





Correlación entre espacios de características acústicas del habla y trastornos clínicos de la voz en pacientes con disfagia

Correlation Between Speech-Related Feature Spaces and Clinical Voice Disorders in Patients with Dysphagia

 Andrés Felipe Flórez-Gómez¹;
 Juan Rafael Orozco-Arroyave²;
  Sebastián Roldán-Vasco³

¹ Instituto Tecnológico Metropolitano, Medellín-Colombia,
andresflorez223360@correo.itm.edu.co

² Universidad de Antioquia, Medellín-Colombia, Friedrich-Alexander-Universität, Erlangen Nürnberg-Alemania,
rafael.orozco@udea.edu.co

³ Instituto Tecnológico Metropolitano, Universidad de Antioquia, Medellín-Colombia,
sebastianroldan@itm.edu.co

ISSN-p: 0123-7799
ISSN-e: 2256-5337

Vol. 24, nro. 53, e2220, 2021

Recibido: 28 octubre 2021
Aceptado: 30 marzo 2022
Disponible: 5 abril 2022

©Instituto Tecnológico Metropolitano
Este trabajo está licenciado bajo
una Licencia Internacional
Creative Commons Atribución
(CC BY-NC-SA)



Cómo citar / How to cite

A. F. Flórez-Gómez; J. R. Orozco-Arroyave; S. Roldán-Vasco, “Correlación entre espacios de características acústicas del habla y trastornos clínicos de la voz en pacientes con disfagia”, *TecnoLógicas*, vol. 25, nro. 53, e2220, 2022. <https://doi.org/10.22430/22565337.2220>

Resumen

La disfagia se define como la dificultad para transportar un bolo alimenticio de forma segura y efectiva desde la cavidad oral hasta el estómago. En la actualidad, los métodos para el diagnóstico de la disfagia son invasivos y altamente dependientes de la experiencia del personal asistencial cualificado. El estudio de las bioseñales, como lo son los registros de voz y habla, ha sido propuesto con el fin de desarrollar herramientas complementarias al diagnóstico. De esta manera, el presente trabajo tuvo como objetivo explorar, en características extraídas en señales de voz y habla, la capacidad de discriminación entre personas sanas y pacientes con trastornos deglutorios. Para ello se registraron señales en un grupo de 30 personas sanas y 45 pacientes diagnosticados con disfagia. Los participantes realizaron diferentes tareas de voz (vocales sostenidas) y de habla (texto leído, monólogo y ejercicios diadococinéticos). Los registros de los pacientes fueron etiquetados en tres condiciones clínicas: voz húmeda, voz disfónica y voz con alteración no determinada. Se evaluaron espacios de características clásicas asociadas al análisis de voz y habla a través de pruebas estadísticas hallándose que las características relacionadas a la fonación, prosodia y diadococinesia tienen potencial como biomarcadores para la discriminación de diferentes alteraciones en pacientes con disfagia. Este trabajo constituye una aproximación preliminar basada en el estudio de señales de voz y habla para un diagnóstico no invasivo y objetivo de la disfagia.

Palabras clave

Disfagia, análisis de voz, análisis del habla, procesamiento de bioseñales, extracción de características, análisis estadístico.

Abstract

Dysphagia is defined as the difficulty to transport an alimentary bolus from the oral cavity to the stomach in a safe and effective way. Currently, dysphagia-related diagnosis methods are invasive and highly dependent on the examiner's experience. Biosignal-based studies, such as those on voice and speech records, have been proposed to develop complementary diagnostic tools. Likewise, this study explores, in features extracted from voice and speech signals, the capacity to discriminate between healthy subjects and patients with swallowing disorders. For this purpose, the signals were recorded in a group of 30 healthy individuals and 45 dysphagic patients. The participants performed different voice tasks (sustained vowels) and speech tasks (text reading, monologue, and diadochokinetic exercises). The patient records were assigned labels of three clinical conditions: wet voice, dysphonic voice, and voice with undetermined alteration. Classical voice- and speech-related feature spaces were assessed using statistical tests, and it was found that features related to phonation, prosody, and diadochokinesia have potential as biomarkers for the discrimination of different alterations in patients with dysphagia. This is a preliminary study based on voice and speech signals for a non-invasive and objective diagnosis of dysphagia.

Keywords

Dysphagia, Speech analysis, Voice analysis, Biosignal processing, Feature extraction, Statistical analysis.

1. INTRODUCCIÓN

1.1 Disfagia y diagnóstico

La disfagia es la alteración del proceso deglutorio, lo que puede generar desnutrición, deshidratación, neumonía e incluso la muerte [1]. Sus causas se dividen en dos grupos: trastornos neuromotores y lesiones obstructivas [2]. Las personas con esta afección han tenido que adoptar distintas medidas en sus hábitos alimenticios, de sueño y del habla para poder contrarrestar los efectos de la disfagia, algo que no solo afecta al paciente, sino también a su círculo familiar y social inmediato [3]. Está asociada a diversas enfermedades neurológicas de base y tiene incidencia, principalmente, en la población de edad avanzada [4].

En Colombia, la población mayor a 59 años pasó de dos millones en el año 1985, a seis millones en 2018, con un crecimiento anual del 3,5 %, superior al 1,7 % de la población total, y se proyecta un aumento de esta población debido a la disminución de la mortalidad y al aumento de la expectativa de vida [5]. Esta población es vulnerable a distintas enfermedades o afecciones como lo son la enfermedad de Parkinson (EP), Alzheimer, esclerosis lateral amiotrófica (ELA), entre otras, las cuales aumentan la incidencia y prevalencia de la disfagia [4].

Los métodos validados clínicamente para la evaluación instrumental de la disfagia son la videofluoroscopia y la endoscopia de fibra óptica (VFSS y FEES, por sus siglas en inglés, respectivamente) [6]. Sin embargo, la VFSS es la prueba de referencia aceptada clínicamente [4]. Ambas técnicas son invasivas y dependientes de la experticia y capacitación del personal que las utiliza [7]. La VFSS, además de ser invasiva, es costosa –con un valor aproximado de medio millón de pesos colombianos–, lo que tiene un impacto sobre el sistema de salud.

Debido a las limitaciones presentadas arriba, distintas investigaciones han propuesto aproximaciones basadas en señales no invasivas (ej. electromiografía de superficie [8] o auscultación cervical con acelerometría [9]), aunque aún no han sido implementadas en el consultorio para tratamiento, diagnóstico o seguimiento de la disfagia. Una de las aproximaciones que podrían complementar el diagnóstico es el estudio de las señales de voz y habla, debido a que estas guardan estrecha relación, tanto a nivel estructural como neurológico, con el proceso deglutorio [10], [11].

Debido a esta relación, algunos pacientes con disfagia presentan trastornos de la voz y del habla, tales como disfonía y la voz húmeda. La disfonía ocurre cuando los músculos alrededor de la laringe se encuentran muy tensos durante el habla y se manifiesta como voz ronca, débil, tensa, entrecortada y/o áspera. Algunos pacientes manifiestan además opresión e incluso dolores musculares en la garganta [12]. Por otra parte, la voz húmeda es la presencia de material extraño en el área de la laringe, normalmente restos de alimentos sólidos o líquidos, y se caracteriza por ser una voz gorgoja o con sonidos guturales anormales [13], [14].

1.2 Análisis de la voz y del habla en pacientes con disfagia

El análisis de voz se refiere al estudio de las características acústico-fonatorias en vocales sostenidas. Por otro lado, el análisis del habla hace referencia al estudio de las características asociadas al habla continua.

Estos análisis se han abordado preliminarmente en estudios de pacientes con disfagia y otras comorbilidades [15]. Lo anterior debido a que este tipo de análisis presenta baja intrusión y fácil implementación en la práctica clínica. Una correlación significativa entre las características de voz contrastada con el análisis de VFSS podría llegar a avances

importantes en un diagnóstico más simple y menos estresante para los pacientes, además de un menor costo e impacto en el sistema de salud [16]. Además, cuando se logra una automatización lo suficientemente robusta, la evaluación basada en el habla se podría llevar a cabo de manera remota, reduciendo el costo para los pacientes y ampliando el espectro de atención hacia personas que viven en áreas rurales.

