Génération de symptômes flous par FFT et SWT : applications à la détection de vibrations

Samir Taleb

▶ **To cite this version:**

Samir Taleb. Génération de symptômes flous par FFT et SWT : applications à la détection de vibrations. Automatique / Robotique. Institut National Polytechnique de Grenoble - INPG, 2007. Français. NNT : . tel-00239310

HAL Id: tel-00239310
https://theses.hal.science/tel-00239310

Submitted on 5 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

N° : Attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|

THÈSE

Présentée par

Samir TALEB

Pour obtenir le grade de

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

Dans le cadre de l'école doctorale

Électronique, Électrotechnique, Automatique, Télécommunications, Signal

Spécialité : Automatique - Productique

GÉNÉRATION DE SYMPTÔMES FLOUS PAR FFT ET SWT APPLICATIONS À LA DÉTECTION DE VIBRATIONS

Soutenance publique le 14 Février 2007 devant le jury

Président	M^{r}	MONTMAIN Jacky	Professeur, EMA
Rapporteur	M^{r}	CHARARA Ali	Professeur, UTC
Rapporteur	M^{r}	SAUTER Dominique	Professeur, UHP
Directeur de thèse	M^{me}	GENTIL Sylviane	Professeur, INPG
Co-directeur	M^{me}	LESECQ Suzanne	Professeur, UJF

Thèse préparée au sein du Laboratoire d'Automatique de Grenoble

Remerciements

Je remercie tout d'abord le Laboratoire d'Automatique de Grenoble, particulièrement son responsable de formation doctorale Mr Christian CAUMAULT, ainsi que son staff administratif et technique, de m'avoir accueilli pendant ces trois années de thèse.

J'exprime naturellement ma gratitude envers les membres de mon jury : d'abord mes deux rapporteurs, le professeur Dominique SAUTER à l'Université Henri Poincaré Nancy et le professeur Ali CHARARA à l'Université de Technologie de Compiègne, pour le travail qui ont fait, le président du jury, le professeur Jacky MONTMAIN, ainsi que mes deux directeurs de thèse LESECQ Suzanne et GENTIL Sylviane.

Je témoigne d'une grande reconnaissance envers mon troisième encadreur, Mr Matija EXEL, Ingénieur de Recherche, pour sa grande disponibilité et son intérêt envers mon travail. J'ai été sensible à son soutien, à son encadrement régulier et à l'ambiance conviviale qu'il a su créer pour qu'une bonne partie de la thèse se déroule dans les meilleurs conditions.

J'adresse mes remerciements à l'industriel métallurgique SMS-Demag AG de nous avoir donné libre accès à ses données, particulièrement Mr Jörng DECKERS de nous avoir fourni des informations essentiels à la bonne compréhension des données.

Je remercie vivement tous les membres du projet européen MAGIC, en particulier : FARARGASAN Ioana, TOUAF Samir, GARCIA-BELTRAN Carlos, BELMAN Pascal et MARTIN Albert, pour les conseils et discussions autour de ce projet. Je remercie également mes collègues et amis : Ahmad, Lyes, Zahra, Joumana, Maher, Fethi et Mounir pour l'ambiance conviviale.

Enfin, un grand merci à ma famille qui m'a toujours soutenu et encouragé dans mes études.

Samir

Table des matières

Table des matières ..... **V**
Table des figures ..... **IX**
Liste des Tableaux ..... **XIII**
Introduction Générale **1**
Chapitre 1 Généralités sur le Diagnostic **5**
1 Introduction 5
2 Supervision et diagnostic 6
3 Caractéristiques souhaitables d'un système de diagnostic 9
3.1 Rapidité et sensibilité de détection 9
3.2 Robustesse 9
3.3 Identification de nouveaux défauts 9
3.4 Capacité de localisation 10
3.5 Adaptabilité 10
3.6 Identification des défauts multiples 10
3.7 Estimation d'erreur de classification 10
3.8 Facilité d'explication 10
4 Transformation des mesures dans un système de diagnostic 11
5 Classification des méthodes de diagnostic 12
6 Approche à base de modèle qualitatif 13
6.1 Modèles causaux 14
6.1.1 Graphes d'influence ou graphe causaux 14
6.1.2 Arbres de défaillances 15
6.1.3 Méthodes à base de physique qualitative 16
6.2 Abstractions hiérarchiques 17
6.2.1 Abstractions hiérarchiques structurelle ou matérielle 17
6.2.2 Abstractions hiérarchiques fonctionnelles 17
7 Approche à base de modèle quantitatif 19
7.1 Equations de redondance analytique, espace de parité 20
7.2 Bancs d'observateurs 20
7.3 Estimation de paramètres 21
8 Approche à base d'extraction de caractéristiques qualitatives 22
8.1 Systèmes experts 22
8.2 Tendances qualitatives 24
9 Approche à base d'extraction de caractéristiques quantitatives 25
9.1 Les méthodes statistiques 25
9.1.1 Tests d'Hypothèses 25
9.1.2 Classification probabiliste 27
9.1.3 Analyse en composante principale & Moindres Carrés Partiels 29
9.2 Les méthodes non statistiques 30
9.2.1 Classification floue 30
9.2.2 Réseaux de neurones 32
9.2.3 Méthodes de signature 32

9.2.4 Méthodes fréquentielles, temps-fréquence et temps-échelle 33
10 Conclusion 34
Chapitre 2 Diagnostic et Traitement du Signal 35
1 Introduction 35
2 Analyse temporelle 36
2.1 Transformée continue de Dirac 36
2.2 Signaux discrets et échantillonnés 36
2.3 Tests statistiques 36
2.4 Méthodes statistiques en diagnostic 38
2.4.1 Test progressif de Wald 38
2.4.2 Application à la détection de saut de variance connu 40
3 Analyse fréquentielle 41
3.1 Transformée de Fourier 41
3.2 Transformée de Fourier à temps discret 41
3.3 Transformée de Fourier discrète 42
3.4 Transformée de Fourier rapide 42
3.5 Précision et résolution spectrales 43
3.6 Effet du fenêtrage sur la résolution 43
4 Analyse temps-fréquence 44
4.1 La Transformée de Fourier à Court Terme 44
4.2 Interprétation en terme de filtrage 44
4.3 Résolution dans le plan temps-fréquence 44
4.4 Rôle de la fenêtre d'analyse 46
4.4.1 Taille de la fenêtre 47
4.4.2 Forme de la fenêtre 47
4.4.3 Conservation de l'énergie et spectrogramme 49
4.5 Transformée de Fourier à court terme discrète 49
5 Analyse temps-échelle 50
5.1 La Transformée en Ondelettes Continue 50
5.2 Interprétation en terme de filtrage 51
5.3 Résolution dans le plan temps-échelle 51
5.4 Bases discrètes d'ondelettes 53
5.4.1 L'analyse multirésolution 53
5.4.1.1 Définition 53
5.4.1.2 Fonction d'échelle et espaces d'approximation 54
5.4.1.3 Fonction ondelette et espaces de détails 54
5.4.1.4 Relations à deux échelles 55
5.4.2 Transformée discrète dyadique en ondelettes 55
5.4.2.1 Calcul des coefficients de la DWT : Algorithme pyramidal 56
5.4.2.2 Filtrage par bandes et résolution dans le plan temps-fréquence 57
5.4.3 Transformée en Ondelettes Stationnaire (Représentation continue dyadique) 58
5.4.3.1 Mise en œuvre de l'algorithme à trous 60
5.4.3.2 Interprétation en banc de filtres 61
5.5 Principales caractéristiques des ondelettes 61
5.5.1 Le cône d'influence 62
5.5.2 Localisation 62
5.5.3 Compacité du support 62
5.5.4 Régularité 62
5.5.5 Nombre de moments nuls 63
5.5.6 Symétrie 63

5.5.7 Orthogonalité 63
6 Conclusion 63
Chapitre 3 Détection de Défauts par Génération de Symptômes flous 65
1 Introduction 65
2 Raisonnement approximatif 66
2.1 Propositions floues 66
2.1.1 Propositions floues élémentaires 66
2.1.2 Propositions floues générales 66
2.1.2.1 Négation de proposition floue 67
2.1.2.2 Combinaisons de propositions floues 67
2.2 Opérateurs d'agrégation 67
2.2.1 Conjonction de propositions floues 67
2.2.2 Disjonction de propositions floues 68
2.2.3 Implication entre propositions floues 68
2.3 Décision multicritères 68
3 Détection de défauts par génération de symptômes flous 70
3.1 Principe de génération de symptômes flous 71
3.2 Validité du modèle 73
3.3 Test de cohérence 74
3.4 Niveau de confiance d'un symptôme et prise de décision 75
4 Génération de symptômes flous par STFT [80] 77
4.1 Extraction des attributs 77
4.2 Construction du modèle (gabarit) 78
4.3 Validité du modèle (gabarit) 82
4.4 Test de cohérence 83
4.4.1 Fuzzification 83
4.4.2 Agrégation 84
4.5 Génération des symptômes 85
5 Génération de symptômes flous par SWT [81] 86
5.1 Extraction des attributs 86
5.2 Seuillage des coefficients d'ondelettes 87
5.3 Validité du modèle (seuils) 89
5.4 Test de cohérence par évaluation des dépassements de seuils 89
5.4.1 Fuzzification 90
5.4.2 Agrégation 90
5.5 Génération des symptômes 92
6 Exemples 93
7 Conclusion 108
Chapitre 4 Application en Détection de Vibrations 109
1 Introduction 109
2 Détection de vibrations torsionnelle dans un laminoir 110
2.1 Description technique du laminoir 112
2.2 Les vibrations et leurs effets sur la qualité de la tôle 113
2.3 Description des données et acquisition 115
3 Résultats avec le test de détection à base de STFT 117
4 Résultats avec le test de détection à base de SWT 123
5 Conclusion 130
Conclusion Générale 131
Bibliographie 135

Table des figures

Figure 1.1 — Types de défauts 7
Figure 1.2 — Transformation des mesures dans un système de diagnostic 11
Figure 1.3 — Les méthodes de diagnostic 12
Figure 1.4 — Principe de génération de résidus 19
Figure 1.5 — Les primitives et la représentation d'une sinusoïde par TDL 24
Figure 1.6 — Test statistique de deux hypothèses 27
Figure 1.7 — Classification de données (ou reconnaissance de formes) 27
Figure 2.1 — Résultats du test de détection de saut de variance connu avec le test de Wald 40
Figure 2.2 — Pavage temps-fréquence associé à la STFT 45
Figure 2.3 — Spectrogrammes d'un signal somme d'une impulsion et d'une sinusoïde 46
Figure 2.4 — Impact de la taille de la fenêtre sur la localisation temps-fréquence 47
Figure 2.5 — Fenêtres d'analyse (représentation temporel, représentation fréquentiel) 48
Figure 2.6 — Exemples d'ondelettes mères 50
Figure 2.7 — Pavage temps-fréquence associé à la Transformée en ondelettes 51
Figure 2.8 — Algorithme pyramidal d'analyse–reconstruction 56
Figure 2.9 — Filtres "miroirs" 57
Figure 2.10 — DWT : Partition fréquentielle et coefficients de détails – largeur de bandes de détails et approximation – résolution temps fréquence associé aux $d_x(j,k)$ 58
Figure 2.11 — SWT : Partition fréquentielle et coefficients de détails – largeur de bandes de détails et approximation – résolution temps fréquence associé aux $cd_x(j,k)$ 59
Figure 2.12 — Décomposition d'un créneau et de sa version translatée 60
Figure 2.13 — Algorithme de décomposition–reconstruction par SWT 60
Figure 2.14 — Les filtres pour SWT 61
Figure 3.1 — Principe de détection de défauts par génération de symptômes flous 71
Figure 3.2 — La prise de décision incertaine 72
Figure 3.3 — Test de cohérence 72
Figure 3.4 — Opérateurs d'agrégation pour la génération des symptômes 76
Figure 3.5 — Les étapes de génération de symptômes par STFT 77
Figure 3.6 — Signal de référence x_1^{ref} 78
Figure 3.7 — Spectres de puissance de référence 79
Figure 3.8 — Génération des enveloppes 80
Figure 3.9 — Construction du gabarit – Etape 1 81
Figure 3.10 — Construction du gabarit – Etape 2 81
Figure 3.11 — Construction du gabarit – Etape 3 82
Figure 3.12 — Règle de fuzzification 83
Figure 3.13 — Signal avec défaut x^{obs} 83
Figure 3.14 — Fuzzification des dépassements du gabarit pour x^{obs} à l'instant $t_k = 1.2[\mathrm{s}]$ 84
Figure 3.15 — Résultat du test de cohérence pour x^{obs} 85
Figure 3.16 — Génération des symptômes par les différents opérateurs pour x^{obs} 85
Figure 3.17 — Les étapes de génération de symptômes par SWT 86
Figure 3.18 — Décomposition par SWT de x^{ref} et calcul des seuils 87
Figure 3.19 — Seuillage doux des coefficients d'ondelettes 88

Figure 3.20 — Décomposition par SWT et seuillage des coefficients de x^{obs} 89
Figure 3.21 — Règle de fuzzification des coefficients seuillés....90
Figure 3.22 — Fuzzification des coefficients seuillés pour x^{obs} 90
Figure 3.23 — Agrégation multi-échelle....91
Figure 3.24 — Résultat de l'agrégation multi-échelle pour x^{obs} 91
Figure 3.25 — Résultat du test de cohérence pour x^{obs} avec D_{coh} 92
Figure 3.26 — Génération des symptômes....92
Figure 3.27 — Signal de référence – signal, Analyse de *Haar*, analyse de *Daubechie6*....94
Figure 3.28 — Exemple 1 – Signal (zoom), coefficients d'ondelettes et symptômes 95
Figure 3.29 — Exemple 2 – Signal (zoom), coefficients d'ondelettes et symptômes 96
Figure 3.30 — Exemple 3 – Signal (zoom), coefficients d'ondelettes et symptômes 98
Figure 3.31 — Exemple 4 – Signal (zoom), coefficients d'ondelettes et symptômes 99
Figure 3.32 — Exemple 5 – Signal (zoom), coefficients d'ondelettes et symptômes 100
Figure 3.33 — Exemple 6 – signal (zoom), coefficients d'ondelettes et symptômes.... 101
Figure 3.34 — Exemple 7 – Signal (zoom), coefficients d'ondelettes et symptômes 102
Figure 3.35 — Exemple 8 – Signal (zoom), coefficients d'ondelettes et symptômes 103
Figure 3.36 — Exemple 9 – Signal (zoom), coefficients d'ondelettes et symptômes 104
Figure 3.37 — Exemple 10 – Signal (zoom), coefficients d'ondelettes et symptômes.... 105
Figure 3.38 — Exemple 11 – Signal (zoom), coefficients d'ondelettes et symptômes.... 106
Figure 3.39 — Exemple 12 – Signal (zoom), coefficients d'ondelettes et symptômes.... 107
Figure 4.1 — Effets des vibrations sur la forme du couple 111
Figure 4.2 — Réduction de la vitesse de laminage à cause du broutement torsionnel 111
Figure 4.3 — Schéma du laminoir dégrossisseur réversible (SMS-Demag AG) [77] 113
Figure 4.4 — Effets des vibrations sur la qualité du produit 114
Figure 4.5 — Photo du capteur de couple [77] 115
Figure 4.6 — Forme du couple haut durant un fonctionnement normal 116
Figure 4.7 — Signal sans défaut.... 117
Figure 4.8 — Spectres des signaux de référence 117
Figure 4.9 — Construction du Gabarit.... 118
Figure 4.10 — La comparaison par rapport au gabarit du spectre d'un signal avec défaut 118
Figure 4.11 — Résultat du test de cohérence, 1er fichier.... 119
Figure 4.12 — Génération des symptômes, 1er fichier 119
Figure 4.13 — Résultat du test de cohérence, 2ème fichier.... 120
Figure 4.14 — Génération des symptômes, 2ème fichier 120
Figure 4.15 — Résultat du test de cohérence, 3ème fichier.... 121
Figure 4.16 — Génération des symptômes, 3ème fichier 121
Figure 4.17 — Résultat du test de cohérence, 4ème fichier.... 122
Figure 4.18 — Génération des symptômes, 4ème fichier 122
Figure 4.19 — Approximation de la fréquence centrale de l'ondelette *Symmlet6*.... 123
Figure 4.20 — Coefficients SWT : signal sans défaut et calcul des seuils.... 124
Figure 4.21 — Coefficients SWT: signal sans défaut, 1er fichier 124
Figure 4.22 — Résultats de l'agrégation multi-échelles, 1er fichier.... 125
Figure 4.23 — Génération des symptômes, 1er fichier.... 125
Figure 4.24 — Coefficients SWT: signal avec défaut, 2ème fichier.... 126
Figure 4.25 — Résultats de l'agrégation multi-échelles, 2ème fichier.... 126
Figure 4.26 — Génération des symptômes, 2ème fichier 127
Figure 4.27 — Coefficients SWT: signal avec défaut, 3ème fichier.... 127
Figure 4.28 — Résultats du test de cohérence, 3ème fichier 128
Figure 4.29 — Génération des symptômes, 3ème fichier 128

Figure 4.30 — Coefficients SWT: signal avec défaut, 4ème fichier 129
Figure 4.31 — Résultats du test de cohérence, 4ème fichier 129
Figure 4.32 — Génération des symptômes, 4ème fichier 130

Liste des Tableaux

Tableau 1.1 — Décision statistique 26
Tableau 2.1 — Fenêtres d'analyses les plus couramment utilisées 48
Tableau 3.1 — Types de défauts simulés 93

Introduction Générale

Les entreprises de nombreux secteurs d'activité (automobile, sidérurgie, chimie, pétrochimie,…) sont de plus en plus soumises à la concurrence du marché. Pour assurer leur avenir, les entreprises doivent désormais faire face à différents enjeux socio-économiques, ce qui les amène pour pouvoir atteindre des objectifs de plus en plus exigeants à avoir des systèmes industriels de plus en plus complexes (automatisation de boucles de contrôle, introduction de microcalculateurs à différents niveaux et, plus récemment, informatisation hiérarchisée et distribuée), pas seulement à cause de la complexité physique des installations et des différentes procédures pour la fabrication de produits de qualité, mais aussi de par la croissante quantité d'information qu'il est nécessaire de gérer pour en garantir le fonctionnement correct. Sur le plan technique, les principales contraintes portent sur la diversité, la flexibilité, la complexité des produits. Des développements au niveau des technologies de l'informatique et de l'automatisation deviennent nécessaires.

Malgré le progrès et les avances technologiques dans la commande distribuée et l'automatisation des procédés complexes, la supervision des procédés reste une tâche très importante qui est encore en grande partie une activité manuelle, exécutée par les opérateurs, notamment lorsqu'il s'agit de répondre aux événements anormaux. Cette activité peut selon la nature et la criticité des anomalies ou dysfonctionnements, avoir un impact économique, environnemental et de sécurité plus ou moins significatif, non seulement au niveau des équipements mais aussi des opérateurs. D'autant plus que le fait de gérer une grande quantité d'information et d'avoir besoin d'agir vite peut mener les opérateurs à prendre des décisions incorrectes, dégradant encore plus la situation. De ce fait, les systèmes de supervision qui intègrent des outils de surveillance, de détection et de diagnostic sont nécessaires afin de fournir à l'opérateur des critères suffisants pour la prise de décisions.

Dans ce contexte, de nombreuses approches sont développées, en vue de la détection de défaillances et du diagnostic, par les différentes communautés scientifiques. Les méthodes se différencient par rapport au type de connaissances a priori sur le procédé qu'elles nécessitent. Ainsi elles peuvent être classées, de façon générale, comme des méthodes à base de modèles, et des méthodes à base d'historiques de données. Les méthodes à base de modèles considèrent un modèle structurel du comportement du procédé basé sur des principes physiques fondamentaux. Ces modèles peuvent être de type quantitatif, exprimés sous forme d'équations mathématiques ou bien de type qualitatif, exprimés par exemple sous forme de relations logiques. Les méthodes à base d'historiques de données cherchent à extraire des informations, sous forme d'exemples type ou tendances, au sein des mesures venant des capteurs et des actionneurs, pouvant identifier le comportement du procédé. Ces méthodes comprennent, parmi d'autres, les méthodes statistiques et de classification (ou reconnaissance de formes).

Il est évident que toutes ces différentes méthodes de détection et de diagnostic ont leurs points forts et leurs faiblesses. Ainsi il est possible que certaines méthodes puissent se complémenter offrant alors de meilleurs systèmes de diagnostic. L'intégration des diverses caractéristiques complémentaires est une façon de développer des systèmes "hybrides" permettant de surmonter les limitations des stratégies individuelles de chaque méthode. De ce fait, plusieurs initiatives de collaboration entre les industriels et les communautés de recherche scientifique conduisent au développement d'outils pour améliorer la supervision des unités industrielles. Une de ces initiatives est le projet Européen MAGIC (*Multi-Agents-Based Diagnostic Data Acquisition And Management In Complex Systems*) [34], auquel une partie importante de la thèse a été consacrée.

Le but principal de ce projet était de regrouper à partir d'une intégration modulaire d'outils logiciels, des techniques d'acquisition, de supervision et d'aide à la décision, notamment de détection et de diagnostic, développées par différentes équipes de recherche européennes. Cette intégration modulaire permet le choix et la combinaison des techniques les plus appropriées, selon les connaissances disponibles sur le procédé à superviser, afin de donner à l'opérateur des informations représentatives pour la prise de décisions.

Dans ce contexte, notre travail porte sur la détection des défauts par génération de symptômes flous, quand l'obtention d'un modèle mathématique n'est pas toujours possible. Notre contribution a donc consisté à proposer une méthode permettant, à partir des données disponibles et à l'aide de méthodes d'extraction de caractéristiques, par outils de traitement du signal, de construire un modèle comportemental du procédé. La mise en place de ce modèle est réalisée avec l'objectif de suivre en ligne les évolutions du procédé à surveiller, afin d'aider l'opérateur dans la prise de décisions, surtout lorsque ces situations anormales ont pour origine des dysfonctionnements.

Ce mémoire de thèse est structuré en quatre chapitres :

Le premier chapitre propose un tour d'horizon sur la supervision et les différentes approches de détection et de diagnostic de défaillances développées par les différentes communautés de recherche. Il rappelle dans un premier temps, les différentes notions utilisées en diagnostic, les caractéristiques souhaitables d'un système de diagnostic et les transformations des mesures dans un système de diagnostic. La deuxième partie fait l'objet d'une classification des différentes approches de diagnostic.

Le deuxième chapitre propose un bref tour d'horizon sur les différentes stratégies permettant le suivi du comportement d'un procédé et la détection de défauts, en s'appuyant sur des données historiques jugées par un expert et des données en ligne. Ce chapitre décrit également différentes méthodes d'extraction de caractéristiques qui peuvent être utilisées pour la mise en place de méthodes de détection de défauts, dont la Transformée de Fourier Rapide et la Transformée en Ondelettes Stationnaires sur lesquelles nous avons appuyé nos travaux.

Dans le troisième chapitre, nous abordons les différentes étapes à suivre pour l'élaboration d'une méthode de détection de défauts à partir d'outils d'analyse fréquentielle et analyse temps-échelle. Ainsi, la première phase consiste à extraire des attributs (caractéristiques) pour caractériser les différents états de fonctionnement du procédé à surveiller. Une description des éléments de paramétrage nécessaires à l'élaboration d'un modèle de comportement du procédé est aussi réalisée. Puis nous proposons l'évaluation des attributs par un mécanisme de raisonnement approximatif afin de générer des symptômes flous permettant la détection des défauts tout en prenant en considération la validité du modèle utilisé.

Le chapitre 4 est dédié dans un premier temps à la présentation des vibrations dans les laminoirs, leur impact sur la qualité des produits et la durée de vie du laminoir, qui peuvent avoir des conséquences économiques néfastes. Puis nous présentons dans un deuxième temps les résultats obtenus en appliquant les méthodes de détection présentées au chapitre 3 sur des mesures de couple dont nous disposons et qui ont été enregistrées sur un laminoir réversible au sein de l'industrie métallurgique SMS-Demag AG, partenaire du projet européen MAGIC.

Ce mémoire s'achève par une conclusion générale où quelques perspectives sont évoquées.

Chapitre 1

Généralités sur le Diagnostic

1 Introduction

De nos jours, les systèmes industriels modernes deviennent de plus en plus complexes et leurs commandes de plus en plus sophistiquées. En même temps la fiabilité, disponibilité et sûreté de fonctionnement sont devenues très importantes ; elles constituent des véritables enjeux des entreprises actuelles. Le diagnostic des systèmes est apparu dans le but d'améliorer les points précédents. Terme peu répandu pour les domaines techniques, il y a une trentaine d'année, le diagnostic est devenu un sujet important tant sur le plan théorique que pratique.

Quelle que soit la branche d'industrie concernée, aujourd'hui les procédés industriels sont couplés à un calculateur numérique qui ne se contente pas de faire l'acquisition des données mais qui est chargé de la mise en œuvre de l'automatisation. Automatiser peut avoir des objectifs divers. Les plus fréquents sont d'augmenter les performances du système de production, de garantir la qualité du produit fabriqué, ou de diminuer les coûts de fabrication. Mais un autre objectif est de plus en plus pris en compte par les systèmes de contrôle-commande : améliorer la sécurité de l'installation industrielle, des hommes qui y travaillent et de son environnement. D'autres objectifs peuvent s'y rajouter tels que garantir la satisfaction des normes de qualité. On arrive ainsi à une automatisation qui ne s'attache pas seulement au maintien de quelques variables à des valeurs de référence, mais qui s'intéresse au procédé dans sa globalité, et dans ses divers modes de fonctionnement.

Ce chapitre est organisé de la façon suivante. Les différentes notions utilisées en diagnostic (défauts, détection, localisation, identification, résidu, symptôme) sont rappelées dans la section 2. La section 3 rappelle les caractéristiques souhaitables d'un système de diagnostic. La section 4 présente les transformations des mesures dans un système de diagnostic. Dans la section 5 une classification des approches de diagnostic sera présentée (approches à base de modèle quantitatif ; de modèle qualitatif ; approches à base d'extraction des caractéristiques quantitatives ; caractéristiques qualitatives), puis détaillée dans les sections 6, 7, 8 et 9. Ce chapitre se termine par une conclusion.

2 Supervision et diagnostic

La surveillance industrielle joue un rôle primordial dans l'augmentation de la disponibilité des installations industrielles, réduisant ainsi les coûts directs et indirects de la maintenance des équipements de production. Les coûts directs de cette maintenance sont ceux relatifs aux diverses pièces de rechange, main d'œuvre, etc. par contre, les coûts indirects sont essentiellement dus au manque à gagner engendré par un arrêt de la production. Par conséquent, les entreprises soucieuses d'avoir une meilleure maîtrise des coûts de maintenance ont intérêt à adopter une politique de surveillance.

La sûreté est définie comme un ensemble de propriétés indispensables à une installation industrielle. La surveillance se réfère à la capacité de reconnaître un comportement anormal et de le signaler. La supervision est fondée sur la surveillance et la capacité d'agir, pour éviter des dommages sur l'installation et pour continuer la production, si cela est possible. Une défaillance est définie comme l'inaptitude d'un composant à remplir toute ou partie de sa fonction. Un défaut est un écart entre un comportement attendu et un comportement observé. Constaté grâce à un indicateur de défaut, il révèle la présence d'une défaillance. L'indicateur de défaut est en pratique élaboré grâce aux variables mesurables de l'installation.

Un système physique, ou un procédé technologique, est conçu pour réaliser une mission donnée. Cette mission n'est pas toujours remplie de la même façon, elle peut être totalement ou partiellement remplie avec différents niveaux de qualité. L'évolution de ce procédé (ou système) au cours du temps peut être décrite par différents modes de fonctionnement. On distingue en général :

- Les modes normaux sous lesquels la mission est remplie, l'un d'eux est le mode nominal où la mission est correctement remplie,
- Les modes évolutifs qui caractérisent des modes, souvent transitoires, de passage d'un mode de fonctionnement à un autre, les modes de démarrage et d'arrêt, qui sont évidemment pris en compte au moment de la conception du système de contrôle-commande.
- Les modes anormaux sous lesquels la mission est partiellement ou totalement non remplie, incluant les modes défaillants, les modes interdits, qui sont les différents états que peut prendre un procédé sous l'effet d'une panne ou d'un dysfonctionnement. Ces modes particuliers peuvent nécessiter soit d'arrêter immédiatement l'installation, soit de basculer la commande vers un mode de repli, ou encore vers un mode dégradé qui consiste à modifier les consignes des boucles locales ou même carrément l'architecture de la commande ou du procédé.

Le diagnostic d'un système est l'identification du mode de fonctionnement, à chaque instant, par ses manifestations extérieures (symptômes). Son principe général consiste à confronter les données relevées au cours du fonctionnement réel du système avec la connaissance que l'on a de son fonctionnement normal ou défaillant. Si le mode de fonctionnement identifié est un mode défaillant, le système de diagnostic devrait localiser sa cause. Il est évident que si l'on veut représenter tous les défauts possibles d'une installation industrielle, la tâche peut être considérable. En conséquence, la plupart des méthodes que nous évoquerons ci-dessous se

fondent sur les connaissances du comportement normal de l'installation. Mais il est bien clair que si l'on a une bonne connaissance des anomalies possibles, il faut l'utiliser pour améliorer la surveillance et le diagnostic.

On distingue généralement, dans la procédure de diagnostic, deux étapes principales : détection et localisation, plus l'identification pour certaines méthodes. Les définitions précises de ces notions ainsi que les différents types de défauts et la définition des termes "résidu" et "symptôme" (souvent utilisés en diagnostic) sont rappelés dans la suite de cette section.

Types de défauts : Un défaut peut être défini comme une modification suffisamment importante et permanente des caractéristiques physiques d'un système ou d'un composant. Un système défaillant (qui subit un défaut) est incapable d'accomplir correctement sa ou ses fonctions avec les performances requises. Comme le montre la Figure 1.1, trois types de défauts sont distingués.

- Les défauts actionneurs,
- Les défauts procédé,
- Les défauts capteurs.

Il est clair ainsi que pour s'attaquer à ce problème de diagnostic, il faut des connaissances approfondies sur l'installation : connaissances de son comportement normal ou anormal. Souvent un défaut est modélisé avec les mêmes outils que ceux utilisés pour représenter le procédé en état normal.

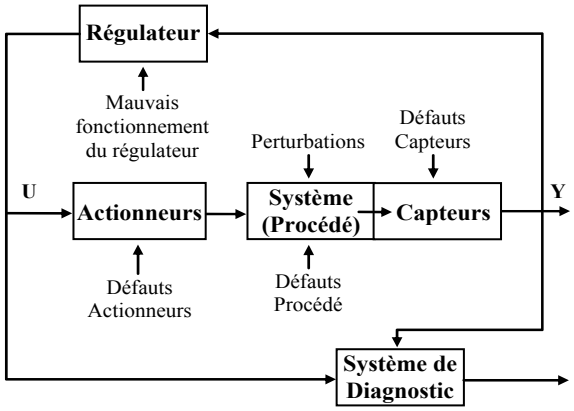


Figure 1.1 — Types de défauts

Deux approches sont adoptées pour la modélisation de défauts dans la littérature [28][39]. Dans la première, les défauts sont considérés comme des termes additionnels dans le modèle d'état, par exemple, il s'agit d'un signal parasite qui vient se superposer en un point du schéma fonctionnel. Un signal additif sur un capteur peut modéliser son *offset* (signal constant). De même, un signal additif sur un actionneur peut être utilisé pour modéliser un défaut. Ceci est raisonnable pour les défauts agissant sur les actionneurs et les capteurs et non pas les défauts agissant sur le système lui-même qui sont plutôt multiplicatifs, et qui font l'objet de l'autre approche consistant à modéliser les défauts comme des changements dans les paramètres du

système [28]. Il ne faut pas confondre aussi un défaut dû à une défaillance imprévue et un mauvais réglage des boucles de régulations, qui pourrait entraîner un pompage de certaines variables.

Détection : La détection (*Fault detection*) est la première décision que doit prendre un système de surveillance. C'est le fait de décider qu'un indicateur de défaut est pertinent et donc de signaler si le système est défaillant ou pas. La détection revient à choisir entre deux hypothèses H_n et H_d. La première correspond au fonctionnement normal, la deuxième correspond au fonctionnement défaillant. Pour détecter les défauts d'un système, il faut donc être capable de classer les situations observables comme étant normales ou anormales.

Localisation : La localisation (*Fault isolation*) est le fait de déterminer le(s) sous-ensemble(s) fonctionnel(s) défaillant(s). Cet ensemble a intérêt à être le plus réduit possible. Dans le cas où les défauts à détecter sont définis, ils constituent autant d'hypothèses H_{di} entre lesquelles le système de surveillance doit décider. La localisation est plus difficile dans le cas de défauts multiples où plusieurs défauts peuvent avoir lieu en même temps.

Identification : L'identification ou l'estimation (*Fault evaluation*) vise à fournir des informations quantitatives sur les caractéristiques du défaut : instant d'apparition, valeur de l'écart d'un paramètre permettant de le quantifier. Dans le cas de défauts procédé, l'identification revient à connaître les valeurs de différents paramètres du système.

Résidu : Les résidus (*Residual*) sont définis comme étant des indicateurs de défauts. Ils expriment l'écart par rapport au fonctionnement normal dans un espace de représentation donné. Cet espace dépend de la nature des informations manipulées (variable numérique, booléenne, variable qualitative, forme). C'est sur la base de leurs valeurs que la décision de détection est prise ou pas. Souvent, les résidus sont considérés nuls (statistiquement) en l'absence de défaut.

Symptôme : Un symptôme (*Symptom*) est un phénomène (caractère) qui révèle un trouble fonctionnel. Les alarmes sont des symptômes de comportement anormal souvent utilisés pour facilité la surveillance et la supervision. Dans un système traditionnel, il s'agit de surveiller simplement que les variables restent à l'intérieur d'un domaine de valeurs caractéristiques du fonctionnement normal. On fixe donc un seuil haut SH et un seuil bas SB pour chaque variable et l'on vérifie qu'à chaque acquisition la variable appartient à l'intervalle [SB, SH]. On peut remarquer qu'implicitement, cela revient à utiliser un modèle fort simple pour chaque variable : elle doit rester dans l'intervalle défini par les seuils. Les alarmes conviennent donc essentiellement à la surveillance du procédé en régime établi autour d'un point de fonctionnement constant. Sur certaines variables, très importantes, on peut fixer quatre valeurs de seuils : deux seuils d'alarme, définis comme précédemment et deux seuils de pré-alarme (l'un plus faible que SH et l'autre plus fort que SB), qui sont utilisés pour focaliser l'attention de l'opérateur humain sur cette variable, lorsqu'elle commence à s'écarter de sa valeur nominale. Des seuils peuvent être placés, de la même façon, sur la variation des variables [14].

3 Caractéristiques souhaitables d'un système de diagnostic

Dans le paragraphe précédent, le problème général de diagnostic des défauts a été présenté. Afin de pouvoir comparer diverses approches de diagnostic, il est utile d'identifier un ensemble de caractéristiques souhaitables qu'un système de diagnostic devrait posséder [28]. Bien que ces caractéristiques ne soient habituellement possédées par aucune méthode de diagnostic particulière, elles sont utiles pour évaluer diverses méthodes en termes d'information *a priori* qui doit être fournie, de rapidité, de robustesse, d'adaptabilité, etc. Chaque fois qu'une anomalie se produit dans un procédé, un système de diagnostic fournit un ensemble d'hypothèses de défauts qui explique les observations. L'efficacité d'un système de diagnostic demande que le défaut réel soit un sous-ensemble de l'ensemble des défauts suspectés. La résolution d'un système de diagnostic demande que l'ensemble des défauts fourni soit aussi minimal que possible. Les paragraphes suivants présentent un ensemble de caractéristiques souhaitables qu'un système de diagnostic doit posséder.

3.1 Rapidité et sensibilité de détection

Un système de diagnostic devrait répondre rapidement en détectant et en diagnostiquant les mauvais fonctionnement du procédé. Cependant, les réponses rapides du diagnostic et les performances tolérables durant un fonctionnement normal sont deux buts différents. Un système qui est conçu pour détecter un défaut rapidement sera sensible aux influences de toutes les erreurs (bruit, incertitudes…). Ceci rend le système sensible au bruit et peut déclencher fréquemment des fausses alarmes pendant un fonctionnement normal.

3.2 Robustesse

On voudrait que le système de diagnostic soit robuste aux divers bruits et incertitudes. On voudrait que ses performances se dégradent lentement au lieu d'échouer totalement et abruptement. La robustesse est incompatible avec les systèmes très sensibles aux seuils. En présence du bruit, ces seuils doivent être choisis raisonnablement. Ainsi, les besoins de robustesse doivent être équilibrés avec ceux de sensibilité.

3.3 Identification de nouveaux défauts

Une des fonctionnalités minimales qu'un système de diagnostic doit posséder, est de pouvoir décider, étant donné des conditions courantes de fonctionnement, si le procédé fonctionne normalement ou anormalement, et s'il est anormal, si la cause est un défaut de fonctionnement connu ou un inconnu. Ce critère est connu comme identification de nouveauté (nouveaux événements). En général, on dispose de données suffisantes pour décrire le comportement normal du procédé. Cependant, typiquement on n'a pas de tels historiques de données disponibles pour modéliser les modes de fonctionnement anormaux d'une manière satisfaisante. Seuls quelques gabarits de données peuvent être disponibles pour couvrir une partie du fonctionnement anormal. Ainsi, il est possible qu'une grande partie de la région de dysfonctionnement ne puisse avoir été modélisée. Ceci posera des défis sérieux pour l'identification de nouvelles situations. Même dans ces conditions difficiles, on voudrait que le système de diagnostic puisse identifier l'occurrence des nouveaux défauts et ne les classifie pas comme d'autres défauts de fonctionnement connus ou comme fonctionnement normal.

3.4 Capacité de localisation

La localisation est la capacité du système de diagnostic à distinguer entre différents défauts. Dans des conditions idéales exemptes de bruit et d'incertitudes de modélisation, ceci revient à dire que le système de diagnostic devrait pouvoir générer un ensemble d'hypothèses qui ne contient pas les défauts absents. Naturellement la capacité de concevoir de tels systèmes dépend en grande partie de la connaissance de la structure et des caractéristiques du procédé.

3.5 Adaptabilité

En général les procédés évoluent en raison des changements des entrées externes ou des changements structurels. Les conditions de fonctionnement du procédé peuvent changer non seulement en raison des perturbations mais également en raison des changements des conditions environnementales telles que les changements des quantités de production avec des demandes variantes, des changements de qualité de matière première, etc. Ainsi le système de diagnostic devrait être adaptable aux changements. Il devrait être possible de développer graduellement la portée du système pendant que les nouvelles situations émergent grâce à la disponibilité d'informations supplémentaires.

3.6 Identification des défauts multiples

La capacité d'identifier les défauts multiples est un objectif important, mais difficile, à cause de l'influence simultanée de la plupart des défauts. Dans un système non linéaire général, des interactions se manifestent, et par conséquent un système de diagnostic ne peut pas employer les différents modèles de chaque défaut pour modéliser l'effet combiné des défauts. Par ailleurs, l'énumération des différentes combinaisons de défauts multiples et l'analyse de leurs effets devient combinatoire pour des procédés de grande échelle.

3.7 Estimation d'erreur de classification

Une condition pratique importante pour un système de diagnostic, est l'établissement des niveaux de confiance. Ceci pourrait être considérablement facilité si le système de diagnostic pouvait fournir une évaluation a priori sur l'erreur de classification qui peut se produire. De telles mesures d'erreur seraient utiles pour projeter des niveaux de confiance sur les décisions produites par le système donnant à l'utilisateur une meilleure sensation à propos de la fiabilité des recommandations proposées par le système de diagnostic.

3.8 Facilité d'explication

En dehors de la capacité d'identifier la source d'un défaut de fonctionnement, un système de diagnostic devrait également fournir des explications sur la façon dont le défaut a commencé et a évolué vers la situation actuelle. C'est un facteur très important lorsque l'on conçoit des systèmes interactifs d'aide à la décision en ligne. Ceci exige la capacité de raisonner au sujet de la cause et des effets dans un procédé. Un système de diagnostic doit justifier ses recommandations de sorte que l'opérateur puisse en conséquence évaluer et agir en utilisant son expérience. On voudrait que le système de diagnostic justifie non seulement pourquoi certaines hypothèses ont été proposées, mais expliquer également pourquoi certaines autres hypothèses n'ont pas été proposées.

4 Transformation des mesures dans un système de diagnostic

En général, le processus de diagnostic peut être défini comme un processus décisionnel. La décision finale est constituée de l'ensemble des défauts possibles. La Figure 1.2, montre les diverses transformations que les données du procédé subissent pendant le processus de diagnostic. L'espace des mesures est défini par les sorties du système, ces sorties peuvent avoir subi des prétraitements (élimination de données aberrantes, filtrage,…). Ce sont les entrées du système de diagnostic.

L'espace des caractéristiques est l'espace obtenu en fonction des mesures en utilisant une connaissance a priori du problème. Ici, les mesures sont analysées et combinées à l'aide d'une connaissance a priori du procédé pour extraire les caractéristiques utiles concernant le comportement du procédé, pour faciliter le diagnostic. Il y a deux manières de développer l'espace des caractéristiques à partir de l'espace des mesures, à savoir, la sélection et l'extraction des caractéristiques. Dans la sélection des caractéristiques, on choisit simplement quelques mesures importantes de l'espace original de mesure. L'extraction de caractéristiques est un procédé qui facilite, par l'utilisation de la connaissance antérieure du problème, une transformation de l'espace de mesure dans un espace de dimension inférieure. Par exemple, si l'existence d'une relation entre deux mesures est connue, le rôle de cette étape est d'identifier ce rapport. Une fois que ce rapport est identifié, plutôt que d'avoir deux ensembles de paramètres caractérisant les deux dimensions, il peut être possible de les représenter avec un unique ensemble.

Le passage de l'espace des caractéristiques à l'espace des décisions est habituellement fait pour satisfaire une certaine fonction objectif (telle que réduire au minimum la non détection). Cette transformation est réalisée en employant une fonction discriminante, dans certains cas en employant des fonctions simples de seuil. L'espace des décisions est l'espace correspondant au nombre de variables de décision obtenu par des transformations appropriées de l'espace des caractéristiques. L'espace des classes correspond au nombre de classes de défaut. En classant les défauts, nous indiquons catégoriquement à quelle classe un défaut appartient. L'espace des classes est ainsi l'interprétation finale du système de diagnostic fourni à l'utilisateur. La transformation de l'espace des décisions en espace de classes est faite en utilisant des fonctions de seuil, des gabarits ou un raisonnement symbolique, selon les circonstances.

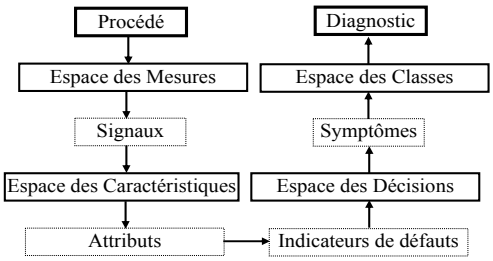


Figure 1.2 — Transformation des mesures dans un système de diagnostic

5 Classification des méthodes de diagnostic

Compte tenu de l'importance des enjeux en termes de productivité (arrêt inutile des installations), de sécurité (anomalie non détectée) ou de qualité de production (mesure incorrecte d'une grandeur à contrôler), de nombreuses approches ont été utilisées pour apporter une contribution à la solution du problème de diagnostic. On distingue cependant, parmi les différentes méthodes, deux familles principales ; celles qui utilisent un modèle du système à surveiller, et utilisent généralement les techniques de l'automatique ; et celles pour qui seules les données acquises sur le procédé considéré, qui se limitent à son observation passée et présente, permettent de caractériser son mode de fonctionnement (voir [27], et les nombreuses références contenues dans cet article). Parmi les méthodes basées sur l'utilisation d'un modèle du procédé, deux sous-familles importantes peuvent également être dégagées ; celles utilisant un modèle qualitatif et celles recourant à un modèle quantitatif (dit aussi modèle numérique). Pour certaines méthodes, cette distinction est cependant sans fondement, car elles utilisent de manière mixte les deux types de modèles en essayant de tirer profit des avantages et s'affranchir des limitations de chacun d'entre eux.

Toutes les méthodes nécessitent une phase d'apprentissage qui permet de connaître une référence de fonctionnement normal ou dégradé. Cette référence permet de générer en temps réel les symptômes lors de la phase de diagnostic. Les sections suivantes présentent les grandes familles de méthodes de diagnostic représentées par la Figure 1.3.

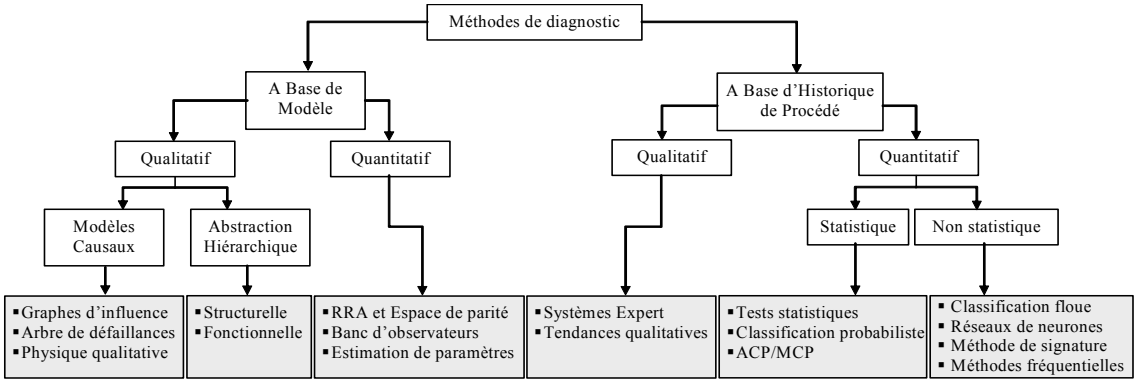


Figure 1.3 — Les méthodes de diagnostic[1]

- La section 6 présente des méthodes basées sur des modèles qualitatifs,
- La section 7 présente des méthodes à base de modèles quantitatifs,
- La section 8 présente l'approche à base de méthodes d'extraction des caractéristiques qualitatives,
- La section 9 présente l'approche à base de méthodes d'extraction des caractéristiques quantitatives.

[1] RRA : Relations de Redondance Analytique
ACP/MPC : Analyse en Composantes Principale / Moindres Carrés Partiels

6 Approche à base de modèle qualitatif

Dans certaines situations, il n'est pas possible d'obtenir un modèle mathématique précis de l'installation à diagnostiquer. Cependant, on possède des connaissances qualitatives sur son comportement. Cette connaissance s'exprime avec des symboles et peut être traitée, en simulant la façon dont un humain raisonne logiquement, par des techniques informatiques relevant de l'intelligence artificielle. Pour construire des représentations qui permettent une meilleure compréhension des phénomènes physiques mis en jeux dans un procédé, l'intelligence artificielle s'intéresse au milieu des années 1980 au raisonnement qualitatif. Le développement de techniques de représentation qualitative du monde pour en comprendre schématiquement le fonctionnement est une idée très naturelle et par conséquent très ancienne.

Le diagnostic est typiquement un processus causal, puisqu'il consiste à découvrir les composants défaillants qui sont la cause première du dysfonctionnement observé. La phase de localisation consiste à chercher quel composant d'un système ne fonctionne pas, en utilisant la connaissance sur la structure du système, ses défaillances potentielles, et les observations disponibles. C'est pourquoi l'approche causale peut être utilisée pour structurer des connaissances de diagnostic, qui après seront exploitées soit par un parcours de graphe, soit par une base de règles. La propagation d'un défaut suit une chaîne de sous effets jusqu'à provoquer un symptôme observable. Cette chaîne peut être connue pour des raisons physiques (quand la vanne s'ouvre, le débit augmente, etc.).

Quelquefois, ce sont des observations heuristiques qui permettent de relier défauts et symptômes. On peut s'intéresser au raisonnement causal sur le fonctionnement anormal de l'installation, dans lequel on associe pannes, effets non observables et effets observables (symptômes). On peut utiliser les analyses réalisées dans le cadre des études de sûreté et valoriser la connaissance heuristique sur les défaillances acquise au cours de l'expérience (règle empirique, intuition statistique, expérience compilée).

On peut encore utiliser le raisonnement causal sur le fonctionnement normal de l'installation. Le principe sous-jacent au raisonnement est alors que, si une chaîne causale est un modèle de la propagation des phénomènes physiques dans le processus, c'est non seulement une bonne représentation de la propagation des effets de changements normaux (commandes, perturbations mesurables), mais aussi un bon moyen de suivre la propagation des effets des défaillances. On peut ainsi regrouper entre eux des défauts dus à une même cause première. Evidement, ce dernier raisonnement suppose que la défaillance n'ait pas modifié la chaîne de propagation des phénomènes.

Un système de diagnostic comporte deux composants importants : le domaine de connaissance a priori et la stratégie de recherche. La connaissance a priori de base qui est nécessaire pour le diagnostic, est l'ensemble des défauts, plus le rapport entre les observations (symptômes) et ces défauts. Les modèles qualitatifs peuvent être développés en tant que modèles ou hiérarchies d'abstraction causales.

6.1 Modèles causaux

La notion de causalité a été largement discutée, dans ce paragraphe nous ne l'aborderons qu'à travers l'intérêt qu'elle a suscité durant ces dernières années dans le domaine de la supervision. En effet, le diagnostic est typiquement un processus causal, car il consiste à chercher les composants défectueux qui peuvent expliquer les dysfonctionnements observés [21]. Qu'il s'agisse de simulation, d'explication ou de diagnostic, la causalité apparaît comme un principe de base. Si l'on ajoute l'idée de dynamique de propagation des effets d'une commande, d'une perturbation ou d'un mode de défaillance, la notion de graphe causal dynamique pour la supervision s'impose.

La procédure de diagnostic peut être fondée sur des symptômes heuristiques ou de la connaissance qualitative sur le fonctionnement du procédé en défaut. Les symptômes heuristiques peuvent être des observations faites par des opérateurs (bruit, oscillations, couleur, odeur). Des observations quantitatives peuvent être transformées en symboles comme "vitesse élevée", "température faible". Cette transformation peut se faire de façon élémentaire, avec des seuils fixes, ou par fuzzification. Dans ce dernier cas, un sous-ensemble flou est défini comme la description symbolique d'un concept, par exemple "température faible", "température forte".

Le principe de base du diagnostic est le suivant : à partir d'un ensemble de symptômes observés, il faut construire une chaîne d'inférence pour obtenir un ensemble de défaillances, à l'origine de ces symptômes.

6.1.1 Graphes d'influence ou graphe causaux

Le graphe causal peut être utilisé pour représenter le fonctionnement normal du procédé. Les nœuds représentent les variables. Ils sont reliés par des arcs orientés. Un arc orienté de x vers y à l'instant t dépend de l'état de x à un instant précédent ; x est la cause, y est l'effet. On essaye d'expliquer les déviations des variables par rapport à leur comportement normal par un minimum de défauts sources. Un arc peut porter un certain nombre d'informations qualitatives, comme le sens de variation des variables, l'ordre de grandeur de l'effet, et même des informations quantitatives comme des fonctions de transfert. Le modèle causal du fonctionnement normal d'une installation peut être obtenu soit à partir de l'expertise sur le procédé, soit par une analyse des équations formelles du modèle lorsqu'elles sont connues [15].

Les variables de détection sont les variables dont l'évolution est considérée comme anormale. Deux types de détection peuvent être rencontrés.

- L'expérience permet la définition d'un seuil d'anormalité pour chaque variable (principe des systèmes classique de traitement d'alarmes) ;
- L'anormalité est déterminée relativement à un modèle de comportement normal.

Les variables sources sont des variables dont l'anormalité permet d'expliquer toutes les autres anormalités rencontrées dans le graphe, au sens des relations portées par les arcs. Ce sont les variables affectées les premières par la défaillance. Le graphe suspect est le sous graphe de propagation des effets de la défaillance : il inclut toutes les variables de détection ; ses racines sont les variables sources.

Une approche fondée sur un graphe intégrant des connaissances sur la dynamique du système a été introduite dans [22]. Une application est décrite en détail dans [12]. Cette approche permet de tenir compte du fait que les défaillances se propagent dans un procédé avec un certain temps de réponse, ou encore des retards, qu'il faut prendre en compte si l'on veut garantir le suivi dynamique des défauts et la création d'alarmes intelligentes (c'est-à-dire reliées entre elles dans le temps quand elles ont une même cause première).

La structure la plus pauvre de graphe causal est celle du graphe signé (*SDG : Signed Directed Graph*). Les nœuds d'un SDG correspondent à l'état des variables, et les arcs orientés portent les signes des influences correspondantes : le signe est « + » lorsque les variables se rapportant à l'arc évoluent dans le même sens, et le signe « - » lorsqu'elles évoluent en sens opposé. Les approches qui utilisent un graphe causal paraissent bien adaptées aux tâches de compréhension du comportement normal ou anormal du procédé, de par la représentation graphique implicite du raisonnement causal qu'elles proposent.

6.1.2 Arbres de défaillances

L'arbre de défaillance est l'un des outils majeurs d'analyse du risque technologique. A l'origine c'est une méthode développée à Bell Telephone Laboratories en 1961 aux Etats-Unis [30]. C'est la méthode la plus couramment utilisée dans les analyses de fiabilité, de disponibilité ou de sécurité des systèmes. Il s'agit d'une méthode optimisée qui permet la détermination des chemins critiques dans un système. Cette méthode a pour objet de déterminer les diverses combinaisons possibles d'événements qui entraînent la réalisation d'un événement indésirable unique. Elle permet donc d'identifier les points faibles de la conception.

C'est une méthode déductive dont la représentation graphique des combinaisons est réalisée par une structure arborescente (arbre), permettant un traitement à la fois qualitatif et quantitatif. Cet arbre est établi sous forme d'un diagramme logique et comporte au sommet l'événement indésirable. Les causes immédiates qui produisent cet événement sont ensuite hiérarchisées à l'aide de symboles logiques "*ET*" et "*OU*". De cette façon, l'arbre est créé pas à pas pour atteindre, à la base, un ensemble d'événements considérés comme élémentaires. Une extension des arbres de défaillances est la méthode du diagramme de causes−conséquences [30]. Cette méthode combine les principes utilisés par l'analyse déductive de l'arbre des défaillances et l'analyse inductive de l'arbre des conséquences.

Le principal inconvénient des arbres de défaillances est que le développement est sensible aux erreurs à différentes étapes. En effet, l'arbre construit est seulement aussi bon que le modèle mental de son créateur. Pour exécuter un diagnostic correct à partir des arbres de défaillances, ceux-ci doivent largement représenter toutes les relations causales du procédé, c'est-à-dire qu'ils doivent être capables d'expliquer tous les scénarios de défauts possibles. De plus, l'emploi de cette méthode se révèle difficile pour les systèmes fortement dépendants du temps. Enfin, il n'y a pas de méthode formelle pour vérifier l'exactitude de l'arbre développé [32].

6.1.3 Méthodes à base de physique qualitative

Dans certains cas, il devient très difficile de disposer des connaissances complètes pour faire un modèle analytique du procédé à cause de sa complexité structurelle ou de son comportement. Une alternative pour traiter les connaissances incomplètes est de faire des abstractions selon le principe du raisonnement qualitatif [26]. Les modèles décrivent alors la structure du procédé à diagnostiquer. Généralement, il s'agit des connexions entre les composants. L'approche qualitative exprime explicitement le lien entre un composant et les formules décrivant son comportement.

La physique qualitative a pour but de représenter un système physique et d'expliquer son comportement, en se basant à la fois sur le raisonnement causal de sens commun utilisé par l'être humain pour analyser qualitativement son environnement, et sur les connaissances scientifiques implicitement utilisées par les ingénieurs. La représentation des variables continues doit être guidée par le principe suivant : les distinctions faites par une quantification doivent être pertinentes pour le type de raisonnement effectué.

La simulation qualitative du comportement du système physique se distingue de la simulation numérique classique, entre autres parce que le temps y est échantillonné selon l'occurrence d'événements intéressants, plutôt que selon un incrément régulier et fixé. Ces remarques expliquent pourquoi la physique qualitative semble appropriée à la supervision des procédés continus, dont un des objectifs est d'expliquer le fonctionnement d'un procédé aux opérateurs, avec un critère de pertinence plus que de précision, afin de les assister dans leurs prises de décision.

