

HAROLD ARMANDO ROMO ROMERO

APLICACIÓN DE WAVELETS EN EL ANÁLISIS Y LA
EXTRACCIÓN DE CARACTERÍSTICAS PARA LA
CLASIFICACIÓN DE SEÑALES EMGS

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Maestría en Ingeniería
Área: Electrónica y Telecomunicaciones

Popayán
2008

HAROLD ARMANDO ROMO ROMERO

APLICACIÓN DE WAVELETS EN EL ANÁLISIS Y LA
EXTRACCIÓN DE CARACTERÍSTICAS PARA LA
CLASIFICACIÓN DE SEÑALES EMGS

Tesis presentada a la Facultad de Ingeniería
Electrónica y Telecomunicaciones de la
Universidad del Cauca para la obtención del
Título de Magíster en Ingeniería

Área:

Electrónica y Telecomunicaciones

Director:

Pablo Emilio Jojúa Gómez

Popayán

2008

*Dedico este trabajo a
mi hijo Juan David
y a mi esposa Esmeralda,
quienes son merecedores
de todo mi esfuerzo, empeño
y superación personal
como ejemplo de vida.*

Agradecimientos

Agradezco al Ph.D Andrés Vivas, quien permitió abordar mi tesis de maestría como parte fundamental en el desarrollo del macroproyecto *“modelado, diseño y simulación de una mano virtual controlada a partir de señales electromiográficas”*, al Ph.D Pablo Jojóa, quien con su acompañamiento y dirección me permitió alcanzar las metas propuestas, y un agradecimiento especial al Ph.D. Kevin Englehart del Instituto de Ingeniería Biomédica de la Universidad de New Brunswick, quien me permitió utilizar sus bases de datos con los registros de las señales electromiográficas (EMGS).

Resumen

Las señales Electro-MioGráficas Superficiales (EMGS) obtenidas de manera no invasiva por medio de electrodos ubicados sobre la piel, constituyen una valiosa fuente de información de la actividad muscular, razón por la cual, estas señales tienen hoy en día gran aceptación para sus aplicaciones en el control de prótesis de mano.

El objetivo central de este trabajo es, mediante el análisis de señales EMGS con técnicas basadas en *wavelets*, extraer características que ayuden a mejorar el desempeño en la clasificación de señales EMGS.

El problema de clasificación de señales EMGS se resuelve mediante un adecuado procesamiento de señales en las fases de: extracción de características, reducción de dimensionalidad y de reconocimiento de patrones. Debido a la interdependencia de estas fases se requiere la implementación de una solución integrada. Esto es posible mediante una adecuada caracterización tiempo-frecuencia de las señales EMGS.

En este trabajo se muestra que la representación simultánea de una señal en los dominios tiempo-frecuencia, lograda mediante la *transformación wavelets*, permite extraer características con mayor capacidad discriminante para mejorar su clasificación. Sin embargo, la transformada wavelet produce una representación bidimensional de la señal con una dimensión comparable a la de su dominio original, por lo que se requiere reducir su dimensionalidad. Para esto, en este trabajo se utiliza la técnica de selección de características con función de costo basada en el *índice de separabilidad de clases* (CSI), junto con una adecuada selección de sub-bandas

para la *transformada wavelet discreta* (DWT) y un algoritmo de selección de las mejores bases para la *transformada de paquetes wavelets* (WPT).

Finalmente se muestra que en la representación de señales EMGS, la combinación de características temporales con características espectral basadas en *wavelets*, permite alcanzar un buen desempeño en la clasificación de señales EMGS. Para validar los resultados se utiliza como clasificadores: una *red neuronal probabilística* (PNN), una *máquina de soporte vectorial* (SVM) y un *perceptrón multicapas* (MLP).

Palabras Clave: Señales EMGS, Análisis Wavelets, Clasificación de patrones.

Abstract

Surface Electro-myographics Signals (EMGS), obtained in non-invasive way by electrodes located over the skin, constitute a valuable information source of the muscular activity. For this reason, these signals are widely accepted to develop the hand prosthesis control.

The main objective of this work is to improve the performance of an EMGS feature extraction system based on *wavelets*.

The problem of EMGS signal classification is solved by an appropriate signal processing in the following phases: feature extraction, dimensionality reduction and pattern recognition. Since these phases are interdependent, the implementation of an integrated solution is required. This signal processing is possible by means of an appropriate time-frequency characterization of the EMGS signals.

This work shows that the simultaneous representation of a signal in the time and frequency domains, achieved by a *wavelet transform*, allows a more discriminative feature extraction to improve the pattern classification. Since the wavelet transform gives a bi-dimensional representation whose length is the same as the input signal, a dimensionality reduction is required. Here is used a feature selection technique based on class separability index (CSI) with an appropriate sub-band selection, for the *discrete wavelet transform* (DWT), and a best basis selection algorithm, for the *wavelet packet transform* (WPT).

Finally, it is shown that the combination of temporal features with spectral features based on wavelets achieves a good performance for the EMGS signal classification.

To validate the results, classifiers based on probabilistic neural network (PNN), the support vector machine (SVM) and the multi-layer perceptron (MLP) are used.

Key words: EMGS Signals, Wavelet Analysis, Pattern Classification.

Contenido

	Pág.
Lista de Figuras	xvii
Lista de Tablas	xxi
Capítulo 1. Introducción	1
Capítulo 2. Generalidades sobre las Señales EMG	5
2.1 Electromiografía	5
2.1.1 La Neurona	5
2.1.2 Los Músculos	6
2.1.3 La Unidad Motora	7
2.1.4 El Potencial de Acción	7
2.1.5 El Potencial de Acción de la Unidad Motora	7
2.2 Generación de Señales EMG	8
2.3 Detección de Señales EMG	9
2.4 Las Señales EMG y el Control de Prótesis de Mano	12
Capítulo 3. Representación y Clasificación de Señales EMGS	19
3.1 Extracción de Características	20
3.1.1 Análisis Temporal	22
3.1.2 Análisis Espectral de Fourier	24
3.1.3 Análisis Espectral con Wavelets	25
3.1.4 Extracción de características basadas en wavelets	34

3.2	Reducción de Dimensionalidad	37
3.2.1	Selección de características	38
3.2.2	Proyección de Características	41
3.3	Clasificación de Patrones	43
3.3.1	El Clasificador MLP	44
3.3.2	El Clasificador PNN	46
3.3.3	El Clasificador SVM	48
Capítulo 4.	Extracción de Características DT, DWT y WPT	53
4.1	Descripciones Preliminares	54
4.1.1	Señales EMGS	54
4.1.2	Preprocesamiento	55
4.1.3	Posición y número de electrodos	55
4.1.4	El estado estacionario de las señales EMGS	56
4.1.5	Segmentación de los registros de las señales EMGS	57
4.1.6	Selección de la base de datos de señales EMGS	58
4.1.7	Selección de canales y características mediante el índice CSI	60
4.2	Características Temporales DT	62
4.2.1	Método 1. Evaluación de parámetros temporales	62
4.2.2	Método 2. Evaluación de momentos temporales	65
4.3	Características Espectrales DWT	68
4.3.1	Nivel de descomposición DWT	69
4.3.2	Selección de la función wavelet	70
4.3.3	Selección de subbandas DWT	71
4.3.4	Método 1. Descomposición de valores singulares SVD	72
4.3.5	Método 2. Evaluación de parámetros espectrales	74
4.3.6	Método 3. Evaluación de momentos espectrales	77
4.4	Características Espectrales WPT	82
4.4.1	Nivel de descomposición WPT	83
4.4.2	Selección de la función wavelet	84
4.4.3	Método 1. Evaluación de parámetros espectrales	85
4.4.4	Método 2. Evaluación de momentos espectrales	85
4.5	Conformación de los vectores híbridos DT-DWT y DT-WPT	89
4.6	Implementación de Algoritmos y Funciones	91
4.6.1	Evaluación de índices CSI	92

4.6.2	Extracción de Características DT	92
4.6.3	Extracción de Características DWT	93
4.6.4	Extracción de Características WPT	93
Capítulo 5.	Clasificación de Patrones EMGS	95
5.1	Bases de Datos de Señales EMGS	96
5.2	Selección de Canales para la Clasificación de las Señales EMGS	96
5.3	Características DT-DWT y DT-WPT	98
5.4	Pruebas de Clasificación con Análisis PCA	100
5.5	Error de clasificación para las combinaciones de características DT-DWT y DT-WPT	101
5.6	Implementación de Algoritmos y Funciones	103
Capítulo 6.	Conclusiones	107
6.1	Aportes Originales	110
6.2	Trabajos Futuros	111
Anexo A.	113
A.1	Análisis Espectral de Fourier	113
A.2	Transformada Wavelet	118
A.2.1	Transformada Wavelet Continua (CWT)	119
A.2.2	Transformada Wavelet Discreta (DWT)	124
A.2.3	Transformada de Paquetes Wavelet (WPT)	133
A.3	Algoritmo: Selección de las Mejores Bases según Coifman-Wickerhauser136	
Anexo B.	141
B.1	Algoritmo: Índice de Separabilidad de Clases (CSI)	141
B.2	Algoritmo: Análisis de Componentes Principales (PCA)	143
Anexo C.	147
C.1	El Clasificador Bayesiano	147
C.2	El clasificador Bayesiano Gausiano	149
C.3	Redes Neuronales Artificiales	151
Referencias	157

Lista de Figuras

	Pág.
2.1 Estructura de la Neurona (Reischl, et al., 2006)	6
2.2 Generación y detección de la señal EMGS (De Luca, 1978)	9
2.3 Secuencia temporal EMG (arriba) y su espectro de frecuencias (abajo)	10
3.1 Sistema de clasificación de patrones EMGS	20
3.2 Localización temporal y limitación en banda de las funciones <i>wavelets</i>	27
3.3 Partición del espectro según la STFT (arriba) y según la DWT (abajo)	30
3.4 Descomposición DWT a un nivel (J=1)	30
3.5 Descomposición DWT a tres niveles (J=3)	31
3.6 Descomposición DWT a tres niveles de la señal <i>chirp lineal</i>	31
3.7 Descomposición <i>wavelet</i> a tres niveles: a) DWT, b) WPT	32
3.8 Análisis de la señal " <i>chirp lineal</i> " mediante descomposición WPT . . .	33
3.9 Resolución del análisis espectral mediante a) STFT, b) DWT y c) WPT	34
3.10 Conformación de la matriz de coeficientes <i>wavelet</i> DWT	36
3.11 Diagrama esquemático del sistema de clasificación	37
3.12 Arquitectura del Perceptrón Multicapas MLP	44
3.13 Arquitectura de una red PNN, (Georgiou V. et al. , 2004)	48
3.14 a) Separación lineal de clases por una red SVM, b) Hiperplano con máximo margen de separación de clases (Alfred M. , 1999)	49
4.1 Error de clasificación según la longitud de los segmentos (Englehart, et al., 2001)	57
4.2 Disposición de electrodos para el registro de señales EMGS (Har- grove, et al., 2005)	58

4.3	Movimientos seleccionados para su clasificación (Hargrove, et al., 2005)	59
4.4	Relación del número de características y el error de clasificación (Englehart, 1998)	60
4.5	Traslape o separabilidad de clases según el índice CSI	61
4.6	Promedios CSI para los parámetros <i>VMA</i> (caract. 1 y 2), <i>LO</i> (caract. 3 y 4), <i>CC</i> (caract. 5 y 6), <i>CSP</i> (caract. 7 y 8).	64
4.7	Promedios CSI para los parámetros <i>VMA</i> (caract. 1 y 2), <i>LO</i> (caract. 3 y 4), <i>CC</i> (caract. 5 y 6)	64
4.8	Promedios CSI por paciente para 4 y 6 características DT	65
4.9	Promedios CSI para los momentos M_1 (caract. 1 y 2), ... M_6 (caract. 11 y 12)	67
4.10	Promedios CSI para Momentos M_2 (caract.1 y 2) y M_4 (caract.3 y 4)	67
4.11	Promedios CSI por paciente según los métodos de: Parámetros y Momentos DT	68
4.12	Indices CSI para 3 y 4 niveles de descomposición DWT	69
4.13	Indices CSI según las tres mejores familias wavelet madre para DWT	70
4.14	Indices CSI según la selección de subbandas DWT	71
4.15	Comparación de índices CSI bajo el análisis PCA	72
4.16	Promedios CSI para los valores singulares sobre la matriz de coeficientes DWT	73
4.17	Promedios CSI para los parámetros <i>AVG</i> (caract. 1 y 2), <i>VAR</i> (caract. 3 y 4), <i>ENER</i> (caract. 5 y 6) y <i>SVD</i> (caract. 7 y 8), sobre la matriz de coeficientes DWT	76
4.18	Promedios CSI para los parámetros <i>AVG</i> (caract. 1 a 4), <i>VAR</i> (caract. 5 a 8), <i>ENER</i> (caract. 9 a 12) y <i>SVD</i> (caract. 13 a 16) sobre cada subbanda de coeficientes DWT	77
4.19	Promedios CSI por paciente según las opciones 1 y 2 del método de parámetros espectrales	77
4.20	Promedios CSI para los momentos espectrales M_1 (caract. 1 y 2), ... M_6 (caract. 11 y 12), sobre la matriz de coeficientes DWT	79
4.21	Promedios CSI para los momentos espectrales M_2 (caract. 1 y 2), M_4 (caract. 3 y 4) y M_6 (caract. 5 y 6), sobre la matriz de coeficientes DWT	80
4.22	Indices CSI promedios para los momentos M_1 (caract. 1, 2, 3 y 4), ... M_6 (caract. 21, 22, 23 y 24) sobre cada subbanda D_2 y D_3 de coeficientes DWT	80
4.23	Promedios CSI para los momentos espectrales M_2 (caract. 1 y 2), M_4 (caract. 3 y 4) y M_6 (caract. 5 y 6), sobre las subbandas D_2 y D_3 de coeficientes DWT	81

4.24 Promedios CSI por paciente según las opciones 1 y 2 del método de momentos	81
4.25 Promedios CSI por paciente para los tres métodos DWT	83
4.26 Índices CSI para 3, 4 y 5 niveles de descomposición WPT	84
4.27 Índices CSI según las tres mejores familias wavelet madre para WPT	84
4.28 Promedios CSI para los parámetros AVG (caract. 1 y 2), VAR (caract. 3 y 4), ENER (caract. 5 y 6) y SVD (caract. 7 y 8) sobre las mejores subbandas WPT	86
4.29 Índices CSI promedios para los momentos M_1 (caract. 1 y 2), ... M_6 (caract. 11 y 12), sobre las mejores subbandas WPT	87
4.30 Índices CSI promedios para los momentos M_2 (caract. 1 y 2), M_4 (caract. 3 y 4) y M_6 (caract. 5 y 6), sobre las mejores subbandas WPT	88
4.31 Promedios CSI por paciente para los métodos de parámetros y momentos WPT	88
4.32 Promedios CSI por paciente para los mejores métodos en las técnicas DT, DWT y WPT	90
4.33 Promedios CSI por paciente para las combinaciones DT-DWT y DT-WPT	92
5.1 Error de clasificación de los tres mejores índices CSI para DT-DWT	97
5.2 Error de clasificación de los tres mejores índices CSI para DT-WPT	97
5.3 Error de clasificación PNN para las combinaciones híbridas DT-DWT	99
5.4 Error de clasificación SVM para las combinaciones híbridas DT-DWT	99
5.5 Error de clasificación PNN para las combinaciones híbridas DT-WPT	99
5.6 Error de clasificación SVM para las combinaciones híbridas DT-WPT	100
5.7 Error de clasificación bajo la condición del análisis PCA sobre los coeficientes DWT	101
5.8 Error de clasificación según el clasificador PNN en los vectores DT-DWT y DT-WPT con y sin análisis PCA	101
5.9 Error de clasificación para las combinaciones de características DT-DWT y DT-WPT	103
5.10 Error de clasificación por paciente de las técnicas DT-DWT y DT-WPT	103
A.1 a) Base estandar (impulsos Dirac); b) Base de Fourier (sinusoides)	115
A.2 Resolución de análisis tiempo-frecuencia	116
A.3 Forma de onda de la señal <i>chirp lineal</i>	117
A.4 Representación STFT de la señal <i>chirp lineal</i> . En a) con $L = 32$ y $K = 1$; b) $L = 32$ y $K = 32$	117
A.5 Funciones wavelets más comunes	120

A.6	Modo de operación de la transformada wavelet continua CWT (De Castro , 2002)	122
A.7	Anidamiento de subespacios wavelets $\Omega_{j,1}$ y escalas $\Omega_{j,0}$ para $J = 3$.	126
A.8	Descomposición DWT a un nivel (J=1)	126
A.9	Descomposición DWT a tres niveles (J=3)	127
A.10	Partición diádica del espectro de frecuencias con DWT	128
A.11	a) Wavelet madre db3, b) wavelet padre db3	129
A.12	Análisis DWT en el plano bidimensional tiempo-frecuencia	130
A.13	Descomposición wavelet WPT a dos niveles (J=2)	133
A.14	Resoluciones tiempo-frecuencia a) Análisis DWT y b) Análisis WPT .	134
A.15	Estructura de subespacios resultantes en el árbol binario WPT	135
A.16	Vectores base paquetes wavelets con índices (j,k,n) = (escala, frecuencia y posición)	135
A.17	Selección de las mejores bases del árbol binario WPT	138
A.18	Respuesta tiempo-frecuencia de la señal <i>chirp lineal</i>	138
A.19	Error de aproximación en compresión de señales EMGS	139
A.20	Comparación entre la señal EMG original y su reconstrucción	140
C.1	Arquitectura del Perceptrón	152

Lista de Tablas

	Pág.
4.1 Mejores pares de canales según el índice CSI para 4 características DT extraídas a partir de los parámetros temporales <i>VMA</i> y <i>LO</i>	66
4.2 Mejores pares de canales según el índice CSI para 4 características DT extraídas a partir de los momentos temporales M_2 y M_4	68
4.3 Mejores pares de canales según el índice CSI para 8 características DWT extraídas a partir de los valores singulares SVD sobre la matriz de coeficientes DWT	74
4.4 Mejores pares de canales según el índice CSI para 8 características DWT extraídas a partir de los parámetros espectrales <i>AVG</i> , <i>VAR</i> , <i>ENER</i> y SVD sobre la matriz de coeficientes DWT	78
4.5 Mejores pares de canales según el índice CSI para 6 características DWT extraídas a partir de los momentos espectrales M_2 , M_4 y M_6 sobre la matriz de coeficientes DWT	82
4.6 Mejores pares de canales según el índice CSI para 8 características WPT extraídas a partir de los parámetros <i>AVG</i> , <i>VAR</i> , <i>ENER</i> y <i>SVD</i> sobre las <i>mejores subbandas</i> WPT	87
4.7 Mejores pares de canales según el índice CSI para 6 características WPT extraídas a partir de los momentos M_2 , M_4 y M_6 sobre las <i>mejores subbandas</i> WPT	89
4.8 Mejores pares de canales según índices CSI promedios comunes a las técnicas DT, DWT y WPT	91
5.1 Mejores pares de canales por paciente según el mínimo error de clasificación	98
5.2 Error de clasificación para las técnicas DT-DWT y DT-WPT	102

Lista de Abreviaturas

ANN: (*Artificial Neuronal Network*), Redes Neuronales Artificiales

AP: (*Action Potencial*), Potencial de Acción

AR: (*Auto Regresive*), Modelo Autoregresivo

CSI: (*Class Separability Index*), Indice de Separabilidad de Clases

DOFS: (*Degree Of Fredoms*), Grados de Libertad

EMG: (*ElectroMyographic Signals*), Señales Electromiográficas

EMGS: (*EMG Surface*), EMG Superficiales

FCM: (*Fuzzy C-Means*), Clasificación Fuzzy C-Medias

FT: (*Fourier Transform*), Transformada de Fourier

FKNN: (*Fuzzy K-Nearest*), Clasificación Fuzzy K-Vecinos más Cercanos

DWT: (*Discrette Wavelet Transform*), Transformada Wavelet Discreta

LDA: (*Linear Discriminant Analysis*), Análisis Discriminante lineal

MMSE: (*Mean Minimum Squared Error*), Mínimo Error Cuadrático Medio

MLP: (*Multi Layer Perceptron*), Perceptrón Multicapas

MU: (*Motor Unit*), Unidad Motora

MUAP: (*Motor Unit Action Potential*), Potencial de Acción de la Unidad Motora

PCA: (*Principal Component Analysis*), Análisis de Componentes Principales

PCM: (*Possibilistic C-Means*), Clasificación Posibilística C-means

PNN: (*Probabilistic Neuronal Network*), Red Neuronal Probabilística

RBFN: (*Radial Basis Function Network*), Redes de Funciones de Base Radial

STFT: (*Short Time Fourier Transform*), Transformada de Fourier de Tiempo Corto

STTT: (*Short Time Thompson Transform*), Transformada de Thompson de Tiempo Corto

SVD: (*Singular Value Decomposition*), Descomposición de Valores Singulares

SVM: (*Support Vector Machines*), Máquina de Soporte Vectorial

TD: (*Time Domain*), Dominio Temporal

TMUAP: (*Train MUAP*), Tren de MUAPs

WPT: (*Wavelet Packet Transform*), Transformada de Paquetes Wavelets

WT: (*Wavelet Transform*), Transformada Wavelet

Capítulo 1

Introducción

La clasificación de *señales electromiográficas superficiales* (EMGS) es ampliamente utilizada en el diseño de prótesis de control mioeléctrico y su desempeño está fuertemente influenciado por el tipo de análisis que se de a estas señales. Su análisis consiste básicamente en buscar una representación equivalente y adecuada de las señales mediante un conjunto reducido de características que las describa, esto es, que contengan la información más relevante, necesaria y suficiente para su discriminación y clasificación con el mínimo error posible. Este trabajo de investigación presenta la *teoría de wavelets* y su aplicación en el análisis de las señales EMGS para su representación, buscando mediante las técnicas de extracción de características temporales y características espectrales basadas en wavelets, conformar vectores de descripción híbridos, que optimicen la clasificación de patrones EMGS. La representación de las señales EMGS comprende las fases de *extracción de características* y de *reducción de dimensionalidad*.

En la fase de *extracción de características* se identifican varias técnicas, destacándose como pionera la basada en el *análisis temporal* por su sencillez y facilidad, por no requerir de ningún tipo de transformación sobre la señal (Hudgins, et al., 1993), luego aparecen las basadas en el *análisis espectral*, las cuales requieren de alguna transformación de la señal, por ejemplo mediante la *Transformada de Fourier de Tiempo Corto* (STFT - *Short Time Fourier Transform*) (Zecca, et al., 2002; Sijiang, 2003; Reaz, et al., 2006). A pesar de suponer esta transformada

la condición de estacionariedad de la señal, la cual no se cumple para las señales EMGS, su aplicación en clasificación de patrones mioeléctricos ha tenido una buena aceptación. Pero la naturaleza aleatoria y no estacionaria de las señales EMGS motivó a la exploración de nuevas alternativas de análisis y representación de este tipo de señales, buscando mejorar cada vez más el desempeño en la clasificación de patrones EMGS. Como resultado de esto surge la *teoría de wavelets* con la *transformada wavelet discreta* (DWT - *Discrete Wavelet Transform*) y la *transformada de paquetes wavelets* (WPT - *Wavelet Packet Transform*). Estas transformadas permiten una representación de señal en los dominios tiempo y frecuencia de manera simultánea y mucho más compacta que la lograda hasta entonces con la transformada STFT (Sijiang, 2003). Así, esta nueva forma de representación permitiría extraer características complementarias a las temporales y mediante su combinación estratégica buscar minimizar el error de clasificación de patrones EMGS.

En la fase de *reducción de dimensionalidad* se busca optimizar el conjunto de características extraídas de acuerdo a alguna función de costo. Aquí se identifican dos técnicas ampliamente reconocidas y muy utilizadas en la representación de señales: una por *selección de características* y la otra por *proyección de características*. La primera busca reducir la dimensión del conjunto original mediante la selección del mejor subconjunto o combinación de sus características, se considera un procedimiento supervisado (Englehart, 1998), y debe obedecer a algún criterio de evaluación como la pertenencia de clase. La segunda busca una representación de los datos en una forma más compacta y reducida mediante alguna transformación del conjunto original de características y se considera un procedimiento no supervisado. En esta segunda categoría, la técnica más utilizada por su capacidad de compactación es la del *análisis de componentes principales* (PCA - *Principal Component Analysis*) (Haykin, 1994), que opera sobre la totalidad de las características originales y no involucra la pertenencia de clase en su criterio. En este trabajo se consideran estas técnicas para determinar las dimensiones adecuadas de los conjuntos de características temporales y espectrales basadas en wavelets, teniendo en cuenta como función de costo la medida del *índice de separabilidad de clases* (CSI - *Class Separability Indice*).

Finalmente para evaluar los procesos de extracción de características y de reducción de dimensionalidad en la conformación de los vectores descriptores con fines de clasificación de señales EMGS, se someten estos vectores a la fase de

clasificación de patrones. Para ello existe una gran variedad de clasificadores, de los cuales se pueden destacar como los más utilizados los estadísticos y los basados en: redes neuronales, lógica difusa y neuro-fuzzy (Zecca, et al., 2002; Reaz, et al., 2006). En este trabajo se implementan tres clasificadores a saber: la *red neuronal probabilística* (PNN - *Probabilistic Neuronal Network*), la *máquina de soporte vectorial* (SVM - *Support Vector Machines*) y el *perceptrón multicapas* (MLP - *Multi Layer Perceptron*), con los cuales se evalúa el error de clasificación de las señales EMGS correspondientes a cuatro clases de movimientos de la mano.

Contribución Original

Son muchos los trabajos de investigación desarrollados al rededor de esta temática, sin embargo cada aplicación demanda de una cuidadosa selección de los métodos de procesamiento de las señales EMGS que responda a los requerimientos particulares. En este trabajo se exploran diferentes métodos de análisis y representación de señales, se investiga sobre técnicas de extracción de características temporales como características espectrales basadas en wavelets mediante las transformadas DWT y WPT, se evalúa la eficiencia del criterio de separabilidad de clases como función de costo, basado en el índice CSI, utilizado en este trabajo como herramienta de reducción de dimensionalidad por selección de características, se implementan sus algoritmos respectivos y se comparan, para finalmente determinar y proponer un método sistemático de representación de señales EMGS que busque minimizar el error en su clasificación. Así mismo, se propone un procedimiento para optimizar la correcta ubicación de los electrodos sobre el antebrazo del paciente, que garantice una buena detección de las señales EMGS. De esta manera se busca implementar la mejor solución al problema planteado, innovando e introduciendo mejoras sobre los métodos existentes, adaptada a los requerimientos particulares que demande la implementación de una prótesis de mano controlada a partir de señales mioeléctricas EMGS, cuyo desempeño busque un equilibrio entre multifuncionalidad, velocidad de respuesta, precisión en la discriminación de movimientos y sencillez de implementación.

La solución que se proponga, la cual consistirá en la mejor combinación de características temporales y espectrales basadas en wavelets, será evaluada mediante la clasificación de señales EMGS correspondientes a cuatro clases de movimientos o estados de la mano (mano relajada, mano abierta, agarre fino y agarre grueso).

Los registros de estas señales se encuentran en la base de datos “*levi_data.zip*”, suministrada por el Dr. Kevin Englehart del Biomedical Engineering Institute, Canadá, de quien se obtuvo la respectiva autorización para ser usada en el presente trabajo.

Estructura de la Tesis

En el presente capítulo se hizo una breve introducción al trabajo propuesto, destacando los aspectos más importantes a ser desarrollados, las contribuciones originales al mismo y la estructura del presente documento. En el Capítulo 2 se presentan las características fisiológicas de las señales EMG, su generación y detección, así como una revisión del estado del arte de sus aplicaciones en control de prótesis de mano. En el capítulo 3 se presenta una descripción general de las técnicas de análisis y extracción de características tanto temporales como espectrales basadas en wavelets mediante sus transformadas DWT y WPT para efectos de clasificación de patrones EMGS. Así mismo, se hace una revisión de las técnicas de reducción de dimensionalidad y las técnicas de clasificación de patrones más apropiadas para los propósitos de este trabajo. En el capítulo 4 se presenta las fases de implementación y desarrollo de los algoritmos para la extracción de características y la reducción de dimensionalidad en la representación de señales EMGS. En el capítulo 5 se evalúa el desempeño de los algoritmos desarrollados en el capítulo anterior, de acuerdo al error en la clasificación de los patrones EMGS y en el capítulo 6 se presenta las conclusiones del trabajo realizado, las contribuciones originales y los trabajos futuros en la misma línea de investigación que pueden ser desarrollados como una continuación y extensión a los logros alcanzados en este trabajo.

Capítulo 2

Generalidades sobre las Señales EMG

El procesamiento de señales biomédicas hace parte fundamental de la ciencia de extracción de información clínicamente significativa a partir de señales fisiológicas. En particular, la *Electromiografía* mide y registra la actividad eléctrica generada en el músculo en respuesta a una estimulación nerviosa, los registros de dicha actividad eléctrica se denominan señales *electromiográficas* (Reaz, et al., 2006). En los últimos años estas señales han sido ampliamente investigadas para su utilización como señales de comando en prótesis mioeléctricas multifuncionales y en aplicaciones de tele-operación de dispositivos robotizados (Alonso, et al., 2002).

2.1 Electromiografía

“La Electromiografía es el estudio electrofisiológico del sistema neuromuscular y consiste en una prueba que mide la respuesta de los músculos a los estímulos nerviosos mediante el registro de acción provenientes de las células excitables”. Para entender la fisiología de las señales EMG, se presentan los conceptos de las unidades fisiológicas que intervienen en el proceso (Valls, 2006):

2.1.1 La Neurona

La neurona es la unidad celular del sistema nervioso central, tienen tres componentes principales: las dendritas, el cuerpo de la célula o soma y el axón. Las dendritas son el árbol receptor de la red y son fibras nerviosas que cargan de señales

eléctricas el cuerpo de la célula, el cuerpo de la célula realiza la suma de estas señales eléctricas de entrada y el axón es quien se especializa en la conducción de dichas señales acumuladas desde el cuerpo de la célula hacia otras neuronas (ver figura 2.1). Estructural y funcionalmente, la neurona es una *unidad celular* (entidad completa y autónoma), con la capacidad de conectarse con otras ya sea para inhibirlas, excitarlas o simplemente para re-transmitirles un impulso nervioso o señal electroquímica que viene desde el cerebro y cuyo destino son las *unidades motoras*.

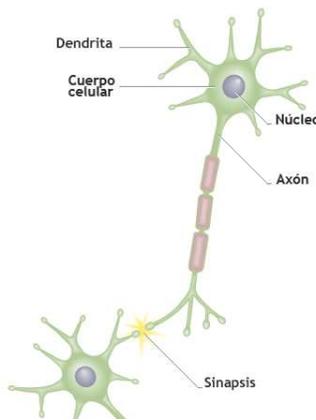


Figura 2.1. Estructura de la Neurona (Reischl, et al., 2006)

Hay tres tipos de neuronas: *las sensoriales*, *las motoras* y *las interneuronas*. Las *neuronas sensoriales* conducen impulsos informativos (visión, sonido, tacto, dolor, etc.) de los receptores (por ejemplo la piel) hacia la médula espinal y el cerebro. Las *neuronas motoras* conducen los impulsos del cerebro y la médula espinal hasta los receptores (por ejemplo los músculos y glándulas exocrinas), es decir en sentido contrario a las sensoriales. Y las *interneuronas* son células nerviosas multipolares cuyo cuerpo y procesos se ubican exclusivamente en el sistema nervioso central (específicamente en el cerebro), no tienen contacto directo con estructuras periféricas (receptores o transmisores).

2.1.2 Los Músculos

Son los órganos que se encargan de la movilidad y la estabilidad del cuerpo. Poseen las propiedades de: contracción (poder acortar sus fibras), elasticidad (poder recuperar su forma después de una contracción) y excitabilidad (responder a los

estímulos). Los músculos están formados por células largas con múltiples núcleos a los que se les llama “fibras musculares”. Cada extremo de estas fibras se inserta en los tendones, y estos a su vez en los huesos cruzando las articulaciones. Estas fibras musculares contienen miles de “miofibrillas”, donde se encuentra la unidad contráctil del músculo y la sinapsis neuromuscular. Juntos llevan a cabo los procesos electroquímicos para producir la contracción muscular.

2.1.3 La Unidad Motora

Un mismo músculo recibe varias fibras nerviosas motoras, la unión entre una neurona motora y las fibras musculares que inerva se llama *unidad motora (MU - Motor Unit)*. Estas pueden variar de tamaño, desde una neurona que inerva 10 fibras musculares (como en el globo ocular), hasta una neurona que inerva 200 o más fibras musculares (como en los músculos de las extremidades). En el primer caso los movimientos son finos y de poca potencia, en el segundo caso son burdos y de gran potencia. Las neuronas motoras se originan en la corteza cerebral motora, luego se interconectan con otras neuronas motoras del tronco cerebral (Tálamo y Bulbo Raquídeo) y finalmente con las que se encuentran en la médula espinal, específicamente en las astas anteriores terminando en su órgano receptor que son los músculos.

2.1.4 El Potencial de Acción

El proceso por el cual el axón transmite una señal impulsiva se denomina *potencial de acción (AP - Action Potencial)*, que consiste en una onda eléctrica auto-regenerada que se propaga a lo largo del axón. Los potenciales de acción alcanzan una amplitud máxima de unos 100 mV, duran aproximadamente 1 ms y la tasa de generación de AP's está limitada a unos 200 impulsos por segundo o menos.

2.1.5 El Potencial de Acción de la Unidad Motora

Se denomina *potencial de acción de la unidad motora (MUAP - Motor Unit Action Potencial)* a la combinación de *potenciales de acción (APs)* de las fibras musculares en una unidad motora aislada. El tamaño de un *MUAP* está relacionado con el diámetro del axón motor, su grosor de mielina, la velocidad de conducción de la fibra

nerviosa, el umbral de despolarización y el tipo de fibras musculares inervadas. Las principales características morfológicas de los MUAPs son:

- **Duración.**- Se define como el tiempo comprendido entre los instantes de la deflexión y el retorno al punto de reposo inicial. Refleja el número y dispersión espacial de las fibras musculares de la *unidad motora* (MU). La duración media de los MUAPs se encuentra entre 5 y 15 mseg, ya que la duración varía para cada músculo, edad y temperatura.
- **Amplitud.**- Se mide entre picos de la señal y refleja la actividad de un número relativamente pequeño de fibras musculares, normalmente mide entre 200 μV y 3 mV.
- **Fases.**- Corresponden a la cantidad de transiciones a uno y otro lado del nivel de reposo.
- **Estabilidad.**- Consiste en la conservación morfológica de la señal durante sucesivas excitaciones. Se cuantifica mediante la medición sistematizada del jitter.

2.2 Generación de Señales EMG

En respuesta a un potencial de acción de la neurona motora una fibra muscular se despolariza, contrayéndose mientras la señal se propaga a lo largo de su superficie. Esta despolarización genera un campo eléctrico en la vecindad de las fibras musculares, el cual puede ser detectado mediante un electrodo de superficie ubicado sobre la piel en cercanía de este campo (ver figura 2.2), o mediante un electrodo de aguja insertado en el músculo.

Todas las fibras musculares asociadas a la unidad motora son excitadas al tiempo, por lo que la excitación repetida de la unidad motora crea un tren de impulsos conocidos como *Tren de MUAPs* (TMUAP). Y a la combinación aditiva de la actividad eléctrica creada por cada unidad motora excitada al momento de realizar un movimiento o contracción muscular se conoce como señales *electromiográficas* (EMG - ElectroMyographic Signals) (De Luca, 1978).

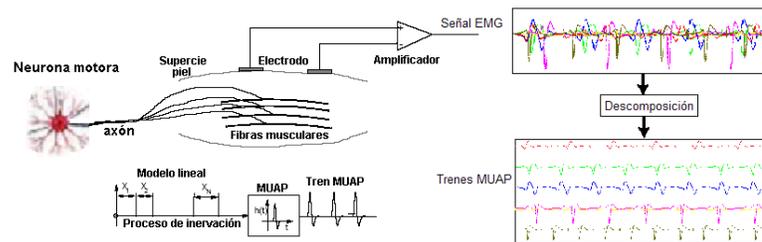


Figura 2.2. Generación y detección de la señal EMGS (De Luca, 1978)

2.3 Detección de Señales EMG

El control de la fuerza en la actividad muscular depende básicamente de dos elementos: el *reclutamiento* de mayor o menor cantidad de unidades motoras activas y su composición, y de la *tasa de disparo* de los impulsos de excitación sobre las neuronas motoras individuales. Según como aumente la señal cerebral para producir una contracción muscular, también aumenta el número de unidades motoras reclutadas para ejecutar la contracción y la frecuencia de los impulsos de excitación. Todas las células musculares dentro de una misma unidad motora se activan al mismo tiempo y mediante la variación del número de unidades motoras activas el cuerpo puede controlar la fuerza de la contracción muscular. Así mismo, para sostener una contracción muscular las unidades motoras se deben excitar repetidamente. Según como aumente la tasa de excitación de la unidad motora, la contracción asociada con cada impulso se sumará para lograr cada vez fuerzas mayores (Valls, 2006).

La tecnología electromiográfica permite detectar la actividad eléctrica creada por la *unidad motora* (MU) sobre un músculo entero o una porción de este; ya sea de manera invasiva con agujas o mediante la ubicación de electrodos en la superficie de la piel. En la detección de señales *EMG superficiales* (*EMGS - EMG Surface*), la ubicación de los electrodos es muy importante, pues de ello depende la detección de información utilizable en la discriminación de movimientos (Sijiang, 2003). Así, una señal EMGS es la suma de la actividad eléctrica temporal y espacial de *unidades motoras* (MUs) individuales dentro de múltiples músculos ubicados en la vecindad de los electrodos de detección. La señal tiene una amplitud típica entre 0 y 10 mV, con un espectro de frecuencias utilizable entre los 0 y 500 Hz, con mayor concentración de energía entre 50 y 150 Hz, como se puede apreciar en la figura 2.3.

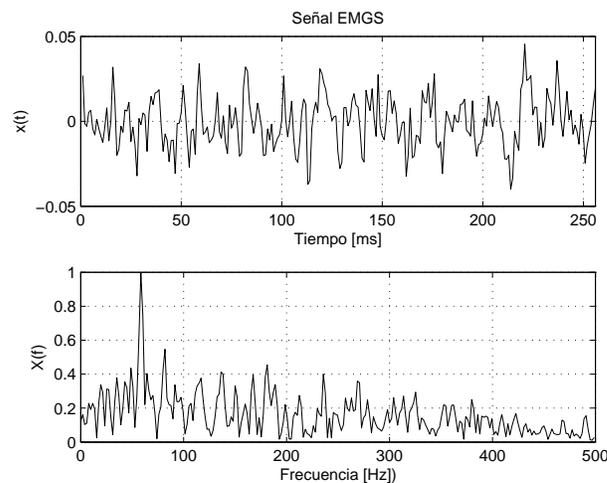


Figura 2.3. Secuencia temporal EMG (arriba) y su espectro de frecuencias (abajo)

Sijiang (2003) afirma que el valor instantáneo de una señal EMG no contiene información, sin embargo su amplitud obedece por naturaleza a un proceso estocástico que puede ser representado razonablemente mediante una función de distribución Gausiana. Según este, los factores que afectan a una señal EMG se pueden clasificar de la siguiente manera:

- i) **Factores causativos.**- Estos afectan directamente a la señal, pueden ser *extrínsecos* e *intrínsecos*. Los *extrínsecos* se deben a la estructura de los electrodos y su ubicación; incluyen factores como: el área de la superficie de detección, la ubicación del electrodo con respecto a los puntos motores en el músculo, la ubicación del electrodo sobre el músculo con respecto a la terminación lateral del mismo y la orientación de las superficies de detección con respecto a las fibras musculares que generan la señal EMG. Los *intrínsecos* se deben a factores fisiológicos, anatómicos y biomédicos; tienen lugar debido a factores como: el número de unidades motoras activas, la composición y tipo de fibra muscular, el flujo sanguíneo, el diámetro de la fibra, la profundidad y ubicación de fibras activas y la cantidad de tejido entre la superficie del músculo y el electrodo.
- ii) **Factores intermedios.** Son fenómenos físicos y psicológicos influenciados por uno o más factores causativos. Las razones pueden estar en el efecto

de filtro *pasa-banda* de un electrodo por su volumen de detección, la superposición de potenciales de acción en la señal EMG detectada, la velocidad de conducción del potencial de acción sobre la membrana de la fibra muscular y el “crosstalk” por cercanía entre músculos.

iii) Factores determinísticos. Influenciados por factores intermedios como el número de unidades motoras activas, la tasa de excitación motora, la interacción mecánica entre fibras musculares (que tienen relación directa entre la información de la señal EMG y la información de la fuerza acumulada), así como la amplitud, la duración y la forma del potencial MUAP.

Según Reaz, et al. (2006), para mejorar la calidad de la señal EMG se debe alcanzar la mejor relación señal a ruido. La amplitud de la señal EMG está entre 0 y 10 mV (-5 mV a +5 mV) antes de ser amplificada, pero es contaminada por ruido mientras se transporta por los tejidos. A continuación se presenta las diferentes categorías de ruido con sus respectivas características.

- **Ruido inherente al equipo de medida.**- Todo equipo electrónico genera ruido que no puede ser eliminado completamente.
- **Ruido ambiental.**- Debido a la radiación electromagnética sobre nuestros cuerpos, cuyas amplitudes fácilmente pueden superar los niveles de las señales EMG.
- **Ruido por movimiento.**- Cuando los dispositivos de captura son removibles (como los electrodos sobre la piel) introducen ruido por contacto de superficies en movimiento.
- **Ruido inherente a la estabilidad de la señal.**- La aleatoriedad en la amplitud EMG, debida a la variación en la tasa de excitación de la unidad motora se encuentra entre 0 y 20 Hz. Otro aspecto importante es evitar al máximo distorsiones de la señal, por ejemplo evitar distorsionar los picos de la señal EMG con filtraje innecesario o no recomendado, como ocurre con los filtros tipo “notch”.

2.4 Las Señales EMG y el Control de Prótesis de Mano

El reto científico y de ingeniería es lograr una replica de las funciones sensorio-motrices de la mano humana. Su consecución implica un gran número de *grados de libertad* (DoFs - *Degree Of Fredoms*), pues esta tiene 22 DoFs controlados por más o menos 38 músculos de la mano. Lograr esto en un sistema de prótesis para ejecutar movimientos precisos, delicados y de gran fuerza con capacidad sensorial, será un gran reto para el hombre. Pero aún lográndose, resultaría en un sistema demasiado complejo y su control por parte del usuario se tornaría difícil o casi imposible, excepto si se usa una codificación compleja de los movimientos lo que implicaría demasiado entrenamiento (Zecca, et al., 2002). Según Zecca, et al. (2002), en 1948 Reiter R. fue el primero en utilizar señales EMG para controlar un simple dispositivo de prótesis. En nuestros días, todos los dispositivos de prótesis usados en prácticas clínicas tienen uno o dos DoFs, directamente controlados a partir de un par de electrodos ubicados en dos músculos antagonistas, ambos trabajando en modo on/off.

Hacia el año de 1975, algunos grupos de investigación lograron el modelado de señales EMG para el control de dispositivos de prótesis con más de un DoFs, sin embargo los recursos software y hardware disponibles en aquella época no permitieron la realización de tales dispositivos para prácticas clínicas. En 1982, Graupe, et al. (1982) con modelado *autoregresivo* (AR - *Auto Regressive*) y discriminación Bayesiana, fueron capaces de discriminar seis clases de movimientos con un éxito de 99%. Desafortunadamente estos resultados solo se lograban después de muchas horas de entrenamiento del usuario que controlaba el dispositivo, degradándose su desempeño significativamente con el tiempo debido a los cambios en la señales EMG producidos por el mismo usuario. Posteriormente, el uso de *redes neuronales artificiales* (ANN - *Artificial Neuronal Network*) permitió reducir el tiempo requerido para el entrenamiento, pero los problemas relacionados a las variaciones de la señal con el tiempo y de persona a persona aún continuaban presentes (Zecca, et al., 2002).

