

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



TRABAJO FIN DE MÁSTER

**Máster Universitario en
Ingeniería de Telecomunicación**

**Análisis biométrico de escritura para estudio de trastornos
neurodegenerativos.**

María Leal Salido

Junio de 2017

**Análisis biométrico de escritura para estudio de trastornos
neurodegenerativos.**

AUTOR: María Leal Salido
TUTOR: Aythami Morales Moreno

Dpto. Biometric Recognition Group - ATVS
Escuela Politécnica Superior
Universidad Autónoma de Madrid

A todas las personas que me han ayudado y han hecho que esto sea posible.
Gracias por todo vuestro apoyo.

Agradecimientos

Quisiera agradecer a varias personas la ayuda que me han prestado en la realización de este Trabajo Fin de Máster. Entre ellas, y en primer lugar, a mi tutor de proyecto Aythami, por todo lo que me han enseñado durante la asignatura de PASM, así mismo como por su ayuda, sin la cual hubiera sido muy difícil la realización de este trabajo. Destacar que siempre ha estado dispuesto a dedicarme su tiempo y su conocimiento.

Me gustaría agradecer a Daniel su ayuda. Sin duda alguna, es la persona que más desinteresadamente me ha ayudado, y siempre ha estado ahí cuando le he necesitado. Gracias por ser un gran amigo, y mejor persona.

Para finalizar, mostrar mi agradecimiento también a todas aquellas personas que han ayudado en mayor o menor medida a la realización de este proyecto.

Resumen

El presente proyecto analiza la escritura manuscrita como biomarcador de trastornos neurodegenerativos como el Parkinson. Se ha empleado la base de datos PaHaW, con 75 sujetos que realizan 8 tareas de escritura simples. De cada tarea se han obtenido 100 características globales y 91 características neuromotoras.

Se analiza el poder discriminante de las características en base a tres criterios: 1) distancia entre distribuciones de sujetos con trastorno y sujetos control; 2) diferencias estadísticas significativas y 3) correlación entre características. Con cada uno de los conjuntos se han realizado pruebas de clasificación binaria empleando SVM. El mejor resultado de clasificación ha sido del 22.5% de Equal Error Rate, se ha conseguido combinando las 8 tareas de escritura y empleando características globales con diferencias estadísticas significativas.

Como conclusiones, destacar que en la actualidad no hay ninguna prueba fiable que permita objetivar el estado real de pacientes con Parkinson. Entre las más de 100 características barométricas evaluadas en este trabajo, se ha observado que algunas de ellas pueden resultar de utilidad como biomarcador de este tipo de trastornos. Sin embargo, la elevada dependencia de usuario (edad, medicación, estadio de la enfermedad,..) complica el modelado automático. Los resultados obtenidos en este proyecto animan a seguir investigando en esta línea.

Abstract

This project analyze the handwritten as biomarker of neurodegenerative disorders like Parkinson. It has been used the PaHaW database, with 75 subjects that make 8 task of simple scripture. Of each task it has been obtained 100 global characteristics and 91 neuromotor characteristics.

It's analyze the discriminating power of the characteristics based on three judgments: 1) distance between distributions of subjects with disorder and control subjects; 2) Statistically significant differences; and 3) correlation between characteristics. With each one of the sets, it has been realizes binary classification tests using SVM. The best classification result has been 22.5% of Equal Error Rate, It has been obtained combining the 8 scripture tasks and using global characteristics with statistically significant differences.

To conclude, we can stand out that, in the actuality there isn't any reliable proof which permits objectify the real state of the subjects with Parkinson. Between more of 100 barometrics characteristics evaluated in this project, it has been observed that some of that can be useful as biomarker of this type of illness. However, high user dependency (age, treatment, stage of the disease,...) complicates the automatic model. The results obtained in this project encourage to investigate in this line.

Acrónimos

Símbolo	Nombre completo
PD	Parkinson Disease
PaHaW	Parkinson's Disease Handwriting Database
NPC	Número de cambios en la presión
H	Healthy
UPDRS III	Unified Parkinson's Disease Rating Scale III
SVM	Support Vector Machines
RBF	Radial Basis Function
FAR	Tasa de falsa aceptación
FRR	Tasa de falso rechazo
EER	Tasa de igual error
KNN	K-Nearest Neighbors

Índice general

Índice general	11
I Memoria	1
1. Introducción	3
1.1. Motivación	6
1.2. Objetivos y metodología	7
1.3. Organización de la memoria	8
2. Características biométricas de la escritura como biomarcador de apoyo al diagnóstico de Parkinson	9
2.1. Base de datos	9
2.1.1. Pruebas	11
2.2. Análisis de características globales	15
2.3. Análisis de características neuromotoras	21
2.3.1. Caracterización neuromotora de los usuarios	26
3. Experimentos y resultados	33
3.1. Combinación de características globales	33
3.2. Combinación de características neuromotoras	40
3.3. Clasificación basada en Máquinas de Vectores de Soporte	45

3.4. Resultados de clasificación	48
4. Conclusiones	51
4.1. Líneas futuras	52
Referencias	55
II Anexo	59

Parte I

Memoria

Introducción

El Parkinson (PD) es el segundo trastorno neurodegenerativo en prevalencia e incidencia después del Alzheimer [1]. Se caracteriza por la pérdida progresiva de la coordinación de movimientos, lo que se traduce en rigidez, lentitud y temblores en la realización de los mismos. Además aparecen otros síntomas como depresión, trastornos del sueño, pérdida de reflejos, incontinencia urinaria, dificultades para tragar, masticar o hablar, o escritura irregular. Dicho trastorno se origina cuando las neuronas no producen suficiente cantidad de dopamina, una sustancia química neurotransmisora cuya función principal es el control de los movimientos.

Un problema que afecta al tratamiento precoz y al seguimiento de este tipo de trastorno es que, en la actualidad no existen análisis o pruebas para diagnosticar el avance de la enfermedad con precisión. Además el diagnóstico puede confundirse con otras patologías principalmente en las primeras fases de desarrollo. Las pruebas actuales permiten estimar el nivel de desarrollo de la enfermedad de forma cualitativa. A día de hoy, las pruebas que permiten objetivar el estado real del mismo se realizan una vez fallecido el paciente a través del estudio patológico del cerebro durante la autopsia.

En la actualidad este trastorno no tiene cura, sólo se puede intentar reducir el avance del mismo mediante tratamientos médicos o estimulación cerebral profunda. Dentro de los tratamientos médicos se encuentra la levodopa, que es el más eficaz para pacientes con PD, puesto que se trata de un precursor metabólico de la dopamina. La estimulación cerebral profunda consiste en la implantación quirúrgica en el cerebro de unos electrodos, que estimulan la zona cerebral encargada del control de los movimientos.

Cabe destacar que la enfermedad afecta principalmente a personas de avanzada edad, pero no es raro encontrar a personas menores de 40 años que ya manifiesten síntomas. En el estudio [2] realizado por el NINDS (National Institute of Neurological Disorders and Stroke) se reportan tasas de incidencia estandarizada en los países desarrollados de 14 personas por cada 100.000 individuos año, pero si se restringe la población de

estudio a individuos con edad superior a 65 años la incidencia es de 160 por cada 100.000 individuos año. Los estudios basados en registros clínicos hospitalarios, hablan de una prevalencia mundial entre 100-300 casos por 100.000 habitantes [3]. En cuanto al origen de la enfermedad es desconocido en un 85 % de los casos, en un 10 % tiene origen genético y un 5 % se podría deber a las condiciones ambientales o a un origen tóxico [4].

Los accidentes cerebrales y los trastornos neurodegenerativos guardan cierta relación. En [5], se ha demostrado que antes de sufrir un accidente cerebral se muestran acontecimientos o síntomas con los que se podría detectar de manera precoz. Por ejemplo en [5] se señala que los factores de riesgo de sufrir un accidente cerebro vascular están asociados con el deterioro de características cognitivas y psicomotoras. Esto mismo sucede con otras enfermedades como Parkinson o esquizofrenia donde un análisis de patrones finos del movimiento se emplea para prevenir dichas enfermedades [6]. Los patrones finos del movimiento son acciones innatas que no se aprenden puesto que están integradas en el ser humano.

La escritura es un biomarcador que permite caracterizar a las personas. Mediante el estudio biométrico de la escritura, se puede modelar el estado de los pacientes que presentan problemas neurodegenerativos y objetivarlo a partir de parámetros cuantificables. Las anomalías pueden aparecer años antes de la realización de un diagnóstico clínico, y por tanto puede ser uno de los primeros signos de enfermedad inminente, facilitando así tratamientos preventivos [7]. El uso de la escritura como biomarcador se fundamenta en que el sujeto ha aprendido a escribir antes de sufrir los efectos de la enfermedad, y esa escritura tras la enfermedad, comienza a manifestar alteraciones [8]. Además facilita la temprana detección de trastornos. Por ser un método no invasivo y por permitir un diagnóstico rápido. Se podría aplicar como método inicial para remitir al paciente a un especialista, que diagnosticase y tratase su caso. Además de para la monitorización del estado de los mismos. Los pacientes con PD tienden a realizar una escritura de tamaño más pequeño de lo normal, con continuas correcciones, con picos de aceleración variables e incluso pausas prolongadas entre una palabra y la siguiente [9].

En [10] se emplea la escritura como biomarcador para caracterizar a las personas con PD. Se emplea la base de datos pública PaHaW (Parkinson's Disease Handwriting Database) para obtener la información de posicionamiento absoluta del lápiz, así como su movimiento sobre la superficie del dispositivo de adquisición cuando se realiza presión y cuando no se realiza. Caracterizan la escritura empleando la presión y las características cinemáticas de cada una de las pruebas. Las características fundamentales extraídas de la presión son la velocidad con la que la presión cambia con respecto al tiempo, el valor de presión medio durante la realización de la prueba, el Número de Cambios en la Presión (NPC) y el NPC relativo. Como características cinemáticas se encuentran velocidad del

trazo, aceleración, sacudidas, descomposición de la velocidad en componente horizontal y vertical, y el número de cambios en la dirección de la velocidad, entre otras. Además obtienen 6 coeficientes estadísticos (media, mediana, desviación estándar, 1^{er} percentil, percentil 99 y percentil 99 - 1^{er} percentil) para captar la relación entre las características de presión y las cinemáticas. Para la construcción del clasificador binario realizaron pruebas mediante SVM (Máquina Vector Soporte), clasificador AdaBoost con un estimador de árbol de decisión y el algoritmo KNN (K-Nearest Neighbors). Las características empleadas para la clasificación reportaron diferencias estadísticas significativas ($p < 0.05$) entre enfermos y sanos. El valor de p se define como la probabilidad de obtener un resultado al menos tan extremo como el que se ha obtenido, suponiendo que la hipótesis nula es cierta. Los resultados no mostraron diferencias significativas entre las tareas de la 1 a la 4 (espiral, “l”, “le” y “les”, respectivamente), la prueba 6 (“porovnat”) fue la que obtuvo máxima precisión de clasificación, $Pacc = 74.2\%$ (Probabilistic Accuracy), empleando características de presión, mientras que empleando las características cinemáticas obtuvieron una precisión máxima de 74.9% en la tarea 8 (frase). Por otro lado la fusión de ambas características no mejoró los resultados de clasificación, pero la fusión de todas las tareas mejora notablemente la precisión del clasificador, lo que se puede deber a la distinta naturaleza de todas las pruebas. El mejor resultado de precisión fue de 87.4% que se obtuvo con la fusión de todas las pruebas y el empleo del clasificador SVM, el cual reportó mejores resultados en términos de precisión, especificación y sensibilidad en comparación con AdaBoost y KNN.

En este estudio [11] realizado por Drotár se emplean marcadores de escritura basados en la entropía, la energía y las medidas extrínsecas de la escritura a mano, sobre la base de datos PaHaW. Mediante las coordenadas cartesianas y el sello temporal obtienen diversas características cinemáticas y espacio temporales. Dentro de las características cinemáticas se emplea la velocidad y todas sus variantes, así como aceleración o sacudidas. Como características espacio temporales se obtienen anchura y longitud del trazo, duración de la escritura, longitud de la escritura y ratio entre movimientos en el aire y sobre la superficie. Además emplean las características de presión comentadas con anterioridad, como cambios de presión respecto al tiempo, NPC (Número de Cambios en la Presión) y NPC relativo, junto con 6 coeficientes de correlación entre presión y velocidad/aceleración (vertical/horizontal). Para modelar la complejidad de la escritura manuscrita, proponen características de escritura no estándar propuestas en [12], como son la entropía y la entropía de Shannon, además de características basadas en la energía y en la cantidad de ruido en relación con la señal de escritura útil. Por último obtienen características intrínsecas, junto con las 6 funciones estadísticas descritas con anterioridad. Al igual que en [10], emplean sólo las características que presentan diferencias estadísticas significativas. Para la clasificación optaron por emplear SVM, obteniéndose el mejor resultado de clasificación

del 89.09% empleando características en la superficie de escritura, y combinando todas las tareas.

1.1. Motivación

El Parkinson es una enfermedad con elevada incidencia en la sociedad actual, aun así no existen pruebas médicas para detectar el estado real del paciente. Por tanto es muy importante desarrollar nuevas técnicas que permitan el estudio de la enfermedad, para mejorar la detección temprana y el seguimiento de la misma.

En el presente estudio se va a emplear la escritura como biomarcador para caracterizar este tipo de trastornos neurodegenerativos, mediante su información biométrica. Cuando se posea un estudio completo de la escritura como biomarcador, se podrán crear herramientas de diagnóstico para este tipo de trastornos. Éstas consistirán en pruebas de escritura simples realizadas en un dispositivo táctil que permita capturar las coordenadas del bolígrafo, los datos de presión y las muestras de tiempo, para que un médico pueda estudiarlas, hacer un diagnóstico temprano de la enfermedad o monitorizar la evolución del paciente.

Alguna de las ventajas que supone el empleo de las pruebas de escritura para la caracterización del trastorno son:

- Realización de pruebas sencillas con las que los pacientes están habituados.
 - Detección temprana de la enfermedad debido a que es sabido que los primeros síntomas se revelan en forma de micro-cambios.
 - Seguimiento o monitorización de la enfermedad, puesto que se obtiene mayor cantidad de información disponible para el personal médico que realiza un seguimiento más constante.
 - Mayor comodidad para los pacientes y cuidadores, los cuales no tienen que asistir al hospital para la realización de las pruebas de seguimiento, tan sólo es necesario un dispositivo que recoja la presión realizada en el trazo y una aplicación software, para la realización de las pruebas y el posterior envío al médico asignado.
 - Reduce los gastos de sanidad, facilitando la labor del personal médico.
 - Sirve para analizar la respuesta individual a determinados tratamientos.
-

Todo esto puede mejorar la calidad de vida de los pacientes a través de tratamientos más personalizados, además de reducir los gastos en sanidad asociados a dicho trastorno.

1.2. Objetivos y metodología

El objetivo principal de este TFM es:

- Analizar el potencial de las características biométricas asociadas a la escritura como biomarcador de la enfermedad de Parkinson. Para ello se van a emplear una serie de pruebas simples de escritura habituales en los estudios relacionados con este trastorno. De cada una de las pruebas se va a analizar el rendimiento de las características biométricas basadas en firma manuscrita, las cuales fueron propuestas en [13]. Además para el reconocimiento de identidad se estudiará un nuevo conjunto de características basadas en el modelado neuromotor del usuario [14].

Asociado a este objetivo, existen una serie de objetivos secundarios que ayudarán a la consecución del primero:

- Analizar el potencial de las diferentes pruebas mediante el pre-procesado de la información. Uno de los ejercicios más populares para detectar temblor, y por tanto esta enfermedad, es la espiral de Arquímedes. Otro ejercicio habitual es la escritura de palabras que contengan la letra “l”. Con dichos ejercicios se consigue evaluar las habilidades neuromotoras de los sujetos [10].
- Análisis de nuevas características relacionadas con el modelado neuromotor de la escritura, junto con las características desarrolladas en [13]. Para ello se van a combinar las características entre sí, además de realizar un algoritmo de clustering para la agrupación de sujetos.
- Desarrollo de un clasificador biclase que permita detectar usuarios con trazas de la enfermedad. Para la realización del mismo se emplean las muestras de la base de datos (PaHaW [10]) con personas diagnosticadas con PD, las cuales tienen distintos grados de enfermedad, y con personas de control que no poseen la enfermedad.

En el caso del presente estudio se va a emplear la base de datos PaHaW (Parkinson’s Disease Handwriting Database), al igual que los estudios anteriores. Las características globales de escritura empleadas en este estudio son las definidas en [13], como duración total de la prueba, velocidad/aceleración (vertical/horizontal), sacudidas/temblor (eje x/

eje y), entre otras. Por otro lado se van a obtener características neuromotoras de la escritura empleando el modelo Sigma-Lognormal, de las cuales se van a calcular una serie de parámetros estadísticos (media, mediana, desviación estándar, 1er percentil, percentil 99 y percentil 99 - 1^{er} percentil). Mediante el algoritmo KNN se realizará una agrupación de sujetos para conocer mejor la base de datos, y las características discriminantes para detectar la presencia o ausencia de enfermedad. Para la clasificación se va a emplear SVM, puesto que según el estudio [10] es el que mejor resultados produce en términos de exactitud.

1.3. Organización de la memoria

La memoria del presente trabajo se organiza en los siguientes 5 capítulos:

- Capítulo 1: Introducción. En este capítulo se da a conocer el trastorno, la tasa de incidencia en la sociedad actual y algunos aspectos clave. Además se incluye una revisión bibliográfica de este tipo de estudios, los objetivos del proyecto y la motivación del mismo.
 - Capítulo 2: Características biométricas como biomarcador de apoyo al diagnóstico de Parkinson. En este capítulo se presenta la base de datos empleada para el estudio, y todas las características empleadas, ya sean globales como neuromotoras. De cada tipo de pruebas se muestran algunas distribuciones relevantes de los datos.
 - Capítulo 3: Experimentos. En este capítulo se pueden observar la combinación de características tanto globales como neuromotoras y una agrupación de sujetos en distintos grupos. Además en el tercer apartado se ha definido el método de clasificación empleado.
 - Capítulo 4: Resultados. En este capítulo se muestran los resultados obtenidos de las distintas pruebas de clasificación.
 - Capítulo 5: Conclusiones. En este capítulo se exponen tanto las conclusiones como las líneas futuras del proyecto.
-

Características biométricas de la escritura como biomarcador de apoyo al diagnóstico de Parkinson

2.1. Base de datos

La base de datos empleada para este trabajo es PaHaW [10], adquirida para desarrollar modelos predictivos para el diagnóstico del PD. Fue adquirida por el Departamento de Telecomunicaciones de la Brno University of Technology, en colaboración del First Department of Neurology, Masaryk University y St. Anne's University Hospital in Brno, Czech Republic.

La población bajo estudio es de un total de 75 sujetos, de ellos 37 son personas enfermas (18 hombres y 19 mujeres) y 38 personas sanas (20 hombres y 18 mujeres). La edad mínima es de 36 años, la máxima de 92 años y la media de 65 ± 11.6 años. Los sujetos sanos, o de control, tienen una media de edad de 61 ± 11.4 años, mientras que los sujetos enfermos tienen una media más elevada, de 68 ± 11 años. Todos los sujetos que participaron en la construcción de la base de datos son diestros, y reportaron como lengua materna el checo, además de poseer al menos 10 años de educación. Para la realización de la base de datos se tuvieron que asegurar de que ningún sujeto presentase síntomas psiquiátricos o alguna enfermedad que afectase al sistema nervioso central, además los sujetos de control tuvieron que pasar un examen médico que descartaba cualquier tipo de enfermedad neurológica, o desorden del movimiento que pudiese interferir en los resultados de escritura. Por otro lado todos los enfermos están tomando tratamiento de levodopa para reducir los efectos del trastorno, lo que se traduce en resultados más optimistas.

La figura 2.1 se corresponde con la distribución de edades entre los sujetos, en la

misma se observa que los pacientes enfermos tienen una mayor edad que los sujetos de control. En la tabla 2.1 se muestra un resumen de información disponible de la población estadística, mientras que la información detallada se puede encontrar en el anexo II.

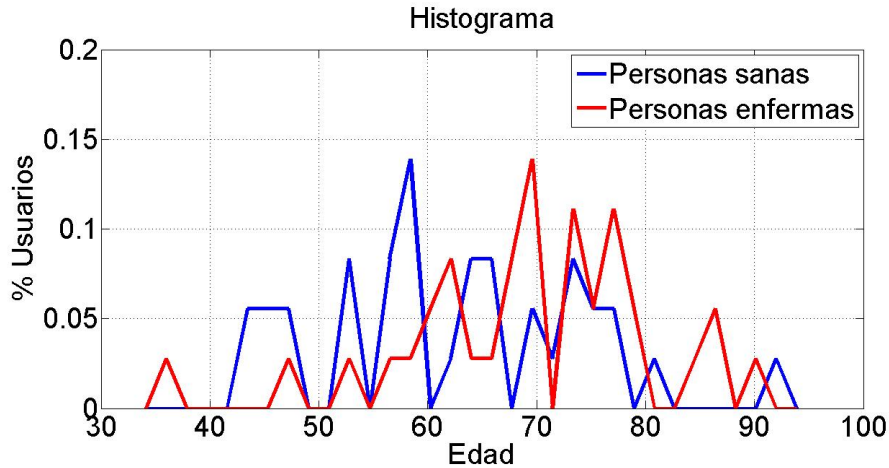


Figura 2.1: Distribución de edades de los sujetos participantes en la construcción de la base de datos.

Tabla 2.1: Resumen de la información de PaHaW.

	Enfermos	Sanos
Promedio de edad	68.4 ± 11	61.4 ± 11.4
Promedio de medicación	1266.5 ± 704.8	-
Promedio de años que llevan enfermos	7 ± 4.8	-
Promedio de escala UPDRS II	2.1 ± 0.8	-

El dispositivo empleado para la adquisición de las muestras de la base de datos era una Tablet digitalizadora Wacom (M Intuos 4), con una frecuencia de muestreo de 200 Hz. En la pantalla del dispositivo se mostraba una plantilla de dibujo compuesta por un cuadrado para el dibujo de la espiral y unas líneas para la realización de las pruebas restantes, como la figura 2.2 pero sin el patrón de letras. Para que el sujeto pudiese visualizar el dibujo realizado se empleó una pluma de tinta especial, lo que permitía una visualización simultánea en la pantalla del dispositivo, como si se tratase de un documento en papel. La frecuencia de adquisición era de 100 muestras/segundo y el sistema era capaz de capturar todos los movimientos del lápiz, cuando se ejercía presión sobre la superficie y cuando estaba el bolígrafo o pluma en el aire.

