

TESIS DOCTORAL

Diagnosis de modos de fallo en máquinas rotativas mediante técnicas basadas en redes neuronales artificiales y en la transformación con ondículas.

-Fco. Javier Sanz Corretge, Ingeniero Superior Industrial.

-Escuela Técnica Superior de Ingenieros Industriales de la Universidad Nacional de Educación a Distancia.

-Año 2007.

- Departamento de Construcción y Fabricación, Escuela Técnica Superior de Ingenieros Industriales.
- Título: Diagnóstico de modos de fallo en máquinas rotativas mediante técnicas basadas en redes neuronales artificiales y en la transformación con ondículas.
- Autor: Fco. Javier Sanz Corretge, Ingeniero Superior Industrial.
- Directores de Tesis: Ricardo Perera Velamazán y Consuelo Huerta Gómez.
- Tutor: Juan José Benito Muñoz.
- Escuela Técnica Superior de Ingenieros Industriales de la Universidad Nacional de Educación a Distancia.
- Año 2007.

AGRADECIMIENTOS

El trabajo presentado en esta tesis fue realizado con la ayuda y colaboración de los profesores:

Dr. Ricardo Perera y Dra. Consuelo Huerta.

A ambos me gustaría expresarles mi más sincero agradecimiento por su profunda implicación en este trabajo. Su aportación de conocimientos y experiencia ha resultado vital para el desarrollo del mismo. Desde mi más sincera opinión, creo que son dos de las personas más competentes con las que me haya podido encontrar, he aprendido mucho con ellos y no solo en el plano técnico sino también en el personal, creo que puedo decir que con este trabajo, no sólo completo una tesis sino que además tengo dos amigos.

Por otro lado quiero agradecer todo el apoyo recibido por mis padres y familia. Esta tesis, sin duda, también se debe a ellos.

Por último me gustaría agradecer al profesor Dr. Juan José Benito todo el apoyo ofrecido con su tutela desde la UNED.

Pamplona, 17 de mayo de 2007.

Javier Sanz Corretge.

Para María.

ÍNDICE

LISTA DE SIMBOLOS, ABREVIATURAS Y SIGLAS.	8
LISTA DE TABLAS Y FIGURAS.	10
1. INTRODUCCIÓN.	14
1.1. ANTECEDENTES.	14
1.2. ESTADO DEL ARTE.	15
1.3. OBJETIVOS.	18
1.4. ORGANIZACIÓN DE LA TESIS.	20
2. FALLOS EN MÁQUINAS ROTATIVAS: DETECCIÓN Y DIAGNÓSTICO.	22
2.1 INTRODUCCIÓN.	22
2.2 CLASIFICACIÓN Y DESCRIPCIÓN DE MODOS DE FALLO EN MÁQUINAS ROTATIVAS.	24
2.2.2. Clasificación.	24
2.2.3. Descripción.	25
2.3. PROCESOS VIBRO-ACÚSTICOS EN MÁQUINAS ROTATIVAS.	33
2.3.1. Clasificación de las distintas fuentes de vibración.	33
2.3.2. Vibraciones originadas por la acción de engrane.	34
2.2.2.1. Modelo matemático.	38
2.3.3. Vibraciones originadas por los rodamientos.	38
2.2.3.1. Modelo matemático.	41
2.4. CARACTERÍSTICAS DINÁMICAS DE LAS SEÑALES REGISTRADAS.	41
2.4.1. Transmisión de las vibraciones.	41
2.4.2. Rasgos fundamentales.	42
2.5. ANÁLISIS DE LA SEÑAL REGISTRADA.	42
2.5.1. Análisis en frecuencia.	43
2.5.2. Análisis temporales-estadísticos.	44
2.5.3. Análisis tiempo-frecuencia.	45
3. SISTEMA DE DETECCIÓN Y DIAGNOSIS BASADO EN ONDÍCULAS Y REDES NEURONALES ARTIFICIALES.	46
3.1. INTRODUCCIÓN.	46
3.2. PRETRATAMIENTO DE SEÑAL CON ONDÍCULAS.	52
3.2.1. Definición de la transformación con ondículas.	53
3.2.2. Señales de detalle y aproximación.	56
3.2.3. Elección de la familia de ondículas.	61
3.2.3.1. Momentos nulos.	
3.2.3.2. Soporte.	
3.2.4. Caracterización y localización de la regularidad de una señal mediante WT.	65
3.2.4.1. Definición matemática de regularidad de una señal.	
3.2.4.2. Herramientas de cuantificación.	
3.2.4.3. Influencias sobre el escalograma.	
3.2.4.4. Influencias originadas por el ruido de fondo.	
3.2.5. Implementación mediante dos ejemplos numéricos.	78
3.3. DIAGNOSIS MEDIANTE REDES NEURONALES ARTIFICIALES.	87
3.3.1. Fundamentos de las redes neuronales artificiales.	88
3.3.1.1. Breve introducción biológica.	
3.3.1.2. Modelo de neurona artificial.	
3.3.1.3. Estructura de la red, modos de operación.	

3.3.1.4. Beneficios del empleo de redes neuronales artificiales.	
3.3.2. Aplicación de redes neuronales en problemas de regresión.	96
3.3.2.1. Implementación mediante redes RBF.	
3.3.2.2. Implementación mediante redes MLP.	
3.3.2.3. Comparativa de topologías MLP y RBF.	
3.3.3. Aplicación de redes neuronales en la detección noval de fallos en maquinaria rotativa.	105
3.3.3.1. Implementación con redes autoasociativas.	
3.4. DISEÑO DEL SISTEMA DE DETECCIÓN.	116
3.4.1. Clasificación de Rytter de los sistemas de diagnóstico.	117
3.4.2. Características del sistema de diagnóstico.	118
3.4.2.1. Objetivos y rangos de aplicación.	
3.4.2.2. Diseño.	
3.4.2.3. Estrategias planteadas.	
3.4.2.4. Ventajas e inconvenientes del sistema de diagnóstico propuesto.	
4. RESULTADOS.	129
4.1. Detección y diagnóstico de daño:	129
4.1.1. Experimento numérico.	130
4.1.1.1. Preprocesamiento de señal.	
4.1.1.2. Entrenamiento.	
4.1.1.3. Reconocimiento.	
4.1.2. Estación de bombeo con picaduras.	142
4.1.2.1. Preprocesamiento de señal.	
4.1.2.2. Entrenamiento.	
4.1.2.3. Reconocimiento.	
4.2. Cuantificación:	150
4.2.1. Modelo numérico de caja de engranajes.	150
4.2.1.1. Modelo matemático.	
4.2.1.2. Modelo dañado.	
4.2.1.3. Detección y diagnóstico de fallos.	
4.2.1.4. Cuantificación.	
5. CONCLUSIONES .	
5.1. Aportaciones.	176
5.2. Desarrollo futuro.	179
BIBLIOGRAFÍA.	181

LISTA DE SIMBOLOS, ABREVIATURAS Y SIGLAS

W_{ij}	<i>Sinapsis en las conexiones de las neuronas 'i' y 'j'.</i>
\bar{w}	<i>Vector de pesos sinápticos.</i>
w	<i>Frecuencia.</i>
\bar{x}_n	<i>Vector de entradas de patrón enésimo.</i>
\bar{y}_n	<i>Vector constituido por las salidas estimadas por la red en el patrón enésimo.</i>
\bar{t}_n	<i>Vector constituido por las salidas observadas en el patrón enésimo.</i>
θ_i	<i>Potencial de activación de la neurona 'i', factor de sesgo.</i>
$h_i(t)$	<i>Regla de propagación de la neurona 'i' en función de la señal 't'.</i>
$y_i(t)$	<i>Función de activación o transferencia de la neurona 'i' en función de la señal 't'.</i>
σ	<i>Varianza.</i>
t	<i>Tiempo.</i>
t_o	<i>Instante temporal concreto.</i>
$\bar{\lambda}$	<i>Vector de índices noales.</i>
$\bar{\delta}$	<i>Vector de índices umbrales.</i>
$\delta(t)$	<i>Función impulso de Dirac.</i>
α	<i>Exponente Hölder / Coeficiente Lipschitz.</i>
S	<i>Nivel de descomposición.</i>
j	<i>Nivel de descomposición.</i>
$E[\]$	<i>Operador esperanza matemática.</i>
E	<i>Función de error.</i>
$Wf[\]$	<i>Operador transformada con ondículas de la función 'f'.</i>
ψ_s	<i>Operador transformada con ondículas para la escala 's' y el tiempo 't'.</i>

$f(t)$	<i>Registro, histórico temporal.</i>
$\hat{f}(w)$	<i>Transformada de Fourier del histórico temporal.</i>
$A_j(t)$	<i>Subseñal de aproximación para un nivel de descomposición 'j'.</i>
$D_j(t)$	<i>Subseñal de detalle para un nivel de descomposición 'j'.</i>
a	<i>Dilatación de la ondícula.</i>
b	<i>Traslación de la ondícula.</i>
$\psi(t)$	<i>Ondícula madre.</i>
$\Psi(w)$	<i>Transformada de Fourier de la ondícula madre.</i>
Ψ_{ab}	<i>Ondícula correspondiente a los parámetros de dilatación y Traslación definidos por los parámetros 'a' y 'b'.</i>
C_{ab}	<i>Coefficiente de transformación con ondícula.</i>
$\phi_{jk}(t)$	<i>Función de escalado para el par (j, k) escala-tiempo.</i>
R	<i>Conjunto formado por los números reales.</i>
$l(R)^2$	<i>Espacio definido por las señales de energía finita.</i>
WT	<i>Transformación con ondículas.</i>
CWT	<i>Transformación continua con ondículas.</i>
DWT	<i>Transformación discreta con ondículas.</i>
FWT	<i>Transformada rápida con ondículas.</i>
SOM	<i>Mapas auto-organizados.</i>
RBF	<i>Funciones de base radial.</i>
RBFNN	<i>Redes neuronales basadas en funciones de base radial.</i>
MLP	<i>Perceptrón multicapa.</i>
MLPNN	<i>Red Neuronal basada en el perceptrón multicapa.</i>
TSA	<i>Promediado síncrono.</i>
OLS	<i>Algoritmo de mínimos cuadrados ortogonales.</i>
RP	<i>Algoritmo de propagación resiliente.</i>
PCA	<i>Análisis de componentes principales.</i>

LISTAS.

NN	<i>Redes neuronales.</i>
ANN	<i>Redes neuronales artificiales.</i>
AANN	<i>Redes neuronales autoasociativas / Redes de cuello de botella.</i>
TE	<i>Error en la transmisión.</i>
BPF	<i>Frecuencia de paso de cada elemento de rodadura en un rodamiento.</i>
FFT	<i>Transformada rápida de Fourier.</i>
STFT	<i>Transformada corta de Fourier.</i>
FTF	<i>Frecuencia de giro del retén.</i>
BSF	<i>Frecuencia de giro del rodillo.</i>
PSD	<i>Espectro de potencia</i>

LISTA DE TABLAS Y FIGURAS.

Tabla 1. Clasificación de modos de fallo en función de la frecuencia.

Tabla 2. Tabla que indica el número de señales detectadas como anómalas sobre 100 señales sanas.

Tabla 3. Tabla que indica el número de señales detectadas como anómalas sobre 100 señales degradadas con distinto soporte.

Tabla 4. Tabla que indica el número de señales detectadas como anómalas sobre 100 señales degradadas con distinta amplitud.

Tabla 5. Número de capas y de neuronas por capa para cada red neuronal. Las redes neuronales están definidas para cada sensor y nivel de descomposición.

Tabla 6. Tabla que indica el número de señales detectadas como anómalas sobre 100 señales sanas. Los resultados se expresan por nivel de descomposición y por caso de carga.

Tabla 7. Tabla que indica el número de señales detectadas como anómalas sobre 100 señales degradadas. Los resultados se expresan por nivel de descomposición y por caso de carga.

Tabla 8. Número de detecciones realizadas con las AANN para cada nivel de descomposición y estado de degradación.

Tabla 9. Errores promedio cometidos con las estimaciones en la red para cada nivel de degradación. Valores obtenidos para patrones entrenados.

Tabla 10. Errores promedio cometidos con las estimaciones en la red para cada nivel de degradación. Valores obtenidos para patrones no entrenados.

Figura 1. Fotos de engranajes con micropicaduras.

Figura 2. Fotos de picaduras en engranajes y rodamientos.

Figura 3. Picadura con desarrollo progresivo.

Figura 4. Engranaje con flancos descascarillados.

Figura 5. Rodamiento con problemas de spalling.

Figura 6. Subcaso de fatiga.

Figura 7. Foto de fallo de engranaje por fatiga a flexión.

Figura 8. Vibraciones ocasionadas por: (a) Variación en amplitud de la carga, (b) variación en dirección de la carga, (c) variación en la posición de la línea de acción sobre el flanco.

Figura 9. Separación de la punta del diente.

Figura 10. Efectos del fallo de un rodamiento en el histórico de una señal.

Figura 11. Esquema de la transformada corta de Fourier.

Figura 12. Esquema de la transformada con ondículas.

Figura 13. Esquema ilustrativo de la transformación WT.

Figura 14. Esquema de descomposición de una señal mediante la transformación DWT.

Figura 15. Esquema de descomposición de una señal en subseñales de detalle y de aproximación.

Figura 16. Pulso cuadrático con discontinuidades.

Figura 17. Transformación WT del pulso cuadrático empleando la familia Daubechies con distinto número de momentos nulos.

Figura 18. Perturbación generada en una señal temporal "suave".

Figura 19. Transformada CWT de la señal perturbada.

Figura 20. Proceso aleatorio de banda ancha.

Figura 21. Señal temporal ejemplo 1.

Figura 22. Transformada CWT de la señal del ejemplo 1.

Figura 23. Mapa de relieves 3D de los coeficientes CWT de la señal del ejemplo 1.

Figura 24. Señal temporal con ruido.

Figura 25. Detalle de los coeficientes CWT de la señal con ruido.

Figura 26. Ilustración de neurona biológica.

Figura 27. Modelo de neurona artificial.

Figura 28. Esquema de transporte de información en el sistema nervioso.

Figura 29. Esquema del modelo organizativo de las neuronas.

Figura 30. Esquema de red RBF.

Figura 31. Modelo de neurona empleada en las redes RBF.

Figura 32. Modelo de neurona MLP.

Figura 33. Red neuronal MLP.

Figura 34. Combinación de capas de neuronas MLP.

Figura 35. Redes neuronales autoasociativas.

Figura 36. Señales temporales del problema de detección noval.

Figura 37. Introducción de series temporales anómalas.

Figura 38. Residuos de señales “sanas” (a) y de las señales con anomalías (b).

Figura 39. Técnica TSA.

Figura 40. Diagrama de bloques del módulo de detección y localización.

Figura 41. Diagrama de bloques del módulo de cuantificación.

Figura 42. Señales vibratorias originales. El símbolo '□' Representa la señal 's₁', los símbolos '*' y 'o' representan a las señales 's₂' y 's₃' respectivamente.

Figura 43. Perturbaciones introducidas sobre las señales con ruido. El símbolo '□' Representa la señal 's₁', los símbolos '*' y 'o' representan a las señales 's₂' y 's₃' respectivamente.

Figura 44. Comparativa de los coeficientes DWT de entrada y salida de la red AANN, las partes (a) y (b) corresponden al nivel 2 de descomposición, las partes (c) y (d) al nivel 3, las partes (e) y (f) al nivel 4 y por último las partes (g) y (h) al nivel 5.

Figura 45. Vectores de índices noales. El símbolo '♦' representa el índice noval para cada coeficiente y la línea continua representa los valores umbrales. El recuadro (a) indica el 5 nivel de descomposición, y los recuadros (b), (c), (d) corresponden a los niveles 4, 3, 2 y 1 respectivamente.

Figura 46. Estación de bombeo en Lemmer, holanda.

Figura 47. Esquema de tren de potencia empleado como banco de ensayos de engranajes por Dalpiaz

Figura 48. Esquema de masas, muelles y amortiguadores empleado en el modelo de parámetros concentrados.

Figura 49. Modelo de rigidez variable para el par de engrane.

Figura 50. Espectro de potencia obtenido para la aceleración de la rueda (9).

Figura 51. Esquema representativo del modo de fallo.

Figura 52. Modelo de rigidez de engrane variable para distintos grados de fallo.

Figura 53. Espectro de potencia obtenido para la aceleración de la rueda (9) para distintos grados de degradación.

Figura 54. Espectro de potencia obtenido para la aceleración de la rueda (9) y espectro reconstruido a partir de los coeficientes DWT seleccionados

Figura 55. Representación de los residuos generados para el nivel D5 y por todos los patrones degradados para cada estado de degradación.

Figura 56. Representación de los residuos generados para el nivel D6 y por todos los patrones degradados para cada estado de degradación.

Figura 57. Estadísticas obtenidas para el nivel de descomposición D5, sobre los distintos patrones degradados.

Figura 58. Estadísticas obtenidas para el nivel de descomposición D6, sobre los distintos patrones degradados.

Figura 59. Esquema de generación de patrones como entrada a la red MLPNN.

Figura 60. Barras de errores cometidos con las estimaciones en la red para cada nivel de degradación. Valores obtenidos para patrones entrenados.

Figura 61. Barras de errores cometidos con las estimaciones en la red para cada nivel de degradación. Valores obtenidos para patrones no entrenados

CAPITULO.1

INTRODUCCIÓN

1.1. ANTECEDENTES.

Las máquinas rotativas juegan un papel crucial en aplicaciones industriales que afectan a sectores como el aeronáutico, el naval, el energético, etc.

La necesidad de incrementar la productividad de dichas máquinas y la fiabilidad de las mismas ha atraído el interés de la comunidad científica en lo referente al diseño de sistemas de diagnóstico que permitan efectuar predicciones en tiempo real sobre su estado de “salud”.

Se puede afirmar que la capacidad de realizar pronósticos ajustados sobre el estado de funcionamiento de dichas máquinas supone un aspecto clave en su mantenimiento. Así, hay que indicar que el propósito de cualquier sistema de diagnóstico es el análisis de información externa relevante de la maquinaria, generalmente a través de medidas indirectas como vibraciones, emisiones acústicas, control de partículas, etc., para establecer un plan de mantenimiento que o bien disminuya el coste de la reparación asociado a un modo de fallo, o bien alargue la vida efectiva de los componentes afectados por el fallo. De esta forma, se minimizarían los perjuicios provocados por una parada o por el funcionamiento anormal de la máquina.

La consecución del fin último que es la estimación de la vida útil de los componentes degradados o prognosis, implica no sólo la identificación y localización de daño sino también su cuantificación.

La necesidad de un sistema de diagnóstico eficiente se ve acrecentada en sectores industriales en los que el fallo pueda llegar a afectar a la seguridad física de las personas, especialmente en aquéllos relacionados con los medios de transporte (sector aeronáutico, ferroviario, automoción, etc.).

Como ejemplo de muestra se puede citar una investigación llevada a cabo en 1989 según la cual el 32% de los accidentes en helicópteros fue causado por defectos en el motor y en los componentes de transmisión de potencia [1].

1.2. ESTADO DEL ARTE

La implementación de tales sistemas de diagnóstico requiere habitualmente emplear un sistema sofisticado de instrumentación, adquisición y tratamiento de señal. Además, la mayoría de los sistemas de detección actuales no son capaces de realizar un control del estado de funcionamiento de la maquinaria en tiempo real. Generalmente las medidas registradas requieren de una fase de postproceso e interpretación por parte de expertos en tratamiento de señal [2]. Por otra parte, el cambio en las condiciones operativas afecta negativamente en numerosas ocasiones mediante el incremento del número de falsas alarmas detectadas con estos sistemas [3].

Parece interesante reseñar que el índice de detección de los sistemas comerciales actuales de diagnóstico, basados en análisis derivados de medidas mecánicas (vibración), es de solo un 60%, [3], por lo que si además se considera el índice de falsas alarmas, resulta evidente la baja eficacia de tales sistemas. Además, la mayoría de estos sistemas comerciales de diagnóstico únicamente generan indicaciones de alarma y generalmente ni localizan, ni cuantifican el daño en la maquinaria rotativa.

TIPOS DE MEDICIONES ACTUALES.

Actualmente se emplean tres alternativas de medición para inferir una diagnosis sobre el estado de “salud” de la maquinaria. Éstas son las mediciones acústicas, las dinámicas (vibratorias) y el control de partículas.

Las mediciones de respuesta dinámica (medidas vibratorias) se hallan muy extendidas desde el punto de vista comercial e industrial, a pesar de que los análisis acústicos y los basados en el control de partículas, históricamente, se han empleado y se siguen empleando con profusión; esto se debe, sin duda, a su facilidad aplicación y al coste asociado.

Además, entre las ventajas que presenta este tipo de medición, se pueden citar las siguientes:

1-Los análisis acústicos son mucho más sensibles al ruido de fondo del entorno de la maquinaria.

2-El análisis de partículas únicamente da resultados satisfactorios cuando hay arranque de materia. Por tanto, los análisis basados en el control de partículas no consideran fallos como el agrietamiento estructural o el desgaste del material por gripado, quedando su ámbito de detección más reducido que en el caso vibratorio que puede considerarse a tal efecto como más genérico.

Una discusión más detallada de las distintas alternativas se llevará a cabo en el capítulo 2 de esta tesis.

TIPOS DE ANÁLISIS DINÁMICOS ACTUALES.

En esta última década, se ha realizado un esfuerzo investigador notable para desarrollar técnicas de detección y diagnosis basadas en medidas vibratorias.

Estas técnicas se pueden aplicar en el dominio temporal, en el dominio de la frecuencia o en el dominio tiempo-frecuencia.

Los análisis más sencillos son aquéllos basados en medidas temporales. Estos sistemas emplean habitualmente medidas estadísticas efectuadas sobre las historias temporales, con el fin de establecer parámetros de tendencia que permitan de detectar la presencia de un modo de fallo.

Entre estas técnicas temporales se debe destacar, dado su éxito, el promediado síncrono (TSA) ([8] y [16]).

Las técnicas basadas en análisis realizados en frecuencia utilizan como rasgos fundamentales para fijar medidas de tendencia, las amplitudes de los armónicos dominantes en la respuesta así como los anchos de banda asociados. Sin embargo, este tipo de análisis no es capaz de detectar fallos locales ya que la transitoriedad de estos eventos en el dominio temporal queda enmascarada en el espectro obtenido al realizar la transformación en frecuencia. Por consiguiente, el seguimiento tanto de los anchos de banda como de los armónicos afectados se ve, en caso de fallos locales, seriamente dificultado.

Una alternativa para resolver este problema se encuentra en el empleo de análisis tiempo-frecuencia, los cuales ofrecen una medida de la distribución de energía de la señal en ambos dominios simultáneamente (pero con distinta resolución de acuerdo con el principio de incertidumbre de Heisenberg). Dentro de estas técnicas se encuentran la transformada corta de Fourier (STFT), las distribuciones Wigner-Ville [4] y Choi-Williams [5] y, por último, la transformación con ondículas (WT). Esta última ha alcanzado bastante auge en la última década, por su capacidad para trabajar con transitorios y periodicidades.

Una descripción exhaustiva de estos tipos de análisis se efectuará en el capítulo 3.

1.3. OBJETIVOS

En virtud de las necesidades expuestas y del estado actual del arte, se pueden establecer el objetivo global y aquellos objetivos más específicos de esta tesis.

OBJETIVO GLOBAL:

Como respuesta a las carencias de los sistemas de diagnosis actuales, el objetivo fundamental que se persigue en la tesis es el desarrollo e implementación de un nuevo sistema de detección, localización y cuantificación del nivel de degradación en una máquina rotativa a partir de mediciones vibratorias indirectas.

La consecución de este objetivo implicará las siguientes tareas:

a) Diseño del sistema en base a una estrategia de detección, b) implementación mediante una herramienta de programación informática (aunque también se podría implementarlo sobre un soporte físico electrónico) y, por último, c) su aplicación en entornos simulados y reales.

OBJETIVOS ESPECÍFICOS:

A la consecución del objetivo global fundamental se le han de añadir una serie de objetivos específicos cuyo fin es aumentar la capacidad del sistema de diagnosis. Entre éstos, se encuentran los siguientes:

- Detección automática y en tiempo real: El postproceso se debe realizar de forma automatizada sin necesidad del diagnóstico a posteriori realizado por expertos. Aunque la inspección visual de los rasgos de las señales vibratorias, es suficiente a veces para establecer el diagnóstico, éste no debería depender de la pericia del experto y, además, cualquier anomalía debería ser detectada en el menor intervalo de tiempo posible.

- Predicciones efectuadas a partir de mediciones vibratorias obtenidas en la carcasa: Los registros se obtendrán sin necesidad de desmantelar la maquinaria o pararla, aumentando su productividad.
- Diagnósis realizada en distintas condiciones operativas, distinto par y velocidad angular: El cambio de patrón de vibración en las máquinas rotativas, debido a su elevada no linealidad, hace que las características de los registros obtenidos en distintas condiciones de funcionamiento cambien notablemente, por este motivo la referencia de “normalidad” cambia con cada punto de funcionamiento y es muy complicado hacer pronósticos en estas condiciones que por otra parte son muy habituales.
- Localización “parcial” del componente mecánico (engranaje, rodamiento, etc.) dañado: Resulta muy interesante detectar el componente dañado, especialmente en maquinaria pesada donde sea posible su desmontaje parcial, ya que directamente supone un ahorro en medios y en tiempo de reparación.
- Robustez ante la presencia de “ruido” contaminante: Todos los registros obtenidos en maquinaria en condiciones normales (fuera de laboratorio) son susceptibles de ser contaminados por cierto ruido de fondo que deberá de ser tenido en cuenta en la confección del nuevo sistema de diagnóstico.
- Cuantificación de la severidad del daño ante un modo de fallo concreto: Muy poco trabajo se ha efectuado para determinar el grado de severidad de un fallo en una máquina rotativa, a pesar de que éste facilitaría un mantenimiento preventivo y evitaría fallos inesperados. Por este motivo en esta tesis se intentará establecer un módulo de cuantificación automatizada.

De todo lo anterior, se deduce que lo que se busca es un sistema de diagnóstico que “dispare” sus alarmas ante cualquier anomalía en una maquinaria rotativa, que indique el componente defectuoso y que estime el grado de severidad del fallo asociado, además se pretende que las inferencias se efectúen en tiempo real.

1.4. ORGANIZACIÓN DE LA TESIS

El trabajo se ha organizado en cuatro fases que corresponden a los cuatro capítulos de los que consta la tesis y cuyo resumen se expone a continuación:

CAPITULO 2

En este capítulo, se realiza primeramente una descripción y clasificación de los modos de fallo más relevantes en maquinaria rotativa y se estudia la transmisibilidad de vibración desde el origen del fallo hasta la carcasa de la maquinaria, donde se recoge la señal.

A continuación, se estudian las principales características de los registros obtenidos y, por último, se hace un repaso de las técnicas de procesamiento de señal.

CAPITULO 3

Este capítulo constituye, sin duda, la base teórica sobre la que se fundamenta el sistema de diagnóstico desarrollado y abarca los siguientes temas:

- La transformación con ondículas empleada como herramienta de preproceso de la señal y representación de los datos.
- El reconocimiento y clasificación de patrones mediante redes neuronales artificiales, utilizadas tanto en tareas de detección novial como en tareas de cuantificación y clasificación.

Por último, se estudia cómo se articulan estas herramientas para diseñar un procedimiento de diagnóstico y detección.

CAPÍTULO 4

En este capítulo se implementa y valida el sistema de diagnóstico propuesto en dos etapas:

1-Detección y diagnóstico.

La capacidad del módulo de detección para detectar anomalías en el funcionamiento de la maquinaria rotativa se evalúa en dos casos diferentes. El primero corresponde a una simulación numérica mientras que en el segundo se estudia el estado de una estación de bombeo en Lemmer (Holanda) ensayada experimentalmente.

2-Cuantificación

En esta etapa, se valida el módulo de cuantificación de daño propuesto empleando los resultados obtenidos con un modelo matemático de un sistema de transmisión de potencia.

CAPITULO 5

En este último capítulo, en base a los resultados obtenidos y a los objetivos fijados inicialmente, se extraen las conclusiones principales haciendo especial hincapié en las ventajas y limitaciones del nuevo sistema de diagnóstico propuesto.

Por último, se plantean posibles temas de desarrollo para futuros trabajos que resultan interesantes para completar y extender el trabajo efectuado en esta tesis.

CAPITULO.2

FALLOS EN MÁQUINAS ROTATIVAS: DETECCIÓN Y DIAGNÓSTICO.

2.1. INTRODUCCIÓN

A partir de la normativas internacionales vigentes ANSI/AGMA 1010-E95 e ISO 15243:2004 (E) se ha efectuado un listado de modos de fallo en rodamientos y engranajes. El empleo masivo de estos componentes en maquinaria rotativa es predominante por lo que su estudio queda plenamente justificado. Sin embargo, la extrapolación a otros elementos de maquinaria rotativa, como por ejemplo ejes, volantes, etc., se podría llevar a cabo dependiendo de cada caso particular (esto dependerá, como se verá posteriormente, de que el fenómeno de degradación posea los rasgos de transitoriedad y periodicidad); no obstante, las limitaciones en tiempo y en extensión hacen que el estudio y desarrollo del sistema de diagnosis desarrollado en esta tesis se aplique solamente a rodamientos y engranes.

Normalmente, la degradación de un componente raramente se debe a un solo mecanismo de fallo sino que, en general, obedece a una combinación de estos mecanismos como por ejemplo fenómenos de oxidación combinados con fatiga, abrasión más oxidación más fatiga, etc.

En esta tesis se persigue el diseño de un sistema de diagnosis asociado a aspectos mecánicos del fallo por lo que todo el ámbito de estudio quedará reducido a su vertiente mecánica; además, si se supone la maquinaria funcionando en condiciones operativas normales (no excepcionales como pudiera ser una carga de impacto, etc.), y se acepta un diseño y montaje con una tolerancia y un mantenimiento adecuados, la fatiga, sin duda, constituye una de las principales causas de fallo, sino la primera.

Por ello, en todos los estudios llevados a cabo, se considerarán solamente los modos de fallo asociados a fenómenos de fatiga en rodamientos y engranajes.

Por otro lado, para su detección, se suelen utilizar tres tipos de aproximaciones: análisis acústico de la señal, control de partículas y análisis vibratorio.

De las tres aproximaciones, los análisis vibratorios se han convertido en los más populares a causa de su facilidad de medición [8].

Las mediciones realizadas con análisis acústicos, aunque cubren un espectro de frecuencias más amplio, son más propensas a la contaminación y, por tanto, es más complejo su empleo en “campo”, entendida esta medición como los registros obtenidos fuera de laboratorio, donde el entorno no es controlable.

Las predicciones basadas en el control de partículas se basan en los registros obtenidos con sensores de tipo inductivo. Éstos contabilizan por un lado el número de partículas suspendidas en el lubricante y, por otro, su tamaño, determinándose a partir de estas mediciones la masa acumulada [3]; a pesar de que son muy útiles en la detección de cierto tipo de fallo asociado con el desprendimiento de material (modo de fallo asociado a picaduras como se verá a lo largo de este capítulo), sólo realizan predicciones a nivel de alarma y no sirven para localizar el componente o componentes dañados. Además, defectos asociados con agrietamientos no se pueden detectar con esta última técnica.

Por ello, en virtud de los objetivos expuestos en esta tesis, para la detección de los posibles modos de fallo, se utilizarán análisis vibratorios sobre medidas registradas (aceleraciones) directamente en la carcasa de la maquinaria sin que sea por tanto necesario el desmontaje de la misma de cara a la realización de predicciones sobre su estado de “salud”.

El contenido de la tesis se apoyará en la relación de causalidad entre el mecanismo de fallo y las vibraciones.

Por tanto, en primer lugar, parece interesante describir y clasificar los modos de fallo más relevantes en maquinaria rotativa, para, posteriormente, estudiar la forma en la que la energía vibratoria asociada al modo de fallo se transmite a la carcasa de dicha maquinaria; en particular, se analizarán y modelizarán matemáticamente las contribuciones a esta señal vibratoria provocadas por los defectos situados en los distintos componentes mecánicos que constituyen dicha maquinaria, en particular los rodamientos y engranajes.

A continuación se ilustrarán las principales características de dichas contribuciones y, por último, se hará un repaso de las técnicas de procesado de señal más habituales

2.2. CLASIFICACIÓN Y DESCRIPCIÓN DE MODOS DE FALLO EN MÁQUINAS ROTATIVAS

2.2.1. CLASIFICACIÓN

La norma ANSI/AGMA 1010-E95 clasifica dos tipos de “patologías” en engranajes según el siguiente esquema:

a. Fatiga de contacto:

 Microescala:

 Micropicaduras (“Micropitting”)

 Macroescala:

 Picaduras (“Macropitting”):

 Estado inicial.

 Estado progresivo o destructivo.

 Desconchamiento (“Flake pitting”).

 Astillaje (“Spalling”).

 Subcaso de fatiga.

b. Fatiga de flexión.

Fallo en la base del diente.

Fallo en el flanco del diente

Fallo en la punta del diente.

Para el caso de los rodamientos la clasificación de los posibles modos de fallo según la normativa ISO 15243:2004 (E) se basa en aspectos visibles en las superficies de contacto o en otras superficies funcionales

Así mismo esta normativa distingue entre dos tipos de “patologías” asociadas a la fatiga de contacto en rodamientos que son:

a. Picaduras por iniciación de grieta subsuperficial.

b. Picaduras por iniciación de grieta superficial.

2.2.2. DESCRIPCIÓN.

a. Mecanismos asociados a la fatiga de contacto.

La fatiga de contacto se halla relacionada con la acción de contacto alternado entre dos superficies.

Durante el contacto, las presiones de contacto hertzianas crean un pico de cortadura en las proximidades de dichas superficies; este fenómeno de cizalla alterna es el generador de pequeñas grietas (por fatiga) que, dependiendo de la presencia de imperfecciones, inclusiones, concentradores geométricos, etc., se propagarán de distinta forma y acabarán saliendo a la superficie.

En los casos donde además de rodadura pura haya cierto deslizamiento, las fuerzas de fricción desplazarán el máximo de las tensiones de cortadura moviéndolo hacia la superficie. Así, en

función del origen de la grieta, se pueden distinguir dos modos de fallo, los de origen superficial y los de origen subsuperficial.

Así mismo, como se estableció en la clasificación interna de modos de fallo en engranajes por fatiga de contacto, existe una segunda subdivisión basada en la escala a la que afecta el fenómeno, así se habla de una macroescala y una microescala.

Hay que indicar que tanto los mecanismos de formación como propagación asociados a la fatiga de contacto, son idénticos tanto para rodamientos como para engranajes.

Antes de continuar, conviene realizar una breve descripción de la caracterización de los defectos asociados a la fatiga de contacto:

a.1. Micropicaduras

Para un tamaño promedio de picadura en torno a las veinte micras, la superficie de contacto adquiere un aspecto mate. La aparición de esta “patología” está directamente relacionada con la rugosidad superficial de las áreas de contacto además de con la dureza superficial.

Esta rugosidad superficial y las altas presiones de contacto provocan, en función de la ductilidad del material, la aparición de microgrietas que tienden a propagarse en dirección paralela a la superficie de contacto formando microagujeros que confieren a la superficie el aspecto mate anteriormente citado (figura 1).

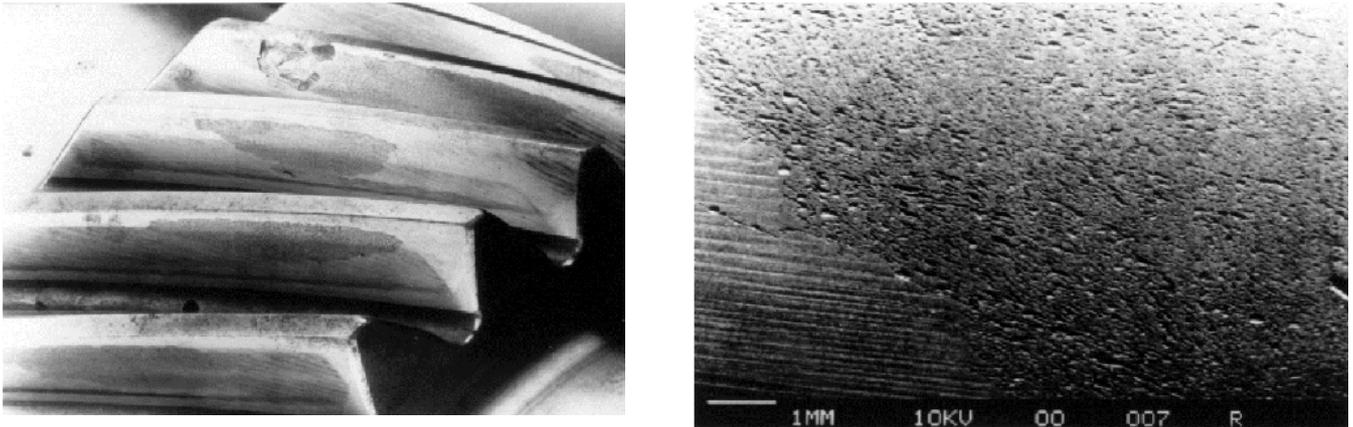


Figura 1. Fotos de engranajes con micropicaduras.

En las fotografías de la figura 1 se puede apreciar el aspecto que presenta este modo de fallo, la foto de la izquierda muestra una visión global de varios dientes afectados por este defecto mientras que la de la derecha representa un detalle de las zonas afectadas.

a.2. Macropicaduras (figura 2)

Según el grado de severidad de este fenómeno la normativa diferencia los siguientes estados de picadura:

Picadura inicial.

Se distingue porque los tamaños de cráteres son inferiores a 1 mm. de diámetro.

Ocurre en áreas localizadas de los flancos de los dientes próximas al plano de presión y, según la norma ANSI, tiende a estabilizarse ya que las picaduras tienden a redistribuir las cargas de forma más homogénea por el flanco del diente.

En rodamientos, es habitual encontrar este tipo de fallo en la pista externa, especialmente en la zona de carga del rodamiento, aunque puede darse en otro lugar.

Además, el proceso se ve acelerado por la presencia de indentaciones provocadas a su vez por la presencia de contaminantes o por la manipulación de los rodamientos en las fases de fabricación y montaje.

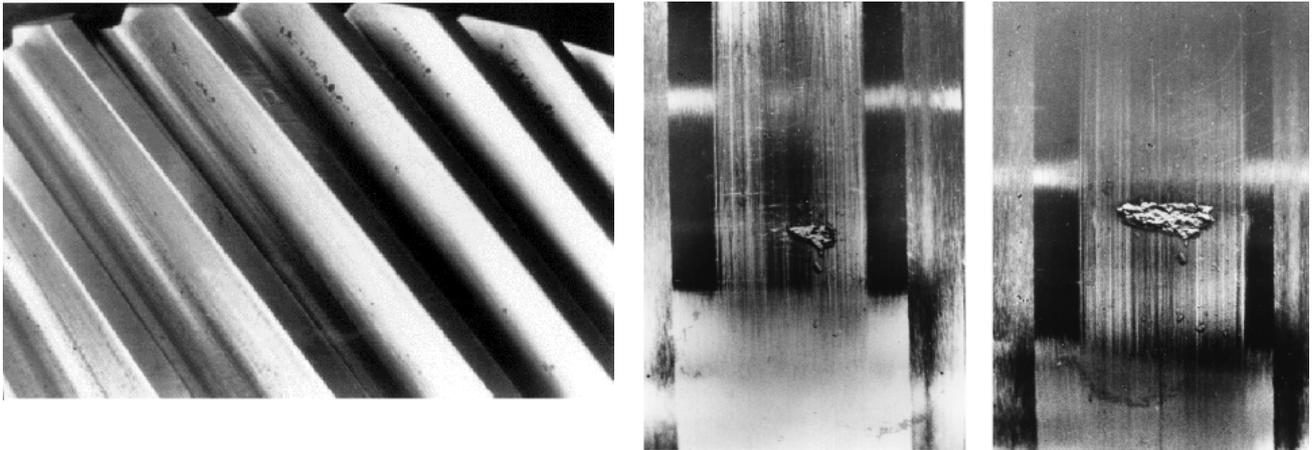


Figura 2. Fotos de picaduras en engranajes y rodamientos.

Picadura progresiva.

La picadura inicial, cuando se supera cierto tamaño en el diámetro de los cráteres, continúa propagándose sin estabilización. Este estado se caracteriza por estar las superficies de los flancos de los dientes cubiertas de picaduras (la superficie se asemeja a la piel de una naranja tal y como se puede apreciar en la foto de la figura 3) de distinto diámetro en toda su extensión.

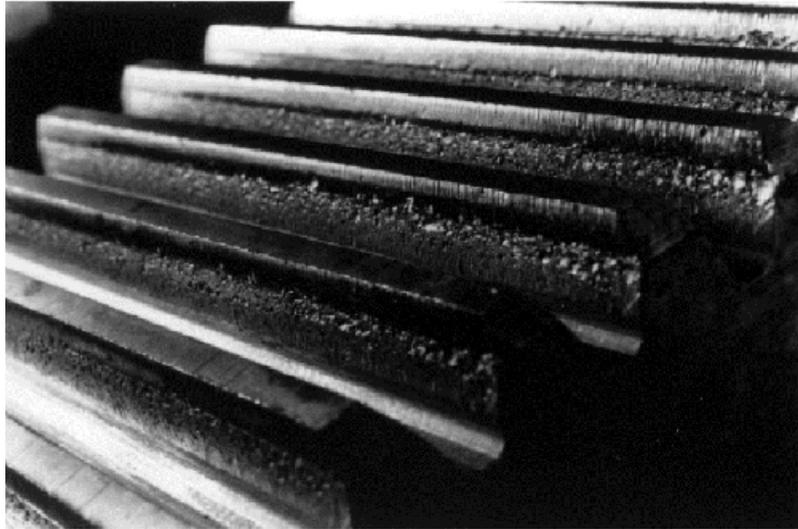


Figura 3. Picadura con desarrollo progresivo.

Este fenómeno es completamente análogo en rodamientos.

Picadura con efecto de descascarillado (figura 4)

Se caracteriza porque las picaduras afectan a áreas más bastas de la superficie del flanco del diente o de la pista de rodadura, en el caso de los rodamientos, descascarillando parte de la superficie. La grieta inicial se propaga a modo de ventilador hasta formar cráteres triangulares de diverso tamaño.

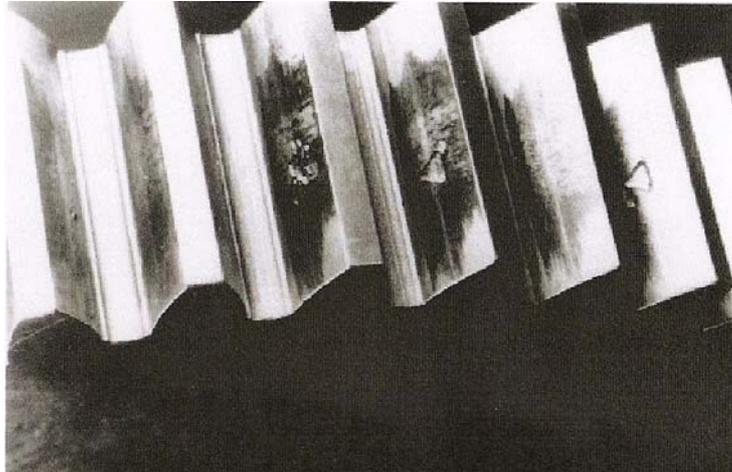


Figura 4. Engranaje con flancos descascarillados.

Esta fase corresponde a un grado de degradación que se puede calificar de severo, al igual que el caso siguiente como se especifica a continuación.

Spalling (figura 5)

Corresponde a una fase en la que las picaduras se colapsan y unen formando cráteres irregulares que cubren un área significativa del diente o del rodamiento dando lugar a una pérdida notable de rigidez.

