

Des réseaux de neurones pour la détection et la
classification des défauts mécaniques des
machines tournantes.

1^{er} juin 2004

Résumé

Ce travail a pour objectif d'étudier l'applicabilité des réseaux de neurones pour la détection et la classification des défauts mécaniques combinés. En adoptant une approche expérimentale par simulation de différents défauts, il a voulu dégager au travers des résultats de classification les indicateurs de l'analyse vibratoire qui ont déterminé les performances de classification.

Une analyse en composante principale a servi non seulement à la réduction dimensionnelle de l'espace des indicateurs mais aussi à résoudre la question de savoir quels indicateurs sont les plus pertinents.

Mots-clés : Analyse vibratoire, Réseau de neurones, Analyse en Composantes Principales, Maintenance conditionnelle, Diagnostic des défaillances.

*A Espérance, Noerlyne et Sylvie
qui ont daigné comprendre la raison de mon absence.*

Remerciements

Mes remerciements s'adressent, en premier lieu, au Professeur P. Dehombreux, promoteur de ce travail, dont la cordialité, le positivisme et la lucidité m'ont soutenu avant et pendant l'exécution de ce travail. Je lui reconnais de m'avoir formé à la recherche et d'avoir éveillé en moi le goût de l'analyse vibratoire appliquée à la maintenance.

Merci à Mme Guanglin Hou pour m'avoir initié à l'utilisation du système d'acquisition OROS. Elle a fait preuve d'une grande disponibilité en acceptant de lire les épreuves de ce travail et d'y apposer des remarques très pertinentes.

Je remercierai ensuite tous les membres du Service de Génie Mécanique de la Faculté Polytechnique de Mons (FPMs) pour cette sympathie spontanée dont j'ai été couvert. Ma gratitude s'adressera particulièrement à M. Vergari pour son apport dans la conception du dispositif expérimental.

Je remercie le corps enseignant de la FPMs avec qui il fut agréable d'apprendre tout le long de mon programme de DEA.

Toute ma gratitude à la Coopération Technique Belge (CTB) qui m'a donné l'opportunité de cette formation en m'accordant la bourse d'étude.

Que tous mes amis trouvent ici le résultat de leur soutien. Je pense particulièrement à Jean Tshimanga, à François Ntambwe et à Kimmy Spadon.

Merci enfin aux autorités de l'Université de Mbujimayi (U.M.) qui ont permis et accepté mon détachement. Je leur reconnais tout le sens de dévouement et de conception équilibrée de l'avenir de l'UM.

Table des matières

Résumé	i
	ii
<i>Remerciements</i>	iii
0.1 La surveillance	2
0.2 Le diagnostic	2
0.3 Les techniques de diagnostic	3
0.3.1 Les méthodes basées sur le modèle du système ou méthodes internes	4
0.3.2 Les méthodes basées sur l'historique du système ou méthodes externes	5
0.4 Cadre du travail	6
I Fondements théoriques et analyse bibliographique	9
1 Les réseaux de neurones	10
1.1 Généralités	10
1.2 Le neurone biologique	10
1.2.1 Présentation du neurone biologique	10
1.2.2 Fonctionnement	12

1.3	Historique	13
1.4	Le réseau de neurones formels	13
1.5	Modélisation mathématique du réseau de neurones	14
1.5.1	Le neurone simple	14
1.5.2	Les fonctions d'activation	15
1.5.3	Le neurone à entrée vectorielle	15
1.5.4	Couche de neurone	17
1.5.5	Erreur de sortie	18
1.6	Classification des réseaux de neurones	19
1.6.1	Types d'apprentissage	20
1.6.2	Types d'architecture	23
1.7	Synthèse	24
2	Extraction des indicateurs pour une analyse vibratoire	25
2.1	Des indicateurs	26
2.2	Notions de statistique	27
2.3	Indicateurs scalaires du domaine temporel	28
2.4	Les indicateurs spectraux	30
2.5	Interprétation d'un spectre de Fourier. Signatures vibratoires.	32
2.6	Conclusions	34
II	Aspects expérimentaux	35
3	Les essais	36
3.1	Introduction	36
3.2	Le dispositif expérimental	36

3.3	Procédure des essais	39
3.3.1	Les précautions	39
3.4	Codification des défauts et des fichiers	40
3.5	Préparation des données	41
3.5.1	Traitement dans le domaine temporel	41
3.5.2	Traitement dans le domaine fréquentiel	42
3.6	Effet de la variation de vitesse	44
3.7	Caractéristiques spectrales des défauts traités	45
3.7.1	Le balourd	45
3.7.2	Le frottement	48
3.7.3	L'ovalisation de la bague externe du roulement	48
3.7.4	Le délignage	51
3.8	Conclusion	53
4	Première phase expérimentale.	
	Simulation de deux défauts : le balourd et le frottement	54
4.1	Construction du réseau de neurones	55
4.2	Le vecteur d'entrée. Domaine temporel	58
4.3	Conduite de l'apprentissage et du test	59
4.3.1	Bruitage des données	60
4.4	Résultats et discussion	61
4.4.1	Détection des deux types de défauts	61
4.4.2	Signal original non filtré 4 canaux	62
4.4.3	Signal filtré 1 canal	63
4.4.4	Signal filtré 4 canaux	64
4.5	Conclusion	66

5 Réduction dimensionnelle de l'espace des indicateurs	67
5.1 Rapport de dimensionnalité	67
5.2 L'analyse en composantes principales (ACP)	68
5.2.1 Normalisation	70
5.2.2 Saturations, communalités et variance expliquée	70
5.2.3 Choix de la dimension effective	71
5.3 Les rotations en analyse en composantes principales	73
6 Seconde phase expérimentale.	
Simulation de tous les défauts.	74
6.1 Introduction	74
6.2 Architecture du réseau	75
6.3 Les indicateurs du domaine temporel	75
6.3.1 Construction	75
6.3.2 Apprentissage et Analyse des performances	77
6.4 Les indicateurs du domaine fréquentiel	80
6.4.1 Construction	80
6.4.2 Apprentissage et Analyse des performances	80
6.5 Les indicateurs mixtes	83
6.5.1 Construction	83
6.5.2 Apprentissage et Analyse des performances	83
6.6 Synthèse et discussion	83

7	Analyse de l'influence des indicateurs.	88
7.1	Démarche	88
7.2	Analyse des indicateurs temporels	89
7.2.1	Matrice des saturations	89
7.2.2	Rotation	89
7.2.3	Matrice de transformation	93
7.2.4	Examens des poids synaptiques	94
7.2.5	Vecteurs construits avec seuls les indicateurs pertinents	98
7.3	Analyse des indicateurs fréquentiels	101
7.4	Résumé	103
A	Le banc d'essais	114
B	Les défauts traités expérimentalement sur le banc d'essais	115
C	Quelques spectres	119
D	La méthode <i>Varimax</i> en Analyse en Composantes Principales	121
E	L'algorithme de rétropropagation du gradient	124
F	Méthodes spectrales paramétriques	128
F.1	Les indicateurs basés sur les paramètres autorégressifs (AR)	129
G	Statistiques d'ordre supérieur :Bispectre et Trispectre	131

Introduction

Actuellement, dans beaucoup d'applications industrielles, on assiste à une demande croissante en matière de remplacement des politiques de maintenance corrective systématique par des stratégies de maintenance conditionnelle. Celle-ci est basée sur le principe que 99% des défaillances mécaniques sont précédés d'un signe avant coureur [1]. Elle nécessite une surveillance continue ou périodique de certains indicateurs caractéristiques de l'état de santé de la machine. Un indicateur est un quantificateur plus ou moins élaboré issu d'une grandeur dont l'acquisition est le plus souvent possible en fonctionnement. Son évolution dans le temps doit être significative de l'apparition ou de l'aggravation d'une dégradation ou d'un dysfonctionnement [3].

Pour mener à bien une tâche de maintenance conditionnelle d'un système il s'agira tout à la fois :

1. d'extraire des observations des objets étudiés, des caractéristiques qui soient pertinentes pour la *surveillance* ;
2. d'établir des critères de positionnement des capteurs ou de sélection de mesures adéquats pour la surveillance ;
3. de concevoir des stratégies de décision pour la détection des endommagements et anomalies , et pour leur *diagnostic*.

A travers ceci, la démarche qui se dégage est celle qui passe par deux étapes importantes qui sont la **surveillance** et le **diagnostic**. Ces deux étapes peuvent être précédées par la phase non moins importante de l'acquisition de la signature de l'équipement [2].

0.1 La surveillance

La surveillance consiste à prélever sur un équipement, de manière continue (on-line) ou périodique (off-line), des mesures d'indicateurs pertinents de l'évolution de la machine. Elle devra permettre, si l'on dispose d'une signature ou de niveau de seuil, de décider si les mesures acquises sont toujours décrites convenablement par la signature ou s'il s'est produit un changement significatif de comportement. La finalité d'une surveillance est l'émission d'alarme à partir d'événements. Ces alarmes sont censées provoquer une réaction humaine ou automatique.

Les indicateurs ainsi surveillés peuvent provenir de différents types d'analyses :

1. surveillance tribologique par analyse physico-chimique des lubrifiants : taux de concentration dans le lubrifiant de particules métalliques et leur spectre dimensionnel ;
2. thermographie : par exemple la surveillance de la température d'un palier ;
3. évolution du rendement mécanique, électrique ou thermodynamique de la machine ;
4. mesures acoustiques et vibratoires.

Il est très important d'avoir à l'esprit qu'un système de surveillance peu performant, qui provoque de fausses alarmes ou tarde à déclencher les alertes, peut conduire à une politique de maintenance conditionnelle qui se révèle plus coûteuse qu'une politique corrective ou préventive systématique.

0.2 Le diagnostic

Un système complet de maintenance conditionnelle doit permettre non seulement de donner une réponse en tout-ou-rien (le système a ou n'a pas changé) grâce aux alarmes, mais aussi de localiser le changement détecté, et de procéder au diagnostic proprement dit.

Le diagnostic industriel est défini selon la norme internationale comme l'identification de la cause probable de la (ou des) défaillance(s) à l'aide d'un raisonnement logique fondé sur un ensemble d'informations provenant d'une inspection,

d'un contrôle ou d'un test. Il ressort de cette définition les deux tâches principales du diagnostic, à savoir [8] :

- observer les symptômes de la défaillance,
- identifier la cause de la défaillance à l'aide d'un raisonnement logique fondé sur des observations.

Le problème du diagnostic peut se poser en ces termes :

étant donné l'espace X des paramètres inconnus \mathbf{x}_i (toutes les causes) et l'espace Y des grandeurs observables \mathbf{y}_i (tous les symptômes), déterminer la solution \mathbf{x}_i de \mathbf{y}_i connaissant la fonction F :

$$F(\mathbf{x}_i) = \mathbf{y}_i.$$

Le diagnostic consiste donc à répondre à la question : quel(s) est (sont) les paramètre(s) \mathbf{x} responsable(s) du déclenchement de l'alarme globale ? La solution n'est en réalité pas aussi aisée que peut le laisser supposer la forme de l'équation car, d'une part il faut définir des modèles pour chacune des n causes \mathbf{x}_i , pour chacun des symptômes \mathbf{y}_i comme pour chacune des fonctions F_i et d'autre part il faut que le problème soit bien posé.

Pour arriver à résoudre ce problème, de nombreuses méthodes ont été développées, parmi lesquelles on distingue les méthodes internes et les méthodes externes.

0.3 Les techniques de diagnostic

Les méthodes de diagnostic se différencient non seulement dans la manière dont les données du système sont utilisées, mais aussi dans la forme requise pour le traitement de ces données. Nous allons présenter ici une classification basée sur la forme des données. On distingue les techniques basées sur le modèle du système (méthodes internes) de celles qui reposent sur l'historique du système (méthodes externes) [9].

0.3.1 Les méthodes basées sur le modèle du système ou méthodes internes

La source de connaissance dans ces méthodes est la compréhension profonde du système. Le principe est de définir un ensemble de relations qui décrivent les interactions entre les différentes variables du système. Dans cette classe de méthodes on peut distinguer deux catégories :

- les méthodes qualitatives
- les méthodes quantitatives.

Les méthodes qualitatives

La stratégie employée est basée sur la notion de causalité dans le comportement du système. Les plus utilisées de ces méthodes sont les arbres de défaillance et la méthode AMDEC (Analyse des Modes des Défaillances, de leurs Effets et de leur Criticité). La méthode de l'arbre de défaillance est une méthode avec laquelle on identifie toutes les combinaisons possibles d'événements qui entraînent la réalisation d'un événement unique indésirable. Elle utilise une chaîne qui remonte jusqu'à un événement primaire considéré comme cause de la déviation observée dans le comportement du système [7]. L'AMDEC est une technique déductive et qualitative avec laquelle les effets (conséquences) des causes de défaillances des composants élémentaires sont systématiquement identifiés [8].

Cependant une sérieuse limitation à l'utilisation de ces méthodes est la génération d'un grand nombre d'hypothèses, ce qui rend le processus de décision incertain. Ceci est dû aux ambiguïtés qualitatives qu'impliquent ces méthodes.

Les méthodes quantitatives

Ces méthodes reposent sur les relations mathématiques qui existent entre les variables et que l'on essaie de modéliser de la façon la plus proche possible du processus réel. Les modèles sont développés soit en utilisant les principes fondamentaux qui régissent la physique du système soit en utilisant les données d'entrée et de sortie

du processus. Ces modèles peuvent être dynamiques, statiques, linéaires ou non linéaires. Parmi ces méthodes, nous citerons la méthode du modèle, les méthodes par identification des paramètres, les méthodes par estimation du vecteur d'état et les méthodes par modélisation des signatures [8].

0.3.2 Les méthodes basées sur l'historique du système ou méthodes externes

Quand aucun modèle n'est disponible, les techniques basées sur les modèles du système sont inapplicables et on met en oeuvre des techniques dites externes. La seule connaissance repose alors sur l'expertise humaine confortée par un solide retour d'expérience. Dans cette catégorie, on peut aussi distinguer des méthodes quantitatives et des méthodes qualitatives.

Les méthodes qualitatives

Il existe :

1. les méthodes basées sur des règles : C'est le cas des systèmes experts. L'historique du système est enregistrée sous forme de couple (antécédent, conséquence) lié à une défaillance connue. Ceci exige une correspondance explicite des symptômes connus aux causes primaires.
2. l'analyse qualitative des tendances : Elle utilise la présence d'information sur la tendance d'évolution dans une mesure donnée. Elle a deux étapes fondamentales : l'identification des tendances dans les mesures et l'interprétation des tendances en terme de scénarios de défaillances. La procédure d'identification est donc tenue à être robuste vis-à-vis des variations du signal dues aux bruits et de ne capturer que les grosses variations [25].

Les méthodes quantitatives

Les plus connues de ces méthodes sont les techniques de reconnaissance des formes, les techniques statistiques et les techniques basées sur l'intelligence artifi-

cielle. Ces dernières constituent des outils privilégiés pour la catégorisation de signaux issus du système à surveiller en fonctionnement, c'est-à-dire l'affectation de la catégorie de défauts à ces signaux [10].

Les techniques statistiques, notamment la statistique multivariable avec l'analyse en composantes principales, peuvent constituer un précieux outil dans le processus de diagnostic des défaillances. On peut leur associer des techniques de classification automatiques telles que les réseaux de neurones. L'idée à la base de l'utilisation des variables statistiques est le fait que l'évolution de l'état d'une machine ou d'un équipement n'est pas déterministe. Le système étant sous influence des facteurs aléatoires, il est logique de lui formuler une approche probabiliste. Ainsi donc, les séries temporelles des mesures seront caractérisées par des distributions de probabilité correspondant à l'état sain ou à un état de défaillance.

0.4 Cadre du travail

Ce propos adresse une méthode automatique de diagnostic externe : les réseaux de neurones. Il en explore l'utilisabilité dans le cas des défauts concourants et s'efforce de dégager pour une telle application les indicateurs les plus pertinents. Il se base sur les données d'analyse vibratoire récoltées sur un banc d'essai simulateur de défauts mécaniques de machines tournantes, chaque défaut ayant plusieurs niveaux d'expression.

Le travail comprend deux grandes parties : une théorique et bibliographique, d'une part et une pratique expérimentale, d'autre part. Les 7 chapitres suivants sont repartis entre les deux parties en dehors de l'introduction et la conclusion :

Chapitre 1 : Ce chapitre introduit la notion de réseaux de neurones. Il en retrace le contenu historique et en présente la modélisation mathématique. La question des architectures et des algorithmes d'apprentissage des réseaux de neurones y est abordée de manière succincte à travers une brève classification des réseaux de neurones.

Chapitre 2 : Une revue de la littérature est faite concernant les différentes possibilités d'extraction des indicateurs issus d'une analyse vibratoire et destinés à

la construction du vecteur d'entrée du réseau de neurones. Plusieurs possibilités se révèlent, notamment les indicateurs statistiques du domaine temporel les spectres fréquentiels, les statistiques d'ordre supérieur et l'analyse temps-fréquence.

Chapitre 3 : Ce chapitre introduit les aspects expérimentaux du travail. Les différents défauts simulés sont présentés ainsi que leur nomenclature. On y trouve une description du dispositif expérimental représentant le banc d'essais simulateur de défauts.

Chapitre 4 : Ici est décrite notre première phase expérimentale au cours de laquelle nous n'avons simulé que deux défauts : le balourd et le frottement. Cette phase avait pour objectif l'acquisition d'une vision première en ce qui concerne la conception du réseau de neurones et la construction des espaces d'indicateurs. Les résultats obtenus dans cette phase ont largement influencé la seconde phase en termes d'architecture du réseau de neurones et de la préparation des données.

Chapitre 5 : La réduction dimensionnelle de l'espace d'indicateurs est un des aspects introduits dans l'étape de la préparation des données. L'analyse en composantes principales est appliquée à l'espace des indicateurs tant temporels que fréquentiels. Des notions permettant d'analyser l'influence des indicateurs sur les performances de classification sont introduites au cours de ce chapitre, en particulier les saturations et les communalités.

Chapitre 6 : Une étude de plusieurs types de défauts est menée. Le balourd, le frottement, l'ovalisation de la bague externe de roulement et le désalignement sont simulés à leurs différents niveaux d'expression. La particularité dans l'architecture du réseau de neurones, dans cette phase, tient du fait qu'il est disposé une sortie pour chaque niveau de défaut et non une sortie par défaut comme dans la première phase. Ceci a l'avantage de permettre l'utilisation des fonctions sigmoïdes à la sortie et d'y imposer seules les deux valeurs 1 et 0. Les différentes sorties correspondant à un défaut donné sont ensuite passés à une fonction de compétition pour assigner la valeur 1 seulement à la sortie qui présente la plus grande valeur.

Chapitre 7 : Dans ce chapitre, en nous basant sur les matrices de saturation, les communalités, les matrices de transfert et les poids synaptiques, nous discutons

sur les indicateurs pour dégager ceux d'entre eux qui semblent avoir un grand contenu discriminatoire et qui influencent ainsi de façon prépondérante la classification.

Première partie

Fondements théoriques et analyse bibliographique

Chapitre 1

Les réseaux de neurones

1.1 Généralités

La technique des réseaux de neurones appliquée aux problèmes de diagnostic industriel a attiré beaucoup d'attention ces dernières années. La littérature abonde sur son application dans ce domaine et des études ont montré que les réseaux de neurones artificiels sont une bonne technique de classification [25].

Les réseaux de neurones artificiels sont un modèle du neurone biologique dont il ne retiennent qu'une vision fort simplifiée.

1.2 Le neurone biologique

1.2.1 Présentation du neurone biologique

Le cerveau humain contient trois couches successives :

- le cerveau reptilien, siège des réflexes instinctifs ;
- l'hippocampe, siège de l'émotivité et de la sensibilité ;
- le cortex, spécifique aux mammifères et siège du raisonnement et de la pensée.

Dans chacune de ces couches on trouve deux types de cellules : les gliales et les neurones [8].

Tous les neurones sont constitués de 3 parties (figure 1.1) :

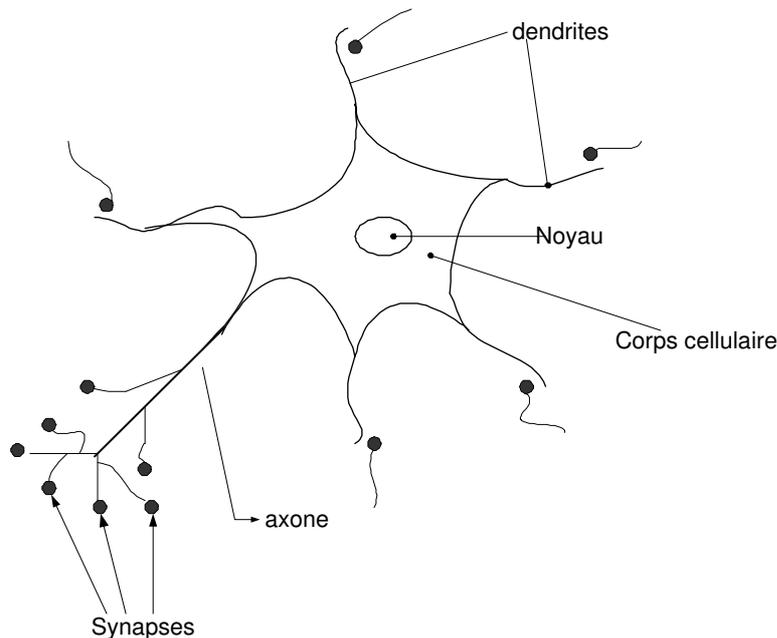


FIG. 1.1 – Le neurone biologique

- le corps proprement dit de la cellule où se situe le mécanisme de déclenchement du neurone ;
- son prolongement l'axone (qui peut atteindre plusieurs dizaines de centimètres) le long duquel voyagent les impulsions nerveuses dirigées vers la périphérie et de minces ramifications ;
- les dendrites, qui reçoivent des informations (impulsions nerveuses) des autres neurones et les apportent vers le corps cellulaire.

Le corps cellulaire, appelé aussi soma, constitue l'élément principal du neurone. Il contient le noyau cellulaire renfermant lui-même le génome, c'est à dire l'information génétique de l'organisme dont le neurone est issu.

L'axone est le prolongement cellulaire principal et c'est grâce à lui que le neurone va transmettre les informations. A son extrémité se trouve l'arborisation terminale où se trouve un nombre plus ou moins important de terminaisons synaptiques et qui est le pôle émetteur de la cellule.

Les dendrites représentent le deuxième type de prolongement cellulaire et se trouvent en grande quantité autour du soma, c'est pourquoi on parle souvent d'arborisation dendritique. Les dendrites constituent le pôle récepteur de la cellule et

c'est à cet endroit que vont se faire les connexions avec les autres neurones.

Les neurones sont interconnectés entre eux par des synapses qui sont des jonctions entre deux neurones.

1.2.2 Fonctionnement

Les axones sont recouverts de cellules, dites cellules de Schwann. Lorsque ces cellules s'enroulent autour de l'axone, il se forme un revêtement appelé gain myélinique qui présente des points découverts ou noeuds de Ranvier.

L'impulsion nerveuse voyage par saut d'un noeud de Ranvier à un autre à une vitesse d'environ 200 m/s. Quand l'axone d'un neurone rencontre un autre neurone, le contact s'établit par l'intermédiaire du synapse. Le synapse est une sorte de bulbe. Entre la surface du bulbe (bouton terminal) et celle du nouveau neurone, se trouve un espace dit espace synaptique. Pour que l'impulsion nerveuse qui est un signal de nature électrique, puisse franchir cet espace, des substances électrochimiques entre en action, les neurotransmetteurs. Le processus de communication entre les neurones se fait donc par voie électrochimique aux niveaux des synapses par des neurotransmetteurs. L'action des neurotransmetteurs est extrêmement brève pour empêcher que la stimulation nerveuse ne se prolonge au-delà du temps nécessaire au transfert de l'impulsion nerveuse.

De nombreuses théories décrivent les mécanismes d'excitation et de réponse des neurones. La théorie électrique sur laquelle se base les réseaux de neurones artificiels considère la jonction synaptique comme le lieu où le neurone perçoit un stimulus par voie électrochimique. La membrane synaptique lorsqu'elle ne reçoit pas d'excitation, est polarisé à une tension d'environ 60 mV. A l'apparition d'excitations qui se traduisent au niveau du noyau cellulaire par une tension supérieur à un seuil, le neurone est alors activé et cela déclenche un processus de dépolarisation : le neurone émet sur l'axone une tension positive d'environ +60 mV en se comportant comme un système non linéaire.

C'est sur base du neurone biologique que les pères du calcul connexionniste ont conçu le neurone artificiel, voyons-en d'abord l'historique.

1.3 Historique

Dans la période de 1940 à 1956, les sciences et techniques de la cognition s'élaborent peu à peu et l'ancienne cybernétique donna naissance aux sciences cognitives.

L'apparition des réseaux de neurones artificiels s'inscrit complètement dans cette genèse des sciences cognitives, et les origines dans la volonté de modéliser, de façon mathématique, les neurones biologiques.

Le premier modèle fut proposé par Mac Culloch et Pitts en 1943 : c'est le neurone de Mc Cullosch-Pitts ou automate à seuil [28]. Quelques années après, en 1949, le neurophysicien Hebb propose une formulation du mécanisme d'apprentissage sous forme d'une règle de modification des connexions synaptiques (règle de Hebb). Cette règle, basée sur des données biologiques, modélise le fait que si des neurones, de part et d'autre d'un synapse, sont activés de façon synchrone et répétée, la force de connexion synaptique va aller croissant.

Les résultats de ces travaux ont été à l'origine de modèles connexionnistes de plus en plus élaborés. En 1958 apparaît, proprement dit, le premier réseau de neurones artificiels grâce aux travaux de Frank Rosenblatt qui conçoit le Perceptron. Le Perceptron modélise les facultés de reconnaissance de forme du système visuel. Il possède une couche de neurones d'entrée ("perceptive") ainsi qu'une couche de neurones de sortie ("décisionnelle"). En parallèle vers les années 60 Widrow met au point l'Adaline (Adaptative Linear Neuron) et le Madaline. Cette étape sera suivie d'une période morte par suite des travaux de Minsky et Papert qui ont montré que le modèle du Perceptron ne présentait aucun intérêt à cause de son incapacité de résoudre le problème du 'ou exclusif'.

Il faudra attendre les travaux des chercheurs comme Hopfield pour revoir un regain d'intérêt aux réseaux de neurones.

1.4 Le réseau de neurones formels

Le neurone artificiel, encore appelé neurode, noeud ou unité neuronale, est une modélisation mathématique du neurone biologique. Il en reprend les grands prin-

cipes de fonctionnement, notamment la sommation des entrées. Les réseaux de neurones artificiels sont inspirés du système nerveux. Ils sont constitués de neurones interconnectés entre eux.

Chaque neurone reçoit en entrée des informations venant de plusieurs neurones à travers les connexions synaptiques. Chacune de ces entrées est pondérée par une valeur w , dite poids synaptique. Les valeurs ainsi pondérées sont ensuite sommées avant de passer en argument à une fonction f appelée fonction d'activation ou fonction de transfert. Les valeurs des poids synaptiques sont déterminées à la suite d'un processus d'apprentissage au cours duquel le neurone ajuste ses forces de connexion afin d'obtenir une sortie désirée.

1.5 Modélisation mathématique du réseau de neurones

Dans cette section nous nous proposons de présenter la modélisation du neurone et celle des réseaux de neurones artificiels. Nous noterons les grandeurs vectorielles et les matrices par des lettres en gras et les grandeurs scalaires par des lettres en italiques.

1.5.1 Le neurone simple

Par neurone simple nous entendons un neurone auquel n'est transmis en entrée qu'une seule valeur scalaire p [29]. Cette valeur est multipliée dans la connexion par

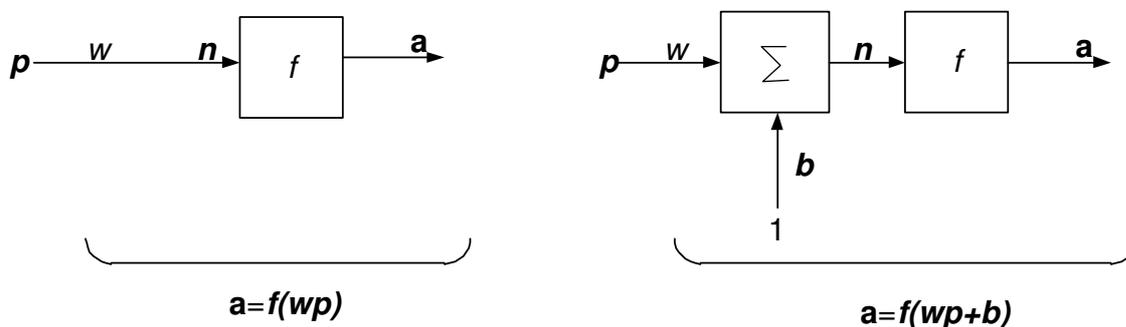


FIG. 1.2 – Le neurone formel simple sans biais (à gauche) et avec biais (à droite)

le poids w , pour former le produit wp qui est aussi un scalaire. Le neurone peut avoir ou non un biais b (figure 1.2). S'il y a un biais, celui-ci sera sommé à la valeur pondérée wp avant de passer en argument au niveau de la fonction d'activation f . L'argument de f est $n = wp + b$ et la valeur a que retourne la fonction f est la sortie du neurone. On aura donc

$$a=f(wp+b) \quad (1.1)$$

Le biais b peut être vu comme un décalage de la fonction f . Il peut tout aussi bien être appréhendé comme un poids synaptique qui s'applique à une entrée unitaire. Les valeurs w et b sont des paramètres scalaires ajustables de ce neurone simple. L'idée centrale est de faire présenter au neurone un comportement désiré par le seul ajustement des paramètres w et b . On peut ainsi donc entraîner un réseau de plusieurs neurones interconnectés à exécuter une tâche particulière en ajustant ces paramètres ou en laissant le réseau les ajuster seul.

1.5.2 Les fonctions d'activation

Les fonctions d'activation sont de nature très variées : elle peuvent être déterministes, continues, discontinues ou aléatoires [29]. Les fonctions les plus utilisées sont :

1. la fonction à seuil de Heaviside ;
2. la fonction sigmoïde de type Log ;
3. la fonction sigmoïde de type tangente hyperbolique ;
4. la fonction linéaire sans saturation ;
5. la fonction linéaire à seuil ou saturation.

1.5.3 Le neurone à entrée vectorielle

Contrairement au neurone simple présenté précédemment, l'entrée du neurone peut être un vecteur de dimension R (figure 1.5).

$$\mathbf{p} = \{p_1, p_2, \dots, p_R\} \quad (1.2)$$

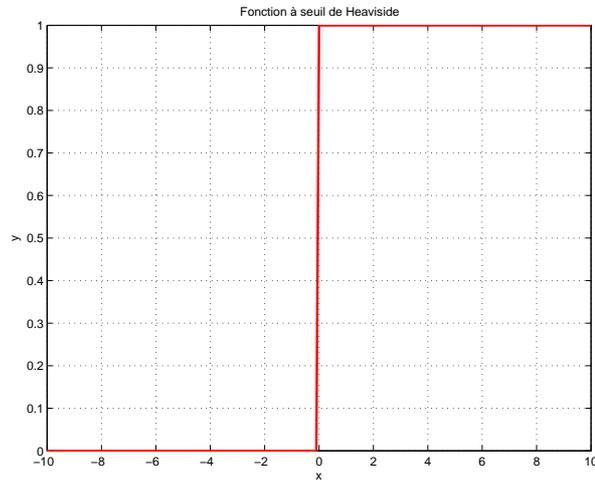


FIG. 1.3 – Fonction à seuil de Heaviside

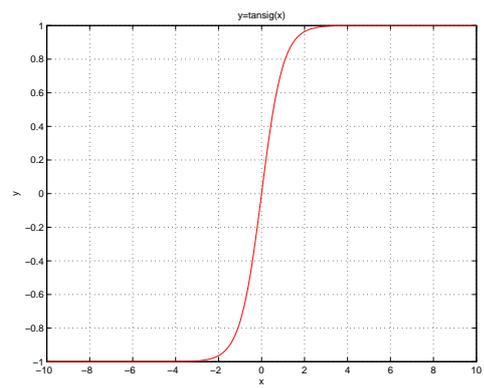
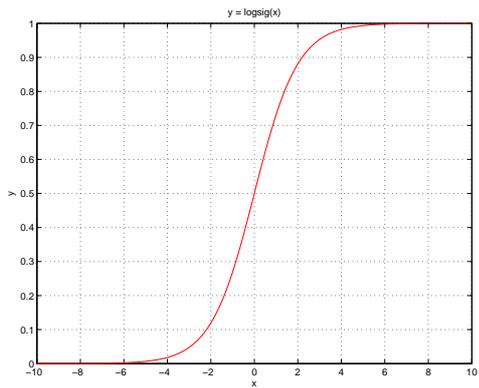


FIG. 1.4 – Fonctions sigmoïde log et tan

Chaque entrée individuelle p_i est multipliée par le poids $w_{1,i}$, l'indice 1 dans le poids indiquant le neurone auquel sont transmises ces entrées. La somme des valeurs pondérées est le produit scalaire des vecteurs \mathbf{W} et \mathbf{p} .