Parmi les méthodes qui utilisent les modèles physiques de type qualitatif pour la détection et le diagnostic des défauts, se trouve la prédiction du comportement à partir des équations différentielles qualitatives (QDES). Ces QDES sont des abstractions des équations différentielles ordinaires qui utilisent des intervalles de variables et des paramètres. L'algorithme QSIM représente un système comme un ensemble d'équations différentielles qualitatives (QDES) couplées. Le comportement du système est alors représenté par une séquence d'états.

Selon l'approche qualitative, le diagnostic est basé sur la théorie du raisonnement logique [25][26]. Pour le diagnostic, les termes *suspect* et *candidat* sont utilisés. Le premier fait référence à tout composant, que le système de diagnostic a identifié comme responsable possible de l'apparition d'une divergence. Tandis que le deuxième fait référence au composant, ou ensemble de composants, dont le fonctionnement incorrect expliquerait toutes les divergences observées. Une caractéristique de cette technique est que le diagnostic se fait de façon itérative en trois phases :

- La détection de conflits: recherche de l'ensemble de composants suspects en fonction des divergences observées ;
- Le diagnostic : recherche des candidats parmi les suspects ;

- La discrimination des hypothèses : raffinement de l'ensemble de candidats dans le cas où il y en aurait plusieurs.

6.2 Abstractions hiérarchiques

Les méthodes d'abstraction hiérarchique pour le diagnostic sont des méthodes issues des études de sûreté de fonctionnement, qui peuvent être utilisées pour identifier les causes des défaillances d'un procédé industriel. Il s'agit, en fait, de l'analyse fonctionnelle et structurelle du système.

6.2.1 Abstractions hiérarchiques structurelle ou matérielle

Parmi ces méthodes, la méthode SADT (*Structured Analysis and Design Technique*) qui est une marque déposée de la société Softech Inc (USA) et d'IGL Technologie (France). Elle a été développée à la fin des années 70 et elle est particulièrement adaptée à l'analyse et à la conception des systèmes [32]. Elle facilite la communication entre spécialistes de disciplines différentes. Elle fournit des outils pour :

- Concevoir d'une façon structurée des systèmes vastes et complexes ;
- Communiquer des résultats de l'analyse et de la conception dans une notation claire et précise ;
- Contrôler l'exactitude, la cohérence et de façon générale, la qualité de manière continue et systématique par des procédures particulières de critiques et d'approbations ;
- Documenter l'analyse du système, la chronologie de l'étude, les décisions et les résultats courants.

6.2.2 Abstractions hiérarchiques fonctionnelles

La décomposition fonctionnelle arborescente fait appel aux concepts de la programmation structurée en utilisant une représentation descendante. Pour chaque fonction principale (niveau 0) réalisée par un système on identifie les fonctions de niveau inférieur (niveau -1) en utilisant éventuellement des symboles logiques ″ET″ et ″OU″ pour prendre en compte les redondances éventuelles.

Cette décomposition est ensuite répétée au niveau inférieur pour faire apparaître les fonctions de niveau -2. Le degré de profondeur de la décomposition est fonction du niveau de détail où l'on souhaite réaliser le diagnostic des défaillances. Les arbres fonctionnels présentent des avantages indiscutables pour faciliter la recherche des liens entre les causes et les effets de défaillances :

- Un outil flexible et modulaire qui permet d'arrêter l'analyse dans les différentes branches à un niveau désiré,
- Un outil qui apporte au concepteur une visualisation qui formalise le raisonnement,
- Un outil qui se prête à l'informatisation : le caractère arborescent de l'analyse permet de l'organiser facilement pour réaliser ultérieurement des études de fiabilité et de disponibilité,
- Un outil adapté au problème de diagnostic car il permet des approches inductives et déductives.

La connaissance sur les modes de défaillance d'un procédé est quelquefois établie de manière systématique, dès leur conception, grâce à une AMDEC (Analyse des Modes de Défaillance, de leurs Effets et leur Criticité), effectuée par des ingénieurs en sûreté. Il est alors astucieux de

capitaliser cette connaissance d'une manière systématique. Elle peut facilement s'exprimer sous la forme d'un graphe défaut défaillance. Dans cette représentation, les racines du graphe sont les défaillances primaires, et les feuilles sont les symptômes observables. On parle d'arbre d'événement quand on représente la connaissance à partir des symptômes vers les défaillances. La connaissance exprimée par ce graphe peut être manipulée soit par un algorithme de parcours de graphe, soit par un système à base de règles.

Parmi ces méthodes l'AMDE (*Analyse des Modes de Défaillance, et leurs Effets*), fût employé pour la première fois, à partir des années 1960, dans le domaine de l'aéronautique pour l'analyse de la sécurité des avions. Cette méthode permet une analyse systématique et très complète, composant par composant, de tous les modes de défaillance possibles et précise leurs effets sur le système global [28].

La démarche consiste d'abord à définir le système, ses fonctions et ses composants. Ensuite, l'ensemble des modes de défaillances des composants doit être établi. Pour chaque mode de défaillance, sont recherchées ensuite les causes possibles de son apparition. Finalement, une étude des effets sur le système et sur l'opérateur est faite pour chaque combinaison (cause, mode de défaillance). Les résultats sont présentés sous forme de tableau.

L'AMDE est parfois complétée par une analyse de criticité ; elle devient AMDEC (*Analyse des Modes de Défaillances de leur Effet et de leur Criticité*). La criticité permet d'extraire les modes de défaillance les plus critiques. L'analyse de criticité est évaluée à partir du couple probabilité–gravité. L'analyse de fonctionnement (HAZOP : *Hazard and Operability Stydy*) est une autre extension de l'analyse des modes de défaillances (AMDE). La méthode HAZOP explore systématiquement l'aspect fonctionnel d'un système en identifiant à la fois les dangers et les dysfonctionnements d'une installation. Elle est mise en œuvre à la fin de la phase de conception puisqu'elle s'appuie sur les schémas détaillés du système.

L'utilisation des tableaux d'AMDE(C) à des fins de diagnostic industriel conduit à utiliser une procédure déductive, c'est-à-dire à utiliser ces tableaux comme un outil d'identification des causes de défaillances à partir des effets observés [32]. La modélisation des relations cause à effet réalisée par l'AMDEC rend cette démarche très puissante pour la résolution des problèmes de diagnostic de défaillances de procédés industriels.

L'AMDE est donc très répandue dans de nombreux domaines industriels, l'aéronautique en particulier ; pourtant elle est lourde et insuffisante. En effet, il est nécessaire d'identifier a priori les défauts et/ou les dysfonctionnements pouvant apparaître, à un certain moment, dans le système supervisé ; le recensement préalable des défaillances et leurs relations éventuelles ne peut jamais être exhaustif et requiert en général une longue expérience. En plus, toute modification ou évolution du système nécessite une réécriture du tableau. Enfin, cette méthode ne peut traiter les cas de défaillances multiples et intégrer l'aspect fonctionnel.

7 Approche à base de modèle quantitatif

L'utilisation de modèle mathématique pour le diagnostic est très largement répandue. Les méthodes de surveillance à base de modèle (*Model-based- FDI : Fault Detection and Isolation*) utilisent un modèle et reposent sur la génération et l'étude d'un signal particulier appelé "indicateur de défaut" ou "résidu". Un résidu est un signal qui reflète la cohérence des données mesurées vis-à-vis d'un modèle comportemental du système. Les modèles utilisés peuvent être de nature et de complexité différentes. Ils peuvent être : à temps continu ou à temps discret, linéaires ou non linéaires, représentant le bon fonctionnement ou tenant compte des défaillances.

Les sorties du modèle se représentent classiquement sous la forme de valeurs numériques. Chaque relation du modèle peut être associée à un ou plusieurs composants industriels qui les sous tendent. Ils constituent le support de la relation. La localisation qui suit la détection peut être effectuée à l'aide d'une table de signature ou matrice d'incidence. Les colonnes de cette table sont représentatives des différents défauts et les lignes des différents résidus. Un 1 représente le symptôme qui indique que le résidu est sensible au défaut, un 0 le contraire. La signature d'un défaut (colonne de la matrice) définit l'état des symptômes lorsque ce défaut affecte le système (l'état est une grandeur logique ou symbolique).

Classiquement, en Automatique, des modèles dits de bon fonctionnement sont utilisés. Ils caractérisent le comportement normal du système, c'est-à-dire lorsque aucune défaillance n'est présente. En surveillance, par contre, il est parfois nécessaire de compléter le modèle afin de caractériser le comportement défaillant du système. Suivant la classe de modèle considérée, les techniques de génération de résidus seront différentes. Suivant le niveau de connaissance sur les défaillances, les résultats de la procédure de surveillance seront plus ou moins précis et complets.

Génération de résidus : Les résidus en diagnostic reflètent les défauts potentiels d'un système. L'étape suivante est de confirmer la présence d'un défaut et de l'identifier. C'est-à-dire, détecter et localiser le défaut. Afin de localiser le défaut, il est nécessaire de produire des résidus qui sont non seulement sensibles au défaut mais également sélectifs pour le défaut. À cet effet, le générateur de résidu doit pouvoir produire un ensemble de résidus plutôt qu'un seul, et construire des résidus qui répondent sélectivement à chaque défaut potentiel. Les résidus produits peuvent servir ainsi non seulement comme détecteur de défaut mais également de localisateur de défaut.

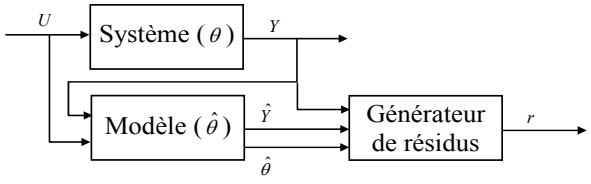


Figure 1.4 — Principe de génération de résidus

Il y a eu des travaux de conception de générateurs de résidus capables de produire des résidus qui favorisent l'analyse des défauts. Deux de ces méthodes sont l'approche résidus directionnels et l'approche résidus structurés. Les générateurs de résidus structurés sont conçus de telle manière que chaque résidu répond sélectivement à un sous-ensemble de défauts. Les différentes signatures de défaut permettent de bien les localiser. Les générateurs de résidus directionnels sont capables de produire des résidus qui sont dans une direction spécifique du défaut dans l'espace de résidus. Dans la suite nous détaillons les principales méthodes de diagnostic à base de modèle quantitatif. La Figure 1.4 schématise le principe de ces méthodes.

7.1 Equations de redondance analytique, espace de parité

Depuis les travaux fondateurs de Potter et Suman [40], le problème de génération de relations de redondance analytique a fait l'objet de très nombreuses études [35][37][38]. Dans la suite, on donne un petit rappel sur le principe de cette méthode.

La redondance d'information et de données est une source de modèles de bon et de mauvais fonctionnement. Les relations de redondance analytique notées (RRA) sont obtenues à l'aide du modèle mathématique du système physique. Une RRA est une relation R (déduite par la combinaison de relations du modèle) qui lie entre elles des variables nécessairement mesurées représentées par l'ensemble Y et telle que $R(Y) = 0$.

La mise en oeuvre de la redondance analytique nécessite donc de construire des modèles testables, c'est à dire ne faisant intervenir que des variables disponibles (mesurées ou connues à l'avance). L'existence de tels modèles est conditionnée par celle d'une redondance d'informations sur le système. Une RRA peut être vérifiée pour diagnostiquer les composants physiques qui constituent son support. Des RRA de support différent (sensible à des défauts particuliers) peuvent être générées, en effectuant des combinaisons judicieuses des relations de modèle.

Lorsqu'un modèle comportemental est disponible, il est possible de l'exploiter pour tester la cohérence des informations prélevées en ligne sur le système. On parle alors de redondance analytique. Le problème de génération de relations de redondance analytique a fait l'objet de très nombreuses études. Une relation de redondance analytique est une équation dans laquelle toutes les variables sont connues. La génération de telles relations permet d'engendrer des résidus. En l'absence de défaut, ce résidu est statistiquement nul. Lors de l'apparition d'un défaut, son amplitude évolue de manière significative.

Une autre approche plus classique est dite espace de parité. Les relations de parité utilisent la redondance directe au moyen de relations algébriques statiques liant les différents signaux ou la redondance temporelle issue de l'utilisation de relations dynamiques. Pour ces méthodes, le nombre de mesures doit être grand et les résidus sont issus de la comparaison de mesures redondantes. Ces notions ont ensuite été généralisées pour la redondance temporelle.

7.2 Bancs d'observateurs

Les observateurs d'état sont des algorithmes, fondés sur le modèle du procédé, chargés de poursuivre l'état de celui-ci. On essaye, en général de rendre l'observateur indépendant des

perturbations non mesurées et dépendant de certains défauts. Un aspect intéressant des observateurs est la possibilité de découpler leur réponse des entrées inconnues du système. Les observateurs ou filtres sont des outils bien connus des automaticiens à des fins de commande.

Le principe général est de concevoir un système dynamique permettant de donner une image, ou estimation, de certaines variables, ou combinaisons de variables, nécessaires au bouclage. Lorsque le système est dynamique et que certaines variables (conditions initiales) sont inconnues, l'estimation n'est correcte qu'après un certain temps de convergence, fixé par la dynamique de l'observateur. Ces outils ont été adaptés à des fins de diagnostic et les travaux utilisant ces approches sont nombreux. Le principe général consiste à comparer des fonctions de sorties estimées avec les mêmes fonctions des sorties mesurées. L'écart entre ces fonctions est utilisé comme résidu.

L'idée de base des méthodes de diagnostic par observateurs est d'estimer les sorties du système à partir des mesures (ou une partie de mesures) en utilisant par exemple les observateurs de Luenberger dans le cas déterministe ou les filtres de Kalman dans le cas stochastique. Ensuite l'erreur d'estimation de la sortie est utilisée comme un résidu. Souvent, lors de la modélisation d'un système, on tient compte des entrées qui ne sont pas mesurables et qui influencent son état. La reconstruction d'état de ces systèmes ne peut se faire que sous certaines conditions ; les observateurs utilisés portent le nom d'observateurs à entrées inconnues.

Dans certains cas, l'utilisation d'un seul observateur ne suffit pas pour arriver à la localisation des défauts. Plusieurs observateurs, mis en batterie, peuvent alors être utilisés pour garantir la localisation des défauts. Dans la littérature [10][19][39], on peut trouver aussi les observateurs adaptatifs qui estiment l'état et quelques paramètres internes des systèmes. Ces observateurs peuvent être utilisés pour la détection et la localisation des défauts du procédé.

Notons que l'utilisation de la méthode de diagnostic à base d'observateur est conditionnée par l'existence de ce dernier. Pour les systèmes non linéaires, la construction d'observateur est difficile même en l'absence de défauts. En général les conditions nécessaires et suffisantes pour la construction des observateurs pour ces systèmes ne sont pas connues. La construction d'observateur est possible pour certains systèmes non linéaires possédant des structures assez particulières (systèmes qui peuvent être transformés par un difféomorphisme en des systèmes linaires, systèmes possédant une structure triangulaire, systèmes affines en entrée…).

7.3 Estimation de paramètres

Le diagnostic des défauts à base de modèle peut être effectué en utilisant les techniques d'identification. Cette approche est basée sur l'hypothèse que les défauts se reflètent sur les paramètres physiques du système. Elle suppose l'existence d'un modèle paramétrique décrivant le comportement du système et que les valeurs nominales des paramètres soient connues. Elle consiste alors à identifier les paramètres caractérisant le fonctionnement réel, à partir des mesures des entrées et des sorties du système.

On suppose que l'on connaît un modèle du procédé dont les paramètres caractérisent un

fonctionnement normal. On estime les paramètres de ce modèle ; quand les valeurs des paramètres ont sensiblement varié, on décide que cela est dû à un défaut. Si l'on peut, de plus, établir un lien entre les valeurs des paramètres et certains composants précis, on a là une méthode de localisation. Cette méthode est surtout intéressante lorsque le modèle est un modèle continu fondé sur des connaissances premières sur le procédé. Elle est donc bien adaptée aux défauts multiplicatifs, lorsque la structure du modèle est bien connue et que les paramètres ont un sens physique [14]. Sinon, le lien entre la modification des paramètres du modèle et les composants devient difficile et la localisation s'en trouve fortement compliquée.

On dispose ainsi d'une estimation des paramètres du modèle, effectuée à partir des mesures prises sur le système et leurs valeurs théoriques. Pour détecter l'apparition de défaillances dans le système, il faut effectuer la comparaison entre les paramètres estimés et les paramètres théoriques. La théorie de décision sert alors à déterminer si l'écart observé est dû à des aléas normaux du fonctionnement ou à des défaillances. Cette approche utilise souvent le modèle mathématique d'entrée sortie du système sous la forme suivante :

$$y(t) = f\left(\theta, u(t), y(t-1), ..., y(t-k)\right) \tag{1.1}$$

où θ est le vecteur des paramètres qui est relié directement aux paramètres physiques du système. La fonction $f(.,.)$ peut prendre une forme linéaire ou non linéaire.

Pour générer les résidus par cette méthode, un algorithme d'identification en ligne des paramètres doit être utilisé. Si on a l'estimation du vecteur des paramètres $\hat{\theta}_k$ à l'instant k, le vecteur des résidus peut être défini de la façon suivante :

$$r(k) = \hat{\theta}_k - \theta_0 \tag{1.2}$$

où θ_0 est le vecteur nominal des paramètres. Evidement, comme pour tout problème d'estimation de paramètres, si le procédé est mal excité, l'estimation converge mal ou pas du tout, et les résultats ne sont pas interprétables [14].

8 Approche à base d'extraction de caractéristiques qualitatives

Ces méthodes ont l'avantage de ne pas nécessiter la connaissance d'un modèle mathématique ou structurel du procédé, contrairement aux méthodes à base de modèles. Seulement, la disponibilité de grandes quantités de données historiques du procédé est nécessaire. Il y a diverses façons de traiter ces données et de les présenter en tant que connaissance *a priori* pour le diagnostic. Ceci est connu comme l'extraction de caractéristiques. Parmi les méthodes d'extraction de caractéristiques qualitatives, se trouvent les méthodes à base de règles (systèmes experts), et les méthodes de représentation et analyse qualitatives des tendances.

8.1 Systèmes experts

Le diagnostic industriel est devenu depuis les deux dernières décennies un des domaines privilégiés d'applications pour les systèmes experts ou systèmes à base de connaissances. En effet, ils possèdent la propriété essentielle de pouvoir restituer à des non experts les

connaissances acquises par les spécialistes d'un domaine technique précis. Dans la plupart des cas, les connaissances utilisées pour le développement d'un système expert d'aide au diagnostic reposent sur l'apprentissage des relations entre les causes et les effets observées pour chaque défaillance. Avec ce mode de connaissance, les systèmes experts d'aide au diagnostic appartiennent à la famille des méthodes sans modèles. Cependant, l'utilisation de modélisation fonctionnelle décrivant les comportements détaillés des composants de systèmes complexes permet également le développement de systèmes experts appartenant à la famille des méthodes de diagnostic à base de modèles [32].

Un système expert est un système informatique destiné à résoudre un problème précis à partir d'une analyse et d'une représentation des connaissances et du raisonnement d'un (ou plusieurs) spécialiste(s) de ce problème. Les systèmes experts sont utilisés de façon privilégiée dans des domaines où l'homme manipule des informations factuelles, des données partielles et des stratégies de raisonnement. Le système expert doit fournir des réponses associées à une situation donnée sachant que la complexité du domaine étudié ne permet pas d'établir a priori toutes les configurations possibles des situations. Ainsi un spécialiste du dépannage réalise son diagnostic à l'aide d'une description partielle de la situation.

A partir de l'ensemble de symptômes à sa disposition, il déduit toutes les conclusions possibles, élabore de nouvelles hypothèses et approfondit son diagnostic en exploitant des informations supplémentaires collectées sur le système à diagnostiquer. Par leurs caractéristiques intrinsèques, les systèmes experts traditionnels mettent en œuvre des raisonnements heuristiques qui peuvent se définir comme des parcelles de connaissance capables de sélectionner les meilleures actions dans une situation particulière. Les systèmes experts, contrairement aux systèmes informatiques algorithmiques classiques séparent les connaissances de leurs algorithmes d'utilisation. La difficulté spécifique de la mise en œuvre des systèmes experts est la formalisation de la démarche cognitive qui a pour objectif, à partir d'une situation donnée, de définir et de décrire le raisonnement associé.

Pour ces systèmes, les connaissances représentées sont alors subdivisées en deux types. Les connaissances descriptives, les assertions (symptômes, événements), peuvent prendre à peu près n'importe quelle forme. Elles sont stockées dans une base de données appelée base de faits (BF). Dans les représentations les plus simples, un fait est représenté par un symbole auquel on attribue une valeur de vérité {VRAI/FAUX}. Les connaissances opératoires prennent la forme de règles de production. Elles sont stockées dans une base de données appelée base de règles, (BR). L'expression générale d'une règle de production est donc :

SI cond1 ET cond2 ET (cond3 OU cond4) ALORS conclusion1 ET conclusion2

Des conditions (prémisses, antécédents) portent sur l'existence de faits dans la base de faits. Les conclusions (conséquents) sont dans le cas courant simple des ajouts de faits dans la base de faits ou des exécutions de procédure externes ; le système met en permanence à jour la base de faits à l'aide de la base de règles, jusqu'à ce que le but soit atteint (ici, l'ensemble des défaillances possible déterminé). Les règles peuvent porter sur des variables, que le moteur d'inférence remplacera au moment de l'exécution par les faits de la base de faits. Ceci permet

d'écrire facilement la base de règles, mais consomme plus de temps d'exécution.

Dans la phase préliminaire d'exploitation d'un système expert, les faits statiques sont introduits afin de décrire la situation associée à la résolution d'un problème donné. Les valeurs du fait peuvent être numériques, symboliques, multiples, incertaines ou indéterminées. Ensuite, lors du déclenchement du mécanisme d'inférence, de nouveaux faits peuvent être créés dynamiquement par déduction ou introduits par questionnement de l'utilisateur. En général, la structure de la base de faits n'est pas organisée a priori et son contenu s'enrichit de façon progressive.

8.2 Tendances qualitatives

Un formalisme général pour la représentation et l'analyse des signaux d'un procédé est la représentation par tendances qualitatives, qui a été proposée par [7]. Ils ont proposé le concept de tendance comme une série d'épisodes avec une représentation triangulaire. En effet, dans ce formalisme chaque épisode est représenté par sa pente initiale, sa pente finale (à chaque point critique) et un segment de droite reliant les deux points critiques. Janusz [16] a proposé une description qualitative des signaux (*TDL : Trend Description Language*) se composant de primitives, d'épisodes, de tendances et de profils. Les primitives sont basées sur le signe des dérivées premières et deuxième (positif, zéro, négatif). Ceci donne un ensemble de 9 primitives (voir Figure 1.5).

Un épisode est défini par une primitive et une durée. Le nombre de fois consécutives qu'une primitive est présente définie la durée de l'épisode. La tendance d'un signal consiste en une série d'épisodes, et un profil s'obtient en ajoutant des informations qualitatives aux tendances. Par la suite il y a eu plusieurs travaux qui ont amélioré le langage de description des tendances (TDL) au niveau du bruit [41], de sa représentation et de son interprétation [20][8] pour l'utiliser ensuite en combinaison avec d'autre techniques, surtout des techniques de reconnaissance de formes [4].

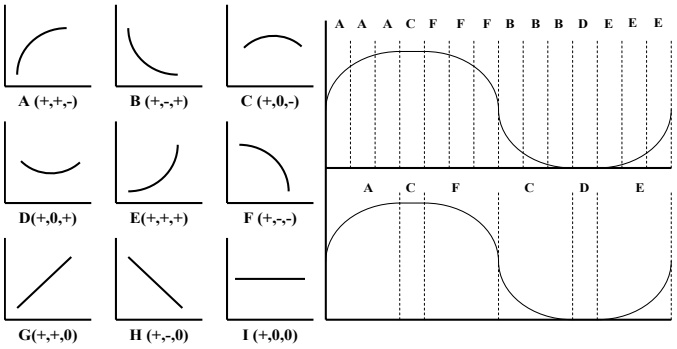


Figure 1.5 — Les primitives et la représentation d'une sinusoïde par TDL

Afin d'être plus générale, l'approche décrite par [20] a élargi les formalismes précédents vers des contextes qualitatifs et quantitatifs. Cette approche permet en effet de construire des épisodes à partir de toute caractéristique extraite des variables du procédé. Selon ce formalisme, une nouvelle représentation permet de décrire les tendances d'un signal selon sa dérivée seconde. La dérivée première qualifiée au début et à la fin de chaque épisode est employée pour

obtenir une représentation plus significative de la tendance du signal. Ainsi, un ensemble de 13 primitives est obtenu. Un avantage important de cet ensemble d'épisodes pour la surveillance est que les discontinuités et les périodes de stabilité (habituellement dans des situations de défaillance et dans des situations normales respectivement) sont explicitement représentées en utilisant 5 types d'épisodes. Cette approche a été utilisée en combinaison avec d'autres méthodes pour le diagnostic d'unités industrielles [36].

L'abstraction qualitative permet une représentation compacte de la tendance en ne prenant en compte que les événements significatifs. Pour des tâches telles que le diagnostic, la représentation qualitative des tendances fournit souvent des informations valables qui facilitent le raisonnement sur le comportement du procédé. Dans la plupart des cas, les dysfonctionnements du procédé produisent une tendance distinctive dans les signaux issus des capteurs. Ces tendances peuvent être employées pour l'identification des situations anormales dans le procédé. Ainsi, une analyse appropriée des tendances peut permettre de détecter plut tôt une défaillance.

9 Approche à base d'extraction de caractéristiques quantitatives

Dans un certain nombre de cas, il est impossible d'obtenir le modèle mathématique du procédé, ou d'une partie de celui-ci, reliant les signaux d'entrée à ceux de sortie. Dans la section précédente, on a vu quelques méthodes basées sur l'extraction de caractéristiques qualitatives. Dans cette section, on va évoquer les méthodes qui extraient les caractéristiques quantitatives. On peut analyser les signaux issus de l'instrumentation et proposer pour chacun de ces signaux un modèle. Ce modèle est en fait une référence du comportement de ce signal, qui peut être représentative de l'état normal de l'installation ou d'un défaut particulier. Ces caractéristiques peuvent être statistiques. Ceci fait l'objet du paragraphe 9.1. Une autre méthode de représentation pour un signal qui peut très bien être utilisée comme référence pour le diagnostic, est sa représentation par un spectre de fréquences, une signature graphique, ou encore l'extraction d'attributs (ou caractéristiques), à partir des données suivie de leur classification, qui permettent, pour des procédés dont on veut analyser plusieurs signaux liés entre eux, de discriminer divers modes de fonctionnement, dont le mode normal et certains modes de défaillance. Ces méthodes dites non statistiques sont évoquées au paragraphe 9.2.

9.1 Les méthodes statistiques

Les signaux sont caractérisés par certaines propriétés statistiques. L'étude de ces propriétés est le traitement du signal le plus classique. Une loi de densité de probabilité est obtenue en analysant les signaux pendant le fonctionnement normal. La détection consiste à déterminer si le signal obéit à cette loi ou s'en écarte significativement. Une difficulté de cette méthode est de disposer d'un historique suffisamment grand permettant l'estimation de grandeurs statistiques telles qu'une moyenne ou un écart type. Une autre difficulté est d'être représentatif de tous les modes de fonctionnement. Il est aussi difficile d'interpréter physiquement le résultat de la détection.

9.1.1 Tests d'Hypothèses

Dans cette approche, on analyse en général les signaux issus du procédé pendant son

fonctionnement normal et on en déduit une loi de densité de probabilité pour chaque signal analysé. La détection consiste à déterminer ensuite en temps réel si le signal obéit bien à cette loi, ou s'en écarte significativement. La difficulté provient justement de la nécessité de traitement en temps réel, quand on sait qu'en général l'estimation de grandeurs telles que moyenne, et écart type nécessitent une fenêtre temporelle longue. Notons que l'on peut très bien appliquer cette méthode à un fonctionnement anormal de l'installation et modéliser les signaux dans certains modes de défaillance.

	H_0 **vraie**	H_1 **vraie**
H_0 **retenue**	$P_{00} = 1-\alpha$	$P_{01} = \beta$
H_1 **retenue**	$P_{10} = \alpha$	$P_{11} = 1-\beta$

Tableau 1.1 — Décision statistique

Les seuils qui permettent de déterminer que le signal suit son modèle sont, en général, fondés sur des tests statistiques. On parle de tests d'hypothèses entre elles. L'hypothèse H_0 se réfère au cas du fonctionnement normal et on la compare à l'hypothèse H_1 de fonctionnement en défaut. On se trouve alors devant quatre situations de décision, décrites par le Tableau 1.1. Dans ce tableau, β représente la probabilité du manque à la détection et α la probabilité du taux de fausses alarmes.

Prenons un exemple simple pour illustrer ces grandeurs. Soit r la variable scalaire sur laquelle on veut faire la détection. On cherche le domaine D_0 et son complément D_1 tels que :

$$r \in D_0 \rightarrow H_0, r \in D_1 \rightarrow H_1 \tag{1.3}$$

Si l'on connaît la densité de probabilité $p(r)$ de r, $r_{\lim}$ est la limite entre deux domaines, telle que :

$$P(r < r_{\lim}) = \int_{-\infty}^{r_{\lim}} p(r)\,dr \tag{1.4}$$

$$\beta = \int_{-\infty}^{r_{\lim}} p(r/H_1)\,dr \tag{1.5}$$

$$\alpha = \int_{r_{\lim}}^{\infty} p(r/H_0)\,dr \tag{1.6}$$

La Figure 1.6 illustre les valeurs du manque à la détection β et du taux de fausses alarmes α dans le cadre d'une distribution gaussienne de la variable r, qui change de moyenne sous l'effet d'un défaut. Cette figure s'applique quand on attend un saut de moyenne positif. C'est un test unilatéral. Si l'on ne fait pas d'hypothèse sur le sens de la variation, on met en place un test bilatéral qui utilise deux limites. La valeur des limites est calculée à partir de la valeur souhaitée pour la probabilité de fausse alarme. De nombreux test ont été proposés, en particulier des tests de saut de moyenne, connu, ou inconnu a priori (on suppose que sous l'effet d'un défaut, la moyenne du signal change de valeur), et des tests de saut de variance (dans ce cas, c'est l'écart type du signal qui change sous l'effet d'un défaut) [19][10].

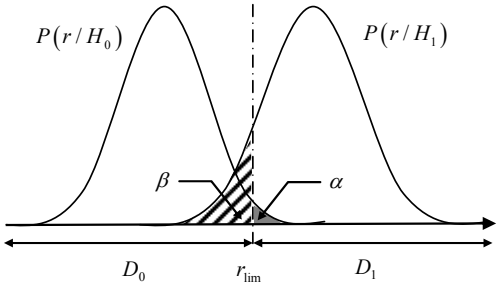


Figure 1.6 — Test statistique de deux hypothèses

9.1.2 Classification probabiliste

L'approche de diagnostic par classification (ou reconnaissance de formes) est basée essentiellement sur des observations et des expériences numériques et humaines accumulées sur le système. Elle ne fait pas l'hypothèse de l'existence d'un modèle mathématique décrivant le comportement du système. La reconnaissance de formes a pour but la reconnaissance d'une forme parmi différentes possibilités à partir d'observations bruitées. Il s'agit de trouver à quelle forme-type une forme observée ressemble le plus. Les méthodes de diagnostic basées sur la reconnaissance de formes constituent des outils importants pour la classification automatique des signatures associées aux différentes modes de fonctionnement normaux ou anormaux d'un système. Dans la suite, une description abrégée de cette approche est donnée, pour plus de détails, le lecteur pourrait consulter les références [9][11][13].

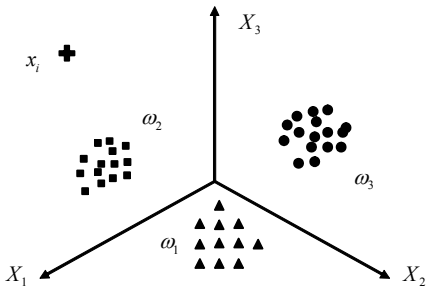


Figure 1.7 — Classification de données (ou reconnaissance de formes)

En reconnaissance de forme, il existe plusieurs approches probabilistes et non probabilistes (floue par exemple) basées sur des concepts différents. Dans le paragraphe suivant, on se concentrera sur l'approche probabiliste.

Considérons un problème de diagnostic par reconnaissance de formes où on a M classes connues $\omega_1, \omega_2, ..., \omega_m$. Pour tout vecteur forme x de $\mathbb{R}^d$, une classe doit être associée. On cherche donc une application D qui à x associe un entier i tel que:

- si $D(x) = 1$, x est associé à la classe ω_1;
- si $D(x) = 2$, x est associé à la classe ω_2; etc.;
- si $D(x) = M$, x est associé à la classe ω_M,
- si $D(x) = 0$, x est associé à la classe ω_0, classe de rejet en distance;

– si $D(x)=-1$, x est rejeté en ambiguïté.

Les probabilités a priori des classes $\omega_1,\omega_2,...,\omega_m$, $P(\omega_i)$, $i=1,...,M$ sont connues avec :

$$\sum_{i=1}^{M} P(\omega_i) \le 1 \tag{1.7}$$

La probabilité de la classe ω_0 peut être déduite par la relation suivante :

$$\sum_{i=0}^{M} P(\omega_i) = 1 \tag{1.8}$$

La loi de probabilité du vecteur x, quand on ne connaît pas sa classe, est donnée par la loi de mélange $P(x)$:

$$P(x) = \sum_{i=0}^{M} P(\omega_i) P(x/\omega_i) \tag{1.9}$$

où les lois conditionnelles $P(x/\omega_i)$ sont supposées connues. La loi spécifique à la classe ω_0, $P(x/\omega_0)$ est choisie par l'utilisateur.

Les décisions prises seront associées à un critère de performance, généralement la probabilité d'erreur. A chaque décision est associé un coût de décision, $C(i,\omega_i)$, coût de décider ω_i quand ω_j est la vraie classe de x. En général, on considèrera un coût de rejet d'ambiguïté dont la valeur sera identique quelle que soit la vraie classe ω_i de x, $C(-1,\omega_i)$.

Pour un vecteur x donné, le risque associé à la décision $D(x)$ s'écrit :

$$R(x) = \sum_{j=0}^{M} C\left(D(x),\omega_j\right) P\left(x/\omega_j\right) \tag{1.10}$$

Le risque moyen R est considéré comme un indicateur de performance. La probabilité a posteriori de la classe ω_i peut s'évaluer aisément à l'aide de la formule de Bayes :

$$P(\omega_i/x) = \frac{P(x/\omega_i)P(\omega_i)}{P(x)} \tag{1.11}$$

Le risque moyen R est obtenu en moyennant les risques associés aux décisions relatives à chaque vecteur x :

$$R = \int R(x) P(x)\, dx \tag{1.12}$$

Considérons maintenant un cas particulier lorsque aucune décision n'est prépondérante par rapport à une autre; les coûts dans ce cas sont fixés de la manière suivante:

$$\begin{aligned} C(i,\omega_i) &= \begin{cases} 1 & i \neq j \\ 0 & i = j \end{cases} \qquad i,j = 0,...M \\ C\left(-1,\omega_j\right) &= a \qquad j = 0,...,M \end{aligned} \tag{1.13}$$

Dans ce cas le risque associé à la décision $D(x)$ sera :

– si $D(x)=-1, R(x)=a$;

– si $D(x)=i,\ i=0,...,M,\ R(x)=\sum_{j=0}^{M} C(i,\omega_j) P(\omega_j/x) = \sum_{j=0,i\neq j}^{M} P(\omega_j/x)$, $R(x)=1-P(\omega_j/x)$

La règle de décision est alors facile à déduire :

– x est affecté à ω_i si :

$$P(\omega_i / x) = \max_{j=0}^{M}\left[P(\omega_j / x)\right] \text{ et } P(\omega_i / x) \geq 1 - a \tag{1.14}$$

x est rejeté en ambiguïté si :

$$P(\omega_i / x) = \max_{j=0}^{M}\left[P(\omega_j / x)\right] \text{ et } P(\omega_i / x) < 1 - a \tag{1.15}$$

Par conséquent, plusieurs cas peuvent donc se produire:

- Le vecteur x est associé à une des classes ω_i ,
- Le vecteur x est associé à la classe de rejet en distance ω_0 ,
- Le vecteur x est rejeté en ambiguïté entre deux ou plusieurs classes connues ou entre une ou plusieurs classes connues et la classe de rejet en distance.

9.1.3 Analyse en composante principale & Moindres Carrés Partiels

Une autre méthode est l'analyse en composante principale (*ACP*), (*PCA : Principal Components Analysis*), qui est une technique statistique multivariable. Les techniques statistiques multivariables sont des outils puissants capables de comprimer des données et de réduire leur dimension de sorte que l'information essentielle soit maintenue et plus facile à analyser que dans l'ensemble original des données. Ces techniques peuvent également manipuler le bruit et la corrélation pour extraire l'information efficacement. La fonction principale de ce type de techniques est, par le biais d'une procédure mathématique, de transformer un certain nombre de variables corrélées en un ensemble plus petit de variables non corrélées. L'ACP est essentiellement basée sur une décomposition orthogonale de la matrice de covariance des variables du processus le long des directions qui expliquent la variation maximale des données. Cette méthode recherche une projection des observations sur des axes orthogonaux. De ce fait, le premier axe contient la plus grande variation. Le deuxième axe contiendra la seconde plus grande variation orthogonale au premier.

Le but principal en utilisant l'ACP est de retrouver un ensemble de facteurs (composantes) qui ait une dimension inférieure à celle de l'ensemble original de données, et qui puisse décrire correctement les tendances principales. L'ACP est une procédure qui ne prend en compte que les variables du procédé. Par exemple en étude de qualité il est souhaitable d'inclure aussi l'ensemble de variables de qualité du produit, et d'utiliser de cette façon les variables du procédé pour prédire et détecter des changements dans les variables de qualité du produit. Pour cela, la méthode des Moindres Carrés Partiels (*MCP*) (*PLS : Partial Least Squares*), peut être utilisée. Cette méthode modélise le rapport entre deux blocs de données tout en les comprimant simultanément. Elle est employée pour extraire les variables latentes qui expliquent la variation des données du procédé.

Divers travaux menés par MacGeorge, [18][23] ont montré l'application des méthodes statistiques multivariables telles que les ACP et MCP, dans l'analyse et la commande de procédés, la détection et le diagnostic de défauts, aussi bien dans le cas de procédés continus que dans les cas des procédés continus par lots. Une limitation importante de la surveillance basée sur l'ACP est que le représentation obtenue est invariante dans le temps, tandis que la plupart des procédés réels évoluent au cours du temps. Par conséquent, la représentation issue de l'ACP nécessite également d'être mise à jour périodiquement. Un autre inconvénient est

qu'elle ne possède pas des propriétés de signature pour le diagnostic, ce qui rend la localisation des défauts difficile.

9.2 Les méthodes non statistiques

La reconnaissance de formes non probabiliste par extraction d'attributs à partir des données suivie de leur classification, telles la classification floue, ou les réseaux de neurones est une autre façon de traiter les données et de les présenter en tant que connaissance *a priori* non statistiques pour le diagnostic. On peut encore représenter les signaux par signature et contenu spectral, qui seront évoquées dans la suite.

9.2.1 Classification floue

En reconnaissance de formes par approche floue, le cadre de la modélisation de l'imprécision (incertitude) est différent que ce soit pour construire l'espace de représentation ou pour prendre la décision. Une fonction d'appartenance quantifie le degré d'appartenance $\mu_\omega(x) \in [0,1]$ de chaque vecteur x à la classe ω. Une classe floue $(x, \mu_\omega(x))$ est un sous ensemble flou de l'espace Ω. Les différentes classes floues sont construites par apprentissage. M fonctions d'appartenances aux M classes sont associées à chaque forme x. Ces fonctions d'appartenances caractérisent la similitude entre x et le prototype de chaque classe floue. L'espace de classification est donc défini d'une façon graduelle qui est très intéressante pour le diagnostic en ligne:

$$\sum_{i=1}^{M} \mu_\omega(x) = 1 \tag{1.16}$$

La mise en oeuvre d'une méthode de classification floue implique deux étapes: la construction des fonctions d'appartenance et la définition des règles de décision.

Fonctions d'appartenance : L'algorithme c-moyennes floues (FCM : Fuzzy C-Means) est l'un des premiers algorithmes proposés pour la construction de fonctions d'appartenance d'une façon automatique. Il s'agit dans cet algorithme d'une méthode non supervisée qui consiste à chercher les prototypes, à partir des formes $x_1, \ldots x_n$ en minimisant itérativement le critère suivant :

$$J = \sum_{k=1}^{n} \sum_{i=1}^{M} \mu_i^m(x_k)\, d_i^2(x_k, v_i) \tag{1.17}$$

avec les conditions suivantes :

$$\mu_i(x_k) \in [0,1] \quad \forall i,k \tag{1.18}$$

$$\sum_{k=1}^{M} \mu_i(x_k) = 1 \quad \forall i \tag{1.19}$$

$$\sum_{i=1}^{n} \mu_i(x_k) > 0 \quad \forall k \tag{1.20}$$

où m est un paramètre qui contrôle le caractère "flou" de la partition, v_i sont les centres de classes et d est la distance euclidienne classique. La solution s'obtient d'une manière itérative. La matrice d'appartenance $U(M,n)$ est initialisée avec des valeurs μ_i et à chaque étape les nouveaux centres de classe sont calculés :

$$v_i = \frac{\sum_{k=1}^{n} \mu_i^m(x_k)\, x_k}{\sum_{k=1}^{n} \mu_i^m(x_k)} \quad i = 1,...,M \tag{1.21}$$

Ensuite les éléments de la matrice U sont actualisés comme suit :

$$\mu_i(x_k) = \left(\sum_{j=1}^{M} \left(\frac{d_i}{d_j} \right)^{2/m-1} \right)^{-1} \tag{1.22}$$

L'algorithme de calcul des prototypes et des fonctions d'appartenance s'arrête quand la matrice U change peu entre deux itérations :

$$\left\| U^j - U^{j-1} \right\| < \varepsilon \tag{1.23}$$

Notons que plusieurs optimisations peuvent être faites en changeant l'initialisation de U (problème des minimums locaux) et que le choix de M et m n'est pas facile (généralement $m = 2$).

Décision floue : Avec l'approche floue, les règles de décision se définissent dans l'espace des fonctions d'appartenance. La règle la plus simple consiste à choisir la classe ayant le plus haut degré d'appartenance :

$$x \to \omega_i \quad si \quad \mu_i(x) = \max_{j=1,M} \mu_j(x) \tag{1.24}$$

Afin de prendre en compte les notions de rejet en distance et d'ambiguïté, il est possible d'utiliser un seuil θ_k pour chaque classe. Ce seuil est soit défini a priori, soit déterminé à partir de l'ensemble d'apprentissage :

$$\theta_k = \min_{x_i \in \omega_k} \mu_k(x_i) \tag{1.25}$$

Si l'on considère $A = \{\gamma_a, \gamma_1, ..., \gamma_M, \gamma_d\}$ l'ensemble des actions possibles incluant l'affectation à une classe $\omega_i(\gamma_i)$, à un rejet d'ambiguïté γ_a et à un rejet en distance γ_d, pour chaque vecteur d'entrée x, on obtient un ensemble $R(x)$ des résultats des actions obtenues :

$$R(x) = \{k \in \{1,...,M\} \mid \mu_k(x) > \theta_k\} \tag{1.26}$$

Une règle simple consiste à rejeter x en ambiguïté si x a un degré d'appartenance significatif à plusieurs classes, et à choisir l'option de rejet de distance si $R(x)$ est vide :

$$\gamma(x) = \begin{cases} \gamma_k & si \quad R(x) = \{k\} \;\; k \in \{1,...,M\} \\ \gamma_d & si \quad R(x) = \varnothing \\ \gamma_a & si \quad |R(x)| > 1 \end{cases} \tag{1.27}$$

L'inconvénient de la règle précédente est le fait que les deux rejets sont contrôlées par le même paramètre θ_k. La règle dite du rapport d'appartenance, proposée par Frélicot [13] se base sur le rapport :

$$Q = \frac{\mu_m(x)}{\mu_p(x)} \tag{1.28}$$

avec :

$$\mu_p(x) = \max_{k \in R(x)} \mu_k(x) \tag{1.29}$$

$$\mu_m(x) = \max_{k \in R(x) \setminus p} \mu_k(x) \tag{1.30}$$

On a évidemment $0 \le Q \le 1$ et son interprétation est la suivante :

Si Q est proche de 0, alors le degré d'appartenance μ_p est très grand par rapport aux autres degrés d'appartenance. L'action $\gamma_p(x)$ sera la plus favorable,

Si Q est proche de 1, alors au moins deux actions différentes $\gamma_p(x)$ et $\gamma_m(x)$ ont des degrés d'appartenance presque identiques. Cette situation correspond à un rejet d'ambiguïté.

Pour prendre la décision finale, ce rapport d'appartenance Q est alors comparé à un seuil S. Si $Q > S$, alors x est rejeté en ambiguïté, sinon, l'action $\gamma_p(x)$ est sélectionnée.

9.2.2 Réseaux de neurones

Les réseaux de neurones artificiels sont des outils particulièrement adaptés aux problèmes de classification. Pour ces raisons, de nombreuses études pilotes sont entreprises dans le secteur du diagnostic pour évaluer leurs apports avant leur mise en œuvre de façon opérationnelle.

Les réseaux de neurones sont des outils de l'intelligence artificielle, souvent utilisés pour des tâches de classification. Leur principe de fonctionnement est inspiré de celui des neurones biologiques. Le principal avantage des réseaux de neurones par rapport aux autres outils est leur capacité d'apprentissage et de généralisation de leurs connaissances à des entrées inconnues. Les réseaux de neurones peuvent être également implémentés dans des circuits électroniques offrant ainsi la possibilité d'un traitement temps réel. Le processus d'apprentissage est une phase très importante pour la réussite d'une telle opération. Plusieurs types de réseaux de neurones et plusieurs algorithmes d'apprentissage existent dans la littérature [9][11].

Une des qualités de ce type d'outil est son adéquation pour la mise au point de systèmes de surveillance modernes, capables de s'adapter à un système complexe avec reconfigurations multiples. Un réseau de neurones désigne un ensemble de neurones interconnectés. Le type de connexion varie d'une architecture à une autre. Les Perceptrons Multicouches (*PMC* et les Réseaux à Fonctions de base Radiales (*RFR*) sont les deux architectures les plus utilisées en surveillance industrielle. L'expert humain joue un rôle primordial dans ce genre d'outils.

Chaque mode de fonctionnement du système est caractérisé par un ensemble de données. Cette association (ensemble de données-modes de fonctionnement) sera apprise par le réseau de neurones durant la phase d'apprentissage. Après cette phase d'apprentissage, le réseau de neurones associera les classes représentant les modes de fonctionnement aux formes d'entrée caractérisées par les données du système.

9.2.3 Méthodes de signature

Le diagnostic de défauts par signatures fait partie des méthodes sans modèle. Il s'effectue à partir des symptômes externes observés sur les systèmes. Grâce à ces symptômes, le spécialiste déduit directement la nature de la défaillance sans se préoccuper des relations physiques qui lient les défaillances à leurs symptômes. Le diagnostic par signature nécessite une phase d'apprentissage qui permet de définir des signatures de référence qui correspondent à l'état normal du système à diagnostiquer et des signatures types pour des défaillances connues

sur le système. Une fois l'apprentissage fait, le diagnostic par signature s'effectue suivant la procédure [32][31] :

- Construction des signatures : la construction de signatures peut se réaliser avec les cinq sens humains, à l'aide de capteurs spécialisés ou après un traitement des données collectées sur le système. C'est une opération destinée à donner une information utile aux experts de diagnostic.
- Comparaison des signatures avec les signatures de référence : dans cette étape, les signatures actuelles sont comparées aux signatures de référence qui correspondent à l'état normal et à des états défaillants.
- Décision : une fois la comparaison faite, la phase de décision vient pour déclarer les signatures normales ou anormales. Cette étape de décision conduit à définir des seuils qui permettront d'accepter avec un risque admissible une non détection ou une fausse alarme.
- Analyse : Cette étape se déclenche si les signatures actuelles ne correspondent pas aux signatures de référence qui caractérisent l'état normal du système. Pour savoir la nature du défaut constaté il faut faire appel aux signatures types définies pendant la phase d'apprentissage. S'il n'y a pas des signatures types qui correspondent aux signatures actuelles on peut conclure que le fonctionnement est anormal mais inconnu.

Différentes catégories de signatures sont utilisées pour le diagnostic des systèmes industriels. Ces signatures contiennent les informations jugées pertinentes par les spécialistes pour la détection et la localisation des défaillances. La complexité des signatures dépend de la nature des systèmes et des matériels à diagnostiquer et varie avec l'anomalie recherchée.

9.2.4 Méthodes fréquentielles, temps-fréquence et temps-échelle

Le diagnostic et la surveillance des systèmes industriels nécessitent parfois l'interprétation des signaux mesurés tant dans le domaine temporel que fréquentiel. Dans ce second domaine plusieurs outils mathématiques sont souvent utilisés tels que la transformée de Fourier et la transformée en Ondelettes. L'analyse fréquentielle des signaux de vibration permet d'isoler des composantes spectrales : la répartition de l'énergie du signal est analysée en fonction de la fréquence. La variation de ce spectre permet de détecter des défauts voire d'identifier la défaillance.

Dans ce cadre, on effectue l'analyse spectrale du signal (répartition de son énergie en fonction de la fréquence). Celle-ci peut changer à l'occurrence d'un défaut. La variation du spectre est donc un moyen correct de détecter et même parfois de localiser un défaut. Par exemple, dans le cas des entraînements électriques, différents défauts (rupture de barres, court-circuit de spires, jeu dans les engrenages, etc.) donnent naissance à des raies spectrales bien particulières. Notons qu'il n'est pas toujours facile d'automatiser une analyse spectrale et que bien souvent c'est un expert qui décide que l'apparition de telle ou telle raie dans le spectre est significative. Pour déterminer le spectre du signal, on se sert de la transformée de Fourier rapide dont plusieurs versions sont disponibles dans les bibliothèques de traitement du signal. Pour garder un caractère temps réel à la détection, on utilise plus volontiers une analyse temps-fréquence. L'analyse est alors faite sur une fenêtre temporelle glissante.

Dans [24] des filtres d'ondelettes sont utilisés pour détecter le vieillissement rapide des isolants et le diagnostic des défauts paramétriques de la machine. Les filtres d'ondelettes réalisent une projection du signal sur des espaces d'analyse possédant des propriétés particulières. Cela conduit à une présentation de l'information qui permet de mettre en évidence des comportements du signal qui n'était pas forcément détectable "à l'œil nu". La difficulté de ces méthodes consiste à automatiser l'analyse du spectre ou des coefficients d'ondelettes.

10 Conclusion

L'objectif de ce premier chapitre était d'une part d'identifier une problématique générale au niveau du fonctionnement des applications industrielles, d'autre part de faire ressortir la nécessité de mettre en place des systèmes de supervision pour réduire les coûts économiques et répondre aux contraintes environnementales et de sécurité. Nous avons proposé un rapide état de l'art des différentes méthodes de diagnostic de défauts pour des applications industrielles.

Quatre catégories de méthodes ont été présentées : les méthodes à base de modèles quantitatifs, les méthodes à base de modèles qualitatifs, les méthodes à base d'historique de données qualitatif, et les méthodes à base d'historique de données quantitatif. Le choix d'une de ces méthodes dépend essentiellement des connaissances disponibles sur le procédé, sans oublier les considérations techniques et économiques. Néanmoins, nous avons constaté que ces méthodes ont des limitations et qu'un cadre de travail concernant la résolution des problèmes de façon collective en utilisant des raisonnements différents et parallèles s'avère être attractive pour relever les défis du diagnostic d'unités industrielles complexes.

Dans le cadre de cette problématique, notre travail porte sur le développement de méthodes pour la caractérisation du comportement du procédé. Ceci est réalisé au moyen des méthodes de traitement du signal de façon à exploiter les caractéristiques des données issues des capteurs et à prendre en compte les connaissances des ingénieurs et opérateurs pour faire le lien entre des caractéristiques particulières et les anomalies qu'elles caractérisent. Nous nous sommes intéressés aux méthodes de traitement de signal en considérant le fait que même si il n'y a pas une connaissance exhaustive de tous les modes de fonctionnement normaux ou anormaux du système, ces méthodes doivent être en mesure de détecter des nouveaux phénomènes et les prendre en compte de façon à les caractériser et à les détecter. Il apparaît donc clairement que le fait d'être capable de détecter la situation dans laquelle se trouve le procédé à chaque instant est indispensable pour apporter des informations supplémentaires sur les causes d'une défaillance et pour aider dans la prise de décision des actions à mener sur le système.

Ainsi, dans le chapitre suivant nous présenterons les méthodes de construction d'un modèle de comportement d'un procédé à travers l'extraction des caractéristiques de ses sorties (mesures). Ce modèle est obtenu à partir des données en utilisant des techniques statistiques, fréquentielles et temps-fréquence.

Chapitre 2

Diagnostic et Traitement du Signal

1 Introduction

Les signaux fournis par les différents capteurs contiennent des informations pertinentes disponibles pour le diagnostic. Le diagnostic associe, suite à l'analyse de ces signaux, des signatures correspondant au fonctionnement normal ou anormal du procédé ou du système à surveiller. La précision de l'analyse effectuée est essentielle pour une bonne décision de diagnostic. Il s'agit de détecter toute déviation par rapport à un comportement de référence, puis de diagnostiquer les causes de cette modification.

Le diagnostic des systèmes industriels nécessite parfois l'utilisation de méthodes issues du traitement du signal [44]. C'est en particulier le cas lorsqu'un modèle de comportement ou de connaissance de l'installation est, soit difficilement exploitable, soit irréaliste à obtenir compte tenu de la complexité de l'installation. Comme pour d'autres méthodes de diagnostic, le signal extrait de l'installation (avec ou sans prétraitement) contient de l'information sur l'état du procédé à surveiller. L'apparition d'un défaut se manifeste par l'occurrence d'une modification de la caractéristique extraite du signal analysé (par exemple, moyenne, écart type, spectre, etc.). La procédure de détection doit donc mettre en évidence ces modifications des caractéristiques analysées.

Ce chapitre présente quelques méthodes de traitement du signal employées couramment à des fins de diagnostic. La première partie introduit les notions de signal à temps continu, discret et échantillonné. Dans la deuxième partie sont rappelées quelques méthodes pour l'analyse temporelle des signaux, telles que l'étude des lois de probabilité et des tests de changement des caractéristiques statistiques. La troisième partie fait l'objet de l'analyse fréquentielle ou spectrale : nous rappelons la transformée de Fourier continue, discrète, et l'algorithme de calcul de la transformée de Fourier rapide. La section 4 s'intéressera à l'analyse temps-fréquence et la section 5 à l'analyse temps-échelle.

2 Analyse temporelle

Dans cette section, nous rappelons les représentations naturelles d'un signal (à temps continu, à temps discret), sans insister sur tous les aspects mathématiques. Nous donnerons aussi une brève présentation du principe des tests statistiques basés sur le rapport de vraisemblance, pour ensuite nous concentrer sur le test progressif de Wald, en donnant un exemple pour le test de saut de variance connu. Certains de ces tests ont été implantés dans la bibliothèque de routine de détection de défauts du projet européen MAGIC.

2.1 Transformée continue de Dirac

Le traitement du signal consiste à manipuler des enregistrements de données pour en extraire des informations jugées pertinentes. Le modèle temporel d'un signal est constitué des valeurs affectées à chacun des instants que le dispositif physique fournit. Cette représentation correspond à une décomposition sur la base continue des distributions de Dirac :

$$\begin{aligned} & L^2(\mathbb{R}) \to L^2(\mathbb{R}) \\ & x(t) \quad \to DT_x(t) = \int_{\mathbb{R}} x(u)\delta(t-u)\,du = \langle x, \delta_u \rangle \text{ avec } \delta_u = \delta(t-u) \end{aligned} \tag{2.1}$$

Elle donne une information précise en temps : la valeur $x(t)$ indique l'intensité du signal à l'instant t.

2.2 Signaux discrets et échantillonnés

De nos jours, les systèmes à base de micro-controlleurs ou de microprocesseurs dominent le domaine du traitement du signal. Les manipulations de signaux se réduisent à celles de tableaux de données constitués des valeurs extraites par échantillonnage des signaux analogiques. Une classe de signaux discrets correspond aux signaux échantillonnés à partir des signaux à temps continu, avec la période d'échantillonnage $T_e = \Delta t$:

$$\begin{aligned} & L^2(\mathbb{R}) \to L^2(\mathbb{Z}) \\ & x(t) \quad \to x[k] = \sum_{k=-\infty}^{+\infty} x[k\Delta t]\,\delta[t-k\Delta t] \end{aligned} \tag{2.2}$$

La distinction entre les signaux discrets et les signaux échantillonnés réside dans la possibilité de reconstruire le signal continu à partir des valeurs $x[k]$. Dans le cas d'un signal échantillonné, cette reconstruction est possible sous certaines hypothèses.