De otra parte, en la generación de la señal EMG debida a una contracción muscular se identifican dos fases o estados: uno, el *estado transitorio* que comprende los primeros instantes de la generación de la señal y dos, el *estado estacionario* que comprende los instantes posteriores al transitorio y que se genera cuando el

movimiento es sostenido. Los trabajos iniciales se realizaron analizando la fase *estacionaria* de la señal EMG, para lo cual era necesario pedir al usuario que ejecute un movimiento y lo sostenga durante algunos segundos. Esta forma de análisis simplificó el control de los dispositivos de prótesis, aún sabiendo que cierta cantidad de información de la señal EMG era ignorada y que la fase estacionaria de la señal tiene una estructura temporal muy corta, en parte debido a los cambios dinámicos en el reclutamiento de unidades motoras y el patrón de los impulsos eléctricos necesarios para una contracción sostenida (De Luca, 1979). Por su parte, en 1993 Hudgins, et al. (1993) fueron los primeros en proponer una nueva estrategia de control de dispositivos artificiales utilizando el estado *transitorio* de la señal EMG, en su análisis observaron que existía cierta estructura en la señal (durante el establecimiento de la contracción) diferenciable para varios movimientos del antebrazo, la cual podría ser utilizada como fuente de información para clasificación de señales EMG. De esta manera lograron discriminar cuatro movimientos diferentes (extensión del codo, flexión de muñeca, y rotaciones humeral radial y humeral lateral) a partir de un solo electrodo bipolar. Todo esto con evaluación de características temporales y clasificación con una red ANN de dos capas.

Simultáneamente a la experimentación con las dos fases de la señal EMG, se investigaba en diversas técnicas de análisis de la señal, tanto para la extracción de características, la reducción de dimensionalidad, como su clasificación. En este sentido, Kang, et al. (1966) utilizaron el modelo AR para representar la señal EMG y Chang, et al. (1966), aprovecharon la información de los *coeficientes cepstrales* de la señal EMG como comandos de control para una interfaz hombre-maquina. Englehart, et al. (1999a), propusieron la clasificación de señales EMG, en su fase transitoria, aplicando por primera vez la *transformada wavelet* en la representación tiempo-frecuencia de la señal. Aplicaron las transformadas STFT, DWT y la WPT, compararon su desempeño con características del *dominio temporal* (TD - Time Domain) propuestas por Hudgins y concluyeron que la precisión en la clasificación se logra en el orden progresivo TD \rightarrow STFT \rightarrow DWT \rightarrow WPT. Estos resultados se aplicaron en la discriminación de cuatro patrones de señales EMGS correspondientes a movimientos del codo y antebrazo, colectadas del bíceps y del tríceps mediante un par de electrodos.

En el mismo año Englehart, et al. (1999b) propusieron un esquema de clasificación continua basada en wavelets para el control mioeléctrico multifuncional, afir-

mando que una efectiva extracción de características es crucial para una adecuada clasificación. Para ello extrajeron características wavelets WPT, con reducción de dimensionalidad PCA y un clasificador basado en el *análisis discriminante lineal* (LDA - *Linear Discriminant Analysis*). Sus resultados fueron aplicados en la discriminación de seis movimientos de mano y muñeca (mano cerrada, mano abierta, flexión de muñeca, extensión de muñeca, desviación cubital y desviación radial). Concluyeron que utilizando cuatro canales se obtiene mejores resultados que con uno o dos, puesto que se dispone de un mayor número de combinaciones posibles de patrones EMG utilizables eficientemente en la discriminación. Demostraron además, que se puede lograr una sorprendente precisión en la discriminación utilizando señales EMG en su estado estacionario, pues sus trabajos previos se basaban en análisis de señales EMG en su fase transitoria. De esta manera, con la combinación WPT/PCA/LDA el error de clasificación era de 6.25% comparado al error obtenido por Hudgins del 9.25%.

En el año 2001 Reischl, et al. (2006) desarrollaron en el Instituto Forschungszentrum Karlsruhe la mano artificial *FZK-Hand* con la habilidad de mover todos los dedos independientemente. Mediante la combinación de una red neuronal artificial y lógica difusa se logró de manera automática la determinación de la intensidad de movimiento del usuario mediante un continuo escaneo de la activación muscular, con lo cual se logró ejecutar cinco diferentes tipos de agarre (gancho, pinza, pellizco lateral, esférico, apuntamiento con dedo índice). Adicionalmente se le incorporó sensores de realimentación para controlar la fuerza aplicada según se requiera para diferentes objetos. Alonso, et al. (2002), en este mismo año desarrollaron un *Entrenador Mioeléctrico de Prótesis para Amputados de Brazo y Mano*, el cual consiste en un sistema (software y hardware) para adecuar las prótesis a personas amputadas y posibilitar un aprendizaje previo de su manejo. Para esto, inicialmente se entrena la prótesis con personas normales que ejecutan cuatro movimientos (apertura de mano, cierre de mano y giro de muñeca en ambos sentidos), utilizando tan solo la señal de dos electrodos diferenciales ubicados en los músculos antagonistas del brazo. Para la extracción de características utilizaron el método de "histograma de Zardoshti" y la clasificación de patrones se basó en redes neuronales.

Por otra parte, conscientes de la fatiga o cansancio que puede producir la portabilidad y el control de un dispositivo de prótesis, Ferguson y Dunlop (2002), desarrollaron un sistema que permite identificar cuatro tipos de agarre complejos (esférico,

cilíndrico, de pinza y lateral) a partir de señales EMGS, la cual tenía como propósito evitar la fatiga del usuario. En la extracción de características utilizaron la transformada de Fourier STFT, el modelado AR, la transformada wavelet DWT y la deconvolución, y para el reconocimiento de patrones probaron métodos estadísticos (“Distancias Euclídea y de Mahalanobis”) y redes neuronales. Sus resultados concluyeron en la afirmación de que la misma combinación de técnicas de extracción de características y clasificación de patrones no tiene el mismo desempeño para todos los pacientes, por lo que el dispositivo deberá adaptarse para cada caso con las técnicas que mejor desempeño provean.

El tener que adecuar la prótesis a cada paciente, dada la incertidumbre en el patrón de la señal EMG para los diferentes movimientos, pasó a ser un aspecto de interés y de investigación. Su solución se vislumbró en el uso de redes neuronales con lógica difusa, afirmando que es una técnica muy ventajosa en procesamiento y clasificación de señales biomédicas, dado que estas no son señales estrictamente repetibles y muchas veces aún contradictorias, de tal manera que los sistemas con lógica difusa podrían tolerar dichas contradicciones, además de permitir discriminar patrones en los datos que con otras técnicas no sería tan fácil. En este sentido, Zhang, et al. (2002), propusieron una técnica de clasificación neuro-fuzzy basada en wavelets para control EMG aplicada al control de prótesis con múltiples DoFs. La combinación DWT/Valores Singulares (SVD - Singular Value Decomposition) para extracción de características y una red neuro-fuzzy para clasificación fueron aplicadas al reconocimiento de seis movimientos (extensión de codo, flexión del codo, pronación de muñeca, supinación de muñeca, mano abierta y mano cerrada) con mejores resultados que los obtenidos hasta entonces con métodos tradicionales. Así mismo Yücel y Mehmet (2002), en este mismo año propusieron la combinación DWT/PCA y clasificación fuzzy con los algoritmos: “fuzzy c-means” (FCM), “possibilistic c-means” (PCM) y “fuzzy K-nearest” (FKNN) para señales EMG. Su método se aplicó en la ejecución de cuatro movimientos (flexión y extensión de codo, pronación y supinación de antebrazo) sobre un dispositivo de prótesis, concluyendo que sus resultados eran mejores que los logrados con las características temporales, los coeficientes AR o los coeficientes cepstrales.

Betancourt, et al. (2004), presentaron una metodología para el reconocimiento de patrones de movimientos (flexión, supinación, extensión y pronación), a partir de señales EMG del sistema fisiológico brazo-antebrazo. Su técnica de procesamiento

y caracterización se basó en: aproximación temporal, modelado AR, transformaciones STFT y wavelets para construir un conjunto híbrido de características, y una red neuronal en configuración “back-propagation” para clasificación de los patrones de movimientos, con un acierto del 95% para flexión y extensión y del 90% para supinación y pronación. En este mismo año, un estudio comparativo del desempeño de las diferentes familias wavelets disponibles en el toolbox de labView con las transformadas DWT y WPT fue desarrollado por Kilby y Hosseini (2004), para ello utilizaron señales EMGS tomadas de un músculo bajo contracción durante 4 segundos y con diferentes cargas. Con base en el error de aproximación obtenido tras la descomposición y reconstrucción de la señal concluyen que las familias wavelets como: “Daubechies”, “Symmlet” o “Coiflet” son las más adecuadas para el análisis de señales EMGS.

Si bien la transformada *wavelet* ha demostrado en los últimos trabajos ser la más adecuada para el análisis y representación de las señales EMGS, el espacio de características que resulta de los coeficientes wavelet es un espacio de dimensión muy grande. A este problema ya se habían sugerido como soluciones las técnicas de reducción de dimensionalidad, entre ellas el análisis PCA y la *descomposición de valores singulares* (SVD - *Singular Value Decomposition*). Pero los investigadores y científicos continuaron innovando y proponiendo estrategias diferentes. Un ejemplo de esto es el trabajo de Rodriguez y Vuskovic (2005), quienes desarrollaron un método de extracción de características basado en el cálculo de los tres primeros “momentos” a partir de los coeficientes wavelets. Su método se aplicó en la clasificación de patrones EMG de aprehensión o agarre (cilíndrico, esférico, lateral y de precisión), mostrando ser significativamente mejor que el basado en la *transformada de Thompson de tiempo corto* (STTT - *Short Time Thompson Transform*) de tres ventanas. Para la etapa de clasificación utilizaron la distancia “Mahalanobis” basada en clasificadores ARTMAP (“Adaptive Resonant Theory - based algorithm for supervised incremental learning and classification”).

En el 2005 varios trabajos fueron realizados con la combinación de las mejores técnicas descubiertas hasta el momento. Así, Wang y Cui (2005), usaron la combinación DWT/ANN en señales EMGS de cuatro electrodos que registraron la actividad de cuatro músculos. Se evaluaron los “valores propios” de los coeficientes wavelets de seis niveles de descomposición para obtener los “vectores propios”,

los cuales se sometieron a una red neuronal “Elman”¹. Su técnica se aplicó en la identificación y clasificación de ocho patrones diferentes de movimiento con una alta tasa de identificación. Rodriguez y Vuskovic (2006), proponen la evaluación de los primeros cinco momentos sobre las secuencias de coeficientes wavelets de tercer nivel como las características descriptoras. Su nueva propuesta se aplicó en clasificación de cuatro tipos de agarre de mano (cilíndrico, esférico, de precisión y lateral). Se presenta como una extensión al trabajo de Englehart con las transformadas DWT y WPT, pero reduciendo la dimensionalidad con la evaluación de momentos en lugar de usar PCA. Para la clasificación utilizaron la red SVM de “Vapnik”, la cual ofrece buen desempeño en casos de tener espacios sobredimensionados de características y pocas muestras de entrenamiento. Por su parte Hargrove, et al. (2005), presentan un estudio de comparación de desempeño entre la clasificación de señales EMG superficiales y señales EMG intramusculares, bajo la premisa de que las señales intramusculares tienen menos “crosstalk” muscular, lo cual permitiría puntos de control más independientes. Sin embargo, los resultados experimentales aplicados a la clasificación de seis patrones de movimiento mostraron que no hay diferencia significativa en la precisión de clasificación. Este resultado es particularmente importante, puesto que implica que para aplicaciones de control de prótesis se puede trabajar eficientemente con señales EMG superficiales sin perder precisión en la clasificación, además resulta más sencillo de obtener las señales sin invadir el organismo del paciente. La extracción de características se basó en estadísticos temporales, en el modelo AR de 6^{to} orden y el valor *rms* de la señal. Para el reconocimiento de patrones utilizaron el clasificador estadístico LDA y la red neuronal MLP.

Los trabajos previamente realizados, independientemente sobre el estado transitorio o sobre el estado estacionario de las señales EMG, advierten que cada estado contiene algún tipo de información útil para clasificación de señales EMG asociadas a clases de movimientos. Esto hace pensar que la combinación de esta información puede ayudar a mejorar las técnicas de reconocimiento y clasificación de patrones EMG como lo muestran Liu, et al. (2005), quienes desarrollaron en el 2005 un método de aprendizaje en tiempo real aplicando la transformada *wavelet* en la extracción de características y una red neuronal en la clasificación de señales EMG de cuatro canales. El método se aplicó para discriminación de cinco dife-

¹Red de Elman: red neuronal de dos capas, con algoritmo de entrenamiento en backpropagation.

rentes clases de movimientos y controlar en tiempo real una prótesis. Analizaron tanto la fase transitoria como la estacionaria de la señal EMG en la ejecución de un movimiento, mostrando que la discriminación de movimientos puede ser mejorada sustancialmente con esta forma combinada de análisis.

Como hemos visto, las señales EMG han sido ampliamente investigadas, con propósitos de aplicación en control mioeléctrico de prótesis multifuncionales y en tele-operación de dispositivos robotizados. Sin embargo, aún no se ha logrado conseguir el control óptimo de una mano multifuncional. El mayor problema es la variación en el tiempo de las características de las señales EMGS, debido a los cambios fisiológicos del músculo y a los cambios en el acoplamiento entre la piel y los electrodos. Un problema igualmente importante es la naturaleza estocástica de estas señales, resultando en errores de estimación de parámetros que a su vez causan dificultades de clasificación y/o de control. Adicionalmente, algunos errores de control son introducidos por la incapacidad del usuario para generar o reproducir idénticas contracciones cada vez que repite un movimiento, lo cual se traduce en un error del operador. Una posible solución a esta dificultad es la implementación de un módulo de aprendizaje en línea ya sea supervisado o no, para adaptar continuamente los parámetros del clasificador (Zecca, et al., 2002).

La motivación de este trabajo parte de que siendo conscientes de las muchas limitaciones intrínsecas de una prótesis normal, su desempeño puede incrementarse significativamente mediante la implementación de mejores técnicas de análisis de señales EMGS para su representación y optimizar la clasificación de sus respectivos patrones de movimiento.

Capítulo 3

Representación y Clasificación de Señales EMGS

El procesamiento de señales implica aplicar técnicas de análisis para propósitos específicos como: la eliminación de ruido e interferencias, la compresión, la clasificación, etc. En este trabajo se plantea el análisis temporal y espectral basado en *wavelets* de señales EMGS para su representación, con miras a minimizar el error de clasificación de patrones de movimientos mioeléctricos.

Actualmente la clasificación de señales EMGS es ampliamente utilizada en el control de prótesis de miembros superiores como la mano. Este proceso consiste en identificar a partir de estas señales la intención de un movimiento determinado. Según Sijiang (2003), esto es posible gracias a la estructura diferenciable en la forma de onda temporal de las señales EMGS, las cuales son generadas durante la contracción muscular en el establecimiento y/o sostenimiento de un movimiento determinado. Sin embargo, hoy en día se sabe que la estructura espectral de una señal puede contener también información valiosa para su clasificación.

Englehart, et al. (1999a), describen la clasificación de patrones como un proceso de tres fases: la *extracción de características*, la *reducción de dimensionalidad* y la *clasificación de patrones* como se muestra en la figura 3.1. La primera fase consiste en lograr una adecuada caracterización de las señales y será determinante para una buena clasificación. Las características pueden extraerse del dominio temporal de la señal a partir de sus parámetros estadísticos y/o del dominio espectral de la señal mediante transformaciones como la *transformada de Fourier STFT* o

las *transformadas wavelets* DWT y WPT. La segunda fase consiste en minimizar el número de características descriptoras. Esto puede hacerse de dos maneras: por *selección de características* y por *proyección de características*; con la selección de características se busca el mejor subconjunto de características a partir del conjunto original y con la proyección de características se busca la mejor combinación de estas mediante alguna transformación aplicada sobre el conjunto original de características. La tercera y última fase consiste en discriminar a partir de vectores conformados por las características extraídas, la clase del movimiento al que pertenece cada patrón de prueba EMGS. Existen clasificadores estadísticos, clasificadores estructurales y los más modernos denominados de aprendizaje (Englehart, 1998). A continuación se describe con mayor detalle cada una de las fases del sistema de clasificación de patrones EMGS.

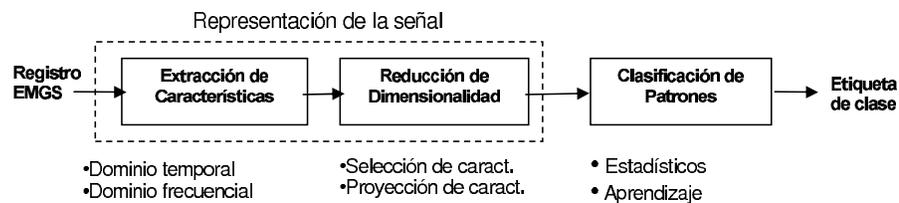


Figura 3.1. Sistema de clasificación de patrones EMGS

3.1 Extracción de Características

Para entender el proceso de extracción de características es necesario definir primero algunos términos a ser utilizados en todo el documento (Marques de Sa, 2001):

- **Clase:** Estado natural o categoría de objetos asociados a conceptos o prototipos.
- **Patrón:** Cualquier cosa que sirva para modelar una colección de objetos que van a ser observados.
- **Característica:** medida, atributo o primitiva para caracterizar un patrón.

La extracción de características con fines de clasificación de señales EMGS consiste en conformar patrones de N variables $\{x_i\}_{i=1}^N$, que se pueden asociar como componentes de un vector \mathbf{x} , el cual debe pertenecer a una de K clases, denotadas como y_k . Así, la clasificación puede ser vista como una función $d : X \rightarrow Y$, donde X es el *espacio de la señal de entrada* y Y es el *espacio de la señal de salida* consistente en K etiquetas de clase, tal que a cada patrón $\mathbf{x} \in X \subseteq R^N$ se le asocia una etiqueta de clase $\mathbf{y} \in Y = \{y_1, y_2, \dots, y_K\}$ (Englehart, 1998).

La conformación directa de un vector descriptor \mathbf{x} , con las muestras de la señal en su espacio original, no siempre resulta conveniente por las siguientes razones: primero, porque el espacio de la señal de entrada generalmente es de gran dimensión como sucede con las señales EMGS, por lo que conviene procesarlas en secciones cortas (comúnmente de 256 muestras), segundo, porque la presencia de ruido en los registros de las señales EMGS puede dificultar la clasificación, y tercero, porque en el espacio original de la señal puede haber información redundante. Ante esto, se prefiere extraer del espacio original de la señal en R^N un conjunto de características componentes de un vector descriptor $\mathbf{v} \in V \subseteq R^M$ de menor dimensión, esto es $M \leq N$. Ahora, si el espacio de dimensión R^M aún resulta ser demasiado grande, se puede introducir una fase intermedia entre la extracción de características y la clasificación llamada fase de reducción de dimensionalidad, con la que se busca extraer las características mínimas necesarias en un vector $\mathbf{z} \in Z \subseteq R^L, L < N$. Esta fase mejora el desempeño del clasificador y reduce la carga computacional.

Partiendo del hecho de que el valor instantáneo de las señales EMGS no contiene información, y según los modelos aceptados para su modelamiento, resultan ser estocásticas (De Luca, 1979), solo es posible obtener información de estas señales mediante las *técnicas de extracción de características*, dentro de las cuales se destacan además de ser pioneras, las basadas en el análisis temporal por su facilidad de evaluación debido a que no requieren ningún tipo de transformación sobre la señal. En este sentido, Hudgins, et al. (1993) encontrando en las señales EMGS una estructura inherente para cada tipo de contracción, extrajeron esta información estructural segmentando la forma de onda y determinando un conjunto de características a partir de la evaluación de algunos parámetros estadísticos sobre cada segmento. También están las técnicas basadas en el análisis espectral, que implican una transformación del dominio original de la señal en otro. Inicialmente

se aplicó la transformada de Fourier STFT, con la cual se obtenía información de la señal en los dominios tiempo-frecuencia (Zecca, et al., 2002; Sijiang, 2003; Reaz, et al., 2006), sin embargo y a pesar de haberse conseguido buenos resultados con esta transformada, la naturaleza estocástica y no estacionaria de las señales EMGS dio lugar a la técnicas de análisis basada en la *transformada wavelet (WT - Wavelet Transform)*, con la cual se obtiene información de la señal en los dominios tiempo-frecuencia más completa y en forma mucho más compacta que la conseguida con la transformada STFT (Sijiang, 2003).

La *representación* de una señal puede clasificarse según la técnica en dos grupos: *métodos lineales* (como la transformada de Fourier y la transformada wavelet) y *métodos cuadráticos* (como la distribución Wigner-Ville)(Englehart, 1998). Los métodos lineales consisten en la descomposición de una señal $x(t)$ en una serie de componentes llamadas *funciones base* de la siguiente manera:

$$x(t) = \sum_{k=0}^{N-1} a_k \phi_k(t), \quad (3.1)$$

donde $\phi_k(t)$ son las *funciones base* y a_k sus respectivos coeficientes. Estas funciones base deben satisfacer ciertas características como ortogonalidad y buena localización tiempo-frecuencia. Los métodos cuadráticos por su parte, se basan en la estimación del espectro de potencia instantáneo (o distribución de energía) mediante alguna operación bilineal sobre la señal, con la desventaja de implicar un alto costo computacional.

A continuación se describen algunas de las técnicas más utilizadas en la fase de *extracción de características* para conformación de vectores descriptores.

3.1.1 Análisis Temporal

Las características extraídas a partir de estadísticos temporales se evalúan sobre segmentos muy breves de tiempo, típicamente 256 ms. El mismo conjunto de características se evalúa sobre cada segmento y representa el patrón de la señal EMGS. Las características temporales con mayor capacidad discriminante introducidas inicialmente por Hudgins, et al. (1993) y posteriormente utilizadas en: Englehart, et al. (1999a, 2001); Yücel y Mehmet (2002); Betancourt, et al. (2004) y Hargrove, et al. (2005), son las siguientes:

- **Valor medio absoluto (VMA).** Estima el valor medio absoluto \bar{X}_i sobre cada segmento i de N muestras de longitud de la señal, mediante la expresión:

$$\bar{X}_i = \frac{1}{N} \sum_{k=1}^N |x_k|, \quad i = 1, 2, \dots, I, \quad (3.2)$$

donde x_k es la $k^{\text{ésima}}$ muestra en el segmento i , siendo I el número total de segmentos.

- **Valor medio absoluto de pendientes (VMAP).** Consiste en la diferencia entre las sumas de segmentos adyacentes, i e $i+1$, definido por:

$$\Delta \bar{X}_i = \bar{X}_{i+1} - \bar{X}_i, \quad i = 1, \dots, I - 1. \quad (3.3)$$

- **Cruces por cero (CC).** Es una forma simple de medir la frecuencia de la señal, puede determinarse contando el número de veces que la forma de onda cruza por cero. Se requiere un umbral para reducir el número de cruces por cero inducidos por ruido, este se selecciona con respecto al voltaje máximo de la señal (su valor típico es 0.01 del valor pico de la señal). El contador de cruces por cero se incrementa según la comparación entre dos muestras consecutivas x_k y x_{k+1} así:

$$\begin{aligned} x_k > 0 \quad y \quad x_{k+1} < 0 \quad o \quad x_k < 0 \quad y \quad x_{k+1} > 0 \\ y \quad |x_k - x_{k+1}| \geq 0.01V. \end{aligned} \quad (3.4)$$

- **Cambios de signo de pendiente (CSP).** Provee otra forma de medir el contenido de frecuencias en la señal mediante el conteo del número de veces que la pendiente de la forma de onda cambia de signo. Igualmente requiere de un umbral para reducir los cambios de signo de pendiente inducidos por el ruido. El contador de cambios de signo en la pendiente se incrementa según las relaciones:

$$\begin{aligned} x_k > x_{k-1} \quad y \quad x_k > x_{k+1} \quad o \quad x_k < x_{k-1} \quad y \quad x_k < x_{k+1} \\ y \quad |x_k - x_{k+1}| \geq 0.01V \quad o \quad |x_k - x_{k-1}| \geq 0.01V \end{aligned} \quad (3.5)$$

- **Longitud de la Forma de Onda (LO).** Provee información de la complejidad de la forma de onda en cada segmento y está dada por la siguiente expresión:

$$l_o = \sum_{k=1}^N |\Delta x_k|, \quad (3.6)$$

donde $\Delta x_k = x_k - x_{k-1}$ (diferencia entre valores de muestras consecutivas). Su valor da una idea de la amplitud, la frecuencia y su duración de la onda de la señal en un solo parámetro.

Otra técnica de análisis temporal, utilizada en: Sijiang (2003); Alonso, et al. (2002); Rodriguez y Vuskovic (2006); Yücel y Mehmet (2002) y Wang y Cui (2005), es la basada en el *modelado paramétrico Auto Regresivo (AR)*, el cual permite modelar series de tiempo mediante la expresión:

$$x(k) = - \sum_{i=1}^P a_i x(k-i) + e(k), \quad (3.7)$$

donde a_i son los coeficientes *AR*, P el orden del modelo y $e(k)$ un término de error. En este caso los coeficientes a_i forman el conjunto de características descriptoras. Hargrove, et al. (2005), fueron los primeros en clasificar señales EMGS utilizando coeficientes *AR* como características para clasificación a partir de un solo canal, sin embargo este parámetro tiene el inconveniente de suponer la condición de estacionariedad en la señal.

3.1.2 Análisis Espectral de Fourier

La mayoría de transformadas asumen que la señal a analizar es estacionaria, esto es, que las propiedades estadísticas de la señal no involucran al tiempo, siendo la *transformada de Fourier (FT - Fourier Transform)* la más utilizada. Pero esta representación es apropiada siempre que la señal tenga componentes de frecuencia estacionarios, pero cambios abruptos en el tiempo de una señal se verán dispersos en todo el rango de frecuencias de $X(f)$ sin ninguna resolución. De aquí que esta transformada no es la más indicada para señales no estacionarias. En su lugar, se consideró la STFT buscando una representación simultánea de la señal en los dominios tiempo-frecuencia con cierto grado de resolución, utilizando para ello una

función ventana $g(t)$, con la cual se “mira” a la señal en el tiempo. Así para una señal $x(t)$ la cual se supone estacionaria, se define su transformada STFT de la siguiente manera (Englehart, 1998):

$$STFT(t, f) = \int x(\tau)g^*(\tau - t)e^{-j2\pi f\tau} d\tau. \quad (3.8)$$

La resolución del análisis en el plano bidimensional tiempo-frecuencia depende exclusivamente de la ventana $g(t)$ y su respectiva transformada Fourier $G(f)$, la cual se selecciona de acuerdo a la aplicación en particular y teniendo en cuenta aspectos como: localización, eficiencia computacional y supresión de dispersiones espectrales. Así por ejemplo, una ventana tipo *gausiana*¹, brinda una buena resolución tiempo-frecuencia. Sin embargo su resolución siempre estará acotada inferiormente por la desigualdad conocida como principio de *incertidumbre de Heisenberg* expresado como:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi}, \quad (3.9)$$

lo que significa que se puede intercambiar resolución temporal por resolución frecuencial.

A pesar de la madurez teórica de la transformada STFT y sus algoritmos rápidos, no deja de producir una segmentación uniforme y cuadrículada del plano bidimensional en todo el rango temporal y espectral de la señal, con una relación fija en el producto de las resoluciones tiempo-frecuencia, lo cual no corresponde con las señales físicas. Ante esto aparece la transformada *wavelet*, con resultados sorprendentes en el análisis de este tipo de señales. Mayor información de la transformada STFT y su aplicación en el análisis de señales se presenta en el Anexo A.

3.1.3 Análisis Espectral con Wavelets

Desde la perspectiva del procesamiento de señales, la *teoría de wavelets* puede considerarse como una herramienta matemática útil para el análisis, representación o segmentación de señales. Las características propias de la *transformada wavelet* permiten representar señales no estacionarias en diferentes niveles de resolución,

¹La transformada STFT con ventana tipo Gausiana se denomina *transformada de Gabor* (Gabor, 1946).

con una buena localización temporal de las componentes de frecuencia que se presentan en la señal. Por esta razón, se considera a la *transformada wavelet* como la herramienta más apropiada para la representación de señales EMGS con fines de clasificación. A continuación se presenta una breve descripción del proceso de descomposición de una señal mediante las transformadas DWT y WPT, el cual será aplicado a las señales EMGS en la fase de extracción de características. En el Anexo A, se presenta un tratamiento más completo sobre la teoría de *wavelets*.

La *transformada wavelet* consiste en una transformación lineal, similar a la transformada de Fourier, pero con mejor resolución de análisis simultaneo en los dominios tiempo-frecuencia. Particularmente en la STFT, el tamaño fijo de la ventana de observación supone una limitación en la resolución de análisis, dado que puede resultar demasiado grande para analizar frecuencias altas o demasiado pequeña para analizar frecuencias bajas. En el caso de la *transformada wavelet* su funcionamiento se basa en el análisis “multiresolución”, el cual consiste en multiplicar la señal por funciones ventana de diferentes escalas o resoluciones llamadas *wavelets*. Así, para analizar una señal o secciones de esta de alta frecuencia se emplean ventanas *wavelets* estrechas, mientras que para analizar la señal o secciones de esta de baja frecuencia se emplean ventanas *wavelets* mas anchas (Kaiser , 1994).

El término *wavelet* se define como una “onda pequeña” o función localizable en el tiempo. Estas funciones forman *bases de wavelets* $\Psi = \{\psi_k | k \in Z\}$, con tres características fundamentales a saber (Englehart, 1998):

- i) *Las wavelets son bloques de construcción para funciones en general.* Cualquier función $f \in L^2(R)$ se puede expresar como una serie infinita de *wavelets*, de la forma:

$$f = \sum_k c_k \psi_k; \quad k \in Z, \quad (3.10)$$

donde c_k son los coeficientes wavelets y ψ_k las bases wavelets.

- ii) *Las wavelets tienen localización tiempo-frecuencia.* Localización en tiempo por ser de *soporte compacto*², y localización en frecuencia porque su transformada de Fourier resulta de banda limitada (ver figura 3.2). Además, tanto la resolución en el tiempo Δt como la resolución en la frecuencia Δf pueden variar en el plano bidimensional.

²Soporte compacto: significa que la función es no nula en el intervalo [a,b] y es nula fuera de este.

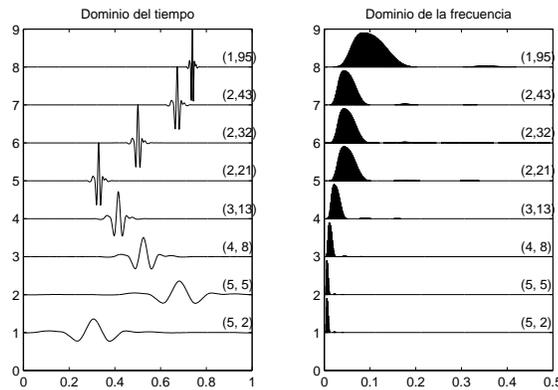


Figura 3.2. Localización temporal y limitación en banda de las funciones *wavelets*

iii) *Las wavelets tienen transformada rápida.* Su algoritmo piramidal tiene orígenes en procesamiento de imágenes y fue adaptado para el análisis de señales por Stepahne Mallat e Yves Meyer en 1986 (Mackenzie D., 2001). Consiste en una serie de filtros lineales pasa-bajas y pasa-altas que descomponen la señal según sus componentes de bajas y altas frecuencias. Introduce la operación de *submuestreo* en el decimado de la señal, lo cual acelera el algoritmo al reducir el número de operaciones geoméricamente con cada iteración. De esta manera, la complejidad computacional de la *transformada wavelet* será de $O(CN)$ operaciones, donde C depende de la *wavelet* seleccionada y N es el número de muestras, mientras que para transformada de Fourier es del orden de $O(N \log N)$ (Englehart, 1998).

En la *teoría de wavelets*, desarrollada inicialmente por Grossmann y Morlet (1984), se pueden encontrar diferentes formas de transformaciones *wavelets*, y según la aplicación específica se puede seleccionar una u otra. Entre estas se encuentran principalmente: la *transformada wavelet continua* (CWT - *Continuous Wavelet Transform*), la *transformada wavelet discreta* (DWT - *Discrete Wavelet Transform*) y la *transformada de paquetes wavelet* (WPT - *Wavelet Packet Transform*). A continuación, se describe brevemente cada una de estas, haciendo mayor énfasis en las transformadas DWT y WPT por su aplicación en este trabajo.

Transformada wavelet continua CWT (Burrus, et al. , 1999).

Esta transformada permite un cubrimiento variable del plano tiempo-frecuencia y se define como:

$$CWT_x(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad a, b \in R; a \neq 0, \quad (3.11)$$

donde la continuidad de la transformada radica en que la variable de *escala* a y la variable de *traslación* b son continuas. La función ventana (o prototipo) $\psi(t)$ se denomina *wavelet madre* y el análisis *wavelet* determina la correlación entre la señal de análisis y las versiones *escaladas* y *trasladadas* de la función ventana. La implementación digital de la transformada *CWT* puede hacerse mediante la integral de convolución, sin embargo, esta produce demasiados coeficientes con alta tasa de información redundante debido a la no ortogonalidad de sus funciones base. Por lo que resulta más práctico a nivel computacional implementar su versión discretizada como se presenta a continuación.

Transformada wavelet discreta (DWT) (Burrus, et al. , 1999).

La discretización de la transformada *CWT* consiste en asociar valores discretos a las variables de *escala* a y *traslación* b según una escala diádica, esto es, haciendo $a = 2^j$ y $b = n2^j$, de tal manera que el conjunto de *funciones wavelets*:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi \left(\frac{t-b}{a} \right); \quad a, b \in R, \quad (3.12)$$

adquieren la forma

$$\psi_{j,n}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - n); \quad j, n \in Z, \quad (3.13)$$

donde el factor $2^{-j/2}$ se introduce para garantizar la condición de ortonormalidad de las *bases wavelets*. Con esto se logra mayor eficiencia computacional y una representación de la señal más compacta, pues la ortogonalidad de las bases evitan la redundancia entre sus coeficientes.

Una función *wavelet* $\psi(t)$ tiene asociada consigo una función *escala* $\phi(t)$, y la *DWT* permite un análisis multiresolución basado en la diferencia de información que existe entre versiones de la misma señal en dos resoluciones

sucesivas (2^j y 2^{j+1}). A esta diferencia de información se le denomina “detalles” de la señal, de tal manera que es posible aproximar cualquier función $f(t) \in L^2(\mathbb{R})$ en términos de estas funciones mediante la expresión:

$$f(t) = \sum_n \sum_j d_{j,n} \psi_{j,n}(t) + \sum_n c_{J,n} \phi_{J,n}(t) \quad j, n \in \mathbb{Z}, \quad (3.14)$$

donde $d_{j,n}$ representan los coeficientes de “detalles” y $c_{j,n}$ representan los coeficientes de “aproximaciones” en la descomposición *wavelet DWT* de la señal. Estos coeficientes se pueden obtener mediante etapas de filtrado sucesivo a través de “bancos de filtros” pasa-altas y pasa-bajas respectivamente.

Antes de describir el proceso de la descomposición DWT, es importante anotar que una función *wavelet madre* puede tener muchas formas, y que el término “escala” está íntimamente relacionado con la frecuencia. Por tal razón, cuando se trata con la transformada *wavelet*, se puede hablar de resolución en frecuencia sin que eso implique haber dejado de lado el concepto de escala. Para ilustrar esto, se puede considerar una función *wavelet madre* de la forma:

$$\psi(t) = g(t)e^{-j2\pi f_0 t}, \quad (3.15)$$

donde $g(t)$ es la ventana de análisis de la STFT y f_0 la frecuencia fundamental de análisis STFT. Esto implica que la frecuencia análoga a la escala de la DWT es $f = af_0 = 2^j f_0$, lo que significa que el ancho de banda de la ventana de análisis ortogonal para la DWT es función de su frecuencia central f_0 y aumenta con la escala (Englehart, 1998). Según esto la relación ancho de banda a frecuencia central C , dada por

$$C = \frac{\Delta f}{f} = \frac{2^j}{2^j f_0}, \quad (3.16)$$

se mantiene constante, lo que significa que el ancho de banda de las ventanas de análisis para la DWT se distribuye logarítmicamente con la frecuencia y no linealmente como en la STFT, como se muestra en la figura 3.3.

Descomposición DWT (Saito , 1994). Suponiendo una señal $x = x[n]$ en el espacio original $\Omega_{0,0}$ de N muestras, la descomposición *DWT* de primer

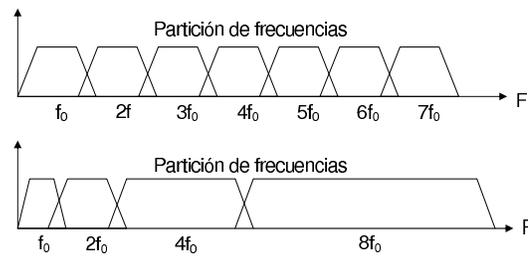


Figura 3.3. Partición del espectro según la STFT (arriba) y según la DWT (abajo)

nivel consiste en someter la señal original a dos operadores G y H separadamente. El primero consiste en un filtraje tipo pasa-altas a través de $g = \{g[l], l = 0, 1, \dots, L - 1\}$ y el segundo en un filtraje tipo pasa-bajas a través de $h = \{h[l], l = 0, 1, \dots, L - 1\}$. Así, el espacio original $\Omega_{0,0}$ queda representado por los nuevos espacios $\Omega_{1,1}$ y $\Omega_{1,0}$ respectivamente. Este proceso se muestra en la figura 3.4, donde el operador “ $\downarrow 2$ ” implica una decimación por 2.

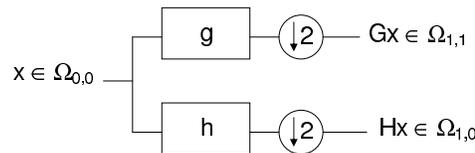


Figura 3.4. Descomposición DWT a un nivel ($J=1$)

Toda esta operación se puede expresar como:

$$(Hx)_n = \sum_{k=0}^{L-1} h[k]x[2n - k] \quad y \quad (Gx)_n = \sum_{k=0}^{L-1} g[k]x[2n - k], \quad n = 0, 1, \dots, N - 1 \quad (3.17)$$

Esto significa que las “*aproximaciones*” de la señal se obtienen mediante filtros *pasa-bajas* y los “*detalles*” de la señal mediante filtros *pasa-altas*. Se puede observar también que si la señal a analizar consiste en un vector de N muestras de longitud, las secuencias Gx y Hx son vectores de longitud $N/2$.

Una segunda descomposición DWT de la señal consistirá en aplicar el mismo proceso de filtraje sobre los coeficientes de la banda de bajas frecuencias Hx para obtener las nuevas secuencias H^2x y GHx de longitud $N/4$ cada una.

Este mismo proceso puede iterarse hasta un nivel $J \leq \log_2 N$, con lo que se consigue un vector de secuencias de coeficientes wavelets DWT de la forma $\{Gx, GHx, GH^2, \dots, GH^{J-1}, H^J x\}$. En la figura 3.5 se muestra el proceso de descomposición para el caso de una descomposición a nivel tres ($J = 3$), para el cual se obtiene las secuencias de coeficientes $\{D1, D2, D3, A3\}$.

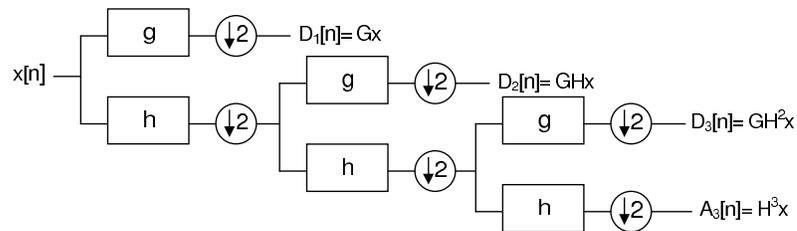


Figura 3.5. Descomposición DWT a tres niveles ($J=3$)

Una manera práctica de comprobar la capacidad de la *transformada wavelet* en detectar los cambios en una señal, debidos a su no estacionariedad, es aplicando la descomposición DWT sobre una sección de 256 muestras de la señal “chirp lineal” muestreada a una frecuencia de 1000 muestras por segundo. En la figura 3.6 se observa: a la izquierda, la descomposición temporal de la señal según las secuencias de coeficientes $\{D1, D2, D3, A3\}$ que resultan de la descomposición DWT de la señal a nivel 3, y a la derecha las respectivas subbandas con las componentes de frecuencia presentes en cada porción temporal de señal.

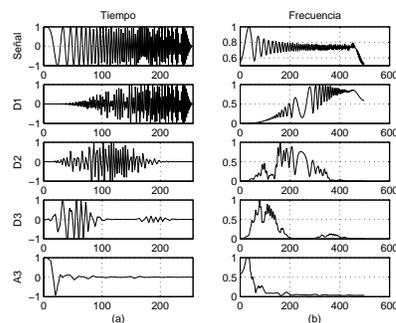


Figura 3.6. Descomposición DWT a tres niveles de la señal *chirp lineal*

Transformada de Paquetes Wavelets WPT (Burrus, et al. , 1999)

La *transformada de paquetes wavelets* WPT es una generalización de la transformada DWT, donde la descomposición en cada nivel no se limita solamente a las bandas de aproximación, figura 3.7.a, sino también sobre las bandas de detalles respectivas, figura 3.7.b. De esta manera, una señal $x[n]$ con N muestras de longitud genera en el primer nivel de descomposición, mediante filtraje pasa-bajas y pasa-altas, dos subbandas de longitud $N/2$ conteniendo la mitad del espectro de la señal cada una (igual que la descomposición DWT). En el segundo nivel de descomposición el filtraje se realiza sobre estas dos subbandas, generando cada una dos subbandas más, obteniéndose cuatro subbandas de longitud $N/4$ cada una, conteniendo cada una $1/4$ del espectro de frecuencias. Así, del tercer nivel de descomposición resultan ocho (8) subbandas con los coeficientes *wavelets* identificados como $\{W_{3,0}, W_{3,1}, \dots, W_{3,7}\}$ (Burrus, et al. , 1999).

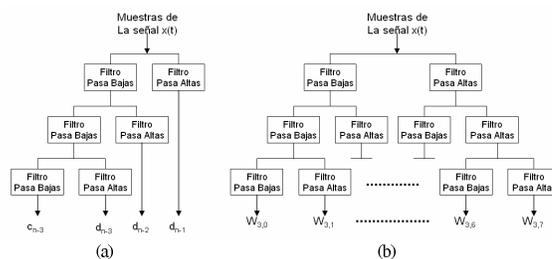


Figura 3.7. Descomposición *wavelet* a tres niveles: a) DWT, b) WPT

Este proceso de descomposición puede realizarse hasta un nivel máximo $J \leq \log_2 N$, resultando en JN coeficientes *wavelets* en total, y a pesar de que este procedimiento de descomposición es más completo que el de la descomposición DWT, su costo computacional aún resulta ser menor que el requerido por la transformada STFT, esto es $O(JN) \leq O(N \log_2 N)$ (Wickerhauser, 1994).

Retomando el ejemplo de la señal “*chirp lineal*” en la cual la frecuencia aumenta linealmente con el tiempo, se puede comprobar que el análisis WPT permite detectar estos incrementos de frecuencia con mejor detalle y resolución que con el análisis DWT, figura 3.8.a, donde la localización tiempo-frecuencia de cada subbanda es mucho mejor, puesto que en cada subbanda los coeficientes mas significativos son localizados temporalmente justo en el intervalo donde la señal pasa a través del rango de frecuencias correspondientes a la

subbanda. En la figura 3.8.b, se observa la presencia de coeficientes significativos en cada subbanda de la descomposición para cada nivel, lo que significa que cada subbanda tiene componentes de frecuencia significativas.

