
Universidad Central “Marta Abreu” de Las Villas
Facultad de Ingeniería Eléctrica
Departamento Biomédica



TRABAJO DE DIPLOMA

*“Análisis Computacional de la Voz para Identificar
Deterioro Cognitivo Moderado”*

Autor: Alex Romero Baluja.

Tutor: Dr.C. Eduardo Gonzalez Moreira.

Santa Clara
2015
"Año 57 de la Revolución"

Universidad Central “Marta Abreu” de Las Villas
Facultad de Ingeniería Eléctrica
Departamento Biomédica



TRABAJO DE DIPLOMA

“Análisis Computacional de la Voz para Identificar Deterioro Cognitivo Moderado”

Autor: Alex Romero Baluja.

e-mail: aromero@uclv.edu.cu

Tutor: Dr.C. Eduardo Gonzalez Moreira

e-mail: moreira@uclv.edu.cu

Santa Clara
2015
"Año 57 de la Revolución"



Hago constar que el presente proyecto de 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 Biomédica, 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 Autor

Firma del Jefe de Departamento
donde se defiende el trabajo

Firma del Responsable de
Información Científico-Técnica

PENSAMIENTO

“El lenguaje ha de ser matemático, geométrico, escultórico. La idea ha de encajar exactamente en la frase, tan exactamente que no pueda quitarse nada de la frase sin quitar eso mismo de la idea.”

José Martí

DEDICATORIA

- **A mi familia por ser el soporte y la guía en mi vida.**
- **A mis amistades por su apoyo y mantenerse siempre a mi lado.**

.

AGRADECIMIENTOS

- **A mis padres y familia por apoyarme en todos mis años de vida y formación como profesional.**
- **A mi tutor Eduardo González Moreira por el tiempo que me dedicó, sus sabios consejos en mi investigación y mi formación como ingeniero.**
- **A todos mis compañeros de aula incluyendo a los que por cuestiones de la vida no pueden estar con nosotros pero que influyeron en mi vida universitaria.**
- **A aquellos profesores que participaron en mi formación como profesional.**
- **A todas las amistades y personas que han influido en mi vida.**
- **A mi novia por apoyarme y estar presente en el transcurso de mis estudios.**

TAREA TÉCNICA

- ❖ Realización de un análisis crítico del estado del arte acerca de estudios realizados sobre mediciones de deterioro cognitivo moderado mediante el procesamiento digital de voz.
- ❖ Grabación mediante sistemas de medición a pacientes con deterioro cognitivo moderado y participantes sanos.
- ❖ Diseño de una base de casos de pacientes con deterioro cognitivo moderado y participantes sanos.
- ❖ Obtención de rasgos prosódicos en el dominio espectral para la clasificación de deterioro cognitivo moderado.
- ❖ Evaluación de los resultados obtenidos para la identificación de deterioro cognitivo moderado y algoritmos clasificadores empleados en la clasificación binaria.
- ❖ Redacción del informe.

Firma del Autor

Firma del Tutor

RESUMEN

Estudios realizados en los últimos años indican la importancia de las características del habla para diferenciar las personas sanas de enfermas con demencia. El presente trabajo se ha centrado en el estudio de técnicas que permitan la detección temprana de la demencia moderada 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 demencia y otro con sujetos sanos. El objetivo, pues, 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.

En primer lugar, se realiza una revisión del estado del arte relacionada con la detección de demencia mediante el procesamiento digital de la voz. Esta revisión permitió conocer las pruebas con mejores resultados.

Posteriormente se procedió a la creación de la base de datos para luego realizar la extracción de los parámetros de cada grupo y una comparativa de la media resultante en cada uno de los mismos, con el objetivo de observar las posibles diferencias existentes.

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 varias opciones de un clasificador para comprobar la eficiencia discriminativa de los rasgos.

TABLA DE CONTENIDO

Contenido

PENSAMIENTO.....	I
DEDICATORIA.....	II
AGRADECIMIENTOS	III
RESUMEN	V
INTRODUCCIÓN.....	VIII
CAPÍTULO 1. FUNDAMENTOS TEÓRICOS SOBRE LA DEMENCIA Y SU RELACIÓN CON EL HABLA.....	1
1.1 Definición de demencia.	1
1.2 Diagnóstico de deterioro cognitivo moderado.....	3
1.3 Prevalencia e incidencia en las demencias	5
1.4 Análisis perceptivo del habla	7
1.5 Análisis instrumental del habla	8
1.6 Estudios relacionados con la detección instrumental de la demencia moderada	9
1.7 Consideraciones finales del capítulo 1.....	13
CAPÍTULO 2. ANÁLISIS COMPUTACIONAL PARA LA MEDICIÓN DE DETERIORO COGNITIVO MODERADO	15
2.1 Creación de la base de datos de Deterioro Cognitivo Moderado.....	16
2.1.1 Pasos a seguir en el proceso de grabación.....	16
2.1.2 Corpus de voz	22
2.1.3 Software DCGrab v3.0.....	24
2.2 Extracción de rasgos de la base de datos de Deterioro Cognitivo Moderado	27
2.2.1 Análisis prosódico de la voz	27
2.2.2 Herramientas Praat, Prosogram y Matlab.....	28
2.2.3 Base de Casos de DCM.....	30
2.3 Rasgos Prosódicos para la Caracterización de la Demencia	31
2.4 Clasificación por Máquinas de Soportes Vectorial	32
2.4.1 Máquinas de soporte vectorial para clasificación binaria.....	33

CAPÍTULO 3. RESULTADOS.....	36
3.1 Descripción estadística de los rasgos extraídos.....	36
3.2 Obtención de rasgos significativos	38
3.2.1 Prueba de Wilcoxon.....	38
3.3 Clasificación de los pacientes	39
3.4 Clasificación de rasgos prosódicos empleando SVM	41
3.4.1 Clasificación de los rasgos prosódicos según su significancia	41
3.4.2 Clasificación de los rasgos prosódicos según posibles combinaciones	43
3.5 Conclusiones del capítulo	45
CONCLUSIONES.....	46
RECOMENDACIONES.....	47
REFERENCIAS BIBLIOGRÁFICAS.....	48

INTRODUCCIÓN

INTRODUCCIÓN

En las últimas décadas, hemos sido testigos de impresionantes avances en el desarrollo social humano. Uno de los más importantes es el envejecimiento demográfico, el cual ocurre en todas las regiones del mundo, particularmente en los países de mediano y bajos ingresos. Se estima un incremento de la población mundial de 2.7 billones, es decir, de 6.5 billones de personas en el 2005, este número se incrementará a 9.2 billones en el 2050. Ello se acompañará de un dramático incremento en el número y la proporción de adultos mayores. El número de personas de 60 años y más en el mundo, se triplicará, desde 606 millones en el 2000 a 1.9 billones para el 2050 [7].

Esta transición demográfica, que ocurre particularmente, en América Latina, China y la India, es el resultado del declinar de las tasas de fecundidad y de reemplazo, la disminución de la mortalidad, y el aumento de la expectativa de vida. En la última mitad del siglo XX la expectativa de vida al nacer en América Latina y el Caribe se incrementó en aproximadamente 20 años, de 50 a 70 años, y la tasa de fecundidad decreció en la mitad, aproximadamente de 6 a menos de 3 niños [7].

Paralelamente con el envejecimiento demográfico, está aconteciendo un proceso conocido como resultado de los cambios demográficos, sociales y económicos en la población mundial, en los cuales las enfermedades infecciosas son gradualmente reemplazadas por las enfermedades crónicas no transmisibles como principal causa de muerte [7].

Las enfermedades crónicas no transmisibles constituyen en la actualidad la principal causa de muerte en todas las regiones del mundo, con excepción del África Sub – Sahariana. De los 35 millones de muertes en el año 2005 debido a enfermedades crónicas no transmisibles, 80% ocurrieron en países de bajo y mediano ingreso. Esto se debe en parte a que la inmensa mayoría de los adultos mayores viven en estas regiones - 60% hoy día y alcanzarán un 80% en el año 2050 [7].

A pesar del interés creciente en colocar las enfermedades crónicas en particular en las agendas y estrategias de salud nacional y global, las enfermedades mentales en general y las demencias en particular no constituyen una prioridad en la mayoría de los países en vías de desarrollo. Sin embargo, las demencias, son el mayor contribuyente de discapacidad, dependencia y mortalidad, en los ancianos. Cuba es un país en vías de desarrollo con indicadores de salud similares a los de países desarrollados y con un rápido envejecimiento de su población total, con un 17,9 % de su

INTRODUCCIÓN

población que supera los 60 años y una esperanza de vida al nacer de 77.97 años, 76 años para los hombres y 80.02 para las mujeres y de 22.09 años a los 60 años y 8.8 años más a los 80 años. Se estima que para el año 2020 Cuba se convertirá en el país de Latinoamérica con mayor proporción de adultos mayores, el 25% de la población con 60 años y más de edad. De acuerdo con estimados recientes y basados en revisiones sistemáticas de datos de prevalencia y un consenso de expertos, se calculan 36 millones de personas que sufren de demencia en todo el mundo, con 4.6 millones de nuevos casos anualmente es similar a la incidencia anual de ictus no fatal.

Esta cifra se duplicará cada 20 años hasta alcanzar los 80 millones de personas con demencia en el año 2040 en todo el mundo, incremento que será más marcado en las regiones en vías de desarrollo que en las regiones desarrolladas.

El crecimiento en el número de personas con demencia entre el año 2000 y el 2020 será de un 120% en América Latina, por lo que sobrepasará al de cualquier otra región del mundo. De 2 millones de personas con demencia en América Latina en la actualidad la cifra se incrementará a 4.1 millones para el 2020 y 9.1 millones en el 2040, es decir será similar al de Norteamérica. En Cuba la prevalencia de síndrome demencial oscila entre 6.4 y 10.2 % en la población de 65 años y más, con un predominio en el sexo femenino, constituyendo la enfermedad de Alzheimer su causa más frecuente seguida por las demencias vasculares. Si tenemos en cuenta el acelerado envejecimiento de la población cubana se estima una cifra de 130 000 personas con enfermedad de Alzheimer u otro tipo de demencia, cifra que debe duplicarse para el año 2020. Sin una intervención efectiva, es decir si no se logra una cura para esta enfermedad en los próximos años, el número de cubanos con demencia se incrementará en 2.3 veces para el 2040, es decir a 300 000 personas con demencia, lo que significa el 2.7 % de la población cubana. En los próximos 30 años, la demanda de cuidados a largo plazo para personas con demencia se incrementará en 10 veces las necesidades actuales. De acuerdo con estos resultados estimamos en 28 750 nuevos casos de demencia por año en Cuba. De mantenerse estas proyecciones, el número de casos nuevos de 65 años y más con demencia se incrementará en 2.5 veces para el 2040, es decir, de 28 670 casos nuevos por año o un caso nuevo cada 5 minutos a 71675 casos nuevos por año o un caso nuevo cada 3 minutos en el 2040.