Los datos obtenidos de las muestras de la base de datos son:

- Coordenadas X.
- Coordenadas Y.

- Sello de tiempo.
- Estado ON/OFF (0 si el bolígrafo estaba en el aire y 1 si el mismo estaba en la superficie).
- Azimut.
- Altitud.
- Precisión.

Con los datos se pueden obtener los biomarcadores de la escritura que caracterizan la enfermedad, para ello se van a emplear unas características globales de escritura y unas características neuromotoras, que serán detalladas a continuación.

Un problema del empleo de la base de datos PaHaW es que los enfermos están medicándose para reducir los efectos del trastorno, lo que se traduce en menor cantidad de síntomas aparentes para detectar la enfermedad. Por tanto aparecen sujetos enfermos que no presentan síntomas a simple vista, y personas sanas que presentan muchos temblores, lo que puede ser debido a la edad y el deterioro neuromotor asociado al envejecimiento.

2.1.1. Pruebas

La base de datos PaHaW cuenta con 8 pruebas de escritura simples realizadas por cada uno de los 75 sujetos. Un ejemplo de las distintas pruebas se puede ver en la figura 2.2. Destacar que los sujetos pudieron practicar antes de realizar la prueba definitiva.

Una de las pruebas más populares empleadas para detectar temblor, y por tanto esta enfermedad, es la espiral de Arquímedes. Otro ejercicio habitual es la escritura de palabras que contengan la letra “l” de forma repetida. Ambas pruebas, cuando se trate de una persona sana puede realizarlos en un mismo trazo, sin necesidad de levantar el lápiz, pero cuando aparecen los problemas neuromotores se suelen apreciar temblores en el trazo, discontinuidades y rectificaciones en la escritura.

Se realizan 8 pruebas distintas, la primera consiste en la espiral de Arquímedes, la cual se ha realizado desde dentro a fuera, pero sin ninguna otra restricción. Las pruebas 2, 3 y 4 es la escritura repetida de las letras “l”, “le” y “les”, respectivamente. La prueba 5 consiste en escribir dos veces “lektorka” que en español significa “profesora”, la prueba 6 es la palabra “porovnat” que traducida al español es “comparar” y la palabra de la prueba 7 es “nepopadnout” que en castellano es “no coger”. Por último la prueba 8 consiste en escribir la frase “Tramvaj dnes u^z nepojede” cuyo significado es: “el tranvía no va a ir

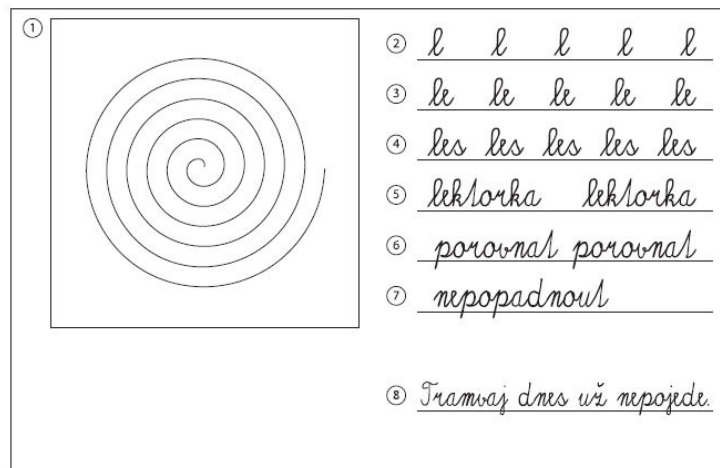


Figura 2.2: Ejemplo de las 8 pruebas realizadas en la base de datos PaHaw.

hoy”. Con una frase se puede obtener más información como el tiempo de transición entre palabras. Todas las pruebas han sido diseñadas para aportar información del trastorno además de ser sencillas.

Se ha elegido 3 participantes distintos para comparar visualmente las distintas pruebas. El primer sujeto elegido es el número 7, es una mujer de 79 años que lleva enferma desde hace 19 años y tiene un grado 3 en la escala UPDRS III. El segundo sujeto es el número 74, un varón de 53 años que lleva enfermo 9 años y por último el sujeto 67, un paciente de control de 59 años.

Cabe destacar que las figuras representadas a continuación muestran la posición “x” e “y” en cada instante de tiempo, sin tener en cuenta si el bolígrafo está apoyado o levantado de la superficie de adquisición.

Prueba 1:

A simple vista observando la figura 2.3, no se ven diferencias claras en la forma de la espiral entre el sujeto sano y el número 74, mientras que en la persona que lleva más tiempo enferma se puede observar el temblor en las líneas, además de cambios de dirección.

Por otro lado los vectores de velocidad tanto horizontal como vertical son muy diferentes para los sujetos comparados. Para una persona sana, la velocidad se puede descomponer como una suma de lognormales, tal y como se observa en la figura del sujeto 67, mientras que en los enfermos esa distribución no se aprecia de manera tan clara.

En estos tres casos se observa que el tiempo que tarda el sujeto en realizar la prueba podría ser un buen indicativo de enfermedad, puesto que la persona sana tarda menos

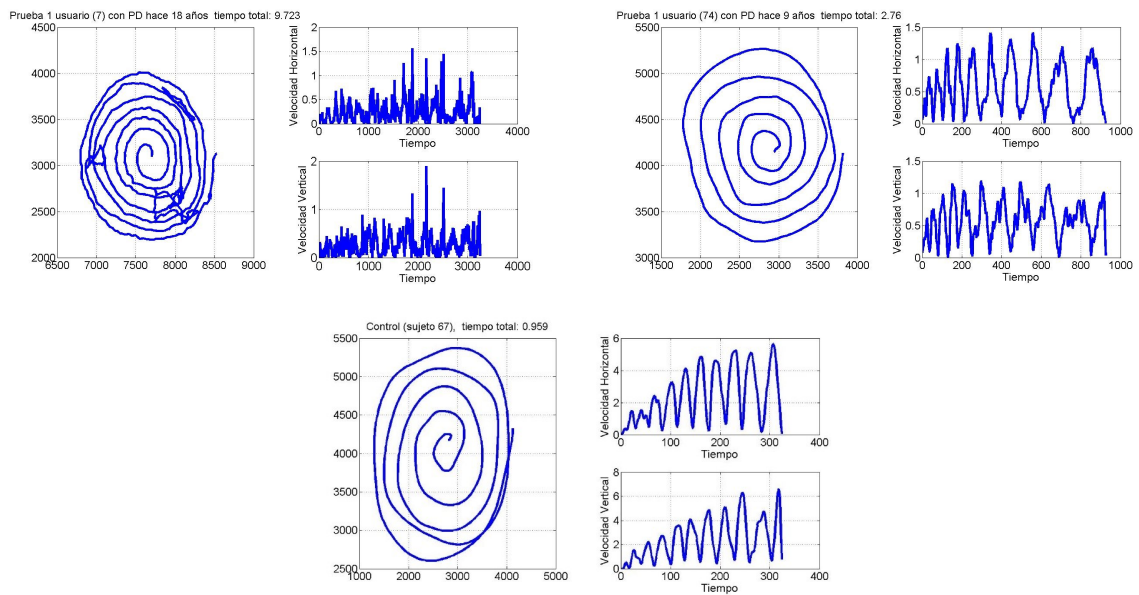


Figura 2.3: Representación de la prueba 1, a tres sujetos distintos (2 enfermos y uno sano). La figura muestra la representación gráfica de la prueba y los vectores de velocidad tanto horizontal como vertical.

de 1 segundo en realizar la espiral y los enfermos más tiempo, dependiendo del grado de enfermedad.

Prueba 2:

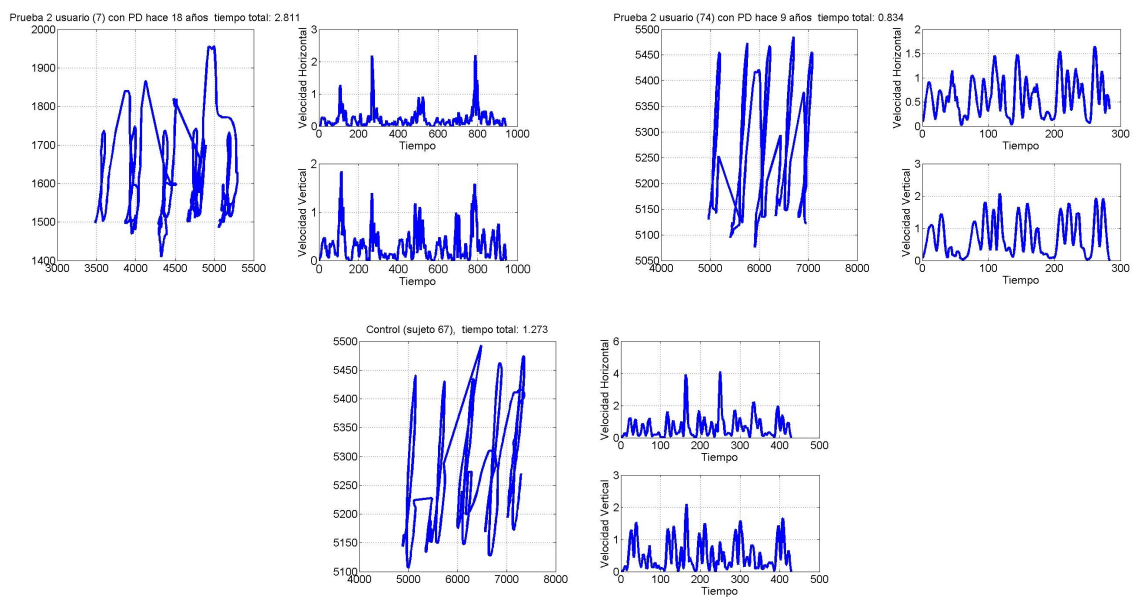


Figura 2.4: Representación de la prueba 2 (escritura de la letra “l”), a tres sujetos distintos (2 enfermos y uno sano). La figura muestra la representación gráfica de la prueba y los vectores de velocidad tanto horizontal como vertical.

Observando la figura 2.4 nos damos cuenta que la segunda prueba (escritura de la letra “l”) es peor que la primera, para detectar pacientes con problemas neurodegenerativos, porque no se puede distinguir de manera tan clara un paciente sano de uno enfermo, como sucedía en la prueba anterior.

En esta prueba el tiempo empleado para la realización de la misma no es una característica discriminante, puesto que el sujeto 74 tarda menos tiempo en realizarla que el de control. En la gráfica realizada por el enfermo 7, se observa que la realización de una prueba larga le supone un esfuerzo lo que se traduce en un deterioro de la caligrafía.

Pruebas 3, 4, 5, 6 y 7:

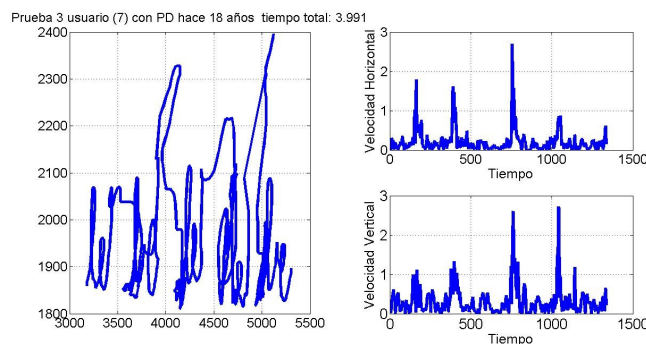


Figura 2.5: Representación de la prueba 3 (“le”), al sujeto 7 que lleva 18 años enfermo de PD. La figura muestra la representación gráfica de la prueba y los vectores de velocidad tanto horizontal como vertical. El paciente muestra fatiga muscular en la realización de la prueba.

Las conclusiones que se pueden sacar de estas pruebas son similares a las obtenidas para la prueba número 2, lo único que en la figura 2.5 del paciente 7 se observa de manera más clara su dificultad para terminar la prueba. Mientras que las diferencias entre los pacientes 74 y 67 no son significativas.

Prueba 8:

La prueba 8 consiste en escribir una frase y debería ser la prueba que mayor cantidad de información nos aporte, pero mediante el parámetro de distancia media se ha descubierto que no es la mejor prueba si se emplean características globales.

Las pocas diferencias encontradas, en la figura 2.6, entre el paciente que lleva 9 años enfermo y el sano se pueden deber al tratamiento médico, que controla los temblores y alivia los síntomas de la enfermedad.

Se ha observado que el sujeto 7 tiene problemas de visión espacial, puesto que no sigue las líneas de la plantilla de respuesta. Como se puede ver en la prueba 2, la línea de respuesta está en el eje y en la posición 5150 y el sujeto ha escrito su prueba en la

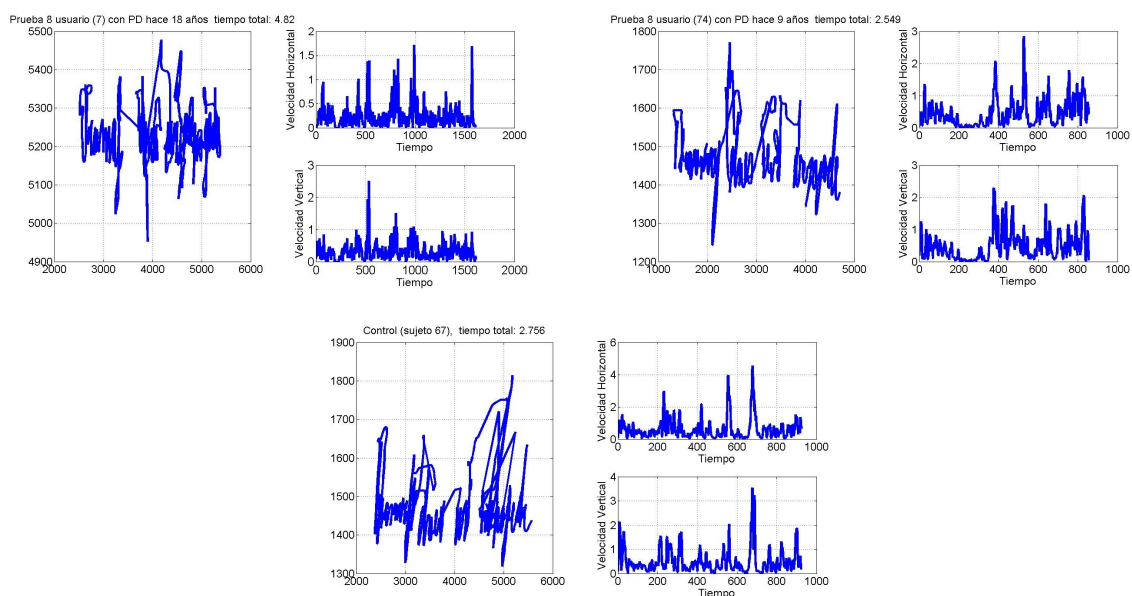


Figura 2.6: Representación de la prueba 8 (frase), a tres sujetos distintos (2 enfermos y uno sano). La figura muestra la representación gráfica de la prueba y los vectores de velocidad tanto horizontal como vertical.

posición 1500, al igual que en las demás pruebas donde tampoco ha seguido la plantilla.

2.2. Análisis de características globales

El primer análisis realizado en el presente proyecto sobre las pruebas de escritura se basa en el conjunto de características propuestas en [13]. Estas características fueron desarrolladas para el reconocimiento biométrico de la identidad asociada a la firma manuscrita. Dada una secuencia temporal extraída de una muestra de escritura, se obtienen 100 características globales.

Las características globales se pueden dividir en 4 categorías:

- Tiempo: relativo a la duración de la prueba, o eventos como los máximos locales en los que el bolígrafo se encuentra levantado.
- Velocidad y aceleración: se obtienen de las derivadas de primer y segundo orden de la posición en función del tiempo.
- Dirección: extraídas de la trayectoria.
- Geometría: asociada con los trazos y con la relación de aspecto.

Las primeras 18 características globales extraídas de las pruebas de escritura son las mostradas en la tabla 2.2. De las mismas se hará una selección de las mejores, en términos de distancia, eliminando las que sean nulas o las que no aporten información relevante.

Tabla 2.2: Selección de las 18 primeras características globales aplicadas a las pruebas de escritura, véase [13] para tabla completa.

Característica	Descripción	Característica	Descripción
1	Duración total T_s	2	$\frac{(\text{duración pen down } T_w)}{T_s}$
3	$\frac{(1st\ t(v_{max}))}{T_w}$	4	$\frac{T(v_x > 0)}{T_w}$
5	$\frac{T(v_x < 0)}{T_w}$	6	$\frac{T(v_y > 0)}{T_w}$
7	$\frac{T(v_y < 0)}{T_w}$	8	$\frac{T(v_x > 0 pen\ up)}{T_w}$
9	$\frac{T(v_x < 0 pen\ up)}{T_w}$	10	$\frac{T(v_y > 0 pen\ up)}{T_w}$
11	$\frac{T(v_x < y pen\ up)}{T_w}$	12	$\frac{T(1st\ pen\ up)}{T_w}$
13	$\frac{T(2nd\ pen\ up)}{T_w}$	14	$\frac{T(2nd\ pen\ down)}{T_s}$
15	$\frac{T(3rd\ pen\ down)}{T_s}$	16	$\frac{(1st\ t(v_{y,max}))}{T_w}$
17	$\frac{(1st\ t(v_{y,min}))}{T_w}$	18	$\frac{(1st\ t(v_{x,max}))}{T_w}$

Para cuantificar la calidad de las distintas pruebas que componen la base de datos y las características que mejor describen el trastorno, se ha calculado una medida de distancia entre distribuciones de usuarios con trastorno y usuarios control. Este valor de distancia sirve para cuantificar de manera numérica la separación existente entre las distribuciones de las personas enfermas y de control. Dicho parámetro se calcula mediante la siguiente fórmula 2.1:

$$D = \frac{|\bar{f}_{control} - \bar{f}_{enfermos}|}{\sigma_{control} + \sigma_{enfermos}} \quad (2.1)$$

Donde $\bar{f}_{control}$ es la media de las muestras obtenidas para una prueba y una característica, de todas las personas sanas que han participado el experimento, $\bar{f}_{enfermos}$ es la media pero en este caso de los sujetos enfermos. El valor $\sigma_{control}$ es la desviación estándar de las muestras empleadas anteriormente para los sujetos de control y $\sigma_{enfermos}$ para los sujetos que están diagnosticados con el trastorno. Al aplicar la fórmula 2.1 a cada característica y a cada prueba distinta se ha obtenido la tabla 2.2 de distancias mostrada en el anexo, en la misma se ha señalado en rojo las 5 características de cada prueba que

tenían mayor distancia. En la tabla 2.3 se ha realizado un resumen de la tabla general de distancias, donde la segunda columna calcula la distancia media de las características destacadas, las 5 características que tiene mayor distancia, de cada una de las pruebas y en la tercera columna se calcula la distancia media de todas las características por prueba. En ambos casos la prueba que ha reportado mayor distancia es la número 1, la espiral de Arquímedes, por tanto al emplear las características globales la prueba 1 es idónea para poder clasificar entre personas sanas y enfermas.

Tabla 2.3: Resumen del parámetro distancia calculado para todas las pruebas y todas las características globales. La columna “todas” se ha calculado como la media de todas las distancias obtenidas de las 100 características para cada una de las pruebas y la columna “destacadas” se ha calculado como la media de la distancia de las 5 características con mayor valor. La prueba con mayor valor de distancia es la espiral (prueba número 1).

	Características	
	Todas	Destacadas
Prueba 1	0,115	0,384
Prueba 2	0,08	0,315
Prueba 3	0,1	0,314
Prueba 4	0,074	0,318
Prueba 5	0,084	0,270
Prueba 6	0,072	0,275
Prueba 7	0,071	0,275
Prueba 8	0,095	0,308

A continuación se va a mostrar algunas distribuciones de cada una de las pruebas.

Prueba 1 (espiral): se han obtenido las distribuciones de las características destacadas, en el caso de la prueba 1 las 5 mejores son: 22 ($\frac{1st\ t(x_{max})}{T_w}$), 23 ($\frac{2nd\ t(x_{max})}{T_w}$), 42 (promedio de las sacudidas), 95 (N(máximos locales x)) y 98 (N(máximos locales y)).

En las distribuciones de la figura 2.7 se observa que por lo general hay poca separación entre clases. La distribución de sujetos sanos parece estar más concentrada, mientras que la de enfermos tiene más variabilidad. Como cabía esperar, los sujetos de control producen menor número de máximos locales tanto en el eje x como en el eje y, al igual que sucede con las sacudidas, donde los sujetos sanos tienen las muestras distribuidas en torno a 0.1 aproximadamente y los enfermos un espectro más ancho y con valores más elevados. La característica 1 (duración total) no es una característica muy discriminante. Aunque pueda parecer que los enfermos suelen tardar más tiempo en la realización de una prueba, no sucede así, en concreto se encuentra en la posición 28 de distancia respecto del total de características. Además observando la característica 1 no se aprecian muchas diferencias entre los distintos grupos.

