Universidad Central "Marta Abreu" de Las Villas

Facultad de Ingeniería Eléctrica

Departamento de Biomédica

TRABAJO DE DIPLOMA

Estudio de Técnicas de Tratamiento Digital de la Voz para la Identificación de Deterioro Cognitivo Leve

Autor: Alexander León Macías

Tutor: Dr.C. Eduardo González Moreira

Santa Clara

2014

"Año 56 del triunfo de la Revolución"

Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Departamento de Biomédica

TRABAJO DE DIPLOMA

Estudio de Técnicas de Tratamiento Digital de la Voz para la Identificación de Deterioro Cognitivo Leve

Autor: Alexander Leon

almacias@uclv.edu.cu

Tutor: Dr.C. Eduardo Gonzalez Moreira

moreira@uclv.edu.cu

Santa Clara

2014

"Año 56 del triunfo de la Revolución"

Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central "Marta Abreu" de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería en Informática, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

_____

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

_____ _____

Firma del Tutor

Firma del Jefe de Departamento donde se defiende el trabajo

_____

Firma del Responsable de Información Científico-Técnica

PENSAMIENTO

Escriba aquí el texto del pensamiento (Opcional)

DEDICATORIA

Escriba aquí el texto de la dedicatoria (Opcional)

AGRADECIMIENTOS

Escriba aquí el texto de los agradecimientos (Opcional)

TAREA TÉCNICA

- Realización de un análisis crítico del estado del arte acerca de estudios realizados sobre mediciones de deterioro cognitivo leve mediante el procesamiento digital de voz.
- Diseño del protocolo de grabación para las mediciones de deterioro cognitivo leve.
- Obtención de rasgos prosódicos para la clasificación de deterioro cognitivo leve a través del procesado del habla.
- Evaluación de los resultados obtenidos para la identificación de deterioro cognitivo leve y algoritmos clasificadores empelados en la clasificación binaria.
- Redacción del informe.

Firma del Autor

Firma del Tutor

RESUMEN

El envejecimiento poblacional es hoy en día una tendencia innegable en nuestro país. En este contexto socio-demográfico, se constata que en las consultas de neurología el deterioro cognitivo constituye el segundo motivo de consulta más frecuente. Ante esta situación, se hace necesaria la detección precoz de este tipo de problemas, especialmente mediante pruebas de cribado, con el objetivo de intentar actuar a tiempo para prevenir la evolución hacia la cronicidad en este tipo de patologías. La discriminación entre ancianos sanos y ancianos con deterioro cognitivo se ha hecho en muchas ocasiones en función del rendimiento obtenido por los ancianos en determinadas pruebas cognitivas específicas, entre ellas, el Minimental State Examination (MMSE). Estas técnicas son consumen tiempo y requieren la presencia de un especialista para realizar la evaluación.

El presente trabajo se ha centrado en el estudio de técnicas que permitan la detección temprana del deterioro cognitivo leve (DCL) por medio de la extracción y el análisis de parámetros asociados a la prosodia de la voz. Para ello, se ha dispuesto de una base de datos consistente de dos grupos diferenciados: uno con sujetos diagnosticados con DCL y otro con sujetos sanos. El objetivo de este documento es describir los pasos, estudios y análisis realizados sobre esta base de datos mediante el empleo de parámetros prosódicos, así como las conclusiones extraídas a partir de los resultados.

Para la confección del contenido de este trabajo nos hemos enfocado en estudios de técnicas que nos puedan permitir una posible detección temprana del DCL mediante la extracción de los parámetros prosódicos de la voz y el análisis de los mismos. Para lograr la confección de nuestro trabajo, se ha confeccionado en el transcurso de la misma una base de datos de dos grupos de personas, personas sanas y personas con DCL.

Por último, se estudió mediante métodos estadísticos la importancia de los parámetros prosódicos extraídos con anterioridad a la hora de emplearse como discriminativos entre sujetos sanos y enfermos, y se implementaron varios clasificadores para comprobar la eficiencia discriminativa de los rasgos.

TABLA DE CONTENIDOS

PENSAMIENTO i

DEDICATORIA ii

AGRADECIMIENTOS iii

TAREA TÉCNICA iv

RESUMEN v

INTRODUCCIÓN 1

Organización del informe 2

CAPÍTULO 1. DCL y su relación con el habla. 4

1.1 Definición del DCL. 4

1.2 Criterios y pruebas para el diagnóstico de DCL. 5

1.3 Ventajas del diagnóstico del DCL mediante el habla. 9

1.4 Análisis del habla. 10

1.5 Estudios relacionados con la detección del DCL. 11

1.6 Conclusiones parciales. 16

CAPÍTULO 2. Extracción de Datos Prosódicos de la Base de Datos de DCL. 17

2.1 Análisis Prosódico de la Voz. 17

2.2 Desarrollo de una base de datos de DCL. 18

2.3 Orden en el proceso de grabación. 18

2.4 Corpus de voz. 22

2.5 Software DCGrab v3.0 25

2.6 Herramientas: Praat, Prosogram y Matlab. 26

2.7 Tipos de Segmentación. 29

2.8 Base de Datos de DCL. 30

2.9 Obtención de Parámetros Prosódicos. 31
CAPÍTULO 3. RESULTADOS Y DISCUSIÓN 33
3.1 Descripción estadística de los rasgos extraídos. 33
3.2 Obtención de rasgos significativos 35
3.2.1 Prueba de Wilcoxon 36
3.2.2 Obtención mediante trazado de curvas ROC 37
3.3 Comparativa de resultados 42
3.4 Clasificación de los pacientes 43
3.5 Clasificador LDA 44
3.5.1 Clasificador LDA con rasgos significativos 45
3.5.2 Clasificador LDA con todos los rasgos 46
3.6 Clasificador RBF. 47
3.6.1 Clasificador RBF con los rasgos significativos. 49
3.6.2 Clasificador RBF con todos los rasgos. 50
3.7 Conclusiones del capítulo 51
CONCLUSIONES Y RECOMENDACIONES 52
Conclusiones 52
Recomendaciones 52
REFERENCIAS BIBLIOGRÁFICAS 53

INTRODUCCIÓN

El envejecimiento poblacional es hoy en día una tendencia innegable [1]. En este contexto sociodemográfico, en las consultas para el deterioro cognitivo constituye un motivo de consulta frecuente [2]. Ante esta situación, se hace necesaria la detección precoz de este tipo de problemas, especialmente mediante pruebas, con el objetivo de intentar actuar a tiempo para prevenir la evolución hacia una actividad crónica en este tipo de patologías.

En Latinoamérica se observan prevalencias que van desde el 1,8 hasta el 11,5% [3-6]. En un estudio multicéntrico desarrollado en Uruguay, Chile, Brasil, Venezuela, Cuba y Perú, se encontró una prevalencia del 7,1% en personas mayores de 65 años, siendo la más frecuente entre los deterioros cognitivos la demencia tipo Alzheimer [7]

Una posible confusión entre ancianos sanos y ancianos con deterioro cognitivo se ha hecho en muchas ocasiones en función del rendimiento obtenido de ellos en determinadas pruebas cognitivas específicas [8], entre ellas, el Minimental State Exami nation (MMSE) [9]. A pesar de las dificultades que entraña el proceso de cribado en el deterioro cognitivo leve [10], se sabe que algunas pruebas de cribado son útiles para la identificación de los ancianos en riesgo de padecer deterioro cognitivo en la vejez.

A pesar de la gran cantidad de pruebas al uso, el problema es que gran parte de las pruebas de cribado del deterioro cognitivo actualmente en uso son pruebas realizadas en otros países cuyos ancianos poseen mayor nivel educativo, basadas en gran parte en tareas de memoria, fluidez verbal y orientación espacial, por lo que requieren cierto nivel educativo y determinadas habilidades gráficas. Esto hace que en el momento actual, en nuestro país, se busquen estrategias evaluativas que, además de cumplir requisitos de brevedad, aceptabilidad, facilidad y validez , sean simples y libres, en la medida de lo posible, de

cualquier influencia cultural, de manera que se minimice el riesgo de utilizar instrumentos de evaluación poco sensibles para ancianos con un nivel educativo alto o con un nivel educativo muy bajo, lo que redundaría, en última instancia, en un cribado del deterioro cognitivo leve impreciso (alta tasa de falsos negativos y alta tasa de falsos positivos)

Con el fin de lograr dar solución al problema existente se plantean los siguientes objetivos de investigación:

Objetivo General:

- Desarrollar un análisis computacional de la voz para identificar DCL.

Objetivos Particulares:

- Identificar los rasgos característicos presentes en la voz de los pacientes con DCL descritos en la literatura.
- Realizar grabaciones con el sistema de medición a pacientes que presenten afectación de DCL.
- Evaluar la capacidad de los rasgos prosódicos para la identificación de DCL mediante técnicas de clasificación.

Organización del informe

El informe del trabajo está estructurado de la siguiente manera: introducción, capitulario, conclusiones, recomendaciones, referencias bibliográficas. En la introducción se dejará definida la importancia, actualidad y necesidad del tema que se aborda y se dejarán explícitos los elementos del diseño teórico, además se plantea la necesidad de la investigación, antecedentes y las características de las mediciones realizadas. En el Capítulo I se realiza la fundamentación teórica sobre el DCL. Se analizan diferentes estudios realizados en el mundo con respecto al tema. En el Capítulo II se analizan las características del sistema utilizado para las mediciones y se caracteriza el proceso de obtención de las muestras para la realización del estudio. El Capítulo III se dedicará a analizar los resultados obtenidos a partir de los rasgos prosódicos y sus análisis estadísticos, teniendo como objetivo final identificar DCL mediante el habla para su posterior uso en las

consultas de neurología. A modo de conclusiones se exponen los principales resultados que deben dar respuesta a los objetivos planteados a inicio del trabajo. Las recomendaciones indicarán las mejoras palpables para futuros trabajos en el tema. Por último la bibliografía muestra todo el material consultado para la elaboración del trabajo.

CAPÍTULO 1. DCL y su relación con el habla.

En este capítulo se realiza la fundamentación teórica sobre el DCL, así como conceptos básicos de esta patología para la realización de este trabajo. También, se le ofrece un recorrido por las diferentes pruebas que existen para la Detección del DCL. Además se pretende puntualizar en la ventaja que presenta el diagnóstico del mismo y las áreas con las que vamos a trabajar.

1.1 Definición del DCL.

Ya desde mediados del siglo pasado, Kral propuso el concepto clínico de olvido benigno del anciano como una alteración aislada de la memoria; posterior-mente otros autores han elaborado distintas nomenclaturas para un concepto casi nosológico, tales como deterioro de memoria asociado a la edad. Por fin, a finales de los noventa surge el concepto de deterioro cognitivo leve definido por Petersen [11], esbozando las diversas subclasificaciones y es actualmente el más usado. En estas denominaciones están incluidos aquellos que no presentan criterios clínicos de ninguna entidad neurodegenerativa, pero que presentan un rendimiento anormal en las baterías neuropsicológicas. Las distintas definiciones tienen muchos puntos en común: alteración subjetiva de la memoria, su confirmación objetiva en baterías neuropsicológicas generales y preservación de las actividades habituales. De esas definiciones y una de las más usadas es la que a continuación se muestra.

Deterioro cognitivo es la disminución mantenida del rendimiento de las funciones intelectuales o cognitivas desde un nivel previo más elevado. Las causas son múltiples: demencia, depresión, alteración por la edad, accidentes o procesos cerebro-vasculares,

alcoholismo, daño cerebral traumático, cirugía del cerebro, etc. El deterioro cognoscitivo o cognitivo leve (DCL) es conceptualizado como una frontera entre el envejecimiento normal y la demencia, agrupa a pacientes con un declinar cognoscitivo cuya cuantía sobrepasa el envejecimiento normal

Actualmente hay subtipos de DCL que son bien reconocidos [12].

- **DCL del tipo Amnésico**: Son individuos con alteraciones importantes de memoria, que no llenan los criterios para demencia. Se considera un estado precursor de Enfermedad de Alzheimer (EA). En ocasiones al realizar una valoración neuropsicológica cuidadosa se puede encontrar que existen otras áreas afectadas además de la memoria
- **DCL de áreas o dominios múltiples**: Los individuos presentan mínimas alteraciones en las diferentes áreas de funciones cognoscitivas y de conducta. Pueden tener pequeños dificultades con la realización de las actividades de vida diarias pero no cumplen con los criterios para demencia. Usualmente estos individuos pocas veces progresan a EA o demencia Vascular. Es una condición poco estable y hasta puede observarse a individuos regresar a su estado funcional previo.
- **DCL de un solo dominio o área no amnésica**: Trastorno aislado de alguna función ejecutiva., lenguaje o habilidad viso especial. Dependiendo del área afectada se puede progresar a Demencia Frontotemporal, Afasia Primaria Progresiva o Demencia de Cuerpos de Lewy.

1.2 Criterios y pruebas para el diagnóstico de DCL.

Criterios:

- Quejas de pérdida de memoria, preferiblemente corroborado por otra persona (familiar o cuidador).
- Alteración de la memoria con respecto a lo que correspondería para la edad y el nivel educativo.
- Función cognitiva general conservada.

