

**UNIVERSITAT POLITÈCNICA DE CATALUNYA**

*Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial*

**ANALISIS DE LA ACTIVIDAD  
MUSCULAR RESPIRATORIA  
MEDIANTE TECNICAS TEMPORALES,  
FRECUENCIALES Y ESTADISTICAS**

Autor: Miguel Angel Mañanas Villanueva  
Director: Pere Caminal Magrans

Juny de 1999

# Capítulo 7

## Análisis multivariable en pacientes con EPOC y grupo control

### 7.1 Introducción

En el presente capítulo se determina una metodología de análisis estadístico de datos en el ámbito de la ingeniería biomédica y su implementación en el caso de tests respiratorios para evaluar la actividad muscular respiratoria en pacientes con EPOC y personas sanas [Mañanas y otros, 1996(b)] [Serena y otros, 1999]. Se evalúa diversas técnicas estadísticas multivariadas, y se escoge y ordena las más apropiadas para aplicar a los datos obtenidos de la caracterización de las señales miográficas en los test respiratorios.

Mediante este análisis se pretende interpretar el comportamiento del músculo, y diferenciar el correspondiente a una persona sana y un enfermo con problemas respiratorios [Mañanas y otros, 1996]. De esta manera se obtiene información que sirve de ayuda al diagnóstico médico. Además, debido a la larga duración de las pruebas respiratorias, la cantidad de parámetros calculados es considerable. Se pretende reducir dicho número de datos determinando los de mayor importancia para el análisis, y reducir el número de pruebas respiratorias a realizar por el individuo con el correspondiente ahorro de tiempo, dinero y recursos.

El análisis que se realiza se engloba dentro de la estadística inferencial. Uno de los objetivos es utilizar las técnicas de análisis multivariable de la varianza para medidas repetidas como una técnica exploratoria para seleccionar las variables de interés. Esto permite abordar los problemas de trabajar con un elevado número de parámetros y de medidas repetidas de dichos parámetros, así como la gran variabilidad presentada por los datos analizados. A continuación, con las variables escogidas se realiza un análisis factorial con el doble objetivo de explorar las relaciones entre ellas, y reducir su número. Por último, se realiza un análisis discriminante, como técnica confirmatoria, del que se obtiene una función discriminante de ayuda al diagnóstico.

## 7.2 Definición de la metodología

El método propuesto de análisis de datos se divide en tres fases con los siguientes objetivos:

### 1. Análisis multivariable de la varianza de medidas repetidas:

- Cuantificar e interpretar la evolución del comportamiento del músculo ECM durante los diferentes protocolos.
- Reducir el número de variables obtenidas eliminando información redundante y seleccionando parámetros de interés a extraer de las señales biomédicas.
- Generar y escoger las variables que mejor diferencian los sujetos sanos de los enfermos con EPOC y las que mejor explican la función muscular durante la prueba.

### 2. Análisis factorial de componentes principales:

- Explorar de forma global las relaciones entre las variables escogidas en el análisis multivariable de la varianza (MANOVA) de medidas repetidas.
- Encontrar combinaciones lineales óptimas para reducir el número de datos manteniendo la máxima información.
- Creación de nuevos factores con una alta interpretabilidad funcional del músculo durante la prueba respiratoria.

### 3. Análisis discriminante:

- Encontrar funciones lineales que separen grupos de individuos sanos y pacientes con EPOC, así como pacientes con bajo y alto nivel de obstrucción respiratoria.
- Evaluar las componentes que más discriminan entre poblaciones y, por tanto, conocer qué efectos o instantes del protocolo son más determinantes en dicha discriminación.
- Analizar las pruebas de mayor interés e importancia en la separación de poblaciones y, por consiguiente, reducir el número de tests a realizar por el sujeto.

## 7.3 Análisis multivariable de la varianza de medidas repetidas

### 7.3.1 Introducción

Se considera medidas repetidas cuando la misma variable se mide en varias ocasiones para cada sujeto [Hand y Taylor, 1993]. En el presente estudio, los parámetros se calculan

para cada ciclo respiratorio durante el protocolo. Este tipo de análisis requiere un menor número de sujetos en la muestra para obtener resultados similares. En este trabajo el grupo de variables dependientes son las treinta y seis variables extraídas de las señales biomédicas, y el atributo que se desea medir es la evolución del nivel de actividad y la fatiga muscular respiratoria durante el protocolo ventilatorio.

Cuando en un diseño experimental, como en este caso, a cada individuo se le realiza mediciones de diversas variables múltiples veces, algunos autores lo denominan diseño de medidas repetidas doblemente multivariable para resaltar esta doble multiplicidad del modelo: en el número de variables y en el número de veces que se miden [Norusis, 1992]. El análisis es similar al análisis de la varianza sin medidas repetidas. La principal diferencia se encuentra en la realización del análisis del modelo. Este se realiza mediante la transformación ortonormalizada de cada variable dependiente y los siguientes análisis se basan en estas nuevas variables transformadas.

## 7.3.2 Modelo de análisis de la actividad muscular respiratoria

### 7.3.2.1 Diseño de medidas repetidas

En el MANOVA de medidas repetidas se diferencia dos tipos de factores en base a su diferente comportamiento y tratamiento:

- *Factores intra-sujeto.*

Son aquellos en que todas las combinaciones de sus diferentes niveles ocurren dentro del individuo, y en cada nivel se miden las mismas variables.

- *Factores entre-sujeto.*

Dividen la muestra en varios grupos discretos de individuos, y sobre cada uno de estos grupos independientes están definidos los factores intra-sujeto.

Para el estudio del modelo se utiliza dos análisis que han de conducir a resultados coherentes aunque algo diferenciados:

- *Análisis multivariable.*

Considera que la muestra proviene de una distribución multivariada y no hace suposiciones sobre la matriz de covarianza de las variables del estudio.

- *Análisis univariable.*

Es más eficiente, especialmente en muestras pequeñas, si se reúne las condiciones de aplicación. Es decir, si existe diferencias es más probable que las detecte. Los resultados multivariados promediados, es decir, realizar un promedio de los resultados univariados para todas las variables, equivalen a los del análisis multivariable.



Se analiza variables transformadas ( $T_i$ ,  $i = 1, \dots, J$ ), que son combinaciones lineales de las variables originales (vector  $\bar{v}$ ) mediante  $J$  vectores de contrastes ( $\bar{c}_i$ ) estadísticamente independientes y ortonormales ( $J = M - 1$  siendo  $M$  el número de niveles del factor):  $T_i = \bar{c}_i \cdot \bar{v}^T$ .

Se evalúa la hipótesis nula que su valor medio es cero mediante los siguientes contrastes (se supone  $M=5$ ):

- *Repetido.*

Compara cada nivel del factor con el nivel que le precede.

$$\begin{array}{ll} \bar{c}_1 = [-1 \ 1 \ 0 \ 0 \ 0]/\sqrt{2} & \bar{c}_3 = [0 \ 0 \ -1 \ 1 \ 0]/\sqrt{2} \\ \bar{c}_2 = [0 \ -1 \ 1 \ 0 \ 0]/\sqrt{2} & \bar{c}_4 = [0 \ 0 \ 0 \ -1 \ 1]/\sqrt{2} \end{array}$$

- *Diferencia.*

Compara cada nivel del factor con la media de los niveles que le preceden.

$$\begin{array}{ll} \bar{c}_1 = [-1 \ 1 \ 0 \ 0 \ 0]/\sqrt{2} & \bar{c}_3 = [-1 \ -1 \ -1 \ 3 \ 0]/\sqrt{12} \\ \bar{c}_2 = [-1 \ -1 \ 2 \ 0 \ 0]/\sqrt{6} & \bar{c}_4 = [-1 \ -1 \ -1 \ -1 \ 4]/\sqrt{20} \end{array}$$

- *Helmert.*

Compara cada nivel del factor con la media de los niveles siguientes.

$$\begin{array}{ll} \bar{c}_1 = [4 \ -1 \ -1 \ -1 \ -1]/\sqrt{20} & \bar{c}_3 = [0 \ 0 \ 2 \ -1 \ -1]/\sqrt{6} \\ \bar{c}_2 = [0 \ 3 \ -1 \ -1 \ -1]/\sqrt{12} & \bar{c}_4 = [0 \ 0 \ 0 \ -1 \ 1]/\sqrt{2} \end{array}$$

- *Polinómico.*

Compara si los distintos niveles del factor se comportan de forma lineal, parabólica, cúbica, etc.

$$\begin{array}{ll} \bar{c}_1 = [-2 \ -1 \ 0 \ 1 \ 2]/\sqrt{10} & \bar{c}_3 = [-1 \ 2 \ 0 \ -2 \ 1]/\sqrt{10} \\ \bar{c}_2 = [2 \ -1 \ -2 \ -1 \ 2]/\sqrt{14} & \bar{c}_4 = [1 \ -4 \ 6 \ -4 \ 1]/\sqrt{70} \end{array}$$

### 7.3.2.2 Condiciones de aplicación

Para realizar el análisis mediante el análisis univariable se ha de cumplir dos hipótesis:

1. *Todas las variables transformadas o contrastes ( $T_i$ ) tienen la misma varianza y sus covarianzas son nulas para cada factor.*

Esta condición se estudia mediante el test de esfericidad de Mauchly que evalúa la hipótesis nula ( $H_0$ ) que la matriz de covarianza de las variables transformadas es  $\sigma$ -diagonal. Por tanto, la hipótesis se cumple cuando no se rechaza la  $H_0$  del test de Mauchly ( $p > 0.05$ ). Cuando el nivel de significación es bajo (violación de la esfericidad) se realiza un ajuste de los grados de libertad del numerador y del

denominador del estadístico  $F$ , mediante la multiplicación de ambos por el ratio  $\varepsilon$ , y evaluando la significación del nuevo estadístico  $F$ . El producto por el ratio  $\varepsilon$  de Huynh-Feldt es una aproximación para corregir el producto por el ratio  $\varepsilon$  de Greenhouse-Geisser, que tiende a ser demasiado conservador debido a su poca potencia, sobretodo con muestras pequeñas [Norusis, 1992]. Por ello, el primero es el utilizado en los resultados del presente estudio.

2. *Las matrices de covarianza han de ser iguales en todos los niveles del factor entre-sujeto para cada efecto en particular.*

Esta suposición, llamada condición de simetría, se ha de cumplir cuando hay varias poblaciones de individuos. Se evalúa mediante la generalización multivariada del test  $M$  de Box basado en el determinante de la matriz de covarianza de todas las celdas entre-sujeto del diseño. Si no se cumple la condición de simetría, los estadísticos  $F$  de los resultados univariados y multivariados promediados pueden no ser correctos. Este test para medidas repetidas, está basado en todas las variables originales de los efectos intra-sujeto y no en las variables transformadas. En el presente estudio, debido al gran número de variables consideradas, la matriz de covarianzas tiende a la singularidad impidiendo la ejecución del test. Sin embargo, uno de los principales objetivos de esta primera parte del estudio es seleccionar variables diferenciadoras de las poblaciones sin pretender realizar una valoración global de los distintos parámetros ni obtener resultados finales. Por tanto, la homogeneidad multivariable no es crítica en este caso y no se tiene en consideración esta condición. En contrapartida, se evalúa la igualdad de la varianza en cada variable original en ambos grupos de individuos mediante el cálculo del test de homogeneidad de la varianza de Levene.

### 7.3.2.3 Efecto tramo, ciclo y grupo

Para alcanzar los objetivos enumerados en el Apartado 7.2, mediante el MANOVA de medidas repetidas, se propone considerar los siguientes factores:

- *TRAMO.*

Es un factor intra-sujeto que tiene tantas categorías o niveles como número de intervalos se ha decidido dividir la prueba respiratoria. De esta manera, existe cinco categorías en las pruebas mantenidas y seis en las incrementales. La MCS varía dependiendo del individuo y de la prueba incremental. Por tanto, el número de cargas diferentes soportadas durante el test varía y puede diferir al valor de seis fijado para todos los casos. Para resolver esta dificultad y uniformizar el estudio en todos los individuos se calcula el valor de los parámetros en instantes equiespaciados a lo largo de la prueba mediante la interpolación lineal de los parámetros medidos a las diferentes cargas soportadas. Se espera, en general, que el valor de los parámetros varíen entre los distintos tramos de las pruebas y, por tanto, encontrar significación estadística ya que el nivel de actividad o fatiga del músculo varía durante el desarrollo de la prueba respiratoria.

- *CICLO.*

Este factor intra-sujeto consta de cinco categorías o niveles ya que cada tramo en que se ha dividido la prueba contiene cinco ciclos respiratorios. En cada variable, el valor del factor corresponde al parámetro calculado en el ciclo respiratorio. Si las características fisiológicas y funcionales del músculo ECM no varían significativamente de un ciclo respiratorio a otro durante un mismo tramo, el factor no presenta significación estadística.

- *GRUPO.*

Este factor entre-sujeto divide la muestra de doce individuos en dos niveles: un grupo de seis pacientes con EPOC y un grupo control de seis personas sanas. Presenta significación estadística cuando existe diferencias entre ambos grupos.

Además, se ha de destacar el especial interés de la interacción de los factores Grupo y Tramo (Grupo\*Tramo). Cuando en algún parámetro dicha interacción es significativa ( $p < 0.05$ ) indica que variaciones del parámetro entre ciertos tramos aparecen de forma diferente en la población de pacientes respecto al grupo control. Por tanto, estas variaciones o pendientes de ciertos parámetros permiten separar ambas poblaciones. Debido al reducido número de individuos en las poblaciones se tiene en cuenta en el presente trabajo valores de significación de  $p < 0.1$  si la interpretación del parámetro está relacionada directamente con la función muscular. Esto no significa que dicho parámetro diferencie poblaciones, aunque es posible que con una población mayor el nivel de significación aumente, sino que se considera y no se descarta en los posteriores análisis estadísticos.

Mediante el factor Ciclo y sus interacciones con otros factores es posible cuantificar si existe variabilidad entre ciclos durante un mismo tramo. Visualmente se observa que los parámetros tienen un valor casi constante en los cinco ciclos respiratorios escogidos como representativos de una carga determinada en el protocolo incremental o intervalo de la prueba mantenida.

Para cada parámetro en estudio se analiza el efecto Ciclo y su interacción Grupo\*Ciclo mediante los siguientes análisis:

- Multivariable mediante cuatro estadísticos de contraste: Traza de Pillai, Lambda de Wilks, Traza de Hotelling y Raíz mayor de Roy.
- Univariados promediados.
- Contraste repetido generando cuatro variables transformadas.

Todos ellos coinciden en que se puede aceptar la igualdad entre los ciclos ( $p > 0.6$ ), y entre los grupos para las variaciones entre los ciclos ( $p > 0.49$ ). Esto significa que además de no apreciarse diferencia significativa entre cinco ciclos respiratorios consecutivos, tampoco se aprecia entre ambos grupos en ninguna de las variaciones entre los ciclos.

A partir de los resultados, se realiza la media de los valores en los cinco ciclos respiratorios para conseguir un único valor promedio en cada tramo de las pruebas. De esta forma, se consigue reducir en cinco veces el volumen de datos con una pérdida de información estadísticamente despreciable. Además, esta media se considera una medida más robusta que escoger un único ciclo respiratorio en que las señales miográficas pudieran tener artefactos debido, por ejemplo, a movimientos del sujeto.

#### 7.3.2.4 Cálculos realizados

Se realiza el MANOVA de medidas repetidas en cada uno de los tres grupos de parámetros de análisis de la actividad muscular (temporales, de correlación y frecuenciales) para cada una de las pruebas (ver Apartado 2.4.4):

1. *Evaluación de las condiciones de aplicación.*

Se realiza mediante el test de homogeneidad de la varianza de Levene y la condición de esfericidad con el test de Mauchly.

2. *Efecto del factor Grupo.*

Se calcula los resultados multivariados y univariados que permite determinar el parámetro cuyo valor medio durante toda la prueba separa de forma estadísticamente significativa ambas poblaciones.

3. *Efecto del factor Tramo.*

Se calcula los siguientes resultados:

- Multivariados.
- Multivariados promediados mediante la Traza de Pillai, Lambda de Wilks, Traza de Hotelling y Raíz mayor de Roy.
- Univariados promediado.
- Aplicación de los contrastes Repetido, Diferencia, Helmert y Polinómico en los diferentes tramos de la prueba.

Estos últimos resultados permiten en cada parámetro agrupar los tramos de una prueba mediante el promedio de sus valores para dividir el test en tres intervalos: inicial, central y final. De esta forma, variaciones entre estos intervalos están relacionados con la adaptación al esfuerzo (inicial-central) y fatiga muscular (central-final).

4. *Efecto de la interacción de los factores Grupo y Tramo.*

Se calcula las mismas pruebas estadísticas mencionadas en el punto anterior para el efecto de la interacción Grupo\*Tramo. Dichos resultados permiten seleccionar variaciones de interés de los parámetros entre los tramos o intervalos de la prueba cuando son estadísticamente diferentes entre ambas poblaciones ( $p < 0.05$  o en ciertas situaciones  $p < 0.1$ ).

En los siguientes Apartados se comenta e interpreta los resultados y conclusiones más relevantes de dicho análisis, así como las variables seleccionadas para el siguiente análisis estadístico para diferenciar las dos poblaciones en estudio. Los resultados estadísticos completos para cada prueba respiratoria se pueden consultar en [Serena y Mañanas, 1999].

### 7.3.3 Protocolo incremental

#### 7.3.3.1 Resultados

A continuación se compara ordenadamente los resultados de los estudios estadísticos realizados en los protocolos IC e IR.

#### PARÁMETROS TEMPORALES

##### Efecto Grupo

En ambas pruebas se observa que ningún parámetro presenta en su valor promedio diferencia significativa entre grupos indicando un comportamiento similar durante la prueba: el valor de los parámetros aumenta al incrementarse la carga ventilatoria (ver figuras 7.1 y 7.2).

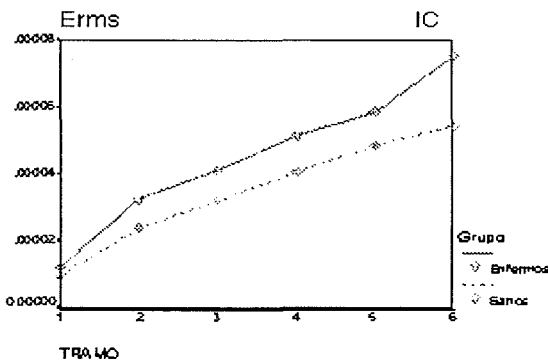


Figura 7.1: RMS de la señal EMG en la prueba IC en ambos grupos de individuos.

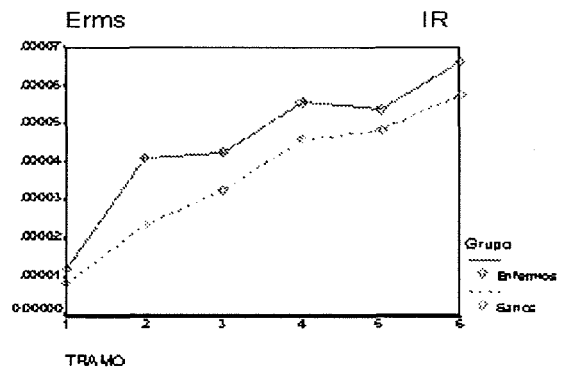


Figura 7.2: RMS de la señal EMG en la prueba IR en ambos grupos de individuos.

##### Efecto Tramo

El contraste lineal permite evaluar la evolución creciente de los parámetros cuando aumenta la carga soportada por el paciente. El número de parámetros en que el contraste lineal es significativo es mayor en la prueba IC que en la IR. Estos parámetros corresponden a los provenientes del RMS de las señales EMG y VMG ( $p < 0.005$ ), tanto en valor absoluto como normalizado por el máximo o por el valor medio de toda la prueba.

Al evaluar el resto de contrastes, las variables transformadas indican diferencias más significativas entre los tramos iniciales caracterizados por bajas cargas ventilatorias ( $p < 0.02$ ).

#### Efecto Grupo\*Tramo

Es importante señalar la significación del contraste lineal del RMS de la señal EMG en la prueba IR para diferenciar los grupos ( $p < 0.001$ ) que no sucede en la prueba IC ( $p < 0.29$ ). Esto es debido a la evolución creciente más uniforme del parámetro en los dos grupos en esta última cuando no hay descanso entre cargas (ver figuras 7.1 y 7.2). Sin embargo, la prueba IC muestra una diferencia significativa entre los grupos en la evolución entre el intervalo central (promedio de los tramos 2, 3, 4 y 5) y el final (tramo 6) en el RMS de la señal EMG normalizado respecto al valor medio de toda la prueba ( $p < 0.02$ ).

Respecto a los parámetros relacionados con la duración del período inspiratorio se observa en ambas pruebas que el ritmo inspiratorio varía dependiendo de la población ante cargas próximas a la máxima carga soportada como muestra la variación de los siguientes parámetros (ver figuras 7.3 y 7.4):

- Tinsciclo en la prueba IC entre el intervalo central (promedio de los tramos 3 a 5) y el final (tramo 6) ( $p < 0.01$ ).
- Tins en la prueba IR entre el intervalo central (promedio de los tramos 2 a 5) y el final (tramo 6) ( $p < 0.002$ ).

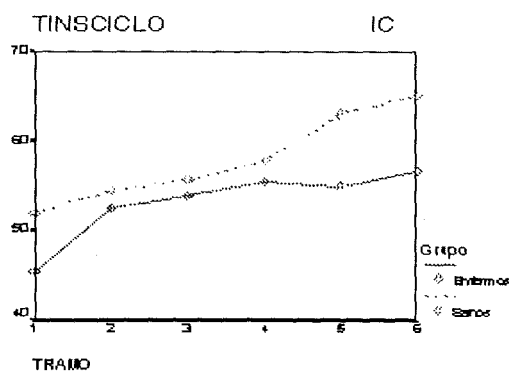


Figura 7.3: Duración del período inspiratorio normalizado respecto a la duración total del ciclo en la prueba IC.

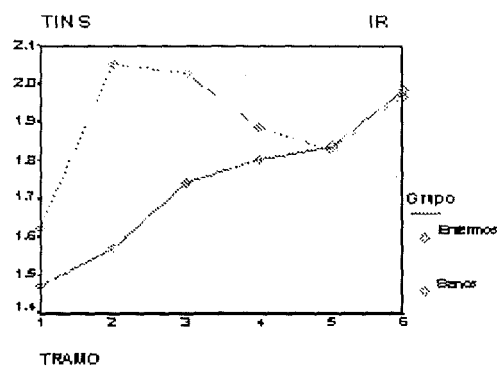


Figura 7.4: Duración del período inspiratorio en la prueba IR en ambos grupos de individuos.

## PARÁMETROS DE CORRELACIÓN

### Efecto Grupo

En ambas pruebas ventilatorias, el valor promedio del TRM para retardos positivos normalizado respecto al período inspiratorio (ver figuras 7.5 y 7.6) diferencia ambos grupos ( $l_{pi}$ ,  $p < 0.04$ ), así como su valor ponderado por el área positiva de la FCC ( $l_{piar}$ ,  $p < 0.04$ ) y calculado a partir de la FCC promedio ( $l_{picp}$ ,  $p < 0.04$ ).

Además, en la prueba IR aparecen diferencias significativas en dos parámetros más que la prueba IC:

- Tiempo de retardo medio para retardos negativos normalizados respecto al período inspiratorio (lni,  $p < 0.005$ ). En la figura 7.8 se observa que en la prueba IR, exceptuando el primer tramo, los valores del parámetro lni en el grupo de enfermos es claramente mayor que en el grupo de control, mientras que en la prueba IC esta diferencia entre grupos en los tramos 2 a 4 no existe (ver figura 7.7).
- Tiempo de retardo medio para retardos positivos de la FCC promedio (lpc,  $p < 0.05$ ).

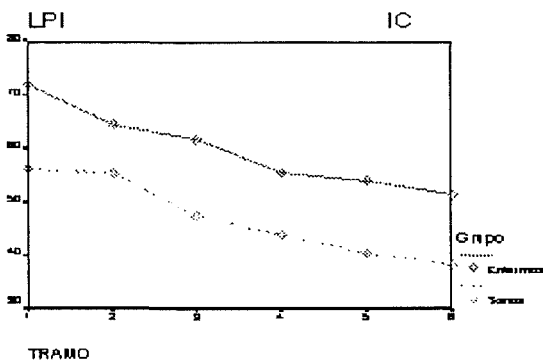


Figura 7.5: TRM para retardos positivos normalizado respecto al periodo inspiratorio de la FCC en la prueba IC.

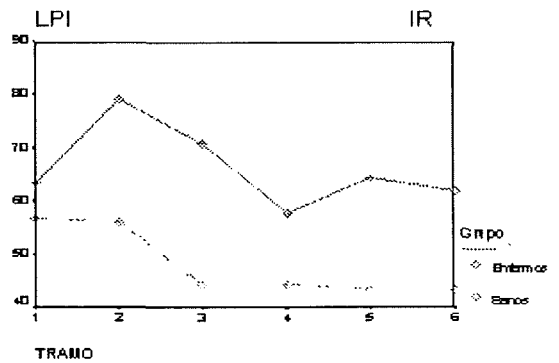


Figura 7.6: TRM para retardos positivos normalizado respecto al periodo inspiratorio de la FCC en la prueba IR.

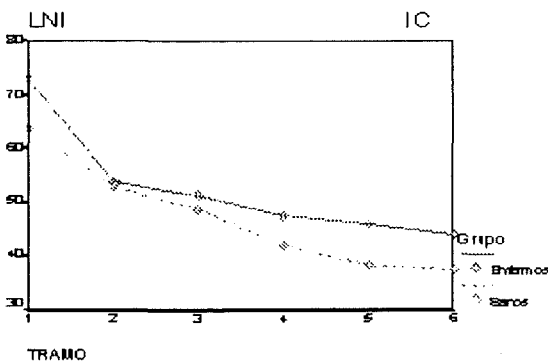


Figura 7.7: TRM para retardos negativos normalizado respecto al periodo inspiratorio de la FCC en la prueba IC.

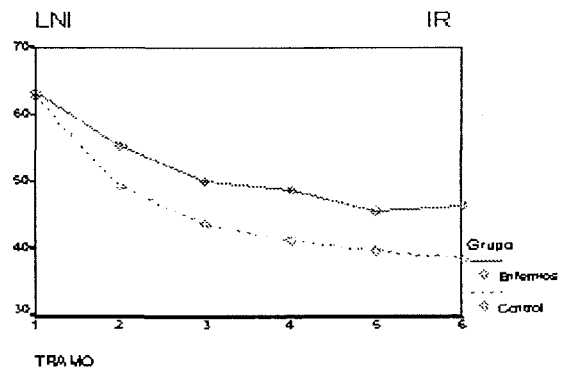


Figura 7.8: TRM para retardos negativos normalizado respecto al periodo inspiratorio de la FCC en la prueba IR.

### Efecto Tramo

En la tabla 7.1 se observa que los contrastes lineales para la mayoría de los parámetros en la prueba IC, correspondientes a tiempos de retardo medio tienen significación estadística, mientras que en la prueba IR la significación es menor o no existe. Estos resultados son similares a los obtenidos en los parámetros temporales.

Parámetro	IC (p<)	IR (p<)
ln	0.03	0.1
lnc	0.001	0.02
lni	0.0005	0.005
lp	0.002	0.8
lpc	0.001	0.7
lpi	0.001	0.03
lpiar	0.001	0.02

Tabla 7.1: *p*-valor del contraste lineal para diferentes parámetros de correlación evaluado mediante el efecto tramo en ambas pruebas respiratorias.

#### Efecto Grupo\*Tramo

En la prueba IC, los parámetros correspondientes a las áreas de la FCC para retardos positivos presentan diferencias significativas en ambos grupos entre tramos intermedios, en especial arcp3-4 ( $p < 0.08$ ), y sobre todo, arlpi4-5 ( $p < 0.04$ ). Entre los tramos intermedios también aparece significación estadística entre el tramo 3 y 4 en el parámetro ltot ( $p < 0.05$ ).

En las figuras 7.9 y 7.10 se observa el diferente comportamiento de las dos poblaciones de individuos respecto a las áreas de la FCC para retardos positivos en las pruebas.

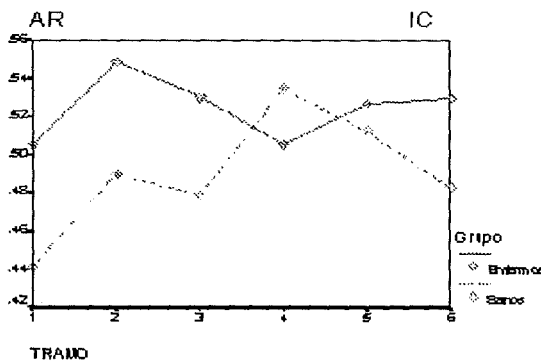


Figura 7.9: Área de la FCC para retardos positivos en la prueba IC.

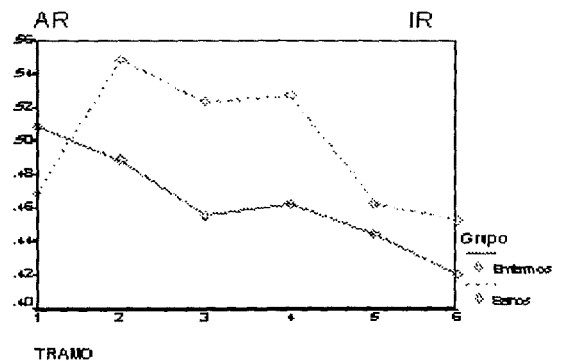


Figura 7.10: Área de la FCC para retardos positivos en la prueba IR.

En la prueba IR, los parámetros correspondientes a tiempos de retardo medio presentan significación estadística. En concreto, la variación entre el intervalo central (promedio de los tramos 2 a 4) y el final (promedio de los tramos 5 y 6) de los parámetros lp ( $p < 0.08$ ) y ln ( $p < 0.06$ ). El parámetro lpi4-5 también indica la diferente evolución entre los grupos al final de la prueba ( $p < 0.02$ ).



Por último, en ambas pruebas destaca la diferencia entre grupos en la variación de apnea a volumen corriente en dos parámetros en ambas pruebas:

- Parámetro  $ltot0-1$  ( $p < 0.04$  en la prueba IC y  $p < 0.02$  en la prueba IR).
- Parámetro  $arplp0-1$  ( $p < 0.02$  en ambas pruebas).

## PARÁMETROS FRECUENCIALES

### Efecto Grupo

Ningún parámetro de la DEP de las señales miográficas de las dos pruebas muestran diferencias significativas entre los dos grupos de individuos. En las figuras 7.11 y 7.12 se observa, como ejemplo, la evolución de la frecuencia central en la señal EMG. El parámetro tiene valores muy similares en ambos grupos a volumen corriente y se diferencian progresivamente con cargas superiores.

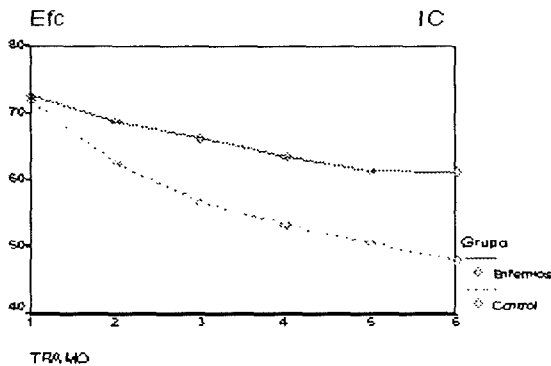


Figura 7.11:  $fc$  de la DEP de la señal EMG en ambos grupos en la prueba IC.

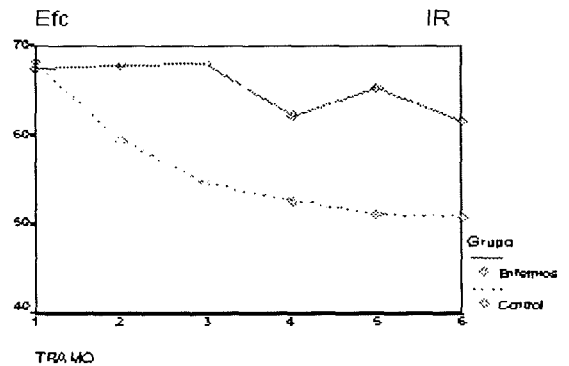


Figura 7.12:  $fc$  de la DEP de la señal EMG en ambos grupos en la prueba IR.