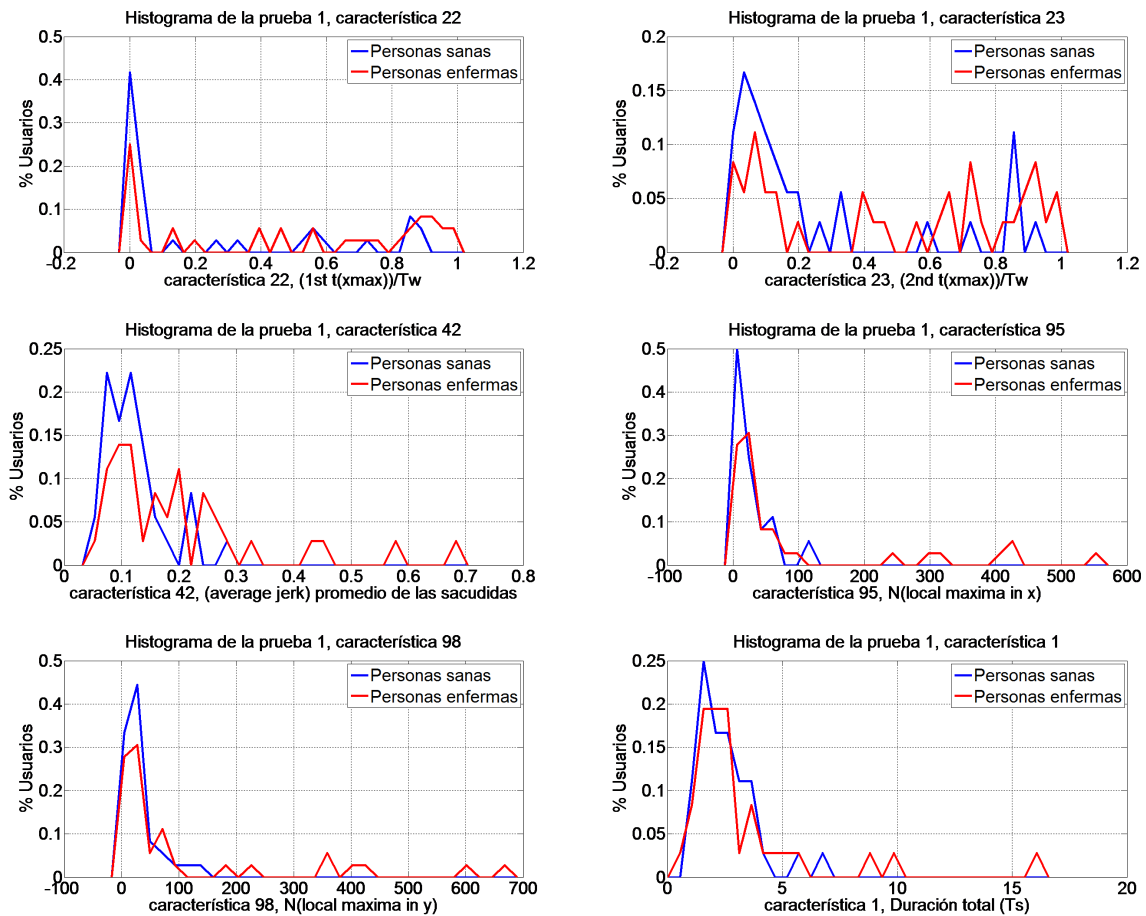


Figura 2.7: Características globales destacadas de la prueba 1 y característica 1 (duración total).

La prueba 2 se corresponde con la escritura repetida de la letra “i”, donde las 5 características con mayor valor de distancia se relacionan con la velocidad y la aceleración, estas son: 36 (integrated abs pentri acc. alc)/amax, 37 ($\frac{\text{correlación velocidad } v_{x,y}}{v^2_{max}}$), 38 (desviación estándar de vx), 45 (jerk máximo) y 93 (histograma espacial t3). Las características 36 y 37 muestran distribuciones solapadas, aunque los enfermos tienden a valores más elevados. Es curioso observar que los sujetos sanos en esta prueba producen menor cantidad de sacudidas que los enfermos. En el caso de la característica 1, se pueden ver más separados las distribuciones en comparación con la prueba anterior, aunque no sería de las mejores características para distinguir entre los grupos. En la figura 2.8 se muestran algunas distribuciones de dicha prueba.

La prueba 4 se corresponde con la escritura de la sílaba “les” de manera repetida. Las características destacadas de la misma son: 5 ($\frac{T(v_x < 0)}{T_w}$), 20 ($T((dy/dt)/(dx/dt) > 0) / T((dy/dt)/(dx/dt) < 0)$), 64 (dirección histograma s3), 91 (histograma espacial t1) y 94 (histograma espacial t4), las mismas relacionadas con la velocidad y los histogramas es-

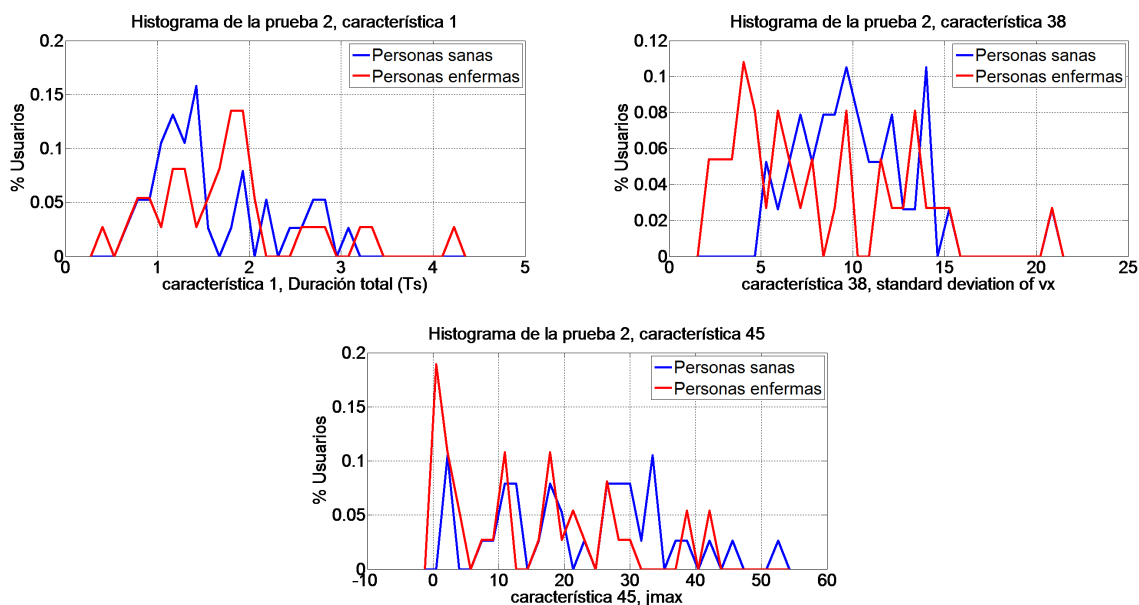


Figura 2.8: Distribuciones obtenidas de la prueba 2 (escritura letra “l”), mediante características globales.

paciales. Destacar que no aparecen características relacionadas con las sacudidas dentro de las destacadas, como sucedía en las pruebas anteriores. A simple vista, las mejores características para separar ambas distribuciones en esta prueba son las mostradas en la figura 2.9.

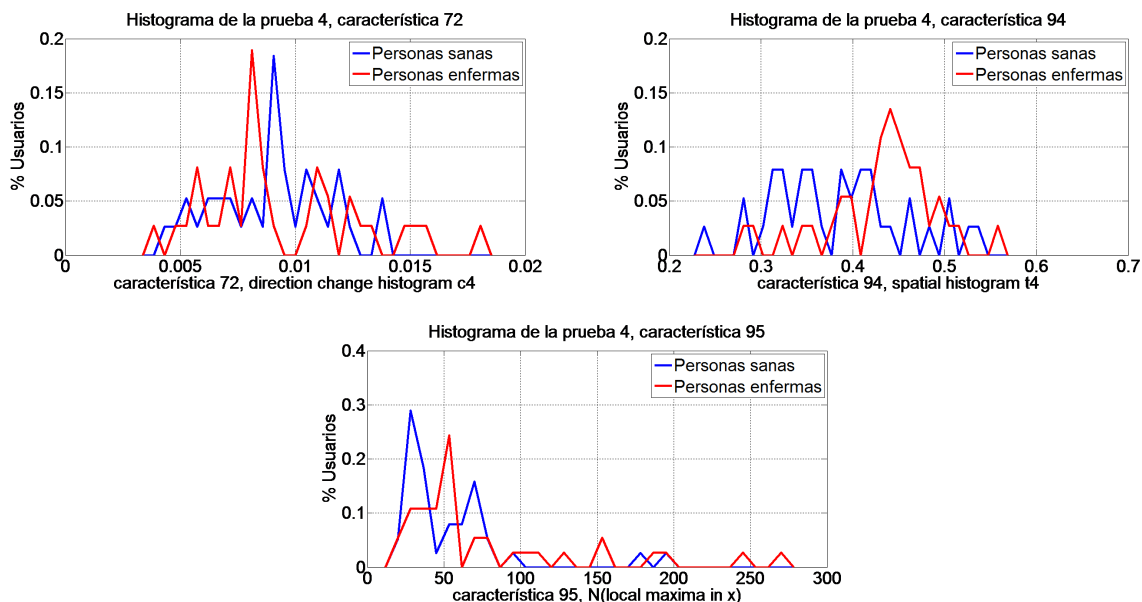


Figura 2.9: Distribuciones obtenidas de la prueba 4 (escritura sílaba “les”), mediante características globales.

Las características destacadas de las pruebas 5 y 6 no aportan mucha información

puesto que las distribuciones están muy solapadas. Mientras que las destacadas de la prueba 7 (escritura de la palabra “nepopadnout”) sí muestran diferencias entre los grupos, estas son: 20 ($(T((dy/dt)/(dx/dt)>0) / T((dy/dt)/(dx/dt)<0))$), 53 ($N(\text{cambios de signo de } dx/dt \text{ y } dy/dt)$), 54 ($(T((dx/dt)/(dy/dt)>0) / T((dx/dt)/(dy/dt)<0))$), 62 (dirección del histograma s1) y 95 ($N(\text{máximos locales en } x)$), relacionadas con el número de máximos locales y con la geometría. Los histogramas de la prueba 7 que muestran mayor separación entre las distribuciones son los mostrados en la figura 2.10.

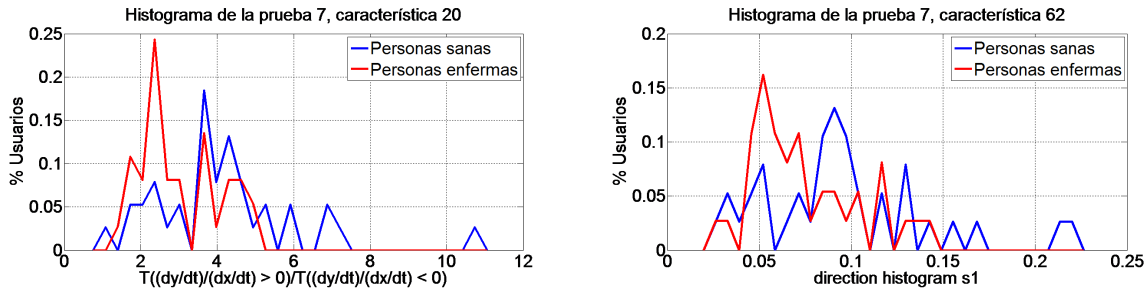


Figura 2.10: Distribuciones obtenidas de la prueba 7 (palabra “nepopadnout”), mediante características globales.

Las distribuciones destacadas, de la prueba 8, según el parámetro de distancia son los siguientes: 5 ($\frac{T(v_x < 0)}{T_w}$), 42 (promedio de las sacudidas), 62 (dirección del histograma s1) y 88 ($(x_{\max} - x_{\min}) / x_{\text{acquisitionrange}}$). Destacar que la normalización de la media de la posición x es la característica más idónea (número 90) para distinguir a los sujetos en esta prueba, mientras que en las anteriores pruebas esta característica está en casi todas por encima de la posición 75, esta diferencia de posición se puede deber a que la prueba 8 tiene una naturaleza totalmente distinta a las pruebas anteriores. Y la característica 88 relativa a la posición en el eje x es la segunda mejor característica de la prueba, las distribuciones se pueden observar en la figura 2.11. Que las dos mejores características estén relacionadas con la geometría se podría deber a que los sujetos enfermos suelen escribir más juntas las palabras casi sin espacio entre ellas, por tanto la normalización del eje x es distinta para ambos grupos y se aprecia más en la construcción de una frase.

En conclusión, las características destacadas que se repiten en distintas pruebas son la 5 ($\frac{T(v_x < 0)}{T_w}$), la 20 ($T((dy/dt)/(dx/dt)>0) / T((dy/dt)/(dx/dt)<0)$), la 38 (desviación estándar de v_x), la 62 (dirección del histograma s1) y la 94 (histograma espacial t4). Están relacionadas con la velocidad, la desviación estándar y los histogramas. Los enfermos producen valores de desviación estándar menor que los sujetos de control, aunque lógicamente se podría pensar que fuera al contrario.

Sin analizar los datos, se podría decir que la duración total en realizar la prueba sería una característica discriminante, pero se ha demostrado que la misma no lo es. En

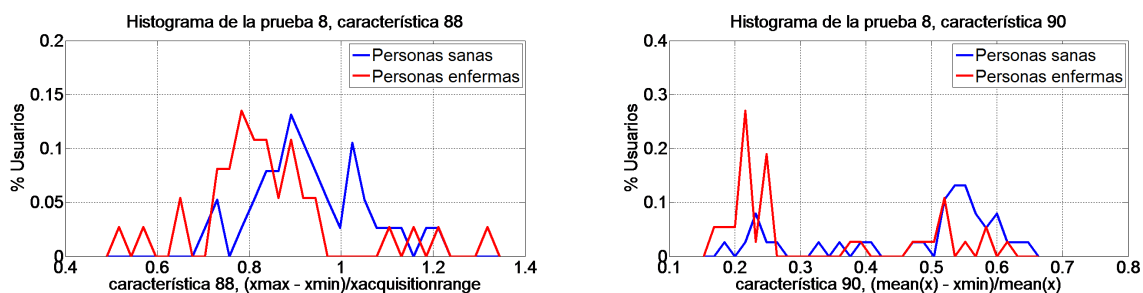


Figura 2.11: Distribuciones obtenidas de la prueba 8 (frase), mediante características globales.

la prueba 7 dicha característica se encuentra en la posición 15 del vector de distancias respecto al total, en las demás pruebas su posición es superior.

Una característica empleada para detectar la enfermedad es el temblor y tan sólo en las pruebas 1 (espiral), 2 (“l”), 3 (“les”) y 8 (frase) aparece algún tipo de jerk (sacudidas) como característica con mayor distancia entre las distribuciones de los grupos analizados.

En cada prueba las características destacadas se relacionan con alguna de las 4 categorías distintas comentadas con anterioridad, esto se podría deber a la naturaleza heterogénea de cada prueba. Debido a dicha naturaleza, la mejor característica de la prueba 8 se encuentra en posiciones muy bajas del rango de distancias en las demás pruebas. Como se ha comentado anteriormente, la geometría del eje x en la prueba 8 es relevante para distinguir ambos grupos, esto se podría deber a que los sujetos enfermos suelen escribir las palabras más juntas.

2.3. Análisis de características neuromotoras

En este proyecto se propone estudiar características neuromotoras derivadas de la escritura. En los últimos 20 años numerosos estudios ([15], [16], [17], [18], [19]) del control del movimiento han creado un modelado de los mismos acorde con la teoría cinemática de los movimientos humanos rápidos [20]. Esta teoría propone que los movimientos humanos se componen de una serie de comandos motores que se envían a través del sistema nervioso y cada comando se puede modelar como una contribución lognormal. Los movimientos se consideran el resultado sinérgico de dos sistemas neuromusculares, el primero el agonista que se encarga de iniciar el movimiento, y el segundo, el antagonista es el sistema opuesto que se encarga de parar el movimiento. Con la suma de esas contribuciones lognormales se reconstruye el perfil de velocidad de la escritura [5].

En este trabajo se va a emplear el modelo Sigma-Lognormal, que se emplea para

modelar movimientos arbitrarios complejos como puede ser el caso de palabras, una frase o movimientos oscilatorios en una dimensión. Se ha elegido este modelo puesto que se encarga de representar la información de los comandos motores y las propiedades temporales del sistema neuromuscular involucrado en la producción de movimientos complejos [21].

El modelo Sigma-Lognormal considera que la velocidad resultante de una acción del sistema neuromuscular se puede descomponer en lognormales escaladas por un parámetro de orden D y desplazadas en el tiempo por un orden t_0 [14]. El vector de velocidad de un trazo j está definido por la suma vectorial de lognormales, tal y como muestra la ecuación 2.2.

$$|\vec{v}_j(t; P_j)| = \frac{D_j}{\sigma(t - t_{0j})\sqrt{2\pi}} \exp\left\{\frac{[\ln(t - t_{0j}) - \mu_j]^2}{-2\sigma^2_j}\right\} \quad (2.2)$$

Todos los parámetros que componen el modelo son:

- D_j : longitud del trazo.
- t_{0j} : parámetro de desplazamiento de tiempo.
- μ_j : retrasos lógicos de tiempo, los delays de tiempo de los sistemas neuromusculares expresados en escala logarítmica de tiempo.
- σ_j : tiempo de respuesta lógica, son los tiempos de respuesta del sistema neuromuscular expresados en escala logarítmica de tiempo.
- θ_{sj} : ángulo inicial de la trayectoria.
- θ_{ej} : ángulo final de la trayectoria.

Por tanto cada trazo está descrito por los parámetros anteriores ($P_j = (t_{0j}, D_j, \mu_j, \sigma_j, \theta_{sj}, \theta_{ej})$). Con la suma vectorial de todas las lognormales, que componen cada trazo, desplazados en el tiempo se obtiene una aproximación de la escritura compleja, mediante la ecuación 2.3.

$$\vec{v}(t) = \sum_{j=1}^M \vec{v}_j(t) \quad (2.3)$$

Para la obtención de los parámetros Sigma-Lognormales que componen cada trazo se ha empleado el software ScriptStudio, se trata de un algoritmo robusto que calcula

dichos parámetros empleando el modelo descrito con anterioridad. Se encarga de reconstruir la trayectoria del vector de velocidad con alto valor de SNR y extraer los parámetros lognormal de los trazos que componen el vector de velocidad. La figura 2.12 muestra la pantalla de ScriptStudio, donde se ve la representación de la letra “l” en la parte inferior derecha, en la parte superior el vector de velocidad original y la señal reconstruida mediante el algoritmo, y en la parte izquierda el modelado de la señal con todos los parámetros correspondientes de cada lognormal.

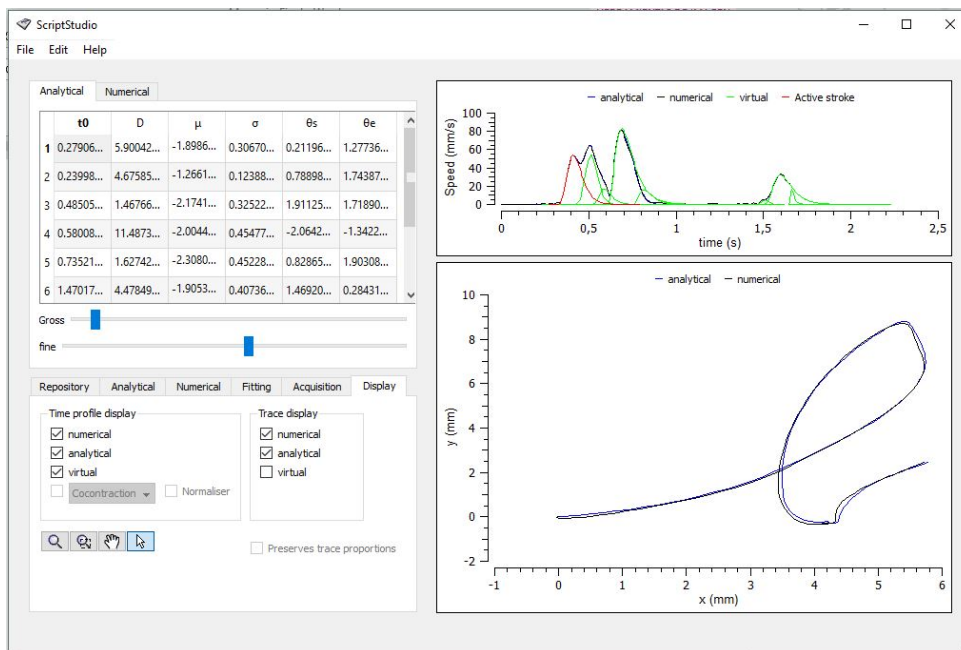


Figura 2.12: Interfaz de pantalla de ScriptStudio, con el ejemplo de la letra “l”.

Se va a analizar mejor el comportamiento de los pacientes mediante el análisis de la prueba 2, que consiste en la escritura repetida de la letra “l” minúscula. Se ha elegido dicha prueba porque reporta mayor distancia entre las distribuciones de ambos grupos al emplear las características neuromotoras, posteriormente se mostrará la tabla con el parámetro de distancia.

Tal y como se muestra en la figura 2.13, la letra “l” minúscula se suele componer de 3 trazos, donde el primer trazo está señalado con rojo, el segundo con verde y el último con azul. Por tanto el vector de velocidad estará compuesto como mínimo por 3 lognormales. Los espacios entre letras se suelen representar por una lognormal de mayor energía (señal con menor desviación y mayor amplitud, que caracteriza movimientos rápidos).

Según el grado de la enfermedad y el estado del paciente, estas señales se pueden ver afectadas. Para comprobar esto se van a comparar los resultados de pacientes sanos y enfermos. En las figuras posteriores, las flechas en rojo señalan la posición de los espacios entre letras y los rectángulos azules la formación de las distintas letras “l”.

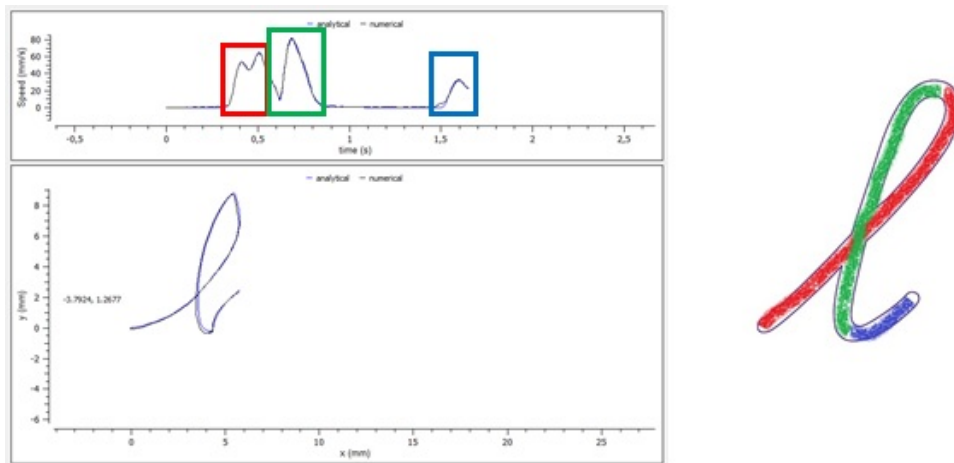
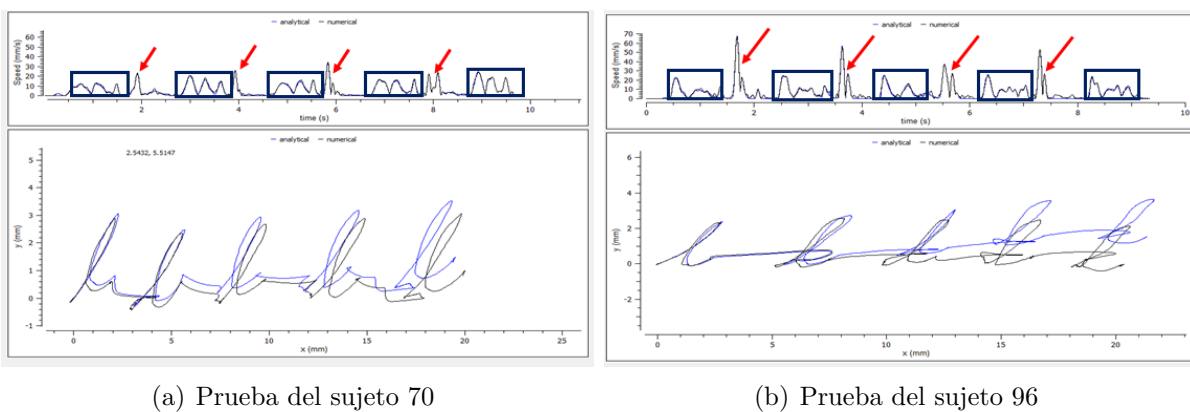


Figura 2.13: Letra “l”. La lognormal de color rojo se compone de otras dos con menor energía correspondientes a trazos distintos, la separación de la lognormal azul con la verde se puede deber a una pausa realizada por el sujeto antes de dibujar el último trazo.

