

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MENTOURI - CONSTANTINE**

**FACULTE DES SCIENCES DE L'INGENIEUR
DEPARTEMENT D'ELECTROTECHNIQUE**

N° d'ordre :
Série :



MEMOIRE

*Présenté pour obtenir le diplôme de Magister
En Electrotechnique*

OPTION

Modélisation et commande des machines électriques

Par

Abla Bouguerne

THEME

**Diagnostic automatique des défauts
des moteurs asynchrones**

Soutenu le : --/--/2009

Devant le Jury :

Président :	Monsieur Aissa Bouzid	Prof. (Univ. Constantine)
Rapporteur :	Monsieur Abdessalem Lebaroud	MC (Univ. Skikda)
Examineurs :	Monsieurs Abdelmalek Khazzar	Prof. (Univ. Constantine)
	Ammar Bentounsi	Prof. (Univ. Constantine)



Sommaire

Tables des matières

Introduction générale	5
Chapitre I . La machine asynchrone et les méthodes de diagnostic	
1.1 Introduction.....	8
1.2 La machine asynchrone.....	9
1.2.1 Constitution de la machine asynchrone.....	9
1.2.1.1 Le stator	10
1.2.1.2 Le rotor à cage.....	11
1.2.1.3 Les organes mécanique.....	12
1.2.2 Etude statistique des défaillances de la machine asynchrone.....	13
1.2.2.1 Défaillances au stator.....	15
1.2.2.2 Défaillances au rotor.....	17
1.2.2.3 Défaillances mécaniques.....	20
1.3 Méthodes de surveillance et de diagnostic de la machine asynchrone.....	21
1.3.1 Méthodes basées sur une approche système.....	22
1.3.1.1 Réseaux de neurones.....	22
1.3.1.2 Reconnaissance de forme (RdF).....	24
1.3.1.3 Analyse temps-fréquence : outil privilégié de détection.....	25
1.3.2 Méthodes de prises de décision (classification).....	27
1.3.2.1 Discriminant Linéaire – ACP.....	27
1.3.2.2 Les K plus proches Voisins (K- Nearest Neighbors).....	28
1.3.2.3 Réseaux de Neurones	28
1.3.2.4 Méthodes de coalescence (clustering).....	29
1.4 Choix de la méthode de classification.....	30
1.4.1 Représentation temps-fréquence et classification.....	31
1.4.2 RTF dite "Dépendante de la Classe du Signal (DCS) ".....	32
1.4.3 La décision.....	33
1.5 Conclusion.....	33

Chapitre II. Extraction des paramètres de diagnostic par la représentation temps-fréquence optimisée

2.1 Introduction.....	35
2.2 La représentation temps-fréquence (RTF).....	35
2.2.1 Représentation temps-fréquence à noyau paramétrique.....	37
2.2.2 Représentation temps-fréquence à noyau non paramétrique.....	38
2.3 Choix de la méthode de diagnostic.....	38
2.3.1 Représentation temps-fréquence et classification.....	38
2.3.2 RTF dite "Dépendante de la Classe du Signal (DCS) ".....	40
2.3.3 La décision.....	40
2.4 Vers une méthode temps-fréquence de classification.....	41
2.4.1 La classe de Cohen.....	41
2.4.1.1 La classe de Cohen par la RTF.....	42
2.4.1.2 La classe de Cohen par la fonction d'ambiguïté.....	43
2.4.2 Existence et signification des interférences.....	45
2.4.3 Conception des noyaux pour la classification.....	45
2.4.4 Le critère de décision.....	46
2.5 RTF optimisée pour la classification.....	47
2.6 Application de la RTF optimisée pour la classification des défauts de la machine asynchrone.....	50
2.6.1 Application du RDF dans le plan Doppler-retard.....	50
2.6.2 Discrimination optimale des classes par le critère de Fisher.....	51
2.6.2.1 Expression du Contraste de Fisher de la troisième variante.....	53
2.6.2.2 Conception des noyaux discriminants de Fisher.....	54
2.6.2.3 Résultats de calcul.....	57
2.6.3 Résultats expérimental.....	58
2.7 Conclusion	61

Chapitre III. Méthodes de la prise de décision

3.1 Introduction.....	63
3.2 Détermination d'une méthode de décision.....	63
3.3 Méthodes de prise de décision.....	63

3.3.1 Réseau de neurone.....	63
3.3.1.1 Mise en œuvre d'un modèle de réseaux de neurones.....	64
3.3.1.2 Apprentissage du réseau : Algorithme de rétropropagation	
de gradient.....	65
A. Mise en équation	65
B. Principe de la rétropropagation	65
C. Accélération de la rétropropagation.....	67
3.3.1.3 L'application de réseau de neurone pour la classification des défauts de	
la machine asynchrone.....	67
3.3.2 Règle des k – plus proches voisins (k – ppv).....	71
3.3.2.1 Choix de la distance dans la règle des k – ppv.....	71
3.3.2.2 Choix du nombre de plus proches voisins k	72
3.3.2.3 Règle des k – plus proches voisins avec rejet.....	72
3.3.2.3.1 Rejet d'ambiguïté.....	75
3.3.2.3.2 Rejet de distance.....	75
3.3.2.4 La prise de décision par les k-ppv pour les nouvelles	
observations..	76
3.3.2.5 Classification non exclusive.....	81
3.3.2.5.1 Rejet d'ambiguïté.....	84
3.3.2.5.2 Rejet d'appartenance.....	85
3.3.2.5.3. La décision pour les nouvelle observation	86
3.4 Perspectives et conclusion.....	91
Conclusion générale.....	93
Bibliographie.....	95



Introduction générale

Introduction générale

Le diagnostic automatique consiste en la détection d'un changement anormal dans le comportement ou dans l'état d'un système et dans la localisation de sa cause. Les systèmes de diagnostic automatique peuvent être mis en œuvre en cours d'exploitation (en ligne) ou pendant les phases normales d'arrêt d'exploitation. En cours d'exploitation, le but est de garantir la sécurité et la continuité de service et d'enregistrer les événements utiles pour la maintenance curative ou le retour d'expérience. En arrêt d'exploitation (hors ligne), l'objectif est d'évaluer le degré de vieillissement de certains éléments, de décider d'une opération de maintenance préventive ou curative.

Le développement de système de diagnostic automatique paraît une solution un peu plus coûteuse à l'investissement, mais elle s'amortit sur le long terme. Les signaux mesurables tels que les courants, les tensions, la vitesse, les vibrations ou bien encore la température peuvent fournir des informations significatives sur les défauts et ainsi servir à déterminer un ensemble de paramètres représentant les signatures de défauts du moteur. A partir de ces paramètres, la mise en place de méthodes décisionnelles peut permettre de concevoir des systèmes de diagnostic performants.

La représentation temporelle classique du signal ne donne pas une bonne perception des composantes oscillantes multiples, tandis que la représentation fréquentielle (transformée de Fourier) ne permet pas la localisation temporelle de ces composantes. Ainsi, partant des propriétés de ces signaux et des limitations de la transformée de Fourier (TF), il est naturel de s'orienter vers un schéma d'analyse temps-fréquence. Afin de rendre cette approche automatique, il lui faut rajoutée des méthodes ou techniques de prise de décision.

C'est dans cette perspective que ce mémoire est orienté. Il s'intéresse au diagnostic automatique par classification des défauts de la machine asynchrone et plus particulièrement sur les méthodes de prise de décision.

Ce mémoire s'articule autour de trois chapitres :

Le premier chapitre présente la constitution de la machine asynchrone et aussi leurs différentes défaillances qui peuvent se produisent sur ce type de machine. Ce chapitre se décompose en deux partie : la première partie fait état des différentes origines (électriques,

mécaniques,...) des défaillances qui peuvent se produire sur la machine asynchrone; la deuxième partie traite des techniques qui permettent la mise en oeuvre d'une procédure de diagnostic. On s'est étalé sur deux approches principales, à savoir l'approche signal et l'approche système, l'approche privilégiée est détaillée dans le second chapitre.

Le deuxième chapitre propose une approche système qui consiste en l'automatisation de la procédure de détection des défauts à partir de la représentation Temps-fréquence; notre approche est basée sur la classification de signaux après apprentissage. La représentation temps-fréquence employée dans ce chapitre est dite "dépendante de la classe du signal (DCS)" dont le plan d'ambiguïté est lissé par un noyau conçu spécialement afin de réaliser une séparation maximale entre le défaut et la machine saine. Le contraste de Fisher est utilisé pour la séparation de classe. Des vecteurs formes représentatives des différents défauts sont extraits pour les besoins de la prise de décision du troisième chapitre.

Le troisième chapitre est dédié aux méthodes de prise décision. Deux méthodes sont traités à savoir les réseaux de neurone et les k-plus proche voisins. L'étude est plus concentré sur la méthode les k-plus proches voisins vu leurs simplicités et leurs performances.



Chapitre I

**La machine asynchrone et les
méthodes de diagnostic**

1.1 Introduction

Le diagnostic est une composante principale d'un module de supervision. Il consiste à déterminer à chaque instant le mode de fonctionnement dans lequel le système se trouve. Il s'appuie sur une connaissance a priori des modes de fonctionnement et sur une connaissance instantanée matérialisée par une nouvelle observation de l'état du système. Il existe plusieurs approches pour réaliser le diagnostic, le choix d'une approche est lié au mode de représentation de la connaissance.

La première partie de ce chapitre fait état des différentes origines (électriques, mécaniques,...) des défaillances qui peuvent se produire sur la machine asynchrone. Débute par un rappel sur la constitution des machines électriques, et plus particulièrement sur celle de la machine asynchrone à cage d'écureuil. Nous évoquerons essentiellement des défauts qui peuvent survenir tant au stator qu'au rotor et d'enroulements de la machine.

Dans la deuxième partie de ce chapitre, nous arborons sur les différentes approches de diagnostic est leurs domaines d'applications. A partir de ces considérations, nous proposons une classification non exhaustive des méthodes de diagnostic selon trois axes : les méthodes reposant sur une approche signal, les méthodes basées sur une approche système et les méthodes de prises de décision.

1.2 La machine asynchrone

Le moteur asynchrone est la machine la plus utilisée dans le domaine des puissances supérieures à quelques kilowatts car il présente de nombreux avantages tels que sa puissance massique, sa robustesse, sa facilité de mise en oeuvre, son faible coût, etc.... L'apparition dans les années 1980 des variateurs permettant de faire varier la fréquence de rotation dans une large gamme a grandement favorisé son développement. En effet, il entre dans la conception de nombreux procédés industriels associant des convertisseurs statiques et des machines électriques (traction électrique, laminoirs, levage, pompage, etc....).

Bien que la machine asynchrone ait la réputation d'être robuste, elle peut présenter comme toute autre machine électrique, des défaillances d'ordre électrique ou mécanique. Ainsi, en raison des conséquences importantes et coûteuses que peut engendrer l'apparition d'un défaut sur les processus industriels, le diagnostic des défauts fait l'objet d'un engouement prononcé depuis les deux dernières décennies.

1.2.1 Constitution de la machine asynchrone

La machine asynchrone, souvent appelée moteur à induction comprend un stator et un rotor, constitués de tôles d'acier au silicium et comportant des encoches dans lesquelles on place les enroulements. Le stator est fixe ; on y trouve les enroulements reliés à la source. Le rotor est monté sur un axe de rotation. Selon que les enroulements du rotor sont accessibles de l'extérieure ou sont fermés sur eux mêmes en permanence, on définit deux types de rotor : bobiné ou à cage d'écureuil. Toutefois, nous admettrons que sa structure est électriquement équivalente à celle d'un rotor bobiné dont les enroulements sont en court-circuit. Dans ce travail, nous nous intéressant à la machine asynchrone à cage d'écureuil.

Les éléments de constitution d'une machine asynchrone à cage d'écureuil sont illustrés dans la Figure 1.1.

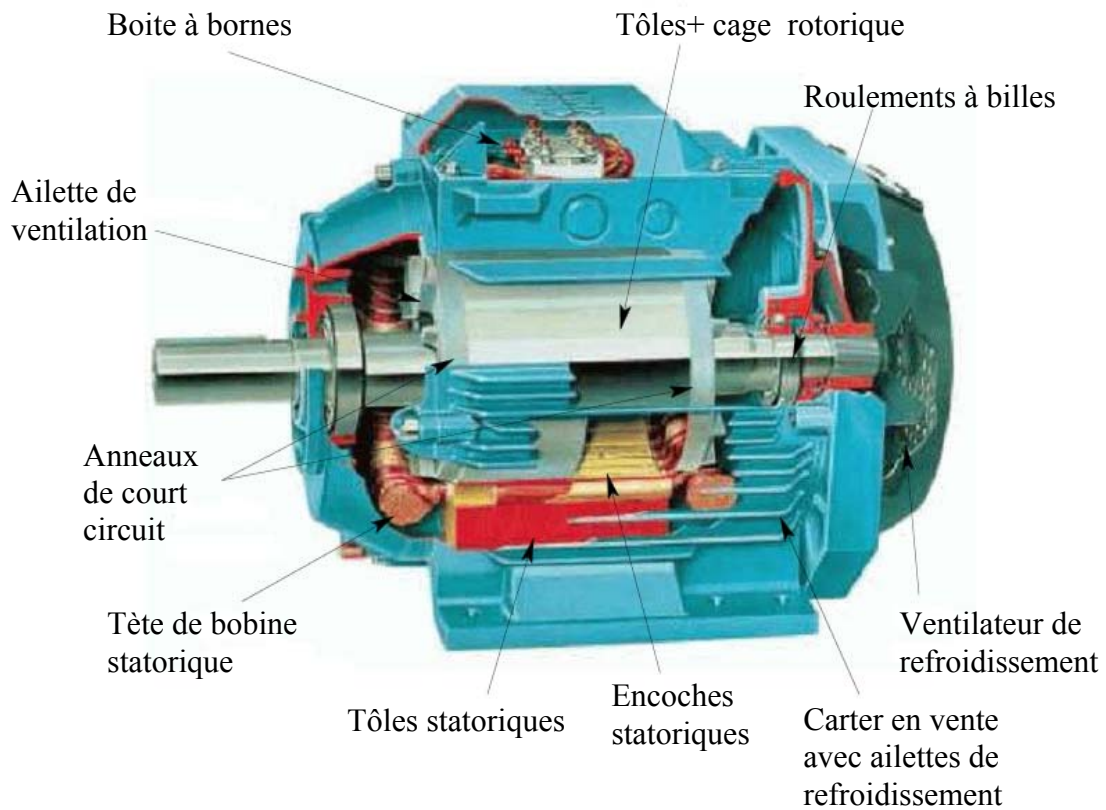


Figure 1.1. Eléments de constitution d'une machine asynchrone à cage d'écureuil

1.2.1.1 Le stator

Il est constitué d'un enroulement bobiné réparti dans les encoches du circuit magnétique. Ce circuit magnétique est constitué d'un empilage de tôles dans lesquelles sont découpées des encoches parallèles à l'axe de la machine (figure 1.2). Le bobinage statorique peut se décomposer en deux parties : les conducteurs d'encoches et les têtes de bobines.

Les conducteurs d'encoches permettent de créer dans l'entrefer le champ magnétique à l'origine de la conversion électromagnétique. Les têtes de bobines permettent, quant à elles, la fermeture des courants en organisant la circulation judicieuse des courants d'un conducteur d'encoche à l'autre. L'objectif est d'obtenir à la surface de l'entrefer une distribution de courant la plus sinusoïdale possible, afin de limiter les ondulations du couple électromagnétique.

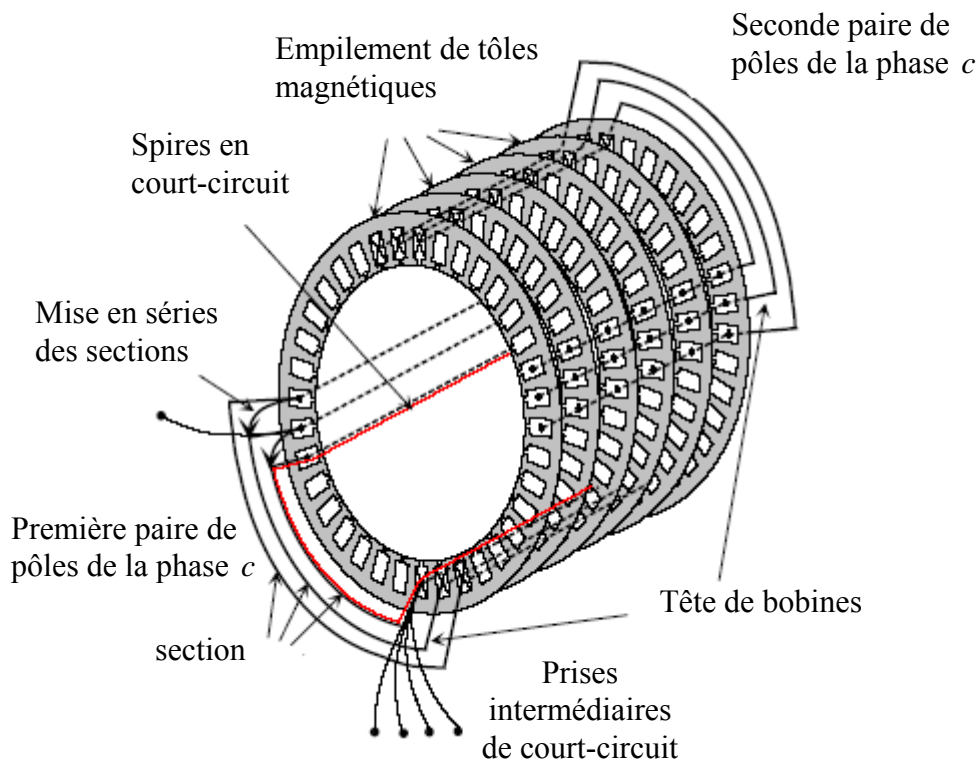


Figure 1.2. *vue schématique en perspective du stator(circuit magnétique, conducteurs d'encoches, tête de bobines).(enroulement statorique d'une machine a 4 pôles).[SCH99]*

1.2.1.2 Le rotor à cage

Dans le rotor à cage, les anneaux de court-circuit permettent la circulation des courants d'un conducteur d'encoche (barre rotoriques) à l'autre. Ces barres conductrices sont régulièrement réparties, et constituent le circuit du rotor (figure 1.3). Cette cage est insérée à l'intérieur d'un circuit magnétique constitué de disques en tôles empilés sur l'arbre de la machine analogue à celui du moteur à rotor bobiné.

Dans le cas de rotors à cage d'écureuil, les conducteurs sont réalisés par coulage d'un alliage d'aluminium, ou par des barres massives de cuivre préformées et frettées dans les tôles du rotor. Il n'y a généralement pas, ou très peu, d'isolation entre les barres rotoriques et les tôles magnétiques, mais leur résistance est suffisamment faible pour que les courants de fuite dans les tôles soient négligeables, sauf lorsqu'il y a une rupture de barre [Ker 84]. Le moteur à cage d'écureuil est beaucoup plus simple à construire que le moteur à rotor bobiné et, de ce fait, son prix de revient est inférieur. De plus, il dispose d'une plus grande robustesse. Il constitue la plus grande partie du parc de moteurs asynchrones actuellement en service.

Son inconvénient majeur est qu'il a, au démarrage, de mauvaises performances (courant élevé et faible couple). C'est pour remédier à cette situation qu'ont été développés deux autres types de cages (rotor à double cage et rotor à encoches profondes).

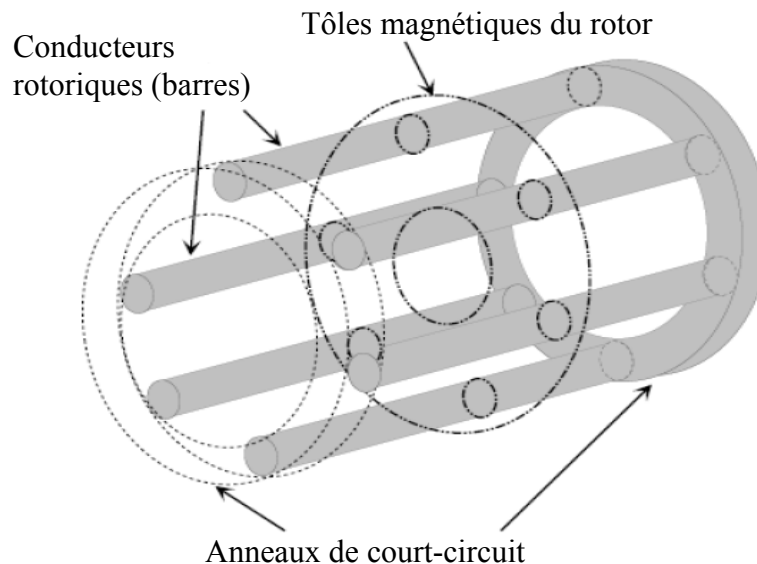


Figure 1.3 vue schématique en perspective du rotor (tôles magnétiques, conductrices d'encoches (barres) et anneaux de court-circuit [SCH99].

1.2.1.3 Les organes mécaniques

La carcasse sert de support, elle joue le rôle d'enveloppe et assure la protection contre l'environnement extérieur. L'arbre est un organe de transmission. Il comprend une partie centrale qui sert de support au corps du rotor et un bout d'arbre sur lequel est fixé un demi accouplement. Il est généralement constitué en acier moulé ou forgé. Son dimensionnement est fonction des efforts de flexion (force centrifuge qui s'exerce sur lui, attraction magnétique radiale, etc...), des efforts radiaux et tangentiels dus aux forces centrifuges, des efforts de torsion (couple électromagnétique transmis en régime permanent, transitoire). Il est supporté par un ou plusieurs paliers. Ces paliers soutiennent le rotor et assurent la libre rotation. Le second palier est libre pour assurer les dilatations thermiques de l'arbre. Une isolation électrique de l'un des paliers assure l'élimination des courants dans l'arbre dû aux dissymétries des réluctances du circuit magnétique. Ils sont généralement à roulements pour les machines de petite et moyenne puissance.

Afin d'assurer la variation de vitesse des moteurs électriques mais aussi la conversion de l'énergie, de nombreuses machines se voient associer un convertisseur. Ainsi, de nos jours, on trouve dans un grand nombre de procédés industriels, des associations entre des convertisseurs statiques et des machines électriques (pompage, levage,...). Cela permet de fournir aux moteurs des tensions et des courants de fréquence variable servant à commander la machine en couple ou en vitesse.

1.2.2 Etude statistique des défaillances de la machine asynchrone

Les défaillances peuvent être d'origines diverses, électriques, mécaniques ou bien encore magnétiques. Leurs causes sont multiples et peuvent se classer en trois groupes [O'Do 85] :

- les générateurs de pannes ou initiateurs de défauts : surchauffe du moteur, défaut électrique (court-circuit), survoltage d'alimentation, problème d'isolation électrique, usure des éléments mécaniques (roulements à billes), rupture de fixations, etc.
- les amplificateurs de défauts : surcharge fréquente, vibrations mécaniques, environnement humide, échauffement permanent, mauvais graissage, vieillissement, etc.
- les vices de fabrication et les erreurs humaines : défauts de fabrication, composants défectueux, protections inadaptées, mauvais dimensionnement de la machine, etc.

Une étude statistique, effectuée en 1988 par une compagnie d'assurance allemande de systèmes industriels [All 88] sur les pannes des machines asynchrones de moyenne puissance (de 50 kW à 200kW) a donné les résultats suivants (figure 1.4) :

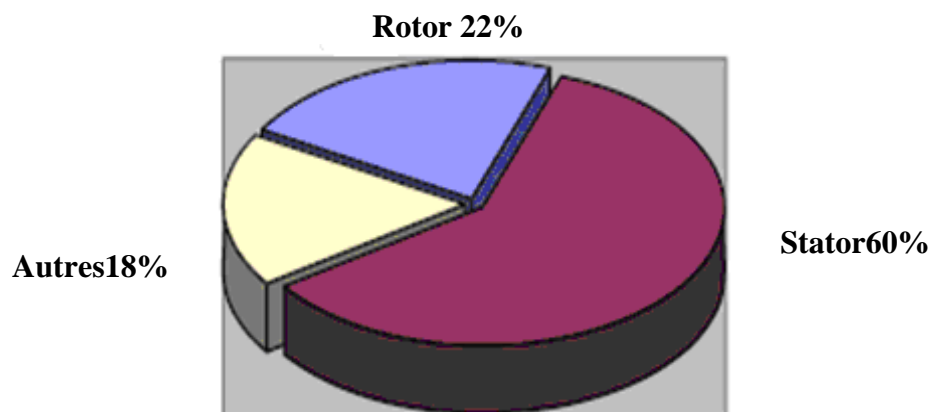


Figure 1.4. Proportion des défauts

La répartition des pannes dans les différentes parties du moteur est présentée sur la figure 1.5:

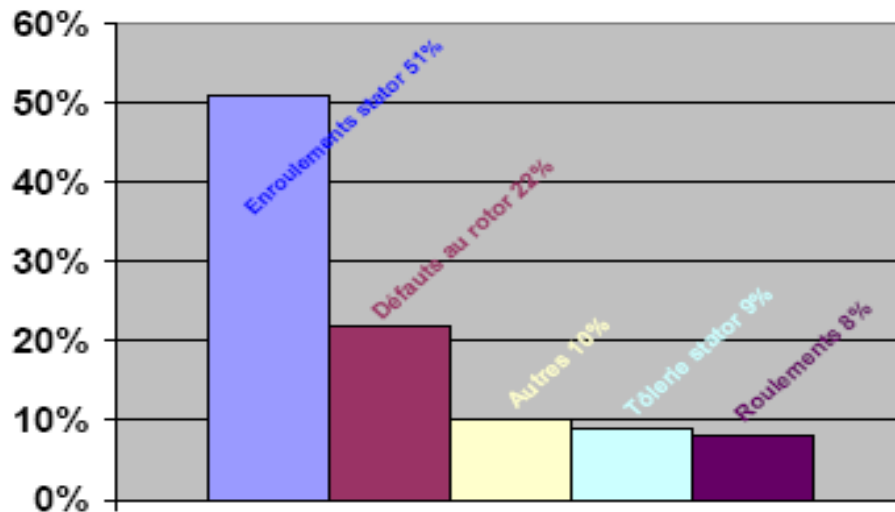


Figure 1.5. Répartition des pannes sur les machines de faibles et moyennes puissances

Une autre étude statistique [Tho 95] faite sur des machines de grande puissance (de 100 kW à 1 MW) donne des résultats qui sont présentés sur la figure 1.6.

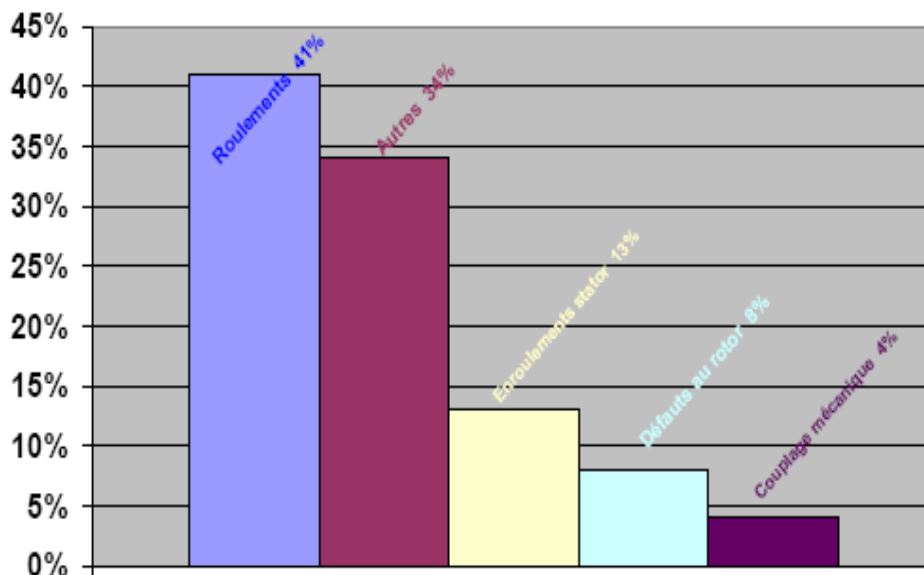


Figure 1.6. Répartition des pannes sur les machines de fortes puissances

Les contraintes mécaniques sont plus grandes pour ces types de machines ce qui explique le taux élevé des pannes dues aux roulements. Celles-ci exigent une maintenance mécanique accrue.

