ANEXO B

CLASIFICACIÓN DE SEÑALES BIOELÉCTRICAS

OSCAR HERNÁN PARUMA PABÓN

EDGAR BOLIVAR MUÑOZ BURBANO

UNIVERSIDAD DEL CAUCA

FACULTAD DE INGENIERÍA ELECTRÓNICA

PROGRAMA DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES

DEPARTAMENTO DE ELECTRÓNICA, INSTRUMENTACIÓN Y CONTROL

POPAYÁN

2003

ANEXO B

CLASIFICACIÓN DE SEÑALES BIOELÉCTRICAS

OSCAR HERNÁN PARUMA PABÓN

EDGAR BOLIVAR MUÑOZ BURBANO

Director
JUAN FERNANDO FLOREZ MARULANDA
Ingeniero en Electrónica y Telecomunicaciones

UNIVERSIDAD DEL CAUCA

FACULTAD DE INGENIERÍA ELECTRÓNICA

PROGRAMA DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES

DEPARTAMENTO DE ELECTRÓNICA, INSTRUMENTACIÓN Y CONTROL

POPAYÁN

2003

TABLA DE CONTENIDO

pág.

B. CLASIFICACIÓN DE SEÑALES BIOELÉCTRICAS 3

B.1 CLASIFICACIÓN DE PATRONES MIOELÉCTRICOS 4

B.2 EXTRACCIÓN DE CARACTERÍSTICAS 6

B.2.1 Selección y proyección. 6

B.2.2Estimación de la fuerza. 7

B.2.3 Parámetros en el dominio del tiempo. 9

B.2.4 Información frecuencial. 11

B.2.5 Representaciones de la señal en tiempo y en frecuencia. 12

B.2.6 Evaluación de las características. 15

B.3 CLASIFICACIÓN DE CARACTERÍSTICAS 15

B.4 CLASIFICADOR ESTADÍSTICO 17

B.4.1 Proceso de diseño del clasificador estadístico. 17

B.4.2 Limitaciones del análisis discriminante lineal. 27

B.5 VALIDACIÓN DEL CLASIFICADOR 28

BIBLIOGRAFÍA 30

B. CLASIFICACIÓN DE SEÑALES BIOELÉCTRICAS

La clasificación de señales, su compresión y la eliminación de ruido en ellas son ejemplos de problemas clásicos en la teoría de la señal. El problema de la clasificación de señales bioeléctricas, cae dentro del dominio de la teoría de la señal, por lo que a continuación se presenta el problema de una manera más formal.

En este anexo se analiza la forma en la cual pueden ser extraídas las características mas importantes de estas señales y la información irrelevante ser descartada, todo ello enmarcado en el contexto de la clasificación de señales, campo de estudio que cubre áreas como la estadística, la ingeniería, la inteligencia artificial, las ciencias de la computación, la sicología y la fisiología.

La clasificación de señales involucra las siguientes fases:

- **Formulación del problema:** permite fijar los objetivos de la investigación y planear las fases siguientes.

- **Recolección de datos:** se realizan medidas de variables apropiadas y se almacenan sistemáticamente.

- **Procesamiento inicial de los datos:** se normalizan los datos para que queden ubicados dentro de un rango que facilite el trabajo computacional.

- **Extracción de características:** se seleccionan las variables de las medidas realizadas que sean apropiadas para el trabajo. Se pueden obtener nuevas variables a través de transformaciones lineales o no lineales de los datos originales.

- **Clasificación de patrones:** se aplica un procedimiento de discriminación en el que se compara el parámetro seleccionado del patrón de entrada con los parámetros de referencia establecidos con anterioridad (proceso de entrenamiento) para determinar la correspondencia del patrón con uno de los estados posibles.

- **Evaluación de los resultados:** El clasificador entrenado es probado con un conjunto de patrones de entrada.

- **Interpretación de los resultados:** con base en la clasificación del patrón de entrada se toman decisiones y se ejecutan acciones.

B.1 CLASIFICACIÓN DE PATRONES MIOELÉCTRICOS

Un sistema clasificador de señales realiza funciones de extracción de características y la posterior clasificación a partir de las características obtenidas.

Se define **patrón** como un vector de **N** variables $\vec{\mathbf{x}} = [\mathbf{x}(1), \mathbf{x}(2), ..., \mathbf{x}(\mathbf{N})]$. Estas variables son las **N** muestras consecutivas de la señal. Las muestras se almacenan durante un tiempo **T** = 250 ms debido a que durante este periodo la señal EMG presenta características determinísticas que facilitan su procesamiento, para una frecuencia de muestreo $\mathbf{f_s}$ = 1024 muestras por segundo se obtienen **N** = 256 muestras por patrón, por lo tanto, los patrones tienen una longitud de **N** = 256 muestras. Cada patrón puede asociarse a una de **K** clases posibles $\mathbf{y} = \mathbf{y_0}, \mathbf{y} = \mathbf{y_1}, ..., \mathbf{o}\ \mathbf{y} = \mathbf{y_K}$. Se considerará K = 3, en donde $\mathbf{y_0}$ será el estado de no actividad, $\mathbf{y_1}$ será el estado de contracción media y $\mathbf{y_2}$ será el estado de contracción fuerte.[3]

Sea el espacio de entrada **X** el conjunto de valores que puede tomar el vector patrón $\vec{\mathbf{x}}$ y sea el espacio de la señal de salida **Y** el conjunto de valores que pueda tomar $\vec{\mathbf{y}}$. El espacio **X** tiene dimensión **N** y el espacio **Y** tiene dimensión **K**.

En el control discreto, en el cual la señal mioeléctrica es clasificada en una de las **K** clases posibles definidas para modificar alguno de los estados de los dispositivos controlados, se impone a $\vec{\mathbf{y}}$ una restricción, y es que no puede ser una combinación lineal de los vectores de su espacio, sino que ha de tomar uno de los valores discretos que se proponen, llamados **clases**. Formalmente, si $\vec{\mathbf{x}} \in \mathbf{X} \subseteq \Re^N, \vec{\mathbf{y}} \in \mathbf{Y} = \{\vec{\mathbf{y}}_0, \vec{\mathbf{y}}_1, ..., \vec{\mathbf{y}}_\mathbf{K}\}$. De manera menos formal, se podrá escribir que las clases son $\mathbf{y_0}, \mathbf{y_1}, \ldots \mathbf{y_K}$ sin la notación vectorial.

Clasificar una señal $\vec{\mathbf{x}}$ es establecer una relación entre **X** e **Y**, que asigne a cada patrón de entrada $\vec{\mathbf{x}}$ una clase $\mathbf{y}$. La voluntad del usuario se traduce en unos patrones a partir de los cuales el clasificador debe determinar su clase de salida.

El **rendimiento del clasificador** hace referencia al porcentaje de aciertos. Un porcentaje de aciertos inferior al 80% se considera deficiente. Evidentemente, basta con reducir el

tamaño **K** (es decir, el número de estados) para que el rendimiento aumente, pero esta solución no es la adecuada.

Establecer una relación directa de los datos de entrada con el espacio de la señal de salida no es razonable. Se podría clasificar la salida **y** como función de las **N** componentes de $\vec{\mathbf{x}}$, pero teniendo en cuenta que **N** = 256 en el sistema propuesto, esta opción se debe descartar.

El espacio de entrada es realmente grande para ser tratado directamente. Se considerará un número de datos como grande si no es posible procesarlo en tiempo real, pero también se considerará grande si han de proporcionársele demasiados patrones de entrenamiento de manera supervisada, tantos como para cansar al ser humano que tenga que generarlos.

Por todo lo anterior se define un espacio de características **F** de dimensión **M**, entre el espacio de la señal de entrada y el espacio de la señal de salida, donde **M** < **N**.

Un **extractor de características** se define como el sistema que establece la relación de **X** a **F**, y el **clasificador** como el que lo hace entre **F** e **Y**, aunque por extensión a todo el proceso se le conocerá como “clasificador”.

Se contará con una colección de datos de entrenamiento, que consiste en **P** pares de señales de entrada y de clases de salida $(\vec{\mathbf{x}}, \mathbf{y}): (\vec{\mathbf{x}}_1, \mathbf{y}_1), (\vec{\mathbf{x}}_2, \mathbf{y}_1), (\vec{\mathbf{x}}_3, \mathbf{y}_2), \ldots$. Con el fin de evaluar la generalización de la capacidad de un clasificador, se asume que esta colección de datos proviene de un modelo probabilístico, y que se trata de realizaciones de procesos estocásticos. En la práctica serán más interesantes los pares de entrenamiento $(\vec{\mathbf{f}}, \mathbf{y})$ ya que el clasificador habrá de operar sobre las características del espacio **F**.

Se define el **número de canales** como **L** si hay **L** amplificadores de instrumentación. Así, en un momento dado se contará en realidad con los patrones $\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \ldots, \vec{\mathbf{x}}_L$, donde los subíndices denotan el canal del que procede el patrón. El número considerado de canales es **L** = 1 a través del cual se introducirá la señal EMG captada de músculos en el antebrazo, en el bíceps o el tríceps. No se recomienda usar demasiados canales ya que esto implica aumentar también el número de electrodos lo cual resulta incómodo para el usuario.

Para que el sistema opere en tiempo real y entregue resultados satisfactorios, el clasificador no puede decidir a partir del patrón de entrada directamente, sino que debe hacerlo a partir de un conjunto de características representativas de dicho patrón.