Dentro de las personas sanas hemos analizado mediante ScriptStudio la segunda prueba de los sujetos número 70 y 96, la figura 2.14(a) representa la prueba realizada por una mujer sana de 47 años, donde su señal de velocidad está compuesta por 37 segmentos, y la figura 2.14(b) corresponde con una mujer de 77 años de edad, su vector de velocidad se ha descompuesto en 35 lognormales.



(a) Prueba del sujeto 70

(b) Prueba del sujeto 96

Figura 2.14: Representación de la prueba 2 para pacientes sanos, mediante ScriptStudio.

Los sujetos enfermos elegidos para la comparación son el número 3 y el 4. El número 3 es una mujer de 69 años de edad, que lleva 7 años enferma, su señal de velocidad está compuesta por 58 trazos, la figura 2.15(a) representa la prueba realizada por dicho sujeto. La figura 2.15(b) muestra la prueba del paciente número 4, el cual tiene 79 años de edad, lleva 12 años enfermo y su señal de velocidad se ha dividido en 73 lognormales.

En la figura 2.15(a) se puede observar una envolvente de la señal más abrupta, no es tan suavizada como en los casos anteriores. Estos picos se pueden deber a que la paciente

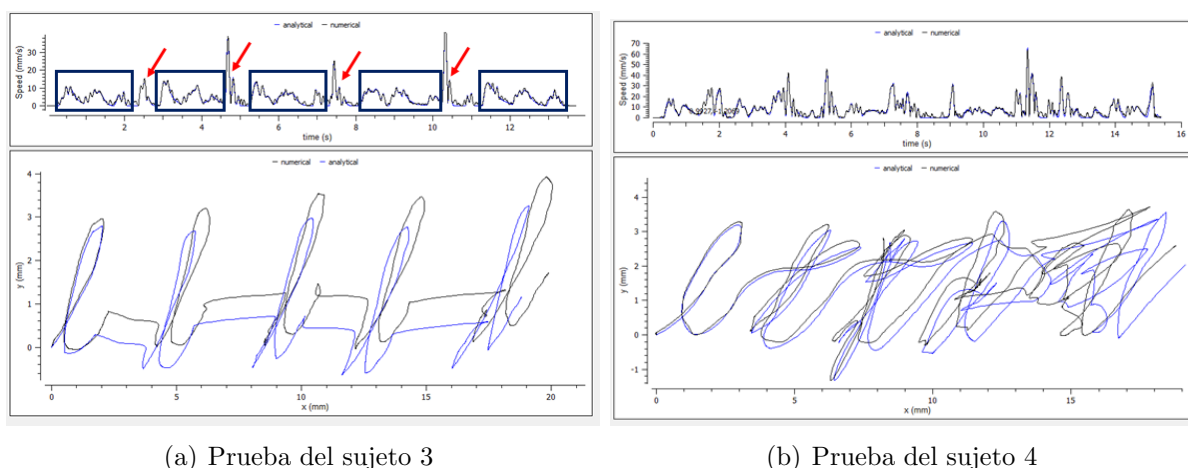


Figura 2.15: Representación de la prueba 2 para pacientes enfermos, mediante ScriptStudio.

ha realizado las letras con trazos pequeños y no de manera uniforme, una característica que comparten algunos enfermos.

A simple vista en la figura 2.15(b) se ve la dificultad que tiene el sujeto para la finalización de la prueba 2. No se puede distinguir ni la separación entre las letras ni los espacios, porque es una escritura con muchos trazos. Al principio de la prueba se observa la construcción de las letras perfectamente, pero conforme transcurre el tiempo de la prueba la escritura se vuelve más irregular, esto es debido a la fatiga muscular que sufren los enfermos de Parkinson.

En resumen, los pacientes enfermos suelen tener vectores de velocidad cuya envolvente no es tan suavizada como en los pacientes sanos, esto se puede deber a una mayor cantidad de trazos. Además en los sujetos sanos se suele distinguir perfectamente los trazos que componen las letras y los espacios en blanco, lo que no sucede en todos los pacientes enfermos. En algunos pacientes que sufren el trastorno se aprecia la fatiga muscular, puesto que al comienzo de la prueba se distinguen mejor las letras, pero conforme transcurre el tiempo de la misma la escritura va empeorando.

Como cabría esperar en los sujetos enfermos la señal de velocidad está compuesta por mayor cantidad de segmentos, puesto que no realizan la escritura empleando trazos continuos. Además al comparar los resultados obtenidos de ambos grupos bajo estudio, se observa que si la persona lleva enferma pocos años las diferencias son mínimas y conforme aumenta los años enfermos la señal se va degradando.

2.3.1. Caracterización neuromotora de los usuarios

Mediante el modelo neuromotor de cada prueba adquirido de ScriptStudio se obtienen 6 parámetros (t_0 , D , μ , σ , θ_s , θ_e) de cada lognormal que representan cada trazo. Con esos 6 parámetros se obtienen 18 características psicomotrices [22], las cuales se muestran en la tabla 2.4. Sin embargo las características neuromotoras empleadas para el estudio de los sujetos se obtienen de la combinación de los parámetros estadísticos (media, desviación estándar, 1^{er} percentil, percentil 99 y mediana) con las 18 características anteriores, de esta combinación se obtienen las 91 ($18 \times 5 + 1$) características neuromotoras mostradas en la tabla 2.5.

Tabla 2.4: Características psicomotrices cognitivas, obtenidas del modelo Sigma-Lognormal.

Características del trazo		Características de tiempo	
1	D_i	9	$v_2 = v_i(t_{2i}) $
2	μ_i	10	$v_3 = v_i(t_{3i}) $
3	σ_i	11	$v_4 = v_i(t_{4i}) $
4	$\sin(\theta_{si})$	12	$\delta t_{05} = t_{5i} - t_{0i-1}$
5	$\cos(\theta_{si})$	13	$\delta t_{15} = t_{5i} - t_{1i-1}$
6	$\sin(\theta_{ei})$	14	$\delta t_{13} = t_{3i} - t_{1i-1}$
7	$\cos(\theta_{ei})$	15	$\delta t_{35} = t_{5i} - t_{3i-1}$
8	$\Delta t_0 = t_{0i} - t_{0i-1}$	16	$\delta t_{24} = t_{4i} - t_{2i-1}$
		17	$\Delta t_1 = t_{1i} - t_{1i-1}$
		18	$\Delta t_3 = t_{3i} - t_{3i-1}$

Tabla 2.5: Características neuromotoras obtenidas de aplicar los parámetros estadísticos a las características psicomotrices cognitivas.

Características	Descripción
[1 - 18]	Media de las 18 características psicomotrices
[19 - 36]	Desviación estándar de las 18 características psicomotrices
[37 - 54]	Percentil 1 de las 18 características psicomotrices
[54 - 72]	Percentil 99 de las 18 características psicomotrices
[73 - 90]	Mediana de las 18 características psicomotrices
91	Número de lognormales que componen la señal de velocidad

Para cada una de las prueba y de las características neuromotoras se ha obtenido el parámetro de distancia, con el cual se ha completado la tabla 4.2, mostrada en el anexo. La tabla 2.6 muestra un resumen de la tabla completa de distancias. La prueba que ha reportado mayor distancia es la número 2 (“1”) por tanto al emplear las características neuromotoras dicha prueba es idónea para poder clasificar entre personas sanas y enfer-

mas. Comparando esta tabla con la obtenida de las características globales se observa que las pruebas suelen producir mayor distancia empleando las características neuromotoras.

Tabla 2.6: Resumen del parámetro distancia calculado para todas las pruebas y todas las características neuromotoras. La columna “todas” se ha calculado como la media de todas las distancias obtenidas de las 100 características para cada una de las pruebas y la columna “destacadas” se ha calculado como la media de la distancia de las 5 características con mayor valor. La prueba con mayor valor de distancia es la número 2 (escritura repetida de la letra “l”).

	Características	
	Todas	Destacadas
Prueba 1	0,07	0,292
Prueba 2	0,154	0,491
Prueba 3	0,11	0,409
Prueba 4	0,09	0,288
Prueba 5	0,0835	0,199
Prueba 6	0,1	0,316
Prueba 7	0,088	0,301
Prueba 8	0,114	0,48

A continuación se van a mostrar algunas distribuciones de cada una de las pruebas.

Las características destacadas de la prueba 1 (espiral) están relacionadas con la mediana, las cuales son: 2 ($\text{mean } \mu_i$), 74 ($\text{median } \mu_i$), 85 ($\text{median}(\delta t_{15} = t_{5i} - t_{1i-1})$), 87 ($\text{median}(\delta t_{35} = t_{5i} - t_{3i-1})$) y 88 ($\text{median}(\delta t_{24} = t_{4i} - t_{2i-1})$). En todos ellos las distribuciones de sujetos enfermos se concentran en torno a valores más reducidos en comparación con el grupo de control. Destacar que en esta prueba los histogramas obtenidos de los percentiles 1 y 99 no aportan mucha información puesto que las distribuciones en la mayoría de los casos están totalmente solapadas. Sin embargo, el número de lognormales, en el caso de la espiral si es una característica discriminante tal y como se observa en la figura 2.16, en concreto se encuentra en la posición 7 según el parámetro de distancia.

La prueba 2 (“l”) es la que mayor valor de distancia ha reportado entre las distribuciones. Las características destacadas son la número 1 ($\text{mean}(D_i)$), la 8 ($\text{mean}(\Delta t_0 = t_{0i} - t_{0i-1})$), la 9 ($\text{mean}(v_2 = |v_i(t_{2i})|)$), la 10 ($\text{mean}(v_3 = |v_i(t_{3i})|)$) y la 73 ($\text{median}(D_i)$), las mismas relacionadas con la media. La figura 2.17 muestra datos muy distribuidos, pero los enfermos tienen valores más reducidos.

La figura 2.18 muestra otras características que podrían ser empleadas para clasificar entre personas sanas y enfermas, como el número de lognormales, al igual que en el caso anterior, y la desviación estándar de la distancia (D_i).

En el caso de la prueba 3 (“e”) las características que mejor se comportan para

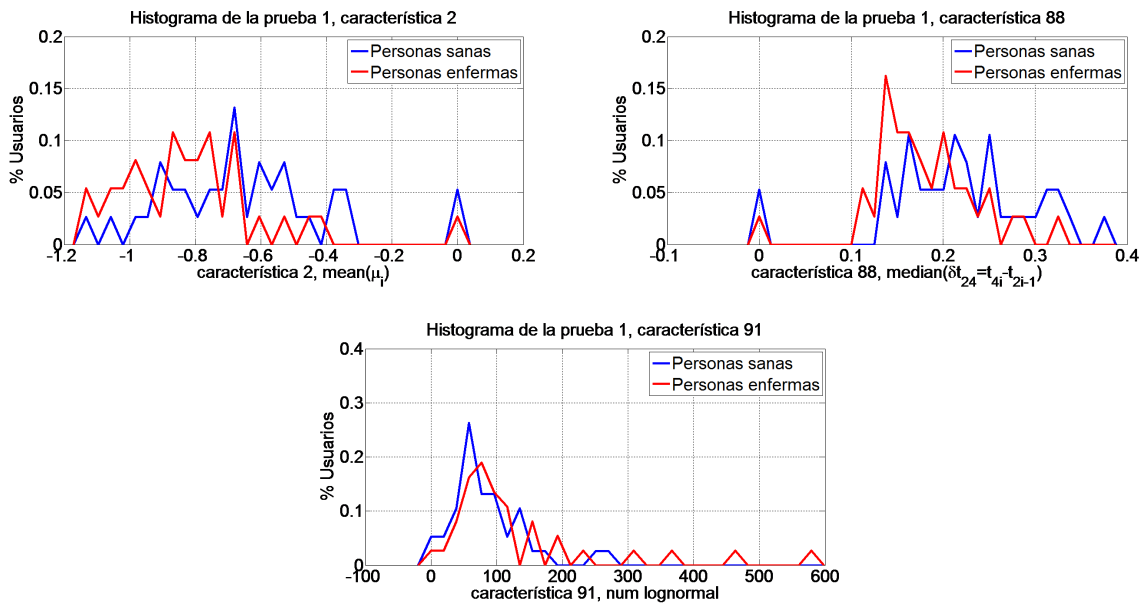


Figura 2.16: Distribuciones obtenidas de la prueba 1 (espiral), mediante características neuromotoras.

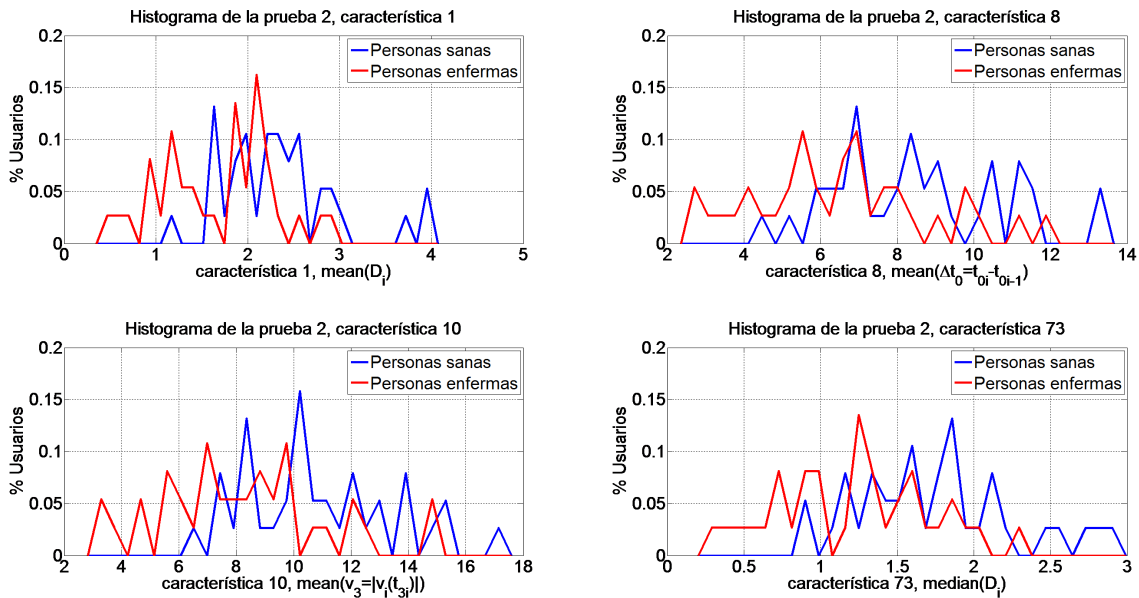


Figura 2.17: Características neuromotoras destacadas de la prueba 2 (“l”).

la distinción entre enfermos y sanos son las de desviación estándar, como es el caso de las características 22 ($\text{std}(\sin(\theta_{si}))$), 26 ($\text{std}(\Delta t_0 = t_{0i} - t_{0i-1})$) y 27 ($\text{std}(v_2 = |v_i(t_{2i})|)$), tal y como se puede ver en la figura 2.19. El número de lognormales en esta prueba se encuentra entre las 5 características con mayor distancia.

Al comparar las distribuciones obtenidos en la prueba 4 (“les”) con la número 3 se observa que la primera se comporta peor para distinguir entre los grupos estudiados.

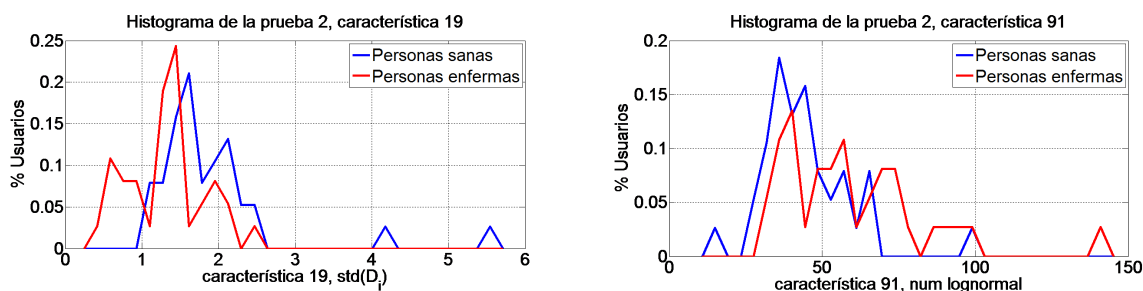


Figura 2.18: Distribuciones obtenidas de la prueba 2 (“l”), mediante características neuromotoras.

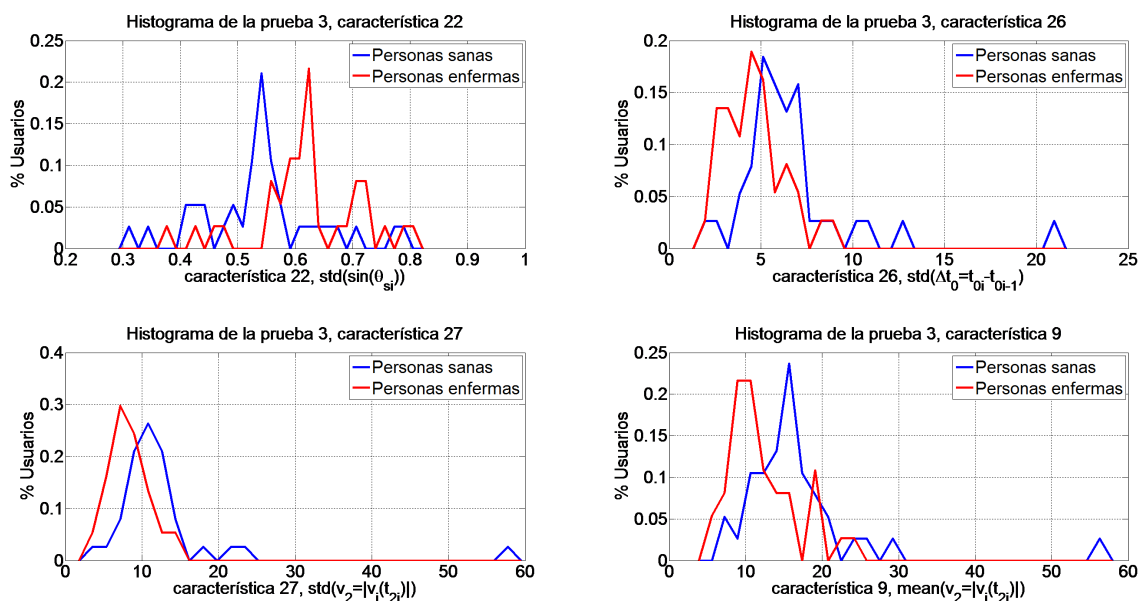


Figura 2.19: Distribuciones obtenidas de la prueba 3 (“le”), mediante características neuromotoras.

Las distribuciones de datos normalmente están más dispersas, o como en el caso de la característica 60 los histogramas están solapados. El ejemplo mostrado en la figura 2.20 se repite en algunos casos al emplear tanto el percentil 1 como el 99, normalmente estas características no serán buenas para realizar la clasificación biclase. Con las características destacadas de la prueba 5 (“lektorka”) sucede lo mismo que en el caso de la prueba 4, de la misma señalar que los sujetos sanos tienden a producir mayor cantidad de lognormales que los sanos, tal y como se puede observar en la figura 2.20.

Las características destacadas de la prueba 6 (“porovnat”) se relacionan en su mayoría con valores de percentil, pero dos de las mismas no aportan información debido al solapamiento entre los datos, mientras que la característica 55 (percentil99 (D_i)) si aporta mayor información. Otras características relacionadas con la distancia que se consideran discriminantes es la número 1 (mean (D_i)) y la 19 (std(D_i)), mostradas en la figura 2.21.

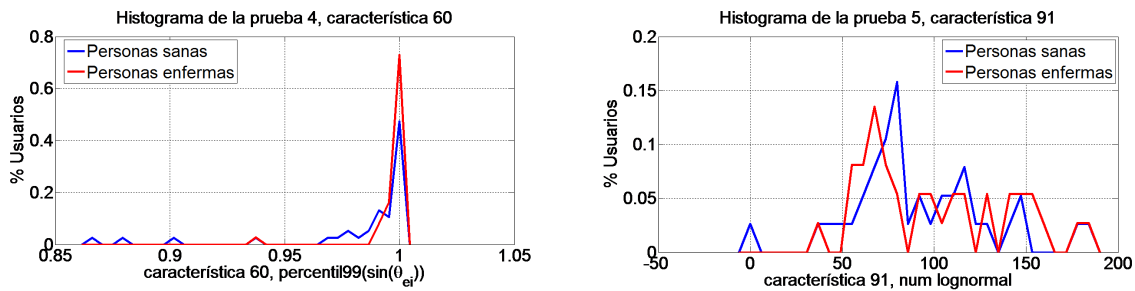


Figura 2.20: Distribuciones obtenidas de las pruebas 4 (“les”) y 5 (“lektorka”), mediante características neuromotoras.

Las características destacadas de la prueba 7 se relacionan con la media, destacar que la número 58 (percentil99($\sin(\theta_{si})$)) tiene una distribución muy solapada como sucede con la mayoría de las distribuciones de percentiles. En las demás, los datos están muy distribuidos.

