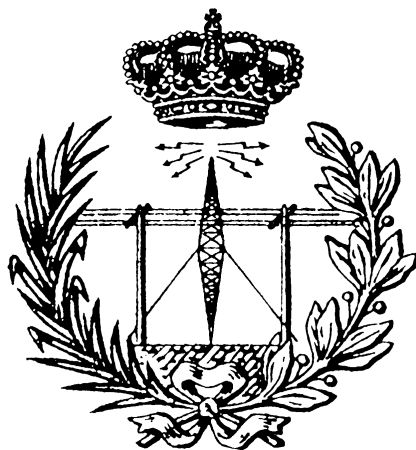


ESCUELA DE INGENIERIA DE TELECOMUNICACIÓN Y ELECTRÓNICA



TRABAJO FIN DE GRADO

SISTEMA DE IDENTIFICACIÓN DE PERSONAS MEDIANTE SENSORES BIOMÉDICOS

Titulación: Grado en Ingeniería en Tecnologías de la
Telecomunicación

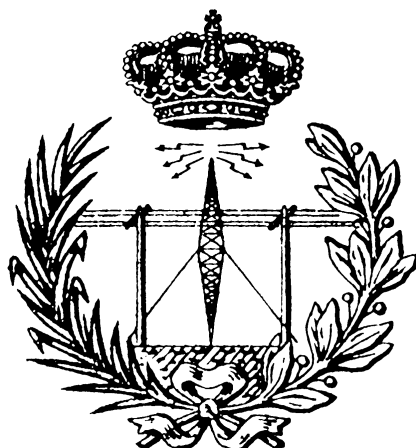
Mención: Sistemas de Telecomunicación

Autor: Alain David Díaz Alonso

Tutores: Dr. D. Carlos Manuel Travieso González
Dr. D. Jesús Bernardino Alonso Hernández

Fecha: Julio de 2016

ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA



SISTEMA DE IDENTIFICACIÓN DE PERSONAS MEDIANTE SENSORES BIOMÉDICOS

HOJA DE EVALUACIÓN

Calificación: _____

Presidente

Fdo.:

Vocal

Fdo.:

Secretario/a

Fdo.:

Fecha: Junio de 2016

“Me gustaría agradecer, en primer lugar, a todas aquellas personas que han participado libremente en la toma de muestras. A mi padre, por haberme educado siempre a través del ejemplo, así como a mi familia y amigos, por su incansable apoyo a lo largo de estos años. A Argentina, por ofrecerme su ayuda en todo momento. Por último, a mis tutores, que han contribuido en gran medida al desarrollo de este Trabajo

Fin de Grado.

A todos ellos, Gracias”

ÍNDICE DE CONTENIDOS

I. MEMORIA	1
Capítulo 1. Introducción.....	3
1.1 Definición e historia de los sistemas de identificación	3
1.2 Motivación y problemática	5
1.3 Estado del arte	6
1.4 Objetivos.....	8
1.5 Estructura del documento.....	9
Capítulo 2. Construcción de la base de datos.....	11
2.1 Descripción de dispositivos utilizados.....	11
2.1.1 Placa Arduino.....	11
2.1.2 Plataforma e-Health.....	12
2.1.3 Sensor de flujo de aire.....	13
2.1.4 Sensor de respuesta galvánica de la piel	14
2.1.5 Sensor de electrocardiograma	14
2.1.6 Sensor de temperatura.....	15
2.1.7 Sensor de electromiograma	15
2.1.8 Sensor de pulso y oxígeno en sangre	15
2.2 Consideraciones de diseño de la base de datos.....	16
2.3 Características de las señales a extraer	16
2.3.1 Flujo de aire	17
2.3.2 Respuesta Galvánica de la Piel	17
2.3.3 Electrocardiograma	18
2.3.4 Temperatura	18
2.3.5 Electromiograma.....	19
2.3.6 Pulso y oxígeno en sangre.....	19
2.4 Proceso de toma de muestras.....	19
Capítulo 3. Parametrización y sistema de clasificación.....	25
3.1 Introducción a Máquinas de Vectores de Soporte.....	25
3.2 Descripción de parámetros útiles	27
3.3 Introducción al Análisis de Componentes Principales	29
3.4 Desarrollo del sistema de clasificación	31
Capítulo 4. Experimentos y resultados.....	33
4.1 Metodología experimental	33
4.2 Resultados de la fase inicial.....	34
4.3 Modificaciones del sistema.....	37
4.4 Resultados de la fase final	40
4.5 Análisis de los resultados	43
Capítulo 5. Conclusiones y líneas futuras	45
5.1 Conclusiones.....	45
5.2 Líneas futuras.....	46
ANEXOS	49
BIBLIOGRAFÍA	59
II. PLIEGO DE CONDICIONES	61
III. PRESUPUESTO	65
P1. Desglose del Presupuesto.....	68
P2. Recursos materiales.....	68
P3. Trabajo tarifado por tiempo empleado.....	70
P4. Costes de redacción del Trabajo Fin de Grado.....	72
P5. Material fungible	73

P6. Derechos de visado del COIT	73
P7. Gastos de tramitación y envío.....	74
P8. Aplicación de impuestos.....	74

ÍNDICE DE FIGURAS

Figura 1-1. Tareas a realizar para el cumplimiento de los objetivos.....	9
Figura 2-1. Arduino UNO Revisión 3	12
Figura 2-2. Plataforma e-Health.....	12
Figura 2-3. Sensores biométricos y plataforma e-Health.....	13
Figura 2-4. Sensor de flujo de aire.....	13
Figura 2-5. Sensor GSR.....	14
Figura 2-6. Sensor ECG conectado a e-Health.....	14
Figura 2-7. Sensor de temperatura	15
Figura 2-8. Pulsioxímetro	15
Figura 2-9. Estructura de la base de datos.....	16
Figura 2-10. Medidas de flujo de aire.....	17
Figura 2-11. Fases de un ECG sano	18
Figura 2-12. Grupos de señales por sujeto.....	19
Figura 2-13. Posición del sensor de flujo de aire.....	20
Figura 2-14. Posición de sensor GSR.....	20
Figura 2-15. Posición de los electrodos	21
Figura 2-16. Posición de sensor temperatura	21
Figura 2-17. Posición del sensor de EMG	22
Figura 2-18. Señal EMG de 30 segundos.....	22
Figura 2-19. Posición del pulsioxímetro	23
Figura 2-20. Señales del pulsioxímetro.....	23
Figura 3-1. Hiperplanos de separación	26
Figura 3-2. Posibles conjuntos de datos.....	26
Figura 3-3. Solución de función kernel.....	27
Figura 3-4. Transformación de coordenadas.....	30
Figura 4-1. Resultados de señales individuales	34
Figura 4-2. Comparación de combinaciones y EMG.....	37

ÍNDICE DE TABLAS

Tabla 4-1. Resultados de señales individuales	34
Tabla 4-2. Resultados de combinaciones de señales	36
Tabla 4-3. Resultados de parámetros de señales	38
Tabla 4-4. Resultados de combinación de parámetros	39
Tabla 4-5. PCA en combinación 4	40
Tabla 4-6. PCA en combinación 5	41
Tabla 4-7. PCA en combinación 6	41
Tabla 4-8. Combinación de 12 elementos	42
Tabla 4-9. PCA en combinación de 12 elementos	42
Tabla 5-1. Comparación con el estado del arte	46
Tabla P.1. Costes de herramientas hardware	69
Tabla P.2. Costes de herramientas software	70
Tabla P.3. Factor de corrección en función del número de horas invertidas	71
Tabla P.4. Costes de ejecución material	72
Tabla P.5. Costes del material fungible.....	73
Tabla P.6. Costes totales del TFG	75

LISTADO DE ACRÓNIMOS

BD = Base de datos.

ECG = Electrocardiograma.

EEG = Electroencefalograma.

EITE = Escuela de Ingeniería de Telecomunicación y Electrónica.

EMG = Electromiograma.

GSR = Galvanic Skin Response (Respuesta Galvánica de la Piel).

GB = Gigabyte.

IDE = Integrated Development Environment (Ambiente de desarrollo integrado).

MG = Megabyte.

MLP = Multilayer Perceptron (Perceptrón Multicapa)

PCA = Principal Component Analysis (Análisis de componentes principales).

RAM = Random Access Memory (Memoria de Acceso Aleatorio).

RBF = Radial Basis Function (Función de Base Radial)

SRM = Structural Risk Minimization (Minimización del riesgo estructural).

SVM = Support Vector Machines (Máquinas de Vectores de Soporte).

TFG = Trabajo de Fin de Grado.

I. MEMORIA

Capítulo 1. Introducción

En este capítulo se llevará a cabo una explicación introductoria de los sistemas de identificación, su definición, su historia y sus diversas aplicaciones. Se realizará una explicación de la motivación y problemática que han dado lugar a este estudio, continuando con el estado del arte dentro de los sistemas biométricos de identificación. Se abordarán distintas propuestas para la resolución del problema y se definirán los objetivos a conseguir. Por último se establecerá la estructura de la que dispondrá el presente documento.

1.1 Definición e historia de los sistemas de identificación

Desde el origen de la sociedad, la correcta identificación de individuos ha contado con un papel importante ligado estrechamente a la seguridad. Para el correcto funcionamiento de las leyes y la sociedad cívica era necesario castigar y perseguir a aquellos que rompían la ley o intentaban burlarla. Lamentablemente, debido al carácter primitivo de aquellos tiempos y de sus medios de comunicación, el único medio de identificación utilizado era la asociación de un nombre a la memoria visual, el cual podía burlarse con gran facilidad. Debido a la debilidad de este sistema, se aplicaban penas de gran brutalidad con la intención de disuadir a aquellos que quisieran romper las leyes.

Junto con los brutales castigos surgió la marca de fuego, también llamado tatuaje judicial, que además de formar parte de la pena, cumplía la función de identificar al delincuente.

Así, y de diversas formas, este método fue utilizado para proteger a la sociedad de aquellos que no querían acatar sus normas. En Grecia y Roma, marcaban la frente con hierros candentes o tatuaban representaciones de animales, armas o letras de tal forma que fuese imposible su disimulo. En Francia, se empezó marcando en la frente una flor de lis para reconocer al delincuente, más adelante se sustituyó por otras marcas, algunos ejemplos son la letra V para los ladrones, F para los falsificadores o W para ladrones reincidentes. En la India, se realizaba en la frente de los delincuentes una

marca distinta según el delito cometido y además, se marginaba al delincuente de tal forma que incluso eran abandonados por sus parientes.

Con el paso del tiempo y el avance de la ciencia se desarrollaron métodos más complejos de identificación. En los inicios de la "policía científica" se realizaban listados de las marcas y cicatrices, se tenían en cuenta las medidas de partes características del cuerpo y, llegado el momento, se empezó a aplicar la fotografía judicial (la cual ha prevalecido hasta nuestros días). El interés por la antropometría asociada a la identificación dio lugar a diversos y novedosos métodos, entre los más importantes podríamos destacar los siguientes:

- Método dentario de Amoedo: por el cual se trataba de identificar a una persona por su impresión dental.
- Craneográfico de Alfonso: mediante un aparato llamado craneógrafo, se tomaba una medida del perfil craneal, desde la raíz de la nariz hasta la cresta occipital.
- Dactiloscopia: quizás la más conocida y usada de las anteriormente nombradas, la dactiloscopia se basa en la identificación humana mediante las huellas digitales.

Diversos métodos y técnicas de reconocimiento han sido desarrollados a lo largo de los años (especialmente en el ámbito policial), pero la mayoría no eran fiables en muchos casos y podían ser burlados en muchos otros con cierta facilidad, ya sea mediante el uso de la ciencia, debido a un accidente, una enfermedad o la vejez [1].

Actualmente vivimos en una sociedad basada en la información y la tecnología, la cual ha dado lugar a métodos de identificación destinados a nuevos tipos de servicios. Muchos de estos métodos están basados en biometría, considerada actualmente como el método ideal de identificación humana, que consiste en la medición e identificación automática mediante técnicas matemáticas y estadísticas de alguna característica propia de una persona [2]. De esta forma, ejemplos de sistemas de identificación biométrica serían aquellos basados en: análisis de voz, escáner de retina o geometría de la mano [3].

1.2 Motivación y problemática

Como se ha mencionado anteriormente, el ámbito policial ha sido la principal fuente de desarrollo de los sistemas de identificación, dada la importancia que alberga esta tarea dentro de las funciones policiales. No obstante, existen otras áreas donde la aplicación de una correcta identificación supone gran importancia.

Los hospitales son lugares que cuentan habitualmente con una gran cantidad de personas, tanto pacientes como empleados, en los que se producen un considerable número de situaciones en las que la identificación juega un papel delicado. En estos centros la identificación suele realizarse mediante pulseras y listados con la información del paciente, dichos elementos pueden ser fácilmente extraviados, falsificados ó confundidos de paciente. Existen infinidad de antecedentes sobre incidentes por confusión de pacientes, denominados también de "paciente equivocado", estos incidentes provocan errores con consecuencias de distinta magnitud, desde errores leves, como puede ser suministrar un medicamento equivocado, a otros más graves como puede ser suministrar medicamentos que además estén contraindicados, realizar diagnósticos equivocados, transfusiones de sangre erróneas, etc., empeorando así la salud del paciente [4]. En casos más extremos, se ha llegado a realizar cirugías a pacientes equivocados, este tipo de errores son menos habituales que los anteriormente mencionados, pero suceden con inquietante frecuencia [5].

Los problemas anteriores, derivados de una mala identificación, junto con otros como el tratamiento de pacientes con problemas de salud mental o la gestión de pacientes en hospitales de gran superficie, hacen evidente la necesidad de un sistema de identificación fiable y automatizado, que elimine la dependencia de pulseras y listados.

Además, el campo de la identificación biométrica destinada a entornos hospitalarios no está tan explotado como en otros sectores, por lo que se trata de un campo con un desarrollo potencial.

1.3 Estado del arte

En los últimos años se han llevado a cabo diversos trabajos y propuestas destinados a solventar este problema, a continuación se citan algunos de estos trabajos y en que se basan:

Código de barras ligado a base de datos: esta propuesta plantea el uso de códigos de barras en las pulseras convencionales, dichos códigos de barras contendrían información codificada sobre el paciente. Además los códigos de barras podrían usarse conjuntamente con una base de datos, donde la información codificada serviría como acceso a información más detallada del paciente [6]. Esta propuesta cuenta con serias limitaciones, por un lado el sistema resulta ilegible a primera vista, siendo necesario el uso de lectores, que supondrían un coste importante al tener que abarcar la totalidad del establecimiento, también presentaría problemas el uso en neonatos, donde el tamaño y curvatura de la pulsera pueden ser demasiado pequeñas para permitir la lectura del código de barras. Por último, este sistema, si bien mejora la identificación del paciente, no soluciona los problemas de las pulseras convencionales, pudiéndose extraviar, falsificar y confundir de paciente con relativa facilidad.

Identificación por radiofrecuencia: este tipo de identificación se basa en la transferencia de datos entre un lector y un dispositivo denominado "etiqueta" por medio de radiofrecuencias. La etiqueta no es más que un transpondedor que emitirá una señal fija. Existen dos tipos de etiquetas, por un lado están las activas, que cuentan con su propia fuente de alimentación, son más fiables y son capaces de transmitir señales más potentes, por otro lado están las pasivas, las cuales poseen un rango de actuación inferior a las activas pero cuentan con la ventaja de no depender de una fuente de alimentación interna, además su información almacenada no puede ser modificada, a diferencia de las etiquetas activas. La propuesta ofrecida mediante este sistema, es el de insertar estas etiquetas bajo la piel del paciente, asociando de esta forma la identidad del paciente al de la señal transmitida por la etiqueta [7]. Esta propuesta cuenta con diversos inconvenientes, sus costes son superiores al de los códigos de barras, además se trata de un método invasivo y su reutilización conlleva una gran dificultad, puesto que los dispositivos deberían ser extraídos, higienizados y reasignados a los nuevos pacientes, habiéndose antes desasociado del paciente anterior en la base de datos.

Existen otros trabajos y estudios basados en la identificación mediante biometría, que si bien no fueron diseñados para este propósito concreto, sí podrían adaptarse adecuadamente ante una problemática de estas características. A continuación se citan los más destacados:

En el artículo publicado en 2006 denominado *EEG Based Biometric Framework for Automatic Identity Verification* [8], se propone el uso del encefalograma (EEG) como parámetro biométrico para la identificación automática del individuo. Para ello, se registraba las señales de alta frecuencia procedentes del encefalograma obtenido mientras los sujetos observaban dibujos en blanco y negro de objetos comunes. Los resultados obtenidos fueron prometedores, el sistema logró un porcentaje de éxito de $98.56 \pm 1.87\%$ en la identificación de 40 sujetos.

En *Biometric Human Identification Based on ECG* [9], realizado en 2005, se demuestra como el electrocardiograma (ECG) de una persona depende de las características anatómicas del cuerpo y el corazón del individuo, pudiendo ser usado como elemento caracterizador de la persona. En este estudio, se tomaron registros de 90 voluntarios y se desarrolló con éxito un sistema de identificación con un porcentaje de éxito del 96%.

