Análisis Acústico de la Voz para la Detección del Deterioro Cognitivo (Acoustic Speech Analysis for the Assessment of Patients with Cognitive Impairment)

Lixania Hernández, Nuria Calet [a] y Jose. A. Gonzalez-Lopez[b]

a Departamento de Psicología Evolutiva y de la Educación, Universidad de Granada, Facultad de Ciencias de la Educación, 18071 Granada, España

b Departamento de Teoría de la Señal, Telemática y Comunicaciones, Universidad de Granada, ETS de Ingenierías Informática y de Telecomunicación, 18071 Granada, España

Email: lixaniahernandez@gmail.com; ncalet@ugr.es; joseangl@ugr.es

Autor de correspondencia: Jose A. Gonzalez (joseangl@ugr.es) Departamento de Teoría de la Señal, Telemática y Comunicaciones, Universidad de Granada, ETS de Ingenierías Informática y de Telecomunicación, 18071 Granada, España.

Financiación

La presente investigación no ha recibido ayudas específicas provenientes de agencias del sector público, sector comercial o entidades sin ánimo de lucro.

Conflicto de intereses

Los autores son responsables de la investigación aquí descrita y han participado en el diseño, análisis e interpretación de los datos, redacción o revisión del manuscrito, y han aprobado el manuscrito tal como se presenta. Los autores no tienen ningún conflicto de intereses que pueda interpretarse como una influencia en la investigación.

Análisis Acústico de la Voz para la Detección del Deterioro Cognitivo

Resumen

Antecedentes y objetivo: Investigaciones recientes señalan que el análisis acústico de la voz es una herramienta valiosa tanto para la evaluación objetiva del deterioro cognitivo, como para la monitorización de la progresión de la enfermedad. El objetivo de este estudio es determinar si el análisis automático de la voz es también útil en el diagnóstico del deterioro cognitivo.

Materiales y métodos: Se trata de un estudio descriptivo correlacional transversal en el que se hace una comparativa entre un grupo experimental compuesto por 10 participantes con deterioro cognitivo y un grupo control con 10 participantes asintomáticos. Se recogieron grabaciones de voz de ambos grupos mientras realizaban 4 tareas cognitivas: conteo hacia atrás (desde el 305 hasta el 285), descripción de una lámina y dos tareas de fluidez verbal (fonológica y semántica). Las muestras de voz fueron posteriormente analizadas acústicamente para extraer de ellas variables predictoras del estado cognitivo del sujeto.

Resultados: Los resultados muestran que las variables acústicas analizadas son eficaces para las detección precoz del deterioro cognitivo, logrando una tasa de clasificación del 100% al predecir el estado cognitivo de los sujetos de la muestra. Parece que las tareas de fluidez verbal presentan mayor eficacia que las de conteo hacia atrás y descripción de una imagen.

Conclusiones: A la luz de los resultados encontrados consideramos que el análisis automático de la voz podría ser una herramienta de evaluación objetiva adicional para mayores con deterioro cognitivo. Se discuten las implicaciones de los resultados encontrados.

Palabras clave: Evaluación del deterioro cognitivo; análisis acústico de la voz; demencia; aprendizaje computacional.

Abstract

Background and aim: Recent research has shown that acoustic voice analysis is a valuable tool for both the objective assessment of cognitive impairment and the monitoring of disease progression. The aim of this study is to determine whether automatic voice analysis is also useful for diagnosis of cognitive impairment.

Materials and methods: This is a descriptive cross-sectional correlational study in which a comparison is made between an experimental group composed of 10 participants with cognitive impairment and a control group with 10 healthy participants. Voice recordings were collected from both groups while they performed 4 tasks: counting backwards (from 305 to 285), description of an image and two tasks of verbal fluency (phonological and semantic). The voice samples were later acoustically analyzed to extract from them variables predictive of the cognitive state of the subject.

Results: The results show that the acoustic variables are effective for early detection of cognitive impairment, achieving a classification rate of 100% when predicting the cognitive status of the subjects in our sample. From the results it is clear that verbal fluency tasks are more effective than counting backwards and describing an image.

Conclusions: In light of our results, we consider that automatic voice analysis could be an additional objective assessment tool for elderly people with cognitive impairment. The implications of the found results are discussed.

Keywords: Cognitive impairment assessment; acoustic speech Analysis; dementia; machine learning.

Introducción

El concepto más utilizado para referirse al estado entre el envejecimiento y la demencia es el de deterioro cognitivo leve (DCL) (Junqué & Jurado, 2009; Petersen et al., 1997). El DCL es una condición patológica que se caracteriza por cierto grado de déficit cognitivo cuya severidad resulta insuficiente para cumplir criterios de demencia, ya que las personas que lo sufren poseen independencia en las actividades de la vida diaria (Mora-Simón et al., 2012). Además, el deterioro cognitivo es uno de los primeros síntomas que se puede detectar acompañado de alteraciones de memoria reciente y conductas repetitivas, principalmente a personas mayores de 65 años (Velasquez-Perez, Guerrero-Camacho, Rodriguez-Agudelo, Alonso-Vilatela, & Yescas-Gomez, 2008).

El DCL agrupa a personas que presentan déficits cognitivos sin demencia pero con alto riesgo de evolucionar en ella (Serrano, Dillon, Leis, Taragano, & Allegri, 2013). Con el fin de identificar a los pacientes, el DCL ha sido clasificado en tres subtipos (Lopez et al., 2003): el DCL amnésico (DCL-a), caracterizado por un déficit de la memoria; el DCL multidominio (DCL-mult), que implica un déficit leve de más de un dominio cognitivo (puede incluir la memoria), pero sin cumplir criterios para el diagnóstico de demencia; y el DCL monodominio no amnésico (DCL-mnoa), que representa la afección de un solo dominio distinto de la memoria. Este último constituiría el estado prodrómico de demencias no Alzheimer, tales como las frontotemporales, demencia por cuerpos de Lewy o vascular, mientras que el DCL-a evoluciona generalmente a enfermedad de Alzheimer (EA) (Migliacci, Scharovsky, & Gonorazky, 2009).

A diferencia del DCL, la demencia se define como una pérdida de capacidades cognitivas (memoria y otras funciones tales como el lenguaje y la capacidad ejecutiva), de tal gravedad como para interferir en las actividades cotidianas de la persona (Donoso & Vásquez, 2002). El tipo de demencia más frecuente es el Alzheimer (Acarin, 2010). Se estima que hay 35,6 millones de personas en todo el mundo con esta enfermedad y esta cifra sigue en aumento (Kunz et al., 2017). La EA se diagnostica cuando ha alcanzado la etapa en que los síntomas cognitivos y neuropsiquiátricos interfieren con el funcionamiento social o las actividades de la vida diaria. En la fase inicial de la enfermedad se produce una alteración de la fluidez semántica y posteriormente la afectación de la denominación (anomia) y trastornos de comprensión (Facal et al., 2009). Más tarde, aparecen los trastornos del lenguaje, la pérdida de habilidades motoras, desorientación y al final el paciente termina sin lenguaje y totalmente dependiente.

