TRABAJO DE FIN DE GRADO

MEDIDAS DE INTELIGIBILIDAD PARA PREDICCIÓN DEL GRADO DE PARKINSON

Grado en Ingeniería de Sistemas de Comunicaciones

Autor: Blanca Valdivielso Paño
Tutor: Ascensión Gallardo Antolín

Agradecimientos

Me gustaría mostrar mi agradecimiento a todas las personas que, directa o indirectamente, han formado parte en la realización de esta memoria de trabajo de fin de grado.

En primer lugar, a mi tutora, Ascen, por confiar en mí para este proyecto, por el compromiso y la dedicación adquiridos durante todos estos meses, que no han sido pocos.

A BIP, la empresa en la que llevo trabajando ya dos años, en los que he estado compaginando estudios y trabajo. Por haberme apoyado, comprendido y ayudado en todo lo posible para poder finalizar mis estudios. A los compañeros de la empresa, de la cual me llevo muy buenos amigos, pues todos ellos han colaborado a hacerme una mejor profesional.

A mis amigos y compañeros de la carrera. Ha sido un largo periodo en el que hemos pasado muchas horas juntos, tanto fuera como dentro de la universidad. Todos nos hemos ayudado mutuamente, y colaborando entre nosotros para terminar esta etapa de la vida.

A ti, Javier, por haber sido una parte fundamental en mi día a día. Comenzamos como compañeros de clase, y ahora, como pareja, me has ayudado tanto en lo académico como en lo personal; apoyándome durante todo el tiempo que he tenido que dedicar al TFG, a los exámenes y otro aspectos de mi vida.

Y por último, como no, a mi familia, en especial, a mis padres y hermanas. Ha sido un largo periodo, no solo durante este TFG, sino durante toda la carrera; hemos compartido buenas y malas situaciones, y en todas, han estado ahí. Por darme la educación que ha permitido convertirme en la persona que soy hoy en día.

A todo vosotros, gracias.

Resumen

La comunicación ha sido un instinto básico en el desarrollo del hombre, las personas tendemos a interactuar con el medio, y, por tanto, con nuestros iguales, es por ello, que es imprescindible lograr un proceso comunicativo donde prime el entendimiento. Unos de los factores para conseguir un correcto entendimiento entre interlocutores a través de la comunicación oral, es la inteligibilidad del habla, que en ocasiones puede verse afectada a causa de la denominada disartria.

A lo largo de esta memoria, se hablará de dicha disartria y de las implicaciones que tiene en personas con enfermedad de Parkinson. Es la segunda enfermedad más extendida después del Alzheimer, y por tanto, afecta a más de 300.000 personas tan solo en España. Cifra que irá aumentando debido al envejecimiento de la población.

Con este Trabajo Fin de Grado, se pretende elaborar un predictor que sea capaz de estimar el grado de inteligibilidad de señales de voz. Se ha utilizado la base de datos "Universal Access" que contiene audios de diversos interlocutores con disartria y sus correspondientes etiquetas con el grado de inteligibilidad que se obtuvieron de forma subjetiva por una serie de evaluadores. La disartria se presenta como síntoma habitual en personas con Parkinson, por ello se ha elegido esta base de datos para el desarrollo y evaluación del sistema.

El sistema predictor de inteligibilidad que se ha desarrollado consta de una serie de procesos como la extracción de las características acústicas o *features*, selección de características, regresión y evaluación de los resultados, entre otros. Tras insertar las señales por el predictor, se obtiene una salida concreta con la predicción del grado de inteligibilidad del paciente, que se evalúa en base a la correlación de Pearson y la raíz del error cuadrático medio.

Se han realizado diferentes tipos de pruebas, comparadas con artículos relacionados o de forma independiente. En todas ellas, los resultados han presentado un alto grado de aproximación, alcanzando los objetivos planteados en el proyecto.

.

Abstract

Communication has been a basic instinct in the development of human, people tend to interact with the environment, and therefore with our peers, that is why it is essential to achieve a communicative process where the understanding prevails. One of the factors to achieve a correct understanding between interlocutors through oral communication is speech intelligibility, which can sometimes be affected by the so-called dysarthria.

Throughout this report, we will discuss such dysarthria and the implications it has on people with Parkinson's disease. It is the second most widespread disease after Alzheimer's disease, and therefore affects more than 300,000 people just in Spain. This figure will increase due to the aging of the population.

With this Final Degree Project, we pretend to elaborate a predictor that is capable of estimating the degree of intelligibility of speech signals. We have used the “Universal Access” database that contains audios of several speakers with dysarthria and their corresponding labels with the intelligibility score that were subjectively obtained by a set of evaluators. Dysarthria presents as a common symptom in people with Parkinson's disease, so this database has been chosen for the development and assessment of the system.

The intelligibility prediction system that has been developed consists of several processes as the extraction of acoustic characteristics or features, feature selection, regression and results evaluation, among others. After feeding the signals into the predictor, we obtain an output with the prediction of the intelligibility degree of the patient, which is evaluated according to the Pearson correlation and the root mean square error.

Different types of tests have been performed, compared to related papers or independently. In all of them, the results have presented a high degree of approximation, achieving the objectives of the project..

Índice general

Índice de figuras

Índice de tablas

Índice de ecuaciones

Palabras clave

Intelligibility
Dysarthria
Acoustic features
SVR Regression

Acrónimos

PD	**P**arkinson **D**isease
SVM	**S**upport **V**ector **M**achine
SVR	**S**upport **V**ector **R**egression
TFG	**T**rabajo de **F**in de **G**rado
LPC	**L**inear **P**redictive **C**oding

1. Introducción

El objetivo de este primer capítulo es situarnos en el contexto de este trabajo de fin de grado. Constará de diferentes partes, se comenzará con una introducción donde se describirá el problema en cuestión, el impacto y cómo se va a llevar a cabo. Después se procederá a exponer la motivación que me ha llevado a realizar este proyecto en concreto, junto con los objetivos que se esperan conseguir de él. Se añadirá un apartado con la estructura de todo el trabajo, parea que el lector comprenda y vea de forma clara todos los argumentos o secciones. Y, por último, se añadirá el contexto o impacto socio económico.

1.1. Introducción

Desde siempre, la comunicación ha sido un instinto básico en el desarrollo del hombre. Por naturaleza, las personas tendemos a interactuar con el medio, y, por tanto, con nuestros iguales. A lo largo del tiempo, han existido diferentes modos a través de los cuales nos hemos podido comunicar, empezando por gruñidos, señales de humo, pinturas, etc. y terminando en lo que actualmente conocemos como los sonidos articulados, es decir, el lenguaje. Por ello es imprescindible lograr un proceso comunicativo donde prime el entendimiento.

Unos de los factores para conseguir un correcto entendimiento entre interlocutores a través de la comunicación oral, es la inteligibilidad del habla, que en ocasiones puede verse afectada a causa de la denominada disartria. Ésta consiste en un conjunto de alteraciones en el habla, producidas por una lesión neurológica. Dichas alteraciones se producen mayoritariamente en el control muscular, afectando a partes fundamentales involucradas en la producción del habla, lo que provoca un problema para el entendimiento del paciente.

Al estar provocada por lesiones neurológicas, la disartria es una característica muy común en enfermedades tales como el Alzheimer, parálisis cerebral, esclerosis múltiple, etc. En este caso, y como bien indica el título del proyecto, nos vamos a centrar en ver cómo afectaría a personas con la enfermedad de Parkinson, ya que es uno de los síntomas más habituales. Se estima que son afectados entre un 60% / 80% de los pacientes que sufren esta enfermedad, manifestándose por un tono bajo, una reducción del ritmo y la fluidez del habla, o incluso por repeticiones de sílabas o frases.

En los exámenes médicos que tienen por objeto determinar el grado de severidad de la enfermedad de Parkinson en un paciente, uno de los factores que se miden de forma subjetiva es el nivel de inteligibilidad de su habla. Esto requiere que el paciente se desplace de forma periódica al hospital o centro médico para la realización

de estas pruebas, con toda la problemática asociada (dificultad de desplazamiento, esperas, etc.). Por tanto, sería de gran utilidad que este tipo de seguimiento pudiera hacerse de forma remota. En este contexto, el objetivo de este trabajo es el diseño y desarrollo de un sistema automático que determine el grado de inteligibilidad de un paciente mediante el análisis de varias de sus elocuciones.

Este sistema podría determinar de una forma objetiva, si un paciente, tras haberse sometido a tratamiento, consigue mejorar o empeorar su inteligibilidad a lo largo del tiempo, lo que proporcionaría una información muy útil para inferir si dicho tratamiento está produciendo mejoras en su enfermedad o si, por el contrario, es necesario modificarlo. En definitiva, se cree que sería de gran utilidad para el análisis, detección y seguimiento de la evolución de la disartria, y lo se ha querido aplicar en este caso a personas con enfermedad de Parkinson, por el elevado número de pacientes que lo padecen en la actualidad y muy posiblemente, lo padecerán en el futuro.

Desde el punto de vista práctico, el sistema diseñado en este trabajo se ha implementado usando el programa de cálculo científico Matlab. Para su desarrollo, se ha utilizado una base de datos con señales de voz pronunciadas por una serie de locutores con diferentes grados de inteligibilidad. Ha sido probado en diversos experimentos para para predecir el grado de inteligibilidad de nuevos locutores. En este proceso de evaluación del sistema se han conseguido resultados muy prometedores y mejores que los de otro sistema similar.

1.2. Estructura

Este punto nos sirve para exponer de forma global, como va a estar estructurado nuestro proyecto. Es una manera de entender mejor todo el desarrollo que se ha llevado a cabo y situar al lector para su correcto entendimiento.

En el **Capítulo 1,** se comienza con una introducción general sobre todo el trabajo de fin de grado, situándolo en su contexto, con sus principales objetivos y unas breves palabras sobre cómo se va a llevar a cabo. A su vez, el capítulo se divide en varios apartados, en los cuales se profundiza de manera específica en cada uno de ellos. En el apartado **1.3, "**Objetivos y motivación", se plantean todos los objetivos que tiene el proyecto, además de las motivaciones que me han llevado a realizar este trabajo en concreto.

Tenemos un pequeño apartado, **"**Requisitos y restricciones", donde se detallan brevemente los requisitos y algunas de las restricciones que se han encontrado a la hora de la elaboración del trabajo.

Por último, en los epígrafes, **"Entorno socio-económico"** y "**Marco regulador**", se expone de un modo más técnico, cómo se regulan todos los aspectos de los que habla el trabajo, y cuál es su entorno. En él, entre otros aspectos, se pueden encontrar datos numéricos referentes al proyecto.,

Pasando al **Capítulo 2**, encontramos el **"Estado del arte"**. Consiste esencialmente, en explicar de un modo más profundo, los aspectos que envuelven el proyecto. En este caso, extraer características acústicas para predecir el grado de inteligibilidad en pacientes con disartria, por tanto, se tratará de definir este concepto, a quién afecta, tipos, etc., del mismo modo, nos referiremos al Parkinson, a sus síntomas, dando cifras, posibles diagnósticos y soluciones. Para finalizar el capítulo, encontramos un apartado titulado **"Estudios relacionados",** donde se encuentran algunos artículos de otros autores, que nos han servido de apoyo a la hora de elaborar este proyecto. Se han utilizado como punto de partida para el diseño de nuestro sistema, y como referencia a la hora de evaluar resultados.

El **Capítulo 3** consiste en explicar el **"Diseño de la solución"**. En él, se pueden encontrar cada una de las "partes" principales de las que constará el sistema (base de datos, features, etc.) y la justificación de cada una de ellas.

Siguiendo con el **Capítulo 4**, se encuentra la **"Implementación"** donde se relatan los detalles técnicos de cada uno de los módulos del sistema desarrollado. Se encuentran, tanto las funciones, como los procesos oportunos que se han tomado para acondicionar la base de datos, los criterios y relaciones que nos han llevado a elegir un determinado número de parámetros acústicos. También se explica cómo se han evaluado los resultados y las herramientas utilizadas.

Uno de los capítulos más significativos en el proyecto es el **Capítulo 5 "Pruebas y resultados"**. Aquí se explican cada una de las pruebas que se han realizado. Consta de dos apartados, uno contiene las pruebas realizadas que tienen como base de referencia un artículo/ensayo de otro autor, del cual hemos utilizado un experimento similar y buscamos obtener unos mejores resultados. Además, se encuentra el apartado con las pruebas independientes, con todas aquellas, que al contrario de las comentadas anteriormente, no encuentran un referente en artículo alguno; se han llevado a cabo de forma exclusiva para este trabajo. Junto con la descripción detallada de estos experimentos, se expone una breve conclusión de cada uno de ellos, es decir, si los resultados han sido positivos, o por el contrario, no se ha logrado alcanzar el éxito deseado.

Después de explicar las pruebas del proyecto, pasamos a un **Capítulo 6** donde se detallan aspectos generales del trabajo, como son la "**Planificación y presupuesto**". En él, se explican cuáles han sido las fases en la elaboración del trabajo, con un breve resumen sobre lo que abarca cada una de ellas. Se puede ver en la **Figura 18** un diagrama Gantt que ilustra los periodos y las fases el proyecto. Una vez definidos los hitos o fases, se indican los recursos utilizados en toda la elaboración del trabajo, divididos en recursos físicos, software y recursos físicos. Se describen cada uno de los grupos, y se muestran los importes que aplican en sus correspondientes tablas. Además de las tablas desglose, se muestra una tabla final (**Tabla 9**) con el presupuesto global para la financiación de este TFG.

Tras haber descrito todas las secciones del trabajo que han permitido seguir un orden lógico de los procesos, y por tanto, han ayudado a comprender mejor el orden y la finalidad del proyecto, se termina con el **Capítulo 7** de **"Conclusiones y líneas**

futuras". Como el nombre indica, se exponen de forma global las conclusiones extraídas de todo el trabajo, más concretamente, si cada uno de los objetivos planteados en el primer capítulo se han llevado a cabo. Se plantean además, posibles alternativas que han ido surgiendo a lo largo del proceso, y que podrían ser interesantes para futuros estudios.

Para terminar, en el **Capítulo 8** se detallan todas las referencias que se han necesitado para la elaboración de la memoria. Se han tenido que consultar diferentes artículos y ensayos para recaudar toda la información y poder redactar la memoria. Todas ellas forman la Bibliografía de este TFG.

1.3. Objetivos y motivación

El objetivo principal de este trabajo es la realización de un sistema que nos permita estimar el grado de inteligibilidad de un locutor mediante el análisis de varias señales de su voz. Para ello, se extraerán una serie de características acústicas de dichas señales, que estén correladas con el grado de inteligibilidad del paciente.

Actualmente son muchas las enfermedades que tienen entre sus síntomas, problemas de inteligibilidad, ya sea por dificultades en la respiración, neurológicas, auditivas, o en zonas como la lengua o la laringe. Como hemos comentado, la comunicación es fundamental hoy en día, y por ello se considera que podría ser una herramienta muy útil e interesante para llevar a cabo un control objetivo de la evolución del habla de un paciente, que se supone, estará correlada con la evolución de la enfermedad que produce dichos problemas del habla. Hay que tener en cuenta que habitualmente, lo realmente interesante en el sector médico, no es una medida exacta del grado de inteligibilidad del paciente, sino su utilidad como herramienta para poder medir la evolución temporal de dicha inteligibilidad.

La motivación que me ha llevado a realizar este trabajo, además de hacer un proyecto técnico, estudiando diferentes parámetros y características del habla, es que dicho estudio, pueda tener una finalidad médica.

Sería gratificante poder ver que un programa elaborado con Matlab pueda tener un impacto directo en el seguimiento de una persona con Parkinson. Incluso podría ser un avance a la hora de las revisiones médicas ya que, con un programa como este, el paciente podría auto-evaluarse desde su domicilio siguiendo una serie de pautas y obtener el resultado a los pocos minutos. Además de la comodidad que supone esto para personas con plena capacidad de movimiento, las personas que sufren de Parkinson, suelen tener problemas de movilidad, y por tanto, en muchos casos evitaría que personas con sillas de ruedas o gente de avanzada edad, tuviera que verse obligada a asistir periódicamente a un hospital.

Concluir que, a pesar de todas las herramientas que se usen y de los parámetros que se consigan extraer de las señales de voz consiguiendo un éxito en el

proyecto, lo más importante y motivador es el impacto humano que se podría obtener, consiguiendo mejoras de peso en el sector médico.

1.4. Requisitos y restricciones

En este apartado se expondrán los requisitos principales que son necesarios para el sistema, junto con las restricciones correspondientes.

Según las especificaciones iniciales de nuestro sistema, hay que tener presente que éste trabaja de la siguiente manera. En primer lugar, es necesario introducir una señal de voz a la entrada para el análisis. Ésta presenta como requisito, estar grabada en mono y con una frecuencia de muestreo de 16KHz. Tras esta entrada, el sistema lo analizará, obteniendo una salida, a través de la cual se obtendrá el grado de inteligibilidad de dicha elocución. En cuanto a restricciones, será necesario que las señales de voz a la entrada del sistema no contengan segmentos largos de silencio o ruido, ya que, en caso contrario, podrían falsear los resultados. Por este motivo, será necesario realizar un preprocesamiento de dichas señales de entrada para que cumplan este requisito. Este proceso se comentará en capítulos posteriores con más detalle.