Los resultados obtenidos por los dos trabajos anteriores son bastante prometedores y demuestran que elementos propios de un hospital pueden ser usados como medio para obtener señales útiles en la identificación de personas. Lamentablemente, estos sistemas también cuentan con inconvenientes, dado que la maquinaria necesaria para su funcionamiento suele ser cara, necesita de un cuidadoso mantenimiento y es difícil de manejar, necesitando personal especializado.

Por último, en el artículo *Biometric Approach Based on Physiological Human Signals* [10], desarrollado en la ULPGC, se hace uso de una combinación de tres señales fisiológicas: electrocardiograma, electromiograma (EMG) y respuesta galvánica de la piel (GSR, por sus siglas en inglés). Mediante la combinación de estas señales y el uso de *redes neuronales* como elemento clasificador, se consiguió un porcentaje de éxito del 85.55%, valor destacable, dado que las señales utilizadas fueron recogidas mediante sensores de bajo coste y calidad, usándose únicamente 3 electrodos para capturar las

señales de EMG y ECG, mientras que en los trabajos anteriores se hace uso de equipos médicos profesionales, los cuales son más costosos y emplean un mayor número de electrodos.

1.4 Objetivos

El objetivo de este trabajo es diseñar e implementar un sistema automático de identificación de personas a partir de señales biométricas, que además solvante los problemas encontrados en los sistemas del apartado 1.3, por lo tanto, el sistema debe tener un porcentaje de éxito razonable, ser de bajo coste, no necesitar personal especializado, no contar con métodos invasivos y ser invulnerable a falsificaciones o pérdidas. Para ello, se creará una base de datos (BD) a partir de las señales de 25 sujetos. Para cada uno de estos 25 sujetos se extraerá dos grupos de señales, uno de estos grupos, al que llamaremos *grupo de referencia*, será usado para definir los sujetos ante el sistema, el otro grupo, al que llamaremos *grupo de test*, será usado para determinar la eficacia del sistema. Por tanto, la tasa de éxito del sistema de identificación vendrá dada por la cantidad de grupos de test que puedan ser relacionados adecuadamente con su respectivo grupo de referencia.

Una vez creada la base de datos, se seleccionará un clasificador con el que se estudiará cada señal extraída, con el propósito de revelar la eficacia de cada señal a la hora de definir al sujeto. Conocida la eficacia de cada señal, se realizarán diversas combinaciones de las mismas, buscando aquella que obtenga mejores resultados. Además, se tratará de mejorar el sistema, mediante la extracción de parámetros útiles, pertenecientes a dichas señales. Finalmente, se tratará de mejorar aún más el resultado, aplicando *Análisis de Componentes Principales* (PCA).

Para llevar a cabo la realización de los objetivos se sigue el siguiente diagrama de bloques:



Figura 1-1. Tareas a realizar para el cumplimiento de los objetivos

1.5 Estructura del documento

El documento se encuentra dividido en tres bloques: Memoria, Pliego de Condiciones y Presupuesto.

El bloque Memoria, consta a su vez de 5 capítulos en los que se tratan diferentes aspectos acerca del presente Trabajo Fin de Grado:

Capítulo 1. Introducción.

En este capítulo se inicia al lector en una breve explicación de los sistemas de identificación, su historia, y la importancia que ha tenido a lo largo de los años. También se explicará las causas motivacionales que han dado lugar a la realización de este trabajo, así como los antecedentes más destacados al respecto. Finalmente, se resumirá la estructura que seguirá el presente trabajo.

Capítulo 2. Construcción de la base de datos

Esta capítulo comienza con la descripción de los elementos utilizados en la realización de la toma de muestras. Además se expondrán las consideraciones establecidas para el diseño de la base de datos y se explicará las características de los parámetros a extraer. Para terminar se describirá con detalle el procedimiento realizado para la toma de muestras.

Capítulo 3. Parametrización y sistema de clasificación

La función de este capítulo será la descripción de todos aquellos parámetros útiles que pueden ser extraídos de las señales previamente obtenidas en la toma de muestras. Se continuará con breves introducciones a las *Máquinas de Vectores de Soporte* (SVM, por sus siglas en inglés) y al *Análisis de Componentes Principales*. Para terminar se explicará de que forma se ha desarrollado el sistema de clasificación.

Capítulo 4. Experimentos y resultados

En este capítulo se extraerán los resultados finales. En primer lugar se obtendrán resultados para señales sin procesar, con el fin de conocer las capacidades de cada una. Posteriormente se obtendrán los resultados de diferentes combinaciones de señales, así como de la aplicación de parametrización y PCA.

Capítulo 5. Conclusiones y líneas futuras

En este capítulo se evaluará cuán eficaz es el sistema diseñado en este trabajo y hasta que punto se ha conseguido el fin por el cual fue planteado, exponiendo una comparación con los sistemas existentes y anteriormente descritos. Finalmente, se añade un apartado con algunas sugerencias de las posibles líneas futuras de actuación que podrían complementar este trabajo.

Como parte de la memoria también se incluye como anexo un artículo escrito en relación a este Trabajo Fin de Grado y la bibliografía consultada para la realización del conjunto de tareas. En el bloque referente al presupuesto se muestra el coste de realización del presente trabajo y previamente se encuentra el pliego de condiciones.

Capítulo 2. Construcción de la base de datos

En este capítulo se abordarán las plataformas y sensores utilizados para la extracción de las señales, las cuales serán usadas en la posterior construcción de la base de datos. Además se presentarán las condiciones de diseño con las que debe contar la base de datos y se expondrán las características de las señales a extraer. Posteriormente se describirá de forma detallada el proceso utilizado para capturar cada una de las señales.

2.1 Descripción de dispositivos utilizados

Para llevar a cabo la extracción de señales, se ha hecho uso de una plataforma *e-Health* acoplada a una placa *Arduino*. A continuación, se describirá en detalle cada uno de los elementos que forman este sistema.

2.1.1 Placa Arduino

Arduino es una plataforma de hardware libre, basada en una placa de bajo coste con un microcontrolador y un entorno de desarrollo (IDE, por sus siglas en inglés), diseñado para facilitar el uso de la tecnología en proyectos multidisciplinarios [11].

Su hardware, consistente en una placa de circuito impreso, puede intercambiar información con el exterior mediante sus puertos analógicos y digitales de entrada/salida. De esta forma, Arduino permite conectar extensiones (también llamadas *shields*) a sus puertos, modificando o ampliando sus características. Además, Arduino cuenta con su propio entorno de desarrollo, el cual está basado en el entorno *Processing* y un lenguaje de programación basado en *Wiring*. El microcontrolador de la placa puede ser programado por un ordenador mediante comunicación serie.

En este trabajo se ha empleado un modelo *Arduino UNO Revisión 3* (ver figura 2-1).

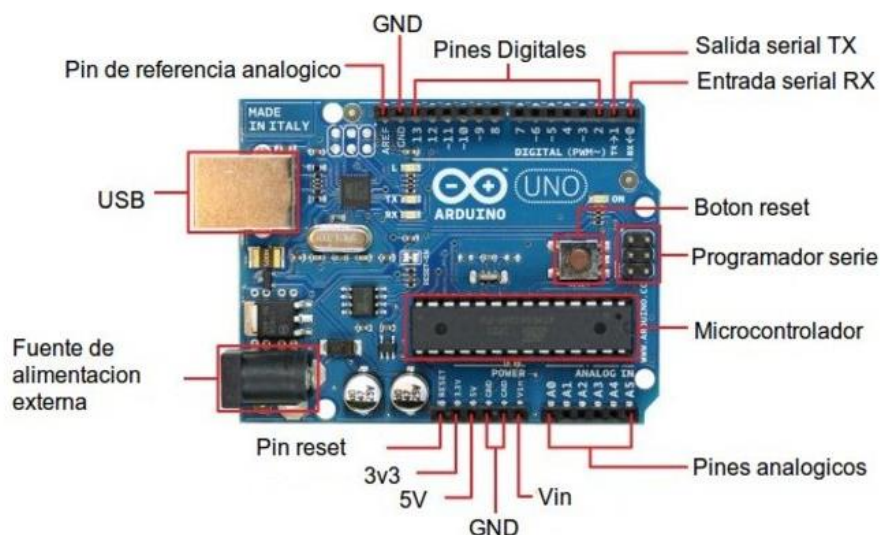


Figura 2-1. Arduino UNO Revisión 3

2.1.2 Plataforma e-Health

La plataforma e-Health, mostrada en la *figura 2-2*, es un dispositivo desarrollado por la empresa *Libelium*, el cual sirve como extensión para Arduino y que permite a los usuarios el desarrollo de aplicaciones relacionadas con la medicina y la biometría [12].

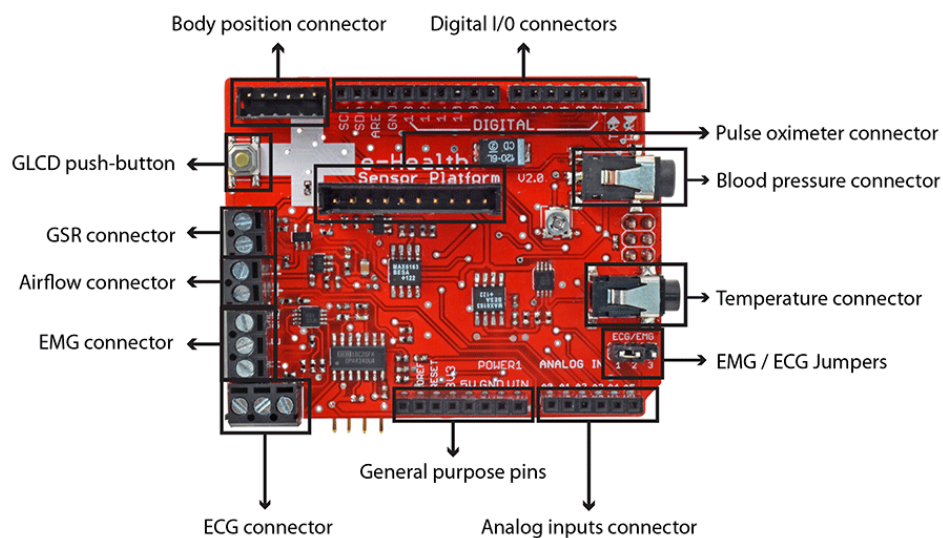


Figura 2-2. Plataforma e-Health

Esta plataforma cuenta con 9 sensores que permiten extraer diversas señales del cuerpo humano, pudiéndose monitorear en tiempo real el estado de un paciente o capturar datos sensibles para un posterior análisis médico. De entre los 9 sensores que componen la plataforma, se ha prescindido de 3 de ellos por considerarse ineficaces

para la identificación biométrica o invasivos, estos son: el sensor de posición del paciente (que consiste en un acelerómetro), sensor de presión sanguínea y glucómetro. Los sensores que sí se han utilizado son: sensor de flujo de aire, sensor de respuesta galvánica de la piel, sensor de electrocardiograma, sensor de temperatura, sensor de electromiograma y sensor de pulso y oxígeno en sangre. En la *figura 2-3* se muestra la plataforma e-Health con todos sus sensores conectada a la placa Arduino.



Figura 2-3. Sensores biométricos y plataforma e-Health

A continuación, se describirá en detalle cada uno de los sensores utilizados y su modo de funcionamiento.

2.1.3 Sensor de flujo de aire

El sensor de flujo de aire es un dispositivo que permite medir la frecuencia y amplitud de la respiración del paciente (ver figura 2-4). Consiste en un cable flexible que consta de dos pequeños dientes metálicos que se colocan bajo los orificios nasales. Este sensor no es más que un termopar [13], es decir, un transductor formado por la unión de dos metales distintos que generan una pequeña diferencia de potencial la cual varía en función de la temperatura. De esta forma es posible medir el flujo de aire de un paciente a partir de la temperatura presente en los orificios nasales.

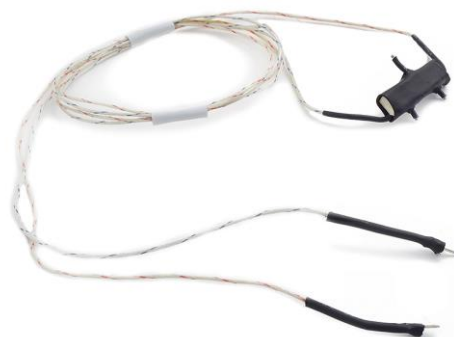


Figura 2-4. Sensor de flujo de aire

2.1.4 Sensor de respuesta galvánica de la piel

La respuesta galvánica de la piel, también llamada conductancia de la piel, es el método por el que se mide la conductancia eléctrica entre dos puntos del cuerpo humano. Este sensor se ha diseñado para realizar dicha medida conectando un cable al dedo índice y otro al dedo medio, calculando cuanta corriente pasa entre estos puntos, al igual que haría un óhmetro en un circuito eléctrico [14].

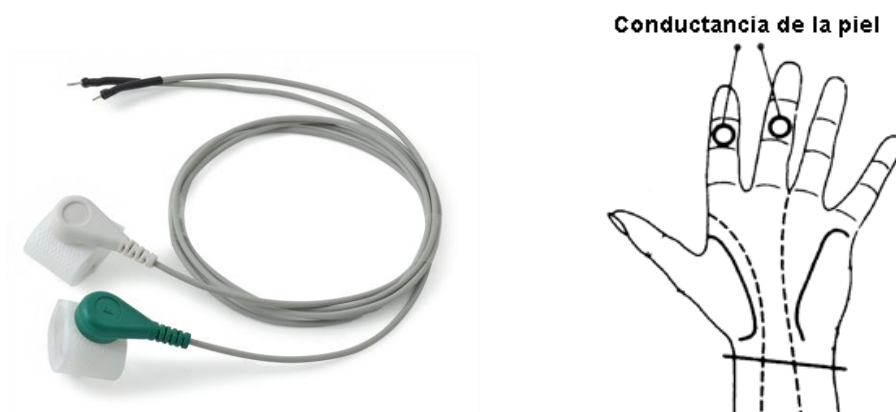


Figura 2-5. Sensor GSR

2.1.5 Sensor de electrocardiograma

Quizás el dispositivo más usado en las pruebas médicas de la medicina moderna. El sensor de ECG de la plataforma e-Health, aun siendo un dispositivo de bajo coste, desempeña su trabajo de manera más que aceptable, permitiendo evaluar de forma continua la actividad eléctrica del corazón [15].

Existen diferentes formas de realizar un ECG, en los que se emplean más o menos electrodos, en este caso el sistema destaca por su simpleza, empleando únicamente tres electrodos (positivo, negativo y neutro) como puede apreciarse en la figura 2-6.

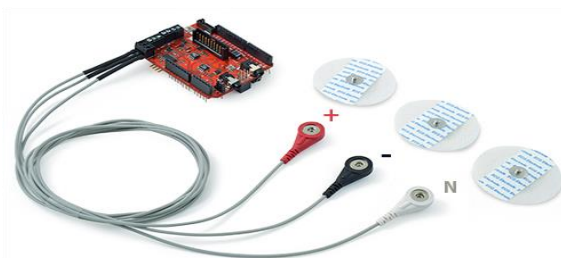


Figura 2-6. Sensor ECG conectado a e-Health

2.1.6 Sensor de temperatura

Este dispositivo realiza la misma función que un termómetro digital, haciendo uso de un semiconductor del cual su resistencia varía en función de la temperatura del cuerpo al que ha sido conectado. De esta forma es posible convertir calor en tensión, y esta a su vez usar un convertidor analógico/digital para pasar la información a números binarios.



Figura 2-7. Sensor de temperatura

2.1.7 Sensor de electromiograma

De forma similar al sensor de ECG, este sensor dotado de 3 electrodos (positivo, negativo y neutro) permite registrar la actividad eléctrica, en este caso de músculos esqueléticos, ya sea durante su etapa de contracción o reposo. La representación de esta actividad eléctrica en el tiempo es llamada electromiograma [16].

2.1.8 Sensor de pulso y oxígeno en sangre

El sensor de pulso y oxígeno en sangre (ver figura 2-8), también llamado pulsioxímetro, se trata de un dispositivo no invasivo que proporciona, además del pulso,



la saturación de hemoglobina (SaO_2) en la sangre [17]. Para ello se vale de las características de dos moléculas distintas presentes en la sangre, la hemoglobina oxigenada, también llamada oxihemoglobina (HbO_2) y la no oxigenada, llamada desoxihemoglobina (Hb).

Figura 2-8. Pulsioxímetro

Cada molécula absorbe una longitud de onda diferente, la oxihemoglobina absorbe las señales de 920 nanómetros (espectro infrarrojo) y la desoxihemoglobina señales de 660 nanómetros (espectro rojo). El pulsioxímetro cuenta con un emisor de ambas longitudes de onda y un fotodetector, colocando uno y otro en lados opuestos de una parte del cuerpo relativamente translúcida y con buen flujo sanguíneo (por ejemplo

un dedo de la mano, pie o en el lóbulo de la oreja) puede calcularse la cantidad de señal que ha conseguido pasar en ambos casos. Conociendo estos datos, la saturación de la hemoglobina puede calcularse según la *ecuación 2.1*:

$$SaO_2 = \frac{HbO_2}{(HbO_2 + Hb)} \quad (2.1)$$

2.2 Consideraciones de diseño de la base de datos

Previamente a la creación de la base de datos, se han establecido algunas consideraciones para su diseño. Se ha acordado que la base de datos esté compuesta de al menos 25 sujetos para asegurar la solidez del sistema. Además la captura de cada señal deberá durar al menos 1 minuto, con el fin de garantizar la suficiente recogida de datos en cada señal. La base de datos estará compuesta por 25 carpetas correspondientes a los 25 sujetos, dichas carpetas contendrán cada una de las medidas realizadas para el correspondiente sujeto, tal y como puede verse en la *figura 2-9*.

