens Y positifs pour les circulations ME1 (essieu ME1 en tête) et dans le sens des Y négatifs pour les circulations ME2 (essieu ME2 en tête). Les deux balises photoélectriques sont placées devant chaque extrémité de la plaque voir figure 3.2(a). Ainsi, la distance entre les deux cellules est égale à 2.18 m qui représente la longueur de la plaque. Elles détectent, cette fois-ci, le passage des essieux et non pas les galets comme dans l’installation sur la voie. Le système d’acquisition fournit, pour un passage, 9 signaux vibratoires et 2 signaux fournis par les deux balises photoélectriques. Toutefois, seuls les deux axes horizontal (Y) et vertical (Z) sont utilisés dans la suite des traitements. La figure 3.3 montre les deux sens de circulations ME1 et ME2 ainsi que le positionnement des accéléromètres et des cellules sur la zone d’acquisition.

La plaque d’aiguillage est une structure finie contrairement au rail qui est supposé infini. Cela implique la présence de fréquences de résonances engendrées par la plaque. Ces résonances durent plus longtemps que le passage d’un couple de galets et peuvent éventuellement fausser la qualification d’un couple de galets. Cette problématique nous a inquiété au début des essais. Toutefois, l’idée d’instrumenter la plaque d’aiguillage reste attractive. D’un point de vue technique, les accéléromètres sont mieux protégés contre les parasites électriques et électromagnétiques. Comme

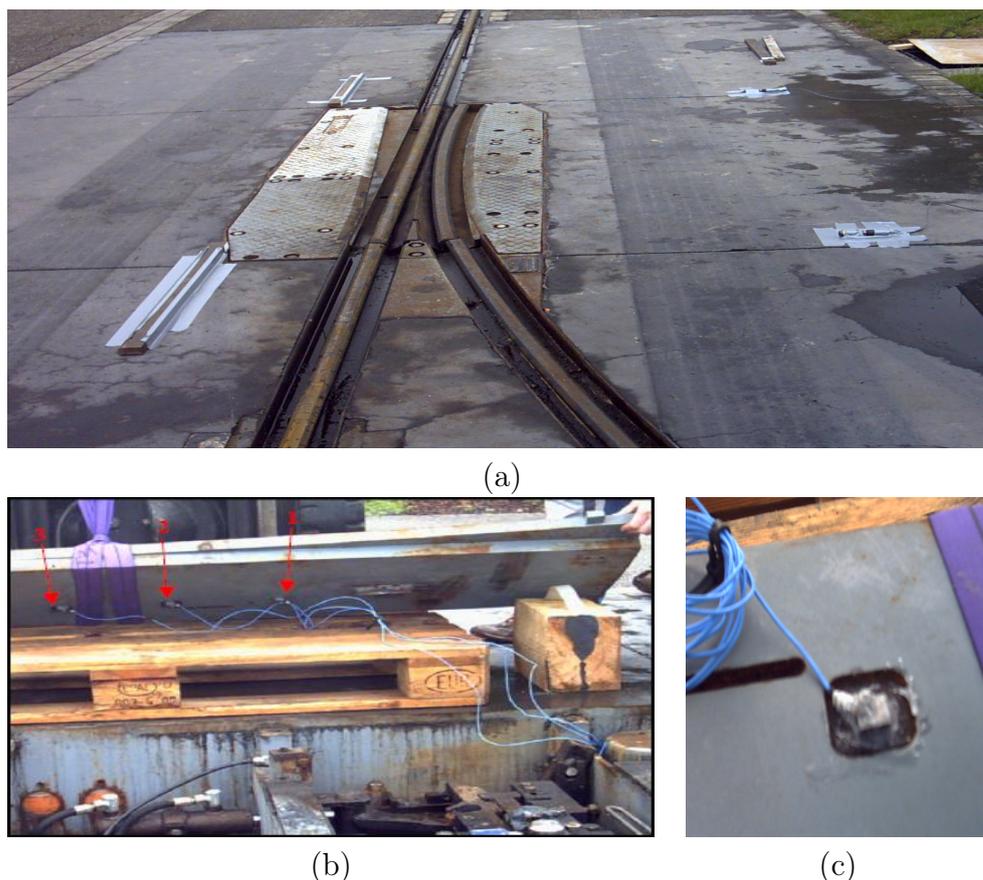


FIG. 3.2 – (a) Système d'acquisition sur site : Installation sur l'appareil de voie. (b) Pose des accéléromètres sous la plaque d'aiguillage. (c) Emplacement réservé pour un accéléromètre.

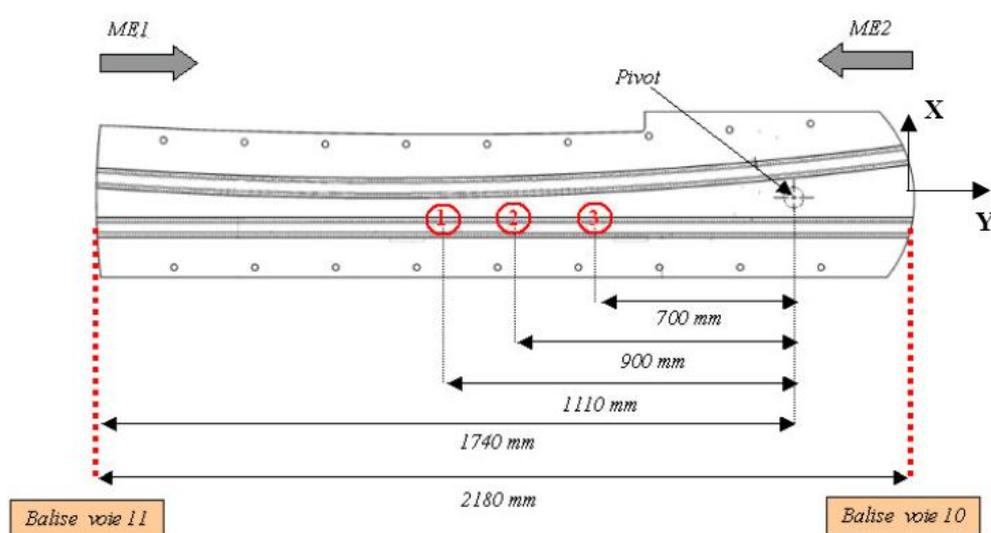
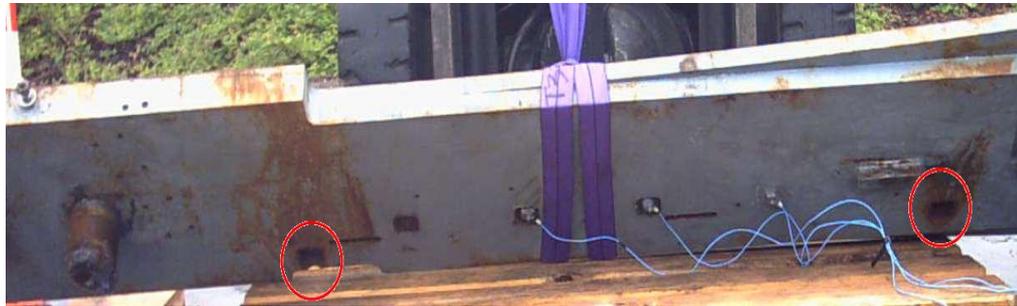


FIG. 3.3 – L'installation sur l'appareil de voie.

l'indique la figure. 3.2(a), les cellules photoélectriques sont positionnées sur le sol. Plus tard elles seront aussi protégées et intégrées dans des emplacements appropriés, figure 3.4(a) sur l'ADV. Par conséquent, les signaux obtenus seront plus ou moins propres et peu bruités ainsi le diagnostic ne sera que plus fiable. La figure 3.5 montre les deux configurations utilisées pour les deux sens ME1 et ME2.



(a)



(b)

FIG. 3.4 – (a) Emplacements réservés aux cellules photoélectriques sur l'appareil de voie. (b) Pivot d'appuis sous la plaque d'aiguillage.

3.3 Étude de faisabilité

3.3.1 Premiers essais, premières difficultés

Les premiers essais SYSCOGA-I et SYSCOGA-II ont eu lieu respectivement aux mois de Janvier et d'Avril 2004. Ces deux premiers essais ont permis la vérification de l'existence de vibrations, significatives, lors du passage du tramway sur la zone d'acquisition, le test de la chaîne d'acquisition et de la configuration matérielle choisies : la disposition des accéléromètres et des cellules photoélectriques sur la voie. Une première possibilité d'intégration des accéléromètres dans le rail est exploitée, figure 3.6. Suite à ces deux campagnes d'essais une première version du détecteur SYSCOGA a été testée et retenue pour la suite du projet.

Les essais SYSCOGA-I ont été réalisés avec le premier démonstrateur Translohr (roulement par lignes aériennes de contact (LAC) et batterie), voir figure 3.7(a). La configuration matérielle utilisée pour l'acquisition des signaux est constituée de :

- huit accéléromètres (4 verticaux et 4 latéraux) collés sur la semelle du rail,

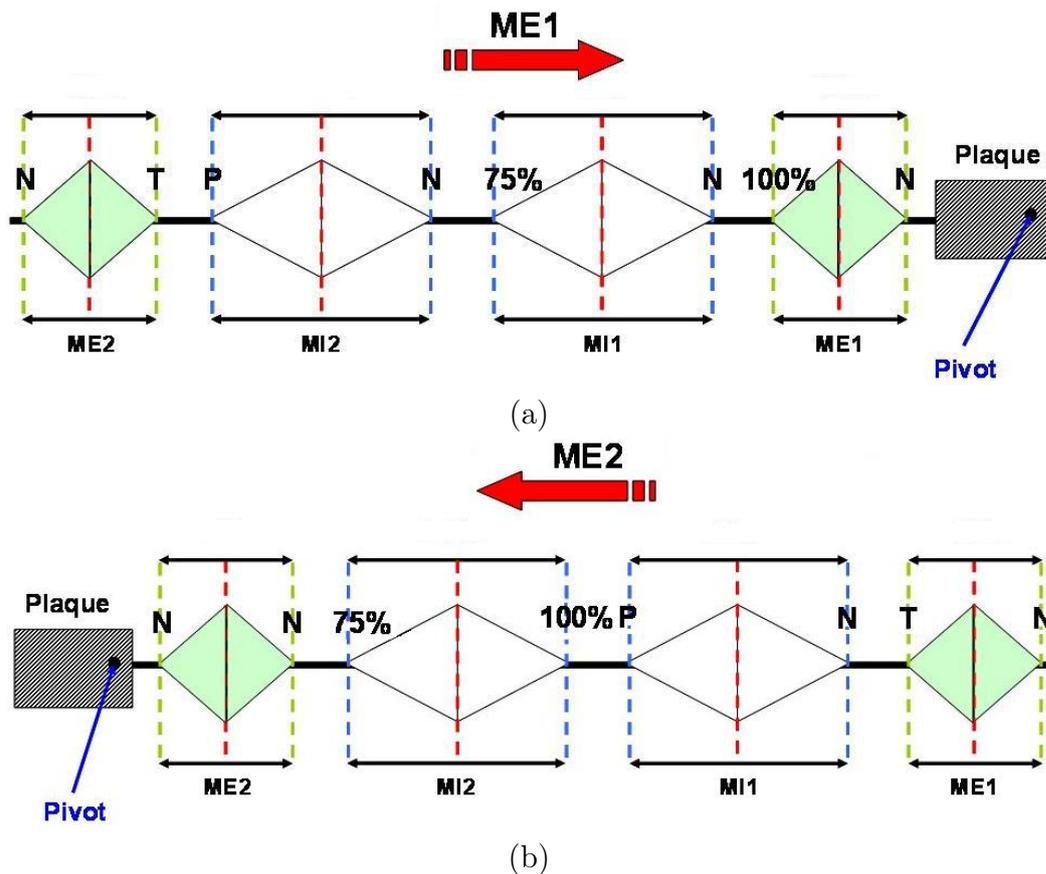


FIG. 3.5 – (a) La configuration 1 du montage des galets pour le sens $ME1$. (b) La configuration 2 du montage des galets pour le sens $ME2$.

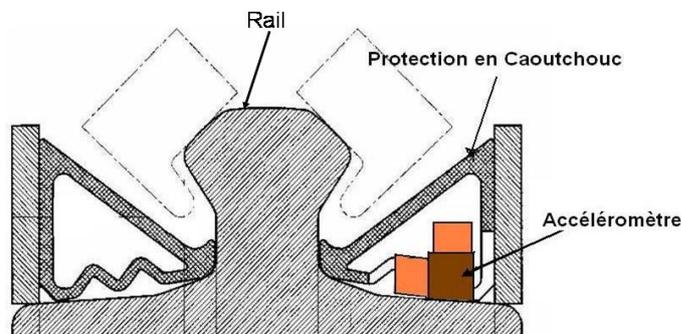


FIG. 3.6 – Emplacement de l'accéléromètre dans le rail.

figure 3.7(b) ;

- deux microphones posés sur la voie près du rail, figure 3.7(c), pour l'acquisition des signaux acoustiques. L'idée d'utiliser le signal acoustique comme complément d'information a été rapidement abandonnée par la suite ;
- une cellule photoélectrique et deux scotchs réfléchissants, positionnés à l'avant et l'arrière du tramway pour détecter l'entrée et la sortie du tramway dans la zone d'acquisition, figure 3.7(d) et (e). La distance entre les deux scotchs réfléchissants est de 12.32 m.

Le schéma 3.8(a) représente l'ensemble de l'installation matérielle et les distances



(a)



(b)



(c)



(d)



(e)

FIG. 3.7 – (a) Le premier démonstrateur Translohr utilisé lors des essais SYSCOGA-I. (b) Huit accéléromètres superposés 2 à 2 (4 latéraux et 4 verticaux). (c) Deux microphones pour l’acquisition des signaux acoustiques. (d) Cellule photoélectrique. (e) Le trait vertical en blanc représente le réflecteur positionné sur le tramway.

entre les dispositifs de mesure (en cm). La cote "A" a eu les valeurs suivantes : 13.5 cm, 10 cm et 7 cm.

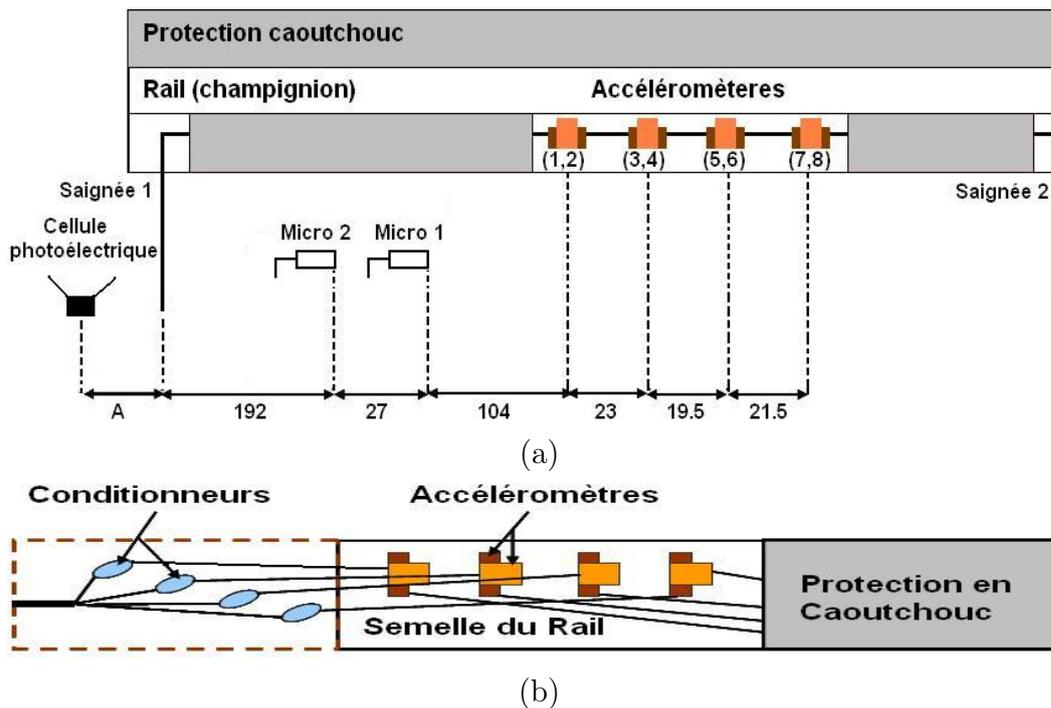


FIG. 3.8 – (a) Configuration matérielle SYSCOGA-I. (b) Position des conditionneurs pour accéléromètres sous la gaine en caoutchouc.

3.3.2 Configuration technique du véhicule et des défauts

Deux configurations de défauts ont été testées :

- (i) configuration C1 : guidage en parfait état technique (roulement, patins,...) et présence de défaut uniquement sur les galets avec configuration symétrique AVD/AVG
 - passages en mode LAC (MGC) ;
 - passages en mode batterie (MGA) ;
- (ii) configuration C2 : la même configuration des galets utilisée dans C1 avec la mise en place d'un défaut unique sur chaque essieu (patin, roulement, vérin,...)
 - passages en LAC avec frein électrique dans la zone de mesure en marche avant (AV) ;
 - passages en LAC à vitesse stabilisée en marche arrière (AR) ;
 - véhicule en position basse.

3.3.3 Problèmes rencontrés

Dans cette étape de recueil de données, nous avons été confrontés à un ensemble de problèmes liés directement à la nouveauté et la méconnaissance du système de guidage, aux conditionnements de l'environnement électrique et électromagnétique

du tramway qui influent sur la qualité des signaux acquis par les accéléromètres et les cellules photoélectriques. D'autres problèmes sont liés plutôt aux conditions climatiques (pluie, neige, humidité, etc.).

(a) Problèmes techniques

Lors de cette première campagne d'essais, nous avons rencontrés plusieurs problèmes dont la plupart purement mécaniques ayant pour cause :

- signaux d'accéléromètres obtenus avec les passages en mode LAC très bruités ;
- bruits électrique dus au retour de courant qui ont été évité par la suite grâce à l'accélération par l'alimentation par caténaire (LAC) mais avec frein électrique à 100% dans la zone de mesure pour renvoyer la tension dans la LAC et enfin un redémarrage de l'alimentation. Cette solution a améliorée nettement la qualité des signaux et c'est celle utilisée dans la suite de la première campagne d'essai SYSCOGA-I avec des mesures en marche avant (AV) seulement ;
- bruit dû au chargeur de la batterie qui est moins important que celui dû à la caténaire ;
- problèmes de mesures sur les accéléromètres dus à l'humidité dans la gaine en caoutchouc des conditionneurs situés proches des accéléromètres, figure 3.8(b) ;
- positionnement de la balise et des deux réflecteurs, figure 3.7(d) et (e) ;
- déclenchement de la balise lors du passage du tramway à des grandes vitesses $\gg 40$ km/h ;
- erreur de vitesse obtenue à partir du compteur Translohr ($V = 40$ km/h) et celle mesurée avec la cellule photoélectrique ($V = 35$ km/h).

Certains de ces problèmes ont été résolus lors de la deuxième et la troisième campagne d'essais SYSCOGA-II et SYSCOGA-IIbis détaillées dans l'annexe B. La version finale du détecteur SYSCOGA a été validée lors de la quatrième campagne d'essais SYSCOGA-III.

(b) Problèmes climatiques

Le problème le plus gênant et le plus souvent rencontré est la présence d'eau sur la voie et sur l'appareil de voie en raison de chutes de pluie. Cependant, dans l'installation sur l'appareil de voie, l'eau, contrairement à l'installation sur la voie, n'est pas en contact direct avec les accéléromètres. Étant donné que les rails de la plaque ne sont pas en contact avec le rail de la voie ; l'eau s'évacue à travers les disjonctions de la plaque avec le rail aux deux extrémités. Toutefois, quelques signaux vibratoires fournis par les accéléromètres sont bruités. Les cellules photoélectriques sont à leurs tour affectées. Par conséquent, la qualité de la détection a été dégradée et d'autres créneaux parasites apparaissent de part et d'autre des passages d'essieux en raison de l'eau projetée par le passage du Translohr.

3.3.4 Premières analyses

À l'issue de l'étape d'acquisition on dispose, pour chaque passage du tramway, de six signaux fournis par les trois accéléromètres (horizontal & vertical) et de deux signaux fournis par les deux cellules photoélectriques. Les signaux vibratoires représentent l'accélération en ms^{-2} et les signaux des cellules photoélectrique représentent leur tension en volt. Si la cellule photoélectrique détecte à un instant t le passage du tramway, la valeur correspondante est supérieure à zéro, $V_{cell} > 0$, sinon $V_{cell} = 0$.

Durant une grande partie du projet, la phase de prétraitement comprenant le filtrage et l'extraction du segment correspondant à chaque couple de galets à partir du signal complet a été effectuée par la société SD. Un couple de galets est représenté par une matrice CGk , de valeurs en dB, de dimension (8×512) , dans SYSCOGA-I, et (6×1024) dans SYSCOGA-III. Les colonnes représentent les accéléromètres. Les lignes représentent les fréquences ν avec $\nu \in [0 - 22007]$ Hz avec un pas de 44 Hz pour SYSCOGA-I et $\nu \in [0 - 6000]$ Hz pour SYSCOGA-III. Les matrices CGk résultent de la transformée de fourier à fenêtre, du signal vibratoire, appliquée par la société SD dans cette phase de paramétrisation des signaux.

Chaque couple de galets est donc caractérisé par un vecteur forme ($x = x_1, \dots, x_p$) avec $p = 4096$, dans SYSCOGA-I, et $p = 6144$ variables (descripteurs) dans SYSCOGA-III. Pour les essais SYSCOGA-III, nous avons étudié essentiellement les 4 classes : Sans défaut, avec trous, l'usure 100% et l'usure patatoïde. Pour chaque classe, 100 passages valides ont été retenus. Vu la différence entre les passages allers et les passages retours une base de données est construite pour chaque sens de marche, $N_{sys-III A}$ et $N_{sys-III R}$ de dimension (100×1024) , à partir des vecteurs formes. Ces bases d'apprentissage sont utilisées ensuite dans l'étape de classification pour définir l'état de chaque couple de galets. Malheureusement le nombre important des p variables caractéristiques par couple de galet ne facilite pas l'étape de classification et nécessite donc une étape de réduction de dimension. Pour réduire la dimension des p variables nous avons utilisés plusieurs techniques notamment de sélection de variables tels que les algorithmes séquentiels (SFFS et SFBS). Le chapitre 5 détaille l'ensemble des méthodes les plus citées dans la littérature et celles utilisées dans notre travail.

L'objectif de cette première partie du projet était de trouver les fréquences ou proprement dit la gamme de fréquences représentatives pour chaque classe. Cependant, la sélection de variables n'a pas permis de séparer nettement les différentes classes étudiées. Cet échec est dû essentiellement aux raisons suivantes :

- Le mélange de deux types d'usure sur le même couple de galets ;
- Le signal vibratoire (l'accélération) est de nature non-stationnaire, les phénomènes existants dans le signal ne peuvent pas être détectés par une simple transformée de Fourier. Ce point est amplement détaillé dans la partie 4.1.1 du chapitre 4 ;
- Deux types de défauts peuvent être rencontrés, locaux et structuraux, (cf. 1.5.3). Le défaut local touche la surface du galet (bandage) et se manifeste en hautes fréquences. Par contre, le deuxième type se manifeste en basses fré-

quences. Un même galet peut comporter les deux types de défauts en même temps. Une transformée de Fourier (FFT) ne permet pas d'avoir les deux informations à la fois ;

- Les deux raisons les plus importantes de l'échec à ce stade, à notre avis, sont la méconnaissance du comportement du système de guidage et les différents états de fonctionnement anormaux (défauts) : système nouveau, l'ignorance et le manque d'une partie de compréhension et d'analyse des signaux.

3.3.5 Redirection de la thèse

À ce moment de la thèse, nous avons fixé d'autres objectifs et la problématique du projet et de la thèse s'est réorientée d'une approche d'apprentissage pure vers "*l'analyse des signaux non-stationnaires multi-composantes*". Ce changement de vision vis-à-vis de la nature des signaux traités impose une redirection ou une reformulation du projet. Le signal temporel, et non plus des FFTs, est utilisé cette fois-ci comme matière première pour effectuer l'étape de diagnostic. Ainsi, la partie de génération des signatures et le repérage des segments correspondants aux passages des couples de galets sont intégrés par la suite dans notre travail de thèse.

L'objectif est de bien décrire les phénomènes existants dans le signal afin de détecter aussi bien un défaut local (hautes fréquences) qu'un défaut structurel (basses fréquences). La question qui vient à l'esprit maintenant est la suivante : "*partant du signal vibratoire, existe-il une représentation avec laquelle on peut comprendre son comportement et suivre son évolution ?*". Plusieurs méthodes existent dans la littérature et permettent de répondre à cette question. Le choix étant difficile et strictement lié aux attentes fixées et l'application elle-même, le chapitre 4 détaille cet aspect du travail.

3.4 Première analyse des signaux

L'ensemble des signaux obtenus lors des campagnes d'essais sont prétraités (débruitage, filtrage, etc.), dans un premier temps, par la société SD. Dans la suite, on considère la moyenne du signal latéral et vertical par capteur. En conséquence, chaque passage est représenté par seulement trois signaux correspondant aux trois accéléromètres. Ensuite, les signaux résultants subissent un découpage temporel afin de localiser les segments du signal associés aux couples de galets. Ainsi, pour chaque passage, on obtient huit segments correspondants aux huit couples de galets. L'identification de l'état d'un couple de galets donné est ensuite assurée par le module de diagnostic. Toutefois, avant de décrire en détail cette procédure de segmentation, il semble intéressant de souligner quelques remarques qui ont attiré notre attention sur l'aspect physique et les points caractéristiques des signaux vibratoires pour les deux installations, sur la voie et sur l'ADV.

3.4.1 Descriptif des signaux SYSCOGA-I à IV (sur la voie)

À l'issue de l'étape d'acquisition on dispose, pour chaque passage du tramway, de six signaux fournis par les trois accéléromètres (horizontal & vertical) et de deux signaux fournis par les deux cellules photoélectriques, figure 3.9.

Au passage d'un essieu, le signal issue d'une cellule photoélectrique se décompose en trois zones correspondant respectivement aux détections du couple de galets avant puis des roues pneumatiques et enfin du couple de galets arrière ce qui donne pour 4 essieux 4 groupes de 3 créneaux (1 par essieu), figure 3.10(a). Ainsi, pour chaque couple de galets, la première cellule photoélectrique fournit la date d'entrée dans la zone d'acquisition tandis que la deuxième cellule photoélectrique fournit la date de sortie, figure 3.10(b). Cette information sera utile pour la correspondance temps-position afin de localiser les segments du signal associés aux couples de galets.

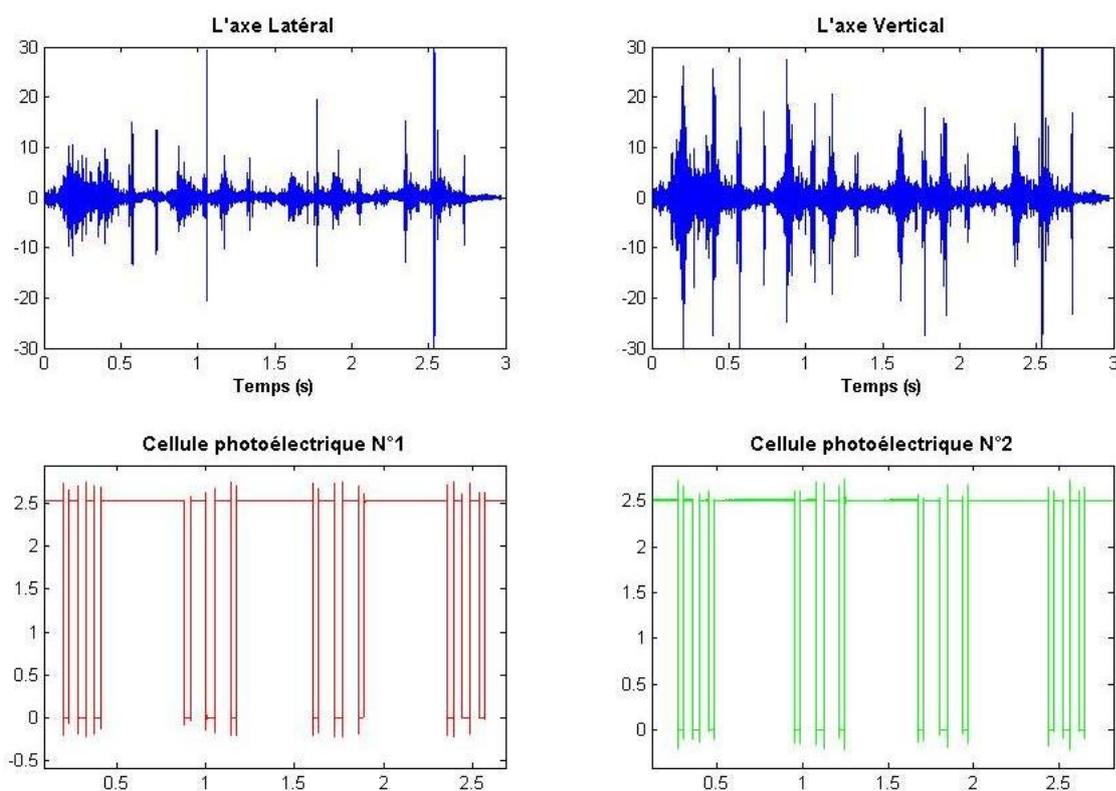
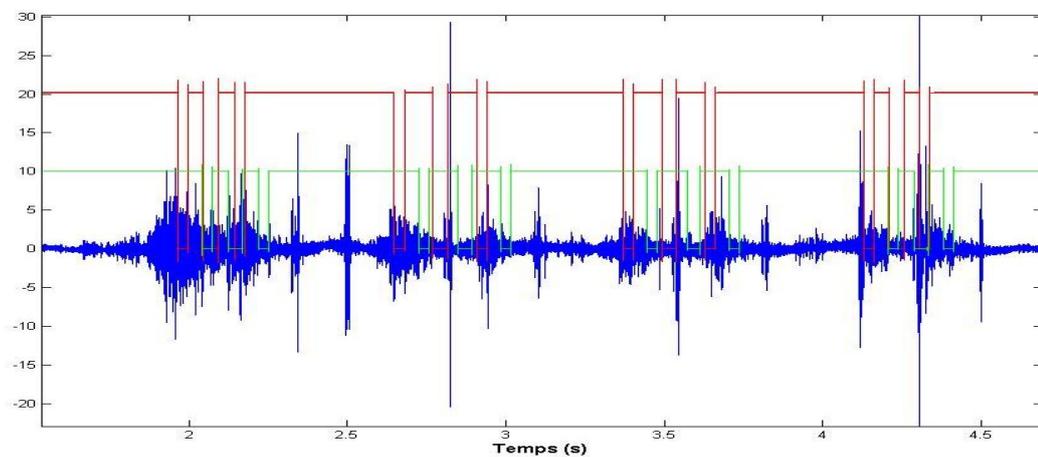


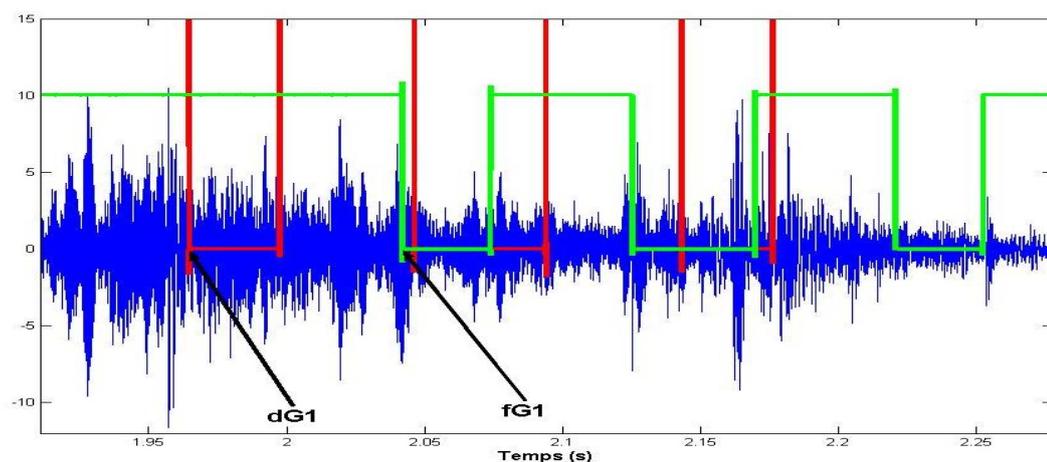
FIG. 3.9 – Les signaux vibratoires en unités arbitraires acquis avec l'installation sur la voie, SYSCOGA I-IV : l'axe latéral et l'axe vertical d'un capteur. Les deux cellules photoélectriques fournissent les dates d'entrée et de sortie des couples de galets et des roues.

Découpage temporel

Pour chaque passage du tramway, le signal acquis représente l'information superposée provenant simultanément des huit couples de galets CG_k , $k = 1, \dots, 8$, figure 3.10(a). Ainsi, au passage d'un essieu, le signal se décompose en trois zones, figure 3.10(b), correspondant respectivement aux passages du couple de galets avant



(a)



(b)

FIG. 3.10 – (a) Le signal vibratoire acquis avec l'installation sur la voie : trait rouge représente la cellule photoélectrique N°1 et le trait vert représente la cellule photoélectrique N°2. (b) Zoom sur un essieu : l'intervalle $[dG1, fG1]$ représente respectivement les moments d'entrée, repérés par la 1ère cellule, et de la sortie, repérés par la 2ème cellule, du premier couple de galets dans la zone d'acquisition.

puis des roues pneumatiques et enfin du couple de galets arrière. "*Comment peut-on avoir une correspondance temps-position afin de localiser les segments du signal temporel associés aux couples de galets ?*".

Pour chaque capteur, on s'attend à ce que le signal à un instant donné soit dominé par le comportement du couple de galet le plus proche du capteur à cet instant, figure 3.11. Une segmentation temporelle basée sur les dates de début et fin de passage des galets/roues, fournies par les deux cellules photoélectriques, permet de repérer la position exacte de chaque couple de galet sur le signal fourni par un capteur k donné. Si un couple de galets est plus proche du capteur k que des autres, le signal correspondant à la tranche $[\alpha_{k-1}, \alpha_k]$ lui est affecté, voir figure 3.11(a).

Introduisons les notions suivantes :

- t_{Gk}^i représente le début d'acquisition pour le ki -ème galet donné par la première cellule photoélectrique ;
- t_{Gk}^f est le top de la fin d'acquisition du ki ème fournit par la seconde cellule photoélectrique ;
- $t_{Gk}^i = (T1_{debut} + T0_{debut})/2$;
- $t_{Gk}^f = (T1_{Fin} + T0_{Fin})/2$;
- la vitesse est donnée par $V = L/T$. Tels que L_{Gk} , V_{Gk} et $T_{Gk} = t_{Gk}^f - t_{Gk}^i$ représentent respectivement la longueur parcourue, la vitesse du tramway et enfin la durée d'acquisition pour le ki ème couple de galets.

L'objectif est de trouver les segments du signal $[t_{Gk}^i, t_{Gk}^f]$, correspondants aux k couples de galets $k \in 1, \dots, 8$, à partir des dates de référence fournies par les deux cellules photoélectriques voir figure 3.11(b) et (c). Pour ce faire, on procède comme suit :

Posant X_{Gk} la position du ki ème couple de galets par rapport à un capteur donné, $X_{Gk} \in [0, L]$, figure 3.11(a), avec $t \in [t_{Gk}^i, t_{Gk}^f]$. Pour chaque capteur, on extrait le segment du signal correspondant à la position X_{Gk} du couple de galets Gk considéré. Par conséquent, trois possibilités se présentent :

- **si Gk est en face du premier capteur a alors**
 $t \in [a_k^i, a_k^f] \Leftrightarrow X_{Gk} \in [\alpha_0, \alpha_1] \Leftrightarrow t \in [t_{Gk}^i + \alpha_0/L_{Gk} * T_{Gk}, t_{Gk}^i + \alpha_1/L_{Gk} * T_{Gk}]$;
- **si Gk est en face du deuxième capteur b alors**
 $t \in [b_k^i, b_k^f] \Leftrightarrow X_{Gk} \in [\alpha_1, \alpha_2] \Leftrightarrow t \in [t_{Gk}^i + \alpha_1/L_{Gk} * T_{Gk}, t_{Gk}^i + \alpha_2/L_{Gk} * T_{Gk}]$;
- **enfin, si Gk est en face du dernier capteur c alors**
 $t \in [c_1^i, c_1^f] \Leftrightarrow X_{Gk} \in [\alpha_2, \alpha_3] \Leftrightarrow t \in [t_{G1}^i + \alpha_2/L_{Gk} * T_{Gk}, t_{Gk}^i + \alpha_3/L_{Gk} * T_{Gk}]$.

On se basant sur ces intervalles, on réalise le découpage temporel sur le signal vibratoire. Ainsi huit segments en résultent et correspondent aux huit couples de galets, figure 3.12.

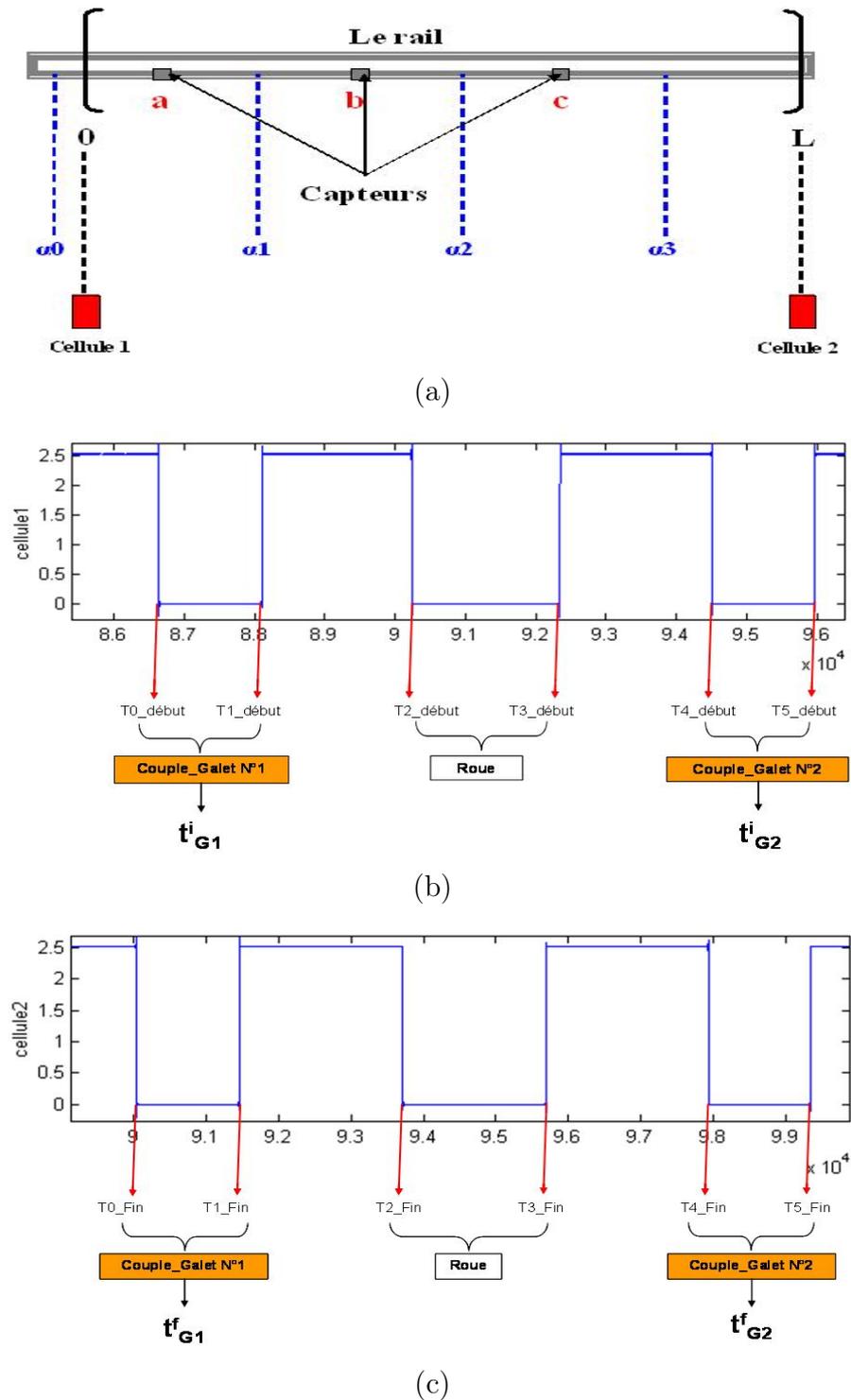


FIG. 3.11 – (a) Le découpage temporel : Le premier capteur, "a", est espacé de 6.5cm de la cellule $N^{\circ}1$ et du capteur $N^{\circ}2$, "b", de 21.1cm. La distance entre le capteur $N^{\circ}2$ et le capteur $N^{\circ}3$, "c", est égale à 21cm. Enfin la distance entre le capteur $N^{\circ}3$ et la cellule $N^{\circ}2$ est égale à 25.6cm. Les tranches $[\alpha_{k-1}, \alpha_k]$ représentent les segments correspondants au passage d'un galet devant un capteur k . La longueur totale $L = 74.2\text{cm}$. Dates d'entrées (b) et de sorties (c), par essieu, respectivement du couple de galets $N^{\circ}1$, de la roue et enfin du du couple de galets $N^{\circ}2$ données par les deux cellules photoélectriques 1&2.

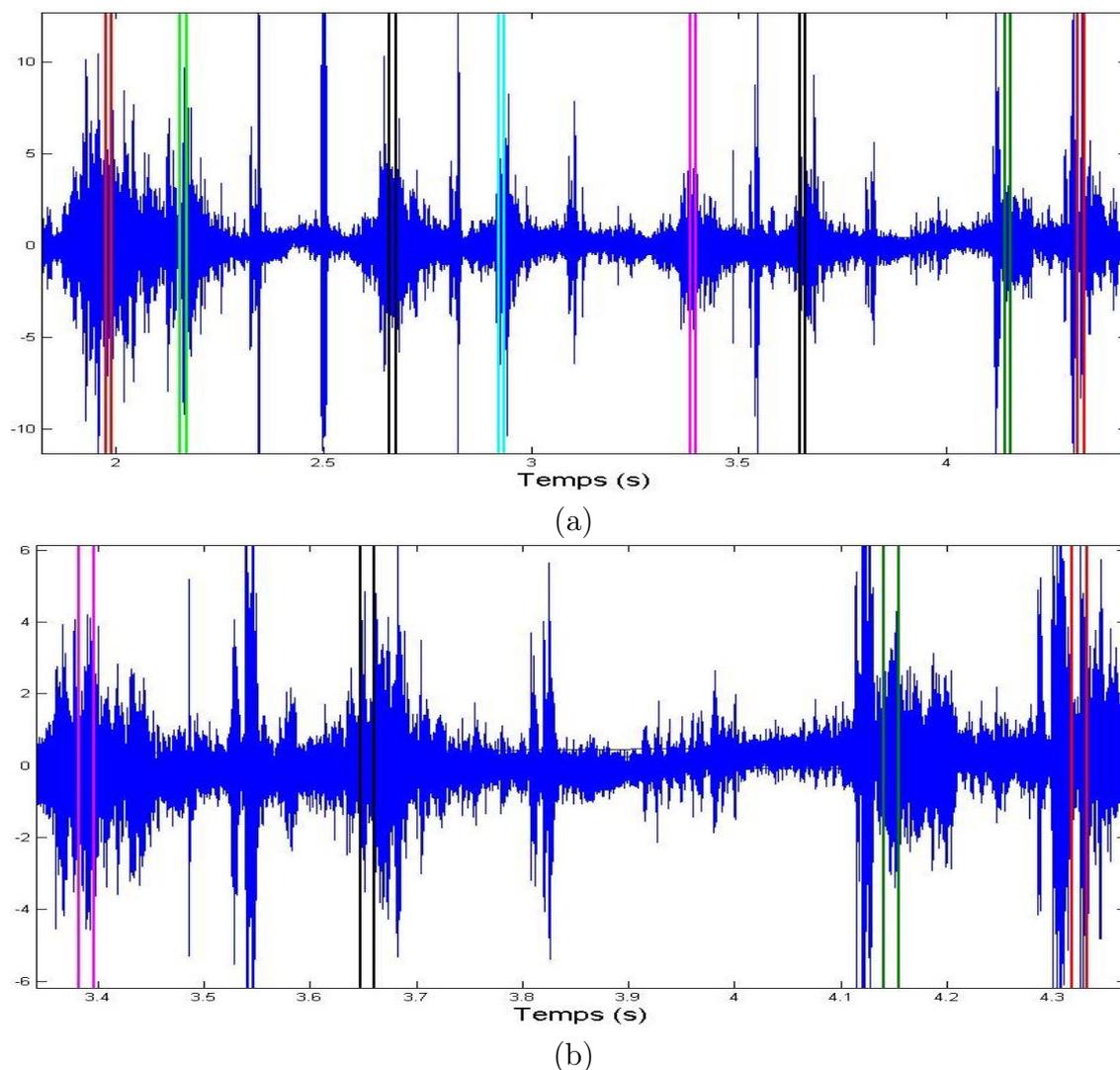


FIG. 3.12 – (a) Découpage temporel du signal en 8 segments correspondant aux 8 couples de galets : Installation sur la voie. Chaque essieu est composé de deux couples de galets (b).

On remarque, sur l'ensemble des signaux acquis avec l'installation sur la voie, une différence entre les passages allers et retours. Le signal vibratoire obtenu en sens retour est beaucoup plus amplifié que celui d'un passage aller. En fait, tout ce qu'on peut dire sur ces différences de comportement, pour l'instant, est qu'elles peuvent provenir soit de la position des galets sur le translohr, la structure du système de guidage et la manière dont les couples de galets enserrant le rail ou soit quelles sont tout simplement spécifiques à la piste d'essai et notamment à la position des accéléromètres par rapport aux joints entre les dalles.

On constate couramment, sur l'ensemble des signaux vibratoires, la présence de pics énergétiques hors synchronisme exact galet/capteur. Ces phénomènes peuvent s'expliquer par le passage des pneus sur les joints de dalles créant des vibrations qui se propagent ensuite dans le rail et/ou la voie. Ces chocs, figure 3.13 (pics A, B et C) sont aussi bien présents sur l'axe vertical que sur l'axe latéral, par contre ils sont moins importants sur ce dernier. Cette variation énergétique apparaît nettement sur la figure 3.13. Il est envisageable d'approfondir ce point, dans les prochains essais, pour mieux comprendre le comportement à la fois du défaut et du système de guidage

lui même. À ce stade du projet, on ne peut pas se prononcer sur les causes exactes de certains phénomènes rencontrés lors des essais. Toutefois, tous les points notés ici seront analysés lors des prochains essais. Les réponses à ces questions faciliteront la tâche de diagnostic.

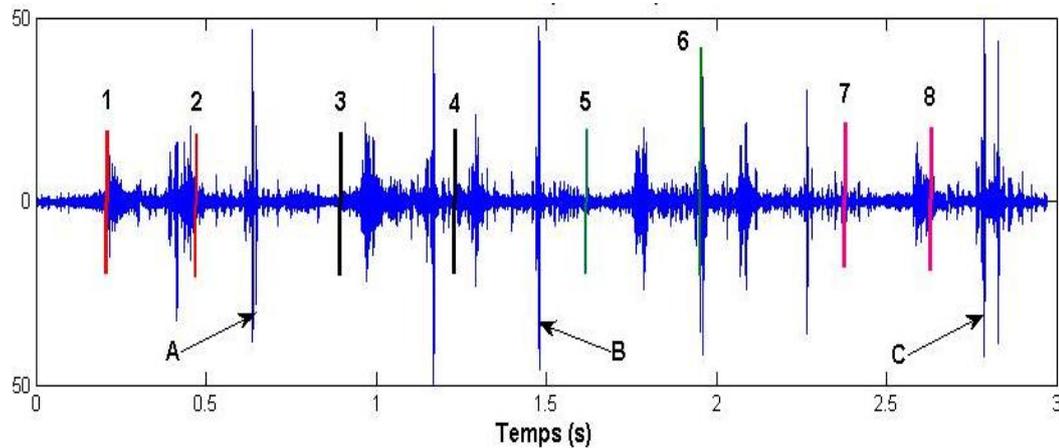


FIG. 3.13 – L’existence de pics parasites, (A, B,C), sur l’accélération latérale & verticale pour un passage donné. Les deux ensembles [1,3,5,7] et [2,4,6,8] représentent respectivement les entrées et les sorties des 4 essieux.

3.4.2 Descriptif des signaux SYSCOGA-V (sur l’ADV)

Au passage du tramway, le signal issu d’une cellule photoélectrique se décompose en quatre zones correspondant à la détection des quatre essieux, figure 3.14(a). Contrairement à l’installation sur la voie la première cellule photoélectrique fournit cette fois-ci la date d’entrée de chaque essieu dans la zone d’acquisition tandis que la deuxième cellule photoélectrique fournit la date de sortie. On aura donc, pour les 4 essieux, 4 groupes d’un seul créneau (1 par essieu). Cette information sera utile pour extraire les segments du signal correspondants aux essieux et ensuite ceux correspondants aux couples de galets. La correspondance temps-position est illustrée dans la partie réservée au découpage temporel.

Malgré nos réticences a priori, l’utilisation de la plaque d’aiguillage a permis de répondre à plusieurs questions posées préalablement dans l’installation sur la voie. Avec cette configuration, les variations en fonction du sens de circulation, figure 3.15(a), semblent plus compréhensibles contrairement à l’installation sur la voie. De même, avec cette installation, le signal vibratoire obtenu dans le sens ME2, figure 3.15(c), c’est à dire avec l’essieu ME2 en tête (coté pivot) est plus amplifié que celui d’un sens ME1 avec l’essieu ME1 en tête, figure 3.15(b). Cette fois-ci, la cause de ces différences de comportement est connue : elle est due à la dissymétrie et à la manière dont la plaque est fixée, notamment son point d’appui au sol (le pivot). Comme le montre la figure 3.4(b) la plaque d’aiguillage repose sur un pivot qui permet le changement de voie de roulement. Le pivot, dans cette configuration, sert simplement d’appui sans aucune fixation.

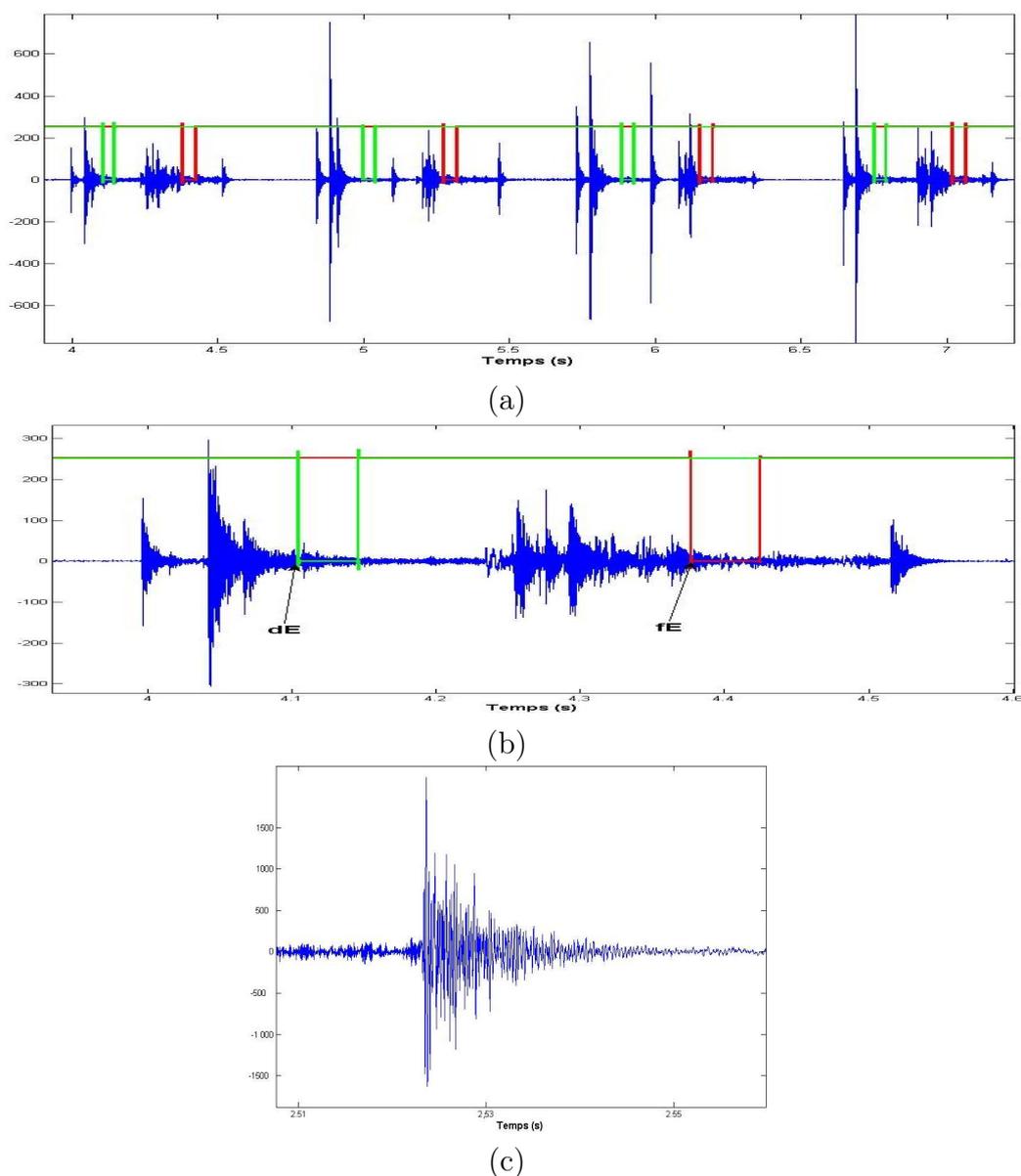


FIG. 3.14 – (a) Le signal vibratoire acquis avec l’installation sur l’appareil de voie : trait vert représente la cellule photoélectrique N°1 et le trait rouge représente la cellule photoélectrique N°2. (b) Zoom sur un essieu : l’intervalle $[dE, fE]$ représente respectivement le moment d’entrée, repérée par la 1ère cellule, et de la sortie, repérée par la 2ème cellule, d’un essieu dans la zone d’acquisition. (c) Un choc amorti. Zoom sur un choc provoqué par l’entrée ou la sortie d’un galet.

Un autre point positif concerne la localisation des couples de galets qui devient naturelle et instinctive. Le découpage des signaux est réalisé à moitié par la plaque elle-même, les essieux sont nettement séparés. On remarque sur la figure. 3.16(a), l’existence de quatre groupes concentrés et bien hétérogènes représentant chacun le passage d’un essieu sur la plaque d’aiguillage. Cela va simplifier considérablement la localisation des couples de galets et ainsi le découpage temporel. Cependant, certains points restent à explorer telles que les variations de la réponse de la plaque en fonction de la position d’un même couple de galets sur la rame.

Cependant, on constate également avec cette installation la présence de chocs sur

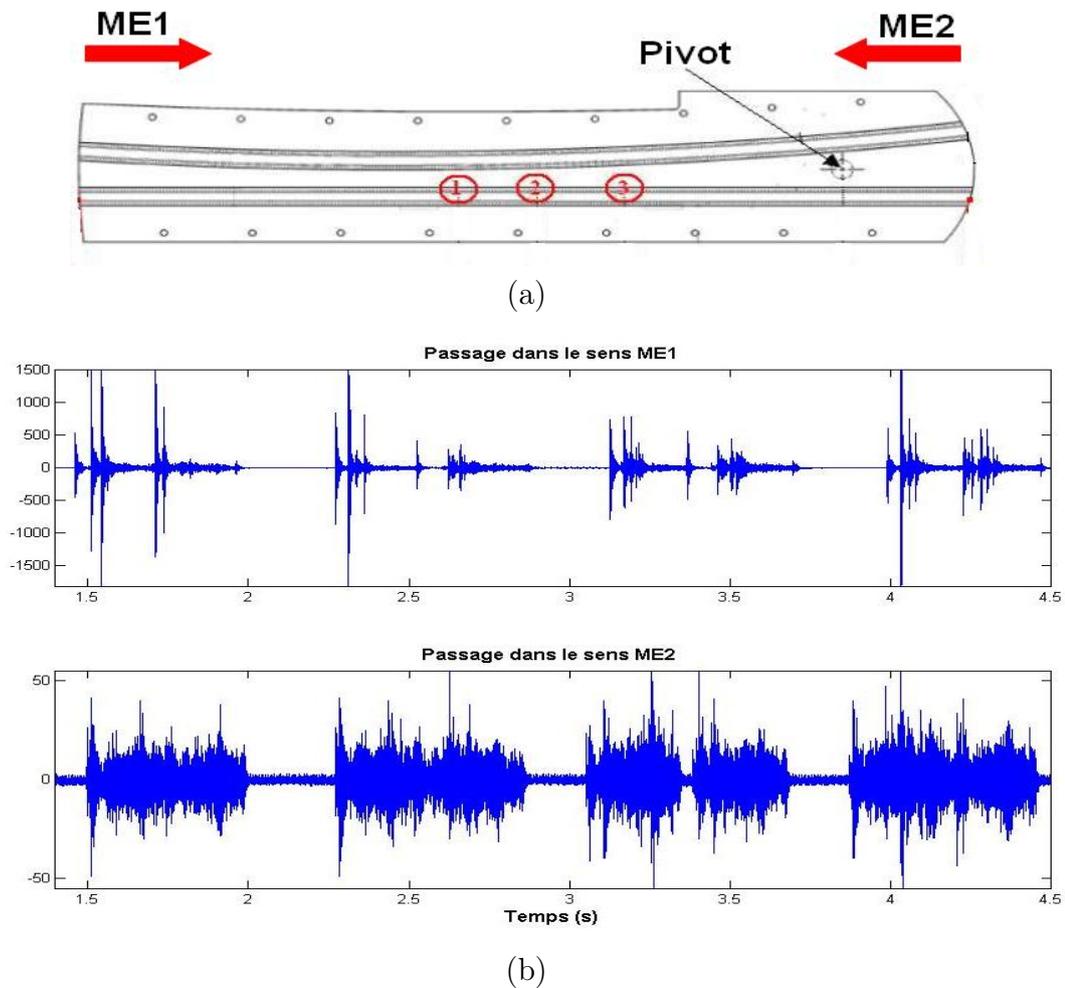


FIG. 3.15 – (a) Le sens de circulation du Translohr : installation sur l'appareil de voie. (b) Signal vibratoire, en unités arbitraires, acquis dans les deux sens ME1 et ME2.

les signaux vibratoires comme ceux déjà rencontrés dans l'installation sur la voie. Ces chocs sont présents en dehors du passage des galets sur les accéléromètres et peuvent s'expliquer, en plus des raisons préalablement cités dans la partie 3.4.1, par la résonance de la plaque d'aiguillage. Ces chocs (2, 3 et 6), figure 3.16(b), sont aussi bien présents dans le sens ME2 que dans le sens ME1. Ils représentent un phénomène type de choc puis oscillation amortie exponentiellement, figure 3.14(c).