Avec \mathbf{W} donné par

$$\mathbf{W} = \{w_{1,1}, w_{1,2}, \dots, w_{1,R}\} \quad (1.3)$$

L'argument n de la fonction d'activation f est

$$n = w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b \quad (1.4)$$

ou vectoriellement

$$n = \mathbf{W}\mathbf{p} + b \quad (1.5)$$

et la sortie du neurone a sera évidemment calculée par $a = f(n)$.

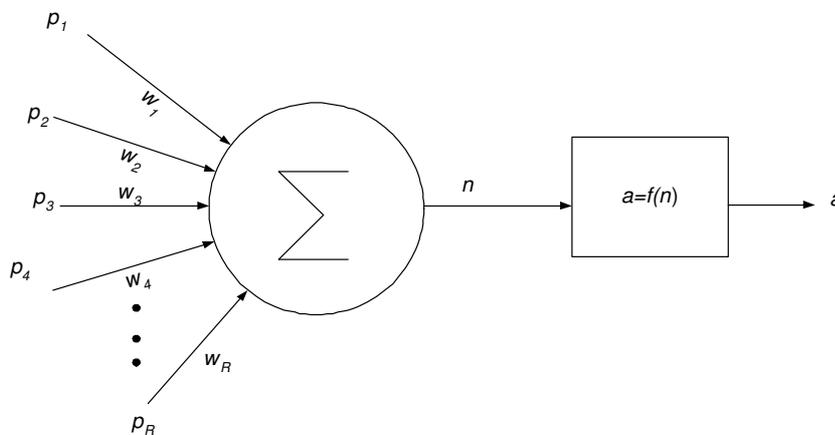


FIG. 1.5 – Neurone à entrée vectorielle

1.5.4 Couche de neurone

Nous inspirant du neurone présenté ci-haut, nous pouvons définir une couche comme l'espace où l'on trouve les poids synaptiques et le biais, où s'effectue les

opérations d'addition et de multiplication, et où finalement la fonction d'activation prend en argument la somme des entrées pondérées augmentée du biais. Les entrées ne sont pas comptées dans une couche.

En généralisant l'architecture par combinaison de plusieurs neurones dans une couche et en considérant plusieurs couches disposées les unes après les autres, nous obtenons ce qu'on appelle réseau de neurones. Une couche sera donc caractérisée par le nombre S de ses neurones, un vecteur biais \mathbf{b} de dimension S , une matrice \mathbf{W} des poids synaptiques et une série de S fonctions d'activation identiques ou non pour chaque neurone. La matrice \mathbf{W} est une matrice de dimension $S \times R$, R étant le nombre d'entrées pour la couche considérée ou la dimension du vecteur d'entrée \mathbf{p} .

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,R} \\ \vdots & \vdots & & \vdots \\ w_{S,1} & w_{S,2} & \cdots & w_{S,R} \end{bmatrix} \quad (1.6)$$

Le premier indice (indice de ligne) des éléments de la matrice \mathbf{W} indique le neurone auquel le synapse aboutit, et l'indice de colonne l'origine du synapse. Ainsi, par exemple, $w_{1,2}$ est la force du signal provenant du second élément d'entrée et destiné au premier neurone. Il faudra faire une distinction entre la matrice de poids connectés aux entrées et les matrices de poids de connexion entre couches. Pour cela convenons d'appeler \mathbf{IW} (Input weights) la matrice de poids d'entrée et $\mathbf{LW}_{j,i}$ (Layer weights), la matrice de connexion entre la couche j et la couche i . La couche de départ est i et celle de destination j .

La figure 1.6 illustre un réseau de neurones à trois couches et à S_i neurones dans la couche i . Dans cet exemple, toutes les fonctions d'activation d'une couche sont identiques. La sortie \mathbf{a}^i d'une couche i est un vecteur dont les composantes sont

$$\mathbf{a}^i = \{a_1^i, a_2^i, \dots, a_{S_i}^i\} \quad (1.7)$$

1.5.5 Erreur de sortie

D'une manière générale, la sortie \mathbf{a} du réseau de neurones est comparée à une sortie attendue \mathbf{t} appelée "cible". L'écart entre la sortie obtenue et la sortie attendue est

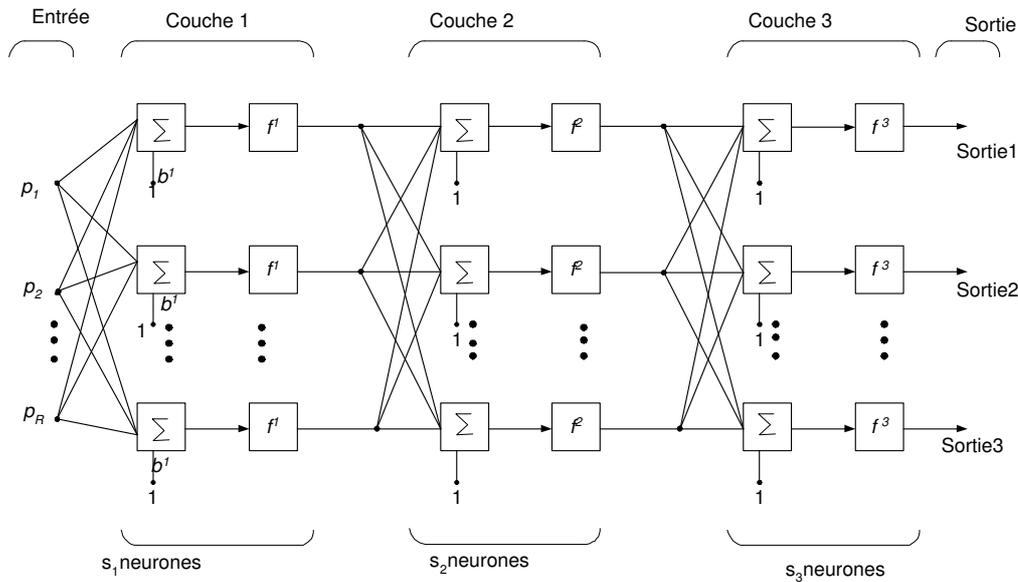


FIG. 1.6 – Réseau de neurones multicouche

représentée par une fonction d'erreur dont la définition dépend du type de problème. L'erreur est fonction des valeurs des poids synaptiques et des biais. La fonction d'erreur la plus utilisée est la "moyenne des carrés de l'erreur" (mse : mean squared error).

Si l'on dispose de m exemples de vecteurs d'entrée \mathbf{x}_q ($q = 1 \dots m$) auxquels on associe m vecteurs de sortie attendus \mathbf{t}_q ($q = 1 \dots m$), la fonction d'erreur "mse" est définie comme suit :

$$mse(\mathbf{W}) = \frac{1}{m} \sum_{q=1}^m (\mathbf{t}_q - \mathbf{a}_q)^2 \quad (1.8)$$

1.6 Classification des réseaux de neurones

Il serait prétentieux pour nous de vouloir dresser une classification complète des réseaux de neurones, cela dépasserait le cadre de ce travail. Il existe différentes façons de classer les réseaux de neurones selon les critères choisis, tels que : type d'apprentissage, architecture des connexions, forme pour traiter et représenter les données, etc. Pour notre part, nous estimons qu'une brève classification s'impose

pour mieux comprendre les avantages et les inconvénients d'un choix d'architecture ou d'algorithme par rapport à un autre.

1.6.1 Types d'apprentissage

En fonction des influences venant des entrées et leurs poids respectifs, la fonction de transfert d'un neurone en détermine l'activation. Les changements apportés aux poids synaptiques de la structure du réseau sont responsables des changements de comportement. Le réseau peut ainsi être capable d'associer des stimuli aux réponses afin de bien résoudre un problème. La méthode utilisée pour modifier le comportement d'un réseau de neurones est la *règle d'apprentissage*. Les réseaux de neurones ont donc la capacité d'apprendre des règles à partir des exemples, et sur base des règles apprises, certains réseaux peuvent faire une généralisation. La généralisation est la capacité du réseau à réagir correctement lorsqu'on lui présente des entrées non vues au cours des séances d'apprentissage. L'apprentissage est un processus itératif au cours duquel les poids sont modifiés plusieurs fois avant d'atteindre leurs valeurs finales [4].

On distingue trois types principaux d'apprentissage :

Apprentissage supervisé : l'utilisateur dispose d'un comportement de référence précis qu'il désire faire apprendre au réseau. Les poids synaptiques sont déterminés à partir d'exemples étiquetés de formes auquel ont été associées des réponses désirées également étiquetées (cibles). Pendant l'apprentissage les poids sont ajustés dans le sens de minimiser l'écart entre les cibles et les sorties calculées. Ce processus peut prendre un grand nombre d'itérations avant de converger, parce que certaines règles d'apprentissage (telle que la rétropropagation du gradient) peuvent converger vers des minima locaux sur la surface d'erreur [11] (figure1.7).

Apprentissage semi-supervisé : l'utilisateur ne dispose que des indications imprécises (par exemple, échec/succès du réseau) sur le comportement final désiré. Ces techniques d'apprentissage sont aussi appelées apprentissage par renforcement (reinforcement learning).

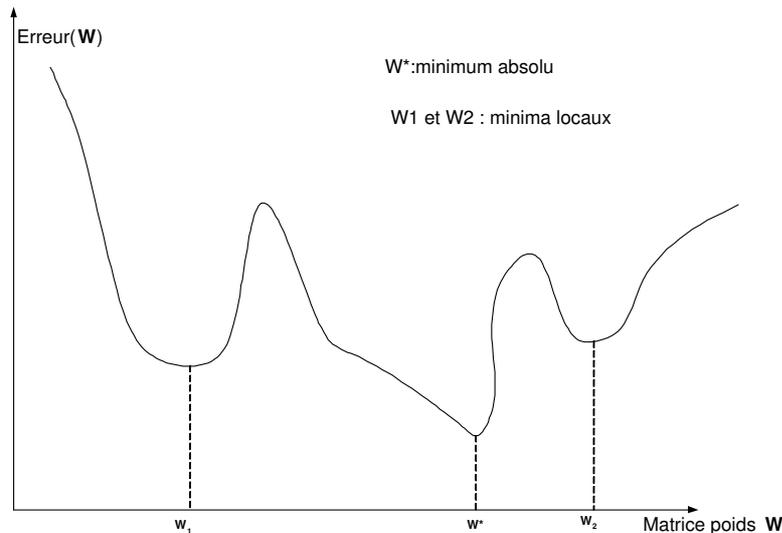


FIG. 1.7 – La surface "Erreur" en fonction de la matrice des poids

Apprentissage non-supervisé : ce processus d'apprentissage est beaucoup utilisé dans des problèmes de classification où les cibles sont non connues a priori. Le système essaie alors de générer un ensemble unique de poids pour une classe particulière d'objets. L'objectif de l'apprentissage non-supervisé est d'ajuster les poids de manière autonome jusqu'à l'équilibre c'est-à-dire jusqu'à ce qu'aucune modification des poids ne soit nécessaire. L'apprentissage non-supervisé associe donc une classe d'objets à une classe de poids. Les comportements résultant de ces apprentissages sont appelés *auto-organisation* (Self-Organization). Les exemples typiques des réseaux de neurones à apprentissage non-supervisé sont les réseaux de Hopfield, les mémoires associatives et les réseaux cognitifs [5].

Généralement, l'apprentissage d'un réseau de neurones demande une grande quantité de données que l'on regroupe dans un ensemble d'*exemples d'apprentissage*. Il est souvent utile de disposer d'autres ensembles de données pour tester la validité du comportement appris par le réseau neuronal. Ces ensembles sont appelés ensembles d'*exemples de test ou de généralisation*. C'est surtout les capacités de généralisation d'un réseau de neurones qui déterminent son efficacité en pratique.

Il peut arriver qu'un réseau de neurones se spécialise trop à un ensemble de données d'apprentissage. Dans ce cas on parle d'apprentissage par *coeur* ou surap-

apprentissage (*overfitting*), qui a pour conséquence une mauvaise généralisation. Pour éviter ce type de problème, une façon pratique consiste à partitionner les données en trois ensembles : le premier pour l'apprentissage, le second pour la validation et le troisième pour le test. Il arrive souvent que l'on ne constitue que deux ensembles : apprentissage et test. L'ensemble de test permet de contrôler l'erreur de généralisation et de déterminer le bon moment pour arrêter le processus d'apprentissage. La figure 1.8 illustre l'évolution de l'erreur en fonction du nombre d'itérations ou époques, au-delà d'une certaine limite l'erreur de généralisation se met à croître.

Le phénomène de surapprentissage peut aussi être évité en utilisant lors de l'entraînement des séries de données différemment bruitées. De cette manière, le risque de mémorisation des exemples par le réseau de neurones est presque nul.

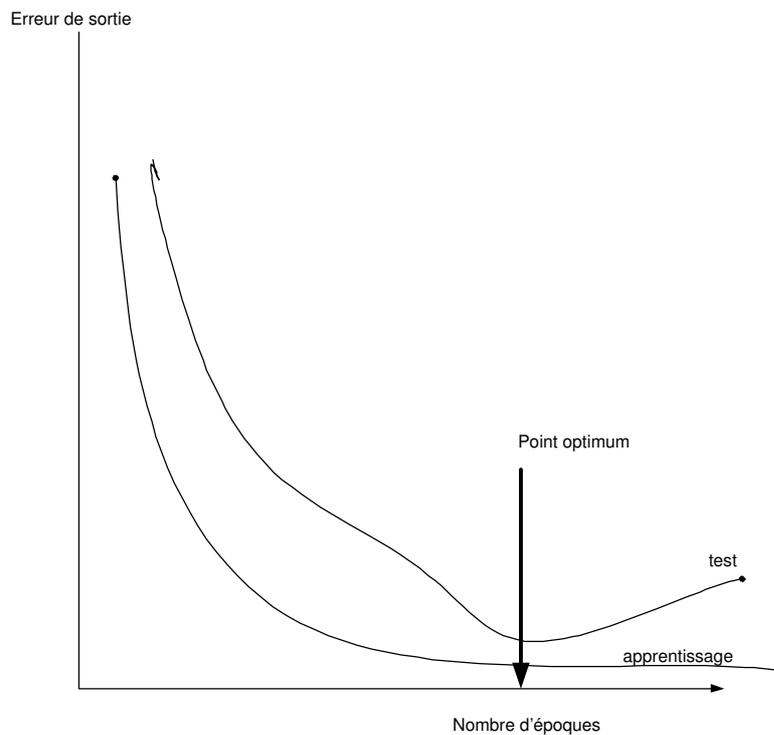


FIG. 1.8 – Erreur dans l'ensemble d'apprentissage et de test

Selon la manière dont les données d'apprentissage sont présentées au réseau de neurones on peut parler de :

- *Apprentissage instantané* au cours duquel l'ensemble des données d'apprentissage est analysé en une seule fois et les poids synaptiques ajustés d'un seul

coup. On ne fait donc qu'un seul passage sur les données, ce mode est aussi appelé *one-shot-learning* et n'est pas très utilisé.

- *Apprentissage incrémental ou continu* : les poids du réseau sont ajustés à chaque passage sur une entrée c'est-à-dire que l'algorithme prend en compte continuellement les exemples qui lui arrivent (*continuous /on-line learning*). Les exemples sont alors présentés séquentiellement [29].
- *Apprentissage par lots (paquets)* : dans ce type d'apprentissage les poids synaptiques ne sont ajustés qu'après présentation de tous les exemples. On effectue souvent plusieurs présentations de façon à optimiser les poids et à réduire l'erreur en sortie. Chaque présentation de l'ensemble complet des données d'apprentissage est appelée *époque* (epoch) [4]. L'ordre des exemples de l'ensemble d'apprentissage peut aussi être manipulé, ce qui peut avoir une influence sur l'évolution de l'apprentissage. Ce type d'apprentissage est le plus utilisé et est aussi connu sous le nom de *batch training* ou *batch learning*.

Les règles d'apprentissage les plus utilisées sont ([5] [4] [8]) :

- les méthodes de correction de l'erreur, telles que la descente de gradient sur une surface : Adaline, Madaline, Rétro-propagation du gradient, Cascade-Correlation, CMAC (Cerebellar model articulation controller) ;
- les méthodes d'apprentissage par renforcement ;
- les méthodes d'apprentissage par compétition ou par auto-organisation : Kohonen feature Map, ART ;
- les méthodes d'apprentissage par création de prototypes ou de noyaux : RBF ;
- les méthodes d'apprentissage basées sur des mémoires associatives (auto-associatives ou hétéro-associatives) : modèles de Hopfield, ... ;
- les méthodes d'apprentissage temporel (réseaux récurrents).

Ces méthodes ont largement été utilisées dans la détection des défauts mécaniques.

1.6.2 Types d'architecture

Outre le fait que les unités neuronales peuvent différer selon le type de fonction d'activation, leur mode de connexion, leur architecture peuvent varier.

Les architectures les plus importantes sont :

- réseaux à seule couche : les neurones sont tous sur le même niveau et directement connectés aux entrées. Les connexions latérales peuvent y être aussi organisées. Un exemple typique est le 'Kohonen Feature Map'.
- réseaux à couches unidirectionnels (feedforward) : On y a plusieurs couches, chaque unité d'une couche reçoit ses entrées de la couche précédente et envoie ses sorties vers la couche suivante ;
- réseaux récurrents : ils peuvent avoir une ou plusieurs couches avec la particularité de possibilité de connexion au sein d'une couche ou en boucle vers une couche inférieure. Ce type d'architecture permet la modélisation des comportements dynamiques. Le réseau de Hopfield est un exemple de réseau récurrent.

1.7 Synthèse

Nous avons, dans ce chapitre, touché à l'historique, aux types et aux architectures des réseaux de neurones artificiels après en avoir montré le lien avec le réseau de neurones biologiques. Pour une application pratique, il est toujours inévitable de se questionner sur ce que doit être l'alimentation du réseau de neurones pour s'assurer des résultats satisfaisants. Dans notre cas où les données d'analyse vibratoires sont utilisées pour la classification, il faut opérer un choix sur les indicateurs à fournir au réseau. Le chapitre qui suit discute des différentes manières possibles de construire les entrées d'un réseau de neurones à partir de l'analyse vibratoire.

Chapitre 2

Extraction des indicateurs pour une analyse vibratoire

Plusieurs travaux sur la surveillance conditionnelle des machines, ont utilisé l'analyse vibratoire combinée aux réseaux de neurones. Les machines tournantes n'ont pas été les seules à être concernées. Des applications telles que la surveillance des transformateurs électriques ont été abordées dans [15]. Un domaine qui a été suffisamment exploré est celui de la surveillance de l'usure d'outils en usinage et en formage ([16] [17] [20]). Les travaux comme [21] et [23] ont plutôt utilisé la force de coupe, mais dans tous les cas d'autres indicateurs devaient être fournis pour caractériser complètement une condition de coupe : les vitesses d'avance et de coupe, la géométrie de la pièce, l'émission acoustique, les propriétés des matériaux (outil, lubrification, pièce), ...

Une question majeure, parce qu'elle dépend de la qualité et des performances d'un système d'aide au diagnostic (réseau de neurones par exemple), est de choisir les indicateurs à utiliser qui soient très représentatifs de l'évolution de la machine.

Il n'est généralement pas possible de faire une classification basée sur un échantillon unique de données vibratoires.

2.1 Des indicateurs

L'extraction des indicateurs peut être vue comme un processus qui consiste à ne retenir que les informations discriminatoires entre classes et à réduire sensiblement les informations communes dans un ensemble de mesures faites sur plusieurs classes [25]. Les indicateurs développés dans cette philosophie pourront bien montrer qu'une mesure est tout à fait représentative d'une classe donnée plutôt que d'une autre.

Les données vibratoires sont les plus utilisées dans la surveillance des machines à cause de leur sensibilité et de leur capacité à prédire assez précocement le développement des défauts ([6] [22]). Comparées aux autres techniques, elles sont, selon [12], les seules à donner la mesure la plus pertinente de l'état d'une machine tournante.

Le choix des indicateurs est une phase cruciale dont dépend de manière non négligeable la phase de classification par le réseau de neurones [24]. Comme nous l'avons dit plus haut, les indicateurs doivent suffisamment être représentatifs, dans leur évolution, de l'état de la machine. Ils devront, non seulement, être en mesure de provoquer une alarme lors de l'initialisation d'un défaut, mais aussi de permettre un diagnostic assez précis de la machine.

Dans l'utilisation des réseaux de neurones pour le diagnostic des défauts basé sur l'analyse vibratoire, on a le choix entre les indicateurs scalaires dans le domaine temporel et indicateurs spectraux dans le domaine fréquentiel.

Une forte préférence penche du côté de l'analyse dans le domaine fréquentiel où les défauts sont révélés même quand les niveaux vibratoires sont faibles. L'avantage de ceci est que les défauts peuvent être détectés dès leur phase d'initialisation et ainsi être surveillés pendant leur évolution. Cette corrélation mène à une meilleure analyse que l'on ne peut faire dans le domaine temporel.

Dans ce qui suit nous allons parcourir les indicateurs vibratoires utilisés pour alimenter un réseau de neurones destiné au diagnostic mécanique, mais avant cela, introduisons d'abord quelques notions de statistique.

2.2 Notions de statistique

Nous avons souligné dans la section 0.3.2 qu'il était justifié d'associer à un système sous influence aléatoire des variables statistiques.

Si la série temporelle de la vibration est regardée comme une variable aléatoire X , la répartition de la probabilité est définie comme

$$F(x) = P(X < x) \quad (2.1)$$

la fonction de densité de probabilité est

$$f(x) = \frac{dF(x)}{dx} \quad (2.2)$$

l'espérance mathématique d'une fonction $g(x)$ est l'intégrale

$$E\{g(x)\} = \int_{-\infty}^{+\infty} g(x)f(x)dx \quad (2.3)$$

On définit la fonction caractéristique de la variable aléatoire comme la transformée de Fourier de sa fonction de densité de probabilité $f(x)$ moyennant un changement de signe. Elle est donnée par [38] :

$$\Phi(\omega) = \int_{-\infty}^{+\infty} e^{j\omega x} f(x)dx = E\{e^{j\omega x}\} \quad (2.4)$$

cette fonction est aussi appelée *fonction de génération* des moments statistiques. En effet, les moments sont les coefficients des puissances de $j\omega$ dans son développement en série de Taylor

$$E\{e^{j\omega x}\} = E\left\{1 + (j\omega)x + \frac{(j\omega)^2 x^2}{2!} + \frac{(j\omega)^3 x^3}{3!} + \dots\right\} \quad (2.5)$$

soit encore

$$E\{e^{j\omega x}\} = 1 + (j\omega)E\{x\} + \frac{(j\omega)^2}{2!}E\{x^2\} + \frac{(j\omega)^3}{3!}E\{x^3\} + \dots \quad (2.6)$$

Les moments d'ordre n sont

$$E\{x^n\} = m_n \quad (2.7)$$

On pourra aisément remarquer que les moments sont des dérivées de la fonction caractéristique par rapport à $j\omega$ et calculées au point $\omega = 0$

$$m_n = \frac{d^n \Phi(0)}{d(j\omega)^n} \quad (2.8)$$

On définit une autre fonction $\Psi(\omega)$ comme *la seconde fonction caractéristique* ou *fonction de génération des cumulants*. Elle est donnée par

$$\Psi(\omega) = \ln \Phi(\omega) \quad (2.9)$$

Les cumulants sont alors

$$c_n = \frac{d^n \Psi(0)}{d(j\omega)^n} \quad (2.10)$$

Cependant, on se souviendra que l'évolution des signaux vibratoires relevés sur les machines tournantes n'est pas aléatoire. Les événements se reproduisent cycliquement et de façon "à peu près" identique [31]¹.

2.3 Indicateurs scalaires du domaine temporel

Dans le domaine temporel, il existe des indicateurs statistiques de longue tradition qui sont basées sur les notions de moments et de cumulants statistiques [38]. La série temporelle la plus utilisée en surveillance vibratoire est l'accélération. Bien que l'on se trouve dans le domaine temporel, il est important d'avoir à l'esprit que les valeurs des indicateurs associés à un signal vibratoire sont fonction de la plage fréquentielle choisie pour l'analyse. Il est donc indispensable de toujours associer à chaque indicateur la bande passante dans laquelle il a été calculé ou mesuré.

Les plus utilisés de ces indicateurs sont :

La valeur efficace (RMS) : c'est un indicateur temporel qui mesure le contenu énergétique dans une signature vibratoire. C'est un bon indicateur dans le suivi du niveau vibratoire, mais ne peut efficacement être utilisé pour un diagnostic. En

¹Le terme "à peu près" constitue la non-stationnarité ; il existe une similitude de la réponse vibratoire d'un cycle de fonctionnement à l'autre. Les paramètres statistiques des signaux varient périodiquement, on parle alors de cyclostationnarité.

effet, il représente la puissance totale du signal vibratoire induite par le fonctionnement de la machine et non uniquement la puissance vibratoire induite par le défaut recherché. Dans le cas de déséquilibre des systèmes rotatifs, son utilisation dans la détection est performante. Pour une série temporelle x_n de longueur N , la valeur efficace s'exprime par

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2} \quad (2.11)$$

La valeur RMS est la racine carrée du deuxième moment statistique m_2 .

La valeur efficace calculée dans des bandes fréquentielles données peut servir d'indicateur à très faible investissement qu'il est possible de suivre et ainsi estimer le moment d'intervenir sur la machine.

Kurtosis : il est un indicateur de la proportion d'échantillons qui s'écartent faiblement de la valeur moyenne comparée à ceux qui s'en écartent largement. Il caractérise l'allure aplatie d'une distribution par rapport à la distribution normale. Il peut être calculé de deux manières différentes :

$$Kurt = \frac{m_4}{m_2^2} - 3 \quad (2.12)$$

ou

$$Kurt = \frac{c_4}{c_2^2} \quad (2.13)$$

Ces deux définitions portent les mêmes informations. La présence d'un terme de puissance 4 donne un poids considérable aux amplitudes élevées. Le kurtosis permet de détecter l'apparition et de suivre l'évolution des défauts induisant des forces impulsionnelles périodiques.

Facteur de vrillage (Skewness) : Il caractérise le degré de dissymétrie d'une distribution autour de sa moyenne. Il est calculé par

$$Skew = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_n - \mu_x}{\sigma_x} \right)^3 \quad (2.14)$$

Facteur de crête : Le facteur de crête est un des indicateurs spécifiques qui accentuent l'apparition de chocs dans les vibrations. Les pics dans le signal temporel correspondront à un accroissement du facteur de crête. Cet indicateur, comme le

kurtosis, est utilisé pour la détection des anomalies qui se traduisent par des vibrations impulsives telles que la rupture de dents d'engrenage ou un défaut les éléments d'un roulement. A l'apparition du défaut, il reste constant, puis avec l'augmentation des chocs le niveau de crête devient plus élevé alors que la valeur efficace n'est pas influencée. Il en résulte une augmentation significative du facteur de crête. Quand la valeur efficace rejoint la valeur de crête par suite d'augmentation du nombre de chocs par cycle, le facteur de crête retombe presque à sa valeur initiale.

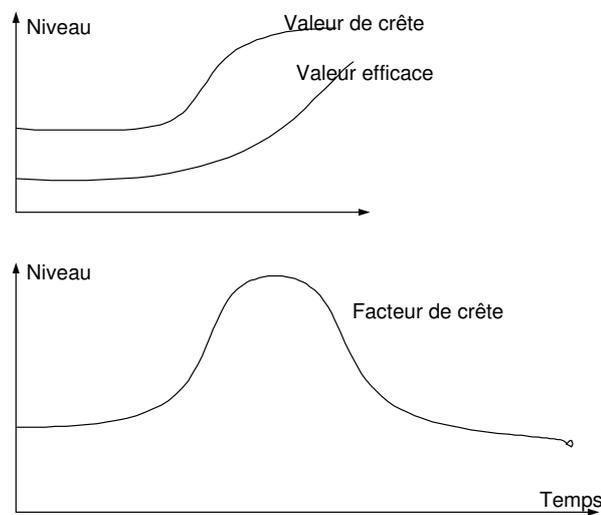


FIG. 2.1 – Evolution des valeurs de crête, efficace et du facteur de crête (adapté de [18])

L'utilisation des statistiques d'ordre supérieur peut aussi donner de bons indicateurs pour le diagnostic [14]. Ces indicateurs ont l'avantage de contenir l'information concernant les écarts d'une distribution relativement à un modèle gaussien et les phases de linéarité du processus vibratoire. L'annexe G présente les notions de bispectre et trispectre.

2.4 Les indicateurs spectraux

La surveillance basée sur les indicateurs *scalaires* calculés ou mesurés dans une gamme de fréquence déterminée présente trois inconvénients majeurs :

- la nécessité de définir au préalable, de mesurer et donc de gérer un grand nombre d'indicateurs destinés chacun à la surveillance d'un nombre restreint de défauts. La définition préalable de ces indicateurs entache la fiabilité de la surveillance, puisque l'apparition de tout défaut ou symptôme non prévu lors du paramétrage de la surveillance peut ne pas être détectée ou ne l'être que très tardivement ;
- l'impossibilité de surveiller le système dont le fonctionnement normal génère des chocs périodiques d'amplitudes élevées (compresseurs à vis, à piston, ...) susceptibles de masquer l'apparition de défauts induisant eux aussi des forces impulsionnelles (jeux, écaillage, ...), mais d'amplitude souvent nettement plus faibles ;
- l'impossibilité de détecter et de suivre l'évolution de défauts induisant une puissance vibratoire faible, voire très faible, mais susceptible de modifier sensiblement la forme du signal [3].

Il est apparu nécessaire de rendre plus fiable la surveillance par l'utilisation d'autres types d'indicateurs. Les indicateurs spectraux sont sensibles non seulement à l'augmentation de la puissance du signal, mais également aux évolutions de sa forme ou des amplitudes de chacune de ses composantes. Pour déterminer la structure du signal, la méthode générale consiste à en estimer le spectre de puissance par une Transformée de Fourier Discrète (DFT). D'autres méthodes spectrales sont aussi définies pour l'analyse fréquentielle des signaux vibratoires. L'annexe F.8 décrit l'analyse spectrale paramétrique qui est une alternative à la transformée de Fourier.

Pour un signal temporel $x(t)$ décrit par une série discrète de N valeurs x_n exprimées aux instants t_n , la DFT est définie comme suit

$$X_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N} \quad (2.15)$$

Cette relation est établie sur une fenêtre de N valeurs échantillonnées produisant ainsi des raies fréquentielles avec une résolution spectrale de f_{ech}/N , où f_{ech} est la fréquence d'échantillonnage.

Le vecteur d'entrée du réseaux de neurones peut alors être construit sur base de ces valeurs spectrales après, bien sûr, réduction dimensionnelle. Rangwala et Dornfeld [16], dans leurs travaux sur la surveillance d'outil d'usinage, ont appliqué cette procédure.

Si le signal contient des composantes aléatoires, la représentation classique de Fourier devient inappropriée à cause du continuum de fréquences dans le signal. On calcule alors le densité spectrale de puissance (PSD) par moyennage des énergies des DFT sur différents blocs du signal. La subdivision en blocs a le double avantage de réduire le nombre de raies spectrales et d'améliorer le rapport signal-bruit dans le spectre. Javadpour et Knapp [34] ont utilisé le vecteur de densité spectrale de puissance (PSD) pour nourrir un réseau de neurone de type ARTMAP. Mais il est aussi déploré le fait que le moyennage peut mener au masquage de certains phénomènes de courte durée [35].

Une autre difficulté d'application de la transformée de Fourier traditionnelle est rencontrée quand le signal est instationnaire. Une approche alternative pour l'analyse des signaux instationnaires et transitoires est l'analyse temps-fréquence. Différentes méthodes ont été utilisées dans la détection des défauts mécaniques. Koo et Kim [36] dans leur étude sur le diagnostic des pompes de centrale nucléaire ont appliqué la distribution de Wigner-Ville. Ils ont, pour alimenter un réseau de neurones à rétro-propagation du gradient, procédé d'abord à une réduction des données sur les deux axes (temps et fréquence) en ne s'intéressant qu'à des fréquences susceptibles de porter une information. Le vecteur d'entrée du réseau est ainsi de dimension $k = m \times n$, m et n étant les nombres réduits de données sur les axes de temps et de fréquence respectivement.

L'analyse par ondelettes est aussi d'application en surveillance et diagnostic par réseau de neurones des machines, particulièrement dans la reconnaissance des défauts des roulements [37] et dans le suivi de l'usure d'outil en usinage [20].

2.5 Interprétation d'un spectre de Fourier. Signatures vibratoires.

L'interprétation d'un spectre consiste à trouver l'origine cinématique des composantes les plus énergétiques. Afin de pouvoir déchiffrer des informations délivrées par la machine, il est conseillé de procéder à une analyse préalable aux mesures sur site afin de prédire quelles seront les composantes dominantes du signal : lorsqu'une composante apparaîtra sur le spectre, il sera alors possible de l'interpréter [19].