2.3 Tests statistiques

Afin de détecter la modification de la moyenne ou de l'écart type d'un signal ou d'un résidu, d'estimer l'instant d'occurrence de ce changement et d'identifier la nature et l'amplitude du changement dans le domaine temporel, les tests de saut de moyenne et de variance ont plus particulièrement retenu notre attention. L'ensemble de ces méthodes est basé sur la théorie de la décision et plus particulièrement sur celle des tests d'hypothèses.

En tout premier lieu, on peut s'interroger, après une acquisition complète des données, sur le caractère aléatoire. L'hypothèse la plus communément utilisée à propos des bruits de mesures est de considérer qu'ils sont distribués selon une loi normale. Cette hypothèse est justifiée par le

théorème centrale limite qui démontre que la distribution d'une variable aléatoire résultant de l'addition de multiples variables aléatoires de distributions quelconques, indépendantes et de même ordre de grandeur, tend vers la loi normale.

On doit s'interroger, en second lieu sur la distribution de probabilité des valeurs et en particulier, tenter de vérifier sa normalité. Si la séquence analysée est réputée normale, on peut ensuite comparer les paramètres caractéristiques de la distribution (le plus souvent la moyenne et l'écart type) à des valeurs données. Si l'on dispose de connaissances plus précises relatives aux distributions de probabilité en fonctionnement normal et en fonctionnement anormal, on peut construire des tests plus performants.

En pratique, il est intéressant de prendre des décisions en temps réel, au fur et à mesure de l'acquisition des observations. Les tests séquentiels d'hypothèses comme le test de Wald ou le test de Page-Hinkley permettent de prendre de telles décisions, tout en garantissant des risques d'erreur inférieurs à certains seuils.

Un test statistique est une règle qui permet de prendre une décision à partir de résultats expérimentaux [3]. Son exécution suppose les cinq opérations principales suivantes :

- Définition de l'hypothèse H_0 à contrôler, dite hypothèse nulle (l'hypothèse alternative H_1 est généralement définie comme la négation de H_0) ;
- Choix d'une fonction discriminante Y, fonction des résultats expérimentaux dont la distribution de probabilité sous l'hypothèse H_0, $p(Y \mid H_0)$ est connue (analytiquement ou par estimation) ;
- Choix d'un seuil de confiance α et de la taille de l'échantillon N ;
- Définition de la région de rejet de l'hypothèse H_0 ;
- Evaluation de la fonction discriminante sur la base des échantillons observés.

Si sa valeur est située dans la région de rejet, on rejette l'hypothèse H_0 ; si elle est à l'extérieur de cette région, l'hypothèse H_0 est acceptée relativement au seuil de confiance α choisi.

On considère maintenant que l'on connaît les densités de probabilités conditionnelles des observations $y_k = y[k]$ relativement aux deux hypothèses H_0 et H_1. L'objectif du test consiste toujours à décider, à partir des observations effectuées sur le système, selon quelle hypothèse le système a fonctionné. Notons dorénavant Y le vecteur de dimension N des observations y_k.

On suppose l'espace d'observation divisé en deux sous-espaces disjoints Y_0, $\{Y_0 : y < \gamma\}$, et Y_1, $\{Y_1 : y > \gamma\}$. Si l'observation courante appartient à Y_0, on choisit l'hypothèse H_0, sinon l'hypothèse H_1 est retenue. En réalité, quatre décisions D_{ij} peuvent être prises pouvant ainsi se résumer par :

- D_{00} : l'hypothèse H_0 est vraie et H_0 a été effectivement choisie ;
- D_{01} : l'hypothèse H_0 est vraie mais H_1 a été choisie ;
- D_{11} : l'hypothèse H_1 est vraie et H_1 a été effectivement choisie ;
- D_{10} : l'hypothèse H_1 est vraie mais H_0 a été choisie.

Deux types d'erreurs peuvent être commises ; Il y a donc deux situations de mauvaise décision (D_{01} correspond à une erreur dite de type I et D_{10} à une erreur de type II). La probabilité de commettre une erreur de type I est égale à α ; celle de commettre une erreur de type II est égale à β. De manière idéale, il faudrait spécifier les valeurs de α et β. Les deux probabilités α et β étant liées, pour un nombre fixe d'observations, une diminution de α entraîne une augmentation de β.

Si l'on considère que H_1 est associée à la présence d'un défaut, la situation D_{11} est relative à une bonne détection (avec une probabilité $P_D = 1-\beta$), la situation D_{01} à une fausse alarme (avec une probabilité $P_F = \alpha$) et D_{10} à un défaut non détecté (avec une probabilité $P_{ND} = \beta$). Les densités de probabilités conditionnelles des observations y_k relativement aux deux hypothèses H_0 et H_1 étant connues, ces trois situations sont caractérisées par les probabilités (cf. Figure 1.6) :

$$\begin{aligned} P_D &= 1-\beta = p\left(H_1 \text{ choisie}|\, H_1 \text{ vraie}\right) = \int_{\gamma}^{\infty} p\left(Y \mid H_1\right) dY \\ P_F &= \alpha = p\left(H_1 \text{ choisie}|\, H_0 \text{ vraie}\right) = \int_{\gamma}^{\infty} p\left(Y \mid H_0\right) dY \\ P_{ND} &= \beta = p\left(H_0 \text{ choisie}|\, H_1 \text{ vraie}\right) = \int_{-\infty}^{\gamma} p\left(Y \mid H_1\right) dY \end{aligned} \tag{2.3}$$

Le problème posé consiste donc à rechercher le minimum de P_{ND} sous la contrainte $P_F - \alpha = 0$. Cela revient à déterminer l'extremum de [19] :

$$J = P_{ND} + \lambda\left(P_F - \alpha\right) \tag{2.4}$$

La règle de décision s'exprime alors de la manière suivante :

$$\Lambda\left(Y\right) = \frac{p\left(Y \mid H_1\right)}{p\left(Y \mid H_0\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{2.5}$$

Le rapport ainsi formé est appelé rapport de vraisemblance des densités de probabilité conditionnelles et λ le seuil de détection. La notation employée signifie que si ce rapport est inférieur à λ, on accepte l'hypothèse H_0 ; dans le cas contraire, on accepte H_1.

2.4 Méthodes statistiques en diagnostic

Dans le cadre du projet MAGIC [56], quelques tests de saut de moyenne et d'écart type ont été implantés :

- Saut de moyenne connue (Neyman Pearson, Page-Hinkley et Wald),
- Saut de moyenne inconnue (Neyman Pearson et Wald),
- Saut de variance connue (Wald).

Nous présentons maintenant le test progressif de Wald.

2.4.1 Test progressif de Wald

Les tests progressifs ou tests séquentiels permettent de prendre une décision dès que le nombre d'observations est suffisant pour garantir l'infériorité des probabilités d'erreurs à des valeurs imposées [19]. Pour le test de Wald, ou test séquentiel du rapport de vraisemblance,

l'objectif est toujours de décider entre deux hypothèses concernant un paramètre θ :

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned} \tag{2.6}$$

Mais, dans ce type de test, on calcule au fur et à mesure de l'obtention des données, une fonction de ces dernières qui est confrontée à deux limites correspondant au refus et à l'acceptation d'une hypothèse. Après chaque nouvelle acquisition, une des trois décisions suivantes est prise :

- accepter H_1,
- accepter H_0.
- n'accepter ni H_1 ni H_0

L'acceptation de l'une ou l'autre des hypothèses ou la non prise de décision s'effectuent en fonction de la valeur du rapport de vraisemblance défini précédemment. Soit $y_1, y_2, ..., y_k$ la séquence d'observations disponible à l'instant k. Le rapport de vraisemblance s'écrit :

$$\Lambda(y_1, y_2, ..., y_k) = \frac{p(y_1, y_2, ..., y_k \mid H_1)}{p(y_1, y_2, ..., y_k \mid H_0)} \tag{2.7}$$

Fixons alors la valeur des risques d'erreur α et β de première et seconde espèces. Pour retenir l'hypothèse H_1, il faut que ce rapport de vraisemblance soit tel que le risque d'erreur de première espèce soit inférieur ou égal à α si l'on retient H_1 alors que H_0 est vraie.

Si γ_1 est le seuil de décision correspondant, on doit avoir :

$$\Lambda(y_1, y_2, ..., y_k) \geq \gamma_1 \tag{2.8}$$

Le seuil γ_1 peut être déterminé en considérant le cas limite où l'on a égalité des deux membres, c'est-à-dire :

$$\int_{\gamma_1}^{\infty} p(y_1, y_2, ..., y_k \mid H_1)\, dy_1 dy_2, ...dy_k = \gamma_1 \int_{\gamma_1}^{\infty} p(y_1, y_2, ..., y_k \mid H_0)\, dy_1 dy_2, ...dy_k \tag{2.9}$$

ou encore:

$$1 - \beta = \alpha\gamma_1 \tag{2.10}$$

Le seuil de décision s'exprime donc en fonction des risques d'erreur sous la forme :

$$\gamma_1 = (1-\beta)/\alpha \tag{2.11}$$

De manière duale, l'hypothèse H_0 sera retenue, si le rapport de vraisemblance est tel que le risque d'erreur de seconde espèce est inférieur ou égal à β si l'on retient H_0 alors que H_1 est vraie. Une démarche tout à fait analogue à la précédente permet de déterminer un second seuil de décision γ_0 qui s'exprime également en fonction des risques d'erreur :

$$\gamma_0 = \beta/(1-\alpha) \tag{2.12}$$

La règle de décision est alors la suivante :

- si $\Lambda(y_1, y_2, ..., y_k) \leq \beta/(1-\alpha)$, on accepte H_0 ;
- si $\Lambda(y_1, y_2, ..., y_k) \geq (1-\beta)/\alpha$, on accepte H_1 ;
- si $\beta/(1-\alpha) \leq \Lambda(y_1, y_2, ..., y_k) \leq (1-\beta)/\alpha$, on poursuit la collecte des données jusqu'à l'obtention du premier rapport dont la valeur est extérieure aux limites.

Grâce à sa formulation séquentielle, le test de Wald est très efficace pour détecter des ruptures dans le comportement d'un signal au cours du temps. Cette propriété est particulièrement intéressante dans le contexte de la détection de défauts.

2.4.2 Application à la détection de saut de variance connu

Dans ce cas on cherche à décider entre les deux hypothèses :

$$\begin{aligned} H_0 &: y_k = m_0 + b_k, \quad b_k \text{ est de variance } \sigma_0^2 \text{ connue} \\ H_1 &: y_k = m_1 + b_k, \quad b_k \text{ est de variance } \sigma_1^2 \text{ connue} \end{aligned} \tag{2.13}$$

Les bruits b_k qui affectent les mesures sont supposés suivre une loi normale centrée. Le logarithme du rapport de vraisemblance, pour N mesures, s'écrit :

$$\ln\left(\Lambda\left(y_1, y_2, ..., y_k\right)\right) = N \log\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2\sigma_1^2} \sum_{i=1}^{N} \left(y_i - m_0\right)^2 \tag{2.14}$$

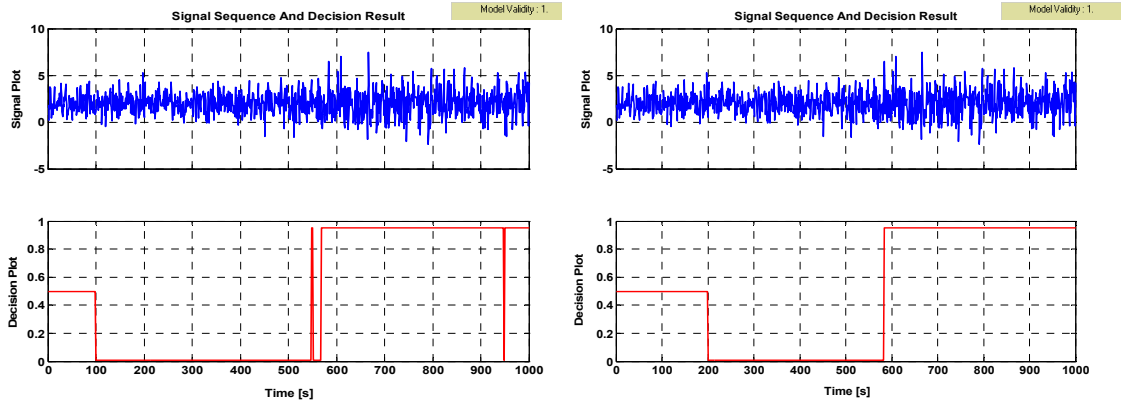
Le test de Wald est donc décrit par la condition d'acceptation de l'hypothèse H_0 :

$$\sum_{i=1}^{N} \left(y_i - m_0\right)^2 \underset{H_0}{\lessgtr} \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2} \left(\ln\left(\frac{\beta}{1-\alpha}\right) + N \ln\left(\frac{\sigma_1}{\sigma_0}\right) \right) \tag{2.15}$$

De façon analogue, la condition d'acceptation de l'hypothèse H_1 :

$$\sum_{i=1}^{N} \left(y_i - m_0\right)^2 \underset{H_1}{\gtrless} \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2} \left(\ln\left(\frac{1-\beta}{\alpha}\right) + N \ln\left(\frac{\sigma_1}{\sigma_0}\right) \right) \tag{2.16}$$

Exemple : Soit un signal aléatoire contenant un saut de variance. Les caractéristiques du signal sont : sa moyenne de référence est $m_0 = 2$, son écart type est $\sigma_0 = 1$, le nouvel écart type est $\sigma_1 = 1.5$, sachant que le signal de référence a un Rapport Signal sur Bruit de $RSB = 6\ dB$. Le signal contient 1000 échantillons, le saut se produit à l'échantillon 500. Les paramètres du test sont un taux de fausse alarme $\alpha = 0.01$, un taux de manque à la détection $\beta = 0.05$, et une fenêtre de longueur $N = 100$, $N = 200$ respectivement. Le signal et les résultats du test sont présentés sur la Figure 2.1. Le test détecte le saut avec un retard qui dépend de la longueur de la fenêtre N. (retard = 50 points pour $N = 100$, retard = 80 points pour $N = 200$).



a) $\alpha = 0.01,\ \beta = 0.05,\ N = 100$ b) $\alpha = 0.01,\ \beta = 0.05,\ N = 200$

Figure 2.1 — Résultats du test de détection de saut de variance connu avec le test de Wald

3 Analyse fréquentielle

Le diagnostic et la surveillance des systèmes industriels nécessitent parfois l'interprétation des signaux mesurés tant dans le domaine temporel que fréquentiel. Dans ce second domaine, c'est l'aspect périodique des signaux qui est privilégié. Un des outils mathématiques adapté à l'analyse des signaux périodiques est la Transformée de Fourier.

3.1 Transformée de Fourier

Une alternative à la représentation temporelle du signal consiste à représenter son contenu fréquentiel. Celui-ci est obtenu en calculant sa transformée de Fourier (FT : *Fourier Transform*), définie pour un signal $x(t)$ d'énergie finie par [48] :

$$\begin{aligned} & L^2(\mathbb{R}) \to L^2(\mathbb{R}) \\ & x(t) \quad \to FT_x(\nu) = \int_{\mathbb{R}} x(t) e^{-i2\pi\nu t} dt = \left\langle x, e^{i2\pi\nu t} \right\rangle \end{aligned} \tag{2.17}$$

Cette transformée étant inversible on a :

$$x(t) = \int_{\mathbb{R}} FT_x(\nu) e^{i2\pi\nu t} d\nu \tag{2.18}$$

La FT mesure donc un indice de similarité entre le signal et une collection d'ondes (exponentielles complexes) de longueur infinie, $FT_x(\nu)$ représentant la contribution de l'onde de fréquence ν au signal $x(t)$. La FT est en fait une transformée en ondes du signal dans laquelle toute information temporelle disparaît [48].

3.2 Transformée de Fourier à temps discret

Par définition, la transformée de Fourier à temps discret (DTFT : *Discrete Time Fourier Transform*) d'une suite $\{x[k]\}$ est la fonction de ν, périodique de période 1 :

$$X(\nu) = \sum_{k=-\infty}^{+\infty} x[k] e^{-i2\pi k\nu} \tag{2.19}$$

C'est une fonction à fréquence continue, c'est-à-dire $\nu \in \mathbb{R}$ [43]. Il est d'usage de la représenter sur les intervalles $[-1/2, 1/2]$ ou $[0,1]$, du fait de sa périodicité.
On déduit la formule inverse :

$$x[k] = \int_{-1/2}^{1/2} X(\nu) e^{i2\pi k\nu} d\nu \tag{2.20}$$

La fonction $|X(\nu)|^2$ est désignée par le terme de spectre. Dans la littérature ce terme est aussi associé à la fonction $|X(\nu)|$ (spectre d'amplitude). Les deux expressions deviennent équivalentes si l'on utilise l'échelle des décibels en posant $S_{dB}(\nu) = 20\log_{10}|X(\nu)|$. Comme le premier membre de (2.21) est par définition l'énergie du signal, $|X(\nu)|^2$ s'interprète comme la distribution de l'énergie le long de l'axe des fréquences :

$$\sum_{k=-\infty}^{+\infty} |x[k]|^2 = \int_{-1/2}^{1/2} |X(\nu)|^2 d\nu \tag{2.21}$$

3.3 Transformée de Fourier discrète

Le calcul sur ordinateur de la DTFT à partir de la valeur des échantillons de $x[k]$ impose une charge de calcul infinie puisque la série comporte un nombre infini de termes et que la fréquence ν varie continûment sur l'intervalle $[0,1]$. C'est la raison pour laquelle la notion de transformée de Fourier discrète (DFT : *Discrete Fourier Transform*) a été introduite [43]. Son calcul est limité à un nombre fini de valeurs de k et à un nombre fini de valeurs de ν. Son intérêt pratique est très largement dû à la découverte d'une méthode de calcul rapide connue sous le nom de transformée de Fourier rapide (FFT : *Fast Fourier Transform*).

Considérons la suite finie $\{x[k]\}_{k=0:N-1}$. Le plus souvent, cette suite provient de l'échantillonnage d'un signal à temps continu. Partant de la définition (2.17), la DTFT s'écrit :

$$X(\nu)=\sum_{k=0}^{N-1}x[k]e^{-i2\pi k\nu} \tag{2.22}$$

où $\nu\in[0,1]$ est la fréquence normalisée. En se limitant à un nombre fini de L de valeurs de la fréquence, à savoir $\nu=n/L$ avec $n=0:L-1$, on aboutit à la notion de transformée de Fourier discrète. L et N ne jouent pas le même rôle. L est le nombre de points de calcul de la DTFT et N est le nombre de points de la suite temporelle. Comme nous le verrons par la suite, L influe sur la précision du tracé du spectre, tandis que N est lié à la résolution. Toutefois prendre $L=N$ ne restreint en rien la généralité de la notion introduite. Cela aboutit à la définition de la transformée de Fourier discrète pour la suite $\{X[n]\}$:

$$X[n]=\sum_{k=0}^{N-1}x[k]\Phi_N^{nk},\ n=0:N-1 \text{ avec } \Phi_N^{nk}=e^{-i2\pi nk/N} \tag{2.23}$$

La formule inverse est :

$$x[k]=\frac{1}{N}\sum_{n=0}^{N-1}X[n]e^{i2\pi nk/N}=\frac{1}{N}\sum_{n=0}^{N-1}X[n]\Phi_N^{-nk},\ n=0:N-1 \tag{2.24}$$

3.4 Transformée de Fourier rapide

La transformée de Fourier discrète (DFT) d'un jeu de $N=2^p$ échantillons $x[k],k=0,...,N-1$, est un ensemble de N coefficients :

$$X[n]=\sum_{k=0}^{N-1}x[k]e^{-i2\pi nk/N},\ n=0:N-1 \tag{2.25}$$

La DFT est donc (à un facteur de normalisation près) une transformation orthogonale associée à une matrice $N\times N$. Son application à un vecteur nécessite en théorie, $\mathcal{O}(N^2)$ additions et multiplications. Cooley et Tukey [46] ont découvert que ce calcul peut être fait de manière récursive avec seulement $\mathcal{O}(N\log_2 N)$ additions et multiplications. Le principe est le suivant. Notons $x_{pair}[k]=x[2k],\ x_{impair}[k]=x[2k+1],\ k=0:N/2-1$. Pour $n=0:N/2-1$ on a :

$$\begin{aligned}X[n]&=X_{pair}[n]+e^{-i2\pi n/N}X_{impair}[n],\\ X[n+N/2]&=X_{pair}[n]+e^{-i2\pi n/N}X_{impair}[n].\end{aligned} \tag{2.26}$$

Le célèbre algorithme de transformée de Fourier rapide (*FFT : Fast Fourier Transform*) de calcul de DFT est fondé sur l'utilisation récursive de ces formules. Lorsque N n'est pas une

puissance de deux, des calculs analogues [43] en utilisant une décomposition en facteurs premiers conduisent au même ordre de grandeur $\mathcal{O}(N \log_2 N)$ du nombre de multiplications et additions. Beaucoup de travaux ont contribué à l'optimisation de cet algorithme, parmi lesquels on notera le programme FFTW (*Fastest Fourier Transform in the West*) disponible sur le Net [47].

3.5 Précision et résolution spectrales

Le fait de ne disposer que d'une séquence de durée finie fait apparaître des ondulations dans la DTFT du signal. Comme la DFT correspond à l'échantillonnage de la DTFT aux points de fréquence n/N, elle est en général constituée de valeurs différentes de zéro, sauf si ν_0 est exactement un multiple de $1/N$ [43]. Pour des fréquences qui ne sont pas égales à un multiple de $1/N$, où N est le nombre de points de FFT calculés, une sinusoïde pure apparaît sous forme de plusieurs valeurs non nulles, dont la plus importante en module est proche de la vraie fréquence.

Si N désigne le nombre de points de calcul de la DFT, la précision en fréquence est égale à $1/N$. Pour des signaux échantillonnés à la fréquence F_e (en [Hz]) cela donne une précision de F_e/N [Hz]. La précision ne doit pas être confondue avec la résolution qui est le pouvoir de distinguer (on dit aussi résoudre ou séparer) deux fréquences voisines dans un signal [58]. Il est commode de définir la résolution par l'écart minimum en fréquence qu'il faut mettre entre deux sinusoïdes d'amplitudes différentes pour observer sur le spectre de leur somme un creux de plus de $3dB$ entre les deux maxima.

Le fait d'avoir limité le nombre de valeurs traitées à N conduit à l'apparition de lobes secondaires dans le spectre d'une sinusoïde. Le lobe principal a une largeur égale à $2/N$. Il s'ensuit que, si $x[k]$ contient deux sinusoïdes dont les fréquences sont séparées de moins de $1/N$, les lobes principaux de chacune d'elles seront si proches qu'il sera difficile de les distinguer avec certitude. Et ceci est d'autant plus vrai que les deux amplitudes sont très différentes.

3.6 Effet du fenêtrage sur la résolution

La fait de limiter à N le nombre d'échantillons d'un signal peut être vu comme la multiplication terme à terme de la totalité du signal par la suite $\phi_N[k] = 1[k]_{k=0:N-1}$ à qui on donne le nom de fenêtre rectangulaire. Cette multiplication est équivalente, d'un point de vue spectral, à convoluer la DTFT de $x[k]$ avec la DTFT $\Phi_N(\nu)$ de la suite $\phi_N[k]$. Cette opération de convolution a pour effet d'introduire des ondulations dans le spectre. On appelle fenêtre la suite des coefficients utilisés pour pondérer le signal avant d'en calculer le spectre. Une description succincte de différentes fenêtres que l'on rencontre souvent est donnée dans la section suivante. En règle générale, la résolution en fréquence est d'autant meilleure que le lobe principal (central) est étroit et que les lobes secondaires (latéraux) sont bas. Malheureusement la réduction en hauteur des lobes secondaires s'accompagne toujours de l'élargissement du lobe principal. Il faut donc accepter un compromis entre ces deux effets. Le choix de la fenêtre de pondération est un problème délicat. On doit faire un compromis entre la largeur du lobe principal et l'importance des lobes secondaires, ce problème sera abordé dans la section 4.4.

4 Analyse temps-fréquence

Dans le cas d'un signal non stationnaire, le contenu fréquentiel du signal est susceptible de varier au cours du temps. On décompose alors le signal en une combinaison linéaire de fonctions oscillantes ayant un support temporel limité. Cette décomposition fournit une représentation bidimensionnelle du signal, fonction du temps et d'un paramètre correspondant directement à la fréquence ou pouvant être interprété en terme de fréquence. Les fonctions d'analyse utilisées peuvent ainsi être vues comme une collection d'atomes temps-fréquence pour lesquels la transformée mesure la contribution de chaque atome à la constitution du signal.

4.1 La Transformée de Fourier à Court Terme

Les représentations temps-fréquence ne se réduisent pas bien sûr à la transformée de Fourier à Court Terme, mais celle-ci présente l'avantage de la simplicité pour la présentation de cette approche, même s'il peut sembler réducteur de s'y limiter [45]. La transformée de Fourier à Court Terme *(STFT : Short Time Fourier Transform)* réalise une décomposition de Fourier sur un horizon temporel limité, imposé par une fenêtre d'analyse ϕ centrée sur l'instant t. Elle est définie par :

$$\begin{aligned} & L^2(\mathbb{R}) \to L^2(\mathbb{R}^2) \\ & x(t) \quad \to STFT_x(t,\nu) = \int_{\mathbb{R}} x(u)\phi^*(u-t)e^{-i2\pi\nu u}du \end{aligned} \tag{2.27}$$

Ceci revient à comparer le signal à une collection de fonctions dont la forme dépend d'une fenêtre d'analyse qui subit une translation temporelle t et fréquentielle ν :

$$\begin{aligned} STFT_x(t,\nu) &= \int_{\mathbb{R}} x(u)\phi_{t,\nu}{}^*(u)du = \langle x, \phi_{t,\nu} \rangle \\ \phi_{t,\nu}(u) &= \phi(u-t)e^{i2\pi\nu u} \end{aligned} \tag{2.28}$$

Ainsi, $STFT_x(t,\nu)$ mesure la contribution de la fréquence ν au signal $x(t)$ à l'instant t. Les fonctions de base $\phi_{t,\nu}$ dépendent de la fenêtre ϕ (Gaussienne, Hamming, etc.).

4.2 Interprétation en terme de filtrage

Les définitions précédentes basées sur des produits scalaires peuvent être reformulées en utilisant des produits de convolution, qui permettent d'interpréter les lignes du plan temps-fréquence comme autant de signaux résultant du filtrage du signal original [45]. Ainsi, à partir de la définition (2.27), la transformée de Fourier à court terme peut s'exprimer par :

$$STFT_x(t,\nu) = e^{-i2\pi\nu t}\int_{\mathbb{R}} x(u)\overline{\phi}^*(t-u)e^{i2\pi\nu(t-u)}du \tag{2.29}$$

où $\overline{\phi}(u) = \phi(-u)$. Si l'on considère $x_{\nu_0}(t)$ le signal à temps continu obtenu en modulant par $e^{i2\pi\nu_0 t}$ une coupe horizontale du plan temps-fréquence à la fréquence ν_0, ce signal résulte du filtrage du signal original $x(t)$. La réponse fréquentielle du filtre pour une fenêtre ϕ réelle est $|\Phi(\nu-\nu_0)|$ où $\Phi(\nu)$ est la transformée de Fourier de $\phi(t)$. Ainsi tous les filtres considérés sont de type passe-bande et ont la même largeur de bande, centrés sur la fréquence analysée.

4.3 Résolution dans le plan temps-fréquence

La transformée STFT utilise des fonctions dont l'enveloppe constante et le nombre

d'oscillations augmente avec la fréquence analysée par action de l'opérateur de modulation temporelle (translation fréquentielle) [52]. Ceci a pour effet de conserver une largeur de bande d'analyse fréquentielle constante quelle que soit la fréquence des oscillations de la fonction analysante (cf. Figure 2.2).

Le coefficient $STFT_x(t,\nu)$ ne décrit pas le contenu du signal strictement à l'instant t et à la fréquence ν, dans la mesure où pour évaluer le contenu spectral à cette date, il faut nécessairement observer le signal sur un horizon non nul autour de cette date. L'atome d'analyse $\phi_{t,\nu}$ est ainsi caractérisé par ses extensions conjointes temporelle Δt et fréquentielle $\Delta\nu$ [42], celles-ci étant définies pour une fonction ϕ d'énergie unité et centrée à l'origine, par :

$$\begin{cases} \Delta t^2 = \int_{\mathbb{R}} t^2 \left|\phi(t)\right|^2 dt \\ \Delta\nu^2 = \int_{\mathbb{R}} \nu^2 \left|\Phi(\nu)\right|^2 d\nu \end{cases} \tag{2.30}$$

L'atome d'analyse mélange donc l'information contenue dans le signal entre les dates $t \pm \Delta t/2$ et dans la bande de fréquence $\nu \pm \Delta\nu/2$. Le principe d'incertitude d'Heisenberg-Gabor traduit le fait que la localisation idéale temps-fréquence $(\Delta t = 0,\ \Delta\nu = 0)$ est impossible et que la résolution conjointe temps-fréquence est bornée inférieurement (l'égalité étant obtenue avec une fonction ϕ gaussienne) :

$$\Delta t \Delta\nu \geq \frac{1}{4\pi} \tag{2.31}$$

Les résolutions temporelle $\Delta t_{\phi_{\nu,t}}$ et fréquentielle $\Delta\nu_{\phi_{\nu,t}}$ ne sont pas modifiées par l'action des opérateurs de translation en t et en ν :

$$\Delta t_{\phi_{\nu,t}} = \Delta t_{\phi}\,;\Delta\nu_{\phi_{\nu,t}} = \Delta\nu_{\phi} \tag{2.32}$$

Il est alors naturel de représenter, dans le plan temps-fréquence des pavés élémentaires $\Delta t\Delta\nu$ de taille et de forme constantes quelle que soit la fréquence analysée (cf. Figure 2.2).

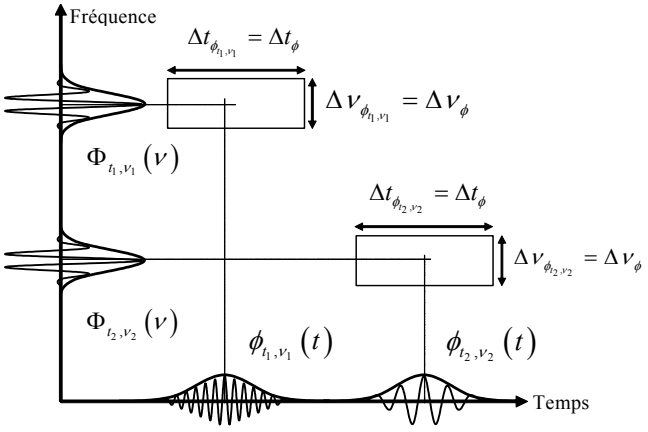


Figure 2.2 — Pavage temps-fréquence associé à la STFT

4.4 Rôle de la fenêtre d'analyse

Commençons par considérer deux exemples académiques :

Exemple 1 : Si $x(t)=\delta(t-t_0)$ alors $STFT_x(t,\nu)=\phi^*(t_0-t)e^{-i2\pi\nu t_0}$

Exemple 2 : Si $x(t)=e^{i2\pi\nu_0 t}$ alors $STFT_x(t,\nu)=\Phi^*(-(\nu-\nu_0))e^{-i2\pi(\nu-\nu_0)t}$

Le premier exemple nous montre que la transformée de Fourier à court terme d'une fonction parfaitement concentrée temporellement autour de l'instant t_0 est dispersée autour de son maximum en t_0, avec une dispersion Δt_ϕ. On constate grâce au second exemple que la transformée de Fourier à court terme d'une sinusoïde complexe présente un pic en ν_0 avec une dispersion $\Delta\nu_\phi$. Ces deux situations sont illustrées sur la Figure 2.3. Le signal analysé $x(t)$ est constitué d'une sinusoïde à la fréquence ν_0 = 150 Hz et d'un Dirac d'amplitude 1 apparaissant à l'instant $t_0=0.5[s]$. On représente les spectrogrammes $|STFT_x(t,\nu)|^2$ pour des fenêtres de Hanning de longueur différente 16 et 512 respectivement, où la représentation temporelle du signal analysé $x(t)$ est affichée en haut, associée au spectrogrammes en bas.

On constate clairement que le choix d'une fenêtre "étroite" (16 échantillons) permet de mieux détecter les ruptures et les discontinuités dans les signaux. L'instant d'apparition du Dirac est bien localisé par contre la fréquence est mal localiser. En revanche une fenêtre "large" (256 échantillons) localise bien la composante périodique à 150 Hz, mais pas du tout la discontinuité.

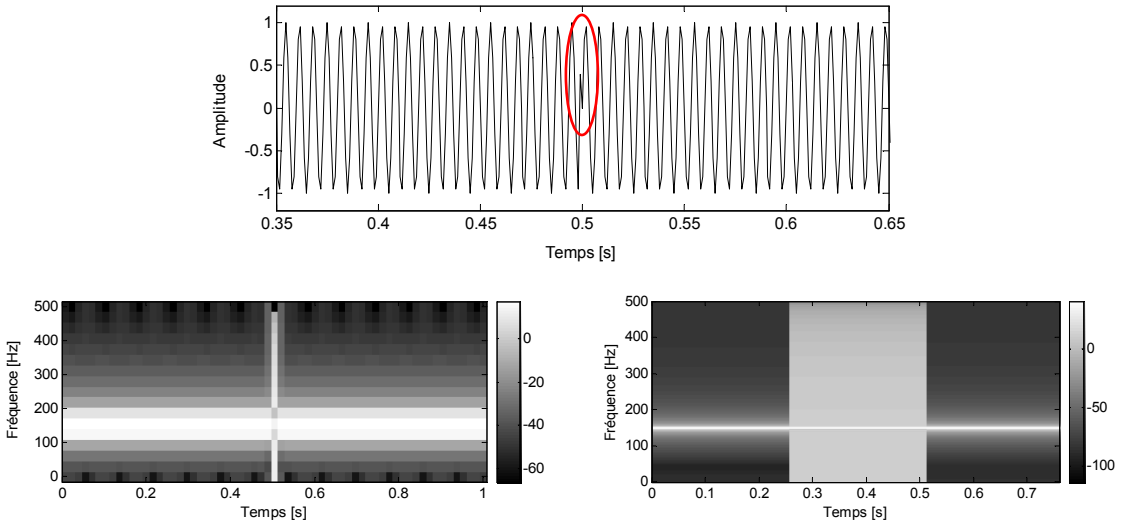


Figure 2.3 — Spectrogrammes d'un signal somme d'une impulsion et d'une sinusoïde

Le choix d'une fenêtre unique pour représenter tous les signaux de $L^2(\mathbb{R})$ par un spectrogramme impose donc un compromis entre la préservation de la localisation temporelle et celle de la localisation fréquentielle. Ce compromis est dû au principe d'incertitude de Heisenberg-Gabor, que l'on a évoqué ci-avant. Nous présentons maintenant plus en détail ce principe.

4.4.1 Taille de la fenêtre

On vient de voir que la taille (ou échelle) de la fenêtre joue un rôle fondamental dans le compromis entre les localisations temporelle et fréquentielle. Considérons la famille des fenêtres :

$$\phi_a(t) = \frac{1}{\sqrt{a}}\phi\left(\frac{t}{a}\right) \tag{2.33}$$

obtenue par dilatation $(a \geq 1)$ ou contraction $(0 < a \leq 1)$ de ϕ. Ces fenêtres définissent autant de transformée de Fourier à court terme $STFT_x^{\phi_a}$. La localisation temps-fréquence de ϕ_a est caractérisée par $a\Delta t$, $\Delta\nu/a$. Ainsi, lorsque a est petit, ϕ_a est bien localisée temporellement, mais dispersée en fréquence. La Figure 2.4 illustre le rapport entre la localisation temps-fréquence et le coefficient d'échelle a.

Si $x(t)$ est la superposition de fonctions bien localisées temporellement en des instants t_k, alors $\left|STFT_x(t,\nu)\right|^2$ sera également bien localisée au voisinage des instants t_k à condition que a soit suffisamment petit :

$$a\Delta t_\phi \ll \left|t_k - t_{k+1}\right| \tag{2.34}$$

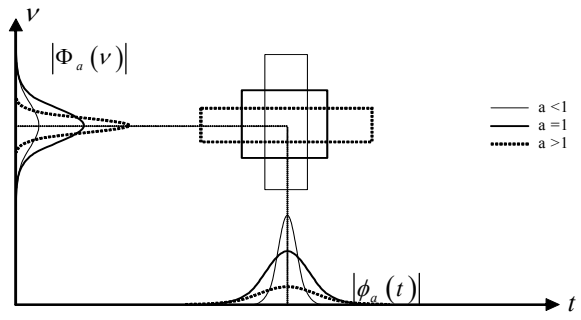


Figure 2.4 — Impact de la taille de la fenêtre sur la localisation temps-fréquence

Par contre si $x(t)$ est la superposition de différentes composantes fréquentielles $e^{i2\pi\nu_n t}$, il faudra que l'échelle a soit grande pour que la transformée de Fourier à court terme permette de résoudre ces composantes. On devra donc avoir:

$$\Delta\nu_\phi / a \ll \left|\nu_n - \nu_{n+1}\right| \tag{2.35}$$

Pour l'analyse d'un signal $x(t)$, le choix de l'échelle a de la fenêtre ϕ_a est donc crucial et dépend des structures que l'on s'attend à trouver dans le signal. La forme de la fenêtre est également un facteur important.

4.4.2 Forme de la fenêtre

En pratique, on fait appel à une fenêtre d'analyse à support compact [52]. Considérons par exemple la fenêtre rectangulaire $\phi(t) = 1_{[-1/2,1/2]}(t)$ dont la transformée de Fourier est :

$$\Phi(\nu) = sinc(\pi\nu) = \frac{\sin(\pi\nu)}{\pi\nu}$$

La transformée de Fourier à court terme d'une sinusoïde pure avec une telle fenêtre donne $\left|STFT_x(t,\nu)\right|^2 = sinc^2\left(\pi(\nu - \nu_0)\right)$. On peut observer sur la Figure 2.5 que cette fonction admet un

maximum absolu à la fréquence $\nu = \nu_0$, ce qui permet d'estimer la fréquence de la sinusoïde analysée par détection du maximum. Cependant ce maximum est entouré de lobes secondaires qui peuvent engendrer des artefacts lors de traitements utilisant la détection de pics.

Pour éviter les fausses détections de pics dus aux lobes secondaires, on évitera l'emploi d'une fenêtre rectangulaire. On cherchera plutôt à faire appel à une fenêtre dont l'amplitude relative des lobes secondaires est faible. Par ailleurs, si le lobe principal de Φ est trop large, il sera difficile de détecter deux composantes fréquentielles proches, car leurs pics risquent de fusionner. On privilégiera donc une faible bande passante BP_{-3dB} du lobe principal, définie par l'atténuation à $-3dB$. Enfin, si l'ordre r de la décroissance asymptotique $|\Phi(\nu)| = \mathcal{O}\left(|\nu|^{-r-1}\right)$ est faible, une composante de forte amplitude à la fréquence ν_0 pourra masquer les composantes éloignées en fréquence mais de beaucoup plus faible amplitude.

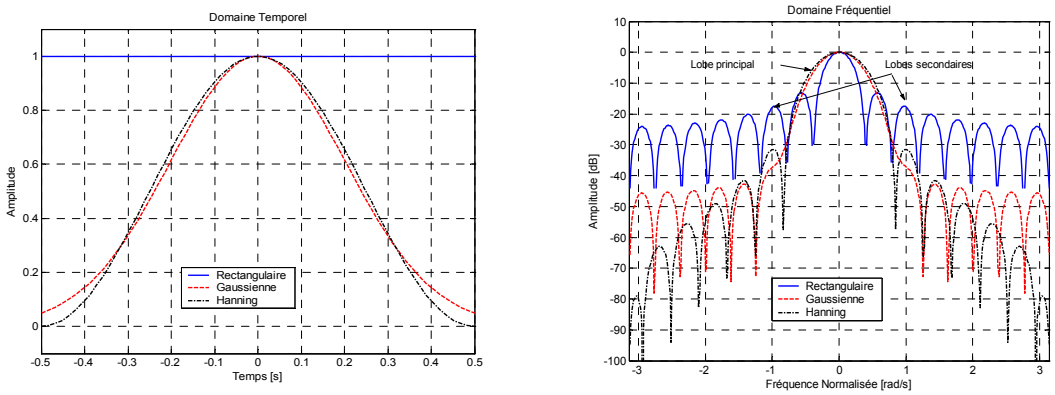


Figure 2.5 — Fenêtres d'analyse (représentation temporel, représentation fréquentiel)

Le Tableau 2.1 donne la description de différentes fenêtres d'analyse à support compact, aussi que les propriétés que l'on vient de décrire. On remarquera que la fenêtre de Hanning, qui atténue considérablement les lobes secondaires par rapport à la fenêtre rectangulaire, voit cependant la largeur de son lobe principal presque doublée. Les fenêtres (rectangulaire, Hamming, Gaussienne tronquée) pour lesquelles $r = 1$ sont discontinues en $t = \pm 1/2$. Les fenêtres de Hanning et Blackman sont, elles, continûment dérivables. En définitive, le choix d'une fenêtre (et de son paramètre d'échelle a) pour une application donnée dépend des caractéristiques attendues du signal à analyser. Il peut être crucial pour une bonne détermination des pics de la Transformée de Fourier à Court Terme.

Nom	Expression sur [-1/2,1/2]	A(dB)	BP_{-3dB}	r
Rectangle	1	-13	0.89	1
Hanning	½[1 + cos ($2\pi t$)]	-32	1.44	3
Hamming	0.54+0.46 cos ($2\pi t$)	-43	1.36	1
Gaussienne	exp (-18t^2)	-55	1.55	1
Blackman	0.42 +0.5 cos ($2\pi t$) + 0.8 cos ($4\pi t$)	-58	1.68	3

Tableau 2.1 — Fenêtres d'analyses les plus couramment utilisées

4.4.3 Conservation de l'énergie et spectrogramme

On peut représenter l'énergie de $x(t) \in L^2(\mathbb{R})$ sous la forme :

$$E_x = \int_{\mathbb{R}} |x(t)|^2 \, dt = \int\int_{\mathbb{R}\,\mathbb{R}} |STFT_x(t,\nu)|^2 \, d\nu dt = \int\int_{\mathbb{R}\,\mathbb{R}} |\langle x, \phi_{t,\nu} \rangle|^2 \, d\nu dt \tag{2.36}$$

La preuve de cette identité fait appel à l'égalité de Parseval-Plancherel [52]. Cette propriété de conservation de l'énergie justifie l'emploi de la transformée de Fourier à court terme pour définir une carte d'énergie temps-fréquence de $x(t)$: le spectrogramme $|STFT_x^{\phi}(t,\nu)|^2$ est en quelque sorte la densité d'énergie de $x(t)$ à l'instant t et à la fréquence ν.

4.5 Transformée de Fourier à court terme discrète

La discrétisation et le calcul rapide de la STFT relève des mêmes idées que la discrétisation de la transformée de Fourier décrite précédemment. On considère des signaux discrets de longueur N. La fenêtre $\phi[k]$ est un signal discret symétrique, de longueur N et de norme unité $\|\phi\| = 1$ [52]. Les atomes de Fourier à court terme résultants sont :

$$\phi_{k,n}[l] = \phi[l-k] e^{i2\pi nl/N} \tag{2.37}$$

La transformée de Fourier discrète de $\phi_{k,n}$ est :

$$\Phi_{k,n}[m] = \Phi[m-n] e^{-i2\pi k(m-n)/N} \tag{2.38}$$

La transformée de Fourier à court terme discrète de $x[k]$ est définie par :

$$STFT_x[k,n] = \langle x, \phi_{k,n} \rangle = \sum_{l=0}^{N-1} x[l] \phi[l-k] e^{-i2\pi nl/N} \tag{2.39}$$

Pour chaque $0 \le k < N$, $STFT_x[k,n]$ se calcule pour $0 \le n < N$ par transformée de Fourier discrète de $x[l]\phi[l-k]$. Ce calcul est effectué avec N FFT de taille N, ce qui demande un total de $\mathcal{O}(N^2 \log_2 N)$ opérations.

La formule de reconstruction s'écrit :

$$x[k] = \frac{1}{N} \sum_{k=0}^{N-1} \phi[l-k] \sum_{n=0}^{N-1} STFT_x[k,n] e^{i2\pi nl/N} \tag{2.40}$$

A l'instant k fixé, la seconde somme est la transformée de Fourier discrète inverse en n de $STFT_x[k,n]$. Notons qu'une transformée de Fourier à court terme discrète est une image $STFT_x[k,n]$ de taille N^2, qui est extrêmement redondante puisque déterminée par le signal x de taille N.

La transformée de Fourier à court terme que nous venons de présenter et la transformée en ondelettes que nous évoquerons dans la section suivante sont deux variantes de représentations temps-fréquence linéaires. Elles se calculent en corrélant le signal $x(t)$ avec des familles d'atomes temps-fréquence. La résolution de ces transformées est donc limitée par la résolution des atomes correspondants. Il existe aussi des représentations quadratiques, par exemple, la distribution de Wigner-Ville qui se calcule en corrélant le signal avec sa version translatée en temps et en fréquence. Cohen a introduit également une classe de distributions temps-fréquence quadratiques obtenues par moyenne pondérée de la distribution de Wigner-Ville. Ces représentations ne sont pas décrites dans ce manuscrit et le lecteur pourra se référer à [45].

5 Analyse temps-échelle

Pour analyser des composantes transitoires de durées différentes, il est nécessaire d'utiliser des atomes dont les supports temporels ont des tailles variables. Pour cela la transformée en ondelettes décompose les signaux sur une famille d'ondelettes translatées et dilatées [52]. Dans cette section, nous rappelons les propriétés de la transformée continue en ondelettes, en particulier en terme de résolution temps-fréquence et de filtrage, puis nous présentons brièvement la théorie de l'analyse multirésolution qui, d'une part formalise la définition de bases discrètes d'ondelettes obtenues par échantillonnage dyadique du facteur d'échelle, et d'autre part permet de mettre en évidence les liens existant entre bases d'ondelettes et bancs de filtres. La transformée en ondelettes stationnaire peut alors être définie comme une extension de la transformée discrète en ondelettes, c'est une représentation intermédiaire entre la transformée continue et la transformée discrète.

5.1 La Transformée en Ondelettes Continue

La transformée continue en ondelettes (CWT : *Continuous Wavelet Transform*) réalise une projection sur une base de fonctions dont la construction diffère de celle de la STFT, et auxquelles on réserve classiquement le terme d'ondelettes. Les ondelettes sont construites à partir d'une ondelette mère $\psi(t)$ à l'aide d'un opérateur de translation temporelle et d'un opérateur de changement d'échelle. La transformée continue en ondelettes est définie par :

$$\begin{aligned} & L^2(\mathbb{R}) \to L^2(\mathbb{R}^2) \\ & x(t) \quad \to CWT_x(a,t) = \int_{\mathbb{R}} x(u)\psi_{a,t}^*(u)\,du = \langle x, \psi_{a,t} \rangle \end{aligned} \tag{2.41}$$

avec :

$$\psi_{a,t}(u) = \frac{1}{\sqrt{a}}\psi\left(\frac{u-t}{a}\right) \tag{2.42}$$

On peut montrer que la transformée est inversible si l'ondelette mère vérifie une condition dite d'admissibilité :

$$\int_{\mathbb{R}} |\Psi(\nu)|^2 \frac{d\nu}{|\nu|} = C_\psi < \infty \; ; \; \int_{\mathbb{R}} \psi(t)\,dt = \Psi(0) = 0 \tag{2.43}$$

où $\Psi(\nu)$ est la transformée de Fourier de $\psi(t)$. Pour assurer la convergence de l'intégrale, il faut à la fois que $\Psi(\nu)$ décroisse suffisamment vite quant $\nu \to +\infty$ et que $\Psi(0) = 0$. Cette dernière condition peut se réécrire $\int_{\mathbb{R}} \psi(t)\,dt = 0$. On obtient alors la formule de reconstruction de $x(t)$ à partir des CWT_x :

$$x(t) = \frac{1}{C_\psi} \int_{\mathbb{R}} \int_{\mathbb{R}} CWT_x(a,u)\psi_{a,t}(u) \frac{dadu}{a^2} \tag{2.44}$$

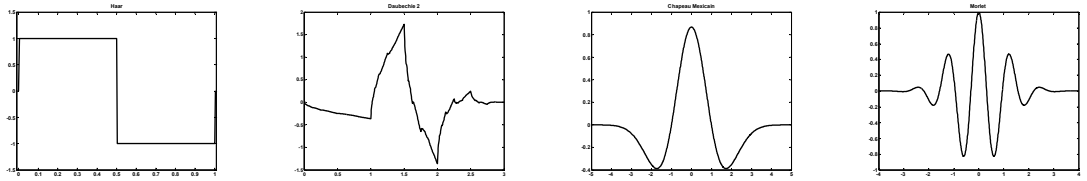


Figure 2.6 — Exemples d'ondelettes mères

La CWT possède la propriété de covariance par translation en temps et dilatation en échelle :

$$y(t)=\frac{1}{\sqrt{a}_0}x\left(\frac{t-t_0}{a_0}\right) \quad \Rightarrow \quad CWT_y(a,t)=CWT_x\left(\frac{a}{a_0},\frac{t-t_0}{a_0}\right)$$

Dans cette approche, la spécificité de la transformée ne dépend pas d'une fonction fenêtre, mais de l'ondelette mère qui peut avoir des formes très variées (cf. Figure 2.6).

5.2 Interprétation en terme de filtrage

Les définitions précédentes basées sur des produits scalaires peuvent être reformulées en utilisant des produits de convolution, qui permettent d'interpréter les lignes du plan temps-échelle comme autant de signaux résultants du filtrage du signal original [55]. Ainsi, à partir de la définition (2.41), la transformée en ondelettes peut être exprimée sous la forme d'un produit de convolution :

$$CWT_x(a,t)=\int_{\mathbb{R}} x(u)\overline{\psi}_a^*(t-u)\,du=\left(x*\overline{\psi}_a^*\right)(t) \tag{2.45}$$

où $\overline{\psi}(t)=\psi(-t)$. Si l'on considère $x_a(t)$ le signal obtenu en réalisant une coupe horizontale du plan temps-échelle à une échelle a fixée, ce signal résulte ainsi directement du filtrage du signal $x(t)$. Dans le cas d'une ondelette à valeurs réelles, la réponse fréquentielle du filtre est $\left|\Psi_a(\nu)\right|$ où $\Psi_a(\nu)$ est la transformée de Fourier de $\psi_a(t)$. La condition d'admissibilité $\Psi(0)=0$ donne à ces filtres la qualité de filtres passe-bande. Nous verrons au paragraphe suivant que leur largeur de bande dépend de la fréquence analysée.

5.3 Résolution dans le plan temps-échelle

La transformée CWT utilise des ondelettes dont l'enveloppe rétrécit avec l'échelle tandis que le nombre d'oscillations reste constant (cf. Figure 2.7). En effet l'opération de changement d'échelle de facteur a est équivalente à une homothétie de paramètre a sur l'axe des temps et de $1/\sqrt{a}$ sur celui des amplitudes. Dans ce cas, on obtient une largeur de bande qui augmente avec la fréquence, avec une résolution fréquentielle relative constante. Les pavages des plans temps-fréquence et temps-échelle, indicateurs des zones d'influence des différents atomes d'analyse, vont donc être fondamentalement différents.

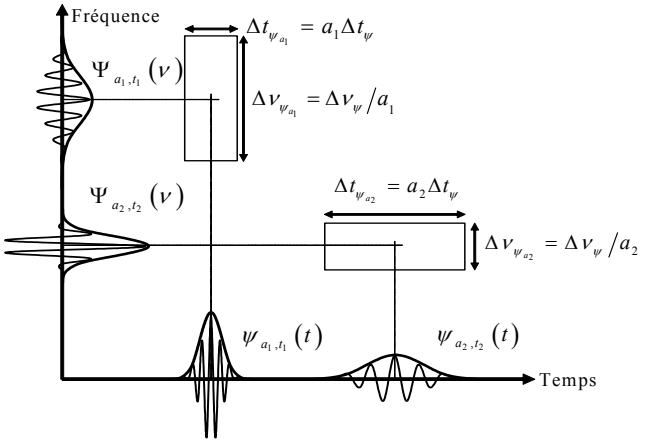


Figure 2.7 — Pavage temps-fréquence associé à la Transformée en ondelettes

On a vu que $CWT_x(a,t)$ pouvait être interprété comme résultant du filtrage du signal $x(t)$ par l'ondelette "retournée", celle-ci ayant un caractère passe-bande. Ceci permet de lui associer une fréquence caractéristique (dite centrale) $\nu_\psi = \int_0^{+\infty} \nu \left|\Psi(\nu)\right|^2 d\nu$, où $\Psi(\nu)$ est la transformée de Fourier de l'ondelette mère $\psi(t)$. La transformée de Fourier de l'ondelette dilatée $\psi_a(t)$ à l'échelle a, s'écrit :

$$\Psi_a(\nu) = \sqrt{a}\ \Psi(a\nu) \tag{2.46}$$

ce qui met en évidence le fait que les filtres associés aux différentes échelles ont une largeur de bande dépendant de l'échelle et que l'opérateur de dilatation translate la fréquence centrale selon $\nu_{\psi_a} = \nu_\psi / a$. Cette relation entre échelle et fréquence permet de regraduer en fréquence l'axe des échelles par la transformation $a \to \nu = \nu_\psi / a$, et de voir le plan temps-échelle comme une exploration particulière du plan temps-fréquence. Les résolutions temporelle et fréquentielle de l'ondelette ψ_a vérifient alors :

$$\begin{cases} \Delta t_{\psi_a} = a\Delta t_\psi \\ \Delta \nu_{\psi_a} = \Delta \nu_\psi / a \end{cases} \tag{2.47}$$

La résolution fréquentielle dépend donc de la fréquence centrale de l'ondelette et le facteur de qualité Q du filtre ondelette est invariant par action de l'opérateur de dilatation :

$$Q_{\psi_a} = \frac{\nu_{\psi_a}}{\Delta \nu_{\psi_a}} = \frac{\nu_\psi}{\Delta \nu_\psi} = Q_\psi \tag{2.48}$$

La CWT travaille donc à résolution fréquentielle relative constante. Cela implique que :

- a petit, correspond aux hautes fréquences, et donne une "bonne" résolution temporelle et une "mauvaise" résolution fréquentielle.
- a grand, correspond aux basses fréquences, avec une "mauvaise" résolution temporelle et une "bonne" résolution fréquentielle.

En terme de pavage temps-échelle (ou fréquence), cela se traduit par des pavés élémentaires qui se déforment lorsque l'échelle d'analyse varie, leur surface demeurant constante (cf. Figure 2.7). Dans cette représentation, chaque cellule correspond à la résolution d'analyse pour le point du plan situé en son centre, les lignes horizontales délimitant les bandes passantes des filtres associés.

La CWT, peut être échantillonnée, c'est-à-dire calculée pour des valeurs discrètes de a et de t. L'implantation de la transformée continue correspond généralement à un échantillonnage uniforme du temps calqué sur celui du signal, et à la sélection d'un nombre constant d'échelles par bandes d'octaves. Cette implantation conserve la caractéristique de redondance attachée à la transformée continue, tant du point de vue du temps que de l'échelle. Nous allons nous intéresser à deux transformées discrètes ne présentant pas de redondance sur l'axe des échelles : la transformée en ondelettes discrète et la transformée en ondelettes stationnaires (dite continue-discrète).

5.4 Bases discrètes d'ondelettes

La transformée en ondelettes discrète consiste en la projection du signal sur une base discrète d'ondelettes réalisant un échantillonnage critique du plan temps-échelle (a,t), c'est-à-dire tel qu'un échantillonnage plus serré conduirait à une redondance d'information alors qu'un échantillonnage plus lâche ne permettrait plus la reconstruction du signal. Une première approche consiste à rechercher, à partir de l'ensemble continu des fonctions de base utilisées dans la CWT (transformée redondante dans la mesure où deux coefficients voisins partagent de l'information) une famille discrète possédant les mêmes propriétés [42]. Cependant, la définition de telles bases par échantillonnage critique de la CWT ne fournit pas forcément de "bonnes" bases d'ondelettes, en particulier la base duale nécessaire à la reconstruction du signal n'est pas obligatoirement une base d'ondelettes. Une autre approche consiste à définir les conditions pour engendrer directement des bases discrètes d'ondelettes .