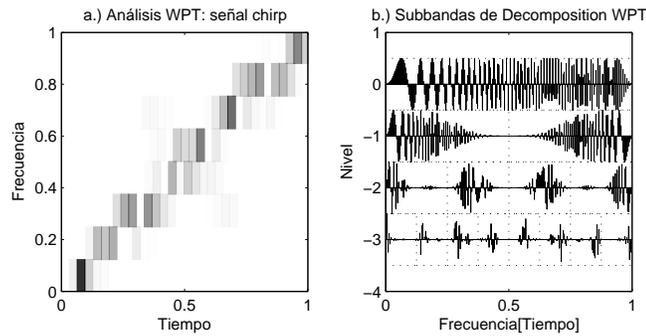


Figura 3.8. Análisis de la señal “*chirp lineal*” mediante descomposición WPT

Sin embargo, en aplicaciones prácticas no siempre será necesario considerar todas las subbandas resultantes de la descomposición WPT ya que la distribución de la energía en la señal no siempre será uniforme. En este caso, interesa únicamente analizar y procesar aquellas subbandas del árbol de descomposición WPT que concentren la mayor cantidad de energía o información relevante para la extracción de características descriptoras de la señal. El problema ahora es entonces encontrar la “*mejor base*” de representación de las señales EMGS para los propósitos de clasificación de patrones mioeléctricos.

Para determinar la *mejor base* de las $2(2^J - 1)$ bases ortonormales posibles, es necesario evaluar y comparar su eficacia según la aplicación específica. El algoritmo de *selección de la mejor base* propuesto por Coifman y Wickerhauser (1992), el cual se utiliza en este trabajo, se basa en un proceso de podado de las ramas del árbol binario WPT. Partiendo desde el nivel más bajo se comienza a eliminar ciertas ramas que conduzcan a encontrar la solución óptima para la aplicación específica. La función de costo que se asocia al algoritmo de podado se basa en una medida de entropía, con la cual se busca maximizar la cantidad de información a partir del conjunto de bases seleccionadas. La medida más natural de entropía es la de *Shannon*:

$$H(\mathbf{p}) \doteq \sum_i p_i \log_2 p_i, \quad (3.18)$$

donde, $\mathbf{p} = \{p_i\}$ es una secuencia no negativa tal que $\sum_i p_i = 1$. Mayor información sobre el algoritmo de selección de la mejor base de Coifman-Wickerhauser se encuentra en el Anexo A.

Finalmente se puede observar que, según el tipo de transformada que se utilice, se obtiene una resolución de análisis y una forma particular de partición del plano bidimensional tiempo-frecuencia. Así, mientras que con la transformada STFT se consigue una partición uniformemente cuadrículada, figura 3.9.a, con la transformada DWT se consigue una partición no uniforme y rectangular, con buena resolución temporal para altas frecuencias y buena resolución en frecuencia para bajas frecuencias, figura 3.9.b, y con la transformada WPT se puede configurar una partición del plano tiempo-frecuencia no uniforme (cuadrícula configurable), esto es, analizar componentes o rangos de frecuencia al nivel de resolución que se desee, figura 3.9.c. No obstante, las ventajas que ofrece la transformada WPT sobre las otras dos implica también: un mayor costo computacional, una mayor complejidad de procesamiento y una mayor dimensionalidad del espacio transformado.

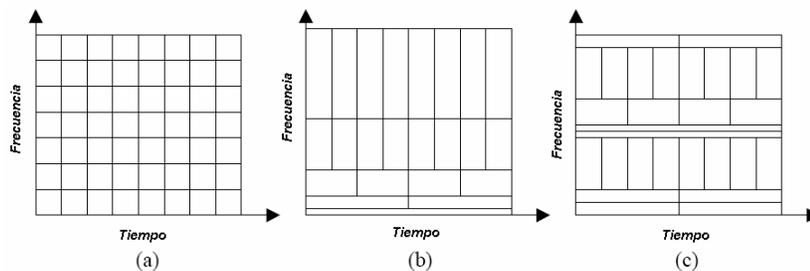


Figura 3.9. Resolución del análisis espectral mediante a) STFT, b) DWT y c) WPT

3.1.4 Extracción de características basadas en wavelets

Con el análisis espectral basado en *wavelets* se busca extraer características discriminantes cuya información esté contenida simultáneamente en el plano bidimensional tiempo-frecuencia, información que con un simple análisis temporal o espectral de Fourier sería imposible detectar. Esto se puede lograr directamente a partir

de los *coeficientes wavelets* o procesándolos antes de extraer las características deseadas. A continuación se presentan los métodos más destacados en este proceso, los cuales serán objeto de análisis y comparación para determinar entre ellos la mejor opción para la representación y discriminación de señales EMGS.

- **Método 1. Cuadrado de los coeficientes wavelets.** Consiste en tomar el cuadrado de los coeficientes *wavelets* y someterlos a las técnicas de reducción de dimensionalidad, bien sea por *selección de características* de acuerdo al índice de separabilidad de clases CSI, o por *proyección de características* mediante el análisis de componentes principales PCA, como lo propuso Englehart (1998).
- **Método 2. Descomposición de valores singulares.** Consiste en aplicar la *descomposición de valores singulares SVD* a la matriz de coeficientes *wavelets* DWT, la cual contiene en sus filas las secuencias de coeficientes *wavelets* que resultan de la descomposición DWT de la señal original como se muestra en la figura 3.10. Siendo A la matriz de *coeficientes wavelets*, los valores singulares se pueden determinar como la raíz cuadrada de los valores propios de $A^T A$, los cuales son utilizados como características de descripción (Zhang, et al., 2002).

La extracción de características a partir de la descomposición SVD fue una técnica utilizada inicialmente por Zhang, et al. (2002), quienes realizaron la descomposición SVD sobre la matriz de coeficientes conformada por las subbandas de la descomposición DWT de nivel 4. Posteriormente esta técnica fue utilizada en trabajos de investigación en clasificación de patrones como en Zhang, et al. (2004); Zecca, et al. (2002), gracias a su buen desempeño como características descriptoras de las señales mioeléctricas.

- **Método 3. Evaluación de parámetros.** Consiste en extraer un conjunto de características a partir de la evaluación de los siguientes parámetros (Zhang, et al. , 2004):

i) *El promedio absoluto de coeficientes:*

$$AVG_j = \frac{1}{K} \sum_{k=1}^K |c_{jk}|, \quad (3.19)$$

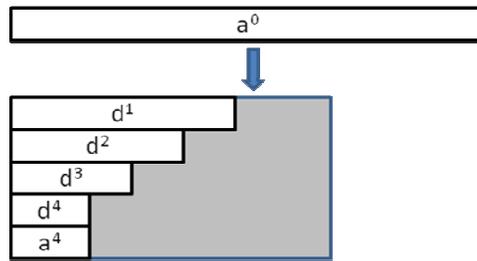


Figura 3.10. Conformación de la matriz de coeficientes *wavelet* DWT

donde c_{jk} es el k -ésimo coeficiente *wavelet* de la escala j y K el número de coeficientes.

ii) *La energía de escala:*

$$ENER_j = \frac{1}{2} \sum_k c_{jk}^2, \quad (3.20)$$

donde c_{jk} es el k -ésimo coeficiente *wavelet* de la escala j .

iii) *El máximo en la escala:*

$$MAX_j = \max_j (|c_{jk}|), \quad (3.21)$$

donde c_{jk} es el k -ésimo coeficiente *wavelet* de la escala j .

iv) *Valores singulares:*

$$SVD = \sqrt{\lambda(C^T C)}, \quad (3.22)$$

donde $\lambda(C^T C)$ significa las raíces características de la matriz $(C^T C)$ (C es una matriz cuyas filas se forman con las respectivas subbandas de coeficientes *wavelets*).

v) *Coefficientes Autoregresivos AR:*

$$x_n = - \sum_{k=1}^p a_k x_{n-k} + e_n, \quad (3.23)$$

donde x_n es la señal temporal a analizar, a_k el k -ésimo coeficiente del modelo AR, e_n es ruido blanco y p el orden del modelo AR. La desventaja de esta técnica es que supone la estacionariedad de la señal, lo cual no corresponde con la naturaleza de las señales EMGS.

- **Método 4. Evaluación de momentos.** Es una de las formas más recientes de extracción de características basadas en *wavelets*, y consiste en la evaluación

de los momentos M_m sobre las secuencias de *coeficientes wavelets* dados por la expresión:

$$M_m = \sum_{n=0}^{N_j-1} \left(\frac{n}{N_j} \right)^{m-1} S[n] \quad m = 1, 2, .. \quad (3.24)$$

donde $S[n]$ son las secuencias de coeficientes en cada banda de la descomposición *wavelet*. Rodriguez y Vuskovic (2005) evaluaron inicialmente los tres (3) primeros momentos $m = 1, 2, 3$ y posteriormente en 2006, utilizaron los seis (6) primeros momentos $m = 1, 2, 3, 4, 5, 6$, mejorando aun más la clasificación.

Excepto por el método de descomposición SVD (aplicable solo para la descomposición DWT), los métodos descritos aplican para las técnicas DWT y WPT, con la diferencia de que para la descomposición WPT, antes de evaluar las características respectivas, se somete la estructura completa del árbol de descomposición a algún algoritmo de *selección de la mejor base*. Este algoritmo seleccionará las mejores subbandas de la descomposición según su contenido de información discriminante, sobre las cuales se procede a extraer las características.

3.2 Reducción de Dimensionalidad

Dependiendo de la naturaleza del problema de clasificación (tipo de datos, conjunto de características y el clasificador seleccionado), reducir la dimensión de los vectores descriptores puede ser fundamental para el buen desempeño del clasificador. En el diagrama esquemático de la figura 3.11 se identifican las dimensiones respectivas de los espacios resultantes en cada fase de procesamiento. El espacio de entrada que contiene los registros de la señal $\mathbf{x} \in X \subseteq \mathbb{R}^N$ y el espacio de la respuesta del sistema $y \in Y \subseteq \{y_1, \dots, y_k\}$ ya fueron definidos en la sección 3.1.

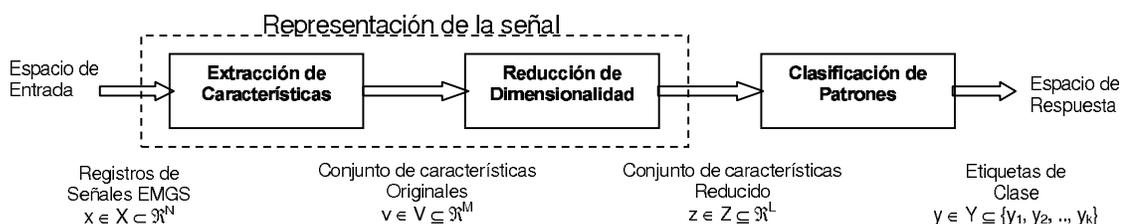


Figura 3.11. Diagrama esquemático del sistema de clasificación

De la fase de extracción de características se obtiene un conjunto de características $\mathbf{v} \in V \subseteq R^M$, que puede consistir en las muestras temporales de la señal, los coeficientes de alguna transformación o la evaluación de parámetros discriminantes (temporales o espectrales). En cualquier caso, cuando su dimensión de este conjunto original resulte demasiado grande, lo más recomendable es aplicarle alguna técnica de *reducción de dimensionalidad* para obtener un nuevo conjunto de características $\mathbf{z} \in Z \subseteq R^L$ de menor dimensión, con $L < M$.

Con la reducción de la dimensionalidad se busca principalmente preservar la información más relevante para la discriminación, desechando la redundancia; adicionalmente se consigue reducir la carga computacional del clasificador, puesto que entre más reducido sea el tamaño del vector descriptor que entra al clasificador, menor será el número de parámetros de adaptación a determinar y mayor será su capacidad de generalización. Así, mientras más eficiente sea el procesamiento, más trivial será la tarea del clasificador (Englehart, et al., 1999a). Las técnicas de reducción de dimensionalidad se clasifican en dos categorías: por **selección de características** y **proyección de características**.

3.2.1 Selección de características

Esta técnica busca reducir la dimensión del espacio que contiene al conjunto de características originales mediante la selección de algunas de estas y la eliminación de otras, buscando el mejor *subconjunto* de características de acuerdo a algún criterio de selección. Este procedimiento involucra: primero, establecer un criterio mediante el cual sea posible juzgar cuando un subconjunto de características es mejor que otro, y segundo, seguir un procedimiento sistemático en la búsqueda del mejor subconjunto. Idealmente el criterio de selección debería ser el de la *probabilidad de error de clasificación*, pero en la práctica este criterio resulta demasiado complejo y solo queda aplicar criterios más simples como el de *separabilidad de clases*.

En una situación ideal, el procedimiento de búsqueda debería ser exhaustivo con todos los subconjuntos posibles, pero estos procedimientos resultan ser complejos y de gran carga computacional. Ante esto se opta por manejar criterios simples y procedimientos de búsqueda sencillos que no siempre resultarán en subconjuntos óptimos (Fukunaga , 1990). Un criterio sencillo y muy utilizado de reducción de dimensionalidad por selección de características, y que será utilizado en este trabajo,

es el criterio de *índice de separabilidad de clases (CSI - Class Separability Index)* (Englehart, et al., 1999a), el cual se fundamenta en la separabilidad de clases por “distancia Euclídea” y es considerado un procedimiento supervisado.

Criterio del Índice de separabilidad de clases CSI.

Para explicar este criterio considérese un problema de dos (2) clases de patrones, con un conjunto de características originales representadas en un vector $\mathbf{v} \in V \subseteq R^M$. Y para simplificar el problema, se va a suponer que el subconjunto candidato contiene una *única característica* a la vez, en lugar de una combinación de estas. Esto implica aplicar el criterio de selección a cada característica individualmente y ordenar las M características en términos de su capacidad discriminante. Al final, las $L \leq M$ características más discriminantes serán seleccionadas para conformar el subconjunto de características buscado (Englehart, 1998).

Sean p y q las dos clases, cada una con P y Q patrones respectivamente, organizados en forma matricial así:

$$[\mathbf{v}^{(p,1)} \mathbf{v}^{(p,2)} \dots \mathbf{v}^{(p,P)}] \quad y \quad [\mathbf{v}^{(q,1)} \mathbf{v}^{(q,2)} \dots \mathbf{v}^{(q,Q)}], \quad (3.25)$$

donde el m -ésimo patrón de la clase p se denota por:

$$\mathbf{v}^{(p,m)} = [\mathbf{v}_1^{(p,m)} \mathbf{v}_2^{(p,m)} \dots \mathbf{v}_M^{(p,m)}]^T. \quad (3.26)$$

La capacidad discriminante se evalúa sobre cada característica, esto es, la discriminabilidad de las características i -ésimas denotadas como p_i y q_i respectivamente se aplica a la fila i -ésima de cada matriz. Esta medida se denota como $D(p_i, q_i)$ y representa qué tanta separabilidad hay entre p_i y q_i .

En este trabajo, se evalúa el índice CSI ³ según la expresión (Zhang, et al., 2004):

$$CSI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{ij}) = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \left(\frac{S_i + S_j}{D_{ij}} \right), \quad (3.27)$$

donde K es el número de clases, R_{ij} es el índice de similaridad entre las clases i y j , S_i y S_j describen el grado de dispersión entre las clases i y j , y D_{ij} es la distancia media entre las clases i y j . Estas medidas de dispersión y distancia pueden evaluarse de la siguiente manera:

³Índice CSI, conocido también como índice de Davies-Bouldin

$$S_i = \left[\frac{1}{N_i} \sum_{j=1}^{N_i} (x_j - m_i)^T (x_j - m_i) \right]^{1/2} \quad (3.28)$$

y

$$D_{ij} = \left[(m_i - m_j)^T (m_i - m_j) \right]^{1/2}, \quad (3.29)$$

donde N_i representa el número de muestras de la clase i , x_j la j -ésima muestra de entrada en la clase i , y m_i es la media de las muestras de entrada de la clase i .

Ahora si la medida discriminante es aditiva, se puede evaluar para subconjuntos como:

$$D(\{p_i, p_j, p_k\}, \{q_i, q_j, q_k\}) = D(p_i, q_i) + D(p_j, q_j) + D(p_k, q_k). \quad (3.30)$$

Por tanto, se puede definir un conjunto de M características como $\mathbf{p} = \{p_i\}_{i=1}^M$ y evaluar la discriminabilidad de estos arreglos de características como:

$$D(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^M D(p_i, q_i). \quad (3.31)$$

Y para el caso de más de dos clases se cambia la notación de una característica por $p_i^{(k)}$, $k = 1, \dots, K$, para denotar la i -ésima característica de la clase k -ésima de las K clases, y por $\mathbf{p}^{(k)} = \{p_i^{(k)}\}_{i=1}^M$ al conjunto de M características. Entonces, la medida de discriminación D entre la i -ésima característica de cada clase se obtiene al considerar todas las posibles combinaciones de parejas mediante la expresión

$$D\left(\left\{p_i^{(k)}\right\}_{i=1}^K\right) \doteq \sum_{m=1}^{K-1} \sum_{n=m+1}^K D\left(p_i^{(m)}, p_i^{(n)}\right). \quad (3.32)$$

De igual manera para un arreglo de características se obtiene la medida

$$D\left(\left\{\mathbf{p}^{(k)}\right\}_{k=1}^K\right) \doteq \sum_{m=1}^{K-1} \sum_{n=m+1}^K D\left(\mathbf{p}^{(m)}, \mathbf{p}^{(n)}\right). \quad (3.33)$$

El procedimiento de selección de características debería considerar todos los posibles subconjuntos de características, pero esto resulta en un proceso demasiado complejo, aunque dichas combinaciones puedan proveer información conjunta que aisladamente no pueda detectarse. Y al contrario, dos características que individualmente guardan información significativa pueden estar tan correlacionadas que juntas contribuyan con menos información que si se tratan aisladamente (Englehart,

1998). Idealmente debería utilizarse como criterio, el error de clasificación, pero llevar el proceso de selección hasta la fase de clasificación se tornaría demasiado dispendioso y complejo. En su lugar y en particular para este trabajo, se utiliza el índice *CSI* como herramienta de selección de aquellas características que se suponen con mayor capacidad discriminante, bajo la premisa de que los criterios de separabilidad son muy útiles para comparar conjuntos de características de igual tamaño. La mayor desventaja de esta técnica es que algunas de las características originales y por tanto la información que contienen serán desechadas definitivamente. Esto es un problema si la información de separabilidad de clases se encuentra dispersa en todo el conjunto de características original. Por esta razón, el método de selección de características no es adecuado en situaciones de alta dimensionalidad con datos altamente redundantes (caso de señales grabadas y no procesadas). A menos que los datos tengan un buen comportamiento, la información en espacios de gran dimensión tiende a dispersarse y la probabilidad de que unas pocas características contengan la mayor información de separabilidad de clases es pequeña. Esto implica que los métodos de selección de características son indicados en conjuntos de características debidamente identificadas o que concentran las características invariantes del espacio de entrada.

3.2.2 Proyección de Características

Esta técnica consiste en determinar la mejor *combinación*, a partir de la transformación del conjunto original de características, para obtener un nuevo conjunto pero de menor dimensión. La técnica más conocida y eficiente de reducción de dimensionalidad por proyección consiste en el *análisis de componentes principales (PCA-Principal Component Analysis)*, la cual produce un conjunto de características decorrelacionadas mediante la proyección de los datos sobre los vectores propios de la matriz de covarianza. Es un procedimiento no supervisado en razón de que ningún criterio de pertenencia de clase califica los datos cuando se especifican los vectores propios (Englehart, et al., 1999a).

Análisis de Componentes Principales PCA.

Consiste en una transformación lineal utilizada a menudo para extracción de características o compresión de datos. En reconocimiento de patrones la efectividad

de *PCA*⁴ está en eliminar dependencias lineales y ruido no correlacionado en los datos (Huber , 1985). Por otra parte, descarta aquellos términos que tienen varianzas pequeñas y retiene solo aquellos términos que tienen varianzas mayores (Kocyigit y Korurek , 2002). En compresión de señales el análisis *PCA* provee un mapeo lineal que ayuda a minimizar el criterio del *mínimo error cuadrático medio* (*MMSE - Mean Minimun Squared Error*). Así, entre más pequeño sea el conjunto de características que permitan reconstruir de manera aceptable la señal original más efectivo será dicho conjunto.

Para un conjunto de P vectores descriptores (o patrones de una clase) $\mathfrak{S} \equiv [\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(P)}]$, donde cada vector se compone de M características $\mathbf{v}^{(p)} = [v_1^{(p)} \dots v_M^{(p)}]^T$, el objetivo del análisis *PCA* es mapear cada vector original $\mathbf{v}^{(p)} \in V \subseteq R^M$ en un nuevo vector de dimensión reducida $\mathbf{z}^{(p)} = [z_1^{(p)} \dots z_L^{(p)}]^T \in Z \subseteq R^L$, $L < M$, de acuerdo a algún criterio como el MMSE. El algoritmo *PCA* se describe con mayor detalle en el anexo B.

En la práctica, el algoritmo para *PCA* procede sustrayendo del conjunto de datos su media $\bar{\mathbf{v}} = \sum_{p=1}^P \mathbf{v}^{(p)}$, para luego evaluar la matriz de covarianza y sus autovalores. Los autovectores correspondientes a los L autovalores mas grandes son retenidos y el conjunto de vectores descriptores original \mathfrak{S} es proyectado sobre los L autovectores para reducir la dimensión del conjunto original M a uno de dimensión $L < M$.

En el contexto de reconocimiento de patrones, los elementos del vector $\mathbf{z}^{(p)}$ son vistos como el conjunto reducido de características. Este nuevo conjunto tiene las siguientes características (Fukunaga , 1990):

i) La efectividad de cada característica, en términos de la representación del vector \mathbf{v} , está determinada por su correspondiente autovalor. Las características son ordenadas y aquellas con los autovalores más grandes conforman el conjunto de características.

ii) Los valores de las características son mutuamente decorrelacionados.

⁴En teoría de comunicaciones *PCA* se conoce como *Transformación Karhunen-Loéve*

3.3 Clasificación de Patrones

La clasificación de patrones es clave para muchas aplicaciones con procesamiento de señales. Una buena clasificación está supeditada a un adecuado proceso de representación de señales, lo que significa que su desempeño dependerá fuertemente de la relevancia de las características extraídas. La clasificación de patrones según los métodos utilizados se pueden agrupar en tres grandes categorías: Los clasificadores *estadísticos* (o de decisión teórica), los *estructurales* (o sintácticos) y los de *aprendizaje* (o neuronales).

Los *clasificadores estadísticos* se basan en el análisis de las características estadísticas de los datos. Los datos se asignan a una clase particular de acuerdo a un modelo probabilístico de su función de densidad de probabilidad en un espacio N-dimensional, el cual se divide en regiones correspondientes a cada clase. Los clasificadores más difundidos en esta clase, por su precisión en la clasificación, se encuentra el Bayesiano y el analizador discriminante lineal (LDA), sin embargo, sus algoritmos tiene un costo computacional muy elevado.

Los *clasificadores estructurales o sintácticos* se basan en la utilización de la estructura de patrones y la interrelación entre componentes de un patrón. El reconocimiento de patrones estructurales o sintácticos involucra la identificación de componentes principales o “primitivas” de los patrones y el desarrollo de una sintaxis formal o “gramática” que describa la síntesis de los patrones a partir de sus primitivas, lo que los hace ver más como una forma de representación de señales que como una tarea de clasificación (Fukunaga , 1990).

Y los *clasificadores de aprendizaje o neuronales*, utilizados en este trabajo, se basan en redes neuronales artificiales, son considerados como determinísticos debido a que un algoritmo de aprendizaje no asume nada de las características estadísticas de los patrones de una clase (como lo hacen los estadísticos), a pesar de ser los dos muy similares en su forma y objetivos (Fukunaga , 1990). Para efectos de evaluar el proceso de representación de las señales EMGS mediante el error de clasificación, se implementaron como clasificadores: el *perceptrón multicapas* (MLP - *Multi Layer Perceptron*), una *red neuronal probabilística* (PNN - *Probabilistic Neuronal Network*) y una *máquina soporte vectorial* (SVM - *Support Vector Machine*).

La mayoría de los clasificadores tienen sus principios de operación en la clasificación estadística y particularmente en la regla de clasificación de Bayes (mayor

información sobre el clasificador de Bayes se presenta en el anexo C), la cual busca minimizar la probabilidad de error de clasificación asignando un patrón dado a la clase cuya probabilidad *a posteriori* $P(y_k|\mathbf{x})$ sea la mayor. En este sentido el clasificador MLP es capaz de aproximar esta probabilidad *a posteriori* mediante funciones discriminantes parametrizadas, ajustando sus parámetros mediante un proceso de aprendizaje a partir de los datos preclasificados, mientras que los clasificadores PNN y SVM son redes capaces de estimar las distintas funciones de densidad de probabilidad que permitan caracterizar probabilísticamente el problema y diseñar el clasificador óptimo (Rosa y Gil , 2006). A continuación se presentan las características de clasificación de cada uno.

3.3.1 El Clasificador MLP

El perceptrón multicapas MLP se explica desde la teoría de las redes neuronales ANN, que son una sistematización de la estructura de las redes neuronales biológicas y de sus características de aprendizaje. Una red ANN se considera un clasificador de entrenamiento supervisado que requiere de un conjunto de datos de *entrenamiento* \mathfrak{S} , con P pares de muestras del espacio de entrada y sus respectivas etiquetas de clase. Su desempeño y su capacidad de generalización en la clasificación se evalúa mediante un conjunto de *datos de prueba* $\tilde{\mathfrak{S}}$.

El perceptrón multicapas MLP, está formado por una o varias capas de neuronas de tipo *perceptrón*⁵, dispuestas de forma secuencial de modo que las salidas de las neuronas de una capa son las entradas de las neuronas de la capa siguiente, figura 3.12.

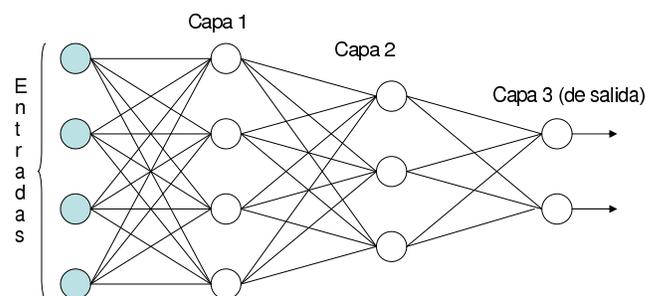


Figura 3.12. Arquitectura del Perceptrón Multicapas MLP

⁵Perceptrón: modelo de sistema capaz de aprender mediante entrenamiento supervisado, contiene un conjunto de pesos ajustables y una función de activación tipo escalón.

Cada neurona utiliza una función de base *lineal*, que consiste en una suma “pesada” de acuerdo a los pesos w_i , de las N componentes del vector de entrada $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$, a la cual se le agrega un valor inicial o “sesgo” de peso $w_0 = \Theta$, según la expresión:

$$s(x) = \sum_{i=1}^N w_i x_i + \Theta = \mathbf{w}^T \mathbf{x} + \Theta. \quad (3.34)$$

Esta función de base *lineal* define un hiperplano en el espacio de entrada. El resultado de la suma $s(x)$, se pasa por una función de activación *no lineal* $f(\bullet)$, cuyo valor es proporcional a la distancia del vector de entrada \mathbf{x} al hiperplano. La salida u , puede ser una etiqueta de clase o un nuevo valor de entrada para la siguiente capa de la red, de acuerdo a la expresión:

$$u = f(s(x)). \quad (3.35)$$

Las funciones de activación *no lineales* más empleadas son las *sigmoides*, entre las que se encuentran: la función *logarítmica* y la *tangente hiperbólica*, dadas por las expresiones

$$f_{\log}(s) = \frac{1}{1 + e^{-\beta s}} \quad f_{\tanh}(s) = \frac{e^{\beta s} - e^{-\beta s}}{e^{\beta s} + e^{-\beta s}}, \quad (3.36)$$

respectivamente.

Esta configuración permite una partición del espacio con fronteras de decisión arbitrariamente complejas. Este es el caso del clasificador *MLP* con algoritmo de aprendizaje en modo “*backpropagation*”, el cual opera bajo la condición de que las funciones de activación sean diferenciables para poder ajustar los valores de los pesos.

Así pues, para un vector $\mathbf{x} \in R^N$, una red de N entradas, V capas con W_v neuronas en la capa v^{esima} y por tanto $M = W_v$ salidas, representa una función $F(x) : R^N \rightarrow R^M$. El vector de salida y_1 de la primera capa puede expresarse como (Rosa y Gil, 2006):

$$y_1 = f_1(\Omega_1 \mathbf{x} + \Theta_1), \quad (3.37)$$

donde Ω_1 es una matriz con N columnas y W_1 filas, que contiene los pesos de las neuronas de la primera capa, Θ_1 es un vector columna con los valores iniciales o

sesgos de las neuronas de la primera capa y $f_1(\bullet)$ es una función sin memoria que comprende las W_1 funciones de activación de las neuronas de la primera capa.

Para simplificar el análisis, la salida de la red *MLP* de dos capas $V = 2$ tendrá la forma:

$$y = f_2(\Omega_2 \cdot f_1(\Omega_1 \mathbf{x} + \Theta_1) + \Theta_2). \quad (3.38)$$

Al utilizar funciones de activación tipo *sigmoide*, cada una de las neuronas de la primera etapa divide el espacio de observación en dos regiones claramente diferenciadas: una, en la que la función de activación obtiene los valores más altos y otra, en la que la función de activación obtiene los valores más bajos. Dichas regiones están delimitadas por la ecuación del hiperplano definida por los pesos y los sesgos de la neurona en cuestión. Los valores de la región próximos al hiperplano dan lugar a las salidas de las neuronas intermedias entre el valor máximo de salida y el mínimo. El tamaño de la región de esta zona intermedia viene definido, entre otras cosas, por el módulo del vector de pesos. La segunda capa de la red combina los subespacios generados por cada una de las neuronas de la primera capa. De este modo, se puede ver que las fronteras de decisión finales estarán formadas por la combinación no lineal de hiperplanos, y esto justifica claramente que con un número suficiente de hiperplanos se puede construir cualquier tipo de frontera de decisión (Rosa y Gil , 2006). El número de capas necesarios en una red *MLP* depende de la forma de la frontera de decisión requerida para una adecuada clasificación.

Aspectos igualmente importantes de funcionamiento e implementación de una red neuronal *MLP* son: la *tasa de aprendizaje*, que controla la velocidad de búsqueda del error mínimo del algoritmo; el *criterio de parada*, que determina cuando se alcanzó el umbral mínimo en el error deseado y los *nodos de capa oculta*, que dependen de la naturaleza de los datos y determina la capacidad de la red. Estos son aspectos que permiten optimizar la configuración del clasificador en cuanto a su capacidad de generalización, eficiencia computacional y precisión en la clasificación. Mayor información sobre las redes ANN y MLP se encuentra en el Anexo C.

3.3.2 El Clasificador PNN

La red neuronal probabilística *PNN* fue introducida por Specht y en la actualidad es ampliamente utilizada en el área de reconocimiento de patrones. Pertenece a

la familia de las *redes de funciones de base radial (RBFN - Radial Basis Function Network)*, en las cuales la estimación de las funciones de densidad de probabilidad (funciones no paramétricas de Parzen) de cada clase, se realiza utilizando funciones de base *radial* centradas sobre cada patrón de entrenamiento (Georgiou V. et al. , 2004).

Mientras una red neuronal *MLP* usa funciones de base *lineal* (suma pesada de las entradas), una red neuronal tipo RBFN utiliza una función de base *radial* de la forma:

$$s(\mathbf{x}) = |\mathbf{x} - \mathbf{z}|_{\Sigma}^2 = (\mathbf{x} - \mathbf{z})^T \Sigma^{-1} (\mathbf{x} - \mathbf{z}), \quad (3.39)$$

conocida como el cuadrado de la distancia de “Mahalanobis”. Esto es, la distancia de un patrón de entrada \mathbf{x} a un vector de referencia \mathbf{z} , con respecto a una matriz de ponderación Σ . Así, en las redes neuronales RBFN, la activación de cada neurona depende de la distancia que separa el vector de entrada y un conjunto de vectores de referencia o prototipos, por lo que resulta conveniente utilizar funciones de activación con carácter localizado, como las funciones Gaussianas (Rosa y Gil , 2006).

Una red *PNN* está conformada por cuatro capas, como en la figura 3.13. La primera capa (o de entrada) contiene tantos nodos como características tenga el vector o patrón de entrada. La segunda capa (o de patrones) contiene tantas neuronas como patrones de entrenamiento, divididas en K grupos (uno por clase), de tal manera que para un vector de entrada $\mathbf{x} = (x_1, x_2, \dots, x_N)^T \in R^N$, cada una de sus N componentes son sometidas a cada neurona con función de base *radial*. De esta manera, la i -ésima neurona del k -ésimo grupo evalúa su salida mediante un *kernel*⁶ Gaussiano de la forma:

$$F_{k,i}(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_{k,i}\|^2}{2\sigma^2}\right), \quad (3.40)$$

donde $\mathbf{x}_{k,i}$ es el centro del *kernel* y σ un parámetro de dispersión “spread” que determina el tamaño del campo receptivo o esfera de influencia del respectivo *kernel*.

La tercera capa (o de sumarización), con tantas neuronas como el número de

⁶Funciones kernel. modelado de funciones de densidad de probabilidad como combinación de núcleos Gaussianos.

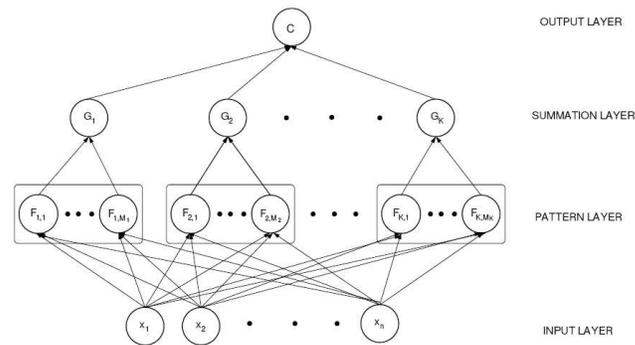


Figura 3.13. Arquitectura de una red PNN, (Georgiou V. et al. , 2004)

clases, evalúa las funciones de probabilidad condicionales mediante una combinación lineal de las densidades evaluadas en la capa anterior, según la expresión:

$$G_k(\mathbf{x}) = \sum_{i=1}^{M_k} w_{ki} F_{k,i}(\mathbf{x}), \quad k \in \{1, 2, \dots, K\}, \quad (3.41)$$

donde M_k es el número de neuronas patrón de la clase k , y w_{ki} son coeficientes positivos que satisfacen la condición $\sum_{i=1}^{M_k} w_{ki} = 1$. Entonces, el vector de entrada \mathbf{x} se clasificará en la clase que corresponda a la unidad de salida con la mayor suma, cuya etiqueta la entrega el nodo de la cuarta capa (o de salida).

En el caso de la red *PNN*, su entrenamiento requiere de una sola pasada sobre los patrones de entrenamiento y produce una configuración única. Esto la hace más rápida en cuanto al tiempo de entrenamiento que las redes neuronales con retroalimentación.

3.3.3 El Clasificador SVM

Las máquinas de soporte vectorial SVM son un método de clasificación basado en el conocimiento, desarrollado por Vladimir Vapnik, que usan: las herramientas estadísticas de optimización y las máquinas de aprendizaje. Tienen características importantes como: la ausencia de mínimos locales, la alta capacidad de generalización y de control, y no requieren hipótesis alguna sobre la densidad de probabilidad de los datos (Alfred M. , 1999).

La forma de clasificación de una red *SVM* es como sigue (Alfred M. , 1999): la red primero mapea los puntos de entrada a un espacio de características de di-

menor con el fin de separar las muestras de entrenamiento en dos regiones separables linealmente mediante un hiperplano como se muestra en la figura 3.14.a.

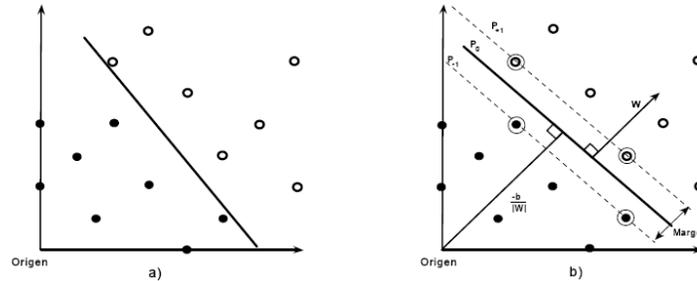


Figura 3.14. a) Separación lineal de clases por una red SVM, b) Hiperplano con máximo margen de separación de clases (Alfred M. , 1999)

Así, partiendo de un conjunto de L vectores de entrenamiento \mathbf{x}_l , cada uno de dimensión N , con su respectiva etiqueta y_l :

$$\mathbf{x}_l \in R^N \quad y \quad \mathbf{y}_l \in \{+1, -1\} \quad l = 1, \dots, L, \quad (3.42)$$

se busca estimar una función f tal que para una entrada en R^N produzca una salida en $\{\pm 1\}$

$$f : R^N \rightarrow \{+1, -1\}. \quad (3.43)$$

Así se puede clasificar correctamente un nuevo dato considerando su salida $y = f(\mathbf{x})$ con la misma distribución de probabilidad $P(\mathbf{x}_l, y_l)$, de los datos de entrenamiento.

Los clasificadores SVM se fundamentan en la obtención de hiperplanos que separen los datos de entrenamiento en dos grupos. Entre cada una de las clases etiquetadas como $\{-1, +1\}$, existe un único hiperplano con máximo margen de separación, figura 3.14.b. Con la intención de forzar la generalización de la máquina de aprendizaje, se busca que la distancia entre el hiperplano óptimo y el patrón de entrenamiento más cercano sea máxima (Alfred M. , 1999). El hiperplano óptimo de separación tiene la forma:

$$g(\mathbf{x}) = (\mathbf{w} \cdot \mathbf{x}) + b = 0. \quad (3.44)$$

Para maximizar el margen de separación, se proponen dos planos paralelos que contienen los puntos más cercanos al hiperplano óptimo, llamados *vectores soporte*.

Asumiendo como $1/\|\mathbf{w}\|^7$, la distancia entre el punto más cercano y el hiperplano óptimo, las ecuaciones de los dos planos paralelos estarán dadas por:

$$p_{+1} = (\mathbf{w} \cdot \mathbf{x}) + b = +1 \quad (3.45)$$

$$p_{-1} = (\mathbf{w} \cdot \mathbf{x}) + b = -1. \quad (3.46)$$

El margen entre estos dos planos es entonces $2/\|\mathbf{w}\|$, y la distancia del origen al hiperplano óptimo es $b/\|\mathbf{w}\|$.

Como por definición, entre los dos hiperplanos no deben existir datos de entrenamiento, los datos deben cumplir con las siguientes condiciones:

$$(\mathbf{w} \cdot \mathbf{x}) + b \geq +1 \quad \text{para } y_i = +1 \quad y \quad (3.47)$$

$$(\mathbf{w} \cdot \mathbf{x}) + b \leq -1 \quad \text{para } y_i = -1, \quad (3.48)$$

por tanto puede observarse que la función de decisión $f_{w,b}(\mathbf{x}) = y_i$, corresponde al signo que resulta de evaluar un dato en la ecuación (3.44) del hiperplano óptimo, por lo que puede escribirse como

$$f_{w,b}(\mathbf{x}) = \text{sign}[g(\mathbf{x}_i)]. \quad (3.49)$$

Es así como el algoritmo de aprendizaje de la red SVM tiene que resolver un problema de optimización cuadrático para encontrar la solución óptima que separa con un margen máximo el conjunto de muestras.

Multclasificación con SVM (Alfred M. , 1999). El planteamiento anterior solo aborda un problema de biclasificación, (clases con valores de ± 1), pero muchos problemas involucran más de dos clases. Para resolver este problema se admiten dos tipos de arquitectura: las *multclasificadoras* y las *biclasificadoras generalizadas*. La máquina *multclasificadora* construye una función clasificadora global, considera todas las posibles clases modificando la función objetivo de la SVM. Su principal desventaja es la complejidad algorítmica. Por su parte, la máquina *biclasificadora generalizada* construye una función clasificadora global a partir de un conjunto de funciones biclasificadoras, con la ventaja de resolver el problema con mayor simplicidad y con menor tiempo de respuesta que la anterior, además tiene dos técnicas

⁷La distancia perpendicular de un punto cualquiera a un plano esta dada por: $\frac{(\mathbf{w} \cdot \mathbf{x}) + b}{\|\mathbf{w}\|}$

de descomposición: la técnica *uno contra uno*, que consiste en construir un clasificador binario por cada par de clases diferentes, lo cual resulta muy dispendioso y poco práctico, y la técnica *uno contra todos*, que busca construir un clasificador binarios para cada una de las clases, tratando de decidir entre una de las clases y el resto, optando después por la clase en la cual la *SVM* obtenga la mayor salida. Las propiedades más importantes de las redes *SVM* son (Rosa y Gil , 2006):

- i)* Inducen un hiperplano lineal en el espacio de entrada, por lo que pertenecen a la familia de los métodos denominados "*linear learning machines*".
- ii)* El aprendizaje no lineal con *SVM* se consigue mediante el uso de las denominadas funciones kernel, que permiten transformar el espacio de entrada en un espacio de trabajo de dimensión mayor.
- iii)* Entre todos los posibles hiperplanos separadores, las *SVM* eligen el de margen máximo. Este sesgo inductivo de aprendizaje, que complementa el habitual modo de minimizar el error en el conjunto de aprendizaje, contribuye a disminuir el riesgo de sobreaprendizaje. Esto permite trabajar a las redes *SVM* en espacios de gran dimensión con garantía de robustez.
- iv)* El aprendizaje de las *SVM* se presenta como un problema de optimización con restricciones, que se puede resolver utilizando técnicas de aproximación cuadrática. La convexidad garantiza una solución única (esto supone una gran ventaja con respecto al modelo general de redes neuronales).

Capítulo 4

Extracción de Características DT, DWT y WPT

Para la extracción de características se dispone de una base de datos consistente en los registros mioeléctricos EMGS de cuatro tipos de movimientos de la mano. Estos registros son capturados mediante electrodos de superficie, por lo que se conocen como señales *electromiográficas superficiales* EMGS. Los registros fueron tomados de seis sujetos saludables quienes ejecutaron cada movimiento de manera sostenida y en secuencia, durante un intervalo de tiempo suficiente para capturar el estado estacionario de las señales EMGS.

Independientemente de la técnica de análisis, temporal o espectral basado en wavelets, las características extraídas deben proveer una buena separabilidad de clases para lograr una buena precisión en la clasificación de patrones y requerir el mínimo costo computacional para reducir los retardos en la selección de estados (Hudgins, et al., 1993). Teniendo en cuenta esto, se analiza inicialmente la eficiencia del conjunto de características extraídas según los índices *CSI* respectivos, sin recurrir a la evaluación del error de clasificación para cada combinación de características posibles. De esta forma se busca establecer un procedimiento sistemático de representación de señales EMGS eficiente y de baja complejidad, dejando la medida del error de clasificación solamente para evaluar el desempeño de las combinaciones de características *DT-DWT* y *DT-WPT* en el reconocimiento de patrones EMGS.

En este capítulo se describe y se justifica el preprocesamiento de los datos, el cual involucra principalmente la amplificación y segmentación de los registros, antes de someterlos a la fase de extracción de características. En el proceso de

representación de las señales EMGS, la fase de extracción de características es acompañada por la técnica de reducción de dimensionalidad mediante la *selección de características*, esto para optimizar la dimensionalidad de los vectores descriptores de los patrones EMGS, utilizando como herramienta de evaluación de todo este proceso el *índice de separabilidad de clases CSI*.