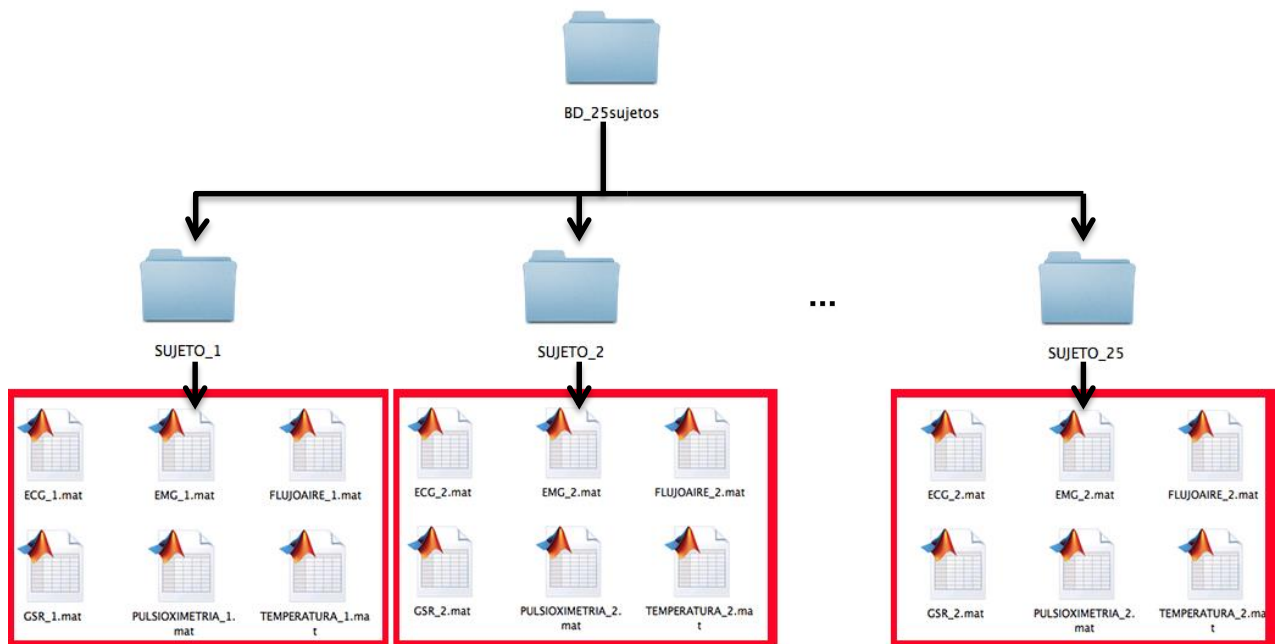


Figura 2-9. Estructura de la base de datos

2.3 Características de las señales a extraer

En este apartado se describirán las implicaciones de cada una de las señales a extraer, sus características y que importancia pueden tener a la hora de identificar al sujeto.

2.3.1 Flujo de aire

Una respiración anormal es un claro indicador de inestabilidad fisiológica, permitiendo detectar casos de hipoxemia o apnea. La amplitud y frecuencia respiratoria pueden aportar información de la capacidad pulmonar y la fuerza del diafragma de la persona, en este trabajo se tratará de demostrar que esta información constituye un fuerte elemento identificador. En la *figura 2-10* puede observarse dos medidas de flujo de aire, las cuales han sido realizadas a sujetos en reposo y con una duración de 30 segundos.

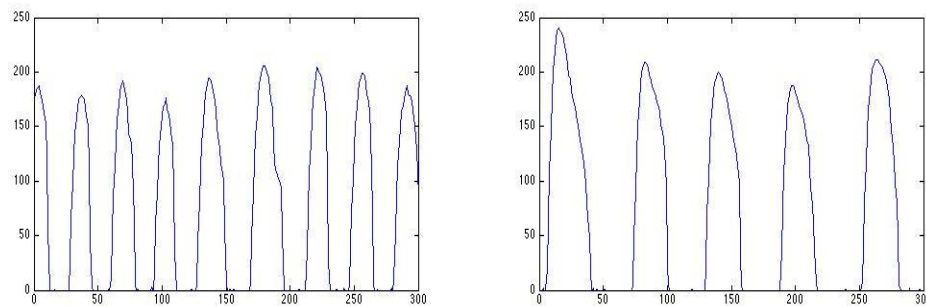


Figura 2-10. Medidas de flujo de aire

En este ejemplo es fácil ver la diferencia entre ambas gráficas, de las que podemos deducir que la gráfica de la derecha corresponde a una persona con mayor capacidad pulmonar que la de la izquierda.

2.3.2 Respuesta Galvánica de la Piel

La conductancia de la piel depende en gran medida de la sudoración del sujeto, a su vez, la sudoración está fuertemente ligada a emociones tales como excitación, estrés o miedo. Esto hace que la Respuesta Galvánica de la Piel cobre gran importancia en las investigaciones relacionadas con temas emocionales y fisiológicos, llegándose a usar esta medida en polígrafos. Si bien existen unos valores que pueden considerarse normales para esta medida, existen diferencias entre personas, por lo que dos sujetos que se encuentren en el mismo estado emocional no tienen porque presentar la misma conductancia. En este trabajo se tratará de usar esta información como elemento identificador.

2.3.3 Electrocardiograma

El electrocardiograma es la representación gráfica en el tiempo de la actividad eléctrica del corazón. Esta medida es de gran importancia en medicina ya que proporciona información de gran utilidad a la hora de diagnosticar problemas tales como enfermedades cardiovasculares, alteraciones metabólicas o antecedentes de cardiopatías.

Esta medida tiene una forma característica la cual puede verse en la *figura 2-11*, que presenta ciertas diferencias entre pacientes sanos, como por ejemplo, la duración del ciclo cardíaco. En pacientes con problemas cardíacos, las diferencias son aún más notables, en ambos casos y como ya se ha demostrado en estudios anteriores, el ECG es un buen elemento identificador. En este Trabajo de Fin de Grado (TFG), el sensor usado es sencillo y de bajo coste, por lo que aunque no se esperan los mismos resultados que con equipos más profesionales, si se espera obtener información importante para el proceso de identificación.

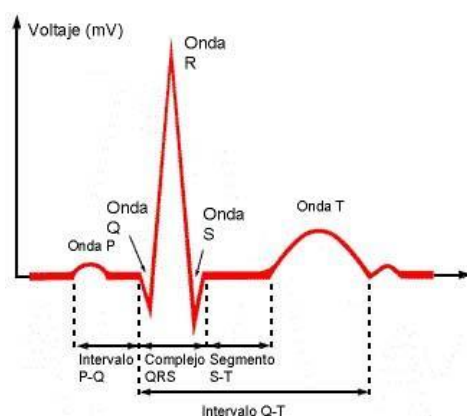


Figura 2-11. Fases de un ECG sano

2.3.4 Temperatura

La temperatura corporal depende del lugar del cuerpo en el que ha sido tomada, su importancia radica en el hecho de que una gran número de enfermedades vienen acompañadas de cambios en la temperatura. En lo que respecta a la identificación de pacientes no proporciona demasiada información, ya que el cuerpo humano trabaja de forma normal dentro de un rango de valores reducido. No obstante, la temperatura corporal presenta variaciones muy pequeñas en el tiempo, por lo que se tratará de usar estas variaciones de temperatura como elemento identificador.

2.3.5 Electromiograma

El electromiograma es usado para monitorear la actividad de los músculos de un paciente, así como para el estudio del sistema neuromuscular. Al igual que el ECG el electromiograma consta de gran importancia como elemento identificador, dado que la respuesta dada es distinta de un sujeto a otro. Esta diferencia será mayor o menor en dependiendo del músculo monitorizado, atendiendo a la fisionomía del sujeto.

2.3.6 Pulso y oxígeno en sangre

Los pulsioxímetros son comúnmente vistos en salas de cuidados intensivos, durante operaciones ó salas de recuperación, entornos donde existen casos de inestabilidad en la saturación de oxígeno en sangre. En este TFG se pretende estudiar si la correlación entre pulso y saturación de oxígeno en sangre es capaz de aportar información útil para un sistema de identificación.

2.4 Proceso de toma de muestras

El proceso de toma de muestras está dividida en dos etapas. En la primera etapa se recoge las muestras de 25 sujetos y se utilizan para la creación de la base de datos anteriormente explicada. En la segunda etapa se vuelven a recoger muestras de los 25 sujetos anteriores, y se crea una nueva base de datos de igual estructura que la primera. De esta forma cada sujeto tendrá dos grupos de señales, uno por etapa, como podemos ver en la *figura 2-12*.

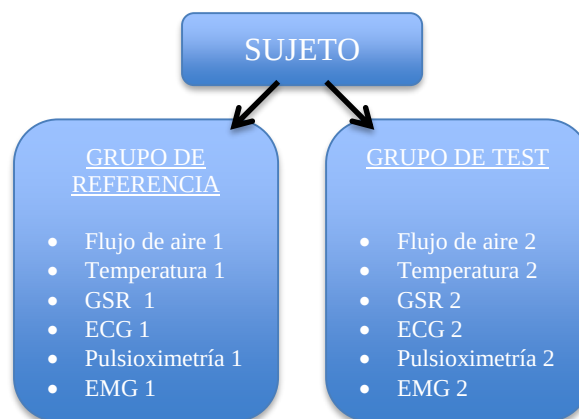


Figura 2-12. Grupos de señales por sujeto

Al primero de estos grupos se le llamará *grupo de referencia*, y será usado para definir al sujeto dentro de nuestro sistema de identificación. El segundo grupo será

llamado *grupo de test* y será usado para comprobar la calidad de nuestro sistema. De esta forma, la eficacia de nuestro sistema de identificación vendrá dado por la cantidad de grupos de test que es capaz de asociar con su correspondiente grupo de referencia.

A continuación, se describe de forma clara el procedimiento seguido para la captura de cada una de las señales anteriormente descritas. Como ya explicó en la sección 2.2, cada captura tendrá una duración de 1 min.

Muestra 1: Flujo de aire

Para esta captura el sujeto debe estar sentado con el cuerpo en reposo y la espalda erguida. El sensor debe ir colocado bajo los orificios nasales, pasando el resto del cable del sensor tras las orejas del sujeto, a modo de sujeción (ver figura 2-13). El sujeto debe respirar con tranquilidad y no hablar durante el proceso, puesto que interferiría en la captura del flujo de aire.

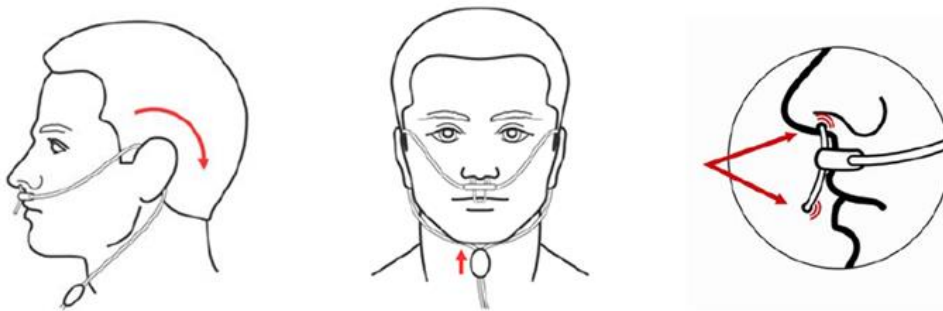


Figura 2-13. Posición del sensor de flujo de aire

Muestra 2: Respuesta Galvánica de la piel

Para que la captura de esta muestra no difiera de un sujeto a otro, se medirá siempre la conductancia entre el dedo índice y el dedo medio de la mano derecha, como se muestra en la *figura 2-14*. Esta medida varía en función del estado emocional del sujeto, por lo que el sujeto debe estar calmado y no recibir influencias externas.

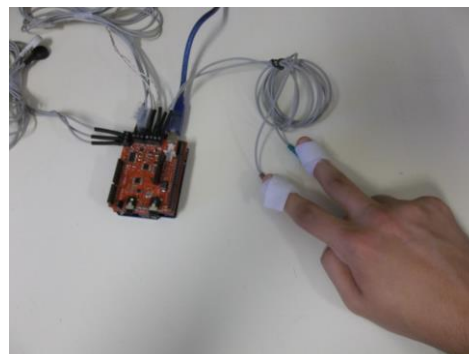


Figura 2-14. Posición de sensor GSR

Muestra 3: Electrocardiograma

El electrocardiograma obtenido en esta medida será muy simple, el sensor cuenta únicamente con tres electrodos, los cuales serán colocados en la parte frontal del cuerpo tal y como se muestra en la *figura 2-15*. Los electrocardiogramas suelen realizarse con el sujeto acostado en decúbito supino, en este TFG por el contrario, se ha optado por adoptar la misma postura que en el resto de pruebas, esto es, sentado y en reposo. El sujeto no debe moverse ni hablar durante la medida, debiendo respirar normalmente mientras se realiza, de otro modo, los movimientos del diafragma o del propio cuerpo podrían modificar el ECG.

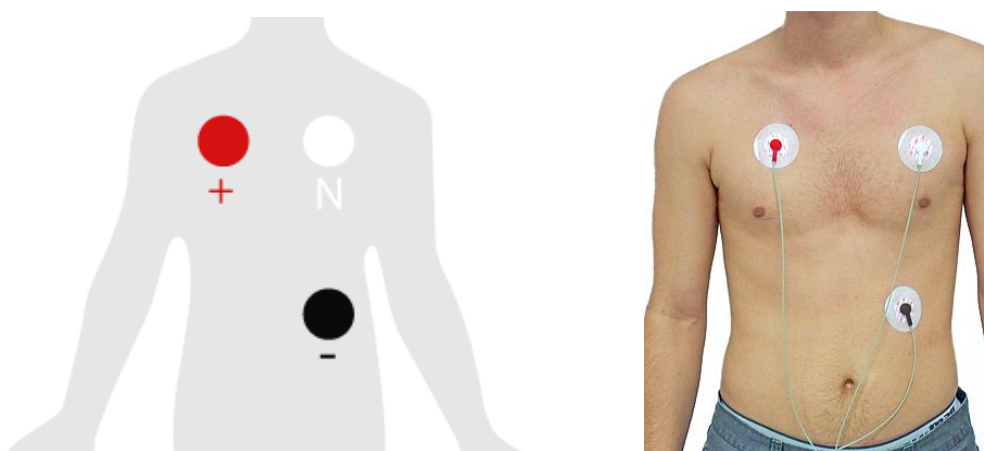


Figura 2-15. Posición de los electrodos

Muestra 4: Temperatura

Para que la captura de esta muestra no difiera de un sujeto a otro, se realizará midiendo siempre la temperatura del dedo índice de la mano izquierda, tal y como se muestra en la *figura 2-16*.

El sujeto debe dejar el brazo en reposo durante la medida, además el sensor puede ser unido mediante cinta adherente para evitar errores en la medida.



Figura 2-16. Posición de sensor temperatura

Muestra 5: Electromiograma

El músculo elegido en este TFG para la realización del electromiograma es el bíceps del brazo derecho en el caso de gente diestra, y el izquierdo en el caso de los zurdos. Al igual que en el caso del ECG, este sensor cuenta con tres electrodos (positivo, negativo y neutro), dichos electrodos son colocados como se muestra en la 2-17. Para esta medida se hará uso de una aplicación móvil que simula el sonido de un metrónomo, dicha aplicación es programada para dar 20 pulsos por minuto.

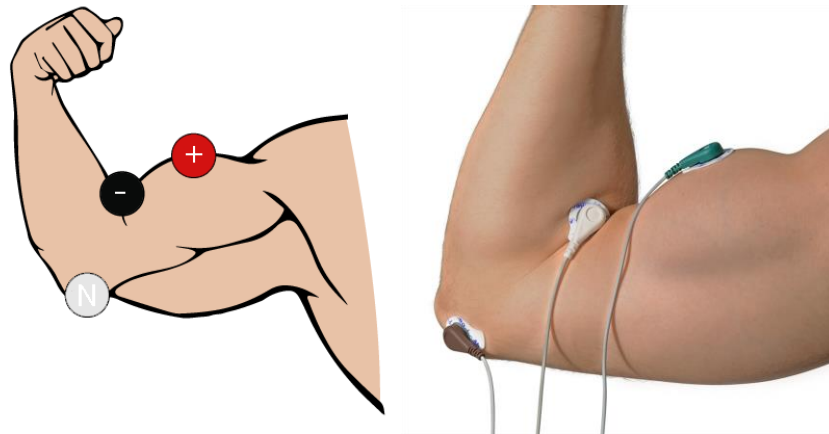


Figura 2-17. Posición del sensor de EMG

El sujeto comenzará con el brazo relajado encima de una superficie plana, al oír el primer pulso contraerá el bíceps y lo relajará con el siguiente pulso. Dado que la prueba tiene una duración de 1 minuto, el sujeto contraerá y relajará el bíceps un total de 10 veces. El resultado obtenido es una señal con la forma de la *figura 2-18*, la cual corresponde a un secuencia de 30 segundos, donde podemos ver las 5 contracciones realizadas en ese tiempo.