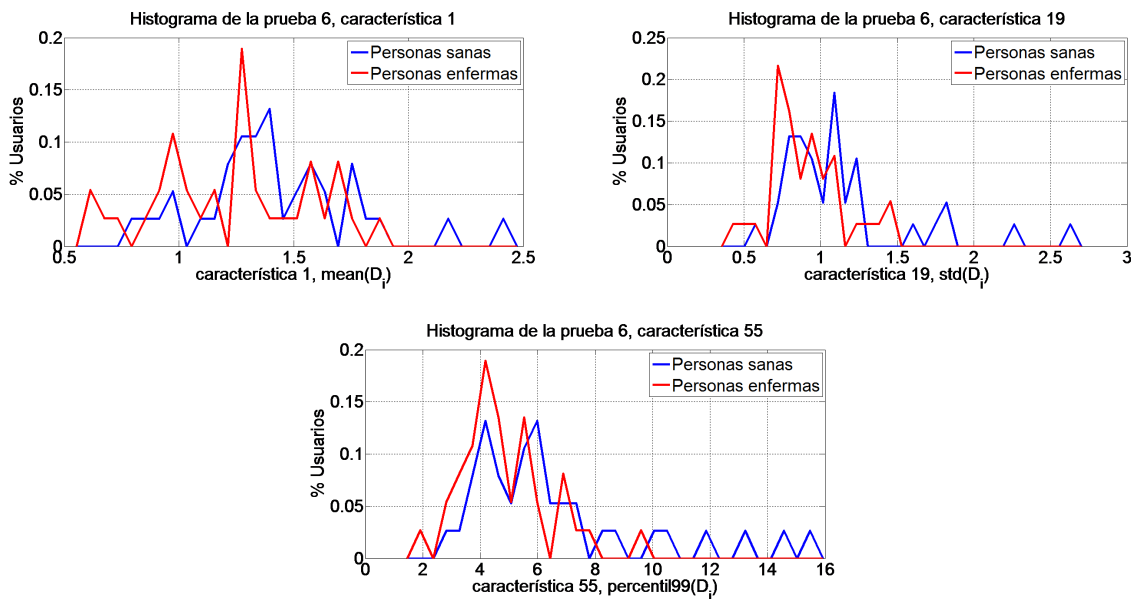


Figura 2.21: Distribuciones obtenidas de las pruebas 6 (“porovnat”), mediante características neuromotoras.

La prueba 8 (frase) puede ser empleada para distinguir entre ambos grupos, puesto que sus distribuciones están bastante separados al emplear las características neuromotoras, tal y como se puede observar en la figura 2.22. En todos ellos los datos de los sujetos enfermos se encuentran con valores inferiores que en el caso de los sujetos de control.

En conclusión, debido a la distinta naturaleza de las pruebas los valores estadísticos que se comportan mejor en cada una de ellas son distintos, por ejemplo en la prueba 1 las características destacadas se basan en la mediana de los datos, mientras que en la prueba

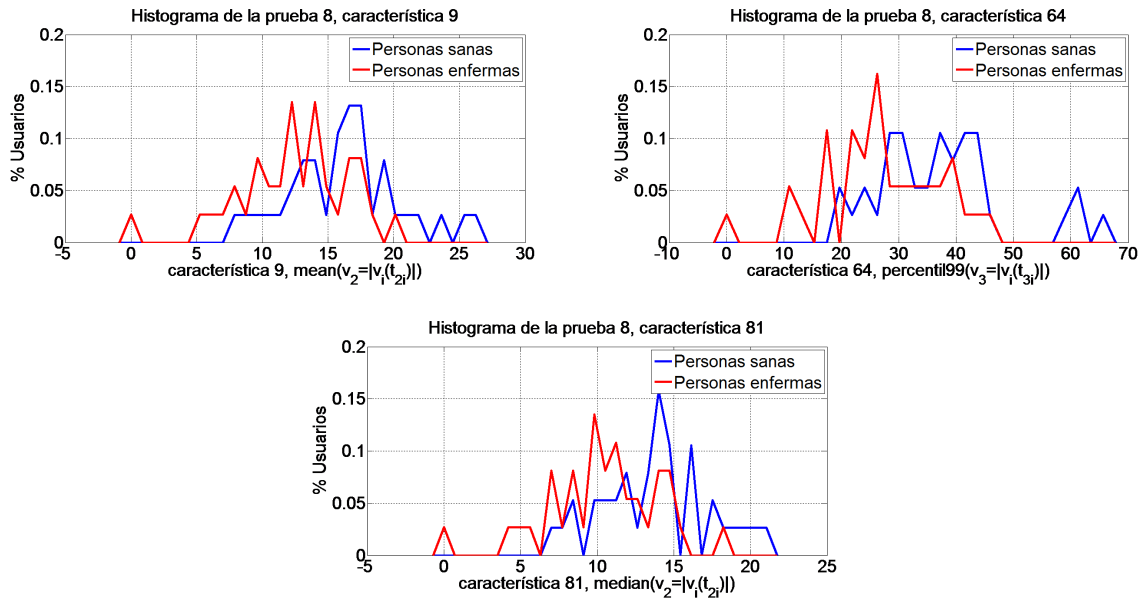


Figura 2.22: Distribuciones obtenidas de las pruebas 8 (frase), mediante características neuromotoras.

3 se basan en desviación estándar. Por otro lado a simple vista se pueden distinguir que pruebas se comportan mejor al emplear las características neuromotoras, como puede ser la número 3.

Las características destacadas que se repiten en las distintas pruebas son: 2 ($\text{mean}(\mu_i)$), 8 ($\text{mean}(\Delta t_0 = t_{0i} - t_{0i-1})$), 9 ($\text{mean}(v_2 = |v_i(t_{2i})|)$), 10 ($\text{mean}(v_3 = |v_i(t_{3i})|)$), 22 ($\text{std}(\sin(\theta_{si}))$), 60 ($\text{percentil99}(\sin(\theta_{ei}))$), 73 ($\text{median}(D_i)$), 74 ($\text{median}(\mu_i)$), 80 ($\text{median}(\Delta t_0 = t_{0i} - t_{0i-1})$) y 81 ($\text{median}(v_2 = |v_i(t_{2i})|)$). La número 73 se repite un total de 4 veces, lo que nos puede indicar que es una característica idónea para la distinción de los sujetos y la número 80 un total de 3 veces, ambas son mediana de los datos. Por tanto, la media y la mediana son los valores estadísticos más idóneos cuando se está trabajando con características neuromotoras, mientras que los valores de los percentiles 1 y 99 no, porque la mayoría de las distribuciones muestran un solapamiento considerable entre los datos.

Se observa que en la mayoría de las distribuciones se sitúan los valores de los sujetos enfermos por debajo de los de control. Claro ejemplo se dá con la media de la longitud del trazo ($\text{mean}(D_i)$), donde los sujetos al producir trazos más discontinuos el valor de la media de la longitud del trazo disminuye, al igual que la desviación estándar de la longitud del trazo es mayor en los sujetos de control.

Otra característica importante para la distinción entre los grupos es el número de lognormales que componen el vector de velocidad, como cabría esperar en la mayoría de

los casos los sujetos enfermos producen mayor cantidad de las mismas, pero en el caso de la prueba 5 sucede lo contrario, tal y como se puede observar en la figura 2.20 (b).

Experimentos y resultados

3.1. Combinación de características globales

Combinando las características se puede obtener información complementaria de los sujetos, siempre y cuando las características combinadas no estén correladas. Se analizará primero la complementariedad por partes.

Agrupación cualitativa

Para ver qué características se complementan bien se han seleccionado las 5 mejores de la prueba con mayor distancia, la espiral, y se han analizado dibujando en una gráfica cada una de ellas en un eje. Siendo las mejores características de la primera prueba la 22 ($\frac{1st\ t(x_{max})}{T_w}$), la 23 ($\frac{2nd\ t(x_{max})}{T_w}$), la 42 (promedio de las sacudidas), la 95 (N(máximos locales x)) y la 98 (N(máximos locales y)).

A partir de las gráficas obtenidas se propone dividir el conjunto de población en 4 grupos distintos. La figura 3.1 se corresponde con la agrupación de los sujetos al emplear las características 42 (promedio de las sacudidas) x 22 ($\frac{1st\ t(x_{max})}{T_w}$). Y la figura 3.2 se corresponde con la agrupación de las características 22 ($\frac{1st\ t(x_{max})}{T_w}$) x 95 (N(máximos locales x)). Los números indicados en las gráficas se corresponden con la posición concreta del vector de características tanto de enfermos como de sanos, y no con el identificador de la base de datos.

Al comparar las tablas 3.1 y 3.2 se observa que la mayoría de los sujetos en los distintos grupos son los mismos, a excepción de algunos sujetos que varían de grupo. Esto indica que, independientemente de las características que empleemos para realizar la agrupación, los grupos van a ser muy parecidos.

Conclusiones:

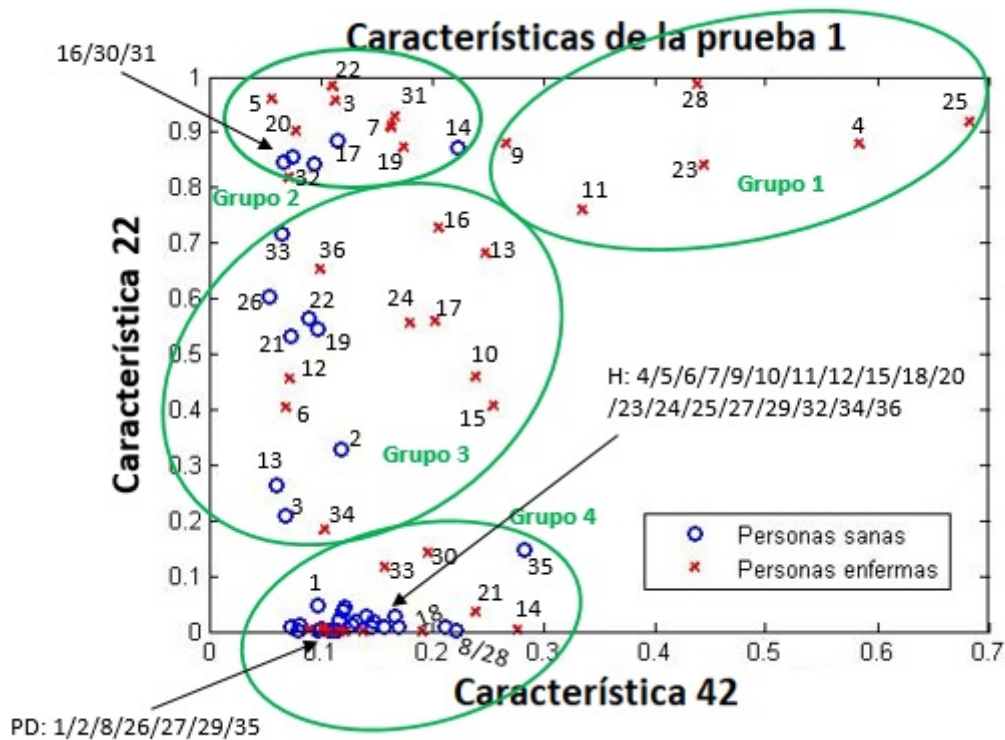


Figura 3.1: Combinación de características 42x22 de la prueba 1.

Tabla 3.1: Agrupación cualitativa de los sujetos en 4 grupos distintos, combinando las características 42 y 22. El primer grupo se corresponde con sujetos enfermos, el segundo agrupa pacientes con síntomas de enfermedad, el tercero es una mezcla entre personas con temblores aparentes y sin ellos, y el cuarto grupo se corresponde con los sujetos que no presentan síntomas de enfermedad.

Característica 42 x Característica 22		
Grupos	Enfermos	Sanos
Grupo 1	4 9 11 23 25 28	
Grupo 2	3 5 7 19 20 22 31 32	14 16 17 30 31
Grupo 3	6 10 12 13 15 16 17 24 34 36	2 3 13 19 21 22 26 33
Grupo 4	1 2 8 14 18 21 26 27 29 30 33 35	1 4 5 6 7 8 9 10 11 12 15 18 20 23 24 25 27 28 29 32 34 35 36

- Grupo 1:** en este grupo se puede distinguir dos tipos de enfermos, los que tienen un grado 1 o 2 en la escala UPDRS (sujetos 5, 11, 23 y 25), los mismos realizan la espiral con una velocidad relativamente constante, pero muy reducida. Aun así se distingue que son enfermos por su elevado temblor en el dibujo. Mientras que los sujetos 4, 7 y 28 al llevar más años enfermos se aprecian los efectos de la enfermedad a simple vista.

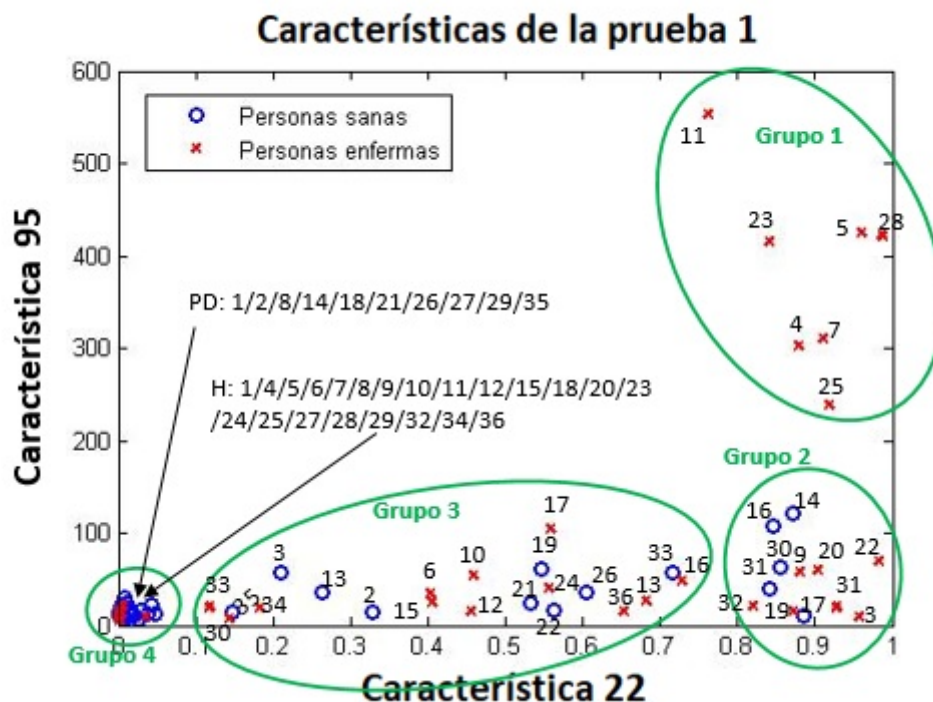


Figura 3.2: Combinación de características 22x95 de la prueba 1.

Tabla 3.2: Agrupación cualitativa de los sujetos en 4 grupos distintos, combinando las características 22 y 95. El primer grupo se corresponde con sujetos enfermos, el segundo agrupa pacientes con síntomas de enfermedad, el tercero es una mezcla entre personas con temblores aparentes y sin ellos, y el cuarto grupo se corresponde con los sujetos que no presentan síntomas de enfermedad.

Característica 22 x Característica 95		
Grupos	Enfermos	Sanos
Grupo 1	4 5 7 11 23 25 28	
Grupo 2	3 9 19 20 22 31 32	14 16 17 30 31
Grupo 3	6 10 12 13 15 16 17 24 30 33 34 36	2 3 13 19 21 22 26 33 35
Grupo 4	1 2 8 14 18 21 26 27 29 35	1 4 5 6 7 8 9 10 11 12 15 18 20 23 24 25 27 28 29 32 34 36

Por tanto se puede decir que este grupo de personas son las que más síntomas tienen de la enfermedad, además no aparece ninguna persona sana en dicho grupo. La media de edad de estos pacientes es de 75 años.

- **Grupo 2:** en este grupo se pueden distinguir dos tipos de sujetos enfermos. Las personas que tienen velocidad casi nula en la construcción de la espiral (20, 22, 31, y 32), son sujetos que llevan más de 9 años enfermos y una escala de grado

UPDRS III (Unified Parkinson's Disease Rating Scale III) mayor a 2, y los demás que tienen un vector de velocidad más irregular, llevan menos años enfermos y tienen un grado menor de enfermedad. Pero todos ellos presentan temblores a simple vista, en mayor o menor medida, como en el caso del sujeto 9, que aunque lleve tan solo 3 años enfermo presenta muchos síntomas.

Por otro lado la espiral realizada por el paciente sano número 14, se asimila a la de una persona que tiene la enfermedad en estado avanzado, esto se puede deber a que el paciente tiene 80 años y los temblores en el pulso y en la escritura se pueden deber a la edad. Los sujetos 16 y 31 también presentan rasgos de enfermedad, aunque no padezcan Parkinson, mientras que en las gráficas de los sujetos 17 y 30, no se aprecian temblores aparentes. La edad media de todos los sujetos pertenecientes al grupo es de aproximadamente 67 años.

- **Grupo 3:** en este grupo hay sujetos enfermos que muestran muchos temblores y síntomas como son los sujetos número 15, 16, 17 y 34, que presenten tantos síntomas se puede deber a que llevan bastantes años enfermos, el que menos lleva es 6 años y el que más 18 años, además los valores de escala UPDRS III son elevados. Los demás sujetos presentan síntomas, pero no excesivos en la mayoría de los casos.

Dentro de los sujetos sanos no han realizado una espiral con temblores, a excepción de los sujetos con mayor edad (sujeto 2 con 92 años y sujeto 35 con 77 años). Destacar que la diferencia de edad media entre sanos y enfermos en este grupo es de aproximadamente 8 años y una media de edad del conjunto de 67 años.

- **Grupo 4:** se podría decir que este grupo está compuesto por las personas que no presentan síntomas aparentes de enfermedad. Aunque haya sanos con temblores, no son temblores muy pronunciados para indicar que estuviesen enfermos. Igual que los temblores de los enfermos no son muy graves para diagnosticarles enfermedad. La media de edad del grupo es de 63 años, menor que en los grupos anteriores.

Algoritmo de agrupación automática

Para contrastar los datos de agrupación cualitativa se ha realizado una agrupación automática, para lo cual se ha empleado el algoritmo de Kmeans. Este es un algoritmo de agrupación simple y popular que divide el conjunto de datos en K grupos o cluster. Fijándose este parámetro K a priori, en nuestro caso se evaluarán 4 grupos.

El algoritmo Kmean define K centroides, pertenecientes a cada grupo, y cada dato se agrupará con el centroide que se encuentre en la posición más cercana. Por tanto cada

grupo se caracteriza por su centroide, que se encuentra en el centro de los elementos que componen el grupo [23]. El algoritmo se realiza en los siguientes pasos:

- 1.- Se eligen de forma aleatoria los centroides. Cada centroide inicial debe ser un valor de x_i sobre el conjunto de datos $D_n = (x_1, x_2, \dots, x_n)$.
- 2.- Para cada dato se asigna el centroide que esté más próximo según una medida de distancia (distancia Euclídea en nuestro caso).
- 3.- Una vez que se han asignado todos los datos a un centroide se recalculan los mismos, para que el centroide quede en la posición central del cluster.
- 4.- Se repiten los pasos 2 y 3 hasta que converge el algoritmo, es decir que no haya cambios en la agrupación de datos de una iteración anterior a la siguiente.

La ventaja de la representación mediante centroides es que tiene un significado gráfico y estadístico inmediato. Sin embargo, el inconveniente principal es que necesita una definición a priori del número de cluster, mientras que otros algoritmos no requieren esa condición.

La figura 3.3(a) se ha obtenido al aplicar el algoritmo de Kmeans con $K=4$. Para la obtención de esa agrupación se han empleado los datos de las 5 mejores características de la espiral y se ha realizado una representación 2D empleando las características 42 y 22 de todos los sujetos enfermos. Se observa en la figura 3.3(a) que la mayoría de los sujetos se han clasificado como pertenecientes al grupo 1. La agrupación cualitativa y la automática son muy distintas, las diferencias producidas se pueden deber a que la clasificación cualitativa se ha realizado en base a 2 características y no a 5 como al emplear el algoritmo Kmeans, además en el caso de la agrupación automática se han empleado solamente los sujetos enfermos. Al introducir al algoritmo toda la información de los sujetos, como edad, años que lleva enfermo, escala UPDRS III y la cantidad de medicación, se obtiene la agrupación mostrada en la figura 3.3(b).

Tabla 3.3: Agrupación automática de los sujetos enfermos en 4 grupos distintos, combinando las 5 mejores características y la información disponible de los sujetos.

Característica 42 x Característica 22	
Grupos	Enfermos
Grupo 1	3 4 10 28 29 31
Grupo 2	2 9 12 18 24 30 33 34
Grupo 3	5 7 8 11 17 19 23 25 27
Grupo 4	1 6 13 14 15 16 20 21 22 26 32 35 36

De la tabla 3.3 se pueden obtener las siguientes conclusiones:

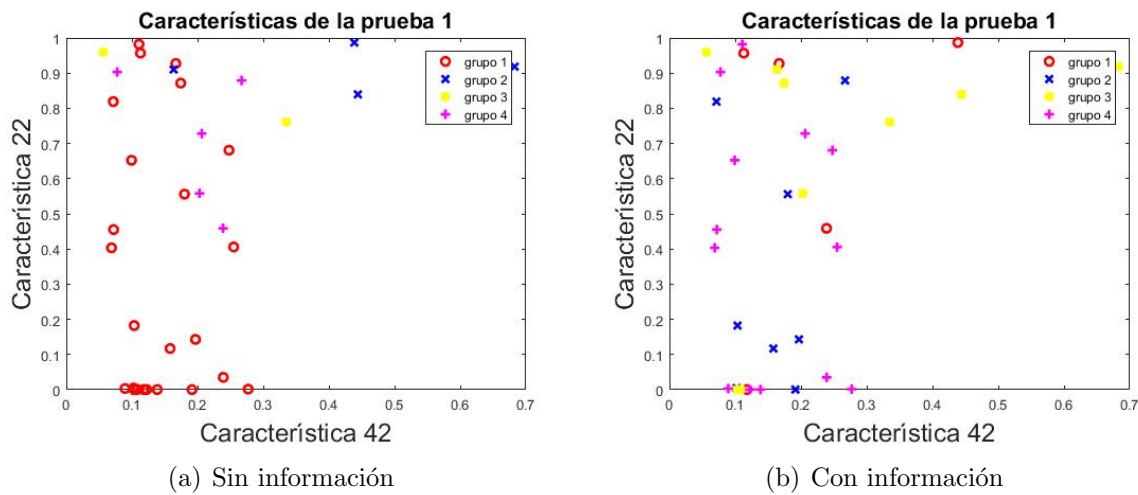


Figura 3.3: Agrupación automática, de los sujetos enfermos, con Kmeans para la agrupación de características 42 y 22. La figura (a) no emplea información de los sujetos para realizar la agrupación, la figura (b) introduce la información disponible de los sujetos como edad, medicación, sexo, etc para realizar dicha agrupación.