Nature du défaut	Fréquence dominantes	Directions
Déséquilibre	f	Radiale et axiale
Délignage	f et $2f$, parfois $3f$ et $4f$	Radiale et axiale
Fléxion de l'arbre	f et $2f$, parfois $3f$ et $4f$	Radiale et axiale
Fissuration de l'arbre	f et $2f$	Radiale et axiale
Roulement	Fréquences d'impact liées aux éléments constitutifs du roulement	Radiale et axiale
Paliers hydrodynamiques	entre $0,4$ et $0,5f$	Radiale
Engrenages	Liées aux nombres de dents (hautes fréquences)	Radiale et axiale
Jeux, contacts intermittents	$2f$	Radiale et axiale
Forces électromagnétiques	Fréquence d'alimentation et harmoniques (fonction du nombre de pôles)	Radiale et axiale
Cavitation, turbulence	hautes fréquences	Radiale et axiale

TAB. 2.1 – Fréquences (en fonction de la fréquence f de rotation du rotor) et directions caractéristiques de défauts affectant les machines tournantes.

2.6 Conclusions

La variété des indicateurs d'analyse vibratoire utilisés dans les différentes études citées ici montrent qu'en effectuer le choix reste une étape très liée à la nature du problème traité. Dans certains cas une normalisation des valeurs d'alimentation du réseau de neurone est appliquée pour améliorer la vitesse d'apprentissage et pour que les indicateurs à haute valeur ne viennent masquer les autres indicateurs (cf. 5.2.1).

Dans la seconde partie de ce propos nous abordons les aspects expérimentaux de nos travaux. Nous y discutons, en nous basant sur les performances de classification, de l'influence de différents indicateurs. Cette influence est déduite d'une analyse en composantes principales opérée sur l'espace des indicateurs dans le but d'en réduire la dimension .

Deuxième partie

Aspects expérimentaux

Chapitre 3

Les essais

3.1 Introduction

En vue d'explorer les possibilités et les performances d'une utilisation des réseaux de neurones pour la détection et la classification des défauts mécaniques complexes sur machines tournantes, nous avons conduit une série d'essais sur un banc d'essais.

La campagne complète a porté sur quatre défauts de natures différentes, chacun ayant plusieurs niveaux d'expression. Il s'agit de :

- défaut de balourd : 5 niveaux ;
- défaut de frottement : 3 niveaux ;
- ovalisation de la bague extérieure du roulement : 3 niveaux ;
- défaut d'alignement de la ligne d'arbre : 3 niveaux.

3.2 Le dispositif expérimental

Le dispositif expérimental consiste en un moteur électrique accouplé élastiquement à un arbre portant deux disques entre ses deux paliers. Sur les deux disques sont réalisés des trous taraudés de façon à pouvoir y fixer des masses pour créer des balourds de niveaux différents. Un levier portant un galet peut être appliqué sur un des disques afin d'introduire un frottement. Pour modifier le niveau de frottement, on varie la force de contact par suspension d'une masse différente sur le levier.

3.2 Le dispositif expérimental

Une vis de pression peut être serrée dans un palier pour créer des défauts d'ovalisation de la bague extérieure du roulement tandis que des défauts d'alignement sont obtenus en déplaçant un des paliers.

Les figures 3.1 et 3.2 montrent une photographie du banc et son illustration schématique.



FIG. 3.1 – Banc d'essais simulateur de défauts mécaniques.

Nous nous sommes proposés d'analyser le comportement du dispositif sur 3 vitesses de rotation différentes obtenues à l'aide d'un variateur de vitesse de type "EUROTHERM 601" contrôlant le moteur. Les fréquences de rotation retenues sont : 25 Hz, 22,5 Hz et 20 Hz.

L'acquisition des signaux vibratoires a été effectuée à l'aide du système d'acquisition OROS à quatre canaux comprenant :

- un PC portable Compaq HP Presario 2510EA/P4 2.4/256 MB/30 GB
- une carte d'acquisition, un analyseur FFT en temps réel 4x20 kHz OR24-PCP-4
- un accéléromètre industriel (PCB piezotronics) M 622A01 (SN 7470) de sensibilité $10,3 \text{ mV}/(\text{m}/\text{s}^2)$

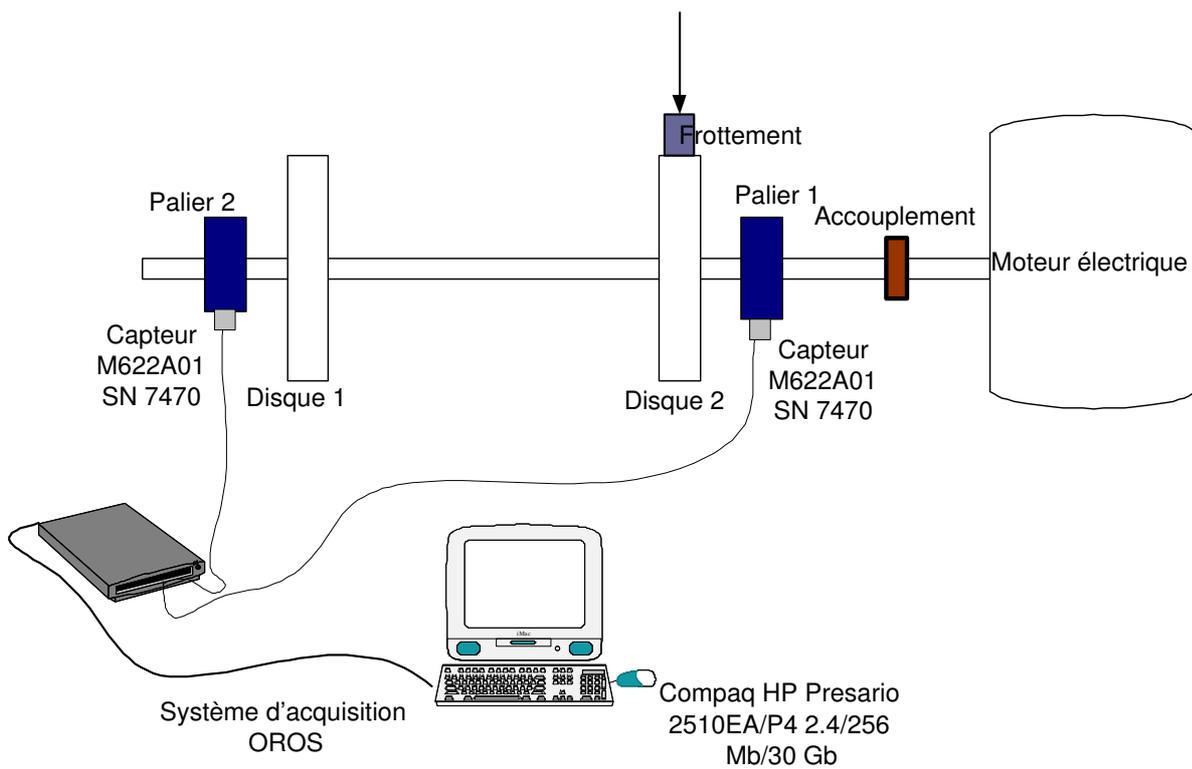


FIG. 3.2 – Représentation schématique du banc.

3.3 Procédure des essais

Après avoir imposé un défaut et mis en marche le banc d'essai, les mesures sont relevées sur chaque palier suivant deux directions radiales perpendiculaires (Z : verticale et Y : horizontale). Les points de mesure sont notés de la manière suivante :

- Z1 : direction verticale sur le palier 1
- Y1 : direction horizontale sur le palier 1
- Z2 : direction verticale sur le palier 2
- Y2 : direction horizontale sur le palier 2

Une condition de fonctionnement est complètement caractérisée par la vitesse de rotation, les défauts simulés et leurs niveaux respectifs.

Pour chaque condition, nous avons enregistré un fichier temporel dont les principaux attributs sont :

- gamme fréquentielle : 0-2000 Hz
- durée de l'enregistrement T :20 secondes
- taille de l'échantillon :102400
- pas de temps $\Delta t : 1,9531 \times 10^{-4}$ s

Des spectres fréquentiels provenant d'une FFT en temps réel ont été enregistrés suivant deux gammes de fréquences en chaque point de mesure et pour chaque condition :

1. Gamme 1 : 0-200 Hz
 - Résolution fréquentielle :0,125 Hz
2. Gamme 2 : 0-2000 Hz
 - Résolution fréquentielle :1,25 Hz

3.3.1 Les précautions

Afin de conduire à bien les mesures et de s'assurer une provision de répétabilité des essais, les précautions suivantes ont dû être observées :

- observation systématique d'un temps stabilisation à chaque changement de condition avant de procéder à la mesure, ceci permet d'éviter les phénomènes transitoires ;

- soin particulier dans la fixation du capteur et meilleure immobilisation du câble de transmission du signal pour minimiser les bruits de mesure.

3.4 Codification des défauts et des fichiers

Pour faciliter le traitement, un système de codification a été adopté. Il comporte toute l'information devant caractériser complètement la condition simulée. Le tableau 3.1 reprend tous les défauts qui font l'objet de la campagne complète d'essais.

Défaut	Niveau	Signification	Désignation
Balourd	0	machine supposée exempte de défaut de balourd	B0
	1	Masse de 20 gr sur le premier disque	B1
	2	Masse de 40 gr sur le premier disque	B2
	3	Masse de 20 gr sur le second disque	B3
	4	Masse de 40 gr sur le second disque	B4
Frottement	0	Machine supposée sans défaut de frottement	F0
	1	Force normale modérée au point de contact	F1
	2	Force normale élevée au point de contact	F2
Défaut de roulement (ovalisation)	0	Machine supposée sans ce défaut	R0
	1	Vis de pression modérément serrée	R1
	2	Vis de pression suffisamment serrée	R2
Désalignement	0	Pas de défaut d'alignement	A0
	1	Faible désalignement	A1
	2	Désalignement relativement grand	A2

TAB. 3.1 – Défauts simulées

La codification comprend aussi une information sur le niveau de fréquence de rotation et la gamme de fréquences de mesure. Les fréquences de rotation du moteur sont désignées par les chiffres 1, 2 et 3

- fréquence 1 : 25 Hz
- fréquence 2 : 20 Hz

- fréquence 3 : 22.5Hz

Ainsi, par exemple, un fichier noté H1B1F2R0A0 désignera un enregistrement haute fréquence (0-2 kHz) à la fréquence de rotation de 25 Hz sur une condition de balourd1, de frottement2 et sans défaut de roulement ni d'alignement.

L'inventaire complet des combinaisons des défauts avec leur différents niveaux donne 135 conditions à simuler par niveau de vitesse soit au total 405 essais. Les 135 conditions correspondent à 5 cas de balourd x 3 cas de frottement x 3 cas de défauts de roulement x 3 cas d'alignement, y compris le niveau "zéro" pour chaque type de défauts.

Nous avons, au lieu de simuler toutes ces conditions, conduit des essais seulement sur 162 conditions choisies de façon à avoir une représentabilité de chaque niveau de défaut. L'annexe B reprend toutes les conditions traitées dans cette étude.

3.5 Préparation des données

Le système OROS permet l'obtention en temps réel de spectres fréquentiels. Pour des raisons de facilité de traitement numérique sous Matlab, nous n'avons exploité que les enregistrements temporels données par OROS sous format ".wav" après leur conversion en format ".mat".

Pour le stockage des 4 enregistrements (Z1, Y1, Z2, Y2), une matrice dont les colonnes reprennent chacun des 4 enregistrements a été construite pour chaque condition de fonctionnement.

3.5.1 Traitement dans le domaine temporel

Il a été procédé, grâce aux codes écrits sous Matlab, au calcul des indicateurs du domaine temporel dans des gammes de fréquence spécifiques après filtrage numérique. Les signaux dans les gammes suivantes ont été obtenus avec des filtres numériques de Butterworth d'ordre 5 :

- 0-30 Hz
- 30-300 Hz

- 300-2000 Hz

Chaque série temporelle a été subdivisée en 4 blocs de taille 25600 points sans recouvrement afin d'introduire l'effet du bruit de mesure. Cette procédure permet d'analyser la robustesse du réseau de neurones face à de tels bruits et d'augmenter les tailles des ensembles d'apprentissage et de test.

Le vecteur d'entrée du réseau de neurones est constitué, pour une condition donnée, des indicateurs statistiques calculés du signal brut et/ou de signaux filtrés dans les trois gammes ci-dessus.

Une réduction dimensionnelle est opérée par analyse en composantes principales pour faciliter l'apprentissage du réseau de neurones.

3.5.2 Traitement dans le domaine fréquentiel

Dans le domaine fréquentiel, nous avons utilisé la transformée discrète de Fourier. Nous avons calculé, pour chaque condition de fonctionnement, un spectre de 512 lignes obtenues par moyennage d'énergie sur de fenêtres sans recouvrement de longueur 1024 points [39].

Puisqu'on ne peut s'imaginer nourrir un réseau de neurone avec les vecteurs de valeurs spectrales dont la taille est très grande, il s'imposait une réduction de la dimension pour laquelle deux choix étaient possibles :

1. L'extraction des raies caractéristiques : on forme un vecteur d'entrée constitué des valeurs correspondant aux seules fréquences caractéristiques, notamment les multiples de la fréquence de rotation. L'idée n'est pas d'extraire exactement les raies des multiples de la fréquence de rotation, mais plutôt les raies les plus grandes dans les intervalles judicieusement choisis autour des fréquences multiples de la fréquence de rotation.
2. La réduction dimensionnelle par une analyse multivariée : on procède à une analyse en composantes principales pour ne retenir que les directions principales qui recèlent une quantité d'information élevée.

L'organigramme de la figure 3.3 montre, sous une forme condensée, le traitement de préparation subi par les données avant l'alimentation du réseau de neurones.

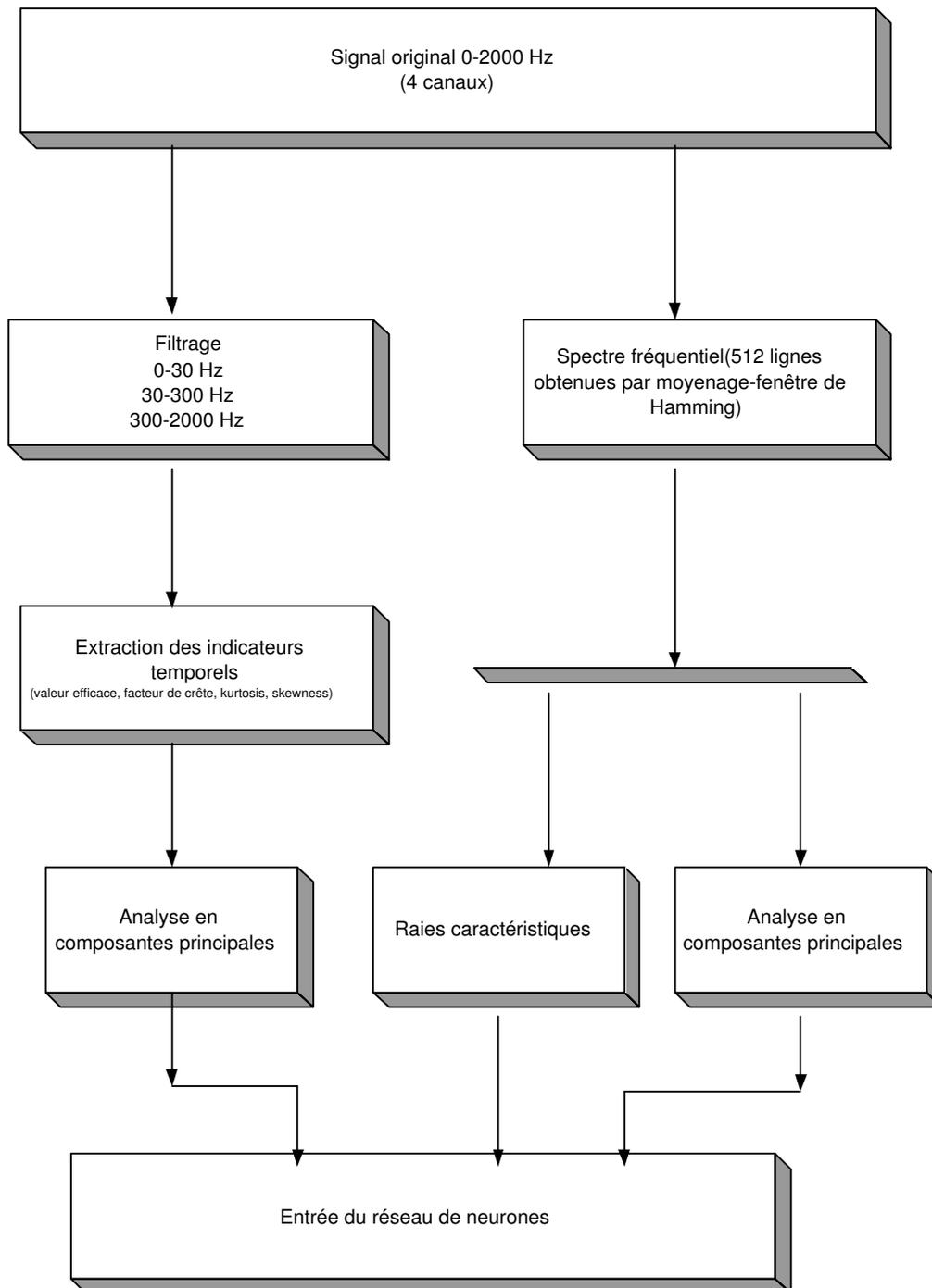


FIG. 3.3 – Traitement de préparation des données

3.6 Effet de la variation de vitesse

La variation de la fréquence de rotation est un phénomène dont il faut tenir compte dans toute procédure de détection de défauts mécaniques sur les machines tournantes. Deux effets importants sont envisageables : d'une part, la position des composantes spectrales synchrones avec la vitesse de rotation peut varier dans des proportions importantes et d'autre part, une vibration est l'image du produit de convolution d'une force d'excitation par la réponse impulsionnelle de la structure sur laquelle est fixée le capteur. Dans le domaine fréquentiel, les grandeurs caractéristiques du signal vibratoire dépendent du produit de l'amplitude de chaque composante spectrale de la force d'excitation par les valeurs de la fonction de transfert associées aux fréquences de ces mêmes composantes [3].

Or, d'une manière générale, l'amplitude des composantes spectrales de la force d'excitation croissent avec la vitesse de rotation. Dans le cas particulier des balourds, cette amplitude croît proportionnellement au carré de la fréquence de rotation et l'on peut ainsi s'affranchir de l'effet de variation de vitesse sur l'amplitude en normalisant les valeurs spectrales par rapport au carré de la fréquence de rotation. Ceci n'est pas nécessairement vrai pour les autres types de défauts.

Pour un diagnostic basé sur la comparaison des spectres obtenus aux spectres gabarits, la considération des variations de vitesse est impérative. Une variation inférieure à 10% peut généralement être compensée et une comparaison des spectres est admissible, mais pour une variation de plus de 10%, il faut nécessairement un autre gabarit [30].

Dans cette étude où la variation maximale de vitesse est de 25%, du fait de la présence de la fréquence intermédiaire (22,5 Hz), les performances de détection peuvent être assurées pour toute fréquence de rotation comprise entre la fréquence maximale de 25 HZ et la fréquence minimale de 20 Hz. Les conditions de fonctionnement aux fréquences de rotation en dehors de la plage 20-25 Hz mais pour lesquelles la variation reste inférieure à 10% peuvent aussi, en principe, être reconnues.

3.7 Caractéristiques spectrales des défauts traités

3.7.1 Le balourd

Nous illustrons sur les figures 3.4 à 3.6 les spectres aux 3 vitesses des conditions de balourd de niveau 1 et de niveau 2 et de la condition normale, un zoom sur la zone 0-200 Hz est faite afin de rendre visibles ces phénomènes qui se passent dans les basses fréquences. On sait y remarquer l'effet du changement de vitesse par la différence des niveaux.

On peut aussi reconnaître l'effet de l'excentricité du rotor du moteur électrique à deux fois la fréquence du réseau électrique soit à 100 Hz.

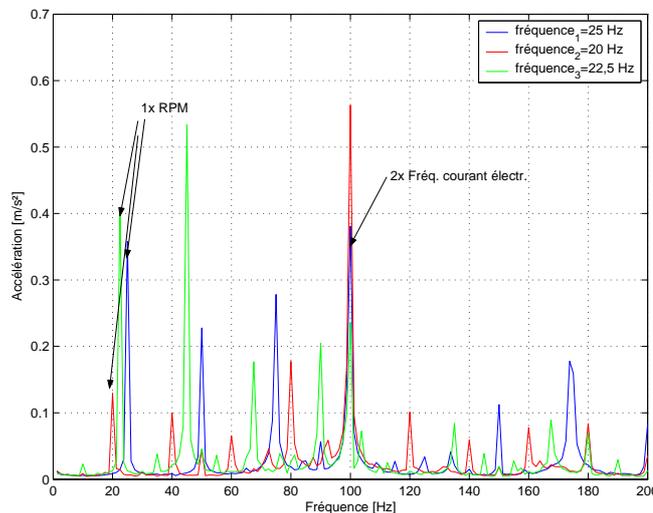


FIG. 3.4 – Zoom sur la gamme 0-200 Hz pour le balourd B1 aux 3 fréquences de rotation (Canal Z2).

Un autre fait remarquable est qu'à la fréquence de rotation de 22,5 Hz les niveaux sont plus élevés qu'à la fréquence de 25 Hz. Un tel fait peut être expliqué par la possibilité que cette fréquence soit proche d'une résonance.

A ce niveau, nous pouvons nous questionner sur la possibilité pour un classificateur automatique de distinguer les différents niveaux de balourd. La figure 3.7 montre une forte similarité entre les spectres des balourds et fait prédire une relative difficulté de diagnostic pour ce cas.

3.7 Caractéristiques spectrales des défauts traités

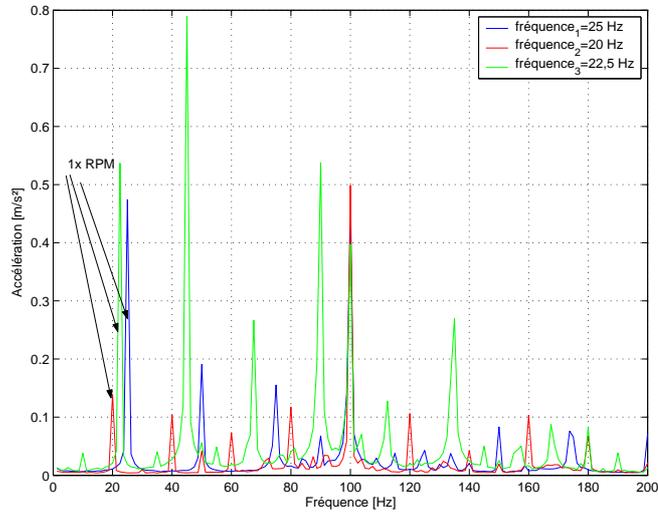


FIG. 3.5 – Zoom sur la gamme 0-200 Hz pour le balourd B2 seul aux 3 fréquences de rotation (Canal Z2).

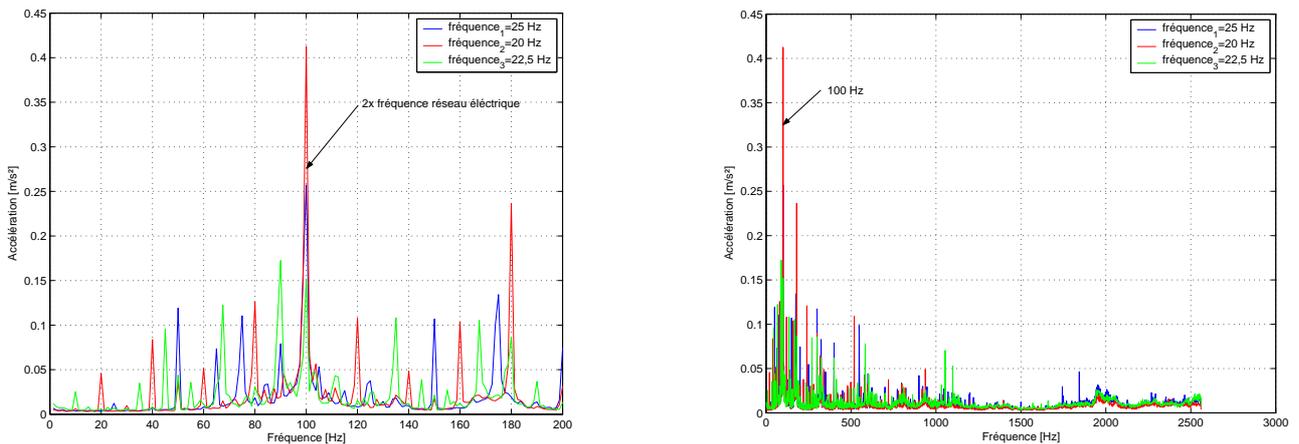


FIG. 3.6 – La condition normale aux 3 fréquences de rotation (Canal Z2). Zoom sur la gamme 0-200 Hz (à gauche) et Spectre global (à droite)

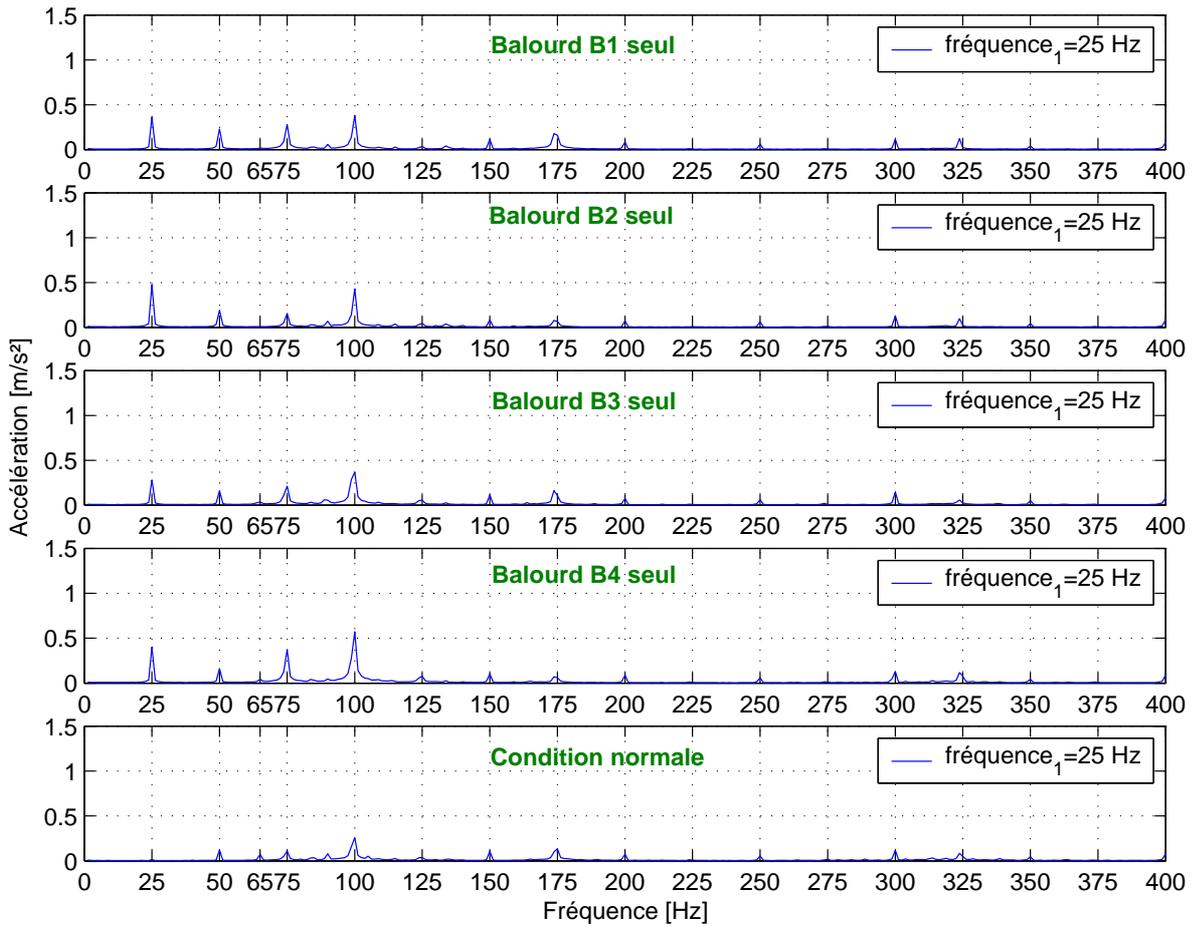


FIG. 3.7 – Balourds de niveaux différents : les spectres se ressemblent.

3.7.2 Le frottement

Sur la figure 3.8 est représentée la condition de frottement sévère F2 pour le signal relevé au point Z2. Nous remarquons que dans la zone des basses fréquences rien de particulier n'est observable excepté l'effet de la fréquence du courant électrique à 100 Hz, mais l'examen du spectre global révèle des pics de niveaux élevés en hautes fréquences. Le bruit de fond y est aussi élevé comparé au cas de la condition normale de la figure 3.6b.

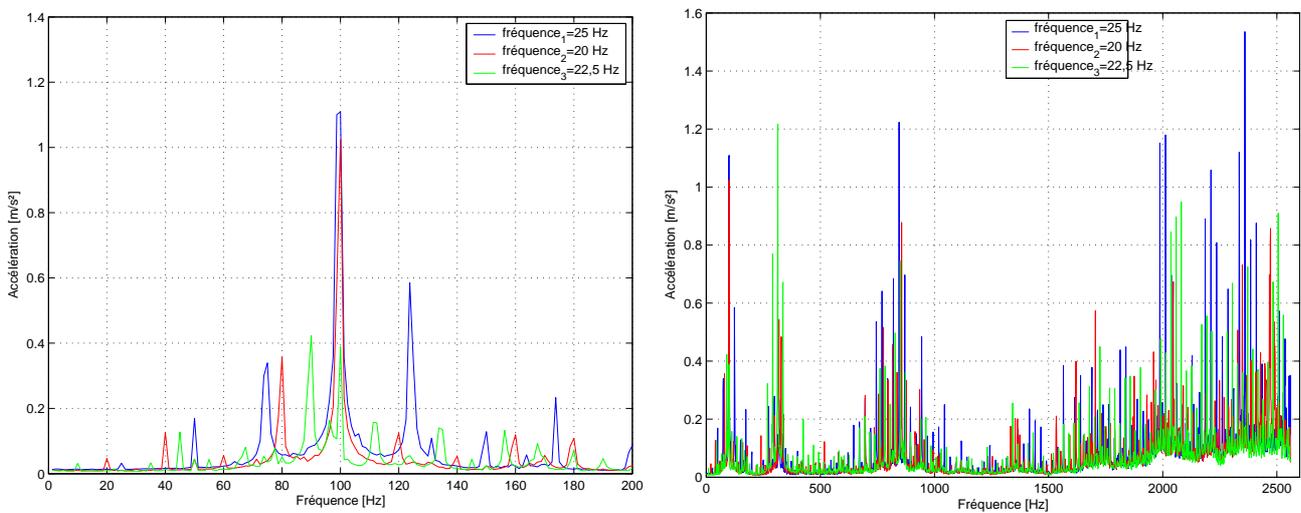


FIG. 3.8 – La condition F2 aux 3 fréquences de rotation (Canal Z2). Zoom sur la gamme 0-200 Hz (à gauche) et Spectre global (à droite)

3.7.3 L'ovalisation de la bague externe du roulement

Les spectres ci-dessous sont associés à la condition R1 de déformation de niveau 1 de la bague externe du roulement. La figure 3.9 révèle dans les basses fréquences l'apparition de nouvelles raies non synchrones avec les fréquences de rotation et un autre phénomène se répétant avec des raies latérales remarquables facilement dans la zone hautes fréquences.

En examinant séparément les spectres correspondant à chaque vitesse (figures 3.10 à 3.12), nous voyons la présence d'un pic à environ $2,6 \times RPM$ qui doit correspondre à la fréquence caractéristique de la bague externe du roulement.