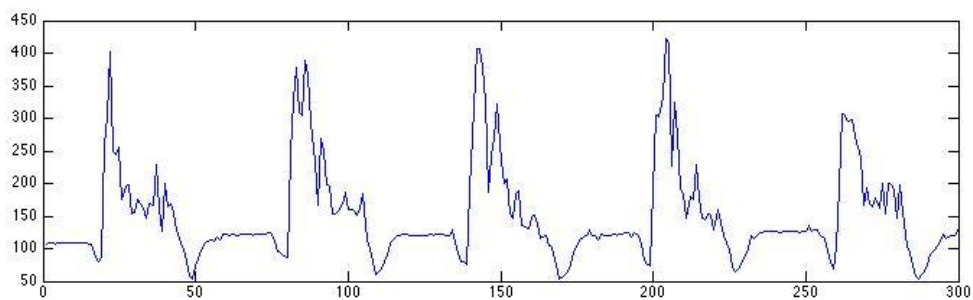


Figura 2-18. Señal EMG de 30 segundos.

Muestra 6: Pulsioximetría

Esta captura se realizará colocando el sensor en el dedo índice de la mano izquierda, tal y como se muestra en la *figura 2-19*. El sujeto debe permanecer sentado con la espalda erguida, estar relajado y no hablar durante la captura de la señal.



Figura 2-19. Posición del pulsioxímetro

Para este TFG, el sistema se ha diseñado de forma que recoja simultáneamente los datos del pulso y los de la saturación de oxígeno en sangre. El resultado obtenido será una señal, conjunto de las dos anteriores, cuyos elementos pares corresponden a la señal del pulso y los impares a los de la saturación de oxígeno en sangre (ver figura 2-20). Durante la realización de este TFG se tratará de sacar el máximo provecho de cada una de las tres señales.

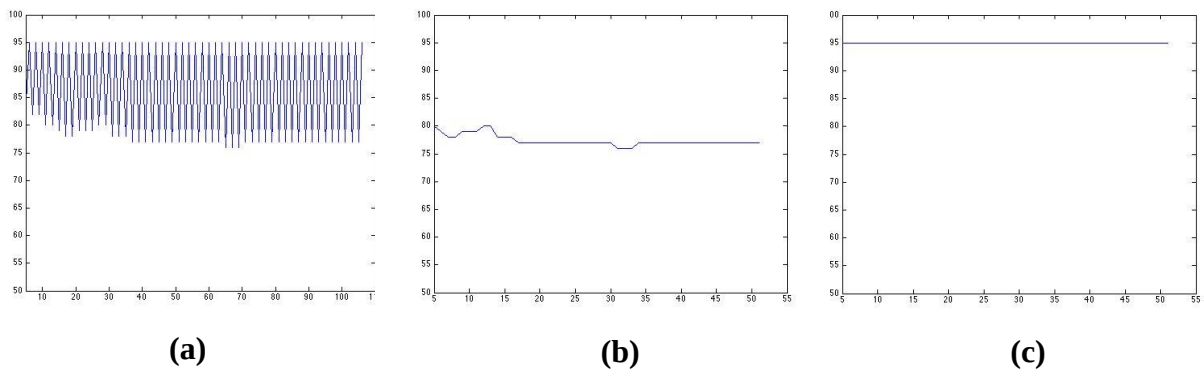


Figura 2-20. Señales del pulsioxímetro

(a) Pulsioximetría , (b) Pulso y (c) Saturación

Capítulo 3. Parametrización y sistema de clasificación

En este capítulo se abarca el sistema de clasificación que va a ser utilizado en este TFG, el cual se basa en Máquinas de Vectores de Soporte. Se explicará el funcionamiento de las Máquinas de Vectores de Soporte, así como sus características. Seguidamente se explicarán diversos parámetros útiles, los cuales pueden aportar información adicional de las señales previamente obtenidas durante la toma de muestras. Finalmente se realizará una introducción al Análisis de Componentes Principales y se explicará como se ha realizado el diseño del clasificador.

3.1 Introducción a Máquinas de Vectores de Soporte

Las Máquinas de Vectores de Soporte son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y basados en la idea de minimización del riesgo estructural (SRM, por sus siglas en inglés) [18].

Las SVM son ampliamente utilizadas en aplicaciones de clasificación y reconocimiento de patrones, donde han demostrado grandes resultados. Mediante estos algoritmos es posible etiquetar las clases de un conjunto de muestras y entrenar una SVM para construir un modelo que prediga a que clase pertenece una nueva muestra [19]. En el caso de este TFG las clases etiquetadas serán cada uno de los grupos de referencia pertenecientes a los sujetos medidos, siendo objetivo del clasificador predecir a que clase corresponde los grupos de test obtenidos en la etapa 2 de la toma de muestras. A continuación se explica el funcionamiento de las SVM más detalladamente.

Dado un conjunto separable de datos, es posible definir un hiperplano de separación como una función lineal que es capaz de separar dicho conjunto sin error (ver figura 3-1.a). Para ello, las SVM pueden proyectar el conjunto de datos sobre un espacio de dimensionalidad mayor y luego tratar de buscar un hiperplano óptimo que separe y maximice la distancia entre las clases en el nuevo espacio (ver figura 3-1.b).

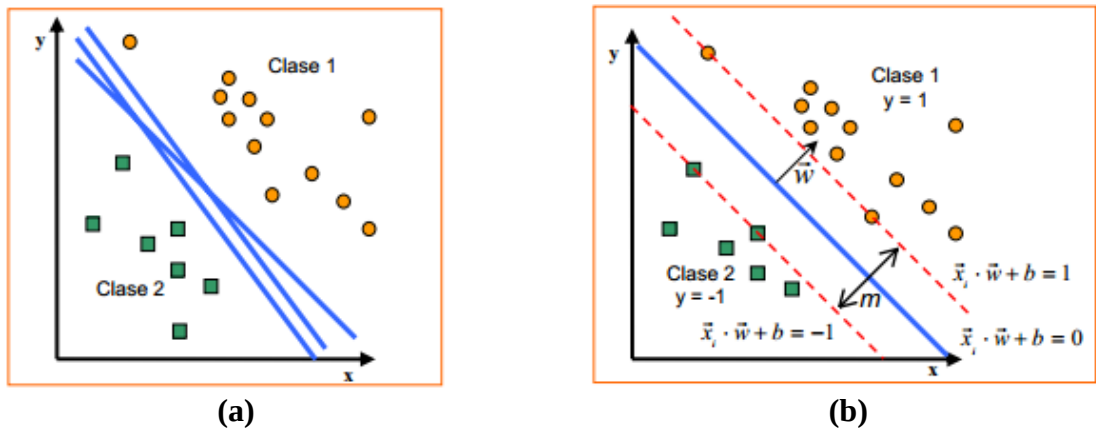


Figura 3-1. Hiperplanos de separación

(a) Posibles hiperplanos, (b) Hiperplano óptimo.

Por tanto, el principal problema de las SVM es encontrar aquel hiperplano óptimo que permita separar los conjuntos de datos de la mejor manera posible. Dentro de este planteamiento el sistema puede encontrarse con dos casos principales: linealmente separable y no linealmente separable [20].

Caso linealmente separable:

En este caso los conjuntos de datos están claramente diferenciados, tal y como puede verse en la *figura 3-2.a*. Las SVM en este caso separan sin esfuerzo los diversos conjuntos de datos en subgrupos y los etiquetan, todo esto mediante un hiperplano óptimo con la intención de forzar la generalización de la máquina que se esté construyendo.

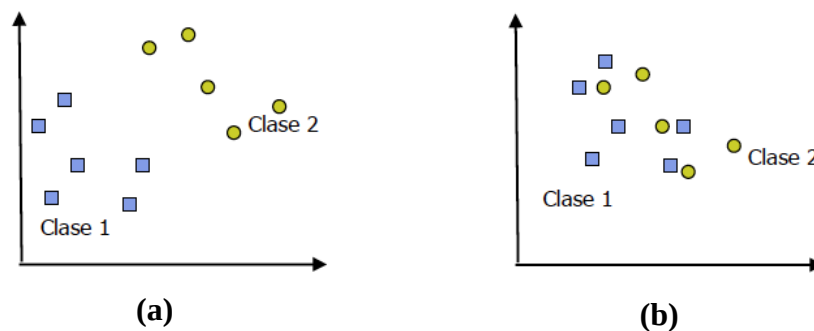


Figura 3-2. Posibles conjuntos de datos

(a) Linealmente separable, (b) Linealmente no separable.

Caso no linealmente separable:

En este caso la separación entre conjuntos de datos no está bien definida, por lo que a priori resulta imposible diferenciarlos mediante una línea, tal como podemos ver en la *figura 3-2.b*. Una de la formas que SVM tiene para solucionar este tipo de casos es el uso de *funciones kernel* [21].

Las funciones *kernel* son funciones matemáticas que permiten convertir un caso no linealmente separable en el espacio dimensional original, a uno más sencillo y linealmente separable en un espacio de dimensionalidad mayor (ver figura 3-3).

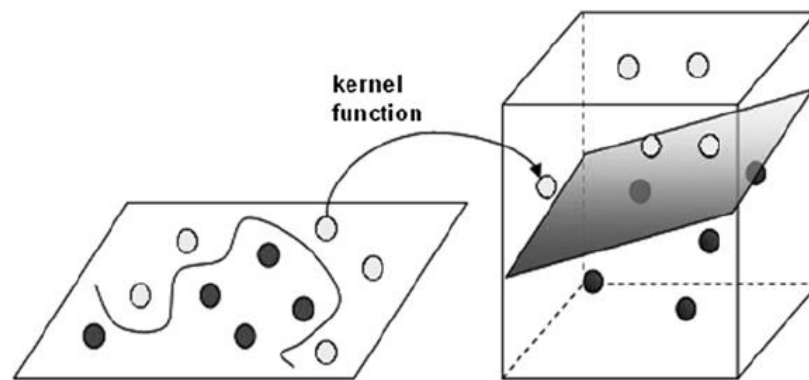


Figura 3-3. Solución de función kernel

Debido a que en este Trabajo Fin de Grado se utilizan diversos grupos de señales con valores muy cercanos entre sí, la separación lineal puede no ser la mejor de las opciones, es por esto que se aplicarán diversas funciones kernel a nuestro clasificador. Mas concretamente las funciones *kernel* a usar serán las siguientes: lineal, cuadrática, polinomial, función de base radial gaussiana (RBF por sus siglas en inglés) y perceptrón multicapa (MLP por sus siglas en inglés). De esta forma se podrá determinar que caso se ajusta más a este trabajo una vez se obtengan los resultados. Más detalles sobre estas funciones pueden ser consultadas en [22].

3.2 Descripción de parámetros útiles

Durante la toma de muestras se han obtenido grupos de señales de 25 sujetos diferentes y se han almacenado dichas señales en forma de vectores. A su vez, estos grupos de vectores serán usados para definir a su sujeto correspondiente.

Una vez hecho esto, es posible extraer diversos parámetros que aporten información adicional y de gran relevancia al sistema de identificación. De esta forma se pueden crear vectores compuestos de parámetros de señales, en lugar de señales completas, y usarlos para definir a los sujetos usando menos valores. A continuación se detallan los parámetros utilizados en este Trabajo Fin de Grado:

Entropía

La entropía es una medida estadística que proporciona información de los cambios producidos en un sistema, desde su estado inicial hasta su estado final [23]. De esta manera, es posible ver la cantidad de “desorden” existente en la señal.

La entropía puede ser definida mediante la *ecuación 3.1*.

$$H = - \sum_{i=0}^N P_i * \log_2 P_i \quad (3.1)$$

Media

La media aritmética es la suma de un conjunto de elementos dividido entre el número de elementos del conjunto. Se espera que este parámetro sea de gran relevancia en aquellas señales con más periodicidad. La media puede ser definida mediante la *ecuación 3.2*.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n a_i = \frac{a_1 + a_2 + \dots + a_n}{n} \quad (3.2)$$

Coefficiente de asimetría o Skewness

Este parámetro mide la falta de simetría en una distribución, entendiendo la simetría como la similitud entre izquierda y derecha del punto central de la distribución [24]. El coeficiente de asimetría para un conjunto de datos invariable se define según la *ecuación 3.3*.

$$Skewness = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^3 / N}{s^3} \quad (3.3)$$

Donde Y_i es el elemento, $\bar{Y}$ es la media, N es el número de elementos y s el número de elementos de la distribución.

Curtosis

La curtosis permite medir cuán escarpada o achatada está una curva o distribución, permitiendo medir la cantidad de datos cercanos a la media. Si el coeficiente es positivo, indicará mas concentración de datos en torno a la media. Por el contrario si es negativo indica una menor concentración de datos en torno a la media. Si el coeficiente es nulo, la distribución se dice normal [25]. La curtosis para un conjunto de datos invariable se define según la *ecuación 3.4*.

$$Kurtosis = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4 / N}{s^4} \quad (3.4)$$

Donde Y_i es el elemento, $\bar{Y}$ es la media, N es el número de elementos y s el número de elementos de la distribución.

Desviación estándar

Esta medida revela la cantidad de variación o dispersión existente en un conjunto de datos. Se define como la raíz cuadrada de la varianza y puede obtenerse mediante la *ecuación 3.5*.

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.5)$$

Donde x_i es el elemento y $\bar{x}$ es la media.

3.3 Introducción al Análisis de Componentes Principales

Cuando se trata de encontrar conexiones entre variables pueden presentarse dos problemas principales. Por un lado, una gran cantidad de variables significa considerar una gran cantidad de coeficientes de correlación, lo cual hace difícil observar las conexiones entra variables. Por otro lado, aparece en ocasiones una fuerte correlación entre variables, esta alta correlación supone menos variabilidad y hace más difícil la identificación. Para tratar de solucionar estos problemas, es considerablemente útil reducir el número de variables mediante el análisis de componentes principales (PCA por sus siglas en inglés).

El análisis de componentes principales es una técnica estadística utilizada para reducir las dimensiones de un conjunto de datos. Además permite acentuar la variabilidad del conjunto de datos, hallando primero sus causas y ordenándolas luego por importancia.

Para llevar acabo esta reducción de dimensionalidad, se realiza una transformación del sistema de coordenadas original a través de una rotación de magnitud θ que dará como resultado un nuevo sistema en el que los datos se encuentran incorrelados, tal y como puede verse en la *figura 3-4*. En este nuevo sistema, la varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (Primer Componente Principal), la segunda varianza de mayor tamaño se establecerá como segundo eje y así sucesivamente. Esta característica del sistema constituye su mayor ventaja, ya que permite retener los conjuntos de datos que más contribuyen a la varianza dentro de las componentes principales, es decir, se mantiene la mayor parte de la información en pocas dimensiones [26].

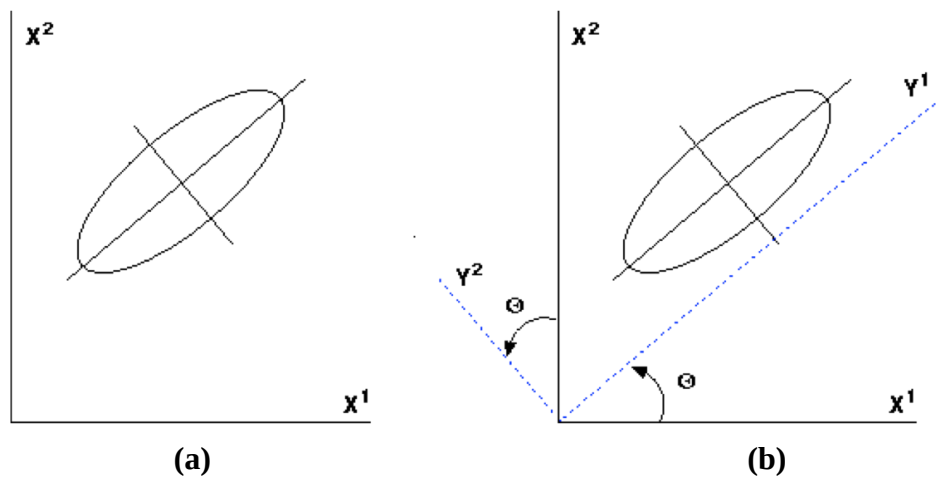


Figura 3-4. Transformación de coordenadas

(a) X^1 y X^2 están correladas (b) Y^1 e Y^2 están incorreladas.

De esta forma, al aplicar PCA sobre los datos con los que se ha trabajado previamente, se consigue dar importancia a aquellos que realmente aportan información útil en la identificación, pudiendo además, descartar aquellos que aportan poco o nada al sistema, ya que actuarían como ruido y disminuirían la eficacia del sistema. Por tanto, serán los datos almacenados en las componentes principales los que se usarán durante el desarrollo de la fase de clasificación.

3.4 Desarrollo del sistema de clasificación

En el desarrollo del sistema de identificación se ha elegido MATLAB como entorno de trabajo, dado que su amplia versatilidad a la hora de trabajar con vectores lo convierte en una herramienta adecuada para esta situación. En la etapa de toma de muestras, los datos recogidos por la placa Arduino son transferidos mediante conexión puerto serie y procesados por MATLAB. De esta forma, la captura de señales puede ser representada gráficamente y en tiempo real, permitiendo visualizar si la captura se ha realizado correctamente, en caso de ser así, las señales son almacenadas como vectores siguiendo la estructura descrita en el apartado 2.2.