En cuanto a otro tipo de restricciones más generales, sería interesante encontrar unos métodos de trabajo y algoritmos que no fueran muy costosos a nivel computacional, ya que de por sí, el programa requiere gran cantidad de tiempo en procesar el nivel de datos disponible.

1.5. Entorno socio-económico

La enfermedad de Parkinson se sitúa como el segundo trastorno más extendido después del Alzheimer. Se estima que ya hay más de 6 millones de personas diagnosticadas en todo el mundo, número que asciende cada año debido al envejecimiento de la población.

Este apartado, por tanto, tiene una estrecha relación con la motivación que me ha llevado a realizar el trabajo. Analizando esta cifra, la solución que planteamos al desarrollar este proyecto, tendría un grandísimo impacto social, tanto en las personas diagnosticadas de Parkinson, como en las que padecen disartria a causa de otras muchas enfermedades, las que todavía no han sido diagnosticadas o aquellas que están en unas primeras fases.

Tan solo en España [1], se estima que más de 300.000 personas lo padecen, y que, por tanto, la mayoría presentarán disartria, ya que es uno de sus principales síntomas. Se ha estimado que en España, el gasto por paciente asciende a 17.000€, cifra que se irá incrementando por una mayor esperanza de vida.

Con un sistema como el planteado en este proyecto se colabora en el ámbito de la salud y podría repercutir positivamente en el aspecto económico, tanto a nivel individual del paciente como a nivel estatal.

1.6. Marco regulador

En este apartado se pretenden exponer regulaciones o leyes aplicables a cualquier aspecto de nuestro proyecto.

Para la realización del proyecto, ha sido necesario disponer de una base de datos de voz de personas con disartria, lo cual sería la parte más conflictiva en caso de haberla tenido que obtener de manera exclusiva de algún hospital o laboratorio. En esta ocasión no ha sido así, ya que la base de datos utilizada ha sido generada por la Universidad de Illinois y es de acceso público para otras universidades y centros de investigación. A pesar de esto, existen una serie de arreglos que se tienen que llevar a cabo. En primer lugar, la tutora responsable de la supervisión del proyecto, tuvo que solicitar dicha base de datos a la entidad responsable, indicando que ésta iba a ser utilizada únicamente con fines de docencia e investigación y no iba a ser distribuida a terceros. Por otra parte, las personas cuyas voces aparecen en la base de datos, tuvieron que firmar unos documentos con la entidad, con la conformidad sobre la distribución de dichos audios a otras instituciones con fines de investigación. Existen varios locutores que no estuvieron de acuerdo con ello, es por eso que no disponemos de la base de datos al completo.

Nuestro trabajo ha sido elaborado con la herramienta MATLAB y ayudado por el pack Microsoft Office, ambos disponibles para todos los alumnos por la Universidad Carlos III de Madrid, por tanto, no ha sido necesario la obtención de ningún permiso, ni necesidad de regulación.

Conforme se han ido exponiendo las premisas y toda la información general acerca de nuestro proyecto, ha sido necesario acceder a diferentes ensayos o *papers* para detallar dicha información. Nos hemos basado en los trabajos de otros expertos en la materia, y se aplica la regulación sobre la propiedad intelectual. En cualquier apartado que se hayan utilizado, ya sea de forma literal, o no, los estudios o trabajos de otro, han quedado debidamente referenciados. Esto se rige por el Real Decreto Legislativo 1/1996 del 12 de abril.

Por otra parte, a nivel nacional, también interviene la Ley Orgánica 15/1999, de 13 de diciembre*, de Protección de Datos de Carácter Personal* y a su Reglamento de desarrollo, aprobado por Real Decreto 1720/2007, de 21 de diciembre.

En la página web de la Universidad Carlos III también podemos encontrar las Directivas Comunitarias sobre la materia, especialmente la Directiva 2001/29/CE sobre derechos de autor y derechos afines en la sociedad de la información.

2. Estado del arte

Con el fin de comprender mejor este trabajo de fin de grado, procedemos a describir varios aspectos determinantes. En primer lugar, es necesaria una introducción a la enfermedad del Parkinson, con las características físicas que conlleva, y que por tanto nos serán definitivas para el análisis. Tras ello, hablaremos de la disartria, ya que tiene una relación directa y se trata de uno de los principales síntomas que se presenta en los pacientes. También enunciaremos diferentes artículos y ensayos llevados a cabo que guardan cierta similitud con este proyecto. Con ellos hemos podido conocer mejor la enfermedad y sus síntomas, y, por tanto, hemos podido usarlo a la hora de descartar opciones que ya habían sido estudiadas y guiarnos por características que habían resultado exitosas.

2.1. Enfermedad de Parkinson

2.1.1. Introducción

La enfermedad de Parkinson es una de las cuales afectan al sistema nervioso, mayormente a las capacidades motoras. Es una enfermedad degenerativa, y a pesar de los estereotipos, no es una enfermedad fatal, es decir, la causa de muerte del afectado, no será el Parkinson.

Se debe a la muerte de unas neuronas específicas, que producen dopamina, que se encarga de controlar el movimiento en el organismo. Al producirse un descenso en los niveles de dopamina, el organismo no es capaz de controlar correctamente los movimientos, lo que se traduce en síntomas como temblores, rigidez, etc. Que a su vez afecta a partes como la lengua y laringe, y se traduce en problemas en el habla, provocando la denominada disartria, que tomará más protagonismo en los próximos capítulos de esta memoria.

Se estima que hay aproximadamente unos 6 millones de personas diagnosticadas de Parkinson en todo el mundo, y 300.000 tan solo en España, cifras que se cree que irán aumentando debido al envejecimiento de la población.

La causa de esta enfermedad es todavía desconocida, aunque se han planteado algunas teorías al respecto. Entre ellas se baraja la posibilidad de que estar expuesto a toxinas ambientales [2] pueda ser dicha causa. También plantean que puede ser debido a estrés oxidativo, procesos que son parte natural del envejecimiento, o simplemente que la persona tiene predisposición genética.

2.1.2. Síntomas

Esta enfermedad presenta muchos y muy diversos síntomas, los que se han querido separar en los siguientes [2] grupos: motores y no motores. A continuación se explica con detalle cada uno de ellos, para poder comprender como funciona la enfermedad y poder relacionarlo con las pruebas.

Motores

Entre los síntomas más comunes se encuentra el temblor, la rigidez, y falta de movimiento. Todos los síntomas por lo general, suele iniciarse en un lado cuerpo, pero con el tiempo suele extenderse al otro. Vamos a enumerar algunos de los síntomas motores y a entrar en detalle.

- Reducción del movimiento: seguramente cuando oigamos enfermedad de Parkinson, se relacione directamente con una persona que presenta dificultad de movimiento, o lo hace muy lentamente. Aunque no es el único, sí que es uno de los síntomas clave y más frecuentes. El Parkinson produce una rigidez en los músculos de todo el cuerpo, lo que afecta a acciones como andar, o a la hora de comer o hablar.

  Es por esto que muchas veces los pacientes se ven en la necesidad de utilizar silla de ruedas o herramientas como andadores o bastón. Imaginarse a una persona mayor de 60 años en silla de ruedas, se podría decir que es más habitual, sin embargo, nos cuesta imaginar a personas de 30 en esta situación, a las que también afecta esta enfermedad.

  La voz de un paciente con estas características se vuelve monótona y carece de expresividad o entonación. Puede producirse repetición de sílabas o tartamudeo.

- Temblores: El temblor en las personas con Parkinson también es muy frecuente. Suele dar en extremidades superiores, y se hace muy notable, sobretodo, cuando se encuentra en estado de reposo. También puede suceder, que el nivel de temblor aumente en situaciones de estrés. Esto puede deberse a los niveles de adrenalina.

- Inestabilidad postural: Esto se hace un síntoma inevitable después de saber que presenta problemas musculares y de movimiento. Al producirse rigidez muscular, se hace complicado mantener una postura vertical, lo que puede suponer cierto desequilibrio provocando alguna caída. Algunos pacientes describen el término “congelación”, refiriéndose a la sensación de tener los pies anclados al suelo, y ser incapaz de seguir adelante.

No motores

Además de los síntomas motores, se ha demostrado [3] que los no motores contribuyen al deterioro en la salud, es posible que no se reflejen a simple vista, pero deben ser tenidos en cuenta a la hora de aplicar una u otra terapia. Entre los síntomas más frecuentes se encuentran:

- Cognitivos: Suele afectar especialmente a la memoria y el sueño. En muchos casos pueden producirse ataques de sueño, o en el momento inesperado. También afecta a los estados de ánimo, provocando apatía y cansancio.
- Psiquiátricos: Como en muchas enfermedades, ésta puede dar lugar a depresión, pudiendo acentuar los síntomas motores.
- Sistema autónomo: Es el encargado de controlar las funciones involuntarias de nuestro cuerpo, como, por ejemplo, la respiración, digestión, micción, etc. Éstos pueden verse afectados en gran manera por el deterioro de este sistema, aunque suele ocurrir en fases avanzadas de la enfermedad.

2.1.3. Diagnosis y tratamiento

No existe una prueba o test concreto que permita el diagnóstico del Parkinson. Especialistas en el tema se basan en el historial médico conocido y en la observación del paciente, para poder determinarlo. En ocasiones pueden recomendar pruebas médico como un análisis de sangre o un escáner cerebral, aunque no son determinantes, se tienen que analizar en conjunto con el resto de síntomas, tanto físicos como neurológicos.

En cuanto al tratamiento, recordar que la enfermedad es provocada por un fallo en el nivel de dopamina en el cuerpo, por tanto, las medicaciones que suelen recetarse, procuran sustituir o compensar este fallo. Son medicaciones que pueden tener algunos efectos secundarios, pero han resultado ser efectivas.

Además, existen otros métodos de tratamiento como la cirugía, que procura controlar el movimiento para evitar los temblores y movimiento involuntarios. Esta opción siempre es la más arriesgada.

2.2. Disartria

Tras haber descrito la enfermedad de Parkinson, se puede ver que la disartria juega un papel mayor en estos pacientes, ya que es uno de los síntomas más comunes y con mayor notoriedad. Es por ello, por lo que vamos describir la disartria, sus

síntomas, características y tipos, para poder ver de qué manera afecta a esta enfermedad.

2.2.1 Introducción

"La disartria es un trastorno del habla de origen neurológico donde algunas de sus características son, la lentitud de movimiento, debilidad, imprecisión, incoordinación, movimientos involuntarios y/o alteración del tono de la musculatura implicada en el habla" [4]. Podemos resumir, por tanto, que la disartria es una alteración motora. Esta misma, puede ser clasificada según diversos criterios, como la edad de inicio, severidad, características perceptuales, etc.

2.2.2 Criterios y componentes

El estudio de la disartria se ha basado en estos citados criterios [4] para poder estimar el origen o la causa. Puede verse iniciada de forma congénita, o por algún tipo de trastorno traumático, infeccioso o degenerativo entre otros. A su vez, la disartria puede seguir cursos evolutivos diferentes, puede mantenerse estable a lo largo del tiempo, sin conseguir mejorar o empeorar con rehabilitación, siendo el caso de parálisis cerebrales o también puede tratarse de una disartria regresiva, efecto de una etapa post traumática, donde el paciente presenta un nivel que consigue disminuir al cabo del tiempo. En contraposición, puede tratarse de una disartria progresiva, la cual se ve incrementada desde sus primeros síntomas, esto puede darse en enfermedades como el ELA o el Parkinson.

Al igual que en la mayoría de enfermedades, existe diferentes grados, desde uno leve, a uno grave, siendo éste, quedarse sin habla, es decir, disartria en su máxima expresión. El grado suele ir directamente relacionado con el nivel de lesión en el sistema nervioso central o periférico y como tal, pueden producirse diferentes alteraciones como la espasticidad, temblor, rigidez o una combinación de todos ellos.

Es necesario nombrar, cuáles son los **procesos motores** [4] que intervienen a la hora de hablar, y que, por tanto, alguno o varios de ellos, se verán afectados por la disartria. El primero y el más importante es la respiración, ya que el aire afecta a las cuerdas vocales haciéndolas vibrar, de tal manera que se produce la voz, conduciéndonos al segundo proceso que es la fonación. El segundo de ellos es la resonancia, que permite aumentar o disminuir el tono vocal. Un ejemplo de resonador es la laringe. También tenemos la articulación, que nos permite modificar el sonido a través de unos articuladores, como puede ser la lengua. Y, por último, tenemos la prosodia, que estudia los rangos sonoros de la voz, o aspectos melódicos, donde podemos percibir las emociones del paciente, que presenta patrones de ritmo y entonación.

A su vez, existen unos **componentes funcionales** [4] en la voz, que ayudan al estudio y clasificación de la disartria, estos son, naturalidad, inteligibilidad, velocidad del habla y compresibilidad.

- Naturalidad: Se considera un habla natural cuando la voz sigue los estándares de entonación y ritmo.
- Inteligibilidad: Lo que se consigue entender, independientemente de entonación o tono. Lo que se quiere reflejar es la capacidad del interlocutor para hacerse entender, usando estrategias como por ejemplo pausas, para compensar el resto de carencias de la voz.
- Velocidad del habla: Se refiere al número de palabras por minuto que el interlocutor es capaz de trasmitir. (En un habla fluida se estiman unas 150 palabras por minuto).
- Comprensibilidad: Este concepto une la inteligibilidad, es decir, lo que somos capaces de entender vocalmente hablando, con el resto de aspectos que nos permiten comprender el mensaje, ya sea acompañando con gestos o expresiones.

2.2.3 Tipos

Tras analizar detenidamente todas las características anteriormente nombradas, los expertos clasifican las disartrias en los siguientes tipos:

- Disartria flácida: Es debido un daño en las neuronas de los nervios craneales, que puede ser provocado por algún proceso degenerativo, accidente cerebrovascular o enfermedades congénitas entre otros. Cuando alguna parte de la unidad motora inferior se ve dañada, puede suponer alguna alteración en los movimientos, tanto los automáticos como los voluntarios o reflejos. Dichas alteraciones pueden manifestarse como un tipo de parálisis flácida, que afecta principalmente a la respiración y al habla. Por estos motivos pueden producirse síntomas como la hipernasalidad en el habla, dificultad a la hora de tragar, problemas de respiración o babeo, entre otras muchas.

- Disartria espástica: en este caso, se debe al daño en las neuronas motoras superiores, más concretamente, en las vías de activación directa o indirecta, provocando una rigidez en los músculos, especialmente en la lengua y labios, afectando a la respiración y la fonación. Es por ello que pueden producirse dificultades en la articulación de palabras y con mayor lentitud, además de con voz ronca y en ocasiones, alguna alteración emocional.

- Disartria atáxica: este tipo de disartria se debe a lesiones en el cerebelo, quien regula los movimientos procedentes de otros sistemas motores. Afecta principalmente a los músculos, produciendo lentitud en movimientos y con la intensidad inadecuada. Aunque en mayor medida, también presenta desajustes en el habla, como distorsiones, o énfasis en sílabas equivocadas.

- Disartria hipercinética: Se asocia con aumento en la velocidad y el número de movimientos involuntarios.

- Disartria hipocinética: A diferencia de la hipercinética, supone la disminución de la velocidad y el número de movimiento, donde una causa muy común es la enfermedad de Parkinson.

- Disartria mixta: Como su propio nombre indica, consiste en una combinación de característica de las previamente descritas.

2.3. Estudios relacionados

En este apartado se quiere mencionar algunos estudios que ya se han llevado a cabo sobre temas relacionados con algunos aspectos de este proyecto. Debido a que las condiciones de los experimentos de los artículos y las del proyecto no eran siempre las mismas, no se puede comparar de forma directa los resultados de ambos, aunque sí estimar qué parámetros o características ayudarían a mejorar los sistemas.

El caso que se estudia durante este TFG es algo diferente a lo que se suele ver en los estudios, por tanto, se destacarán los aspectos similares que se han tratado en otros ensayos y que hayan podido ayudar a realizar este nuevo estudio. Los artículos que se han considerado interesantes destacar y de los que se hablará a continuación, serán los siguientes:

1. "Dysarthria Intelligibility Assessment in a Factor Analysis Total Variability Space" [5]
2. "Spectral Features for Automatic Blind Intelligibility Estimation of Spastic Dysarthric Speech" [6]
3. "Dysarthric Speech Database for Universal Access Research " [7]
4. "Automated Dysarthria Severity Classification for Improved Objective Intelligibility Assessment of Spastic Dysarthric Speech" [8]

Todos los artículos mencionados anteriormente, se asemejan a nuestro caso de estudio en mayor o menor medida. Uno de los más destacables es el artículo 1, donde el objetivo es obtener el grado de inteligibilidad en personas con disartria, al igual que en el caso de este proyecto, y además, se ejecuta sobre la misma base de datos que se plantea en este TFG. La diferencia mayoritaria de este artículo con respecto a las pruebas actuales, reside en las características acústicas extraídas. Usa dos predictores, uno lineal y otro basado en máquinas de vectores soporte ("Support Vector Regression", SVR), y los resultados se evalúan en función de las correlaciones de Pearson, Spearman y la raíz cuadrada del error cuadrático medio ("Root-Mean-Square Error", RMSE). A pesar de que las características extraídas son diferentes, y que por tanto no se pueden comparar, los resultados obtenidos en este proyecto, resultan mejores que los planteados en el artículo [5].