3.7 Caractéristiques spectrales des défauts traités

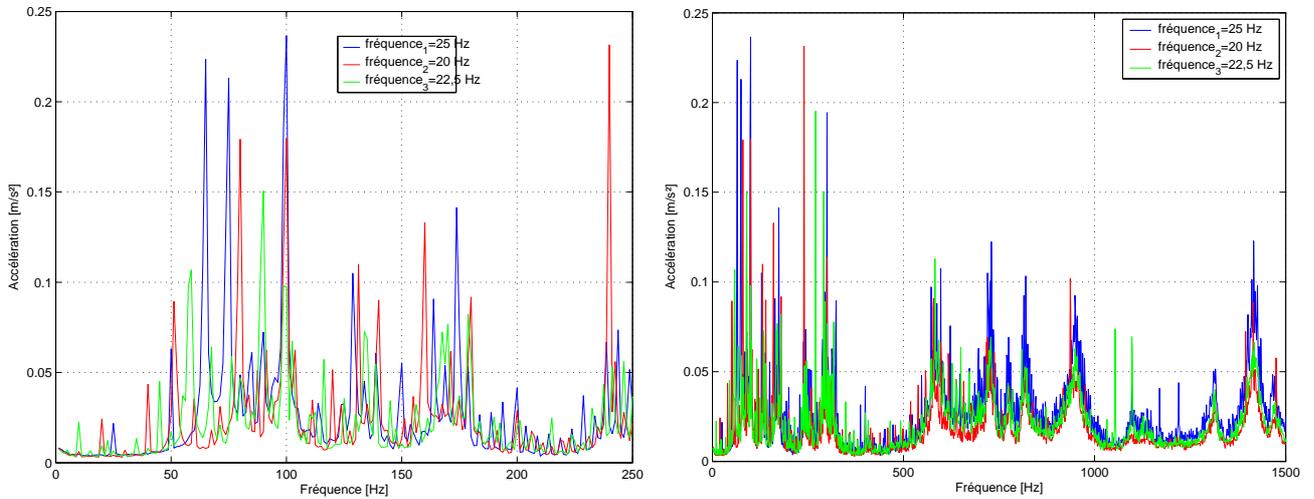


FIG. 3.9 – La condition R1 aux 3 fréquences de rotation (Canal Z2). Zoom sur la gamme 0-250 Hz (à gauche) et Spectre global (à droite)

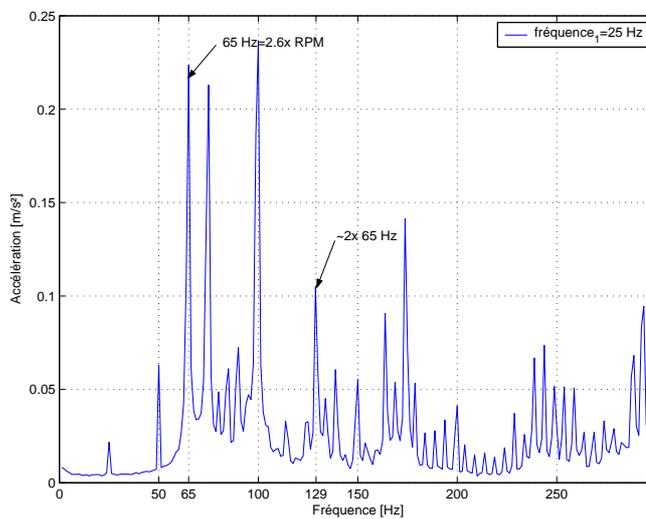


FIG. 3.10 – Zoom sur la gamme 0-300 Hz pour le défaut R1 seul à la fréquence de rotation de 25 Hz (Canal Z2).

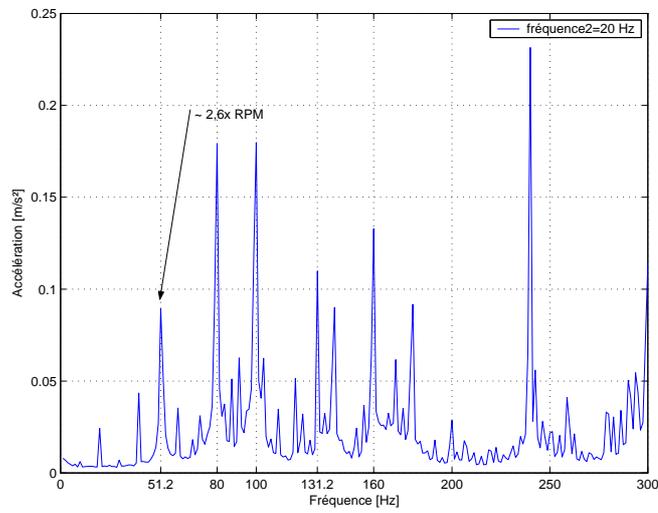


FIG. 3.11 – Zoom sur la gamme 0-300 Hz pour le défaut R1 seul à la fréquence de rotation de 20 Hz (Canal Z2).

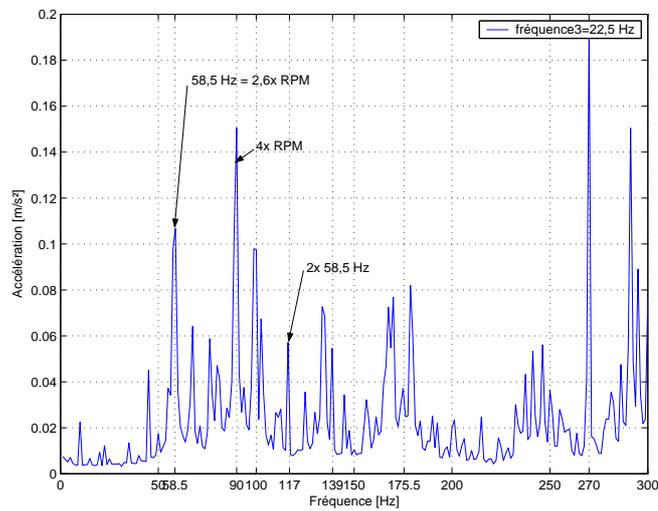


FIG. 3.12 – Zoom sur la gamme 0-300 Hz pour le défaut R1 seul à la fréquence de rotation de 22,5 Hz (Canal Z2).

3.7.4 Le délignage

Comme nous n'avons pas, dans la série de nos essais, traité isolément un cas de désalignement, nous examinerons une combinaison de ce type de défaut avec un autre, par exemple le balourd. Nous allons nous intéresser à la condition B4F0R1A1 c'est-à-dire un balourd de niveau 4 avec une ovalisation du roulement et un désalignement de niveau 1. Nous examinons, sur les figures 3.13 à 3.15, successivement l'évolution de la structure spectrale quand un défaut est ajouté. On part du balourd B4 seul auquel on ajoute le défaut de roulement R1 puis le désalignement A1.

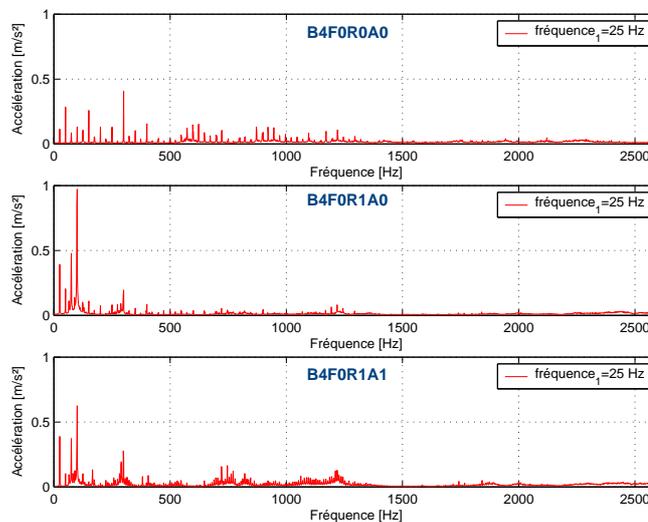


FIG. 3.13 – Spectres globaux à la fréquence de rotation de 25 Hz (Canal Z1).

L'examen dans les basses fréquences à partir des figures 3.14 et 3.15 montre que l'ajout du défaut de roulement amplifie les raies déjà présentes avec le balourd et fait apparaître d'autres pics, particulièrement celui qui a été identifié à $2,6 \times RPM$, donc à 65 Hz pour la fréquence de rotation de 25 Hz. Cette raie à 65 Hz est tout de même présente en condition sans défaut de roulement mais à avec un niveau très faible. Sa mise en évidence est rendue possible avec l'utilisation d'une échelle logarithmique. On peut déduire de tout ceci qu'avant d'imposer une ovalisation au roulement, celui-ci présentait déjà de petits chocs.

Le défaut d'alignement vient, par contre, atténuer les niveaux qui ont été amplifiés par le défaut de roulement dans les basses fréquences.

3.7 Caractéristiques spectrales des défauts traités

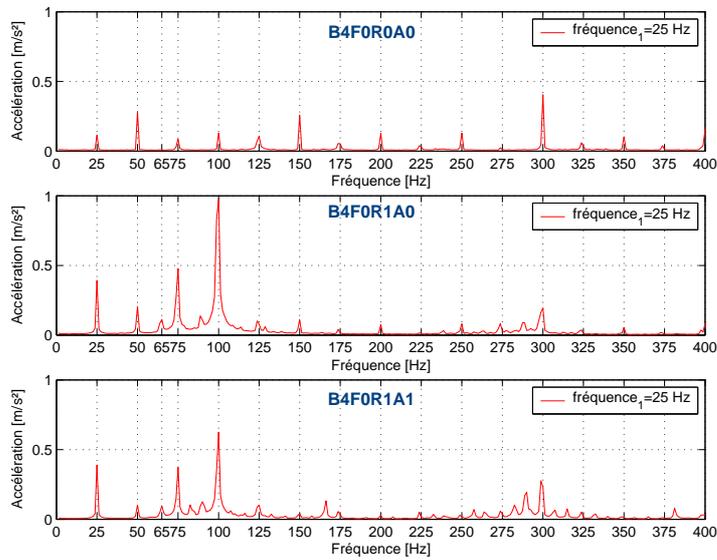


FIG. 3.14 – Zoom sur la gamme 0-400 Hz à la fréquence de rotation de 25 Hz (Canal Z1).

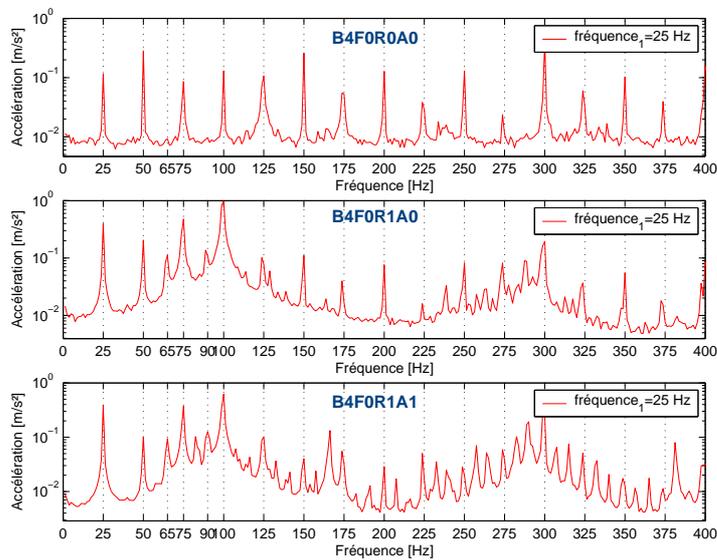


FIG. 3.15 – Zoom sur la gamme 0-400 Hz à la fréquence de rotation de 25 Hz en échelle logarithmique pour amplifier les phénomènes de faibles amplitudes (Canal Z1).

3.8 Conclusion

Nous avons aussi observé que le deuxième niveau de désalignement atténuait les amplitudes à de nombreuses fréquences tout en en faisant apparaître à d'autres fréquences. La figure 3.16 fait ressortir cette observation.

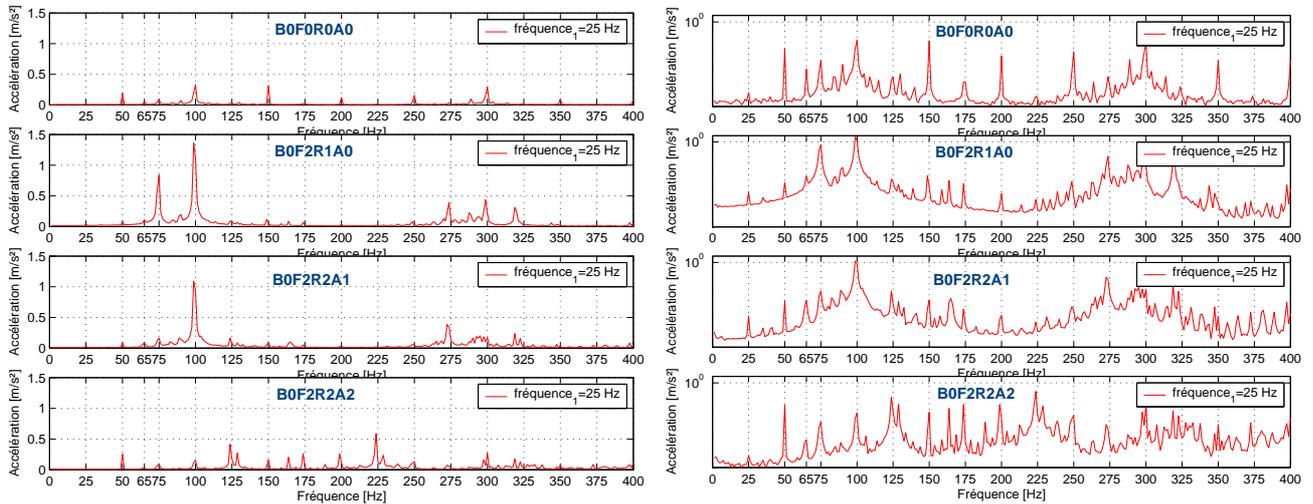


FIG. 3.16 – Effet de niveau de désalignement en échelles linéaire (à gauche) et logarithmique (à droite).

3.8 Conclusion

Nous venons de présenter les défauts que nous avons traité sur le banc simulateur des défauts mécaniques. Les caractéristiques spectrales présentées ici n'ont eu pour objectif que de mettre en évidence les différences apparaissant dans la forme des spectres fréquentiels. C'est ainsi qu'il peut déjà être prédit à partir de la figure 3.7 que les différents balourds pourront être distingués de manière relativement difficile par un classificateur.

Notre étude expérimentale a été conduite en deux phases qui sont décrites dans les chapitres qui suivent. La première phase peut être vue comme une mise en train dans le processus du traitement des données et la construction du réseau de neurones. Aussi n'avons-nous considéré d'abord que deux types de défauts. La deuxième phase a bénéficié de toutes les leçons tirées au cours de la première phase. On remarquera donc que les techniques de traitement des données ainsi que l'architecture des réseaux de neurones sont très différentes.

Chapitre 4

Première phase expérimentale. Simulation de deux défauts : le balourd et le frottement

Dans une première phase, seuls deux types de défauts ont été considérés : le balourd et le frottement. Les deux défauts ont été simulés aussi bien seuls que combinés en tenant compte de leurs différents niveaux d'expression. Le tableau 4.1 donne une caractérisation des défauts qui ont été simulés au cours de cette phase.

Défaut	Niveau	Signification
Balourd	0	machine supposée exempte de défaut de balourd
	1	Masse 1 sur le premier disque
	2	Masse 2 sur le premier disque
	3	Masse 1 sur le second disque
	4	Masse 2 sur le second disque
Frottement	0	Machine supposée sans défaut de frottement
	1	Force normale modérée au point de contact
	2	Force normale élevée au point de contact

TAB. 4.1 – Les défauts simulés au cours de la première phase d'essais

Cette première étape de l'expérimentation a consisté en l'étude de la détectabilité de deux défauts concourants : le balourd et le frottement. Elle nous a aussi permis

Condition	Codification	Sortie Désirée
Normale	B0F0R0A0	0000
Frottement1	B0F1R0A0	0100
Frottement2	B0F2R0A0	0200
Balourd1	B1F0R0A0	1000
Balourd1 et frottement1	B1F1R0A0	1100
Balourd1 et frottement2	B1F2R0A0	1200
Balourd2	B2F0R0A0	2000
Balourd2 et frottement1	B2F1R0A0	2100
Balourd3	B3F0R0A0	3000
Balourd3 et frottement1	B3F1R0A0	3100
Balourd4	B4F0R0A0	4000
Balourd4 et frottement1	B4F1R0A0	4100

TAB. 4.2 – Conditions étudiées dans la phase1

de réaliser des essais sur l'architecture du réseaux de neurones et sur le traitement des données. Le tableau 4.2 fournit, en regard de chaque condition, la sortie désirée pour le réseau de neurones. Les quatre chiffres représentent, en effet, les quatre composants du vecteur "cible" du réseau de neurones et correspondent au niveau de chacun de quatre défauts dans la condition de fonctionnement considérée. Remarquons que les vecteurs "cibles" du réseau caractérisent les défauts quelle que soit la vitesse de rotation du banc de simulation.

4.1 Construction du réseau de neurones

La construction d'un réseau de neurone est, dans un large nombre de cas, un processus par essai et erreur. Pour concevoir un réseau de neurone de haute performance les paramètres relatifs aussi bien à l'apprentissage (taux d'apprentissage, algorithmes, ...) qu'à la structure du réseau de neurones doivent être regardés simultanément.

Pour l'apprentissage, il existe des paramètres de nuisance tels que l'ajustage

initial des poids synaptiques, les tailles et leur rapport pour les ensembles d'apprentissage et de test. Une bonne conception doit être robuste vis-à-vis de ces paramètres.

Dans la tâche de construction du réseau de neurones, les degrés de liberté ajustables sont :

1. Type de réseau

Nous avons choisi d'utiliser le réseau à couches unidirectionnel et rétropropagation du gradient (multilayers feedforward) qui est le type de réseau le plus utilisé.

2. Architecture

Un réseau de neurones est d'autant plus puissant qu'il a des couches cachées et un nombre élevé de noeuds dans ces couches. Les fonctions sigmoïdes sont les plus indiquées pour mieux approcher toute non linéarité.

(a) *Nombre de couches*

Il a été retenu, après plusieurs essais, une structure à 4 couches.

(b) *Nombre de neurones par couches*

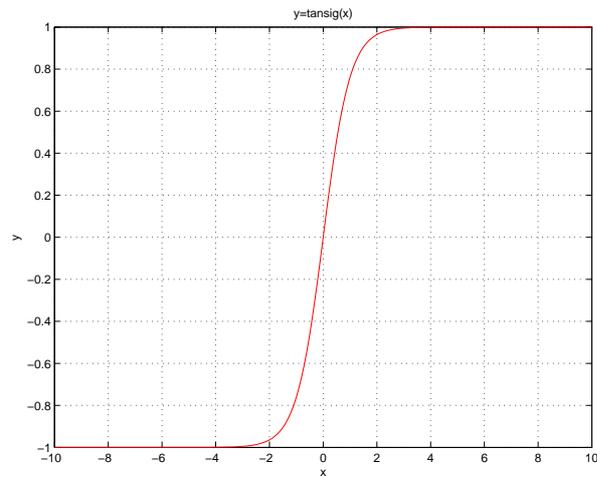
Un réseau avec plusieurs neurones apprend mieux mais présente un risque assez grand de surapprentissage.

Le nombre de neurones dans la couche de sortie est fixé par la dimension du vecteur "cible" qui est, dans cette étude, 4.

(c) *Fonctions d'activation des couches*

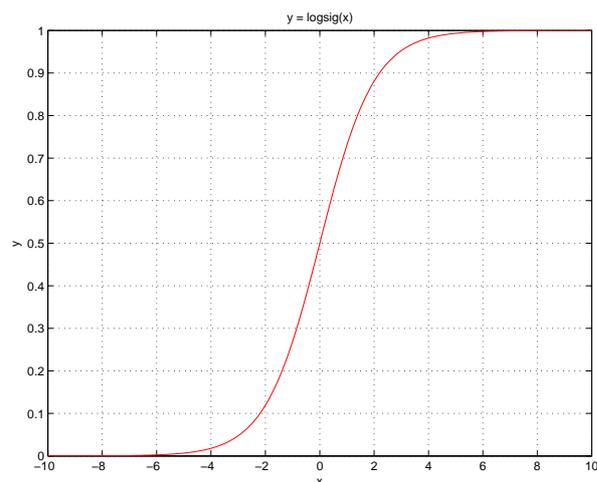
Dans les couches cachées nous avons utilisé les fonctions sigmoïdes et à la couche de sortie une fonction linéaire non saturée positivement. La fonction tangente sigmoïde est donnée par

$$\text{tansig}(n) = \frac{2}{1 + \exp(-2n)} - 1 \quad (4.1)$$



et la fonction sigmoïde log par

$$\text{logsig}(n) = \frac{1}{1 + \exp(-n)} \quad (4.2)$$



Les fonctions linéaires sont

$$\text{purelin}(x) = x \quad (4.3)$$

$$\text{poslin}(x) = \begin{cases} x & \forall x \geq 0 \\ 0 & \text{ailleurs} \end{cases} \quad (4.4)$$

(d) Algorithmes d'apprentissage

Deux algorithmes de rétropropagation du gradient implémentés dans le "neural network toolbox" de Matlab ont été expérimentés. Il s'agit de :

- l'algorithme du gradient conjugué "trainscg" : cet algorithme est bien dédié aux problèmes de classification dans le cas de grands réseaux et présente l'avantage d'avoir moins d'exigence de stockage en mémoire pendant le calcul et d'être suffisamment rapide.
- l'algorithme de la descente du gradient à taux variable d'apprentissage "traingdx" : celui-ci converge lentement. Rappelons qu'il peut parfois être préférable d'avoir une convergence lente ; c'est le cas des situations où l'on désire obtenir une erreur de validation la plus faible possible ;

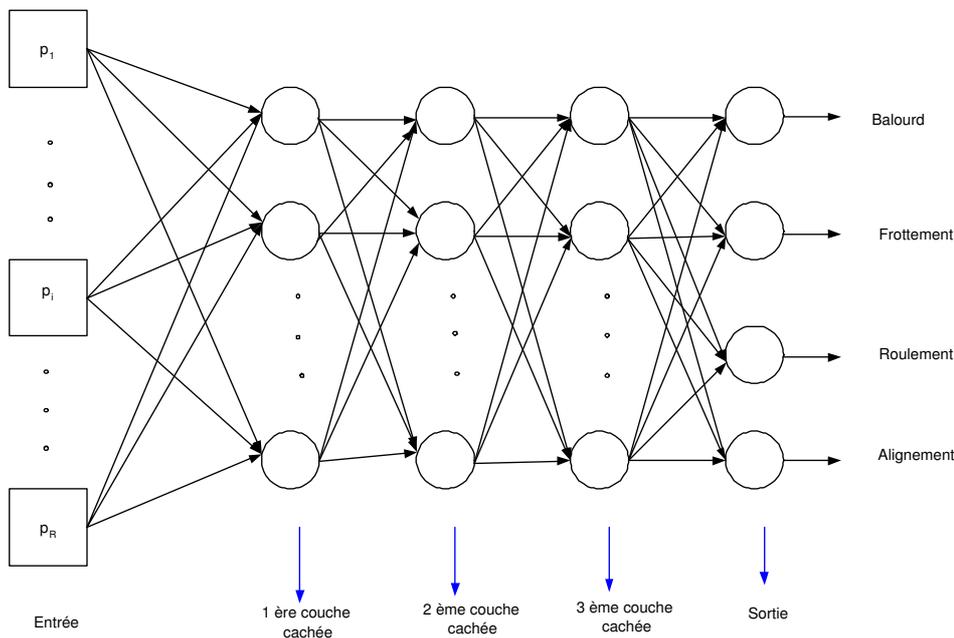


FIG. 4.1 – Structure du réseau de neurones

4.2 Le vecteur d'entrée. Domaine temporel

A ce stade, nous nous proposons d'analyser deux axes importants à savoir :

- la conception du réseau de neurones pour de meilleures performances ;
- et l'étude de la sensibilité des réponses du réseau de neurones par rapport aux différents indicateurs.

Une évolution simultanée sur les deux axes est moins évident parce que chaque axe est multifactoriel. C'est ainsi que nous avons décidé de figer la structure du ré-

seau à celle qui présente, avec "un minimum de noeuds", le meilleur score de classification pour une famille d'indicateurs initiaux préalablement choisis. Ceci nous permet d'investiguer l'influence des indicateurs sur les performances du réseau de neurones.

Les indicateurs que nous avons utilisés sont extraits des statistiques du domaine temporel. Les familles suivantes ont été testées :

1. *signal original non filtré* : valeur efficace, facteur de crête, kurtosis et skewness calculés pour les quatre points de mesure. Le vecteur ainsi construit est un vecteur de dimension 16 (4 indicateurs par point de mesure) ;
2. *signaux filtrés* : analyse conduite avec le seul canal Z1. Les indicateurs calculés sur les signaux filtrés ont considérés seuls ou joints à ceux provenant du signal original pour former le vecteur d'entrée ;
3. *signaux non filtrés* : analyse conduite avec le seul canal Z1 pour mieux comparer les résultats obtenus avec les signaux filtrés à ceux issus du signal original non filtré sur ce canal.

4.3 Conduite de l'apprentissage et du test

En divisant chaque signal en 4 blocs, nous avons constitué un ensemble de 144 signaux à partir duquel devaient être construits les ensembles d'apprentissage et de test. Trois quarts des données ont servi à l'apprentissage et un quart au test. Ces ensembles ont été construits par extraction des colonnes régulièrement espacées de la matrice des données P qui, elle, a 144 colonnes. La même démarche a été suivie pour la matrice des cibles T. L'extrait de code suivant a servi à cet effet

```
[R Q]=size(P);  
indi_app=[1:4:Q 3:4:Q 4:4:Q];  
indi_tst=2:4:Q;  
  
% ensemble de données d'apprentissage  
Ptrain=P(:,indi_app);
```

```
% ensemble de sorties désirées pour l'apprentissage
Ttrain=T(:,indi_app);

% ensemble de données de test
Ptest=P(:,indi_tst);

% ensemble de sorties désirées pour le test
Ttest=T(:,indi_tst);
```

4.3.1 Bruitage des données

Lors de l'apprentissage un bruitage a été effectué sur les données par addition d'un bruit blanc. L'objectif poursuivi est de comparer les performances du réseau entraîné dans les conditions normales avec celles du réseau entraîné dans les conditions de bruit. Le bruitage a permis aussi de prévenir le phénomène de mémorisation lors de l'apprentissage du réseau de neurones vu que sa complexité (nombre de couches cachées et de neurones) n'était pas négligeable.

Pour cela, dix ensembles des données bruitées ont été présentés successivement au réseau après passage sur les données non bruitées. Voici un extrait de code :

```
%copie du réseau préalablement entraîné sans bruit
netn=net;
% nouvel ensemble "cible"
Tn=[Ttrain Ttrain Ttrain Ttrain];
for passage=1:10
    fprintf('Passage=%.0f\n',passage)

%addition d'un bruit de niveau 1
    pn1=(Ptrain+randn(size(Ptrain))*0.1);

%addition d'un bruit de niveau 2
    pn1=(Ptrain+randn(size(Ptrain))*0.2;
```

```
%nouvel ensemble d'apprentissage
    Pn=[Ptrain,Ptrain,pn1 , pn2];

%apprentissage
    [netn,tr]=train(netn,Pn,Tn);
end
```

Il s'avérait ensuite fondé de refaire un passage sur les données non bruitées pour s'assurer qu'elles restaient bien catégorisées.

Les sorties du réseaux de neurones sont arrondies à l'entier le plus proche pour permettre leur comparaison aux cibles qui sont des entiers.

4.4 Résultats et discussion

Nous allons analyser les performances de classification atteintes en fonction des indicateurs utilisés et du prétraitement des données.

4.4.1 Détection des deux types de défauts

Dans tous les cas, la détection du balourd et du frottement, sans considération de niveau d'expression, s'est réalisée de manière satisfaisante.

La présence du frottement a été détectée en moyenne dans 98% des cas tant avec le réseau entraîné sous bruit qu'avec le réseau sans bruit.

La présence du balourd a été détectée dans 93% des cas pour le réseau sans bruit et dans 94% des cas pour le réseau avec bruit.

La condition supposée normale de la machine a été, elle, assez mal catégorisée : 64% dans le cas sans bruit et 67% dans le cas avec bruit. Elle a été classée comme une condition de balourd de niveau 1 ou de niveau 2. Ce phénomène s'explique très logiquement par le fait que ce que nous avons considéré comme condition normale n'est pas rigoureusement exempt de déséquilibre.

Le tableau 4.3 illustre de manière condensée les performances obtenues dans la détection des deux classes sans différenciation de niveau.

Classe	Réseau entraîné avec bruit	Réseau entraîné sans bruit
Balourd	94%	93%
Frottement	98 %	98 %
Normale	67%	64%

TAB. 4.3 – Capacité de détection des classes

4.4.2 Signal original non filtré 4 canaux

Une première étude a été menée avec les données non filtrées globales des 4 canaux afin de déterminer une structure satisfaisante du réseau de neurones.

Nous avons abouti au choix d'une structure à 4 couches avec respectivement 100, 50, 50 et 4 noeuds dans les couches. Quant aux fonctions de transfert, l'utilisation de la fonction tangente sigmoïde dans la deuxième couche s'est révélée préférable à une structure avec seulement des fonctions logsig ou à une structure avec la fonction tangente sigmoïde dans la première couche.

Le tableau 4.4 illustre cet effet. Il reprend les performances obtenues avec un vecteur d'entrée construit avec la valeur efficace, le facteur de crête, le kurtosis et le skewness des données non filtrées des quatre points de mesure. Convenons de noter R1 le réseau entraîné avec les données non bruitées et R2 celui entraîné avec bruit.

Cas	Structure du réseau	Performance de R1	Performance de R2
1	100/50/50/4 tansig/logsig/logsig/purelin traingdx	72%	67%
2	100/50/50/4 logsig/logsig/logsig/purelin traingdx	56%	64%
3	100/50/50/4 logsig/ tansig/logsig/purelin traingdx	72%	72%

TAB. 4.4 – Effet de choix des fonctions d'activation

L'apprentissage du réseau de neurone avec l'algorithme "trainscg" plutôt que "traingdx" a semblé améliorer les résultats. Sa rapidité de convergence nous a permis d'imposer une erreur encore plus faible pour les sorties au cours de l'apprentissage. Avec une erreur quadratique moyenne $mse = 0,01$ nous avons obtenu 72% pour R1 et 69% pour R2. En imposant mse à 0,001 les performances sont passées à 72% pour R1 et 75% pour R2 (tableau 4.5).

Niveau d'erreur maximal	R1	R2
Erreur moyenne quadratique $mse=0,01$	72%	69%
Erreur moyenne quadratique $mse=0,001$	72 %	75 %

TAB. 4.5 – Performances en fonction de l'objectif d'erreur

Un vecteur d'indicateurs composé de la valeur efficace, du facteur de crête, du skewness et de la valeur de crête positive a donné les résultats suivants : 69% pour R1 et 56% pour R2.

4.4.3 Signal filtré 1 canal

Pour analyser le comportement du réseau de neurones vis-à-vis des indicateurs issus des signaux filtrés, il a d'abord été considéré le seul point de mesure Z1. Le vecteur d'entrée est constitué de la valeur efficace, le facteur de crête, kurtosis et le skewness calculés pour les signaux filtrés dans les 3 gammes de fréquence (0-30, 30-300, 300-2000 Hz) et pour le signal brut.

Les taux de bonne prédiction atteints dans ce cas sont 50% pour R1 et 47% pour R2. Et en omettant le skewness, on obtient 50% et 44% pour R1 et R2 respectivement.

Pour comparer les performances des signaux filtrés à celles des signaux non filtrés, il fallait aussi pour ce dernier conduire un calcul sur le seul canal Z1. Et dans ce cas les performances n'ont été que de 31% pour R1 et 33% pour R2.

On remarquera que les indicateurs calculés dans des gammes de fréquence spécifiques donnent une prédiction nettement meilleure.

4.4.4 Signal filtré 4 canaux

En considérant à la fois les 4 points de mesure, les indicateurs issus des signaux filtrés auxquels sont joints ceux issus du signal brut donnent des performances qui témoignent de l'importance d'une telle procédure. Le tableau 4.6 dégage aussi l'effet du choix des indicateurs dans ce cas.

Indicateurs	Dimension du vecteur	Performance de R1	Performance de R2
RMS FC KURT SKEW	64	50%	75%
RMS KURT SKEW	48	72%	83%
RMS KURT	32	64%	83%
RMS FC	32	61%	58%

TAB. 4.6 – Résultats obtenus avec des signaux filtrés 4 canaux
 RMS :valeur efficace ; FC :facteur de crête ; KURT :kurtosis ; SKEW :skewness

Ces résultats sont améliorés par rapport d'une part à ceux obtenus en 4 canaux avec les signaux non filtrés et d'autre part à ceux obtenu en 1 seul canal avec filtrage. Il a en outre été remarqué une réduction sensible du temps de calcul et du nombre d'itérations avant convergence.

En investiguant les sensibilités de la prédiction par rapport aux indicateurs, nous avons procédé à l'omission d'un ou de deux indicateurs et observé la tendance qu'affichaient les résultats. Nous en sommes arrivés à la conclusion surprenante que la présence du facteur de crête rendait très imprécise la classification.

Comme dans notre démarche nous travaillons sur des signaux divisés en blocs, il est intuitivement logique d'admettre que les indicateurs fiables doivent présenter

une certaine stabilité de niveau pour toute taille de blocs considérée. En observant la figure 4.2 qui fournit les niveaux des indicateurs en fonctions de la taille de l'échantillon, on peut remarquer que le facteur de crête est le seul indicateur qui ne se stabilise que pour des échantillons de taille très élevée. Donc pour une démarche comme celle adoptée dans cette étude et qui consiste en la division de la série temporelle en plusieurs blocs, le facteur de crête semble ne pas être un bon indicateur. Cette figure 4.2 est tracée sur base du signal en Z1 de la condition H1B0F0R0A0. Les évolutions restent les mêmes pour les autres points de mesure et pour toutes les conditions.