5.4.1 L'analyse multirésolution

L'analyse multirésolution (MRA : *Multiresolution Analysis*) donne un cadre formel à l'élaboration de ces bases d'ondelettes et montre que la seule restriction à l'obtention de celles-ci porte sur le motif générateur, l'ondelette mère $\psi(t)$ (notée $\psi_{0,0}(t)$), qui doit vérifier une condition d'admissibilité équivalente à celle de la transformation continue (2.43). Cette théorie définit la grille dyadique (j,k) correspondant à l'échantillonnage du plan (a,t) aux points $(a_j,t_k)=(2^j,2^j k)$. La collection $\left\{\psi_{j,k}(t)=2^{-j/2}\psi_{0,0}\left(2^{-j}t-k\right),\ (j,k)\in\mathbb{Z}^2\right\}$ constitue alors une base (non nécessairement orthogonale) multirésolution de $L^2(\mathbb{R})$. La base duale $\left\{\tilde{\psi}_{j,k}(t),\ (j,k)\in\mathbb{Z}^2\right\}$, définie par $\left\langle\psi_{i,k},\tilde{\psi}_{j,k}\right\rangle=\delta_{ij}$, est une base d'ondelettes qui peut être facilement obtenue, ce qui rend possible la reconstruction exacte du signal.

5.4.1.1 Définition

L'analyse multirésolution revient à projeter le signal $x(t)$ dans une série de sous-espaces d'approximations V_j et de sous-espaces de détails W_j. Les projections successives du signal sur les espaces V_j en constituent des approximations de plus en plus grossières. L'information perdue entre deux approximations est collectée dans les signaux de détails, qui rendent alors compte du comportement du signal à des résolutions différentes.

Une analyse multirésolution de $L^2(\mathbb{R})$ est une suite $\{V_j\}_{j\in\mathbb{Z}}$ de sous-espaces fermés de $L^2(\mathbb{R})$ ayant les propriétés suivantes [50]:

i) $\bigcap_{j\in\mathbb{Z}} V_j=\{0\}$, $\bigcup_{j\in\mathbb{Z}} V_j$ est dense dans $L^2(\mathbb{R})$ et $V_{j+1}\subset V_j$;

ii) Pour toute fonction $x(t)$ de $L^2(\mathbb{R})$ et tout $j\in\mathbb{Z}$, $x(t)\in V_j \Leftrightarrow x\left(2^j t\right)\in V_0$;

iii) Pour toute fonction $x(t)$ de V_0 et tout $k\in\mathbb{Z}$, $x(t-k)\in V_0$; (2.49)

iv) Il existe une fonction $\phi(t)$ (notée $\phi_{0,0}(t)$) de V_0 telle que l'ensemble $\{\phi_{0,0}(t-k),\ k\in\mathbb{Z}\}$ constitue une base inconditionnelle de V_0.

Comme $V_{j+1}\subset V_j$, la projection dans V_{j+1} constitue une approximation plus grossière du signal que la projection dans V_j. Donc l'information contenue dans V_j est plus riche que celle contenue dans V_{j+1} (cf. propriété i) ci-avant). La propriété ii) exprime la possibilité de passer d'un espace V_j à un espace V_i par un simple changement d'échelle. le point iii) montre que

chaque espace V_j est invariant par translation temporelle. Enfin la propriété iv) traduit le fait que l'on va engendrer V_0 par simple translation temporelle d'un même motif de base $\phi_{0,0}(t)$.

5.4.1.2 Fonction d'échelle et espaces d'approximation

Le sous-espace d'approximation V_0 est défini à l'aide de la fonction d'échelle $\phi_{0,0}(t)$ et de ses translatées. Les sous-espaces V_j sont engendrés par les dilatées d'un facteur 2^j de cette fonction et de leurs translatées $\left\{\phi_{j,k}(t)=2^{-j/2}\phi_{0,0}\left(2^{-j}t-k\right),\ (j,k)\in\mathbb{Z}^2\right\}$ [42]. Ces fonctions constituent une base de Riesz de V_j :

$$V_j\left(\phi_{0,0}\right)=\left\{x:x(t)=\sum_{k\in\mathbb{Z}}a_x(j,k)2^{-j/2}\phi_{0,0}\left(2^{-j}t-k\right);\ \left\{a_x(j,k)\right\}\in l^2\left(\mathbb{Z}^2\right)\right\} \tag{2.50}$$

Pour un signal $x(t)$ d'énergie finie, les coefficients d'approximation au niveau j sont définis par :

$$a_x(j,k)=\int_{\mathbb{R}}x(t)\phi_{j,k}^*(t)\,dt=\left\langle x,\phi_{j,k}\right\rangle \tag{2.51}$$

L'approximation du signal $x(t)$ à la résolution 2^{-j} correspond à sa projection dans V_j :

$$A_jx(t)=\sum_{k\in\mathbb{Z}}a_x(j,k)\ \tilde{\phi}_{j,k}(t) \tag{2.52}$$

5.4.1.3 Fonction ondelette et espaces de détails

Après avoir défini l'approximation d'un signal à la résolution 2^{-j}, il semble naturel de s'interroger sur la perte d'information qu'occasionne cette approximation [42]. C'est pour cela que l'on définit la famille des espaces de détails $\left\{W_j\right\}_{j\in\mathbb{Z}}$ par :

$$V_{j-1}=V_j\oplus W_j \tag{2.53}$$

où W_j est le supplémentaire orthogonal de V_j dans V_{j-1}. Un des grands résultats de la MRA est la preuve de l'existence d'une fonction $\psi(t)$, appelée ondelette mère et notée $\psi_{0,0}(t)$, construite à partir de la fonction d'échelle et telle que la famille $\left\{\psi_{j,k}(t)=2^{-j/2}\psi_{0,0}\left(2^{-j}t-k\right),\ (j,k)\in\mathbb{Z}^2\right\}$ constitue une base de Riesz de W_j.

$$W_j\left(\psi_{0,0}\right)=\left\{x:x(t)=\sum_{k\in\mathbb{Z}}d_x(j,k)2^{-j/2}\psi_{0,0}\left(2^{-j}t-k\right);\left\{d_x(j,k)\right\}\in l^2\left(\mathbb{Z}^2\right)\right\} \tag{2.54}$$

Comme dans le cas des espaces d'approximation, on définit les coefficients de détail au niveau j par :

$$d_x(j,k)=\int_{\mathbb{R}}x(t)\psi_{j,k}^*(t)\,dt=\left\langle x,\psi_{j,k}\right\rangle \tag{2.55}$$

et le signal de détail associé comme la projection de $x(t)$ dans W_j :

$$D_jx(t)=\sum_{k\in\mathbb{Z}}d_x(j,k)\ \tilde{\psi}_{j,k}(t) \tag{2.56}$$

où $\tilde{\psi}_{j,k}(t)$ est l'ondelette duale. L'équation (2.53) signifie que la projection du signal dans l'espace W_j correspond à l'information perdue quand on passe de l'approximation dans V_{j-1} à l'approximation dans V_j :

$$A_{j-1}x(t)-A_jx(t)=D_jx(t) \tag{2.57}$$

On montre également que, si $I<J$, l'approximation dans V_I peut être obtenue en sommant les détails contenus dans les $\left\{W_j\right\}_{j\in\{I+1,\ldots,J\}}$ à l'approximation au niveau J :

$$A_I x(t) = A_J x(t) + \sum_{j=\{I+1,\ldots,J\}} D_j x(t) \tag{2.58}$$

Par construction, les W_j sont orthogonaux entre eux et leur somme directe couvre $L^2(\mathbb{R})$:

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j \tag{2.59}$$

La formule de reconstruction est donnée par :

$$x(t) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} d_x(j,k)\ \tilde{\psi}_{j,k}(t) \tag{2.60}$$

Du fait de l'emboîtement des espaces d'approximation et de détails, il existe des relations entre fonctions d'échelles de deux espaces V_j et V_{j+1} ainsi qu'entre fonction d'échelle et ondelette. Nous allons les expliciter, dans la mesure où ces relations sont, avec les précédentes, à la base des algorithmes utilisés pour le calcul de la DWT mais aussi de la SWT.

5.4.1.4 Relations à deux échelles

On parle de relations à deux échelles les relations qui explicitent le lien entre les fonctions se situant dans deux espaces consécutifs. Comme par définition $V_1 \subset V_0$, on peut exprimer les éléments de V_1 comme combinaison linéaire des vecteurs de base de V_0 et comme par ailleurs $\phi_{1,0} = 1/\sqrt{2}\,\phi_{0,0}(t/2) \in V_1$, on a donc :

$$\phi_{1,0}(t) = \sum_{k\in\mathbb{Z}} h(k)\ \phi_{0,0}(t-k) = h * \phi_{0,0}(t) \tag{2.61}$$

De même, avec $W_1 \subset V_0$ et $\psi_{1,0} = 1/\sqrt{2}\,\psi_{0,0}(t/2) \in V_1$ on peut écrire :

$$\psi_{1,0}(t) = \sum_{k\in\mathbb{Z}} g(k)\ \phi_{0,0}(t-k) = g * \phi_{0,0}(t) \tag{2.62}$$

Ces relations mettent en évidence l'existence des filtres h et g, dont la connaissance est équivalente à la connaissance des fonctions $\phi_{0,0}$ et $\psi_{0,0}$. Le filtre h est un filtre passe-bas alors que le filtre g est un filtre passe-haut. Nous reviendrons ultérieurement sur les propriétés de ces filtres.

5.4.2 Transformée discrète dyadique en ondelettes

Elle est définie par la collection des $\{d_x(j,k)\}_{(j,k)\in\mathbb{Z}^2}$ projections orthogonales de $x(t)$ dans les espaces d'ondelettes $\{W_j\}_{j\in\mathbb{Z}}$ définis par la MRA :

$$\begin{aligned} &L^2(\mathbb{R}) \to l^2(\mathbb{Z}^2) \\ &x(t) \ \ \to d_x(j,k) = \langle x, \psi_{j,k} \rangle \end{aligned} \tag{2.63}$$

Cette transformée peut être inversée à l'aide d'une somme discrète qui met en jeu la base duale $\{\tilde{\psi}_{j,k}\}_{(j,k)\in\mathbb{Z}^2}$ [42]:

$$x(t) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} d_x(j,k)\ \tilde{\psi}_{j,k}(t) \tag{2.64}$$

Contrairement à la situation où l'on échantillonne la transformée continue réalisée avec une ondelette non issue d'une multirésolution :

- la MRA permet d'expliciter la base duale ;
- elle rend utilisable la formule de reconstruction exacte ;
- la base duale est une base d'ondelettes ;
- dans le cas d'ondelettes orthogonales, la base duale est la base de départ:

$$\tilde{\psi}_{j,k}(t) = \psi_{j,k}(t) \quad \forall (j,k) \in \mathbb{Z}^2 \tag{2.65}$$

Notons qu'on a une équivalence entre CWT et DWT à condition que l'ondelette d'analyse soit issue d'une analyse multirésolution. Dans ce cas la DWT correspond à l'échantillonnage sur la grille dyadique de la CWT calculée sur $\tilde{x}$, projection de x sur V_0 :

$$d_x(j,k) = CWT_{\tilde{x}}\left(2^j, k2^j\right) \tag{2.66}$$

Remarque : contrairement à la CWT, la DWT n'est pas covariante par action de l'opérateur de translation en temps : si $y(t) = x(t-t_0)$, on n'a pas nécessairement $d_y(j,k) = d_x(j,k-k_0)$. Cette constatation nous conduira à utiliser une autre transformée en ondelettes, à savoir la transformée stationnaire qui possède la propriété de covariance par translation dans le temps.

5.4.2.1 Calcul des coefficients de la DWT : Algorithme pyramidal

Les coefficients de la DWT s'obtiennent grâce à l'algorithme pyramidal rapide proposé par S. Mallat [52]. Cet algorithme s'appuie sur les relations à deux échelles (2.61) (2.62), pour effectuer un calcul récursif des $d_x(j,k)$ utilisant des filtres déduits de h et g. Si l'on définit l'opérateur de décimation d'un facteur 2 par :

$$y(k) = \downarrow 2[x](k) = x(2k) \quad \forall k \in \mathbb{Z} \tag{2.67}$$

on montre que les coefficients d'approximation et de détail à un niveau j sont obtenus par filtrage et décimation des coefficients d'approximation au niveau $j-1$

$$a_x(j,k) = \downarrow 2\left[\tilde{h}(.) * a_x(j-1,.)\right](k) \tag{2.68}$$

$$d_x(j,k) = \downarrow 2\left[\tilde{g}(.) * a_x(j-1,.)\right](k) \tag{2.69}$$

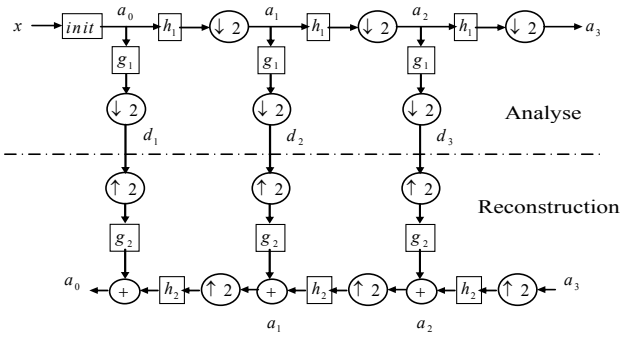


Figure 2.8 — Algorithme pyramidal d'analyse–reconstruction

Les filtres $\tilde{h}$ et $\tilde{g}$ utilisés à tous les niveaux de la décomposition sont définis à partir de h et g, et seront notés h_1 et g_1 :

$$\begin{cases} h_1(k) = \tilde{h}(k) = h(-k) \\ g_1(k) = \tilde{g}(k) = g(-k) \end{cases} \tag{2.70}$$

On obtient ainsi l'algorithme pyramidal schématisé sur la Figure 2.8.

L'étape d'initialisation consiste à projeter le signal continu $x(t)$ sur V_0. Dans la pratique, on dispose du signal échantillonné $x[k]$, et une solution simple consiste à prendre $a_0[k] = x[k]$.

Une analyse multirésolution d'un signal de N points se fera sur au maximum $J = \log_2(N)$ niveaux et fournira une représentation de N points. En effet, le nombre de coefficients de détails au niveau j est $N/2^j$ et le dernier niveau fournit 1 coefficient de détail et 1 coefficient d'approximation (la composante la plus basse fréquence du signal), ce qui nous donne un total de $N-1$ coefficients de détails et 1 coefficient d'approximation.

Pour la reconstruction, la DWT possède un algorithme rapide basé en particulier sur les relations (2.53) et (2.57), qui utilise l'opérateur d'interpolation d'un facteur 2 défini par :

$$\forall k \in \mathbb{Z} \begin{cases} y(2k) &= \uparrow 2[x](2k) = x(k) \\ y(2k+1) &= \uparrow 2[x](2k+1) = 0 \end{cases} \tag{2.71}$$

Selon cet algorithme, chaque niveau d'approximation est reconstruit par sommation de l'approximation plus grossière et du détail correspondant, préalablement interpolés et filtrés :

$$a_x(j-1,k) = \left(h_2(.) * \uparrow 2\left[a_x(j,.)\right]\right)(k) + \left(g_2(.) * \uparrow 2\left[d_x(j,.)\right]\right)(k) \tag{2.72}$$

Les filtres utilisés notés h_2 et g_2, sont liés aux fonctions des bases duales et définis par :

$$\begin{cases} h_2(k) = \tilde{h}(k) \\ g_2(k) = \tilde{g}(k) \end{cases} \text{avec} \begin{cases} \tilde{\phi}_{1,0}(t) = \tilde{g} * \tilde{\phi}_{0,0}(t) \\ \tilde{\psi}_{1,0}(t) = \tilde{h} * \tilde{\phi}_{0,0}(t) \end{cases} \tag{2.73}$$

Dans le cas de bases orthogonales d'ondelettes, où les fonctions et leurs duales sont identiques, on a :

$$\begin{cases} h_2(k) = h(k) \\ g_2(k) = g(k) \end{cases} \tag{2.74}$$

5.4.2.2 Filtrage par bandes et résolution dans le plan temps-fréquence

On a vu qu'une MRA est complètement définie par ses fonctions échelle ϕ et ondelette mère ψ, ou de manière équivalente, par les filtres h et g. Ceux-ci vérifient les conditions de reconstruction parfaite du signal correspondant au schéma illustré sur la Figure 2.9. H_1, H_2 représentent les transformées en z des filtres h_1, h_2 et G_1, G_2 celles de g_1, g_2. Ces transformées en z doivent vérifier :

$$\begin{cases} H_1(-z)H_2(z) + G_1(-z)G_2(z) = 0 \\ H_1(z)H_2(z) + G_1(z)G_2(z) = 2 \end{cases} \tag{2.75}$$

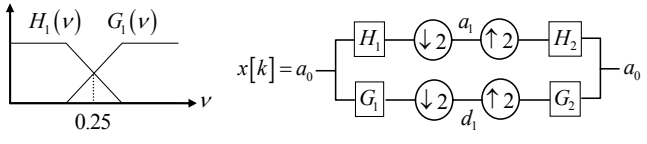


Figure 2.9 — Filtres "miroirs"

Dans le cas d'une base d'ondelettes orthogonale, H_1 et G_1 sont des filtres dits miroirs en quadrature [54] caractérisés par :

$$\begin{cases} H_1(z)H_1(z^{-1}) + H_1(-z)H_1(-z^{-1}) = 1 \\ G_1(z) = zH_1(-z^{-1}) \quad \text{soit} \quad g_1(k) = (-1)^k h_1(1-k) \end{cases} \tag{2.76}$$

Les conditions de reconstruction parfaite (2.74) impliquent que :

$$\begin{cases} H_2(z) = H_1(z^{-1}) \\ G_2(z) = G_1(z^{-1}) \end{cases} \tag{2.77}$$

Les réponses fréquentielles des filtres (cf. Figure 2.9) vérifient :

$$|G_1(\nu)| = |H_1(0.25-\nu)| \text{ et } |H_1(\nu)|^2 + |G_1(\nu)|^2 = 1 \tag{2.78}$$

$$H_1(\nu = 0) = \sqrt{2} \text{ et } G_1(\nu = 0) = 0 \tag{2.79}$$

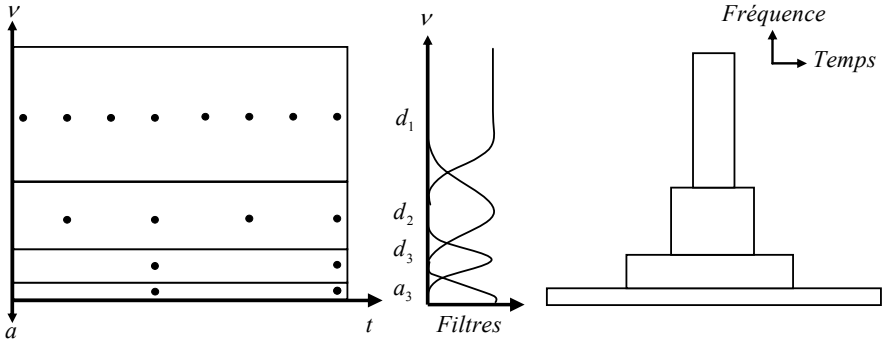


Figure 2.10 — DWT : Partition fréquentielle et coefficients de détails – largeur de bandes de détails et approximation – résolution temps fréquence associé aux $d_x(j,k)$

Les opérations successives de filtrage et de décimation répétées tout au long de la décomposition réalisent ainsi une partition de l'axe fréquentiel sous la forme d'un banc de filtres à structure dyadique, selon le schéma de la Figure 2.10. Chaque détail d_j contient de l'information relative à tout le signal dans une bande de fréquence particulière indicée par j et dont la largeur est égale à $2^{-(j+1)}$. A chaque détail d_j, les coefficients $d_x(j,k)$ ont une résolution temporelle en 2^j. On peut ainsi remarquer que bien que provenant d'opérations successives de filtrage, ces coefficients ne peuvent pas être assimilés à des signaux filtrés en raison des opérations de décimation qu'ils ont subies.

5.4.3 Transformée en Ondelettes Stationnaire (Représentation continue dyadique)

La transformée en ondelettes stationnaire (SWT : *Stationary Wavelet Transform*) constitue, en terme de redondance, une représentation intermédiaire entre la CWT à forte redondance et la DWT non redondante. Elle conserve un échantillonnage dyadique des échelles, et maintient de ce fait une non redondance fréquentielle, mais réalise un échantillonnage quasi continu et uniforme du temps [57][59]. Elle présente un intérêt dans le cadre de notre étude car, contrairement à la DWT, les coefficients de sa décomposition ont le statut de signaux filtrés. La SWT est définie par le calcul des coefficients $\{cd_x(j,k)\}_{(j,k)\in\mathbb{Z}^2}$ tels que :

$$cd_x(j,k) = \int_{\mathbb{R}} x(t)\psi'^{*}_{j,k}(t)\,dt = \left\langle x, \psi'_{j,k} \right\rangle \tag{2.80}$$

$$\text{où} \quad \psi'_{j,k}(t) = 2^{-j/2}\psi_{0,0}\left(2^{-j}(t-k)\right) \tag{2.81}$$

La relation entre les bases d'ondelettes utilisées dans la DWT et la SWT est mise en œuvre en utilisant la formulation suivante :

$$\begin{gathered} \psi_{j,k}(t) = \psi_{j,0}\left(t - 2^j k\right) \qquad \psi'_{j,k}(t) = \psi_{j,0}(t-k) \\ \text{avec} \qquad \psi_{j,0}(t) = 2^{-j/2}\psi\left(t/2^j\right) \end{gathered} \tag{2.82}$$

Ainsi, la SWT coïncide avec la CWT sur la grille uniforme et la DWT sur la grille dyadique (cf. Figure 2.11) où les $cd_x(j,k)$ ont la même résolution fréquentielle que les $d_x(j,k)$, mais partagent de l'information sur l'axe des temps :

$$\begin{aligned} cd_x(j,k) &= CWT_x\left(a = 2^j, t = k\right) \\ cd_x\left(j, 2^j k\right) &= d_x(j,k) \end{aligned} \tag{2.83}$$

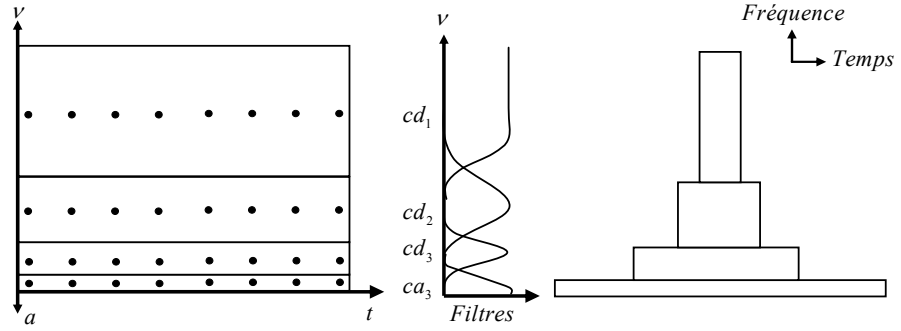


Figure 2.11 — SWT : Partition fréquentielle et coefficients de détails – largeur de bandes de détails et approximation – résolution temps fréquence associé aux $cd_x(j,k)$

La SWT présente l'avantage, par rapport à la DWT, d'être covariante par action de l'opérateur de translation en temps. Ceci résulte du fait qu'elle présente un échantillonnage en temps identique à celui du signal. C'est également ce qui explique que les cd_j aient la propriété d'être des signaux filtrés.

Un exemple simple pour illustrer cette caractéristique apparaît dans l'analyse d'un signal $x[k]$ de type créneau. Il est constitué de la manière suivante : 128 valeurs nulles, puis 255 valeurs égales à 1 et pour terminer, 129 valeurs nulles. A l'aide de la DWT, avec l'ondelette de Haar, on trouve, à gauche de la Figure 2.12 l'analyse MRA au niveau 1 de $x[k]$. Au milieu on a représenté l'analyse MRA au niveau 1 de $x[k]$ translaté d'un pas de temps et à droite l'analyse par la SWT.

Les deux analyses MRA sont très différentes, comme le montrent les détails obtenus. En particulier, aucune ne repère simultanément les deux discontinuités. Pour la première (à gauche), le détail est fort au voisinage de l'instant 128 et nul au voisinage de l'instant 384 alors que pour la seconde (au milieu) c'est l'inverse. Par contre les deux discontinuités sont repérées

simultanément par l'analyse SWT (à droite). Cette propriété est intéressante dans plusieurs domaines d'application comme par exemple, la détection d'instants de "ruptures" dans un signal.

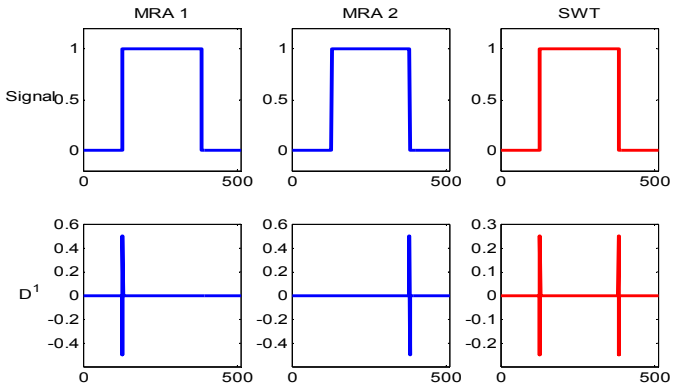


Figure 2.12 — Décomposition d'un créneau et de sa version translatée

5.4.3.1 Mise en œuvre de l'algorithme à trous

Lorsque l'ondelette d'analyse $\psi_{0,0}$ est construite à partir d'une MRA associée aux filtres h et g, cette décomposition peut être effectuée par un algorithme récursif, l'algorithme pyramidal, à la différence notable que, les coefficients de la SWT respectent un échantillonnage temporel uniforme. Le passage à une échelle inférieure ne s'accompagne pas d'une opération de décimation des coefficients d'approximation et de détail, mais de l'interpolation des filtres utilisés. En définissant, par analogie avec la DWT, les coefficients d'approximation de cette transformée :

$$ca_x(j,k) = \int_{\mathbb{R}} x(t)\phi'^{*}_{j,k}(t)\,dt = \left\langle x, \phi'_{j,k} \right\rangle \tag{2.84}$$

les $cd_x(j,k)$ et les $ca_x(j,k)$ s'obtiennent à partir du niveau précédent grâce à des versions interpolées de h_1 et g_1 selon la définition (2.83), c'est-à-dire résultant de l'insertion de zéros entre les coefficients de ces filtres (d'où le nom d'algorithme à trous) :

$$\begin{aligned} ca_x(j,k) &= \left[\uparrow 2^{j-1}[h_1] * ca_x(j-1,.)\right](k) \\ cd_x(j,k) &= \left[\uparrow 2^{j-1}[g_1] * ca_x(j-1,.)\right](k) \end{aligned} \tag{2.85}$$

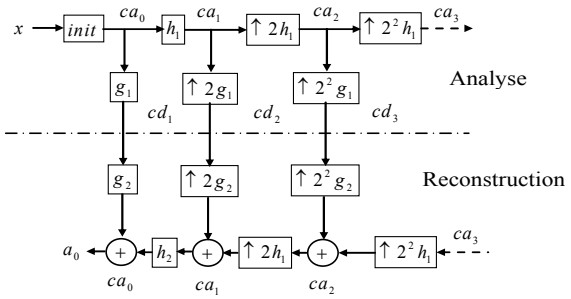


Figure 2.13 — Algorithme de décomposition–reconstruction par SWT

La structure complète de cet algorithme est donnée sur la Figure 2.13, en utilisant les notations :

$$\begin{aligned} ca_j &= \left\{ ca_x(j,k) \right\}_{k\in\mathbb{Z}} \\ cd_j &= \left\{ cd_x(j,k) \right\}_{k\in\mathbb{Z}} \end{aligned} \tag{2.86}$$

Dans le cas d'ondelettes orthogonales, la base duale est identique à la base de départ à un facteur multiplicatif près [42]. Il existe un algorithme de reconstruction équivalent, basé sur l'utilisation des séquences h_2 et g_2 interpolées [42]. Le schéma simplifié d'analyse-reconstruction à un niveau correspondre alors à celui de la Figure 2.14.

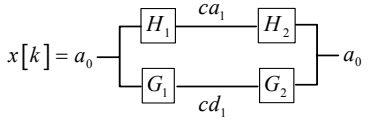


Figure 2.14 — Les filtres pour SWT

5.4.3.2 Interprétation en banc de filtres

Un des avantages de cette décomposition réside dans son interprétation en terme de banc de filtres. On peut en effet exprimer l'équation (2.83) sous la forme :

$$\begin{aligned} ca_x(i,k) &= \uparrow 2^{i-1}[h_1] * \left(\uparrow 2^{i-2}[h_1] * ... * \uparrow 2[h_1] * [h_1] * x(k) \right) \\ cd_x(i,k) &= \uparrow 2^{i-1}[g_1] * \left(\uparrow 2^{i-2}[h_1] * ... * \uparrow 2[h_1] * [h_1] * x(k) \right) \end{aligned} \tag{2.87}$$

En définissant de manière itérative à partir de h_1 et g_1 deux séquences de filtres de longueur croissante :

$$\begin{cases} \tilde{h}_{i+1}(k) = \left(\uparrow 2^i [h_1] * \tilde{h}_i \right)(k) \\ \tilde{g}_{i+1}(k) = \left(\uparrow 2^i [g_1] * \tilde{h}_i \right)(k) \end{cases} \text{ soit } \begin{cases} \tilde{H}_{i+1}(z) = H\left(z^{2^i}\right) \tilde{H}_i(z) \\ \tilde{G}_{i+1}(z) = G\left(z^{2^i}\right) \tilde{H}_i(z) \end{cases} \tag{2.88}$$

avec la condition initiale $\tilde{H}_0(z) = 1$, on peut exprimer [42] les signaux de détails et d'approximation comme résultant directement du filtrage du signal x par ce banc de filtres :

$$\begin{aligned} ca_x(j,k) &= \left(\tilde{h}_i * x \right)(k) \\ cd_x(j,k) &= \left(\tilde{g}_i * x \right)(k) \end{aligned} \tag{2.89}$$

La formulation banc de filtres peut être intéressante lorsque l'on souhaite calculer uniquement un signal (ca_J ou cd_J) : il est alors plus économique de déterminer directement le filtre correspondant.

5.5 Principales caractéristiques des ondelettes

Nous avons déjà rencontré, au fil de cet exposé, certaines des propriétés souhaitées des ondelettes. Nous allons ici faire une liste de leurs principales caractéristiques.

5.5.1 Le cône d'influence

L'extension avec l'échelle d'analyse du support du motif analysant a pour effet, aux grandes échelles, de délocaliser l'information [42]. A la verticale de points où le signal est nul, des coefficients peuvent, pour les échelles les plus grandes, être non nuls, simplement parce que l'ondelette mesure le comportement sur une portion très étendue du signal, où celui-ci n'est plus forcément nul. Ce phénomène, désigné par le terme de cône d'influence, est évidemment une conséquence de l'utilisation de l'opérateur de dilatation et explique la structure en points de la plupart des images de scalogrammes (densité d'energie en temps-échelle). Avec la DWT, la notion de cône d'influence reste effective, quoique masquée par l'échantillonnage dyadique du plan : un coefficient des grandes échelles peut être non nul même si à sa verticale le signal est rigoureusement nul.

5.5.2 Localisation

On a déjà indiqué que l'analyse en ondelettes consiste en une décomposition reconstruction du signal à l'aide de briques élémentaires dans les domaines temporel et spectral simultanément limités. La qualité de cette localisation conjointe est mesurée par rapport à la borne du principe d'incertitude de Gabor-Heisenberg. L'ondelette de Morlet, dont l'enveloppe est gaussienne, atteint cette borne [42]. Les ondelettes B-spline tendent également très rapidement vers celle-ci. La localisation conjointe des ondelettes de Daubechies tend vers cette valeur idéale quand on augmente la taille de leur support temporel.

5.5.3 Compacité du support

On impose parfois à l'ondelette d'avoir un support fini, dans l'un ou l'autre domaine (temporel ou fréquentiel). Cela permet de garantir efficacité et précision dans le calcul effectif des coefficients et évite les problèmes de troncature.

5.5.4 Régularité

Dans l'analyse en ondelettes, on lit les coefficients comme le résultat de l'interaction (produit scalaire) entre le signal et une forme d'analyse. Dans la synthèse en ondelettes, on écrit le signal sous forme d'une somme pondérée de formes de synthèse ; dans les deux cas, il est naturel de souhaiter, pour faciliter l'interprétation, que l'ondelette présente certaines qualités de régularité (continue, dérivable,…). On peut, par exemple, souhaiter que les r-premières dérivées de l'ondelette soient bornées :

$$\forall n = 0,...,r,\ \left|\frac{d^n}{dt}\psi(t)\right| \le C_n \tag{2.90}$$

Ou bien imposer des contraintes plus sévères de décroissance exponentielle dans le domaine temporel des dérivées :

$$\forall n = 0,...,r,\ \left|\frac{d^n}{dt}\psi(t)\right| \le C_n\left(1+|t|\right)^{-m}, m \ge 0 \tag{2.91}$$

Satisfaire de telles contraintes est équivalent à spécifier le comportement de décroissance asymptotique de la transformée de Fourier $\Psi(\nu)$ de ψ. Dans le premier cas par exemple, cela équivaut à forcer $\left|\Psi(\nu)\right| \sim \nu^{-(r+1)}, \nu \to +\infty$. La régularité temporelle de l'ondelette est donc liée à

sa localisation fréquentielle. L'inverse est également vrai [52].

5.5.5 Nombre de moments nuls

Le nombre de moments nuls de l'ondelette est l'entier $\mathcal{N}$ tel que :

$$\forall k \in \{0,...,\mathcal{N}-1\}, \int_{\mathbb{R}} t^k \psi(t)\, dt = 0 \tag{2.92}$$

Cet entier joue un rôle essentiel dans l'utilisation de l'analyse en ondelettes. En effet, il implique que la transformée en ondelettes est aveugle aux portions douces d'un signal, c'est-à-dire celles qui peuvent être approchées par un comportement polynomial. Supposons qu'un signal $x(t)$ puisse être (localement au moins) approximé par un polynôme d'ordre p, $x_p(t)$, plus un reste, non polynomial, $x_r(t)$. Si $p < \mathcal{N}$, alors les coefficients de la partie polynomiale seront nuls $CWT_{x_p}(a,t) \equiv 0$, et les coefficients de la partie non polynomiale $CWT_{x_r}(a,t) \equiv CWT_x(a,t)$. Les coefficients de la décomposition en ondelettes ne capturent donc que les comportements les plus singuliers du signal. Ce nombre gouverne également la décroissance à l'origine de la transformée de Fourier de l'ondelette :

$$\forall k \in \{0,...,\mathcal{N}-1\}, \frac{d^k \Psi}{d\nu^k}(0) = 0 \tag{2.93}$$

5.5.6 Symétrie

On souhaite parfois que les ondelettes présentent des propriétés de symétrie temporelle afin qu'elles constituent des bancs de filtres n'induisant pas de distorsion de phase [55]. Notons que, dans le cas d'ondelettes issus de l'analyse multirésolution la symétrie par rapport à $t = 1/2$ et non à $t = 0$.

5.5.7 Orthogonalité

Enfin, la dernière propriété est à l'origine de la popularité des ondelettes qui offrent des bases non redondantes et orthogonales d'atomes temps-fréquence. Ce succès a conduit à restreindre abusivement l'utilisation du terme d'ondelettes aux ondelettes orthogonales. Notons que la propriété d'orthogonalité ne figure pas comme une nécessité dans la liste des préoccupations qui ont présidé à l'élaboration de la philosophie temps-échelle. Elle n'existe notamment pas (contrairement aux propriétés précédentes) dans les transformées continue ou redondante puisqu'elle est liée à la notion de base. Selon les cas, elle pourra se révéler un atout important ou un inconvénient rédhibitoire dans la mesure, par exemple, où cette propriété contraint forcément la forme temporelle de l'ondelette [42]. Il n'y a alors que peu de latitude pour la création d'ondelettes et le choix se restreint souvent à quelques motifs très particuliers (cf. Figure 2.6).

6 Conclusion

Notre étude est orientée vers le diagnostic de défauts. Nous rappelons que le diagnostic s'effectue en trois étapes : la détection, la localisation et l'identification des défauts. Nous nous intéressons, dans notre cas, à la détection des changements dans la structure d'un signal. Les changements de structure que l'on a considéré sont des changements de caractéristiques statistiques, des dérives lentes, ou bien des changements du contenu spectral. Le but est de détecter l'apparition de ces changements au cours du temps. L'apparition de ces types de

défauts ne peut pas être mise en évidence avec les mêmes méthodes de traitement du signal.

Dans ce chapitre, nous avons présenté différentes méthodes de traitement du signal. Dans un premier temps, nous avons vu quelques méthodes classiques d'analyse temporelle des signaux, notamment les tests statistiques tels que les saut de moyenne ou de variance suivant le paramètre considéré. D'autres techniques non présentées dans cet exposé (par exemple la corrélation) peuvent également être utilisés.

Si le signal a une structure périodique, alors la transformée de Fourier sera la méthode d'analyse la plus appropriée. En effet, elle projette le signal sur un espace de fonctions de type sinus ou cosinus. La transformée de Fourier offre uniquement une estimation de la puissance spectrale d'un signal. Dans le cas où on s'intéresse au contenu spectral d'un signal, cette méthode présente l'avantage d'un calcul précis et rapide à une résolution fréquentielle pré-choisie. Par contre, elle ne peut pas localiser en temps les changements se produisant dans le signal au cours du temps. Cette tâche peut être réalisé par l'analyse temps-fréquence.

La Transformée en ondelettes projette le signal sur un espace de fonctions localisées en temps et de préférence à support compact. Pour la transformée en ondelettes continue, le signal de départ est convolué avec l'ondelette d'analyse. Cette méthode s'avère plus adaptée que l'analyse temps-fréquence dans le cas où on souhaite avoir conjointement de "bonnes" localisations temporelle et fréquentielle. La Transformée en ondelette discrète est implantée en utilisant des filtres numériques passe-bas pour obtenir les approximations et passe-haut pour calculer les détails. Le filtrage du signal se fait récursivement. Par contre elle présente un inconvénient de covariance par translation. En conséquence, pour mieux détecter les ruptures dans un signal la transformée en ondelettes stationnaire est préférée.

La Transformée de Fourier à fenêtre glissante ne peut être utilisée que si on suppose le signal stationnaire sur la fenêtre d'analyse. La transformée en ondelettes offre la liberté de choisir la base de fonctions sur laquelle le signal est projeté, fonctions dont la durée temporelle varie. Elle permet de localiser en temps et en fréquence les changements dans la structure d'un signal.

Chapitre 3

Détection de Défauts par Génération de Symptômes flous

1 Introduction

Dans le milieu industriel, les opérations de maintenance préventive sont remplacées au fur et à mesure par des opérations de maintenance prédictive. Les installations sont surveillées en temps réel afin de prédire un dysfonctionnement avant que ses conséquences soient importantes. Les méthodes de traitement du signal analysent les signaux mesurés pour en extraire des informations pertinentes concernant l'évolution du système. Ces informations, combinées avec d'autres connaissances sur le bon ou le mauvais fonctionnement du système, sont utilisées dans la procédure de génération de symptômes pour la détection des défauts.

Nous proposons d'utiliser uniquement les informations concernant le bon fonctionnement du système, car une connaissance complète sur l'ensemble des cas de mauvais fonctionnement est difficile à acquérir. Cependant, la procédure de détection aura un caractère général. Nous avons vu précédemment que la Transformée de Fourier et la Transformée en Ondelettes sont deux méthodes d'extraction et de représentation des informations contenues dans les signaux mesurés. La procédure de génération de symptômes pour la détection construite à partir des informations ainsi obtenues est très flexible par rapport au type d'application et de défaut. Les choix de la longueur et du type de la fenêtre de pondération pour la FFT, de l'ondelette et du nombre de niveaux de décomposition pour la transformée en ondelettes stationnaire, adaptés à l'application, ainsi que la manière dont les comportements du système sont associés au bon ou au mauvais fonctionnement, constituent les degrés de liberté de la procédure proposée.

Ce chapitre est organisé de la façon suivante. La section 2 rappelle les notions de raisonnement approximatif, les opérateurs d'agrégation floue et la décision multicritères. Dans la section 3 nous décrirons le principe de la méthode proposée, puis sa transposition pour la transformée de Fourier à court terme et la transformée en ondelettes stationnaire (Sections 4 et 5). Nous concluons ensuite ce chapitre.

2 Raisonnement approximatif

Depuis que les ensembles flous ont été introduits par Zadeh en 1965, de nombreuses applications de ce "nouveau" concept d'ensemble ont été développées dans plusieurs domaines. En fait, toute application qui s'intéresse à la manipulation d'informations vagues ou incertaines peut faire appel à la théorie des ensembles flous. D'une façon générale, la représentation et la mise en oeuvre de connaissances vagues, imprécises et/ou incertaines par la théorie des ensembles flous sont connues sous le nom de raisonnement approximatif.

Dans cette section, nous introduisons brièvement les concepts de base de la logique floue pour la mise en œuvre d'un mécanisme de raisonnement approximatif, c'est-à-dire d'un mécanisme capable d'utiliser et de prendre en compte des connaissances imprécises ou incertaines afin de produire de nouvelles connaissances, tel que le raisonnement humain est capable de le faire. De la même manière que la logique classique, la logique floue se base sur la définition de propositions floues qui, à l'instar des propositions en logique classique, seront associées à une valeur de vérité.

2.1 Propositions floues

Une variable linguistique est représentée par un triplet (V, X_V, T_V) dans lequel [68] :

- V est le nom de la variable ;
- X_V est l'univers des valeurs prises par V ;
- T_V est l'ensemble de sous-ensembles flous de X_V, utilisés pour caractériser V.

2.1.1 Propositions floues élémentaires

Une proposition floue élémentaire est définie à partir d'une variable linguistique (V, X_V, T_V), elle représente une qualification de cette variable : "V est A". Dans le cadre de la logique classique, chaque proposition est attachée à une valeur de vérité. Une proposition p peut être soit vraie soit fausse. On peut alors noter t_p cette valeur de vérité : $t_p \in \{0,1\}$.

Dans le cadre de la logique floue, on retrouve une problématique équivalente qui est de déterminer des degrés de vérité pour des propositions floues. La valeur de vérité de la proposition floue élémentaire "V est A" est donnée par la fonction d'appartenance μ_A à A. Cette fois, nous possédons donc une fonction $\mu_A : X_V \to [0,1]$ qui représente le degré avec lequel chaque valeur de X_V est susceptible de confirmer la proposition.

2.1.2 Propositions floues générales

A partir de plusieurs propositions floues élémentaires, une proposition plus complexe peut être construite : une proposition floue générale est une composition de propositions floues élémentaires. Soient deux variables linguistiques (V, X_V, T_V) et (W, X_W, T_W) associées aux propositions floues élémentaires "V est A" et "W est B". Notons p_A la première proposition, et p_B la seconde. Comme on l'a vu, p_A est associée à la valeur de vérité donnée par la fonction d'appartenance μ_A à A, et p_B est associée à μ_B à B.

2.1.2.1 Négation de proposition floue

La plus simple des propositions floues est obtenue à partir d'une seule proposition, par exemple p_A. Elle consiste en la négation de cette proposition : "V est $\bar{A}$ ". Le degré de vérité $\mu_{\bar{A}}$, $\mu_{\bar{A}} = n(\mu_A)$ où n est une fonction de $[0,1]$ dans $[0,1]$ qui associe la sémantique de la négation $(n(vrai) = faux, \ldots)$. Il y a différentes façons de choisir n en fonction des propriétés que l'on veut voir respectées par cette fonction [68]. Comme dans le cas des ensembles, selon le choix qui sera fait des propriétés de la logique classique à conserver, une famille de fonctions sera possible. Ainsi, la fonction de négation la plus communément utilisée est $n(x) = 1 - x$ pour tout $x \in [0,1]$.

2.1.2.2 Combinaisons de propositions floues

Pour combiner deux propositions comme p_A et p_B, on utilise un opérateur de combinaison. Les opérateurs les plus courants sont la conjonction ("et", noté $\wedge$ en logique), la disjonction ("ou", noté $\vee$), et l'implication ("implique", noté $\rightarrow$).

Chacune de ces combinaisons permet donc de construire une nouvelle proposition p à laquelle est associée aussi un degré de vérité, dépendant des degrés de vérité de p_A, de p_B et de la composition mise en œuvre. Ainsi, de façon générale, si nous notons $\oplus$ l'opérateur de combinaison choisi, le degré de vérité de la proposition $p_{A \oplus B}$ obtenue par la combinaison de p_A et de p_B à l'aide de $\oplus$, $p_{A \oplus B} = p_A \oplus p_B$ (littéralement : "V est $A \oplus W$ est B ") dépendra donc de μ_A, de μ_B et de la sémantique que l'on souhaite associer à l'opérateur $\oplus$, ou qui est déjà associée à cet opérateur en logique classique si $\oplus$ est un opérateur comme $(\wedge, \vee, \ldots)$: $p_{A \oplus B}$ est vraie selon que p_A et p_B sont vraies.

Pour calculer le degré de vérité de $p_{A \oplus B}$, on agrège les degrés de vérité μ_A et μ_B pour obtenir un degré de vérité unique $\mu_{A \oplus B}$. D'autre part, $\mu_{A \oplus B}$ est défini sur $X_V \times X_W$ et est à valeurs dans $[0,1]$. On définit donc $\mu_{A \oplus B} : X_V \times X_W \rightarrow [0,1]$ telle que $\forall x \in X_V$ et $\forall y \in X_W$:

$$\mu_{A \oplus B}(x, y) = \Phi_{\oplus}\left(\mu_A(x), \mu_B(y)\right) \tag{3.1}$$

où $\Phi_{\oplus}$ est une fonction $[0,1] \times [0,1] \rightarrow [0,1]$ qui correspond à la sémantique de l'opérateur $\oplus$. Elle généralise la fonction de logique classique qui va de $\{0,1\} \times \{0,1\}$ vers $\{0,1\}$.

2.2 Opérateurs d'agrégation

Les opérations généralisées entre ensembles flous sont définies par des opérateurs flous qui agissent sur les fonctions d'appartenance. Des résultats différents sont obtenus suivant le choix des opérateurs utilisés ; chaque choix est adapté à un problème particulier. Divers opérateurs flous sont présentés dans la littérature [64] pour les opérations ensemblistes courantes. Dans cette section, nous présentons les opérations entre ensembles flous ainsi que quelques exemples parmi les opérateurs les plus courants. Les opérateurs flous sont des fonctions dont les arguments sont des valeurs d'appartenance.

2.2.1 Conjonction de propositions floues

La valeur de vérité de la conjonction ($\oplus$ est $\wedge$) de deux propositions floues doit généraliser la conjonction dans le cadre classique. $p_{A \wedge B}$ est vraie si et seulement si p_A et p_B

sont vraies ensemble. Dans le cadre flou, on peut par exemple considérer que la valeur de vérité de $p_{A\wedge B}$ doit être égale à la plus petite des valeurs de vérité de p_A et de p_B. On agrégera donc ces deux degrés à l'aide de l'opérateur $\Phi_\wedge$ pris comme étant le minimum $\mu_{A\wedge B} = \min(\mu_A, \mu_B)$. Plus généralement, on montre que les propriétés des t-normes en font d'excellents opérateurs $\Phi_\wedge$ [68].

2.2.2 Disjonction de propositions floues

De la même manière, la valeur de vérité de la disjonction ($\oplus$ est $\vee$) de deux propositions floues sera telle que $p_{A\vee B}$ est vraie si et seulement si p_A est vraie ou si p_B est vraie. On peut par exemple considérer que la valeur de vérité de $p_{A\vee B}$ doit être égale à la plus grande des valeurs de vérités de p_A et de p_B. On agrégera donc ces deux degrés à l'aide de l'opérateur $\Phi_\vee$ pris comme étant le maximum $\mu_{A\vee B} = \max(\mu_A, \mu_B)$. Plus généralement, on montre que les propriétés des t-conormes font d'excellents opérateurs $\Phi_\vee$ [68].

2.2.3 Implication entre propositions floues

La valeur de vérité de l'implication de deux propositions floues est beaucoup moins évidente à déterminer. En effet, comme dans le cas des autres opérateurs, il n'existe pas de solution unique, mais un ensemble de familles de fonctions possédant certaines des propriétés de l'implication classique ainsi que d'autres propriétés qui peuvent être requises. Ainsi, par exemple, on sait que l'implication classique $a \to b$ peut s'écrire de différentes façons à partir des opérateurs de conjonction, de disjonction, et de négation.

2.3 Décision multicritères

Quand le raisonnement diagnostic est exécuté, une décision symbolique doit être prise (notamment le processus est/n'est pas défectueux) à partir de données numériques (moyenne, écart type, spectre, etc.). Cette décision peut être basée sur des concepts vagues (semblables, ressembler à, etc.) [88]. Ceci justifie l'utilisation de la prise de décision floue pour la décision de diagnostic. L'outil flou constitue un outil simple pour créer l'interface entre les données numériques de bas niveau et la connaissance symbolique de haut niveau, parce qu'il explique la nature continue des variables manipulées dans une représentation symbolique. L'idée est de réaliser un diagnostic robuste en évitant les discontinuités soudaines qui surgissent si des limites précises sont introduites.

Quelques résultats classiques dans la prise de décision floue sont maintenant récapitulés. La logique floue permet une modélisation formelle de la prise de décision pour des conditions imprécises et incertaines afin de choisir une solution caractérisée par des points de vue partiels, appelés critères partiels. Dans un environnement connu, chaque décision $d \in D$ (où D est l'ensemble de décisions possibles) est évaluée à l'aide des objectifs partiels G_i. Ces objectifs sont considérés en tant qu'ensembles flous, définis par leur fonction d'appartenance qui mesure la compatibilité entre l'objectif partiel et la décision d par une série de valeurs $\left[c_1(d), c_1(d), ..., c_p(d)\right]$ où $c_i(d)$ mesure la décision d dans le sens du critère i. d est défini comme le sous-ensemble flou obtenu par l'agrégation des p critères partiels. Ainsi, la fonction d'appartenance μ_d est telle que :

$$\mu_d = \Phi_\oplus\left(c_1(d), c_1(d), ..., c_p(d)\right) \tag{3.2}$$

où $\Phi_{\oplus}$ est un opérateur flou à déterminer. Les conditions nécessaires sur l'opérateur $\Phi_{\oplus}$ sont [64] :

- $\Phi_{\oplus}$ est une fonction continue;
- $\Phi_{\oplus}(0,0,...,0)=0$ et $\Phi_{\oplus}(1,1,...,1)=1$;
- $\forall(u_i,v_i)\in[0,1]^2$ si $u_i \geq v_i$ alors $\Phi_{\oplus}(u_1,u_2,...,u_p)\geq\Phi_{\oplus}(v_1,v_2,...,v_p)$.

Trois attitudes principales de prise de décision peuvent être modélisées en utilisant la fonction d'agrégation conjonctive $(\wedge)$, la fonction d'agrégation de compromis et la fonction d'agrégation disjonctive $(\vee)$. Par exemple, supposant qu'à l'aide d'une température "haute" et un débit "bas", nous voulions analyser l'état "défectueux" d'un échangeur de chaleur.

i) L'opérateur $\Phi_{\oplus}$ peut exprimer que tous les critères doivent être satisfaits simultanément. L'état est défectueux si en même temps la température est "haute" et le débit est "bas". Par conséquent un axiome naturel est :

$$\forall(u_1,u_2,...,u_p),\ \Phi_{\oplus}(u_1,u_2,...,u_p)\leq\min_i(u_i),\ i=1:p \tag{3.3}$$

La décision globale ne peut pas être plus grande que la plus petite valeur des évaluations partielles. Ces opérateurs correspondent aux conjonctions. Tous les opérateurs conjonctifs sont couverts par la famille d'opérateur de Yager [68] :

$$Y_q(u_i)=1-\min\left(1,\left(\sum_{i=1}^{p}(1-u_i)^q\right)^{1/q}\right),\ \text{où } q\geq 0 \tag{3.4}$$

Notons que :

$$\begin{aligned}&\lim_{q\to\infty}Y_q(u_i)=\min_i(u_i),\ i=1:q\\&Y_1(u_i)=\max\left(0,\sum_{i=1}^{p}u_i-p+1\right)\end{aligned} \tag{3.5}$$

ii) L'opérateur $\Phi_{\oplus}$ peut exprimer qu'un seul critère est suffisant pour prendre la décision. L'état est défectueux si la température est "haute" ou le débit est "bas". Par conséquent un axiome naturel est :

$$\forall(u_1,u_2,...,u_p),\ \max_i(u_i)\leq\Phi_{\oplus}(u_1,u_2,...,u_p),\ i=1:p \tag{3.6}$$

La décision globale est déterminée par la valeur la plus élevée des évaluations partielles. Ces opérateurs sont des disjonctions :

$$CY_q(u_i)=\min\left(1,\left(\sum_{i=1}^{p}u_i^q\right)^{1/q}\right),\ \text{où } q>0 \tag{3.7}$$

Parmi les choix possibles, les plus communs sont :

$$\begin{aligned}&\lim_{q\to\infty}CY_q(u_i)=\max_i(u_i),\ i=1:q\\&CY_1(u_i)=\min\left(1,\sum_{i=1}^{p}u_i\right)\end{aligned} \tag{3.8}$$

iii) L'opérateur $\Phi_{\oplus}$ peut exprimer un compromis quand l'axiome suivant est vérifié :

$$\forall(u_1,u_2,...,u_p),\ \min_i(u_i)\leq\Phi_{\oplus}(u_1,u_2,...,u_p)\leq\max_i(u_i),\ i=1:p \tag{3.9}$$

Les opérateurs de compromis peuvent être exprimés par :

$$Ym_q(u_i) = \left(\frac{\sum_{i=1}^{p} u_i^q}{p} \right)^{1/q} \tag{3.10}$$

Quand q est égal à 1, (3.10) correspond à la moyenne arithmétique $Ym_1(u_i)$; quand q est égal à -1, (3.10) correspond à la moyenne harmonique $Ym_{-1}(u_i)$ et quand q est égal à 0, (3.10) correspond à la moyenne géométrique $Ym_0(u_i)$.

En conclusion, l'agrégation floue de critères permet de donner la sémantique adaptée au type de décision qui est souhaitée (nécessité, possibilité, compromis) et propose plusieurs manières de coder mathématiquement cette sémantique. Dans la prochaine section, les choix des opérateurs d'agrégation seront justifiés par la sémantique recherchée.

L'application de la prise de décision floue à l'évaluation de la confiance qui peut être attachée au modèle conçu avec le prototype de comportement des signaux de référence est analysée dans le paragraphe suivant. Quelques critères partiels permettant cette évaluation sont décrits. La modification en ligne de cet indicateur de confiance a priori employant d'autres critères selon le fonctionnement du procédé est alors expliquée.

Le problème de décision multicritère peut avoir plusieurs interprétations. L'une d'entre elles est la fusion d'informations imprécises, qui sera abordée dans les prochaines sections.

3 Détection de défauts par génération de symptômes flous

Les diverses approches de diagnostic à base de traitement de signal utilisent un prototype représenté par quelques caractéristiques particulières qui constituent une référence pour les futures analyses. Le but est alors de décider si le signal observé a le même comportement que le prototype. Ces méthodes sont fondées sur les bases théoriques bien maitrisées et les données utilisées comme prototype doivent être extrêmement fiables, ce qui n'est pas toujours le cas quand l'acquisition est faite dans un contexte industriel. L'objectif de ce travail est l'évaluation de la validité d'un prototype et son utilisation dans la décision de diagnostic. L'analyse des caractéristiques d'un ou plusieurs signaux de référence, considérés comme prototypes d'un comportement spécifique est utilisée pour établir un modèle qui représente par exemple le comportement normal du procédé ou un comportement spécifique d'un défaut. La validité de ce modèle est évaluée a priori. Pendant le diagnostic en ligne, la cohérence entre les attributs des données observées et le modèle est calculée grâce au calcul d'une distance appropriée. La validité du modèle peut être modifiée pour obtenir la validité du modèle a posteriori. La cohérence et la validité du modèle a posteriori conduisent au calcul de la décision de diagnostic. Le cadre général pour l'évaluation de la validité du modèle et la mesure de cohérence est inspiré par la formulation d'une décision floue multicritères. La méthodologie proposée est en particulier illustrée avec une transformée de Fourier rapide glissante et la transformée en ondelettes stationnaire.