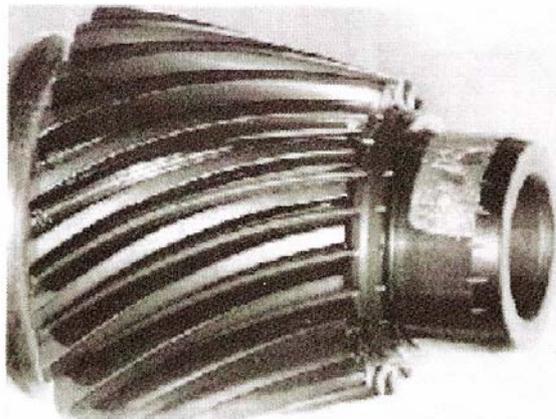


Figura 5. Rodamiento con problemas de spalling.

Subcaso de fatiga (figura 6)

Se suele originar en engranajes (este tipo de fallo no aparece recogido en la norma ISO 15243:2004 (E) para rodamientos) cuyos dientes se han endurecido por tratamientos mecánicos como por ejemplo carbonitruración, temple por inducción, etc. La grieta se genera en la zona de transición entre el núcleo del diente y la zona superficial afectada por el tratamiento de endurecimiento y progresa paralela a la superficie del flanco hasta que se ramifica a la superficie; las grietas ramificadas pueden aparecer como grietas longitudinales finas en unos pocos dientes.

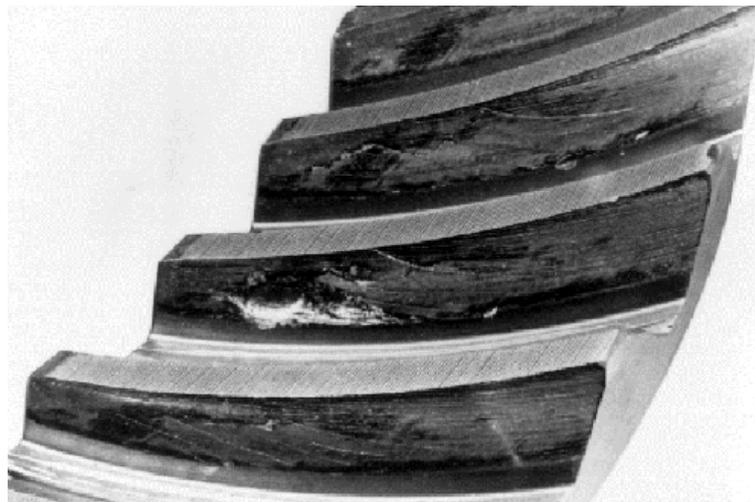


Figura 6. Subcaso de fatiga.

b. Mecanismos asociados a la fatiga de flexión

Este tipo de fallo no es aplicable a rodamientos, por lo que únicamente se realizará una descripción para el caso de los engranajes (figura 7).

El inicio de grieta a flexión se asocia a los concentradores de tensión geométricos como pueden ser la base del diente o la punta del mismo y su progresión se ve dominada por el modo I de propagación de grieta. Este tipo de fallo es bastante común; no obstante, el agrietamiento puede aparecer en cualquier lugar debido a tensiones mecánicas residuales generadas durante el proceso de fabricación, por efectos térmicos, etc.

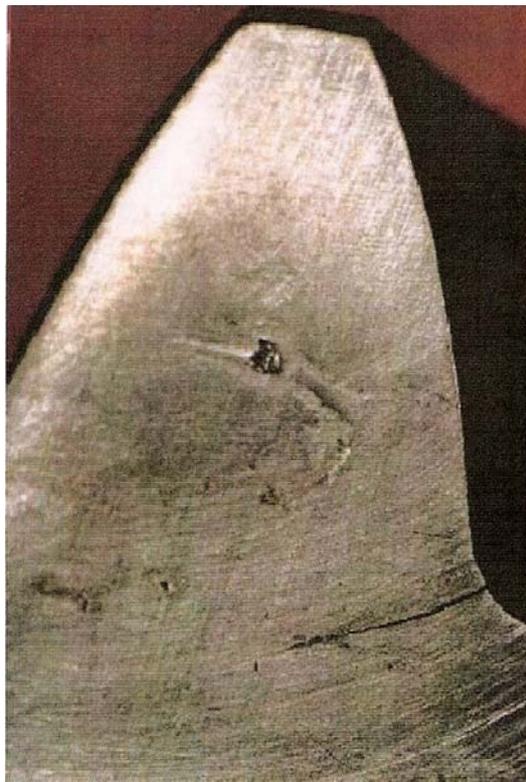


Figura 7. Foto de fallo de engranaje por fatiga a flexión.

2.3. PROCESOS VIBRO-ACÚSTICOS EN MÁQUINAS ROTATIVAS.

2.3.1. CLASIFICACIÓN DE LAS DISTINTAS FUENTES DE VIBRACIÓN.

La vibración puede entenderse como una fluctuación u oscilación registrada en una determinada magnitud física de interés (una fuerza, un movimiento, etc.).

La forma más habitual de clasificar las señales vibratorias se basa en su propia naturaleza, así cabe distinguir [9]:

- Señales vibratorias armónicas. (Desequilibrado de maquinaria, desalineamiento de ejes, etc.)
- Señales vibratorias tipo pulso. (Fenómenos de impacto asociados a defectos en engranajes y rodamientos).
- Señales vibratorias de naturaleza aleatoria. (Fenómenos de cavitación en aceite lubricante de caja de engranajes).

Una forma rápida de detección del fenómeno de fallo a partir del espectro de vibración se puede llevar a cabo heurísticamente mediante la tabla 1 [10].

Tipo de fallo	Frecuencia de fallo
Desequilibrado	$f_{\text{giro_eje}}$
Desalineamiento	$Mu \times f_{\text{giro_eje}}$ Siendo $Mu = 1,2,3,4.$
Pérdidas mecánicas	$Mu \times f_{\text{giro_eje}}$ Siendo $Mu = 0.5,1.5,2,5.$
Defectos en rodamientos.	Pulsos con frecuencias comprendidas entre 20 – 60 kHz.
Defectos en engranajes.	A la frecuencia de engrane y sus armónicos. $f_{\text{engrane}} = N_{\text{Dientes_corona}} \times f_{\text{giro_corona}}$

Tabla 1. Clasificación de modos de fallo en función de la frecuencia.

Se puede apreciar que distintos modos de fallo están asociados a bandas de frecuencia específicas.

Hay que indicar que en esta tesis se van a estudiar fenómenos asociados a bandas de frecuencias intermedias comprendidas en el intervalo de 100 Hz a 10 kHz aproximadamente, en concreto se estudiarán modos de fallo asociados a engranajes y rodamientos aunque en virtud de la tabla 1 pudieran estudiarse otros fenómenos como el desalineamiento de ejes, desequilibrado, etc.

2.3.2. VIBRACIONES ORIGINADAS POR LA ACCIÓN DE ENGRANE.

Durante la acción de engrane, la vibración se genera por una variación en la fuerza de contacto entre los flancos de los dientes del piñón impulsor y de la rueda o corona impulsada. Estas variaciones se deben a tres motivos principales (figura 8):

- a- Cambio en la amplitud de esta acción.
- b- Modificación de la dirección de la línea de acción entre los dientes que entran en contacto.
- c- Variación en la posición de dicha línea de acción.

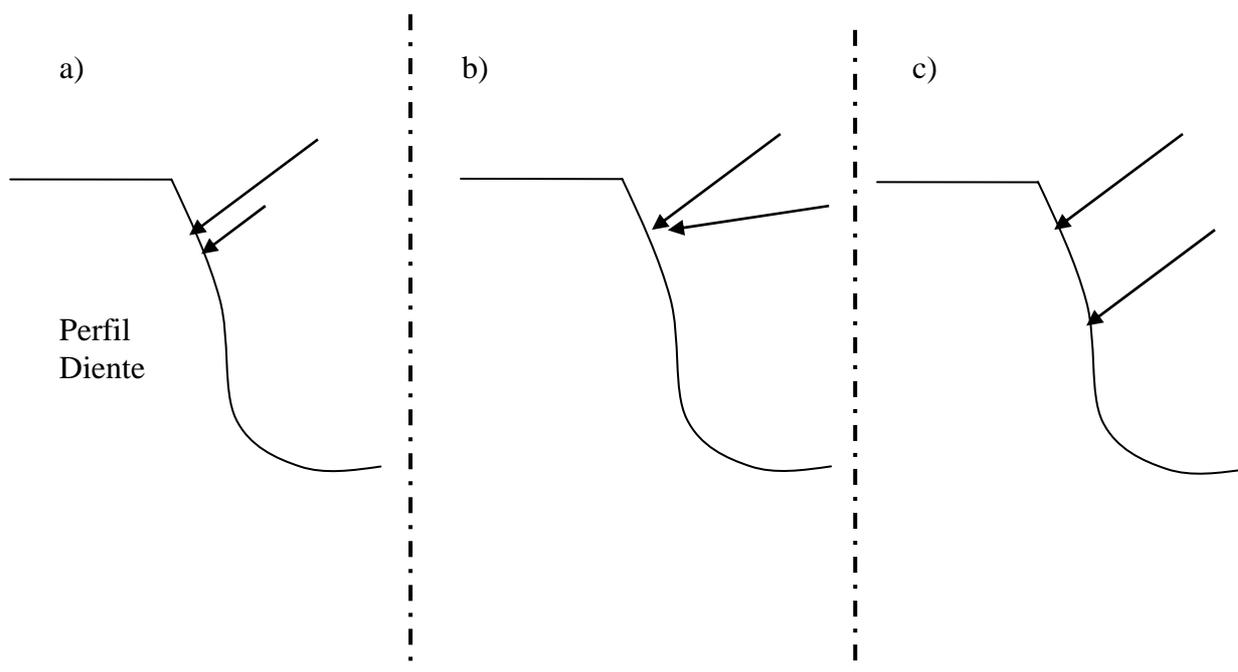


Figura 8. Vibraciones ocasionadas por: (a) Variación en amplitud de la carga, (b) variación en dirección de la carga, (c) variación en la posición de la línea de acción sobre el flanco.

Para involutas o perfiles conjugados la principal causa de vibración se asociada principalmente al caso (a), es decir a una variación en la amplitud de la carga de contacto.

En perfiles de involuta el diseño de su geometría hace que se pueda considerar fija la línea de acción, por lo cual se puede descartar el efecto indicado en el punto (c) frente al comentado anteriormente; además, la máxima variación angular de la fuerza de acción entre dientes es aproximadamente de unos tres grados para coeficientes de fricción típicos (del orden de 0.05) por lo que también, para este tipo de involuta, se puede despreciar la causa indicada en (b), [11]. Por tanto, según lo anterior, si se produce una rodadura perfecta entre las superficies de contacto, cabría preguntarse a qué se debe la variación en la amplitud de la fuerza de engrane. Por cuestiones relacionadas con la filosofía del diseño, el perfil del diente en su parte superior se aparta de la involuta perfecta en lo que se llama “separación de la punta”. Esto se debe a que las deformaciones elásticas de los perfiles podrían provocar el enclavamiento de dichas

puntas de los dientes; si a este defecto (hay que entender que se trata de un defecto provocado) se le añaden los errores de fabricación, de montaje, el desgaste de dientes, las deformaciones de origen térmico, etc., se tiene una combinación de factores cuyo efecto es el despegue de la superficies de rodadura durante el engrane de cada diente. Esto es lo que se conoce como **error en la transmisión (TE)**. Dicho error puede entenderse como un desplazamiento relativo angular entre los ejes impulsor e impulsado (figura 9).

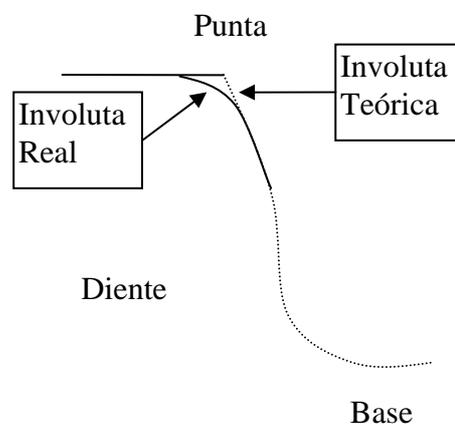


Figura 9. Separación de la punta del diente.

Esta diferencia se suele multiplicar bien por el radio primitivo o bien por el radio base (entendido éste como el radio sobre el que se traza la evolvente del perfil del diente) para obtener una medida lineal del desplazamiento; como este desplazamiento se suele tomar en el plano de presión se opta por multiplicar por el radio base.

Se ha de comentar también que es frecuente el empleo de diseños de perfiles que contemplen además de la “separación de la punta”, una “separación de la base del diente”.

En cualquier caso, el origen de la vibración en cualquier par cinemático de engrane se debe al error en la transmisión y cualquier defecto quedará englobado, por tanto, dentro de este término.

Como consecuencia se tienen una serie de pulsos originados por la sucesión de impactos debidos al contacto y despegue de los perfiles de engrane que se propagan por el eje de la máquina y se transmiten por los alojamientos de los rodamientos a la carcasa de la maquinaria.

2.3.2.1. Modelo matemático

La señal vibratoria registrada en la carcasa debida a la acción de engrane, tanto para engranajes “sanos” como para los defectuosos, se puede expresar mediante la siguiente relación [12]:

$$(2.3.1) \quad z(t) = \sum_{k=1}^K A_k(t) \cos[2\pi k f_{engrane} t + \phi_k(t)]$$

Siendo ‘ $A_k(t)$ ’ y ‘ $\phi_k(t)$ ’ la modulación instantánea de amplitud y fase, respectivamente, debido a la carga transmitida y ‘ K ’ el número de armónicos necesarios para representar esta contribución a la señal vibratoria total.

Un espaciado no constante en los dientes o una velocidad angular no estacionaria da lugar a modulaciones en la fase, mientras que las modulaciones en la amplitud se originan por irregularidades en la superficie de contacto de los dientes.

Los fallos en los engranajes producen una modulación tanto en magnitud como en fase y, al igual que en el caso anterior, las contribuciones debidas a dichos defectos se pueden determinar mediante las funciones de modulación de amplitud y fase ‘ $a_m(t)$ ’ y ‘ $b_m(t)$ ’, respectivamente [13]:

(2.3.2)

$$z(t) = \sum_{k=1}^K A_k(t)(1 + a_m(t)) \cos[2\pi k f_{engrane} t + \phi_k(t) + b_m(t)]$$

2.3.3. VIBRACIONES ORIGINADAS POR LOS RODAMIENTOS

En este caso, a diferencia de los engranajes, la contribución a la señal vibratoria registrada en la carcasa de la maquinaria debida a un rodamiento teórico “sano” es nula. Por tanto, la contribución a la señal vibratoria registrada debido a estos componentes solamente se produce cuando se presenta un defecto en los mismos.

Se pueden considerar tres escenarios de fallo asociados a la posición donde se produce la degradación en el rodamiento.

Escenario I: Degradación en pista externa

Normalmente, el defecto se localizará en la zona de carga del rodamiento, por tanto, cada vez que un elemento de rodadura (esferas, conos, etc.) pase por dicha zona se producirá un impulso. Se tendrán unos registros en el histórico de la señal vibratoria similares a los indicados en la parte superior izquierda de la figura 10, siendo ‘BPF’ la frecuencia de paso de cada elemento de rodadura.

VIBRATION SIGNALS FROM FAULTS
IN ROLLING ELEMENT BEARINGS

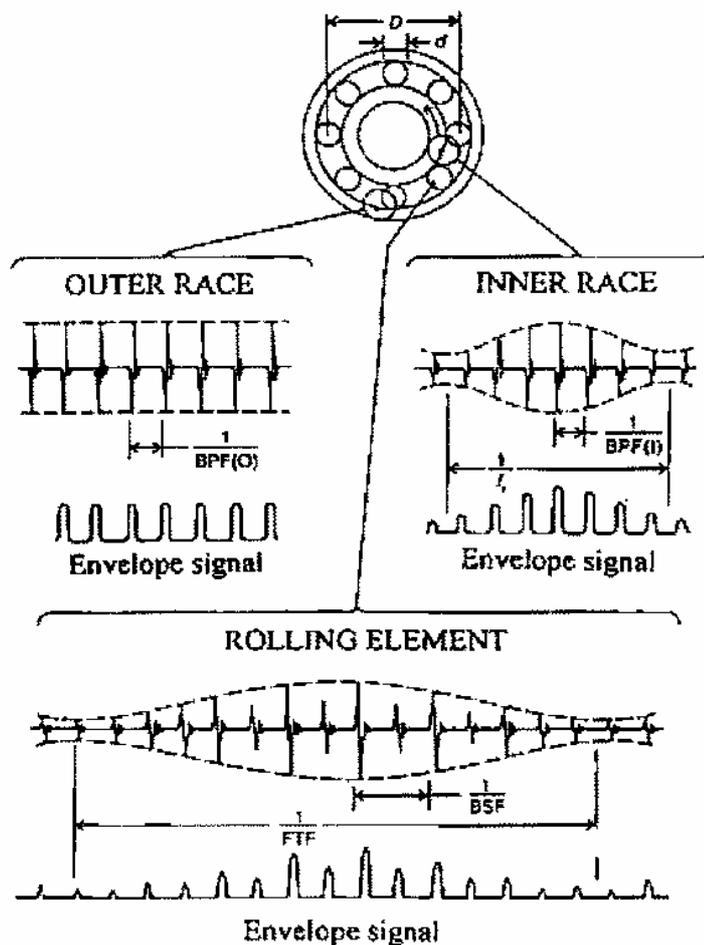


Figura 10. Efectos del fallo de un rodamiento en el histórico de una señal.

Escenario II: Degradación en pista interna

Los trenes de pulsos vendrán modulados esta vez por la frecuencia de paso de la pista interna por la zona de carga, por tanto, estarán modulados por una frecuencia igual a la frecuencia de giro del eje: Esto se puede comprender mejor tras analizar la parte derecha superior de la figura 10.

Escenario III: Degradación en elemento de rodadura

En este caso, la modulación está gobernada por la velocidad de giro del retén o jaula (es un elemento que se emplea para mantener distanciadas las bolas o rodillos) además de por la frecuencia de giro de los rodillos.

Este caso se muestra en la parte inferior de la figura 10 siendo FTF la frecuencia de paso del retén y BSF la frecuencia de rotación del rodillo.

2.3.3.1. Modelo matemático

Tal como se ha comentado anteriormente, en el caso de rodamientos dañados se producen una serie de pulsos modulados por una envolvente cuyas características dependen de la zona donde se produzca la falla en el rodamiento.

Hay que indicar que la contribución a la señal vibratoria en el caso de contar con rodamientos defectuosos se puede modelar mediante series de Fourier. En concreto, si se produce un defecto en la pista exterior la modulación de la señal se puede expresar mediante la relación [14]:

$$(2.3.2) \quad r(t) = \sin(2\pi f_{bpf} t)(1 + \beta f_r t)$$

2.4. CARACTERÍSTICAS DINÁMICAS DE LAS SEÑALES REGISTRADAS.

2.4.1. TRANSMISIÓN DE LAS VIBRACIONES.

Según lo visto, los defectos asociados tanto a rodamientos como a engranajes producen una serie de trenes de pulsos que se transmiten desde la fuente de excitación hasta el lugar de la carcasa donde se colocan los transductores; en el caso de los engranajes, este camino de vibración involucra al eje de giro donde se tienen instalados la corona o piñón, el alojamiento del rodamiento donde se sustenta el eje de giro, el rodamiento y por último la carcasa. En el caso del rodamiento el camino de vibración únicamente está constituido por los dos últimos componentes citados.

Estos pulsos excitan una amplia banda de frecuencias y, por lo tanto, su contenido viene determinado por las frecuencias de resonancia del sistema excitadas por esta banda. Según esto último, la transmisión de la vibración desde la fuente de excitación al lugar donde se haya colocado el sensor se puede modelar como un filtro lineal (para cada condición de operación) con una determinada función de transferencia [15].

Las características modales de la estructura causan este filtrado, además los modos de vibración de la estructura en relación con la posición espacial en la carcasa y su orientación (dirección de medición) determinarán el grado de amplificación o decrecimiento con el que estas vibraciones son transmitidas.

Por consiguiente, la respuesta observada en los transductores consistirá en una serie de trenes de pulsos, cuya magnitud decae temporalmente en forma exponencial de acuerdo con el amortiguamiento estructural de la máquina en estudio.

2.4.2. RASGOS FUNDAMENTALES.

A tenor de lo explicado en el punto anterior y lo argumentado en el apartado 2.4, es posible modelar la señal vibratoria registrada en un punto de la carcasa como:

$$(2.4.1) \quad u(t) = z(t) + r(t) + w(t)$$

siendo 'z(t)' el componente asociado a la acción de engrane, 'r(t)' la contribución debida a defectos en el rodamiento y 'w(t)' la componente asociada a excitaciones de origen aleatorio.

Se puede concluir que en maquinas rotativas las contribuciones debidas a los defectos asociados tanto a rodamientos como a engranajes tienen como características fundamentales su carácter transitorio (el debido a fenómenos de impacto) y periódico (repetición del fenómeno asociada al modo de fallo en concreto).

Esto hace que sea muy interesante una representación del registro vibratorio en el dominio temporal y en frecuencia al mismo tiempo.

2.5. ANÁLISIS DE LA SEÑAL REGISTRADA

Una vez establecidas las causas de vibración en máquinas rotativas y su transmisión, se hace necesario el tratamiento de la misma a fin de extraer sus rasgos y características principales para su posterior análisis.

Existen numerosas técnicas de tratamiento disponibles. Estas se pueden clasificar en función del dominio de análisis en los siguientes grupos:

- 1-Análisis en frecuencia (“Cepstrum”).
- 2-Análisis temporal-estadístico.
- 3-Análisis tiempo-frecuencia

Una descripción detallada de estas técnicas se puede encontrar en las referencias [8] y [16]. En esta tesis se realizará una breve descripción de dichas técnicas haciendo especial hincapié en las ventajas y limitaciones de las mismas.

2.5.1. ANÁLISIS EN FRECUENCIA

El análisis espectral es una herramienta clásica empleada en el análisis de fallos. Por mera comparativa entre los espectros de la máquina dañada y sin dañar se puede, en algunos casos, establecer una diagnosis sobre el estado de la maquinaria [17].

El análisis “Cepstrum” consiste fundamentalmente en la transformada inversa de Fourier del logaritmo del espectro de potencia a fin de destacar las periodicidades que se hayan registrado en la medida “vibratoria”.

Según las referencias anteriormente mencionadas ([8], [16]) este tipo de análisis puede ser útil cuando la máquina rotativa a analizar no contenga demasiados pasos de reducción, sin embargo su aplicación a máquinas más complejas deja de ser eficiente por el número de componentes a analizar; además, tiene el inconveniente de no indicar la localización del defecto.

Esta técnica es sensible tanto a la evolución del daño como a la localización de los sensores.

2.5.2. ANÁLISIS TEMPORALES-ESTADÍSTICOS

Se basan en la comparación de promedios síncronos (en determinado régimen de vueltas de la maquinaria) de una señal de referencia obtenida a priori cuando el par cinemático de engrane esta “sano” (se tiene la certeza de que está sin ningún tipo de daño) frente a la señal obtenida en condiciones de funcionamiento de la máquina (funcionando en el mismo punto de operación).

Existen a su vez dentro de este tipo de análisis las siguientes subcategorías:

a. -Técnicas de demodulación de amplitud y fase del residuo [13].

Esta técnica se basa en un filtrado de la señal “pura” (sin tratar) alrededor de la frecuencia de engrane y de sus armónicos (hay que seleccionar un ancho de banda) para posteriormente sustraer dicho armónico fundamental y recuperar la envolvente de esta señal residual (tanto en amplitud como en fase) en el dominio temporal con el objetivo de captar el efecto fundamental anteriormente comentado de modulación de la señal.

Hay que destacar que esta técnica es muy sensible a la selección del ancho de banda (si no se elige correctamente se puede llegar a perder cierta información de interés).

b. Análisis estadísticos.

Se basan en realizar estadísticas sobre el residuo de la señal obtenida al sustraer la señal leída en condiciones operativas de la máquina respecto a la de referencia.

A grandes rasgos todas las técnicas arriba mencionadas tienen la ventaja de que al promediar se elimina parte del ruido “contaminante” de las señales; no obstante, hay que asegurar muy bien la sincronización puesto que de no ser así se podrían eliminar ciertos eventos periódicos que tuvieran relevancia en el dictamen de la diagnosis de la maquinaria rotativa.

Como ventaja cabe citar que los análisis temporales se han mostrado útiles tanto en la detección como en la localización del daño y además son sensibles a la evolución de la degradación.

2.5.3 ANÁLISIS TIEMPO-FRECUENCIA

Como se ha comentado anteriormente cualquier perturbación añadida a una señal vibratoria implica tanto un efecto “*transitorio*”, por cada revolución del eje, como un efecto “*consistente*”, es decir, que se repite revolución tras revolución por lo que existe cierta periodicidad de dicha perturbación.

Un tratamiento óptimo de la señal vibratoria sería aquel que contemplara este doble aspecto de la señal (periódico-transitorio). Para tal propósito, se ha empleado históricamente la transformación STFT (transformada corta de Fourier). Sin embargo, al igual que con la técnica de demodulación, se tiene el inconveniente de la selección de la ventana temporal; esto se superó a finales de los años ochenta mediante la técnica de transformación con ondículas implementada mediante el algoritmo FWT (transformada rápida con ondículas).

Dicha técnica ofrece un marco muy completo para el tratamiento de señal como se verá en detalle en los siguientes capítulos de la tesis.

CAPITULO.3

SISTEMA DE DETECCIÓN Y DIAGNOSIS BASADO EN ONDÍCULAS Y REDES NEURONALES ARTIFICIALES.

3.1. INTRODUCCIÓN.

El principal propósito de cualquier sistema de diagnosis es el análisis de información relevante con objeto de juzgar bajo qué condiciones los componentes internos inaccesibles se están degradando sin que para ello sea necesario el desmontaje de la maquinaria.

Aunque la degradación del componente rotativo aparece claramente reflejada en la señal vibratoria recogida en la carcasa de la máquina, sus rasgos característicos suelen estar escondidos en la señal, por consiguiente, es necesaria una técnica lo suficientemente sensible a la hora de establecer el sistema de diagnosis y detección.

La mayoría de las técnicas de detección conocidas (capítulo 2) representan la señal vibratoria bien en el dominio temporal o bien en frecuencia. Las técnicas basadas en el promediado sincrónico TSA proporcionan un promedio de la señal temporal de cada par de engrane sobre un gran número de ciclos; este tipo de técnicas han sido reconocidas como una herramienta potente en la detección de fallos en engranajes y rodamientos ([18] a [21]), debido a que son capaces de eliminar el ruido de fondo y todos los eventos periódicos que no sean exactamente sincrónicos con el par de engrane de interés.

La utilización de la señal residual obtenida al restar sobre la señal TSA los armónicos derivados del par de engrane, constituye una de las técnicas más habituales en la detección de defectos

en maquinaria puesto que la señal residuo contiene esencialmente la porción de señal anómala generada por la degradación del par de engrane defectuoso.

El análisis tiempo-frecuencia ofrece un método alternativo en el ámbito del procesamiento de señal puesto que representa la información en ambos dominios simultáneamente. El método conocido como transformación de Gabor [22] o transformada corta de Fourier (STFT), es el método tiempo frecuencia más conocido; este método consiste en la aplicación de la transformada de Fourier previo filtrado de la señal temporal mediante ventanas de anchura temporal fija centradas en un instante de tiempo determinado.

De esta forma, se obtiene el contenido de frecuencias en ese instante y podría obtenerse para cualquier otro instante de tiempo repitiendo el mismo proceso, solo que centrando la ventana en ese tiempo concreto.

La resolución de este método depende del tamaño de la ventana, y como éste es constante, una resolución alta en tiempo y frecuencia no se puede obtener simultáneamente; el ancho de banda de la ventana se diseña con el objetivo bien de detectar picos elevados en la señal residual o bien anomalías a bajas frecuencias pero nunca para conseguir estos dos objetivos a la vez, por tanto su resolución es a menudo insatisfactoria (figura 11).

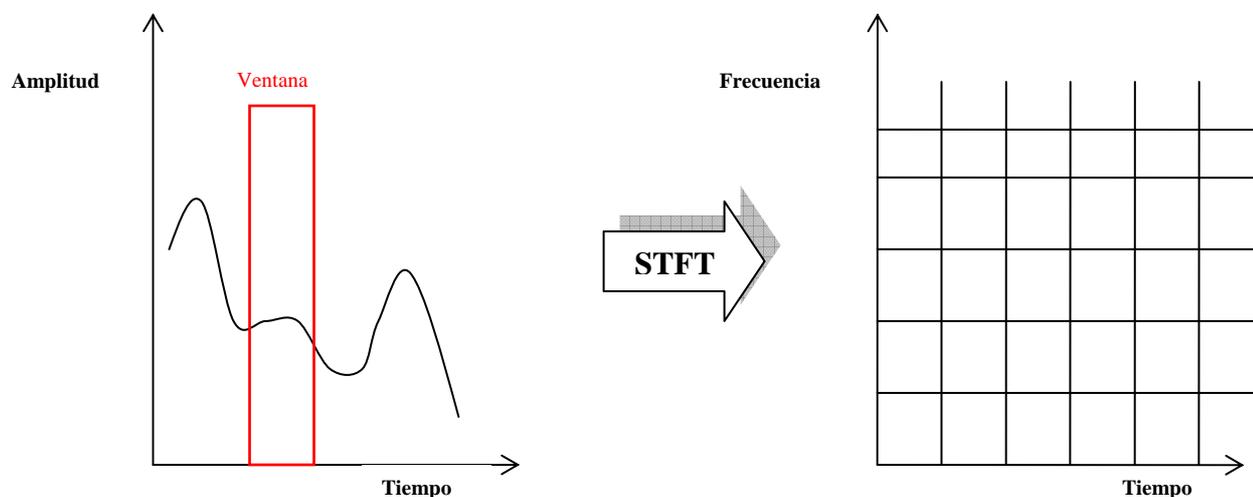


Figura 11. Esquema de la transformada corta de Fourier.

Se hace necesario, por tanto, una técnica más flexible. El tratamiento de señal basado en la transformación con ondículas (WT) se desarrolló originalmente a finales de la década de los 80 [23].

Se ha reconocido la utilidad de esta transformación (WT) en el tratamiento de señales transitorias como medio de detección de anomalías tanto en maquinaria ([24] a [27]) como en estructuras de ingeniería civil ([28] a [31]) y de otros campos de la ingeniería [32].

La ventaja del tratamiento con ondículas en contraposición al análisis de Fourier, es que las ondículas descomponen una señal mediante una serie de funciones base localizadas en el eje temporal [37] mientras que la transformada de Fourier usa como funciones base sinusoides de duración infinita, por tanto se pierde toda la información temporal.

Todas las funciones base de la transformada WT se generan mediante la dilatación y traslación por el eje temporal de una ondícula “madre”. Por tanto, la transformada WT es una transformada bidimensional. Para historias temporales, sus dos parámetros son la escala y el tiempo (figura 12).

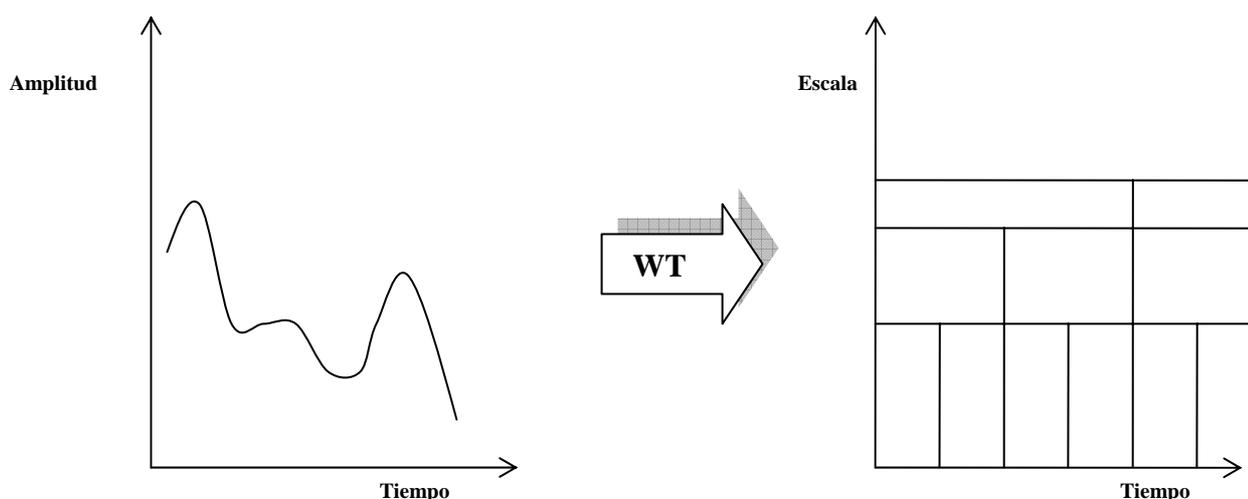


Figura 12. Esquema de la transformada con ondículas.

Esta descomposición local proporciona la capacidad de analizar fenómenos locales en señales vibratorias, como por ejemplo aquéllos debidos a la presencia de fallos, dado que la transformación WT produce un efecto “lupa” sobre cualquier instante de tiempo.

La transformación con ondículas es, por consiguiente, capaz de revelar algunos aspectos de las señales vibratorias que otras técnicas no son capaces de desvelar; consecuentemente, el método desarrollado en esta tesis utilizará esta herramienta para representación de las señales vibratorias obtenidas.

Sin embargo, en lo que respecta al análisis y a la toma de decisiones, aunque la inspección visual de ciertos rasgos de la transformada con ondículas pueda ser suficiente en aplicaciones particulares, se hace necesario el desarrollo de una herramienta de diagnóstico fiable y automatizada. Las redes neuronales se han mostrado como una excelente herramienta de reconocimiento de patrones tanto en la teoría como en aplicaciones prácticas. A causa de esto, modelos basados en redes neuronales se han aplicado recientemente en la detección y diagnóstico de componentes rotativos defectuosos ([33] a [36]).

Las redes neuronales artificiales (ANN) pueden ser entrenadas con medidas efectuadas sobre prototipos “sanos” o dañados con objeto de determinar el actual estado de la maquinaria en estudio.

La arquitectura y complejidad de la red neuronal dependerá del nivel de identificación requerido.

Para detectar la ocurrencia de daño, una red neuronal basada en la detección noval puede ser utilizada [48]. El objetivo de este tipo de redes es extraer patrones de funcionamiento cuando la máquina opera en condiciones normales. Estos patrones se utilizan como entradas y salidas de la red neuronal. Si una señal difiere de lo observado, la ocurrencia de una novedad o anomalía implica el disparo de una alarma por la red.

La filosofía implícita en la detección noval se puede implementar con las redes auto-asociativas (AANN), como se verá en más detalle a lo largo de este tema.

Por otro lado, dada la complejidad que supone el empleo de modelos matemáticos que tengan en cuenta la degradación de la maquinaria, se puede pensar también, siempre que se disponga de prototipos con distintos grados de daño, en el empleo de redes neuronales supervisadas como la perceptrón multicapa (MLP) o las basadas en funciones de base radial (RBF) para

tratar de cuantificar o clasificar la severidad del daño ante la presencia de un determinado tipo de fallo.

La organización de este capítulo se ha enfocado de la siguiente manera. En primer lugar se detallarán algunos aspectos teóricos de la transformación con ondículas que serán útiles para su aplicación como herramienta de preproceso de la señal y representación de los datos. Posteriormente, se describirá cómo las redes neuronales artificiales pueden ser empleadas tanto como detectores novales como en tareas de cuantificación e interpolación para determinar grados de severidad en la degradación de la máquina en estudio.

Por último, se verá como articular estas herramientas para diseñar un procedimiento de diagnóstico y detección.

3.2. PRETRATAMIENTO DE SEÑAL CON ONDÍCULAS

Uno de los aspectos clave en el tratamiento de señal es la descripción de sus propiedades tanto a nivel local como global.

Este aspecto es esencial para un sistema de diagnóstico y detección de fallos en máquinas rotativas, ya que su objetivo es la detección de eventos (fallos) que tienen lugar a determinadas frecuencias y en localizaciones específicas de la posición angular del eje de giro. Estos fallos originan pequeñas perturbaciones en las señales vibratorias que se registran mediante transductores localizados en la carcasa de la máquina rotativa y que se repiten con cada revolución del eje de giro de la misma. Además, tal y como se ha descrito en el capítulo anterior, fallos diferentes se asocian a diferentes bandas de frecuencia; como consecuencia, **se hace necesario pretratar las señales con objeto de aislar, resaltar y representar eventos a nivel local-global en la escala tiempo-frecuencia.**

El objetivo de esta sección es la presentación y justificación de la transformación con ondículas como herramienta de análisis y localización de perturbaciones en señales vibratorias. Para ello, se va a dar una breve introducción a los aspectos teóricos base de dicha transformación omitiendo las pruebas de los teoremas y evitando entrar en la descripción de los detalles matemáticos de los mismos.

En primer lugar se definirá la transformada con ondículas (WT) y se verá como una señal se puede descomponer utilizando dos alternativas, bien mediante transformación continua (CWT) o bien mediante transformación discreta (DWT). Una vez introducidas las bases del análisis multiresolución se darán pautas acerca de cómo seleccionar el nivel de descomposición requerido. A continuación se indicará la forma de seleccionar la familia de ondículas apropiada para efectuar tal descomposición. Posteriormente se comentará la forma de caracterizar

matemáticamente un pulso y su influencia en el escalograma y, por último, se ilustrará todo lo anterior con algún ejemplo numérico.

3.2.1. DEFINICIÓN DE LA TRANSFORMACIÓN CON ONDÍCULAS

La transformación con ondículas consiste en realizar una convolución de la señal original con una onda denominada ondícula “madre” en dos etapas; en la primera de ellas la convolución se realiza mediante traslación de la ondícula “madre” a lo largo del eje de tiempos (ver figuras 13.a y 13.b), abarcando todo el soporte temporal de la señal original; en la segunda fase, se repite la etapa anterior pero con versiones dilatadas de dicha ondícula “madre” (ver figura 13.c), obteniéndose finalmente los coeficientes de la transformación.

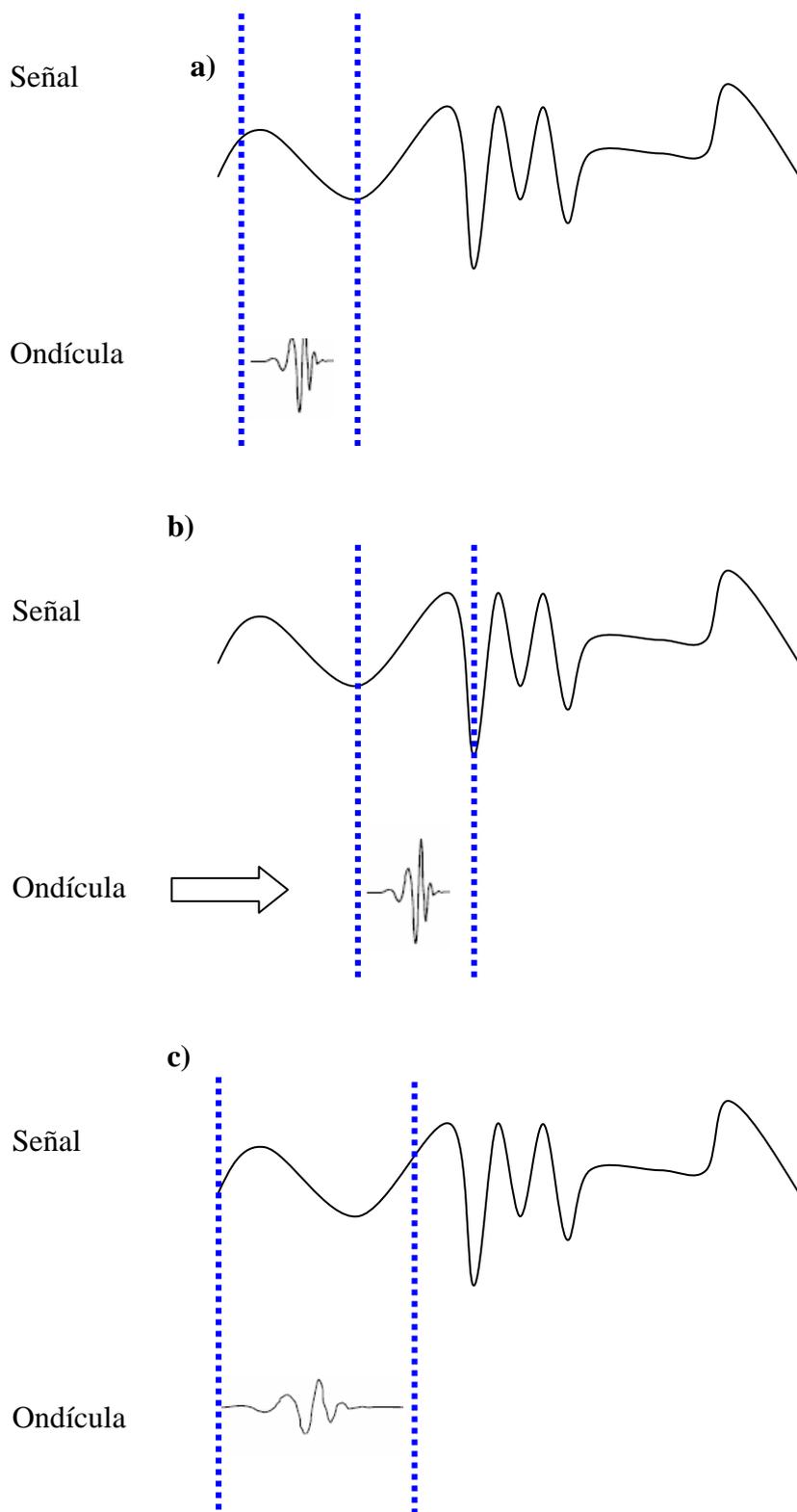


Figura 13. Esquema ilustrativo de la transformación WT.

Si se define la ondícula “madre” de la siguiente forma:

$$(3.2.1) \quad \psi_{ab} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$$

siendo ‘a’ y ‘b’ los coeficientes que representan la dilatación y traslación, respectivamente, la convolución que permite obtener los coeficientes de transformación viene dada por la expresión:

$$(3.2.2) \quad C_{ab} = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$

Se puede pensar en estos coeficientes como indicadores del grado de similitud entre la ondícula “madre” y la señal “bruta”; de hecho, se puede demostrar que si la energía de la señal a analizar es la unidad, entonces su transformación con ondículas determina los factores de correlación entre dicha señal y la ondícula “madre”.

La transformación (3.2.2) se denomina transformación continua (CWT) y transforma la señal original ‘f(t)’ definida en la escala tiempo, en una superficie 2D o escalograma que es, en cierto modo, una representación redundante de la información original.

Igualmente, en sentido recíproco, se puede reconstruir la señal original a partir de los coeficientes anteriores mediante la siguiente expresión:

$$(3.2.3) \quad f(t) = \frac{1}{K_{\Psi}} \int_{a=-\infty}^{\infty} \int_{b=-\infty}^{\infty} C_{ab} \psi_{ab}(t) db \frac{da}{a^2}$$

siendo K_{Ψ} una constante dependiente del tipo de ondícula.

El grado de redundancia, anteriormente referido, implica que no todos los coeficientes son necesarios para reconstruir la señal original. Como se está planteando el desarrollar un sistema de identificación en tiempo real, se va a utilizar una versión discreta de la transformación CWT, denominada DWT; en esta última versión, la señal original 'f(t)' se proyecta en subespacios de distinto soporte temporal, lo que implica menor esfuerzo computacional sin pérdida de información aparente.