4.1 Descripciones Preliminares

4.1.1 Señales EMGS

Las señales *Electromiográficas Superficiales* EMGS son registros unidimensionales, sobre los cuales es posible aplicar procesamiento digital de señales como la extracción de características y el reconocimiento de patrones para la clasificación de movimientos mioeléctricos. Todo esto es posible gracias a que las señales motoras EMGS son señales que proveen información sobre la actividad neuromuscular que los origina. Estas señales son capturadas típicamente mediante electrodos bipolares de superficie ubicados sobre la piel. Hargrove, et al. (2005), realizaron un trabajo de comparación entre clasificación de señales mioeléctricas superficiales e intramusculares, concluyendo que la información extraída de las dos maneras permiten una buena clasificación, no encontrándose diferencias significativas en su capacidad de discriminación. Este resultado es importante, ya que brinda la seguridad de continuar utilizando las señales EMGS en futuros trabajos de investigación y aplicaciones.

Para la adquisición de las señales EMGS, las cuales son generadas por contracción muscular, es necesaria la identificación de las regiones musculares más comprometidas en la ejecución de los movimientos a clasificar para determinar la mejor ubicación de los electrodos. Debido a la elevada resistencia eléctrica natural de la piel, se recomienda la aplicación de un gel para disminuir dicha resistencia además de buscar una buena superficie de contacto y adherencia. A pesar de estas disposiciones, las señales EMGS aún son muy débiles para ser procesadas directamente, por lo que se requiere de una etapa de preprocesamiento que involucra la amplificación y segmentación de los registros EMGS. Estos aspectos se detallan a continuación.

4.1.2 Preprocesamiento

La amplitud típica de las señales EMGS es de 0-6 mV por lo que se requiere pasarlas directamente por un amplificador diferencial de alta ganancia (unas 5000 veces), cuidadosamente diseñado para evitar que distorsione la información de la señal. Una vez amplificada la señal, se debe considerar la eliminación de componentes de ruido tanto de alta frecuencia como la componente de ruido de baja frecuencia introducida por la red de distribución eléctrica de 60 Hz. Para eliminar esta última componente de ruido, lo más común es utilizar un filtro tipo “notch” aunque no sea lo más recomendable, pues este podría eliminar o atenuar también componentes de frecuencia con información útil dada la mayor concentración de energía de las señales EMGS entre los 50 Hz y 150 Hz, aunque su canal de información vaya de los 20 Hz a los 500 Hz (Graupe, et al., 1982).

4.1.3 Posición y número de electrodos

La posición de los electrodos es muy importante para una adecuada captura de los registros EMGS. Para lograr esto se recomienda la asistencia de personal médico especializado, quienes pueden identificar con mayor precisión las regiones musculares con mayor compromiso en la ejecución de los movimientos de interés. Hargrove, et al. (2005), demuestran que unas regiones son mejores que otras para la ubicación de los electrodos. Disponiendo para ello de 16 electrodos alrededor del antebrazo y por comparación del error de clasificación identificaron las mejores posiciones. Además, la generalización de dichas posiciones no es recomendable para todos los pacientes, por lo que la identificación de las mejores regiones musculares debe aplicarse a cada paciente.

El número de electrodos corresponde directamente con el número de canales a procesar, este es un parámetro esencial que incidirá en la precisión de clasificación y en el número de movimientos diferentes a discriminar. Para desarrollar una prótesis mioeléctrica multifuncional, es necesario extraer la mayor cantidad de información de cada canal o asignar una función de control a una combinación específica de señales de un sistema multicanal (Hudgins, et al., 1993). De esta manera, se consigue aumentar el número de movimientos en el sistema de prótesis a partir de la disposición de un mayor número de canales. Englehart, et al. (2001), mostraron que la precisión en la clasificación de patrones EMGS para el control mioeléctrico

multifuncional es mayor con cuatro canales que con solo dos o uno. Sin embargo, con un mayor número de fuentes de señales se presentan factores degradantes como: el aumento en el tiempo de respuesta del sistema por la gran cantidad de información que hay que procesar y la interferencia entre canales por la presencia de “crosstalk” entre músculos. Este tipo de crosstalk se debe a factores como: el arreglo geométrico de los músculos y la propagación dispersiva de la señal en el volumen conductor (Lowery, et al. , 2005; Nishikawa, et al., 1999). Estos son fenómenos que en lugar de mejorar la clasificación podrían empeorarla, por lo que es necesario mediar entre precisión, multifuncionalidad y velocidad de respuesta.

4.1.4 El estado estacionario de las señales EMGS

En la generación de señales *mioeléctricas* por contracción muscular se identifican dos fases o estados: uno, el estado *transitorio* que corresponde a los instantes de la ráfaga de actividad mioeléctrica que acompaña al esfuerzo muscular repentino en la ejecución de un movimiento, y dos, el estado *estacionario* que corresponde al tiempo del esfuerzo muscular realizado durante un movimiento sostenido (Englehart, et al., 1999a). Algunos de los trabajos de investigación como los de: Graupe y Cline (1975); Hudgins, et al. (1993); Englehart, et al. (2001) y Liu, et al. (2005), están enfocados en la clasificación de acciones discretas como las generadas por la muñeca, los dedos o el codo, utilizando el estado transitorio de las señales EMGS. Comprobaron la existencia de una estructura determinística en las señales durante la fase inicial de una contracción muscular, la cual contiene información discriminante muy valiosa. De igual manera se han realizado trabajos en los cuales se utiliza el estado estacionario de las señales EMGS como los de: Englehart, et al. (1999a, 2001) y Crawford, et al. (2005) con el propósito de simplificar el control de los dispositivos de prótesis, aún sabiendo que cierta cantidad de información de la señal es ignorada y que su fase estacionaria tiene una estructura temporal muy corta debida a los cambios dinámicos en el reclutamiento de unidades motoras y el patrón de los impulsos eléctricos necesarios para una contracción sostenida. Particularmente Englehart, et al. (2001), obtienen resultados que son concluyentes al demostrar que se puede lograr un desempeño excepcionalmente bueno en la clasificación continua de patrones de movimientos usando el estado estacionario de las señales EMGS como lo muestra la figura 4.1.

4.1.5 Segmentación de los registros de las señales EMGS

La introducción del procesamiento digital de señales para el análisis de señales EMGS sugiere un proceso de muestreo para su digitalización. La tasa de muestreo práctica para este tipo de señales es de 1000 Hz, con el fin de registrar componentes de frecuencia en la señal hasta los 500 Hz según el teorema de Nyquist ¹, o muestrear a una frecuencia de 500 Hz para registrar componentes de frecuencia hasta los 250 Hz. No se requiere de mayores tasas de muestreo, puesto que la mayor concentración de energía y por tanto información útil en estas señales se encuentra entre los 50 y 150 Hz (Hudgins, et al., 1993). De cualquier manera se debe tener en cuenta que la tasa de muestreo incidirá directamente en la longitud de los registros de la señal a procesar, lo que implicará mayor o menor tiempo de cómputo sobre cada registro, comprometiendo con ello el retardo en la respuesta del sistema de clasificación.

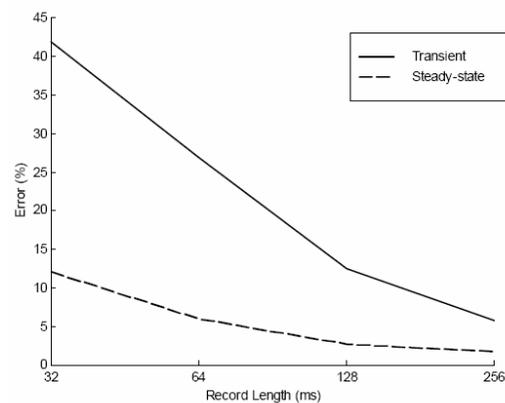


Figura 4.1. Error de clasificación según la longitud de los segmentos (Englehart, et al., 2001)

Una vez aclarado el proceso de detección de los registros EMGS se preparan los datos para su procesamiento. La preparación consiste básicamente en segmentar los registros de las señales en secciones de longitud 256 muestras (256 ms de grabación), para luego aplicar las técnicas de extracción de características a cada segmento como en: Hargrove, et al. (2005); Englehart, et al. (1999a, 2001) y Ferguson y Dunlop (2002). Por su parte Englehart, et al. (2001) compararon el desempeño en la clasificación de seis movimientos diferentes de la mano, variando la longitud de

¹La frecuencia Nyquist corresponde al doble de la frecuencia máxima registrada en la señal

los segmentos en 32, 64, 128 y 256 muestras en cada ensayo. Así, mostraron que el error de la clasificación disminuye cuando aumenta la longitud de cada segmento, figura 4.1. Otros trabajos como en: Rodriguez y Vuskovic (2005, 2006) y Zhang, et al. (2004), utilizan secuencias de 200 y 400 muestras (200 y 400 ms), obteniendo igualmente resultados satisfactorios. Sin embargo, cabe anotar que longitudes mayores de 256 muestras (o 256 ms) implicará mayor tiempo de procesamiento y por tanto mayor retardo en la respuesta de los sistemas de prótesis.

4.1.6 Selección de la base de datos de señales EMGS

Las señales EMGS utilizadas para el desarrollo de este trabajo son tomadas de la base de datos “*LeviData.zip*” de Levi Hargrov, suministrada y autorizada por Kevin Englehart Ph.D., P. Eng. del *Instituto de Ingeniería Biomédica de la Universidad de New Brunswick*. Para su construcción Hargrov dispuso alrededor del antebrazo de los pacientes un arreglo lineal de 16 electrodos con espaciamiento de 2 cm, figura 4.2. Este arreglo es conectado a un sistema de grabación con preamplificación y configurado para mediciones diferenciales entre electrodos adyacentes. Cada canal (proveniente de cada electrodo) es filtrado entre 10 y 500 Hz usando un filtro pasa-banda tipo *Besel* de 4^{to} orden con ganancia de 33 dBs (2000 veces). La salida analógica de la etapa de amplificación se pasa a la tarjeta de conversión A/D a 12 bits *DAQCard-6024E* de la *National Instrument*, con tasa de muestreo 1024 Hz (Hargrove, et al., 2005).

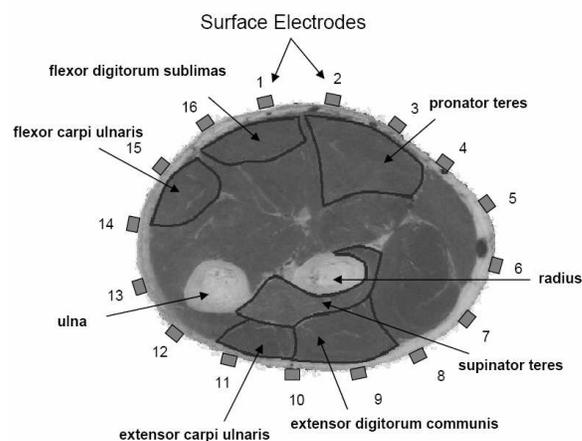


Figura 4.2. Disposición de electrodos para el registro de señales EMGS (Hargrove, et al., 2005)

La base de datos original “LeviData.zip” contiene los registros EMGS de diez (10) movimientos mioeléctricos que involucran los dedos de la mano y la muñeca. Los pacientes quienes ejecutaron los movimientos eran sujetos normales y saludables, con edades comprendidas entre 23 y 30 años. Para cada movimiento realizaron una contracción isométrica completa de mediana fuerza durante 5 segundos seguida de un breve periodo de descanso, repitiendo cada secuencia dos veces. En este orden fueron realizadas dos sesiones de grabación, una para obtener la base de datos de entrenamiento y la otra para obtener la base de datos de prueba (Hargrove, et al., 2005).

Para este trabajo se seleccionaron cuatro (4) movimientos básicos de la mano: el reposo, la mano abierta, el agarre fino y el agarre grueso, por considerarse los más útiles y prácticos para ser implementados en una prótesis de mano, figura 4.3.



Figura 4.3. Movimientos seleccionados para su clasificación (Hargrove, et al., 2005)

Así mismo, se propone la selección de los dos (2) mejores canales, considerados como suficientes para la discriminación y clasificación de cuatro movimientos de la mano, para lo cual se tiene en cuenta los efectos de la interferencia intramuscular y efecto cocanal presente entre canales adyacentes, además de la redundancia en información que pueda presentarse en dos canales consecutivos.

En cuanto a la extracción y selección de las características *temporales* y las características *espectrales* basadas en wavelets para la conformación de los vectores descriptores de las señales EMGS, es necesario tener en cuenta los efectos de la dimensión de los vectores sobre el error de clasificación. Englehart (1998) encontró que unas cuantas características no son suficientes para una representación con buena capacidad discriminante, y que demasiadas características degradan la capacidad de generalización. En la figura 4.4, se puede observar que el error de clasificación se minimiza para una dimensión de alrededor de 20 características.

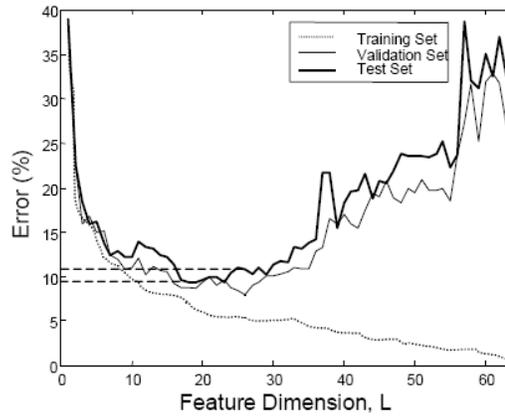


Figura 4.4. Relación del número de características y el error de clasificación (Englehart, 1998)

4.1.7 Selección de canales y características mediante el índice CSI

Teniendo en cuenta que las propiedades más deseables del conjunto de características para una buena clasificación son: que la distancia estadística entre clases debe ser maximizada, y que el conjunto de características debe ser optimizado eliminando aquellas que sean redundantes (Englehart, 1998), en este trabajo se utiliza el índice de separabilidad de clases *CSI* como herramienta de reducción de dimensionalidad por selección de características y en la selección del mejor par de canales para cada uno de los pacientes. El índice *CSI*² consiste en una medida relativa del grado de traslape entre clases con respecto a una característica seleccionada, aspecto muy útil en aplicaciones de clasificación (Zhang, et al. , 2004).

En la figura 4.5 se puede apreciar una condición de traslape entre clases (a la izquierda) y una condición de buena separabilidad de clases (a la derecha), esta última condición puede lograrse seleccionando: el mejor par de canales y las características más discriminantes. El índice de separabilidad de clases *CSI* se obtiene mediante la siguiente expresión (Zhang, et al. , 2004):

$$CSI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{ij}) = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \left(\frac{S_i + S_j}{D_{ij}} \right), \quad (4.1)$$

donde K es el número de clases, R_{ij} es el índice de similitud entre las clases i y j , S_i y S_j describen el grado de dispersión entre las clases i y j y D_{ij} es la

²El índice *CSI* se conoce también como índice de Davies-Bouldin

distancia media entre las clases i y j . Estas medidas de dispersión y distancia pueden evaluarse de la siguiente manera:

$$S_i = \left\{ \frac{1}{N_i} \sum_{j=1}^{N_i} (x_j - m_i)^T (x_j - m_i) \right\}^{1/2} \quad (4.2)$$

y

$$D_{ij} = \left\{ (m_i - m_j)^T (m_i - m_j) \right\}^{1/2}, \quad (4.3)$$

donde N_i representa el número de muestras de la clase i , x_j la muestra de entrada y m_i la media de las muestras de entrada de la clase i .

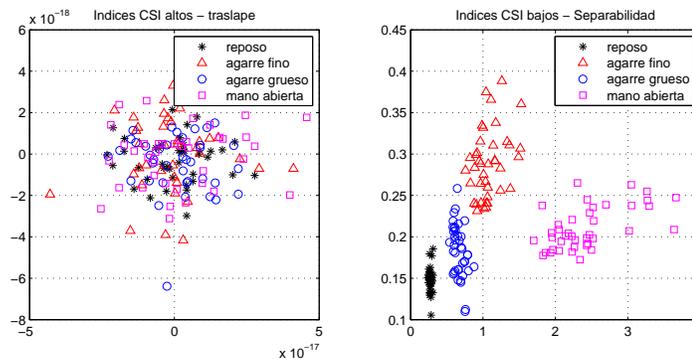


Figura 4.5. Traslape o separabilidad de clases según el índice CSI

Partiendo de la hipótesis de que los 16 canales (o electrodos dispuestos al rededor del antebrazo) son fuente de información discriminante de los movimientos mioeléctricos, la extracción de características temporales (DT) y espectrales (DWT y WPT), mediante los diversos métodos descritos, se realiza sobre todas las posibles combinaciones de pares de canales, ejecutando en cada caso los procesos:

- a) **Extracción y selección de características.** Según el método implementado se obtiene un conjunto original de características y luego se hace una selección de aquellas con mayor capacidad discriminante, para esto, se promedia para cada una los índices CSI sobre todas las combinaciones de pares de canales descartando aquellas que resulten con los índices más altos en cada caso.
- b) **Selección de canales.** Una vez seleccionadas las mejores características de cada método, se determinan los dos (2) mejores pares de canales por

paciente, seleccionando los dos mejores índices *CSI* promediados sobre el conjunto de características seleccionadas.

Se realizó un análisis previo para identificar los métodos tempo-espectrales a implementar, y de cada uno seleccionar los mejores subconjuntos de características a partir de los conjuntos originales, con la colaboración de Gasca y Rojas (2008), en el desarrollo de su trabajo de grado “*Extracción de características descriptoras de movimientos de la mano a partir de señales electromiográficas (EMG) aplicando técnicas wavelets*”. Los resultados se publicaron en (Romo, et al., 2007).

4.2 Características Temporales DT

La variabilidad intra-clases de las señales EMGS tiene un comportamiento aleatorio, por lo que intentar clasificar sus patrones directamente de las formas de onda muestreadas resultaría en un pobre desempeño en la clasificación. Sin embargo, a pesar de esta variabilidad el comportamiento estadístico de la forma de onda podría ser suficientemente estable para una adecuada clasificación de patrones (Hudgins, et al., 1993). Para la extracción de características temporales se implementaron dos (2) métodos: ***evaluación de parámetros*** y ***evaluación de momentos***. Los resultados se presentan a continuación.

4.2.1 Método 1. Evaluación de parámetros temporales

De los parámetros temporales descritos en el capítulo 2, se seleccionaron: el *valor medio absoluto* (VMA), la *longitud de onda* (LO), los *cruces por cero* (CC) y los *cambios de signo de pendiente* (CSP) por su mayor capacidad discriminante. A continuación se ejecutan los procesos de: extracción y selección de características, y selección de canales.

a) Extracción y selección de características

Los parámetros *VMA*, *LO*, *CC* y *CSP* se evalúan mediante las siguientes expresiones:

– ***Valor medio absoluto VMA:***

$$\bar{X}_i = \frac{1}{N} \sum_{k=1}^N |x_k|, \quad i = 1, 2, \dots, I, \quad (4.4)$$

donde x_k es la $k_{ésima}$ muestra en el segmento i , el número total de segmentos es I .

– **Longitud de la forma de onda LO:**

$$l_o = \sum_{k=1}^N |\Delta x_k|, \quad (4.5)$$

donde, $\Delta x_k = x_k - x_{k-1}$ (diferencia entre valores de muestras consecutivas).

– **Cruces por cero CC.** El contador de cruces por cero se incrementa según la comparación entre dos muestras consecutivas x_k y x_{k+1} así:

$$\begin{aligned} x_k > 0 \quad y \quad x_{k+1} < 0 \quad o \quad x_k < 0 \quad y \quad x_{k+1} > 0 \\ y \quad |x_k - x_{k+1}| \geq 0.01V. \end{aligned} \quad (4.6)$$

– **Cambios de signo de pendiente CSP.** El contador de cambios de signo en la pendiente se incrementa según las relaciones:

$$\begin{aligned} x_k > x_{k-1} \quad y \quad x_k > x_{k+1} \quad o \quad x_k < x_{k-1} \quad y \quad x_k < x_{k+1} \\ y \quad |x_k - x_{k+1}| \geq 0.01V \quad o \quad |x_k - x_{k-1}| \geq 0.01V. \end{aligned} \quad (4.7)$$

De la evaluación de estos *parámetros* resulta un conjunto de 8 características temporales DT(4 parámetros x 2 canales), y para cada una se promedia el índice *CSI* sobre todas las combinaciones de pares de canales. Estos promedios se presentan en la figura 4.6, donde se observa que las características 7 y 8 correspondientes al parámetro *CSP* en cada canal presentan índices *CSI* promedio relativamente altos, por lo que se procede a eliminarlas.

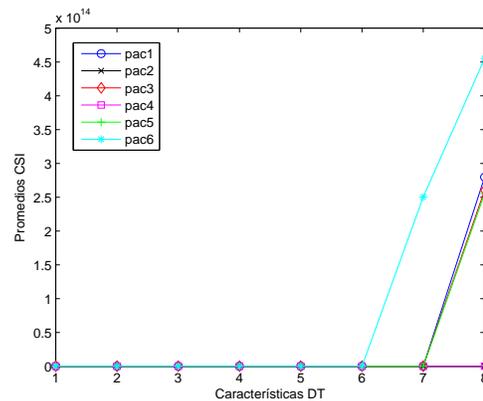


Figura 4.6. Promedios CSI para los parámetros *VMA* (*caract. 1 y 2*), *LO* (*caract. 3 y 4*), *CC* (*caract. 5 y 6*), *CSP* (*caract. 7 y 8*).

En la figura 4.7 se presentan los índices *CSI* promedio para las características correspondientes a los *parámetros VMA, LO y CC*. Ahora los índices *CSI* promedio en general son relativamente bajos, excepto en las características 1 y 2 del paciente 2 y la característica 1 del paciente 5. Sin embargo, eliminando el parámetro *CC* se mejoran los índices *CSI* promedio para los seis pacientes.

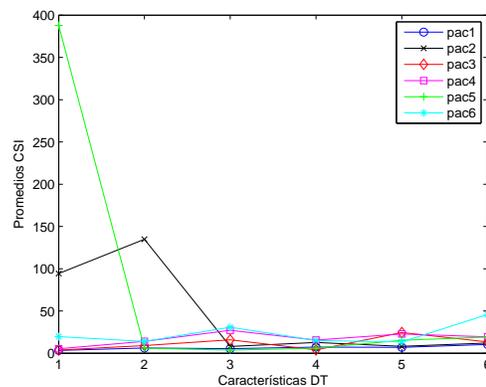


Figura 4.7. Promedios CSI para los parámetros *VMA* (*caract. 1 y 2*), *LO* (*caract. 3 y 4*), *CC* (*caract. 5 y 6*)

En la figura 4.8 se comparan los índices *CSI* promedio por paciente para cuatro (4) características DT (*VMA y LO*) y seis (6) características DT (*VMA, LO y CC*), observándose mínimas diferencias por lo que el considerar 4 o 6 carac-

terísticas dependerá de la dimensión de los vectores descriptores y del error de clasificación que se logre en cada caso.

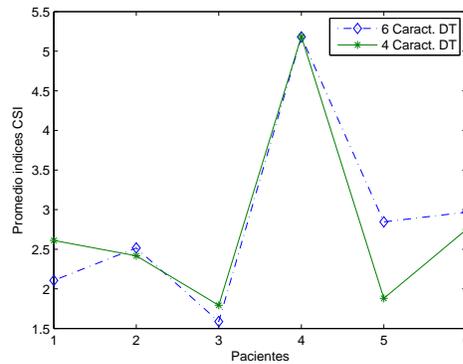


Figura 4.8. Promedios CSI por paciente para 4 y 6 características DT

b) Selección de canales

Una vez seleccionadas como las mejores características las correspondientes a los parámetros VMA y LO , se determinan los dos (2) mejores pares de canales por paciente seleccionando los dos mejores índices CSI promedios sobre las cuatro (4) características seleccionadas. Los resultados se resumen en la tabla 4.1.

4.2.2 Método 2. Evaluación de momentos temporales

La evaluación de momentos temporales M_m sobre las muestras de los registros EMGS se introduce bajo la suposición de la existencia de una estructura estadística en la señal. Para este trabajo se consideran seis (6) momentos (M_1, \dots, M_6) sobre cada sección de 256 muestras. A continuación se ejecutan los procesos de: extracción y selección de características, y selección de canales.

a) Extracción y selección de características

Los momentos temporales se evalúan mediante la expresión:

$$M_m = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^m \quad m = 1, 2, 3, 4, 5, 6, \quad (4.8)$$

Tabla 4.1. Mejores pares de canales según el índice CSI para 4 características DT extraídas a partir de los parámetros temporales *VMA* y *LO*.

Pacientes	Mejores canales	Promedios CSI
Paciente 1	6 - 15	0.6788
	3 - 15	0.8185
Paciente 2	1 - 4	0.8816
	4 - 15	0.9782
Paciente 3	4 - 6	0.7494
	1 - 4	0.8200
Paciente 4	11 - 13	2.3448
	4 - 13	2.4501
Paciente 5	6 - 8	1.5298
	8 - 16	1.6998
Paciente 6	10 - 12	1.7565
	9 - 12	1.9670

De la evaluación de estos *momentos* resulta un conjunto de 12 características temporales DT (6 momentos x 2 canales), y para cada una se promedia el índice *CSI* sobre todas las combinaciones de pares de canales. Estos promedios se presentan en la figura 4.9, donde se observa que las características 1, 2, 5, 6, 9, 10, 11 y 12 correspondientes a los momentos *M1*, *M3*, *M5* y *M6* presentan índices *CSI* promedios relativamente altos, por lo que se procede a eliminarlos.

En la figura 4.10 se presentan los índices *CSI* promedios para las características correspondientes a los dos mejores momentos (*M2* y *M4*), donde se observa que los índices *CSI* promedios son en general bajos, excepto por la característica 3 del paciente 6. A pesar de esto, serán conservados estos dos momentos (*M2* y *M4*) para el conjunto de características.

b) Selección de canales

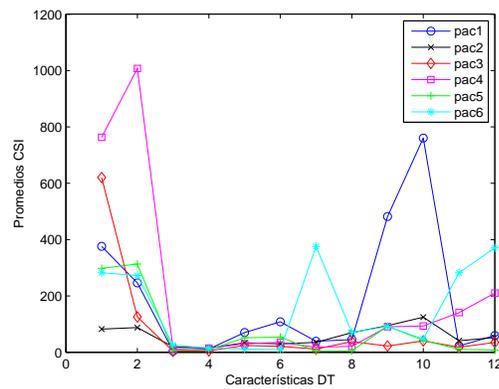


Figura 4.9. Promedios CSI para los momentos M_1 (caract. 1 y 2), ... M_6 (caract. 11 y 12)

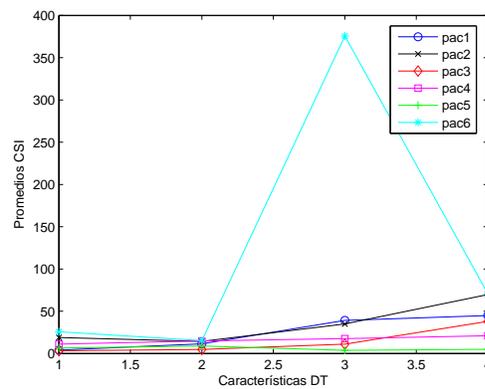


Figura 4.10. Promedios CSI para Momentos M_2 (caract.1 y 2) y M_4 (caract.3 y 4)

Una vez seleccionadas como las mejores características las correspondientes a los a los momentos M_2 y M_4 , se determinan los dos (2) mejores pares de canales por paciente seleccionando los dos mejores índices CSI promedios sobre las cuatro (4) características seleccionadas. Los resultados se resumen en la tabla 4.2.

Para finalizar el análisis temporal, en la figura 4.11, se comparan los índices CSI promedio sobre el conjunto de características seleccionadas en cada método. Se puede observar en general (excepto para el paciente 4), que el método de *parámetros temporales* es mejor que el método de *momentos temporales*.

Tabla 4.2. Mejores pares de canales según el índice CSI para 4 características DT extraídas a partir de los momentos temporales M_2 y M_4

Pacientes	Mejores canales	Promedios CSI
Paciente 1	4 - 15	0.8465
	1 - 15	0.8520
Paciente 2	4 - 15	1.1202
	12 - 15	1.1899
Paciente 3	4 - 6	1.0173
	6 - 12	1.1126
Paciente 4	3 - 13	1.9524
	7 - 13	2.1641
Paciente 5	6 - 16	1.7144
	3 - 16	1.7446
Paciente 6	9 - 12	1.9985
	6 - 12	2.2149

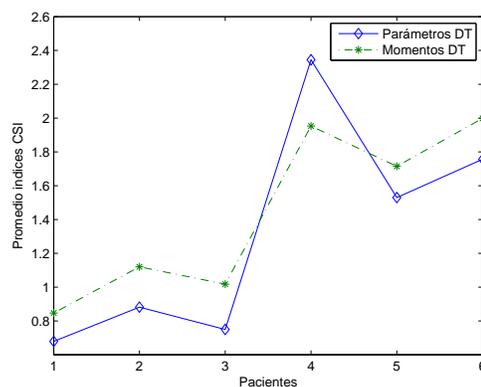


Figura 4.11. Promedios CSI por paciente según los métodos de: Parámetros y Momentos DT

4.3 Características Espectrales DWT

La extracción de características espectrales DWT parte de la descomposición wavelet sobre registros EMGS de 256 muestras, para lo cual es necesario identificar el nivel óptimo de descomposición y la función wavelet más adecuada. Adicionalmente

se puede seleccionar solamente las subbandas con mayor concentración de energía o información discriminante para los registros EMGS. Esto hace más eficiente el proceso de extracción de las características DWT como se describe a continuación.

4.3.1 Nivel de descomposición DWT

El nivel máximo J , de descomposición DWT sobre una secuencia temporal de longitud N está dado por $J = \log_2 N$, lo que significa que para los registros EMGS de 256 muestras su descomposición se puede realizar hasta el nivel $J = 8$. Sin embargo, la descomposición a este nivel como en (Rodríguez y Vuskovic, 2006), no es lo más común, pues la mayoría de autores realizan la descomposición a nivel $j = 4$ como en (Zhang, et al., 2002) y a nivel $j = 3$ como en (Kilby y Hosseini, 2004). En general cada trabajo de investigación selecciona el nivel de descomposición DWT según el procesamiento a realizar sobre los coeficientes wavelets.

En Romo, et al. (2007), se comprobó que entre los niveles de descomposición 3 y 4, el nivel óptimo de descomposición es el nivel 3, figura 4.12, pues permite obtener mejores índices de separabilidad de clases CSI a la vez que se gana tiempo de procesamiento.

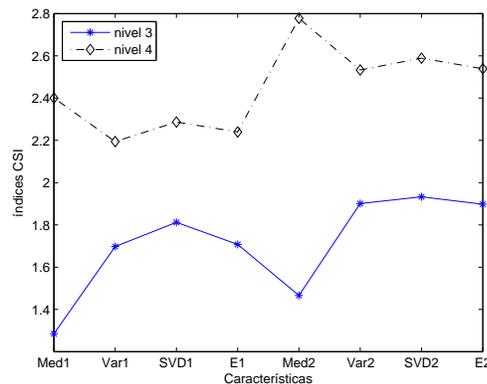


Figura 4.12. Índices CSI para 3 y 4 niveles de descomposición DWT

Estos resultados se justifican en que el ancho de banda de las señales EMGS va de 0 a 500 Hz, por lo que las secuencias de coeficientes de *aproximación* y de *detalles* de una descomposición wavelet DWT de nivel 4 quedan distribuidas aproximadamente en las siguientes subbandas de frecuencias:

D1: 250 - 500 Hz

D2: 125 - 250 Hz

D3: 62.5 - 125 Hz

D4: 32.25 - 62.5 Hz

A4: 0 - 32.25 Hz

Sin embargo, la mayor concentración de energía (o información útil) de estas señales se encuentra en el rango de frecuencias entre 50 y 150Hz, encontrándose además que son fuertemente afectadas por el ruido de frecuencias inferiores a los 60 Hz (De Luca, 1978). Por lo anterior no se justifica la descomposición de las subbandas inferiores A4 y D4, siendo suficiente una descomposición de nivel 3.

4.3.2 Selección de la función wavelet

La selección de la función *wavelet* para la descomposición DWT, consiste en determinar la familia de funciones wavelets que mejor se acople a la forma de onda de las señales EMGS. Para ello, en Romo, et al. (2007), se compara el desempeño de las familias wavelets más utilizadas en los trabajos relacionados con señales mioeléctricas, entre las cuales se destacan familias como: la *Daubechies* (*db*), la *Coiflet* (*Coif*) y la *Symlet* (*Sym*). La comparación de las mejores *wavelet* en cada familia muestra que con la *wavelet* “Daubechies” de orden 3 (*db3*) se obtiene los índices *CSI* más bajos, figura 4.13.

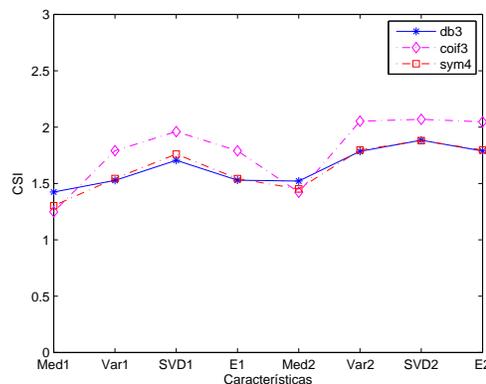


Figura 4.13. Índices CSI según las tres mejores familias wavelet madre para DWT

4.3.3 Selección de subbandas DWT

De la descomposición wavelet DWT de nivel 3 resultan las secuencias de coeficientes $\{D_1, D_2, D_3, A_3\}$, con la distribución de coeficientes en las siguientes subbandas de frecuencias:

D1: 250 - 500 Hz

D2: 125 - 250 Hz

D3: 62.5 - 125 Hz

A3: 0 - 62.5 Hz

La selección de subbandas consiste en eliminar la secuencia de coeficientes de aquellas subbandas donde no hay concentración de energía o información discriminante para las señales EMGS. La comparación de los índices *CSI* por característica, cuando se consideran las cuatro subbandas $\{D_1, D_2, D_3, A_3\}$ y cuando se consideran las subbandas $\{D_2, D_3\}$, excluyendo A_3 y D_1 por su bajo contenido de información discriminante, muestra menores índices *CSI* para el caso de las dos subbandas (D_2 y D_3), figura 4.14.

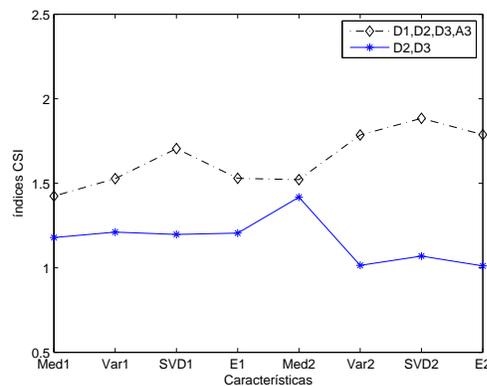


Figura 4.14. Índices CSI según la selección de subbandas DWT

Así, esta selección de subbandas se puede ver como una forma eficiente de reducción de dimensionalidad. Pues la comparación de los índices *CSI* promedio por paciente, sobre un conjunto de características (*parámetros espectrales*) evaluadas: en primer lugar sobre la matriz que resulta de someter los coeficientes wavelets (con descomposición DWT a nivel 3) al análisis PCA, y en segundo lugar, directamente sobre las subbandas D_2 y D_3 , muestra menores índices *CSI* para el segundo caso,

es decir, sin análisis PCA, figura 4.15. Además de que resulta de menor complejidad computacional la reducción de dimensionalidad por selección de subbandas que por proyección de características mediante el análisis PCA.

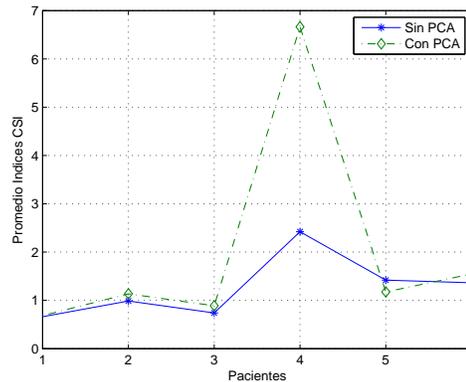


Figura 4.15. Comparación de índices CSI bajo el análisis PCA

Una vez determinado el nivel óptimo de descomposición wavelet DWT, la función wavelet más adecuada y las subbandas con mayor información discriminante se procede a extraer las características espectrales DWT, lo cual consiste en manipular de alguna manera los coeficientes que resultan de la descomposición DWT de la señal EMGS. En el capítulo 2 se describieron cuatro (4) métodos de extracción de características a partir de los coeficientes DWT, de los cuales se excluyó el método de (*cuadrados de los coeficientes wavelets*), por implicar una gran dimensionalidad el trabajar con la totalidad de los coeficientes DWT. Así, para la extracción de características espectrales DWT se implementaron los métodos: *evaluación de valores singulares*, *evaluación de parámetros* y *evaluación de momentos*. Los resultados se presentan a continuación.

4.3.4 Método 1. Descomposición de valores singulares SVD

En este trabajo la descomposición SVD se realiza sobre la matriz A de coeficientes de la descomposición wavelet DWT de nivel 3, conformada por las subbandas de descomposición $\{D1, D2, D3, A3\}$, asumiendo que:

$$A = \sum_{i=1}^k \lambda_i \mu_i v_i^T, \quad (4.9)$$

donde λ_i son los valores singulares, los cuales se pueden encontrar como la raíz cuadrada de los valores propios de $A^T A$ y que son utilizados como características de descripción. A continuación se ejecutan los procesos de: extracción y selección de características, y selección de canales.

a) Extracción y selección de características

De la evaluación los *valores singulares* como la raíz cuadrada de los valores propios de $A^T A$ resultan 8 características espectrales DWT (4 valores singulares x 2 canales), y para cada una se promedia el índice *CSI* sobre todas las combinaciones de pares de canales. Estos promedios se presentan en la figura 4.16, donde se puede observar que en general para todos los pacientes las características presentan índices *CSI* promedio relativamente bajos, excepto por las características 3 y 4 para el paciente 6 y la característica 8 para el paciente 5, sin embargo son casos aislados que no justifican la eliminación de ninguna característica. Por tanto, se decide conservar todas las características que resultan de la descomposición SVD.

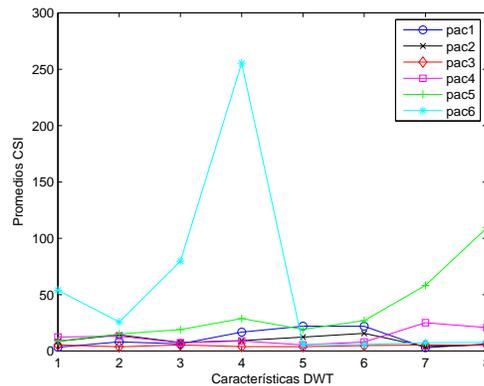


Figura 4.16. Promedios CSI para los valores singulares sobre la matriz de coeficientes DWT

b) Selección de canales

Después de comprobar que todos los valores singulares extraídos de la matriz de coeficientes DWT son útiles como características de descripción, se determinan los dos (2) mejores pares de canales por paciente seleccionando los

dos mejores índices CSI promedios sobre las cuatro (4) características seleccionadas. Los resultados se resumen en la tabla 4.3.

Tabla 4.3. Mejores pares de canales según el índice CSI para 8 características DWT extraídas a partir de los valores singulares SVD sobre la matriz de coeficientes DWT

Pacientes	Mejores canales	Promedios CSI
Paciente 1	6 - 15	0.7765
	9 - 15	0.8898
Paciente 2	4 - 15	0.9628
	1 - 4	1.0748
Paciente 3	4 - 13	0.8146
	13 - 15	0.9076
Paciente 4	3 - 13	1.9402
	4 - 13	2.3987
Paciente 5	8 - 12	1.9726
	1 - 8	2.3008
Paciente 6	10 - 12	1.7565
	9 - 12	1.8390

4.3.5 Método 2. Evaluación de parámetros espectrales

Para este trabajo se seleccionaron los parámetros: el *promedio absoluto de coeficientes* (*AVG*), la *energía de escala* (*ENER*), los *valores singulares* (*SVD*) y la *varianza* (*VAR*). A continuación se ejecutan los procesos de: extracción y selección de características, y selección de canales.

a) Extracción y selección de características

Los parámetros *AVG*, *VAR*, *ENER* y *SVD* se evalúan mediante las siguientes expresiones:

– **El promedio absoluto de coeficientes AVG:**

$$AVG_j = \frac{1}{K} \sum_{k=1}^K |c_{jk}|. \quad (4.10)$$

– **La varianza VAR:**

$$VAR_j = \frac{1}{K} \sum_{k=1}^K (c_{jk} - \bar{c}_{jk})^2. \quad (4.11)$$

– **La energía de escala ENER:**

$$ENER_j = \frac{1}{2} \sum_k c_{jk}^2. \quad (4.12)$$

– **Valores singulares SVD:**

$$SVD = \sqrt{\lambda(C^T C)}. \quad (4.13)$$

donde c_{jk} es el k^{esimo} coeficiente wavelet de la escala j , $\lambda(C^T C)$ significa las raíces características de la matriz $(C^T C)$, donde C es una matriz cuyas filas se forman con las respectivas subbandas de coeficientes wavelets.

La evaluación de estos cuatro (4) parámetros puede realizarse de dos maneras: sobre la matriz de coeficientes DWT compuesta con las subbandas seleccionadas o sobre cada una de estas subbandas. Luego para cada una de las características obtenidas en cada caso, se promedia el índice CSI sobre todas las combinaciones de pares de canales. A continuación se presentan los resultados obtenidos.

Opción 1. Evaluación de parámetros sobre la matriz de coeficientes DWT.

En este caso se evalúan los parámetros AVG , VAR , $ENER$ y SVD sobre la matriz de *coeficientes wavelets* conformada con las subbandas de mayor concentración de energía D_2 y D_3 . De la evaluación de estos *parámetros* resultan 8 características espectrales DWT (4 parámetros x 2 canales), y para cada una se promedia el índice CSI sobre todas las combinaciones de pares de canales. Estos promedios se presentan en la figura 4.17, donde se observa en general que los índices CSI promedios son relativamente bajos para todos los pacientes, excepto por las características 3 y 4 para el paciente 4 y las características 7 y 8 para los pacientes 1 y 6. Sin embargo, son casos aislados y muy particulares que no justifican la eliminación de ninguna característica.

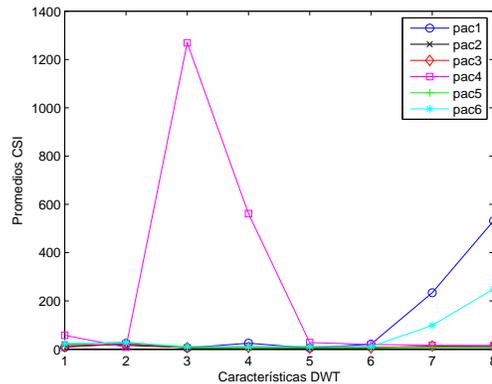


Figura 4.17. Promedios CSI para los parámetros AVG (caract. 1 y 2), VAR (caract. 3 y 4), ENER (caract. 5 y 6) y SVD (caract. 7 y 8), sobre la matriz de coeficientes DWT

Opción 2. Evaluación de parámetros sobre cada sub-banda de coeficientes DWT.

De la evaluación de estos *parámetros* sobre cada una de las subbandas D_2 y D_3 , resultan 16 características DWT (4 parámetros x 2 subbandas x 2 canales), y para cada una se promedia el índice *CSI* sobre todas las combinaciones de pares de canales. Estos promedios se presentan en la figura 4.18, donde se puede observar en general que los índices *CSI* promedios son relativamente bajos para todos los pacientes, excepto por las características 9 y 10 para el paciente 3, y las características 8 y 9 para los pacientes 5 y 6. Sin embargo justifican su eliminación.