B.2 EXTRACCIÓN DE CARACTERÍSTICAS

Extraer características significa reducir el patrón de dimensión N a otro de dimensión M sin perder información significativa. El problema es que el patrón está mezclado con el ruido y al elegir un conjunto de características de tamaño $M < N$ se perderá irremediablemente parte de la información útil para clasificar los patrones. Escoger el conjunto óptimo de características F es un aspecto muy importante que depende, en gran medida, del clasificador seleccionado. Las características que son óptimas para un clasificador pueden no serlo para otro. El objetivo es extraer las características que mejor preserven la separabilidad de las clases para un clasificador dado.

Para empezar, se asume que el clasificador de referencia es el bayesiano; de modo que maximizar la separabilidad es minimizar el error bayesiano. El vector de características $\vec{f}$ idóneo en el caso bayesiano son las $K-1$ probabilidades a posteriori $P(y_k \mid \vec{x})$. Son $K-1$ componentes y no K porque la última se deduce al instante sabiendo que todas han de sumar la unidad. Así que la mejor medida de la separabilidad de dos clases es el error bayesiano que arroja su clasificación; pero como no se conocen las probabilidades a posteriori esto sólo será una medida teórica, y el error se calculará experimentalmente o se darán medidas de la separabilidad de las clases. En el caso de que las características se extrajeran como combinación lineal de las muestras del patrón, y dado un criterio de separabilidad de clases, sería muy sencillo extraer de manera automática la combinación óptima de muestras que mejor separabilidad presente. Sin embargo los parámetros significativos que representan la señal no están linealmente relacionados con sus muestras. Y no existe un método para extraer características no lineales de un patrón de manera sistemática, de modo que la extracción de características es algo que depende de la naturaleza de cada problema, y es preciso tantear con las distintas posibilidades. Si se conociera de manera determinística las ecuaciones que rigen la señal EMG, bastaría con conocer los parámetros, pero dichas ecuaciones no son conocidas.

En esta sección se mencionan algunas características que pueden ser empleadas para clasificar los patrones mioeléctricos. Se postula un método que demuestra que la relación entre la señal EMG y la fuerza está representada por la varianza lo cual justifica su escogencia cómo el parámetro para controlar interfaces hombre – máquina en el caso particular de este proyecto. Se indican algunas características que se pueden extraer a partir de la representación temporal y a partir de la representación frecuencial de la señal. Finalmente se exploran los métodos en tiempo - frecuencia. Además se explican los conceptos de selección y de proyección de características.

B.2.1 Selección y proyección. El conjunto de características que se extrae inicialmente se puede dejar tal cual está o bien puede ser manipulado para mejorar la separabilidad de las clases. Por ejemplo, si se toma como características a la varianza y al número de cruces por cero de la señal, pueden dejarse en el vector $\vec{f}$ (o mejor dicho, $\vec{x}$) estas dos

características tal cual están (entonces el procedimiento se llama selección de características) o bien se pueden combinar de manera que se mejore la distinción de clases (se habla entonces de proyección de características).

La proyección de características se aplica sobre un conjunto de características ya elegido, de modo que la selección ha de realizarse en todo caso, no se puede eludir; y la proyección no es más que una mejora. La proyección puede ser, por ejemplo, ponderar de diversas maneras las componentes del vector $\vec{x}$, comprobar el error del clasificador para cada caso, y escoger la combinación óptima. Por lo tanto, el procedimiento es escoger un conjunto de características, tan amplio como sea posible, quedarse con un subconjunto (selección) y después combinar los más relevantes (proyección). La proyección es una mejora importante sobre la selección si no se tienen en cuenta las covarianzas, ya que puede hacer no correlacionadas a las características, además tiene la ventaja que es un método no supervisado.

B.2.2 Estimación de la fuerza. Las primeras características que se presentaron como base para el clasificador fueron estimaciones de la fuerza ya que todos los movimientos se producen con la fuerza de los músculos y no con otro parámetro. En este proyecto, la fuerza se considera como una característica especialmente significativa a partir de la cual se realizará la clasificación de las señales EMG detectadas.

B.2.2.1La fuerza se manifiesta en la varianza. Para demostrar que la fuerza se manifiesta en la varianza, el modelo clásico supone a la señal $\boldsymbol{x}(\boldsymbol{t})$ como un proceso aleatorio gaussiano de media cero y varianza $\sigma(\boldsymbol{F}(\boldsymbol{t}))$, donde $\boldsymbol{F}(\boldsymbol{t})$ es una medida de la fuerza del usuario. Ello es aceptable bajo la consideración de que la señal es gaussiana por la suma de otros muchos procesos aleatorios con la misma media y varianza y no necesariamente la misma distribución. El sistema $\boldsymbol{H}$ representa el filtrado que modela al ruido blanco gaussiano de media cero. Ese filtro es dependiente de los tejidos y de la posición de los electrodos. El sistema $\sigma(\boldsymbol{F})$ oculta la relación entre la fuerza y la amplitud de la señal EMG. Es decir, se representa el origen de la señal como un ruido blanco gaussiano pasado por un filtro desconocido, y modulado en amplitud por una desconocida función $\sigma(\boldsymbol{F})$. Nada hace pensar que la fuerza haya de manifestarse en el parámetro varianza de la señal y no en cualquier otro, como por ejemplo la frecuencia de la señal, por lo tanto se debe realizar la demostración.

La actividad mioeléctrica es un proceso no estacionario, ya que la fuerza del músculo varía, pero considerando períodos suficientemente cortos no va a ser relevante la no estacionareidad. La función de densidad de probabilidad de una muestra es normal:

$$\boldsymbol{p}(\boldsymbol{x}(\boldsymbol{n})) = \frac{1}{\sqrt{2\pi}\sigma(\boldsymbol{F})} \exp\left[-\frac{\boldsymbol{x}^2(\boldsymbol{n})}{2\sigma^2(\boldsymbol{F})}\right] \qquad \text{(B.1)}$$

La dependencia temporal entre muestras sucesivas queda determinada por la función de autocorrelación, o, equivalentemente, por la densidad espectral de potencia. La densidad espectral de potencia de la señal es:

$$\boldsymbol{S}_{\boldsymbol{x}}(\boldsymbol{f}) = |\boldsymbol{H}(\boldsymbol{f})|^2 \, \sigma^2(\boldsymbol{F}) \tag{B.2}$$

A partir de la función de autocorrelación (o si se prefiere, la densidad espectral de potencia) y la función de densidad de probabilidad de cada muestra descrita por $\boldsymbol{p}(\boldsymbol{x}(\boldsymbol{n}))$ se debe poder extraer la función de similitud para un patrón, $\boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{F}) = \boldsymbol{p}(\boldsymbol{x}(1), \boldsymbol{x}(2), ..., \boldsymbol{x}(\boldsymbol{N}) \mid \boldsymbol{F})$. Esa es la probabilidad de que, dada una fuerza $\boldsymbol{F}$, la señal sea $\vec{\boldsymbol{x}}$. Pero lo que se desea saber es: dada una señal $\vec{\boldsymbol{x}}$, que es la que se ha medido, cuál es la fuerza estimada $\hat{\boldsymbol{F}}$ para la cual la probabilidad es mayor. Se trata, de maximizar $\boldsymbol{p}(\vec{\boldsymbol{x}} \mid \hat{\boldsymbol{F}})$. Esta vez no se acude al teorema de Bayes, y para maximizar se deriva y se iguala a cero:

$$\frac{\boldsymbol{d}}{\boldsymbol{dF}} \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{F})\Big|_{\boldsymbol{F}=\hat{\boldsymbol{F}};\ \vec{\boldsymbol{x}}=\vec{\boldsymbol{x}}_{\boldsymbol{ob}}} = 0 \tag{B.3}$$

Tomando logaritmo:

$$\frac{\boldsymbol{d}}{\boldsymbol{dF}} [\ln \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{F})] = 0 \tag{B.4}$$

No se puede suponer que todas las muestras son no correlacionadas y que las probabilidades de cada muestra son independientes así que para aceptar la hipótesis, se ha de aplicar un filtro de blanqueo. Un filtro de blanqueo es aquel cuya función de transferencia es inversa al espectro de la señal que ha de pasar por ella. Así, tras el filtro de blanqueado, la señal presenta un espectro plano, y una correlación nula entre dos muestras distintas cualquiera. Entonces, una vez blanqueada la señal, la probabilidad conjunta queda como una sumatoria de probabilidades:

$$\boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{F}) = \sum_{\boldsymbol{n}=1}^{\boldsymbol{N}} \boldsymbol{p}(\boldsymbol{x}(\boldsymbol{n}) \mid \boldsymbol{F}) \tag{B.5}$$

A partir de $\boldsymbol{p}(\boldsymbol{x}(\boldsymbol{n}))$ y $\boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{F})$, se tiene:

$$\ln \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{F}) = \sum_{\boldsymbol{n}=1}^{\boldsymbol{N}} \ln \boldsymbol{p}(\boldsymbol{x}(\boldsymbol{n}) \mid \boldsymbol{F}) \tag{B.6}$$

es decir:

$$\ln \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{F}) = -\boldsymbol{N} \ln \sqrt{2\pi} - \boldsymbol{N} \ln \sigma(\boldsymbol{F}) - \frac{1}{2} \sigma^{-2}(\boldsymbol{F}) \sum_{\boldsymbol{n}=1}^{\boldsymbol{N}} \boldsymbol{x}^2(\boldsymbol{n}) \tag{B.7}$$