En 2004, se realizó un estudio en 93 pacientes para determinar el valor diagnóstico del análisis de voz para la detección de aspiraciones laríngeas en pacientes con alto riesgo [17]. Se midieron cinco variables acústicas pre y pos VFSS: el valor medio de la frecuencia fundamental de vibración de los pliegues vocales (F_0), la perturbación promedio relativa (RAP, por sus siglas en inglés), el porcentaje de Shimmer, la relación ruido-armónico y el índice de turbulencia de voz. Se observó que el RAP, el Shimmer, la relación ruido-armónico y el índice de turbulencia permiten detectar pacientes con alto riesgo de aspiración, en especial el RAP.

En una investigación posterior [15], se analizaron señales de voz obtenidas de pacientes con desórdenes neurológicos de diferentes etiologías -EP, accidente cerebrovascular y ELA-, bajo tres condiciones de evaluación: actividades predeglutorias, después de tragar una solución líquida y después de tragar una solución pastosa. Adicionalmente, utilizaron tres enfoques: análisis acústico, basado en F_0 , Jitter y Shimmer; un método basado en dinámica no lineal en vocales; y un análisis de entropía relativa entre los grupos de señales. Los autores encontraron que únicamente el análisis de dinámica no lineal arrojó diferencias estadísticamente significativas en los grupos.

Otro estudio evaluó un grupo de pacientes con disfagia y un grupo de control en el que se analizó la producción vocal a partir del registro de la vocal sostenida 'a' antes y después de la ingesta de sustancias pastosas mediante la escala GRBAS –siglas en inglés para grado, aspereza, respiración, astenia y tensión– y el parámetro de voz húmeda. Los autores observaron una disminución significativa en el grado y la astenia y un aumento importante de la tensión en las personas con disfagia, mientras que los controles no presentaron variación en estos parámetros después de la deglución. Por otra parte, la voz húmeda no mostró variaciones entre los grupos [18].

En el 2016 se realizó un estudio en el que extrajeron características lineales y no lineales en señales de habla continua, que pueden ser interpretados como indicadores o predictores confiables en la determinación de alteraciones de deglución en pacientes con EP, los cuales también presentaron un alto riesgo de aspiración relacionado a la disfagia. Los autores extrajeron múltiples características como la clasificación sonora-insonora, la relación ruido-armónico, la frecuencia fundamental y sus variaciones, Jitter, Shimmer, coeficientes cepstrales de frecuencia en la escala Mel (MFCC, por sus siglas en inglés), y la entropía de permutación. Se halló que, además de la frecuencia fundamental, se presentaron fuertes variaciones en la relación ruido-armónico y en la entropía de permutación en pacientes después de la deglución de una solución pastosa [16].

1.3 Contribuciones del trabajo

En este trabajo se buscó establecer una correlación entre las características del habla y condiciones de voz húmeda y disfonía, en pacientes diagnosticados con disfagia. Para ello, se realizó un análisis de las dimensiones del habla con el fin de contribuir a un diagnóstico de la disfagia oportuno, no invasivo, con menor sesgo del evaluador y costo-efectivo.

2. MATERIALES Y MÉTODOS

2.1 Base de datos

Se construyó una base de datos con registros de voz de 45 pacientes con disfagia (23 hombres y 22 mujeres, edad promedio de $59,7 \pm 12,3$ y 30 personas de control (13 hombres y 17 mujeres, edad promedio de $63,2 \pm 9,9$ años). Los voluntarios firmaron consentimiento informado aprobado por el Comité de Ética de la Universidad Pontificia Bolivariana, el cual está definido siguiendo los lineamientos del acuerdo de Helsinki.

La muestra de pacientes se dividió en tres subgrupos diferenciados por condición clínica de la voz: voz húmeda (VH), voz disfónica (VDf) y voz normal o no determinada (ND). Dicha condición es etiquetada por una fonoaudióloga con experiencia certificada en pacientes con disfagia. La Tabla 1 muestra la distribución demográfica de la base de datos, en donde la edad está dada en años y presentada en media y desviación estándar.

Tabla 1. Distribución demográfica de la base de datos. Fuente: elaboración propia.

Grupo	Condición de Voz	Sexo [M / F]	Edad [M / F] (años)
Pacientes	VH	3 / 4	$67,33 \pm 5,51$ / $53,75 \pm 9,53$
	VDf	13 / 10	$60,85 \pm 12,93$ / $65,10 \pm 8,10$
	ND	7 / 8	$54,57 \pm 10,55$ / $55,63 \pm 17,25$
Controles	-	16 / 14	$63,88 \pm 10,46$ / $62,36 \pm 9,45$

Nota: VH: voz húmeda; VDf: voz disfónica; ND: condición no determinada

Con el fin de evitar sesgos, el grupo de control se dividió en otros dos, uno con siete voluntarios para comparaciones con el grupo de VH y otro con 23 voluntarios para comparaciones con el grupo de VDf. La selección de ambos subgrupos fue realizada teniendo en cuenta un emparejamiento en términos de edad y género (prueba t con $p = 0,82$, prueba χ^2 con $p = 1,00$ para comparaciones con VH; prueba t con $p = 0,7$, prueba, χ^2 con $p = 1,00$ para comparaciones con VDf).

2.2 Protocolo de grabación

En este trabajo se pretende cubrir el análisis de voz y habla en pacientes con disfagia. Para el análisis de voz se consideró la producción de vocales sostenidas, y, para el análisis del habla, se consideraron tareas de habla continua y diadococinesia. Para la adquisición de las señales se utilizó la diadema Logitech H390. El micrófono se posicionó a una distancia aproximada de 5 cm de la boca, como se observa en la Figura 1.

Las grabaciones fueron adquiridas con el *software* Audacity con una frecuencia de muestreo de 44,1 kHz y exportadas en formato .wav de 16 bits con signo.

Se solicitó a cada voluntario realizar las siguientes tareas de voz orientadas a la evaluación de tres dimensiones del habla: fonación, articulación y prosodia [19]:

Vocales sostenidas: consistió en pedirle al participante que pronunciara las vocales de forma sostenida durante al menos tres segundos o hasta que se le agote el aire:

\a\, \e\, \i\, \o\, \u\. Repitiendo el ejercicio tres (3) veces.



Figura 1. Posicionamiento de la diadema. Fuente: elaboración propia.

Tareas diadococinéticas (DDK): se le solicitó al participante que repitiera rápidamente las siguientes palabras y sílabas:

- | | |
|---------------|---------|
| a) \pa-ta-ka\ | d) \pa\ |
| b) \pa-ka-ta\ | e) \ta\ |
| c) \pe-ta-ka\ | f) \ka\ |

Lectura: se le requirió al participante que leyera el siguiente párrafo, haciendo las respectivas pausas y entonaciones donde indiquen los signos de puntuación, interrogación y exclamación:

“Ayer fui al médico. ¿Qué le pasa? Me preguntó. Yo le dije: ¡Ay, doctor! Donde pongo el dedo me duele. ¿Tiene la uña rota? Sí. Pues ya sabemos qué es. Deje su cheque a la salida”.

Esta lectura es fonéticamente balanceada y contiene todos los fonemas del español hablado en Colombia [20].

Monólogo: consistió en que el participante hablara acerca de las actividades que realiza en un día común.

Las señales de los pacientes fueron etiquetadas por el experto en fonoaudiología de acuerdo con sospecha de voz húmeda, disfonía o voz con alteración no determinada.

2.3 Preprocesamiento

Con el fin de eliminar sesgo a causa del ambiente acústico, se aplicó una normalización de canal basada en la compresión GSM full-rate, utilizando el *software* Sound eXchange (SoX). Su tasa de bits se redujo a 13 bps, submuestreó a 8 kHz, y se filtró entre 0,2 kHz y 3,4 kHz.

2.4 Extracción de características

Se extrajeron distintas características de los registros de habla asociados a las dimensiones del habla mediante Python y la librería Parselmouth [21], la cual permite

acceder de manera indirecta a la interfaz de Praat [22]. A continuación, se describen las características asociadas a cada dimensión del habla.