INTRODUCCIÓN

En la mayoría de los estudios antiguos se han utilizados técnicas hechas por psicólogos profesionales donde se hacen cuestionarios, para hacer apuntaciones específicos para poder diagnosticar si una persona exhibe síntomas de la enfermedad. Sin embargo estos métodos no son muy precisos ni eficientes, dado que ellos son sometidos también a observaciones visuales de un observador (psicólogo), por lo tanto puede resultar como un diagnóstico prejuiciado e incorrecto. Por esta razón avances hoy en día han tratado de mejorar el diagnóstico para que estas no se limiten a resultados por observaciones de personal capacitado. Así que nuestro trabajo es implementar una técnica más fiel y concreto para analizar dicha enfermedad, al fin de poder detectar los síntomas tempranos.

Estudios han mostrado que una de las áreas cognitivas más afectadas por el deterioro cognitivo es el lenguaje. La demencia afecta el lenguaje a dos niveles: nivel lingüístico, contenido verbal y el nivel paralingüístico, como una persona habla. Al nivel lingüístico, en la mayoría de los casos, existe una disminución en la fluidez semántica y fonológica, y dificultades de la denominación de objetos.

Además, se han observado que los casos de demencia de tipo Alzheimer son acompañados de una pérdida de la capacidad oral y escrita para comunicarse, Afasia. Errores en el procesamiento semántico, un vocabulario poco profundo y dificultades a la hora de encontrar palabras son algunos síntomas de la afasia, que a su vez llevan a un deterioro del habla espontánea.

La naturaleza progresiva, y en la mayoría de casos, la evolución continua del deterioro cognitivo hace difícil una diferenciación entre la normalidad, con la pérdida irremediable de facultades debidas a la edad, el deterioro cognitivo y la demencia. Por ello se fijan dos objetivos, primero detectar si hay deterioro cognitivo o no, y segundo discriminar la gravedad entre normalidad, deterioro cognitivo o demencia.

La ocurrencia de esta enfermedad ha conducido a efectos desfavorables en la sociedad, afecta tanto a nivel social como familiar. Además, el número de casos seguirá aumentando acorde incrementamos también la esperanza de vida. Por ello, cada día se aumentan las investigaciones destinadas a hallar curas a sus síntomas, pese a que todavía no se ha logrado identificar su causa.

INTRODUCCIÓN

Debido a la naturaleza degenerativa de la enfermedad de Alzheimer, cobra una importancia primordial su detección precoz. En etapas tempranas se puede optar a terapias más eficaces, logrando controlar los síntomas de la enfermedad y ralentizar su avance.

Es importante hablar del debate que existe en el diagnóstico precoz del Alzheimer. Es un hecho consensuado que el diagnóstico temprano de la enfermedad hace que pueda ser llevada mejor, viéndose disminuidos sus síntomas cuanto antes sean tratados. Pero puesto que no existe un tratamiento que permita curar al enfermo, y el conocimiento de la enfermedad provoca gran angustia tanto al paciente como a la familia, existen partidarios de evitar esta detección precoz; afirmando que puesto que ésta no llegará a curar al paciente, es mejor evitar el sufrimiento que conlleva su conocimiento temprano. Pese a estas afirmaciones, si se realiza una evaluación objetiva de los pros y contras del diagnóstico temprano, se puede concluir que resulta más beneficioso que perjudicial.

Por tanto, con el fin de lograr dar solución al problema existente nos hemos planteado los siguientes objetivos para la presente investigación:

Objetivo General:

- Desarrollar un análisis computacional de la voz para identificar Deterioro Cognitivo Moderado (DCM).

Objetivos Específicos:

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

INTRODUCCIÓN

Organización del informe:

El informe del trabajo está estructurado de la siguiente manera: introducción, capitulo, 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 deterioro cognitivo. 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 Deterioro Cognitivo Moderado 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. FUNDAMENTOS TEÓRICOS SOBRE LA DEMENCIA Y SU RELACIÓN CON EL HABLA

En el presente capítulo se realiza la fundamentación teórica sobre la demencia y su relación con la voz. Igualmente, se ofrece una pequeña introducción al análisis prosódico de la voz. Y por último, se analizan diferentes estudios realizados en el mundo, relacionados con la medición del deterioro cognitivo, en los cuales se destaca su utilidad en el análisis de los resultados obtenidos y las diferencias entre ellos.

1.1 Definición de demencia.

La palabra demencia, etimológicamente derivada del latín demens (de: fuera, mens: espíritu, mente) indica un debilitamiento progresivo, más o menos global, de las capacidades intelectuales y afectivas, grave y definitivo [1].

"Demencia es una afección ordinariamente sin fiebre y crónica, caracterizada por un debilitamiento de la sensibilidad, la inteligencia y la voluntad, con incoherencia de ideas, defecto de la espontaneidad intelectual y moral y tales son los signos de esta afección. El hombre que está en la demencia ha perdido la facultad de percibir espontáneamente los objetos, captar sus relaciones, comparar, preservar el recuerdo por completo, de lo que resulta la imposibilidad de razonar con éxito"[1]

Durante el siglo XIX, el término de demencia empieza a ser usado para referirse principalmente a cuadros de trastornos cognitivos adquiridos y en la segunda mitad del siglo XIX se restringe casi exclusivamente a trastornos irreversibles que afectan

predominantemente a los ancianos. Se mantuvo una visión sindrómica de las demencias, incluyendo entre otros las demencias vesánicas, es decir, estados terminales de diferentes enfermedades mentales. Esquirol define las demencias como un estado de pérdida de la capacidad de razonar y diferencia las demencias agudas de las crónicas y las seniles. Las demencias agudas equivalen a un estado confusional, mientras que las seniles son el resultado de la edad y consisten en una pérdida de la capacidad de entendimiento, las demencias crónicas son irreversibles y causadas por múltiples causas, tales como manías, estados de melancolías, accidentes vasculares y epilepsia [2].

Es una enfermedad degenerativa crónica del sistema nervioso central, con una elevada prevalencia y asociada a una elevada morbilidad. Es preciso descartar siempre causas secundarias y tratables. Un cuadro de confusión o desorientación en las personas de edad avanzadas puede ser debido a la descompensación de una neupatía obstructiva crónica, de una cardiopatía o de una insuficiencia hepática o renal, a una depresión, a la ingesta de ciertos fármacos como antihistamínicos, antipsicóticos, anticolinérgicos, hipnosedantes, antiinflamatorios, diuréticos, corticoides, hipoglucemiantes, a un proceso infeccioso o a una alteración hidroeléctrica, deshidratación en caso de infección o inducida por diuréticos. Es preciso por tanto, buscar la causa y tratarla.

Se caracteriza por un deterioro de las capacidades cognitivas y se manifiesta por la pérdida de memoria y de capacidad de razonamiento y la alteración del lenguaje. La demencia más frecuente es la enfermedad de Alzheimer, cuyo diagnóstico clínico se realiza sobre la base de los resultados de tests psicométricos, pero el diagnóstico de certeza solo es anatomopatológico. Aunque a veces es difícil, hay que distinguir la enfermedad de Alzheimer de deterioro cognitivo propio del envejecimiento y de la pseudodemencia [3].

En nuestros días aparece como grave problema en el área de salud, y sobre todo en los países más avanzados, el aumento vertiginoso de casos de sujetos afectados por una enfermedad terriblemente discapacitante como es la demencia. Es una enfermedad ligada a la edad, por lo que uno de los factores que han contribuido a incrementar su incidencia y prevalencia en las últimas décadas se deriva del envejecimiento progresivo de la población. La demencia es

un síndrome que ha de entenderse como un declinar de las funciones superiores, entre ellas la memoria, en relación con el nivel previo de los pacientes. Adicionalmente, completando el cuadro sintomático, se suelen sumar alteraciones psicológicas y del comportamiento, produciendo en el sujeto una discapacidad progresiva [4].

La demencia es una de las enfermedades más temidas y caras de la sociedad actual. Es un síndrome clínico de deterioro cognitivo adquirido que determina disminución de la capacidad intelectual suficiente como para interferir en el funcionamiento social y funcional del individuo y en su calidad de vida. Muchas patologías pueden causar demencia, y algunas de ellas pueden ser reversible.

1.2 Diagnóstico de deterioro cognitivo moderado

El diagnóstico en el área de salud es de gran utilidad en nuestros días para la detección de enfermedades tanto clínicas como psicológicas y con el fin de encontrar patologías en pacientes antes que estas puedan llegar a ser graves o irreversibles. Es por ello la importancia que tiene el diagnóstico de deterioro cognitivo moderado.

El deterioro cognitivo moderado es cuando condiciona incapacidad para vivir de manera independiente. El individuo no recuerda información básica acerca de su vivienda, actividades recientes o el nombre de ciertas personas familiares a él.

Dada la importancia que tiene la detección temprana de síntomas que indiquen un posible deterioro cognitivo, es importante elegir adecuadamente qué prueba o conjuntos de pruebas debemos utilizar. En la actualidad, se siguen utilizando herramientas de evaluación que fueron creadas a partir de la conceptualización que anteriormente se hacía del DCM y en las que, por tanto, el peso que se atribuye al rastreo de todas las funciones cognitivas superiores no es equitativo. Gran parte de los tests evalúan más profundamente aspectos relacionados con la memoria o la orientación temporoespacial, minimizando la exploración de funciones superiores, como podrían ser el lenguaje o las funciones ejecutivas. Es prioritario no perder la ocasión, cuando existe la sospecha, para aplicar las pruebas de rastreo cognitivo que permiten, con mayor o menor eficacia, detectar la presencia de alteraciones cognitivas.

Aunque el diagnóstico siempre va a conllevar una evaluación más exhaustiva y pormenorizada de las funciones mentales superiores, a ser posible por parte de un especialista, la aplicación de pruebas de rastreo cognitivo puede ayudar a detectar síntomas de forma temprana y así poder iniciar un tratamiento completo para el paciente.

Puede darse la situación de que se sospeche que el paciente presente alguna afectación y no se disponga de ningún test que facilite la detección del DCM en la consulta o, especialmente, en una visita en el domicilio del paciente o en una atención en urgencias. En estos casos, y de forma excepcional, se podría aplicar alguna técnica informal. Por ejemplo, si sospechamos que el paciente tiene un problema cognitivo, podríamos pedirle identificar un objeto, esto mostraría su capacidad de reconocimiento y de lenguaje, luego al final de la consulta pedirle que recuerde el objeto mostrado, observando la capacidad de la memoria. Sin embargo, esto debería aplicarse en situaciones puntuales y siempre contrastadas con una posterior evaluación estandarizada y completa, aplicando las pruebas de rastreo seleccionadas, que permita objetivar la posible alteración.

Una de las pruebas utilizadas es el MMSE de Folstein (por las siglas en inglés de *Mini-mental State Examination*) que tiene una sensibilidad de 90% y una especificidad de 75% para detectar deterioro cognitivo. La prueba es un recurso adecuado para la evaluación temprana del deterioro cognitivo. Tiene una puntuación máxima de 30 puntos, que valora diferentes dominios como son orientación, registro o fijación, atención y cálculo, memoria reciente, lenguaje y construcción visual. El resultado deberá ser controlado por la escolaridad de los sujetos. Con una media para normalidad de 24 puntos o más para individuos de 5 a 8 años de escolaridad y de 22 puntos para los que tienen de 0 a 4 años de escolaridad. En personas con más de 8 años de escolaridad el MMSE establece como deterioro cognitivo

- Sin deterioro: 24 puntos o más.
- Leve: entre 19 y 23 puntos.
- Moderado: entre 14 y 18 puntos.
- Grave: menos de 14.