- Actividades de la vida diaria intactas.
- No cumple criterios para diagnóstico de demencia.
- Evidencia objetiva de defectos de memoria inferior a 1,5 desviaciones estándar en relación a la media establecida para su edad.

Diagnóstico diferencial:

Los principales procesos con los que se debe hacer el diagnóstico diferencial son:

- Delírium: el curso clínico suele ser agudo-subagudo, con fluctuaciones importantes en el estado mental, y alteraciones en el grado de atención y en el nivel de conciencia. Es importante su detección porque indica la existencia de enfermedades médicas subyacentes con eleva-da mortalidad si no se tratan a tiempo.
- Depresión: influye negativamente sobre las funciones cognitivas, y puede confundirse con una demencia, aunque también pueden coexistir en el mismo paciente. Los pacientes con depresión suelen tener más quejas subjetivas de pérdida de memoria, con frecuencia tienen enlentecimiento psicomotor y escasa motivación en la realización de los test.

Pruebas para el diagnóstico de DCL:

La valoración de las funciones cognitivas es fundamental en el diagnóstico de deterioro cognitivo. Debe incluir examen de la atención, orientación, lenguaje, memoria, habilidades constructivas, cálculo, razonamiento y capacidad de abstracción. La evaluación del diagnóstico del deterioro cognitivo debe fundamentar en una historia clínica detallada. Es detectado mediante pruebas neuropsicologías si realmente hay deterioro cognitivo. La evaluación se inicia con pruebas de screening o cribaje que se suelen aplicar en consultas ambulatorias de atención primaria o en una primera fase de exploración por un especialista (neurólogo, psiquiatra, neuropsicólogo).

En el mundo son usadas diferentes tipos de pruebas como son:

1- Pruebas de rendimiento cognitivo global: La más utilizada es el Examen de Estado Mental (MMSE, del inglés Mini Mental State Examination). Su rango es 0-30 puntos,

su punto de corte para probable deterioro cognitivo es de 23/24 .Valora orientación espacial, temporal, memoria inmediata, recuerdo diferido, atención, praxias, etc. El test MMSE fue desarrollado en 1975 por M. Folstein como una herramienta para evaluar el estado cognitivo de las personas mayores. Probablemente es una de las herramientas más establecidas para evaluar el estado cognitivo, aunque hay que destacar que no es una herramienta de diagnóstico de demencia. Evalúa orientación, atención, denominación, lenguaje y la capacidad para seguir órdenes verbales y escritas. Tiene numerosas ventajas, una de las principales es que puede ser administrado por médicos, enfermeros o personal desentrenado obteniendo la misma calidad de ejecución, sin importar quien realice el test. Además solo lleva 5 o 10 minutos realizar el test, y suele ser muy bien aceptado por los pacientes a los que se le realiza. Algunas de sus desventajas es que las personas con un nivel de educación inferior obtienen menor puntuación que las de un mayor nivel, y la repercusión de algunos déficits en la visión o en la capacidad auditiva en la puntuación [13].

2- Auditory Verbal Learning Test-potencial de aprendizaje (AVLT-PA) [14]. Adaptado al español por Calero et al, consiste en la presentación de una lista de palabras en seis veces consecutivas, debiendo recordar el mayor número de ellas.

3- FVS [15]. Tarea simple de recuerdo de palabras de una categoría dada que consiste en pedirle a la persona que diga tantas palabras como pueda de la categoría ‘animales’. El evaluador anota las palabras válidas citadas por el evaluado durante un minuto.

4- Tarea de evaluación de la memoria de trabajo [16]. Mide la amplitud de la memoria de trabajo a partir de la presentación de tarjetas con tres números cada una, de manera que la persona debe leer en voz alta los números de cada tarjeta y recordar el último número, ya que luego tendrá que reproducirlos una vez que se le hayan presentado las tarjetas. El número de tarjetas se va incrementando en función del rendimiento de la persona, de tal manera que la tarea empieza con dos tarjetas y acaba con cinco, siempre que la persona supere las fases anteriores.

5- Tarea de atención sostenida (TAS) [17]. Tarea informatizada, desarrollada a partir del test de ejecución continua de Conners, que consiste básicamente en una tarea de vigilancia diseñada para evaluar la atención sostenida en la que aparecen estímulos

numéricos del 1 al 9 que son presentados aleatoriamente en intervalos de 1.000 ms. Los números son presentados en blanco sobre un fondo negro. Al individuo se le pide que presione la barra espaciadora del ordenador inmediatamente después de observar la secuencia 3-6. Hay un bloque inicial de entrenamiento y cuatro bloques correspondientes al test, cada uno de los cuales dura un minuto. En este intervalo, la secuencia 3-6 aparece aleatoriamente en 15 ocasiones sobre 60 estímulos posibles. Los 45 estímulos restantes son considerados distractores. Al finalizar la tarea, un total de 60 secuencias correctas habrán aparecido sobre un total de 240 estímulos disponibles. La variable dependiente en esta tarea fue el número total de aciertos por participante.

6- El Test 7Minutos: Se diseñó para la detección de la demencia .Es un test que permite obtener una puntuación global a partir de varios subtest. La orientación temporal de Benton que incluye las preguntas habituales de orientación más la hora actual, la fluidez semántica (nombres de animales en un minuto) ,el Test del Reloj (dibujar un reloj a la orden que explora organización visuoespacial y funciones ejecutivas) y el subtest de memoria visual (presentación de 16 objetos que explora de nominación y recuerdo inmediato).

7- El Short Portable Mental Status Questionnaire (SPMSQ) es una opción muy adecuada para personas de bajo nivel educativo. Evalúa la orientación, la memoria, el cálculo y la atención. El punto de corte para deterioro cognitivo son 3 errores.

8- El Memory Impairment Screen de Buschke (B-MIS) y el Test de las fotos que valoran la memoria mediante el aprendizaje incidental de 4 palabras y 4 fotos respectivamente.

9- El test ‘Wechsler Logical Memory’ (WLM), muy utilizado en algunas de las líneas de trabajo citadas posteriormente, es una prueba neuropsicológica diseñada para medir diferentes funciones de la memoria de un individuo. Se emplea para personas con edades comprendidas entre los 16 y los 90 años y su última versión es el WLM-IV. La puntuación final de este test se obtiene a través de cinco secciones: memoria auditiva, memoria visual, memoria de trabajo visual, memoria inmediata y memoria retardada.

10- El test Barcelona conlleva alrededor de 3 horas, y el test Barcelona abreviado, unos 45 minutos aproximadamente. Evalúa la orientación, lenguaje, lectura, escritura,

reconocimiento visual, memoria y abstracción en un total de 55 ítems y una puntuación máxima de 110 puntos.

Los resultados de los test deben interpretarse en conjunción con los datos de la historia. Si ésta sugiere deterioro cognitivo pero la exploración neuropsicológica está por encima del punto de corte, puede deberse a un nivel educativo alto, depresión o deterioro cognitivo leve. Una puntuación baja sin datos en la historia de deterioro puede indicar un bajo nivel cultural, pero también un escaso reconocimiento de los síntomas por parte del paciente o la familia. Los pacientes con resultados por debajo del punto de corte requieren una evaluación neuropsicológica detallada, que proporcione información sobre el perfil neuropsicológico, lo que puede contribuir al diagnóstico etiológico.

1.3 Ventajas del diagnóstico del DCL mediante el habla.

El diagnóstico del DCL permite que el evaluador sea detallado, rápido, preciso y coherente en la descripción de las características del habla del individuo. Revela información sobre el mecanismos subyacente que causa la desviación del habla. El diagnóstico del DCL se ha convertido en una herramienta estándar de la clasificación de patologías del habla y ha abierto muchos campos de investigación basado en la descripción de la voz.

En el mundo las áreas comúnmente evaluadas son: memoria, atención, lenguaje, funciones ejecutivas, gnosias, praxias, habilidades visuo-espaciales y velocidad de procesamiento que pueden alterarse en un DCL de forma aislada o en un DCL multi-dominio. En el caso de este trabajo nos enfocaremos en la memoria, la atención y primordialmente en el lenguaje.

El lenguaje es una de las áreas más deterioradas en el DCL. La denominación de objetos es la función lingüística que se altera de forma más frecuente, generalmente en el contexto de un deterioro cognitivo amnésico multi-dominio. También podemos encontrar disminuciones en la fluidez tanto semántica como fonológica, aunque esta alteración está más ligada a un problema del control ejecutivo. En un DCL raramente encontraremos trastornos importantes de la comprensión de órdenes o de la repetición. Por ello, para la correcta evaluación del lenguaje es imprescindible la evaluación de la denominación y la fluidez verbal, sobre todo.

1.4 Análisis del habla.

Una herramienta muy útil para la detección del DCL u otros tipos de demencia son los test relacionados con el lenguaje. Este tipo de prueba es capaz de realizar un análisis del discurso y obtener información que permita detectar deterioro cognitivo o posibles demencias. Los test indicados para la detección del DCL o demencias pueden dividirse en dos tipos:

- Lingüísticos
- Prosódicos

Estos dos tipos de test se diferencian en el tipo de parámetros que utilizan para la evaluación, aunque cabe destacar que en numerosas ocasiones se evalúan ambos aspectos.

El elemento más sensible para la evaluación de las habilidades lingüísticas, según se describe en los artículos, es la entrevista al paciente. El discurso espontáneo del paciente es el mejor elemento para evaluar un posible deterioro del lenguaje. Para conseguir que el paciente hable sobre una situación y obtener un discurso fluido, los métodos más empleados son la entrevista al paciente, la narración de historias y la descripción fotográfica. Se muestra al paciente una fotografía y éste debe tratar de describir qué sucede, qué le sugiere o simplemente qué es lo que observa. En el caso Estudio de la Aplicación de Técnicas de Tratamiento Digital de la Voz para la Detección de Patologías de la narración, el paciente lee un fragmento de texto que cuenta una historia, y tras ello, debe volver a contar la historia con sus propias palabras. Del discurso realizado por el paciente se evalúan, dependiendo del estudio, un determinado número de palabras, por ejemplo, sus primeras cien palabras. De estas palabras cabe destacar en la evaluación, debido a sus apariciones reiteradas, las siguientes características léxicas:

- Número de nombres
- Número de pronombres
- Número de adjetivos
- Número de verbos

El discurso del paciente, para la evaluación lingüística, debe ser transcrito tal y como va sucediendo, es decir, deben incluirse repeticiones, palabras incompletas o parafasias. El objetivo es transcribir exactamente el discurso del paciente para poder proceder a su evaluación y su procesamiento mediante diferentes algoritmos y, finalmente, realizar la interpretación de los resultados.

El procesamiento se realiza, en la mayoría de los casos, de forma manual, dando paso a un posible análisis automático, y la interpretación de los resultados la realiza el especialista. Esto ocurre tanto para el análisis lingüístico como para el análisis prosódico.

Por otro lado la prosodia es una rama de la lingüística que analiza y representa formalmente aquellos elementos de la expresión oral, tales como el acento, los tonos y la entonación. Existen también otros parámetros que evalúan los aspectos prosódicos del lenguaje como son los siguientes:

- Número de pausas
- Duración media de pausas
- Duración media de la expresión
- Duración total de pausas
- Duración total de expresiones

1.5 Estudios relacionados con la detección del DCL.

En la actualidad la tendencia en el procesamiento de voces patológicas es el de utilizar sistemas donde se combinen valoraciones objetivas y subjetivas para apreciar la severidad de los rasgos de la voz, en dependencia de su precisión. Existen varios trabajos reportados en la literatura para mejorar la clasificación de las disartrias.

Uno de estos trabajos reportados ‘’Predicción del deterioro cognitivo en ancianos mediante el análisis del rendimiento en fluidez verbal y en atención sostenida’’ en el cual se hizo una investigación en la cual participaron un total de 264 ancianos con una edad media de 78,08 ± 7,97 años (rango: 65-96 años), de los cuales el 38,5% eran hombres y el 61,5%, mujeres. El 30,4% eran amas de casa; el 58,2% habían sido obreros sin cualificar; el 9,1%,

profesional de grado medio, y el 2,3%, profesionales de grado superior. Por su parte, el 29,3% residía en su propio domicilio, mientras que el 70,7% vivía en varias residencias de ancianos, en todos los casos, en la zona geográfica de Andalucía oriental. Los participantes en este estudio proceden de residencias de ancianos, centros de día y una universidad de mayores de la zona de Andalucía oriental. Se seleccionaron ancianos que cumplieran los siguientes requisitos: no padecimiento de enfermedades relevantes o limitaciones físicas o psíquicas, y ausencia de déficits motores y sensoriales.

En el artículo ‘‘Computerized Analysis of Speech and Language to Identify Psycholinguistic Correlates of Fronto-Temporal Lobar Degeneration’’ [18] tuvo como como objetivo la descripción de un sistema semiautomático para la evaluación del discurso y las características del lenguaje en pacientes con degeneración lobular frontotemporal (FTLD siglas en ingles). Este es el nombre que recibe un conjunto de desórdenes clínica, patológica y genéticamente heterogéneos asociados a una atrofia de los lóbulos frontal y temporal del cerebro, y más moderadamente, de los lóbulos parietal y occipital en el cual se graban muestras de voz de 38 pacientes diagnosticados con alguno de los tres síndromes que componen el FTLD, de tres centros médicos participantes en el estudio. Previamente, cada paciente se sometió a una batería de pruebas neuropsicológicas. Las muestras de voz son la respuesta al estímulo de la conocida lámina del “ladrón de galletas”, una lámina que representa diferentes acciones y que se muestra a continuación:

Fig. 1.1 Ladrón de galletas.