Fonación: estudiada principalmente en las vocales sostenidas, se entiende como el análisis que se realiza al primer momento de la producción vocal [23]. En ocasiones referido como el análisis acústico de la voz, los descriptores que se evaluaron en este estudio son: F_0 , Jitter, Shimmer, los cocientes de perturbación temporal o de amplitud (PPQ y APQ, respectivamente) y la intensidad de la señal o energía en forma logarítmica, estudiados en [19].

F_0 es la frecuencia fundamental de vibración de los pliegues vocales [24]. Para hallarla, se utilizó el método desarrollado por Paul Boersma [25], [26]. Adicionalmente, se hallaron la primera y segunda derivadas de F_0 , es decir, ΔF_0 y $\Delta^2 F_0$, respectivamente, donde ΔF_0 representa el cambio entre ventanas, mientras que $\Delta^2 F_0$ representa el cambio entre ventanas de la primera derivada, que se interpretan como la velocidad y la aceleración de la característica, respectivamente [27], [28].

La intensidad o energía de la señal es extraída de forma logarítmica, tal como se describe en (1). Representa la intensidad de la señal y está expresada en decibelios (dB).

$$\log E_i = \log \left(\sum_{k=1}^N |s_i(k)|^2 \right) \quad (1)$$

donde N es el número de muestras en la i -ésima ventana y s es el valor de la señal en el k -ésimo punto.

El Jitter permite medir las perturbaciones temporales que se presentan en las señales de voz. La ecuación (2) describe la forma de calcular Jitter.

$$Jitter (\%) = \frac{100}{N \cdot M_f} \sum_{k=1}^N |F_0(k) - M_f| \quad (2)$$

donde N corresponde al número de ventanas, M_f es el valor máximo del *pitch* de la señal y $F_0(k)$ es el valor del *pitch* en la k -ésima ventana.

Por otro lado, el Shimmer permite estimar las perturbaciones en la amplitud y se calcula con (3).

$$Shimmer (\%) = \frac{100}{N \cdot M_a} \sum_{k=1}^N |A(k) - M_a| \quad (3)$$

donde N corresponde al número de ventanas, M_a es el valor máximo de la amplitud de la señal y $A(k)$ es el valor de la amplitud en la k -ésima ventana.

Los cocientes de perturbación, ya sea de amplitud o de F_0 (APQ o PPQ), son medidas empleadas para estimar la variabilidad de la amplitud pico a pico o F_0 de la señal, y se calcula con (4),

$$PQ = \frac{1}{L} \sum_{i=1}^L \frac{\left| \frac{1}{K} \sum_{j=1}^K D(i+j-1) - D(i+m) \right|}{\left| \frac{1}{M} \sum_{n=1}^M M \cdot D(i) \right|} \quad (4)$$

donde $L = M - (K - 1)$, $D(i)$ corresponde a la secuencia del período de F_0 o amplitud según corresponda (PPS, cuando se calcula PPQ, y PAS, cuando se calcula APQ), M es la longitud de la PPS o PAS, K es la longitud de la media móvil (normalmente 11 para APQ y 5 para PPQ), y $m = (K - 1)/2$.

Para estos descriptores se empleó el método de ventana deslizante con un tamaño de ventana de 40 ms y una superposición del 50 %, a excepción del APQ y PPQ, en donde se utilizó un tamaño de ventana de 150 ms.

Articulación: esta dimensión hace referencia al movimiento muscular de los articuladores del tracto vocal, en el que los cambios de posición de estos producen diferentes sonidos y resonancias [29]. En el estudio de la articulación se emplearon tanto acercamientos espectrales como cepstrales y se realizó sobre vocales sostenidas y en habla continua:

a) Articulación en vocales sostenidas: en el dominio espectral se determinaron los formantes $F1$ y $F2$ que permiten rastrear la apertura mandibular y posición lingual, respectivamente [30]. En términos de estas características, las vocales a , i y u representan los movimientos articulatorios extremos, por lo que reciben el nombre de “vocales de las esquinas del triángulo vocal”, estudiadas en [31], [32], [33]. En cuanto al dominio cepstral en el análisis de las vocales sostenidas, se calcularon los coeficientes cepstrales de frecuencia de Mel (MFCC, por sus siglas en inglés), que se pueden entender como la representación del habla que se basa en la percepción auditiva humana [34].

El proceso comúnmente empleado para determinar los MFCC se ilustra en la Figura 2, en donde se aplica un método de ventana deslizante a la señal, posteriormente se calcula el espectro de potencia con la transformada rápida de Fourier (FFT) sobre cada uno de las ventanas, luego se le aplica un banco de filtros a los espectro de potencia y se suma la energía de cada uno, se lleva esta energía a forma logarítmica, después se realiza una transformada discreta coseno (DCT) sobre las energías, para finalmente obtener los coeficientes de frecuencia deseados [35], [36].

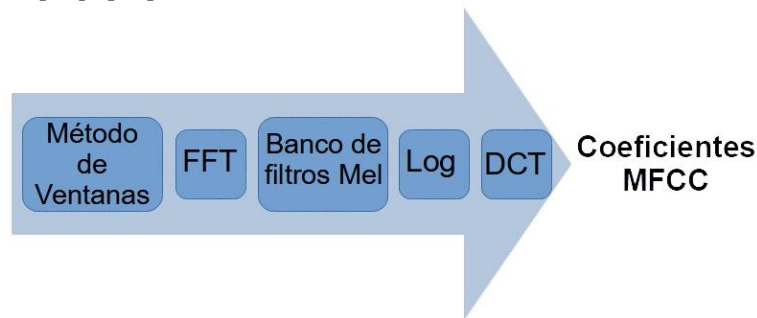


Figura 2. Esquema de detección de los MFCC. Fuente: elaboración propia.

Para el caso del presente estudio solamente se analizaron los primeros 12 MFCC, que son los que se encuentran dentro del rango de frecuencias del habla humana [37]. Para hallar los valores de frecuencia f en Hertz a escala Mel M y viceversa se emplea (5):

$$M(f) = 1125 \ln \left(1 + \frac{f}{700} \right) \quad (5)$$

b) Articulación en habla continua: contrario al análisis en vocales sostenidas, para el habla continua se calcularon las primeras 22 bandas de energía de Bark (BBE, por sus siglas

en inglés) descritas en [38], una escala basada en las transiciones sonoro-insonoro e insonoro-sonoro *–offset y onset*, respectivamente– [39]. Se plantea el análisis de estas transiciones debido a la producción de sonidos anormales o a la dificultad para iniciar/detener la vibración de las cuerdas vocales que presentan algunos pacientes con otras enfermedades neurológicas de base, como en el caso del párkinson [19], [40].

Las BBE se diferencian de los MFCC en la escala y en las frecuencias utilizadas: mientras que los MFCC emplean la frecuencia de Mel y su comportamiento es netamente logarítmico, las BBE emplean la frecuencia de Bark y su comportamiento es logarítmico en sus primeras bandas y posteriormente lineal, de acuerdo con la ecuación (6):

$$Bark(f) = 13 \tan^{-1}(7,6 \times 10^{-4}f) + 3,5 \tan^{-1} \left(\frac{f}{7500} \right)^2 \quad (6)$$

Finalmente, se determinó el contorno del operador de energía de Teager-Kaiser (TKEO, por sus siglas en inglés), tanto para las vocales sostenidas, como para el habla continua. El TKEO es muy sensible a los cambios en la energía de la señal y resulta muy útil al momento de determinar los instantes de activación en una señal [41], [42]. La ecuación (7) se emplea para calcular este descriptor.

$$\Psi[s(n)] = s^2(n) - s(n+1)s(n-1) \quad (7)$$

donde $s(n)$ es el valor de la amplitud de la señal en el n -ésimo punto de la señal.

Prosodia: hace referencia a la entonación y sincronización en la producción del habla natural [43]. Los descriptores que se decidieron extraer para esta dimensión son la media, desviación estándar y el máximo de F_0 y la intensidad de la señal. También se compararon segmentos sonoros e insonoros, en donde se evaluó la tasa sonora y la tasa de silencios, (8) y (9) respectivamente, así como la media y la desviación estándar de la duración de los segmentos sonoros e insonoros.

$$\text{Tasa Sonora} = \frac{\text{Duración total de los segmentos sonoros}}{\text{Duración total de la grabación}} \quad (8)$$

$$\text{Tasa Silencios} = \frac{\text{Duración total de los segmentos insonoros}}{\text{Duración total de la grabación}} \quad (9)$$

Diadococinesia - DDK: se emplea para estudiar movimientos repetitivos de los órganos articuladores involucrados en el habla [40]. Los descriptores que se calcularon son similares a los de la dimensión de la prosodia, excepto que en lugar de determinar las tasas sonoras se determinaron las tasas DDK.