En la transformación DWT, los parámetros de dilatación y traslación, 'a' y 'b', son enteros; habitualmente, se utiliza un muestreo diádico para dichos parámetros lo que significa que la dilatación viene definida como 'a = 2^j' mientras que la traslación viene dada por los valores 'b = k 2^j', siendo 'k' y 'j' valores enteros.

3.2.2. SEÑALES DE DETALLE Y APROXIMACIÓN

En el caso de muestreo diádico, la transformación discreta se puede definir formalmente a partir de (3.2.2) como:

$$(3.2.4) \quad C_{jk} = 2^{-j/2} \int_{-\infty}^{\infty} f(t) \psi(2^{-j}t - k) dt$$

Asimismo, de igual forma que para la transformación CWT, la señal se puede reconstruir a partir de los coeficientes de la transformación, mediante una transformación inversa discreta IDWT:

$$(3.2.5) \quad f(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} C_{jk} \psi_{jk}(t)$$

Si en (3.2.5) se fija la escala 'j' y se suman todas las traslaciones 'k' a lo largo de la escala de tiempos se obtiene la subseñal de detalle para esta escala:

$$(3.2.6) \quad D_j(t) = \sum_{k \in \mathbb{Z}} C_{jk} \psi_{jk}(t)$$

De la misma forma, si se suman todas las subseñales de detalle a lo largo de todas las escalas 'j' se recupera la señal original:

$$(3.2.7) \quad f(t) = \sum_{j \in \mathbb{Z}} D_j(t)$$

De acuerdo a esta última expresión, la señal original se cubre en el dominio temporal mediante ondículas trasladadas mientras su espectro finito lo hace con el espectro de ondículas "dilatadas". Esto se puede interpretar como una operación de filtrado con una serie de bancos de filtros, ya que la convolución consiste realmente, en el ámbito de procesamiento de señal, en una operación de filtrado. Sin embargo, mientras la traslación de las ondículas viene

limitada por la duración temporal de la señal de energía finita, sería necesario un número infinito de escalas para recuperar todo el espectro de la señal original.

Si sólo se pudiese realizar esta transformación hasta un número finito de escalas ' $j < J$ ' y se quisiese reconstruir la señal original con las escalas disponibles, sería necesario complementar la información correspondiente a ' C_{jk} ' para escalas superiores a J ; en ese caso se tendría:

$$(3.2.8) \quad f(t) = \sum_{j \leq J} D_j(t) + \text{complemento} = \sum_{j \leq J} D_j(t) + A_j(t)$$

siendo ' $A_j(t)$ ' la subseñal de aproximación. De la ecuación anterior, resulta obvio que las subseñales de aproximación se hallan relacionadas a diferentes escalas mediante:

$$(3.2.9) \quad A_{j-1} = A_j + D_j$$

En un análisis ortogonal, en el que la familia de ondículas ' $\psi_{jk}(t)$ ', cumple los requisitos de ortonormalidad, la subseñal de aproximación constituye el complemento ortogonal de la subseñal de detalle.

Para obtener esta subseñal de aproximación se recurre a otras funciones base ' $\phi_{jk}(t)$ ', denominadas funciones de escalado, de tal forma que (3.2.5) se reescribe como:

$$(3.2.11) \quad f(t) = \sum_{k=-\infty}^{\infty} C_{Jk}^A \phi_{Jk}(t) + \sum_{j \leq J} \left(\sum_{k=-\infty}^{\infty} C_{jk} \psi_{jk}(t) \right) = A_J + \sum_{j \leq J} D_j$$

Los coeficientes de la señal de aproximación se obtienen mediante la expresión:

(3.2.12)
$$C_{Jk}^A = \int_{-\infty}^{\infty} f(t) \phi_{Jk}(t) dt$$

En la ecuación (3.2.11), las ondículas cubren el espectro hasta la escala 'J', el resto de información se obtiene de la función de escalado; por tanto, se tiene ahora un número finito y limitado de ondículas, lo que conduce al llamado análisis multiresolución porque genera una descomposición jerárquica de la señal en señales de detalle y de aproximación, que proporcionan información referente a las fluctuaciones y tendencias registradas en diferentes escalas (frecuencias), respectivamente [37].

Así, para un nivel de descomposición '3' se generaría el esquema de descomposición mostrado en la figura 14 en la cual 'S' representa la señal original y los coeficientes nombrados con las letras 'A' y 'D' indicarían los coeficientes de aproximación y de detalle, respectivamente. En esta figura, cada número representa el nivel de descomposición empleado.

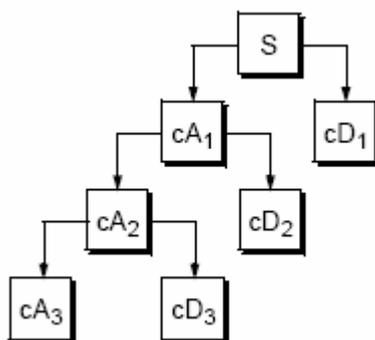


Figura 14. Esquema de descomposición de una señal mediante la transformación DWT.

Aplicando esto mismo a una señal cualquiera, se obtiene el esquema indicado en la figura 15 (**). En este caso, el árbol jerárquico de descomposición se ha obtenido con la familia de ondículas de Daubechies de tercer orden con dos niveles de descomposición.

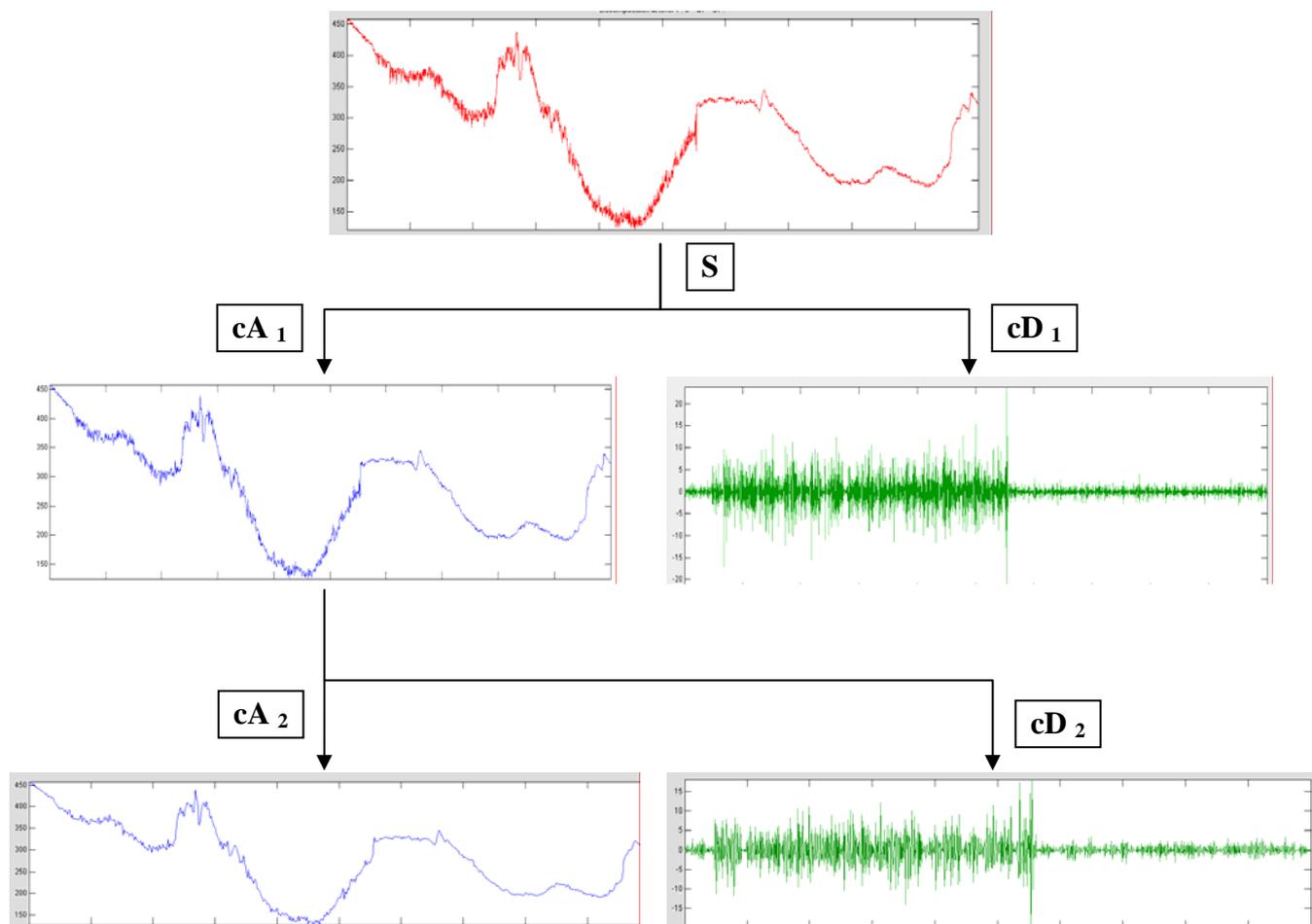


Figura 15. Esquema de descomposición de una señal en subseñales de detalle y de aproximación.

** Para la descomposición se eligió la señal “nelec” del ejemplo de ondículas suministrado por MATLAB.

Como para la mayor parte de las funciones las transformaciones con ondículas no tienen solución analítica, éstas se calculan numéricamente mediante el algoritmo de la transformada rápida FWT desarrollado por Mallat (1989); como la transformación con ondículas se puede interpretar como una operación de filtrado, tal y como se comentó anteriormente, este algoritmo utiliza una serie de filtros paso bajo y paso banda para obtener los coeficientes de las subseñales de aproximación y de detalle, respectivamente [38].

De cara a un proceso de diagnóstico de daño, las señales más atractivas son las señales de detalle dado que son más sensibles a la presencia de fluctuaciones en el soporte de la señal original provocadas por la presencia de fallos en la maquinaria. Por otro lado, las frecuencias más interesantes son aquellas que sean múltiplos naturales de la frecuencia de fallo; de esta forma, el nivel de descomposición alcanzado en el análisis multiresolución se determinará por aquellas escalas inferiores a la frecuencia característica de fallo de la máquina.

3.2.3. ELECCIÓN DE LA FAMILIA DE ONDÍCULAS

La elección de la familia de ondículas adecuada para transformar la señal original se basa principalmente en dos propiedades fundamentales, el número de momentos nulos y el tamaño del soporte de la ondícula, aunque la regularidad de la señal a analizar también se debe considerar.

Con la transformación WT se persigue que cualquier fluctuación de la señal original, ocasionada por un defecto, proporcione coeficientes de transformación lo más altos posibles, mientras que el resto de la señal no perturbada dé lugar a coeficientes lo más bajos posibles. Este efecto se halla directamente ligado a las dos propiedades anteriormente mencionadas.

3.2.3.1 Momentos nulos: El número de momentos nulos de una ondícula determina el orden del polinomio que se puede aproximar. Se dice que una ondícula ' ψ ' tiene ' p ' momentos nulos si se cumple:

$$(3.2.13) \quad \int_{-\infty}^{\infty} t^k \psi(t) dt = 0 \quad \text{para } 0 \leq k < p$$

Por tanto, ondículas con ' p ' momentos nulos generarán coeficientes WT muy bajos para señales polinomiales de orden ' $p-1$ ' y cualquier pico de orden superior producirá coeficientes elevados.

3.2.3.2 Soporte: Se define el soporte de una ondícula como el tramo temporal en que su valor es distinto a cero.

Esta propiedad permite determinar el número de coeficientes WT afectados por una perturbación contenida en el soporte de la ondícula; a mayor soporte mayor número de coeficientes afectados. Por otra parte, según el teorema de Parseval de conservación de la energía de la señal, una mayor dispersión de coeficientes influenciados implicará menor magnitud en los mismos. Por tanto, interesan soportes lo más compactos posibles sobre la irregularidad local de la onda.

Estas dos propiedades son independientes entre sí; sin embargo, para la familia de ondículas ortogonales (que es la que se empleará aquí por ser la que menos coeficientes emplea para reconstruir la señal original) se puede demostrar que si ' ψ ' tiene ' p ' momentos nulos entonces su soporte es al menos del tamaño ' $2p-1$ ' [42].

Como consecuencia, ondículas ortogonales muy compactas implican muy pocos momentos nulos y, por tanto, hay que llegar a una solución de compromiso una vez se tenga identificada la regularidad de la señal a transformar.

La familia de ondículas de Daubechies [42] tiene el mínimo soporte para un determinado número de momentos nulos; de ahí que esta familia resulte idónea para el sistema de detección y diagnóstico.

No obstante, si se conociesen de antemano las características del pulso perteneciente a la señal degradada se podría optar por alguna otra familia.

A fin de ilustrar mejor las dos propiedades citadas, considérese un pulso cuadrático (figura 16) con discontinuidades en sus extremos (lo que implica exponentes inferiores a la unidad y una alta influencia a bajas escalas).

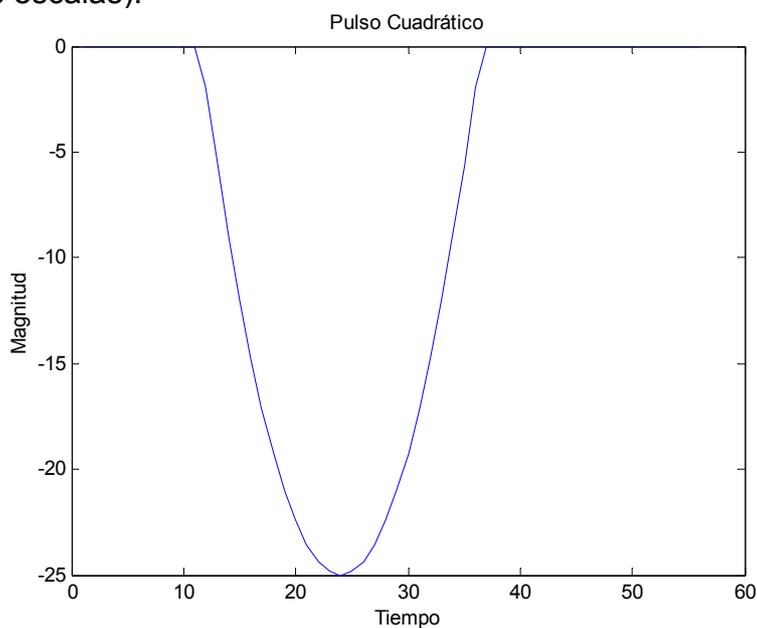


Figura 16. Pulso cuadrático con discontinuidades.

Esta señal se ha transformado con ondículas Daubechies de órdenes '1', '2' y '3', respectivamente, obteniéndose los resultados mostrados en la figura 17, donde se ilustran los escalogramas para cada transformación. En abscisas se representa el eje de tiempos y en ordenadas las escalas.

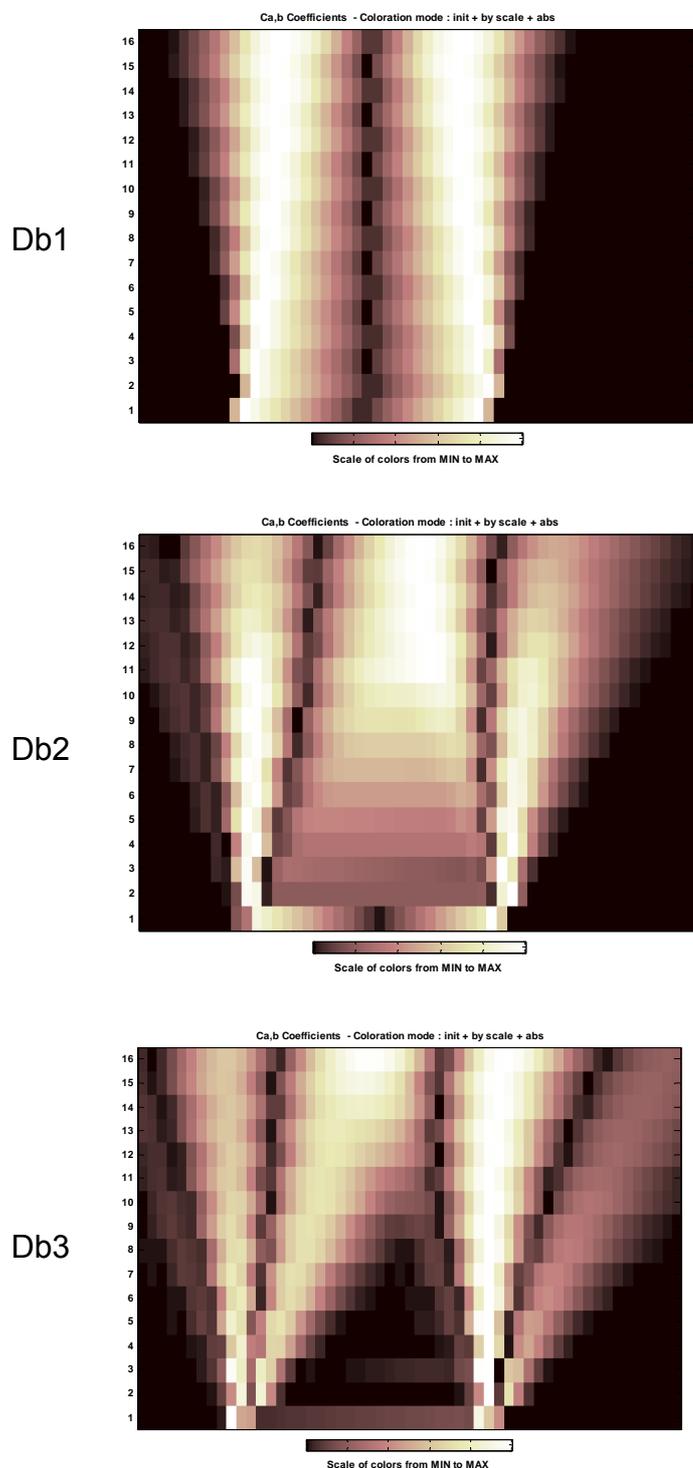


Figura 17. Transformación WT del pulso cuadrático empleando la familia Daubechies con distinto número de momentos nulos.

De estas gráficas se pueden extraer las siguientes conclusiones:

a) A menor soporte, menor dispersión de coeficientes afectados por un evento determinado.

Todas las familias empleadas detectan las singularidades de los extremos del pulso, la diferencia estriba en la anchura de los conos de influencia de cada transformación.

b) El pulso cuadrático no es detectado con la transformación 'Db1' dado que ésta únicamente posee un momento nulo; sin embargo, dicho pulso sí es detectado en las transformaciones restantes.

En este caso, si interesara resaltar este pulso cuadrático se elegiría la familia de ondículas 'Db2' que es la que presenta una mayor concentración de coeficientes afectados por la regularidad que interesa resaltar.

3.2.4. CARACTERIZACIÓN Y LOCALIZACIÓN DE LA REGULARIDAD DE UNA SEÑAL MEDIANTE WT.

De cara a alcanzar el objetivo de esta Tesis, es esencial entender cómo caracterizar desde un punto de vista matemático un pulso y ver sus efectos de propagación sobre los coeficientes obtenidos mediante la transformación WT en el escalograma; además, parece interesante considerar los efectos de ruido de fondo que pueden aparecer en una adquisición real.

3.2.4.1. Definición matemática de regularidad de una señal.

La regularidad de una señal se evalúa habitualmente mediante los exponentes de Lipschitz ya que, de alguna forma, dichos exponentes dan una medida de la diferenciabilidad de la función con ciertas sutilezas como se verá a continuación.

Sea 'n' un entero positivo tal que ' $n \leq \alpha \leq n+1$ '. Se dice de una función que es "Lipschitziana" de orden ' α ', si y sólo si existen dos constantes 'A' y ' $h_0 > 0$ ', y un polinomio de orden 'n', ' $P_n(t)$ ', tal que para ' $h < h_0$ ', se tiene:

$$(3.2.14) \quad |f(t_0 + h) - P_n(h)| \leq A|h|^\alpha$$

La regularidad "Lipschitziana" de ' $f(t)$ ', se caracteriza mediante el límite superior del exponente ' α ' en (3.2.14). Si el exponente de Lipschitz es inferior a la unidad en un punto se dice que la función es singular en dicho punto.

Una función acotada en un instante 't' pero discontinua en dicho instante es Lipschitziana de orden cero. Si ' $\alpha < 1$ ' en 't', entonces ' $f(t)$ ', no es diferenciable en 't' y este exponente ' α ' caracteriza la medida de la singularidad.

Como conclusión, la información que se puede extraer de los exponentes de Lipschitz se refiere a la diferenciabilidad de una función, define si la derivada de orden ' α ' está acotada pero no la continuidad de la misma.

Sea ' α ' un número real no entero y el intervalo $[a,b] \in \mathfrak{R}$. La función ' $f(t)$ ', se dice que es uniformemente "Lipschitziana" de orden ' α ' en el intervalo $]a,b[$, si y sólo si su primitiva es uniformemente "Lipschitziana" de orden ' $\alpha + 1$ ' en dicho intervalo.

Hay que advertir que esta definición es global al referirse a intervalos y no a puntos.

3.2.4.2. Herramientas de cuantificación.

a) Herramientas de cuantificación clásicas.

Una vez establecida la definición formal de regularidad de una función o de una señal, la siguiente cuestión radica en cómo cuantificar o evaluar las irregularidades cuando aparecen.

Una herramienta clásica para medir la regularidad de Lipschitz de una función es medir el decaimiento asintótico de su transformada de Fourier. Se puede demostrar que una señal de energía finita ' $f(t) \in L(\mathfrak{R})^2$ ', es "Lipschitziana" de orden ' α ' si cumple la siguiente relación:

$$(3.2.15) \quad \int_{-\infty}^{+\infty} |\hat{f}(w)| (1+|w|^\alpha) dw < +\infty$$

Esta última ecuación, aunque no se va a demostrar, se puede justificar de una forma no demasiado ortodoxa desde el punto de vista matemático.

Anteriormente, se vio que la diferenciabilidad de una función se puede determinar mediante los coeficientes de Lipschitz; por tanto, un exponente α que verifique " $n \leq \alpha \leq n + 1$ " indicaría que la derivada de orden 'n' de esta función está acotada y por tanto $f^n(t) \in L(\mathfrak{R})^2$.

Por otro lado, derivar una función en el dominio temporal equivale a multiplicar su transformada en el dominio de Fourier por términos $'i\omega'$; atendiendo a estas consideraciones previas se puede concluir que la ecuación (3.2.15) expresa que una señal $'f(t)'$ y su derivada de orden α son señales de energía finita y por tanto están acotadas.

Hay que advertir que la condición expresada según la ecuación (3.2.14) no es una condición necesaria pero sí suficiente y además mide la regularidad de la señal globalmente.

Una señal de ruido blanco, que tiene un espectro uniforme y abarca todo el ancho de frecuencias; requiere un exponente de Lipschitz negativo para que se cumpla (3.2.15). Por tanto, representa una función singular.

Por el mismo motivo la función impulso de Dirac definida de la forma:

$$\delta(\tau) \ni \int_{-\infty}^{\infty} \delta(\tau) d\tau = 1$$

es también singular.

Este aspecto puede resultar muy interesante en lo referente al filtrado del ruido como se verá más adelante

b) Herramientas de cuantificación basadas en la transformación WT.

A continuación se va a tratar de caracterizar la regularidad a través de los coeficientes de Lipschitz, previo empleo de la transformada WT. Se demostrará que esta herramienta facilita la estimación de estos coeficientes con la ventaja de que dichas estimaciones se efectúan a nivel local, bien por intervalos, bien por puntos.

Sea $f(t) \in L(\mathfrak{R})^2$ y $[a,b]$ un intervalo de \mathfrak{R} . Considérese $0 \leq \alpha \leq 1$. Para cualquier $\varepsilon > 0$, $f(t)$ es uniformemente “Lipschitziana” de orden ‘ α ’ en el intervalo ‘ $]a + \varepsilon, b - \varepsilon[$ ’, si y sólo si existe una constante ‘A’ para cada $\varepsilon > 0$ y para cada punto ‘ t_0 ’ del intervalo tal que:

$$(3.2.16) \quad |Wf(s, t_0)| \leq A_\varepsilon S^\alpha$$

La demostración de la expresión (3.2.16) puede encontrarse en la referencia [39].

Sin embargo, se puede llegar a esta última expresión de una forma intuitiva partiendo de la ecuación (3.2.15); la condición de suficiencia (3.2.15) indica un decaimiento de la transformada de Fourier $\hat{f}(w)$ para grandes frecuencias, superior a $1/w^\alpha$; por tanto, la ecuación (3.2.16) es equivalente a (3.2.15) si se consideran las escalas ‘s’ semejantes a las inversas de las frecuencias. Sin embargo, a diferencia de la condición (3.2.15), establecida sobre la transformada de Fourier de la señal, la condición (3.2.16) tiene el carácter de necesaria y suficiente y además es una restricción que al definirse sobre intervalos permite determinar la regularidad de las señales a analizar sobre dichos intervalos.

En virtud de lo anterior se puede plantear el determinar la regularidad (o lo que es lo mismo, los coeficientes de Lipschitz) de una señal en un intervalo temporal analizando el decaimiento logarítmico de los coeficientes obtenidos en su transformación con ondículas. Esto resulta muy interesante ya que con un tratamiento de Fourier no se podría analizar la regularidad sobre intervalos temporales sino sólo a nivel global.

Por otra parte, si se detectan perturbaciones se caracterizarán por exponentes ‘ α ’ grandes lo cual implica que se produzcan coeficientes de transformación elevados a escalas crecientes; sin embargo, las singularidades se verán atenuadas a escalas crecientes al ser sus exponentes

inferiores a la unidad. Esto es lo que anteriormente se denominó efecto “lupa” y supone un rasgo de las WT fundamental a efectos de detección de daño.

La expresión (3.2.16) muestra el efecto de una perturbación de regularidad ‘ α ’ en el plano escala frecuencia. Sin embargo, esta caracterización se da a intervalos no en un punto temporal preciso. El teorema que se va a enunciar a continuación proporciona la herramienta complementaria necesaria para determinar la regularidad de una señal en un instante determinado.

Sea ‘ $f(t) \in L(\mathbb{R})^2$ ’, y ‘ n ’ un entero positivo correspondiente al número de momentos nulos de la ondícula ‘ $\psi(t)$ ’. Considérese que $\alpha \leq n$; si ‘ $f(t)$ ’ es “Lipschitziana” de orden ‘ α ’ en ‘ t_0 ’, entonces existe una constante A tal que para cada instante de tiempo alrededor del momento considerado ‘ t_0 ’ y para cualquier escala se verifica:

$$(3.2.17) \quad |Wf(s, t_0)| \leq A (S^\alpha + |t - t_0|^\alpha)$$

Por otra parte si $\alpha \leq n$ pero ‘ n ’ no es un entero, se dice que la función ‘ $f(t)$ ’ es “Lipschitziana” de orden ‘ α ’ si cumple estos dos prerequisites:

1. Existen un valor de ‘ $\varepsilon > 0$ ’ y una constante ‘ A ’ para los cuales, en la proximidad del instante considerado ‘ t_0 ’ y para cualquier escala, se cumple:

$$(3.2.17i) \quad |Wf(s, t)| \leq AS^\varepsilon$$

2. Existe un valor de la constante 'B' para el que en la proximidad del instante considerado 't₀' y para cualquier escala se cumple:

$$(3.2.17ii) \quad |Wf(s,t)| \leq B \left(S^\alpha + \frac{|t-t_0|^\alpha}{|\log|t-t_0||} \right)$$

La demostración de este teorema se puede encontrar en la referencia [40].

A partir de las ecuaciones (3.2.16) y (3.2.17) se dispone de las herramientas necesarias para evaluar la regularidad de cualquier señal en un instante de tiempo, en un intervalo, o globalmente en todo el soporte temporal en el que dura la señal.

3.2.4.3. Influencias sobre el escalograma.

Una vez cuantificada la regularidad de una señal con los coeficientes de la transformada, la siguiente cuestión radica en cómo determinar la influencia de este evento de regularidad 'α' en el escalograma.

a) Localización temporal:

Supóngase que interesa determinar cuáles son los coeficientes afectados por una perturbación caracterizada por su exponente de Lipschitz 'α' en el eje de tiempos. Si este evento tiene lugar en el tiempo 't₀' y la ondícula madre 'ψ(t)' cubre el intervalo [-t₁,t₂], entonces su versión escalada 'ψ_{j₀}(t)', en el caso de optar por la transformación DWT, estará localizada en el rango [-t₁2^j, t₂2^j] y por tanto 'ψ_{j_k}(t)', estará soportada en el intervalo [(-t₁+k)2^j, (t₂+k)2^j].

Por lo tanto, los coeficientes afectados por la perturbación serán los localizados en el rango temporal:

$$(3.2.18) \quad (-t_1 + k)2^j \leq t_0 \leq (t_2 + k)2^j$$

Esto es lo que se llama cono de influencia de la perturbación, y en el escalograma aparece como una especie de embudo cuyo vértice está localizado en el origen de la perturbación. (Ver figuras 18 y 19).

En las figuras 18 y 19 se analiza, previo empleo de la transformada CWT, una señal

$f(t) = \left| \frac{1}{x^2 - 4} \right| + \text{sen}(x)$ caracterizada por la existencia de un pico en una localización determinada de la escala temporal (transformada Morlet compleja).

Se puede apreciar cómo la perturbación genera en el escalograma el cono de influencia anteriormente comentado.

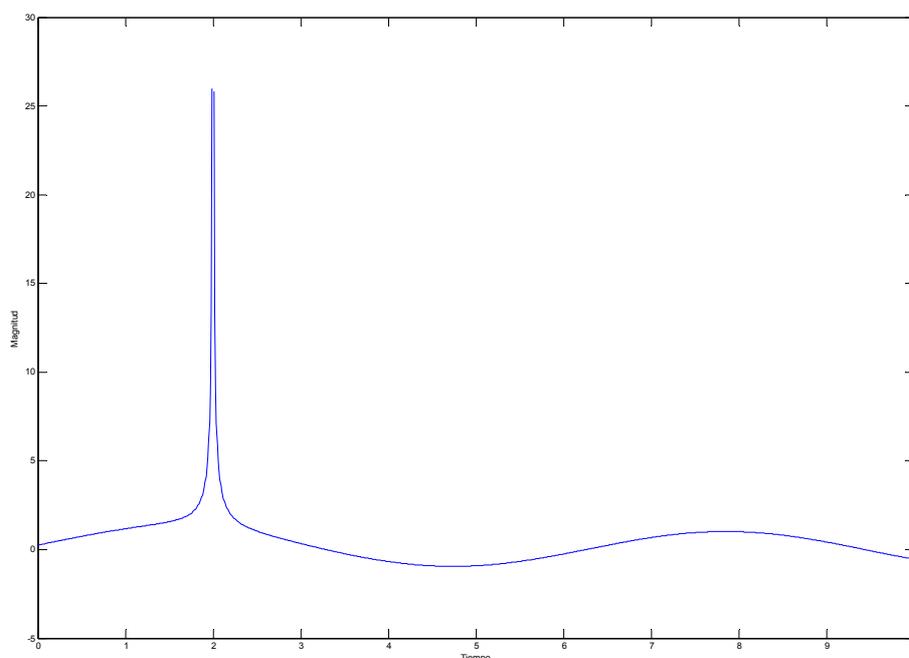


Figura 18. Perturbación generada en una señal temporal “suave”.



Figura 19. Transformada CWT de la señal perturbada.

b) Localización en frecuencia:

Dado que la transformada de Fourier de la ondícula madre $\psi_{jk}(t) = 2^{-j/2}\psi(2^{-j}t - k)$ viene dada por $2^{j/2}\Psi(2^{-j}\omega)e^{-i2^j k\omega}$, de (3.2.4), aplicando la fórmula de Parseval, se tiene:

$$(3.2.19) \quad C_{jk} = \frac{1}{2\pi} 2^{m/2} \int_{-\infty}^{\infty} F(\omega) \Psi^*(2^j \omega) e^{-i2^j k\omega} d\omega$$

siendo 'F(w)' y ' $\Psi^*(w)$ ', las transformadas de Fourier de la señal temporal y de la ondícula madre, respectivamente. El símbolo '*' indica la operación de conjugación compleja.

Supóngase ahora que el soporte temporal de la ondícula madre viene dado por el ancho de

banda $[\omega_{\min}, \omega_{\max}]$; a la escala 'j' este ancho de banda será $[\frac{\omega_{\min}}{2^j}, \frac{\omega_{\max}}{2^j}]$.

Por consiguiente, una componente espectral ' ω_0 ' de la señal original afecta a la escala ' j ' de la transformada DWT en el ancho de banda:

$$(3.2.20) \quad \frac{\omega_{\min}}{2^j} \leq \omega_0 \leq \frac{\omega_{\max}}{2^j}$$

y, por tanto, el siguiente rango de escalas se verá influenciado:

$$(3.2.21) \quad \log_2 \left(\frac{\omega_{\min}}{\omega_0} \right) \leq j \leq \log_2 \left(\frac{\omega_{\max}}{\omega_0} \right)$$

3.2.4.4. Influencias originadas por el ruido de fondo

Una vez caracterizada la regularidad de una fluctuación y su efecto sobre el escalograma, parece interesante caracterizar también los distintos tipos de ruido que se pueden superponer a la señal medida y su influencia al pasar al dominio escala-tiempo mediante la WT.

El término ruido se utiliza comúnmente para designar las señales aleatorias no deseadas que aparecen en los sistemas y sobre las que no se tiene ningún control.

Normalmente se caracterizan por una distorsión o "contaminación" que se superpone sobre la señal determinista de interés.

Existen dos tipos de ruido:

1. Ruido externo al sistema: ocasionado por causas ajenas e independientes al sistema como por ejemplo las provocadas por el entorno ambiental, por acciones humanas, etc.
2. Ruido interno del sistema: por ejemplo el ruido eléctrico de los sensores empleados para instrumentar, etc.

En el caso considerado en esta tesis, el ruido interno es el que más afecta ya que se van a registrar aceleraciones; si se estuviese interesado en registrar medidas acústicas habría que considerar además el efecto del ruido externo ya que los factores ambientales sí que podrían jugar un papel sumamente importante en dichos registros.

Los efectos generados por el ruido pueden suponer una limitación en la capacidad de detección del sistema de diagnóstico debido a la pérdida de sensibilidad de medición de la instrumentación que se vaya a utilizar y a la “contaminación” de las medidas.

Por ello, es muy importante considerar y analizar en detalle los efectos del ruido eléctrico interno porque afectarán a la robustez del sistema de detección y constituyen un factor que incidirá directamente tanto en las falsas alarmas como en la pérdida de “sensibilidad de detección”.

Este ruido eléctrico tiene dos principales contribuciones que son el ruido impulsivo y el ruido térmico.

Además, según el teorema central límite, se pueden considerar conjuntamente contribuciones en lo que se denomina ruido de banda ancha de naturaleza Gaussiana.

Cualquier proceso aleatorio es estacionario y ergódico si es independiente del tiempo y, además, cualquier muestra es representativa de todo el conjunto de patrones muestra. El ruido asociado a la señal vibratoria a analizar implica frecuencias elevadas por lo que su período característico es muchísimo menor que los periodos asociados a cada vuelta del eje y a cada paso de diente de los engranajes. Por tanto, se puede considerar dicho ruido como

estacionario y, además, se va a suponer que cualquier muestreo de ruido sobre una vuelta del eje va a ser representativo del ruido presente durante todo el tiempo de adquisición de datos.

Evidentemente, al ser los sistemas de adquisición digitales la máxima frecuencia asociada a este ruido vendrá limitada por el mínimo periodo de muestreo del sistema de adquisición.

Aparte de la ergodicidad y estacionariedad, se puede considerar el ruido aleatorio como un proceso de banda ancha, lo que en teoría de señal suele denominarse como ruido “rosa” [41], estando su espectro caracterizado por la historia temporal y el espectro ilustrados en la figura 20.

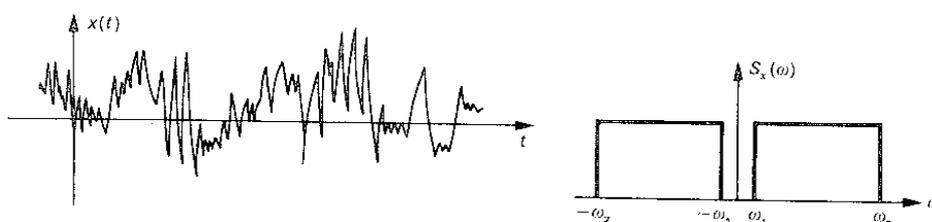


Figura 20. Proceso aleatorio de banda ancha.

A continuación, se va a analizar la transformación con ondículas de una señal real $f(t) \in L(\mathbb{R})^2$, con un ruido de fondo superpuesto:

$$(3.2.22) \quad f(t) = g(t) + h(t)$$

en donde $g(t) \in L(\mathbb{R})^2$ representa la señal vibratoria determinista subyacente y $h(t)$ el ruido aleatorio contaminante. Se va a demostrar que el grado de influencia de las contribuciones de la señal determinista y del ruido dará lugar a aportaciones bien diferenciadas dependientes de la escala seleccionada.

El valor cuadrático de la señal previa vendrá dado por la relación:

$$(3.2.23) \quad f^2(t) = g^2(t) + h^2(t) + 2g(t)h(t)$$

Teniendo en cuenta que la transformación con ondículas es un operador lineal, se tiene que:

$$(3.2.24) \quad Wf^2(s,t) = Wg^2(s,t) + Wh^2(s,t) + 2(Wg(s,t) \cdot Wh(s,t))$$

Evaluando la esperanza matemática de la expresión anterior y suponiendo que el ruido tiene un promedio nulo se tiene:

$$(3.2.25) \quad \begin{aligned} E\left[|Wf(s,t)|^2\right] &= |Wg(s,t)|^2 + E\left[|Wh(s,t)|^2\right] \\ E\left[|Wh(s,t)|^2\right] &= E\left[|h(t) * \psi_s(t)|^2\right] = E\left[h(t)^2 * \psi_s^2(t)\right] = E\left[h(t)^2\right] * E\left[\psi_s^2(t)\right] \end{aligned}$$

Dado que el valor cuadrático medio del ruido es un valor constante característico del mismo, la anterior convolución se transforma en un producto y finalmente el valor cuadrático medio de la señal con ruido superpuesto vendrá dado por:

$$(3.2.26) \quad E\left[|Wf(s,t)|^2\right] = |Wg(s,t)|^2 + \frac{E\left[h(t)^2\right]}{s^2} \|\psi\|_2^2$$

Siendo $\|\psi\|_2^2$, el valor cuadrático de la norma Euclídeana de la ondícula “madre”.

Para cualquier ruido de banda ancha de naturaleza Gaussiana la anterior ecuación queda reducida a:

$$(3.2.27) \quad E\left[|Wf(s,t)|^2\right] = |Wg(s,t)|^2 + \frac{\sigma^2}{s^2} \|\psi\|_2^2$$

De las ecuaciones (3.2.26) y (3.2.27) se pueden extraer dos conclusiones importantes:

- El ruido de banda ancha genera una contribución que hace que los coeficientes de la transformada crezcan de forma explosiva cuando la escala tiende a cero.
- La varianza del ruido está asociada a la energía de la señal dado que se puede obtener a partir del área obtenida de su espectro de potencia, el cual es un indicador de la energía de la señal y lo habitual es que ésta sea muy pequeña. Por tanto para escalas crecientes, este ruido tiende a desvanecerse.

3.2.5. IMPLEMENTACIÓN MEDIANTE DOS EJEMPLOS NUMÉRICOS

Con objeto de ilustrar los aspectos expuestos en el apartado anterior, se van a mostrar dos ejemplos numéricos. En ambos casos, se va a emplear la transformada continua aunque se obtendrían conclusiones análogas si se realizase una transformación discreta.

Ejemplo 1.

Como primer ejemplo, se va a tomar una función de respuesta en frecuencia de un sistema (figura 21) que viene caracterizada por la existencia de grandes pulsos en las frecuencias de resonancia. Como se sabe, la altura de estos “picos” de la respuesta estructural, es decir la regularidad de la señal, dependerá del valor del amortiguamiento de la estructura.

El interés en este estudio radica en aplicar las técnicas indicadas anteriormente para tratar de identificar las singularidades, si es que existen, su localización y el valor de los exponentes de Holder en estas localizaciones.

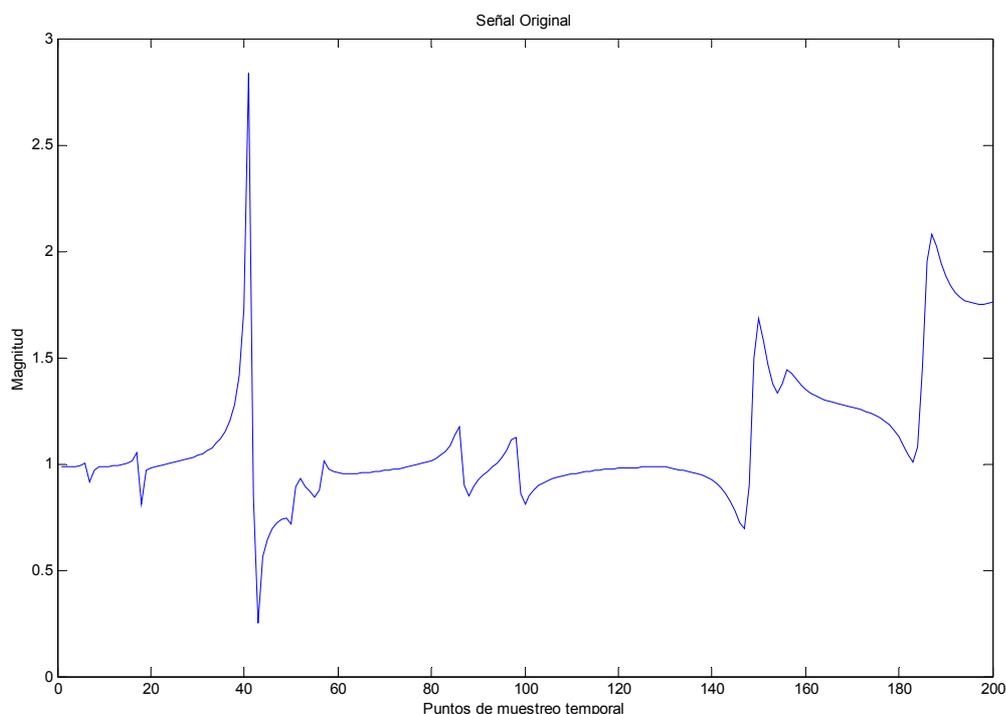


Figura 21. Señal temporal ejemplo 1.

Para el estudio realizado, se va a considerar la figura 21 como un histórico temporal de una señal vibratoria.

El estudio de esta señal se ha efectuado mediante transformación con ondículas de la familia 'Daubechies' de orden '4', lo que implica que se pueden caracterizar coeficientes de Holder o Lipschitz hasta de cuarto orden.

El escalograma resultante de la transformación CWT, obtenido mediante el programa MATLAB, se muestra en la figura 22. En el eje de abscisas se ha representado el dominio temporal y en el de ordenadas las escalas. El mapa de colores indica la magnitud que alcanza cada coeficiente de la transformación WT.