Trabajar con las señales de forma directa supondría un proceso de clasificación lento y torpe, debido a la gran cantidad de información y a la correlación presente entre los señales. Debido a esto, de las señales almacenadas se extraerán aquellos parámetros útiles que se hayan probado más eficientes para la identificación, posteriormente, se definirá a cada sujeto mediante un vector compuesto por estos parámetros en orden de mayor a menor importancia. Para simplificar el manejo de los datos, los vectores serán agrupados en forma de matriz, donde las filas representarán sujetos y las columnas parámetros útiles. Dado que cada sujeto posee dos grupos de señales, una de referencia y otra de test, se obtendrá dos matrices de sujetos, *matriz de referencia* y *matriz de test*.

Para aplicar PCA al sistema es necesario obtener la matriz de transformación que permita el cambio de coordenadas. Dicha matriz se obtendrá a partir de la matriz de referencia mediante la función “*pca*” en Matlab, la matriz de transformación resultante será usada para generar el nuevo sistema de coordenadas tanto de la matriz de referencia como de la matriz de test, tal como se describe en la *ecuación 3.6*.

$$Y = W^T * X \quad (3.6)$$

Donde X corresponde al sistema de coordenadas original, W^T es la matriz de transformación e Y es el nuevo eje de coordenadas resultante.

Tras aplicar PCA, las matrices resultantes contendrán la mayor parte de la información en sus componentes principales, pudiendo desechar el resto. Por tanto, se

reducirá la longitud del vector en una cantidad que se haya probado permita conservar únicamente aquellas componentes que supongan el grueso de la información.

Como ya se ha explicado anteriormente, se ha optado por el uso de SVM como clasificador en el sistema de identificación. Este clasificador necesitará ser entrenado primero antes de poder desarrollar un modelo predictivo, para ello se etiquetará cada una de las filas de la matriz de referencia mediante la función “*svmtrain*”. Durante esta fase de entrenamiento se aplicará la función *kernel* que se haya demostrado más eficiente para los datos con los que se trabaja.

Por último, se aplicará la función “*svmclassify*”, la cual comparará cada una de las filas de la matriz de test con todas las etiquetadas durante la fase de entrenamiento. Dicha función dará un valor tras cada comparación, que será mayor o menor en función de la similitud entre los datos comparados. Por tanto, se espera que el mayor valor obtenido de comparar una fila de la matriz test con las de la matriz de referencia, corresponda al caso en que las dos filas pertenecen a la misma persona.

Para determinar la eficacia del sistema, se evaluará en cuantos casos se cumple que el mayor valor obtenido procede de comparar una fila de la matriz de test con su correspondiente en la matriz de referencia. En el siguiente diagrama, se muestra cada una de las fases del sistema de clasificación:

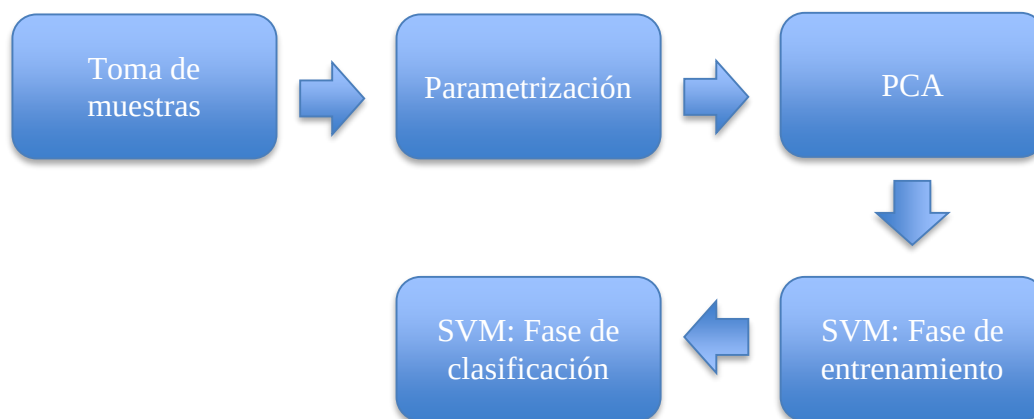


Figura 3-5. Diagrama del sistema de clasificación

Capítulo 4. Experimentos y resultados

En este capítulo se explicará la metodología experimental usada en este Trabajo Fin de Grado y se expondrán los resultados obtenidos por el clasificador en conjunto con aquellos elementos aplicados para la mejora del sistema, los cuales ya han sido descritos en el *capítulo 3*. Se comenzará con los resultados obtenidos de las señales sin procesar, valorándose su eficacia ya sea de forma individual o en grupo. Posteriormente, se estudiarán los resultados obtenidos al usar únicamente parámetros útiles extraídos de la señales anteriores. Por último, se presentará aquellos resultados obtenidos de aplicar PCA en el caso anterior.

4.1 Metodología experimental

En primer lugar, se pretende determinar hasta que punto es eficaz el sistema de identificación cuando trabaja únicamente con las señales capturadas durante la toma de muestras. Para ello, se procederá primero entrenando el sistema con cada una de la señales de forma individual, de aquellas que resulten más relevantes, se crearán diversas combinaciones de señales con la intención de mejorar el resultado anterior. Durante cada una de las pruebas, se probarán los distintos tipos de funciones *kernel* explicados anteriormente, para ver cual se ajusta mejor a cada caso.

A continuación, se tratará de mejorar el sistema trabajando con vectores compuestos de un número menor de elementos. Mientras que en el caso anterior se usan señales compuestas de cientos de elementos, en este se usarán parámetros extraídos de dichas señales. Para determinar que parámetros aportan más información en cada señal se entrenará el sistema con cada una de ellas. Al igual que en el caso anterior se realizarán distintas combinaciones pretendiendo mejorar el resultado individual y se variará la función *kernel*.

Para finalizar, se tratará de reducir la dimensionalidad de la mejor combinación obtenida. Se aplicará PCA a dicha combinación y se mostrarán los resultados obtenidos de reducir en cantidades varias la dimensionalidad de la combinación, con motivo de encontrar la cantidad más efectiva de componentes principales.

4.2 Resultados de la fase inicial

A continuación se presentan los resultados obtenidos para cada tipo de señal y función *kernel*. En verde, los mejores resultados para cada señal.

	LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
EMG	40	56	24	64	0
TEMP	0	16	0	60	4
FA	8	8	4	4	0
GSR	4	20	12	32	0
PA	28	16	24	16	0
PULSO	16	16	16	20	4
SAT	12	0	0	16	4
ECG	4	4	4	4	8

Tabla 4-1. Resultados de señales individuales

Siendo TEMP la temperatura, FA el flujo de aire, PA la pulsioximetría y SAT la saturación. Otra forma de comparar los resultados sería de forma gráfica, tal y como puede observarse en la siguiente figura en la siguiente figura:

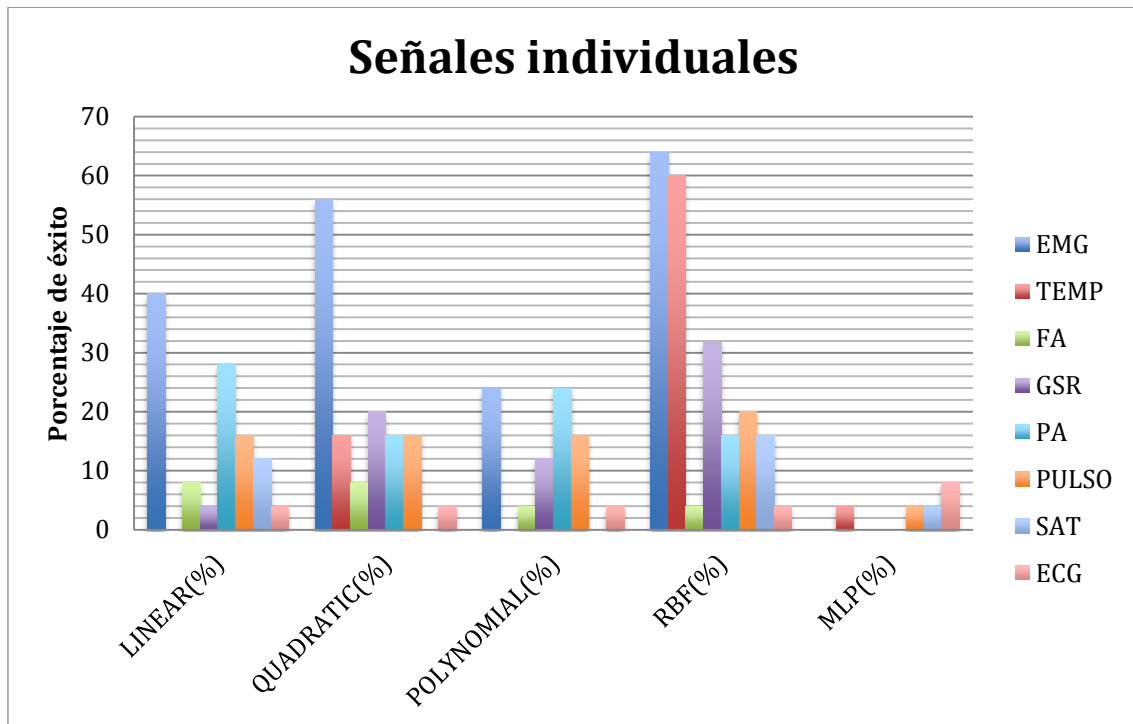


Figura 4-1. Resultados de señales individuales

En esta gráfica puede verse que la señal de EMG es la que más información ofrece indistintamente de la función kernel usada, obteniendo un 64% de porcentaje de éxito como valor máximo. Por el contrario, el flujo de aire se muestra poco eficiente en la identificación, logrando apenas un 8% como valor máximo. Además puede observarse que, el uso de RBF como función *kernel* permite al clasificador diferenciar las señales individuales mejor que el resto de funciones, a excepción de los casos de flujo de aire y pulsioximetría.

Tras analizar las señales de forma individual, se procede a entrenar el clasificador con combinaciones de señales, se han realizado 3 tipos de combinaciones:

Combinación 1:

Formada por 5 señales elegidas aleatoriamente para ver el comportamiento del clasificador al aumentar el número de señales. La combinación contiene las siguientes señales y en el siguiente orden: EMG, flujo de aire, temperatura, GSR y ECG.

Combinación 2:

En esta combinación se incluye la totalidad de las señales capturadas, las cuales han sido ordenadas en función de su máximo porcentaje de éxito, colocándose primero aquellas con mejor resultado. La combinación contiene las siguientes señales y en el siguiente orden: EMG, temperatura, GSR, pulso, saturación, pulsioximetría, flujo de aire y ECG.

Combinación 3:

Esta combinación está formada por aquellas señales cuyos valores máximos son obtenidos mediante la función kernel RBF(la mayoría de ellas), siendo descartadas aquellas que obtienen sus valores máximos en otro tipo de funciones. La combinación contiene las siguientes señales y en el siguiente orden: EMG, temperatura, GSR, pulso y saturación.

		LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
Comb. 1	EMG, FA, TEMP, GSR, ECG	32	20	4	4	0
Comb. 2	EMG, TEMP, GSR, PULSO, SAT, PO, FA, ECG	48	24	4	4	0
Comb. 3	EMG, TEMP, GSR, PULSO, SAT	68	68	36	28	0

Tabla 4-2. Resultados de combinaciones de señales

Los resultados de la combinación 1 reflejan que el clasificador puede dar mejores resultados al recibir más información mediante el uso de combinaciones de señales en lugar de señales individuales, ya que como puede verse, los resultados obtenidos superan al de muchas señales individuales. No obstante, no se supera el valor de la señal temperatura y ECG, cuyo resultado máximo supera en hasta un 32% a los obtenidos en la combinación 1.

En la combinación 2 se ha usado la totalidad de la señales, se espera que al contar con aún más información, el clasificador pueda discriminar más fácilmente a un sujeto de otro y por consiguiente, obtener mejores resultados. Los resultados confirman una mejora con respecto a la combinación 1, pero aún sigue sin superarse el valor alcanzado en la señal temperatura y EMG de forma individual.

Por último, en la combinación 3 se han utilizado solamente aquellas señales cuyo valor máximo es conseguido en la misma función kernel, descartándose las otras. En este caso sí se ha superado los resultados obtenidos en la totalidad de las señales individuales. Es destacable como esta combinación supera los resultados de la combinación 2, esto puede deberse a que la combinación 2 contaba con señales que no solo no aportaban demasiada información, sino que entorpecían la tarea del clasificador, al prescindir de estas señales para este último caso los resultados obtenidos se han visto reforzados.

Cabe resaltar que, mientras que en los casos individuales la función *kernel* RBF reflejaba un mejor manejo de los datos, en el caso de combinaciones de señales no resulta así, siendo la función Lineal la que mejor se ha adaptado para estos conjuntos de señales.

Por último se presenta de forma gráfica los resultados obtenidos por las tres combinaciones en comparación con la obtenida por la señal EMG.

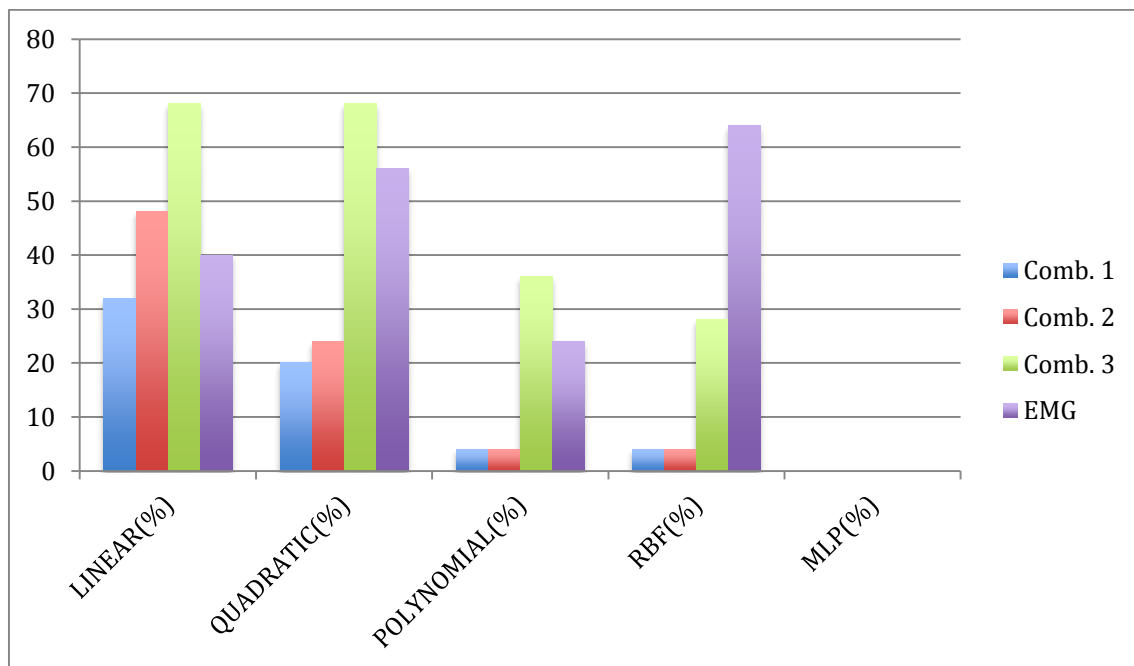


Figura 4-2. Comparación de combinaciones y EMG

4.3 Modificaciones del sistema

En este apartado se estudia la eficacia de los distintos parámetros aplicados a cada señal, posteriormente se crearán combinaciones de estos parámetros al igual que se hizo en el caso anterior, tratando de superar los resultados individuales. En la *tabla 4-3* se muestran los resultados obtenidos para cada señal, además se marca en verde el mayor resultado obtenido para cada función *kernel*, en caso de haber varios resultados iguales dentro de una misma función, se seleccionará solo uno de ellos.

		PARÁMETRO	LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
EMG	Entropía	4	12	16	24	4	
	Media	8	12	24	16	20	
	Skewness	8	24	12	20	16	
	Curtosis	8	12	4	12	8	
	Std	8	12	20	16	4	
TEMP	Entropía	0	0	0	36	4	
	Media	8	12	20	20	20	
	Skewness	4	4	8	8	4	
	Curtosis	0	8	4	12	8	
	Std	4	12	12	8	8	
FA	Entropía	12	12	12	16	0	
	Media	8	4	20	16	4	
	Skewness	8	12	0	4	4	
	Curtosis	0	8	16	8	4	
	Std	8	12	8	12	4	
GSR	Entropía	8	16	16	24	4	
	Media	4	12	12	16	12	
	Skewness	4	12	12	8	4	
	Curtosis	4	4	0	0	4	
	Std	8	4	8	12	0	
PA	Entropía	0	0	0	36	4	
	Media	4	24	20	20	16	
	Skewness	8	16	8	8	8	
	Curtosis	4	12	20	12	8	
	Std	8	8	4	8	12	
PULSO	Entropía	0	0	0	40	4	
	Media	8	12	8	20	12	
	Skewness	4	8	16	4	4	
	Curtosis	4	8	12	0	8	
	Std	4	12	16	12	8	
SAT	Entropía	0	0	0	36	4	
	Media	4	8	16	12	12	
	Skewness	0	4	0	4	0	
	Curtosis	4	4	0	4	0	
	Std	4	12	16	4	12	
ECG	Entropía	12	24	16	12	16	
	Media	8	16	12	8	4	
	Skewness	8	12	4	16	16	
	Curtosis	8	20	16	24	12	
	Std	4	12	16	20	20	

Tabla 4-3. Resultados de parámetros de señales

Siendo Std es la desviación estándar.