Para que el MMSE tenga una adecuada confiabilidad es necesario que el paciente no esté cursando con delirium, alteraciones metabólicas o comorbilidad que pudiera alterar su estado de alerta o percepción sensorial [5]

Otra prueba utilizada es la del dibujado del reloj, es una herramienta bien establecida y válida debido a su rápida y fácil aplicación, se lleva 2 minutos aproximadamente, además tiene una buena aceptación en los pacientes de edad avanzada. El instrumento de escrutinio o prueba del dibujado del reloj evalúa la función ejecutiva, construcción visuo-espacial y habilidades cognitivas, ha sido validado en muestras comunitarias extensas para diferenciar pacientes con demencia de sanos, los valores de sensibilidad son de 80% y especificidad de 60%. La evaluación demuestra la funcionalidad de lóbulo central y temporo-parietal, se requiere diferentes habilidades cognitivas incluyendo la comprensión auditiva y visual, alteraciones visuo-espaciales, abstracción y función ejecutiva [5].

1.3 Prevalencia e incidencia en las demencias

La población mundial está envejeciendo aceleradamente, lo que se debe en gran parte a la mejoría en la atención de la salud durante el último siglo, traducida en vidas más largas y saludables. Sin embargo, este logro también ha tenido como resultado un aumento en el número de personas con enfermedades no transmisibles, incluida la demencia. Las proyecciones de prevalencia e incidencia indican que el número de personas con demencia continuará creciendo particularmente entre los más ancianos y que los países en proceso de transición demográfica son los que experimentarán el mayor aumento de casos. Se ha estimado que, a nivel global, en 2010 vivían 35.6 millones de personas con demencia, y se prevé que esta cifra se duplicará cada 20 años, alcanzando 65.7 millones en 2030 y 115.4 millones en 2050. Anualmente, el número total de nuevos casos de demencia en el mundo es de casi 7.7 millones, lo que significa un nuevo caso cada cuatro segundos[6].

En 2010, el costo total estimado de la demencia fue de 604 mil millones de dólares al año. En los países de ingresos altos, la atención informal (45%) y la atención formal (40%) representan la mayor parte del costo, mientras que la contribución proporcional del costo médico directo (15%) es mucho menor. En cambio, en los países de ingresos bajos y medianos, los costos sociales directos son pequeños mientras que predominan los costos de los cuidados informales [6].

En Cuba la prevalencia de síndrome demencial oscila entre 6.4 y 10.2 % en la población de 65 años y más, con un predominio en el sexo femenino, constituyendo la enfermedad de Alzheimer su causa más frecuente seguida por las demencias vasculares. Si tenemos en cuenta el acelerado envejecimiento de la población cubana se estima una cifra de 130 000 personas con enfermedad de Alzheimer u otro tipo de demencia, cifra que debe duplicarse para el año 2020. Sin una intervención efectiva, es decir si no se logra una cura para esta enfermedad en los próximos años, el número de cubanos con demencia se incrementará en 2.3 veces para el 2040, es decir a 300 000 personas con demencia, lo que significa el 2.7 % de la población cubana. En los próximos 30 años, la demanda de cuidados a largo plazo para personas con demencia se incrementará en 10 veces las necesidades actuales.

El estudio longitudinal prospectivo sobre envejecimiento y Alzheimer realizado en Ciudad Habana y Matanzas, con un periodo de seguimiento de 4.1 años en personas de 65 años y más mostró una incidencia de demencia de 9.3 por 1000 personas/año y 21.2 por 1000 personas/año, asociándose la demencia a diversos factores de riesgo, en particular edad avanzada, ser portador de uno o dos alelos de la apolipoproteína E4 un marcador de susceptibilidad, signos de Parkinsonismo, menor nivel educacional y Deterioro Cognitivo Leve. La incidencia de demencia encontrada en nuestro estudio fue muy similar a la reportada por el Estudio Canadiense de Envejecimiento y Salud, el más extenso de los estudios longitudinales realizado en adultos de 65 años y más en el mundo, que reportó una tasa de incidencia anual de 21.8 por 1000 para las mujeres y 19.1 por 1000 para los hombres y ligeramente superior a la reportada por el Reino Unido en el Estudio de Envejecimiento y Función Cognitiva (CFAS).

En la actualidad, Cuba es el segundo país más envejecido de América Latina, con 18,6% de población adulta mayor y aumentará un 25% en el 2020, pronosticándose que 1 de cada 4 cubanos tendrá 60 años o más y medio millón de personas superará los 80 años.

La demencia es el mayor contribuyente de años vividos con discapacidad en las personas de 60 años o más, dando un 11,2%, tiene una proporción mayor que las demás enfermedades

como la cerebro vascular (9,5%), las enfermedades músculo esqueléticas (8,9%), la enfermedad cardiovascular (5,0%) y el cáncer (2,4%) [8].

1.4 Análisis perceptivo del habla

Análisis perceptivo del Habla (PAS, por las siglas en inglés de Perceptual Analysis of Speech). Se fundamenta en las percepciones de los médicos sobre lo que oyen en las voces de las personas con deterioro cognitivo de acuerdo a sus capacidades de escucha, formación, experiencia y conocimiento sobre la enfermedad evaluada.

En la literatura, la fiabilidad de los jueces o los médicos es evaluada usando intra e inter juicios. Intra juicio depende de las pruebas repetidas sobre una muestra de un juez e inter juicio depende de las comparaciones entre los jueces de una misma muestra.

Según el estudio de (Kent R.) él definió el análisis de la percepción como la primera opción del clínico en el tratamiento de trastornos del habla [9].

Además, los métodos de dicha percepción llevan ventajas fuertes de conveniencia, economía, y robustez. Hay que usar regularmente en la práctica clínica, el diagnóstico provisional, sin los beneficios de datos adicionales. A pesar de esas ventajas, el método de PAS es susceptible a una variedad de errores y prejuicios. La conciencia de estas amenazas a la validez y la fiabilidad es un paso principal en el uso efectivo y afinado de los métodos de dicha percepción.

Existen preocupaciones sobre los juicios de percepción del habla en la práctica clínica, dado que las suposiciones basadas sobre el uso de dichos juicios no siempre caractericen la realidad. Por lo tanto los problemas enfrentados por el análisis son:

- Los jueces no parecen tener definiciones equivalentes de las dimensiones a ser evaluadas. Este fenómeno es como consecuencia de que en las evaluaciones del habla, los especialistas no están de acuerdo sobre cuáles de los aspectos de voz son más importantes en la valoración normal y trastornos del habla. Por lo tanto no logran alcanzar un consenso fijo para el análisis de percepción de habla.

- Luego las valoraciones perceptivas de varios aspectos de voz no están correlacionadas.
- Entonces como resultado de lo indicado anteriormente, varios aspectos perceptivos no son evaluadas con fiabilidad uniforme.

En fin las diferencias entre jueces expertos son mayores de las diferencias necesarias para la clasificación diagnóstica o los efectos de intervención [10].

1.5 Análisis instrumental del habla

Se puede definir el análisis instrumental del habla como la realización de un análisis fisiológico o acústico con la ayuda de herramientas y algoritmos para medir algunos fenómenos en el proceso de producción del habla. Generalmente, cada instrumento o algoritmo es creado para ofrecer información sobre un aspecto particular de la producción del habla, como por ejemplo; movimientos estructurales, corriente y presión de aire generado en diversas partes del mecanismo del habla y actividades musculares.

Con respecto al mecanismo del habla, dicho análisis instrumental ha sido muy efectivo y ha mostrado muchas diversidades de herramientas y operaciones que están diseñadas para ofrecer información acerca de aspectos concretos de la producción del habla como por ejemplo las funciones de articulación, laringe, las vías respiratorias y el puerto velo-faríngeo [11].

Además, el análisis instrumental plantea un análisis acústico que se usa para relacionar medidas acústicas con relación a varios aspectos del motor control del habla. Dependiendo de la muestra de discurso recogida y de la naturaleza del análisis, pueden ser calculadas un número de diversas medidas acústicas. Por lo tanto, no hay un sistema estándar de medidas para una observación específica. La mayoría de los análisis acústicos reportados en la literatura centran su estudio sobre la frecuencia fundamental, energía, prosodia y articulación.

Estudios de análisis instrumental enseñan que dicho análisis prevalece en algunas de las limitaciones del análisis perceptivo.

- Proveer medidas objetivas en un amplio rango de diferentes parámetros del habla para evaluación a corto plazo, así como el seguimiento a largo plazo del aparato del habla.

Además puede ampliar la precisión del diagnóstico a través de una descripción más detallada de los rasgos anormales [11].

A pesar de las ventajas indicadas anteriormente, existen limitaciones en la aplicación del análisis en la práctica clínica.

- Existe falta de disponibilidad de equipos instrumentales en la práctica cotidiana.
- También existe una falta de coincidencia entre las mediciones directas del análisis instrumental y el análisis perceptivo hecho por el clínico. Esto se debe porque la técnica instrumental requiere más de dos observaciones para describir los rasgos perceptivos.

Además la interacción de los diferentes subsistemas afectados pueden afectar las mediciones instrumentales [12].

1.6 Estudios relacionados con la detección instrumental de la demencia moderada

El primer estudio desarrollado por el Doctor Serguei Pakhomov, Gleen E. Smith, Dustin Chacon, Yara Feliciano y otros investigadores de la Universidad de Minnesota que tuvo como objetivo evaluar el uso de un sistema semiautomático para la medida de características del habla y el lenguaje en pacientes con degeneración del lóbulo fronto-temporal, para la realización de esta investigación, participaron 38 personas a las que les fueron administradas una batería de test neuropsicológicos, cuyos resultados fueron evaluados por neurólogos de avalado prestigio.

La valoración psicolingüística computarizada se realiza basándose en la técnica de Reconocimiento Automático del Habla conocido por sus siglas en inglés ASR para el análisis del habla y del lenguaje. Pese a que la aplicación de estas técnicas a las pruebas psicométricas

fue innovadora, la investigación lo expone como una extensión natural de las capacidades permitidas por estas tecnologías.

En el estudio se procesaron las grabaciones de voz de pacientes realizando la parte de descripción de imagen Cookie-Theft del Boston Diagnostic Aphasia Examination.

El audio fue transcrito de forma manual, con anotaciones sobre el propio audio y mediante tecnologías ASR, alineando automáticamente transcripción y audio (para lo cual fue necesario el uso de modelos acústicos y de lenguaje). Esta alineación permitió una medida precisa y una cuantificación de la duración y frecuencia de los distintos niveles, enunciados (dos o más palabras coarticuladas), palabras y fonemas.