Las características del habla afectada en personas con deterioro cognitivo parecen ser las relacionadas con la articulación y prosodia en términos de medidas temporales y acústicas, que incluye alteraciones del ritmo (capacidad de variar el nivel de tono, modulación del tono, reducción o tasa fluctuante de salida del lenguaje, frecuentes pausas para encontrar palabras, falta de iniciativa y lentitud) (Horley, Reid, & Burnham, 2010; Martínez-Sánchez, Meilán, Pérez, Carro, & Arana, 2012). En etapas posteriores estas personas también presentan déficits en la fluidez verbal y razonamiento sintáctico (Taler & Phillips, 2008). Algunas de estas alteraciones podrían detectarse mediante el análisis automático de la voz de los sujetos mientras estos realizan tareas cognitivas o incluso durante conversaciones informales (König et al., 2015).

Es importante realizar un diagnóstico precoz de estas enfermedades neurocognitivas para poder iniciar una intervención preventiva cuanto antes; sin embargo estas enfermedades son complejas y varían en sus síntomas de persona a persona y de etapa a etapa. Para su detección existen gran variedad de herramientas, pero un diagnóstico confiable sólo puede realizarse a través de evaluaciones en profundidad, de una combinación integral de evaluaciones conductuales (p. ej. pruebas psicométricas) y en vivo (p. ej. formación de imágenes cerebrales funcionales). Las evaluaciones del comportamiento típicamente consisten en entrevistas estructuradas y también pueden incluir una serie de tareas bien definidas para evaluar aspectos particulares de la cognición, como la memoria y la función ejecutiva. Algunas de éstas son las denominadas tareas de Fluidez Verbal Semántica (p. ej. decir animales en 1 minuto) y de Fluidez Verbal Fonológica (p. ej. decir palabras que empiecen por /S/ en un minuto) (Auriacombe et al., 2006; Pakhomov, Eberly, & Knopman, 2016; Raoux et al., 2008).

Además de estas técnicas, investigadores y personal clínico buscan otro tipo de herramientas no intrusivas, sencillas y poco costosas para evaluar la gravedad y progresión de la enfermedad desde sus etapas iniciales. La detección temprana del DCL es necesaria para optimizar la atención del paciente y para proporcionar mejores herramientas para la investigación clínica (Martínez-Sánchez et al., 2012). En este sentido, el lenguaje hablado es el método de comunicación más espontáneo, intuitivo y eficiente que revela el estado cognitivo y emocional de una persona, por lo cual, el déficit en este dominio demuestra ser un fuerte predictor para la progresión de la enfermedad (Satt, Hoory, König, Aalten, & Robert, 2014).

Una de las herramientas más usadas para obtener información objetiva sobre la voz es el análisis de determinadas variables acústicas (Delgado, León, Jiménez, & Izquierdo, 2017). Ésta es una técnica objetiva, eficiente y no invasiva basada en el procesamiento digital de la señal de voz. Diversos estudios muestran que esta aproximación es satisfactoria para el diagnóstico de ciertos tipos de patologías de la voz (Godino-Llorente & Gomez-Vilda, 2004; Sáenz-Lechón, Godino-Llorente, Osma-Ruiz, & Gómez-Vilda, 2006). Para su interpretación se requiere como referencia los valores de normalidad y su fiabilidad depende de factores como el tipo de micrófono, el ruido ambiental, el software de análisis y los parámetros acústicos utilizados (Delgado et al., 2017). Estas técnicas de análisis de la voz no requieren una amplia infraestructura o la disponibilidad de equipos médicos, y la obtención por estos medios es fácil, rápida y económica. Asimismo, el análisis por métodos automáticos de habla espontánea, posiblemente combinada con otras metodologías, tiene el potencial de convertirse en un método útil y eficaz para el diagnóstico del deterioro cognitivo (López-de-Ipiña et al., 2013).

También es de gran interés proporcionar métodos fiables para evaluar la progresión del deterioro cognitivo. En este sentido también destaca el análisis automático de la voz, ya que éste es capaz de evaluar con precisión a los pacientes en tiempo real. Incluso, usar situaciones de la vida cotidiana y aplicar métodos menos intrusivos que no requieren personal especializado (König et al., 2015).

Debido a las potenciales ventajas de estas técnicas, en los últimos años diversos estudios han explorado el uso del análisis acústico de la voz como herramienta para la detección y diagnóstico del DCL y otras demencias como la EA (Godino-Llorente & Gomez-Vilda, 2004; Sáenz-Lechón et al., 2006). En estos estudios se explota el hecho de que estos trastornos afectan al habla y el lenguaje de los

sujetos que los padecen y, por tanto, es posible detectar estas patologías a través de un análisis objetivo de la señal de voz emitida por los sujetos. En términos generales, los marcadores que han demostrado verse afectados por estas patologías se pueden clasificar en dos grupos (Weiner, Engelbart, & Schultz, 2017; Weiner & Schultz, 2018): acústicos y lingüísticos. Los marcadores lingüísticos reflejan lo que el sujeto ha dicho, mientras que los acústicos miden cómo habla. Estudios previos han mostrado que algunos de los principales marcadores acústicos afectados en el DCL y la demencia son la proporción de pausas y/o silencios en el habla del sujeto (König et al., 2015; Satt et al., 2014; Weiner et al., 2017; Weiner, Herff, & Schultz, 2016), la tasa de locución (medida como el número de palabras emitidas por minuto) (Weiner & Schultz, 2018) y las características espectrales de la voz (Weiner & Schultz, 2018). Los marcadores lingüísticos, por otro lado, vienen a medir cambios en el vocabulario y la estructura sintáctico-semántica de las oraciones causados por la demencia. Investigaciones recientes han demostrado que algunos de los marcadores lingüísticos afectados en estas patologías son la riqueza léxica (Weiner et al., 2017), co-ocurrencia de palabras en el discurso (Mirheidari et al., 2018) y los roles gramaticales de las palabras en las oraciones (Mirheidari, Blackburn, Reuber, Walker, & Christensen, 2016; Weiner & Schultz, 2018). En cualquier caso, los marcadores lingüísticos se calculan a partir de las transcripciones del habla del sujeto. Estas transcripciones se han realizado de forma manual tradicionalmente, si bien recientemente se han empleado sistemas de reconocimiento automático de voz con resultados muy prometedores (Mirheidari et al., 2016; Weiner et al., 2016).

En base a todo lo anterior, y teniendo en cuenta el impacto social causado por el deterioro cognitivo, en este trabajo se analiza el uso del análisis automático de la voz como herramienta en el diagnóstico de este trastorno. Aunque existen trabajos previos en este ámbito, la mayoría se ha realizado en otras lenguas distintas del español (p. ej. König et al., 2015; Satt et al., 2014; Weiner et al., 2017; Weiner, Herff, & Schultz, 2016). Por otro lado, hasta donde sabemos, hay muy pocos estudios sobre predicción de puntuaciones clínicas a partir de muestras de audio (Yancheva, Fraser, & Rudzicz, 2015). Por tanto, el objetivo principal de este estudio es aportar evidencia de que el análisis automático mediante grabaciones de voz en tareas cognitivas aporta información para la evaluación y diagnóstico del deterioro cognitivo en español. Para ello en nuestro trabajo se analizan las grabaciones de voz de un conjunto de sujetos mientras estos realizaban una serie de tareas cognitivas. De estas grabaciones se calcularon una serie de variables acústicas (marcadores) relacionadas con la continuidad del habla del sujeto, esto es, la proporción y longitud de los silencios. Finalmente, se usaron técnicas de aprendizaje computacional para intentar predecir el estado cognitivo del sujeto y su gravedad a partir de las variables acústicas. Para ello, se evalúa la fiabilidad de la predicción de las puntuaciones obtenidas por los sujetos en el Mini-Examen Cognoscitivo (MEC) (Lobo, Escobar, Ezquerra, & Seva Díaz, 1980).