Découpage temporel

Pour chaque signal on regarde les points caractéristiques essieu par essieu. On donne une estimation, en secondes et en mètres, de la distance entre chaque point et le début du signal. Pour illustrer la procédure du découpage temporel, prenons les signaux vibratoires présentés dans les figures 3.18(a) et 3.19(a), acquis respectivement dans le sens ME1 et le sens ME2 avec les configurations 1 et 2. Généralement, les essieux ME1, MI1, MI2 et ME2 se comportent de la même manière dans un sens ou dans un autre. Cependant, l'étalement, l'amplitude des signaux ainsi que la propagation des vibrations dépendent, entre autres, de la position du pivot par

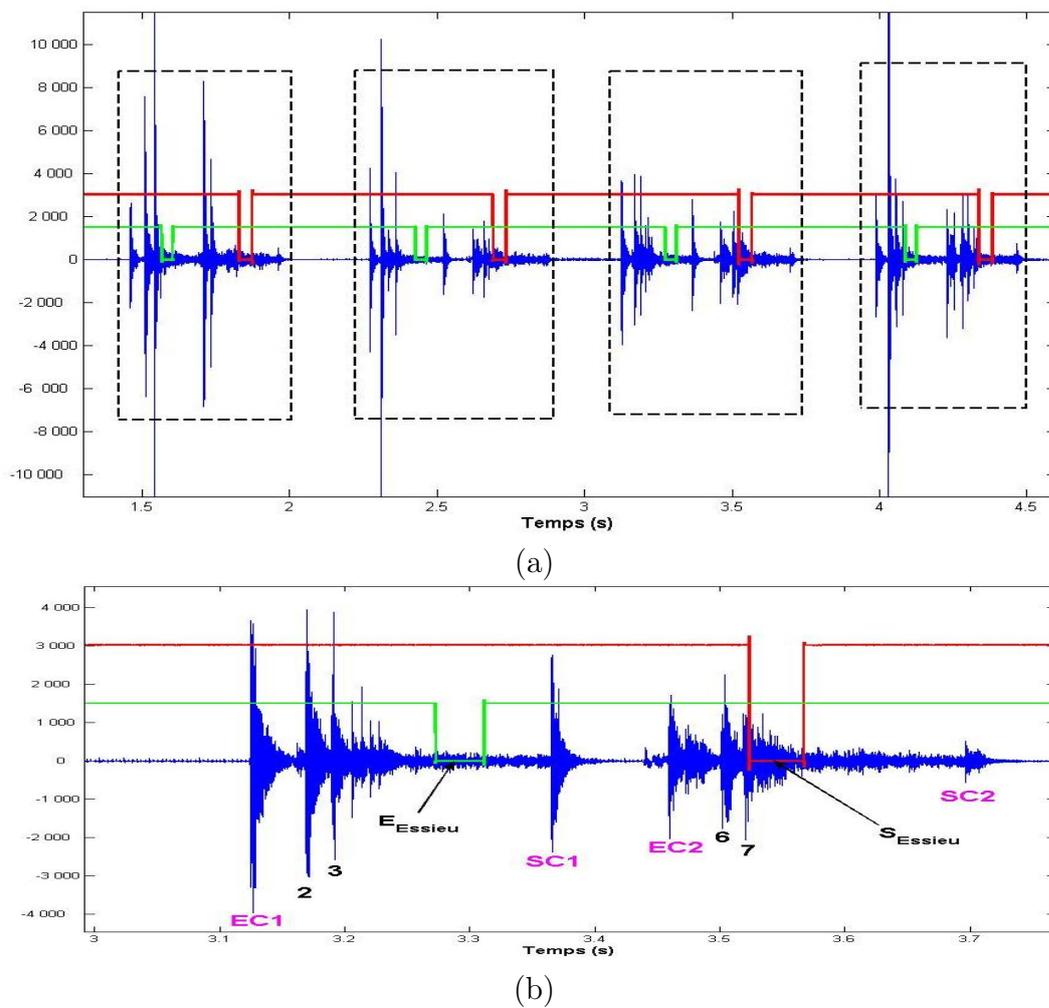


FIG. 3.16 – (a) Le passage des quatre essieux sur la plaque d'aiguillage. (b) Pics parasites sur le segment du signal correspondant au sens ME1C1 Essieu MI2. E_{Essieu} , EC1, S_{Essieu} et SC1 représentent respectivement l'entrée de l'essieu et du premier couple de galets et la sortie des deux de la zone d'acquisition (ADV).

rapport à l'essieu et du mécanisme de guidage.

Rappelons que la vitesse $V = L/T$ où L et T représentent respectivement, la distance réelle entre les deux balises (2.18 m), et le temps nécessaire pour parcourir cette distance. Dans le sens ME1 la distance entre la première balise (voie 11) et le pivot est de 1.74 m soit ≈ 0.23 s. Par contre dans le sens ME2, le pivot est à 0.44 m soit ≈ 0.058 s de l'entrée de l'ADV (la deuxième balise, voie 10). À partir de là nous pouvons calculer la vitesse du tramway à l'intérieur de l'essieu et effectuer ensuite le découpage temporel. Par exemple, pour l'essieu ME1, figure 3.18(b), la différence entre la première coupure balise 1 (premier trait rouge) et la première coupure balise 2 (premier trait vert) est égale à 0.27 s. Nous estimons donc la vitesse du tramway dans l'essieu ME1 à 8.01 m.s^{-1} (29 km.h^{-1}).

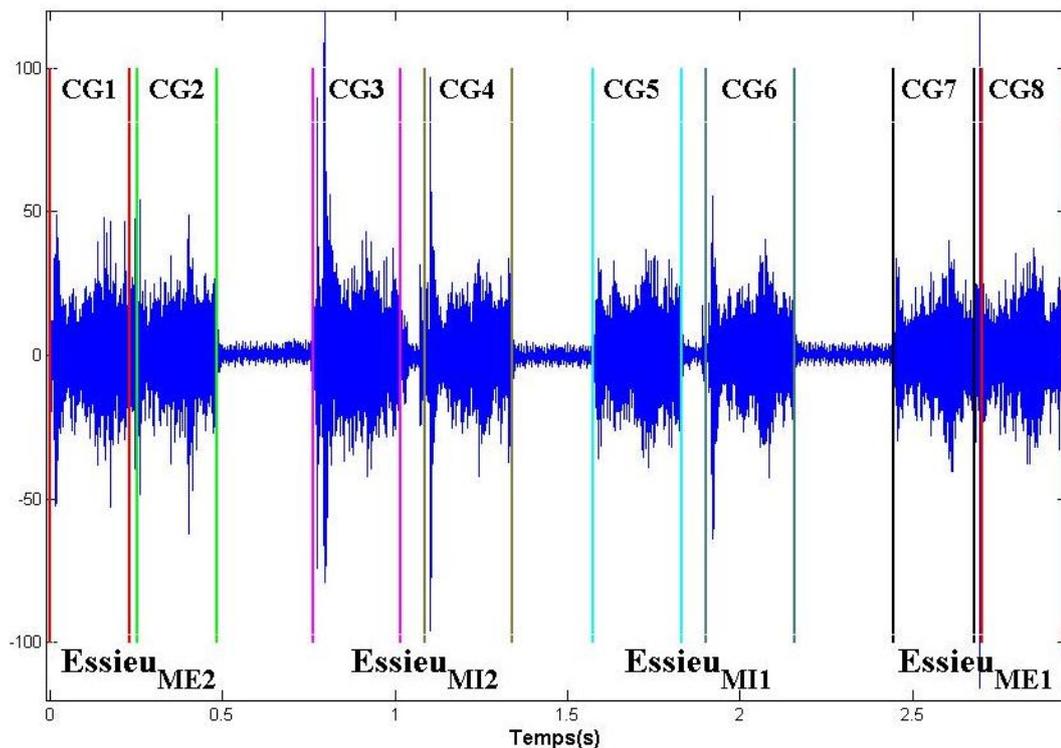


FIG. 3.17 – Découpage temporel du signal en 8 segments correspondant aux 8 couples de galets : installation sur l'ADV.

Nous nous basons sur ces données et nous découpons les signaux de façon à isoler chaque couple de galets. Nous considérons alors le moment où il est seul sur la plaque. D'après la distance entre le premier pic et le passage de l'essieu on peut estimer la distance intergalets à 1.98 m pour les essieux extrêmes et 2.78 m pour les essieux intermédiaires. L'écart entre cette distance estimée et la distance intergalets réelle est égale à -38 cm. Nous pensons que cette distance peut correspondre à la taille du patin, figure 1.2(c). Les signaux sont découpés ainsi :

Essieux extrêmes

Les 2 premiers mètres correspondent au passage seul du premier couple de galets et les 18 cm suivants correspondent à la sortie du premier couple de galets

et à l'entrée du deuxième couple de galets. Les 2 derniers mètres correspondent au seul passage du second couple de galets ;

Essieux intermédiaires

Pour les essieux intermédiaires, les 2.18 premiers mètres correspondent au seul passage du premier couple de galets. Durant les 60 cm suivants rien ne roule sur la plaque. Enfin, les 2.18 derniers mètres correspondent au passage seul du second couple de galets.

En procédant ainsi, les couples de galets appartenant aux essieux extrêmes et aux essieux intermédiaires sont représentés respectivement par des segments de signal $SCGk$, $k \in 1, \dots, 8$, correspondant à un déplacement du tramway de 2 m et de 2.18 m.

Quelques points caractéristiques

Globalement, pour un sens donné ME1 ou ME2, les essieux extrêmes ou intermédiaires se comportent de la même manière au contact de la plaque d'aiguillage. Toutefois, nous avons relevé quelques points caractéristiques pour chaque essieu et chaque couple de galets :

- ◇ l'existence d'un choc (pic) qui vient juste après l'entrée des couples CG1 et CG2 sur la plaque :
 - **dans le sens ME1**, figure 3.18(b) et (c), la distance moyenne entre :
 - l'entrée du CG1, pic N°1, et le deuxième pic N°2 est de 0.045 s soit 0.35 m ;
 - l'entrée du deuxième couple CG2, pic N°5, et le deuxième pic N°6 est de 0.017 s soit 0.12 m.
 - **par contre dans le sens ME2**, figure 3.19(b), la distance moyenne entre :
 - l'entrée du CG1, pic N°1, et le deuxième pic (N°2, *) est de 0.009 s soit 0.075 m ;
 - l'entrée du deuxième couple CG2, (*, pic N°3), et le deuxième pic N°4 est de 0.009 s soit environ 0.08 m ;
 - le pivot est à 0.051 s, soit 0.44 m, de l'entrée du premier couple CG1, pic 1, ce qui correspond \approx à la sortie du premier couple de galets CG1.
- ◇ dans le sens ME1, que ce soit dans un essieu extrême ou intermédiaire figure 3.18(b) et (c), l'entrée des deux couples de galets CG1 et CG2 sur l'ADV est représentée par 3 pics successifs et leur sortie par un seul pic. On peut estimer la moyenne de propagation de l'ensemble des pics à 0.035 s, soit 0.28 m. Néanmoins, il existe une différence significative au niveau de leur amplitude. Les chocs (pics) correspondants à l'entrée du couple CG1 sont beaucoup plus forts que ceux à la fois de la sortie de ce dernier ainsi que l'entrée et la sortie du deuxième couple CG2.

À notre avis, cela est dû aux caractéristiques propres à chaque couple de galets et de la plaque d'aiguillage. Étant donné que la plaque d'aiguillage est fixée du côté du pivot et libre ailleurs, plus la distance entre le couple de galet et le pivot est grande plus le choc est fort.

D'un autre point de vue, ce qui peut expliquer aussi la différence d'amplitude entre le passage respectif des couples CG1 et CG2 sur l'ADV est le mécanisme de guidage lui même. Chaque essieu est composé de deux couples de galets, CG1 et CG2, disposés symétriquement par rapport à l'axe de l'essieu. Il y a un système de vérins qui permet d'activer le couple situé en avant de l'essieu, le CG1, dans le sens de la marche et de laisser un jeu latéral à l'organe de guidage situé derrière, le deuxième couple CG2. Dans le sens de la marche, le premier couple CG1 est donc le couple de galets directeur et le couple CG2 est suiveur. Il est fort probable, que la façon dont le couple de galets enserre le rail, directeur ou suiveur, se traduit par une différence d'amplitude du signal correspondant à chaque couple.

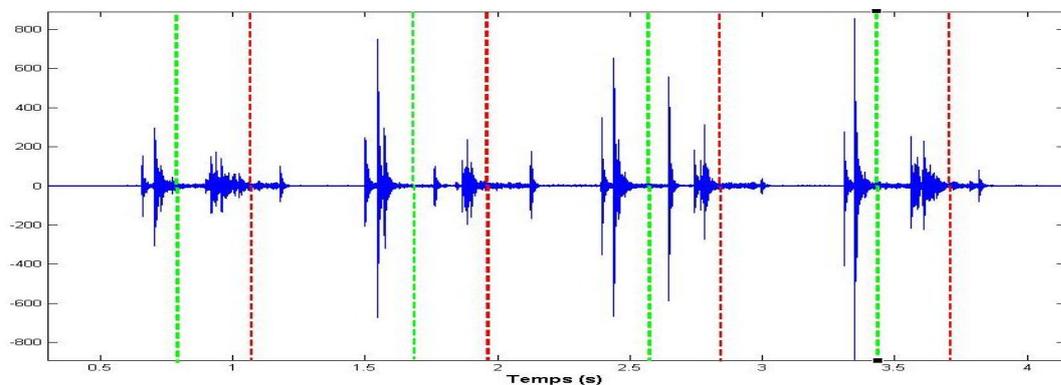
Les supports de chaque couple de galets, pour les essieux intermédiaires, portent des patins de retour de courant. Tandis que, les essieux extrêmes sont équipés de patins de mise à la masse, qui servent également de chasse-pierres. Cette différence peut aussi avoir une influence sur la nature du signal acquis. Ces supports ne sont peut être pas situés aux mêmes endroits relativement aux galets¹ ;

- ◇ l'existence, dans les deux sens ME1 et ME2, de pics parasites dont la présence ne peut être justifié pour l'instant. Par exemple, les pics (2, 3 et 6), essieu ME1 figure 3.18(b), et les pics ([2, 3];[6, 7]), essieu MI1, figure 3.18(c) ;
- ◇ dans le sens ME2 configuration 2 le signal est plat plus longtemps entre MI1 arrière/MI1 avant que entre MI2 arrière/MI2 avant. Il y a 0.30 m entre MI2 avant et MI2 arrière et 0.61 m entre MI1 avant et MI1 arrière. Ceci n'est vrai que pour le sens ME2 avec la configuration 2, dans les autres cas nous avons toujours de l'ordre de 0.60 m entre un MI avant et un MI arrière, figure 3.20.

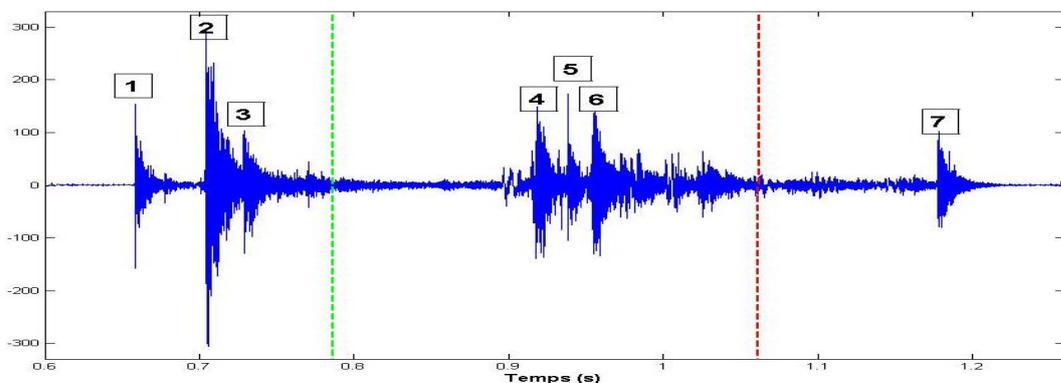
D'autres points caractéristiques ont aussi retenu notre attention lors de cette étape d'analyse des signaux. Nous avons remarqué quelques différences entre le passage du Translohr dans un sens ou dans un autre (ME1/ME2). Dans le sens ME1 la distance entre l'entrée de l'ADV et le pivot est de 1.74 m soit ≈ 0.23 s. Par contre dans le sens ME2, le pivot est à 0.44 m soit ≈ 0.06 s de l'entrée de l'ADV. Étant plus près du pivot de la plaque d'aiguillage, dans le sens ME2, l'ADV se cale (stabilise) dès l'entrée du premier couple de galets. De ce fait, l'étalement, la forme, l'amplitude des signaux ainsi que la propagation des vibrations sont différents, (cf. 3.4.2).

Certains points restent aussi à explorer tels que l'évolution du niveau d'usure survenue notamment dans le sens ME2. Ainsi que les variations de la réponse de la plaque en fonction de la position d'un même couple de galets sur la rame (galets directeur ou suiveur). Dans le sens de la marche, le système de vérin active le couple situé en avant de l'essieu et laisse un jeu latéral à l'organe de guidage situé en arrière, le deuxième couple. Nous pensons que la façon dont le couple de galets enserre le rail, directeur ou suiveur, se traduit par une différence d'amplitude du signal

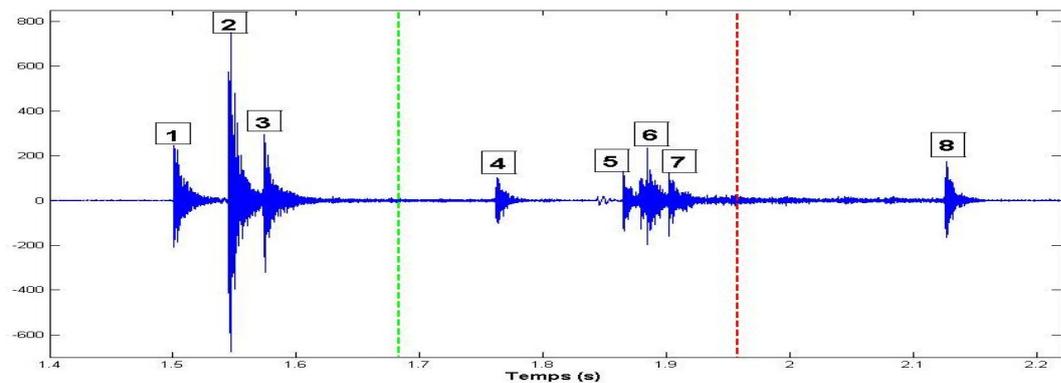
¹Cette hypothèse n'est pas vérifiée à cause de la méconnaissance des détails du système de guidage tenus par le secret industriel.



(a)

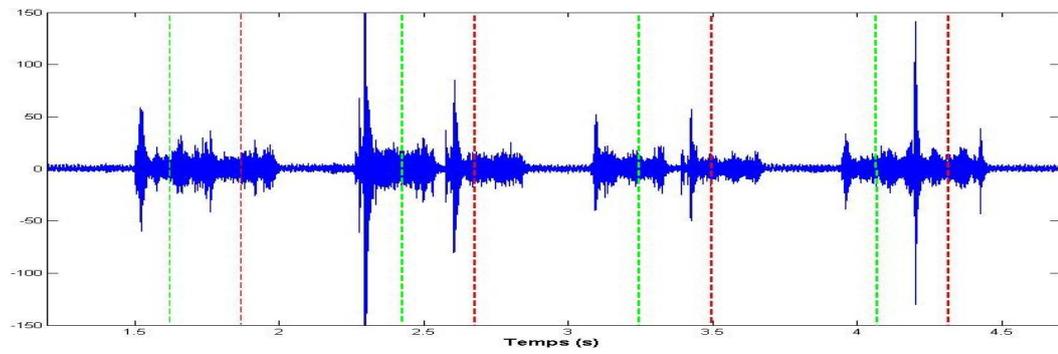


(b)

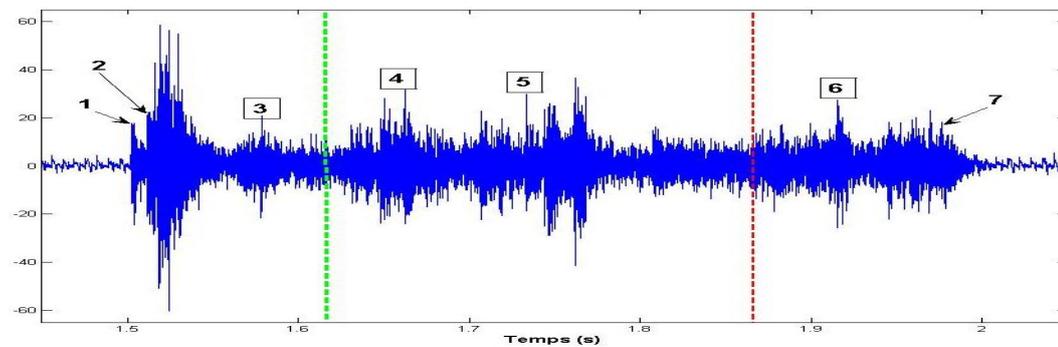


(c)

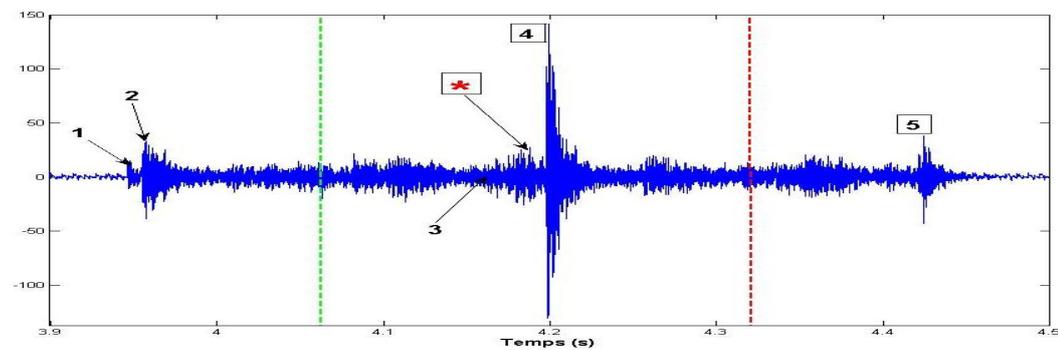
FIG. 3.18 – (a) Le signal correspondant au premier passage et la première voie dans le sens ME1. Extraits des signaux correspondants aux essieux (b) ME1 et (c) MI1. Les chiffres encadrés représentent le début de chaque choc pouvant correspondre à l'entrée ou la sortie du galet.



(a)



(b)



(c)

FIG. 3.19 – (a) Le signal correspondant au passage N°6 voie 1 dans le sens ME2. Extraits des signaux correspondants aux essieux (b) ME2 et (c) ME1.

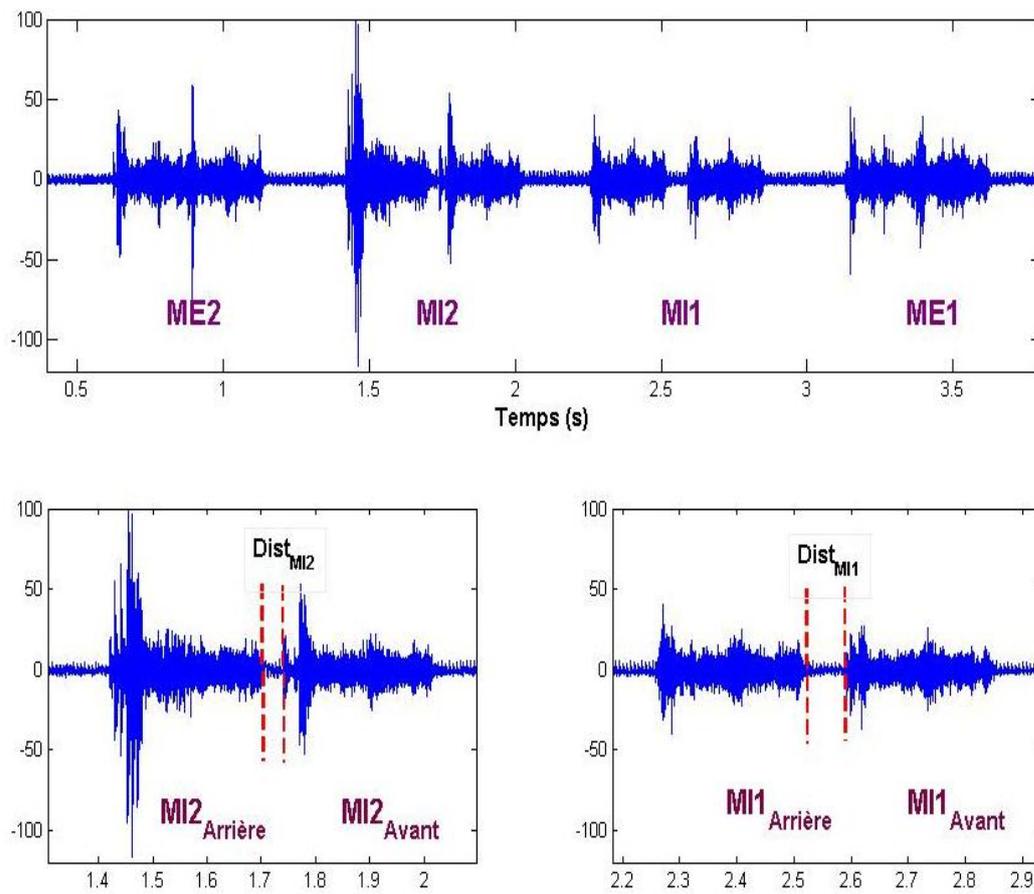


FIG. 3.20 – Différence entre MI1 et MI2 dans le sens ME2 configuration 2.

correspondant à chaque couple. Cela nous laisse croire aussi qu'un même défaut sur un couple de galets directeurs sera mieux détecté que celui sur un couple suiveur (la force et le type d'appui). Suite à cette étape d'analyse des signaux, (cf. 3.4.2), deux hypothèses h_1 et h_2 ont été posées selon lesquelles le taux de classification sera meilleur :

- h_1 : dans le sens ME2;
- h_2 : avec la configuration C2, à savoir les défauts sont montés sur les essieux avant, couple de galets directeurs. Les sections suivantes vont confirmer ou infirmer ces hypothèses.

Nous allons confirmer ou infirmer ces constats lors de l'étape de la classification et qui sera traitée dans le chapitre suivant.

Conclusion

Le but de ce chapitre était de choisir la bonne configuration matérielle pour enregistrer les signaux vibratoires. La difficulté était de trouver un compromis entre l'environnement acoustique du tramway, la dimension, la sensibilité et l'isolation des composants constituant le système d'acquisition. Pour répondre à ces besoins nous avons, dans un premier temps, détaillé les deux configurations matérielles utilisées pour l'enregistrement des signaux vibratoires. Dans la première configuration, le système d'acquisition a été placé sur la voie, (cf. 3.2.1). Alors que dans la seconde configuration, il a été installé sur l'appareil de voie (ADV), (cf. 3.2.2). Dans les deux cas, une structure multi-capteur a été utilisée pour enregistrer les vibrations du rail au passage du Translohr. Pour la suite du projet, c'est l'installation sur l'ADV qui est retenue pour l'intégration sur site. Par conséquent, les chapitres suivants détaillent les signaux associés. Néanmoins, quelques points caractéristiques des signaux issus de l'installation sur la voie ont été donnés à titre comparatif.

Dans un second temps, nous avons procédé à une première analyse des signaux vibratoires bruts avec une description du système "Tramway/Rail". Cette étape est indispensable à la bonne compréhension des données expérimentales. Elle nous a permis de segmenter le signal en temps. Permettant ainsi d'associer un segment de signal distinct à chaque couple de galets et éviter les interférences. Les signaux résultants de cette étape de perception représentent l'information a priori sur le fonctionnement du système de guidage et sont utilisés dans le module de diagnostic par reconnaissance de formes.

La suite du travail s'attachera à mettre au point une démarche de diagnostic de l'état des galets par RdF à partir de ces signaux. Cependant, la nature non stationnaire du signal vibratoire nécessite des méthodes temps-fréquence (temps-échelle) bien appropriées pour une meilleure exploration de ces propriétés. Il est donc utile de trouver l'approche la plus appropriée. Le chapitre suivant expose les méthodes de traitement de signal les plus utilisées.

4

ANALYSE TEMPS-ÉCHELLE DES SIGNAUX

En exagérant à peine, nous introduirons presque autant d'algorithmes d'analyse qu'il y a de signaux [178].

L'objectif du système de diagnostic est de déterminer le ou les galets défectueux (localisation temporelle) et de définir la nature du défaut (caractérisation fréquentielle). Nous avons donc besoin d'une information temps-fréquence, d'où l'utilisation d'une analyse temps-fréquence ou temps-échelle.

Nous présentons dans ce chapitre une synthèse sur les méthodes temps-fréquence en résumant les résultats importants d'une théorie qui a connu un fantastique essor au milieu des années 80. Nous rappellerons dans un premier temps quelques résultats sur la transformée de Fourier, pour motiver l'introduction des représentations temps-fréquence et pour fixer nos notations. Les éléments classiques sur les représentations conjointes en temps et en fréquence sont détaillés dans [88, 115, 170]. Ces méthodes sont utilisés en particulier dans le domaine des transports ferroviaires pour la détection des usures des rails [279, 69, 24] ainsi que dans le secteur industriel pour le diagnostic des machines tournantes et moteurs à combustion interne [274].

La seconde partie montre quelques résultats sur les représentations temps-fréquence (RTF) pour introduire la transformée en ondelettes. La dernière partie du chapitre est réservée à la transformée en ondelettes continue ou la représentation temps-échelle (RTE) et rappelle les propriétés fondamentales pour justifier son utilisation pour l'analyse de nos signaux. Ce chapitre est essentiellement bibliographique, il a été alimenté par la lecture de [52, 88, 201, 53, 170, 14].

4.1 Les représentations temps-fréquence et temps-échelle

4.1.1 La transformée de Fourier

Considérons le signal $x(t)$ à temps continu. Si x est d'énergie finie sa transformée de Fourier (TF) à la fréquence ν existe et appartient à L_2 [37]

$$X(\nu) = \int_{-\infty}^{+\infty} x(t)e^{-i2\pi\nu t} dt, \quad (4.1)$$

son inverse est donnée par :

$$x(t) = \int_{-\infty}^{+\infty} X(\nu)e^{i2\pi\nu t} d\nu. \quad (4.2)$$

Sa transformée de Fourier discrète (TFD) de N points avec une période d'échantillonnage T est donnée par

$$X\left[\frac{\nu}{NT}\right] = \frac{1}{N} \sum_{k=0}^{N-1} x(kT)e^{-\frac{i2\pi\nu kT}{NT}}, \quad (4.3)$$

et sa relation inverse est donnée par

$$x[kT] = \frac{1}{N} \sum_{k=0}^{N-1} X\left[\frac{\nu}{NT}\right] e^{\frac{i2\pi\nu kT}{NT}}. \quad (4.4)$$

Le signal sous sa forme fréquentielle $X(\nu)$ est le plus souvent étudié par l'analyse de la courbe $|X(\nu)|^2$ définissant l'amplitude du spectre à chaque fréquence. Cette transformation fait clairement ressortir le contenu fréquentiel du signal $x(t)$. En effet, la TF décompose le signal $x(t)$ sur une base de vecteurs propres sinusoïdaux $\{e^{i2\pi\nu t}\}_{\nu \in \mathbb{R}}$.

Malheureusement, ceci convient pour des signaux stationnaires dont les propriétés statistiques sont invariantes dans le temps. Cependant, en pratique la grande majorité des signaux rencontrés sont non-stationnaires et la notion de décomposition sur une somme infinie d'exponentielles complexes propres à la TF s'avère inadéquate. Pour mieux décrire ce type de signaux, on fait appel à des représentations faisant intervenir conjointement le temps et la fréquence. Le changement ainsi opéré nous permettra d'analyser les lois du comportement fréquentielle du signal au cours du temps [10]. Dans le cas de l'existence de sauts, d'impulsions ou de changements de fréquence par exemple, la TF ne permet pas de localiser ces événements transitoires qui surviennent dans le signal. Ces phénomènes sont non-stationnaires et du fait qu'ils sont brefs et souvent à caractère oscillatoire, leur contenu spectral est difficile à mettre en évidence. Dans ce cas c'est le changement instantané de la

fréquence tout au long du temps qui nous intéresse et pas nécessairement la fréquence elle-même. Il faut donc mettre à profit les caractéristiques et les propriétés statistiques du signal afin de faire apparaître les événements le constituant.

Suite aux travaux fondamentaux de Gabor [95], Ville [266] et de Wigner [271] sur les représentations temps-fréquence (RTF), le sujet est quasiment devenu une discipline scientifique à part entière. Les RTF sont utilisées dans de nombreux domaines d'applications [31, 32, 230, 53, 88, 82, 140], pour la détection [11, 228] et pour la classification [64]. Plus récemment, la transformée en ondelettes a gagné en popularité et est devenue l'analyse la plus utilisée surtout dans le domaine industriel. L'objectif commun de toutes ces représentations est d'offrir une description détaillée et pertinente des signatures temps-fréquence composant le signal. Parmi les méthodes susceptibles de répondre à nos attentes, seules les plus utilisées vont retenir notre attention.

4.1.2 La transformée de Fourier à fenêtre glissante

Les atomes temps-fréquence

Contrairement à la TF qui évalue le contenu fréquentiel d'un signal $x(t)$ en l'étudiant sur tout son domaine temporel, la transformée de Fourier à fenêtre glissante (TFFG) estime le contenu fréquentiel du signal $x(t)$ dans une fenêtre localisée dans le temps [88, 61]. Elle peut s'interpréter comme la projection du signal analysé sur des atomes temps-fréquences de Gabor. Ces atomes sont construits par une translation en temps d'une fenêtre unique $h(t)$ et une modulation

$$h_{t,\nu}(s) = h(s - t)e^{+i2\pi\nu s}. \quad (4.5)$$

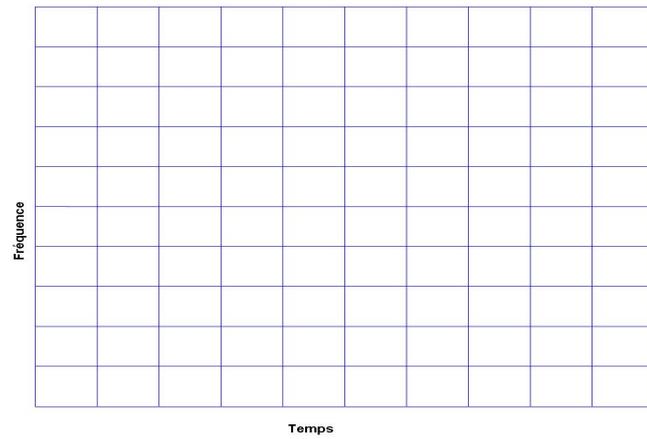
L'énergie de $h_{t,\nu}$ est concentrée au voisinage de t sur un intervalle de longueur T_h . Sa transformée de Fourier s'obtient en translatant de ν la transformée de Fourier \hat{h} de h

$$\hat{h}_{t,\nu}(w) = \hat{h}(w - \nu)e^{-i2\pi t(w - \nu)}. \quad (4.6)$$

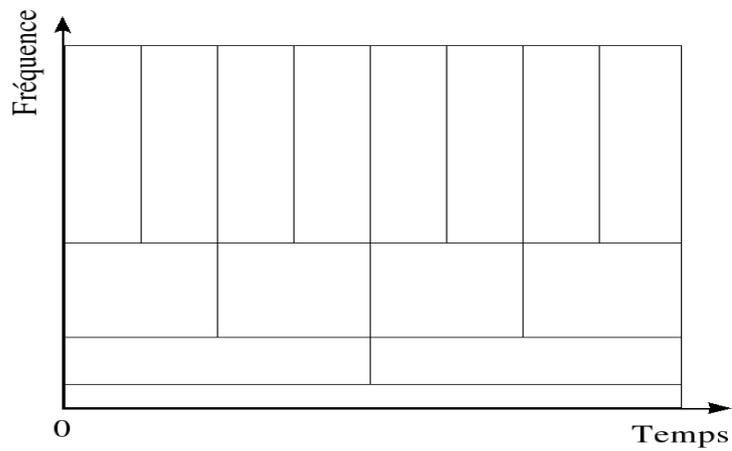
L'énergie de $\hat{h}_{t,\nu}$ est localisée au voisinage de la fréquence ν sur un intervalle de largeur B_h .

Dans le plan temps-fréquence (s, w) , l'étalement de l'énergie de l'atome est représenté par un rectangle d'Heisenberg. Ce rectangle a pour centre (t, ν) avec une largeur temporelle T_h et une largeur fréquentielle de B_h , figure 4.1(a).

Le principe de la TFFG consiste à effectuer une TF du signal $x(t)$ multipliée au préalable par la fenêtre $h(s - t)$. Chaque fenêtre h est lissée, bien localisée et de courte durée T_h pendant laquelle $x(t)$ est considéré comme stationnaire. h est en plus réelle, symétrique : $h(t) = h(-t)$ et normalisée à $\|h\| = 1$, afin d'avoir $\|h_{t,\nu}\| = 1$ pour tout $(t, \nu) \in \mathbb{R}^2$.



(a)



(b)

FIG. 4.1 – (a) Pavage de la transformée de Fourier à fenêtre glissante (TFFG). (b) Pavage de la Transformée en ondelettes (TO).

L'application de la TFFG sur le signal $x(t)$ fournit une représentation à deux dimensions dépendant du temps et de la fréquence [95]. Sa formule est

$$F_x(t, \nu) = \langle x, h_{t, \nu} \rangle = \int_{-\infty}^{+\infty} x(s)h(s-t)e^{-i2\pi\nu s} ds, \quad (4.7)$$

où la fonction qui associe $h(s-t)$ à t est la fenêtre qui sélectionne la durée T_h autour de l'instant t . Cette intégrale de Fourier peut être exprimée en une intégrale en fréquence grâce à l'identité de Parseval

$$F_x(t, \nu) = e^{-i2\pi\nu t} \int_{-\infty}^{+\infty} X(\xi)H^*(\xi - \nu)e^{i2\pi\xi t} d\xi. \quad (4.8)$$

La transformée de Fourier $H(\nu)$ de $h(t)$ joue le rôle d'une fenêtre spectrale qui sélectionne la bande de fréquence B_h autour de la fréquence ν appliquée au spectre $X(\nu)$ du signal $x(t)$. La transformée $F_x(t, \nu)$ ne dépend que des valeurs $x(t)$ et $X(\nu)$ dans les voisinages temporels et fréquentiels où sont concentrées les énergies de $h_{t, \nu}$ et $H_{t, \nu}$.

La densité d'énergie E_x , notée $|F_x(t, \nu)|^2$, est appelée spectrogramme. Elle mesure l'énergie de $x(t)$ dans le voisinage temps-fréquence de (t, ν) défini par la boîte d'Heisenberg de $h_{t, \nu}$ centrée en (t, ν) et de dimension (T_h, B_h) . La taille de cette boîte ne dépend pas de (t, ν) , cela veut dire que les dimensions de la fenêtre d'analyse sont constantes dans tout le plan temps-fréquence, figure 4.1(a).

Le passage d'un axe unique de temps (resp. de fréquence) à un plan temps-fréquence fournit une meilleure structuration de l'information sur le signal, certes au prix d'une augmentation de redondance, mais en offrant une possibilité de lecture différente et complémentaire du signal analysé [88]. Ainsi, une lecture intelligente de l'image du plan temps-fréquence (spectrogramme), nous permet de bien comprendre le comportement du signal et une interprétation directe sur le nombre de composantes et sur leur contenu temps-fréquence.

La résolution temps-fréquence dans la TFFG

Bien que délivrant une information temporelle et fréquentielle à la fois, la TFFG s'avère toutefois insuffisante pour certaines applications. En fait, la rigidité de l'enveloppe de la fenêtre d'analyse utilisée limite l'adaptabilité de l'analyse par la TFFG et cela se traduit par des problèmes de résolution.

La résolution en temps et en fréquence de la TFFG dépend de l'étalement de la fenêtre en temps et en fréquence. Cela peut se mesurer par la surface $T_h B_h$ de la boîte d'Heisenberg. Le théorème d'incertitude d'Heisenberg montre que la surface de ce rectangle ne peut être petite : $T_h B_h \geq \frac{1}{4\pi}$. Cette inégalité nous informe que la cellule du plan temps-fréquence a une surface minimale et que cette borne minimale de $\frac{1}{4\pi}$ est atteinte par $h(t)$ de forme gaussienne. Lorsque ce type de fenêtre est choisi, on parlera de l'analyse de Gabor. Compte tenu de ce principe d'incertitude, il est impossible donc d'avoir à la fois une localisation parfaite en temps et en fréquence c'est-à-dire un pavé réduit à un point. Il subsiste donc un problème de résolution

temporelle et fréquentielle évoluant en inverse l'une de l'autre. La conséquence d'une mauvaise résolution temporelle (resp. fréquentielle) est l'étirement du motif dans la dimension temporelle (resp. fréquentielle). Le principe d'Heisenberg impose donc un choix dans la sélection de la résolution de la fenêtre.

4.1.3 Distribution de Wigner-Ville

La densité d'énergie temps-fréquence

Par opposition aux représentations atomiques, telle que la TFFG, qui procédaient à une décomposition linéaire du signal en composantes élémentaires ; Ville, ingénieur en traitement du signal, a proposé d'étudier les propriétés du signal $x(t)$ à partir de sa densité d'énergie temps-fréquence [266]. Il proposait de prendre la quantité $P_x(t, \nu)$, dépendant du temps et de la fréquence, et satisfaisant les contraintes de distributions marginales [88]. Cette distribution est interprétée comme étant une densité de probabilité qui peut prendre, à l'inverse de toute densité de probabilité, des valeurs négatives.

Contrairement à la TFFG qui se calcule en corrélant le signal $x(t)$ avec des familles d'atomes temps-fréquence, la distribution de Wigner-Ville (DWV) se calcule en le corrélant avec une translation en temps et en fréquence de lui même. La fenêtre d'analyse n'est rien d'autre que le signal lui même inversé. Aucun choix de type de fenêtre d'analyse ni de sa largeur n'est imposé, le temps et la fréquence jouent des rôles symétriques. La formule de la DWV est

$$W_x(t, \nu) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi\nu\tau} d\tau. \quad (4.9)$$

La distribution de Wigner-Ville est réelle et représente la transformée de Fourier de $x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})$ qui possède une symétrie hermitienne en t .

D'une manière générale, l'utilisation d'une classe de Cohen entraîne l'apparition de termes d'interférences externes en présence de signaux multi-composantes et de termes d'interférences interne en présence d'une composante dont la fréquence instantanée est non linéaire en fonction du temps [230]. Faisant partie de la distribution, ces termes interférentiels sont porteurs d'une information sur le signal analysé en particulier de phase [88], et peuvent s'avérer utiles pour la détection des signaux ou la classification des signaux. Une bonne connaissance du signal analysé et de la nature de l'oscillation de ces composantes interférentielles permet, plus au moins, de les distinguer sur le plan temps-fréquence [115]. En revanche leur importance en nombre et en amplitude perturbe l'analyse des signaux à composantes multiples ou à modulation de fréquence non-linéaire.

Plusieurs travaux ont été consacrés aux modifications à apporter à la définition initiale pour améliorer la lisibilité de la DWV sans trop perdre en propriétés théoriques. L'objectif étant d'introduire une opération de lissage $\Pi(t, \nu)$ dans le plan temps-fréquence, qui appliquée à la DWV conduit à l'évaluation de l'expression suivante : $\Pi(t, \nu) * W_x(t, \nu)$. Selon un choix particulier de Π , différentes distributions

d'énergies ont été proposées, on cite la distribution de Page en 1952 et celle de A.W. Rihadaczek en 1968, la réallocation [88, 45], la pseudo Wigner-Ville lissée (PWVL) introduite par Hlawatsch et Boudreaux-Bartels [115]. L'objectif étant de trouver un lissage $\Pi(t, \lambda)$ permettant la suppression des termes d'interférences.

4.1.4 La transformée en ondelettes continue

L'adaptabilité en temps et en fréquence

Si un signal $x(t)$ est la somme de plusieurs composantes dont les dérivées de fréquence instantanée sont très différentes, il est impossible de visualiser de façon concentrée toutes les composantes avec la TFFG. Dans ce cas il serait nécessaire que la taille de la fenêtre soit variable suivant la position dans le plan temps-fréquence, et c'est justement le principe de la transformée en ondelettes.

Les composantes d'un même signal n'évoluent pas forcément de la même manière, leur comportement varie selon qu'elles soient en hautes ou en basses fréquences. Naturellement, une composante de type basse fréquence nécessite par essence un temps d'observation suffisamment long pour pouvoir être correctement analysé. A l'inverse, une composante de type haute fréquence évolue rapidement et nécessite un temps d'observation plus réduit pour être analysé. C'est justement l'adaptabilité de l'analyse de la transformée en ondelettes qui permettra d'analyser différemment ces composantes.

Il existe un grand nombre d'ondelettes offrant une grande diversité d'analyse. Il existe également diverses transformées continues ou discrètes. Ainsi, la transformée en ondelettes permet une représentation temps-échelle ou temps-fréquence du signal dans un même plan. Les méthodes d'analyse temps-échelle sont très utilisées pour la détection de ruptures dans les signaux [44, 225].

Atomes temps-fréquence affines

C'est comme une alternative à la transformée de Fourier fenêtrée que la transformée en ondelettes continue [128, 104] a été développée afin de pallier les difficultés de résolution temps-fréquence. J. Morlet en 1983 [186], pour l'étude haute résolution des signaux sismiques, propose une transformée où la taille de la fenêtre est variable, ceci grâce à un paramètre d'échelle. Cette transformée en ondelettes est une décomposition atomique dont les atomes sont issus d'une même fonction, l'ondelette mère, par opérations de translation et dilatation. Cette ondelette, appelée ainsi car représentant des oscillations, comme une onde, mais limitée en temps d'où le nom de la petite onde. Elle doit être bien localisée en temps et en fréquence.

On conduit l'analyse en ondelettes de la même manière que l'analyse par la TFFG. Le signal est multiplié par une fonction semblable à la fonction de fenêtrage de la TFFG. La fenêtre modulée $h(t)e^{2i\pi\nu t}$ de la TFFG a été remplacée par

une ondelette $\frac{1}{\sqrt{a}}\psi\left(\frac{s-t}{a}\right)$ [170, 88]. La translation est liée à la localisation de la fenêtre, à mesure que cette fenêtre est décalée sur l'étendue du signal. L'échelle peut s'interpréter comme l'inverse d'une fréquence $1/\nu$.

De même que la TFFG, la transformée en ondelette permet de mesurer les variations temporelles des composantes spectrales de $x(t)$. Cependant, il existe une grande différence entre les deux transformées. Pour analyser des composantes transitoires de durées différentes, la transformée en ondelettes décompose les signaux sur une famille d'ondelettes translatées et dilatées. La largeur de la fenêtre est donc modifiée pour chacune des composantes spectrales individuellement, ce qui se traduit par une résolution temps-fréquence différente de celle de la TFFG. Ce dernier point est la caractéristique la plus significative de la transformée en ondelettes.

Une famille d'atomes temps-échelle s'obtient à partir d'une ondelette ψ unique, en la dilatant d'un facteur d'échelle a et en la translatant de t

$$\psi_{t,a}(s) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{s-t}{a}\right). \quad (4.10)$$

La transformée en ondelettes de $x \in L^2(\mathbb{R})$ au temps t et à l'échelle a se calcule en projetant $x(t)$ sur la famille d'ondelettes $\psi_{t,a}(s); t \in \mathbb{R}$ et $a \in \mathbb{R}^+$. La représentation temps-échelle obtenue est appelée transformée en ondelettes continue (TOC), sa définition est

$$T_x(t, a) = \langle x, \psi_{t,a} \rangle = \int_{-\infty}^{+\infty} x(s) \frac{1}{\sqrt{|a|}} \psi^*\left(\frac{s-t}{a}\right) ds. \quad (4.11)$$

Sa version fréquentielle est donnée par :

$$T_x(t, a) = |a|^{1/2} \int_{-\infty}^{+\infty} \chi(\nu) \psi^*(a\nu) e^{2i\pi\nu t} d\nu. \quad (4.12)$$

La fonction $t \mapsto \psi^*\left(\frac{s-t}{a}\right)$, dans l'équation (4.11), sélectionne la durée aT_ψ autour de t , T_ψ étant la durée de l'ondelette ψ . Pour une ondelette $\psi(\nu)$ de fréquence centrale ν_0 , la fonction $\nu \mapsto \psi^*(a\nu)$, dans l'équation (4.12), sélectionne par contre la bande $\frac{B_\psi}{a}$ autour de $\frac{\nu_0}{a}$, B_ψ étant la largeur de la bande de l'ondelette ψ . Le module de $T_x(t, a)$ mesure alors l'amplitude du signal dans la fenêtre centrée sur $(t, \nu) = (t, \frac{\nu_0}{a})$ et de dimension $(aT_\psi, \frac{B_\psi}{a})$ dans le plan temps-fréquence. L'image $|T_x(t, a)|^2$ est appelée le scalogramme.

La transformée est donc une fonction de deux paramètres, la translation temporelle t et l'échelle a . $\psi(t)$ est l'ondelette mère, et doit être localisée aussi bien en temps qu'en fréquence. Pour cela elle doit satisfaire certaines conditions d'admissibilité :

– $\psi(t)$ doit être absolue et de carré intégrable

$$\int |H(\nu)|^2 \frac{d\nu}{|\nu|} = 1. \quad (4.13)$$

Cette première propriété impose que l'ondelette $\psi(t)$ ait un spectre, $H(\nu)$, dont la décroissance soit au moins en $|\nu|^2$,

- elle doit être de moyenne nulle, autrement dit, le spectre de $\psi(t)$ s'annule à l'origine : $\int \psi(t)dt = H(0) = 0$.

Ces conditions garantissent que l'ondelette mère a une énergie finie. D'autres conditions supplémentaires telles que la régularité, la décroissance rapide ou encore la compacité peuvent être imposées suivant les besoins de l'application [177].

Résolution temps-échelle en ondelette

Contrairement au comportement rigide de la TFFG la transformée en ondelettes adopte un comportement dynamique, lié au zoom adaptatif de l'échelle a , pour suivre les singularités et les irrégulières du signal analysé. Avec une ondelette ψ réelle, les variations brutales du signal $x(t)$ produisent de grands coefficients d'ondelettes et la singularité locale est liée à la décroissance de $|T_x(t, a)|$ lorsque a tend vers 0.

Du fait de son extinction à l'origine des fréquences et sa décroissance à l'infini, le coefficient $T_x(t, a)$ peut être interprété, pour chaque échelle a , comme le résultat d'un filtrage passe-bande du signal $x(t)$. Ainsi, la transformée en ondelettes est considérée comme un banc continu de filtres à surtension constante, connu sous le nom du Q-constant.

Dans l'équation (4.12), on remarque qu'une valeur $a \neq 1$ du paramètre d'échelle définit un filtre dans la bande $[(\nu_0 - B_\psi/2)/a, (\nu_0 + B_\psi/2)/a]$. Il en résulte un gabarit modifié dans sa fréquence centrale et sa largeur de bande, mais inchangé dans son facteur de surtension Q défini comme l'inverse de sa bande relative [88]

$$Q_a = \frac{\Delta\nu}{\nu} = \frac{B_\psi/a}{\nu_0/a} = \frac{B_\psi}{\nu_0}. \quad (4.14)$$

De ce fait, la transformée en ondelettes présente une résolution absolue qui dépend du point d'évaluation et varie en fonction de la fréquence. Par contre, sa résolution fréquentielle relative $\frac{\Delta\nu}{\nu}$ est constante.

La notion de représentation temps-échelle ou espace-échelle permet de voir la transformée en ondelettes comme un microscope mathématique, le zoom étant l'échelle et l'optique, l'ondelette. Le paramètre échelle est très similaire à la notion d'échelle pour les cartes géographiques. En tant qu'opération mathématique, les grandes échelles, ou basses fréquences, dilatent le signal et correspondent à des vues plus globales non détaillées du signal. Les petites échelles, ou hautes fréquences, correspondent à des vues détaillées et donnent des informations détaillées sur un motif caché dans le signal. Généralement ces composantes de faibles échelles n'existent que sur de courtes durées du signal, elle apparaissent de temps en temps associées à des pics transitoires. A l'inverse, les grandes échelles existent typiquement sur toute la durée du signal.

4.2 Choix de la représentation temps-fréquence

Nous avons présenté trois types de transformations temps-fréquence, la TFFG, la WV et la TO. Le choix est en fait beaucoup plus vaste. Meyer a écrit à ce sujet : "*En exagérant à peine, nous introduirons presque autant d'algorithmes d'analyse qu'il y a de signaux* [178]". Le choix de la transformation, comme s'accordent à dire les spécialistes temps-fréquence [115, 88] et les physiciens [4], dépend de l'application et du signal étudié. Le problème revient donc à spécifier les critères sur lesquels on peut se baser pour choisir une représentation en adéquation avec la problématique à résoudre. L'ensemble des propriétés est synthétisé dans [88, 82, 178], de manière générale une représentation pertinente doit être : locale, adaptative, lisible et dans certain cas orthogonale.

L'objectif de notre application est de pouvoir repérer l'apparition de défauts qui vont du millimètre, défaut local qui touche la surface du bandage, à la dizaine de centimètres, défaut structurel et qui modifie la structure même du galet. Un défaut local peut se manifester en hautes fréquences et nécessite un temps court pour être détecté. Par contre, un défaut structurel engendre des vibrations en basses fréquences et nécessite un temps suffisamment long pour être détecté. Ceci nécessite donc un tour de galet, voire plus, pour décider que le défaut appartient bien à ce dernier type de défaut. Pour avoir une bonne résolution en basses et en hautes fréquences l'analyse temps-échelle est la plus adaptée à notre problème puisque la précision relative est constante sur toute la bande de fréquences analysées ($\Delta\nu/\nu$ est constant).

Pour nos traitements, nous préférons utiliser la transformée en ondelettes continue (TOC) plutôt que la TFFG ou une distributions d'énergie de la classe de Cohen. Sa linéarité permet de ne pas avoir de termes d'interférences croisés, tout en obtenant des motifs assez concentrés en hautes et en basses fréquences grâce à la résolution temps-fréquence adaptée. Concernant le choix de l'ondelette ψ , notre choix s'est porté sur l'ondelette de type chapeau mexicain car elle est bien adaptée à la détection de régimes transitoires et approche la borne d'Heisenberg-Gabor ce qui fait d'elle une fenêtre de bonne résolution temps-fréquence. Du fait qu'elle soit symétrique en une dimension, elle permet de ne pas avoir de déphasage dans la transformée en ondelettes. L'ondelette en chapeau mexicain normalisée est donnée par l'équation suivante [170]

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}} \left(\frac{t^2}{\sigma^2} - 1 \right) \exp \left(\frac{-t^2}{2\sigma^2} \right). \quad (4.15)$$

Le scalogramme, figure 4.2, représente une image en niveau de gris de la densité d'énergie de la transformée temps-échelle, $|T_x(a, t)|^2$, d'un signal issu des essais SYSCOGA-III sur la voie. Les axes verticaux et horizontaux représentent respectivement $\log(a)$ et t . On remarque bien que la transformée en ondelettes continue, de type chapeau mexicain, choisie est complète, préserve l'énergie et est bien adaptée à la détection de transitions brutales. Appliquée à nos signaux, elle fait apparaître les zones énergétiques sous forme de pics facilement identifiables sur le scalogramme. Ces pics transitoires peuvent être associés à l'existence d'usure.