Por tanto, se evalúa el efecto grupo considerando sólo el intervalo central de las pruebas mediante el promedio de los tramos 3 a 5:

- En la prueba IC, la frecuencia central ( $Efc$ ,  $p < 0.09$ ) y la frecuencia media ( $Efm$ ,  $p < 0.1$ ) en la señal EMG (ver figura 7.11).
- En la prueba IR, la frecuencia central ( $Efc$ ,  $p < 0.04$ ), la frecuencia media ( $Efm$ ,  $p < 0.07$ ) y el ratio H/L ( $Ehl$ ,  $p < 0.06$ ) en la señal EMG; y la frecuencia central en la señal VMG ( $Vfc$ ,  $p < 0.03$ ) (ver figuras 7.12 y 7.14).

Se observa, por tanto, mucha menor significación estadística en los parámetros frecuenciales de la señal VMG respecto a la señal EMG.

### Efecto Tramo

De igual forma que en los parámetros temporales y de correlación, el comportamiento lineal es evidente especialmente en la prueba IC. Dicho comportamiento es decreciente a medida que aumenta la carga. En la señal EMG se obtiene para los tres parámetros una significación de  $p < 0.0005$  en ambas pruebas (ver figuras 7.11 a 7.14). En la señal VMG, la frecuencia central presenta  $p < 0.0005$  en la prueba IC y  $p < 0.07$  en la prueba IR, y la frecuencia media  $p < 0.05$  en ambas pruebas.

### Efecto Grupo\*Tramo

En primer lugar, la tendencia lineal de ciertos parámetros difiere entre ambas poblaciones. En la prueba IC, la frecuencia central en la señal EMG ( $p < 0.1$ , ver figura 7.11) y la frecuencia media en la señal VMG ( $p < 0.07$ ) son los parámetros más importantes. En la prueba IR, destaca la frecuencia central ( $p < 0.01$ , ver figura 7.12) y el ratio H/L ( $p < 0.05$ , ver figura 7.14) en la señal EMG. Estas variables reflejan las diferentes pendientes durante toda la prueba en el desplazamiento de la DEP de las dos señales en ambos grupos. En las figuras 7.13 y 7.14 se muestra la evolución del ratio H/L en la señal EMG en ambas pruebas.

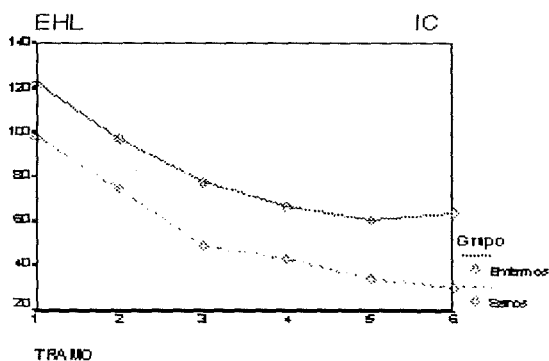


Figura 7.13: Ratio H/L de la DEP de la señal EMG en la prueba IC.

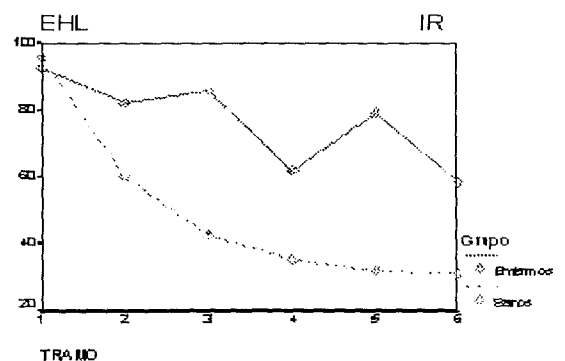


Figura 7.14: Ratio H/L de la DEP de la señal EMG en la prueba IR.

Aparece, también, variación de los parámetros frecuenciales entre los dos últimos tramos de ambas pruebas que difiere entre el grupo de pacientes y grupo control. En la prueba IC, son los parámetros Efc5-6 ( $p < 0.04$ , ver figura 7.11) y Efm5-6 ( $p < 0.04$ ). En la prueba IR, el parámetro Ehl5-6 ( $p < 0.07$ , ver figura 7.14) es el que muestra más el diferente comportamiento de los dos grupos ante cargas ventilatorias altas.

Por último, en la prueba IR, existe dos parámetros en la señal EMG cuyas variaciones entre el intervalo inicial (promedio de los tramos 1 y 2) y central (promedio de los tramos 3 a 5) es diferente entre ambas poblaciones: frecuencia central ( $p < 0.01$ , ver figura 7.12) y ratio H/L ( $p < 0.03$ , ver figura 7.14).

### 7.3.3.2 Selección de parámetros

Se pretende seleccionar las variables con mayor interpretación sobre el comportamiento muscular de los individuos durante la prueba, y que mejor diferencien entre los dos grupos según los resultados anteriores. A estas variables se les aplica posteriormente un análisis factorial que proporcione el menor número de combinaciones posible de éstos, y que expliquen el máximo nivel del comportamiento del músculo ECM durante la prueba.

En la tabla 7.2 se muestra los parámetros seleccionados en ambas pruebas incrementales, juntamente con su nivel de significación estadística.

Parámetros temporales	Prueba IC		Prueba IR	
	Erme234-56	p<0.02	Erms_lin	p<0.001
Tinsciclo345-6	p<0.01	Tins2345-6	p<0.002	
Parámetros de correlación	lpi	p<0.04	lpi	p<0.04
			lni	p<0.005
	ltot0-1	p<0.04	ltot0-1	p<0.02
	arlp0-1	p<0.02	arlp0-1	p<0.02
	ltot3-4	p<0.05	lp234-56	p<0.08
	arcp3-4	p<0.08	ln234-56	p<0.06
Parámetros frecuenciales	arlp4-5	p<0.04	lpi4-5	p<0.02
	Efc345	p<0.09	Efc345	p<0.04
	Efm345	p<0.1	Efm345	p<0.07
			Ehl345	p<0.06
			Vfc345	p<0.03
	Efc5-6	p<0.04	Ehl5-6	p<0.07
	Efm5-6	p<0.04		
	Efc_lin	p<0.1	Efc_lin	p<0.01
	Vfm_lin	p<0.07	Ehl_lin	p<0.05
			Efc12-345	p<0.01
		Ehl12-345	p<0.03	

Tabla 7.2: Parámetros seleccionados en las pruebas incrementales con el nivel de significación estadística.

### 7.3.3.3 Conclusiones

El comportamiento de los parámetros en ambas pruebas incrementales tiene una tendencia marcadamente lineal como lo demuestra el contraste correspondiente en el efecto tramo. A medida que se aumenta la carga ventilatoria es creciente para los parámetros temporales y decreciente para los de correlación y frecuenciales. El primer caso indica un aumento de la actividad eléctrica muscular debido a un reclutamiento de fibras y aumento de la frecuencia de activación. Los tiempos de retardo medio de la FCC reflejan

un mayor acoplamiento entre la actividad mecánica y eléctrica, y los parámetros frecuenciales muestran un desplazamiento de la DEP hacia bajas frecuencias a medida que se incrementa la carga. Las tendencias lineales son más uniformes en la prueba IC que en la IR debido a la naturaleza de las pruebas en que la primera no presenta descansos entre cargas ventilatorias.

Respecto a los parámetros que diferencian grupos de sujetos enfermos y grupo control, destaca el tiempo de retardo medio para retardos positivos normalizado respecto al período inspiratorio junto a otros parámetros calculados a partir de la FCC de las señales miográficas. El valor medio de este parámetro es significativamente superior en el grupo de pacientes respecto al grupo control indicando que la energía de la señal VMG se encuentra globalmente más retardada respecto a la señal EMG, concretamente durante el período espiratorio, en la primera población.

La variación del RMS en la señal EMG entre el intervalo central y final de la prueba IC indica una mayor actividad eléctrica necesaria en el grupo de pacientes respecto al de sujetos sanos cuando se alcanza cargas ventilatorias muy elevadas y el músculo se encuentra próximo a la fatiga.

La variación de parámetros provenientes de la duración del período inspiratorio entre dichos intervalos muestra que ante cargas cercanas a la máxima carga sostenida el ritmo inspiratorio varía de forma diferente tratándose de sujetos sanos o pacientes con EPOC.

Los parámetros frecuenciales, principalmente en la señal EMG, muestra valores superiores en la población de pacientes que en sujetos sanos debido a un mayor número de componentes frecuenciales relacionado, posiblemente, con un mayor número de fibras activas. Dichos parámetros varían significativamente entre los últimos tramos de las pruebas de forma que el desplazamiento en frecuencia con cargas muy elevadas está relacionado con la fatiga muscular cuyo efecto aparece de diferente manera entre los sujetos sanos y pacientes con EPOC.

Por último, hay que destacar las variaciones entre la prueba de apnea y el primer tramo de las pruebas incrementales (a volumen corriente) que se detecta con el efecto Grupo\*Tramo en los tiempos de retardo medio y el área de la FCC.

Las diferencias halladas entre las dos pruebas son atribuibles a la diferente naturaleza de las pruebas y, posiblemente, al aprendizaje de los individuos durante la primera de ellas (IC).

### **7.3.4 Protocolo mantenido**

#### **7.3.4.1 Resultados**

De igual forma que se ha realizado para las pruebas incrementales, a continuación se comenta los resultados de los estudios realizados en ambas pruebas mantenidas.

## PARÁMETROS TEMPORALES

### Efecto Grupo

En ninguna de las dos pruebas se observa significación en el efecto grupo. Ello puede ser debido a que el sujeto soporta una carga muy elevada que produce niveles de energía similares en pacientes con EPOC y sujetos sanos en las señales miográficas en ambas pruebas.

Sólo es destacable que en ambas pruebas la duración del período inspiratorio respecto a la duración total del ciclo es mayor en sujetos sanos que en pacientes con EPOC (ver figuras 7.15 y 7.16). Es decir, que estos últimos inspiran con mayor rapidez que los individuos sanos, aunque no se alcanza significación estadística ( $p < 0.2$  en la prueba M1 y  $p < 0.1$  en la prueba M2).

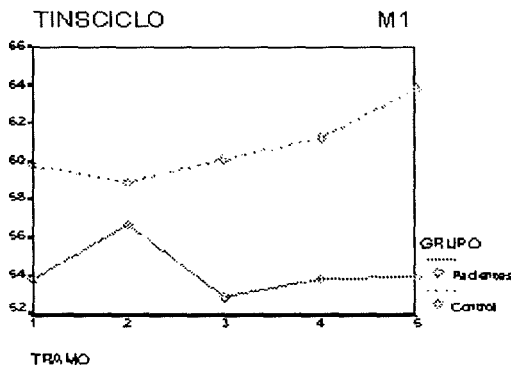


Figura 7.15: Duración del período inspiratorio respecto al total del ciclo en la prueba M1.

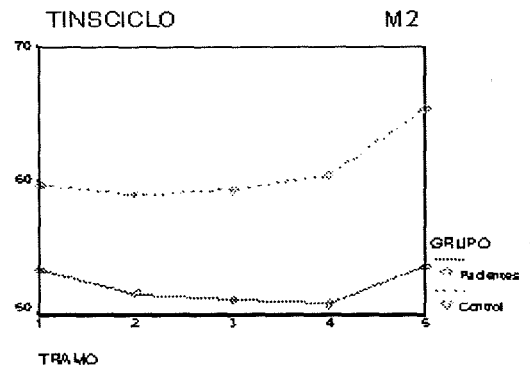


Figura 7.16: Duración del período inspiratorio respecto al total del ciclo en la prueba M2.

Relacionado con este efecto, se evalúa el factor Grupo en el intervalo central de la prueba M2 (promedio de los tramos 2 a 4) en el parámetro duración del período inspiratorio con un nivel de significación estadístico de  $p < 0.07$  (ver figura 7.17)

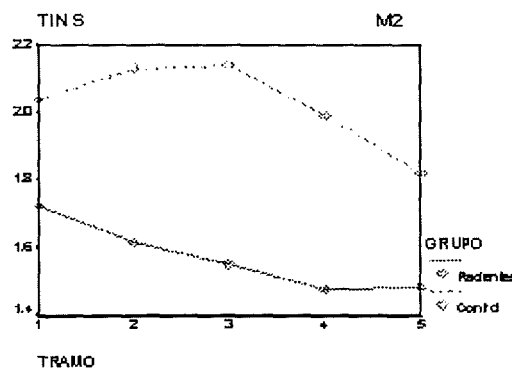


Figura 7.17: Duración del período inspiratorio en la prueba M2.

### Efecto Tramo

En la prueba M2 se encuentra más parámetros y variables transformadas con significación estadística que en la prueba M1. Ello se produce especialmente en la parte final de las pruebas en los siguientes parámetros:

- RMS de la señal EMG y sus diferentes normalizaciones ( $p < 0.07$  en la prueba M1 y  $p < 0.04$  en la prueba M2).
- Cociente de los RMS de la señal EMG y la señal VMG, y sus diferentes normalizaciones en la prueba M1 ( $p < 0.07$ ).
- Duración del período inspiratorio normalizado respecto a la duración total del ciclo en la prueba M2 ( $p < 0.02$ ).

Además, en la prueba M2 se observa un comportamiento lineal de los diferentes parámetros relacionados con el RMS de la señal EMG ( $p < 0.01$ ), mientras que en la prueba M1 dicha tendencia no es tan significativa ( $p < 0.1$ ).

### Efecto Grupo\*Tramo

Aparecen menos variables con mayor efecto Grupo\*Tramo en la prueba M1 que en la M2. En esta última, el RMS normalizado respecto al valor medio diferencia poblaciones en los siguientes instantes de la prueba:

- Entre el intervalo inicial (tramo 1) y el central (promedio de los tramos 2 a 4) en la señal EMG ( $p < 0.01$ ).
- Entre el intervalo central (promedio de los tramos 3 y 4) y final (tramo 5) en la señal VMG ( $p < 0.04$ ).

Al final de la prueba M1, la variación del RMS de la señal EMG entre los tramos 4 y 5 presenta significación estadística como puede observarse en la figura 7.18 ( $p < 0.05$ ). Las tendencias entre los dos últimos tramos también son opuestas en el RMS de la señal VMG en ambas poblaciones, pero sin alcanzar la significación estadística,  $p < 0.2$ , debido a la gran variabilidad entre individuos (ver figura 7.19).

Un parámetro que reúne información correspondiente a ambas señales miográficas es el EVrms. Dicho parámetro indica la relación entre la energía de la señal EMG y la de la señal VMG. Se puede interpretar como el efecto entre la eficiencia eléctrica y mecánica. Se observa en la prueba M1, en la evolución de las medias de cada Tramo (ver figura 7.20), tendencias diferentes en ambas poblaciones al final de la prueba, indicando diferencias en la fatiga muscular, aunque sin alcanzar significación estadística ( $p < 0.2$ ). Este efecto aparece también en la prueba M2 (ver figura 7.21) donde tampoco se obtiene significación estadística ( $p < 0.2$ ).

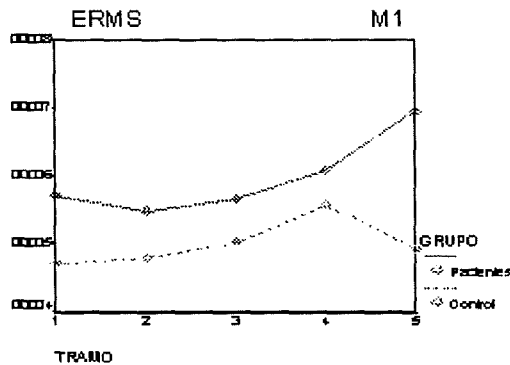


Figura 7.18: RMS de la señal EMG durante la prueba M1.

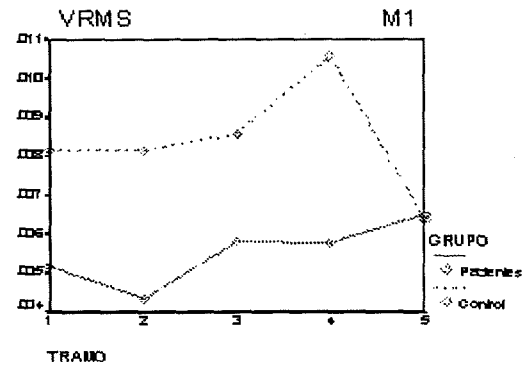


Figura 7.19: RMS de la señal VMG durante la prueba M1.

Esta falta de significación puede ser debida a la dispersión de los valores y al bajo número de sujetos en los grupos. Una posible manera de aumentar el número de datos es reuniendo los resultados de ambos protocolos ventilatorios y evaluar el parámetro EVrms del tramo 4 al tramo 5. En los tests de normalidad de Kolmogorov-Smirnov y Shapiro-Wilks (ver tabla 7.3) se observa que los datos en la población de pacientes son normales, sin embargo, no es así en el grupo control. Por consiguiente, se utiliza una prueba no paramétrica: U de Mann-Whitney. No obstante, la diferencia encontrada ( $p < 0.05$ ) incluye el efecto de pruebas diferentes. Por tanto, la prueba de Wilcoxon es estrictamente más correcta ya que compara dos muestras relacionadas [Ferrán, 1996]. En este caso se obtiene  $p < 0.1$  que, aunque no es estadísticamente significativo, es suficientemente pequeño para ser tenido en consideración.

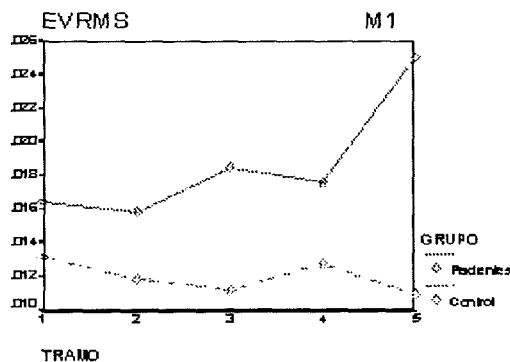


Figura 7.20: Cociente de RMS entre las señales miográficas en ambas poblaciones durante la prueba M1.

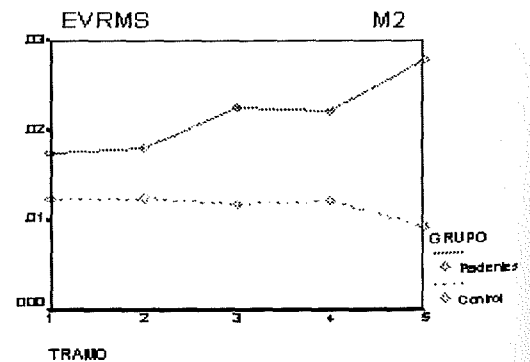


Figura 7.21: Cociente de RMS entre las señales miográficas en ambas poblaciones durante la prueba M2.

Relacionado con esta circunstancia, se selecciona la variación de RMS entre tramos 4 y 5 mediante los parámetros Erms y Vrms de la prueba M1 en que este efecto se visualiza con mayor énfasis. Además, la variable transformada que mayor nivel de significación presenta es la variación del cociente normalizado respecto al valor máximo entre el intervalo central (promedio de los tramos 3 y 4) y el final (tramo 5) en la prueba M2 ( $p < 0.04$ ).

Grupo	Kolmogorov-Smirnov	Shapiro-Wilks
Pacientes	0.200	0.716
Control	0.000	0.010

Tabla 7.3: Pruebas de normalidad para la diferencia entre los tramos 4 y 5 del parámetro  $EV_{rms}$ .

Al observar gráficamente la evolución del RMS normalizado respecto al valor medio de todo el protocolo en la señal EMG (ver figuras 7.22 y 7.23) se comprueba que dicha evolución es similar para los pacientes en ambas pruebas ventilatorias, pero difiere en el grupo de control, sobre todo al final de la prueba.

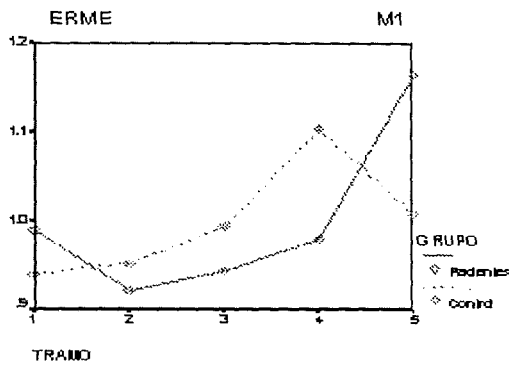


Figura 7.22: RMS de la señal EMG, normalizado respecto a la media de todo el protocolo, para la prueba M1.

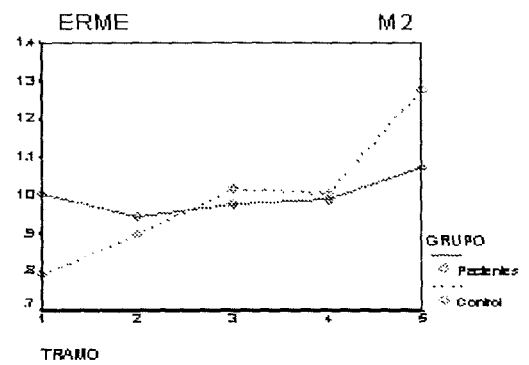


Figura 7.23: RMS de la señal EMG, normalizado respecto a la media de todo el protocolo, para la prueba M2.

Se observa el mismo efecto en la señal VMG (ver figuras 7.24 y 7.25) debido posiblemente a un efecto de aprendizaje en los individuos sanos. Por tanto, es lógico que el contraste lineal en ambas señales para los RMS normalizados sea significativo en la prueba M2 ( $p < 0.03$ ), mientras que no lo es en la M1. Además, el contraste lineal del parámetro  $Erms$  también presenta alta significación estadística en la prueba M2 ( $p < 0.02$ ).

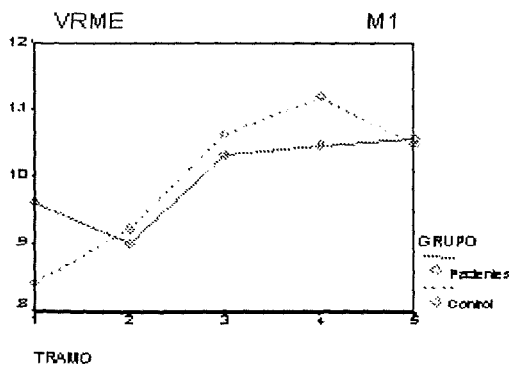


Figura 7.24: RMS de la señal VMG, normalizado respecto a la media de todo el protocolo, para la prueba M1.

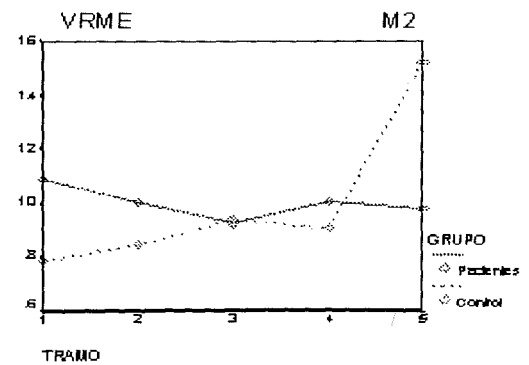


Figura 7.25: RMS de la señal VMG, normalizado respecto a la media de todo el protocolo, para la prueba M2.



Referente a los parámetros procedentes de la duración de períodos inspiratorios tiene especial interés su variación normalizada respecto a la duración del ciclo respiratorio entre los tramos 2 y 3 en la prueba M1 ( $p < 0.02$ ). En individuos enfermos, el porcentaje de tiempo de inspiración respecto al total del ciclo disminuye en comparación con los sujetos sanos (ver figura 7.15).

En las figuras 7.26 y 7.27 se observa la evolución de la duración del período inspiratorio en las pruebas M1 y M2, respectivamente, entre los tramos agrupados. La variación entre intervalos inicial (promedio de tramos 1 y 2 en la prueba M1, y tramo 1 en la M2) y central (promedio de tramos 3 y 4 en la prueba M1, y tramos 2 a 4 en la M2) presentan significación estadística ( $p < 0.07$  en la M1 y  $p < 0.005$  en la M2).

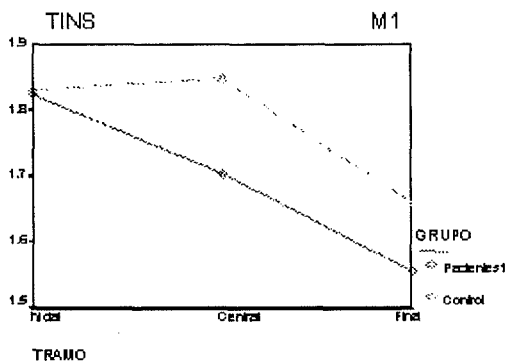


Figura 7.26: Variación del período inspiratorio entre intervalos inicial, central y final en la prueba M1.

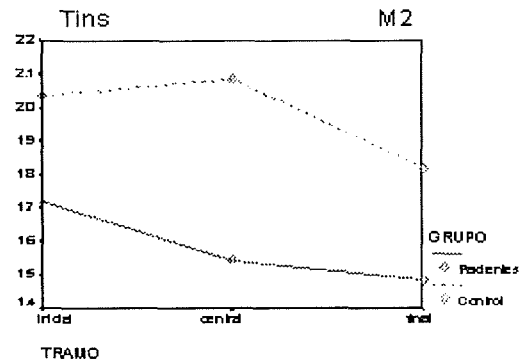


Figura 7.27: Variación del período inspiratorio entre intervalos inicial, central y final en la prueba M2.

## PARÁMETROS DE CORRELACIÓN

### Efecto Grupo

En ambas pruebas ventilatorias, el TRM para retardos positivos normalizado respecto a la duración del período inspiratorio diferencia ambas poblaciones (ver figuras 7.28 y 7.29). El valor promedio de los intervalos centrales (tramos 2 a 4 en la prueba M1, 2 y 3 en la M2) presenta una diferencia estadísticamente significativa ( $p < 0.07$  en la prueba M1 y  $p < 0.005$  en la M2).

### Efecto Tramo

Aparece diversas variables transformadas con significación estadística, especialmente las correspondientes a la segunda mitad de la prueba. En concreto, en la tabla 7.4 se observa las significaciones estadísticas de los parámetros con  $p < 0.05$  mediante análisis univariable. Estos corresponden a los tiempos de retardo medio. Sin embargo, la significación desaparece totalmente cuando dichos parámetros se normalizan respecto a la duración del período inspiratorio o a la total del ciclo respiratorio. Por tanto, las variaciones significativas entre tramos se deben más a la variación del ritmo inspiratorio y respiratorio que al propio parámetro de correlación. Esta misma circunstancia aparece también con el contraste lineal (ver tabla 7.4).

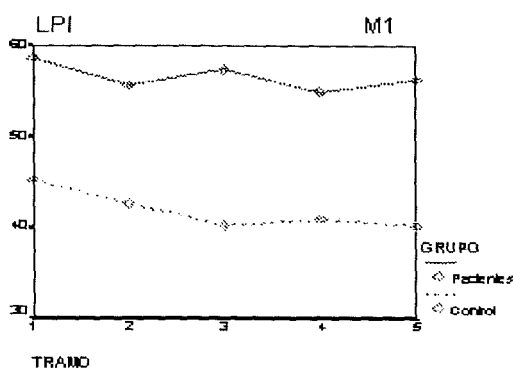


Figura 7.28: TRM para retardos positivos normalizado respecto a la duración del período inspiratorio en la prueba M1.

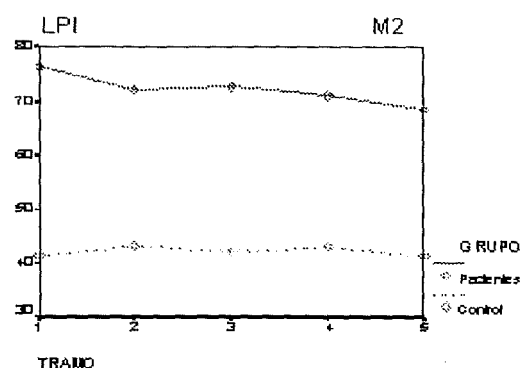


Figura 7.29: TRM para retardos positivos normalizado respecto a la duración del período inspiratorio en la prueba M2.

Parámetro	Prueba M1		Prueba M2	
	Univariable (p<)	Lineal(p<)	Univariable(p<)	Lineal(p<)
ln	0.0005	0.009	0.003	0.01
lp	0.01	0.002	0.01	0.02
lppar	0.0005	0.006	0.003	0.01

Tabla 7.4: Niveles de significación para diversos parámetros de correlación mediante el análisis univariable y los contrastes lineales.

#### Efecto Grupo\*Tramo

Los parámetros que presentan significación estadística provienen, básicamente, de los tiempos de retardo medio con variaciones al inicio y al final de las pruebas. En la prueba M1 se observa en la figura 7.28 diferencia entre grupos ante la variación de dichos tiempos normalizados respecto a la duración del período de inspiración entre los tramos 3 y 4 ( $p < 0.03$ ).

Respecto a los parámetros provenientes del área, se alcanza la mayor significación de  $p < 0.09$  en el parámetro arlpi entre el intervalo inicial (promedio de tramos 1 y 2) y central (promedio de tramos 3 y 4).

En la prueba M2 se obtiene nivel de significación de  $p < 0.05$  en la variación del TRM entre los tramos 1 y 2 y en el contraste parabólico de dicho parámetro ( $p < 0.05$ ). Este contraste (ver figura 7.30) indica el diferente retardo entre actividad mecánica y eléctrica del músculo durante la prueba entre ambas poblaciones cuando los tramos extremos tienen una importancia inversa a la de los centrales.

Por último, el parámetro lppar presenta  $p < 0.06$  en su variación entre el intervalo central (promedio de tramos 3 y 4) y el tramo final.

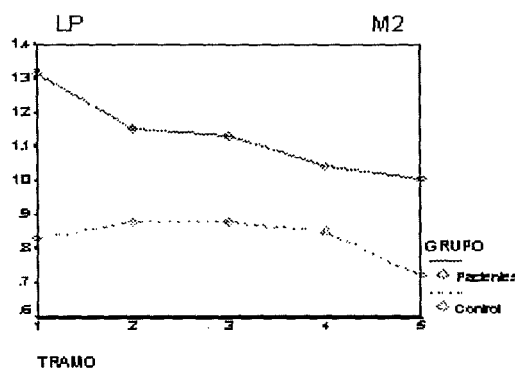


Figura 7.30: TRM para retardos positivos durante la prueba M2 para cada grupo.

## PARÁMETROS FRECUENCIALES

### Efecto Grupo

En las pruebas M1 y M2, los parámetros calculados a partir de la DEP de la señal EMG diferencian claramente ambas poblaciones, sobre todo la frecuencia central ( $p < 0.01$  y  $p < 0.02$ , respectivamente) y frecuencia media ( $p < 0.02$  y  $p < 0.01$ , respectivamente); mientras que el ratio H/L obtiene una significación de  $p < 0.08$  y  $p < 0.06$  en las pruebas M1 y M2, respectivamente (ver figuras 7.31 a 7.34). En cambio, no aparece significación en los parámetros frecuenciales de la señal VMG.

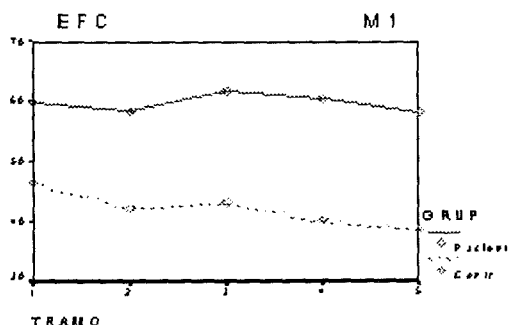


Figura 7.31: Frecuencia central de la señal EMG en la prueba M1.