2.5 Pruebas estadísticas

Se aplicó la prueba estadística no paramétrica Mann-Whitney U con un nivel de significancia del 5 % ($\alpha = 0,05$) sobre los grupos de comparación establecidos en tres escenarios propuestos (Ver Figura 3).

Escenario 1: comparación entre pacientes con una determinada condición clínica vs. los demás pacientes con cualquier otra condición, ej., pacientes con voz húmeda vs. los demás pacientes (voz disfónica y condición de voz no determinada). Lo anterior se ilustra en la Figura 3a).

Escenario 2: comparación entre pacientes con alguna de las condiciones clínicas conocidas (VH o VDf) vs. los pacientes con condición no determinada. El proceso se ilustra en la Figura 3b).

Escenario 3: comparación entre pacientes de alguna de las condiciones clínicas conocidas vs. los sujetos control. La Figura 3c) ilustra el esquema de este escenario.

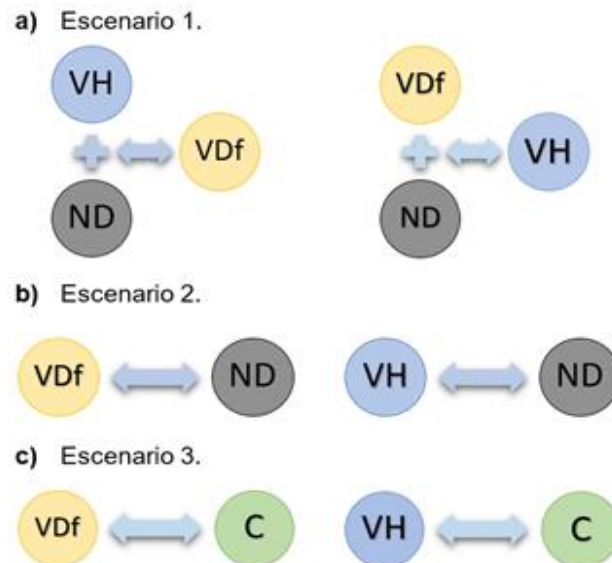


Figura 3. Esquema de los escenarios propuestos. Fuente: elaboración propia.

3. RESULTADOS Y DISCUSIÓN

Con el fin de hacer más sencilla la lectura e interpretación de los resultados, en la Tabla 2 se observa un listado de sufijos empleados en el reporte de características y su respectivo significado.

Tabla 2. Sufijos empleados en las características. Fuente: elaboración propia.

Sufijo	Significado
_m	Media
_sd	Desviación Estándar
_sk	Asimetría
_kt	Curtosis
_on	Onset
_off	Offset
_máx	Máximo

3.1 Fonación en vocales sostenidas

Para el reporte de las características en la dimensión de la fonación se seleccionaron aquellas con diferencias estadísticamente significativas (DES) en dos o más vocales en cualquiera de los escenarios y con p – valor $< \alpha$.

En la Tabla 3 se muestran las características con el criterio de reporte establecido, en donde cada uno de los grupos evaluados presentaron características estadísticamente diferentes entre sí, es decir, el grupo Vdf reporta DES en la media de ΔF_0 y el Shimmer y la asimetría de la ΔF_0 , mientras que el grupo VH las exhibe en la media del Jitter y en la asimetría y curtosis de la energía.

Tabla 3. Características de la fonación con DES en los escenarios evaluados. Fuente: elaboración propia.

Grupo	Característica	Esc. 1	Esc. 2	Esc. 3
Vdf	ΔF_{0_m}	\a\, \u\	\a\, \u\	\a\, \u\
	ΔF_{0_sk}	\a\, \u\	\a\, \u\	\a\
	Shimmer_m	\e\, \o\	\e\	\a\, \e\, \i\, \o\, \u\
VH	Jitter_m	\a\, \o\, \u\	\a\, \u\	-
	Energía_sk	\a\, \e\, \i\	\a\, \e\, \i\	-
	Energía_kt	\a\, \e\, \o\	\a\, \e\	\i\

La Figura 4 ilustra la distribución de la media de ΔF_0 de la vocal \a\ del grupo Vdf a lo largo de los tres escenarios, en el que se puede apreciar que la mediana de este grupo se encuentra por debajo de la de los datos de los otros grupos de comparación, y que la distribución del grupo Vdf en el escenario 2 se encuentra por debajo de la mediana del grupo ND, además de presentar un p-valor inferior al de los otros dos escenarios, lo que sugiere un posible biomarcador. Se plantea que la variación de F_0 en el grupo Vdf es menor debido a que la disfonía se relaciona con una disminución en el control de la vibración de las cuerdas vocales [12].

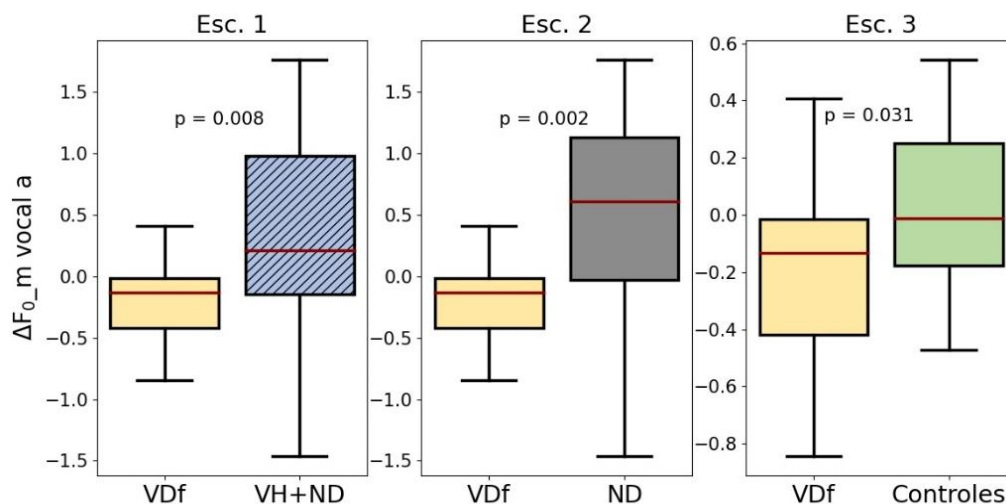


Figura 4. Distribución de ΔF_{0_m} para la vocal \a\ del grupo de pacientes con Vdf en los escenarios evaluados. Fuente: elaboración propia.

En contraste con lo hallado en [15], se determinó que hay características acústicas o fonatorias con potencial capacidad de diferenciar entre pacientes y controles, hallazgo similar a lo reportado en [44], además de una posible diferenciación entre pacientes con distintos trastornos de voz. En un análisis preliminar también se emplearon las características que se usaron en este estudio de fonación y en el que presentaron que las combinaciones entre estas características muestran gran potencial en su uso como factores de diferenciación en pacientes con disfagia [45].

3.2 Articulación en vocales sostenidas

En vista del volumen de información a raíz del espacio de características (176 por vocal), se aplicó como criterio de reporte mostrar aquellas características con DES en al menos dos de las vocales de las esquinas del triángulo vocal.

Siguiendo este criterio de inclusión, no se encontraron características del grupo VDf en los escenarios 1 y 2, lo que sugiere que esta dimensión no suministra información para diferenciar entre pacientes con VDf y pacientes con otras condiciones. Sin embargo, se encontraron múltiples características en el escenario 3, como la media y la desviación estándar de los formantes y sus respectivas derivadas, la desviación estándar del TKEO, la media del segundo al séptimo MFCC, entre otros. Esto indica que esta dimensión, más que diferenciar entre condiciones específicas de voz en pacientes, tiene capacidad de discriminación entre pacientes con disfagia y personas sanas.