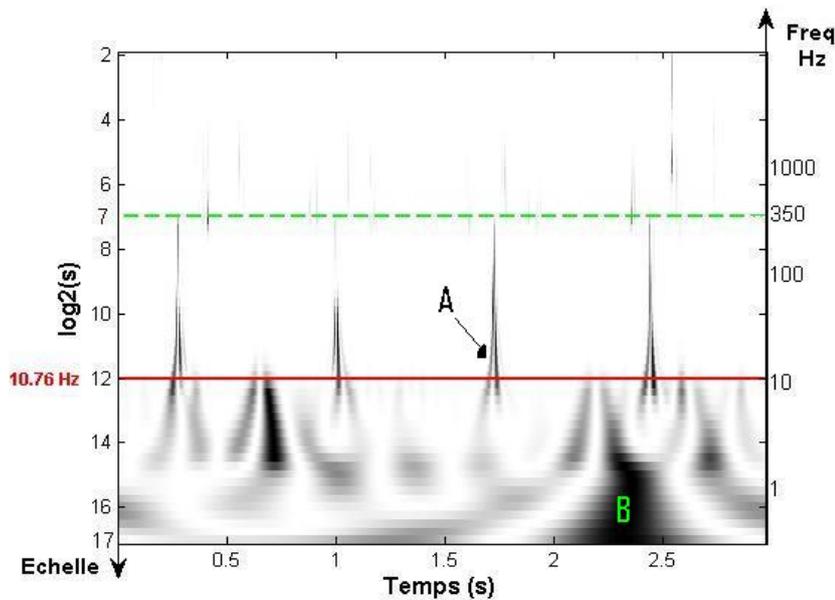


FIG. 4.2 – La représentation temps-échelle du signal (scalogramme). L'axe vertical de droite représente l'axe des fréquences.

4.3 Étude des scalogrammes

4.3.1 Description qualitative des scalogrammes

La transformée continue en ondelettes, figure 4.2, fait apparaître des motifs énergétiques assez concentrés en hautes et en basses fréquences. Quelques informations sont directement évidentes. Notons F_{rot} la fréquence de rotation des galets calculée à partir de la vitesse du tramway et de la circonférence du galet. Elle est représentée sur le scalogramme, figure 4.2, par un trait horizontal continu au niveau de l'échelle $a = 12$ correspondant approximativement à la fréquence $F_{rot} \approx 10.76$ Hz. Aux fréquences $\nu > F_{bruit}$, avec $F_{bruit} \simeq 350$ Hz correspondant à l'échelle $a \simeq 7$ représentée sur le scalogramme par un trait pointillé, les vibrations s'atténuent rapidement et l'analyse révèle que cette bande ne contient que très peu d'information. Les fréquences associées correspondraient à des défauts d'une taille inférieure au centimètre. Le scalogramme est ainsi clairement divisé en trois parties.

Deux types principaux de défauts peuvent être distingués. Le premier type correspond à des vibrations aux fréquences intermédiaires $F_{rot} \leq \nu \leq F_{bruit}$ (petites échelles, de ~ 1 cm à ~ 10 cm) et caractérise les défauts surfaciques locaux. La figure 4.2 représente un passage du tramway où :

- les essieux arrières ME1, MI1, MI2 et ME2 portent un couple de galets avec un défaut local, respectivement, de type : trous, usure 50%, usure 100% et enfin usure 100% avec des trous ;
- les essieux avant ME1 et MI2 portent des galets sans défauts ;
- enfin, les essieux MI1 et ME2 avant portent respectivement des galets avec une usure patatoïde et avec un vis fixé sur le bandage.

Les pics, figure 4.2(zone A), se trouve exactement aux mêmes endroits et peuvent

donc être associés aux défauts locaux cités ci-dessus.

Le second type correspond aux défauts structuraux des galets qui produisent des vibrations à basses fréquences $\nu < F_{rot}$ (grandes échelles, $\gtrsim 10$ cm). Ces structures basses fréquences, figure 4.2(zone B), additionnent les effets des défauts de déformation des galets et des effets d'instabilité dynamique du chariot du guidage. Dans ce travail nous ne traitons que les défauts locaux associés aux fréquences $F_{rot} \leq \nu \leq F_{bruit}$, c'est-à-dire dans la bande de fréquence $[10.76 - 350]$ Hz. L'exploitation de la bande des basses fréquence est plus délicate et nécessiterait d'avantage d'informations sur le comportement du système de guidage. Cependant, ces données ne sont pas tout à fait acquise ce qui rend l'étude de la bande des basses fréquences pour l'instant difficile. Les deux bandes $\nu > F_{bruit}$ et $\nu < F_{rot}$ ne seront pas prises en compte dans la suite des traitements.

4.3.2 Découpage temps-échelle des scalogrammes

Le découpage temporel, présenté dans la partie 3.4.2, appliqué aux scalogrammes fournit une matrice de coefficients d'ondelettes CWT_k pour chaque couple de galets CGk . Cette même procédure, illustrée dans la figure 4.3, est appliquée sur les signaux issus de l'installation sur l'ADV :

- la figure 4.3(a) représente le signal initial acquis sur l'ADV ;
- l'extraction du signal utile allant de la détection de l'entrée du premier couple de galets $CG1$ jusqu'à la sortie du dernier couple $CG8$, figure 4.3(b) ;
- l'application de la transformée en ondelettes continue sur le nouveau signal ainsi que le découpage temporel, figure 4.3(c).

Exemple 4.3.1

La figure 4.4 représente un extrait du signal correspondant au deuxième essieu intermédiaire (MI2). Sur l'avant de l'essieu, $CG3$, est monté un galet sans défaut et un galet avec défaut, 100% d'usure. L'arrière de l'essieu, $CG4$, comporte un galet sans défaut et un galet usé à 75%. Au niveau de la bande intermédiaire, $[10.76 - 350]$ Hz ($[7 \leq a \leq 12]$), nous remarquons l'existence de perturbations (pics). Dans la zone correspondante au couple de galets défectueux à 100% elles sont beaucoup plus marquées que celles propre au couple de galets usé à 75%.

À une vitesse moyenne $\simeq 30$ km/h et $Fe = 44100$ Hz, à chaque couple CGk ($k = 1, \dots, 8$) est associé un secteur rectangulaire du scalogramme correspondant à une matrice de coefficients d'ondelette CWT_k , figure 4.3(c), de taille (10584×64) pour un essieu extrême et (11466×64) pour un essieu intermédiaire. Ces matrices sont utilisées par la suite en entrée du module de diagnostic. L'identification de l'état d'un couple de galets (sans défaut, ou avec défaut) se résume en une tâche de classification (chapitre 6).

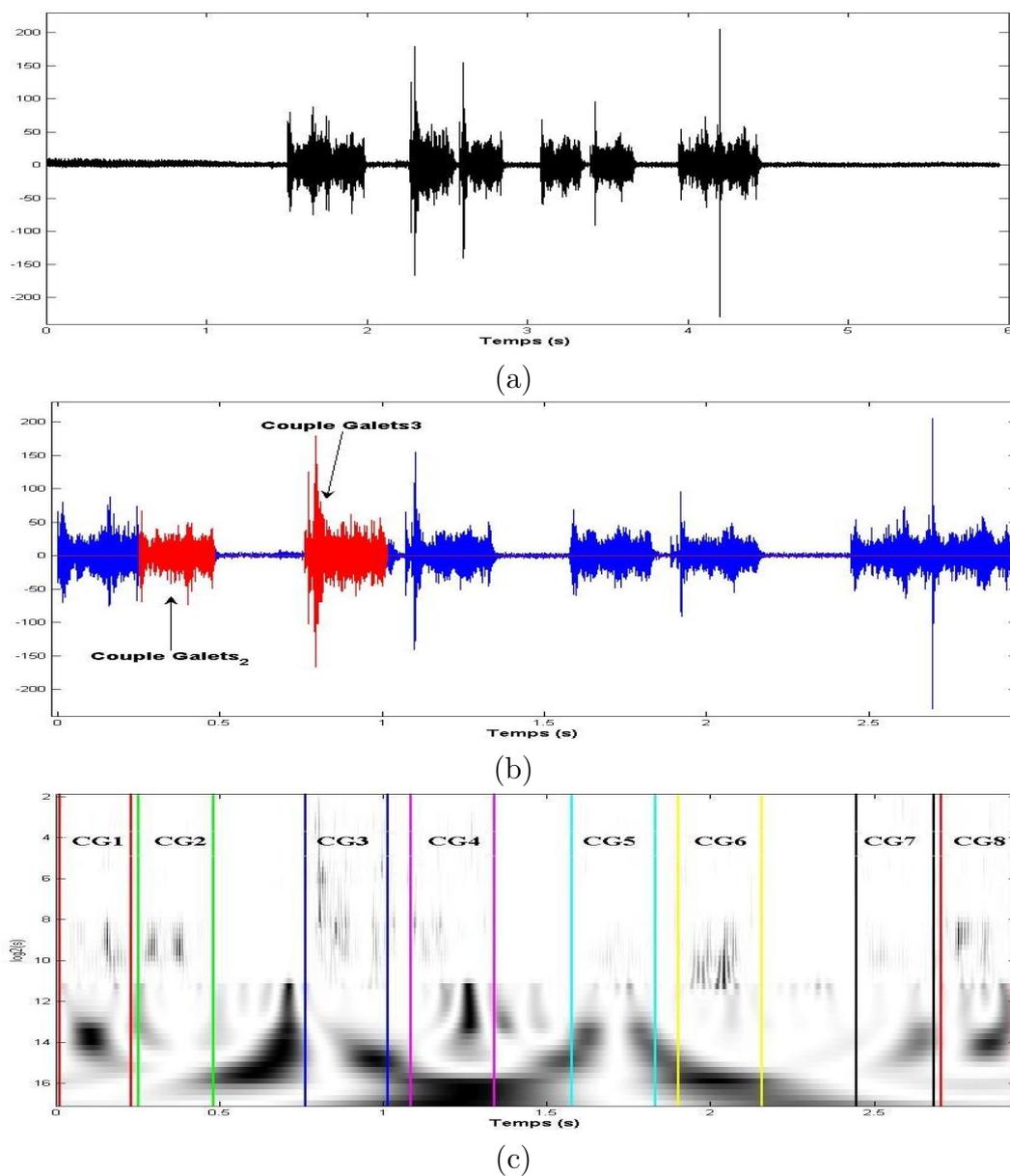


FIG. 4.3 – (a) Le signal temporel complet. (b) Extraction du signal utile, de l'entrée du premier couple de galets CG1 jusqu'à la sortie du dernier couple. Le couple CG2 représente des galets sans défaut (neufs) et le couple CG3 représente une usure à 100%. (c) Découpage temporel pour la localisation des couples de galets. Les 8 zones représentent les couples de galets CG_k , $k = 1, \dots, 8$.

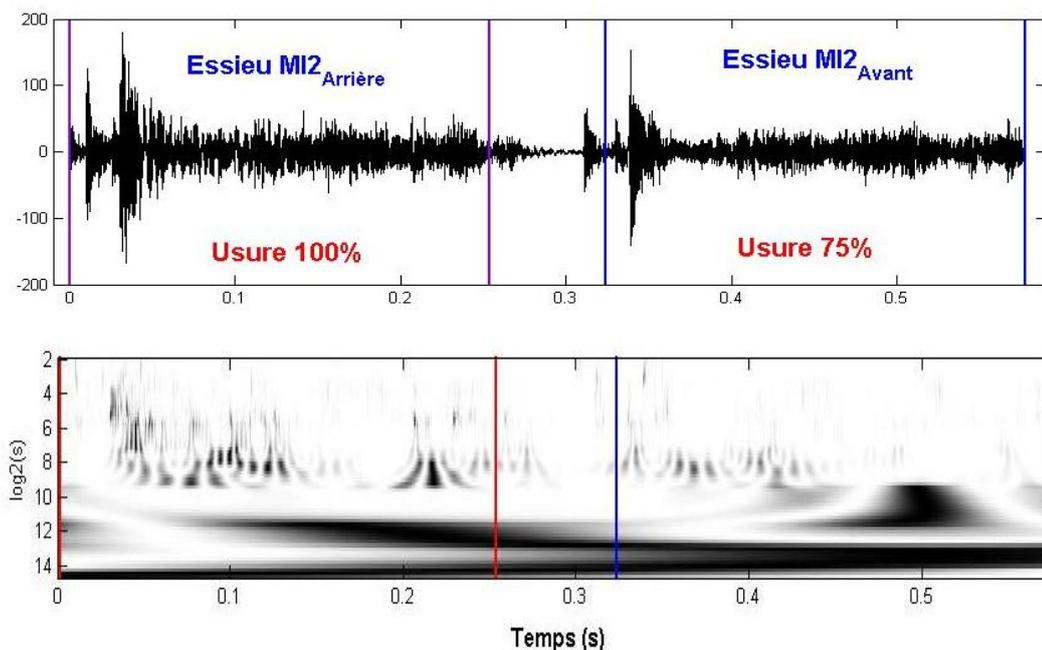


FIG. 4.4 – Extrait du signal correspondant à l’essieu MI2 et sa transformée en ondelette continue. Le premier couple de galet (MI2 arrière) représente un galet utilisé à 100%, par contre le deuxième couple (MI2 avant) est utilisé à 75%.

4.4 Vers la classification

4.4.1 La dimension de l’espace de représentation d’un couple de galets

Dans un problème de reconnaissance de formes, Il s’agit de construire les vecteurs formes $x_k = (x_{k1}, x_{k,2}, \dots, x_{k,p})$, $k \in [1, n]$. En statistique un tel tableau est considéré comme la réalisation d’un échantillon de taille n d’un vecteur aléatoire x_k de dimension p . Chaque vecteur correspond à un individu dont on connaît les p variables descriptives. À la fin de l’étape de génération de données, paragraphe 4.3.2, chaque couple de galet CGk est représenté par une matrice de données $CWTk_{(i,j)}$ de dimension (10584×64) , pour un essieu extrême, et (11466×64) pour un essieu intermédiaire. Donc un couple de galets est représenté par un vecteur forme $x_k = (x_1, x_2, \dots, x_p)$, $p \in [677376, 7733824]$. La dimension p des données disponibles à la fin de cette étape de génération est très importante. En plus ces p variables ne sont pas toutes porteuses d’informations et peuvent correspondre à du bruit. Certaines peuvent aussi être corrélées ou non pertinentes pour la tâche de classification. Il est donc nécessaire de réduire cette dimension afin d’obtenir de meilleures performances en classification.

Le besoin de réduire la dimension de l’espace, dans lequel une forme est décrite, est tout d’abord une nécessité physique. Des problèmes de stockage, de transmission, de représentation et même de perception se posent en dimension élevée. Malheureusement, l’augmentation du nombre de variables n’améliore pas systématiquement la qualité de l’apprentissage, au contraire elle fait croître le nombre de paramètres à

estimer.

En effet, on est confronté au phénomène appelé "the curse of dimensionality" [22]. Ce problème provient du fait qu'un concept avec une quantité N de données en dimension p , nécessaire pour apprendre, croît exponentiellement en fonction de p [2]. La relation exacte entre la probabilité d'erreur de classification, le nombre de variables p , la taille n_i de l'ensemble d'apprentissage et la complexité de la règle de décision est très difficile à établir [224]. Dans le cas général, il est conseillé de disposer par classe d'un nombre n_i d'exemples d'apprentissage dix fois plus grand que le nombre de variables ($n_i/p > 10$) pour obtenir un discriminateur avec de bonnes performances [135]. Plus le modèle est complexe, plus le rapport n_i/p doit être grand, pour éviter le problème de la malédiction de dimensionnalité [223].

4.4.2 Interprétation des représentations temps-fréquence

Il existe plusieurs travaux traitant le problème d'extraction de l'information pertinente à partir du plan temps-fréquence [8, 158, 229]. Certains auteurs se sont focalisés sur l'interprétation des RTFs de la classe de Cohen [51, 214]. D'autres sur l'extraction d'un ensemble de mesures statistiques de signaux par apprentissage pour la détection d'anomalies sur les moteurs électriques [276].

Dans [15, 216], les auteurs considèrent la RTF comme étant une distribution de probabilité. L'entropie de Rényi d'une distribution de Wigner-Ville est utilisée comme un indicateur sur la complexité d'un signal. Une autre idée est exploitée dans [215] et consiste à utiliser la théorie des graphes des représentations minimales pour l'extraction du squelette de la RTF. Alors que Barre a présenté dans [214] une autre approche basée sur l'extraction de signatures à partir du calcul de distances et des moments statistiques.

D'autres auteurs [117, 118] considèrent la RTF comme un ensemble de réalisations de variables aléatoires, en partant de l'hypothèse que le signal analysé contient une partie aléatoire. Les auteurs considèrent que le motif spectral est un ensemble de réalisations de variables aléatoires décrites par les paramètres de sa loi. L'estimation de ces paramètres associe à la localisation de la composante temps-fréquence une caractérisation de son comportement énergétique.

Une autre approche consiste, de manière la plus naturelle, à considérer la RTF comme étant l'image du signal acquis, ainsi, plusieurs techniques appliquées sur les images pour la reconnaissance de formes, peuvent être exploitées pour l'interprétation temps-fréquence :

- l'utilisation des modèles déformables pour le diagnostic de sténoses [129],
- la morphologie mathématique pour la détection des motifs dans les RTFs [204, 159, 42], etc. On trouve une synthèse de ces méthodes dans le livre de Martin et Doncarli [81].

Dernièrement, la décomposition en valeurs singulières (DVS) de la matrice du plan temps-fréquence du signal pour une caractérisation globale de la RTF a fait

l'objet de plusieurs travaux [105, 106, 126, 108, 263, 264]. Certains auteurs se sont concentrés sur l'extraction de caractéristiques en se basant seulement sur les valeurs singulières des distributions temps-fréquence des signaux [126]. Marinovic et Eichmann [173], ont utilisé la décomposition en valeurs singulières de la distribution de Wigner en se basant seulement sur les valeurs singulières pour déterminer les informations pertinentes pour la classification. Alors que d'autres estiment que dans certain cas, les valeurs singulières apportent peu d'informations et préfèrent utiliser les vecteurs singuliers pour la caractérisation des signatures à partir d'une distribution temps-fréquence [106, 105, 108, 263, 264].

Selon la façon de voir les matrices CWT_k correspondant aux couples de galets CG_k , il existe plusieurs techniques de réduction de dimension pour la classification. Nous y reviendrons dans le chapitre 5. La première idée qui nous est venue à l'esprit, exploitée dans le paragraphe suivant 4.5, est de considérer les scalogrammes comme étant des images en niveau de gris. La tâche de diagnostic reviendrait alors à segmenter ces images et extraire les zones d'intérêts pouvant correspondre à l'existence de défauts sur un couple de galets. Seules les zones vérifiant certaines conditions sont retenues pour l'étape de la classification. Les points forts la méthode de segmentation d'image par binarisation, testée sur les scalogrammes, et ses limites sont rapidement illustrées à travers quelques exemples expérimentaux.

4.5 Tentative de segmentation des images

4.5.1 Segmentation des scalogrammes par binarisation

En traitement d'images, il est souvent impraticable de manipuler les points initiaux de l'image à cause de sa taille. d'autre part, l'information véhiculée par un pixel est souvent peu précise et redondante. Il est alors préférable d'extraire de l'image des primitives apparaissant peu souvent dans les images et contenant beaucoup d'informations (les segments de contour, les régions). Ces primitives sont ensuite décrites par un ensemble de caractéristiques permettant de les discriminer (la longueur du contour, la surface de la région). Cette phase d'extraction des primitives est appelée segmentation. On trouve les caractéristiques traditionnellement extraites, dans les domaines de l'analyse d'images et du signal sonore, recensés dans [130, 258, 253]. Cette phase de segmentation et extraction des caractéristiques est souvent précédée d'une phase de pré-traitement pour améliorer la qualité des images et réduire le bruit généré lors de l'étape d'acquisition (amélioration du contraste, filtrage de bas niveau, égalisation de l'histogramme,...).

Pour la segmentation des images nous avons utilisé la binarisation d'images à niveaux de gris par la méthode d'OTSU [193], méthode rapide et simple à mettre en oeuvre. Elle se base sur l'histogramme de l'image en optimisant un seuil tout en minimisant la variance intra classe. Cette méthode sépare les pixels de l'images en deux classes (noir et blanc) en ne gardant que ceux dont l'intensité est supérieure (ou inférieure) à une valeur seuil optimale [172, 259]. La binarisation est appliquée après une égalisation de l'histogramme du scalogramme. Du fait de la différence

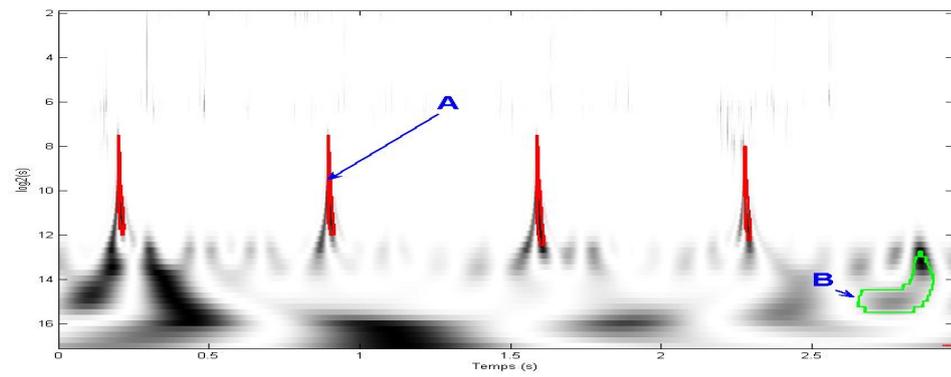
entre les trois bandes de fréquences : $\nu < F_{rot}$, $[F_{rot} \leq \nu \leq 350]$ Hz et $\nu > 350$ Hz, section 4.3.1, un seuillage local est appliqué dans chaque bande de fréquences.

Reprenant le scalogramme de la figure 4.5(a). Rappelons que F_{rot} est la fréquence de rotation des galets. En bas, la bande $\nu < F_{rot}$, représente les basses fréquences. La région entourée sur la figure 4.5(a)(zone B), peut correspondre à une déformation structurelle. Elle est représentée sur l'image binarisée, figure 4.5(c)(région B), par une région large avec un étalement latéral. Plus haut la bande des hautes fréquences, $\nu > F_{bruit}$ avec $F_{bruit} \simeq 350$ Hz, est représentée sur l'image binaire par une bande presque uniforme, figure 4.5(b)(rectangle pointillé). Les pics noirs au niveau de la bande intermédiaire $F_{rot} \leq \nu \leq 350$ Hz, figure 4.5(a)(pic A), sont représentés sur l'image binarisée par des régions avec un étalement vertical, figure 4.5(c)(région A). Ils peuvent être associés aux défauts locaux sur la surface du bandage.

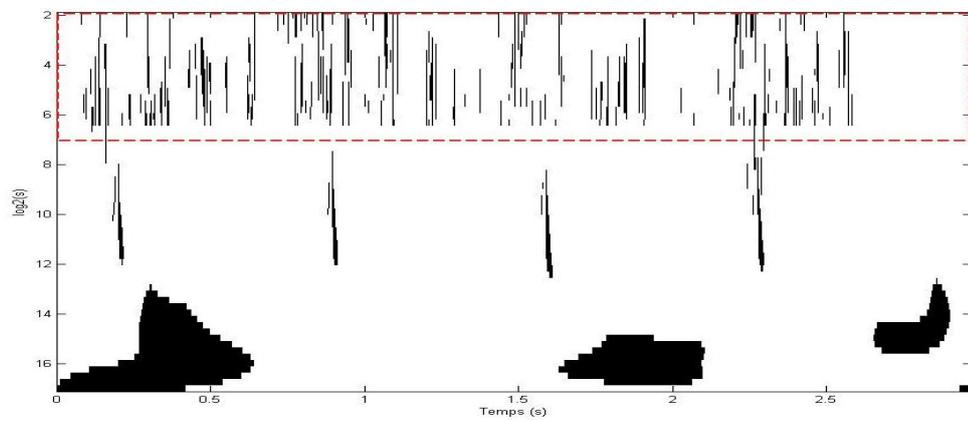
Les étapes de la segmentation

Les différentes étapes du processus de segmentation des scalogrammes par binarisation sont illustrées dans la figure 4.6 :

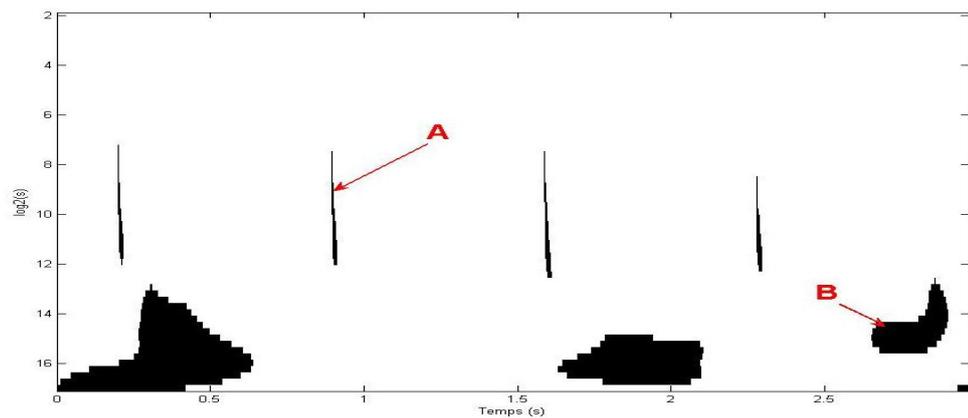
- la première étape du processus consiste à calculer, à partir du scalogramme son histogramme égalisé, figure 4.6(a) ;
- l'étape de l'égalisation de l'histogramme est suivie d'un seuillage automatique locale approprié à chacune des trois bandes de fréquences : $\nu < F_{rot}$, $[F_{rot} \leq \nu \leq 350]$ Hz et $\nu > 350$ Hz, figure 4.6(b). L'égalisation d'histogramme a pour but d'harmoniser la répartition des niveaux de luminosité de l'image, de telle manière à tendre vers un même nombre de pixel pour chacun des niveaux de l'histogramme. Cette opération vise à augmenter les nuances dans l'image ;
- ensuite, on calcule pour chaque zone d'intérêt (région noire) ses caractéristiques telles que l'aire, le contour, le centre de gravité, les moments statistiques, etc., en se basant sur ses coordonnées, un appariement est établi entre une régions de l'image binaire, figure 4.6(b), et sa correspondante sur le scalogramme, figure 4.6(d) ;
- une fois l'emplacement de l'ensemble des régions définit, on calcule l'énergie existante dans chacune d'elles. Seules les régions vérifiant certaines critères sur l'aire, la position et l'énergie sont retenues, figure 4.6(c). Chaque zone d'intérêt, affectée à un couple de galets donné, est représentée par un ensemble de paramètres constituant le vecteur forme. Ce vecteur caractéristique est utilisé ensuite dans l'étape de diagnostic pour l'identification de l'état du couple de galets. La figure 4.6(e) montre les régions finales extraites à partir du scalogramme initial.



(a)



(b)



(c)

FIG. 4.5 – (a) La représentation temps-échelle (scalogramme). L'image binaire du scalogramme avant (b) et après (c) extraction de zones d'intérêts.

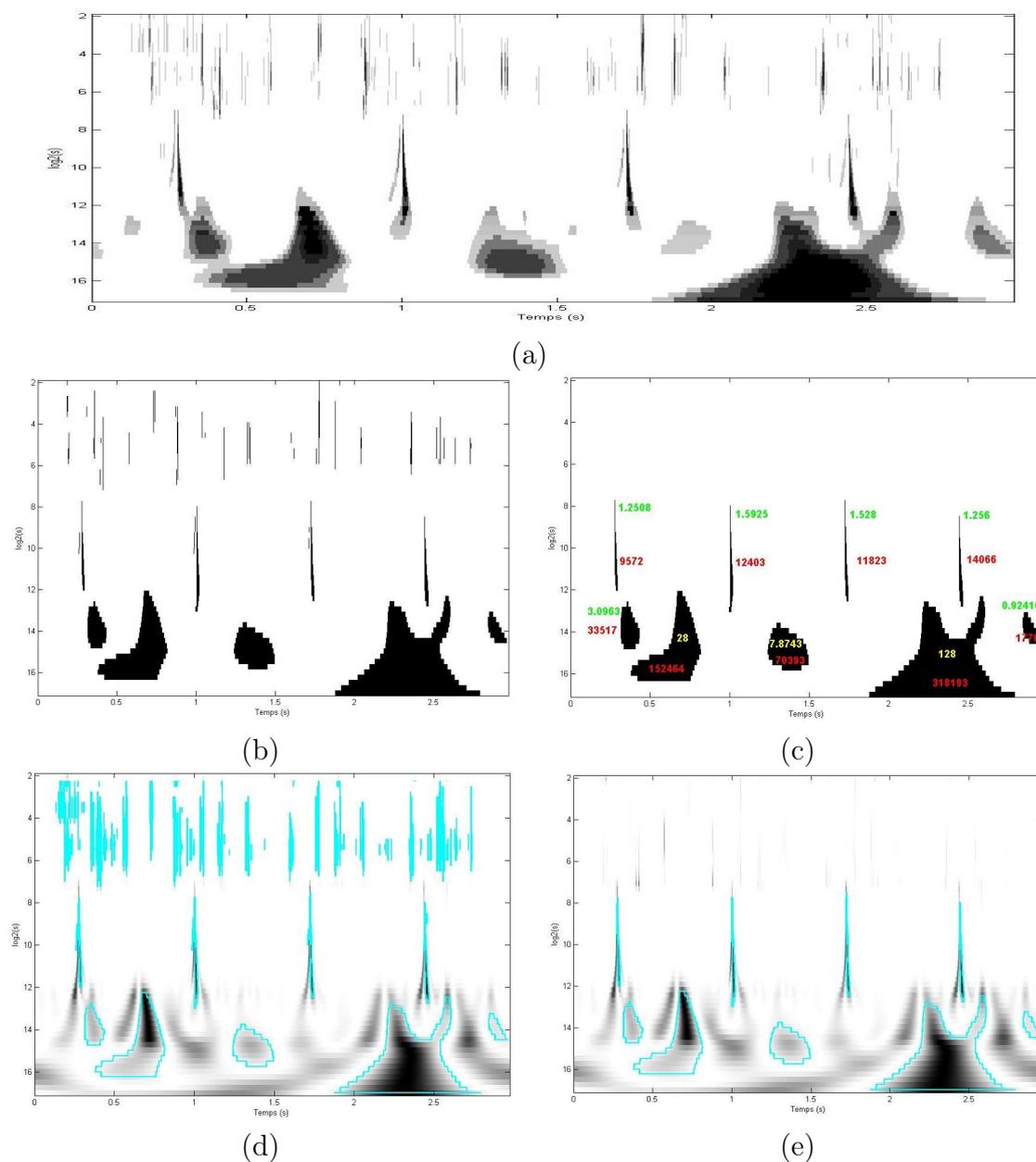


FIG. 4.6 – (a) Scalogramme après égalisation de son histogramme. (b) Seuillage local de l'image de l'histogramme égalisé. (c) L'image binaire après extraction de zones d'intérêts : la surface de la région (l'aire) et l'énergie. Extraction des contours des zones correspondantes sur le scalogramme avant (d) et après (e) l'extraction de zones d'intérêts.

4.5.2 Inconvénients et avantages de la binarisation des scalogrammes

Inconvénients de la binarisation

Les deux figures 4.7(a) et 4.7(b) représentent le passage de galets sans défaut montés sur l'ensemble des essieux fournis, respectivement, par le premier et le deuxième capteur. On remarque clairement une différence au niveau de la bande des hautes fréquences $\nu > 350$ Hz. Cette différence inter capteur, constatée sur la majorité des passages, est caractérisée par la présence de zones parasites dans le premier scalogramme figure 4.7(a) et leur absence sur la deuxième figure 4.7(b). Ces zones, avec un étalement latéral, sont caractérisées par des traits fins presque parallèles, peu distants et peu énergétiques. De telles bandes, si elles existent, sont plus marquées et plus denses dans un milieu d'acquisition bruité (chutes de pluie, l'existence de l'humidité au niveau des capteurs, etc.). En se basant sur ces données, la bande de fréquence $\nu > 350$ Hz est jugée peu porteuse d'informations et substituée à du bruit. Cette constatation a été confirmée par la suite par la société SD en adoptant une stratégie complètement différente de la nôtre. Pour cette raison, elle ne sera pas prise en compte dans la suite des traitements.

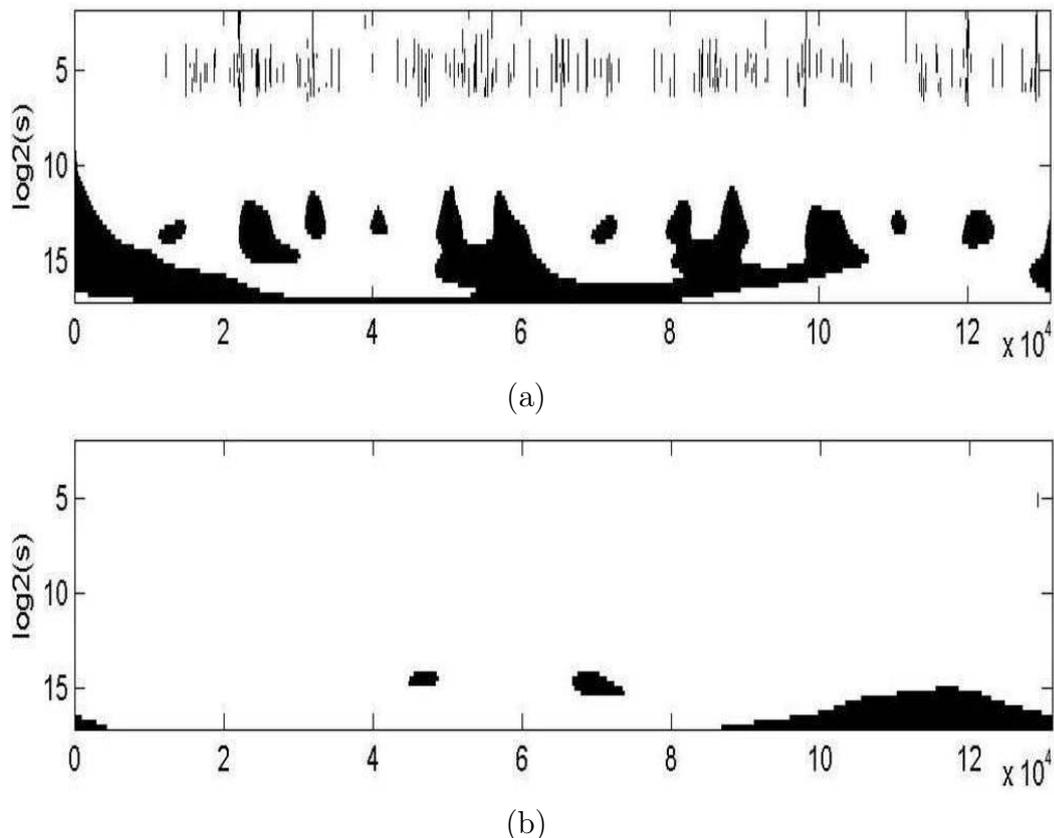


FIG. 4.7 – Images binaires du scalogramme d'un passage avec des galets neufs sur l'ensemble des essieux fournis par le premier capteur (a) et le deuxième capteur (b).

Un couple de galets sans défauts est toujours représenté sur le scalogramme segmenté par une zone blanche. En revanche, la discrimination entre les classes d'usures avec deux types de défauts différents, par exemple trous et usure à 100%,

n'est pas facile. Un autre problème est souvent rencontré et se caractérise par la présence de régions sous forme de dômes au niveau des essieux, figure 4.8(a). Ce phénomène est présent dans deux cas, le premier correspondant aux passages acquis durant la chute de pluie. Tandis que le deuxième se manifeste de temps en temps et lié à un dysfonctionnement du guidage lui même. L'explication de ces phénomènes n'est pas tout à fait évidente et leur existence n'a jamais été justifiée expérimentalement. Dans d'autres cas, le résultat de la segmentation est trop bruité et complètement illisible, figure 4.8(b), même après un bon nettoyage de l'image, figure 4.8(c).

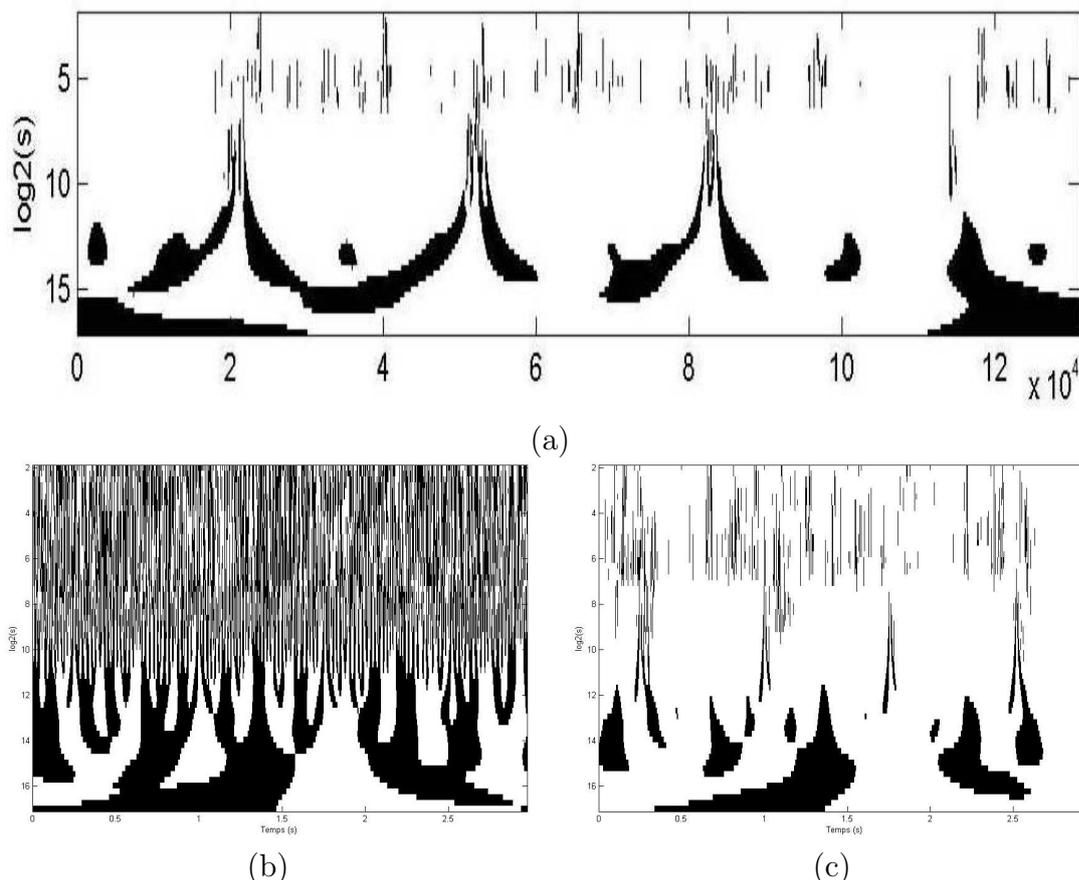


FIG. 4.8 – (a) Régions parasites au niveau des essieux (chocs). Images binaires bruitées (b) avant et (c) après l'extraction de zones d'intérêts.

Avantages de la binarisation

Malgré les difficultés rencontrées lors de l'étape de segmentation, cette première tentative n'est pas tout à fait un échec puisqu'elle nous a permis de faire une première réduction de dimension. Dans la suite du projet, seule la bande de fréquences $[F_{rot} \leq \nu \leq 350]$ Hz est jugée porteuse d'informations et est retenue pour l'étape de classification, figure 4.9(a). Comme indiqué précédemment, si un couple de galets ne se fait attribuer aucune région ce dernier est classé directement sans défaut. Dans le cas contraire, si une région, figure 4.9(b), est affectée à un couple de galets donné, il est jugé avec défaut. Dans ce cas, la classification se fait en deux temps. L'avantage ici est que l'identification du type du défaut, local ou structurel, se fait

naturellement grâce à la position de la coordonnée y du centre de gravité de la région (l'échelle), figure 4.9(a), par rapport à la fréquence de rotation du galet F_{rot} . Cependant la nature du défaut (trous, arrachement du bandage, etc.) est loin d'être facile à identifier.

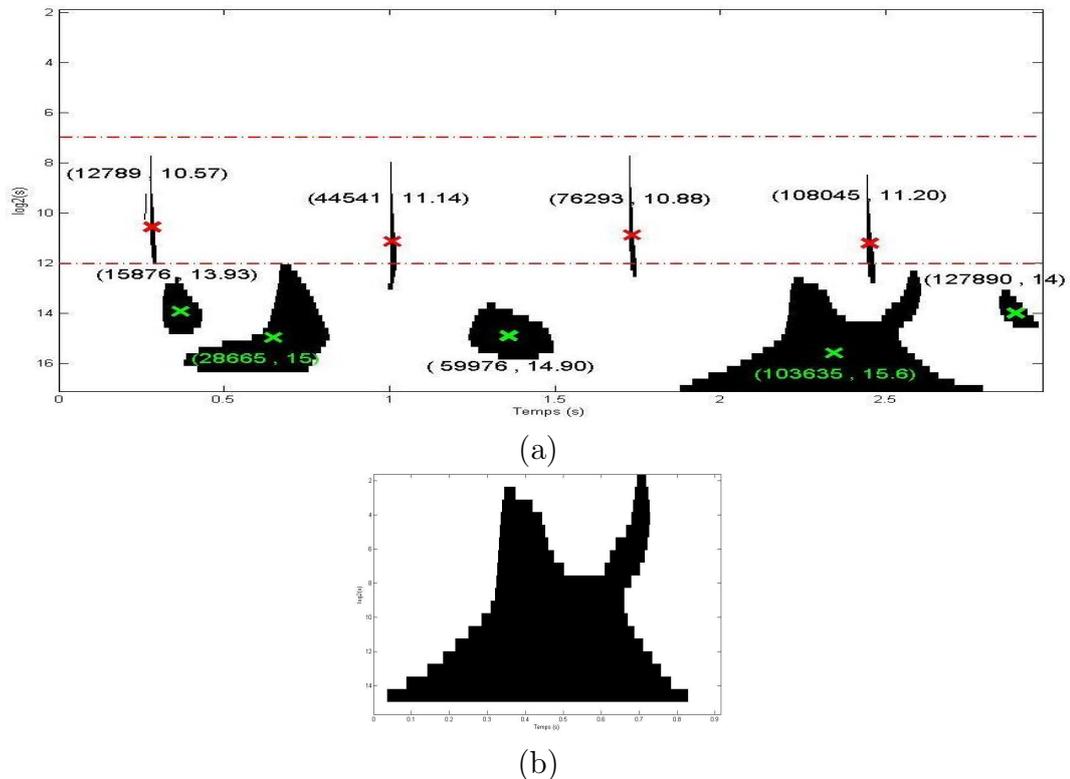


FIG. 4.9 – (a) L'ensemble de zones d'intérêts avec les coordonnées du centre de gravité. (b) Exemple de zone d'intérêt pouvant correspondre à un défaut structurel.

Bien que la méthode de binarisation d'images ait fourni des résultats partiellement satisfaisants, elle nécessite le réglage d'un grand nombre de paramètres puisqu'elle dépend fortement du signal traité. De plus, ces méthodes qui considèrent le plan temps-fréquence comme une image ne fournissent aucune description du contenu énergétique de la représentation temps-échelle (resp. temps-fréquence). Dans notre cas, pour avoir cette information nous devons faire un appariement entre l'image binaire et le scalogramme, ce qui prend du temps. En plus, la perte d'information lors de l'étape de la segmentation par binarisation rend la tâche de l'identification (classification) de la nature du défaut difficile. Toutefois, une bonne connaissance du comportement de galet sans défaut ou avec défaut au niveau des trois bandes de fréquences pourrait améliorer le résultat de la segmentation.

Conclusion

Le but de ce chapitre était de trouver l'approche la plus appropriée pour analyser les signaux vibratoires. De nature non stationnaire, ce type de signaux nécessite des méthodes temps-fréquence ou temps-échelle bien appropriées pour mieux décrire ces composantes. Pour atteindre cet objectif, nous avons présenté dans la première partie du chapitre les méthodes temps-fréquence les plus citées dans la littérature. Nous avons rappelé les éléments classiques et les limites des représentations temps-fréquence (RTF) pour introduire la transformée en ondelettes. Trois types de représentations temps-fréquence ont été discutées. La deuxième partie du chapitre a été réservée à la transformée en ondelettes continue et rappelle les propriétés fondamentales de la représentation temps-échelle (RTE) pour justifier son utilisation pour l'analyse de nos signaux. En effet, deux types principaux de défauts ont été distingués. Le premier type correspond aux défauts structuraux des galets et se manifeste en basses fréquences (ou en grandes échelles). Le second type correspond aux défauts surfaciques locaux et produit des vibrations à hautes fréquences (petites échelles). Pour avoir une bonne résolution en basses et en hautes fréquences l'analyse temps-échelle est la plus adaptée à notre problème puisque la précision relative est constante sur toute la bande de fréquences analysées ($\Delta\nu/\nu$ est constant). Une fois la transformée continue en ondelettes effectuée, le signal vibratoire est représenté par une image en niveau de gris, dite scalogramme, sur lequel les deux types de défauts sont clairement identifiables. Ensuite, grâce à un découpage temporel, chaque couple de galets CGk est représenté par une matrice de coefficients d'ondelettes CWT_k . Enfin, l'identification de l'état d'un couple de galets CGk se résume en une tâche de classification. Malheureusement, la dimension des matrices données (CWT_k) disponibles à la fin de l'étape de génération est très importante. Il est donc nécessaire de réduire cette dimension afin d'obtenir de meilleures performances en classification.

Plusieurs travaux traitent le problème d'extraction de l'information pertinente à partir du plan temps-fréquence et on fait l'objet de la troisième partie du chapitre. Nous avons essayé de recenser, dans d'un premier temps, les interprétations des représentations temps-fréquence (RTF) les plus utilisées. La manière la plus naturelle, était de considérer la RTF comme étant l'image du signal acquis, ainsi, plusieurs techniques appliquées sur les images pour la reconnaissance de formes, peuvent être exploitées pour l'interprétation temps-fréquence. De la même manière, nous avons proposé de procéder à une segmentation par binarisation de l'image temps-fréquence par l'algorithme d'Otsu. Cette segmentation permet une séparation automatique et efficace des différentes composantes contenues dans le signal quand celles-ci n'interfèrent pas entre elles dans l'image temps-fréquence. Cette première tentative nous a fourni des résultats partiellement satisfaisants. Toutefois, puisqu'elle considère le plan temps-fréquence comme une image, cette méthode ne fournit aucune description du contenu énergétique de la représentation temps-échelle. En plus, la perte d'information lors de l'étape de la segmentation par binarisation rend la tâche de l'identification (classification) de l'état du couple de galets difficile.

Nous allons présenter, dans le chapitre suivant 5, les deux techniques utilisées dans le cadre de la réduction de la dimension. La sélection de variables et l'approche

d'extraction de variables.

RÉDUCTION DE LA DIMENSION POUR LA CLASSIFICATION

Tableau. Représentation en deux dimensions de quelque chose d'inintéressant en trois. [Ambrose Bierce]

5.1 Introduction

La transformée en ondelettes continue (TOC) adopte un comportement dynamique lié au zoom adaptatif de l'échelle a pour suivre les singularités et les irrégularités du signal analysé. La notion de zoom permet d'avoir la même information pour une composante du signal dans plusieurs niveaux d'échelle. Cette redondance dans le plan temps-échelle défini par la TOC représente une propriété importante pour comprendre le comportement des composantes du signal. Barlow dans [19] considère que le concept de la redondance est fondamental à l'étude et affirme que c'est la redondance qui permet au cerveau de construire ses "cartes cognitives" ou "les modèles" de son environnement. Dans notre cas, chaque couple de galet extrême ou intermédiaire est représenté respectivement par 677376 et 733824 variables. Malheureusement, l'augmentation du nombre de variables n'améliore pas systématiquement la qualité de l'apprentissage ni celle de la classification. Le problème est que les bases d'apprentissage de taille finie n'arrivent pas à couvrir de manière exhaustive un espace de représentation de dimension élevée (curse of dimensionality) [22]. Pour obtenir une classification avec de bonnes performances, on est amené à réduire le nombre de variables sans pour autant perdre en performances. Ainsi, l'utilisation d'un nombre de descripteurs minimal à partir de la RTE peut faciliter énormément l'étape de la classification.

L'objectif fondamental de la réduction des variables est d'accentuer et compacter les informations importantes portées par la RTE (resp. RTF), sans pour autant perdre ses propres notions de temps, énergie et de fréquence utiles pour la manipulation des signaux non-stationnaires. Autrement dit, l'approche choisie pour l'extraction ou la sélection de variables doit conserver l'information temporelle et spectrale contenue dans la RTE (resp. RTF). Tout dépend de la façon dont on considère une représentation temps-fréquence (resp. RTE). Un ensemble de techniques ou d'approches de réduction de variables : statistiques, appliquées en imagerie, etc. peuvent

être alors utilisées. La diversité des travaux dans ce domaine est liée au problème traité et à l'application elle-même. Les méthodes de réduction de dimension, les plus utilisées dans la littérature, sont illustrées dans les paragraphes suivants. Dans un problème de classification, l'objectif est de définir un sous-espace de dimension q inférieure à p pour mieux représenter l'ensemble d'apprentissage [102].

La figure 5.1 illustre et rappelle les étapes de fonctionnement d'un système de RdF [218].

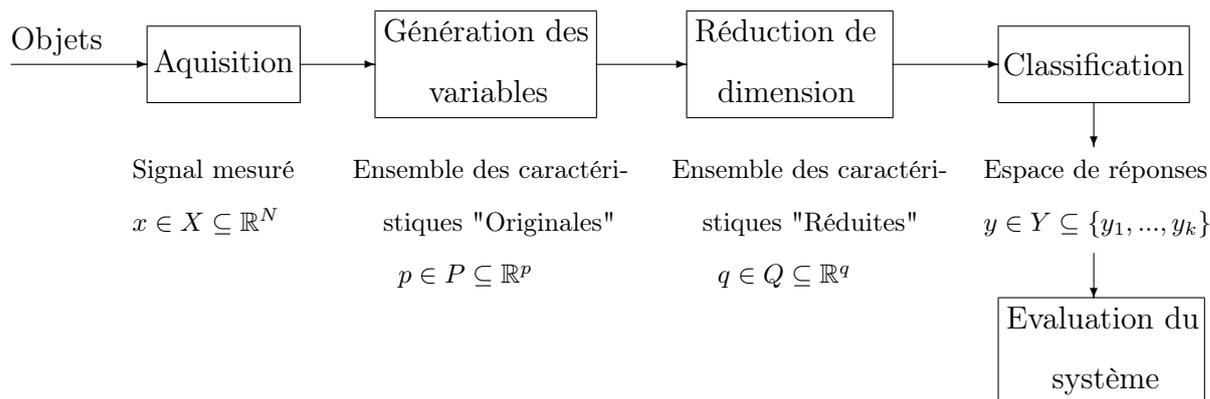


FIG. 5.1 – Système de reconnaissance de formes.

L'étape de réduction de dimension vient tout de suite après l'étape de génération de données et opère sur les segments du scalogramme résultant de cette étape, (cf. 4.3.2). Elle peut être obtenue par l'extraction des variables en projetant les p variables initiales le mieux possible dans un espace de plus petite dimension. Où bien par la sélection d'un sous-ensemble Sq de q variables dans l'ensemble de p variables d'origine Sp , ($Sq \subseteq Sp$). En mode d'apprentissage, l'étape d'extraction/sélection permet de définir l'espace de représentation le plus approprié pour la reconnaissance. C'est dans cet espace de représentation que sont définies les formes à reconnaître.

Nous traitons, dans ce chapitre, ce problème de réduction de dimension dans le cadre de la reconnaissance des formes statistiques et plus particulièrement dans le cadre de la classification supervisée. Nous présentons dans les sections suivantes quelques techniques d'extraction/sélection de variables pour la réduction de la dimension. Dans un premier temps nous allons aborder le problème de la réduction de dimension à l'aide de la méthode de sélection de variables. Les méthodes d'extraction de données sont discutées dans un deuxième temps. La méthode que nous avons finalement retenue est la DVS. La dernière section du chapitre est réservée à la décomposition en valeurs singulières (DVS) ainsi que sa validation à travers quelques expérimentations. Concernant la classification, deux tentatives sont étudiées. Dans la première, seules les valeurs singulières sont utilisées. Dans la deuxième, nous illustrons la possibilité d'exploiter les vecteurs singuliers plutôt que les valeurs singulières [105, 108].

5.2 Sélection de variables

5.2.1 Principe de la sélection

En partant d'un ensemble P des p variables d'origine, l'objectif est de sélectionner un sous-ensemble Q de q variables dans un espace des caractéristiques de taille minimale avec $q \leq p$ vérifiant un certain critère préalablement défini. Dans ce nouvel espace de caractéristiques, les frontières de décision entre les classes sont plus nettes.

La sélection des q caractéristiques pertinentes se déroule en deux étapes. La première étape consiste à explorer l'espace des différentes combinaisons possibles de variables pour trouver le sous-ensemble candidat des caractéristiques en utilisant une procédure de recherche et en se basant sur un critère d'arrêt. La deuxième étape consiste à choisir un certain critère d'évaluation J , avec lequel le sous-ensemble candidat est jugé pertinent ou pas. Le problème se résume donc à choisir un sous-ensemble ($Q \subseteq P$) tel qu'il optimise le critère d'évaluation J . Selon le critère choisi, J peut être une fonction à maximiser ou à minimiser.

En classification supervisée, la sélection d'un sous-ensemble de variables de taille minimale permet de réduire la complexité du problème en s'assurant que :

- les performances de la règle de classement ne diminuent pas trop voire même augmentent ;
- la distribution des données, dans l'espace restreint, doit être aussi proche que possible de la distribution des données dans l'espace des p variables d'origine.

5.2.2 Les étapes de la sélection de variables

Algorithmes de recherche

Le problème de sélection d'un sous-ensemble de q variables parmi les p variables avec $p \gg q$ est un problème combinatoire avec une complexité Np . Plusieurs études ont été effectuées pour tenter de réduire cette complexité, en considérant par exemple, les combinaisons possibles comme étant un graphe. L'espace de recherche est réduit en utilisant des heuristiques pour parcourir ce graphe [145]. Ces heuristiques retournent, sous certaines conditions, une solution optimale du problème ou une solution sous-optimale (ex. les algorithmes *séquentiels* [68], les algorithmes *flottants* [210]).

(a) Les méthodes séquentielles

Les algorithmes séquentiels consistent à rajouter ou à éliminer itérativement des variables. On peut, commencer avec un ensemble de variables vide et ajouter, à chaque étape, une variable à celles déjà sélectionnées (*Sequential Forward Selection*, SFS). On peut aussi, en partant de l'ensemble de toutes les variables,

éliminer à chaque étape une variable parmi celles déjà sélectionnées (*Sequential Backward Selection*, SBS).

Ces deux méthodes sont connues pour leur simplicité de mise en oeuvre et leur rapidité, la complexité étant polynomiale $O(p^2)$. La SFS commence la sélection avec un ensemble vide de variables, elle est donc moins complexe que l'algorithme SBS pour des problèmes de grande dimension. Cependant, les deux méthodes ne permettent pas d'explorer tous les sous-ensembles possibles de variables et ne permettent pas un retour arrière pendant la recherche. Dans les deux cas, la solution obtenue n'est pas optimale. Une façon pour réduire ce problème de sous-optimalité des algorithmes SFS et SBS est d'alterner les procédures SFS et SBS permettant ainsi d'ajouter des variables et puis d'en éliminer d'autres ;

(b) Les méthodes flottantes

Les méthodes flottantes (*Sequential Floating Search methods*) consistent à appliquer après chaque étape *forward* (resp. *backward*) autant d'étapes *backward* (resp. *forward*) nécessaires pour que le sous-ensemble résultant améliore le critère d'évaluation. Le nombre de variables ajoutées ou éliminées à chaque étape est déterminé dynamiquement en fonction de la valeur du critère d'évaluation J . Par conséquent, aucun paramètre n'est à régler au préalable.

D'autres méthodes de recherche tentent de réduire le temps de calcul, soit en incluant une recherche aléatoire de sous-ensembles de variables [114, 161, 162]. Soit en insérant le processus de sélection dans l'algorithme d'apprentissage. Dans cette dernière approche, les variables sont supposées indépendantes et l'évaluation d'une variable à chaque niveau de l'arbre ne tient donc pas compte de son interaction avec d'autres variables du graphe.

En reconnaissance des formes, les algorithmes de sélection sont généralement groupés en deux catégories : les algorithmes optimaux et les algorithmes sous-optimaux. Ces méthodes peuvent plus ou moins être associées à n'importe quelle mesure d'évaluation [148, 150].