Lors du diagnostic, on essaye généralement d'évaluer l'information contenue dans un signal représenté par un ou plusieurs attributs. Il existe différentes techniques de prise de décision lors de l'évaluation de ces attributs. Il peut s'agir d'une prise de décision basée sur un simple test logique (comparaison à un seuil), d'une prise de décision statistique (test de Bayes), de technique de reconnaissance des formes ou de technique de prise de décision basée sur la logique floue.

La principale difficulté lors de l'évaluation des attributs est que ces attributs varient souvent sous l'influence des phénomènes normaux. Lorsque le résultat de la comparaison à un seuil est booléen (alarme ou non), la décision peut être complètement différente suivant que l'attribut est juste au-dessous ou juste au-dessus du seuil qui a été fixé. L'utilisation des seuils nets conduit à des difficultés lors de la prise de décision : il n'est pas possible de représenter un doute. En revanche, l'utilisation de la logique floue [88] permet de construire une décision plus progressive. Le but de cette méthode est la détection des défauts par génération de symptômes flous, c'est-à-dire des indicateurs de présence de défaut progressive compris en 0 et 1.

3.1 Principe de génération de symptômes flous

Un test de cohérence sur un signal particulier détermine la cible (un composant ou un ensemble de composants) et dépend des hypothèses sous-jacentes au test. La fonction de test de cohérence a pour but de vérifier que le comportement observé est cohérent avec un comportement de référence. La fonction du test est dite valide si et seulement si une contrainte construite à partir de certaines caractéristiques élémentaires est satisfaite.

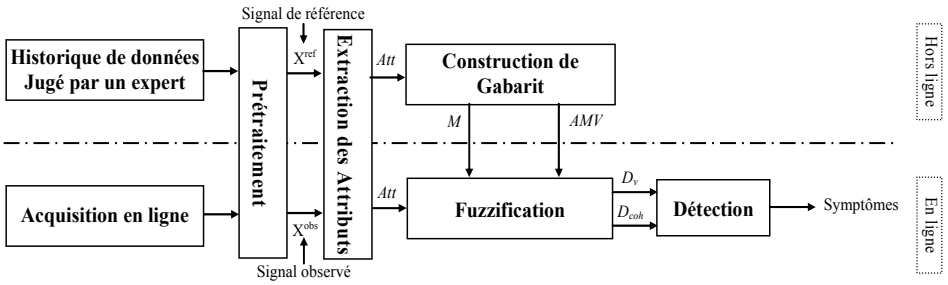


Figure 3.1 — Principe de détection de défauts par génération de symptômes flous

Soit un test de détection admettant une fonction de vérité à valeur soit dans $\{vrai, faux\}$ soit dans $[0,1]$ s'il s'agit d'un ensemble flou. Le test de détection est défini par :

$$Test : \begin{cases} D_{coh} = \mu_{coh} \\ D_v = \mu_v \end{cases} \tag{3.11}$$

où :

- D_{coh} représente le résultat du test de cohérence. Il s'agit d'évaluer la cohérence des attributs extraits en ligne d'un signal observé avec un modèle de référence caractérisant le comportement normal d'un composant (ou un ensemble de composants).

La vérité de cette décision n'est pas forcément booléenne. Un test de cohérence peut conduire à une décision incertaine dans le cas où la valeur de D_{coh} appartient à l'intervalle $[0,1]$. $D_{coh} = 1$

représente que le test détecte une incohérence d'une manière sûre et $D_{coh} = 0$ le cas contraire, c'est-à-dire, que le test confirme la cohérence des observations avec les valeurs de référence. $D_{coh} = 0.5$ représente le cas du doute dans la prise de décision (impossibilité de prendre une décision entre une cohérence ou une incohérence). Entre les deux bornes on peut avoir un test qui soit vrai à 20% ou par complémentarité faux à 80%. μ_{coh} représente le degré d'appartenance de la décision d'un test de détection à une cohérence. Ces décisions peuvent être obtenues grâce à 2 seuils sur les valeurs des attributs (cf. Figure 3.2).

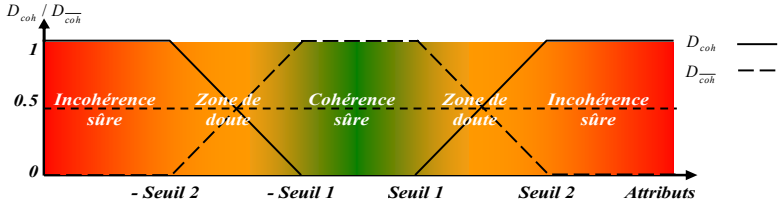


Figure 3.2 — La prise de décision incertaine

Le seuil 1 et le seuil 2 sont fixés de telle sorte que si les attributs produits par le test de détection se trouvent dans l'intervalle $[-Seuil1, Seuil1]$ cela représente qu'il n'y a pas d'alarme d'une manière sûre. Autrement dit, le comportement est cohérent avec celui de référence. Si les attributs sont supérieurs à ($Seuil2$) ou inférieurs à ($-Seuil2$) alors la décision prise par le test est la détection d'incohérence. Ce qui revient à dire que le comportement est incohérent d'une manière sûre. En revanche, si les attributs sont entre $[Seuil1, Seuil2]$ ou $[-Seuil2, -Seuil1]$ alors la décision est incertaine. En effet, cette zone représente la zone de doute où il n'est pas possible de décider entre une cohérence ou une incohérence.

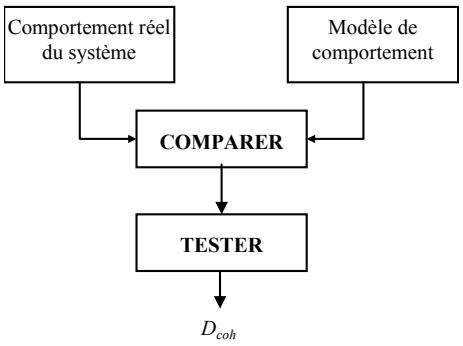


Figure 3.3 — Test de cohérence

Un modèle de comportement n'est pas universel ; il n'est généralement valide que dans certaines plages de fonctionnement, décrites par des conditions de validité. Un test de détection ne conduit pas toujours à des décisions sûres.

- D_v représente la fonction de validité du test de détection. Il s'agit de l'union des ensembles de contraintes de validités issues de certaines contraintes comportementales.

De même que le résultat du test de cohérence, la fonction de validité peut aussi prendre ses valeurs dans l'intervalle $[0,1]$.

En effet, si un test a pour résultat D_{coh} égal à 1 et valide alors un ou plusieurs composants sur lesquels il se base, sont forcément dans un état anormal. Si test a pour résultat D_{coh} nul et valide, cela n'implique pas forcément que les composants sont dans un état normal. En revanche, si le test de détection est invalide alors les hypothèses reliées au test peuvent être soit vraies, soit fausses : on ne peut conclure.

L'incertitude se caractérise par l'incapacité de savoir si un test de détection est vrai ou faux. Le test de détection peut conduire à des décisions comprises entre 0 et 1 où 1 représente une alarme sûre et 0 représente l'absence d'alarme d'une manière sûre. Nous allons montrer qu'il est possible d'appréhender l'incertitude du test de cohérence afin de l'exploiter durant la phase de génération des symptômes. La méthode proposée permet de fusionner la décision du test de cohérence avec celle de sa validité.

3.2 Validité du modèle

La validité de modèle a priori $AMV \in [0,1]$ mesure le niveau de confiance qui peut être attribué au modèle utilisé pour la décision de diagnostic, ce qui influence évidemment la confiance en la décision produite par le test de détection. $AMV = 1$ signifie que la confiance attachée à ce modèle est complète, tandis que $AMV = 0$ correspond à un modèle inapproprié. $AMV \in [0,1]$ est calculé hors ligne pendant la configuration du test de détection.

En fait, le calcul de la validité du modèle a priori tient compte de la confiance qui peut être mise dans le signal de référence x^{ref} et dans l'évaluation de ses attributs. Les hypothèses sur le signal de référence (par exemple stationnarité, normalité, etc.) sont prises afin de fournir des attributs fiables. Ces hypothèses constituent p suppositions AS_i qui sont évaluées avec un indicateur partiel $c_i = c_i(AS_i)$. Quand l'évaluation d'une supposition ne peut pas être faite (par exemple, il n'y a pas assez d'échantillons pour évaluer la stationnarité d'un signal), une valeur heuristique peut être choisie pour c_i.

Puis, la validité du modèle a priori AMV est calculée grâce à une agrégation floue de ces critères partiels :

$$AMV = \Phi_{\otimes}\left(c_1(AS_1),...,c_n(AS_n)\right) \tag{3.12}$$

Par conséquent, la fonction d'appartenance mesure la pertinence des hypothèses pour le signal de référence x^{ref}. L'opérateur "min" (3.3) a été choisi pour cette étude. Il permet de tenir compte du plus mauvais critère partiel, qui est une vue pessimiste de la pertinence d'hypothèse :

$$D_v = AMV = \min_i \left\{c_i = c_i(AS_1)\right\}_{i=1:p} \tag{3.13}$$

Un autre indicateur (noté $SV(x^{obs}) \in [0,1]$) pourrait être lié au capteur utilisé pour l'acquisition du signal. Même pendant un mode de fonctionnement sans défaut, les caractéristiques du capteur changent. Par exemple, le biais présent dans les mesures peut changer en raison du vieillissement du capteur ou encore le changement de sa sensibilité par des conditions externes.

Une autre situation classique dans un contexte industriel est que le capteur utilisé pour l'acquisition des signaux de référence x_i^{ref} a été remplacé par un autre, dont les caractéristiques sont légèrement différentes (par exemple la linéarité du capteur actuel est différente de la linéarité du capteur utilisé pendant l'acquisition de x_i^{ref}). Si ces changements sont quantifiables, la validité du modèle a priori AMV doit être modifiée pour être associée au signal actuellement enregistré, menant à une nouvelle validité de modèle PMV .

Notez que $SV\left(x^{obs}\right)$ peut être évaluée avec une agrégation floue des critères partiels. La validité du modèle a posteriori PMV peut être calculée en utilisant une agrégation floue de AMV et des autres critères partiels tels que par exemple, la $SV\left(x^{obs}\right)$:

$$D_v = PMV = \Phi_{\oplus}\left(SV\left(x^{obs}\right), AMV\right) \tag{3.14}$$

où $\Phi_{\oplus}$ est la fonction d'agrégation. Notez qu'un opérateur conjonctif semble tout à fait logique parce qu'il diminue la valeur de la validité du modèle :

$$D_v = PMV = \min\left(SV\left(x^{obs}\right), AMV\right) \tag{3.15}$$

Ces concepts sont transposés dans les sections suivantes où sera présenté le calcul de la validité du modèle pour des tests de détection à base de transformée de Fourier rapide glissante et transformée en ondelettes stationnaire.

3.3 Test de cohérence

Il est évident que quand les caractéristiques des signaux observés sont légèrement ou complètement différentes du modèle, la décision de diagnostic ne puisse pas avoir la même valeur. Une distance entre le modèle et les attributs du signal observé peut être définie. Dans un cas plus simple, le modèle peut être défini comme un simple intervalle. Par conséquent, le calcul de décision tiendra compte de la distance entre l'attribut extrait du signal observé et les limites supérieure et inférieure de l'intervalle. De cette manière, un indicateur progressif peut être généré, ce qui est fortement souhaitable dans un contexte de diagnostic. Par exemple, considérons $M_{att} = \left[att_{\min}, att_{\max}\right]$ l'intervalle d'appartenance de la valeur d'un attribut att_x , du signal observé x^{obs} pour que la décision de diagnostic soit " le comportement de x^{obs} est cohérent avec le comportement du signal de référence x^{ref} ". $att_{ref} = \left(att_{\max} + att_{\min}\right)/2$ est le centre de M_m , tandis que $L_m = att_{\max} - att_{\min}$ est sa largeur. Ainsi, des critères partiels peuvent être définis comme suit :

$$c(att^{obs}) = \begin{cases} \min\left(1, \dfrac{\left|att^{obs} - att_{\max}\right|}{L_m/2}\right) & \text{si} \quad att^{obs} > att_{\max} \\ \min\left(1, \dfrac{\left|att^{obs} - att_{\min}\right|}{L_m/2}\right) & \text{si} \quad att^{obs} < att_{\min} \\ 0 & \text{si} \quad att^{obs} \in M_{att} \end{cases} \tag{3.16}$$

Si le modèle se compose des attributs $Att = (att_1, att_2, ..., att_n)$, la décision peut être calculée à l'aide d'une agrégation floue des critères partiels (3.16) :

$$D_{coh} = \Phi_{\otimes}\left(c^1\left(att_1^{obs}\right), c^2\left(att_1^{obs}\right), ..., c^n\left(att_n^{obs}\right)\right) \tag{3.17}$$

Dans (3.17), $\Phi_{\oplus}$ est conforme avec les opérateurs rappelés dans la section 2.1. Par exemple, un opérateur "$\min$" veut révéler qu'une anomalie dans l'un des attributs est suffisante pour augmenter le degré d'appartenance à l'ensemble flou "les attributs sont incohérents".

La décision de diagnostic utilise la cohérence des attributs et la confiance dans le modèle pour évaluer la présence d'un défaut à l'aide d'une fonction d'agrégation pour générer la décision finale :

$$D_D = \Phi_{\oplus}\left(D_{coh}, D_v\right) \tag{3.18}$$

Dans (3.18), $\Phi_{\oplus}$ peut être un compromis, une conjonction ou une disjonction, exprimant respectivement le souhait d'avoir une décision sensible ou robuste.

3.4 Niveau de confiance d'un symptôme et prise de décision

Les procédures classiques de décision prennent en compte les incertitudes par l'intermédiaire des seuils (fixes) définis sur la différence entre comportement réel et celui de référence. La difficulté est alors de déterminer la valeur du seuil. Si elle est trop petite, il peut y avoir des fausses alarmes. Si le seuil est trop grand, on aboutit à des manques à la détection. Les tests de détection génèrent des symptômes qui contiennent des résultats bruts. En effet, chaque symptôme se base sur le résultat du test de cohérence, c'est-à-dire, la décision du test de cohérence noté D_{coh} et la décision de validité du modèle (du test) notée D_v.

La décision du test de cohérence D_{coh} ainsi que le décision de validité du test D_v peuvent prendre leurs valeurs dans l'intervalle $[0,1]$. Si le résultat d'un test de détection est égal à 1 ($D_{coh} = 1$) alors le test est incohérent sans qu'aucun doute ne soit permis, sinon, si le résultat du test de détection est égal à 0 ($D_{coh} = 0$), alors le test est obligatoirement cohérent. En revanche, si le résultat du test est égal à 0.5 ($D_{coh} = 0.5$) alors la décision est peu fiable.

La validité possède aussi un degré de vérité. En effet, un test de détection donné repose sur des modèles élémentaires avec lesquels il est construit. Ces modèles peuvent être valides ou non valides pour un ensemble de valeurs courantes de leurs variables. Cependant, ces résultats bruts ne sont pas directement utilisés pour la génération des symptômes. Tout d'abord, pour chaque test de détection, le résultat de la décision D_{coh} et la validité du test D_v sont fusionnés ensemble afin de pouvoir générer la décision finale du test de détection. Cette valeur calculée correspond au degré de vérité, noté par D_D du test de détection lié aux hypothèses correspondantes.

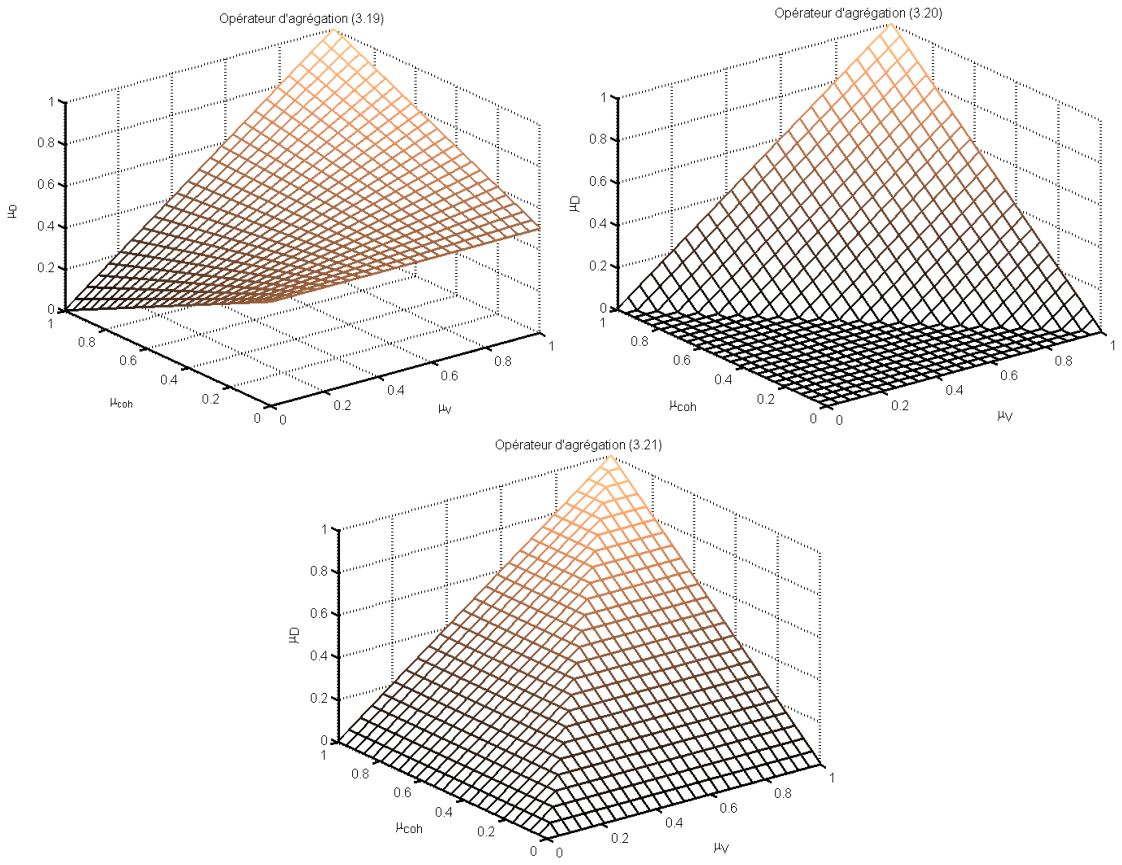


Figure 3.4 — Opérateurs d'agrégation pour la génération des symptômes

Les résultats partiels D_{coh} et D_v seront par la suite fusionnés pour générer un symptôme $D_D \in \{AL, OK\}$ où AL signifie qu'une alarme doit être déclenchée et OK son complément. Cette fusion a été faite en utilisant des opérateurs de conjonction. Par ailleurs, la fonction de fusion peut être donnée par plusieurs relations différentes (Figure 3.4) :

$$\mu_{D1} = \left(1 + \left(2\mu_{coh} - 1\right)\mu_v\right)/2 \tag{3.19}$$

$$\mu_{D2} = \max\left(\mu_{coh} + \mu_v - 1, 0\right) \tag{3.20}$$

$$\mu_{D3} = \min\left(\mu_{coh}, \mu_v\right) \tag{3.21}$$

La première relation (3.19), proposée dans [62] s'exprime littéralement ainsi : si la fonction d'appartenance de la validité du modèle est nulle alors ceci implique qu'on ne peut rien conclure et par conséquent la décision est égale à 0.5. La fonction de fusion de la relation (3.19) illustre que le niveau de confiance μ_D représenté sur la Figure 3.4 sera égal à la valeur de décision μ_{coh} si la validité est égale à 1. Quand la validité du modèle du test de détection décroît alors le niveau de confiance tend vers 0.5, c'est-à-dire vers plus d'incertitude. Néanmoins, il existe toute une famille d'opérateurs susceptibles d'être utilisés pour décrire une conjonction, mentionnons la t-norme de Lukasiewics (3.20) et la t-norme de Zadeh (3.21).

4 Génération de symptômes flous par STFT [80]

Le traitement des indicateurs spectraux permet de définir des indicateurs portant une information complémentaire à celle des indicateurs mesurés ou calculés dans le domaine temporel. La technique qui consiste à comparer l'évolution d'indicateurs spectraux n'est pas aussi simple qu'elle ne le paraît à première vue à cause par exemple de variations de vitesse de rotation des moteurs ou de charge. Même dans le cas d'installations entraînées par des moteurs à fréquence d'alimentation fixe, les variations de vitesse qui résultent des variations de charge entraînent de petites variations de la vitesse de rotation de la machine. Ainsi la position des composantes spectrales fréquence par fréquence peut ne pas être strictement fixe. De ce fait, toute comparaison de spectres par simple soustraction fréquence par fréquence devient impossible. C'est pourquoi une technique de surveillance consiste le plus souvent à comparer à un instant donné un spectre avec un spectre gabarit déterminés à partir d'un ou de plusieurs spectres de même nature dit spectres de référence, calculés dans des conditions identiques à un moment où la machine peut être considérée en bon état.

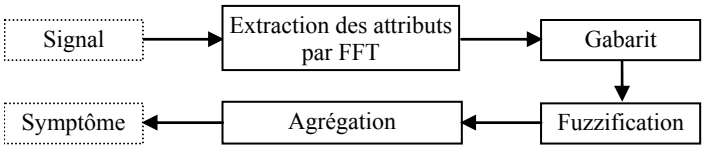


Figure 3.5 — Les étapes de génération de symptômes par STFT

La méthode proposée précédemment est illustrée dans cette section avec la transformée de Fourier à court terme. Un modèle comportemental est conçu à partir du spectre d'un ou de plusieurs signaux de référence, représentant le comportement normal du système. Pour cela le spectre du signal observé doit être cohérent avec le spectre du signal de référence. Ce concept de "cohérence" doivent être transformé en relations mathématiques facilement implantables et pouvant être utilisées en ligne. Par ailleurs, dans un contexte industriel, l'indicateur de détection du défaut ne devrait pas être booléen avec des réponses telles que le procédé est en défaut/le procédé est sain. L'utilisation d'un point de vue graduel permet de focaliser l'attention sur un composant avant qu'un défaut ne soit complètement installé. Rappelons que plusieurs concepts souvent utilisés pour faire un diagnostic sont vagues. Mesurer la cohérence entre les caractéristiques d'un signal observé et celles d'un modèle de référence est un autre concept qui doit être graduel afin de pouvoir détecter les défauts progressifs. La décision que l'on va générer utilise le résultat de la validité du modèle.

4.1 Extraction des attributs

De nombreux signaux présentent des caractéristiques non stationnaires, par exemple le changement du contenu fréquentiel. La transformée de Fourier à court terme a été choisie pour sa popularité industrielle due à son interprétation simple. Pour un signal discret $x[k]$, une transformée de Fourier rapide est calculée à chaque instant par :

$$STFT_x[k,n] = \sum_{l=0}^{N-1} x[k+l]\phi[l]e^{-i\omega_n l} \tag{3.22}$$

où ϕ est la fenêtre d'analyse (par exemple rectangulaire, Hanning, Hamming, etc.), N est la longueur du signal discret $x[k]$, $\omega_n = 2\pi n/N$ où $n = 0:N-1$ est le nombre discret d'échantillons de la transformée de Fourier qui correspond à la fréquence (en Hz) :

$$\nu_n = \nu_e \frac{\omega_n}{2\pi} \tag{3.23}$$

avec ν_e la fréquence d'échantillonnage. Rappelons qu'une bonne résolution en temps exige une fenêtre d'analyse ϕ "courte", tandis qu'une bonne résolution en fréquence implique une fenêtre "longue". Généralement, N est une puissance de 2. À chaque instant $t_k = kT_e$, la transformée de Fourier discrète du signal fenêtré est calculée, et le spectre de puissance résultant est considéré pour la procédure de détection, avec $n = 0:N/2$:

$$PS[k,n] = \frac{1}{N}\left|STFT_x[k,n]\right|^2 \tag{3.24}$$

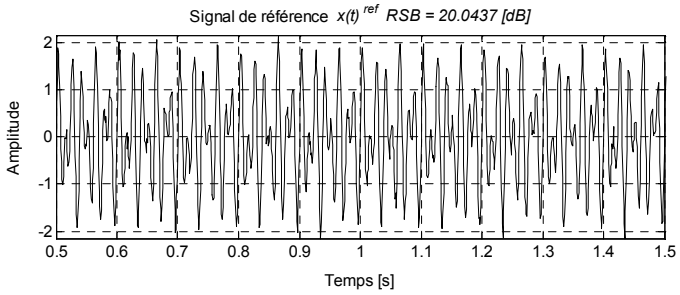


Figure 3.6 — Signal de référence x_1^{ref}

Un exemple académique illustrera les différentes étapes du calcul du Gabarit (modèle comportemental) et de l'évaluation de sa validité a priori. Trois signaux de référence x_i^{ref}, $i = 1:3$, sont simulés avec :

$$x_i^{ref}[t_k] = \sin[2\pi\nu_1 kT_e] + \sin[2\pi\nu_2 kT_e] + b[kT_e] \tag{3.25}$$

où, $T_e = 1/\nu_e = 10^{-3}$ [s], $k = 0:2047$, ν_1 = 50 [Hz], ν_2 = 80 [Hz], et b est un bruit blanc gaussien de moyenne nulle et de variance telle que le Rapport Signal sur Bruit $RSB \approx 20[dB]$.
Un exemple de x_i^{ref} est donné sur la Figure 3.6.

4.2 Construction du modèle (gabarit)

Un gabarit est une fonction "en escalier" représentant, à un facteur d'amplitude près, une enveloppe plus ou moins simplifiée du spectre de référence dont il est issu. La méthode de surveillance proposée permet de déterminer automatiquement à partir d'un spectre de référence, un gabarit dont la forme dépend de trois paramètres fixés par l'utilisateur :

- les pourcentages d'évolution relative fixant les hauteurs des "marches" ;
- le seuil en dessous duquel aucune composante du spectre de référence n'est prise en compte dans la détermination de la hauteur des "marches" ;
- la largeur des marches exprimée en multiple entier de la résolution d'analyse $\Delta\nu$.

Ce dernier paramètre qui représente plusieurs fois la résolution d'analyse $\Delta\nu$ limite considérablement les effets induits par des petites variations autour de la fréquence d'interet. Cependant, le choix de la largeur des marches conditionne la forme du gabarit de surveillance. Une forme trop simple peut ne pas signaler par une alarme l'apparition d'une composante spectrale. Un gabarit constitué d'un nombre trop important de marches risque de générer des fausses alarmes au moindre décalage d'un pic de fréquence par rapport au pic de référence. Ce type de décalage peut par exemple intervenir lorsque la vitesse de rotation d'un moteur n'est pas parfaitement stabilisée, sans pour autant que cela soit synonyme de l'apparition d'un défaut.

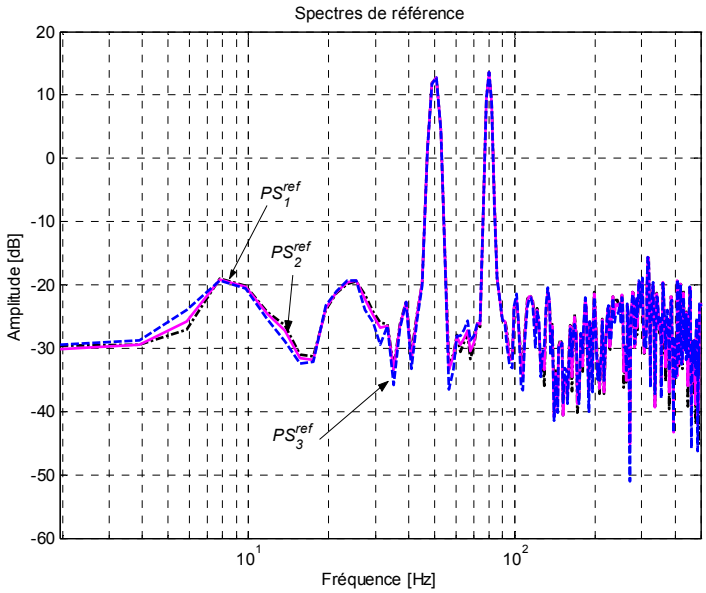


Figure 3.7 — Spectres de puissance de référence

Enfin, la hauteur de chaque marche doit être compatible avec les amplitudes maximales admissibles des composantes contenues dans la bande fréquentielle qu'elle délimite. L'évolution relative maximale admissible dépend donc à la fois de l'amplitude initiale dans le spectre de référence, et de l'amplitude maximale admissible en l'abscence de défaut, pour une composante fréquentielle donnée. Les spectres gabarits ne peuvent pas être définis à partir d'un seul facteur multiplicatif, ce facteur multiplicatif dépendant de la bande fréquentielle considérée.

Pour les signaux de référence x_i^{ref}, les spectres de puissance PS_i^{ref} sont d'abord calculés. La Figure 3.7 montre les spectres de puissance pour les signaux définis par (3.25) avec $N = 512$ et une fenêtre ϕ de Blackman. Puis, les valeurs maximales et minimales de ces spectres sont calculées pour chaque fréquence $\nu_n = n\nu_e / N$ afin de générer les enveloppes "min" et "max" (3.26) :

$$\forall \nu_n \begin{cases} S_{\min}[\nu_n] = \min\left\{PS_i^{ref}[\nu_n]\right\} \\ S_{\max}[\nu_n] = \max\left\{PS_i^{ref}[\nu_n]\right\} \end{cases} \tag{3.26}$$

Les enveloppes $S_{\min}$ et $S_{\max}$ représentent le comportement de référence du signal (cf. Figure 3.8).

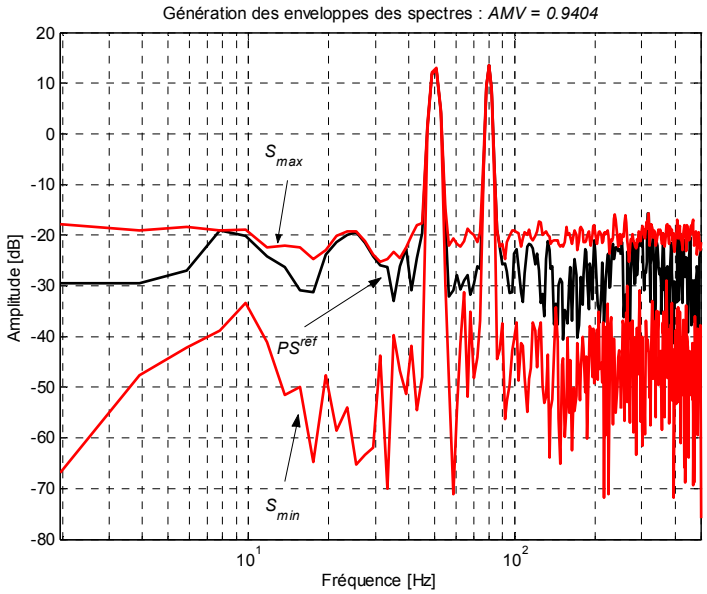


Figure 3.8 — Génération des enveloppes

La construction du modèle est effectuée en trois étapes ; chacune d'entre elles implique quelques paramètres à régler. Le but de la première étape est de définir les pourcentages relatifs fixant les hauteurs des marches. Deux coefficients de pourcentage $\{\alpha_{\inf}, \alpha_{\sup}\}$ permettent de définir une sous-estimation et une sur-estimation des hauteurs des marches à partir des enveloppes :

$$\forall \nu_n \quad \begin{cases} P^1_{\inf}[\nu_n] = (1-\alpha_{\inf}) S_{\min}[\nu_n] \\ P^1_{\sup}[\nu_n] = (1+\alpha_{\sup}) S_{\max}[\nu_n] \end{cases} \tag{3.27}$$

Pour le signal de référence (3.25) $\alpha_{\sup} = 0.2$ et $\alpha_{\inf} = 1$ ce qui conduit à $P^1_{\inf} = 0$ pour toutes les fréquences ν_n. On a reporté sur la la Figure 3.9 le tracé de $P^1_{\sup}$.

Remarque 1 : les coefficients $\{\alpha_{\inf}, \alpha_{\sup}\}$ peuvent ne pas être identiques pour toutes les raies du spectre. Ils peuvent être changés, si on attribue plus ou moins d'importance à des raies spécifiques qui peuvent caractériser par exemple des défauts spécifiques.

Remarque 2 : le choix de $\alpha_{\inf} = 1$, est fait pour ne pas utiliser le gabarit inférieur, ce gabarit inférieur permet par exemple de surveiller des attributs caractéristiques du point de fonctionnement, du système (par exemple : une fréquence correspondant à la vitesse de rotation d'une machine). Il peut être utilisé pour réévaluer la validité du modèle en-ligne, ce point n'a pas été abordé dans cette étude.

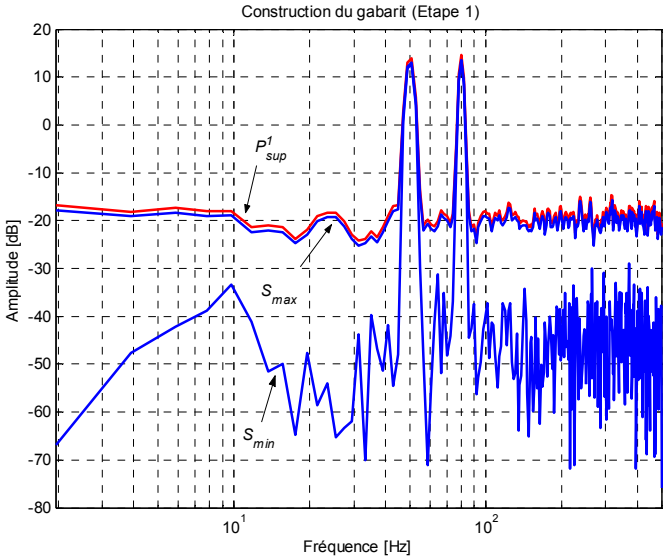


Figure 3.9 — Construction du gabarit – Etape 1

La deuxième étape définit la largeur des marches par l'introduction d'un facteur multiplicatif de la résolution d'analyse q, qui prend en compte l'incertitude sur le contenu fréquentiel liée à la résolution du spectre. Le gabarit résultant devient :

$$\forall \nu_n \quad \begin{cases} P^2_{\inf}[\nu_n] = \min\left\{P^1_{\inf}[\nu_m] \ \text{ pour } \ \nu_{n-q} \le \nu_m \le \nu_{n+q}\right\} \\ P^2_{\sup}[\nu_n] = \max\left\{P^1_{\sup}[\nu_m] \ \text{ pour } \ \nu_{n-q} \le \nu_m \le \nu_{n+q}\right\} \end{cases} \tag{3.28}$$

La Figure 3.10 illustre cette deuxième étape avec $q = 2$ pour les signaux de référence donnés par (3.25). Notons que $P^2_{\inf} = 0$ parce que $P^1_{\inf} = 0$.

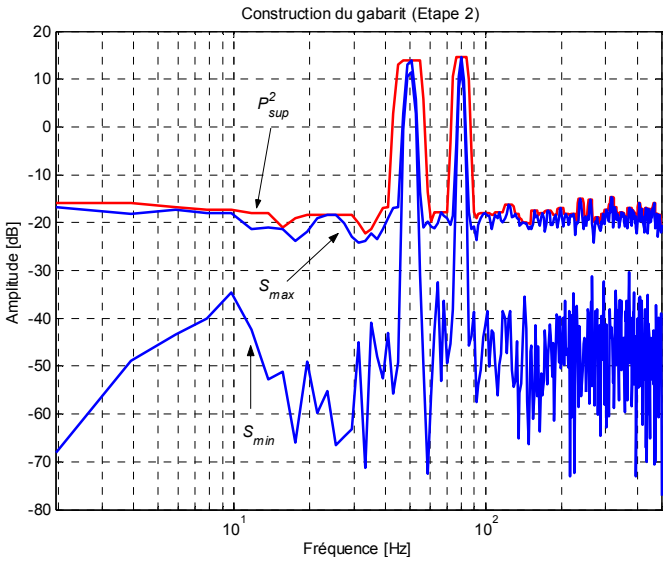


Figure 3.10 — Construction du gabarit – Etape 2

Dans la troisième étape, on introduit le seuil $P_{\min}$, en dessous duquel aucune composante du spectre de référence n'est prise en compte dans la détermination de la hauteur des marches. Il représente une valeur minimale nécessaire pour le gabarit afin de prendre en compte uniquement le contenu fréquentiel d'intérêt. Par conséquent, toutes les fréquences correspondant au contenu fréquentiel supposé sans intérêt ou à la bande fréquentielle correspondant au bruit ne sont pas prises en compte dans la procédure de détection.

Ainsi, la construction du gabarit s'achéve par :

$$\forall \nu_n \begin{cases} P_{\inf}[\nu_n] = P_{\inf}^2[\nu_n] \\ P_{\sup}[\nu_n] = \begin{cases} P_{\sup}^2[\nu_n] \ \text{ pour } \ P_{\sup}^2[\nu_n] > P_{\min} \\ P_{\min} \ \text{ pour } \ P_{\sup}^2[\nu_n] \le P_{\min} \end{cases} \end{cases} \tag{3.29}$$

Pour les signaux académiques (3.25), le gabarit est donné sur la Figure 3.11 avec $P_{\min} = -13$ dB.

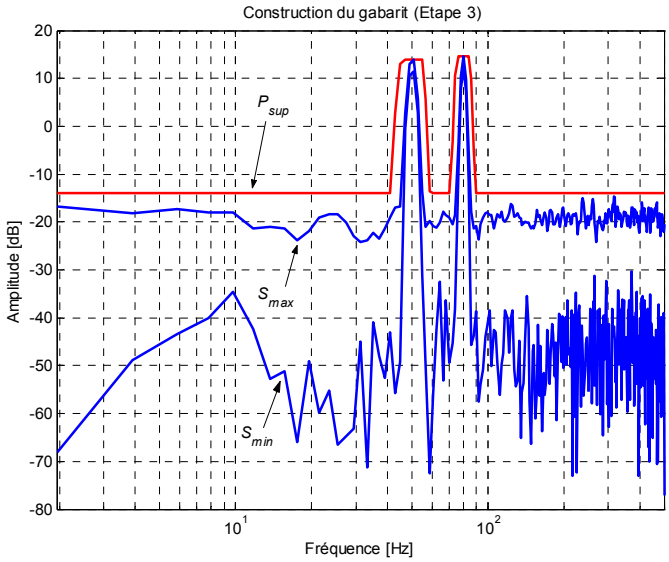


Figure 3.11 — Construction du gabarit – Etape 3

4.3 Validité du modèle (gabarit)

Comme on l'a vu précédemment, la validité du modèle (dans ce cas le gabarit) tient compte de la confiance qui peut être mise dans les signaux de référence x_i^{ref}, $i = 1:K$, et de l'évaluation de leurs attributs. Pour ce test de détection à base de STFT, les attributs extraits de x_i^{ref} sont les spectres de puissance. Ici, $D_v = AMV$ (*A Priori Model Validity*) mesure la cohérence entre les différents spectres de puissance de référence. Plusieurs approches peuvent être suivies. Dans cette étude, la distance entre les enveloppes $S_{\max}$ et $S_{\min}$ définie par (3.26) est évaluée. En fait, cette distance est normalisée afin d'obtenir une $AMV \in [0,1]$. Plusieurs indicateurs ont été comparés sur des exemples académiques et des données industrielles. Pour l'applications du chapitre 4 (et pour des raisons de robustesse), la formule suivante a été choisie :

$$D_v = AMV = 1 - \min\left\{1, \left\|S_{\max} - S_{\min}\right\|_2 \Big/ \left\|S_{\max}\right\|_2\right\} \ \text{ avec } \ \left\|S_{\max}\right\|_2 \neq 0 \tag{3.30}$$

4.4 Test de cohérence

La distance entre le modèle du spectre $\{P_{\inf}, P_{\sup}\}$ et le spectre de puissance PS^{obs} du signal observé x^{obs} est évaluée avec la règle décrite ci-dessous (cf. Figure 3.12). Supposons d'abord que $P_{\inf} \neq 0$, c'est-à-dire $\alpha_{\inf} \neq 1$. Pour toute fréquence ν_n, $PS^{obs}[\nu_n]$ est comparé à $P_{\inf}[\nu_n]$ et $P_{\sup}[\nu_n]$ grâce à la règle floue représentée sur la Figure 3.12.

4.4.1 Fuzzification

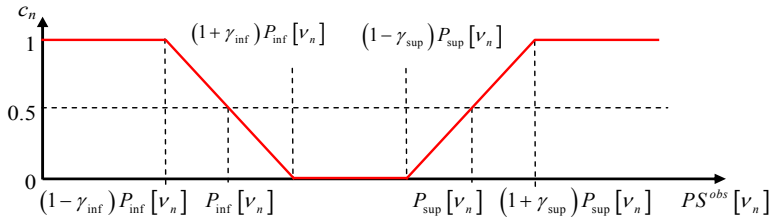


Figure 3.12 — Règle de fuzzification

Plus formellement, le critère partiel c_n est évalué pour chaque fréquence ν_n avec :

$$\begin{cases} PS^{obs}[\nu_n] < (1-\gamma_{\inf})P_{\inf}[\nu_n] & \Rightarrow c_n = 1 \\ (1-\gamma_{\inf})P_{\inf}[\nu_n] \le PS^{obs}[\nu_n] \le (1+\gamma_{\inf})P_{\inf}[\nu_n] & \Rightarrow c_n = \dfrac{(1+\gamma_{\inf})P_{\inf}[\nu_n] - PS^{obs}[\nu_n]}{2\gamma_{\inf}P_{\inf}[\nu_n]} \\ (1+\gamma_{\inf})P_{\inf}[\nu_n] < PS^{obs}[\nu_n] \le (1-\gamma_{\sup})P_{\sup}[\nu_n] & \Rightarrow c_n = 0 \\ (1-\gamma_{\sup})P_{\sup}[\nu_n] \le PS^{obs}[\nu_n] \le (1+\gamma_{\sup})P_{\sup}[\nu_n] & \Rightarrow c_n = \dfrac{PS^{obs}[\nu_n] - (1-\gamma_{\sup})P_{\sup}[\nu_n]}{2\gamma_{\sup}P_{\sup}[\nu_n]} \\ PS^{obs}[\nu_n] > (1+\gamma_{\sup})P_{\sup}[\nu_n] & \Rightarrow c_n = 1 \end{cases} \quad (3.31)$$

où $\gamma_{\inf}$ et $\gamma_{\sup}$ définissent la fonction d'appartenance. Cette fonction est conçue afin d'avoir $c_n = 0.5$ quand $PS^{obs}[\nu_n] = P_{\inf}[\nu_n]$ ou $PS^{obs}[\nu_n] = P_{\sup}[\nu_n]$. $c_n = 0$ signifie que "$PS^{obs}[\nu_n]$ est strictement à l'intérieur du modèle" tandis que $c_n = 1$ signifie que "$PS^{obs}[\nu_n]$ est en dehors et loin du modèle". Notons que $\gamma_{\inf} = \alpha_{\inf}/(1+\alpha_{\inf})$ et $\gamma_{\sup} = \alpha_{\sup}/(1+\alpha_{\sup})$ conduit à une fonction d'appartenance triangulaire avec :

$$(1+\gamma_{\inf})P_{\inf}[\nu_n] = (1-\gamma_{\sup})P_{\sup}[\nu_n] \quad (3.32)$$

Notons que avec $P_{\inf}[\nu_n] = 0$, (3.31) devient une fonction d'appartenance plus simple.

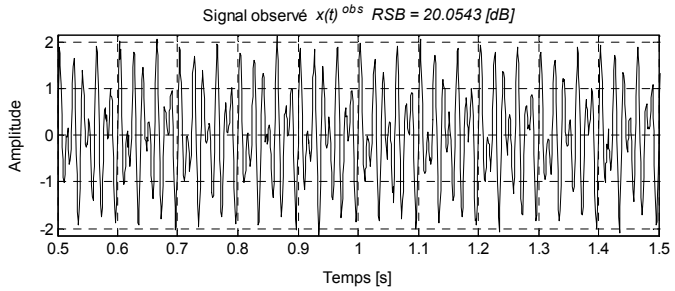


Figure 3.13 — Signal avec défaut x^{obs}

Sur la Figure 3.13 on donne le signal observé pour l'exemple académique, le signal x^{obs} est généré par :

$$x^{obs}[t_k] = \sin[2\pi\nu_1 kT_e] + \sin[2\pi\nu_2 kT_e] + b[kT_e] + 0.25\sin[2\pi\nu_3 kT_e].u[(k-1000)T_e] \quad (3.33)$$

où, $T_e = 1/\nu_e = 10^{-3}[\text{s}]$, $k = 0:2047$, ν_1 = 50 [Hz], ν_2 = 80 [Hz], ν_3 = 40 [Hz] et b est un bruit blanc gaussien de moyenne nulle et de variance telle que le Rapport Signal sur Bruit $RSB \approx 20[dB]$.

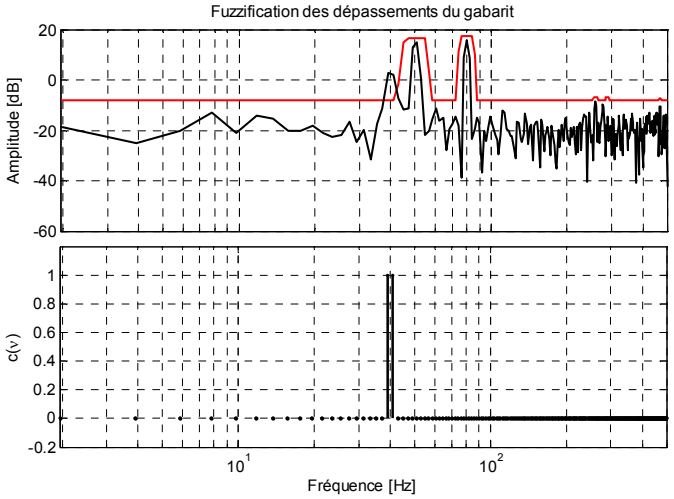


Figure 3.14 — Fuzzification des dépassements du gabarit pour x^{obs} à l'instant $t_k = 1.2[\text{s}]$

Le défaut simulé est une fréquence additif qui apparaît à partir de $t_k = 1[\text{s}]$, on voit sur la Figure 3.14 le spectre du signal observé qui dépasse le gabarit à la fréquence 40 Hz et le critère partiel correspondant à l'instant $t_k = 1.2[\text{s}]$.

4.4.2 Agrégation

Le résultat du test de cohérence D_{coh} est calculé par une agrégation floue des critères partiels c_n :

$$D_{coh} = \Phi_{\otimes}(c_n) \quad (3.34)$$

En fait, un poids $w_n \in [0,1]$ peut être défini pour chaque fréquence ν_n afin d'attacher plus d'importance pour des bandes de fréquence particulières (qui peuvent caractériser la présence d'un défaut par exemple). Ainsi, la fonction d'agrégation pour générer le résultat du test de cohérence devient :

$$D_{coh} = \Phi_{\otimes}(w_n c_n) \quad (3.35)$$

Dans cette méthode l'opérateur disjonctif "max" a été choisi pour une simple raison de robustesse. Par conséquent, le critère partiel $w_n c_n$ le plus élevé sera pris en compte. Ceci correspond à la fréquence ν_n qui donne la plus grande distance entre $PS^{obs}[\nu_n]$ et le modèle (Gabarit). Sur la Figure 3.15, on a tracé le résultat du test de cohérence, au début le résultat affiche 0.5 (synonyme de non décision) à cause d'un nombre insufisant d'échantillons pour calculer le premier spectre. Durant la première partie du signal où il ne se passe rien la $D_{coh} = 0$,

elle passe à 1 dès l'apparition de la fréquence ν_3 = 40 [Hz], mais avec un retard de 1.2[s] lié à longueur de la fenetre d'analyse (ici 512 points).

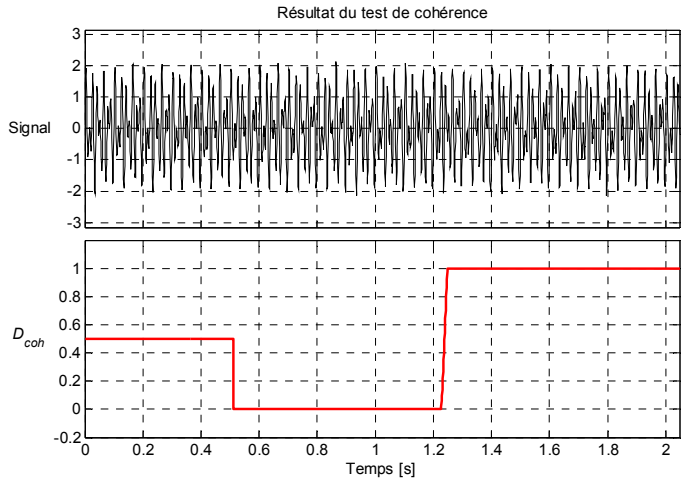


Figure 3.15 — Résultat du test de cohérence pour x^{obs}

4.5 Génération des symptômes

Afin de détecter les défauts progressifs plutôt que les changements brusques, la décision de détection basée sur l'évaluation de la cohérence (entre le gabarit et le le spectre observé) et la validité du modèle est donnée par agrégation floue :

$$D_D = \Phi_{\oplus}\left(D_{coh}, D_v\right) \tag{3.36}$$

où $\Phi_{\oplus}$ est un opérateur conjonctifs présentés précédemment (3.19) à (3.21).

Sur la Figure 3.16, on a reporté le signal x^{obs} ainsi que les décisions de détection $D1_D$, $D2_D$ et $D3_D$ calculées respectivement par (3.19), (3.20) et (3.21).

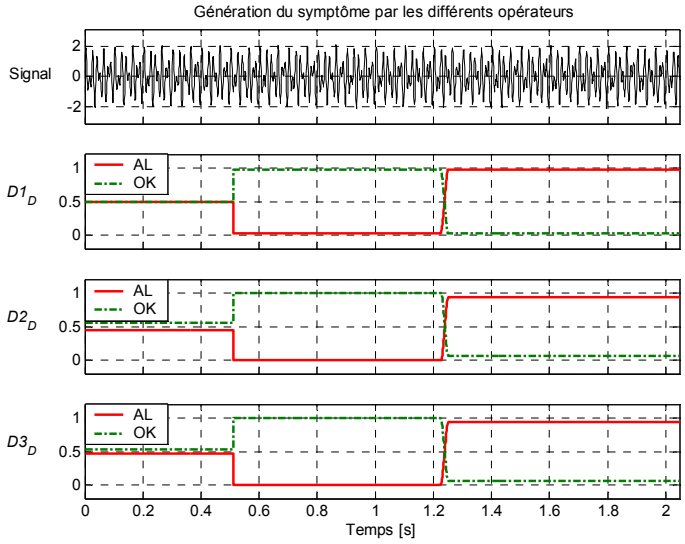


Figure 3.16 — Génération des symptômes par les différents opérateurs pour x^{obs}

5 Génération de symptômes flous par SWT [81]

Dans cette section, l'analyse par ondelettes est utilisée pour détecter des défauts particuliers. La méthode de détection est basée sur la transformée en ondelettes stationnaire. Les coefficients d'ondelette permettent d'analyser les changements dans un signal à différentes échelles. Les coefficients d'ondelettes seront par la suite fuzzifiés, afin d'obtenir des critères partiels, relatifs au comportement des coefficients sur différentes échelles. L'agrégation floue de ces critères partiels donne le résultat de détection. La procédure dépend de différents paramètres qui doivent être réglés. Plusieurs opérateurs d'agrégation sont appliqués à des exemples académiques pour illustrer les étapes et les résultats de la méthode.

L'occurrence d'un défaut est supposée être indiquée par une singularité dans le signal [70]. La méthode de détection proposée a pour but de détecter ces aspects de singularité. Elle se compose de quatre étapes (cf. Figure 3.17). La première étape transforme le signal en coefficients d'ondelettes, en décomposant le signal sur différentes échelles. Cette étape permet également de caractériser les fréquences locales dans le signal analysé. La deuxième étape correspond à la discrimination entre les coefficients qui caractérisent l'existence, ou la non-existance d'un phénomène singulier. Cette étape peut être faite à l'aide d'un seuillage des coefficients d'ondelettes. La troisième étape est celle de la fuzzification. En fait, elle permet de normaliser entre 0et 1 la distance entre le coefficient d'ondelettes et son seuil. Afin de donner un indicateur unique, mais également progressif, la quatrième étape considère les divers coefficients fuzzifiésés en tant que critères partiels et différents opérateurs d'aggrégation sont considérés.

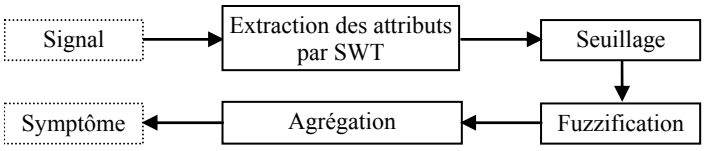


Figure 3.17 — Les étapes de génération de symptômes par SWT

5.1 Extraction des attributs

Les analyses spectrales traditionnelles fournissent une bonne description des signaux stationnaires. Malheureusement, elles ont des inconvénients. Le premier est que la transformée de Fourier ne peut pas bien analyser et représenter un signal avec des caractéristiques non périodiques (par exemple, un signal transitoire). Une autre difficulté est son incapacité à fournir des informations au sujet de l'évolution du contenu fréquentiel en fonction du temps. Dans ce cas-ci, des méthodes de type temps-fréquence doivent être utilisées au lieu de l'analyse spectrale traditionnelle [44][45]. La difficulté principale de ces méthodes est leur résolution temps-fréquence constante selon le principe d'incertitude de Heisenberg-Gabor. Pour obtenir des résolutions temporelle et fréquentielle variables, on utilise des méthodes temps-échelle telles que la transformée en ondelettes (WT) qui a été présentée dans le chapitre 2. Rappelons que dans ce travail, on s'est intéréssé à la transformée en ondelettes stationnaire.

Son principal avantage [57][59] est sa propriété de covariance par translation dans le temps : les coefficients d'un signal retardé sont une version retardée des coefficients du signal non retardé.

Pour illustrer les étapes de la méthode de détection à base de SWT et évaluer les différents opérateurs d'agrégation floue pour la génération de symptôme, un exemple académique sera utilisé par la suite. On effectura une décomposition en ondelettes sur $J = 5$ niveaux. Le signal $x^{ref}[t_k]$ est considéré sans défaut (3.25). Il est simulé par (voir Figure 3.6) :

$$x_1^{ref}[t_k] = \sin[2\pi f_1 k T_e] + \sin[2\pi f_2 k T_e] + b[k T_e]$$

où $T_e = 10^{-3}[\text{s}]$, $k = 0:2047$, ν_1 = 50 [Hz], ν_2 = 80 [Hz] et b est un bruit blanc gaussien de moyenne nulle et de variance telle que le Rapport Signal sur Bruit $RSB \approx 20[dB]$.

Sur la Figure 3.18, on a représenté le signal x^{ref} ainsi que sa décomposition en ondelettes stationnaire avec une ondelette de *Haar*.

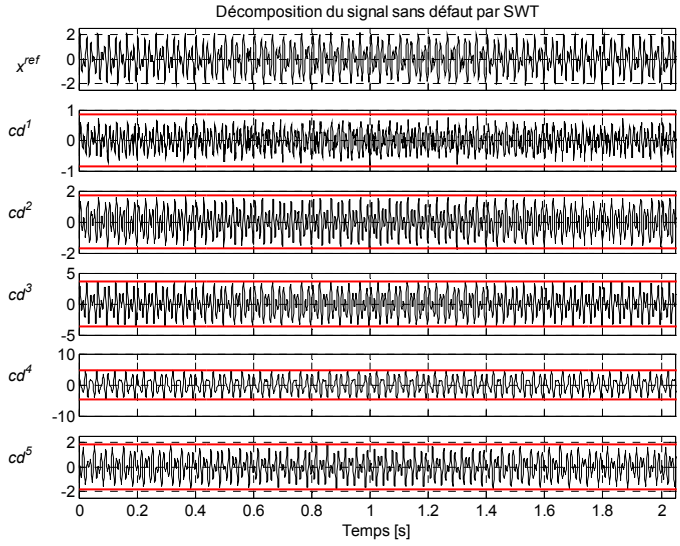


Figure 3.18 — Décomposition par SWT de x^{ref} et calcul des seuils

5.2 Seuillage des coefficients d'ondelettes

L'occurrence de singularité dans le comportement d'un signal est indiquée par la taille des coefficients d'ondelette [54][70]. Dans un contexte de débruitage, l'objectif est de trouver quand les données bruitées changent de comportement, quelle que soit l'hypothèse au sujet de la forme paramétrique de la fonction g qui contient ce changement :

$$x[t_k] = g[t_k] + b[t_k], \quad k = 1:n \tag{3.37}$$

Dans (3.37), b est habituellement un bruit normalement distribué centré de moyenne nulle et de variance σ^2. Le but est de distinguer entre les coefficients d'ondelettes contenant des informations sur la singularité, des autres qui décrivent le comportement normal du signal. Donoho [72] propose d'extraire ces coefficients significatifs par un seuillage doux :

$$\delta_k^j = \begin{cases} cd_k^j + \lambda^j & cd_k^j < -\lambda^j \\ 0 & 0 \le \left| cd_k^j \right| \le \lambda^j \\ cd_k^j - \lambda^j & cd_k^j > \lambda^j \end{cases} \tag{3.38}$$

où δ_k^j sont les coefficients seuillés du niveau de décomposition j, cd_k^j sont les coefficients de détails donnés par (2.80) et λ^j est la valeur du seuil. Ainsi, les coefficients au-dessous de leur seuil sont mis à zéro (en considérant qu'ils représentent le comportement normal). Tout les coefficients qui excèdent le seuil sont supposés indiquer l'occurrence d'un comportement anormal dans le signal (cf. Figure 3.19). Notons que, l'occurrence d'une singularité peut affecter seulement quelques niveaux de décomposition. Par conséquent, λ^j doit dépendre de chaque niveau de décomposition.