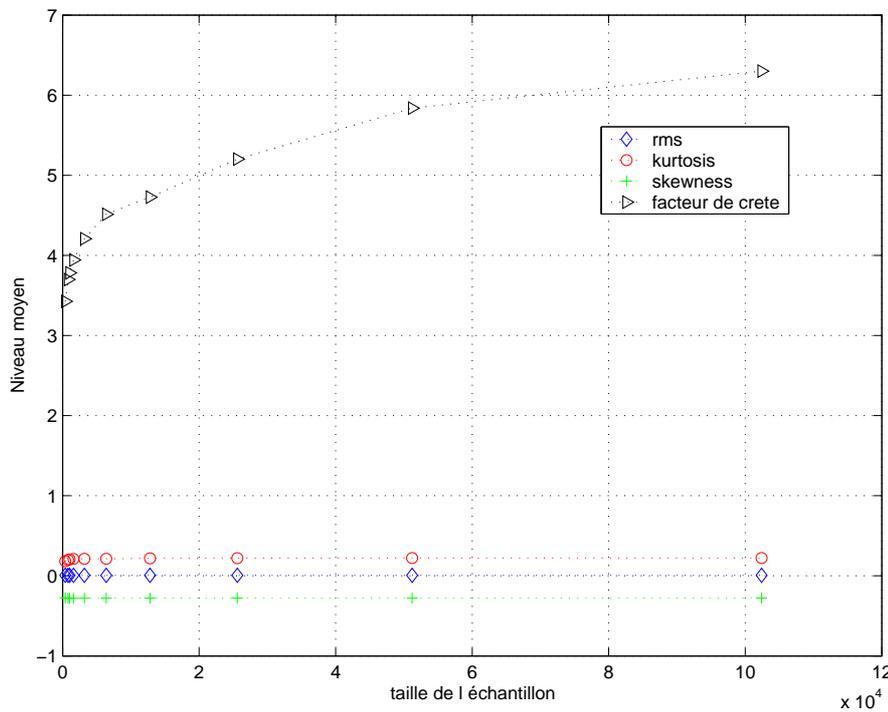


FIG. 4.2 – Niveaux moyens des indicateurs temporel en fonction de la taille de l'échantillon.

4.5 Conclusion

De l'étude faite dans ce chapitre, nous retenons les aspects suivants :

1. le filtrage des signaux dans des gammes de fréquences spécifiques et choisies adéquatement permet de déduire dans le domaine temporel des indicateurs conduisant à une distinction améliorée des classes et niveaux de défauts ;
2. l'exploitation des données provenant de plusieurs points de mesure conduit à des résultats meilleurs par rapport à ceux obtenus à partir des données d'un seul canal ;
3. la structure du réseau de neurones utilisée ici est très complexe, ceci peut s'expliquer par la taille de l'échantillon qui n'est pas très grande. Un réseau très simple a alors difficile à découvrir des règles dans ces données ;
4. il est important de réduire la dimension de l'espace d'indicateurs pour faciliter l'apprentissage ;
5. bien que le facteur de crête et le kurtosis soient tous sensibles à la forme du signal, le kurtosis est un meilleur indicateur puisque la dispersion des résultats obtenus par mesures successives (ou par blocs successifs de signal), est plus faible pour le kurtosis.

Chapitre 5

Réduction dimensionnelle de l'espace des indicateurs

5.1 Rapport de dimensionnalité

Il a été prouvé expérimentalement qu'en augmentant le nombre d'indicateurs les performances de classification peuvent se trouver sensiblement dégradées [43]. Les capacités de généralisation d'un classificateur opérant dans un espace d'indicateurs de grande dimension sont moins bonnes par rapport à celles d'un classificateur opérant dans un espace de faible dimension.

C'est seulement quand le nombre de cas, n , est largement supérieur au nombre d'indicateurs, d , que l'on peut espérer que le classificateur aura une bonne capacité de généralisation. Le rapport n/d est appelé *rapport de dimensionnalité*.

Le choix de ce rapport a été étudié par plusieurs auteurs. Nous pouvons retenir qu'en général un rapport de dimensionnalité au moins égal à 3 est recommandé [40] [41] [42] [43].

Il est nécessaire, pour une meilleure classification, de procéder à une réduction dimensionnelle de l'espace d'indicateurs. Deux voies se présentent pour la réduction du nombre d'indicateurs :

- effectuer un choix de quelques indicateurs sur un ensemble d'indicateurs préalablement sélectionnés. Le principe est de partir d'un espace d'indicateurs de grande dimension et d'en réduire la dimension en ne retenant que

les indicateurs qui présentent des aptitudes de discrimination. Cette procédure n'est pas évidente en pratique, car on ne sait jamais définir a priori les indicateurs pertinents.

- la seconde voie consiste en une projection de l'espace d'indicateurs de manière à en réduire la dimension. La technique la plus utilisée est l'analyse en composantes principales.

5.2 L'analyse en composantes principales (ACP)

L'ACP est une méthode de transformation des données qui consiste à trouver une base orthogonale de l'espace des indicateurs dont les dimensions sont déterminées par les directions selon lesquelles les données changent rapidement. L'ACP permet donc d'abord de passer d'un ensemble \mathbf{X} de vecteurs X_1, X_2, \dots, X_m à un ensemble \mathbf{Y} de vecteurs Y_1, Y_2, \dots, Y_m , tous de dimension p . Les composantes des vecteurs X_i sont les variables d'origine et celles des vecteurs Y_i , les *facteurs* ou *scores factoriels*. Il importe que les nouvelles variables ne présente aucun degré de redondance dans l'information qu'elles véhiculent. Il s'agira ensuite de ne retenir dans \mathbf{Y} que quelques vecteurs qui répondront à un critère informationnel.

L'ACP procède à cette transformation de la façon la plus simple possible, c'est à dire linéairement. Les facteurs sont, en effet, construits comme combinaisons linéaires des variables et réciproquement. Dans ce contexte linéaire, la condition de non redondance de la liste de facteurs s'exprime alors par la condition de non-corrélation des facteurs entre eux ; l'utilisation de la *décomposition spectrale*, appliquée à la matrice des covariances de \mathbf{X} , Σ .

L'idée de base dans une ACP est qu'une grande information correspond à une grande variance. Il faudra transformer \mathbf{X} en $\mathbf{Y} = \mathbf{A}^T \mathbf{X}$, en choisissant \mathbf{A} de telle manière que \mathbf{Y} ait la plus grande variance possible. La matrice \mathbf{A} contient les coefficients des combinaisons transformant les variables de départ en facteurs. On choisit \mathbf{A} comme la matrice orthogonale qui intervient dans la décomposition spectrale de la matrice Σ , c'est donc la matrice des vecteurs propres [26].

On montre que la direction de variance maximum est parallèle au vecteur propre correspondant à la plus grande valeur propre de la matrice Σ . Les vecteurs propres

de Σ peuvent donc servir à la construction de nouveaux axes vus comme une rotation du système original.

La variance liée au premier axe sera la plus grande, elle sera suivie de celle du second axe et ainsi de suite.

En pratique la matrice Σ n'est pas connue, on utilise alors la matrice de covariance S calculée sur l'échantillon.

$$S = \frac{1}{m-1} \sum_{j=1}^m (X_j - \bar{X})(X_j - \bar{X})^T \quad (5.1)$$

où $\bar{X} = \frac{1}{m} \sum_j X_j$

Si l'on note

- les valeurs propres de S par $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$
- les vecteurs propres correspondant à $\lambda_1, \lambda_2, \dots, \lambda_p$ par e_1, e_2, \dots, e_p respectivement

Les axes principaux sont choisis de façon à avoir un système orthonormé, c'est-à-dire

$$e_i e_j^T = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

La transformation est définie comme suit

$$\mathbf{Y} = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \vdots \\ \mathbf{e}_p^T \end{bmatrix} \mathbf{X} = \mathbf{E}^T \mathbf{X} \quad (5.2)$$

où \mathbf{E} est la matrice des vecteurs propres et correspond à la matrice \mathbf{A} décrite ci-haut.

Le vecteur Y_i est tel que ses composantes ne sont pas corrélées entre elles. L'ACP peut être donc vue comme un moyen d'obtenir des indicateurs non corrélés.

Les vecteurs Y_i sont caractérisés par le fait que la plus grande partie du contenu informationnel est stockée dans les quelques premières composantes seulement.

On peut donc ne considérer qu'un nombre relativement réduit d'indicateurs pour la classification et améliorer ainsi les performances du classificateur.

5.2.1 Normalisation

Le processus d'apprentissage des réseaux de neurones peut être plus efficace si certains prétraitements sont effectués sur les entrées et sur les cibles. La normalisation est un traitement qui permet d'éviter que les indicateurs qui ont de grandes valeurs n'aient d'influence plus grande les autres. La normalisation s'impose encore d'avantage si les indicateurs n'ont pas tous la même unité.

Un autre avantage de la normalisation, c'est l'allègement de la tâche d'apprentissage qui se traduit par une réduction sensible du temps d'apprentissage. Deux manières de normaliser les données sont possibles :

1. La normalisation proprement dite :

Les entrées et les cibles sont toutes remises dans la plage $[-1, 1]$. Le vecteur normalisé x_n est obtenu du vecteur original x par

$$x_n = 2 \frac{[x - \min(x)]}{[\max(x) - \min(x)]} - 1 \quad (5.3)$$

$\min(x)$ et $\max(x)$ sont respectivement la plus petite et la plus grande composante de x

2. La standardisation :

Cette approche normalise la moyenne et l'écart-type de l'ensemble d'apprentissage de telle manière qu'après le traitement la moyenne soit nulle et l'écart-type égale à 1. Le vecteur standardisé x_n est obtenu du vecteur original x par

$$x_n = \frac{[x - \text{moy}(x)]}{\text{std}(x)} \quad (5.4)$$

où $\text{moy}(x)$ est la moyenne de x et $\text{std}(x)$, son écart-type.

Dans une analyse en composantes principales la standardisation est un préalable.

5.2.2 Saturations, communalités et variance expliquée

La relation entre les variables brutes ou standardisées et les facteurs est explicitée dans une matrice dite *de structure* ou *des saturations* (factor matrix) de composantes

$L_{j\alpha}$ (factor loadings). Les composantes de \mathbf{L} sont les corrélations entre les variables brutes ou standardisées et les facteurs.

La matrice des saturations \mathbf{L} a les propriétés suivantes :

1. $h_j^2 = \sum_{\alpha=1}^p L_{j\alpha}^2 = 1$ ($j = 1, \dots, p$)

La somme des carrés de toutes les saturations pour une variable donnée vaut 1. Cette quantité est appelée *communalités* ou *communauté*. Comme le plus souvent, les objectifs de l'ACP sont la condensation de l'information, on ne conservera pas tous les p facteurs, mais seulement les k premiers ($k < p$), ceux dont la variance λ_α est la plus grande. De cette façon les communalités sont alors inférieures à l'unité. De fait, h_j^2 mesure la proportion de variance de la variable j expliquée par les k premiers facteurs, et constitue une mesure locale de la qualité d'une ACP, variable par variable ;

2. $\sum_{j=1}^p L_{j\alpha}^2 = \lambda_\alpha$ ($\alpha = 1, \dots, p$)

La somme des carrés de toutes les saturations pour un facteur donné vaut λ_α , sa variance.

5.2.3 Choix de la dimension effective

Le choix du nombre de facteurs à retenir est un compromis à adopter entre les objectifs opposés de "conservation de l'information" et "de simplification de l'information". Il existe plusieurs façons de déterminer le nombre de composantes requis pour garder une représentation adéquate des données. Nous en citerons quelques unes :

1. proportion expliquée de la variance par les k premiers facteurs : cette méthode propose de sommer les variances des k premières composantes principales et de diviser cette somme par la somme de toutes les variances. En général, les composantes qui donnent une proportion d'au moins 75% sont suffisantes pour représenter les données :

$$\frac{\sum_{\alpha=1}^k \lambda_\alpha}{\sum_{\alpha=1}^p \lambda_\alpha} \geq 0,75 \quad (5.5)$$

2. proportion expliquée de la variance par le facteur α : on sélectionne les composantes principales qui expliquent un certain pourcentage de la variance totale, par exemple 2% au minimum. C'est le critère implémenté dans la fonction "prepca" du "Neural Network Toolbox" de Matlab que nous avons utilisé dans nos travaux :

$$frac_var(\alpha) = \frac{\lambda_\alpha}{\sum_{\alpha=1}^p \lambda_\alpha} \quad (5.6)$$

Pout illustrer cette méthode, la figure 5.1 les résultats obtenus avec nos données traitées dans le domaine temporel.

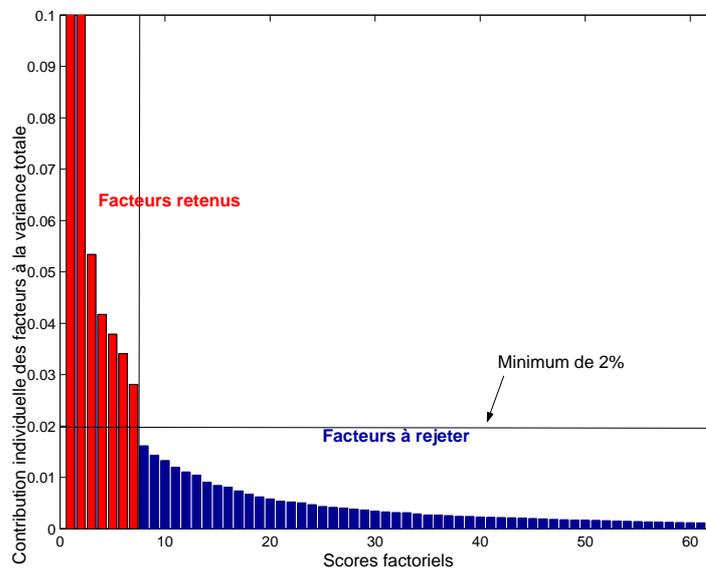


FIG. 5.1 – Proportion de la variance totale individuellement expliquée par chacun des facteurs.

3. moyenne des variances : la direction α sera considérée sans importance si $\lambda_\alpha < \bar{\lambda}$, où $\bar{\lambda} = \frac{1}{p} \sum_{\alpha=1}^p \lambda_\alpha$ est la moyenne des variances ou simplement ce que serait la variance si toutes les directions étaient également importantes. C'est le critère de *Guttman-Kaiser*.

4. scree Test : on utilise, ici, le graphique des λ_α en fonction de α . Là où la courbe devient plate, c'est là que la variation devient aléatoire, donc sans structure. on ne gardera alors que les valeurs propres apparaissant avant le coude (critère de *Cattell*)

5.3 Les rotations en analyse en composantes principales

Les facteurs sont des abstractions mathématiques et ne possèdent aucune signification intuitive. Leur interprétation n'est pas souvent aisée, mais elle peut être sérieusement facilitée si les saturations sont soit grandes (c'est à dire proche de 1), soit proches de zéro. Dans ce cas on saurait à quelle(s) variable(s) correspondrait principalement un facteur.

Pour ce faire, on peut procéder à une nouvelle transformation par rotation des facteurs pour avoir chacune des variables associée à un seul facteur. Plusieurs algorithmes ont été développés pour l'obtention de la matrice de rotation, les plus utilisés portent les noms de *varimax*, *quartimax*, ou *equamax*. Nous avons utilisé dans ce travail la méthode de rotation *varimax* proposée par *Kaiser* en 1985 (Annexe D).

L'ACP, par ses différentes qualités, s'impose donc dans le prétraitement des données destinées à l'alimentation d'un réseau de neurones. Elle permet aussi l'analyse des contributions des indicateurs sur les performances de classification.

Chapitre 6

Seconde phase expérimentale. Simulation de tous les défauts.

6.1 Introduction

Dans le chapitre 4 nous avons étudié le cas de deux défauts, le balourd et le frottement. Cette étude nous a permis d'établir partiellement les effets des indicateurs, particulièrement le rôle perturbateur du facteur de crête. Nous en avons aussi conclu le besoin d'alléger la complexité du réseau de neurones en terme de nombre de couches et de neurones. L'utilisation de la fonction linéaire en sortie du réseau devrait aussi être repensée pour une utilisation de la fonction logsig qui est la mieux adaptée à des problèmes de classification.

Nous avons alors, dans le domaine temporel, construit un espace d'indicateurs sans le facteur de crête. Seuls la valeur efficace, le kurtosis et le skewness de l'accélération ont été utilisés. En ce qui concerne la sortie du neurone, les cibles devraient être construites de manière à permettre l'utilisation de la fonction logsig, c'est ainsi que nous avons disposé une sortie pour chaque niveau de défaut.

Dans le présent chapitre, nous allons considérer tous les 4 défauts prévus pour la campagne d'essais. Nous n'avons pas procédé à la simulation de toutes les combinaisons possibles, mais des cas représentatifs de tous les défauts ont été étudiés. Nous avons ainsi simulé 162 conditions au lieu de 540 suggérés par le décompte de

toutes les combinaisons possibles. Ces 162 signaux ont été subdivisés chacun, comme dans le chapitre 4, en quatre blocs, ce qui donne 648 signaux pour 162 conditions.

Trois quarts de ces données ont servi à l'apprentissage et un quart au test.

6.2 Architecture du réseau

L'architecture du réseau de neurones a été élaborée progressivement en observant ses performances. Nous avons retenu un réseau de neurones à rétropropagation avec une seule couche cachée contenant 80 neurones. Pour bénéficier complètement de la puissance de la fonction *logsig*, nous n'avons plus constitué une sortie par défaut, mais une sortie par niveau de défaut. De cette manière les sorties sont réglées à 1 ou 0 selon qu'il y a ou non existence de ce niveau de défauts. Mais il peut arriver qu'en réponse, le réseau de neurones reconnaisse sur une même condition à la fois deux niveaux différents d'un même défaut. Par exemple, le réseau entraîné sur les indicateurs temporels bruités donne pour la condition B3F0R2A2 la sortie B3B4F0F1R2A2. Pour palier cet inconvénient et par conséquent améliorer les performances du réseau de neurones, nous avons opté pour une compétition à la sortie de chaque type de défaut. Les sorties obtenues, au lieu de subir une opération d'arrondissement, sont plutôt passées en argument à une fonction qui retourne une valeur de 1 là où la sortie a sa valeur maximale et zéro ailleurs. Cette opération est conduite par lot sur des sorties correspondant à un seul défaut.

L'architecture utilisée pour le réseau de neurones est représentée sur figure 6.1. On y remarque une seule couche cachée à 80 neurones et une couche de sortie à 14 neurones correspondant chacun aux 14 sorties différentes pour les 4 types de défauts.

6.3 Les indicateurs du domaine temporel

6.3.1 Construction

Comme le fait ressortir la figure 3.3, des signaux filtrés dans les gammes de fréquence différentes sont extraits les indicateurs qui sont la valeur efficace, le kurtosis

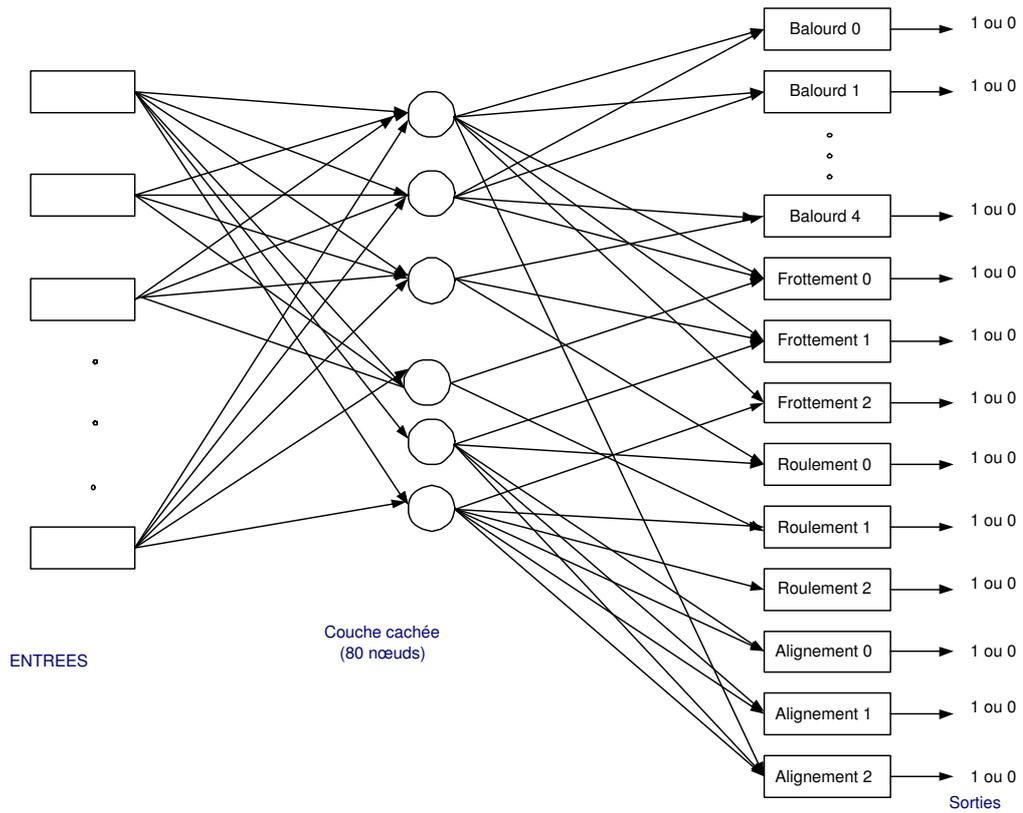


FIG. 6.1 – Architecture du réseau de neurones

et le skewness . Ces indicateurs considérés pour tous les points de mesure servent à construire le vecteur d'entrée de dimension 48. Pour réduire la dimension du vecteur d'entrée une analyse en composantes principales est effectuée de manière à produire des données non corrélées entre elles, ensuite il n'est retenu que les directions principales qui contribuent le plus à la variation dans l'ensemble des données. Généralement l'analyse en composantes principales exige une standardisation préalable des données (entrées et cibles). Mais, dans ce cas, comme les cibles sont binaires (0 ou 1), il ne s'est avéré utile de standardiser que les entrées avant l'analyse en composantes principales.

Le choix de la dimension finale à retenir a été déterminé par le principe de la contribution individuelle des valeurs propres à la variation totale des données, c'est ce que nous avons appelé proportion expliquée de la variance par le facteur α . Ici la contribution fixée à 2% a assuré aussi le respect de certains critères plus élaborés tel que celui de la proportion expliquée de la variance (80% ici) ou le critère de Kaiser. Ainsi, le vecteur d'entrée se trouve-t-il réduit à la dimension 13.

6.3.2 Apprentissage et Analyse des performances

Données non bruitées

Conduit avec des données non bruitées, l'apprentissage du réseau de neurones s'arrête à 635 époques par atteinte du gradient minimum de 10^{-6} . L'erreur d'apprentissage est de 0,001765. La figure 6.2 donne l'évolution de l'erreur d'apprentissage en fonction des époques.

145 des 162 conditions de fonctionnement prévues pour tester les performances du réseau de neurones sont bien classifiées. Le taux de réussite s'élève à 89,5%.

On peut analyser les performances du réseau de neurones en examinant sortie par sortie comment les cibles sont suivies par les réponses du réseau. Pour ce faire, une régression linéaire est effectuée sur les deux séries de valeurs. En effet, si A est le vecteur des réponses obtenues sur une sortie et T le vecteur des valeurs attendues sur cette même sortie, la régression linéaire consiste à trouver la meilleure droite qui représente le nuage des points (A, T). L'écart de cette droite par rapport à la droite

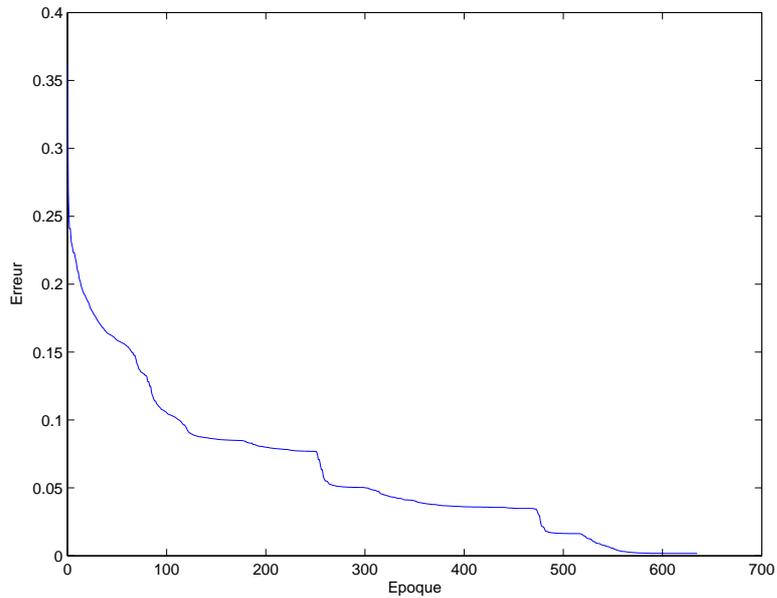


FIG. 6.2 – Erreur d'apprentissage

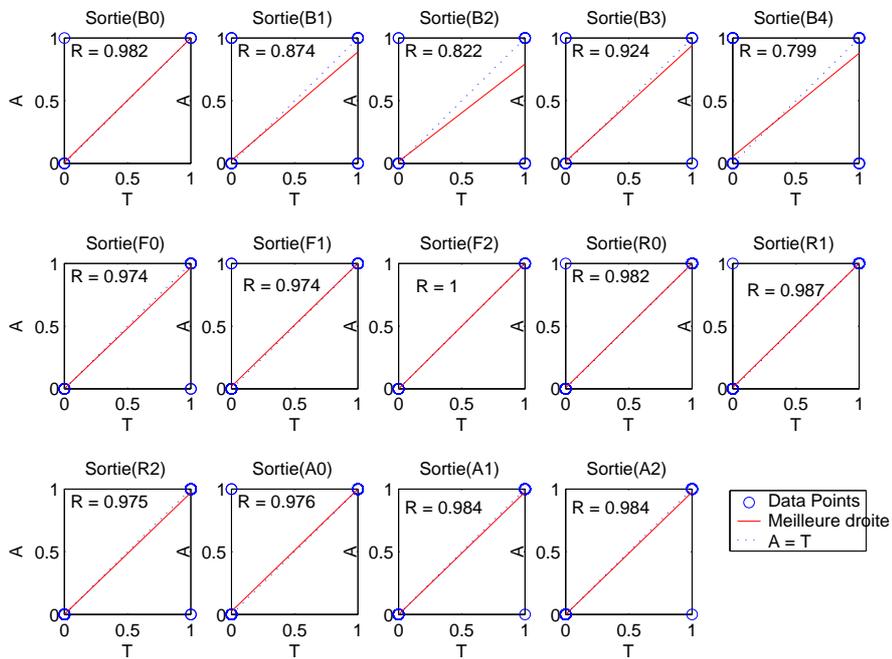


FIG. 6.3 – Corrélations entre entrées et sorties. Indicateurs temporels non bruités

de pente unité ($A=T$) constitue une mesure de la qualité de reconnaissance du défaut associé à la sortie en question. Cet écart est exprimé par le coefficient de corrélation entre les deux séries.

On remarque, dans ce cas, un très bon suivi pour les autres sorties que le balourd. En effet, pour le balourd on a une corrélation moyenne de 0,88 entre les sorties et les cibles tandis que pour les autres conditions les corrélations sont quasi unitaires. Nous remarquerons que dans cet espace d'indicateurs c'est le niveau de balourd B4 qui est le moins bien catégorisé, la figure 6.3 illustre bien ce constat.

Sur cette figure, les points situés sur la première diagonale sont les points correctement représentés :

- valeur attendu= 0 et valeur obtenu= 0 ou
- valeur attendu= 1 et valeur obtenu= 1 ;

tandis que ceux situés sur la seconde diagonale correspondent aux échecs de reconnaissance :

- valeur attendu= 0 et valeur obtenu= 1 ou
- valeur attendu= 1 et valeur obtenu= 0.

Données bruitées

Le réseau de neurones déjà entraîné avec des données non bruitées subit ensuite un entraînement avec 10 séries différentes de données bruitées. Pour s'assurer que les vecteurs non bruités restent toujours bien catégorisés, cet apprentissage est clôturé par un passage sur les données non bruitées.

Une légère dépréciation du taux de bonne classification est observée par rapport au réseau sans bruit : on passe de 89,5% à 88,3%.

Contrairement au cas précédent, le niveau B4 est relativement bien classifié et présente un coefficient de corrélation de 0,810 contre 0,799 avant, mais il reste la sortie qui présente une faible performance (figure 6.4).

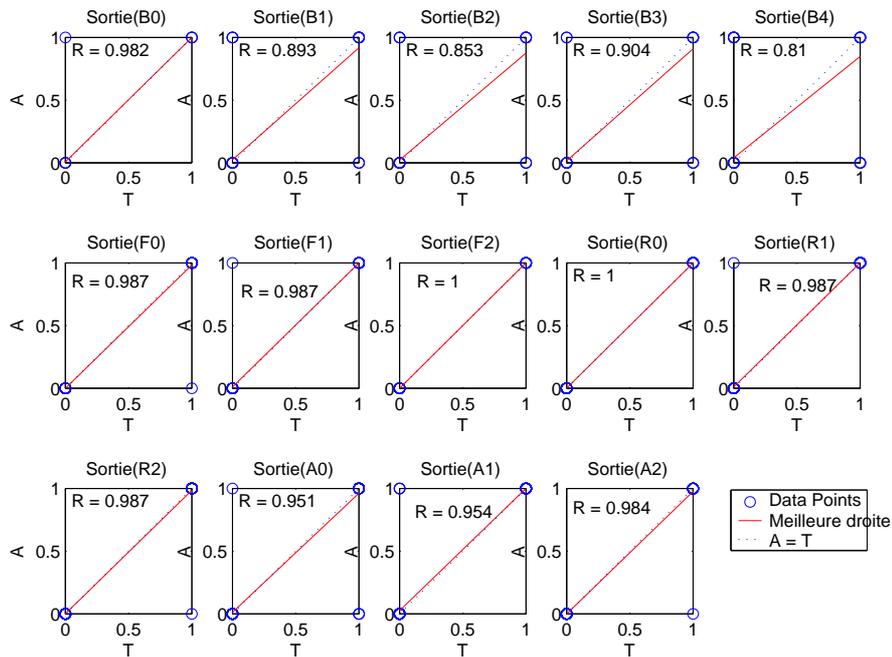


FIG. 6.4 – Corrélations entre entrées et sorties. Indicateurs temporels bruités

6.4 Les indicateurs du domaine fréquentiel

6.4.1 Construction

Des spectres fréquentiels de 512 lignes sont calculés pour chaque signal. Le vecteur d'entrée pour une condition donnée est construit avec les valeurs spectrales issues des 4 points de mesure. Après réduction dimensionnelle sur le même principe de l'analyse en composantes principales, on passe d'un vecteur de 2048 composants à un vecteur de 7 composants. Ceci prouve une forte corrélation entre les spectres fréquentiels des signaux vibratoires mesurés en ces 4 points Z1, Y1, Z2 et Y2.

6.4.2 Apprentissage et Analyse des performances

Données non bruitées

L'apprentissage conduit sur un réseau de neurones avec des données non bruitées a convergé en 424 époques vers l'objectif d'erreur fixé à 0,001. La figure 6.5

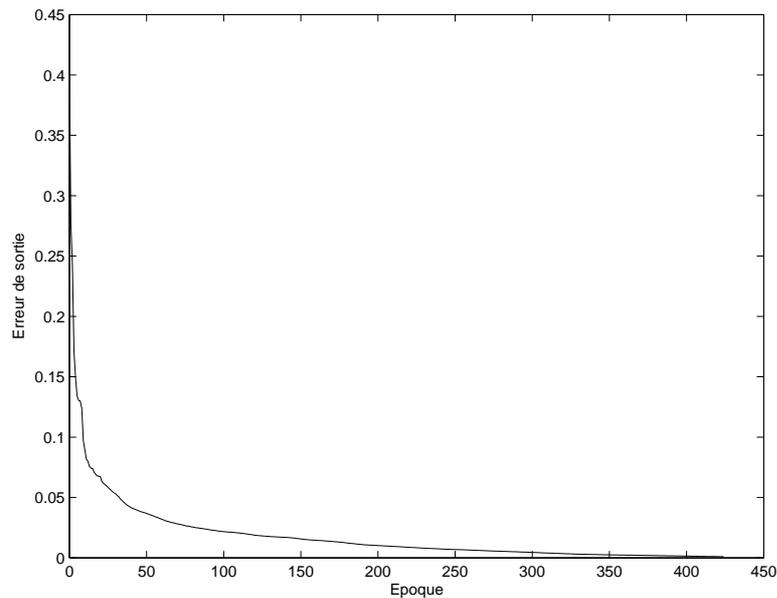


FIG. 6.5 – Erreur d'apprentissage

montre l'évolution de l'erreur d'apprentissage aux cours des époques.

Le taux de réussite dans la classification pour ce cas est de 96,91%. L'analyse de la réponse du réseau de neurones aux données de test peut être réalisée en faisant une régression linéaire entre les sorties obtenues et les sorties désirées (cibles). On peut remarquer que toutes les sorties, exceptées les sorties B1 et B2 et B3, présentent une corrélation proche de 1. Néanmoins, les balourds B1 et B2 ne peuvent pour autant pas être considérés comme mal reconnus et mal classifiés parce que les corrélations y correspondant sont de 0,946 et 0,88 respectivement, donc assez élevées. La figure 6.6 illustre comment les sorties suivent les cibles.