Por otra parte, el artículo 2 también tiene una estrecha relación, ya que pretende de igual forma, predecir la inteligibilidad de pacientes con disartria. Se

consideró interesante, porque en él se hace referencia al proceso de entrenamiento. Para estimar la inteligibilidad, es posible entrenar previamente el predictor mediante la comparación de elocuciones de habla disártrica con su versión de referencia (es decir, con pronunciaciones de las mismas palabras o frases sin ninguna alteración debida a la disartria). Sin embargo, no es habitual disponer de dichas señales de referencia, por lo que es necesario adoptar la estrategia denominada "ciega", como es el caso de este TFG.

El artículo 3 trata sobre la base de datos que se usa en esta memoria. Se describe como está compuesta, las palabras y los tipos de pacientes que participan en el estudio. Ésta se diseñó específicamente con el objetivo de crear un sistema de reconocimiento automático de voz disártrica, aunque ha sido utilizada en estudios posteriores para la predicción automática del grado de disartria.

Por último, se encuentra el artículo 4, que corresponde al que se utilizará en mayor medida a lo largo de la memoria, como artículo de referencia. En él, como viene siendo habitual, se pretende predecir la inteligibilidad en pacientes con disartria. Está basado en un subconjunto de 10 interlocutores de la base de datos de Universal Access [7], nombrada anteriormente en el artículo 3, de la cual, se extraerán las características acústicas para ser evaluadas. El experimento del artículo 4 se realiza con un vector de 6 características, que también se han evaluado en la primera prueba de este proyecto y cuyo resultado nos ha servido como experimento de referencia. En ambos casos, se evaluaron los resultados en base a la correlación de Pearson y de la raíz del error cuadrático medio, pero a pesar de que los métodos de regresión utilizados fueron diferentes en los dos casos, el sistema desarrollado en este TFG ha logrado una mejoría con respecto al artículo de referencia tal y como explica en el capítulo 5.

3. Diseño de la solución

3.1. Introducción

Como hemos dicho en varias ocasiones, la idea clave de este trabajo, es conseguir predecir el grado de inteligibilidad en diversas señales de voz. Para ello habrá que analizar dichas señales, extraer y seleccionar las características acústicas de las mismas que son útiles para realizar la predicción de su inteligibilidad.

Para el desarrollo del sistema y las pruebas experimentales, contaremos con la base de datos "Universal Access" [7], que consta de un conjunto de ficheros de audio de 15 interlocutores que padecen disartria. Ésta es provocada por parálisis cerebral, pero debido a las semejanzas en cuanto a los síntomas con respecto al Parkinson, se ha utilizado para realizar las pruebas.

Cada uno de los interlocutores pronuncia 255 palabras diferentes. También disponemos de los "scores" reales de inteligibilidad de los pacientes, es decir, junto con la base de datos, se proporciona una lista, donde se asigna a cada paciente una puntuación de inteligibilidad (en el rango de 0 a 100) de acuerdo al estudio realizado por especialistas. Por lo tanto, nuestro objetivo es diseñar un sistema que, a partir de un conjunto de parámetros acústicos extraídos de las señales de voz de un paciente, obtenga una medida de su inteligibilidad que esté correlada con el valor real de inteligibilidad proporcionado en la base de datos.

El sistema propuesto consta de dos componentes fundamentales: la extracción de características acústicas ("features") y el regresor que relaciona dichas características con la medida predicha de inteligibilidad. En este capítulo, aparte de la descripción de la base de datos utilizada, se comentarán los features elegidos para esta tarea, que se han elegidos a partir de estudios sobre las particularidades de la voz disártrica [4], y los algoritmos de aprendizaje máquina utilizados para realizar el proceso de regresión. Asimismo, se describirán las medidas consideradas para evaluar las prestaciones del sistema. Finalmente, se hará una breve descripción del sistema propuesto, cuyos detalles se explican con más profundidad en el siguiente capítulo.

3.2. Estructura y justificación

En este apartado se hará una descripción detallada de todos los elementos que toman parte a la hora de cumplir el objetivo de este estudio. También se hará un análisis con las justificaciones y los motivos de las elecciones de los parámetros acústicos.

3.2.1. Base de Datos

Para toda prueba o estudio, tener una buena base de datos es determinante. En nuestro caso, como se ha comentado ya brevemente, disponemos de la base de datos Universal Access [7], que consta señales de voz de 15 interlocutores diferentes (4 mujeres y 11 hombres) que padecen la disartria provocada por parálisis cerebral. A pesar de no padecer la enfermedad de Parkinson, según se ha visto en el Capítulo del Estado del arte, la disartria es uno de los síntomas principales de personas con Parkinson, y por tanto, se plantea utilizar la base de datos con interlocutores con parálisis cerebral, ya que puede ser una aproximación fiel a una situación de pacientes con Parkinson.

Los datos que disponemos sobre ellos son, si corresponde a un hombre o una mujer, y el grado de inteligibilidad. La documentación viene organizada en forma de carpetas nombradas con el código de interlocutor y que contienen los audios de cada una de las 255 palabras diferentes que se han grabado de cada uno de ellos. Entre estas palabras se encuentran los diez dígitos, las veintiséis palabras del alfabeto radio, diecinueve comandos de ordenador, cien palabras comunes en inglés y otras cien no comunes.

A su vez, esta base de datos se ha dividido en dos grupos. Por un lado, el grupo de entrenamiento, con ciento cincuenta y cinco palabras de las anteriores mencionadas, a excepción de la cien no comunes, que si tenemos en cuenta los quince interlocutores obtendremos 2325 palabras. Por otro lado, tenemos el grupo de test, donde ahora sí, consta de las cien palabras no comunes, que de la misma forma, resultan 1489, ya que once registros no pudieron grabarse correctamente. Estas serán las señales que utilizaremos para el desarrollo (grupo de entrenamiento) y evaluación (grupo de test) de nuestro sistema.

A continuación se muestra en la **Figura 1**, proporcionada junto con las señales de voz, en la que se aprecia las características de cada interlocutor.

Speaker Label	Age	Speech Intelligibility (%)	Dysarthria Diagnosis
M01	>18	very low (15%)	Spastic
M04	>18	very low (2%)	Spastic
M05	21	mid (58%)	Spastic
M07	58	low (28%)	Spastic
M08	28	high (93%)	Spastic
M09	18	high (86%)	Spastic
M10	21	high (93%)	Mixed
M11	48	mid (62%)	Athetoid
M12	19	very low (7.4 %)	Mixed
M14	40	high (90.4%)	Spastic
M16	-	low (43%)	Spastic
F02	30	low (29%)	Spastic
F03	51	very low (6%)	Spastic
F04	18	mid (62%)	Athetoid
F05	22	high (95%)	Spastic

Figura 1: *Características de los interlocutores*

3.2.2. Justificación elección base de datos

En este trabajo, se pretende realizar un programa automatizado con la finalidad de colaborar en el sector médico y ayudar a las personas con disartria, en su calidad de vida. Es cierto que existen muchas enfermedades donde se padece, y que ha sido complicado elegir a un sector tan específico de entre todos los que hay, pero en nuestro caso, hemos decidido realizar este análisis para aplicarlo a personas con la enfermedad de Parkinson.

Además de esto, los requisitos de la base de datos necesarios para llevar a cabo este proyecto en concreto, fueron muy específicos y por tanto, fue necesario encontrar una que cumpliera con las expectativas. En primera lugar, era necesaria una base de datos de acceso público, que nos permitiera poder trabajar con ella sin ningún tipo de restricción. Se ha comentado que fue necesario un acuerdo, para el uso exclusivo de la misma para trabajos de investigación, lo cual en nuestra situación, nos permite libertad para realizar las pruebas pertinentes. Por otro lado, era conveniente un sistema que tuviera una carga de datos adecuada, es decir, un número razonable de interlocutores, y a su vez, un número suficiente de ficheros de audio por interlocutor. Se consideró ésta correcta, basándonos en otros artículos similares en los que se hablan durante la memoria. Se puede observar como en algunos de ellos se trabaja con 10 o 19 interlocutores, mientras que en nuestro caso, eran 15. Por último y más importante, ante el objetivo de obtener un sistema que prediga el grado de inteligibilidad de una señal de voz, se necesita un punto de referencia para evaluar si

es correcto o no. Por tanto, fue necesario elegir una base de datos que tuviera unas etiquetas predefinidas con el grado de inteligibilidad hallado subjetivamente, a modo de comparación.

No es sencillo encontrar unos datos que cumplan todas las expectativas, es por ello que a pesar de haber muchas disponibles de diferentes clases, esta fue de las pocas que se ajustaba a las necesidades del proyecto.

3.2.3. Características acústicas

A continuación, expondremos el apartado más importante del proyecto, que son las características acústicas. En primer lugar, fue necesario decidir una serie parámetros a extraer de nuestras señales de voz, que nos permitieran compararlas y caracterizar los valores que nos serán interesantes. Con esta información, podremos sacar las conclusiones pertinentes. Los parámetros considerados se dividen en cuatro grupos principales. Todos estos features se extraen a nivel de trama (en concreto, cada 10 ms) y seguidamente se calculan sobre ellos, una serie de estadísticos (media, desviación típica, etc.) a nivel de todo el fichero de audio, tal y como se describe a continuación. Los grupos previamente citados son los siguientes:

- **LHMR**: el primer feature que hemos usado ha sido "*Low-to-High Modulation energy Ratio*" (LHMR). Como se ha descrito anteriormente, parte de nuestras pruebas se han basado un artículo [8] donde se ha utilizado este parámetro para realizar el experimento. De forma práctica, lo que hace el LHMR, es comparar la energía de modulación a varias frecuencias. Como se puede observar en la fórmula para el cálculo del LHMR, tenemos un parámetro K, que corresponde al índice del filtro de modulación, es decir, se modifica para establecer el límite entre cada rango de frecuencias de modulación que van a compararse. En la propuesta del artículo original, se calcula el cociente de energía de modulación entre frecuencias por encima y por debajo de los 4Hz, que corresponden con K=4. Sin embargo, en este caso se plantea obtener varios features de LHMR variando este parámetro, dándole valores a K desde 1 hasta 7, obteniendo de este modo diferentes comparaciones de energías de modulación en diferentes rangos de frecuencia. En total, obtenemos 7 características acústicas relacionadas con el LHMR.

$$\mathrm{LHMR} = \frac{\sum_{k=1}^{K^*}\sum_{j=1}^{23}\bar{E}_{j,k}}{\sum_{k=K^*+1}^{8}\sum_{j=1}^{23}\bar{E}_{j,k}}$$

***Ecuación 1**: Fórmula de LHMR; [6]*

- **Δc0:** El segundo feature que se ha usado ha sido la primera derivada del coeficiente cepstral de orden 0, que está relacionado con la log-energía de la señal. Los coeficientes conocidos como cepstrales, son comunes en los campos que buscan reconocer tanto el habla como sonidos ambientales [9]. En el artículo propuesto, se pretende obtener la tasa de cambio de la señal log-energía, y para ello se ha calculado la desviación típica de este parámetro. A su vez, y con el fin de conseguir mejores resultados además de la desviación, hemos procedido a calcular tanto la media como la kurtosis y skewness (simetría). Se han obtenido, por tanto, 4 características acústicas derivadas de Δc0, correspondientes a cada uno de los estadísticos antes mencionados.

- **f0:** este feature corresponde con la frecuencia fundamental de la señal de voz es decir, la frecuencia a la que vibran las cuerdas vocales de un hablante. Nuestro artículo de referencia ha extraído 3 diferentes parámetros, la desviación estándar, en rango, es decir el máximo menos el mínimo, y, por último, el porcentaje de segmentos sonoros. Por completitud, también hemos añadido otras medidas como la media, la kurtosis y la simetría, obteniendo finalmente 6 ficheros para el análisis.

- **LPC:** como cuarto y último feature, se ha calculado el residuo de la predición lineal ("Linear Predictive Coding", LPC) de la señal de voz que suele utilizarse para estudiar la excitación de la fuente vocal y para codificación del habla. Del mismo modo que antes, se obtuvieron varias medidas. La primera fue la kurtosis, ya que según vimos en el artículo de referencia, es un feature bueno para el análisis, y para completar, volvimos a calcular, la media, la desviación y la simetría, obteniendo finalmente 4 características acústicas derivadas del residuo LPC.

3.2.4. Justificación de la elección de los features

A continuación, exponemos la relación que presentan las características mencionadas en el apartado anterior con la inteligibilidad de la voz. Es necesario tener en cuenta, que la relación entre la característica acústica extraída y alguna de las dimensiones de las disartrias, no es ni mucho menos exacta, ya que se están utilizando señales de voz de personas con enfermedades reales, y que por tanto, cada voz presenta unos rasgos diferentes, que en muchas ocasiones no pueden compararse ya que no siguen el mismo comportamiento. Para dar dicha explicación, nos basamos en la Tabla 1 del artículo [4] "*Las disartrias*" (**Figura** y **Figura**) donde se muestran las diferentes dimensiones utilizadas para el estudio de la disartria.

Vamos a utilizar una estructura similar al apartado anterior, enumerando los features y en qué nos hemos basado a la hora de relacionarlos en parámetros del habla.

En primer lugar, hemos hablado del **LHMR** (*Low-to-High Modulation Energy Ratio),* que es una medida relacionada con el espectro de modulación de la señal de voz, es decir, con la variación temporal de la energía de la señal de voz en diferentes bandas de frecuencia. Observando la tabla de las disartrias, hemos podido relacionar este parámetro con: la velocidad del habla, el aumento de la velocidad en segmentos, el aumento de la velocidad en general y las interrupciones en la voz, así como los aspectos relacionados con la intensidad.

A su vez, tenemos el segundo feature, **Δc0,** que al estar relacionado con la log-energía de la señal, puede indicar la presencia de anormalidades en la intensidad: monointensidad, excesiva variación en la intensidad, disminución de la intensidad, intensidad alternada y nivel de intensidad general.

Una persona con una disartria avanzada, suele tener una voz con una misma intensidad, lo que se traduciría en cambios insignificantes de energía. De la misma manera, las personas con Parkinson, suelen tener síntomas como temblores, lo que provocaría ciertas interrupciones en el habla o cambios bruscos de energía. Con esto, a pesar de que como ya hemos dicho, estas deducciones no son del todo precisas, la energía y su variación en determinadas bandas de frecuencia están directamente relacionadas con aspectos que caracterizan el habla disártrica.

Siguiendo con el feature 3**, f0**, la frecuencia fundamental, lo hemos relacionado con aspectos tonales de la voz [10]. Cuando existe disartria, la voz adquiere monotonía y, por tanto, podemos decir, que la variación de su frecuencia fundamental es pequeña. Respecto a este feature, vemos que ha sido utilizado en las pruebas del artículo [8], más concretamente, se ha obtenido la desviación estándar de la frecuencia fundamental. Esto nos indica la variabilidad o cuanto se desvían los resultados de un valor concreto, es decir, si obtenemos una baja desviación, es que se mantiene un valor muy cercano a la frecuencia fundamental, y por tanto, es el indicativo de una voz monótona. Del mismo modo, el rango dinámico de f0 puede relacionarse con el quiebre del tono. Finalmente, el porcentaje de segmentos sonoros puede ser indicativo de la presencia de distorsión vocálica.

Por último, tenemos nuestro feature **LPC**, que, según la descripción, está relacionado con la excitación de la señal de voz. Siguiendo esta línea, lo hemos asociado a la obtención de parámetros como voz ronca, voz soplada, etc.