Como la derivada era cero:

$$\frac{\boldsymbol{d}}{\boldsymbol{dF}}\left[\ln\ \boldsymbol{p}\left(\vec{\boldsymbol{x}}\mid \boldsymbol{F}\right)\right]=-\boldsymbol{N}\sigma'\left(\boldsymbol{F}\right)\sigma^{-1}\left(\boldsymbol{F}\right)+\sigma'\left(\boldsymbol{F}\right)\sigma^{-3}\left(\boldsymbol{F}\right)\sum_{\boldsymbol{n}=1}^{\boldsymbol{N}}\boldsymbol{x}^{2}\left(\boldsymbol{n}\right)=0 \qquad \text{(B.8)}$$

Donde las primas representan derivación respecto a $\boldsymbol{F}$. Sacando factor común:

$$\sigma'\left(\boldsymbol{F}\right)\sigma^{-2}\left(\boldsymbol{F}\right)\left(\sigma^{-1}\left(\boldsymbol{F}\right)\sum_{\boldsymbol{n}=1}^{\boldsymbol{N}}\boldsymbol{x}^{2}\left(\boldsymbol{n}\right)-\boldsymbol{N}\,\sigma\left(\boldsymbol{F}\right)\right)=0 \qquad \text{(B.9)}$$

Por lo tanto:

$$\sigma^{2}\left(\boldsymbol{F}\right)=\frac{\sum_{\boldsymbol{n}=1}^{\boldsymbol{N}}\boldsymbol{x}^{2}\left(\boldsymbol{n}\right)}{\boldsymbol{N}} \qquad \text{(B.10)}$$

Si se denota la fuerza como estimación que es, y se representa la inversa de la función desconocida $\boldsymbol{g}=\sigma\left(\boldsymbol{F}\right)$ como $\boldsymbol{F}=\sigma^{-1}\left(\boldsymbol{g}\right)$:

$$\hat{\boldsymbol{F}}=\sigma^{-1}\left[\sqrt{\frac{1}{\boldsymbol{N}}\sum_{\boldsymbol{n}=1}^{\boldsymbol{N}}\boldsymbol{x}^{2}\left(\boldsymbol{n}\right)}\right] \qquad \text{(B.11)}$$

Lo que está entre corchetes es la varianza de las muestras observadas. Es decir, que la fuerza se manifiesta únicamente a través de la varianza de la señal. La fuerza no se podía manifestar en parámetros frecuenciales porque la salida del filtro blanqueador es 'blanca', y como la función de distribución de probabilidad era gaussiana por la hipótesis de que hay muchos procesos aleatorios, entonces sólo esta última distribución iba a caracterizar la señal. Y la distribución normal queda completamente determinada por su media y su varianza. Siendo la media cero para toda señal EMG, el único parámetro en que podía manifestarse la fuerza era en la varianza. A pesar de que la relación entre varianza y fuerza es desconocida, una vez realizada esta demostración será tarea sencilla tomar datos experimentales y ajustar la relación. Este resultado teórico es muy importante: una vez blanqueada la señal, la fuerza $\boldsymbol{F}$ puede estimarse a partir de la desviación típica de la señal. Se deben tomar las muestras, calcular su desviación típica y de ahí inferir la relación (por ejemplo, se ha tomado a menudo la relación $\sigma\left(\boldsymbol{F}\right)=\boldsymbol{F}^{\,1{,}75}$).

B.2.3 Parámetros en el dominio del tiempo. Aplicando algunas operaciones matemáticas sobre las muestras obtenidas, es posible calcular los siguientes parámetros en el dominio temporal:

- **Valor absoluto medio (MAV – Mean Absolute Value).** También es llamado IAV (Integral Absolute Value). Se calcula como:

$$MAV = \frac{1}{N}\sum_{n=1}^{N}|x(n)| \tag{B.12}$$

- **Varianza.**

$$VAR = \sigma^2 = \frac{1}{N}\sum_{n=1}^{N}x^2(n) \tag{B.13}$$

Si la función de densidad de probabilidad de las amplitudes de las muestras fuera gaussiana, debería cumplirse que:

$$VAR = \sqrt{\frac{\pi}{2}}MAV \tag{B.14}$$

- **Momentos de orden superior.**

$$\sigma^3 = \frac{1}{N}\sum_{n=1}^{N}|x(n)|^3 \tag{B.15}$$

e igualmente para ordenes superiores hasta orden 5.

- **Longitud de la señal.**

$$WL = \frac{1}{N}\sum_{n=1}^{N}(x(n) - x(n-1)) \tag{B.16}$$

- **Amplitud Willison.**

$$WAMP = \sum_{n=1}^{N} f\left(|x(n) - x(n+1)|\right) \tag{2.17}$$

con

$$f(x)=\begin{cases}1 & \textit{si } x > \textit{umbral}\\ 0 & \textit{en otro caso.}\end{cases}$$

B.2.4 Información frecuencial. En este apartado se mencionan los cruces por cero, la frecuencia mediana, los momentos frecuenciales de orden superior y la transformada de Fourier como orígenes de características susceptibles de ser seleccionadas.

- **Cruces por cero (ZC – Zero Crossings).**

$$ZC=\frac{1}{N}\sum_{n=1}^{N}\operatorname{sgn}\left(-x(n)\,x(n+1)\right) \qquad \text{(B.18)}$$

En presencia de ruido, éste y otros parámetros que representen la frecuencia introducen un cierto error. La frecuencia media presenta mayor solidez ante el ruido que el número de cruces con cero y la frecuencia mediana.

- **Frecuencia mediana.** El valor mediano de la frecuencia puede ser usado como característica y se calcula por métodos analógicos o mediante la Transformada Rápida de Fourier (FFT).

- **Momentos frecuenciales de orden superior.** Si la frecuencia media es el momento de frecuencia de orden 1, también los momentos superiores pueden aportar información, y también se han usado para el análisis de la señal mioeléctrica. Pueden expresarse como:

$$M_k=\frac{\int_0^F f^k P(f)\,df}{\int_0^F P(f)\,df} \qquad \text{(B.19)}$$

- **Transformada de Fourier.** La transformada de Fourier ofrece información sobre el patrón mioeléctrico, pero esta información no es fiable por no ser estacionaria la señal. La transformada de Fourier continua se define como:

$$X(f)=\int_{-\infty}^{+\infty}x(t)\,e^{-j2\pi f t}\,dt \qquad \text{(B.20)}$$

Como no se va a trabajar con funciones continuas, sino discretas se podría pensar en hallar la DFT (transformada de Fourier discreta) de cada patrón:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} \qquad \text{(B.21)}$$

Los coeficientes de la DFT no sirven tal cual están como características del patrón pues son tan numerosas como las muestras de la señal. Sin embargo, se puede pensar en aplicar algún tipo de reducción más eficaz a partir de la DFT. El número de operaciones para realizar una DFT es del orden de N^2, es decir, unas 65000. Si la DFT se realiza sobre segmentos de un número potencia de 2, entonces es aplicable un algoritmo que reduce notablemente las operaciones a realizar, la FFT. Con la FFT sólo hace falta una cantidad aproximada de $N \log_2 N$ operaciones, es decir, de 2000. 50000 operaciones está rozando el límite de lo razonable para que los equipos actuales pudieran trabajar en tiempo real.

Estas transformaciones tiene sentido aplicarlas si las señales son estacionarias, si no evolucionan en el tiempo. En caso de que varíen en el tiempo, carecerían de sentido. Si se aplica una FFT sobre un patrón en el que al principio se estaba en reposo y al final ya hay una contracción, el resultado tendrá características extrañas. Tomando patrones suficientemente cortos, habrá un gran número de ellos que sean válidos y sólo algunos serán no válidos. Pero hay mejores maneras de solucionar este problema.

B.2.5 Representaciones de la señal en tiempo y en frecuencia. Las representaciones de la señal en tiempo - frecuencia combinan el análisis de la señal en ambos dominios, temporal y frecuencial. Ha sido utilizada en casi todos los campos del procesado de la señal, como el de compresión de la señal, la codificación, el filtrado, la eliminación de ruido, o los problemas de detección, clasificación y visualización. En este proyecto es de interés porque es una estupenda caracterización de la señal que facilitará la clasificación de los patrones mioeléctricos.

Las representaciones en tiempo - frecuencia se dividen en métodos lineales (como la STFT y la transformada Wavelet) y los cuadráticos (distribución Wigner-Ville). Es una teoría complicada y extensa que desborda los propósitos de este proyecto, pero que no puede dejar de ser mencionada como tendencia más importante en la extracción de características para la clasificación de patrones mioeléctricos.

- **Transformada STFT.** Hacer una STFT (Short Time Fourier Transform) de un patrón equivale a hacer DFTs en pequeños trocitos del patrón tras ser enventanados. Aquí se verá cómo y porqué utilizar esta transformada.