1.2.2.1 Défaillances au stator

Pour le stator, les défaillances sont principalement dues à un problème :

- thermique (surcharge,...)
- électrique (diélectrique,...)
- mécanique (bobinage,...)
- environnemental (agression,...)

Les défauts qui sont les plus récurrents, localisés au niveau du stator, peuvent être définis comme suit :

A. Défauts d'isolant dans un enroulement

La dégradation des isolants dans les enroulements peut provoquer des courts-circuits. En effet, les différentes pertes (Joule, fer, mécanique,...) engendrent des phénomènes thermiques se traduisant par une augmentation de la température des différents constituants du moteur. Or les matériaux d'isolation ont une limite de température, de tension et mécanique. De ce fait, si l'environnement de travail d'un matériau d'isolation dépasse une de ces limites, ce matériau se dégrade de manière prématurée ou accélérée, puis finit par ne plus assurer sa fonction. Les différentes causes pour ce type de défaut sont :

- dégradation de l'isolant à la fabrication.
- tension de l'enroulement supérieure à la limite du matériau d'isolation.
- vibrations mécaniques.
- courant élevé dans l'enroulement dû à un court-circuit, un défaut du convertisseur, une surcharge. Ceci entraîne une élévation de la température dégradant prématurément le matériau d'isolation.
- vieillissement naturel des isolants. Tous les matériaux isolants ont une durée de vie limitée. Même dans une utilisation 'normale', l'isolant finit naturellement par se dégrader.
- fonctionnement dans un environnement sévère.

B. Court-circuit entre spires

Un court-circuit entre spires de la même phase est un défaut assez fréquent. Cette défaillance a pour origine un ou plusieurs défauts d'isolant dans l'enroulement concerné. Il entraîne une augmentation des courants statoriques dans la phase affectée, une légère variation de l'amplitude sur les autres phases, modifie le facteur de puissance et amplifie les courants dans le circuit rotorique. Ceci a pour conséquence une augmentation de la température au niveau du bobinage et, de ce fait, une dégradation accélérée des isolants, pouvant provoquer ainsi, un défaut en chaîne (apparition d'un 2ème court-circuit).

Par contre, le couple électromagnétique moyen délivré par la machine reste sensiblement identique hormis une augmentation des oscillations proportionnelle au défaut.

C. Court-circuit entre phases

Ce type de défaillance peut arriver en tout point du bobinage, cependant les répercussions ne seront pas les mêmes selon la localisation. Cette caractéristique rend difficile une analyse de l'incidence de ce défaut sur le système.

L'apparition d'un court-circuit proche de l'alimentation entre phases, induirait des courants très élevés qui conduiraient à la fusion des conducteurs d'alimentation et/ou à la disjonction par les protections. D'autre part, un court-circuit proche du neutre entre deux phases engendre un déséquilibre sans provoquer la fusion des conducteurs.

Les courants statoriques sont totalement déséquilibrés et ce déséquilibre est proportionnel au défaut qui apparaît. Les courants dans les barres ainsi que dans les anneaux sont augmentés lors de l'apparition de ce défaut. La détection de ce type de défaut peut reposer sur le déséquilibre des courants de phases.

D. Court-circuit phase/bâti

Le bâti a généralement un potentiel flottant, mais pour des raisons de liaisons mécaniques, il est souvent relié à la masse. Si le potentiel est flottant, un court-circuit entre l'enroulement et le bâti n'a pas d'importance du point de vue matériel, excepté les effets capacitifs, le bâti prend alors le potentiel de l'enroulement à l'endroit du court-circuit.

Par contre, au niveau de la sécurité des personnes, ce type de défaut peut être très dangereux et il est alors nécessaire de mettre en place des dispositifs de protection (disjoncteurs différentiels). En présence de ce type de défaillance, la tension de la phase concernée ne change pas.

Cependant le courant circulant dans cette phase augmente avec la réduction de la résistance et de l'inductance. Cette augmentation du courant se traduit par une augmentation de la température pouvant entraîner des défauts d'isolant dans l'enroulement. De plus, cette défaillance va générer une composante homopolaire entraînant l'apparition d'un couple pulsatoire. Une mesure du courant de fuite pourrait permettre de détecter ce type de défaut.

E. Défaits de circuit magnétique

Ces défauts aboutissent dans la plupart des cas à une dissymétrie au niveau du fonctionnement de la machine, qui à son tour peut accentuer le problème par des phénomènes de surchauffe, de surtension, d'élévation importante du courant, etc.

1.2.2.2 Défaillances au rotor

Pour le rotor, les défaillances sont essentiellement dues à un problème :

- thermique (surcharge,...)
- électromagnétique (force en $B^2(t)$,...)
- résiduel (déformation,...)
- dynamique (arbre de transmission,...)
- mécanique (roulement,...)
- environnemental (agression,...)

Les défauts qui sont les plus récurrents, localisés au niveau du rotor, peuvent être définis comme suit :

A. Ruptures de barres [Nai 00], [Elk 92]

La cassure ou rupture de barre est un des défauts les plus fréquents au rotor. Elle peut se situer soit au niveau de son encoche soit à l'extrémité qui la relie à l'anneau rotorique. La détérioration des barres réduit la valeur moyenne du couple électromagnétique et augmente l'amplitude des oscillations, qui elles-mêmes provoquent des oscillations de la vitesse de rotation, ce qui engendre des vibrations mécaniques et donc, un fonctionnement anormal de la machine. La grande amplitude de ces oscillations accélère la détérioration de la machine.

Ainsi, le couple diminue sensiblement avec le nombre de barres cassées induisant un effet cumulatif de la défaillance. L'effet d'une cassure de barre croît rapidement avec le nombre de barres cassées.

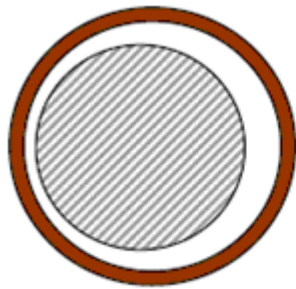
B. Ruptures d'anneaux

La rupture de portion d'anneau est un défaut qui apparaît aussi fréquemment que la cassure de barres. Ces ruptures sont dues soit à des bulles de coulées ou aux dilatations différentielles entre les barres et les anneaux. Comme il est difficile de le détecter [Yah 95], ce défaut est généralement groupé, voir confondu, avec la rupture de barres dans les études statistiques. Ces portions d'anneaux de court-circuit véhiculent des courants plus importants que ceux des barres rotoriques. De ce fait, un mauvais dimensionnement des anneaux, une détérioration des conditions de fonctionnement (température, humidité,...) ou une surcharge de couple et donc de courants, peuvent entraîner leur cassure.

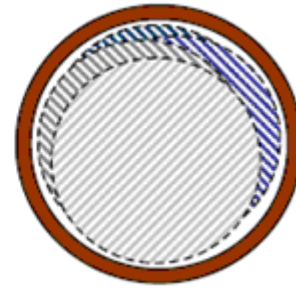
La rupture d'une portion d'anneau déséquilibre la répartition des courants dans les barres rotoriques et de ce fait, engendre un effet de modulation d'amplitude sur les courants statoriques similaire à celui provoqué par la cassure de barres.

C. Excentricité statique et dynamique

Parfois, la machine électrique peut être soumise à un décentrement du rotor, se traduisant par des oscillations de couple (décalage entre le centre de rotation de l'arbre et le centre du rotor, figure 1.8).



Excentricité statique



Excentricité dynamique (plusieurs positions du rotor au cours de la rotation)

Figure 1.8. Modélisation schématique de l'excentricité statique et dynamique

Ce phénomène est appelé excentricité (statique et dynamique) dont l'origine peut être liée à un positionnement incorrect des paliers lors de l'assemblage, à un défaut roulement (usure), à un défaut de charge, ou à un défaut de fabrication (usinage).

Trois cas d'excentricité [Sob 98], [Dev 01] sont généralement distingués :

- l'excentricité statique, le rotor est déplacé du centre de l'alésage stator mais tourne toujours autour de son axe.
- l'excentricité dynamique, le rotor est positionné au centre de l'alésage mais ne tourne plus autour de son axe.
- l'excentricité qu'on pourrait qualifier de 'mixte', associant les deux cas précédemment cités.

Ce défaut modifie le comportement magnétique ainsi que mécanique de la machine. En effet, l'augmentation de l'excentricité dans l'entrefer induit une augmentation des forces électromagnétiques qui agissent directement sur le noyau statorique ainsi que l'enroulement correspondant, ce qui engendre une dégradation de son isolation. D'autre part, cette augmentation peut avoir comme conséquence des frottements entre le stator et le rotor en raison des forces d'attraction magnétique qui déséquilibrent le système. Ceci donne naissance à des niveaux de vibration considérables dans les enroulements.

1.2.2.3 Défaillances mécaniques

A. Défauts roulements

Les roulements à billes jouent le rôle d'interface électromécanique entre le stator et le rotor. En outre, ils représentent l'élément de maintien de l'axe de la machine permettant d'assurer une bonne rotation du rotor. Dans l'article [Bon 93], l'auteur présente la plupart des défauts survenant dans les roulements des moteurs à induction ainsi que les raisons de leur vieillissement. Comme il a été présenté précédemment, ce type de défaut est le plus fréquent sur les machines de fortes puissances. Il est généralement lié à l'usure du roulement et plus précisément une dégradation des billes, ou de la bande de roulement. Ses causes possibles :

- l'usure due au vieillissement
- la température de fonctionnement élevée
- la perte de lubrification
- l'huile contaminée (par des paillettes métalliques issues de la dégradation des billes ou de la bande de roulement)
- le défaut de montage
- les courants d'arbres (Shaft Current)

Les conséquences directes de cette défaillance sur les roulements sont :

- des trous dans les gorges de roulement intérieures et extérieures
- l'ondulation de leur surface de roulement
- l'attaque des billes
- la corrosion due à l'eau
- défaut de graissage, problème dû à la température
- décollement, effritement de surface provoqué par une surcharge

Sur le système, ce type de défaut se traduit par des oscillations du couple de charge, une apparition de pertes supplémentaires et un jeu entre la bague interne et la bague externe du roulement entraînant des vibrations par les déplacements du rotor autour de l'axe longitudinale de la machine. Dans le cas le plus défavorable, la présence d'un roulement défectueux peut amener au blocage du moteur.

B. Autres défaillances mécaniques

Au stator, il n'y a pas de pièces mobiles donc a priori pas de défaillances mécaniques. Cependant, il peut apparaître des phénomènes d'oxydation liés à l'environnement de la machine et plus précisément au taux de salinité qui influe sur l'étanchéité et les contacteurs.

1.3 Méthodes de surveillance et de diagnostic de la machine asynchrone

La maintenance nécessite de connaître les régimes de fonctionnement des systèmes alors que la surveillance donne une réponse aux besoins d'information de la maintenance dans le but de garantir le bon fonctionnement des systèmes. Le diagnostic s'intègre à la surveillance en ayant pour objectif de déterminer les causes de dysfonctionnement.

La surveillance et le diagnostic des défauts nécessitent d'obtenir des symptômes caractéristiques du fonctionnement de l'entraînement surveillé et de les analyser pour en déduire l'état du système. L'établissement des symptômes se fait souvent en référence à la connaissance du comportement sain dont on dispose.

La forme et la méthode de génération des symptômes sont tributaires de la forme prise par cette connaissance. Par exemple si le fonctionnement défaillant est mis en évidence par l'apparition des nouvelles raies spectrales dans une grandeur électrique, une analyse en traitement de signal permettra une acquisition des symptômes. Le mode de traitement des symptômes dépend de leur type mais aussi de la forme prise de la connaissance.

La classification des méthodes de surveillance et le diagnostic se fait généralement en fonction de la forme prise par la connaissance et bien entendu, en fonction de la méthode d'analyse qui en découle.

Traditionnellement, les symptômes du défaut sont recherchés dans les signaux mesurés: ce sont les méthodes reposant sur une approche signal. Elles traitent en grande partie des informations issues de l'expérience. Les méthodes de surveillance et de diagnostic mises en œuvre s'orientent donc naturellement sur l'analyse du signal (sans ajout de capteur).

Une autre approche, appelée « système » basée sur des méthodes heuristiques, issues de l'expérience et de l'expertise. Des techniques d'aide à la génération et la formalisation des symptômes (mesures) ont permis de fournir une aide à l'expertise en quantifiant les symptômes (analyse de signaux électrique, vibratoire, thermique) mais l'analyse reste une affaire de règles et de classification (reconnaissance de forme, système expert.). Ces méthodes de surveillance nécessitent une bonne connaissance du système et de ces composants

1.3.1 Méthodes basées sur une approche système

Les méthodes dites « système » repose sur une connaissance du comportement du système sain et défaillant c'est-à-dire l'effet d'un défaut sur les variables mesurées.

Ces méthodes exploitent les paramètres issus des méthodes dites « signal ». Ces paramètres constituent alors une signature caractéristique du défaut. Elles nécessitent une base d'apprentissage intégrant des signatures du système dans ses différents modes de fonctionnement. Un nouveau point de fonctionnement est comparé avec les points de la base de fonctionnement.

Plusieurs méthodes ont été développées dans la littérature basée sur les réseaux de neurones [Rai 00], la reconnaissance de formes (RDF) [Cas 03], [Ond 06], le neuro-floue, les algorithmes génétiques [Der 02], les représentations temps fréquence (RTF). Toutes ces méthodes tendent à automatiser le processus de diagnostic depuis l'acquisition des données sur la machine jusqu'à la prise de décision sans l'intervention d'un expert. Dans cette partie, nous allons présenter quelques unes de ces techniques ainsi que leur application dans le domaine du diagnostic de la machine électrique.

1.3.1.1 Réseaux de neurones

Le principe des réseaux de neurones s'inspire des neurones biologiques. le modèle du neurone formel se comporte comme un opérateur effectuant une somme pondérée de ses entrées, suivie d'une non linéarité, appelée fonction d'activation, responsable du comportement du neurone.

Un réseau de neurones comprend un certain nombre de couches de neurones interconnectés. Couches d'entrées, couches cachées et couches de sortie. L'entrée représentée par les nœuds de sources fournit des poids d'accès physiques pour l'application des signaux d'entrées. Les neurones dans la couche cachée, se comportent en détecteurs de caractéristique, tandis que, les neurones dans la couche de sortie présentent à l'utilisateur les conclusions obtenues par le réseau, ou la réponse aux signaux d'entrées (figure 1.9). Une loi d'apprentissage permet d'ajuster les poids des connexions afin d'avoir les performances désirées.

L'initialisation du réseau de neurones est un processus itératif d'ajustement des poids et des biais de manière à minimiser un critère d'erreur. Elle constitue la phase d'apprentissage.

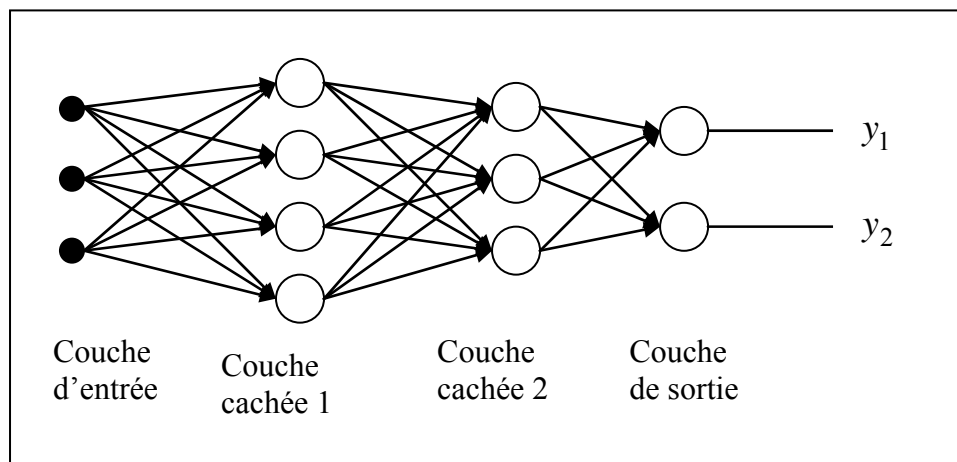


figure 1.9 la structure générale de réseau de neurone

Les avantages des réseaux de neurones

- la rapidité, très utile lorsque le diagnostic doit être conduit en ligne.
- un réseau de neurone est robuste surtout vis-à-vis du bruit.

Toutefois un réseau de neurone présente des inconvénients [Rai 00]

- l'apprentissage peut être long et difficile.
- l'apprentissage est gourmand en temps de calcul. Par ailleurs, il doit être réalisé sur toutes les données à la fois, avec le risque que le réseau oublie les résultats précédents.
- il est nécessaire de posséder des jeux de données sur toutes les conditions de fonctionnement et en particulier pour les fonctionnements défaillants.

- un réseau de neurones ne fonctionnera pas forcément correctement hors de sa plage d'apprentissage.
- une fois entraîné, un réseau de neurone n'est pas flexible. Si de nouvelles données interviennent, il faudra recourir à un nouvel apprentissage. Cependant celui-ci peut être effectué en ligne.

L'intérêt des réseaux de neurones dans le domaine de diagnostic se résume en deux points :

- le premier point est la faculté de réseau de neurones d'être utilisé en tant que règle de décision dans un processus d'automatisation de l'opération du diagnostic,
- le second point est la faculté d'apprentissage et de mémorisation d'un grand volume d'information.

D'ailleurs, en ce qui concerne le diagnostic des machines électriques, cette méthode a été largement utilisée. [Sit 04] a proposé un système d'identification et de classification des défauts de la machine asynchrone. Ce système est basé sur les réseaux de neurones à base radiale (RBF).

Le système développé par cet auteur, permet la détection des défauts électriques et mécaniques de la machine asynchrone. Quatre vecteurs caractéristiques ont été extraits à partir des spectres de puissance des signaux vibratoires de la machine. Ces vecteurs sont fournis à l'entrée du réseau de neurones. Le système proposé permet également d'estimer la sévérité des défauts détectés.

[Nej 00] expérimente la détection du déséquilibre de l'alimentation alors que [Sch 95] s'intéresse à l'excentricité. Dans ces deux cas, leurs essais prouvent l'intérêt des réseaux de neurones pour la classification et la prise de décision.

[Fil 95], a utilisé les réseaux de neurones, notamment pour la détection et l'estimation du nombre de barres cassées.

1.3.1.2 Reconnaissance de forme (RdF)

La Reconnaissance des Formes (RdF) repose sur le classement des objets ou formes en les comparant à des formes-types. De manière générale, deux types de RdF se distinguent :

- la RdF structurelle qui se base sur une représentation des formes à l'aide de grammaires [Fu 74]. Cette approche ne sera pas développée par la suite,
- la RdF statistique qui s'appuie sur une représentation numérique des formes [Dub 90].

L'auteur [Cas 03] a appliqué la RdF, pour le diagnostic des défauts de la machine asynchrone. Un vecteur de paramètres, appelé vecteur forme, a été extrait de chacune des mesures effectuées sur la machine. Les règles de décisions utilisées ont permis de classer les observations décrites par le vecteur forme, par rapport aux différents modes de fonctionnement connus avec ou sans défaut. Cette approche permet une classification avec un taux d'erreur presque nul. Toutefois les performances de cette méthode sont liées au choix du vecteur forme. [Ond 06] propose une méthode de sélection des paramètres basée sur un l'algorithme génétique. Il permet une réduction notable de la dimension de ce vecteur sans perte sensible d'information.

L'approche RdF est une méthode bien adaptée au domaine du diagnostic des défauts. Elle requiert une phase d'apprentissage qui peut être non supervisé. Mais ces méthodes requièrent des volumes de calcul important, parfois incompatible avec une exploitation temps réel.

Nous avons fait évoluer ces méthodes de classification dans un souci de réduction du temps de calcul et d'efficacité. Nous avons contribué au développement d'une nouvelle approche, système, exploitant une représentation Temps-fréquence et un critère de décision. Cette méthode est souple, facile à mettre en œuvre. Elle se base uniquement sur le traitement du courant de phase statorique. Par conséquent, cette méthode se prête bien pour un diagnostic en ligne.

1.3.1.3 Analyse Temps-fréquence : outil privilégié de détection

La représentation temporelle du signal ne donne pas une bonne perception des composantes oscillantes multiples, tandis que la représentation fréquentielle (transformée de Fourier) ne montre pas clairement la localisation temporelle de ces composantes. Il semble donc naturel de penser que ces limitations pourraient être surmontées par une analyse temps-fréquence où le signal est représenté comme une fonction conjointe du temps et de la fréquence- c'est-à-dire dans un plan « temps-fréquence » plutôt qu'une fonction du temps ou de la fréquence.

Une telle analyse devrait constituer un outil majeur pour la compréhension de nombreux processus et phénomène, dans des problèmes d'estimation, de détection ou de classification.

Cependant, Le temps-fréquence se trouve confronter au problème de la transformation mathématique qui permet de passer du signal analysé à sa représentation temps-fréquence. C'est une limitation fondamentale, connue sous le nom de principe d'incertitude d'Heisenberg, qui exclut toute localisation temporelle précise d'une fréquence.

Cette ambiguïté qui a donné lieu à la définition d'une grande diversité de transformation temps-fréquence.

Les représentations temps-fréquences (RTF), présentées dans ce contexte, sont issues de la classe de Cohen [Coh 95]. Elles décrivent le contenu énergétique pour une fréquence à un instant donné.

$$RTF_x^\phi(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\xi, \tau) A_x(\xi, \tau) e^{j2\pi(t\xi - f\tau)} d\xi d\tau \quad (1.1)$$

Les RTF basée sur la classe de Cohen sont caractérisées par un noyau Φ que l'on peut choisir pour une application particulière, ce qui permet une interprétation visuelle intuitive facilitant la conception de procédures de traitement automatique. De nombreux travaux s'appuient d'ailleurs sur cette qualité pour proposer des algorithmes de décision temps-fréquence [Fla 88], [Boa 90], [Hei 95], [Lem 95]

a) Représentation temps-fréquence à noyau paramétrique

La recherche en temps-fréquence a essentiellement portée sur la conception de fonction de lissage ou des noyaux Φ paramétrés qui permettent notamment la réduction maximale des interférences quadratiques [Coh 95], qui rendent difficile l'interprétation de l'image. Bien qu'une partie de ces RTF puisse offrir des avantages pour la classification de certains types de signaux [Wan 95],[Ben 98], la conception du système de détection ou de classification a rarement eu pour but la conception de noyau. Ces quelques méthodes qui proposent d'optimiser le noyau pour la classification contraignent la forme du noyau aux fonctions paramétriques prédéfinies avec les symétries qui peuvent ne pas être appropriées à la détection ou à la classification [Hei 95], [Dav 98].

b) Représentation temps-fréquence à noyau non paramétrique

Pour raisonner en termes de diagnostic et de classification, il n'est pas nécessairement souhaitable de représenter avec précision la distribution d'énergie d'un signal en temps-fréquence. En fait, une telle représentation peut être à l'encontre de l'objectif de la classification.

Il peut être avantageux de concevoir une RTF qui accentue spécifiquement les différences entre les classes (le terme « classe » est utilisé pour définir un groupe, arbitraire ou autre, des données "semblables").

Cette approche a été employée dans les domaines du radar, du sonar, de l'astronomie par radio, des communications et des systèmes optiques .

1.3.2 Méthodes de prises de décision (classification)

De manière générale, on peut diviser les méthodes de classification entre méthodes avec apprentissage supervisé et méthodes avec auto-apprentissage. Les méthodes avec apprentissage supervisé ont besoin de connaître *a priori* les classes associées aux données d'entraînement.

Dans tous les cas, un vecteur résumant l'information (individu) qui correspond aux échantillons du système, est analysé pour trouver les caractéristiques des classes. La différence entre les deux types de méthodes est la phase d'apprentissage.

Dans le contexte des méthodes basées sur la fouille de données, les méthodes qui se situent dans le "*Machine Learning*" permettent d'obtenir des règles ou des expressions de classification plus simples, de telle sorte que les résultats obtenus soient facilement interprétables. Un des objectifs de ces méthodes est que l'intervention humaine dans la phase d'entraînement soit minimale et qu'en même temps, les classes obtenues soient facilement interprétables.

1.3.2.1 Discriminant Linéaire – ACP

Il s'agit d'un algorithme standard et qui est le plus utilisé dans différents systèmes de diagnostic industriel.

Le but principal de l'ACP est donc de trouver un ensemble de facteurs (composantes) qui ait une dimension inférieure à celle de l'ensemble original de données et qui puisse décrire correctement les tendances principales. L'ACP est une procédure qui ne prend en compte que les variables du processus. Parfois, un ensemble additionnel de données est disponible.

Le discriminant linéaire [Fis 36] suppose une distribution normale des données avec des moyennes différentes. Ce type de distribution est justifié quand il y a une grande quantité de données. L'objectif est de trouver les hyperplans qui divisent l'espace des données. En principe, un hyperplan de dimension d (avec d égal au nombre de descripteurs) est tracé de telle manière qu'il passe par le milieu de la ligne qui unit les centres de chaque classe.

L'algorithme cherche à obtenir des fonctions discriminantes paramétriques (hyperplans), à partir de la minimisation de l'erreur quadratique de classification. Par conséquent, c'est une technique qui requiert de connaître *a priori* les classes de l'ensemble des données d'entraînement. Une fois la partition de l'espace obtenue, de nouvelles données peuvent être classées.

1.3.2.2 Les K plus proches Voisins (K- Nearest Neighbors)

C'est la première méthode de classification qui ne suppose pas un type de distribution des données et c'est la première à proposer une analyse de similitude entre des données en utilisant la distance entre elles [Fix 51]. L'algorithme fait un calcul de distance entre tous les individus et chaque objet est classé dans le groupe où se trouvent ses K plus proches voisins.

Les k plus proches voisins sont des individus de l'ensemble des données d'entraînement. Habituellement, la distance employée pour l'analyse est la distance euclidienne. Normalement le nombre des K voisins est un ou trois, mais cette valeur dépend de l'application [Dud 01].

1.3.2.3 Réseaux de Neurones

La structure générale d'un réseau de neurones utilisée pour le diagnostic de systèmes, est généralement constituée de trois couches : une couche d'entrée, une couche cachée et une couche de sortie (Figure 1.10).

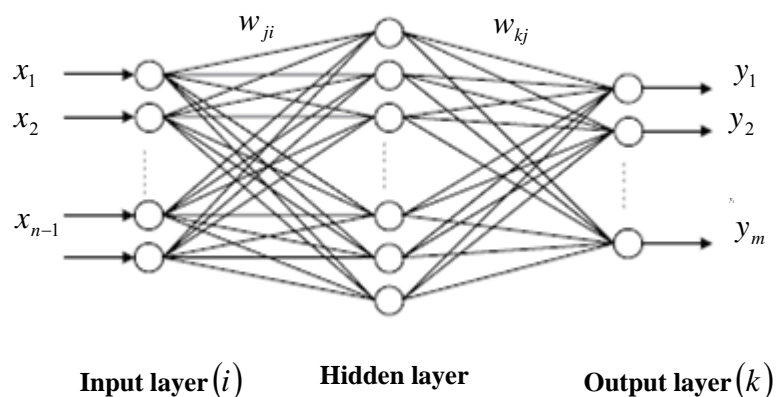


Figure I.10 les trios couches de RNA

Le nombre de neurones de la première couche correspond au nombre de descripteurs du problème (d=nombre de descripteurs de chaque individu x_n) et le nombre de neurones de la couche de sortie correspond aux K classes. Normalement, les performances du réseau sont directement liées à la taille de la couche cachée. L'équation générale d'un neurone est la suivante :

$$y_i(t) = \sum_{j=1}^N w_{ij} x_j(t) \dots\dots\dots(1-2)$$

$$u_i(t) = f(y_i(t))$$

$y_i(t)$ est l'état interne du neurone i et $u_i(t)$ est son état de sortie ou l'activation de sortie.