Nº	DIMENSIÓN	DESCRIPCIÓN
1	Nivel del tono	El tono de la voz aparece siempre demasiado bajo o demasiado alto para la edad y el sexo del individuo.
2	Quiebre del tono	El tono de la voz muestra una variación súbita e incontrolada (interrupciones en falsete).
3	Monotonía	La voz se caracteriza por monotonía. Carece del tono normal y de variaciones de inflexión. Tiende a mantenerse en un nivel de tono.
4	Temblor de la voz	La voz muestra temblores o vibraciones.
5	Monointensidad	La voz presenta monotonía de intensidad. Carece de las variaciones normales de intensidad.
6	Excesiva variación en la intensidad	La voz muestra alteraciones súbitas e incontroladas en la intensidad: a veces se vuelve demasiado fuerte y a veces demasiado débil.
7	Disminución de la intensidad	Hay una progresiva disminución de la intensidad.
8	Intensidad alternada	Se observan cambios de intensidad alternados.
9	Nivel de intensidad, general	La voz es insuficiente o excesivamente fuerte.
10	Voz áspera	La voz es áspera, ronca y raspante.
11	Voz ronca (húmeda)	Hay una ronquera húmeda, con "sonido líquido".
12	Voz soplada (continua)	La voz es siempre soplada, débil y fina.
13	Voz soplada (transitoria)	La voz soplada es transitoria, periódica e intermitente.
14	Voz forzada-estrangulada	La voz (fonación) tiene un sonido forzado o estrangulado (como si pasara con gran esfuerzo por la glotis).
15	Interrupciones de la voz	Se producen interrupciones súbitas del flujo respiratorio vocal (como si un obstáculo del aparato vocal impidiera el flujo del aire por un momento).
16	Hipernasalidad	La voz suena demasiado nasal. Una excesiva cantidad de la corriente de aire es resonada en la cavidad nasal.
17	Hiponasalidad	La voz es poco nasal.
18	Emisión nasal	Hay emisión nasal de la corriente de aire.

***Figura 2**: Dimensiones para estudio de disartria – Parte 1*

Nº	DIMENSIÓN	DESCRIPCIÓN
19	Inspiración-espiración forzada	El habla es interrumpida por suspiros de inspiración y espiración, súbitos y forzados.
20	Inspiración audible	Hay una inspiración audible, jadeante.
21	Gruñido al final de la espiración	Se produce un gruñido al terminar la espiración.
22	Velocidad	La velocidad del habla es lenta o rápida en grado anormal.
23	Frases cortas	Las frases son cortas (quizás porque las inspiraciones se producen con mayor frecuencia de lo normal). El emisor parece haberse quedado sin aliento. Al final de la frase puede producir un jadeo entrecortado.
24	Aumento de la velocidad en segmentos	La velocidad aumenta en forma progresiva dentro de determinados segmentos del habla.
25	Aumento de la velocidad general	La velocidad aumenta en forma progresiva desde el comienzo hacia el final del enunciado.
26	Acentuación reducida	El habla muestra reducción de la acentuación apropiada o de los patrones de énfasis.
27	Velocidad variable	La velocidad alterna de lenta a rápida.
28	Intervalos prolongados	Hay prolongación de los intervalos entre las palabras o entre las sílabas.
29	Silencios inadecuados	Hay intervalos de silencio inadecuados.
30	Breves precipitaciones al hablar	Hay breves precipitaciones del habla separadas por pausas.
31	Acentuación excesiva y uniforme	Hay una acentuación excesiva sobre partes del lenguaje que suelen ser átonas, es decir, palabras monosilábicas y sílabas átonas de palabras polisilábicas.
32	Distorsión consonántica	Los sonidos consonánticos carecen de precisión. La articulación es superficial. Muestra falta de nitidez, distorsiones y de fuerza. Hay falta de agilidad al pasar de un sonido consonántico a otro.
33	Sonidos prolongados	Hay prolongación de los sonidos.
34	Sonidos repetidos	Hay repetición de sonidos.
35	Quiebres articulatorios	Hay quiebres no sistemáticos en la exactitud de la articulación.
36	Distorsión vocálica	Los sonidos vocálicos están distorsionados en su duración total.
37	Inteligibilidad (general)	Se refiere a cuánto entiende el receptor en relación al habla que produce el paciente.
38	Carácter extraño (naturalidad)	Corresponde al grado en que el habla llama la atención a causa de sus características inusuales, peculiares y poco naturales.

***Figura 3:** Dimensiones para estudio de disartria – Parte 2*

3.2.5. Regresores

Este apartado corresponde al otro componente fundamental de nuestro sistema: el algoritmo de regresión utilizado para realizar la predicción de inteligibilidad. En particular, hemos optado por probar cuatro regresores distintos. Depende de cuál utilicemos, se obtendrán resultados más aproximados que otros. En el apartado de "Resultados" detallaremos cuál de ellos ha resultado más favorable, y se verá reflejado en los resultados numéricos.

Todos los regresores considerados están basados en máquinas de vectores soporte ("Support Vector Machine", SVM). Este método puede ser utilizado tanto en clasificación como en regresión, y donde otros procuran minimizar el error, este por el contrario, busca la disminución del riesgo estructural [11]. En resumen, "la idea es seleccionar un hiperplano de separación que equidista de los ejemplos más cercanos de cada clase para, de esta forma, conseguir lo que se denomina, un margen máximo a cada lado del hiperplano". Es decir, lo que se pretende es conseguir una separación clara entre los dos tipos de datos, para que, cuando llegue una nueva muestra, el predictor sea capaz de decidir o asignarlo a una de las dos clases. Esta clase de barrera o separación que se quiere establecer, puede estimarse de diversas maneras, no siempre de forma lineal. De hecho, en nuestro caso, se utilizan tanto, kernels lineales como polinómicos de diferentes grados (en concreto, de orden 2 y 3) y gaussiano.

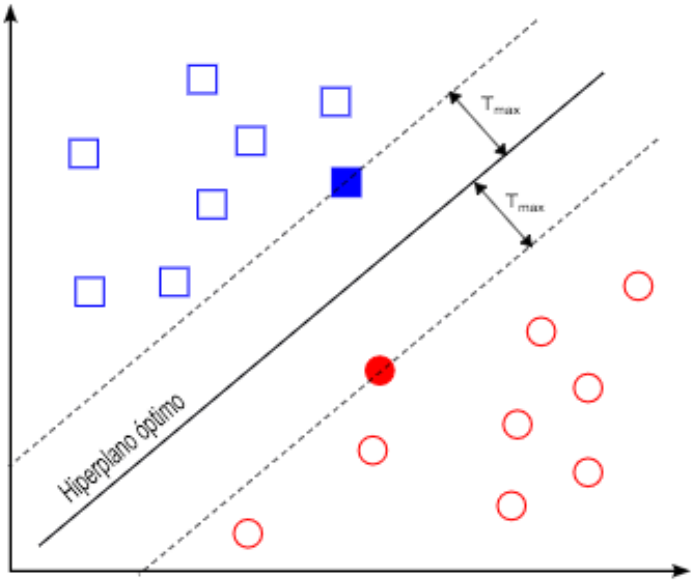

***Figura 2**: SVM – Margen de un hiperplano de separación. [11]*

3.2.6. Correlaciones

Para determinar los features más adecuados para la tarea en cuestión, se decidió utilizar la correlación de Pearson como criterio de selección, aunque tambi´

Para la evaluación de las prestaciones del sistema se consideraron las siguientes medidas: correlación de Pearson, correlación de Spearman y error cuadrático medio. Procedemos a detallar cada una de ellas:

- **Correlación de Pearson**: Mide el grado de covariación que existe entre dos variables relacionadas linealmente. Es importante hacer hincapié en linealmente, ya que puede haber ocasiones en que haya una relación estrecha, pero no sea lineal. En estos casos, no se aconseja obtener el coeficiente de Pearson. El valor de esta correlación de encuentra entre 0 y 1 en valor absoluto (o bien entre -1 y 1). Cuanto más cercano a 1 se encuentre (en valor absoluto), más importancia tendrá para nosotros, ya que indica que tiene una gran correlación.

$$r = \frac{S_{XY}}{S_X S_Y}$$

***Ecuación 2**: Coeficiente de correlación de Pearson*

Como se ha mostrado, la **Ecuación 2**, se obtiene el coeficiente de correlación de Pearson, el cual, pretende establecer la relación entre dos variables. Sxy representa la covarianza de la variable X e Y, mientras que Sx e Sy, representan la desviación típica de X e Y respectivamente.

- **Correlación de Spearman**: Es la relación entre dos variables aleatorias continuas. Esta es calculada en base a una serie de rangos asignados. Además, sigue las mismas normas de interpretación que Pearson, ya que varía entre -1 y 1, y por tanto, el 0 indica una no correlación.

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{r\,(r^2 - 1)}$$

***Ecuación 3:** Coeficiente de correlación de Spearman*

De igual manera, para el cálculo de la correlación de Spearman, se utiliza el parámetro d, que corresponde a los estadísticos de orden x – y, es decir, la distancia que existe entre las posiciones relativas, mientras que “r” es el número de puntuaciones.

- **Raíz del error cuadrático medio**: Consiste en medir el promedio de los errores al cuadrado y calcular su raíz cuadrada.

Además de las correlaciones, también hemos calculado el error cuadrático medio, que nos servirá de referencia para sacar las conclusiones de nuestras pruebas y se calcula de la siguiente manera.

$$x_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N}}\ .$$

Ecuación 4: *Raíz del error cuadrático medio*

A pesar de haber calculado las correlaciones anteriores, es necesario analizar si los resultados obtenidos son estadísticamente significativos o no para nuestras pruebas. Se usan los P-valores para determinar qué variables son significativas en el estudio, y por tanto, deben mantenerse en el modelo de regresión. Cuanto más próximo a cero es el valor, mayor significancia. Este valor, es un indicativo de cuanto de valiosa es la variable para el estudio, es decir, si conviene descartarla o sería útil para obtener unos mejores resultados.

Nótese que para la selección de las características acústicas más adecuadas a nuestra tarea, se ha utilizado como criterio la correlación de Pearson calculada entre cada característica individual y los “scores” reales de inteligibilidad.

4. Implementación

Habiendo visto el diseño, ya podemos hacernos una idea general de los pasos y etapas principales del sistema desarrollado. A continuación, procederemos a detallar cada uno de ellos, explicando las modificaciones o ajustes que hemos llevado a cabo para la correcta ejecución de las pruebas. En la **Figura 3** se presenta un diagrama de bloques general de todos los procesos involucrados en nuestro sistema, que iremos desglosando paso a paso.

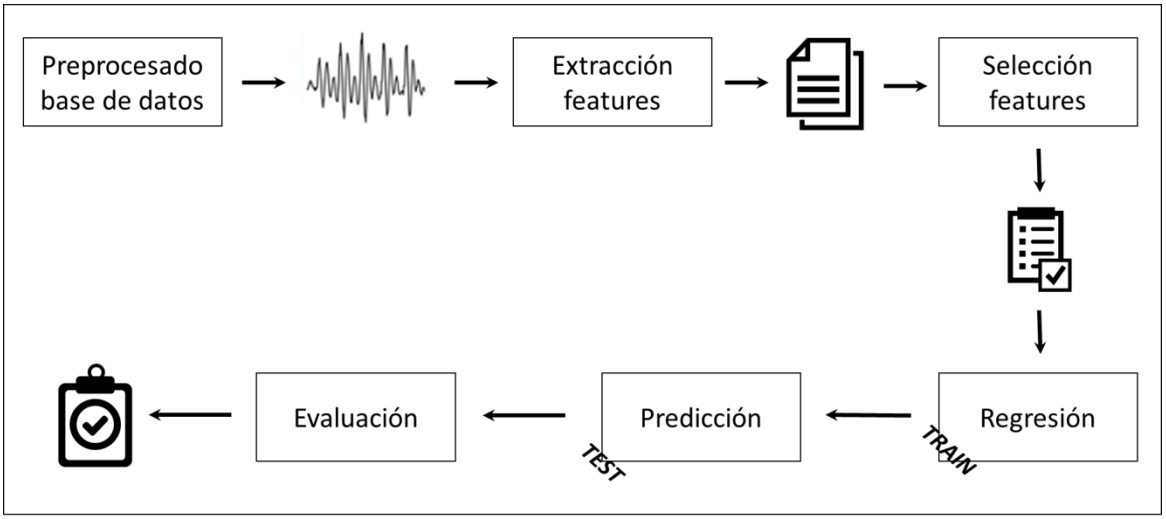

Figura 3: *Diagrama de bloques*

4.1. Preprocesado de la base de datos

Una de las etapas más críticas del sistema es el preprocesado de todas las señales de voz que conforman la base de datos. El sistema exige que haya una señal a la entrada, para ser analizada y poder obtener una salida de la cual podamos obtener la inteligibilidad. Es por esto de la importancia que tiene el preprocesado, ya que, con una mala entrada, puede verse perjudicado todo el sistema al obtener unos resultados poco satisfactorios y por tanto, unas conclusiones erróneas.

Para determinar el preprocesado más adecuado, en primer lugar, se analizó el contenido y calidad (en cuanto a las condiciones de grabación) de la base de datos. Como hemos comentado previamente, la base de datos está organizada de modo que se dispone de las carpetas por interlocutor, con los audios de cada una de las palabras pronunciadas por ellos. Dichos audios, tiene una duración de 3 segundos aproximadamente. Existen algunas excepciones, por ejemplo, de 10 segundos, ya que esto depende tanto de la palabra, como de la dificultad del paciente a la hora de

hablar. Ante esta posible variación, fue necesario encontrar un ajuste apropiado para cualquiera de las situaciones posibles.

Una forma de analizar estos audios fue representando gráficamente algunos de ellos. Observando las imágenes, se pudo distinguir claramente los intervalos de silencio, y en los que se produce sonido, tanto si es inteligible como si no. Con esto pudimos establecer un patrón y darnos cuenta que todos ellos contaban con tramos de silencio (o ruido de fondo) de longitud variable en los que no había presencia de voz. Se concluyó que estos tramos silenciosos o de ruido podrían interferir en las pruebas "falseando" los resultados y, por tanto, se decidió aplicar un detector de actividad vocal ("Voice Activity Detector", VAD) sobre todos los ficheros de voz. De esta forma, el VAD descarta los segmentos de silencio/ruido de forma que sólo la parte del audio que contiene voz es la es posteriormente procesada por el sistema. De entre las múltiples opciones de VAD existentes, nos inclinamos por utilizar el algoritmo propuesto por Sohn [12] debido a su buen funcionamiento y a que su uso está muy extendido en aplicaciones relacionadas con el procesamiento de voz.

Desde el punto de vista práctica, el VAD se implementó utilizando la función *vadsohn*. Esta pertenece a la ToolBox "voicebox" de MATLAB [13], y consigue establecer un vector de ceros y unos, según si la trama correspondiente es de silencio/ruido o voz. De esta forma, hemos podido procesar únicamente los tramos con voz, que son los que realmente aportan la información valiosa.

A continuación, presentamos información gráfica al respecto. En primer lugar tenemos la comparativa aplicando o no la función vadsohn, en un interlocutor con un 29% de inteligibilidad, pronunciando la palabra "command".

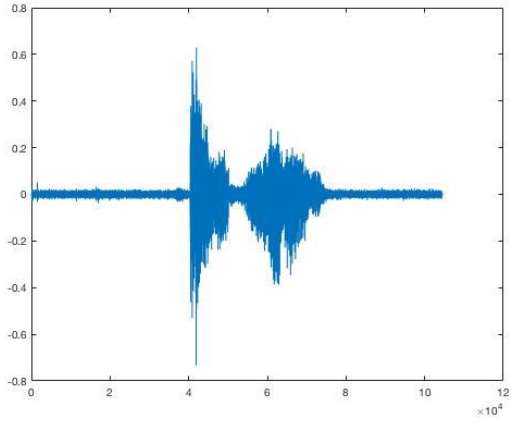

***Figura 6**: Ejemplo interlocutor (29%) sin vadsohn*

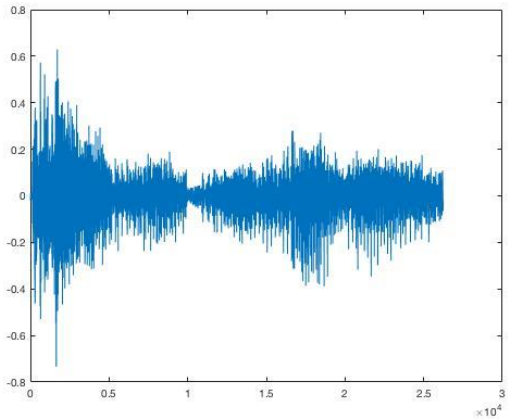

***Figura 7**: Ejemplo interlocutor (29%) con vadsohn*

Como se puede observar en la **Figura 6** y en la **Figura** , se consiguen eliminar adecuadamente las partes en las que no se aprecia sonido, resultando una señal para analizar, una cuarta parte más pequeña que la original.