Basándose en la alineación entre audio, transcripciones manuales y anotaciones, se definieron 15 variables, como la pausa entre palabras, variación de la frecuencia fundamental, perplejidad del discurso, etc. Sobre estas variables se realizaron análisis de factores para examinar la relación entre subconjuntos, desestimando la utilidad de una variable. Finalmente, se realizaron análisis de varianza para identificar las características diferenciadoras de las distintas enfermedades degenerativas del lóbulo frontotemporal (FTLD).

Como resultados de la investigación, se obtuvo que las medidas entre pausa y palabra y entre nombre y pronombre podían utilizarse para discriminar entre las variantes de FTLD. Además, afirman que a partir del estudio de las 15 variables se pueden sugerir cuatro componentes (longitud, vacilación, contenido vacío y gramaticalidad) que definen las distintas variantes de enfermedades FTLD.

Las conclusiones del estudio fueron que el análisis semiautomático del lenguaje y el habla era una forma prometedora de aproximación a la valoración neuropsicológica, ofreciendo valorables contribuciones a las herramientas que utilizan los investigadores de demencias y otros desórdenes neurodegenerativos [13]

El segundo estudio realizado conjuntamente por Dilek Hakkay-Tür del centro International Computer Science Institute (ICSI), Berkeley, y por Dimitra Vergyri y Gokhan Tur del SRI

International, Speech Technology and Research Lab tiene como objetivo la usabilidad de métodos automáticos para la valoración del estado cognitivo a través de test de evaluación verbal en personas de edad avanzada, los métodos de valoración automática evaluados en la investigación podrían ser utilizados como pruebas de cribaje, frecuentes, no intrusivas y de bajo costo; y permitir la monitorización objetiva y longitudinal del estado cognitivo, que podría complementar a las visitas clínicas regulares, e incluso ser útil para la detección temprana de patologías asociadas con el lenguaje y la comunicación.

Para realizar el estudio se recopilaron grabaciones de 100 personas, en su mayoría mayores de 70 años realizando dos tipos de test. Uno de recuento de una historia (parte del subtest de memoria lógica del Wechsler Memory Scale, WMS), utilizado para la valoración funcional de memoria y lenguaje y otro test de descripción de una imagen (test estándar como instrumento de valoración de la Afasia, como el WAB test), utilizado para valorar la información contenida en el habla.

Esta investigación realiza su propuesta basándose en el hecho de que el criterio principal de valoración clínica para el estado cognitivo se basa en unidades de contenido semántico. Por tanto, el objetivo de esta valoración automática es detectar la cantidad de unidades de contenido semántico que el hablante menciona.

En los test de recuento de una historia se extrajeron 35 unidades de contenido semántico, y se evaluaban en las transcripciones del test las menciones que hacían los pacientes de estas unidades de contenido, obteniendo un punto por cada unidad mencionada.

Por otro lado, y debido a que no existe una descripción escrita de la imagen, en los test descriptivos se realizó una evaluación basándose en el número de unidades de información que los pacientes utilizaban al describir la imagen de una lista total de 36, divididas en cuatro categorías, sujetos, lugares, objetos y acciones.

Además, se destaca la naturaleza no trivial de la detección de las unidades de contenido debido a las variaciones lingüísticas que existen en el lenguaje natural. La alta correlación de

la valoración automática podría alcanzarse si las unidades puntuadas cubren todos los conceptos semánticos.

Para automatizar la puntuación de las menciones utilizaron técnicas de recuperación de información y métricas para el resumen del habla. Contabilizando las menciones, la precisión de éstas, además del análisis de unigramas y bigramas (frecuencia de aparición de una palabra en el texto o dos respectivamente), en el recuento de la historia y únicamente menciones en la descripción de imagen.

Puesto que existen numerosos factores que afectan la realización de estos test como el nivel educativo, el estado cognitivo o complementación de los test con el estado de salud del sujeto, el objetivo de la investigación fue únicamente probar la usabilidad de las tecnologías del habla para realizar estos test, mientras que la interpretación de los resultados serían realizadas por clínicos.

Finalmente, la investigación concluyó mostrando alta correlación entre los resultados obtenidos manualmente y los obtenidos con las métricas automáticas, utilizando cualquiera de las transcripciones del habla, manual o automática [14]

El tercer estudio desarrollado por el Doctor Serguei Pakhomov, Universidad de Minnesota, investigador también del estudio anterior, se basa en un nuevo método que utiliza tecnologías del habla para realizar una detección temprana de los cambios cognitivos, analizando información sobre la duración de las palabras y el contenido semántico de éstas, la motivación recae en que pese a que los test de fluidez verbal son muy utilizados para el diagnóstico de la Enfermedad de Alzheimer o E.A, y son fáciles de administrar, los métodos de puntuación estándar de estos test no detectan cambios en las fases muy tempranas de estados cognitivos previos a la demencia.

Se planea desarrollar un sistema especial de reconocimiento de voz que realizará transcripciones de la voz de una persona que padezca la E.A, midiendo además la duración de las palabras con gran precisión de forma que se pueda realizar un análisis de la vacilación,

entonación y el tono de voz, obteniendo de esta forma información sobre el estado cognitivo y emocional del paciente.

Para validar el instrumento se utilizarán los datos de un estudio clínico sobre envejecimiento con una muestra de 90 voluntarios de diferentes sectores que padecen la E.A. y DCL de la Memory Clinic de la Universidad de Minnesota. Su comportamiento se basará en la correlación de las medidas semánticas con marcadores neurodegenerativos obtenidos mediante fMRI.

El instrumento podría ser implementado en plataformas móviles y telefónicas, permitiendo una colección de voz remota, útil para realizar pruebas clínicas con una observación a largo plazo y desde el hogar, además de a gran escala. Podría ser un método adecuado para caracterizar los cambios cognitivos asociados con el riesgo de progreso a Alzheimer [15]

1.7 Consideraciones del capítulo

La demencia moderada tiene su nivel de complejidad de acuerdo a las características fundamentales que la forman. En varios estudios 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.

En la actualidad se trabaja con métodos donde se combina el análisis perceptual y objetivo como consecuencia de las limitaciones 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. ANÁLISIS COMPUTACIONAL PARA LA MEDICIÓN DE DETERIORO COGNITIVO MODERADO

El estudio expuesto a lo largo de capítulo se centrará en el análisis prosódico de la voz de un grupo de personas diagnosticadas con deterioro cognitivo moderado y de otro grupo que no padece dicha enfermedad, con el objeto de encontrar disparidades en determinadas características entre ambos que permitan diferenciar entre una persona sana y otra enferma sin más pruebas que el estudio de una señal de voz.

Estos resultados están destinados a un posible diagnóstico temprano de la demencia severa, la cual puede derivar en la enfermedad del Alzheimer (EdA). El sistema desarrollado para el análisis semiautomático del habla y del discurso se describe en la figura 2.1 que aparece a continuación.

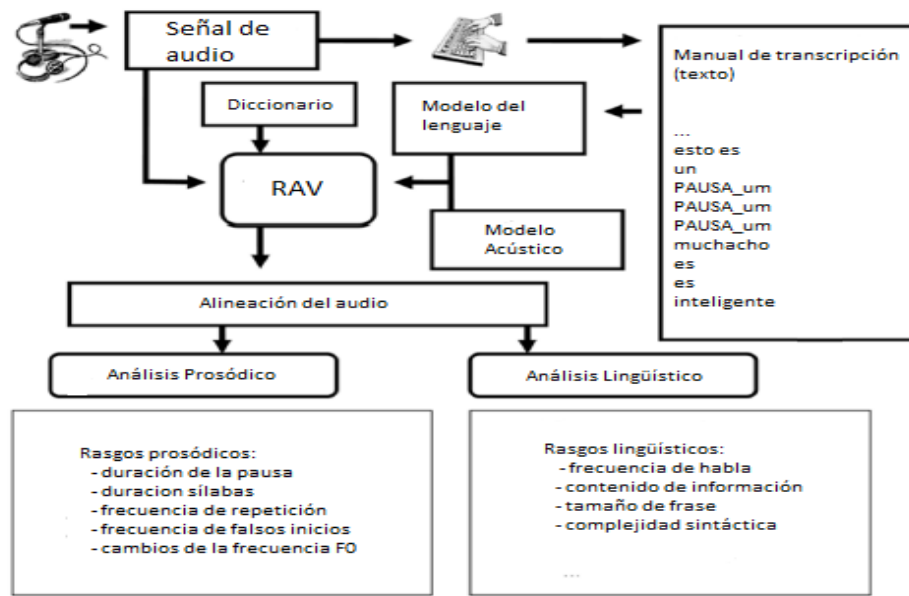


Figura 2.1 Diagrama de flujo que ilustra el funcionamiento del sistema de automatización para la evaluación del habla y el lenguaje

2.1 Creación de la base de datos de Deterioro Cognitivo Moderado

El análisis del habla de grabaciones de voces de personas, para su adecuada utilización en los estudios de detección de patologías, debe tener en cuenta la voz de pacientes que sufren de la enfermedad y la voz de pacientes sanas como grupo de control. A continuación proponemos los criterios generales para el diseño y especificación de la base de datos de pacientes con demencia y el grupo de control que fue grabada con el objeto de posibilitar el trabajo desarrollado en esta Tesis, relacionado con la búsqueda de rasgos para la detección automática mediante el análisis del habla de casos de DCM.

2.1.1 Pasos a seguir en el proceso de grabación

Para esta investigación se desarrolló un protocolo de grabación que siguió los siguientes pasos:

- Análisis del diálogo durante el MMSE
- Análisis de la descripción del dibujo Ladrón de Galletas

- Análisis del proceso de lectura, párrafo “El Abuelo”

El MMSE es un test psicométrico del diálogo entre clínico y paciente que puntúa sobre un máximo de 30 puntos, agrupando los ítems en 5 apartados, el punto de corte que establece la probable demencia se establece habitualmente en 24 puntos, pero depende de factores como el grado de estudios del paciente. Además el clínico obtendrá información como el grado orientación, memoria inmediata, denominación, el tiempo de respuesta del paciente, capacidad de comprensión de la pregunta, veces de repetición de la pregunta, ritmo de respuesta, etc., es decir, un análisis prosódico y lingüístico aplicado al campo de la conversación [16].

Al comprobar el grado de orientación se hacen al paciente preguntas como: ¿Cuál es la fecha de hoy? o ¿Dónde vive?, entre otras. Para la comprobación de la “Memoria inmediata” se intenta que el sujeto aprenda tres palabras, donde el clínico se las repite 3 veces para que el paciente se aprenda las palabras si le es posible, porque se le volverán a preguntar en el apartado de memoria diferida.

A continuación se presenta en la tabla 2.1 las diferentes preguntas y apuntaciones del MMSE.

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 va quedando? Detenga la prueba tras 5 sustracciones. (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

Después de la evaluación con el MMSE al paciente se le da una imagen llamada el 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 fijar las acciones sobre las que se puede actuar, como el agua cayendo al piso u el niño sobre el taburete inestable [17]. Por otra parte los pacientes que exhiben síntomas de deterioro cognitivo,

usualmente no se dan cuenta de estos puntos principales de la imagen. Aún más aquellos que experimentan dificultades a la hora de organizar los conceptos o a veces no se pueden encontrar las palabras apropiadas para expresarse.