Una vez conocidos los resultados para cada caso, se procede a formar combinaciones de elementos. En este caso se crearán 3 combinaciones, formadas por los parámetros más representativos de las señales en cada una de las 3 funciones kernel que mejor resultados han dado. Esto es, por orden: RBF, Polinomial y Cuadrática.

La composición de cada una de estas 3 nuevas combinaciones es la siguiente:

Combinación 4:

Esta combinación está formada por los parámetros de cada señal que mejor resultado han dado en la función RBF. Está compuesta por: entropía del pulso, entropía de la saturación, entropía de la pulsioximetría, entropía de la temperatura, entropía del EMG, entropía de GSR, curtosis del ECG y entropía del flujo de aire.

Combinación 5:

Esta combinación está formada por los parámetros de cada señal que mejor resultado han dado en la función Polinomial. Está compuesta por: media de EMG, media de la temperatura, media del flujo de aire, media de la pulsioximetría, entropía de GSR, skewness del pulso, media de la saturación y entropía del ECG.

Combinación 6:

Esta combinación está formada por los parámetros de cada señal que mejor resultado han dado en la función Cuadrática. Está compuesta por: skewness del EMG, media de la pulsioximetría, entropía del ECG, entropía de GSR, media de la temperatura, entropía del flujo de aire, media del pulso y desviación estándar de la saturación.

Los resultados obtenidos para cada combinación son expuestos en la *tabla 4-4*.

	LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
Comb. 4	36	48	52	56	20
Comb. 5	52	60	78	72	0
Comb. 6	60	76	76	76	4

Tabla 4-4. Resultados de combinación de parámetros

En este caso, los resultados muestran de una forma más clara como el uso de combinaciones de parámetros permite obtener mejores resultados que con parámetros individuales. Además, este sistema ha resultado ser el más eficaz hasta este punto, a excepción de la combinación 4, logrando superar ampliamente los resultados obtenidos en la combinación 3. Por otra parte, cabe destacar que, mientras que la combinación 3 hacía uso de señales de cientos de elementos cada una, en la combinación 4 se utiliza únicamente 8 elementos, lo cual facilita la operación del clasificador y reduce drásticamente el tiempo de procesado.

4.4 Resultados de la fase final

En esta última fase se aplicará PCA a las combinaciones anteriores, con la intención de reducir aún más el número de elementos y mejorar la eficacia del sistema. Así pues, en *tabla 4-5*, *tabla 4-6* y *tabla 4-7* se muestran los resultados obtenidos de reducir cada una de las combinaciones de 8 elementos, a un número N inferior a 8.

Comb. 4	LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
N=1	8	16	16	24	16
N=2	20	40	52	64	24
N=3	28	56	60	68	4
N=4	32	48	48	56	4
N=5	32	48	48	56	4
N=6	32	48	48	56	4
N=7	32	48	48	56	4

Tabla 4-5. PCA en combinación 4

Comb. 5	LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
N=1	12	12	20	20	12
N=2	8	40	44	56	16
N=3	24	68	68	60	0
N=4	44	76	80	84	4
N=5	36	88	88	88	4
N=6	56	80	76	76	0
N=7	44	76	76	80	4

Tabla 4-6. PCA en combinación 5

Comb. 6	LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
N=1	8	16	12	16	16
N=2	12	28	40	60	8
N=3	12	44	64	72	4
N=4	32	68	72	80	8
N=5	52	80	84	80	4
N=6	40	72	72	72	8
N=7	52	64	64	64	8

Tabla 4-7. PCA en combinación 6

Como puede observarse en los resultados, la aplicación de PCA ha permitido no solo obtener mejores resultados en cada una de las tres combinaciones, sino que además ha permitido reducir la dimensionalidad. El caso más destacado es la combinación 5, donde se ha logrado reducir su dimensionalidad de 8 a 5 elementos y aumentar en un 10% su porcentaje de éxito, pasando de 78% a 88%.

La cantidad de combinaciones que pueden construirse es inmensa, durante el desarrollo de este Trabajo Fin de Grado se han realizado numerosas pruebas con diferentes opciones, además de las ya mencionadas. En el momento de redacción del presente trabajo, la combinación más exitosa esta formada por 12 elementos y ha logrado un porcentaje de éxito del 92%. A continuación se presenta la composición y resultados de dicha combinación, antes y después de aplicar PCA.

La combinación está formada por los siguientes 12 elementos: skewness de saturación, curtosis de pulsioximetría, entropía del EMG, media de pulsioximetría, media de saturación, curtosis del pulso, desviación estándar de EMG, desviación estándar del flujo de aire, desviación estándar de la temperatura, desviación estándar del ECG, curtosis del ECG y media del ECG. Mediante estos 12 elementos, sin aplicar PCA, se obtienen los resultados representados en la *tabla 4-8*.

	LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
12 elementos	64	60	64	76	4

Tabla 4-8. Combinación de 12 elementos

Aunque su valor máximo es ligeramente inferior al de la combinación 5, en este caso se cuenta con un mayor número de elementos, lo cual conlleva mayor probabilidad de que alguno de los elementos afecte negativamente al resultado. Esta afirmación puede corroborarse mediante la aplicación de PCA, donde se reducirá la dimensionalidad de la combinación a una cantidad de elementos N inferior a 12.

12 elementos	LINEAL(%)	CUADRÁTICA(%)	POLINOMIAL(%)	RBF(%)	MLP(%)
N=1	8	16	24	20	8
N=2	8	24	24	40	16
N=3	20	56	72	60	0
N=4	52	68	80	84	0
N=5	56	68	72	80	4
N=6	60	76	76	72	0
N=7	68	80	80	80	8
N=8	76	72	80	72	0
N=9	92	72	76	80	4
N=10	84	84	92	84	4
N=11	80	92	88	88	0

Tabla 4-9. PCA en combinación de 12 elementos

Los resultados obtenidos revelan de forma clara un aumento en el porcentaje de éxito, logrando un valor superior al de casos anteriores y reduciendo el número de elementos a 9 en el mejor de los casos.

4.5 Análisis de los resultados

Tras la extracción de resultados, surgen una serie de revelaciones dignas de análisis. Así pues, en las siguientes líneas se muestran aquellos hechos encontrados de forma tanto implícita como explícita en los resultados.

Se puede apreciar como la señal de ECG en sí misma proporciona poca información a la identificación, lo que conlleva resultados bastante bajos. El motivo de que esto ocurra tiene lugar en el método de captura de la señal, realizado únicamente por 3 electrodos, y que da lugar a una señal de baja calidad. Por el contrario, la señal EMG parece definir adecuadamente a los sujetos, esto resulta interesante, dado que la captura de dicha señal también se realiza mediante 3 electrodos. Esta diferencia entre señales puede deberse al nivel de complejidad de cada una, mientras que el corazón consta de movimientos rápidos e independientes de la voluntad del sujeto, el EMG presenta un carácter mucho más pausado, con menos repeticiones y en donde el sujeto ejerce control absoluto sobre la intensidad del movimiento. Por este motivo, una señal simple y clara como la del EMG, permite en este caso aportar más información a la identificación que la señal ECG, la cual requiere de un método de captura más preciso y complejo, que permita recoger aquellas características del ECG que distinguen a un usuario de otro. No obstante, una señal ECG de baja calidad sigue teniendo utilidad, dado que se ha demostrado que, tras aplicar parametrización, se obtienen resultados iguales o superiores que en el resto de señales.

Es fácilmente observable que, en aras de obtener un mayor porcentaje de éxito en la identificación, no basta con entrenar el clasificador con la mayor cantidad de información posible, sino con la información adecuada. Esto se ve de forma clara al comparar las combinaciones 2 y 3, mientras que en la combinación 2 se hacía uso de todas las señales disponibles, en la 3 se empleaban únicamente aquellas que habían dado buenos resultados. En consecuencia, los valores alcanzados por la combinación 3

mejoraban con respecto a la 2, demostrando que incluir señales que aporten poca o nula información al clasificador, empeora el porcentaje de éxito obtenido.

Lo afirmado anteriormente cobra fuerza de nuevo en la fase de parametrización, donde además de superarse los valores de la fase previa, se consigue reducir el número de elementos necesarios de forma abismal, de cientos a únicamente 8 elementos, creando combinaciones de parámetros en lugar de señales completas, como puede verse en los resultados de las combinaciones 5 y 6. Vuelve a resaltar la importancia de entrenar el clasificador con la información correcta y queda patente que determinados parámetros consiguen aportar más información que la señal en sí misma.

Cabe destacar el buen funcionamiento que presenta el análisis de componentes principales aún en pequeñas cantidades de datos. Durante la fase final, PCA ha sido aplicado a combinaciones de 8 y 12 elementos y en todos ellos se ha conseguido mejorar el porcentaje de éxito y además reducir aún más el número de elementos necesarios. Es necesario resaltar el resultado obtenido al aplicar PCA sobre la combinación 4, ya que si bien posee un porcentaje de éxito inferior al resto de casos de la fase final, lo consigue mediante el uso de tan sólo 3 elementos. En lo referente al resto de casos, PCA ha permitido reducir en 3 su número de elementos y conseguir un porcentaje de éxito de 92% en el mejor de los casos.

Capítulo 5. Conclusiones y líneas futuras

En este capítulo se evaluará hasta que punto se han alcanzado los objetivos planteados en este Trabajo Fin de Grado, así como las ventajas y desventajas del sistema de identificación diseñado en comparación con los anteriormente mencionados en la historia del arte. Además, se presenta un apartado dedicado a exponer diversas líneas futuras de actuación que podrían complementar a éste estudio.

5.1 Conclusiones

Valorando los resultados obtenidos en el presente estudio, se puede concluir que **se han alcanzado los objetivos inicialmente planteados**. En primer lugar se pretendía diseñar un sistema de identificación fiable, que funcionase mediante señales de fácil obtención en hospitales y sin necesidad de métodos invasivos. En este Trabajo Fin de Grado el porcentaje de éxito logrado ha sido del 92%, valor más que aceptable y muy prometedor para futuras líneas de estudio, además, la totalidad de las señales de la base de datos han sido extraídas mediante métodos no invasivos y que pueden ser encontrados en hospitales con cierta facilidad. Cabe destacar como ventaja, la posibilidad de detectar ciertos problemas de salud en el paciente, ya que en caso de producirse cambios drásticos en las señales capturadas se obtendrían fallos en el sistema de identificación, lo cual alertaría a los operarios de dichas irregularidades en el sujeto.

Otro requisito fue no necesitar personal especializado para el manejo del sistema. El sistema desarrollado en este estudio es altamente automatizable, por lo que tanto el proceso de toma de muestras como el de identificación puede ser realizado por cualquier personal sanitario. Respecto a la comparación con otros trabajos, si bien no se ha logrado el nivel de porcentaje de éxito conseguido en los estudios con señales ECG y EEG, cabe destacar como ventaja del presente trabajo que se trata de un sistema de bajo coste y que, al no depender de una única señal, puede considerarse más estable ante fallos en los instrumentos de captura y procesado individual de cada señal. En lo referente al proyecto realizado en la ULPGC, dependiente de 3 señales y con un porcentaje de éxito del 85.55%, se ha conseguido superar con éxito, mejorando el resultado y aumentando la estabilidad del sistema al apoyarse sobre 6 señales.

Por último, en la *tabla 5-1* se presenta una breve comparación entre el sistema realizado en este estudio y los anteriormente mencionados.

	Precio	Eficiencia	Dificultad de Manejo	Dependencia
Sistema ECG	Alto	96%	Difícil	1 señal
Sistema EEG	Alto	98.56% $\pm$ 1.87	Difícil	1 señal
Sistema de 3 señales	Bajo	85.55%	Fácil	3 señales
Sistema de 6 señales	Bajo	92%	Fácil	6 señales

Tabla 5-1. Comparación con el estado del arte

5.2 Líneas futuras

A continuación se plantean varios puntos que se consideran posibles vías de actuación para la mejora del actual estudio:

- Aumentar el número de señales utilizado por el sistema de identificación. Esto aportaría más información al sistema y permitiría el correcto desarrollo de su funcionamiento aún en el caso de ocasionarse fallos puntuales en la toma de una o varias muestras, además de mejorar el porcentaje de éxito.
- Estudiar la eficacia de parámetros distintos a los usados en el presente estudio, con la intención de extraer la máxima cantidad de información de las señales. Esta línea permitiría incorporar al sistema señales que, aunque a priori parezcan no aportar información relevante, puedan contener alguna propiedad que sea útil en la identificación.
- Reducir los tiempos de captura del sistema.
- Ampliar el número de funciones *kernel* utilizadas en la experimentación, con la intención de encontrar aquella que mejore la distinción de datos por parte de la máquina de vectores de soporte.
- Desarrollar un sistema que permita capturar las señales de forma simultánea.
- Considerar otros sistemas de clasificación que no sean SVM y observar su respuesta.

- Adaptar el sistema de modo que pueda alertar de cambios bruscos en las señales del paciente, sin que se produzcan fallos en el proceso de identificación.
- Aumentar el tamaño de la base de datos. Esta línea permitiría medir con mayor precisión la calidad del sistema de identificación.

ANEXOS

ANEXOS

Con motivo de la celebración del congreso 2016 *Ninth International Conference on Contemporary Computing (IC3)*, patrocinado por el IEEE y que se celebrará entre los días 11 y 13 de agosto de 2016 en el Jaypee Institute of Information Technology, situado en Noida, India. Se ha redactado un artículo referente al estudio realizado en el presente Trabajo Fin de Grado. Dicho artículo se ha enviado y está pendiente de contestación, para más detalles puede visitarse la siguiente web: <http://www.jiit.ac.in/jiit/ic3/>

En la siguientes páginas se presenta el citado artículo.

Biometric personal identification system using biomedical sensors

Alain D. Díaz Alonso, Carlos M. Treviso, Jesús B. Alonso

Signal and Communication Department, Institute for Technological Development and Innovation on Communications. University of Las Palmas de Gran Canaria, Campus Universitario de Tafira, sn, Ed. de Telecomunicación, Pabellón B, Desp. 111, E35017, Las Palmas de Gran Canaria, Spain.

Alain.diaz101@alu.ulpgc.es, carlos.travies@ulpgc.es, jesus.alonso@ulpgc.es

Abstract—This paper proposes a biometric system to identify people using common biomedical sensors. A dataset is built by groups of signals acquired from 25 people, and is used in the system. The present proposal applies a combination of principal components analysis and support vector machines to identify people by a group of biometric signals. The testing results have achieved a 92% of correct identification rate.

Keywords—Personal identification; Biometric; Biomedical sensors; Arduino; Signal processing;

I. INTRODUCTION

From the beginning of the civilization, the right people identification has been related to security, where the main purpose was detecting those who broke the law. Since that time, people identification has evolved from visual memory and marking criminals to recognise them, until the use of fingerprints patterns in the last centuries [1].

At present, where the world is based on information and technology, identification systems have evolved to satisfy new kinds of services. Thus, there are several identification systems based on biometrics such as: voice analysis, retina scan and hand geometry [2].

At hospitals some severe problems have happened due to a wrong identification. Among these problems it is possible to mention such as operations to wrong people, errors in drug administration or treatment with mental health problems [3].

The motivation behind this study is solving those problems related to identification in these centres, where patient's identification is commonly performed with wristbands and lists (which are easy to remove or forge). The aim is to design an identification system intended for hospital patients with stable vital signs.

There are previous researches on biometric identification, which are suitable to be use for hospital patients. Among these researches is remarkable the use of the electrocardiogram [4] and the electroencephalogram [5], but these systems are usually very expensive, need a continuous maintenance and are difficult to manage.

This work tries to provide a low cost alternative with an easier to manage system and based on the use of different biometric signals, which gives the system more stability. It also exists a precedent of a study with similar features, this study presents a low cost system based on a group of biometric signals, and achieved an 85.55% of correct identification rate using a combination of galvanic skin response, electrocardiogram and electromyogram [6]. So the aim of this study is also surpasses that rate.

The system must have an appropriate grade of reliability and will be based on biomedical sensors easily founded at any medical centre.

Firstly, it is necessary to create an appropriate dataset for the identification system. This dataset will be performed through an Arduino platform [7] with the medical sensor platform “e-Health” attached. The “e-Health” platform has several sensors intended to capture biomedical signals as hospital's equipment do.

Secondly, data's properties will be analysed in order to detect which of them are more suitable to be use in the identification system.

Finally, these properties will be use to define people by means of their biometric signals and developing the identification system.

II. DESCRIPTION OF PLATFORM AND SENSORS

The e-Health Sensor Shield is a platform designed to perform biometric and biomedical applications, which allows Arduino users capture different kinds of signals from a human body in real time [8]. In this section, the different signals provided by the sensors are described in detail.