De igual forma, se ha analizado otra señal de voz, pero en este caso, de un interlocutor con un 95% de grado de inteligibilidad pronunciando la misma palabra "command", resultando lo siguiente:

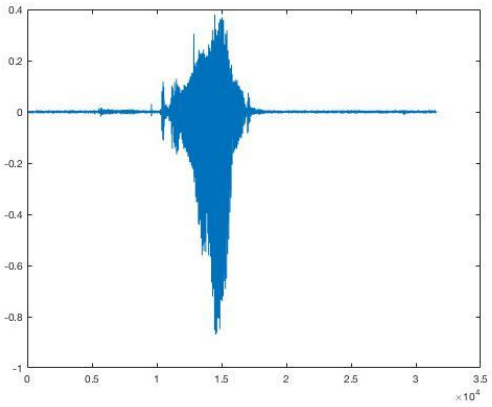

***Figura 4**: Ejemplo interlocutor (95%) sin vadsohn*

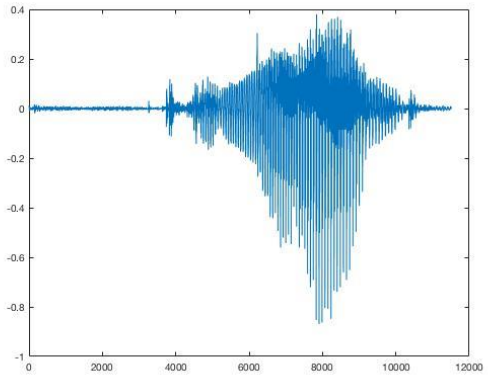

***Figura 5**: Ejemplo interlocutor (95%) con vadsohn*

Observando ahora la **Figura 4** y en la **Figura 5**, podemos ver, en primer lugar, que se ha eliminado correctamente la parte de la señal considerada como silencio. Es cierto que en esta ocasión, ha considerado la parte entre el instante 0,5 y 1, como voz, cuando en realidad podría aproximarse más al tramo de silencio. A pesar de esto, ha reducido la señal para el análisis de forma considerable..

Además de esto, la representación gráfica nos ayuda a apreciar la diferencia en las señales originales, donde el interlocutor primero parece que habla en dos tiempos, en el segundo se ve claramente como tiene un único cambio brusco en la voz indicando que habla con mayor fluidez. Esto es muy útil especialmente cuando se conoce la palabra que se está representando.

4.2. Extracción de *features*

Después de haber adaptado la base de datos de forma que tengamos las entradas del programa de acuerdo a nuestras necesidades, pasamos al segundo bloque, donde es preciso extraer los features que serán útiles para nuestro objetivo.

Comenzaremos realizando un nuevo *script* de MATLAB, que se usarán para incluir todas las funciones que servirán para calcular los features. Recordar que

nuestra variable de entrada, será a la que previamente habremos aplicado la función *vadsohn,* es decir, las señales de voz que no contienen tramos de silencio. Se separarán los features en cuatro grupos, siendo estos los citados en el punto 3.2.3, LHMR, c0, f0 y LPC, de los cuales, obtendremos las siguientes 21 características:

Características acústicas	
LHMR K=1	*f0 media*
LHMR K=2	*f0 desviación*
LHMR K=3	*f0 simetría*
LHMR K=4	*f0 kurtosis*
LHMR K=5	*f0 % tramos sonoros*
LHMR K=6	*f0 rango*
LHMR K=7	*LPC media*
Δc0 media	*LPC desviación*
Δc0 desviación	*LPC simetría*
Δc0 simetría	*LPC kurtosis*
Δc0 kurtosis	

***Tabla 1**: Características acústicas*

Los features serán obtenidos tras el análisis de toda nuestra base de datos, que serán algo más de 3.000 registros. Esto implica una gran carga computacional, y, por tanto, es un proceso de extracción lento.

A continuación se explicarán las funciones que se han utilizado a para extraer todos los features anteriormente nombrados. Se comenzó con la extracción del LHMR, para ello, se utilizó la función "SRMR" del SRMR ToolBox [14]. Con esta función, se reciben dos parámetros, el primer, LHMR que es el que nos interesa, y el segundo es el espectro de modulación, que no nos ha sido útil en este caso. A continuación, para entrar más en detalle, se muestran los parámetros que necesita dicha función en la **Figura 6**.

```
[ratio, energy] = SRMR(s, fs, 'fast', 0, 'norm', 0, 'minCF', 4, 'maxCF', 128)
```

***Figura 6:** Función SRMR de Matlab*

La mayor parte de los parámetros, los hemos adaptado a nuestra situación, mientras que se ha dejado el valor por defecto en el caso 'fast', 0, siendo este implementación rápida desactivada y en 'norm', 0, desactivando la normalización del espectro de modulación. El valor "s" corresponde con los valores de entrada que van a ser analizados, en este caso se introducirán las muestras preprocesadas por la función

vadsohn. “fs” es la frecuencia de muestreo, se ha fijado a 16000Hz. Por último se tienen ‘minCF’ y ’maxCF’, correspondiendo a la frecuencia central del primer y último filtro respectivamente. Los valores por defecto son 4 y 128, que han sido sustituidos en este caso por 2 y 64.

En segundo lugar, se obtuvo el feature “Δc0” (la primera derivada del coeficiente cepstral de orden 0). Para ello se tuvo que utilizar la función de Matlab “melceps” [15] de la toolbox Voicebox [16]. Esta función admite diversos parámetros que para nuestro caso fueron innecesarios. Como muestra la **Figura 7**, existen 9 parámetros que pueden ser modificados, mientras que en este caso, se han utilizado únicamente los primeros 4. Estos corresponden a las muestras de entrada (“s”), extraídas tras utilizar “vadsohn”, la frecuencia fundamental como “fs” fijada a 16000Hz, “w” (modo de operación) admite varias posibilidades donde se ha elegido ‘Ed’ que indica que se incluye la log-energia y la primera derivada de los parámetros mel-cepstrales. “nc” se refiere al máximo orden de los coeficientes mel-cepstrales a extraer. En este caso, se establece a 0 ya que sólo se necesita el coeficiente de orden 0. Para el resto de parámetros de configuración se han utilizado los valores por defecto.

```
function [c,tc]=melcepst(s,fs,w,nc,p,n,inc,fl,fh)
```

***Figura 7**: Función melceps completa*

La tercera función se ha utilizado para obtener la frecuencia fundamental, y se ha hecho mediante “fxrapt” [17] de la toolbox Voicebox. Ésta devuelve dos variables, la frecuencia fundamental de cada frame, y el inicio y fin del mismo. Además, admite varios parámetros, esto son: las muestras de entrada (“s”), la frecuencia de muestreo (“fs”), el modo, que en este caso se ha utilizado ‘u’, siendo esta un indicativo para incluir las tramas sordas a las que se asigna un valor de frecuencia fundamental de “NaN” que hay que filtrar antes de calcular los estadísticos de f0 correspondientes. Por último, el parámetro “q”, que indica la estructura de los datos, no ha sido necesario incluirlo. Todo esto se puede encontrar en la **Figura 8**.

.

```
[fx,tt]=fxrapt(s,fs,mode,q)
```

***Figura 8**: Función fxrapt de Matlab*

En cuarto y último lugar, hemos obtenido el residuo del LPC utilizando la función “lpcresidual” de la toolbox COVAREP [18]. Viene definida en la **Figura 9**. Sus parámetros de salida son los coeficientes y el residuo LPC, en este caso es útil el residuo. Y como parámetros de entrada admite, de igual forma que antes, las muestras de la señal de voz tras ser procesada por la función vadsohn (“x”), “L” corresponde con

el tamaño de la ventana en muestras, definido como 0,02 por la frecuencia de muestreo (que corresponde con 20 ms), el "shift" es el desplazamiento de ventana en muestras, fijado a 0,01 por la frecuencia de muestreo (que corresponde con 10 ms). El último parámetro de entrada es el orden de la predicción, definido a 12.

```
[res,LPCcoeff] = lpcresidual(x,L,shift,order);
```

Figura 9: *Función lpcresidual de Matlab*

4.3. Selección *features*

De antemano, no podemos predecir qué *feature* va a ser más útil para estimar el grado de inteligibilidad de la voz, por tanto, es necesario establecer un criterio que nos permita seleccionar los mejores de entre el conjunto de features inicialmente calculados. El criterio utilizado ha sido el de la correlación de Pearson. Por tanto, se calcula dicha correlación entre cada uno de features individuales y los scores de inteligibilidad reales y se seleccionan las características que presentan mayor valor absoluto de índice de correlación. Este proceso se realiza utilizando únicamente los ficheros de audio del conjunto de entrenamiento.

Desde el punto de vista práctico, tras la extracción del conjunto inicial de features, se obtiene una matriz con tantas filas como señales de voz, y tantas columnas como parámetros hayamos programado, en nuestro caso 21. Esto supone una gran cantidad de datos, y como hemos dicho, no todos podrían ser apropiados o válidos para nuestros resultados, por ello, es necesario establecer unos criterios que nos permitan escoger los más adecuados.

Para hacer dicho filtro en la elección de features, hemos calculado las correlaciones individuales de todos ellos, de Pearson y Spearman sobre el conjunto de entrenamiento. Una vez calculadas, se decidió comenzar eligiendo un mínimo de 6 características basándonos en la correlación de Pearson, lo que supone escoger aquellas cuya correlación en valor absoluto, se aproxime más a 1. Según se detallará más adelante, se seguirá utilizando este criterio progresivamente en el desarrollo de las pruebas, con objeto de hacer experimentos con subconjuntos de features de mayor tamaño.

Figura 10*: Diagrama bloques - Selección features*

4.4. Análisis de los regresores

Para poder realizar nuestro predictor y poder estimar el grado de inteligibilidad de nuevas señales, se han utilizado los cuatro tipos de regresores que se comentaron en el apartado 3.2.5.

En primer lugar fue necesario el proceso de entrenamiento de dichos regresores. Esto se llevó a cabo con unas señales de voz concretas de la base de datos, contenidas en el conjunto de entrenamiento. Nótese que de todas las señales disponibles, se escogieron unas determinadas para proceder con el entrenamiento, las cuales no serán usadas cuando tengamos que realizar las pruebas de evaluación del predictor.

Para realizar el entrenamiento, se elaboró un fichero "txt" con los nombres de los ficheros elegidos y su ubicación, éste se debe introducir en el sistema el cual se encargará de procesarlos, y así poder obtener el sistema "entrenado" para predecir las futuras señales de entrada.

Se ha comentado que es posible utilizar diferentes regresores para este proceso, por tanto, se utilizó la función de Matlab "fitsrvm", que entrena cada uno de estos modelos de regresión. El entrenamiento se realizó mediante el procedimiento de validación cruzada, que consiste en dividir el conjunto de datos de entrenamiento en dos subconjuntos, entrenar el regresor sobre uno de los subconjuntos con ciertos valores de configuración y utilizar el otro (subconjunto de validación) para validar los resultados. Este proceso se repite para un barrido determinado de los valores de configuración del regresor y finalmente, se selecciona aquélla que ha producido los mejores resultados sobre el conjunto de validación.

Esta función admite una serie de parámetros entre los cuales es imprescindible definir, tanto la variable a predecir como la respuesta, 'KernelFunction' que indica el tipo de kernel del regresor, donde para las pruebas se usarán los tipos 'linear', polynomial' y 'gaussian', también se tendrá que establecer el parámetro 'Standardize' como true, ya que se pretende normalizar los datos de entrenamiento, y por último, en el caso de elegir una regresión polinominal, se tendrá que definir el 'polynomialOrder', siendo es este caso 2 y 3. A continuación se muestra en la

Figura **11** una parte del código del entrenamiento del predictor, donde se aprecian los parámetros de la función anteriormente citada.

```
switch(regr_SVR)
    case 0
       fprintf('LINEAR SVR regression\n');
       svrModel = fitrsvm(feat_train,lab_train(:,3),'KernelFunction','linear',...
          'Standardize',true,'OptimizeHyperparameters','auto',...
          'HyperparameterOptimizationOptions',...
          struct('AcquisitionFunctionName','expected-improvement-plus'))

    case 1
       fprintf('POLINOMIAL SVR regression - ORDER 2\n');
       pol_order_svr = 2;
       svrModel = fitrsvm(feat_train,lab_train(:,3),'KernelFunction','polynomial',...
           'PolynomialOrder',pol_order_svr,'Standardize',true,...
           'OptimizeHyperparameters','auto','HyperparameterOptimizationOptions',...
            struct('AcquisitionFunctionName','expected-improvement-plus'))

    case 2
       fprintf('POLINOMIAL SVR regression - ORDER 3\n');
       pol_order_svr = 3;
       svrModel = fitrsvm(feat_train,lab_train(:,3),'KernelFunction','polynomial',...
           'PolynomialOrder',pol_order_svr,'Standardize',true,...
           'OptimizeHyperparameters','auto','HyperparameterOptimizationOptions',...
            struct('AcquisitionFunctionName','expected-improvement-plus'))
    case 3
       fprintf('GAUSSIAN SVR regression\n');
       svrModel = fitrsvm(feat_train,lab_train(:,3),'KernelFunction','gaussian',...
           'Standardize',true,'OptimizeHyperparameters','auto',...
           'HyperparameterOptimizationOptions',...
           struct('AcquisitionFunctionName','expected-improvement-plus'))
```

Figura 11: Código para el entrenamiento de los diferentes regresores considerados

4.5. Predicción y evaluación

Una vez se hayan utilizado las señales “train” con el predictor, ya lo tendremos preparado para introducir nuevas señales y predecir la inteligibilidad. Para ello, se tienen las señales de voz denominadas “test”, también pertenecientes a la base de datos nombrada, y en ningún caso se utilizará una señal usada en el proceso de entrenamiento anterior para esta fase.

Al igual que para el entrenamiento, se elaboró un listado con los nombres de los ficheros, de las señales que se evaluarán con el sistema. Éstas se introducen en el sistema, de las cuales obtendremos los resultados que nos permitirán evaluar el corrector funcionamiento del predictor, y por tanto, ver si cumple nuestro objetivo.

Para poder realizar dicha evaluación, las medidas que fueron calculadas, con el fin de evaluar el proceso, fueron la raíz del error cuadrático medio y la correlación de Pearson. Con elloa, se pudieron obtener las conclusiones pertinentes.

5. Pruebas y resultados

Como se ha comentado anteriormente, el objetivo de cualquiera de nuestras pruebas es, tras introducir una señal de entrada en nuestro predictor, obtener como salida un parámetro de inteligibilidad de la misma. Estos resultados se irán comparando progresivamente con el resto de pruebas que se realicen. Se usarán las correlaciones y la raíz del error cuadrático medio para poder evaluar los valores obtenidos. Se van a separar cada una de las pruebas por tipo, ya que se comienza con un ensayo de referencia, y a partir de ahí, se hacen unas pruebas independientes.

5.1. Experimento base y pruebas comparativas

Este punto abarca las pruebas ejecutadas partiendo de algún ensayo o experimento ya realizado, buscando mejorar los resultados existentes. Después de recaudar información sobre el tema, se encontraron diversos artículos en los que se llevaron a cabo experimentos de todo tipo, de los cuales pudimos usar algunos como referencia para nuestro trabajo.

En primer lugar, quisimos hacer un experimento similar al que se realizó en el *paper* [8] "Automated Dysarthria Severity Classification for Improved Objetive Intelligibility Assessment of Spastic Dysarthric Speech" donde se propuso un sistema para predecir el grado de inteligibilidad del habla de pacientes con disartria espástica. Para ello, utilizaron un vector de seis características acústicas, que corresponde con un subconjunto de las 21 que fueron extraídas en nuestro sistema previamente. Estas fueron: la kurtosis del residuo de la predicción lineal, la desviación estándar de la primera derivada del coeficiente cepstral de orden 0 (que está relacionado con la log-energía de la señal), el LHMR con K = 4, la desviación estándar de la frecuencia fundamental, el rango de la frecuencia fundamental, y, por último, el porcentaje de tramas sonoras.

.

$$S_1 = \{\mathcal{K}_{LP},\ \sigma_{\Delta c_0},\ \mathrm{LHMR},\ \sigma_{f_0},\ \Delta_{f_0},\ \%\mathcal{V}\}.$$

***Figura 12**: Vector de features para el experimento [8]*

Los datos que fueron analizados en esta ocasión, fueron 10 locutores pertenecientes a la base de datos "Universal Access", donde cada uno tuvo que leer 765 palabras. Se utilizaron diferentes tipos de regresión: lineales, cuadráticas y con la distancia Mahalanobis, y el sistema se evaluó en función de la raíz del error cuadrático

medio y la correlación de Pearson. Los resultados obtenidos en función de la correlación de Pearson pueden observarse en la **Tabla 2**.

.

	Clasificador		
	Lineal	**Cuadrático**	**Mahalanobis**
S1	0,839	0,825	0,835

***Tabla 2**: Resultados numéricos [5]*

A continuación se pueden observar los resultados obtenidos tras extraer los mismos 6 features del artículo citado anteriormente, pero siendo aplicados para nuestra base de datos.

	Tipo de regresor			
	SVM Lineal	**SVM Poli 2**	**SVM Poli 3**	**SVM Gaussiano**
RMSE	23.5	22.41	22.32	19.59
Pearson (R)	0.8073	0.8178	0.8142	0.8621

***Tabla 3:** Resultados con features del paper [8]*

En vista de los resultados anteriores, hay que destacar que su base de datos tenía 10 interlocutores, mientras que la que se está analizando en este TFG tenía 15.