Évaluation des variables

Pour sélectionner un sous-ensemble de variables, il faut disposer d'une mesure permettant d'évaluer sa pertinence. Cependant, la notion de pertinence est une notion entièrement relative au problème traité [21]. L'objectif, selon [138] et [145], est de trouver toutes les variables fortement pertinentes et une partie des variables faiblement pertinentes. Malheureusement, en pratique il est impossible de faire la nuance entre pertinence forte et pertinence faible. Généralement, deux approches sont couramment utilisées pour évaluer la pertinence d'un sous-ensemble de variables sélectionnées [30] :

- dans les approches de type *filtrage* (*filters*), la sélection se fait en utilisant des critères d'évaluation fondés uniquement sur les données et indépendamment

du classifieur. Les variables sont filtrées avant les phases d'apprentissage et de classification. Ces critères sont souvent regroupés en trois catégories : les mesures de distance, les mesures d'information et les mesures de dépendance ;

- contrairement aux premières approches, les approches *enveloppantes* (*wrappers*) tiennent également compte de la règle de classement (classifieur) dans le calcul du critère d'évaluation. Elles intègrent la méthode de sélection dans la construction du classifieur. Les paramètres nécessaires à la règle de classement sont estimés à partir de l'ensemble des variables sélectionnées. Le critère d'évaluation est simplement la probabilité d'erreur et dépend fortement du couple (données, règle de classement). De ce fait, le critère choisi diminue la capacité en généralisation des approches enveloppantes. Un autre inconvénient concerne le coût de calcul qui est très élevé, dans ce cas, car la sélection boucle sur le processus d'apprentissage.

Les filtres n'ont pas les défauts des wrappers. Ils sont beaucoup plus rapides et reposent sur des considérations plus théoriques. Ils permettent de mieux comprendre les relations de dépendance entre variables. Mais, comme ils ne prennent pas en compte les biais de l'algorithme de classification, les sous-ensembles de variables générés donnent un taux de reconnaissance plus faible.

Pour donner un score à un sous-ensemble, une première solution est de donner un score à chaque variable indépendamment des autres et de faire la somme de ces scores. Pour évaluer une variable, l'idée est de déterminer sa corrélation avec la variable de classe. Malheureusement cette approche dite *feature ranking* n'élimine pas les variables redondantes, d'autre part il est possible que des variables peu corrélées avec la classe deviennent utiles lorsqu'on les considère dans le contexte des autres variables. L'autre solution, dite *subset ranking*, consiste à évaluer un sous-ensemble dans sa globalité. On se rapproche ici de l'apprentissage de la structure de réseau bayésien.

Il existe un intermédiaire entre *feature ranking* et *subset ranking* utilisé avec de bons résultats dans le cadre de la CFS (correlation based feature selection) par Mark Hall [109]. C'est cette méthode qui sera utilisée pour la sélection de variables dans la partie ??.

Le score d'un sous-ensemble est construit en fonction des corrélations variable-classe et des corrélations variable-variable selon la formule suivante :

$$r_{\theta_s} = \frac{k\bar{r}_{\theta_i}}{\sqrt{k + k(k-1)\bar{r}_{ij}}}$$

avec r_{θ_s} le score du sous-ensemble de cardinal k , \bar{r}_{θ_i} la moyenne arithmétique des corrélations entre θ et les variables i et \bar{r}_{ij} la moyenne des k^2 intercorrélations entre variables. Cette équation exprime que le score du sous-ensemble augmente si les variables sont fortement corrélées avec θ et diminue si elles sont fortement corrélées entre elles.

De plus cette formule est valable dans le cadre de scores normalisés (i.e. de variance unitaire). L'idée est de dire qu'un bon sous-ensemble est constitué de variables

hautement corrélées avec la classe (pour ne pas garder les indépendantes), et peu corrélées entre elles (pour éviter la redondance). Il s'agit d'une approximation car on ne prend en compte que les interactions d'ordre 1. Pour plus d'information le lecteur peut se référer à l'article [109].

La corrélation entre deux variables peut être définie de plusieurs façons. On peut utiliser par exemple le coefficient de corrélation statistique ou bien un test d'indépendance statistique comme le test du chi-carré χ^2 . Il est également possible d'utiliser la notion d'information mutuelle. D'autres méthodes d'évaluation de sous-ensembles sont possibles. Des méthodes récentes combinant les wrappers et les filtres sont présentées dans [107].

Dernièrement, Pena et al ont proposé en 2007 un nouvel algorithme sous contraintes basé sur les Réseaux Bayésiens pour l'apprentissage de la couverture de Markov dans de "très" gros volumes de données. Une évaluation empirique a été menée sur plusieurs bases de données synthétiques et réelles de UCI Repository, dont la base Thrombin constituée de 140000 variables! Il lève donc un verrou scientifique : la passer à l'échelle. Aux JFRB 2008 à Lyon 29-30 Mai, Rodrigues de Morais et al. montrent également comment passer à l'échelle avec ce genre d'algorithmes en améliorant encore les performances quand les nombres d'instance est faible (data-efficiency). Les variables sont discrètes mais il existe des façons simples de discrétiser les variables continues en fonction de l'objectif (discrétisation supervisée).

Critères d'arrêt

Une fois l'algorithme de recherche et le critère d'évaluation choisis, il faut un critère pour arrêter la recherche et définir le sous-ensemble de variables sélectionnées. Le choix de ce critère d'arrêt est très important car les résultats de la méthode de sélection en dépendent fortement. Cependant, ce choix reste très difficile à faire en pratique. Il existe plusieurs manières pour le fixer. On peut, par exemple, décider d'arrêter la procédure de recherche lorsqu'un nombre maximum de variables à sélectionner ou d'itérations à exécuter est atteint [148]. Néanmoins, fixer préalablement le nombre de variables à sélectionner est pratiquement difficile, ce qui rend son utilisation comme critère d'arrêt rare.

Une autre façon pour fixer le critère d'arrêt, est de se baser sur le critère d'évaluation en fixant un seuil sur J . Dans ce cas, la procédure de recherche est arrêtée lorsque ce dernier est atteint ou lorsque l'ajout ou la suppression d'une nouvelle variable, dans les algorithmes séquentiels, n'améliore pas la mesure J [189]. Si le critère d'évaluation utilisé est monotone (croissant ou décroissant) vis-à-vis du nombre de variables, on peut choisir d'arrêter la recherche par exemple lorsque son évolution $\theta = \frac{J(S_{q-1})}{J(S_q)}$ tend vers 1.

Lorsque la dimension des données n'est pas trop élevée, une bonne heuristique, dont la complexité est suffisamment raisonnable est de calculer pour les différents sous-ensembles sélectionnés une estimation de la probabilité d'erreur. L'ensemble des variables sélectionnées est celui qui donne les meilleures performances.

La probabilité d'erreur peut être estimée comme dans les approches *enveloppantes* à l'aide d'une procédure de validation croisée ou de Bootstrap. Cette stratégie nécessite alors un ré-apprentissage du modèle pour chaque sous-ensemble de variables candidat.

5.3 Extraction des variables

La méthode de sélection de variables cherche à sélectionner le meilleur sous-ensemble dans l'espace original de variables qui maximise la séparabilité entre les classes. L'extraction des variables permet de déterminer la meilleure combinaison des variables originales en projetant l'espace d'origine de dimension p dans un sous-espace de dimension réduite q . Parmi ces méthodes on cite :

- l'analyse en composantes principales (ACP) ;
- la décomposition en valeurs singulières (DVS) ;
- la poursuite de projection (PP).

Dernièrement la décomposition en valeurs singulières (DVS) a été largement utilisée dans les domaines de traitement du signal et l'imagerie pour la réduction de la dimension pour la classification [105, 106, 126, 108, 263, 264, 173]. De ce fait, on s'intéresse dans ce qui suit à présenter et argumenter la méthode de réduction de dimension à l'aide de la DVS.

5.3.1 Décomposition en valeurs singulières

Définitions

La décomposition en valeurs singulières (DVS) est très populaire dans l'analyse statistique de données et le traitement du signal. L'existence de la DVS a été établie par le géomètre Italien Beltrami en 1873, 20 ans seulement après la conception de la matrice comme un multiple de quantité par Cayley.

Pour toute matrice $X \in \mathbb{R}^{M \times N}$, rectangulaire réelle, il existe des matrices orthogonales $U \in \mathbb{R}^{M \times M}$ et $V \in \mathbb{R}^{N \times N}$, [100, 249], telles que

$$U^t X V = \Sigma \quad (5.1)$$

où Σ est une matrice pseudo-diagonale. À partir de l'équation (5.1), X peut s'écrire comme suit

$$X = U \Sigma V^t. \quad (5.2)$$

L'équation (5.2) représente la décomposition en valeurs singulières de la matrice X où U et V contiennent respectivement les vecteurs singuliers de gauche et droite de X . Cette décomposition est unique et cela est assuré par la contrainte d'orthogonalité

des matrices U et V . La matrice $\Sigma \in \mathbb{R}^{M \times N}$ est pseudo diagonale et possède sur sa diagonale les valeurs singulières σ_{ii} de la matrice X ordonnées de façon décroissante [100] tel que :

$$UU^t = I,$$

$$VV^t = I,$$

$$\sigma_{11} \geq \sigma_{22} \geq \dots \geq \sigma_{pp} \geq 0, \text{ avec } p = \min(m, n).$$

I est matrice identité et $\sigma_{ij} = 0$ si $i \neq j$ et $\sigma_{ij} \geq 0$ sinon; pour simplifier la notation on note la i -ème valeur singulière σ_{ii} par σ_i . Les σ_i sont les racines carrées non négatives des valeurs propres de la matrice $X^t X$.

Une fois la décomposition réalisée, le rang r de la matrice décomposée est égal au nombre de σ_i non nulles, telque :

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_p = 0,$$

ceci définit alternativement le rang de l'espace nul de X , et qui est égal au nombre de valeurs singulières nulles. On peut alors noter la matrice X comme suit :

$$X = \sum_{i=1}^r \sigma_i (u_i v_i^t), \quad (5.3)$$

où u_i et v_i sont respectivement les colonnes et les lignes des matrices U et V . Les σ_i sont les valeurs singulières associées à ces vecteurs. On peut voir, dans l'expression (5.3), que la DVS décompose la matrice X en une somme pondérée de matrices de rang 1. Une matrice de rang p peut donc se décomposer, de façon unique, en une somme de r matrices de rang 1.

La DVS permet ainsi d'estimer le rang et l'identification de l'espace nul de la matrice. Cet espace peut être défini pour les vecteurs singuliers de gauche et de droite, et l'espace engendré par ces vecteurs correspond aux valeurs singulières pour lesquelles $\sigma_i = 0$.

En plus, comme la DVS décompose une matrice qui n'est pas forcément carrée, ainsi, il est possible de décomposer directement la représentation de distribution temps-échelle sans l'utilisation de la matrice de covariance. Ainsi, la décomposition DVS directe maintient toutes les informations appropriées sur les espaces nul, lignes et colonnes de la matrice de données dans une forme compacte. Comme la DVS décompose une matrice pas forcément carrée sans avoir besoin du calcul de la matrice de covariance, la base résultante n'est pas aussi sensible aux problèmes de gamme dynamique que l'ACP. Théoriquement, il n'est pas possible d'inverser une matrice qui n'est pas carrée. Ainsi, l'implémentation de la DVS est un compromis entre la définition théorique et la forme souple en pratique, en introduisant la notion du pseudo-inverse d'une matrice.

5.4 La décomposition en valeurs singulières des matrices des coefficients en ondelettes CWTk

L'analyse des signaux temps-échelle par la DVS repose sur le type de la distribution temps-échelle utilisée. De préférence, cette distribution doit être linéaire et avec une bonne résolution. Certains auteurs ont jugés que l'utilisation des valeurs singulières suffit largement pour représenter les distributions temps-fréquence (DTF) des signaux [79, 126]. Alors que d'autres, évoquent la limite des valeurs singulières du fait qu'elles n'apportent pas d'information significative concernant le comportement des composantes spectrales dans la matrice temps-fréquence. Pour y remédier, ils préfèrent utiliser les vecteurs singuliers pour la caractérisation des signatures à partir d'une distribution temps-fréquence et estiment qu'ils peuvent apporter plus d'informations que les valeurs singulières [106, 105, 108].

C'est la propriété d'orthogonalité mutuelle des matrices U et V qui les a poussé à s'intéresser à ces vecteurs singuliers au lieu des valeurs singulières. Les valeurs singulières σ_i représentent l'importance des vecteurs singuliers dans la composition de la matrice. D'une autre manière, les vecteurs singuliers correspondant aux valeurs singulières élevées ont plus d'informations pertinentes sur la structure de la RTF que les autres vecteurs singuliers. Cette approche basée sur les vecteurs singuliers utilise une estimation de la fonction de distribution de ces vecteurs associée à la distribution temps-fréquence des signaux pour l'extraction de caractéristiques pertinentes. Ce problème de choix d'utilisation des valeurs ou des vecteurs singuliers, a été traité et amplement détaillé dans [13] pour l'analyse des signaux Doppler. Nous nous sommes inspirés de ce travail dans notre étude. Ainsi, dans un premier temps, seules les valeurs singulières ont été utilisées comme vecteurs forme caractéristiques pour les matrices temps-échelle CWT_k . Ces vecteurs formes sont utilisés dans l'étape de classification, (cf. 6). Dans un deuxième temps, l'intérêt des vecteurs singuliers par rapport aux valeurs singulières pour l'extraction de caractéristiques à partir des RTE des signaux vibratoires a été vérifiée. Ils existe différentes façons de manipuler les vecteurs singuliers obtenus à partir d'une distribution temps-fréquence. Nous présentons quelques approches dans le paragraphe 5.5.

Les éléments utilisés dans la suite sont essentiellement alimenté par la lecture de [79, 126, 106, 105, 108, 43, 13].

5.4.1 Formulation du problème

Rappelons que la matrice CWT_k , de dimension $M \times N$ avec $M \gg N$, est la matrice des coefficients d'ondelettes associée au couple de galets CGk, (cf. 4.3.2). Chaque matrice CWT_k peut être décomposée en

$$CWT_k = U\Sigma V^t,$$

où U de taille $M \times M$ et V ($N \times N$) sont deux matrices orthogonales appelées respectivement matrice de gauche et de droite. Σ de dimension $M \times N$ est la matrice

diagonale telle que $\Sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$. La composante $\sigma_{ij} = 0$ si $i \neq j$ et $\sigma_{ii} \geq 0$ avec $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N$. Les σ_i représentent les valeurs singulières de la matrice CWT_k .

Lorsqu'on décompose une représentation temps-fréquence X quelconque, les colonnes u_i de la matrice U et les colonnes v_i de V sont décorréélées deux à deux et représentent respectivement les caractéristiques temporelles et fréquentielles de X [43]. L'écriture

$$X = \sum_{i=1}^{\min(n,m)} \sigma_i u_i v_i^t.$$

montre que la représentation temps-fréquence s'écrit sous forme de somme de représentations de type $u_i v_i^T$ auxquelles on associe une amplitude σ_i . Chaque caractéristique fréquentielle u_i correspond donc à une caractéristique temporelle v_i et à une énergie σ_i^2 .

Pour simplifier les calculs et réduire le nombre de composantes mises en jeu on ne conserve qu'un certain nombre de caractéristiques fréquentielles et temporelles qui préserve la plus grande partie de l'énergie du signal, par exemple 90%. Cela revient à considérer maintenant

$$S = U_q \Sigma_q V_q^t,$$

où Σ_q est la matrice carrée diagonale contenant les q premières colonnes et lignes de Σ , et U_q et V_q sont les matrices contenant les q premières colonnes de U et V , où q vérifie

$$\frac{\sum_{i=1}^q \sigma_i^2}{\sum_{i=1}^{\min(n,m)} \sigma_i^2} \geq 90\%.$$

5.4.2 Étude de la décomposition en valeurs singulières d'un passage complet

La DVS d'un passage sur la voie

Pour valider le choix de l'analyse par SVD discutée précédemment, prenons un scalogramme, avec les trois bandes de fréquences, d'un signal vibratoire issu de l'installation sur la voie, figure 5.2 (au dessus). La reconstruction de la matrice en utilisant les dix premières valeurs singulières et les vecteurs singuliers correspondants de gauche et de droite de la décomposition SVD du scalogramme, figure 5.2 (au dessous), est représentée sur la figure 5.2 (au dessous).

Nous remarquons clairement que la compression des données (réduction) ne modifie quasiment pas l'aspect du scalogramme initial ($\approx 99\%$ d'énergie est retenue). Les deux matrices de gauche et de droite représentent, dans notre cas, respectivement les échelles et le temps et les σ_i représentent leur poids.

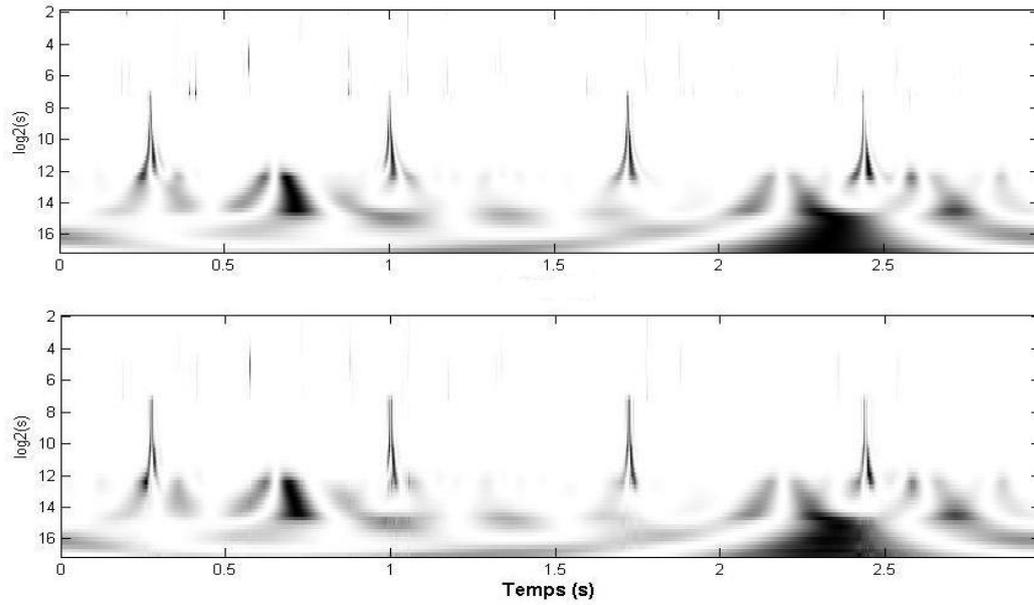


FIG. 5.2 – Scalogramme initial d'un passe (au dessus) et le même scalogramme après avoir conservé les dix composantes ($p = 10$).

Considérons maintenant seulement la bande de fréquences intermédiaire [$F_{rot} \leq \nu \leq 350$] Hz, figure 5.3(a), du même scalogramme. Cette fois-ci, les 5 premières valeurs singulières et les vecteurs singuliers correspondants des caractéristiques spectrales et temporelles suffisent pour reconstruire et avoir toute l'information portée sur le scalogramme, figure 5.3(b). Nous retrouvons les même pics avec pratiquement la même intensité que ceux existants sur le scalogramme, figure 5.3(a).

Voyons maintenant ce qui se passe, de plus près, pour les caractéristiques spectrales U et temporelles V issues de la SVD. La figure 5.4(b) représente les deux premiers vecteurs des caractéristiques spectrales (échelles) et temporelles.

Le scalogramme, figure 5.4(a), correspond à un signal vibratoire acquis avec l'installation sur la voie. Sur chaque essieu arrière est monté un couple de galets avec un défaut local : trous, usure 50%, usure 100% et enfin usure 100% avec des trous. Les essieux avant sont tous composés de couples de galets sans défauts à part les essieux MI1 et ME2 qui comportent une usure structurelle respectivement l'usure patatoïde et l'existence de vis fixé sur le bandage. Un défaut local se manifeste au niveau de la bande de fréquences $10.76 \leq \nu \leq 350$ Hz. Sur le scalogramme correspondant dans cette même bande de fréquence, figure 5.4(a), il existe un ensemble de pics au nombre de 4 au niveau des échelles ($7 \leq a \leq 12$) qui caractérise ces défauts locaux. Le comportement spectral des défauts locaux est bien reflété par les deux premiers vecteurs des caractéristiques spectrales, figure 5.4(b), qui montrent une augmentation significative d'énergie dans le même intervalle ($7 \leq a \leq 12$). La localisation temporelle et la durée sont aussi montrées par les deux premiers vecteurs de droite. De plus, la différence en amplitude entre les 4 types de défauts locaux est aussi bien montrée sur la figure 5.4(b).

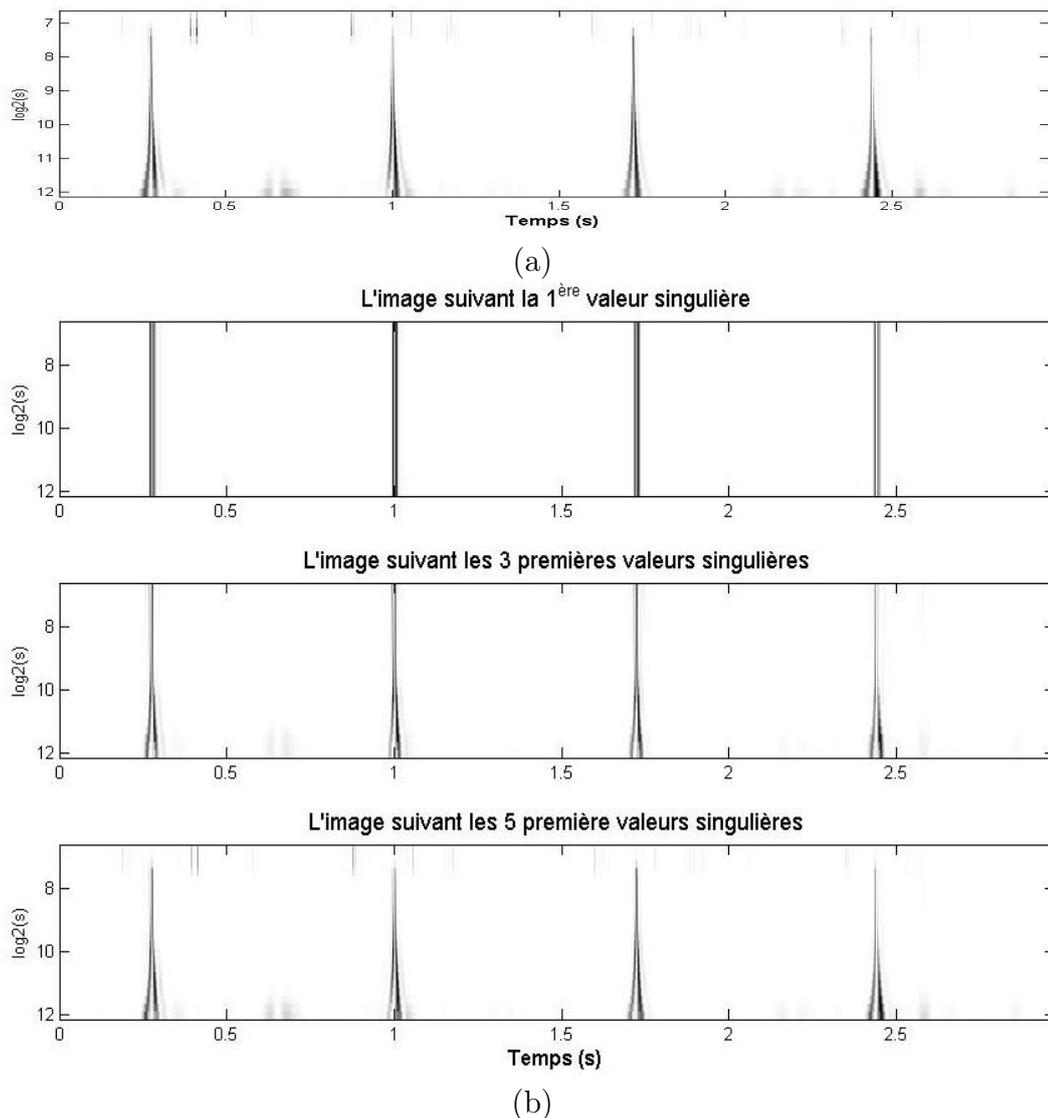


FIG. 5.3 – (a) La bande de fréquences intermédiaire (défauts locaux). (b) Reconstruction du scalogram correspondant à la bande de fréquences intermédiaire à partir des 5 premières composantes.

La DVS d'un passage sur l'ADV

La figure 5.5(a) représente un signal vibratoire acquis avec l'installation sur l'ADV et le scalogramme associé. Le passage est effectué dans le sens ME2 avec l'essieu ME2 en tête (côté pivot) et en configuration 2, (cf. 3.4.2). Les essieux ME2 avant et arrière, MI1 arrière et ME1 arrière sont composés de couples de galets sans défauts. Par contre, sur l'ensemble des essieux MI2 avant et arrière, MI1 avant et ME1 avant est monté, respectivement, un défaut avec usure : à 75%, à 100%, patatoïde et un défaut de type trous.

Les deux premiers vecteurs singuliers des caractéristiques spectrales, figure 5.5(b), représentent l'évolution d'énergie dans la bande intermédiaire ($7 \leq a \leq 12$) correspondant aux défauts locaux. Regardons maintenant ce qui se passe au niveau des deux premiers vecteurs de droite qui représentent le comportement temporel au

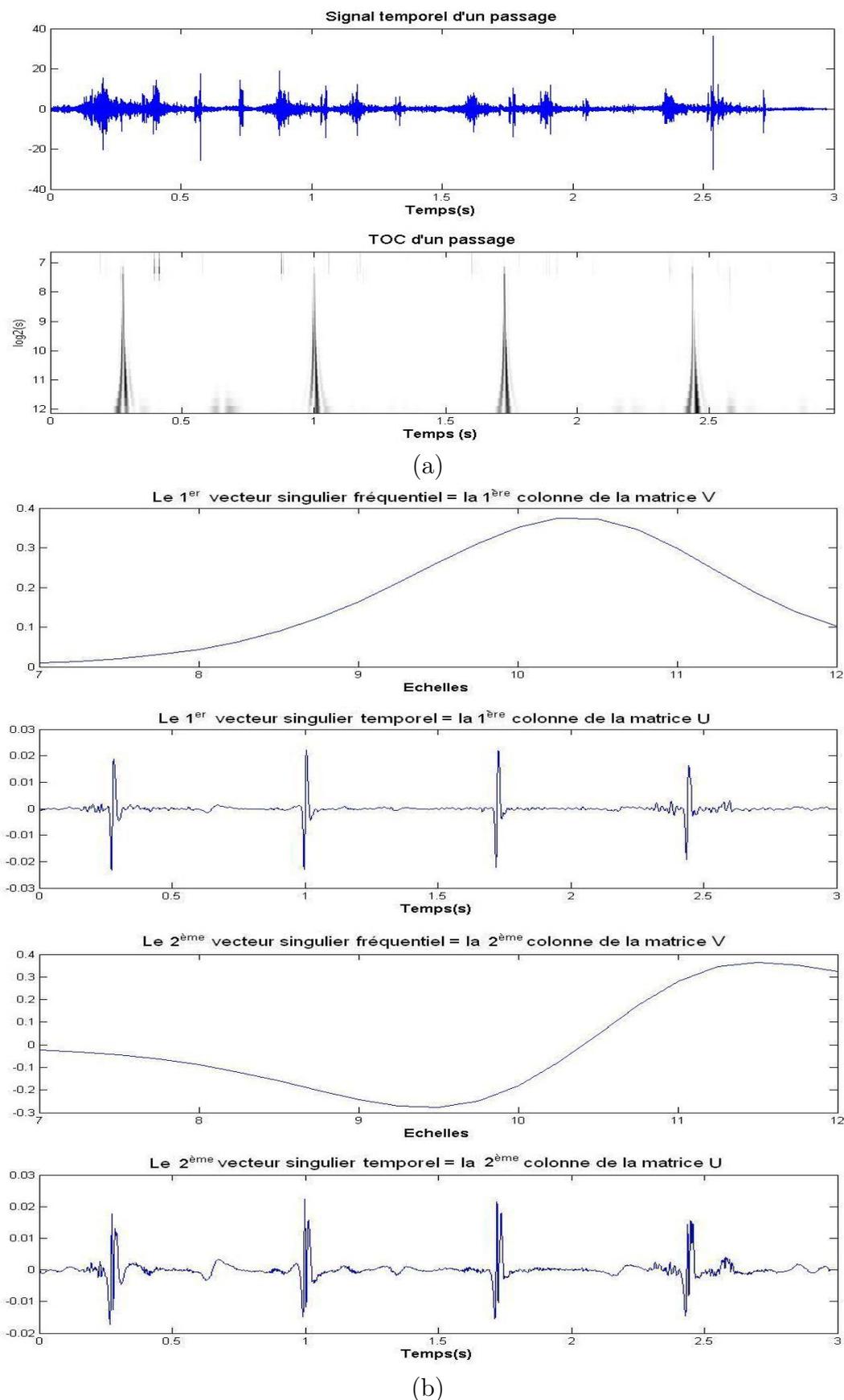
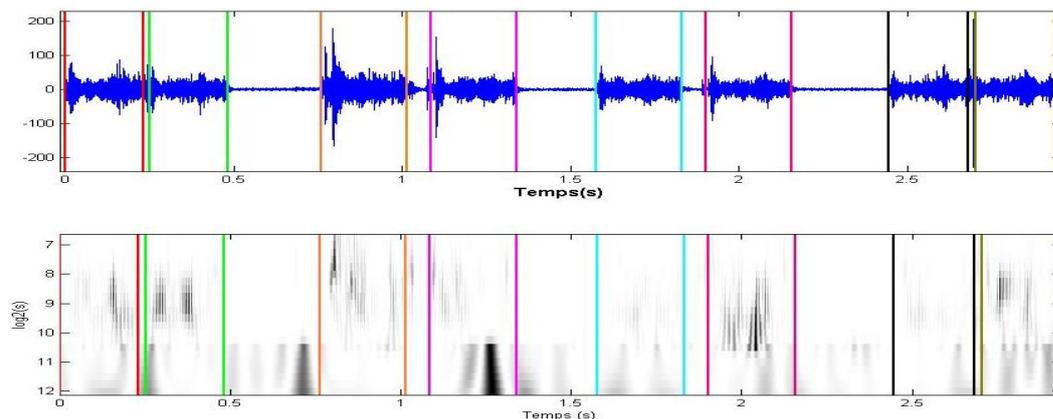


FIG. 5.4 – (a) Le signal vibratoire pour un passage acquis avec l'installation sur la voie et son scalogramme approprié. (b) Les deux premières caractéristiques spectrales et les poids temporels correspondants issues de la SVD.

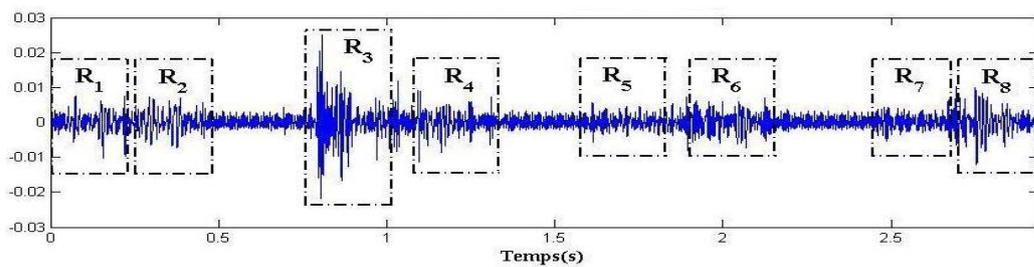
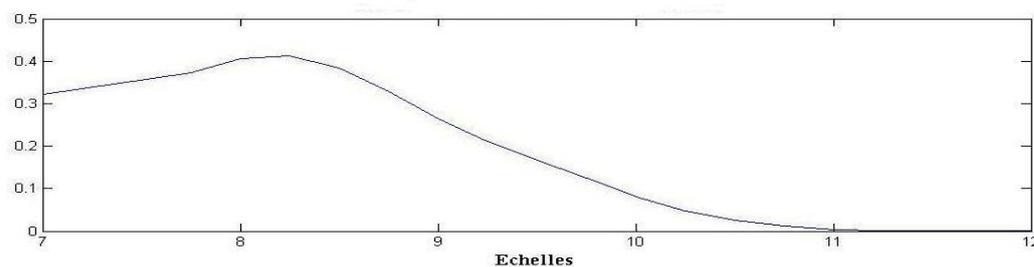
niveau de chaque essieu. Notons R1, R2, R3, R4, R5, R6, R7 et R8 l'ensemble des régions correspondant aux couples de galets CGk , $k = 1, \dots, 8$ sur le premier et le deuxième vecteur singulier des caractéristiques temporelles tels que CG3, CG4, CG6 et CG8 sont avec défaut de type, respectivement, usure à 100%, usure à 75%, usure patatoïde et trous. Nous distinguons, d'une manière générale, que dans les régions R3, R4, R6 et R8 le signal est plus amplifié. Les perturbations remarquées dans ces régions sont justifiées par l'importance du défaut. Les deux régions les plus marquées sont R3 et R8 et correspondent respectivement à une usure à 100% et les trous. Les autres régions R1, R2, R5 et R7 correspondent aux couple de galets sans défauts.

Le comportement temporel d'un galet sans défaut par rapport à un galet avec défaut ainsi que son importance sont bien reflétés par le premier vecteur singulier des caractéristiques temporelles. Sur la figure 5.6 nous remarquons bien une différence nette entre les segments correspondant aux couples de galets sans défauts, figure 5.6(R1, R5 et R7), et ceux correspondants aux couples de galets défectueux, figure 5.6(R3, R4 et R8).

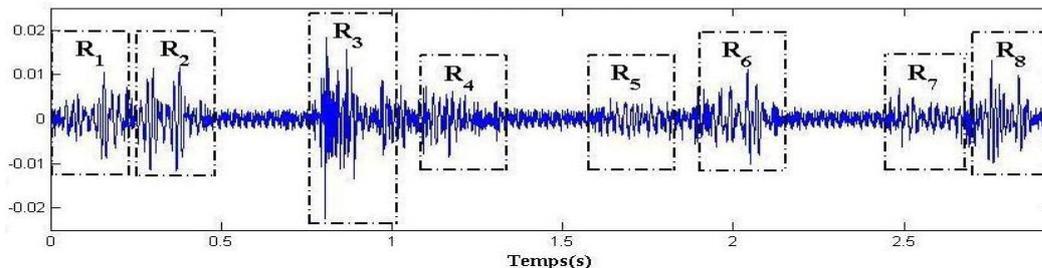
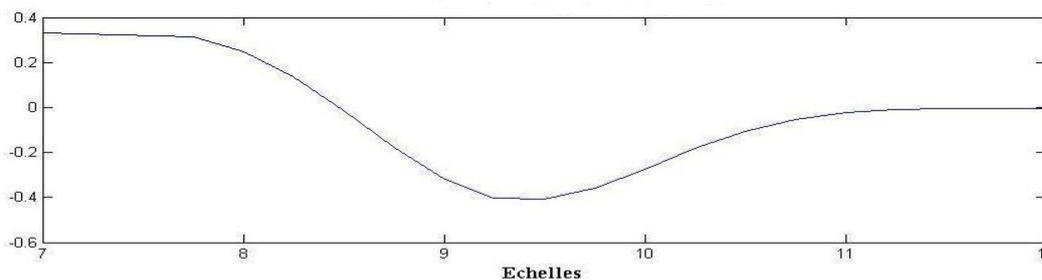
À ce stade, le comportement des vecteurs singuliers se montre assez prometteur pour donner de bons résultats de classification. Pour l'étape de classification, nous appliquons la DVS, cette fois-ci, sur les segments des scalogrammes SCG_k correspondants aux couples de galets CGk , $k \in 1, \dots, 8$. Nous nous attendons à ce que ce même comportement et la différence entre un galet sans et avec défaut soit bien reflétée à la fois par les valeurs et les vecteurs singuliers. Nous espérons ainsi que les résultats obtenus à partir de la décomposition en valeurs singulières des segments SCG_k puissent être utilisés comme un moyen de caractérisation des signaux vibratoires. Un ensemble réduit de caractéristiques pertinent peut même augmenter la séparabilité entre les classes d'usures et ainsi, fournir un meilleur taux de classification avec un coût d'analyse de données réduit [132]. Cette hypothèse est vérifiée et illustrée dans le chapitre 6.



(a)



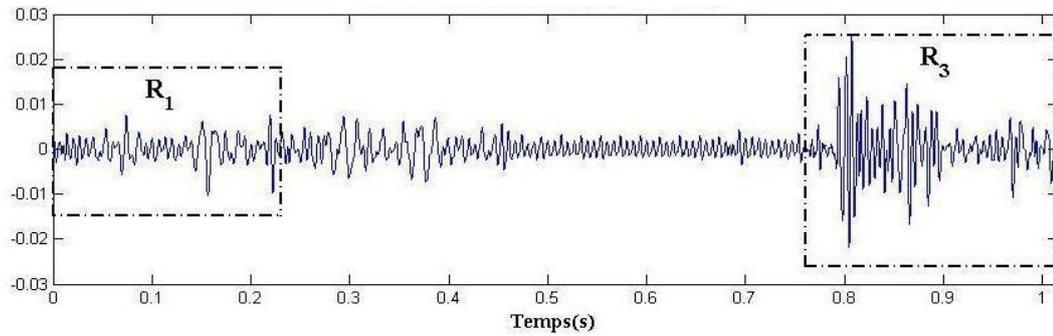
(b)



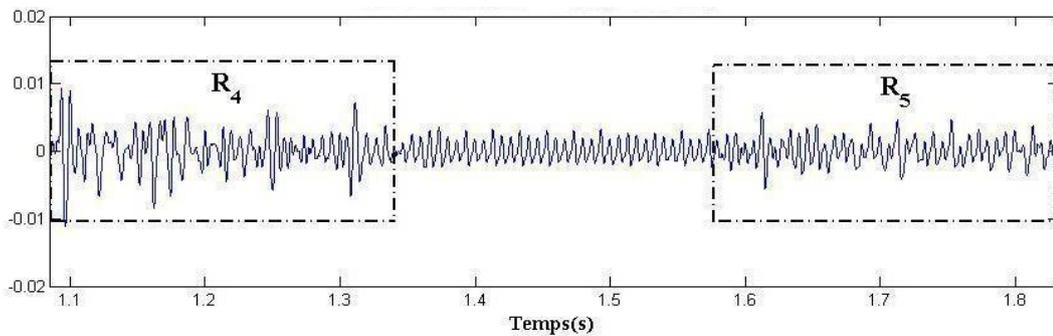
(c)

(b)

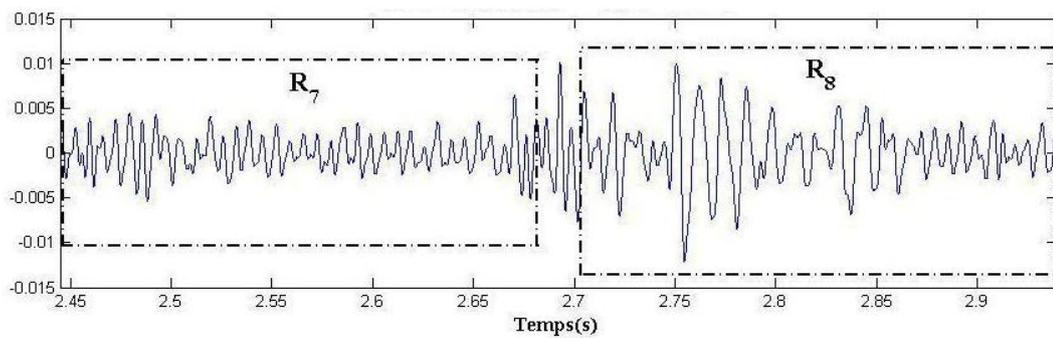
FIG. 5.5 – (a) Le signal vibratoire acquis avec l’installation sur l’ADV et son scalogramme associé. (b) Les deux premières caractéristiques spectrales et les poids temporels correspondants issues de la SVD.



(a)



(b)



(c)

FIG. 5.6 – Segments du premier vecteur singulier des caractéristiques temporelles correspondant aux couples de galets : (a) R_1 sans défaut et R_3 usure à 100%, (b) R_4 usure à 75% et R_5 sans défaut, (c) R_7 sans défaut et R_8 avec trous.

5.5 Définition de signatures basées sur les vecteurs singuliers

5.5.1 Introduction

La classification des signaux non-stationnaires pour la détection et l'identification des défauts est devenue, ces dernières années, le problème le plus considérablement étudié. D'un point de vue purement statistique, le test du Ratio de Wald [268] peut être considéré comme une contribution importante au développement des techniques en-ligne. Il convient également de mentionner les travaux de Basseville et Nikiforov [20]. Cependant, la plupart des méthodes de détection des défauts disponibles dans la littérature [202] sont fondées sur le calcul du résiduel où l'on dispose d'un modèle pour le système et on peut comparer la sortie réelle avec celles produites par le modèle. Dans le cas où le modèle du système n'est pas disponible, de nombreux chercheurs préconisent la construction d'un modèle mesuré en utilisant les données d'entrée et de sortie (RdF). Dans de telles situations, ces travaux se basent sur les techniques de traitement de signal pour accentuer l'effet du défaut et le rendre plus apparent à l'opérateur.

F. Chowdhury et J.L Aravena dans [50], ont examiné le cas particulier des systèmes de transmission de puissance et ont proposé une approche de modèle libre basée sur le traitement de banc de filtres et la définition de "vecteurs d'information instantanée" pour la détection et l'identification des défauts. Dans une autre référence [12], les mêmes auteurs ont montré qu'un banc de filtres peut séparer un signal en plusieurs composantes orthogonales et permet la définition de bandes de fréquences étroites où l'effet d'un défaut est facilement repérable. Par contre, cette décomposition ne se prête pas facilement à la mise en oeuvre d'un algorithme automatique de détection des changements. Dans leurs récents travaux de recherches, ces auteurs se basent sur l'utilisation de pseudos-signatures de puissance pour la détection et l'identification de signaux non-stationnaires. Cette méthode a été introduite par Venkatachalam Vidya en 1998 et amplement détaillée dans son rapport de thèse [265]. L'originalité de ces travaux réside dans le fait que l'auteur a introduit la notion d'indépendance entre les signatures calculées pour une classe de signaux non-stationnaires, la position et la longueur de l'événement (défaut).

Dans ces premiers travaux [263], l'auteur a proposé une nouvelle méthode basée sur la transformation en ondelettes continue et en définissant des représentations appelées "pseudo-signatures de puissance". Ces signatures capturent l'information à différents échelles, indépendamment de la longueur (la durée), l'endroit et la grandeur du signal (ou le défaut). Ces signatures caractérisent, essentiellement, la distribution de l'énergie de l'échelle d'une manière indépendante du temps. Les signatures sont alors invariantes aux intervalles temporels et les segments du signal sont tous caractérisés par la même signature. Ceci est une caractéristique extrêmement importante, et qui n'est généralement pas considéré dans les travaux de classification. Dans [264], l'auteur a fait le rapprochement entre la DVS et l'aspect théorique de sa méthode. Ainsi, il a proposé une approche simple pour déterminer cette signature en utilisant une décomposition en valeurs singulières (SVD) de la transformation en

ondelettes continue (TOC), et en extrayant les composantes principales [53].

Pour un signal $x \in L^2(\mathbb{R})$ avec CWT , $c_\psi^x \in H = L^2(\mathbb{R}^2, C_\psi^{-1} \frac{da db}{a^2})$, où ψ est admissible, la méthode consiste à approximer $c_\psi^x(a, b)$ par le produit de deux éléments séparable de la forme

$$c_\psi^x(a, b) = s_\psi^x(a) r_\psi^x(b).$$

où $s_\psi^x \in S = L^2(\mathbb{R}, C_\psi^{-1} \frac{da}{a^2})$, et $r_\psi^x \in R = L^2(\mathbb{R}, db)$. La fonction s_ψ^x correspond à la pseudo signature de puissance du signal x .

Plusieurs auteurs se sont inspirés de ces travaux pour développer des détecteurs de changement. Dans [167], M. Luo, J.L Aravena et Fahmida N. Chowdhury ont proposé une méthode basée sur la définition de coefficients de corrélations basée sur les pseudo-signatures de puissance (PPS) dans le domaine de la sécurité des avions. Le banc d'essai est le modèle de la NASA pour un avion de type Boeing 747 implémenté dans l'environnement DASMAT par Gary Balas de l'université du Minnesota. La méthode utilisée est décrite dans la partie 5.5.3.

Une autre technique basée sur la fonction de distribution de probabilité des vecteurs singuliers (FDPVs) a été introduite par Groutage dans [105]. Puisque les vecteurs singuliers sont orthonormés, leurs éléments élevés au carré peuvent être vus comme des valeurs de la fonction de densité de probabilité (FDP). Cette dernière est utilisée comme la signature caractéristique du signal. Cette technique est illustrée dans le paragraphe 5.5.4.

Dans le même esprit, nous avons testé l'efficacité de cette méthode sur nos signaux. Nous avons essayé de trouver une signature commune pour une classe donnée. Nous avons vu, dans la phase du découpage temporel, (cf. 3.4.2), que la longueur du signal d'un même défaut dépend non seulement de sa position - le signal d'un défaut monté sur un essieu extrême est plus long que celui monté sur un essieu intermédiaire - mais aussi de la vitesse du tramway ($L = V/T$). Nous pensons par conséquent que l'utilisation de cette approche, basée sur des signatures indépendamment de la durée du signal, est un bon moyen pour unifier les signatures pour une classe donnée et améliorer les performances du classifieur. Les paragraphes suivants illustrent les méthodes citées préalablement sans trop les détailler.

5.5.2 La pseudo signature de puissance (PPS)

Rappels et définitions

L'idée de base est de créer des signatures qui caractérisent la distribution de l'énergie d'un signal qui sont indépendantes de la durée effective de chaque composante du signal (défaut). Cette méthode se base sur le concept de la distribution spectrale de l'énergie existante dans le scalogramme issu de la transformée en ondelettes continue (TOC) du signal. Elle propose une représentation qui permet de

définir une "distribution d'énergie instantanée", appelé la pseudo signature de puissance.

Rappelons que la transformée en ondelettes continue (TOC) est destinée à fournir une analyse multirésolution naturelle d'une fonction $x \in L^2(\mathbb{R})$ donnée dans le plan temps-échelle grâce aux paramètres de temps, b , et d'échelle a , (cf. 4.1.4). La TOC de la fonction $x \in L^2(\mathbb{R})$ est donnée par

$$c_{\psi}^x(a, b) = \langle x, \psi^x(a, b) \rangle = \int_{-\infty}^{+\infty} x(t) \overline{\psi_{a,b}(t)} dt. \quad (5.4)$$

où $\psi_{a,b}$ est la fonction dilatée et translatée de l'ondelette mère $\psi(t) \in L^2(\mathbb{R})$. Elle est définie par $\psi_{a,b} = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right)$. Formellement la TOC peut être représentée comme une projection

$$\Gamma : L^2(\mathbb{R}) \rightarrow H = L^2\left(\mathbb{R}^2, c_{\psi}^{-1} \frac{dad b}{a^2}\right),$$

de rang M où M est un sous-ensemble fermé de H [61]. La propagation de l'énergie de la TOC dans le plan temps-échelle est définie par la distribution d'énergie appelée scalogramme.

L'énergie du signal sur un support, R , donnée est défini comme

$$E_{\psi}^{x(R)} = C_{\psi}^{-1} \int \int_{(a,b) \in \mathbb{R}^2} SC_{\psi}^x(a, b) \frac{dad b}{a^2}. \quad (5.5)$$

Le scalogramme de la fonction $x(t)$, $c_{\psi}^x(a, b)$, est donné par

$$SC_{\psi}^x(a, b) = |c_{\psi}^x(a, b)|^2. \quad (5.6)$$

Ainsi, le scalogramme peut être utilisé comme la fonction de densité d'énergie temps-échelle.

Le principe de la méthode

Considérons le signal $x \in L^2(\mathbb{R})$ avec la TOC $c_{\psi}^x(a, b)$ où ψ est une ondelette admissible. Le scalogramme SC_{ψ}^x associé à c_{ψ}^x peut être considéré comme la fonction de puissance temps-échelle correspondante et la fonction $SC_{\psi}^x(., b)$ comme "la distribution de puissance d'échelle à l'instant b ".

Examinons maintenant l'intervalle $[b_1, b_2]$ sur lequel $SC_{\psi}^x(a, .)$ est continue. En se basant sur le théorème de la valeur moyenne pour l'intégrale¹, Il existe alors

¹Si la fonction $f(.)$ est continue sur un intervalle $[a, b]$, alors il existe un nombre $\mu \in [a, b]$ tel que $f(\mu)(b - a) = \int_a^b f(t) dt$.

$b_0 \in [b_1 - b_2]$ telle que

$$SC_{\psi}^x(a, b_0) = \int_{b_1}^{b_2} \frac{|c_{\psi}^x(a, b)|^2}{b_2 - b_1} db.$$

Si l'ondelette ψ a un support compact, alors pendant que l'échelle diminue la valeur de $SC_{\psi}^x(a, b_0)$ est essentiellement indépendante des valeurs de $x(t)$ en dehors de l'intervalle $t_1 < b < t_2$. Ainsi la distribution de puissance d'échelle peut être estimée par les scalogrammes des petits segments du signal. Pour un ensemble petit d'échelles, elle serait essentiellement indépendante de la longueur de l'enregistrement (signal) [265]. L'idéal serait de disposer d'une ondelette telle que, pour une classe donnée de signaux, la transformée en ondelettes correspondante est séparable, c'est-à-dire

$$c_{\psi}^x(a, b) = s_{\psi}^x(a)r_{\psi}^x(b). \quad (5.7)$$

Par conséquent, la distribution de puissance d'échelle à l'instant b_0 pourrait être donnée par

$$SC_{\psi}^x(a, b_0) = |s_{\psi}^x(a)|^2 \int_{b_1}^{b_2} \frac{|r_{\psi}^x(b)|^2}{b_2 - b_1} db. \quad (5.8)$$

Il est clair que la distribution normalisée serait essentiellement indépendante de b_0 . Ainsi, la fonction d'échelle $s_{\psi}^x(a)$, convenablement normalisées, pourrait être utilisée comme une signature de puissance indépendante du temps pour caractériser une classe de signaux.

La faisabilité de ce concept est renforcée par le fait que si l'on s'éloigne des signaux de $L^2(\mathbb{R})$, on peut trouver des fonctions dont la TOC "formelle" est séparable. Cela revient à déterminer alors une fonction d'ondelette $\psi \in L^2(\mathbb{R})$ admissible qui admet un résultat similaire pour des fonctions de $L^2(\mathbb{R})$ telle que notée dans l'équation (5.7). Comme il a été démontré par le théorème suivant [265],

Théorème 5.5.1

Pour n'importe quelle fonction non triviale $x \in L^2(\mathbb{R})$, et une fonction d'ondelette admissible (non triviale) donnée $\psi \in L^2(\mathbb{R})$, l'espace des transformées en ondelettes continue (TOCs) de x ne contient aucun élément de la forme donnée par $c_{\psi}^x(a, b) = s(a)r(b)$, où $s \in L^2(\mathbb{R}, C_{\psi}^{-1} \frac{da}{a^2})$, et $r \in L^2(\mathbb{R}, db)$.

Il a été prouvé dans [264] que théoriquement il n'existe aucune ondelette admissible qui admet une fonction séparable de la transformée en ondelette continue (TOC). Par conséquent, le théorème établit la nécessité d'obtenir seulement une fonction d'approximation séparable pour une TOC donnée. Il faut donc utiliser des décompositions qui satisferont l'équation (5.7) d'une manière approximative. La

distribution d'énergie instantanée fournit par $s_\psi^x(a)$, convenablement normalisée, est ensuite définie comme "**la pseudo-signature de puissance**" du signal.

Dans la pratique, la génération de ces pseudos signatures de puissance repose sur la décomposition en valeurs singulières (SVD) des matrices de dimensions finies. Cette méthode est dérivée de la décomposition de la TOC d'un signal comme une somme de termes séparables. Cette analyse se fonde sur les résultats suivants [265].

Proposition 5.5.2

La transformée en ondelettes continue peut toujours être exprimée sous la forme

$$c_\psi^x(a, b) = \sum_i \sigma_i s_i(a) r_i(b). \quad (5.9)$$

Pour chaque i , $s_i(a) \in H = L^2(\mathbb{R}, C_\psi^{-1} \frac{da}{a^2})$, $r_i(b) \in L^2(\mathbb{R}, db)$. L'ensemble des fonctions $\{s_i\}_i$ et $\{r_i\}_i$ sont complètes dans respectivement H et L^2 .

La première composante principale de c_ψ^x , notée $PC [c_\psi^x]$, est donnée par $\sigma_1 s_1(a) r_1(b)$. La fonction s_1 peut ensuite être utilisée pour définir la pseudo signature de puissance associée au signal $x(t)$.

La proposition 5.5.2 fournit ainsi, une décomposition de la fonction $c_\psi^x(a, b)$ comme une somme de termes séparables. Cette décomposition est similaire à la décomposition DVS appliquée aux matrices de dimensions finies. Le calcul est considérablement simplifié si la détermination des pseudos signatures de puissance est réduite à un problème standard de matrice DVS. Dans [265], l'auteur a montré que sous certaines conditions, une discrétisation de la matrice TOC dans le temps et en échelle appropriée peut être effectuée de la manière suivante

$$[U, S, V] = SVD(TOC).$$

où la TOC devient la matrice d'exemples, V la matrice d'échelle et U la matrice du temps. En particulier la première colonne de la matrice V devient une approximation discrète de la pseudo signature de puissance du signal $x(t)$. Il convient de noter ici que cette discrétisation de la TOC ne correspond pas à une transformation en ondelettes discrète. La transformation discrète conventionnelle créée par une analyse multi-résolution a tendance à ignorer les détails du signal. En fait, pour la détection des défauts, les détails sont souvent les plus utiles pour l'étude du comportement des composantes du signal. La méthode de discrétisation présentée dans cette section, permet de se concentrer sur n'importe quelle gamme d'échelle et en se basant sur les détails du signal.

La mise en oeuvre de la méthode

Pour illustrer cette procédure sur nos signaux, prenons par exemple la moyenne de 50 pseudo-signatures de puissance pour les trois classes : "sans défaut", "trous" et

"usure à 75%". Les vecteurs moyens résultants pour chaque classe et chaque capteur sont illustrés sur les figures 5.7(a), (b) et (c). Nous constatons en premier lieu que pour une classe donnée, les trois capteurs ont pratiquement le même comportement. Les trois pseudo-signatures correspondant à une classe sont globalement semblables. En deuxième lieu, les trois pseudo-signatures moyennes correspondantes aux trois classes ont un comportement différent. Nous pensons, par conséquent, que l'utilisation de ces PPS dans l'étape de la classification pourra augmenter le taux de bonne classification.

Regardons maintenant les deux figures 5.8 et 5.9. La classe Usure normale regroupe les deux classes de défaut Trous et Usure 75%. Nous pouvons remarquer que sur certains intervalles d'échelles, les PPS correspondant aux couples de classes : (Sans usure, usure normale) et (Trous, usure75%) prises pour un capteur donné, sont facilement séparables. Ainsi, une sélection intelligente d'échelle dans ces intervalles peut maximiser la séparation entre les trois signatures, donc entre les classes. Le taux de classification ne sera que meilleur. Nous allons voir dans le chapitre 6, si cette hypothèse est vérifiée ou pas. Une deuxième méthode de réduction de dimension peut être considérée; basée sur les méthodes automatiques de réduction de variables décrites dans la partie 5.2. La sélection de variables est appliquée sur chaque capteur par capteur et par expert afin de maximiser le taux de reconnaissance.

Pour le premier expert et pour chaque capteur, figure 5.8(a), figure 5.8(b) et figure 5.8(c), la PPS est la moyenne de 100 PPS correspondantes aux couples de galets non défectueux. La deuxième signature est la moyenne calculée à partir d'un mélange de 100 PPS correspondant à la classe Trous et la classe Usure 75% (50 PPS de chacune), figure 5.9(a), figure 5.9(b) et figure 5.9(c). Nous remarquons que, pour les trois capteurs, les deux signatures sont facilement séparables dans l'intervalle d'échelle [9 – 11]. La même observation est constatée pour le deuxième expert.