Données bruitées

L'apprentissage après bruitage sur les indicateurs du domaine fréquentiel donne une performance de classification de 97,53%. Comparé au cas des indicateurs non bruités, une légère dépréciation est notée sur la sortie B0 tandis que les sorties B2 et B4 sont améliorées. (figure 6.7)

6.4 Les indicateurs du domaine fréquentiel

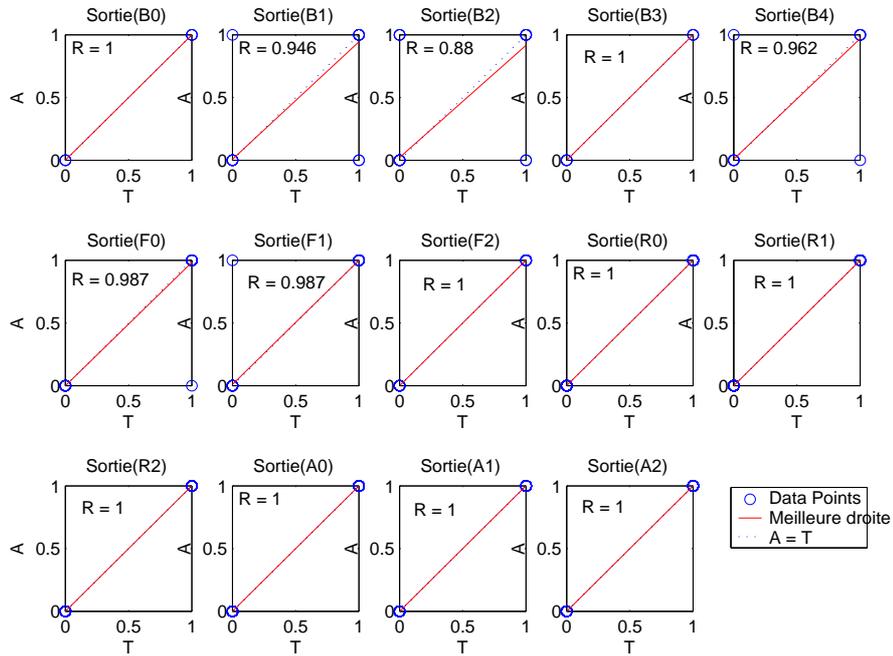


FIG. 6.6 – Corrélations entre entrées et sorties. Indicateurs fréquentiels non bruités

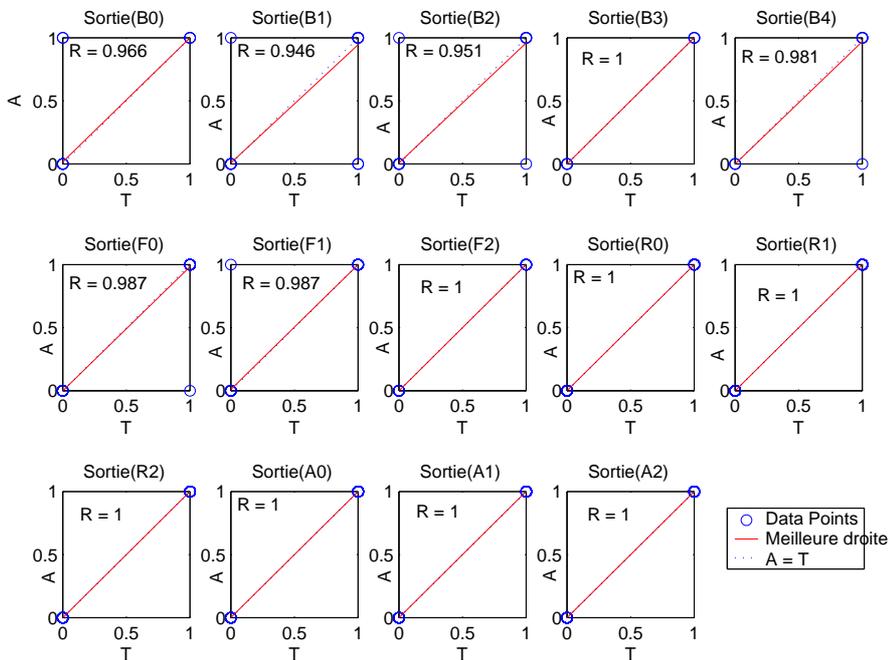


FIG. 6.7 – Corrélations entre entrées et sorties. Indicateurs fréquentiels bruités

6.5 Les indicateurs mixtes

6.5.1 Construction

Pour construire un vecteur d'entrée qui tiendrait compte des caractéristiques extraites et du temporel et du fréquentiel, nous concaténons les deux vecteurs d'indicateurs temporels et fréquentsiels pour en faire un seul de dimension 2096. La déconsidération, après analyse en composantes principales, des directions principales qui ne recèlent pas une grande quantité d'information (c'est à dire qui contribuent pour moins de 2% à la variation) nous amène à un vecteur d'entrée de dimension 7 .

6.5.2 Apprentissage et Analyse des performances

Données non bruitées

L'apprentissage conduit sur un réseau de neurones avec des données non bruitées a convergé en 487 époques vers l'objectif d'erreur fixé. La figure 6.8 montre l'évolution de l'erreur d'apprentissage en fonction des époques.

Le taux global de bonne classification est de 95,06%. De la régression linéaire faite sur les sorties et les cibles, on remarquera ce qui suit : sur les sorties B0, B1, B2 et B4 on observe quelques erreurs. B1 et B2 sont encore les seuls à présenter les corrélations les plus faibles comme dans le cas des indicateurs fréquentsiels.

Données bruitées

Avec le bruitage on obtient le taux de classification de 95,68% mais avec une légère dépréciation de la corrélation sur les sorties de balourd F1 et F2, les balourds sont mieux classés par rapport au cas précédent.(figure 6.10)

6.6 Synthèse et discussion

L'examen de ces résultats montre que, dans ce cas, le bruitage n'apporte pas un changement sensible à la performance du réseau de neurones. Ce constat est une

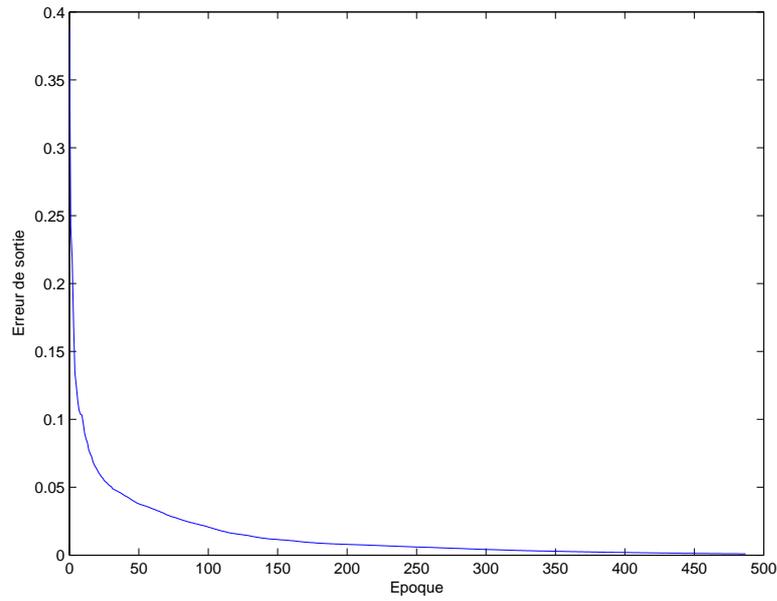


FIG. 6.8 – Erreur d'apprentissage

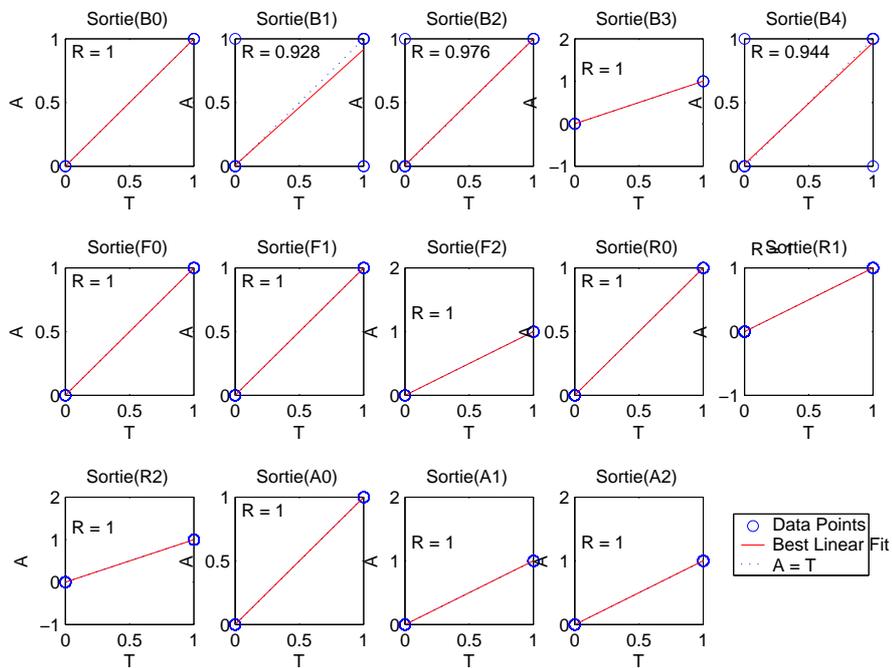


FIG. 6.9 – Corrélations entre entrées et sorties. Indicateurs mixtes non bruités

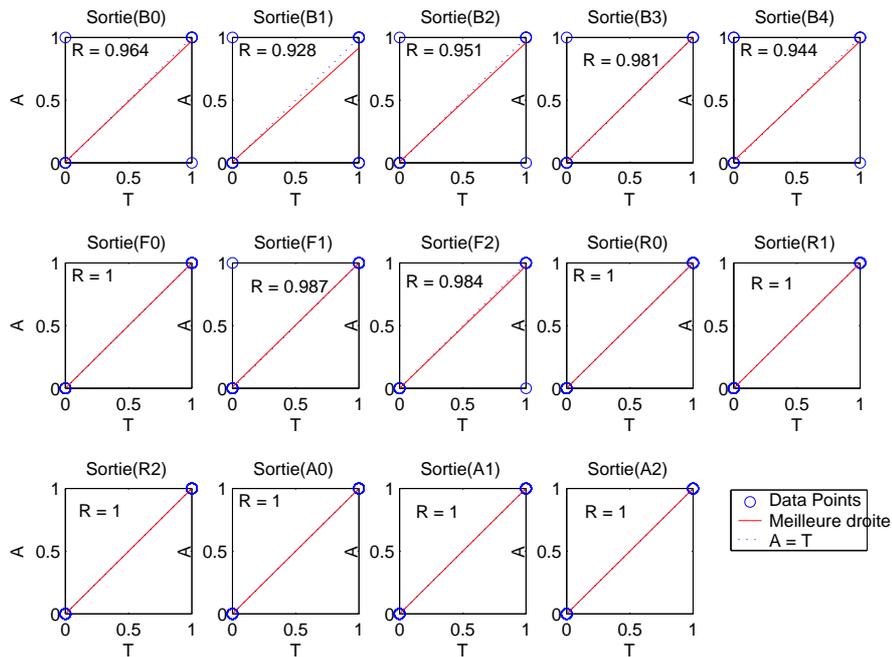


FIG. 6.10 – Corrélations entre entrées et sorties. Indicateurs mixtes bruités

preuve d’une bonne cohérence des données. En effet dans le cas où les données seraient bruitées ou tronquées, le réseau bruit afficherait de meilleurs résultats.

Les indicateurs issus du domaine fréquentiel assurent une classification meilleure que ceux issus du domaine temporel. Les vecteurs d’indicateurs mixtes semblent avoir hérité des capacités des deux types d’indicateurs : les temporels et les fréquentiels ; ils donnent des résultats intermédiaires mais sont plus performants par rapport aux temporels. Les vecteurs des indicateurs fréquentiels et mixtes, après réduction dimensionnelle, présentent une dimension de 7 alors que celui des indicateurs temporels a une dimension de 13. Ceci laisse supposer une corrélation entre certains indicateurs temporels et les indicateurs fréquentiels.

Le tableau 6.1 synthétise les résultats obtenus dans les différents espaces d’indicateurs avec des données bruitées ou non.

Comparés aux résultats de la première phase expérimentale, les résultats obtenus ici sont très améliorés. L’organisation d’une sortie par niveau pour chaque type de défaut et l’utilisation des fonctions sigmoïdes que cela entraîne, explique en grande partie cette amélioration. Il faut aussi souligner la taille de l’ensemble des

Indicateurs	Pk(temporel)		Pxx(fréquentiel)		Pmixt(mixte)	
	Normales	Bruitées	Normales	Bruitées	Normales	Bruitées
B0	0,982	0,982	1	0,966	0,947	0,964
B1	0,874	0,893	0,946	0,946	0,872	0,928
B2	0,822	0,853	0,88	0,951	0,906	0,951
B3	0,924	0,903	1	1	1	0,981
B4	0,779	0,810	0,962	0,981	0,962	0,944
F0	0,974	0,987	0,987	0,987	1	1
F1	0,974	0,987	0,987	0,987	1	0,987
F2	1	1	1	1	1	0,984
R0	0,982	1	1	1	1	1
R1	0,987	0,987	1	1	1	1
R2	0,975	0,987	1	1	1	1
A0	0,976	0,951	1	1	1	1
A1	0,984	0,954	1	1	1	1
A2	0,984	0,984	1	1	1	1
Taux(%)	89,51	88,28	96,91	97,53	95,06	95,68

TAB. 6.1 – Taux de classification et Corrélation entre sorties désirées et sorties obtenues.

données et l'architecture du réseau de neurones parmi les facteurs qui ont influencé les performances de la classification.

Chapitre 7

Analyse de l'influence des indicateurs.

Comme les réseaux de neurones construits présentent des capacités de classification satisfaisantes, nous allons analyser l'espace des indicateurs pour déterminer les variables les plus prépondérantes, donc les indicateurs les plus pertinents pour la reconnaissance des défauts.

7.1 Démarche

La démarche que nous nous sommes proposée de suivre est basée sur l'analyse en composantes principales. Elle consiste à rechercher les contributions des variables initiales aux scores factoriels (variables transformées). Pour ce faire, nous analyserons la manière dont les scores factoriels utilisés comme entrée des réseaux de neurones sont corrélés avec les variables brutes. En effet, les scores factoriels qui sont des combinaisons linéaires des variables initiales normalisées ne laissent pas transparaître à vue d'oeil les contributions des variables initiales ; seul l'examen des *saturations* peut permettre l'interprétation des scores factoriels. Les *saturations* sont les corrélations entre les variables d'origines normalisées et les scores factoriels.

Une rotation judicieusement choisie des axes principaux permet d'obtenir des saturations proches de 1, -1 ou 0, ce qui facilite l'interprétation des facteurs obtenus.

7.2 Analyse des indicateurs temporels

7.2.1 Matrice des saturations

Une première analyse sur la matrice des saturations des indicateurs du domaine temporel conduit à relever que les facteurs 1, 2, 3, 4 et 5 présentent les plus fortes corrélations avec les variables de départ tandis que les autres facteurs sont les moins corrélés avec les variables initiales. Les graphiques de la figure 7.1 représentent la corrélation de chaque facteur avec les variables d'origine.

Un examen poussé des graphiques de la figure 7.1 montre que seul le facteur 2 est le plus corrélé avec des variables initiales suivi du facteur 1.

En analysant, par exemple, le facteur 2 (figure 7.2), on peut remarquer que les variables 5, 17, 29, 41 d'une part et 6, 18, 30, et 42 d'autre part sont les plus significatifs et s'opposent sur l'axe de ce facteur. La première série de variables qui présente des saturations négatives correspond aux kurtosis de la gamme 0-30 Hz pour les 4 points de mesure tandis que la deuxième série qui présente des saturations positives correspond aux skewness dans la même gamme fréquentielle pour les 4 points de mesure.

L'analyse du facteur 1 révèle, quant elle, une forte contribution des valeurs efficaces des hautes fréquences et du signal global. La contribution du point de mesure Y2 est négligeable dans ce cas par rapport aux autres. Les saturations liées à ces valeurs efficaces sont toutes négatives.

7.2.2 Rotation

Nous procédons ici à l'extraction des informations non plus à partir de la matrice des saturations initiale mais à partir de celle obtenue après rotation par le critère *varimax*. La rotation améliore la représentation des variables dans l'hypersphère des corrélations. Les variables les mieux représentées auront une communalité proche de l'unité. En observant les graphiques de la figure 7.4, nous remarquons que par rapport à la représentation avant rotation d'autres variables viennent se révéler avec

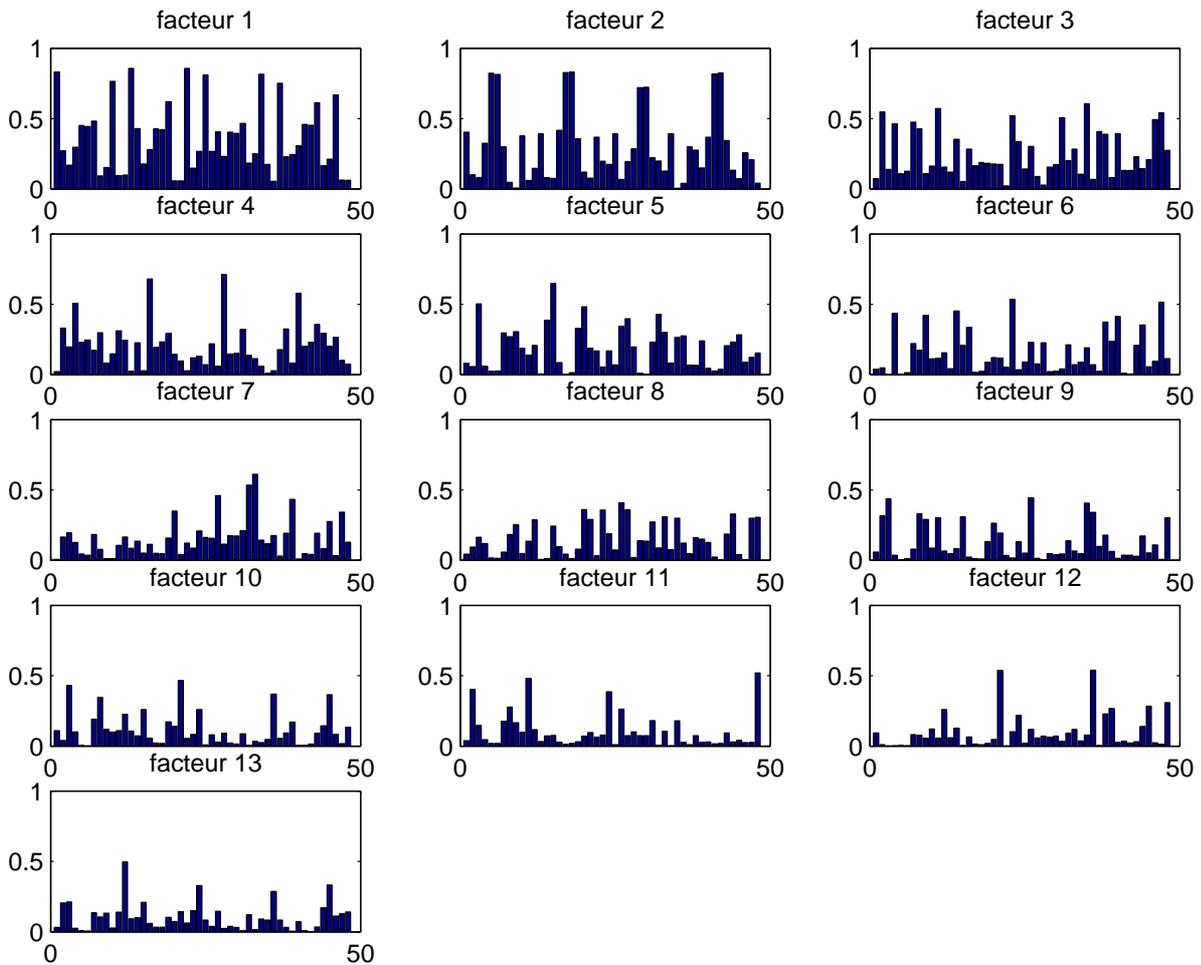


FIG. 7.1 – Valeurs absolues des saturations. En abscisse sont repris les indicateurs temporels et en ordonnées les valeurs absolues des corrélations de ces indicateurs avec le facteur considéré

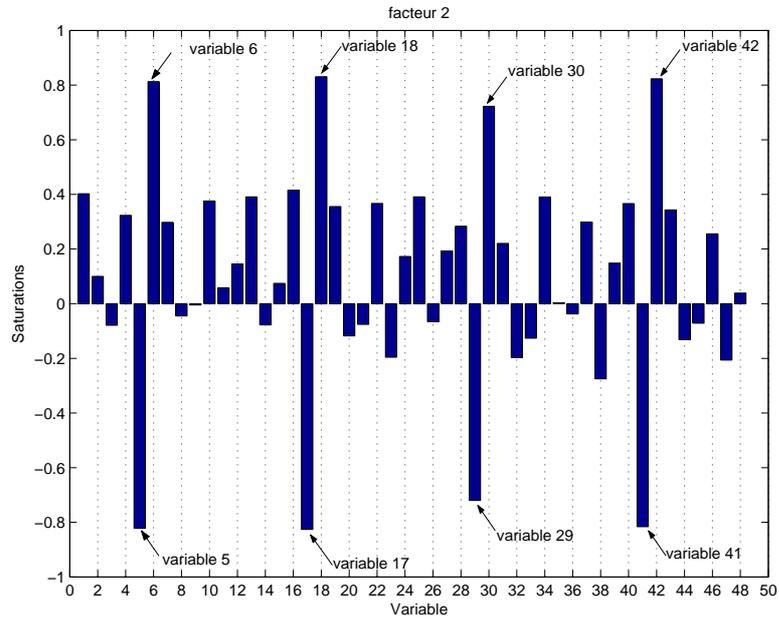


FIG. 7.2 – Saturations pour le facteur2

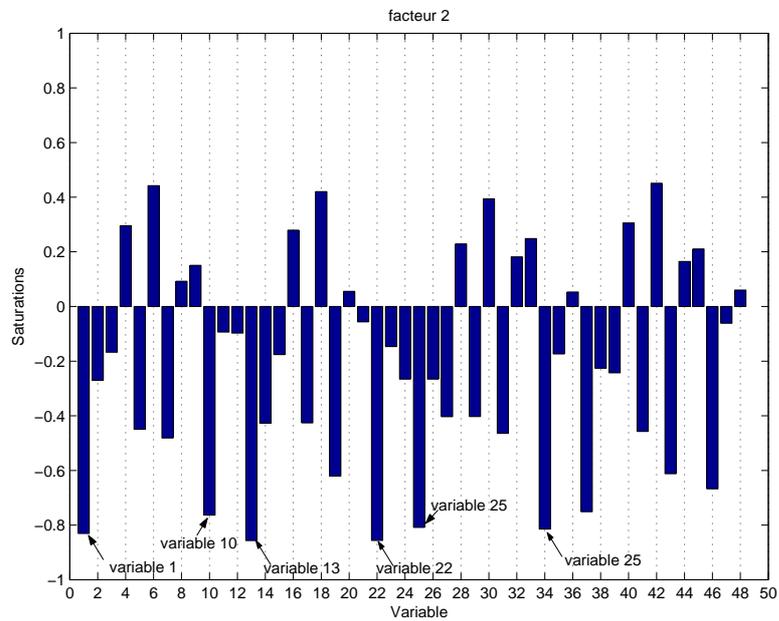


FIG. 7.3 – Saturations pour le facteur1

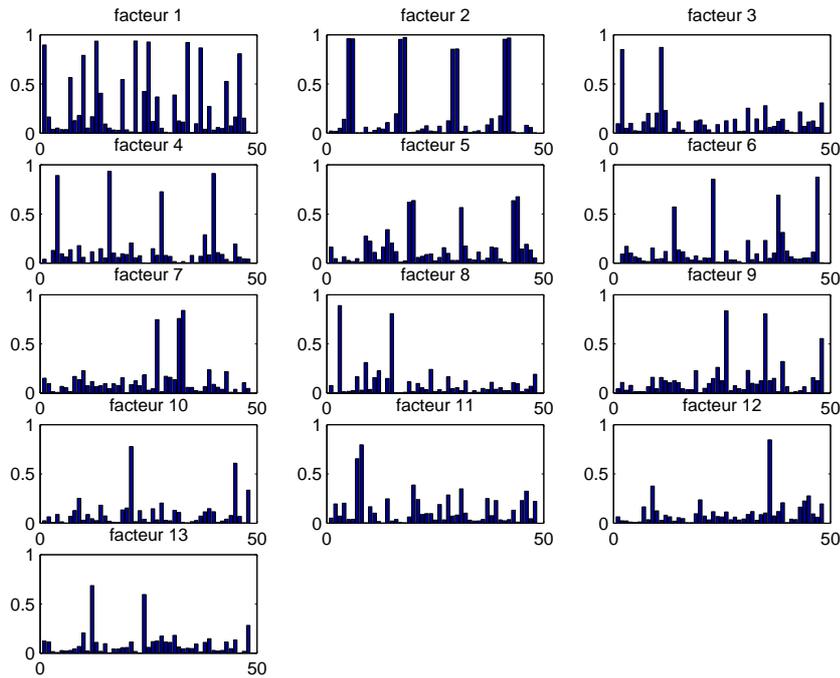


FIG. 7.4 – Saturations en valeurs absolues après rotation

Facteur	Variables	Indicateurs
1	1,10,13,22,25,34,37,46	RMS global et RMS haute fréquence(300-2000 Hz)
2	5,6,17,18,29,30,41,42	Kurtosis et Skewness entre 0-30 Hz
3	2,11	Kurtosis global et haute fréquence suivant Z1
4	4,16,40	RMS 0-30 Hz
5		Pas de bonne corrélation
6	23, 47	Kurtosis Haute fréquence Y1 et Y2
7	33	Skewness 30-300 Hz sur Z2
8	3,15	Skewness global Z1 et Y1
9	26,35	Kurtosis global et haute fréquence Z2
10	21	Skewness 30-300 Hz sur Y1
11	8	Kurtosis 30-300 Hz Z1
12	36	Skewness haute fréquence Z2
13	12,24	Skewness haute fréquence Z1 et Y1

TAB. 7.1 – Variables à fortes saturations après rotation

de bonnes saturations. Le tableau 7.1 représente les différentes variables fortement corrélées avec les scores factoriels.

Si nous adoptons comme critère de choix des variables pertinentes une saturation de plus de 90% en valeur absolue, nous ne retiendrons que les variables 5, 6, 17, 18, 41, 42 toutes expliquées par le facteur 2, les variables 13, 22, 25, 34 expliquées par le facteur 1 et les variables 16 et 40 expliquées par le facteur 4. Les mêmes conclusions peuvent être tirées en examinant les communalités des variables de la figure 7.5



FIG. 7.5 – Les communalités

7.2.3 Matrice de transformation

L'analyse en composantes principales est une méthode linéaire car les scores factoriels sont des combinaisons linéaires des variables initiales. Une autre manière d'examiner la contribution des variables initiales sur les axes à fort contenu informationnel consiste à analyser les coefficients de cette combinaison linéaire. Ces coefficients sont contenus dans une matrice dite matrice de transformation. C'est cette matrice qui effectue la projection de l'espace initial d'indicateurs sur le nouvel espace.

En regardant les éléments de la matrice de transformation facteur par facteur, on peut déterminer les indicateurs qui contribuent pour le plus à la reconnaissance et à classification des défauts. Les graphiques de la figure 7.6 font ressortir des scores factoriels les indicateurs ci-dessous repris dans le tableau 7.2. Les indicateurs retrouvés dans chacun des facteurs par l'examen de la matrice de transformation correspondent bien à ceux déterminés à l'aide des saturations. On remarquera que plus le contenu informationnel est grand dans le facteur, plus la cohérence entre les variables qu'il contient est grande.

Il importe, pour déterminer les indicateurs pertinents pour la détection et la classification des défauts mécaniques, de pouvoir être en mesure de choisir de prime abord les facteurs à étudier. La cohérence dans les variables contenues par le facteur représenterait un premier critère dans le cas de notre étude. Par cohérence, nous entendons le fait que les variables présents dans un facteur permettent leur regroupement en type bien déterminé. Dans la suite, nous allons tenir compte, pour le choix des facteurs à examiner, des poids synaptiques qui leur sont assignés lors de l'apprentissage du réseau de neurones.

7.2.4 Examens des poids synaptiques

L'analyse des saturations, communalités et des éléments de la matrice de transformation permet de déceler variables les mieux représentées par la projection. Dans l'analyse qui précède, nous avons considéré, comme critère pour déterminer les indicateurs pertinents, le degré de représentation de la variable au travers des scores factoriels. Nous ne devons cependant pas négliger le fait que lors de l'apprentissage, le réseau de neurones attribue aux différents facteurs, à l'entrée, des poids qui correspondent dans une certaine mesure à leurs contributions à la réponse du réseau de neurone.

Il nous semble donc logique de proposer la construction d'une mesure d'importance des scores factoriels qui tiendrait compte d'une part du fait que leur importance informationnelle est représentée par leur valeur propre, et d'autre part, du fait qu'en tant qu'entrée du réseau de neurones, ces facteurs ont une contribution qui dépend des poids qui leur sont attribués lors de l'apprentissage.

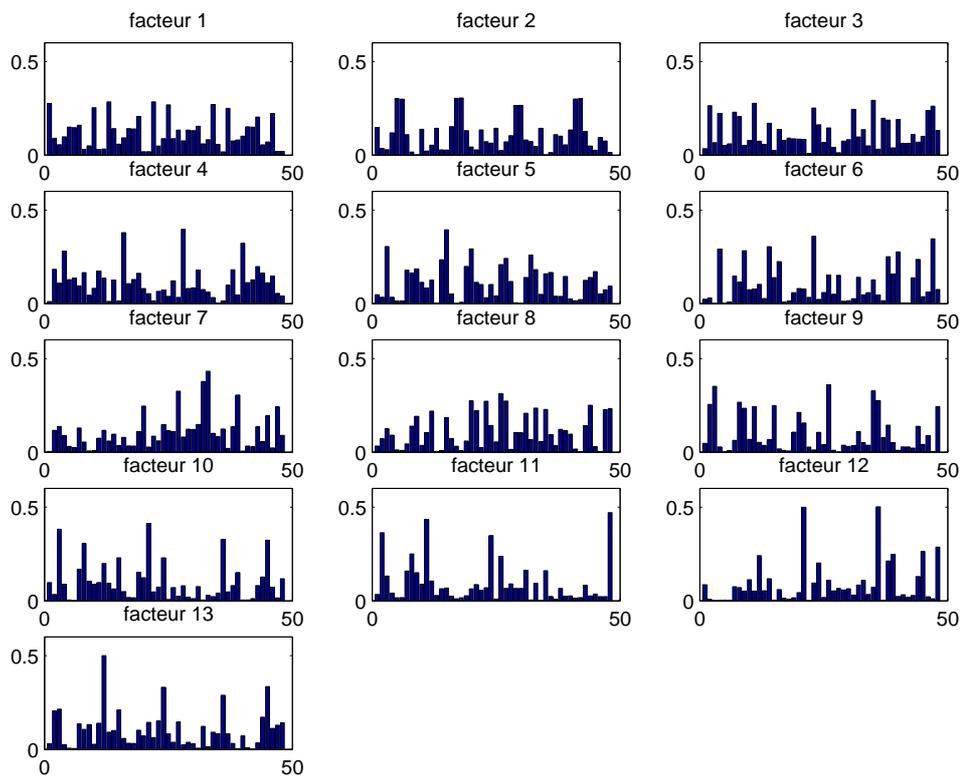


FIG. 7.6 – Eléments de la matrice de transformation facteur par facteur (valeurs absolues).

Facteur	Variabes	Indicateurs
1	1,10,13,22,25,34,37	RMS global e RMS haute fréquence
2	5,6,17,18,29,30,41,42	Kurtosis et Skewness dans la gamme 0-30 Hz
3	2,11,35,47	Kurtosis global et haute fréquence
4	4,16,28,40	RMS 0-30 Hz
5	3,15,20,32	Skewness global et Kurtosis 0-30 Hz
6	4,9,14,23,40,47	Pas de grande cohérence
7	27,32,33,39	Pas de grande cohérence
8	20,23,26,27	Pas de grande cohérence
9	2,3,8,26,35	Pas de grande cohérence
10	3,8,21,36,45	Pas de grande cohérence
11	2,8,11,24,48	Pas de grande cohérence
12	21,36	Pas de grande cohérence
13	12,24,36,45	Pas de grande cohérence

TAB. 7.2 – Variables ayant des coefficient supérieurs à 0,25 dans les combinaisons linéaires représentant les scores factoriels.

Pour analyser l'importance des poids synaptiques, nous avons calculé, pour chaque facteur, la moyenne des poids, ceux-ci considérés en valeur absolue. Nous avons, en outre, estimé qu'une entrée très pondérante devait non seulement présenter des poids assez élevés en valeurs absolues mais aussi avoir ces valeurs suffisamment bien regroupées autour de leur moyenne. L'écart-type, étant une mesure de la dispersion autour de la moyenne, peut servir pour la détermination du niveau de regroupement des valeurs absolues des poids autour de leur moyenne. Plus l'écart-type est élevé plus la dispersion est grande, ceci nous suggère donc de diviser la moyenne par l'écart-type pour avoir une mesure du degré d'importance.