Como se puede apreciar, los mejores resultados provienen, en caso del artículo citado, de la regresión lineal, y en nuestro caso es más favorable con el regresor gaussiano. El mejor de ambos resulta ser el caso gaussiano con un 0,8621 como correlación de Pearson frente a 0,839. Aunque, como se ha comentado, existe una diferencia considerable en cuanto a la base de datos. Es por eso, por lo que a pesar de haber conseguido unos mejores resultados, no es posible realizar una comparación directa entre ellos.

.

5.2. Pruebas incrementando el número de features seleccionados

El objetivo principal de esta sección, es basarse en los resultados obtenidos anteriormente, en el punto 5.1, donde se escogieron los 6 features propuestos del ensayo y fueron aplicados a la base de datos actual. A este experimento se le llamará en adelante "prueba base".

Comentar de antemano, que para todos los experimentos se analizaron lo p-valores. Según se ha explicado en el epígrafe 3.2.5, es el valor que indica si los resultados son significativos o no. En cualquiera de las situaciones, resultaba muy próximo a cero, indicativo de que todos ellos presentan gran significancia estadística.

Los features utilizados en la prueba base cuyos resultados pueden verse en la **Tabla 3,** fueron los mismos 6 que se indicaban en el artículo. En el siguiente experimento, también se utilizaron 6 features, pero, en este caso, se seleccionaron de entre el conjunto inicial de 21 características siguiendo el criterio de las correlaciones individuales. Es cierto que, aunque la correlación de un parámetro resulte favorable de forma individual, no es garantía de que sea efectivo al añadirlo a otro u otros parámetros ya que, podría ser redundante o empeorar el resultado al aumentar la dimensionalidad del vector de características. A pesar de esto, como primera prueba y con el propósito de utilizar la misma dimensionalidad que en el artículo de referencia [5], se utilizaron los 6 features con mejor correlación de Pearson. Estos 6 fueron la media y la desviación de $\Delta c0$, la media, desviación y rango de la frecuencia fundamental (f0), y por último, la simetría del residuo LPC[1]. Con estos nuevos features, los resultados de la regresión, se muestran en la **Tabla 4**.

	Tipo de regresor			
	SVM Lineal	**SVM Poli 2**	**SVM Poli 3**	**SVM Gaussiano**
RMSE	19,69	17,89	16,75	15,06
Pearson (R)	0,8554	0,8995	0,9174	0,9303

Tabla 4: *Resultados con 6 mejores features*

Observando ambas tablas (**Tabla 3** y **Tabla 4**) se ve de forma clara, primero, que en ambos casos, el mejor regresor es el gaussiano, y en segundo lugar, que el criterio de selección de features, mediante correlaciones individuales, presenta mejores resultados que con los features de la prueba base.

[1] *Linear Predictive Coding*

Se presentan unas gráficas comparativas como evidencia clara de los resultados anteriores, utilizando cada uno de los regresores.

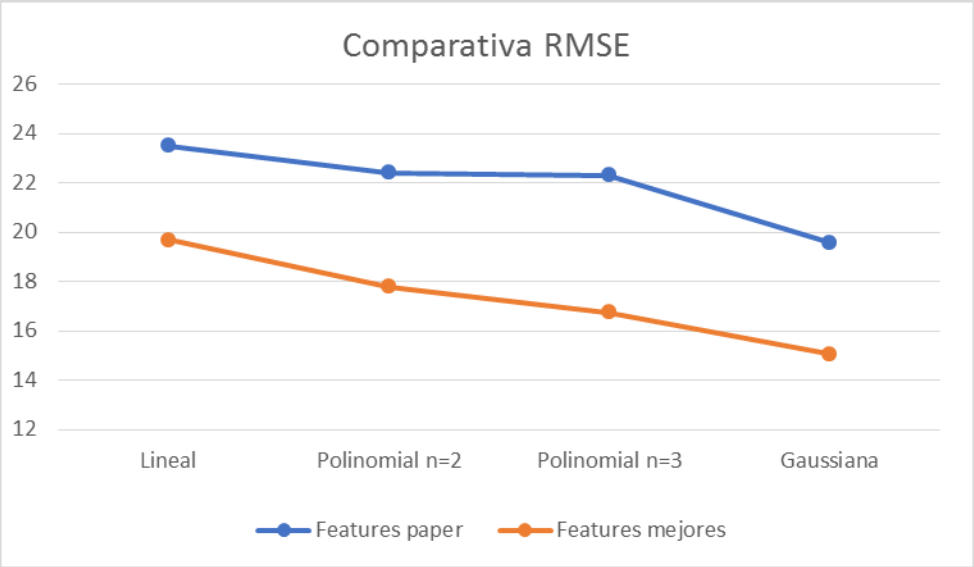

Figura 13: *Comparativa regresión (RMSE)*

En la **Figura 13**, se pueden observar los resultados en cuanto a la raíz del error cuadrático medio ("Root mean square error" - RMSE), que como es natural, resulta más favorable cuanto más cercano a cero sea; por tanto, se puede ver un mejor resultado eligiendo los features con correlación individual, que con los utilizados en la prueba base. Por el contrario, al analizar la correlación de Pearson, al ser valores entre 0 y 1, se ve un resultado más favorable, cuando el valor sea más cercano a 1.

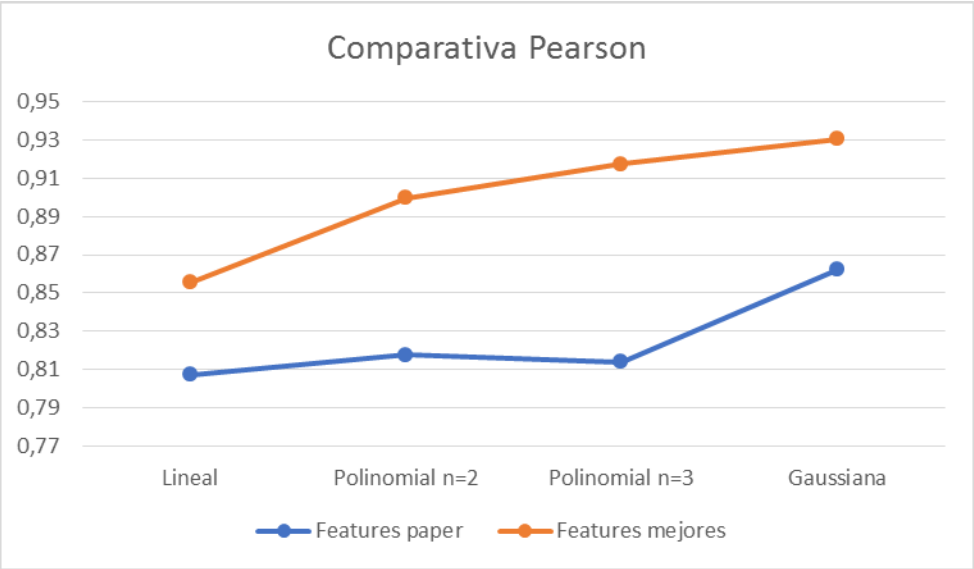

Figura 14*: Comparativa regresión (Pearson)*

En la **Figura 14** se vuelve a confirmar, que los resultados son más exitosos en la segunda prueba que en el caso de la prueba base mostrándose en la comparativa de la correlación de Pearson.

Siguiendo con el orden lógico, se realizó la prueba con todos los regresores y con todos los features incrementándolos de uno en uno, es decir, desde 6 hasta 21 de forma progresiva en pruebas diferentes.

A continuación, se vuelven a presentar en la **Figura 15** y en la **Figura 16**, unas gráficas comparativas, tanto del RMSE como de la correlación de Pearson utilizando el resultado de cada una de las pruebas, es decir, según el número de features introducidos en cada prueba.

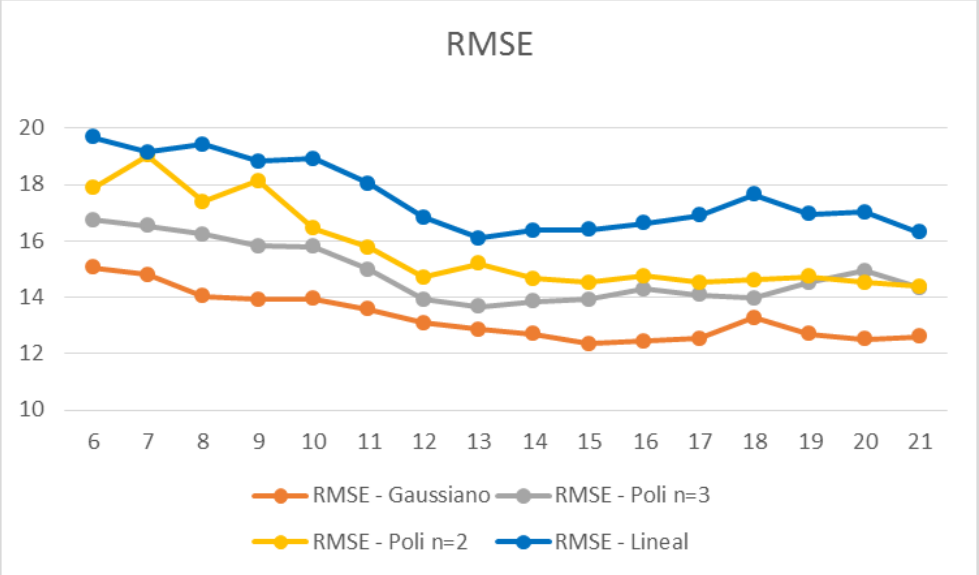

***Figura 15**: Comparativa features RMSE*

Según se puede ver en la **Figura 15**, el mejor resultado obtenido, según la raíz del error cuadrático medio y el **regresor gaussiano**, es en la prueba donde se introducen los 15 mejores features. Éste resultado es muy próximo al obtenido con todos los features (12,35 frente a 12,62).

Además se observa, una tendencia descendente en cuanto al valor del RMSE según se introducen más features, a excepción del intervalo de 16 a 18 features introducidos, donde se aprecia una subida. Esto es un indicativo, de que los features utilizados en esas ocasiones, podrían eliminarse de las predicciones ya que proporcionan un resultado negativo en las pruebas.

Por otro lado, podemos ver el gráfico comparativo según la correlación de Pearson en la **Figura 16**.

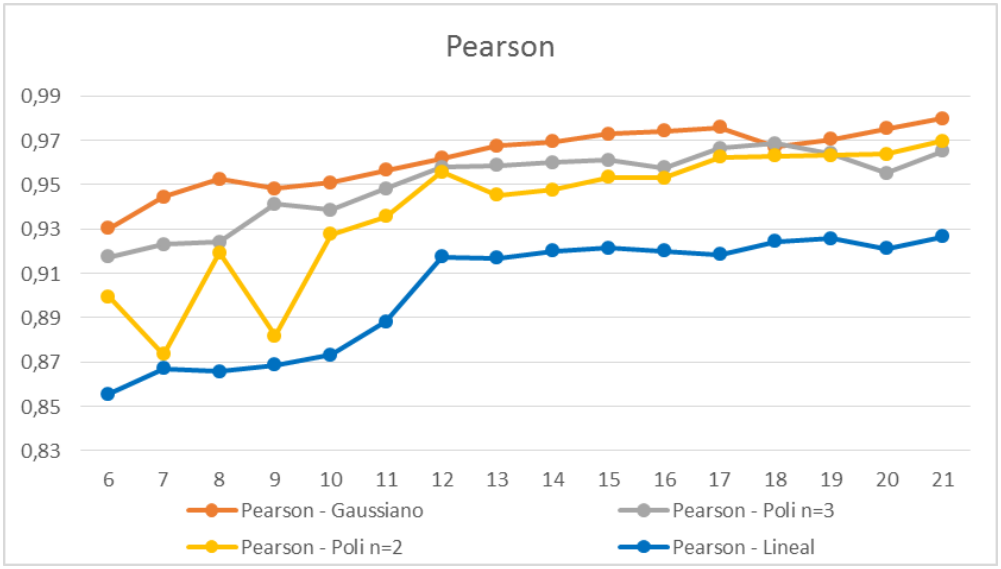

***Figura 16**: Comparativa features Pearson*

Al observar la gráfica, se puede ver que en esta ocasión, basándose en la correlación de Pearson, el mejor resultado obtenido objetivamente, es en la última prueba, donde se introducen todos los features.

En esta ocasión, se aprecia una tendencia ascendente, salvo en dos ocasiones, lo que nos vuelve a indicar un feature no favorable para los resultados, que podría tratar de eliminarse para obtener un mejor resultado conjunto.

Analizando las pruebas de forma global, se ha visto que por una parte, la situación más favorable resulta en el experimento con 15 features según RMSE, y con 21 features según Pearson. En la mayoría de experimentos, la situación ideal que pretende encontrarse es un mejor resultado con la menor carga computacional posible, por tanto, viendo el resultado según RMSE, conviene fijarse en qué indica la tabla de Pearson, en el caso de introducir 15 features. Se puede observar que se obtiene un valor de 0,9729, mientras que el mejor, con 21 sería de 0,9799. La diferencia entre estos dos valores, no supone una gran mejoría de forma global, ya que ambas presentan una alta correlación. Por tanto, y como conclusión final, se puede decir, que el resultado más adecuado, valorando la carga computacional y los resultados numéricos, es el caso de introducir los 15 mejores features, reflejados en la **Tabla 5** según la correlación individual y utilizando una regresión gaussiana.

Features con mejor resultado	
LHMR K=1	*f0 media*
LHMR K=5	*f0 desviación*
LHMR K=6	*f0 simetría*
LHMR K=7	*f0 kurtosis*
Δc0 media	*f0 rango*
Δc0 desviación	*LPC simetría*
Δc0 simetría	*LPC kurtosis*
Δc0 kurtosis	

***Tabla 5**: Features utilizados para mejor resultado*

A modo de resumen, se muestra a continuación la **Figura 17** como comparativa de todas las pruebas que conviene destacar. Entre ellas se encuentran, la prueba realizada con los 6 features presentados en el artículo [8], la realizada con los 6 mejores features según el criterio de correlación individual, el caso completo con los 21 features extraidos, y por último, el caso más favorable, con los 15 features de la **Tabla 5**. Los resultados están basados en el regresor gaussiano, ya que como se ha visto en las figuras anteriores, siempre ha presentado unos mejores resultados.

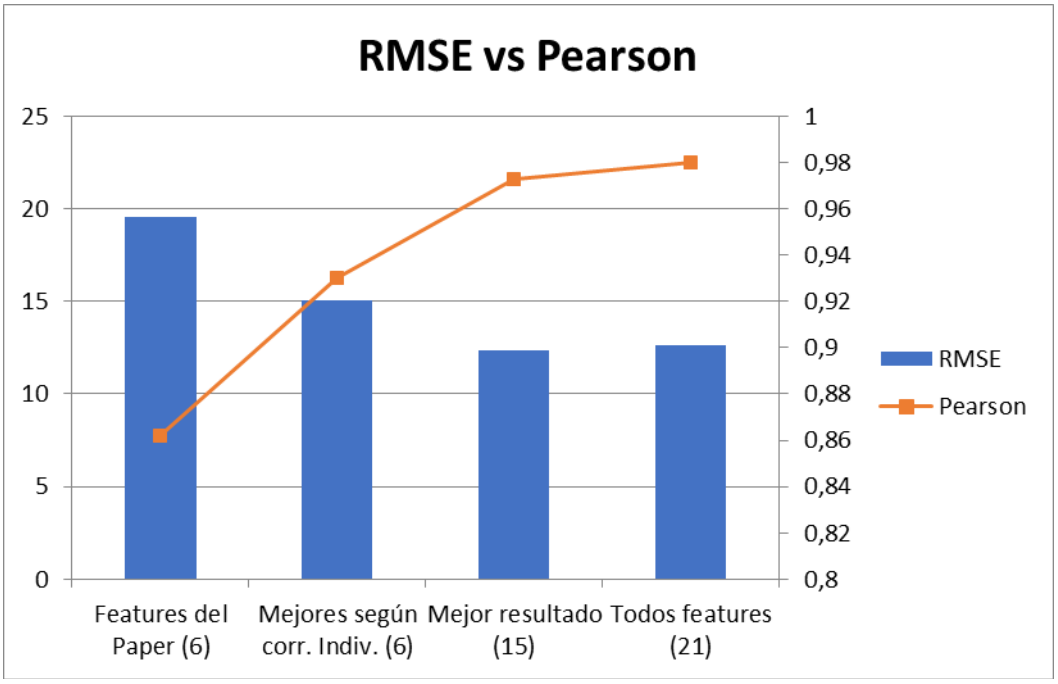

***Figura 17**: Comparativa global - Reg. Gaussiano*

En la **Figura 17** se han mostrado los resultados de las 4 situaciones anteriores tanto a nivel de RMSE (eje izquierdo) como a nivel de correlación de Pearson (eje derecho).

6. Planificación y presupuesto

A continuación, se detallará como se han planificado las diferentes etapas y fases del proyecto, incluyendo las fechas, documentos gráficos, tablas ilustrativas y los costes.