Para el caso del grupo VH, se encontró que solamente la desviación estándar de $\Delta^2\text{MFCC1}$ cumplió con el criterio de reporte y está presente en todas las vocales de las esquinas del triángulo vocal en los escenarios 1 y 2, mientras que sólo está presente en la vocal \a\ en el escenario 3. Esta característica tiene una DES en todas las vocales de los escenarios 1 y 2 con excepción de la vocal \o\. No se aceptan tablas incrustadas como imagen.

En la Figura 5 se ilustra la distribución de esta característica en la vocal \a\ en los tres escenarios de este grupo, en donde se observa que el comportamiento de las distribuciones es muy similar puesto que las de los datos del grupo VH se encuentran por debajo de la mediana de la distribución de los grupos de comparación en todos los escenarios.

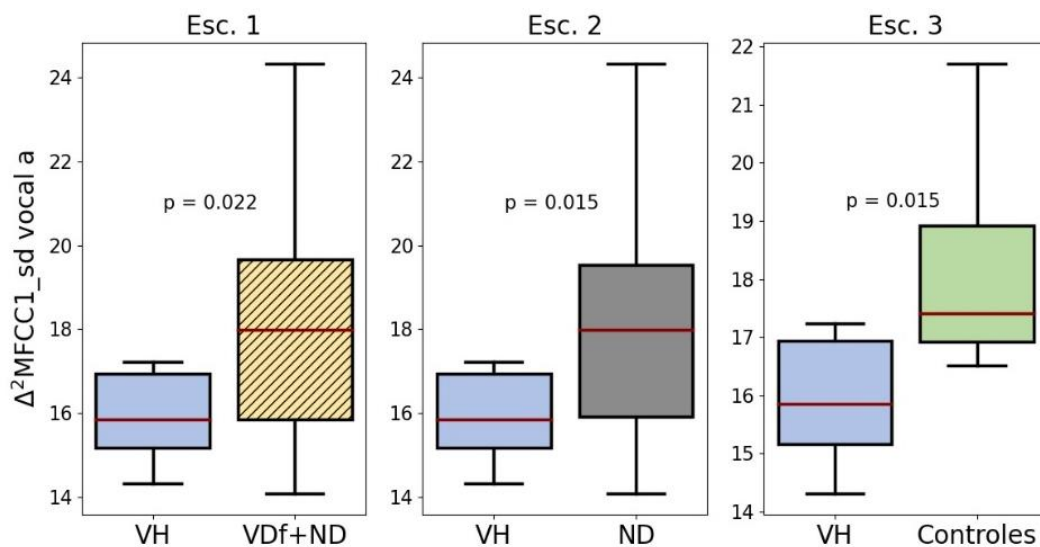


Figura 5. Distribución de $\Delta^2\text{MFCC1_sd}$ de la vocal \a\ en el grupo VH en escenarios evaluados

Fuente: elaboración propia.

Si bien cada una de las vocales también revelaron múltiples características en el grupo VH, muchas de ellas están presente en una o dos vocales a lo largo de los tres escenarios propuestos, por lo que no cumplieron los criterios de reporte establecidos.

3.3 Articulación en habla continua

La Tabla 4 muestra las características relacionadas al habla continua con una DES. Como criterio de reporte se estableció que la característica presentara DES significativa en al menos dos escenarios, en cualquier tarea de voz, sea monólogo o lectura.

Tabla 4. Características de la articulación del habla continua con DES en los escenarios evaluados
Fuente: elaboración propia.

Grupo	Característica	Esc. 1	Esc. 2	Esc. 3
VDf	BBE1_on_m	M	M, L	M, L
	BBE1_off_m	M, L	M, L	M, L
	BBE14_off_m	L	-	M, L
	BBE2_on_sd	L	-	M, L
	BBE5_on_sd	M, L	M, L	-
VH	BBE1_off_m	-	L	M, L
	BBE5_off_m	M, L	M, L	-
	BBE12_off_sd	M	M	M
	BBE13_off_sd	M	M	M

Nota: M: Monólogo; L: Lectura.

Se puede apreciar que los grupos VDf y VH solo tienen como característica común la media de la energía de la primera banda de Bark en *offset* (BBE1_off_m). Mientras que en VDf se presenta en las dos tareas en todos los escenarios, el grupo VH solo reportó diferencia en la lectura del escenario 2 y en ambas tareas del escenario 3. Adicionalmente, el monólogo muestra ser la tarea más robusta y consistente, lo cual puede dar lugar a evaluaciones no intrusivas.

La Figura 6 y la Figura 7 ilustran el comportamiento estadístico, esta característica en común en la lectura de texto realizada por los grupos VDf y VH, respectivamente. Se observa que para ambos grupos la mediana de la distribución se encuentra por encima de sus respectivos grupos de comparación, especialmente en el escenario 3, donde la distribución de los grupos VDf y VH se encuentran sobre la mediana de sus grupos de control. Para el caso del grupo VH en el escenario 1, si bien la mediana se encuentra por encima de la del grupo de comparación, la diferencia presentada no es estadísticamente significativa. Estos resultados son similares a las comparaciones en pacientes con EP en [46], indicando una posible extensión de dichos hallazgos en pacientes con disfagia caracterizados por VH y VDf.

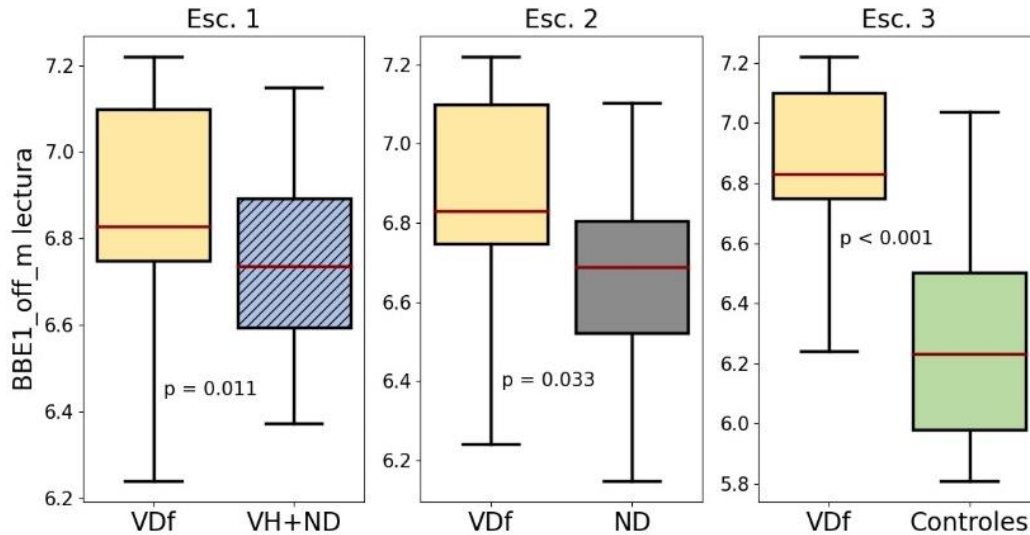


Figura 6. Distribución de BBE1_off_m de la lectura del grupo Vdf en los escenarios evaluados
Fuente: elaboración propia.

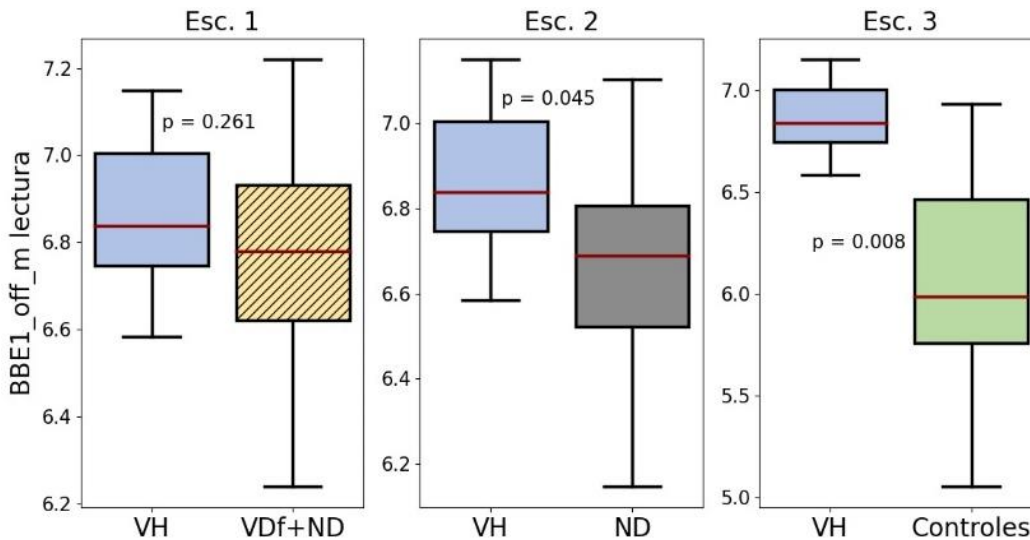


Figura 7. Distribución de BBE1_off_m de la lectura del grupo VH en los escenarios evaluados
Fuente: elaboración propia.