El objetivo es la descripción de la imagen por parte del paciente. El paciente debe contar qué es lo que observa en dicha imagen, y esa muestra de habla es la que posteriormente será analizada.

El audio correspondiente a cada paciente es analizado por el sistema para obtener un total de 15 variables del lenguaje, tanto lingüísticas como prosódicas:

1. Ratio palabra-pausa.
2. Varianza de la frecuencia fundamental.
3. Perplejidad de parte del discurso.
4. Perplejidad del nivel de palabras
5. Ratio pronombres-sustantivos.
6. Número de sustantivos.
7. Duración total del discurso.
8. Media de la duración de la frase prosódica.
9. Unidades de información correcta.
10. Número de pausas largas normalizadas (silencios >400 ms de duración).
11. Número de pausas completas normalizadas.
12. Número de silencios normalizados (silencios >150 ms de duración).
13. Número de falsos inicios normalizados.
14. Número de repeticiones normalizadas.
15. Número de disfluencias normalizadas (pausas muy largas, falsos comienzos y repeticiones).

Las mediciones de las características del habla y el lenguaje utilizados por este enfoque semiautomático se basan en la alineación entre el texto de la descripción de imagen transcrito y el audio de la señal representada. NLP y ASR son un conjunto de técnicas empleadas para el análisis del lenguaje y del discurso. La aplicación de estas técnicas para pruebas psicométricas es relativamente novedosa, pero corresponde a una extensión de las

capacidades que ofrecen estas tecnologías. El sistema desarrollado por este grupo para el análisis semiautomático del habla y del discurso se describe en la figura que aparece a continuación:

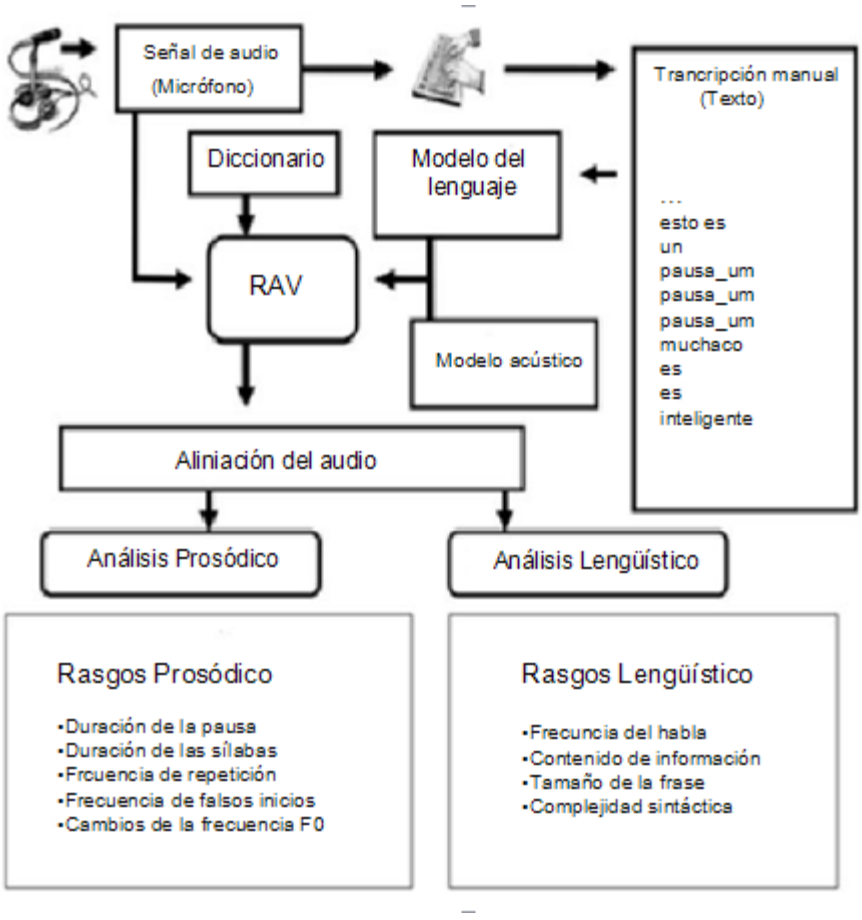

Fig. 1.2 Diagrama de flujo del sistema de automatización para la evaluación del habla y el lenguaje.

En cuanto a mediciones semiautomáticas de la relación de pausa a palabra y la relación pronombre-sustantivo fueron capaces de discriminar entre algunas de las variantes del FTLD. El análisis de las componentes principales de las variables sugeridas da como resultado 4 variables significativas: duración del discurso, dificultad, contenido vacío y gramaticalidad, correspondientes a la fenomenología de variantes del FTLD.

En el artículo ‘‘Automatically Derived Spoken Languages Markers for Detecting Mild Cognitive Impairment’’ [19] se realizó un estudio que en cual se plasma la viabilidad para discriminar entre grupos de pacientes que poseen algunos marcadores que se encuentran en el lenguaje, observados al realizar a los pacientes test neuropsicológicos. Del discurso de

los pacientes tras la realización de estos test, se obtienen automáticamente, a partir de audio transcrito, determinadores marcadores que se emplean para distinguir entre dos grupos: individuos sanos frente a individuos con deterioro cognitivo leve (DCL). Se comprueba la efectividad de la obtención de forma automática de los marcadores frente a la obtención de dichos marcadores de forma manual.

Para este estudio se recogen muestras de audio de 44 exámenes neuropsicológicos, que se distinguen por pertenecer a dos grupos: adultos sanos y adultos que padecen DCL. De las personas incluidas en el estudio se toman diferentes medidas como la edad o el nivel educativo. Además, se pasan diferentes test como el MMSE o el Wechsler Logical Memory.

En el estudio se aplica un tipo de test y se obtiene el siguiente resultado: en la mayoría de los casos no hay apenas diferencia entre los valores extraídos de forma manual y los extraídos automáticamente. Es decir, los marcadores son extraídos de forma automática con alta precisión, y esto supone encontrar diferencias entre los diferentes grupos de individuos sometidos al estudio.

En el artículo ‘‘Syntactic Complexity Measures for Detecting Mild Cognitive Impairment’’ [20] en el cual se consideran útiles algunas medidas sintácticas complejas del lenguaje para el diagnóstico del paciente, diferenciando entre muestras de voz de pacientes sanos y pacientes que sufren deterioro cognitivo. Se evalúan las diferencias que aparecen al obtener esas medidas de forma manual, construyendo árboles de análisis, y de forma automática. Se presentan métodos para la medida automática de la complejidad sintáctica del lenguaje a través de muestras del discurso de pacientes de edad avanzada, sometidos a examen neuropsicológico, y se trata de buscar rasgos diferenciadores entre el grupo de control y el grupo que padece deterioro cognitivo, a través de medidas del lenguaje. Establece como pruebas más fiables para la evaluación la lista de palabras para memorizar o el recuerdo narrativo (Wechsler Logical Memory I y Wechsler Logical Memory II), considerándolas más fiables que el MMSE en la detección de deterioro cognitivo, aunque este último test también lo pasan los pacientes. Se centra principalmente en la evaluación de medidas de complejidad sintáctica.

Establecen dos claves a considerar en la evaluación de medidas de complejidad sintáctica, en relación con la evaluación psicométrica: alta utilidad discriminativa, debido a que el objetivo es diferenciar ambos grupos con estas medidas, y fidelidad de la medida al derivar las transcripciones del análisis automático.

Para ello se recogen muestras de 55 test neuropsicológicos. Tras pasar estos test neuropsicológicos, se realiza la prueba de recuerdo narrativo. El texto tiene únicamente 3 frases, que el paciente debe leer y volver a contar posteriormente. Se transcribe la respuesta, anotando manualmente los árboles de análisis. Los algoritmos para la extracción automática de marcadores de complejidad sintáctica de árboles de análisis procesan los árboles anotados manualmente, para la obtención de las medidas automáticas.

Se obtuvo como resultados una alta correlación entre los marcadores tomados manualmente y los obtenidos de forma automática. Esto implica alta fiabilidad de los analizadores automáticos para la obtención de estas medidas. Las técnicas de procesado del lenguaje natural pueden emplearse para detectar de forma automática marcadores útiles para discriminar entre sujetos sanos y pacientes con deterioro cognitivo.

1.6 Conclusiones parciales.

El Deterioro Cognitivo Leve tiene su nivel de complejidad de acuerdo a las características fundamentales que lo se forman. En la literatura se ha reportado varios métodos para el análisis de las voces relacionado con el diagnóstico de deterioro cognitivo, uno de estos métodos es el análisis instrumental y el análisis perceptual del habla.

Hoy día se trabaja con métodos donde se combina el análisis perceptual y objetivo como consecuencia de las restricciones que presentan los sistemas reportados. Esta nueva combinación debe ser más eficiente mediante el uso de métodos de computación.

Adicionalmente hay estudios que sugieren que los métodos de inteligencia artificial podrían ofrecer buenos resultados relacionados con una predicción de diagnóstico mucho más precisa.

CAPÍTULO 2. Extracción de Datos Prosódicos de la Base de Datos de DCL.

En este Capítulo tenemos como objetivo intentar detectar automáticamente las evidencias en el habla de los locutores con DCL, describiremos las técnicas empleadas sobre la base de datos de DCL para la extracción de datos prosódicos.

2.1 Análisis Prosódico de la Voz.

Esta es una rama de la lingüística que principalmente se basa en el aprendizaje de los rasgos fónicos que afectan directamente a la expresión oral, como por ejemplo, la tonalidad, la cadencia o la acentuación. Cada individuo posee rasgos característicos de si y, por tanto, estos rasgos varían de persona en persona. La edad, el nivel escolar alcanzado por la misma, el estado emocional en que se encuentran, estos son factores influyentes esas características prosódicas de la persona.

Según estudios que se han realizado han demostrado que ciertas personas que padecen de algunas enfermedades específicas, como el caso del Alzheimer, estas tienen presentes algunas diferencias significativas en ciertos rasgos melódicos y fónicos del habla con respecto a los de personas que no padecen de enfermedades, es decir personas sanas. [21] [22]

A lo largo de este Capítulo nos centraremos principalmente en el análisis prosódico de la voz y para ello presidiremos de personas diagnosticadas de Alzheimer y de personas que no la padezcan, personas sanas, con el primordial objeto de encontrar diferencias en ciertas características entre ellas las cual nos permitan notar una diferencia entre una persona sana y otra enferma solo tomando como estándar el estudio de una señal de voz.

Los resultados que obtendremos están directamente destinados a un posible diagnóstico temprano del deterioro cognitivo leve (DCL), que como dijimos en el Capítulo anterior el DCL es una fase clínica previa a la demencia, en el cual si de los síntomas el que comienza a predominar es la pérdida gradual de memoria, puede terminar en la enfermedad del Alzheimer. Si es posible una detección tempranera del DCL antes de que este pueda convertirse en demencia, y un tratamiento prematuro, se puede evitar que esta enfermedad pueda seguir su progresión y con esto poder ofrecerle al paciente una oportunidad de tener una mejor calidad de vida. [23].

2.2 Desarrollo de una base de datos de DCL.

A continuación proponemos los criterios en los cuales nos basamos para el diseño y especificación de la base de datos de pacientes con DCL y el grupo de personas sanas que fue grabada con el objeto de permitirnos la continuación de esta Tesis, la cual se relaciona con la búsqueda de rasgos prosódicos para la detección automática mediante el análisis del habla de casos de personas con DCL.

2.3 Orden en el proceso de grabación.

Para la continuación del desarrollo de esta investigación se tomo un orden de grabación la cual se muestra a continuación:

- El Análisis del Test MMSE.
- El Análisis del Test de lectura del texto del “El Abuelo”.
- El Análisis Test de descripción de la imagen “Ladrón de Galletas”.

Test de cribaje mini mental

La evaluación consiste en un breve test de cribaje. A continuación se puede observar las distintas preguntas de este test de cribaje realizado, con sus correspondientes puntuaciones para cada apartado.