Los coeficientes de la transformada discreta de Fourier, los $X(k)$, dan la distribución de la señal en el dominio de la frecuencia en toda la señal, sin dar resolución temporal, y las

muestras $x(n)$ de la señal dan información sólo en el dominio temporal sin dar información frecuencial. La ventaja de la STFT es que da información de la señal tanto el dominio temporal como en el frecuencial. La STFT considera a la señal como estacionaria sólo durante intervalos limitados de tiempo, en esto supera a la DFT, y se define como:

$$STFT(t, f) = \int x(\tau)\, g^*(\tau - t)\, e^{-2j\pi f \tau} d\tau \qquad \text{(B.22)}$$

La STFT direcciona la señal en un plano bidimensional t y f, aplicando una transformada de Fourier a los pedacitos tras un enventanado. Hay dos parámetros en los que se debe centrar la atención, que son la resolución temporal Δt y la resolución frecuencial Δf. La resolución indica con qué precisión se conocerá la información en un dominio o en otro; la mejora de una resolución implica el empeoramiento de la otra. Si se toman fragmentos grandes de la señal en el tiempo, la DFT tendrá más puntos y arrojará más precisión, pero a costa de hacer la partición en el tiempo más gruesa; y viceversa, si se toman fragmentos pequeños de la señal en el tiempo dará sólo una tosca DFT. Las resoluciones dependen de la **función ventana** $g(t)$. Dada una función ventana $g(t)$ y su transformada de Fourier $G(f)$ se puede tomar como medida de la resolución frecuencial al ancho de banda medio cuadrado:

$$\Delta f^2 = \frac{\int f^2 |G(f)|^2 df}{\int |G(f)|^2 df} \qquad \text{(B.23)}$$

y la resolución temporal es el ancho cuadrático medio de la señal en el dominio del tiempo:

$$\Delta t^2 = \frac{\int t^2 |g(t)|^2 dt}{\int |g(t)|^2 dt} \qquad \text{(B.24)}$$

Una u otra resolución pueden ser arbitrariamente pequeñas, pero el producto de ambas resoluciones está limitado por el principio de incertidumbre (desigualdad de Heisemberg):

$$\Delta f \; \Delta t \geq \frac{1}{4\pi} \qquad \text{(B.25)}$$

La ventana gaussiana es la que consigue llegar a la igualdad, siendo el mejor caso. En este caso particular, la transformada STFT se denomina transformada Gabor. La ventana que se elija, en todo caso, debe de poder ser calculada con rapidez. Ahora se va a escribir la STFT de una manera más conveniente. La DFT puede ser reescrita como:

$$X(m) = \sum_{i=1}^{L-1} x[i]\, e^{-2j\pi (mF)(iT_S)} \qquad \text{(B.26)}$$

donde T_s y f_s son el período y la frecuencia de muestreo, respectivamente. $F = \frac{1}{LT_s}$ es el tamaño del paso de la frecuencia de muestreo. Así la STFT queda como una serie de DFTs:

$$STFT(k, m) = STFT(kT_s, mF) = \sum_{i=1}^{L-1} x[i]\, g[i-k]\, e^{-2j\pi mi/L} \qquad \text{(B.27)}$$

$T = KT_s$ es el tamaño del paso del muestreo temporal, el equivalente en el tiempo a F. Ahora sí se puede escribir la STFT en función de estos dos parámetros fundamentales, los pasos de muestreo temporal - frecuenciales. Se escribe como:

$$STFT(k, m) = STFT(nT, mF) = \sum_{i=1}^{L-1} x[iT_s]\, g[iT_s - nT]\, e^{-2j\pi mi/L} \qquad \text{(B.28)}$$

La STFT tiene muchas propiedades útiles incluyendo una teoría desarrollada bien conocida, una interpretación sencilla y una buena eficiencia computacional. Pero tiene el fallo de que la rejilla espacio temporal es fija.

Se define espectrograma como la magnitud cuadrada de la STFT. Da información de la energía, mucho mejor para ser representada ya que no incluye partes real e imaginaria.

Transformada Wavelet. La transformada Wavelet (WT) no tiene fija las resoluciones Δt y Δf en el plano tiempo - frecuencia, como la STFT. Para analizar la señal EMG, en un momento dado quizás se pueda suponer que la contracción es constante o el reposo es prolongado, de modo que la resolución temporal interese poco y la frecuencial pueda ser afinada más. Y en otros momentos, quizá sepamos que se está realizando un movimiento, se está iniciando una contracción y la señal varía su estructura con rapidez. Entonces lo adecuado sería reducir al mínimo el tamaño de las ventanas sobre las que analizar la frecuencia. Esta flexibilidad sólo la puede proporcionar la transformada Wavelet y sus familiares. Desde que los computadores tienen capacidad para calcular en tiempo real la transformada Wavelet de patrones mioeléctricos, y siendo ésta una herramienta más potente, ha sido utilizada para diagnóstico médico (por ejemplo se ha usado para analizar la fatiga muscular) y para control.

La transformada Wavelet continua (CWT) cubre el espacio tiempo – frecuencia de una manera variable y se define como:

$$CWT(\tau, a) = \frac{1}{\sqrt{a}} \int x(t)\, \psi^*\left(\frac{t-\tau}{a}\right) \qquad \text{(B.29)}$$

Donde ψ es una función ventana prototipo que se llama función Wavelet madre. El análisis determina la correlación de la señal con versiones de la señal escaladas por a y desplazadas en τ.

La implementación discreta de la CWT se puede calcular directamente convolucionando la señal con una versión escalada y dilatada de la Wavelet madre. En la versión discreta, $\boldsymbol{a} = \boldsymbol{a}_0^{\boldsymbol{j}}$ y $\tau = \boldsymbol{n}\boldsymbol{a}_0^{-\boldsymbol{j}}$ con $\boldsymbol{j}$ y $\boldsymbol{n}$ enteros; y entonces ya no se llama CWT sino DWT (transformada Wavelet discreta). Escoger $\boldsymbol{a} = 1$ y $\tau = \boldsymbol{n}\, 2^{-\boldsymbol{j}}$ equivale al caso continuo; lo habitual es elegir $\boldsymbol{a} = 2^{\boldsymbol{j}}$ y $\tau = \boldsymbol{n}\, 2^{-\boldsymbol{j}}$ (base diádica). Cuando se habla de transformada Wavelet, se sobreentiende que se habla de la CWT con base diádica.

La transformada Wavelet packet es una versión generalizada de la CWT y de la DWT. La transformada es redundante y permite elegir muchas bases ortonormales simultáneamente, de manera que la división en tiempo – frecuencia es configurable, la partición se ajusta a las necesidades de la señal.

B.2.6 Evaluación de las características. Para evaluar el grado de acierto en la elección de características un buen método sería comparar las distintas tasas de errores que originan en la salida del clasificador según alguno de los métodos ya comentados. Sin embargo, la opción más común es aprovechar alguna medida de la separabilidad de las clases que ya esté bien establecida, como es la distancia de Bhattacharyya.

Distancia de Bhattacharyya. La distancia de Bhattacharyya es un caso particular de la distancia de Chernoff. La distancia entre dos clases $\boldsymbol{k}$ y $\boldsymbol{j}$ se define como:

$$\mu(1/2) = \frac{1}{8}(\boldsymbol{m}_{\boldsymbol{k}} - \boldsymbol{m}_{\boldsymbol{j}})^{T}\left(\frac{\boldsymbol{C}_{\boldsymbol{k}} + \boldsymbol{C}_{\boldsymbol{j}}}{2}\right)^{-1}(\boldsymbol{m}_{\boldsymbol{k}} - \boldsymbol{m}_{\boldsymbol{j}}) + \frac{1}{2}\ln\frac{\left|\dfrac{\boldsymbol{C}_{\boldsymbol{k}} + \boldsymbol{C}_{\boldsymbol{j}}}{2}\right|}{\sqrt{|\boldsymbol{C}_{\boldsymbol{k}}||\boldsymbol{C}_{\boldsymbol{j}}|}} \qquad \text{(B.30)}$$

La distancia de Bhattacharyya ha sido ampliamente utilizada para la medida de la separación de las distribuciones de los patrones mioeléctricos.

B.3 CLASIFICACIÓN DE CARACTERÍSTICAS

La teoría de clasificación distingue entre tres tipos de clasificadores. En primer lugar se encuentra la aproximación estadística (o teoría de la decisión), a continuación se tiene la aproximación estructural (sintáctica) y finalmente se encuentra la aproximación basada en aprendizaje (redes neuronales).

- **Aproximación estadística.** Se basa en el análisis estadístico de los datos a ser clasificados. A partir de la colección de datos de entrenamiento se intenta inferir la función de densidad de probabilidad para cada clase. Todos los patrones que se sabe que pertenecen a una clase (porque son patrones de entrenamiento y por lo tanto son conocidos) no son más que realizaciones de un proceso estocástico. A partir de esto, según algún criterio predeterminado, se divide el espacio de la señal de entrada **X**. Todo patrón $\vec{\mathbf{x}}$ que se desee clasificar, será asignado a una clase $\mathbf{y}$ u otra según pertenezca a una región del espacio o a otra. Hay varios criterios para dividir el espacio de la señal de entrada, pero en general se distinguirá entre clasificadores paramétricos y clasificadores no paramétricos. Los clasificadores paramétricos asumen una función de densidad de probabilidad parametrizada para los datos, mientras que los no paramétricos no preasumen forma alguna de la función. Debido a que este es el método seleccionado para realizar la clasificación de características en este proyecto, se profundizará más en esta temática en el siguiente apartado.

- **Aproximación sintáctica.** Está basada en utilizar la estructura de los patrones y la interrelación entre las componentes de un patrón. El reconocimiento sintáctico de patrones implica identificar componentes significativas o primitivas de los patrones y desarrollar una sintaxis formal o gramática describiendo la síntesis de los patrones a partir de sus primitivas. Se puede realizar un paralelismo con la teoría del lenguaje en donde primitivas = palabras, señales = sentencias, descripción estructural = gramática. Se debe tener en cuenta que lo que es natural para los humanos es muy complejo de procesar en las máquinas y es difícil programar analizadores sintácticos frente a los métodos estadísticos que se prestan perfectamente a la computación. El problema realmente difícil es extraer las características de la señal que permitan este análisis, la clasificación luego debería ser más sencilla. Tampoco está claro que la señal esté sujeta a un tipo de estructura, por lo que estos métodos de reconocimiento de patrones no son en principio muy apropiados en este sistema y se comentan con el fin de dar generalidad a las soluciones presentadas. Esta aproximación tan sólo se utiliza en técnicas aisladas de procesado de la señal de vıdeo y similares, donde hay formas bien delimitadas y bordes que han de ser reconocidos. Un clasificador de aproximación sintáctica puede ser implementado utilizando lógica difusa.