La comparación de los índices *CSI* promedio por paciente para las dos formas de evaluar los *parámetros espectrales*, figura 4.19, muestra que la evaluación de los parámetros *AVG*, *VAR*, *ENER* y *SVD* sobre la matriz de coeficientes DWT, es mejor opción que la evaluación sobre cada una de las subbandas D_2 y D_3 .

b) Selección de canales

Después de comprobar que todos los parámetros *AVG*, *VAR*, *ENER* y *SVD* se pueden utilizar como características de descripción y que los índices mas bajos se logran cuando se evalúan sobre la matriz (opción 1), se determi-

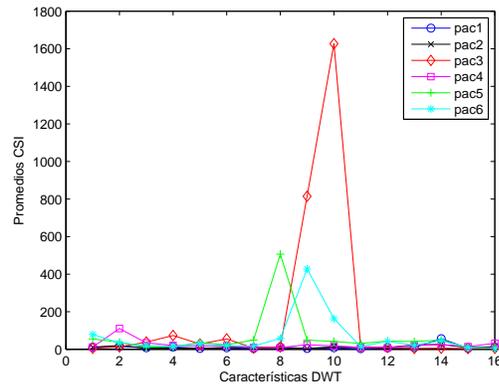


Figura 4.18. Promedios CSI para los parámetros *AVG* (caract. 1 a 4), *VAR* (caract. 5 a 8), *ENER* (caract. 9 a 12) y *SVD* (caract. 13 a 16) sobre cada subbanda de coeficientes DWT

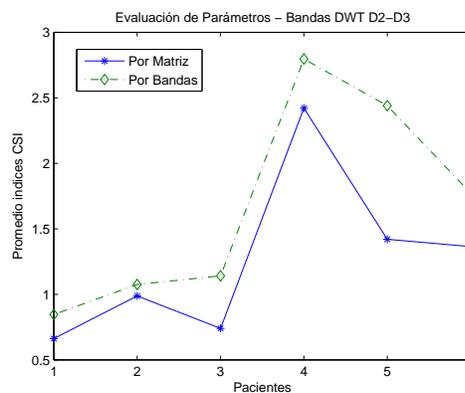


Figura 4.19. Promedios CSI por paciente según las opciones 1 y 2 del método de parámetros espectrales

nan los dos (2) mejores pares de canales por paciente seleccionando los dos mejores índices *CSI* promedios sobre las ocho (8) características seleccionadas. Los resultados se resumen en la tabla 4.4.

4.3.6 Método 3. Evaluación de momentos espectrales

Para este trabajo se considera la evaluación de seis (6) momentos M_i sobre los coeficientes wavelets DWT mediante la siguiente expresión:

Tabla 4.4. Mejores pares de canales según el índice CSI para 8 características DWT extraídas a partir de los parámetros espectrales AVG, VAR, ENER y SVD sobre la matriz de coeficientes DWT

Pacientes	Mejores canales	Promedios CSI
Paciente 1	3 - 15	0.6627
	6 - 15	0.6828
Paciente 2	4 - 15	0.9876
	1 - 4	1.0709
Paciente 3	4 - 6	0.7396
	6 - 12	0.8638
Paciente 4	10 - 13	2.4213
	7 - 13	2.6644
Paciente 5	8 - 12	1.4203
	6 - 8	2.3630
Paciente 6	8 - 12	1.3648
	10 - 12	1.4491

$$M_m = \sum_{n=0}^{N_j-1} \left(\frac{n}{N_j} \right)^{m-1} S[n] \quad m = 1, \dots, 6, \quad (4.14)$$

donde $S[n]$ son las secuencias de coeficientes de la descomposición wavelet DWT.

A continuación se ejecutan los procesos de: extracción y selección de características, y selección de canales.

a) Extracción y selección de características

La evaluación los *momentos* puede realizarse de dos maneras: una, sobre la matriz de coeficientes DWT compuesta por las subbandas seleccionadas D_2 y D_3 , y dos, sobre cada una de las subbandas D_2 y D_3 . A continuación se presentan los resultados obtenidos.

Opción 1. Evaluación de momentos espectrales sobre la matriz de coeficientes DWT.

En este caso se conforma la matriz de coeficientes con las subbandas de mayor concentración de energía D_2 y D_3 . De la evaluación estos *momentos* resultan 12 características espectrales DWT (6 momentos x 2 canales), y para cada una se promedia el índice *CSI* sobre todas las combinaciones de pares de canales. Estos promedios se presentan en la figura 4.20, donde se observa que los índices *CSI* promedios para las características 1, 2, 5, 6, 9 y 10 correspondientes a los momentos M_1 , M_3 y M_5 respectivamente son relativamente altos, por lo que se procede a eliminarlos.

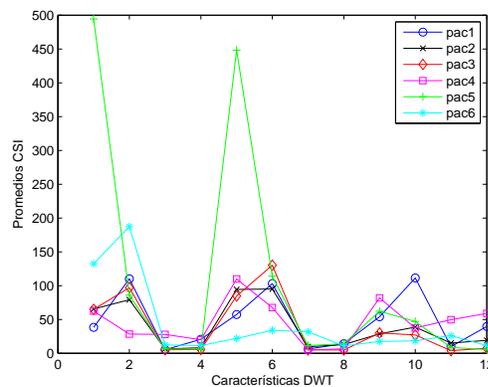


Figura 4.20. Promedios CSI para los momentos espectrales M_1 (caract. 1 y 2), ... M_6 (caract. 11 y 12), sobre la matriz de coeficientes DWT

En la figura 4.21, se muestran los índices *CSI* promedios para las características correspondientes a los momentos M_2 , M_4 y M_6 , después de eliminar los demás, donde se observa índices *CSI* promedios relativamente bajos, por lo que se consideran estos tres momentos (M_2 , M_4 y M_6) para el conjunto de características basado en momentos espectrales DWT.

Opción 2. Evaluación de momentos espectrales sobre cada subbanda de coeficientes DWT.

En este caso se evalúan los *momentos* sobre cada una de las subbandas de mayor concentración de energía D_2 y D_3 . De la evaluación estos *momentos* resultan 24 características espectrales DWT (6 momentos x 2 subbandas x 2 canales), y para cada una se promedia el índice *CSI* sobre todas las combinaciones de pares de canales. Estos promedios se presentan en la figura 4.22, donde se puede observar que los índices *CSI* promedios de las características

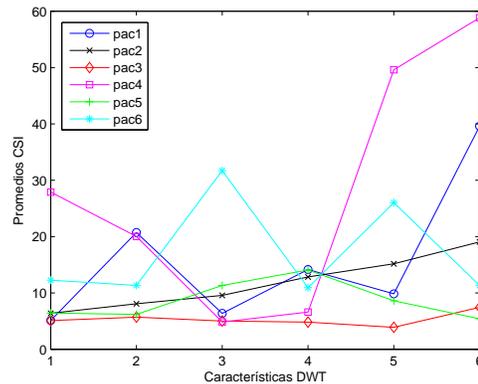


Figura 4.21. Promedios CSI para los momentos espectrales M_2 (caract. 1 y 2), M_4 (caract. 3 y 4) y M_6 (caract. 5 y 6), sobre la matriz de coeficientes DWT

1, 2, 3 y 4 (momento M_1), 9, 10, 11 y 12 (momento M_3), 17, 18, 19 y 20 (momento M_5) son relativamente altos en general para todos los pacientes, por lo que se procede a eliminarlos.

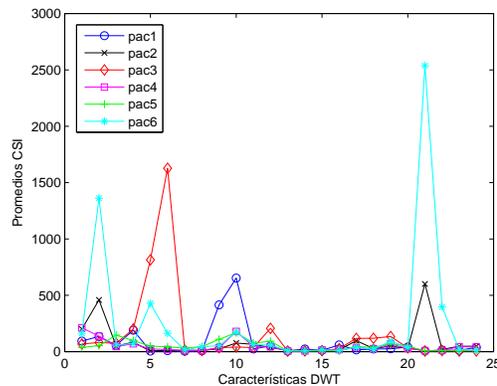


Figura 4.22. Índices CSI promedios para los momentos M_1 (caract. 1, 2, 3 y 4), ... M_6 (caract. 21, 22, 23 y 24) sobre cada subbanda D_2 y D_3 de coeficientes DWT

En la figura 4.23 se muestran los índices CSI promedios para los momentos M_2 , M_4 y M_6 , después de eliminar los demás, donde se observa índices CSI promedios bajos, en general para todos los pacientes, excepto por las características 1 y 2 para el paciente 5, la característica 9 para el paciente 2 y las características 9 y 10 para el paciente 6. Sin embargo, son casos aislados que no justifican la eliminación de dichas características.

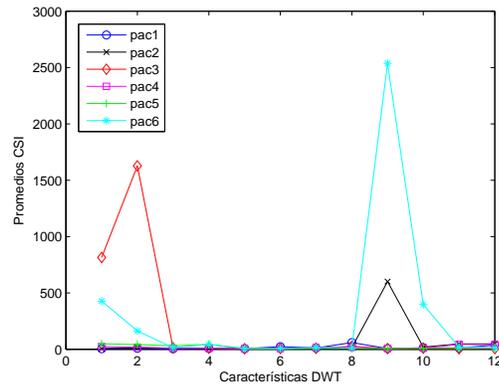


Figura 4.23. Promedios CSI para los momentos espectrales M_2 (caract. 1 y 2), M_4 (caract. 3 y 4) y M_6 (caract. 5 y 6), sobre las subbandas D_2 y D_3 de coeficientes DWT

La comparación de los índices CSI promedio por paciente para las dos formas de evaluar los *momentos espectrales*, figura 4.24, muestra que la evaluación de los momentos (M_2 , M_4 y M_6) sobre la matriz de coeficientes DWT es mejor opción que la evaluación sobre cada una de las subbandas D_2 y D_3 .

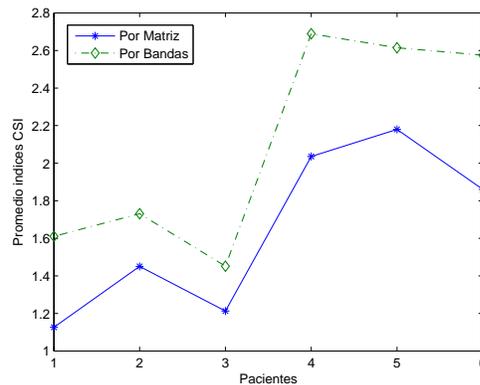


Figura 4.24. Promedios CSI por paciente según las opciones 1 y 2 del método de momentos

b) Selección de canales

Después de determinar los momentos M_2 , M_4 y M_6 , evaluados sobre la matriz de coeficientes DWT, como los mejores parámetros de extracción de características DWT, se determinan los dos (2) mejores pares de canales por

paciente seleccionando los dos mejores índices *CSI* promedios sobre las seis (6) características seleccionadas. Los resultados se resumen en la tabla 4.5.

Tabla 4.5. Mejores pares de canales según el índice *CSI* para 6 características DWT extraídas a partir de los momentos espectrales M_2 , M_4 y M_6 sobre la matriz de coeficientes DWT

Pacientes	Mejores canales	Promedios <i>CSI</i>
Paciente 1	4 - 15	1.1260
	5 - 15	1.1477
Paciente 2	4 - 15	1.4503
	3 - 15	1.6790
Paciente 3	4 - 12	1.2129
	10 - 12	1.3533
Paciente 4	11 - 13	2.0356
	7 - 13	2.4000
Paciente 5	8 - 12	2.1798
	8 - 10	2.2946
Paciente 6	8 - 13	1.8603
	8 - 12	1.9869

Para finalizar, en la figura 4.25 se comparan los índices *CSI* promedio por paciente para los tres métodos espectrales DWT descritos (con las mejores opciones en cada uno), donde se aprecia que el método de *parámetros*, produce índices *CSI* más bajos que los métodos de *valores singulares SVD* o de *momentos M_n* , al menos para cuatro de los seis pacientes.

4.4 Características Espectrales WPT

Dado que la descomposición WPT a nivel j resulta en $2(2^j - 1)$ subbandas, de las cuales algunas no contienen información útil para discriminar señales EMGS, es necesario aplicar un algoritmo de *selección de las mejores bases* para identificar las mejores subbandas, tal como se hizo con la descomposición DWT al seleccionar

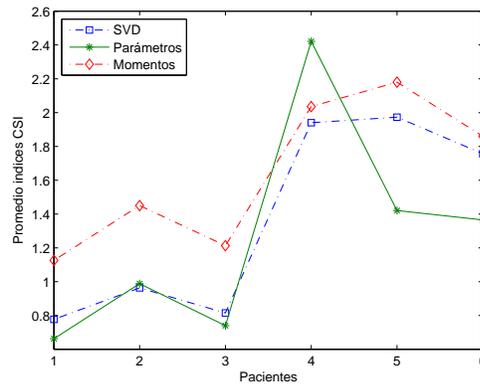


Figura 4.25. Promedios CSI por paciente para los tres métodos DWT

las subbandas D_2 y D_3 . Por esta razón, de los tres (3) métodos aplicados a la descomposición DWT, se descarta para la descomposición WPT el método de *valores singulares*, ya que solo aplica sobre la totalidad de coeficientes que resultan de la descomposición wavelet.

La descomposición wavelet WPT se realiza sobre registros EMGS de 256 muestras, y al igual que con la transformada DWT es necesario identificar el nivel óptimo de descomposición y la función wavelet más adecuada. En este trabajo se utiliza el algoritmo de *selección de las mejores bases de "Coifman-Wickerhauser"* (Coifman y Wickerhauser, 1992). Su algoritmo se describe en el anexo A.

4.4.1 Nivel de descomposición WPT

Al igual que con la transformada wavelet DWT, el nivel máximo de descomposición WPT sobre una secuencia temporal de longitud N está dado por $J = \log_2 N$. Para los registros EMGS de 256 muestras, su descomposición se puede realizar hasta el nivel $J = 8$, sin embargo, la descomposición a este nivel como en (Rodríguez y Vuskovic, 2006), no es lo más común, pues la mayoría de autores realizan la descomposición a nivel $j = 4$ como en Zhang, et al. (2002) y a nivel $j = 3$ como en Kilby y Hosseini (2004).

En Romo, et al. (2007), se comprobó que entre los niveles de descomposición 3, 4 y 5, el nivel óptimo de descomposición es el nivel 3, pues permite obtener mejores índices de separabilidad de clases CSI (figura 4.26), a la vez que se gana tiempo de procesamiento.

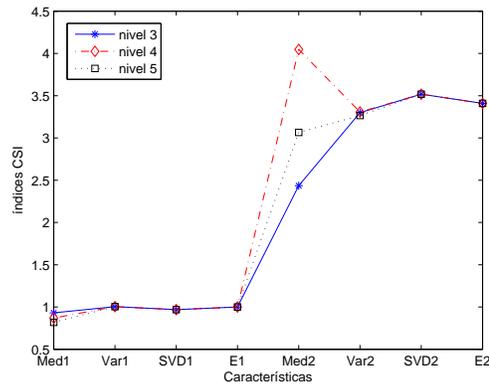


Figura 4.26. Índices CSI para 3, 4 y 5 niveles de descomposición WPT

4.4.2 Selección de la función wavelet

Para la selección de la función *wavelet* en la descomposición WPT, en Romo, et al. (2007) se compara el desempeño de las familias wavelets más utilizadas en los trabajos relacionados con señales mioeléctricas, entre ellas se destacan familias como: la *Daubechies* (*db*), la *Coiflet* (*Coif*) y la *Symlet* (*Sym*). La comparación de las mejores *wavelet* en cada familia muestra un comportamiento similar para las tres funciones aunque con una leve favorabilidad para la *Coif4*, (figura 4.27), razón por la cual se seleccionó para este trabajo la *wavelet madre Coiflet* de orden 4 (*Coif4*).

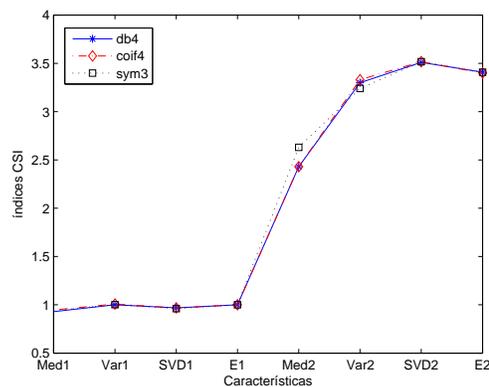


Figura 4.27. Índices CSI según las tres mejores familias wavelet madre para WPT

Una vez se ha determinado el nivel óptimo de descomposición wavelet WPT y la función wavelet más adecuada se procede a extraer las características espectrales

WPT, lo cual consiste en manipular de alguna manera los coeficientes que resultan de la descomposición WPT de la señal EMGS. Para la extracción de características espectrales WPT se implementaron los métodos: ***evaluación de parámetros*** y ***evaluación de momentos***. Los resultados se presentan a continuación.

4.4.3 Método 1. Evaluación de parámetros espectrales

Los parámetros seleccionados fueron: el *promedio absoluto de coeficientes* (*AVG*), la *energía de escala* (*ENER*), los *valores singulares* (*SVD*) y la *varianza* (*VAR*), los cuales son evaluados sobre los coeficientes wavelets de las subbandas que resultan de someter la descomposición WPT al algoritmo de *selección de las mejores bases de "Coifman-Wickerhauser"*. A continuación se ejecutan los procesos de: extracción y selección de características, y selección de canales.

a) Extracción y selección de características

Los parámetros *AVG*, *VAR*, *ENER* y *SVD* se evalúan mediante las expresiones (4.10) a (4.13) sobre las mejores bases WPT, resultando un conjunto de 8 características espectrales WPT (4 parámetros x 2 canales), y para cada una se promedia el índice *CSI* sobre todas las combinaciones de pares de canales. Estos promedios se presentan en la figura 4.28, donde se observa que los índices *CSI* promedios son bajos en general para todos los pacientes, excepto por las características 1 y 2 en los pacientes 2 y 5. Pero estos dos casos no justifican la eliminación de ninguna característica.

b) Selección de canales

Después de comprobar que todos los *parámetros* evaluados pueden utilizarse como características de descripción, se determinan los dos (2) mejores pares de canales por paciente seleccionando los dos mejores índices *CSI* promedios sobre las cuatro (4) características seleccionadas. Los resultados se resumen en la tabla 4.6.

4.4.4 Método 2. Evaluación de momentos espectrales

En este caso se considera la evaluación de seis (6) momentos M_i sobre los coeficientes de las mejores bases WPT mediante la expresión (4.14). A continuación se

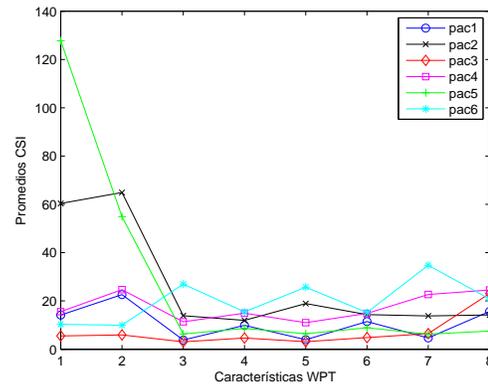


Figura 4.28. Promedios CSI para los parámetros AVG (caract. 1 y 2), VAR (caract. 3 y 4), ENER (caract. 5 y 6) y SVD (caract. 7 y 8) sobre las mejores subbandas WPT

ejecutan los procesos de: extracción y selección de características, y selección de canales.

a) Extracción y selección de características

Los momentos M_i se evalúan sobre los coeficientes WPT de las subbandas con mayor concentración de energía o información discriminante obtenidas de someter la descomposición wavelet WPT al algoritmo de selección de las mejores bases de “Coifman-Wickerhauser”. De la evaluación de estos *momentos* resultan 12 características espectrales WPT (6 momentos x 2 canales), y para cada una se promedia el índice *CSI* sobre todas las posibles combinaciones de pares de canales. Estos promedios se presentan en la figura 4.29, donde se puede observar que los índices *CSI* promedios para las características 1, 2, 5, 6, 9 y 10 (momentos M_1 , M_3 y M_5), son relativamente altos para todos los pacientes, por lo que se procede a eliminarlos.

En la figura 4.30 se muestran los índices *CSI* promedios para los momentos M_2 , M_4 y M_6 , después de eliminar los demás. Ahora se observa que los índices *CSI* promedios ya son relativamente bajos para todos los pacientes, excepto por las características 3 y 4 (momento M_2) para el paciente 3, pero es un caso muy particular que no justifica la eliminación de estas características.

Tabla 4.6. Mejores pares de canales según el índice CSI para 8 características WPT extraídas a partir de los parámetros AVG , VAR , $ENER$ y SVD sobre las mejores subbandas WPT

Pacientes	Mejores canales	Promedios CSI
Paciente 1	6 - 15	0.6734
	4 - 15	0.7383
Paciente 2	4 - 15	0.8922
	4 - 12	0.9889
Paciente 3	4 - 6	0.7642
	6 - 12	0.8840
Paciente 4	3 - 13	1.5763
	4 - 13	2.0276
Paciente 5	6 - 16	2.2103
	3 - 16	2.2914
Paciente 6	9 - 12	1.5592
	10 - 12	1.9234

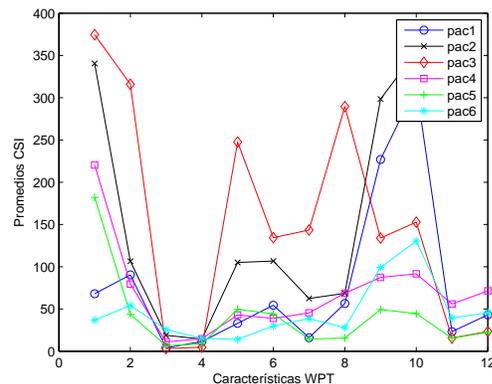


Figura 4.29. Índices CSI promedios para los momentos $M1$ (caract. 1 y 2), ... $M6$ (caract. 11 y 12), sobre las mejores subbandas WPT

b) Selección de canales

Después de determinar los momentos $M2$, $M4$ y $M6$, evaluados sobre las mejores bases WPT, como los mejores momentos para la extracción de ca-

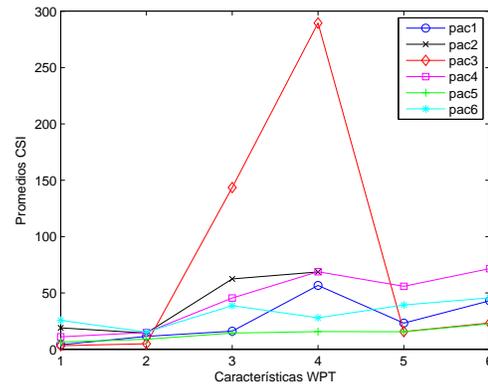


Figura 4.30. Índices CSI promedios para los momentos M_2 (caract. 1 y 2), M_4 (caract. 3 y 4) y M_6 (caract. 5 y 6), sobre las mejores subbandas WPT

racterísticas WPT, se determinan los dos (2) mejores pares de canales por paciente seleccionando los dos mejores índices CSI promedios sobre las seis (6) características seleccionadas. Los resultados se resumen en la tabla 4.7.

En la figura 4.31 se compara de los índices CSI promedio por paciente para los dos métodos espectrales WPT descritos, donde se aprecia que el método de *parámetros*, produce índices CSI más bajos que el método de *momentos* M_n para cinco de los seis pacientes.

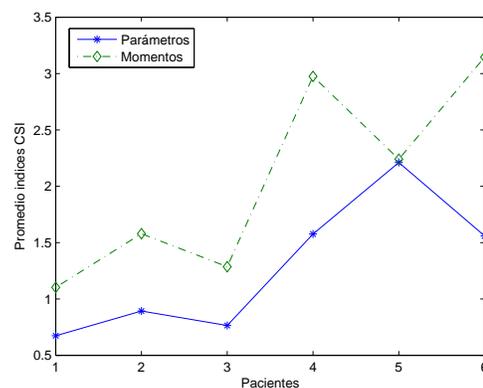


Figura 4.31. Promedios CSI por paciente para los métodos de *parámetros* y *momentos* WPT

Finalmente, la comparación de los mejores métodos en cada una de las técnicas analizadas (temporal y espectrales DWT y WPT), según sus índices CSI más bajos

Tabla 4.7. Mejores pares de canales según el índice CSI para 6 características WPT extraídas a partir de los momentos M_2 , M_4 y M_6 sobre las mejores subbandas WPT

Pacientes	Mejores canales	Promedios CSI
Paciente 1	4 - 15	1.1260
	5 - 15	1.1477
Paciente 2	4 - 15	1.4503
	3 - 15	1.6790
Paciente 3	4 - 12	1.2129
	10 - 12	1.3533
Paciente 4	11 - 13	2.0356
	7 - 13	2.4000
Paciente 5	8 - 12	2.1798
	8 - 10	2.2946
Paciente 6	8 - 13	1.8603
	8 - 12	1.9869

observados en cada caso, figura 4.32, refleja la misma tendencia en las tres curvas de índices CSI sobre los seis (6) pacientes, lo cual comprueba la eficiencia de los métodos analizados. Los índices CSI son mas bajos para los pacientes 1, 2 y 3 y aumentan un poco para los pacientes 4, 5, y 6, lo cual puede obedecer a la calidad de los registros EMGS.

4.5 Conformación de los vectores híbridos DT-DWT y DT-WPT

La conformación de los vectores descriptores para la clasificación de señales EMGS finalmente consiste en la combinación de características temporales y espectrales wavelets, dando lugar a los vectores híbridos: *DT-DWT* y *DT-WPT*. Estos vectores son evaluados mediante el error producido en la clasificación de las señales EMGS.

Según el índice CSI se determinó que el mejor conjunto de características en cada técnica de análisis son: los *parámetros DT, DWT y WPT*. Con su combinación

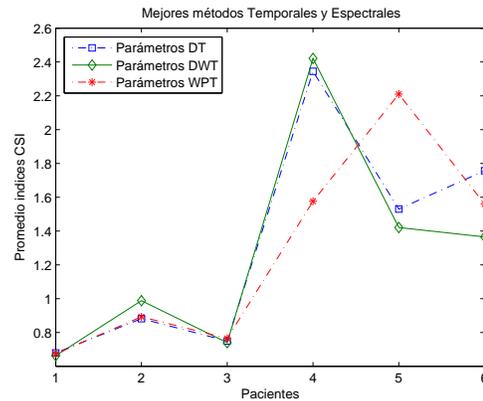


Figura 4.32. Promedios CSI por paciente para los mejores métodos en las técnicas DT, DWT y WPT

se identifica para cada paciente el mejor par de canales común a las tres técnicas de análisis (DT, DWT y WPT), pero ahora teniendo en cuenta el error de clasificación. Hasta ahora se han identificado los dos mejores pares de canales por paciente, los cuales no necesariamente coinciden para los seis pacientes, por lo que para la fase de clasificación será necesario considerar al menos tres de los mejores pares de canales según los menores índices CSI. De esta manera, serán las pruebas de clasificación de señales EMGS las que finalmente determinen el mejor par de canales por paciente. Las ternas de los mejores canales por paciente se registran en la tabla 4.8, con sus índices CSI promedio respectivos.

Para concluir, en la figura 4.33 se comparan los índices CSI promedio por paciente para las dos formas de combinación de características: *DT-DWT* y *DT-WPT*, con los *parámetros DT, DWT y WPT*. Se observa que para los pacientes 1 y 3 los índices CSI coinciden, para los pacientes 2 y 4 la combinación *DT-WPT* tiene índices CSI más bajos y para los pacientes 5 y 6 la combinación *DT-DWT* tiene índices CSI más bajos. De aquí se puede concluir que en principio las dos formas de combinación de características tienen un buen desempeño como vectores híbridos de descripción de señales EMGS. Sin embargo, serán las pruebas de clasificación de patrones EMGS las que finalmente determinen cual de estas dos combinaciones resulta ser la mejor opción.

Tabla 4.8. Mejores pares de canales según índices CSI promedios comunes a las técnicas DT, DWT y WPT

Paciente	Canales	CSI DT	CSI DWT	CSI WPT	CSI Promedio
Paciente 1	6 - 15	0.679	0.683	0.673	0.678
	3 - 15	0.818	0.663	0.755	0.745
	4 - 15	6.011	0.786	0.738	2.512
Paciente 2	4 - 15	0.978	0.988	0.892	0.953
	1 - 4	0.882	1.071	0.993	0.982
	4 - 12	2.248	1.955	0.989	1.731
Paciente 3	4 - 6	0.749	0.740	0.764	0.751
	6 - 12	0.935	0.864	0.884	0.894
	1 - 4	0.820	1.168	1.028	1.005
Paciente 4	11 - 13	2.345	2.775	2.291	2.470
	3 - 13	3.482	2.876	1.576	2.645
	7 - 13	3.251	2.664	2.200	2.705
Paciente 5	6 - 8	1.530	2.363	3.541	2.478
	8 - 16	1.700	3.429	3.237	2.789
	6 - 16	1.761	4.617	2.210	2.863
Paciente 6	10 - 12	1.756	1.449	1.924	1.710
	9 - 12	1.967	1.691	1.559	1.739
	8 - 12	3.239	1.365	6.177	3.594

4.6 Implementación de Algoritmos y Funciones

Los algoritmos y funciones implementadas para llevar a cabo las pruebas del sistema de clasificación de señales MEGS en cada una de sus fases de: extracción de características, reducción de dimensionalidad y clasificación de señales EMGS, fueron implementadas en el paquete de simulación de Matlab 7.0, con la utilización de las librerías del “toolbox” de *wavelet* versión 3.0.3 y las librerías del paquete *wavelab850*.

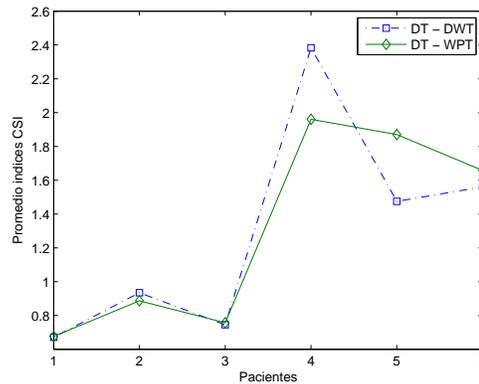


Figura 4.33. Promedios CSI por paciente para las combinaciones DT-DWT y DT-WPT

4.6.1 Evaluación de índices CSI

Para la evaluación de los índices CSI por característica se implementó la función:

- ***evalua_CSI.m***. Con sintaxis:

indices_csi = evalua_csi(caracteristicas).

La variable de entrada “*caracteristicas*”, consiste en un arreglo vectorial de las características extraídas en cada método. La salida de esta función es un vector con los índices *CSI* de cada característica.

4.6.2 Extracción de Características DT

Para la extracción de características temporales DT se implementó la función:

- ***caract_tempo.m***. Con sintaxis:

caracteristicas_DT = caract_tempo(ch1,ch2,metodo).

Las variables de entrada “*ch1*” y “*ch2*” son las secciones de 256 muestras de los registros EMGS de cada canal, la variable de entrada “*metodo*”, identifica el método de extracción de características temporales (*parámetros* o *momentos*). La salida es un vector de características DT, extraídas según el método seleccionado.

4.6.3 Extracción de Características DWT

Para la extracción de características espectrales basadas en wavelets DWT se implementaron las siguientes funciones:

- ***caract_dwt.m***. Para extraer las características DWT, con sintaxis:

caracteristicas_DWT = caract_dwt(ch1,ch2,metodo).

Las variables de entrada “*ch1*” y “*ch2*” son las secciones de 256 muestras de los registros EMGS de cada canal, la variable de entrada “*metodo*”, identifica uno de los tres métodos espectrales DWT: *valores singulares*, *parámetros* y *momentos*. La salida es un vector de características DWT, extraídas según el método seleccionado.

- ***WTAnalysis.m***. Para la descomposición wavelet DWT, con sintaxis:

dwt = WTAnalysis(emgA(:,j),n,w).

Las variables de entrada: “*emgA*” representa un segmento de la señal EMGS de longitud 256 muestras, “*n*” el nivel de la descomposición DWT y “*w*” la función wavelet (*db3* en este caso). La salida “*dwt*” es un arreglo matricial con los coeficientes que resultan de la descomposición wavelet DWT.

4.6.4 Extracción de Características WPT

Para la extracción de características espectrales basadas en wavelets WPT se implementaron las siguientes funciones:

- ***caract_wpt.m***. Para extraer las características WPT, con sintaxis:

caracteristicas_WPT = caract_wpt(ch1,ch2,metodo).

Las variables de entrada “*ch1*” y “*ch2*” son las secciones de 256 muestras de los registros EMGS de cada canal y la variable de entrada “*metodo*”, identifica el método de extracción de características espectrales WPT (*parámetros* o *momentos*). La salida es un vector de características WPT, extraídas según el método seleccionado.

- **WPAanalysis.m**. Para la descomposición wavelet WPT, con sintaxis:

wpt = WPAanalysis(emgA(:,j),n,qmf).

Las variables de entrada: “*emgA*” representa un segmento de señal EMGS de longitud 256 muestras, “*n*” el nivel de la descomposición WPT y “*qmf*” los coeficientes de los filtros wavelets “*coif4*”. La salida “*wpt*” es un arreglo matricial con los coeficientes en la estructura de árbol que resulta de la descomposición wavelet WPT.

- **BestBasisV2**. Para la selección de las mejores bases WPT, con sintaxis:

[mejoresBases,btree,vtree]= BestBasisV2(ent_nodos,wpt,TE).

Las variables de entrada: “*ent_nodos*” representa el árbol binario con las entropías de cada nodo, “*wpt*” la matriz de coeficientes de la descomposición wavelet WPT y “*TE*” el tipo de entropía (en este caso de “Shannon”). La salida contiene un arreglo vectorial con: las “*mejores bases WPT*”, el árbol binario de las mejores bases “*btree*” y el árbol de entropías de las mejores bases “*vtree*”.

Capítulo 5

Clasificación de Patrones EMGS

En el presente capítulo se someten los vectores híbridos de descripción *DT-DWT* y *DT-WPT* de las señales EMGS a la fase de clasificación de patrones para evaluar la fase de representación de señales EMGS. Para esto se dispone de una base de datos de entrenamiento y otra de pruebas. Como clasificadores se implementaron: una *red neuronal probabilística* PNN por sus características híbridas de los clasificadores estadísticos y los basados en aprendizaje neuronal, una *máquina de soporte vectorial* SVM por ser un clasificador de aprendizaje automático, y una red tipo *perceptrón multicapas* MLP por ser un clasificador con capacidad de aprendizaje basado en redes neuronales. Los tres son clasificadores modernos, eficientes y muy populares en aplicaciones de reconocimiento de patrones (Rosa y Gil , 2006).

Inicialmente se utilizan los clasificadores PNN y SVM, gracias a que su desempeño depende únicamente de sus parámetros de diseño y con una sola pasada sobre los datos de entrenamiento queda la configuración definitiva para la clasificación de los datos de prueba. Esta característica permite realizar una evaluación objetiva (no dependiente del esfuerzo de entrenamiento) de la fase de representación de las señales EMGS mediante los vectores híbridos *DT-DWT* y *DT-WPT*. El clasificador MLP se utiliza al final de las pruebas, donde se compara el desempeño de los tres clasificadores sobre las mejores combinaciones de características *DT-DWT* y *DT-WPT*. Los parámetros de diseño y configuración óptimos de los tres clasificadores se lograron con la colaboración de Rengifo y Belalcazar (2008), en el desarrollo de su trabajo de grado "*Clasificación de señales electro-miográficas para una prótesis de mano*".

5.1 Bases de Datos de Señales EMGS

Las señales EMGS a ser clasificadas corresponden a cuatro clases de movimientos mioeléctricos de la mano, extraídas de la base de datos “*LeviData.zip*”. Para la construcción de esta base de datos el paciente ejecutó para cada movimiento una contracción isométrica completa de mediana fuerza durante 5 segundos seguida de un breve periodo de descanso. Cada movimiento fue ejecutado uno tras otro repitiendo esta secuencia dos veces. En este orden fueron realizadas dos sesiones de grabación, una para obtener la base de datos de *entrenamiento* y la otra para obtener la base de datos de *prueba*. El formato de presentación de las bases de datos obedece a los siguientes datos: 5 segundos de grabación por movimiento con frecuencia de muestreo de 1024 para 4 movimientos en 2 sesiones. Los registros de cada movimiento se segmentan en secciones de 256 muestras obteniéndose un arreglo matricial de dimensiones 256x160 para cada paciente. Esto es 160 muestras EMGS para entrenamiento y 160 muestras EMGS diferentes para pruebas (40 patrones EMGS por clase o movimiento).

5.2 Selección de Canales para la Clasificación de las Señales EMGS

En la tabla 8 del capítulo 4 se presentó para cada paciente los tres mejores pares de canales según los índices CSI promedio mas bajos observados para las características DT, DWT y WPT. A partir de esta tabla se evalúan las combinaciones de características *DT-DWT* y *DT-WPT*, con base en el error de clasificación producido por los clasificadores PNN y SVM.

Para la combinación DT-DWT, figura 5.1, se observa que para el clasificador PNN el error de clasificación promedio sobre los 6 pacientes aumenta conforme aumenta el índice CSI, sin embargo para el clasificador SVM se tiene el menor error de clasificación para el segundo mejor índice CSI.

Para la combinación DT-WPT, figura 5.2, se observa que ni para el clasificador PNN ni para el clasificador SVM, el error de clasificación promedio sobre los 6 pacientes aumenta conforme aumentan los índices CSI. Pues mientras que para el clasificador PNN el menor error de clasificación se presenta para el primer índice CSI, para el clasificador SVM el menor error de clasificación se da para segundo

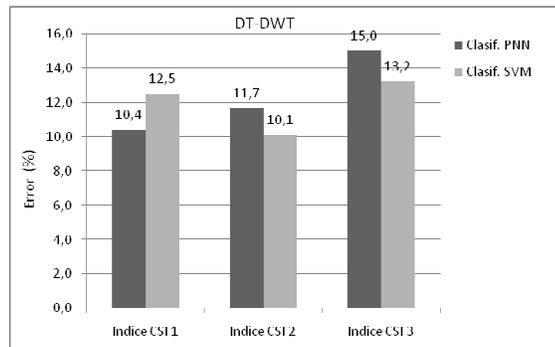


Figura 5.1. Error de clasificación de los tres mejores índices CSI para DT-DWT

mejor índice CSI. Esto justifica por que considerar por lo menos los tres mejores índices CSI en la fase de representación de las señales, pues no siempre el menor índice CSI sera garantía del menor error de clasificación para cualquier tipo de clasificador. A pesar de esto, se demuestra que el índice de separabilidad de clases CSI sigue siendo una herramienta útil en la fase de representación de señales EMGS, pues garantiza un desempeño aceptable en la clasificación de patrones EMGS. Por lo anterior, la selección del mejor par de canales se debe realizar individualmente para cada paciente y de acuerdo al menor error de clasificación sobre los tres (3) mejores índices CSI. Siguiendo este proceso se encontraron los mejores canales por paciente, los cuales se consignan en la tabla 5.1.

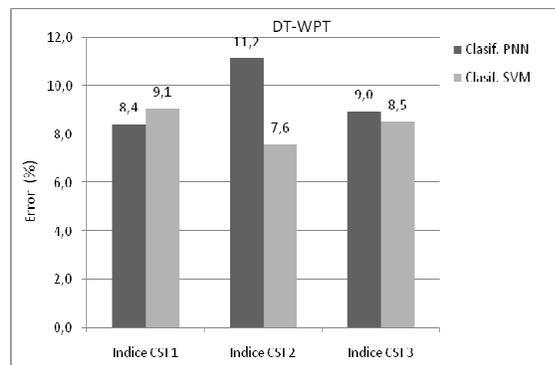


Figura 5.2. Error de clasificación de los tres mejores índices CSI para DT-WPT

Tabla 5.1. Mejores pares de canales por paciente según el mínimo error de clasificación

Paciente	Mejores canales
Paciente 1	6 - 15
Paciente 2	4 - 15
Paciente 3	1 - 4
Paciente 4	3 - 13
Paciente 5	6 - 8
Paciente 6	8 - 12

5.3 Características DT-DWT y DT-WPT

A pesar de que los mejores índices *CSI* se presentaron para las combinaciones de características: *parámetros temporales DT* y *parámetros espectrales (DWT y WPT)*, es necesario comparar los errores de clasificación para cada una de las combinaciones de características *DT-DWT* y *DT-WPT*, según los métodos de extracción de características temporales (*parámetros y momentos*), espectrales DWT (*Valores singulares, parámetros y momentos*) y espectrales WPT (*parámetros y momentos*), analizados en el capítulo 4.

Así, los errores de clasificación para la combinación de características *DT-DWT*, según los métodos descritos y para los clasificadores PNN y SVM, se presentan en las figuras 5.3 y 5.4 respectivamente, donde se puede observar que para los dos clasificadores el menor error de clasificación se presenta para la combinación de los métodos: *momentos DT* y *valores singulares DWT*, seguido de la combinación *parámetros DT* y *parámetros DWT*, con una diferencia en sus errores de clasificación que solo implica 3 patrones demás en la base de datos de prueba.

Y los errores de clasificación para la combinación de características *DT-WPT*, según los métodos descritos y para los clasificadores PNN y SVM, se presentan en las figuras 5.5 y 5.6 respectivamente, donde se puede observar que para los dos clasificadores el menor error de clasificación se presenta para la combinación de los métodos: *parámetros DT* y *parámetros WPT*, coincidiendo con la mejor combinación, según el índice *CSI*, encontrada durante la fase de representación de las

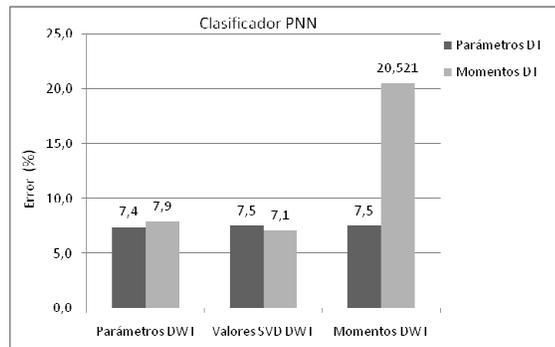


Figura 5.3. Error de clasificación PNN para las combinaciones híbridas DT-DWT

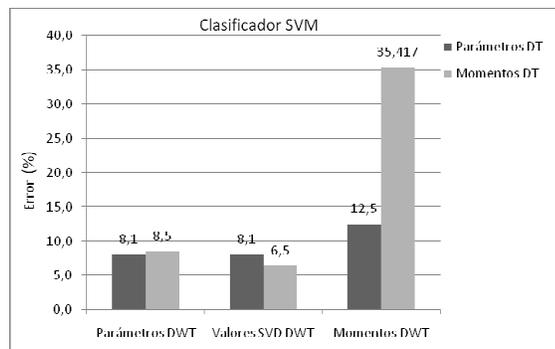


Figura 5.4. Error de clasificación SVM para las combinaciones híbridas DT-DWT

señales EMGS.

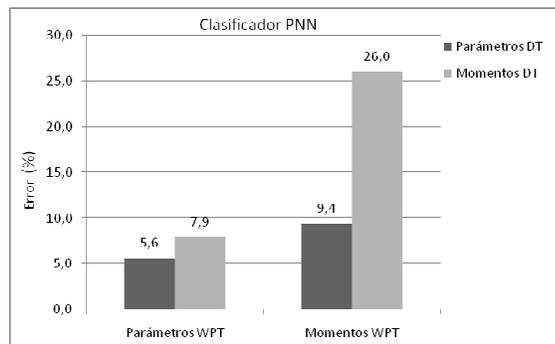


Figura 5.5. Error de clasificación PNN para las combinaciones híbridas DT-WPT

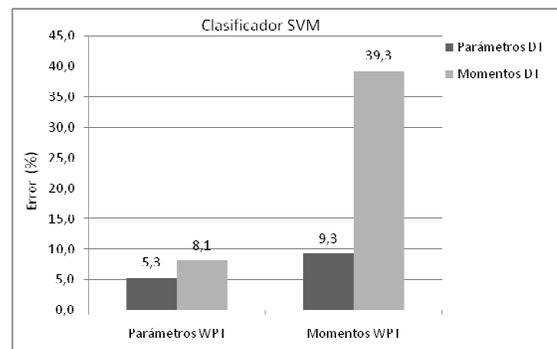


Figura 5.6. Error de clasificación SVM para las combinaciones híbridas DT-WPT

5.4 Pruebas de Clasificación con Análisis PCA

La reducción de dimensionalidad mediante *análisis* PCA consiste en reducir la dimensión del espacio transformado antes de extraer las características descriptoras. Aquí se compara el error de clasificación producido por el clasificador PNN sobre los vectores híbridos *DT-DWT* combinando características obtenidas de los *parámetros temporales DT* y los *parámetros espectrales DWT*, estos últimos evaluados de dos maneras: una, evaluándolos sobre la matriz de coeficientes wavelet DWT sometida al algoritmo PCA, y dos, evaluándolos sobre los coeficientes wavelet DWT de las subbandas D_2 y D_3 , eliminando las subbandas D_1 y A_3 . Los resultados obtenidos se presentan en la figura 5.7, donde se observa que la simple eliminación de las subbandas D_1 y A_3 de la matriz de coeficientes DWT produce errores de clasificación comparables a los logrados mediante la reducción con PCA. Sin embargo, la simple eliminación de subbandas implica menor complejidad que el algoritmo PCA y por tanto menor tiempo de procesamiento.