- **Grupo 1:** la mayoría de las pruebas de la espiral de los sujetos pertenecientes al grupo 1 presentan temblores e incluso en algunos casos aparecen pen up (han levantado el lápiz de la superficie de adquisición) en la realización de la prueba. Otras características comunes al grupo es que la mayoría de los sujetos son varones, y la medicación ronda aproximadamente 1600, así mismo, la media de edad es de 75 años, el grado UPDRS medio es 2.4 y llevan enfermos una media de 6.6 años.
- **Grupo 2:** en este grupo el sujeto 9 realiza la espiral con elevados temblores, y los demás sujetos presentan temblores moderados. Algunos de los vectores de velocidad tienen una envolvente relativamente suavizada.

La edad media es de 70 años, el grado medio UPDRS III es 2.4 y el tiempo que llevan enfermos es 9 años de media aproximadamente, además toman de medicación entre 2000 y 2500.

- **Grupo 3:** Los sujetos que pertenecen a este grupo se caracterizan por tener un vector de velocidad casi nulo o en otros casos no presentan apenas síntomas de enfermedad en la realización de la espiral. Sin embargo en la espiral de los sujetos 7 y 19 a simple vista se aprecian los síntomas de la enfermedad, el primero caso se podría deber a que lleva 19 años enfermos y en el caso del segundo sujeto que tiene 86 años aunque tan solo lleve 6 años enfermo.

Los sujetos de este grupo tienen un grado de enfermedad medio de 1.8, una edad media de 71 años y los años que llevan enfermos de media son 4.6. En este caso la medicación del grupo es menor en comparación con los demás grupos, puesto que

ronda entre 400 y 1000 aproximadamente.

- **Grupo 4:** este grupo se puede separar entre sujetos que parecen sanos, ya que no presentan síntomas de enfermedad como son los sujetos: 1, 6, 14, 20, 21, 26, 32, 35 y 36. Y los que sí presentan síntomas que son: 13, 15, 16 y 22.

La edad media es de 55 años aproximadamente, el grado UPDRS III medio es de 2.2 y el tiempo que llevan enfermos es 7.5 años de media aproximadamente, además toman una medicación media de 1300.

A continuación se ha realizado una agrupación con toda la información de los sujetos y las características destacadas de la espiral, pero realizando una clasificación con dos grupos ($K=2$) y empleando tanto a sujetos sanos como enfermos. Se ha obtenido la figura 3.4

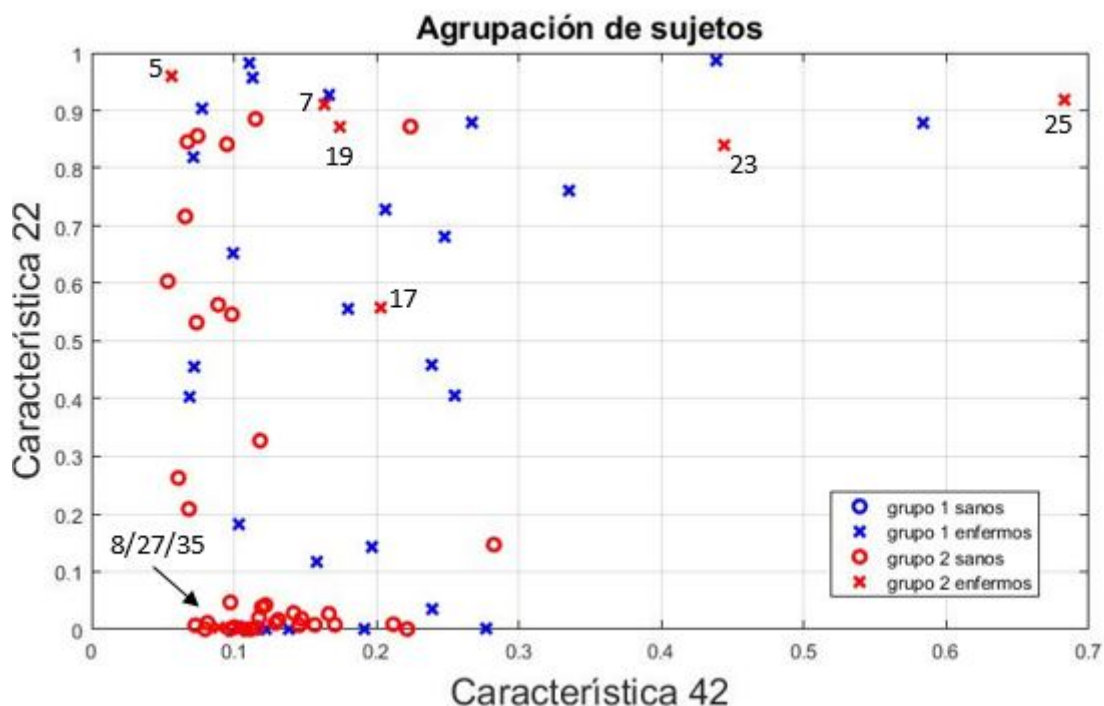


Figura 3.4: Agrupación automática $k=2$ de la prueba 1, y toda la información disponible de los sujetos, enfermos y sanos, junto con las 5 mejores características de dicha prueba.

El grupo 1 se corresponde con los pacientes enfermos y el grupo 2 se corresponde con los pacientes sanos, aunque en el segundo grupo se han introducido pacientes enfermos. Los sujetos enfermos etiquetados como sanos no presentan síntomas, podrían pasar por personas sanas, a excepción de 2 sujetos, los cuales tienen alto grado de enfermedad. Las características que tienen en común dichos sujetos son que toman poca cantidad de dopamina entre 397.5 a 800, tienen una edad media de 66 años, un grado UPDRS III

medio de 2, llevan de media 6 años enfermos, además el 80 % de los sujetos son de género femenino.

3.2. Combinación de características neuromotoras

Aplicando el mismo procedimiento anterior, se va a realizar una agrupación cualitativa y posteriormente una automática. Para la realización del mismo se ha empleado la prueba 2, la cual reporta mayor distancia media, combinando las características 9 y 10.

Agrupación cualitativa

Para ver qué características se complementan bien, se han representado las mejores características de la prueba 2, las cuales son: la 1 ($\text{mean}(D_i)$), la 8 ($\text{mean}(\Delta t_0 = t_{0i} - t_{0i-1})$), la 9 ($\text{mean}(v_2 = |v_i(t_{2i})|)$), la 10 ($\text{mean}(v_3 = |v_i(t_{3i})|)$) y la 73 ($\text{median}(D_i)$). Se ha elegido la representación de la característica 10 en el eje x y la 1 en el eje y. En la gráfica obtenida se observa claramente como el conjunto de población se puede dividir en 3 grupos, tal y como se muestra en la figura 3.5. Los números indicados en las gráficas se corresponden con la posición concreta del vector de características tanto de enfermos como de sanos, y no con el identificador de la base de datos.

Tabla 3.4: Agrupación cualitativa de los sujetos en 3 grupos distintos, combinando las características neuromotoras 10 y 1. El primer grupo se corresponde con sujetos enfermos, el segundo agrupa pacientes enfermos y sanos, y el cuarto grupo se corresponde con los sujetos que no presentan síntomas de enfermedad.

Característica 10 x Característica 1	
Grupos	Enfermos Sanos
Grupo 1	2 3 5 12 17 22 25 28 29 31 36
Grupo 2	4 6 7 8 9 10 11 13 14 15 16 18 19 20 21 24 26 27 30 32 33 34 35 37 37
Grupo 3	1 23 4 5 7 8 10 12 13 19 23 30 31 34 38

Conclusiones:

- **Grupo 1:** está compuesto por pacientes que llevan enfermos 6 ± 4 años de media, tienen una escala UPDRS media de 2.4 ± 0.67 y una edad media de 73.5 ± 11 años.

La prueba 2 de estos enfermos se puede dividir en dos conjuntos. El primer conjunto (realizado por los sujetos 2, 3 y 5), se corresponde con pruebas que no presentan

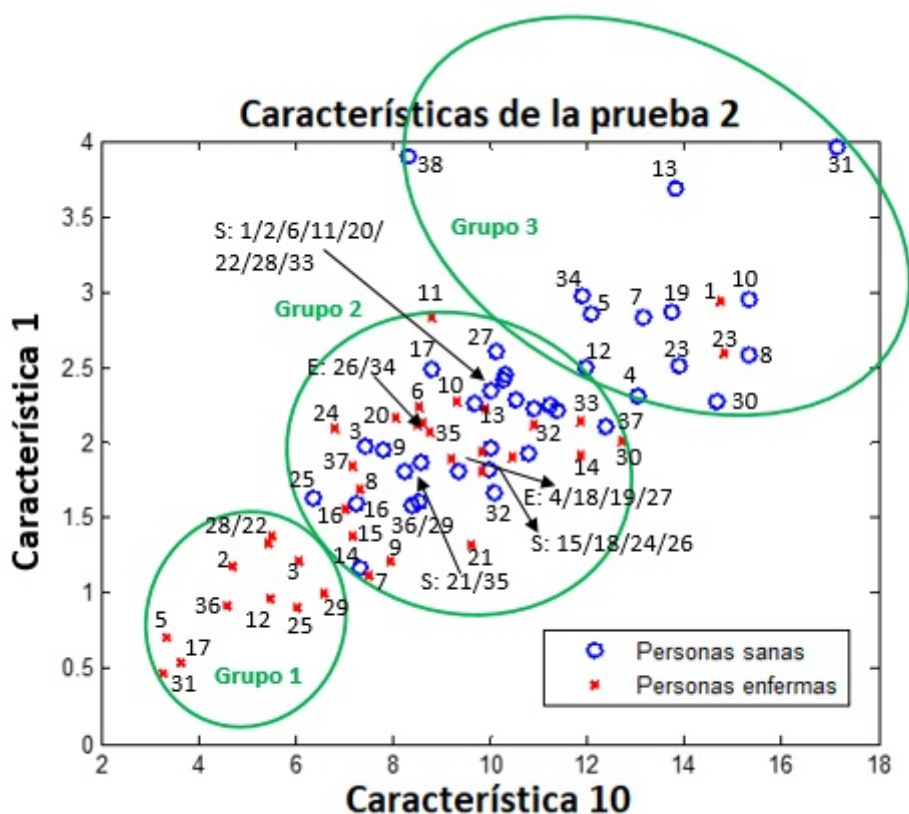


Figura 3.5: Combinación de características 10x1 de la prueba 2 para características neuromotoras.

temblor aparente, en el vector de velocidad se puede distinguir entre las distintas letras y los espacios en blanco y tienen una envolvente de la señal relativamente suavizada. Mientras que el otro conjunto se corresponde con pruebas con mayores síntomas de enfermedad, puesto que aparecen temblores, las letras no se aprecian bien y los vectores de velocidad tienen una envolvente muy picuda, incluso en algunos casos las letras no están bien formadas, como en el caso de las pruebas realizadas por los sujetos 17 y 28.

- **Grupo 2:** se corresponde con una mezcla de sujetos enfermos y sanos. Los sujetos de control tienen una media de edad de 64 ± 11 años, mientras que los sujetos enfermos tienen una edad media de 66 ± 11 años, una medicación media de 1260 ± 660 , una escala UPDRS media de 2 ± 0.9 y una media de edad de 7.8 ± 5.2 años.

En el caso de los pacientes sanos es curioso observar como algunos sujetos realizan la prueba con trazos muy suavizados, sin temblor y con una caligrafía muy cuidada, mientras que otros presentan muchos temblores y trazos discontinuos. El sujeto 2, aunque no padece el trastorno presenta muchos síntomas de enfermedad en la

realización de la prueba que se pueden deber a que tiene 92 años, al igual que sucede con el paciente 14 de 80 años. Algunas pruebas de los sujetos enfermos de este grupo presentan vectores de velocidad muy picudos, con temblores en la escritura de las letras y con elevados pen up entre las mismas.

- **Grupo 3:** se corresponde con sujetos sanos, a excepción de los números 1 y 23. Los pacientes que sufren el trastorno llevan enfermos 4 y 6 años, tienen un 2 en la escala UPDRS y una edad media de 65 años, además como cabría esperar no presentan síntomas de enfermedad. La edad media del grupo es de 58 ± 10 años, inferior que los grupos anteriores.

A simple vista no se presentan elevados temblores y síntomas de enfermedad, a excepción de los sujetos 31 y 38 que podrían pasar como sujetos enfermos, aunque uno de ellos tenga 44 años.

Algoritmo de agrupación automática

Para contrastar los datos de agrupación cualitativa se ha realizado un algoritmo de Kmean para la agrupación de los sujetos enfermos, en k conjuntos definidos previamente. Se ha aplicado el algoritmo Kmean con $k=3$, sobre las 5 mejores características de la prueba 2, obteniéndose la figura 3.6(a) en la que no se introduce información de los sujetos, como edad, medicación, sexo, etc y la figura 3.6(b) donde sí se introduce dicha información, la representación de ambas figuras se ha realizado en 2D empleando las características 1 y 10.

Tabla 3.5: Agrupación automática de los sujetos enfermos en 3 grupos distintos, combinando las 5 mejores características neuromotoras de la prueba 2 y la información disponible de los sujetos.

Característica 10 x Característica 1 (Con información)	
Grupos	Enfermos
Grupo 1	1 3 4 6 10 12 14 15 16 20 21 22 26 28 37
Grupo 2	2 9 18 24 29 30 31 32 33 34 36
Grupo 3	5 7 8 11 13 17 19 23 25 27 35

Comparando la agrupación cualitativa con la agrupación automática se observa que los sujetos se han agrupado de manera totalmente distinta, tan solo coinciden algunos sujetos.

De la tabla 3.5 se pueden obtener las siguientes conclusiones:

- **Grupo 1:** la edad media de este conjunto es de 69 años, toman una medicación

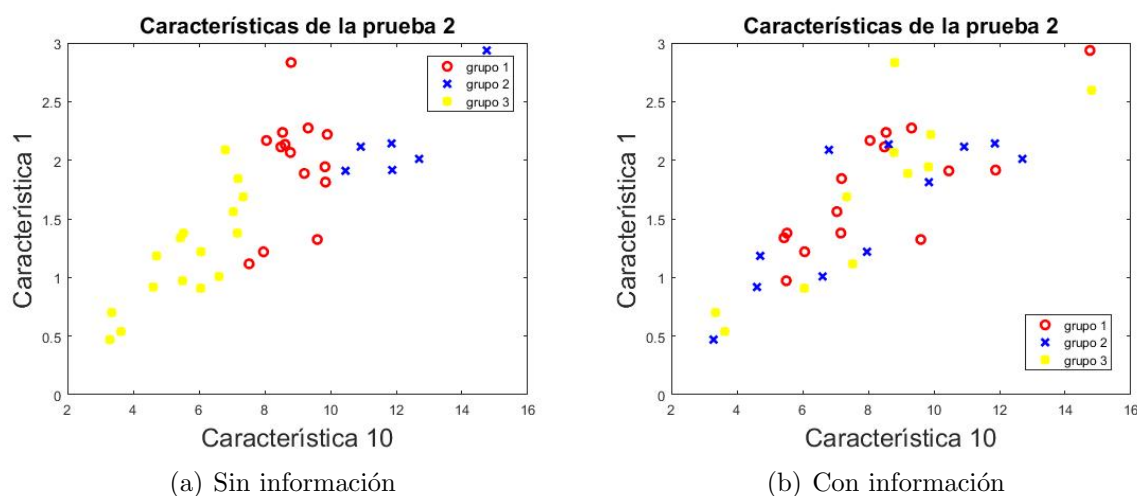


Figura 3.6: Agrupación automática de los sujetos enfermos con Kmeans para la agrupación de características neuromotoras 10 y 1. La figura (a) no emplea información de los sujetos para realizar la agrupación, la figura (b) introduce la información disponible de los sujetos como edad, medicación, sexo, etc para realizar dicha agrupación.

media de 1300, un grado UPDRS III medio de 2.4 y de media llevan enfermos aproximadamente 8 años.

En este grupo la mayoría de los enfermos no presentan síntomas a simple vista ni en el vector de velocidad ni en la realización de las letras, pero hay algunos que si presentan elevados síntomas como en el caso de los sujetos 4, 15 y 16, los cuales llevan enfermos más de 12 años.

- **Grupo 2:** la edad media de este conjunto de enfermos es de 70, toman una medicación media de aproximadamente 2200, llevan enfermos una media de 7 años y un valor de 2 en la escala UPDRS III.

Este conjunto se puede dividir entre los sujetos que presentan síntomas a simple vista y los que no, ya sea por temblor o por irregularidades en el vector de velocidad. Aunque hay pacientes que llevan más de 8 años enfermos no presentan tantos síntomas como los que llevan menos años enfermos, por ejemplo el sujeto 9 que lleva 3 años enfermo presenta más síntomas que el número 33 que lleva 12 años enfermo.

- **Grupo 3:** la medicación en el tercer grupo es inferior que en los anteriores, en este caso está entre 400 y 1000. La edad media del grupo es de 66 años, llevan enfermos una media de 6 años y tiene de media un grado 1.8 en la escala UPDRS III.

La mayoría de los sujetos no presentan síntomas de enfermedad a simple vista, por tanto se podría decir que este grupo es de enfermos con menor avance del trastorno. A excepción de los sujetos 7, 17 y 25, el primero lleva 19 años enfermo y los otros dos tienen una edad de 86 y 90, respectivamente.

A continuación se ha realizado una agrupación automática con todos los sujetos de la base de datos, con toda la información de los participantes y las características destacadas de la prueba 2, pero realizando una clasificación en dos grupos ($K=2$). Se ha obtenido la figura 3.7

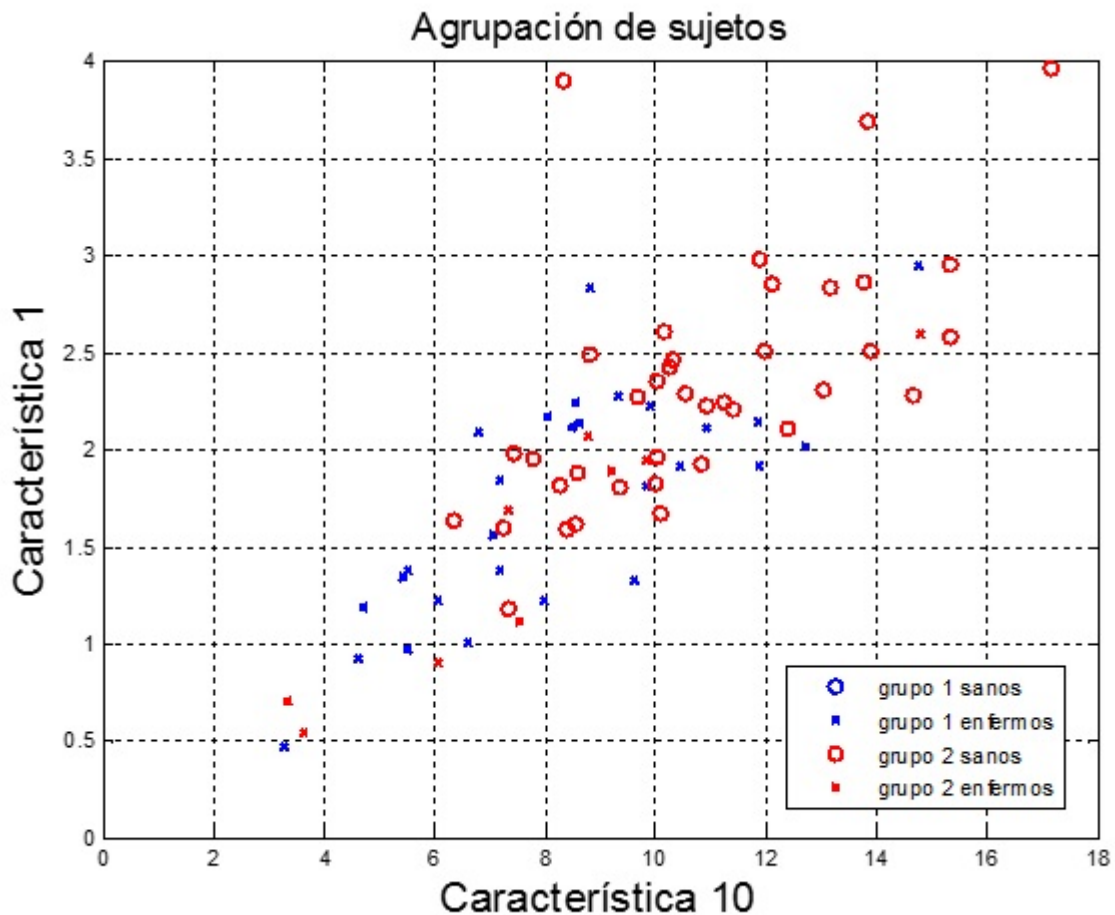


Figura 3.7: Agrupación automática con $k=2$ de la prueba 2 con características neuromotoras, y toda la información disponible de los sujetos junto con las 5 mejores características de dicha prueba.

En la figura 3.7, el grupo 1 se corresponde con los sujetos enfermos y el grupo 2 con los sujetos sanos, aunque hay sujetos enfermos que se han introducido en el grupo de sujetos sanos. Estos no presentan síntomas y podrían pasar por personas sanas, a excepción de los sujetos 7 y 17, el primero lleva enfermo 19 años y otro segundo lleva enfermo 6 años y tiene 86 años. Las características que tienen en común dichos sujetos son que toman poca cantidad de dopamina, tienen una edad media de 66 ± 16 años, un grado UPDRS III medio de 1.8 ± 0.6 y llevan de media 6 ± 5.8 años enfermos.

3.3. Clasificación basada en Máquinas de Vectores de Soporte

Otro objetivo es construir un modelo que nos permita distinguir entre personas sanas y enfermas, es decir realizar un clasificador binario. Para ello se ha empleado SVM (support vector machines), la elección de dicho método se debe a que es el que mejores resultados en términos de rendimiento ha reportado en [10].

SVM es un algoritmo de aprendizaje supervisado, que construye una separación óptima entre clases mediante el hiperplano de separación en el espacio [24]. Esto permite crear un modelo, basado en los datos de entrenamiento, que prediga la clase a la que pertenece una nueva muestra introducida al conjunto de entrenamiento.