6.1. Fases

Detallaremos las fases del proyecto, con las respectivas funciones en cada una de ellas. Debido al tiempo disponible y estimación temporal de realización del trabajo, el objetivo fue presentarlo a tribunal en octubre, es decir, entrega final el 20 de septiembre. Se comenzó a finales de mayo, y hubo en compromiso de entrega de 3 meses, a fecha de 30 de agosto, con un margen de 3 semanas para posible evaluación del tutor y posibles modificaciones. Para ello, tuvimos que invertir un gran número de horas para entrar en plazo.

El proyecto se dividió en:

- **Toma de contacto**: Se comienza el proyecto con varias reuniones con el tutor, para una primera toma de contacto. En ellas se expone el tipo de proyecto que va a realizarse, el tiempo estimado, tanto disponible como de carga de trabajo, y la clase de herramientas necesarias y disponibles.
- **Búsqueda de información**: De estas primeras reuniones, se concluye que es necesario hacer una primera fase de análisis, donde se procede a la búsqueda y puesta en común de la información que se dispone. En las primeras semanas, y hasta el inicio de las pruebas, la búsqueda es más exhaustiva ya que hay que establecer el enfoque que se le dará al proyecto. Una vez que se comienza con las pruebas y se tiene una base, la búsqueda continua, ya que es necesario redactar con detalle todo el desarrollo del trabajo, aunque, esto se realiza de forma menos intensiva que al inicio.
- **Pruebas**: Tras recopilar información y obtener los datos que se van a usar, procedemos a comenzar las pruebas. Como se ha comentado anteriormente en diferentes apartados del documento, existe una gran carga computacional, por lo que esta etapa duró algo más de lo esperado. Se realizó con el programa MATLAB, y hubo diferentes reuniones con el tutor para la validación de las mismas.
- **Redacción**: Una vez finalizadas las pruebas, se pasa a la fase más larga, que es la redacción del proyecto. Esto se combina con la búsqueda de

información, ya que para diversos apartados es necesario una información general y el contexto de algunos aspectos técnicos para el correcto entendimiento.

- **Entrega provisional/revisión**: Una vez hechos todos los apartados del trabajo, se procede a la entrega provisional al tutor para una primera revisión. Se realizan diferentes modificaciones tras las cuales vuelve a ver revisiones hasta la entrega final.
- **Entrega final:** Por último, tras todas las revisiones y dejarlo todo a punto, se abre el plazo de entrega final para la corrección del tribunal.

Concluimos, por tanto, que toda la realización del proyecto se ha llevado a cabo en un total de **18 semanas**.

Fases	Periodo	Fecha Inicio	Fecha Fin
Toma de contacto	2 semanas	29/05/2017	12/06/2017
Búsqueda de información	12 semanas	12/06/2017	30/08/2017
Pruebas	3 semanas	03/07/2017	24/07/2017
Redacción	11 semanas	10/07/2017	20/09/2017
Entrega provisional/ Revisión	3 semanas	04/09/2017	20/09/2017
Entrega final	1 semana	20/09/2017	28/09/2017

Tabla 6: *Detalle de fases y fechas de proyecto*

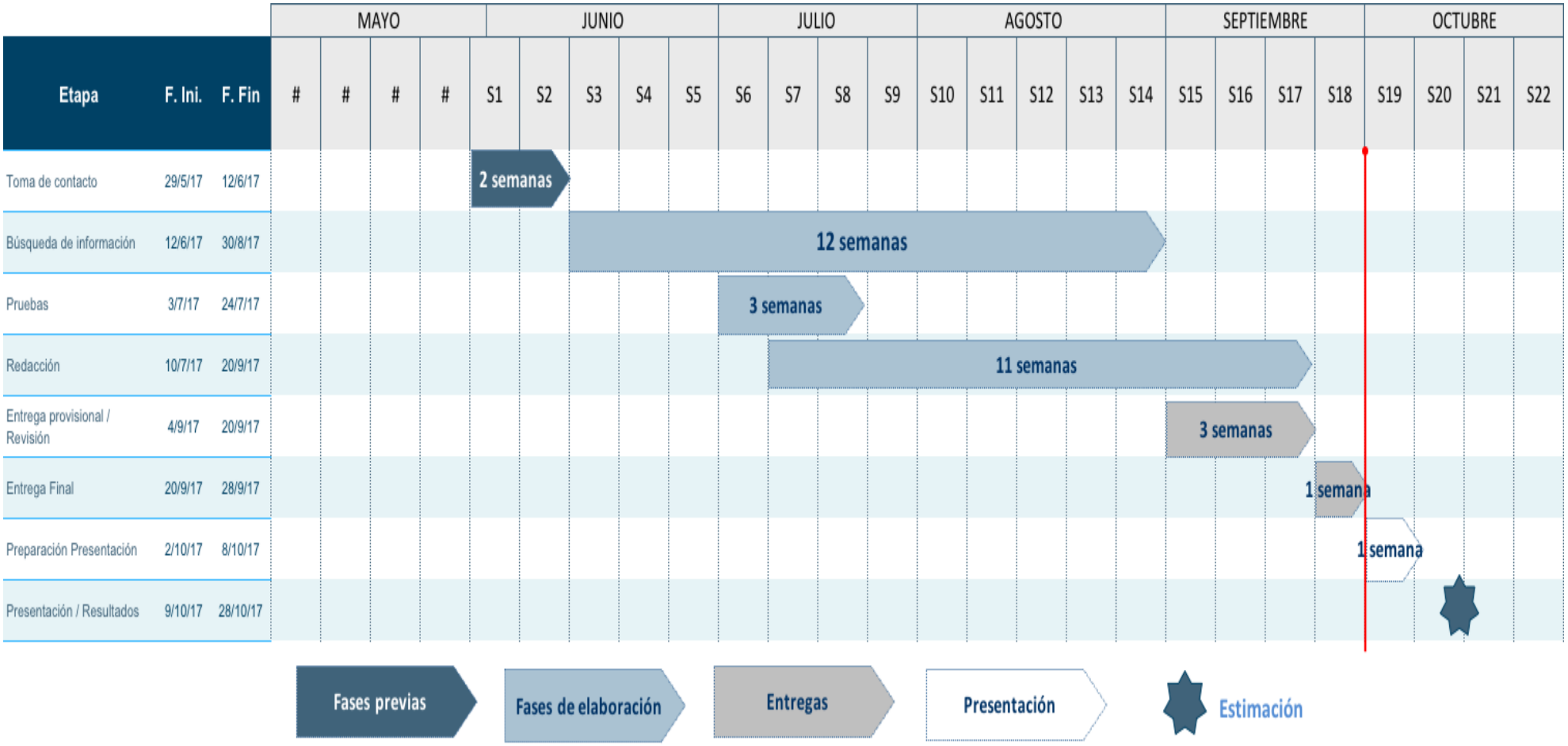

***Figura 18**: Diagrama Gantt*

6.2. Presupuesto

En este capítulo se detallarán los costes empleados para la realización del proyecto, tras lo que indicaremos el presupuesto total del mismo. Para ello, mostraremos unas listas diferenciadas por tipos de recurso.

Recursos físicos

En primer lugar, se tiene el ordenador desde el cual se ha realizado la mayor parte del trabajo, tanto de búsqueda de información como de redacción. Por tanto, se incluye un coste de 1.000€.

En segundo lugar, también se ha utilizado una impresora HP valorada aproximadamente en 100€ para la impresión de documentación como ensayos, artículos o gráficas comparativas entre otros. Además del material necesario para la impresora, valorado en 40€, como papel de impresora y tinta.

Para todo el proceso de elaboración del proyecto, se ha usado una memoria USB de 8Gb estimada en 10€.

Consideramos que es conveniente añadir una serie de costes categorizados como "Otros" en los cuales se engloban cosas mínimas como podrían ser maletín portaordenador, medio de transporte para reuniones, material de oficina, recursos de conexión a internet, conexión de luz, etc. Todo ello ha sido valorado en 40€.

Recurso	Coste
Ordenador MacBook Air	1.000€
Impresora HP	140€
Memoria USB	10€
Otros	40€

***Tabla 7**: Costes de recursos físicos*

Recursos humanos

Por otra parte, encontramos lo más importante, que es el coste por persona implicada en el proyecto, en este caso ha sido una ingeniera senior y una junior. La tarifa de los mismos se valora en 25€/h en el caso del ingeniero senior, y 15€/h en el caso del junior. El cálculo del importe total, se lleva a cabo teniendo en cuenta las 18 semanas de trabajos, siendo una media de 5 horas de trabajo semanales para Ascensión Gallardo, tutora e ingeniera senior del trabajo, y 25 horas de trabajo

semanales para Blanca Valdivielso, alumna y autora del trabajo. El resultado de los cálculos será de 2.250€ por parte del ingeniero senior, y 6.750€ por parte del junior. Estos costes se ven reflejados en la **Tabla 8**.

Recurso	Horas	Tarifa	Coste
Ingeniero Senior	90	25€/h	2.250 €
Ingeniero Junior	450	15€/h	6.750 €

Tabla 8: *Coste de recursos humanos*

Recursos software

Para la redacción y pruebas en nuestro trabajo, se tuvieron que utilizar diferentes recursos software, entre ellos se encuentra el paquete Office, del cual utilizamos los programas Excel, Word y Powerpoint, sobre todo en la parte de la redacción y elaboración de algunas tablas. Por otra parte, y como herramienta principal, se utilizó MATLAB, que nos sirvió de base para la programación de funciones que nos dieron los resultados de las pruebas necesarias. También pudimos extraer diferentes gráficos comparativos que se ilustran a lo largo de la memoria. Además de todo esto, es necesaria la instalación del sistema operativo en el ordenador en cuestión, que en nuestro caso fue uno de marca Apple, el cual utiliza macOS. Este último coste no es imputable en esta ocasión ya que queda reflejado en la parte del dispositivo físico. El resto de licencias y herramientas, han sido proporcionadas por la Universidad Carlos III con coste 0€, es por esto, que la parte de Software, no supone ningún coste adicional al proyecto.

Por tanto, teniendo en consideración todos los conceptos anteriormente citados, el presupuesto para la realización de todo este TFG asciendo a **diez mil ciento noventa euros**, viéndose reflejado en la **Tabla 9**.

Concepto	Coste
Recursos físicos	1.190 €
Recursos humanos	9.000 €
Recursos software	0 €
Total	**10.190 €**

Tabla 9: *Presupuesto total del proyecto*

7. Conclusiones y líneas futuras

En este capítulo de cierre, se pretende relacionar los objetivos planteados en el Capítulo 1, analizarlos de forma específica y concluir si el trabajo realizado ha tenido un resultado exitoso. Además, como es común, durante todo el proyecto, se han planteado una serie de pruebas, de las cuales solo una parte se ha podido llevar a cabo, lo que plantea una serie de líneas futuras, que podrían ser interesantes para estudios próximos.

7.1. Conclusiones

En primer lugar, es necesario hacer referencia al Capítulo 1 del trabajo, donde se detallan los objetivos concretos que se querían conseguir y se analizan de manera individual con el fin de comprobar si se han resuelto con éxito.

El objetivo principal que se quería conseguir con este trabajo, era la realización de un predictor, que introduciendo una señal de voz de una persona con disartria, procesara las señales, y con esto se obtuviera un resultado de inteligibilidad del habla, para ser aplicado a personas con Parkinson. Al tener las etiquetas de referencia con el valor de inteligibilidad predefinido para cada interlocutor, fue necesario realizar las comparaciones con respecto a éstas, para poder indicar si el predictor se había creado correctamente. Para dicha comparación, se utilizaron, la raíz del error cuadrático medio (RMSE) y la correlación de Pearson. De ésta última, se obtuvieron resultados del orden de 0,9, lo que indicó una correlación alta y por tanto, que el predictor fue realizado de forma adecuada, obteniendo unos resultados superiores al experimento base de referencia. Además, el resultado por parte del RMSE también presenta una situación favorable para admitir las pruebas como válidas, ya que presenta valores entre 12 y 14, considerándose aceptable.

En base a esto, viendo que se predice con un alto nivel de correlación la inteligibilidad en señales de voz, se ha conseguido que funcionara en personas con disartria provocada por una parálisis cerebral. Por tanto, ante el éxito obtenido en la aplicación del predictor en estas grabaciones, se puede concluir que el método muy posiblemente se puede aplicar de igual manera o con pequeñas variaciones, a señales de voz con disartria provocada por el Parkinson.

El sistema desarrollado consta básicamente de dos módulos: extracción de características y el de regresión. Con respecto al primero, se han probado diferentes parámetros acústicos derivados del valor LHMR, derivada del coeficiente cepstral de primer orden, frecuencia fundamental y residuo de predicción lineal y se ha mostrado que es posible seleccionar las características más relevantes para nuestra tarea utilizando el criterio de la correlación individual de Pearson. Con respecto al segundo

módulo, para todas las pruebas se han aplicado diferentes regresores basados en SVM ("Support Vector Machine"), entre los cuales se encuentran el lineal, polinomial de órdenes dos y tres y el gaussiano. En todos los experimentos realizados, se ve reflejada con claridad, la superioridad del gaussiano respecto al resto, habiéndose extraído todas las conclusiones en base a esta opción.

En definitiva, se han obtenido unos resultados exitosos tanto en la automatización del proceso, como en los resultados, ya que se demuestra que es posible predecir el grado de inteligibilidad del habla en paciente con disartria con un alto grado de correlación.

7.2. Líneas futuras

La elaboración de un trabajo de fin de grado, conlleva la elección de un tema en concreto que tiene que encajar dentro de la normativa de elaboración de la memoria. Esto en muchas ocasiones imposibilita extenderse mucho en cuanto a explicaciones y realizar un número de pruebas menor al deseado. El tener que concretar tanto en estos aspectos, plantea un gran número de líneas de investigación alternativas que podrían estudiarse basándose en este trabajo.

Por una parte, sería interesante poder crear una base de datos de habla con disartria en idiomas diversos, y realizar las mismas pruebas que hasta ahora se han realizado con señales de voz de habla inglesa. Que los resultados hayan sido positivos anteriormente, no supone que al analizarlo en un idioma diferente vaya a resultar igualmente positivo, ya que cada lengua tiene diferentes entonaciones y son fonéticamente diferentes, por ello, se considera interesante poder realizar este tipo de pruebas en el futuro.

Además de hacer variaciones en cuanto al idioma, también podría ser interesante, aplicar dicho predictor a personas con diferentes enfermedades. El experimento ha resultado exitoso, ha conseguido hacer una predicción suficientemente precisa de la disartria de cada paciente, y en consecuencia, como síntoma principal, aplicable a la enfermedad de Parkinson, cosa que podría ser útil en otro tipo de patología, en la cual los pacientes también padezcan disartria.

La base de datos utilizada, estaba compuesta tanto por hombres como por mujeres, y se han realizado las pruebas de forma conjunta, sin distinción alguna. Como es sabido, el tono vocal o frecuencia de la voz en hombres y mujeres, es diferente, incluso varía según la edad. Ante esto, se valora la posibilidad de obtener unos resultados alternativos a los mostrados, y que podrían ser objeto de estudio.

En el apartado de pruebas, se especifica que el sistema se ha basado en un conjunto de features seleccionados según su correlación individual. Esto fue una elección inicial de diseño que podría ser modificada en el futuro, ya que podrían utilizarse otros criterios como por ejemplo, la información mutua. Además, las pruebas se realizaron de forma progresiva, utilizando en primer lugar 6 features, y subiendo

hasta los 21 basándose en la correlación individual. Otra posible forma de hacerlo, hubiera sido probar todas las posibles combinaciones, hasta encontrar la que tuviera el mejor resultado. Esto supondría un proceso laborioso que implicaría mucha carga computacional, por lo que podría ser un buen método para un proyecto de mayor alcance y con más recursos para afrontar esta carga de cómputo.

8. Bibliografía

[1] R. García-Ramos, E. López Valdés, L. Ballesteros, S. Jesús y P. Mir, «The social impact of Parkinson's disease in Spain: Report by the Spanish Foundation for the Brain,» *Report by the Spanish Foundation for the Brain. Neurología (English Edition),* 2016.

[2] T. Perera y W. Thevathasan, «An Introduction to Parkinson's Disease».

[3] Centro de investigación Biomédica en Red Enfermedad Neurodegenerativas - "Síntomas no motores del Parkinson" - https://ciberned.es/noticias/blog/411-sintomas-no-motores-del-parkinson-.html.

[4] R. A. González y J. A. Bevilacqua R., «Las disartrias,» 2012.

[5] D. Martinez, P. Green y H. Christensen, «Dysarthria Intelligibility Assessment in a Factor Analysis Total Variability Space,» *Proceedings of Interspeech.,* 2013.

[6] R. Hummel, W.-Y. Chan y T. H. Falk, «Spectral Features for Automatic Blind Intelligibility Estimation of Spastic Dysarthric Speech,» *En Twelfth Annual Conference of the International Speech Communication Association.,* 2011.