- **Aproximación basada en aprendizaje.** Los algoritmos capaces de “aprender” toman la forma de redes neuronales. Las Redes Neuronales Artificiales son un método determinístico porque, en oposición a los métodos estadísticos, los algoritmos de aprendizaje no asumen nada acerca de las propiedades estadísticas de los patrones del espacio de la señal de entrada.

En el presente proyecto, como clasificador de características se utilizará un clasificador estadístico debido a su facilidad de implementación, a las ventajas que presenta en cuanto a procesamiento computacional y a los resultados que entrega.

B.4 CLASIFICADOR ESTADÍSTICO

Inicialmente se debe escoger un grupo de características $\vec{f}$ que representen a la señal. La entrada del clasificador está representada por estas características $\vec{f}$ y el clasificador es el bloque que entrega una salida y a partir de una serie de características $\vec{f}$. El clasificador no actúa sobre el espacio de entrada sino sobre un extracto del mismo.

B.4.1 Proceso de diseño del clasificador estadístico. Existe un proceso bien establecido para diseñar el clasificador [3]. En el diseño del clasificador se deben seguir los siguientes pasos:

- **Estimación no paramétrica del error sobre el espacio de entrada X.** En primer lugar, se recogen los datos y las muestras son normalizadas. Después se mide la separabilidad de clases entre los datos recogidos mediante la estimación del error de Bayes en el espacio de la señal de entrada. En principio no se asume forma matemática alguna para la estructura de los datos, por lo que la estimación debe ser no paramétrica. Si el error de Bayes es mayor que el error del clasificador que se desea hacer, entonces es mejor no seguir; el extractor de características y el clasificador empeorarán aun más la situación.

- **Análisis de la estructura de datos y estimación del error sobre el espacio de características.** Si de momento el error es tolerable, se procede a extraer características, asignar grupos y hacer análisis estadísticos. Cada vez que se extraiga un conjunto de características se estimará el error de Bayes sobre él y se comparará con el error de Bayes sobre el espacio de la señal de entrada. La diferencia entre ambos da una medida de la bondad del extractor de características, procediéndose a su rediseño si tal diferencia no fuera tolerable.

- **Diseño del clasificador y error final.** Una vez que se ha entendido la estructura de los datos, se escoge un clasificador. El clasificador normalmente es paramétrico, que siendo más sencillo es más apropiado para operar en tiempo real. Finalmente se compara el error final con el error de Bayes sobre el espacio de características; y si la diferencia es inaceptablemente alta debe rediseñarse el clasificador.

B.4.1.1 Aproximación bayesiana. El problema en el reconocimiento de patrones estadísticos es hallar las funciones de decisión óptimas, que se acerquen más al rendimiento del 100 % y que hagan coincidir la voluntad del usuario con la respuesta del clasificador. Con el clasificador bayesiano se consigue maximizar el rendimiento.

Las medidas de $\vec{x}$ e y son consideradas en un marco probabilístico en esta aproximación y son vistas como observaciones de las variables aleatorias **X** e **Y**. Lo que se desearía conocer es la probabilidad condicionada de que conociendo un patrón observado $\vec{x}$ pertenezca a la clase y_K. Procediendo así con todas las clases, se escogerá la que arroje una probabilidad de acierto mayor. Las probabilidades condicionadas a posteriori para cada clase se denotan como

$$P(y_K \mid \vec{x}), k = 1,2,..., K \tag{B.31}$$

Dichas probabilidades no se conocen, pero a partir de un conjunto de pares de entrenamientos se pueden inferir.

La aproximación bayesiana exige que las funciones de densidad de probabilidad sean constantes en el tiempo. Esta hipótesis no es aceptada en cierto sentido por las aproximaciones sintácticas, donde la relación entre entrada y salida no es unívoca y un mismo patrón podría ser asignado a clases distintas en momentos diferentes, o por la aproximación de aprendizaje donde el algoritmo se podría adaptar a las nuevas situaciones. En la práctica, las funciones de densidad de probabilidad no son constantes. La aproximación bayesiana por definición maximiza el rendimiento, pero las hipótesis de partida no se verifican y por tanto no se puede garantizar ese máximo en la realidad. En todo caso, aunque permanecieran constantes, no se dispone de la colección completa de pares de entrenamiento, con lo que el conocimiento de las funciones de densidad de probabilidad tampoco sería perfecto. En principio se supondrá que las funciones de densidad de probabilidad son las mismas durante la operación del sistema y durante el entrenamiento. Las probabilidades a posteriori $P(y_K \mid \vec{x})$ no se pueden estimar directamente. Esto se hará a partir de las probabilidades a priori $P(y_K)$ y la función de densidad $p(\vec{x} \mid y_k)$ usando el teorema de Bayes:

$$P(y_K \mid \vec{x}) = \frac{P(y_k)p(\vec{x} \mid y_K)}{p(\vec{x})} \tag{B.32}$$

donde $p(\vec{x})$ es la función de densidad de probabilidad de los vectores del espacio de entrada. Se calcula como:

$$p(\vec{x}) = \sum_{j=1}^{K} P(y_j)p(\vec{x} \mid y_j) \tag{B.33}$$

Dado que permanece constante para todas las $P(y_K \mid \vec{x})$, puede ser ignorado para efectos de discriminación y simplificación.

Si se desea saber si un patrón pertenece a una clase $k = 1$ o a una clase $k = 2$, se suele recurrir a la **tasa de similitud**, que es:

$$\ell(\boldsymbol{x}) = \frac{\boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_1)}{\boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_2)} \tag{B.34}$$

y se compara con el cociente $\boldsymbol{P}(\boldsymbol{y}_2)/\boldsymbol{P}(\boldsymbol{y}_1)$. Más convenientemente se puede definir la tasa menos-logarítmica de similitud como $\boldsymbol{h}(\vec{\boldsymbol{x}}) = -\ln \ell(\vec{\boldsymbol{x}})$. Esta $\boldsymbol{h}(\vec{\boldsymbol{x}})$ es la función discriminante.

Si se desea realizar la clasificación entre más clases, el criterio habitual para dos clases ya no sirve. Si se extiende a más variables, se puede tomar como función discriminante a:

$$\boldsymbol{h}_{\boldsymbol{k}}(\vec{\boldsymbol{x}}) = \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_{\boldsymbol{k}})\boldsymbol{P}(\boldsymbol{y}_{\boldsymbol{k}}), \boldsymbol{k} = 1, 2, ..., \boldsymbol{K}. \tag{B.35}$$

y el criterio para decidir es escoger la $\boldsymbol{h}_{\boldsymbol{k}}(\vec{\boldsymbol{x}})$ mayor.

El criterio bayesiano garantiza que se ha minimizado la probabilidad de error. Esa probabilidad de error se llama **error bayesiano** y está dado por:

$$\varepsilon = \sum_{\boldsymbol{k}=1}^{\boldsymbol{K}} \boldsymbol{P}(\boldsymbol{y}_{\boldsymbol{k}})\varepsilon_{\boldsymbol{k}} \tag{B.36}$$

con

$$\varepsilon_{\boldsymbol{k}} = \int_{\boldsymbol{L}_{\boldsymbol{k}}} \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_{\boldsymbol{k}})\boldsymbol{dx} = \sum_{\vec{\boldsymbol{x}} \subseteq \boldsymbol{L}_{\boldsymbol{k}}} \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_{\boldsymbol{k}}) \tag{B.37}$$

donde $\boldsymbol{L}_{\boldsymbol{k}}$ indica la región de $\boldsymbol{x}$ para la cual se debería haber escogido $\boldsymbol{y}_{\boldsymbol{k}}$ y no se hizo.

Cuando no se conoce la distribución de probabilidad de las clases, las probabilidades a priori se hacen iguales: $\boldsymbol{P}(\boldsymbol{y}_{\boldsymbol{k}}) = 1/\boldsymbol{K}$ con $\boldsymbol{k} = 1, 2, ..., \boldsymbol{K}$ y entonces la función discriminante se reduce a conocer $\boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_{\boldsymbol{k}})$.

No debería ser muy difícil estimar las $\boldsymbol{P}(\boldsymbol{y}_{\boldsymbol{k}})$ observando las acciones más frecuentes que realiza el usuario y observar con qué frecuencia invoca a una u otra función de la interfaz. Pero en general, y dado que no es posible suponer que el usuario realizará más frecuentemente una acción que otra se va a suponer que las clases son equiprobables.

B.4.1.2Clasificadores gaussianos bayesianos. El clasificador gaussiano bayesiano propone que las funciones de densidad de probabilidad condicional $p(\vec{x} \mid y_k)$ son normales multivariadas, y que la observación de algunos pares de entrenamiento sirve para fijar los parámetros que las modelan, es decir, el vector de medias y la matriz de covarianza. Aunque para algunas colecciones de datos es difícil hacer estas suposiciones, la distribución normal es muy apropiada en la práctica para señales EMG como resultado del teorema del límite central el cual dice que si $x_1, x_2, ..., x_n$ son variables aleatorias (discretas o continuas), con idéntico modelo de probabilidad de valor medio μ y varianza σ^2 entonces la distribución de la variable

$$Z = \frac{(x_1 + x_2 + ... + x_n) - n\mu}{\sigma\sqrt{n}} = \frac{\left(\sum_{i=1}^{n} x_i\right) - n\mu}{\sigma\sqrt{n}} \qquad \text{(B.38)}$$

se aproxima a la de una variable normal tipificada con media 0 y varianza 1, $N(0,1)$, mejorándose la calidad de la aproximación a medida que n aumenta.