Les applications des RNA dans le cas du diagnostic automatique montrent que grâce à la capacité de généralisation de cette technique, les résultats sont satisfaisants. Les inconvénients avec les RNA sont le temps d'apprentissage et la difficulté, d'interprétation de la structure des classes, interprétation souvent nécessaire dans l'implémentation d'une technique de classification pour le diagnostic.

1.3.2.4 Méthodes de coalescence (clustering)

Ces algorithmes ont comme paramètre d'entrée le nombre de classes. Ils partagent l'ensemble de N objets en K groupes, ainsi la similarité à l'intérieur d'un même groupe est élevée mais faible entre les différentes classes. Pour ce faire, ces algorithmes itèrent en deux étapes, d'abord ils calculent les centres des groupes et deuxièmement ils assignent chaque objet au centre le plus proche.

Chaque classe est caractérisée par le centre ou prototype et par ses éléments. Le prototype des classes est le point dans l'espace de dimension d (d correspond au nombre de descripteurs) où la somme des distances à tous les objets d'un même groupe est minimale.

K-Means est sans nul doute l'algorithme le plus connu Parmi les différentes méthodes basées sur la distance. Il effectue une séparation stricte de l'espace (c'est-à-dire chaque objet n'est associé qu'à une seule classe). La mesure de similarité utilisée est la distance, pour le calcul plusieurs mesures peuvent être utilisées (e.g. euclidienne, Mahalanobis, etc.), par conséquent la forme géométrique change en fonction de la mesure employée (e.g. pour le cas euclidien la géométrie des groupes est une hypersphère) [Gen 07].

le Fuzzy K-Means (FKM) qui permet d'avoir un regroupement de type flou. FKM est un processus itératif qui permet d'obtenir comme résultat la matrice de partition floue ($U = [u_{kn}]$) laquelle contient les degrés d'appartenance de chaque objet x_n à chacune des K classes.

GK-Means [Gus 92] est Un autre algorithme, reconnu pour ses très bonnes performances. Cet algorithme utilise une norme adaptative de la distance, en obtenant des classes de géométrie hyper-ellipsoïdale avec des axes qui ne sont pas nécessairement parallèles aux axes de l'espace des données.

Ceci constitue un avantage face à l'algorithme FKM parce que le GK-Means permet à chaque groupe d'adapter sa norme de distance en fonction de la structure topologique locale des données. Chaque groupe a une matrice de norme induite A_k qui permet d'obtenir la norme de distance D_{nk}^2 propre à chaque classe.

1.4 Choix de la méthode de classification

Parmi les méthodes « système » exposées ci-dessus, notre choix s'est porté pour l'analyse temps fréquence à noyau non paramétrique en tant qu'outil de surveillance et de diagnostic. Celui-ci est justifié pour les raisons suivantes :

- la méthode est pratique à mettre en œuvre,
- elle est économe en temps de calcul,

1.4.1 Classification par la représentation temps-fréquence

Dans la classification traditionnelle, les données étaient souvent transformées en une Représentation Temps-Fréquence (RTF) standard (par exemple, le spectrogramme ou le Wigner-Ville TFR); puis, une projection était appliquée à la RTF pour aboutir à un espace de dimension réduite [Hay 97]. Le choix d'une RTF et d'un algorithme de projection doivent être conjointement optimisés. Pour diminuer le temps de calcul, il est plus pertinent d'utiliser une RTF optimisée par un noyau approprié et qui puisse être directement classifiée sans réduction de la dimension de la RTF.

Un signal peut être représenté par un nombre infini de RTF (figure 1.11 a). Cependant, une RTF optimale pour une tâche de classification s'avère nécessaire pour l'extraction des motifs. Il a été montré que, par exemple, le spectrogramme (une des RTF très utilisée), n'est pas optimal en terme de classification de signaux [Dav 00].

En effet, l'objectif d'un spectrogramme est de décrire la densité d'énergie d'un signal simultanément dans des domaines de temps et de fréquence alors que le but d'une classification RTF est de maximiser la séparabilité des signaux de différentes classes [Wan 04]. Par conséquent, il est souhaitable de concevoir une RTF qui permet une classification optimale c'est-à-dire une *RTF* (figure 1.11 b) qui maximise les distances entre classes.

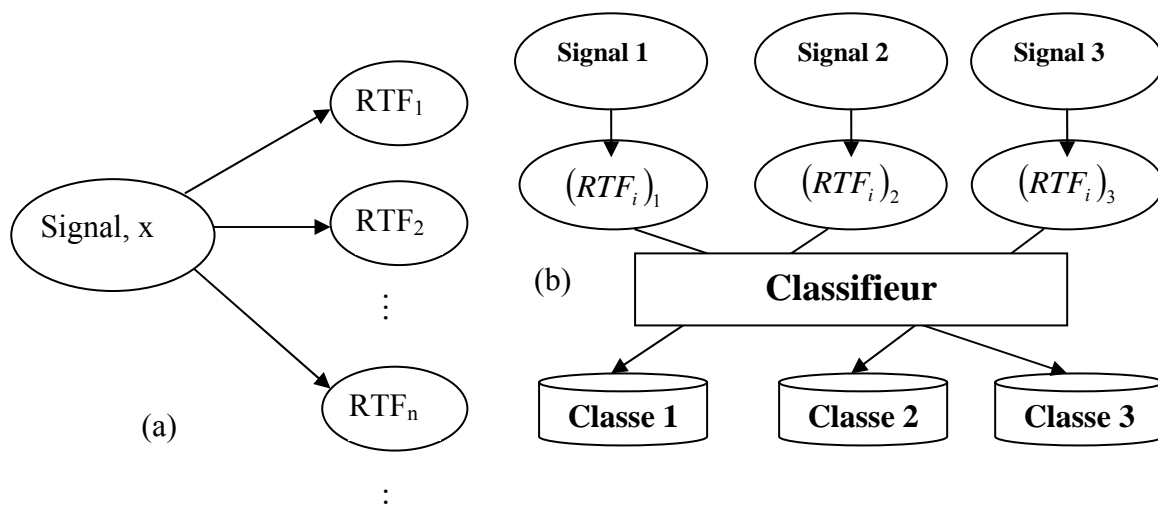


Figure 1.11 RTF orientée vers la classification

1.4.2 RTF dite "Dépendante de la Classe du Signal (DCS) "

La RTF dite "dépendante de la classe du signal " n'exige à priori aucune forme paramétrique et prédéfinie du noyau. La méthode proposée nécessite deux étapes principales : l'extraction de motif et la classification. L'extraction de motif saisit l'information pertinente des signaux à classer et fournit les motifs comme des entrées pour le système de classification.

Dans notre cas la classification repose sur le calcul des distances de Mahalanobis entre le vecteur motif à classer et les vecteurs motifs d'apprentissage des différentes classes.

Nous proposons une méthode pour concevoir des noyaux optimisés, et par conséquent des RTF optimisées pour la discrimination entre les ensembles prédéfinis de classes. Les noyaux résultants ne sont restreints à aucune fonction prédéfinie mais, plutôt, sont arbitraires dans leur forme. Au lieu de faire des hypothèses à priori au sujet du type de lissage temps fréquence, notre approche établit le lissage approprié pour réaliser la meilleure performance de la classification.

1.4.3 La décision

Automatiser implique comprendre, détecter, reconnaître ou classer, ce sont autant de formes courantes d'une problématique unique : la classification de signaux. En effet, la compréhension fait appel implicitement à un apprentissage et une comparaison [Dud 73], la détection d'un événement consiste à affecter le signal à la classe événement présent ou la classe absent [Van 68], etc. dans ce type de procédure, ce qui est attendu est une décision, une affectation à une classe.

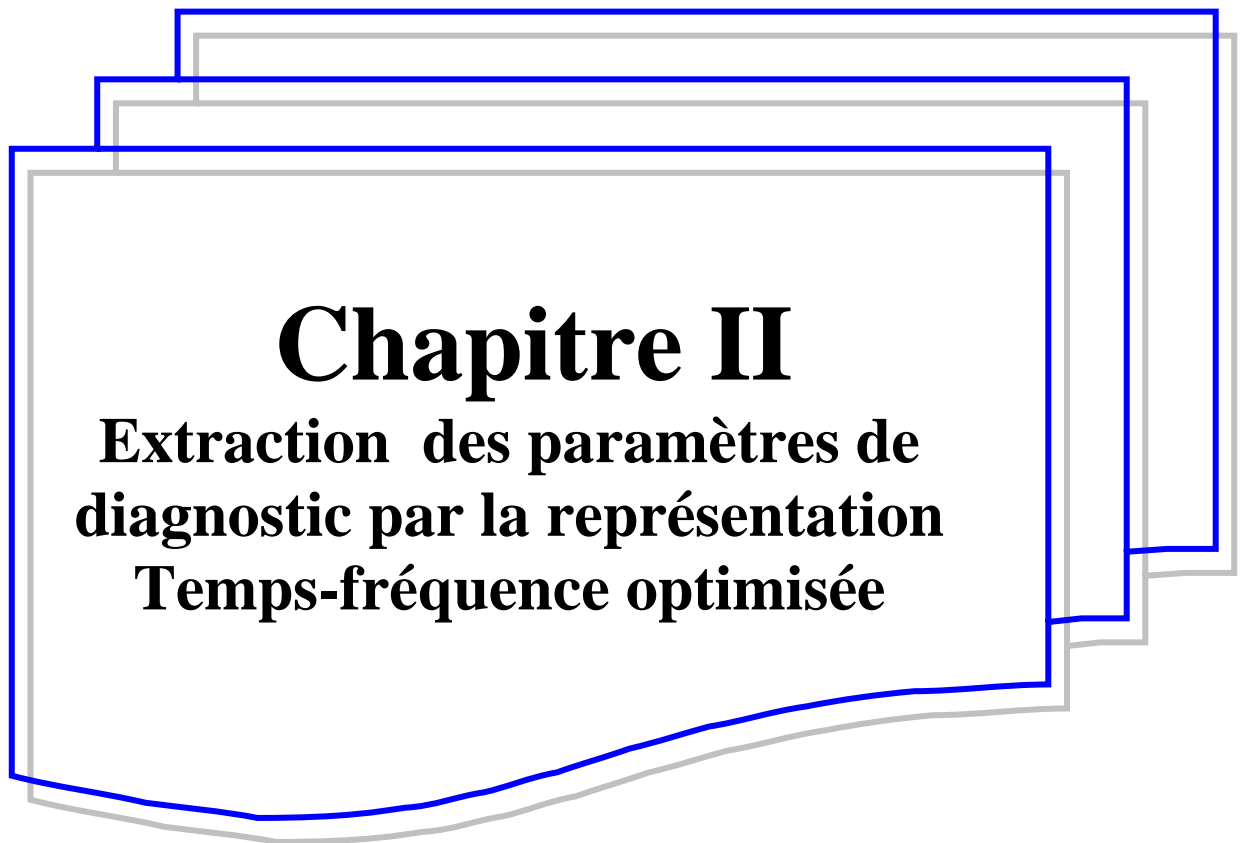
Une décision nécessite un classifieur. Celui-ci fait appel à des indices de divergences. Il existe plusieurs classifieurs. Nous traiterons particulièrement la classification par les réseaux de neurones et les k-plus proches voisin.

La première phase du processus de classification est une phase d'apprentissage à partir des observations effectuées sur des signaux dont la nature est connue. La seconde phase d'un processus de classification consiste à la prise de décision concernant l'affectation d'un signal non classé.

1.5 Conclusion

Ce chapitre est essentiellement consacré à l'état de l'art de la machine asynchrone et sur les méthodes de diagnostic. Nous avons scindé ce chapitre en deux parties: la première concerne la constitution de la machine asynchrone et leurs défaillances ; la deuxième partie traite des techniques qui permettent la mise en oeuvre d'une procédure de diagnostic. Les techniques et autres outils de diagnostic étant très nombreuses, nous les avons classées en deux approches: (i) une approche signal basée sur le traitement de signal des grandeurs mesurables, à savoir le courant et la tension, (ii) une approche système.

Parmi les méthodes « système » exposées ci-dessus, notre choix s'est porté pour l'analyse temps fréquence à noyau non paramétrique en tant qu'outil de surveillance et de diagnostic. Celui-ci est justifié pour les raisons suivantes : la méthode est pratique à mettre en oeuvre et elle est économe en temps de calcul.



Chapitre II

**Extraction des paramètres de
diagnostic par la représentation
Temps-fréquence optimisée**

2.1 Introduction

La représentation temporelle du signal ne donne pas une bonne perception des composantes oscillantes multiples, tandis que la représentation fréquentielle (transformée de Fourier) ne montre pas clairement la localisation temporelle de ces composantes. Il semble donc naturel de penser que ces limitations pourraient être surmontées par une analyse temps-fréquence où le signal est représenté comme une fonction conjointe du temps et de la fréquence- c'est-à-dire dans un plan « temps-fréquence » plutôt qu'une fonction du temps ou de la fréquence. Une telle analyse devrait constituer un outil majeur pour la compréhension de nombreux processus et phénomène, dans des problèmes d'estimation, de détection ou de classification par extraction des paramètres pertinents au diagnostic.

2.2 La Représentation temps-fréquence (RTF)

Il est aujourd'hui bien admis que les représentations d'un signal conjointement en temps et en fréquence offrent un réel intérêt : elles permettent une description des signaux non stationnaires, c'est à dire l'analyse des lois de comportement fréquentiel du signal au cours du temps. Suite aux travaux fondamentaux de Ville et de Wigner sur les RTF, le sujet est quasiment devenu une discipline scientifique a part entière [Boa 03],[Boa 92], [Che 02], [Coh 98], [Fla 93], [Hla 05], [Pap 02].

Par RTF, nous entendons la mise en correspondance d'une fonction mono-variable (en t) avec une fonction a deux variables : le temps t et la fréquence ν . Contrairement a la TF, qui donne le contenu fréquentiel global d'un signal, les RTF situent l'énergie dans le plan temps-fréquence [Fla 93], [Hla 05].

Parmi les RTF, la classe de Cohen tient une place particulière [Coh 95]: «[elle] peut être construite de manière objective et [elle] contient l'ensemble des représentations bilinéaires covariantes par translations dans le plan temps-fréquence » (extrait de P. Flandrin [Fla 93], page 108).

L'expression générique de la classe de Cohen peut s'écrire [Coh 95], [Fla 93], [Hla 05] :

$$C_x(t, \nu; \Pi) = \iint \Pi(u - t, \xi - \nu) w_x(u, \nu) du d\xi \dots\dots\dots(2.1)$$

ou $w_x(u, \nu)$ est la Distribution de Wigner-Ville (WVD):

$$w_x(t, \nu) = \int x\left(t + \frac{\tau}{2}\right) \cdot x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi\nu\tau} d\tau \dots\dots\dots(2.2)$$

Π est le noyau de la représentation temps-fréquence. Le choix de Π détermine les propriétés de $C_x(t, \nu, \Pi)$: conditions marginales, FI . . .

Ainsi, toute représentation de la classe de Cohen s'obtient par la double convolution temps-fréquence de la WVD qui apparaît comme la distribution quadratique centrale pour l'ensemble de la classe de Cohen.

Il existe de nombreuses distributions appartenant à la classe de Cohen comme par exemple le spectrogramme, la WVD (et ses extensions : PWVD et SPWVD), la distribution de Choi-Williams, celle de Born-Jordan, ou la B-distribution de Barkat-Boashash [Boa 03]...

Dans l'absolu, cette description simultanée en temps et en fréquence rend l'interprétation plus facile. Toutefois, de par leur bilinéarité, ces distributions engendrent des termes d'interférences pouvant nuire à la lisibilité des représentations obtenues. Ainsi, la WVD fournit d'excellents résultats pour des signaux mono-composante mais pour les signaux à composantes multiples, elle présente des interférences indésirables. Supposons que x_1 et x_2 soient deux composantes d'un seul signal x . La WVD est alors :

$$w_{x_1+x_2}(t, \nu) = w_{x_1}(t, \nu) + w_{x_2}(t, \nu) + 2\Re(w_{x_1, x_2}(t, \nu)) \dots\dots\dots(2.3)$$

avec :

$$w_{x_1, x_2}(t, \nu) = \int x_1\left(t + \frac{\tau}{2}\right) \cdot x_2^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi\nu\tau} d\tau \dots\dots\dots(2.4)$$

$\Re(w_{x_1, x_2}(t, \nu))$ présente une structure oscillante contrairement aux composantes propres du signal en général plus réguliers.

On peut montrer que les structures interférentielles sont placées à mi-distance des composantes x_1 et x_2 et oscillent suivant l'axe des temps et/ou des fréquences [Fla 93]. La règle du point milieu résume la contribution des interférences : deux points du plan interfèrent pour créer une contribution en un troisième, leur milieu géométrique.

En conséquence, lorsque le signal est « complexe », les interférences compliquent l'analyse et la réduction de celles-ci est réalisée de différentes manières [Bas 02], [Coh 95], [Fla 93].

L'une des solutions consiste à appliquer un lissage dans le plan temps-fréquence afin de les atténuer. Ainsi la distribution pseudo Wigner-Ville (PWVD) effectue un lissage en fréquence tandis que la pseudo Wigner-Ville lissée (SPWVD) applique un lissage bidimensionnel séparable en temps et en fréquence. Une autre solution, plus « géométrique », permettant d'améliorer la lisibilité des distributions de la classe de Cohen (et de la classe affine) est basée sur la méthode de la réallocation [Cha 98].

Le principe est de réarranger le plan temps-fréquence pour y améliorer la localisation des composantes d'un signal.

De manière générale, ces représentations, quoique différentes reposent sur un concept identique à savoir la re-écriture du noyau de lissage selon des critères prédéfinis à l'avance. Ces critères sont souvent choisis selon les caractéristiques connues des signaux à analyser ou de la problématique envisagée (détection [Lem 95], [Ric 98], classification [Dav 00] . . .).

Des approches ont été proposées pour choisir automatiquement, parmi un ensemble de RTF, celle qui décrit de manière optimale le signal selon un critère donné [Coh 95], [IOA 03] [Mal 00]. Pour illustrer certaines RTF, nous définissons un signal synthétique $x(n)$ comme étant la combinaison linéaire de quatre « atomes gaussiens élémentaires ». Un « atome gaussien élémentaire » x_{τ_i, ν_i} s'exprime comme le produit d'une fonction sinusoïdale modulée par une gaussienne :

$$x_{\tau_i, \nu_i}(n) = \cos(2\pi\nu_i(n - \tau_i)) \cdot \exp\left(-\pi\left(\frac{n - \tau_i}{2\sqrt{N}}\right)^2\right) \dots\dots\dots(2.5)$$

avec N le nombre de points.

2.2.1 Représentation temps-fréquence à noyau paramétrique

La recherche en temps-fréquence a essentiellement porté sur la conception de fonction de lissage ou des noyaux ϕ paramétrés qui permettent notamment la réduction maximale des interférences quadratiques [Coh 95], qui rendent difficile l'interprétation de l'image.

Bien qu'une partie de ces RTF puisse offrir des avantages pour la classification de certains types de signaux [Wan 95],[Ben 98], la conception du système de détection ou de classification a rarement eu pour but la conception de noyau. Ces quelques méthodes qui proposent d'optimiser le noyau pour la classification contraignent la forme du noyau aux fonctions paramétriques prédéfinies avec les symétries qui peuvent ne pas être appropriées à la détection ou à la classification [Hei 95], [Dav 98].

2.2.2 Représentation temps-fréquence à noyau non paramétrique

Pour raisonner en termes de diagnostic et de classification, il n'est pas nécessairement souhaitable de représenter avec précision la distribution d'énergie d'un signal en temps-fréquence. En fait, une telle représentation peut être à l'encontre de l'objectif de la classification.

Il peut être avantageux de concevoir une RTF qui accentue spécifiquement les différences entre les classes (le terme « classe » est utilisé pour définir un groupe, arbitraire ou autre, des données "semblables").

Cette approche a été employée dans les domaines du radar, du sonar, de l'astronomie par radio, des communications et des systèmes optiques .

2.3 Choix de la méthode de diagnostic

Parmi les méthodes « système » exposées ci-dessus, notre choix s'est porté pour l'analyse temps fréquence à noyau non paramétrique en tant qu'outil de surveillance et de diagnostic. Celui-ci est justifié pour les raisons suivantes :

- la méthode est pratique à mettre en œuvre,
- elle est économe en temps de calcul,

2.3.1 Représentation temps-fréquence et classification

Dans la classification traditionnelle, les données étaient souvent transformées en une Représentation Temps-fréquence (RTF) standard (par exemple, le spectrogramme ou le Wigner-Ville TFR); puis, une projection était appliquée à la RTF pour aboutir à un espace de dimension réduite [Hay 97].

Le choix d'une RTF et d'un algorithme de projection doivent être conjointement optimisés (figure 2.1 (a)). Pour diminuer le temps de calcul, il est plus pertinent d'utiliser une RTF optimisée par un noyau approprié et qui puisse être directement classifiée (figure 2.1 (b)) sans réduction de la dimension de la RTF.

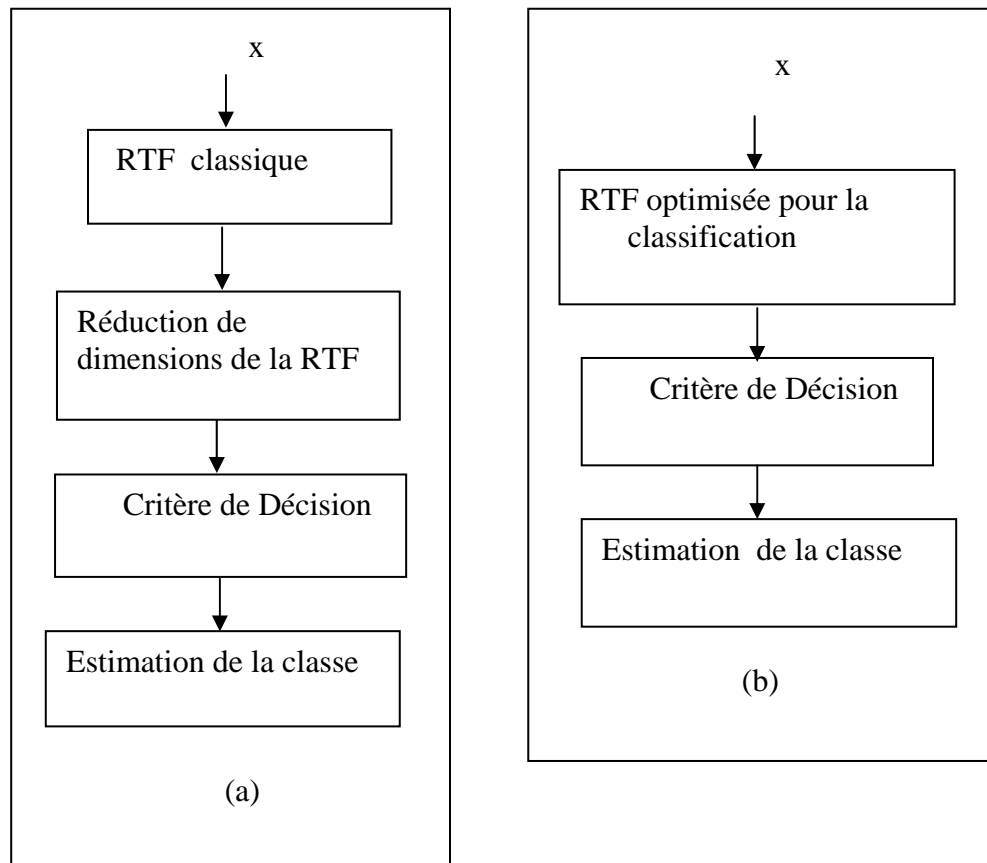


Figure 2.1 Procédure de classification par la RTF

Un signal peut être représenté par un nombre infini de RTF (figure 2.2(a)). Cependant, une RTF optimale pour une tâche de classification s'avère nécessaire pour l'extraction des motifs. Il a été montré que, par exemple, le spectrogramme (une des RTF très utilisée), n'est pas optimal en terme de classification de signaux [Dav 00]. En effet, l'objectif d'un spectrogramme est de décrire la densité d'énergie d'un signal simultanément dans des domaines de temps et de fréquence alors que le but d'une classification RTF est de maximiser la séparabilité des signaux de différentes classes [Wan 04]. Par conséquent, il est souhaitable de concevoir une RTF qui permet une classification optimale c'est-à-dire une RTFi (figure 2.2(b)) qui maximise les distances entre classes.

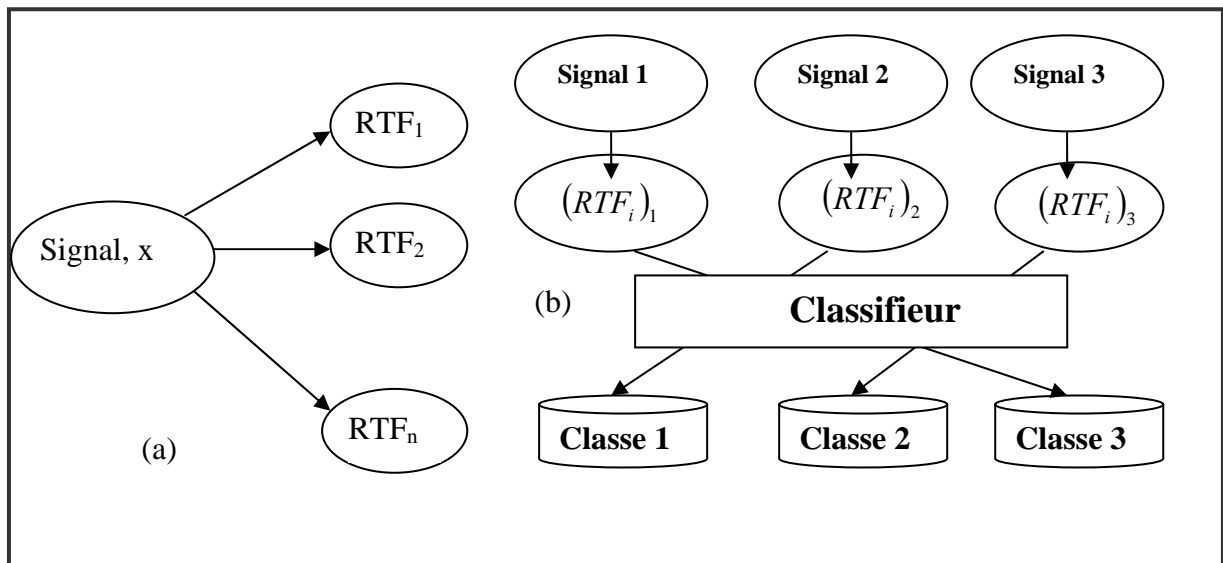


Figure 2.2 RTF orientée vers la classification

2.3.2 RTF dite "Dépendante de la Classe du Signal (DCS) "

La RTF dite "dépendante de la classe du signal " n'exige à priori aucune forme paramétrique et prédéfinie du noyau. La méthode proposée nécessite deux étapes principales : l'extraction de motif et la classification. L'extraction de motif saisit l'information pertinente des signaux à classifier et fournit les motifs comme des entrées pour le système de classification.