Las principales aportaciones de este estudio son: (i) la validación de variables acústicas propuestas en otros trabajos en población castellanoparlante, (ii) el uso de un conjunto rico y cognitivamente desafiante de tareas para la evaluación del DCL y (iii) la predicción de la puntuaciones del test MEC usando las variables acústicas mencionadas. Éste último es un test muy conocido para *screening* y seguimiento de demencias.

Método

Diseño

Se trata de un estudio descriptivo correlacional transversal de casos y controles, en el que las voces de los participantes asintomáticos y con deterioro cognitivo se evaluaron objetivamente mediante el análisis acústico de la voz mientras realizaban 4 tareas cognitivas.

Participantes

En este estudio participaron 20 sujetos, 10 de ellos con deterioro cognitivo (grupo experimental) y 10 asintomáticos (grupo de control). De los 10 sujetos del grupo experimental, 8 eran mujeres y 2 eran hombres. La edad media de los hombres era de 85 años (DT: .00) y la de las mujeres de 85.88 años (DT: 5.11; rango: 80-97 años). Dentro de este grupo, 7 de los participantes padecían DCL, 2 de ellos EA leve y 1 EA moderada. Los criterios de inclusión para los participantes de este grupo fueron: tener el español como lengua materna, tener un diagnóstico de deterioro cognitivo emitido por el servicio de neurología, no tener otras enfermedades de interés y no estar recibiendo tratamiento logopédico.

Por otro lado, el grupo de control se componía de 1 hombre de 89 años de edad y 9 mujeres. La edad media de las mujeres era 82 años (DT = 3.57; rango = 78-89 años). Los criterios de exclusión para este estudio fueron: tener estudios superiores, presentar antecedentes de trastornos de voz, obtener resultados compatibles con deterioro cognitivo en el Mini-Examen Cognoscitivo (MEC), no tener el español como lengua materna y haber recibido terapia en voz. Los participantes fueron reclutados de dos centros gerontológicos de Granada.

Todos los sujetos de este estudio contaban con estudios primarios. Previamente a las evaluaciones se solicitó una autorización firmada para participar en este estudio a través de un consentimiento informado. El estudio contó con la aprobación de la Comisión de Ética en Investigación Humana de la Universidad de Granada.

Procedimiento

En nuestro estudio la voz de los participantes fue grabada con una grabadora Olympus WS-110 situada a unos 30 cm de la cavidad oral mientras realizaban las siguientes tareas:

1. Tarea de fluidez semántica en la que los sujetos debían decir todos los animales que se acordasen en 1 minuto.
2. Tarea de fluidez fonológica en la que los sujetos debían decir palabras (no nombres propios) que comenzasen por la letra /s/ en 1 minuto.
3. En la tercera actividad los participantes debían describir la lámina del “robo de las galletas” del Test de Boston (Goodglass, Kaplan, & Barresi, 2001) con todos los detalles que pudieran y sin límite de tiempo.
4. En la cuarta y última actividad los sujetos tenían que realizar el conteo hacia atrás desde el 305 al 285. Esta última tarea se eligió porque en el estudio de König et al (2015) demostró ser de las tareas más fiables para discriminar entre sujetos asintomáticos y en fase de predemencia.

Instrumentos

Se recogieron una serie de variables sociodemográficas (edad, género y nivel de estudios) a través de un breve cuestionario diseñado para este fin. Por otro lado, el estado cognitivo general de los sujetos se evaluó mediante pruebas neuropsicológicas ampliamente usadas en la literatura como son el Mini-Examen Cognoscitivo (MEC), la Escala de Deterioro Global (GDS) y pruebas de fluidez verbal (fonológica y semántica). Asimismo, se administraron las tareas de conteo hacia atrás y de descripción de una lámina del test de Boston (lámina del "robo de las galletas"). Las grabaciones realizadas por los sujetos fueron analizadas acústicamente y de ellas se extrajeron una serie de variables que posteriormente se usaron en los análisis estadísticos. A continuación se describen con más detalle los instrumentos y los análisis acústicos realizados.

Mini-Examen cognoscitivo (MEC) (Lobo et al., 1980)**.** Consiste en una prueba *screening* del deterioro cognitivo conformada por 35 ítems agrupados en seis dominios: orientación temporal (cinco ítems), orientación espacial (cinco ítems), fijación (un ítem), concentración y cálculo (dos ítems), memoria (un ítem) y lenguaje y construcción (once ítems). El rango de puntuación abarca de 0 a 35 puntos. Se considera que existe deterioro cognitivo si la puntuación es menor a 23/24 puntos. El coeficiente de fiabilidad de la prueba es de KR= .637 (IC 95%= .596 - .678; z= 12.655; $p < .01$).

Escala de Deterioro Global (GDS) (Reisberg, Ferris, de Leon, & Crook, 1982). Esta escala permite realizar una valoración global de las etapas de la función cognitiva para aquellos que padecen de una demencia degenerativa primaria. Se divide en 7 etapas diferentes, ordenadas desde la normalidad hasta los grados más severos de la demencia de enfermedad de Alzheimer. Las etapas 1-3 son las etapas previas a la demencia, mientras que las etapas 4-7 son las etapas de la demencia (p. ej. en la etapa 5 un individuo ya no puede sobrevivir sin asistencia). La condición del paciente de EA en cada etapa es: estadio 1 (normal), estadio 2 (queja subjetiva de memoria), estadio 3 (deterioro cognitivo leve), estadio 4 (demencia leve), estadio 5 (demencia moderada), estadio 6 (demencia moderadamente severa) y estadio 7 (demencia severa). Los cuidadores pueden tener una idea aproximada de dónde se encuentra la persona en el proceso de la enfermedad observando las características de comportamiento y comparándolas con el GDS.

Test de Fluidez Verbal Fonológica y Semántica (Marino & Alderete, 2009). Las Pruebas de Fluidez Verbal (PFV) son técnicas neuropsicológicas que conllevan la actividad de diversos procesos cognitivos, como operaciones ejecutivas, mecanismos de control, atencionales y memoria semántica. Estas pruebas consisten en la producción de determinadas palabras en un tiempo estipulado. La evaluación de la fluidez verbal se divide en dos pruebas: una denominada fluidez verbal semántica (FVS), donde se pide al sujeto que nombre elementos dentro de una categoría semántica determinada (p. ej. animales, frutas); y otra prueba denominada fluidez verbal fonológica (FVF), en la que se pide al sujeto que diga todas las palabras que comiencen con una letra determinada. Las dos pruebas exigen una serie de demandas ejecutivas que no implican los mismos procesos y estrategias cognitivas.

Análisis acústico de la voz. Las grabaciones de voz realizadas por los participantes se analizaron sin ningún tipo de pre-procesamiento. A partir de las mismas, se calcularon las variables acústicas que se

describen brevemente a continuación (ver el Apéndice A para una descripción en profundidad de las mismas).