Tabla 2.1 Prueba de MMSE

ORIENTACIÓN TEMPORAL			
1	¿En qué día estamos?	0 - 1	x/3
2	¿De qué mes?	0 - 1	
3	¿Y de qué año?	0 - 1	
ORIENTACIÓN ESPACIAL			
4	¿En qué hospital (o edificio) estamos?	0 - 1	x/5
5	¿En qué planta?	0 - 1	
6	¿Ciudad (o municipio)?	0 - 1	
7	¿Provincia?	0 - 1	
8	¿Y en qué país (o nación)?	0 - 1	
FIJACIÓN – RECUERDO INMEDIATO *Nombre tres palabras siguientes y pida al paciente que las repita. Esta primera repetición otorga la puntuación, 1 punto por cada palabra correcta, pero continúe diciéndolas hasta que el sujeto repita las 3, hasta 6 veces.*			
9	Mesa	0 - 1	x/3
10	Lápiz	0 - 1	
11	Árbol	0 - 1	
ATENCIÓN – CÁLCULO *Si tiene 30 y va quitando 3, ¿cuánto le van quedando? Detenga la prueba tras 5 substracciones. (Si el sujeto no puede realizar esta prueba, pídale que deletree la palabra MUNDO al revés).*			
12	27 (O)	0 - 1	x/5
13	24 (D)	0 - 1	
14	21 (N)	0 - 1	
15	18 (U)	0 - 1	
16	15 (M)	0 - 1	
RECUERDO DIFERIDO *Preguntar por las palabras mencionadas anteriormente.*			
17	Mesa	0 - 1	x/3
18	Lápiz	0 - 1	
19	Árbol	0 - 1	
DENOMINACIÓN *Mostrar los siguientes objetos y preguntar qué son.*			
20	Lápiz	0 - 1	x/3
21	Reloj	0 - 1	
22	Sortija	0 - 1	
REPETICIÓN *Pedirle que repita la siguiente frase, o si lo prefiere, la frase que está entre paréntesis.* *Puntúa si la repite correctamente*			
23	"Ni sí, ni no, ni pero" (o "En un trigal había 5 perros")	0 - 1	x/1

ASOCIACIÓN			
24	¿Qué son la manzana y la pera?	0 - 1	x/3
25	¿Y el perro y el gato?	0 - 1	
26	¿Y el rojo y el verde?	0 - 1	
ÓRDENES *Pedirle que siga la orden: "coja un papel con la mano derecha, dóblelo por la mitad, y póngalo en el suelo".*			
27	Coge el papel	0 - 1	x/3
28	dobla por la mitad	0 - 1	
29	pone en el suelo	0 - 1	
LECTURA *Escriba legiblemente en un papel "Cierre los ojos". Pídale que lo lea y haga lo que dice la frase.*			
30	Lee y ejecuta la frase correctamente	0 - 1	x/1
PUNTUACIÓN TOTAL			
Puntuaciones de referencia	27 o más: *Normalidad* 24 a menos: *Sospecha patológica* de 24 a 12 deterioro cognitivo de 11 a 9 demencia		x/30

De esta parte de la evaluación se recogerá una grabación de audio que cubrirá la totalidad de la duración del test mini mental, incluyendo las preguntas del psicólogo y las respuestas del paciente, así como toda la conversación que tenga lugar durante el test. En el archivo de audio grabado, se marcarán los tiempos en los que tienen lugar las diferentes preguntas y respuestas de manera automática.

Test de lectura del texto "El abuelo"

Tras el test mini mental, la evaluación contínua con una lectura por parte del paciente de un breve texto utilizado en estudios de disartrias. La utilización de dicha lectura para la valoración de habla y de lenguaje fue originalmente diseñado para obtener una muestra de lectura oral, y desde su principio que ha provisto un métrico para analizar muchos aspectos de lectura oral y de funcionamiento motora de habla [24] [25].

El Abuelo

¿Tú quieres saber todo acerca del abuelo? Pues bien, él tiene casi 83 años de edad y aún conserva su mente tan brillante como siempre. Él usa un viejo saco negro al que generalmente le faltan algunos botones. Su barba larga y colgante produce en quienes lo observan un profundo sentimiento de máximo respeto. Cuando el habla, su voz es un poco rajada y tiembla ligeramente. Dos veces al día toca un pequeño órgano con excelencia y virtuosismo. Cada día él hace unas caminatas cortas al aire fresco, excepto en el invierno cuando la nieve o el hielo lo impide. Nosotros siempre lo exhortamos a que camine más y que fume menos, pero él siempre responde "Aceite de plátano". Al abuelo le gusta ser moderno en su lenguaje.

Tabla Texto de lectura de la segunda parte de la evaluación.

Test de descripción de la imagen "Ladrón de Galletas".

Después la evaluación con el MMSE el paciente es dado una imagen llamada Ladrón de Galletas para describir. Esta imagen muestra varias acciones como por ejemplo; un niño alcanzando por un pote de galleticas desde un taburete que está a punto de caerse y una niña se las está pidiendo. Mientras tanto una mujer lavando platos no ve que está sucediendo detrás su espalda y, mientras pasa esto, el agua se está desbordando del fregadero.

Luego el paciente debe contar qué es lo que observa en dicha imagen, y esa muestra de habla es la que posteriormente será analizada. La descripción normal de la imagen por los pacientes sanos, es fijarse las acciones sobre las que se puede actuar, como el agua cayendo al piso u el niño sobre el taburete inestable [26]. Por otra parte los pacientes que exhiben síntomas de deterioro cognitivo, usualmente no dan cuenta de estos puntos principales de la imagen. Aún más ellos experimentan dificultades a la hora de organizar los conceptos o a veces no se pueden encontrar las palabras apropiadas para expresarse

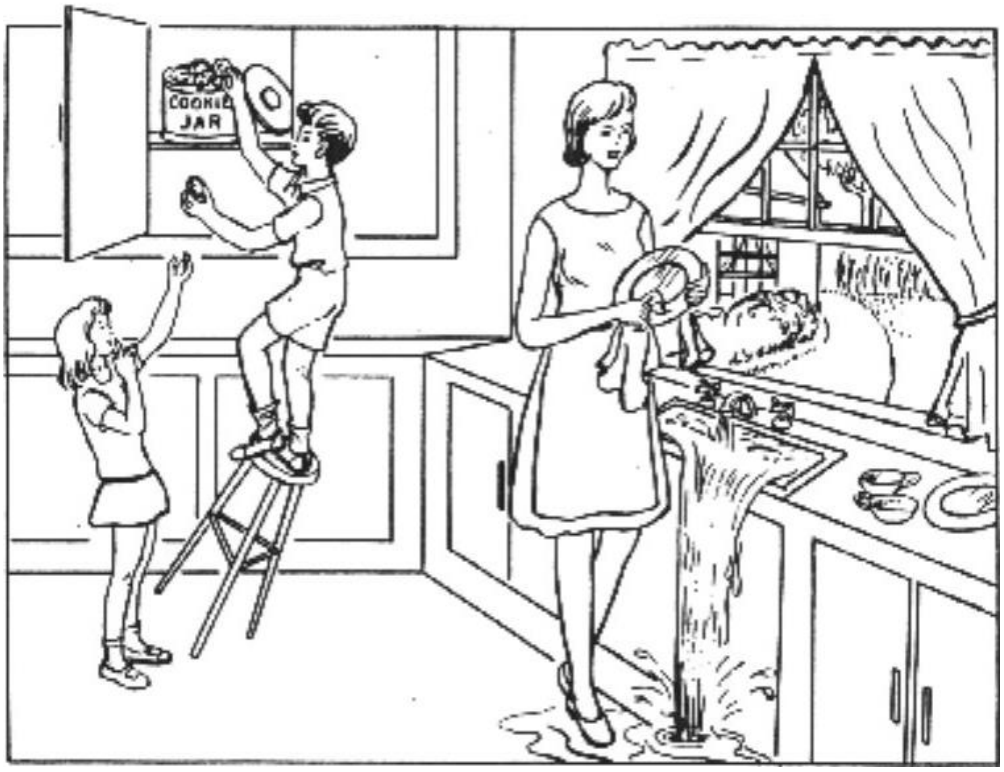

Fig. 2.1 Ladrón de galletas.

2.4 Corpus de voz.

Las muestras de voz de cada paciente fueron grabadas en una sala del Hogar del Adulto Mayor #2, intentando buscar un nivel aceptable de aislamiento acústico. La frecuencia de muestreo empleada fue de 44 kHz para posteriormente ser re muestreadas a 16kHz. El equipo de grabación utilizado estaba formado por un portátil ASUS A55V, dos micrófonos de alta calidad, de tipo diadema, modelo profesional AKG C520L, y un convertidor A/D M-Audio M-Track 2-Channel Portable USB Audio.

Todo el equipamiento ha sido suministrado por la colaboración con el proyecto internacional ‘‘CMC-V2: Caracterización, Modelado y Compensación de Variabilidad en la Señal de Voz’’, aprobado bajo el código TEC2012-37585-C02-02 y dirigido por el profesor Dr.C. Luis A. Hernández Gómez, Jefe del Grupo de Aplicaciones de Procesado de Señales1 (GAPS) de la Universidad Politécnica de Madrid [27].

Fig. 2.2 Micrófono AKG C520L

Micrófono AKG C520L Headworn Condenser: El micrófono Headworn Condenser tiene máximo rango dinámico para sonido vocal perfecto. Su cardioid patrón polar rechaza eficazmente ruido ambiental no deseado. El brazo del micrófono se monta en ambos lados, izquierda o derecha de la banda cabezal según se desee.

El protector de humedad impide traspiración de penetrar el elemento del transductor, para asegurar un grado alto de protección de humedad para el micrófono

Fig. 2.3 Convertidor A/D M-Audio M-Track

La interfaz de M-Track de audio a M-audio® transforma la computadora en un estudio de grabación digital. Un solo cable USB provee el poder y una conexión de datos hacia y desde la computadora portátil o el computador de escritorio. El formato de bajo perfil da control profesional en un paquete portátil.

M-Track ofrece monitoreo ultra-bajo de latencia de la señal entrada a través de los audífonos. El botón de la mezcla del monitor ajusta el balance entre las entradas directas y el sonido pregrabado del software de la computadora.

Todo este proceso de grabación fue controlado por un experto, que se asesoró que las señales de audio tuviesen un nivel de intensidad adecuado sin llegar a estar saturada y que no tuviesen ruidos externos. Además, para disminuir la variabilidad en las grabaciones de distintos locutores, se solicitó a los sujetos leer las frases con un mismo patrón rítmico.

2.5 Software DCGrab v3.0

Para este proceso de grabación se empleó el software DCGrab v3.0 el cual fue desarrollado en el Centro de Estudio de Electrónica y Tecnologías de la Información (CEETI) [28]. El DCGrab v3.0 esta es una herramienta que nos permite la grabación de archivos de audio para el apoyo al diagnóstico del DCL a través de la voz.

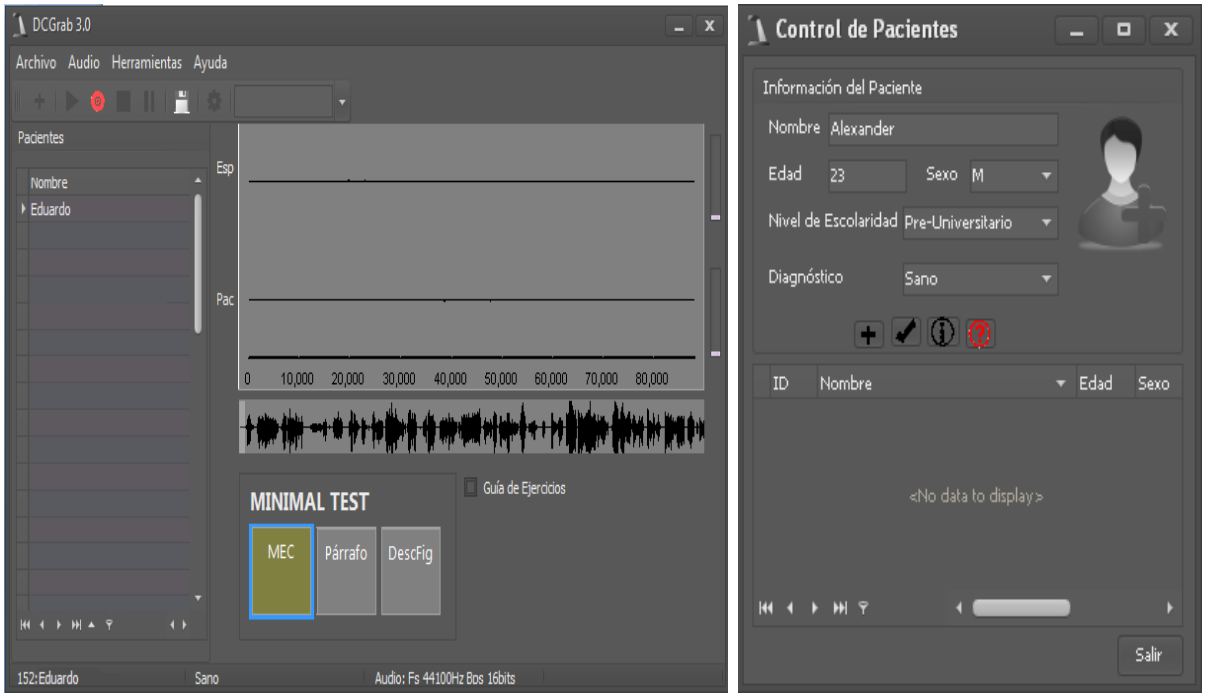

Fig. 2.4 Programa DCGrab 3.0.

Este programa fue diseñado de tal modo que se utiliza el Microsoft Access para crear una Base de Datos con los datos de los pacientes de los cuales fuimos recogiendo sus datos personales y ciertas informaciones clínicas a medida que se producía la grabación de los pacientes. Adicionalmente las señales de audio correspondientes a cada paciente por cada encuesta realizada son guardadas dentro archivos con el mismo sistema.