Nous proposons une méthode pour concevoir des noyaux optimisés, et par conséquent des RTF optimisées pour la discrimination entre les ensembles prédéfinis de classes. Les noyaux résultants ne sont restreints à aucune fonction prédéfinie mais, plutôt, sont arbitraires dans leur forme. Au lieu de faire des hypothèses à priori au sujet du type de lissage temps fréquence, notre approche établit le lissage approprié pour réaliser la meilleure performance de la classification.

2.3.3 La décision

Automatiser implique comprendre, détecter, reconnaître ou classer, ce sont autant de formes courantes d'une problématique unique : la classification de signaux.

En effet, la compréhension fait appel implicitement à un apprentissage et une comparaison [Dud 73], la détection d'un événement consiste à affecter le signal à la classe événement présent ou la classe absent [Van 68], etc. dans ce type de procédure, ce qui est attendu est une décision, une affectation à une classe.

Une décision nécessite un classifieur. Celui-ci fait appel à des indices de divergences. Il existe plusieurs classifieurs. Nous traiterons particulièrement : la méthode de réseau de neurone et la méthode de k- plus proche voisin (chapitre 3).

2.4 Vers une méthode temps-fréquence de classification

2.4.1 La classe de Cohen

la formulation historique de la classe de Cohen [Coh 66] consiste à définir un ensemble de distributions conjointes temps-fréquence, à l'image de distributions de probabilité conjointes de deux variables aléatoires.

La construction de Cohen repose sur cette idée: le spectre d'énergie $|X(f)|^2$ et l'énergie instantanée $|x(t)|^2$ d'un signal sont interprétés comme des distributions de probabilité, qui peuvent être vues comme les densités marginales d'une distribution conjointe temps-fréquence, dépendante du signal. Dans le cas de la fréquence, la valeur moyenne d'une application $g(f)$ est le moment du premier ordre du carré de la transformée de Fourier d'un signal x :

$$\langle g(f)x, x \rangle = \int_{-\infty}^{+\infty} g(f) |X(f)|^2 dt df = \langle g(F) \rangle_x \dots \dots \dots (2.6)$$

Dans le cas d'une distribution conjointe des variables t et f pour un signal x (notée $C_x(t, f)$), la valeur moyenne d'une application g du temps et de la fréquence s'écrit :

$$\langle g(t, f)x, x \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(t, f) C_x(t, f) dt df \dots \dots \dots (2.7)$$

La distribution $C_x(t, f)$ est ici inconnue, et l'on cherche à la construire. On peut à présent se poser la question des règles de correspondance. Comment associer $g(t, f)$ à un éventuel opérateur dans le domaine dual ?

L'idée originale de Wigner reprise et étendue par Cohen [Coh 66] est de proposer d'associer à $g(t, f)$ l'opérateur $G(t, f)$ (dont la forme est pour l'instant inconnue) dans les domaines temporel et fréquentiel de la façon suivante :

$$\langle g(t, f)x, x \rangle = \langle G(T, F)x, x \rangle = \langle G(T, F)X, X \rangle \dots \dots \dots (2.8)$$

2.4.1.1 La classe de cohen par la RTF

La localisation temps-fréquence se trouve confrontée au problème de la transformation mathématique qui permet de passer du signal analysé à sa représentation temps-fréquence. C'est une limitation fondamentale, connue sous le nom de principe d'incertitude, qui exclut toute localisation temporelle précise d'une fréquence.

Il a pour conséquence qu'on ne pourra jamais considérer une transformation comme la seule transformation temps-fréquence correcte parce que la localisation temps-fréquence ne peut pas être vérifiée de manière exacte. Cette ambiguïté qui a donné lieu à la définition d'une grande diversité de transformations temps-fréquence se caractérise aujourd'hui par une multitude de transformations (ou représentation) différentes.

Elles sont regroupées dans différentes classes, en l'occurrence la classe affine, la classe Cohen, la classe hyperbolique. Parmi ces classes, nous avons choisi la classe de Cohen qui admet une formulation en termes de noyau. La classe de Cohen est particulièrement intéressante car elle fournit un ensemble illimité de représentations connues et paramétrables par la simple donnée d'une fonction bidimensionnelle (le noyau). L'expression générale de la classe est :

$$C_x^\phi(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(s-t, \nu-f) w_x(s, \nu) ds d\nu \dots \dots \dots (2.9)$$

où ϕ est le noyau de la représentation temps-fréquence, et $w_s(t, f)$ est la distribution de Wigner-Ville. Dans le cas de deux signaux x et y , cette dernière s'écrit :

$$w_s(t, f) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) y^*\left(t + \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \dots \dots \dots (2.10)$$

Toute représentation temps-fréquence s'obtient par la double convolution temps-fréquence de la distribution de Wigner-Ville. On remarque que le choix du noyau ϕ est arbitraire; de ce choix découlent les propriétés de la représentation temps-fréquence de Wigner-Ville $w_s(t, f)$.

Le respect des propriétés particulières de la distribution de $w_s(t, f)$, est conditionné par des contraintes imposées au noyau. Pour que ce soit effectivement des distributions d'énergie il faut vérifier la condition [Leb 06]:

$$\int_{-\infty-\infty}^{+\infty+\infty} \int \phi(t; f) dt df = 1 \dots\dots\dots(2.11)$$

2.4.1.2 La classe de Cohen par la fonction d'ambiguïté

La définition de la classe de Cohen par la fonction d'ambiguïté est très intéressante car elle permet en pratique l'optimisation du noyau, rendant la longueur des calculs acceptable. En effet, dans le plan des ambiguïtés, on peut écrire :

$$C_x^\phi(t, f) = \int_{-\infty-\infty}^{+\infty+\infty} \int \Phi(\xi, \tau) A_x(\xi, \tau) e^{j2\pi(t\xi - f\tau)} d\xi d\tau \dots\dots\dots(2.12)$$

Où Φ est la transformation de Fourier de ϕ et $A_x(\xi, \tau)$ la fonction d'ambiguïté du signal, définie par :

$$\begin{aligned} A_x(\xi, \tau) &= \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi\xi t} dt \dots\dots\dots(2.13) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w_x(t; f) e^{j2\pi(-t\xi + f\tau)} d\xi d\tau \end{aligned}$$

Ainsi, le produit de convolution temps-fréquence de l'équation (2.9) est diagonalisé : il s'exprime comme le produit simple de A_x et Φ obtenus par transformations de Fourier directes et inverses de w_x et de ϕ .

En pratique, il est éminemment plus rapide de passer dans le plan des ambiguïtés, multiplier A_x par ϕ et revenir dans le plan temps-fréquence par transformations de Fourier.

La figure 2.3 donne les notations et les règles de passage entre les domaines temps-fréquence et doppler-retard [Leb 06]. Les domaines temps-fréquence et doppler-retard sont duaux, mais on utilise souvent en dernier lieu le domaine temps-fréquence.

L'une des raisons majeures est que la fonction d'ambiguïté est généralement à valeurs complexes, le module et la phase portant des informations de nature différentes, alors que l'on peut s'arranger pour obtenir des représentations temps-fréquence à valeurs réelles.

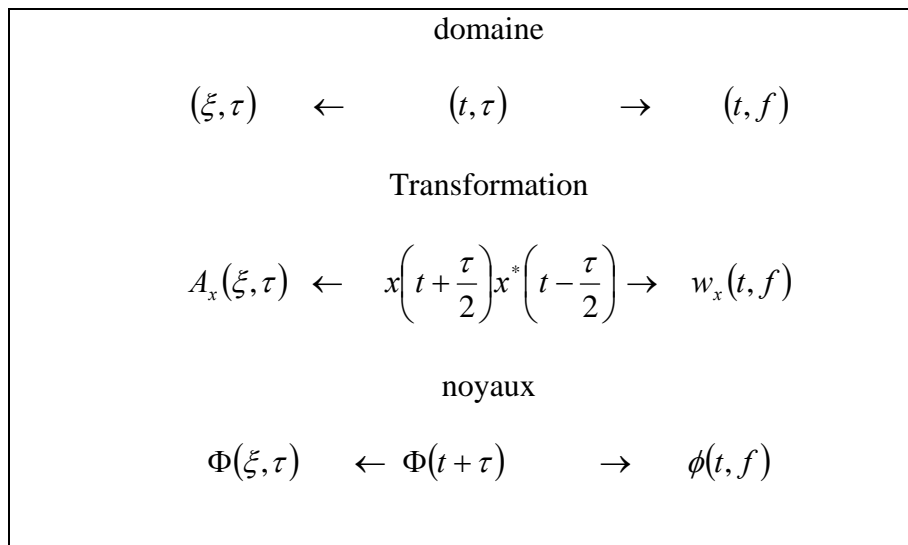


Figure 2.3 Règle des passage et notations pour les domaines temps-fréquence et doppler-retard (les flèches indiquent les transformées de Fourier directes) [Dav 00].

En pratique, il est éminemment plus rapide de passer dans le plan des ambiguïtés, et revenir dans le plan temps-fréquence par transformations de Fourier. L'atout principal du plan doppler-retard est son invariance en module aux translations temporelles et fréquentielles des signaux, tout en contenant la même information que le plan temps-fréquence c'est-à-dire en tenant compte cette fois du module et de la phase [Leb 06].

Dans le plan d'ambiguïté le temps de calcul est peu coûteux car la dimension de l'espace de représentation est réduite à $N_{point}/4$; en effet, la fonction d'ambiguïté d'un signal est symétrique par rapport au point $(\eta = 0; \tau = 0)$, elle est donc redondante.

2.4.2 Existence et signification des interférences

La représentation d'un signal égal à une somme de deux signaux n'est pas égale à la somme des représentations de chacun des deux signaux :

$$\text{Si } y(t) = x_1(t) + x_2(t)$$

$$RTF_y(t, f) = RTF_{x_1}(t, f) + RTF_{x_2}(t, f) + RTF_{x_1 x_2}(t, f) + RTF_{x_2 x_1}(t, f) \dots \dots \dots (2.14)$$

$$RTF_{x_2 x_1}(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(t - \nu, \tau) x_1\left(\nu + \frac{\tau}{2}\right) x_2^*\left(\nu - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\nu d\tau \dots \dots \dots (2.15)$$

Les termes supplémentaires $RTF_{x_1 x_2}(t, f)$ et $RTF_{x_2 x_1}(t, f)$ de l'équation (2.14) sont appelés termes croisés d'interférence, alors que les termes dus aux composantes prises indépendamment sont appelés les termes propres ou motif. La représentation temps-fréquence conserve l'énergie du signal, ces termes croisés d'interférence distribuent le produit scalaire des deux composantes ainsi :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} RTF_y(t, f) dt df = \int_{-\infty}^{+\infty} |x_1(t) + x_2(t)|^2 dt \dots \dots \dots (2.16)$$

Donc

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [RTF_y(t, f) - RTF_{x_1}(t, f) - RTF_{x_2}(t, f)] dt df = 2 \operatorname{Re} \left\{ \int_{-\infty}^{+\infty} x_1(t) x_2^*(t) dt \right\} \dots \dots \dots (2.17)$$

Leur présence est donc une condition sine qua non pour que la propriété de conservation de l'énergie soit vérifiée [Hla 97]. Mais la localisation et l'importance de ces termes supplémentaires constitue souvent un gêne qui ne facilite pas l'utilisation et l'interprétation de ces représentations, et peut même l'interdire lorsque le signal comporte un grand nombre de composantes.

2.4.3 Conception des noyaux pour la classification

Supposant qu'il y ait N exemples d'apprentissage dans un ensemble d'apprentissage pour une classe particulière. Notons que $y_i^{(c)}$ décrit le ième exemple d'apprentissage de la cth classe.

Le but est de concevoir un classifieur pour déterminer l'appartenance de classe d'un vecteur x . Ceci est fait en utilisant les exemples d'apprentissage, sachant que le vecteur x ne figure pas dans l'ensemble de d'apprentissage.

Théoriquement, le classifieur est réalisé par l'intermédiaire d'une fonction de décision $\alpha(x) \in \{1 \dots c \dots C\}$ [Dud 73]. Celle-ci est habituellement réalisée, dans la pratique, par l'extraction des motifs à partir d'une RTF standard (par exemple, le spectrogramme) et par la classification directe de ces motifs. Cela suppose des hypothèses implicites concernant le type de lissage temps-fréquence requis pour la classification. Comme nous l'avons indiqué, ceci peut dégrader les performances de la classification si la RTF standard n'est pas régulière.

Nous proposons de concevoir un classifieur directement dans le plan d'auto ambiguïté. Puisque toutes les RTF peuvent être dérivées du plan d'auto-ambiguïté, aucune hypothèse à priori n'est faite concernant le lissage exigé pour la classification.

Le noyau $\phi_{DCS}[\eta, \tau]$ est conçu pour chaque tâche spécifique de classification. Ce procédé de conception utilise $A_{y_i^{(e)}}$ (la représentation de matrice de la fonction d'auto-ambiguïté calculée à partir des exemples $y_i^{(e)}$) pour déterminer directement le lissage exigé. x est classé via $\alpha(\phi_{DCS} \circ A_x) \in \{1 \dots c \dots C\}$, où \circ est le produit élément par élément.

2.4.4 Le critère de décision

Pour valider les méthodes de décision mises en place, nous avons utilisé deux méthodes de classification l'une et la méthode de réseau de neurone et l'autre et la méthode de k- plus proche voisin (chapitre 3) :

- La décision par réseau de neurone est également justifiée par d'autres propriétés intéressantes, entre autres, une convergence d'apprentissage rapide et stable et les développements mathématiques qui lui ont été consacrés.
- La décision par les k-plus proche voisin consiste à associer une mesure à la classe majoritairement représentée parmi les mesures voisines les plus proches (au sens d'une distance qui peut être euclidienne ou de Mahalanobis). L'association à cette règle, d'une fonction d'appartenance définie pour chaque classe, permet de déterminer un indice de fiabilité (confiance) à l'affectation réalisée.

2.5 RTF optimisée pour la classification

La méthode choisie est la RTF dite "dépendante de la classe du signal (DCS)". Celle-ci n'exige à priori aucune forme paramétrique et prédéfinie du noyau. La méthode proposée nécessite deux étapes principales : l'extraction de motif et la classification.

L'extraction de motif saisit l'information pertinente des signaux à classer et fournit les motifs comme des entrées pour le système de classification. Plusieurs classifieurs sont adaptés au procédé de la classification; nous opterons ici pour la distance Mahalanobis, entre le vecteur forme du signal à classer et ceux des classes des défauts, en tant que critère de décision.

Nous proposons une méthode pour concevoir des noyaux optimisés, et par conséquent des RTF optimisées pour la discrimination entre les ensembles prédéfinis de classes. Les noyaux résultants ne sont restreints à aucune fonction prédéfinie mais, plutôt, sont arbitraires dans la forme. Au lieu de faire des hypothèses à priori au sujet du type de lissage temps-fréquence, notre approche établit le lissage approprié pour réaliser la meilleure performance de la classification.

La conception de la RTF résulte d'une transformée de Fourier inverse discrète en n (temps discret) appliquée à la fonction d'auto-corrélation instantanée :

$$R[n, \tau] = x^*[n] x[(n + \tau)_N] \dots\dots\dots (2.18)$$

Ainsi, la fonction d'auto ambiguïté est définie par :

$$A[\eta, \tau] = F_{n \rightarrow \eta}^{-1} \{R[n, \tau]\} = \sum_{n=0}^{N-1} R[n, \tau] e^{-j \frac{2\pi}{N} n \eta} \dots\dots\dots (2.19)$$

η et τ sont respectivement les Doppler et retard discrets. La RTF correspondante qui est une version discrète de la RTF du Rihaczek est donnée par :

$$R[n, k] = F_{\eta \rightarrow n}^{-1} \{F_{\tau \rightarrow k} \{A[\eta, \tau]\}\} = \frac{1}{N} \sum_{\eta=0}^{N-1} \sum_{\tau=0}^{N-1} A[\eta, \tau] e^{-j(2\pi/N)nk} e^{j(2\pi/N)\eta m} \dots\dots\dots (2.20)$$

Où k est la fréquence discrète et N le nombre d'échantillons de l'enregistrement. La fonction caractéristique [Coh 95] de la RTF discrète du Rihaczek est $A[\eta, \tau]$.

Il existe un noyau $\phi[\eta, \tau]$ qui opère par la multiplication bidimensionnelle sur la fonction d'auto-ambiguïté [Cla 80]. La RTF correspondante est donnée par :

$$G[n, k] = F_{\eta \rightarrow n}^{-1} \{ F_{\tau \rightarrow k} \{ \phi[\eta, \tau] A[\eta, \tau] \} \} = \frac{1}{N} \sum_{\eta=0}^{N-1} \sum_{\tau=0}^{N-1} \phi[\eta, \tau] A[\eta, \tau] e^{-j(2\pi/N)k\tau} e^{j(2\pi/N)\eta n} \dots\dots(2.21)$$

La fonction caractéristique de $G[n, k]$ est $\phi[\eta, \tau] A[\eta, \tau]$. Tout élément non nul du noyau $\phi[\eta, \tau]$ en termes de η et/ou τ peut effectuer un lissage sur la RTF originale (équation 2.17) de Rihaczek, $R[n, k]$, en temps et/ou en fréquence respectivement. $G[n, k]$ est une version lissée de $R[n, k]$.

L'avantage du noyau est que toutes les RTF peuvent être obtenues à partir de $R[n, k]$ par l'application d'un noyau approprié. Ainsi $G[n, k]$ peut être considéré comme RTF généralisée [Coh 95].

Le plan d'ambiguïté a des propriétés très appropriées pour la classification. Une seule position (η_i, τ_i) dans ce plan offre des informations "globales" sur la structure temps-fréquence du signal.

Si $\phi[\eta, \tau] = 0$ pour toutes les valeurs exceptées celles sur l'axe $\eta = 0$, alors toute l'information temporelle est lissée, et seulement l'information de la fréquence stationnaire est maintenue dans $G[n, k]$ (RTF résultante lissée).

Si $\phi[\eta, \tau] = 0$ pour toutes les valeurs exceptées celles sur l'axe $\tau = 0$, alors toute l'information spectrale est lissée et seulement l'information temporelle est retenue. Les points qui ne sont pas sur l'un ou l'autre axe correspondent à une structure temps-fréquence non stationnaire.

Quand le noyau $\phi[\eta, \tau]$ est conçu dans le but de la classification, nous nous référons à lui en tant que noyau dépendant de la classe du signal (DCS). Ce noyau est décrit par $\phi_{DCS}[\eta, \tau]$.

En outre, nous nous référons à la RTF correspondante, $RTF_{DCS}[\eta, \tau]$ comme la RTF dépendante de la classe du signal; elle est décrite par :

$$RTF_{DCS}[n, k] = F_{\eta \rightarrow n}^{-1} \left\{ F_{\tau \rightarrow k} \left\{ \phi_{DCS}[\eta, \tau] A[\eta, \tau] \right\} \right\} \dots \dots \dots (2.22)$$

Il est possible de visualiser la RTF dépendante de la classe du signal et d'observer la structure de temps-fréquence exploité par un classifieur.

La RTF de Rihaczek sert de représentation de base puisque la fonction d'ambiguïté et la fonction caractéristique de la RTF Rihaczek sont les mêmes. Généralement, d'autres RTF peuvent être utilisés comme des représentations de base.

Pour décrire cette généralisation, définissons un noyau multiplicatif de transformation $\phi_T[\eta, \tau]$ qui transforme la RTF de Rihaczek en une autre RTF quadratique: $\phi_T[\eta, \tau] A[\eta, \tau]$ est la fonction caractéristique de cette nouvelle RTF. Puisque le noyau de transformation est imposé avant l'application du noyau dépendant de la classe du signal, le "nouveau" noyau dépendant de la classe du signal $\phi_{DCS}[\eta, \tau]$, lié à la nouvelle représentation de base, est donné par :

$$\phi_{DCS}[\eta, \tau] = \frac{\phi_{DCS}[\eta, \tau]}{\phi_T[\eta, \tau]} \dots \dots \dots (2.23)$$

Ce noyau produit une RTF dépendante de la classe du signal par :

$$RTF_{DCS}[n, k] = F_{\eta \rightarrow n}^{-1} \left\{ F_{\tau \rightarrow k} \left\{ \phi_{DCS}[\eta, \tau] (\phi_T[\eta, \tau] A[\eta, \tau]) \right\} \right\} \dots \dots \dots (2.24)$$

D'après (2.23), l'unique critère pour toute RTF de base est que

$$|\phi_T[\eta, \tau]| \neq 0 \dots \dots \dots (2.25)$$

pour tout η et τ , préservant toute l'information dans la fonction originale d'auto-ambiguïté.

2.6 Application de la RTF optimisée pour la classification des défauts de la machine asynchrone

Le noyau DCS souffre de manque de précision car il ne prend pas en considération la variance intra classe, ce qui peut engendrer une classification incorrecte. Le noyau discriminant linéaire offre une forte discrimination de classe et s'approche ainsi du critère optimal de Bayes; mais il nécessite des grands ensembles d'apprentissages, ce qui est rare dans les applications industrielles. La conception d'un noyau par le rapport discriminant de Fisher (RDF) offre l'avantage de traiter des exemples d'apprentissage de nombre limité et de prendre en considération à la fois la séparation entre classes et la variance intra classe [Leb 08], [Leb 09].

Notre démarche consiste à constituer un ensemble d'apprentissage à partir des signaux de courant statorique. Ceux-ci représentent les différents états de fonctionnement de la machine saine et en défaut. Cet ensemble d'apprentissage servira comme base de données qui permettra l'affectation d'un nouveau signal non classé, par le biais d'un classifieur.

2.6.1 Application du RDF dans le plan Doppler-retard

Le contraste de Fisher pour les s points qui maximise $J_F(s)$ est :

$$J_F(s) = \frac{(m_1 - m_2)^2}{V_1^2 + V_2^2} \dots\dots\dots(2.26)$$

m représente la moyenne des classes 1 et 2,

V^2 représente la variance des classes 1 et 2

Soit un ensemble de signaux qui représentent les courants statoriques dans le cas sain, déséquilibre rotorique et déséquilibre statorique. Un nombre total de $N - 1$ noyaux doit être conçu pour N classes.

Dans notre cas nous avons quatre classes: classe moteur sain, classe défaut rotor, classe défaut stator et classe défaut de roulement. Ces classes sont conçues pour un régime de fonctionnement du moteur à vide et en charge. La discrimination entre classes par le critère de Fisher est élucidée par la figure 2.4.

L'objectif est de concevoir des noyaux qui permettent de maximiser l'écart entre classes et de minimiser la variance intra classe. Supposant qu'il y ait n classes et un total de N_i exemples d'apprentissage pour la classe i . La notation $A_{ij}[\eta, \tau]$ représente le plan d'ambiguïté du j th exemple d'apprentissage dans la i th classe.

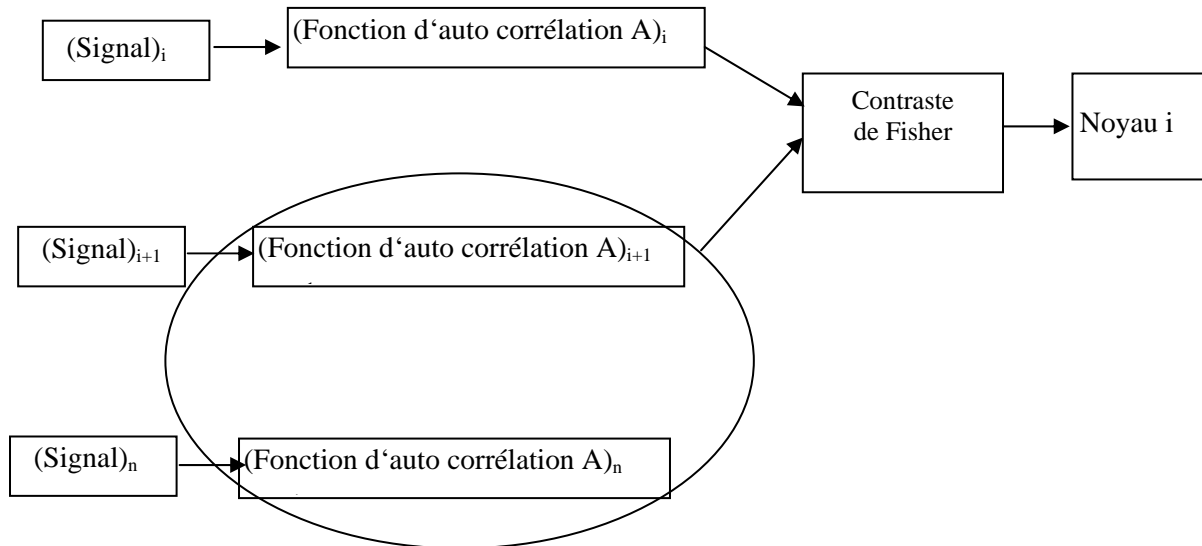


Figure 2.4 Conception des noyaux

2.6.2 Discrimination optimale des classes par le critère de Fisher

Un nombre total de $N-1$ noyaux doit être conçu pour un système de classification des défauts de N classe. Un Noyau K_S peut être soit séparateur de simple classe ou un séparateur de groupe de classe. Dans le cas de séparateur de simple classe, le noyau K_i est conçu pour discriminer la classe i des classes $\{i+1, \dots, N\}$. Dans le cas de séparateur de groupe de classe, le noyau K_i est conçu pour discriminer un groupe de classe $\{i, i+1, \dots, i+m\}$ de toutes les classes restantes $\{i+m+1, i+m+2, \dots, N\}$. [Leb 08], [Leb 09].

Pour le séparateur de simple classe. Trois variantes sont proposées pour une discrimination optimale des classes. Dans notre étude quatre classes ont été envisagées à savoir :

- classe défaut rotor,
- classe défaut stator,
- classe défaut roulement et
- classe moteur sain.

La discrimination entre différentes classes est faite par la séparation de la classe i de toutes les classes restantes $\{i+1, \dots, N\}$ (figure 2.5). Dans ce cas de figure le noyau défaut stator est conçu pour discriminer la classes défaut stator du reste des classes $\{\text{défaut rotor}, \text{défaut roulement et moteur sain}\}$.

Le noyau défaut rotor est conçu pour discriminer la classe défaut rotor du reste des classes $\{\text{défaut roulement et moteur sain}\}$. Le noyau défaut roulement est conçu pour discriminer la classe défaut roulement de la classe moteur sain. L'avantage de la méthode réside dans la séparation optimale entre les différentes classes.