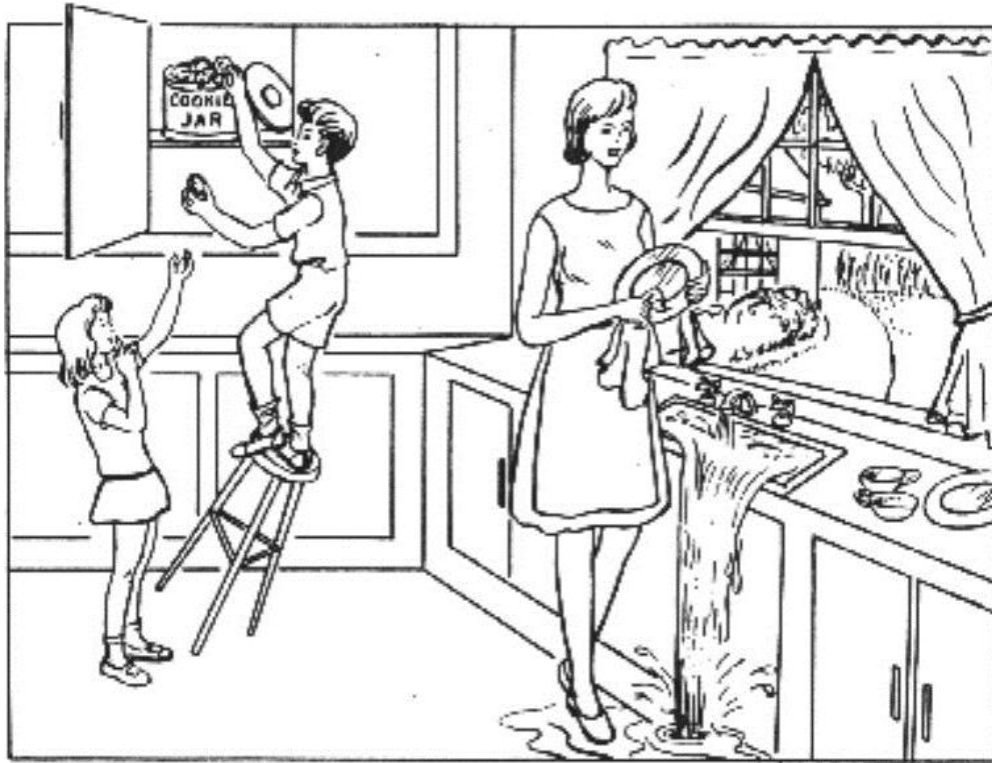


Figura 2.2 Ladrón de galletas.

Por último se realiza la lectura “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 él 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”. El abuelo le gusta ser moderno en su lenguaje.* 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. Además de sus aplicaciones de investigación y clínicas, “El Abuelo” es también empleado oficialmente por la Oficina de Medicina Aeroespacial de las Fuerzas Armadas de Estados Unidos como una herramienta cribadora para la inteligibilidad entre astronautas, pilotos, y los controladores de tráfico aéreo (U.S. Department of Transportation, 2008) [18] [19]

2.1.2 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 44kHz. 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.



Figura 2.3 Portátil ASUS A55V.

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 [20].

AKG Micrófono C520L Headworn Condenser: El micrófono Headworn Condenser tiene máximo rango dinámico para sonido vocal perfecto. Su cardioid patrón polar eficazmente rechaza ruido

Un choque de transductor reduce el ruido a un mínimo. El protector de humedad impide transpiración de penetrar el elemento del transductor, para asegurar un grado alto de protección de humedad para el micrófono.



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 es equipada con un arreglo de entradas versátiles para proveer los mejores resultados de cualquier tipo de fuente audio. Cada canal ofrece a una entrada XLR y un 1/4 una entrada balanceada. La entrada XLR es diseñado para trabajar con micrófonos u otras fuentes de $Lo Z$. El interruptor de poder da 48V a las entradas XLR para el uso con micrófonos profesionales del condensador. La entrada 1/4 " puede estar dispuesto a recibir señal del nivel de la línea.

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.

Medición de LED multicolor provee información visual instantánea de los niveles principales de salida.

El proceso de grabación fue controlado en todo momento por un experto, que cuidó que las señales de audio tuviesen un nivel de intensidad adecuado sin llegar a estar saturadas y que no tuviesen, en la medida de lo posible, 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.

2.1.3 Software DCGrab v3.0

Para el proceso de grabación se empleó el software DCGrab v3.0 desarrollado en el Centro de Estudio de Electrónica y Tecnologías de la Información (CEETI) para la presente investigación [21]. El DCGrab v3.0 es una herramienta que permite la grabación o la importación de los archivos de audio para el apoyo al diagnóstico de la Demencia a través de la voz.

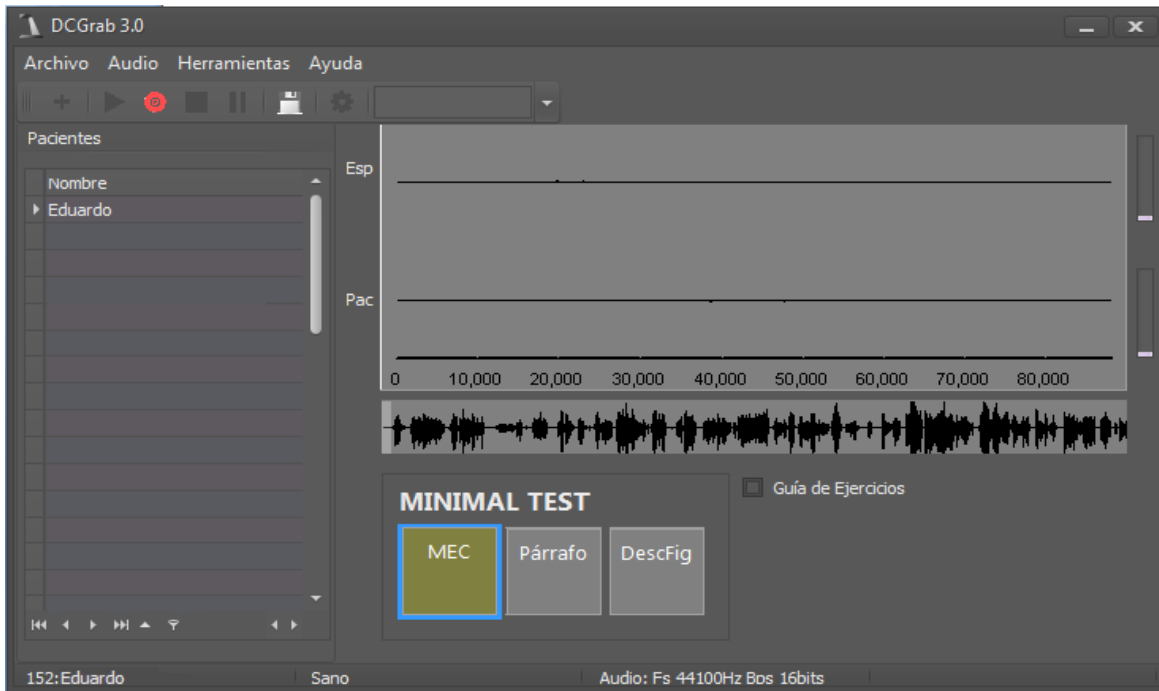


Figura 2.6 Ventana principal del software DCGrab v3.0

Este sistema fue programado de modo tal que se utiliza el Microsoft Access para crear una Base de Datos, el cual consiste de ciertas informaciones personales y valoraciones clínicas de los pacientes encuestados. Adicionalmente las señales de audio correspondientes a cada paciente por cada encuesta realizada son guardadas dentro archivos con el mismo sistema donde en su ventana principal, hay dos canales; uno para el especialista y el otro para el paciente. Además, cuando se añaden los pacientes, el nombre de cada paciente va a salir en la parte izquierda de la ventana principal, como se muestra en la Figura 2.6.

El “archivo” es estructurado de tal manera para realizar las funciones siguientes:

- **Agregar Paciente:** Mediante esta opción se agregan nuevos pacientes al sistema de base de datos. Es posible insertar varios pacientes o insertar solamente uno y comenzar a aplicar la encuesta (Figura 2.7).
- **Buscar Pacientes:** Mediante esta opción es posible buscar un sujeto en el sistema de base de datos ya sea mediante su número identificador o su nombre y activarlo en el sistema.

- Listar Paciente: Mediante esta opción es posible listar todos los sujetos que están incluidos en el sistema en un documento para imprimir o salvar la información en algún documento con formato conocido (.doc, .pdf, etc.)
- Salir: Cerrar DCGrab v3.0

Figura 2.7 Ventana de control de pacientes.

El “audio” contiene los controles para la manipulación de las señales de audio. Herramientas permite manipular la información de la Base de Datos y la configuración de los parámetros de audio.

La “ayuda” provee acceso a la ayuda del sistema.

En la parte inferior de dicha ventana principal, hay una barra de estado que contiene información acerca del sujeto activo en el sistema en el momento actual y la F0 con que se lo graba.

Finalmente para comenzar las grabaciones, selecciona una prueba del “MINIMENTAL TEST” de la misma ventana principal (Figura 2.6)

2.2 Extracción de rasgos de la base de datos de Deterioro Cognitivo Moderado

En la sección presente, se tratará automáticamente las evidencias en el habla de los locutores con DCM, y describir los métodos utilizados para extraer los datos prosódicos de la base de datos de DCM.

2.2.1 Análisis prosódico de la voz

Una de las áreas que se encuentra en la lingüística es el estudio de la prosodia que da información sobre los rasgos fónicos dado que influye la expresión oral, tales como la acentuación, la tonalidad o la cadencia. Dichos rasgos no son iguales en cada persona, por lo tanto se puede distinguir los rasgos de una persona a otra. Además existen factores tales como, la edad, nivel cultural, el estado físico e emocional de cada individuo que influyen en las características prosódicas.

Estudios modernos han presentado que personas que sufren ciertas enfermedades, como el síndrome de la apnea-hipopnea durante el sueño (o SAHS) o el Alzheimer, exhiben diferencias significativas en determinados rasgos fónicos y melódicos del habla con respecto a los de personas sanas [22] [23].

Para este trabajo se destacará el análisis prosódico de la voz de dos grupos de pacientes; pacientes que muestran signos de la enfermedad de deterioro cognitivo y otro grupo no diagnosticados de dicha enfermedad, de modo que se puedan comparar para tratar encontrar diferencias en características específicas entre ambos casos. Luego si hay diferencias entre ellos, nos permitan distinguir entre una persona sana y otra enferma sin la utilización de más pruebas, solo con el procesamiento de una señal de voz.

Estos resultados posiblemente van a ser útiles para la detección temprano del deterioro cognitivo, una fase clínica previa a la demencia, de modo que si el síntoma predominante es la pérdida gradual de memoria, puede deducir en la enfermedad del Alzheimer (EdA). Una detección temprana del

deterioro cognitivo antes de que evolucione en demencia, y su consiguiente tratamiento precoz, puede evitar su progresión y ofrecer a los pacientes mejor calidad de vida, además de reducir costes sanitarios [7].