Formalmente la función es:

$$p(\vec{x} \mid y_k) = \frac{1}{2\pi^{N/2}|C_k|^{1/2}} \exp\left[-\frac{(\vec{x} - \vec{m}_k)^T C_k^{-1} (\vec{x} - \vec{m}_k)}{2}\right], k = 1, 2, ..., K \qquad \text{(B.39)}$$

que queda completamente determinada con el vector de medias $\vec{m}_k$ y la matriz de covarianza C_k, definidos por:

$$\vec{m}_k = E_k[\vec{x}] \qquad \text{(B.40)}$$

y

$$C_k = E\left[(\vec{x} - \vec{m}_k)(\vec{x} - \vec{m}_k)^T\right] \qquad \text{(B.41)}$$

donde $E_k[.]$ denota el operador esperanza (o valor esperado) sobre los patrones de clase y_k y $|C_k|$ representa el determinante de la matriz C_k. La covarianza entre dos variables $x_k(i)$ y $x_k(j)$ expresa su tendencia a variar conjuntamente, y está acotado entre $-\sigma_k(i)\sigma_k(j)$ y $+\sigma_k(i)\sigma_k(j)$, tomando el valor 0 cuando las variables aleatorias son independientes.

La matriz de covarianza es la matriz que contiene todas las relaciones entre las características del patrón mioeléctrico, e indicará su grado de independencia o su grado de correlación. Al multiplicar por la matriz de covarianza se está haciendo una transformación lineal con el fin de normalizar. En definitiva, conceptualmente no es más que una función de densidad de probabilidad gaussiana para cada clase definida, sólo que en vez de ser función de una única variable, es función de un vector de variables, la media es sustituida por un vector de medias y la varianza es sustituida por la matriz de covarianza. La media y la covarianza deben estimarse a partir de los patrones de entrenamiento. Teniendo $\boldsymbol{P}$ patrones de entrenamiento en una clase, se podría tomar como estimador de la media de esa clase a:

$$\vec{\boldsymbol{m}}_{\boldsymbol{k}} = \frac{1}{\boldsymbol{P}} \sum_{\boldsymbol{p}=1}^{\boldsymbol{P}} \vec{\boldsymbol{x}}_{\boldsymbol{k}\,\boldsymbol{p}} \tag{B.42}$$

a partir del cual se estimaría la covarianza. La varianza de la estimación decrece con el número de patrones de ejemplo, entre mayor sea el número de muestras mejor es la estimación.

La matriz de covarianza se puede calcular como:

$$\boldsymbol{C}_{\boldsymbol{k}} = \frac{\sum_{\boldsymbol{p}=1}^{\boldsymbol{P}} \left(\vec{\boldsymbol{x}}_{\boldsymbol{pk}} - \vec{\boldsymbol{m}}_{\boldsymbol{k}}\right)\left(\vec{\boldsymbol{x}}_{\boldsymbol{pk}} - \vec{\boldsymbol{m}}_{\boldsymbol{k}}\right)^{\boldsymbol{T}}}{\boldsymbol{P}-1} \tag{B.43}$$

Hay que tener mucho cuidado al decidir el número de patrones $\boldsymbol{P}$ de entrenamiento. $\vec{\boldsymbol{x}}$ y $\vec{\boldsymbol{m}}_{\boldsymbol{k}}$ son vectores de $\boldsymbol{N}$ componentes pero $\boldsymbol{C}_{\boldsymbol{k}}$ es una matriz de $\frac{\boldsymbol{N}\times\boldsymbol{N}}{2}$ componentes independientes. Para que la estimación de $\boldsymbol{C}_{\boldsymbol{k}}$ sea buena son necesarios al menos $\boldsymbol{P} = \frac{\boldsymbol{N}\times\boldsymbol{N}}{2}$ patrones de entrenamiento y no $\boldsymbol{N}$ como se podría pensar. Eso aumenta el gasto computacional y en cuanto crezca $\boldsymbol{N}$ aumentará el número de patrones de entrenamiento necesarios. Si no se extrajeran características del patrón $\boldsymbol{N}$ y se consideraran las 256 componentes, harían falta como mínimo unos 32000 patrones de entrenamiento por cada clase y el entrenamiento supondría muchas horas.

La aproximación bayesiana toma como criterio el cálculo para cada clase $\boldsymbol{K}$ de $\boldsymbol{h}_{\boldsymbol{k}}(\vec{\boldsymbol{x}}) = \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_{\boldsymbol{k}})\boldsymbol{P}(\boldsymbol{y}_{\boldsymbol{k}})$ y ahora se ha supuesto $\boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_{\boldsymbol{k}})$ como gaussiano. Para hacerlo más manejable, se utiliza la versión menos-logarítmica de las funciones:

$$\boldsymbol{h}_{\boldsymbol{k}}(\vec{\boldsymbol{x}}) = \ln \boldsymbol{p}(\vec{\boldsymbol{x}} \mid \boldsymbol{y}_{\boldsymbol{k}}) + \ln \boldsymbol{P}(\boldsymbol{y}_{\boldsymbol{k}}) \tag{B.44}$$

Con lo que la función discriminante queda como:

$$\boldsymbol{h_k}(\vec{\boldsymbol{x}}) = \ln \boldsymbol{P}(\boldsymbol{y_k}) - \frac{\boldsymbol{N}}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{C_k}| - \frac{1}{2} \left[(\vec{\boldsymbol{x}} - \vec{\boldsymbol{m}}_{\boldsymbol{k}})^{\boldsymbol{T}} \boldsymbol{C_k}^{-1} (\vec{\boldsymbol{x}} - \vec{\boldsymbol{m}}_{\boldsymbol{k}}) \right] \qquad \text{(B.45)}$$

El término $\frac{\boldsymbol{N}}{2} \ln 2\pi$ es igual para cada clase, de modo que no ofrece discriminación y puede ser eliminado. Esta función discriminante $\boldsymbol{h_k}$ será calculada para cada clase, escogiendo como más viable la función que sea mayor (o escogiendo la función menor cuando se multiplique a la función por -1). Es posible realizar simplificaciones en esta función, obteniendo expresiones cada vez más sencillas.

- **Función discriminante cuadrática.** $\boldsymbol{P}(\boldsymbol{y_k})$ se tomará constante para todas las clases. Multiplicando también a todas las clases por -2, se obtiene la función discriminante QDF (Quadratic Discriminant Function):

$$\boldsymbol{h_k}(\vec{\boldsymbol{x}}) = \ln |\boldsymbol{C_k}| + (\vec{\boldsymbol{x}} - \vec{\boldsymbol{m}}_{\boldsymbol{k}})^{\boldsymbol{T}} \boldsymbol{C_k}^{-1} (\vec{\boldsymbol{x}} - \vec{\boldsymbol{m}}_{\boldsymbol{k}}), \boldsymbol{k} = 1, 2, ..., \boldsymbol{K} \qquad \text{(B.46)}$$

La superficie que divide el espacio de entrada es la que minimiza las probabilidades de error y es una superficie cuadrática (paraboloide, elipsoide o hiperboloide).

Como se ha multiplicado por un número negativo, al evaluar todas las funciones discriminantes debe escogerse la menor y no la mayor como antes. Por eso en vez de funciones discriminantes se puede hablar de distancias como sinónimo.

- **Función discriminante de Mahalanobis.** El segundo término de la función QDF por sí solo se utiliza también como distancia entre un patrón y la distribución que caracteriza a un grupo, llamándose distancia de Mahalanobis:

$$\boldsymbol{h_k^2}(\vec{\boldsymbol{x}}) = (\vec{\boldsymbol{x}} - \vec{\boldsymbol{m}}_{\boldsymbol{k}})^{\boldsymbol{T}} \boldsymbol{C_k}^{-1} (\vec{\boldsymbol{x}} - \vec{\boldsymbol{m}}_{\boldsymbol{k}}), \boldsymbol{k} = 1, 2, ..., \boldsymbol{K} \qquad \text{(B.47)}$$

Lo único que se ha hecho es eliminar un término de energía de la QDF que al fin y al cabo no dependía del patrón. La distancia de Mahalanobis tiene la ventaja de que todos los parámetros quedan normalizados, no importando la escala en que estén (es posible que unos parámetros del patrón estén en microvoltios y otras en voltios, por ejemplo.) y no importando la correlación de las componentes. Ello es debido a que se conserva la transformación lineal normalizadora que supone multiplicar por la covarianza.