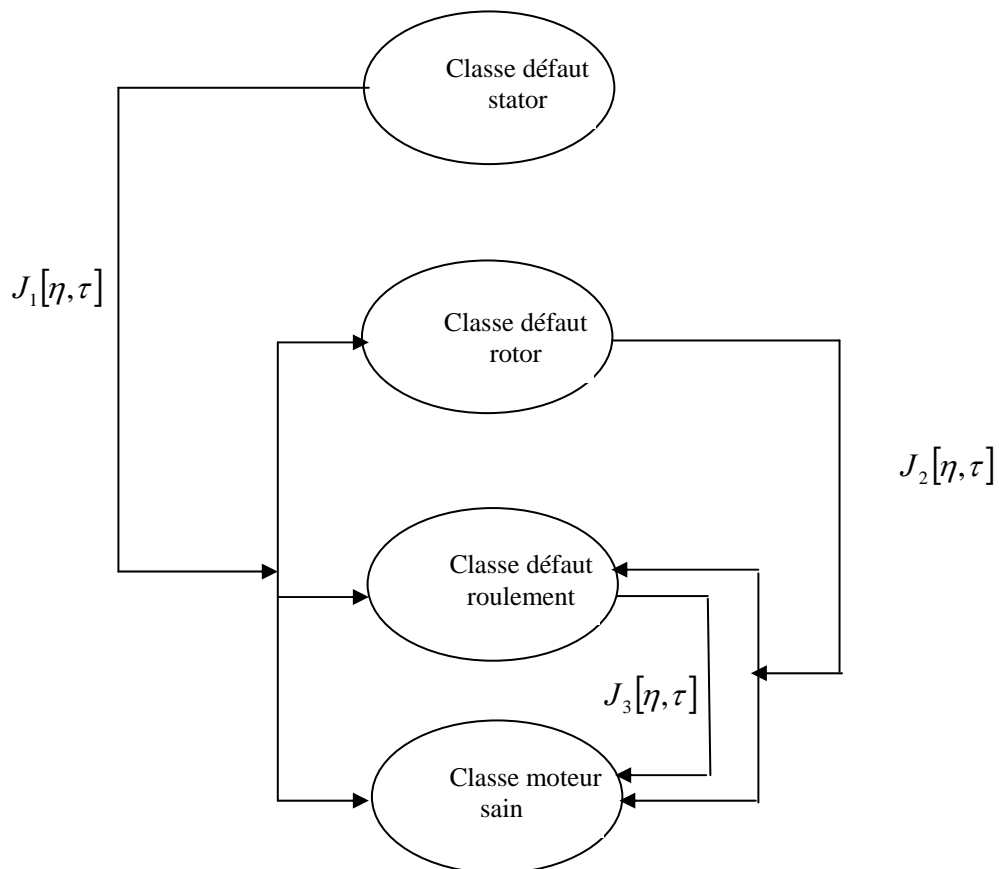


Figure 2.5 Conception des noyaux du premier niveau par la séparation des classes

2.6.2.1 Expression du Contraste de Fisher

Le contraste de Fisher s'exprime alors par :

$$J_i(\eta, \tau) = \frac{(m_i[\eta, \tau] - m_{i-\text{res tan } t}[\eta, \tau])^2}{V_i^2[\eta, \tau] + V_{i-\text{res tan } t}^2[\eta, \tau]} \dots\dots\dots (2.27)$$

où

$m_i[\eta, \tau]$ et $m_{i-\text{res tan } t}[\eta, \tau]$ représentent deux moyennes dans le plan Doppler- retard (η, τ)

$$m_i[\eta, \tau] = \frac{1}{N_i} \sum_{j=1}^{N_i} A_{ij}[\eta, \tau] \dots\dots\dots (2.28)$$

$$m_{i-\text{res tan } t}[\eta, \tau] = \frac{\sum_{k=i+1}^4 \sum_{j=1}^{N_k} A_{kj}[\eta, \tau]}{\sum_{k=i+1}^4 N_k} \dots\dots\dots (2.29)$$

$V_i^2[\eta, \tau]$ et $V_{i-\text{res tan } t}^2[\eta, \tau]$ représentent deux variances dans le plan Doppler- retard (η, τ)

$$V_i^2[\eta, \tau] = \frac{1}{N_i} \sum_{j=1}^{N_i} (A_{ij}[\eta, \tau] - m_i[\eta, \tau])^2 \dots\dots\dots (2.30)$$

$$V_{i-\text{res tan } t}^2[\eta, \tau] = \frac{\sum_{k=i+1}^4 \sum_{j=1}^{N_k} (A_{kj}[\eta, \tau] - m_{i-\text{res tan } t}[\eta, \tau])^2}{\sum_{k=i+1}^4 N_k} \dots\dots\dots (2.31)$$

La troisième variante est optimale par rapport aux deux autres variantes, car elle permet une discrimination non redondante entre les différentes classes.

2.6.2.2 Conception des noyaux discriminants de Fisher

La conception des Noyaux Discriminants de Fisher (NDF) s'effectue en plusieurs étapes (figure 2.6) :

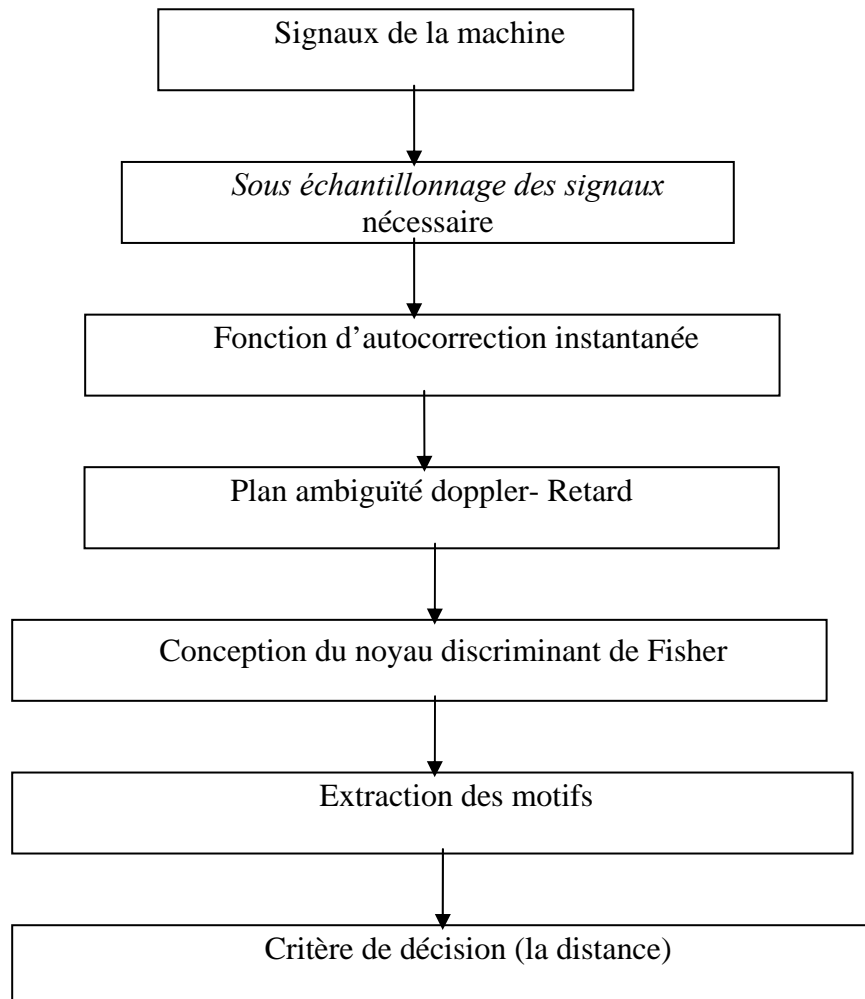


Figure 2.6 Démarche de la classification

a. Signaux pour apprentissage

L'ensemble des signaux d'apprentissage a été acquis avec une fréquence d'échantillonnage de 400 Hz et une durée d'observation de 10s. De ce fait, le nombre d'échantillons par signal s'élève à $N=4000$. Le moteur est alimenté par deux types d'alimentation: une alimentation réseau ou une alimentation convertisseur. La dimension de la matrice d'autocorrection instantanée comportera la dimension de $N^2 = 1610^6$.

La fréquence d'échantillonnage est compatible avec les fréquences caractéristiques des défauts. En effet, la fréquence des ruptures de barres se situe dans la plage fréquentielle 35 à 50 Hz, et la fréquence de défaut statorique correspond, en module, à la fréquence fondamentale 50Hz.

b. Sous échantillonnage des signaux

Le sous échantillonnage consiste à réduire la fréquence d'échantillonnage initiale si la fréquence d'acquisition est trop élevée (certaines de nos mesures ont été effectuées à 10 kHz); celle-ci contient des fréquences qui ne sont pas utiles à notre étude, seule la plage des fréquences recherchée est préservée.

La fréquence des ruptures de barres se situe dans la plage fréquentielle 35 à 50 Hz, et la fréquence de défaut statorique correspond, en module, à la fréquence fondamentale 50Hz. Afin de conserver toutes ces fréquences caractéristiques, le signal original est sous échantillonné à 25 fois la fréquence d'échantillonnage initiale.

Par conséquent, la dimension de signal va être considérablement réduite, ainsi que le temps de calcul. Par ce procédé, nous réduisons de 500 fois le signal initial; de même, chaque signal de l'ensemble d'apprentissage comportera 200 points avec la conservation des signatures fréquentielles des signaux originaux.

c. Fonction instantanée d'auto corrélation et le plan d'ambiguïté

Nous calculons la fonction instantanée d'auto corrélation $R[n,t]$ de signal i (2.32) par l'expression (2.19); le signal de courant est donné par le vecteur :

$$i = [i_1 \ i_2 \ i_3 \ \dots \ i_{n-1} \ i_n] \dots\dots\dots (2.32)$$

Le plan d'auto-ambiguïté du signal est obtenu par la transformée de Fourier inverse de la fonction d'auto corrélation instantanée (2.11). La dimension du plan d'auto ambiguïté est 200x200. Nous choisissons des points directement de N à partir de ce plan en tant que vecteur de motif. Les points sur l'axe $\eta = 0$ résultent une structure temps-fréquence stationnaire dans le temps.

De même, les points sur l'axe $\tau = 0$ correspondent à une structure de temps- fréquence stationnaire en fréquence [Gil 99].

d. Conception du noyau discriminant de Fisher

Les plans d'auto-ambiguïtés sont distincts pour chaque signal de courant (sain et avec défaut). Nous voulons déterminer les coordonnées des N' points du plan d'ambiguïté $N \times N$, telles que les valeurs en ces endroits soient très semblables pour des signaux de la même classe, alors qu'elles changent de manière significative pour des signaux de différentes classes.

La conception du noyau est basée sur la fonction discriminante de Fisher pour obtenir les coordonnées de N' points du plan d'ambiguïté. Ici le noyau est conçu par les signaux d'apprentissage de chaque classe.

La fonction de Fisher est maximisée quand la séparation entre les moyennes des classes est grande et la variance interclasse est faible. La fonction de Fisher est conçue pour un ensemble d'apprentissage dans un plan de dimension $N \times N$.

Nous transformons la fonction de Fisher en matrice binaire (le noyau) en remplaçant les points les plus élevés par 1 et les autres points par 0. L'opération de transformation est automatique, elle est faite par le choix de N premiers points de plus fort contraste de la matrice de Fisher, ces positions de points seront maintenu par '1' le reste seront annulés par 'zéro'.

e. Extraction des motifs

Des motifs peuvent être extraits directement à partir de $A[\eta, \tau] \varphi_i[\eta, \tau]$. Le noyau $\varphi_i[\eta, \tau]$ a les mêmes dimensions que le plan d'ambiguïté. Dans cette application, le noyau $\varphi_i[\eta, \tau]$ est défini comme une matrice binaire où chaque élément est soit 1 ou 0 :

$$A[\eta, \tau] \varphi_i[\eta, \tau] = \begin{cases} A[\eta, \tau], & \text{si } \varphi_i[\eta, \tau] = 1 \\ 0, & \text{si } \varphi_i[\eta, \tau] = 0 \end{cases} \dots\dots\dots (2.33)$$

Les points de motif sont les points du plan d'ambiguïté de position (η, τ) quand $\varphi_i[\eta, \tau] = 1$. Par conséquent, le processus d'extraction de motifs doit choisir les points $A[\eta, \tau]$ qui sont optimaux pour la tâche de la classification dans le plan d'ambiguïté [Wan 04].

En multipliant la forme binaire de la fonction de Fisher avec le plan de l'ambiguïté d'un certain signal, nous trouverons N points de motif pour ce signal qui sont présentés par un ordre décroissant dans un vecteur motif (figure 2.7).

On peut interpréter la sélection de points dans le plan Doppler-retard comme le masquage de la fonction d'ambiguïté d'un signal par une fonction binaire adaptée, laquelle peut être vue comme un noyau ϕ optimal. Cette méthode est peu coûteuse en termes de temps de calcul, puisqu'elle réduit l'espace de représentation à $N_{point}/2$ dimensions, soit ici 3 points par noyau.

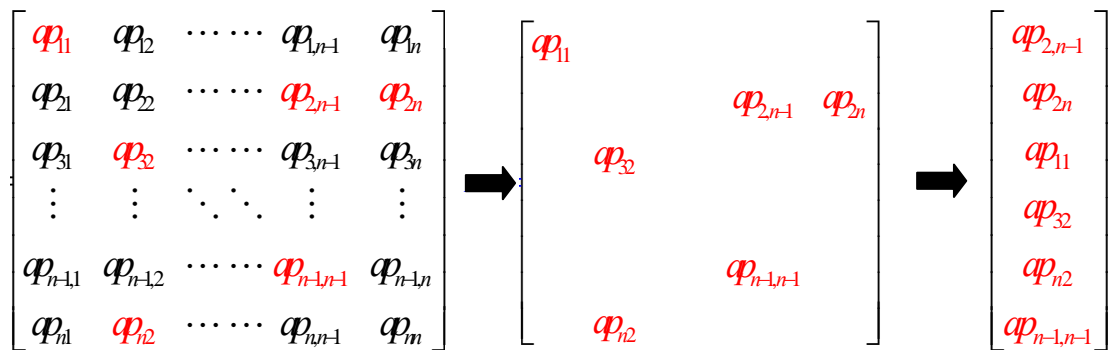


Figure 2.7 Illustration d'extraction du vecteur motif

2.6.2.3 résultats de calcul

L'ensemble d'acquisition des données sur la machine est constitué de 15 exemples de courant du stator, enregistrés à différents niveaux de charge (0%, 25%, 50%, 75%, 100%) ceci pour chaque état de fonctionnement de la machine ; saine, défaut stator, défaut de rupture de barre et défaut de roulement. La base d'apprentissage est réalisée par 10 exemples de courant, les cinq derniers exemples de courant servent pour le test.

2.6.3 Résultats expérimentaux

Un banc moteur a été monté pour effectuer différents essais (figure 2.8), il est constitué du moteur asynchrone de 5.5 KW, 230V, 50 Hz, $p=4$, 1470 t/Mn.

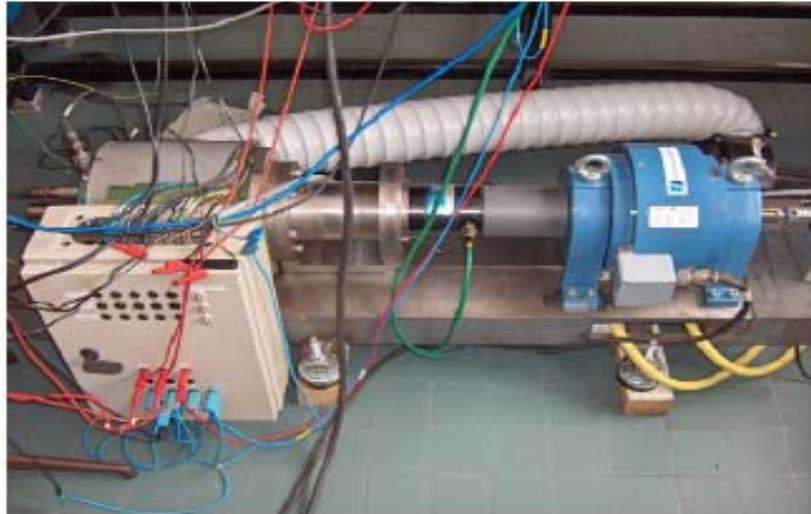


Figure 2.8 Banc d'essai de la machine à induction

L'acquisition des signaux des courants statoriques a été réalisée par une fréquence d'échantillonnage de 10Khz et un temps d'acquisition de 10 s. ceci afin d'obtenir une très bonne résolution spectrale $\pm 0.1\text{Hz}$.



Figure 2.9 défaut roulement



Figure 2.10 rupture de barre

Le processus de conception des noyaux du premier niveau est montré dans la figure 2.6.

Trois noyaux sont nécessaires pour le premier niveau de classification, un noyau pour le défaut stator, un noyau pour le défaut rotor et un noyau pour le défaut de roulement.

La base d'apprentissage du défaut stator a été constituée par 10 exemples de courant à vide et 10 autres exemples en charge. Nous avons donc 20 exemples d'apprentissage pour chaque défaut. La conception des noyaux a été réalisée selon la troisième variante (2.6.2.1).

Le noyau défaut stator est donc conçu afin d'obtenir une séparation maximale entre la classe défaut stator et le reste des classes, à savoir, défaut rotor, défaut roulement et la classe saine.

Le noyau défaut rotor est conçu afin d'obtenir une séparation maximale entre la classe défaut rotor le reste des classes, à savoir, défaut roulement et la classe saine.

Enfin le noyau défaut roulement est conçu afin d'obtenir une séparation maximale entre la classe défaut roulement et la classe saine.

La dimension du plan Doppler-retard contient initialement $200 \times 200 = 40000$ points; vu la symétrie par rapport à l'origine, nous prenons le quart des valeurs de la matrice ce qui correspond à $N=10000$.

Les positions des points de Fisher de fort contraste sont présentées dans le plan Doppler retard figure 2.11.

Nous avons obtenu, après sélection, trois positions de points pour chaque noyau $\{(\xi, \tau), \dots, (\xi, \tau)_s\}$. Les positions et les contrastes de ces points caractéristiques sont montrés dans le tableau 2.1.

Ces derniers servent à l'apprentissage des positions caractéristiques du défaut. La sélection des positions est réalisée sur la base de deux critères : sur la valeur du contraste et sur la base de rapprochement des positions dans le plan Doppler-retard.

La sélection des positions caractéristiques est basée sur le principe de Fisher, c'est à dire maximisation de la séparation interclasse et la minimisation de la variance l'intra classe.

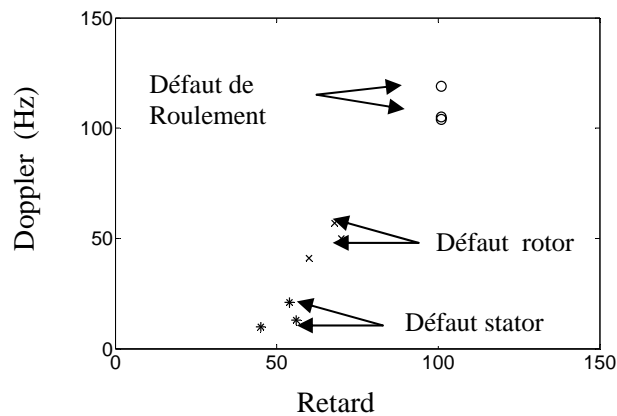


Figure 2.11 Plan d'ambiguïté lissé par trois noyaux

Défauts	Points du vecteur forme	Positions Doppler-retard	Contraste
Défaut roulement	3	(101,104)	22.18
	4	(101,105)	21.81
	5	(101,119)	17.64
Défaut stator	4	(45,10)	9.75
	5	(54,13)	10.2
	6	(56,11)	9.2
Défaut rotor	2	(60,41)	12.91
	3	(68,49)	15.63
	4	(70,57)	14.65

Tableau 2.1. Les positions et les contrastes des points caractéristiques

2.7. Conclusion

Dans ce chapitre nous avons proposé une méthode de représentation temps-fréquence dite "dépendante de la classe du signal" pour l'extraction des paramètres de diagnostic. Nous avons montré que les RTF classiques ont des noyaux paramétriques et à priori prédéfinis donc inappropriés à la classification, d'où l'orientation vers le plan d'ambiguïté doppler-retard où toutes les RTF peuvent être dérivées par un choix approprié d'un noyau. Dans cette application, le choix du noyau est crucial car il permet d'avoir une RTF optimisée et par conséquent une classification précise des signaux. Nous avons opté pour le noyau discriminant de Fisher. Les positions et les contrastes des points caractéristiques des paramètres de diagnostic des trois défauts : statorique, rotorique et de roulement sont résumés dans le tableau 2.1.



Chapitre III

Méthodes de la prise de décision

3.1 Introduction

Le diagnostic automatique nécessite au moins deux étapes principales, la première consiste à l'extraction de l'information pertinente du signal à traiter, chose qui a été faite dans le deuxième chapitre via la représentation temps-fréquence. La seconde étape consiste à la prise de décision, les détails de cette étape feront l'objet du troisième et dernier chapitre. Nous avons présenté dans ce chapitre des méthodes de prise de décision, à savoir les réseaux de neurones et la méthode statistique des k-plus proches voisins.

3.2 Détermination d'une méthode de décision

Soit $X_a = (\underline{X}_1, \underline{X}_2, \dots, \underline{X}_N)$ l'ensemble d'apprentissage initial et $(\Omega_1, \Omega_2, \dots, \Omega_M)$ l'ensemble des M classes connues. Le problème posé est alors d'affecter une nouvelle observation \underline{X}_u à l'une des M classes à l'aide d'une règle de décision.

Toutefois, l'affectation ne doit pas être systématique car des erreurs peuvent être commises. L'algorithme de décision doit donc bénéficier de solutions alternatives afin de diminuer le risque de mauvaise classification. Ces solutions sont fournies par les options de rejets d'ambiguïté et de distance. Il s'agit en réalité de différer la décision car les observations dont l'appartenance aux classes est ambiguë ou insuffisante, pourraient être précurseur de l'apparition de nouvelles classes.

3.3 Méthodes de prise de décision

3.3.1 Réseau de neurone

Un réseau de neurones est un outil d'analyse statistique permettant de construire un modèle de comportement à partir de données qui sont des exemples de ce comportement. Un réseau de neurones est constitué d'un graphe pondéré orienté dont les noeuds symbolisent les neurones. Ces neurones possèdent une fonction d'activation qui permet d'influencer les autres neurones du réseau. Les connexions entre les neurones, que l'on nomme liens synaptiques,

propagent l'activité des neurones avec une pondération caractéristique de la connexion. On appelle poids synaptique la pondération des liens synaptiques.

Notre but ici est l'utilisation des réseaux de neurones en tant qu'outil de prise de décision pour la classification de trois types de défaut de la machine asynchrone à savoir le défaut roulement, défaut statorique et défaut rotorique.

3.3.1.1 Mise en œuvre d'un modèle de réseaux de neurones

Les neurones peuvent être organisés de différentes manières, c'est ce qui définit l'architecture et le modèle du réseau. L'architecture la plus courante est celle dite du perceptron multicouche. Avant de pouvoir utiliser les capacités de classification d'un réseau de neurones, il faut le construire. Ceci se déroule en quatre temps:

1. La construction de la structure du réseau (généralement empirique).
2. La constitution d'une base de données de vecteurs représentant au mieux le domaine à modéliser. Celle-ci est scindée en deux parties : une partie servant à l'apprentissage du réseau (on parle de base d'apprentissage) et une autre partie aux tests de cet apprentissage (on parle de base de test).
3. Le paramétrage du réseau par apprentissage. Au cours de l'apprentissage, les vecteurs de données de la base d'apprentissage sont présentés séquentiellement et plusieurs fois au réseau. Un algorithme d'apprentissage ajuste le poids du réseau afin que les vecteurs soient correctement appris. L'apprentissage se termine lorsque l'algorithme atteint un état stable.
4. La phase de reconnaissance qui consiste à présenter au réseau chacun des vecteurs de la base de test. La sortie correspondante est calculée en propageant les vecteurs à travers le réseau. La réponse du réseau est lue directement sur les unités de sortie et comparée à la réponse attendue. Une fois que le réseau présente des performances acceptables, il peut être utilisé pour répondre au besoin qui a été à l'origine de sa construction.

La décision par réseaux de neurones implique deux étapes majeures :

- L'apprentissage du réseau de neurones, afin d'apprendre les défauts par l'accouplement (association) des entrées et des sorties désirées.
- Le test du réseau de neurones (évaluation) afin d'estimer l'exactitude du modèle proposé pour des événements qui n'ont pas été présentés lors de la phase d'apprentissage.

3.3.1.2 Apprentissage du réseau : Algorithme de rétropropagation de gradient

La rétropropagation est actuellement l'outil le plus utilisé dans le domaine de réseaux de neurones. C'est une technique de calcul des dérivées qui peut être appliquée à n'importe quelle structure de fonctions dérivables.

Mathématiquement, cette méthode est basée sur l'algorithme de descente du gradient et utilise les règles de dérivation des fonctions dérivables. Dans cette méthode, l'erreur commise en sortie du réseau sera rétro propagée vers les couches cachées d'où le nom de rétropropagation.

A. Mise en équation

Avant de définir la règle d'apprentissage, on doit définir la relation entre les sorties du réseau d'une part, et les entrées et les poids d'autre part. Dans un réseau à (l) couches ayant (n) entrées et (m) sorties les états des différents neurones sont donnés par :

$$U_i^l(k) = f^l(P_i^l(k)) \dots \dots \dots (3.1)$$

Avec

$$P_i^l(k) = \sum_{j=1}^{N_{l-1}} W_{ij}^l U_j^{l-1}(k) \dots \dots \dots (3.2)$$

Ou $i = 1, 2, \dots, N_l$

et $j = 1, 2, \dots, N_{l-1}$

et $q = 1, 2, \dots, N_{l+1}$

N nombre de neurones dans la couche 1 .

N_{l-1} nombre de neurones dans la couche l-1 .

N_{l+1} nombre de neurones dans la couche l+1 .

L nombre de couches .

$U_i^l(k)$ sortie du neurone i de la couche l .

$W_{ij}^l(k)$ coefficient synaptique (poids) de la $j^{ème}$ entrée du neurone (i) de la couche (l) .

$$U_i^0(k) = X_i(k) \quad i = 1, 2, \dots, n \dots\dots\dots(3.3)$$

$$U_i^0(k) = Y_i(k) \quad i = 1, 2, \dots, m \dots\dots\dots(3.4)$$

Où : $X_i(k)$ et $y_i(k)$ sont respectivement les entrées et les sorties du réseau .

L'objectif de la méthode de la rétropropagation est d'adapter les paramètres W_{ij}^l de façon a minimiser une fonction de coût donné par :

$$E(W) = \sum_{P=1}^T E_P(W) \dots\dots\dots(3.5)$$

avec

$$E_P(W) = \frac{1}{2} \sum_{i=1}^m [y_i^d(k) - y_i(k)]^2 \dots\dots\dots(3.6)$$

où $y_i^d(k)$ est le vecteur de sortie désiré , $y_i(k)$ le vecteur de sortie de réseau et T le nombre d'exemples ou longueur de l'ensemble d'entraînement.

B. Principe de la rétropropagation

L'approche la plus utilisée pour la minimisation de la fonction E est basée sur la méthode du gradient. On commence l'entraînement par un choix aléatoire des vecteurs initiaux du poids.

On présente le premier vecteur d'entrée, une fois on a la sortie du réseau, l'erreur correspondante et le gradient de l'erreur par rapport à tous les poids sont calculés. Les poids sont

alors ajustés. On refait la même procédure pour tous les exemples d'apprentissage. Ce processus est répété jusqu'à ce que les sorties du réseau soient suffisamment proches des sorties désirées.

C. Accélération de la rétropropagation

Bien que l'algorithme de rétropropagation soit le plus utilisé pour l'apprentissage supervisé des MLP, son implantation se heurte à plusieurs difficultés techniques. Il n'existe pas de méthodes permettant de :

- Trouver une architecture appropriée (nombres de couches, nombre de neurones).
- Choisir une taille et une qualité adéquate d'exemples d'entraînement.
- Choisir des valeurs initiales satisfaisantes pour les poids, et des valeurs convenables pour les paramètres d'apprentissage permettant d'accélérer la vitesse de convergence.
- Problème de la convergence vers un minimum local, qui empêche la convergence et cause l'oscillation de l'erreur.

Plusieurs approches ont été proposées pour remédier à ces problèmes, Une des techniques d'accélération est celle de la création dynamique des neurones, un neurone est ajouté chaque fois que l'erreur se stabilise à un niveau inacceptable.