5.5.3 Coefficient de corrélation et indicateur de distance basée sur la PPS

Définition de la méthode

Dans leurs article [167], M. Luo, J.L Aravena et Fahmida N. Chowdhury proposent l'utilisation de la technique de la matrice DVS (PPS) décrite plus haut, (cf. 5.5.2), pour déterminer l'état d'un vol. Ils supposent que les données pour un vol normal sont disponibles et peuvent être traitées afin de déterminer les pseudo-signatures de puissance pour des conditions normales. Une première alternative consistait à calculer, au cours d'un vol réel, la TOC pour les données issues d'un capteur et, pour un instant t donné, déterminer un coefficient de corrélation avec la pseudo-signature de puissance. Si la valeur du coefficient de corrélation est élevée, par conséquent, la distribution de puissance d'échelle est semblable à celui de la pseudo signature puissance d'un état normal de vol. Par opposition, un petit coefficient de corrélation indique une situation anormale. Le coefficient de corrélation est

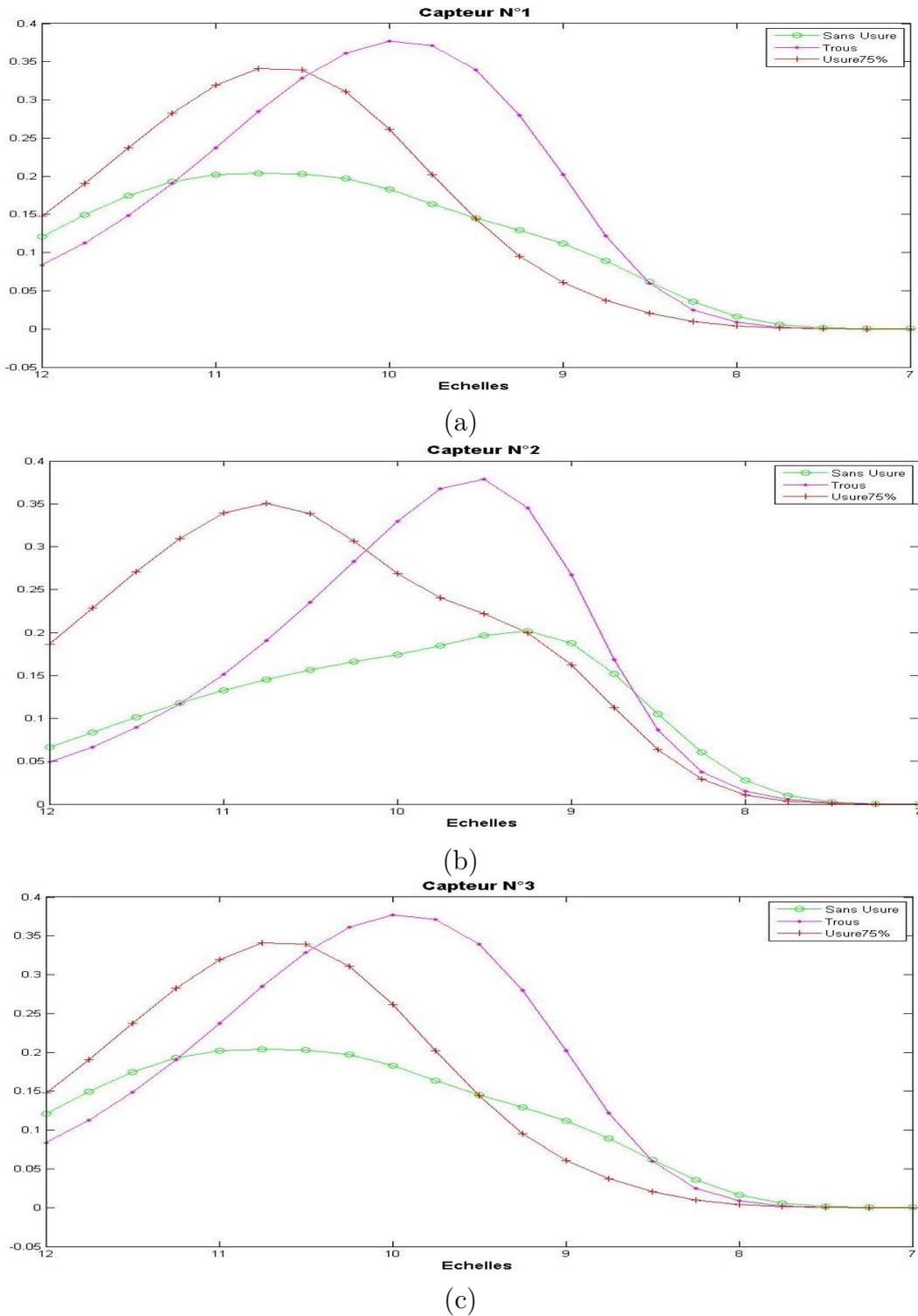
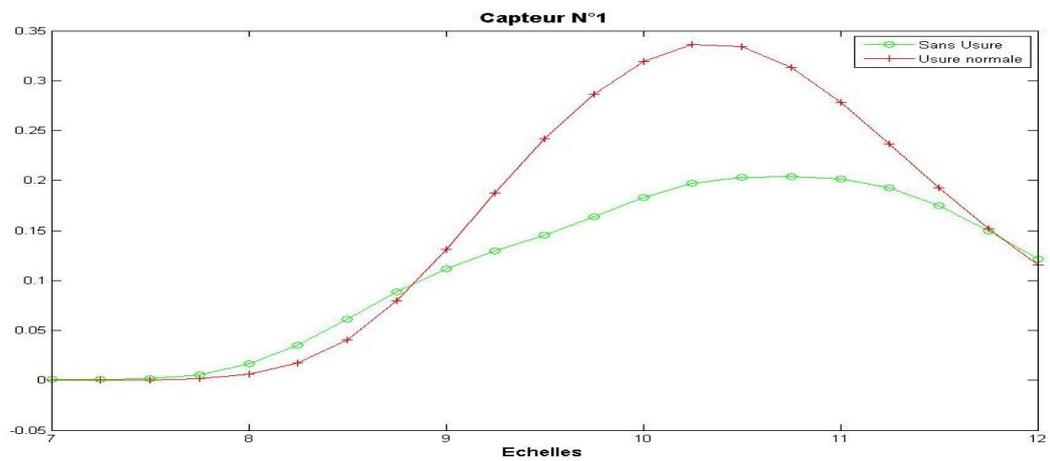
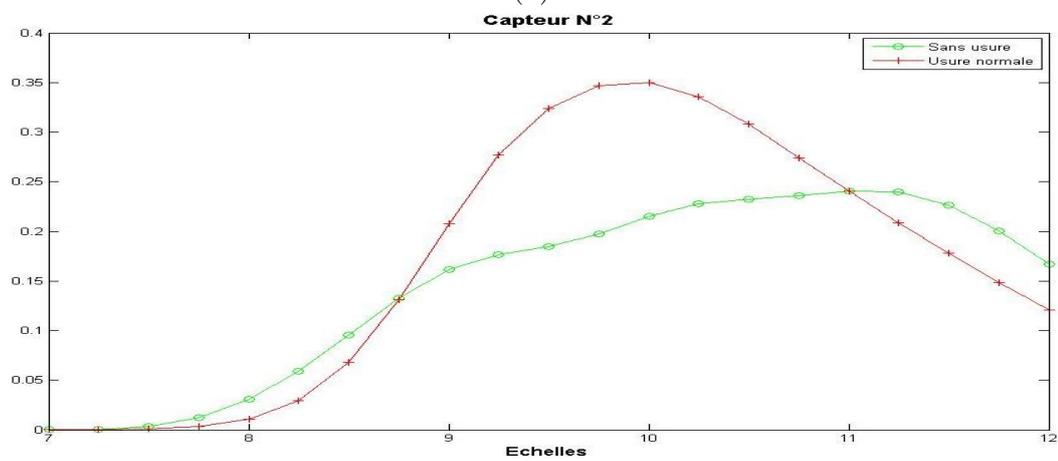


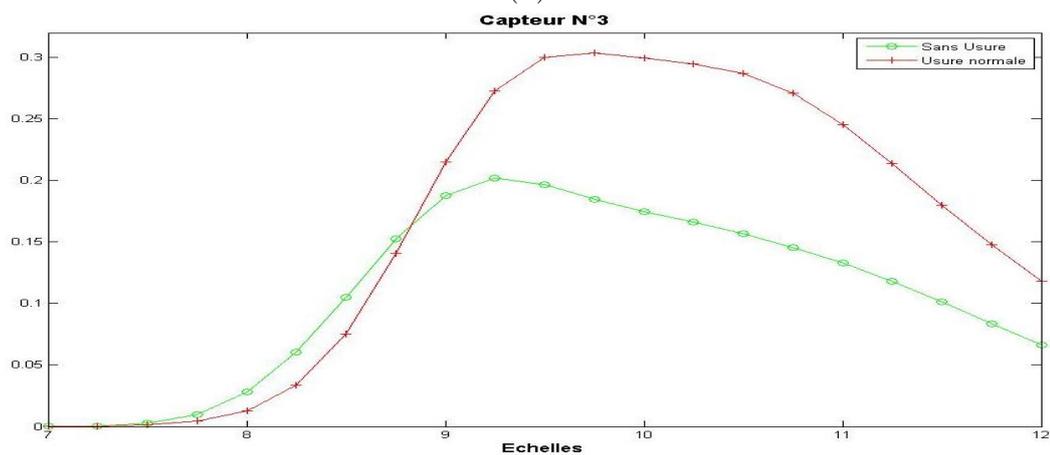
FIG. 5.7 – Vecteur moyen de 50 vecteurs singuliers des caractéristiques spectrales (échelles) pour le premier (a), le deuxième (b) et le troisième (c) capteur.



(a)

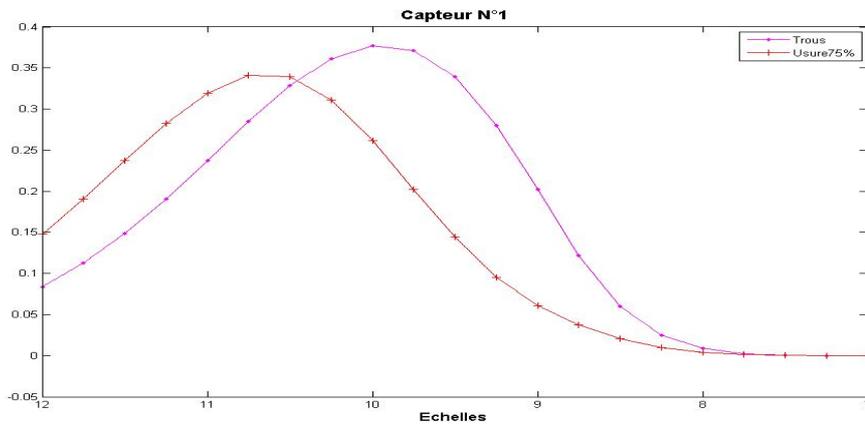


(b)

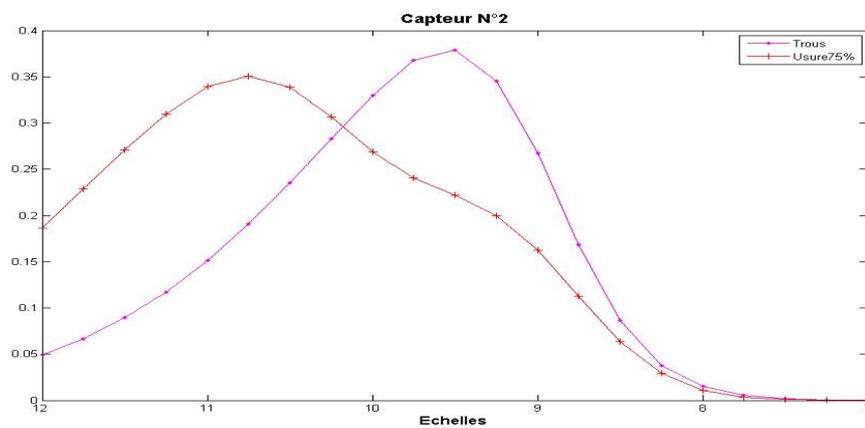


(c)

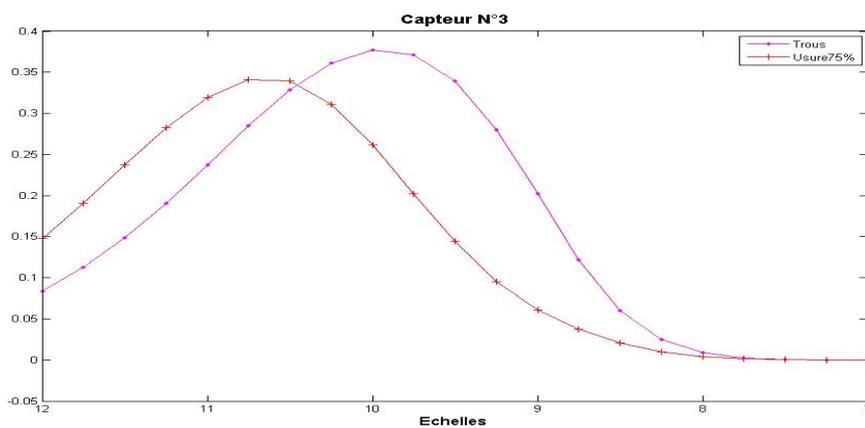
FIG. 5.8 – Vecteur moyen de 50 vecteurs singuliers des caractéristiques spectrales (échelles) pour les classes : sans défaut et usure normale pour le premier (a), le deuxième (b) et le troisième (c) capteur.



(a)



(b)



(c)

FIG. 5.9 – Vecteur moyen de 50 vecteurs singuliers des caractéristiques spectrales (échelles) pour les classes : sans défaut, trous et usure à 75% pour le premier (a), le deuxième (b) et le troisième (c) capteur.

défini et calculé comme suit

$$\rho = \frac{\langle Y.V_1 \rangle}{\|Y\|}. \quad (5.10)$$

où Y est la transformée en ondelettes continue (TOC) du vol de test. La matrice V est la matrice d'échelle et V_1 représente le premier vecteur singulier d'un vol normal, c'est à dire sa pseudo signature de puissance. Cependant, l'utilisation du premier vecteur singulier V_1 comme la pseudo-signature de puissance pour l'état du vol a donné une approximation très pauvre par rapport à la TOC. Ce résultat souligne la nécessité d'approximer la transformée en ondelettes continue avec une somme de termes séparables. Une deuxième alternative a été ensuite proposée pour améliorer les PPS en envisageant des approximations avec des termes supplémentaires. Cette approche peut être considérée comme la définition d'un "sous-espace de signatures". Pour simplifier les calculs, seulement les deux premiers termes d'approximation ont été utilisés, mais cela n'empêche pas l'extension de la méthode à plus de termes.

Supposons que la TOC peut être exprimée sous la forme

$$c_\psi^x(a, b) = \sigma_1 s_1(a) r_1(b) + \sigma_2 s_2(a) r_2(b).$$

À n'importe quel instant b_0 nous pouvons constater que la fonction $c_\psi^x(., b_0)$ est une combinaison linéaire des deux fonctions $s_1(.)$ et $s_2(.)$.

Dans un état normal, sans défaut, la TOC doit appartenir au sous-espace de signatures correspondant à un état normal. Ainsi, le calcul de la distance à ce sous-espace peut être utilisé pour mesurer les écarts. Plus l'écart est important plus l'état dans lequel évolue le système s'éloigne de la normale.

Prenons maintenant les deux vecteurs singuliers V_1 et V_2 issus de la décomposition en valeurs singulières de la TOC d'un état normal. Ces deux vecteurs forment une base orthogonale du sous-espace des signatures. Pour chaque vecteur Y la projection de Y sur le plan des signatures de l'état normal est de la forme $\alpha_1 V_1 + \alpha_2 V_2$. Les coefficients sont déterminés comme des projections sur les vecteurs de base peuvent être déterminés à partir de

$$\langle Y.V_1 \rangle = \alpha_1, \text{ et } \langle Y.V_2 \rangle = \alpha_2.$$

La distance entre Y et le plan des signatures de l'état normal est ensuite obtenue à l'aide de la formule suivante

$$d = \sqrt{\|Y\|^2 - (\alpha_1^2 + \alpha_2^2)}.$$

Les auteurs de cet article soulignent l'importance des échelles et du choix du type d'ondelettes utilisée. La transformée en ondelettes qui a donné de bons résultats sur leurs tests est de type Daubechies Db8 avec 128 échelles.

5.5.4 Fonction de distribution de probabilité des vecteurs singuliers

Définition

Une autre approche a été présentée par Groutage dans [106] pour l'extraction de caractéristiques des signaux non-stationnaires. Cette approche est basée sur l'approximation de la RTF en un nombre fini de rectangles. Dans la RTF, la région possédant une densité d'énergie uniforme est représentée par un rectangle. Ces rectangles sont identifiés par des vecteurs de 5 éléments $[\bar{t}; \bar{f}; \hat{t}; \hat{f}; \hat{\sigma}]$, où \bar{t} et \hat{t} représentent la localisation et la durée en temps respectivement, \bar{f} et \hat{f} représentent la localisation et la largeur en fréquence dans la RTF ; et $\hat{\sigma}$ représente l'importance du rectangle dans la composition de la RTF. La position et les dimensions des rectangles sont calculés à partir des premiers et seconds moments de la fonction de densité extraite des vecteurs singuliers de la RTF. Cependant cette approche, selon H. Hassanpour, M. Mesbah et B. Boashash, possède trois inconvénients [108] :

1. elle utilise un nombre fixe de caractéristiques (rectangles) pour caractériser une région de la RTF. Or, en utilisant un nombre limité de rectangles, la méthode peut ne pas être adéquate pour identifier la totalité des caractéristiques des RTF ;
2. s'il y a plus d'un maximum local dans la fonction de densité, le premier et le second moments de la fonction de densité ne peuvent pas montrer la position et la largeur du maximum local. Cette technique peut être efficace si :
 - (a) le plan temps-fréquence est simple pour l'approximation par des rectangles en nombre limité ;
 - (b) la densité d'énergie du signal n'est pas uniformément concentré à différentes localisations de la RTF.
3. il peut y avoir plus d'un vecteur singulier de gauche et de droite qui sont nécessaires pour l'approximation de la RTF. Donc, le moment extrait à partir de seulement un vecteur singulier de gauche et un de droite n'est pas suffisant pour trouver la localisation, la durée, et la bande de fréquences dans le plan temps-fréquence. Par conséquent, dans une autre référence [105], les auteurs améliorent leur approche en respectant les trois critères cités précédemment.

Formulation

Posons CWT_k la matrice des coefficients d'ondelettes correspondant à un couple de galets CGk . En appliquant une décomposition en valeurs singulières, cette matrice peut être écrite comme suit

$$CWT_k = U\Sigma V^T.$$

où $U(M \times M)$, $\Sigma(M \times N)$ et $V(N \times N)$. La fonction de distribution de probabilité (FDP) peut être composée à partir des colonnes individuelles des matrices U et V associées aux vecteurs singuliers des caractéristiques temporelles et spectrales

(échelles). Par exemple, la FDP correspondant à la première colonne de la matrice U , f_{U1} , est donnée par

$$f_{U1} = \{u_{11}^2, u_{12}^2, \dots, u_{1M}^2\},$$

où u_{1i} représente le i ème élément de $U1$ et $\sum_{i=1}^M u_{1i}^2 = 1$. La fonction de distribution de probabilité correspondante est calculée comme suit

$$F_{U1} = \{v_1, v_2, \dots, v_M\},$$

$$\text{où } v_j = \sum_{i=1}^j u_{1i}^2, \text{ pour } j = 1, \dots, M.$$

Contrairement à la première, cette approche se base sur les deux matrices U et V pour approximer la matrice TOC. Pour illustrer cette sélection de caractéristiques, prenons la DVS d'un ensemble de matrices temps-fréquence correspondant aux signaux vibratoires avec et sans défauts.

Les figures 5.10 et 5.11 présentent les fonctions de distribution calculées à partir des deux premiers vecteurs fréquentiels (échelles) et temporels pour respectivement un couple de galet sans défaut et un couple de galet avec défaut (usé à 75%). Nous avons comparé ensuite les DFP calculée pour l'ensemble des classes : sans défaut, usure à 75%, usure à 100% et Trous deux à deux. Le résultat est présenté dans les figures 5.12 et 5.13. Nous pouvons constater sur ces figures que sur certaine bande d'échelles, les différentes classes sont clairement séparables. Nous pensons qu'une sélection basée sur les échelles peut améliorer nettement les résultats de la classification. Les deux dernières méthodes sont attractives et seront testées ultérieurement dans nos travaux de recherches.

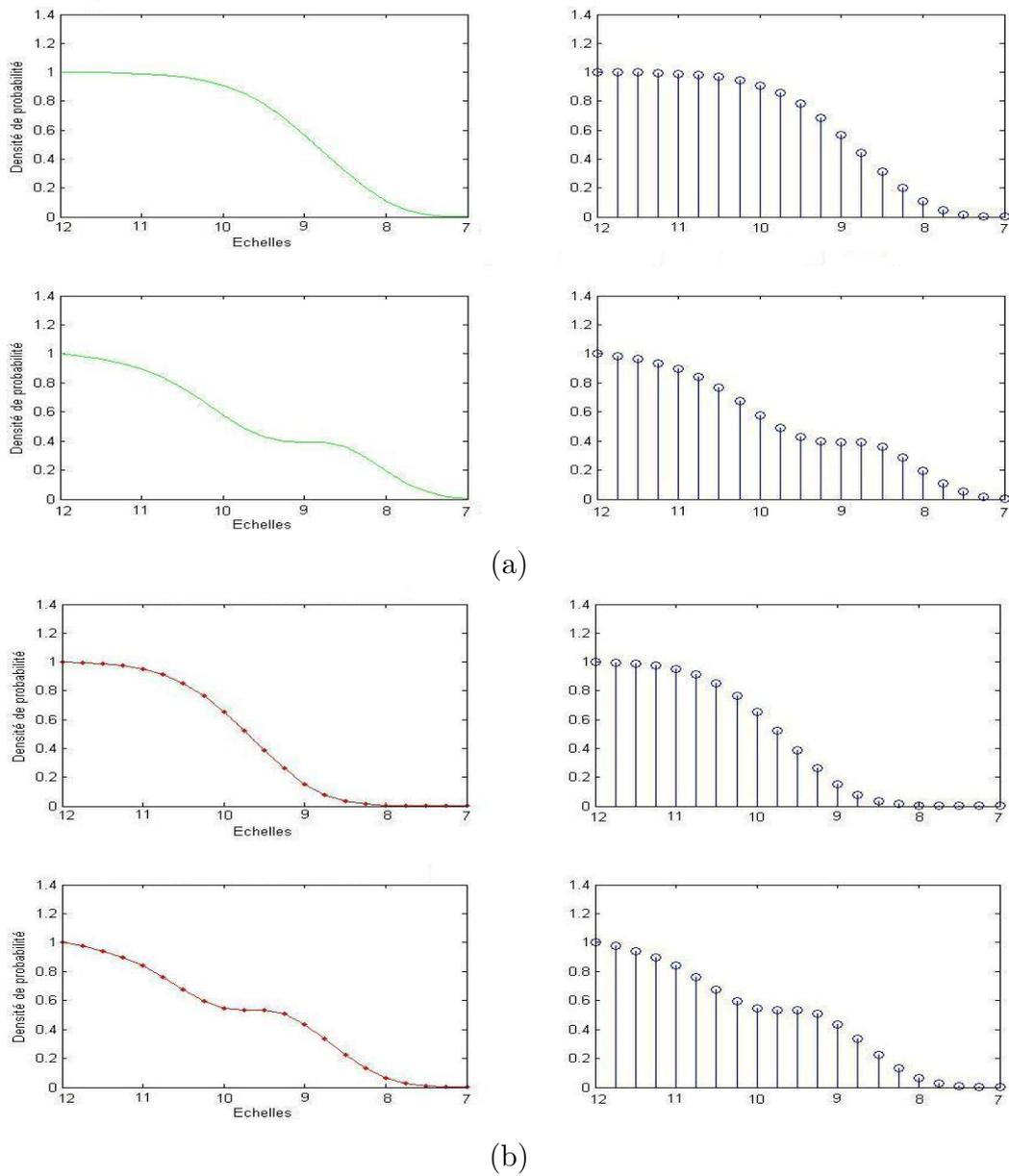


FIG. 5.10 – La FDP calculée à partir des deux vecteurs singuliers des caractéristiques spectrales (échelles) pour (a) un galet sans usure et un galet utilisé à 75% (b).

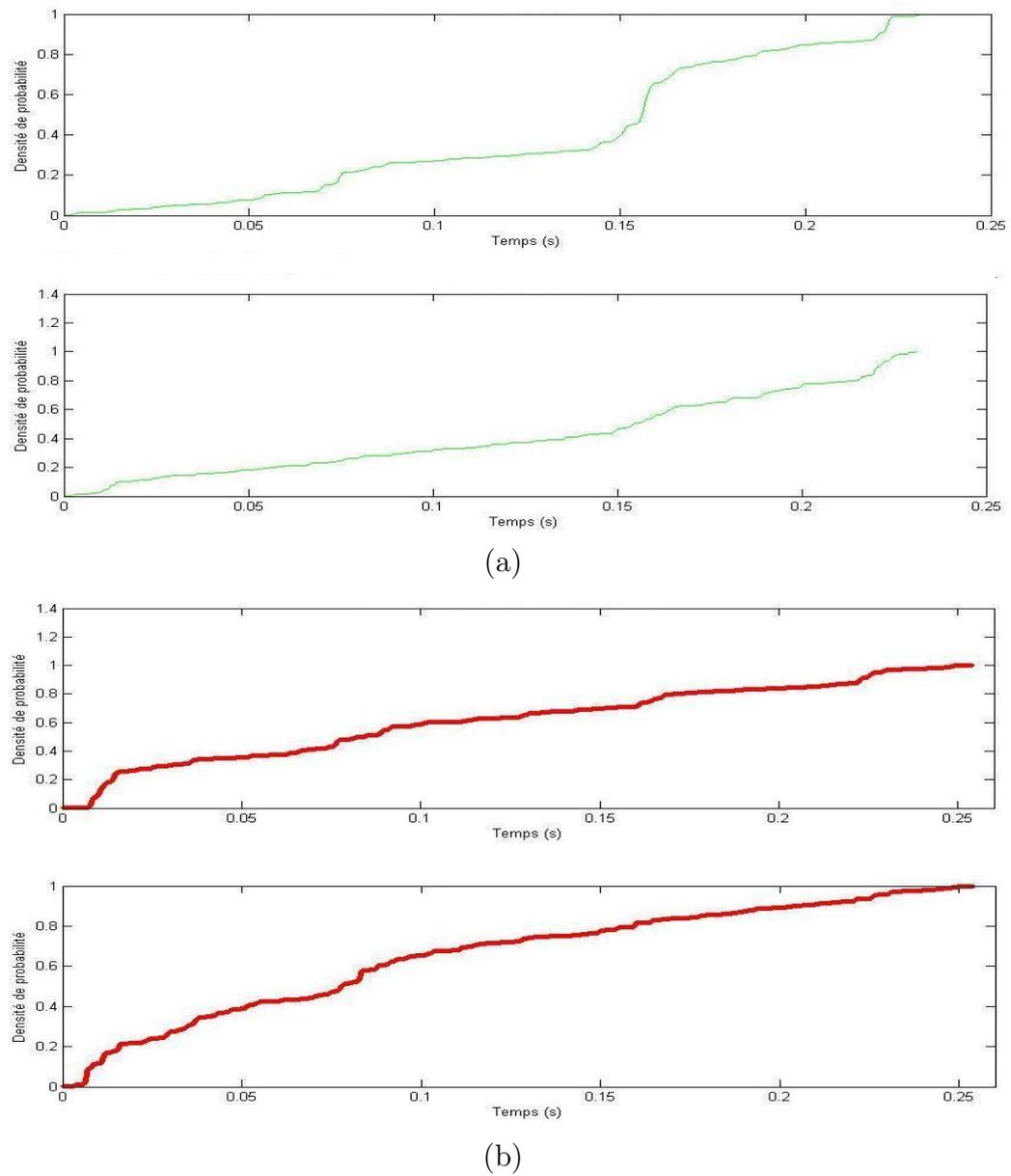


FIG. 5.11 – La FDP calculée à partir des deux vecteurs singuliers des caractéristiques temporelles pour (a) un galet sans usure et un galet usé à 75% (b).

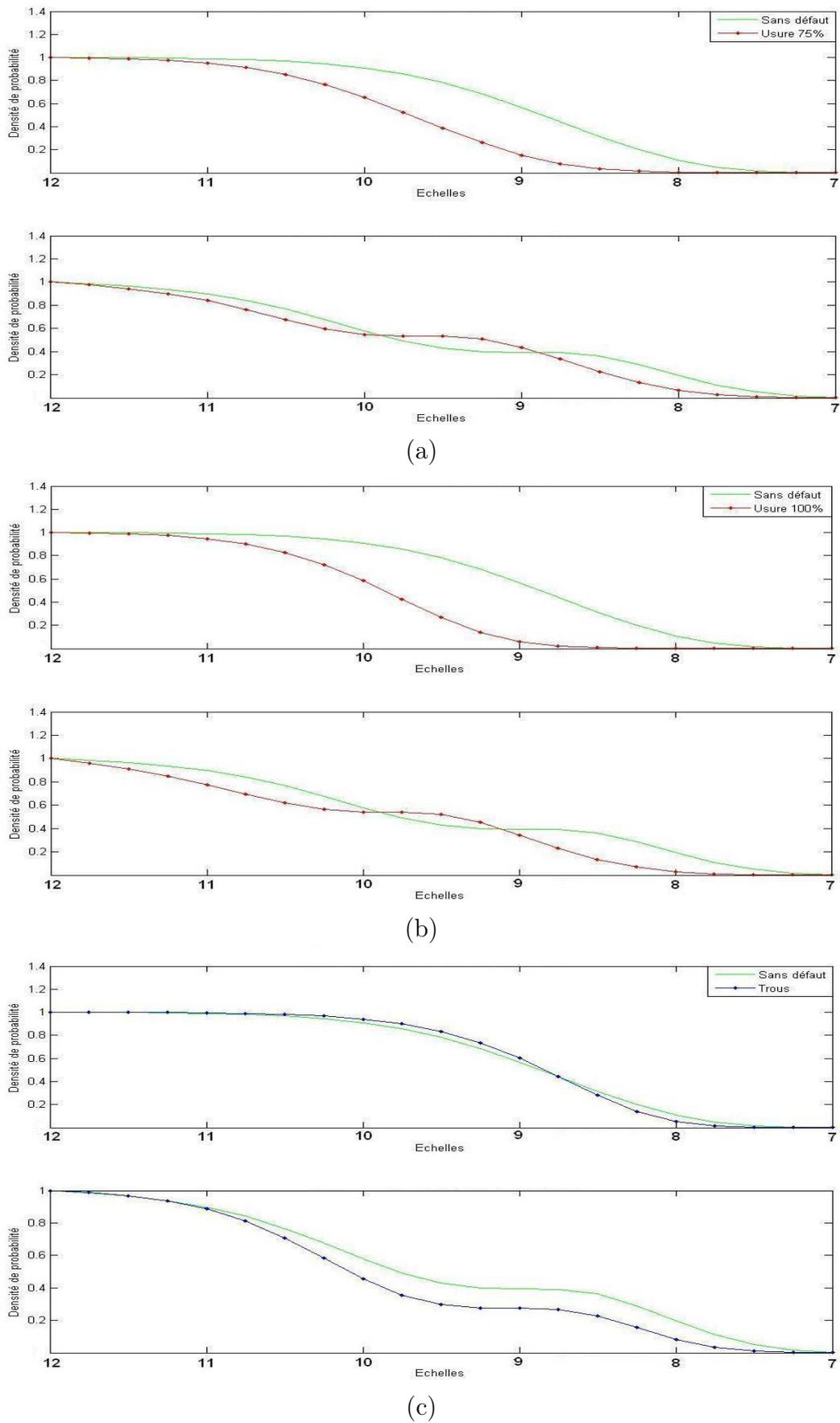


FIG. 5.12 – Comparaison des deux FDPs calculées à partir des deux premiers vecteurs singuliers des caractéristiques spectrales (échelles) pour la classe sans défaut avec celle de la classe d'usure à 75% (a), la classe d'usure à 100% (b) et la classe Trous (c).

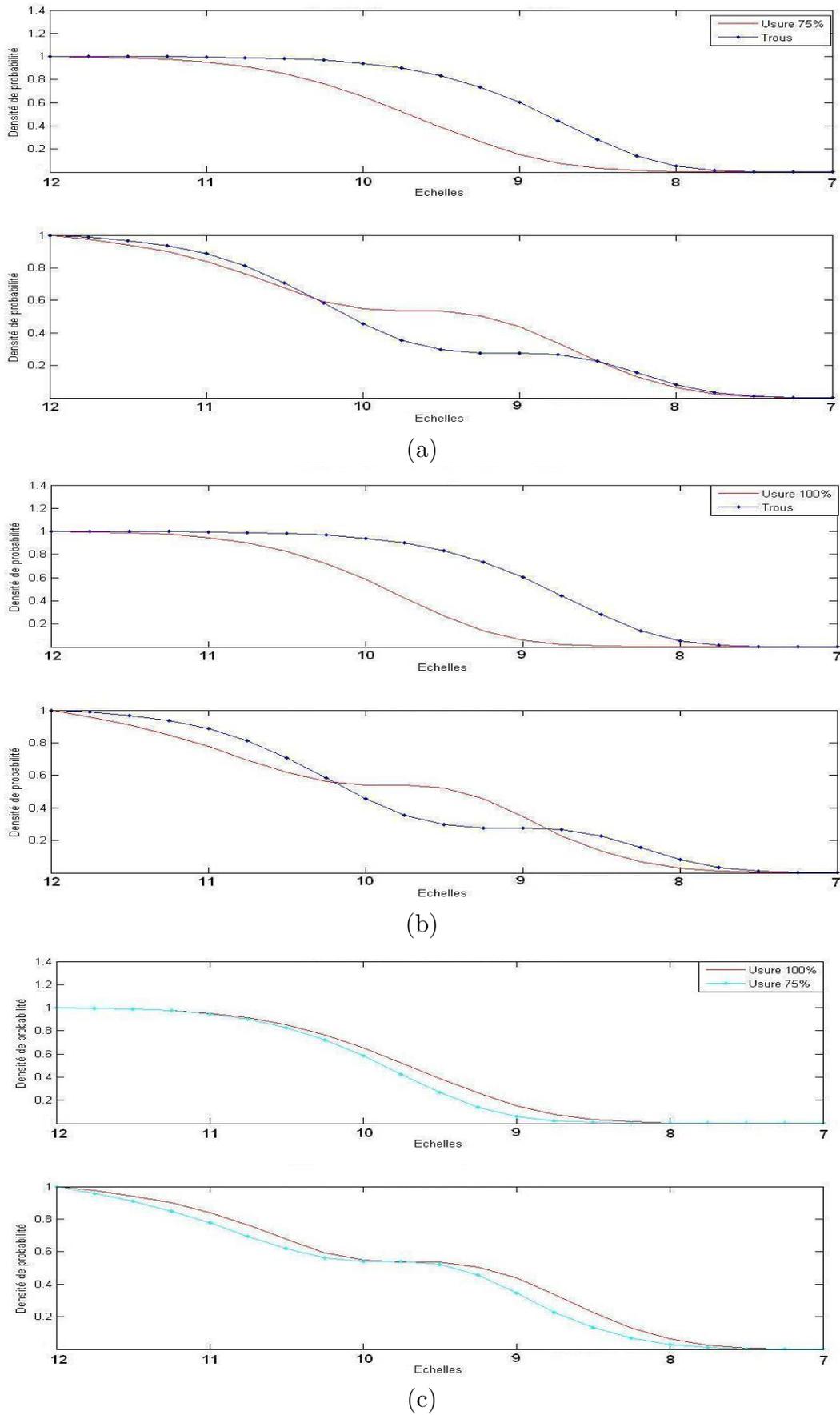


FIG. 5.13 – Comparaison entre les FDPs de l'ensemble des trois classes comparées deux à deux. La classe d'usure à 75% et la classe Trous (a), 100% et Trous (b) et la classe d'usure à 100% et l'usure à 75% (c).

Conclusion

L'objectif de ce chapitre était de trouver la meilleure façon pour réduire la dimension des données associées à chaque couple de galets. Ainsi trouver une signature de taille réduite qui le caractérise, par conséquent le différencie des autres couples de galets appartenant aux autres classes. Pour cela nous avons, dans un premier temps, passé en revue quelques méthodes de sélection de variables pour la réduction de dimension, (cf. 5.2). Les différentes approches de sélection se distinguent par l'algorithme de recherche et le critère d'évaluation utilisés. Les nombreuses études comparatives effectuées dans ce domaine affirment qu'il n'existe pas de méthode idéale. La plupart de ces méthodes s'appuient sur des heuristiques qui parcourent de manière limitée l'espace de toutes les combinaisons de variables possibles, ce qui ne garantit pas toujours de trouver la solution optimale. Les résultats dépendent également du discriminateur utilisé après l'étape de sélection : un sous-ensemble de variables qui est optimal pour un discriminateur donné ne l'est pas nécessairement pour un autre discriminateur.

Dans un second temps, nous avons présenté une deuxième technique de réduction de dimension, l'approche d'extraction de caractéristiques, (cf. 5.3). Nous nous sommes concentrés sur la décomposition en valeurs singulières. Cette dernière repose sur l'étude des matrices U , V , et Σ issues de l'application de la DVS à l'ensemble des scalogrammes obtenus grâce à l'étape de paramétrisation des signaux (cf. 4). La DVS propose un moyen très intéressant de réduction de variables. Pour l'étape de la classification, nous pouvons explorer les résultats de la DVS de deux manières : soit effectuer une classification en se basant sur les premières valeurs singulières σ_p , soit en utilisant les vecteurs singuliers de gauche et de droite correspondants aux premières valeurs singulières σ_p . L'étude de ces vecteurs singuliers donne à première vue un outil performant pour la séparation des classes. Cette approche est basée sur la définition de signatures appelées "pseudo-signatures de puissance" pour les classes de signaux non-stationnaires indépendamment de la longueur du signal (la durée). Nous avons aussi décrit brièvement deux approches qui se basent sur ces signatures PPS de deux manières différentes pour la détection et l'identification des défauts. La première technique décrite dans la partie 5.5.3 propose, dans un premier temps, la définition d'un coefficient de corrélation entre les pseudos signatures d'un état normal et anormal. Dans un deuxième temps, le même auteur, propose le calcul de la distance entre la pseudo signature d'un état normal et le signal test pour identifier l'état dans lequel évolue le système. Une deuxième technique basée sur la fonction de distribution de probabilité des vecteurs singuliers (FDP), proposée dans [188, 105, 108], a été ensuite détaillée dans la partie 5.5.4. Cette dernière se base sur les deux matrices de temps et de l'échelle pour construire les signatures. Les différentes signatures issues de cette étape de réduction de dimension sont ensuite utilisées pour la classification. La réussite de cette dernière repose entièrement sur la qualité de ces signatures.

CLASSIFICATION SUPERVISÉE

6.1 Introduction

Nous consacrons ce dernier chapitre à la phase finale du processus de diagnostic, la classification. Nous nous intéressons plus particulièrement à la classification supervisée à partir des bases de données construites lors de l'étape d'acquisition, (cf. 3). De nombreux tests ont été menés sur plusieurs jeux de données, valeurs et vecteurs singuliers, pour montrer l'efficacité de chacune des deux approches : avec les valeurs singulières et avec les vecteurs singuliers. Nos travaux relatifs à la définition des signatures obtenus à partir des vecteurs singuliers sont basés seulement sur le premier vecteur des caractéristiques spectrales (échelles) selon la méthode décrite dans le précédent chapitre, (cf. 5.5.2). Nous nous sommes concentrés sur les classifieurs de type SVM pour effectuer l'étape de classification. Toutefois, nous confortons notre choix en comparant plusieurs méthodes de classification. Leurs performances ont été estimées à travers les taux de bon classement et d'autres mesures classiques d'erreurs. Plusieurs types de classifieurs ont été mis en oeuvre et comparés : les MLPs (cf. 2.5.3), les RBFs (cf. 2.5.4), les arbres de décision (Id3 et J48), (cf. 2.4.2), le modèle logistique et les SVM 2.6. Enfin pour améliorer la précision des classifieurs, nous avons utilisé deux méthodes de construction d'ensembles de classifieurs : la méthode adaboost (cf. 2.5.5) et le prédicteur bagging proposée par Leo Breiman dans [36].

L'étape de classification a été mise en oeuvre à l'aide du logiciel de data mining Weka[272]. Weka est un open source très riche programmé en java et regroupe un ensemble d'outils qui peuvent être utilisés pour la classification, la régression et le clustering. L'utilisation de cet outil nous a permis de comparer plusieurs méthodes de classification pour pouvoir au final garder la meilleure pour notre système de diagnostic. Comme on vise la performance, pour la partie classification par les SVMs nous avons choisi d'utiliser la boîte à outils SVM développée au laboratoire PSI sous Matlab [164]. Deux types de noyaux ont été testés avec différentes valeurs pour les paramètres afin de choisir ceux qui donnent les meilleures performances. Les résultats obtenus sont détaillés dans la partie 6.6.

Le chapitre est divisé en cinq grandes parties. La première partie rappelle brièvement la démarche globale de diagnostic suivie dans notre travail, (cf. 6.2). Après une brève introduction de l'ensembles des mesures d'erreurs utilisées pour mesurer la per-

formance des classifieurs, la deuxième partie sera divisée en trois principaux sous paragraphes. À défaut d'expérience et de méconnaissance des classes de défauts réelles nous avons utilisé l'ensemble des classes : "Sans défaut, Usure100%, Usure75%, Trous et l'usure Patatoïde". Nous avons effectué une classification 1 – *Contre* – 1 pour comparer les défauts entre eux et mieux comprendre leurs comportements. De façon un peu naïve, nous avons pensé que procéder ainsi va nous permettre de regrouper les semblables et voir en éliminer les fausses classes de défauts. Toutefois, l'expérience nous a montrée qu'il y a une différence entre les passages du tramway dans le sens ME1 et le sens ME2. Ce constat nous a obligé de prendre en compte séparément les signaux résultant de chaque sens. Dans les deux cas les deux configurations sont considérées, la configuration C2 - où les défauts sont montés sur les essieux avants (galets directeurs) - et la configuration C1 - où les défauts sont montés sur les essieux arrières (galets suiveurs). Par conséquent, l'étape de la classification a été mise en oeuvre pour chaque sens de marche. Les deux configurations C1 et C2 ont été prises, dans un premier temps, séparément, (cf. 6.3.2(A)), et mélangées dans un deuxième temps, (cf. 6.3.2(B)). Nous avons vérifié, dans le dernier paragraphe de cette partie, si le sens de la marche avait réellement un impact sur les résultats de la classification, (cf. 6.3.2(C)). Dans ces trois cas, seules les valeurs singulières ont été utilisées en entrée du classifieur.

Lors d'une réunion effectuée le 06 Juillet 2006, l'exploitant - la T2C - a précisé, qu'à ce stade du projet et sans connaissance a priori des défauts pouvant surgir, l'idéal est d'avoir un outil de diagnostic et d'aide à la décision qui permet de détecter la présence d'un défaut. Par conséquent, nous avons traduit ce besoin en diagnostic en deux phases. Détecter l'existence du défaut dans une première phase, et après l'identification du type de défaut dans la deuxième phase. L'ensemble des classes de défauts et la nouvelle démarche de diagnostic suivie sont présentés dans la troisième partie du chapitre, (cf. 6.4).

Entre temps, d'autres travaux de recherches ont montrés l'intérêt des vecteurs singuliers issus de la SVD, dans le domaine des signaux vibratoires, par rapport aux valeurs singulières. Cherchant à améliorer notre méthode de diagnostic pour trouver la meilleure signature représentant un défaut, nous avons essayé d'utiliser les vecteurs singuliers pour la classification et en tirer les avantages de cette méthode. Ainsi, nous avons comparé dans la quatrième partie, (cf. 6.5), les résultats de la classification obtenus à l'aide des différents classifieurs en utilisant, dans un premier temps, les valeurs singulières en entrée du classifieur, (cf. 6.5.1), et dans un deuxième temps, les pseudo-signatures de puissance, (cf. 6.5.2), définies dans la partie 5.5.2.

Enfin, nous avons consacré la dernière partie du chapitre à l'étude des classifieurs de type SVM [260]. Leur performance a été comparée à celle des classifieurs utilisés précédemment : MLP, RBF.

6.2 Démarche utilisée pour le diagnostic

Rappelons que le détecteur SYSCOGA sur l'ADV, (cf. 3.2.2), est constitué de trois capteurs permettant chacun de mesurer les accélérations lors du passage du Translohr. Ce dernier, circule dans les deux sens ME1 avec l'essieu ME1 en tête et ME2 avec l'essieu ME2 en tête. Deux montages de couples de galets, configurations C1 et C2 tableau 6.1, ont été utilisés pour les deux sens de marche. L'ensemble des défauts étudiés sont montés sur l'essieu arrière dans la configuration C1 et sur l'essieu avant dans la deuxième configuration sauf les deux défauts "Usure100%" et "Trous". Le premier est toujours monté sur un essieu arrière (ME1_{AR} et MI2_{AR}) par contre le second est monté sur un essieu avant (ME2_{AV} et ME1_{AV}).

		Configuration C1		Configuration C2	
		Gauche	Droite	Gauche	Droite
ME1	Avant	Sans défauts	Sans défauts	Trous	Sans défauts
ME1	Arrière	Sans défauts	Usure100%	Sans défauts	Sans défauts
MI1	Avant	Sans défauts	Sans défauts	Patate	Sans défauts
MI1	Arrière	Sans défauts	Usure75%	Sans défauts	Sans défauts
MI2	Avant	Sans défauts	Sans défauts	Usure75%	Sans défauts
MI2	Arrière	Sans défauts	Patate	Usure100%	Sans défauts
ME2	Avant	Sans défauts	Trous	Sans défauts	Sans défauts
ME2	Arrière	Sans défauts	Sans défauts	Sans défauts	Sans défauts

TAB. 6.1 – Configurations C1 et C2 utilisées sur site.

Reprenons le schéma global du processus de diagnostic, partant de l'étape d'acquisition jusqu'à la classification, figure 6.1 :

1. pour chaque passage du Translohr, le système d'acquisition fournit trois signaux vibratoires. Chacun résulte de la somme des deux axes latéral et vertical d'un capteur, figure 6.1(1) ;
2. ensuite, une transformée en ondelettes continue est appliquée sur chaque signal (module traitement du signal), figure 6.1(2) ;
3. une fois calculée, nous procédons au découpage temporel et fréquentiel de la matrice des coefficients d'ondelettes obtenue. À l'aide des dates de références fournies par les deux cellules photoélectriques, (cf. 3), nous appliquons le découpage temporel et fréquentiel, décrit dans la partie 4.3.2, pour trouver la position exacte des 8 couples de galets sur le signal complet. En conséquence, nous obtenons, pour chaque capteur, l'ensemble des segments $S_i = \{SCG1, \dots, SCGk\}$ correspondants aux couples de galets CGk , avec $k = 1, \dots, 8$. À l'issue de cette étape, à un couple de galets CGk correspondent 3 matrices de coefficients d'ondelettes $CWTCap_i$, $i = 1, 2, 3$ représente les numéros des capteurs, figure 6.1(3) ;
4. avant d'entamer l'étape de classification une phase de réduction de dimension par décomposition en valeurs singulières est appliquée sur chaque matrice $CWTCap_i$. Seules les premières valeurs singulières σ_p sont retenues pour la classification. Nous récupérons à la fin de cette étape, pour chaque CGk , trois vecteurs formes x_{1k} , x_{2k} et x_{3k} correspondant respectivement aux capteurs 1, 2 et 3, avec $k = 1, \dots, 8$, figure 6.1(4) ;

5. en se basant sur ces vecteurs formes en entrée, le module de classification fournit, pour un couple de galets CGk , l'ensemble de décisions $DCap_iCGk$ avec $k = 1, \dots, 8$ et $i = 1, \dots, 3$, figure 6.1(5) ;
6. la décision finale Df pour un CGk est la fusion des trois décisions issues des trois capteurs $Df = \bigcup \{DCap_1CGk, DCap_2CGk, DCap_3CGk\}$, figure 6.1(6). Cela nécessite l'emploi d'une stratégie permettant de fusionner les décisions prises par les trois capteurs, pour un galet donné CGK , pour parvenir à une décision finale. La procédure de vote majoritaire a été choisie pour nos traitements. Les ambiguïtés, dans notre cas, sont prises comme des erreurs de classification. Si par exemple, deux capteurs sur les trois jugent que le premier couple de galets est "Sans défaut", alors cette classe a reçu 2 votes, et elle l'emporte sur les autres classes. Ainsi le premier couple de galets est classé Sans défaut.

Les étapes de 4 à 6 sont effectuées pour les k couples de galets. Les deux dernières étapes, 5 et 6, sont illustrées dans ce présent chapitre.

Les usures ou les défauts utilisés durant les essais ont été élaborés manuellement. L'ensemble des classes testées sont :

- classe de couples de galets sans défauts (neuf ou non usé) ;
- l'usure parallèle qui représente un gommage du bandage : à 75% et à 100% ;
- l'usure aléatoire qui regroupe les trous et l'usure papatoïde.

La base de données complète contient un ensemble de signatures propres aux galets sans défauts et celles correspondant aux 4 défauts : Usure100%, Usure75%, Trous et l'usure Patatoïde. Chaque défaut est représenté par une matrice de coefficients d'ondelettes CWT ($N \times M$), constituée de 677376 variables si le défaut est monté sur un couple de galet extrême et de 733824 variables s'il est monté sur un essieu intermédiaire. Ce nombre trop élevé de paramètres est ensuite réduit grâce à l'application de la DVS sur les matrices CWT. Chaque défaut est représenté, comme mentionné précédemment, par les cinq première valeurs singulières. Deux sens de marches ME1 et ME2 avec les deux configurations C1 et C2 ont été testés. Le nombre d'exemples utilisés pour les 5 classes est identique. Quatre bases de données sont construites et utilisées dans le paragraphe suivant :

- ME1C1 (400×5) ;
- ME1C2 (250×5) ;
- ME2C1 (255×5) ;
- ME2C2 (250×5).

Nous avons comparé pour un sens donné ME1 ou ME2, dans un premier temps, le comportement des deux à deux. Nous exposons quelques résultats de la classification multi-classe, en adoptant l'approche "1-vs-1", obtenus avec les différentes méthodes utilisées par classe, par sens et par configuration, (cf. 6.3.2(A) et (B)). Pour la partie 6.3.2(A) les deux configurations ont été prises séparément. Le comportement d'un même défaut a été testé en étant monté sur un essieu arrière (couple suiveur), avec C1, ensuite sur un essieu avant (couple directeur), avec C2. Dans le paragraphe 6.3.2(B), nous avons mélangé les deux configurations C1 et C2 pour voir si cela influe sur le taux global de la classification.

Enfin, les défauts à détecter en priorité, leurs formes et leurs types ont été grossièrement définis par l'exploitant. Nous discutons ce point en illustrant la stratégie de diagnostic adoptée dans la partie 6.4.

6.3 Analyses et résultats

La technique la plus populaire pour l'estimation de l'erreur de généralisation est la validation croisée qui est utilisée indépendamment de la nature de la machine d'apprentissage utilisée. Pour nos tests nous avons utilisé une procédure simple d'estimation de l'erreur de généralisation dite k -fold et décrite dans la partie 2.3.3. Elle consiste à diviser l'ensemble des données en k sous ensembles mutuellement exclusifs de taille approximativement égale. L'étape d'apprentissage est effectuée en utilisant $k - 1$ sous-ensembles et le test est effectué sur le sous-ensemble restant. Cette procédure est répétée k fois et chaque sous-ensemble est utilisé une fois pour le test. La moyenne des k taux d'erreur obtenus estime l'erreur de généralisation. Dans ce qui suit, nous présentons les résultats de l'identification de l'état des galets de guidage (classification) en utilisant la procédure 5-fold. Nous avons utilisé l'approche de classification multi-classes "1-contre-1", (cf. 6.3.2).

Les paragraphes suivants resument les performances de bonne classification obtenues pour chaque sous-classifieur. Pour tous les classifieurs, le taux de bonne détection est estimé par la procédure 5-fold. Pour chacun des sous-classifieurs sont indiqués les valeurs du taux de bonne classification (CC). Pour comparer leur performance nous avons utilisé des critères classiques comme le coefficient Kappa (κ), les mesures d'exactitude : les vrais et les faux positifs (TP et FP), les vrais et les faux négatifs (TN et FN), la précision et le rappel (Recall). L'ensemble de ces mesures sont détaillés en annexe, (cf. C.1 et C.2). Néanmoins, nous les introduisons brièvement dans le paragraphe suivant pour faciliter la compréhension du chapitre.

6.3.1 Performance des classifieurs

1. **Le taux de bonne classification** ($CC\%$) : c'est le nombre d'exemples bien classés en pourcentage du nombre total d'exemples ;
2. **Kappa statistic** (κ) : ce coefficient mesure le degré de concordance de deux ou de plusieurs juges. Le jugement représente la classe d'un exemple et les juges représentent le classifieur et la classe réelle de l'exemple. "L'accord/désaccord" entre les juges se lit directement dans la matrice de confusion. Ce coefficient est maximal, $\kappa = 1$, quand les deux jugements sont les mêmes ;
3. **Matrice de confusion** : pour une classe, un classifieur, et un exemple donné, quatre cas peuvent se présenter :
 - l'exemple est de cette classe, et le classifieur ne se trompe pas : c'est un **vrai positif** (TP) ;
 - l'exemple est de cette classe, mais le classifieur se trompe : c'est un **faux négatif** (FN) ;

- l'exemple n'est pas de cette classe, mais le classifieur la lui attribue quand même : c'est un *faux positif* (FP) ;
- l'exemple n'est pas de cette classe, et le classifieur ne le range pas non plus dans cette classe : c'est un *vrai négatif* (TN).

Ces quatre valeurs sont regroupées dans la matrice de confusion (coïncidence). Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée, et chaque ligne représente le nombre d'occurrences d'une classe réelle. À partir de la matrice de confusion on peut calculer d'autres mesures comme par exemple le TP Rate et FP Rate ;

4. **Rapport des vrais positifs ou la sensibilité** (TP Rate) : c'est le rapport entre le nombre des exemples bien classés et le nombre total d'éléments qui devraient être bien classés. Il correspond à 1 si les exemples d'une classe donnée sont bien classés ;
5. **Rapport des faux positifs ou la spécificité** (FP Rate) : il correspond à $1 - \text{FP Rate}$;
6. **La précision** : c'est le rapport entre le nombre de vrais positifs et la somme des vrais positifs et des faux positifs. Si elle vaut 1, cela veut dire que tous les exemples classés positifs l'étaient vraiment ;
7. **Le rappel** : si tous les exemples positifs ont été trouvés, le rappel vaut 1 ;
8. **La quantité "F-Measure"** : elle permet de regrouper en un seul nombre les performances du classifieur pour une classe donnée ;
9. **l'indice de Youden** : il est égal à $(\text{sensibilit} + \text{spcificit} - 1)$. Si cette indice se rapproche du 1 donc le test (classification) est "efficace" ;
10. **le coefficient Q de Yule** : il mesure l'intensité de la liaison entre les deux variables (Classes/Signes). Il est égal à $Q = (TP * TN - FP * FN) / (TP * TN + FP * FN)$ et l'intensité est :
 - nulle si $Q = 0$;
 - négligeable si $Q \in [0.01 - 0.09]$;
 - légère si $Q \in [0.10 - 0.29]$;
 - modérée si $Q \in [0.30 - 0.49]$;
 - forte si $Q \in [0.50 - 0.69]$;
 - très forte si $Q \in [0.70 - 1]$.

Pour plus de détails sur le calcul de ces mesures le lecteur peut se référer au annexe C.

6.3.2 La classification multi-classe avec l'approche 1-vs-1

Quelque soit le sens et la configuration utilisées les sous-classifieurs de type MLP fournissent, dans la majorité des cas, les meilleurs résultats. Pour la classe sans défaut, chaque configuration comportait un mélange des deux cas, tableau 6.1. Dans la première configuration C1, il y avait 4 couples de galets sans défaut dont un seul

est suiveur ($ME2_{arr}$). Dans la C2, il y avait éventuellement 4 couples de galets sans défaut, par contre cette fois-ci, avec un seul couple directeur ($ME2_{av}$). Dans l'ensemble des exemples de classification, (cf. 6.3.2), nous n'avons pas fait la différence entre un couple de galets directeur (CGD) ou un couple de galets suiveur (CGS).

A - Les deux configurations C1 et C2 sont prises séparément

Dans ce premier cas, une base de données est construite par sens, pour une configuration, un capteur et une classe donnée. Nous avons constaté, qu'avec l'ensemble des sous-classifieurs, la classe "Usure 100%" est reconnue globalement à 100% indépendamment du sens de la marche. Cette classe est facilement détectable bien qu'elle soit toujours montée sur un essieu arrière, couple de galets suiveur dans les deux configurations. En fait, cette usure "maxi" (sans bandage) est tout de suite ressentie par le conducteur lui-même ainsi que les usagers. Par conséquent, le détecteur SYSCOGA ne trouve aucun problème pour la séparer des autres classes même d'une usure à 75%. Dans le sens ME1, avec les deux configuration C1 et C2, l'ensemble des classifieurs : le modèle logistique, MLP, RBF, Adaboost, Bagging et les arbres de décision (J48 et Id3) séparent l'usure 100% des autres classes - l'usure 75%, Trous et l'usure patatoïde - avec un taux global de classification égal à 100%, (cf. tableau 6.4), sauf dans certains cas.

La classe "Sans défaut" est aussi séparée de la classe "Usure 100%" à 100% sauf pour les deux classifieurs RBF et l'arbre de décision de type J48 qui fournissent, avec la configuration C1, 99.37% (1 erreur) de taux de reconnaissance, tableau ???. Adaboost et Bagging fournissent, avec la configuration C1, 98.75%. Enfin, les arbres de décisions de type Id3 fournissent 97.5% de taux de reconnaissance. Dans le sens ME2, avec la configuration C2, le classifieur de type MLP donne les meilleurs résultats. Il sépare nettement la classe d'usure 100% des autres classes avec un taux de classification égal à 100%.