L'introduction de l'aspect de contenu informationnel déduit de l'analyse en composantes principales pourra se faire en pondérant les valeurs obtenues à l'étape précédente par la contribution individuelle des facteurs à la variance totale, par exemple. On aboutit ainsi au degré d'importance des indicateurs défini comme suit :

$$d_i = \frac{\text{mean}[abs(IW_i)]}{\text{std}[abs(IW_i)]} \cdot \frac{\lambda_i}{\sum \lambda} \quad (7.1)$$

où

- $mean$, std et abs représentent respectivement le calcul de la moyenne, de l'écart-type et de la valeur absolue ;
- IW_i , le vecteur des poids synaptiques associés au facteurs i ;
- d_i , le degré d'importance du facteur i ;
- λ_i , la valeur propre correspondant à la direction principale i .

Les degrés d'importance pour chaque facteur sont donnés dans la dernière colonne du tableau 7.3 et représentés sur la figure 7.7.

Ainsi défini, le degré d'importance permet de déterminer les scores factoriels les plus pondérant dans la classification, donc les scores dont on extraira les indicateurs pertinents pour la détection des défauts. Les facteurs 1 et 2 sont encore les seuls lus influants dans la classification.

De ceci nous pouvons donc conclure que les indicateurs représentés suivant ces deux axes sont les plus pertinents. Il s'agit de :

- la valeur efficace du signal global
- la valeur efficace du signal haute fréquence
- du kurtosis en basse fréquence (0-30 Hz)
- et du skewness en basse fréquence (0-30 Hz)

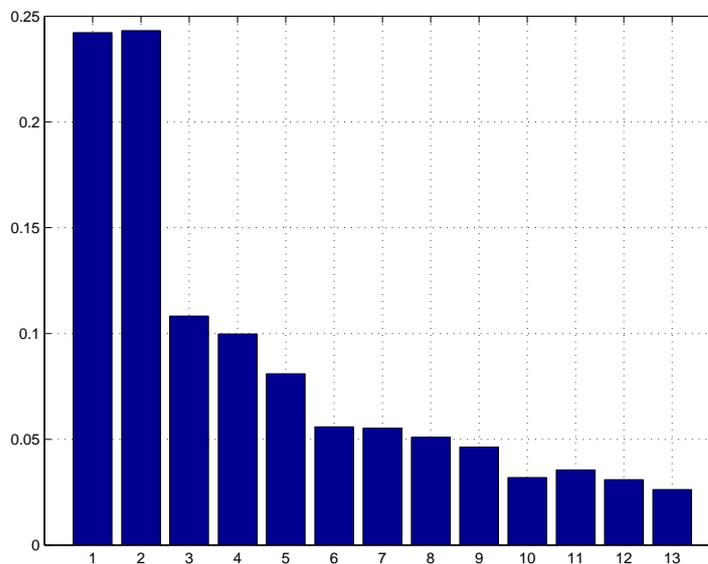


FIG. 7.7 – Degré d'importance des facteurs

	Moyenne	Variance	Ecart-type	d_i
FACTEUR 1	1,6248	1,6128	1,2700	0,2422
FACTEUR 2	1,1071	0,4954	0,7038	0,2432
FACTEUR 3	1,6095	1,7560	1,3252	0,1082
FACTEUR 4	1,7416	1,3623	1,1672	0,0998
FACTEUR 5	1,5640	1,1867	1,0894	0,0810
FACTEUR 6	1,4622	1,4602	1,2084	0,055868
FACTEUR 7	1,5147	1,308	1,1437	0,055219
FACTEUR 8	1,6603	1,3646	1,1682	0,050911
FACTEUR 9	1,7596	1,4725	1,2135	0,046273
FACTEUR 10	1,4442	1,437	1,1988	0,031932
FACTEUR 11	1,7067	1,5019	1,2255	0,035498
FACTEUR 12	1,2873	1,0064	1,0032	0,030932
FACTEUR 13	1,5157	1,3985	1,1826	0,026217

TAB. 7.3 – Importance des facteurs avec considération des poids synaptiques

7.2.5 Vecteurs construits avec seuls les indicateurs pertinents

Ayant déterminé les indicateurs pertinents de l'analyse des saturations, de la matrice de transformation et des poids synaptiques, nous allons examiner les capacités d'un classificateur qui serait construit dans un espace d'indicateurs composé uniquement de ces indicateurs. Ici, il s'agira donc de construire des vecteurs d'entrée dans l'espace des valeurs efficaces des signaux globaux et haute fréquence (300-2000 Hz), des Kurtosis et Skewness dans la gamme 0-30 Hz.

Alimenté par les variables normalisées contenues dans les facteurs 1 et 2 (tableau 7.1, le réseau de neurones présente avec des données non bruitées les performances suivantes :

1. taux de bonne classification 93,8% ;
2. les corrélations entre les sorties obtenues et les cibles sont illustrées figure 7.8.

La classification obtenues est même meilleure que celle réalisée avec les scores factoriels. L'espace d'indicateurs ainsi construit est libéré des bruits informationnels insérés par les variables moins pertinents, ce qui explique l'amélioration des résultats.

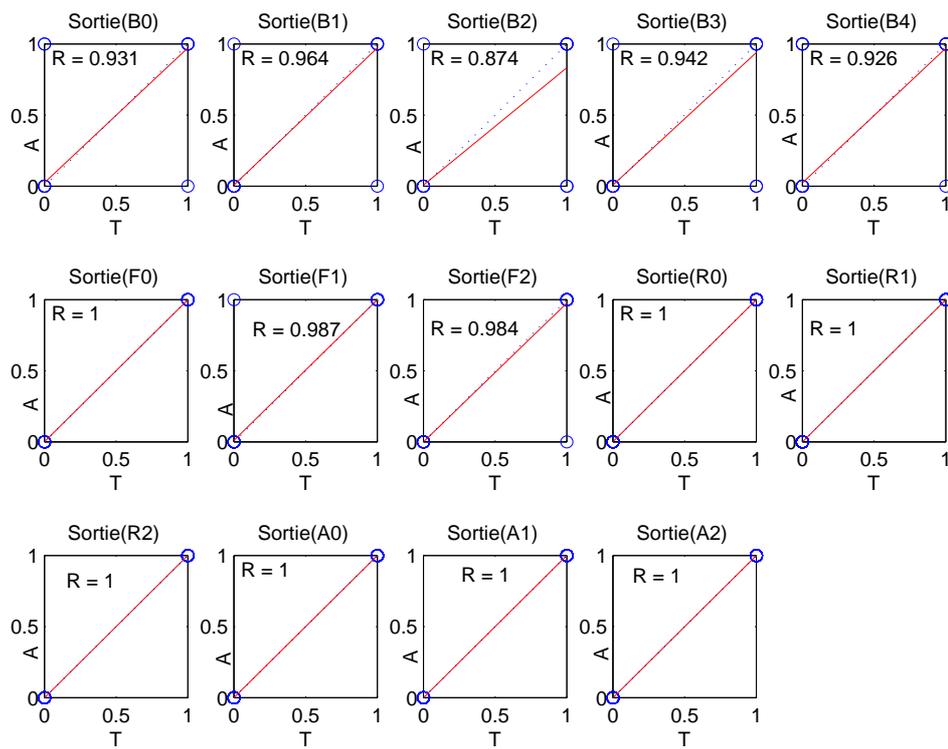


FIG. 7.8 – Corrélations entre sorties et cibles dans le cas des variables temporelles jugées pertinentes

Entraîné avec des données bruitées, le réseau de neurones présente une classification de 95% dans cet espace d'indicateurs. La manière dont les sorties suivent les cibles est représentée sur la figure 7.9

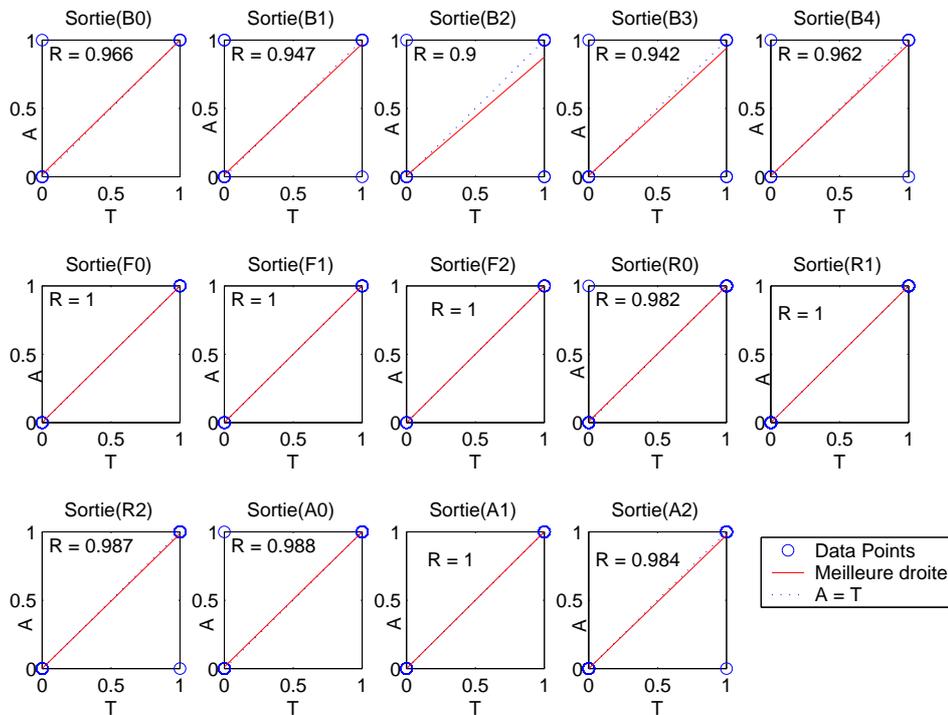


FIG. 7.9 – Corrélations entre sorties et cibles dans le cas des variables temporels jugés pertinents. Données bruitées.

Dans le cas bruité et non bruité les corrélations sont différentes de l'unité pour les sorties de balourd. Ceci pourrait s'expliquer par le fait que la plus grande énergie due à ce phénomène est contenue dans les basses fréquences. Donc, comme les indicateurs pertinents repris ici sont tirés du signal global et du signal haute fréquence, les phénomènes ayant lieu en basse fréquence seront moins bien représentés par rapport aux autres.

7.3 Analyse des indicateurs fréquentiels

Des graphiques de la figure 7.10 on voit que les facteurs 1 et 2 sont les plus corrélés avec les variables d'origine.

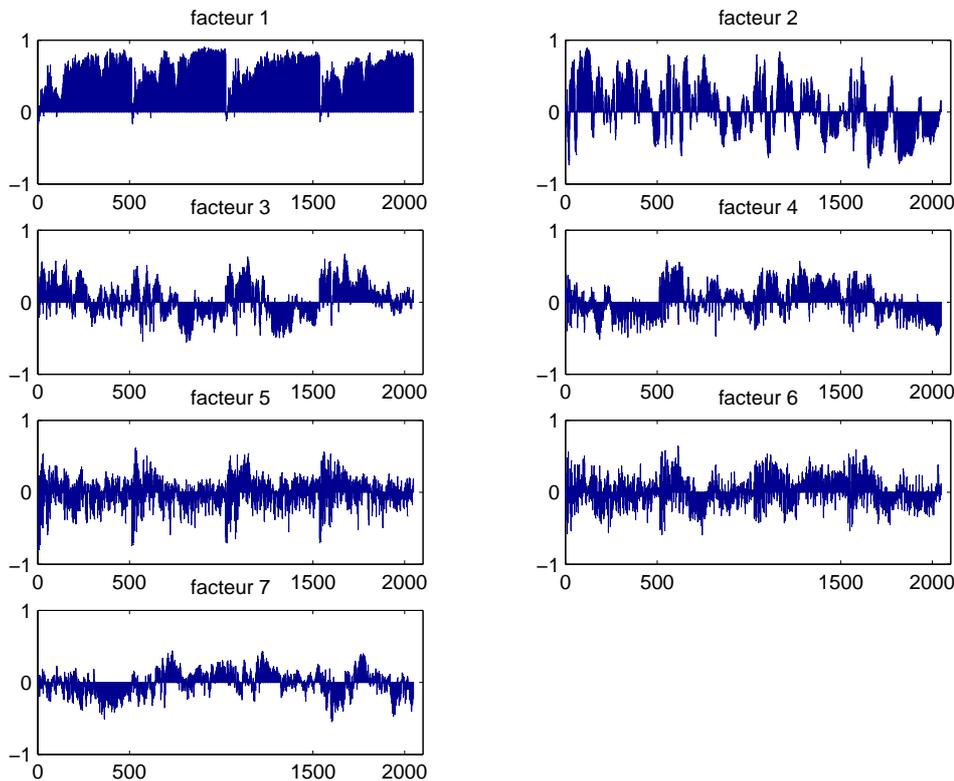


FIG. 7.10 – Saturations des variables spectrales

La figure 7.11 qui reprend uniquement le facteur 1 révèle une meilleure représentation des composantes haute fréquence (saturations supérieures à 0,8 en valeur absolue). Ces composantes s'opposent sur l'axe de ce facteur aux composantes basse fréquence dues aux balourds ($1 \times$, $2 \times f_{rot}$; f_{rot} étant la fréquence de rotation).

Le facteur 2 contient plus les basses et les moyennes fréquences, mais leur représentation est moins bonnes par rapport aux phénomènes haute fréquence contenus dans le facteur 1 (figure 7.12).

Une étude d'indicateurs pertinents comme celle menée dans le domaine temporel serait dénuée de tout sens dans le domaine fréquentiel vu le nombre de variables

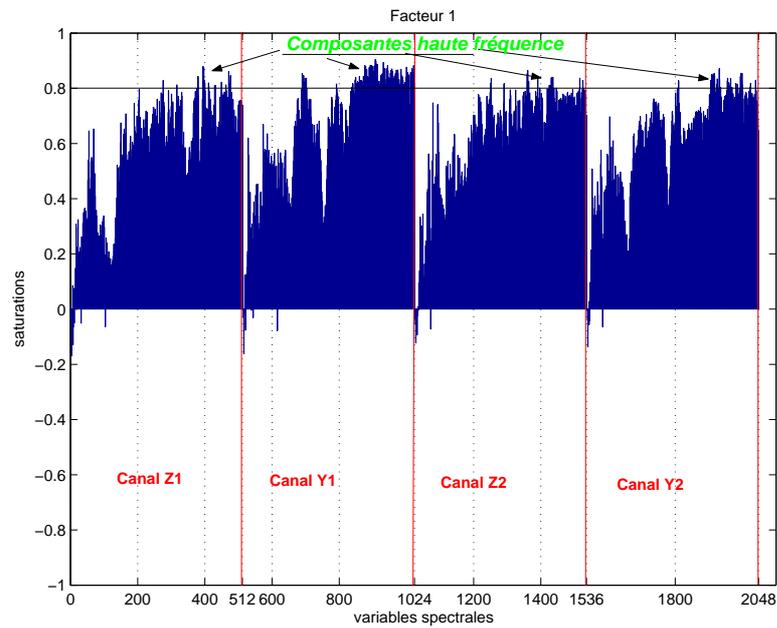


FIG. 7.11 – Saturations facteur 1. Remarquons que les hautes fréquences sont les mieux représentées.

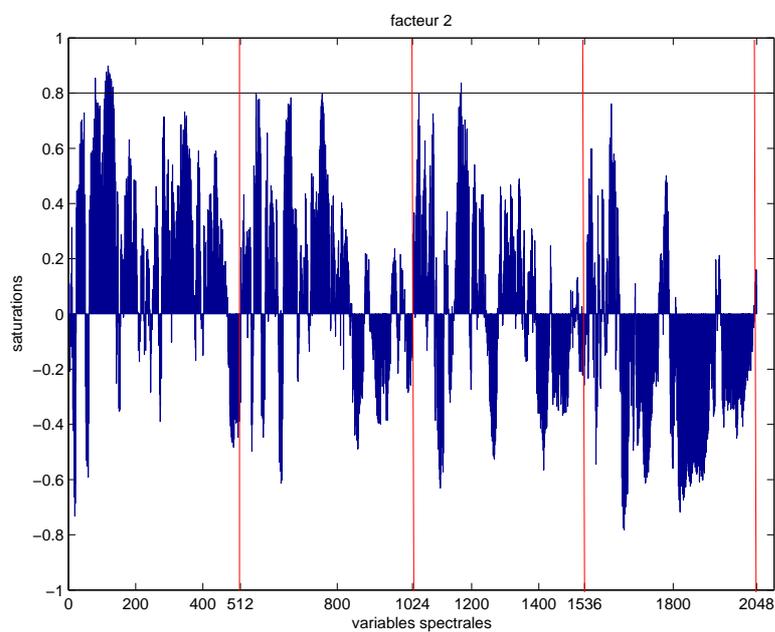


FIG. 7.12 – Saturations facteur 2

spectrales. Ici, on peut se contenter de la seule réduction dimensionnelle obtenue avec l'analyse en composantes principales au lieu de tenter une quelconque extraction des variables pertinentes, qui sont ici des raies spectrales.

Nous allons, néanmoins, étudier les degrés d'importance des scores factoriels afin de vérifier les observations établies dans la section précédente. En utilisant cette notion de degré d'importance, nous rejoignons ce que nous avons déjà établi sur base des saturations : que le facteur 1 est celui dont on peut analyser le contenu informationnel afin de voir quelles variables sont prépondérantes. Mais rappelons que cette étude n'a pas cet objectif dans le domaine fréquentiel ; en la menant nous voulions vérifier si le degré d'importance défini permet effectivement le choix des facteurs à examiner.

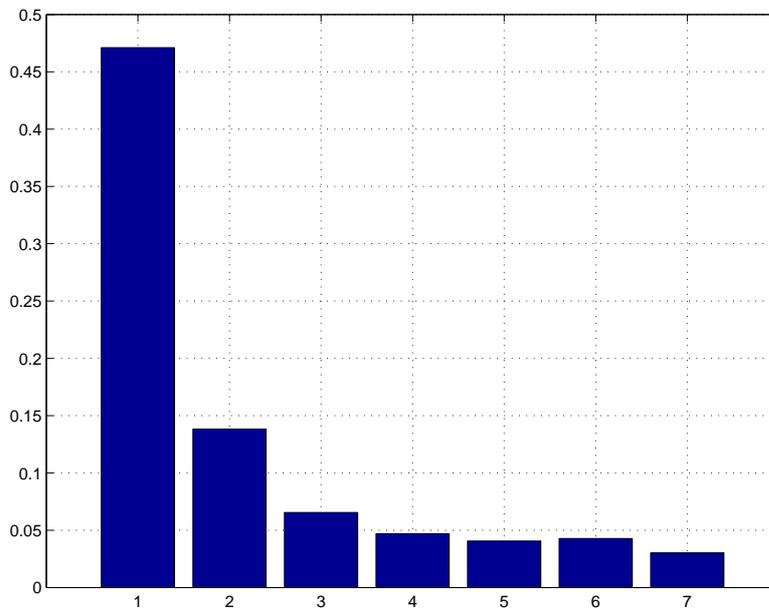


FIG. 7.13 – Degré d'importance des facteurs

7.4 Résumé

L'étude qui a été effectuée sous ce chapitre avait pour objectif de déterminer dans un espace d'indicateurs donné les variables qui influencent le plus la détection et la classification des défauts mécaniques. Cette étude recèle un sens seulement dans

Facteur	Moyenne	Variance	Ecart-type	d_i
Facteur 1	0,81972	0,43239	0,65756	0,47105
Facteur 2	0,99103	0,94424	0,97172	0,13825
Facteur 3	1,129	0,84786	0,92079	0,065452
Facteur 4	0,99463	0,77717	0,88157	0,047071
Facteur 5	0,6884	0,41071	0,64086	0,04066
Facteur 6	0,83844	0,44694	0,66854	0,042774
Facteur 7	1,0855	1,0079	1,0039	0,030398

TAB. 7.4 – Importance des facteurs avec considération des poids synaptiques

le domaine temporel où nous avons dégagé que pour les défauts traités seuls les indicateurs suivants étaient pertinents, il s'agit de :

- la valeur efficace du signal global ;
- la valeur efficace du signal filtré en haute fréquence (300-2000 Hz) ;
- le Kurtosis du signal filtré entre 0 et 30 Hz ;
- le Skewness du signal filtré entre 0 et 30 Hz.

L'examen du domaine fréquentiel a révélé que la zone haute fréquence avait la plus grande influence. On peut établir un lien entre les conclusions tirées dans le domaine temporel et celles provenant du domaine fréquentiel. En effet la valeur efficace dans une bande fréquentielle est liée au niveau d'énergie contenu dans cette bande. Il est donc compréhensible que l'information véhiculée par les valeurs efficaces haute fréquence se retrouve dans les raies spectrales à haute fréquence.

Conclusions et perspectives

Conclusions

Les objectifs poursuivis dans ce travail étaient d'une part, d'étudier l'applicabilité des réseaux de neurones aux fins de détection et de classification des défaillances mécaniques combinées, et d'autre part de dégager les indicateurs qui sont pertinents pour une telle application. Dans ce but, nous avons d'abord survolé, dans la littérature, les différentes études similaires en établissant au préalable une base de compréhension du concept de réseau de neurones. Nous nous sommes particulièrement attardé sur le problème crucial du choix d'indicateurs issus de l'analyse vibratoire et destinés à alimenter un réseau de neurones. Une fois tous ces aspects circonscrits, nous avons présenté et discuté les résultats de nos expériences conduites sur un banc d'essais simulateur de défauts mécaniques.

Les résultats obtenus ont dégagé, en premier lieu l'aspect que l'utilisation des réseaux de neurones dans le contexte du diagnostic mécanique nécessite la considération des mesures récoltées en plusieurs points de la machine. Un diagnostic basé sur un seul point de mesure conduit à des performances médiocres. L'analyse dans le domaine temporel s'est révélée intéressante seulement si elle est conduite sur des signaux filtrés dans des gammes de fréquence adéquatement choisies. En effet, à l'issue de la première phase de cette étude, les performances obtenues dans la classification, et reprises dans le tableau ci-dessous, justifient qu'au cours de la seconde phase nous n'ayons travaillé qu'avec les signaux filtrés et considéré tous les quatre canaux.

	1 Canal	4 Canaux
Filtré	50 %	83 %
Non filtré	33 %	75 %

Le domaine fréquentiel a été expérimenté dans la seconde phase de l'étude. Il a fourni des capacités de reconnaissance de défauts très satisfaisantes : 97,5 % de reconnaissance. L'écart d'avec le taux de classification de 89,5 % obtenu lors de cette phase pour les indicateurs temporels est très remarquable. La construction d'un espace d'indicateurs composé à la fois des indicateurs spectraux et temporels a abouti à un taux de classification intermédiaire entre ceux donnés par les deux types d'indicateurs précédents : 95,7 %.

Dans la seconde phase, le bruitage des données à l'entrée du réseau de neurones n'a pas aussi influencé les résultats qu'au cours de la première phase. Pour les indicateurs fréquentiels, par exemple, l'amélioration du taux de classification n'a été que de 0,6 % alors que des améliorations de 15% ont été possibles avec les données de la première phase. Cependant, rappelons qu'aucun résultat obtenu lors de la première phase n'a égalé ceux de la seconde. Nous avons supposé que cette faible influence du bruitage devrait être due à une bonne cohérence des données vu que la taille de l'ensemble d'apprentissage a été sensiblement augmentée par rapport à la première phase : 144 contre 648.

La structure du réseau de neurones utilisé ainsi que le prétraitement des données se sont affirmés comme deux facteurs très influents dans l'optimisation des performances de classification. La structure disposant une sortie binaire pour chaque niveau de défauts et appuyée par un processus de compétition est la mieux adaptée dans ce type de problème. Le prétraitement qui abouti à une réduction dimensionnelle a allégé l'apprentissage et amélioré les résultats.

L'analyse en composantes principales qui a conduit à cette réduction dimensionnelle, a aussi permis un examen des espaces d'indicateurs afin de déterminer les indicateurs qui disposaient d'une grande capacité de discrimination de défauts et de leurs niveaux. Cet examen a été procédé en nous basant sur les notions de saturations, de communalités et de matrices de transformation. Une approche intégrant les poids synaptiques a aussi été proposée dans l'analyse de l'influence des indicateurs. Les conclusions auxquelles nous avons abouti sont identiques pour les différentes approches et ont révélé, dans le domaine temporel, les indicateurs suivants comme portant le plus grand poids d'information discriminatoire entre classes de défauts :

- la valeur efficace du signal global ;
- la valeur efficace du signal filtré en haute fréquence (300-2000 Hz) ;

- le Kurtosis du signal filtré entre 0 et 30 Hz ;
- le Skewness du signal filtré entre 0 et 30 Hz.

La question de la taille de l'échantillon pour la mise en oeuvre d'un diagnostic basé sur un réseau de neurones a été mis en évidence en comparant les résultats obtenus dans la première phase expérimentale à ceux de la seconde phase. Il est donc souhaitable de considérer un assez vaste jeu de données pour s'assurer la construction d'un réseau de neurones efficace, ce qui peut être une difficulté dans certains contextes industriels.

Une autre difficulté dans l'application industrielle des réseaux de neurones tels qu'utilisés dans notre travail est le fait que l'état d'une machine évolue avec le temps. Il peut alors être intéressant de construire un réseau de neurones pour chaque élément de machine qui nécessite une surveillance particulière au lieu d'un réseau capable de détecter globalement les défaillances de divers éléments. Cependant, on gardera toujours à l'esprit le fait que, dans leurs états avancés, les défauts peuvent interagir et rendre ainsi la classification incertaine. En exemple, nous pouvons citer le cas où un désalignement conduit à un balourd.

Une signature d'un défaut donné récoltée à une certaine période sera logiquement sans commune mesure avec ce même défaut exprimé à un âge avancé de la machine. En effet, il est très fréquent que certains défauts interagissent entre eux à leurs états avancés.

Perspectives

A terme, il serait intéressant qu'une étude similaire à celle-ci soit menée avec d'autres méthodes de diagnostic afin de comparer les résultats à ceux auxquels a abouti ce travail. On pourra utiliser des méthodes de classification basées sur l'intelligence artificielle ou non telles que les systèmes experts, la logique floue, les arbres de décisions, ...

L'exploitation de ces méthodes combinées à d'autres types d'indicateurs qui n'ont pas fait l'objet de ce travail peut être souhaitable dans le but de déterminer,

pour un type de problème donné, les indicateurs les mieux dédiés. Les spectres paramétriques (Annexe F) et les statistiques supérieures (Annexe G) pourront être considérés à cet effet.

Les études pourront ainsi concerner des cas industriels réels pour lesquels, actuellement, la disponibilité d'une large base de données n'est peut-être pas aussi illusoire que l'on pourrait le penser. En effet, l'introduction de la Gestion de la Maintenance Assistée par Ordinateur (GMAO) a systématisé les rapports sur les analyses de défaillances. En outre, la surveillance des machines par l'analyse vibratoire s'est considérablement développée dans beaucoup d'industrie ; les enregistrements périodiques systématiques de plusieurs indicateurs sont en principe sauvegardés. Hélas, la mise en relation de ces deux jeux de données n'est pas réalisée en milieu industriel à l'heure actuelle.

Le potentiel est large car les installations industrielles comptent plusieurs machines de même type. Bien que chaque machine ait un comportement unique, au lieu de considérer chaque machine individuellement, on peut concevoir une exploitation des mesures répétées sur plusieurs machines semblables dans diverses conditions de fonctionnement. Les informations ainsi récoltées seront alors regroupées et traitées avec des techniques d'analyse "intelligentes" des données adaptées aux échantillons de grande taille. Parmi celles-ci, les techniques de "Data Mining" basées sur des méthodes d'apprentissage fournissent un moyen d'extraire les masses d'informations implicites souvent emprisonnées sous forme brute dans les bases des données. La capacité de dégager les anomalies, les classes et les tendances s'avère prometteuse pour la détection et le diagnostic des défauts mécaniques des machines tournantes.

Bibliographie

- [1] D.J.Edwards,G.D.Holt : Predicting mechanical reliability using artificial intelligence :Neural networks handle multi-variable data easily. School of Engineering and Built environnement, University of Wolverhampton,U.K
- [2] M.Baseville, A.Benveniste, Q.Zhang : Surveillance d'installations industrielles : démarche et conception de l'algorithme, Rapport de recherche n°2889, INRIA, 1996
- [3] A.Boulenger,C. Pachaud : Analyse vibratoire en maintenance :Surveillance et diagnostic des machines,2^eédition, Dunod, Paris, 2003
- [4] F.S.Osório : Un système hybride neuro-symbolique pour l'apprentissage automatique constructif. thèse de doctorat, Institut National Polytechnique de Grenoble, 1998.
- [5] Amit Konar : Artificial intelligence and soft computing : Behavioural and cognitive modelling of the human brain.CRC Press,NY,2000.
- [6] M.Serridge : Ten crucial concepts behind trustworthy fault detection in machine condition monitoring, Proceedings of the 1st International machinery monitoring and diagnostics conference and exhibit, Las Vegas, NV, 1989, 722-727
- [7] Lapp S.A..Powers G.A. : Computer-aided synthesis of fault-trees, IEEE Trans.Reliability,37, 2-13,1977.
- [8] Zwingelstein G. : Diagnostic des défaillances :Théorie et pratique pour les systèmes industriels,Traités des Nouvelles Technologie, série Diagnostic et Maintenance,Hermès,1995
- [9] Sourabh D.,Venkatsubramanian V. : Challenges in the industrial applications of faults diagnostic system,

- [10] Dujardin A.S. : Pertinence d'une approche hybride multineuronale dans la resolution de problèmes liés au diagnostic industriel ou médical
- [11] Zahner D.A. : Micheli-Tzanakou E. Artificial neural networks : definitions, Methods, Applications.
- [12] M.Angels : Choosing accelerometers for machinery health monitoring, Sound and Vibration (December 1990)20-24
- [13] A.Ypma : learning methods for vibration analysis and health monitoring, PhD thesis, Technische universiteit Delft, 2001
- [14] A.C.McCormick : Cyclostationarity and higher-order stitiscal signal processing for machine condition monitoring, PhD thesis, University of Strathclyde,1998
- [15] C.Both,J.R.McDonald : The use of artificial networks for condition monitoring of electrical power transformers, Neurocomputing 23 (1998) 97-109
- [16] S.Rangwala,D.Dornfeld : Sensor integration using neural networks for intelligent tool condition monitoring, Transactions of the ASME, Journal of Engineering for Industry, vol. 112, 219-228, 1990
- [17] R.J.Kuo,P.H.Cohen : Intelligent tool wear estimation system through artificial neural networks and fuzzy modelling, Artificial Intelligence in Engineering, 12(1998), 229-242
- [18] P.Dehombreux : Fiabilité et maintenance des équipements industriels, notes de cours, Faculté polytechnique de Mons, 2003.
- [19] Guanglin Hou, P.Dehombreux, O. Basile : Intégration des concepts d'optimisation dans la gestion de la maintenance basée sur la fiabilité. Rapport scientifique OPTIMAIN, janvier 2004.
- [20] Issam Abu-Mahfouz : Drilling wear detection and classification using vibration signals and artificial neural network, International Journal of Machine Tools and Manufacture XX (2003) XXX-XXX
- [21] L.X.Kong, S.Nahavandi : On-line tool condition monitoring and control system in forging processes, Journal of Materials Processing Technology 125-126 (2002) 464-470
- [22] D.Chang-Ching Lin, B.Hsu-Pin Wang : Performance analysis of rotating machinery using enhanced cerebellar model articulation (E-CEMAC) neural networks. Computers ind.Engng Vol.30, N°2, pp227-242, 1996

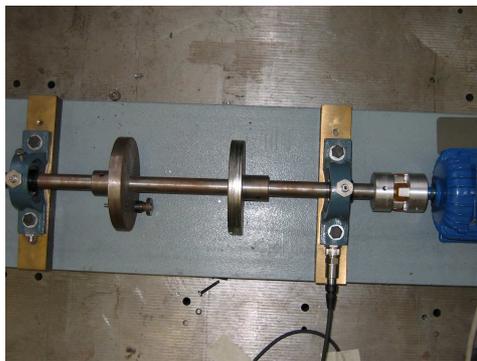
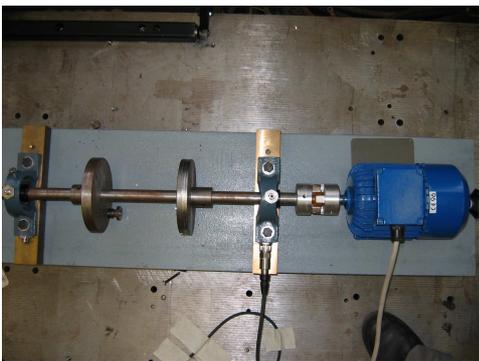
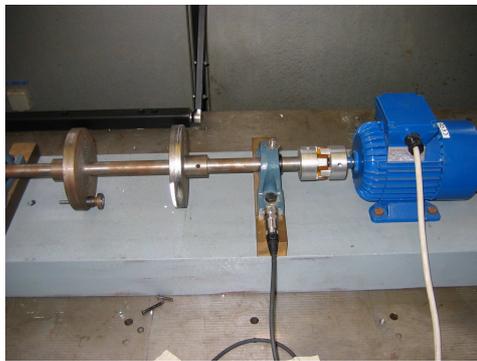
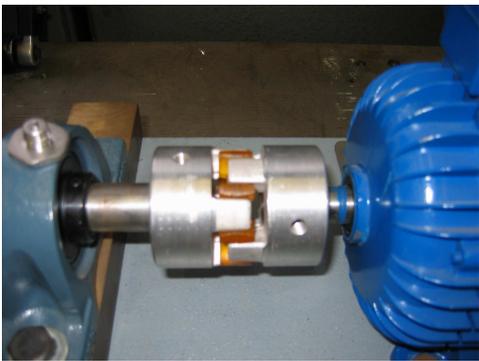
- [23] S.C.Lin, R.J.Lin : Tool wear monitoring in face milling using force signals, *Wear*, Vol.198, Nos.1-2, pp.136-142 (1996)
- [24] A.C.McCormick, A.K.Nandi : Classification of the rotating machine condition using artificial neural networks, *Proceedings of IMechE :PartC*, pp 439-450, Vol 211(6), 1997
- [25] Venkatsubramanian V., R.Rengaswamy, S.N.Kavuri, K.Yen : A review of process fault detection and diagnosis :Process history based methods, *Computers and Chemical Engineering* 27 (2003) 327-346
- [26] Michael Berthold, David J.Hand : *Intelligent Data Analysis*. 2nd ed, Springer, Berlin, 2003.
- [27] A.C.McCormick, A.K.Nandi : Neural network autoregressive modelling of vibrations for condition monitoring shafts. Signal Processing Division, Department of Electronic and Electrical Engineering, University of Strathclyde,U.K.
- [28] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, vol. 9,127-147,1943.
- [29] H.Demuth, M.Beale, *Neural Network Toolbox for use with Matlab*. The Math-Works, Natick, USA, 2003
- [30] Joëlle Courrech, Ronald Eshleman : Condition monitoring of machinery. In *Harris' shock and vibration handbook*, p.16.1-16.25, 5th edition, 2002
- [31] C.Breneur : *Eléments de maintenance préventive de machines tournantes dans le cas de défauts combinés d'engrenages et de roulements*, Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, décembre 2002.
- [32] S.Marple : *Digital spectral analysis with applications*. Prentice-Hall, 1987.
- [33] J.K.Spoerre : Application of the cascade correlation algorithm (CCA) to bearing fault classification problems. *Computers in Industry* 32 (1997) 295-304.
- [34] R.Javadpour, G.M.Knapp : A fuzzy neural network approach to machine condition monitoring. *Computers & Industrial engineering* 45 (2003) 323-330
- [35] G.K. Singh, S.A.Kazzaz : Induction machine drive condition monitoring and diagnostic research-a survey. *Electric Power Systems Research* 64(2003) 145-158
- [36] I.S.Koo, W.W.Kim : The development of reactor coolant pump vibration monitoring and a diagnostic system in nuclear power plant. *ISA transactions* 39 (2000) 309-316

- [37] A.Ypma, R.Ligteringen, E.E.E.Frietman, R.P.W.Duin : Recognition of bearing failures using wavelets and neural networks. Computational physics group, Delft University of Technology, Netherlands.
- [38] A.Papoulis : Probability, Random Variables, and Stochastic Processes. McGraw-Hill, 1965.
- [39] Signal Processing Toolbox for use with Matlab. The MathWorks, Natick,USA, 2002
- [40] Fukunaga K. : Introduction to Statistical Pattern Recognition. Academic Press, 1990.
- [41] Fukunaga K, Hayes RR : Effect of sample size in classifier performance. IEEE Tr.Patt.Anal.Mach.Intel., 11 :1087-1101, 1989.
- [42] Jain AK, Chandrasekaran B : Dimensionality and sample size considerations in ptttern recognition. In :Krishnaiah Pr, Kanal LN (eds) Handbook of statistics, 2, North Holland Pub.Co.,pp 835-855, 1989.
- [43] Marquès de Sá J.P. : Applied Statistics using SPSS, STATISTICA and MATLAB. Springer, Berlin 2003.
- [44] Holstrom L, Koistinen P. : Using additive noise in backpropagation training. IEEE Trans Neural Networks 3(1) :24, 1992.
- [45] Gilles Fleury : Analyse spectrale. Méthodes non-paramétriques et paramétriques. Technosup, Ellipses Editions, Paris, 2001 ;

Annexes

Annexe A

Le banc d'essais



Annexe B

Les défauts traités expérimentalement sur le banc d'essais

Toutes les conditions de fonctionnement qui ont été traitées sur le banc d'essais sont reprises ci-dessous. La vitesse 1 correspond à une fréquence de rotation de 25 Hz, la vitesse 2 à 20 Hz et la vitesse 3 à une fréquence de 22,5 Hz. La codification est expliquée en 3.4.