3.3.1.3 L'application de réseau de neurone pour la classification des défauts de la machine asynchrone

Position du problème

L'objectif de cette phase est l'application des réseaux de neurones pour la classification des nouvelles observations à l'une des classes existantes des défauts de la machine asynchrone (défaut roulement, défaut statorique et défaut rotorique). Le problème consiste à décider ou non, d'affecter une nouvelle observation à l'une des classes connues. Pour se faire nous avons élaborés un programme de réseaux de neurones sur MATLAB et qui consiste notamment à calculer l'erreur quadratique moyenne pour les deux étapes suivantes :

Premier étape est la phase d'apprentissage qui consiste à utiliser les données obtenues dans le second chapitre (Tableau. 3.1) pour construire l'ensemble d'apprentissage.

La seconde étape est la phase de test qui consiste à classifier les nouvelles observations (exemple tableau 3.2).

Défauts	Points du vecteur forme	Positions Doppler-retard	Contraste
Défaut roulement	3	(101,104)	22.18
	4	(101,105)	21.81
	5	(101,119)	17.64
Défaut stator	4	(45,10)	9.75
	5	(54,13)	10.2
	6	(56,11)	9.2
Défaut rotor	2	(60,41)	12.91
	3	(68,49)	15.63
	4	(70,57)	14.65

Tableau. 3.1 Trois noyaux d'apprentissage par type de défaut

Les nouvelles observations	Contraste
$Xu1$	12.36
	20.45
	15.23
$Xu2$	5.7
	15.2
	10.8

Tableau 3.2 : les nouvelles observations Xu

Choix d'une architecture de réseau :

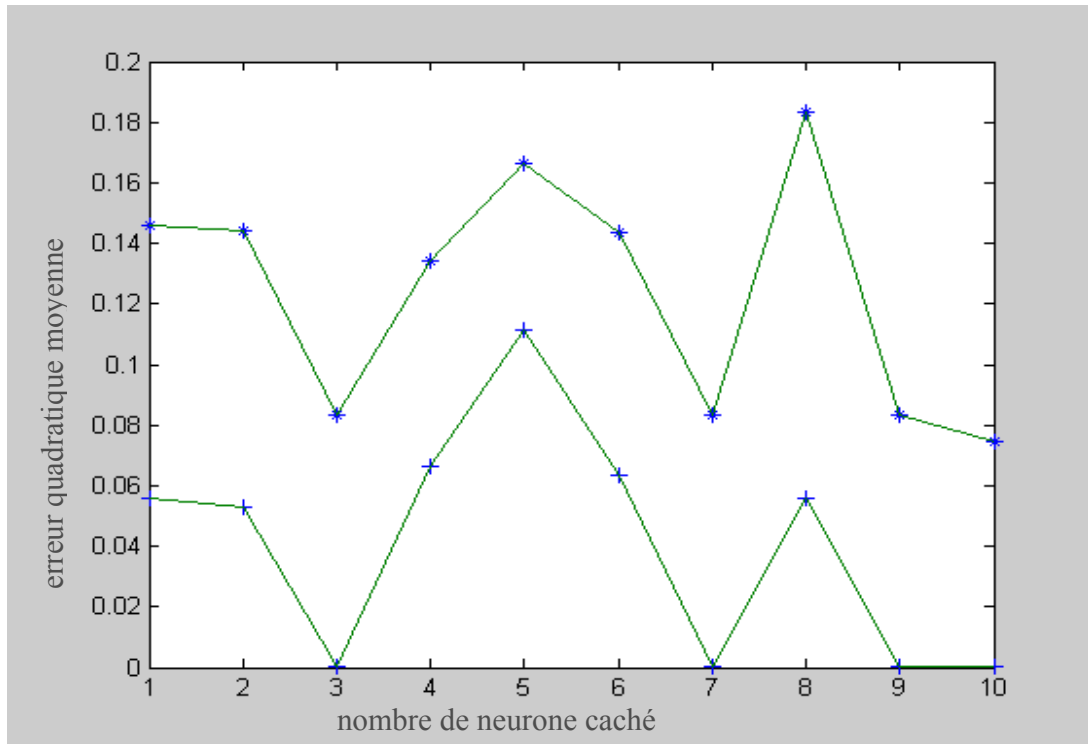


Figure 3.1 *L'erreur quadratique moyenne pour les deux étapes d'apprentissage et de test en fonction de nombre de neurone caché*

Les réseaux de neurones comme il est montré sur la figure 3.1 présentent un comportement aléatoire en fonction du nombre de neurones dans la couche cachée en termes d'erreur. L'augmentation du nombre de neurones cachés (ou augmentation des interconnexions) n'implique pas obligatoirement l'obtention d'une meilleure performance, bien que ce constat va à l'encontre de certains résultats publiés. Donc on choisit une architecture de réseau 7-3 qui nécessite 09 itérations de calcul et une erreur quadratique moyenne minimale par rapport aux autres structures (Figure 3.2, 3.3).

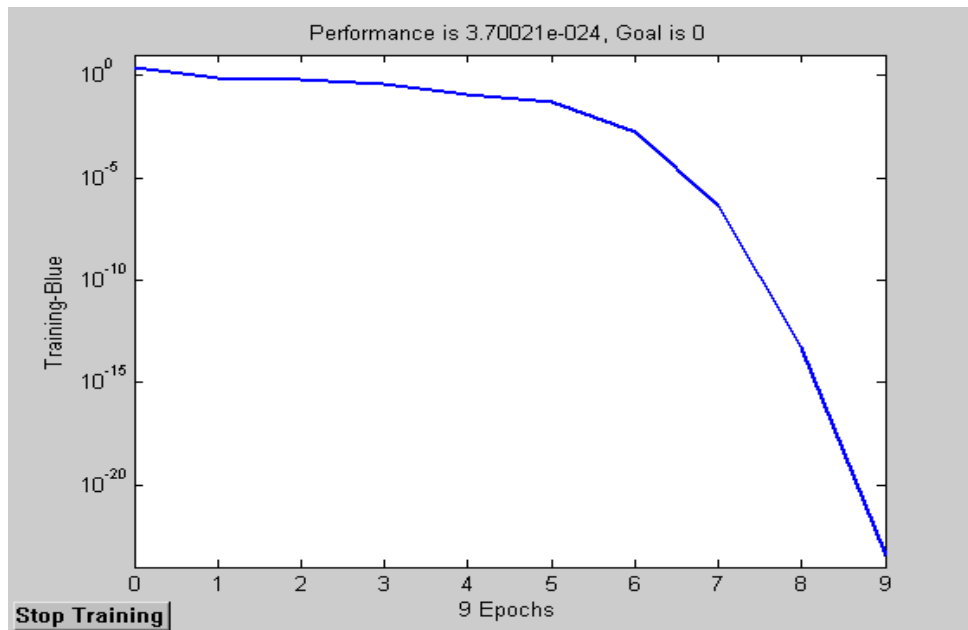


Figure 3.2 l'apprentissage de réseau choisi 7-3

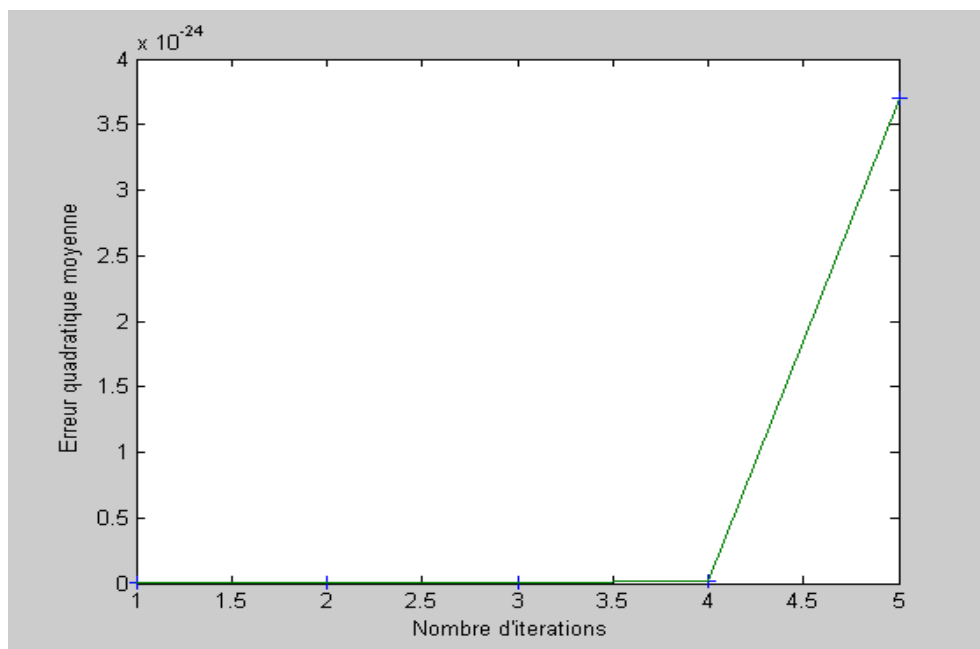


Figure 3.3 l'erreur quadratique moyenne pour l'apprentissage d'un réseau

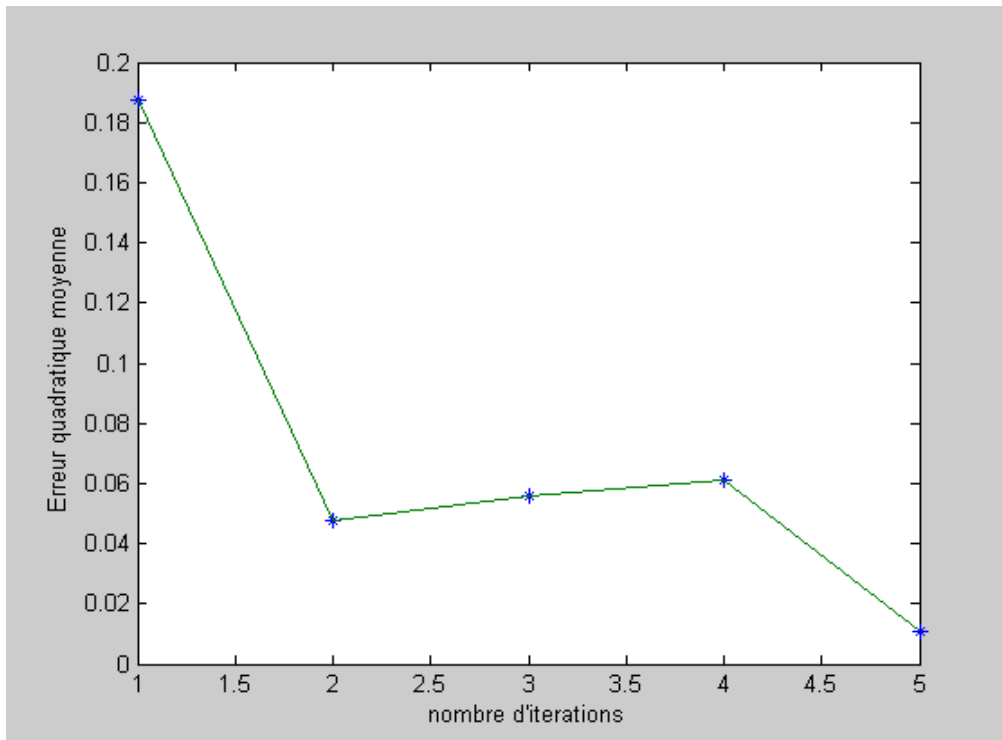


Figure 3.4 l'erreur quadratique moyenne pour le test d'un réseau

La principale critique sur les réseaux de neurones est que les résultats de l'apprentissage, c'est-à-dire le calcul par l'algorithme de rétropropagation, n'est pas interprétable par l'utilisateur : on ne peut pas donner d'explication au calcul d'une sortie sur un vecteur d'entrée. On parle de « boîte noire ». Ceci est la principale différence entre réseaux de neurones et d'autres méthodes de prise de décision. Afin de donner un sens et une lisibilité à la méthode de prise de décision nous proposons une autre méthode à savoir les k plus proches voisins ($k - ppv$).

3.3.2 Règle des k – plus proches voisins ($k - ppv$)

Soit $X_a = (\underline{X}_1, \underline{X}_2, \dots, \underline{X}_N)$ l'ensemble d'apprentissage composé de N vecteurs indépendants, chacun étiqueté à l'une M classes connues. Le principe de la règle des $k - ppv$ est d'affecter une nouvelle observation \underline{X}_u à la classe majoritairement représentée parmi ses k - plus proches voisins.

En réalité, La manière la plus simple d'élaborer cette règle, est de mesurer la distance entre la nouvelle observation et chacun des vecteurs de l'ensemble d'apprentissage. La nouvelle observation sera alors affectée à la classe majoritairement représentée parmi ses $k - ppv$.

La définition de la notion de plus proche voisin est liée au choix d'une distance.

3.3.2.1 Choix de la distance dans la règle des k - ppv

Différentes distances peuvent être utilisées, définies par la formulation générale :

$$d^2(\underline{X}, \underline{Y}) = (\underline{X} - \underline{Y})^t \cdot A(\underline{X} - \underline{Y}) \dots\dots\dots(3.7)$$

avec : A est une matrice définie positive.

$d(\underline{X}, \underline{Y})$: distance entre \underline{X} et \underline{Y}

\underline{X} et \underline{Y} : vecteur de dimension d'

Plusieurs cas particuliers peuvent être tirés de cette formulation :

- La distance *euclidienne* où A est une matrice identité :

$$d_E^2(\underline{X}, \underline{Y}) = (\underline{X} - \underline{Y})^t \cdot A(\underline{X} - \underline{Y}) \dots\dots\dots(3.8)$$

- La distance de *Mahanalobis* où A est l'inverse de la matrice variance- covariance totale Σ :

$$d_M^2(\underline{X}, \underline{Y}) = (\underline{X} - \underline{Y})^t \sum^{-1}(\underline{X} - \underline{Y}) \dots\dots\dots(3.9)$$

L'introduction de la matrice de variance-covariance dans l'expression (3.9) signifie que cette mesure prend en compte la dispersion de l'ensemble d'apprentissage, dans chaque dimension.

3.3.2.2 Choix du nombre de plus proches voisins k

Le choix de k est un choix délicat souvent effectué un peu au hasard. Deux compromis sont à satisfaire cependant : prendre k grand pour obtenir une estimation « fiable », et être certain

que les k voisins sont proches de la nouvelle observation pour s'assurer de ne pas effectuer un diagnostic erroné.

Le choix reste souvent arbitraire, k est pris relativement faible par rapport à la taille de l'ensemble d'apprentissage N . k est toujours ramené à un nombre impair, pour éviter l'ambiguïté d'indécision. k égal à N [Dub 90] réalise un bon compromis.

Il est possible de faire une étude sur k [Ond 06] qui évalue la valeur de k permettant d'obtenir les meilleurs résultats en terme de classification et il est obtenu que quelque soit le nombre k de voisins, le taux de classification est élevé (varie de 98 à 99%). Ceci permet de valider, d'une part, le choix des paramètres formant le vecteur forme optimal. Ces paramètres présentent un fort pouvoir de séparation entre les différentes classes de X_{al} , rendant la classification plus efficace. D'autre part ce fort taux permet de montrer la robustesse de notre approche et notamment de la règle de décision.

3.3.2.3 Règle des k – plus proches voisins avec rejet

En réalité, il existe deux grandes notions de rejet : le rejet en ambiguïté qui concerne une nouvelle observation située entre deux ou plusieurs classes de X_a et le rejet en distance qui correspond à une nouvelle observation située à une distance éloignée des classes de X_a . Ces rejets ont été introduits dans la règle des k – ppv et plus généralement, dans les méthodes de diagnostic afin de diminuer les erreurs de classification et, par conséquent, les erreurs de diagnostic. Globalement, la règle de décision incluant les deux options de rejets sera appliquée pour $(M + 2)$ classes :

- $\underline{X}_u \rightarrow \Omega_c (c = 1, M)$: \underline{X}_u est classé dans Ω_c .
- \underline{X}_u est rejeté en ambiguïté alors : $\underline{X}_u \rightarrow \Omega_0$.
- \underline{X}_u est rejeté en distance alors : $\underline{X}_u \rightarrow \Omega_d$.

avec

Ω_d : classe fictive des observations rejetées en distance

Ω_0 : classe fictive des observations rejetées en ambiguïté

3.3.2.3.1 Rejet d'ambiguïté

Lorsque deux états de fonctionnement du système sont représentés par deux classes proches dans \mathfrak{R}^d et que l'observation \underline{X}_u du système à l'instant t se trouve à la limite de la frontière de décision entre ces deux classes (figure 3.5), il semble intuitivement plus sûr de ne pas prendre de décision sur l'affectation de \underline{X}_u .

Par contre le classifieur indiquera à l'utilisateur qu'il y a ambiguïté sur la décision. Appliqué à la règle des k – ppv, le rejet d'ambiguïté se définit ainsi [Hel, 70], [Tom, 76] :

- \underline{X}_u est associé à la classe Ω_c si k' au moins parmi ses k plus proches voisins appartiennent à Ω_c

$$\begin{cases} \underline{X}_u \rightarrow \Omega_0 & \text{si } k_c = \max_{r=1,M} k_r < k' \\ \underline{X}_u \rightarrow \Omega_c & \text{si } k_c = \max_{r=1,M} k_r \geq k' \end{cases} \dots\dots\dots (3.10)$$

Avec:

k le nombre de plus proches voisins du vecteur \underline{X}_u .

k_c le nombre de plus proches voisins de \underline{X}_u (parmi k) dans la classe Ω_c .

Les observations rejetées en ambiguïté seront affectées à une nouvelle classe fictive Ω_0 , appelée classe de rejet en ambiguïté. Plus k' est proche de k , plus le rejet en ambiguïté est important.

La règle de rejet d'ambiguïté a été par la suite généralisée par [Dev 77] en fixant un nombre minimal de voisin propre à chacune des classes. D'autre part si $k' = 0$, alors le rejet n'est pas appliqué.

3.3.2.3.2 Rejet de distance

Seule, la notion de rejet en ambiguïté est insuffisante. Il faut ajouter la possibilité de décision sur des observations résultant de l'apparition de nouvelles classes.

Ces nouvelles classes correspondent généralement à des zones de l'espace inexplorées, au sens où aucune observation de X_a n'y a été rencontrée. Dans ce cas, une deuxième possibilité de rejet a été utilisée, appelée rejet en distance [Dub 90].

Ce rejet en distance est basé sur la notion de distance aux classes, qui doit être 'grande'. L'idée la plus simple consiste à effectuer un seuillage de distance. L'observation \underline{X}_u est rejetée si la distance à ses k - ppv est supérieure à un seuil T .

Le seuil de distance est défini par l'utilisateur. Il est comparé par exemple :

- à la distance moyenne de \underline{X}_u à ses k plus proches voisins $y_h (h = 1, k)$:

$$\underline{X}_u \rightarrow \Omega_d \text{ si } \frac{1}{k} \sum_{h=1}^k d(\underline{X}_u, \underline{y}_h) > T \dots\dots\dots(3.11)$$

- ou à la distance entre X_u et le centre de gravité de sa classe d'affectation :

$$\underline{X}_u \rightarrow \Omega_d \text{ si } \frac{1}{k} \sum_{h=1}^k d(\underline{X}_u, \underline{m}_c) > T \dots\dots\dots(3.12)$$

m_c étant le centre de gravité de la classe d'affectation de \underline{X}_u

Le seuil T peut dépendre de la classe, donc tenir compte de sa géométrie et être lié au diamètre de la classe où a des considérations statistiques. Ainsi une première solution peut être d'assimiler le seuil de rejet en distance T_c de la classe Ω_c au diamètre de cette classe. Pour cela, il est nécessaire de calculer les distances entre les vecteurs de celle-ci et son centre de gravité m_c :

$$T_c = \beta \cdot \max_{v=1, N_c} [d(\underline{X}_{cv}, \underline{m}_c)] \text{ avec } \beta = 2 \text{ par exemple}$$

Avec :

N_c , le nombre de vecteurs de la classe Ω_c

\underline{X}_{cv} , le $V^{i\text{ème}}$ vecteur de la classe Ω_c

La marge de sécurité (facteur β) permet de prendre en compte les incertitudes sur la dimension exacte de la classe mais aussi sur le bruit lors de l'acquisition des mesures. La deuxième solution consiste à définir les caractéristiques statistiques réelles des classes pour ensuite s'en servir lors de la détermination du seuil.

Finalement, la règle des k – ppv incluant les deux options de rejet peut s'exprimer par :

$$\left\{ \begin{array}{l} \underline{X}_u \rightarrow \Omega_d \text{ si } d(\underline{X}_u, \underline{m}_c) > T_c \\ \underline{X}_u \rightarrow \Omega_0 \text{ si } d(\underline{X}_u, \underline{m}_c) \leq T_c \text{ et } k_c = \max_{r=1,M} k_r \langle k' \dots\dots\dots(3.13) \\ \underline{X}_u \rightarrow \Omega_c \text{ si } d(\underline{X}_u, \underline{m}_c) \leq T_c \text{ et } k_c = \max_{r=1,M} k_r \rangle k' \end{array} \right.$$

3.3.2.4 la prise de décision par les k-ppv pour les nouvelles observations

Il s'agit maintenant de valider la méthode de décision les k-ppv mises en place. Comme précisé au début de ce chapitre, nous utilisons les mesures ne faisant pas partie de l'ensemble d'apprentissage. Ainsi, le diagnostic final de la règle des k – ppv est établi pour les même nouvelles observations illustré dans le tableau 3.2.on utilise les deux distance la distance euclidienne et la distance mahanalobis.

Dans ce travail nous avons choisis trois procédures de décision par les k-ppv. La première, c'est la classification du vecteur test point par point. la seconde classification se contente de vérifier deux points au lieu de tout le vecteur test afin de prendre une décision. La troisième est la classification par le centre de gravité de vecteur test.

- La décision point par point
- La décision par deux points
- La décision par le centre de gravité