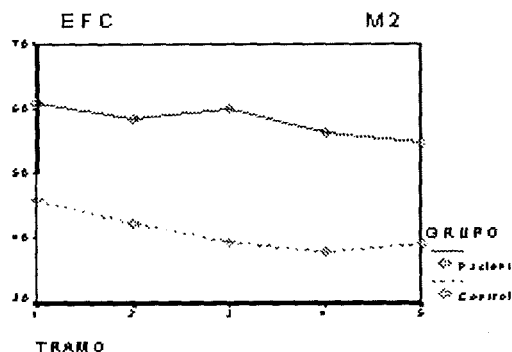


Figura 7.32: Frecuencia central de la señal EMG en la prueba M2.

### Efecto Tramo

Las frecuencias central y media de la señal EMG son los parámetros que presentan más variaciones estadísticamente significativas durante la prueba. Estas variaciones aparecen en los tramos iniciales y finales de las pruebas estando relacionadas con procesos de adaptación al esfuerzo y fatiga muscular, respectivamente. En la tabla 7.5 se observa los niveles de significación de dichos parámetros en las variables transformadas obtenidas mediante los contrastes Diferencia y Helmert en ambas pruebas.

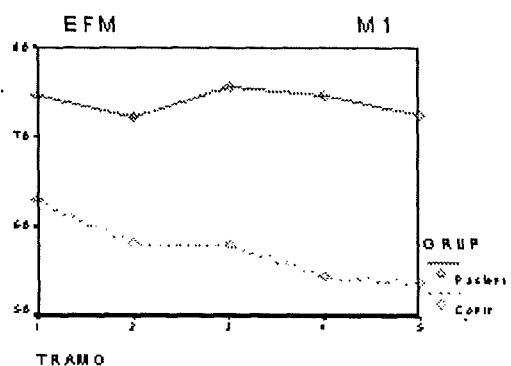


Figura 7.33: Frecuencia media de la señal EMG en la prueba M1.

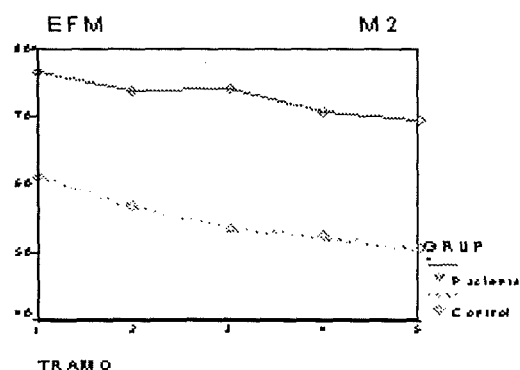


Figura 7.34: Frecuencia media de la señal EMG en la prueba M2.

Parámetro	Prueba M1		Prueba M2	
	Tramos 1-2345	Tramos 1234-5	Tramos 1-2345	Tramos 1234-5
Efc	$p < 0.01$	$p < 0.02$	$p < 0.02$	$p < 0.003$
Efm	$p < 0.008$	$p < 0.03$	$p < 0.005$	$p < 0.004$
Ehl	$p < 0.09$	$p < 0.003$	$p < 0.02$	$p < 0.04$

Tabla 7.5: Niveles de significación para los parámetros frecuenciales de la señal EMG de las variables transformadas obtenidas mediante contraste Helmert (tramo 1 con promedio de los tramos 2 a 5) y Diferencia (promedio de tramos 1 a 4 con tramo 5).

En la señal VMG, la frecuencia media es el parámetro que muestra mayores diferencias entre tramos durante toda la prueba ( $p < 0.04$  y  $p < 0.03$  en las pruebas M1 y M2, respectivamente, mediante análisis univariable promediado).

#### Efecto Grupo\*Tramo

La prueba M1 presenta más parámetros con variaciones entre tramos que diferencian las poblaciones que la prueba M2. Relacionado con la posible fatiga muscular entre los tramos 4 y 5 destacan los parámetros Ehl y Vhl en la prueba M1 ( $p < 0.04$  en ambos casos) como puede observarse en las figuras 7.35 y 7.37.

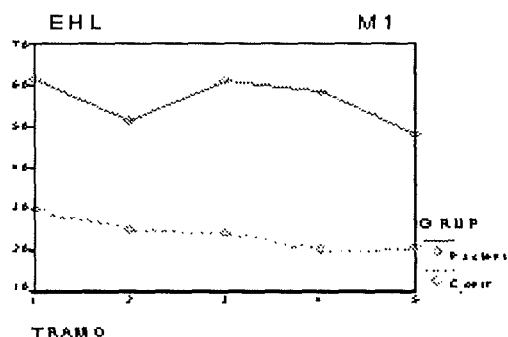


Figura 7.35: Ratio HL de la DEP de la señal EMG en la prueba M1.

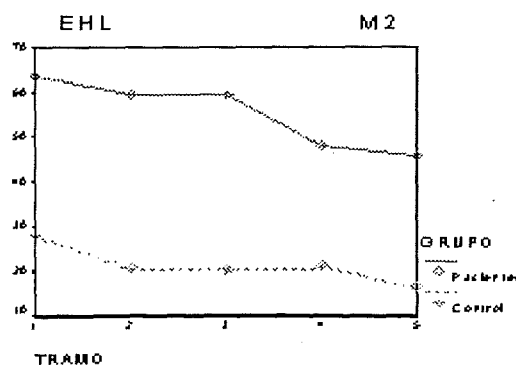


Figura 7.36: Ratio HL de la DEP de la señal EMG en la prueba M2.

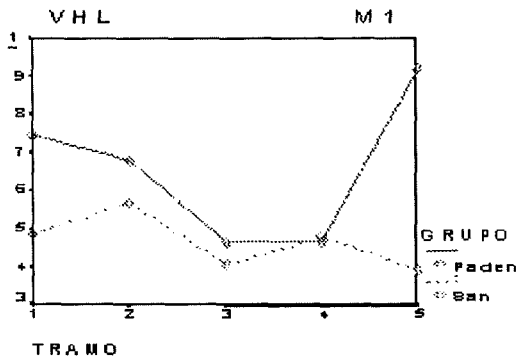


Figura 7.37: Ratio HL de la DEP de la señal VMG en la prueba M1.

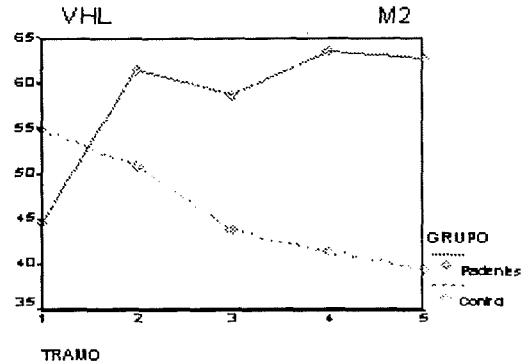


Figura 7.38: Ratio HL de la DEP de la señal VMG en la prueba M2.

Las tendencias entre los tramos 4 y 5 no aparecen en la prueba M2 (ver figuras 7.36 y 7.38). Sin embargo, existe nivel de significación estadística en la variación entre los tramos 3 y 4 del ratio H/L (ver figura 7.36), y tramo 3 con promedio de los tramos 4 y 5 de la frecuencia central en la señal EMG con  $p < 0.01$  en ambos casos (ver figura 7.32).

Diferentes parámetros cuyas variaciones entre los intervalos inicial y central de la prueba M1 presentan nivel de significación estadística son los siguientes:

- Señal EMG: frecuencia central y frecuencia media entre el tramo 1 y promedio de los tramos 2 y 3 ( $p < 0.07$  y  $p < 0.02$ , respectivamente), como puede observarse en la figuras 7.31 y 7.33, respectivamente.
- Señal VMG: frecuencia central entre tramo 1 y promedio de los tramos 2 a 4 ( $p < 0.06$ ).

En la prueba M2, es el ratio H/L en la señal VMG cuya variación entre el tramo 1 y promedio de los siguientes presenta diferencia estadísticamente significativa en la fase inicial del test ( $p < 0.02$ ), como muestra la figura 7.38.

Por último, señalar que la variable transformada procedente del contraste parabólico del ratio H/L en la señal VMG de la prueba M1 presenta un nivel de significación estadística de  $p < 0.04$  (ver figura 7.37).

#### 7.3.4.2 Selección de parámetros

De igual forma que en las pruebas incrementales, se pretende seleccionar los parámetros con mayor interpretación fisiológica que mejor expliquen el comportamiento de los individuos durante la prueba, y que mejor diferencien entre los dos grupos según los resultados anteriores.

En la tabla 7.6 se muestra los parámetros seleccionados en ambas pruebas mantenidas, juntamente con su nivel de significación estadística.

	Prueba M1		Prueba M2	
Parámetros temporales	Erms4-5	p<0.05	VErmx34-5	p<0.04
	Vrms4-5	p<0.2	Vrme34-5	p<0.04
			Erme1-234	p<0.01
			Erms.lin	p<0.02
			Erme.lin	p<0.03
			Vrme.lin	p<0.03
			Tins234	p<0.07
	Tins12-34	p<0.07	Tins1-234	p<0.005
	Tinsciclo2-3	p<0.01		
Parámetros de correlación	lpi234	p<0.07	lpi23	p<0.005
	lpi3-4	p<0.03	lppar34-5	p<0.06
	arplpi12-34	p<0.09	lp1-2	p<0.05
			lp_par	p<0.05
Parámetros frecuenciales	Efc	p<0.01	Efc	p<0.02
	Efm	p<0.03	Efm	p<0.01
	Ehl	p<0.08	Ehl	p<0.06
	Ehl4-5	p<0.04	Efc3-45	p<0.01
	Vhl4-5	p<0.04	Ehl3-4	p<0.01
	Efc1-23	p<0.07		
	Efm1-23	p<0.02		
	Vfc1-234	p<0.06	Vhl1-2345	p<0.02
	Vhl_par	p<0.04		

Tabla 7.6: *Parámetros seleccionados en las pruebas mantenidas con el nivel de significación estadística.*

### 7.3.4.3 Conclusiones

Mediante el efecto tramo se observa variaciones de los parámetros durante la prueba en los tramos iniciales y finales estando relacionadas con procesos de adaptación al esfuerzo y fatiga muscular, respectivamente. En la fase inicial se obtiene variaciones significativas en los parámetros frecuenciales. Sin embargo, es en la segunda parte donde se observa más significación en los tres tipos de parámetros debido al alto esfuerzo que han de realizar los sujetos.

Con el objeto de diferenciar poblaciones se ha encontrado dos tipos de parámetros calculados mediante el promedio durante toda la prueba, que presentan una clara significación estadística en ambas pruebas:

- *Parámetros frecuenciales de la señal EMG.*

Indican un mayor rango de componentes frecuenciales de la DEP de la señal EMG en los pacientes con EPOC respecto a los sujetos sanos, posiblemente mostrando una mayor utilización de fibras musculares, especialmente fibras FT.

- *Tiempos de retardo medio para retardos positivos normalizado respecto a la duración del período inspiratorio.*

El valor de los parámetros de correlación viene determinado por un mayor retardo de la actividad de la señal VMG respecto a la EMG indicando un desacoplamiento entre actividad eléctrica y mecánica. La normalización del TRM respecto a la duración del período inspiratorio permite eliminar el efecto de los diferentes ritmos respiratorios de cada individuo e indica que la actividad de la señal VMG está ubicada durante el período espiratorio en mayor medida en el grupo de pacientes. El mantenimiento de estos patrones temporales de la señal VMG en ambas pruebas reflejan que no depende del posible aprendizaje del sujeto.

La duración del período inspiratorio respecto a la duración del ciclo respiratorio es menor en ambas pruebas en pacientes que en sujetos sanos, aunque no es estadísticamente significativo. Es posible que con un número mayor de sujetos se alcance niveles de significación mayores. Además, los pacientes reducen el período de inspiración al ser sometidos inicialmente a un esfuerzo, mientras que en el grupo control esta reducción aparece avanzada la prueba entre los intervalos central y final.

Es destacable los niveles de significación obtenidos en la variación de parámetros temporales y frecuenciales entre los dos tramos finales de las pruebas, relacionada con la fatiga del músculo ECM:

- Los parámetros procedentes del RMS de la señal EMG y VMG indican que el efecto de la fatiga muscular respiratoria en ambas pruebas produce una pérdida de la eficiencia mecánica-eléctrica mayor en el grupo de pacientes. Mediante el cociente entre el RMS de ambas señales se observa que para una cierta actividad mecánica es requerida mayor actividad eléctrica para soportar el tramo final del test respiratorio.
- Especialmente en la prueba M1, la señal EMG presenta un desplazamiento de la DEP hacia frecuencias inferiores en los enfermos que puede ser debida a una fatiga de las fibras FT. El desplazamiento hacia frecuencias superiores de la DEP de la señal VMG en el grupo de pacientes puede estar reflejando un aumento en la rigidez de las fibras y, por tanto, en la frecuencia de vibración de éstas.

También se observa algunas variaciones de parámetros frecuenciales al inicio de las pruebas que diferencian significativamente ambas poblaciones. Ello hace pensar que la adaptación a la prueba produce un diferente comportamiento del músculo entre pacientes y sujetos sanos. Dichas diferencias se observan en mucha menor medida en la prueba M2 debido, posiblemente, a un aprendizaje de los individuos durante la prueba M1.

### 7.3.5 Discusión y conclusiones

El MANOVA de medidas repetidas se ha aplicado en los datos procedentes de los parámetros calculados en los diferentes protocolos ventilatorios. Mediante diversos efectos o factores se ha analizado de forma exhaustiva, cuantificado e interpretado la evolución de

la función del músculo ECM durante las pruebas. Además, se ha observado importantes diferencias entre el grupo de pacientes y sujetos sanos a partir, tanto de la observación de los resultados multivariantes, como de la interpretación de todos los resultados estadísticos de cada una de las variables transformadas. Las conclusiones extraídas a partir de los diferentes factores son las siguientes:

- *Efecto Tramo.*

Se ha cuantificado los siguientes procesos de la actividad muscular:

- La tendencia lineal que presentan los parámetros durante las pruebas incrementales.

Dichas tendencias son creciente para los temporales, y decreciente para los frecuenciales y de correlación. Los dos primeros, en la prueba IR, muestra diferencias en dichas tendencias entre los pacientes y el grupo control en la señal EMG. De esta forma, el progresivo aumento de la carga en el test produce diferencias en el comportamiento muscular entre ambas poblaciones. Este fenómeno no se produce en la prueba IC debido a su tendencia más uniforme que la IR por la ausencia de períodos de descanso entre cargas. Este efecto tramo ha permitido separar en tres intervalos, inicial, central y final, las diferentes pruebas mediante la agrupación de tramos.

- Adaptación al esfuerzo y fatiga muscular.

En las pruebas mantenidas se ha detectado variaciones más significativas de los parámetros al inicio y al final de éstas. Ello se explica considerando que los individuos, al principio de la prueba, han de verse afectados por un efecto de adaptación al esfuerzo respiratorio que son sometidos al utilizarse una carga elevada correspondiente al 80% de la MCS. También aparece la fatiga muscular al final del protocolo incremental.

- *Efecto Grupo\*Tramo.*

Se observa en el protocolo mantenido que los procesos de adaptación y de fatiga afectan de diferente forma al grupo de pacientes y al grupo control. El primer proceso se detecta mediante parámetros frecuenciales, mientras que el segundo está indicado por parámetros procedentes del RMS y frecuenciales:

- Parámetros temporales.

Muestran una pérdida de la eficiencia mecánica-eléctrica con la patología ya que los pacientes requieren de mayor energía de la señal EMG para valores determinados de señal VMG.

- Parámetros frecuenciales.

Indican un desplazamiento hacia bajas frecuencias de la señal EMG como signo de fatiga, y hacia mayores frecuencias en la señal VMG debido a la rigidez del músculo.



La diferencia entre poblaciones en parámetros al inicio y final de los tests se detecta también en las pruebas incrementales mediante los tres tipos de parámetros.

- *Efecto Grupo.*

Se evalúa la importancia de valores promedio del parámetros durante toda la prueba. Se ha detectado diferencias significativas entre ambas poblaciones en todos los protocolos con los siguientes parámetros:

- Duración del período inspiratorio.

Es menor en pacientes que en sujetos sanos, aunque sólo se alcanza niveles estadísticamente significativos en la prueba M2, así como dicho parámetro normalizado respecto a la duración total del ciclo. Además, esta duración aumenta en menor medida en pacientes ante cargas próximas a la MCS en los protocolos incrementales e incluso disminuye al inicio del protocolo mantenido mientras que los sujetos sanos aumentan su ritmo inspiratorio en instantes más retrasados de la prueba. Por tanto, es importante recomendar que el ritmo respiratorio no sea marcado al sujeto desde el exterior durante la prueba, excepto lo estrictamente necesario para que siga soportando la carga, para que cada individuo los imponga según su capacidad.

- Tiempo de retardo medio de la FCC para retardos positivos.

Ha de estar normalizado respecto a la duración del período de inspiración para que no dependa del diferente ritmo respiratorio de cada individuo, y además tenga como referencia de tiempo el inicio del período de espiración. De esta manera, se observa un mayor retardo en pacientes con EPOC entre la energía de la señal VMG respecto a la EMG debido a una mayor actividad durante el período espiratorio.

- Parámetros frecuenciales de la señal EMG.

La señal EMG presenta mayores componentes frecuenciales en enfermos que en sujetos sanos en todas las pruebas debido, seguramente, a una mayor utilización y activación de las fibras musculares, especialmente de las fibras FT que por la fatiga produce desplazamientos hacia frecuencias inferiores. En este efecto en los parámetros frecuenciales se obtiene mayor significación estadística en la frecuencia central y media, mientras que el ratio H/L indica mejor el proceso de fatiga al final de las pruebas mantenidas.

- *Efecto Ciclo*

Se ha evaluado en los cinco ciclos respiratorios de cada tramo y al no ser significativo se ha reducido el número de datos calculando el valor medio del parámetro.

También se ha reducido el número de datos al seleccionar las variables y sus variaciones que diferencian en mayor medida ambas poblaciones en estudio. De esta forma se acondicionan los datos para la obtención final en análisis posteriores de una función discriminante que separe ambos grupos como ayuda al diagnóstico.

## 7.4 Análisis factorial

### 7.4.1 Introducción

El análisis factorial es una técnica para representar un conjunto de numerosas variables originales u observadas en un espacio de dimensión reducida e interpretar las relaciones entre ellas. Se impone que este nuevo conjunto de variables o factores creado mediante combinación lineal de las variables iniciales suponga una pérdida de información en cantidad pequeña y conocida [Ferrán, 1996]. Con el análisis factorial, se investiga el número de factores subyacentes y, en muchos casos, se identifica la representación conceptual de los factores. Adicionalmente, se calcula puntuaciones factoriales para cada individuo que son utilizadas en un análisis discriminante posterior. Se pretende representar las correlaciones observadas entre las variables originales usando el menor número de factores posibles, y que éstos sean significativos. Conseguir un modelo simple e interpretable es básico para la claridad de resultados y obtención de conclusiones.

En el presente estudio se analiza las variables seleccionadas en los Apartados 7.3.3.2 y 7.3.4.2 mediante el MANOVA de medidas repetidas. Para cada uno de los protocolos ventilatorios se estudia las relaciones entre las variables, se genera nuevos factores y se representa los individuos mediante dichos factores. Finalmente se compara los resultados entre las pruebas.

Las variables seleccionadas de los protocolos son adecuadas para el análisis al ser de tipo cuantitativo con la posibilidad de calcular los coeficientes de correlación de Pearson.

### 7.4.2 Modelo de análisis de la actividad muscular respiratoria

#### 7.4.2.1 Método de componentes principales

En el presente estudio se utiliza el análisis de componentes principales como solución directa del análisis factorial a partir de la matriz de correlación, ya que sus características son las que mejor se adaptan a los objetivos planteados en este proyecto calculando los factores en orden decreciente de la información contenida. Así, los primeros factores o componentes son los que tienen mayor interés. Además, se ha de destacar por su mayor claridad matemática y en la interpretabilidad de los factores [Norusis, 1994].

Este análisis pretende obtener un reducido número de nuevas variables a partir de las observadas, tal que la pérdida de información que tengamos sea mínima. Esto se realiza gracias a la existencia de correlaciones entre las variables, debido a información redundante y a la cual representan los factores comunes.

El modelo lineal a calcular en la descripción de los datos es el siguiente:

$$\begin{aligned}\bar{x}_1 &= a_{11} \cdot \bar{F}_1 + a_{12} \cdot \bar{F}_2 + \dots + a_{1m} \cdot \bar{F}_m + d_1 \cdot \bar{U}_1 \\ \bar{x}_2 &= a_{21} \cdot \bar{F}_1 + a_{22} \cdot \bar{F}_2 + \dots + a_{2m} \cdot \bar{F}_m + d_2 \cdot \bar{U}_2 \\ &\vdots \\ \bar{x}_p &= a_{p1} \cdot \bar{F}_1 + a_{p2} \cdot \bar{F}_2 + \dots + a_{pm} \cdot \bar{F}_m + d_p \cdot \bar{U}_p\end{aligned}$$

donde  $m \leq p$ , y

- $\bar{x}_1, \dots, \bar{x}_p$ : Son  $p$  variables seleccionadas en el MANOVA de medidas repetidas para los  $n$  individuos.
- $\bar{F}_1, \dots, \bar{F}_m$ : Son  $m$  factores comunes correspondiente a las incógnitas del modelo.
- $a_{ij}$ : Se denominan saturaciones, y  $a_{ij}^2$  es la parte de la varianza de  $\bar{x}_i$  debida a la contribución del factor común  $\bar{F}_j$ .
- $\bar{U}_j$ : Son  $p$  factores únicos.
- $d_j$ : Se denominan unicidades, y  $d_j^2$  es la parte de la varianza que se debe a la contribución  $\bar{U}_j$ .

Además:

- $h_i^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2$ : reciben el nombre de comunalidad y representa la contribución de los  $m$  factores a la varianza de  $\bar{x}_i$ .
- Los factores comunes y los únicos son medidas reducidas<sup>1</sup>, de media nula y varianza unidad, y están incorrelados.

Por tanto, las variables se representan mediante los factores comunes calculados en el análisis y los factores únicos que no se superponen entre variables observadas. El conjunto de dichas variables se dispone en una matriz de datos  $\mathbf{X}$  cuyas  $p$  columnas son las variables originales reducidas para cada uno de los  $n$  individuos, correspondiente a las filas. Estas variables pertenecen a un espacio de dimensión  $\max\{n, p\}$ . El objetivo es que los vectores columna de las variables pertenezcan a un espacio de dimensión menor al número de variables. Los ejes de coordenadas de este espacio reducido son los factores comunes, y las variables pueden ser expresadas linealmente en términos de estos factores. Al hacer tal cambio de coordenadas debe recordarse que la configuración geométrica, es decir, la nube de puntos, permanece inalterada [Harman, 1967].

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix} = (\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_p)$$

y se define las componentes principales:

$$\begin{aligned} \bar{F}_1 &= b_{11}\bar{x}_1 + b_{21}\bar{x}_2 + b_{31}\bar{x}_3 + \dots + b_{p1}\bar{x}_p \\ \bar{F}_2 &= b_{12}\bar{x}_1 + b_{22}\bar{x}_2 + b_{32}\bar{x}_3 + \dots + b_{p2}\bar{x}_p \\ &\vdots \\ \bar{F}_m &= b_{1m}\bar{x}_1 + b_{2m}\bar{x}_2 + b_{3m}\bar{x}_3 + \dots + b_{pm}\bar{x}_p \end{aligned} \quad \text{con varianza decreciente.}$$

Siendo la primera componente principal con varianza máxima equivalente a:

<sup>1</sup>Variables reducidas: Cociente entre la variable menos su media, y su desviación estándar.

$$\bar{F}_1 = \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ \vdots \\ f_{n1} \end{bmatrix} = \begin{bmatrix} b_{11}x_{11} + b_{21}x_{12} + b_{31}x_{13} + \dots + b_{p1}x_{1p} \\ b_{11}x_{21} + b_{21}x_{22} + b_{31}x_{23} + \dots + b_{p1}x_{2p} \\ b_{11}x_{31} + b_{21}x_{32} + b_{31}x_{33} + \dots + b_{p1}x_{3p} \\ \vdots \\ b_{11}x_{n1} + b_{21}x_{n2} + b_{31}x_{n3} + \dots + b_{p1}x_{np} \end{bmatrix} = \mathbf{X} \cdot \bar{\mathbf{b}}_1$$

y de esta forma las restantes componentes son:

$$\mathbf{F} \equiv (\bar{F}_1 \bar{F}_2 \dots \bar{F}_m) = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ x_{31} & x_{32} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ b_{31} & b_{32} & \dots & b_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pp} \end{bmatrix} = \mathbf{X} \cdot \mathbf{B}$$

Las varianzas de las componentes principales son las siguientes:

$$\begin{aligned} \text{var}(\bar{F}_1) &= \bar{F}_1^T \cdot \bar{F}_1 = (\mathbf{X} \cdot \bar{\mathbf{b}}_1)^T (\mathbf{X} \cdot \bar{\mathbf{b}}_1) = \bar{\mathbf{b}}_1^T \mathbf{X}^T \mathbf{X} \bar{\mathbf{b}}_1 = \bar{\mathbf{b}}_1^T \mathbf{R} \bar{\mathbf{b}}_1 \\ \text{var}(\bar{F}_2) &= \bar{F}_2^T \cdot \bar{F}_2 = (\mathbf{X} \cdot \bar{\mathbf{b}}_2)^T (\mathbf{X} \cdot \bar{\mathbf{b}}_2) = \bar{\mathbf{b}}_2^T \mathbf{X}^T \mathbf{X} \bar{\mathbf{b}}_2 = \bar{\mathbf{b}}_2^T \mathbf{R} \bar{\mathbf{b}}_2 \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ \text{var}(\bar{F}_m) &= \bar{F}_m^T \cdot \bar{F}_m = (\mathbf{X} \cdot \bar{\mathbf{b}}_m)^T (\mathbf{X} \cdot \bar{\mathbf{b}}_m) = \bar{\mathbf{b}}_m^T \mathbf{X}^T \mathbf{X} \bar{\mathbf{b}}_m = \bar{\mathbf{b}}_m^T \mathbf{R} \bar{\mathbf{b}}_m \end{aligned}$$

Como los vectores  $\bar{\mathbf{b}}_i$  son desconocidos, el problema radica en buscar una función cuadrática a la que se impone dos condiciones:

1. La varianza de los vectores  $\bar{\mathbf{b}}_i$  sea máxima.
2. Los vectores  $\bar{\mathbf{b}}_i$  estén normalizados, es decir,  $\|\bar{\mathbf{b}}_i\| = \bar{\mathbf{b}}_i^T \cdot \bar{\mathbf{b}}_i = 1$

Se trata pues de un problema de máximos condicionados que se resuelve por medio de operadores de Lagrange [Mallo, 1985] mediante:

$$\begin{aligned} \phi &= \bar{\mathbf{b}}_i^T \mathbf{R} \bar{\mathbf{b}}_i - \lambda (\bar{\mathbf{b}}_i^T \cdot \bar{\mathbf{b}}_i - 1) \\ &\quad \downarrow \\ \frac{\delta \phi}{\delta \bar{\mathbf{b}}_i} &= 2 \cdot \mathbf{R} \cdot \bar{\mathbf{b}}_i - 2 \cdot \lambda \cdot \bar{\mathbf{b}}_i = 0 \\ &\quad \downarrow \\ \mathbf{R} \cdot \bar{\mathbf{b}}_i &= \lambda_i \cdot \bar{\mathbf{b}}_i \end{aligned}$$

De esta forma la transformación  $\mathbf{R}$  sobre el vector  $\bar{\mathbf{b}}_i$  corresponde a un nuevo vector colineal con el original. Existe infinitas soluciones, pero todas son proporcionales o paralelas, y sólo se escoge las independientes como componentes principales. La solución, por tanto, consiste en encontrar valores propios ( $\lambda_i$ ) y vectores propios normalizados ( $\bar{\mathbf{b}}_i$ ) de una matriz cuadrada y simétrica correspondiente a la matriz de correlación de las variables reducidas ( $\mathbf{R}$ ).

La variabilidad explicada por el factor común es el valor propio, como demuestra la siguiente expresión:

$$\text{var}(\bar{F}_i) = \bar{F}_i^T \cdot \bar{F}_i = \bar{\mathbf{b}}_i^T \mathbf{R} \bar{\mathbf{b}}_i = \bar{\mathbf{b}}_i^T \cdot \lambda_i \cdot \bar{\mathbf{b}}_i = \lambda_i \cdot \bar{\mathbf{b}}_i^T \cdot \bar{\mathbf{b}}_i = \lambda_i$$

Los componentes principales o factores comunes  $\bar{F}_i$  forman una base ortonormal y los vectores  $\bar{b}_i$  varían la combinación lineal de las variables originales seleccionadas en el MANOVA de medidas repetidas para formar la componente principal  $\bar{F}_i$ .

En las soluciones indirectas, resultantes de una transformación de la solución obtenida de forma directa, el subespacio que forman los factores comunes o componentes principales se transforma mediante una rotación ortogonal u oblicua. La primera tiene la característica de mantener los factores independientes, mientras que en la rotación oblicua se pierde esa independencia a cambio de una mayor interpretabilidad o representatividad de dichos factores. En la presente tesis se escoge la rotación ortogonal Varimax que minimiza el número de variables con saturaciones<sup>2</sup> altas en cada factor, es decir, intenta identificar cada factor común con el menor número de variables originales posibles y que éstas sean las mejor representadas en él. De esta forma, se ayuda a la interpretación de las componentes.

Para la representación de los individuos en el subespacio de las componente principales, se utiliza el método de Anderson-Rubin para estimar los coeficientes de sus coordenadas denominados puntuaciones [Norusis, 1994]. Este método obtiene puntuaciones resultantes de media nula y desviación estándar unidad. Además, minimiza la suma de cuadrados ( $SC$ ) de los factores únicos sobre el rango de las variables y la ausencia de correlación entre los coeficientes permite obtener mejores resultados en el análisis discriminante posterior.

#### 7.4.2.2 Condiciones de aplicación

Para una correcta utilización del método de componentes principales, se comprueba previamente las siguientes hipótesis:

- *La matriz de correlación de las variables observadas reducidas ( $\mathbf{R}$ ) no se ha de aproximar a la matriz identidad.*

Si tiende a la matriz identidad, indica que las variables están poco correladas y se habría de reconsiderar el utilizar el modelo factorial. Esta hipótesis se prueba mediante el test de esfericidad de Bartlett, basado en una transformación Chi-cuadrado del determinante de la matriz  $\mathbf{R}$ . Cuando se rechaza la hipótesis que la matriz es una matriz identidad ( $p < 0.05$ ) indica que el análisis factorial es adecuado.

- *Las correlaciones entre parejas de variables son explicadas por el resto de variables.*

Se prueba mediante la medida de adecuación de la muestra de Kaiser-Meyer-Olkin (KMO). Este es un índice que compara la magnitud de los coeficientes de correlación observados con la magnitud de los coeficientes de correlación parcial. Además, se calcula una medida de la adecuación de la muestra para cada variable individualmente de manera similar, incluyendo en la suma para el cálculo de la magnitud de los coeficientes de correlación sólo aquellos en que está implicada la variable. Cuanto mayor sea este índice, cuyo valor máximo es la unidad, más aconsejable resulta realizar un análisis de componentes principales.

<sup>2</sup>Saturaciones: Proyecciones de las variables originales sobre los factores comunes extraídos.

- *Valores de comunalidad de las variables elevados.*

La comunalidad coincide con el cuadrado del coeficiente de correlación múltiple de la variable observada con el resto de las variables, y su valor máximo es la unidad. Si es pequeño, indica que el nivel de asociación lineal entre las variables es pequeño y sugiere que debería ser eliminado del conjunto de variables analizado.