3.4 Prosodia

Igual que en experimentos anteriores, solo se incluyeron características de prosodia con DES en por lo menos dos de los escenarios evaluados en alguna de las tareas de voz.

La Tabla 5 muestra las características que cumplieron con el criterio de reporte, en el que se puede apreciar que las tres características de la energía presentan DES en alguna de las tareas o escenarios evaluados. La característica con menor incidencia de las tres es la desviación estándar de la energía, pues solo se presenta en el escenario 3 del grupo VH, en ambas tareas. También se puede apreciar que se presenta DES en la energía media del grupo Vdf, en ambas tareas en todos los escenarios.

Tabla 5. Características de la prosodia con DES en los escenarios evaluados. Fuente: elaboración propia.

Característica	Esc. 1		Esc. 2		Esc. 3	
	VDf	VH	VDf	VH	VDf	VH
Energía_m	M, L	-	M, L	-	M, L	M, L
Energía_sd	-	-	-	-	-	M, L
Energía_máx	-	L	M	L	M, L	M, L

Nota: M: Monólogo; L: Lectura.

La distribución de la energía media de la lectura del grupo VDF se ilustra en la Figura 8, donde se ve que la mediana de la distribución de este grupo se encuentra por encima de la mediana de los otros grupos de comparación en todos los escenarios, en especial en el escenario 3, donde la mediana del grupo VDF se encuentra por encima de la distribución del grupo de controles y tiene una fuerte DES. Esto puede estar relacionado con el hecho de que los pacientes con VDF deben de hacer un esfuerzo mayor que los demás individuos para realizar la lectura debido a los cambios de entonaciones y fonemas que esta tarea presenta.

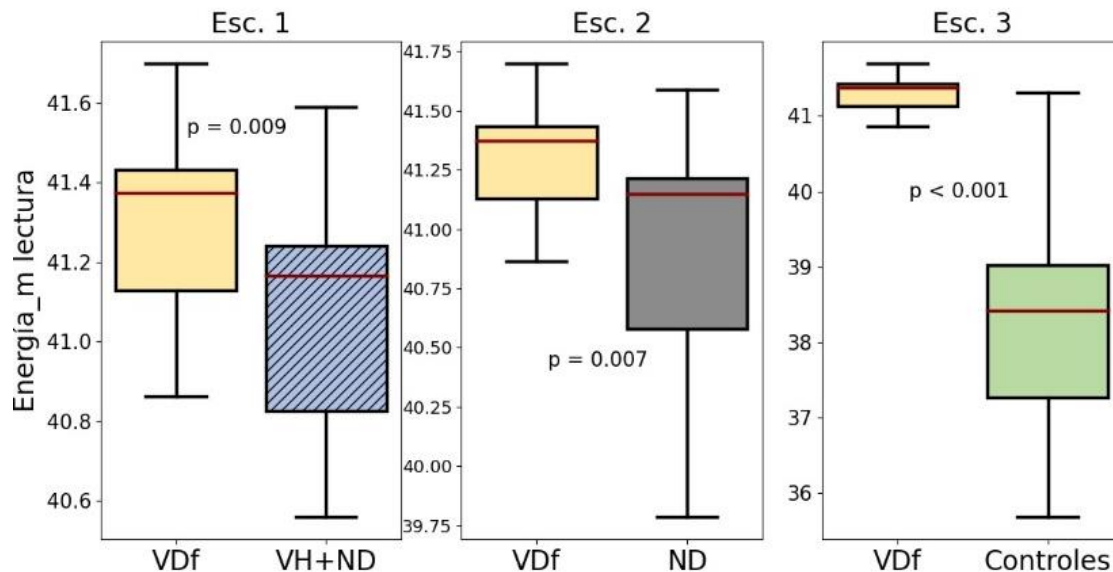


Figura 8. Distribución Energía_m de la lectura del grupo VDF en los escenarios evaluados
Fuente: elaboración propia.

3.5 DDK

En el caso del análisis DDK, el criterio de reporte se estableció en las características con DES en al menos tres de las tareas DDK en cualquiera de los grupos o escenarios evaluados.

En la Tabla 6 se muestran las características que cumplieron con el criterio de selección establecido. Muy similar al análisis de la prosodia, se observa que las tres características de la energía están nuevamente presentes en este análisis, siendo la desviación estándar de la energía la que menos veces presenta DES para alguna de las tareas. También se observó una DES en la media y desviación estándar DDK de las tareas \pe-ta-ka\ y \pa\, pero solamente en los escenarios 1 y 3. Adicionalmente, la energía máxima es la característica que más relevancia presenta en este análisis, teniendo una DES en por lo menos una tarea de los dos grupos en todos los escenarios, exceptuando el escenario 3 del grupo VH.

Tabla 6. Características DDK con DES en los escenarios evaluados. Fuente: elaboración propia.

Característica	Esc. 1		Esc. 2		Esc. 3	
	Vdf	VH	Vdf	VH	Vdf	VH
ΔF_0_m	-	-	-	-	1 – 3	-
$\Delta F_0_máx$	-	-	-	-	-	1 – 3
Energía_m	-	-	-	2, 3	1 – 6	1 – 6
Energía_sd	-	-	-	-	1 – 4	-
Energía_máx	2 – 4	1 – 3, 5	3, 4	1, 5	1, 3 – 6	-
DDK_m	4	4	-	-	1 – 3	3, 4
DDK_sd	4	4	-	-	1 – 3	4
S_sd	-	-	-	-	4 – 6	-

Nota: pa-ta-ka: 1; pa-ka-ta: 2; pe-ta-ka: 3; pa: 4; ka: 5; ta: 6.

La Figura 9 ilustra la distribución de la energía máxima de la tarea \pa\ del grupo Vdf en los escenarios evaluados, en la que se observa que la mediana de la distribución de los datos para este grupo es inferior a la mediana de las de los grupos de comparación, en especial en el escenario 3 donde la distribución del grupo Vdf se encuentra cercanamente por debajo de la mediana de la distribución del grupo de control, además de presentar una fuerte DES.

Se observa que los pacientes tienen una energía menor a la de los grupos de comparación, en especial a la del grupo de control. Esto se relaciona con el hecho de que en los pacientes con Vdf se observó que la intensidad de la señal se reduce rápidamente en las tareas DDK.

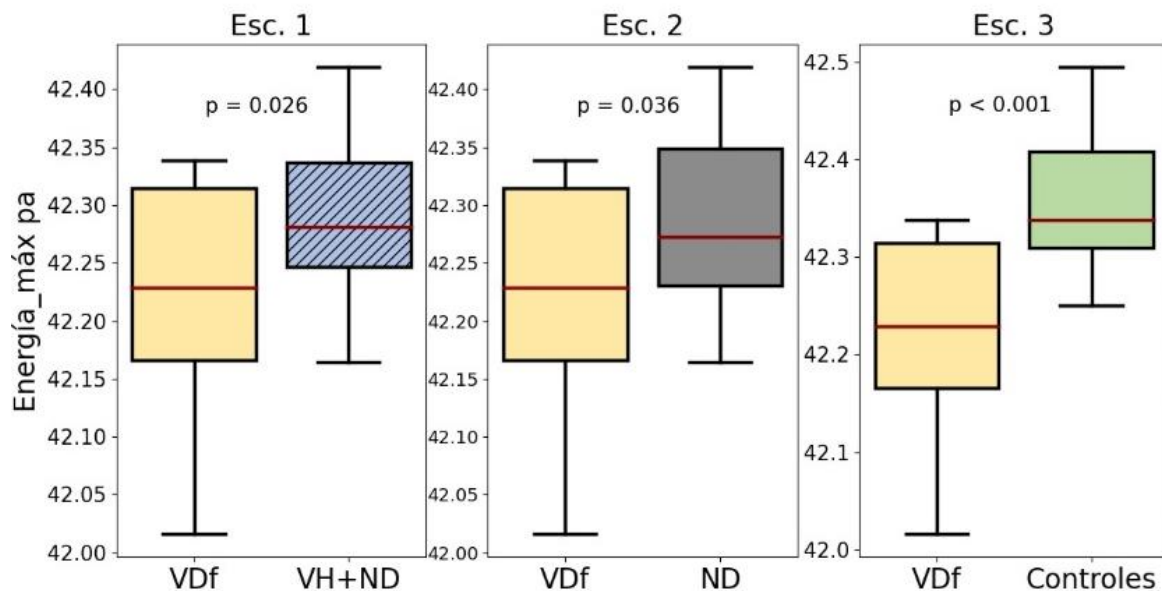


Figura 9. Distribución de Energía_máx de la tarea DDK \pa\ del grupo Vdf en los escenarios propuestos. Fuente: elaboración propia.