Lors des tests, nous avons remarqué qu'un simple classifieur de type arbres de décision donne de bon résultats. Ce dernier est simple à gérer et à mettre en oeuvre, par contre sa complexité dépend fortement de celle du problème posé. Néanmoins, l'objectif de ce classifieur est de trouver le/les test(s) à effectuer sur les valeurs singulières ($\sigma_1, \sigma_2, \dots, \sigma_p$). Nous pouvons ainsi nous baser sur l'ensemble des tests définis pour affecter un vecteur forme donné à la bonne classe d'appartenance. Prenons un exemple de classification du couple de classes (Sans défauts, Usure 100%) effectuée à l'aide d'un arbre de décision de type J48. Dans le sens ME1 avec la configuration C2, pour les deux premiers capteurs, un simple test appliqué sur la première valeur singulière σ_1 suffit pour séparer les deux classes avec un taux de reconnaissance égal à 100%. L'utilisation des arbres de décision facilite largement la tâche de la classification. Cependant, de telles approches ne permettent pas la gestion du caractère évolutif d'un état normal ou anormal du système. Ainsi ce type d'approche n'est pas approprié pour notre cas du fait que le système de guidage est un système dynamique, évolutif. Un comportement jugé anormal, par le système de diagnostic, peut être dû à un comportement normal du système de guidage (l'appui, l'emplacement, sens, etc.).

L'expert "Usure 100%/Usure Patatoïde" fournit pour l'ensemble des classi-

Classification Un-contre-Un (1vs1)							
		Les classifieurs					
		Logistic		MLP		RBF	
Classes	Configurations	1	2	1	2	1	2
		Usure 100%	100	100	100	100	99.37
	Usure 75%	84.37	93	86.87	94	90	94
	Trous	91.25	92	91.25	90	90.62	85
	Patatoïde	81.25	95	88.12	94	76.87	93

		Les classifieurs							
		Adaboost		Bagging		J48		Id3	
Classes	Configurations	1	2	1	2	1	2	1	2
		Usure 100%	98.75	100	98.75	100	100	99	99.37
	Usure 75%	92.5	95	90.62	94	90	95	70.62	95
	Trous	90	89	91.87	86	93.12	85	87.37	86
	Patatoïde	76.87	93	83.75	94	83.75	93	66.25	94

TAB. 6.2 – Taux global de classification pour la classe Sans défaut dans le sens ME1.

		Les classifieurs					
		Logistic		MLP		RBF	
Classes	Configurations	1	2	1	2	1	2
		Usure 100%	98.04	100	97.06	100	97.06
	Usure 75%	86.28	97	85.3	95	83.33	96
	Trous	93.14	91	93.14	91	92.16	92
	Patatoïde	100	99	100	100	100	98

		Les classifieurs							
		Adaboost		Bagging		J48		Id3	
Classes	Configurations	1	2	1	2	1	2	1	2
		Usure 100%	97.06	100	97.06	99	98.04	99	74.51
	Usure 75%	74.51	93	83.3	96	71.5	93	67.6	96
	Trous	93.14	92	92.16	91	92.16	93	86.3	85
	Patatoïde	98.04	100	99.02	100	99.02	99	98.04	96

TAB. 6.3 – Taux global de classification pour la classe Sans défaut dans le sens ME2.

fieurs, dans le sens ME1, 100% de taux de reconnaissance. Ainsi, qu'elle soit montée sur un couple de galets suiveur (l'essieu $MI2_{ARR}$), dans la configuration C1, ou sur un couple de galets directeurs (l'essieu $MI1_{AV}$), dans la configuration C2, l'usure Patatoïde est facilement séparable de l'usure 100%. Par contre, dans le sens ME2, seulement le classifieur de type MLP fournit 100% de taux de bonne classification, voir tableaux 6.4 et 6.5.

Les deux experts "Sans défaut/Usure Patatoïde" et "Usure 75%/Usure Pata-

		Les classifieurs							
		Adaboost		Bagging		J48		Id3	
Classes	Configurations	1	2	1	2	1	2	1	2
		Usure 75%	100	100	100	100	99.37	99	97.87
	Trous	100	100	100	100	98.75	100	96.87	100
	Patatoïde	98.75	100	98.75	100	99.37	99	98.12	100

TAB. 6.4 – Taux global de classification pour la classe Usure100% dans le sens ME1.

		Les classifieurs					
		Logistic		MLP		RBF	
Classes	Configurations	1	2	1	2	1	2
		Usure 75%	98.04	99	98.04	100	98.04
	Trous	99.02	100	98.04	100	96.08	100
	Patatoïde	96.08	98	97.04	100	89.21	99

		Les classifieurs							
		Adaboost		Bagging		J48		Id3	
Classes	Configurations	1	2	1	2	1	2	1	2
		Usure 75%	98.04	100	97.06	99	98.04	99	76.4
	Trous	96.08	99	95	99	97.06	99	72.55	93
	Patatoïde	92.16	98	92.16	98	90.2	98	72.55	99

TAB. 6.5 – Taux global de classification pour la classe Usure100% dans le sens ME2.

toïde" à leur tour séparent facilement la classe Patatoïde de la classe de galet Sans défaut ainsi que la classe des galets usés à 75%. Ils fournissent, dans le sens ME2, un taux de reconnaissance égal à 100% pour l'ensemble des classifieurs et les deux configurations. Le premier expert dans la configuration C2 se trompe seulement de 1% avec le modèle logistique et de 2% avec le classifieur RBF. Les résultats obtenus pour ces deux experts sont illustrés dans les tableaux 6.3 et 6.7.

Concernant la classe "Trous", l'expert "Usure 75%/Trous" dans le sens ME2, avec la configuration C2 (cf. tableau 6.7), représente un bon exemple. Les deux types de défaut sont montés sur des couples de galets directeurs. La force de l'appui ainsi que l'importance du défaut ne changent pas. Les trous sont montés sur l'essieu $ME1_{Av}$ et l'usure 75% est montée sur l'essieu $MI2_{Av}$. Les classifieurs : MLP et RBF ont fourni un taux de classification (CC) égal à 97%, (cf. tableau 6.7). Dans le cas où l'un des deux défauts est monté sur un essieu avant et l'autre sur un essieu arrière, le taux de classification est meilleur. Prenons par exemple le sens ME1 avec la configuration C1, dans ce cas les trous sont montés sur l'essieu $ME2_{Av}$ et l'usure 75% est montée sur l'essieu $MI1_{Att}$. Le classifieur SVM, avec la méthode de résolution SMO, fournit le meilleur résultat avec un taux de classification égal à 99.37%, (cf. tableau 6.6). Les mesures d'erreur ainsi que les matrices de confusions obtenues

pour chaque capteur sont détaillées dans les tableaux 6.10 et 6.11.

Classification Un-contre-Un (1vs1)							
		Les classifieurs					
		Logistic		MLP		RBF	
Classes \ Configurations		1	2	1	2	1	2
	Trous		98.75	95	98.75	93	98.12
Patatoïde		93.12	94	94.37	97	92.5	94

		Les classifieurs							
		Adaboost		Bagging		J48		Id3	
Classes \ Configurations		1	2	1	2	1	2	1	2
	Trous		99.37	90	98.75	89	98.75	92	94.37
Patatoïde		89.37	92	90	91	90	93	84.37	91

TAB. 6.6 – Taux global de classification pour la classe Usure75% dans le sens ME1.

		Les classifieurs					
		Logistic		MLP		RBF	
Classes \ Configurations		1	2	1	2	1	2
	Trous		100	96	99.02	97	97.06
Patatoïde		100	100	100	100	100	100

		Les classifieurs							
		Adaboost		Bagging		J48		Id3	
Classes \ Configurations		1	2	1	2	1	2	1	2
	Trous		98.04	98	99.02	96	99.02	96	90.02
Patatoïde		100	100	100	99	98.04	98	97.06	98

TAB. 6.7 – Taux global de classification pour la classe Usure75% dans le sens ME2.

Classification Un-contre-Un (1vs1)							
Les classifieurs							
Logistic							
MLP							
RBF							
Classes	Configurations	1	2	1	2	1	2
		Patatoïde	84.37	94	82.5	95	85

Les classifieurs									
Adaboost									
Bagging									
J48									
Id3									
Classes	Configurations	1	2	1	2	1	2	1	2
		Patatoïde	80	94	78.75	93	76.87	95	78.75

TAB. 6.8 – Taux global de classification pour la classe Trous dans le sens ME1.

Les classifieurs							
Logistic							
MLP							
RBF							
Classes	Configurations	1	2	1	2	1	2
		Patatoïde	96.08	96	96.8	96	95

Les classifieurs									
Adaboost									
Bagging									
J48									
Id3									
Classes	Configurations	1	2	1	2	1	2	1	2
		Patatoïde	94.12	94	93.13	93	93.13	95	94.12

TAB. 6.9 – Taux global de classification pour la classe Trous dans le sens ME2.

Classifieur : MLP – Configuration C2						
Classe	TP Rate	FP Rate	Précision	Recall	F-Measure	ROC Area
Premier capteur : $CC = 97\%$, $\kappa = 0.94$						
1	0.96	0.02	0.98	0.96	0.97	0.996
2	0.98	0.04	0.96	0.98	0.97	0.996
Deuxième capteur : $CC = 100\%$, $\kappa = 1$						
1	1	0	1	1	1	1
2	1	0	1	1	1	1
Troisième capteur : $CC = 97\%$, $\kappa = 0.94$						
1	0.96	0.02	0.98	0.96	0.97	0.995
2	0.98	0.04	0.96	0.98	0.97	0.995

TAB. 6.10 – Les mesures de l'erreur estimée avec le classifieur MLP pour l'expert "Usure75%/Trous".

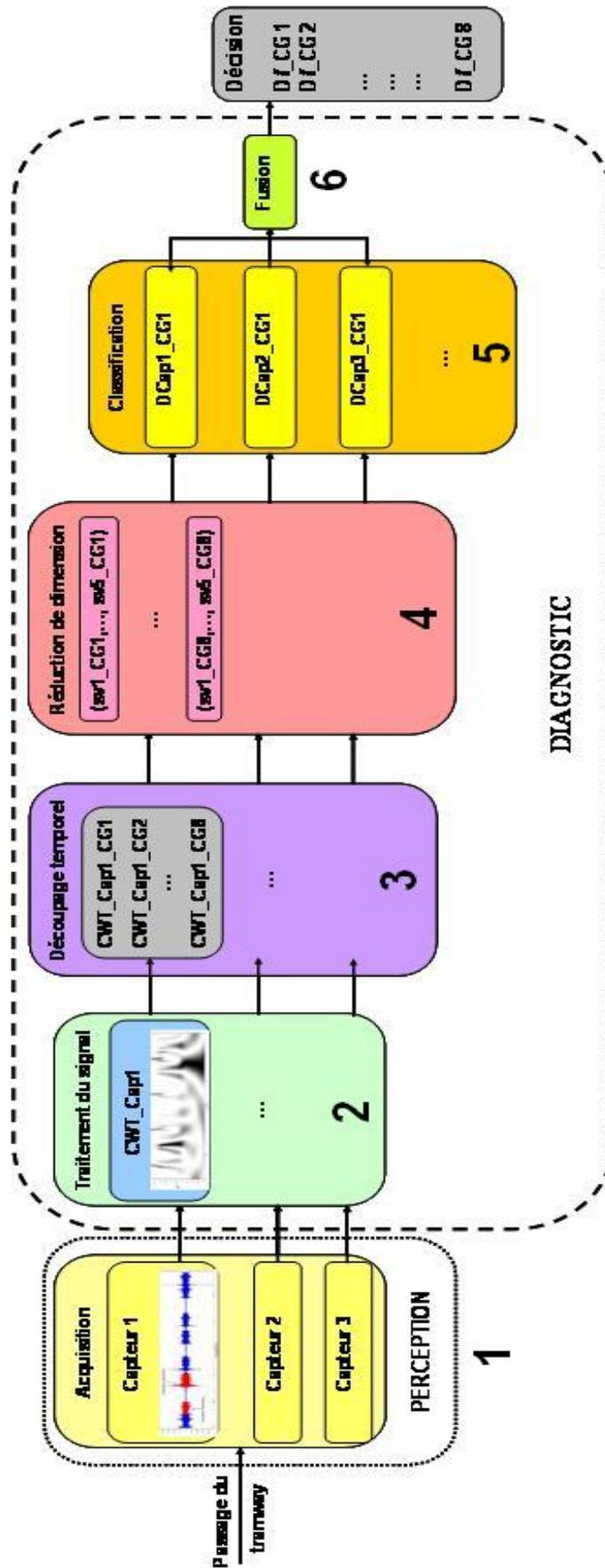


FIG. 6.1 – Les différentes étapes du processus de diagnostic.

Premier capteur		
a	b	← classer comme
48	2	a = Usure 75%
1	49	b = Trous
Deuxième capteur		
a	b	← classer comme
50	0	a = Usure 75%
0	50	b = Trous
Troisième capteur		
a	b	← classer comme
48	2	a = Usure 75%
1	49	b = Trous

TAB. 6.11 – Tableaux des matrices de confusions obtenues, avec le classifieur MLP, pour l'expert "Usure75%/Trous", pour les trois capteurs.

Pour l'expert "Usure à 75%/Trous", pour le premier et le troisième capteurs les 49 exemples (vecteurs formes), de la classe $b = \text{"Trous"}$, sont bien classés donc la valeur du TP Rate = $49/50 = 0.98$. Cependant, un seul exemple est mal classé parmi ceux appartenant à la classe $b = \text{"Trous"}$ et 48 exemples qui n'appartiennent pas à cette classe n'ont pas été reconnus comme tels donc la valeur FP Rate = $2/50 = 0.04$.

B - Influence des variations de configurations

Dans cette partie, nous avons mélangé les deux configurations C1 et C2 pour un sens de marche donné. Par conséquent, une seule base de données est construite par sens et par capteur pour une classe donnée. Le tableau 6.12 expose les résultats de la classification pour la classe sans défauts (Neuf ou non usé) dans les deux sens ME1 et ME2 .

Classification Un-contre-Un (1vs1)			
Classes	Les classifieurs		
	Logistic	MLP	RBF
Dans le sens ME1			
Usure 100%	100	100	99.61
Usure 75%	73.46	88.08	85.4
Trous	86.54	88.07	87.30
Patatoïde	77.30	91.16	78.08
Dans le sens ME2			
Usure 100%	100	100	100
Usure 75%	78.22	91.6	78.22
Trous	91.1	91.6	87.62
Patatoïde	97.03	97.03	97.03

TAB. 6.12 – Les résultats de classification pour la classe "Sans défaut", dans les sens ME1 et ME2 avec les deux configurations 1 et 2 mélangées.

Les résultats de la classification exposés dans le tableau 6.13 montrent qu'à

part pour la classe usure 100% et l'usure patatoïde, le taux de classification est légèrement détérioré. Prenons par exemple les résultats de la classification obtenus à l'aide des sous-classifieurs MLP. Le tableau 6.13 illustre et compare les résultats obtenus avec et sans mélange de configurations dans le sens ME1 et ME2. Dans le premier cas seule la configuration C2 est considérée. Dans le deuxième cas, la base de données testée est constituée du mélange d'un nombre égal d'exemples obtenus avec la configuration C1 (défaut sur les galets suiveurs) et de la configuration C2 (défaut sur les galets directeurs).

Classe : "Sans défaut" – Classifieur : "MLP"					
		Dans le sens ME1		Dans le sens ME2	
Classes	Mélange	Non	Oui	Non	Oui
	Usure 100%		100	100	100
Usure 75%		94	88.08	95	91.6
Trous		90	88.07	91	91.6
Patatoïde		94	91.16	100	97.03

TAB. 6.13 – Comparaison entre les résultats de la classification sans et avec mélange des deux configurations obtenus, avec un MLP, pour la classe "Sans défaut".

Cela peut s'expliquer et se justifier de deux manières :

- Un même défaut se comporte différemment dans une configuration ou une autre, C1 ou C2. Ainsi, l'aspect du signal d'un même défaut, l'amplitude, l'énergie et l'étalement, n'est pas le même et dépend plus au moins de la position du défaut sur l'essieu, avant ou arrière ;
- Les deux classes d'usure patatoïde et 100% sont des usures majeures et facilement repérables. Le fait d'être montée sur un couple suiveur (C1) ou directeur (C2) n'influe pas sur le taux de la classification.

C - Influence du sens de la marche

Dans la suite nous allons vérifier si notre première hypothèse préalablement annoncée dans la partie 3.4.2 à savoir - h_1 : le taux de classification sera meilleur dans le sens ME2 - est confirmée ou pas par les résultats obtenus dans l'étape de la classification. Pour cela, nous avons choisi la configuration 2 où les défauts sont tous montés sur les essieux avant sauf l'usure100% qui est sur le MI2_{arr}. Le tableau 6.14 expose et compare les résultats de la classification pour la classe sans défauts (galets neufs ou non usés) dans les deux sens de marche. Le taux de classification est toujours meilleur dans le sens ME2 mais pas comme nous l'avions imaginé. Nous nous attendions à un grand écart, ce qui n'est pas le cas.

Nous pensons que le taux de la classification dépend plus de la gravité du défaut et moins de son emplacement. Il est évident que la manière dont le couple est positionné, directeur ou suiveur, dégrade légèrement l'importance de l'usure mais à un degré plus au moins variable. Il est tout à fait logique qu'un défaut léger, une usure à 5% par exemple, ne soit pas perçu comme telle s'il est monté sur un couple

Classification Un-contre-Un (1vs1)							
		Les classifieurs					
		Logistic		MLP		RBF	
Classes	Le sens	ME1	ME2	ME1	ME2	ME1	ME2
		Usure 100%	100	100	100	100	100
	Usure 75%	93	97	94	95	94	96
	Trous	92	91	90	91	85	92
	Patatoïde	95	100	94	100	93	100

TAB. 6.14 – Le taux global de classification pour la classe "Sans défaut" : Comparaison entre les deux sens de marche, ME1 et ME2, avec la configuration 2.

suiveur. Toutefois, cette dégradation n'est pas pour l'instant quantifiée. Nous avons vu que l'usure normale à 75%, dans le sens ME1, a été détectée à l'aide du classifieur de type MLP avec un taux de 90% avec la configuration C1 contre 95% avec C2.

À la fin de l'étape d'analyse des signaux et avec le peu d'information que nous avons sur le comportement réel du système de guidage (système de vérins), d'une part, et sur le comportement des défauts eux même d'autre part, nous nous attendions à avoir un écart entre le taux de reconnaissance d'un même défaut en configuration C1 et en configuration C2. Cette hypothèse - h_2 : le taux de classification sera meilleur avec la configuration C2 - annoncée précédemment dans la partie 3.4.2 est certes confirmée mais elle est conditionnée par l'importance du défaut et reste limitée.

Toutefois, nous sommes convaincus qu'afin de maximiser le taux de reconnaissance de n'importe quelle classe sans défaut ou avec défaut, que ce soit minime ou majeur, le détecteur SYSCOGA doit prendre en compte l'hypothèse h_2 où les défauts sont montés sur les essieux avant, couple de galets directeurs. Pour un sens de marche donné, le détecteur doit fournir seulement le rapport d'état des couples de galets directeurs. Autrement dit dans un sens, seuls les couples $GD_{Sens1} = (CG1, CG3, CG5, CG7)$ sont diagnostiqués et l'ensemble $GD_{Sens2} = (CG2, CG4, CG6, CG8)$ sera pris en compte dans l'autre sens. Un même couple de galets sera fort probablement vu plusieurs fois dans la journée par le système de diagnostic, cela dépend de la fréquence de roulement. Pour l'instant c'est le sentiment que nous avons et nous espérons qu'au delà d'un cycle de vie complet du galet (25000 km), d'autres contraintes et paramètres entreront en jeu. L'ensemble des défauts, leur comportement ainsi que leur évolution ne sera plus inconnue.

6.4 Spécifications de la T2C

6.4.1 Types de défauts retenus

Lors de la dernière réunion du 06 Juillet 2006 avec l'exploitant, la T2C, a souligné qu'à ce stade du projet, il n'a aucune connaissance préalable des défauts qui peuvent se manifester dans le cas réel ni de leur évolution dans un environnement spécifique à la piste du site Clermont-Ferrand. La chaussée utilisée est différente de celle de la piste d'essais de la société Lohr Industrie à Duppigheim (Strasbourg). De ce fait, la T2C ne peut pas exprimer ses besoins ni spécifier, à cette date, les types d'usures, leurs degrés et leurs importance. L'essentiel est de pouvoir détecter une usure normale qui peut regrouper, par exemple, l'usure parallèle (gommage du bandage) et les trous et l'usure accidentelle qui peut se manifester par un éclatement du galet (bandage) ou même l'absence du patin, etc. Malheureusement ce dernier type d'usure n'est pas considéré dans ce travail faute d'une base d'exemples de signaux correspondant à ce type d'usure.

L'usure Patatoïde, à notre avis, ne peut pas correspondre à une usure réelle et fort probablement ne sera jamais rencontrée dans un cas réel (bandage brûlé avec un chalumeau pour les essais). L'usure maxi à 100% est un état extrême de l'usure parallèle (gommage) et correspond à l'absence complète du bandage. Sa signature est très énergétique et sa détection ne pose pas de réel problème. Cela se traduit par un taux de classification, dans un sens ou dans un autre, toujours à $\approx 100\%$. Détecter une usure avec cette importance ne constitue, à notre avis, qu'un intérêt limité pour l'exploitant. Le système doit détecter une usure, par exemple parallèle, avant d'atteindre cet état extrême. Il reste cependant à préciser le niveau à partir duquel l'usure est considérée suffisamment importante pour être détectée. L'exploitant et la société Lohr industrie doivent donc désigner un niveau d'usure. Ce réglage doit être ni trop sévère ni assez souple afin de mieux gérer les fausses alarmes. Par exemple si l'usure parallèle à 25% est nocive pour le système de guidage et/ou augmente considérablement le niveau acoustique, ce degré d'usure peut être alors considéré comme un niveau déclencheur. D'une autre manière, l'exploitant peut définir un intervalle de confiance, $[Val_{min} - Val_{max}]$. Une fois la valeur minimale d'usure franchie, le tramway peut rouler sous surveillance. L'usure est dans ce cas considérée comme acceptable jusqu'à ce qu'elle dépasse la borne maximum d'usure tolérée.

6.4.2 Nouvelle démarche suivie et données utilisées

En considérant l'usure patatoïde et l'usure à 100% comme inintéressantes pour notre étude, seules les classes suivantes sont retenues :

- classe sans défaut (galets neufs ou non usés) ;
- classe d'usures normales :
 - usures parallèles : usure 75% ;
 - usures aléatoires : les trous.

Ce problème de classification peut être résolu de deux manières. La plus pratique, est de considérer un problème à 2 classes, sans défaut et usures normales, et répondre ainsi aux besoins actuels de l'exploitant. Dans le cas où la connaissance de la nature de défaut est souhaitée, une deuxième étape peut être mise en oeuvre. La procédure de reconnaissance d'une sous classe de défauts est réalisée à partir du moment où le couple de galets est jugé défectueux. La détection du défaut et l'identification de sa nature, usure75% ou trous, s'effectue d'une manière séquentielle. En conséquence, une structure de détection séquentielle, avec deux sous-détecteurs, a été définie, figure 6.2.

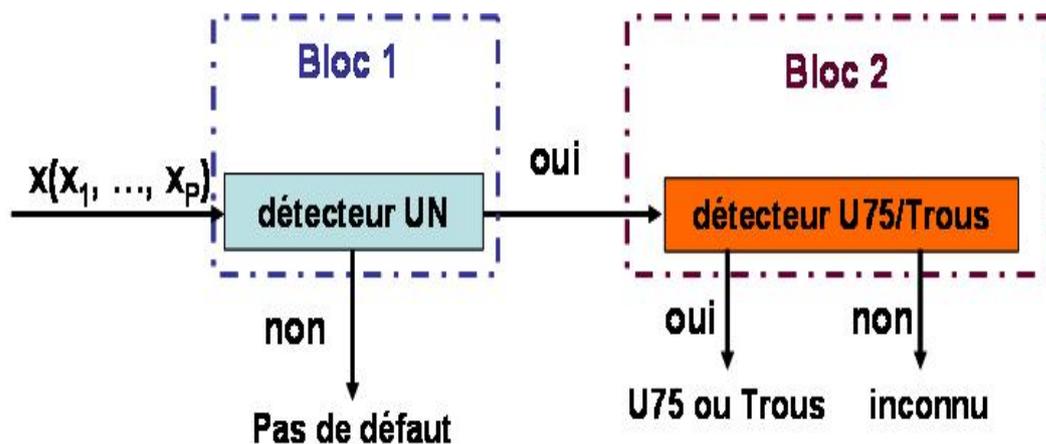


FIG. 6.2 – Le processus de classification choisi.

L'intérêt de cette approche est que la quantité de données, le temps de calculs et de traitements effectués reste proportionnelle à la difficulté du problème de classification à résoudre à chaque niveau. Ce gain se verra mieux avec un nombre de classes important (par exemple > 3).

Pour chaque capteur, nous effectuons le même processus de classification illustré dans la figure 6.2. La décision finale Df pour un CGk est alors le résultat de la fusion des trois décisions issues des trois capteurs, $DCapt_1, DCapt_2, DCapt_3$ suivant la stratégie du vote majoritaire. Les ambiguïtés, dans notre cas, sont prises comme des erreurs de classification.

Le premier bloc effectue une séparation entre la classe sans défaut et la classe d'usures normales. Un galet est tout suite classé sans ou avec défaut. S'il est défectueux, le deuxième bloc précise la nature de l'usure. Le deuxième sous détecteur sert ainsi à repérer les trous des usure75%. Bien que ces deux derniers défauts se manifestent en hautes fréquences, ($10.75 \leq \nu \leq 350$) Hz, ils sont bien séparables. Nous allons dans les paragraphes suivants mettre en oeuvre cette démarche de classification en utilisant, dans un premier temps, les bases de données construites à partir des 5 premières valeurs singulières, (cf. 6.5.1). Dans un deuxième temps, l'étape de classification est effectuée à l'aide des pseudo-signatures calculées à partir du premier vecteur singulier de gauche suivant la méthode de Venkatachalam décrite dans (cf. 5.5.2). Les résultats obtenus sont exposés dans la partie 6.5.2. Nous retenons pour les paragraphes suivants seulement les deux types de classifieurs : MLP et RBF.

6.5 Détection de la présence d'un défaut

6.5.1 Classification avec les valeurs singulières

Pour l'étape de la classification nous avons utilisé les deux bases de données Base1 = Sans défauts/Usure normale et Base2 = Trous/Usure 75% construites à partir des 5 premières valeurs singulières. La classe "Usure normale" regroupe la classe : Usure75% et Trous. La classe sans défaut regroupe les couples de galets montés sur un essieu arrière ($ME1_{arr}$ et $MI1_{arr}$). Pour la configuration C2, seule $ME1_{av}$ comporte un couple de galets sans défaut et sera donc utilisé pour la classification. La première base de données est constituée de 200 exemples d'apprentissage, 100 de chaque classe. Tandis que la deuxième base est composée de 50 exemples de la classe Trous et 50 de la classe Usure 75%, au total 100 exemples. Le tableau 6.15 présente les taux globaux de classification obtenus avec la méthode Leave-One-Out (LOO) à l'aide des classifieurs : MLP et RBF. Nous remarquons que le réseau de neurones de type MLP a donné le meilleur résultat pour les deux experts. Pour le premier, un galet sans défaut est reconnu à 90% et dans le cas "Trous/Usure 75%", l'expert se trompe simplement de 5% entre un galet avec trous et un galet usé à 75%. Voyons maintenant si l'utilisation des pseudo-signatures de puissances, (cf. ??), avec les même classifieurs va améliorer ce taux de classification.

Classes	Les classifieurs	
	MLP	RBF
Configuration 2		
Sans défauts/Usure normale	90	87
Usure75%/Trous	92	94

TAB. 6.15 – Taux global de classification dans le sens ME1 avec la configuration C2.

6.5.2 Classification avec les pseudo-signatures de puissances

Nous avons évoqué dans la partie 5.5.2 la possibilité d'améliorer les résultats de la classification en utilisant les vecteurs singuliers de gauche (échelles) et de droite (temps) issus de la décomposition en valeurs singulières des scalogrammes comme signature d'une classe donnée. Ainsi, au lieu de travailler avec les 5 premières valeurs singulières, nous allons dans la suite utiliser les pseudo-signatures de puissance (PPS), calculées suivant la méthode décrite dans la partie 5.5.2 en entrée du classifieur.

Pour effectuer l'étape de classification à partir des pseudo-signatures deux bases de données ont été construites. Pour la première phase de classification, l'expert "Sans défaut/Usure normale" est entraîné sur les PPS des deux classes sans défaut et usure normale. Cette dernière classe regroupe les deux classes de défaut "Trous" et "Usure 75%". La première base de données résultante est de dimension (200×21) . Cette première phase de classification est suivie d'une deuxième phase de classification qui permet l'identification du type d'usure (Trous ou usure 75%). L'expert

"Trous/Usure 75%" est entraîné seulement sur les données des deux classes considérées. La base de données construite cette fois-ci est de dimension (100×21) . Le nombre d'exemples, N_i , par classe dans les deux cas est petit $N_i \ll 10 * p$, avec N le nombre de vecteurs formes (observations) et p représente le nombre de variables par vecteur. Cela nécessite le passage par une étape de réduction de dimension, (cf. 5.2), pour améliorer les performances des classifieurs et augmenter ainsi le taux de bonne classification. De même, pour effectuer l'étape de sélection nous avons utilisé le logiciel Weka. Quatre méthodes de sélection automatique de variables ont été mises en oeuvre et comparées afin de garder celle qui fournit les meilleurs résultats de classification. Une fois de plus le logiciel Weka nous a permis de comparer plusieurs méthodes de sélections en utilisant plusieurs critères de recherche et d'arrêt. La méthode CFS, quelque soit les critères de recherche et d'arrêt sélectionné, a permis de sélectionner la ou l'ensemble des variables qui maximise la séparation entre les classes apprises par expert. Cette technique évalue la valeur d'un sous-ensemble d'attributs en considérant la capacité prédictive individuelle de chaque variable avec le degré de redondance entre eux. Les sous-ensembles de variables qui sont fortement corrélés avec la classe ayant la basse inter corrélation sont préférés.

Regardons maintenant si cette méthode de sélection automatique de variables donnent le même résultat que la sélection manuelle de l'intervalle de confiance, (cf. 5.5.2), pour les deux experts "Sans usure/Usure normale" et "Trous/Usure 75%". Est-ce que l'intervalle d'échelle [9 – 11], préalablement sélectionné manuellement, sera aussi choisi par les méthode de sélection automatique ?.

La CFS a fourni globalement pour l'ensemble des cas testés, critères de recherche et d'arrêts, les mêmes résultats :

- pour le premier capteur, l'ensemble des valeurs sélectionnées [1, 4, 5, 7, 14] correspondent aux valeurs d'échelles [12, 11.25, 11, 10.5, 8.75] ;
- pour le deuxième capteur les valeurs sélectionnées sont [1, 6, 13, 14]. Elles correspondent aux échelles [12, 10.75, 9, 8.75] ;
- pour le troisième capteur, une seule valeur a été sélectionnée. Elle correspond à l'échelle 10.75.

Les cinq premières valeurs choisies restent cohérentes avec notre intervalle de confiance [9 – 11] préalablement sélectionné à partir des PPS moyennes.

Les résultats de classification obtenus, en utilisant les variables sélectionnées à l'aide de la méthode CFS avec le critère d'arrêt : "Best_{First} bidirectionnelle", sont illustrés dans le tableau 6.16. Les taux de classification obtenus, par capteur, avec le classifieur de type "RBF" sont légèrement meilleurs que ceux obtenus avec un "MLP" pour les deux experts. Non seulement le taux de la classification est meilleur que celui obtenu avec les valeurs singulières, tableau 6.15, mais le temps de calcul est nettement amélioré surtout pour les MLPs. Les deux tableaux 6.17 et 6.18, comparent les temps de calcul passés dans l'étape de classification avec un MLP et un RBF, par expert et pour chaque capteur, dans un premier temps, avec les valeurs singulières et ensuite après une sélection de variables à partir du premier vecteur singulier des caractéristiques spectrales.

Pour le premier expert, le réseau de neurones de type MLP a donné le meilleur

résultat. Un galet sans défaut est reconnu à 97.5%. Pour le deuxième expert, le RBF a fournit le meilleur taux de bonne classification. L'expert se trompe de 2% entre un galet avec trous et un galet usé à 75%, à savoir 98% le taux global de bonne classification. Les résultats obtenus avec le réseau RBF pour le deuxième expert sont détaillés dans les deux tableaux 6.19 et 6.20.

CC% pour le premier expert : "Sans défaut/Usure normale"		
Les classifieurs	MLP	RBF
Les Critères d'arrêt		
$Best_{First}$ B	97.5	97
CC% pour le deuxième expert : "Trous/Usure à 75%"		
Les classifieurs	MLP	RBF
Les Critères d'arrêt		
$Best_{First}$ B	97	98

TAB. 6.16 – Le taux de bonne classification pour les deux experts ("Sans défaut/Usure normale" et "Trous/Usure à 75%") dans le sens ME1 avec la configuration C2, en utilisant le premier vecteur singulier des caractéristiques spectrales après une étape de sélection de variables.

Le temps de calcul CPU (s) : l'expert "Sans défaut/Usure normale"			
Les capteurs	1	2	3
Données en entrée			
σ	1.19	1.11	0.96
CFS(U)	0.56	0.56	0.3
Le temps de calcul CPU (s) : l'expert "Trous/Usure à 75%"			
Les capteurs	1	2	3
Données en entrée			
σ	0.55	0.53	0.51
CFS(U)	0.14	0.16	0.23

TAB. 6.17 – Comparaison entre le temps de calcul de la classification, avec un MLP, en utilisant les 5 premières valeurs singulières et les variables sélectionnées à partir du premier vecteur singulier de gauche (échelles).

Le classifieur RBF pour l'expert "Trous/Usure 75%" se trompe sur la classification, pour le premier capteur, de 4 vecteurs formes sur les 100 existants. Ainsi, il classe deux couples de galets comme étant avec trous, alors qu'ils sont usés à 75%. Donc les deux exemples ne sont pas de la classe $a = \text{"Trous"}$, mais le classifieur la leur attribue quand même, c'est un **Faux Positif**, FP = 2. Il juge que les deux autres exemples sont usés à 75% alors qu'ils sont avec trous. Les deux exemples appartiennent à la classe $a = \text{"Trous"}$ mais le classifieur se trompe, c'est un **Faux Négatif**, FN = 2. De la même manière, pour le deuxième et le troisième capteur, deux couples de galets seulement sont mal classés. Un couple a été classé comme usé à 75% alors qu'il est avec trous et inversement, l'autre couple a été jugé avec trous alors qu'il est usé à 75%, FP = FN = 1. Ces valeurs sont regroupé dans la matrice de confusion, définie en annexe (cf. C.2). En apprentissage supervisé, cette matrice sert à mesurer la qualité d'un système de classification. Chaque colonne de la matrice

Le temps de calcul CPU (s) : l'expert "Sans défaut/Usure normale"			
Données en entrée	Les capteurs		
	1	2	3
σ	0.18	0.24	0.28
CFS(U)	0.2	0.11	0.03
Le temps de calcul CPU (s) : l'expert "Trous/Usure à 75%"			
Données en entrée	Les capteurs		
	1	2	3
σ	0.04	0.04	0.02
CFS(U)	0.02	0.02	0.03

TAB. 6.18 – Comparaison entre le temps de calcul de la classification, avec un RBF, en utilisant les 5 premières valeurs singulières et les variables sélectionnées à partir du premier vecteur singulier de gauche (échelles).

représente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle, (cf. tableau 6.19).

Premier capteur		
a	b	← classer comme
48	2	a = Trous
2	48	b = Usure à 75%
Deuxième capteur		
a	b	← classer comme
49	1	a = Trous
1	49	b = Usure à 75%
Troisième capteur		
a	b	← classer comme
49	1	a = Trous
1	49	b = Usure à 75%

TAB. 6.19 – Les matrices de confusions calculées avec l'expert RBF "Trous/Usure à %75" pour les trois capteurs. Les valeurs utilisées en entrées sont ceux sélectionnées, à partir du premier vecteur singulier, avec la méthode CFS ($Best_{FirstB}$)

Le tableau 6.20 regroupe l'ensemble de mesures de l'erreur, estimées pour l'expert RBF "Trous/Usure à %75", confirme ces résultats. Le coefficient κ est égal à 0.92 pour le premier capteur et 0.96 pour le deuxième et le troisième capteur. Rappelons que ce coefficient mesure le degré de concordance, dans notre cas, de deux juges. C'est à dire la différence entre l'accord constaté entre les deux juges, et l'accord qui existerait si les juges classaient les exemples au hasard. Le jugement, c'est la classe d'un exemple, et les deux juges sont le classifieur et la classe réelle de l'exemple. Pour les trois capteurs le coefficient κ a une grande valeur, donc presque tous les exemples sont sur la diagonale (matrice diagonale), l'accord dans ce cas entre les deux juges est excellent. Le paragraphe C.1, en annexe, donne plus de détails sur le calcul du coefficient kappa κ .

En se basant sur les résultats, présentés dans le tableau 6.19, on peut aussi calculé

les mesures d'exactitude par classe et par expert, (cf. tableau 6.20).

Classifieur : RBF – L'expert "Trous/Usure à %75" - CFS (Best _{FirstB})						
Classe	TP Rate	FP Rate	Précision	Recall	F-Measure	ROC Area
Premier capteur : $\kappa = 0.92$						
1	0.96	0.04	0.96	0.96	0.96	0.942
2	0.96	0.04	0.96	0.96	0.96	0.942
Deuxième capteur : $\kappa = 0.96$						
1	0.98	0.02	0.98	0.98	0.98	0.98
2	0.98	0.02	0.98	0.98	0.98	0.974
Troisième capteur : $\kappa = 0.96$						
1	0.98	0.02	0.98	0.98	0.98	0.976
2	0.98	0.02	0.98	0.98	0.98	0.976

TAB. 6.20 – Les mesures de l'erreur estimées pour le classifieur RBF, avec les valeurs sélectionnées à partir du premier vecteur singulier de gauche obtenus avec la méthode CFS (Best_{FirstB}), de l'expert "Trous/Usure à %75".

Par exemple, pour l'expert "Trous/Usure à 75%", pour le deuxième et le troisième capteurs les 49 exemples (vecteurs formes), de la classe $a = \text{"Trous"}$, sont bien classés donc la valeur du TP Rate = $49/50 = 0.98$. Cependant, un seul exemple est mal classé parmi ceux appartenant à la classe $a = \text{"Trous"}$ et 49 exemples qui n'appartient pas à cette classe n'ont pas été reconnus comme tels donc la valeur FP Rate = $1/50 = 0.02$. L'ensemble de ces mesure de l'erreur sont détaillés en annexe, (cf. C). On peut estimer les sensibilités et spécificités, définis en annexe, (cf. C), correspondant à cet expert "Trous/Usure à 75%" à partir de la matrice de confusion. Prenons par exemple la matrice de confusion du deuxième et du troisième capteur qui est la même, (cf. 6.21).

a	b	← classer comme
49	1	a = Trous
1	49	b = Usure à 75%

TAB. 6.21 – La matrice de confusion calculées avec l'expert RBF "Trous/Usure à %75" pour le deuxième et le troisième capteur. Les valeurs utilisées en entrées sont ceux sélectionnées, à partir du premier vecteur singulier, avec la méthode CFS (Best_{FirstB})

La sensibilité correspond au rapport des vrais positifs (TP Rate) donc le rapport entre le nombre des exemples bien classés et le nombre total d'éléments qui devraient être bien classés. Elle correspond, dans ce cas, à $Se = TP Rate = TP/(TP + FN) = 49/(49 + 1) = 0.98$. Le paramètre de spécificité quand à lui correspond à $Sp = TN/(TN + FP) = 49/(49 + 1) = 0.98$. L'indice de Youden est égal à 0.96. Cet indice se rapproche du 1 donc le test (classification) est "efficace". Pour l'expert "Trous/Usure à 75%", le coefficient de Yule $Q = 1$ donc la liaison entre les résultats attendus (les classes) et les résultats observés est très forte. Cela veut dire que les deux classes "Trous" et "Usure à 75%" sont très différents (séparables) et leur présence est rapidement détectée.

Conclusion sur les méthodes MLP, RBF, ...

Bien que performant sur beaucoup de problèmes, la principale lacune, des méthodes discriminantes telles que le MLP et le réseau RBF, est la difficulté d'adapter adéquatement l'architecture du classifieur (réseau de poids) au problème de classification, (cf. 2.5). L'autre maillon faible des méthodes, est leur sensibilité à la dimensionalité des données. Comme nous l'avons noté dans le paragraphe 2.6, le SVM procure un cadre théorique qui résout le problème du choix du nombre de paramètres puisque celui se fait automatiquement. L'efficacité du SVM à traiter des données de nature et de complexité arbitraire a été établie sur beaucoup d'applications de reconnaissance de formes. Sa robustesse vis à vis de la dimensionalité des données et son pouvoir accru de généralisation, font que le SVM est nettement plus avantageux.

Smola et Wahba [243, 246] ont démontré qu'associer un noyau particulier à un SVM revient à considérer une pénalisation différente de l'erreur d'apprentissage en maximisant la marge. Ce qui nous permet de dire que la maximisation de la marge dans l'espace augmenté est une forme de régularisation de l'apprentissage.

Nous nous intéressons dans la section suivante aux machines à vecteurs supports (SVM). Afin de montrer leur performance, nous allons comparer les taux de classification obtenus avec les MLPs et les RBFs avec celles fournis par les SVMs. Pour la mise en oeuvre des SVMs, nous utilisons la toolbox SVM développée au laboratoire PSI sous Matlab [164] et l'algorithme de base SMO, pour le calcul des SVMs, intégré dans le logiciel Weka.

De la même manière, les taux de classification sont donnés pour les deux experts : "Sans usure/Usure normale" et "Trous/Usure %75". L'étape de la classification est effectuée, dans un premier temps, à l'aide des bases de données construites à partir des 5 premières valeurs singulières, et dans un deuxième temps, avec celles construites à partir des pseudo-signatures de puissances.

6.6 Classification par SVM

Un système de classification par SVM nécessite le réglage de paramètres tels que le paramètre de pénalisation C ou le choix d'un type de noyau et de sa paramétrisation. Ces paramètres sont typiquement réglés au moyen d'une procédure de validation croisée. Cette procédure nécessite l'exécution de plusieurs instances d'apprentissage et de test en explorant un ensemble de valeurs possibles des différents paramètres, pour retenir celles donnant les meilleurs résultats de classification.

Une alternative moins coûteuse et plus avantageuse dans les situations où les données à disposition sont limitées, consiste à trouver les paramètres optimaux à partir d'un unique ensemble d'apprentissage en exploitant des critères qui tentent de prédire le comportement du classificateur en généralisation (c'est-à-dire sur de nouveaux exemples de test). Deux critères de ce type ont été utilisés : l'estimation du

risque réel (l'erreur en généralisation) et la borne VC. Ces deux caractéristiques ont été présentées dans des travaux antérieurs comme efficaces pour l'obtention de bons réglages des classificateurs SVM. Joachims [137] propose de minimiser l'erreur sur un ensemble de valeurs possibles des paramètres pour déterminer les valeurs adéquates du paramètre C et le choix du noyau. Il présente cette erreur comme un critère possible pour sélectionner la valeur de C et les paramètres du noyau. Schölkopf et al. [242] préconisent d'élire pour cela le jeu de paramètres réalisant la plus petite dimension VC (notée h) afin d'obtenir la borne la plus petite sur le risque. Nous envisageons dans la suite l'utilisation du premiers critère pour le réglage des SVM. Les paramètres retenus sont ceux qui minimisent l'erreur de généralisation.

6.6.1 Paramètres d'optimisation du calcul des SVM

Nous commençons par étudier le comportement des SVM vis à vis du paramètre de pénalisation C . De nombreux tests préliminaires ont été effectués en utilisant les deux noyaux polynômial et gaussien et des valeurs de C que nous avons fait varier par puissances de 10, en prenant $C = \{1, 10, 100, 1000, 10000\}$. Cela nous a conduit à restreindre l'étude à des valeurs de C dans l'ensemble $C \in [5 - 100]$. Les mêmes comportements que ceux qui sont décrits ici sont retrouvés pour des valeurs plus grandes de C et d'autres noyaux.

Des expériences préliminaires ont été effectuées sur les trois classes : "Sans défaut", "Trous" et "Usure à 75%" à partir des données des bases d'apprentissage. Deux machines à vecteurs supports sont ainsi apprises pour les paires "Sans défaut/Usure normale" et "Trous/Usure à 75%", pour chaque ensemble de valeurs de C , le degré du polynôme δ et la largeur de la gaussienne σ . Nous obtenons dans chaque cas une estimation de l'erreur en généralisation à l'aide de la procédure Leave-One-Out (LOO). Ainsi nous pouvons en déduire le taux global de la bonne classification.

Nous souhaitons maintenant, en nous fixant une valeur de C , sélectionner le noyau et sa paramétrisation les plus appropriés, en nous basant sur les critères d'erreur. Rappelons qu'à priori les meilleures configurations réalisent une dimension VC minimale et une erreur minimale. Nous considérons dans la suite les noyaux, polynômial

$$k(x, y) = \left(\frac{x \cdot y}{d} \right)^\delta, \quad (6.1)$$

et RBF gaussien

$$k(x, y) = \exp \left(- \frac{\|x - y\|^2}{d\sigma^2} \right). \quad (6.2)$$

Les choix intéressants de σ se situent dans l'intervalle $[0, 1]$. Nous présentons les cas $\sigma \in [0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.5, 0.7, 1]$ qui

sont assez représentatifs. Nous réalisons l'apprentissage des deux SVM ("Sans défaut/Usure normale", "Trous/Usure à 75%") pour les différents choix du paramètre δ , pour le noyau polynômial, et du paramètre σ pour le noyau gaussien, en mesurant l'erreur de généralisation sur les deux bases de données construites. Les deux parties 6.6.2 et 6.6.3 présentent respectivement les résultats de classification obtenus avec la procédure LOO, dans un premier temps, avec les valeurs singulières, et dans un deuxième temps, avec les pseudo-signatures de puissances (PPS).

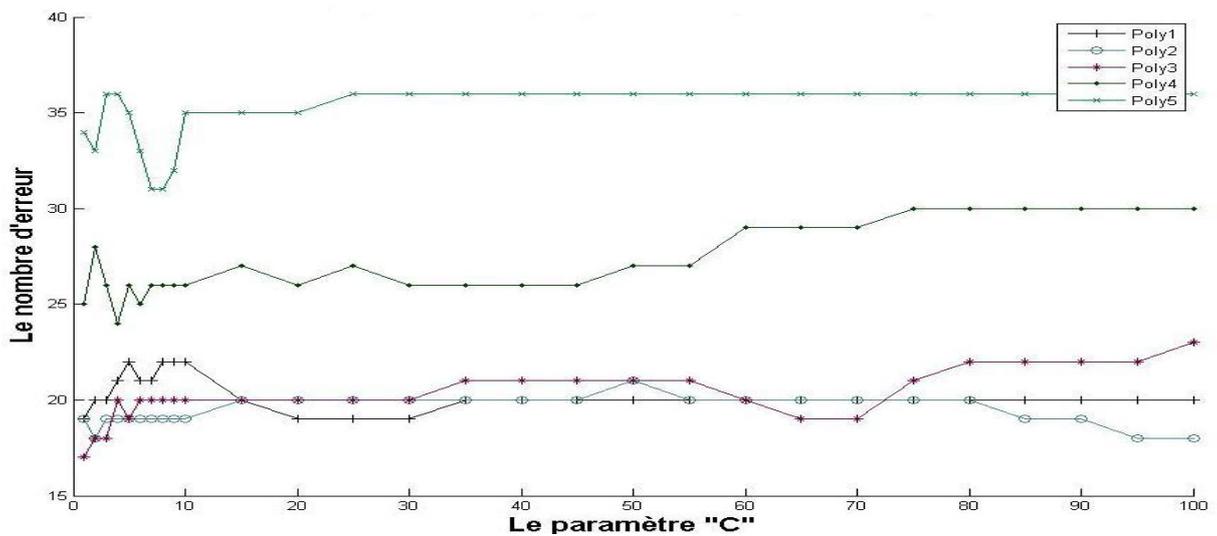
6.6.2 Classification SVM avec les valeurs singulières

Les figures 6.3 et 6.4 illustrent les résultats obtenus avec le noyau polynômial, pour les deux experts. Le nombre d'erreurs de généralisation est donné pour le couple de paramètres (δ, C) avec $\delta \in [1, 2, 3, 4, 5]$ et $C \in [1 - 100]$. Pour le premier expert, "Sans défaut/Usure normale", nous remarquons que l'ensemble des paramètres qui donnent les meilleurs résultats pour le premier, le deuxième et le troisième capteurs sont $(\delta = 2, C = 95)$, $(\delta = 3, C = 60)$ et $(\delta = 1, C = 4)$ correspondant respectivement à 91%, 93.5% et 88% de taux de bonne classification. Pour l'expert "Trous/Usure à 75%", les paramètres $(\delta = 1, C = 30)$, $(\delta = 1, C = 60)$ et $(\delta = 1, C = 7)$ fournissent les meilleurs résultats pour les trois capteurs, avec respectivement un taux de bonne classification de 94% pour le capteur 1, 98% pour le capteur 2 et 96% pour capteur 3. Le nombre d'erreurs est fourni par expert et pour les trois capteurs. Rappelons que les bases de données sont de dimension (200×5) et (100×5) pour respectivement le premier et le deuxième expert.

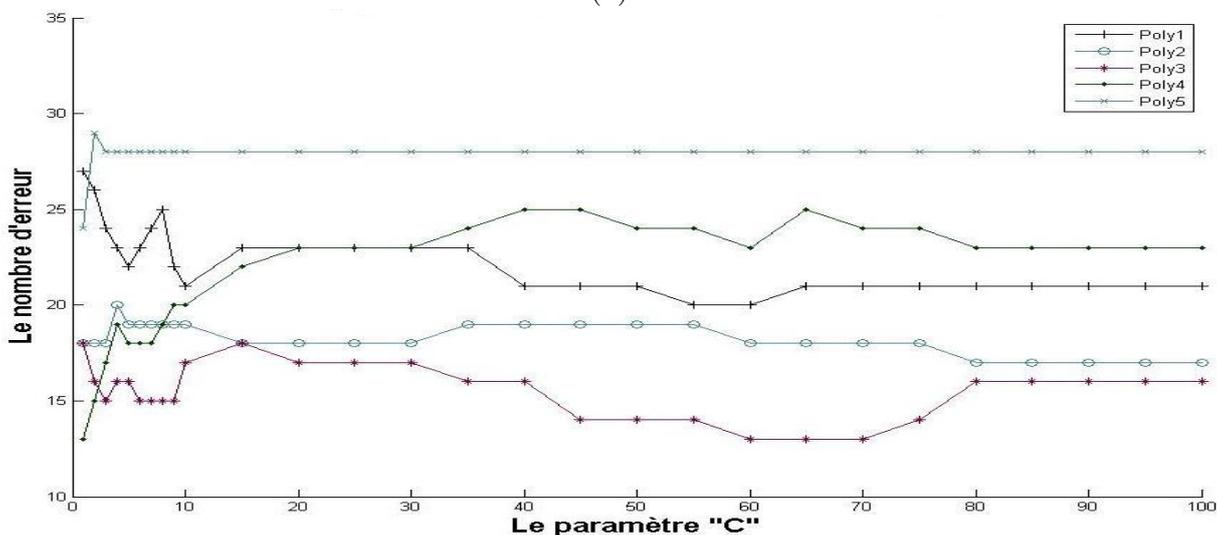
Nous remarquons que le nombre d'erreurs commises par chaque expert :

- est plus lié au paramètre δ du noyau polynômial ;
- varie beaucoup dans le cas d'un expert "Sans défaut/Usure normale" alors qu'il reste assez statique dans le cas de l'expert "Trous/Usure à 75%", (cf. figures 6.3 et 6.4) ;

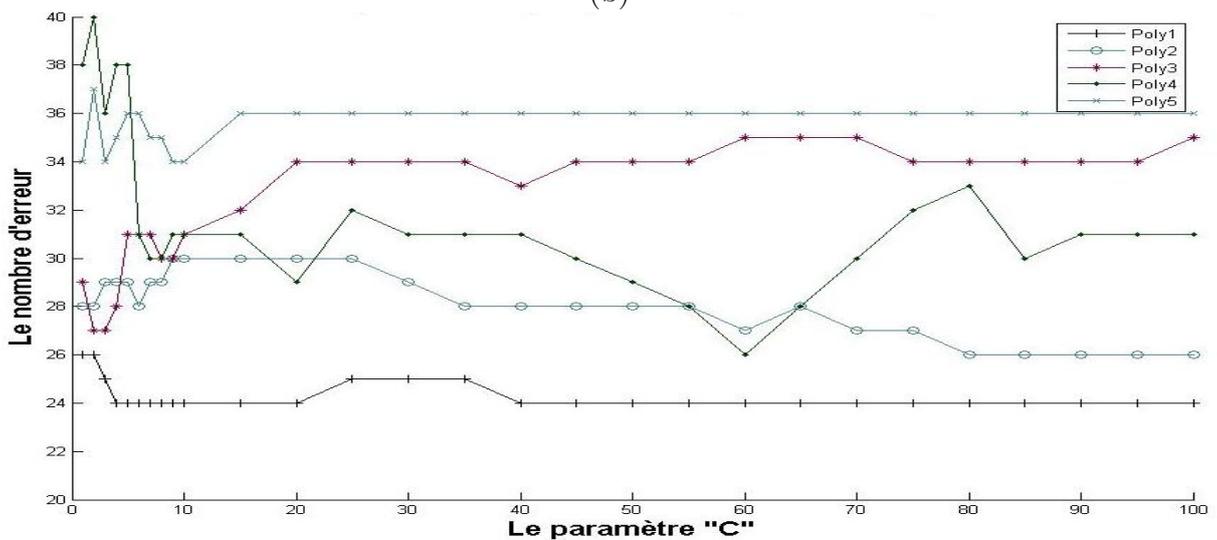
Les tableaux 6.22 et 6.23 résument les résultats obtenus avec un noyau gaussien pour respectivement l'expert "Sans défaut/Usure normale" et l'expert "Trous/Usure à 75%" avec les bases de données construites à partir des 5 premières valeurs singulières. Le nombre d'erreurs est calculé pour l'ensemble de paramètres croisés $\sigma \in [0 - 1]$ et $C \in [5 - 100]$. Nous remarquons, pour les 3 capteurs, que pour n'importe quelle valeur de la largeur du noyau gaussien $\sigma \in [0.02 - 0.1]$ et quelque soit les valeurs du paramètre de pénalisation $C \in [5 - 100]$ le nombre d'erreurs reste statique. Prenons par exemple le premier expert "Sans défaut/Usure normale", pour une valeur de σ égale à 0.06, le classifieur se trompe sur 54 exemples sur les 200 exemples de la base de données du premier capteur. Le nombre d'erreurs reste statique et ne dépend que de la valeur de σ . Tandis que pour des valeurs un peu plus grande de $\sigma \in [0.2 - 1]$, le nombre d'erreurs commises dépend des deux paramètres C et σ . Ce même résultat est constaté sur les 3 capteurs et les deux experts "Sans défaut/Usure normale" et l'expert "Trous/Usure à 75%", (cf. tableaux 6.22 et 6.23).