Otra forma de reducción de dimensionalidad mediante *análisis* PCA consiste en reducir la dimensión de los vectores descriptores, pues la transformación PCA elimina la redundancia o correlación que pueda existir entre las características del vector. Ahora se compara el error de clasificación que produce el clasificador PNN sobre los vectores híbridos *DT-DWT* y *DT-WPT*, combinando características obtenidas de los *parámetros DT* y los *parámetros DWT* y *WPT*, primero sometiéndolos al algoritmo PCA y luego pasándolos directamente al clasificador. Los resultados se presentan en la figura 5.8, donde se observa que el error de clasificación no mejora después de someter al análisis PCA los vectores híbridos de descripción, lo cual

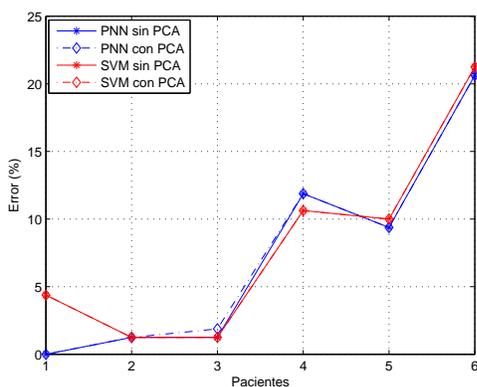


Figura 5.7. Error de clasificación bajo la condición del análisis PCA sobre los coeficientes DWT

significa que la dimensión original de los vectores *DT-DWT* y *DT-WPT* es la más adecuada y optimizada para propósitos de clasificación de patrones EMGS.

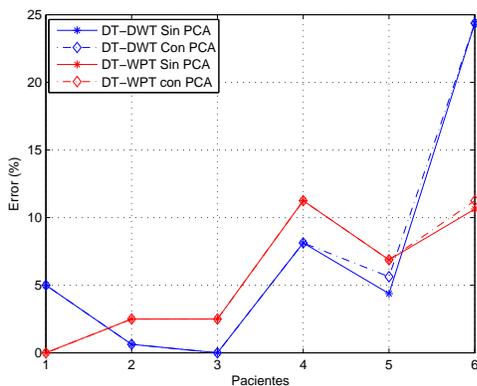


Figura 5.8. Error de clasificación según el clasificador PNN en los vectores DT-DWT y DT-WPT con y sin análisis PCA

5.5 Error de clasificación para las combinaciones de características DT-DWT y DT-WPT

Teniendo en cuenta los mejores canales por paciente según los mejores índices CSI y posteriormente los menores errores de clasificación, se procede a comparar el error de clasificación de patrones EMGS para las combinaciones de características

DT-DWT y DT-WPT. Para esto se utilizan los clasificadores PNN, SVM y MLP. Los dos primeros tienen la característica de que su tasa de error de clasificación se mantiene constante una vez fijados sus parámetros respectivos de diseño, puesto que su entrenamiento solo requiere de una sola pasada sobre los datos de entrenamiento, mientras que con el clasificador MLP una vez fijados sus parámetros de diseño, puede buscarse un error de clasificación cada vez menor con la reiteración del entrenamiento. Los resultados de la clasificación de patrones EMGS bajo la condición de parámetros de diseño optimizados para cada clasificador se presentan en la tabla 5.2.

Tabla 5.2. Error de clasificación para las técnicas DT-DWT y DT-WPT

Paciente	PNN		SVM		MLP	
	DT-DWT	DT-WPT	DT-DWT	DT-WPT	DT-DWT	DT-WPT
Paciente 1	5.000	0.000	0.000	4.375	1.250	0.625
Paciente 2	0.625	2.500	1.875	2.500	0.000	0.625
Paciente 3	0.000	2.500	0.000	1.250	1.250	1.875
Paciente 4	8.125	11.250	6.875	6.875	3.125	3.125
Paciente 5	4.375	6.875	3.750	6.875	3.125	3.750
Paciente 6	24.375	10.625	26.250	10.000	6.875	5.000
	7.083	5.625	6.458	5.312	2.604	2.500

En la figura 5.9 se puede observar que el menor error de clasificación en cada uno de los clasificadores se presenta para la combinación de características DT-WPT, y el menor error de clasificación para las dos formas de combinación de características DT-DWT y DT-WPT se presenta para el clasificador MLP, seguido del clasificador SVM.

Finalmente en la figura 5.10, se puede comprobar en los errores de clasificación un comportamiento similar al de los índices CSI promedios para cada uno de los seis (6) pacientes. Esto permite concluir que el criterio del índice CSI es un buen indicador de la capacidad de discriminación de clases que posee una característica, y que puede utilizarse como herramienta confiable en la fase de reducción de dimensionalidad por selección de características para propósitos de clasificación de patrones EMGS.

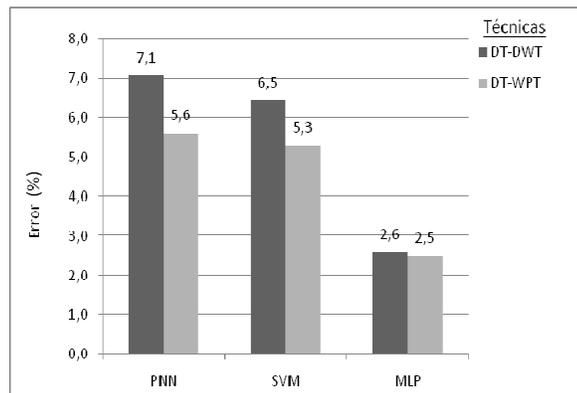


Figura 5.9. Error de clasificación para las combinaciones de características DT-DWT y DT-WPT

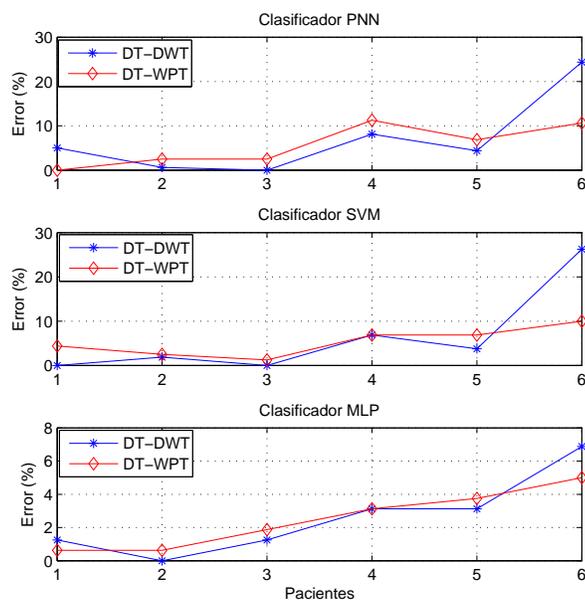


Figura 5.10. Error de clasificación por paciente de las técnicas DT-DWT y DT-WPT

5.6 Implementación de Algoritmos y Funciones

Para la fase de clasificación de patrones EMGS se implementaron las siguientes funciones:

- **corpus_emg.m**. Para extraer los vectores híbridos de características *DT-DWT* y *DT-WPT*, vistos como los patrones EMGS de entrenamiento y de pruebas,

con sintaxis:

[características,indicesCSI] = corpus_emg(ch1,ch2,tipo,paciente).

Las variables de entrada “*ch1*” y “*ch2*” son las secciones de 256 muestras de los registros EMGS de cada canal, la variable de entrada “*tipo*” identifica el tipo de registros EMGS (de *entrenamiento* o de *pruebas*), y la variable “*paciente*” identifica uno de los seis pacientes. La salida son vectores de “*características*” *DT-DWT* y *DT-WPT*, junto con los vectores que contienen los respectivos índices CSI por característica.

- ***clasificador_emg.m.*** Encargada de cargar los patrones EMGS (vectores híbridos de características) bien sea para la fase de entrenamiento de los clasificadores o para la clasificación de los patrones EMGS de prueba. Esta función invoca las respectivas funciones de los clasificadores.
- ***clasifica_con_PNN.m.*** Para clasificar con la *red neuronal probabilística*, con sintaxis:

[resultado,matriz_clasif] = clasifica_PNN(base_datos,patrones_prueba).

Las variables de entrada: “*base_datos*”, representa el arreglo matricial de patrones EMGS de entrenamiento, y “*patrones_prueba*” representa los patrones EMGS para clasificación. Las salidas son: “*resultado*” que contiene el porcentaje del error de la clasificación, y “*matriz_clasif*” un arreglo de patrones EMGS clasificados en cuatro clases.

- ***newpnn.m.*** Para crear y entrenar la red neuronal PNN, con sintaxis:

net = newpnn(P_e,T,spread).

Las variables de entrada: “*P_e*” representa los parámetros EMGS de entrenamiento, “*T*” es un arreglo vectorial que identifica la clase a la que pertenece cada patrón EMGS, y “*spread*” es el parámetro de diseño de la red PNN que indica la dispersión de las funciones de base radial. La salida “*net*” contiene la estructura de la red PNN.

- ***clasifica_con_SVM.m.*** Para clasificar con la *máquina de soporte vectorial*, con sintaxis:

[resultado,matriz_clasif] = clasifica_SVM(base_datos,patrones_prueba, paciente,tecnica).

Las variables de entrada: “*base_datos*”, representa el arreglo matricial de patrones EMGS de entrenamiento, “*patrones_prueba*” representa los patrones EMGS para clasificación, “*paciente*” identifica uno de los seis pacientes para cargar la red entrenada respectiva, y “*tecnica*” identifica la combinación de características *DT-DWT* o *DT-WPT* para cargar la red entrenada respectiva. Las salidas son: “*resultado*” que contiene el porcentaje del error de la clasificación, y “*matriz_clasif*” un arreglo de patrones EMGS clasificados en cuatro clases.

- **LinearSVC.m.** Para crear y entrenar la red SVM, con sintaxis:

[AlphaY, SVs, Bias, Parameters, nSV, nLabel] = LinearSVC(P_e,Tc).

Las variables de entrada: “*P_e*” representa los patrones EMGS de entrenamiento y “*Tc*” las respectivas etiquetas de clase de cada patrón de entrenamiento EMGS. Las salidas: “*AlphaY*” contiene los coeficientes de Lagrange no nulos, “*SVs*” contiene los vectores soporte, “*Bias*” contiene todos los “sesgos” de los clasificadores binarios SVM, “*Parameters*” son los parámetros de salida del entrenamiento, “*nSV*” es el número de vectores soporte en cada clase y “*nLabel*” son las etiquetas en cada clase.

- **clasifica_con_MLP.m.** Para clasificar con la red neuronal *perceptrón multicapas*, con sintaxis:

[resultado,matriz_clasif] = clasifica_MLP(base_datos,patrones_prueba, paciente,tecnica).

Las variables de entrada: “*base_datos*” representa el arreglo matricial de patrones EMGS de entrenamiento, “*patrones_prueba*” representa los patrones EMGS para clasificación, “*paciente*” identifica uno de los seis pacientes para cargar la red entrenada respectiva, y “*tecnica*” identifica la combinación de características *DT-DWT* o *DT-WPT* para cargar la red entrenada respectiva. Las salidas son: “*resultado*” que contiene el porcentaje del error de la clasificación, y “*matriz_clasif*” un arreglo de patrones EMGS clasificados en cuatro clases.

- **newff.m.** Para crear la red MLP, con sintaxis:

net = newff(PR,[S1 S2...SNI],{TF1 TF2...TFNI}).

Las variables de entrada: “PR” contiene los valores mínimo y máximo del arreglo de patrones EMGS de entrenamiento, el arreglo $[S_1 S_2 \dots S_{NI}]$ indica el número de neuronas en cada capa S_i de la red MLP y el arreglo $\{TF_1 TF_2 \dots TF_{NI}\}$ especifica el tipo de función de transferencia en cada capa.

- ***train.m.*** Para entrenar la red MLP creada, con sintaxis:

net = train(net,P_e,Tc).

Las variables de entrada: “net” representa la estructura de la red MLP a entrenar, “P_e” es el arreglo de patrones EMGS de entrenamiento y “Tc” las respectivas etiquetas de clase de los patrones de entrenamiento. La salida “net” contiene la estructura de red MLP diseñada y entrenada.

Capítulo 6

Conclusiones

El propósito de este trabajo fue investigar la aplicación de las transformadas wavelets DWT y WPT en el análisis y extracción de características con fines de clasificación de señales EMGS. Se planteó como objetivo general el analizar señales EMGS con técnicas basadas en wavelets para extraer características espectrales que ayuden a mejorar la clasificación de patrones de movimientos mioeléctricos. En este orden de ideas se revisó el origen fisiológico de las señales EMGS, su generación y detección mediante electrodos de superficie, y la revisión del estado del arte permitió escoger para estos propósitos las señales EMG superficiales (EMGS) en su fase estacionaria, por considerarse la forma más sencilla de obtenerlas y la menos traumática para el paciente.

Se describió en detalle las fases del sistema de clasificación de señales EMGS, que incluye la representación de las señales EMGS y la clasificación de sus patrones respectivos. Dentro de la representación de señales se desarrollaron los procesos de extracción de características y de reducción de dimensionalidad. Para la extracción de características, se seleccionaron e implementaron: los métodos de *parámetros* y *momentos* temporales DT, los métodos de *valores singulares*, *parámetros* y *momentos* espectrales DWT y los métodos de *parámetros* y *momentos* espectrales WPT. Para la fase de reducción de dimensionalidad se aplicó la técnica *selección de características* según la medida del índice *CSI*, algoritmo de "*Davies-Bouldin*", con el fin de dimensionar adecuadamente vectores híbridos de características DT-DWT y DT-WPT.

La fase de clasificación de patrones permitió evaluar la representación de señales

EMGS, utilizando primero los clasificadores PNN y SVM por su relativa sencillez y eficiencia, pues su entrenamiento solo requiere de una sola pasada sobre los datos, lo que genera una solución única de clasificación y los errores dependen más de la fase de representación de las señales que del entrenamiento mismo. Luego se incluye el clasificador MLP con algoritmo de entrenamiento “*back-propagation*”, con el que se consigue minimizar aun más el error de clasificación alcanzado por los clasificadores PNN y SVM.

De la implementación de los diferentes métodos de extracción de características y de acuerdo a la medida del índice *CSI*, se determinó que las características extraídas a partir de: los *parámetros temporales* DT y los *parámetros espectrales* DWT y WPT, son las más representativas de las señales EMGS. Se observó además un comportamiento similar en los índices *CSI* promedio por paciente para las tres técnicas de análisis (DT, DWT y WPT), lo cual indica que son técnicas igualmente eficientes además de complementarias para la representación de estas señales.

La evaluación del índice *CSI*, acompañado de la selección manual de las subbandas D_2 y D_3 en el caso de la descomposición DWT, y acompañado de la selección de las mejores bases, mediante el algoritmo de “*Coifman-Wickerhauser*”, en el caso de la descomposición WPT, demostró ser una herramienta muy eficiente en la reducción de dimensionalidad, utilizado en este trabajo para la conformación de los vectores híbridos DT-DWT y DT-WPT y para la selección de los mejores pares de canales para cada paciente.

Las pruebas de clasificación para los tres mejores índices *CSI* en las combinaciones de características DT-DWT y DT-WPT, mostraron el menor error de clasificación para el mejor índice *CSI* según el clasificador PNN, mientras que para el clasificador SVM mostraron el menor error de clasificación para el segundo mejor índice *CSI*. De aquí se concluye que no necesariamente el menor índice *CSI* implicará el menor error de clasificación con cualquier clasificador, por consiguiente se recomienda considerar al menos dos de los mejores índices *CSI* para asegurar los menores errores de clasificación y garantizar una selección adecuada de los mejores pares de canales por paciente. La selección del mejor par de canales para cada paciente mostró que no siempre son los mismos, por lo que este proceso de selección debe realizarse independientemente para cada paciente.

Las pruebas de clasificación para las combinaciones de características DT-DWT, según los clasificadores PNN y SVM mostraron: el menor error de clasificación para la combinación de los *momentos* DT y los *valores singulares* DWT, y el segundo menor error de clasificación para la combinación de *parámetros* DT y los *parámetros* DWT. Por otro lado, las pruebas de clasificación para las combinaciones de características DT-WPT, según los clasificadores PNN y SVM, si mostraron el menor error de clasificación para la combinación de los *parámetros* DT y los *parámetros* WPT, coincidiendo en este caso la combinación del mejor índice *CSI* en la fase de representación de las señales. De aquí se concluye que hay mayor correspondencia entre el menor índice de separabilidad de clases *CSI* y el menor error de clasificación para la combinación DT-WPT.

Adicionalmente, se realizaron pruebas de clasificación sometiendo tanto la matriz de coeficientes wavelets DWT como los vectores de características DT-DWT y DT-WPT al algoritmo de reducción de dimensionalidad *PCA*, no encontrándose una mejoría que justificara su implementación. De aquí se concluye que los procesos de selección de las mejores subbandas para el caso DWT y el algoritmo de selección de la mejores bases para el caso WPT, acompañados del criterio de separabilidad de clases *CSI*, se constituyen en herramientas igualmente eficientes de selección de características y reducción de dimensionalidad.

Finalmente, la evaluación del error de clasificación mediante los clasificadores PNN, SVM y MLP, para las mejores combinaciones de características DT-DWT y DT-WPT, fueron los siguientes: 7.1% y 5.6% para el clasificador PNN, 6.5% y 5.3% para el clasificador SVM, y 2.6% y 2.5% para el clasificador MLP. De aquí se puede concluir que la combinación de los *parámetros* DT y los *parámetros* WPT tiene mayor éxito en la clasificación de señales EMGS que la combinación de los *momentos* DT y los *valores singulares* DWT. Adicionalmente se concluye que con el clasificador MLP y un algoritmo de entrenamiento "*backpropagation*" se puede minimizar el error de clasificación alcanzando un mejor desempeño que con los clasificadores PNN y SVM.

6.1 Aportes Originales

Los aportes originales del desarrollo de este trabajo son los siguientes:

1. El proceso de representación de señales EMGS aquí propuesto ofrece un mejor desempeño a un menor costo computacional con respecto a trabajos previos, pues los errores promedio de clasificación de 2,5 % y 2.6% obtenidos para los vectores híbridos DT-DWT y DT-WPT respectivamente, representan una mejoría significativa con respecto al error promedio del 6.25%, alcanzado por Englehart sobre su base de datos (la misma que fue utilizada en este trabajo), o del 9.25%, obtenido por Hudgins únicamente con parámetros temporales.
2. Se demostró que el algoritmo del índice de separabilidad de clases *CSI* de “*Davies-Bouldin*”, evita utilizar alguna medida de probabilidad de error de pertenencia de clase e incluso la medida del error de clasificación, durante los procesos de selección de las mejores características para cada técnica de análisis (DT, DWT y WPT) y de los mejores pares de canales por paciente.
3. Se demostró que la extracción de características con la transformada DWT basta realizarse sobre los coeficientes en las subbandas D_2 y D_3 de la descomposición de nivel 3, resultando mucho menos complejo algorítmicamente y tan eficiente como el método de reducción de dimensionalidad con PCA. Aspecto muy importante a considerar al evaluar la velocidad de respuesta de un sistema de clasificación de patrones mioeléctricos.
4. Se propuso evaluar los *momentos* temporales y se comprobó que combinados con los *valores singulares* (SVD) se logra un vector de características DT-DWT con el menor error de clasificación en las demás combinaciones de métodos. Sin embargo, la combinación de los *parámetros* DT y los *parámetros* WPT produce un vector de características DT-WPT aun más representativos para las señales EMGS y con un menor error en su clasificación.
5. Se publicaron dos artículos. El primero, refleja el estado del arte del área objeto de investigación, titulado *Análisis de Señales EMG Superficiales y su*

Aplicación en Control de Prótesis de Mano. Aceptado para la revista *Avances en Sistemas e Informática*, Vol. 4 No. 1, Junio de 2007, Medellín, ISSN 1657-7663. Y el segundo, refleja algunos de los resultados obtenidos durante el desarrollo del presente trabajo, titulado *Evaluación de las Transformadas Wavelet DWT y WPT en la Extracción de Características Descriptoras sobre Señales EMGS.* Aceptado en el XII Simposio de Tratamiento de Señales, Imágenes y Visión Artificial, STSIVA 2007, publicado en las Memorias con ISBN 978-958-8252-59-9 en septiembre de 2007, Universidad del Norte Barranquilla.

6.2 Trabajos Futuros

Como trabajos futuros y de continuación directa al trabajo aquí realizado, se pueden proponer los siguientes:

1. La integración del proceso sistemático de extracción y selección de características, aquí propuesto, con un sistema de adquisición de señales EMGS, cuyo objetivo sería validar los resultados con pacientes diferentes y con los mismos o diferentes movimientos. Incluso, considerando la extensión a cuatro canales (o electrodos) para aumentar el número de movimientos diferentes a clasificar.
2. La implementación software-hardware de la solución planteada como la mejor opción en la representación de señales EMGS, así como un algoritmo de clasificación con capacidad de reaprendizaje y adaptación a la dinámica cambiante de las señales EMGS para un mismo paciente, cuyo objetivo sería aplicar los resultados en el diseño e implementación de dispositivos activos de control mioeléctrico como una prótesis de mano.

Anexo A

A.1 Análisis Espectral de Fourier

La mayoría de transformadas asumen que la señal a analizar es *estacionaria*, esto es, que las propiedades estadísticas de la señal no involucran al tiempo, siendo la transformada de Fourier FT la más utilizada:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt, \quad (\text{A.1})$$

donde la expresión $X(f)$ denota los coeficientes de la distribución de la señal en el dominio de la frecuencia. Esta representación es apropiada siempre que la señal tenga componentes de frecuencia estacionarios, pero cambios abruptos en el tiempo de una señal se verán dispersos en todo el rango de frecuencias de $X(f)$ sin ninguna resolución frecuencial. De aquí que la transformada FT no es la más indicada para señales no estacionarias. En su lugar se consideró la *transformada de Fourier de tiempo corto* STFT, introducida por Gabor, para lograr una representación simultánea de la señal en los dominios tiempo-frecuencia con cierto grado de resolución. Para esto, utiliza una función ventana $x(\tau)g^*(\tau - t)$ con la cual “mira” a la señal en el tiempo, así para una señal $x(\tau)$ la cual se supone estacionaria en el tiempo de duración de la ventana $g(\tau - t)$ centrada en t se define su transformada STFT de la siguiente manera (Englehart, 1998):

$$STFT(t, f) = \int x(\tau)g^*(\tau - t)e^{-j2\pi f\tau} d\tau, \quad (\text{A.2})$$

la cual mapea la señal de su dominio original al plano bidimensional tiempo-frecuencia (t, f) . Un aspecto importante a tener en cuenta es el compromiso entre la resolución temporal Δt y la resolución frecuencial Δf de la STFT. La resolución del análisis en el plano bidimensional tiempo-frecuencia depende exclusivamente de la ventana $g(t)$ y su respectiva transformada Fourier $G(f)$. Una medida de la resolución de frecuencia puede ser el *ancho de banda cuadrático medio*:

$$\Delta f^2 = \frac{\int f^2 |G(f)|^2 df}{\int |G(f)|^2 df}, \quad (\text{A.3})$$

y la correspondiente medida de la resolución temporal es el *ancho de tiempo cuadrático medio*:

$$\Delta t^2 = \frac{\int t^2 |g(t)|^2 dt}{\int |g(t)|^2 dt}. \quad (\text{A.4})$$

Una ventana tipo *Gausiana* brinda una buena resolución tiempo-frecuencia. Sin embargo, esta resolución siempre estará acotada inferiormente por la desigualdad conocida como principio de *incertidumbre de Heisenberg* dada por la expresión:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi}, \quad (\text{A.5})$$

lo que significa que se puede intercambiar resolución temporal por resolución frecuencial. La transformada STFT con una ventana de análisis tipo Gausiana se denomina *transformada de Gabor* (Gabor, 1946). En general, el tipo de ventana a usar se selecciona de acuerdo a la aplicación en particular y teniendo en cuenta aspectos como: localización, eficiencia computacional y supresión de dispersiones espectrales.

Una familia de *secciones* de tiempo-frecuencia, con el producto Heisenberg uniformemente acotado se puede representar por celdas de información de áreas iguales. Una *base* cualquiera de tales *secciones*, corresponde a cubrir el plano tiempo-frecuencia con rectángulos uniformes, mientras que una *base ortonormal* puede ilustrarse como un cubrimiento del mismo plano con rectángulos *disjuntos*. La selección de las bases será la que determine la relación de las celdas tiempo-frecuencia. Así por ejemplo, la *base estándar* (o base Dirac) tiene una localización óptima en tiempo pero no en frecuencia, mientras que las *bases Fourier* (base sinusoidal) tiene óptima localización en frecuencia mas no en tiempo, tal como se ilustra en la figura A.1.

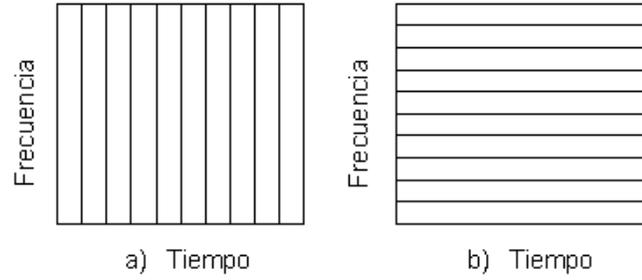


Figura A.1. a) Base estándar (impulsos Dirac); b) Base de Fourier (sinusoides)

En la práctica para analizar una señal física la base más apropiada será la que mejor compromiso ofrezca en la relación de resoluciones tiempo-frecuencia. Y en este sentido la transformada continua STFT produce una representación en el plano tiempo-frecuencia altamente redundante, por lo que se prefiere utilizar su versión discretizada denominada *transformada de Fourier discreta* DFT.

Considérese una señal $x(t)$ muestreada a una frecuencia f_s (período de muestreo T_s) y la ventana $g(t) = g(iT_s)$ con longitud o duración L instantes de muestreo; su transformada DFT en el instante $t = iT_s$ se expresa como

$$X(mF) \equiv X[m] = \sum_{i=1}^{L-1} x[i] e^{-j2\pi(mF)(iT_s)}, \quad (\text{A.6})$$

donde $F = \frac{1}{LT_s}$ es el tamaño de paso en el muestreo de frecuencia. Y la transformada STFT discreta consiste de una serie de DFTs, indexadas con respecto a T_s y F así:

$$STFT[k, m] \equiv STFT[kT_s, mF] = \sum_{i=1}^{L-1} x[i] g[i - k] e^{-j2\pi mi/L}. \quad (\text{A.7})$$

En general, las muestras de tiempo están espaciadas K muestras. En tiempo continuo el tamaño de paso del muestreo temporal es $T = K \cdot T_s$. De esta manera, la transformada STFT puede ser indexada por sus pasos de muestreo tiempo-frecuencia T y F mediante

$$STFT[n, m] \equiv STFT[nT, mF] = \sum_{i=1}^{L-1} x(iT_s) g(iT_s - nT) e^{-j2\pi mi/L}. \quad (\text{A.8})$$

Con $K = 1$, la transformada STFT se evalúa sobre todas las muestras de tiempo y las ventanas de análisis sucesivas se traslapan entre sí, mientras que con $K = L$,

dichas ventanas no se traslapan. En la figura A.2 se muestra la segmentación del plano tiempo-frecuencia resultante del análisis mediante la transformada STFT. Los pasos de muestreo impuestos en el plano tiempo-frecuencia produce una partición tal que cada celda tiene un ancho de tiempo T y una altura de frecuencia F . Así, un coeficiente particular de la transformada contendrá la magnitud de la representación de la señal dentro de los límites de la celda en el plano tiempo-frecuencia.

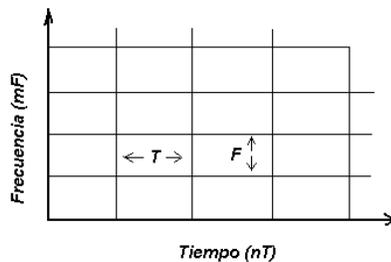


Figura A.2. Resolución de análisis tiempo-frecuencia

Si la ventana $g(t)$ tiene buena localización en tiempo y frecuencia la STFT puede indicar la densidad de energía de la señal simultáneamente en la posición (nT, mF) del plano bidimensional. El área de cada celda será $TF = KT_s \cdot 1/LT_s = K/L$, por lo que si $TF = 1$ se dice que la representación en tiempo-frecuencia es críticamente muestreada y el número de coeficientes es igual al número de muestras en la señal (asumiendo muestreo de la señal a la tasa Nyquist), condición óptima para garantizar la reconstrucción perfecta de la señal. En caso de tener $TF < 1$, la grilla resulta sobremuestreada y la representación será altamente redundante. Por el contrario, si $TF > 1$, la grilla resulta submuestreada y muy pobre en información como para reconstruir la señal. En aplicaciones de clasificación, debe buscarse el mejor compromiso entre la resolución temporal y la frecuencial, esto permitirá una *representación de la señal* con una una partición del plano bidimensional capaz de capturar la estructura tiempo-frecuencia de la señal y el número de coeficientes será el adecuado.

Para ilustrar esto se toma una sección de longitud 256 muestras de la señal *chirp lineal* cuya frecuencia aumenta linealmente con el tiempo de 0 a 500 Hz, como se aprecia en la figura A.3. Esta forma de onda permitirá comprobar la capacidad de localización tiempo-frecuencia de la transformada STFT.

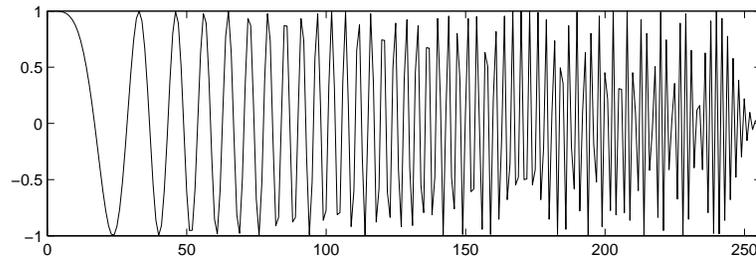


Figura A.3. Forma de onda de la señal *chirp lineal*

En la figura A.4 se pueden diferenciar dos condiciones de representación tiempo-frecuencia mediante la transformada STFT de esta señal. En ambos casos se utilizan como funciones ventana Gausianas de longitud $L = 32$. En la figura 3.5.a, con $K = 1$, se observa una grilla con demasiadas celdas traslapadas en el plano temporal debido al producto $TF = K/L = 1/32 < 1$, presentándose sobremuestreo y por ende un gran número de coeficiente de representación redundantes entre sí (condición no deseada para efectos de clasificación). En la figura 3.5.b, con $K = 32$, se observa una grilla con celdas sin traslape como resultado del producto $TF = K/L = 32/32 = 1$, y son suficientes para capturar la variación lineal de la frecuencia de la señal con el tiempo.

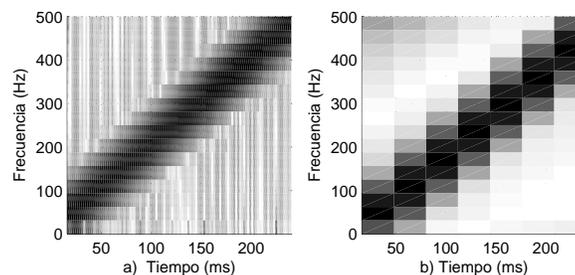


Figura A.4. Representación STFT de la señal *chirp lineal*. En a) con $L = 32$ y $K = 1$; b) $L = 32$ y $K = 32$

Sin embargo y a pesar de la madurez teórica de la transformada STFT y sus algoritmos rápidos de implementación, no deja de producir una segmentación del plano tiempo-frecuencia uniforme y cuadriculada; es decir, la resolución de análisis es uniforme en todo el rango temporal y de frecuencias de la señal, con una relación fija en el producto de la resolución temporal Δt y la resolución frecuencial Δf . Pero la distribución de energía de las señales físicas no siempre es uniforme, ni mantiene

una relación fija en el plano tiempo-frecuencia.

A.2 Transformada Wavelet

De igual manera como la transformada de Fourier descompone una señal en términos de funciones sinusoidales, la transformada wavelet descompone la señal en términos de funciones denominadas *wavelets*. El término *wavelet* se define como una “onda pequeña” o función localizable en el tiempo. Estas funciones forman “**bases**” de wavelets $\Psi = \{\psi_\lambda | \lambda \in \Lambda\}$ con tres características fundamentales a saber (Englehart, 1998):

i) Las wavelets son bloques de construcción para funciones en general.

Cualquier función $f \in L^2(R)$ se puede expresar como una serie infinita de wavelets, de la forma:

$$f = \sum_{\lambda} c_{\lambda} \psi_{\lambda} \quad (\text{A.9})$$

ii) Las wavelets tienen localización tiempo-frecuencia. Localización en tiempo, significa que la mayor parte de la energía de la función wavelet se encuentra en un intervalo finito. Si la función es idénticamente igual a cero fuera de dicho intervalo, se denomina *de soporte compacto*. Localización en frecuencia significa que la transformada de Fourier de la función *wavelet* resulta de banda limitada. El principio de incertidumbre de Heisenberg impone una cota inferior al producto de las variaciones de tiempo y frecuencia. Mientras la transformada de Fourier STFT da lugar a una relación fija en la resolución tiempo-frecuencia, con la transformada wavelet se produce una relación variable de esta relación, obteniéndose una buena localización en frecuencia para el rango de bajas frecuencias (ventanas de mayor duración), y buena localización en tiempo en el rango de altas frecuencias (ventanas de menor duración). Esto produce una segmentación o *enrejado* del plano tiempo-frecuencia mas acorde al comportamiento de la mayoría de señales físicas, especialmente para aquellas de naturaleza transitoria.

iii) Las wavelets tienen transformada rápida. Para muchas aplicaciones, es importante que la transformada sea fácil de implementar, particularmente en

procesamiento de señales en tiempo real. La *transformada rápida wavelet* se obtiene mediante el *análisis multiresolución*; un algoritmo piramidal con orígenes en procesamiento de imágenes adaptado para el análisis wavelet por Stephane Mallat e Yves Meyer en 1986 (Mackenzie D., 2001). Esta transformada usa una serie de filtros lineales paso-bajas y paso-altas para descomponer la señal en componentes de frecuencia de bajas y altas frecuencias. El algoritmo también combina las operaciones de *submuestreo*, con lo que se consigue un decimado de la señal (reduciendo su longitud a la mitad) en cada estado de la descomposición como se verá mas adelante. Esta operación contribuye en la velocidad del algoritmo, por que reduce las operaciones de procesamiento geoméricamente con cada iteración. Así en j iteraciones, el número de muestras manipuladas decrece en 2^j . Esto hace que el algoritmo sea más eficiente. Una transformada wavelet sobre N puntos, tiene una complejidad del orden de $O(N)$, mientras que la transformada de Fourier es del orden de $O(N \log N)$ (Englehart, 1998). Así, la complejidad de la transformada wavelet será CN (C constante que depende de la *wavelet* seleccionada). Las wavelets que constan de pocos coeficientes mejoran la eficiencia de los algoritmos (valores pequeños de C), como es el caso de las funciones *wavelets* desarrolladas por Ingrid Daubechies (Daubechies , 1992).

La *transformada wavelet continua* (CWT), la *transformada wavelet discreta* (DWT) y la *transformada de paquetes wavelet* (WPT) constituyen nuevas herramientas de análisis de señales, que proporcionan información de la señal en los dominios tiempo-frecuencia de manera simultanea. A continuación se describe cada una de estas.

A.2.1 Transformada Wavelet Continua (CWT)

La *transformada wavelet continua* CWT se define como:

$$CWT_x(a, b) = \int x(t)\psi_{a,b}^*(t)dt, \quad (\text{A.10})$$

con

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \quad a, b \in R; a \neq 0 \quad (\text{A.11})$$

Con $\Psi(t)$ como *wavelet madre*, el conjunto $\{\psi_{a,b}(t)\}$ forma una base ortonormal en $L^2(R)$. Esto implica que a partir de este conjunto se puede generar cualquier función en $L^2(R)$. En esta transformada, la función wavelet $\psi(t)$ actúa como *ventana*, donde la variable, a , determina su ancho y la variable, b , determina su ubicación en el dominio del tiempo. De acuerdo a esto, se identifica a la variable, a , como *variable de escala* con la cual se comprime o dilata la función $\psi(t)$ y establece el grado de resolución con el cual se analiza la señal $x(t)$, la variable, b , se identifica como *variable de traslación* y determina la posición de la función *ventana* sobre la señal analizada (Daubechies , 1992). Así, con la CWT se logra expresar una señal continua en el tiempo $x(t)$ como una expansión en términos de versiones escaladas y trasladadas de la función $\psi(t)$, cuyos coeficientes son proporcionales al producto interno entre la señal y las versiones respectivas de $\psi(t)$.

Algunos ejemplos de funciones wavelets se muestran en la figura A.5.

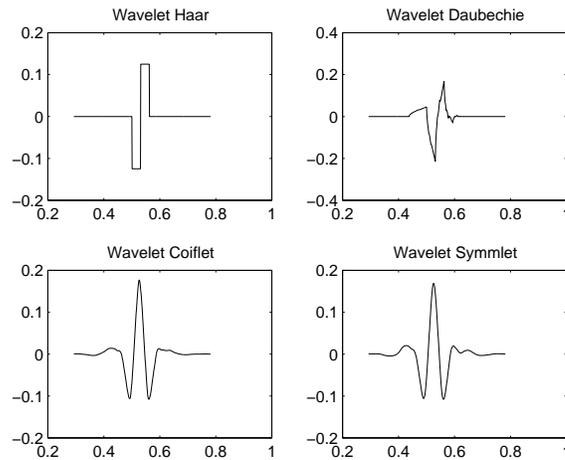


Figura A.5. Funciones wavelets más comunes

De acuerdo a la definición de la CWT puede concluirse que más que una representación tiempo-frecuencia, es una representación tiempo-escala. Esto en razón de que las bajas frecuencias de una señal se analizan con escalas altas, obteniéndose una buena resolución en frecuencia, mientras que los detalles de la señal, correspondientes a las altas frecuencias, se analizan con bajas escalas, obteniéndose una buena resolución en tiempo. En este sentido, se dice que la escala y la frecuencia tienen una relación inversa.

De la limitación en banda de las funciones wavelets se puede decir que estas

operan como filtros pasa-banda y se cumple que:

$$c = \frac{\Delta f}{f}, \quad (\text{A.12})$$

donde Δf es su ancho de banda, f es la frecuencia central de la banda y c es una constante. Desde este punto de vista, las funciones wavelets pueden ser vistas como bancos de filtros con una banda de paso relativamente constante (un análisis de Q-constante) (Englehart, 1998).

Así mismo, recordemos que en el análisis con la STFT se tiene una resolución temporal Δt fija y una resolución en frecuencia Δf restringida a la *desigualdad de Heisenberg*, $\Delta t \cdot \Delta f \geq \frac{1}{4\pi}$, lo que produce una partición regular del plano tiempo-frecuencia. En el análisis wavelets, con una resolución temporal $\Delta t'$ y una resolución de frecuencia $\Delta f'$ de la función wavelet $\psi(t)$, se puede demostrar que para la versión escalada y trasladada $\psi_{a,b}(t)$, las resoluciones temporal y de frecuencia son respectivamente:

$$\Delta t' = |a| \Delta t \quad y \quad \Delta f' = \frac{1}{|a|} \Delta f. \quad (\text{A.13})$$

De manera que la partición del plano tiempo-frecuencia no solo resulta con celdas no uniformes sino que dependen directamente de la escala a . Así, a bajas escalas (altas frecuencias) las celdas son estrechas y altas, mientras que para escalas altas (bajas frecuencias), las celdas son anchas y bajas (ver: figura 4, Capítulo 2). Aun así, la localización tiempo frecuencia de cada función wavelet escalada y trasladada estará restringida por la desigualdad de Heisenberg:

$$\Delta t' \cdot \Delta f' = \Delta t \cdot \Delta f \geq \frac{1}{4\pi} \quad (\text{A.14})$$

Desde un punto de vista intuitivo, la CWT consiste en calcular un índice de semejanza entre la señal que está siendo analizada y la wavelet seleccionada (De Castro, 2002), tal como se muestra en la figura A.6. El proceso de cálculo de la CWT consiste en seleccionar una wavelet madre con un valor inicial para las variables a y b , compararla con un intervalo inicial de la señal y determinar el coeficiente de correlación $c_{a,b}$, entre la wavelet y la sección de la señal bajo análisis. Entre mayor sea el valor del coeficiente mayor será la similitud, de allí que los resultados del análisis dependen mucho de la función wavelet elegida para el análisis. Y mediante variaciones de la variable de traslación b se recorre sucesivamente toda la señal de

análisis para determinar todos los coeficientes del nivel de escala a . Este proceso se repite para cada valor de escala a deseado, según la profundidad de análisis que se desee.

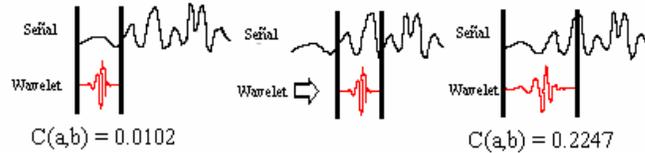


Figura A.6. Modo de operación de la transformada wavelet continua CWT (De Castro , 2002)

Propiedades de las wavelets

Las propiedades más importantes de las wavelets son las condiciones de *admisibilidad* y de *regularidad*, y son las propiedades que le dan a una wavelet su nombre (Alfred M. , 1999). Se puede mostrar que una función $\psi(t)$ cuadrado integrable satisface la condición de *admisibilidad* si se cumple que:

$$\int \frac{|\Psi(w)|^2}{|w|} dw < +\infty, \quad (\text{A.15})$$

por lo que puede ser utilizada para analizar y sintetizar una señal sin pérdida de información. Aquí $\Psi(w)$ representa la transformada de Fourier de $\psi(t)$. La condición de *admisibilidad* implica también que la transformada de Fourier de $\psi(t)$ es nula a frecuencia cero, esto es

$$|\Psi(w)|_{w=0}^2 = 0, \quad (\text{A.16})$$

lo que significa que la wavelet tiene un espectro similar a una banda de paso, característica muy importante para la implementación de la transformada wavelet de manera rápida y eficiente. Que el espectro sea nulo a frecuencia cero significa también que el valor promedio temporal de la wavelet es cero, es decir:

$$\int \psi(t) dt = 0, \quad (\text{A.17})$$

por tanto $\psi(t)$ puede tratarse como una onda oscilatoria.

La condición de *regularidad* establece que la función wavelet debe tener alguna suavidad y concentración tanto en el tiempo como en la frecuencia. Es un concepto

un tanto complejo que puede intentarse explicar desde el concepto de *momentos de desvanecimiento*. Si se expande la transformada wavelet dada por la ecuación (3.2), en series de Taylor hasta el orden n , haciendo $b = 0$ por simplicidad, y para $t = 0$, se llega a

$$CWT_x(a, b) = \frac{1}{\sqrt{a}} \left[\sum_{p=0}^n x^{(p)}(0) \int \frac{t^p}{p!} \psi \left(\frac{t}{a} \right) dt + O(n+1) \right], \quad (\text{A.18})$$

donde $x^{(p)}$ representa la derivada de orden n de x y $O(n+1)$ representa el resto de la expansión.