En las dos tareas de fluidez verbal se anotó de forma manual el tiempo de inicio de las palabras emitidas por los participantes usando el software Audacity versión 2.1.0. También se contabilizó de forma manual el número total de palabras emitidas en el tiempo que duraba cada tarea (1 minuto). A partir de esos datos, se calcularon las siguientes variables acústicas:

- Número de palabras emitidas por el sujeto en 1 minuto (1 variable por tarea).
- La diferencia en segundos entre el inicio de la primera palabra y el resto, hasta un máximo de 9 palabras (8 variables por tarea).
- La posición relativa en la grabación de las primeras 9 palabras, considerando como posición 0 el comienzo de la primera palabra y posición relativa 1 el minuto de estar el sujeto hablando (8 variables por tarea).

En las tareas de conteo hacia atrás y de descripción de la lámina del test de Boston se analizó la continuidad de la voz de los participantes. Se espera que el habla de aquellos con deterioro cognitivo y/o demencia sea menos continua, esto es, contenga más silencios (pausas o interrupciones) que la de los sujetos del grupo de control. Para este análisis se empleó la técnica automática de detección de la actividad de voz (VAD; del inglés *voice activity detection*) descrita en el Anexo A para calcular la duración de los segmentos de voz y silencio en las grabaciones. Asimismo, se midió de forma automática la duración de los segmentos periódicos y no periódicos de la voz a partir de los valores de la frecuencia fundamental (F0) calculados por el software Wavesurfer versión 1.8.8. En definitiva, a partir de las grabaciones de voz se calcularon de forma automática las siguientes medidas:

- Duración en segundos de los segmentos de voz.
- Duración en segundos de los segmentos de silencio.
- Duración en segundos de los segmentos periódicos.
- Duración en segundos de los segmentos aperiódicos.

A partir de las medidas anteriores, se calcularon de forman independiente para cada tarea las siguientes 36 variables acústicas propuestas en König et al. (2015):

1. La duración media de cada tipo de segmento (voz, silencio, periódico y aperiódico) en cada grabación (4 variables por tarea). Se espera que los sujetos del grupo de control presenten duraciones más largas de la voz y longitudes de segmentos periódicos y, por consiguiente, duraciones más cortas para el silencio y las longitudes de segmentos aperiódicos.
2. La relación entre cada par de promedio de duraciones calculadas en el paso anterior, definida como la duración media de la voz / duración media del silencio, duración media del silencio / duración media de la voz, duración periódica media / duración aperiódica media y duración aperiódica media / duración periódica media (4 variables por cada tarea).

3. La mediana de las duraciones para cada tipo de segmento y, de forma similar al paso 2, la relación entre las medianas de las duraciones (8 variables por tarea).
4. La desviación estándar de las duraciones para cada tipo de segmento y la relación entre la desviación estándar de las duraciones (8 variables por tarea).
5. La suma de las duraciones en segundos para cada tipo de segmento y la relación entre la suma de las duraciones (8 variables por tarea).
6. El recuento del número de segmentos de cada tipo (4 variables por tarea)

Análisis estadísticos

Se utilizó el programa SPSS versión 22.0 para el análisis estadístico de los datos. En primer lugar, se procedió al análisis descriptivo de las variables de interés en el estudio. Se realizaron análisis estadísticos no paramétricos debido a que la prueba de Kolmogorov-Smirnov no mostró normalidad para ninguna de las variables. Asimismo, las pruebas de homocedasticidad de Levene mostraron no homogeneidad en las varianzas para algunas de las variables ($p > .05$). Se utilizó, por tanto, la prueba no paramétrica *U* de Mann-Whitney para muestras independientes comparando cada variable entre grupos (experimental y control).

Por otro lado, se utilizó el software de aprendizaje automático Weka versión 3.9.1 para predecir el estado cognitivo de los sujetos (asintomático o con deterioro cognitivo) y su puntuación en el test MEC a partir de las variables acústicas que mostraron diferencias significativas entre grupos en el análisis estadístico. Tanto en la tarea de clasificación como en la de predicción de las puntuaciones del test MEC se utilizó un esquema de validación cruzada de tipo *leave-one-out*: se aleatorizó el orden de los participantes y se utilizaron las variables de voz calculadas para 19 de ellos para entrenar los modelos de clasificación/predicción, evaluando el modelo obtenido en el participante restante. Este procedimiento se repitió 20 veces. Asimismo, se analizó la bondad de ajuste lograda por tres de las técnicas de aprendizaje computacional más conocidas en la literatura: regresión lineal (regresión logística para clasificación), *random forests* (Breiman, 2001) y perceptrón multicapa (Bishop, 2006).

Resultados

Análisis descriptivo

En la Tabla 1 se recogen los estadísticos descriptivos por grupos de los participantes del estudio.

Tabla 1. *Estadísticos descriptivos por grupos.*

Variable	GE (n= 10)	GC (n= 10)	*p*
Género			
Mujer	8	9	

Hombre	2	1	
Edad	85.70 (80-97)	82.70 (78-89)	.11
MEC	18.70 (14-23)	29.60 (28-30)	.00
GDS	3.40 (3-5)	1.30 (1-2)	.00

Nota. GE= Grupo experimental; GC= Grupo control; MEC= Mini-Examen Cognoscitivo; GDS= Escala de Deterioro Global. Los número entre paréntesis corresponden al rango.

En primer lugar, se comprobó mediante la prueba no paramétrica Mann-Whitney que ambos grupos (GE y GC) presentaban diferencias en las medidas MEC (U = .00; p = .000), y GDS (U = .00; p = .000), pero no en la edad (p > .05), corroborando de este modo que la diferencia entre grupos alude solamente al deterioro cognitivo que presenta el GE en comparación con el GC.

En segundo lugar, se representó gráficamente las variables acústicas calculadas, por un lado, para las tareas de conteo hacia atrás y descripción de la imagen del test de Boston y, por otro, para las tareas de fluidez verbal con objeto de analizar si éstas contenían la suficiente información para discriminar entre ambos grupos (GE y GC). Dada el gran número de variables a representar para cada sujeto, se utilizó la técnica *t-distributed stochastic neighbor embedding* (t-SNE) (Maaten & Hinton, 2008) para proyectar las variables de cada sujeto a un espacio de dos dimensiones y así facilitar su visualización. Las proyecciones obtenidas con la técnica t-SNE su muestran en la Figura 1a (variables calculadas para las tareas de imagen y contar) y Figura 1b (variables calculadas para las tareas de fluidez verbal fonológica y semántica). En la gráfica cada punto representa a un sujeto. Como se puede apreciar, los puntos referidos a sujetos del mismo grupo tienden a estar más cerca unos de otros, mientras que puntos de sujetos de distinto grupo tienden a estar alejados. Esto viene a apoyar la idea de que las diferencias existentes entre ambos grupos también se plasman en las variables acústica. También se observa que los grupos son más “puros” en el caso de las tareas de fluidez, mientras que en las tareas de conteo regresivo y descripción de la lámina del test de Boston se producen algunas confusiones, al estar algunos de los puntos situados en las inmediaciones de los agrupamientos del grupo contrario. Esto, como se confirmará más adelante, viene a indicar que las variables calculadas para las tareas de fluidez verbal son mejores predictoras del estado del sujeto.