- **Función discriminante lineal.** Realizando las operaciones de producto en la función QDF se obtiene:

$$h_k(\vec{x}) = \ln|C_k| + \vec{x}^T C^{-1}\vec{x} - 2\vec{x}^T C^{-1}\vec{m}_k + \vec{m}_k^T C^{-1}\vec{m}_k,\ k = 1, 2, ..., K \qquad \text{(B.48)}$$

Si además se asume que la matriz de covarianza es igual para todas las clases $C_k = C$ para $k = 1, 2, ..., K$ entonces los dos primeros términos son iguales para todas las clases. Se tiene por lo tanto la función discriminante LDF (Linear Discriminant Function).

$$h_k(\vec{x}) = \vec{m}_k^T C^{-1}\vec{m}_k - 2\vec{x}^T C^{-1}\vec{m}_k,\ k = 1, 2, ..., K \qquad \text{(B.49)}$$

Se ha llegado a una colección de funciones discriminantes que son lineales en el término $\vec{x}$, Cuando se utiliza el clasificador normal (o gaussiano) bayesiano de esta forma, se dice que se está realizando un análisis discriminante lineal (LDA). Las superficies que dividen el espacio de entrada ahora son planos (hiperplanos), y se puede demostrar que siguen siendo divisiones óptimas incluso para otras funciones de distribución no gaussianas. También puede interpretarse como una medida de la correlación entre el patrón $\vec{x}$ y el patrón representante de la clase $\vec{m}_k$, al que se le debe sumar un término de energía. El análisis discriminante lineal es muy sencillo de interpretar y de implementar y se entrena bien con un número reducido de patrones de entrenamiento, así que será la opción seleccionada para ser implementada en este proyecto.

- **Función discriminante de distancia euclídea.** La función de clasificación de distancia es quizás el método más simple e intuitivo para solucionar el problema. El uso de funciones de distancia como herramientas de clasificación se entiende fácilmente a partir de la noción de que la similitud es una medida de la proximidad. Los resultados que ofrece este clasificador son buenos sólo cuando las clases tienden a estar bien agrupadas, lo cual sucede de una manera muy aproximada en el problema que se trata. Este clasificador se llamará clasificador de mínima distancia. La distancia euclídea entre un vector $\vec{x}$ y uno $\vec{x}_k$ es:

$$d_k(\vec{x})^2 = \| \vec{x} - \vec{x}_k \|^2 = (\vec{x} - \vec{x}_k)^T (\vec{x} - \vec{x}_k) \qquad \text{(B.50)}$$

donde $\vec{x}_k$ es el patrón representante de la clase $\vec{y}_k$.

Un clasificador de mínima distancia calcula la distancia d_k de un patrón desconocido al prototipo de cada clase y le asigna el patrón a la clase más próxima. La distancia no tiene por que ser necesariamente euclídea, puede ser de otro orden distinto de 2. Para aplicar eficientemente la distancia euclídea, conviene que las componentes del patrón $\vec{x}$ estén normalizadas, que estén en el mismo rango. Para ello se puede pensar en escalar cada componente $x(n)$ de $\vec{x}$:

$$x(n) = \frac{x(n) - m_k(n)}{\sigma_k(n)} \qquad \text{(B.51)}$$

El prototipo, el representante de cada clase, es único y en principio será el vector medio de los vectores de cada clase que sirvieron de entrenamiento, es decir, el estimador de la media de la clase. Si es así, y si la distancia es la euclídea, se puede demostrar que el clasificador de mínima distancia no es sino un caso particular del clasificador bayesiano.

Retomando la función discriminante QDF:

$$\boldsymbol{h_k}\left(\vec{\boldsymbol{x}}\right)=\ln\left|\boldsymbol{C_k}\right|+\left(\vec{\boldsymbol{x}}-\vec{\boldsymbol{m}}_{\boldsymbol{k}}\right)^{\boldsymbol{T}}\boldsymbol{C_k}^{-1}\left(\vec{\boldsymbol{x}}-\vec{\boldsymbol{m}}_{\boldsymbol{k}}\right),\ \boldsymbol{k}=1,\ 2,\ ...,\ \boldsymbol{K} \qquad \text{(B.52)}$$

A la suposición de que la distribución de probabilidad es normal en el criterio bayesiano gaussiano, se le añade ahora otra hipótesis: que las características sean mutuamente no correlacionadas, es decir, que las componentes de $\vec{\boldsymbol{x}}$ sean independientes entre sí. En ese caso la matriz de covarianza es una matriz diagonal. Y así

$$\boldsymbol{d_k}\left(\vec{\boldsymbol{x}}\right)=\ln\left(\prod_{\boldsymbol{n}=1}^{\boldsymbol{N}}\sigma_{\boldsymbol{k}}^{2}\left(\boldsymbol{n}\right)\right)+\sum_{\boldsymbol{n}=1}^{\boldsymbol{N}}\left[\frac{\left(\boldsymbol{x}\left(\boldsymbol{n}\right)-\boldsymbol{m_k}\left(\boldsymbol{n}\right)\right)^2}{\sigma_{\boldsymbol{k}}^{2}\left(\boldsymbol{n}\right)}\right] \qquad \text{(B.53)}$$

donde $\sigma_{\boldsymbol{k}}\left(\boldsymbol{n}\right)$ y $\boldsymbol{m_k}\left(\boldsymbol{n}\right)$ son, respectivamente, la desviación típica y la media de la componente n-ésima del patrón de la clase $\boldsymbol{k}$.

Si además el primer término, el productorio, es el mismo para todas las clases, se puede eliminar del mismo modo que se eliminó al pasar de la función QDF a la de Mahalanobis. De modo que si la varianza es la misma en todas las componentes de cada clase, entonces se tiene la distancia euclídea $\boldsymbol{d_k}\left(\vec{\boldsymbol{x}}\right)=\left\|\ \vec{\boldsymbol{x}}-\vec{\boldsymbol{x}}_{\boldsymbol{k}}\ \right\|$.

No se pierde eficacia respecto del clasificador bayesiano gaussiano siempre y cuando las varianzas de cada característica sean aproximadamente iguales. Si en el vector de características se incluyen términos con varianzas muy dispares, como es lo más común, el sistema funcionará muy mal a no ser que se normalicen todos los datos. Este clasificador es lineal al igual que en análisis LDA, y el espacio de la señal de entrada queda también fraccionado por hiperplanos.

2.4.1.3 Realización práctica del LDF. El análisis discriminante lineal es el elegido en el sistema que se propone por su sencillez. Para aplicar la función discriminante lineal a los patrones mioeléctricos se realiza el procedimiento que se describe a continuación, el cual permite obtener una expresión de fácil implementación computacional para la función $\boldsymbol{h_k}\left(\vec{\boldsymbol{x}}\right)$**.**

La función discriminante lineal era:

$$\boldsymbol{h_k}\left(\vec{\boldsymbol{x}}\right) = \vec{\boldsymbol{m}}_{\boldsymbol{k}}^{\boldsymbol{T}}\,\boldsymbol{C}^{-1}\,\vec{\boldsymbol{m}}_{\boldsymbol{k}} - 2\,\vec{\boldsymbol{x}}^{\boldsymbol{T}}\,\boldsymbol{C}^{-1}\,\vec{\boldsymbol{m}}_{\boldsymbol{k}},\ \boldsymbol{k} = 1, 2, ..., \boldsymbol{K} \qquad \text{(B.54)}$$

De manera general se ha obtenido la regla de decisión QDF, que se convertía en la regla de decisión lineal LDF con tan sólo solicitar que la matriz de covarianza fuera igual para todas las clases. En este sistema ya se ha comentado que las matrices de covarianza no son iguales para los distintos estados, y sin embargo sería deseable poder aplicar un clasificador lineal. En la realización práctica se reescribe la función discriminante lineal como:

$$\boldsymbol{h_k}\left(\vec{\boldsymbol{x}}\right) = \boldsymbol{V}^{\boldsymbol{T}}\,\vec{\boldsymbol{x}} + \boldsymbol{v}_0 \qquad \text{(B.55)}$$

Comparando la ecuación anterior con la ecuación LDF se tiene que:

$$\boldsymbol{V} = -2\,\boldsymbol{C}^{-1}\vec{\boldsymbol{m}}_{\boldsymbol{k}} \qquad \text{(B.56)}$$

y

$$\boldsymbol{v}_0 = \vec{\boldsymbol{m}}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\vec{\boldsymbol{m}}_{\boldsymbol{k}} \qquad \text{(B.57)}$$

Lo anterior no es válido porque las matrices de covarianza no son iguales, y el problema es que se ignora qué $\boldsymbol{V}$ y que $\boldsymbol{v}_0$ tomar. El patrón $\boldsymbol{N}$ - dimensional $\vec{\boldsymbol{x}}$ es proyectado en la dirección de un vector $\boldsymbol{V}$ para dar un escalar $\boldsymbol{h}$ y $\boldsymbol{v}_0$ será el umbral para decidir una u otra clase. Si $\boldsymbol{X}$ tiene una función de distribución normal, entonces $\boldsymbol{h}$ también. La esperanza de $\boldsymbol{h_k}\left(\vec{\boldsymbol{x}}\right)$ es:

$$\eta_{\boldsymbol{h_k}} = \boldsymbol{E}\left[\boldsymbol{h_k}\left(\vec{\boldsymbol{x}}\right)\right] = \boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{m_k} + \boldsymbol{v}_0 \qquad \text{(B.58)}$$

y

$$\sigma_{\boldsymbol{h_k}}^2 = \boldsymbol{Var}\left[\boldsymbol{h_k}\left(\vec{\boldsymbol{x}}\right)\right] = \boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{C_k}\boldsymbol{V} \qquad \text{(B.59)}$$