2.6 Herramientas: Praat, Prosogram y Matlab.

Usaremos una herramienta creada idealmente para este tipo de estudio el cual es el análisis prosódico de las señales de audio que usaremos aquí donde trabajaremos con: el programa Praat, desarrollado y diseñado para el análisis fonético del habla por Paul Boersma y David Weenink, del Instituto de Ciencias Fonéticas de la Universidad de Ámsterdam. [29]

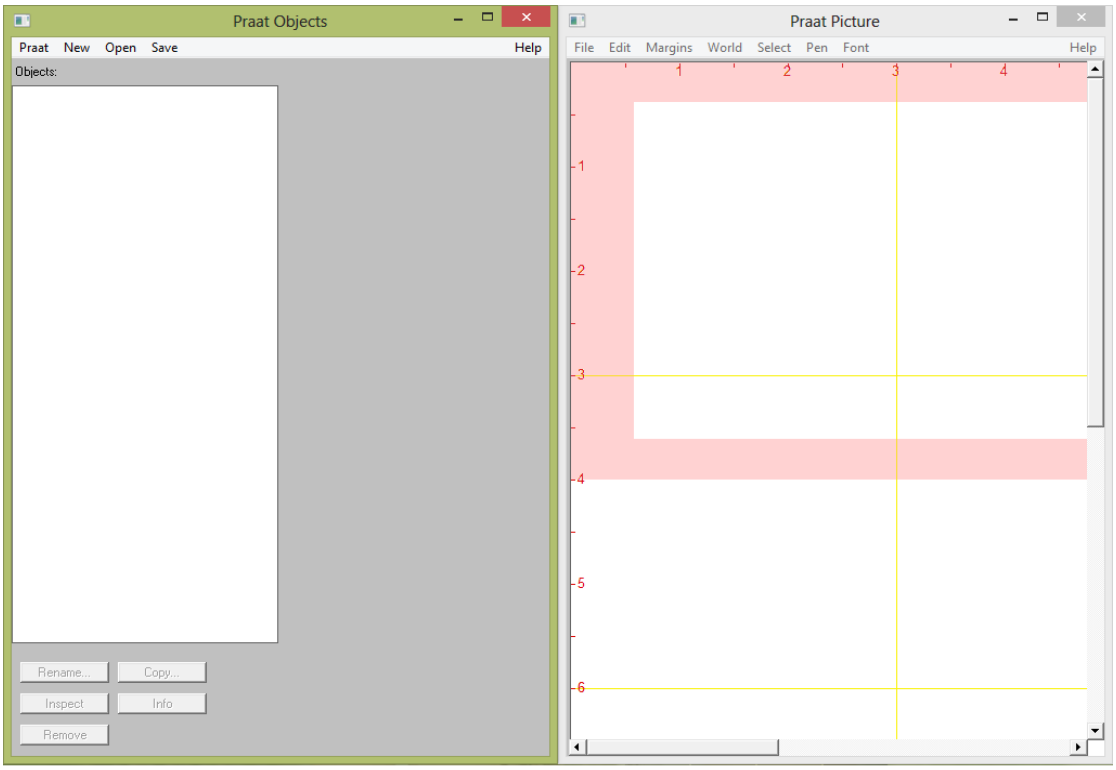

Fig. 2.2 Herramienta Praat.

Todas las opciones tienen asignadas un valor por defecto, de este modo permite que la herramienta pueda ejecutar el análisis prosódico sin verse en la necesidad de que el usuario realice más cambios.

La primera pestaña, Task, permite seleccionar la clase de tarea que deseamos que realice el programa, como realizar todos los cálculos de segmentación, estilización y prosodia, presentándolos de forma gráfica o no; calcular la segmentación del archivo de audio en sílabas y guardarla en un archivo de texto; representar gráficamente el tono del audio en semitonos, sin realizar estilización de frecuencia, etcétera. En función de la tarea escogida,

algunos de los parámetros de análisis siguientes pueden ser irrelevantes, y no será preciso tenerlos en cuenta. Dado que en nuestro estudio utilizaremos la tarea que viene seleccionada por defecto (Prosograma y perfil prosódico), la cual realiza todos los pasos necesarios para calcular los datos prosódicos de la señal de voz y representa gráficamente sobre un eje temporal los núcleos silábicos junto con la progresión de F0 y la intensidad, referimos a continuación los parámetros de análisis que podemos seleccionar para dicha tarea:

Este programa posee su propio lenguaje de programación y debido a esto es posible escribir códigos de acuerdo con las especificaciones que nos hagan falta en ese momento y poder ejecutarlo a través de esta herramienta. Para nuestro trabajo, se usara el código Prosogram, el cual fue desarrollado por Piet Mertens, profesor de lingüística francesa y computacional. [30]

El entendimiento y comprensión de cómo es el funcionamiento debido de estas herramientas complementarias es primordial para la continuación del estudio prosódico que se está llevando a cabo durante este Capítulo, por lo que a continuación se lleva a cabo una descripción breve de los mismos.

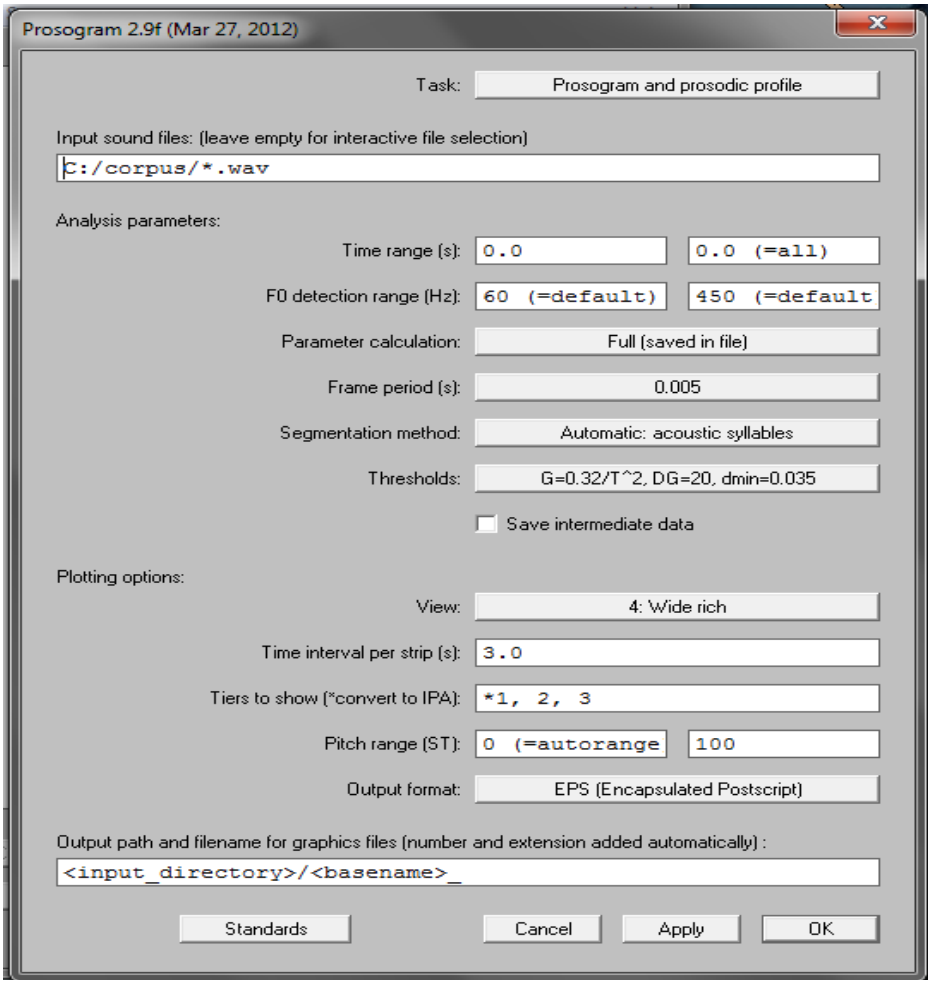

Fig. 2.3 Código Prosogram.

La función de este código es la realización de transcripciones prosódicas de señales de audio a partir de parámetros determinados de entrada los cuales pueden ser definidos por los usuarios. Lo primero que realiza el programa es el cálculo de parámetros acústicos ya sea la frecuencia fundamental, la intensidad, o las partes en los que hay voz. Luego se encarga de obtener la segmentación de dichos parámetros en unidades silábicas o vocálicas, este selecciona sólo aquellas que sobrepasan un determinado umbral de intensidad o volumen sonoro. Por tanto el resultado obtenido es una representación visual sobre un eje temporal de estas unidades junto con las progresiones de intensidad. También, se puede crear un archivo de texto en el cual se pueden incluir diversos datos prosódicos los cuales pueden hacer referencias a cada uno de los núcleos detectados.

Con el uso de la herramienta Matlab vs2010 se puede analizar los datos recogidos a partir de la herramienta *Prosogram* v29f. Matlab es muy útil para el tratamiento de señales y el manejo de datos en forma matricial. Por lo tanto a lo largo de este trabajo se utilizará la herramienta Matlab v2010 para extraer otros datos de interés a partir de los calculados por *Prosogram* v29f y llevaremos a cabo diferentes modelos estadísticos para tratar de obtener diferencias significativas entre las señales de voz de los grupos de DCL y Sanas.

2.7 Tipos de Segmentación.

La herramienta Prosogram nos da la posibilidad de selección de varios tipos de segmentación los cuales son: la segmentación automática, la segmentación externa y la segmentación semiautomática.

- **Segmentación automática:** Esta segmentación deja a propio criterio del programa la detección de los núcleos silábicos del archivo de voz de entrada, el cual se basa en el parámetro de intensidad de las unidades silábicas.
- **Segmentación externa:** Este consiste en otorgarle al programa un archivo de texto con las etiquetas de los fonemas que se encuentran presentes en el audio al cual se va a analizar, mostrando los intervalos de tiempo en donde estos se sitúan. Claramente, al llevarse a cabo la realización de este procedimiento implica que el archivo de voz tenga que analizarse anteriormente por parte del programador, el cual deberá ubicar los fonemas manualmente mediante la utilización de alguna otra herramienta de programación.
- **Segmentación semiautomática:** Esta segmentación utiliza una combinación de las dos segmentaciones anteriores, el cual necesita también un archivo externo de etiquetas.

Antes de comenzar el estudio sobre la base de datos, se realizará una prueba sobre un archivo de audio empleando estos tres métodos de segmentación, para evaluar cada uno para nuestro trabajo.

A modo de resumen de esta sección vamos a escoger la segmentación automática ya que nos da una segmentación fiable en comparación con la externa y la semiautomática.

2.8 Base de Datos de DCL.

Par analizar las características del habla nos basamos previa disponibilidad de grabaciones que propiciamente deben corresponder tanto a voces pertenecientes a pacientes los cuales padezcan de la enfermedad como a voces de personas sanas .En esta parte les mostramos criterios básicos de diseño y especificación de una base de datos de pacientes con DCL y pacientes sanos la cual fue obtenida para poder posibilitar el futuro trabajo desarrollado en este capítulo, el cual se basa en la búsqueda de rasgos para poder dar un diagnóstico a través del análisis del habla de casos con DCL.

Tabla 2.1 Base de datos de DCL

Rasgos		**GRUPO**	
		SANO	**DCL**
Cantidad de participantes		11	10
SEXO	**Mujeres**	18.2%	70.0%
	Hombres	81.8%	30.0%
EDAD (Años)	**Edad media (Mín-Máx)**	76.9 (64-87)	80.3 (65-92)
Nivel de Escolaridad			
Analfabeto		0 %	0 %
Primaria		100 %	70 %
Secundaria		0 %	0 %
Pre-Universitario		0 %	0 %

Técnico Medio	0 %	30 %
Universitario	0 %	0 %

Con la cooperación del Hogar del Adulto Mayor #2 y #3 de la ciudad de Santa Clara fue posible hacer una captura de la base de datos. La base de datos como muestra en la Tabla 2.1, consiste de 11 personas sanas de ellas 9 son hombres y 2 mujereas; 10 personas con DCL de ellas 3 hombres y 7 mujeres. En la Tabla se puede ver además que el promedio de edad de las personas con DCL es mayor que las de las personas sanas (80.3 por 76.9 respectivamente). Las pacientes de ambos grupos fueron seleccionados para tratar de garantizar, en la medida de lo posible, que tuvieran requisitos fisiológicos similares en relación a su edad y nivel de escolaridad. Sea necesaria tener características similares entre las poblaciones de personas Sanas y DCL para que no haya influencia de factores exteriores que predisponen la demencia.

2.9 Obtención de Parámetros Prosódicos.

La herramienta con la que estamos trabajando *Prosogram* tendrá en su salida los parámetros prosódicos previamente definidos, los cuales se refiere a la fluidez y rapidez de lectura, y hemos definido a continuación:

- Tiempo de lectura: Es la duración total que el paciente realiza la lectura, desde que comienza el primer núcleo que el programa detecta hasta la culminación del último.
- Cantidad de pausas: Es la cantidad de periodos de silencio mayores que un cierto umbral.
- Proporción de pausas: La duración de las pausas dividido entre la duración de lectura.
- Tiempo de fonación: Es el tiempo de duración en que son pronunciados los núcleos más los silencios los cuales no tomados como pausas.
- Proporción de fonación: La duración de fonación dividida entre el tiempo de duración de la lectura.