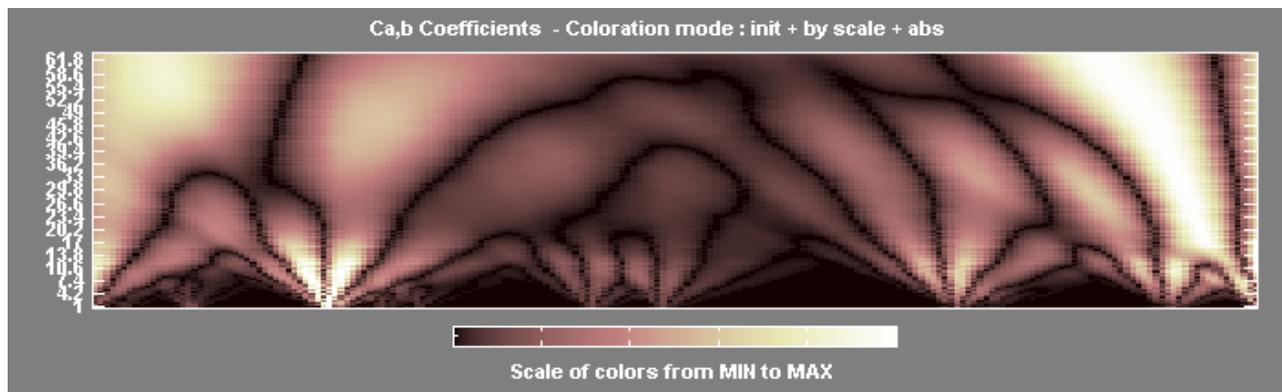


Figura 22. Transformada CWT de la señal del ejemplo 1.

Como cabría esperar, los pulsos en las resonancias se pueden caracterizar por polinomios de alto orden que se propagan en el escalograma a lo largo de su cono de influencia.

Para determinar la regularidad de la señal en estos posibles puntos de resonancia, se va a examinar el valor del decaimiento del módulo de los coeficientes de la transformación en el cono de influencia de dichas regiones.

Para facilitar esta última tarea, se representa el escalograma de la figura 22 en un diagrama de relieves tridimensional (figura 23). La localización de las zonas de máximos locales próximas a escalas nulas (ya que las singularidades corresponden a exponentes negativos y por tanto su influencia debe de ser apreciable a escalas próximas a cero) permite determinar la pendiente y el decrecimiento de estos máximos a fin de estimar los exponentes de Holder o Lipschitz mediante la ecuación (3.2.16). En general, se está hablando de un rango de escalas comprendido entre 0 y 10 unidades, aproximadamente.

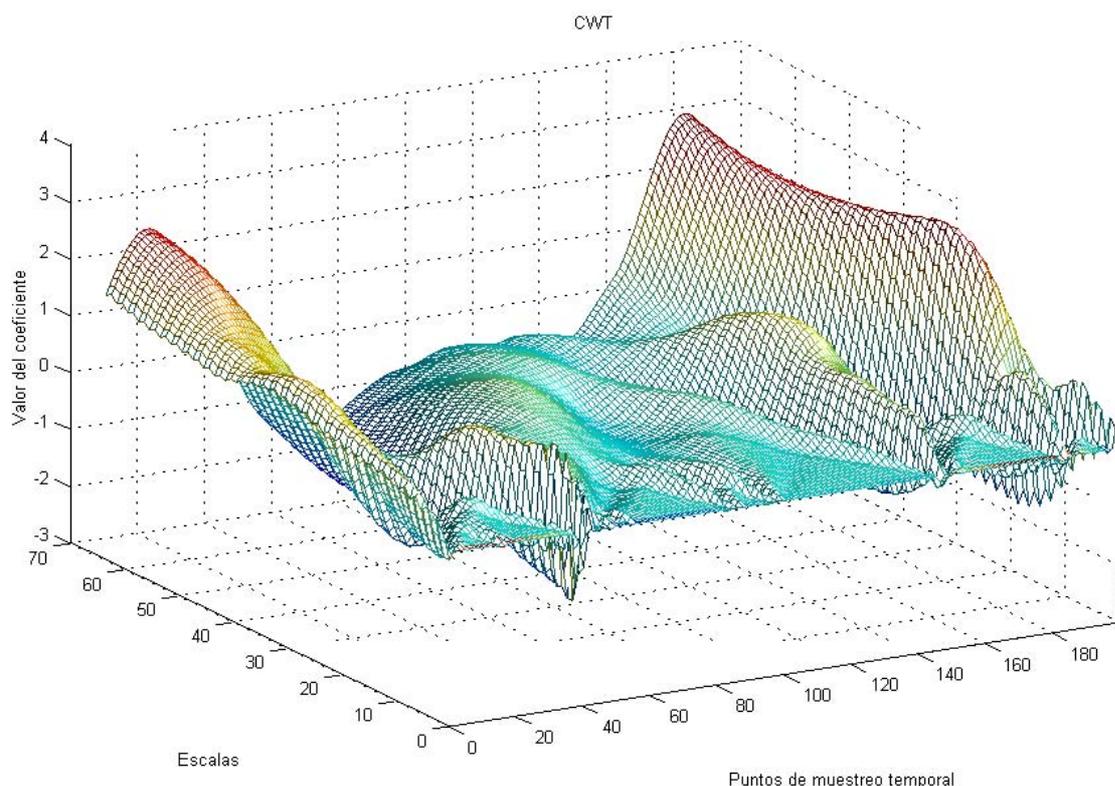


Figura 23. Mapa de relieves 3D de los coeficientes CWT de la señal del ejemplo 1.

Las zonas más próximas a las abscisas temporales 40 y 80-90 representan puntos de singularidad que, sin duda, corresponden a frecuencias de resonancia del sistema estructural. Como los exponentes son superiores a cero (en concreto en ambas localizaciones son 0.8 y

0.7, respectivamente) su efecto se intensifica con la escala, lo cual resulta muy interesante a efectos prácticos ya que la tendencia del ruido, según lo comentado anteriormente, es justamente la contraria, es decir atenuarse al aumentar el nivel de escalas.

Ejemplo 2

Como segundo ejemplo, se ha considerado la misma señal de la figura 21 pero con “ruido” de fondo superpuesto. El objetivo fundamental es mostrar el desvanecimiento del efecto “contaminante” del “ruido” con la escala según las ecuaciones (3.2.26) y (3.2.27). Para cuantificar tal efecto se determina a qué nivel de escala la contribución del ruido al valor absoluto del coeficiente de la transformada es de un orden de magnitud inferior a la aportación debida a la señal subyacente (señal pura sin contaminar por el ruido) dentro del cono de influencia de la singularidad correspondiente a la abscisa ‘40’.

Mediante la herramienta SIMULINK de MATLAB, se ha añadido un ruido Gaussiano a la señal con un ancho de banda comprendido entre cero y cinco hertzios, con un valor promedio nulo y cuyo valor cuadrático medio es de ‘0.1997 unidades’ (este valor ha sido calculado por integración del espectro de potencia en SIMULINK). La señal resultante contaminada se muestra en la figura 24.

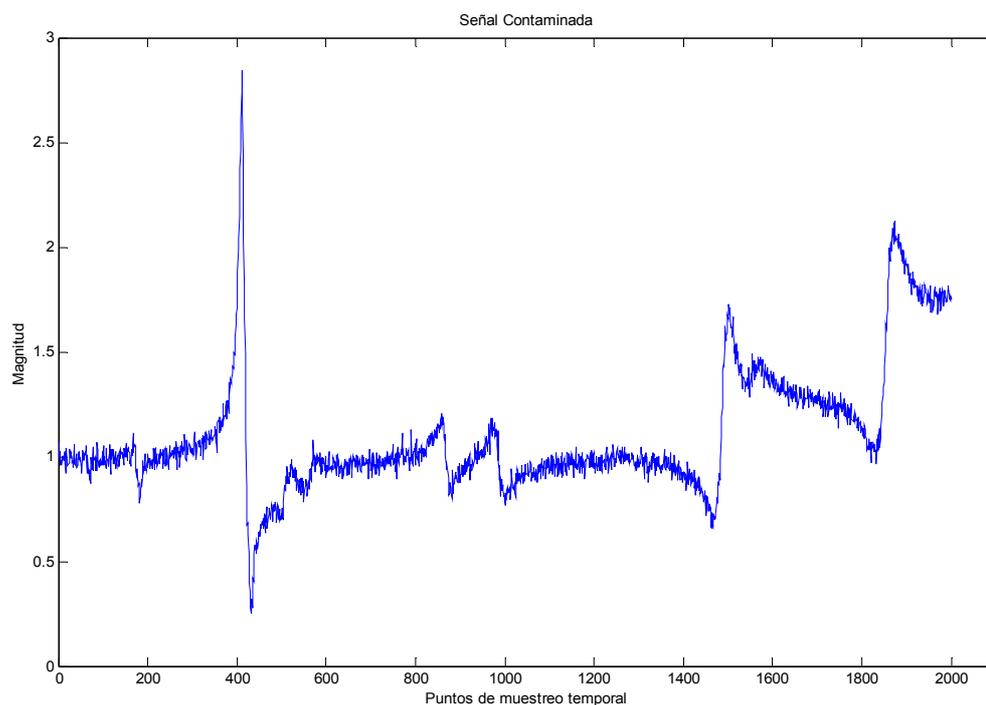


Figura 24. Señal temporal con ruido.

Para cuantificar el desvanecimiento del ruido en el escalograma de la transformación, si se desea que la contribución de dicho ruido a la transformación sea de un orden de magnitud inferior (factor 10) que la aportación debida a la señal subyacente, se tiene de (3.2.17) y (3.2.27) la siguiente relación en el cono de influencia de la singularidad correspondiente a la abscisa '40':

$$(3.2.28) \quad A^2 S^{2\alpha} = 100 \cdot \frac{E[f(t)^2]}{S^2}$$

En esta ecuación los valores de 'A' y 'α' se han extraído del ejemplo anterior mientras que el valor cuadrático medio se ha obtenido por integración directa con SIMULINK. En función de los valores calculados, se puede predecir que el ruido se desvanecerá a partir de la escala:

(3.2.29)

$$S = 3.7$$

Esto se puede comprobar cualitativamente en un mapa de relieves de la transformada de la señal con “ruido”, realizando una aproximación en un diagrama de detalle sobre la zona de escalas próximas al valor anteriormente calculado en la singularidad de abscisa ‘40’ (ver figura 25). Como el número de puntos de muestreo por paso temporal es igual a diez puntos por segundo se tiene un total de dos mil puntos temporales de muestreo en el eje de abscisas. Además, en esta figura se han omitido las escalas próximas a cero ya que en la proximidad de estas zonas la contribución del “ruido” al valor de los coeficientes de la transformación se hace explosiva y caótica tal y como cabría esperarse de las ecuaciones (3.2.26) y (3.2.27)

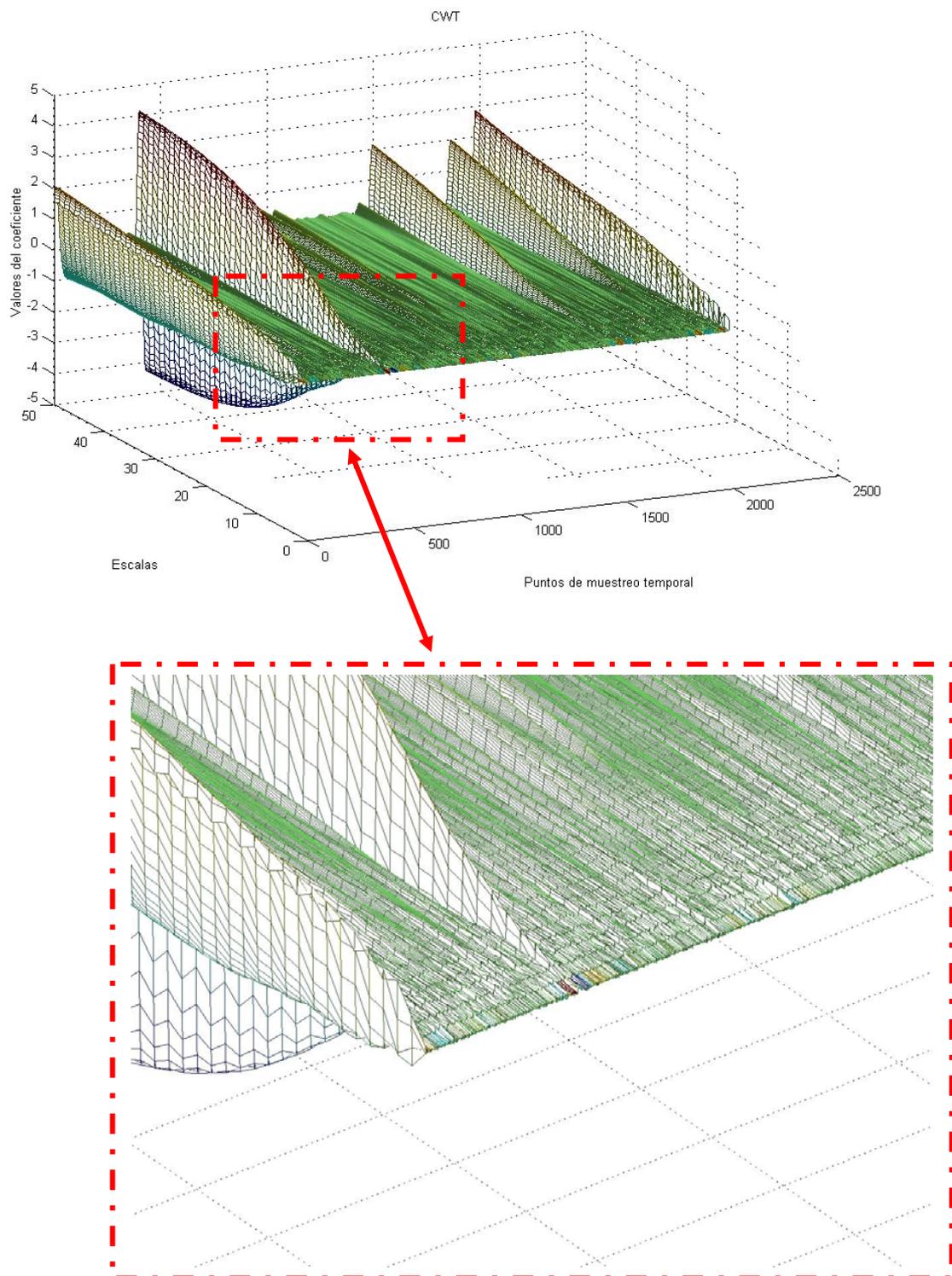


Figura 25. Detalle de los coeficientes CWT de la señal con ruido.

En la figura 25 se puede apreciar cómo el relieve, que es también común al de la figura 23 en la abscisa '40', empieza a “despegar” y a elevarse a partir del nivel de escala '3' tal y como se había predicho en el análisis previo.

Por otra parte, se puede demostrar que la transformada WT de la señal original (empleando la familia de transformación Daubechies de orden tres) sin ruido para la escala 3.7 produce unos coeficientes en la abscisa 40 de aproximadamente 0.23 unidades, mientras que si se realiza la misma transformación únicamente para el ruido en la misma escala y abscisa, los coeficientes toman un valor de 0.03 unidades; por consiguiente, la contribución del ruido a la señal total (señal determinista más el ruido artificial) es de un orden de magnitud inferior en el valor del coeficiente y, por tanto, se pueden dar por correctas las predicciones efectuadas mediante las expresiones (3.2.28) y (3.2.29).

3.3. DIAGNOSIS MEDIANTE REDES NEURONALES ARTIFICIALES.

En esta sección se van a introducir las redes neuronales artificiales como instrumento de clasificación automática, para el reconocimiento de patrones anómalos o novalos en las señales pretratadas y como herramienta de cuantificación automática del grado de severidad de la degradación del componente mecánico dañado

Se puede indicar, por tanto, que constituyen una herramienta fundamental en el diseño del sistema de diagnóstico propuesto en esta tesis.

Esta sección se organizará de la siguiente manera, en una primera parte se introducen los fundamentos de las redes neuronales artificiales y se muestra su analogía con los sistemas biológicos de aprendizaje, en especial con el cerebro. Se presentarán igualmente los modelos estándares de neuronas y los tipos fundamentales de arquitecturas empleadas. A continuación se comentan cuáles son los beneficios del empleo de las redes neuronales artificiales, en especial para el problema de diagnóstico que esta tesis aborda.

Por último, se muestran dos aplicaciones fundamentales para la estrategia de diagnóstico como son la aplicación de las redes en problemas de regresión y la aplicación de las redes en problemas de detección noval; en todos los casos se justificará el empleo de las topologías más interesantes para abordar estos dos tipos de problemas.

En este tema se omiten las demostraciones de algunos teoremas y se evita el entrar en detalles matemáticos; en cualquier caso se pretenden sentar las bases teóricas necesarias para entender las redes neuronales artificiales como instrumento eficaz para tratar la problemática de diagnóstico y detección en tiempo real.

3.3.1. FUNDAMENTOS DE LAS REDES NEURONALES ARTIFICIALES

Las redes neuronales artificiales tratan de emular el tratamiento de información que realizan los sistemas neurobiológicos aunque, como parece evidente, los dispositivos biológicos son muchísimo más efectivos y rápidos en tareas como el reconocimiento de patrones que cualquier ordenador diseñado por el hombre; a modo de ejemplo, el tiempo de reconocimiento de una foto de un familiar por parte de una persona se estima en 20-100 milisegundos, mientras que esta operación en una computadora podría llevar días.

La efectividad y la rapidez de los sistemas neurobiológicos para resolver problemas los hace especialmente atractivos. Estos sistemas operan de forma muy diferente a cualquier ordenador, así, a grandes rasgos, manipulan la información con sistemas de procesamiento no-lineales, sencillos, masivos y en paralelo que les confiere las características de: paralelismo de cálculo, memoria distribuida y adaptabilidad al entorno.

3.3.1.1. Breve introducción biológica.

Antes de proseguir, parece interesante ver como las neuronas biológicas manipulan las señales, con el objetivo de establecer un paralelismo y tratar de emularlas con sistemas artificiales ya sea mediante hardware o software.

La descripción de las neuronas se debe en gran parte al trabajo de Ramón y Cajal [43] por el que obtuvo el premio Nobel (1911); éstas, a efectos funcionales, pueden definirse como procesadores de información sencillos con un canal de entrada de información, las dendritas, un órgano de cómputo, el soma, y un canal de salida, el axón (figura 26) que transmite información a nuevas neuronas mediante lo que se llama sinapsis.

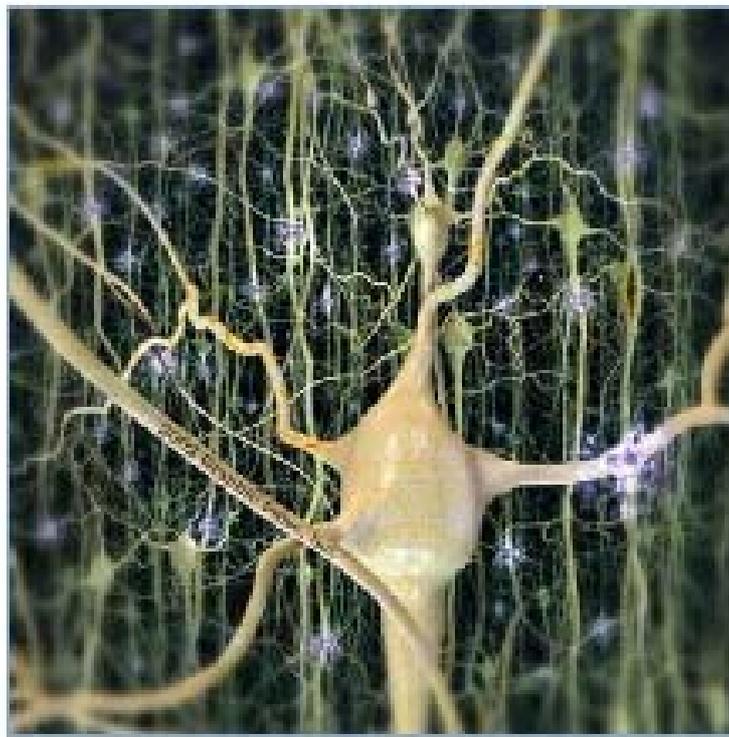


Figura 26. Ilustración de neurona biológica.

En relación a la sinapsis, se habla de neuronas presinápticas (las que envían señales) y postsinápticas (las que realizan las funciones de receptores). La información fluye de forma direccional por lo que las sinapsis son direccionales.

La señal nerviosa puede ser de naturaleza química o eléctrica, en general la transmisión química prevalece fuera de la neurona mientras que la eléctrica lo hace en el interior.

Veamos, de forma somera, cómo tiene lugar la transmisión de una señal nerviosa para una neurona estándar. En estado de reposo la neurona queda cargada con iones $P2^-$ que, dado su tamaño, son incapaces de atravesar su membrana, en desequilibrio con una concentración elevada de iones Na^+ presente en el medio exterior de la neurona y que en principio son

incapaces de atravesar dicha membrana, por tanto el soma hace el papel de condensador manteniendo una diferencia de potencial en reposo de -60 mV, entre ambos medios.

Las neuronas presinápticas, en función del grado de excitación, liberan, lo que se llaman, neurotransmisores que se acoplan en la superficie de la membrana del soma y que la convierten en permeable a los iones de Na⁺, cuando el número de éstos es suficientemente grande en el interior de la neurona, de tal forma que la barrera de potencial decae por debajo de los -45 mV, la diferencia de concentración de iones sodio en el interior y exterior hace que éstos entren en el interior del soma despolarizándolo de forma brusca, de forma que la neurona pasa a tener una diferencia de potencial de + 50 mV. A continuación la neurona vuelve a su potencial de reposo; este proceso da lugar a la generación de una señal eléctrica que se propaga a lo largo del axón.

3.3.1.2. Modelo de neurona artificial

Se llama neurona artificial o unidad de procesamiento, a un dispositivo simple de cálculo que, a partir de un vector de entradas, produce una única salida.

Las partes que constituyen la neurona artificial (figura 27) son las siguientes:

1. Vector de entradas: formado por la señal de excitación, y representado como $x_j(t)$,
2. Pesos sinápticos: Indican el grado de intensidad o excitación de las neuronas presinápticas 'i' en su interacción con la neurona post-sináptica 'j', se representan como 'w_{ij}'.
3. Regla de propagación: Determinan el valor del potencial postsináptico, en función de las señales de entrada y de los pesos. La regla de propagación más habitual viene dada por la suma ponderada de las entradas con los pesos sinápticos:

(3.3.1i)

$$h_i(t) = \sum_j w_{ij} x_j - \theta_i$$

donde 'θ' representa el valor umbral o potencial de activación de cada neurona.

Otro tipo de regla de propagación muy utilizada en las redes, es la basada en la distancia Euclideana:

(3.3.1ii)

$$h_i(t) = \sum_j (x_j - w_{ij})^2$$

4. Función de activación o función de transferencia: Establece el grado de activación de cada neurona en función del estado anterior y del potencial postsináptico actual y establece el valor de la señal de salida (excepto en algunos casos como la máquina de Boltzmann donde a esta función se le aplica otra función estocástica).

Existen numerosos tipos de función de transferencia entre los que caben destacar, por su uso, las funciones de identidad, las funciones sigmoideas y las funciones Gaussianas.

(3.3.2)

$$y_i(t) = f_i(a_i(t-1), h_i(t))$$

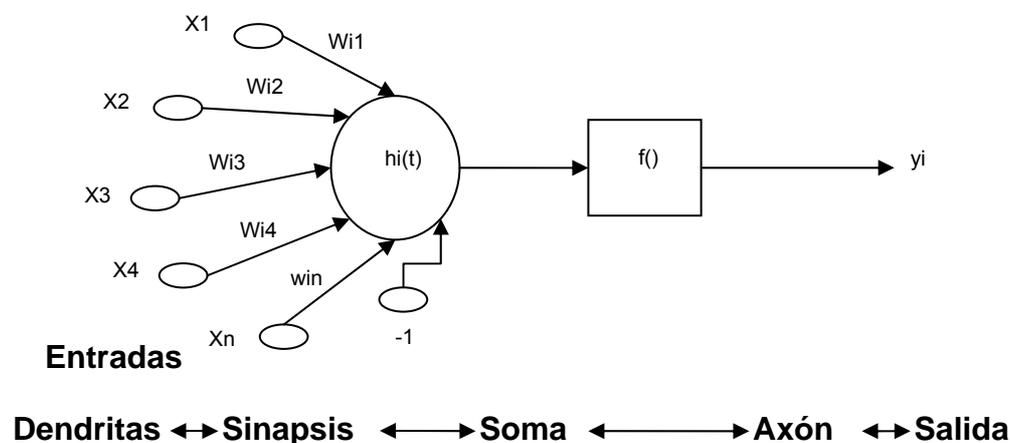


Figura 27. Modelo de neurona artificial.

3.3.1.3. Estructura de la red, modos de operación.

Formalmente una red neuronal se define como un sistema constituido por una serie de procesadores paralelos masivamente distribuidos que tratan de emular al cerebro humano en los dos aspectos siguientes:

El conocimiento se adquiere del entorno a través de lo que se denomina fase de aprendizaje o entrenamiento.

Los pesos sinápticos se emplean para almacenar el conocimiento adquirido.

A partir de las neuronas artificiales se puede pensar en establecer una red artificial, imitando la organización neuronal del cerebro para tratar una determinada información del medio externo.

Éste de forma muy simple utiliza la información capturada por unos receptores y la manipula con un determinado grupo de neuronas agrupadas en torno a capas en áreas específicas y concretas del encéfalo para determinar una decisión que es ejecutada en los actuadores.

El transporte de información y ejecución es unidireccional (figura 28).

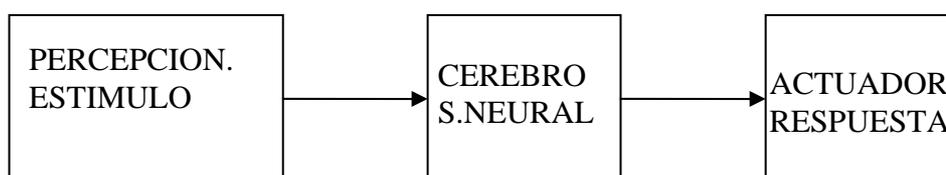


Figura 28. Esquema de transporte de información en el sistema nervioso.

Con este simple esquema es factible la realización de un modelo artificial paralelo, así se puede pensar en la percepción realizada en una capa que directamente proporciona el vector de entradas al resto de la estructura, de igual modo podemos pensar en los efectores como

una capa de neuronas cuyo valor de salida sea la activación de las mismas y que constituyan por tanto la salida de la red, por otra parte es posible el agrupar las neuronas artificiales en capas de un modo similar a las capas del cortex cerebral, por tanto una estructura de red neuronal es la indicada en el esquema de la figura 29.

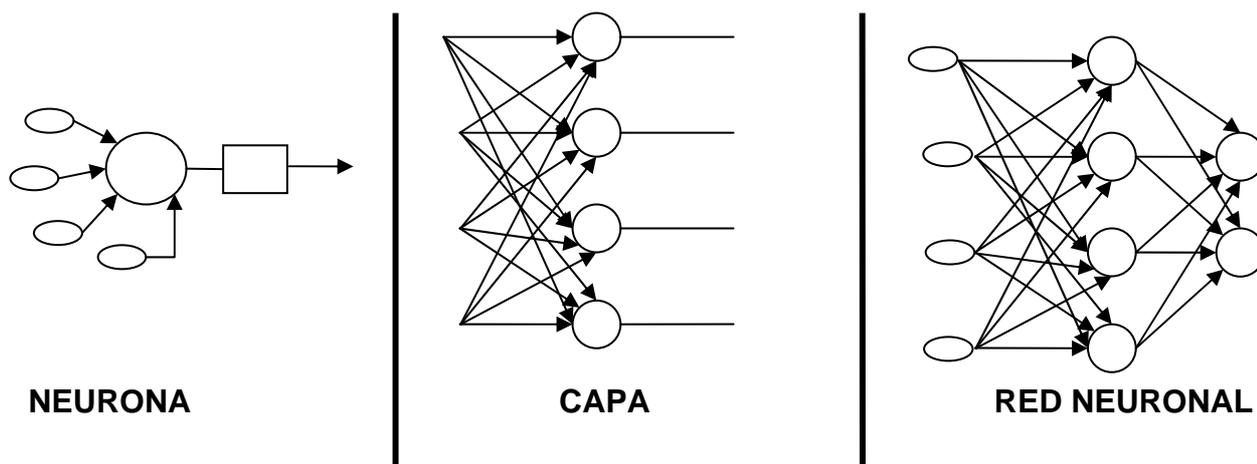


Figura 29. Esquema del modelo organizativo de las neuronas.

Una vez determinada la arquitectura de la red neuronal caben plantearse las dos siguientes preguntas: a) ¿Cómo se establecen las conexiones entre las distintas neuronas artificiales? y b) Una vez configurada toda esta estructura, ¿cómo opera?

La primera pregunta hace referencia a la determinación de los pesos sinápticos de la red; el ajuste de estos pesos se denomina aprendizaje o entrenamiento y se efectúa a partir de una serie de patrones de entrada-salida mediante la minimización de una función de error o coste, que mide la eficacia operativa actual de la red.

Este proceso de aprendizaje es usualmente iterativo, dada la naturaleza no-lineal de las redes artificiales empleadas.

Hay dos tipos de aprendizaje, el supervisado y el no supervisado. En el aprendizaje supervisado, se tutela a la red, mediante patrones constituidos por variables de salida diferentes a las variables de entradas. En el no supervisado se presentan a la red multitud de patrones sin adjuntar ninguna respuesta.

Una vez entrenada la red artificial, el valor de los pesos se “congela” y la configuración de la estructura queda fijada, estando la red neuronal ya dispuesta para tratar la información. Se busca que la red neuronal artificial, realice predicciones razonables ante entradas desconocidas para la misma, es lo que se denomina generalización.

3.3.1.4. Beneficios del empleo de redes neuronales artificiales.

A continuación se van a enumerar algunas de las características fundamentales de las redes neuronales artificiales con objeto de justificar su empleo:

- Las redes neuronales no pretenden determinar el funcional subyacente o generador de datos, cuya relación de causalidad proporciona una determinada salida en el sistema ante una entrada. Únicamente actúan a modo de caja negra proporcionando salidas razonables ante las mismas entradas, ya que éstas cuentan con la experiencia adquirida en la fase de aprendizaje y son capaces de generalizar ante entradas para las cuales no han sido entrenadas.

Esto es muy útil en problemas de regresión complejos, como se verá a continuación, donde no se tiene ningún conocimiento de tal generador de datos, es el caso de numerosos problemas inversos y de clasificación en la ingeniería, ya que no será necesario ni determinar, ni asumir ninguna hipótesis acerca de la función de densidad de probabilidad.

- Una vez entrenadas realizan pronósticos, casi de forma inmediata, ya que implican operaciones muy sencillas entre las neuronas, esto las hace muy apropiadas para su empleo

en problemas donde es necesario realizar una diagnosis en tiempo real, como es el caso que se pretende estudiar en esta tesis.

- Si se utilizan funciones de activación no lineales (en la mayoría de los casos), los pronósticos efectuados por tales redes neuronales serán como es lógico no lineales ya que una red diseñada con interconexiones no lineales, es en si misma no lineal.

La no linealidad es muy importante, especialmente cuando el proceso físico generador de la señal sea no lineal. En el caso de máquinas rotativas con transmisión de potencia mediante engranajes, la fuente de generación de señales vibratorias se debe a un choque o impacto entre las superficies en las que se produce el par de engrane, este impacto, es de naturaleza claramente no lineal.

- Las redes neuronales son sistemas especialmente robustos incluso en las condiciones de funcionamiento más adversas. Si alguna interconexión o alguna neurona resulta dañada, la pérdida de funcionalidad de tal sistema no se ve seriamente afectada. Esto se debe a que el almacenaje de la información se distribuye en toda la red, por consiguiente, el deterioro de una red se da de forma gradual y no de forma catastrófica, como podría ocurrir con un sistema electrónico de diagnosis.

- Las redes neuronales tienen la capacidad de adaptar sus pesos sinápticos a cambios en el entorno en el que han sido entrenadas. En particular una red que ha sido configurada para unas condiciones específicas de operación puede ser rápidamente reentrenada para afrontar pequeños cambios en las condiciones operativas.

3.3.2. APLICACIÓN DE REDES NEURONALES EN PROBLEMAS DE REGRESIÓN

Un problema común en ingeniería consiste en determinar, a partir de la observación de los patrones constituidos por parejas de variables de entrada y salida, un funcional aproximador que establezca la relación de causalidad entre dichas variables. Este proceso, conocido como aproximación funcional en el entorno del análisis numérico, se denomina aprendizaje supervisado en el campo de las redes neuronales artificiales.

El conjunto de patrones de entrenamiento está constituido por pares de variables independientes (variables de entrada) y dependientes (variables de salida).

En esta sección se va a explotar el empleo de redes neuronales artificiales (ANN) como herramienta de “proyección” entre los espacios de variables de entrada y salida.

La idea básica para entender la capacidad aproximadora universal de las redes neuronales consiste en que éstas, si se combinan de forma adecuada, son capaces de construir cualquier función de base localizada en cada punto del espacio de entradas; por tanto, mediante la combinación lineal de dichas funciones base se puede aproximar cualquier hipersuperficie constituida por las variables del espacio de salida ([55] a [58]).

Esta capacidad aproximadora se puede implementar mediante dos topologías o arquitecturas de redes neuronales artificiales, el perceptrón multicapa (MLP), y las redes de función de base radial (RBF).

3.3.2.1. Implementación mediante redes RBF

Las redes de base radial se caracterizan porque están formadas por una única capa oculta (figura 30) y cada neurona de esta capa posee un carácter local, en el sentido de que cada neurona oculta de la red se activa en una región diferente del espacio de patrones de entrada.

Este carácter local viene dado por el uso de las llamadas funciones de base radial, generalmente la función Gaussiana, como funciones de activación.

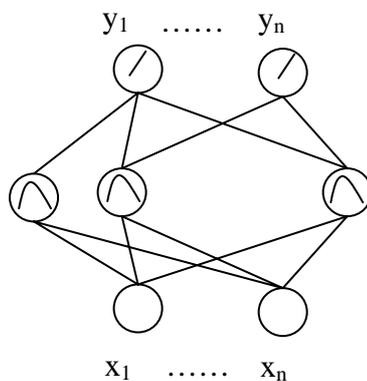


Figura 30. Esquema de red RBF.

El modelo de neurona característico de este tipo de redes se puede representar mediante el diagrama de bloques de la figura 31. La señal de salida se define mediante la siguiente expresión:

$$(3.3.3) \quad y = \phi(a) = \phi(\|dist(\bar{w}, \bar{x})b\|)$$

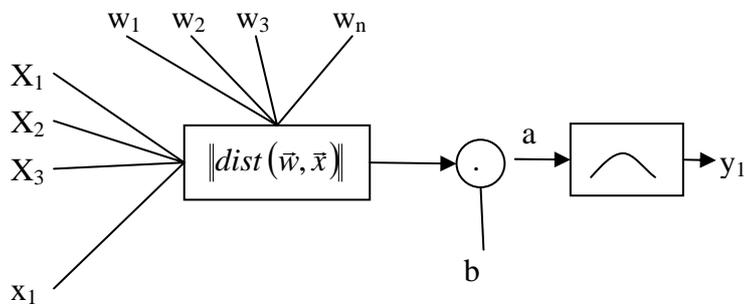


Figura 31. Modelo de neurona empleada en las redes RBF.

El parámetro 'a' de la función de transferencia se determina mediante la distancia entre el vector de entradas y el centro o vector de pesos de la función de base, multiplicada por el parámetro 'b'. La función de transferencia o activación de dicha neurona se obtiene como ya se indicó previamente mediante una función de tipo Gaussiano.

La respuesta de la neurona será tanto mayor cuanto más cercana sea la distancia entre el vector de entradas y el centro de la función base localizada.

En la capa oculta, como se indicó previamente, están embebidos los centros de las funciones base localizadas y la capa de salida está constituida por neuronas con funciones de activación puramente lineales, por tanto se pueden escribir las estimaciones efectuadas por esta red neuronal mediante la siguiente expresión:

$$(3.3.4) \quad y_k(\vec{x}) = \sum_{j=1}^M w_{kj} \phi_j(\vec{x}) + b_k$$

Rescribiendo esta ecuación en notación matricial se obtiene la siguiente relación:

$$(3.3.5) \quad \vec{y}(\vec{x}) = \vec{W} \vec{\phi}$$

En el caso de emplear funciones de base Gaussianas se tiene:

$$(3.3.6) \quad \phi_j(\vec{x}) = \exp\left(-\frac{\|\vec{x} - \vec{w}_j\|^2}{2\sigma_j^2}\right)$$

Los parámetros a ajustar durante el entrenamiento son:

- i) Número de neuronas en la capa oculta.
- ii) Posiciones de los centros de las funciones base de cada neurona (pesos sinápticos de la capa oculta).
- iii) Anchura de las funciones de base 'oj' (que puede ser variable para cada neurona o subconjunto de neuronas, o por el contrario fijo para todas).
- iv) Pesos sinápticos de la capa de salida.

Dentro de las estrategias de entrenamiento de las redes RBF hay dos enfoques claramente diferenciables, el primero adopta la técnica del ajuste de parámetros puramente supervisado (dicha técnica es la empleada por las redes MLP, "Perceptrón multicapa") la cual implica la resolución de un problema de optimización o ajuste no lineal; el segundo enfoque adopta una técnica que algunos califican como híbrida (entre supervisada y no supervisada) que se basa en la diferencia de los papeles que juegan las neuronas de la capa oculta y de salida de la red RBF y que trata de resolver el problema de ajuste en dos etapas, en la primera de ellas se establece el número de centros de neuronas así como sus posiciones (pesos sinápticos de la primera capa) basándose únicamente en los patrones de las variables del espacio de entradas (técnica no supervisada); una vez completada esta fase únicamente queda pendiente el ajuste de los pesos de la capa de salida, el cual, como se verá implica la resolución de un problema de optimización lineal (entrenamiento supervisado).

A grandes rasgos, se pueden intuir los beneficios obtenidos siguiendo la estrategia híbrida ya que no se cuenta con las desventajas del empleo de técnicas de optimización no lineal que implican problemas de velocidad de convergencia, problemas de estancamiento en mínimos locales, etc.

A continuación se va a describir brevemente la filosofía a seguir en cada etapa de la estrategia híbrida indicada anteriormente:

1) Establecimiento del número de neuronas de la capa oculta (centros de Kernel), ajuste de pesos sinápticos de dicha capa y anchura de las funciones base.

Existen numerosos enfoques y técnicas para completar esta etapa, de entre las cuales cabe destacar la técnica OLS (“Orthogonal Least Squares”). Básicamente, la idea de este procedimiento consiste en incrementar secuencialmente el número de funciones base o neuronas, cada una centrada en cada punto de la muestra; esto se realiza por medio de la construcción de un conjunto de vectores ortogonales expandidos por cada dato de los patrones de entrenamiento,

Esto posibilita localizar la siguiente función de base a añadir porque el algoritmo minimiza el error cuadrático medio residual.

(Para obtener más información sobre este algoritmo consultar las referencias [44] y [45].)

Por otra parte, la elección de la anchura de las funciones base, σ , se suele realizar de una forma heurística, aunque se suele elegir igual para todas las neuronas. Este parámetro dictamina el grado de suavidad de las estimaciones efectuadas por la red neuronal. Si este parámetro es lo suficientemente pequeño como para que no exista solape entre las funciones base, entonces las predicciones efectuadas por la red serán casi discontinuas; si, por el contrario, la anchura es muy grande, la predicción efectuada por la red tenderá a un hiperplano, es decir, tenderá a dar una respuesta casi lineal y habrá que llegar por tanto a una solución de compromiso.

2) Ajuste de los pesos de la capa de salida.

Suponiendo completada la fase anterior (elección del número de neuronas de la capa oculta y sus pesos sinápticos) si se minimiza el error cuadrático medio de la relación (3.3.7):

$$(3.3.7) \quad E = \frac{1}{2} \sum_m \{ \bar{y}(\bar{x}^n) - \bar{t}^n \}^2$$

se habrán ajustado los pesos sinápticos de las neuronas de la capa de salida, siendo 't' el vector de salida de cada patrón enésimo para un vector de entradas 'x' e 'y' la estimación efectuada con la red neuronal artificial.

Por tanto, en virtud de las relaciones (3.3.5) y (3.3.7) se obtienen los pesos óptimos que vienen dados por la siguiente ecuación:

$$(3.3.8) \quad \bar{\phi}^T \bar{\phi} \bar{W}^T = \bar{\phi}^T \bar{T}$$

Por último, la solución formal a los pesos vendrá dada por:

$$(3.3.9) \quad \bar{W}^T = \bar{\phi}^+ \bar{T}$$

donde la notación ' $\bar{\phi}^+$ ', indica la pseudoinversa de ' $\bar{\phi}$ '; en la práctica las ecuaciones (3.3.9) se resuelven utilizando una descomposición de valores singulares para evitar problemas de mal condicionamiento de la matriz ' $\bar{\phi}$ '. Por tanto, la segunda capa de pesos se puede encontrar de forma muy rápida si se utilizan técnicas de inversión de matrices apropiadas.

3.3.2.2. Implementación mediante redes MLP

Para este tipo de red el modelo de neurona viene dado mediante el diagrama de bloques de la figura 32. Según esta figura, la salida de las capas ocultas de la red MLP viene dada por la siguiente ecuación:

$$(3.3.10) \quad y = \phi(\bar{w} \bar{x} + b)$$

siendo la función de activación ' ϕ ' una función de tipo sigmoideo.

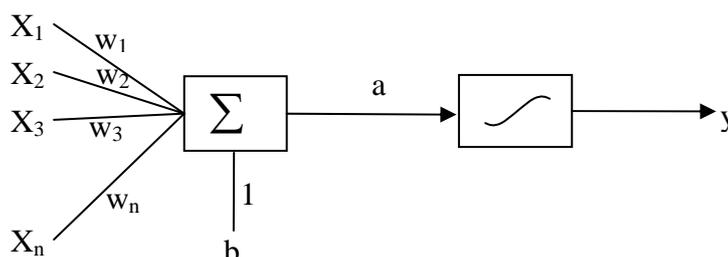


Figura 32. Modelo de neurona MLP.

La arquitectura de la red se organiza mediante una serie de capas ocultas con funciones de transferencia sigmoideas (el número de capas viene determinado por la complejidad del modelo a representar) y una capa de salida con neuronas cuya función de transferencia es puramente lineal (figura 33).

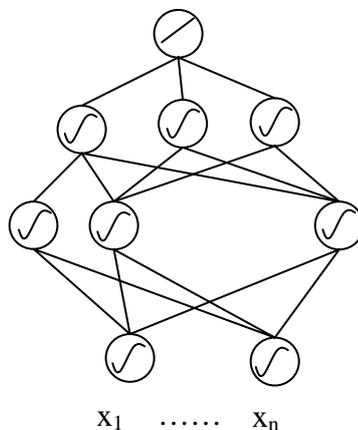


Figura 33. Red neuronal MLP.

Como la función de activación de la neurona da lugar a una respuesta con forma de superficie sigmoidea (figura 34.a), es posible combinar dos neuronas con orientación ortogonal y desplazadas una cierta distancia para dar lugar a una superficie en forma de cresta (figura 34.b); combinando más neuronas con orientaciones ortogonales se puede formar una superficie de respuesta con forma de montículo localizado rodeado de otra serie de crestas que tienden a extenderse hasta el infinito (figura 34.c). Para aislar el montículo central (figura 34.d) es necesario añadir otra capa más a la estructura que actúa a modo de filtro de las respuestas de la capa anterior (se filtra el montículo central y se atenúan el resto de crestas).

Por tanto, en la red MLP se puede sintetizar cualquier función de base localizada con un mínimo de tres capas, es decir, se pueden construir funciones de tipo Kernel y es viable la resolución del problema de regresión que se planteaba al comienzo del capítulo.