### 7.4.2.3 Cálculos realizados

En la realización del análisis factorial se desarrolla las siguientes etapas:

#### 1. *Selección de las variables de mayor interés en el análisis factorial.*

Cuando el número de variables seleccionadas en el MANOVA de medidas repetidas no es menor al número de individuos, doce, se elimina del estudio las variables por su baja comunalidad o interpretación en la funcionalidad del músculo. De esta forma, aunque para la extracción de componentes no es requisito que el número de variables no supere el tamaño de la población, sí se necesita para la aplicación del test de esfericidad de Bartlett y el cálculo del índice KMO que informa de la adecuación del análisis factorial en la muestra.

#### 2. *Evaluación de las condiciones de aplicación.*

En primer lugar se realiza la inspección de las correlaciones entre las variables. Si los coeficientes de correlación de Pearson son pequeños es improbable que existan factores comunes entre las variables observadas. Además, la matriz de correlación es la primera aproximación al estudio de las relaciones entre las variables seleccionadas en el MANOVA de medidas repetidas. Se aplica el test de esfericidad de Bartlett y, por último, se determina la adecuación de la muestra mediante el índice KMO.

#### 3. *Extracción de las componentes principales.*

En el análisis de componentes principales cada nueva componente explica menos correlación que la anterior. Por tanto, se escoge el número de componentes basándose en la varianza acumulada explicada por las componentes elegidas. La extracción de los factores comunes de las variables seleccionadas en los Apartados 7.3.3.2 y 7.3.4.2 mediante el análisis de la varianza se realiza siguiendo dos conceptos o criterios:

- Se calcula las componentes principales de todo el conjunto de variables seleccionadas a partir del MANOVA de medidas repetidas en la prueba. Según este criterio se realiza dos estudios diferenciados según el número de componentes extraídas:
  - Número de componentes principales necesarias para acumular una varianza superior al 85% de la total.
  - Dos componentes principales que se utilizan en el análisis discriminante posterior de todo el protocolo completo (cuatro pruebas conjuntamente).

- Extracción de componentes principales en subgrupos procedentes de las variables seleccionadas en el MANOVA de medidas repetidas en una prueba determinada. Dichos subgrupos se determinan agrupando las variables seleccionadas correspondientes a variables calculadas a partir de:
  - Información extraída al inicio de la prueba.
  - Variaciones de parámetros durante la parte intermedia de la prueba.
  - Información extraída al final de la prueba.
  - Variables resultantes del valor medio de un parámetro durante todo el protocolo.
  - Otras variables que sean de interés como las tendencias lineales de un parámetro durante el desarrollo de la prueba.

De esta forma se obtiene una o dos componentes principales de cada subgrupo asociadas a una información específica y de diferente interpretación de la función muscular entre subgrupos (adaptación al esfuerzo, fatiga muscular, etc.). Por consiguiente, mediante un análisis discriminante posterior, se conoce la importancia del tipo de parámetro y de la información de la prueba para diferenciar las dos poblaciones, y las variables que más aportan a cada componente en los diferentes protocolos.

#### 4. Rotación del nuevo subespacio.

En el presente estudio se utiliza la rotación Varimax para conseguir una buena interpretación de los factores. La varianza total acumulada por el modelo no varía aunque se modifica ligeramente la correspondiente a cada componente. Cuando se extrae menos componentes frecuenciales, la primera componente extraída al rotar adquiere más importancia en porcentaje en el modelo.

#### 5. Extracción de las puntuaciones de los individuos en el nuevo subespacio.

Estas son los valores de cada individuo en el nuevo sistema de coordenadas formado por las componentes principales. Se utiliza para observar su agrupación en el nuevo subespacio y para el análisis discriminante posterior.

### 7.4.3 Protocolo incremental

#### 7.4.3.1 Resultados de la prueba IC

#### ANÁLISIS DEL GRUPO DE VARIABLES CONJUNTAMENTE

##### Selección de variables

En la tabla 7.7 se observa las comunalidades que presenta cada una de las variables seleccionadas en el MANOVA de medidas repetidas para la prueba IC (ver Apartado 7.3.3.2) cuando se extrae cinco componentes principales (con el 94.6% de la varianza acumulada).

Variable	Comunalidad	Variable	Comunalidad
Vfm_lin	0.800	lpi	0.942
Efc_lin	0.805	Tinsiclo345-6	0.950
arcp3-4	0.832	Efc5-6	0.956
Erme2345-6	0.833	Ltot0-1	0.966
ltot3-4	0.834	arlp0-1	0.968
arlp4-5	0.877	Efc345	0.992
Efm5-6	0.932	Efm345	0.994

Tabla 7.7: *Comunalidades de las variables seleccionadas en el MANOVA de medidas repetidas en la prueba IC mediante la extracción de cinco componentes principales.*

Se decide descartar las siguientes variables:

- Efc5-6: Posee comunalidad similar a Efm5-6 (0.956 y 0.932, respectivamente), y tienen prácticamente la misma aportación a la 1ª componente principal ( $b_{10,1} = -0.896$  y  $b_{7,1} = -0.887$ , respectivamente). Además, corresponden al mismo fenómeno de desplazamiento de la DEP de la señal EMG. Se descarta este parámetro respecto a Efm5-6 por tener un p-valor mayor en el MANOVA de medidas repetidas.
- Efc\_lin y Vfm\_lin: presentan las comunalidades más pequeñas y la adecuación muestral de ambas variables (0.1) es bastante menor a la obtenida descartándolas junto a Efc5-6 y considerando las once restantes variables (0.4).

#### Condiciones de aplicación

Respecto a la adecuación muestral y la matriz de correlación de las once variables seleccionadas se evalúa los siguientes apartados:

- El índice KMO de la prueba es 0.448. Este resultado es moderadamente satisfactorio, ya que si el modelo es adecuado para los datos en estudio los resultados son coherentes incluso con un valor más bajo.
- El p-valor que contrasta la hipótesis de esfericidad de Bartlett es 0.000. De esta forma, el modelo es adecuado para los datos.

#### Extracción de componentes principales

En la tabla 7.8 se observa las varianzas explicadas por las cuatro primeras componentes principales en el análisis factorial: 49.4%, 15.6%, 11.6% y 10.0% respectivamente (86.7% en total). Cuando éstas son rotadas para interpretar mejor los factores el porcentaje total es el mismo mientras que su distribución en los cuatro factores cambia a: 36.2%, 25.2%, 13.3% y 11.9%.

Al escoger dos factores, la varianza total explicada es de 65.0%, y tras la rotación se reparte en 40.7% para la primera componente y 24.2% para la segunda.



Componente principal	Autovalor		
	Total	% de varianza	%acumulado
1	5.43	49.43	49.43
2	1.71	15.61	65.05
3	1.28	11.69	76.74
4	1.10	10.04	86.78

Tabla 7.8: *Varianzas explicadas por el modelo en la extracción de componentes principales sin rotación.*

En las tablas 7.9 y 7.10 se observa las contribuciones de las variables a cada componente principal ( $\bar{b}_i$ ) en ambos análisis realizados. En la primera columna se ordena los parámetros según importancia decreciente en cada componente. Mediante líneas horizontales se divide los conjuntos de parámetros que más influyen en los diferentes componentes.

Variable	Componente principal rotada			
	1	2	3	4
Efc345	0.974	-0.037	-0.007	0.170
Efm345	0.903	-0.020	-0.106	0.177
lpi	0.827	-0.420	0.021	-0.210
Efm5-6	-0.711	0.410	-0.101	0.033
ltot3-4	0.626	-0.518	0.294	0.306
arcp3-4	0.573	-0.474	0.401	0.336
arlp0-1	-0.158	0.968	0.106	-0.113
ltot0-1	-0.181	0.908	0.225	-0.121
Erme2345-6	0.122	0.074	0.818	-0.192
arlp4-5	-0.481	0.376	0.666	0.164
tinsciclo345-6	0.119	-0.166	-0.125	0.950

Tabla 7.9: *Contribución de cada variable a la componente principal del modelo para la extracción de cuatro componentes rotadas.*

Variable	C. P. rotada	
	1	2
Efc345	0.900	-0.104
lpi	0.841	-0.271
Efm345	0.819	-0.113
ltot3-4	0.806	-0.296
arcp3-4	0.784	-0.206
Efm5-6	-0.772	0.250
ltot0-1	-0.301	0.865
arlp0-1	-0.330	0.841
arlp4-5	-0.302	0.646
Erme2345-6	0.335	0.580
tinsciclo345-6	0.188	-0.411

Tabla 7.10: *Contribución de cada variable a la componente principal del modelo para la extracción de dos componentes rotadas.*

La reducción a sólo dos factores dificulta la interpretación de las componentes principales, ya que todas las variables se distribuyen en ellas para la realización del modelo. Sin embargo, se observa que las variables que contribuyen a la 1ª componente al extraer dos factores son las mismas que la correspondiente a cuatro factores, variando únicamente la cantidad en su contribución. En la representación en el plano factorial para dos factores (ver figura 7.39) se observa dos grupos de variables en el segundo y cuarto cuadrante que contribuyen a cada una de las componentes y tres variables menos asociadas a una única componente por encontrarse relativamente alejadas de los ejes y más cercanas al origen que el resto: Erme2345-6, Efm5-6, y Tinsciclo345-6.

#### Representación de los sujetos en las componentes principales

En la figura 7.40, se observa cómo se distribuyen los individuos en el nuevo subespacio formado por las componentes principales rotadas en el análisis en que se extraen dos

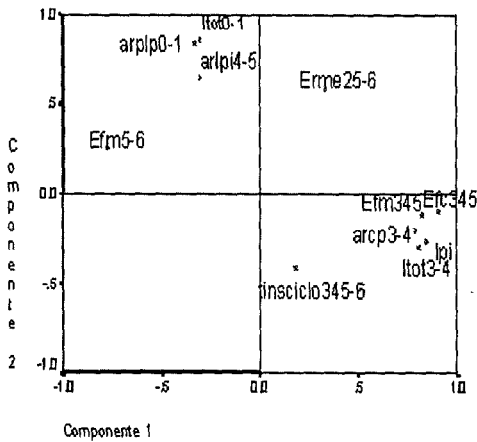


Figura 7.39: Representación de las variables en el plano factorial del análisis para la extracción de dos componentes principales.

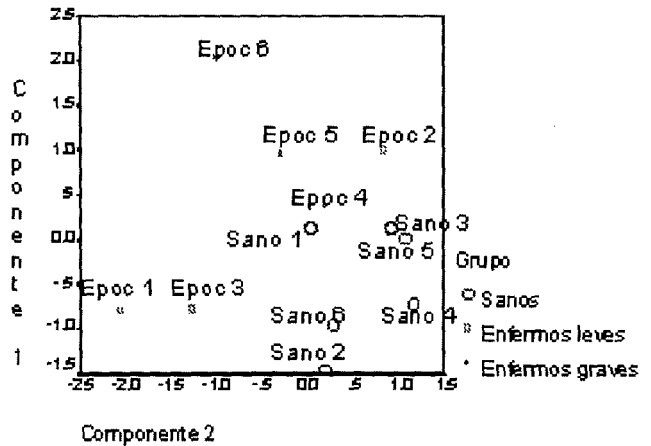


Figura 7.40: Representación de los doce individuos en las dos componentes principales.

componentes. De esta forma se puede observar cómo se agrupan los individuos de ambas poblaciones en estudio. En el caso del primer análisis, al tratarse de cuatro componentes, esta representación se realiza con gráficos bivariantes entre las componentes. En este tipo de gráfico es difícil observar si los individuos se agrupan alrededor de algún punto del subespacio formado por las componentes principales, pues se observa proyecciones sobre cada pareja de componentes principales de la representación de los grupos en el subespacio de factores comunes.

Se observa que en las dos componentes extraídas en el segundo análisis factorial, el grupo control se encuentra agrupado en una zona concreta mientras que el grupo de pacientes se encuentra disperso por el resto del subespacio. Además, se observa que los grupos de pacientes leves y graves se encuentran en regiones diferentes exceptuando el paciente con EPOC número 3.

## ANÁLISIS DE SUBGRUPOS DE VARIABLES SEPARADAMENTE

### Determinación de subgrupos de variables

Los subgrupos de variables formados a partir de las variables seleccionadas en el MANOVA de medidas repetidas en la prueba IC (ver Apartado 7.3.3.2) son los siguientes:

- Variaciones entre apnea y volumen corriente: ltot0-1, arplp0-1.
- Variaciones en la fase media de la prueba: ltot3-4, arcp3-4, arlp4-5.
- Variaciones al final de la prueba: Erme2345-6, Tinsciclo 345-6, Efm5-6, Efc5-6.
- Evolución durante toda la prueba: Efc.lin, Vfc.lin.
- Valores promedio a diferentes cargas: Efc345, Efm345, lpi.

### Inspección de las correlaciones entre variables y extracción de componentes principales

En el subgrupo 'A', el coeficiente de correlación entre ambas variables es elevado: 0.95. Por tanto, es lógico que se extraiga una única componente con una varianza del 97% de la total de la muestra. La componente principal extraída es la siguiente:

$$a1 = 0.988 \text{ arplp0-1} + 0.988 \text{ ltot0-1}$$

Respecto al subgrupo 'B', el coeficiente de correlación entre las variables arcp3-4 y ltot3-4 es 0.710. Sin embargo, la correlación de la tercera variable, arlpi4-5 con las anteriores corresponde a coeficientes pequeños: -0.2 y -0.127, respectivamente. Por consiguiente, se espera que contribuya mucho menos en la componente principal extraída con una varianza del 59% del total de la muestra y que corresponde a la siguiente combinación lineal:

$$b1 = 0.914 \text{ arcp3-4} + 0.895 \text{ ltot3-4} - 0.380 \text{ arlpi4-5}$$

Las componentes principales de los subgrupos 'A' (a1) y 'B' (b1) están relacionadas con el acoplamiento entre la actividad eléctrica y mecánica del músculo al proceder de parámetros de correlación. El primero indica cambios entre apnea y respiración a volumen corriente, y en el caso 'B' de cargas ventilatorias medias a altas.

En la tabla 7.11 se muestra la matriz de correlación de las variables del subgrupo 'C'. La mayor correlación se observa entre los parámetros frecuenciales (coeficiente de 0.82), mientras que el resto tienen valores sensiblemente menores. Además, las otras dos variables corresponden a parámetros temporales. En consecuencia, se extrae las siguientes dos componentes principales rotadas:

$$c1 = 0.956 \text{ Efc5-6} + 0.940 \text{ Efm5-6} + 0.004 \text{ Tinsciclo345-6} + 0.217 \text{ Erme2345-6}$$

$$c2 = 0.186 \text{ Efc5-6} + 0.005 \text{ Efm5-6} - 0.794 \text{ Tinsciclo345-6} + 0.768 \text{ Erme2345-6}$$

	Erme2345-6	tinsciclo345-6	Efm5-6	Efc5-6
Erme2345-6	1.000	-0.236	0.045	0.413
tinsciclo345-6	-0.236	1.000	-0.139	-0.037
Efm5-6	0.045	-0.139	1.000	0.829
Efc5-6	0.413	-0.037	0.829	1.000

Tabla 7.11: Matriz de correlación de las variables del subgrupo 'C'.

Las variables frecuenciales, Efc5-6 y Efm5-6, contribuyen a la 1ª componente (c1), relacionándose con la variación entre cargas ventilatorias muy elevadas de la velocidad de conducción de las fibras musculares. Mientras, las variables calculadas a partir de parámetros temporales contribuyen principalmente a la 2ª componente (c2), asociándose a la diferencia de esfuerzo en parámetros temporales entre cargas medias y elevadas: Tinsciclo345-6 y Erme2345-6.

El coeficiente de correlación entre variables del subgrupo D es muy pequeño, -0.017, siendo lógico al estar calculadas a partir de ambas señales miográficas. En este caso no tiene sentido extraer componentes principales y se decide utilizar los valores originales de las dos variables para representar a los individuos convirtiéndose directamente las variables *Efc\_lin* y *Vfm\_lin* en las componentes *d1* y *d2*, respectivamente.

En el subgrupo E, la correlación entre las variables frecuenciales *Efc345* y *Efm345* es muy alta (coeficiente de 0.93). En cambio, la correlación entre el parámetro *lpi* y los frecuenciales es menor: coeficientes 0.77 y 0.62, respectivamente. Además, corresponde a parámetros de diferentes dominios: temporal y frecuencial. Por tanto, se extrae dos componentes principales que tras la rotación Varimax son las siguientes:

$$e1 = 0.946 Efm345 + 0.843 Efc345 + 0.360 lpi$$

$$e2 = 0.300 Efm345 + 0.515 Efc345 + 0.932 lpi$$

Las dos variables frecuenciales contribuyen principalmente a la primera componente (*e1*), relacionada con valores promedio en frecuencia de la DEP. La variable de la FCC contribuye a la segunda componente (*e2*), pudiéndose relacionar con valores promedio del acoplamiento entre la actividad mecánica y eléctrica del músculo durante toda la prueba.

En resumen, en la tabla 7.12 se observa cómo se reparten las variables en su contribución a las componentes principales y el porcentaje de varianza total explicada en cada modelo de los diferentes subgrupos.

Subgrupo		Total	1ª componente	2ª componente
A	% varianza explicada	97.0	97.0	
Variaciones apnea-vc	Variables que más contribuyen		arlp0-1 ltot0-1	
B	% varianza explicada	59.4	59.4	
Variaciones intermedias	Variables que más contribuyen		arcp3-4 ltot3-4	
C	% varianza explicada	77.5	46.1	31.3
Variaciones finales	Variables que más contribuyen		Efc5-6 Efm5-6	Tinsciclo345-6 Erme2345-6
D	% varianza explicada	...	var. original	var. original
Evolución lineal	Variables que más contribuyen		Vfm_lin	Efc_lin
E	% varianza explicada	98.5	57.8	40.7
Cargas promediadas	Variables que más contribuyen		Efm345 Efc345	lpi

Tabla 7.12: Resultados del análisis de componentes principales para los diferentes subgrupos de variables.

Los parámetros frecuenciales en los casos 'C' y 'E' se agrupan en las primeras componentes, que son las que explican más varianza, mientras los parámetros temporales y de correlación están aportando información a la segunda.

### Representación de los sujetos en las componentes principales

En las dos representaciones de los individuos en los factores comunes extraídos (figura 7.41 para el caso 'C', figura 7.42 para el caso 'E') se observa la necesidad de dos componentes principales para distribuir en dos zonas diferenciadas las poblaciones de pacientes y sujetos sanos. Esto confirma la elección de dos componentes principales para estos dos análisis. También se observa en ambas representaciones que el paciente con EPOC número 3 se encuentra alejado del resto de su grupo de enfermos leves, y que los enfermos leves y graves se localizan en regiones diferentes.

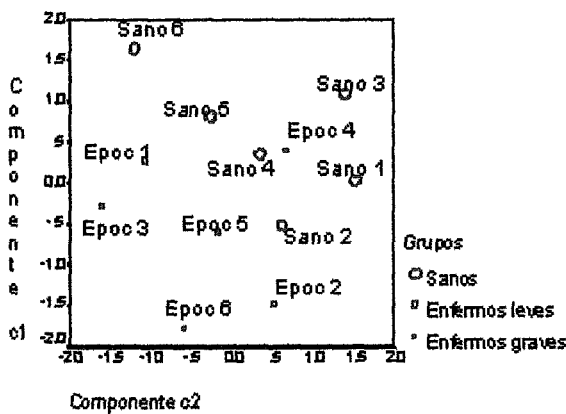


Figura 7.41: Representación de los doce individuos en las dos componentes extraídas en el análisis factorial 'C'.

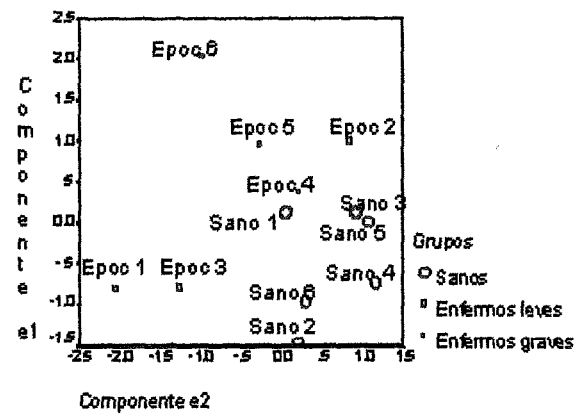


Figura 7.42: Representación de los individuos en las dos componentes extraídas en el análisis factorial 'E'.

#### 7.4.3.2 Resultados de la prueba IR

### ANÁLISIS DEL GRUPO DE VARIABLES CONJUNTAMENTE

#### Selección de variables

En la tabla 7.13 se observa las comunalidades que presenta cada una de las variables seleccionadas en el MANOVA de medidas repetidas para la prueba IR (ver Apartado 7.3.3.2) cuando se extrae cinco componentes principales (con el 96.4% de la varianza acumulada).

Se decide descartar las siguientes variables:

- Vfc345: Tiene la menor comunalidad de todas las variables y, por tanto, queda menos reflejada en el modelo.
- lpi4-5: Es el único parámetro que de forma aislada está relacionado con variaciones entre tramos intermedios de la prueba.
- lni, ln234-56: Resulta más interesantes los TRM para retardos positivos que negativos, ya que se intenta evaluar la cantidad de energía existente en la señal VMG en retraso respecto a la EMG, especialmente durante los períodos espiratorios.

Variable	Comunalidad	Variable	Comunalidad
Vfc345	0.474	lpi	0.887
Erms_lin	0.510	Efc_lin	0.901
Ehl5-6	0.663	Ehl_lin	0.902
Tins2345-6	0.708	Ehl12-345	0.910
lp234-56	0.819	ln234-56	0.920
lpi4-5	0.834	Efc12-345	0.910
lni	0.840	Efm345	0.929
ltot0-1	0.866	Ehl345	0.988
arlp0-1	0.879	Efc345	0.989

Tabla 7.13: *Comunalidades de las variables seleccionadas en el MANOVA de medidas repetidas en la prueba IR mediante la extracción de cinco componentes principales.*

- tins234-56: Se trata de un parámetro en valor absoluto, mientras que en la otra prueba incremental (IC) se escogió normalizado respecto a la duración total del ciclo respiratorio.
- lp234-56: En el MANOVA de medidas repetidas se obtuvo tan sólo un p-valor de 0.08 en el efecto Grupo\*Tramo, el mayor de todos los parámetros seleccionados.
- Ehl12-345: Está asociada a desplazamientos de la DEP de la señal EMG, de igual forma que el parámetro Efc12-345. Sin embargo, éste último presenta un p-valor menor en el MANOVA de medidas repetidas, 0.01, que la variable descartada: 0.03.

#### Condiciones de aplicación

Respecto a la adecuación muestral y la matriz de correlación de las once variables seleccionadas se evalúan los siguientes apartados:

- El índice KMO de la prueba es 0.315. Este resultado no es muy satisfactorio, sobretodo si se compara con el obtenido de la prueba IC, 0.448, pero si el modelo es adecuado para los datos en estudio los resultados son coherentes.
- El p-valor que contrasta la hipótesis de esfericidad de Bartlett es 0.000. De esta forma, el modelo es adecuado para los datos.

#### Extracción de componentes principales

En la tabla 7.14 se observa las varianzas explicadas por las cuatro primeras componentes principales en análisis factorial: 49.8%, 23.0%, 10.7% y 8.7% respectivamente (92.4% en total). Tras la rotación, el porcentaje total es el mismo mientras que su distribución en los cuatro factores cambia a: 38.3%, 24.9%, 19.5%, 9.6%.

Al escoger dos factores, la varianza total explicada es de 72.9%, y tras la rotación se reparte en 40.2% para la primera componente y 32.6% para la segunda. Los porcentajes

Componente principal	Autovalor		
	Total	% de varianza	%acumulado
1	5.48	49.84	49.84
2	2.54	23.09	72.94
3	1.17	10.71	83.65
4	0.96	8.78	92.43

Tabla 7.14: *Varianzas explicadas por el modelo en la extracción de componentes principales sin rotación.*

son muy similares a los obtenidos en el caso de la prueba IC, con la observación que el porcentaje de varianza total explicada por el modelo es ligeramente superior en la prueba IR para ambos análisis factoriales.

En las tablas 7.15 y 7.16 se puede ver las contribuciones de las variables a cada componente principal ( $\bar{b}_i$ ) para la extracción de cuatro y dos factores, respectivamente. En la primera columna se ordena los parámetros según importancia decreciente en cada componente. Mediante líneas horizontales se divide los conjuntos de parámetros que más influyen en los diferentes componentes.

Variable	Componente principal rotada			
	1	2	3	4
Efc345	0.968	0.220	-0.016	-0.016
Ehl345	0.964	0.119	-0.085	0.030
Efm345	0.894	0.296	-0.019	-0.158
lpi	0.821	0.038	-0.412	0.009
Ehl5-6	0.818	0.173	0.054	-0.093
Ehl_lin	0.119	0.956	-0.189	-0.017
Efc12-345	-0.329	-0.898	0.162	-0.025
Efc_lin	0.251	0.824	-0.411	0.165
ltot0-1	-0.099	-0.199	0.966	0.057
arlp0-1	-0.039	-0.345	0.895	-0.122
Erms_lin	-0.107	0.075	-0.037	0.989

Tabla 7.15: *Contribución de cada variable a la componente principal del modelo para la extracción de cuatro componentes rotadas.*

Variable	C. P. rotada	
	1	2
Efc345	0.976	0.146
Ehl345	0.947	0.129
Efm345	0.939	0.152
Ehl5-6	0.841	0.053
lpi	0.774	0.290
Efc_lin	0.303	0.892
arlp0-1	0.035	-0.872
Ehl_lin	0.236	0.806
ltot0-1	-0.058	-0.780
Efc12-345	-0.429	-0.750
Erms_lin	-0.241	0.262

Tabla 7.16: *Contribución de cada variable a la componente principal del modelo para la extracción de dos componentes rotadas.*

De forma similar a la prueba IC, las variables que contribuyen a la 1ª componente al extraer dos factores son las mismas que la correspondiente a cuatro factores, variando únicamente la cantidad en su contribución. En la figura 7.43 se presenta la representación en el plano factorial de las variables en la extracción de dos factores, ya que la interpretación gráfica con cuatro componentes no resulta interesante. Parámetros promedio en el dominio frecuencial y el TRM para retardos positivos se encuentran próximos a la 1ª componente, así como la variación de componentes frecuenciales al final de la prueba. Otras variables como tendencias lineales en parámetros frecuenciales y variación de parámetros

de correlación entre apnea y volumen corriente contribuyen a la segunda componente de forma opuesta entre ellos. El resto de variables, Erms\_lin y Efc12-345, contribuyen de forma desigual a ambas componentes.

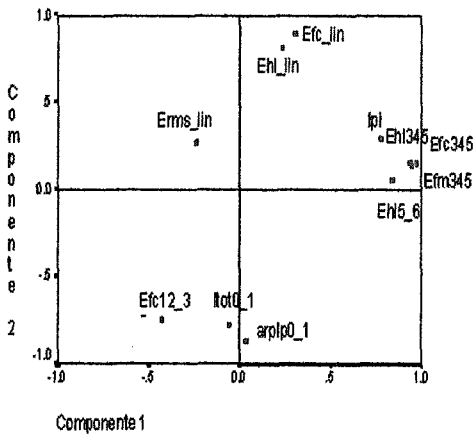


Figura 7.43: Representación de las variables en el espacio factorial de dos componentes principales.

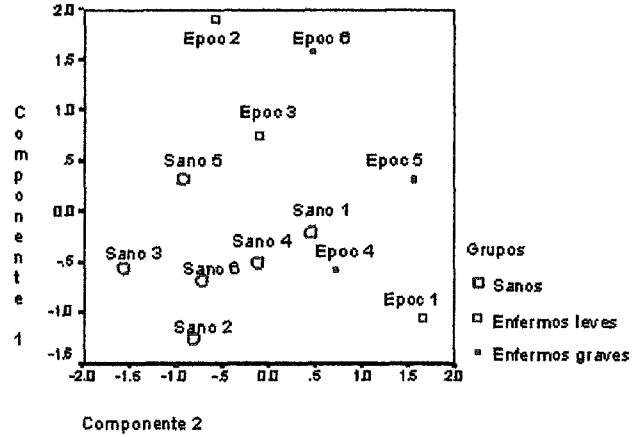


Figura 7.44: Representación de los doce individuos en las dos componentes principales.

Representación de los sujetos en las componentes principales

En la figura 7.44, se observa la distribución de los individuos en el subespacio formado por las dos componentes principales rotadas.

El grupo control está claramente situado en una zona separada del subespacio distinta de la del grupo de pacientes. En cambio, el grupo de pacientes leves no ocupa un lugar muy diferenciado del grupo de pacientes graves. Por consiguiente, se espera una buena función discriminante en la separación de dos poblaciones, pero no muy eficiente en el posterior análisis discriminante para los tres grupos de individuos: sujetos sanos, enfermos leves y graves.

ANÁLISIS DE SUBGRUPOS DE VARIABLES SEPARADAMENTE

Determinación de subgrupos de variables

Los subgrupos de variables formados a partir de las variables seleccionadas en el MANOVA de medidas repetidas en la prueba IR (ver Apartado 7.3.3.2) son los siguientes:

- A. Variaciones entre apnea y volumen corriente: Itot0-1, arplp0-1.
- B. Variaciones en el inicio de la prueba: Efc12-345, Ehl12-345.
- C. Variaciones al final de la prueba: ln234-56, lp234-56, lpi4-5, tins2345-6, Ehl5-6.
- D. Evolución durante toda la prueba: Efc.lin, Ehl.lin, Erms.lin.
- E. Valores promedio a diferentes cargas: Efc345, Efm345, Ehl345, Vfc345, lpi, Ini.



### Inspección de las correlaciones entre variables y extracción de componentes principales

El coeficiente de correlación entre las variables del subgrupo 'A', es alto, 0.923, de igual forma que el obtenido en la prueba IC. Por tanto, también se extrae una única componente con una varianza del 96.1% de la total de la muestra. La componente principal extraída es la siguiente:

$$a1 = 0.981 arplp0-1 + 0.981 ltot0-1$$

En el subgrupo 'B', el coeficiente de correlación entre las variables Ehl12-345 y Efc123-45 es alto, 0.945, y se obtiene una componente que explica el 97.2% de la varianza total:

$$b1 = 0.986 Efc12-345 + 0.895 Ehl12-345$$

La componente principal del subgrupo 'B' (b1) está relacionada con la variación de la velocidad de conducción y desplazamiento de la DEP de la señal EMG al inicio de la prueba ante cargas bajas.

En la tabla 7.17 se muestra la matriz de correlación de las variables del subgrupo 'C'. Se observa la mayor correlación entre los parámetros ln234-56 y lp234-56, coeficiente de 0.746, mientras que el parámetro frecuencial y temporal tienen valores sensiblemente menores alcanzando coeficientes de valor -0.06.

	ln234-56	lp234-56	Ehl5-6	tins234-56	lpi4-5
ln234-56	1.000	0.746	-0.486	-0.087	0.398
lp234-56	0.746	1.000	-0.395	0.158	0.323
Ehl5-6	-0.486	-0.395	1.000	0.158	0.323
tins234-56	-0.087	0.158	-0.064	1.000	0.182
lpi4-5	0.398	0.323	-0.151	0.182	1.000

Tabla 7.17: Matriz de correlación de las variables del subgrupo 'C'.

Respecto al subgrupo 'D', los parámetros frecuenciales presentan la correlación más alta con un coeficiente de 0.905, mientras que la tendencia lineal del Erms presenta bajos coeficientes de correlación con las variables Efc\_lin y Ehl\_lin: 0.202 y 0.045, respectivamente. En la tabla 7.18 se muestra la matriz de correlación de las variables del subgrupo 'E'. Los coeficientes de correlación entre variables frecuenciales son muy altos, superiores a 0.935, mientras que los tiempos de retardo medio presentan coeficientes entre ellos y con los parámetros frecuenciales de valores inferiores.