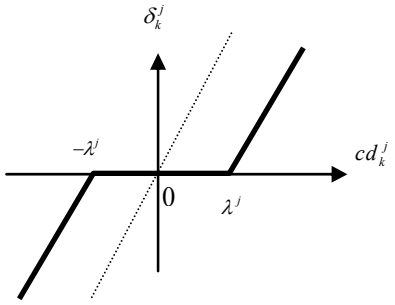


Figure 3.19 — Seuillage doux des coefficients d'ondelettes

Le choix des seuils est délicat. Plusieurs méthodes peuvent être employées et une étude bibliographique indique plusieurs possibilités [67][71][74]. Le choix optimal exige la connaissance (où au moins des hypothèses) sur le signal analysé. Dans cette étude, le seuil qui a été choisi est le *minimax* [71], ce choix est lié à la nature peu bruitée des signaux de l'application industrielle. Sa principale propriété est que la fonction risque :

$$R(g,\hat{g}) = \frac{1}{n} E\left(\sum_{k=1}^{n} \left(g(t_k) - \hat{g}(t_k) \right)^2 \right) \tag{3.39}$$

est minimale, $\hat{g}$ étant le signal débruité.

Pour chaque niveau j, le seuil λ^j est donné par :

$$\begin{aligned} \lambda^j &= \sigma^j \lambda_{n_j}^* \quad \text{si} \quad n_j \ge 64 \\ \lambda^j &= \sigma^j \sqrt{2\log(n_j)} \quad \text{si} \quad n_j < 64 \end{aligned} \tag{3.40}$$

où σ^j est l'écart type des coefficients cd_k^j, n_j est le nombre de coefficients utilisés pour le calcul du seuil et $\lambda_{n_j}^*$ est tabulé et peut être trouvé dans [71].

Les seuils sont calculés à partir d'un signal en absence de défaut. Si cette mesure n'est pas possible et si les connaissance a priori sur l'application permettant de supposer qu'il n'y a pas de défaut dans une portion du signal, alors cette portion est utilisée pour déterminer les valeurs des seuils à chaque échelle. Ces seuils sont fixes au cours du temps. Nous ne nous intéressons pas à la reconstruction du signal, mais à la détection de l'instant quand un changement apparaît dans le signal. De plus, les valeurs des coefficients d'ondelettes n'ont en elles-mêmes aucune

conséquence sur la détection. Seule la comparaison entre leurs valeurs en module et les seuils constitue l'indicateur d'une singularité dans le signal, donc seulement l'évolution des maxima des modules des coefficients de détails est considérée, sans tenir compte de leur signe.

Sur la Figure 3.20, on a représenté le signal x^{obs} (3.33), ainsi que sa décomposition en ondelettes stationnaire avec une ondelette de *Haar*.

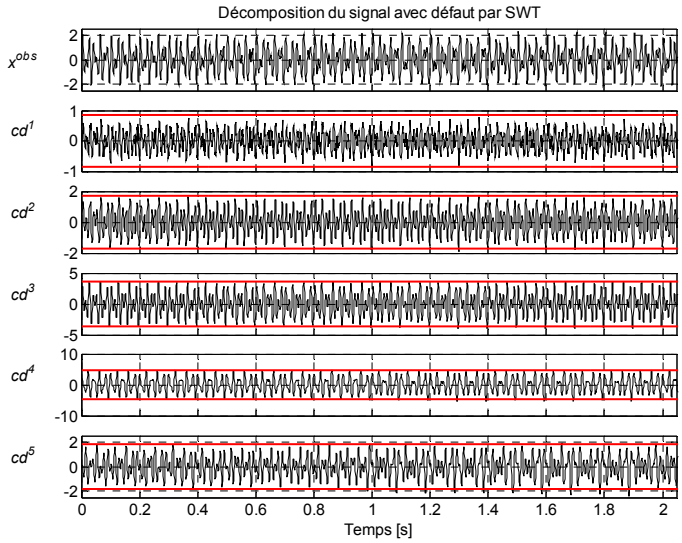


Figure 3.20 — Décomposition par SWT et seuillage des coefficients de x^{obs}

5.3 Validité du modèle (seuils)

Comme on l'a vu précédemment, la validité a priori du modèle tient compte de la confiance qui peut être mise dans les signaux de référence x_i^{ref}, $i = 1:K$, et dans l'évaluation de leur attributs. Pour ce test de détection à base de SWT, les attributs extraits de x_i^{ref} sont les coefficients des détails de la décomposition en ondelettes stationnaire.

Ici, AMV mesure la similarité entre les seuils calculés pour les différents signaux de référence. Plusieurs approches peuvent être suivies. Dans cette étude, on considère $\lambda_{\min}^j = \min_i\left(\lambda^j\left(x_i^{ref}\right)\right)$ et $\lambda_{\max}^j = \max_i\left(\lambda^j\left(x_i^{ref}\right)\right)$ où $\lambda^j\left(x_i^{ref}\right)$ correspond au seuil (3.40) calculé pour le signal de référence x_i^{ref}, $i = 1:K$, et pour le niveau de décomposition j, $j = 1:J$. La similarité entre les seuils pour un niveau est en fait normalisée afin d'obtenir une $AMV \in [0,1]$. Plusieurs opérateurs peuvent être utilisés. Pour cette application (et pour une raison de robustesse), on a choisi :

$$D_v = \Phi_{\otimes}\left(AMV^j\right) \text{ avec } AMV^j = 1 - \min\left\{1, \left\|\lambda_{\max}^j - \lambda_{\min}^j\right\|_2 \Big/ \left\|\lambda_{\max}^j\right\|_2\right\} \tag{3.41}$$

5.4 Test de cohérence par évaluation des dépassements de seuils

Après l'étape de décomposition et celle du seuillage des coefficients d'ondelettes, la troisième étape est l'étape de fuzzification des coefficients seuillés. Cette étape permet de générer un symptôme graduel plus intéressant qu'un indicateur booléen (il y a ou il n'y a pas un défaut).

5.4.1 Fuzzification

La fuzzification des coefficients seuillés est effectuée par (3.42) :

$$c_k^j\left(\delta_k^j\right)=\begin{cases}1 & \left|\delta_k^j\right|\geq 2\gamma^j\lambda^j \\ \dfrac{\left|\delta_k^j\right|}{2\gamma^j\lambda^j} & 0<\left|\delta_k^j\right|<2\gamma^j\lambda^j \\ 0 & \delta_k^j=0\end{cases} \tag{3.42}$$

où γ^j est un paramètre qui définit la fonction d'appartenance des coefficients (cf. Figure 3.21).

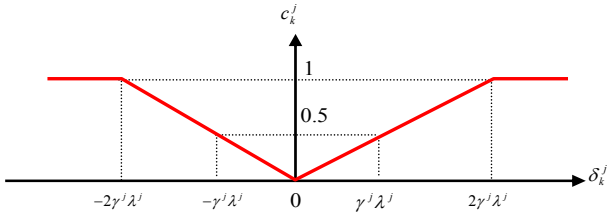


Figure 3.21 — Règle de fuzzification des coefficients seuillés

Sur la figure Figure 3.22, on a reporté les valeurs de c^j pour le signal x^{obs} défini dans (3.33). Dans cette application, on a choisi : $\gamma^j=0.1$, $j=1:J$. On peut constater que dès l'apparition du défaut à l'instant $t_k=1[\mathrm{s}]$ les indicateurs c_k^j prennent des valeurs significatives notamment aux niveaux 3, 4 et 5. Sachant que la fréquence caractéristique du défaut simulé se trouve au 4ème niveau de décomposition, ce fait est lié au choix de l'ondelette ; une ondelette plus régulière serait mieux adaptée pour ce type de défaut.

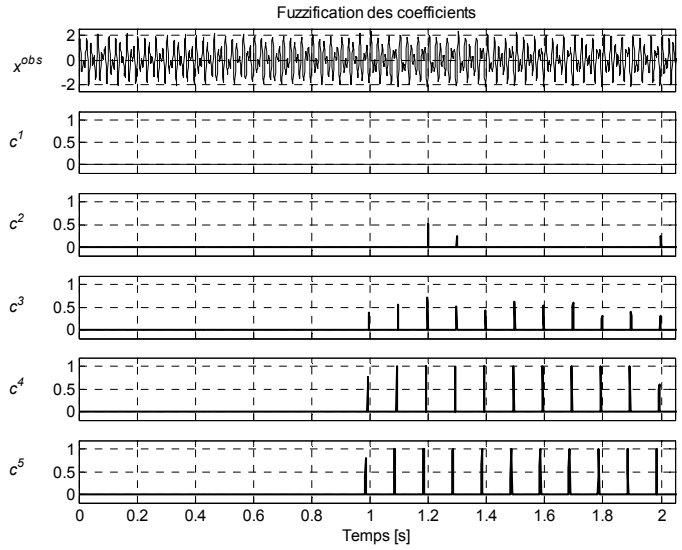


Figure 3.22 — Fuzzification des coefficients seuillés pour x^{obs}

5.4.2 Agrégation

Afin de donner un indicateur unique, mais également progressif, la quatrième étape considère les divers coefficients seuillés et fuzzifiés et assure leur agrégation. Différentes

manières d'agréger on été supposées :

$$D1_{coh} = \min_j (w^j c_k^j);\ j = 1:J \tag{3.43}$$

$$D2_{coh} = \operatorname{mean}_j (w^j c_k^j);\ j = 1:J \tag{3.44}$$

$$D3_{coh} = \max_j (w^j c_k^j);\ j = 1:J \tag{3.45}$$

$$D4_{coh} = \max_j \{\operatorname{mean}(w^j c_{k-l}^j);\ l = 0:N-1\};\ j = 1:J \tag{3.46}$$

$$D5_{coh} = \max_j \{\max(w^j c_{k-l}^j);\ l = 0:N-1\};\ j = 1:J \tag{3.47}$$

(3.43) permet une décision robuste : elle est significative seulement quand la singularité existe à tout les niveaux de décomposition. (3.44) réalise un compromis quand la singularité n'est pas visible sur tous les niveaux. (3.45) permet une décision sensible : elle prend une valeur significative quand un dépassement de seuil sur au moins un niveau de décomposition. Pour améliorer la robustesse des indicateurs (3.43) à (3.45), (3.46) et (3.47) proposent d'agréger les coefficients sur une fenêtre glissante de la longueur N. Ces opérateurs permettent d'amplifier la progressivité des indicateurs précédents.

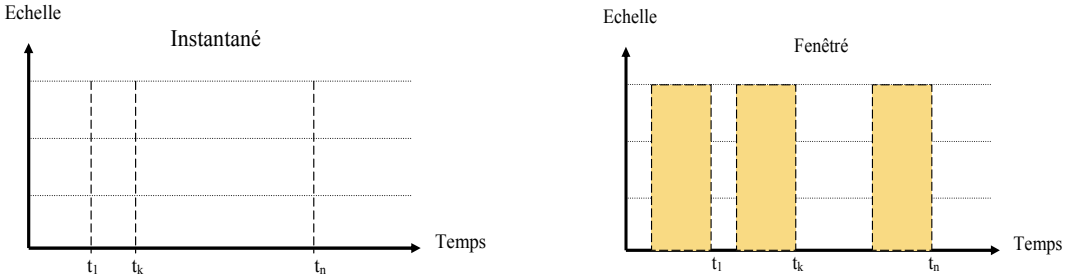


Figure 3.23 — Agrégation multi-échelle

Sur la Figure 3.24, on a repporté l'allure du signal x^{obs} ainsi que le résultat des différents indicateurs $D1_{coh}$ à $D5_{coh}$ avec $N = 2^J = 32$ et $w^j = 1$, $j = 3:J$.

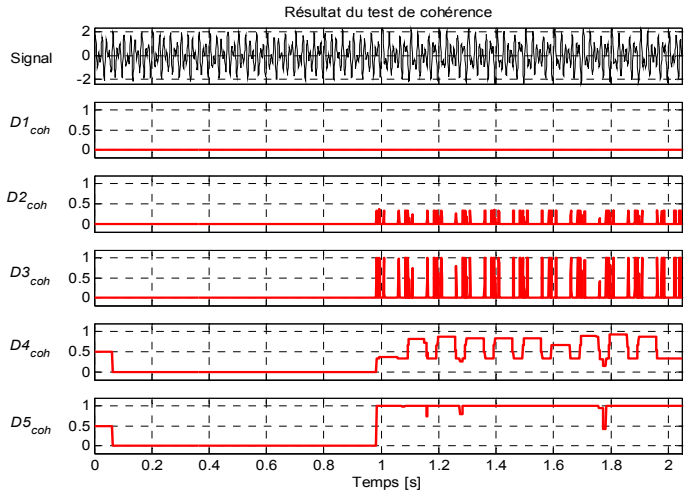


Figure 3.24 — Résultat de l'agrégation multi-échelle pour x^{obs}

Ainsi qu'on pouvait attendre, en observe que dès l'apparition du défaut simulé, par exemple l'indicateur $D3_{coh}$ affiche des valeurs significatifs à partir de l'instant $t_k = 1[\mathrm{s}]$ mais d'une façon discontinue, alors que $D4_{coh}$ et $D5_{coh}$ affichent des valeurs significatifs persistantes. Par conséquent, afin de générer les symptômes on va utiliser l'information donnée par $D4_{coh}$ et $D5_{coh}$ pour générer un indicateur unique D_{coh} par :

$$D_{coh} = \left(D4_{coh} + D5_{coh}\right)/2 \tag{3.48}$$

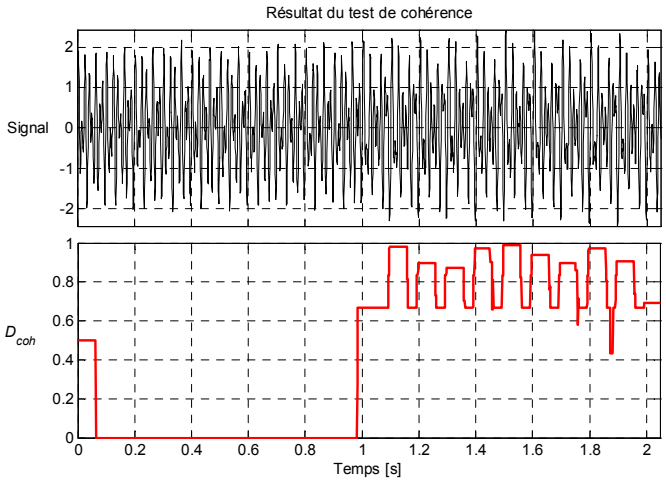


Figure 3.25 — Résultat du test de cohérence pour x^{obs} avec D_{coh}

5.5 Génération des symptômes

A partir de l'indicateurs de cohérence D_{coh} et de la mesure de validité du modèle D_v, on calcule le symptôme D_D par :

$$D_D = \Phi_{\oplus}\left(D_{coh}, D_v\right) \tag{3.49}$$

Sur la Figure 3.26, on a tracé D_D pour les différentes définitions de $\Phi_{\oplus}$, à savoir $D1_D$, $D2_D$ et $D3_D$ calculées respectivement par (3.19), (3.20) et (3.21) avec $D_v = 0.85$.

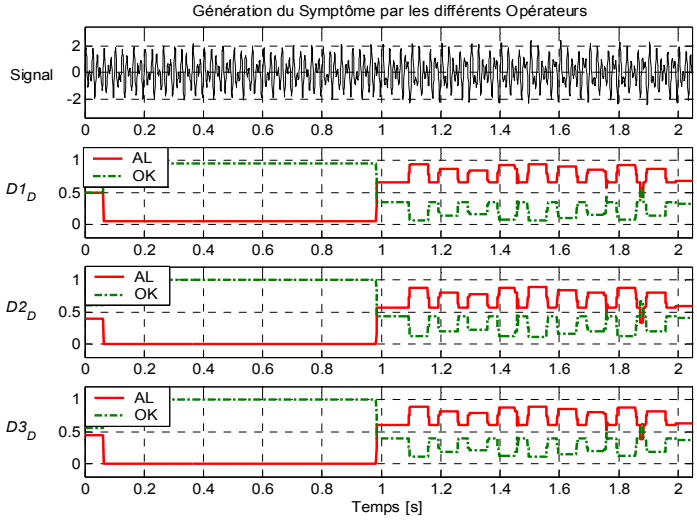


Figure 3.26 — Génération des symptômes

6 Exemples

Dans cette section, les résultats obtenus pour les différentes étapes de la procédure de détection sont présentés pour des exemples académiques. Nous allons considérer plusieurs types de défauts, du type oscillatoire, pseudo périodique, dérive, saut de moyenne, impulsion et modulation. La formulation des signaux considérés est donnée dans le Tableau 3.1. Ces exemples ont été réalisés en utilisant le module développé dans le projet européen MAGIC et le logiciel Matlab® avec la boite à outils Wavelab.

Le signal sans défaut est supposé sinusoïdal bruité par un bruit blanc, de distribution gaussienne, échantillonné à $F_e = 1\,\text{kHz}$. La décomposition en ondelettes est réalisée sur 5 niveaux en utilisant les ondelettes de *Haar* (notée *SWT1*) et de *Daubechies 6* (notée *SWT2*). Ces ondelettes ont été choisie pour leur caractéristiques de localisation, l'ondelette de *Haar* repère bien les défauts discontinus, en revanche la *Daubechies6* est une ondelette plutôt adaptée pour des défauts de type fréquentiel.

Exemple	Type de défaut	Formulation	Figures
	Signal de référence	$x^{ref} = \sin(2\pi 60t) + randn(t)$	Figure 3.27
1	Fréquence additive : 40 Hz	$x_1^{obs} = x^{ref} + \sin(2\pi 40t)u(t-1.024)$	Figure 3.28
2	Fréquence additive : 100 Hz	$x_2^{obs} = x^{ref} + (\sin(2\pi 100t)u(t-1.024))u(t-1.024)$	Figure 3.29
3	Fréquences additives : 40 et 100 Hz	$x_3^{obs} = x^{ref} + (\sin(2\pi 40t) + \sin(2\pi 100t))u(t-1.024)$	Figure 3.30
4	Fréquence additive : 100 Hz	$x_4^{obs} = x^{ref} + 2\sin(2\pi 100t)u(t-1.024)$	Figure 3.31
5	Fréquences additives : 40 et 100 Hz	$x_5^{obs} = x^{ref} + 0.8(\sin(2\pi 40t) + \sin(2\pi 100t))u(t-1.024)$	Figure 3.32
6	Dérive polynomiale d'ordre 2	$x_6^{obs} = x^{ref} + t^2 u(t-1.024)$	Figure 3.33
7	Saut de moyenne de 1	$x_7^{obs} = x^{ref} + u(t-1.024)$	Figure 3.34
8	Impulsion	$x_8^{obs} = x^{ref} + 3\delta(t = 1.024[\text{s}])$	Figure 3.35
9	Pseudo-fréquence 40 Hz	$x_9^{obs} = x^{ref} + 2e^{-0.25(2\pi 40t)}\sin(2\pi 40t)u(t-1.024)$	Figure 3.36
10	Pseudo-fréquence 100 Hz	$x_{10}^{obs} = x^{ref} + 2e^{-0.25(2\pi 100t)}\sin(2\pi 100t)u(t-1.024)$	Figure 3.37
11	Modulation d'amplitude	$x_{11}^{obs} = x^{ref} + \sin(2\pi 100t)\sin(2\pi 40t)u(t-1.024)$	Figure 3.38
12	Modulation de phase	$x_{12}^{obs} = x^{ref} + \sin(2\pi 40t + \sin(2\pi 100t))u(t-1.024)$	Figure 3.39

Tableau 3.1 — Types de défauts simulés

Sur la la Figure 3.27, on a tracé le signal de référence (zoom entre 0.5[s] et 1.5[s]) ainsi que sa décomposition en ondelettes stationnaire avec l'ondelette de *Haar* (au milieu) et de *Daubechies6* (en bas). Les seuils pour les différents niveaux ont été calculés avec (3.40). *AMV* a été obtenue avec (3.41) en considérant 10 signaux de référence.

Analysons maintenant l'influence des différents défauts qui interviennent à partir de l'instant 1.024[s] pour tous les exemples. La décomposition en ondelettes est réalisée avec l'ondelette de *Haar* pour (*SWT1*) et de l'ondelette de *Daubechies6* pour (*SWT2*) sur 5 niveaux de décomposition. Pour la méthode à base de STFT les spectres sont calculés avec une fenêtre de

Blackman de longueur 512 points. Le gabarit est construit suivant la méthode décrite à la section 4.2 avec $\alpha_{\inf} = 1$ et $\alpha_{\sup} = 0.25$. *AMV* a été obtenue avec (3.30) et égale à $D_v = 0.7$. Les symptomes ont été générés en utilisant l'opérateur $D1_D$ (3.19).

Signal de référence :

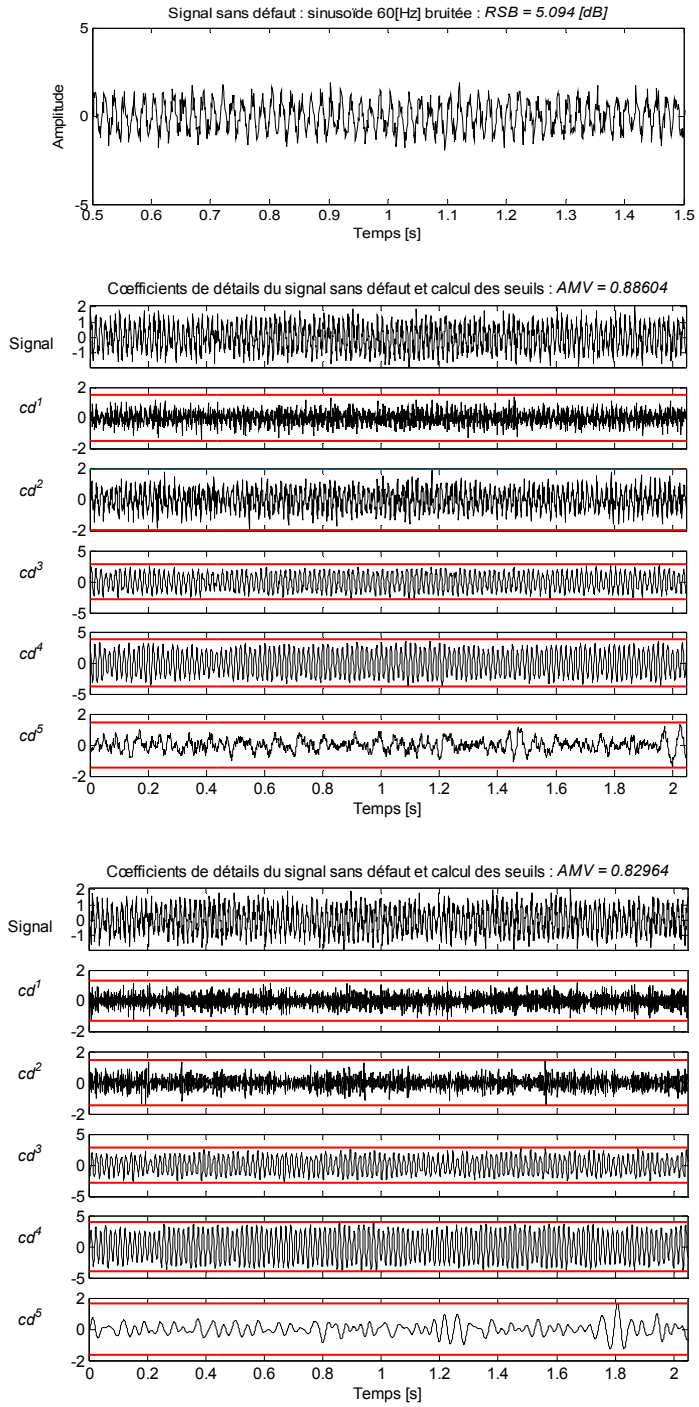


Figure 3.27 — Signal de référence – signal, Analyse de *Haar*, analyse de *Daubechie6*

Exemple 1 :

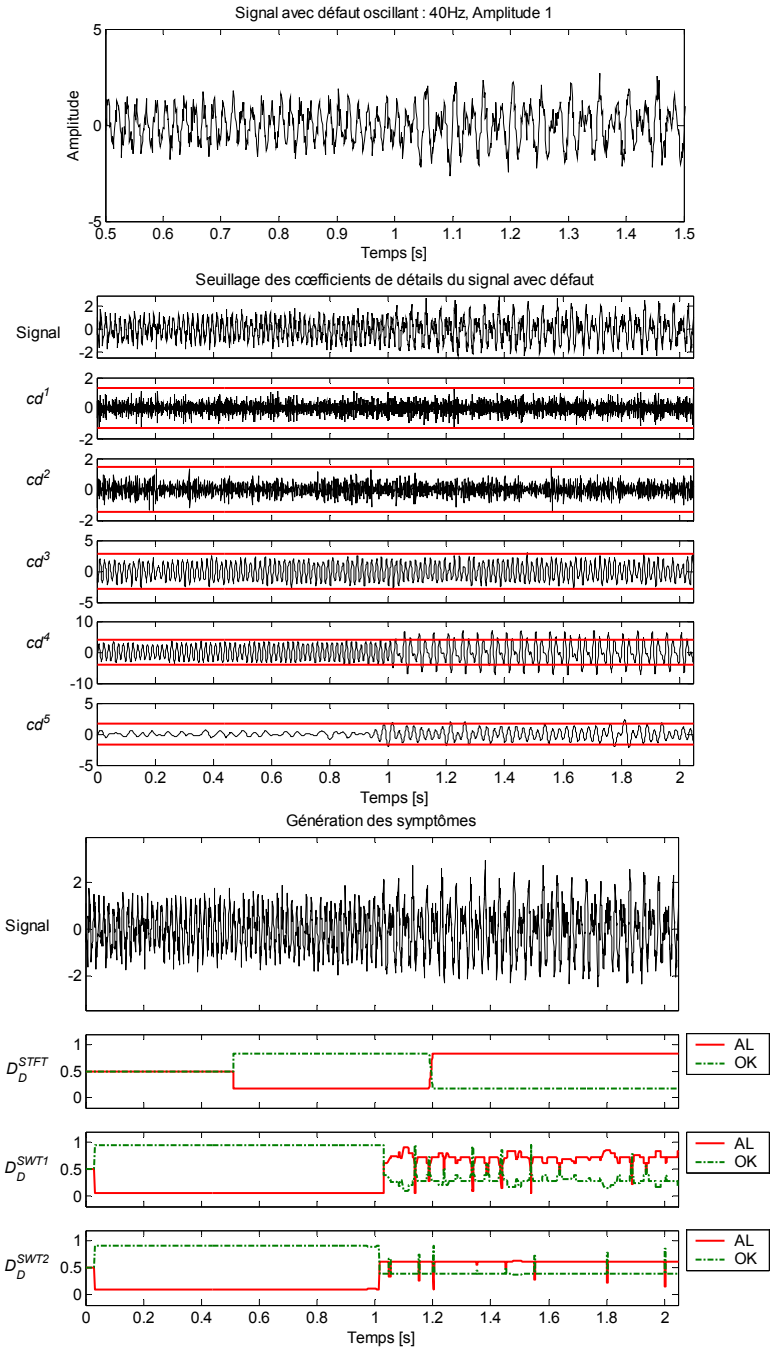


Figure 3.28 — Exemple 1 – Signal (zoom), coefficients d'ondelettes et symptômes

Pour le défaut 1 (Figure 3.28), seul le vecteur de détail cd^4 présente des coefficients qui dépassent le seuil. La fréquence du défaut se trouve dans la moitié haute fréquence de la bande de fréquence indiquée par le vecteur d'approximation ca^3, c'est-à-dire entre 31,25[Hz] et 62,5[Hz]. En effet, la composante fréquentielle du défaut est dans le contenu spectral du vecteur

cd^4. Cette composante est à 40[Hz]. Par conséquent les symptômes affichent des valeurs significatifs à partir de l'instant $t = 1.024[\text{s}]$ pour la méthode à base de SWT, avec des valeurs constantes pour l'analyse *Daubechies6*. En revanche la méthode à base de STFT détecte le défaut avec un retard d'environ $0.8[\text{s}]$.

Exemple 2 :

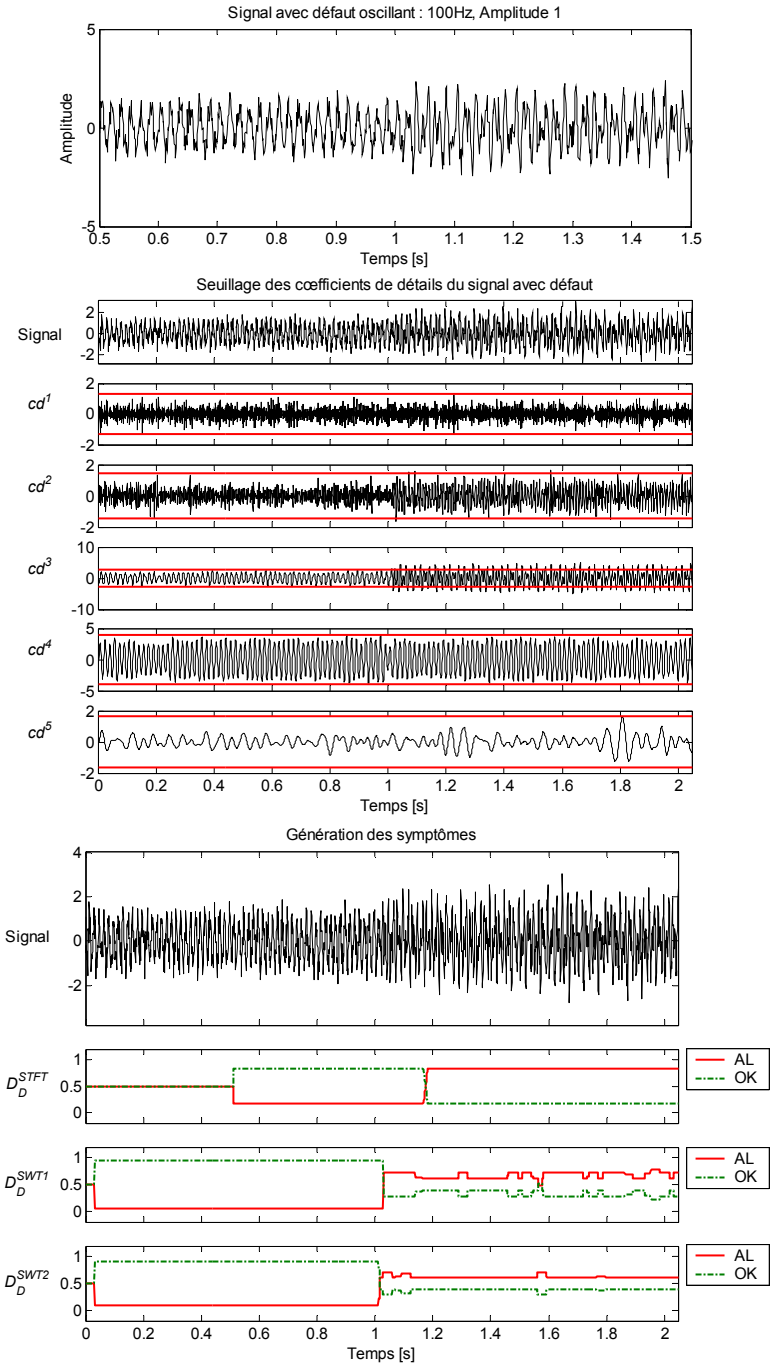
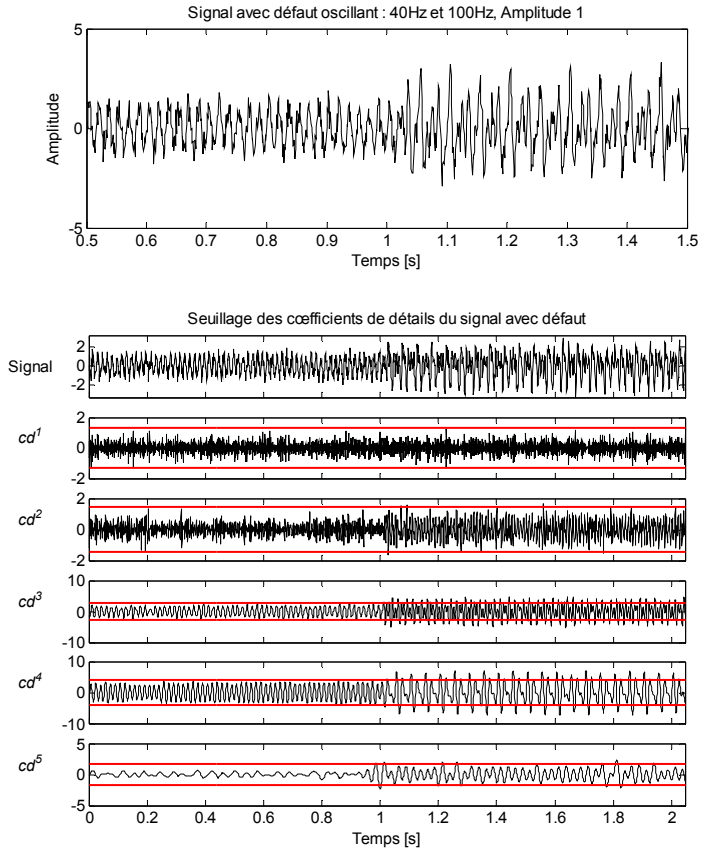


Figure 3.29 — Exemple 2 – Signal (zoom), coefficients d'ondelettes et symptômes

Pour le défaut 2 (Figure 3.29), les coefficients dépassant le seuil peuvent être vus dans le vecteur de coefficients de détail cd^3. Avec les mêmes considérations que pour le défaut précédent, la fréquence du défaut 100 [Hz] se trouve dans contenu spectral de cd^3 et elle est comprise entre 62,5[Hz] et 125[Hz]. Concernant les symptômes, on constate la même chose que l'exemple précédent, en revanche, le symptôme générer à partir de l'analyse de *Haar* cette fois affiche une valeur plutôt constante.

Exemple 3 :



Pour ce troisième exemple (Figure 3.30), le défaut est composé de deux fréquences ; la première 100[Hz] est retrouvée dans le vecteur cd^3 et la deuxième (40 [Hz]) dans le vecteur cd^4. Notons que le symptôme de la méthode de détection par STFT présente toujours un certain retard à la détection compte tenu de la longueur de la fenêtre d'analyse utilisée. En revanche, ce que on peut voir dans ce cas c'est les valeurs des symptômes qui sont nettement plus importantes que les valeurs précédentes, cette amplification est directement liée à l'ampleur du défaut.

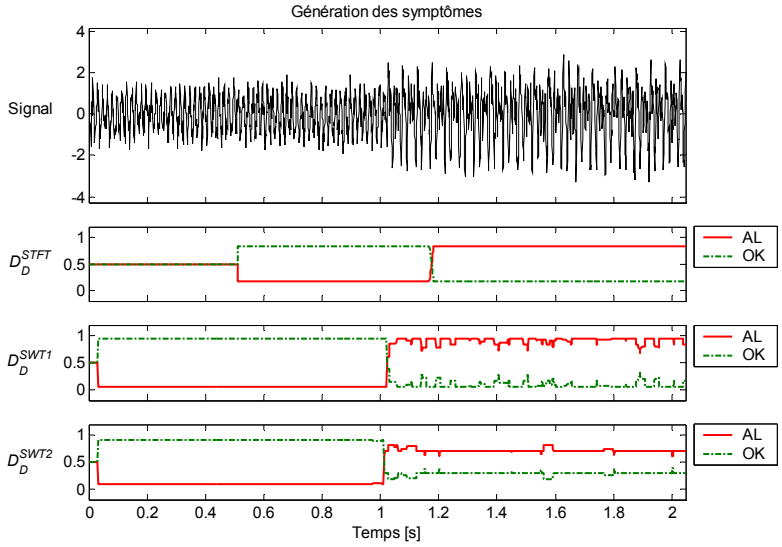
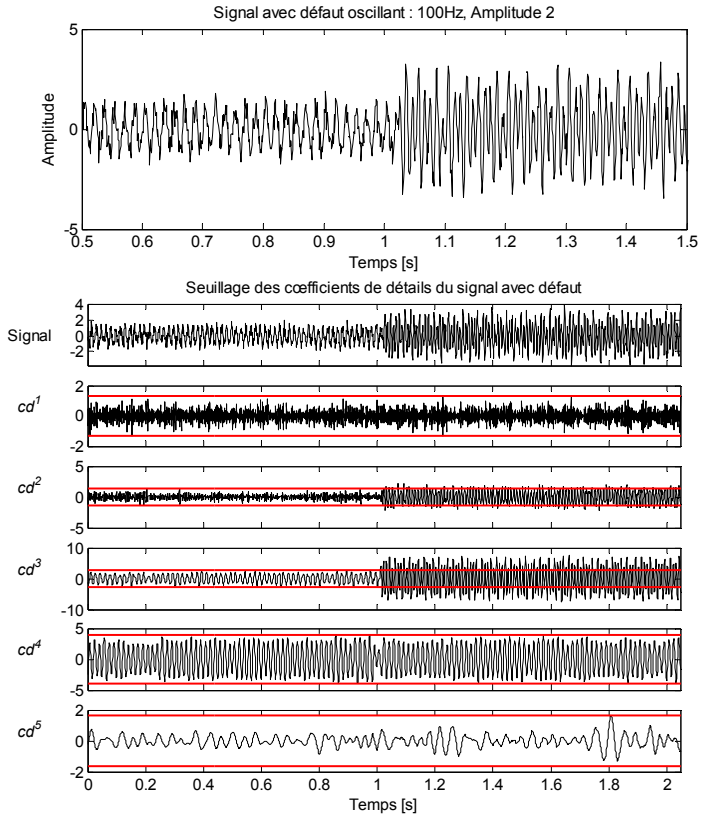


Figure 3.30 — Exemple 3 – Signal (zoom), coefficients d'ondelettes et symptômes

Exemple 4 :



Dans cet exemple (Figure 3.31), on peut constater la même chose que pour le défaut de l'exemple 2, mais avec une amplitude plus importante, les coefficients du détail cd^3 dépassent clairement les seuils. Par conséquent les symptômes présentent des valeurs importantes quasi constantes durant toute la période de présence du défaut. Concernant le symptome générer par la

méthode à base de STFT c'est la présence toujours du retard à la détection.

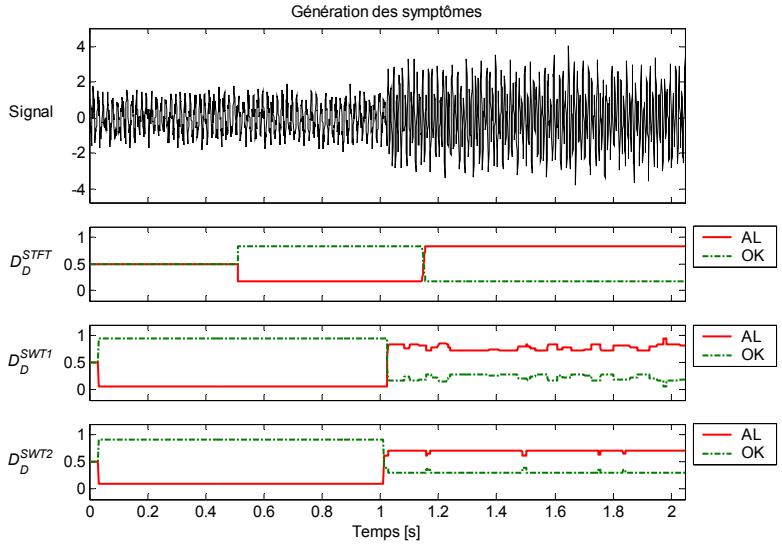
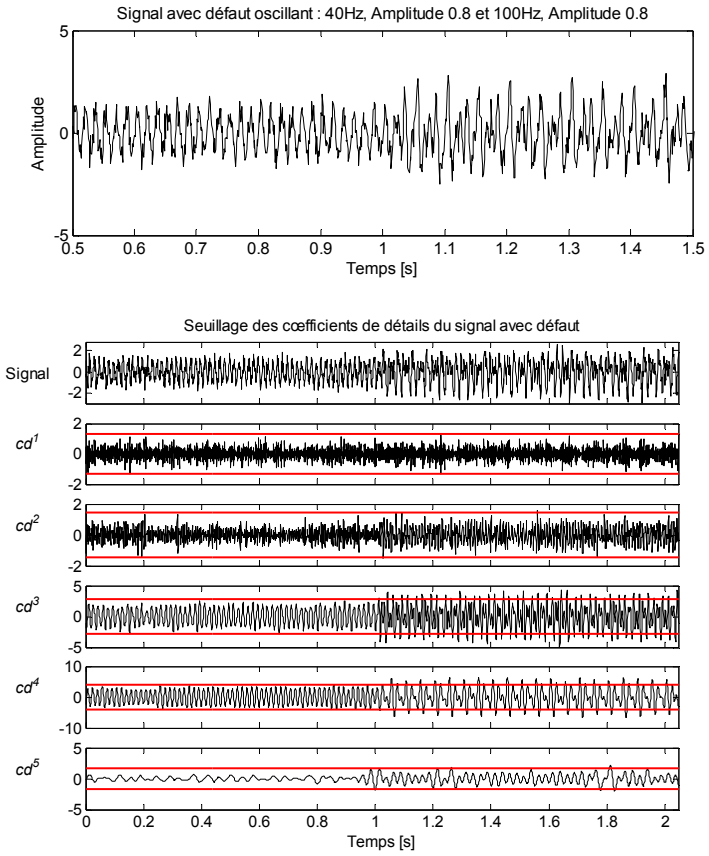


Figure 3.31 — Exemple 4 – Signal (zoom), coefficients d'ondelettes et symptômes

Exemple 5 :



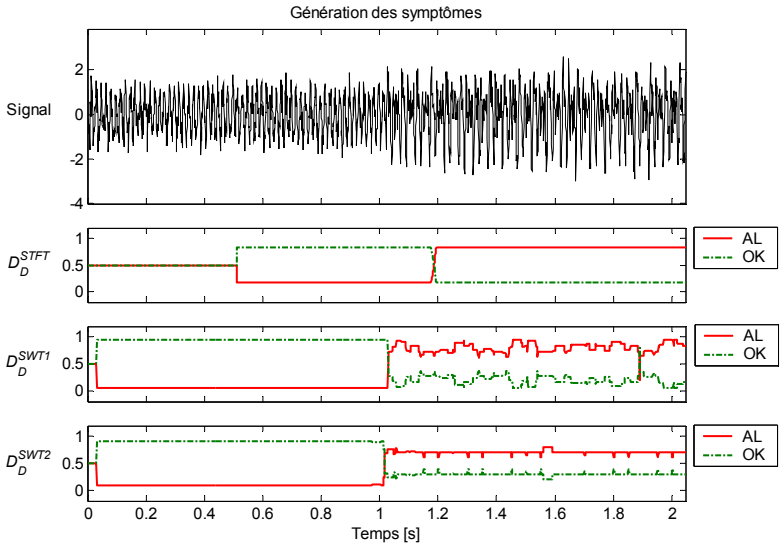
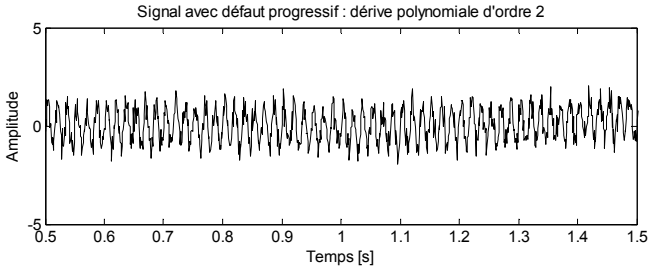


Figure 3.32 — Exemple 5 – Signal (zoom), coefficients d'ondelettes et symptômes

Dans cet exemple (Figure 3.32) également, on peut constater la même chose que pour le défaut de l'exemple 3, mais avec des amplitudes un peu plus faibles : les coefficients des détails cd^3 et cd^4 dépassent clairement les seuils. Par conséquent, les symptômes présentent des valeurs aussi importantes que celles de l'exemple 3 durant toute la période de présence du défaut. Quant à la STFT, on constate ici encore un retard à la détection lié à la longueur de la fenêtre utilisée.

Exemple 6 :



Pour un défaut de type dérive lente (Figure 3.33), la méthode basée sur la transformée en ondelettes présente une limitation, puisqu'elle utilise les coefficients de détails et non pas les coefficients d'approximation, alors que les changements de type dérive lente ce perçoivent généralement dans l'approximation. La méthode basée sur la STFT détecte ce changement puisque le gabarit considéré s'interesse aussi aux bases fréquence.

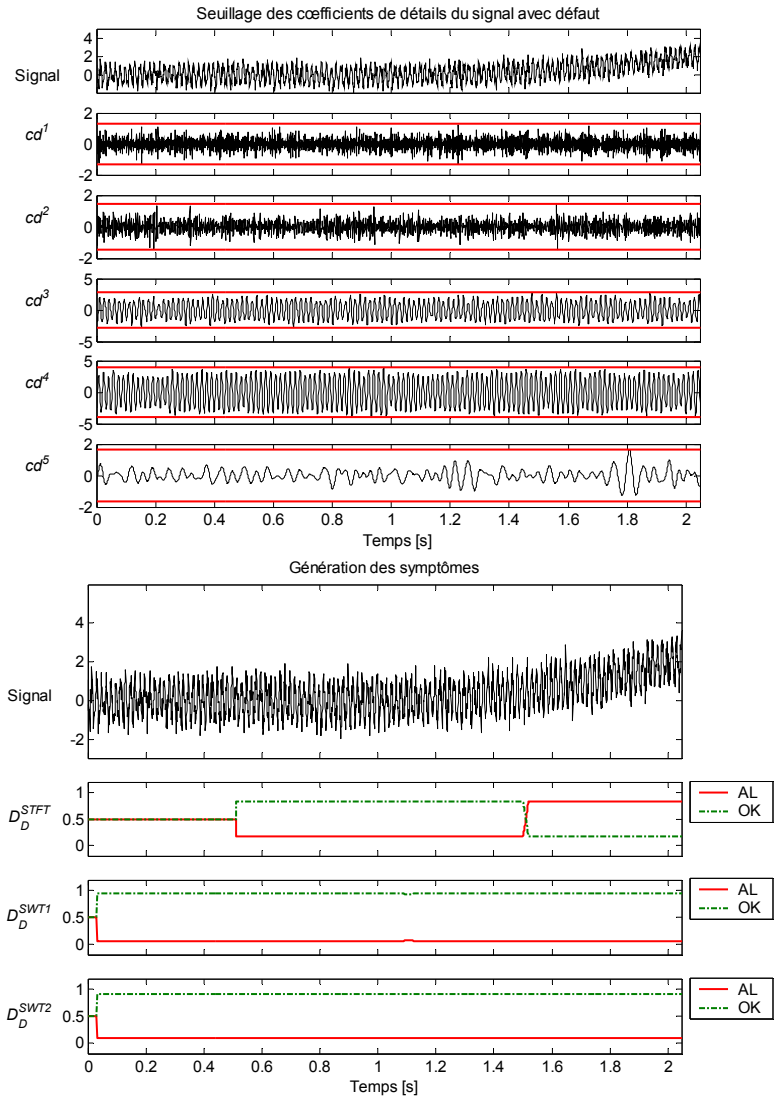
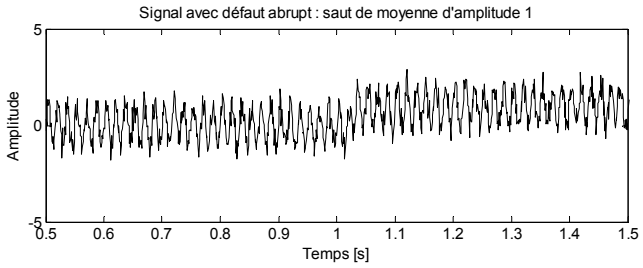


Figure 3.33 — Exemple 6 – signal (zoom), coefficients d'ondelettes et symptômes

Exemple 7 :



Les changements brusques de moyenne (Figure 3.34) peuvent être détectés par la méthode à base de transformée en ondelettes, mais dans ce cas là le choix de l'ondelette est très important. En remarquant les valeurs des symptômes on peut voir que le symptôme généré à partir d'une décomposition en ondelette de Haar (D_D^{SWT1}) détecte le défaut, alors que celui qui a été généré à

partir d'une décomposition en ondelette "régulière" de Daubechies (D_D^{SWT2}) ne le détecte pas. Quant à la STFT, on constate ici encore un retard à la détection.

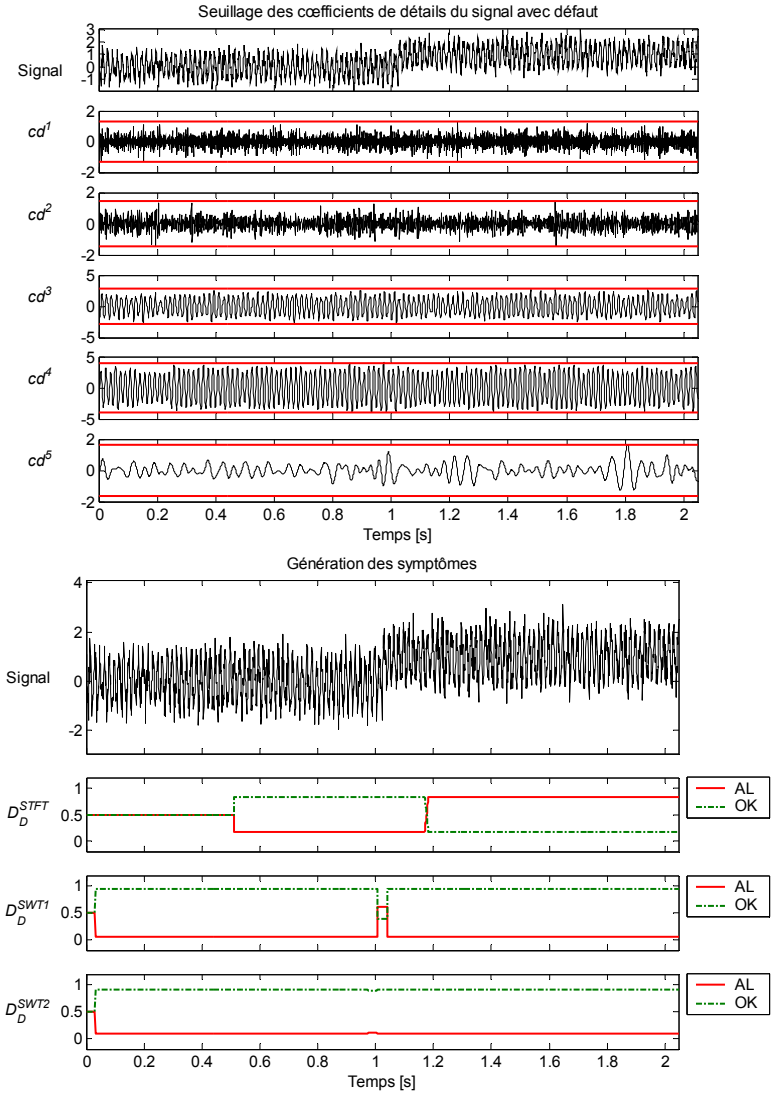
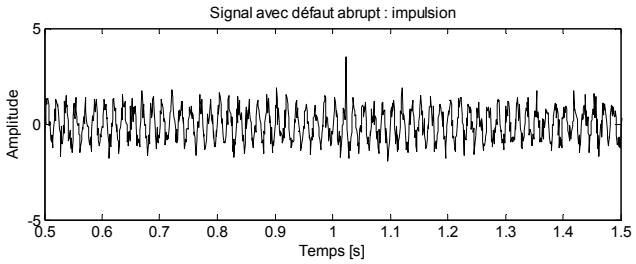


Figure 3.34 — Exemple 7 – Signal (zoom), coefficients d'ondelettes et symptômes

Exemple 8 :



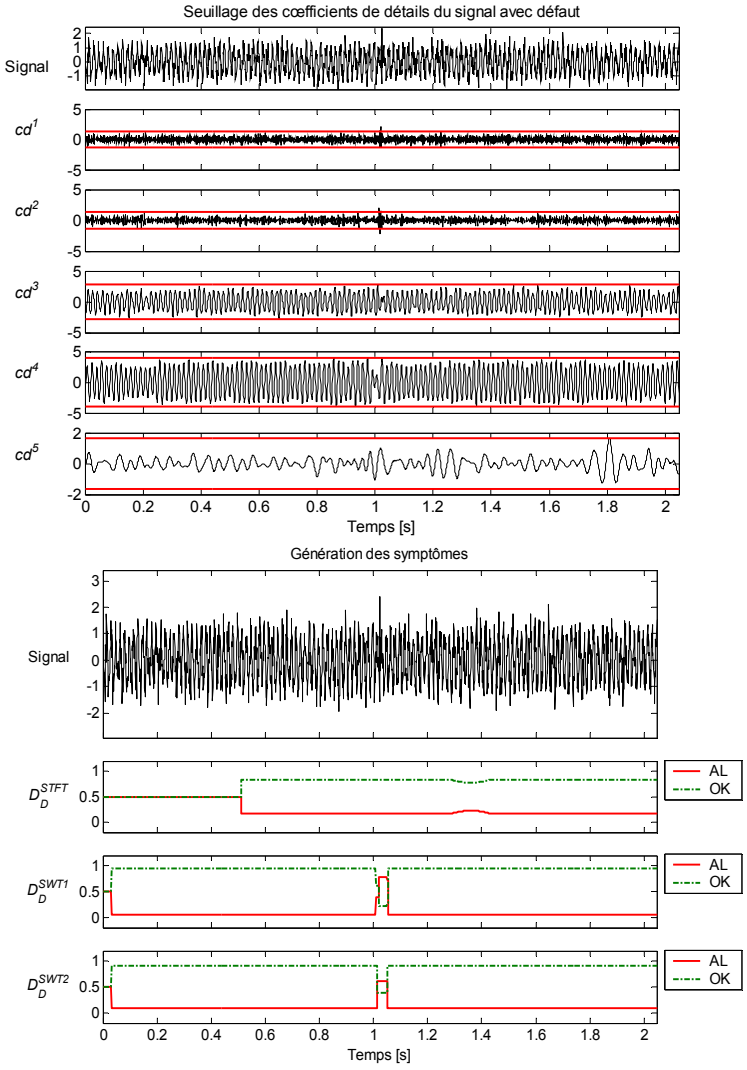
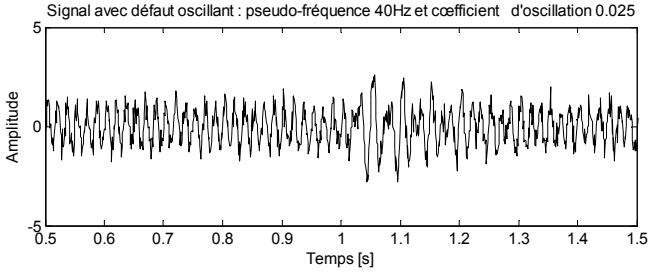


Figure 3.35 — Exemple 8 – Signal (zoom), coefficients d'ondelettes et symptômes

On constate la même chose pour un défaut de type impulsion (Figure 3.35), sauf que dans cet exemple l'amplitude de l'impulsion est suffisamment importante pour être détectée par les deux analyses. En revanche la méthode basée sur la STFT n'est pas adaptée à ce type de défaut.