- Rapidez de elocución: Es la cantidad de sílabas (núcleos) dividido entre el tiempo de duración de lectura.
- Rapidez de articulación: Es la cantidad de sílabas dividido entre la duración de fonación.

Al realizarse la extracción de los valores, debemos tener en cuenta que para llevar a cabo la ejecución de un cálculo correcto realizar un cálculo adecuado respecto a las medidas referentes, a la duración de fonación y pausas, también tenemos que tener en cuenta el valor de gap obtenido a la salida de la herramienta Prosogram para que el núcleo inicias como cero.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

Como recuento del capítulo anterior hemos descrito el análisis prosódico usando las herramientas descritas anteriormente, y la conformación de una base de datos de voz capturada en el Hogar del Adulto Mayor #2 y #3 de Santa Clara para una extracción de los rasgos prosódicos que se van a emplear para el diagnóstico de un posible deterioro cognitivo. Para ello fue necesaria tener dos grupos de sujetos (DCL y de Control) de modo tal que ambos tuviesen características similares, para garantizar una comparación adecuada entre ambos grupos de personas. Por tanto en este capítulo les mostraremos los resultados obtenidos a partir de pruebas estadísticas y de clasificación realizada que se le realizaron a los rasgos prosódicos de dicha base de datos, con un objetivo en específico rasgos discriminantes para separar los dos grupos de pacientes analizados.

3.1 Descripción estadística de los rasgos extraídos.

El tratamiento matemático de los datos prosódicos ha sido una limitante en otros artículos tratados por lo cual llegamos a un nuevo enfoque, de modo más directo y experimental que tal caso y estamos hablando de una representación visual de la distribución temporal de sílabas. Al ver los resultados que hemos obtenido nos damos cuenta de que apoyan a los obtenidos en apartados anteriores de una manera diferente, lo que aumenta aún más la importancia de estos a la hora de encontrar determinadas características que nos permitan una detección precoz del deterioro cognitivo.

Al realizar estos dos tipos de mapas, el que nos resulta más confuso es el de silabas ya que al tener un número muy elevado de núcleos en cada frase los resultados obtenidos mediante observación visual son imprecisos y prácticamente inexistentes. Por eso empleando el código *Prosogram* obtenemos resultados que si son de interés.

La tabla 3.1 muestran los resultados obtenidos a partir de los análisis estadísticos de los rasgos prosódicos anteriormente.

Tabla 3.1 Valores estadísticos de rasgos del grupo Control.

Rasgos Prosódicos	N	Media	Mínimo	Máximo	Desv.est.
Duración de lectura (seg)	11	128.35	59.21	249.76	63.76
Número de pausas	11	61.91	27	124	33.35
Proporción de pausas	11	47.45	30.74	64.86	10.99
Tiempo de fonación (seg)	11	62.15	33.54	107.21	20.64
Proporción de fonación	11	52.55	35.14	69.27	10.99
Velocidad de locución (sílabas/seg)	11	2.57	1.61	3.62	0.61
Velocidad de articulación (sílabas/seg)	11	4.89	4.38	6.39	0.57

En la Tabla 3.2, podemos observar que el grupo de DCL tienen un número de pausas casi el doble que los del grupo de Control. También al visualizar la diferencia en el tiempo de fonación que existe entre los dos grupos, nos damos cuenta de que el grupo de Control posee un tiempo mucho más corto que el del grupo DCL y todo esto debido a que ellos leen con menos balbuceos, más precisión y menos repeticiones.

Tabla 3.2 Valores estadísticos de rasgos del grupo DCL.

Rasgos Prosódicos	N	Media	Mínimo	Máximo	Desv.est.
Duración de lectura (seg)	10	199.77	178.37	220.72	14.61
Número de pausas	10	102.80	84	126.22	12.30
Proporción de pausas	10	60.75	52.79	68.70	5.46
Tiempo de fonación (seg)	10	90.03	78.59	104.20	7.19
Proporción de fonación	10	49.89	43.01	58.42	4.24
Velocidad de locución (sílabas/seg)	10	2.56	2.12	3.02	0.32
Velocidad de articulación (sílabas/seg)	10	5.71	4.49	7.03	0.94

Por consiguiente, lo indicado previamente de que el resultado de la duración de la lectura del grupo de DCL es mucho mayor que la registrada por la del grupo de control, como podemos observar en las tablas la diferencia es poco más de 70 segundos

Por lo tanto como resultado de lo anteriormente dicho las velocidades de locución y articulación de los pacientes de Control son más rápidas que los de DCL.

3.2 Obtención de rasgos significativos

Al terminar de analizar los archivos de audio de nuestra base de datos y culminar la extracción de los datos de mayor interés, nos proponemos comprobar cuáles de ellos pueden ser usados para una distinción entre ambos grupos de sujetos. Esto quiere decir que la diferencia de valores obtenidos para cada grupo es lo suficientemente amplia o considerable como para resultar un factor de exclusión.

Con este objetivo, emplearemos dos métodos distintos que nos permitirá la obtención de estos parámetros significativos .La primera será la prueba de Wilcoxon, y la otra el trazado de curvas ROC.

3.2.1 Prueba de Wilcoxon

La prueba de los rangos con signo de Wilcoxon es una prueba no paramétrica para comparar la mediana de dos muestras relacionadas y determinar si existen diferencias entre ellas. Para ello se utilizó la función ranksum del Matlab v2010 que realiza una prueba de Wilcoxon de dos muestras para comparar las distribuciones de los valores de dos vectores de datos esta prueba la hipótesis nula de que los datos de dos variables son las muestras de distribuciones continuas con igualdad de medianas, contra la alternativa de que no lo son. La prueba se supone que las dos muestras son independientes y estas pueden tener diferentes longitudes. También devuelve un valor lógico que indica la decisión de prueba. El resultado $h = 1$ indica un rechazo de la hipótesis nula por tanto se acepan los parámetros como significativos, y $h = 0$ indica que no se rechaza la hipótesis nula al nivel de significación del 5% por tanto los parámetros no se aceptan como no significativos.

Empleando dicha función con todos los parámetros prosódicos de nuestro estudio, se obtienen los resultados recogidos en la Tabla 3.3.

Tabla 3.3 Valores de *h* y *p* de la prueba de Wilcoxon

Rasgos Prosódicos	*h*	Valores de *p*
Duración de lectura (seg)	1	0.0151
Número de pausas	1	0.0083
Proporción de pausas	1	0.0102
Tiempo de fonación (seg)	1	0.0022
Proporción de fonación	0	0.4597

Velocidad de locución (sílabas/seg)	0	0.9719
Velocidad de articulación (sílabas/seg)	1	0.0378

Con respecto al resultado de la tabla anterior el umbral que tomamos para decidir cuales rasgos son significativos y cuáles no, se establecerá a partir de α igual a 0,05, empleado convencionalmente. De este modo, aquellos datos cuyo valor de p sea igual o inferior al nivel de significación podrán ser considerados como significativos a la hora de ser empleados en una clasificación entre los grupos de DCL y Sanos.

Empleando el criterio anterior se obtienen los siguientes resultados ver Tabla 3.4

Tabla 3.4. Clasificación de los Rasgos prosódicos según su capacidad discriminativa mediante el empleo de la prueba de Wilcoxon.

Rasgos no significativos	**Rasgos significativos**
• Proporción de fonación • Velocidad de locución	• Número de pausas • Velocidad de articulación • Duración de lectura • Proporción de pausas • Tiempo de fonación

3.2.2 Obtención mediante trazado de curvas ROC

Como segundo método de comprobación de parámetros significativos se emplea el trazado de las curvas ROC (Característica Operativa del Receptor) este proporciona herramientas para seleccionar los modelos posiblemente óptimos y descartar modelos que no sean

óptimos independientemente del coste de la distribución de las dos clases sobre las que se decide. La curva ROC es también independiente de la distribución de las clases en la población (en diagnóstico, la prevalencia de una enfermedad en la población). El análisis ROC se relaciona de forma directa y natural con el análisis de coste/beneficio en toma de decisiones diagnósticas, proporcionando una comparación visual directa entre pruebas en una escala común, a diferencia de otro tipo de gráficos, como los diagramas de puntos o los histogramas de frecuencias, que requieren diferentes gráficos cuando difieren las escalas. Además, el trazado de estas curvas permitirá una clasificación de parámetros válidos y no válidos.

La implementación del método se ha llevado a cabo mediante Matlab v2010, mediante una función que calcula la sensibilidad y la especificidad de los datos que pasamos como parámetros, por medio de los cuales podemos representar gráficamente las curvas ROC. Además, este programa considera los valores de los datos de entrada (los datos prosódicos de los dos grupos para una determinada característica) para calcular los umbrales de la clasificación, y de todos ellos, selecciona automáticamente aquel con mayor precisión. Así mismo, calcula, a modo de orientación a la representación visual, el área bajo la curva (AUC, por sus siglas en ingles). Aún más, tiene otra ventaja de que, se lo compara los datos de entrada con el umbral de clasificación de tal manera que considera que aquellos que lo superen se corresponderán con sujetos enfermos, de modo que aquellos parámetros prosódicos cuyo valor alto indica una correspondencia con el grupo asintomático deberán introducirse en el programa con valores negativos.

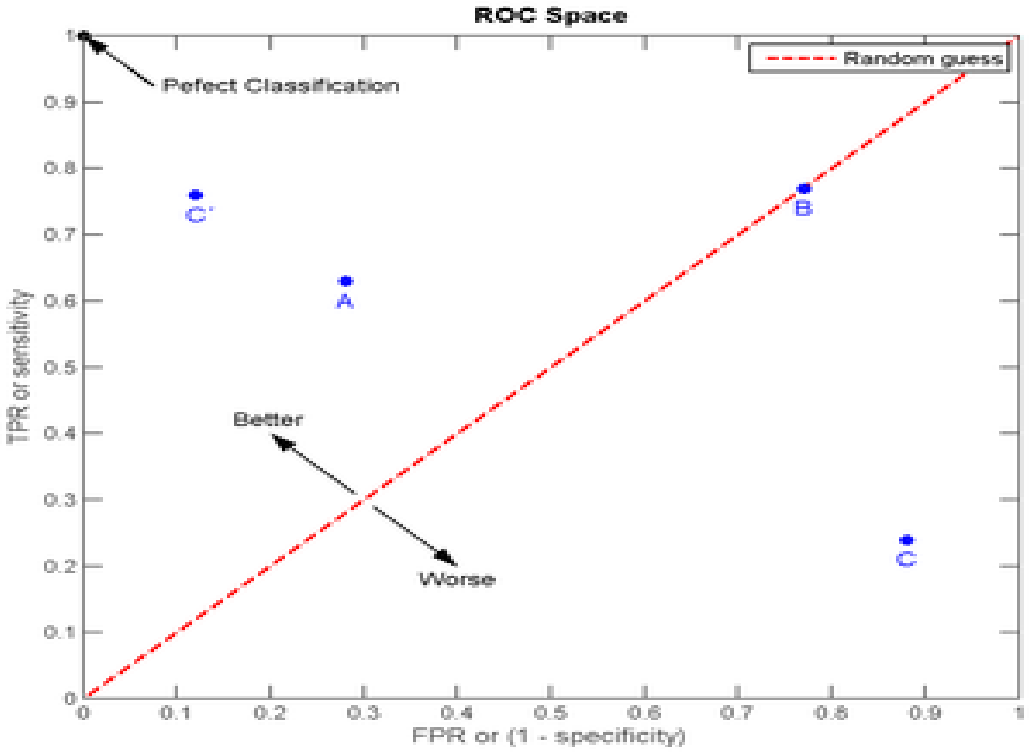

Fig. 3.1 El espacio ROC

En la figura se puede observar que diagonal divide el espacio ROC. Los puntos por encima de la diagonal representan los buenos resultados de clasificación (mejor que el azar), puntos por debajo de la línea de los resultados pobres (peor que al azar). Nótese que la salida de un predictor consistentemente pobre simplemente podría ser invertida para obtener un buen predictor.

Con las curvas ROC se representan la tasa de verdaderos positivos (*sensibilidad*) frente al ratio de falsos positivos (1 - *especificidad*) según se varía el umbral de discriminación. Cada punto de la curva representa un par *S* (*sensibilidad*)/1-*E* (*especificidad*) correspondiente a un nivel de decisión determinado. Una prueba con discriminación perfecta, sin solapamiento de resultados en las dos poblaciones, tiene una curva ROC que pasa por la esquina superior izquierda, donde *S* y *E* toman valores máximos(*S* y *E* = 1).

Si la curva queda por encima de la diagonal de referencia, podremos tomar ese parámetro como bueno a la hora de realizar una clasificación entre los grupos de DCL y Sanos,

mientras que si queda por debajo, lo descartaremos. En la Tabla 3.5 se muestran en forma de tabla los resultados del cálculo de AUC.