[7] H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. Huang, K. Watkin y S. Frame, «Dysarthric Speech Database for Universal Access Research,» 2008.

[8] M. Sarria Paja y T. H. Falk, «Automated Dysarthria Severity Classification for Improved Objective Intelligibility Assessment of Spastic Dysarthric Speech,» de *En Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[9] X. Valero y F. Alías, «Análisis de la señal acústica mediante coeficientes cepstrales bio inspirados y su aplicación al reconocimiento de paisajes sonoros,» *Acústica,* 2012.

[10] C. García y D. T. apias, «"La frecuencia fundamental de la voz y sus efectos en reconocimiento de habla continua"».

[11] E. Suárez y J. Carmona, «Tutorial sobre Máquinas de Vectores Soporte,» 2014.

[12] J. Sohn, N. Kim y W.Sung, «A statistic model-based voice activity detection,» *IEEE signal processing letters,* 1999.

[13] VoiceBox - http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

[14] SRMR ToolBox - https://github.com/MuSAELab/SRMRToolbox.

[15] VoiceBox melcepts - http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/doc/voicebox/melcepst.html.

[16] VoiceBox - http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

[17] VoiceBox fxrapt - http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/doc/voicebox/fxrapt.html.

[18] VoiceBox covareb - https://github.com/covarep/covarep.

[19] E. Peñas Domingo, El libro blanco del Parkinson en España.

[20] T. H. Falk, R. Hummel y W.-Y. Chan, «Quantifying perturbations in temporal dynamis for automated assessment of spastic dysarthric speech intelligibility,» *Acoustics, Speech and Signal Processing (ICASSP),* 2011.

9. Anexos

Anexo 1

Features	RMSE - Gaussiano	RMSE - Poli n=3	RMSE - Poli n=2	RMSE - Lineal
6	15,06	16,75	17,89	19,69
7	14,81	16,54	19,04	19,16
8	14,05	16,25	17,4	19,44
9	13,93	15,82	18,14	18,83
10	13,95	15,81	16,44	18,91
11	13,57	15	15,78	18,05
12	13,09	13,92	14,71	16,84
13	12,86	13,67	15,21	16,09
14	12,7	13,86	14,67	16,38
15	12,35	13,92	14,54	16,4
16	12,45	14,29	14,75	16,64
17	12,53	14,09	14,54	16,9
18	13,29	13,97	14,61	17,65
19	12,7	14,53	14,73	16,96
20	12,52	14,94	14,52	17,03
21	12,62	14,35	14,4	16,3

Figura 19: *Valores según RMSE*

Anexo 2

Features	Pearson - Gaussiano	Pearson - Poli n=3	Pearson - Poli n=2	Pearson - Lineal
6	0,9303	0,9174	0,8995	0,8554
7	0,9444	0,923	0,8735	0,867
8	0,9524	0,9242	0,9193	0,8658
9	0,9481	0,9413	0,8818	0,8686
10	0,951	0,9385	0,9275	0,8732
11	0,9566	0,9482	0,9355	0,8883
12	0,9619	0,9579	0,9557	0,9174
13	0,9674	0,9586	0,9453	0,9169
14	0,9693	0,9601	0,9477	0,9201
15	0,9729	0,961	0,9533	0,9215
16	0,9742	0,9576	0,9531	0,9201
17	0,9758	0,9664	0,9625	0,9186
18	0,967	0,9687	0,963	0,9243
19	0,9705	0,9639	0,9631	0,9258
20	0,9752	0,9551	0,9637	0,9212
21	0,9799	0,965	0,9697	0,9265

Figura 20: *Valores según corr. Pearson*

Apéndices

A. English summary

The aim of this summary is to expose the main parts of this Final Degree Project for a global understanding. It will be composed for different sections, where it is going to be explained with details, the key ideas of each of them.

Chapter 1: Introduction

Introduction

Communication has always been a basic instinct in the development of humans. By nature, people interact with the environment, and therefore with their equals.

One of the factors to achieve a correct understanding between interlocutors through oral communication is the speech intelligibility of these interlocutors, which can sometimes be affected by the so-called dysarthria. Dysarthria consists of a set of alterations in speech, produced by a neurological lesion. These alterations are mainly in the muscular control, affecting fundamental parts involved in the production of the speech, which causes a problem for the understanding of the patient.

Since this is caused by neurological injuries, we know that dysarthria is a very common symptom in diseases such as Alzheimer's, cerebral palsy, multiple sclerosis, etc. In our case, and as the title of the project indicates, we will focus on how dysarthria affects people with Parkinson's disease, as it is one of the most common symptoms.

This system could determine objectively if a patient, after having undergone treatment, improves or worsens their intelligibility over time, which would provide very useful information to infer if such treatment is improving his/her disease or if, on the contrary, it is necessary to modify it. In short, it is believed that it would be very useful for the analysis, detection and monitoring of the evolution of dysarthria, and we wanted to apply it in this case to people with Parkinson's disease, because of the high number of people suffering from it today and, very likely, will suffer in the future.

From the practical point of view, the system designed in this work has been implemented using the Matlab scientific calculation program. For its development, a database with voice signals pronounced by a series of speakers with different degrees of intelligibility has been used. This system has been tested in various experiments to predict the degree of intelligibility of new speakers. In this process of system evaluation, very promising results have been obtained, being better than those achieved by another similar system of the state of the art.

.

Motivation and goals

The main objective of this project is to build a system that allows us to estimate the degree of intelligibility of a speaker by analyzing several speech signals. For this purpose, some acoustic characteristics of speech signals will be extracted and compared with the degree of intelligibility of the patient.

The motivation that has led me to perform this work, besides doing a technical project, studying different parameters and characteristics of speech, is that this study, may have a medical purpose. It would be useful that the implementation of a Matlab intelligibility predictor could have a direct impact on the monitoring of Parkinson in affected people.

Chapter 2: State of the art

First, an introduction to Parkinson's disease with the physical characteristics that it entails is needed. That will be basic for the analysis. After that, we will talk about dysarthria, which has a direct relationship with Parkinson since it is one of its main symptoms.

Parkinson disease

Parkinson is a disease that affects the nervous system focusing mainly on the motor capacity. This causes symptoms such as tremors, muscle stiffness, speech problems, etc. It is estimated that there are approximately 6 million people diagnosed with Parkinson's worldwide, 300,000 just in Spain, and this number will increase over the years due to the aging of the population. The symptoms are divided in two groups, motors and no motors. The most common motor symptoms are tremor, stiffness, and lack of movement. All symptoms usually start on one side of the body, but over time it usually extends to the other. On the other hand, non-motor problems, even though they may not be easily discovered, also contribute to deterioration in health and

should be taken into account when applying either therapy. Among the most frequent non-motors symptoms are cognitive, psychiatric, on the autonomous system, etc.

At the moment, there is no specific test that allows the diagnosis of Parkinson. Specialists in the subject are based on the medical known history and the observation of the patient to be able to determine it. For the treatment, since the disease is caused by a failure in the level of dopamine in the body, the medications that are usually prescribed try to replace or compensate this failure. These are medications that may have some secondary effects but they have proved to be effective.

Dysarthria

After describing Parkinson's disease, it can be seen that dysarthria plays an important role in its patients, since it is one of the most common and most notorious symptoms. Some of its characteristics are: the slow movement, weakness, imprecision, incoordination, involuntary movements and / or alteration of the tone of the musculature involved in speech. It can begin in a congenital way, or by some type of traumatic, infectious or degenerative disorder among others. In turn, dysarthria can follow different ways. It can be maintained stable over time, without getting better or worse with rehabilitation, being the case of cerebral paralysis or may also be a regressive dysarthria, as an effect of a post-traumatic accident, where the patient has a level that decreases over time. In contrast, it can also be a progressive dysarthria, which increases from the first symptoms. This type of dysarthria can occur in diseases such as ALS or Parkinson.

It is necessary to name, the motor processes that take part in the act of speaking, and that, therefore, some or several of them, will be affected by dysarthria. The first and most important is the breath, since the air affects the vocal cords making them vibrate, in such a way that the voice is produced, leading us to the second process that is the phonation. The second of them is the resonance, which allows to increase or decrease the vocal tone. An example of a resonator is the larynx. We also have the joint, which allows us to modify the sound through articulators, such as the tongue. And finally, we have prosody, which studies the sonorous ranges of the voice, or melodic aspects, and allows us to perceive the patient emotions, which presents patterns of rhythm and intonation.

In turn, there are functional components in the voice, which help the study and classification of dysarthria. These components are, naturalness, intelligibility, speech speed and compressibility. And of course, dysarthria is divided in some types, which are: Flaccid, spastic (it will be named along the project), ataxic, hyperkinetic, hypokinetic and mixed.

Chapter 3: Solution design

This section contains a detailed description of all the elements that take part in the accomplishment of the objective of this study. An analysis will also be made with the justifications and the reasons for the choices of acoustic parameters..

Database

We have used the Universal Access database, which consists of voice signals from 15 different speakers (4 women and 11 men) who suffer from dysarthria caused by cerebral palsy. Although not suffering from Parkinson's disease, as has been seen in the State of the Art Chapter, dysarthria is one of the main symptoms of people with Parkinson's, and therefore, it is proposed to use the database with partners with paralysis cerebral, since it can be a faithful approximation to a situation of patients with Parkinson.

The database contains the voice signals of each of 255 different words that have been recorded from each of the speakers. Among these words are the ten digits, the twenty-six words of the radio alphabet, nineteen computer commands, one hundred common words in English, and another one hundred non-common ones.

This database has been divided into two groups. On the first hand, the training group, with 155 words of the aforementioned, except for the one hundred common, that if we take into account the fifteen interlocutors we will get 2325 words. On the other hand, we have the test group, where now, it consists of one hundred common words, which in the same way, result in 1489, since eleven records could not be recorded correctly. These will be the signals that we will use for the training/development (training group) and evaluation (test group) of our system.

On the other hand, a system with a suitable data was desirable, with a reasonable number of speakers, and a sufficient number of audio files per speaker. This was considered correct, based on other similar papers mentioned in this memory. It can be observed how in some of them it works with 10 or 19 interlocutors, whereas in our case, they were 15. Finally and more important, with the objective of obtaining a system that predicts the degree of intelligibility of a voice signal, a benchmark is needed to assess whether it is correct or not. Therefore, it was necessary to choose a database that had predefined labels with the degree of intelligibility found subjectively, by way of human listeners comparison. It is not easy to find data that meet all expectations, which is why despite having many available from different classes, this was one of the few that fit the needs of the project.

Features

One of the most important parts of the project has been the choice of the features that will be used. It was necessary to allow us to compare them and to characterize the values that will be of interest. All these features are extracted at frame level (every 10 ms) and then we calculated on them, some statistics measures (average, standard deviation, etc.) at audio file level. The parameters considered are divided into four main groups, they are: LHMR ("*low-to-high modulation energy ratio")*, Δc0 ("first derivative of the cepstral coefficient of 0 order"), f0 ("fundamental frequency") and the LPC ("Linear Predictive Coding") residual.

For the choice of features, it was necessary to establish a relationship between the extracted features and some vocal characteristics. For that, we used some papers where some dimensions of the dysarthria are found. Some of the characteristics we found in these documents that were of our interest and which we considered to be useful for the results were: speed of speech, variations of intensity, monotony, etc..

Regression

To make the predictions, it was necessary to use some regressors that allowed us to train the system and evaluate the results. The regressors in this case were based on SVM (Support Vector Machines) with linear, polynomial of different orders and Gaussian kernels. As will be seen in the results, there is a great difference in precision, so for the conclusions, only the best of them was taken into account. As mentioned before, all of them are based on the SVM method, which seeks to reduce the structural risk. The aim is to achieve a clear separation between the two types of data, so that when a new sample arrives, the predictor will be able to decide or assign it to one of the two classes.

After that, the measures to evaluate the results were the Pearson coefficient and the RMSE (Root Mean Square Error).

Chapter 4: Implementation

Preprocessing of the database

One of the most critical stages of the system is the preprocessing of all the voice signals that comprise the database. The system requires a signal at the input to be analyzed and to be able to obtain an output from which we can obtain the intelligibility.

In order to determine the most appropriate pre-processing, we first analyzed the content and quality of the database. One way to analyze these audios was to graphically represent some of them. With this we were able to establish a pattern and realize that all of them had sections of silence (or background noise) of variable length in which there was no presence of voice. It was concluded that these silent or noise sections could interfere with the tests by "falsifying" the results and, therefore, it was decided to apply a Voice Activity Detector (VAD) on all voice files to fix the problem.

Feature extraction and selection

After having preprocessed the database so that we have the program input according to our needs, we move to the second block, where we need to extract the features that will be useful for our purpose. Some of the Matlab functions that were used were: ‘fxrapt’, ‘lpcresidual’, ‘melceps’ or ‘SRMR’’.

It was necessary to establish a criterion that allows us to select the best among the set of features initially calculated. The criterion used was the Pearson correlation. Therefore, this correlation between each of individual features and the real intelligibility scores was calculated and the characteristics that have the highest absolute value of correlation index were selected.

Regressors analysis and evaluation

In the first place the process of training of said regressors was carried out. This was done with specific voice signals from the database, contained in the training set. Once the "train" signals have been used with the predictor, we will have it ready to introduce new signals and predict intelligibility. To do this, we have the voice signals called "test". These are introduced into the system, from which we will obtain the results that will allow us to evaluate the corrector performance of the predictor, and therefore, see if it meets our objective.

In order to perform this evaluation, the parameters that were calculated, to evaluate the process, were the root mean square error and the Pearson correlation. With them, relevant conclusions could be obtained.

Chapter 5: Testing

Basic experiment and comparative testing

This point covers the tests that have been elaborated starting from some test or experiment already performed, and its main goal is to improve the existing results.

We wanted to do an experiment similar to the one that was done in paper "Automated Dysarthria Severity Classification for Improved Objective Intelligent Assessment of Spastic Dysarthric Speech" where a system was proposed to predict the degree of speech intelligibility of patients with spastic dysarthria. To do this, they used a vector of six acoustic characteristics, which correspond to a subset of the 21 that were previously extracted in our system.

The data that were analyzed in this occasion were 10 speakers belonging to the database "Universal Access". They used different regression types (linear, quadratic and with the Mahalanobis distance), and the system was evaluated based on the root mean square error and the Pearson correlation.

The best results obtained by both parties, come from the Mahalanobis distance, in the case of the paper, and with the Gaussian regressor in our case. The best of them turns out to be the Gaussian case with a 0.8621 as Pearson's correlation versus 0.835, although there is a considerable difference in the database, that is why, despite having achieved better results, it is not possible to make a successful comparison between them.

Testing increasing the number of features selected

The features used in the baseline experiment- were the same as those indicated in the paper. For the following experiments, it was necessary to establish another criterion for the choice of the features of the following tests. The criterion that was established to choose the features of the tests, was the individual correlations.

With the purpose of using the same dimensionality as in the reference paper, we used the six features with the best Pearson correlation. These six were: the mean and standard deviation of $\Delta c0$, the mean, deviation and range of the fundamental frequency (f0), and finally the LPC residual symmetry. The best regressor is the Gaussian, and the experiment where the criterion of selection of features is individual correlations, presents better results than the features of the baseline experiment.

Following the logical order, the tests were developed increasing one by one the features for the input. This means, input from 6 to 21 features in different tests.

Analyzing the evidence globally, we have seen the most favorable situation results in the experiment with 15 features according to RMSE, and with 21 features according to Pearson.

Therefore, and as a final conclusion, it can be said that the most appropriate result, assessing the computational load and the numerical results, is the case of introducing the 15 best features according to the individual correlation and using a SVM-based regression with Gaussian kernel..

Chapter 6: Planning and budget

The total project was developed in approximately, 18 weeks, starting on May 2017 and the final summit was in September 2017.

As for the budget, taking into account all the resources employed, amounted to the sum of ten thousand one hundred and ninety euros.

Chapter 7: Conclusions and future works

Experiments on speech from people with dysarthria caused by cerebral palsy have shown that that speech intelligibility can be predicted with a high correlation. Therefore, given the success obtained in the application of the prediction intelligibility system to these speech signals, it can be concluded that it is very likely that the method could be applied equally to voice signals from people with dysarthria caused by the Parkinson disease.

On the other hand, in the experimentation, different regressors based on SVM (Support Vector Machine) with: the linear, polynomial of different order and the Gaussian kernels have been applied. In all the tests carried out, the Gaussian kernel superiority is clearly shown with respect to the rest, having drawn all the conclusions based on this option.

Successful results have been obtained both in the automation of the process and in the results, since it is shown that it is possible to predict the degree of intelligibility of a patient with dysarthria with a high degree of correlation.

As an example of future works, it would be interesting to be able to create a speech database with dysarthria in different languages, and to perform the same tests that have so far been done with English-speaking voice signals.


UNIVERSIDAD • CARLOS III • DE MADRID
homo
homini
sacra
res