Ahora si se define los momentos M_p de una wavelet por:

$$M_p = \int t^p \psi(t) dt, \quad (\text{A.19})$$

entonces se puede reescribir la expansión en series de Taylor de la forma:

$$CWT_x(a, b) = \frac{1}{\sqrt{a}} \left[x(0)M_0a + \frac{x^{(1)}(0)}{1!}M_1a^2 + \dots + \frac{x^{(n)}(0)}{n!}M_na^{n+1} + O(a^{n+2}) \right] \quad (\text{A.20})$$

De la condición de *admisibilidad* se tiene que $M_0 = 0$ (momento de orden nulo), por tanto el primer término de la serie en la ecuación (3.12) se elimina. Y si se logra hacer que los momentos restantes hasta el de orden n también sean nulos, entonces los coeficientes de la transformada wavelets $CWT_x(a, b)$ decaerán tan rápido como $a^{(n+2)}$ para una señal suave $x(t)$. Esto se conoce en la literatura como momentos de desvanecimiento u orden de aproximación. De esta manera, si una wavelet tiene N momentos de desvanecimiento, entonces el orden de aproximación de la transformada wavelet será también N . Los momentos no tienen que ser exactamente iguales a cero, un valor suficientemente pequeño también será aceptado. De hecho, investigaciones experimentales sugieren que el número de momentos de desvanecimiento requeridos depende principalmente de la aplicación.

En resumen, la condición de *admisibilidad* determina la forma de onda y la condición de *regularidad* (junto con los momentos de desvanecimiento) determina la velocidad con que La onda decae, y las dos condiciones juntas determinan a la wavelet.

A.2.2 Transformada Wavelet Discreta (DWT)

La transformada CWT presenta ciertas desventajas que la hacen difícil de usar directamente. Entre ellas está la redundancia en sus coeficientes wavelets por la variación continua de las variables de escala y traslación, pero aún sin redundancia resultan infinitas wavelets de la transformada, lo que obliga a restringirla a un número manejable. Por otra parte las funciones wavelets escaladas no conforman una base ortogonal y un inconveniente adicional es que para la mayoría de las funciones la transformada wavelet no tiene solución analítica y debe calcularse numéricamente (Saito , 1994).

En este sentido se han desarrollado algoritmos de análisis de señales de tiempo discreto mediante la *transformada wavelet discreta* DWT, para lo cual se ha discretizado convenientemente los valores de las variables de *escala* y *traslación*. De esta manera, haciendo $a = 2^j$ y $b = n \cdot 2^j$ (j y $n \in \mathbb{Z}$), la expresión de $\psi(t)$ en (A.11) se transforma en:

$$\{\psi_{j,n}(t)\}_{j,n \in \mathbb{Z}} \doteq \{2^{-j/2}\psi(2^{-j}t - n)\} \quad (\text{A.21})$$

A este conjunto de funciones se le conoce como la *versión diádica discreta* de la función *wavelet* o *conjunto de expansión wavelet* y forma una *base ortonormal* de $L^2(\mathbb{R})$, donde el factor $2^{-j/2}$ se denomina *constante de normalización* (Faundez , 2002). Lo anterior significa que las bases wavelets permiten la descomposición ortogonal de cualquier función en $L^2(\mathbb{R})$:

$$L^2(\mathbb{R}) = \oplus_j \Omega_{j,1}, \quad (\text{A.22})$$

donde $\Omega_{j,1}$ es el subespacio generado por $\{\psi_{j,n}\}_{n \in \mathbb{Z}}$. Una *función wavelet* $\psi(t)$ está siempre asociada a una *función escala* $\phi(t)$, comúnmente llamada *wavelet padre*. Y al igual que la función wavelet madre, la función escala

$$\{\phi_{j,n}(t)\}_{j,n \in \mathbb{Z}} \doteq \{2^{-j/2}\phi(2^{-j}t - n)\}, \quad (\text{A.23})$$

genera una base ortonormal de $L^2(\mathbb{R})$. Lo que significa que las *bases escala* permiten también la descomposición ortogonal de cualquier función en $L^2(\mathbb{R})$, esto es:

$$L^2(\mathbb{R}) = \oplus_j \Omega_{j,0}, \quad (\text{A.24})$$

donde $\Omega_{j,0}$ es el subespacio generado por $\{\phi_{j,n}\}_{n \in Z}$.

De igual manera se puede decir que una función $x(t)$ en $L^2(R)$ puede aproximarse en términos de estas dos funciones según la expresión:

$$x(t) = \sum_n \sum_j c_{j,n} \phi(t) + \sum_n \sum_j d_{j,n} \psi(t); \quad j, n \in Z \quad (\text{A.25})$$

La función *escala* $\phi(t)$ induce una cadena de subespacios anidados de la forma:

$$\Omega_{J,0} \subset \Omega_{J-1,0} \subset \dots \subset \Omega_{1,0} \subset \Omega_{0,0} \quad (\text{A.26})$$

La naturaleza de la función *escala* es que una *proyección* de la señal original $x(t)$ en el espacio $\Omega_{j,0}$ es una operación de *paso-bajas*, es decir una *aproximación* de $x(t)$ a escala $a = 2^j$ de la forma:

$$A_j[n] = \langle x(t), \phi_{j,n}(t) \rangle \quad (\text{A.27})$$

Se define $\Omega_{0,0}$ (escala $a = 2^0$) como el espacio de la señal original $x(t)$; esto es, $A_0[n] = x[n] \equiv x(nT_s)$. Según esto, la ecuación (3.18) puede verse como una secuencia de sucesivas aproximaciones burdas de $x(t)$ en el rango de escalas de 0 a J . Los subespacios anidados generados por las funciones wavelet y escala están relacionados así:

$$\Omega_{j,0} = \Omega_{j+1,0} \oplus \Omega_{j+1,1} \quad \text{para } j = 0, 1, \dots, J \quad (\text{A.28})$$

Esto significa que el espacio $\Omega_{j+1,1}$ contiene el *detalle* necesario para ir de una *aproximación* $\Omega_{j+1,0}$ a un nivel de *aproximación* más fino $\Omega_{j,0}$. La componente de *detalle* de $x(t)$ a escala $a = 2^j$ es:

$$D_j[n] = \langle x(t), \psi_{j,n}(t) \rangle, \quad (\text{A.29})$$

que corresponde a una operación *pasa-banda*. Por tanto, la transformada DWT puede verse como una forma de representar el espacio $\Omega_{0,0}$ como una suma directa de subespacios ortogonales como se aprecia en la figura A.7 (para $J=3$). En forma general se expresa mediante la relación:

$$\Omega_{0,0} = \left(\bigoplus_{j=1}^J \Omega_{j,1} \right) + \Omega_{J,0} \quad (\text{A.30})$$

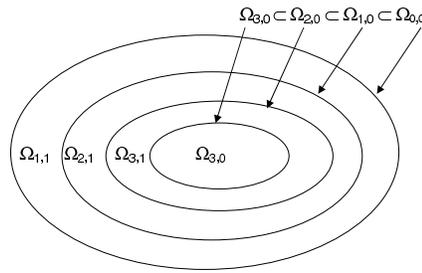


Figura A.7. Anidamiento de subespacios wavelets $\Omega_{j,1}$ y escalas $\Omega_{j,0}$ para $J = 3$

Lo que significa que la transformada DWT analiza una señal por descomposición en sucesivas señales de aproximaciones $A_j[n] \in \Omega_{j,0}$ y detalles $D_j[n] \in \Omega_{j,1}$ (Englehart, 1998). De esta manera, un análisis completo de una señal original $x[n]$ queda representado por las secuencias de coeficientes wavelets:

$$\{D_1[n], D_2[n], \dots, D_J[n], A_J[n]\} \quad (\text{A.31})$$

Así, los coeficientes resultantes de la transformada DWT a un nivel J de descomposición consisten en: J niveles de coeficientes de *detalles* y un nivel (*Jésimo*) de coeficientes de *aproximación*.

Transformada DWT y Bancos de Filtros

Desde el punto de vista del procesamiento digital de señales, la manera práctica y en particular para aplicaciones como la planteada en este trabajo, la transformada DWT se puede implementar a partir de los llamados *bancos de filtros*. Su proceso de descomposición se describe a continuación (Saito, 1994).

La descomposición $\Omega_{j,0} \rightarrow \Omega_{j+1,0}$ será representada por el operador H y la descomposición $\Omega_{j,0} \rightarrow \Omega_{j+1,1}$ será representada por el operador G . Suponiendo una señal $x = x[n] \in \Omega_{0,0}$ de N muestras, una descomposición de la señal $x[n]$ de primer nivel se muestra en la figura A.8.

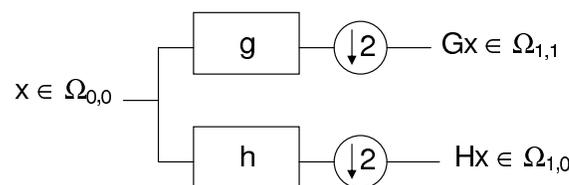


Figura A.8. Descomposición DWT a un nivel ($J=1$)

Esta descomposición consiste en someter la señal original a los operadores G y H separadamente; el primero consiste en un filtraje paso-altas a través de $g = \{g[n], n = 0, 1, \dots, L - 1\}$ y el segundo en un filtraje paso-bajas a través de $h = \{h[n], n = 0, 1, \dots, L - 1\}$, donde el operador “ $\downarrow 2$ ” implica una decimación por 2.

Toda esta operación se puede expresar como:

$$(Hx)_n = \sum_{k=0}^{L-1} h[k]x[2n - k] \quad y \quad (Gx)_n = \sum_{k=0}^{L-1} g[k]x[2n - k], \quad n = 0, 1, \dots, N - 1 \quad (\text{A.32})$$

Esto significa que las *aproximaciones* de la señal se obtienen mediante filtros *paso-bajas* y los *detalles* de la señal mediante filtros *paso-altas*. Se puede observar también que si la señal a analizar consiste en un vector de N muestras de longitud, las secuencias Gx y Hx son vectores de longitud $N/2$. Una segunda descomposición DWT de la señal consistirá en aplicar el mismo proceso de filtraje sobre los coeficientes de la banda de bajas frecuencias Hx para obtener las nuevas secuencias H^2x y GHx de longitud $N/4$ cada una. Este proceso puede iterarse hasta un nivel $J \leq \log_2 N$, obteniéndose un vector de secuencias de la descomposición DWT de la forma $\{Gx, GHx, GH^2, \dots, GH^{J-1}, H^Jx\}$. En la figura A.9 se muestra el proceso de descomposición para el caso de una descomposición a nivel tres ($J = 3$).

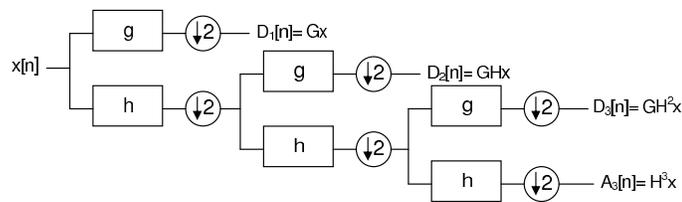


Figura A.9. Descomposición DWT a tres niveles ($J=3$)

El análisis de una señal mediante la transformada DWT particiona su espectro de frecuencias diádicamente, actuando siempre sobre la región de bajas frecuencias en cada nivel de descomposición como se muestra en la figura A.10.

Los operadores G y H son llamados filtros de *reconstrucción perfecta* o *filtros ventana en cuadratura* (QMF - Quadrature Mirror Filter), y satisfacen la siguiente condición de ortogonalidad:

$$HG^* = GH^* = 0 \quad y \quad HH^* = GG^* = I, \quad (\text{A.33})$$

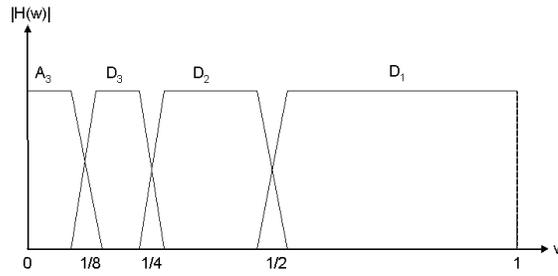


Figura A.10. Partición diádica del espectro de frecuencias con DWT

donde “ $*$ ” es el operador conjugado e I es la matriz identidad. Estas condiciones imponen ciertas restricciones en los coeficientes de los filtros h y g . Sean m_0 y m_1 funciones periódicas acotadas, definidas por:

$$m_0(\xi) = \sum_{n=0}^{L-1} h[n]e^{jn\xi} \quad y \quad m_1(\xi) = \sum_{n=0}^{L-1} g[n]e^{jn\xi} \quad (\text{A.34})$$

Daubechies (1988), demostró que H y G son QMF's si y solo si la matriz

$$\begin{bmatrix} m_0(\xi) & m_0(\xi + \pi) \\ m_1(\xi) & m_1(\xi + \pi) \end{bmatrix} \quad (\text{A.35})$$

es unitaria para todo $\xi \in R$. Otros criterios de diseño y selección de los coeficientes del filtro pasa-bajas h (concernientes a regularidad, simetría, etc.) pueden encontrarse en (Daubechies , 1992). Una vez fijados los coeficientes del filtro h , los coeficientes del filtro g se obtienen haciendo $g[n] = (-1)^n h[L - 1 - n]$ para obtener filtros QMF's.

Los filtros QMF's h y g están íntimamente relacionados con las funciones *wavelet* y *escala* mediante las relaciones (Meyer , 1993):

$$h[n] = \langle \phi(t), \sqrt{2}\phi(2t - n) \rangle \quad y \quad g[n] = \langle \psi(t), \sqrt{2}\psi(2t - n) \rangle \quad (\text{A.36})$$

Esto significa que para obtener en forma rápida la transformación wavelet DWT, basta con pasar la señal a través de este par de filtros, cuyos coeficientes representan a las respectivas funciones *wavelet* y *escala* asociadas. Los filtros QMF's tienen la propiedad de permitir una reconstrucción perfecta de la señal original a partir de los coeficientes wavelet. En el proceso de reconstrucción o síntesis, es necesario un

sobremuestreo (inserción de ceros entre muestras) y un filtraje a través del mismo par de filtros.

Aplicación de la transformada DWT

A manera de ejemplo para ilustrar la aplicación de la transformada DWT se utiliza una sección de longitud $N = 256$ muestras la señal *chirp lineal*¹, muestreada a una tasa de 1000 muestras por segundo durante 256 ms. Esta señal se presta para mostrar las propiedades del análisis DWT en el plano bidimensional tiempo-frecuencia, pues se trata de una señal no estacionaria cuya frecuencia aumenta linealmente con el tiempo.

La función *wavelet* de análisis, *db3*, pertenece a la familia Daubechie (de orden 3). En la figura A.11.a se muestra la función *wavelet madre db3* y en la figura A.11.b su respectiva función de escala o *wavelet padre db3*.

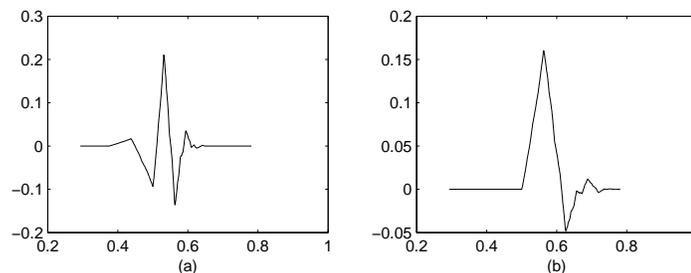


Figura A.11. a) Wavelet madre db3, b) wavelet padre db3

La descomposición DWT para una sección de $N = 256$ muestras se puede llevar hasta un nivel máximo $J = 8$, pudiéndose detener en cualquier nivel $j \leq \log_2 N$ según las necesidades del análisis y de la aplicación en particular. Para este ejemplo, la descomposición se realiza a nivel tres ($j = 3$) obteniéndose las secuencias de coeficientes $\{A3, D3, D2, D1\}$.

En la figura A.12.a, se muestra como el análisis DWT detecta las diferentes componentes de frecuencia en la señal original con aceptable localización en el tiempo según la magnitud de los coeficientes en las bandas de partición del plano bidimensional tiempo-frecuencia normalizado. En este análisis se puede observar que en los primeros instantes, de 0 a 0.125 del tiempo de la señal solo se registran coeficientes wavelet en la banda mas baja de frecuencias, esto es de 0 a 0.125 de $f_s/2$, correspondiente a la banda de aproximaciones $A3$; entre los instantes 0.125 a 0.25

¹La señal *chirp lineal* se encuentra disponible en el toolbox de MatLab

del tiempo de la señal, se registran coeficientes wavelet en la banda de frecuencias de 0.125 a 0.25 de $f_s/2$, correspondiente a la banda de detalles $D3$; entre los instantes 0.25 a 0.5 del tiempo de la señal se registran coeficientes wavelet en la banda de frecuencias de 0.25 a 0.5 de $f_s/2$, correspondiente a la banda de detalles $D2$ y entre los instantes 0.5 a 1.0 del tiempo de la señal se registran coeficientes wavelet en la banda de frecuencias de 0.5 a 1.0 de $f_s/2$, correspondiente a la banda de detalles $D1$.

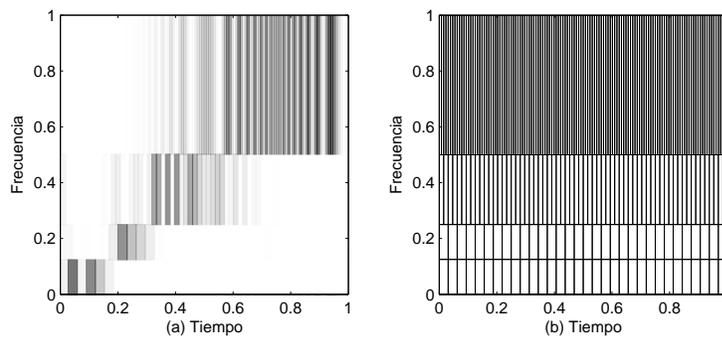


Figura A.12. Análisis DWT en el plano bidimensional tiempo-frecuencia

En la figura A.12.b, se muestra la forma de partición o “grilla” del plano bidimensional tiempo-frecuencia normalizado, que muestra la resolución del análisis DWT de acuerdo al principio de Heisenberg explicado en el capítulo 2, donde se puede apreciar una resolución de frecuencias mas fino hacia las bajas frecuencias que hacia las altas frecuencias.

Esta forma de partición del plano bidimensional tiempo-frecuencia difiere de la partición lograda con la transformada STFT, precisamente por que la transformada DWT tiene resolución de análisis logarítmico y no lineal. Además, la selección de una escala de muestreo diádica ($a = 2^j, b = k \cdot 2^j$) garantiza que la representación de la señal con DWT resulte *críticamente muestreada*, por lo que se encontrarán N celdas discretas en la partición del plano bidimensional tiempo-frecuencia (Englehart, 1998). Cada celda de información, tendrá una resolución temporal $\Delta t = 2^j \cdot T_s$, con T_s *período de muestreo* y una resolución de escala $\Delta a = 2^{J-j}$, que por su relación inversa con la frecuencia, la celda tendrá una resolución de frecuencia $\Delta f = 2^{-j}(f_s/2)$, con f_s como *frecuencia de muestreo*².

²Tasa de Nyquist, f_s , igual al doble de la frecuencia máxima de la señal.

En el ejemplo se tiene entonces $N = 256$ celdas de información en todo el plano bidimensional tiempo-frecuencia. Con $J = 3$, su distribución se analiza de la siguiente manera: para el primer nivel de descomposición ($j = 1$) la resolución temporal es $\Delta t = 2^1 \cdot T_s = 2T_s = 2 \cdot 10^{-3} \text{seg.}$, lo que equivale a tener en una sección de 256 msec. un total de 128 celdas (parte superior del plano, figura A.12.b), con resoluciones de escala $\Delta a = 2^{3-1} = 4$ y de frecuencia $\Delta f = 2^{-1} f_s/2 = 0.5(f_s/2)$ Hz. Para el segundo nivel ($j = 2$), la resolución temporal será $\Delta t = 4 \cdot 10^{-3} \text{seg.}$, o 64 celdas (parte intermedia del plano, figura A.12.b), con resolución de escala $\Delta a = 2$ y de frecuencia $\Delta f = 0.25(f_s/2)$ Hz. Y en el tercer nivel ($j = 3$), la resolución temporal será $\Delta t = 8 \cdot 10^{-3} \text{seg.}$, o 32 celdas (partes inferiores del plano, figura A.12.b), con resolución de escala $\Delta a = 1$ y de frecuencia $\Delta f = 0.125(f_s/2)$ Hz. Este tercer nivel se aplica tanto a la banda de aproximaciones $A3$ como a la de detalles $D3$.

Se puede notar también que el producto $\Delta t \cdot \Delta f = (2^j \cdot T_s) (2^{-j} f_s/2) = 1/2$ resulta ser mayor que la cota mínima establecida por la *desigualdad de Heisenberg*; por tanto el análisis wavelet aunque no se acerca al límite de *Heisenberg*, satisface la desigualdad $\Delta t \cdot \Delta f \geq 1/4\pi$. Esto evita traslapes entre celdas adyacentes en la representación tiempo-frecuencia de la señal, además de una “grilla” críticamente muestreada que asegura una perfecta reconstrucción de la señal original (Englehart, 1998).

Selección de la función wavelet

La selección de una wavelet madre depende mucho de la señal a analizar y de los propósitos del procesamiento de la señal. Recordando que la transformada wavelet produce secuencias de coeficientes $\{D_1[n], D_2[n], \dots, D_J[n], A_J[n]\}$, los cuales actúan sobre los *vectores base* en $L^2(\mathbb{R})$:

$$\left\{ \{\psi_{1,n}\}_{n=0}^{2^{n_0-1}-1}, \{\psi_{2,n}\}_{n=0}^{2^{n_0-2}-1}, \dots, \{\psi_{J,n}\}_{n=0}^{2^{n_0-J}-1}, \{\phi_{J,n}\}_{n=0}^{2^{n_0-J}-1} \right\}, \quad (\text{A.37})$$

cada uno de los cuales genera un subespacio. El conjunto de dichos subespacios mutuamente ortonormales es:

$$\{\Omega_{1,1}, \Omega_{2,1}, \dots, \Omega_{J,1}, \Omega_{J,0}\}, \quad (\text{A.38})$$

donde cada subespacio $\Omega_{j,k}$ ³ es generado por 2^{n_0-j} vectores base. Por facilidad de escritura, se utiliza la siguiente notación para representar los conjuntos de

³ $k = 1$ para subespacios de detalles y $k = 0$ para subespacios de aproximaciones.

vectores base wavelets:

$$\mathbf{w}_{j,k,n} = \{ w_{j,k,n} [i] \}_{i=0}^{N-1} \in \Omega_{j,k}, \quad (\text{A.39})$$

para $k = 0, 1$ en la escala j y posición n , con $0 \leq n \leq 2^{n_0-j}$, $n \in \mathbb{Z}$.

Estos vectores base tienen las siguientes propiedades (Saito , 1994):

i) **Momentos de desvanecimiento.** Según la expresión

$$\sum_{i=0}^{L-1} i^m w_{j,k,n} [i] = 0 \quad \text{para } m = 0, 1, \dots, M-1, \quad (\text{A.40})$$

corresponde al momento de desvanecimiento mas alto que tenga la base y se relaciona con la suavidad de la onda wavelet. En las bases originalmente propuestas por Daubechies (1988), la longitud de los filtros QMF está dada por $L = 2M$, por ejemplo la wavelet Daubechies db4 tiene una longitud $L = 8$, por tanto tiene $M = 4$ momentos de desvanecimiento. Aunque existen otras posibilidades como la familia de wavelets *Coiflets* con $L = 3M$, que son menos simétricas que las wavelets Daubechies.

ii) **Regularidad.** La condición

$$|w_{j,k,n} [i+1] - w_{j,k,n} [i]| \leq c2^{-j\rho}, \quad (\text{A.41})$$

donde $c > 0$ constante y $\rho > 0$ se conoce como *regularidad* de la wavelet. Entre más grande sea el valor de ρ , más regular será la función wavelet. Esta propiedad es muy importante cuando se busca altas tasas de compresión de señales, ya que una baja regularidad podría verse como dispersiones fractales en las señales o imágenes reconstruidas.

iii) **Soporte compacto.** Esta propiedad impone la condición

$$w_{j,k,n} [i] = 0 \quad \text{para } i \notin [2^j n, 2^j n + (2^j - 1)(L - 1)], \quad (\text{A.42})$$

y es muy importante desde el punto de vista de la eficiencia y su implementación numérica exacta.

Según estas propiedades, se puede afirmar que ciertas wavelets son mejores que otras según la aplicación específica. Aunque estas propiedades han sido analizadas y tenidas en cuenta en aplicaciones como compresión de señales y eliminación de ruido, su aplicación en clasificación de señales aún es objeto de evaluación (Englehart, 1998). En todo caso, para cualquier aplicación, se considera las mejores funciones wavelets aquellas que mejor se acoplan a la forma de onda de la señal, evaluado el grado de acoplamiento por la magnitud de los coeficientes wavelet en cada subbanda de descomposición, lo que garantizará un adecuado modelado de la señal original en función de las bases wavelets seleccionadas.

A.2.3 Transformada de Paquetes Wavelet (WPT)

La *transformada de paquetes wavelets* WPT es una generalización de la transformada DWT, donde la descomposición en cada nivel no se limita solamente a las bandas de aproximación, sino también sobre las bandas de detalles respectivas. Como se puede observar en la figura A.13, una señal $x[n]$ con N muestras de longitud genera en el primer nivel de descomposición, mediante filtraje paso-bajas y paso-altas, dos subbandas Hx, Gx de longitud $N/2$ conteniendo la mitad del espectro de la señal cada una (igual que la descomposición DWT), en el segundo nivel de descomposición el filtraje se realiza sobre estas dos subbandas, generando cada una dos subbandas más, obteniéndose cuatro sub-bandas H^2x, GHx, HGx, G^2x de longitud $N/4$ cada una, conteniendo cada una $1/4$ del espectro de frecuencias de la señal original (Burrus, et al. , 1999).

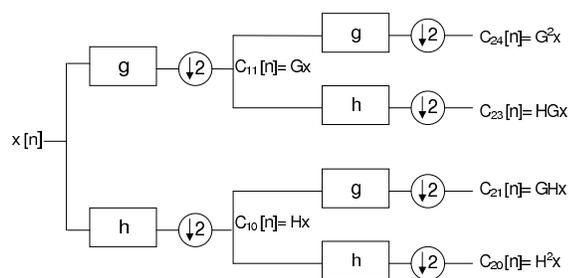


Figura A.13. Descomposición wavelet WPT a dos niveles ($J=2$)

Este proceso de descomposición puede realizarse hasta un nivel máximo $J \leq \log_2 N$, resultando en JN coeficientes en total y a pesar de que este procedimiento

de descomposición es más completo que el de la descomposición DWT, su costo computacional aún resulta ser menor que el requerido por la transformada STFT, esto es $O(JN) \leq O(N \log_2 N)$ (Wickerhauser, 1994).

La descomposición WPT genera una estructura denominada *árbol de paquetes wavelet binario*, donde cada nodo del árbol representa un subespacio con diferentes características de localización en frecuencia. Así mismo, la partición del plano bidimensional tiempo-frecuencia resulta mucho más completo y lo más importante es que puede ser configurable particularmente de acuerdo a las necesidades de cada problema. En la figura A.14, se puede apreciar la diferencia en las resoluciones de análisis DWT y WPT con respecto a las particiones del plano tiempo-frecuencia. En cada caso se considera una descomposición a tres niveles, donde puede observarse que la resolución de análisis para las altas frecuencias es mucho mejor en WPT que en DWT.

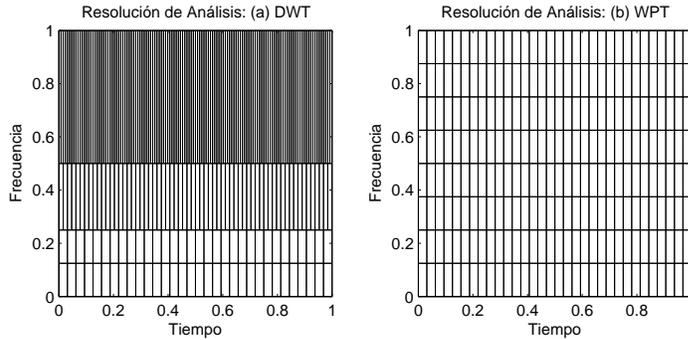


Figura A.14. Resoluciones tiempo-frecuencia a) Análisis DWT y b) Análisis WPT

Para la identificación de las subbandas resultantes de la descomposición WPT se adopta la siguiente notación: el subespacio del nodo raíz del árbol binario se identifica con $\Omega_{0,0}$ y cualquier nodo $\Omega_{j,k}$ será descompuesto en dos subespacios ortogonales mediante los operadores $H : \Omega_{j,k} \rightarrow \Omega_{j+1,2k}$ y $G : \Omega_{j,k} \rightarrow \Omega_{j+1,2k+1}$. Donde j indica la *escala* y k el índice de la subbanda dentro de la respectiva escala. De tal manera que se puede expresar cualquier nodo $\Omega_{j,k}$ como:

$$\Omega_{j,k} = \Omega_{j+1,2k} \oplus \Omega_{j+1,2k+1} \quad \text{para } j = 0, \dots, J \quad \text{y } k = 0, \dots, 2^j - 1 \quad (\text{A.43})$$

En la figura A.15, se muestra el árbol de paquetes wavelet binario de una descomposición a nivel $J = 3$. Sin embargo, esta forma de descomposición resulta en

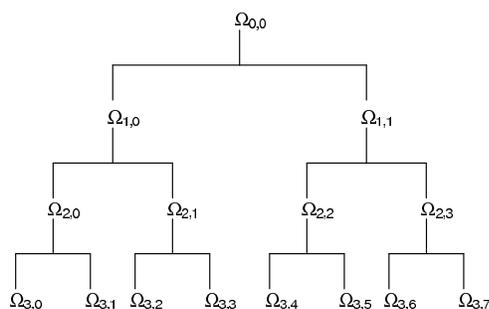


Figura A.15. Estructura de subespacios resultantes en el árbol binario WPT

un conjunto de subespacios altamente redundantes puesto que hay más de $2^{2^{(J-1)}}$ bases ortonormales posibles en este árbol binario (Wickerhauser, 1994). Cada subespacio $\Omega_{j,k}$ es generado por 2^{n_0-j} vectores base $\{W_{j,k,n}\}_{n=0}^{2^{n_0-1}-1}$. En esta familia de *bases de paquetes wavelets* el parámetro j indica la escala, k indican la banda de frecuencias y n el centro de la forma de onda. Así, el vector $W_{j,k,n}$ está centrado al rededor de $2^j n$, tiene una longitud de soporte $\approx 2^j$ y oscila $\approx k$ veces. Para $j = 0$ se tiene el espacio de la señal original (La base Euclídea estándar R^N). Un conjunto típico de vectores base de paquetes wavelets se muestra en la figura A.16.

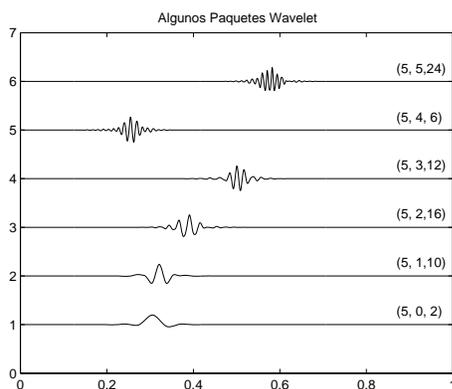


Figura A.16. Vectores base paquetes wavelets con índices $(j,k,n) = (\text{escala, frecuencia y posición})$

A.3 Algoritmo: Selección de las Mejores Bases según Coifman-Wickerhauser

El algoritmo de la mejor base propuesto por Coifman y Wickerhauser (1992) se basa en un proceso de podado de las ramas del árbol binario, que empieza con el árbol de la descomposición WPT completa, partiendo del nivel más bajo se comienza a eliminar ciertas ramas que conduzcan a encontrar la solución óptima para la aplicación específica. La función de costo que se asocia al algoritmo de podado se basa en una medida de entropía, de tal manera que para compresión de señales, se busca maximizar la cantidad de información a partir del conjunto de bases seleccionadas. La medida más natural de entropía es la *Shannon*:

$$H(\mathbf{p}) \doteq \sum_i p_i \log_2 p_i, \quad (\text{A.44})$$

donde, $\mathbf{p} = \{p_i\}$ es una secuencia no negativa tal que $\sum_i p_i = 1$. Aunque existen otras formas de medida de entropía con efectos variantes en la salida del algoritmo (Wickerhauser, 1994).

El algoritmo de *podado* se puede describir de la siguiente manera: supongamos un subespacio $\Omega_{j,k}$ en el árbol binario de paquetes wavelets WPT. Sea $B_{j,k}$ el conjunto de vectores base generador del subespacio $\Omega_{j,k}$, de la forma:

$$B_{j,k} = [\mathbf{W}_{j,k,0}, \mathbf{W}_{j,k,1}, \dots, \mathbf{W}_{j,k,2^{n_0-j}-1}]^T \quad (\text{A.45})$$

Supongamos que $A_{j,k}$ representa la mejor base para la señal restringida a la expansión de $B_{j,k}$ y \aleph la función de costo seleccionada para la medida de información. El siguiente algoritmo “poda” las ramas del árbol binario mediante comparación de la función de costo de cada *nodo padre* con la de sus dos *nodos hijos*.

Dada una señal x :

- i) Se selecciona un método de descomposición tiempo-frecuencia (en este caso WPT), especificando el nivel de descomposición J y la función de costo de la información \aleph .
- ii) Se descompone la señal x en su árbol de paquetes wavelet binario y se obtiene los coeficientes $\{B_{j,k}\mathbf{x}\}$ para $0 \leq j \leq J$ y $0 \leq k \leq 2^j - 1$.

- iii) Empezando en el nivel J : fijar $A_{j,k} = B_{j,k}$ para $k = 0, \dots, 2^j - 1$
- iv) Se determina el mejor subespacio $A_{j,k}$ para $j = J - 1, \dots, 0$, $k = 0, \dots, 2^j - 1$ mediante

$$A_{j,k} = \begin{cases} B_{j,k} & \text{Si } \aleph(B_{j,k} \mathbf{x}) \leq \aleph(A_{j+1,2k} \mathbf{x} \cup A_{j+1,2k+1} \mathbf{x}) \\ A_{j+1,2k} \oplus A_{j+1,2k+1} & \text{en otro caso.} \end{cases} \quad (\text{A.46})$$

Al completarse el algoritmo, se termina en $A_{0,0}$, obteniendo la mejor base para la señal x restringida a la expansión de $B_{0,0} \equiv R^N$. La mejor base seleccionada consiste en un conjunto de subespacios disjuntos, donde cada subespacio $\Omega_{j,k}$ contiene 2^{n_0-j} vectores base. El número total de funciones base es siempre N , donde $N = 2^{n_0}$ es la longitud de la señal x . EL algoritmo puede ser más rápido si la función de costo se considera *aditiva*, esto es $\aleph(\{x_i\}) = \sum_i \aleph(x_i)$ de tal manera que

$$\aleph(A_{j+1,2k}x \cup A_{j+1,2k+1}x) = \aleph(A_{j+1,2k}x) + \aleph(A_{j+1,2k+1}x), \quad (\text{A.47})$$

lo que implica que una simple adición es suficiente en lugar de evaluar el costo de la unión de los nodos. El costo computacional de este algoritmo es $O(N)$. Obteniéndose finalmente un reducido número de coeficientes correspondientes a los subespacios seleccionados, en lugar de tener la cantidad de coeficientes resultado de la descomposición completa en el árbol binario.

Ahora interesa interpretar esta selección de las mejores bases con respecto al análisis y localización en tiempo-frecuencia. Para ello se presenta nuevamente el caso de la señal *chirp lineal*.

En la figura A.17 se muestra la señal *chirp lineal* con sus respectivas sub-bandas de la descomposición WPT de cuatro niveles de profundidad y el árbol resultante después someter el árbol binario completo al algoritmo de selección de mejores bases, el cual se encarga de determinar las subbandas con mayor concentración de energía o información de la señal. Las mejores bases forman un conjunto disjunto mutuamente ortogonal. Para esto, se usó el algoritmo de *podado* de Coifman y Wickerhauser, teniendo en cuenta la función de costo basada en la métrica de información llamada *entropía de Shannon*. Las ramas del árbol muestran como se segmenta el espectro de frecuencias de la señal según las subbandas del árbol, y la altura de cada rama en el árbol es proporcional al incremento en la función de costo obtenido de la ramificación de un nodo padre en sus nodos hijos (Englehart, 1998).

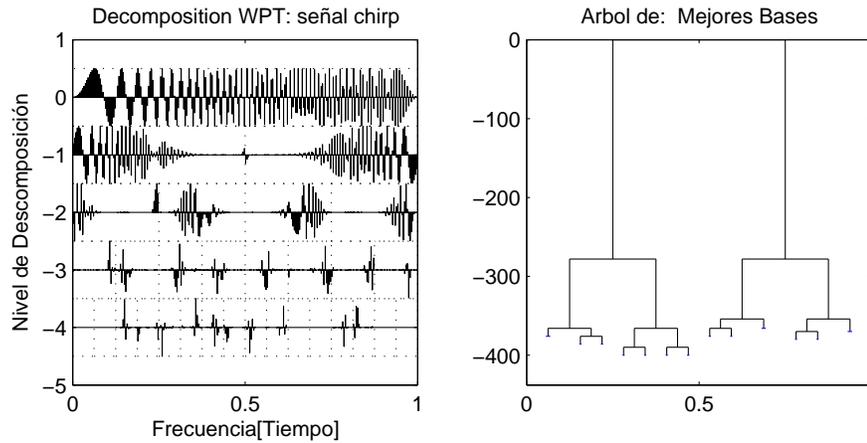


Figura A.17. Selección de las mejores bases del árbol binario WPT

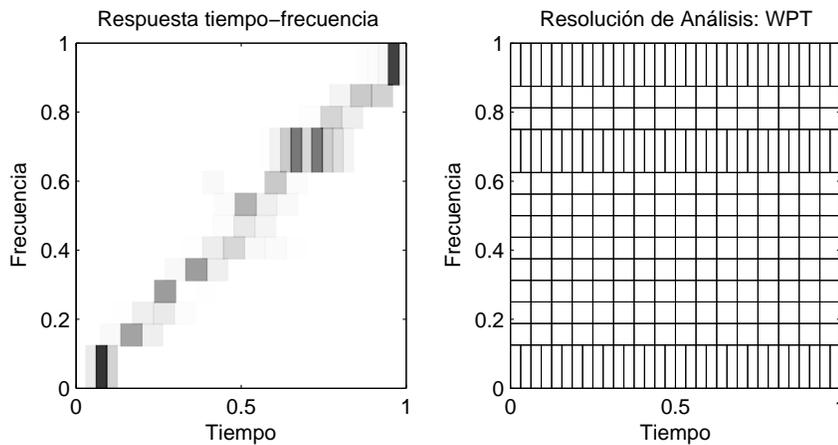


Figura A.18. Respuesta tiempo-frecuencia de la señal *chirp lineal*

En la figura A.18, se puede apreciar que la detección del incremento lineal de la frecuencia con el tiempo de la *señal chirp lineal* es más precisa, gracias a la forma más detallada de la descomposición WPT, lo que permite una mejor resolución de análisis y mejor localización tiempo-frecuencia. Se puede ver también como se ubican los coeficientes de mayor magnitud en las celdas de la diagonal sobre la partición del plano bidimensional tiempo-frecuencia.

Aplicación de la WPT en Compresión de Señales

En aplicaciones de compresión, después de obtener los coeficientes wavelets que resultan de la selección de las mejores bases, estos son sometidos a un proceso de filtraje por umbralamiento. Este proceso consiste en retener solamente los

coeficientes necesarios para lograr una reconstrucción de la señal con cierta medida de calidad a partir de estos. En un esquema de *umbralamiento rígido* o “*hard thresholding*”, únicamente los coeficientes que exceden un umbral especificado se conservan. Una forma de medir la calidad de la compresión es mediante la medida del error de aproximación

$$\epsilon = \frac{\sum (x - \tilde{x})^2}{\sum x^2}, \quad (\text{A.48})$$

donde x es la señal original y \tilde{x} es la señal síntesis, reconstruida a partir de los coeficientes umbralizados. En la figura A.19 se puede observar la curva de error de aproximación contra el número de coeficientes wavelets WPT no nulos (superiores al umbral seleccionado), al descomponer un registro de señal EMGS en un segmento de 256 muestras. Así, para un punto sobre la curva de error por ejemplo para 32 coeficientes (de 256) usados para la reconstrucción de la señal se registra un error de aproximadamente 0.33, lo que significa un factor de compresión 8:1 con un error de aproximación de 0.33.

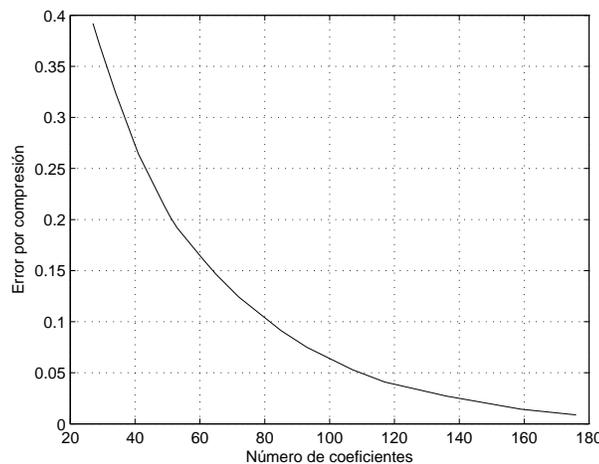


Figura A.19. Error de aproximación en compresión de señales EMGS

Se puede comprobar que el factor de compresión incide directamente en la calidad de la señal reconstruida como se observa en la figura A.20. Cuanto más se comprime la señal, menor será la calidad de la señal a sintetizar y la forma de onda se consigue replicar con buen grado de aceptación, salvo por aquellos detalles mínimos similares a “ruido” en la señal que parecen perderse en el proceso

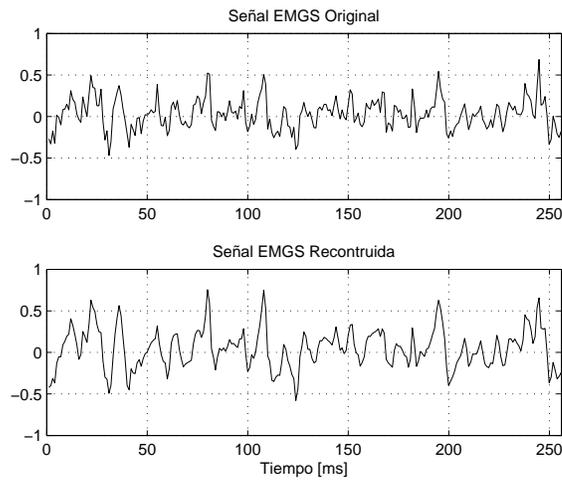


Figura A.20. Comparación entre la señal EMG original y su reconstrucción

análisis-síntesis, lo que demuestra que la representación de señales EMGS basada en wavelets es apropiada para su modelado.

Como se puede observar, la eficiencia la transformada WPT en aplicaciones de compresión se puede comprobar por su capacidad de identificación de las subbandas con la información relevante en la señal. Esta capacidad será aprovechada en este trabajo para extraer características descriptoras a partir de dicha información para propósitos de clasificación de patrones EMGS.