- **Cálculo de $\boldsymbol{V}$ y $\boldsymbol{v}_0$ para el clasificador lineal entre dos clases.** Se tratará de maximizar la separabilidad entre sólo dos clases $\boldsymbol{k} = 1$ y $\boldsymbol{k} = 2$, la extensión a un clasificador de más de dos clases no será sencilla pero tomará como referencia este caso simple que no se puede por tanto eludir. Sea $\boldsymbol{f}\left(\eta_0, \sigma_0, \eta_1, \sigma_1\right)$ la función que se desea maximizar. Derivando $\boldsymbol{f}$ respecto de $\boldsymbol{V}$ y $\boldsymbol{v}_0$ e igualando a 0 se habrá obtenido el máximo, que se demuestra que se da cuando:

$$\boldsymbol{V} = \left[\boldsymbol{s}\boldsymbol{C}_1 + \left(1 - \boldsymbol{s}\right)\boldsymbol{C}_2\right]^{-1}\left(\vec{\boldsymbol{m}}_2 - \vec{\boldsymbol{m}}_1\right) \qquad \text{(B.60)}$$

donde

$$\boldsymbol{s} = \frac{\partial \boldsymbol{f} / \sigma_1^2}{\partial \boldsymbol{f} / \sigma_1^2 + \partial \boldsymbol{f} / \sigma_2^2 +} \tag{B.61}$$

Una vez conocido $\boldsymbol{f}$, $\boldsymbol{v}_0$ se puede calcular a partir de la expresión:

$$\frac{\partial \boldsymbol{f}}{\partial \eta_1} + \frac{\partial \boldsymbol{f}}{\partial \eta_2} = 0 \tag{B.62}$$

Para determinar qué $\boldsymbol{f}$ se debe tomar como medida de separabilidad se tienen dos aproximaciones:

- **Criterio de Fisher.** El criterio de Fisher postula como función a maximizar la siguiente:

$$\boldsymbol{f} = \frac{(\eta_1 - \eta_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{B.63}$$

Operando como se ha descrito, se tiene:

$$\boldsymbol{V} = \left[\frac{\boldsymbol{C}_1 + \boldsymbol{C}_2}{2} \right]^{-1} (\vec{\boldsymbol{m}}_2 - \vec{\boldsymbol{m}}_1) \tag{B.64}$$

Se observa que en el criterio no interviene $\boldsymbol{v}_0$, dado que al restar $\eta_1 - \eta_2$ se elimina en la ecuación de $\eta_{\boldsymbol{h_k}}$. Con tan sólo introducir a $\boldsymbol{V}$ en la ecuación $\boldsymbol{h_k}(\vec{\boldsymbol{x}}) = \boldsymbol{V}^T \vec{\boldsymbol{x}} + \boldsymbol{v}_0$, se obtiene el clasificador lineal de Fisher.

- **Dispersión entre clases.** Se toma a $\boldsymbol{f}$ como la dispersión entre clases normalizada por la dispersión dentro de la clase. Es decir:

$$\boldsymbol{f} = \frac{\boldsymbol{P}(\boldsymbol{y}_1)\eta_1^2 + \boldsymbol{P}(\boldsymbol{y}_2)\eta_2^2}{\boldsymbol{P}(\boldsymbol{y}_1)\sigma_1^2 + \boldsymbol{P}(\boldsymbol{y}_2)\sigma_2^2} \tag{B.65}$$

Aplicando las ecuaciones anteriormente descritas, se llega a que:

$$\boldsymbol{V} = \left[\boldsymbol{P}(\boldsymbol{y}_1)\boldsymbol{C}_1 + \boldsymbol{P}(\boldsymbol{y}_2)\boldsymbol{C}_2\right]^{-1}(\vec{\boldsymbol{m}}_2 - \vec{\boldsymbol{m}}_1) \tag{B.66}$$

y

$$\boldsymbol{v}_0 = -\boldsymbol{V}^T\left[\boldsymbol{P}(\boldsymbol{y}_1)\vec{\boldsymbol{m}}_1 + \boldsymbol{P}(\boldsymbol{y}_2)\vec{\boldsymbol{m}}_2\right] \tag{B.67}$$

Cálculo de V y v_0 para el clasificador lineal multiclase. En este caso, que es el que se usará en el sistema, es demasiado complicado llegar a una solución analítica, y se considerarán dos alternativas:

Promedio. Se puede tomar la ecuación LDF y considerar la matriz de covarianza que debiera ser igual para todas las clases como la media de las matrices de covarianza. El método tan sólo es razonablemente válido si las matrices de covarianza no son muy distintas. Parece una solución sencilla y apropiada al problema al que se enfrenta este proyecto. Es tan sólo calcular $h_k(\vec{x})$ para todas las clases y escoger la mayor:

$$h_k(\vec{x}) = 2\,\vec{m}_k \left(\frac{\sum_{i=1}^{k} C_i}{K} \right)^{-1} \vec{x} - \vec{m}_k^T \left(\frac{\sum_{i=1}^{k} C_i}{K} \right)^{-1} \vec{m}_k \qquad \text{(B.68)}$$

Ajuste de coeficientes. Este método es realmente muy complicado de implementar y no se sugiere su uso, salvo cuando las matrices de covarianza sean muy distintas. En el caso de los patrones mioeléctricos no debería ser necesario implementarlo. El método consiste en hallar las funciones discriminantes lineales para cada par de clases. Es decir, calcular h_{ij} para todo i y para todo j, hallar cada vector V_{ij}^T mediante los métodos expuestos anteriormente para el caso particular de separación entre dos clases. Entonces, el espacio queda dividido por múltiples rectas divisoras y un patrón dado s pertenece a la clase k si en todas y cada una de las combinaciones en las que está implicado sale vencedor. Una de las desventajas del método es que no todos los puntos del espacio de entrada caen en un sitio al que se pueda asignar con seguridad una clase (este espacio se llama región de rechazo). Un refinamiento sería retirar la condición de que el punto se asigna a una clase si es la mejor en todas las combinaciones, e imponer una condición menos restrictiva.

B.4.2 Limitaciones del análisis discriminante lineal. Para que la regla de decisión del discriminante lineal funcione bien, las hipótesis que se han aplicado (funciones de distribución gaussianas, clases equiprobables, matrices de covarianza iguales para cada clase) deberían ser ciertas. Las limitaciones se presentan si hay alguno de estos casos:

Desigualdad de las matrices de covarianza. La hipótesis de que las matrices de covarianza son iguales en todas las clases es bastante mala en el caso que ocupa este proyecto. Un análisis a simple vista de los patrones mioeléctricos revela que la dispersión de los datos no es la misma para cada clase. En general, niveles bajos de contracción muscular implican más agrupamiento, y esfuerzos mayores implican una dispersión mucho mayor. Esta hipótesis, que tan sólo se cumple de una manera burda, ha sido utilizada para llegar a la expresión más sencilla posible.

- **Alta correlación entre las componentes de $\vec{x}$.** O una diferencia de escala muy grande entre las características de X. La solución pasa por emplear el discriminante de Mahalanobis o bien transformar los patrones. Con ese problema se ha de contar en las características que se extraigan de los patrones mioeléctricos.

- **La función de distribución de X no es gaussiana.** Si toma una disposición curva, el discriminante lineal será malo. Puede disminuirse usando el discriminante de Mahalanobis que emplea superficies cuadráticas para dividir el espacio, pero si ello también fracasa habría que recurrir a los algoritmos de aprendizaje, a las redes neuronales.

- **La distribución se agrupa en varios subgrupos.** En ese caso también habría que recurrir al uso de redes neuronales, aunque es un caso que no debería presentarse con las características extraídas de los patrones mioeléctricos.

- **Distribución demasiado compleja.** Si responde a una estructura compleja, es mucho mejor recurrir a un método no estadístico sino sintáctico.

B.5 VALIDACIÓN DEL CLASIFICADOR

Una vez diseñado el clasificador, se debe comprobar cuál es su grado de eficiencia. Probar la eficiencia es ver qué tal clasifica una serie de patrones cuya clase es conocida. Para ello se proponen los siguientes métodos:

- **Prueba sobre los datos de entrenamiento.** Si el clasificador no es capaz de clasificar bien ni siquiera los datos que sirvieron para entrenarle, evidentemente se debe empezar de nuevo.

- **Método holdout.** De los patrones recogidos para el entrenamiento, sólo algunos se emplean para entrenar al clasificador, y se reservan los otros para comprobarlo. El inconveniente es que hay que tomar más patrones de entrenamiento de los normales, o bien desaprovechar los que ya se tienen.

- **Método leave-one-out.** El clasificador se entrena con todos los patrones disponibles excepto con uno, que sirve de comprobación. A continuación se vuelve a incluir al patrón separado, y se excluye a otro distinto para hacer otra comprobación; y así se procede sucesivamente con todos los patrones de entrenamiento. Este método además es una buena cota superior del error bayesiano.

- **Método bootstrapping.** Se toma aleatoriamente alguno de los patrones de entrenamiento para realizar la validación.

- **Método de varios usuarios.** Partiendo de los parámetros obtenidos a partir del entrenamiento del sistema por parte de un primer usuario, se prueba el clasificador para diferentes usuarios sin realizar nuevos entrenamientos y se determina el porcentaje de aciertos en la clasificación de las señales EMG dentro de los tres posibles estados (reposo, contracción media o contracción fuerte) dependiendo de la voluntad de cada persona. Este método de prueba, planteado por los autores de este trabajo, es el que se emplea para validar el clasificador estadístico implementado para el sistema.

BIBLIOGRAFÍA

[1] ALONSO, A.; HORNERO, R.; ESPINO, P.; DE LA ROSA, R.; LIPTAK, L. Entrenador Mioeléctrico de Prótesis para Amputados de Brazo y Mano. Universidad de Valladolid. 2001.

[2] ENGLEHART, Kevin. Signal Representation for Classification of the Transient Myoelectric Signal. Doctoral Thesis. University of New Brunswick (Canada). Octubre de 1998.

[3] RODRÍGUEZ, Víctor. Entrenador para el Control de Prótesis Mioeléctricas. Universidad de Valladolid. Diciembre de 2001.