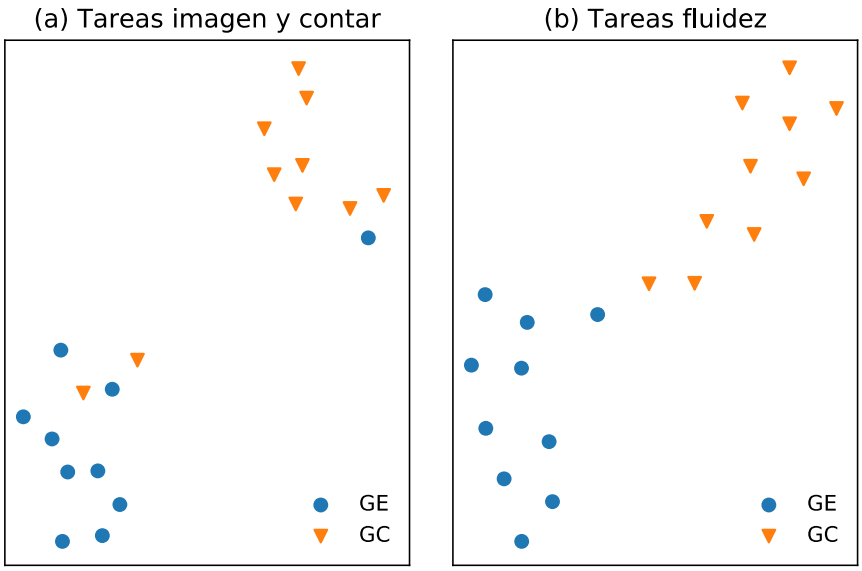

Figura 1. Visualización usando la técnica t-SNE de las variables de voz calculadas para (a) las tareas de descripción de la lámina robo de las galletas del test de Boston (tarea imagen) y de conteo hacia atrás y (b) tareas de fluidez verbal y fonológica. Cada punto representa a un sujeto de la muestra.

Selección de atributos

Posteriormente se realizaron comparaciones entre grupos para detectar en qué variables había diferencias significativas en cada una de las tareas grabadas. Para ello se calculó la significatividad del estadístico U de Mann-Whitney para cada variable, seleccionando aquellas con un valor p<.05. Estas fueron las variables que se usaron posteriormente para predecir el estado del paciente y su puntuación en el MEC

Para las tareas de conteo y descripción de la imagen, el test de Mann-Whitney encontró diferencias significativas entre grupos para todas las variables excepto en la duración media y mediana de la voz, su desviación típica, la suma total de la duración de los silencios, la media y mediana de la duración de los segmentos periódicos, la mediana de los segmentos aperiódicos, la desviación típica de las duraciones de segmentos periódicos y la suma total de duraciones de segmentos aperiódicos. En definitiva, se seleccionaron 27 variables. Las variables para las que las diferencias fueron mayores entre grupos fueron aquellas que aludían a la razón entre medidas (p. ej. cociente entre duración media de voz y silencio), corroborando que los sujetos con deterioro cognitivo tienden a presentar un habla menos fluida y con más pausas.

Para la tarea de fluidez semántica, el análisis estadístico encontró diferencias significativas entre grupos en 12 de las variables: todas excepto la distancia de la 2ª y 3ª palabra y las posiciones relativas de estas palabras. Por otra parte, para la tarea de fluidez fonológica se encontraron diferencias significativas en 14 variables: todas excepto las distancias de las 8ª y 9ª palabras contadas a partir de la 1ª palabra.

Predicción usando técnicas de aprendizaje computacional

En este apartado se presentan los resultados obtenidos en las tareas de clasificación del estado del sujeto (asintomático o con deterioro cognitivo) y la predicción de la puntuación obtenida en el test MEC. Para ello se entrenaron diversas técnicas de aprendizaje computacional usando las variables acústia para las que el análisis estadístico halló diferencias significativas entre grupos. Asimismo, se evaluó el efecto de usar distintos conjuntos de atributos como variables independientes para realizar la predicción en ambas tareas. Para ello se entrenaron 3 conjuntos de modelos con los siguientes atributos:

- Todas las variables seleccionadas por el análisis estadístico (Todos los atributos).
- Únicamente las variables seleccionadas por el análisis estadístico en las tareas de contar y descripción de imagen (Atributos de contar e imagen).
- Únicamente usando como variables el número de palabras dichas por los sujetos en las tareas de fluidez (Atributos de fluidez).

La Tabla 2 muestra los resultados de clasificación (porcentaje de sujetos clasificados correctamente) obtenidos en la tarea de clasificación del estado cognitivo sujeto. Se observa que todas las técnicas de aprendizaje automático son capaces del clasificar correctamente al 100% de los sujetos cuando se usan todos los atributos extraídos en el análisis acústico. Cuando se usan un subconjunto de los atributos, las variables acústicas calculadas en las pruebas de fluidez verbal (atributos de fluidez) muestran ser más eficaces para diagnosticar el estado del sujeto que los extraídos para las tareas de contar y descripción de la lámina del test de Boston. En el caso de los atributos de fluidez, como ya reportó el análisis estadístico, se encuentran diferencias significativas tanto en la tarea de fluidez semántica (6.50±1.84 palabras para GE y 14.00±1.83 palabras para el GC) como en la de fluidez fonológica (2.18±1.18 palabras para GE y 10.40±4.86 palabras para el GC). A pesar de ser menos eficaces que los atributos de fluidez, las variables calculadas en las tareas de contar y descripción de la imagen demuestran también ser buenas predictoras del estado del sujeto, logrando una tasa de clasificación del 90% cuando se usa la técnica de *random forests*.

Tabla 2. *Porcentaje de sujetos correctamente clasificados en función de su estado cognitivo en la tarea de clasificación.*

	Todos los atributos	Atributos de contar e imagen	Atributos de fluidez
Regresión logística	100%	80%	100%
Random forest	100%	90%	100%
Perceptrón multicapa	100%	85%	100%

A continuación, se analizó si las variables acústicas son buenas predictoras de la puntuación obtenida por los sujetos en el test MEC. Para esta tarea se compararon las predicciones realizadas por la técnica de regresión lineal clásica con las predicciones realizadas por dos técnicas de regresión no lineal: los *random forests* y el perceptrón multicapa. Para cada técnica se presentan dos resultados: el coeficiente de correlación de Pearson (*r*) entre el valor real obtenido por el sujeto en el test MEC y el predicho por las técnicas de regresión y la raíz cuadrada del error cuadrático medio (RMSE) entre los valores reales y los predichos. Se usaron estas medidas en lugar del coeficiente de determinación (R_2) al ser las primeras más fiables para evaluar la bondad de ajuste en las técnicas de regresión no lineales (Spiess & Neumeyer, 2010).

La Tabla 3 muestra los resultados obtenidos en la tarea de predicción de las puntuaciones MEC. De nuevo, los atributos de fluidez demuestran ser más eficaces a la hora de realizar la predicción que los atributos acústicos de las tareas de contar e imagen. En cuanto a la comparativa entre técnicas de aprendizaje automático, se observa que las técnicas de regresión no lineal (*random forests* y perceptrón multicapa) obtienen mejores ajustes que la regresión lineal clásica. El mejor resultado se obtiene al usar el perceptrón multicapa y atributos de fluidez, obteniendo una correlación de *r*= .91 y un error de RMSE= 2.34 entre los valores predichos por la técnica y los valores reales obtenidos por los sujetos en el MEC.