Por consiguiente, se extrae dos componentes principales en estos tres subgrupos que corresponde a las siguientes combinaciones lineales:

$$c1 = 0.929 \ln234-56 + 0.832 \lp234-56 - 0.682 Ehl5-6 - 0.087 tins2345-6 + 0.458 lpi4-5$$

$$c2 = -0.024 \ln234-56 + 0.216 \lp234-56 + 0.025 Ehl5-6 + 0.927 tins2345-6 + 0.513 lpi4-5$$

	Efm345	Efc345	Ehl345	lpi	lni
Efm345	1.000	0.960	0.984	0.787	0.562
Efc345	0.960	1.000	0.935	0.662	0.494
Ehl345	0.984	0.935	1.000	0.803	0.619
lpi	0.787	0.662	0.803	1.000	0.719
lni	0.562	0.494	0.619	0.719	1.000

Tabla 7.18: Matriz de correlación de las variables del subgrupo 'E'.

$$d1 = 0.980 Ehl\_lin + 0.968 Efc\_lin + 0.061 Erms\_lin$$

$$d2 = -0.021 Ehl\_lin + 0.150 Efc\_lin + 0.997 Erms\_lin$$

$$e1 = 0.948 Efm345 + 0.927 Efc345 + 0.891 Ehl345 + 0.238 lni + 0.566 lpi$$

$$e2 = 0.244 Efm345 + 0.364 Efc345 + 0.429 Ehl345 + 0.938 lni + 0.727 lpi$$

En los subgrupos 'D' y 'E' la primera componente está formada básicamente por los parámetros frecuenciales, mientras que las variables temporales y de correlación se proyectan en el segundo factor, respectivamente. En la tabla 7.19 se presenta un resumen del análisis mediante subgrupos. El modelo de análisis factorial del subgrupo 'C' es el que menos varianza explica y en el que la interpretación de los componentes resulta más compleja, aunque de igual forma que en los subgrupos 'D' y 'E', los parámetros temporales y de correlación se proyectan sobre la 2ª componente.

Subgrupo		Total	1ª componente	2ª componente
A Variaciones apnea-vc	% varianza explicada	96.1	96.1	
	Variables que más contribuyen		ltot0-1 arplp0-1	
B Variaciones iniciales	% varianza explicada	97.2	97.2	
	Variables que más contribuyen		Ehl12-345 Efc12-345	
C Variaciones finales	% varianza explicada	68.1	44.7	23.3
	Variables que más contribuyen		ln234-56 lp234-56 Ehl5-6	Tins2345-6 lpi4-5
D Evolución lineal	% varianza explicada	97.2	63.3	33.9
	Variables que más contribuyen		Efc_lin Ehl_lin	Erms_lin
E Cargas promediadas	% varianza explicada	94.2	58.5	35.6
	Variables que más contribuyen		Efc345 Efm345 Ehl345	lni lpi

Tabla 7.19: Resultados del análisis de componentes principales para los diferentes subgrupos de variables.

Representación de los sujetos en las componentes principales

En las representaciones de los individuos en los factores comunes extraídos para los subgrupos 'C', 'D' y 'E' con dos componentes (ver figuras 7.45 a 7.47) se observa que el grupo de pacientes y de sujetos sanos se encuentran situados en diferentes zonas del subespacio. Respecto a la separación entre enfermos leves y graves con los individuos sanos, la mejor distribución corresponde a la del subgrupo 'E', sin ser altamente satisfactoria, por tanto, se espera mejor clasificación en este caso en el análisis discriminante posterior.

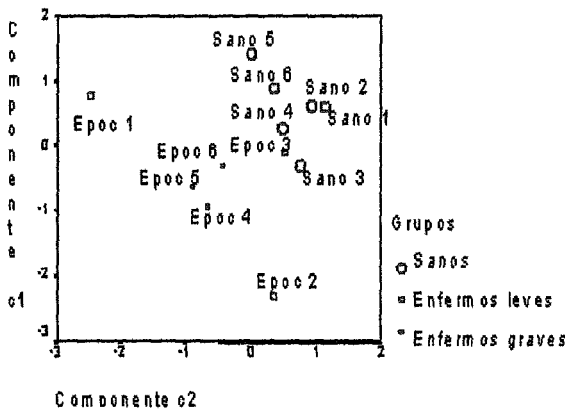


Figura 7.45: Representación de los doce individuos en las dos componentes extraídas en el análisis factorial 'C'.

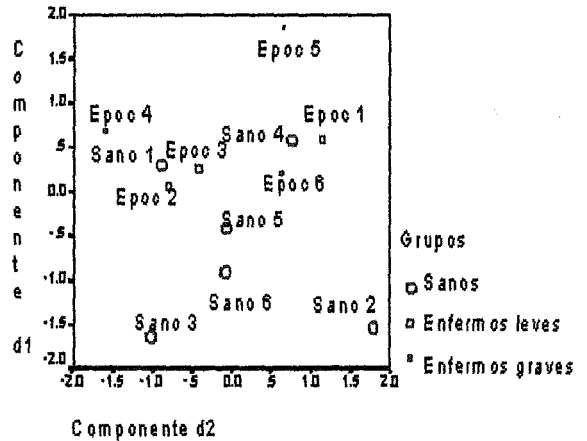


Figura 7.46: Representación de los individuos en las dos componentes extraídas en el análisis factorial 'D'.

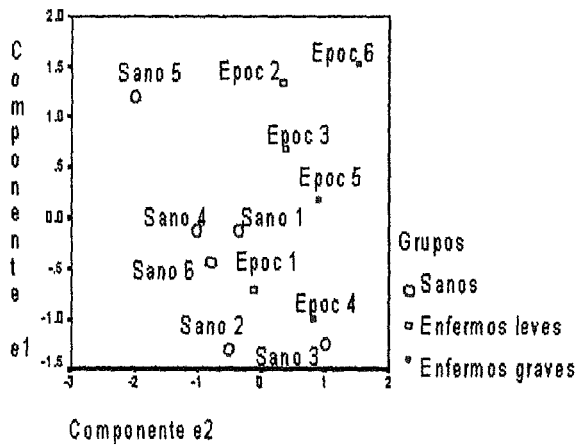


Figura 7.47: Representación de los individuos en las dos componentes extraídas en el análisis factorial 'E'.

### 7.4.3.3 Conclusiones

En el análisis factorial de todo el grupo de variables se observa al comparar las tablas 7.9 y 7.10 con las tablas 7.15 y 7.16 correspondientes a las pruebas IC e IR, respectivamente, los siguientes aspectos:

- La distribución de las variables en las diferentes componentes es muy parecida, exceptuando los parámetros provenientes de los contrastes lineales que fueron descartados en la prueba IC.
- La primera componente en la extracción de cuatro y dos factores en ambas pruebas es prácticamente idéntica mostrando que los parámetros que reúnen mayor cantidad de varianza de la muestra son los valores promedio de las variables frecuenciales y TRM para retardos positivos normalizado respecto a la duración del período inspiratorio.
- En ambas pruebas, también se origina una componente correspondiente únicamente a la variación de parámetros de correlación entre apnea y volumen corriente.

Por consiguiente, se observa una interesante reproducibilidad de los modelos en ambas pruebas incrementales.

Respecto al análisis factorial de subgrupos de variables de forma separada se obtienen las siguientes conclusiones:

- Se ha agrupado las variables seleccionadas en cinco grupos de forma similar en ambas pruebas. El subgrupo 'B' es el único con ciertas diferencias, ya que mientras en la prueba IC corresponde a variaciones intermedias, en la IR se agrupa las variaciones entre el inicio y la mitad de la prueba. Por otra parte, aunque las variables son diferentes en ambos protocolos, la interpretación de las componentes extraídas es similar.
- Los modelos en ambos protocolos coinciden en el número de componentes extraídas: una en los subgrupos 'A' y 'B' y dos en el 'C', 'D' y 'E'.
- Los porcentajes de varianza reunida por las componentes son similares en ambas pruebas a excepción del subgrupo 'B' mencionado anteriormente.
- En cada subgrupo, los parámetros frecuenciales, si existen, se proyectan en la primera componente, mientras que los temporales y de correlación se proyectan en la segunda.

Todas estas semejanzas muestran también la reproducibilidad del modelo cuando el análisis factorial se realiza siguiendo esta metodología de subgrupos separados.

Respecto a la representación de las puntuaciones factoriales de los individuos en los planos factoriales, se ha observado la posibilidad de diferenciar claramente entre el grupo control y el grupo de pacientes al situarse en regiones separadas, y la dificultad de hacerlo entre enfermos graves y leves.

## 7.4.4 Protocolo mantenido

### 7.4.4.1 Resultados de la prueba M1

#### ANÁLISIS DEL GRUPO DE VARIABLES CONJUNTAMENTE

##### Selección de variables

Se selecciona las variables a partir de las comunalidades de las variables seleccionadas en el MANOVA de medidas repetidas para la prueba M1 (ver Apartado 7.3.4.2 y tabla 7.20). Dichas comunalidades se han obtenido a partir de un análisis de cinco componentes principales con el 96.2% de la varianza acumulada.

Variable	Comunalidad	Variable	Comunalidad
lpi234	0.583	Vhl4-5	0.877
arplpi12-34	0.626	Efm1-23	0.896
lpi3-4	0.636	Ehl4-5	0.899
Erms4-5	0.769	Ehl	0.920
Vfc1-234	0.809	Vhl_par	0.922
Efc1-23	0.814	Vrms4-5	0.936
tinsciclo2-3	0.837	Efc	0.965
tins12-34	0.874	Efm	0.983

Tabla 7.20: Comunalidades de las variables seleccionadas en el MANOVA de medidas repetidas en la prueba M1 mediante la extracción de cinco componentes principales.

De igual forma que en las pruebas incrementales, se selecciona once variables descartando las siguientes:

- lpi234, arplpi12-34, lpi3-4, Vfc1-234: Corresponde a las variables con comunalidad menor y, por tanto, que menos quedan explicadas en el modelo de análisis factorial. El parámetro Erms4-5 con baja comunalidad no se descarta porque conjuntamente con la Vrms4-5 están asociadas a la eficiencia mecánica-eléctrica entre energías de ambas señales miográficas.
- Vhl\_par: el contraste parabólico es de difícil interpretación de la función muscular.

##### Condiciones de aplicación

Referente a las condiciones correspondientes a la adecuación muestral y la matriz de correlación de las variables se obtiene los siguientes resultados:

- El índice KMO de la prueba tiene un valor similar al de la prueba IR: 0.332. No resulta demasiado elevado, pero si el modelo es adecuado para los datos en estudio los resultados son coherentes y válidos.
- El p-valor que contrasta la hipótesis de esfericidad de Bartlett es 0.000. De esta forma, el modelo es adecuado para los datos.

## Extracción de componentes principales

En la tabla 7.21 se observa las varianzas explicadas por las cuatro primeras componentes principales en el análisis factorial: 47.7%, 19.2%, 13.2% y 11.1% respectivamente (91.4% en total). Cuando éstas son rotadas para interpretar mejor los factores el porcentaje total es el mismo mientras que su distribución en los cuatro factores cambia a: 35.6%, 20.6%, 18.3% y 16.7%.

Componente principal	Autovalor		
	Total	% de varianza	%acumulado
1	5.25	47.79	47.79
2	2.11	19.20	66.99
3	1.45	13.22	80.22
4	1.23	11.19	91.41

Tabla 7.21: *Varianzas explicadas por el modelo en la extracción de componentes principales sin rotación.*

Al escoger dos factores, la varianza total de la muestra explicada por el modelo es de 66.9%. Tras realizar la rotación, dicha varianza se reparte entre la primera componente, 41.6%, y la segunda: 25.3%.

En las tablas 7.22 y 7.23 se observa las contribuciones de las variables a cada componente principal ( $\bar{b}_i$ ) en ambos análisis realizados. En la primera columna se ordena los parámetros según importancia decreciente en cada componente. Mediante líneas horizontales se divide los conjuntos de parámetros que más influyen en las diferentes componentes.

Variable	Componente principal rotada			
	1	2	3	4
Ehl	0.959	-0.068	-0.016	-0.114
Efm	0.958	-0.176	0.064	-0.123
Efc	0.934	-0.240	0.130	-0.104
Ehl4-5	0.918	-0.232	-0.081	-0.176
Vrms4-5	-0.112	0.948	0.020	0.176
Vhl4-5	-0.328	0.704	-0.382	-0.275
Erms4-5	-0.370	0.667	-0.036	0.460
tins12-34	-0.112	-0.067	0.970	-0.015
tinsciclo2-3	0.141	-0.081	0.928	-0.199
Efm1-23	-0.152	0.025	-0.097	0.952
Efc1-23	0.240	0.526	-0.174	0.710

Tabla 7.22: *Contribución de cada variable a la componente principal del modelo para la extracción de cuatro componentes rotadas.*

Variable	C. P. rotada	
	1	2
Ehl4-5	0.956	-0.062
Efm	0.933	-0.134
Ehl	0.918	-0.014
Efc	0.914	-0.212
Erms4-5	-0.622	0.487
tinsciclo2-3	-0.002	-0.852
tins12-34	-0.283	-0.828
Efc1-23	-0.484	0.622
Vhl4-5	-0.372	0.534
Vrms4-5	-0.426	0.488
Efm1-23	-0.312	0.400

Tabla 7.23: *Contribución de cada variable a la componente principal del modelo para la extracción de dos componentes rotadas.*

Las variables que contribuyen a la 1ª componente al extraer dos y cuatro factores son las mismas a excepción de Erms4-5 que se encuentra casi repartido de igual forma en

ambas componentes. Estas variables corresponden a valores promedio de componentes frecuenciales, así como al desplazamiento de la DEP de la señal EMG al final de la prueba por el efecto de la fatiga muscular.

Las variables que se proyectan en las otras componentes en el análisis de cuatro factores aparecen contribuyendo en la segunda componente en el análisis de dos factores. Por esta razón, la reducción a sólo dos factores dificulta la interpretación de las componentes principales. En el primer análisis, las componentes resultantes corresponden a variables de procedencia similar: variaciones de parámetros frecuenciales y temporales en la 2ª componente, variaciones del ritmo inspiratorio al inicio del test en la 3ª y variaciones de parámetros frecuenciales al inicio del test en la 4ª componente principal.

En la representación en el plano factorial con dos componentes (ver figura 7.48) se observa claramente tres grupos de variables. El primero de ellos está claramente proyectado en la 1ª componente y está compuesto por las variables antes mencionadas. El segundo grupo orientado muy cerca de la segunda componente corresponde a los parámetros de la duración del período de inspiración que forman la 3ª componente en el análisis de cuatro factores. Por último quedan las otras variables que se encuentran en el segundo cuadrante correspondiente a variación de parámetros temporales y frecuenciales al inicio y final de la prueba asociados con efectos de adaptación al esfuerzo y fatiga muscular.

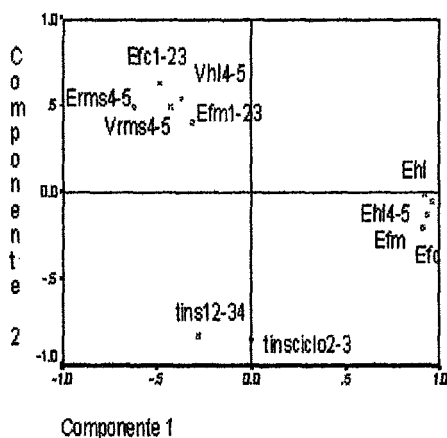


Figura 7.48: Representación de las variables en el plano factorial del análisis para la extracción de dos componentes principales.

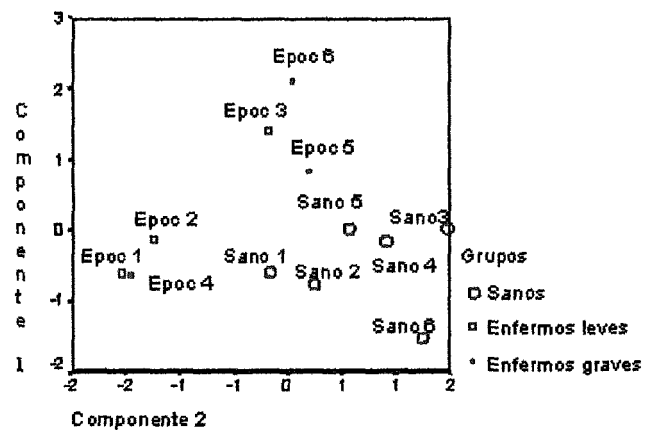


Figura 7.49: Representación de los doce individuos en las dos componentes principales.

#### Representación de los sujetos en las componentes principales

En la figura 7.49, se observa la distribución de los individuos en el nuevo subespacio formado por las componentes principales rotadas en el análisis con dos componentes. De esta forma se observa la agrupación de los individuos de ambas poblaciones en estudio. El grupo control se encuentra sensiblemente separado de la zona del subespacio donde se sitúan los pacientes. Sin embargo, los pacientes graves y leves comparten una misma

región, por lo que no se espera una buena discriminación entre los tres grupos aunque sí entre ambas poblaciones de sujetos sanos y pacientes en general, en el posterior análisis discriminante. En el caso de cuatro componentes, los gráficos bivariantes entre las componentes es de difícil interpretación.

#### ANÁLISIS DE SUBGRUPOS DE VARIABLES SEPARADAMENTE

##### Determinación de subgrupos de variables

Los subgrupos de variables formados a partir de las variables seleccionadas en el MANOVA de medidas repetidas en la prueba M1 (ver Apartado 7.3.4.2) son los siguientes:

- A. *Variaciones al inicio de la prueba:* Efc1-23, Efm1-23, Vfc1-234.
- B. *Variaciones al final de la prueba:* Erms4-5, Vrms4-5, Ehl4-5, Vhl4-5.
- C. *Variaciones en la fase media de la prueba:* Tins12-34, Tinsciclo2-3, lpi3-4, arplpi12-34.
- D. *Valores promedio a diferentes cargas:* Efc, Efm, Ehl, lpi234.

##### Inspección de las correlaciones entre variables y extracción de componentes principales

Los coeficientes de correlación entre las variables del subgrupo 'A' se sitúan en valores relativamente elevados: entre 0.65 y 0.74. Se extrae una única componente que explica el 81.2% de la varianza del subgrupo:

$$a1 = 0.923 Efm1-23 + 0.888 Vfc1-23 + 0.886 Efc1 - 23$$

En la tabla 7.24 se muestra la matriz de correlación de las variables del subgrupo 'B'. Se observa la mayor correlación entre las variaciones del RMS de las señales EMG y VMG entre los tramos 4 y 5: coeficiente de correlación de 0.74. En cambio, el correspondiente a los parámetros frecuenciales es sensiblemente menor, 0.575, y aún más pequeños las correlaciones entre estos últimos y los primeros. Mientras que la componente *a1* se asocia al efecto de adaptación al esfuerzo al inicio de la prueba, se decide extraer en el subgrupo 'B' una única componente principal que esté asociada a las variaciones globales de los parámetros al final de la prueba y, por tanto, a la fatiga del músculo. Dicha componente explica el 66.5% de la varianza del subgrupo y es la siguiente:

$$b1 = 0.873 Vrms4-5 + 0.864 Erms4-5 + 0.717 Vhl4-5 - 0.695 Ehl4-5$$

	Erms4-5	Vrms4-5	Ehl4-5	Vhl4-5
Erms4-5	1.000	0.860	-0.121	-0.449
Vrms4-5	0.860	1.000	0.171	-0.563
Ehl4-5	-0.541	-0.315	1.000	-0.422
Vhl4-5	0.328	0.575	-0.422	1.000

Tabla 7.24: Matriz de correlación de las variables del subgrupo 'B'.



En la matriz de correlaciones de las variables del subgrupo 'C' (ver tabla 7.25) se observa que la variable lpi3-4 es la menos correlada con las otras variables con coeficientes inferiores a 0.17. Las provenientes de la duración del período inspiratorio están altamente correladas entre ellas, coeficiente de 0.86, y el restante parámetro de correlación se encuentra ligeramente correlado con estos últimos. En este caso se extrae dos componentes principales que explican un total del 83.5% de la varianza del subgrupo y la contribución de cada variable a ambas componentes tras la rotación es la siguiente:

$$c1 = 0.945 \text{ tinsciclo2-3} + 0.915 \text{ tins12-34} - 0.734 \text{ arplpi12-34} + 0.044 \text{ lpi3-4}$$

$$c2 = 0.100 \text{ tinsciclo2-3} - 0.231 \text{ tins12-34} - 0.240 \text{ arplpi12-34} + 0.978 \text{ lpi3-4}$$

	tins12-34	tinsciclo2-3	arplpi12-34	lpi3-4
tins12-34	1.000	0.789	-0.541	0.328
tinsciclo2-3	0.789	1.000	-0.315	0.575
arplpi12-34	-0.121	0.171	1.000	-0.144
lpi3-4	-0.449	-0.563	-0.144	1.000

Tabla 7.25: Matriz de correlación de las variables del subgrupo 'C'.

Lógicamente, la variable lpi3-4 se proyecta casi totalmente en la segunda componente, mientras que la primera se compone, principalmente, de los parámetros de variación de la duración del periodo inspiratorio.

En el subgrupo 'D' existe correlaciones muy altas entre los parámetros frecuenciales promediados en todos los tramos de la prueba con coeficientes superiores a 0.90 (ver tabla 7.26). Por ello, al extraer dos componentes principales, estas variables se proyectan casi totalmente en la primera componente principal, mientras la variable del TRM para retardos positivos normalizado respecto a la duración de la inspiración se proyecta en gran medida en la segunda componente. Ambas componentes explican el 96.7% de la varianza total del subgrupo y son las siguientes:

$$d1 = 0.944 \text{ Efm} + 0.929 \text{ Efc} + 0.866 \text{ Ehl} + 0.336 \text{ lpi234}$$

$$d2 = 0.288 \text{ Efm} + 0.325 \text{ Efc} + 0.424 \text{ Ehl} + 0.941 \text{ lpi234}$$

	Efc	Efm	Ehl	lpi234
Efc	1.000	0.969	0.909	0.622
Efm	0.969	1.000	0.912	0.594
Ehl	0.909	0.912	1.000	0.678
lpi234	0.622	0.594	0.678	1.000

Tabla 7.26: Matriz de correlación de las variables del subgrupo 'D'.

En resumen, en la tabla 7.27 se observa el reparto de las variables en las componentes principales y el porcentaje de varianza de las variables de los diferentes subgrupos. Se observa que los parámetros relacionados con los periodos inspiratorios en el subgrupo 'C',

Grupo		Total	1ª componente	2ª componente
A Variaciones iniciales	% varianza explicada	81.2	81.2	
	Variables que más contribuyen		Efm1-23 Vfc1-234 Efc1-23	
B Variaciones finales	% varianza explicada	66.5	66.5	
	Variables que más contribuyen		Vrms4-5 Erms4-5 Vh14-5 Eh14-5	
C Variaciones tramos intermedios	% varianza explicada	83.5	56.6	26.9
	Variables que más contribuyen		Tins12-34 Tinsciclo2-3 arplpi12-34	lpi3-4
D Cargas promediadas	% varianza explicada	96.7	65.4	31.3
	Variables que más contribuyen		Efc Efm Eh1	lpi234

Tabla 7.27: Resultados del análisis de componentes principales para los diferentes subgrupos de variables.

y los parámetros frecuenciales en el caso 'D' se agrupan en las primeras componentes, que son las que explican más varianza, mientras que los parámetros de correlación están aportando información a la segunda componente.

#### Representación de los sujetos en las componentes principales

En las representaciones de los sujetos en dos factores comunes extraídos (ver figura 7.50 para el subgrupo 'C', y figura 7.51 para el caso 'D') se observa la necesidad de dos componentes para distribuir en dos grupos diferenciados las dos poblaciones de enfermos y personas sanas. Esto confirma la elección de dos componentes principales para los dos últimos análisis. Además, en el caso 'C' se observa una definida agrupación por zonas de los individuos sanos y pacientes con alto y bajo nivel de obstrucción.

#### 7.4.4.2 Resultados de la prueba M2

##### ANALISIS DEL GRUPO DE VARIABLES CONJUNTAMENTE

##### Selección de variables

La comunalidades de cada una de las variables seleccionadas en el MANOVA de medidas repetidas para la prueba M2 cuando se extrae cinco componentes principales se presenta en la la tabla 7.28 (ver Apartado 7.3.4.2). La varianza acumulada del modelo es el 95.3%.

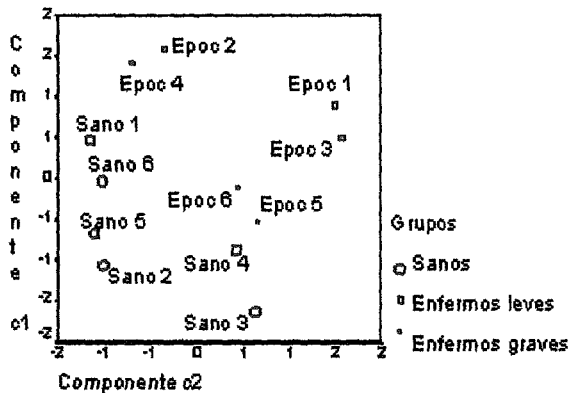


Figura 7.50: Representación de los doce individuos en las dos componentes extraídas en el análisis factorial 'C'.

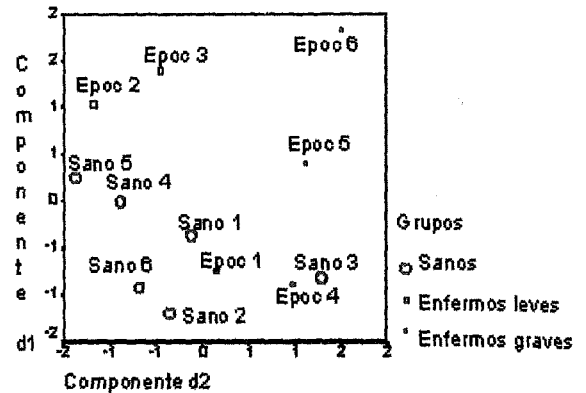


Figura 7.51: Representación de los individuos en las dos componentes extraídas en el análisis factorial 'D'.

Variable	Comunalidad	Variable	Comunalidad
Vhl1-2345	0.663	Vrme_lin	0.905
Erms_lin	0.723	Tins234	0.909
Ehl3-4	0.725	Vrme34-5	0.909
Erme_lin	0.792	lp_par	0.917
Efc3-45	0.828	lp1-2	0.921
VERmx3-45	0.853	Efm	0.947
tins1-234	0.859	Erme1-234	0.957
lppar34-5	0.884	lpi23	0.958
Ehl	0.896	Efc	0.963

Tabla 7.28: Comunalidades de las variables seleccionadas en el MANOVA de medidas repetidas en la prueba M1 mediante la extracción de cinco componentes principales.

Se decide descartar las siguientes variables para realizar el análisis con once variables:

- Vhl1-2345, EVrms\_lin, Ehl3-4: Poseen las comunalidades más bajas y, por tanto, las que menos contribuyen en el modelo obtenido. Por consiguiente, se eliminan inicialmente del análisis.
- lp\_par: el contraste parabólico tiene difícil interpretación de la función muscular en el test mantenido.
- VERmx34-5, Vrme34-5: De entre todos los parámetros procedentes del RMS seleccionados en el MANOVA de medidas repetidas, son los que presentan un p-valor más alto.
- Erme1-234: Es el único parámetro temporal procedente del RMS aparte de los dos contrastes lineales: Erme\_lin y Vrme\_lin. Debido a que de forma aislada ofrece información del inicio de la prueba se decide eliminarlo del análisis.

### Condiciones de aplicación

Respecto a la adecuación muestral y la matriz de correlación de las once variables seleccionadas se evalúa los siguientes apartados:

- El índice KMO de la prueba es 0.443, que aunque no es muy alto, si es más elevado que el correspondiente a la prueba M1.
- El p-valor que contrasta la hipótesis de esfericidad de Bartlett es 0.000. En consecuencia, el modelo es adecuado para los datos.

### Extracción de componentes principales

En la tabla 7.29 se observa las varianzas explicadas por las cuatro primeras componentes principales en el análisis factorial con las que se obtiene una varianza acumulada total superior al 85% (92.4%): 53.5%, 17.9%, 11.8% y 9.1% respectivamente. Cuando éstas son rotadas para interpretar mejor los factores la distribución de la varianza en las cuatro componentes cambia a: 40.2%, 20.1%, 16.9% y 15.0%.

Componente principal	Autovalor		
	Total	% de varianza	%acumulado
1	5.89	53.58	53.58
2	1.97	17.96	71.55
3	1.30	11.82	83.37
4	1.00	9.11	92.48

Tabla 7.29: *Varianzas explicadas por el modelo en la extracción de componentes principales sin rotación.*

En la esta tabla puede observarse que al escoger dos factores, la varianza total explicada es de 71.5%, y tras la rotación se reparte en 42.5% para la primera componente y 29.0% para la segunda.

En las tablas 7.30 y 7.31 se observa las contribuciones de las variables ordenadas en importancia decreciente para cada componente principal ( $\bar{b}_i$ ) en ambos análisis realizados. Mediante líneas horizontales se divide los conjuntos de parámetros que más influyen en las diferentes componentes.

Las variables que forman básicamente la primera componente en ambos análisis son las mismas. La tercera componente está formada por parámetros relacionados con la duración del período inspiratorio y la cuarta con contrastes lineales de los parámetros procedentes del RMS. En la representación en el plano factorial con dos componentes (ver figura 7.52) se observa que las variables de la 1ª componente se hallan claramente proyectadas en el primer factor y las variables Erme.lin y tins234 en el segundo, mientras que el resto se halla más disperso en el segundo y cuarto cuadrante.

Variable	Componente principal rotada			
	1	2	3	4
Ehl	0.958	0.131	0.064	-0.072
Efc	0.917	0.256	-0.242	-0.054
Efm	0.898	0.247	-0.275	0.015
lpi23	0.846	-0.116	-0.325	-0.331
lpi1-2	0.797	0.522	0.127	-0.110
Efc3-45	0.091	0.933	-0.098	-0.098
lppar34-5	-0.383	-0.790	0.342	0.051
tins234	-0.192	-0.152	0.939	0.016
tins1-234	0.403	0.458	-0.572	-0.406
Vrme_lin	-0.305	-0.031	-0.056	0.933
Erme_lin	0.249	-0.258	0.517	0.694

Tabla 7.30: Contribución de cada variable a la componente principal del modelo para la extracción de cuatro componentes rotadas.

Variable	C. P. rotada	
	1	2
Ehl	0.966	-0.022
Efc	0.934	-0.285
Efm	0.915	-0.267
lpi1-2	0.866	-0.226
lpi23	0.798	-0.274
Erme_lin	0.244	0.845
tins1-234	0.436	-0.817
tins234	-0.180	0.699
lppar34-5	-0.483	0.674
Efc3-45	0.223	-0.633
Vrme_lin	-0.283	0.437

Tabla 7.31: Contribución de cada variable a la componente principal del modelo para la extracción de dos componentes rotadas.

Representación de los sujetos en las componentes principales

En las figura 7.53, se observa cómo se agrupan los individuos en el nuevo subespacio formado por dos componentes principales rotadas. En el caso de cuatro componentes, esta representación se realiza con gráficos bivariantes entre las componentes de muy difícil interpretación gráfica.

Se observa que los sujetos sanos se encuentran separados de los pacientes con EPOC, aunque entre ellos se encuentran mezclados los de alto y bajo nivel de severidad obstructiva.

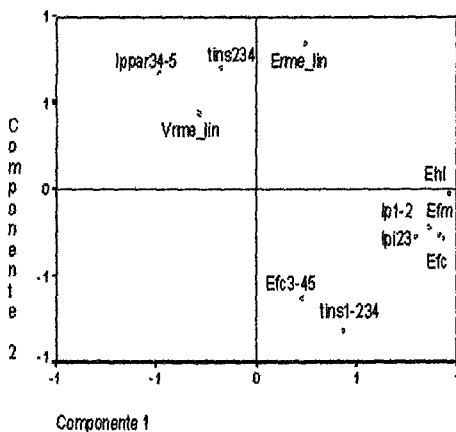


Figura 7.52: Representación de las variables en el plano factorial del análisis para la extracción de dos componentes principales.