Exemple 9 :



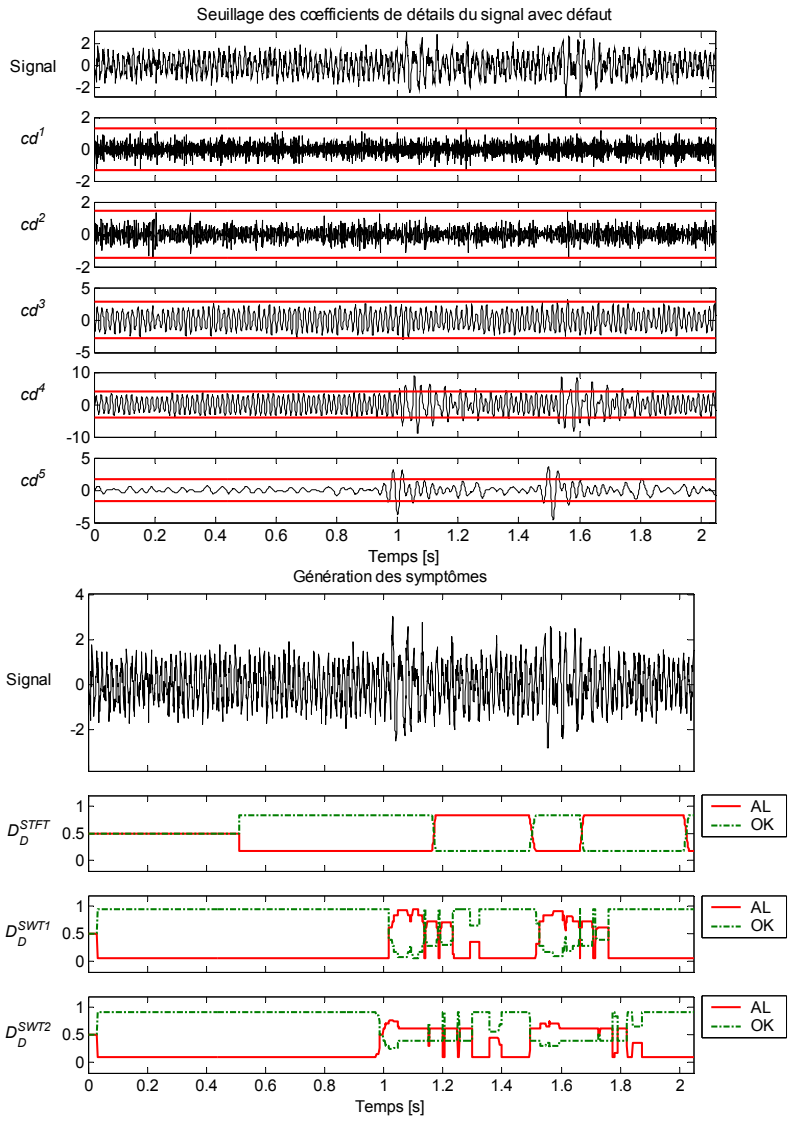


Figure 3.36 — Exemple 9 – Signal (zoom), coefficients d'ondelettes et symptômes

En mécanique on trouve souvent des défauts de type pseudo-oscillatoire, Dans cet exemple (Figure 3.36), on retrouve les coefficients qui dépassent les seuils principalement dans le vecteur cd^4 ceci est du à la fréquence de 40Hz de la pseudo-oscillation. Le symptôme généré à partir d'une décomposition en ondelette de *Haar* présente des valeurs plus importantes que celui de l'ondelette de *Daubechies*. On remarque que les symptômes prennent des valeurs importantes puis diminuent avec l'atténuation des l'oscillations. Concernant la méthode à base de STFT la détection des deux pseudo-oscillations est nettement meilleure mais toujours avec un retard non négligeable.

Exemple 10 :

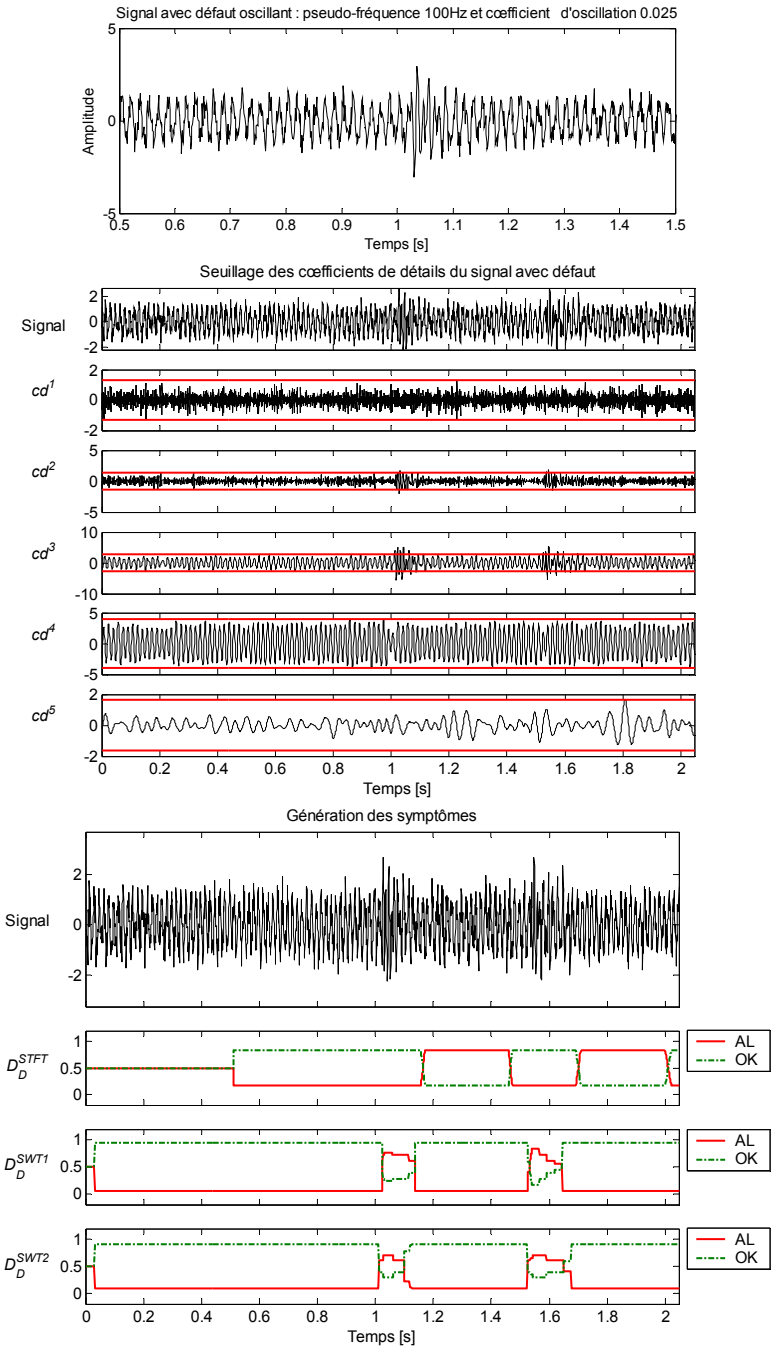


Figure 3.37 — Exemple 10 – Signal (zoom), coefficients d'ondelettes et symptômes

Même remarque que l'exemple précédent, sauf que dans cet exemple (Figure 3.37) les oscillations sont plus rapides, par conséquent les symptômes présentent des valeurs indiquant la présence du défaut durant des périodes mois longues que l'exemple précédent. De plus pour cet exemple les coefficients qui dépassent les seuils sont principalement dans le vecteur cd^3.

Exemple 11 :

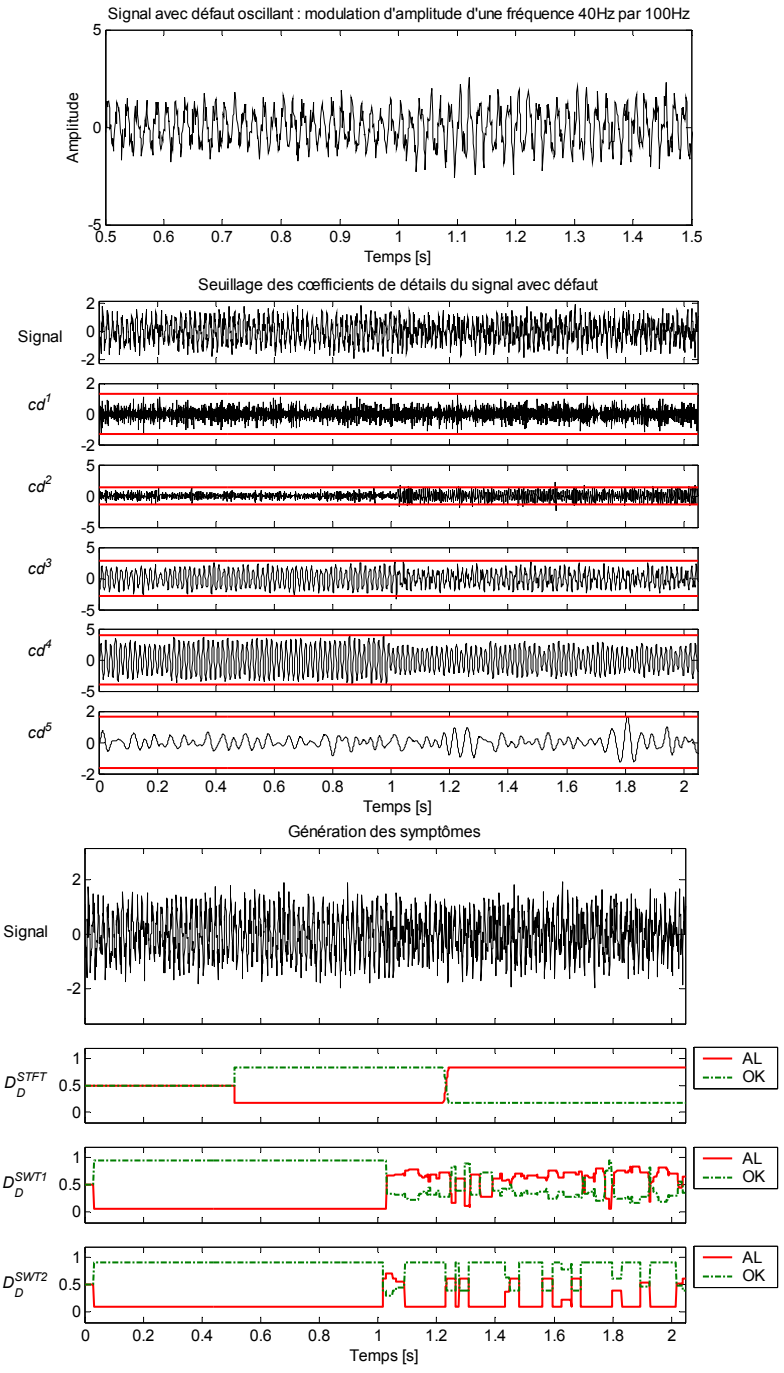


Figure 3.38 — Exemple 11 – Signal (zoom), coefficients d'ondelettes et symptômes

Pour l'exemple 11 (Figure 3.38), le défaut est une modulation d'amplitude d'une fréquence de 40Hz. Le symptôme généré à partir d'une analyse de Haar présente des valeurs importantes, contrairement au symptôme généré à partir d'une analyse de Daubechies. Par contre le symptôme générer par la méthode à base de STFT et plus net.

Exemple 12 :

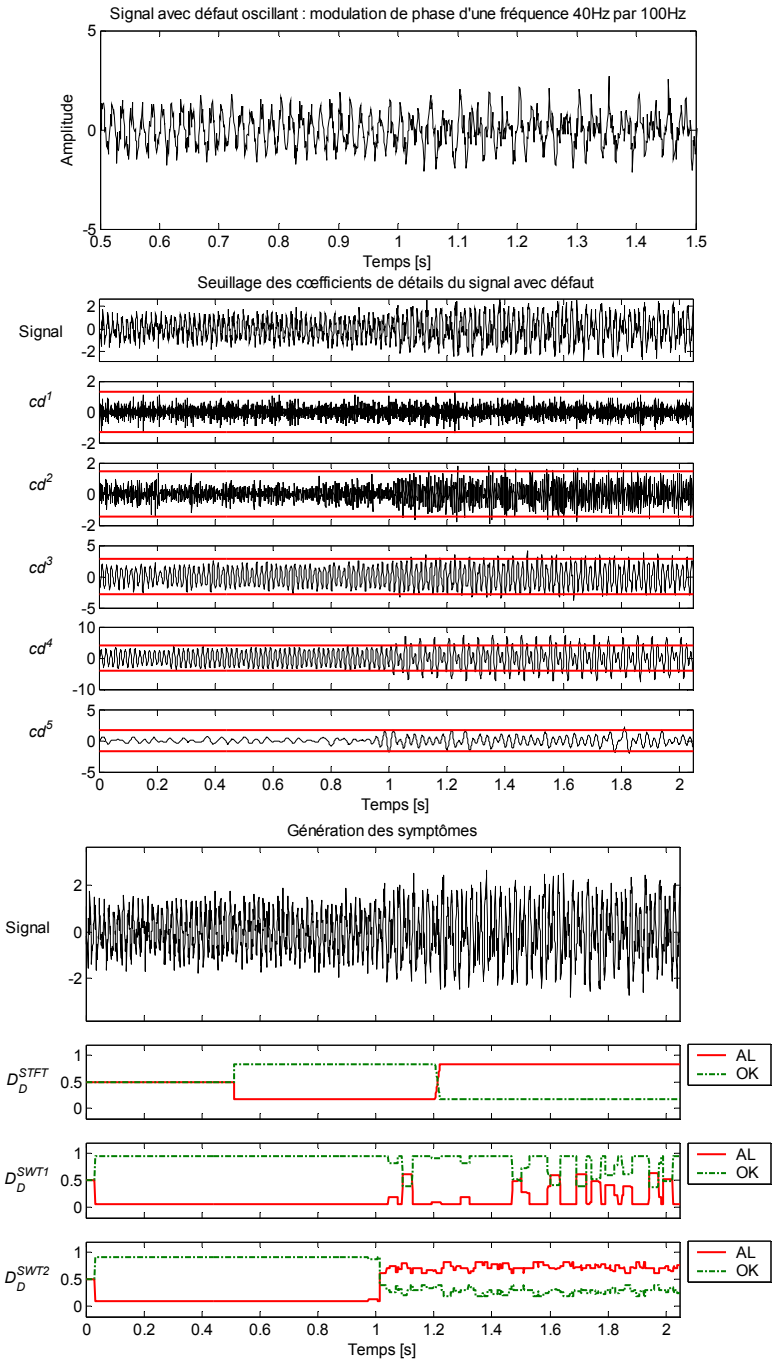


Figure 3.39 — Exemple 12 – Signal (zoom), coefficients d'ondelettes et symptômes

Enfin l'exemple 12 (Figure 3.39) présente un changement de type modulation de fréquence, on peut voir clairement que le choix d'une ondelette régulière est important pour détecter ce type de changement dans les caractéristiques du signal. Le symptôme généré à partir d'une analyse par l'ondelette de Daubechies a des valeurs importantes et régulières, contrairement au

symptôme généré par une décomposition en ondelette de Haar. Concernant la méthode à base de STFT on constate la même chose que les exemples précédents.

7 Conclusion

En diagnostic, il est important de savoir qu'un modèle de comportement n'est pas universel ; il n'est généralement valide que dans certaines plages de fonctionnement, décrites par des conditions de validité. Tous les tests de détection ne conduisent pas toujours à des décisions sûres. L'incertitude se caractérise par l'incapacité de savoir si un test de détection est vrai ou faux. Les tests de détection peuvent conduire à des décisions comprises entre 0 et 1 où 1 représente une alarme sûre et 0 représente l'absence d'une alarme d'une manière sûre. Nous avons montré qu'il était possible d'appréhender des incertitudes de décision au niveau des tests de détection afin de les exploiter durant la phase de génération des symptômes. Le formalisme proposé permet de fusionner la décision du test avec celle de sa validité.

Dans ce chapitre, la validité du prototype a été définie, puis transposé pour deux tests de détection, à base de transformée de Fourier à fenêtre glissante et de transformée en ondelettes stationnaire. Un modèle représentant le comportement normal du procédé est calculé à partir de quelques signaux de référence. La validité de ce modèle AMV est évaluée a priori à l'aide d'une agrégation floue de critères partiels selon les conditions expérimentales observées pendant l'acquisition des données. AMV mesure la confiance qui peut être attribuée au modèle utilisé pour la décision de diagnostic, ce qui influence évidemment la confiance en la décision produite par le test.

Un modèle (gabarit) de spectre de puissance est conçu à partir d'un ou plusieurs signaux référence représentant un comportement particulier (normal) du procédé. Le modèle est par la suite utilisé dans un test de détection par STFT. La validité du modèle AMV a été définie. Elle prend en compte le niveau de confiance qui peut être attribué aux signaux de référence et à l'évaluation de leurs caractéristiques. La cohérence entre le spectre de puissance du signal observé et le modèle (gabarit) est évaluée à l'aide d'un indicateur progressif afin de pouvoir diagnostiquer les défauts progressifs. Cet indicateur est calculé à l'aide de décision floue. La décision finale tient compte de la validité du modèle et de la cohérence avec le spectre de référence. La méthode a été illustrée sur des exemples académiques.

Les possibilités d'utilisation de la transformée en ondelettes stationnaire pour la détection de défauts ont été également étudiées. Une méthode de détection basée sur la transformée en ondelettes stationnaire suivi d'un seuillage des coefficients de chaque niveau a été présentée. Les paramètres de cette méthode sont l'ondelette mère, le nombre de niveaux de décomposition, les seuils et la méthode de seuillage. Les dépassements de seuils sur plusieurs niveaux, associés au comportement anormal sont fuzzifiés, puis agrégés afin de fournir un symptôme progressif. Le choix de l'ondelette dépend des phénomènes qui doivent être détectés dans le signal analysé. Ce choix parfois n'est pas unique. Pour le but de la détection, le choix final est fait afin de maximiser la sensibilité du symptôme.

Chapitre 4

Application en Détection de Vibrations

1 Introduction

La disponibilité des machines est un facteur essentiel de leur qualité. Elle participe à l'économie des entreprises et la sécurité des hommes. Aujourd'hui la connaissance des machines de et leur comportement ne sont plus capitalisés par la mémoire des exploitants. Ce n'est plus l'intuition et la connaissance des individus qui sont utilisées pour comprendre les dérives et évaluer l'état des l'installations, mais des systèmes automatiques de supervision.

La surveillance vibratoire des machines est l'une des techniques qui se répand dans l'industrie ; longtemps cantonnée à l'équilibrage des machines tournantes [86], elle a fait de grands progrès grâce au traitement informatique du signal. Aujourd'hui, par le développement des études sur la pathologie des machines, les effets vibratoires observés peuvent être reliés aux causes matérielles qui les génèrent. Pour l'utilisateur, ces possibilités nouvelles offrent des diagnostics plus sûrs, plus rapides, plus économiques. Mais la complexité des mécanismes en jeu nécessite des éclairages théoriques et pratiques. Les causes ne se limitent pas au simple balourd des machines tournantes. Elle ont aussi pour origine les composants technologiques utilisés, ainsi que les principes physiques sur lesquels leur fonctionnement repose.

Toutes les machines en fonctionnement produisent des vibrations. Images des efforts dynamiques engendrés par les pièces en mouvement, ces vibrations occupent une place privilégiée parmi les paramètres à prendre en considération pour effectuer un diagnostic. Leur place est d'autant plus justifiée que la modification d'une signature vibratoire de machine constitue souvent la première manifestation physique d'une anomalie, cause potentielle de dégradations, voire de panne. Ces caractéristiques font de la surveillance par analyse des vibrations, un outil indispensable pour une maintenance moderne [86], puisqu'elle permet, par un diagnostic approprié des défauts, d'éviter la casse et de n'intervenir sur une machine qu'à bon escient et pendant des arrêts programmés de production.

Dans ce chapitre on va aborder une application concernant les vibrations torsionnelles dans un laminoir réversible (laminage à chaud) pour des produits plats. Ce chapitre se termine par une

conclusion sur l'application des tests de détection proposés au chapitre 3.

2 Détection de vibrations torsionnelle dans un laminoir

Le laminage, opération d'une grande diversité, peut se résumer ainsi : c'est une mise en forme par déformation plastique, destinée à réduire la section d'un produit de grande longueur, par passage entre deux ou plusieurs outils axisymétriques tournant autour de leur axe, la rotation des outils qui entraîne le produit dans l'emprise par l'intermédiaire du frottement [84]. Opération fondamentale en métallurgie, elle voit passer environ 90 % de tout le métal produit, tous métaux et alliages confondus.

Le laminage vient après l'élaboration du métal, puis la coulée, le plus souvent continue maintenant. Si l'on oublie quelques laminages « exotiques » (laminage d'anneaux et de roues, laminage - retour de barres à forger), il se scinde en [83] :

- laminage de produits longs (barres, fils, tubes, poutrelles, rails...), où les deux dimensions de la section, du même ordre de grandeur en général, sont petites devant la longueur ; les outils sont le plus souvent des cylindres cannelés;
- laminage de produits plats (tôles, bandes et feuillards) où l'épaisseur est petite devant la largeur, elle-même très inférieure à la longueur. Les outils sont des objets axisymétriques presque cylindriques (au bombé de rectification près).

Le laminage fournit surtout des produits semi finis, mais avec des exceptions notables : tôles fortes, poutrelles et rails, certains tubes.... Les autres produits longs sont destinés à l'usinage (barres), au tréfilage (fil machine), au forgeage ; quant aux produits plats, ils vont vers des opérations de mise en forme des métaux en feuilles (emboutissage, découpage fin, repoussage et fluotournage).

La plupart des opérations de laminage présentent un régime permanent dominant, voire quasi exclusif. Il n'en reste pas moins que des phases transitoires existent : tête et queue de plaque, tôle, brame ou bande, soudure entre bobines en laminage continu d'acier à froid. Il est d'autre part très difficile de considérer le régime thermique comme permanent, surtout pour les cylindres. Les modèles, pour des raisons d'économie de temps de calcul, se concentrent souvent sur le régime stationnaire – même si le problème à étudier est précisément lié aux transitoires, comme c'est malheureusement souvent le cas.

Les vibrations des laminoirs sont un des problèmes les plus préoccupants aujourd'hui, amplifiés par la croissance de la vitesse de laminage. On connaît en particulier le broutement, vibration qui force l'opérateur à diminuer la vitesse du laminage [77]. Excité par n'importe quelle fluctuation de paramètre d'entrée pour peu que la vitesse de laminage lui donne une amplification, le broutement s'auto entretient du fait de la quadrature de phase entre les vibrations verticales des cylindres de travail et la fluctuation consécutive des tensions de bande (cf. Figure 4.1). Tout ce qui tend à augmenter l'effet des tensions de bande sur la force de laminage va dans le sens de la déstabilisation.

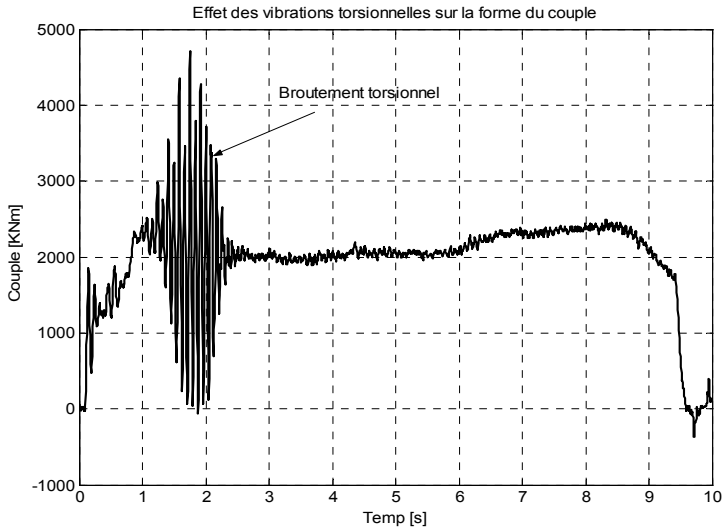


Figure 4.1 — Effets des vibrations sur la forme du couple

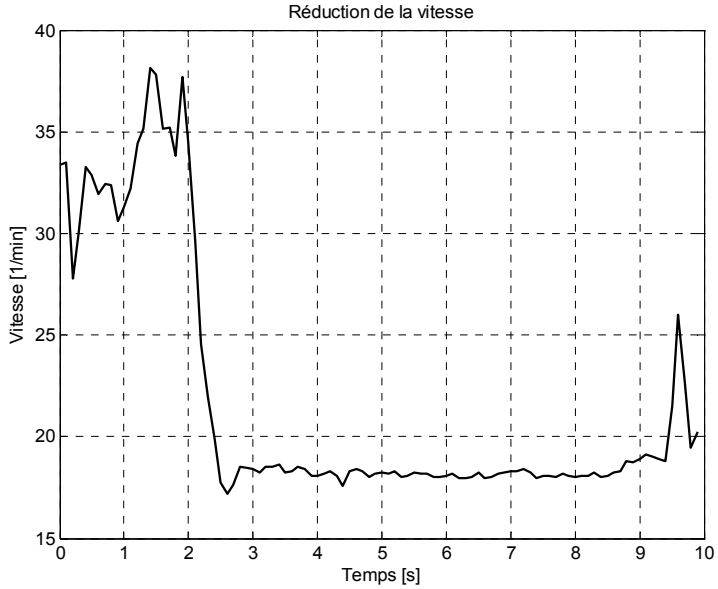


Figure 4.2 — Réduction de la vitesse de laminage à cause du broutement torsionnel

La réponse des industries de laminage à la compétition internationale et à la pression globale tend vers une augmentation de productivité surtout obtenue par une augmentation de la vitesse de laminage. Cela est particulièrement vrai pour les laminoirs à froid dans lesquels les bandes d'acier et d'aluminium sont produites.

Cependant, des vitesses plus élevées augmentent la densité de puissance dans le processus aussi bien que le potentiel du procédé à réagir avec des vibrations. Dans certaines conditions de fonctionnement, des vibrations ou des oscillations peuvent se produire, et provoquent le

phénomène de broutement ou les marques de broutement sur les composants du laminoir. Les marques de broutement sont des défauts périodiques dans l'épaisseur ou la forme de la bande ou des nuances régulières sur la surface de la bande perpendiculaire à la direction de laminage. Les vibrations importantes du laminoir peuvent même causer des ruptures de la bande.

L'analyse, la surveillance et la correction des imperfections de bande provoquées par les vibrations des laminoirs est un champ de recherche actif dans le domaine du laminage. L'industriel métallurgique (SMS-Demag AG) avec lequel nous avons collaboré étudie depuis des années les effets des vibrations dans les laminoirs. Les objectifs principaux de SMS-Demag dans ce domaine de recherche et de développement sont l'analyse du phénomène aussi bien que le développement des systèmes de surveillance assistés par ordinateur, avec lesquels la qualité du produit est continuellement surveillée et assurée.

2.1 Description technique du laminoir

Les moteurs d'entraînement tendent souvent à vibrer ce qui réduit leur durée de vie, la qualité du produit et qui cause fréquemment des défauts. Ceci est particulièrement vrai pour les moteurs d'entraînement principal des laminoirs. Bien que ces entraînements ont une structure adaptée, la capacité de charge des laminoirs toujours plus élevée conduit à une augmentation accrue des efforts et des vibrations, qui ne peuvent plus être tolérés par le procédé. Pour la détection et le diagnostic afin de détecter de telles vibrations des méthodes basées sur l'analyse des signaux sont employées en particulier en analysant les mesures de couples.

Des vibrations torsionnelles des équipements laminoir sont souvent induites par les moteurs d'entraînement et ont un effet sur les composants de la chaîne d'entraînement entière [77]. En outre, ces vibrations s'auto-entretiennent et excitent les fréquences naturelles de vibrations torsionnelles du système. D'autres vibrations peuvent être excitées par des convertisseurs de fréquence ou des contrôleurs mal paramétrés et se propagent par l'intermédiaire des moteurs à la chaîne d'entraînement.

Dans un laminoir, les plus importantes perturbations viennent des impulsions, par exemple comme les changements du couple générées pendant l'entrée de la tôle entre les cylindres. Près de l'excitation des premières fréquences naturelles, ces perturbations causent une baisse de vitesse de l'entraînement, qui doit être accélérée jusqu'à la vitesse nominale par le contrôleur de vitesse. Les forces variables de roulement et l'excitation périodique des excentricités de roulement et le déséquilibre sont des perturbations additionnelles.

Pendant un fonctionnement normal, les vibrations peuvent également augmenter par des défauts des éléments de la machine comme par exemple des défauts de vitesse ou écaillage des roulements, une excentricité des cylindres, un jeu important de denture, un frottement accru ou des vibrations électriques induites par les convertisseurs de puissance.

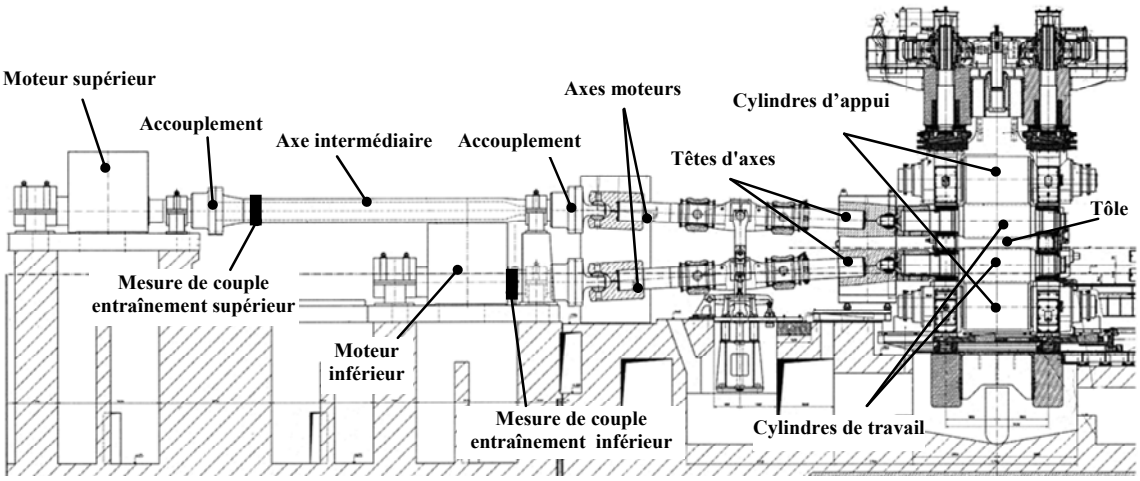


Figure 4.3 — Schéma du laminoir dégrossisseur réversible (SMS-Demag AG) [77]

L'application traitée concerne l'entraînement principal d'un laminoir de tôle, où un système de mesure de couple télémétrique a été installé. L'entraînement des parties supérieure et inférieure est séparé dit *Twin Drive*, avec deux moteurs différents qui entraînent les cylindres de travail supérieur et inférieur directement. Entre les moteurs supérieur et inférieur il n'y a aucun lien mécanique permanent. Pendant le laminage, les deux chaînes d'entraînement sont couplées par la bande laminée. La Figure 4.3 illustre une représentation schématique du *Twin Drive*. Il se compose des moteurs d'entraînement supérieur et inférieur, des axes, des accouplements, des axes d'extrémité et de l'ensemble des roulements intermédiaires.

Puisqu'il y a qu'une seule plateforme de laminage, la tôle est laminée en plusieurs passages jusqu'à son épaisseur finale. La bande peut entrer entre les cylindres de travail des deux côtés de la table de laminage et par conséquent les entraînements principaux du laminoir doivent inverser le sens de rotation des moteurs entre chaque passage de la tôle. Chaque passe a son propre identifiant de passage.

2.2 Les vibrations et leurs effets sur la qualité de la tôle

Les vibrations de la plateforme de laminage affectent la bande laminée, chaque fois que les contours déformés des cylindres de travail se déplacent relativement par rapport à la bande. Selon la direction dans laquelle les cylindres de travail tournent, les défauts se produisent à des intervalles plus ou moins réguliers, perpendiculairement à la direction de laminage (cf. Figure 4.4) [85]. Les défauts de bande représentés sur la Figure 4.4 se produisent rarement sous une forme pure telle que représentée sur la figure.

L'épaisseur, par exemple, demande une vibration d'antiphase de la moitié inférieure du laminoir réglée contre la moitié supérieure. La bande laminée montre un ensemble d'erreurs de bande. L'amplitude et la longueur d'onde des défauts périodiques de bande dépendent du système de vibration et de la fréquence de vibration. Deux gammes de fréquences peuvent être distinguées [82] :

- vibrations verticales à la plateforme, dites vibrations de plateforme: de 1 Hz à 15 Hz ;
- broutement torsionnel, vibrations torsionnelles d'entraînement principal : 5 Hz à 20 Hz ;
- broutement de troisième octave : 100 Hz à 300 Hz ;
- broutement des cylindres ou broutement de cinquième octave : 500 Hz et plus.

En outre, une différentiation des fréquences d'interférence entre celles qui sont proportionnelles à la vitesse et celles qui ne sont pas, peut aider à décrire le phénomène de vibration et ses manifestations [82]. Ainsi on peut avoir des :

Causes proportionnelles à la vitesse :

- dégradation du matériel par le broutement de troisième octave ;
- forme polygonale du cylindre en raison de vibration de la machine de meulage de cylindre ;
- Erreur d'équilibre et d'excentricité des cylindres ;
- rouleaux de roulement et erreur de dents d'engrenage ;
- irrégularité d'entraînement, en raison du déplacement de l'axe ;

Causes indépendantes de la vitesse :

- vibrations naturelles de la plateforme de laminage ;
- fluctuation du couple ;
- non homogénéité du matériau ;
- effets de glissement ;
- vibration de moteur (i.e. régulation du moteur) ;
- marques de broutement dues au mouvement libre excessif.

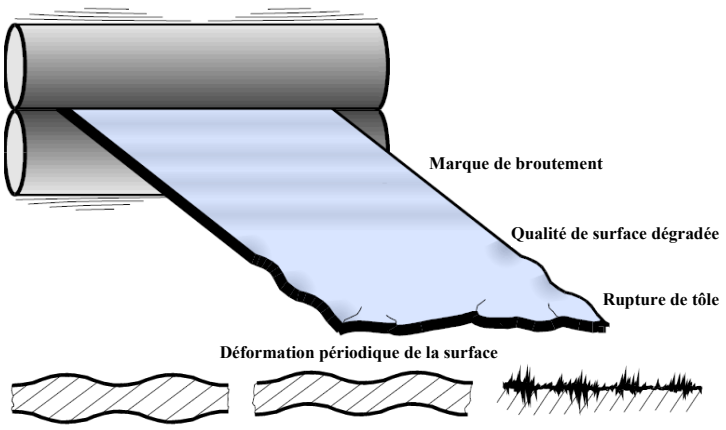


Figure 4.4 — Effets des vibrations sur la qualité du produit

Les vibrations observées dans les plateformes de laminage peuvent également être classées selon les différents types vibratoires connus en physique :

- *les vibrations libres* : se produisent quand une simple impulsion de déclenchement affecte le système oscillatoire [82]. Dès que l'impulsion sera déclenchée, la plateforme de laminage ou des parties mobiles oscillent avec leur fréquence naturelle ;

– *les vibrations excitées séparément* : dans les laminoirs c'est en général l'effet direct d'une plateforme dégradée ou d'un procédé déficient, comme par exemple quand le jeu dans la fixation de cylindres, les rouleaux des roulements ou le train d'entraînement est trop important ;
– *les oscillations auto-excitées* : résultent de l'interaction de divers paramètres. Ce qui est connu comme le broutement de troisième-octave des plateforme peut être classé dans cette catégorie ;
– *des vibrations paramétriquement excitées :* peuvent être liées au changement périodique des paramètres du système (par exemple variation des variables commandées).

2.3 Description des données et acquisition

Les résultats de quelques tests sur différents types de laminoirs ont montré les avantages d'utiliser le couple pour la surveillance dans les laminoirs. Le couple est en effet le paramètre le plus important dans le laminage. La productivité, la qualité du produit et la fiabilité de la plateforme demandent un couple constant pendant le laminage. En pratique, les moteurs d'entraînement principaux du laminoir sont des composants fortement et dynamiquement chargés : l'impact de l'aplatissement, l'inversion du sens de rotation, les vibrations torsionnelles, etc. affectent le produit et la durée de vie des composants.

Pour la surveillance des entraînements du laminoir le couple est mesuré directement aux axes ce qui s'avère plus avantageux que de mesurer le courant du moteur. Le système de mesure de couple pour le laminoir présenté dans ce chapitre a été conçu pour des machines "lourdes", c'est-à-dire entraînements principaux du laminoir. Il se compose d'un capteur de couple et d'un système d'acquisition. Le capteur de couple a été développé pour un fonctionnement permanent et continu et pour de grands diamètres d'axe.

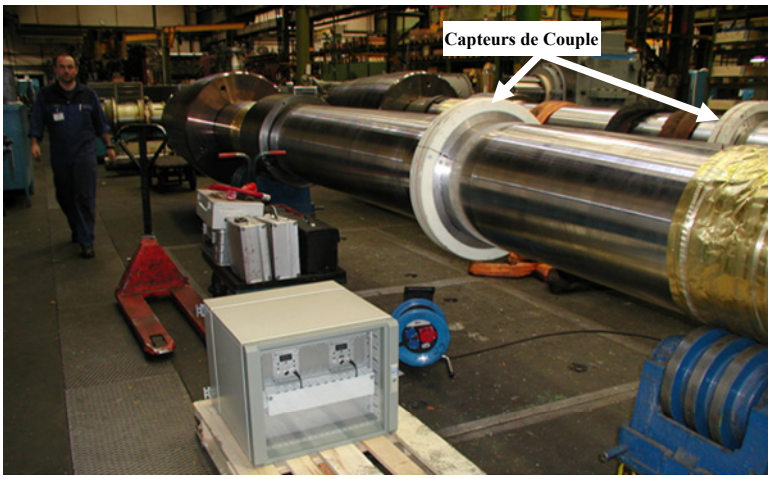


Figure 4.5 — Photo du capteur de couple [77]

Les capteurs de couple utilisés dans les laminoirs doivent être robustes à cause des conditions d'utilisation. La Figure 4.5 montre le capteur de mesure du couple à base de jauges de contrainte [77]. L'expérience a prouvé que la fiabilité des capteurs de couple dans les laminoirs peut

seulement être garantie si ils sont entièrement en contact libre [85], c'est-à-dire, avec transmission télémétrique du signal et alimentation en énergie inductive.

Une information directe peut être extraite des signaux de couple pour l'optimisation du programme de passage de la bande à rouler, l'amélioration de la lubrification et du refroidissement dans l'espace de laminage en plus de la protection contre les temps d'arrêt non programmés, l'estimation de l'état courant et l'évaluation de la durée de vie des composants. Notons que des méthodes de traitement du signal ont déjà été employées pour la détection et le diagnostic des vibrations dans l'industrie métallurgique. Certains concepteurs les ont intégrées dans des systèmes de surveillance assistés par ordinateur [82].

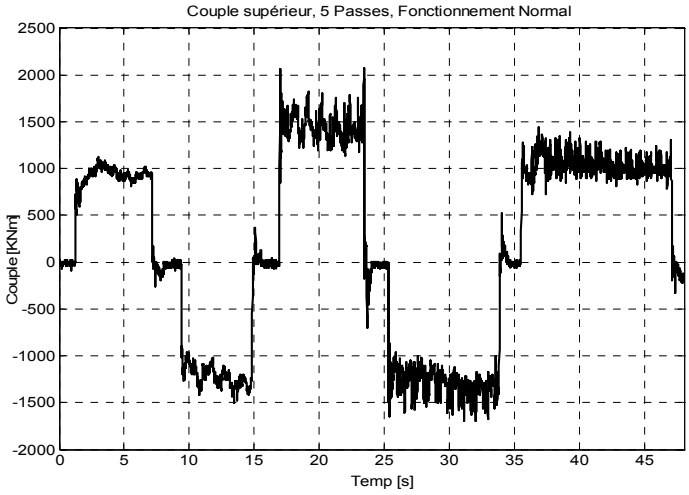


Figure 4.6 — Forme du couple haut durant un fonctionnement normal

Les mesures ont été fournies par SMS-Demag AG [77]. Plusieurs essais ont été enregistrés, chaque essai est composé de 5 enregistrements (nommés passes). Chaque enregistrement contient six mesures (Couple, Courant et Vitesse) pour les parties supérieure et inférieur du laminoir, puisque c'est un laminoir à double entraînement. Les signaux de couple ont été échantillonnés à une fréquence de 250 Hz.

A cause de la fréquence d'échantillonnage, seules les vibrations basse fréquence (c'est-à-dire ici < 100 Hz), peuvent être détectées [78]. Les fichiers de données sont séparés en deux répertoires, les fichiers de comportement normal, avec lesquels nous avons configuré les méthodes de détection, et ceux en fonctionnement anormal.

Seuls les signaux de mesure du couple haut sont présentés dans ce chapitre, en effet une analyse préliminaire des données a montré que les couples dits haut et bas apportaient la même information sur les vibrations présentes ou non. Les vibrations augmentent avec l'ordre croissant des passes, ainsi l'étude sera limitée à la 5ème passe. La Figure 4.6 illustre l'allure du couple haut pour un enregistrement composé de 5 passes. Cette étude a été réalisée en utilisant le module développé dans le projet européen MAGIC et le logiciel Matlab® avec la toolbox

Wavelab.

3 Résultats avec le test de détection à base de STFT

Premièrement, un signal de couple sans défaut est sélectionné pour être utilisé comme signal de référence afin de régler les paramètres de la procédure de détection. Ce signal est fractionné en plusieurs séquences x_j^{ref}, $i=1:10$ (cf. Figure 4.7). Ensuite les spectres de référence PS_i^{ref} sont calculés avec une fenêtre de Blackman (cf. Tableau 2.1), de longueur $N=512$ points. Les PS_i^{ref} pour $i=1:3$ sont illustrés sur la Figure 4.8.

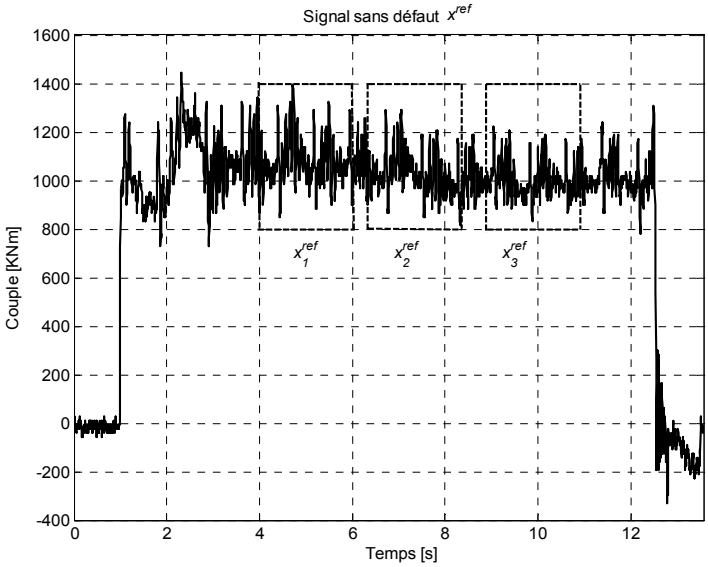


Figure 4.7 — Signal sans défaut

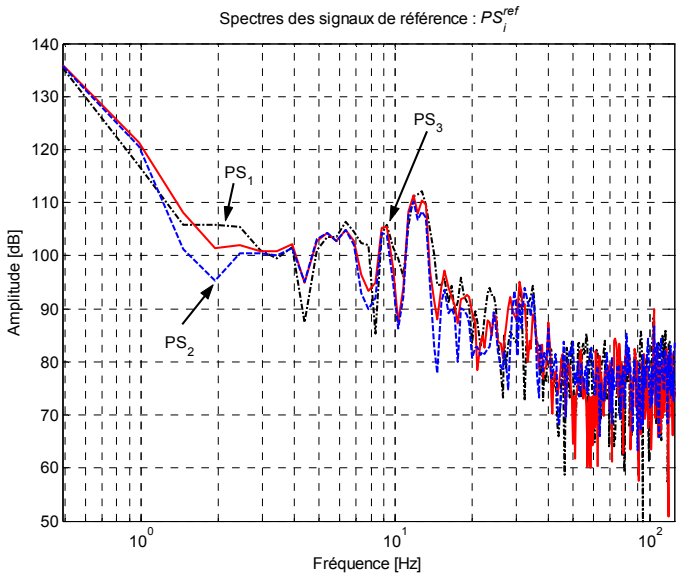


Figure 4.8 — Spectres des signaux de référence

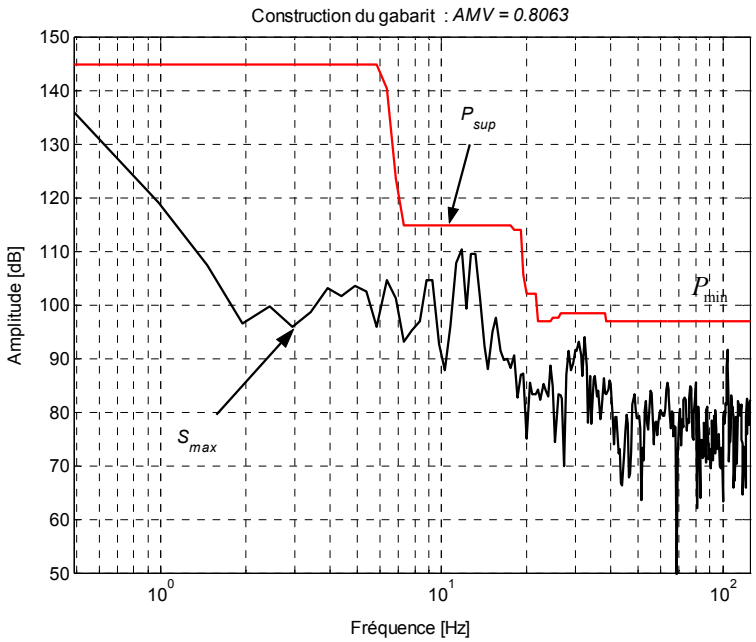


Figure 4.9 — Construction du Gabarit

Après, le gabarit est calculé (Figure 4.9) avec $\alpha_{\inf} = 1$, $\alpha_{\sup} = 0.25$, $P_{\min} = 96$ [dB] et $q = 12$. La validité du modèle (gabarit (3.30)) est $AMV = 0.80$, ce qui veut dire qu'il y a une ressemblance claire entre le contenu fréquentiel des signaux de référence.

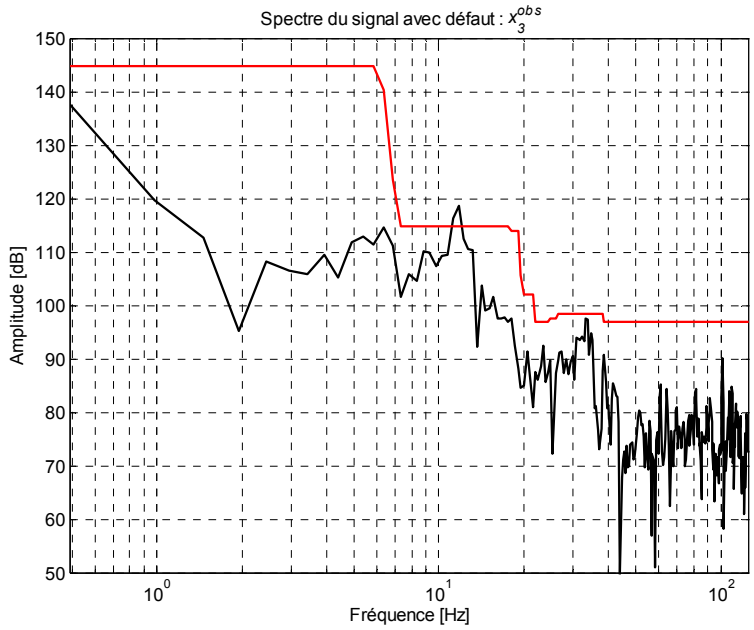


Figure 4.10 — La comparaison par rapport au gabarit du spectre d'un signal avec défaut

Sur la Figure 4.11, on a donnée le résultat de l'agrégation des critères partiels sur chaque fréquence calculé à partir de (3.35). Au tout début de l'enregistrement et à la fin de la passe

$D_{coh} = 0.5$, c'est la phase transitoire du signal. Le test commence après que le couple affranchi une valeur minimale. Cette décision continue à valuer 0.5 à cause nombre insuffisant de points pour calculer le premier spectre.

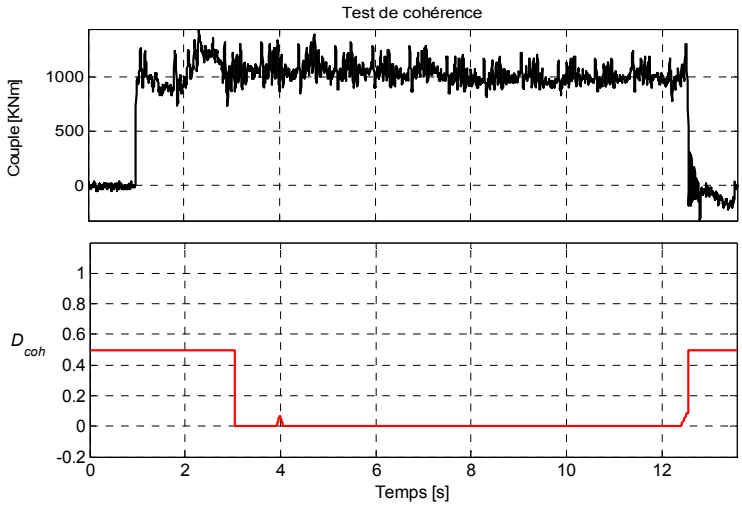


Figure 4.11 — Résultat du test de cohérence, 1^er^ fichier

Notons que ce gabarit a été utilisé pour un test de détection STFT sur un signal considéré sans défaut, et qu'aucune alarme n'a été enregistrée. Cet exemple est illustré sur la Figure 4.12. Sur la partie supérieure de la figure, on a tracé le signal de couple haut, et sur la partie inférieure la génération des symptômes ($D1_D$ à $D3_D$) par les différent opérateurs ((3.19) à (3.21)). Après la phase d'initialisation, de même la valeur des symptômes passe à zéro, ce qui signifie que le contenu spectral du signal observé (*i.e.* le comportement observé) est cohérent avec le contenu spectral du signal de référence. A la fin de la passe (phase transitoire : la tôle quitte les cylindres du laminoir) la valeur des symptômes repasse à 0.5.

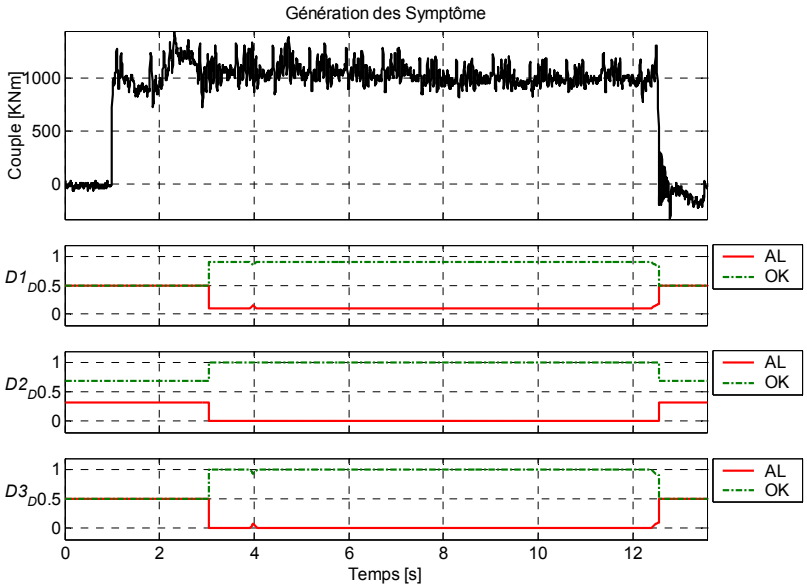


Figure 4.12 — Génération des symptômes, 1^er^ fichier

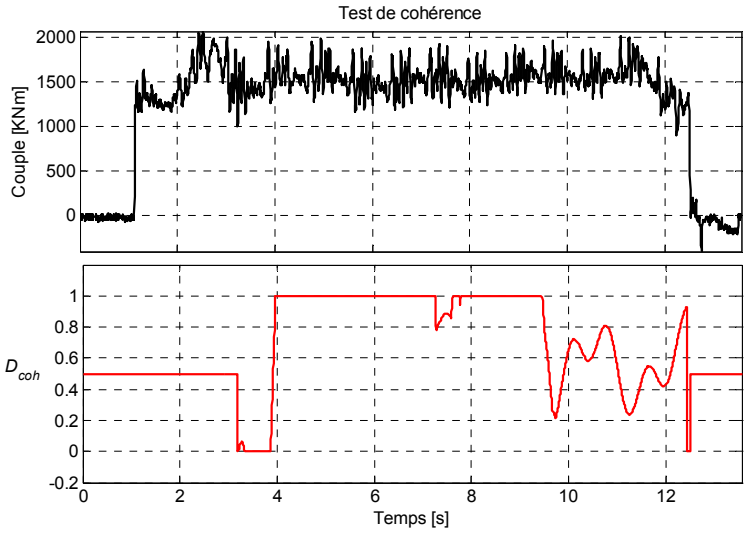


Figure 4.13 — Résultat du test de cohérence, 2ème fichier

Nous avons ensuite procédé à l'analyse d'un fichier correspondant, d'après l'expert, au système en défaut. La Figure 4.13 illustre le couple haut ainsi que le résultat du test de cohérence (3.35), et la Figure 4.14 les valeurs des symptômes, qui au début passent à zéro après la phase transitoire puis repassent à une vleur proche de 1 ($D1_D = 0.9$, $D2_D = 0.8$ et $D3_D = 0.8$) durant une bonne partie de la passe, pour ensuite diminuer progressivement, suite à une diminution de la vitesse du laminoir, jusqu'à repasser à nouveau à zéro avant de passer à 0.5 à la sortie de la tôle d'entre les cylindres du laminoir (fin de passe).

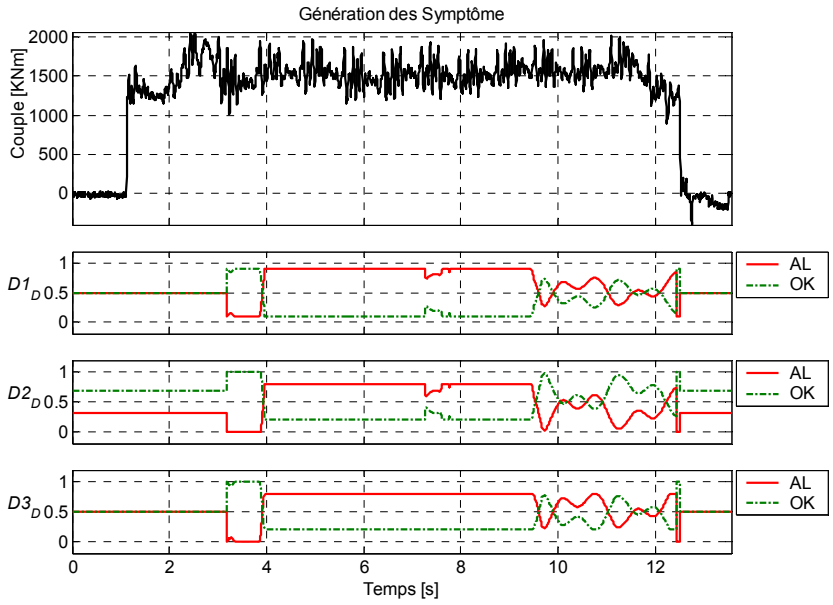


Figure 4.14 — Génération des symptômes, 2ème fichier

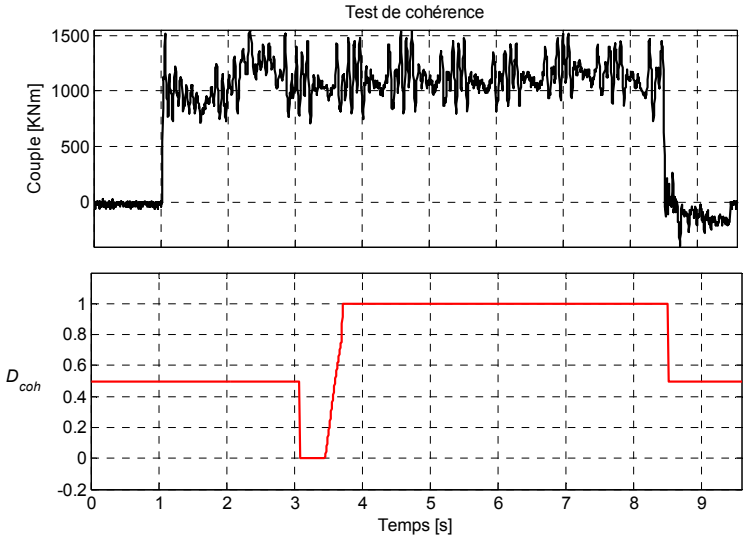


Figure 4.15 — Résultat du test de cohérence, 3ème fichier

Dans le troisième cas, les oscillations deviennent encore plus importantes, le résultat du test de cohérence est quasiment à un depuis le début, il passe à zéro en fin de la période transitoire qui suit l'entrée de la tôle entre les cylindres du laminoir et reste à 1 durant tout le régime permanent avant de repasser à 0.5 quand la tôle quitte les cylindres. Sur la Figure 4.16 les valeurs des symptômes reflètent la situation de défaut installé. En effet, les oscillations n'ont pas tendance à diminuées et les indicateurs d'alarme restent proches de 1. Les valeurs Di_D plus fortes qu'en Figure 4.14.