(a)

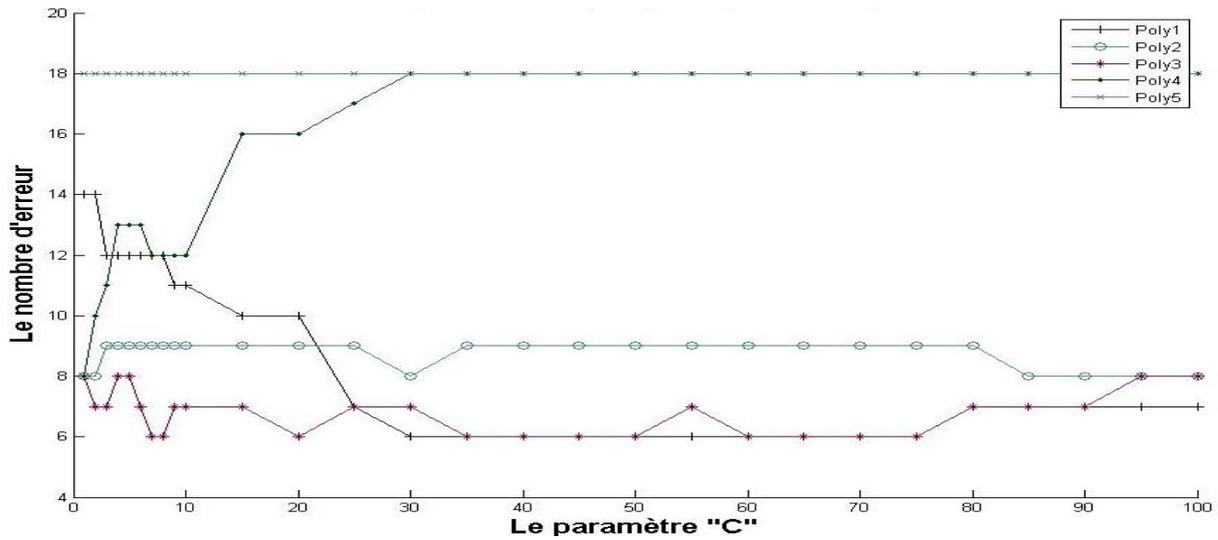


(b)

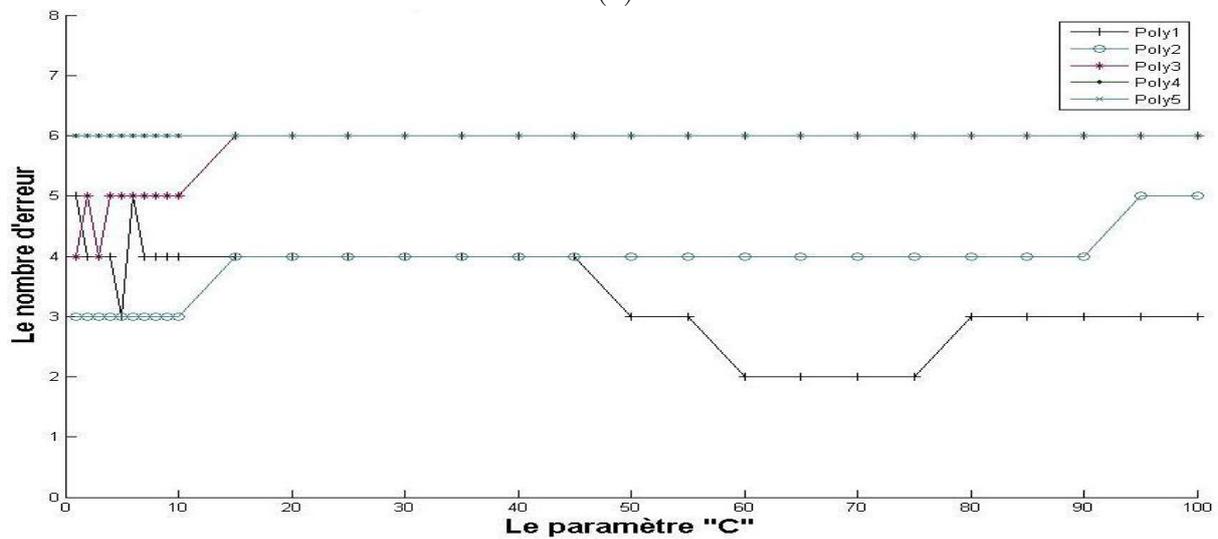


(c)

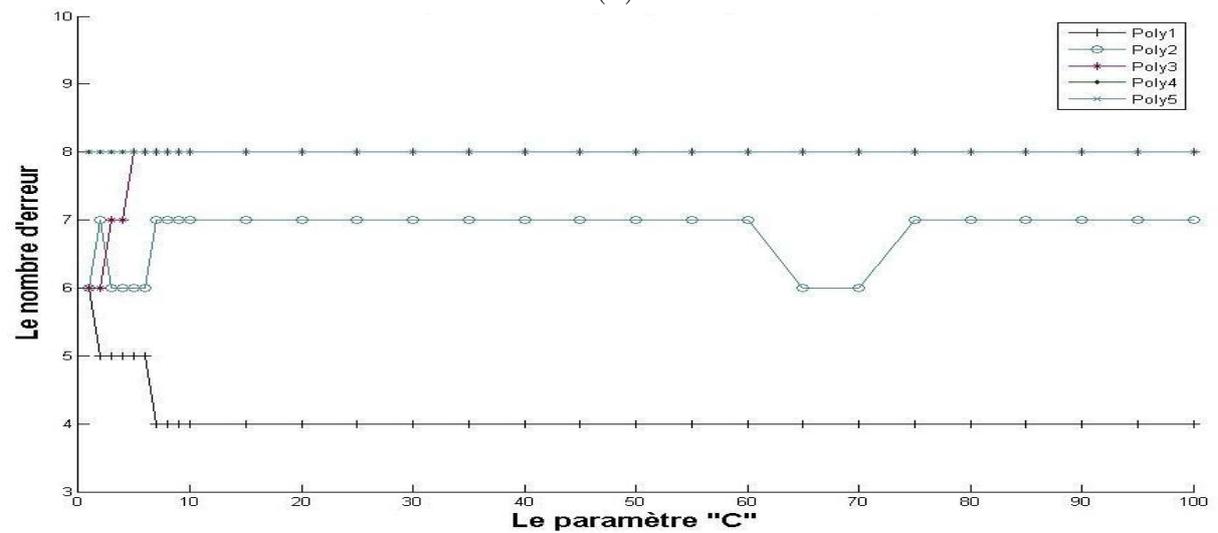
FIG. 6.3 – Le nombre d'erreurs obtenu pour l'expert "Sans défaut/Usure normale", en utilisant les 5 premières valeurs singulières, avec le couple de paramètres (δ, C) pour le premier (a), le deuxième (b) et le troisième (c) capteur.



(a)



(b)



(c)

FIG. 6.4 – Le nombre d'erreurs obtenu pour le deuxième expert "Trous/Usure à 75%", en utilisant les 5 premières valeurs singulières, avec le couple de paramètres (δ, C) pour le premier (a), le deuxième (b) et le troisième (c) capteurs.

Le premier expert : (Sans défaut/Usure normale)						
Les capteurs Sigma (σ)	1		2		3	
	C	Nb Err	C	Nb Err	C	Nb Err
0.02	5-100	196	5-100	198	5-100	190
0.03	5-100	83	5-100	180	5-100	82
0.04	5-100	67	5-100	142	5-100	76
0.05	5-100	62	5-100	103	5-100	63
0.06	5-100	54	5-100	56	5-100	62
0.07	5-100	50	5-100	50	5-100	53
0.08	5-100	48	5-100	41	5-100	45
0.09	5-100	45	5-100	48	5	46
0.1	5-100	41	5-10	45	5-100	46
0.2	5	26	5	31	5	39
0.5	5	22	20	15	5	30
0.7	10	19	30	14	5	29
1	5	20	100	15	5	29

TAB. 6.22 – Nombre d’erreurs de classification pour l’expert "Sans défaut/Usure normale" dans le sens ME1 avec la configuration C2 : 5 premières valeurs singulières - $C \in [5 - 100]$.

Le deuxième expert : (Trous/Usure à 75%)						
Les capteurs Sigma (σ)	1		2		3	
	C	Nb Err	C	Nb Err	C	Nb Err
0.02	5-100	100	5-100	100	5-100	100
0.03	5-100	97	5-100	93	5-100	96
0.04	5-100	78	5-100	39	5-100	82
0.05	5-100	41	5-100	34	5-100	73
0.06	5-100	33	5-100	29	5-100	59
0.07	5-100	30	5-100	25	5-100	50
0.08	5-100	27	5-100	21	5-100	34
0.09	5-100	28	5-100	18	5-100	18
0.1	5-100	28	5-100	16	5-100	17
0.2	5	19	5	10	5	9
0.5	5	12	30-40	3	5	9
0.7	30-40	10	100	3	5	7
1	100	10	100	2	20	4

TAB. 6.23 – Nombre d’erreurs de classification pour l’expert "Trous/Usure à 75%" dans le sens ME1 avec la configuration C2 : 5 premières valeurs singulières - $C \in [5 - 100]$.

Le tableau 6.24 expose les meilleurs résultats de bonne classification obtenus avec les paramètres (σ, C) optimisés avec la procédure LOO.

CC% obtenus avec le noyau gaussien							
Experts		E1			E2		
Les capteurs		1	2	3	1	2	3
Les paramètres							
Sigma (σ)		0.7	0.7	0.7	1	1	1
C		10	30	5	100	100	20
Nb Err		19	14	29	10	2	4

TAB. 6.24 – Comparaison entre les résultats obtenus, avec les valeurs singulières, pour les deux experts E1 : "Sans défaut/Usure normale" et E2 : "Trous/Usure à 75%" dans le sens ME1 avec la configuration C2 : $C \in [5 - 100]$ et $\sigma \in [0 - 1]$.

Les taux de bonne classification ($CC\%$) correspondants, par expert sont les suivants :

- le premier expert :
 - Capteur 1 , $CC = 90.5\%$ avec $(\sigma = 0.7, C = 10)$;
 - Capteur 2 , $CC = 93\%$ avec $(\sigma = 0.7, C = 30)$;
 - Capteur 3 , $CC = 85.5\%$ avec $(\sigma = 0.7, C = 5)$;
- le deuxième expert :
 - Capteur 1 , $CC = 90\%$ avec $(\sigma = 1, C = 100)$;
 - Capteur 2 , $CC = 98\%$ avec $(\sigma = 1, C = 100)$;
 - Capteur 3 , $CC = 96\%$ avec $(\sigma = 1, C = 20)$;

Les classifieurs		MLP	RBF	SMO	SVM
Classes					
Configuration 2					
Sans défauts/Usure normale		90	87	88	90.5
Usure 75%/Trous		92	94	95	96

TAB. 6.25 – Comparaison entre les taux de classification obtenus avec les MLP, RBF, SMO et les SVMs.

Le tableau 6.25 compare les résultats obtenus par un classifieur SVM avec un noyau gaussien par rapport à ceux obtenus à l'aide des MLPs et les RBFs. Le taux de classification obtenu à l'aide des machines SVM avec le noyau polynômial ou gaussien, tableau 6.25, sont meilleurs que ceux obtenus avec les autres classifieurs RBF et SMO. Par contre ce taux de classification est pratiquement le même, 90%, obtenu avec un classifieur de type MLP. Les paramètres du noyau sont optimisés par capteur et pour chaque expert. Un couple de galets sans défaut est reconnu à 90.5% d'un couple de galets avec usures normales (trous et/ou usé à 75%). La séparation entre les couples usés à 75% et les couples avec trous sont nettement meilleure, avec un taux de 96% de reconnaissance obtenu avec un SVM contre 92%, 94% et 95% obtenus respectivement avec un MLP, un RBF et un SMO. Ces résultats confirment l'efficacité des SVMs par rapport aux autres classifieurs étudiés.

6.6.3 Classification SVM avec les pseudo-signatures de puissances

Comme pour les premiers classifieurs, nous allons dans la suite effectuer l'étape de classification en utilisant les mêmes bases de données construites à partir des pseudo-signatures de puissance. Nous nous attendons à une amélioration du taux de classification puisqu'avec les valeurs singulières, les SVMs ont fourni des résultats meilleurs par rapport aux autres classifieurs, MLP et RBF. Dans les deux cas, la classification a été effectuée sans passer par une étape de réduction de la taille du premier vecteur singulier de gauche.

Les figures 6.5 et 6.6 exposent les résultats obtenus avec le noyau polynômial, pour les deux experts. Le nombre d'erreurs de généralisation est donné pour le couple de paramètres (δ, C) avec $\delta \in [1, 2, 3, 4, 5]$ et $C \in [1 - 100]$. Pour l'expert "Sans défaut/Usure normale", les meilleurs résultats pour le premier, le deuxième et le troisième capteurs sont obtenus avec l'ensemble des paramètres $(\delta = 3, C = 5)$, $(\delta = 5, C = 40)$ et $(\delta = 2, C = 10)$ correspondant respectivement à 98%, 99% et 98% de taux de bonne classification. Pour l'expert "Trous/Usure à 75%", les paramètres $(\delta = 3, C = 5)$, $(\delta = 2, C = 10)$ et $(\delta = 1, C = 5)$ fournissent les meilleurs résultats pour les trois capteurs, avec respectivement un taux de bonne classification égal à 99% pour le capteur 1, 99% pour le capteur 2 et 99% pour capteur 3. Le taux global de classification est égal alors à 98% pour l'expert "Sans défaut/Usure normale" et 99% pour l'expert "Trous/Usure à 75%".

Le nombre d'erreurs commises par chaque expert dans ce cas est :

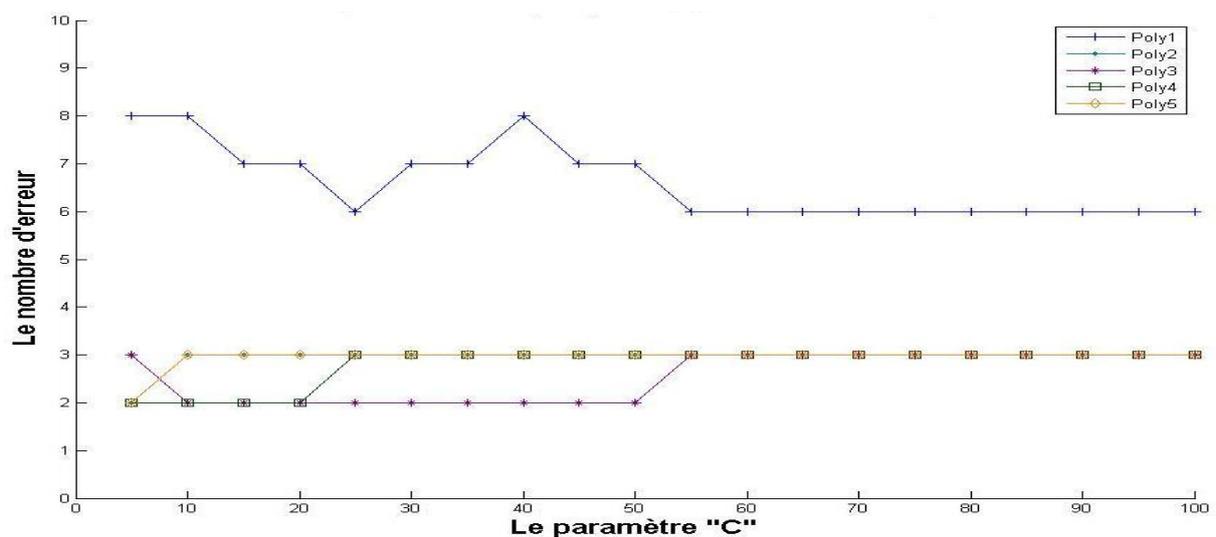
- plus lié au paramètre δ du noyau polynômial qu'au paramètre de régularisation C ;
- contrairement au premier cas, avec valeurs singulières, il reste assez statique pour les deux experts, (cf. figures 6.5 et 6.6).

Enfin, nous remarquons que les résultats obtenus avec le premier vecteur singulier de gauche sont largement meilleurs que ceux obtenus avec les valeurs singulières. Le tableau 6.26 compare les résultats dans les deux cas.

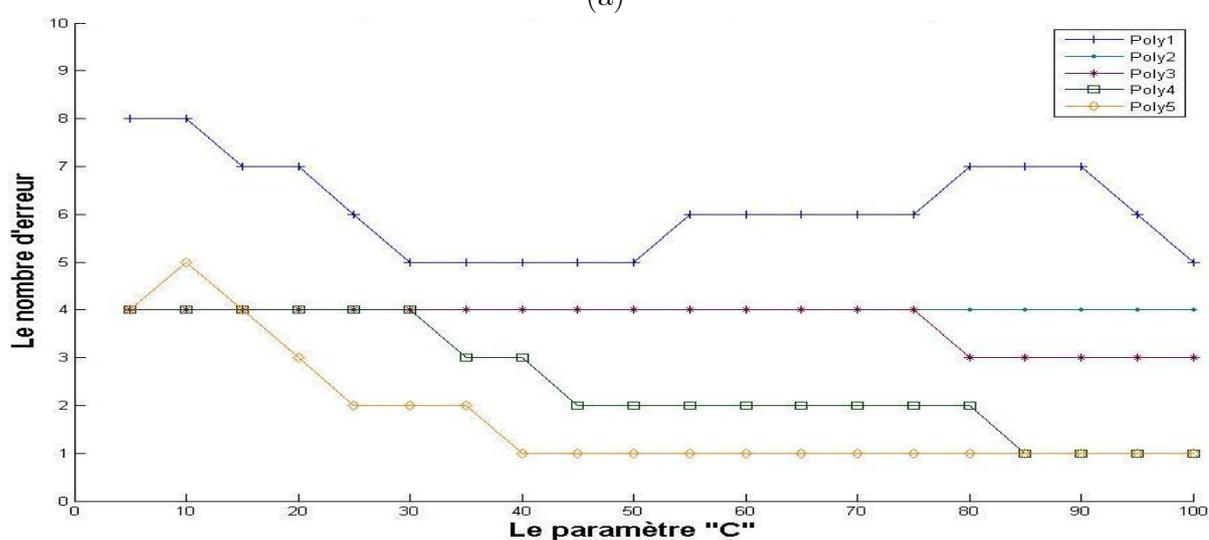
CC% obtenus avec le noyau polynômial		
CC%	E1	E2
Les données		
Sigma (σ)	90.5%	96%
U	98%	99%

TAB. 6.26 – Comparaison entre les résultats obtenus dans le sens ME1 configuration C2, avec les valeurs singulières σ , et ceux obtenus avec le premier vecteur singulier de gauche U (échelles). Avec E1 : "Sans défaut/Usure normale", E2 : "Trous/Usure à 75%", $\delta \in [1, 2, 3, 4, 5]$ et $C \in [1 - 100]$.

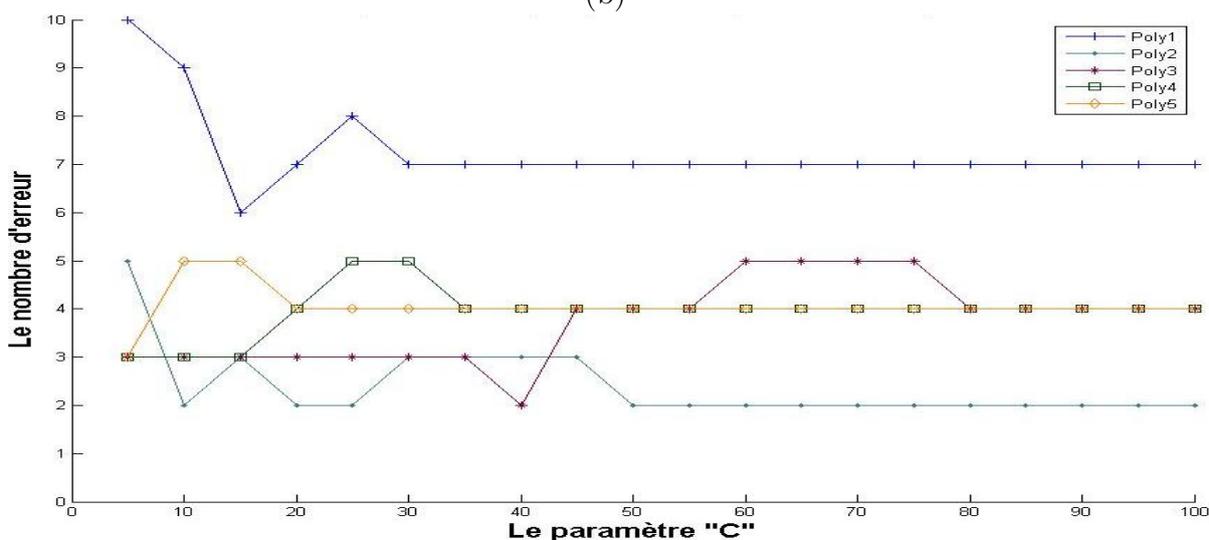
De la même manière, pour le noyau gaussien, le nombre d'erreurs est calculé pour l'ensemble de paramètres croisés $\sigma \in [0 - 1]$ et $C \in [5 - 100]$. Nous remarquons



(a)

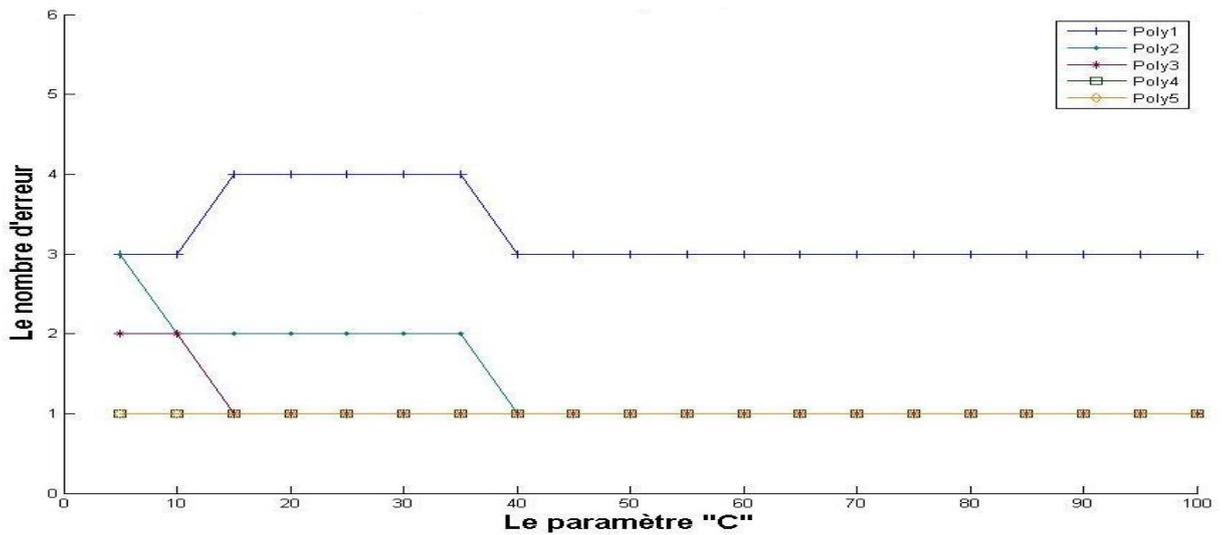


(b)

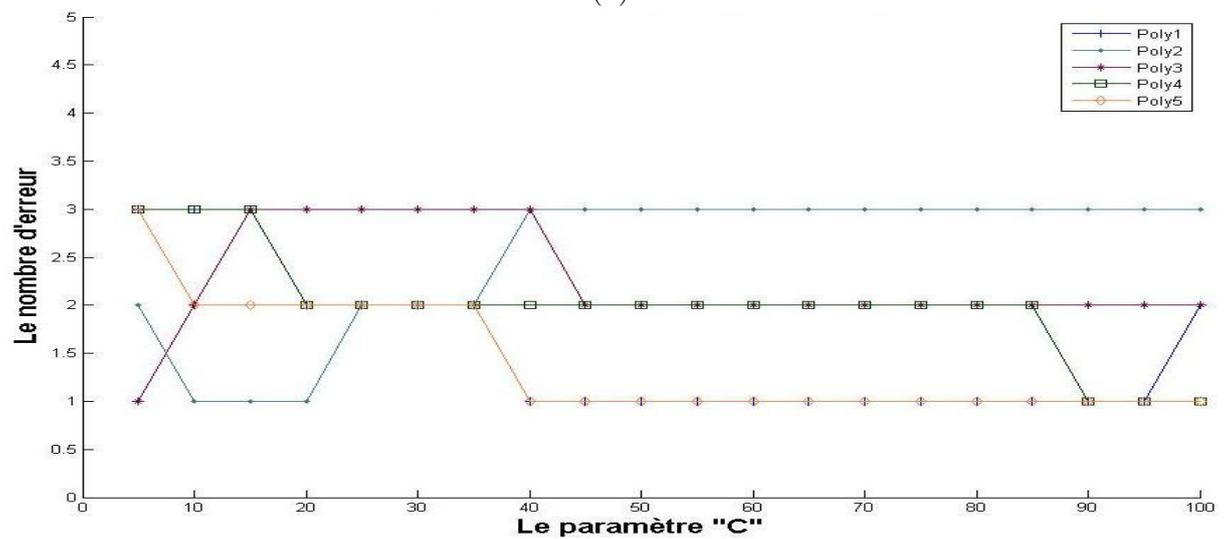


(c)

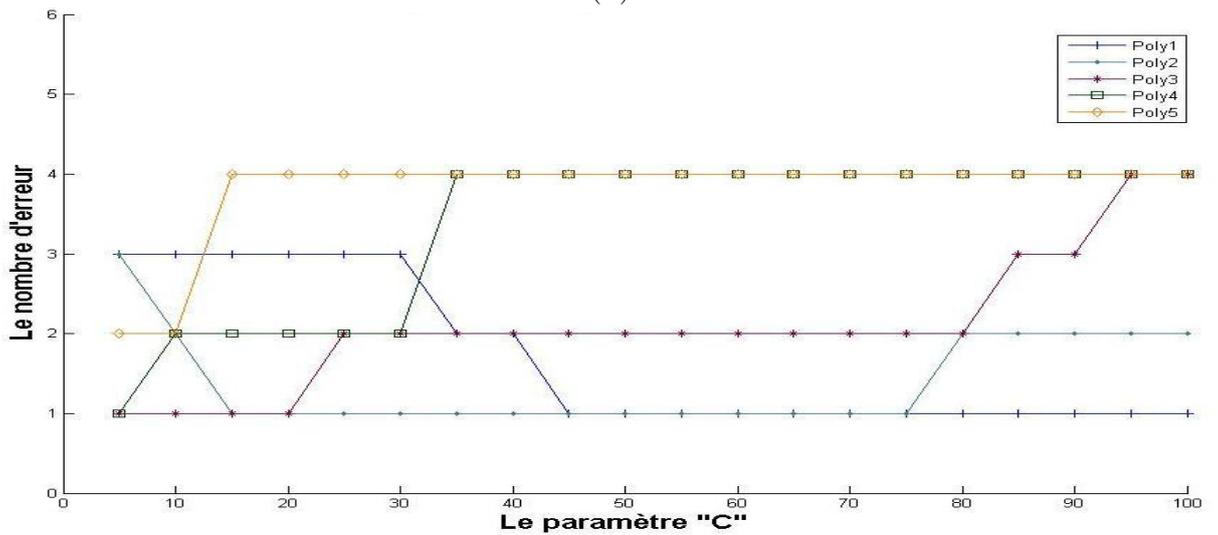
FIG. 6.5 – Le nombre d'erreurs obtenu pour le premier expert "Sans défaut/Usure normale", en utilisant le premier vecteur singulier de gauche (les échelles), avec le couple de paramètres (δ, C) pour le premier (a), le deuxième (b) et le troisième (c) capteur.



(a)



(b)



(c)

FIG. 6.6 – Le nombre d'erreurs obtenu pour le deuxième expert "Trous/Usure à 75%" , en utilisant le premier vecteur singulier de gauche (les échelles), avec le couple de paramètres (δ, C) pour le premier (a), le deuxième (b) et le troisième (c) capteur.

que les deux experts se comporte de la même manière que celle constatée dans le cas des valeurs singuliers. Pour n'importe qu'elle valeur de la largeur du noyau gaussien et quelque soit les valeurs du paramètre de pénalisation $C \in [5 - 100]$ le nombre d'erreurs reste globalement statique. Prenons par exemple le deuxième expert "Trous/Usure à 75%" pour une valeur de σ égale à 0.02 et $\forall C \in [5 - 100]$, le classifieur se trompe sur 4 exemples sur les 100 exemples de la base de données du premier capteur. Le nombre d'erreurs reste généralement statique et ne dépend que des valeurs de σ . Ce même résultat est constaté sur les 3 capteurs et les deux experts.

Le tableau 6.27 expose les meilleurs résultats de bonne classification ($CC\%$) obtenus avec les paramètres (σ, C) optimisés avec la procédure LOO. Les taux de bonne classification et le nombre d'erreurs sont fourni par expert et par capteur. L'étape de la classification a été effectuée dans l'espace initial ($p = 21$) sans passer par une étape de réduction de dimension. Les deux bases de données, Base1 = Sans défauts/Usure normale et Base2 = Trous/Usure 75%, en entrée du classifieur sont donc de dimension (200×21) et (100×21) . Le taux global de classification pour le premier expert est de 98% et pour le deuxième expert est de 99%. Ce taux de classification est le même que celui obtenu avec le noyau polynômial, (cf. 6.26).

CC% obtenus avec le noyau gaussien						
Experts	E1			E2		
Les capteurs	1	2	3	1	2	3
Les paramètres						
Sigma (σ)	0.2	0.5	1	0.2	0.5	0.5
C	5	25	50	25	15	20
Nb Err	1	4	2	1	1	1
CC%	99%	96%	98%	99%	99%	99%

TAB. 6.27 – Comparaison entre les résultats obtenus, avec le premier vecteur singulier de gauche (échelles), dans le sens ME1 configuration C2 pour les deux experts. Avec E1 : "Sans défaut/Usure normale", E2 : "Trous/Usure à 75%", $C \in [5 - 100]$ et $\sigma \in [0 - 1]$.

D'une manière générale, l'ensemble des résultats obtenus, soit avec les valeurs singulières soit avec le premier vecteur singulier de gauche, montrent l'intérêt des classifieurs SVMs par rapport aux autres classifieurs (MLPs, RBFs, etc.). Ils confirment ainsi que les SVMs permettent de répondre aux problèmes centraux de la théorie de l'apprentissage statistique cités préalablement en fin de la partie (cf. 6.5). Sa robustesse vis à vis de la dimensionalité des données de l'espace initial de classification est donc démontrée. Par conséquent, il nous semble convenable de choisir les SVMs pour la construction finale de notre système de classification. En plus de ces performances, il nous paraît plus prometteur et en tout cas mieux justifié d'un point de vue théorique.

Conclusion

Dans ce chapitre, nous avons présenté les performances des différents classifieurs testés pour le diagnostic de l'état des galets de guidage. La première partie du chapitre a été dédiée à l'étude des classes initiales : "Sans défaut", "Usure 100%", "Usure 75%", "Trous", "Usure patatoïde". Suite à un ensemble d'analyses effectuées sur ces classes seulement deux classes de défauts sont considérées recevables. Ainsi seules les classes "Sans défaut", "Usure 75%" et les "Trous" sont retenues pour la démarche finale de diagnostic. Pour effectuer l'étape de classification, l'approche multi-classes "1-Contre-1" a été mise en oeuvre. L'approche de classification "1-Contre-1" permet de construire des experts et d'optimiser ainsi les paramètres des sous classifieurs deux à deux en se basant seulement sur les deux classes prises en compte. C'est cette approche de classification multi-classe qui a été retenue pour la construction du système final de diagnostic. Les performances de plusieurs classifieurs : les MLPs (cf. 2.5.3), les RBFs (cf. 2.5.4), la méthode adaboost (cf. 2.5.5), le prédicteur bagging [36], les arbres de décision (Id3 et J48), (cf. 2.4.2), le modèle logistic, ont été ainsi comparées. Dans cette première partie, seules les valeurs singulières ont été considérées en entrée des classifieurs. Dans la majorité des cas, les MLPs ont fournies les meilleures performances. Nous avons remarqué aussi que, pour un sens de marche donné ME1 ou ME2, les taux de reconnaissance dépendent fortement du type de défaut et de son importance.

L'étape de classification a été mise en oeuvre en utilisant les deux classes : "Sans défaut" et "Usure normale". La classe usure normale regroupe les deux types de défauts "Usure 75%" et les "Trous". Deux experts, "Sans défaut/Usure normale" et "Trous/Usure 75%", ont été définis. Le diagnostic de l'état d'un couple de galets a été effectué en deux phases de classification successives. La première consistait à définir l'état des galets, sans défaut ou défectueux. Dans le cas où l'état du couple de galets est jugé défectueux, la deuxième phase de classification sert à identifier la nature du défaut : "Trous" ou "Usure à 75%".

Une fois les classes de défauts et la démarche finale de classification définis, nous avons introduit, dans la deuxième partie du chapitre, la notion des pseudo-signatures de puissances (PPS). Cette approche d'extraction de caractéristiques basée sur la décomposition en valeurs singulières et qui repose sur l'étude des pseudo-signatures de puissances (vecteurs singuliers) donne des résultats satisfaisants concernant la discrimination des signaux. De plus, elle propose un moyen de réduction de dimension assez important. Seulement le premier vecteur singulier de gauche (échelles), issu de la décomposition en valeurs singulières des matrices de coefficients d'ondelettes (CWT_k), est nécessaire pour représenter l'information utile contenue dans le plan temps-échelle. Par conséquent, ce premier vecteur singulier a été pris en compte pour définir les signatures PPS. Nous avons observé que les taux de reconnaissance étaient systématiquement plus élevés en les comparant avec ceux obtenus en utilisant les premières valeurs singulières en entrée des classifieurs.

La dernière partie du chapitre est consacrée à l'étude des classifieurs SVMs. De la même manière que pour les autres classifieurs (MLP et RBF), nous avons aussi comparé les résultats de la classification obtenus, dans un premier temps, avec

les valeurs singulières et ceux obtenus, dans un deuxième temps, avec le premier vecteur singulier de gauche. Avec les SVM le passage par une deuxième étape de réduction de dimension, pour réduire la taille du premier vecteur singulier de gauche, devient inutile. La robustesse de ces machines vis-à-vis de la dimension de l'espace de représentation les démarque des autres classifieurs tels que les MLPs et les RBFs.

Pour l'optimisation des SVMs, nous avons effectué un test croisé sur les paramètres de régularisation C et le noyau à partir des données d'apprentissage, sans passer par une étape de test. Les paramètres qui minimisent l'erreur en généralisation sont ceux retenus pour la classification. Cette première étape de test consistait à faire un balayage d'un intervalle de valeurs pour le couple de paramètres C et le noyau pour pouvoir retenir celui qui donne les meilleurs résultats de classification. Cette procédure est effectuée par expert et pour chaque capteur. Par conséquent, 6 couples de paramètres (C , Noyau) ont été définis. Nous avons remarqué que des noyaux différents lorsqu'ils sont correctement paramétrés, donnent lieu à des résultats de classification similaires.

Conclusions et perspectives

Dans cette thèse, nous avons apporté une contribution à la problématique du diagnostic dans le domaine du transport ferroviaire. Le travail mené au cours de cette thèse a permis d'obtenir un système de diagnostic automatique de l'état des galets de guidage d'un tramway sur pneumatiques, le "Translohr". Les objectifs de l'automatisation du système de diagnostic sont de détecter précisément le galet qui cause l'anomalie et d'identifier le type de défaut. La nouveauté du système de guidage par rail central, les délais imposés par les industriels ainsi que les modifications successives de l'installation n'ont pas permis la modélisation physique complète du système mécanique directement. Par conséquent, nous avons choisi d'utiliser une approche de surveillance et de diagnostic par reconnaissance de formes à partir de l'enregistrement de signaux vibratoires sur le rail. L'ensemble des signaux utilisés a été enregistré dans des conditions réelles grâce à un système d'acquisition nommé "détecteur SYSCOGA" (SYStème de Contrôle des Galets), propriété industrielle de la société Signal Développement.

Nous avons consacré le premier chapitre à la description du Translohr et de son système de guidage. Ce nouveau mode de transport utilise des galets de guidage par rail central. L'usure du bandage composite des galets peut être la source de nuisances sonores et n'implique en aucun cas - selon le constructeur - une panne du système de guidage. L'enjeu de ce travail de thèse était de proposer un système de surveillance et de diagnostic permettant d'identifier automatiquement les défauts dès leur apparition afin d'éviter des dommages qui pourraient s'aggraver et causer d'importantes nuisances voire l'arrêt de fonctionnement. Le caractère automatique de ce système est un point essentiel en termes de coûts de fonctionnement et de main d'oeuvre. Une des limitations inhérentes au projet et à son cahier des charges provenait d'un manque d'expérience et de connaissance à priori des différents types d'usures pouvant surgir dans un environnement réel. Nous avons dû nous référer à un ensemble de défauts qui ont été élaborés manuellement par les techniciens de la société Lohr Industrie. C'est un exemple parmi d'autres des difficultés rencontrées dans un travail de recherche motivé par une problématique industrielle innovante.

Le deuxième chapitre concernait l'étude des méthodes de diagnostic par reconnaissance des formes. Nous avons présenté, dans un premier temps, les deux étapes principales sur lesquelles repose tout système de surveillance, à savoir la perception et le diagnostic. Nous avons ensuite dressé un état de l'art des méthodes de diagnostic en nous concentrant plus particulièrement sur les méthodes de diagnostic par

reconnaissance des formes. La dernière phase du processus est une étape de classification supervisée les classes de défauts étant supposées connues. Nous avons rappelé les ingrédients des approches de discrimination non-paramétrique les plus utilisées dans la littérature. Il s'agissait de bien connaître les avantages et limitations des différentes méthodes envisageables.

L'usure partielle ou complète du bandage - galet défectueux - engendre un changement de la stimulation mécanique du rail par le système de guidage. Les vibrations engendrées sont généralement plus fortes que celles correspondant à un fonctionnement normal. Il s'avère que l'état d'usure d'un galet se caractérise par une signature vibratoire particulière. L'idée conductrice de ce travail était de construire un système de surveillance et d'aide à la décision permettant de déterminer à l'aide de ces signatures vibratoires si un couple de galets présente ou non un défaut.

Pour l'enregistrement des signaux vibratoires, nous avons utilisé un système d'acquisition, nommé "détecteur SYSCOGA" proposé par la société Signal Développement, partenaire de ce projet. Le détecteur SYSCOGA mesure les accélérations (tridimensionnelles) du rail lors du passage du Translohr. Deux installations de la version finale du détecteur SYSCOGA ont été testées : sur la voie et sur l'appareil de voie (ADV). Suite à des contraintes liées aux normes de sécurité ferroviaire, seule la dernière, sur l'ADV, a été retenue et implémentée sur la piste de Clermont Ferrand. L'analyse des signaux vibratoires nous a révélé plusieurs informations sur le comportement du tramway sur le comportement des couples de galets pour un sens de passage donné. Avec le recul, il apparaît qu'une des difficultés de cette thèse provenait du manque de contrôle que nous avions sur la nature des expériences à effectuer, sur la méthode d'acquisition, ainsi que sur le choix des types de défauts à considérer. Même si nous avons réussi à imposer, à mon avis tardivement, quelques points pour la bonne poursuite du projet, j'ai parfois manqué de poids dans certaines prises de décision.

Deux types de défauts ont été définis et retenus dans notre étude, locaux et structuraux. Un défaut local touche la surface du galet (bandage) et se manifeste en hautes fréquences, soit aux petites échelles inférieures à 10 cm environs. Le deuxième type de défauts se manifeste en basses fréquences et peut correspondre par exemple à une déformation de la structure du galet à l'échelle de son diamètre. Un même galet peut comporter ces deux types de défauts en même temps. Les deux composantes du signal n'évoluent pas forcément de la même manière et leur comportement varie selon qu'elles sont en hautes ou en basses fréquences. Il fallait de plus être capable de localiser dans le temps une information de nature fréquentielle. Nous avons choisi d'utiliser les méthodes temps-échelles, et c'est la transformée en ondelettes continues qui a été retenue dans notre étude. Grâce à son comportement dynamique, lié au zoom adaptatif de l'échelle, la transformée en ondelette continue permet de suivre les singularités et les instationnarités du signal analysé. Un état de l'art des méthodes temps-fréquences les plus adaptées pour l'analyse des signaux non stationnaires a fait l'objet du chapitre 4. Nous avons comparé les propriétés de la transformée en ondelettes continue à celles de deux autres méthodes, la transformée de Fourier à

fenêtre glissante et l'analyse de Wigner-Ville. La transformée en ondelette continue assure une résolution fréquentielle relative constante à toutes les fréquences, basses ou hautes, c'est la raison principale de notre choix.

Un premier découpage fréquentiel du scalogramme, issu de l'application de la transformée en ondelette au signal vibratoire, nous a permis de spécifier les bandes d'échelles utiles pour la suite des traitements. Nous avons aussi procédé à un découpage temporel pour déterminer la position exacte de chaque couple de galets sur le signal vibratoire complet. À l'issue de cette étape, à chaque couple de galets correspond une matrice de coefficients d'ondelettes (un extrait du scalogramme). Nous avons dans un premier exploré une méthode de segmentation par binarisation d'images selon la méthode d'Otsu, mais sans succès. Nous avons préféré nous tourner vers les méthodes de classification à partir de vecteurs formes issus d'une réduction de dimension basée sur la sélection et/ou l'extraction de variables.

Les trop grandes dimensions des matrices de coefficients d'ondelette, essentiellement dues à une fréquence d'acquisition trop grande (choisie par SD), les rendaient peu intéressantes en vue de la classification. Le chapitre 5 revient sur la sélection et/ou l'extraction de variables. Nous nous sommes intéressés aux méthodes d'extraction de caractéristiques et plus particulièrement à la décomposition en valeurs singulières (DVS) des matrices temps-échelle. Une fois appliquée, elle décompose la matrice de données en valeurs et vecteurs singuliers. Plusieurs travaux antérieurs ont proposé différentes manières d'exploiter ces deux informations (valeurs et vecteurs singuliers) pour la caractérisation de la matrice de données. Nous avons utilisé, dans un premier temps, les valeurs singulières, et dans un deuxième temps les vecteurs singuliers en entrée du module de classification. Un intérêt particulier a été réservé aux vecteurs singuliers. L'utilisation de ces descripteurs en classification s'est avérée très efficace.

La dernière partie du manuscrit est consacrée à la présentation de notre système complet de diagnostic automatique et d'aide à la décision. De nombreuses expérimentations menées sur nos données - matrices temps-échelle - ont permis de montrer que notre méthode, la combinaison de la décomposition en valeurs singulières et la transformée en ondelette continue, est capable de fournir des variables pertinentes pour la classification des signaux vibratoires avec plusieurs méthodes, et notamment les machines à noyau SVM. Même si cette combinaison n'est pas une nouveauté en soi, sa mise en oeuvre rigoureuse, son utilisation dans le domaine ferroviaire et la nouveauté du système de guidage étudié rend notre application tout à fait originale. Pour effectuer l'étape de classification, les performances de plusieurs méthodes ont été comparées afin de retenir la meilleure.

Plusieurs outils de classification ont été testés sur nos données. Un intérêt très particulier a été réservé aux machines à vecteurs de support (SVMs). Leurs performances ont été comparées à celles des méthodes connexionnistes, tels que les MLPs et les RbFs. Le point intéressant concernant les SVMs est quelles effectuent systé-

matiquement une réduction de données en entrées pour résoudre un grand nombre de problèmes de classification. Cela est obtenu grâce au processus de projection des données de l'espace initial à un espace augmenté de dimension supérieure. Il ressort de notre étude que les meilleurs résultats ont été obtenus avec les SVMs. Pour l'optimisation des SVMs, nous nous sommes intéressés à l'optimisation du paramètre de régularisation C et des paramètres du noyau pour la sélection du meilleur modèle. Globalement la valeur du paramètre C sert à fixer le compromis entre la minimisation de l'erreur d'apprentissage et la maximisation de la marge. L'influence de ce paramètre s'est avérée limitée dans notre contexte.

L'étape de classification a été exécutée grâce à un ensemble de sous classifieurs avec l'approche "Un-Contre-Un". En entrée des classifieurs, nous avons utilisé les valeurs singulières puis, dans un deuxième temps, les vecteurs singuliers. Les résultats obtenus sont très satisfaisants et sont exposés dans la partie 6.5.2.

6.6.4 Perspectives

Le travail effectué dans cette thèse laisse un nombre de perspectives ouvertes. Malgré les difficultés rencontrées au démarrage du projet, et tout au long de cette thèse, nous nous sommes appliqués à ce que notre système réponde aux exigences imposées par les industriels. Un stage ingénieur de 5 mois a d'ailleurs eu lieu en 2006 avec pour but de fournir une maquette logicielle complète de la chaîne de traitement, indépendante de tout logiciel commercial. Une maquette d'interface destinée à la T2C (Transports en Commun Clermontois) a aussi été élaborée. La responsabilité du développement final et de la mise en oeuvre industrielle revient à la société Signal Développement.

Il est probable que le signal vibratoire peut être exploité pour comprendre le comportement d'autres composantes du système de guidage et peut-être même du système de roulement du tramway. De manière générale, comme beaucoup d'application de la reconnaissance des formes au diagnostic, nous nous sommes heurtés au problème de l'insuffisance de la connaissance a priori sur le système étudié. Dans une application réelle, certains défauts ne peuvent pas être étudiés pour des raisons de sécurité d'une part et de coûts d'autre part. En effet, le manque d'un couple de galets, qui est un état dangereux, constitue un vrai état de défaillance et peut certainement provoquer des dommages irréversibles. Malheureusement, nous n'avons pas pu tester notre système de diagnostic sur ce type de défaut. Il pourrait être judicieux de prendre en compte ce type de défaut dans les futures applications.

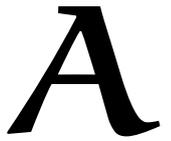
Idéalement, un processus de diagnostic par reconnaissance des formes doit être adaptatif et prédictif. On pourra souhaiter passer du diagnostic simple et instantané à un diagnostic tenant compte de l'historique du système. L'adaptabilité du système permettrait de suivre l'évolution d'un défaut d'une classe à une autre. Il peut s'agir d'une évolution vers une classe connue ou inconnue. La plupart du temps l'évolution du système sera prévisible, par exemple d'une classe "Usure à 25%" à une classe "Usure à 50%" et ainsi de suite jusqu'à atteindre le degré maximal donc l'"Usure

à 100%". Il serait bon que le système de diagnostic puisse prendre en compte ce comportement évolutif du défaut. L'apprentissage d'un processus répétitif peut être utilisé pour la prédiction et aussi pour la prise de décision de changer un couple de galets avant d'atteindre l'usure maximale. Quand à l'Usure à 100%, on peut supposer par exemple qu'elle engendre une fissure du pivot et à terme sa casse. Dans ce cas l'évolution de l'usure correspond à la création d'une nouvelle classe de défaut. En effet, le système de diagnostic doit être capable, dans ce cas, de détecter ce nouvel état ce qui va permettre à l'expert humain d'effectuer une autre phase d'apprentissage afin d'enrichir la base connaissance a priori sur le système de guidage. Cette notion d'adaptabilité devient intéressante dans le cas où un ensemble d'état d'usures réelles sont définis. Cela nécessite d'avoir au moins un cycle de vie d'un galet, ce qui n'est pas le cas dans notre application.

En plus d'être évolutif, le système de diagnostic doit aussi être capable d'anticiper l'évolution d'un mode de fonctionnement normal, galets sans défauts, vers un mode anormal, galets défectueux : on accèderait alors à une fonction de pronostic. Cet aspect prédictif permet d'éviter l'apparition des dommages irréversibles voire l'arrêt du tramway. Nous pensons que la connaissance de la probabilité qu'un état de fonctionnement évolue d'une classe à une autre, peut être utilisé afin de prédire son évolution. Les chaînes de Markov cachées (HMM) par exemple peuvent être combinées au classifieur utilisé, les SVMs, pour pouvoir calculer la matrice des probabilités. Si une "Usure à 100%" provoque, par exemple, à 80% la casse du pivot on peut décider de changer le couple de galets usés une fois cette usure est atteinte ou bien avant. Le système de diagnostic, prend en compte dans ce cas le passé du système surveillé pour pouvoir prédire l'état ou la classe future. (à valider par Alex)

Nous pensons que la gestion du caractère évolutif de l'usure dans le cas d'un système dynamique - le système de guidage - pour pouvoir faire un pronostic serait très intéressant. Ce travail de thèse peut être considéré comme un premier travail sur ce nouveau système de guidage et pourrait certainement être adapté à d'autres situations similaires, pour d'autres systèmes sur rail par exemple.

PLANNING PRÉVISIONNEL DU PROJET SYSCOGA



- **Chef de projet LOHR** : Jean Luc ANDRE ;
- **Responsable projet SD** : Olivier COSTE ;
- **Responsable projet SMTC** : Patrice CHARLAT ;
- **Doctorant LIMOS et chargé de mission SMTC** : Zahra HAMOU MAMAR ;
- **Encadrement scientifique LIMOS** : Alexandre AUSSEM et Pierre CHAINAIS.

Le planning prévisionnel donné en Octobre 2003

Phase 1 : Développement de la méthode "classique"

- Deux campagnes d'une semaine d'essai à Duppigheim, Strasbourg (LOHR et SD) ;
- Disponibilité du TRANSLOHR (LOHR) ;
- Fourniture des accéléromètres (LOHR) ;
- Mise en place des capteurs sur rail (LOHR) ;
- Préparation des galets avec défauts (LOHR) ;
- Réalisation des mesures (SD) ;
- Analyse des mesures et mise au point de la méthode classique (SD) ;
- Analyse par des méthodes statistiques "évoluées" (DOCTORANT).

Phase 2 : Suivi de la thèse : " méthodes statistiques évoluées"

- 1 réunion par mois sur 3 ans (SD et DOCTORANT) ;
 - Rédaction de la thèse (DOCTORANT) ;
 - Soutenance de la thèse prévue pour le dernier trimestre 2006 (DOCTORANT).
-

Phase 3 : Développement du détecteur prototype Syscoga

- Spécifications matériel (cartes, capteurs, etc.) (SD) ;
- Achats de matériel (cartes, capteurs, etc.) (SD) ;
- Montage et test en labo du prototype (SD) ;
- Intégration éventuelle de la méthode " évoluée " dans le détecteur (SD et DOCTORANT) ;
- Intégration du détecteur prototype dans le système TRANSLHOR (fixation mécanique sur le rail, câblage, communication des résultats) (LOHR).

Phase 4 : Mise en place du détecteur Syscoga à Clermont-Ferrand,

- Mise en service du détecteur prototype (LOHR et SD),
- Suivi et retour d'expérience (LOHR et SD).

Phase 5 : Généralisation aux autres sites TRANSLOHR,

- Commercialisation (LOHR),
 - Intégration mécanique, communication des défauts (LOHR),
 - Fourniture " matériel et logiciel ", mise en service du détecteur (SD).
-

CAMPAGNES D'ESSAIS

B.1 Essais Syscoga-II

La campagne d'essais syscoga-II a eu lieu les 03 et 04 mars 2004, avec le démonstrateur translohr, figure B.1(a), et les même configurations de montage des galets par essieu utilisées dans syscoga-I. L'objectif étant de pallier les difficultés rencontrés lors de la première campagne d'essai. La configuration utilisée est composée de deux cellules photoélectriques posées sur la chaussée, figure B.1(b), et six accéléromètres (3 latéraux et 3 verticaux) fixés sous la semelle du rail, figure B.1(c). Chaque cellule est composée d'un émetteur avec faisceau infrarouge invisible, placé entre le rail et la roue, et d'un réflecteur placé à l'extérieur de la roue, figure B.1(a). L'idée d'utiliser les signaux acoustique comme information sur l'état des galets de guidage en plus de l'information vibratoires a été abandonnée lors de ces essais.

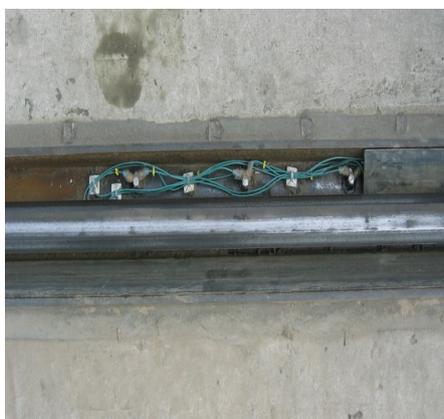
B.2 Essais Syscoga-IIbis

La campagne d'essais syscoga-IIbis a été organisée par la société SD en mois de juillet 2004. L'objectif étant de tester une nouvelle configuration matérielle et de trouver une bonne isolation par rapport au courant circulant dans le rail. Pour mesurer de vibrations du rail, trois types de capteurs accélérométriques ont été utilisés :

- un accéléromètre tri-axe comportant une pastille isolante, vissé sur le rail,
 - un accéléromètre bi-axe, collé sur une pastille de mica elle-même collée sur le rail,
 - un accéléromètre mono-axe, collé directement sur le rail, figure B.2
 - deux barrières optiques, chacune composée d'une cellule photoélectrique et d'un réflecteur positionnées au ras du sol pour détecter le passage des galets/roues.
-



(a)



(b)



(c)

FIG. B.1 – (a) Le Translohr pour les essais Syscoga-II. (b) Positions des accéléromètres sur le rail. (c) Le système d'acquisition sur site : Installation sur la voie.

B.3 Essais Syscoga-IV

Cette campagne d'essais a eu lieu du 04-06 Avril 2005, la configuration matérielle utilisée est la même que celle de syscoga-III avec un test d'identification de la rame et la possibilité de l'intégrer dans le logiciel finale de diagnostic, figure B.3

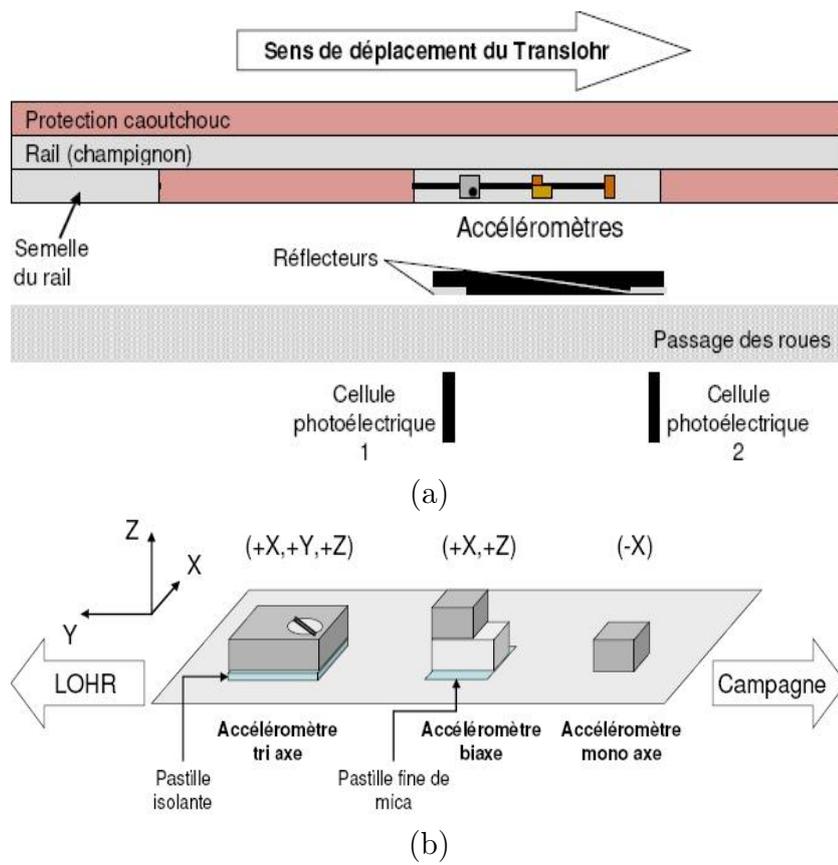


FIG. B.2 – (a) Installation matérielle sur la voie : les essais Syscoga-IIbis. (b) Positionnement des accéléromètres sur la voie.

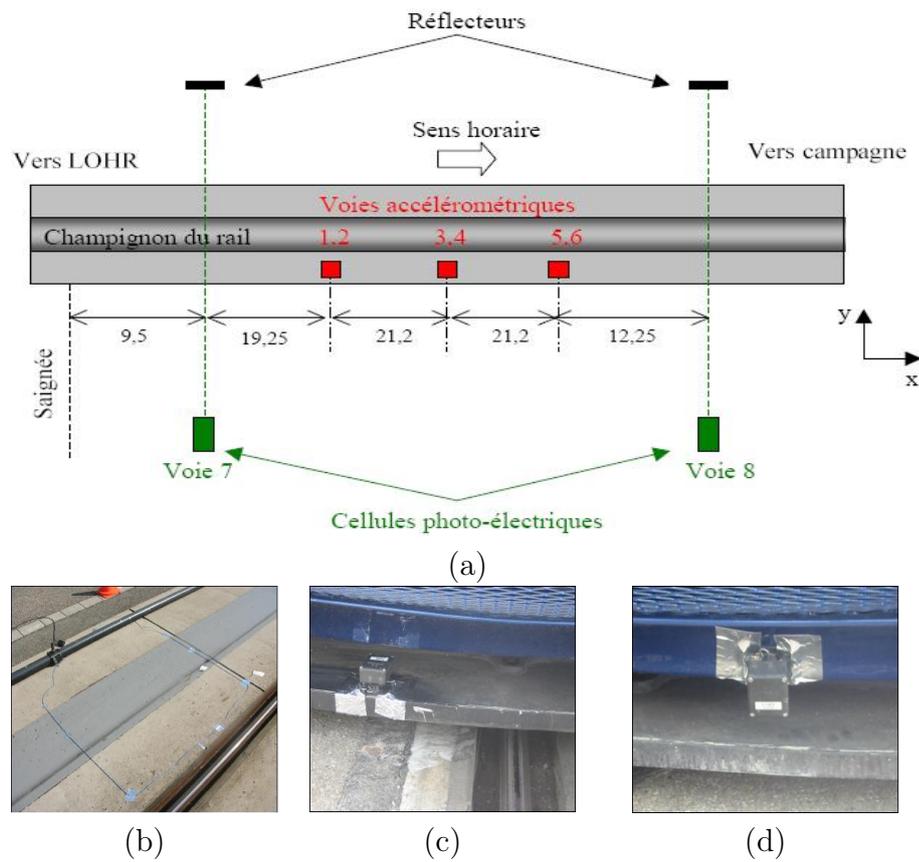


FIG. B.3 – (a) Installation matérielle pour les essais Syscoga-IV. (b) Emplacement en pleine voie de la boucle coaxiale du système d'identification du numéro de la rame. (c) et (d) Deux transpondeurs montés à l'avant et à l'arrière de la rame.

LES MESURES DE L'ERREUR ESTIMÉE POUR UN CLASSIFIEUR

C

Lorsqu'on construit un classifieur, on peut en même temps calculer un ensemble de mesures de l'erreur estimée. Ces mesures sont classées dans deux catégories : les mesures générales et les mesures d'exactitude par classe.

C.1 Les mesures générales

1. *Le taux de bonne classification* (Correctly Classified Instances (CC))

C'est le nombre d'exemples bien classés en pourcentage du nombre total d'exemples.