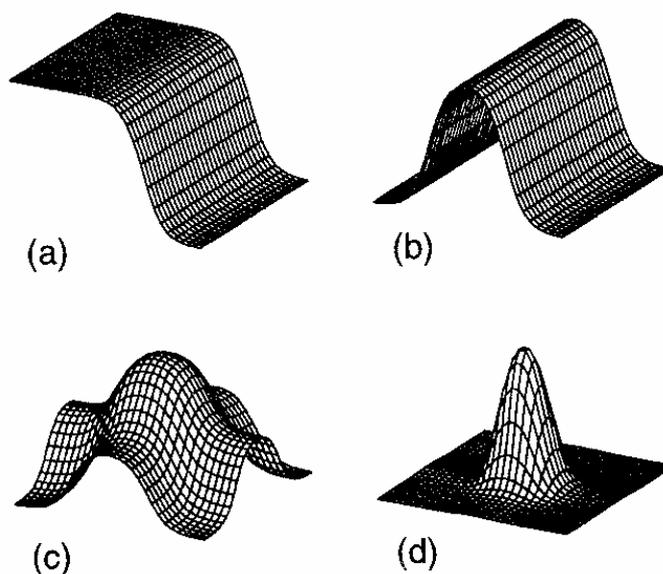


Figura 34. Combinación de capas de neuronas MLP.

El entrenamiento de este tipo de red se suele efectuar con cualquier algoritmo de optimización no lineal como el de Newton, Cuasi-Newton, Levenberg-Marquadt, etc. ([46] y [47]).

3.3.2.3. Comparativa de topologías MLP y RBF

A pesar de la diferencia en la estructura de ambos tipos de redes, en la actualidad y para numerosas aplicaciones, estas dos topologías se emplean habitualmente para problemas de proyección o regresión no lineal, por tanto parece conveniente plantear sus principales diferencias a fin de justificar la elección de uno u otro tipo.

A grandes rasgos, se pueden destacar las siguientes diferencias [48]:

1. Los parámetros libres o parámetros a ajustar en el caso de las redes MLP se obtienen mediante una estrategia global de optimización no lineal, mientras que en el caso de las redes RBF se obtienen en dos fases, una de ellas emplea una técnica no supervisada mientras la otra implica la resolución de un problema lineal

Se puede por tanto concluir que al aplicar redes MLP en la resolución de problemas de optimización no lineal con superficie de respuesta muy plana se pueden tener problemas de estancamiento en mínimos locales y dificultades con la velocidad de convergencia. En este caso sería preferible el empleo de redes RBF.

2. La estructura de la red RBF implica dos capas (figura 31) mientras que la estructura MLP ofrece un complejo patrón de conectividad con más de una capa. A pesar de este hecho, algunos autores consideran que el número de neuronas requerido en un problema de regresión no lineal, en el caso de las redes RBF es muy superior al de las empleadas para el mismo problema con una topología MLP [49], dado que el número de neuronas en el caso RBF es proporcional al espacio constituido por las variables de entrada, lo cual supone una mayor eficiencia de las redes MLP en lo referente a tiempos de computación (estando la red ya entrenada), además supone un ahorro si se piensa en su implementación directa mediante soporte electrónico.

3.3.3. APLICACIÓN DE REDES NEURONALES EN LA DETECCIÓN NOVAL DE FALLOS EN MAQUINARIA ROTATIVA

La detección noval consiste en la identificación de nuevos patrones sobre una señal o señales que un sistema de aprendizaje (en este caso una red neuronal) no ha conocido durante la fase de entrenamiento.

Puede entenderse el problema de la detección noval como un problema de clasificación binario que distingue las señales procedentes de la máquina a analizar entre dos estados de funcionamiento, el estado normal y el anómalo o noval.

En esta tesis la detección noval se realizará sobre vectores constituidos no por la señal vibratoria original, sino sobre los coeficientes obtenidos de su transformación DWT.

En este caso y a diferencia de otros “detectores noval” no se ha empleado un único parámetro para determinar las condiciones de funcionamiento degradadas en la maquinaria, sino que se emplea un vector de índices noval que aportará información de anormalidad en lo referente a la localización tiempo-frecuencia del “daño”, lo cual servirá en algunos casos para localizar el componente defectuoso en la máquina rotativa.

El tipo de red neuronal que se va a emplear para tales propósitos es una red no supervisada. La ventaja que presenta este tipo de redes es que, a priori, no necesitan clasificar las distintas condiciones de operación de la maquinaria, ya que esto implicaría los siguientes perjuicios en el sistema de detección:

1. Conocer de antemano todas las condiciones de operación de la máquina.
2. Realizar el entrenamiento separando las señales obtenidas según las distintas condiciones operativas.
- 3-Se necesitaría una red para cada condición de operación.

Las redes no supervisadas, sin embargo, no necesitan conocer las distintas condiciones en las que funciona la máquina, de hecho, llegan a aprender cuáles son dichas condiciones; no obstante, conviene indicar que durante el entrenamiento se consideren todos los puntos de operación de la maquinaria, de no ser así, un punto que no estuviera asociado a ningún estado de degradación podría ser detectado como tal, ya que sería diagnosticado como un estado noval a partir de la definición que hemos expuesto al comienzo de esta sección.

En el estudio llevado a cabo, se han utilizado para la implementación de los detectores novales redes no supervisadas autoasociativas o de cuello de botella, dada su simplicidad de implementación tanto a nivel programación como en soporte electrónico, aunque cualquier otro tipo de red no supervisada, como por ejemplo las redes SOM (mapas autoorganizados), pudiera ser igualmente efectivo.

3.3.3.1. Implementación con redes autoasociativas

Este tipo de red corresponde a un tipo de red MLP no supervisada configurada para reproducir en la salida los patrones de datos más relevantes o comunes del espacio de las variables de entrada. Para ello, estas redes adoptan como variables objetivo las mismas variables de entrada, es decir tratan de proyectar los vectores de entrada sobre sí mismos.

Antes de proseguir detallando las posibilidades de las redes autoasociativas, parece interesante ver cómo se aborda esta tarea de reconocimiento de rasgos nuevamente desde la perspectiva estadística, la cual se apoya en una técnica llamada análisis de componentes principales (PCA o transformación Karhunen-Loève); posteriormente se indicará su implementación directa mediante las redes autoasociativas [50].

Con el análisis de componentes principales se reduce el número de rasgos necesarios para una representación efectiva del espacio de entrada; esto matemáticamente equivale a buscar una proyección del espacio de entrada en un subespacio de menor dimensión de tal forma que, cuando se realice la transformación inversa, el vector sintético generado de esta manera sea lo más semejante posible al vector original. Por tanto, en lugar de manejar los datos del espacio original de entrada, se utilizan sus proyecciones en el espacio reducido (rasgos principales) ya que previamente se ha asegurado que la pérdida de información derivada de esta compresión sea lo más pequeña posible.

Para realizar esta compresión, se determinan primeramente los autovectores y autovalores de la matriz de covarianzas de las muestras constituidas por los vectores del espacio de entrada, para después proyectar los vectores de entrada, ortogonalmente, en el subespacio constituido por los autovalores dominantes. Esta operación se denomina codificación.

Se puede indicar que si se expande de nuevo el vector proyectado al espacio de entrada, mediante una transformación inversa (decodificación), su diferencia respecto al vector original es mínima.

La implementación más simple de la técnica PCA mediante redes neuronales podría realizarse de la siguiente manera: a) La etapa de codificación se implementaría utilizando una capa oculta con un número de neuronas inferior al del espacio de entrada; esto equivaldría a la proyección del vector de entrada (proyección lineal o no lineal dependiendo de la función de activación de la neurona) en un espacio de menor dimensión. b) La decodificación se llevaría a cabo previo empleo de una capa de salida con el mismo número de neuronas (con funciones de activación lineales) que el vector de entrada.

La estructura de este tipo de red se muestra en la figura 35 a.

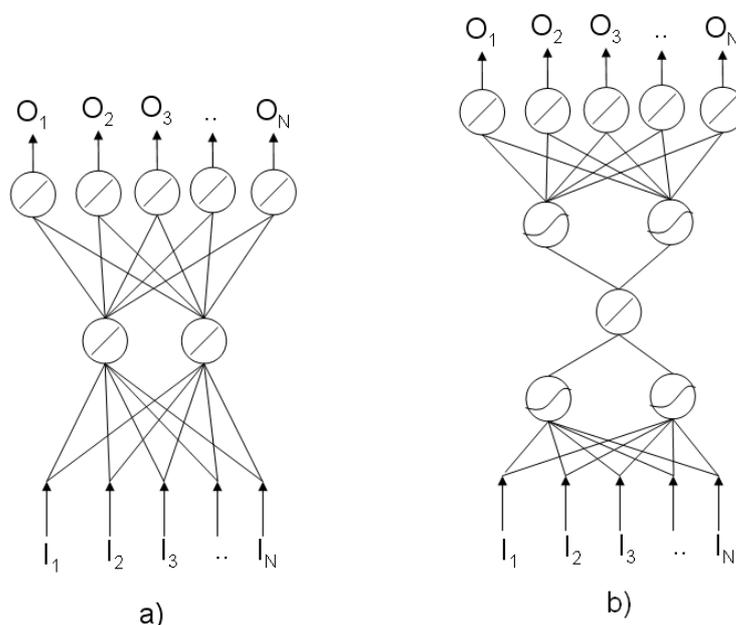


Figura 35. Redes neuronales autoasociativas.

Si las funciones de activación en la capa oculta son lineales se puede demostrar que el ajuste de los pesos durante la fase de entrenamiento tiene un mínimo global; la consecución de este mínimo equivale a la proyección del espacio de entrada en un subespacio de menor dimensión utilizando la técnica PCA lineal.

Este resultado no es sorprendente puesto que tanto el análisis de componentes principales como la red neuronal, realizan una reducción de la dimensión del espacio de entrada previa minimización de la misma función objetivo [51].

De cualquier forma, cualquier reducción del espacio mediante proyecciones lineales tiene numerosas limitaciones; por tanto, cabría pensar que si se utilizaran funciones de activación sigmoideas como en una red convencional MLP (modificando el esquema (a) de la figura 35 con funciones de activación tangente-sigmoideas) se superarían dichas limitaciones. Sin

embargo, se ha demostrado [51] que tal red no produce las expectativas depositadas en ella y los resultados no son muy diferentes a los obtenidos mediante PCA lineal.

Una mejora sustancial sí se consigue cuando se añade más de una capa oculta, empleando una estructura de tipo “cuello de botella” tal como se muestra en la figura 35 b [52] De cara a conseguir los objetivos de esta tesis, se va a emplear esta configuración, la cual tiene como parámetros de ajuste el número de neuronas en las capas de codificación y decodificación así como los distintos pesos sinápticos.

Los pesos se ajustan durante la fase de entrenamiento con un algoritmo llamado de propagación de resiliencia [53]. Este es un método bastante efectivo y rápido, especialmente cuando los tamaños de las redes son grandes (superan las veinte neuronas en las capas de codificación y decodificación) ya que en ese caso la velocidad de convergencia puede llegar a ser muy lenta por los problemas mencionados cuando se describieron los algoritmos de entrenamiento de las redes MLP.

Dado que las funciones de activación sigmoideas se caracterizan porque sus pendientes tienden a cero ante entradas grandes, el método del descenso del gradiente no es adecuado durante la fase de entrenamiento ya que, al ser los gradientes muy pequeños, existe un claro problema de estancamiento.

Para eliminar este efecto perjudicial en las magnitudes de las derivadas parciales, el algoritmo de propagación de resiliencia únicamente emplea el signo de estas derivadas para determinar la dirección de la actualización de los pesos, la magnitud de la derivada no tiene ningún efecto en la actualización de dichos pesos. Para tal propósito se emplea un incremento por un factor constante si el signo de la derivada en dos iteraciones sucesivas no cambia, si por el contrario ocurre un cambio de signo se emplea un decremento de la misma cuantía.

Como en este caso no se almacenan los valores de las derivadas sino que únicamente se evalúa el signo de las mismas, se acelera el proceso de entrenamiento de forma notable.

Por último, para el ajuste del número de neuronas en las capas de codificación y decodificación se opta por una técnica de crecimiento, consistente en incrementar progresivamente el número de neuronas a partir de un valor pequeño inicial en función de los errores cometidos con cada red, hasta alcanzar aproximadamente un 70% del número de neuronas de las capas de entrada y salida. Finalmente se selecciona la red que haya dado el menor de los errores totales.

Los errores que se adoptan para esta calibración son el error de sesgo o truncamiento y el error de dispersión y se definen de la forma siguiente:

$$(3.3.11) \quad \begin{aligned} \text{Error Total} &= \text{Error Truncamiento} + \text{Error Dispersión} \\ \text{Error Truncamiento} &= \left[E \left[\bar{y}^n - \bar{x}^n \right] \right]^2 \\ \text{Error Dispersión} &= E \left[\left(\bar{y}^n \right) - E \left[\bar{y}^n \right] \right]^2 \end{aligned}$$

En donde 'E' representa el operador esperanza matemática.

Mientras que el error de dispersión proporciona una medida de la dispersión de las estimaciones pertenecientes a una misma clase, ya que se busca una uniformidad en la predicción dentro de cada clase, el error de truncamiento especifica el error cometido en la "autoproyección".

A mayor complejidad en el modelo menor error de truncamiento y mayor error de dispersión (las predicciones tienden a ajustarse al ruido de fondo), [54].

Una vez configurada la red autoasociativa, se debe plantear la manera de definir los índices novalés.

Tal como se indicó al inicio del capítulo, el objetivo es definir un sistema automático de reconocimiento de anomalías en una serie de muestras; en el caso particular de esta tesis, las muestras consisten en los coeficientes proporcionados por la transformación mediante ondículas de un conjunto de series temporales.

Los índices noales se establecen de acuerdo con los siguientes pasos:

1. Se determina el vector residual a partir de la diferencia entre el vector de entrada, constituido por el vector de coeficientes de la transformación con ondículas, y el vector de salida de la red autoasociativa.

$$(3.3.12) \quad \lambda^k(y) = |y_o^k - y_i^k|$$

En esta expresión los superíndices indican la componente de cada vector mientras los subíndices 'o' e 'i' se refieren al vector de salidas y entradas de la red neuronal, respectivamente.

2. Se establece el vector residual promedio y se comprueba la distribución de los distintos residuos con el residuo promedio.

En la mayoría de los casos, la distribución de estos residuos o "ruido" sigue una distribución normal Gaussiana, en especial cuando hay numerosas fuentes de contribución a la aleatoriedad (fuentes de ruido); de hecho, se puede demostrar (utilizando el teorema central del límite) que la suma de estas contribuciones sigue una distribución de este tipo.

3. En función de la distribución de los residuos, se determinan unos valores umbrales. En el caso de que dicha distribución pueda ser modelada mediante una distribución Gaussiana (caso que ocupa esta tesis) se establecerá el siguiente criterio umbral:

$$(3.3.13) \quad \delta_{\lambda}^k = \bar{\lambda}^k + 4\sigma_{\lambda}^k$$

siendo 'k' el k-ésimo componente del vector residual definido en el punto (1), ' $\bar{\lambda}$ ' el valor promedio de tal componente y ' σ ' la desviación estándar del mismo.

La ventaja de tratar con un vector de índices noales es que la condición de alarma se establece sobre cada coeficiente de la transformación, por tanto se están señalando las perturbaciones que se están produciendo en la maquinaria en la escala tiempo-frecuencia.

A continuación se va a realizar un ejemplo sencillo en el cual se va a aplicar este sistema de detección noval mediante redes de cuello de botella.

En este ejemplo se busca una red que sea capaz de clasificar unas series temporales, bien como normales o bien como noales y que, además, especifique en qué instantes de tiempo se producen dichas condiciones noales.

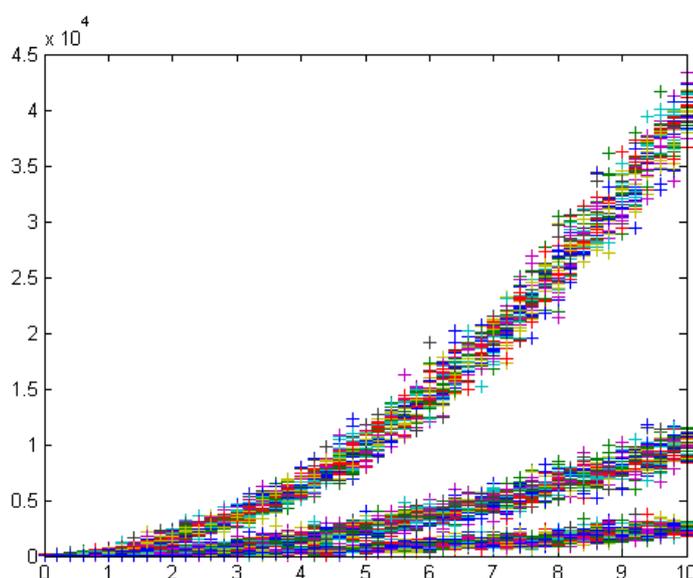


Figura 36. Señales temporales del problema de detección noval.

En la figura 36 se ilustran cuarenta muestras (representadas por distintos colores) correspondientes a tres series temporales derivadas de tres condiciones de operación “sanas” que pudieran darse en una máquina. Las tres condiciones operativas son claramente distinguibles por tres ramas parabólicas de distinta pendiente. Se dispone de un total de 120 señales.

Por tratarse de series temporales, el eje de abscisas representa el tiempo y en ordenadas se muestra la amplitud de la señal recogida en cada instante de tiempo.

Cada muestra contiene 52 registros correspondientes a 52 distintos instantes de tiempo. Por tanto cada muestra está constituida por un vector de dimensión 52.

De las 120 muestras, 100 se emplean para entrenar a una red de cuello de botella, mientras que las 20 restantes junto con otras 20 muestras correspondientes a condiciones operativas anómalas se emplearán como instrumento de validación de la red. Para representar las condiciones anómalas se emplean también series temporales parabólicas pero con una pendiente intermedia respecto a las observadas anteriormente (figura 37).

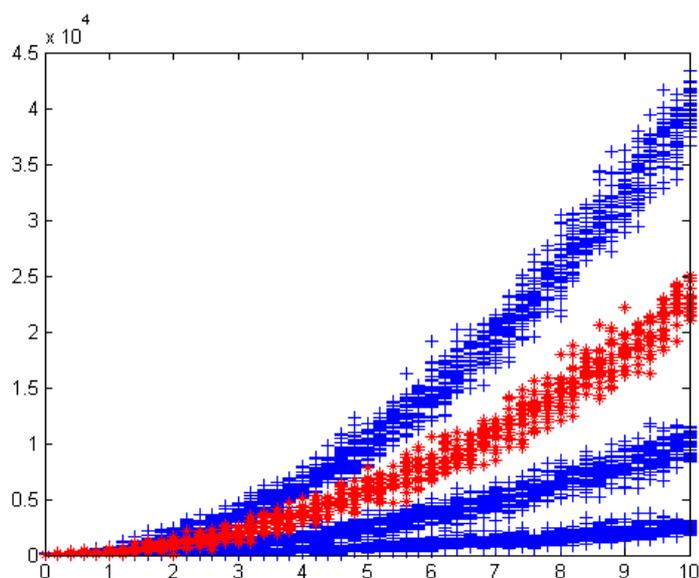


Figura 37. Introducción de series temporales anómalas.

En la figura 37 se muestran en rojo las series correspondientes a unas hipotéticas condiciones anómalas en la maquinaria.

En primer lugar, se configura la red autoasociativa ajustando sus pesos sinápticos mediante el algoritmo RP, partiendo de una red de 10 neuronas en la etapa de codificación y decodificación, y utilizando el algoritmo de crecimiento comentado anteriormente hasta llegar al 70% del número de neuronas en las capas de entrada y salida.

En función de este algoritmo se determina una topología óptima según la ecuación (3.3.26) con 25 neuronas en las etapas de codificación y decodificación.

A continuación se establecen los valores umbrales en función de los residuos.

En las figuras 38.a y 38.b se representan los residuos de las muestras no utilizadas durante el entrenamiento y de las muestras correspondientes a las condiciones anómalas, respectivamente. En ambos casos se muestran los valores umbrales de los residuos mediante dos líneas discontinuas de color rojo.

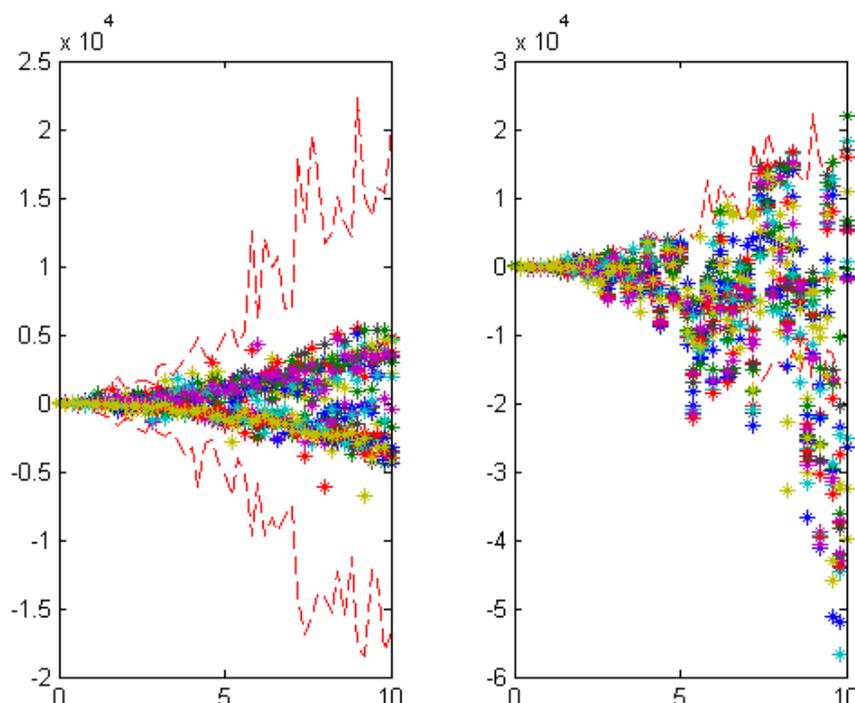


Figura 38. Residuos de señales "sanas" (a) y de las señales con anomalías (b).

Se puede apreciar que la red no proporciona falsas alarmas porque no se superan los valores umbrales. Sin embargo se han detectado las condiciones anómalas en todos los casos (100% de detección).

Se puede apreciar que esto ocurre a partir de la abscisa 2.5 ya que la red confunde las condiciones anómalas como normales hasta este punto temporal. Esto se debe a que si se observan las series correspondientes a las condiciones degradadas (figura 37), éstas entran dentro de la dispersión observada para las condiciones normales o “sanas”.

Por tanto, se puede concluir que esta red artificial generaliza bien porque clasifica bien patrones que no ha conocido durante su entrenamiento.

Hay que indicar que si se hubiera tratado de resolver este problema empleando técnicas como la del TSA (promediado síncrono), se tendrían que realizar promedios sobre cada una de las ramas de las parábolas correspondientes a cada condición de operación y establecer umbrales sobre cada una de ellas; la ventaja del método aquí presentado (utilización de redes autoasociativas) es que emplea una única red para tratar todas las hipotéticas condiciones de operación en una máquina.

3.4. DISEÑO DEL SISTEMA DE DETECCIÓN.

Una vez establecidas las bases para el pretratamiento de las señales y las herramientas de clasificación y cuantificación automáticas, queda ver cómo se articulan y cómo tratan la información proporcionada por las señales vibratorias con objeto de establecer un sistema de diagnóstico y alarma; de hecho, en esta sección se pretende dar respuesta a las siguientes dos

cuestiones fundamentales: a) ¿Qué tipo de sistema de diagnóstico se quiere desarrollar? b) ¿Cómo se puede desarrollar dicho sistema?

3.4.1. CLASIFICACIÓN DE RYTTER DE LOS SISTEMAS DE DIAGNOSIS.

En la tesis de Rytter [59], se establecen cuatro etapas en la determinación e identificación de daño en estructuras que se van a extrapolar aquí para la determinación del daño en máquinas rotativas. Estas cuatro etapas son las siguientes:

1-Detección: el método debe ser capaz de identificar de forma cualitativa la existencia o no de daño en el sistema.

2-Localización: El método debe proporcionar información referente a la posible localización espacial del daño.

3-Evaluación: Se refiere a la estimación y cuantificación del daño generado.

4-Prognosis: Capacidad de estimación de la vida residual del sistema.

El cumplimiento de la primera etapa se puede llevar a cabo mediante la medición de alguna propiedad (por ejemplo alguna propiedad mecánica) del sistema en condiciones operativas “normales”, con el objetivo de establecer un criterio de alarma o aviso que indique cuando la medida obtenida en cualquier momento corresponde a un sistema defectuoso; en resumidas cuentas, en esta fase únicamente se hace necesario establecer la medición de algún rasgo sensible a la degradación del sistema.

Para la segunda y la tercera etapa se necesita contar bien con modelos matemáticos que representen y simulen el funcionamiento del sistema en condiciones sanas o degradadas o bien con mediciones y ensayos sobre prototipos reales tanto en condiciones normales como anómalas.

La última etapa precisa de un modelo matemático capaz de predecir la evolución del modo de fallo en el prototipo con objetivo de establecer la vida residual del mismo [60].

3.4.2. CARACTERÍSTICAS DEL SISTEMA DE DIAGNOSIS.

3.4.2.1-Objetivos y rangos de aplicación

En este apartado se van a especificar las características deseadas del sistema de detección; para ello, por un lado, se tendrá en cuenta la clasificación de Rytter y, por otra parte, se tendrá en consideración la velocidad de generación de decisiones en el sistema de diagnóstico; así se distinguirán los juicios realizados en tiempo real (los realizados automáticamente al recibir la señal procedente del registro correspondiente a cada vuelta de la máquina) de los realizados a posteriori.

Se pretende conseguir una detección, localización y cuantificación automática en tiempo real, siempre y cuando el sistema sea lo suficientemente sensible para apreciar que algo no está funcionando correctamente; esto se debe llevar a cabo para las distintas condiciones de operación de la máquina rotativa.

Además, se persigue determinar qué componente rotativo está fallando en la máquina; esta última tarea se realizará posteriormente a la fase de alarma y sus pronósticos no se darán en tiempo real.

Hay que indicar que la capacidad de localización será selectiva cuando se empleen bandas de frecuencias para determinar el componente o componentes dañados ya que puede ocurrir que defectos diferentes en distintos componentes tengan armónicos que se solapen y sea por consiguiente imposible distinguir su procedencia.

Una diagnosis de cuarto nivel requiere realizar inferencias que se escapan de los objetivos de esta tesis ya que son específicas de la máquina a evaluar.

En esta tesis se pretende diseñar un procedimiento de detección que sea generalizable y extrapolable para cualquier máquina rotativa; no obstante, para cada caso concreto sería de verdadero interés el confeccionar un plan de mantenimiento a partir de la información proporcionada por el sistema de diagnosis desarrollado en esta tesis en combinación con otros aspectos (como pudieran ser el coste económico de la parada de la máquina, el nivel o severidad de riesgo humano al no efectuar la parada, etc.) para minimizar los efectos adicionales originados por las condiciones de operación en las condiciones de funcionamiento del sistema degradado.

En resumidas cuentas, el sistema de detección debe ser capaz de detectar, localizar y cuantificar una anomalía en la máquina rotativa de manera automática y en tiempo real y, dependiendo del tipo de máquina, debe también predecir el componente que está fallando.

3.4.2.2. Diseño.

Anteriormente, cuando se indicaban las características deseadas del sistema de diagnosis, se destacaba su capacidad de reconocimiento automático de patrones. Watanabe [64] define el patrón como lo opuesto al caos, como una entidad definida de forma muy somera a la que se le podría designar con un determinado nombre o etiqueta. Ejemplos de patrones pudieran ser por ejemplo una huella dactilar, una letra, una cara, etc.

Existen dos alternativas en las tareas de clasificación; éstas son, respectivamente, la clasificación supervisada y la no supervisada.

En el caso supervisado, cada patrón es identificado como miembro de una clase predefinida mientras que en el caso no supervisado, cada patrón se asocia a una clase inédita o novel.

A grandes rasgos, el diseño de cualquier sistema de reconocimiento de patrones implica esencialmente tres aspectos fundamentales:

1) Adquisición y preprocesamiento de los datos, 2) representación del conjunto de datos y, por último, 3) toma de decisiones.

La clave del éxito del sistema de reconocimiento de patrones radica en que éste sea capaz de crear una separación entre las distintas clases muy superior al grado de aleatoriedad interna de cada clase; esto sin duda ofrecerá una representación de los patrones compacta y conducirá a un estrategia de decisión simple y eficaz.

Los tres puntos anteriormente citados establecen las pautas para diseñar cualquier sistema de diagnóstico. De hecho, si se examinan los métodos clásicos de detección (ver capítulo 2) se puede comprobar que se sigue esta filosofía.

Así, por ejemplo, las técnicas temporales estadísticas como por ejemplo las basadas en el TSA (promediado síncrono) realizan las siguientes tareas (figura 39):

a) Adquisición de señales vibratorias en puntos sensibles de la carcasa de la maquinaria

b) Mediante un pretratamiento de las señales realizando promedios por vuelta de la maquina se obtiene una señal prototipo.

c) A cada señal a analizar, se le resta la señal prototipo generándose un residuo que, en teoría, debiera contener el ruido más alguna anomalía (si es que la hubiera).

d) Por último, se establecen unos umbrales en función de la densidad de probabilidad observada, bien de la magnitud de la señal o bien de su fase y a partir de estos umbrales se establece un criterio de decisión para determinar si efectivamente se está detectando una anomalía en la señal originaria.

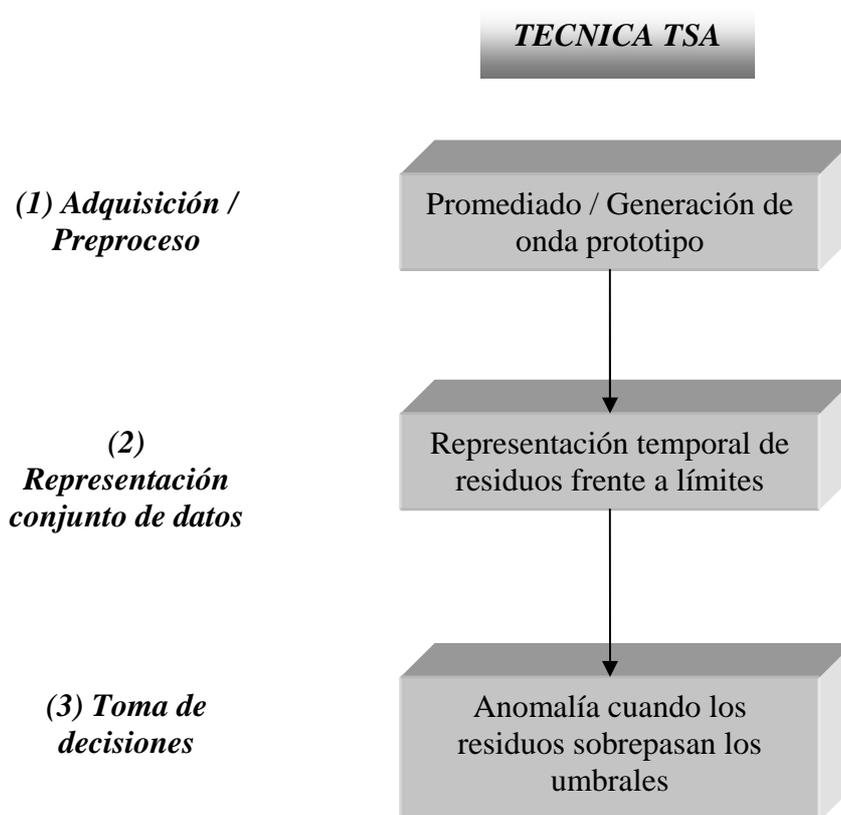


Figura 39. Técnica TSA.

En esta tesis se va a seguir un enfoque totalmente análogo al clásico, adaptado a las técnicas y herramientas comentadas aquí lo que, sin duda, le conferirá unas posibilidades superiores a las proporcionadas por los métodos convencionales, tal y como se demostrará a lo largo de esta sección.

3.4.2.3. Estrategias planteadas.

La estrategia seguida para diseñar el sistema de diagnosis se basa en el principio de “divide y vencerás”. Según eso, el sistema se va a dividir en dos módulos de funcionamiento jerárquico; esta jerarquía implicará que el segundo módulo se active condicionado a las diagnosis efectuadas en el primer módulo.

Así pues, el primer módulo del sistema de detección, el que le confiere capacidad predictiva de nivel 2 (según la escala Rytter) en tiempo real, sigue las siguientes etapas (figura 40):

Pretratamiento de la señal mediante su transformación con ondículas discretas (DWT); esto permite representar la información vibratoria en el dominio tiempo-frecuencia.

Entrenamiento y configuración de una red neuronal autoasociativa a partir de los vectores generados con los coeficientes de la transformación DWT.

Generación de los residuos sustrayendo a cada vector su salida de la red neuronal.

Obtención de vectores con coeficientes umbrales, suponiendo que el ruido “vibratorio” o residuos obtenidos en el punto anterior son de origen Gaussiano.

Toma de decisiones cuando los vectores residuo superen en alguno de sus coeficientes los valores umbrales.

Por último, análisis de los coeficientes afectados y su influencia en el plano tiempo-frecuencia como método de localización de daño.

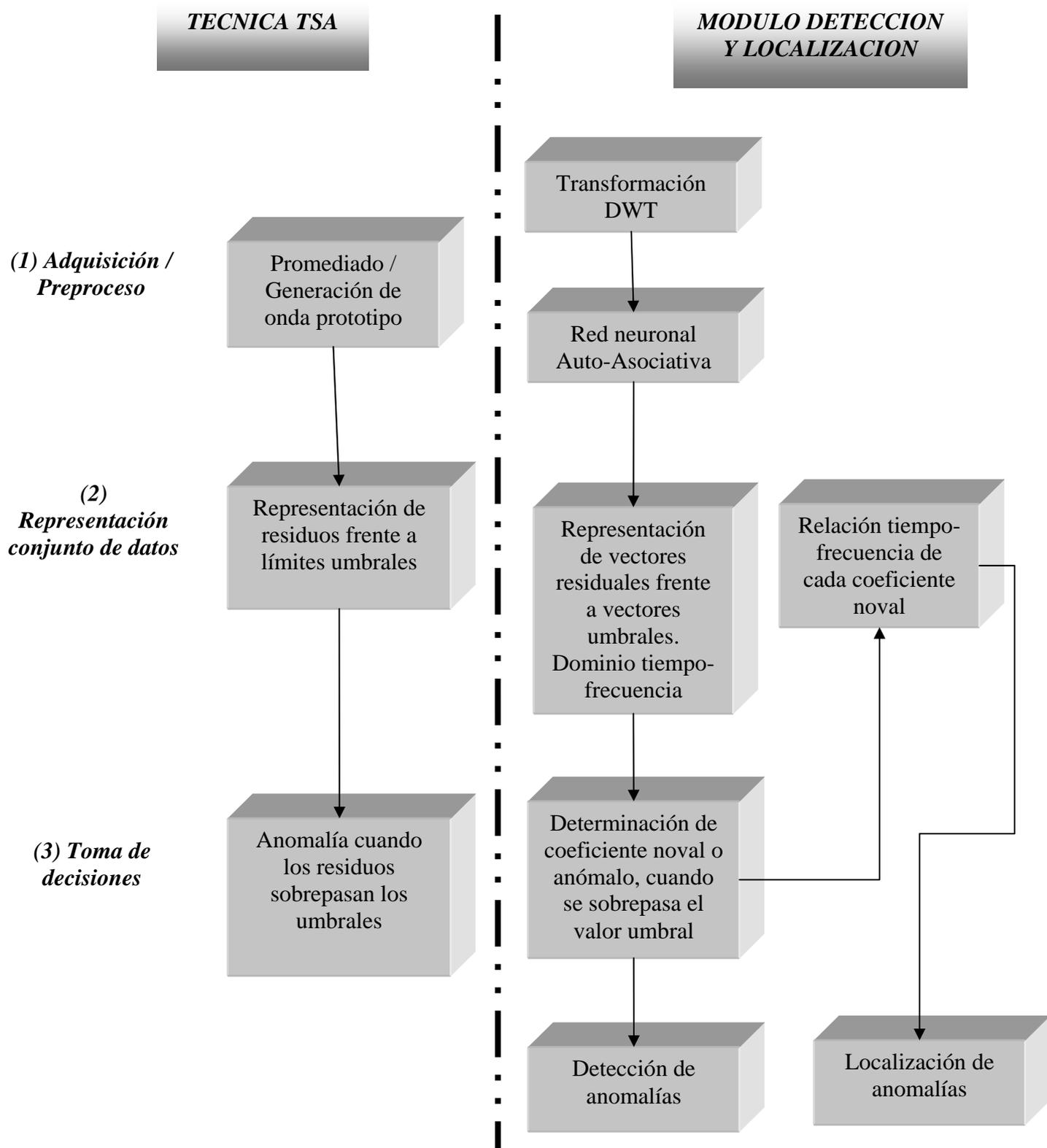


Figura 40. Diagrama de bloques del módulo de detección y localización.

El segundo módulo del sistema de diagnóstico, que sólo se activaría en el caso de que el módulo anterior registrase alguna perturbación, se llevará a cabo en las siguientes etapas (figura 41):

Realización de ensayos sobre distintos prototipos dañados con el mismo tipo de fallo y distinto grado de severidad. Si se dispusiera de un modelo numérico tanto del sistema “sano” como del degradado, esta tarea se realizaría llevando a cabo simulaciones con distintos niveles de degradación.

Adquisición y pretratamiento de las señales vibratorias registradas mediante transformación DWT.

Extracción de rasgos sobre los coeficientes DWT afectados por el fallo.

Entrenamiento de una red supervisada tipo RBF o MLP que ligue algunas características de los coeficientes de los distintos niveles con el nivel de degradación. Previamente se debe llevar a cabo un estudio de sensibilidad de dichos coeficientes, con el objetivo de extraer parámetros sensibles al grado de degradación

Realización de predicciones en función de los parámetros anteriores siempre y cuando se haya detectado daño en la primera etapa.

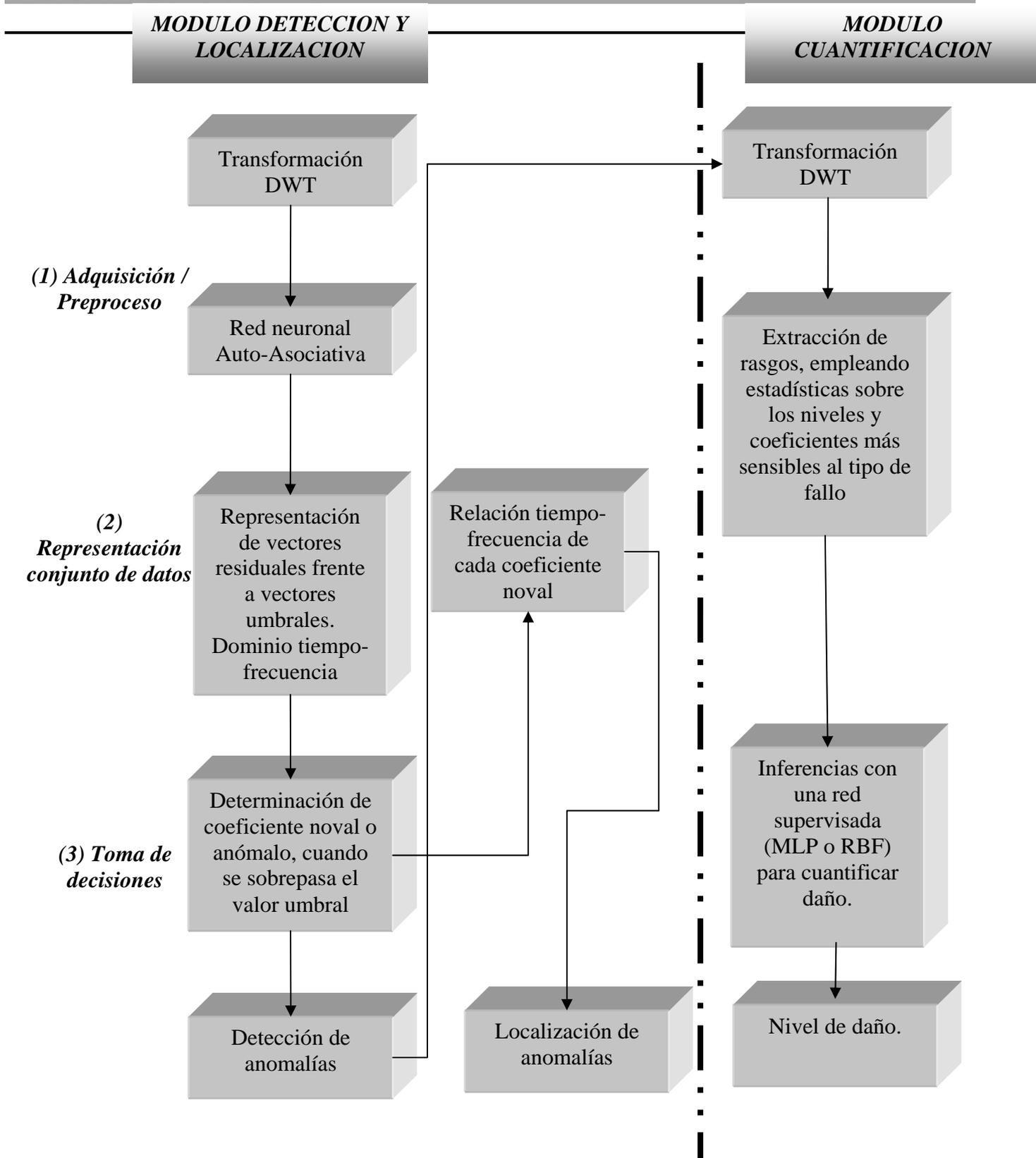


Figura 41. Diagrama de bloques del módulo de cuantificación.

3.4.2.4. Ventajas e inconvenientes del sistema de diagnóstico propuesto.

Módulo de detección y localización:

Hay que mencionar que el método propuesto al emplear la transformada DWT realiza un “rastreo” de anomalías en todo el espectro y todo el intervalo temporal considerado (con distinta resolución para cada intervalo tiempo frecuencia estimado), por lo que el filtrado de la señal original es nulo y a diferencia de otras técnicas actuales permite analizar eventos tanto a altas como a bajas frecuencias.

Otra ventaja de realizar la transformada DWT para cada giro de la máquina, es que no es necesaria la realización de promediado alguno, cuya evaluación puede ser compleja y extremadamente sensible al corte de las señales temporales.

Como no es necesario calcular ningún valor promedio (a diferencia de las técnicas basadas en el TSA) sobre un número grande de revoluciones, las predicciones se establecen en tiempo real y además se evita la pérdida de información debida por ejemplo a la falta de sincronismo.

Hay que indicar que las predicciones en tiempo real (se proporciona esta información a cada vuelta del rotor) permiten un menor almacenamiento de datos y por supuesto una diagnosis más rápida.

Otro aspecto esencial en el sistema de diagnóstico propuesto es la utilización de índices noales vectoriales, a diferencia de los índices escalares propuestos en otros procedimientos, esto proporciona información acerca de dónde se puede encontrar la perturbación detectada (posición angular) y la frecuencia de la misma.

Además, las predicciones se pueden efectuar para las distintas condiciones de operación de la maquinaria en estudio, esto tiene validez para condiciones de operación discretas.

Por otra parte, no es necesario conocer de antemano cuáles son las condiciones de operación de la maquinaria ya que para este módulo se emplean redes no supervisadas.