Dado que no existe un valor concreto de AUC a partir del cual se pueda considerar válido para discriminar entre sujetos enfermos y asintomáticos, suelen tomarse intervalos de AUC para los cuales se considera si el modelo estadístico es válido o no. Para el trabajo tomamos en cuenta que un AUC igual a 0,5, que se correspondería con una curva ROC sobre la diagonal de referencia, indica una clasificación al azar (hay un 50 % de posibilidades de que se clasifique correctamente a un paciente como sano o enfermo), y que un AUC igual a 1 equivaldría a una clasificación perfecta, con un 100 % de acierto, podemos considerar el valor de AUC igual a 0,75 como un umbral intermedio para la clasificación.

Tabla 3.5 Datos relativos de las curvas AUC de rasgos prosódicos

Rasgos Prosódicos	**AUC (%)**
Duración de lectura (seg)	85.71
Número de pausas	90.48
Proporción de pausas	80.95
Tiempo de fonación (seg)	90.48
Proporción de fonación	71.43
Velocidad de locución (sílabas/seg)	61.9
Velocidad de articulación (sílabas/seg)	76.19

De este modo, podríamos definir un modelo válido a partir de un AUC de 0,75. No obstante, como ya se ha indicado, este valor sería orientativo, ya que, al fin y al cabo, este dato no hace otra cosa que señalar la tasa de éxito al clasificar correctamente a un sujeto en

un grupo determinado empleando un rasgo prosódico concreto. Si tenemos una tasa del 80 %, hay un 20 % de posibilidades de que el modelo falle, aunque bien es cierto que cuanto mayor sea nuestro valor de AUC, más posibilidades hay para una clasificación correcta.

De este modo, definimos los siguientes parámetros para el análisis:

- AUC < 0,75: El rasgo puede considerarse poco significativo.
- 0,75 < AUC < 0,8: El rasgo puede considerarse significativo.
- 0,8 < AUC: El rasgo puede considerarse muy significativo.

Según los intervalos de valores del AUC establecidos, podemos ver en la Tabla 3.5 los parámetros prosódicos extraídos en nuestro estudio que podemos considerar como significativos, que son la duración total de la tarea de lectura, las proporciones de fonación y pausas (cuyos valores coinciden al tratarse de parámetros complementarios), y la velocidad de elocución, correspondientes a los valores medios de los grupos. De todos estos, entran en el intervalo de parámetros muy significativos el número de pausas y la velocidad de articulación (ver Tabla 3.6).

Tabla 3.6 Clasificación de los Rasgos prosódicos según su capacidad discriminativa mediante el empleo de curvas AUC

Rasgos no significativos	**Rasgos posiblemente significativos**	**Rasgos significativos**
• Velocidad de locución • Proporción de fonación	• Velocidad de articulación	• Número de pausas • Duración de lectura • Proporción de pausas • Tiempo de fonación

3.3 Comparativa de resultados

Una vez obtenidos los resultados podemos comprobar que son casi equivalentes en ambos métodos. La diferencia estriba en el carácter impreciso del método de las curvas ROC, al no tener un umbral fijo de probabilidad para decidir si el valor de AUC es discriminativo o no. De esta limitación se debe el hecho de clasificar como posible parámetro significativo la duración de lectura y la velocidad de locución, cuyo valor de AUC no supera nuestro umbral elegido pero tampoco se encuentra muy alejado como para decidir descartarlo del todo.

Con el método de *Wilcoxon*, obtenemos resultados directos sin tener que aplicar un criterio de decisión propio para discernir qué probabilidades son válidas, salvo por el nivel de significación, que resulta ser más preciso. En este caso, observamos que la duración de lectura, proporción de pausas, proporción de fonación y velocidad de locución, que calificábamos como indeterminadas mediante el método AUC, son clasificados como no significativos directamente. A la vista de estos resultados, optamos por clasificar entonces dichos rasgos como no significativos.

Así que, los rasgos extraídos en nuestro estudio quedan clasificados tal y como se indica a continuación en la Tabla 3.7.

Tabla 3.7 Clasificación final de los rasgos prosódicos según los métodos estadísticos

Parámetros no significativos	**Parámetros significativos**
• Proporción de fonación • Velocidad de locución	• Número de pausas • Velocidad de articulación • Duración de lectura • Proporción de pausas • Tiempo de fonación

3.4 Clasificación de los pacientes

Lo que vamos hacer a continuación es centraremos en el desarrollo de métodos que nos brinde una posibilidad de una clasificación atinada de un paciente en un grupo determinado. Los siguientes clasificadores fueron utilizados para el presente trabajo:

- Clasificador Análisis Discriminante Lineal (LDA siglas en ingles).
- Clasificador Función de Base Radial (RBF siglas en ingles)

Para validar los resultados obtenidos por los dos clasificadores se empleó la técnica de validación cruzada dejando uno fuera, lo cual implica separar los datos de forma que para cada iteración tengamos una sola muestra para los datos de prueba y todo el resto conformando los datos de entrenamiento.

La validación cruzada, es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre datos de entrenamiento y prueba. Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar cómo de preciso es un modelo que se llevará a cabo a la práctica. Es una técnica muy utilizada en proyectos de inteligencia artificial para validar modelos generados.

La validación cruzada proviene de la mejora del método de retención. Este consiste en dividir en dos conjuntos complementarios los datos de muestra, realizar el análisis de un subconjunto (denominado datos de entrenamiento), y validar el análisis en el otro subconjunto (denominado datos de prueba), de forma que la función de aproximación sólo se ajusta con el conjunto de datos de entrenamiento y a partir de aquí calcula los valores de salida para el conjunto de datos de prueba (valores que no ha analizado antes). La ventaja de este método es que es muy rápido a la hora de computar. Sin embargo, este método no es demasiado preciso debido a la variación de los resultados obtenidos para diferentes datos de entrenamiento. La evaluación puede depender en gran medida de cómo es la división entre datos de entrenamiento y de prueba, y por lo tanto puede ser significativamente diferente en función de cómo se realice esta división. Debido a estas carencias aparece el concepto de validación cruzada [31].

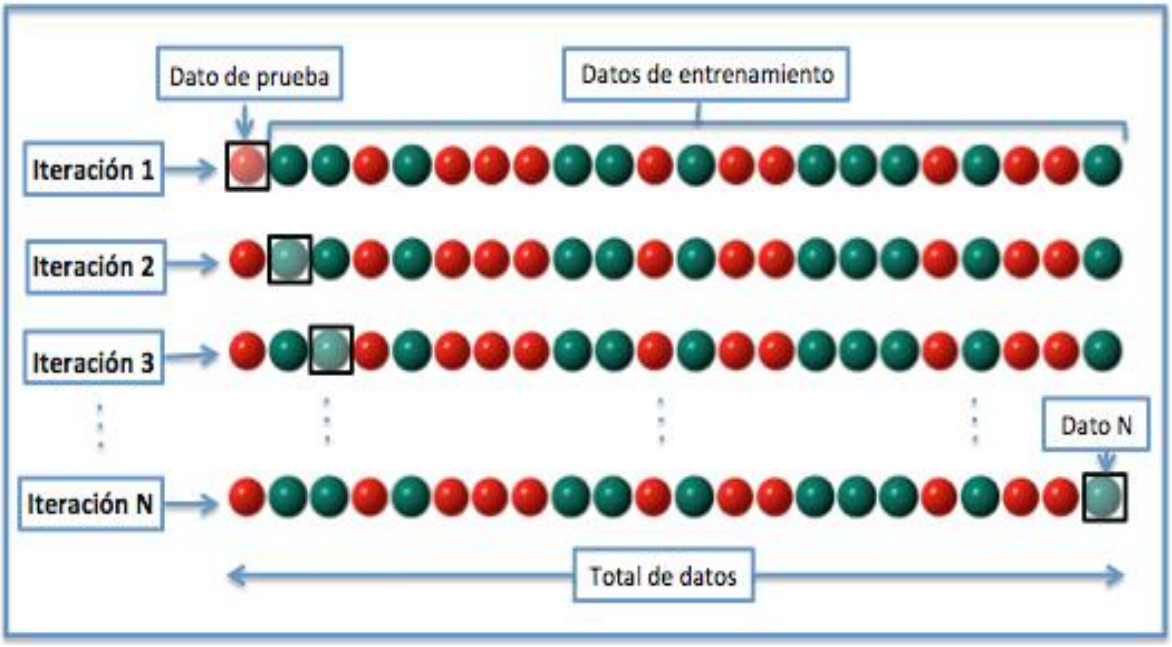

Fig. 3.2 Validación cruzada dejando uno fuera

En la validación cruzada dejando uno fuera se realizan tantas iteraciones como muestras (*N*) tenga el conjunto de datos. De forma que para cada una de las *N* iteraciones se realiza un cálculo de error. El resultado final lo obtenemos realizando la media aritmética de los *N* valores de errores obtenidos, según la fórmula:

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i.$$

Donde se realiza el sumatorio de los *N* valores de error y se divide entre el valor de *N*.

3.5 Clasificador LDA

Este tipo de clasificador toma decisiones a partir de la combinación lineal de sus características, mediante una función discriminante de tipo lineal.

Para el empleo del clasificador LDA se ha usado la función *classify* de Matlab. Debido a esta función, introducimos los parámetros que definirán la clasificación de un archivo de

voz en el grupo de DCL o en el grupo de personas Sanas a partir de los valores de dichos parámetros del propio audio.

El proceso de clasificación mediante el LDA se realiza de dos maneras distintas: utilizando como características únicamente los parámetros obtenidos como significativos y utilizando todos los rasgos.

3.5.1 Clasificador LDA con rasgos significativos

Primeramente se diseñó el clasificador haciendo uso de los cinco rasgos prosódicos extraídos según los análisis estadísticos para la realización de este estudio que dieron como resultado significativos. El principal motivo para la realización de este clasificador es la comprobación de una posible mejoría en la clasificación al tener en cuenta un mayor número de rasgos diferentes. Los rasgos empleados para el diseño son, entonces, los siguientes:

1. Número de pausas
2. Velocidad de articulación
3. Duración de lectura
4. Proporción de pausas
5. Tiempo de fonación

En este caso, el LDA tendrá menos datos para realizar la clasificación que en el otro caso en el cual se emplearan todos los rasgos, pero como los que vamos a utilizar son los rasgos significativos deberían ser suficientes para la realización de un clasificador válido. Realizando a continuación el método de validación cruzada, se obtienen los resultados para la probabilidad de clasificación correcta que se muestran en la Tabla 3.8.

Tabla 3.8 Resultados del clasificador LDA para rasgos significativos.

Grupos	Correctos	Erróneos	Tasa de Clasificación Correcta (%)
Sanos	9	2	81.82
DCL	8	2	80.00
Total	17	4	80.95

Estos resultados de un 80.9 % de clasificación correcta son comparables a los reportados en la literatura. La técnica de clasificación mediante LDA es bastante simple y linear lo cual un resultado con un error menor al 20 % puede considerarse adecuado para la data y estudio bajo análisis.

3.5.2 Clasificador LDA con todos los rasgos

Ahora se realizarán los mismos pasos que en el apartado anterior que son: *classify* y validación cruzada, pero introduciéndole al LDA todos los rasgos prosódicos extraídos en nuestro trabajo. Lo que se propone con esta parte es que el resultado debe mejorar al tener en cuenta en el clasificador un mayor número de rasgos. Los rasgos empleados para el diseño son los siguientes:

1. Duración de lectura
2. Número de pausas
3. Proporción de pausas
4. Tiempo de fonación
5. Proporción de fonación
6. Velocidad de locución
7. Velocidad de articulación

Aquí tomaremos todos los parámetros que habíamos extraído de cada uno de los sujetos de la base de datos. De este modo, para diseñar el clasificador LDA, se introducen los sietes rasgos prosódicos extraídos: siete rasgos por cada uno de los ocho sujetos del grupo de DCL y por supuesto sietes rasgos por cada uno de los diez sujetos del grupo de personas Sanas. A continuación, se introducen los valores de dichos parámetros de un sujeto en concreto. El clasificador LDA asignará al paciente a uno de los dos grupos en función del parecido de sus parámetros con los introducidos como entrenamiento. Los resultados son mostrados en la Tabla 3.9 a continuación.

Tabla 3.9 Resultados del clasificador LDA para todos los rasgos prosódicos.

Grupos	Correctos	Erróneos	Tasa de Clasificación Correcta (%)
Sanos	8	3	81.82
DCL	10	0	100.00
Total	18	3	85.71

Como se puede apreciar cuando hacemos una comparación entre ambas tablas nos damos cuenta de que entre mayor sea la cantidad de rasgos prosódicos que sean extraídos para este clasificador LDA mejor va a ser nuestra clasificación ya que puede darse el caso que se quede algún rasgo que sea significativo dentro de los que no lo son y esto puede traer consigo cambios en nuestra prueba y cómo podemos ver en las tablas la tasa de error de clasificación disminuye a 14.29 % lo que significa una reducción del 25.0 %

3.6 Clasificador RBF.

La RBF están formada con una arquitectura rígida de tres capas: la de entrada, la oculta y la de salida la cual está diseñada con neuronas en la capa oculta activadas mediante funciones radiales de carácter no lineal con sus centros gravitacionales propios y en la capa

de salida mediante funciones lineales,. Estas tienen un mejor desempeño con un mayor volumen de datos de entrenamiento

Tal como ya se dijo anteriormente, una RBF, está conformada de tres capas.