Tabla 3. *Resultados obtenidos en la tarea de predicción de las puntuaciones MEC usando distintas técnicas de regresión.*

	Todos los atributos		Atributos de contar e imagen		Atributos de fluidez	
	r	RMSE	*R*	RMSE	*r*	RMSE
Regresión lineal	.32	6.76	.65	4.91	.78	3.71
Random forest	.88	2.96	.68	4.23	.89	2.62
Perceptrón multicapa	.85	3.18	.71	4.06	.91	2.34

Discusión

El objetivo de este estudio era determinar si el análisis automático de voz es útil para la detección precoz del deterioro cognitivo. Para ello se recogieron muestras de grabaciones de la voz de 20 sujetos, 10 con deterioro cognitivo y 10 asintomáticos, mientras realizaban 4 tareas de distinta complejidad cognitiva.

En primer lugar se corroboraron las diferencias entre los grupos en las puntuaciones del test de *screening* de demencia MEC, lo cual era esperable dado el diagnóstico de los participantes. La muestra

seleccionada para este estudio, por tanto, es apropiada ya que sólo se diferencian en relación al deterioro cognitivo que presenta el GE en comparación con el GC, no habiendo diferencias entre ellos en la edad o el nivel de estudios. Asimismo, el análisis estadístico halló diferencias significativas entre grupos en casi todas las variables acústicas analizadas en cada una de las tareas grabadas. Esto viene a confirmar que los sujetos con DCL presentan alteraciones en su habla y que el deterioro cognitivo (o demencia, según el caso) afecta a sus capacidades cognitivas, haciendo que haya diferencias en las tareas de fluidez. Esto pone de manifiesto el potencial de estas medidas en la valoración de los pacientes con estas características tal y como se demuestra en estudios previos (Canning, Leach, Stuss, Ngo, & Black, 2004; Forbes-McKay & Venneri, 2005; König et al., 2015).

En consonancia con lo hallado en otros estudios recientes (König et al., 2015; Mirheidari et al., 2016; Satt et al., 2014; Weiner et al., 2016; Weiner & Schultz, 2018), nuestro estudio viene a mostrar que las tecnologías del habla pueden ser unas herramientas valiosas tanto en la detección del deterioro cognitivo y como en la monitorización de su progresión. En este sentido se observa que las técnicas de aprendizaje automático son capaces de clasificar el 100% de participantes según tengan deterioro cognitivo o no usando las variables acústicas extraídas de las grabaciones de estas tareas. Dicho de otra forma, estas variables proporcionan la información suficiente para clasificar a los participantes en sus grupos de diagnóstico con una gran precisión. Esto va en la línea de estudios recientes (Godino-Llorente & Gomez-Vilda, 2004; König et al., 2015; Linz et al., 2017; Satt et al., 2014) que apuntan que la tecnología de procesamiento de la voz podría ser un valioso método de apoyo para el personal clínico, el cual demanda herramientas objetivas y automatizadas adicionales para el diagnóstico del deterioro cognitivo.

En las tareas de conteo hacia atrás y descripción de la imagen las características óptimas fueron aquellas que reflejaron la continuidad del habla. Así, el análisis estadístico encontró una mayor continuidad del habla en los participantes asintomáticos y una menor para el grupo con deterioro cognitivo, mostrando de esta forma en los sujetos del grupo de control segmentos de voz continuos más largos y segmentos de silencio más cortos y segmentos periódicos continuos más largos y segmentos aperiódicos más cortos. Esto concuerda con lo hallado en otros estudios similares (König et al., 2015; Satt et al., 2014). Por otro lado, en las tareas de fluidez verbal semántica y fonológica la mayor contribución a la precisión de la clasificación se obtuvo de las posiciones (en el tiempo) de las palabras y el número de palabras emitidas por los participantes en el tiempo que duraba la tarea. Como aparece recogido en la revisión de Taler y Phillips, (2008) los déficits de lenguaje encontrados en sujetos con DCL y EA posiblemente podrían ser producto del deterioro del conocimiento semántico.

Otro de los resultados relevantes obtenidos en el presente estudio es que las variables acústicas calculadas en las tareas de fluidez verbal semántica y fonológica demuestran ser más eficaces para predecir el estado del sujeto que las variables analizadas en las tareas de conteo y de descripción de la imagen. Esto podría deberse a la complejidad cognitiva de las tareas.

Asimismo, los análisis de predicción de las puntuaciones MEC reflejan que las variables acústicas son también eficaces para evaluar el estado cognoscitivo del sujeto con deterioro cognitivo. En este sentido, se encontró que las predicciones realizadas por las técnicas de aprendizaje automático a partir de las

variables tienen una alta correlación (hasta *r*=0.91 cuando se emplean perceptrones multicapa) con las notas reales obtenidas por los sujetos en este test. Esto concuerda con lo hallado en otros estudios recientes. Así, en Bucks, Singh, Cuerden y Wilcock, (2000) se encontró que distintos parámetros acústicos extraídos de grabaciones de pacientes con EA tenían una alta relación con la nota del examen del test MEC en su versión inglesa.

A pesar de los valiosos resultados hallados en este estudio, también presenta algunas limitaciones entre las que se encuentra el número de participantes, el cual es escaso debido a la dificultad de conseguir participantes que cumplieran con los requisitos exigidos. No obstante, para futuros estudios se espera ampliar la muestra con el fin de obtener mayor representatividad. Además, se espera conseguir mayor homogeneidad en la muestra en cuanto al diagnóstico y poder realizar comparaciones entre grupos con Alzheimer o DCL. Por otro lado, cabe destacar que las diferentes tareas también requieren niveles de esfuerzo cognitivo considerables y diversos, por ejemplo, contar hacia atrás es cognitivamente más exigente que describir una imagen por lo tanto, estas diferencias podrían explicar en parte la alta sensibilidad de los análisis de voz.

Conclusiones

Tras la investigación realizada, se concluye que el análisis automático de la voz parece una herramienta eficaz en la detección del deterioro cognitivo. Pensamos que la detección precoz de estos trastornos podría permitir una intervención más temprana con los consecuentes beneficios para el paciente. En definitiva, el uso de programas simples como el de análisis de voz podría facilitar la evaluación del lenguaje oral en los parámetros específicos requeridos y contribuir al diagnóstico del deterioro cognitivo para, desde el punto de vista logopédico, poder comenzar cuando antes un tratamiento con el fin de ralentizar el progreso de la enfermedad, sobre todo en el ámbito del lenguaje y memoria.

En un futuro, sería interesante ampliar el alcance de la investigación, recopilando datos en una escala más amplia y utilizando tareas cognitivas que sean más desafiantes, como la descripción de un recuerdo (vida cotidiana). Esperamos que las nuevas propuestas de tareas cognitivas aumenten la precisión discriminativa.

Bibliografía

Acarin, N. (2010). *Alzheimer: Manual de instrucciones*. Madrid: RBA.

Auriacombe, S., Lechevallier, N., Amieva, H., Harston, S., Raoux, N., & Dartigues, J.-F. (2006). A Longitudinal Study of Quantitative and Qualitative Features of Category Verbal Fluency in Incident Alzheimer's Disease Subjects: Results from the PAQUID Study. *Dementia and Geriatric Cognitive Disorders*, *21*(4), 260–266. https://doi.org/10.1159/000091407

Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., … Reynolds, D. A. (2004). A tutorial on text-independent speaker verification. *EURASIP Journal on Advances in Signal Processing*, *2004*(4), 430–451.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bucks, R. S., Singh, S., Cuerden, J. M., & Wilcock, G. K. (2000). Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance. *Aphasiology*, *14*(1), 71–91. https://doi.org/10.1080/026870300401603

Canning, S. J. D., Leach, L., Stuss, D., Ngo, L., & Black, S. E. (2004). Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia. *Neurology*, *62*(4), 556–562. https://doi.org/10.1212/WNL.62.4.556

Delgado, J., León, N. M., Jiménez, A., & Izquierdo, L. M. (2017). Análisis acústico de la voz: Medidas temporales, espectrales y cepstrales en la voz normal con el Praat en una muestra de hablantes de español. *Revista de Investigación En Logopedia*, *7*(2), 108–127.