Para realizar el modelo de clasificación es necesario emplear un Kernel, en nuestro caso el Kernel empleado es el RBF (Radial Basis Function), el cual tiene forma de campana. Los dos parámetros que definen el kernel son: “ γ ” y “ σ ”. El parámetro de regulación γ permite aumentar o disminuir la complejidad del modelo, si γ es bajo el modelo es simple y no evita los errores de clasificación producido en los datos de entrenamiento, mientras que si el parámetro γ es alto el algoritmo evita los datos clasificados erróneamente, lo que puede llegar a producir overfitting (sobreajuste). El overfitting es una situación que se produce cuando el modelo de clasificación se basa en observaciones particulares en vez de la estructura general de los datos [25], lo que se traduce en errores de rendimiento al introducir nuevas muestras al sistema. Y el parámetro σ se corresponde con el ancho de banda del kernel. El ajuste de estos parámetros puede empeorar o mejorar los resultados de entrenamiento del clasificador.

Teniendo definidos los parámetros del kernel para el sistema SVM se procede a realizar la clasificación. Se van a realizar dos clasificadores distintos, uno con características globales y otro con neuromotoras, los pasos a seguir para realizar los mismos y obtener los datos de rendimiento son los siguientes:

- De todas las características disponibles, ya sean neuromotoras o globales, se coge una selección de las mejores, puesto que al emplear todas se consigue empeorar los resultados ya que algunas son nulas o no aportan información. Para realizar la selección de las mejores se han realizado varias pruebas hasta determinar los 3 conjuntos distintos:
 - Mejores características en términos de distancia, en este conjunto de características se emplean las mejores según el parámetro de distancia obtenido

entre las distribuciones de datos, los valores de distancia obtenidos en cada prueba se pueden observar en las tablas 4.1 y 4.2 de los anexos. Tras realizar varias pruebas se ha definido que el vector de características globales empleado para la clasificación está compuesto por las 30 mejores características, en cada una de las pruebas, es decir según la prueba definida para realizar el clasificador el vector de características empleado será las 30 mejores características de dicha prueba. Mientras que en el caso de las características neuromotoras el sistema empleará las 50 mejores características.

- Características con menor correlación, de todas las características se ha calculado la correlación entre los sujetos sanos y enfermos para cada una de las pruebas y se han seleccionado las que tiene un coeficiente de correlación menor a 0.05 ($\rho < 0.05$).
 - Características significativas, se ha calculado que características muestran diferencias estadísticas significativas entre los conjuntos de sujetos enfermos y sanos. Para ello se ha calculado el valor de p y se han elegido las que $p < 0.05$. El valor de p se define como la probabilidad de obtener un resultado al menos tan extremo como el que se ha obtenido, suponiendo que la hipótesis nula es cierta.
- Tras seleccionar un conjunto de características ya sea global o neuromotor se obtienen los datos de las clases, las cuales se van a tratar por separado, es decir por un lado el vector de sujetos enfermos y por otro el de control.
 - Será necesario definir el tamaño de las muestras de test y de entrenamiento, además de las repeticiones realizadas por el algoritmo. En las pruebas realizadas el tamaño de las muestras de entrenamiento es de 80 % y el de test de 20 %.
 - Se eligen de manera aleatoria los sujetos sanos y enfermos que van a ser empleados en los datos de test y entrenamiento, teniendo en cuenta el tamaño de las muestras. Después se crea el vector de etiquetas de las clases, donde una tendrá valor -1 y la otra valor 1.
 - Los datos de entrenamiento se corresponden con los sujetos restantes. También es necesario definir el vector de etiquetas de estos datos.
 - Para crear el modelo los datos de entrenamiento son introducidos en SVM, empleando un kernel RBF cuyos parámetros son: $\gamma = 20$ y $\sigma = 100$.
 - Cuando ya se tiene el modelo se introduce el mismo junto con los datos de test para obtener las puntuaciones de las muestras de test.
-

- Para presentar los valores de rendimiento es necesario identificar que puntuaciones son genuinos y cuales impostores, para ello se tiene que emplear el vector de etiquetas de las muestras de test.
- Con las puntuaciones positivas y negativas (sujetos sanos clasificados como enfermos, y viceversa) se calculan las curvas FAR y FRR. FAR, tasa de falsa aceptación del sistema, es decir es la probabilidad de que un sujeto impostor sea clasificado como genuino según un umbral de decisión. FRR, tasa de falso rechazo del sistema, es decir es la probabilidad de que un sujeto genuino sea clasificado como impostor según un umbral de decisión. El punto de intersección de ambas curvas se conoce como EER (tasa de igual error), es decir el punto donde la tasa de falso rechazo y falsa aceptación son la misma, el valor de EER es un medidor del rendimiento general del sistema.

El proceso se repite de igual manera para cada clasificador, es decir utilizando las características globales y neuromotoras. Con las puntuaciones de ambos sistemas se hace una combinación de los mismos para ver si mejora el resultado general de clasificación debido al empleo de información complementaria sobre los patrones de clasificación. Pero la principal razón para combinar clasificadores es que la combinación de estos puede aumentar la eficiencia y la precisión [26]. Esta combinación es útil cuando se emplean diferentes conjuntos de características o diferentes conjuntos de entrenamiento, en nuestro caso se van a emplear características distintas. Como un sistema produce mejores resultados de clasificación, se van a combinar los clasificadores con la técnica de suma ponderada. En la cual se ha empleado pesos, para darle más importancia a los datos obtenidos del mejor clasificador, se han realizado varias pruebas hasta definir que el peso es $w=0.3$ para el clasificador 1 de características globales y $w=0.7$ para el de características neuromotoras.

Se han realizado pruebas preliminares para definir todos los parámetros del sistema, como puede ser el tamaño de los vectores de características en términos de distancia. Tras varias pruebas se ha definido que el tamaño del mismo es de 30 y 50 características en los clasificadores de características globales y neuromotoras respectivamente. También se han probado varios parámetros del kernel, hasta optar el empleo de $\gamma = 20$ y $\sigma = 100$.

Otro parámetro a definir es el peso de los sistemas a combinar, para ello se han realizar varias pruebas con distintos pesos ($w=0.3$, 0.5 y 0.7). Se ha comprobado que el peso idóneo es de $w=0.3$ para el caso de sistema de características global y de $w=0.7$ para el de características neuromotoras, por tanto se da mayor importancia a los resultados obtenidos del segundo clasificador.

3.4. Resultados de clasificación

Empleando todos los parámetros comentados en el apartado anterior, se obtienen los resultados mostrados en las tablas 3.6, 3.7 y 3.8, se han señalado en rojo los mejores resultados en términos de rendimiento. La diferencia entre estas tablas se corresponde con el conjunto de características empleadas.

En la tabla 3.6 se muestran los resultados obtenidos al emplear las características con mayor distancia. La prueba 1 era la que mayor distancia media obtenía al emplear las características globales, pero al clasificar es la prueba 8 la que mejor rendimiento obtiene, y en el caso de las características neuromotoras es la prueba 3 la que mejores resultados obtiene. Por tanto el menor error de clasificación al emplear estas características es de 26.1786 %, obtenido al combinar los sistemas empleando las muestras de la prueba 3.

Tabla 3.6: Resultados de clasificación obtenidos al emplear las características con mayor distancia.

Clasificador 1 (global)			Clasificador 2 (neuromotor)			Combinación
Nº prueba	Características	EER	Nº prueba	Características	EER	EER
1	30 mejores	38.8214	1	50 mejores	42.9286	38.8929
2	30 mejores	42.0714	2	50 mejores	38.5	38.2143
3	30 mejores	36.5	3	50 mejores	31.8571	26.1786
4	30 mejores	30.8571	4	50 mejores	45.3929	40.7857
5	30 mejores	44.7857	5	50 mejores	52.7143	49.9643
6	30 mejores	36.8571	6	50 mejores	37.4286	34.6786
7	30 mejores	48.3929	7	50 mejores	51.6071	48.7143
8	30 mejores	29.8571	8	50 mejores	36.75	29.0714

La tabla 3.7 muestra también los resultados de rendimiento, pero ahora empleando las características sin correlación entre sanos y enfermos. En este caso el mínimo error es de 33.7857 % si se emplea la prueba 2 en el clasificador 1. Al comparar estos resultados con los anteriores se observa que son peores en términos de rendimiento.

La tabla 3.8 muestra los resultados al emplear las características con diferencias significativas, en el caso de la prueba 5 no hay características que tengan diferencias significativas. Al igual que sucedía en los casos anteriores el sistema 2 con las muestras obtenidas de la prueba 3 da los mejores resultados de rendimiento, en torno al 28.7857 %, pero los mismos peores que el caso de emplear las características con mayor distancia.

De las tablas anteriores se han elegido los mejores resultados de rendimiento y se han realizado distintas pruebas combinando los mejores clasificadores para ver si mejoran los mismos al combinarlos. Estos resultados son los mostrados en la tabla 3.9.

Tabla 3.7: Resultados de clasificación obtenidos al emplear las características sin correlación.

Clasificador 1 (global)			Clasificador 2 (neuromotor)			Combinación
Nº prueba	Características	EER	Nº prueba	Características	EER	EER
1	Sin correlación	42.6071	1	Sin correlación	42.2857	38.2143
2	Sin correlación	38.5	2	Sin correlación	39.4643	36.8929
3	Sin correlación	38.3929	3	Sin correlación	33.7857	33.7857
4	Sin correlación	52.4286	4	Sin correlación	50.75	46.4643
5	Sin correlación	45.8214	5	Sin correlación	49.5	50.4286
6	Sin correlación	43.1429	6	Sin correlación	45.9286	47.5
7	Sin correlación	46.5	7	Sin correlación	53.0714	50.9286
8	Sin correlación	41.0714	8	Sin correlación	38.3571	37.3571

Tabla 3.8: Resultados de clasificación obtenidos al emplear las características con diferencias significativas.

Clasificador 1 (global)			Clasificador 2 (neuromotor)			Combinación
Nº prueba	Características	EER	Nº prueba	Características	EER	EER
1	Significativas	37.25	1	Significativas	36.5714	32.3571
2	Significativas	38.4286	2	Significativas	32.75	33.3929
3	Significativas	34.4286	3	Significativas	35.1786	28.7857
4	Significativas	32.0714	4	Significativas	41.6429	38.1786
5	Significativas	33.1786	5	Significativas		
6	Significativas	39.9286	6	Significativas	37.2314	33.3929
7	Significativas	40.8214	7	Significativas	44.1786	40.9643
8	Significativas	29.8571	8	Significativas	32.3571	29.2857

Tabla 3.9: Resultados de clasificación obtenidos al realizar distintas combinaciones.

Clasificador 1 (global)			Clasificador 2 (neuromotor)			Combinación
Nº prueba	Características	EER	Nº prueba	Características	EER	EER
8	30 mejores	29.75	3	50 mejores	31.3571	28.1786
4	30 mejores	31.7143	3	50 mejores	29.8214	29.2143
3	30 mejores	35.7857	3	50 mejores	31.0714	26.7143
8	Significativas	28.8571	3	Significativas	35.25	31.4286

El mejor resultado de rendimiento se obtiene al combinar las pruebas número 3 de ambos sistemas al emplear las mejores características según la distancia entre las clases. Dicho rendimiento se encuentra en torno al 26% de error. También se puede observar que la combinación de clasificadores no siempre produce resultados mejores que si se empleasen los clasificadores individualmente.

Otra prueba realizada ha consistido en definir un clasificador individual por cada una de las tareas de escritura y combinar sus puntuaciones mediante suma ponderada,

siendo el peso de cada sistema igual para todos, es decir $w=1/8$. Además se ha realizado la suma ponderada de los 16 clasificadores distintos, es decir 8 clasificadores empleando características globales y otros 8 clasificadores neuromotores, para cada una de las pruebas, el peso en esta combinación es $w=1/16$. Los resultados obtenidos son los mostrados en la tabla 3.10. Donde los mejores resultados se obtienen al combinar las 8 pruebas mediante características significativas globales, los resultados están entorno al 22.5 % de error, inferior a los obtenidos anteriormente.

Tabla 3.10: Resultados de clasificación obtenidos al combinar todas las pruebas de escritura.

Combinación de clasificadores			
Características	8 globales	8 neuromotores	8 globales + 8 neuromotores
Mejores	26.2857	32.4286	33.8036
Sin correlación	38.4286	38.6429	39.5338
Significativas	22.5714	30.4286	30.2857

Conclusiones

Al observar todas las pruebas y graficas obtenidas de los sujetos que han participado en el estudio se observa que el diagnóstico de pacientes enfermos de PD implica que cada persona responde de manera diferente al medicamento, los síntomas son distintos para cada persona, cualquier otro tipo de trastorno puede afectar a la escritura e incluso la edad produce temblores en la escritura que pueden ser considerados PD. Por tanto hay sujetos que a simple vista parecen que sufren la enfermedad, con un alto grado UPDRS, aunque en la realidad no sufran la misma. Incluso hay pacientes que llevan años enfermos y al tomar medicación no se aprecian síntomas a simple vista. Por todos estos motivos en la actualidad no existe ninguna prueba fiable que permita detectar el Parkinson, y las únicas que pueden confirmar la enfermedad se realizan en la autopsia.

Las pruebas mostradas en este estudio podrían ser empleados como método de diagnóstico de la enfermedad. Con las pruebas de escritura estudiadas y un dispositivo táctil con acceso a Internet el paciente podría realizar las pruebas desde su hogar, sin necesidad de desplazarse al hospital. Para el personal médico se podría tratar de una herramienta de apoyo para realizar un seguimiento más exhaustivo de la enfermedad y un tratamiento más personalizado. Otro aspecto importante observado es que los pacientes con PD sufren fatiga muscular, por tanto realizando pruebas con extensión temporal mayor podríamos intentar detectar la enfermedad con un porcentaje menor de error.

No todas las pruebas y características tienen el mismo nivel discriminante para caracterizar la enfermedad, por tanto es muy importante realizar una selección de las que mejor caracterizan la enfermedad para intentar obtener los mejores resultados en el clasificador biclase diseñado mediante SVM.

Los mejores resultados en términos de clasificación se han obtenido de combinar las 8 pruebas distintas al emplear las características globales, el rendimiento del sistema es de aproximadamente 22.5% de EER. Aunque estos resultados se encuentran por debajo de los obtenidos en [10], los cuales ronda el 18%, esto se podría deber al empleo de

características distintas, además de que en [10] solamente combinan las pruebas de la 2 a la 8.

Estos resultados son bastante favorables puesto que los pacientes diagnosticados con PD toman medicación para combatir los efectos del trastorno. Por tanto si se emplean estas pruebas como dispositivo preventivo los resultados de rendimiento del sistema final aumentarían ya que los pacientes enfermos no estarían tomando medicación.

En general se trata de un estudio bastante complejo, debido a la falta de diagnóstico de la enfermedad, no se puede determinar de forma concreta qué pacientes están enfermos. Sin embargo, sí que se ha podido comprobar que existen diferencias significativas entre los sujetos de este estudio que estaban enfermos y sanos que nos permiten separar por dichas clases. Por ello se podría decir que este estudio nos permite reconocer algunos de los síntomas de esta enfermedad, y por tanto, facilitar el diagnóstico de la misma.

4.1. Líneas futuras

Lo más importante para poder evolucionar con el tratamiento de pacientes enfermos de Parkinson, es seguir estudiando métodos que nos permitan detectar la enfermedad, y por consiguiente aumentar la calidad de vida de los pacientes que la sufren. En este estudio se han mostrado pruebas de escritura simples que podrían detectar la enfermedad, ahora toca seguir investigando para que estas pruebas ganen relevancia puesto que no son invasivas y ayudan a desempeñar la labor del personal sanitario.

Para seguir investigando se necesitaría aumentar la población bajo estudio, incluso realizar otras pruebas distintas con mayor duración temporal para ver cómo se comportan los sujetos. Otra idea importante es realizar de manera periódica las pruebas de diagnóstico a los sujetos para ver la evolución de la enfermedad y obtener mayores datos de comparación.

Otro aspecto importante es crear la aplicación en dispositivos táctiles para la realización de las pruebas, para ello será necesario que el dispositivo sea capaz de obtener los valores de presión ejercidos sobre la pantalla, además de un sistema de envío por Internet que mande los datos al personal médico asignado a cada paciente para que realice un estudio exhaustivo de las mismas.

Sería necesario encontrar pruebas que permitieran definir de una forma más clara y concisa el grado de enfermedad de cada sujeto, para así poder aplicar los medicamentos necesarios desde un primer momento, evitando así la demora que se puede producir en estudios que necesiten muchas pruebas para hallar el grado real del paciente.

Referencias

- [1] Wirdefeldt, K., Adami, H. O., Cole, P., Trichopoulos, D., & Mandel, J. (2011). “*Epidemiology and etiology of Parkinson’s disease: a review of the evidence*”. *European journal of epidemiology*, 26(1), 1-58.
- [2] Hirtz, D., Thurman, D. J., Gwinn-Hardy, K., Mohamed, M., Chaudhuri, A. R., & Zalutsky, R. (2007). “*How common are the “common” neurologic disorders?*”. *Neurology*, 68(5), 326-337.
- [3] García-Ramos, R., Valdés, E. L., Ballesteros, L., Jesus, S., & Mir, P. (2013). “*Informe de la fundación del cerebro sobre el impacto social de la enfermedad de Parkinson en España*”. *Neurología*.
- [4] “*Parkinson, una de las enfermedades con mayores expectativas en tratamientos futuros, señalan desde la SEN*” [en línea]. *Psiquiatria.com* 10 abril de 2013 [fecha consulta: 3 marzo 2017]. Disponible desde Internet: http://www.psiquiatria.com/parkinson_y_trastornos_del_movimiento/parkinson-una-de-las-enfermedades-con-mayores-expectativas-en-tratamientos-futuros-senalan-desde-la-sen.
- [5] O’Reilly, C., Plamondon, R., & Lebrun, L. H. (2014). “*Linking brain stroke risk factors to human movement features for the development of preventive tools*”. *Frontiers in aging neuroscience*, 6, 150.
- [6] Hernandez, A. B., Fischer, A., & Plamondon, R. (2015). “*Omega-lognormal analysis of oscillatory movements as a function of brain stroke risk factors*”. In 17th Biennial Conference of the International Graphonomics Society.
- [7] Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I., & Schlesinger, I. (2013). “*Handwriting as an objective tool for Parkinson’s disease diagnosis*”. *Journal of neurology*, 260(9), 2357-2361.

-
- [8] Ünlü, A., Brause, R., & Krakow, K. (2006). “*Handwriting analysis for diagnosis and prognosis of Parkinson’s disease*”. In International Symposium on Biological and Medical Data Analysis (pp. 441-450). Springer Berlin Heidelberg.
- [9] San Luciano, M., Wang, C., Ortega, R. A., Yu, Q., Boschung, S., Soto-Valencia, J., ... & Saunders-Pullman, R. (2016). “*Digitized Spiral Drawing: A Possible Biomarker for Early Parkinson’s Disease*”. PloS one, 11(10).
- [10] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2016). “*Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson’s disease*”. Artificial intelligence in medicine, 67, 39-46.
- [11] Drotár, P., Mekyska, J., Smékal, Z., Rektorová, I., Masarová, L., & Faundez-Zanuy, M. (2015). “*Contribution of different handwriting modalities to differential diagnosis of Parkinson’s Disease*”. In Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium on (pp. 344-348). IEEE.
- [12] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2015). “*Decision support framework for parkinson’s disease based on novel handwriting markers*”. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 23(3), 508-516.
- [13] Martinez-Diaz, M., Fierrez, J., Krish, R. P., & Galbally, J. (2014). “*Mobile signature verification: feature robustness and performance comparison*”. IET Biometrics, 3(4), 267-277.
- [14] O’Reilly, C., & Plamondon, R. (2009). “*Development of a Sigma–Lognormal representation for on-line signatures*”. Pattern Recognition, 42(12), 3324-3337.
- [15] Georgopoulos, A. P., Kalaska, J. F., & Massey, J. T. (1981). “*Spatial trajectories and reaction times of aimed movements: effects of practice, uncertainty, and change in target location*”. J Neurophysiol, 46(4), 725-743.
- [16] Abend, W., Bizzi, E., & Morasso, P. (1982). “*Human arm trajectory formation*”. Brain: a journal of neurology, 105, 331-348.
- [17] Lacquaniti, F., Licata, F., & Soechting, J. F. (1982). “*The mechanical behavior of the human forearm in response to transient perturbations*”. Biological Cybernetics, 44(1), 35-46.
- [18] Atkeson, C. G., & Hollerbach, J. M. (1985). “*Kinematic features of unrestrained vertical arm movements*”. Journal of Neuroscience, 5(9), 2318-2330.
-

-
- [19] Nagasaki, H. (1989). “*Asymmetric velocity and acceleration profiles of human arm movements*”. *Experimental brain research*, 74(2), 319-326.
- [20] Plamondon, R. (1995). “*A kinematic theory of rapid human movements*”. *Biological cybernetics*, 72(4), 295-307.
- [21] Galbally, J., Fierrez, J., Martinez-Diaz, M., & Plamondon, R. (2011). “*Quality analysis of dynamic signature based on the sigma-lognormal model*”. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on* (pp. 633-637). IEEE.
- [22] Fischer, A., & Plamondon, R. (2015). “*A dissimilarity measure for on-line signature verification based on the sigma-lognormal model*”. In *17th Biennial Conference of the International Graphonomics Society*.
- [23] Cambronero, C. G., & Moreno, I. G. (2006). “*Algoritmos de aprendizaje: knn & kmeans*”. *Inteligencia en Redes de Comunicación*, Universidad Carlos III de Madrid.
- [24] Bennett, K. P., & Bredensteiner, E. J. (2000). “*Duality and geometry in SVM classifiers*”. *ICML* (pp. 57-64).
- [25] Bailey, D. H., Borwein, J. M., de Prado, M. L., & Zhu, Q. J. (2014). “*Pseudomathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance*”. *Notices of the AMS*, 61(5), 458-471.
- [26] Kittler, J., Hatef, M., Duin, R. P., & Matas, J. (1998). “*On combining classifiers*”. *IEEE transactions on pattern analysis and machine intelligence*, 20(3), 226-239.
-

Parte II

Anexo

Anexo

En esta parte se muestra las tablas con las distancias obtenidas de cada una de las características. De cada prueba se han señalado con rojo las 5 mejores características. La tabla 4.1 se corresponde con la tabla de distancia de las características globales y la tabla 4.2 de las características neuromotoras.

En una tabla 4.3 se muestra toda la información disponible de los sujetos de la base de datos.