➤ *La décision point par point*

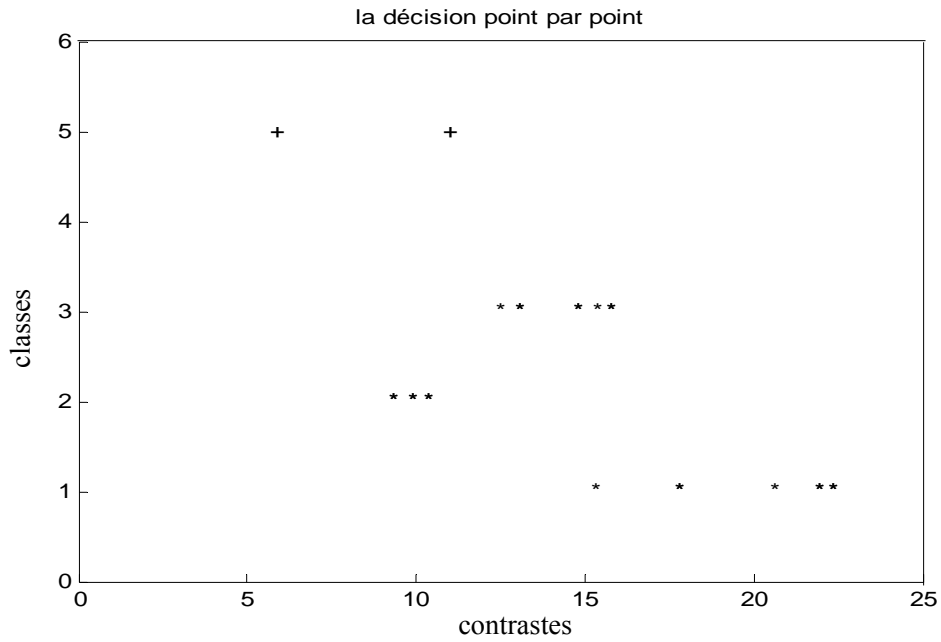


Figure 3.5 la décision point par point par la distance euclidienne

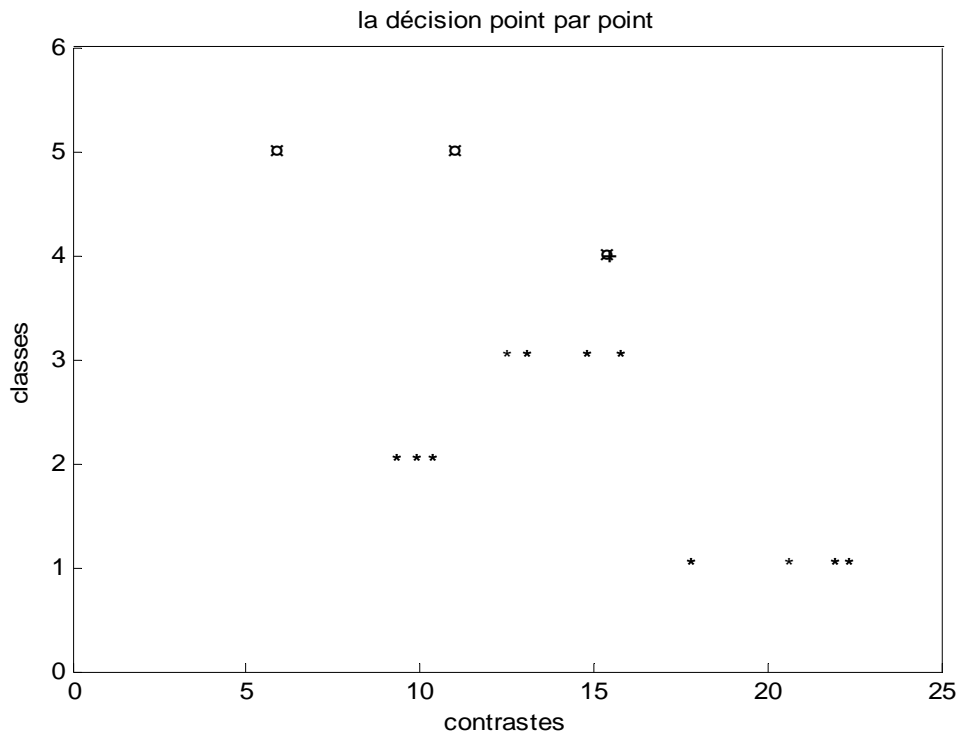


Figure 3.6 la décision point par point par la distance Mahanalobis

➤ *La décision par deux points*

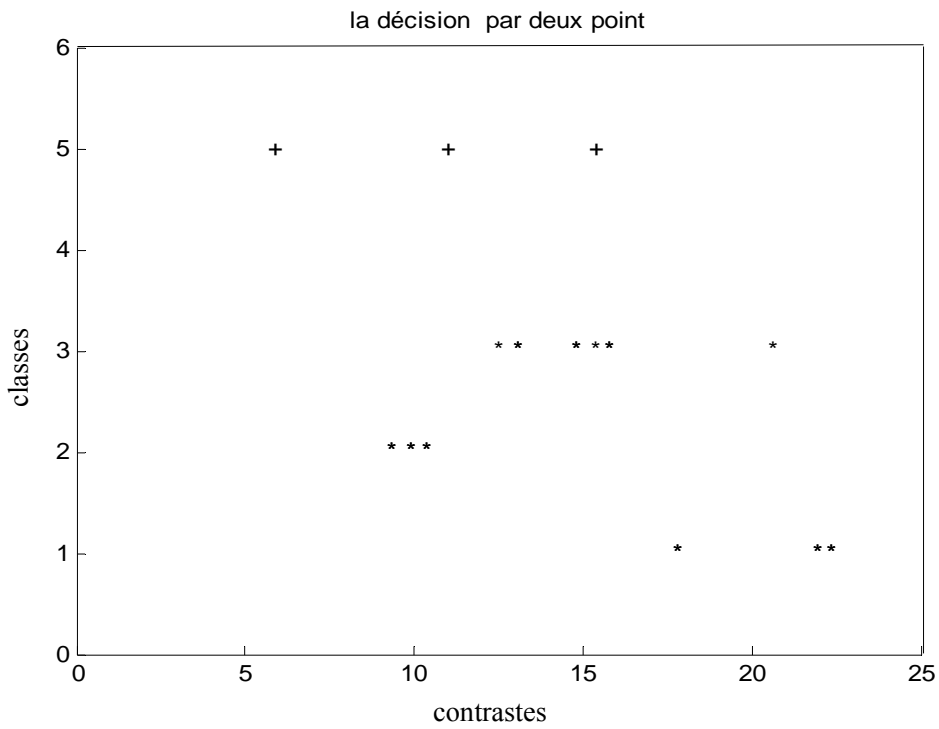


Figure 3.7 La décision par deux point satisfait par la distance euclidienne

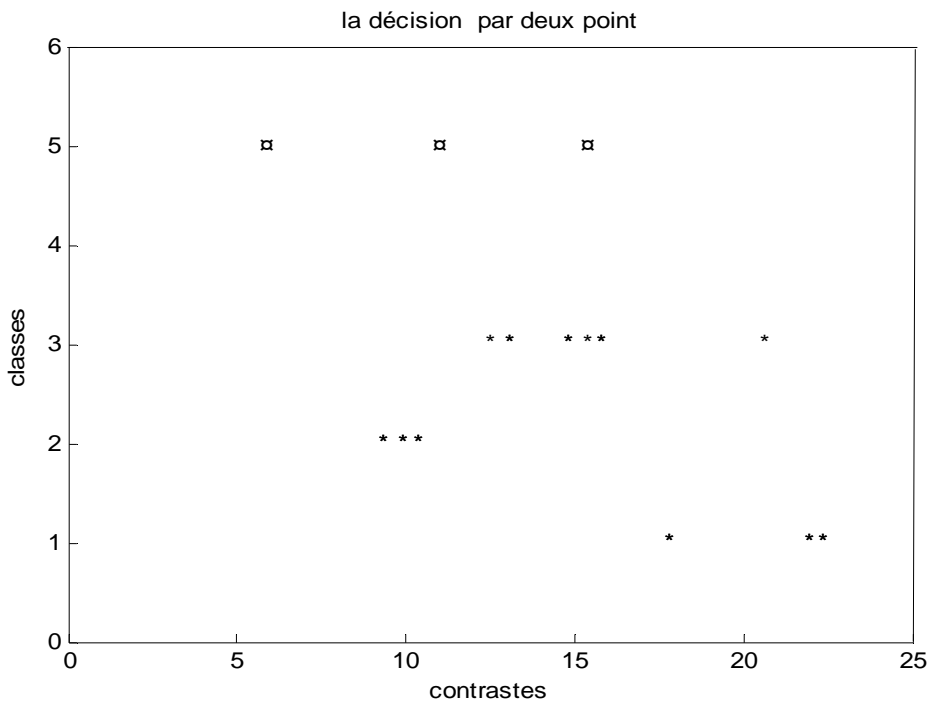


Figure 3.8 La décision par deux point satisfait par la distance Mahalanobis

➤ *La décision par le centre de gravité*

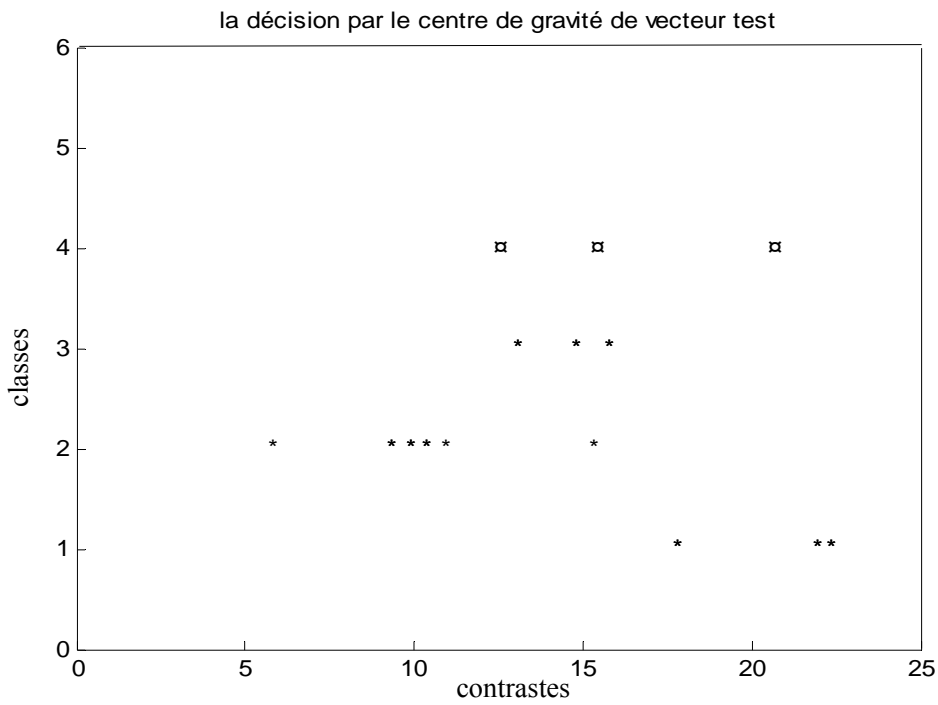


Figure 3.9 La décision par le centre de gravité par la distance euclidienne

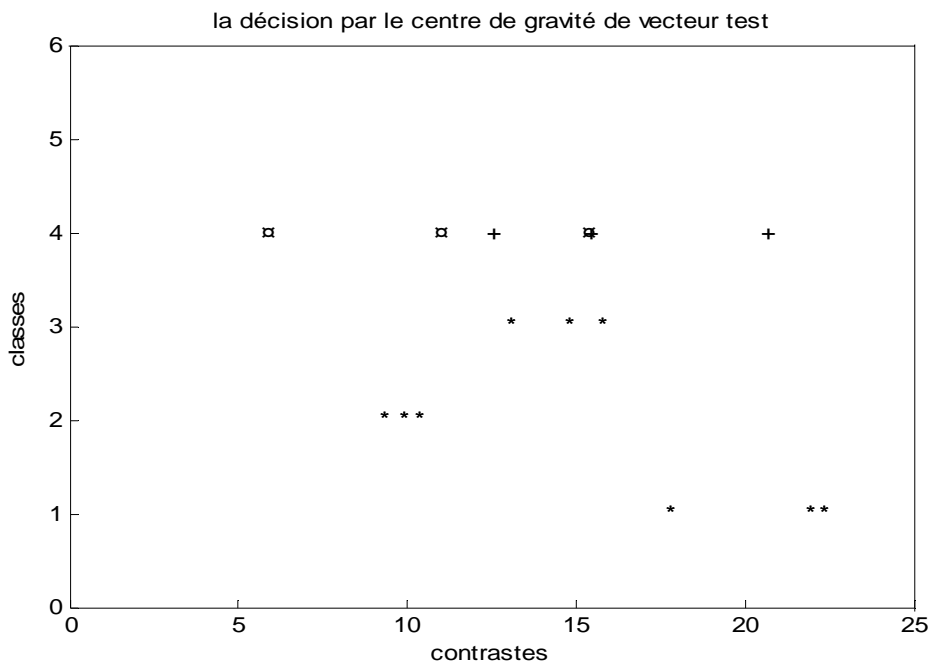


Figure 3.10 La décision par le centre de gravité par la distance Mahalanobis

les figure 3.5 présente la classification de deux nouvelles observation Xu1 et Xu2 par la procédure « point par point » ceci par l'utilisation de la distance euclidienne. Nous constatons que le premier point du vecteur Xu1 est affecté a la classe Ω_3 qui correspond au défaut rotorique mais les deux autres points sont rejetés en ambiguïté, de ce fait la décision est la rejeté de vecteur par ambiguïté avec un taux d'erreur 33% (le taux d'erreur correspond au nombre des points mal classés sur le nombre des points du vecteur observé), en revanche la deuxième point du vecteur Xu2 est affecté à la classe Ω_2 c'est à dire le défaut statorique et les autres points sont rejetés en distance, par conséquent le vecteur est rejeté en distance avec un taux d'erreur 33%.

La figure 3.6 présente la classification de deux nouvelles observations Xu1 et Xu2 par la procédure « point par point » mais cette fois-ci par l'utilisation de la distance Mahanalobis. Nous constatons que la décision est le même que la distance euclidienne, c'est-à-dire le vecteur Xu1 est rejeté en ambiguïté et le vecteur Xu2 est rejeté en distance. Cette similitude des résultats entre les deux distances s'explique par le principe de Fisher, la séparation entre les classes est plus grand et les point de la même classe est plus proche. Les points d'apprentissage sont issus du principe Fisher de plus fort contraste.

Les figures 3.7et 3.8 illustrent les cas de deux points quelconques prisent des vecteurs Xu1 ou Xu2, par conséquent le vecteur Xu1 est affecté à la classe Ω_3 et donc la classe de défaut rotorique mais avec un taux d'erreur 66% et le vecteur Xu2 est rejeté en distance avec un taux d'erreur de 33%.

Dans la troisième étape, nous avons utilisés le centre de gravité de chaque vecteur. Les figures 3.9 et 3.10 illustrent la classification des vecteurs dans les différentes classes possible. Nous observons que le vecteur Xu1 est rejeté en ambiguïté, ce résultat correspond à la classification point par point comme nous l'avons vu précédemment et le vecteur Xu2 est rejeté en ambiguïté avec un taux d'erreur 66% ceci est à cause de l'éloignement entre les points de vecteur Xu2.

le tableau (3.3) résumé l'affectation des nouvelles observations aux différentes classes de l'ensemble d'apprentissage : la classe défaut roulement Ω_1 , la classe défaut statorique Ω_2 , la classe défaut rotorique Ω_3 , la classe de rejet en ambiguïté Ω_a et la classe de rejet en distance Ω_d .

	La décision par la distance Euclidienne				La décision par la distance Mahalanobis			
	Par Point	Le vecteur	Deux points	Le centre de gravité	Par point	Le vecteur	Deux points	Le centre de gravité
Xu1	$point 1 \rightarrow \Omega_3$ $point 2 \rightarrow \Omega_a$ $point 3 \rightarrow \Omega_a$	Ω_a	Ω_3	Ω_a	$point 1 \rightarrow \Omega_3$ $point 2 \rightarrow \Omega_1$ $point 3 \rightarrow \Omega_a$	Ω_a	Ω_3	Ω_a
Xu2	$point 1 \rightarrow \Omega_d$ $point 2 \rightarrow \Omega_1$ $point 3 \rightarrow \Omega_d$	Ω_d	Ω_d	Ω_2	$point 1 \rightarrow \Omega_d$ $point 2 \rightarrow \Omega_a$ $point 3 \rightarrow \Omega_d$	Ω_d	Ω_d	Ω_a

3.3.2.5 Classification non exclusive

Le fait d'utiliser une classification exclusive renseigne sur l'appartenance d'une observation à une classe, mais ne donne aucune indication sur sa localisation dans la zone de la classe concernée. En effet, une observation située au centre d'une classe et une observation située en bordure (proche du seuil limite de rejet en distance) de celle-ci n'ont pas la même représentativité, et pourtant, ces deux observations sont affectées de la même manière à cette classe.

De même, une observation située en bordure de classes contiguës, ou entre deux classes disjointes, peut signifier un état transitoire entre ces deux classes. Quantifier cette information permet de donner la probabilité d'appartenance à la classe retenue et, de ce fait, fournit un degré de confiance dans la classification réalisée. Cela permet aussi de suivre l'évolution d'observations successives à l'intérieur d'une classe, c'est-à-dire de donner une mesure de l'évolution du système vis-à-vis des états connus.

Pour ces raisons, la mise en place d'une règle de décision non exclusive peut être intéressante. Nous disposerons de plus d'informations sur l'appartenance d'une observation, ce qui permettra d'une part, de donner un diagnostic plus en adéquation avec la réalité d'évolution du processus et d'autre part, d'associer un degré de confiance à la classification.

Le passage de l'affectation exclusive à l'affectation non exclusive doit se caractériser par une gradation entre la non appartenance absolue et l'appartenance totale. Pour répondre à ces contraintes, plusieurs solutions sont envisageables. Une des solutions possibles, qui n'a pas été utilisée par la suite, se base sur l'utilisation d'une fonction gaussienne centrée sur le centre de gravité de la classe. L'expression de cette fonction pour notre problème prendrait la forme suivante :

$$f_c(\underline{X}_u) = \frac{1}{\sigma_c \cdot \sqrt{2\pi}} \cdot e^{-d\left(\frac{\underline{X}_u, \underline{m}_c}{2\sigma_c}\right)^2} \dots\dots\dots(3.14)$$

Avec σ_c , l'écart type de la classe Ω_c .

En supposant que les classes soient caractérisées par une distribution gaussienne (ce qui est généralement le cas), il est possible de fixer d_{seuil_c} , correspondant à une appartenance totale, à trois fois l'écart type (ce qui englobe 99% des observations de la classe) et ensuite, de fixer le seuil T_c pour obtenir une appartenance décroissante jusqu'au rejet en distance.

Ceci est représenté sur la figure 3.11. Ainsi, la fonction d'appartenance de la classe Ω_c pourrait se définir comme suit :

$$\text{Si } d(\underline{X}_u, \underline{m}_c) \leq 3 \times \sigma \quad \text{alors } \mu_{\Omega_c}(\underline{X}_u) = 1$$

Si $3 \times \sigma \leq d(\underline{X}_u, \underline{m}_c) \leq \beta \cdot d_{seuil}$ alors $\mu_{\Omega_c}(\underline{X}_u) = f_c(\underline{X}_u)$ (3.15)
 0 sinon

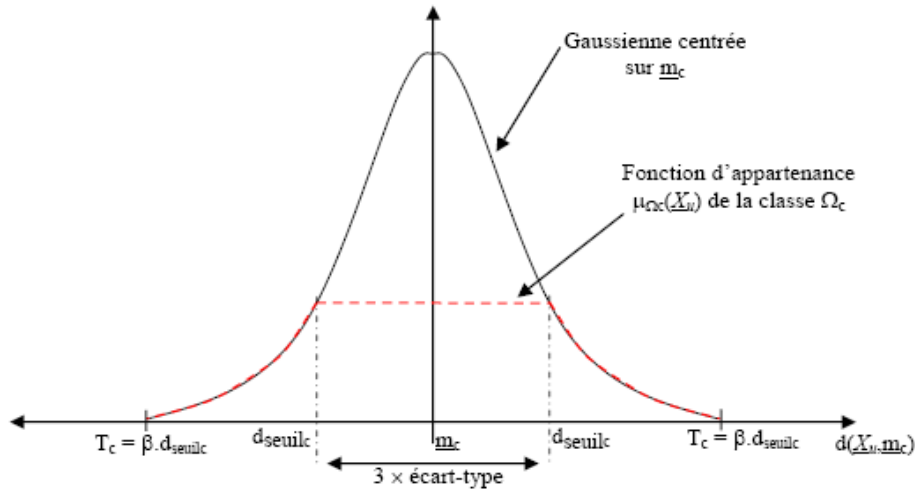


Figure 3.11 Fonction d'appartenance définie à l'aide d'une fonction gaussienne centrée sur m_c [ond 06].

La solution qui a été retenue par la suite définit une fonction d'appartenance à l'aide d'une fonction π multidimensionnelle. Cette fonction d'appartenance est décrite par la formulation suivante (3-16) :

$$\mu_{\Omega_c}(\underline{X}_u) = \begin{cases} 1 - (1 - \eta) \left(\frac{d(\underline{X}_u, \underline{m}_c)}{d_{seuil_c}} \right)^g & \text{si } 0 \leq d(\underline{X}_u, \underline{m}_c) \leq d_{seuil_c} \\ \eta \left(\frac{\beta}{\beta - 1} \right)^g \left(\frac{\beta \cdot d_{seuil_c} - d(\underline{X}_u, \underline{m}_c)}{\beta \cdot d_{seuil_c}} \right)^g & \text{si } d_{seuil_c} \leq d(\underline{X}_u, \underline{m}_c) \leq \beta \cdot d_{seuil_c} \dots(3.16) \\ 0 & \text{si non} \end{cases}$$

avec :

$d(\underline{X}_u, \underline{m}_c)$ est la distance entre une observation \underline{X}_u et le centre de gravité m_c de la classe Ω_c .

d_{seuil_c} est le paramètre qui définit la position des points d'inflexion. Il correspond à la distance maximale entre le centre de gravité de la classe Ω_c et le point le plus éloigné de cette même classe.

$$d_{seuil_c} = \max_{v=1, N_c} [d(\underline{X}_{cv}, \underline{m}_c)] \dots \dots \dots (3.17)$$

$$\mu_{\Omega_c}(\underline{X}_u) = \eta \text{ si } d(\underline{X}_u, \underline{m}_c) = d_{seuil_c} \dots \dots \dots (3.18)$$

η est le paramètre servant à fixer le plateau de la fonction π . β est un paramètre ($\beta \in \mathfrak{R}^{+*} \setminus 1$) qui permet de définir la largeur de bande de la fonction π (les points d'arrêt de la fonction π). g est un paramètre permettant de faire varier la pente de la fonction.

Cette règle n'autorise pas la notion de rejet d'ambiguïté. Pour diminuer le risque de mauvaise classification, il convient donc de définir de nouvelles règles de décision non exclusive permettant de contrôler ce type de rejet.

3.3.2.5.1 Rejet d'ambiguïté

Une règle de décision non exclusive incluant la notion de rejet d'ambiguïté doit être basée sur une analyse non binaire des fonctions d'appartenance.

L'idée la plus simple consiste à décider de l'affectation d'une observation \underline{X}_u aux différentes classes par comparaison des valeurs des fonctions $\mu_{\Omega_c}(\underline{X}_u)$ ($c=1, M$) à un seuil d'appartenance. Pour que ce rejet puisse être appliqué, il est nécessaire que les k plus proches voisins soient issus d'au moins deux classes différentes. Ainsi, une nouvelle observation est rejetée en ambiguïté. Les équations régissant ce rejet sont :

$$\text{Si } \mu_{\Omega_c}(\underline{X}_u) \neq 0 \dots \dots \dots (3.19)$$

Alors

$$\{\underline{X}_u \in \Omega_0 \text{ si } \mu_{\Omega_c}(\underline{X}_u) - \mu_{\Omega_r}(\underline{X}_u) \leq T_{ne} \dots \dots \dots (3.20)$$

Remarque : Plus $T_{ne} \rightarrow 0(1)$, moins (plus) les observations sont rejetées en ambiguïté.

Une information supplémentaire est apportée pour l'utilisateur concernant l'incertitude sur l'affectation d'une nouvelle observation. En effet, une règle peut être émise pour alerter sur la classification réalisée. Elle ne s'applique que lorsque les k plus proches voisins sont issus d'au moins deux classes. Elle indique que les degrés d'appartenance de l'observation aux classes sont inférieurs à l'incertitude, ce qui laisse penser que l'observation se situe dans une zone de l'espace faiblement représentée par les classes de X_a . La règle émet donc une alerte si :

$$\sum_{c=1}^M \mu_{\Omega_c}(\underline{X}_u) \leq 1 - \sum_{c=1}^M \mu_{\Omega_c}(\underline{X}_u) \dots\dots\dots (3.21)$$

3.3.2.5.2 Rejet d'appartenance

Un vecteur se situant loin d'une classe doit avoir un faible degré d'appartenance à cette classe. La notion de rejet de distance peut donc être exprimée en terme d'appartenance. Nous introduisons donc la notion de rejet d'appartenance, comparable au rejet de distance, mais spécialement dédiée aux règles de décision non exclusive. Il peut être défini comme suit :

- Une observation est rejetée en appartenance si son appartenance à toutes les classes possibles est inférieure à un seuil d'appartenance, noté T_{ra}

$$\underline{X}_u \in \Omega_d \text{ si } \mu_{\Omega_c}(\underline{X}_u) \leq T_{ra} \quad \forall c = 1, M \dots\dots\dots (3.22)$$

Une observation \underline{X}_u n'est pas rejetée en appartenance s'il existe au moins une classe Ω_c telle que $\mu_{\Omega_c}(\underline{X}_u) \geq T_{ra}$. L'équation 3.48, représentant le rejet en appartenance, peut donc s'écrire :

$$\underline{X}_u \in \Omega_d \text{ si } \max_{c=1, M}(\mu_{\Omega_c}(\underline{X}_u)) \leq T_{ra} \dots\dots\dots (3.23)$$

Finalement, la règle des $k - ppv$ non exclusive incluant les deux options de rejet peut être formulée de la manière suivante :

$$\left\{ \begin{array}{l} \underline{X}_u \in \Omega_c \text{ si } \mu_{\Omega_c}(\underline{X}_u) = \max_{r=1,M}(\mu_{\Omega_r}(\underline{X}_u)) \text{ et } \mu_{\Omega_c}(\underline{X}_u) \geq T_{ra} \text{ et 3.46 soit satisfaite} \\ \underline{X}_u \in \Omega_d \text{ si } \max_{c=1,M}(\mu_{\Omega_c}(\underline{X}_u)) \leq T_{ra} \\ \underline{X}_u \in \Omega_0 \text{ si } \mu_{\Omega_c}(\underline{X}_u) - \mu_{\Omega_r}(\underline{X}_u) \leq T_{ne} \\ \hspace{15em} r \neq c \end{array} \right. \dots\dots\dots (3.24)$$

La gradualité de la fonction d'appartenance permet d'une part, de quantifier la fiabilité de la classification et d'autre part, de suivre l'évolution des nouvelles observations dans l'espace de représentation pour éventuellement construire un pronostic.

3.3.2.5.3. La décision pour les nouvelles observations

La prise de décision par les k-plus proche voisin non exclusif est réalisée comme les k-ppv simple et nous avons utilisés les même nouvelles observations donc les vecteur motifs Xu1 et Xu2 de tableau 3.2.