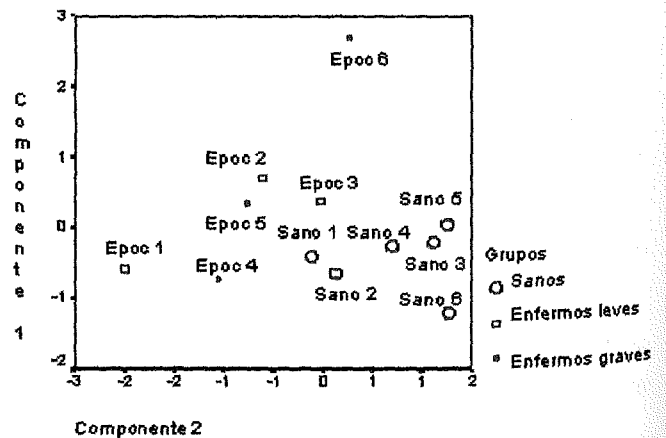


Figura 7.53: Representación de los doce individuos en las dos componentes principales.

## ANALISIS DE SUBGRUPOS DE VARIABLES SEPARADAMENTE

## Determinación de subgrupos de variables

Los subgrupos de variables formados a partir de las variables seleccionadas en el MANOVA de medidas repetidas en la prueba M2 (ver Apartado 7.3.4.2) son los siguientes:

- A. Variaciones al inicio de la prueba: Erme1-234, Tins1-234, Vhl1-23, lp1-2.
- B. Variaciones al final de la prueba: Efc3-45, lppar34-5, VErmx34-5, Vrme34-5.
- C. Evolución durante toda la prueba: EVrms\_lin, Vrme\_lin, Erme\_lin.
- D. Valores promedio a diferentes cargas: Efc, Efm, Ehl, lpi23, Tins234.

## Inspección de las correlaciones entre variables y extracción de componentes principales

En la tabla 7.32 se presenta la matriz de correlación de las variables correspondientes al subgrupo 'A'. Los parámetros temporales tienen el mayor coeficiente de correlación (0.65), mientras que el frecuencial es el que menos lo está con el resto de variables. El restante parámetro de correlación está ligeramente correlado con el correspondiente a la duración del período inspiratorio. Por tanto, se decide la extracción de dos componentes principales que reúne el 78.3% de la varianza total del subgrupo y son los siguientes:

$$a1 = 0.908 \text{ Erme1-234} + 0.872 \text{ tins1-234} + 0.053 \text{ Vhl1-23} + 0.320 \text{ lp1-2}$$

$$a2 = -0.058 \text{ Erme1-234} + 0.352 \text{ tins1-234} - 0.830 \text{ Vhl1-23} + 0.786 \text{ lp1-2}$$

	Erme1-234	tins1-234	lp1-2	Vhl1-23
Erme1-234	1.000	0.650	0.116	-0.088
tins1-234	0.650	1.000	0.571	-0.174
lp1-2	0.116	0.571	1.000	-0.356
Vhl1-23	-0.088	-0.174	-0.356	1.000

Tabla 7.32: Matriz de correlación de las variables del subgrupo 'A'.

Respecto al subgrupo 'B', en la tabla 7.33 se muestra la matriz de correlación de las variables. La mayor correlación se observa entre los parámetros temporales, 0.682, así como entre las dos restantes variables: -0.742. Sin embargo, entre estos últimos y los primeros no hay prácticamente correlación. De esta forma se extrae dos componentes principales que reúnen el 85.5% de la varianza total repartida casi equitativamente: 43.0% y 42.5%. Las componentes son las siguientes:

$$b1 = -0.929 \text{ Efc3-45} + 0.927 \text{ lppar3-45} + 0.061 \text{ VErmx34-5} - 0.060 \text{ VErme34-5}$$

$$b2 = -0.025 \text{ Efc3-45} - 0.024 \text{ lppar3-45} + 0.917 \text{ VErmx34-5} + 0.917 \text{ VErme34-5}$$

	Vrme34-5	VErmx34-5	lppar34-5	Efc3-45
Vrme34-5	1.000	0.682	-0.055	0.035
VErmx34-5	0.682	1.000	0.020	-0.076
lppar34-5	-0.055	0.020	1.000	-0.724
Efc3-45	0.035	-0.076	-0.724	1.000

Tabla 7.33: Matriz de correlación de las variables del subgrupo 'B'.

En el subgrupo 'C', la mayor correlación se encuentra entre la variable Erms\_lin y Vrme\_lin con un coeficiente de valor -0.654. Esta última variable también está correlada con un coeficiente similar, 0.504, con la variable Erme\_lin. Sin embargo, dicha variable con la primera mencionada se encuentran prácticamente incorreladas con un coeficiente de valor -0.059. Por tanto, se extrae dos componentes que explica el 93.3% de la varianza total y corresponde a las siguientes combinaciones lineales:

$$c1 = 0.966 \text{ Erms\_lin} - 0.788 \text{ Vrme\_lin} - 0.064 \text{ Erme\_lin}$$

$$c2 = 0.051 \text{ Erms\_lin} + 0.523 \text{ Vrme\_lin} + 0.980 \text{ Erme\_lin}$$

Se muestra en la tabla 7.34 los coeficientes de correlación entre las variables del subgrupo 'D'. Existe una alta correlación entre los parámetros frecuenciales y el parámetro de correlación, coeficientes de valor superior a 0.77, mientras que el parámetro temporal aparece ligeramente correlado con el resto de variables, coeficientes de valor -0.4, a excepción del parámetro Ehl. Se decide extraer una única componente que reúne el 75.6% de la varianza total del subgrupo. La extracción de una segunda componente principal para considerar parte de la varianza de la variable tins234 no justifica la complicación del modelo. La componente, que obtiene una gran proyección de los parámetros frecuenciales y de tiempo de retardo medio, y en menor medida, de la duración del período inspiratorio, es la siguiente:

$$d1 = 0.974 \text{ Efc} + 0.973 \text{ Efm} + 0.909 \text{ lpi23} + 0.901 \text{ Ehl} - 0.500 \text{ tins234}$$

	tins234	lpi23	Efc	Efm	Ehl
tins234	1.000	-0.447	-0.411	-0.437	-0.178
lpi23	-0.447	1.000	0.830	0.822	0.774
Efc	-0.411	0.830	1.000	0.977	0.887
Efm	-0.437	0.822	0.977	1.000	0.875
Ehl	-0.178	0.774	0.887	0.875	1.000

Tabla 7.34: Matriz de correlación de las variables del subgrupo 'D'.

En resumen, en la tabla 7.35 se observa cómo se reparten las diferentes variables temporales, frecuenciales y de correlación en su contribución a las componentes principales y el porcentaje de varianza de las variables de los diferentes subgrupos.

Grupo		Total	1ª componente	2ª componente
A Variaciones iniciales	% varianza explicada	78.3	42.2	25.1
	Variables que más contribuyen		Erme1-234 tins1-234	Vhl1-23 lp1-2
B Variaciones finales	% varianza explicada	85.5	43.0	42.5
	Variables que más contribuyen		Efc3-45 lppar34-5	VErmx34-5 Vrme34-5
C Evolución lineal	% varianza explicada	93.3	52.2	41.1
	Variables que más contribuyen		Erms_lin Vrme_lin	Erme_lin
D Cargas promediadas	% varianza explicada	75.6	75.6	
	Variables que más contribuyen		Efc Efm lpi23 Ehl tins234	

Tabla 7.35: Resultados del análisis de componentes principales para los diferentes subgrupos de variables.

#### Representación de los sujetos en las componentes principales

Las representaciones de los sujetos en dos factores comunes extraídos se muestran en las figuras 7.54 a 7.56 para los subgrupos 'A', 'B' y 'C', respectivamente. Sólo en los casos 'A' y 'B' se observa una total separación entre las zonas de agrupación de los individuos sanos y pacientes con EPOC, aunque sólo el primero separa pacientes con alto y bajo nivel de severidad obstructiva. Esto hace suponer que no se conseguirá de forma general una perfecta discriminación al intentar separar las tres poblaciones con estos factores.

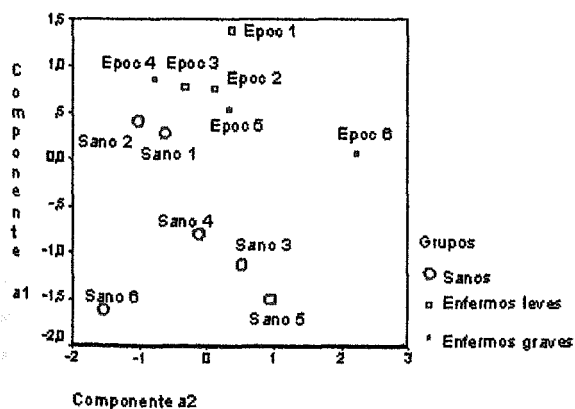


Figura 7.54: Representación de los doce individuos en las dos componentes extraídas en el análisis factorial 'A'.

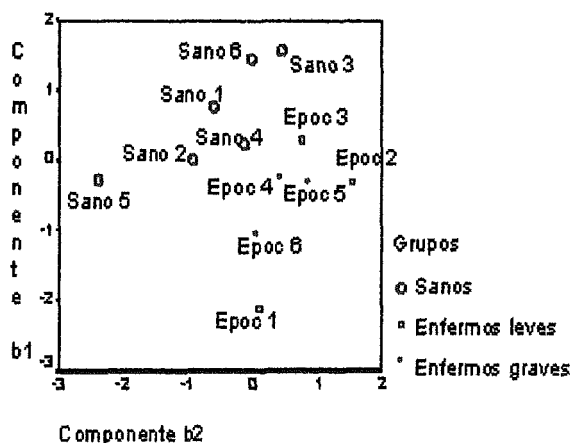


Figura 7.55: Representación de los individuos en las dos componentes extraídas en el análisis factorial 'B'.



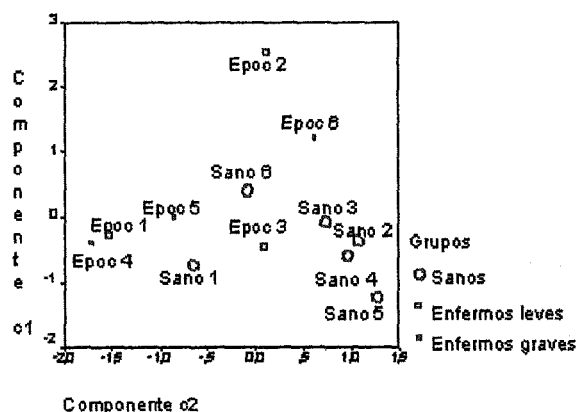


Figura 7.56: Representación de los individuos en las dos componentes extraídas en el análisis factorial 'C'.

#### 7.4.4.3 Conclusiones

Correspondiente al análisis factorial de todo el grupo de variables de forma conjunta, se compara las tablas de resultados 7.22 y 7.23 con las tablas 7.30 y 7.31 correspondientes a las pruebas M1 y M2, respectivamente y se obtiene las siguientes conclusiones:

- La proyección de las variables en las diferentes componentes principales para la extracción de cuatro y dos factores es muy similar en ambas pruebas a excepción de los siguientes parámetros:
  - variación frecuencial al inicio y contrastes lineales de parámetros temporales que no fueron seleccionados en el MANOVA de medida repetidas en las pruebas M1 y M2, respectivamente.
  - el tiempo de retardo medio para retardos positivos de la FCC no aparece en el modelo factorial de la prueba M1 ya que fue eliminado inicialmente del análisis.
- La primera componente en la extracción de cuatro y dos factores en ambas pruebas corresponde, básicamente, a valores promedio de las variables frecuenciales.
- En las pruebas M1 y M2, se forma una componente en la extracción de cuatro factores correspondiente a la variación del rimo inspiratorio al inicio de la prueba y otra formada a partir de variación de parámetros al final de la prueba por el efecto de la fatiga muscular.

Debido a todo ello, se comprueba una reproducibilidad en ambas pruebas mantenidas reflejada en sus modelos factoriales.

Referente al análisis de subgrupos de variables de forma separada se concluye los siguientes aspectos:

- Las variables seleccionadas en el MANOVA de medidas repetidas se han agrupado en cinco grupos de forma similar en ambas pruebas. El subgrupo 'C' es el único diferente: en la prueba M1 corresponde a variaciones en la fase intermedia y en la M2 se agrupa los contrastes lineales de los parámetros procedentes del RMS.
- Los modelos en ambos protocolos no coinciden en el número de componentes extraídas ni, por tanto, en los porcentajes de varianza que reúne cada componente.

Por consiguiente, con la metodología de análisis en subgrupos de variables no se obtiene modelos tan similares en ambas pruebas. En consecuencia el modelo factorial ya no es tan reproducible debido, seguramente, a un efecto de aprendizaje del sujeto en la realización del protocolo mantenido.

Mediante la representación de las puntuaciones factoriales de los individuos en los planos factoriales, se ha observado en todos los casos la posibilidad de diferenciar claramente entre el grupo control y el grupo de pacientes mediante un análisis discriminante posterior, al situarse en regiones separadas. También se detecta la dificultad de hacerlo de forma general entre enfermos graves y leves, aunque en la prueba M1 se diferencia claramente las tres poblaciones en componentes extraídas del análisis mediante subgrupos 'D' y en mayor medida 'C' donde la segunda componente está formada en ambos casos por parámetros de tiempo de retardo medio de la FCC para retardos positivos.

#### 7.4.5 Discusión y conclusiones

Al realizar el análisis factorial mediante la extracción de dos factores y de un número de componentes que reúna una varianza superior al 85% de la total de la muestra, cuatro en todos los protocolos, se obtiene las siguientes conclusiones:

- La proyección de las variables seleccionadas en el MANOVA de medidas repetidas en cada factor es, en general, muy similar en todas las pruebas:
  - La primera componente principal, correspondiente al factor que reúne mayor varianza de la muestra, en ambos análisis y para las cuatro pruebas está formada básicamente por las mismas variables: valores promedio durante la prueba de los parámetros frecuenciales y TRM de la FCC para retardos positivos normalizado respecto a la duración del período inspiratorio.
  - Las restantes componentes en el análisis de cuatro factores se proyectan de forma desigual en su contribución en la segunda componente cuando se extrae dos factores dificultando su interpretación de la función muscular en este caso.

- Se forma una componente en la extracción de cuatro factores correspondientes a parámetros relacionados con la variación del ritmo inspiratorio en tres de las cuatro pruebas: IC, M1 y M2. En la prueba IR fue eliminado inicialmente del análisis.
  - Variaciones de parámetros procedentes del dominio temporal, frecuencial y de la FCC entre tramos al final de la prueba forman en tres de las cuatro pruebas (IC, M1 y M2) una componente en el análisis con cuatro factores. Dicha componente está asociada a la fatiga muscular del sujeto alcanzada en los protocolos que, por tanto, queda reflejada mediante variables de diferente naturaleza.
- Sin embargo, en los modelos factoriales con cuatro componentes se observa ciertas diferencias entre los protocolos:
    - En ambas pruebas del protocolo incremental se forma una componente basada en variaciones de los parámetros de correlación entre apnea y volumen corriente. Dicha componente no aparece en el protocolo mantenido por no tenerse en cuenta el test de apnea.
    - Las variables correspondientes a contrastes lineales se distribuyen entre los factores en determinadas componentes de forma aislada en las segundas pruebas de ambos protocolos: IR y M2. No aparece en la prueba IC porque fueron eliminadas inicialmente del análisis por su baja comunalidad, ni en la prueba M1 por no seleccionarse en el MANOVA de medidas repetidas.

Mediante el análisis factorial de forma separada en subgrupos determinados según el instante de la prueba en que se ha calculado las variables se resalta los siguientes aspectos:

- Se forma los subgrupos dependiendo del instante de la prueba que se ha calculado las variables. De esta forma se obtiene componentes principales con una alta interpretación de la función muscular que permitirá mediante un análisis discriminante posterior analizar si las cargas ventilatorias bajas, medias o altas, los efectos de adaptación al esfuerzo y fatiga muscular, o la evolución durante la prueba son más o menos determinantes para separar el grupo de pacientes y el de control.
- Los protocolos incrementales constan de cinco subgrupos y los mantenidos de cuatro debido a que no aparece variación entre apnea y volumen corriente en estos últimos. Los restantes subgrupos son:
  - Variaciones al inicio de las pruebas IC, M1 y M2.
  - Variaciones en la fase intermedia de la pruebas IC y M1.
  - Variaciones al final de las cuatro pruebas.
  - Evolución durante toda la prueba en todas ellas excepto en el test M1.
  - Valores promedio a diferentes cargas en las cuatro pruebas.

- Aparece diferencias características en los siguientes protocolos:
  - Protocolo incremental: Los modelos coinciden en ambas pruebas en el número de componentes extraídas en cada subgrupo y en los porcentaje de varianza reunida por cada una de ellas demostrando su reproducibilidad.
  - Protocolo mantenido: Los modelos presentan un número de componentes por cada subgrupo y porcentaje de varianza explicada en las componente diferentes en las pruebas M1 y M2. En este caso el modelo no es reproducible en ambos tests debido a un posible efecto de aprendizaje del sujeto durante el protocolo.
- Cuando en un subgrupo compuesto por dos componentes existen parámetros de diferente procedencia, los parámetros frecuenciales se proyectan en la primera de ellas que explica mayor cantidad de varianza del modelo y los temporales y de correlación en la segunda. Ello, además de permitir que el modelo mantenga una alta interpretabilidad de la función del músculo ECM, demuestra que cada tipo de parámetro está asociado con información diferente del comportamiento muscular.

Se ha observado mediante la representación de los individuos en el plano factorial generado con las nuevas componentes principales que los individuos sanos y pacientes con EPOC se agrupan en diferentes zonas del espacio, comprobando que los datos se encuentran preparados para obtener buena separación de ambas poblaciones en el análisis discriminante posterior. No sucede lo mismo entre los enfermos con alto y bajo nivel de severidad obstructiva al encontrarse generalmente mezclados. En la prueba M1 se presenta una buena distribución entre las tres poblaciones (sujetos sanos y enfermos leves y graves) que favorece su separación en subespacios formados por componentes de subgrupos, donde la segunda de ellas corresponde al TRM para retardos positivos de la FCC normalizado respecto a la duración del período inspiratorio.

## 7.5 Análisis discriminante

### 7.5.1 Introducción

El análisis discriminante es útil para las situaciones en que se desea construir un modelo de pronóstico de pertenencia al grupo de enfermos o sujetos sanos basándose en las características observadas para cada caso o individuo analizado. El análisis genera una función discriminante, en el caso de más de dos grupos un conjunto de funciones discriminantes, basándose en la combinación lineal de las variables predictoras o independientes que proporcionan la mejor discriminación entre los grupos. Las funciones se generan a partir de una muestra de casos para los que se conoce la pertenencia al grupo. Posteriormente, se aplica las funciones a nuevos casos con medidas de las variables independientes para predecir la pertenencia al grupo no conocida a priori.

En el presente estudio se busca y compara funciones discriminantes calculadas a partir de diferentes componentes extraídas del análisis factorial del Apartado 7.4. En cada protocolo ventilatorio, se compara los resultados cuando se utiliza puntuaciones factoriales, provenientes de componentes principales extraídas globalmente de todos los parámetros de la prueba, o calculadas a partir de diferentes subgrupos de variables según su interpretación.

Se han utilizado diferentes métodos para validar la función discriminante calculada [Cuadras, 1991]. Finalmente, se compara las funciones discriminantes correspondientes a cada protocolo ventilatorio, para analizar la importancia de los diferentes tests en la separación de poblaciones.

## 7.5.2 Modelo de análisis de la actividad muscular respiratoria

### 7.5.2.1 Descripción del método

Dada una variable dependiente cualitativa,  $n$  individuos o casos y  $p$  variables independientes cuantitativas se pretende obtener funciones lineales que sean combinación de las variables independientes que permita clasificar a los  $n$  individuos en uno de los  $k$  grupos establecidos por los valores de la variable dependiente cualitativa.

A partir de los  $n \times p$  coeficientes, se trata de extraer un nuevo espacio de pequeña dimensión tal que, al proyectar la nube de puntos, los correspondientes a individuos del mismo grupo estén próximos y los de distintos grupos estén alejados. El sistema de ejes de este nuevo espacio es la función discriminante, y cada uno de los ejes corresponde a las funciones canónicas discriminantes.

El número de funciones discriminantes a calcular,  $s$ , es el número de grupos menos uno ( $s=k-1$ ), excepto cuando algunos grupos se encuentren sobre el mismo eje del nuevo espacio, alineados, o cuando  $p < k-1$ , caso que no se produce en el presente estudio.

Las puntuaciones discriminantes para el  $i$ -ésimo individuo, calculadas a partir de la función discriminante son las siguientes:

$$d_{is} = c_{s1}x_{i1} + c_{s2}x_{i2} + \dots + c_{sp}x_{ip} + p_{s0} \quad \text{donde:}$$

- $x_{i1}, \dots, x_{ip}$ : Son las variables independientes a utilizar en el análisis.
- $c_{s1}, \dots, c_{sp}$ : Son los coeficientes de la función discriminante 's'.
- $p_{s0}$ : Es el término independiente del valor de la función en el origen de la abscisa.

Los coeficientes  $c_{si}$  se calculan para que cumplan las siguientes especificaciones:

- los valores de  $d_{is}$  tengan la mayor diferencia posible entre grupos, es decir, para que el ratio  $SC_{entre\ grupos} / SC_{intra\ grupos}$  en las puntuaciones discriminantes sea máximo.

- la condición de estandarización para todos los individuos de la muestra, es decir, el valor medio de las puntuaciones discriminantes ha de ser cero, y su desviación estándar uno.

Para evaluar la información de cada función discriminante en particular, se calcula el denominado coeficiente de correlación canónica ( $C_c$ ). Este mide las desviaciones de las puntuaciones discriminantes entre grupos respecto a las desviaciones totales sin distinguir grupos. Representa la proporción de la varianza total atribuible a las diferencias entre grupos. Es, por tanto, una medida del grado de asociación entre las funciones discriminantes y los grupos. De esta forma, si es próxima a la unidad la dispersión será debida a dichas diferencias y, en tal caso, dicha función discriminará eficientemente.

Además, es interesante, a partir de la puntuación de cada uno de los individuos, realizar un análisis de la varianza para calcular el p-valor. Cuánto menor sea éste, el análisis discrimina con una significación estadística mayor.

Este método es utilizado para escoger ordenadamente las variables independientes que forman la función discriminante de forma que correspondan con las que más expliquen las diferencias entre los grupos. Existen diferentes criterios para la inclusión de las diferentes variables en la función discriminante [Norusis, 1994]. En el presente estudio se utiliza la Lambda de Wilks ( $\lambda_W$ ) utilizando el método paso a paso para mantener la coherencia de todos los cálculos realizados hasta el momento. Este se basa en el cálculo inicial de la  $\lambda_W$  para las variables independientes. La  $\lambda_W$  mide las desviaciones dentro de cada grupo respecto a las desviaciones totales sin distinguir grupos en el espacio p-dimensional generado por los valores de las  $p$  variables independientes:

$$\lambda_W = \frac{SC_{intra\ grupos}}{SC_{intra\ grupos} + SC_{entre\ grupos}}$$

Si su valor es:

- pequeño: la variabilidad total es debida a las diferencias entre grupos, y las variables discriminan los grupos.
- próximo a la unidad: los grupos están mezclados.

El criterio que se escoge para evaluar la entrada de una variable en el modelo es la disminución de la  $\lambda_W$ . Si esta disminución es significativa, es decir, la variable aporta información significativa, el estadístico F de cambio es grande. Este valor se calcula mediante la siguiente expresión donde  $m$  es el número del paso en el método:

$$F_{cambio} = \left( \frac{n - k - m}{k - 1} \right) \frac{1 - \frac{\lambda_{wm} + 1}{\lambda_{wm}}}{\lambda_{wm} + \frac{1}{\lambda_{wm}}}$$

De esta forma se fija un estadístico F de entrada ( $F_E$ ), cuyo valor ha de ser superado por la variable candidata para entrar en el modelo. Por otra parte, que la  $\lambda_W$  tome

un valor pequeño para una variable candidata a entrar en el modelo no quiere decir que sea al final significativa en el análisis, ya que puede existir correlación entre las variables y al entrar otra, en un paso posterior, la significación de lo explicado por las otras variables introducidas previamente se modifica. Además, el modelo puede ajustarse menos a la población real que se da en la sociedad debido al bajo número de individuos. Por tanto, antes de seleccionar una variable se trata de eliminar alguna de las previamente seleccionadas mediante la evaluación del incremento de la  $\lambda_W$  si la variable correspondiente es eliminada. Si el valor del estadístico F calculado es menor que el de la F de salida ( $F_S$ ) el incremento no es significativo y la información que se perdería es poca, con la consiguiente salida de la variable del modelo.

Por último, para evitar un proceso cíclico se establece un límite de número de pasos, que se escoge igual al doble del número de variables independientes.

### 7.5.2.2 Condiciones de aplicación

Para que la función discriminante sea óptima, es decir, proporcione unas reglas de clasificación que minimice la probabilidad de errar en dicha clasificación, se evalúa las siguientes hipótesis:

- Matrices de covarianza intra-grupos iguales en todos los grupos. Si no son demasiado diferentes, la función lineal discriminante se genera correctamente, especialmente en muestras pequeñas [Wahl y Kronmal, 1997]. Se evalúa mediante el test M de Box ya comentado en el Apartado 7.3.
- La pertenencia al grupo es exclusiva, es decir, ningún caso pertenece a más de un grupo y exhaustiva de modo colectivo, es decir, todos los casos son miembros de un grupo.
- La pertenencia al grupo es una variable verdaderamente categórica como en el presente estudio: sujeto sano y paciente con EPOC o enfermo grave y leve. En este caso, el modelo es más efectivo. Si la pertenencia al grupo se basa en los valores de una variable continua se tendría que utilizar la regresión lineal múltiple para aprovechar la información más completa ofrecida por la variable continua.

### 7.5.2.3 Validación del método

A partir de las puntuaciones discriminantes se estima la probabilidad de pertenecer a cada grupo mediante el teorema de Bayes. Esta regla probabilística calcula la probabilidad que un sujeto con una puntuación discriminante  $d$  pertenezca a un grupo  $i$ , y se expresa de la siguiente forma:

$$P(G_i/d) = \frac{P(d/G_i) \cdot P(G_i)}{\sum_{i=1}^k P(d/G_i) \cdot P(G_i)} \quad \text{donde:}$$

- $P(G_i)$  se denomina probabilidad previa. Es la probabilidad que un sujeto pertenezca a un grupo particular cuando no se dispone de información sobre él. Habitualmente, cuando el número de individuos en ambos grupos es el mismo, como sucede en el presente estudio, se escoge equiprobable para todos los grupos. De esta manera, se considera que una persona que se dirige al servicio neumológico tiene la misma probabilidad, a priori, de estar sana o enferma. En el caso de estudios epidemiológicos, donde la muestra está escogida según proporciones reales de las poblaciones en la sociedad, la probabilidad depende del tamaño del grupo.
- $P(d/G_i)$  se denomina probabilidad condicional. Si las puntuaciones están normalmente distribuidas es posible calcular la probabilidad de obtener una determinada puntuación si el sujeto es miembro del grupo  $i$ , y corresponde a la siguiente expresión:  $P(d/G_i) = \exp\left(-\frac{x_i^2}{2}\right)$ , siendo  $x_i$  la distancia entre el punto del individuo correspondiente localizado a partir de sus puntuaciones discriminantes, y el centroide del grupo  $i$ .
- $P(G_i/d)$  se denomina probabilidad posterior. Es un estimador de la calidad como miembro de cada uno de los grupos, ya que los sujetos se clasifican basándose en su puntuación discriminante en el grupo cuya probabilidad posterior es mayor.

El análisis discriminante es un método de diagnóstico. Por ello, es importante evaluar su eficiencia mediante los siguientes parámetros:

- *Sensibilidad ( $S$ )*.

Muestra la bondad de la técnica en la detección de pacientes con EPOC en la población.

$$S = \frac{\text{Numero de enfermos correctamente detectados}}{\text{Numero total de enfermos}}$$

- *Especificidad ( $E_p$ )*.

Muestra si la técnica ha catalogado a los sujetos sanos correctamente.

$$E_p = \frac{\text{Numero de controles correctamente detectados}}{\text{Numero total de controles}}$$

- *Exactitud ( $E_x$ )*.

La sensibilidad y especificidad son medidas sobre los grupos específicos y no dan una medida global del rendimiento de la técnica. Por ejemplo, con niveles conservadores, se puede llegar a no obtener ningún sujeto sano clasificado como enfermo obteniendo una  $E_p=100\%$ . Sin embargo, esto se consigue a costa de una menor sensibilidad. La exactitud da una medida global del rendimiento del análisis:

$$E_x = \frac{\text{Numero de diagnosticos correctos}}{\text{Numero total de individuos}}$$



Los tres parámetros se relacionan mediante la siguiente expresión:

$$E_x = \frac{N^{\circ} \text{ de enfermos}}{N^{\circ} \text{ total de individuos}} \cdot S + \frac{N^{\circ} \text{ de controles}}{N^{\circ} \text{ total de individuos}} \cdot E_p$$

En el presente estudio, tienen especial interés los porcentajes de casos bien clasificados correspondiente a la exactitud y la sensibilidad. Se considera más importante detectar a todos los enfermos, aunque algún sujeto sano pueda estar mal clasificado, que conseguir un porcentaje muy elevado de individuos bien clasificados.

La efectividad de la función discriminante se evalúa calculando el ratio de clasificaciones bien realizadas en otra muestra distinta a la utilizada para el cálculo de la función. Una posible solución es dividir la muestra en dos: una para construir la función discriminante y otra para validarla. Cuando el número de casos o sujetos disponibles es reducido, como así ocurre en el presente estudio, este método no es práctico pues al estimarse la función de discriminación con la mitad de los individuos, ésta resultaría poco fiable.

Mediante el método de la validación cruzada ('Jack-knife') se utiliza la clasificación dejando un caso o sujeto fuera del análisis para construir la función discriminante ('leaving-one-out'). A continuación, mediante esta función se clasifica el individuo que queda fuera. Este proceso se realiza con todos y cada uno de los sujetos. Es decir, se clasifica cada caso del análisis mediante la función derivada a partir del resto de casos. De esta forma, se obtiene un estimador menos sesgado que el calculado a partir del modelo construido incluyendo dicho sujeto y los porcentajes calculados se acercan más a los obtenidos si la función discriminante se aplicara sobre los individuos de toda la sociedad.

#### 7.5.2.4 Cálculos realizados

En el análisis discriminante se realiza los siguientes cálculos:

1. *Evaluación de la condición de aplicación.*

Análisis de la condición de homogeneidad de la varianza entre los grupos mediante el test M de Box.

2. *ANOVA de cada variable independiente separadamente.*

Se realiza un ANOVA de cada una de las variables predictoras del análisis de forma separada y se obtiene el nivel de significación de la discriminación entre poblaciones. La variable que obtenga un p-valor menor es la que entra en primer lugar en el método paso a paso.

3. *Cálculo de funciones discriminantes a partir de diferentes variables independientes.*

Para cada prueba respiratoria, se halla las funciones discriminantes en los siguientes casos en que las variables independientes o predictoras son las siguientes:

- puntuaciones que representan a los individuos en las componentes principales extraídas de ambos análisis factoriales realizados previamente (ver Apartado 7.4):
  - Cuatro componentes principales de las variables seleccionadas del MANOVA de medidas repetidas.
  - Componentes principales extraídas a partir de subgrupos de variables seleccionadas en el MANOVA de medidas repetidas.
- variables originales seleccionadas en el MANOVA de medidas repetidas.