4. CONCLUSIONES

Se encontraron múltiples características asociadas a las dimensiones del habla con potencial uso, como biomarcadores para la discriminación de diferentes alteraciones de la voz y del habla en pacientes con disfagia. Adicionalmente, se encontró que las características asociadas a la fonación, la prosodia y al análisis DDK son las que mejor permiten diferenciar entre los grupos estudiados, lo que sugiere que un análisis multidimensional logra detectar las variaciones de las cualidades acústicas del habla bajo condiciones de disfagia. Lo anterior es promisorio para realizar abordajes no invasivos y objetivos que mejoren y complementen los métodos disponibles de evaluación de trastornos deglutorios.

5. TRABAJOS FUTUROS

Se plantea ampliar la base de datos construida para reducir sesgos por la diferencia en el número de muestras de los diferentes grupos de comparación establecidos. También se deberá caracterizar el comportamiento bajo condiciones de disartria, condición que se asocia a varias patologías de base que desencadenan disfagia. Adicionalmente, se plantea realizar una evaluación más robusta en la capacidad discriminatoria de las características con DES, modelos basados en reconocimiento de patrones. Finalmente, se podrán establecer escenarios de comparación pre y posdeglutoria, y el seguimiento de las alteraciones de las cualidades para fines de complementación diagnóstica.

6. AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente en el marco de la convocatoria “Jóvenes Investigadores e Innovadores ITM 2020”, así como por Minciencias - República de Colombia, proyecto No. 825-2017. Los autores agradecen, además, a OFA IPS por su aporte material e intelectual en el desarrollo de este trabajo.

CONFLICTOS DE INTERÉS DE LOS AUTORES

Los autores declaran no tener conflictos de interés.

CONTRIBUCIÓN DE LOS AUTORES

Andrés Felipe Flórez-Gómez: contribuyó principalmente en la escritura del artículo, la conceptualización, metodología, además del desarrollo de los algoritmos para la extracción de características y de las pruebas estadísticas, y en la obtención de los resultados.

Juan Rafael Orozco-Arroyave: contribuyó en la conceptualización, metodología, revisión y análisis de los resultados.

Sebastián Roldán-Vasco: contribuyó al desarrollo de los algoritmos para la extracción de características, conceptualización, metodología, además de la revisión, análisis y validación de los resultados.

7. REFERENCIAS

- [1] L. Sura; A. Madhavan; G. Carnaby; M. Crary, “Dysphagia in the elderly: management and nutritional considerations”, *Clin. Interv. Aging*, vol. 2012, no. 7, pp. 287-298, Jul. 2012. <https://doi.org/10.2147/CIA.S23404>
- [2] D. C. Wolf, “Dysphagia”, en *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3a ed., Eds. Boston: Butterworths, 1990. <https://www.ncbi.nlm.nih.gov/books/NBK408/>
- [3] A. Farri; A. Accornero; C. Burdese, “Social importance of dysphagia: its impact on diagnosis and therapy”, *Acta Otorhinolaryngol Ital*, vol. 27, no. 2, pp. 83-6, Abr. 2007. <http://www.ncbi.nlm.nih.gov/pubmed/17608136>
- [4] O. Ortega; A. Martín; P. Clavé, “Diagnosis and Management of Oropharyngeal Dysphagia Among Older Persons, State of the Art”, *J. Am. Med. Dir. Assoc.*, vol. 18, no. 7, pp. 576-582, Jul. 2017. <https://doi.org/10.1016/j.jamda.2017.02.015>
- [5] Ministerio de Salud y Protección Social Oficina de Promoción Social, “Sala situacional de la Población Adulta Mayor”, *Minist. Salud y Protección Soc.*, pp. 1-8, 2018. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/PS/sala-situacion-envejecimiento-2018.pdf>
- [6] S. E. Langmore, “Evaluation of oropharyngeal dysphagia: which diagnostic tool is superior?”, *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 11, no. 6, pp. 485-489, Dic. 2003. <http://journals.lww.com/00020840-200312000-00014>
- [7] T. Warnecke *et al.*, “The Safety of Fiberoptic Endoscopic Evaluation of Swallowing in Acute Stroke Patients”, *Stroke*, vol. 40, no. 2, pp. 482-486, Feb. 2009. <https://doi.org/10.1161/STROKEAHA.108.520775>
- [8] S. Restrepo-Agudelo; S. Roldan-Vasco; L. Ramirez-Arbelaez; S. Cadavid-Arboleda; E. Perez-Giraldo; A. Orozco-Duque, “Improving surface EMG burst detection in infrahyoid muscles during swallowing using digital filters and discrete wavelet analysis”, *J. Electromyogr. Kinesiol.*, vol. 35, pp. 1-8, Aug. 2017. <https://doi.org/10.1016/j.jelekin.2017.05.001>
- [9] C. M. Steele *et al.*, “Development of a Non-invasive Device for Swallow Screening in Patients at Risk of Oropharyngeal Dysphagia: Results from a Prospective Exploratory Study”, *Dysphagia*, vol. 34, no. 5, pp. 698-707, Oct. 2019. <https://doi.org/10.1007/s00455-018-09974-5>
- [10] D. H. McFarland; P. Tremblay, “Clinical implications of cross-system interactions”, *Semin. Speech Lang.*, vol. 27, no. 4, pp. 300-310, 2006. <https://doi.org/10.1055/s-2006-955119>
- [11] D. Farneti, “Voice and Dysphagia”, en *Dysphagia: Diagnosis and Treatment*, O. Ekberg, Ed. Cham: Springer International Publishing, 2017, pp. 257-274. https://doi.org/10.1007/174_2017_110
- [12] A. E. Aronson, *Clinical voice disorders*. Thieme Inc., 1990.
- [13] T. Warms; J. Richards, ““Wet Voice” as a Predictor of Penetration and Aspiration in Oropharyngeal Dysphagia”, *Dysphagia*, vol. 15, no. 2, pp. 84-88, Mar. 2000. <https://doi.org/10.1007/s004550010005>
- [14] S. Murugappan; S. Boyce; S. Khosla; L. Kelchner; E. Gutmark, “Acoustic characteristics of phonation in ‘wet voice’ conditions”, *J. Acoust. Soc. Am.*, vol. 127, no. 4, pp. 2578-2589, Abr. 2010. <https://doi.org/10.1121/1.3308478>
- [15] M. E. Dajer; P. R. Scalassara; J. L. Marrara; J. C. Pereira, “Voice analysis of patients with neurological disorders using acoustical and nonlinear tools”, *IEEE Int. Work. Mach. Learn. Signal Process. MLSP*, 2012. <http://dx.doi.org/10.1109/mlsp.2012.6349803>
- [16] K. López-De-Ipiña *et al.*, “Advances in a multimodal approach for dysphagia analysis based on automatic voice analysis”, en *Smart Innovation, Systems and Technologies*, 2016, vol. 54, pp. 201-211. https://doi.org/10.1007/978-3-319-33747-0_20
- [17] J. S. Ryu; S. R. Park; K. H. Choi, “Prediction of laryngeal aspiration using voice analysis”, *Am. J. Phys. Med. Rehabil.*, vol. 83, no. 10, pp. 753-757, Oct. 2004. <http://dx.doi.org/10.1097/01.PHM.0000140798.97706.A5>
- [18] K. W. Dos Santos; B. Scheeren; A. C. Maciel; M. Cassol, “Vocal variability post swallowing in individuals with and without oropharyngeal dysphagia”, *Int. Arch. Otorhinolaryngol.*, vol. 19, no. 1, pp. 61-66, 2015. <https://doi.org/10.1055/s-0034-1394129>
- [19] J. R. Orozco-Aroyave *et al.*, “NeuroSpeech: An open-source software for Parkinson’s speech analysis”, *Digit. Signal Process. A Rev. J.*, vol. 77, pp. 207-221, Jun. 2018. <https://doi.org/10.1016/j.dsp.2017.07.004>
- [20] J. R. Orozco-Aroyave; J. D. Arias-Londoño; J. F. Vargas-Bonilla; M. C. González-Rátiva; E. Nöth, “New Spanish speech corpus database for the analysis of people suffering from Parkinson’s disease”, *Proc. 9th Int. Conf. Lang. Resour. Eval. Lr. 2014*, pp. 342-347, 2014. http://www.lrec-conf.org/proceedings/lrec2014/pdf/7_Paper.pdf
- [21] Y. Jadoul; B. Thompson; B. de Boer, “Introducing Parselmouth: A Python interface to Praat”, *J. Phon.*, vol. 71, pp. 1-15, Nov. 2018. <https://doi.org/10.1016/j.wocn.2018.07.001>