Fig 1. Biometric sensors

The platform has 10 different sensors, for this study, it will be use only 6 to capture the following signals:

Pulse and oxygen in blood (SPO2): Oxygen saturation is a measurement of the amount of oxygen dissolved in blood, based on the detection of Haemoglobin (Hb) and Deoxyhemoglobin (HbO2). Two different light wavelengths are used to measure the actual difference in the absorption spectrum of HbO2 and Hb.

This sensor provides a signal that it will be called “pulsyoximetry”, which it is composed of the oxygen saturation signal and the pulse signal interleaved. Each signal will be used as the best appropriate way.

Airflow (breathing): breathing rate is measured by a set of two prongs, which are placed in the nostrils.

Body temperature: the sensor turns heat to voltage, so body temperature can be measured.

Electrocardiogram (ECG): by means of connecting three electrodes (positive, negative and neutral) to the body, muscular functions of the heart are routinely captured. The projection of this captures through the time is named Electrocardiogram.

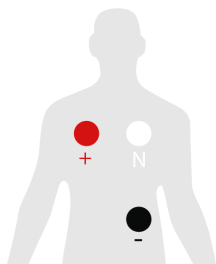


Fig 2. Electrodes positioning

Electromyography (EMG): applying the same process used in ECG, it can be capture muscular functions of other muscles. In this study, the electromyography has been taken from the right biceps.

Galvanic skin response (GSR – sweating): Skin conductance, also known as galvanic skin response (GSR), is a method to measure the electrical conductance between two points of skin, which varies with its moisture level. In this study, the electrical conductance will be measured between forefinger and middle finger.

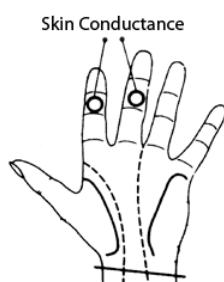


Fig 3. GSR sensor positioning

III. DATASET AND PROCESSING

The dataset has been built from samples of 25 subjects. Every signal is captured from each subject twice, in different periods of time. Thus, it will be obtained two sets of signals for subject (figure 4). One of these sets will be used as a reference (it will define the subject), and the other one will be used to test. Identification system quality will be measured according to how many test sets can be associated with the pertinent reference set.

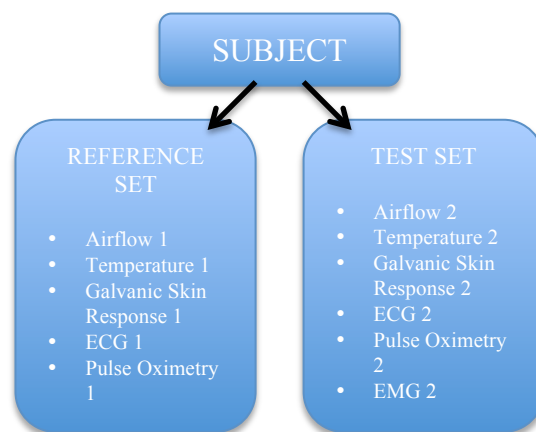


Fig 4. Signals per subject

Each signal capture will last 1 minute, so a set will last 6 minutes. Therefore, every subject measuring will last around 15 minutes (including connections and disconnection times).

The measurement of each feature is performed according to the following instructions:

Pulse and oxygen in blood (SPO2): in sitting or lying position, and the pulsioximeter sensor placed on the forefinger.

Airflow (breathing): in a sitting or lying position, the sensor is placed under the nostrils and the flexible threads positioned behind the ears, as in figure 5. The subject mustn't talk during this capture.



Fig 5. Airflow sensor positioning

Body temperature: with one arm in relaxed position, the sensor is attached to the forefinger using adhesive tape.

Electrocardiogram (ECG): in a sitting or lying position, electrodes are connected according to figure 2. The subject must be relaxed and quiet.

Electromyography (EMG): with the sensor placed as in figure 6 (where M, E and GND are positive, negative and neutral respectively), the arm initially at rest position and extended and palm up.

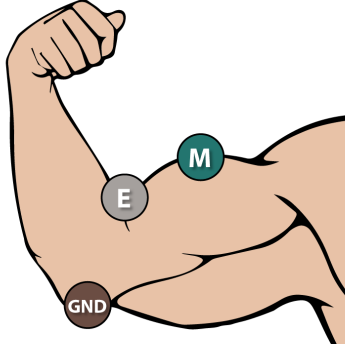


Fig 6. EMG sensor positioning

With the help of a metronome, the subject must contract the biceps when a metronome beat sounds, and stretch it with the next beat. The metronome has a rate of 20 beats per minutes, as every signal capture lasts 1 minute, the subject will performs 10 contractions.

Galvanic skin response (GSR – sweating): with the hand in relaxed position, the sensor is connected according to figure 3. The subject must be relaxed and quiet.

IV. PARAMETRIZATION, FEATURE REDUCTION AND CLASSIFYING SYSTEM

In this section, features are extracted from signals, and later, a feature reduction block is applied, before to use a classification system.

From the previous signals, captured from the subjects and stored as vectors, it is possible to extract several parameters that will be useful to define its pertinent subject.

The following parameters are those chosen for this study depending on the signal nature:

Entropy: is a statistical measure of changes in systems from the previous state to the final state [9]. Entropy is defined as in equation 1.

$$H = - \sum_{i=0}^N P_i * \log_2 P_i \quad (1)$$

Where H is entropy, P_i is the element probability and N the number of elements.

Mean: The arithmetic mean of a finite set of numbers is the characteristic value of a series of quantitative data object of study that is based on the principle of expectation or expected value. The definition is shown in equation 2

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n a_i = \frac{a_1 + a_2 + \dots + a_n}{n} \quad (2)$$

Skewness: the skewness measures the lack of symmetry of a distribution [10]. Skewness is defined for unvarying data as in equation 3.

$$Skewness = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^3 / N}{s^3} \quad (3)$$

Where Y is the element, $\bar{Y}$ the mean, N is the number of elements and s the standard deviation.

Kurtosis: is the measure of the height and sharpness of a peak relative to the rest of the data [11]. Higher values indicate a higher, sharper peak. Lower values indicate a lower, less distinct peak.

Kurtosis is defined for unvarying data as in equation 4.

$$Kurtosis = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4 / N}{s^4} \quad (4)$$

Where Y is the element, $\bar{Y}$ the mean, N is the number of elements and s the standard deviation.

Standard Deviation (Std): is the measure of the amount of variation or dispersion of a set of values. It is calculated according to equation 5.

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5)$$

Now it can be defined the subjects with a vector, which it is compound by the signal parameters previously mentioned. So, each subject will have two vectors, one corresponding to the set of reference signals, and another one corresponding to the set of test signals. In addition, it is possible to put together the vectors as a matrix, where the rows will be subjects and the columns will be parameters. Thus, the system can count on a reference matrix and a test matrix.

When it is tried to find connections among variables, it is possible to see two important problems. On the one hand, having to many variables means it must be considered an important amount of possible correlation coefficients, which makes difficult to visualize the connections among variables. On the other hand, sometimes appears a strong correlation among variables, which supposes less variability and makes more difficult the identification.

In order to solve these problems, it is highly useful to reduce the number of variables using Principal Component Analysis (PCA).

PCA is a technique used to reduce the dimensionality of a data set and emphasize its variation [12]. Its operation consists in finding a new set of orthogonal axis where the data are uncorrelated, this means the data variance is maximum (figure 7).

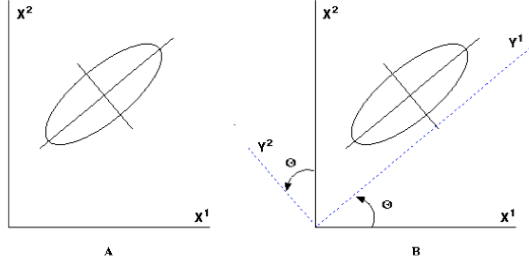


Fig 7. A) X^1 and X^2 are correlated. B) Y^1 and Y^2 uncorrelated

With PCA, it can be got a transformation matrix based on the reference matrix, this transformation matrix will be use to create a new reference matrix and a new test matrix where data are uncorrelated. So, according to equation 6.

$$Y = W^T * X \quad (6)$$

Where X is the matrix with the original axis, W^T is the transformation matrix and Y is the new matrix with new axis where data are uncorrelated.

The new matrix Y has two important advantages. Firstly, its variance is higher than the original one. Secondly, most of the information in the matrix Y is contained in fewer values than the original matrix, so the dimensionality has been reduced.

After extracting the most important parameters of the signals, and reducing its dimensionality, the data is suitable to be processed by the classification system.

In this study, the classification system has been performed using Support Vectors Machines (SVM) [13].

SVM are a group of supervised learning algorithms used in classification and regression analysis. SVM allows marking each class of a training sample set and building a predictive model, which will predicts the class of a new sample. In this case, the reference set will be used as training samples, where the classes are the rows, this is, the subjects. Thus, the aim of SVM is to predict which class the rows of the test set belong to.

Obviously, part of the prediction efficiency depend on the data used as training samples. That is why extracting the most important parameters of the signals and reducing the dimensionality is necessary to make easier the classification process between the reference set and the test set.

V. EXPERIMENTAL METHODOLOGY AND RESULTS

In this section, methodology, different experiments and their pertinent results are explained in detail.

In this study, MATLAB has been used as the working environment. Each capture has been

processed by the Arduino platform, at the same time, data acquisition was transmitted to MATLAB through a serial port communication. Thus, it is possible to visualise data acquisition in real time, to check if the capture has been appropriately performed and to save the signals as vectors. In addition, it will be possible to work with the captures and create the identification system by using MATLAB.

As mentioned above, SVM is chosen as classifier and the system have two sets of signal. One set will be use for training the classifier (reference set), and the other one for testing.

Three experiments have been performed for this study, applying parameterization and PCA as it is observed in section IV, moreover, it has been modified the kernel function one by one to verify the best one in each case.

Kernel functions are a kind of algorithms for pattern analysis used in SVM. Kernel functions allow turning a problem of nonlinear classification in the original dimensional space to an easier linear classification problem in a higher dimensional space. In these experiments, this proposal will use five types of kernel functions: linear, quadratic, polynomial, Gaussian radial basis function (RBF) and multilayer perceptron kernel (MLP). More details about these kernels can be found in [14].

A. Experiment I. Classification with raw signals

In the first experiment, SVM have been trained with vectors composed of concatenated raw signals. The raw signals have too many elements and that complicates the classification process and makes it inefficient. After trying various combinations, vectors composed just by its EMG signal were the most successful option. Thus, for EMG signal and different kernel function it is obtained the results of figure 8.

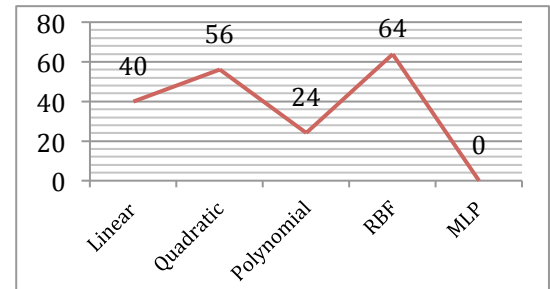


Fig 8. Results in experiment I

For this experiment the best result obtained was 64% of correct identification rate using RBF as kernel function.

B. Experiment II. Classification using parameterization

In the second experiment, the vectors were composed by a combination of 12 parameters extracted from the different signals, as mentioned in section IV, this reduces the number of elements in the vectors and makes the classification process easier.

The parameters have been studied individually, so vectors were composed by the parameters with better results, in order from highest to lowest. In this way, the most successful combination has been the following:

1. Skewness (Oxygen saturation)
2. Kurtosis (Pulsioximetry)
3. Entropy (EMG)
4. Mean (Pulsioximetry)
5. Mean (Oxygen saturation)
6. Kurtosis (Pulse)
7. Std (EMG)
8. Std (Airflow)
9. Std (Temperature)
10. Std (ECG)
11. Kurtosis (ECG)
12. Mean (ECG)

Thus, for this combination and different kernel function it is obtained the results of figure 9.

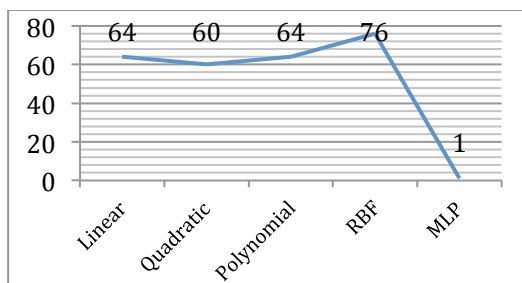


Fig 9. Results in experiment II

For this experiment the best result obtained was 76% of correct identification rate using RBF as kernel function.

C. Experiment III. PCA and parameterization combination

Finally, in the last experiment, PCA was applied to the vectors from the second experiment, reducing the dimensionality and improving the variance. PCA allows containing most of the information in few values. So, the proposal started with vectors of 12 values, but with PCA, it was reduced to a number "N" of values, which is smaller than the previous one.

After a few test, it is determined N=8,9,10 as the best reductions (see figure 10).

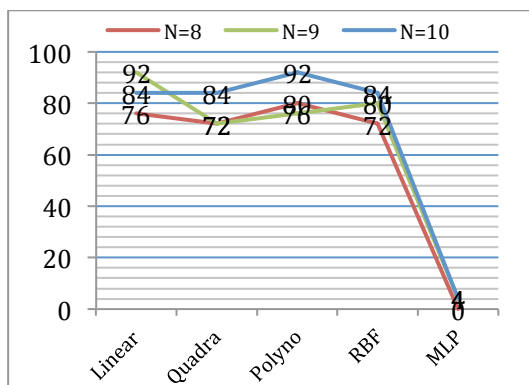


Fig 10. Results in experiment III

For this last experiment, it has been obtained the same result of 92% of correct identification rate using N=9 with linear, and N=10 with polynomial. The aim of PCA is reducing the dimensionality as much as possible, so the case of N=9 with linear is considered as the best result obtained for this experiment.

Finally, it is clearly observed in figure 11, the improvement of experiment III respect to the other two.

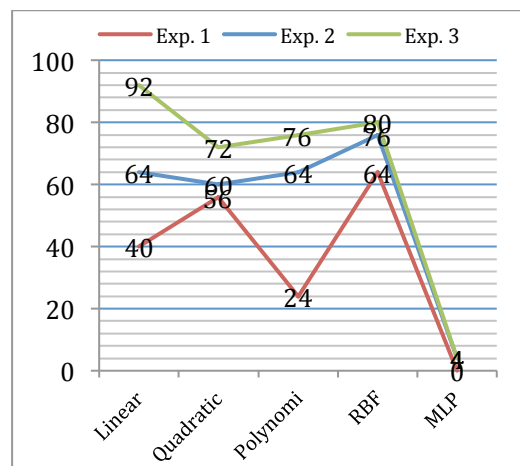


Fig 11. Experiments results comparison.

VI. DISCUSSION

All in all, it is possible to compare advantages and disadvantages of the system respect to other previously mentioned in section I.

On the one hand, in terms of efficiency, the proposed system got a promising result of 92% of correct identification rate using a combination of 6 signals, surpassing the mentioned precedent of section I, which got an 85.55% of correct identification rate with a combination of 3 signals. However, the proposed system doesn't achieve the rates of the systems based on EEG ($98.56 \pm 1.87\%$) and ECG (96%).

On the other hand, EEG and ECG systems are expensive and difficult to manage, in this aspect, the proposed system is much more cheap and easy to use. Furthermore these systems depend on just one signal, on the contrary, this system uses a combination of 6 signals, which provides more stability. All these differences are shown in table 1.

	Price	Efficien.	Handling difficulty	Dependency
ECG system	High	96%	Hard	1 signal
EEG system	High	98.56 ± 1.87	Hard	1 signal
3 Signals System	Low	85.55%	Easy	3 signals
6 Signals System	Low	92%	Easy	6 signals

Table 1. Systems comparison

VII. CONCLUSIONS

The proposed biometric identification system has reached a respectable correct identification rate by using signals easily measurable in a hospital and a low cost platform. The system is highly effective for patient with stable vital signs and very appropriate for patient with physiological pathologies where signals stand out from the common values. In addition, some health problems can be detected because failures would take place in the identification system, as a consequence of signal chances.

In future works, this identification system can be improved by using other kind of biometric signals, so the system can do both, resisting extreme changes on patient health and warning about these changes. Other lines of improvement are using different classification systems and other parameterization stage.

Thus, the system would be suitable for every kind of patient independently of the health status.