Por último, reseñar que las predicciones efectuadas son robustas a la presencia de ruido contaminante siempre que la variabilidad de éste sea de un orden de magnitud similar a la diferencia entre las clases de funcionamiento normal y anómalo.

La principal desventaja del método propuesto es su extrema sensibilidad a la ubicación de los acelerómetros en la carcasa de la maquinaria, ya que ésta actúa a modo de filtro cuya función de transferencia está altamente influenciada por la posición espacial de dichos transductores.

Otro inconveniente de este método es que la fase de entrenamiento debe de ser realizada para todas las condiciones de operación de la maquinaria ya que si se descarta alguna de ellas puede ser detectada como anómala.

Módulo de cuantificación:

La principal ventaja de este módulo es que, debido al empleo de redes neuronales artificiales, no se requiere una modelización de la máquina rotativa, ni en condiciones normales ni en condiciones degradadas, lo cual como se verá posteriormente es extremadamente complejo debido al comportamiento altamente no lineal de la maquinaria a analizar.

La mayoría de los estudios actuales ([60] a [63]) sólo contemplan modelos matemáticos de degradación para engranajes rectos, helicoidales y en montajes relativamente sencillos y claramente repetitivos (bancada de ensayos). Su extrapolación a montajes industriales parece inviable puesto que se requeriría un nuevo modelo matemático para las nuevas condiciones de funcionamiento (condiciones de montaje, entorno, etc.). El procedimiento desarrollado en esta tesis es aplicable en principio en cualquier entorno estable, matizando el término de estabilidad como la situación que garantice las condiciones mínimas necesarias para la obtención de un patrón de funcionamiento asociado a estas nuevas condiciones de funcionamiento normales,

puesto que únicamente se requeriría un reentrenamiento de las redes neuronales para dichas condiciones.

Una característica que conviene comentar es que, una vez entrenado el sistema de diagnóstico, las predicciones de cuantificación de daño se efectuarán en tiempo real (al igual que en el módulo anterior).

Como principal desventaja de la técnica propuesta en esta tesis en lo referente al módulo de cuantificación, es que modos de fallo distintos con grados de severidad muy diferentes pueden dar lugar a señales semejantes, lo que daría lugar a predicciones muy alejadas de la realidad.

Para solventar este problema se podría pensar en emplear un sistema previo de clasificación o de discriminación en función del tipo de defecto detectado, aunque esto es todavía más complejo puesto que generalmente se suelen dar mecanismos de degradación combinados en el deterioro de un componente. Además, si se emplearan redes neuronales para llevar a cabo tal tarea, se necesitaría un gran número de prototipos degradados que contemplaran los distintos modos de fallo a estudiar, lo cual sin duda encarecería en gran medida tal estudio.

Lo que cabe plantear es realizar un estudio estadístico en el que se analice tanto la criticidad como la frecuencia de ocurrencia de cada modo de fallo y, en función de dicho estudio, considerar sólo los modos o el modo de fallo más relevante; en este último caso tendría validez el módulo de cuantificación desarrollado en esta tesis.

Otra inconveniente es que, al igual que en el caso de detección y localización, el empleo de redes neuronales artificiales, al tratarse de una técnica de tipo “caja negra”, no permite acercarse a la realidad del fenómeno físico intrínseco al problema analizar.

CAPÍTULO 4

RESULTADOS

Una vez presentado el método de detección, se ha procedido a su implementación a fin de comprobar su validez y sus limitaciones tanto en problemas simulados numéricamente como en ensayos experimentales. En la primera parte, sección 4.1, se ha hecho hincapié en el módulo de detección cuyo esquema de funcionamiento básico fue descrito según el diagrama de bloques de la figura 40 del capítulo 3.3.2.

Posteriormente, en la sección 4.2 se ha implementado el módulo de cuantificación, detallado en el esquema de la figura 41 del mismo capítulo 3.3.2. Como se comentó en dicho capítulo este procedimiento de cuantificación tendrá un funcionamiento jerárquico y únicamente se activará cuando se haya detectado alguna anomalía previamente con el sistema de detección, dependerá, por tanto, de las alarmas mostradas con anterioridad.

4.1. DETECCIÓN Y DIAGNÓSTICO DE DAÑO

La capacidad del módulo de detección para detectar anomalías en el funcionamiento de la maquinaria rotativa se ha evaluado en dos casos diferentes. El primero corresponde a una simulación numérica mientras que en el segundo caso se aplica a los resultados de un ensayo experimental.

En el primer estudio se representan tres condiciones de operación altamente no lineales mediante tres señales registradas en una determinada máquina hipotética y se evalúa la capacidad de detección y localización en tiempo real del módulo de detección para cada una

de las condiciones. En el segundo estudio, se aplica esta metodología sobre maquinaria rotativa real consistente en una caja reductora de una estación de bombeo afectada por una “picadura” que, como se explicó en el capítulo 2.2, se asocia a un mecanismo de fallo relacionado con la fatiga de contacto.

En ambos casos la robustez del método se evalúa comprobando el índice tanto de falsas alarmas como de alarmas registradas para señales provenientes de la máquina degradada en diferentes puntos de operación.

Conviene señalar además que todos las predicciones se han efectuado por vuelta del eje de giro que contiene el par cinemático de engrane ya sea el piñón o la corona impulsada.

Todos los procesos implicados en el procedimiento de diagnosis se han implementado con MATLAB.

4.1.1. EXPERIMENTO NUMÉRICO

Para el experimento numérico, se han empleado tres señales diferentes representativas de medidas en distintas condiciones de operación y registradas por cada vuelta de una hipotética máquina rotativa. Estas señales, ‘s1’, ‘s2’ y ‘s3’, se caracterizan por su amplitud (10, 35 y 50 unidades) y por su frecuencia (10, 7 y 1 Hz).. En el dominio temporal los registros se representan mediante una rotación del piñón que incluye 51 puntos temporales de muestreo, cada 0.5 unidades angulares (figura 42).

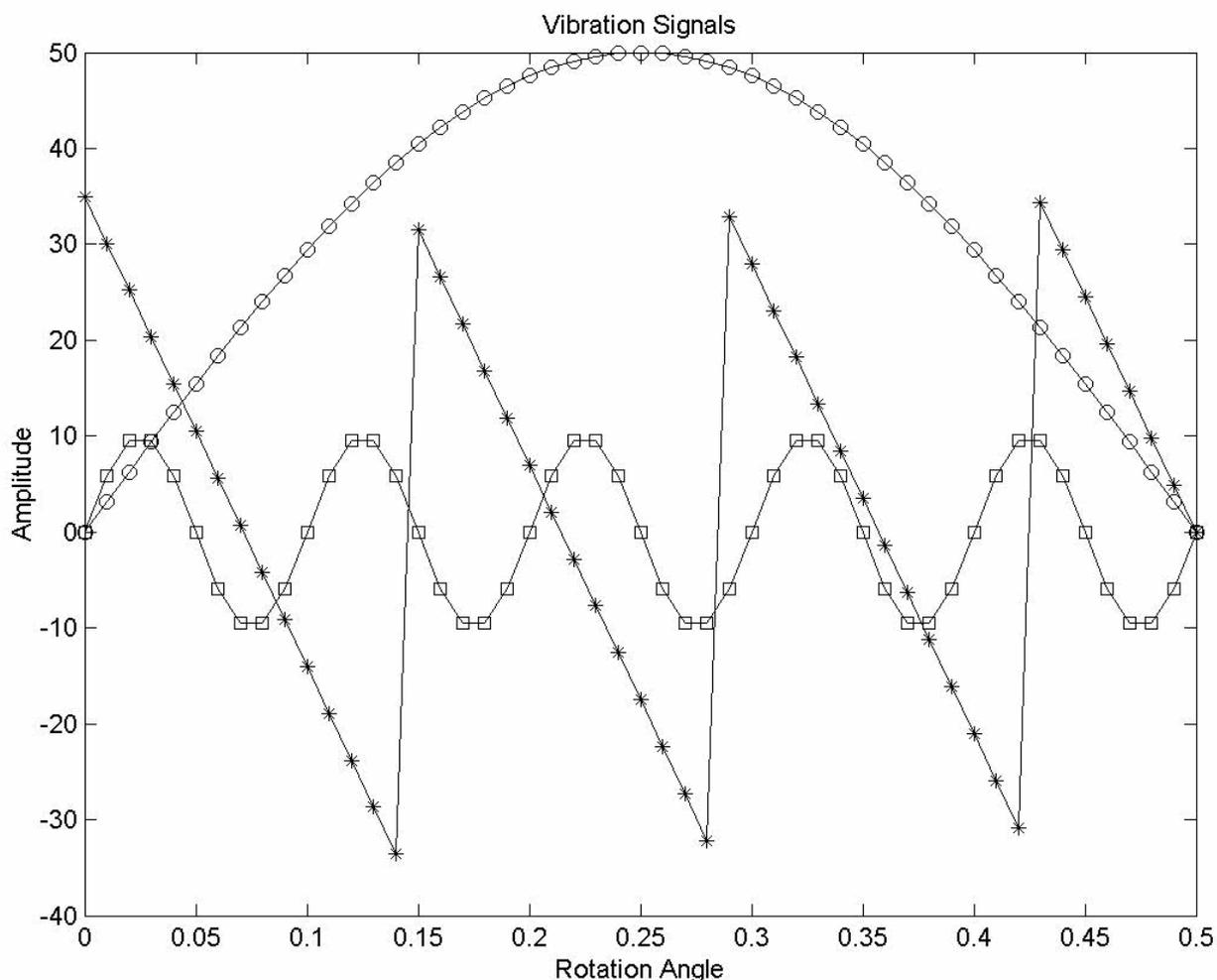


Figura 42. Señales vibratorias originales. El símbolo '□' Representa la señal ' s_1 ', los símbolos '*' y '○' representan a las señales ' s_2 ' y ' s_3 ' respectivamente.

Con objeto de estudiar la funcionalidad del sistema de detección, se han considerado 600 patrones mediante copias de las señales correspondientes a las condiciones normales de la maquinaria distorsionadas mediante la adición de distintos vectores de ruido superpuesto, a fin de ser congruentes con la realidad física experimental, ya que los registros siempre se ven afectados por distintas fuentes de ruido. Este ruido añadido a cada punto temporal es de tipo Gaussiano con desviación estándar igual a uno y promedio nulo (figura 43).

De las 600 señales patrones, 400 se emplearon en la fase de entrenamiento y validación de los parámetros de las redes auto-asociativas AANN mientras que las restantes 200 se usaron para determinar el índice de falsas alarmas de las redes AANN.

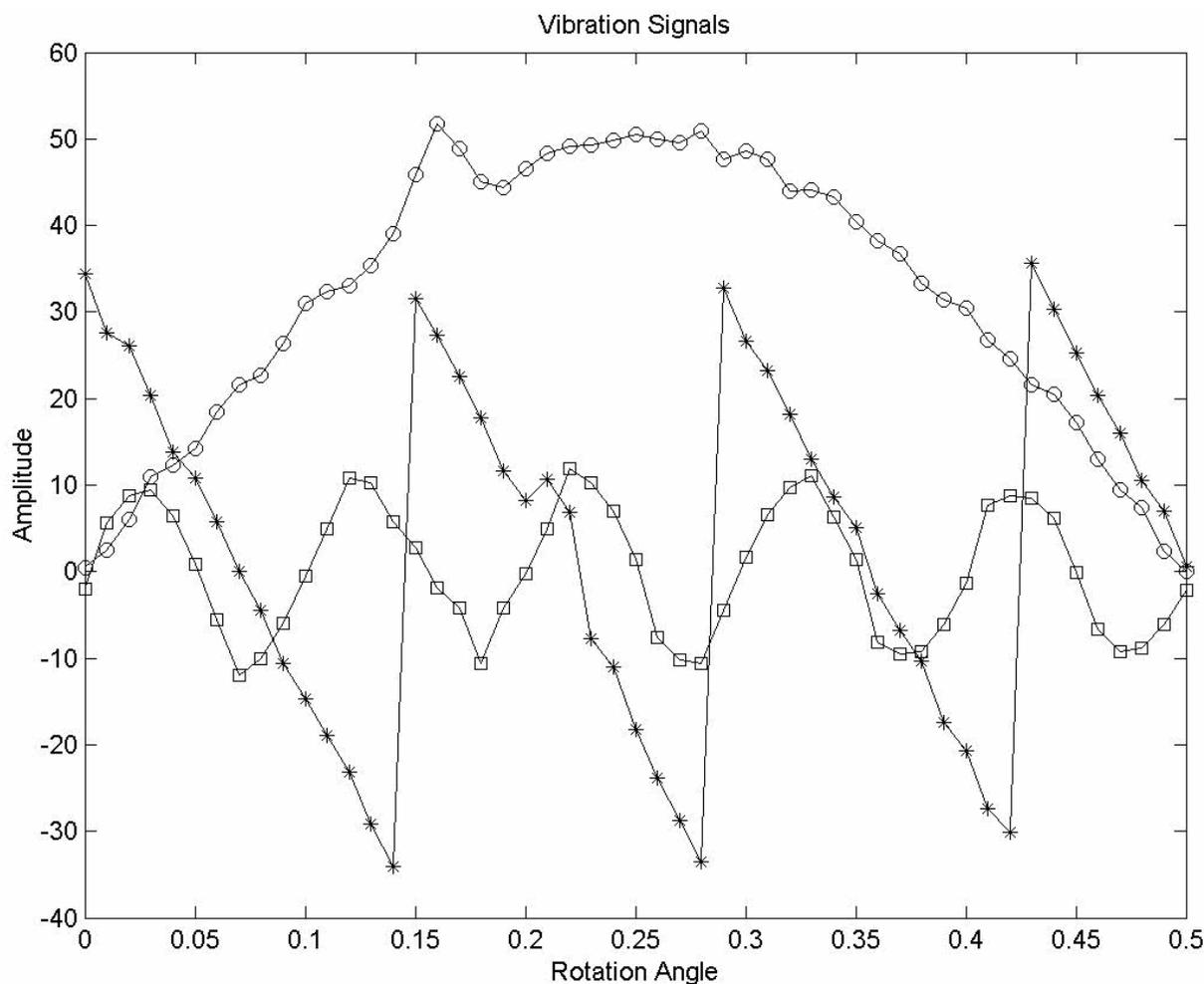


Figura 43. Perturbaciones introducidas sobre las señales con ruido. El símbolo '□' Representa la señal 's₁', los símbolos '*' y '○' representan a las señales 's₂' y 's₃' respectivamente.

Por último, indicar que la anomalía en la señal se simuló, añadiendo a la señal “ruidosa” de la figura 43 una perturbación en forma de trazador cúbico, localizado en diferentes posiciones angulares del eje de giro. Esta perturbación se simuló con trazadores de distinta amplitud (valores de 1 a 12 unidades) y de distinta duración (soportes angulares de 0.03 a 0.05

unidades angulares), generando un total de 1404 señales perturbadas las cuales se emplearon para comprobar el índice de alarmas establecido con el sistema de detección.

4.1.1.1 Preprocesamiento de señal

En esta fase los coeficientes generados a través de la transformación DWT se calculan con el propósito de ser utilizados como vectores de entrada de las redes AANN. Para conseguir esto, las señales con y sin perturbaciones se transfieren a un ordenador que las transforma al dominio de ondículas (escala-tiempo) mediante un programa implementado en MATLAB. Para realizar esta transformación DWT, es necesario seleccionar previamente la familia de ondículas más apropiada, así como el nivel de descomposición requerido; esto se realiza mediante un procedimiento de prueba-error ya que no hay ningún criterio preestablecido, a pesar de que es necesario conocer las características de la señal vibratoria (su regularidad) y de la perturbación incorporada cuando ésta sea conocida.

Asumiendo como posibles alternativas aquellas familias susceptibles del empleo del algoritmo FWT (transformada rápida con ondículas), dentro de este grupo de ondículas deberán de estudiarse las propiedades más importantes de cara a realizar la selección de la familia más apropiada; entre estas propiedades se deben considerar la regularidad, el número de momentos nulos y el soporte temporal.

De acuerdo con la primera propiedad, las ondículas Haar quedan descartadas debido a la irregularidad inherente a la propia ondícula que provoca la aparición de discontinuidades artificiales en la señal original.

Por otro lado hay que considerar el número de momentos nulos ya que determina el orden polinomial que puede ser aproximado, así ondículas de $k+1$ momentos nulos producen valores

muy cercanos a cero o nulos para los coeficientes resultantes de la transformación de un polinomio de orden k .

Como conclusión, interesarán aquellas familias que produzcan pequeños coeficientes sobre señales no perturbadas y coeficientes lo más altos posibles cuando se presenten señales con anomalías. En el problema planteado inicialmente, las señales originales (figura 43) se pueden aproximar mediante el empleo de polinomios lineales o cuadráticos mientras que las perturbaciones son de tipo cúbico, por consiguiente interesará emplear ondículas con tres momentos nulos.

Como se estudió en el capítulo 3, el soporte de una función se define como el intervalo definido por un conjunto de puntos temporales fuera del cual el valor de dicha función es nulo. Este rasgo permite identificar aquellos coeficientes que se verán afectados por un particular evento en el dominio temporal o en frecuencia. En este estudio se ha preferido emplear las ondículas más compactas temporalmente a costa de perder soporte en el dominio de la frecuencia.

Teniendo en consideración los criterios anteriormente mencionados, se seleccionaron las ondículas biortogonales 3.9 para ser empleadas en la transformación DWT, donde 3 y 9 son los órdenes de las ondículas para reconstrucción y modificación, respectivamente. Estas ondículas tienen la característica de tener tres momentos nulos y de poseer un soporte compacto en el dominio temporal.

Una vez que han sido seleccionadas las ondículas, es necesario establecer el número de niveles de descomposición requerido, basándose en la naturaleza de la señal a analizar. Como las perturbaciones se han introducido con soportes angulares comprendidos entre 0.03 y 0.05 unidades, las correspondientes frecuencias de interés estarán incluidas en el intervalo de 20-60 Hz.

Por tanto de acuerdo con la ecuación (3.2.20), las frecuencias afectadas implican extraer coeficientes a los niveles de descomposición 2, 3, 4 y 5 para la ondícula seleccionada resultando un total de 27, 23, 21 y 20 coeficientes, respectivamente para cada nivel.

4.1.1.2 Entrenamiento

Después de la extracción de rasgos utilizando como base la transformación DWT, el siguiente paso consiste en el entrenamiento de redes AANN con estos coeficientes. Dado que se precisan cuatro niveles de descomposición, se configurarán cuatro redes neuronales de reconocimiento diferentes, cada una de ellas “especializada” en un determinado nivel de descomposición.

Como se indicó en el capítulo 3, las redes AANN se caracterizan porque utilizan los mismos vectores de entrada como salida de la red. Estos vectores están constituidos por los coeficientes DWT pertenecientes a un determinado nivel de descomposición. Por consiguiente, para configurar la topología de las redes únicamente se deben establecer el número de neuronas empleadas en la etapa de codificación y decodificación, ya que las capas de entrada y salida quedan fijadas por el número de coeficientes DWT de cada nivel de descomposición seleccionado, además el “cuello de botella” (figura 35b) está formado por una única neurona.

En general, no resulta sencillo establecer el mejor tamaño de la red para un determinado sistema, este dimensionado se puede realizar a través de un procedimiento de prueba-error que consiste en configurar distintas arquitecturas AANN con diferente número de neuronas, finalmente seleccionando aquellas configuraciones que resulten más óptimas.

Para conseguir tal propósito, una estrategia consiste en empezar aleatoriamente con un número determinado de neuronas (en las etapas de codificación y decodificación) y ajustar los

pesos sinápticos de esta arquitectura inicial mediante un algoritmo RP, posteriormente se valida con un conjunto de patrones no visto por la red y se registran dos tipos de errores, el de ajuste y el de varianza. A continuación se incrementa el número de neuronas en las etapas de codificación y decodificación y se repite el proceso anterior, eligiendo finalmente aquella arquitectura que minimiza la función de error ajuste-varianza.

Esta estrategia iterativa de crecimiento se ha seguido hasta alcanzar un número de neuronas en las etapas de codificación y decodificación del 70%, aproximadamente, del número de neuronas de entrada-salida.

Finalmente las siguientes topologías se han adoptado para cada nivel de descomposición:

-Nivel 2: 27x5x1x5x27

-Nivel 3: 23x9x1x9x23

-Nivel 4: 21x10x1x10x21

-Nivel 5: 20x10x1x10x20

Una de las ventajas fundamentales del tipo de red AANN consiste en su habilidad de extraer los rasgos fundamentales de todos los patrones de entrenamiento, lo cual supone una reducción del ruido superpuesto en los mismos.

Este efecto queda claramente patente en la figura 44, en la cual se muestran los coeficientes correspondientes a los cuatro niveles de descomposición como entrada (figuras 44.a, c, e y g) y salida (figuras 44.b, d, f y h) de la red AANN.

Es evidente, comparando ambas figuras, que las redes AANN son capaces de extraer los rasgos característicos de los tres tipos de señales, lo cual representa, sin duda, una clara ventaja a la hora de detectar cualquier perturbación debida a la presencia de un fallo en la maquinaria.

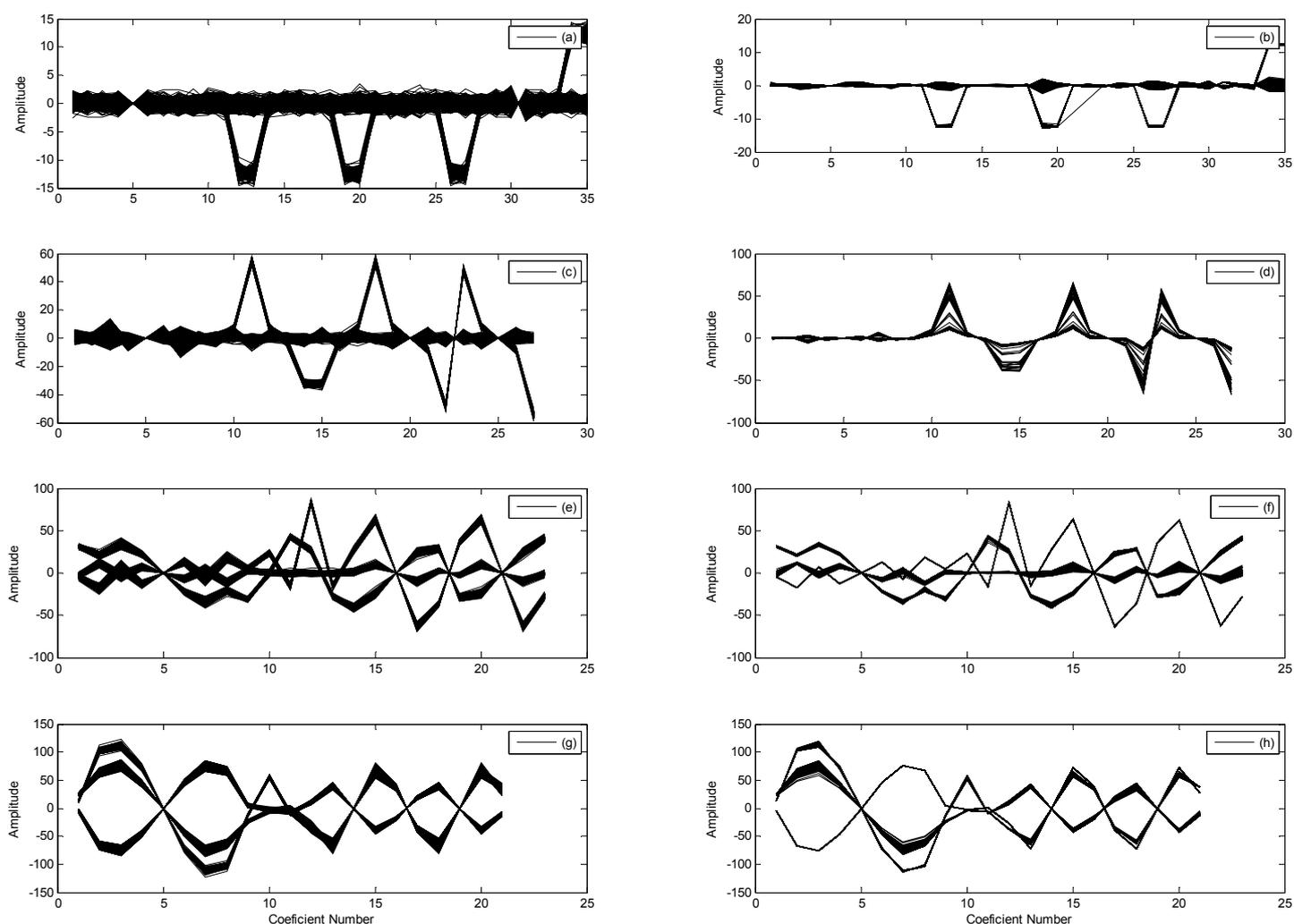


Figura 44. Comparativa de los coeficientes DWT de entrada y salida de la red AANN, las partes (a) y (b) corresponden al nivel 2 de descomposición, las partes (c) y (d) al nivel 3, las partes (e) y (f) al nivel 4 y por último las partes (g) y (h) al nivel 5.

Una vez entrenadas las cuatro redes, es necesario establecer los umbrales, es decir los índices noales (ver capítulo 3). Esto se realiza para cada coeficiente de entrada a la red neuronal, por tanto se obtiene un vector de índices noales para cada nivel de descomposición. Para determinar los vectores noales, una vez ajustados los pesos de las redes en la fase de entrenamiento, los mismos vectores que se emplearon en dicho entrenamiento se vuelven a introducir en la red y la diferencia entre las entradas y las salidas de las redes AANN da lugar a una serie de vectores residuales. La distribución estadística de estos residuos determina

finalmente los valores umbrales, es decir, los límites a partir de los cuales los residuos son muy improbables.

4.1.1.3 Reconocimiento

En esta fase, se validan las redes AANN entrenadas empleando 200 patrones correspondientes a condiciones normales de la maquinaria y 1404 patrones correspondientes a condiciones anómalas en dicha maquinaria.

En primer lugar, se examinan las 200 señales “sanas” con ruido, con el propósito de establecer los índices de falsas alarmas. Para ello, una vez que estas señales son preprocesadas empleando los cuatro niveles de descomposición, los coeficientes DWT generados se utilizan como entrada a las redes AANN y el porcentaje de falsas alarmas es el indicado en la tabla 2. Se puede decir, a partir de estos resultados, que este porcentaje es muy bajo especialmente para los tres niveles superiores, esto asegura la fiabilidad de la red en el caso de que se presenten señales no perturbadas.

PORCENTAJE DE FALSAS ALARMAS	
NIVEL 2	6.50%
NIVEL 3	2.50%
NIVEL 4	4.50%
NIVEL 5	2.50%

Tabla 2. Número de señales detectadas como anómalas sobre 100 señales sanas.

Posteriormente, con el objetivo de examinar la capacidad de detección del método propuesto, se validaron las 1404 señales “perturbadas” . Como se indicó anteriormente, las perturbaciones consisten en trazadores cúbicos añadidos a las señales originales en diferentes localizaciones situadas entre 0.1 y 0.4 unidades angulares, con soportes angulares comprendidos entre 0.03 y 0.05 unidades angulares y con amplitudes incluidas en el rango 1-12.

Los resultados obtenidos para las tres señales, en función de las longitudes de los soportes y en función de las amplitudes, se muestran en las tablas 3 y 4, respectivamente.

En estas tablas se indica el porcentaje de detección realizado con éxito.

En vista de los resultados obtenido, se pueden realizar las siguientes conclusiones:

- a) En general, el funcionamiento de las redes AANN como detectores es satisfactorio.
- b) La duración de la perturbación (soporte) apenas afecta a los resultados.
- c) La sensibilidad del método depende de la amplitud de la perturbación ya que, cuando éste está próximo a los niveles de amplitud del ruido, los pronósticos se convierten en poco fiables. No obstante, hay que señalar, que las redes son capaces de detectar aproximadamente la mitad de los casos con señales degradadas cuando la amplitud de la perturbación es igual a tres.

PORCENTAJE DE DETECCIÓN SOBRE SEÑALES PERTURBADAS DE DISTINTO SOPORTE			
SOPORTE ANGULAR	S1	S2	S3
0.03	70.51%	75.54%	80.13%
0.04	83.33%	82.69%	84.61%
0.05	83.33%	88.46%	85.25%

Tabla 3. Número de señales detectadas como anómalas sobre 100 señales degradadas con distinto soporte.

PORCENTAJE DE DETECCIÓN SOBRE SEÑALES PERTURBADAS DE DISTINTA AMPLITUD			
AMPLITUD DE LA PERTURBACIÓN	S1	S2	S3
1	25.64%	30.76%	35.89%
2	38.46%	51.28%	33.33%
3	48.71%	48.71%	76.92%
4	64.10%	69.23%	71.79%
5	87.17%	94.87%	92.30%
6	97.43%	100%	100%
7	100%	100%	100%
8	100%	100%	100%
9	100%	100%	100%
10	100%	100%	97.43%
11	100%	100%	100%
12	100%	100%	100%

Tabla 4. Número de señales detectadas como anómalas sobre 100 señales degradadas con distinta amplitud

Es necesario remarcar que con este sistema de detección, se determina un índice noval para cada coeficiente DWT generado para cada nivel de descomposición. Esto, sin duda, constituye una ventaja en comparación con otras técnicas de detección ya que produce una distribución de las alarmas de la maquinaria recogida en el dominio tiempo-frecuencia, que permite localizar las anomalías en posición angular y en frecuencia.

En la figura 45 se representan los índices noval obtenidos para nueve señales dañadas correspondientes a perturbaciones de 9 a 11 unidades de amplitud introducidas sobre señales de tipo 's₃' (figura 43). Estas perturbaciones son introducidas en la posición angular 0.25 unidad, es decir, aproximadamente en la mitad de la vuelta del eje de giro.

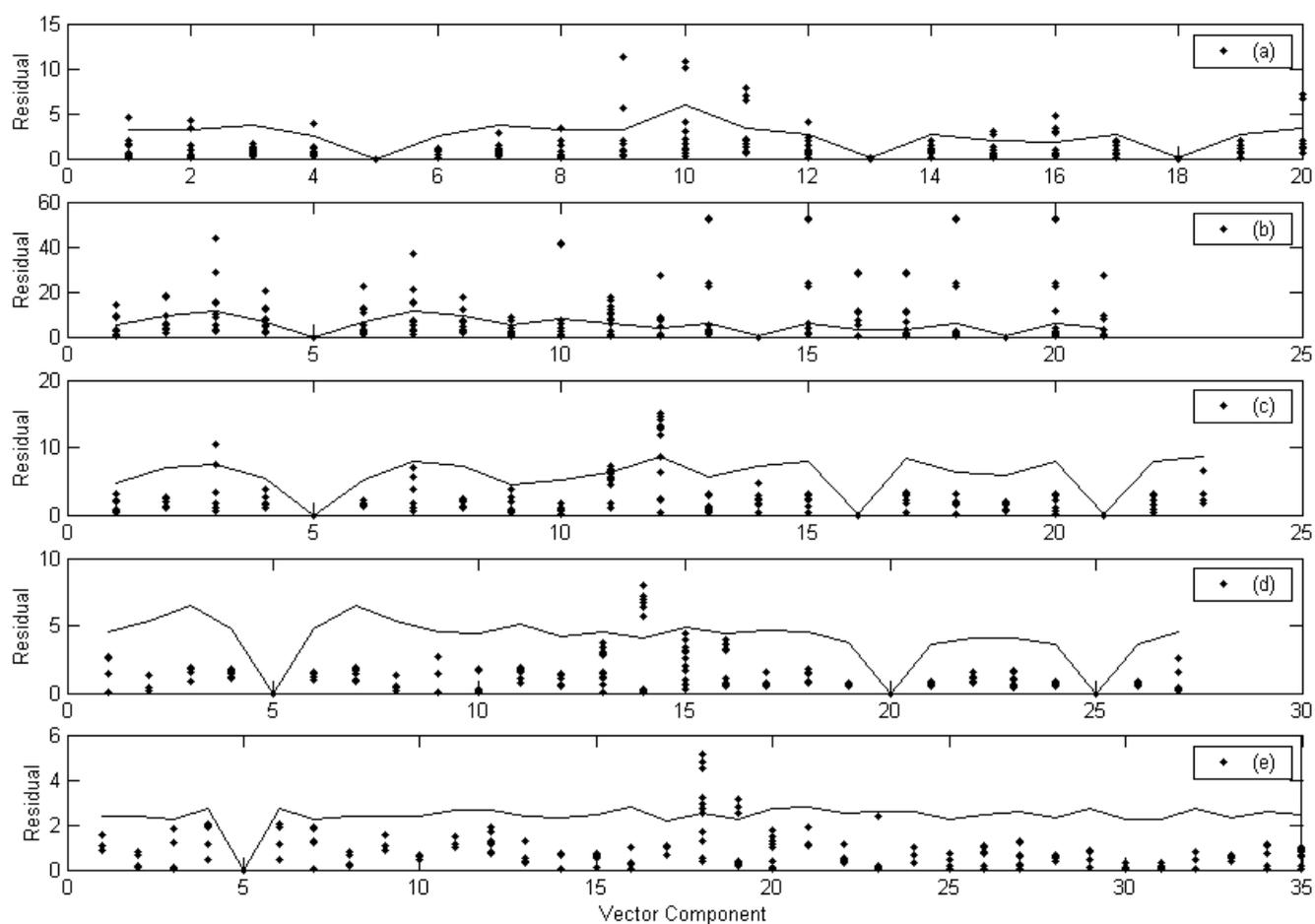


Figura 45. Vectores de índices noales. El símbolo '♦' representa el índice noal para cada coeficiente y la línea continua representa los valores umbrales. El recuadro (a) indica el 5 nivel de descomposición, y los recuadros (b), (c), (d) corresponden a los niveles 4, 3, 2 y 1 respectivamente.

Como se puede apreciar en la figura 45 los residuos más excitados son aquellos que corresponden a los coeficientes DWT situados en la parte central del histórico del registro, por tanto vemos que el sistema de detección realiza una buena localización del daño.

4.1.2. ESTACIÓN DE BOMBEO CON PICADURAS

Las señales vibratorias para este segundo ejemplo se obtuvieron de la referencia Ypma [65]. Estos registros fueron obtenidos de las mediciones realizadas sobre la carcasa de dos bombas hidráulicas idénticas, accionadas cada una de ellas por un motor eléctrico a través de una caja reductora con dos pasos de reducción.

La velocidad de funcionamiento de las máquinas en todos los casos se mantiene constante a 1000 rpm, estando el primer armónico del primer y segundo paso de reducción en 333.33 Hz y 666.6 Hz. Respectivamente (figura 46).



Figura 46. Estación de bombeo en Lemmer, Holanda.

Mientras una de las bombas está funcionando en condiciones normales, la otra funciona con problemas de picadura en ambos pasos de reducción

Las señales se obtuvieron a partir de siete acelerómetros situados en diferentes posiciones de la carcasa de la maquinaria. Los dos primeros se montaron cerca del eje impulsor en dirección

radial y separados un ángulo de 90°, el tercero se instaló en dirección axial y los otros cuatro restantes se montaron radialmente en diferentes posiciones de la carcasa de la maquinaria.

El conjunto de medidas se registró en dos condiciones de operación correspondientes a condiciones de baja y alta carga (mínima y máxima potencia entregada por el motor eléctrico).

Por otra parte, las señales se filtraron mediante filtros paso-bajo con frecuencias de corte de 1 y 5 kHz, respectivamente, empleando un muestreo de 3.2 y 12.8 kHz, dando lugar a un conjunto de medidas agrupadas a su vez en cuatro subconjuntos.

En esta tesis, sólo las señales registradas con el filtro de 1 kHz se transformaron mediante DWT con objeto de extraer los rasgos principales de dichas señales como entradas a las redes neuronales de acuerdo con el procedimiento descrito en detalle a lo largo del capítulo 3.3.2.

Como las señales de baja frecuencia se registraron durante 24.3 segundos y la velocidad angular de la máquina se mantuvo a 1000 rpm, se obtuvieron un total de 405 giros del eje de dicha máquina, es decir 405 señales disponibles para cada sensor.

4.1.2.1 Preprocesamiento de señal

En esta fase, al igual que en el caso del experimento numérico (sección 4.1.1), es necesario seleccionar tanto la familia de ondículas como el nivel de descomposición requerido.

Como se ha reiterado en varias ocasiones en esta tesis, la presencia de un modo de fallo en los engranajes tendrá una importante influencia en las medidas vibratorias adquiridas a través de sensores instalados en la carcasa de la maquinaria rotativa, modulando estas señales a la frecuencia de engrane y sus armónicos asociados. Este es el aspecto clave a la hora de seleccionar el nivel de descomposición DWT necesario.

Únicamente se consideraron las ondículas ortogonales y biortogonales porque estas familias permiten el empleo del algoritmo FWT. Dentro de las familias ortogonales las ondículas de Haar fueron eliminadas puesto que como se vio anteriormente introducen irregularidades artificiales en la señal original.

Finalmente, a pesar de que había otras alternativas dentro del grupo de selección anteriormente comentado, se optó por la familia de ondículas Daubechies con seis momentos nulos ('db6').

Por otro lado, como las frecuencias de engranaje de ambos pasos de reducción son 333.33 Hz y 666.6 Hz, respectivamente, se necesitaron tres niveles de descomposición para analizar el contenido de frecuencias influenciado por la presencia de un fallo en el engrane de la maquinaria.

Para cada uno de estos tres niveles, se obtuvieron 105, 58 y 34 coeficientes DWT que constituyen los vectores de entrada de las redes auto-asociativas AANN.

4.1.2.2 Entrenamiento

En este caso, se necesitaron configurar 21 redes AANN ya que las señales se adquirieron a través de siete transductores y a que además se llevó a cabo la descomposición DWT en tres niveles.

De las 810 señales registradas en la bomba no dañada, 650 se escogieron como patrones de entrenamiento de las redes AANN mientras que las restantes 160 se emplearon para comprobar el índice de falsas alarmas obtenido con las redes entrenadas.

Tanto el entrenamiento como la configuración de las redes AANN se realizó de idéntica manera que en el experimento numérico (sección 4.1.1), las arquitecturas empleadas son las indicadas en la tabla 5.

	TOPOLOGIA DE LAS REDES AANN		
	Nivel D1	Nivel D2	Nivel D3
Sensor 1	105x90x1x90x105	58x33x1x33x58	34x17x1x17x34
Sensor 2	105x45x1x45x105	58x31x1x31x58	34x18x1x18x34
Sensor 3	105x81x1x81x105	58x32x1x32x58	34x12x1x12x34
Sensor 4	105x59x1x59x105	58x37x1x37x58	34x16x1x16x34
Sensor 5	105x87x1x87x105	58x38x1x38x58	34x20x1x20x34
Sensor 6	105x86x1x86x105	58x37x1x37x58	34x20x1x20x34
Sensor 7	105x86x1x86x105	58x39x1x39x58	34x19x1x19x34

Tabla 5. Número de capas y de neuronas por capa para cada red neuronal. Las redes neuronales están definidas para cada sensor y nivel de descomposición.

Una vez que las redes fueron configuradas y entrenadas, los patrones de entrenamiento se introdujeron nuevamente en las redes para determinar los vectores residuales y, de esta forma, los vectores de índices novales.

4.1.2.3 Reconocimiento

En primer lugar se emplearon los 160 patrones obtenidos de la bomba no dañada y no empleados durante el entrenamiento para determinar el índice de falsas alarmas generadas con las distintas redes. Este índice se indica en la tabla 6 para las dos condiciones de carga y para los siete sensores.

Se puede comprobar que la fiabilidad de las redes es altamente satisfactoria cuando se emplean señales correspondientes a mediciones sobre la máquina sin dañar.

Por último, las 810 señales obtenidas de la bomba con picaduras se introdujeron en las redes AANN. El porcentaje de casos satisfactorios es el indicado en la tabla 7.

Se puede extraer como conclusión principal, a raíz de los resultados obtenidos en la tabla 7, que la capacidad del método es muy sensible a la localización de los transductores y de las condiciones operativas de la máquina rotativa. Se obtuvieron resultados muy buenos para los sensores radialmente instalados 4,6 y 7, para los dos primeros niveles de descomposición y bajo condiciones de baja carga (baja potencia).

Esta variabilidad en los resultados indicados en la tabla 7 se puede deber a la función de transferencia entre el par cinemático de engrane y la localización en la carcasa del acelerómetro donde la señal vibratoria es adquirida. Teniendo en consideración esta última aseveración, las mejores localizaciones serían aquellas que fueran lo más sensible posible a la frecuencia de fallo y sus armónicos, además hay que tener en cuenta que esta función de transferencia también cambiaría con las condiciones operativas de la maquinaria.

INDICE DE FALSAS ALARMAS			
BAJA CARGA		ALTA CARGA	
Nivel D1	Índice	Nivel D1	Índice
Sensor 1	3,95 %	Sensor 1	0,246 %
Sensor 2	4,19 %	Sensor 2	13,82 %
Sensor 3	3,95 %	Sensor 3	2,46 %
Sensor 4	6,17 %	Sensor 4	8,88 %
Sensor 5	2,46 %	Sensor 5	2,46 %
Sensor 6	0 %	Sensor 6	4,19 %
Sensor 7	6,17 %	Sensor 7	11,35 %
Nivel D2	Índice	Nivel D2	Índice
Sensor 1	1,72 %	Sensor 1	11,6 %
Sensor 2	1,48 %	Sensor 2	1,48 %
Sensor 3	1,23 %	Sensor 3	0 %
Sensor 4	3,4 %	Sensor 4	0 %
Sensor 5	8,14 %	Sensor 5	6,91 %
Sensor 6	0,74 %	Sensor 6	10,37 %
Sensor 7	8,39 %	Sensor 7	10,12 %
Nivel D3	Índice	Nivel D3	Índice
Sensor 1	1,23 %	Sensor 1	0,74 %
Sensor 2	0,49 %	Sensor 2	0 %
Sensor 3	4,93 %	Sensor 3	0 %
Sensor 4	10,12 %	Sensor 4	6,17 %
Sensor 5	4,93 %	Sensor 5	4,93 %
Sensor 6	1,23 %	Sensor 6	0 %
Sensor 7	0,246 %	Sensor 7	0 %

Tabla 6. Número de señales detectadas como anómalas sobre 100 señales sanas. Los resultados se expresan por nivel de descomposición y por caso de carga.

ÍNDICE DE DETECCIÓN			
BAJA CARGA		ALTA CARGA	
Nivel D1	Índice	Nivel D1	Índice
Sensor 1	43,456 %	Sensor 1	0 %
Sensor 2	67,65 %	Sensor 2	7,65 %
Sensor 3	2,46 %	Sensor 3	3,7 %
Sensor 4	97,03 %	Sensor 4	63,2 %
Sensor 5	0,7 %	Sensor 5	0 %
Sensor 6	86,67 %	Sensor 6	0 %
Sensor 7	100 %	Sensor 7	100 %
Nivel D2	Índice	Nivel D2	Índice
Sensor 1	14,56 %	Sensor 1	0,246 %
Sensor 2	28,14 %	Sensor 2	3,20 %
Sensor 3	3,2 %	Sensor 3	0 %
Sensor 4	99,012 %	Sensor 4	2,71 %
Sensor 5	27,65 %	Sensor 5	13,82 %
Sensor 6	88,64 %	Sensor 6	5,18 %
Sensor 7	100 %	Sensor 7	83,2 %
Nivel D3	Índice	Nivel D3	Índice
Sensor 1	0,7 %	Sensor 1	0 %
Sensor 2	0,25 %	Sensor 2	1 %
Sensor 3	0 %	Sensor 3	0 %
Sensor 4	38,02 %	Sensor 4	5,67 %
Sensor 5	2,46 %	Sensor 5	0
Sensor 6	0,49 %	Sensor 6	0 %
Sensor 7	19,75 %	Sensor 7	1,48 %

Tabla 7. Número de señales detectadas como anómalas sobre 100 señales degradadas. Los resultados se expresan por nivel de descomposición y por caso de carga.