- [22] P. Boersma; D. Weenink, “Praat: doing phonetics by computer [Computer program]”. 2001, [En línea]. Disponible en: <http://www.praat.org/>
- [23] J. C. Catford; J. H. Esling, “Phonetics, Articulatory”, en *Encyclopedia of Language & Linguistics*, Elsevier, 2006, pp. 425–442. <https://doi.org/10.1016/B0-08-044854-2/00002-X>
- [24] F. R. Bach; M. I. Jordan, “Discriminative Training of Hidden Markov Models for Multiple Pitch Tracking [speech processing examples]”, en *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, 2005, vol. 5, pp. 489–492. <http://doi.org/10.1109/ICASSP.2005.1416347>
- [25] P. Boersma, “Acurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound”, *IFA Proc.* 17, pp. 97–110, 1993. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.218.4956&rep=rep1&type=pdf>
- [26] S. Strömbergsson, “Today’s most frequently used F0 estimation methods, and their accuracy in estimating male and female pitch in clean speech”, *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 525–529, Sep. 2016. <http://dx.doi.org/10.21437/Interspeech.2016-240>
- [27] S. Basu; J. Chakraborty; M. Aftabuddin, “Emotion Recognition from Speech using Convolutional Neural Network with Recurrent Neural Network Architecture”, en *2nd International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore 2017, pp. 333–336. <https://doi.org/10.1109/CESYS.2017.8321292>
- [28] A. Shukla; R. Tiwari; R. Kala, “Speech Signal Analysis”, en *Studies in Computational Intelligence*, vol. 307, Springer, Berlin, Heidelberg, 2010, pp. 111–128. https://doi.org/10.1007/978-3-642-14344-1_5
- [29] S. Skodda; W. Visser; U. Schlegel, “Vowel articulation in parkinson’s disease”, *J. Voice*, vol. 25, no. 4, pp. 467–472, Jul. 2011. <https://doi.org/10.1016/j.jvoice.2010.01.009>
- [30] G. Fant, *Acoustic theory of speech production*. The Hague: Mouton, 1960.
- [31] K. N. Stevens; A. S. House, “Development of a Quantitative Description of Vowel Articulation”, *J. Acoust. Soc. Am.*, vol. 27, no. 3, pp. 484–493, May. 1955. <https://doi.org/10.1121/1.1907943>
- [32] M. Blomgren; M. Robb; Y. Chen, “A note on vowel centralization in stuttering and nonstuttering individuals”, *J. Speech, Lang. Hear. Res.*, vol. 41, no. 5, pp. 1042–1051, Oct. 1998. <https://doi.org/10.1044/jslhr.4105.1042>
- [33] M. Guzmán, “Acústica Del Tracto Vocal”, 2010. <https://www.logopediapsicologia.com/wp-content/uploads/acustica-del-tracto-vocal.pdf>
- [34] S. Davis; P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”, *IEEE Trans. Acoust.*, vol. 28, no. 4, pp. 357–366, Ago. 1980. <https://doi.org/10.1109/TASSP.1980.1163420>
- [35] L. Moro-Velázquez; J. A. Gómez-García; J. I. Godino-Llorente; J. Villalba; J. R. Orozco-Arroyave; N. Dehak, “Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson’s Disease”, *Appl. Soft Comput.*, vol. 62, pp. 649–666, Jan. 2018. <https://doi.org/10.1016/j.asoc.2017.11.001>
- [36] F. O. López-pabón; T. Arias-vergara; J. R. Orozco-Arroyave, “Cepstral Analysis and Hilbert- Huang Transform for Automatic Detection of Parkinson ’ s Disease”, *Tecnológicas*, vol. 23, no. 47, pp. 93–108, Jan. 2020. <https://doi.org/10.22430/22565337.1401>
- [37] B. B. Monson; E. J. Hunter; A. J. Lotto; B. H. Story, “The perceptual significance of high-frequency energy in the human voice”, *Front. Psychol.*, vol. 5, no. 587, pp. 1–11, Jun. 2014. <https://doi.org/10.3389/fpsyg.2014.00587>
- [38] E. Zwicker, “Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)”, *J. Acoust. Soc. Am.*, vol. 33, no. 2, pp. 248, feb. 1961. <https://doi.org/10.1121/1.1908630>
- [39] E. Zwicker; E. Terhardt, “Analytical expressions for critical-band rate and critical bandwidth as a function of frequency”, *J. Acoust. Soc. Am.*, vol. 68, no. 5, pp. 1523–1525, Aug. 1998. <https://doi.org/10.1121/1.385079>
- [40] J. R. Orozco-Arroyave et al., “Automatic detection of Parkinson’s disease in running speech spoken in three different languages”, *J. Acoust. Soc. Am.*, vol. 139, no. 1, pp. 481–500, Jan. 2016. <https://doi.org/10.1121/1.4939739>
- [41] P. Maragos; T. F. Quatieri; J. F. Kaiser, “Speech nonlinearities, modulations, and energy operators”, in *[Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing*, Toronto, 1991. <https://doi.org/10.1109/ICASSP.1991.150366>
- [42] R. B. Randall; W. A. Smith, “Application of the Teager Kaiser Energy Operator to Machine Diagnostics”, en *Conference: Tenth DST Group International Conference on Health and Usage Monitoring Systems (HUMS)*, Melbourne, 2017, pp. 26–28. <https://www.researchgate.net/publication/316284738>
- [43] M. Tatham; K. Morton, “Speech Production: Prosody”, en *Speech Production and Perception*, London: Palgrave Macmillan UK, 2006, pp. 121–163. https://doi.org/10.1057/9780230513969_5
- [44] S. Roldan-Vasco; A. Orozco-Duque; J. C. Suarez-Escudero; J. R. Orozco-Arroyave, “Machine learning based analysis of speech dimensions in functional oropharyngeal dysphagia”, *Comput. Methods Programs*

- Biomed.*, vol. 208, p. 106248, Sep. 2021. <https://doi.org/10.1016/j.cmpb.2021.106248>
- [45] K. López-de-Ipiña *et al.*, “Automatic voice analysis for dysphagia detection”, *Speech, Lang. Hear.*, vol. 21, no. 2, pp. 86–89, 2018. <https://doi.org/10.1080/2050571X.2017.1369017>
- [46] J. R. Orozco-Arroyave; N. García; J. F. Vargas-Bonilla; E. Nöth, “Automatic Detection of Parkinson’s Disease from Compressed Speech Recordings”, en *Text, Speech, and Dialogue. TSD 2015. Lecture Notes in Computer Science*, Springer, Cham, 2015, pp. 88–95. http://dx.doi.org/10.1007/978-3-319-24033-6_10