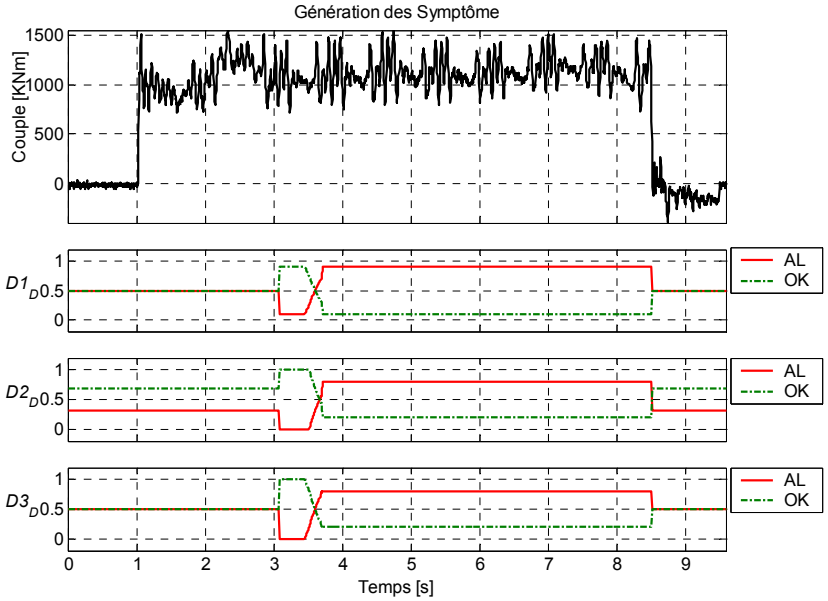


Figure 4.16 — Génération des symptômes, 3ème fichier

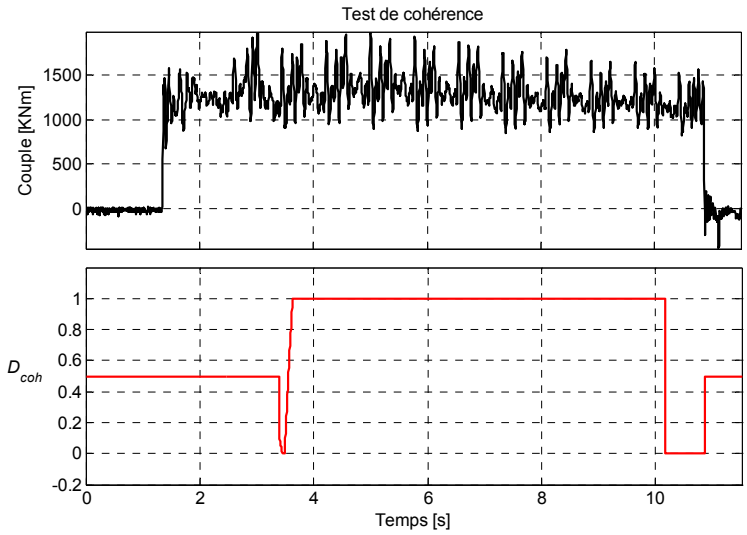


Figure 4.17 — Résultat du test de cohérence, 4ème fichier

Enfin, pour le quatrième fichier, les fluctuations du couple sont les plus importantes, leur valeur crête à crête et de l'ordre de 1000 [KNm] alors que le couple nominal et de l'ordre de 1200 [KNm] ce qui veut dire près de 83%. Les valeurs des symptômes reflètent bien la situation puisque leur valeur passe immédiatement à un et garde cette valeur durant la quasi-totalité de la passe avant de repasser à la fin de la passe à zéro suite à une diminution de l'amplitude des oscillations. Cette diminution a été obtenue par l'opérateur en diminuant la vitesse du laminage.

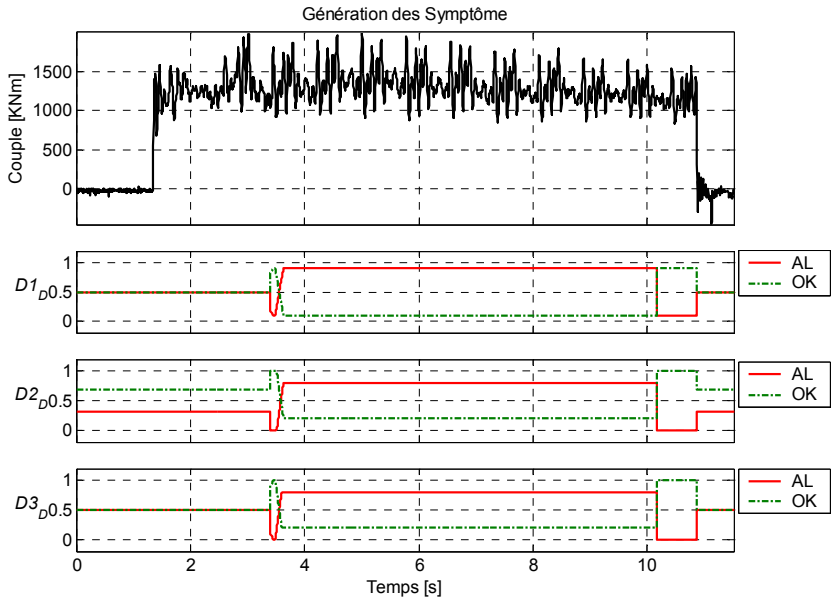


Figure 4.18 — Génération des symptômes, 4ème fichier

4 Résultats avec le test de détection à base de SWT

Quand on utilise les méthodes temps-échelle, une relation entre l'échelle et la fréquence est exprimée à travers une pseudo-fréquence ν_a d'oscillation de l'ondelette correspondant à une échelle particulière *a*. ν_a est la fréquence d'oscillation de l'ondelette à l'échelle *a,* elle peut être calculée à l'aide de la fréquence centrale normalisée de l'ondelettes ν_c et la période d'échantillonnage du signal à analyser :

$$\nu_a = \frac{\nu_\psi}{a \times T_e} \quad \text{avec} \quad T_e = 1/\nu_e \tag{4.1}$$

L'idée principale est d'associer à l'ondelette une sinusoïde de fréquence ν_ψ qui maximise le module de sa transformée de Fourier [42]. La Figure 4.19, illustre l'ondelette *Symmlet6* et la sinusoïde associée. Comme on peut le voir, l'approximation de la fréquence centrale capture l'oscillation principale de l'ondelette. Ainsi la fréquence centrale est une simple caractérisation de la fréquence dominante de l'ondelette.

Une analyse fréquentielle préliminaire a montré que les vibrations torsionnelles qui doivent être détectées se produisent principalement autour de 11.6 Hz. Par conséquent, l'ondelette la mieux adaptée doit avoir une pseudo-fréquence proche de cette fréquence. L'ondelette "*Symmlet6*" "oscille" avec une fréquence très proche au 4ème niveau de décomposition. Ce choix n'est pas unique (l'ondelette Coiflet2 possède la même propriété) ; de plus cette fréquence est aussi liée à la période d'échantillonnage. Le nombre maximal de niveau de décomposition doit donc être supérieur à quatre. Mais il doit être limité en raison du coût calcul. Pour cette application, quelques expériences ont montré que 5 niveaux de décomposition sont suffisants pour garantir une bonne détection des oscillations concernant les vibrations dans les signaux avec défaut.

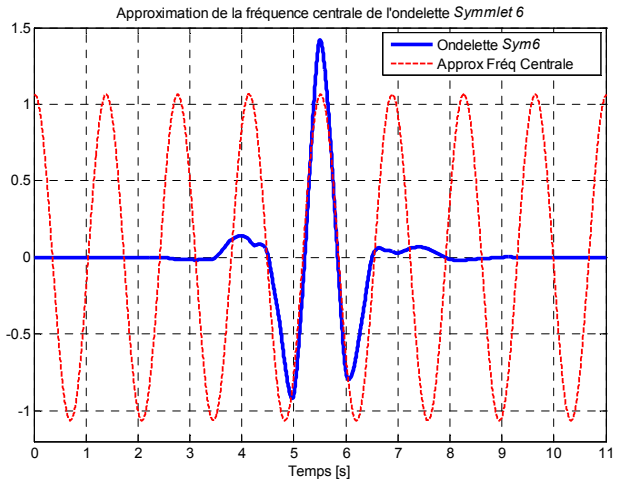


Figure 4.19 — Approximation de la fréquence centrale de l'ondelette *Symmlet6*

La Figure 4.20 illustre le signal du couple haut de la 5ème passe pour un enregistrement sans défaut (cf. Figure 4.7), la décomposition du signal en coefficients d'ondelettes (5 niveaux de

décomposition) ainsi que les seuils calculés comme précédemment (3.40). On voit que la valeur de la validité du modèle (3.41) *AMV = 0.822* est très proche de celle calculée précédemment pour la méthode à base de STFT.

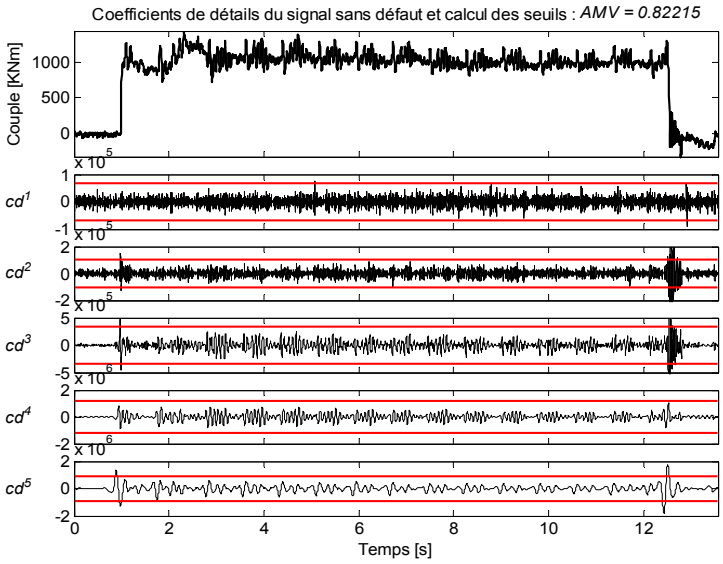


Figure 4.20 — Coefficients SWT : signal sans défaut et calcul des seuils

La Figure 4.21 affiche les coefficients du signal observé.

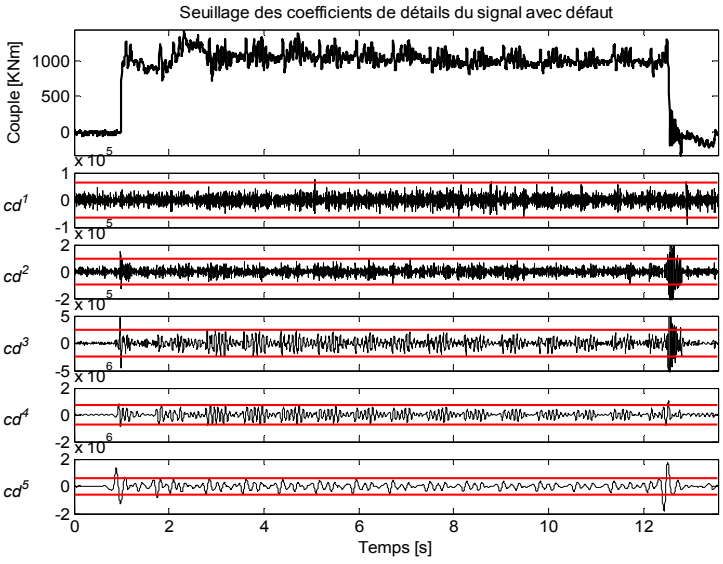


Figure 4.21 — Coefficients SWT: signal sans défaut, 1er fichier

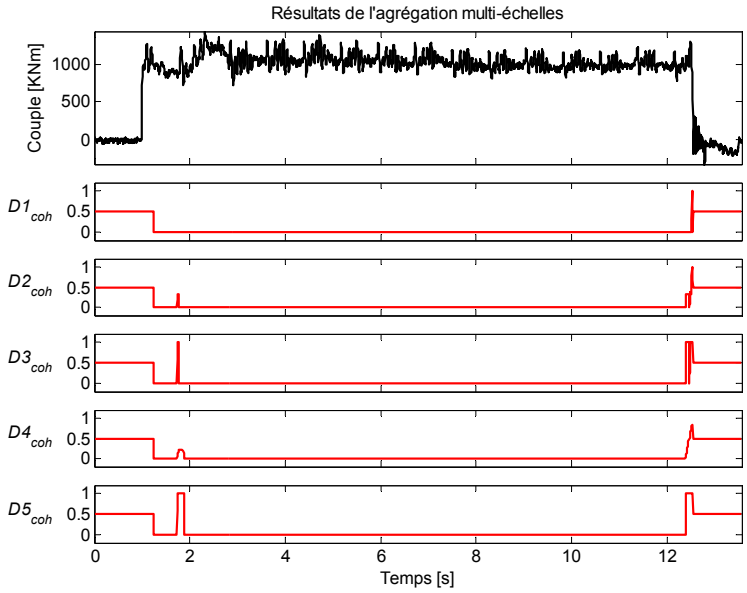


Figure 4.22 — Résultats de l'agrégation multi-échelles, 1er fichier

La Figure 4.22 illustre les résultats de l'agrégation multi-échelles avec les opérateurs (3.43) à (3.47) en considérant seulement les niveaux $j = 3:5$ ($w^j = 1;\ j = 3:5$ et $w^j = 0;\ j = 1:2$). Comme on peut le voir, les indicateurs sont partout à zéro à part quelques transitoires au début et à la fin de la passe. En conséquence les symptômes calculés par (3.19) à (3.21) reflètent bien la situation sans défaut (cf. Figure 4.23).

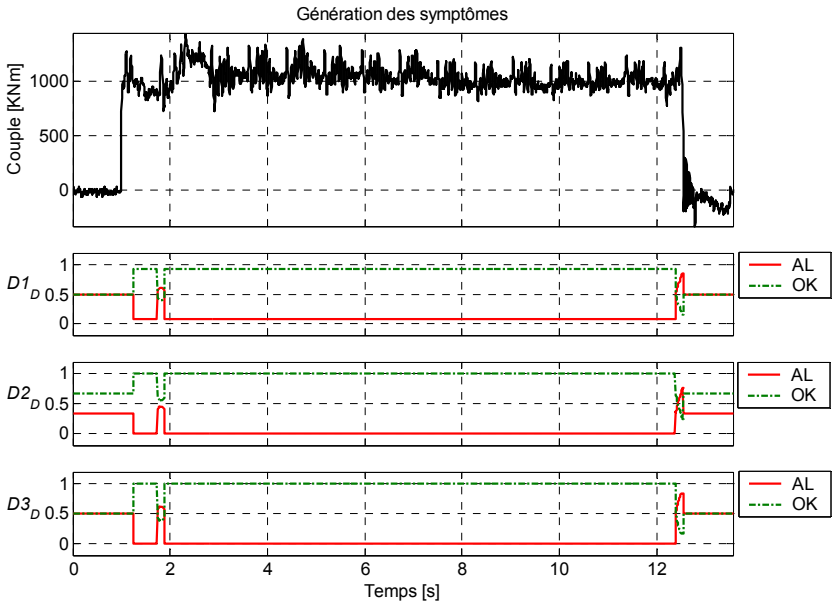


Figure 4.23 — Génération des symptômes, 1er fichier

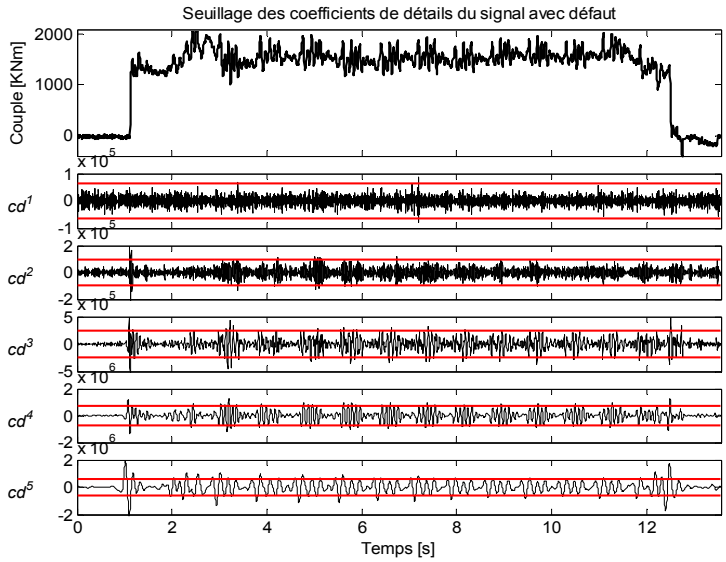


Figure 4.24 — Coefficients SWT: signal avec défaut, 2ème fichier

Le deuxième fichier est représenté sur la Figure 4.24 avec ses coefficients d'ondelettes qui dépassent les seuils notamment aux niveaux 3, 4 et 5. Les indicateurs de cohérence (cf. Figure 4.25) illustrent bien l'apparition des vibrations. On voit que les indicateurs $D1_{coh}$ et $D2_{coh}$ ne sont pas sensible par contre $D3_{coh}$ et $D5_{coh}$ sont très sensibles contrairement à $D4_{coh}$ qui à une allure intermédiaire.

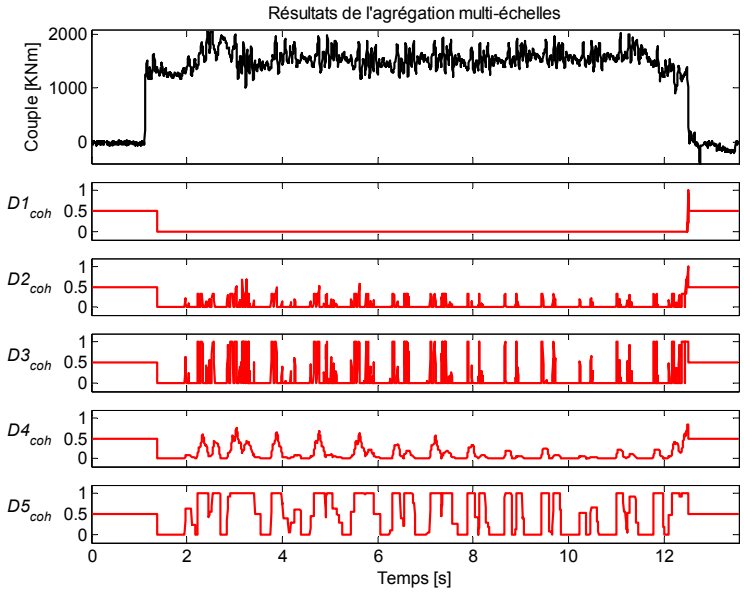


Figure 4.25 — Résultats de l'agrégation multi-échelles, 2ème fichier

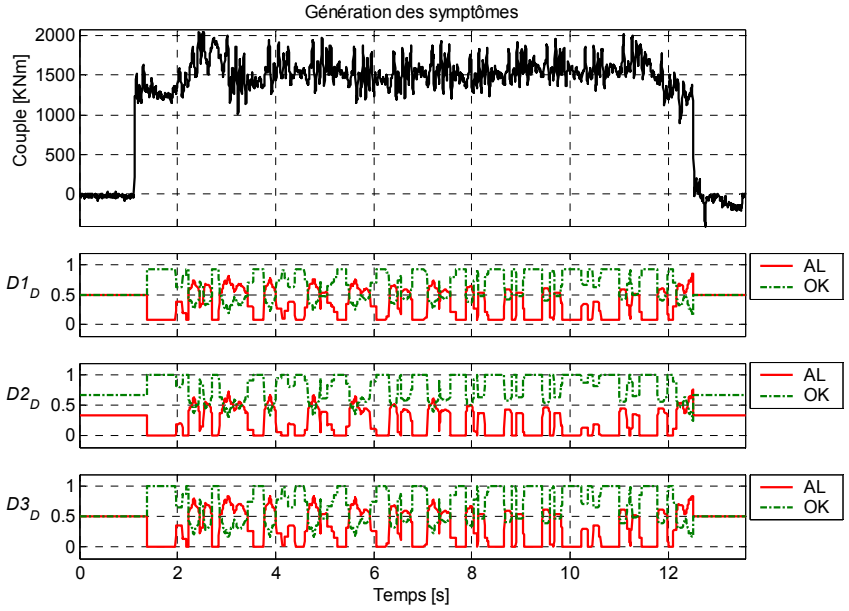


Figure 4.26 — Génération des symptômes, 2ème fichier

Pour le troisième fichier, l'amplitude des vibrations a un peu plus augmenté, on peut le constater à partir des coefficients des détails qui dépassent les valeurs de seuil calcules précédemment (cf. Figure 4.27). La Figure 4.28 illustre la persistance des indicateurs de présence de vibrations générés à partir de l'agrégation multi-échelle des coefficients fuzzifiés. Cependant leur valeur est de moins en moins affectée par les variations locales. L'évolution des indicateurs reflète la pseudo-période des fluctuations du couple autour de sa valeur nominale durant cette passe.

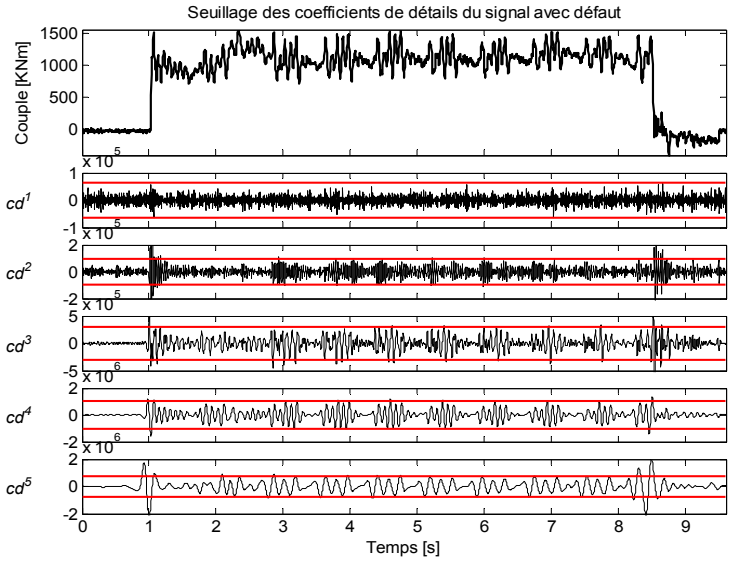


Figure 4.27 — Coefficients SWT: signal avec défaut, 3ème fichier

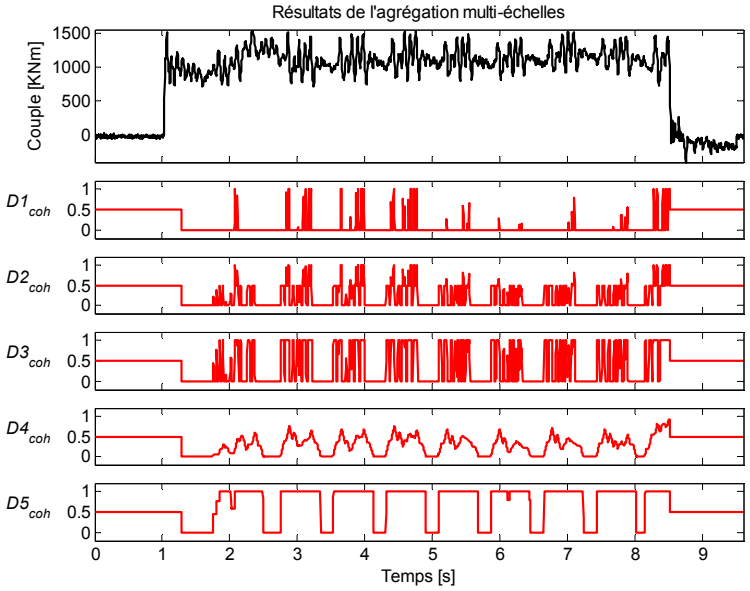


Figure 4.28 — Résultats du test de cohérence, 3ème fichier

Par conséquent, les indicateurs sont de plus en plus persistants (Figure 4.28) et on peut voir que l'indicateur $D5_{coh}$ est à 1 à chaque pseudo-période, $D4_{coh}$ également mais avec une amplitude moins importante. Sur la Figure 4.29, on peut voir le niveau des symptômes qui reflète bien la situation.

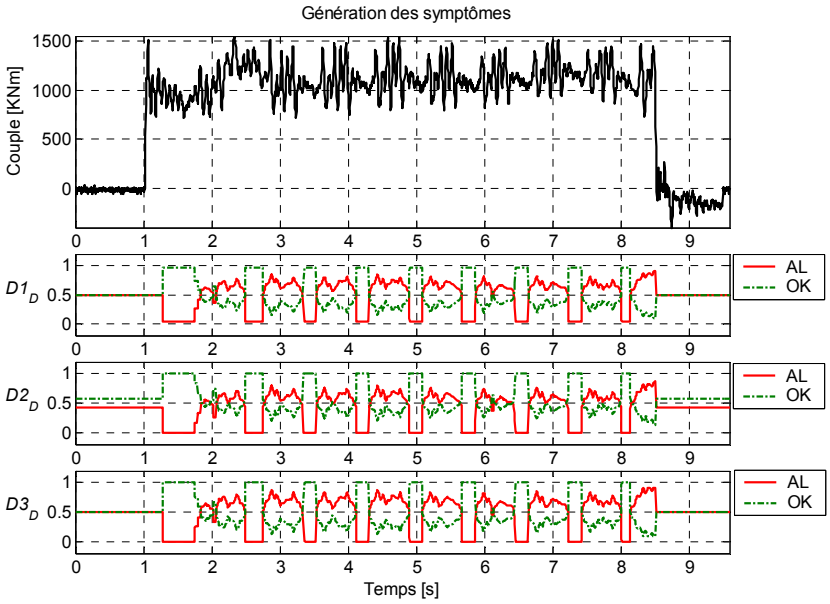


Figure 4.29 — Génération des symptômes, 3ème fichier

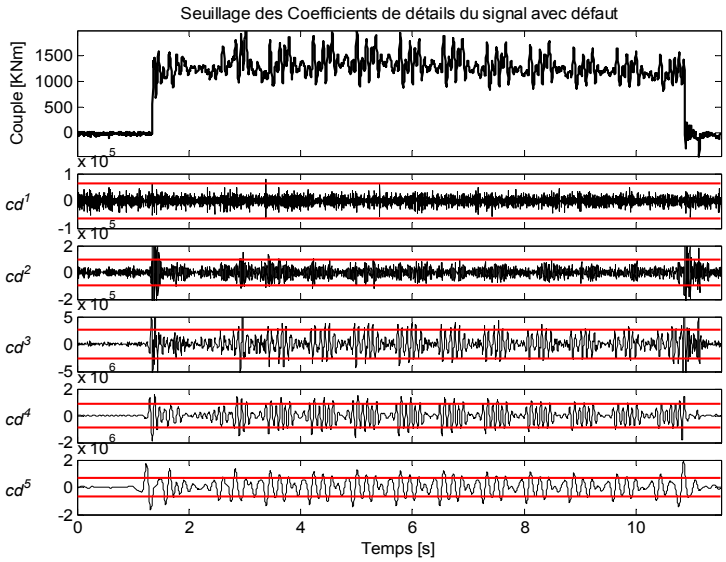


Figure 4.30 — Coefficients SWT: signal avec défaut, 4ème fichier

Dans le dernier fichier où les vibrations sont les plus importantes, leur ampleur est nettement caractérisée par les coefficients en dépassement de seuil, notamment sur le 4ème niveau où se trouve la principale fréquence caractéristique des vibrations torsionnelles mais également sur le 3ème et 5ème niveau (cf. Figure 4.30). Ce constat (Figure 4.31) explique clairement l'amplitude des indicateurs, même les indicateurs instantanées ($D1_{coh}$ à $D3_{coh}$), et par conséquent les valeurs des symptômes qui sont quasiment constants et "grands" durant toute la passe (Figure 4.32).

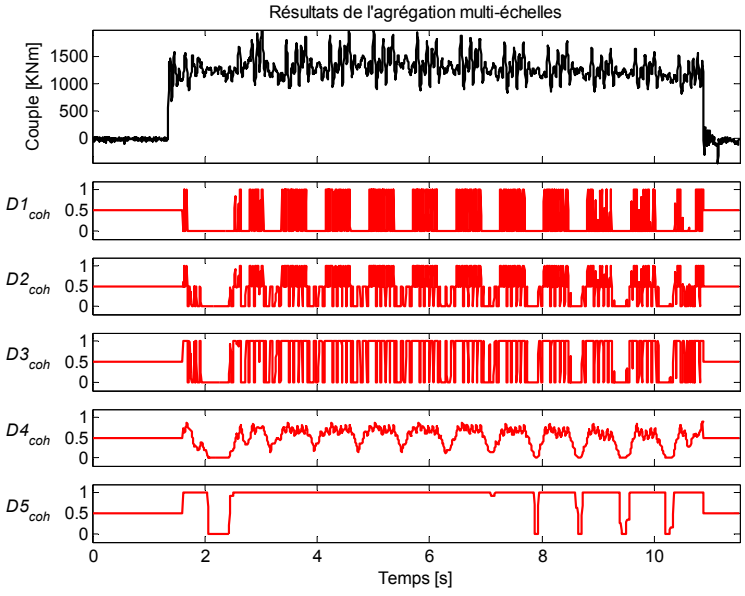


Figure 4.31 — Résultats du test de cohérence, 4ème fichier

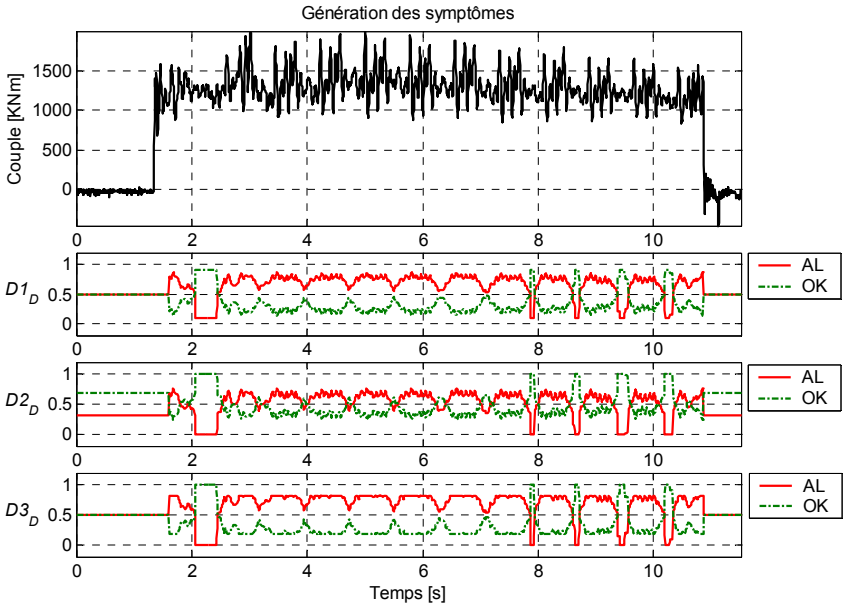


Figure 4.32 — Génération des symptômes, 4ème fichier

5 Conclusion

Le phénomène de vibrations torsionnelles se produisant dans les laminoirs est une cause principale non seulement de dégradation de la qualité du produit, mais également des composants du laminoir et par conséquent il diminue leur durée de vie. Ce phénomène s'auto-entretient, et par conséquent force l'opérateur à diminuer la vitesse du laminoir. En employant des méthodes classiques de traitement du signal (analyse spectrale), on peut déterminer les caractéristiques principales de ces vibrations. Le but de ce travail est de détecter au plus tôt leur apparition avant qu'elles ne s'auto-entretiennent et causent des dégradations importantes.

La méthode que nous avons élaborée (cf. chapitre 3) répond au but fixé. Elle nécessite un minimum d'information sur les signaux analysés, mais elle présente plusieurs paramètres à régler, concernant la fenêtre de pondération pour la STFT, l'ondelette et le nombre de niveaux de décomposition pour la SWT, ainsi que les seuils et les paramètres de fuzzification. En effet, le bon choix de ces paramètres conditionne la robustesse ou la sensibilité de la méthode pour les différents types de changement de comportement des signaux analysés.

La méthode de détection s'est montrée efficace pour tous les fichiers que nous avons analysés, dans le cas où l'ampleur des vibrations a permis une analyse pertinente. Nous avons présenté ici les résultats obtenus en appliquant notre méthode aux signaux mesurés de couple pour différents fichiers. Nous avons décidé de présenter dans ce chapitre uniquement les résultats concernant le couple haut, mais on obtient les mêmes résultats avec les signaux de couple bas. La détection a été réalisée en respectant les contraintes temporelles, à savoir durant la durée de la passe et avant que l'ampleur des vibrations ne soit "trop" importante.

Conclusion Générale

Notre travail s'est focalisé sur la conception de méthodes d'aide à la décision dans le cadre de la détection de défauts. Ces méthodes ont été appliquées à la détection des vibrations torsionnelles dans un laminoir dégrossisseur réversible à chaud.

Un test de détection repose nécessairement sur un composant ou un ensemble de composant (un sous-système), il faut adjoindre au modèle de ce sous-système un domaine de validité correspondant à l'intersection de la validité de ses relations comportementales élémentaires. Le domaine de validité d'un modèle est important pour la fiabilité d'un test de détection.

Nous avons montré qu'il était possible d'appréhender des incertitudes de décision au niveau des tests de détection. Tous les tests ne conduisent pas toujours à des décisions sûres. L'incertitude se caractérise par l'incapacité de savoir si un test de détection est vrai ou faux. Les tests de détection mis en œuvre fournissent des décisions comprises entre 0 et 1 où 1 représente une alarme sûre et 0 représente qu'il n'y a aucune alarme d'une manière sûre.

Nous avons utilisé des méthodes classiques de traitement du signal, pour extraire les caractéristiques générales des signaux étudiés, telles que la transformée de Fourier rapide glissante et la transformée en ondelettes stationnaire. Ces informations permettent de caractériser les signaux en l'absence de défauts et nous ont servi pour établir les paramètres à régler, tel que le type et la longueur de la fenêtre de pondération pour la FFT glissante, le nombre de niveaux de décomposition ainsi que l'ondelette mère pour la transformée en ondelettes stationnaire.

Ces méthodes permettent la localisation autant temporelle que fréquentielle des changements produits dans les signaux analysés. La transformée de Fourier détecte facilement les composantes fréquentielles. En contrepartie, la transformée en ondelettes stationnaire permet des décompositions successives du signal avec différentes résolutions temporelles et fréquentielles. Suivant l'ondelette mère utilisée, elle permet une bonne localisation temporelle et fréquentielle.

Dans le but du diagnostic que nous nous sommes fixé, une analyse fixe des caractéristiques des signaux en présence de défaut est nécessaire. La transformée en ondelettes continue peut nous fournir une telle analyse, mais elle engendre aussi une redondance de l'information. Le besoin d'information dans le diagnostic d'un défaut ne justifie pas le coût très élevé de calcul induit par la redondance de la transformée en ondelettes continue.

Nous avons donc choisi de faire appel à la transformée en ondelettes discrète pour analyser l'évolution des signaux mesurés sans et/ou avec défaut au cours du temps. Parmi les différents types de transformée en ondelettes discrètes, la transformée en ondelettes stationnaire présente un intérêt particulier, car elle est facilement implantable avec de filtres numériques ; de plus elle présente la propriété de covariance par translation temporelle contrairement à la transformée en ondelettes discrète dyadique. Une condition d'efficacité de la procédure de détection impose que les filtres d'analyse soient suffisamment courts pour permettre une implantation en ligne rapide.

Si des connaissances a priori sur le type de défaut à détecter sont disponibles, alors ces connaissances imposent le nombre de niveaux de décomposition. Sinon, le nombre de niveaux de décomposition est imposé par la période d'échantillonnage et la vitesse de calcul, en tenant compte du nombre d'opérations maximum à effectuer pour obtenir le symptôme. Les informations obtenues par l'analyse sont contenues dans les coefficients d'ondelettes calculés au cours du temps à chaque échelle (ou niveau) de décomposition.

La détection en ligne des défauts implique le suivi de l'évolution du signal à chaque instant, c'est à dire à chaque échantillon mesuré. Un seuil adapté à chaque échelle de décomposition sépare les coefficients associés au défaut des autres coefficients. Ce seuil est fixé en fonction de l'écart-type et de la longueur des vecteurs de coefficients de détail.

La fuzzification des coefficients de décomposition de chaque niveau permet ensuite de les agréger en ligne pour la détection, afin de générer un symptôme non nul dès l'apparition d'un changement de comportement du signal. La rapidité de la détection est fortement liée à la propriété de localisation temporelle de l'ondelette mère choisie pour l'analyse.

La méthode de détection que nous avons élaborée est applicable à plusieurs types de situations. Elle nécessite un minimum d'information sur les signaux analysés, mais elle présente plusieurs degrés de liberté en ce qui concerne les choix de l'ondelette mère, du nombre de niveaux de décomposition, et de la méthode de seuillage des coefficients d'ondelettes.

En effet, cette dernière caractérise la sensibilité de la méthode de détection à différents types de changements de comportement du signal analysé. La méthode de détection s'est montrée efficace pour tous les signaux que nous avons analysés, dans le cas où leur niveau a permis une analyse pertinente.

En ce qui concerne la méthode de détection, d'autres techniques de seuillage pourraient également être employées. Le choix de la méthode de seuillage dépend essentiellement de l'écart entre les valeurs des coefficients d'ondelettes associées à l'apparition d'un défaut et les valeurs de tous les autres coefficients.

De plus, s'il existe des informations concernant la structure des changements induits dans le signal par le défaut, celles-ci doivent être exploitées pour choisir une ondelette mère adaptée au

type de défaut attendu.

Une solution future est de proposer une méthode de changement de modèles lorsque celui-ci n'est plus valide par suivi de point de fonctionnement, pour garder une décision progressive même lorsque le modèle initial n'est plus valide. Plus une méthode de localisation et d'identification de défauts dans le cas de multiples défauts.

Bibliographie

[1] Basseville, M. (1988). Detecting changes in signals and systems—a survey. Automatica 24 (3), 309–326.

[2] Basseville, M., & Benveniste, A. (1986). Detection of abrupt changes in signals and dynamic systems (Lecture Notes in Control and Information Sciences: 77). Berlin: Springer-Verlag.

[3] Basseville, M., & Nikiforov, I. V. (1993). Detection of abrupt changes—theory and application. Information and system sciences series. Prentice Hall.

[4] Bakshi, B., Stephanopoulos, G. (1992). Temporal representation of process trends for diagnosis and control. In IFAC symposium on online fault detection and supervision in the chemical process industry, Newark, DE, USA, 69–74.

[5] Blanke, M. Kinnaert, M. Lunze, J. Staroswiecki, M. (2003). Diagnosis and Fault-tolerant Control. Springer Verlag.

[6] Chen, J., & Patton, R. J. (1999). Robust model-based fault diagnosis for dynamic systems. Massachusetts: Kluwer Academic Publishers.

[7] Cheung, J. T., & Stephanopoulos, G. (1990). Representation of process trends part I. A formal representation framework. Computers and Chemical Engineering 14 (4–5), 495–510.

[8] Charbonnier, S. Garcia-Beltran, C. Cadet, C. Gentil, S. (2005). Trends Extraction and analysis for complex systems monitoring and decision support. Engineering Applications of Artificial Intelligence 18 (1), 21–36.

[9] Dubuisson, B. (1990). Diagnostic et reconnaissance des formes. Paris, Hermès Science Publications.

[10] Dubuisson, B. (2001). Automatique et statistiques pour le diagnostic. Paris, Hermès Science Publications.

[11] Dubuisson, B. (2001). Diagnostic, intelligence artificielle et reconnaissance des formes. Paris, Hermès Science Publications.

[12] Evsukoff, A. (1998). Le raisonnement approché pour la surveillance de procédés. PhD thesis, INPG, Grenoble, France.

[13] Frélicot, C. (1992). Un système adaptatif de diagnostic prédictif par reconnaissance de formes floues. PhD thesis, UTC, Compiègne, France.

[14] Gentil, S. (2003). Les procédés agroalimentaires. 2, Commande et supervision; Chapitre 4 : Supervision et Diagnostic. Hermès, Paris.

[15] Heim, B. (2003). Approches ensembliste et par logique floue pour le diagnostic causal de procédés de raffinage. Application à un pilote de FCC. PhD thesis, INPG, Grenoble, France.

[16] Janusz, M., & Venkatasubramanian, V. (1991). Automatic generation of qualitative description of process trends for fault detection and diagnosis. Engineering Applications of Artificial Intelligence 4 (5), 329–339.

[17] Lesecq, S. Barraud, A. (2000). Fault detection using on-line wavelet analysis: application to induction motor. In 4th Symposium on Fault Detection, Supervision and Safety for Technical Process, Hongrie.

[18] MacGregor, J. F., & Kourti, T. (1995). Statistical process control of multivariate processes. Control Engineering Practice 3 (3), 403–414.

[19] Maquin, D. Ragot, J. (2000). Diagnostic des Systèmes linéaires. Paris, Hermès Science Publications.

[20] Melendez J, Colomer J, "Episodes representation for supervision. Application to diagnosis of a level control system", in preprints of DX'01, Twelfth International Worksop on principles of diagnosis, Eds: S. McIlraith & D.T. Dupre, March 2001.

[21] Montmain, J. Gentil, S. (2000). Dynamic causal model diagnostic reasoning for on-line technical process supervision. Automatica, 36:1137–1152.

[22] Montmain, J. Gentil, S. (1991). Qualitative event analysis for fault diagnosis. Proc. 1st IFAC Safeprocess, Baden-Baden, Germany.

[23] Nomikos, P., & MacGregor, J. (1994). Monitoring batch processes using multiway principal component analysis. American Institute of Chemical Engineers Journal 40 (8), 1361−1375.

[24] Petropol, S. (2001). Ondelettes et diagnostic, application aux defauts dièlectriques et électriques des machines tournantes. PhD thesis, INPG, Grenoble, France.

[25] Travé-Massuyès, L. Dague, P. (2003). Modèles et raisonnements qualitatifs. Lavoisier, Paris.

[26] Travé-Massuyès, L. Dague, P. Guerrin, F. (1998). Le raisonnement qualitatif. Hermès, New York.

[27] Venkatasubramanian, V. Rengaswamy, R. Yin, K. Kavuri, S.N. (2003). A review of proccess fault detection and diagnosis Part I: Quantitative model-based methods. Computers and Chemical Engineering. 27, (3), 293-311.

[28] Venkatasubramanian, V. Rengaswamy, R. Kavuri, S.N. (2003). A review of process fault detection and diagnosis Part II : Qualitative models search strategies. Computers and Chemical Engineering. 27, (3), 313-326.

[29] Venkatasubramanian, V. Rengaswamy, R. Kavuri, S.N. (2003). A review of process fault detection and diagnosis Part III : Procédé history based methods. Computers and Chemical Engineering. 27, (3), 327-346.

[30] Villemeur, A. (1988). Sureté de fonctionnement des systèmes industriels. Fiabilité Facteurs humains Informatisation. Ed. Eyrolles.

[31] Youssef, B. Alamir, M. (2003). Generic signature generation tool for diagnosis and parametric estimation of multi-variable dynamical nonlinear systems. In Proceedings of the IEEE Conference on Decision and Control, Hawaï, USA.

[32] Zwingelstein, G. (1995). Diagnostic des Défaillances – Théorie et Pratique pour les Systèmes Industrielles. Paris, Hermès Science Publications.

[33] Köppen-Seliger, B., Frank, P.M., & Wolff A., (1995). Residual evaluation for fault detection and isolation with RCE neural networks. In Proceedings of the American control conference, Seattle, Washington, pp. 3264–3268.

[34] Köppen-Seliger, B. Marcu, T. Capobianco, M. Gentil, S. Albert, M. Latzel, S. (2003). Magic: an integrated approach for diagnostic data management and operator support, IFAC Symposium Safeprocess 2003, Washington, USA, 9-11 June 2003.

[35] Chow, E. Y., & Willsky, A. S. (1984). Analytical redundancy and the design of robust failure detection systems. IEEE Transactions on Automatic Control 29 (7), 603–614.

[36] Colomer J. & Melendez J., "A family of FIR differentiators based on a polynomial least squares estimation", Proceeding of the European Control Conference (ECC2001), pp: 2802-2807, Oporto (Portugal), 4-7 Sept, 2001.

[37] Gertler, J., & Luo, Q. (1989). Robust isolable models for failure diagnosis. AIChE 31 (11), 1856–1868.

[38] Gertler, J., Fang, X., & Luo, Q. (1990). Detection and diagnosis of plant failures: the orthogonal parity equation approach. Control and Dynamic Systems 37, 159–216.

[39] Patton, R., Frank, P., & Clark, R. (1989). Fault Diagnosis in Dynamic Systems: Theory and Applications. New York: Prentice Hall.

[40] Potter, J.E. & Suman, M.C. (1977). Thresholdless redundancy management with arrays of skewed instruments. Integrity in Electronic Flight Control Systems AGARDOGRAPH-224, 15-11 to 15-25.

[41] Vedam, H., Venkatasubramanian, V. (1997). A wavelet theory-based adaptive trend analysis system for process monitoring and diagnosis. In American control conference (pp. 309–313).

[42] Abry, P. (1997). Ondelettes et Turbulences, Multirésolutions, algorithms de decomposition, invariance d'échelle et signaux de pression. Diderot, Arts et Sciences, Paris, New York.

[43] Blanchet, G. Charbit, M. (1998). Traitement numérique du signal, simulation sous Matlab®. Hermès, Paris.

[44] Boashash, B. (edited by). (2003). Time Frequency Signal Analysis and Processing, a comprehensive reference. Elsevier.

[45] Flandrin, P. (1998). Temps-Fréquence. Hermès, Paris.

[46] Cooley, J.W. Tuckey, J.W. (1965). An algorithm for the machine computation of the complex Fourier series. Math. Comput. Vol, 19. p, 297-301.

[47] http://www.fftw.org/

[48] Max, J. Lacoume, J.L. (2004). Méthodes et techniques de traitement du signal, Paris, Dunod.

[49] Hlawatsch, F., & Auger, F. (2005). Temps-fréquence, concepts et outils. Lavoisier, Paris.

[50] Mallat, S. (1989). A theory for multiresolution signal decomposition : the wavelet representation, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 11, p. 674–693.

[51] Vetterli, M. Herley, C. Wavelets and Filter Banks ; Theoty and Design. IEEE Transaction on Signal Processing, vol. 40(9), p. 2207–2232.

[52] Mallat, S. (2000). Une exploration des signaux en ondelettes. Editions de l'école polytechnique, Palaiseau.

[53] Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J-M. (2003). Les ondelettes et leurs applications. Lavoisier, Paris.

[54] Strang, G., & Nguyen, T. (1996). Wavelets and filter bancs. Wellesley-Cambridge Press.

[55] Truchetet, F. (1998). Ondelettes pour le signal numerique. Hermès, Paris.

[56] SATE. (2005). GPMAS User's Manual. S.A.T.E. S.r.l. Systems & Advanced Technologies Engineering. EU-IST-2000-30009 MAGIC.

[57] Pesquet, J.C., Krim, H., Carfatan, H. (1996), Time-Invariant orthonormal wavelet representations, IEEE Trans. Signal Processing, Vol. 44, pp. 1964-1970.

[58] Bellanger, M. (2002). Traitement numérique du signal. Paris, Dunod.

[59] Nason, G.P. Silverman, B.W. (1995). The stationary wavelet transform and some statistical applications. Lecture Notes in Statistics, 103, pp. 281-299.

[60] Daubechies, I. (1992). Ten Lectures On Wavelets. SIAM, Philadelphia.

[61] Vidakovic, B. (1999). Statistical Modeling by Wavelets. John Wiley & Sons, Inc.

[62] Touaf, S. (2005). Diagnostic logique des systèmes complexes dynamiques dans un contexte Muli-Agent. PhD thesis, UJF, Grenoble, France.

[63] Zimmermann, H-J. (1996). Fuzzy Set Theory and its Applications. Kluwer Academic Publishers.

[64] Dubois, D. Prade, H. (1985). A review of fuzzy set aggregation connectives. Information Sciences, 36, pp. 85-121.

[65] Lesecq, S. Gentil, S. Barraud, A. (2005). Diagnostic Symptom Generation Aggregating Model Validity and Residual Evaluation. 44th IEEE Conference on Decision and Control and European Control Conference ECC'05, Seville, Spain.

[66] Lesecq, S. Taleb, S. Gentil, S. Fault detection based on wavelet transform and fuzzy decision making. Advanced Control and Diagnosis, ACD' 05. Mulhouse, France.

[67] Taleb, S. (2003), Diagnostic et transformée en ondelettes, DEA Report, LAG, INPG, Grenoble, France.

[68] Bouchon-Meunier, B. (2003). Logique floue, principes, aide à la décision. Paris, Hermès science, Lavoisier.

[69] Mallat, S. Hwang, W.L. (1992), Singularity Detection and Processing with Wavelets, IEEE Trans. on Information Theory, Vol. 38(2), pp. 617-643.

[70] Mallat, S. Zhong, S. (1992), Characterization of Signals from Multiscale Edges, IEEE Trans. on Pattern Anal. and Machine Intelligence, Vol. 14, pp. 710-732.

[71] Donoho, D.L. Johnstone, I.M. (1994), Ideal Spatial Adaptation by Wavelet Shrinkage, Biometrika, Vol. 81(3), pp 425-455.

[72] Donoho, D.L. (1995), De-Noising by soft-Thresholding, IEEE Trans. On Information Theory, Vol. 41(3), pp. 613-627.

[73] Coifman, R.R, Donoho, D.L. (1995), Translation Invariant De-noising, series Lecture Notes in Statistics, New York Springer-Verlag, Vol. 103, pp. 125-150.

[74] Donoho, D.L. Johnstone, I.M. (1995), Adapting to Unkown Smoothness via Wavelet Shrinkage, Biometrika, Vol. 90(432), pp 1200-1224.

[75] Ravier, Ph. (1998). Détection de Transitoires par Ondelettes adaptées. PhD thesis, LIS-INPG, Grenoble, France.

[76] Taleb, S. Lesecq, S. Stuecher, R. (2004). Torsional vibration detection using torque measurement of a main drive of a hot rolling mill with wavelet transform, MMM'04, 11th IFAC Symposium on Automation in Mining, Mineral and Metal processing, Nancy, France.

[77] Deckers, J. (2003). Measured data for signal based Diagnostic Agents, SMS-Demag. MAGH003R01, EU-IST-2000-30009 MAGIC.

[78] Deckers, J. Jepsen, O. Latzel, S. Metz, H. Stuecher R. (2003). Condition monitoring and failure diagnosis in plants of the metal industry. IFAC Symposium Safeprocess June 2003, Washington, USA, 9-11.

[79] Deckers, J. (2001). Entwicklung einer Low-Cost Körperschallsensorik zur Überwachung des Verschleißverhaltens von wälz- oder gleitgelagerten Kreiselpumpen kleiner Leistung. Doktor-Ingenieur genehmigte Dissertation. Gerhard-Mercator-Universität Duisburg.

[80] Lesecq, S. Gentil, S. Taleb, S. Fagarasan, I. (2005). Diagnostic Decision Algorithm Based On Spectral Analysis and Model Validity. IEEE Conference on Control Applications, CCA'05, Toronto, Ontario, Canada.

[81] Taleb, S. Lesecq, S. Gentil, S. (2006). Fault Detection Based on Wavelet Transform Application to a Roughing Rolling Mill. IFAC Safeprocess'06, Benjing, P.R. China.

[82] Seelinger, A. Mackel, J. Georges, D. (2002). Measurement and diagnosis of preprocess-disturbing oscillations in high-speed rolling plants, IMEKO, Tampere, Finland.

[83] Montmitonnet, P. Laminage, Objectifs et modélisation. Techniques de l'ingénieur, traité Materiaux métaliques, M3065.

[84] Montmitonnet, P. Laminage à chaud, Théorie du laminage. Techniques de l'ingénieur, traité Materiaux métaliques, M7840.

[85] Mackel, J. Flieweger, M. Asch, A. (2002), Maintenance and quality related monitoring of rolling mill main drives, SARUC, Vanderbiljpark, South Africa, 2002.

[86] Boulenger, A. Pachaud, C. (2003). Analyse vibratoire en maintenance: surveillance et diagnostic des machines. Ed, Dunod.

[87] Combastel, C. Lesecq, S. Petropol, S. Gentil, S. (2002). Model-based and wavelet approaches for induction drive on-line fault detection. Control Engineering Practice, Vol. 10, p. 493-509.

[88] Montmain, J. Gentil, S. (1996). Operation support for alarm filtering. "CESA'96, IEEE-IMACS", Multiconferences on "Computational Engineering in Systems Applications", Lille, France.

GÉNÉRATION DE SYMPTÔMES FLOUS PAR FFT ET SWT

RÉSUMÉ

L'objectif de ce travail est la détection des défauts à travers la détection des changements ou des anomalies dans les signaux mesurés. Notre travail présente une méthode permettant de construire, à partir d'historique de données (signaux de référence), un modèle comportemental du procédé, puis de détecter des situations anormales issues des dysfonctionnements. Cette méthode consiste à générer des symptômes flous en ligne permettant de suivre l'évolution temporelle du fonctionnement du système à surveiller. Ces symptômes prennent en considération les incertitudes sur la décision en fusionnant le résultat du test de détection avec celui de sa validité. Par leurs propriétés de localisation fréquentielle et temps-fréquence, la transformée de Fourier à court terme et la transformée en ondelettes stationnaires sont deux méthodes adaptables au diagnostic. Les choix de la longueur et du type de la fenêtre de pondération pour la STFT, de l'ondelette mère et du nombre de niveaux de décomposition pour la SWT sont les degrés de liberté qui leur permettent d'extraire des informations pertinentes des signaux analysés. Un symptôme graduel est obtenu en fonction du type et de l'ampleur du défaut. La méthode de détection conçue a été testée sur des signaux contenant des défauts simulés, puis a été ensuite validée sur des données industrielles pour la détection des vibrations torsionnelles dans un laminoir réversible (laminage à chaud). Les degrés de liberté de la méthode proposée confèrent aux méthodes développées la flexibilité par rapport à différentes applications.

Mots-clés

Détection de défaut, Transformée de Fourier à Court Terme, Transformée en Ondelettes Stationnaires, Agrégation floue, Aide à la décision floue, Gabarit, Seuillage, Laminoir réversible à chaud.

FUZZY SYMPTOMS GENERATION BY FFT AND SWT

ABSTRACT

The purpose of this work is fault detection through changes detection or anomalies in the measured signals. Our work presents a method making possible to build a behavioral model of the process, using historical process data (reference signals), in order to detect abnormal situations resulting from dysfunctions. This method consists in generating on line progressive symptoms, making possible to follow the temporal evolution of the system operation to be supervised. These symptoms take into account uncertainties on the decision by aggregating the result of the detection test with that of its validity. By their properties of frequency and time-frequency localization, Short Time Fourier Transform and Stationary Wavelets Transform are two methods adaptable to fault diagnosis. The choices of the weighting window length and type for the STFT, of the mother wavelet and the number of decomposition levels for the SWT are the freedom degrees which allow this method to extract relevant information from the analyzed signals. A gradual symptom is obtained according to the type and from of the extent fault. The fault detection method conceived was initially tested on signals containing simulated faults was then validate on industrial signals to detect the torsional vibrations in a reversing rolling mill (hot rolling). The freedom degrees of the suggested method confer flexibility to the developed methods with regard to different applications.

Keywords

Fault Detection, Short Time Fourier Transform, Stationary Wavelets Transform, Fuzzy Aggregation, Fuzzy decision making, Spectrum pattern, Thresholding, Hot Rolling Mill.