#### 4. Cálculo de funciones discriminantes para discriminar dos y tres poblaciones.

Cada estudio o cálculo de las funciones discriminantes se realiza para dos casos de variable dependiente:

- Sano - epoc ( $k=2$ ).  
Tenemos un grupo control de 6 individuos sanos y otro de 6 pacientes con EPOC (ver Apartado 2.4.3).
- Sano - epoc leve - epoc grave ( $k=3$ ).  
En el que el grupo de pacientes se ha dividido en dos subgrupos, uno de enfermos leves (Epoc: 1, 2 y 3) y otro de graves (Epoc: 4, 5 y 6) en base a los historiales clínicos que presentan los sujetos referente a su nivel de severidad en la obstrucción respiratoria: enfermos leves con  $VEF_1 = 34.33 \pm 9.4$  y enfermos graves con  $VEF_1 = 68.33 \pm 15.1$  en valor medio y desviación estándar.

El primer caso resulta de mayor interés, ya que el segundo es tan sólo orientativo debido al escaso número de individuos existentes en cada grupo de enfermos. Los resultados provenientes de una muestra tan reducida difícilmente se pueden extrapolar a la población real que se da en la sociedad. Sin embargo, tienen interés ya que a partir de los datos que se dispone indican la tendencia en la separación de enfermos dependiendo de su gravedad en la patología.

#### 5. Cálculo de funciones discriminantes en el protocolo completo.

Mediante las ocho componentes principales extraídas de las cuatro pruebas respiratorias mediante un análisis factorial previo de dos componentes (ver Apartado 7.4) para separar dos y tres poblaciones de individuos: sujetos sanos, enfermos graves y leves.

#### 6. Evaluación de la clasificación de los individuos y eficiencia en la discriminación.

Se calcula los parámetros de exactitud ( $E_x$ ) y sensibilidad ( $S$ ) en la clasificación de los individuos mediante el método 'leaving-one-out'. Se halla el coeficiente de correlación canónica ( $C_c$ ) y el p-valor obtenido al utilizar las puntuaciones discriminantes de cada individuo para evaluar la separación de las poblaciones mediante la función discriminante.

Los resultados han sido extraídos de los listados de resultados que aparecen en [Serena y Mañanas, 1999].

### 7.5.3 Protocolo incremental

#### 7.5.3.1 Resultados de la prueba IC

#### MEDIANTE CUATRO COMPONENTES PRINCIPALES

##### Condición de aplicación

Las variables independientes corresponden a las cuatro componentes principales (comp1, comp2, comp3 y comp4) extraídas en el análisis factorial realizado a las variables seleccionadas en el MANOVA de medidas repetidas de forma conjunta (ver Apartado 7.4.3.1). Dichas componentes cumplen el test M de Box para la hipótesis de homogeneidad multivariable de la varianza con un p-valor de 0.97.

##### Determinación de las funciones discriminantes

Las funciones discriminantes canónicas obtenidas mediante coeficientes estandarizados<sup>3</sup> para separar dos y tres poblaciones de sujetos son las siguientes:

$$f_{2\_grupos} = 0.899 \cdot comp2 - 0.845 \cdot comp1$$

$$f_{3\_grupos} = 1.000 \cdot comp3$$

Se observa que mientras las dos primeras componentes forman la función que discrimina dos grupos de individuos, para separar tres poblaciones (sujetos sanos y enfermos graves y leves) la tercera componente es la única de interés para la función discriminante. Además, al ser necesaria una única función para la discriminación, los tres grupos se encuentran alineados. Los tres primeros factores están formados por variables correspondientes a los siguientes efectos respectivamente:

- Valores promedio de parámetros frecuenciales y TRM para retardos positivos de la FCC.
- Variación de parámetros de correlación entre apnea y volumen corriente.
- Variaciones en el dominio temporal y de la FCC al final de la prueba.

Se obtiene p-valores al realizar un ANOVA de variables separadamente de 0.07 y 0.11 en los dos primeros factores, respectivamente, para dos poblaciones y 0.02 en la tercera componente con tres grupos.

##### Clasificación de los individuos y discriminación de la función

En la tabla 7.36 se muestra los resultados relacionados con la clasificación de individuos en dos y tres poblaciones.

<sup>3</sup>Se denomina estandarizado pues se ha de multiplicar por las variables independientes reducidas o tipificadas aunque en este caso ya lo son al tratarse de componentes principales.

	Dos grupos	Tres grupos
$E_x$ (%)	75.0	58.3
$S$ (%)	66.6	66.6
$C_c$	0.71	0.74
$p <$	0.03	0.02

Tabla 7.36: Resultados de la clasificación utilizando las cuatro componentes principales en la prueba IC.

## MEDIANTE COMPONENTES PRINCIPALES A PARTIR DE SUBGRUPOS

### Condición de aplicación

Las variables independientes corresponden a las seis componentes principales (a1, b1, c1, c2, d1 y d2) extraídas en el análisis factorial realizado a cuatro subgrupos de variables seleccionadas en el MANOVA de medidas repetidas de forma separada (ver Apartado 7.4.3.1). Dichas componentes cumplen el test M de Box para la hipótesis de homogeneidad multivariable de la varianza con un p-valor de 0.13.

### Determinación de las funciones discriminantes

La función discriminante canónica obtenida mediante coeficientes estandarizados para separar dos poblaciones de sujetos es la siguiente:

$$f_{2\_grupos} = 1.000 \cdot b1$$

Se observa la introducción de una única variable en el método paso a paso. Por tanto, la variable que más discrimina es el factor relacionado con el acoplamiento entre la actividad eléctrica y mecánica del músculo al pasar de cargas media a altas. Al realizar un ANOVA de las variables de forma separada se obtiene p-valores significativos en la discriminación de dos grupos en los factores a1 ( $p < 0.03$ ) y c1 ( $p < 0.03$ ), correspondientes a variaciones de parámetros de correlación entre apnea y volumen corriente y variaciones de parámetros frecuenciales a elevadas cargas, respectivamente. Estos parámetros no se introducen en la función discriminante al encontrarse muy correlacionados con la componente b1 (coeficientes de correlación de Pearson de valores -0.67 y -0.58, respectivamente, ambos significativos al nivel 0.05).

Las funciones discriminantes obtenidas para separar tres poblaciones son las siguientes:

$$f_{3\_grupos}^1 = 0.993 \cdot b1 + 0.069 \cdot d1$$

$$f_{3\_grupos}^2 = -0.137 \cdot b1 + 1.000 \cdot d1$$

Para tres grupos, la componente b1 continúa siendo la primera en introducirse en el modelo mediante el método paso a paso. El factor d1 es directamente la variable  $V_{fm\_lin}$ . Los p-valores al realizar un ANOVA de variables separadamente en los factores b1 y d1 son: 0.003 y 0.09, respectivamente. Las componentes a1 y c1, altamente correladas con b1, ofrecen p-valores menos significativos al separar las tres poblaciones:  $p < 0.09$  y  $p < 0.16$ , respectivamente.

### Clasificación de los individuos y discriminación de la función

En la tabla 7.37 se muestra los resultados relacionados con la clasificación de individuos en dos y tres poblaciones.

	Dos grupos	Tres grupos
$E_x$ (%)	91.6	75.0
$S$ (%)	83.3	66.6
$C_c$	0.67	0.85 , 0.63
$p <$	0.01	0.004 , 0.03

Tabla 7.37: Resultados de la clasificación utilizando las componentes principales de cuatro subgrupos de variables en la prueba IC.

La primera función canónica es la que discrimina más los grupos ( $p < 0.004$ ). Sin embargo, en este caso los grupos no están alineados pues la segunda función canónica también discrimina sensiblemente las poblaciones ( $p < 0.03$ ).

### MEDIANTE VARIABLES SELECCIONADAS ORIGINALES

#### Condición de aplicación

El análisis discriminante se realiza directamente a partir de las catorce variables seleccionadas mediante el MANOVA de medidas repetidas (ver Apartado 7.3). De esta forma se compara los resultados con los análisis discriminantes anteriores en que se utiliza puntuaciones factoriales. Las variables cumplen el test M de Box para la hipótesis de homogeneidad multivariable de la varianza con un p-valor de 0.11.

#### Determinación de las funciones discriminantes

La función discriminante canónica obtenida mediante coeficientes estandarizados para separar dos poblaciones de sujetos es la siguiente:

$$f_{2\_grupos} = 1.296 \cdot Efm5-6 + 1.504 \cdot Vfc.lin + 0.870 \cdot arlpi4-5 + 0.908 \cdot Erme2345-6$$

Las variables han sido introducidas en el método paso a paso en el mismo orden que aparecen en la función, por orden decreciente de discriminación. No aparece las variables arlpi0-1, ni ltot0-1, que presentan p-valor de discriminación de grupos mediante un ANOVA de las variables separadamente de  $p < 0.02$  y  $0.04$ , respectivamente. Su no inclusión es debido a la correlación existente con los siguientes parámetros:

- Efm5-6: coeficientes de Pearson de 0.494 y 0.325, respectivamente, y
- arlpi4-5: coeficientes de correlación de valor 0.508 y 0.606, respectivamente.

Además, el parámetro lpi obtiene un p-valor de 0.04 pero también está correlado con la variable Efm5-6: coeficiente de valor -0.853.

Respecto a la discriminación en tres grupos, las funciones discriminantes obtenidas según el orden de introducción de las variables son las siguientes:

$$f_{3\_grupos}^1 = 1.028 \cdot arcp3-4 + 0.872 \cdot arplpi4-5 - 0.743 \cdot Efc\_lin$$

$$f_{3\_grupos}^2 = 0.490 \cdot arcp3-4 - 0.603 \cdot arplpi4-5 + 0.592 \cdot Efc\_lin$$

Al comparar la separación de dos y tres grupos destaca el cambio de variables en la constitución de la función discriminante. En el caso de dos grupos, enfermos y grupo control, los parámetros frecuenciales tienen más importancia que los de correlación, siendo al revés en la separación de tres grupos. Las variables temporales aparecen en último lugar en la función para discriminar entre dos grupos.

En la separación de tres poblaciones, pacientes con diferente grado de obstrucción y sujetos sanos, el parámetro Efm5-6 no aparece en la función discriminante a pesar de presentar un p-valor de 0.04 por estar altamente correlado con arpcp3-4: coeficiente de Pearson de -0.580 siendo significativo al nivel 0.01. El parámetro ltot3-4 presenta la misma situación con un p-valor < 0.02 y un coeficiente de correlación de -0.726.

#### Clasificación de los individuos y discriminación de la función

En la tabla 7.38 se muestra los resultados relacionados con la clasificación de individuos en dos y tres poblaciones.

	Dos grupos	Tres grupos
$E_x$ (%)	91.6	83.3
$S$ (%)	100	83.3
$C_c$	0.95	0.91 , 0.80
p<	0.001	0.001 , 0.01

Tabla 7.38: Resultados de la clasificación utilizando las variables originales seleccionadas en la prueba IC.

#### 7.5.3.2 Resultados de la prueba IR

#### MEDIANTE CUATRO COMPONENTES PRINCIPALES

##### Condición de aplicación

Las cuatro componentes principales (comp1, comp2, comp3 y comp4) extraídas en el análisis factorial realizado con las variables seleccionadas en el MANOVA de medidas repetidas de forma conjunta (ver Apartado 7.4.3.2) son las variables independientes del análisis. Se cumple la hipótesis de homogeneidad multivariable de la varianza evaluada mediante el test M de Box con un p-valor de 0.43.

### Determinación de las funciones discriminantes

Las funciones discriminantes canónicas obtenidas mediante coeficientes estandarizados para separar dos y tres poblaciones de sujetos son las siguientes:

$$f_{2\_grupos} = -0.978 \cdot comp3 + 0.960 \cdot comp1 - 0.822 \cdot comp2$$

$$f_{3\_grupos} = 1.000 \cdot comp3$$

En la primera función discriminante se muestra las componentes que contribuyen en orden decreciente en la discriminación. La tercera componente, formada por variación de parámetros de correlación entre apnea y volumen corriente, es la que mejor separa las poblaciones en dos y, especialmente, tres grupos de individuos. Después, el primer factor corresponde a valores promedio de parámetros frecuenciales y TRM, mientras que el segundo está formado básicamente, por contrastes lineales de variables frecuenciales. Se obtiene unos p-valores al realizar un ANOVA de las variables separadamente de 0.08, 0.10 y 0.14 respectivamente. Por tanto, la combinación lineal de la función mejora la separación al obtener  $p < 0.01$  mediante las puntuaciones discriminantes. Los tres grupos se encuentran alineados al aparecer una única función discriminante.

### Clasificación de los individuos y discriminación de la función

Los resultados relacionados con la clasificación de individuos en dos y tres poblaciones se presenta en la tabla 7.39.

	Dos grupos	Tres grupos
$E_x$ (%)	91.6	50.0
$S$ (%)	100	33.3
$C_c$	0.84	0.58
$p <$	0.01	0.1

Tabla 7.39: Resultados de la clasificación utilizando las cuatro componentes principales en la prueba IR.

## MEDIANTE COMPONENTES PRINCIPALES A PARTIR DE SUBGRUPOS

### Condición de aplicación

Las variables predictoras del análisis son las ocho componentes principales ( $a_1$ ,  $b_1$ ,  $c_1$ ,  $c_2$ ,  $d_1$ ,  $d_2$ ,  $e_1$  y  $e_2$ ) extraídas en el análisis factorial realizado a cinco subgrupos de variables seleccionadas en el MANOVA de medidas repetidas (ver Apartado 7.4.3.2). Estas componentes cumplen la hipótesis de homogeneidad multivariable de la varianza con un p-valor de 0.65 en el test M de Box.

### Determinación de las funciones discriminantes

La función discriminante para separar dos poblaciones de sujetos es una combinación lineal de las siguientes componentes presentadas en el orden de introducción en el método paso a paso:

$$f_{2\_grupos} = 3.825 \cdot b_1 - 2.163 \cdot e_2 + 2.397 \cdot c_2 + 3.041 \cdot d_1 - 0.904 \cdot e_1$$

La primera variable que entra y, por tanto, que más discrimina es la componente b1 correspondiente a desplazamientos de la DEP de la señal EMG al inicio de la prueba. Las siguientes componentes en orden decreciente de importancia son:

- e2: compuesta principalmente por el promedio de parámetros de correlación en toda la prueba.
- c2: relacionada con cambios en el ritmo inspiratorio al final de la prueba.
- d1: compuesta por contrastes lineales de parámetros frecuenciales en la señal EMG.
- e1: corresponde a valores promedio de parámetros frecuenciales de la señal EMG.

Todas las componentes, excepto e1, presentan p-valores en ANOVA de las variables predictoras separadamente, inferiores o igual a 0.02. La componente a1 y c1 también obtienen una significación alta ( $p < 0.02$  y  $0.03$ , respectivamente), pero no aparecen en la función discriminante ya que están muy correladas con la componente b1, que entra en primer lugar en el análisis: coeficientes de correlación de Pearson de  $-0.670$  y  $-0.587$ , respectivamente, siendo significativo al nivel de  $0.05$ . La componente a1 está relacionada con variaciones de parámetros de la FCC entre apnea y volumen corriente, y la b1 con cambios de parámetros de la FCC y frecuenciales al final de las pruebas.

Las funciones discriminantes obtenidas para separar tres poblaciones son las siguientes:

$$f_{3\_grupos}^1 = 1.025 \cdot e2 + 0.982 \cdot d1$$

$$f_{3\_grupos}^2 = -0.539 \cdot e2 + 0.613 \cdot d1$$

En la separación de tres grupos la componente e2 es la primera en introducirse en el modelo y en segundo lugar la d1 obteniendo p-valores obtenidos con el ANOVA de variables separadamente de  $0.03$  y  $0.06$ , respectivamente. Para separar dos poblaciones dichos factores alcanzan un p-valor también significativo de  $0.02$ .

#### Clasificación de los individuos y discriminación de la función

En la tabla 7.40 se muestra los resultados relacionados con la clasificación de individuos en dos y tres poblaciones.

	Dos grupos	Tres grupos
$E_x$ (%)	100	75.0
$S$ (%)	100	100
$C_c$	0.98	0.89 , 0.06
$p <$	0.0003	0.008 , 0.8

Tabla 7.40: Resultados de la clasificación utilizando las componentes principales de cinco subgrupos de variables en la prueba IR.



## MEDIANTE VARIABLES SELECCIONADAS ORIGINALES

### Condición de aplicación

Se muestra, a continuación, el análisis discriminante realizado directamente a partir de las dieciocho variables seleccionadas mediante el MANOVA de medidas repetidas (ver Apartado 7.3). Las variables cumplen el test M de Box para la hipótesis de homogeneidad multivariable de la varianza con un p-valor de 0.25.

### Determinación de las funciones discriminantes

La función discriminante canónica obtenida mediante coeficientes estandarizados para separar dos poblaciones de sujetos se presenta a continuación en el mismo orden que entraron las variables en el método paso a paso:

$$f_{2\_grupos} = 2.338 \cdot lni - 2.097 \cdot tins2345-6 + 0.829 \cdot Ehl5-6 - 0.786 \cdot Efc12-345$$

Sin embargo, otras variables presentan significación con p-valor menor de 0.05 en el ANOVA separado de valores medios (Ehl.lin, arplpi0-1, Efc.lin y Ehl12-35), pero no aparecen en la función discriminante debido a estar correlacionadas con las primeras en entrar en la función.

Las funciones discriminantes obtenidas para separar tres grupos, según el orden de introducción de las variables, son las siguientes:

$$f_{3\_grupos}^1 = 1.216 \cdot lpi4-5 + 1.054 \cdot Efc12-345$$

$$f_{3\_grupos}^2 = -0.382 \cdot lpi4-5 + 0.717 \cdot Efc12-345$$

El parámetro de mayor importancia es lpi4-5 que por sí solo obtiene un nivel de significación de 0.003 en un ANOVA entre tres grupos. Esta variable no forma parte de la función discriminante para dos grupos a pesar de obtener por sí sola  $p < 0.02$  en el ANOVA de forma separada, por estar muy correlada con la variable más importante en este caso, lni, con un coeficiente de valor -0.760 significativo al nivel 0.01. El parámetro Efc12-345 aparece en los dos casos, sin embargo para la discriminación de tres grupos sólo alcanza  $p < 0.06$  en el ANOVA de variables independientes de forma separada.

### Clasificación de los individuos y discriminación de la función

En la tabla 7.41 se muestra los resultados relacionados con la clasificación de individuos en dos y tres poblaciones.

La clasificación entre enfermos graves y leves tiene mayor dificultad y se refleja en un p-valor más elevado. Todos los sujetos sanos se clasifican bien, pero un enfermo grave se clasifica como leve.

	Dos grupos	Tres grupos
$E_x$ (%)	100	91.6
$S$ (%)	100	83.3
$C_c$	0.98	0.93 , 0.40
$p <$	0.00003	0.001 , 0.2

Tabla 7.41: Resultados de la clasificación utilizando las variables originales seleccionadas en la prueba IR.

### 7.5.3.3 Conclusiones

Referente a la clasificación de los individuos obtenida mediante funciones discriminantes en el protocolo incremental se obtiene las siguientes conclusiones (ver tablas 7.36 a 7.41):

- En ambas pruebas se observa mayor dificultad para discriminar entre tres grupos de individuos (sujetos sanos, enfermos leves y enfermos graves) frente a la discriminación de dos grupos (pacientes con EPOC y grupo control) en todos los tipos de análisis discriminante.
- Se obtiene valores más elevados de exactitud y sensibilidad, 100% o próximos, en los análisis realizados sobre datos de la prueba IR que en la prueba IC.
- Al comparar los valores de  $E_x$  y  $S$  obtenidos en los análisis realizados con diferentes variables independientes, se concluye lo siguiente:
  - En la prueba IC se alcanza valores máximos al utilizar directamente las variables originales seleccionadas en el MANOVA de medidas repetidas. Dichos resultados son inferiores, aunque muy próximos, a los resultados obtenidos mediante factores provenientes de subgrupos. Por último se consigue la peor clasificación utilizando cuatro componentes principales.
  - En la prueba IR se logra las mejores clasificaciones mediante las componentes principales provenientes de subgrupos de parámetros y directamente de las variables originales seleccionadas. Se destaca la primera de ellas por alcanzar una sensibilidad del 100% en la separación de dos y tres grupos. La peor discriminación, especialmente con tres poblaciones, se obtiene al utilizar las cuatro componentes principales como variables predictoras.

Las funciones discriminantes son combinaciones lineales de variables independientes que son introducidas en un orden determinado mediante el método paso a paso. De esta forma, se observa la importancia de los parámetros en la separación de poblaciones y clasificación de individuos. A partir de estos resultados se obtiene las siguientes conclusiones:

- En general, se observa que las funciones discriminantes de dos y tres grupos de sujetos son diferentes por las variables que la componen y su combinación lineal.

- Los parámetros de correlación, especialmente el TRM de la FCC para retardos positivos normalizado respecto la duración del período inspiratorio, presentan un papel relevante en la discriminación en diferentes aspectos:
  - Variaciones entre apnea y volumen corriente:
 

Son introducidas en primer lugar en ambas pruebas en el análisis discriminante de dos grupos, así como al separar tres grupos en la prueba IR al utilizar las cuatro componentes principales como variables independientes.

En los otros dos tipos de análisis con diferentes variables predictoras, dichos parámetros no aparecen en la función discriminante aunque obtienen un p-valor estadísticamente significativo en el ANOVA de variables de forma separada debido a la elevada correlación con otras componentes o variables que son introducidas en primer lugar.
  - Valores promedio entre diferentes cargas y variaciones al final de la prueba:
 

Dichos parámetros aparecen en todas las funciones de los tres tipos de análisis discriminante de dos y tres grupos de sujetos. En especial cuando se considera como variables predictoras los factores provenientes de subgrupos: en la prueba IC las variaciones al final de la carga son introducidas en primer lugar en el modelo al separar dos y tres grupos, de igual forma que los valores promedio del TRM en la prueba IR.
- Además de la importancia de las variaciones de parámetros de correlación al final de la prueba mencionada anteriormente, la aparición en ocasiones de factores y variables en las funciones discriminantes de variaciones de variables temporales y frecuenciales al final de la prueba muestra el diferente comportamiento entre enfermos y sujetos sanos del músculo ECM ante esfuerzos elevados con cargas inspiratorias altas. Estos últimos parámetros aparecen especialmente en las funciones que discriminan dos poblaciones y en los casos que no es así, es debido a la correlación existente con otras variables o factores, ya que obtienen un p-valor en el ANOVA de forma separada estadísticamente significativo.

## 7.5.4 Protocolo mantenido

### 7.5.4.1 Resultados de la prueba M1

#### MEDIANTE CUATRO COMPONENTES PRINCIPALES

##### Condición de aplicación

Las variables independientes del análisis son las cuatro componentes principales (comp1, comp2, comp3 y comp4) extraídas en el análisis factorial realizado a las variables seleccionadas en el MANOVA de medidas repetidas de forma conjunta (ver Apartado 7.4.4.1). Mediante el test M de Box se comprueba que cumplen la hipótesis de homogeneidad multivariable de la varianza con un p-valor de 0.59.

### Determinación de las funciones discriminantes

Las funciones discriminantes canónicas para separar dos y tres poblaciones de sujetos son las siguientes según el orden de introducción del método paso a paso :

$$f_{2\_grupos} = 1.548 \cdot comp3 - 1.442 \cdot comp4 + 1.429 \cdot comp1 - 0.783 \cdot comp2$$

$$f_{3\_grupos}^1 = 1.246 \cdot comp3 + 1.134 \cdot comp1 - 1.185 \cdot comp4$$

$$f_{3\_grupos}^2 = -0.507 \cdot comp3 + 0.689 \cdot comp1 + 0.082 \cdot comp4$$

La tercera componente es la que mejor separa las poblaciones en dos y tres grupos de individuos. Dicha componente corresponde a variaciones del ritmo inspiratorio al inicio de la prueba. Las componentes 1ª y 4ª también aparecen en ambos casos y están formadas por valores promedio de parámetros frecuenciales y variaciones de éstos al inicio del test, respectivamente.

### Clasificación de los individuos y discriminación de la función

En la tabla 7.42 se presenta los resultados relacionados con la clasificación de individuos en dos y tres poblaciones.

	Dos grupos	Tres grupos
$E_x$ (%)	100	50
$S$ (%)	100	0
$C_c$	0.94	0.91 , 0.28
$p <$	0.01	0.01 , 0.71

Tabla 7.42: Resultados de la clasificación utilizando las cuatro componentes principales en la prueba M1.

### MEDIANTE COMPONENTES PRINCIPALES A PARTIR DE SUBGRUPOS

#### Condición de aplicación

Las seis componentes principales (a1, b1, c1, c2, d1 y d2) extraídas en el análisis factorial realizado a cuatro subgrupos de variables seleccionadas en el MANOVA de medidas repetidas (ver Apartado 7.4.4.1) son las variables independientes. Estas componentes cumplen la hipótesis de homogeneidad multivariable de la varianza con un p-valor de 0.87 en el test M de Box.

### Determinación de las funciones discriminantes

En la separación de dos y tres grupos de individuos, las funciones discriminantes obtenidas se presentan a continuación según el orden de introducción de la componente en el método paso a paso:

$$f_{2\_grupos} = -0.907 \cdot a1 + 0.837 \cdot d1 + 0.684 \cdot c1$$

$$f_{3\_grupos}^1 = 0.175 \cdot d2 + 1.415 \cdot c1 + 1.097 \cdot c2 + 0.630 \cdot d1 - 0.333 \cdot a1$$

$$f_{3\_grupos}^2 = -1.302 \cdot d2 + 0.453 \cdot c1 + 1.455 \cdot c2 - 0.775 \cdot d1 + 0.749 \cdot a1$$

En la discriminación de dos grupos, la componente a1 compuesta de variaciones de los parámetros frecuenciales al inicio de la prueba es la primera en entrar en la función:  $p < 0.01$  en el ANOVA de variables separadamente. Sin embargo, dicha componente entra en el cuarto paso en la función discriminante de tres poblaciones a pesar de obtener  $p < 0.03$  en el ANOVA de forma separada. Esto es debido a estar sensiblemente correlada con los tres factores que entran previamente en el método paso a paso: d2, c1 y c2 con coeficientes de correlación de -0.38, -0.41 y -0.47, respectivamente. Por esta razón, el estadístico F de cambio no es significativo. En la separación de dos grupos de sujetos sanos y pacientes con EPOC, las componentes que entran en el modelo a continuación de a1 son d1, asociada al valor medio de parámetros frecuenciales, y c1.

La componente b1 correspondiente a variaciones de parámetros al final de la prueba no aparece en la función discriminante a pesar de obtener  $p < 0.03$  debido a la alta correlación existente con la componente d1, coeficiente de Pearson -0.618 significativo al nivel 0.05 y la componente a1 con un coeficiente de valor 0.441. Ambas son introducidas en primer lugar en el modelo.

La componente que más discrimina tres poblaciones de individuos es d2 asociada al valor medio del TRM de la FCC para retardos positivos normalizado respecto al periodo inspiratorio: obtiene  $p < 0.03$  en el ANOVA de forma separada. A continuación, entran en el modelo las componentes c1 y c2 relacionadas con variaciones en el ritmo de inspiración y parámetros de correlación al inicio de la prueba, respectivamente.

#### Clasificación de los individuos y discriminación de la función

En la tabla 7.43 se muestra los resultados relacionados con la clasificación de individuos en dos y tres poblaciones.

	Dos grupos	Tres grupos
$E_x$ (%)	100	100
$S$ (%)	100	100
$C_c$	0.89	0.94 , 0.89
$p <$	0.003	0.002 , 0.02

Tabla 7.43: Resultados de la clasificación utilizando las componentes principales de cuatro subgrupos de variables en la prueba M1.

Es destacable que los valores de exactitud y sensibilidad es del 100% en el caso de dos y tres grupos.

#### MEDIANTE VARIABLES SELECCIONADAS ORIGINALES

##### Condición de aplicación

A continuación se realiza el análisis discriminante mediante las quince variables seleccionadas en el Apartado 7.3. De esta manera se compara los resultados con los análisis discriminantes anteriores. Dichas variables cumplen la hipótesis de homogeneidad multivariable de la varianza con un p-valor de 0.92 en el test M de Box.

### Determinación de las funciones discriminantes

En la discriminación de dos y tres grupos de sujetos, se obtiene las siguientes funciones canónicas según el orden de introducción de las variables en el método paso a paso obtenidas mediante coeficientes estandarizados:

$$f_{2\_grupos} = -0.880 \cdot Efc1-23 + 0.971 \cdot tinsciclo2-3 + 0.778 \cdot Efm$$

$$f_{3\_grupos}^1 = -0.400 \cdot lpi234 + 1.279 \cdot tinsciclo2-3 + 1.461 \cdot lpi3-4$$

$$f_{3\_grupos}^2 = 1.269 \cdot lpi234 + 0.230 \cdot tinsciclo2-3 + -0.572 \cdot lpi3-4$$

Comparando ambos casos, se observa que la variable relacionada con la variación del ritmo inspiratorio, normalizados respecto a la duración del ciclo respiratorio, aparece en ambas funciones en segundo lugar. También es destacable la aparición de parámetros frecuenciales en primer y tercer lugar en la función que separa dos grupos y, en cambio, parámetros de correlación al separar tres poblaciones. Por tanto, la diferencia entre las personas sanas y enfermas se puede detectar mejor a través de los primeros parámetros, mientras que las diferencias entre pacientes con diferente grado de obstrucción y los sujetos sanos se manifiestan en los parámetros de correlación.

### Clasificación de los individuos y discriminación de la función

Relacionado con la clasificación de individuos en dos y tres poblaciones se presenta la tabla 7.44.

	Dos grupos	Tres grupos
$E_x$ (%)	100	91.6
$S$ (%)	100	100
$C_c$	0.92	0.93 , 0.83
$p <$	0.001	0.0003 , 0.008

Tabla 7.44: Resultados de la clasificación utilizando las variables originales seleccionadas en la prueba M1.

#### 7.5.4.2 Resultados de la prueba M2

### MEDIANTE CUATRO COMPONENTES PRINCIPALES

#### Condición de aplicación

Las cuatro componentes principales (comp1, comp2, comp3 y comp4) extraídas en el análisis factorial de las variables seleccionadas en el MANOVA de medidas repetidas (ver Apartado 7.4.4.2) son las variables predictoras. Dichas componentes cumplen la condición de homogeneidad multivariable de la varianza con un p-valor de 0.21 en el test M de Box.

### Determinación de las funciones discriminantes

Las funciones discriminantes obtenidas en la separación de dos y tres poblaciones de sujetos son las siguientes presentando los factores según el orden de introducción en el método paso a paso:

$$f_{2\_grupos} = 1.214 \cdot comp4 - 1.071 \cdot comp1 - 0.981 \cdot comp2 + 0.946 \cdot comp3$$

$$f_{3\_grupos}^1 = 1.216 \cdot comp4 - 1.051 \cdot comp1 - 0.938 \cdot comp2 + 0.947 \cdot comp3$$

$$f_{3\_grupos}^2 = 0.057 \cdot comp4 + 0.653 \cdot comp1 - 0.642 \cdot comp2 + 0.074 \cdot comp3$$

Las cuatro componentes forman parte de las funciones en ambos casos observando que además de entrar en el modelo por el mismo orden, las funciones  $f_{2\_grupos}$  y  $f_{3\_grupos}^1$  corresponden prácticamente a la misma combinación lineal debido a una alineación de las poblaciones en el subespacio de componentes principales.

El factor más relevante es la 4ª componente relacionada con tendencias lineales del RMS en las señales miográficas. A continuación, el método introduce la 1ª componente asociada con valores promedio de parámetros frecuenciales y TRM de la FCC para retardos positivos. En último lugar aparece el tercer factor relacionado con el ritmo inspiratorio.

### Clasificación de los individuos y discriminación de la función

Se muestra los resultados relacionados con la clasificación de individuos en dos y tres poblaciones en la tabla 7.45.

	Dos grupos	Tres grupos
$E_x$ (%)	83.3	41.7
$S$ (%)	83.3	0
$C_c$	0.90	0.91 , 0.32
$p <$	0.007	0.08 , 0.84

Tabla 7.45: Resultados de la clasificación utilizando las cuatro componentes principales en la prueba M2.

Se obtiene una mala clasificación en la separación de tres grupos, donde el p-valor obtenido no es significativo y la sensibilidad está afectada porque los tres pacientes graves se clasifican como enfermos con severidad obstructiva leve.