- La capa de entrada que sirve para los ejemplos o patrones de entrenamiento y prueba.
- La capa oculta completamente interconectada entre todos sus nodos con la capa de entrada y activada a través de la función radial (gaussiana).
- La capa de salida, también completamente interconectada a la capa oculta y activada a través de una función lineal contínua.

El entrenamiento, es solamente hacia adelante. De este modo, la salida z de una RBF, en general, está influenciada por una transformación no lineal originada en la capa oculta a través de la función radial y una lineal en la capa de salida a través de la función lineal continua.

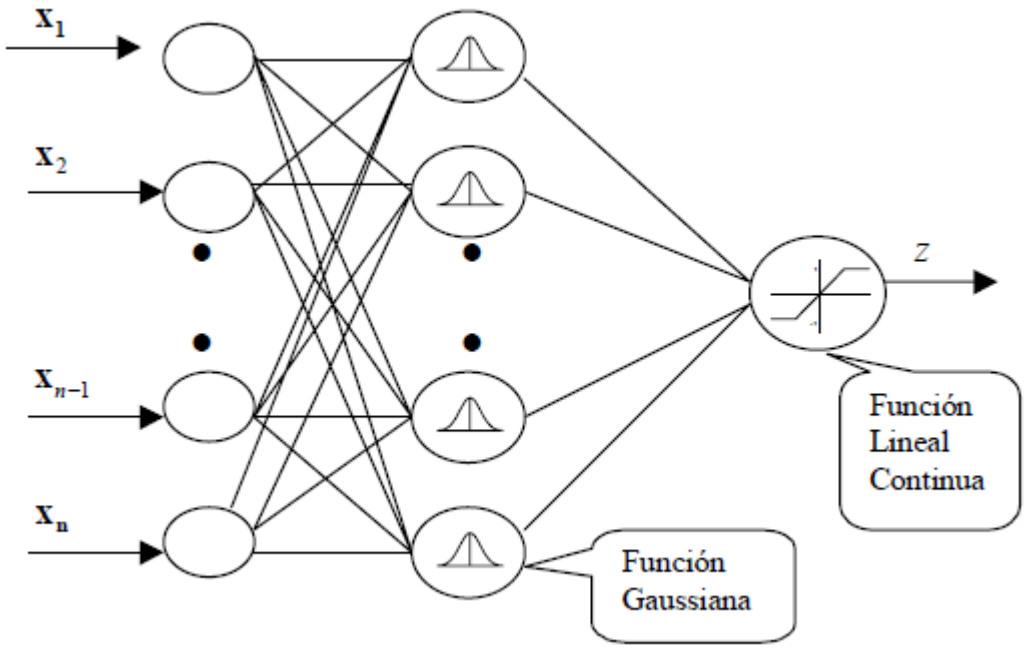

Fig. 3.3 Representación de las 3 capas de la RBF.

3.6.1 Clasificador RBF con los rasgos significativos.

Realizando los mismos pasos que usamos con el clasificador LDA pero en este caso haciendo uso de la función en Matlab *newrb* para la realización de el mismo y haciendo uso de los cinco rasgos prosódicos extraídos según los análisis estadísticos para la realización del estudio que dieron como resultado significativos. El principal motivo para la realización de este clasificador es que es un poco mejor y con esto deben mejorar la clasificación. Los rasgos empleados para el diseño son, entonces, los siguientes:

1. Número de pausas
2. Velocidad de articulación
3. Duración de lectura
4. Proporción de pausas
5. Tiempo de fonación

Como en el otro caso este va a tener menos rasgos ya que solo vamos a usar los que son significativos solamente. Ver Tabla 3.10.

Tabla 3.10 Resultados del clasificador RBF para rasgos significativos.

Grupos	Correctos	Erróneos	Tasa de Clasificación Correcta (%)
Sanos	8	3	81.82
DCL	10	0	100.00
Total	18	3	85.71

Estos resultados de un 85.7 % de clasificación correcta son comparables a los mejores resultados reportados en la literatura. Dado el bajo tamaño de la data para entrenar la red los resultados pueden considerarse muy alentadores.

3.6.2 Clasificador RBF con todos los rasgos.

Ahora se realizarán los mismos pasos que en el apartado anterior que son: *newrb*, pero introduciéndole al RBF todos los rasgos prosódicos extraídos en nuestro trabajo. Lo que se propone con esta parte es que el resultado debe mejorar al tener en cuenta en el clasificador un mayor número de rasgos. Los rasgos empleados para el diseño son los siguientes:

1. Duración de lectura
2. Número de pausas
3. Proporción de pausas
4. Tiempo de fonación
5. Proporción de fonación
6. Velocidad de locución
7. Velocidad de articulación

Aquí tomaremos todos los parámetros que habíamos extraído de cada uno de los sujetos de la base de datos. De este modo, para diseñar el clasificador RBF, se introducen los sietes rasgos prosódicos extraídos: siete rasgos por cada uno de los ocho sujetos del grupo de DCL y por supuesto sietes rasgos por cada uno de los diez sujetos del grupo de personas Sanas. A continuación, se introducen los valores de dichos parámetros de un sujeto en concreto. Los resultados son mostrados en la Tabla 3.11 a continuación.

Tabla 3.11 Resultados del clasificador RBF para todos los rasgos prosódicos.

Grupos	Correctos	Erróneos	Tasa de Clasificación Correcta (%)
Sanos	10	1	90.91
DCL	10	0	100.00
Total	20	1	95.24

Si realizamos una comparación entre las tablas de los clasificadores nos damos cuenta que el clasificador RBF nos da una tasa de clasificación mucho mejor que el LDA en cuanto a los rasgos significativos que varía de un 80.95% a un 85.71% y haciendo una comparación usando todos los rasgos nos da un incremento de un 10 % (85.71% vs 95.24%). Estos resultados muy buenos y superiores a los resultados publicados en la literatura para otras bases de datos.

3.7 Conclusiones del capítulo

A lo largo de este capítulo hemos realizado distintos tipos de análisis de ciertos rasgos prosódicos a partir de una base de datos previamente conformada la cual nos brinda la posibilidad de ver una diferencia entre personas con DCL y personas Sanas.

Se realizó un análisis estadístico de los siete rasgos prosódicos seleccionados para determinar su nivel de relevancia para el estudio bajo análisis obteniendo una evaluación de significancia para cada rasgo que luego fue empleada en la clasificación final para determinar la presencia o no de DCL.

La realización de los dos clasificadores que fueron usados revela un porcentaje a la hora de distinguir entre pacientes sanos y enfermos lo bastante buena como para alentar el desarrollo de nuevos estudios sobre estos parámetros prosódicos.

A forma de conclusión, los resultados obtenidos animan a un estudio más profundo de los rasgos prosódicos como una herramienta sencilla y apenas molesta para la detección precoz del DCL

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- Se realizó un estudio de los métodos descritos en la literatura para el análisis computacional de la voz en la identificación de deterioro cognitivo leve.
- Se diseñó y desarrolló una base de datos de voces de pacientes con deterioro cognitivo leve y sanos en el Hogar del Adulto Mayor #2 de Santa Clara.
- Se realizó la clasificación de los dos grupos (Sano y DCL) mediante diferentes técnicas de clasificación (LDA y RBF) empleando siete rasgos prosódicos del habla, los resultados alcanzados están al nivel de los reportados en la literatura para casos de estudios similares.

Recomendaciones

Para continuar desarrollando del tema de esta investigación se recomienda:

- Aumentar el número de muestras de la base de datos, realizando grabaciones en otros Hogares del Adulto Mayor de la Ciudad de Santa Clara.
- Aumentar el número de rasgos (prosódicos y lingüísticos) para realizar una distinción más profunda entre los dos grupos estudiados en este trabajo.
- Crear una herramienta automática para identificar deterioro cognitivo leve mediante el procesamiento digital de la voz.

REFERENCIAS BIBLIOGRÁFICAS

[1] Eurostat. Europe in figures, Eurostat Yearbook 2006-2007, 2007

[2] Barrero-Hernández FJ, Vives-Montero F, Morales-Gordo B. Evaluación de la versión española del Memory Impairment Screen. Rev Neurol, 43: 15-9, 2006.

[3] Kalaria R, Maestre G, Arizaga R, Friedland R, Galasko D, Hall K, et al. Alzheimer's disease and vascular dementia in developing countries: prevalence, management, and risk factors. Lancet, 7: 812-26, 2008.

[4] Molero A, Pino G, Maestre G. High prevalence of dementia in Caribbean population. Neuroepidemiology ; 29: 107-12,2007.

[5] Llibre J, Férnandez Y, Marcheco B, Contreras N, López A, Otero M, et al. Prevalence of dementia and Alzheimer's disease in a Havana municipality: a community-based study among elderly residents. MEDICC Review ; 11: 29-35,2009.

[6] Herrera E, Caramelli P, Barreiros A, Nitrini R. Epidemiologic survey of dementia in a community dwelling Brazilian population. Alzheimer Dis Assoc Disord ; 16: 241-6,2002.

[7] Nitrini R, Bottino C, Albala C, Custodio N, Ketzoian C, Llibre J, et al. Prevalence of dementia in Latin America: a collaborative study of population-based cohorts. Int Psychogeriatr ; 21: 622-30,2009.

[8] Conde-SalaJL, Garre-Olmo J, Vilalta-Franch J, Llinàs-Reglà J,Turró-Garriga O, Lozano-Gallego M, et al. Predictors of cognitive decline in Alzheimer's disease and

mild cognitive impairment using the CAMCOG: a five-year follow-up. Int Psychogeriatr ; 26: 1-11,2012**.**

[9] Folstein M, Folstein S, McHugh P. Minimental State. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res ,12: 189-98,1975.

[10] Petersen RC, Smith GE, Waring SC, Ivanik RJ, Tangelos E, Kokmen E. Mild cognitive impairment: clinical charac-terization and outcome. Arch Neurol; 56 (3): 303-8, 1999

[11] Winblad B, Palmer K, Kivipeto M, Jelic V, Fratiglioni L, Wahlund LO, et al. Mild cognitive impairment-beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. J Inter n Med ; 256: 240-6, 2004.

[12] Pérez-Martínez DA, Baztán JJ, González-Becerra M, Socorro A. Evaluación de la utilidad diagnóstica de una adaptación española del Memory Impairment Screende Buschke para detectar demencia y deterioro cognitivo. Rev Neurol ; 40: 644-8,2005.

[13] Wiedl KH, Wienöbst J, Schöttke H. Estimating rehabilitation potential in schizophrenic subjects. In Brenner HD, Boker W, Gennes R, eds. The treatment of schizophrenia: status and emerging trends. Bern: Hogrefe & Hunber, p. 88-103, 1999.

[14] Spreen O, Strauss E. A compendium of neuropsychological test. New York: Oxford University Press; 1991

[15] Oakhill J, Yuill N, Parkin A. Working memory, comprehension ability and the resolution of text anomaly. Br J Psychol; 80: 351-61,1989.

[16] Calero MD, Salguero A. TAS: tarea de atención sostenida. Granada: Sinder; 2009.

[17] Pakhomov, S., Smith, G., Chacon, D. Computerized analysis of speech and language to identify psycholinguistic correlates of frontotemporal lobar degeneration, 2010.

[18] Roark, B., Hosom, J., Mitchell, M. Automatically derived spoken language markers for detecting mild cognitive impairment, 2007.

[19] Roark, B., Mitchell, M., & Hollingshead, K. Syntactic complexity measures for detecting mild cognitive impairment, .2007.

[20] Serguei VS Pakhomov,Computerized analysis of speech and language to identify psycholinguistic correlates of fronto-temporal lobar degeneration. 2010, recogido en National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/

[21] Lehr M, Prud'hommeaux E, Shafran I. y Roark B., Fully Automated Neuropsychological Assessment for Detecting Mild Cognitive Impairment. Interspeech, Décimo-tercera Conferencia Anual del International Speech,Communication Association en Portland, septiembre de 2012.

[22] Ayuntamiento de Madrid, Salud. Centro de Prevención de Deterioro Cognitivo. Preguntas Frecuentes: http://www.madrid.es/

[23] Darley, F.L., Aronson, A. E., & Brown, J. R., "Motor speech disorders", 3rd ed. 1975.

[24] Kelman, A.Y.G.a.l., "A Milt Gross Comic Reader", pp. 1–54, 2010.

[25] Órgano Oficial de expresión de la Fundación OMIE, Avances en Salud Mental Relacional. Revista Internacional On-line, vol.3, 2004.

[26] "Grupo de Aplicaciones de Procesado de Señales", www.gaps.ssr.upm.es, revisado 6th diciembre, 2013.

[27] Boza, D.T., "Manual de Sistema de Grabación para el diagnóstico de Deterioro Cognitivo DCGrab v3.0", Universidad de Santa Clara, 2013.

[28] Boersma P y Weenink D, del Instituto de Ciencias Fonéticas de la Universidad de Ámsterdam. Praat: doing phonetics by computing: http://www.fon.hum.uva.nl/praat/

[29] Mertens P. Prosogram, v2.9, Transcription of prosody using pitch contour stylization based on a tonal perception model and automatic or annotation-based segmentation: http://bach.arts.kuleuven.be/pmertens/prosogram/