Se puede apreciar que los mejores resultados fueron obtenidos para las condiciones de baja carga, esto puede deberse a que las condiciones operativas de baja carga se encuentran más alejadas del punto de funcionamiento nominal de la máquina, donde por otra parte, se esperan menores valores del error en la transmisión y por tanto menor nivel de vibraciones registradas. Por tanto, en condiciones de baja carga, parece claro que se dan valores mayores en el error de transmisión y por tanto un incremento en las aceleraciones registradas.

Por otra parte, para este último caso en particular, dado que el número de dientes en un paso de reducción es múltiplo del otro, no es posible localizar el componente que falla y por tanto sólo se pueden proporcionar alarmas.

4.2. CUANTIFICACIÓN

En este apartado se presentan resultados relativos a la cuantificación de modos de fallo en maquinaria rotativa utilizando para ello un modelo numérico de una caja de engranajes.

Se pretende ampliar la capacidad del sistema de detección según lo descrito en el capítulo 3, con el propósito de que éste sea capaz de establecer también la severidad del daño lo que equivale a un nivel de detección 3 según la escala de Rytter [59].

4.2.1. MODELO NUMÉRICO DE CAJA DE ENGRANAJES

Como es bien sabido, las cajas reductoras y multiplicadoras constituyen uno de los componentes más habituales en maquinaria rotativa.

Todos los datos de esta sección han sido recopilados a partir del modelo numérico que Dalpiaz empleó para caracterizar las vibraciones a torsión de un sistema de transmisión de potencia, ([66], [67]).

El sistema de transmisión de potencia consiste en un banco de pruebas de engranajes y está constituido por dos ejes paralelos conectados en sus extremos mediante dos cajas de engranajes idénticas, el par se transmite por medio de una polea instalada en la parte central del eje impulsor (figura 47)

Cada caja reductora tiene un único paso de reducción de módulo 3 mm. y está formada por engranajes rectos con un número de dientes de 28 y 55 dientes para piñón y rueda, respectivamente.

Una descripción más rigurosa y detallada del sistema puede encontrarse en [68].

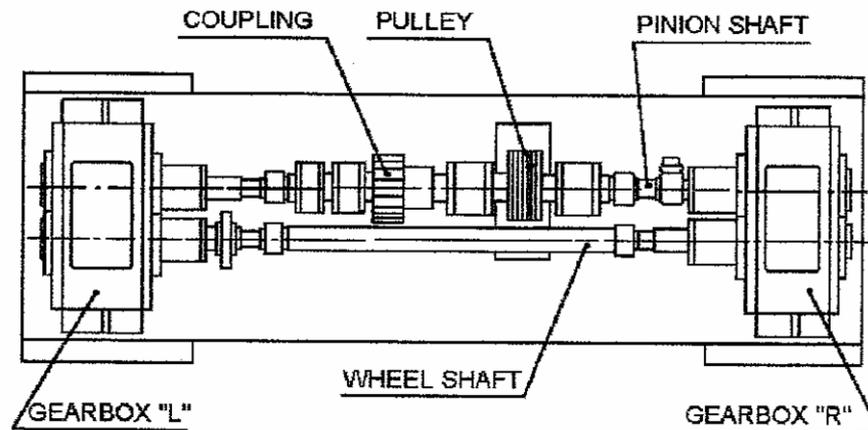


Figura 47. Esquema de tren de potencia empleado como banco de ensayos de engranajes por Dalpiaz.

Las condiciones de operación del banco de pruebas corresponden a un par nominal de 1139 Nm y a velocidades de rueda y piñón de 254.5 y 500 rpm respectivamente.

Por consiguiente estas condiciones implican una frecuencia de engrane de 233.33 Hz. Las señales generadas a partir del modelo numérico representan las aceleraciones tangenciales medidas en el círculo base de la rueda 9 (figura 48).

Además, para ser congruentes con las condiciones operativas reales a este par constante se le ha superpuesto un ruido blanco de promedio nulo y varianza $0.1 \text{ (Nm)}^2/\text{Hz}$.

4.2.1.1. Modelo matemático.

Con objeto de simular las características dinámicas del banco de ensayos, se empleó un modelo no lineal de parámetros concentrados desarrollado por Dalpiaz ([66], [67]). En este modelo solamente se consideraron las vibraciones a torsión generadas por la acción de engrane, por consiguiente sólo fueron tenidas en cuenta las rigideces y amortiguamientos a torsión de los ejes y rodamientos tal y como puede apreciarse en el esquema de la figura 48.

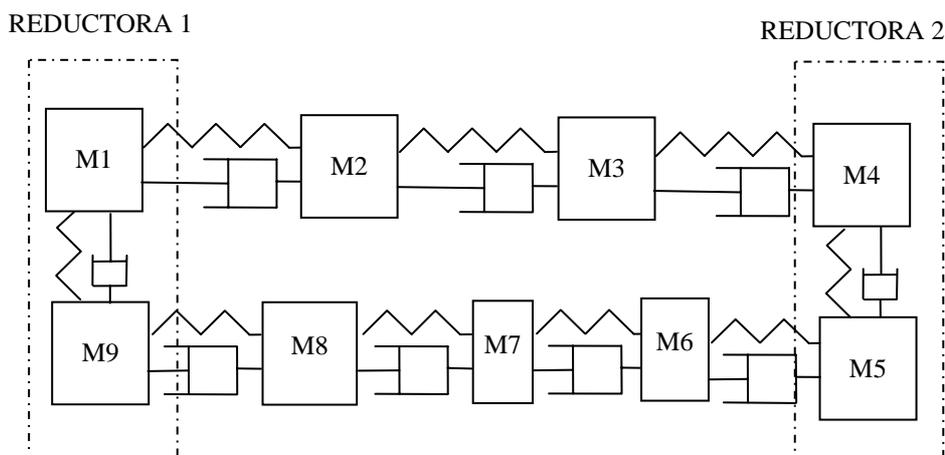


Figura 48. Esquema de masas, muelles y amortiguadores empleado en el modelo de parámetros concentrados.

Como se ha comentado en capítulos anteriores la principal fuente de vibración es la generada por cada par de engrane y ésta ha sido modelizada mediante una función dependiente del tiempo. La rigidez de cada par de engrane viene representada en el esquema de la figura 48 por los muelles situados entre las masas 1 y 9 y entre las masas 4 y 5.

Dicha rigidez se ha simulado en esta tesis mediante un término sinusoidal variable superpuesto a un término constante, a diferencia del modelo Dalpiaz que consideró el término variable constituido por dos mitades de dos funciones sinusoidales ([66], [67]).

Este modelo de rigidez de engrane no contempla las contribuciones en el error de la transmisión originadas por defectos de diseño, fabricación y montaje; únicamente tiene en cuenta la variación de rigidez debida a la variación del punto de contacto entre dientes sobre el plano de presión.

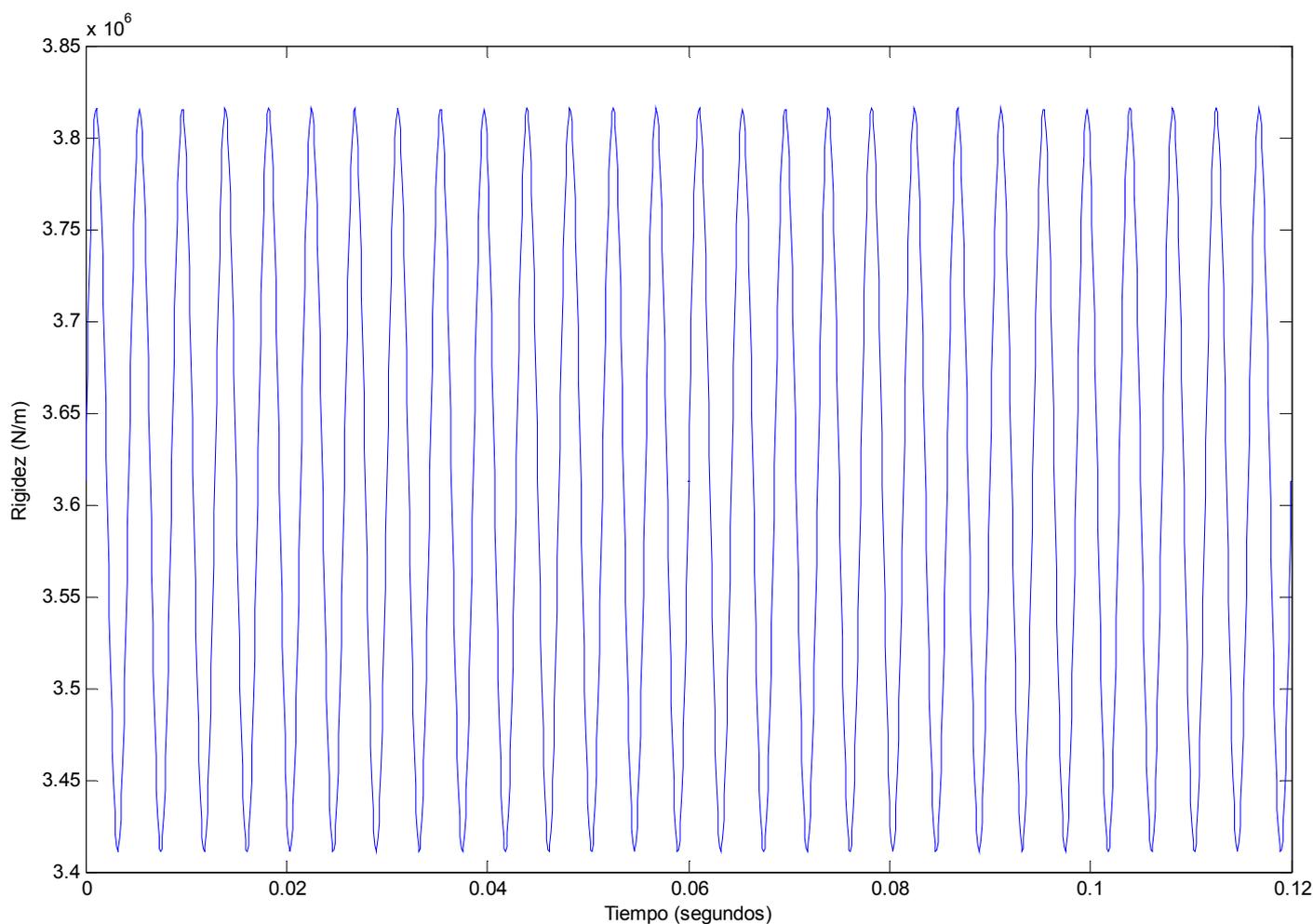


Figura 49. Modelo de rigidez variable para el par de engrane.

Por otra parte aunque las dos cajas reductoras sean idénticas existe un desfase entre ellas, es muy improbable que entren los dientes en contacto al mismo tiempo. Dicho desfase fue establecido disminuyendo las desviaciones observadas entre el modelo numérico y los resultados experimentales especificados por Dalpiaz ([66], [67]).

El modelo numérico se ha implementado mediante la interfaz SIMULINK del programa MATLAB y las ecuaciones de movimiento se han resuelto mediante el algoritmo de diferencias finitas de Dormand-Prince [69] obteniéndose el espectro de respuesta mostrado en la figura 50 para la aceleración tangencial de la rueda 9.

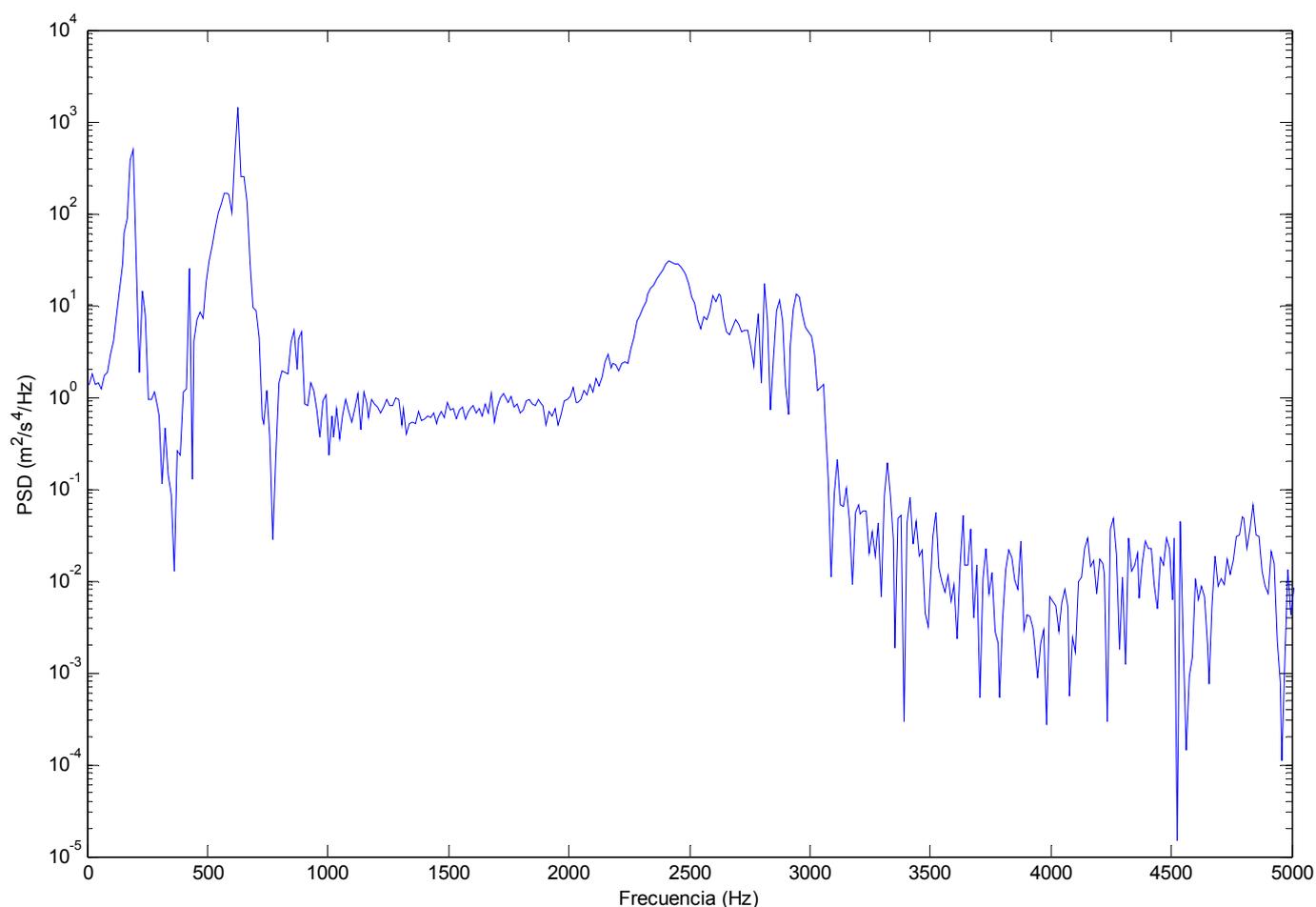


Figura 50. Espectro de potencia obtenido para la aceleración de la rueda 9.

A pesar de su simplicidad, el modelo parece ajustarse bastante bien a los resultados experimentales por debajo de 1 kHz. ([66], [67]).

Por otra parte, a pesar de que el modelo ofrezca una adecuada simulación dinámica del prototipo experimental, cabe destacar el hecho de que el principal objetivo a perseguir es el empleo de este modelo numérico como banco de ensayos “virtual” para poder introducir defectos simulados con distinto grado de severidad; de esta forma se puede comprobar la funcionalidad del módulo de cuantificación desarrollado en esta tesis.

4.2.1.2. Modelo dañado

Como se comentó en capítulos anteriores cualquier defecto en un par de engrane implica un efecto de modulación sobre el registro de la señal vibratoria ya sea tanto en magnitud como en fase. Este efecto tiene los rasgos de transitoriedad y periodicidad.

En esta tesis se ha considerado como escenario de daño un defecto local, como podría ser la presencia de una grieta, en la base de un diente (figura 51).

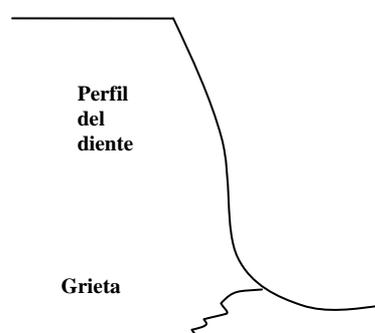


Figura 51. Esquema representativo del modo de fallo.

Este tipo de defecto tiene su origen en los mecanismos de fatiga a flexión del diente y es originado por los concentradores geométricos de tensión en el radio de acuerdo de la base del diente.

La degradación se ha incorporado al modelo numérico antes mencionado (figura 49) mediante una reducción en la rigidez torsional de engrane; dicha reducción se ha conseguido, a pesar de que en teoría debiera producirse de forma abrupta y discontinua, modulando el término constante de la figura 49 con un término sinusoidal que afecte a la mitad del periodo de giro del eje. Como consecuencia de esto, la función de rigidez torsional de engrane ha quedado modificada dando lugar a la onda rizada de la figura 52.

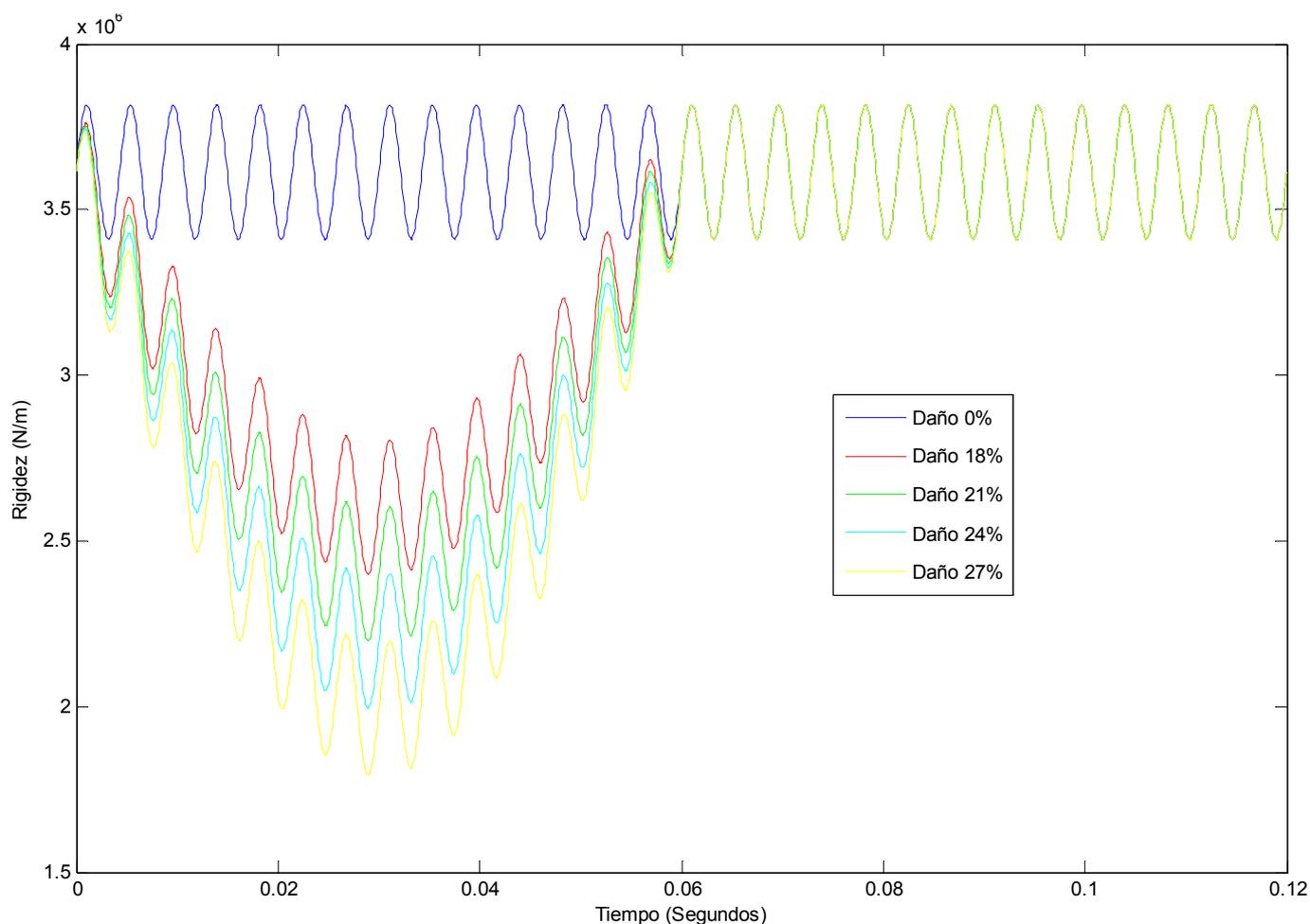


Figura 52. Modelo de rigidez de engrane variable para distintos grados de fallo.

En los estudios realizados en esta tesis se han considerado 10 niveles de severidad del daño, con reducciones del 3 al 30% de la rigidez torsional promedio respecto al caso “sano”.

Hay que indicar que algunos autores han simulado defectos en engranajes previo empleo de series de Fourier, dada la naturaleza armónica del tipo de fallo [13].

Antes de proseguir, conviene señalar las siguientes limitaciones importantes de este modelo de degradación:

- 1-La severidad del daño únicamente se ha introducido previa disminución en la amplitud de la rigidez torsional despreciando cualquier cambio en el periodo de engrane cuando,

como es sabido, los periodos de apertura y cierre de la grieta dependerán del grado de severidad del daño.

2-No se ha considerado ninguna modificación en los valores del amortiguamiento, a pesar de que es de esperar una disipación de energía muy diferente cuando en el sistema aparezcan grietas.

Además, conviene señalar que un modo de fallo como el que se pretende simular en este estudio, agrietamiento de la base del diente, origina cambios mucho más abruptos en la señal temporal registrada que los provocados por el modelo de daño propuesto en esta sección, siendo dichos cambios más fáciles de detectar ya que implicarán grandes oscilaciones en las amplitudes registradas. Por tanto cabe esperar que el sistema desarrollado en esta tesis detecte y cuantifique estos cambios reales de forma mucho más eficiente que con el escenario de fallo descrito en esta sección.

Los espectros de potencia para los 10 niveles de degradación considerados se muestran en la figura 53.

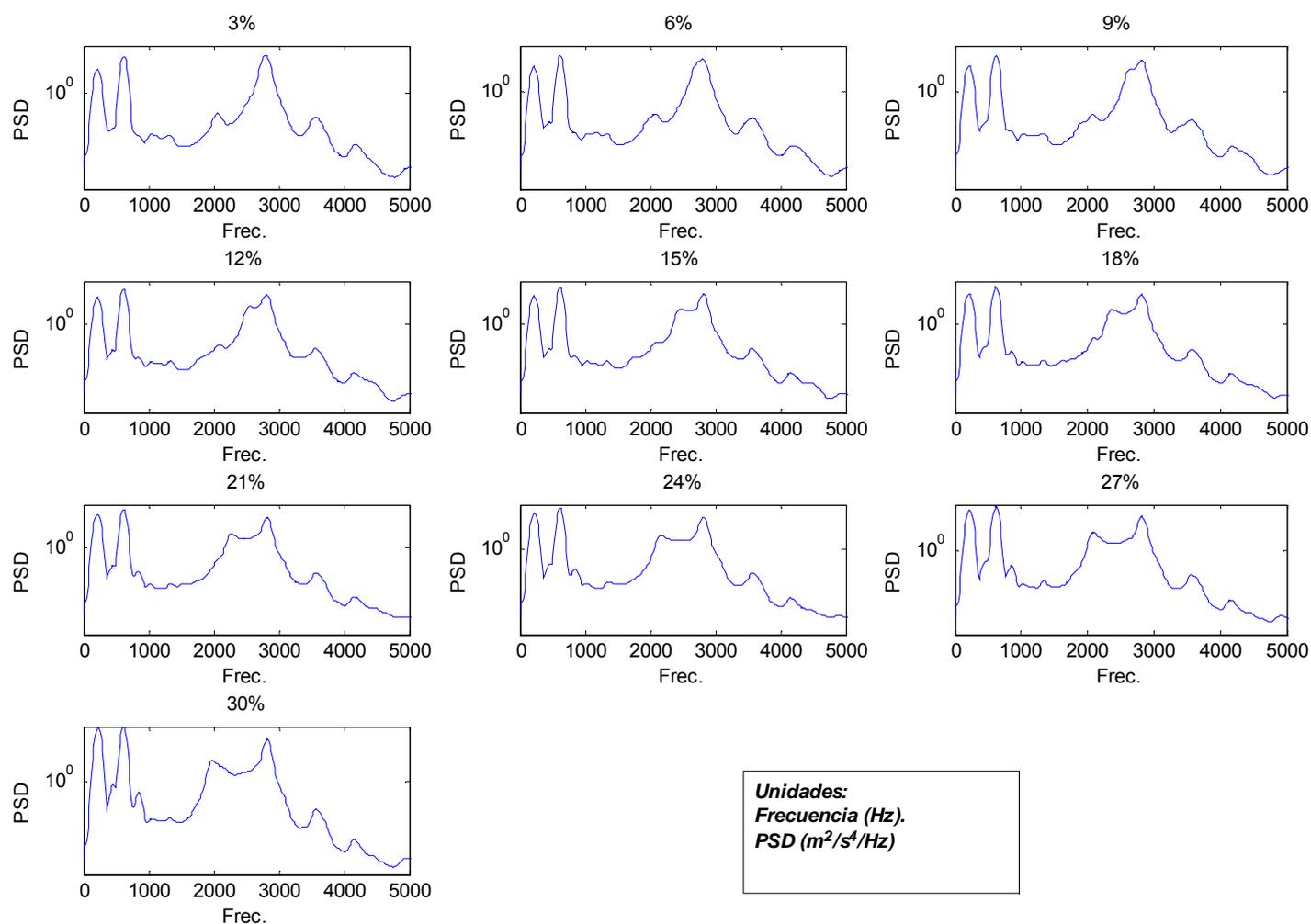


Figura 53. Espectro de potencia obtenido para la aceleración de la rueda (9) para distintos grados de degradación.

Por mera comparación en la evolución del espectro de potencia PSD de la figura 53 ante los distintos grados de degradación, puede apreciarse que la mayor diferencia estriba en el rango de frecuencias comprendido entre 1 y 2 kHz. No obstante, teniendo en cuenta la correlación del modelo numérico con el experimental, solamente se toma en consideración el rango 0-1 kHz, siendo la detección de defectos en dicho intervalo de frecuencias bastante compleja, salvo en los casos más extremos.

4.2.1.3. Detección y diagnóstico

La fase previa a la cuantificación consiste en la detección y localización del modo de fallo, tal y como se describe en el capítulo 3, para posteriormente una vez detectado el fallo proceder a estimar la severidad del daño.

Por ello, en este módulo se va a proceder a realizar un estudio similar al desarrollado para los casos estudiados anteriormente (experimento numérico y estación de bombeo). Además, dicho estudio proporcionará información relevante, como posteriormente se verá, en lo referente a la extracción de rasgos sensibles al daño.

Por tanto el procedimiento de detección es el descrito en el punto 3.3.2.3 según el esquema de bloques de la figura 3.3.2 y es totalmente análogo al implementado en el experimento numérico o en la estación de bombeo.

Este algoritmo de detección se comprobó mediante señales derivadas (aceleraciones tangenciales de la rueda nueva) del modelo de parámetros concentrados descrito en el punto 4.2.1.1.

La simulación se realizó durante 110 segundos obteniéndose un total de 465 señales “sanas” o, lo que es lo mismo, se generaron 465 patrones en condiciones “sanas”.

a) Preprocesamiento de señal:

En esta etapa se seleccionó el tipo de ondícula más apropiado para la detección así como el nivel de descomposición requerido.

En este caso se emplearon ondículas de la familia Daubechies con tres momentos nulos (db3) y para la descomposición como en el resto de casos se empleo el algoritmo de transformación rápida (FWT) desarrollado por Mallat [37].

En lo referente al análisis multiresolución se emplearon 8 niveles de descomposición aunque únicamente se emplearon los coeficientes correspondientes a las subseñales de detalles de los niveles 5 a 8 y los coeficientes correspondientes a la subseñal de aproximación.

Esta selección de niveles fue necesaria dado el elevado número de coeficientes implicados en la transformación DWT.

Con objeto de validar esta reducción de datos, con los coeficientes seleccionados se procedió a reconstruir la señal original mediante la aplicación de la transformación inversa IDWT y se comparó dicha señal sintética con la señal original obteniéndose el espectro de potencia PSD indicado por la línea roja de la figura 54 donde se ha superpuesto la señal original en azul.

De dicha figura se puede apreciar que la selección de coeficientes recoge bien la información de los datos hasta 470 Hz aproximadamente, lo cual resulta muy interesante puesto que mantiene la información referente al primer y segundo armónicos de la frecuencia de engrane (figura 54)

Esta aproximación resulta muy adecuada de cara a los propósitos de cuantificación, puesto que los armónicos de la frecuencia de engrane serán muy sensibles a la degradación de la máquina.

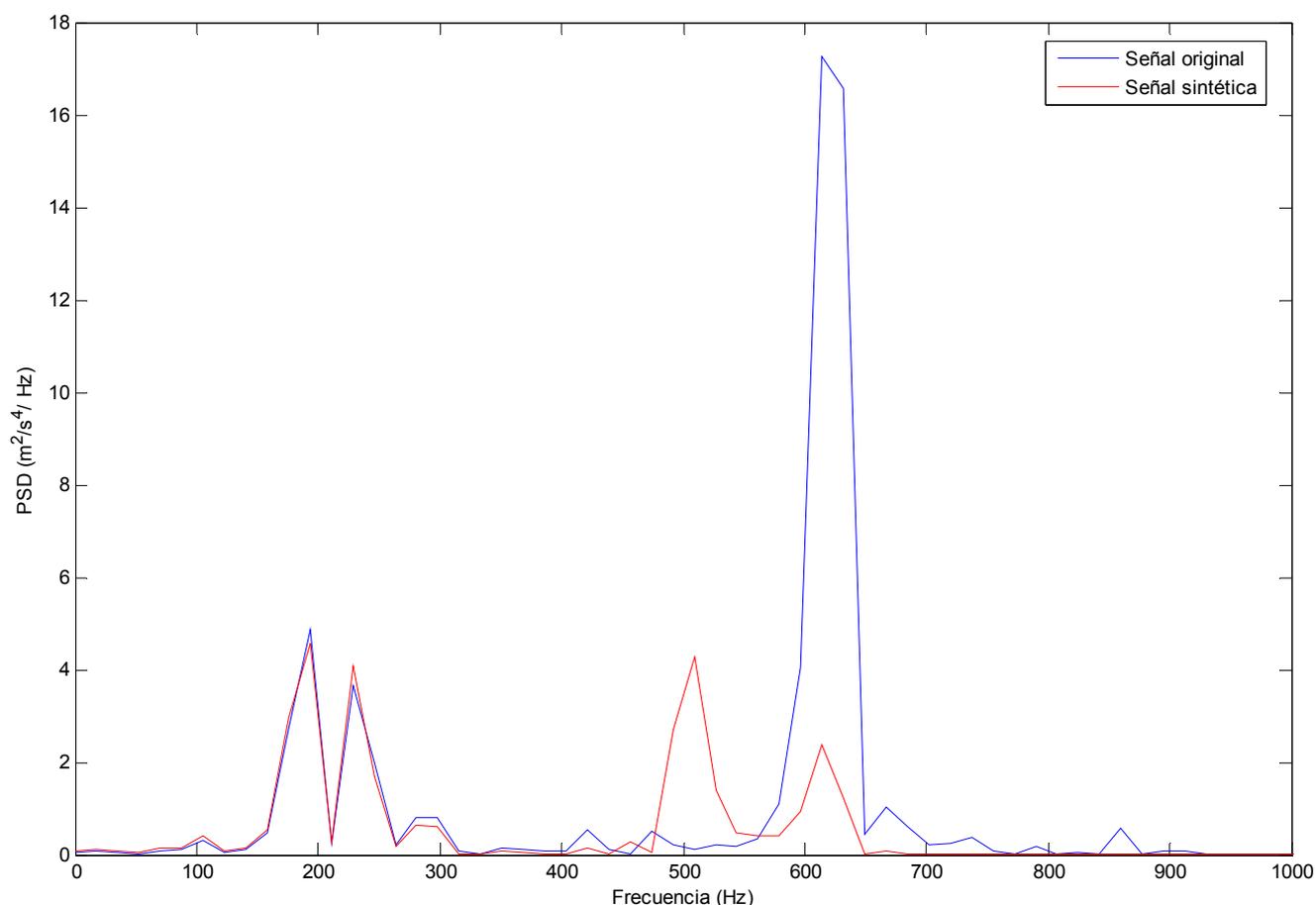


Figura 54. Espectro de potencia obtenido para la aceleración de la rueda (9) y espectro reconstruido a partir de los coeficientes DWT seleccionados

b) Entrenamiento

En este caso fueron necesarias cinco redes neuronales auto-asociativas para la fase de detección, cuatro para los coeficientes de los niveles de detalle elegidos y otra para los coeficientes de aproximación.

Hay que indicar que de las 465 señales “sanas” registradas, 300 fueron empleadas para entrenar a estas redes y 165 fueron utilizadas para determinar el índice de falsas alarmas generadas por dichas redes.

Para configurar las arquitecturas de las redes neuronales se empleó el mismo procedimiento que en los casos 4.1.1 y 4.1.2.

Finalmente las siguientes topologías para las cinco AANN fueron adoptadas:

-Nivel D5: 137x50x1x50x137

-Nivel D6: 71x25x1x25x71

-Nivel D7: 38x15x1x15x38

-Nivel D8: 21x5x1x5x21

-Nivel A8: 21x10x1x10x21

donde los subíndices 'D' y 'A' indican si los coeficientes pertenecen a las subseñales de detalle o de aproximación respectivamente.

Una vez entrenadas las redes, los patrones de entrenamiento se volvieron a introducir en las AANN para determinar los vectores residuales y, de esta forma, obtener los valores umbrales.

c) Reconocimiento

En esta fase, en primer lugar, se comprobaron los índices de falsas alarmas generados a partir de las 165 señales "sanas" que no se habían empleado en la fase de entrenamiento. Los resultados referentes a este parámetro se muestran en la tabla 7, donde además se indican las predicciones erróneas obtenidas para las 300 señales de entrenamiento.

NÚMERO DE PREDICCIONES FALSAS EFECTUADAS SOBRE SEÑALES "SANAS"						
NIVELES DE DESCOMPOSICIÓN	D5	D6	D7	D8	A8	NUMERO DE SEÑALES
ENTRENADAS	16	3	9	8	12	300
SIN ENTRENAMIENTO	8	2	4	2	6	165

Tabla 7. Número de falsas alarmas realizadas con las AANN.

De los resultados recopilados en la tabla 7 puede comprobarse la robustez de las predicciones efectuadas con las redes neuronales.

Por último, para comprobar la capacidad de detección del sistema de diagnóstico, se volvieron a generar 1050 nuevas señales con el modelo de parámetros concentrados, esta vez asociadas

a un patrón de fallo con distintos niveles de severidad. Estas señales se dividieron en grupos de 105 registros para cada nivel de fallo (3%, 6%, 9%, 12%, 15%, 18%, 21%, 24%, 27%, 30%), cuantificado este como el cambio porcentual en la rigidez de engrane frente a esta misma rigidez en condiciones “sanas”.

Los resultados obtenidos para cada estado de degradación, mostrados en la tabla 8, demuestran que hay un alto índice de detección para estados de degradación superiores al 18%.

NIVEL DE DESCOMPOSICIÓN	PREDICCIONES EFECTUADAS SOBRE SEÑALES DEFECTUOSAS									
	REDUCCIÓN PORCENTUAL DE RIGIDEZ									
	3%	6%	9%	12%	15%	18%	21%	24%	27%	30%
D5	1	8	11	18	31	47	66	86	101	104
D6	2	5	6	18	25	50	80	97	104	104
D7	6	4	5	6	6	13	26	20	28	41
D8	9	8	9	19	7	14	30	29	38	52
A8	13	13	12	19	16	18	29	37	34	44
NUMERO DE SEÑALES	105	105	105	105	105	105	105	105	105	105

Tabla 8. Número de detecciones realizadas con las AANN para cada nivel de descomposición y estado de degradación.

4.2.1.4. Cuantificación

Una vez detectado el daño el siguiente paso consiste en determinar el grado de severidad del modo de fallo asociado.

Este estudio de cuantificación resulta substancialmente útil para realizar tareas de control de degradación una vez que el daño se ha detectado en el sistema, especialmente para establecer planes de mantenimiento o determinar un criterio de parada en la máquina rotativa.

La cuantificación de daño se puede interpretar como un problema inverso de ingeniería en el que, a partir de la respuesta estructural del sistema (aceleraciones registradas), se ha de

pronosticar cuál es el cambio en el parámetro (en este caso la rigidez torsional) que provoca ese nivel en la medida de la respuesta observada.

Por tanto, dicho problema constituye un claro problema de clasificación, en el que a cada nivel de aceleración (efecto) se le asocia un determinado rango de rigidez torsional (causa); por tal motivo, en esta tesis, la cuantificación se abordó empleando redes neuronales supervisadas del tipo perceptrón multicapa (MLPNN) con un algoritmo de retropropagación.

En términos generales, el vector de entrada a la red fue proporcionado a través de la señal de aceleración y la salida de dicha red se formó con un único parámetro asociado a la rigidez torsional de engrane.

El éxito del sistema de cuantificación basado en este tipo de redes radica en obtener un determinado número de parámetros (a través de la señal registrada) que caractericen la respuesta dinámica del sistema y sean muy sensibles a las condiciones operativas y por tanto al estado de degradación. Esto, sin duda, constituye la tarea más complicada del procedimiento de cuantificación y es a lo que se refiere el estudio posterior.

a) Extracción de rasgos mediante empleo de coeficientes de la transformación DWT.

En la sección 4.2.1.3. se estudió cómo “comprimir” la información tiempo-frecuencia de la señal de aceleración, obtenida a través del modelo numérico de parámetros concentrados, mediante la transformación DWT, reduciéndose así la información contenida en el alto número de puntos de muestreo, necesarios para asegurar la estabilidad del algoritmo de integración.

Sin embargo, si se adoptan como parámetros de entrada a la red MLPNN los coeficientes de la transformación DWT correspondientes a los niveles de descomposición 5 a 8, el número de

rasgos de entrada (137 en el nivel 5, 71 en el nivel 6, 38 en el nivel 7, 21 en el nivel 8 de detalle y 21 en el nivel 8 de aproximación) continuaría siendo extremadamente alto.

Por ello, otra estrategia de reducción es claramente necesaria para evitar este efecto de excesiva dimensionalidad [48].

a.1) Estudio de sensibilidad

Con objeto de disminuir el número de parámetros en la capa de entrada de la red MLPNN a través de la información proporcionada mediante los coeficientes DWT, parece interesante realizar previamente un estudio de sensibilidad del sistema de detección del daño en función de la severidad del mismo. Esto permitirá determinar los niveles más sensibles al daño y por consiguiente los más susceptibles de ser empleados como entrada a la red MLPNN.

Si se analizan los resultados obtenidos en el sistema de detección, recopilados por grado de severidad en la tabla 8, parece claro que por encima de una degradación del 20% los niveles más sensibles son los niveles 5 y 6.

Por consiguiente, en primer lugar, sólo los coeficientes de estos niveles de descomposición (5 y 6) serán seleccionados, a pesar de que su número continúa siendo elevado.

A continuación se examinan, para los niveles elegidos, los vectores residuales o vectores de índices noales obtenidos a través de las redes AANN, con el propósito de poder realizar una discriminación de los coeficientes dentro de un mismo nivel (5 y 6) eligiendo aquellos que proporcionen mayores residuos.

Las figuras 55 y 56 muestran estos residuos junto con los valores umbrales (línea roja), los cuales, como ya se explicó en capítulos anteriores, son valores límites que indican la presencia de anomalías.

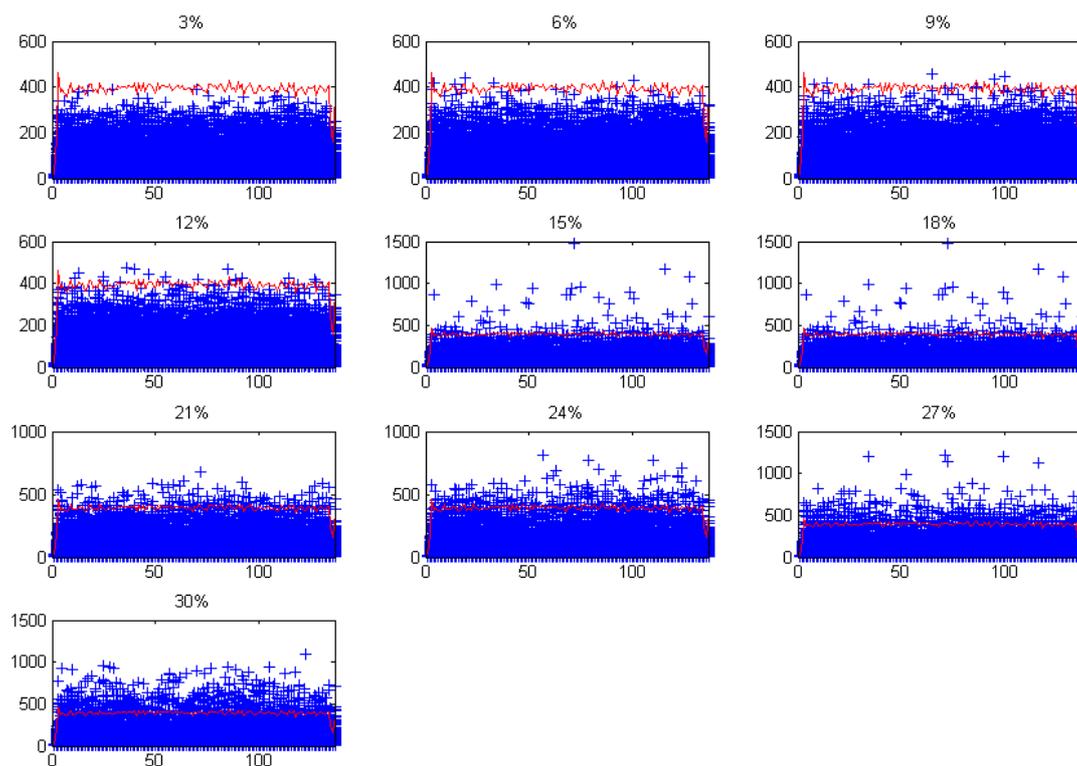


Figura 55. Representación de los residuos generados para el nivel D5 y por todos los patrones degradados para cada estado de degradación.