### MEDIANTE COMPONENTES PRINCIPALES A PARTIR DE SUBGRUPOS

#### Condición de aplicación

Las variables independientes corresponden a las siete componentes principales ( $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $c_1$ ,  $c_2$  y  $d_1$ ) extraídas en el análisis factorial realizado a cuatro subgrupos de variables seleccionadas en el MANOVA de medidas repetidas (ver Apartado 7.4.4.2). En el test M de Box se obtiene un p-valor de 0.81. Por tanto, los factores cumplen la hipótesis de homogeneidad multivariable de la varianza.

### Determinación de las funciones discriminantes

Las funciones discriminantes obtenidas que separan dos y tres poblaciones de sujetos se componen de los siguientes factores según su orden de introducción en el método paso a paso:

$$f_{2\_grupos} = 0.413 \cdot a1 + 1.018 \cdot d1 - 0.806 \cdot c2$$

$$f_{3\_grupos}^1 = 0.128 \cdot a1 + 1.212 \cdot d1 - 1.077 \cdot c2$$

$$f_{3\_grupos}^2 = 1.142 \cdot a1 - 0.656 \cdot d1 + 0.813 \cdot c2$$

Las componentes que intervienen en el modelo y su orden de introducción son los mismos para la separación entre sujetos sanos y pacientes, así como individuos sanos con enfermos leves y graves. Las componentes en orden decreciente de importancia son las siguientes:

- a1: compuesta por parámetros temporales de variaciones al inicio de la prueba.
- d1: correspondiente a valores promedio de parámetros frecuenciales, TRM de la FCC para retardos positivos y duración del período inspiratorio.
- c2: asociada a la variación lineal del RMS de la señal EMG.

El factor a1 obtiene en un ANOVA de forma aislada  $p < 0.005$  y  $p < 0.01$  para dos y tres grupos, respectivamente. De esta forma, dicha componente es la que mejor discrimina y, por tanto, aparece en primer lugar en el modelo.

La componente b1 asociada a variaciones de parámetros al final de la prueba no aparece en la función discriminante porque está muy correlada con el factor a1, coeficiente de correlación de  $-0.630$  significativa al nivel de  $0.05$ , y el factor d1 con un coeficiente de valor  $0.459$ . Ambos entran en primer lugar en el modelo. Al realizar un ANOVA de forma separada de cada componente se obtiene un p-valor de  $0.02$ ,  $0.005$  y  $0.01$  para la discriminación de dos grupos en las componentes b1, a1 y d1, respectivamente, así como  $0.09$ ,  $0.01$  y  $0.03$  en tres poblaciones, respectivamente.

### Clasificación de los individuos y discriminación de la función

Para dos y tres poblaciones, en la tabla 7.46 se muestra los resultados de clasificación de individuos.

### MEDIANTE VARIABLES SELECCIONADAS ORIGINALES

#### Condición de aplicación

Se realiza el análisis discriminante mediante las dieciocho variables seleccionadas con el MANOVA de medidas repetidas (ver Apartado 7.3). Dichas variables predictoras cumplen la condición de homogeneidad multivariable de la varianza al obtener  $p < 0.55$  en el test M de Box.



	Dos grupos	Tres grupos
$E_x$ (%)	91.7	75.0
$S$ (%)	100	83.3
$C_c$	0.94	0.90 , 0.06
$p <$	0.001	0.01 , 0.25

Tabla 7.46: Resultados de la clasificación utilizando las componentes principales de cinco subgrupos de variables en la prueba M2.

#### Determinación de las funciones discriminantes

A continuación, se presenta las funciones discriminantes canónicas para separar dos y tres poblaciones de sujetos donde las variables aparecen en el mismo orden que entraron en el modelo:

$$f_{2\_grupos} = -8.530 \cdot tins1-234 + 19.889 \cdot lpi23 + 24.233 \cdot Efc3-45 \\ -17.065 \cdot Erme\_lin + 12.963 \cdot VErms34-5 - 2.819 \cdot Vhl1-23 + 3.585 \cdot lp1-2$$

$$f_{3\_grupos}^1 = -5.935 \cdot tins1-234 + 16.137 \cdot lpi23 + 21.687 \cdot Efc3-45 \\ -14.029 \cdot Erme\_lin + 10.385 \cdot VErms34-5 + 3.193 \cdot Vhl1-23 + 2.566 \cdot lp\_par$$

$$f_{3\_grupos}^2 = 0.617 \cdot tins1-234 - 1.924 \cdot lpi23 + 0.192 \cdot Efc3-45 \\ +0.182 \cdot Erme\_lin + 1.040 \cdot VErms34-5 + 0.205 \cdot Vhl1-23 + 1.045 \cdot lp\_par$$

Como en los análisis discriminante anteriores mediante puntuaciones factoriales, las variables involucradas y su orden de importancia en el análisis son prácticamente los mismos cuando se discrimina dos poblaciones en sujetos sanos y enfermos así como en tres grupos dividiendo los pacientes en graves y leves. La única variable diferente es el parámetro de correlación  $lp\_par$ , que es el que menos contribuye a la función discriminante, correlación canónica de la variable con la función de 0.011, que en el caso de dos grupos es sustituido por el parámetro  $lp1-2$ .

#### Clasificación de los individuos y discriminación de la función

Se muestra en la tabla 7.47 los resultados relacionados con la clasificación de individuos en dos y tres grupos.

	Dos grupos	Tres grupos
$E_x$ (%)	100	75
$S$ (%)	100	50
$C_c$	0.99	0.99 , 0.7
$p <$	0.000001	0.003 , 0.67

Tabla 7.47: Resultados de la clasificación utilizando las variables originales seleccionadas en la prueba M2.

En el caso de tres poblaciones, todos los sujetos sanos se clasifican bien, pero un enfermo leve y dos graves son intercambiados de grupo.

### 7.5.4.3 Conclusiones

Tras la observación de las tablas 7.42 a 7.47 correspondientes a la clasificación de los individuos mediante funciones discriminantes en protocolo mantenido se deduce las siguientes conclusiones:

- Se obtiene, en general, en los diferentes análisis discriminantes peores valores de  $E_x$  y  $S$  en la discriminación de tres grupos de individuos respecto a dos poblaciones.
- Se observa valores ligeramente más elevados de exactitud y sensibilidad, normalmente iguales o cercanos al 100%, en los análisis realizados en la prueba M1 respecto a la prueba M2.
- Mediante la comparación de valores de  $E_x$  y  $S$  obtenidos en los análisis realizados con diferentes variables predictoras, se obtiene las siguientes conclusiones:
  - En la prueba M1 se alcanzan los valores máximos del 100% en ambos índices y para la discriminación de dos y tres poblaciones cuando se utiliza los factores obtenidos mediante subgrupos de variables. Para separar tres grupos de sujetos, se alcanza valores del 91.6% y 100% de  $E_x$  y  $S$ , respectivamente, mientras que al utilizar cuatro factores se empeora ostensiblemente la clasificación.
  - En la prueba M2 se obtiene las mejores clasificaciones mediante las componentes principales provenientes de subgrupos de parámetros y directamente de las variables originales seleccionadas. En el primer análisis se obtiene una  $S$  de 83.3% al discriminar tres grupos, y en el segundo 50%. La peor clasificación se obtiene al utilizar las cuatro componentes principales como variables predictoras.

Al analizar la importancia de los diferentes parámetros en la separación de poblaciones y clasificación de individuos se concluye los siguientes aspectos:

- Las funciones discriminantes son similares por las variables que las componen y su combinación lineal al discriminar dos y tres grupos de sujetos. Esto sucede especialmente en la prueba M2 en que dichas funciones son prácticamente idénticas para los tres tipos de análisis discriminantes realizados según las variables predictoras utilizadas.
- El parámetro de correlación correspondiente al valor promedio del TRM de la FCC para retardos positivos normalizado respecto la duración del período inspiratorio es importante en ambas pruebas en la discriminación de dos y, especialmente, tres poblaciones:
  - En la prueba M1 entra en primer lugar en el modelo de separación de tres poblaciones cuando se utiliza como variables predictoras las seleccionadas en el MANOVA de medidas repetidas y factores procedentes de subgrupos de variables.

- En la prueba M2 entra en el método paso a paso en segundo lugar en los tres tipos de análisis discriminante de dos y tres grupos de individuos.
- Las variaciones de los parámetros temporales, especialmente el ritmo inspiratorio, y en menor medida frecuenciales (en la prueba M1) al inicio de la prueba aparecen en primer lugar en la función discriminante de dos grupos de individuos en ambas pruebas mantenidas y en la separación de tres grupos en segundo y primer lugar en la prueba M1 y M2, respectivamente.
- Otros parámetros de interés en la discriminación son los siguientes:
  - Valores promedio de componentes frecuenciales de la señal EMG.  
Aunque visual y estadísticamente en el MANOVA de medidas repetidas parecía muy importante en la separación de poblaciones, en el modelo discriminante aparece sólo en segundo plano en la prueba M1, especialmente.
  - Variaciones de parámetros al final de la prueba.  
En este caso se demuestra en ambas pruebas una considerable discriminación sobre todo de dos poblaciones de sujetos mediante el ANOVA de variables independientes de forma separada. Sin embargo no son introducidas en el modelo debido a su correlación con valores promedio de parámetros frecuenciales y especialmente variaciones de variables al inicio de la prueba. Estas dos últimas componentes entran previamente en el método paso a paso.

## 7.5.5 Protocolo completo

### 7.5.5.1 Resultados de las cuatro pruebas

#### Condición de aplicación

Para comparar de forma objetiva y directa los cuatro protocolos respiratorios se pretende realizar el siguiente análisis discriminante: las variables independientes son las puntuaciones factoriales de los doce individuos en las dos componentes principales extraídas en el análisis factorial de dos factores (ver Apartados 7.4.3.1 a 7.4.4.2). Es decir, se utiliza ocho variables predictoras que cumplen la condición de homogeneidad multivariable de la varianza con un p-valor de 0.47 en el test M de Box.

Analizando la relación entre las diferentes componentes y a partir de la matriz de correlación se observa los siguientes aspectos (ver tabla 7.48):

- Las dos componentes de cada protocolo están totalmente incorreladas ya que se trata de factores ortonormales: 1-M1 y 2-M1, 1-M2 y 2-M2, 1-IC y 2-IC, 1-IR y 2-IR.
- Las primeras componentes de cada protocolo están correladas entre ellas (1-IC, 1-IR, 1-M1, 1-M2), así como las segundas componentes (2-IC, 2-IR, 2-M1, 2-M2).

- La máxima correlación en cada componente aparece con su componente homónima correspondiente a la otra prueba del mismo protocolo incremental o mantenido. Esto sucede en todos los casos a excepción de los factores 1-IC y 1-IR cuyos coeficientes de correlación, aunque no son los máximos entre ellos, 0.746, se acerca mucho: 0.781 y 0.822, respectivamente.

	1-IC	2-IC	1-IR	2-IR	1-M1	2-M1	1-M2	2-M2
1-IC	1	0	0.746	0.250	0.602	-0.195	0.781	-0.120
2-IC	0	1	-0.017	-0.672	-0.345	0.486	-0.234	0.622
1-IR	0.746	-0.017	1	0	0.687	-0.194	0.822	0.014
2-IR	0.250	-0.672	0	1	0.196	-0.604	0.133	-0.801
1-M1	0.602	-0.345	0.687	0.196	1	0	0.885	-0.009
2-M1	-0.195	0.486	-0.194	-0.604	0	1	-0.088	0.891
1-M2	0.781	-0.234	0.822	0.133	0.8851	-0.088	1	0
2-M2	-0.120	0.622	0.014	-0.801	-0.009	0.891	0	1

Tabla 7.48: *Matriz de correlación de las componentes principales de las cuatro pruebas respiratorias.*

#### Determinación de las funciones discriminantes

En la discriminación de dos poblaciones de sujetos sanos y pacientes con EPOC, mediante el ANOVA de cada factor de forma separada, se halla que las componentes principales que mejor diferencian los dos grupos de individuos son las segundas componentes de los dos protocolos mantenidos: 2-M1:  $p < 0.007$ , 2-M2:  $p < 0.006$ . En tercer y cuarto lugar se obtiene las componentes 2-IC y 2-IR con  $p < 0.02$ .

La función discriminante obtenida es la siguiente:

$$f_{2\_grupos} = -1.417 \cdot 2-M2 + 0.790 \cdot 1-IR - 1.417 \cdot 1-M1$$

Se presenta la combinación lineal de las componentes según el orden de entrada en el modelo mediante el método paso a paso. La no inclusión de la 2 componente de la primera prueba mantenida en la función discriminante es debida a su alta correlación con la componente 2-M2, no a que no separe las poblaciones pues obtiene  $p < 0.007$  en el ANOVA de forma separada. La 1ª componente de la prueba M1 sólo aparece como tercera variable en importancia, pues entra primero la 2ª componente de la prueba M2 que está muy correlada con ésta.

En la separación de tres grupos, sujetos sanos y enfermos con nivel de severidad obstructiva respiratoria leve y grave, la función discriminante calculada es la siguiente:

$$f_{3\_grupos}^1 = 1.085 \cdot 1-IC - 2.157 \cdot 1-IR + 1.520 \cdot 2-M2$$

$$f_{3\_grupos}^2 = 1.669 \cdot 1-IC - 1.130 \cdot 1-IR - 0.091 \cdot 2-M2$$

Se presenta la combinación lineal de los factores en la función según su orden de entrada en el modelo. Sin embargo, la introducción de las componentes ha sido diferente en el

método paso a paso como se observa en la tabla 7.49 al calcularse un estadístico F cuando cada variable es eliminada del modelo, que si es inferior a la  $F_s$  fijada con el valor 1.3 sale de la función.

Paso	Variable	F a eliminar	$\lambda_w$	Paso	Variable	F a eliminar	$\lambda_w$
1	2-M1	6.278		5	2-M1	0.035	0.040
2	2-M1	11.329	0.700	5	1-M1	0.701	0.051
	1-M1	5.137	0.418		1-IC	5.297	0.124
3	2-M1	9.548	0.473		1-IR	3.540	0.096
	1-M1	2.672	0.224		2-M2	1.601	0.065
	1-IC	1.544	0.183	6	1-M1	0.914	0.053
4	2-M1	8.107	0.242	1-IC	6.368	0.126	
	1-M1	1.699	0.102	1-IR	6.440	0.127	
	1-IC	5.049	0.175	2-M2	14.969	0.242	
	1-IR	2.817	0.127	7	1-IC	7.810	0.170
				1-IR	9.793	0.200	
				2-M2	15.353	0.284	

Tabla 7.49: *Introducción y eliminación de las variables para formar la función discriminante para tres grupos por el método paso a paso.*

La 2ª y 1ª componente de la prueba M1 son las primeras en entrar en la función discriminante, aunque en el paso seis son eliminadas del modelo debido las correlaciones con las otras componentes. La componente 2-M1 obtiene  $p < 0.02$  en el ANOVA de forma separada. La 1ª componente de la prueba IR queda al final en segundo lugar de la función discriminante de igual manera que en la separación de dos grupos de individuos.

#### Clasificación de los individuos y discriminación de la función

Se muestra en la tabla 7.50 los resultados relacionados con la clasificación de individuos en dos y tres grupos.

	Dos grupos	Tres grupos
$E_x$ (%)	100	83.3
$S$ (%)	100	83.3
$C_c$	0.92	0.92 , 0.78
$p <$	0.001	0.001 , 0.02

Tabla 7.50: *Resultados de la clasificación utilizando ocho componentes principales de las cuatro pruebas.*

#### Determinación de la prueba principal

Se decide realizar dos análisis sin las componentes de la segunda prueba mantenida ya que aportan información redundante a las de la primera. En el primer análisis se tiene en cuenta las componentes de la prueba M1 e IC y en el segundo los factores de los tests M1 e IR. Las funciones discriminantes obtenidas en ambos casos para la separación de dos y tres grupos son las siguientes según el orden de introducción en el modelo:

$$f_{2\_grupos} = 1.238 \cdot 2-M1 - 1.091 \cdot 1-M1$$

$$f_{3\_grupos}^1 = 1.209 \cdot 2-M1 - 1.034 \cdot 1-M1$$

$$f_{3\_grupos}^2 = 0.371 \cdot 2-M1 + 0.729 \cdot 1-M1$$

Se observa que en la separación de dos y tres poblaciones, los mejores resultados se obtienen con la función discriminante formada únicamente por las componentes de la primera prueba mantenida.

### 7.5.5.2 Conclusiones

De los resultados obtenidos al realizar el análisis discriminante de factores provenientes de las cuatro pruebas conjuntamente se obtienen dos conclusiones importantes:

- Los modelos obtenidos en las pruebas son similares, en especial entre pruebas mantenidas, así como entre incrementales debido a las correlaciones entre los factores.
- La prueba respiratoria que más importancia tiene en la separación de poblaciones es la M1.

## 7.5.6 Discusión y conclusiones

Mediante el análisis discriminante se ha conseguido obtener, en general, una buena clasificación de los individuos cuando se ha intentado separar en dos poblaciones de pacientes con EPOC y sujetos sanos, así como en tres grupos correspondientes a personas sanas, enfermos con nivel de severidad obstructiva respiratoria leve y grave. Dicha clasificación se ha evaluado mediante la correlación canónica, el p-valor obtenido mediante el ANOVA de las puntuaciones discriminantes de cada individuo, la exactitud y la sensibilidad. A partir de estos dos últimos índices se obtiene las siguientes conclusiones:

- Se observa mayor dificultad para discriminar entre tres grupos de individuos en los diferentes tipos de análisis discriminante: a partir de las componentes principales y de las variables seleccionadas en el MANOVA de medidas repetidas. Una de las causas de este fenómeno es el bajo número de individuos en los dos grupos de enfermos:
  - En la discriminación entre dos grupos de individuos, en todas las pruebas se consigue unos porcentajes de exactitud y sensibilidad del 100% y a excepción de la prueba IC que se obtiene el 92% y 100%, respectivamente.
  - En la discriminación entre tres grupos de individuos, se observa peores porcentajes y sólo en la prueba M1 se alcanza valores de exactitud y sensibilidad del 100%. En cualquier caso, estos resultados son orientativos debido al reducido número de individuos en cada grupo de enfermos.

- Las mejores pruebas para separar poblaciones es la prueba M1 y la IR, permitiendo la eliminación de las otras dos con el considerable ahorro de tiempo, dinero y principalmente esfuerzo que realiza el paciente. En concreto, es en la prueba M1 que se manifiesta mayor diferencia de comportamiento del músculo ECM entre sujetos sanos y enfermos con diferente grado de obstrucción respiratoria al alcanzar valores de sensibilidad y exactitud del 100%,  $p < 0.003$  y  $p < 0.002$  en la discriminación de dos y tres poblaciones, respectivamente.
- Realizando un estudio comparativo de los diferentes tipos de variables independientes utilizadas en el análisis discriminante se observa los siguientes aspectos:
  - Se consigue los mejores porcentajes de exactitud y, en especial, de sensibilidad al utilizar las componentes principales obtenidas del análisis factorial de subgrupos de variables. Esto sucede en todas las pruebas a excepción de la IC en que los porcentajes son muy cercanos a los obtenidos mediante las variables seleccionadas que son los más altos en este caso. En el resto de pruebas estas últimas variables también obtienen valores de  $S$  y  $E_x$  iguales o próximos al 100%. Sin embargo, al utilizar cuatro factores los resultados empeoran considerablemente, en especial en la discriminación de tres poblaciones.
  - La interpretación del análisis discriminante se realiza mucho mejor mediante componentes principales provenientes de subgrupos que con las variables originales ya que una única variable, por azar, puede separar bien una población y enmascarar otros efectos ya que, muy probablemente debido a altas correlaciones entre ellas, parámetros de interés no son incluidos en la función discriminante.

Al analizar las funciones discriminantes obtenidas, las variables predictoras que intervienen y su inclusión en el método paso a paso se deduce las conclusiones siguientes:

- Mientras que en los protocolos incrementales las funciones discriminantes de dos y tres grupos son diferentes en su composición y combinación lineal, en las pruebas mantenidas son muy similares, especialmente en la prueba M2 que son prácticamente idénticas. Puede ser debido a un efecto de aprendizaje del sujeto en la realización de todo el protocolo respiratorio.
- El tipo de parámetro con mayor importancia en la discriminación son los parámetros de correlación por los siguientes aspectos:
  - Variaciones entre apnea y volumen corriente.  
Sólo se evalúa en el protocolo incremental, pero produce una buena discriminación entre tres y, especialmente, dos poblaciones de sujetos.
  - Variaciones al final de la prueba.  
Aparece en las funciones discriminantes de dos y tres grupos en el protocolo incremental mientras que no son relevantes en la separación de poblaciones durante el protocolo mantenido.

- Valores promedio del TRM de la FCC para valores positivos normalizado respecto a la duración del período inspiratorio.

Este parámetro, relacionado con el retardo de la energía de la señal VMG respecto a la señal EMG, se introduce en los primeros pasos del método en los protocolos incremental y mantenido contribuyendo a la función discriminante de dos y, muy especialmente, de tres poblaciones.

- Las variaciones de parámetros temporales, en especial del ritmo inspiratorio, al inicio de la prueba se introduce en los primeros pasos en las funciones discriminantes de dos y tres poblaciones en el protocolo mantenido. Dichas variaciones están relacionadas con un efecto de adaptación del sujeto al esfuerzo realizado al soportar el 80% de la MCS.
- Los parámetros frecuenciales sólo aparecen en las funciones discriminantes en ciertos casos: variaciones al inicio de las pruebas IR y M1, y en mayor medida, los valores promedio de componentes frecuenciales de la señal EMG en el protocolo mantenido. En ambos casos se trata de la separación de dos poblaciones de individuos.
- Pensar que la fatiga muscular al final de las pruebas es similar en todos los individuos, independientemente de presentar la enfermedad por no formar parte de las funciones discriminantes en la mayoría de ocasiones, es erróneo. La razón es que dicha componente está correlada con otros factores más discriminantes como las frecuencias promedio de la señal EMG y variaciones de parámetros al inicio del test. El ANOVA de variables separadamente demuestra que presenta discriminación significativa de dos grupos de sujetos sanos y pacientes con EPOC. Por tanto, esfuerzos musculares elevados mediante cargas cercanas a la MCS en el protocolo incremental y en el tramo final del protocolo mantenido produce un diferente comportamiento del músculo ECM entre personas enfermas y sanas, aunque no es el efecto más discriminante.

## 7.6 **Discusión y conclusiones**

En este Apartado se enumera las conclusiones obtenidas respecto a los objetivos iniciales de la tesis a partir de los resultados obtenidos del análisis estadístico multivariable descrito en el presente capítulo:

- **Metodología de análisis.**

Se ha generado una metodología de análisis de gran cantidad de datos que a partir de diferentes técnicas estadísticas multivariables permite los siguientes aspectos:

- Realizar un análisis de los parámetros seleccionando los más significativos para solucionar el problema en estudio.



- Interpretación de la evolución del comportamiento fisiológico o función durante pruebas clínicas de larga duración.
- Reducción del número de datos manteniendo la máxima información.
- Comparar diferentes pruebas o protocolos que obtengan información similar o redundante y decidir cuáles son las más interesantes en la aplicación de ingeniería biomédica.
- Obtener un sistema de ayuda al diagnóstico.

Dicha metodología se ha aplicado en el caso de la EPOC y en el estudio de la actividad del músculo ECM mediante diferentes protocolos respiratorios, y se compone de las siguientes fases:

### 1. Análisis multivariable de la varianza para diseño de medidas repetidas.

Se estudia la evolución de los parámetros a lo largo de los diferentes protocolos para analizar la actividad del músculo ECM durante las pruebas respiratorias mediante:

- La creación de factores Ciclo, Grupo, Tramo, y la interacción Grupo\*Tramo.
- La utilización de contrastes Diferencia, Repetido, Helmert y Polinómico.

Se genera y selecciona un conjunto de nuevas variables procedentes de la combinación de los datos originales denominadas en el estudio variables seleccionadas, con significación estadística en el problema analizado.

### 2. Análisis factorial.

Las variables seleccionadas se representan en unos nuevos ejes de coordenadas o componentes principales caracterizados por contener la mayor variabilidad de los datos en el mínimo número de ejes y reducir la información redundante en el modelo. Se ha diseñado dos estrategias de análisis:

- Análisis de componentes principales de todas las variables seleccionadas conjuntamente, observando su agrupación en los mismos ejes conteniendo información similar.
- Análisis de distintos subgrupos de variables de forma separada. Dichos subgrupos se forman dependiendo del instante de la prueba que se ha calculado las variables. De esta forma, se consigue mantener una alta interpretación de la función muscular al estar asociados los nuevos factores a efectos como la adaptación al esfuerzo, la fatiga muscular, la evolución durante la prueba o las cargas ventilatorias bajas medias y altas.

A continuación, se calcula los valores de la representación de los individuos en estos ejes, denominados puntuaciones factoriales.

### 3. Análisis discriminante.

Se ha estudiado y aplicado la clasificación de casos mediante el método paso a paso basado en la Lambda de Wilks. Las puntuaciones factoriales, calculadas mediante las dos estrategias anteriores, se utilizan en el análisis para calcular la función discriminante que permite los siguientes aspectos:

- Clasificar nuevos casos o individuos.
- Observar qué tipo de variables resultan más efectivas al realizar dicha clasificación.
- Encontrar los protocolos o pruebas clínicas más interesantes como herramienta de ayuda al diagnóstico.

### ● **Análisis del grupo control y el grupo de pacientes con EPOC**

La metodología generada se ha aplicado a la evaluación de la actividad muscular respiratoria comparando un grupo control de personas sanas y un grupo de pacientes con EPOC que, a su vez, se ha dividido en enfermos con nivel de severidad obstructiva respiratoria leve y grave. Se ha obtenido resultados a partir de los datos de las señales biomédicas que diferencian la función muscular y respiratoria entre las diferentes poblaciones en estudio consiguiendo una nueva herramienta de ayuda al diagnóstico.

#### 1. *Análisis multivariable de la varianza de medidas repetidas*

Se ha realizado un análisis exhaustivo de las señales EMG y VMG en el dominio temporal y frecuencial y se ha observado mejores resultados al combinar información de ambas señales miográficas. Mediante la evaluación del efecto Ciclo se ha reducido el número de datos al calcular el promedio de los parámetros en los cinco ciclos seleccionados.

A través de las variables transformadas obtenidas al aplicar los contrastes se puede observar los efectos de adaptación y fatiga durante las pruebas, al inicio y final de éstas respectivamente mediante el efecto Tramo.

Además, se ha observado diferencias significativas entre el grupo de pacientes y el grupo control con el efecto Grupo y su interacción con el efecto Tramo. A partir de este estudio se ha seleccionado los parámetros temporales, frecuenciales y de correlación de mayor interés. Se puede destacar los siguientes:

- Incremento del valor de los parámetros de RMS de la señal EMG reflejado en el contraste lineal durante el protocolo incremental asociada al reclutamiento de fibras musculares cuando aumenta la carga.
- Variación de los parámetros de correlación entre la prueba apnea y la respiración a volumen corriente, así como al final del protocolo.
- Cambios en los parámetros frecuenciales y en el ritmo inspiratorio al inicio del protocolo mantenido indicando una adaptación del individuo al esfuerzo muscular.
- Al final de las pruebas ventilatorias, cambios en los siguientes parámetros:
  - \* Frecuenciales: relacionados con fatiga de las fibras musculares y cambios en la velocidad de conducción.
  - \* Temporales provenientes del RMS: la energía de la señal EMG incrementa en mayor medida que la de la señal VMG indicando una pérdida de la eficiencia mecánica-eléctrica del músculo.

- Valores promedio durante la prueba de los siguientes parámetros:
  - \* Frecuenciales de la señal EMG: indica mayor número de componentes frecuenciales en pacientes con EPOC que en personas sanas.
  - \* TRM de la FCC para retardos positivos normalizado respecto a duración del período inspiratorio: indica un mayor retardo de la energía de la señal VMG respecto a la señal EMG en enfermos que en el grupo control.
  - \* Duración del período de inspiración: el ritmo inspiratorio es mayor en pacientes con EPOC.

## 2. Análisis factorial de componentes principales.

A través de este análisis se ha reducido considerablemente el número de variables al situarse entre cuatro y ocho factores. A partir de los resultados se ha obtenido las siguientes conclusiones:

- En la extracción de factores de todas las variables seleccionadas conjuntamente se observa que parámetros relacionados entre sí se proyectan sensiblemente en las mismas componentes principales demostrando la eliminación de información redundante en el modelo.
- Los factores se distribuyen y componen de forma parecida entre pruebas obteniendo modelos similares en el protocolo incremental y mantenido. Esto demuestra la adecuación del método utilizado.
- Los parámetros se han agrupados en subgrupos según los instantes de la prueba en que han sido calculados:
  - \* Al inicio de la prueba mostrando la adaptación del sujeto al esfuerzo.
  - \* Al final de la prueba relacionado con la posible fatiga muscular respiratoria del individuo.
  - \* En fases intermedias de la prueba.
  - \* Evolución durante toda la prueba mediante tendencias lineales, especialmente interesantes en el protocolo incremental.
  - \* Valores promedio durante la prueba asociado a un efecto permanente durante el protocolo.
- La representación de los individuos en las componentes principales demuestra la posibilidad de discriminación entre grupos de sujetos sanos y enfermos a través de las señales biomédicas.

## 3. Análisis discriminante

A partir de los resultados obtenidos mediante funciones discriminantes se observa los siguientes aspectos:

- Se ha obtenido, en general, buenas clasificaciones de los individuos evaluadas mediante la correlación canónica, el p-valor de las puntuaciones discriminantes, la  $E_x$  y  $S$ . Aparece mayor dificultad en la clasificación de tres poblaciones que en dos grupos debido al reducido número de sujetos.

- Los mejores valores de exactitud y sensibilidad, próximos al 100%, se consigue utilizando como variables independientes las componentes principales extraídas de diferentes subgrupos de parámetros. En segundo lugar, mediante las variables seleccionadas en el MANOVA de medidas repetidas se consigue porcentajes cercanos a los del primer análisis. Por último, los resultados empeoran notablemente al utilizar factores extraídos de todas las variables conjuntamente.
- La utilización de componentes principales como variables independientes del análisis, en lugar de variables originales, facilita la interpretación pues elevadas correlaciones entre las variables originales seleccionadas enmascararían efectos de algunos parámetros de interés.
- Las pruebas que muestran mayor diferencia de comportamiento del músculo ECM entre dos y tres poblaciones son la IR y, especialmente, la M1. En esta última se consigue porcentajes de  $E_x$  y  $S$  del 100% en ambos casos y  $p < 0.003$  y  $p < 0.002$  en la discriminación de dos y tres grupos, respectivamente.
- Los parámetros de mayor importancia en la discriminación son los parámetros de correlación correspondientes a:
  - \* Variaciones entre apnea y volumen corriente en el protocolo incremental en la separación de tres y, especialmente, dos poblaciones.
  - \* Cambios al final de las pruebas durante el protocolo incremental.
  - \* Valores promedio del TRM de la FCC en ambos protocolos para discriminar dos y, especialmente, tres poblaciones.
- También son importantes los parámetros temporales, en especial el ritmo inspiratorio durante la adaptación del sujeto al inicio de la prueba.
- La fatiga muscular, aunque indica un diferente comportamiento del músculo entre individuos de diferentes grupos, no aparece reflejado en las funciones discriminantes debido a la correlación con parámetros de mayor discriminación.
- De esta forma se ha analizado el tipo de variables y protocolos ventilatorios que discriminan mejor las poblaciones, permitiendo la reducción de complejidad computacional y del número de pruebas, con el consiguiente ahorro de tiempo, dinero y especialmente esfuerzo del paciente.

Por todo ello, se ha demostrado que el estudio de señales miográficas es de gran interés en el estudio de la actividad muscular y en la evaluación de la EPOC. También, se ha comprobado la aplicabilidad de la metodología generada en la presente tesis para el análisis exhaustivo de datos procedente de las señales biomédicas, y de los problemas que éstas abordan.