2. *Kappa statistic* (κ)

Le coefficient κ est censé mesurer le degré de concordance de deux ou de plusieurs juges. Dans notre cas, on est toujours dans le cas de deux juges. On mesure la différence entre l'accord constaté entre les deux juges, et l'accord qui existerait si les juges classaient les exemples au hasard. Le jugement, c'est la classe d'un exemple, et les deux juges sont le classifieur et la classe réelle de l'exemple. "L'accord/désaccord" entre les deux juges se lit directement dans la matrice de confusion. C'est une mesure dont la valeur est d'autant plus grande que la matrice est diagonale.

Le coefficient κ se calcule de la façon suivante

$$\kappa = \frac{P_o - P_e}{1 - P_e}, \quad (\text{C.1})$$

où P_o représente la proportion de l'échantillon sur laquelle les deux juges sont d'accord (i.e. la diagonale principale de la matrice de confusion), et

$$P_e = \frac{\sum_i p_i - p_{.i}}{n^2}, \quad (\text{C.2})$$

où

- p_i : somme des éléments de la ligne i ;
- $p_{.i}$: somme des éléments de la colonne i ;
- n : taille de l'échantillon.

Le coefficient κ prend ses valeurs entre -1 et 1

- il est maximal quand les deux jugements sont les mêmes : tous les exemples sont sur la diagonale, et $P_o = 1$;

- il vaut 0 lorsque les deux jugements sont indépendants ($P_o = P_e$);
- il vaut -1 lorsque les juges sont en total désaccord. Certains auteurs (Landis et Koch) ont proposé une échelle de degré d'accord selon la valeur du coefficient κ .

Accord	κ
Excellent	>0.81
Bon	$[0.80-0.61]$
Modéré	$[0.4-0.41]$
Médiocre	$[0.4-0.21]$
Mauvais	$[0.20-0]$
Très mauvais	<0

Sur un exemple dont la matrice de confusion est

a	b	\Leftarrow classer comme
49	1	a = 1
1	49	b = 2

TAB. C.1 – Exemple de matrice de confusion.

On a

$$P_o = \frac{49 + 49}{100} = \frac{98}{100} = 0.98$$

$$P_e = \frac{(50 * 50) + (50 * 50)}{100^2} = \frac{1}{2}.$$

Cette valeur est constante pour tous les classifieurs sur cet ensemble d'exemples et donc

$$\kappa = \frac{0.98 - 0.5}{1 - 0.5} = 0.96.$$

L'accord dans ce cas est excellent.

C.2 Les mesures d'exactitude par classe

Pour une classe donnée, un classifieur, et un exemple, quatre cas peuvent se présenter :

1. l'exemple est de cette classe, et le classifieur ne se trompe pas : c'est un *vrai positif* (TP);
2. l'exemple est de cette classe, mais le classifieur se trompe : c'est un *faux négatif* (FN);
3. l'exemple n'est pas de cette classe, mais le classifieur la lui attribue quand même : c'est un *faux positif* (FP);

	Classe prédite w_i	
	prédite	non prédite
Exemple $x_i \in w_i$	TP	FN
Exemple $x_i \notin w_i$	FP	TN

TAB. C.2 – Les quatre cas possible pour une classification binaire.

4. l'exemple n'est pas de cette classe, et le classifieur ne le range pas non plus dans cette classe : c'est un *vrai négatif* (TN). Le tableau C.2 résume les différentes situations.

D'une manière générale les résultats d'un classifieur sur des données test sont résumés dans la matrice de confusion (coïncidence) :

Classe prédite	Classe réelle	
	TP	FP
	FN	TN

TAB. C.3 – La structure d'une matrice de confusion.

1. **Rapport des vrais positifs** (TP Rate)

C'est le rapport entre le nombre des exemples bien classés et le nombre total d'éléments qui devraient être bien classés. Il correspond à

$$\begin{aligned} \text{TP Rate} &= \frac{\text{Nombre de vrais positifs}}{\text{Nombre de vrais positifs} + \text{Nombre de faux négatifs}}, \\ &= \frac{\text{TP}}{\text{TP} + \text{FN}}. \end{aligned}$$

Sensibilité = TP Rate.

Si les exemples d'une classe donnée sont bien classés, pas de faux négatifs, donc le **TP Rate** = 1.

2. **Rapport des faux positifs** (FP Rate)

Il correspond à

$$\begin{aligned} \text{FP Rate} &= \frac{\text{Nombre de faux positifs}}{\text{Nombre de faux positifs} + \text{Nombre de vrais négatifs}}, \\ &= \frac{\text{Nombre de faux positifs}}{\text{Nombre d'exemples n'étant pas de cette classe}}, \\ &= \frac{\text{FP}}{\text{FP} + \text{TN}}. \end{aligned}$$

$$\text{Spécificité} = 1 - \text{FP Rate} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

La matrice de confusion permet de calculer TP Rate et FP Rate et vis versa. Dans la terminologie de l'apprentissage supervisé, cette matrice connue sous

les termes "matrice d'erreur", "tableau de contingence" ou "matrice d'erreur de classification", est un outil servant à mesurer la qualité d'un système de classification. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée, et chaque ligne représente le nombre d'occurrences d'une classe réelle.

Prenons l'exemple du tableau C.1. Les 49 exemples de la classe a sont bien classés mais le dernier est mal classé donc $TP\ Rate = \frac{49}{50} = 0.98$. Un exemple est mal classé parmi ceux appartenant à la classe a et 49 exemples qui n'appartiennent pas à la classe a n'ont pas été reconnus comme tels $FP\ Rate = \frac{1}{50} = 0.02$.

3. *La précision*

C'est le rapport entre le nombre de vrais positifs et la somme des vrais positifs et des faux positifs.

$$\text{précision} = \frac{TP}{TP + FP},$$

si elle vaut 1, cela veut dire que tous les exemples classés positifs l'étaient vraiment ;

4. *Le rappel*

$$\text{Rappel} = \frac{TP}{TP + FN},$$

si le rappel vaut 1 cela signifie que tous les exemples positifs ont été trouvés ;

5. *La F-Measure*

Cette quantité permet de regrouper en un seul nombre les performances du classifieur pour une classe donnée. Cette mesure est donnée par

$$\begin{aligned} \text{F-Measure} &= \frac{2 * \text{Rappel} * \text{Précision}}{\text{Rappel} + \text{Précision}}, \\ &= \frac{2 * TP}{2 * TP + FP + FN}. \end{aligned}$$

Le coefficient κ peut être ainsi calculé à partir de cet ensemble de mesures comme suit

$$\kappa = \frac{2(TP * TN + FP * FN)}{[(TP + FN)(FN + TN) + (TP + FP)(FP + TN)]}.$$

6. *La courbe ROC*

Un graphe ROC (Receiver Operating Characteristic) représente le compromis entre les bénéfices (Vrais Positifs) et les pertes (Faux Positifs). Elle permet donc d'étudier les variations de la spécificité et de la sensibilité d'un test pour différentes valeurs du seuil de discrimination. Le terme de courbe ROC peut être envisagé comme une "courbe de caractéristiques d'efficacité". Chaque classifieur discret produit un point (taux FP,taux VP), figure C.1.

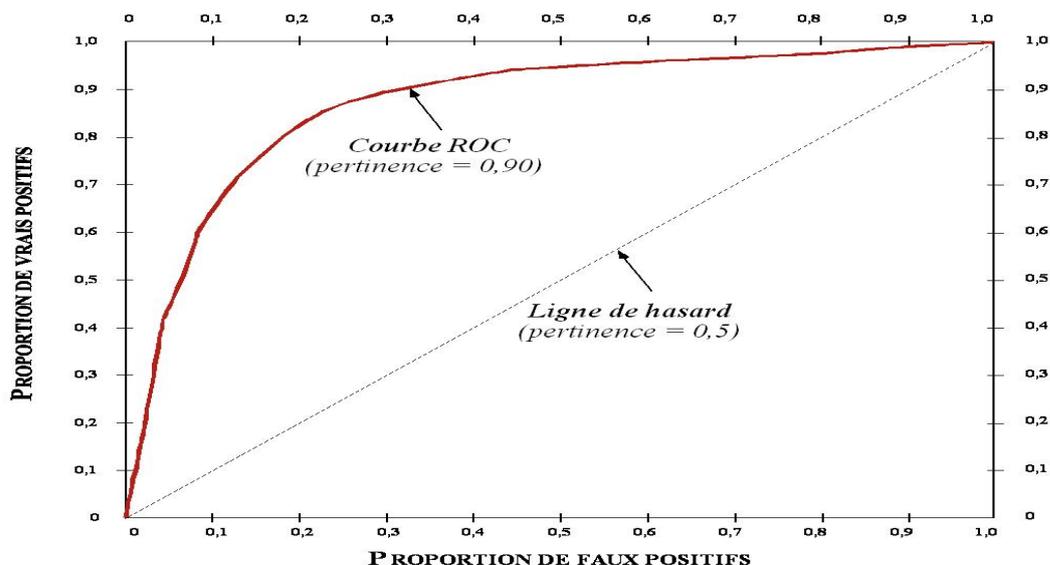


FIG. C.1 – La courbe ROC.

Type d'erreurs :

- erreur de type 1 (alpha), les faux positifs : probabilité d'accepter l'hypothèse alors qu'elle est fausse ;
- erreur de type 2 (beta), les faux négatifs : probabilité de rejeter l'hypothèse alors qu'elle est vraie ;
- comment arbitrer entre ces types d'erreurs ? \Rightarrow la courbe ROC.

L'espace ROC :

- (i) Le point (0,0) prédit toujours négatif : pas de FP ni de VP ;
- (ii) (1,1) prédit toujours positif, donc toute instance est soit FP soit VP ;
- (iii) (0,1) : classification parfaite ;
- (iv) La diagonale (p,p) prédit au hasard positif avec probabilité p ;
- (v) Le triangle sous la diagonale est pire que le hasard. Il est donc souvent vide dans un graphe ROC.

Bibliographie

- [1] R. C. Dubes A. K. Jain and C.-C. Chen. Bootstrap techniques for error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5) :628–633, 1987.
 - [2] R. P. W. Duin A. K. Jain and J. Mao. Statistical Pattern Recognition : A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1) :4–37, 2000.
 - [3] N. M. Laird A. P. Dempster and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1) :1–38, 1977.
 - [4] P. Abry. *Ondelettes et Turbulences*. Diderot Multimedia, Paris, 1997.
 - [5] D. W. Aha and R. L. Bankert. A comparative evaluation of sequential feature selection algorithms. In *D. Fisher and H. Lenz, eds, Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, pages 1–7, Fort Lauderdale, Florida, 1995.
 - [6] H. Almuallim and T. G. Dietterich. Efficient algorithms for identifying relevant features. Technical Report 92-30-03, Oregon State University, Corvallis, OR, USA, 1992.
 - [7] H. Almuallim and T. G. Dietterich. Learning boolean concepts in the presence of many irrelevant features. *Artificial Intelligence*, 69(1-2) :279–305, 1994.
 - [8] R. Altes. Detection Estimation and Classification with Spectrogram. *Journ. Acoust. Soc. Am.*, 67(4) :1232–1246, 1980.
 - [9] J. L. Alty and L. I. Mikulich. *Industrial applications of artificial intelligence*. Kluwer Academic Publishers, 2nd ed., 1991.
 - [10] P.O. Amblard. *Non, non et non, ou quelques études sur et autour des non-propriétés en traitement du signal*. Habilitation à diriger des recherches, INPG, mai 2001.
 - [11] J.C. Cexus et Z. Saidi A.O. Boudraa, C. Pollet. Caractérisation des fonds marins par décomposition modale empirique. In *20ème Colloque GRETSI*, pages 559–562, Septembre 2005.
 - [12] J. L. Aravena and F. Chowdhury. Fault detection of flight critical systems. In *in Proc. 20th Digital Avionics Systems Conference*, volume 1, pages 1C6/1 – 1C6/10, October 2001.
 - [13] S. Assous. *Analyse temps-fréquence par la transformée en S et interprétation des signaux de fluxmétrie laser Doppler : Applications au diagnostic clinique*. Thèse de doctorat, École Nationale Supérieure d’Arts et Métiers, Centre d’Angers, 2005.
-

-
- [14] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Processing*, 43(5) :1068–1089, 1995.
- [15] S. Aviyente and W.J. Williams. Information bounds for random signals in time-frequency plane. In *Proceedings of the IEEE ICASSP-01*, volume 6, pages 3549–3552, Los Alamitos, CA, USA, 2001. IEEE Computer Society.
- [16] N. Ayat. *Sélection de modèle automatique des machines à vecteurs de support : application à la reconnaissance d'images de chiffres manuscrits*. Thèse de doctorat, l'école de technologie supérieure, Montréal, Canada, 2003.
- [17] R. Baker and A. K. Jain. On feature ordering in practice and some finite sample effects. In *Proceedings of the Third International Joint Conference on Pattern Recognition*, pages 45–49, 1976.
- [18] R. G. Baraniuk and D. L. Jones. A signal-dependant time-frequency representation : Optimal kernel design. *IEEE Transactions on Signal Processing*, 41(4) :1589–1601, 1993.
- [19] H. Barlow. Unsupervised learning in neural computation. *Neural Computation*, 1 :295–311, 1989.
- [20] M. Basseville and I. V. Nikiforov. *Detection of abrupt changes - Theory and Application*. Prentice Hall Inc., 1993.
- [21] A. David Bell and Hui Wang. A formalism for relevance and its application in feature subset selection. *Machine Learning*, 41(2) :175–195, November 2000.
- [22] R. Bellman. *Adaptive Control Processes*. Princeton University Press, New-Jersey, 1961.
- [23] M. Ben-Bassat. Pattern recognition and reduction of dimensionality. In *P. R. Krishnaiah and L. N. Kanal, editors, Handbook of Statistics-II*, pages 773–791. New York : North Holland, 1968.
- [24] M. Bentoumi. *Diagnostic de systèmes non linéaires par réseaux de neurones - Application à la détection temps-réel de défauts de rail*. Thèse de doctorat, Université Henri Poincaré de Nancy, 2004.
- [25] M. Bentoumi, G. Millérioux, G. Bloch, L. Oukhellou, and P. Akinin. Classification de défauts de rail par svm, 2004.
- [26] M. Bereau. *Contribution à la théorie des sous-ensembles flous à la règle de discrimination des k plus proches voisins en mode partiellement supervisé*. Thèse de doctorat, Université de Technologie de Compiègne, 1986.
- [27] E. Bernauer. *Les réseaux de neurones et l'aide au diagnostic : un modèle de neurones bouclés pour l'apprentissage de séquences temporelles*. Thèse de doctorat, LAAS/Toulouse, 1996.
- [28] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [29] V. Blanz, B. Schölkopf, H.H. Bulthoff, C. Burges, V. Vapnik, and T. Vetter. Comparison of view-based object recognition algorithms using realistic 3d models. In *ICANN 96 : Proceedings of the 1996 International Conference on Artificial Neural Networks*, pages 251–256, London, UK, 1996. Springer-Verlag.
-

-
- [30] A. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2) :245–271, 1997.
- [31] A. Boashash. *Time-Frequency Signal Analysis and Processing : A Comprehensive Reference*. Elsevier Ltd., Oxford, UK, 2003.
- [32] B. Boashash. *Time-Frequency Signal Analysis-Methods and Applications*. Longman-Cheshire, Melbourne and John Wiley Halsted Press, New York, 1992.
- [33] B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In *COLT '92 : Proceedings of the fifth annual workshop on Computational Learning Theory*, pages 144–152, New York, NY, USA, 1992. ACM.
- [34] N. Boudaoud. *Conception d'un système de diagnostic adaptatif en ligne pour la surveillance des systèmes évolutifs*. Thèse de doctorat, Université de Technologie de Compiègne, 1997.
- [35] E. Boutleux. *Diagnostic et suivi d'évolution de l'état d'un système par reconnaissance des formes floues. Application au modèle du réseau téléphonique français*. Thèse de doctorat, Université de Technologie de Compiègne, 1996.
- [36] L. Breiman. Bagging predictors. *Machine Learning*, 24(2) :123–140, 1996.
- [37] E. Oran Brigham. *Fast Fourier Transform and its Applications*. Prentice Hall, 1988.
- [38] M. Brunet, D. Jaume, M. Labarrere, A. Rault, and M. Verge. *Détection et diagnostic de pannes -Approche par modélisation, Traité des nouvelles technologies, série Diagnostic et Maintenance*. Hermès, Paris, 1990.
- [39] C.J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2) :121–167, 1998.
- [40] P. Burman. A Comparative Study of Ordinary Cross-Validation, V-fold Cross-Validation and the Repeated Learning-Testing Methods. *Biometrika*, 76(3) :503–514, 1989.
- [41] J. Peng C. Domeniconi and D. Gunopulos. Locally Adaptive Metric Nearest-Neighbor Classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9) :1281–1285, 2002.
- [42] N. Martin C. Hory and A. Chehikian. Spectrogram Segmentation by Means of Statistical Features for Non-Stationary Signal Interpretation. *IEEE Trans. on Signal Processing*, 50(12) :2915–2925, 2002.
- [43] M. A. Casey. *Auditory group theory : with application to statistical basis methods for structured audio*. Phd thesis, Massachusetts Institute of Technology Media Laboratory (MIT), Norwich, U.K., 1998.
- [44] M. Chabert. *Détection et estimation de ruptures noyées dans un bruit multiplicatif. Approches classiques et temps-échelle*. Thèse de doctorat, E.N.S.E.E.I.H.T., Toulouse, 1997.
- [45] E. Chassande-Mottin. *Méthodes de réallocation dans le plan temps-fréquence pour l'analyse et le traitement de signaux non stationnaires*. Thèse de doctorat, Université de Cergy-Pontoise, 1998.
- [46] J. N. Chatain. *Diagnostic par système expert, Traité des nouvelles technologies, série Diagnostic et Maintenance*. Hermès, Paris, 1993.
-

-
- [47] Yong-Qing Cheng and al. Human face recognition method based on the statistical model of small sample size. In *SPIE Proceedings of the Intell. Robots and Computer Vision*, volume 1607, pages 85–95, 1991.
- [48] H. Choi and W. Williams. Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 37(6) :862–871, 1989.
- [49] C. K. Chow. On Optimum Recognition Error and Reject Tradeoff. *IEEE Trans. On Information Theory*, 16(1) :41–46, 1970.
- [50] F. Chowdhury and J.L. Aravena. A modular methodology for fast fault detection and classification in power systems. *IEEE International Conference on Control Sys. Tech.*, 6(5) :623 – 634, September 1998.
- [51] M. Coates. *Time-Frequency Modelling*. Phd thesis, University of Cambridge, UK, 1998.
- [52] L. Cohen. Time-Frequency Distributions : A Review . *Proc. of IEEE*, 77(7) :941–981, July 1989.
- [53] L. Cohen. *Time-Frequency Analysis*. Prentice Hall, New York, 1995.
- [54] T. M. Cover and J. M. Van Campenhout. On the possible orderings in the measurement selection problem. *IEEE Transactions on Systems Man and Cybernetics*, SMC-7 :657–661, October 1977.
- [55] T.M. Cover and P.E. Hart. Nearest Neighbor Pattern Classification. *IEEE Trans. On Information Theory*, 13(1) :21–27, 1967.
- [56] N. Cristianini and J. Shawe-Taylor. *An Introduction To Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.
- [57] L. N. Cooper D. L. Reilly and C. Elbaum. A Neural Model of Category Learning. *Biological Cybernetics*, 45 :35–41, 1982.
- [58] L. Personnaz D. Price, S. Kneer and G. Dreyfus. Pairwise Neural Network Classifiers With Probabilistic Outputs. *Neural Information Proceedings Systems. The MIT Press*, 7 :1109–1116, 1995.
- [59] K. C. Daly, E. Gai, and J. V. Harrison. Generalized likelihood test for FDI in redundant sensor configuration. *Journal of Guidance, Control and Dynamics*, 2(1) :9–17, 1979.
- [60] M. Dash and H. Liu. Feature selection for classification. In *Intelligent Data Analysis (IDA 1997) : An International Journal*, volume 1(3), pages 131–156. Elsevier, 1997.
- [61] I. Daubechies. *Ten Lectures On Wavelets*. SIAM, Philadelphia, PA, 1992.
- [62] S. Davies and S. Russell. Np-completeness of searches for smallest possible feature sets. In *Proceedings of the AAAI Fall'94 Symposium on Relevance*, pages 37–39. AAAI Press, New Orleans, 1994.
- [63] A. C. Davison and D. Kuonen. An introduction to the bootstrap with applications in R. *Statistical Computing and Graphics Newsletter*, 13 :6–11, 2002.
- [64] M. Davy. *Noyaux optimisés pour la classification dans le plan temps-fréquence – Proposition d'un algorithme constructif et d'une référence bayésienne basés sur les méthodes MCMC – Application au diagnostic d'enceintes acoustiques*. Thèse de doctorat, Université de Nantes, September 2000.
-

- [65] Association Française de Normalisation. Norme afnor. maintenance - concepts et définitions des activités de maintenance., 1994.
- [66] P. Demartines and J. Héroult. Representation of nonlinear data structures through a fast VQP neural network. In *NeuroNîmes93 (Neural Networks and their applications)*, pages 411–424, Paris, France, October 1993. EC2.
- [67] M. N. Desai, J. C. Deckert, and J. J. Deyst. Dual sensor failure identification using analytic redundancy. *Journal of Guidance, Control and Dynamics*, 2(3) :213–220, 1979.
- [68] P.A. Devijver and J. Kittler. *Pattern Recognition : A Statistical Approach*. Prentice Hall, Englewood Cliffs, New Jersey, 1982.
- [69] F. Lanza di Scalea and J. McNamara. Measuring High-Frequency Wave Propagation In Railroad Tracks By Joint Time-Frequency Analysis. *Journal of Sound and Vibration*, 273 :637–651, 2004.
- [70] E. Diday. Une nouvelle méthode en classification automatique et reconnaissance des formes la méthode des nuées dynamiques. *Revue de Statistiques Appliquées*, 19(2) :19–33, 1971.
- [71] J. Doak. An evaluation of feature selection methods and their application to computer security. Technical report, Department of Computer Science, University of California, Davis, CA, 1992.
- [72] G. Dreyfus and al. *Réseaux de neurones : Méthodologie et applications*. Eyrolles, 2002.
- [73] K. Drouiche, M. Sidahmed, and Y. Grenier. Détection de défauts d’engrenage par analyse vibratoire. *Traitement de Signal*, 8(5) :331–343, 1991.
- [74] B. Dubuisson. *Diagnostic, Intelligence Artificielle et Reconnaissance des Formes*. Editions Hermès, paris, 2001.
- [75] R. P. W. Duin. On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions. *IEEE Trans. Comput.*, 25(11) :1175–1179, 1976.
- [76] S. Dumais. Using svms for text categorization. *IEEE Intelligent Systems*, 13(4), 1998. In : M.A. Hearst, B. Schölkopf, S. Dumais, E. Osuna and J. Platt : Trends and Controversies – Support Vector Machines.
- [77] R. Freund E. Osuna and F. Girosi. Training Support Vector Machines : An Application To Face Detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 130–136, 1997.
- [78] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.
- [79] G. Eichmann and N.M. Marinovic. An expansion of wigner distribution and its applications. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-85*, volume 3, pages 1021–1024, 1985.
- [80] S. Essid. *Classification automatique des signaux audio-fréquences : reconnaissance des instruments de musique*. Thèse de doctorat, l’Université Pierre et Marie Curie, Paris, 2005.
- [81] N. Martin et C. Doncarli. *Décision dans le Plan Temps-Fréquence. Traité IC2 - Information, Commande, Communication. Série Traitement du Signal et de l’Image*. Hermès Sciences, Paris, 2004.
-

-
- [82] F. Hlawatsch et F. Auger. *Temps-Fréquence : Concepts et Outils. Traité IC2, série Traitement du signal et de l'image*. Hermès Sciences, Lavoisier, Paris, 2005.
- [83] M. Basseville et M. O. Cordier. Surveillance et diagnostic de systèmes dynamiques, approches complémentaires du traitement de signal et de l'intelligence artificielle. Technical Report 1004, IRISA, January 1996.
- [84] J. M. Liang F. J. Chang and Y. C. Chen. Flood Forecasting Using Radial Basis Function Neural Network. *IEEE Transactions on Systems, Man and Cybernetics - Part C : Applications and Reviews*, 31(4), 2001.
- [85] M. Hatef F. J. Ferri, P. Pudil and J. Kittler. Comparative study of techniques for large-scale feature selection. In : *Pattern Recognition in Practice IV, Multiple Paradigms, Comparative Studies and Hybrid Systems*, eds. E. S. Gelsema and L. S. Kanal., pages 403–413, 1994.
- [86] H. Farreny. *Les systèmes experts - Principes et exemples*. Cépadues-Editions, Toulouse, 1985.
- [87] T.S. Ferguson. *Mathematical Statistics : A Decision Theoretic Approach*. New York and London : Academic Press. XI, 396 p., 1967.
- [88] P. Flandrin. *Temps-Fréquence*, volume 2ème. Ed. of *Traité des Nouvelles Technologies, série Traitement du Signal*. Hermès, Paris, 1998.
- [89] P. M. Franck. Principles of Model-Based Fault Detection. In *Proc. of the International Symposium on Artificial Intelligence in Real-Time Control*, 1992.
- [90] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy : A survey and some new results. *Automatica*, 26(3) :459–474, 1990.
- [91] Y. Freund and Robert E. Schapire. Experiments With A New Boosting Algorithm. In *International Conference on Machine Learning*, pages 148–156, 1996.
- [92] J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.*, 23(9) :881–890, 1974.
- [93] Keinosuke Fukunaga. *Introduction to statistical pattern recognition (2nd ed.) (Computer Science and Scientific Computing Series)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
- [94] D. Obradovic G. Deco and D. Obradovic. *An Information-Theoretic Approach to Neural Computing*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1996.
- [95] D. Gabor. Theory of Communication. *J. IEE (London)*, 93(3) :429–457, November 1946.
- [96] M. R. Garey and D. S. Johnson. *Computers and Intractability : A Guide to the Theory of NP-Completeness - Series of Books in the Mathematical Sciences*. W. H. Freeman and Company, 1979.
- [97] J. J. Gertler. Survey of model-based failure detection and isolation in complexe systems. *IEEE Control Systems Magazine*, 8(6) :3–11, 1988.
- [98] J. J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Electrical Engineering and Electronics 104. Marcel Dekker, 1998.
-

- [99] J. Ghosh and A. Nag. *An Overview of Radial Basis Function Networks*, volume 2, pages 1–36. Physica-Verlag GmbH, Heidelberg, Germany, 2000.
- [100] G.H. Golub and C.F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, second edition, 1989.
- [101] P. Goncalves and R. G. Baraniuk. Pseudo affine wigner distribution : Definition and kernel formulation. *IEEE Transactions on Signal Processing*, 46(6) :1505–1516, 1998.
- [102] G. Govaert. *Analyse des Données. Traité IC2 - Information, Commande, Communication. Série Traitement du Signal et de l'Image*. Hermès Sciences, Paris, 2003.
- [103] M. Grenier. *Méthodes de détection d'évolutions*. Thèse de doctorat, Université de Technologie de Compiègne, 1984.
- [104] A. Grossmann and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. of Math. An.*, 15 :723–736, 1984.
- [105] D. Groutage and D. Bennink. A New Matrix Decomposition Based on Optimum Transformation of the Singular Value Decomposition Basis Sets Yields Principal Features of Time-Frequency Distributions. In *10th IEEE Workshop on Statistical Signal and Array Processing*, pages 598–602, Pocono Manor, Pa, USA, August 2000.
- [106] D. Groutage and D. Bennink. Feature sets for nonstationary signals derived from moments of the singular value decomposition of cohen-posch (positive time-frequency) distributions. *IEEE Trans. Signal Processing*, 48(5) :1498–1503, 2000.
- [107] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3 :1157–1182, 2003.
- [108] M. Mesbah H. Hassanpour and B. Boashash. Time-frequency feature extraction of newborn EEG seizure using svd-based techniques. *EURASIP Journal on Applied Signal Processing*, 16 :2544–2554, 2004.
- [109] Mark A. Hall. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In *Proc. 17th International Conf. on Machine Learning*, pages 359–366. Morgan Kaufmann, San Francisco, CA, 2000.
- [110] J.K. Hammond and P.R. White. The analysis of non-stationary signals using time-frequency methods. *Journal of Sound and Vibration*, 190(3) :419–447, 1996.
- [111] R. L. Hardy. Multiquadratic Equations of Topography and Other Irregular Surfaces. *Journal of Geophysical Research*, 76(8) :1905–1915, 1971.
- [112] T. Hastie and R. Tibshirani. Classification by pairwise coupling. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. The MIT Press, 1998.
- [113] N. G. Hernandez. *Système de diagnostic par Réseaux de Neurones et Statistiques : Application à la détection d'hypovigilance d'un conducteur automobile*. Thèse de doctorat, LAAS/Toulouse, 1999.
- [114] Robert M. Hierons. Machine learning, by tom m. mitchell, mcgraw-hill, 1997 (book review). *Softw. Test., Verif. Reliab.*, 9(3) :191–193, 1999.
-

- [115] F. Hlawatsch and G. Boudreaux-Bartels. Linear and Quadratic Time-Frequency Signal Representations. *IEEE Signal Process. Mag.*, 9 :21–67, 1992.
- [116] Lasse Holmström, Petri Koistinen, Jorma Laaksonen, and Erkki Oja. Comparison of neural and statistical classifiers – theory and practice. research reports a13., 1996.
- [117] C. Hory. *Mélanges de Distribution du Chi-Square pour l'Interprétation d'une Représentation Temps-Fréquence*. Thèse de doctorat, INPG, Grenoble, 2002.
- [118] C. Hory and N. Martin. Time-frequency modelization as a mixture of chi-square distributions. In *SSP'03 : IEEE Workshop on Statistical Signal Processing*, pages 233–236, Saint Louis, Missouri, 2003.
- [119] H. Hotelling. Analysis of a complex of statistical variables in principal components. *Journal of Educational Psychology*, 24 :417–441, 1933.
- [120] C. Hsu and C.J. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2) :415–425, 2002.
- [121] R. Isermann. Process fault detection based on modeling and estimation methods : A survey. *Automatica*, 20(4) :387–404, 1984.
- [122] R. Isermann and P. Balle. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5(5) :709–719, 1997.
- [123] J. J. Deyst J. C. Deckert, M. N. Desai and A. S. Willsky. F-8 debw sensor failure identification using analytic redundancy. *IEEE Transactions on Automatic Control*, 22(5) :795–803, 1977.
- [124] R. Kohavi J. Dougherty and M. Sahami. Supervised and Unsupervised Discretization of Continuous Features. In *International Conference on Machine Learning*, pages 194–202, 1995.
- [125] S. Beck J. Ghosh and L. Deuser. A neural network based hybrid system for detection, characterization and classification of short-duration oceanic signals. *IEEE Journal of Ocean Engineering*, 17(4) :351–363, October 1992.
- [126] S. Ni J. Gu and J. Yuan. Non-stationary signal analysis and transient machining process condition monitoring. *International Journal of Machine Tools and Manufacture*, 42(1) :41–51, 2002.
- [127] A. Wu J. Meador and T. S. Lin. Fast diagnosis of integrated circuits faults using feedforward neural networks, 1991.
- [128] E. Fourgeau J. Morlet, G. Arens and D. Giard. Wave Propagation and Sampling Theory - Part 1 : Complex Signal and Scattering in Multilayered Media. *Geophysics*, 47(2) :203–221, 1982.
- [129] E. Roy et P. Abraham J. Odobez. Analyse d'images temps-fréquences de signaux doppler du flux sanguin par modèles déformables. In *17ème Colloque GRETSI*, pages 1193–1197, 1999.
- [130] S. Philipp J. P. Cocquerez and al. *Analyse d'Images : Filtrage et Segmentation*. Masson, 1995.
- [131] J. E. Jackson. *A User's Guide to Principal Components*. Wiley Series on Probability and Statistics. John Wiley & Sons, Inc., 1991.
- [132] A. Jain and D. Zongker. Feature selection : Evaluation, application and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2) :153–158, 1997.
-

-
- [133] A. Jain and D. Zongker. Feature selection : Evaluation, application and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2) :153–158, 1997.
- [134] A. K. Jain and R. C. Dubes. *Algorithms For Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.
- [135] A.K. Jain and B. Chandrasekaran. Dimensionality and Sample Size Considerations in Pattern Recognition Practice. In *P. R. Krishnaiah and L. N. Kanal, editors, Handbook of Statistics*, volume 2, pages 835–855. North Holland, 1982.
- [136] T. Joachims. Ransductive Inference for Text Classification Using Support Vector Machines. In *16th International Conference on Machine Learning*, pages 200–209, 1999.
- [137] T. Joachims. Estimating the Generalization Performance of a SVM Efficiently. In *Proceedings of ICML-00, 17th International Conference on Machine Learning*, pages 431–438, San Francisco, US, 2000. Morgan Kaufmann Publishers.
- [138] George H. John, Ron Kohavi, and Karl Pflieger. Irrelevant features and the subset selection problem. In *International Conference on Machine Learning*, pages 121–129, 1994.
- [139] C. Junsheng, Y. Dejie, and Y. YuA. fault diagnosis approach for roller bearings based on emd method and ar model. *Mechanical Systems and Signal Processing*, 20(2) :350–362, 2006.
- [140] S. Kay. *Modern spectral analysis*. Prentice Hall, 1988.
- [141] S. Keerthi, S. Shevade, C. Bhattacharyya, and K. Murthy. Improvements To Platt’s SMO Algorithm For SVM Classifier Design. Technical report, Dept. of CSA, Banglore, India, 1999.
- [142] P. E. Keller, R. T. Kouzes, and L. J. Kangas. Three Neural Network Based Sensor Systems for Environmental Monitoring. In *Electro/94 International Combined Conference Proceedings*, pages 377–382, 13760 Noel Road, Suite 500, Dallas, TX 75240, USA, May 1994. Miller Freeman, Inc.
- [143] J. Kittler. Feature selection and extraction. In T.Y. Young and K. S. Fu, editors, *Handbook of Pattern Recognition and Image Processing*, pages 59–83. Academic Press, New York, 1996.
- [144] J. De Kleer and B. C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32(1) :97–130, 1987.
- [145] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2) :273–324, 1997.
- [146] T. Kohonen. The Self-Organizing Map. *Proceedings of the IEEE*, 78(9) :1464–1480, 1990.
- [147] H. N. Koivo. Artificial neural networks in fault diagnosis and control. *Control Engineering Practice*, 02(1) :89–101, 1994.
- [148] Mineichi Kudo, Petr Somol, Pavel Pudil, Masaru Shimbo, and Jack Sklansky. Comparison of classifier-specific feature selection algorithms. In *SSPR/SPR*, pages 677–686, 2000.
- [149] R. Olshen L. Breiman, J. Friedman and C. Stone. *Classification and Regression Trees*. Chapman & Hall, New York, 1984.
-

- [150] L. Belanche L. C. Molina and A. Nebot. Feature selection algorithms : A survey and experimental evaluation. In *Second IEEE International Conference on Data Mining (ICDM'02)*, pages 306–313, 2002.
- [151] L. GySrfi L. Devroye and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, London, UK, 1996.
- [152] A. Morineau et M. Piron L. Lebart. *Statistique exploratoire multidimensionnelle*. Dunod, Paris, 1995.
- [153] P. Lachenbruch and M. Mickey. Estimation of error rates in discriminant analysis. *A Journal of Statistics for the Physical, Chemical and Engineering Sciences, Technometrics*, 10 :1–11, 1968.
- [154] J. Lampinen, J. Laaksonen, and E. Oja. *Neural Network Systems, Techniques and Applications in Pattern Recognition. Research rept.* 1997.
- [155] P. Langley. Selection of relevant features in machine learning, 1994.
- [156] S. le Cessie and J.C. van Houwelingen. Ridge Estimators in Logistic Regression. *Applied Statistics*, 41(1) :191–201, 1992.
- [157] S.W. Lee and A. Verri. *Pattern Recognition with Support Vector Machines, volume 2388 of LNCS*. Springer Verlag, Berlin, 2002.
- [158] O. Lemoine. *Détection de Signaux non Stationnaires par Représentation Temps-Fréquences*. Thèse de doctorat, Université de Nice, 1995.
- [159] B. Leprettre and N. Martin. Extraction of pertinent subsets from time-frequency representations for detection and recognition purposes. *Signal Processing*, 82(2) :229–238, 2002.
- [160] D. D. Lewis. Feature Selection and Feature Extraction for Text Categorization. In *Proceedings of Speech and Natural Language Workshop*, pages 212–217, San Mateo, California, 1992. Morgan Kaufmann.
- [161] H. Liu and R. Setiono. Feature selection and classification - a probabilistic wrapper approach. In *In Proceedings of the Ninth International Conference on Industrial and Engineering Applications of AI and ES*, pages 284–292, 1996.
- [162] H. Liu and R. Setiono. A probabilistic approach to feature selection - a filter solution. In *International Conference on Machine Learning*, pages 319–327, 1996.
- [163] Huan Liu and Hiroshi Motoda. *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.
- [164] G. Loosli. Fast svm toolbox in matlab based on simplesvm algorithm, 2004. <http://asi.insa-rouen.fr/~gloosli/simplesvm.html>, 2004.
- [165] G. Loosli, S. Canu, S.V.N. Vishwanathan, A. J. Smola, and M. Chattopadhyay. Boîte à Outils SVM Simple et Rapide. *RIA - Revue d'intelligence artificielle*, 19, 2005.
- [166] N. Lopes and B. Ribeiro. Part quality prediction in an injection moulding process using neural networks. In *WMC '99 : Proceedings of the International Workshop on Mobile Commerce*, New York, NY, USA, 1999. ACM.
- [167] M. Luo and J. L. Aravena. Pseudo power signatures for aircraft fault detection and identification. In *in Proc. 21th Digital Avionics Systems Conference*, volume 2, pages 12E3/1 – 12E3/8, October 2002.
-

- [168] I. Djurović M. Daković and L. Stanković. Adaptive local polynomial fourier transform. in *Proceedings of the 11th European Signal Processing Conference (EUSIPCO '02)*, 2 :603–606, 2002.
- [169] M. W. Mak and S. Y. Kung. Estimation of elliptical basis function parameters by the EM algorithm with application to speaker verification. *IEEE Trans. on Neural Networks*, 11(4) :961–969, July 2000.
- [170] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [171] Z. Hamou Mamar, P. Chainais, and A. Aussem. Probabilistic classifiers and time-scale representations : application to the monitoring of a tramway guiding system. In *ESANN*, pages 659–664, 2006.
- [172] M. Mancas, B. Gosselin, and B. Macq. Segmentation using a region growing thresholding. In E. R. Dougherty, J. T. Astola, and K. O. Egiazarian, editors, *Proc. of the Electronic Imaging Conference of the International Society for Optical Imaging, (SPIE/IS&T 2004)*, pages 388–398, San Jose (California), USA, 2005.
- [173] N. M. Marinovic and G. Eichmann. Feature extraction and pattern classification in space-spatial frequency domain. In *SPIE '85 : Proceedings of SPIE Intelligent Robots and Computer Vision*, volume 579, pages 19–25, 1985.
- [174] J. Marques and P. Moreno. A Study Of Musical Instrument Classification Using Gaussian Mixture Models And Support Vector Machines. Technical Report CRL 99/4, Compaq Corporation, Cambridge Research laboratory, Cambridge, US, June 1999.
- [175] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2004.
- [176] J. McNames. A Fast Nearest-Neighbor Algorithm Based on a Principal Axis Search Tree. *IEEE Trans. on Pattern Anal. Mach. Intell.*, 23(9) :964–976, 2001.
- [177] Y. Meyer. *Ondelettes et Opérateurs, I : Ondelettes, II : Opérateurs de Calderon-Zygmund, III : (with R. Coifman), Opérateurs multilinéaires*. Hermann-Editions, Paris, 1990.
- [178] Y. Meyer. *Les Ondelettes : Algorithmes et Applications*. Armand Colin Editeur, Paris, 1992.
- [179] Y. Meyer. Le traitement du signal et l'analyse mathématique. *Ann. Inst. Fourier (Grenoble)*, 50(2) :593–632, 2000.
- [180] C. A. Micchelli. Interpolation of scattered data : distance matrices and conditionally positive definite functions, 1986.
- [181] M. Millgram. *Reconnaissance de formes, méthodes numériques et connexionnistes*. Armand Colin, Dunod, Paris, 1993.
- [182] R. Milne. Strategies for diagnosis. In *IEEE Transactions on Systems, Man and Cybernetics*, volume 17, pages 333–339, 1987.
- [183] T. M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.
- [184] J. Moody and J. Darken. Fast learning in networks of locally tuned processing units. *Neural Computation*, 1 :281–294, 1989.
- [185] Andrew W. Moore and Mary S. Lee. Efficient algorithms for minimizing cross validation error. In *International Conference on Machine Learning*, pages 190–198, 1994.
-

- [186] J. Morlet. Sampling Theory and Wave Propagation. In C. H. Chen, editor, *Issues On Acoustic Signal/Image Processing And Recognition*, volume F1 of *NATO ASI*, pages 233–261. Springer-Verlag, 1983.
- [187] M. S. Mouchaweh. *Conception d'un système de diagnostic adaptatif et prédictif basé sur la méthode Fuzzy Pattern Matching pour la surveillance en ligne des systèmes évolutifs. Application à la supervision et au diagnostic d'une ligne de peinture au trempé*. Thèse de doctorat, Reims Champagne-Ardenne, 2002.
- [188] G. Nakos and D. Joyner. *Linear Algebra with Applications*. Brooks/Cole Publishing Company, 1998.
- [189] Mineichi Kudo Naoto Abe and Masaru Shimbo. Classifier-independent feature selection based on non-parametric discriminant analysis. In *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic and Statistical Pattern Recognition*, pages 470–479, London, UK, 2002. Springer-Verlag.
- [190] P. M. Narendra and K. Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Trans. Comput.*, 26(9) :917–922, 1977.
- [191] A. Nawrocki. *Contribution à la modélisation des câbles monotorons par éléments finis*. Thèse de doctorat, Université de Nantes, École Centrale de Nantes, 1997.
- [192] T. Böhme, C. S. Cox, N. Valentin, and T. Denoeux. *Comparison of Autoassociative Neural Networks and Kohonen Maps for Signal Failure Detection and Reconstruction*, volume 9, pages 637–644. ASME Press, New-York, USA, 1999.
- [193] N. Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1) :62–66, 1979.
- [194] R. B. Ouaghram. *Contribution au diagnostic des machines tournantes complexes. Application à un Laminoir*. Thèse de doctorat, Université de Technologie de Compiègne, 1999.
- [195] L. Oukhellou. *Paramétrisation et Classification de Signaux en Contrôle Non Destructif. Application à la Reconnaissance des Défauts de Rails par Courants de Foucault*. Thèse de doctorat, Université Paris XI - Orsay, 1997.
- [196] Latifa Oukhellou and Patrice Aknin. Hybrid training of radial basis function networks in a partitioning context of classification. *Neurocomputing*, 28(1-3) :165–175, 1999.
- [197] A. Grossmann P. Goupillaud and J. Morlet. Cycle-Octave and related transforms in seismic signal analysis. *Geoexploration*, 23 :85–102, 1984.
- [198] J. Novovičová P. Pudil and J. Kittler. Floating search methods in feature-selection with non-monotonic criterion functions. *PRL*, 15(11) :1119–1125, November 1994.
- [199] J. Novovičová P. Somol, P. Pudil and P. Paclík. Adaptive floating search methods in feature selection. *Pattern Recogn. Lett.*, 20(11-13) :1157–1163, 1999.
- [200] E. Parzen. On The Estimation Of A Probability Density Function And Mode. *Annals of Mathematical Statistics*, 33 :1065–1076, 1962.
- [201] F. Patrick. Quelques méthodes temps-fréquence et temps-échelle en traitement du signal. In *Les ondelettes en 1989 (Orsay, 1989)*, volume 1438 of *Lecture Notes in Math.*, pages 81–92, 205–206. Springer, Berlin, 1990.
-

- [202] X. Pei and F.N.Chowdhury. Unsupervised neural networks for fault detection and classification in dynamic systems. *IEEE International Conference on Control Applications (CCA99)*, 1 :640 – 645, August 1999.
- [203] T. A. Petsche, A. Marcantonio, C. Darken, S. J. Hanson, G. M. Kuhn, and I. Santoso. A neural network autoassociator for induction motor failure prediction. In *Advances in Neural Information Processing Systems*, volume 8, pages 924–930. The MIT Press, 1996.
- [204] V. Pierson. *Extraction de Sous-Ensembles Temps-Fréquences en Vue d'une Prise de Décision en Non stationnaire. Application en Acoustique Sous-Marine*. Thèse de doctorat, INPG, Grenoble, 1997.
- [205] B. Pinkowski. Principal component analysis of speech spectrogram images. *Pattern Recognition*, 30(5) :777–787, 1997.
- [206] J. C. Platt. *Fast Training of Support Vector Machines using Sequential Minimal Optimization*, pages 185–208. MIT Press, Cambridge, MA, USA, 1999.
- [207] S. Ploix. *Diagnostic des systèmes incertains : l'approche bornante*. Thèse de doctorat, l'Institut National Polytechnique de Lorraine, 1998.
- [208] T. Poggio and F. Girosi. A Theory of Networks for Approximation and Learning. Technical Report AIM-1140, Massachusetts Institute of Technology, Cambridge, MA, USA, 1989.
- [209] M. J. D. Powell. *Radial basis functions for multivariable interpolation : a review*, pages 143–167. Clarendon Press, New York, NY, USA, 1987.
- [210] P. Pudil, J. Novovičová, and J. Kittler. Floating search methods in feature selection. *Pattern Recogn. Lett.*, 15(11) :1119–1125, 1994.
- [211] P. Pudil, J. Novovičová, and P. Somol. Feature selection toolbox software package. *Pattern Recogn. Lett.*, 23(4) :487–492, 2002.
- [212] J.R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1) :81–106, 1986.
- [213] Ross Quinlan. *C4.5 : Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.
- [214] M. Coates R. Baraniuk and P. Steeghs. Hybrid linear/quadratic time-frequency attributes. In *ICASSP '00 : Proceedings of the Acoustics, Speech and Signal Processing, 2000. on IEEE International Conference*, pages II681–II684, Washington, DC, USA, 2000. IEEE Computer Society.
- [215] L. Huang R. Carmona and B. Torrèsani. Multiridge Detection and Time-Frequency Reconstruction. *IEEE Trans. on Signal Processing*, 47(2) :480–492, 1999.
- [216] A. J. E. M. Janssen R. G. Baraniuk, P. Flandrin and O. J. J. Michel. Measuring Time-Frequency Contents Using the Rényi Entropie. *IEEE Trans. On Inform. Theory*, 47(4) :1391–1409, 2001.
- [217] M. Coates R. G. Baraniuk and P. Steeghs. Hybrid linear/quadratic time-frequency attributes. *IEEE Transactions on Signal Processing*, 49(4) :760–766, 2001.
- [218] P. E. Hart R. O. Duda and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2nd edition, New York, NY, USA, 2001.
-

- [219] D. Racoceanu R. Zemouri and N. Zerhouni. Application of the dynamique rbf network in a monitoring problem of the production systems, 2002.
- [220] D. Racoceanu R. Zemouri and N. Zerhouni. From the spherical to an elliptic form of the dynamic rbf neural network influence field, 2002.
- [221] D. Racoceanu R. Zemouri and N. Zerhouni. Réseaux de neurones récurrents à fonction de base radiales : Rrfr/application au pronostic, 2002.
- [222] S. Raudys. *Statistical and Neural Classifiers : An Integrated Approach to Design*. Springer-Verlag, London, UK, 2001.
- [223] S. J. Raudys and A. K. Jain. Small Sample Size Effects in Statistical Pattern Recognition : Recommendations for Practitioners. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(3) :252–264, 1991.
- [224] S.J. Raudys and V. Pikelis. On Dimensionality, Sample Size, Classification Error and Complexity of Classification Algorithms in Pattern Recognition. *PAMI*, 2(3) :243–252, May 1980.
- [225] P. Ravier. *Détection de transitoires par ondelettes adaptées - Critères d'adaptation fondés sur les statistiques d'ordre supérieur*. Thèse de doctorat, L'Institut National Polytechnique de Grenoble, 1998.
- [226] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1) :57–95, 1987.
- [227] R. Rengaswamy and V. Venkatasubramanian. A Syntactic Pattern Recognition Approach for Process Monitoring and Fault Diagnosis. *Engineering Applications of Artificial Intelligence Journal*, 8(1) :35–51, 1995.
- [228] C. Richard. *Une méthodologie pour la détection à structure imposée. Applications au plan temps-fréquence*. Thèse de doctorat, Université de Technologie de Troyes, Décembre 1998.
- [229] C. Richard. Time-Frequency-based Detection Using Discret-Time Discret-Frequency Wigner Distributions. *IEEE Trans. on Signal processing*, 50(9) :2170–2176, 2002.
- [230] A. Roueff. *Traitements des Signaux sismiques Multicapteurs et Multicomposantes, utilisant les Représentations Temps-Fréquences*. Thèse de doctorat, Laboratoire des Images et des Signaux de Grenoble, 2003.
- [231] Y. Grandvalet S. Canu and A. Rakotomamonjy. Toolbox svm pour matlab, <http://asi.insa-rouen.fr/arakotom/toolbox/index>, 2002.
- [232] R. T. Kouzes S. Hashem, P. E. Keller and L. J. Kangas. Neural network based data analysis for chemical sensor analysis, 1995.
- [233] C. Bhattacharyya S. Sathiya Keerthi, S. K. Shevade and K. R. K. Murthy. Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, 13(3) :637–649, 2001.
- [234] J. S. Sanchez, F. Pla, and F. J. Ferri. Prototype Selection For The Nearest Neighbour Rule Through Proximity Graphs. *Pattern Recogn. Lett.*, 18(6) :507–513, 1997.
- [235] G. Sandell and W. Martens. Perceptual evaluation of principal-component-based synthesis of musical timbres. *Journal of the Audio Engineering Society*, 43(12) :1013–1028, 1995.
-

-
- [236] C. Saunders, M. O. Stitson, J. Weston, L. Bottou, B. Schölkopf, and A. Smola. Support vector machine - reference manual. Technical Report CSD-TR-98-03, Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, UK, 1998.
- [237] I. P. Schagen. Interpolation in Two Dimensions - A New Technique. *IMA Journal of Applied Mathematics*, 23(1) :53–59, 1979.
- [238] M. Schmidt and H. Gish. Speaker Identification Via Support Vector Classifiers. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-96*, volume 1, pages 105–108, 7-10 May 1996.
- [239] F. Schwenker, H. A. Kestler, and G. Palm. Three learning phases for radial-basis-function networks. *Neural Networks*, 14(4-5) :439–458, 2001.
- [240] B. Schölkopf. *Support Vector Learning*. R. Oldenbourg Verlag, Munich, 1997.
- [241] B. Schölkopf. Svms – a practical consequence of learning theory. *IEEE Intelligent Systems*, 13 :18–21, 1998. In : Trends and Controversies – Support Vector Machines.
- [242] B. Schölkopf, C. Burges, and V. Vapnik. Extracting support data for a given task. In U. M. Fayyad and R. Uthurusamy, editors, *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, Menlo Park, 1995. AAAI Press.
- [243] B. Shölkopf and A. J. Smola. *Learning with kernels*. The MIT Press, 2002.
- [244] W. Siedlecki and J. Sklansky. On automatic feature selection. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 2(2) :197–220, 1988.
- [245] L. Sirovich. Turbulence and the dynamics of coherent structures. *Quarterly Applied Mathematics*, 45(3) :561–590, 1987.
- [246] A.J. Smola and B. Schölkopf. A Tutorial On Support Vector Regression. Technical Report NC2-TR-1998-030, Royal Holloway College, London, UK, 1998.
- [247] J. Sottile and L.E. Holloway. An overview of fault monitoring and diagnosis in mining equipment. *IEEE Transactions on Industry Applications*, 30(5) :1326–1332, Sep/Oct 1994.
- [248] S. D. Stearns. On selecting features for pattern classifiers. In *Proc. Int'l Conf. Pattern Recognition*, pages 71–75, 1976.
- [249] G. W. Stewart. On the early history of the singular value decomposition. *SIAM Rev.*, 35(4) :551–566, 1993.
- [250] M. O. Stitson, J. A. E. Weston, A. Gammerman, V. Vovk, and V. Vapnik. Theory Of Support Vector Machines. Technical Report CSD-TR-96-17, Royal Holloway, University of London, December 1996.
- [251] A. Hoglander T.Braun, C. Diot and V. Roca. An experimental user level implementation of tcp. Rapport de recherche 265, INRIA, Sep 1995.
- [252] G. Terstyanszky and L. Kovacs. Improving fault diagnosis using proximity and homogeneity measure. In *The 15th IFAC World Congress on Automatic Control*, Barcelone, Espagne, July 2002.
- [253] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press Inc., 1999.
-

- [254] Charles W. Therrien. *Decision estimation and classification : an introduction to pattern recognition and related topics*. John Wiley & Sons, Inc., New York, NY, USA, 1989.
- [255] B. Torrèsani. *Analyse continue par ondelettes*. Éditions du CNRS, EDP sciences, 1995.
- [256] S. Touaf. *Diagnostic Logique des Systèmes Complexes Dynamiques dans un Contexte Multi-Agent*. Thèse de doctorat, Joseph Fourier Grenoble 1, 2005.
- [257] D. Tounissoux. *Pseudo-Questionnaires et Information*. Mémoire de 3ème cycle, Université de Lyon 1, 1974.
- [258] O. Trier, A. Jain, and T. Taxt. Feature Extraction Methods for Character Recognition - A Survey. *Pattern Recognition*, 29(4) :641–662, 1996.
- [259] D. Unay and B. Gosselin. Thresholding-Based Segmentation and Apple Grading by Machine Vision. In *Proc. of EUSIPCO*, Antalya, Turkey, 2005.
- [260] V.N. Vapnik. *The Statistical Learning Theory*. Springer, 1998.
- [261] A. T. Vemuri, M. Polycarpou, and S. Diakourtis. Neural Network Based Fault Detection and Accommodation in Robotic Manipulators. *IEEE Transactions on Robotics and Automation*, 14(2) :342–348, 1998.
- [262] A. T. Vemuri and M. M. Polycarpou. Neural Network Based Robust Fault Diagnosis in Robotic Systems. *IEEE Transactions on Neural Networks*, 8(6) :1410–1420, November 1997.
- [263] V. Venkatachalam and J. Aravena. Nonstationary Signal Classification Using Pseudo Power Signatures. In *the 1998 IEEE International Symposium on Circuits and Systems, ISCAS '98*, volume 5, pages 449–452, Monterey, CA, USA, June 1998.
- [264] V. Venkatachalam and J.L. Aravena. Nonstationary signal classification using pseudo power signatures : The matrix SVD approach. *IEEE Transactions on Circuits and Systems II : Analog and Digital Signal Processing*, 46 :1497 – 1505, December 1999.
- [265] V. Vidya. *Pseudo Power Signatures for Nonstationary Signal Analysis and Classification*. Phd dissertation, Louisiana State University, 1998.
- [266] J. Ville. Théorie et Applications de la Notion de Signal Analytique. *Câbles et Transm.*, 2 A(1) :61–74, 1948.
- [267] S. V. N. Vishwanathan, A. J. Smola, and M. N. Murty. Simple SVM. *Proceedings of the Twentieth International Conference on Machine Learning*, pages 760–767, 2003.
- [268] A. Wald. *Sequential Analysis*. John Wiley and Sons, New York, NY, USA, 1947.
- [269] Michael E. Wall, Andreas Rechtsteiner, and Luis M. Rocha. Singular value decomposition and principal component analysis, 2003.
- [270] J. Weston and C. Watkins. Multi-class support vector machines. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, UK, 1998.
- [271] E.P. Wigner. On The Quantum Correction For Thermodynamic Equilibrium. *Phys. Rev.*, 40 :749–760, 1932.
-

- [272] I. H. Witten and Eibe Frank. *Data Mining : Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.
- [273] A. Wu and J. Meador. Measurement selection for parametric ic fault diagnosis. *Journal of Electronic Testing : Theory and Applications*, 5(1) :9–18, 1994.
- [274] J.-D Wu and J.-C Chen. Continuous Wavelet Transform Technique for Fault Signal Diagnosis of Internal Combustion Engines. *NDT&E International Pages*, pages 1–8, 2005.
- [275] L. Atlas Y. Zhao and R. Marks. The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 38(7) :1084–1091, 1990.
- [276] B. Yazici and G.B. Kliman. An Adaptive Statistical Time-Frequency Method for Detection of Broken Bars and Bearing Faults in Motors Using Stator Current. *IEEE Trans. Ind. Appl*, 35(2) :442–452, 1999.
- [277] S. Zieba. *Une méthode de suivi d'un système évolutif. Application au diagnostic de la qualité d'usinage*. Thèse de doctorat, Université de Technologie de Compiègne, 1995.
- [278] D. A. Zighed, S. Rabaséda, and R. Rakotomalala. FUSINTER : A Method for Discretization of Continuous Attributes. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(3) :307–326, 1998.
- [279] G. Zumpano and M. Meo. A New Damage Detection Technique Based On Wave Propagation For Rails. *International Journal of Solids and Structures*, 43 :1023–1046, 2006.
- [280] G. Zwingelstein. *Diagnostic des Défaillances : Théorie et Pratique pour les Systèmes Industriels, Traité des Nouvelles Technologies, Série Diagnostic et Maintenance*. Hermès, Paris, 1995.
-