Donoso, A., & Vásquez, C. (2002). Deterioro Cognitivo y Enfermedad de Alzheimer: Presentación de dos Casos. *Revista de Psicología*, *11*(1), 9–16.

Facal, D., González, M. F., Buiza, C., Laskibar, I., Urdaneta, E., & Yanguas, J. J. (2009). Envejecimiento, deterioro cognitivo y lenguaje: Resultados del Estudio Longitudinal Donostia. *Revista de Logopedia, Foniatría y Audiología*, *29*(1), 4–12. https://doi.org/10.1016/S0214-4603(09)70138-X

Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences*, *26*(4), 243–254. https://doi.org/10.1007/s10072-005-0467-9

Godino-Llorente, J. I., & Gomez-Vilda, P. (2004). Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Transactions on Biomedical Engineering*, *51*(2), 380–384. https://doi.org/10.1109/TBME.2003.820386

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston Diagnostic Aphasia Examination (3a ed.)*. Austin, TX, USA: Pro-Ed.

Horley, K., Reid, A., & Burnham, D. (2010). Emotional Prosody Perception and Production in Dementia of the Alzheimer's Type. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1132–1146. https://doi.org/10.1044/1092-4388(2010/09-0030)

Junqué, C., & Jurado, M. A. (2009). *Envejecimiento, demencias y otros procesos degenerativos*. Madrid, España: Síntesis.

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., … David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring*, *1*(1), 112–124. https://doi.org/10.1016/j.dadm.2014.11.012

Kunz, M., Seuss, D., Hassan, T., Garbas, J. U., Siebers, M., Schmid, U., … Lautenbacher, S. (2017). Problems of video-based pain detection in patients with dementia: A road map to an interdisciplinary solution. *BMC Geriatrics*, *17*(1), 33. https://doi.org/10.1186/s12877-017-0427-2

Linz, N., Tröger, J., Alexandersson, J., Wolters, M., König, A., & Robert, P. (2017). Predicting Dementia Screening and Staging Scores from Semantic Verbal Fluency Performance. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 719–728. https://doi.org/10.1109/ICDMW.2017.100

Lobo, A., Escobar, V., Ezquerra, J., & Seva Díaz, A. (1980). 'El Mini-Examen Cognoscitivo' (Un test sencillo, práctico, para detectar alteraciones intelectuales en pacientes psiquiátricos). *Revista de Psiquiatría y Psicología Médica*, *5*, 39-57.

Lopez, O. L., Jagust, W. J., DeKosky, S. T., Becker, J. T., Fitzpatrick, A., Dulberg, C., … Kuller, L. H. (2003). Prevalence and classification of mild cognitive impairment in the Cardiovascular Health Study Cognition Study: Part 1. *Archives of Neurology*, *60*(10), 1385–1389. https://doi.org/10.1001/archneur.60.10.1385

López-de-Ipiña, K., Alonso, J.-B., Travieso, C., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., … Lizardui, U. M. de. (2013). On the selection of non-invasive methods based on speech analysis oriented to automatic alzheimer disease diagnosis. *Sensors*, *13*(5), 6730–6745. https://doi.org/10.3390/s130506730

Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(Nov), 2579–2605.

Marino, J. C., & Alderete, A. M. (2009). Variación de la actividad cognitiva en diferentes tipos de pruebas de fluidez verbal. *Revista Chilena de Neuropsicología*, *4*(2), 179–192.

Martínez-Sánchez, F., Meilán, J., Pérez, E., Carro, J., & Arana, J. (2012). Expressive prosodic patterns in individuals with Alzheimer's Disease. *Psicothema*, *24*(1), 16–21.

Migliacci, M. L., Scharovsky, D., & Gonorazky, S. E. (2009). Deterioro cognitivo leve: Características neuropsicológicas de los distintos subtipos. *Revista de Neurología*, *48*(5), 237–241.

Mirheidari, B., Blackburn, D., Reuber, M., Walker, T., & Christensen, H. (2016). Diagnosing People with Dementia Using Automatic Conversation Analysis. *Proc. Interspeech 2016*, 1220–1224. https://doi.org/10.21437/Interspeech.2016-857

Mirheidari, B., Blackburn, D., Walker, T., Venneri, A., Reuber, M., & Christensen, H. (2018). Detecting Signs of Dementia Using Word Vector Representations. *Proc. Interspeech 2018*, 1893–1897. https://doi.org/10.21437/Interspeech.2018-1764

Mora-Simón, S., García-García, R., Perea-Bartolomé, M. V., Ladera-Fernández, V., Unzueta-Arce, J., Patino-Alonso, M. C., & Rodríguez-Sánchez, E. (2012). Deterioro cognitivo leve: Detección temprana y nuevas perspectivas. *Revista de Neurología*, *54*(05), 303–310. https://doi.org/10.33588/rn.5405.2011538

Pakhomov, S. V. S., Eberly, L., & Knopman, D. (2016). Characterizing cognitive performance in a large longitudinal study of aging with computerized semantic indices of verbal fluency. *Neuropsychologia*, *89*, 42–56. https://doi.org/10.1016/j.neuropsychologia.2016.05.031

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Kokmen, E., & Tangelos, E. G. (1997). Aging, memory, and mild cognitive impairment. *International Psychogeriatrics*, *9*(S1), 65–69. https://doi.org/10.1017/S1041610297004717

Raoux, N., Amieva, H., Le Goff, M., Auriacombe, S., Carcaillon, L., Letenneur, L., & Dartigues, J.-F. (2008). Clustering and switching processes in semantic verbal fluency in the course of Alzheimer's disease subjects: Results from the PAQUID longitudinal study. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *44*(9), 1188–1196. https://doi.org/10.1016/j.cortex.2007.08.019

Reisberg, B., Ferris, S. H., de Leon, M. J., & Crook, T. (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *The American Journal of Psychiatry*, *139*(9), 1136–1139. https://doi.org/10.1176/ajp.139.9.1136

Sáenz-Lechón, N., Godino-Llorente, J. I., Osma-Ruiz, V., & Gómez-Vilda, P. (2006). Methodological issues in the development of automatic systems for voice pathology detection. *Biomedical Signal Processing and Control*, *1*(2), 120–128. https://doi.org/10.1016/j.bspc.2006.06.003

Satt, A., Hoory, R., König, A., Aalten, P., & Robert, P. (2014). Speech-Based Automatic and Robust Detection of Very Early Dementia. *Proc. Interspeech*, 2538--2542. https://doi.org/10.13140/2.1.1258.8805

Serrano, C. M., Dillon, C., Leis, A., Taragano, F. E., & Allegri, R. F. (2013). Deterioro cognitivo leve: Riesgo de demencia según subtipos. *Actas Españolas de Psiquiatría*, *41*(6), 330–339.