2.2.2 Herramientas Praat, Prosogram y Matlab

A continuación se muestra un análisis prosódico de las voces grabadas, se utiliza una herramienta formada particularmente para este propósito: el programa Praat, diseñado y desarrollado para el análisis fonético del habla por Paul Boersma y David Weenink, del Instituto de Ciencias Fonéticas de la Universidad de Ámsterdam [24].

Para el análisis fonético del habla se utiliza la herramienta Praat. Esta herramienta tiene un lenguaje de programación propio, por tanto nos hace posible escribir código según las especificaciones que nos son necesarias y ejecutarlo por medio de esta herramienta. Para nuestro estudio, se empleará el código Prosogram v29f, desarrollado por Piet Mertens, profesor de lingüística francesa y computacional [11].

A continuación se describe brevemente el funcionamiento de estas herramientas complementarias para el seguimiento del análisis prosódico que se logra durante este trabajo.

El funcionamiento del Prosogram v29f es la implementación de transcripciones prosódicas de señales de audio a partir de unos parámetros específicos de entrada que puede ser definido por el usuario. El primer paso de la implementación por el programa es calcular parámetros acústicos tales como la frecuencia fundamental (F0), la intensidad, o los tramos en los que hay voz. Luego, se adquiere la segmentación de estos parámetros en unidades silábicas o vocálicas, seleccionando sólo las silábicas que son mayores de un cierto umbral de intensidad sonoro, y finalmente se estabiliza su F0 (Figura 2.8).

El resultado está representado en una ventana (figura 2.9) con un eje horizontal en el tiempo de estas unidades (a las que se denomina núcleos) contra las progresiones de F0 e intensidad. Además, es posible generar un archivo de texto en el que se incluyen múltiples datos prosódicos referentes a cada uno de los núcleos detectados.

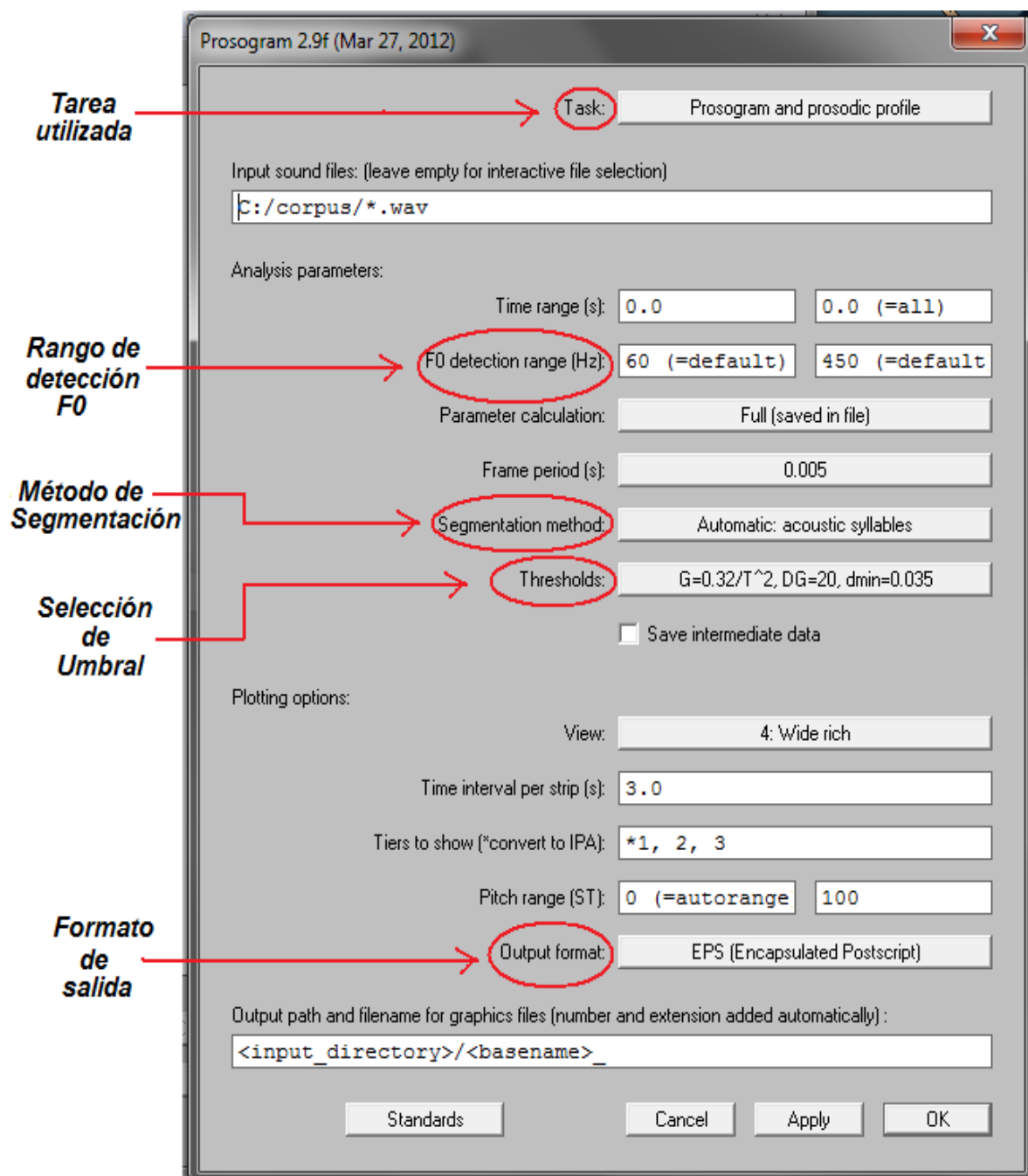


Figura 2.8 Ventana de selección de parámetros de Prosogram v29f

Los núcleos (sílabas) se representan mediante líneas gruesas en color negro, con una longitud equivalente a su duración temporal y orientación según la variación de F0 dentro del mismo. Las trazas en color verde y azul representan las progresiones de intensidad y F0, respectivamente.

Con el uso de la herramienta Matlab vs2013 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 v2013 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 DM y Control.

Adicionalmente, se ha implementado mediante este entorno un pequeño programa para extraer los datos prosódicos de la hoja de cálculo generada por la herramienta *Prosogram* v29f y representarlos de una forma visualmente más cómoda con sus respectivas etiquetas.

2.2.3 Base de Casos de DCM

Para analizar las características del habla nos basamos en grabaciones de voces pertenecientes a pacientes que padecen la enfermedad y a voces de personas sanas. En esta parte les mostramos criterios básicos de diseño y especificación de una base de datos de pacientes con DCM y pacientes sanos la cual fue obtenida para posibilitar el futuro trabajo desarrollado en esta tesis, el cual se basa en la búsqueda de rasgos para poder dar un diagnóstico a través del análisis del habla de casos con DCM.

Tabla 2.2 Base de casos de DCM

Rasgos		GRUPO	
		SANO	DCM
Cantidad de participantes		11	12
SEXO	Mujeres	18.18 %	50 %
	Hombres	81.82 %	50 %

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 2.2, consiste de 11 personas sanas de ellas 9 son hombres y 2 mujeres; 12 personas con DCM de ellas 6 hombres y 6 mujeres.

2.3 Rasgos Prosódicos para la Caracterización de la Demencia

Como ya hemos explicado que se obtiene los parámetros prosódicos que fueron definidos en la Sección 2.2 por la herramienta *Prosogram* v29f, y a partir de estos parámetros, mediante el MatLab se obtiene los rasgos necesario para la creación de la base de datos de deterioro cognitivo moderado. A continuación se muestran los cuatros rasgos prosódicos calculados a partir de los núcleos silábicos

- Frecuencia mínima: Mínimo valor del promedio de la frecuencia mínima y máxima por cada núcleo silábico.
- Frecuencia máxima: Máximo valor del promedio de la frecuencia mínima y máxima por cada núcleo silábico.
- Media de las frecuencias: Promedio de los valores obtenido de los promedios de la frecuencia mínima y máxima por cada núcleo silábico.

- Desviación estándar de las frecuencias: Desviación estándar de los valores obtenidos de los promedios de la frecuencia mínima y máxima por cada núcleo silábico.

2.4 Clasificación por Máquinas de Soportes Vectorial

La teoría de las Máquinas de Soporte Vectorial (SVM por su nombre en inglés Support Vector Machines) es una nueva técnica de clasificación y ha tomado mucha atención en años recientes. La teoría de la SVM está basada en la idea de minimización de riesgo estructural (SRM). En muchas aplicaciones, las SVM han mostrado tener gran desempeño, más que las máquinas de aprendizaje tradicional como las redes neuronales y han sido introducidas como herramientas poderosas para resolver problemas de clasificación [25].

Las máquinas de soporte vectorial surgieron como un método de clasificación basado en la teoría de minimización del riesgo estructural de Vapnik. En la actualidad, tienen numerosas aplicaciones debido a su versatilidad y a sus prestaciones. Las SVM se han utilizado con éxito en campos como la recuperación de información, la categorización de textos, el reconocimiento de escritura o la clasificación de imágenes.

Para poder clasificar con las máquinas de soporte vectorial, se comienza realizando una etapa de aprendizaje. Consiste en encontrar el hiperplano $h(x) = 0$ que mejor separe un conjunto de datos $X \in \mathcal{R}^d$ según la clase $Y \in \{-1, 1\}$ a la que pertenecen. Dicho hiperplano se corresponde con el que maximiza la distancia al punto más próximo de cada clase, por lo tanto, estará a la misma distancia de los ejemplos más cercanos entre ellos de cada categoría.

Según la teoría de Vapnik, el separador lineal que maximiza el margen (2 veces la distancia al punto más próximo de cada clase) es el que nos da la mayor capacidad de generalización, es decir, la capacidad de distinguir características comunes de los datos de cada clase que permitan clasificar imágenes que no sean las del conjunto de entrenamiento.

Para hallarlo, es necesario resolver un problema de optimización usando técnicas de programación cuadrática. A los datos que se utilizan para hallar la frontera de decisión (el hiperplano), se les conoce como vectores de entrenamiento o de aprendizaje.

A partir de unos datos de entrada x_i , las SVM nos proporcionarán su clase según la regla de clasificación $f(x_i) = \text{signo}(h(x_i))$ [26].

2.4.1 Máquinas de soporte vectorial para clasificación binaria

En los procesos de clasificación binaria sólo existen 2 clases: una es considerada como la positiva ($y = 1$) y la otra como la negativa ($y = -1$). Puede ocurrir que los datos no sean linealmente separables o que exista un cierto nivel de ruido. Según esto se pueden emplear distintos tipos de SVM

Tipos de SVM:

- SVM lineal con margen máximo
- SVM para la clasificación no lineal
- SVM con margen blando.

SVM lineal con margen máximo:

Sólo se debe emplear cuando los datos son linealmente separables, es decir, cuando se puede usar como frontera de decisión un hiperplano $h(x)$ tal que:

$$h(x) = w^T x + b = 0$$

Donde w y $x \in \mathbb{R}^d$, siendo d la dimensión del espacio de entrada.