REFERENCE

- [1] Simon A. Cole (2002). Suspect Identities: A history of Fingerprints and Criminal Identification.
- [2] Biometric Identification Systems: <http://www.technovelgy.com/ct/Technology-Article.asp?ArtNum=12> (last visit: May 2016)
- [3] Quarterly Data Summary Issue 10 (Nov 2008): Patient identification errors from failure to use or check ID numbers correctly. <http://www.nrls.npsa.nhs.uk/resources/?EntryId45=59854> (last visit: May 2016)
- [4] Lugovaya T.S. Biometric human identification based on electrocardiogram. [Master's thesis] Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russian Federation; June 2005.
- [5] Ramaswamy Palaniappan. EEG Based Biometric Framework for Automatic Identity Verification. Department of Computer Science, University of Essex; 2 April 2007.
- [6] Néstor M. Alemán-Soler, Carlos M. Travieso, Elyoenai Guerra-Segura, Jesús B. Alonso, Malay Kishore Dutta and Anushikha Singh. Biometric Approach Based on Physiological Human Signals. 3rd IEEE International Conference on Signal Processing & Integrated ,9197-9, Noida, Delhi-NCR, India, 11-12 February 2016 (DOI: 10.1109/SPIN.2016.XXXXXXX) (Accepted – In press)
- [7] What is Arduino?: <https://www.arduino.cc/en/Guide/Introduction> (last visit: May 2016)
- [8] Libelium SL. e-Health Sensor Platform V2.0 : <https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical> (last visit: May 2016)
- [9] P.K. Nag. (2002). Basic and Applied Thermodynamics.
- [10] Measures of Skewness and Kurtosis: <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm> (last visit: May 2016)
- [11] Lawrence T. DeCarlo. (1997). On the Meaning and Use of Kurtosis. Fordham University.
- [12] Lindsay I Smith. (2002). A tutorial on Principal Component Analysis.
- [13] Christopher J. C. Burges. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Bell Laboratories, Lucent Technologies
- [14] Souza, César R. "Kernel Functions for Machine Learning Applications." 17 Mar. 2010. Web. <http://crsouza.blogspot.com/2010/03/kernel-functions-for-machine-learning.html> (last visit: May 2016)

BIBLIOGRAFÍA

- [1] Dr. Sislán Rodríguez : Identidad, concepto y evolución de la identificación. Documento de Internet consultado el 6 de abril de 2016 en <http://www.forensesargentina.com.ar/pdf/Historia-de-la-identificacion.pdf>
- [2] Javier Ortega García, Fernando Alonso Fernández, Rafael Coomonte Belmonte. 2008. *Biometría y Seguridad*. Universidad Politécnica de Madrid.
- [3] *Biometric Identification Systems*. Documento de Internet consultado el 9 de mayo de 2016 en <http://www.technovelgy.com/ct/Technology-Article.asp?ArtNum=12>
- [4] Boletín nº30 (Agosto 2009), SEFH: Sociedad Española de Farmacia Hospitalaria. Documento de Internet consultado el 20 de abril de 2016 en [http://www.sefh.es/sefhpdfs/Boletin30\(Agosto2009\).pdf](http://www.sefh.es/sefhpdfs/Boletin30(Agosto2009).pdf)
- [5] Mark R. Chassin, Elise C. Becher. 2002. <<The Wrong Patient>>. *Annals of Internal Medicine* 2002;136:826-883.
- [6] Doll Martin Associates. 2008. *Technology Solutions to Patient Misidentification*. Australian Commission on Safety and Quality in Healthcare.
- [7] Yuresky Rojas Rincón. 2010. *Innovaciones de la Tecnología de Identificación por Radiofrecuencia –RFID, en el Sector Hospitalario*. Universidad Politécnica de Valencia.
- [8] Ramaswamy Palaniappan, Danilo P. Mandic. 2006. *EEG Based Biometric Framework for Automatic Identity Verification*. Imperial College London, London, UK.
- [9] Tatiana S. Lugovaya. 2005. *Biometric Human Identification Based on ECG*. Electrotechnical University “LETI” , Saint-Petersburg, Russian Federation.
- [10] Néstor M. Alemán-Soler, Carlos M. Travieso, Elyoenai Guerra-Segura, Jesús B. Alonso, Malay Kishore Dutta and Anushikha Singh. <<Biometric Approach Based on Physiological Human Signals>>. 3rd IEEE International Conference on Signal Processing & Integrated ,9197-9, Noida, Delhi-NCR, India, 11-12 February 2016 (DOI: 10.1109/SPIN.2016.XXXXXXX) (Accepted – In press)
- [11] <<What is Arduino?>>. Documento de Internet consultado el 6 de mayo de 2016 en <https://www.arduino.cc/en/Guide/Introduction>
- [12] <<e-Health Sensor Platform V2.0>>. Documento de Internet consultado el 12 de mayo de 2016 en https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical#step3_1
- [13] <<El uso de termopares>>. Documento de Internet consultado el 6 de mayo de 2016 en http://de-de.wika.de/upload/DS_IN0023_es_es_62452.pdf
- [14] Xochiti Gallegos, F. Javier Torres Torrija. 1983. <<Biorretroalimentación de la respuesta galvánica de la piel y entrenamiento en relajación: Un estudio piloto>>. *Revista Latinoamericana de Psicología*. Volumen 15. 259-275.
- [15] <<Electrocardiogram>>. Documento de Internet consultado el 11 de abril de 2016 en

- <http://www.floridahealthfinder.gov/HealthEncyclopedia/Health%20Illustrated%20Encyclopedia/1/003868.aspx>
- [16] Rafael Barea Navarro. <<Instrumentación Biomédica>>. Universidad de Alcalá. Documento de Internet consultado el 13 de abril de 2016 en <http://www.bioingenieria.edu.ar/academica/catedras/bioingenieria2/archivos/apuntes/tema%205%20-%20electromiografia.pdf>
 - [17] M. J. Noguerol Casado, A. Seco González. <<Técnicas en AP: Pulsioximetría>>. Documento de Internet consultado el 6 de mayo de 2016 en <http://www.fisterra.com/material/tecnicas/pulsioximetria/pulsioximetria.pdf>
 - [18] Sewell, Martin. 2008. *Structural Risk Minimization*. Londres: University College London.
 - [19] C. Burges, Christopher. 1998. <<A Tutorial on Support Vector Machines for Pattern Recognition>>. *Data Mining and Knowledge Discovery* nº2. 121-167.
 - [20] Betancourt, Gustavo. 2005. <<Las Máquinas de Soporte Vectorial (SVMs)>>. *Scientia et Technica* Año XI. nº27.
 - [21] Thomas Hofmann, Bernhard Schölkopf, Alexander J. Smola. 2008.<<Kernel Methods in machine learning>>. *The Annals of Statistics* vol. 36, nº3. 1171-1220.
 - [22] Souza, César R. “Kernel Functions for Machine Learning Applications.” 17 Mar. 2010. Documento de Internet consultado el 10 de mayo de 2016 en <http://crsouza.blogspot.com/2010/03/kernel-functions-for-machine-learning.html>
 - [23] P. K. Nag. 2002. *Basic and applied thermodynamics*. Nueva Delhi.
 - [24] <<Measures of Skewness and Kurtosis>> Documento de Internet consultado el 12 de mayo de 2016 en <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm>
 - [25] DeCarlo, Lawrence T. 1997. <<On the Meaning and Use of Kurtosis>>. *Psychological Methods*. Vol 2, nº3. 292-307.
 - [26] Smith, Lindsay I. 2002. *A tutorial on Principal Component Analysis*.

II. PLIEGO DE CONDICIONES

PLIEGO DE CONDICIONES

En este apartado se especifican los requerimientos software y hardware necesarios para la correcta ejecución de las aplicaciones implementadas.

1. Especificaciones hardware.

El sistema mínimo necesario para la correcta ejecución de la aplicación desarrollada es:

- Ordenador Personal: cuyas características principales son un procesador *Intel Core i5* a 2,5 GHz, con 4 GB de RAM y tarjeta gráfica *Intel HD Graphics 4000* de 1GB. La base de datos construida para 25 sujetos y los archivos adicionales no ocupan un tamaño superior a los 250 MB, por lo que el disco duro debe contar con al menos esta capacidad.
- Impresora.
- Placa Arduino UNO Revisión 3.
- Plataforma e-Health.
- Sensor de flujo de aire.
- Sensor de respuesta galvánica de la piel.
- Sensor de electrocardiograma.
- Sensor de temperatura.
- Sensor de electromiograma.
- Sensor de pulso y oxígeno en sangre.

2. Especificaciones software.

- Sistema operativo: OS X 10.9.5
- Matlab 8.1.0.604 (R2013a)
- Microsoft Office 2007: Word (empleado para la redacción del proyecto), Excel (empleado para la representación de gráficos) y PowerPoint (empleado para la presentación).
- Base de datos de señales fisiológicas de 25 usuarios.
- Arduino 1.0.6

III. PRESUPUESTO

PRESUPUESTO

Alain David Díaz Alonso, autor del presenta Trabajo de Fin de Grado, declara que:

El Trabajo de Fin de Grado con título “*Sistema de identificación de personas mediante sensores biomédicos*”, desarrollado en la Escuela de Ingeniería de Telecomunicaciones y Electrónica, de la Universidad de las Palmas de Gran Canaria, tiene un coste de desarrollo total de 16084,74 € correspondiente a la suma de las cantidades consignadas a los apartados considerados a continuación.

Fdo.: Alain David Díaz Alonso



Las Palmas de Gran Canaria a 15 de julio de 2016

P1. Desglose del Presupuesto

Para la realización del presupuesto se han seguido las recomendaciones del Colegio Oficial de Ingenieros de Telecomunicación (COIT) sobre los baremos **orientativos** mínimos para trabajos profesionales en 2009. El presupuesto se ha desglosado en varias secciones en las que se han separado los distintos costes asociados al desarrollo del TFG, fijándose la duración del mismo en 4 meses. Estos costes se dividen en:

- Recursos materiales.
- Trabajo tarifado por tiempo empleado.
- Costes de redacción del Trabajo Fin de Grado.
- Material fungible.
- Derechos de visado del COIT.
- Aplicación de impuestos.

P2. Recursos materiales

Entre los recursos materiales utilizados para la realización de este TFG se incluyen las herramientas software de desarrollo de los algoritmos del sistema, los paquetes software usados para la redacción de la memoria y el sistema operativo bajo el que se ejecutó el trabajo. Asimismo, se incluyen los equipos hardware utilizados para dar soporte a estas herramientas.

Se estipula el coste de amortización para un período de 3 años. Para ello, se utilizará un sistema de amortización lineal o constante, en el que se supone que el inmovilizado material se desprecia de forma constante a lo largo de su vida útil. La cuota de amortización anual se calcula usando la siguiente fórmula:

$$Cuota\ anual = \frac{Valor\ de\ adquisición - Valor\ residual}{Número\ de\ años\ de\ vida\ útil}$$

(PR1)

El “valor residual” es el valor teórico que se supone que tendrá el elemento de su vida útil.

P2.1. Recursos Hardware

Para la ejecución de este estudio las herramientas hardware que se han utilizado son los siguientes:

- Un ordenador portátil MacBook Pro Intel Core i5, con 4 GB de memoria RAM y 500 GB de disco duro.
- Impresora.
- Placa Arduino UNO.
- Plataforma de sensores e-Health V2.0

Coste de las herramientas Hardware			
Descripción	Coste Total	Valor Residual (3 años)	Valor amortización (4 meses)
Ordenador portátil	1049 €	209,8 €	93,24 €
Impresora	68,6 €	25 €	4,84 €
Placa Arduino UNO	9,68 €	1,93 €	0,86 €
Plataforma de sensores e-Health V2.0	450 €	90 €	40€
Total de Costes			138,94 €

Tabla P.1. Costes de herramientas hardware

P2.2. Recursos Software

Las herramientas software utilizadas en el presenta trabajo fueron:

- Matlab® v.8.1.0.604 (R2013a)
- Microsoft Office® 2011 para Mac.
- OSX® v.10.9.5
- Arduino® v.1.0.6
- Base de datos de señales fisiológicas de 25 usuarios.

Coste de las herramientas Hardware			
Descripción	Coste Total	Valor Residual (3 años)	Valor amortización (4 meses)
Matlab® v.8.1.0.604 (R2013a)	2000 €	0 €	222,23 €
Microsoft Office® 2011	250 €	0 €	27,77 €
Total de Costes			250 €

Tabla P.2. Costes de herramientas software

El resto de software citado durante el presente trabajo y que no aparece en la *tabla P.2* no ha tenido ningún tipo de coste, como son los casos de OSX® v.10.9.5, Arduino® v.1.0.6 y la base de datos de 25 usuarios.

P3. Trabajo tarifado por tiempo empleado

En este Trabajo Fin de Grado se han invertido 300 horas en las tareas de especificación desarrollo y documentación necesarias para la elaboración del mismo. El importe de las horas de trabajo empleadas para la realización del estudio se calcula siguiendo las recomendaciones del COIT:

$$H = C_t * 74.88 * H_n + C_t * 96.72 * H_e \quad (\text{PR2})$$

Donde:

- H son los honorarios totales por el tiempo dedicado.
- H_n son las horas normales trabajadas (dentro de la jornada laboral).
- H_e son las horas especiales.
- C_t es un factor de corrección función del número de horas trabajadas.

La totalidad de las 300 horas invertidas han sido realizadas dentro del horario normal. Según el COIT, el coeficiente C_t tiene un valor variable en función del número de horas empleadas de acuerdo con la siguiente tabla:

Horas empleadas	Factor de corrección
Hasta 36 horas	1.00
Desde 36 horas a 72 horas	0.90
Desde 72 horas a 108 horas	0.80
Desde 108 horas a 144 horas	0.70
Desde 144 horas a 180 horas	0.65
<u>Desde 180 horas a 360 horas</u>	<u>0.60</u>
Desde 360 horas a 540 horas	0.55

Tabla P.3. Factor de corrección en función del número de horas invertidas

Como se puede observar el número de horas está en el rango de más de 180 y menos de 360, por lo que según la *tabla P.4* el factor de corrección es de $C_t = 0,60$. Con ello, la ecuación del importe de horas de trabajo resulta de la siguiente forma:

$$H = 0.60 * 74.88 * 300 + 0.60 * 96.72 * 0 = 13478.40\text{€} \quad (\text{PR3})$$

Los honorarios totales por tiempo dedicado libres de impuestos ascienden a *trece mil cuatrocientos setenta y ocho euros con 40 céntimos (13478,40 €)*.

P4. Costes de redacción del Trabajo Fin de Grado

El importe se calcula de acuerdo a la siguiente expresión:

$$R = 0.07 * P * C_n \quad (\text{PR4})$$

Donde:

- P es el presupuesto.
- C_n es el coeficiente de ponderación en función del presupuesto.

En la siguiente tabla se muestra el presupuesto calculado hasta el momento:

Descripción	Costes
Recursos Hardware	138,94 €
Recursos Software	250 €
Trabajo Tarifado por Tiempo Empleado	13478,40 €
Total de Costes	13867,34 €

Tabla P.4. Costes de ejecución material

El presupuesto calculado hasta el momento asciende a 13867,34 €. Como el coeficiente de ponderación para presupuestos menores de 30.050 € viene definido por el COIT con un valor de 1.00, el coste derivado de la redacción del Trabajo Fin de grado es de:

$$R = 0.07 * 13867,34 * 1,00 = 970,70 \text{ €}$$

Por tanto el coste libre de impuestos derivado de la redacción del Trabajo Fin de Grado es de *novecientos setenta con setenta céntimos (970,70 €)*.

P5. Material fungible

Además de los recursos hardware y software, en este trabajo se han empleado otros materiales, como son los folios y el tóner de la impresora entre otros, que se especifican como material fungible. En la tabla P.6. se muestran los costes generados por estos recursos.

Descripción	Costes
Folios	10,00 €
Tóner de la impresora	50,00 €
Encuadernación	10,00 €
Electrodos	28,80 €
Total de Costes	98,80 €

Tabla P.5. Costes del material fungible

P6. Derechos de visado del COIT

Como se mencionó al comienzo del bloque de Presupuesto, todos los cálculos son orientativos, por lo que se han añadido también los derechos de un supuesto visado por el COIT.

Los gastos del visado del COIT se tarifican mediante la siguiente expresión:

$$V = 0.006 * P * C_v \quad (\text{PR5})$$

Donde:

- P es el presupuesto del TFG.
- C_v es el coeficiente reductor en función del presupuesto del trabajo.

El presupuesto P, calculado hasta el momento asciende a la suma de los costes de ejecución material, de redacción y de material fungible.

$$P = 13867,34 + 970,70 + 98,80 = 14936,84 \text{ €}$$

Como el coeficiente de ponderación para presupuestos menores de 30.050 € viene definido por el COIT con un valor de 1.00, el coste de los derechos de visado del trabajo asciende a la cantidad de:

$$V = 0,006 * 14936,84 * 1,00 = 89,62 \text{ €}$$

Por tanto el coste de los derechos de visado del trabajo asciende a *noventa euros con treinta y tres céntimos* (**89,62 €**).

P7. Gastos de tramitación y envío.

Los gastos de tramitación y envío están fijados en 6,01 €.

P8. Aplicación de impuestos.

El coste total del Trabajo Fin de Grado, antes de aplicarle los correspondientes impuestos, asciende a 15032,47 €, a lo que hay que sumarle el 7% de IGIC, con lo que el coste definitivo del Trabajo Fin de Grado es:

Costes Totales del TFG		
Descripción	Coste Parcial	Total
Recursos Materiales		388,94 €
Software	250 €	
Hardware	138,94 €	
Coste de Ingeniería		13478,40 €
Coste de Redacción		970,70 €
Material Fungible		98,80 €
Derechos de Visado		89,62 €
Tramitación y Envío		6,01 €
Subtotal:		15032,47 €
Aplicación de Impuestos (7% I.G.I.C)		1052,27 €
Total de Costes		16084,74 €

Tabla P.6. Costes totales del TFG

El presupuesto total asciende a la cantidad de *dieciséis mil ochenta y cuatro euros con setenta y cuatro céntimos (16084,74 €)*

Autor: Alain David Díaz Alonso

Fdo.:



Las Palmas de Gran Canaria a 15 de julio de 2016