Spiess, A.-N., & Neumeyer, N. (2010). An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach. *BMC Pharmacology*, *10*(1), 6. https://doi.org/10.1186/1471-2210-10-6

Taler, V., & Phillips, N. A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology*, *30*(5), 501–556. https://doi.org/10.1080/13803390701550128

Velasquez-Perez, L., Guerrero-Camacho, J., Rodriguez-Agudelo, Y., Alonso-Vilatela, M. E., & Yescas-Gomez, P. (2008). Conversion of mild cognitive impairment to dementia. *Revista Ecuatoriana De Neurologia*, *17*(1–3), 25–32.

Weiner, J., Engelbart, M., & Schultz, T. (2017). Manual and Automatic Transcriptions in Dementia Detection from Speech. *Proc. Interspeech 2017*, 3117–3121. https://doi.org/10.21437/Interspeech.2017-112

Weiner, J., Herff, C., & Schultz, T. (2016). Speech-Based Detection of Alzheimer's Disease in Conversational German. *Proc. Interspeech 2016*, 1938–1942. https://doi.org/10.21437/Interspeech.2016-100

Weiner, J., & Schultz, T. (2018). Selecting Features for Automatic Screening for Dementia Based on Speech. *International Conference on Speech and Computer*, 747–756.

Yancheva, M., Fraser, K., & Rudzicz, F. (2015). Using linguistic features longitudinally to predict clinical scores for Alzheimer's disease and related dementias. *Proc. 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*.

Apéndice A – Cálculo de las variables acústicas

En este apéndice se detallan, en primer lugar, las variables acústicas calculadas para cada tarea de este estudio y, posteriormente, se describe la técnica de VAD empleada para la detección de voz y silencios en las grabaciones.

Tareas de fluidez verbal

De cada grabación realizada por los participantes en las tareas de fluidez verbal se calcularon las siguientes 17 variables acústicas:

- El número total de palabras emitidas por el sujeto en el minuto que duraba la tarea (1 variable).
- La diferencia en segundos entre el instante de comienzo de la primera palabra emitida por el sujeto y el resto de palabras, contando como mucho 9 palabras (8 variables).
- La posición relativa en un intervalo [0, 1] del comienzo de las 9 primeras palabras emitidas por el sujeto, tomando como posición relativa cero el comienzo de la primera palabra y posición relativa uno el final de la grabación (1 minuto) (8 variables).

El recuento de palabras emitidas en cada grabación y el etiquetado del tiempo de inicio de las palabras se llevó a cabo de forma manual usando el software Audacity versión 2.1.0.

Tareas de conteo hacia atrás y descripción de la lámina del Test de Boston

En las grabaciones realizadas para estas dos tareas se calculó, por un lado, las energías logarítmicas en decibelios (dBs) a nivel de trama (ventana de análisis tipo Hamming con una longitud de 20 ms solapadas 10 ms y un factor de preénfasis igual a 0.97) y, por otro, la frecuencia fundamental de la voz (F0), también a nivel de trama. Ambas medidas se calcularon usando el programa Wavesurfer versión 1.8.8 para MacOS.

En base a las medidas de energía logarítmica obtenidas se aplicó la técnica de VAD descrita más abajo para decidir qué tramas contenían voz y cuáles silencio (no voz). Por otro lado, el valor de F0 se utilizó para clasificar las tramas como segmentos periódicos (F0 > 0) o aperiódicos (F0 = 0), en función de vibración de las cuerdas vocales. Una vez clasificadas las tramas, se calculó de forma automática la duración en segundos de los segmentos del mismo tipo (voz, silencio, periódicos y aperiódicos). En resumen, para cada grabación se obtuvieron una lista con las duraciones de los segmentos de voz, segmentos de silencio, segmentos periódicos y, finalmente, segmentos aperiódicos.

A partir de la lista de duraciones de los segmentos se calcularon las siguientes 36 variables acústicas para cada grabación:

1. Número total de segmentos de cada tipo (voz, silencio, periódicos, aperiódicos) en cada grabación (4 variables por cada grabación).
2. Promedio de las duraciones en segundos para cada tipo de segmento (4 variables por grabación).
3. Mediana de las duraciones en segundos para cada tipo de segmento (4 variables por grabación).

4. Desviación estándar de las duraciones en segundos para cada tipo de segmento (4 variables por grabación).
5. Suma de las duraciones en segundos para cada tipo de segmento (4 variables por grabación).
6. Cociente entre las duraciones promedio para cada par de tipos de segmento, esto es, las siguientes 4 variables:
    a. duración promedio de los segmentos tipo voz/duración promedio de los segmentos tipo silencio,
    b. duración promedio de los segmentos tipo silencio/duración promedio de los segmentos tipo voz,
    c. duración promedio de los segmentos periódicos/duración promedio de los segmentos aperiódicos y
    d. duración promedio de los segmentos aperiódicos/duración promedio de los segmentos periódicos.
7. Lo mismo que el punto 6 pero con las medianas de las duraciones obtenidas en el punto 3 (4 variables por grabación).
8. Lo mismo que el punto 6 pero con las desviaciones típicas de las duraciones obtenidas en el punto 4 (4 variables por grabación).
9. Lo mismo que el punto 6 pero con las suma de las duraciones obtenidas en el punto 5 (4 variables por grabación).

Detección de actividad de voz (VAD)

Para la detección de los segmentos de voz y no voz en las grabaciones se utilizó la técnica de VAD propuesta en (Bimbot et al., 2004) y que, de forma breve, se describe a continuación:

1. En primera lugar, se calcularon las energías logarítmicas en decibelios (dB) a nivel de trama (ventana de análisis tipo Hamming con una longitud de 20 ms solapadas 10 ms y un factor de preénfasis igual a 0.97) usando el programa Wavesurfer versión 1.8.8 para MacOS.
2. Se ajustó un modelo de mezcla de distribuciones normales univariadas con 2 componentes (modelo bi-Gaussiano) a los datos de energía logarítmica calculados en el paso 1 para cada grabación.
3. La componente del modelo con menor valor para la media se consideró que modelaba los segmentos de silencio (no voz), mientras que aquella con mayor valor absoluto en la media se consideró que modelaba los segmentos de voz.
4. La decisión sobre qué segmentos de la grabación se consideraban silencio o voz se basó en un umbral θ calculado de forma automática como el punto medio en donde las dos distribuciones de probabilidad (voz y silencio) se cortan. Este umbral se corresponde con la raíz real situada entre las medias de ambas distribuciones de la siguiente ecuación cuadrática:

$$f(x) = ax^2 + bx + c$$

donde

$$a = \frac{1}{2\sigma_s^2} - \frac{1}{2\sigma_v^2}$$
$$b = \frac{\mu_v}{\sigma_v^2} - \frac{\mu_s}{\sigma_s^2}$$
$$c = \frac{\mu_s^2}{2\sigma_s^2} - \frac{\mu_v^2}{2\sigma_v^2} + \ln\frac{\sigma_s}{\sigma_v}$$

siendo $\mathcal{N}(\mu_s, \sigma_s)$ y $\mathcal{N}(\mu_v, \sigma_v)$ las distribuciones normales que modelan el silencio y la voz (paso 2), respectivamente.

5. Se consideraron como silencio aquellas tramas (segmentos de 10 ms) obtenidas en el paso 1 cuya energía fuese inferior al umbral θ y como voz las tramas con energía superior a este umbral.
6. Finalmente, sobre las decisiones tomadas en el paso 5 se aplicó un filtro de mediana de longitud 0.1 s centrado en cada muestra El objetivo de este filtro fue robustecer la clasificación voz/silencio ante posibles ruidos espontáneos.