SVM para la clasificación no lineal:

Puede ocurrir que los puntos no sean linealmente separables en el espacio de entrada, como se muestra en el ejemplo de la Figura 2.10.

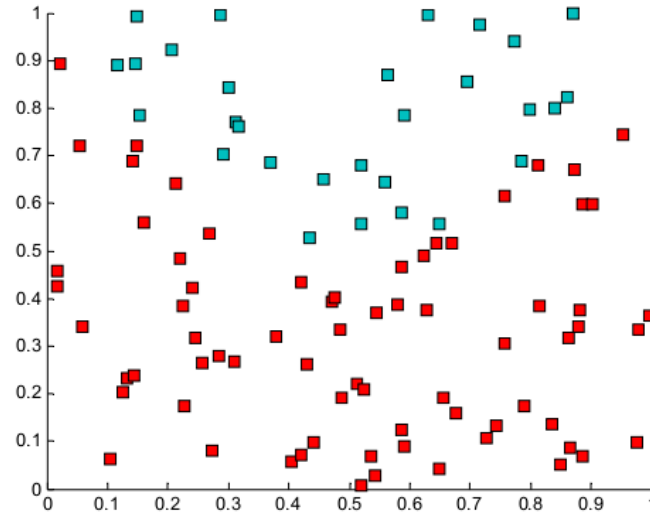


Figura 2.10 Ejemplo de un conjunto de datos no linealmente separables.

Cuando esto sucede, existe la posibilidad de transformar los datos a un espacio $\mathfrak{S}$ de mayor dimensión (el espacio de características) en el que los puntos sí pueden ser separados por un hiperplano (Figura 2.11). Para ello, se utiliza una función Φ , tal que:

$$\begin{aligned} \Phi: \quad \mathfrak{R}^D &\rightarrow \mathfrak{S} \\ \mathbf{x} &\rightarrow \Phi(\mathbf{x}) \end{aligned}$$

La frontera de decisión resultante en el espacio de entrada ya no será lineal.

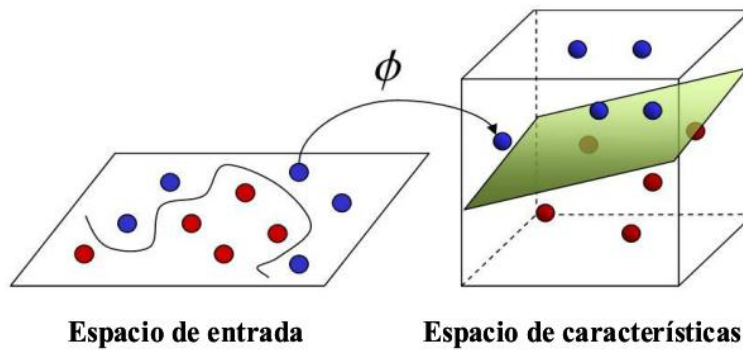


Figure 2.11 Transformación de los datos de entrada a un espacio de mayor dimensión.

Las funciones que se usan para poder realizar esta transformación se llaman funciones núcleo o kernel. Representan el producto vectorial en el espacio de características.

SVM con margen blando:

En algunos casos, puede existir ruido debido a errores en la recogida de datos o por la presencia de algún outlier (dato atípico). En este caso no es conveniente que la SVM se ajuste totalmente a los datos.

En la Figura 2.12 se observan dos conjuntos de datos y la frontera de decisión que se obtendría con la SVM de margen máximo. Cada conjunto está agrupado excepto por un punto que se encuentra muy próximo de los de la otra clase. Éste se puede corresponder con un dato atípico o que ha sido clasificado erróneamente. Este punto no debería ser considerado para hallar la frontera de decisión ya que al ser una excepción podría alterar los resultados deseados y nos llevaría a clasificar incorrectamente nuevos datos. La SVM tiene que ser más robusta para tener una mayor capacidad de generalización [27].

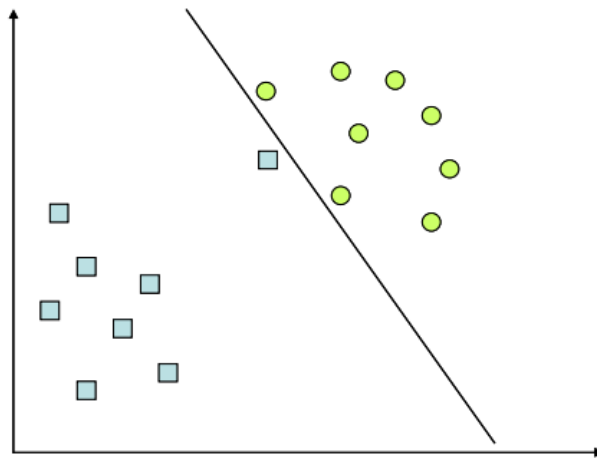


Figura 2.12 Clasificación con SVM de margen máximo de un conjunto de datos en el que existe un outlier.

CAPÍTULO 3. RESULTADOS

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 (DCM 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 la clasificación realizada a los rasgos prosódicos, con un objetivo en específico, obtener rasgos discriminantes para separar los dos grupos de participantes analizados.

3.1 Descripción estadística de los rasgos extraídos

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

Tabla 3.1 Valores estadísticos de rasgos del grupo Control.

Rasgos Prosódicos	N	Mínimo	Máximo	Media	Desviación estándar
Frecuencia mínima	11	60.5	73	64.5000	4.7329
Frecuencia máxima	11	283.5	438	351.7273	50.5556

Media de las frecuencia	11	124.539	202.4321	151.3526	23.0855
Desviación Estándar de las frecuencias	11	34.2428	54.2302	45.2451	6.4072

En la Tabla 3.1 podemos observar que el grupo de control tiene más estabilidad en las frecuencias mínimas con respecto al DCM ya que sus frecuencias se acercan más a la media. Esto se puede observar en el comportamiento de la desviación estándar como es menor la del grupo de control con respecto al DCM, de acuerdo a las frecuencias máximas el grupo de control sí se aleja más de la media, lo que da una mayor desviación estándar 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 DCM.

Rasgos Prosódicos	N	Mínimo	Máximo	Media	Desviación estándar
Frecuencia mínima	12	57	80.5	63.7083	6.7906
Frecuencia máxima	12	316.5	439.5	379.375	40.8312
Media de las frecuencia	12	134.0768	206.3976	174.9274	26.1623
Desviación Estándar de las frecuencias	12	36.3222	67.7891	50.8952	9.6432

En la tabla 3.2 se puede observar cómo la media de las frecuencias tiene una mayor desviación estándar con respecto a la del grupo control debido a la diferencia que existe con respecto a las medias de las frecuencias, siendo la del grupo de DCM con valores más elevados que el grupo de control. En cuanto a la desviación estándar, el grupo DCM tiene valores por encima del otro grupo y un poco más disperso.

3.2 Obtención de rasgos significativos

Al analizar los archivos de audio de nuestra base de datos y terminar la extracción de los datos de interés, se propone comprobar cuáles de ellos pueden tener un mayor grado de significancia de acuerdo a la discrepancia entre ambos grupos de sujetos. Esto quiere decir que la diferencia de valores entre uno y otro es lo suficientemente considerable como para resultar un factor de exclusión.

Con este objetivo, utilizaremos la prueba estadística de Wilcoxon como un método consistente que nos permitirá la obtención de estos parámetros significativos.

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 v2013 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 aceptan 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
Frecuencia mínima	0	0.4449
Frecuencia máxima	0	0.1303

Media de las frecuencias	0	0.0525
Desviación estándar de las frecuencias	0	0.1166

Con respecto al resultado de la tabla anterior se tomaron tres umbrales para decidir cuáles rasgos son significativos, posibles significativos y no significativos, se establecerá para *significativo* valores $p < 0.1$, para *posible significativo* $0.1 \leq \text{valores } p < 0.25$, y para *no significativo* valores $p \geq 0.25$. De este modo, se podrá hacer un mejor estudio con los datos según su grupo de significancia, para obtener mejores resultados a la hora de ser empleados en una clasificación entre los grupos de DCM 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.

Rasgo significativo	Rasgos posibles significativos	Rasgo no significativo
Media de las frecuencias	- Frecuencia máxima - Desviación estándar de las frecuencias	Frecuencia mínima

3.3 Clasificación de los pacientes

A continuación nos centraremos en el desarrollo de métodos que nos brinden una posibilidad de clasificación atinada de un paciente determinado. El clasificador que se utilizó para la presente investigación fue:

- Máquina de Soporte Vectorial (SVM siglas en inglés).

Para validar los resultados obtenidos por el clasificador 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 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 .

3.4 Clasificación de rasgos prosódicos empleando SVM

Se utilizaron varias configuraciones de clasificación brindadas por las SVM, el cual se realizaron en el MatLab v2013 utilizando las funciones `svmtrain` para el entrenamiento y `svmclassify` para la clasificación por el método de validación cruzada dejando uno afuera explicado anteriormente, de acuerdo a los errores de clasificación obtenidos se usó una fórmula estadística para representar la efectividad de la misma.

3.4.1 Clasificación de los rasgos prosódicos según su significancia

Empleando la metodología anteriormente descrita se evaluaron los conjuntos de rasgos y sus combinaciones mediante un clasificador de máquinas de soporte vectorial o como es conocido por sus siglas en inglés SVM. Este permite mediante una función matemática denominada *kernel* buscar la mayor separabilidad de las clases. En la presente investigaciones utilizamos tres tipos de *kernel*: polynomial, cuadrático y base radial. A continuación se muestran los resultados en la Tabla 3.5.

Tabla 3.5 Clasificación de los rasgos prosódicos, posibles combinaciones de significancia

Rasgos	K_F(polynomial)	K_F(quadratic)	K_F(rbf)
S	69,5652	69,5652	73,913

PS	60,8696	52,1739	56,5217
NS	69,5652	30,4348	65,2174
S-PS	65,2174	73,913	69,5652
S-NS	65,2174	65,2174	60,8696
PS-NS	56,5217	69,5652	56,5217
S-PS-NS	73,913	78,2609	73,913

Como se refleja en la tabla anterior, los valores varían de acuerdo a su significancia y la combinación entre ellas, se puede ver cómo la configuración que más se adecua a las pruebas de Wilcoxon es la Kernel Funtion, quadratic. Esta nos brinda un buen resultado de acuerdo a sus rasgos de significancia con un 30.4% para el conjunto de *no significativo* lo cual bajo comparado con el resto de los resultados obtenidos. Se obtiene además un 52.1% para el conjunto de *posible significativo* y un 69.5652% para el conjunto de *significativo*, pero al combinar los conjuntos de rasgos de significancia se obtienen resultados más altos, como el 73.9% de *significativo* con *posible significativo* y el 78.2% con la unión de los tres conjuntos, siendo el valor más alto de clasificación correcta en este estudio. Es interesante notar que la técnica de clasificación utilizada toma ventaja de la inclusión de un mayor número de rasgos para mejorar la clasificación correcta general como era esperado. Los resultados obtenidos pueden considerarse relativamente buenos basados en el número de rasgos, en el tamaño de la data y resultados similares publicados anteriormente como los mostrados en la Tabla 3.6.