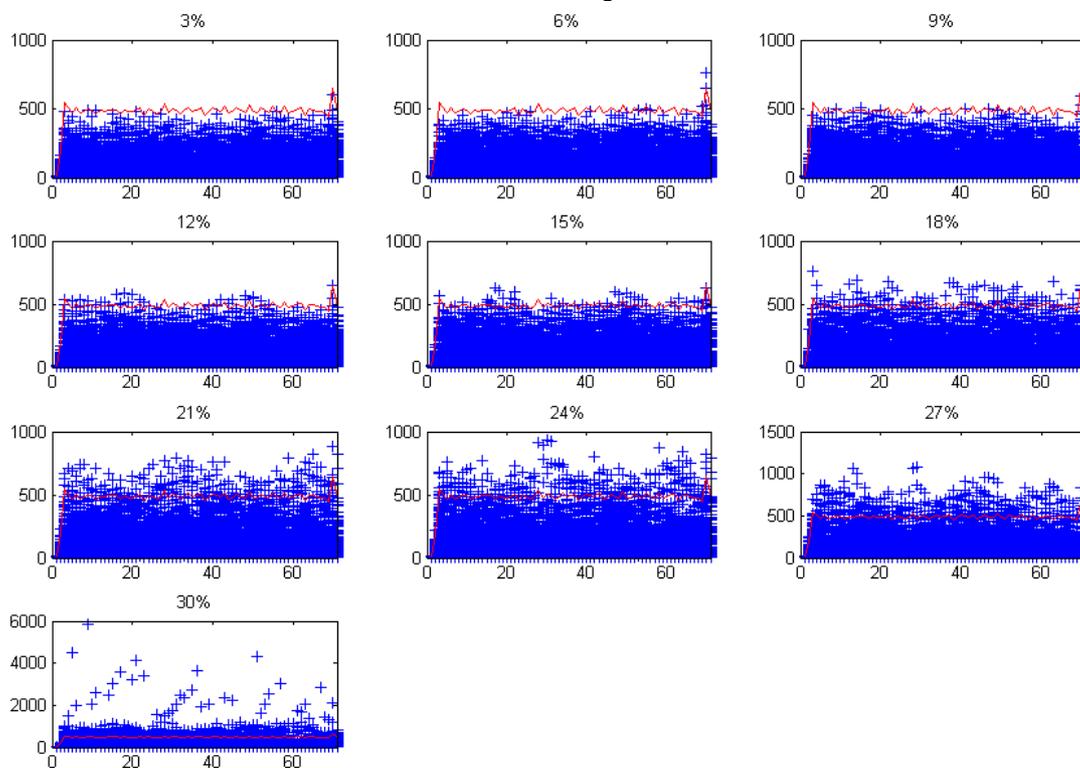


Figura 56. Representación de los residuos generados para el nivel D6 y por todos los patrones degradados para cada estado de degradación.

La figura 55 muestra los residuos obtenidos para el nivel de descomposición 5, para los distintos grados de severidad del daño, mientras que la figura 56 ilustra los residuos correspondientes al nivel 6. Para cada figura, en ordenadas se indican los niveles alcanzados con los residuos y en abcisas los coeficientes DWT.

De la observación de estas figuras, parece evidente que es inviable realizar cualquier tarea de discriminación sobre los coeficientes, por consiguiente es necesario recurrir a otra estrategia de reducción.

Como otra alternativa se ha considerado la evaluación de parámetros estadísticos a partir de los coeficientes DWT correspondientes a los niveles 5 y 6 de descomposición.

a.2) Características estadísticas de la señal registrada

Tal y como se describió en el capítulo 3 con detalle, el hecho de introducir una perturbación en una señal transformada mediante DWT implica que alguno de los coeficientes de dicha transformación destaque sobre el resto (apartado 3.2.4.3 de la tesis), el nivel de descomposición más afectado dependerá del número de momentos nulos de la familia de ondículas empleado y también en la regularidad de la señal perturbada, afectando a aquellos coeficientes situados en lo que se llama cono de influencia.

La magnitud del pico alcanzado en el coeficiente más afectado y la anchura del cono de influencia serán muy sensibles a la regularidad de la perturbación introducida.

Por consiguiente cuando se presenta una perturbación, es de esperar que ciertas medidas estadísticas como la curtosis y la desviación estándar establecidas para la distribución de coeficientes DWT de diferentes señales temporales se vean seriamente afectadas.

Esto es debido a que la curtosis tiende a proporcionar una indicación del pico observado en una distribución, por lo que es previsible que este valor aumente cuando la haga el pico en la

perturbación., Por otra parte, también es lógico suponer un incremento en la anchura del cono de influencia cuando aumenta el grado de severidad del daño, lo que a su vez incrementaría la desviación estándar en la población de coeficientes DWT observados.

En las figuras 57 y 58 se muestran el valor máximo, el valor mínimo, el promedio, la desviación estándar y el valor curtosis determinados para cuatro casos de daño correspondientes a una reducción de la rigidez torsional de 18%, 21%, 24% y 27%, respectivamente. Estos parámetros estadísticos se han obtenido para una población compuesta por los coeficientes DWT extraídos para cada patrón registrado (vuelta de máquina), teniéndose un total de 416 patrones degradados.

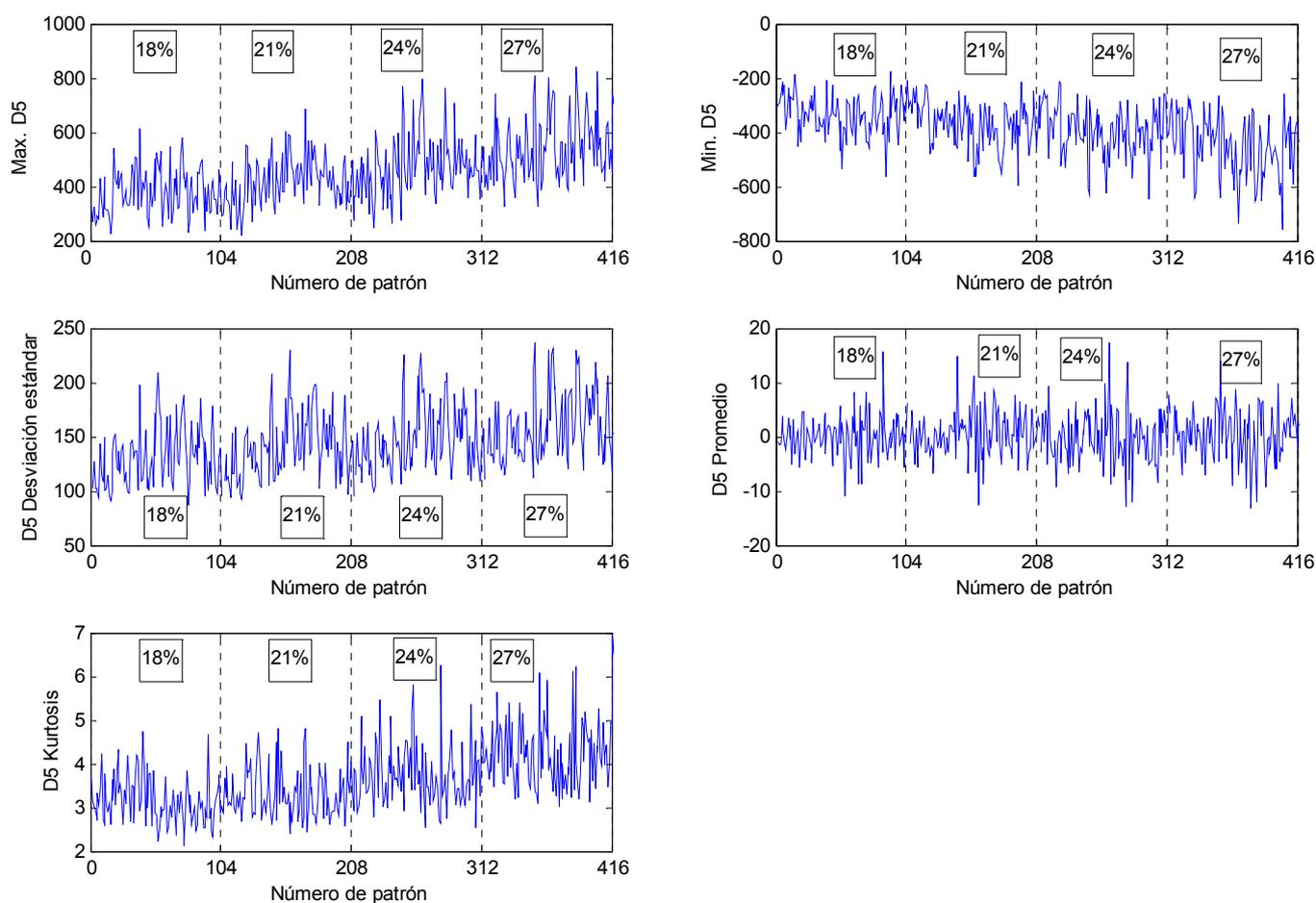


Figura 57. Estadísticas obtenidas para el nivel de descomposición D5, sobre los distintos patrones degradados.

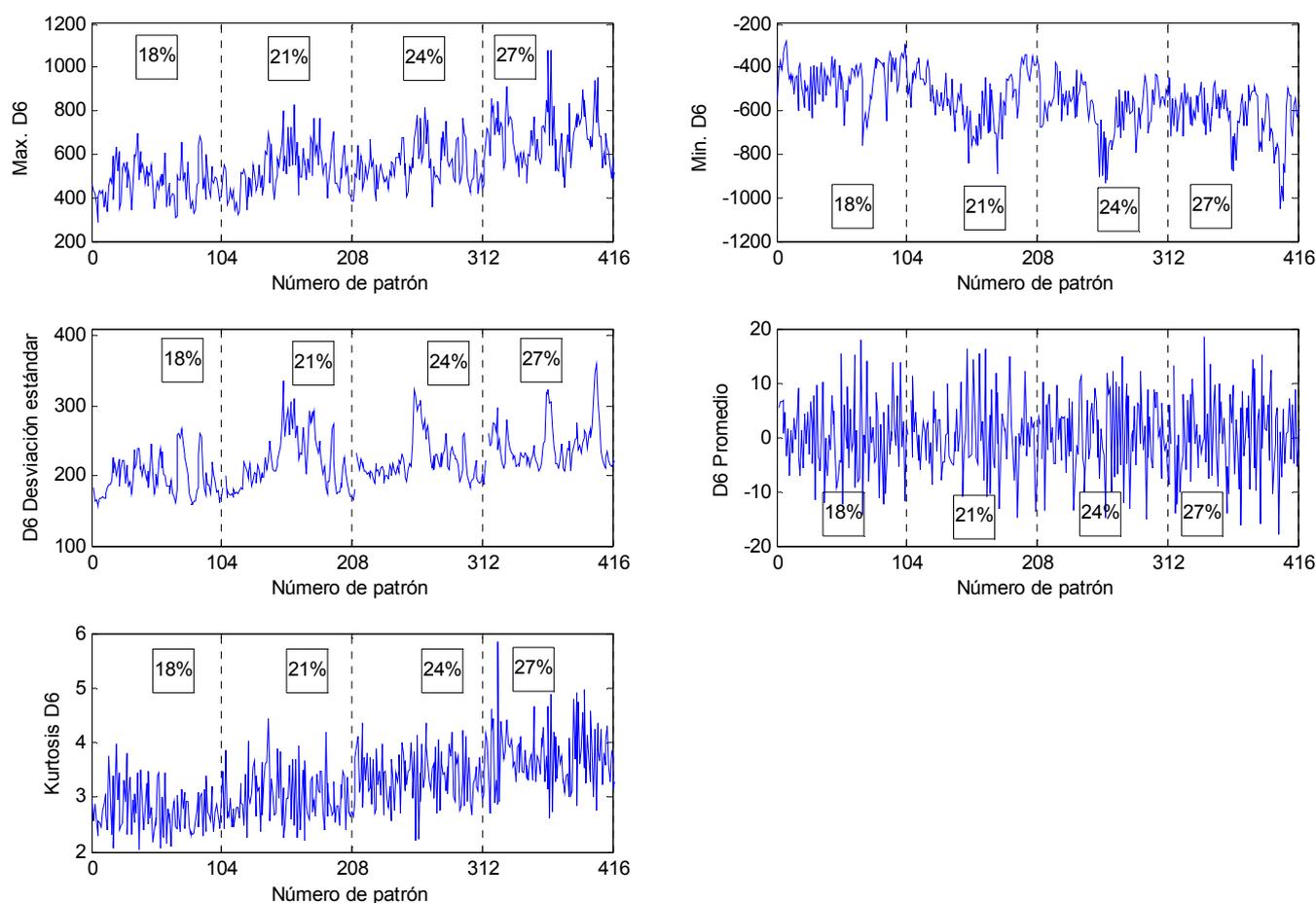


Figura 58. Estadísticas obtenidas para el nivel de descomposición D6, sobre los distintos patrones degradados.

De la observación de estas figuras se puede apreciar que tanto la curtosis como la desviación estándar son sensibles a un incremento en el daño, tal y como se predijo anteriormente, mientras que los valores promedios no son sensibles a la severidad del daño.

Por tanto, como parámetros de entrada a la red se han elegido los valores máximo y mínimo de los coeficientes DWT, además de la curtosis y desviación estándar. Estos cuatro parámetros se han seleccionado para los niveles 5 y 6 de descomposición resultando en un total de 8 parámetros de entrada a la red MLPNN.

Resumiendo, el modulo de cuantificación (se supone que previamente el módulo de detección ha detectado alguna anomalía) funciona extrayendo los parámetros que posteriormente

constituirán el vector de entradas a la red MLPNN de la siguiente forma: En primer lugar se seleccionan los coeficientes DWT correspondientes a los niveles 5 y 6 de la señal de descomposición obtenidos previa transformación de los registros temporales; esto se efectúa por vuelta de maquina. Posteriormente, a partir de estos coeficientes, se extraen los parámetros estadísticos comentados anteriormente y, finalmente, se obtiene el vector de entrada a la red neuronal.

b) Empleo de redes MLP para cuantificación del daño.

b.1) Arquitectura de la red neuronal y entrenamiento.

Hasta el momento, solamente se han ajustado las capas de entrada y salida de la red neuronal, tomando ocho parámetros estadísticos como entrada a la red y un solo parámetro de salida coincidente con la reducción en la rigidez a torsión del par de engrane.

A continuación se indicará cómo se configuran el resto de capas ya que el número de capas ocultas y el número de neuronas presentes en ellas afectará a la capacidad de generalización de la red (sección 3.3.2 de la tesis).

Se puede demostrar que el número de capas ocultas en un problema de regresión o clasificación (como es este caso) debe de ser mayor que dos; por tanto, éste es el mínimo número de capas que debe ser utilizado como punto de partida ([57], [58]).

Partiendo de esta restricción una metodología totalmente heurística se ha aplicado con el objetivo de configurar las denominadas capas “ocultas” de la red.

El entrenamiento de la red MLPNN implica ajustar los valores de los pesos sinápticos mediante la minimización de una función de error definida entre la salida estimada por la red y el

correspondiente valor objetivo. El algoritmo empleado para tal propósito es el RP [53] y la función a minimizar consiste en la desviación cuadrática media del valor objetivo frente a la salida de la red neuronal.

El conjunto de datos necesario para el entrenamiento de la red proviene del modelo numérico de parámetros concentrados, reduciendo la rigidez a torsión del par de engrane en función de la severidad del daño.

La interacción del modelo de parámetros concentrados, como generador de patrones, con la red neuronal, puede entenderse mejor al analizar el diagrama de bloques indicado en la figura 59.

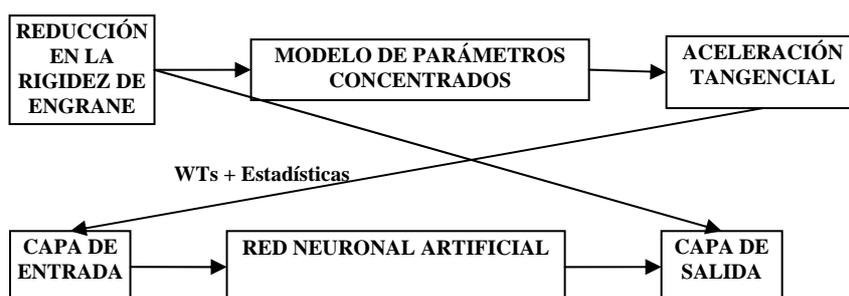


Figura 59. Esquema de generación de patrones como entrada a la red MLPNN.

Una vez se ha completado el entrenamiento de la red, se generan nuevos patrones con un grado de daño diferente al empleado en la fase de entrenamiento con idea de validar y ver cómo generaliza la red.

b.2) Resultados de la cuantificación.

La configuración final de la red corresponde a una arquitectura de neuronas 8x10x4x1, siendo 10 y 4 el número de neuronas en las capas ocultas.

En la fase de entrenamiento, se emplearon 416 patrones obtenidos al introducir en el modelo numérico 4 tipos de daño (18%, 21%, 24% y 27%); además, para ser congruentes con la realidad experimental, se añadió un ruido blanco de las mismas características que el indicado en el punto 4.2.1 de esta tesis. Esto permite obtener diferentes patrones para el mismo tipo de severidad de fallo. Es conveniente reseñar que todos los valores de daño empleados tanto en la fase de entrenamiento como en la de validación son superiores al 18%, ya que éste es el mínimo valor que puede ser detectado por el sistema de identificación.

Los errores de entrenamiento entre los pronósticos de la red y los valores objetivo fueron medidos como porcentaje de la rigidez torsional promediada, dichos errores se muestran en la figura 60 y en la tabla 9, agrupados para cada nivel de fallo.

Nivel de daño	Error promediado
18%	1.5345
21%	1.1033
24%	1.2690
27%	0.9021

Tabla 9. Errores promedio cometidos con las estimaciones en la red para cada nivel de degradación. Valores obtenidos para patrones entrenados.

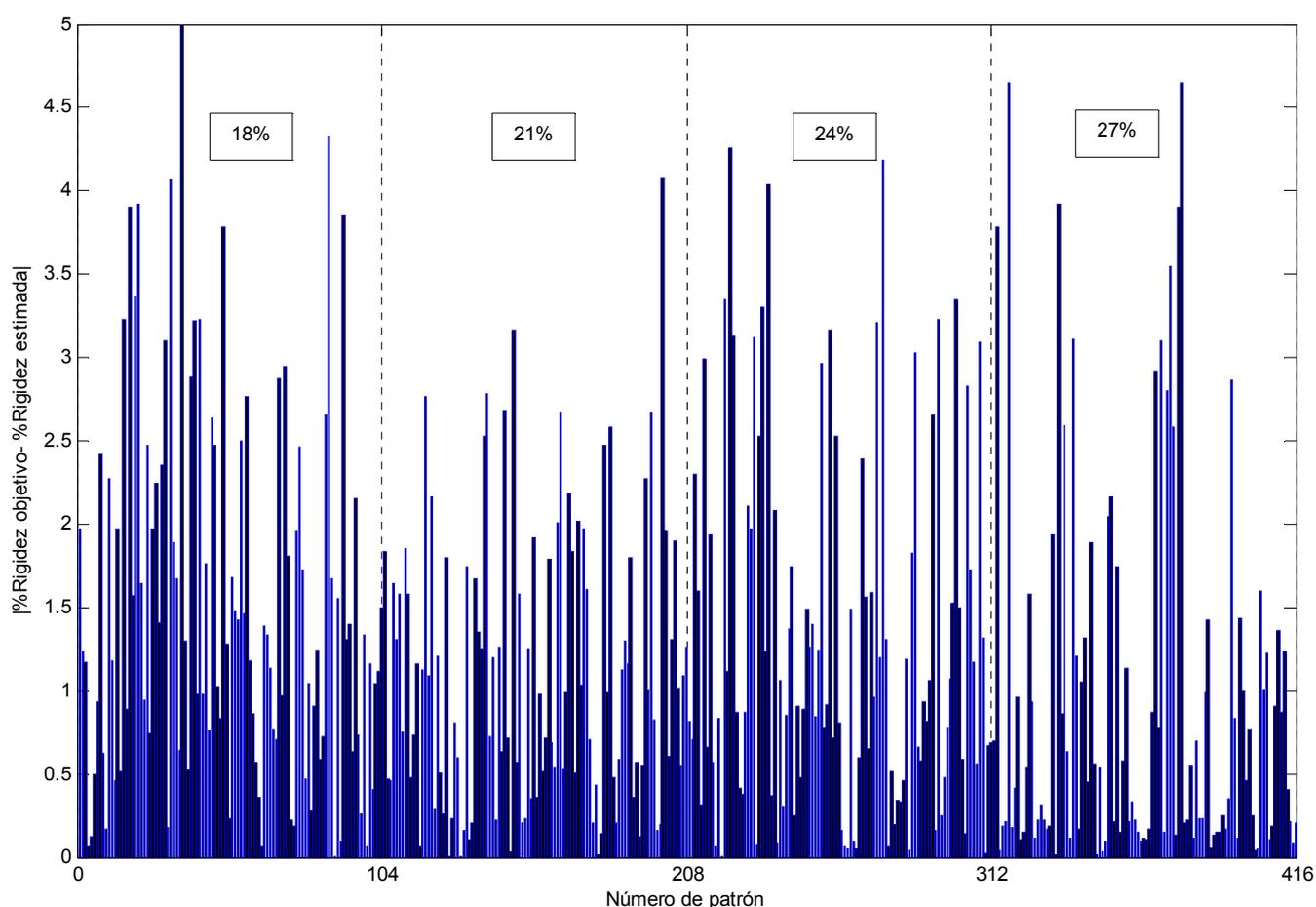


Figura 60. Barras de errores cometidos con las estimaciones en la red para cada nivel de degradación. Valores obtenidos para patrones entrenados

El error promediado para todos los patrones de entrenamiento es 1.2 que es suficientemente pequeño comparado con las variaciones de la rigidez torsionales estudiadas.

Posteriormente, en la fase de validación de la red, se comprobó su funcionalidad, empleando 164 nuevos patrones, diferentes de los empleados en la fase de entrenamiento y correspondientes a los niveles de daño de 18.6%, 22.8%, 25.2% y 28.2%. Los resultados obtenidos, agrupados por nivel de daño, se muestran en la tabla 10.

Asimismo, los errores cometidos con cada patrón se muestran en la figura 61

Nivel de daño	Error promediado
18.6%	1.4250
22.8%	2.1627
25.2%	2.0185
28.2%	1.5686

Tabla 10. Errores promedio cometidos con las estimaciones en la red para cada nivel de degradación. Valores obtenidos para patrones no entrenados.

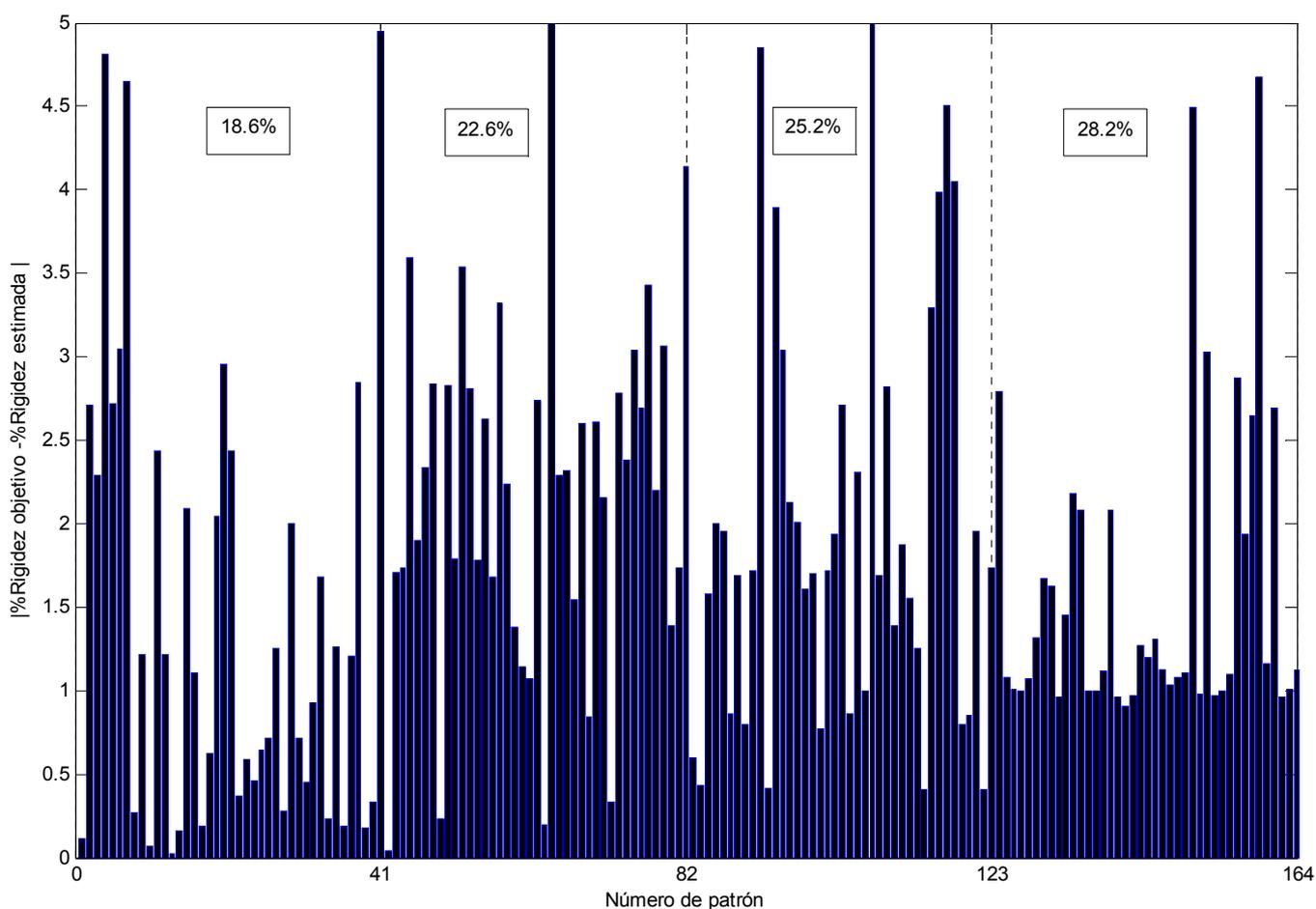


Figura 61. Barras de errores cometidos con las estimaciones en la red para cada nivel de degradación. Valores obtenidos para patrones no entrenados

En este caso, el error total promediado con los 164 patrones es de 1.8, lo cual representa una ligera diferencia respecto del observado con los patrones de entrenamiento por lo que puede

concluirse que la red realiza una buena generalización y puede ser empleada con garantías como herramienta de cuantificación.

Por otra parte, a pesar de que la red pueda ofrecer pronósticos por cada vuelta del eje de la máquina, parece claro de la figura 61 y de la tabla 10, que los pronósticos de la red deberían de ofrecerse al menos promediados cada 40 vueltas con el objetivo de conseguir predicciones más exactas, a causa de la inestabilidad producida por el ruido añadido.

CAPÍTULO 5

CONCLUSIONES

A partir de los resultados obtenidos en las distintas aplicaciones llevadas a cabo con el sistema de diagnóstico desarrollado, y tomando como referencia los objetivos fijados inicialmente y el estado del arte de la tecnología se pueden establecer las principales aportaciones generadas con este trabajo. Asimismo, las limitaciones encontradas durante el desarrollo de este trabajo servirán como punto de partida para futuros trabajos que completen y aumenten las posibilidades del sistema de diagnóstico propuesto.

Puesto que el sistema de detección se ha desarrollado en dos módulos o etapas, una correspondiente a la fase de detección y diagnóstico y otra correspondiente a la fase de cuantificación, las contribuciones y limitaciones se referirán a cada una de estas etapas.

5.1. APORTACIONES

FASE DE DETECCIÓN Y DIAGNOSIS

En esta etapa se ha desarrollado un método de detección y diagnóstico en tiempo real por combinación de las capacidades de las redes neuronales autoasociativas y de la transformación con ondículas discreta.

Los procedimientos de reconocimiento de patrones basados en redes neuronales se han utilizado para comparar los coeficientes de la transformación DWT obtenidos a partir de registros correspondientes a condiciones de funcionamiento normales y anómalas. El residuo

constituido por la diferencia entre coeficientes de salida y entrada a la red se emplea como síntoma de detección de fallo en maquinaria.

Además, el empleo de la transformación DWT permite detectar los rangos de tiempo-frecuencia en los que el fallo se presenta y, por tanto, el procedimiento propuesto puede ser también útil como herramienta de localización del daño (etapa de reducción y dientes afectados, rodamiento dañado, etc.). Sin embargo, en los casos donde los armónicos afectados se solapan (pasos de reducción con relaciones de transmisión múltiplos de otros pasos) no se podrán realizar estas inferencias en lo referente a la localización del daño.

Otra aportación del método propuesto es que la transformación DWT directamente se aplica sobre las señales originarias y no por procesamiento del TSA cuya evaluación puede ser compleja. Como no es necesario calcular el promediado sobre muchas revoluciones no se pierde la información relevante, además las predicciones se efectúan por cada revolución lo cual las hace mucho más rápidas y requiere por otro lado menos espacio de memoria para almacenar las señales registradas (necesario para realizar un promedio).

Por otro lado, el mismo sistema permite realizar predicciones bajo distintas condiciones de operación discretas, sin pérdida significativa de eficiencia en los pronósticos y sin necesidad de modificar parámetros cuando cambien las condiciones de operación o de emplear otro sistema especializado en el nuevo punto de funcionamiento. Esto contribuye, sin duda, a abaratar la implementación del sistema desarrollado en esta tesis.

Por último, hay que indicar que el sistema realiza automáticamente la diagnosis del estado de salud de la maquinaria, evitando, de esta manera, la necesidad de que estas anomalías sean analizadas por un experto. Esto representa una mejora significativa sobre los sistemas actuales de diagnosis.

CUANTIFICACIÓN

Esta etapa complementa a la anterior y sirve para incrementar la capacidad del sistema desarrollado ya que permite determinar el grado de severidad del daño.

En esta fase se emplea también parte de la información (estadísticas) derivada de los coeficientes proporcionados por la transformación DWT mediante su introducción en una red neuronal supervisada (red perceptrón multicapa) que permite determinar el nivel de daño alcanzado.

En el procedimiento no se emplean todos los coeficientes DWT, sino sólo aquellos más sensibles al tipo de daño registrado; para ello, se ha realizado un estudio de sensibilidad para determinar primeramente los niveles de descomposición más sensibles y, posteriormente, seleccionar los coeficientes.

Como parámetro de salida de la red neuronal se emplea el cambio en la rigidez torsional asociada al par de engranes afectados. Este valor se halla directamente relacionado con el tamaño de grieta y, por consiguiente, permite un seguimiento de la evolución en el crecimiento de grieta.

Como principal aportación, este módulo de cuantificación permite realizar las tareas de reparación de una forma programada; además, posibilita la integración de un sistema de valoración de la condición de funcionamiento con el que es factible llevar a cabo una labor de pronóstico para el mantenimiento predictivo de la máquina, permitiendo extender su vida operativa.

5.2. DESARROLLO FUTURO

FASE DE DETECCIÓN Y DIAGNOSIS

De los resultados obtenidos en el capítulo 4.1.2, se pudo comprobar la sensibilidad del sistema a las localizaciones de los transductores sobre la carcasa de la maquinaria ya que ésta actúa a modo de filtro cuya función de transferencia es totalmente dependiente no sólo por la posición espacial de los transductores sino también por las condiciones operativas; por este motivo, sería muy interesante desarrollar un estudio de optimización que en función de las condiciones de funcionamiento de la máquina y de las características tiempo-frecuencia del fallo a detectar permitiese establecer las mejores localizaciones posibles.

Por otra parte, dado que las predicciones sobre la localización del fallo no se realizan en tiempo real, sería interesante considerar un sistema automatizado que en función de los residuos afectados determinara el componente dañado.

Hay que indicar que diferentes tipos de fallo (desgaste, oxidación, picadura, etc.), bien de forma aislada o combinada, pueden dar lugar al mismo patrón de señal anómala; por tal motivo, sería interesante realizar también un estudio de clasificación del tipo de fallo previamente a la predicción de su grado de severidad.

FASE DE CUANTIFICACIÓN

A pesar de que el sistema de diagnosis permite realizar predicciones del nivel de daño en cada vuelta, éstas son sensibles al ruido introducido en el sistema y mejoran mucho cuando se promedian en un determinado número de vueltas (en el caso estudiado en el capítulo 4.2 el promediado se efectuaba cada 40 vueltas aproximadamente). En este sentido, se produce una

limitación del método cuando la velocidad de crecimiento de grieta supera a la velocidad con la que se efectúen las predicciones. Este fenómeno ocurrirá cuando se produzcan cambios notables en la pendiente que determine la velocidad de crecimiento de grieta [70]. En este caso, el fallo no podrá ser controlado y las predicciones de severidad carecerán de sentido, por este motivo sería interesante trabajar para mejorar las velocidades de pronóstico (cuantificación) en el sistema.

BIBLIOGRAFÍA

- [1] D. G. Astringe. **Helicopter Transmissions-Design for Safety and Reliability**. Inst. Mech. Eng. Proc., Pt. G-J Aerosp. Eng, Vol. 203, N° G2, pp 123-138, 1989.
- [2] R. L. Campbell, C. S. Byington, M. S. Lebold. **Generation of Health Usage Monitoring Systems Diagnostic Estimate Using Translational Data**. Proceedings of the 13th International Congress on Condition Monitoring and Diagnostic Engineering Management, Houston, Texas, December 2000.
- [3] P. J. Dempsey. **A comparison of Vibration and Oil Debris Gear Damage Detection Method Applied to Pitting Damage**. NASA TM-210371, December 2000.
- [4] P.D. McFadden, J. Wang. **Analysis of gear vibration signature by the weighted Wigner-Ville distribution**. Institute of Mechanical Engineers International Conference on Vibrations in Rotating machinery 387-393, 1992.
- [5] H. I. Choi, W. J. Williams. **Improved time-frequency representation of multicomponent signals using exponential kernels**. IEEE Transactions on Acoustics, Speech and Signal Processing 37, 862-871, 1989.
- [6] Normativa **ANSI/AGMA 1010-E95**.
- [7] Normativa **ISO 15243:2004 (E)**.
- [8] Wilson Q. Wang, Fathy Ismail, M. Farid Golnaraghi. **Assesment of gear damage monitoring techniques using vibration measurements**. Mechanical Systems and Signal Processing, 15(5), 905-922, 2001.
- [9] R. H. Lyon. **Machinery noise and diagnostics**. Butterworths, Boston, 1987.
- [10] G. A. Fontaine, E. E. E. Frietman, R. P.W. Duin. **Preventive and predictive maintenance using neural networks**. J. Microelectric Systems integration, Volume 4, N°2, 1996.
- [11] J. Derek Smith. **Gear noise and vibration**. Marcel Dekker. New York.2003.
- [12] D. Brie, M. Tomczak, H. oehlmann, A. Richard. **Gear crack detection by adaptative amplitude and phase demodulation**. Mechanical Systems and Signal processing, 11(1):149-167, 1997.
- [13] P. D. McFadden. **Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration**. Journal of Vibration, Acoustics, Stress and Reliability in Design 108, 165-170. 1986.
- [14] P. D. McFadden and J. D. Smith. **Model for the vibration produced by a single point defect in a rolling element bearing**. Mechanical systems and Signal processing, 96(1): 69-82, 1984.
- [15] C. Cempel **Vibroacoustic condition monitoring**. Ellis Horwood. Ltd 1991.
- [16] G. Dalpiaz, A. Rivola, R. Rubini. **Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears**. Academic Press, 2000.
- [17] S. Braun. **Mechanical Signature Analysis. Theory and Applications**. New York: Academic Press, 1986.

- [18] P. D. McFadden. **A revised model for the extraction of periodic waveforms by time domain averaging**. Mech. Syst. Signal Proc. 1(1), 83-95, 1987.
- [19] R. B. Randall. **A new method of modelling gear faults**. Journal of Mechanical Design 104, 259-267, 1982.
- [20] P. D. McFadden. **Examination of a technique for the early detection of failure in gears by signal processing of the time domain average of the meshing vibration**. Mechanical Systems and Signal Processing 1, 173-183, 1987.
- [21] P. D. McFadden, J. Wang. **Decomposition of gear motion signals and its application to gearbox diagnostics**. ASME Journal of Vibration and Acoustics 117, 363-369, 1995.
- [22] D. Gabor. **Theory of communication**. IEEE Journal 93, 429-457, 1946.
- [23] I. Daubechies. **Orthonormal bases of compactly supported wavelets**. Comm. Pure and applied maths XLI, 909-906, 1988.
- [24] D. J. Chen, W. J. Wang. **Pattern Classification of wavelet map using multi-layer perceptron neural network for gear fault detection**. Mechanical Systems and Signal Processing, 2002.
- [25] W. J. Wang, P. D. McFadden. **Application of wavelets to gearbox vibration signals for fault detection**. Journal of Sound Vibration 192, 927-939, 1996.
- [26] D. Boulahbal, M. Farid Golnaghari, F. Ismail. **Amplitude and phase wavelet maps for the detection of cracks in geared systems**. Mechanical Systems and Signal Processing 13, 423-436, 1999.
- [27] J. Sanz, R. Perera, C. Huerta. **Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms**. Journal of Sound and Vibration 302, 981-999, 2007.
- [28] R. E. Douka, S. Loutridis, A. Trochidis. **Crack identification in beams using wavelet analysis**. International Journal of Solids and Structures 40, 3557-3569, 2003.
- [29] A.V. Ovanesova, L.E. Suarez. **Applications of wavelet transforms to damage detection in frame structures**. Engineering Structures 26, 39-49, 2004.
- [30] J. Han, W. Ren, Z. Sun. **Wavelet packet based damage identification of beam structures**. International Journal of Solids and Structures 42, 6610-6627, 2005.
- [31] S.Q. Zhu, S.S. Lu, **Wavelet-based crack identification of bridge beam from operational deflection time history**. International Journal of Solids and Structures 43, 2299-2317, 2006.
- [32] Gurley, A. Kareem. **Application of wavelet transforms in earthquake, wind and ocean engineering**. Engineering Structures 21, 149-167, 1999.
- [33] Wilson Q. Wang, M. Farid Golnaraghi and Fathy Ismail. **Prognosis of machine health condition using neuro-fuzzy systems**. Mechanical Systems and Signal Processing, Volume 18, Issue 4, Pages 813-831, July 2004.

- [34] B. A. Paya, I.I. Esat, M. N. M. Badi. **Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor.** Mechanical Systems and Signal Processing 11, 751-765, 1997.
- [35] B. S. Yang, T. Han, J. L. An. **ART-KOHONEN neural network for fault diagnosis of rotating machinery.** Mechanical Systems and Signal Processing 18, 645-657, 2004.
- [36] Samanta. **Gear fault detection using artificial neural networks and support vectors machines with genetic algorithms.** Mech. Syst. Signal Proc.18, 625-644, 2004.
- [37] S. Mallat. **A theory for multiresolution signal decomposition:the wavelet representation.** IEEE transactions on pattern analysis and machine intelligence, vol 11.Nº 7, july 1989
- [38] G. Strang, T. Nguyen. **Wavelets and filter banks.** Wellesley. MA, Wellesley-Cambridge Press, 1996.
- [39] M. Holschneider and P. Tchamitchian. **Regularite locale de la fonction non-differentiable de Riemann in les ondelettes in 1989.** Lecture notes in Mathematics, Ed. New Cork: Springer-Verlag, 1989.
- [40] S. Jaffard. **Exposants de Holder en des points donnees et coefficients d'ondelettes.** Notes au Compte-Rendu de l'academie Des Scineces, France, vol.308, ser.I, pp,79-81,1989.
- [41] D. E. Newland. **Vibraciones aleatorias y análisis espectral.** Ed.AC 1983.
- [42] S. Mallat. **A wavelet tour of signal processing.** Academic Press, 1998.
- [43] S. Ramón y Cajal. **Texture of Nervous System of Man and Vertebrates.** Springer-Verlag, Viena, Nueva Cork, 1999.
- [44] S. Chen, S. A. Billings, and W. Luo. **Orthogonal least squares method and their application to non linear system identification.** International Journal of control 50 (5), 1873-1896. 1989
- [45] S. Chen, C. F. N. Cowan, and P. M. Grant. **Orthogonal least squares learning algorithm for radial basis function networks.** IEEE Transactions on Neural Networks 2, 302-309, 1991.
- [46] E. M. L. Beale. **A derivation of conjugate gradients. Numerical methods for nonlinear optimization.** London: Acadameic Press, 1972.
- [47] M. T. Hagan and M. Menhaj. **Training feed-forward networks with the Marquardt algorithm.** IEEE Transaction on Neural Networks, vol.5, Nº 6, 1999.
- [48] C. M. Bishop. **Neural Networks for pattern Recognition.** Oxford University Press, 1995.
- [49] P. D. Wasserman. **Advanced methods in neural computing.** New York: Van Nostrand Reinhold, 1993.
- [50] I. T. Jollife. **Principal Component Analysis.** New Cork:Springer-Verlag, 1986.
- [51] H. Boulard, and Y. Kamp. **Autoasotation by multilayer perceptrons and singular value decomposition.** 1988.

- [52] M. A. Kramer. **Nonlinear principal component analysis using auto associative neural networks**. AIChe Journal 37 (2), 233-243, 1991.
- [53] M. Riedmiller and H. Braun. **A direct method for faster backpropagation learning: The RPROP algorithm**. Proceedings of the IEEE International Conference on Neural Networks, 1993.
- [54] S. Geman, E. Bienestock, R. Doursat. **Neural networks and the bias/variance dilemma**. Neural Comput. 4 (1), 1-58, 1992.
- [55] Funahashi. **On the approximate realization of continuous mappings by neural networks**. 2 (3), 183-192, 1989.
- [56] Hetch-Nielsen. **Theory of the back-propagation neural-network**. In Proceedings of the International Joint Conference on Neural Networks, Volume 1, pp. 593-605. San Diego, CA:IEEE, 1989.
- [57] L. K. Jones. **Constructive approximations for neural networks by sigmoidal functions**. Proceedings of the IEEE 78, 1586-1589, 1990.
- [58] E. K. Blum, L. K. Li. **Approximation theory and feed forward networks**. Neural networks 4, 511-515, 1991.
- [59] A .Rytter. **Vibration based inspection of civil engineering structures**. Ph.D. Thesis, Department of Building Technology and Structural Engineering, University of Aalborg, Denmark 1993.
- [60] C. James Li, Hyungdae Lee. **Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics**. 2004.
- [61] Isa Yesilyurt, Fengshou Gu and Andrew D. Ball. **Gear tooth stiffness reduction measurement using modal analysis and its use in wear fault severity assessment of spur gears**. NDT & E International, Volume 36, Issue5, Pages357-372, July 2003.
- [62] Nevzat Ozguven H. and Houser D.R. **Mathematical Models Used in Gear Dynamics – a Review**. Journal of Sound and Vibration, Vol.121, N°3, 1988, pp. 383-411, 1988.
- [63] Q. Hu, Z. He, Z. Zhang, Y. Zi. **Fault diagnosis of rotating machinery based on improved wavelet package transform and SVM'S ensemble**. Mechanical Systems and Signal Processing 21, 688-705, 2007.
- [64] S. Watanabe. **Pattern recognition :Human and mechanical**. New Cork: wiley,1985.
- [65] A. Ypma, R. P. W. Ligteringen, E.E.E. Duin, E.E.E. Frietman. **Pump vibration datasets, Pattern recognition group**. Delft University of Technology, 1999.
- [66] G. Dalpiaz, A. Rivola, R. Rubini, **Dynamic modelling of gear systems for condition monitoring and diagnostics**. Proceedings of the Congress of Technical Diagnostics, Gdansk, Poland 2, 185-192, 1996.
- [67] G. Dalpiaz, A. Rivola, R. Rubini. **A kineto-elastodynamic model of a gear testing machine**, Proceedings of International Conference on Mechanical Transmissions and Mechanisms, MTM '97, Tianjin, China, 549-553, 1997.
- [68] G. Dalpiaz. **Early detection of fatigue cracks in gears by vibrations analysis techniques**. Österreichische Ingenieur und Architekten Zeitschrift (ÖIAZ) 135, 312-317, 1990.

BIBLIOGRAFÍA

- [69] P. J. Prince and J. R. Dormand. *High order embedded Runge-Kutta formulae*. Journal of Computational and Applied Mathematics 7, 67-75, 1981.
- [70] D. Broek. *The practical use of fracture mechanics*. Kluwer Academic Press, 1989.