Anexo B

B.1 Algoritmo: Índice de Separabilidad de Clases (CSI)

Considérese un problema de dos clases de patrones, con un conjunto de características originales representadas en un vector $\mathbf{v} \in V \subseteq R^M$. El criterio de selección a cada característica individualmente. Sean p y q las dos clases, cada una con P y Q patrones respectivamente, organizados en forma matricial así:

$$[\mathbf{v}^{(p,1)} \mathbf{v}^{(p,2)} \dots \mathbf{v}^{(p,P)}] \quad y \quad [\mathbf{v}^{(q,1)} \mathbf{v}^{(q,2)} \dots \mathbf{v}^{(q,Q)}], \quad (\text{B.1})$$

donde el m -ésimo patrón de la clase p se denota por:

$$\mathbf{v}^{(p,m)} = [\mathbf{v}_1^{(p,m)} \mathbf{v}_2^{(p,m)} \dots \mathbf{v}_M^{(p,m)}]^T. \quad (\text{B.2})$$

La capacidad discriminante se evalúa sobre cada característica, esto es, la discriminabilidad de la característica i -ésima denotadas como p_i y q_i respectivamente se aplica a la fila i -ésima de cada matriz. Esta medida se denota como $D(p_i, q_i)$ y representa que tanta separabilidad hay entre p_i y q_i . El índice de separabilidad de clases CSI se puede clasificar en tres tipos de acuerdo a la naturaleza de las características p_i y q_i (Englehart, 1998):

Tipo 1: Los datos para evaluar la medida $D(p_i, q_i)$ son no procesados. En este caso se puede aplicar el índice de separabilidad de clases de Fisher definido como:

$$D(p_i, q_i) \doteq \frac{(\text{mean}(p_i) - \text{mean}(q_i))^2}{\text{var}(p_i) + \text{var}(q_i)}, \quad (\text{B.3})$$

donde $mean(\cdot)$ y $var(\cdot)$ significan la media y varianza de las muestras de la señal en la fila i -ésima de las matrices respectivas.

Tipo 2: Los datos para la medida $D(p_i, q_i)$ representan la *media* o la *energía media* de la i -ésima característica; o pueden representar medidas similares como los momentos de los datos. En este caso, si p_i y q_i son escalares, una rigurosa medida de separabilidad de clases se puede evaluar mediante:

i) **Entropía relativa:**

$$D(p_i, q_i) \doteq p_i \log \frac{p_i}{q_i}, \quad (\text{B.4})$$

la cual mide la diferencia no simétrica entre p_i y q_i .

ii) **Entropía relativa simétrica:**

$$D(p_i, q_i) \doteq p_i \log \frac{p_i}{q_i} + q_i \log \frac{q_i}{p_i}, \quad (\text{B.5})$$

donde la diferencia entre p_i y q_i es simétrica, es decir $D(p_i, q_i) = D(q_i, p_i)$.

iii) **Distancia Euclídea:**

$$D(p_i, q_i) \doteq \|p_i - q_i\| = (p_i - q_i)^2, \quad (\text{B.6})$$

también asimétrica.

Tipo 3: Los datos para la medida $D(p_i, q_i)$ representan la función de densidad de probabilidad de la i -ésima característica. Para esto se requiere estimar la función densidad de probabilidad de cada característica en cada una de las clases, denotadas como $p_i(x)$ y $q_i(x)$. Dos medidas de distancia comunes entre funciones de densidad de probabilidad son:

i) **Entropía relativa:**

$$D(p_i, q_i) \doteq \int p_i(x) \log \frac{p_i(x)}{q_i(x)} dx \quad (\text{B.7})$$

ii) **Distancia de Hellinger:**

$$D(p_i, q_i) \doteq \int \left(\sqrt{p_i(x)} - \sqrt{q_i(x)} \right)^2 dx \quad (\text{B.8})$$

Si se consideran todos los posibles subconjuntos de características el costo computacional para evaluar la medida de discriminación se incrementa sustancialmente. En su lugar se prefieren métodos mas prácticos, aunque no tan óptimos, que no realicen búsquedas tan exhaustivas ni busquen la mejor solución (Englehart, 1998). Algunos de estos métodos son (Fukunaga , 1990):

- a) **Método Add-On:** evalúa una medida de discriminabilidad para cada característica, y la mejor se selecciona como la primera del subconjunto. Las restantes se combinan con la primera para determinar las dos mejores. Este proceso se repite hasta lograr el número de características del subconjunto deseado.
- b) **Método Knock-Out:** Este método es similar al anterior, excepto por que el conjunto de características original es reducido al tamaño del subconjunto deseado.

Cabe notar que la medida de separabilidad de clases no implica desempeño de clasificación. Si fuera posible utilizar un criterio ideal de tasa de error de clasificación, se esperaría que mientras menor sea el número de características retenidas mayor sería la capacidad de generalización del clasificador. Sin embargo, esto continuaría así hasta encontrar un conjunto de características de tamaño óptimo, de un tamaño tan reducido que el desempeño del clasificador empezaría a degradarse. A pesar de todo esto, los criterios de separabilidad de clases son muy útiles para comparar conjuntos de características de igual tamaño (Englehart, 1998).

B.2 Algoritmo: Análisis de Componentes Principales (PCA)

Sea un conjunto de P vectores descriptores $\mathfrak{S} \equiv [\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(P)}]$, patrones de una clase, donde cada vector se compone de M características $\mathbf{v}^{(p)} = [v_1^{(p)} \dots v_M^{(p)}]^T$. El objetivo del análisis PCA es mapear cada vector original $\mathbf{v}^{(p)} \in V \subseteq R^M$ en un nuevo vector de dimensión reducida $\mathbf{z}^{(p)} = [z_1^{(p)} \dots z_L^{(p)}]^T \in Z \subseteq R^L$, $L < M$, de acuerdo a algún criterio como el MMSE. Esto es, un vector de características $\mathbf{v}^{(p)}$ puede ser representado sin error mediante la sumatoria de M vectores ortonormales como:

$$\mathbf{v}^{(p)} = \sum_{i=1}^M z_i^{(p)} \mathbf{u}_i, \quad (\text{B.9})$$

donde los coeficientes $z_i^{(p)}$ se determinan mediante la expresión

$$z_i^{(p)} = \mathbf{u}_i^T \mathbf{v}^{(p)}. \quad (\text{B.10})$$

Suponiendo que se desea retener solamente L de los M vectores base \mathbf{u}_i ($L < M$), la ecuación (2.37) se puede descomponer en dos sumatorias: una, que contiene los L coeficientes $z_i^{(p)}$ y otra donde se reemplazan los coeficientes restantes por coeficientes b_i , como se muestra a continuación:

$$\tilde{\mathbf{v}}^{(p)} = \sum_{i=1}^L z_i^{(p)} \mathbf{u}_i + \sum_{i=L+1}^M b_i \mathbf{u}_i. \quad (\text{B.11})$$

Considerando un conjunto de datos \mathfrak{S} con P vectores descriptores, se debe seleccionar los vectores \mathbf{u}_i y los coeficientes b_i de tal manera que se logre la mejor aproximación de cada vector original del conjunto de datos.

El error cuadrático para todo el conjunto de datos es:

$$E_L \doteq \frac{1}{2} \sum_{p=1}^P \|\mathbf{v}^p - \tilde{\mathbf{v}}^p\|^2 = \frac{1}{2} \sum_{p=1}^P \sum_{i=L+1}^M (z_i^{(p)} - b_i)^2, \quad (\text{B.12})$$

y para minimizarlo con respecto a b_i se se deriva E_L y se iguala a cero, obteniéndose:

$$b_i = \frac{1}{P} \sum_{p=1}^P z_i^{(p)} = \mathbf{u}_i^T \bar{\mathbf{v}}, \quad (\text{B.13})$$

donde $\bar{\mathbf{v}} = \sum_{p=1}^P \mathbf{v}^{(p)}$, es un vector de medias. Ahora, reemplazando b_i en la suma del error cuadrático (ec. 2.40), se tiene:

$$E_L = \frac{1}{2} \sum_{p=1}^P \sum_{i=L+1}^M \{\mathbf{u}_i^T (\mathbf{v}^{(p)} - \bar{\mathbf{v}})\}^2 = \frac{1}{2} \sum_{i=L+1}^M \mathbf{u}_i^T \Sigma_v \mathbf{u}_i, \quad (\text{B.14})$$

donde Σ_v , es la matriz de covarianza del conjunto de vectores \mathbf{v} , definida como:

$$\Sigma_v \doteq \sum_{p=1}^P (\mathbf{v}^{(p)} - \bar{\mathbf{v}}) (\mathbf{v}^{(p)} - \bar{\mathbf{v}})^T. \quad (\text{B.15})$$

En este punto solo queda minimizar el error E_L con respecto a los vectores base \mathbf{u}_i . La minimización de este error de aproximación se conseguirá siempre que los vectores base \mathbf{u}_i satisfagan la condición:

$$\sum_v \mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad (\text{B.16})$$

donde \mathbf{u}_i son los autovectores y λ_i los autovalores de la matriz de covarianza \sum_v . A cada autovector \mathbf{u}_i se le llama “componente principal”.

En la práctica, el algoritmo para PCA procede sustrayendo del conjunto de datos su media $\bar{\mathbf{v}} = \sum_{p=1}^P \mathbf{v}^{(p)}$, para luego evaluar la matriz de covarianza y sus autovalores. Los autovectores correspondientes a los L autovalores mas grandes son retenidos y el conjunto de vectores descriptores original \mathfrak{S} es proyectado sobre los L autovectores para reducir la dimensión del conjunto original M a uno de dimensión $L < M$.

Anexo C

C.1 El Clasificador Bayesiano

Consiste en un reconocimiento de patrones estadístico, basado en el desarrollo de funciones de decisión a partir de un conjunto de datos finito (o patrones muestra) correspondiente a cada clase; de tal manera que estas funciones puedan particionar el espacio de salida en regiones que contengan, cada una, los patrones muestra pertenecientes a cada clase. Recordemos que según el espacio de la señal de entrada $\mathbf{x} \in X \subseteq R^N$ y el espacio de respuesta $y \in Y = \{y_1, y_2, \dots, y_K\}$, la pareja (\mathbf{x}, y) representa el patrón de entrada y su etiqueta de clase respectivamente. Para un clasificador estadístico, las medidas de \mathbf{x} e y pueden ser consideradas en una estructura probabilística y vistas como observaciones simples de las variables aleatorias X e Y . Así, el problema del clasificador estadístico se reduce a determinar la mayor información que pueda conocerse sobre las probabilidades *a posteriori* del espacio de entrada, denotada como:

$$P(y_k|\mathbf{x}) \quad \text{para } k = 1, \dots, K, \quad (\text{C.1})$$

que corresponde a la probabilidad condicional de que dado un patrón \mathbf{x} de entrada, sea la clase y_k a la que pertenece. La decisión se basa en la mayor de estas probabilidades, evaluadas para las K clases. Son probabilidades que no se conocen de antemano, pero que pueden ser estimadas a partir de: la probabilidad *a priori* $P(y_k)$ y la función de densidad de probabilidad condicional $p(\mathbf{x}|y_k)$. Para su estimación se aplica el teorema de Bayes (Papoulis, 1985), el cual establece:

$$P(\mathbf{x}, y_k) = P(y_k)p(\mathbf{x}|y_k) = p(\mathbf{x})P(y_k|\mathbf{x}), \quad (\text{C.2})$$

de donde obtenemos que:

$$P(y_k|\mathbf{x}) = \frac{P(y_k)p(\mathbf{x}|y_k)}{p(\mathbf{x})}, \quad (\text{C.3})$$

con

$$p(\mathbf{x}) = \sum_{j=1}^K P(y_j)p(\mathbf{x}|y_j). \quad (\text{C.4})$$

Aquí, $p(\mathbf{x})$ representa la función de densidad de probabilidad del espacio de entrada y dado que permanece constante para todas las probabilidades $P(y_k|\mathbf{x})$, puede ser ignorada en aplicaciones de clasificación. Por otra parte, cuando la verdadera distribución de clases no se conoce, las probabilidades *a priori* se pueden considerar equiprobables, esto es $P(y_k) = \frac{1}{K}$ para $k = 1, \dots, K$.

De esta manera, la regla de decisión de Bayes se reduce a la implementación de la función de decisión:

$$d_k(\mathbf{x}) = p(\mathbf{x}|y_k)P(y_k), \quad k = 1, \dots, K, \quad (\text{C.5})$$

con la cual se decide asignar un patrón \mathbf{x} a la clase y_i , si para este patrón se cumple que $d_i(\mathbf{x}) > d_j(\mathbf{x})$ para todo $j \neq i$. Esta regla de decisión de Bayes minimiza la *probabilidad de error de clasificación*, haciendo de este un clasificador estadísticamente superior a cualquier otro (Englehart, 1998).

El mayor trabajo radica en estimar de manera confiable las densidades de probabilidad condicional $p(\mathbf{x}|y_k)$ a partir de los datos de entrenamiento; lo cual es difícil sino imposible, siempre que la dimensión del espacio de entrada N sea muy grande con relación al número de muestras de entrenamiento, por lo que se recomienda que el número de muestras de entrenamiento sea superior a la dimensión del espacio de entrada. Además, si se puede hacer algún tipo de suposición sobre la naturaleza de las probabilidades condicionales $p(\mathbf{x}|y_k)$, su estimación se simplifica aún mas. Una forma sencilla y muy efectiva de un clasificador Bayesiano es el clasificador tipo *Gausiano*.

C.2 El clasificador Bayesiano Gausiano

Asumiendo una distribución normal (o Gausiana) para los patrones de muestra en sus respectivas clases, se puede estimar las probabilidades condicionales $p(\mathbf{x}|y_k)$ mediante la expresión:

$$p(\mathbf{x}|y_k) = \frac{1}{(2\pi)^{N/2} |\mathbf{C}_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T \mathbf{C}_k^{-1} (\mathbf{x} - \mathbf{m}_k)\right], \quad k = 1, \dots, K, \quad (\text{C.6})$$

donde cada función de densidad de probabilidad queda completamente especificada por su vector de medias \mathbf{m}_k y su matriz de covarianzas \mathbf{C}_k , definidos como:

$$\mathbf{m}_k = E_k[\mathbf{x}] \quad (\text{C.7})$$

y

$$\mathbf{C}_k = E[(\mathbf{x} - \mathbf{m}_k)(\mathbf{x} - \mathbf{m}_k)^T] \quad (\text{C.8})$$

Aquí, $E_k[\cdot]$ denota el *valor esperado* y $|\mathbf{C}_k|$ denota el determinante de la *matriz de covarianzas* \mathbf{C}_k . Los patrones muestra tomados de una distribución normal tienden a caer dentro de un “cluster” con su centro determinado por el vector de medias y su dispersión definida por la matriz de covarianza. La ubicación geométrica de los puntos cuya densidad de probabilidad es constante describe *hiperelipsoides*, cuyos ejes coinciden con la dirección de los vectores propios de la matriz de covarianza y sus longitudes con los valores propios respectivos (Papoulis, 1985).

Retomando la expresión de la probabilidad condicional $P(y_k|\mathbf{x})$ para la clase y_k , y dada la naturaleza exponencial de la función de densidad $p(\mathbf{x}|y_k)$, resulta mas conveniente trabajar con el logaritmo natural de esta función, obteniéndose:

$$d_k(\mathbf{x}) = \ln[p(\mathbf{x}|y_k)P(y_k)] = \ln p(\mathbf{x}|y_k) + \ln P(y_k), \quad (\text{C.9})$$

y sustituyendo (C.6) en (C.9) se tiene:

$$d_k(\mathbf{x}) = \ln P(y_k) - \frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{C}_k| - \frac{1}{2} [(\mathbf{x} - \mathbf{m}_k)^T \mathbf{C}_k^{-1} (\mathbf{x} - \mathbf{m}_k)], \quad k = 1, \dots, K, \quad (\text{C.10})$$

y mientras el término $\frac{N}{2} \ln(2\pi)$ no dependa de k, puede ser eliminado, resultando finalmente la expresión:

$$d_k(\mathbf{x}) = \ln P(y_k) - \frac{1}{2} \ln |C_k| - \frac{1}{2} [(\mathbf{x} - m_k)^T C_k^{-1} (\mathbf{x} - m_k)], \quad k = 1, \dots, K, \quad (\text{C.11})$$

conocida como *función de decisión de Bayes para distribuciones normales* de datos. Esta función de decisión es *hiper cuadrática* siempre que el último término en (C.11) sea de orden dos en \mathbf{x} . En este caso, lo mejor que un clasificador Bayesiano puede hacer es ubicar una *función discriminante cuadrática* entre clases de patrones. Y en caso de que la suposición de la distribución de los patrones como normal sea acertada, ninguna otra superficie de separación dará mejores resultados (Papoulis, 1985).

La función de decisión cuadrática toma la forma:

$$d_k(\mathbf{x}) = \ln P(y_k) - \frac{1}{2} \ln |C_k| - \frac{1}{2} \mathbf{x}^T C_k^{-1} \mathbf{x} + \mathbf{x}^T C_k^{-1} m_k - \frac{1}{2} m_k^T C_k^{-1} m_k, \quad k = 1, \dots, K, \quad (\text{C.12})$$

y si se asume además, que las matrices de covarianzas son iguales para todas las clases, esto es $C_k = C$ para $k = 1, \dots, K$, se llega a una función de decisión lineal de la forma:

$$d_k(\mathbf{x}) = \ln P(y_k) + \mathbf{x}^T C^{-1} m_k - \frac{1}{2} m_k^T C^{-1} m_k, \quad k = 1, \dots, K \quad (\text{C.13})$$

En este caso, la superficie de decisión es lineal con respecto al espacio de entrada y describe un hiperplano. Este clasificador Bayesiano Gaussiano comúnmente se conoce como clasificador de *análisis discriminante lineal* (LDA). Puede anotarse además que las funciones de decisión cuadráticas y lineales derivadas del clasificador de Bayes son teóricamente óptimas, aún para distribuciones que no sean Gaussianas (Papoulis, 1985).

Coste computacional del clasificador Bayesiano. Considerando el caso en que las matrices de covarianza son calculadas a partir de los datos de entrenamiento (sin suposición alguna), teniendo en cuenta dichos productos matriciales, por cada clase K es necesario calcular N restas, N^2 productos, N^2 ponderaciones con los términos de la matriz de covarianza y $N^2 - 1$ sumas de los términos. Así el número de operaciones simples para clasificar cada patrón de entrada será:

$$N_{op} = K (3N^2 + N) - 1 \quad (\text{C.14})$$

C.3 Redes Neuronales Artificiales

Las redes ANN se pueden definir como: “máquinas de aprendizaje distribuidas, adaptativas y generalmente no lineales formadas por varios elementos de procesamiento sencillos denominados neuronas, donde cada elemento recibe conexiones de otros elementos y/o de sí mismo. La interconectividad de los elementos define la topología de la red” (Rosa y Gil , 2006).

Una red ANN se considera como un clasificador de entrenamiento supervisado, esto es, se requiere de un conjunto de datos de *entrenamiento*, \mathfrak{S} , consistente en P pares de muestras del espacio de entrada y sus respectivas etiquetas de clase, representado como:

$$\mathfrak{S} = \{(\mathbf{x}^{(1)}, y^{(1)}), \dots, (\mathbf{x}^{(P)}, y^{(P)})\} \quad P \text{ finito.} \quad (\text{C.15})$$

Si se dispone de la información de *pertenencia de clase* durante el entrenamiento, un método supervisado será mejor que uno no supervisado, pues esta información permite construir fronteras de discriminación más apropiadas. El desempeño de la red ANN y su capacidad de generalización en clasificación se evalúa mediante un conjunto de *datos de prueba* $\tilde{\mathfrak{S}}$.

En general, si ante un patrón dado, el objetivo es minimizar la probabilidad de error en su clasificación, este debe ser asignado a la clase cuya probabilidad *a posteriori* sea la mayor. Ante este problema, son dos los principales puntos de vista del problema de clasificación (Rosa y Gil , 2006):

- i) La red neuronal es un sistema capaz de estimar las distintas funciones de densidad de probabilidad que permitan caracterizar probabilísticamente el problema y diseñar el detector o clasificador óptimo.
- ii) O para evitar la necesidad de estimar funciones de densidad de probabilidad, pueden utilizarse funciones discriminantes parametrizadas y determinar el valor de los parámetros que las caracterizan mediante un proceso de aprendizaje a partir de un conjunto de datos preclasificados.

La unidad básica de una red ANN es la **neurona artificial**, compuesta por: las *entradas*, los *pesos sinápticos*, w_i , que ponderan cada entrada y que pueden modificarse durante el entrenamiento, una *regla de propagación* $s(x)$ que es función de las

entradas y los pesos de ponderación, llamada suma ponderada, aunque se puede realizar otro tipo de operaciones como la distancia Euclídea (caso de las redes con función de base radial), y una *función de activación* $F(s(x))$ que da la salida de la neurona. Existen varias formas de funciones de activación, entre ellas la *función identidad*, la *función escalón*, la *función lineal a tramos*, la *función Sigmoidea* y la *función Gaussiana* (Rosa y Gil , 2006).

El perceptrón es el primer modelo de sistema capaz de aprender mediante el entrenamiento supervisado y básicamente es una neurona con un conjunto de pesos ajustables y una función de activación tipo escalón, su arquitectura se muestra en la figura C.1.

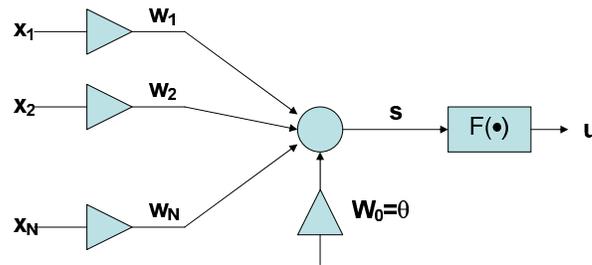


Figura C.1. Arquitectura del Perceptrón

El perceptrón divide el espacio de observación en dos regiones separadas por un hiperplano. La función de base de un perceptrón es una *función de base lineal*, que consiste en una suma “pesada” de acuerdo a los pesos, w_i , de las N componentes del vector de entrada $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$, a la cual se le agrega un valor inicial o “sesgo” de peso $w_0 = \theta$. Una función de base lineal, por tanto, define un hiperplano en el espacio de entrada. El resultado de la suma, s , se pasa por una función de activación *no lineal*, $f(\bullet)$, cuya activación es proporcional a la distancia de la entrada considerada al plano definido por el conjunto de pesos que definen la función. Así, a su salida se tiene como respuesta, u , que puede ser una etiqueta de clase o un nuevo valor de entrada para la siguiente capa de la red, de acuerdo a las expresiones:

$$s(x) = \sum_{i=1}^N w_i x_i + \Theta = \mathbf{w}^T \mathbf{x} + \Theta \quad (\text{C.16})$$

y

$$u = f(s(x)). \quad (\text{C.17})$$

Las funciones de activación *no lineales* más empleadas son las *sigmoides*, entre las que se encuentra la función *logarítmica* dada por la expresión

$$f_{\log}(s) = \frac{1}{1 + e^{-\beta s}} \quad (\text{C.18})$$

y la *tangente hiperbólica* dada por la expresión

$$f_{\tanh}(s) = \frac{e^{\beta s} - e^{-\beta s}}{e^{\beta s} + e^{-\beta s}}. \quad (\text{C.19})$$

El entrenamiento del perceptrón (igualmente aplicable al perceptrón multicapa), consiste en que a partir de un valor inicial del vector de pesos $w(0)$, se van obteniendo nuevas versiones de dicho vector $w(1)$, $w(2)$, ..., de modo que el error se debe ir reduciendo en cada iteración. Para esto los métodos más utilizados y que se basan en la evaluación de las derivadas de dicha función de error son:

- **El método del gradiente.** Donde el vector de pesos de la iteración $i + 1$ se obtiene mediante

$$w^{(i+1)} = w^{(i)} - \eta \nabla(e(w^{(i)})), \quad (\text{C.20})$$

donde ∇ es el operador gradiente, η es el parámetro que controla la velocidad del algoritmo y $e(w)$ la función de error.

- **El método de Newton.** El cual actualiza el vector de pesos mediante de la iteración $i + 1$ mediante

$$w^{(i+1)} = w^{(i)} - H^{-1(i)} \nabla(e(w^{(i)})), \quad (\text{C.21})$$

donde H^i es la matriz Hessiana de la función de error $e(w)$.

- **El método de Gauss-Newton.** El cual actualiza el vector de pesos de la iteración $i + 1$ mediante

$$w^{(i+1)} = w^{(i)} - (J^{(i)T} J^{(i)})^{-1} J^{(i)T} e^{(i)}(w), \quad (\text{C.22})$$

donde $J^{(i)}$ es la matriz Jacobiana de la función de error $e^{(i)}(w)$.

Para entender la forma de clasificación de una red neuronal es necesario recordar al clasificador estadístico Bayesiano, cuyo objetivo principal es estimar la probabilidad *a posteriori* $P(y_k|\mathbf{x})$, a partir de las probabilidades *a priori* $P(\mathbf{x})$ y las funciones de densidad de probabilidad condicionales $p(\mathbf{x}|y_k)$ según la regla de Bayes:

$$P(y_k|\mathbf{x}) = \frac{P(y_k)p(\mathbf{x}|y_k)}{p(\mathbf{x})}. \quad (\text{C.23})$$

La metodología de una red ANN para reconocimiento de patrones consiste en estimar la probabilidad $P(y_k|\mathbf{x})$ directamente, mediante la “aproximación” de una función que el algoritmo de aprendizaje supervisado configure. Así, dado un conjunto de muestras de entrenamiento, se desea entrenar un estimador $\theta_k(\mathbf{x}, \mathbf{w})$ para aproximar la probabilidad $P(y_k|\mathbf{x})$ para cada clase $k = 1, \dots, K$, donde \mathbf{w} representa los parámetros del estimador (o conjunto de pesos de la red ANN). En una situación ideal, el conjunto de entrenamiento consistiría en parejas ordenadas $(\mathbf{x}, P(y_k|\mathbf{x}))$, pero las probabilidades *a posteriori* $P(y_k|\mathbf{x})$ raramente se conocen y en su lugar se utilizan las etiquetas de clase y_k .

Si en la fase de entrenamiento se usa como función de costo el *error cuadrático medio*, se obtendrá como salidas 0's o 1's en lugar de las probabilidades $P(y_k|\mathbf{x})$, sin embargo la solución óptima de los parámetros $\theta_k(\mathbf{x}, \mathbf{w})$ es exactamente la misma que si las probabilidades *a posteriori* fueran las salidas deseadas. En este sentido una red MLP con algoritmo *backpropagation* puede lograr la mejor aproximación de las probabilidades *a posteriori* $P(y_k|\mathbf{x})$ de acuerdo al *error cuadrático medio*.

La no linealidad de la red limita automáticamente a la salida entre 0 y 1, y la suma de la salida de un MLP para un patrón dado tiende a uno (1) (condición de una distribución de probabilidad), lo que hace ver que no solo un adecuado entrenamiento de un clasificador MLP ofrece las mismas capacidades de un clasificador Bayesiano, sino que además provee ponderaciones de la decisión. Para lograr cierta capacidad de generalización de la red, se requiere además de determinar el tamaño óptimo de la red, según el problema en cuestión, disponer de un conjunto suficientemente grande de muestras de entrenamiento.

Los aspectos más importantes de funcionamiento e implementación de una red neuronal MLP son:

- i) **Tasa de aprendizaje.** Puede ser uniforme en toda la red o diferente en cada capa o nodo. En general, es difícil determinar la mejor tasa de aprendizaje,

pero una regla útil es hacer que la tasa de aprendizaje de cada nodo sea inversamente proporcional a la dimensión del vector que alimenta al nodo. Algunos esquemas adaptan la tasa de aprendizaje en función de la curvatura local de la superficie de error seleccionada (Englehart, 1998). Lo más simple y práctico es agregar un término llamado *momentum* de la forma $\alpha(\mathbf{w}_{l,j,k}(k) - \mathbf{w}_{l,j,k}(k-1))$ a cada fase de actualización de pesos, donde $0 < \alpha < 1$. Este término hace de la dirección de búsqueda un promedio exponencialmente pesado de direcciones pasadas, ayudando a conservar el movimiento de los pesos sobre las porciones planas de la superficie de error después de su descenso escalonado.

ii) **Criterio de parada.** El proceso iterativo del cálculo del gradiente y ajuste de pesos se mantiene hasta encontrar un mínimo en la superficie de error (o un punto suficientemente cercano a dicho mínimo). Existen diferentes criterios de parada; uno es, si la magnitud del gradiente cae dentro de un nivel prefijado, el algoritmo se detiene, asumiendo que el mínimo se ha logrado. Sin embargo, el criterio de parada más común es el umbral mínimo en la suma del error cuadrático. Es claro que se requiere conocer previamente un posible valor para el umbral mínimo del error deseado, lo cual no siempre es posible. Otro criterio de parada podría ser el número de iteraciones ejecutadas en la búsqueda de un umbral de error deseado. Este número de iteraciones queda a criterio de ensayo y error sobre la ejecución del algoritmo respectivo. Finalmente, mediante una *validación cruzada* se monitorea la capacidad de *generalización* durante el aprendizaje. Este método requiere particionar la base de datos o patrones en un conjunto de entrenamiento y otro de prueba. En aplicaciones de reconocimiento de patrones, el método de validación cruzada puede extenderse al monitoreo de la tasa de clasificación de la red sobre los dos conjuntos de datos, deteniendo el algoritmo cuando la tasa lograda sobre el conjunto de prueba sea máxima. No hay garantía de que el mejor desempeño de la red con respecto al mínimo error cuadrático (sobre cualquiera de los conjuntos de datos) refleje el mejor conjunto de pesos de la red para un buen desempeño de clasificación. Esta es una desventaja de usar la suma del error cuadrático como función de costo del algoritmo *backpropagation*, por lo que cualquier otra función de costo podría lograr mejores resultados de la red para efectos

de reconocimiento de patrones.

iii) **Nodos de capa oculta.** El número de nodos óptimo en cada capa oculta es difícil determinarse de manera generalizada. Pues este depende fuertemente de la naturaleza de los datos y cada problema demanda diferentes capacidades de la red. Una red muy pequeña no será capaz de formar un modelo adecuado del sistema. Y a la inversa, si es muy grande podría sobredimensionar al mismo, implementando numerosas soluciones, todas consistentes con los datos de entrenamiento pero muchas de las cuales resultan ser aproximaciones muy pobres del problema. En reconocimiento de patrones, se vería que el conjunto de entrenamiento es fácilmente discriminado, mientras que para el conjunto de prueba las fronteras de discriminación tienen una pobre generalización. Finalmente, el número óptimo de nodos en las capas ocultas, será aquel con el cual la red sea capaz de capturar la estructura *esencial* de los datos.

A pesar de que con nada o muy poco conocimiento de la estructura de los datos se puede determinar por ensayo y error el tamaño de la red, resulta mas conveniente seguir algún procedimiento metódico. Por ejemplo, empezando por un número pequeño de nodos e ir incrementándolo hasta conseguir los resultados deseados, con entrenamiento de la red independiente para cada tamaño. Otra forma metódica es mediante el algoritmo de ramificación y podado de ramas de la red (Englehart, 1998), que consiste en agregar o eliminar nodos de acuerdo a como se vayan requiriendo. En todo caso, se recomienda que el número de nodos de las capas ocultas sea siempre menor que el número de muestras de entrenamiento, de lo contrario la red simplemente *memorizará* dichas muestras, resultando en una pobre generalización.

Referencias

- Aldabas E., (2002). Introducción al Reconocimiento de Patrones Mediante Redes Neuronales, Disponible en:<<http://www.jcee.upc.es/JCEE2002/Aldabas.pdf>>. Acceso en 20 jul. 2007.
- Alfred M. (1999). *Signal Analysis Wavelets, Filter Banks, Time-Frequency Transform and Applications*, University of Wollongong, Australia. Jhon Wiley & Sons Ltda. 1999.
- Alonso A., Sanchez H., hurtado E., Steinz D. y Liptak L. (2002). Entrenador Mioeléctrico de Prótesis Para Amputados de Brazo y Mano. Hospital clínico Universitario - Universidad de Valladolid. *Mapfre Medicina*. Vol 13, pp. 11- 19.
- Betancourt G., Giraldo E. y Franco J. (2004). Reconocimiento de Patrones de Movimiento a Partir de Señales Electromiográficas. Grupo de Investigación en Instrumentación y control - Universidad Tecnológica. *Scientia et Technica*. Vol 26, pp. 53 - 58, 2004.
- Betancourt G., (2005). Las máquinas de Soporte Vectorial SVMs. Grupo de Instrumentación y control - Universidad Tecnológica. *Scientia et Technica*. Año XI. No. 27, pp. 67 - 72, 2005.
- Breiman, L., Friedman, H., Olsen, R.A. y C.J. Stone. *Clasification and Regression Trees*, Chapman and Hall, Inc., New York, 1984.
- Burrus S., Gopinath R., Guo H. (1999). *Introduction to Wavelets and Wavelet Transform*. Electrical and Computer Engineering Department and Computer and In-

formation Technology Institute. Rice University. Houston, Texas. Prentice Hall. New Jersey 07458.

Chang G., Kang W., Luh J., Cheng C., Lai J., Chen J., Kuo T. (1966). Real-time Implementation of Electromyogram Pattern Recognition as a Control Command of Man-machine Interface. *Med. Eng. Phys.* Vol. 18, pp. 529 - 537.

Coifman R. y Wickerhauser M. (1992). Entropy-based algorithms for best basis selection. *IEEE. Trans Information Theory*, Vol. 38, No. 2, pp. 713 - 719, 1992.

Crawford B., Miller K., Shenoy P., y Rao R. (2005). Real-Time Classification of Electromyographic Signals for Robotic Control. Department of Computer Science, University of Washington. *Technical Report* No. 2005-03-05.

Daubechies I., (1988) Orthonormal bases of compactly supported wavelets. *Communication on pure and applied mathematics*, 41, pp. 909-996, 1988.

Daubechies I. (1992) *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61, SIAM, Philadelphia, 1992.

De Castro F. (2002). Análisis de la Teoría de Ondículas Orientadas a las Aplicaciones en Ingeniería Eléctrica: Fundamentos. *E.T.S.I. Industriales Departamento de Ingeniería Eléctrica*. Madrid Julio de 2002.

De Luca C. (1978). Physiological Introduction to the Motor Unit. Disponible en: http://nmrc.bu.edu/tutorials/motor_units/index.html. Acceso en 20 jun. 2006.

Deluca C. (1979). Physiology and Mathematics of Myoelectric Signals. *IEEE Trans. Biomed. Eng.* Vol. BME-26, pp:313-325.

Englehart, K. (1998). Ph.D. Thesis. Signal Representation for Classification of the Transient Myoelectric Signal. University of New Brunswick, Fredericton, New Brunswick.

Englehart K., Hudgins B., Parker P. y Stevenson M. (1999a). Classification of the Myoelectric Signal Using Time-frequency Based Representations. Institute of Biomedical Engineering, University of New Brunswick, Canada. 1999a.

- Englehart K., Hudgins B. y Parker P. (1999b). Improving Myoelectric Signal Classification Using Wavelet Packets and Principal Component Analysis. Proc. Of 21st Annual Int. Conf. of the IEEE on Eng. In Med. And Biol. Soc. Atlanta 1999b.
- Englehart K., Hudgins B.y Parker P. (2001). A Wavelet Based Continuous Classification Scheme for Multifunction Myoelectric Control. Department of Electrical and Computer Engineering and the Institute of Biomedical Engineering. University of New Brunswick, Canada. Vol. 48, pp. 302 - 311, 2001.
- Faundez P. (2002). Procesamiento digital de señales acústicas utilizando wavelets. *Instituto de Matemáticas UACH Chile*. 2002.
- Ferguson S. y Dunlop G Reg. (2002). Grasp Recognition From Myoelectric Signals. *Australian conf. on Robotics and Automation, Auckland*. Disponible en: <<http://www.araa.asn.au/acra/acra2002/Papers/Ferguson-Dunlop.pdf>>. Acceso en 17 nov. 2006.
- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, 2nd Ed., Academic press, San Diego, CA 1990.
- Gabor, D.(1946) Theory of communication, *Inst. Elec. Eng.*, 93, 429-457, 1946.
- Gasca D. y Rojas L.(2008). Trabajo de grado. Extracción de características descriptoras de movimientos de la mano a partir de señales electromiográficas (EMG) aplicando técnicas wavelets. Universidad del Cauca, Facultad de Ingeniería Electrónica y Telecomunicaciones. Popayán.
- Georgiou V., Pavlidis N., Parsopoulos K., Alevizos D., Vrahatis M. (2004). Optimizing the Performance of Probabilistic Neural Networks in a Bionformatics Task. Department of Mathematics, University of Patras Artificial Intelligence Research Center.
- Graupe D., Salahi J. y Kohn KH. (1982). Multifunctional Prosthesis and Orthosis Control Via Microcomputer identification of temporal pattern differences in single-site myoelectric signals. *J Biomed Eng*. Vol. 4, pp. 17 - 22.
- Graupe D. y Cline W. (1975). Functional separation of EMG signals via ARMA identification methods for prosthesis control purposes. *In IEEE Trans. Systems Man and Cybernetics*, volume 5, 252-259.

- Grossmann A. y Morlet J. (1984) Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal of Mathematical Analysis*, 15(4),pp. 723-736.
- Hargrove L., Englehart K. y Hudgins B. (2005). A Comparison of Surface and Intramuscular Myoelectric Signal Classification. *Proc. of the 27th Annual Conf. Int.* pag. 5009 - 5012. IEEE EMBS 2005.
- Haykin, S. (1994) *Neural Networks: A comprehensive Foundation*, Maxwell MacMillan Canada, Inc. Don Mill, Ontario, 1994.
- Huber, P.J. (1985) Projection Pursuit (with discution). *The Annal of Statistics*, Vol 13, No. 2, pp 435-475,1985.
- Hudgins B., Parker P. y Scott RN. (1993). A New Strategy for Multifunction Myoelectric Control. *IEEE Trans Biomed Eng.* Vol. 40, pp. 82 - 94.
- Kaiser G. (1994) *A Friendly Guide to Wavelets*. Birkhauser. Boston Basel Berlin. 1994.
- Kang W., Cheng C., Lai J., Shiu J., Kuo T. (1966). A Comparative Analysis of Various EMG Pattern Recognition Methods. *Med. Eng. Phys.* Vol. 18, pp. 390 - 395.
- Kilby J. y Hosseini G. (2004). Wavelet Analysis of Surface Electromyography Signals. Electrotechnology Department, Auckland University of Techology, New Zeland. Proc. Of 26th Annual Int. Conf. of the IEEE EMBS, San Francisco, CA.
- Kocyigit Y. y Korurek M. (2002). EMG Singal Classification using Wavelet Transform and Fuzzy Clustering Algorithms. Istambul Technical University, Department of Electronics and Communication Engineering, Ayazaga, Istambul, Turkey. 2002.
- Liu H., Chen X. y Chen Y. (2005). Wavelet Transform and Real - Time Learning Method for Myoelectric Signal in Motion Discrimination. 7th *International Symposium on Measurement Technology and Inteligent Instruments*. Conf. Ser. Vol. 13, pp. 25 - 253.
- Lowery M., Stoykov N., Taflove A., y Kuiken T. (2005). Tesis profesional - A Multiple-Layer Finite-Element Model of the Surface EMG Signal. *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 5, pp. 446-454, May 2005.

- Mackenzie D. (2001). Wavelets: ver el bosque y los árboles. Disponible en:<http://www7.nationalacademies.org/spanishbeyonddiscovery/mat_008276-01.html>. Acceso en 20 feb. 2005.
- Marques de Sa J.P. (2001). *Pattern Recognition: Concepts, Methods and Application*. Springer Verlag. Berlin Heidelberg, New York, 2001.
- Meyer, Y. (1993) *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, PA., Translated and Revised by R.D. Ryan, 1993.
- Mora, J. (2006). Ph.D. Tesis Doctoral. Localización de Faltas en Sistemas de Distribución de Energía Eléctrica usando Métodos Basados en el Modelo y Métodos de Clasificación Basados en el Conocimiento. *Universitat de Girona*, Girona, España.
- Nishikawa D., Yu W., Yokoi H., y Kakazu Y. (1999). EMG prosthetic hand controller discriminating ten motions using realtime learning method. *IEEE/RSJ IROS*. 1999.
- Papoulis A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1985.
- Reaz M., Hussain M. y Mohd-Yasin F. (2006). Techniques Of Emg Signal Analysis: Detection, Processing, Classification And Applications. *Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia*. Vol. 8, pp. 11 - 35.
- Reischl M., Milkut R., Pylatiuk C., Schulz S. (2006) Control Strategies For Hand Protheses Using Myoelectric Patterns. *Forschungszentrum Karlsruhe. Institute of Applied Computer Science*. Disponible en:<http://www.iai.fzk.de/raw/pages/english/projekte/biosignal/publikationen/control_zit01.pdf>. Acceso en 10 nov. 2006.
- Rengifo, Z. y Belalcazar C.(2008). Trabajo de grado. Clasificación de señales electromiográficas para una prótesis de mano. Universidad del Cauca, Facultad de Ingeniería Electrónica y Telecomunicaciones.Popayán.
- Rodriguez I. y Vuskovic M. (2005). Wavelet_ Based Feature Extraction From Prehensile EMG Singals. *Universidad Pública de Navarra, Pamplona, España*.

Disponible en: <http://medusa.sdsu.edu/Robotics/Neuromuscular/Our_publications/Sweedden2005/Sweden.pdf>. Acceso en 3 nov. 2006.

Rodriguez I. y Vuskovic M. (2006). Wavelet Transform Moments For Feature Extraction From Temporal Signals. *Universidad Pública de Navarra, Pamplona, España*. Disponible en: <http://medusa.sdsu.edu/Robotics/Neuromuscular/Our_publications/Barcelona_2005/Paper575-Revised.pdf>. Acceso en 3 nov. 2006.

Romo H., Gasca D., Rojas L. (2006). Desempeño de las Transformadas Wavelet DWT y WPT en la Extracción de Características Descriptoras a partir de señales EMGS. En el *XII Simposio de Tratamiento de Señales, Imágenes y Visión Artificial, XII STSIVA Sept 2007*. Universidad del Norte.

Rosa M y Gil R. (2006). Ph.D. Thesis. Sistemas de clasificación de blancos radar mediante métodos estadísticos y de inteligencia artificial. *Universidad de Alcalá. Departamento de Teoría de la Señal y Comunicaciones*. 2006.

Saito N., (1994). Ph D. Thesis. Local Feature Extraction and its Application Using a Library of Bases. *Dept. of Mathematics, Yale University, New Haven, CT USA*, December 1994.

Sijiang D. (2003). Thesis profesional. Feature Extraction For Classification of Per-hensile Electromyography patterns. *Master's Degree of Computer Science San Diego State University*. 2003.

Valls J. (2006). Qué es la Electromiografía y para que se utiliza. Disponible en: <<http://personales.ya.com/emgnm>>. Acceso en 3 jun. 2006.

Wang X. y Cui J. (2005). The Application of Wavelet Transform and Neural Network to Surface Electromyographic Signal for Pattern Recognition. *Engineering in Medicine and Biology. Proc. of the 27th Annual Int. Conf. Shanghai, China*. pag. 5009 - 5012. IEEE EMBS 2005.

Wickerhauser M. (1994). *Adapted Wavelet Analysis from Theory to Software*. AK Peter, Ltd. Wellesley, Massachusetts, 1994.

Yücel K. y Mehmet K. (2002). EMG Signal Classification Using Wavelet Transform And Fuzzy Clustering Algorithms. *Istanbul Technical University. Department*

of Electronics and Communication Engineering, Ayazaga, Istanbul, Turkey.
Disponible en:<<http://eleco.emo.org.tr/eleco2003/ELECO2003/bsession/B8-02.pdf>>. Acceso en 12 nov 2006.

Zecca M., Micera S., Carroza M. y Dario P. (2002). Control Of Multifunctional Prosthetic Hands By Processing The Electromyographic Signal. *Critical Reviews in Biomedical Engineering*. Vol. 30, pp. 459 - 485.

Zhang X., Yang Y., Xu X. y Zhang M. (2002). Wavelet Based Neuro-Fuzzy Classification for EMG Control. *Department of Automation, Shanghai Jiaotong University, Shanghai 200030, China*. Vol 2, pp. 1087 - 1089.

Zhang X., Yang Y., Xu X. Zhang J., Gao Z., Hu T., Chen T., Chen Z.(2004) Clinical Detection and Movement Recognition of Neuro Signal. *Journal of Zhenjian University SCIENCE*. pp. 272-279. 2004.