➤ *La décision point par point*

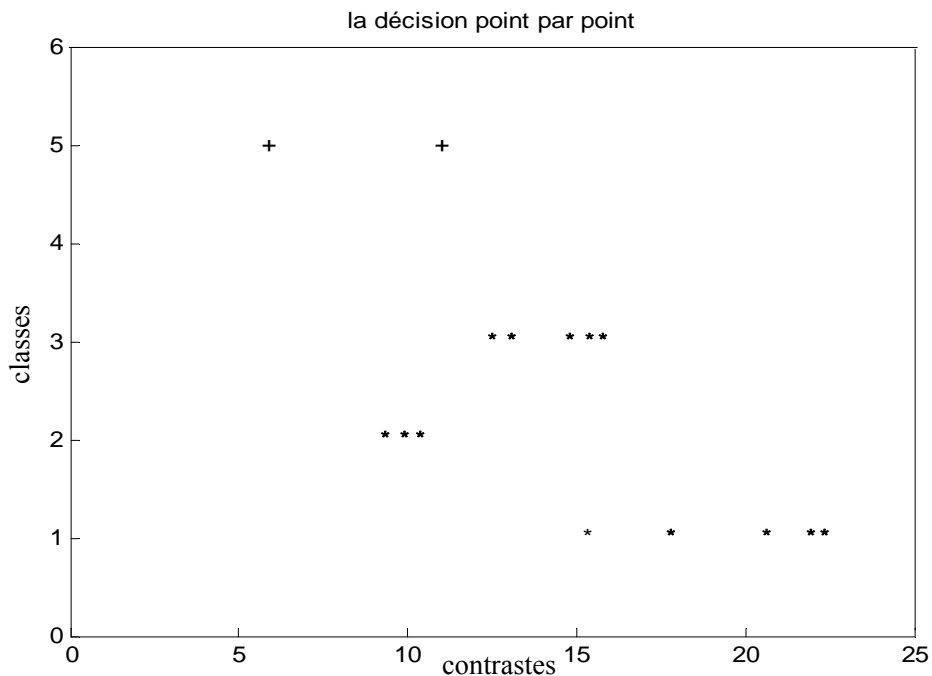


Figure 3.12 la décision point par point par la distance euclidienne

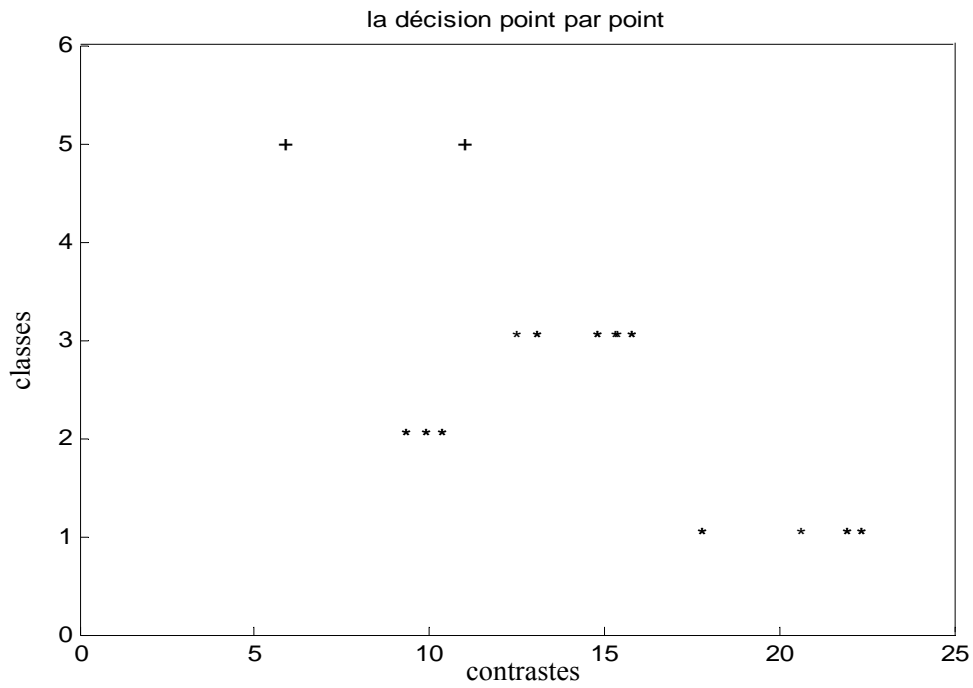


Figure 3.13 la décision point par point par la distance Mahanalobis

➤ *La décision par deux points*

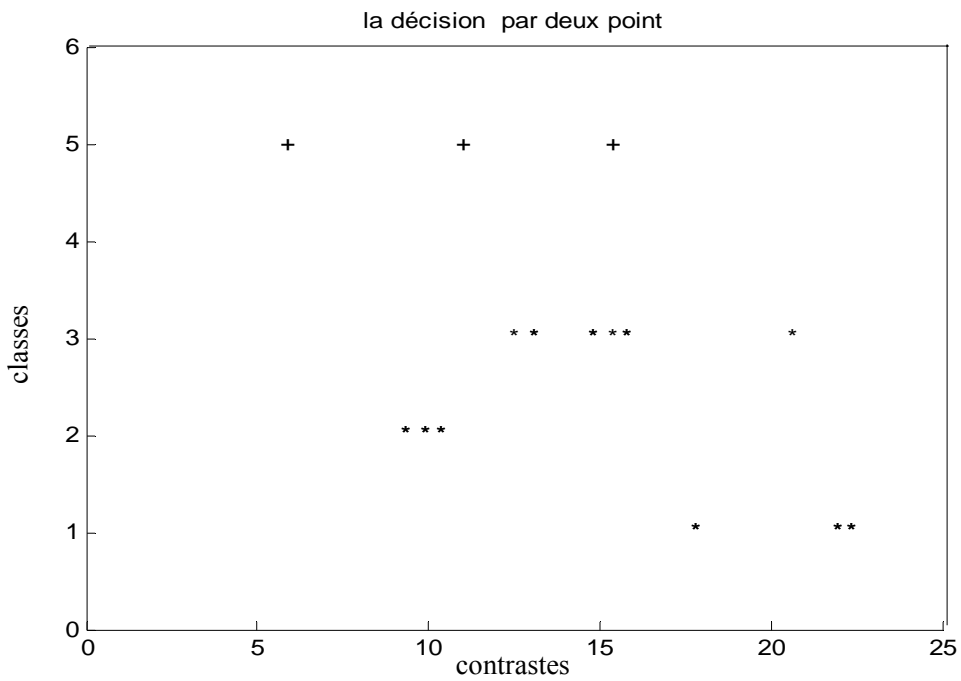


Figure 3.14 La décision par deux points par la distance euclidienne

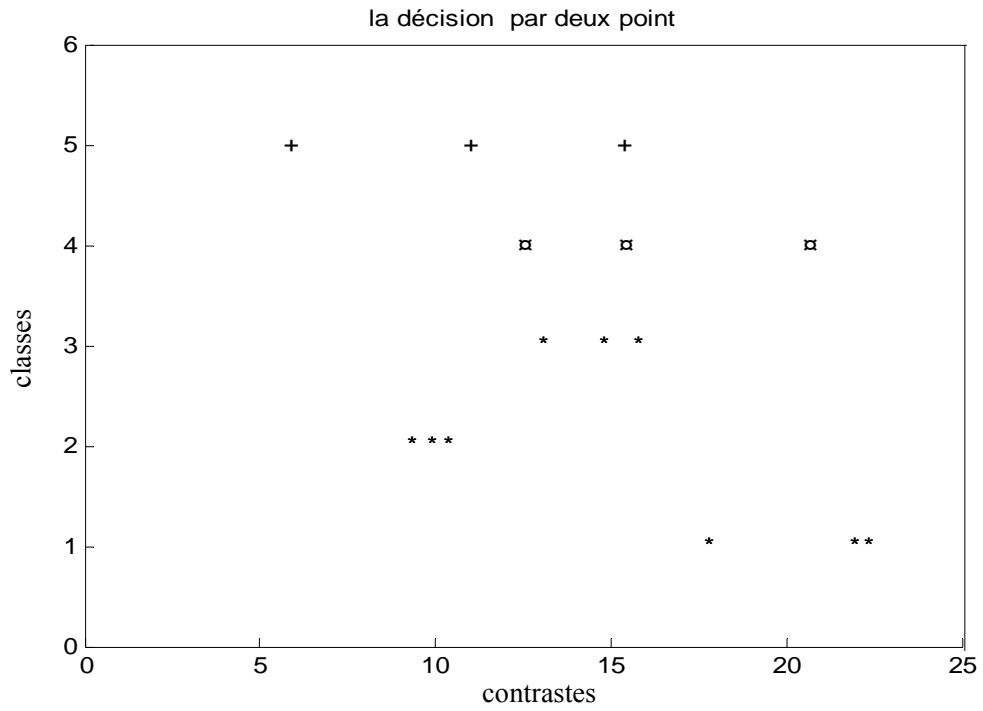


Figure 3.15 La décision par deux point par la distance Mahalanobis

➤ la décision par le centre de gravité

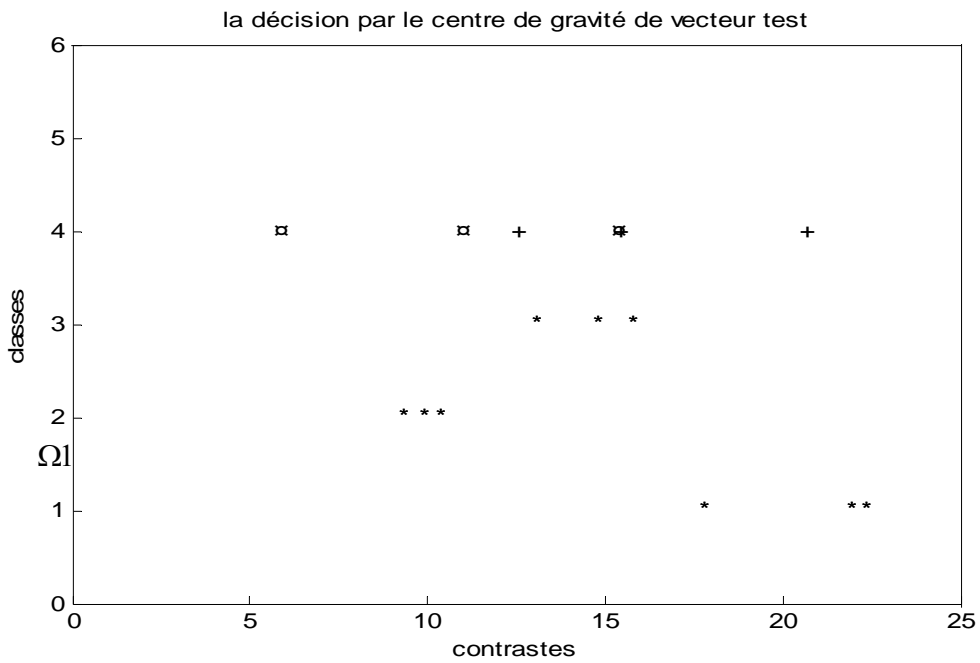


Figure 3.16 La décision par le centre de gravité par la distance euclidienne

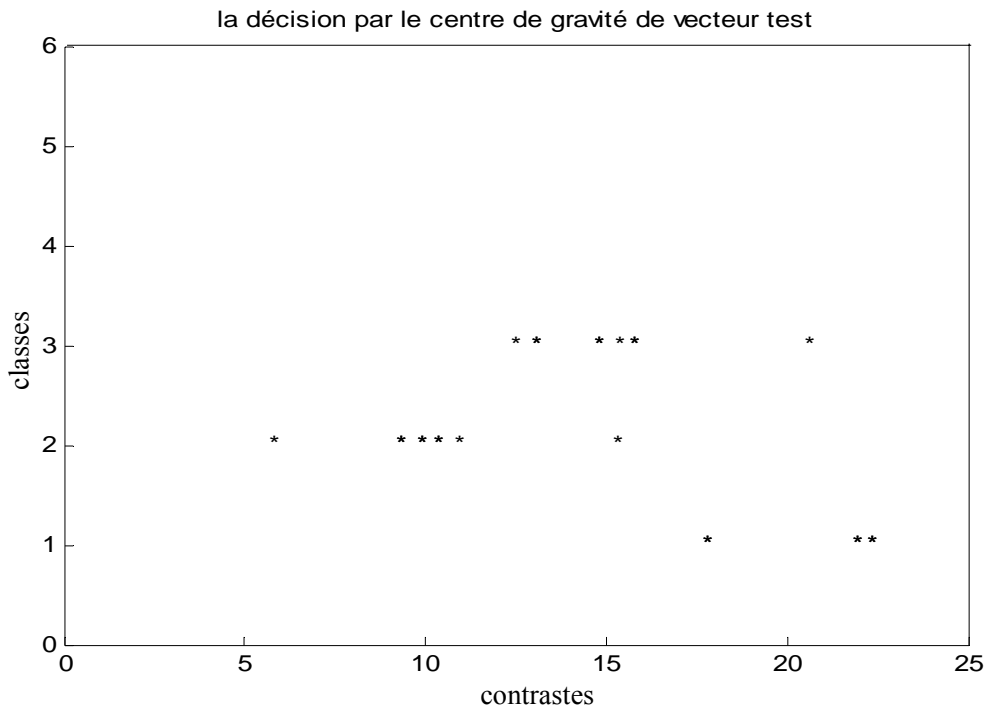


Figure 3.17 La décision par le centre de gravité par la distance Mahalanobis

Dans le cas de la classification par les k-ppv non exclusif, nous avons fait le même travail que les k-ppv simple sauf que nous avons définis une fonction d'appartenance f_p qui donne une performance lors de la prise de décision plus que les k-plus proche voisin simple. Le seuil de rejet en appartenance prend de fois plus que le seuil de rejet en distance cette caractéristique offre une décision plus précise.

Dans les figures 3.12 et 3.13, la classification point par point des nouvelles observations, le vecteur Xu_1 est affecté à la classe Ω_3 c'est-à-dire la classe de défaut rotorique avec un taux d'erreur 33% et le vecteur Xu_2 est rejeté en appartenance avec un taux d'erreur 33%.

Dans les figures 3.14 et 3.15, la classification par deux points est satisfaisante pour la prise de décision. Le vecteur Xu_1 est encore affecté à la classe Ω_3 avec un taux d'erreur 33% et le vecteur Xu_2 est rejeté en appartenance avec un taux d'erreur 33%.

Finalement les figures 3.16 et 3.17 illustrent la classification par le centre de gravité des vecteurs. Le vecteur Xu_1 est rejeté en ambiguïté et Le vecteur Xu_2 est affecté dans la classe Ω_2 .

Méthodes de la prise de décision

Nous observons que la classification par le centre de gravité de chaque vecteur est différente des autres classifications (point par point et deux point de vecteur) ceci est à cause de l'éloignement entre les points dans le même vecteur.

On note que les résultats obtenus par l'utilisation de la distance mahalanobis sont similaires de ceux obtenus par la distance euclidienne car les points de l'ensemble d'apprentissage est les points de plus fort contraste sont relativement proches les unes des autres.

Le tableau (3.4) résume la prise de décision pour la classification des nouvelles observations.

	La décision par la distance Euclidienne				La décision par la distance mahalanobis			
	Par Point	Le vecteur	Deux points	Le centre de gravité	Par point	Le vecteur	deux points	Le centre de gravité
Xu1	$point\ 1 \rightarrow \Omega_3$ $point\ 2 \rightarrow \Omega_1$ $point\ 3 \rightarrow \Omega_3$	Ω_3	Ω_3	Ω_a	$point\ 1 \rightarrow \Omega_3$ $point\ 2 \rightarrow \Omega_1$ $point\ 3 \rightarrow \Omega_3$	Ω_3	Ω_a	Ω_3
Xu2	$point\ 1 \rightarrow \Omega_d$ $point\ 2 \rightarrow \Omega_1$ $point\ 3 \rightarrow \Omega_d$	Ω_d	Ω_d	Ω_2	$point\ 1 \rightarrow \Omega_d$ $point\ 2 \rightarrow \Omega_3$ $point\ 3 \rightarrow \Omega_d$	Ω_d	Ω_d	Ω_2

3.4 Conclusion

L'objectif de ce chapitre est la mise en place d'une règle de décision pour valider la méthode de classification et pour rattacher toute nouvelle observation à l'une des classes existantes des défauts de la machine asynchrone (défaut roulement, défaut statorique et défaut rotorique). Nous avons utilisé deux méthodes de prise de décision, la première méthode est les réseaux de neurones, l'erreur quadratique moyenne obtenue pour les phases d'apprentissage et de test montre que cette méthode est performante à la classification mais elle a un inconvénient majeur en l'occurrence l'interprétation des résultats. Ceci nous a amené à proposer une deuxième méthode statistique à savoir les « k-plus proche voisin ». Nous avons appliqué la méthode pour deux types de classification : exclusif et non exclusif. Nous avons proposé trois procédures pour classer les nouvelles observations ; classification point par point, classification par les deux points et la classification par le centre de gravité. Deux distances ont été utilisées à savoir la distance euclidienne et la distance mahalanobis. Les résultats montrent que la classification par le centre de gravité du vecteur test est la plus simple et la plus performante. La classification par les k-plus proches voisins est la méthode la plus performante et la plus interprétable en termes de résultats de classification.



Conclusion générale

Conclusion générale

Ce travail a porté sur le diagnostic par la classification des défauts de la machine asynchrone. Trois types de défauts ont été traités à savoir le défaut statorique, le défaut rotorique et le défaut de roulement.

Les techniques et autres outils de diagnostic étant très nombreuses, nous les avons classées en deux approches: une approche signal basée sur le traitement de signal des grandeurs mesurables, à savoir le courant et la tension. Une approche système, en l'occurrence la classification par les représentations optimisées temps-fréquence.

Dans ce travail, la classification automatique des défauts de la machine asynchrone a été réalisée à partir de la représentation temps-fréquence dite "dépendante de la classe du signal". Nous avons montré que les RTF classiques ont des noyaux paramétriques et à priori prédéfinis donc inappropriés à la classification, d'où l'orientation vers le plan d'ambiguïté doppler-retard où toutes les RTF peuvent être dérivées par un choix approprié d'un noyau. Après avoir montré les limites du noyau DCS de moyenne carrée et du noyau discriminant linéaire, à cause respectivement du non prise en considération de la variance interclasse pour le premier noyau et la nécessité des grands ensembles d'apprentissage pour le second noyau, nous avons opté pour le noyau discriminant de Fisher afin de palier à ces deux inconvénients.

La première phase du processus de classification est la phase d'extraction des points pertinents. La seconde phase consiste à la prise de décision concernant l'affectation d'un signal non classé. Deux méthodes de la prise de décision ont été utilisées à savoir les réseaux de neurones et les k-plus proches voisin. Cette dernière à été retenus vu ces performances et l'interprétation de ces résultats.



Bibliographie

Bibliographie

- [All 88] "*Présentation des avaries. Machines électriques tournantes*" cahier des préventions. CP2, Allianz, 1988.
- [Bas 02] M.J. Bastiaans, T. Alieva et L. Stankovic : On rotated time-frequency kernels. *IEEE Signal Processing Letters*, 9(11):378–381, 2002.
- [Ben 98] Bentounsi A., Nicolas A, " Transient and steady-state analysis of induction motors with cage faults", IMACS, Multiconference, CESA'98, April 1-4, pp.3511-3514, Tunisia, 1998.
- [Ben 98] Bentounsi A., Nicolas A., " On line diagnosis of defaults on squirrel cage motors using FEM ", *Magnetics, IEEE Transactions on* ,vol. 34 , n° 5 , septembre 1998, p. 3511-3514.
- [Boa 90] B. Boashash and P. O'Shea. A Methodology for Detection and Classification of Some Underwater Acoustic signals Using Time-Frequency Analysis Techniques. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 38(11) :2872_ 2883, Novembre 1990.
- [Boa 92] B. Boashash : *Time-Frequency Signal Analysis-Methods and Applications*. Longman-Cheshire, Melbourne and John Wiley Halsted Press, New York, 1992.
- [Boa 03] A. Boashash : *Time-Frequency Signal Analysis and Processing : A Comprehensive Reference*. Elsevier Ltd., Oxford, UK, 2003.*Time-Frequency Signal Analysis and Processing : A Comprehensive Reference*. Elsevier Ltd., Oxford, UK, 2003.
- [Bon 93] A. H. Bonnett, "Cause and analysis of Anti-Friction Bering Failures in A.C Induction Motors" *IEEE Transactions on Industry Application*, pp 14 - 23, Sept/Oct 1993.
- [Cas 03] R. Casimir, "Diagnostic des défauts des machines asynchrones par reconnaissance des formes" *Thèse de l'Ecole Centrale de Lyon, CEGELY, 2003*.

- [**Cha 98**] E. Chassande-Mottin : Méthodes de réallocation dans le plan temps-fréquence pour l'analyse et le traitement de signaux non-stationnaires. Thèse de doctorat, Université de Cergy-Pontoise, 1998.
- [**Che 02**] V.C. Chen et Hao Ling : Time-Frequency Transforms for Radar Imaging and Signal Analysis. Artech House, Boston, London, janvier 2002.
- [**Cla 80**] Claasen T. A. C. and. Mecklenbräuker W. F. G, "The Wigner distribution—A tool for timefrequency signal analysis—Part III: Relations with other time frequency signal transformations," *Philips J. Res.*, vol. 35, no. 6, pp. 372–389, 1980.
- [**Coh 66**] L. Cohen. Generalised Phase-Space Distribution Functions. *Journal of Mathematical Physics*, 7(5) :781 _ 786, Mai 1966.
- [**Coh 95**] L. Cohen : Time-Frequency Analysis. Englewood Cliffs, NJ : Prentice-Hall, 1995, 1995.
- [**Dav 98**] Davy M. and Doncarli C., "Optimal kernels of time-frequency representations for signal classification," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, 1998, pp. 581– 584.
- [**Dav 00**] Davy M., "Noyaux optimisés pour la classification dans le plan temps-fréquence- Proposition d'un algorithme constructif et d'une référence bayésienne basée sur les méthodes MCMC Application au diagnostic d'enceintes acoustiques", thèses de doctorat, Ecole Centrale de Nantes, 2000.
- [**Der 00**] Derrhi M., Modélisation de la machine asynchrone par réseaux de perméances, validation par le diagnostic, Thèse de doctorat, Université de Picardie Jules Verne, 2000.
- [**Dev 77**] P. A. Devijver, "Reconnaissance des formes par la méthode des plus proches voisins", Thèse de 3ème cycle, Université Paris 6, juin 1977.
- [**Dev 01**] V. Devanneaux, H. Kabbaj, B. Dagues et J. Faucher, "An accurate model of squirrel cage induction machines under static, dynamic or mixed eccentricity" *Proc. IEEE SDEMPED 2001 – Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, pp.121-126, 1-3 Septembre 2001 Grado, Italie

- [Dud 73] Duda R. O. and Hart P.E, "Pattern Classification and Scene Analysis. New York: Wiley, 1973.
- [Dub 90] B. Dubuisson, Diagnostic et Reconnaissance des Formes. Traité des nouvelles technologies. Série diagnostic et Maintenance. Hermès, Paris,1990
- [Dub 90a] B. Dubuisson, "Decision with reject options", Eusipco, Barcelone, 1990.
- [Dud 01] Duda R., Hart P., Stork D., Pattern Classification, Willey Interscience Second edition, 2001.
- [Elk 92] N. M. Elkasabgy, A. R. eastham, G. E. Dawson, "Detection of broken bars in the cage rotor on an induction machine" *IEEE Transactions on Industry Applications*, Vol.28, no.1, Part 1, pp. 165-171, Jan.-Fév. 1992
- [Fil 95] F. Filippetti, G. Fanceschini, C. tassoni, "Neural networks aided on-line diagnostics of induction motor rotor faults" *IEEE Transactions on Industry Applications*, Vol. 31, Issue 4, pp. 892-899, July/August 1995.
- [Fis 36] Fisher R.A., The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v.7, p. 179-188 Cambridge University Press, 1936
- [Fix 51] Fix E., Hodges,J.L., Discriminatory analysis. Non-parametric discrimination. USAF School of Aviation Medicine, Randolph Field, TX, 4 Technical Report 4, 1951.
- [Fla 88] P. Flandrin. A Time-Frequency formulation of Optimal Detection. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 36(9) :1377 _ 1384, Septembre 1988.
- [Fla 93] P. Flandrin : Temps-Fréquence. Editions Hermes, Paris, 1993.
- [Fu 74] K. S. Fu, "*Syntactic Methods in Pattern Recognition*" Academic Press, Inc, New York, 1974.
- [Gen 07] Gentil S., Celse B., Charbonnier S., Cocquempot V., Hamelin F., Lesecq S., Maquin D., Montmain J., Ragot J., Sauter D., *Supervision des Procédés Complexes*, Lavoisier, 2007.

- [**Gil 99**] Gillespie B. W. and Atlas L., "Optimization of time and frequency resolution for radar transmitter identification," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 3, 1999, pp. 1341–1344.
- [**Gus 92**] Gustafson D., Kessel W., Fuzzy Clustering with a Fuzzy Covariance Matrix. IEEE Fuzzy Models for Pattern Recognition, 1992 reprinted from Proc. IEEE CDC 1979.
- [**Hay 97**] Haykin S. and Bhattacharya T. K., "Modular learning strategy for signal detection in a nonstationary environment," IEEE Trans. Signal Processing, vol. 45, pp. 1619–1637, June 1997.
- [**Hei 95**] C. Heitz. Optimum time-frequency representations for the classification and detection of signals. Applied Signal Processing, 3 :124 _ 143, 1995.
- [**Hel 70**] M. E. Hellman, "The nearest neighbor classification rule with a reject option", IEEE. Trans on System, Man and Cybernetics, pp 179 – 185, Vol 6, no 3, July 1970.
- [**Hla 97**] Hlawatsch F, Papandreou-Suppappola A., and. Boudreaux-Bartels G. F. The Hyperbolic Class of Quadratic Time-Frequency Representations - Part II : Subclasses, Intersection with the Affine and Power Class, Regularity and Unitarity. IEEE Transactions on Signal Processing, 45(2) :303 _ 315, Février 1997.
- [**Hla 05**] F. Hlawatsch et F. Auger : Temps-Fréquence : Concepts et Outils. Traite IC2, serie Traitement du signal et de l'image, Hermes, Lavoisier, Paris, France, 2005.
- [**Ioa 03**] C. IOANA : Contribution à la caractérisation optimale des structures temps-fréquence non-linéaires. These de doctorat, Universite de Bretagne Occidentale, septembre 2003
- [**Ker 84**] L. Kerszenbaum et C. F. Landy, "The existence of large inter bar currents in the three phase squirrel cage motors with rotor-bar and en-ring faults" IEEE Transactions on Power Apparatus Systems, Vol. PAS-103, n°7, pp.1854- 1862, July 1984.

- [Leb 06]** A. Lebaroud, "Modélisation et techniques orientées vers le diagnostic de la machine asynchrone associée à des sources variables", thèses de doctorat, Université de Constantine, 2006.
- [Leb 08]** Lebaroud. A, Clerc. G, "Classification of Induction Machine Faults by Optimal Time-Frequency Representations", IEEE Transactions, Industrial Electronics *vol. 55*, no. 12, pp. 4290-4298, December 2008.
- [Leb 09]** Lebaroud. A, Clerc. G, "Accurate Diagnosis of Induction Machine Faults Using Optimal Time-Frequency Representations", ELSEVIER, Engineering Applications of Artificial Intelligence, Volume 22, Issues 4-5, June 2009, Pages 815-822
- [Lem 95]** O. Lemoine : Détection de signaux non stationnaires par représentation tempsfréquence. These de doctorat, Universite de Nice - Sophia antipolis, septembre 1995.
- [Mal 00]** S. Mallat : Une Exploration Des Signaux En Ondelettes. Ellipses, Paris, France, 2000.
- [Naï 00]** M. S. Naït-Saïd, M. E. H. Benbouzid et A. Benchaib, "Detection of broken bars in induction motors using an extended kalman filter for rotor resistance sensorless estimation" *IEEE Transactions on Energy Conversion, Vol.15, no 1, pp.66-70*, Mars 2000.
- [Nej 00]** Nejari H., Benbouzid M.E.H., " Monitoring and diagnosis of induction motors electrical faults using a current Park's vector pattern learning approach ", *Industry Applications, IEEE Transactions on*, vol. 36 , n° : 3 , mai/juin 2000. p.730-735.
- [O'do 85]** P. O'Donnell, "Report of large reliability survey of industrial and commercial installations, Part 1", *IEEE Transaction on Industry Applications*, Vol. IA 21, N° 4, july/august 1985 pp 853 – 864.
- [Ond 06]** O. Ondel, "Diagnostique par Reconnaissance des Formes: Application a un Ensemble Convertisseur-Machine ", Thèse de Doctorat, L'Ecole Centrale de Lyon, October 2006.

- [Pap 02]** A. Papandreou-Suppappola : Applications in Time-Frequency Signal Processing. CRC Press, 2002.
- [Rai 00]** B. Raison, "Détection et localisation de défaillances sur un entraînement électrique" *Thèse de doctorat de l'Institut National Polytechnique de Grenoble*, Septembre 2000.
- [Ric 98]** C. Richard : Une méthodologie pour la détection à structure imposée. Applications au plan temps-fréquence. These de doctorat, Université de Technologie de Troyes, decembre 1998.
- [Sch 95]** R. R. Schoen, T. G. Habetler, "Effect of time-varying loads on rotor fault detection in induction machines" *IEEE Transaction on Industry applications*, Vol. 31, Issue 4, pp.900-906, July-Aug. 1995.
- [Sch 99]** E. Schaeffer, "Diagnostic des machines asynchrones : modèles et outils paramétriques dédiés à la simulation et à la détection des défauts", Thèse de doctorat, Ecole Centrale de Nantes, 1999.
- [Sob 98]** T. J. Sobczyk, A. Izvorski, "Recognition of rotor eccentricity of induction motor based on the fourier spectra of phase currents" *Proc. ICEM'98*, pp. 408 – 413 , Vol 1, september 2 - 4 1998 Istanbul Turkey
- [Tho 95]** O. V. Thorsen, M. Dalva, "A survey of fault on induction motors in offshore oil industry, petrochemical industry, gas terminals, and oil refineries" *IEEE Transactions on Industry Applications*, Vol.31, no.5, pp.1186-1196, September 1995.
- [Tom 76]** I. Tomek, "A generalisation of the kNN rule", *IEEE Trans on System, Man and Cybernetics*, Vol. 6, no 2, pp 121 – 126, february 1976.
- [Van 68]** H. L. Van Trees. Detection, Estimation, and Modulation Theory. Part I. John Wiley & Sons, 1968.
- [Wan 95]** Wang W. J. and McFadden P. D., "Application of orthogonal wavelets to early gear damage detection," *Mech. Syst. Signal Process.*, vol. 9, no. 5, pp. 497–507, 1995.

- [Wan 04] Wang M., Rowe G. I., and Mamishev A. V., "Classification of power quality events using optimal time-frequency representations—Part 2: application," IEEE Trans. Power Delivery, vol. 19, pp. 1496–1503, 2004.
- [Yah 95] H. Yahoui, J. Seetohul, G. Grellet, A. Jammal, "Detection of broken bar or end-ring fault in asynchronous machines by spectrum analysis of the observed electromagnetic torque through supply cable" *Revue Européenne de Diagnostic et Sûreté de Fonctionnement*, Vol.5, no.4, 1995

ملخص

الآلات الكهربائية أصبحت تستعمل أكثر فأكثر لمحركات اللاتزامنية بسبب صلابتها واستطاعتها الكتلية وكلفتها. صيانة وتشخيص الأعطال في هذه المحركات أصبح متطلب اقتصادي . من المهم الكشف المبكر للأعطال التي يمكن أن تظهر في هذه المحركات وبالتالي تطوير طريقة فعالة لمراقبة عملها والصيانة الوقائية.

نعتمد في دراستنا على طريقة آلية للكشف عن الأعطال في المحركات اللاتزامنية باستعمال التمثيل المثالي زمن-تواتر ثم أخذ القرار بالنسبة للظواهر الجديدة بطريقتين : الشبكات العصبية وk-أكثر قربا مجاورة .

الكلمات المفتاحية :

المحركات اللاتزامنية

تشخيص

التمثيل المثالي

زمن-تواتر

الشبكات العصبية

Résumé

Les entraînements électriques utilisent de plus en plus les moteurs asynchrones à cause de leur robustesse, de leur puissance massique et de leur coût. Leur maintenance et leur diagnostic deviennent donc un enjeu économique. Il est important de détecter de manière précoce les défauts qui peuvent apparaître dans ces machines et donc de développer des méthodes de surveillance de fonctionnement ou de maintenance préventive.

Notre approche repose sur l'automatisation d'une procédure de détection des défauts des moteurs asynchrone à partir de la représentation temps- fréquence optimisée puis la prise de décision pour les nouvelles observation par deux méthodes :les réseau de neurone et les k-plus proches voisin.

Mots clés :

les moteurs asynchrones

diagnostic

représentation temps- fréquence

les réseau de neurone

les k-plus proches voisin

Abstract

Electrical tractions use more and more the asynchronous machines because of their robustness, their power weight ratio. Their maintenance as well as their diagnosis then became an economic state. It is important to early detect the faults likely to appear in those motors and therefor to implement a preventive maintenance.

Our approach is based about a procedure of the automation for classification of induction machine faults by optimal time-frequency representation then the decision for the news observations used two methods :the neurons networks and the k-plus nearest neighbor.

Key words:

asynchronous machines

diagnosis

time-frequency representations

neurons networks

k-plus nearest neighbor