Carac global	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 5	Prueba 6	Prueba 7	Prueba 8
1	0,1880	0,0944	0,1795	0,1820	0,1693	0,1339	0,1730	0,1399
2	0,0662	0,0111	0,0888	0,0343	0,0053	0,0042	0,0610	0,0451
3	0,1037	0,0700	0,1242	0,1141	0,2729	0,0001	0,1322	0,1488
4	0,2307	0,0822	0,0515	0,1730	0,0540	0,1170	0,0925	0,0318
5	0,0588	0,1447	0,2704	0,2932	0,2347	0,2499	0,1896	0,2464
6	0,0331	0,1980	0,0281	0,0144	0,0818	0,0653	0,0141	0,1366
7	0,2389	0,0559	0,0650	0,0539	0,1559	0,1231	0,0773	0,0731
8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	0,1916	0,0568	0,1540	0,0301	0,0783	0,0501	0,1122	0,0627
17	0,2610	0,1150	0,1860	0,0211	0,0927	0,0221	0,1191	0,1499
18	0,1209	0,1676	0,2033	0,0789	0,2143	0,1583	0,0274	0,0734
19	0,1534	0,0620	0,1625	0,0633	0,2113	0,2516	0,1113	0,1949
20	0,1770	0,1345	0,1993	0,2632	0,2183	0,2883	0,3601	0,1467
21	0,0640	0,1061	0,0902	0,0802	0,0432	0,1214	0,1255	0,1258
22	0,3688	0,0322	0,0339	0,0139	0,0608	0,2229	0,0820	0,0224
23	0,3671	0,1832	0,0234	0,1143	0,1767	0,1566	0,0307	0,1310

Carac global	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 5	Prueba 6	Prueba 7	Prueba 8
24	0,3067	0,0950	0,1164	0,0998	0,1629	0,1552	0,0323	0,0468
25	0,0895	0,0440	0,0030	0,1163	0,1255	0,1949	0,0798	0,1042
26	0,0586	0,0293	0,1518	0,0767	0,0004	0,0482	0,0358	0,1651
27	0,2230	0,2673	0,1071	0,0649	0,0563	0,0413	0,0878	0,0616
28	0,1887	0,0952	0,1899	0,1998	0,2299	0,1818	0,2447	0,2158
29	0,1838	0,0465	0,1574	0,1729	0,2086	0,1259	0,1886	0,1745
30	0,2174	0,2655	0,1319	0,0472	0,0533	0,0793	0,0928	0,0776
31	0,1864	0,1176	0,0588	0,0142	0,1092	0,1110	0,0517	0,0852
32	0,2332	0,2496	0,0885	0,0861	0,0377	0,0477	0,1021	0,0407
33	0,2437	0,1963	0,0399	0,0560	0,0351	0,0104	0,0089	0,1109
34	0,0083	0,1971	0,0400	0,0316	0,0946	0,0079	0,1171	0,0884
35	0,1134	0,2036	0,0544	0,0341	0,0836	0,0197	0,1144	0,0886
36	0,0574	0,3511	0,1639	0,0316	0,0945	0,0029	0,0411	0,1690
37	0,0112	0,3282	0,1867	0,0277	0,0004	0,0870	0,0210	0,0944
38	0,0152	0,2901	0,3853	0,1055	0,1882	0,2326	0,1637	0,2013
39	0,0452	0,1443	0,0630	0,0171	0,2108	0,0462	0,1530	0,0686
40	0,2682	0,2610	0,3316	0,0493	0,1753	0,1782	0,0162	0,1745
41	0,3099	0,1799	0,0666	0,0176	0,1995	0,1445	0,0435	0,0553
42	0,4095	0,1348	0,1262	0,0231	0,1354	0,0233	0,0569	0,2382
43	0,1031	0,0552	0,0690	0,1420	0,0274	0,0082	0,0127	0,1602
44	0,1253	0,0781	0,0355	0,0873	0,0754	0,0193	0,0802	0,1613
45	0,2667	0,3039	0,2362	0,0083	0,1787	0,0908	0,1406	0,0926
46	0,2370	0,2845	0,2747	0,0109	0,1285	0,1089	0,1230	0,0881
47	0,2864	0,1829	0,1094	0,0129	0,1345	0,1557	0,0652	0,1108
48	0,3555	0,2532	0,2750	0,0219	0,1853	0,0877	0,0900	0,1600
49	0,0906	0,0110	0,1575	0,0432	0,2505	0,0981	0,0852	0,1276
50	0,0038	0,0631	0,2273	0,0514	0,2365	0,1349	0,0990	0,0635
51	0,0540	0,0724	0,0832	0,0699	0,1366	0,0237	0,1856	0,1727
52	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
53	0,3548	0,0735	0,1800	0,1878	0,2280	0,1671	0,2597	0,1749
54	0,2986	0,0328	0,0429	0,1422	0,0595	0,0994	0,2601	0,0275
55	0,0818	0,0922	0,0674	0,0971	0,0678	0,0864	0,0083	0,1053
56	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
57	0,0541	0,2154	0,1251	0,1446	0,0521	0,3391	0,1682	0,1736
58	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
59	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
60	0,0114	0,0424	0,0533	0,1444	0,0273	0,1628	0,0909	0,0786
61	0,0541	0,2154	0,1251	0,1446	0,0521	0,3391	0,1682	0,1736
62	0,1076	0,2472	0,2512	0,2022	0,2639	0,2870	0,2521	0,2626
63	0,0890	0,0745	0,0898	0,0594	0,0711	0,2002	0,1482	0,0828
64	0,0969	0,2034	0,2503	0,2833	0,0932	0,1641	0,1867	0,1567
65	0,1707	0,1137	0,1691	0,2303	0,0674	0,0409	0,0418	0,0909
66	0,2927	0,0631	0,2234	0,1583	0,0949	0,1522	0,1545	0,1918
67	0,1272	0,1722	0,2182	0,1776	0,1752	0,1719	0,2017	0,1825

Carac global	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 5	Prueba 6	Prueba 7	Prueba 8
68	0,2132	0,2350	0,1860	0,1286	0,1322	0,0547	0,0552	0,0072
69	0,2890	0,0895	0,1354	0,2203	0,0837	0,2149	0,0118	0,0091
70	0,1276	0,1034	0,1208	0,0198	0,0960	0,2171	0,0423	0,0180
71	0,1421	0,0956	0,1185	0,0183	0,1013	0,2229	0,0365	0,0116
72	0,1861	0,1521	0,0861	0,0736	0,0342	0,0226	0,0636	0,0759
73	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
74	0,1201	0,0296	0,0627	0,0166	0,1213	0,0390	0,0962	0,0780
75	0,0665	0,0172	0,1174	0,0768	0,1804	0,0042	0,0866	0,1414
76	0,0665	0,0172	0,1174	0,0768	0,1804	0,0042	0,0866	0,1414
77	0,1854	0,0469	0,0208	0,0457	0,0717	0,1271	0,0629	0,0731
78	0,1854	0,0469	0,0208	0,0457	0,0717	0,1271	0,0629	0,0731
79	0,1800	0,0879	0,0622	0,1593	0,0069	0,0208	0,1346	0,1004
80	0,1800	0,0879	0,0622	0,1593	0,0069	0,0208	0,1346	0,1004
81	0,0513	0,0018	0,1988	0,2287	0,1616	0,1648	0,1309	0,1668
82	0,0513	0,0018	0,1988	0,2287	0,1616	0,1648	0,1309	0,1668
83	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
84	0,0893	0,0153	0,0270	0,1061	0,2954	0,1146	0,2007	0,1488
85	0,0951	0,0315	0,0215	0,0982	0,0156	0,1193	0,0083	0,0340
86	0,1593	0,0433	0,1148	0,0750	0,0952	0,0252	0,0363	0,0550
87	0,1311	0,2030	0,2044	0,0766	0,0674	0,1279	0,0728	0,2185
88	0,0704	0,1893	0,1508	0,0238	0,1003	0,0992	0,0816	0,3303
89	0,1261	0,0027	0,1051	0,1135	0,0818	0,0944	0,0450	0,0471
90	0,0154	0,1462	0,0287	0,0488	0,0045	0,0073	0,0201	0,5455
91	0,0970	0,1092	0,0749	0,4052	0,0788	0,1600	0,0410	0,1458
92	0,1375	0,1590	0,0445	0,1834	0,1016	0,1152	0,0556	0,0564
93	0,1078	0,3068	0,2489	0,1261	0,0763	0,1626	0,0005	0,0062
94	0,1575	0,0667	0,3155	0,3654	0,0891	0,1189	0,0197	0,1963
95	0,3903	0,0729	0,1652	0,2434	0,2454	0,2229	0,2587	0,2252
96	0,0121	0,0830	0,1014	0,0697	0,2681	0,1491	0,0952	0,2097
97	0,0589	0,1263	0,0432	0,0362	0,2522	0,2394	0,1338	0,2073
98	0,3881	0,1165	0,1432	0,2008	0,2317	0,1017	0,2062	0,1684
99	0,2419	0,0366	0,0738	0,1369	0,1033	0,0467	0,1892	0,0420
100	0,1664	0,0002	0,0722	0,1242	0,0754	0,0432	0,0071	0,0207

Tabla 4.1: Distancia de características globales.

Carac neuromotora	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 5	Prueba 6	Prueba 7	Prueba 8
1	0,1742	0,5168	0,3098	0,1755	0,1709	0,2624	0,2522	0,4525
2	0,3353	0,1281	0,0412	0,2750	0,1702	0,0250	0,1936	0,0701
3	0,0331	0,0466	0,1375	0,0758	0,1536	0,0255	0,1699	0,0497
4	0,1141	0,2454	0,0362	0,0389	0,0235	0,1568	0,1342	0,1847
5	0,1095	0,1854	0,0321	0,1523	0,1173	0,2075	0,1219	0,1308
6	0,2138	0,1818	0,1021	0,1165	0,0479	0,0502	0,0795	0,0069

Carac neuromotora	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 5	Prueba 6	Prueba 7	Prueba 8
7	0,0613	0,2542	0,0650	0,0814	0,1830	0,1645	0,1761	0,2005
8	0,0345	0,4557	0,3376	0,1337	0,1655	0,1787	0,2718	0,4662
9	0,0269	0,4768	0,3406	0,1337	0,1678	0,1857	0,2677	0,4701
10	0,0219	0,4845	0,3392	0,1324	0,1682	0,1901	0,2637	0,4707
11	0,0695	0,0633	0,1796	0,0725	0,0932	0,0837	0,0087	0,2278
12	0,0454	0,0706	0,1680	0,0524	0,1018	0,0693	0,0625	0,1683
13	0,2367	0,1570	0,1518	0,1398	0,0459	0,0233	0,0026	0,0884
14	0,0004	0,0795	0,1658	0,0490	0,1019	0,0693	0,0642	0,1667
15	0,2175	0,1518	0,1229	0,0881	0,0748	0,0129	0,0584	0,0284
16	0,0640	0,1775	0,1287	0,0491	0,0891	0,1144	0,0398	0,0830
17	0,0758	0,1812	0,1700	0,0517	0,0885	0,1151	0,0331	0,0854
18	0,0792	0,1852	0,1739	0,0471	0,0893	0,1157	0,0317	0,0851
19	0,1507	0,4229	0,2314	0,0539	0,1249	0,3346	0,1555	0,2339
20	0,1518	0,0658	0,2153	0,0077	0,1588	0,2007	0,1049	0,0531
21	0,1314	0,0013	0,0421	0,0470	0,1673	0,0937	0,1599	0,0635
22	0,1088	0,2897	0,4238	0,2910	0,1650	0,1997	0,2427	0,0610
23	0,0009	0,2149	0,3702	0,2951	0,0055	0,2333	0,2475	0,2746
24	0,0534	0,3654	0,2589	0,0996	0,1142	0,1767	0,0733	0,0629
25	0,0624	0,2642	0,2701	0,0887	0,0504	0,2192	0,0897	0,1975
26	0,0051	0,3819	0,4160	0,2131	0,0936	0,1001	0,2107	0,3858
27	0,0180	0,4106	0,3732	0,2191	0,1424	0,1021	0,1974	0,4128
28	0,0296	0,4299	0,3498	0,2156	0,1655	0,1048	0,1859	0,4290
29	0,2272	0,1340	0,1623	0,0383	0,0993	0,0451	0,0175	0,1817
30	0,2228	0,1322	0,1569	0,0358	0,1029	0,0441	0,0276	0,1773
31	0,1534	0,2741	0,1096	0,0375	0,0568	0,1293	0,1029	0,0788
32	0,2252	0,1342	0,1568	0,0359	0,1035	0,0438	0,0273	0,1800
33	0,1004	0,2591	0,0016	0,0086	0,0909	0,0747	0,1057	0,1557
34	0,0135	0,2483	0,1459	0,1662	0,1172	0,1403	0,0904	0,0155
35	0,0035	0,2633	0,1157	0,1186	0,0923	0,1361	0,0713	0,0297
36	0,0034	0,2609	0,1191	0,1103	0,0943	0,1412	0,0694	0,0295
37	0,0558	0,1709	0,2759	0,1138	0,0519	0,0124	0,0575	0,3971
38	0,1517	0,0135	0,0556	0,0877	0,2214	0,1405	0,1433	0,0513
39	0,0840	0,1455	0,2360	0,0262	0,1709	0,1084	0,1726	0,0274
40	0,0776	0,2173	0,3295	0,1351	0,1605	0,2786	0,1928	0,1298
41	0,0657	0,1326	0,1078	0,0912	0,1504	0,0909	0,0126	0,1613
42	0,0664	0,1851	0,2506	0,1903	0,1666	0,3429	0,1867	0,1545
43	0,0617	0,1724	0,2118	0,1365	0,1417	0,1353	0,0319	0,1521
44	0,0425	0,3432	0,3331	0,1609	0,2079	0,0561	0,2252	0,3448
45	0,0353	0,4204	0,3724	0,2455	0,1441	0,1227	0,1871	0,4329
46	0,0418	0,4278	0,3490	0,2712	0,1259	0,1495	0,1826	0,4472
47	0,0629	0,0844	0,0249	0,2492	0,0644	0,0722	0,0519	0,1229
48	0,0506	0,1046	0,0029	0,0617	0,1564	0,1002	0,0943	0,0276
49	0,1175	0,0254	0,0065	0,2735	0,0391	0,1871	0,1049	0,0925
50	0,0695	0,1277	0,0084	0,0347	0,1676	0,0822	0,0794	0,0166

Carac neuromotora	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 5	Prueba 6	Prueba 7	Prueba 8
51	0,0646	0,0823	0,0016	0,1605	0,1266	0,1290	0,0688	0,0629
52	0,1089	0,1566	0,0830	0,1636	0,0439	0,1941	0,0383	0,1097
53	0,1194	0,1703	0,0594	0,1476	0,0178	0,1859	0,0295	0,0873
54	0,1171	0,1702	0,0621	0,1420	0,0187	0,1866	0,0303	0,0920
55	0,1002	0,4048	0,2337	0,0058	0,1012	0,3620	0,0787	0,2167
56	0,1193	0,0120	0,0979	0,1816	0,1140	0,2273	0,0777	0,0386
57	0,0957	0,1073	0,0039	0,0642	0,1179	0,1179	0,2270	0,1709
58	0,0727	0,0050	0,2470	0,2221	0,1248	0,1400	0,3077	0,1005
59	0,0608	0,1043	0,1428	0,2086	0,1596	0,1132	0,0881	0,1653
60	0,0673	0,2350	0,2633	0,2999	0,1870	0,0685	0,1466	0,1485
61	0,0657	0,1059	0,1996	0,0758	0,1531	0,0362	0,0657	0,1621
62	0,0532	0,3588	0,3616	0,1686	0,0427	0,1381	0,1893	0,3970
63	0,0384	0,3738	0,3387	0,1317	0,1216	0,1280	0,1753	0,4619
64	0,0164	0,3676	0,3122	0,1042	0,1588	0,1102	0,1578	0,4878
65	0,1932	0,1300	0,1551	0,0442	0,0502	0,1815	0,0575	0,1859
66	0,1929	0,1284	0,1530	0,0415	0,0537	0,1782	0,0672	0,1797
67	0,0522	0,2043	0,0061	0,0352	0,0699	0,1994	0,0976	0,0681
68	0,1936	0,1299	0,1533	0,0414	0,0540	0,1775	0,0678	0,1892
69	0,0706	0,2180	0,1353	0,0128	0,0819	0,1740	0,1051	0,1232
70	0,0667	0,2533	0,0244	0,0584	0,0789	0,1157	0,1890	0,0600
71	0,0799	0,2929	0,0127	0,0474	0,0985	0,0838	0,1310	0,0378
72	0,1162	0,2954	0,0161	0,0708	0,0757	0,0846	0,1254	0,0042
73	0,1509	0,5254	0,3867	0,1905	0,1566	0,2710	0,2594	0,4929
74	0,3652	0,1799	0,0005	0,2809	0,1114	0,0352	0,2163	0,1318
75	0,0137	0,0212	0,2007	0,0245	0,1187	0,0209	0,1578	0,0667
76	0,0590	0,1415	0,1079	0,0104	0,0433	0,1607	0,1479	0,2927
77	0,0317	0,2119	0,1276	0,1216	0,0975	0,1073	0,0427	0,0078
78	0,1140	0,0896	0,0509	0,1561	0,0680	0,0113	0,0904	0,0177
79	0,0163	0,3441	0,0967	0,0261	0,0345	0,0273	0,1869	0,1046
80	0,0842	0,3635	0,2439	0,0829	0,1973	0,2227	0,3213	0,4747
81	0,0855	0,3749	0,2400	0,0709	0,1856	0,2370	0,3072	0,4655
82	0,0887	0,3529	0,2471	0,0683	0,1725	0,2302	0,2999	0,4584
83	0,2184	0,1511	0,1965	0,1831	0,0382	0,1029	0,1120	0,1788
84	0,2509	0,1050	0,2245	0,1195	0,0301	0,0517	0,0037	0,0769
85	0,2579	0,0290	0,0745	0,2250	0,0271	0,0587	0,0548	0,1415
86	0,2348	0,0746	0,2054	0,1137	0,0347	0,0555	0,0204	0,0411
87	0,2616	0,0029	0,1733	0,1649	0,0239	0,0236	0,0207	0,1030
88	0,2590	0,2345	0,1924	0,0392	0,0317	0,0405	0,0633	0,0311
89	0,2123	0,2468	0,2629	0,1978	0,0082	0,0320	0,0230	0,0588
90	0,1886	0,2692	0,2446	0,2117	0,0018	0,0522	0,0093	0,0543
91	0,2424	0,3891	0,4504	0,2063	0,1147	0,1333	0,1534	0,1548

Tabla 4.2: Distancia de características neuromotas.

ID	Nationality	Sex	Disease	PD status	Age	Dominant hand	LED	UPDRS III	Length of PD
00001	Czech	F	PD	ON	68	R	1115	2	6
00002	Czech	F	PD	ON	78	R	2110	2	8
00003	Czech	F	PD	ON	69	R	1556,6	2	7
00004	Czech	F	PD	ON	79	R	1691	2	12
00005	Czech	F	PD	ON	69	R	600	2	2
00006	Czech	F	PD	ON	57	R	1271,66	2	9
00007	Czech	F	PD	ON	78	R	666,66	3	19
00008	Czech	F	PD	ON	58	R	397,5	1	5
00009	Czech	M	PD	ON	78	R	2066	1	3
00010	Czech	M	PD	ON	74	R	1480	2,5	3
00013	Czech	M	PD	ON	65	R	990	1	2
00014	Czech	M	PD	ON	64	R	1253,33	3	8
00015	Czech	F	PD	ON	69	R	990	2,5	17
00016	Czech	M	PD	ON	67	R	1187,5	2	4
00017	Czech	F	PD	ON	75	R	1370	5	18
00018	Czech	F	PD	ON	76	R	1250	2,5	17
00019	Czech	F	PD	ON	86	R	750	2	6
00020	Czech	F	PD	ON	79	R	2226,66	2	8
00022	Czech	F	PD	ON	67	R	645	2	14
00023	Czech	F	PD	ON	73	R	1235	2	9
00024	Czech	M	PD	ON	70	R	1316,66	4	7
00025	Czech	M	PD	ON	60	R	1143	3	10
00026	Czech	F	H		57	R			
00027	Czech	M	H		92	R			
00028	Czech	F	H		52	R			
00029	Czech	F	H		58	R			
00030	Czech	M	H		69	R			
00031	Czech	M	H		76	R			
00032	Czech	F	H		59	R			
00033	Czech	F	PD	ON	62	R	750	2	4
00034	Czech	M	PD	ON	61	R	2546,66	2	5
00036	Czech	M	PD	ON	90	R	750	2	3
00039	Czech	M	H		65	R			
00040	Czech	M	H		53	R			
00041	Czech	M	H		78	R			
00043	Czech	M	PD	ON	48	R	1080	1	4
00044	Czech	F	PD	ON	62	R	397,5	1	5
00048	Czech	M	PD	ON	87	R	1450	4	12
00049	Czech	M	H		58	R			
00051	Czech	F	H		48	R			
00052	Czech	F	H		44	R			
00053	Czech	M	PD	ON	84	R	1941,66	2	2
00054	Czech	M	PD	ON	69	R	2546	2	10
00055	Czech	M	PD	ON	63	R	1930	2,5	14

ID	Nationality	Sex	Disease	PD status	Age	Dominant hand	LED	UPDRS III	Length of PD
00057	Czech	M	H		80	R			
00060	Czech	M	H		65	R			
00061	Czech	F	H		59	R			
00062	Czech	F	H		65	R			
00066	Czech	F	H		69	R			
00067	Czech	M	H		59	R			
00069	Czech	F	H		74	R			
00070	Czech	F	H		47	R			
00071	Czech	M	H		52	R			
00072	Czech	M	H		45	R			
00073	Czech	F	H		64	R			
00074	Czech	M	PD	ON	53	R	2386,66	2,5	9
00075	Czech	M	PD	ON	73	R	2010	2,5	12
00076	Czech	M	H		56	R			
00077	Czech	M	PD	ON	74	R	2336,66	3	13
00078	Czech	M	PD	ON	36	R	800	2	12
00080	Czech	M	PD	ON	67	R	3544,16	3	5
00082	Czech	M	H		45	R			
00083	Czech	F	H		74	R			
00084	Czech	F	H		62	R			
00085	Czech	F	H		75	R			
00087	Czech	M	H		57	R			
00089	Czech	M	H		63	R			
00090	Czech	M	H		71	R			
00091	Czech	F	H		64	R			
00092	Czech	F	H		58	R			
00094	Czech	M	H		64	R			
00095	Czech	M	H		74	R			
00096	Czech	F	H		77	R			
00097	Czech	M	H		44	R			
00098	Czech	F	PD	ON	77	R	1210	2	6

Tabla 4.3: Información disponible de todos los sujetos.