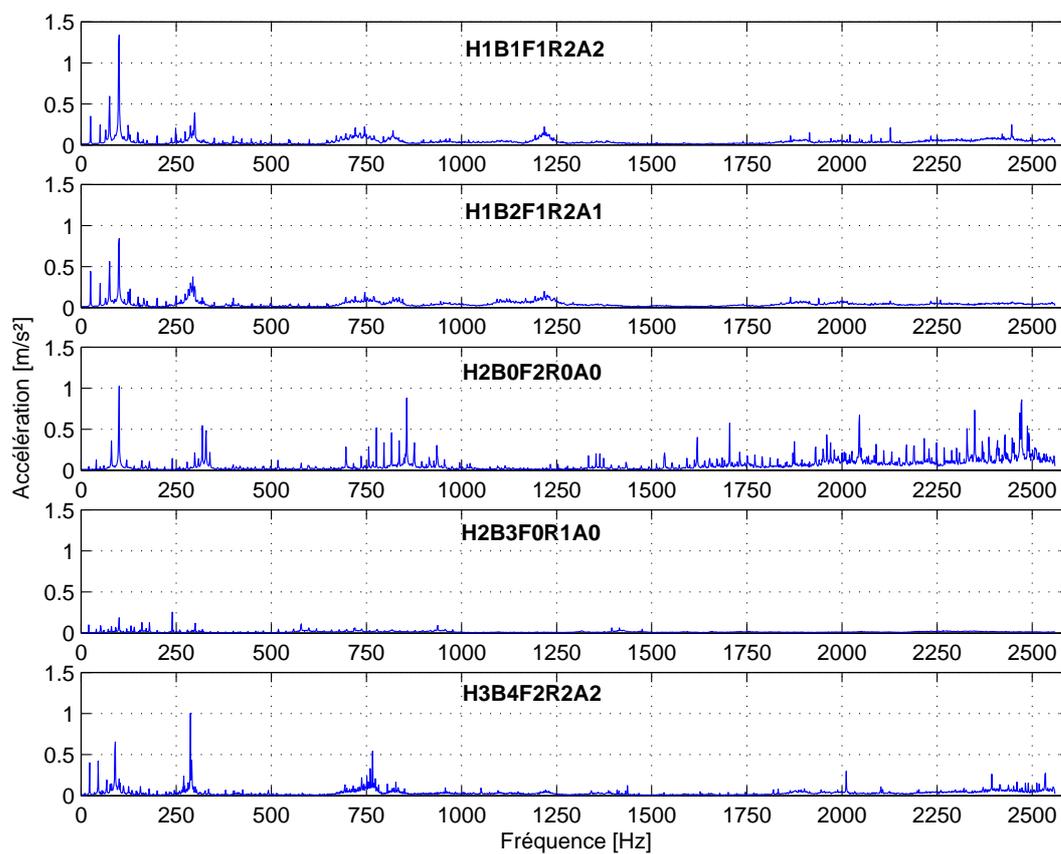
VITESSE 1		VITESSE 2		VITESSE 3	
1	H1B0F0R0A0	1	H2B0F0R0A0	1	H3B0F0R0A0
2	H1B0F0R1A0	2	H2B0F0R1A0	2	H3B0F0R1A0
3	H1B0F0R2A1	3	H2B0F0R2A1	3	H3B0F0R2A1
4	H1B0F0R2A2	4	H2B0F0R2A2	4	H3B0F0R2A2
5	H1B0F1R0A0	5	H2B0F1R0A0	5	H3B0F1R0A0
6	H1B0F1R1A0	6	H2B0F1R1A0	6	H3B0F1R1A0
7	H1B0F1R2A1	7	H2B0F1R2A1	7	H3B0F1R2A1
8	H1B0F1R2A2	8	H2B0F1R2A2	8	H3B0F1R2A2
9	H1B0F2R0A0	9	H2B0F2R0A0	9	H3B0F2R0A0
10	H1B0F2R1A0	10	H2B0F2R1A0	10	H3B0F2R1A0
11	H1B0F2R2A1	11	H2B0F2R2A1	11	H3B0F2R2A1
12	H1B0F2R2A2	12	H2B0F2R2A2	12	H3B0F2R2A2
13	H1B1F0R0A0	13	H2B1F0R0A0	13	H3B1F0R0A0
14	H1B1F0R1A0	14	H2B1F0R1A0	14	H3B1F0R1A0
15	H1B1F0R2A1	15	H2B1F0R2A1	15	H3B1F0R2A1
16	H1B1F0R2A2	16	H2B1F0R2A2	16	H3B1F0R2A2
17	H1B1F1R0A0	17	H2B1F1R0A0	17	H3B1F1R0A0
18	H1B1F1R1A0	18	H2B1F1R1A0	18	H3B1F1R1A0
19	H1B1F1R2A1	19	H2B1F1R2A1	19	H3B1F1R2A1

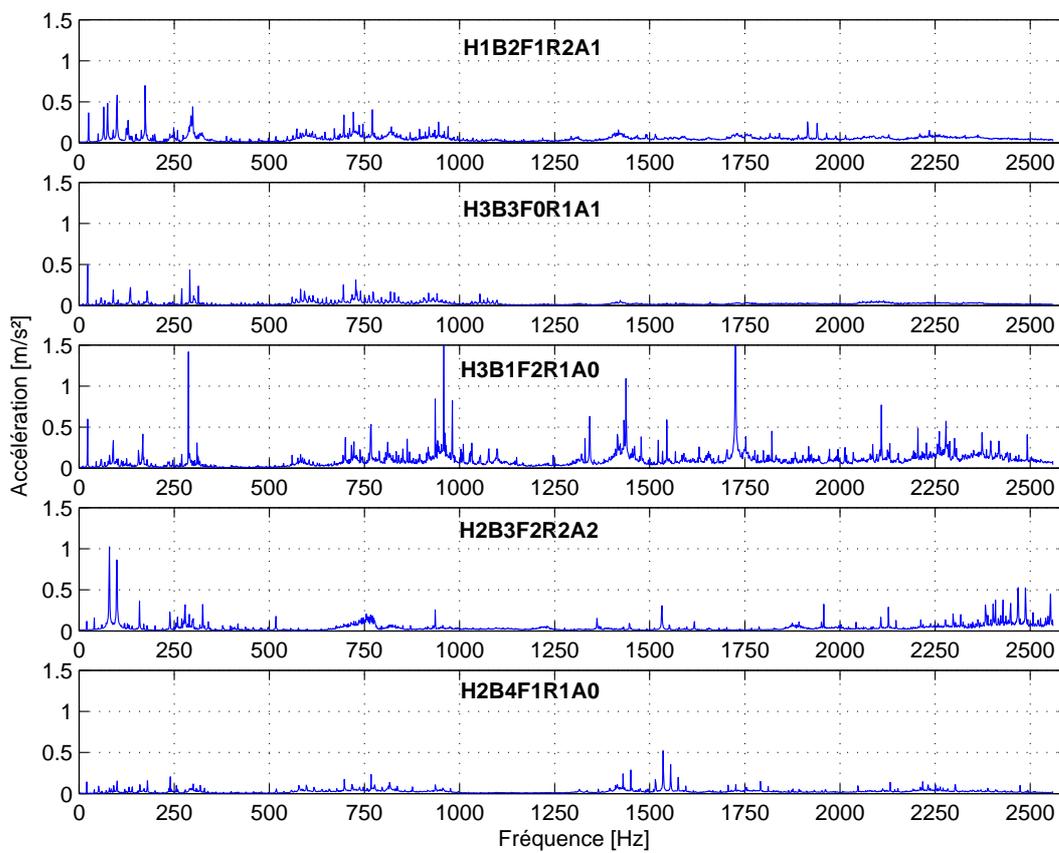
VITESSE 1		VITESSE 2		VITESSE 3	
20	H1B1F1R2A2	20	H2B1F1R2A2	20	H3B1F1R2A2
21	H1B1F2R0A0	21	H2B1F2R0A0	21	H3B1F2R0A0
22	H1B1F2R1A0	22	H2B1F2R1A0	22	H3B1F2R1A0
23	H1B1F2R2A1	23	H2B1F2R2A1	23	H3B1F2R2A1
24	H1B1F2R2A2	24	H2B1F2R2A2	24	H3B1F2R2A2
25	H1B2F0R0A0	25	H2B2F0R0A0	25	H3B2F0R0A0
26	H1B2F0R1A0	26	H2B2F0R1A0	26	H3B2F0R1A0
27	H1B2F0R2A1	27	H2B2F0R2A1	27	H3B2F0R2A1
28	H1B2F0R2A2	28	H2B2F0R2A2	28	H3B2F0R2A2
29	H1B2F1R0A0	29	H2B2F1R0A0	29	H3B2F1R0A0
30	H1B2F1R1A0	30	H2B2F1R1A0	30	H3B2F1R1A0
31	H1B2F1R2A1	31	H2B2F1R2A1	31	H3B2F1R2A1
32	H1B2F1R2A2	32	H2B2F1R2A2	32	H3B2F1R2A2
33	H1B3F0R0A0	33	H2B3F0R0A0	33	H3B3F0R0A0
34	H1B3F0R1A0	34	H2B3F0R1A0	34	H3B3F0R1A0
35	H1B3F0R1A1	35	H2B3F0R1A1	35	H3B3F0R1A1
36	H1B3F0R2A2	36	H2B3F0R2A2	36	H3B3F0R2A2
37	H1B3F1R0A0	37	H2B3F1R0A0	37	H3B3F1R0A0
38	H1B3F1R1A0	38	H2B3F1R1A0	38	H3B3F1R1A0
39	H1B3F1R2A1	39	H2B3F1R2A1	39	H3B3F1R2A1
40	H1B3F1R2A2	40	H2B3F1R2A2	40	H3B3F1R2A2
41	H1B3F2R1A0	41	H2B3F2R1A0	41	H3B3F2R1A0
42	H1B3F2R1A1	42	H2B3F2R1A1	42	H3B3F2R1A1
43	H1B3F2R2A2	43	H2B3F2R2A2	43	H3B3F2R2A2
44	H1B4F0R0A0	44	H2B4F0R0A0	44	H3B4F0R0A0

VITESSE 1		VITESSE 2		VITESSE 3	
45	H1B4F0R1A0	45	H2B4F0R1A0	45	H3B4F0R1A0
46	H1B4F0R1A1	46	H2B4F0R1A1	46	H3B4F0R1A1
47	H1B4F0R2A2	47	H2B4F0R2A2	47	H3B4F0R2A2
48	H1B4F1R0A0	48	H2B4F1R0A0	48	H3B4F1R0A0
49	H1B4F1R1A0	49	H2B4F1R1A0	49	H3B4F1R1A0
50	H1B4F1R1A1	50	H2B4F1R1A1	50	H3B4F1R1A1
51	H1B4F1R2A2	51	H2B4F1R2A2	51	H3B4F1R2A2
52	H1B4F2R1A0	52	H2B4F2R1A0	52	H3B4F2R1A0
53	H1B4F2R1A1	53	H2B4F2R1A1	53	H3B4F2R1A1
54	H1B4F2R2A2	54	H2B4F2R2A2	54	H3B4F2R2A2

Annexe C

Quelques spectres





Annexe D

La méthode *Varimax* en Analyse en Composantes Principales

Dans une ACP l'interprétation des saturations est aisée si les variables peuvent être regroupées en différents ensembles, chacun étant associé à un seul facteur. Pour y arriver, on procède à une rotation des saturations et la méthode *Varimax* est un des algorithmes les plus utilisés.

Pour mieux illustrer, considérons le cas où l'espace des facteurs a 2 dimensions. Une matrice de rotation \mathbf{G} est donnée par

$$\mathbf{G}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad (\text{D.1})$$

représentant une rotation dans le sens horaire du système de coordonnées d'un angle θ . La nouvelle matrice des saturations est par

$$\mathbf{Lr} = \mathbf{LG} \quad (\text{D.2})$$

La méthode *Varimax* essaie de trouver l'angle de rotation qui maximise la somme des variances des carrés des saturations. Ici est repris un code Matlab pour effectuer la rotation *varimax*

```
% Marcus, 1993, in Reyment & Joreskog,  
% Applied Factor Analysis in the Natural Sciences, CUP.
```

```
% This procedure follows algorithm as spelled out in
% Harman (1960) in Chapter 14, section 4. To run the
% program - the loadings are put in an array called
% lding.
% The notation follows Harman. The routine vfunct.m is
% called to compute the variance of the loadings
% squared.

lding=S';
b=lding;
[n,nf]=size(lding);
hjsq=diag(lding*lding'); % communalities
hj=sqrt(hjsq);
vfunct % function to compute
% variances of loadings^2

V0=Vtemp;
for it=1:10; % Never seems to need very many iterations
for i=1:nf-1 % Program cycles through 2 factors
    jl=i+1; % at a time.
    for j=jl:nf
        xj=lding(:,i)./hj; % notation here closely
        yj=lding(:,j)./hj; % follows harman
        uj=xj.*xj-yj.*yj;
        vj=2*xj.*yj;
        A=sum(uj);
        B=sum(vj);
        C=uj'*uj-vj'*vj;
        D=2*uj'*vj;
        num=D-2*A*B/n;
        den=C-(A^2-B^2)/n;
        tan4p=num/den;
        phi=atan2(num,den)/4;
        angle=phi*180/pi;
```

```
[i j it angle];
if abs(phi)>.00001;
    Xj=cos(phi)*xj+sin(phi)*yj;
    Yj=-sin(phi)*xj+cos(phi)*yj;
    bj1=Xj.*hj;
    bj2=Yj.*hj;
    b(:,i)=bj1;
    b(:,j)=bj2;
    lding(:,i)=b(:,i);
    lding(:,j)=b(:,j);
end
end
end;
lding=b;
vfunct;
V=Vtemp;
if abs(V-V0)<.0001;break;else V0=V;end;
end;

%disp(['Varimax Rotated Loadings:']);
%disp([lding]);
```

Le script *vfunct* est donnée ci-dessous

```
% Marcus, 1993, in Reyment & Joreskog,
% Applied Factor Analysis in the Natural Sciences, CUP.

% This little program computes the value of V (Harman's
% notation) given also in formula 7.4 in the text.
bh=lding./(hj*ones(1,nf));
Vtemp=n*sum(sum(bh.^4))-sum(sum(bh.^2).^2);
```

Annexe E

L'algorithme de rétropropagation du gradient

L'algorithme de rétropropagation du gradient suit la démarche de Widrow-Hoff. Définir une notion d'erreur sur un exemple puis calculer la contribution à cette erreur de chacun des poids synaptiques. C'est cette seconde étape qui n'est pas évidente. Elle est parfois désignée sous le nom de "Credit Assignment Problem".

Afin de pouvoir appliquer la méthode du gradient, on a besoin de calculer des dérivées et donc de lisser les calculs. On remplace pour cela la fonction à seuil de Heaviside par une fonction sigmoïde.

Cette fonction est une approximation indéfiniment dérivable de la fonction à seuil de Heaviside, d'autant meilleure que k est grand. On considère généralement $k = 1$

$$\text{sig}(x) = \frac{1}{1 + \exp(-kx)} \quad (\text{avec } k > 0) \quad (\text{E.1})$$

L'algorithme de rétropropagation est une procédure au cours de laquelle l'erreur commise en sortie est propagée vers les couches internes pour modifier les poids synaptiques. C'est une méthode de descente du gradient qui minimise l'erreur.

Soit

$$E_p = \frac{1}{2} \sum_{i=1}^N (T_i - A_i)^2 \quad (\text{E.2})$$

l'erreur associée à l'entrée p . N est le nombre de neurone dans la couche de sortie, T_i est la sortie attendue de la i -ème cellule de la couche de sortie, et A_i la sortie calculée par le neurone i .

$$\Delta w(n) = -\eta \frac{\partial E}{\partial w(n)} \tag{E.3}$$

Soit $E = \sum E_p$ la mesure totale de l'erreur. La méthode de la descente du gradient

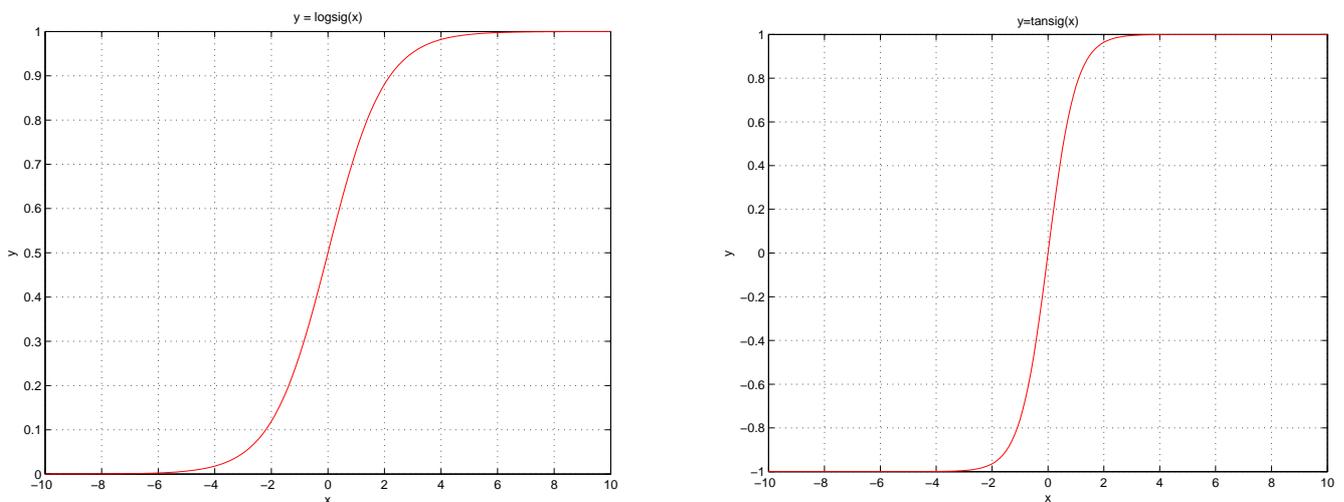


FIG. E.1 – La fonction sigmoïde log à gauche et tan à droite

modifie un poids arbitraire w du réseau de neurones de la manière suivante :

$$w(n + 1) = w(n) + \Delta w(n) \tag{E.4}$$

où n est l'indice d'itération et η une constante d'échelle.

Cette méthode demande le calcul des dérivées $\partial E / \partial w(n)$ pour chaque poids w du réseau de neurones. Pour un noeud arbitraire d'une couche cachée, sa sortie a_j est une fonction non linéaire f de son entrée n_j :

$$a_j = f(n_j) \tag{E.5}$$

où f est la fonction d'activation, généralement de type sigmoïde.

On peut donc écrire

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial n_j} \frac{\partial n_j}{\partial w_{ij}} \tag{E.6}$$

et comme

$$n_j = \sum_{i=1}^n w_{ij} p_i \quad (\text{E.7})$$

nous avons

$$\frac{\partial n_j}{\partial w_{ij}} = p_i \quad (\text{E.8})$$

l'équation (E.6) devient

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial n_j} p_i \quad (\text{E.9})$$

$$\frac{\partial E}{\partial n_j} = \sum_{k=1}^m \frac{\partial E}{\partial n_k} \frac{\partial n_k}{\partial a_j} \frac{\partial a_j}{\partial n_j} \quad (\text{E.10})$$

Comme

$$n_k = \sum_{j=1}^n a_j \quad (\text{E.11})$$

il suit que

$$\frac{\partial n_k}{\partial a_j} = w_{kj} \quad (\text{E.12})$$

et

$$\frac{\partial a_j}{\partial n_j} = f'(n_j) \quad (\text{E.13})$$

par conséquent

$$\frac{\partial E}{\partial n_j} = f'(n_j) \sum_{k=1}^m \frac{\partial E}{\partial n_k} w_{kj} \quad (\text{E.14})$$

Si f est une fonction sigmoïde d'équation (E.1) avec , alors

$$f'(n_j) = A_i(1 - A_i) \quad (\text{E.15})$$

L'équation (E.14) donne l'unique relation qui permet la rétropropagation de l'erreur vers les couches cachées. Pour la couche de sortie

$$\frac{\partial E}{\partial n_j} = \frac{\partial E}{\partial a_j} f'(n_j) \quad (\text{E.16})$$

$$\frac{\partial E}{\partial a_j} = -(T_i - A_i) \quad (\text{E.17})$$

En résumé, les sorties A_i sont d'abord calculées pour chaque neurones. La dérivée de l'erreur nécessaire à la règle de la descente du gradient de l'équation (E.4) est déterminée par

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial n} \frac{\partial n}{\partial w} \quad (\text{E.18})$$

Si j est un neurone de sortie, alors

$$\frac{\partial E}{\partial n_j} = -(T_i - A_i)A_i(1 - A_i) \quad (\text{E.19})$$

Si j est un neurone caché, alors la dérivée de l'erreur est propagée vers l'arrière en utilisant les équations (E.14) et (E.15). Substituant, on obtient

$$\frac{\partial E}{\partial n_j} = A_i(1 - A_i) \sum_{k=1}^m \frac{\partial E}{\partial n_k} w_{jk} \quad (\text{E.20})$$

Enfin, les poids sont modifiés comme dans l'équation (E.4).

Plusieurs modifications à cet algorithme ont été proposées dans le but d'accélérer la convergence. La *convergence* est définie comme la réduction de l'erreur à une valeur minimale fixée d'avance comme objectif.

Une méthode utilisée est l'insertion d'un moment dans l'équation de modification des poids

$$w(n+1) = w(n) - \eta \frac{\partial E}{\partial w_n} + \alpha \Delta w(n) \quad (\text{E.21})$$

η est le taux d'apprentissage et est souvent choisi égal à 0,25. Le terme constant α est un moment qui détermine l'effet des modifications des poids précédents sur la direction de mouvement des poids actuels.

Une autre approche pour améliorer la convergence est l'introduction d'un bruit aléatoire [44].

Annexe F

Méthodes spectrales paramétriques

L'estimation spectrale paramétrique découle d'une démarche de modélisation qui suppose que le signal traité appartient à une famille de signaux f . Cette famille dépend de *paramètres* ϕ qu'il s'agira dans un premier temps d'estimer. Une fois les paramètres estimés, on obtient une estimation du spectre par simple dualité temps-fréquence [45].

Les modèles les plus usuels sont ceux qui résultent d'une modélisation source-filtre. Un bruit blanc ω_k de puissance σ^2 est filtré par un filtre numérique $H(z)$ qui colore le signal selon le principe de la formule des interférences.

Ce filtre s'exprime usuellement selon une fraction rationnelle en z^{-1}

$$H(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_q z^{-q}}{a_0 + a_1 z^{-1} + \dots + a_p z^{-p}} \quad (\text{F.1})$$

Les paramètres ϕ du modèle sont alors :

- la puissance du bruit d'entrée σ^2 ;
- les coefficients du numérateur **b** ;
- les coefficients du dénominateur **a**.

Le cas le plus général est celui où le filtre possède de concert pôles et zéros. Le signal ainsi généré est dit ARMA ou signal Auto-Régressif à Moyenne Ajustée. Dans cette représentation, on extrait deux sous-classes de signaux particuliers. Les signaux AR (autorégressifs) correspondant à des filtres dont le numérateur est réduit à une constante (filtre IIR). Parallèlement, les signaux MA (Moyenne Ajustée), correspondant à des filtres dont le dénominateur est réduit à l'unité (filtre FIR).

F.1 Les indicateurs basés sur les paramètres autorégressifs (AR)

L'avantage de ces méthodes est de permettre la représentation d'un signal temporel aussi bien dans le domaine fréquentiel (PSD) que dans le domaine paramétrique. La représentation spectrale paramétrique peut être très indiquée quand on ne dispose que des courts échantillons des données. Elle est souvent mentionnée comme une alternative avantageuse aux méthodes spectrales non paramétriques (basées sur la transformée de Fourier) [13] [22].

En considérant les signaux vibratoires comme des séries temporelles d'une variable aléatoire $X(n)$, le modèle autorégressif décrit le signal comme la sortie d'un filtre IIR (Infinite Impulse Response) excité par un bruit blanc :

$$X(n) = \sum_{i=1}^p \phi_i X(n-i) + e(n) \quad (\text{F.2})$$

(F.3)

ou encore

$$X(n) - \phi_1 X(n-1) - \phi_2 X(n-2) - \dots - \phi_p X(n-p) = e(n) \quad (\text{F.4})$$

où les ϕ_i ($i = 1, \dots, p$) sont des paramètres et $e(n)$ une distribution gaussienne centrée à variance constante (bruit blanc). Il existe plusieurs techniques pour déterminer les paramètres AR : La méthode de Yule-Walker, la méthode de Burg, la méthode des covariances et la méthode des covariances modifiées pour lesquelles nous pouvons référer le lecteur à [32].

Si l'on utilise le domaine paramétrique chaque paramètre AR constituera une composante de vecteur *indicateur* à l'entrée du réseau de neurones pour le diagnostic. Il est donc capital de bien fixer l'ordre p du modèle AR pour éviter le surapprentissage quand l'ordre est très élevé. Ypma [13] propose une réduction de l'ordre par une analyse en composante principale. Sporre [33] qui applique le modèle AR dans la classification des défauts des paliers propose l'utilisation de deux critères pour la sélection de l'ordre du modèle, il s'agit de : l'erreur de prédiction finale (EPF) et du critère d'information de Akaike (CIA).

Le critère EPF pour un processus AR est défini comme suit

$$EPF(p) = \rho_p \left(\frac{N + (p + 1)}{N - (p + 1)} \right) \quad (\text{F.5})$$

où N est le nombre d'échantillons, p l'ordre et ρ_p l'estimateur de la variance du bruit blanc.

En supposant que le processus est gaussien, le CIA s'exprime pour un processus AR sous la forme suivante

$$CIA(p) = N \ln(\rho_p) + 2p \quad (\text{F.6})$$

Ces deux critères sont asymptotiquement équivalents c'est-à-dire lorsque le nombre d'observées N tend vers l'infini. Le critère EPF surestime l'ordre des signaux réels tandis le CIA le sousstime. Une variante du CIA fut alors développée par Rissanen, qu'il nomma MDL (Minimum Description Length)

$$MDL(p) = N \ln(\rho_p) + p \ln(N) \quad (\text{F.7})$$

La construction du spectre paramétrique d'un modèle AR est réalisée en introduisant les p paramètres ϕ_i dans la fonction théorique de densité spectrale de puissance définie par [22] :

$$P_{AR}(f) = \frac{2\Delta t \sigma^2}{|1 + \sum_{i=1}^p \phi_i e^{-i2\pi f k \Delta t}|} \quad (\text{F.8})$$

$$\begin{aligned} -\frac{1}{2} &\leq f \leq \frac{1}{2} \\ \Delta t &= \frac{1}{S} \end{aligned}$$

où S est la fréquence d'échantillonnage, p , l'ordre du modèle AR et σ^2 , la variance. Les paramètres ϕ_i sont mieux calculés avec la méthode des covariances modifiées et introduits dans l'équation F.8 pour déterminer le spectre du processus aléatoire.

On peut aussi utiliser l'algorithme de Burg pour l'estimation du spectre paramétrique.

Annexe G

Statistiques d'ordre supérieur : Bispectre et Trispectre

La densité spectrale de puissance $S(\omega)$ d'un signal $x(t)$ est égale à la transformée de Fourier de sa fonction d'autocorrélation $R(\tau)$

$$S(\omega) = \int_{-\infty}^{+\infty} e^{-j\omega\tau} R(\tau) d\tau \quad (\text{G.1})$$

où la fonction d'autocorrélation $R(\tau)$, qui peut être vue comme un moment temporel d'ordre 2, est définie par

$$R(\tau) = E\{x(t + \tau)x(t)\} \quad (\text{G.2})$$

Si l'on considère le moment temporel d'ordre 3 c'est à dire la moyenne du produit des valeurs du signal à 3 instants, on a

$$m_3(\tau_1, \tau_2) = E\{x(t)x(t + \tau_1)x(t + \tau_2)\} \quad (\text{G.3})$$

De telles définitions peuvent être faites pour les cumulants aussi.

Le bispectre $B(\omega_1, \omega_2)$ est alors défini comme la transformée de Fourier du cumulant temporel d'ordre 3

$$B(\omega_1, \omega_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} c_3(\tau_1, \tau_2) e^{-j\omega_1\tau_1} e^{-j\omega_2\tau_2} d\tau_1 d\tau_2 \quad (\text{G.4})$$

Le trispectre est

$$T(\omega_1, \omega_2, \omega_3) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} c_4(\tau_1, \tau_2, \tau_3) e^{-j\omega_1\tau_1} e^{-j\omega_2\tau_2} e^{-j\omega_3\tau_3} d\tau_1 d\tau_2 d\tau_3 \quad (\text{G.5})$$

Dans le domaine fréquentiel le bispectre et le trispectre sont tout simplement calculés par

$$B(\omega_1, \omega_2) = E\{X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2)\} \quad (\text{G.6})$$

et

$$T(\omega_1, \omega_2, \omega_3) = E\{X(\omega_1)X(\omega_2)X(\omega_3)X^*(\omega_1 + \omega_2 + \omega_3)\} \quad (\text{G.7})$$