Tabla 3.6 Resumen de resultados alcanzados en trabajos publicados.

Trabajos Publicados	Participantes (Pacientes)	Clasificación correcta
Maiden Lehr [23]	72 (35)	75.4 % - 81.5 %
Calvin Thomas [28]	85 (50)	58.8 % - 75.3 %
Romola Bucks [29]	24 (8)	87.5 %
Presente Investigación	23 (12)	78.26 %

3.4.2 Clasificación de los rasgos prosódicos según posibles combinaciones

Sin embargo, a pesar de que la prueba de Wilcoxon es una de las más utilizadas para la evaluación de significancia de rasgos y la reducción de los mismos, el método no puede garantizar el desempeño óptimo de un clasificador, dado que el conjunto de rasgos que mejor represente una clase no debe ser determinado por niveles de significancia aisladas.

Resumiendo, dado que la pregunta de cómo escoger el mejor conjunto de rasgos no posee una respuesta sencilla, en la presente investigación evaluamos todas las posibles combinaciones de rasgos sin repetición dando lugar a 15 combinaciones. A continuación se muestran los resultados en la Tabla 3.7.

Los resultados que se muestran en la Tabla 3.7 son muy reveladores y confirman lo anteriormente expresado. La mejor clasificación se obtiene para la combinación de los rasgos de **Frecuencia Media** y **Desviación Estándar**, cuya combinación no había sido posible obtener siguiendo la metodología de la prueba de Wilcoxon. El rasgo de **Desviación Estándar** estaba agrupado junto con **Frecuencia Máxima** en el conjunto de *posibles significativos* lo cual impedía la combinación de este de forma aislada con cualquiera de los otros 2 rasgos.

Tabla 3.7 Clasificación de los rasgos prosódicos, combinaciones posibles

Rasgos	K_F(polynomial)	K_F(quadratic)	K_F(rbf)
FMIN	69,5652	30,4348	65,2174
FMAX	52,1739	43,4783	69,5652
FMED	69,5652	69,5652	73,913
FDES	56,5217	60,8696	65,2174
FMIN-FMAX	39,1304	47,8261	52,1739
FMIN-FMED	65,2174	65,2174	60,8696
FMIN-FDES	52,1739	69,5652	65,2174
FMAX-FMED	60,8696	65,2174	73,913
FMAX-FDES	60,8696	52,1739	56,5217
FMED-FDES	86,9565	82,6087	78,2609
FMIN-FMAX-FMED	60,8696	52,1739	73,913
FMIN-FMED-FDES	78,2609	69,5652	69,5652
FMIN-FMAX-FDES	56,5217	69,5652	56,5217
FMAX-FMED-FDES	65,2174	73,913	69,5652
FMIN-FMAX-FMED-FDES	73,913	78,2609	73,913

Resumiendo se obtiene un 86.9% de clasificación correcta, mediante el *kernel Funtion Polynomial* del SVM con solo 3 errores de clasificación lo cual puede considerarse de un muy buen resultado teniendo en cuenta tamaño de la data, número de rasgos y resultados anteriores publicados para clasificaciones similares. Esta forma de buscar la mejor combinación de rasgos que aporten el mejor resultado posible es la de mayor costo computacional, ya que analiza todas las posibles combinaciones, y a medida que aumenta el número de rasgos la cantidad de combinaciones aumenta dado por la formula $\sum_{k=1}^N \frac{N!}{k!(N-k)!}$ donde N es el número total de rasgos y k el tamaño del

conjunto. Debemos resaltar para concluir que siguiendo el anterior método es la única vía para obtener el mejor resultado posible para la data y clasificador bajo análisis.

3.5 Conclusiones del capítulo

Se han realizado distintas pruebas y estadísticas a partir de una base de datos con dos grupos distintos de sujeto. Los resultados han demostrado que existen algunas diferencias a tener en cuenta entre ambos, en datos relativos como las frecuencias de los sujetos. Datos que corroboran estas diferencias, se encuentran los resultados obtenidos mediante diversos métodos.

La realización del clasificador SVM revela porcentajes a la hora de distinguir entre un paciente sano y otro enfermo lo suficientemente elevadas como para alentar el desarrollo de nuevos estudios sobre estos parámetros prosódicos.

También es importante tener en cuenta que todos los resultados obtenidos en este estudio están condicionados por la precisión de la extracción de datos de la herramienta *Prosogram* v29f, que es mediante la cual se detectan los distintos núcleos silábicos que forman las señales de voces grabadas. Aunque el *Prosogram* posee sus limitaciones en cuanto a la detección y reconocimiento de núcleos silábicos sus resultados se han considerado buenos durante la realización de este estudio.

Se realizó en la presente investigación una búsqueda de la combinación de rasgos que aportaran la mejor clasificación general obteniendo un conjunto formado por dos rasgos, incluidos el mejor valorado por la prueba estadística (FMED) y un segundo rasgo que se encontraba agrupado en el conjunto de *posibles significativos* (FDES). Lo anterior demuestra que el tema de selección de rasgos es una asignatura pendiente para los tiempos actuales. A forma de conclusión, los resultados obtenidos animan a un estudio más profundo de los rasgos prosódicos como una herramienta sencilla, rápida y económica para la detección precoz del deterioro cognitivo moderado en la población cubana.

CONCLUSIONES

CONCLUSIONES

- Se identificaron cuatro rasgos prosódicos presentes en la voz de los pacientes con demencia moderada.
- Se creó una base de datos de voces de pacientes con deterioro cognitivo moderado y sanos en el Hogar del Adulto Mayor #2 de Santa Clara formada por 11 participantes sanos y 12 pacientes con deterioro.
- Se realizó la clasificación de los dos grupos, control y pacientes con deterioro cognitivo moderado, mediante varias configuraciones de un clasificador SVM del cual se obtuvieron resultados satisfactorios con respecto a otros estudios publicados.

RECOMENDACIONES

Para continuar el desarrollo del tema de esta investigación se recomienda:

- Aumentar la cantidad de muestras de la base de casos, realizando grabaciones en otros hogares del país.
- Ampliar el número de rasgos prosódicos para realizar una mejor diferenciación de los grupos estudiados en este trabajo.
- Diseñar una herramienta automática para identificar deterioro cognitivo moderado mediante procesamiento digital de voz.

REFERENCIAS BIBLIOGRÁFICAS

- [1] S. Doménech, "Aplicación de un programa de estimulación de memoria a enfermos de Alzheimer en fase leve," Departamento de Psicología Evolutiva y de la Educación, Universidad de Barcelona. Facultad de Psicología, Barcelona. España, 2004.
- [2] A. Slachevsky and F. Oyarzo, "Las demencias: historia, clasificación y aproximación clínica". Buenos Aires, Argentina, 2008.
- [3] Richards SS and H. HC, "Diagnosis, management and treatment of Alzheimer disease", 1999.
- [4] Gil.P and Martin.J, "Demencia," ed, p. 173, 1997.
- [5] Medina.JH, "Diagnóstico y Tratamiento del Deterioro Cognoscitivo en el Adulto Mayor," C. N. d. E. T. e. Salud, Ed., ed. Mexico, 2012.
- [6] Gutiérrez.LM and Arrieta.I. "Plan de acción Alzheimer y otras demencias". Geriatria y Salud, 2014.
- [7] Llibre.JJ., "Envejecimiento y demencia: implicaciones para la comunidad científica, la salud pública y la sociedad cubana". Academia de Ciencias de Cuba, 2012.
- [8] Llibre.JJ., "Epidemiología, diagnóstico e impacto de las demencias y otras enfermedades crónicas no transmisibles en adulto mayores", Academia de Ciencias de Cuba. 2014.
- [9] Kreiman.J and Gerrat.BR, "Perceptual evaluation of voice quality Review: Tutorial and a Framework for future research." vol. 36, ed.,. pp. 21-40, 1993.
- [10] R. Kent, "Hearing and believing: Some limits to auditory-perceptual assessment of speech and voice disorders," J.Speech Hear, 1996.
- [11] Darley.FL, Aronson.AE, and Brown.JR, "Differential diagnostic patterns of dysarthria. J. Speech Hear", 1969.

REFERENCIAS BIBLIOGRÁFICAS

- [12] Castillo.E., "A modem approach to dysarthria classification". UNB., 2002
- [13] Pakhomov.S, Smith.GE, and Chacon.D, "Análisis computerizado del habla y el lenguaje para identificar correlaciones psicolingüísticas de la degeneración del lóbulo fronto-temporal " 2009.
- [14] Hakkay-Tür.D, Vergyi.D, and Tur.G, "Valoración automática del estado cognitivo basada en el habla " 2010.
- [15] Pakhomov.S, "Índices semánticos automáticos para la detección temprana de los cambios cognitivos," 2012.
- [16] Folstein.M.F, Folstein.S.E, and Mchugh.P.R, Mini-Mental state: A Practical Method for Grading the Cognitive State of Patients for the Clinician vol. 12, 1975.
- [17] (2004) Órgano Oficial de expresión de la Fundación OMIE, Avances en Salud Mental Relacional. Revista Internacional On-line.
- [18] Darley.F.L, Aronson.A.E, and Brown.J.R, Motor speech disorders, 1975.
- [19] Kelma.A, A Milt Gross Comic Reader, 2010.
- [20] "Grupo de Aplicaciones de Procesado de Señales," 2013.
- [21] Boza.D.T, "Manual de Sistema de Grabación para el diagnóstico de Deterioro Cognitivo DCGrab v3.0," ed. Universidad de Santa Clara, 2013.
- [22] Roark.P.H, "Análisis Computarizado del Habla y Lenguaje para Identificar Correlaciones Psicolingüísticas de Degeneración del Lobar de Fronto-temporal," 2012.
- [23] Lehr.M, Shafran.I, and Roark.B, "Fully Automated Neuropsychological Assessment for Detecting Mild Cognitive Impairment," presented at the Annual Conference of International Speech Communication Association en Portland, 2012.
- [24] Mertens.P. Transcription of prosody using pitch contour stylization based on a tonal perception model and automatic or annotation-based segmentation, 2012. Available: <http://bach.arts.kuleuven.be/pmertens/prosogram/>
- [25] Betancourt.G.A., "Las Maquinas de Soporte Vectorial (SVMs) ". Scientia Et Technica., 2012.
- [26] Riobó.V, "Reconocimiento de localizaciones mediante Máquinas de Soporte Vectorial," Ingeniería industrial, Universidad Carlos III de Madrid, España, 2011.

REFERENCIAS BIBLIOGRÁFICAS

- [27] Gutiérrez,E, "Aplicación de las Máquinas de Soporte Vectorial para el reconocimiento de matrículas," Ingeniería Industrial, Universidad Pontificia Comillas, España, 2007.
- [28] Thomas, C., Keselj, V., Cercone, N., Rockwood, K., Asp, E. "Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech," Mechatronics and Automation, IEEE International Conference, vol. 3: 1569-1574, 2005.
- [29] Bucks, R. S., Singh, S., Cuerden, J.M., Wilcock, G.K. "Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analyzing lexical performance". Aphasiology, vol. 14: 71-91, 2002.