ARTÍCULO DE INVESTIGACIÓN

Vol. 39 | No. 2 | **MAYO - AGOSTO 2018** | pp 205-216

REVISTA MEXICANA DE Ingeniería Biomédica

dx.doi.org/10.17488/RMIB.39.2.6

Caracterización y Clasificación de Señales de Auscultación Cervical Adquiridas con Estetoscopio para Detección Automática de Sonidos Deglutorios

Characterization and Classification of Cervical Auscultation Signals Acquired with Stethoscope for Automatic Detection of Swallowing Sound

Y. Sánchez-Cardona[1], *A. Orozco-Duque*[1], *S. Roldán-Vasco*[1,2]

[1]Instituto Tecnológico Metropolitano, Medellín, Colombia

[2]Universidad de Antioquia, Medellín, Colombia

RESUMEN

La evaluación automática de sonidos de auscultación cervical (AC) es una herramienta no invasiva para evaluación de la deglución. Sin embargo, los eventos deglutorios pueden verse enmascarados por fuentes de ruido. Este trabajo propone una metodología de caracterización y clasificación de señales de AC con alta resolución temporal a partir de estetoscopio, para discriminar entre sonidos deglutorios y asociados a ruido. Se adquirieron señales de AC en 10 sujetos sanos durante tres pruebas: toma de líquido, pronunciación del fonema /a/ y aclaramiento de garganta. Se extrajeron características de la señal de AC basadas en coeficientes cepstrales en la escala Mel, transformada wavelet discreta y entropía de Shannon. Las características con mayor relevancia fueron utilizadas como entrada a una máquina de vectores de soporte. Utilizando ventanas de 60 ms - alta resolución temporal - y validación cruzada, se obtuvieron exactitudes del 97.7% para detección de eventos acústicos y 91.7% para sonidos deglutorios. El método propuesto permite clasificación de sonidos deglutorios utilizando estetoscopio -dispositivo común en la práctica clínica- con exactitud comparable a otros trabajos que tienen menor resolución temporal o que utilizan otro tipo de sensores. Este trabajo constituye una primera etapa en el desarrollo de un algoritmo robusto para clasificación de sonidos deglutorios asociados a desórdenes de la deglución, a partir de auscultación cervical, para fines de diagnóstico automático.

PALABRAS CLAVE: deglución, sonidos deglutorios, auscultación cervical, estetoscopio, análisis cepstral, algoritmo de clasificación

ABSTRACT

Automatic evaluation of cervical auscultation sounds (AC) is a non-invasive tool for swallowing assessment. However, the swallowing events could be perturbed by acoustic noise. This paper proposes a methodology of characterization and classification of AC signals acquired by stethoscope with high temporal resolution, in order to discriminate between swallowing sounds and other acoustic noise. AC signals from 10 healthy individuals were acquired with stethoscope during three tasks: liquid ingestion, phoneme /a/ pronunciation and throat clearing. Features based in Mel frequency cepstral coefficients, discrete wavelet transform and Shannon entropy, were extracted. Features with highest Fisher's discriminant ratio were used as input of a support vector machine. By application of 60 ms windows and cross validation, the obtained accuracies were 97.7% for acoustic event detection and 91.7% for swallowing sound detection. The proposed method allows classification swallowing sounds with higher temporal resolution than other works but with comparable accuracy. Furthermore, the use of stethoscope could lead to better acceptation than other sensors by physicians, because it is a common device in clinical practice. This work is a first stage in the development of a robust classification algorithm for sounds in swallowing disorders, oriented to automatic diagnosis.

KEYWORDS: swallowing, swallowing sounds, cervical auscultation, stethoscope, cepstral analysis, classification algorithm

Correspondencia

DESTINATARIO: **Andrés Orozco Duque**

INSTITUCIÓN: **Instituto Tecnológico Metropolitano**

DIRECCIÓN: **Carrera 65 #98 A-75, Medellín, Antioquia, Colombia**

CORREO ELECTRÓNICO: **andresorozco@itm.edu.co**

Fecha de recepción:

3 de noviembre de 2017

Fecha de aceptación:

5 de abril de 2018

INTRODUCCIÓN

Diversas patologías asociadas a problemas de funcionamiento muscular o nervioso pueden causar síntomas asociados a desórdenes de la deglución, donde se dificulta o se imposibilita el transporte del bolo alimenticio desde la boca al estómago; dichos síntomas reciben el nombre de disfagia. Enfermedades como Parkinson, Alzheimer y esclerosis lateral amiotrófica, y otros eventos tales como trauma encéfalo craneano y accidente cerebro vascular están fuertemente ligados a la disfagia [1] [2] [3]. Hay varios riesgos asociados a la disfagia, entre los que se encuentran la bronco aspiración, la neumonía por aspiración, malnutrición y deshidratación [1], complicaciones que se suman a la patología de base, deteriorando el estado de salud y afectando el pronóstico del paciente.

El diagnóstico inicial de la disfagia se realiza principalmente mediante valoración clínica, la cual depende de la experiencia del especialista y, por lo tanto, tiene un alto grado de subjetividad. También se cuenta con dos herramientas diagnósticas: la videofluoroscopia [4] y la endoscopia por fibra óptica [2]. Estas dos técnicas tienen la desventaja de ser invasivas. Como técnica no invasiva, la auscultación cervical utilizando estetoscopio (AC-S) es una de las técnicas instrumentales más utilizadas en fonoaudiología para apoyar la evaluación funcional de la disfagia [5]. La AC-S tiene como finalidad detectar los sonidos de la fase faríngea de la deglución, incluyendo sonidos pre y post deglutorios. Esto con el fin de determinar la posibilidad de compromiso de la vía aérea, la probabilidad de penetración/aspiración y la presencia de disfagia [3]. La técnica permite evidenciar la integridad del mecanismo de protección de la vía aérea, es decir, el cierre glótico que constituye el sonido característico de la deglución. Sin embargo, la exactitud de la AC-S es debatible, debido a que la interpretación de las señales es subjetiva y que hay muy pocos estudios de análisis de la correlación entre la información de los sensores y los eventos fisiológicos [6].

Con el fin de aumentar la objetividad en la evaluación de los sonidos deglutorios, en la literatura se han reportado diversos trabajos que utilizan auscultación cervical (AC) digital y métodos de procesamiento de señales para realizar un análisis automático que no dependa de la interpretación del evaluador clínico [7]. La AC digital es en una técnica genérica que hace referencia al análisis acústico no invasivo de la deglución [8], cuya información se puede adquirir mediante distintos dispositivos tales como acelerómetros [9] [10] [11], micrófonos [12] o estetoscopios [13].

Los acelerómetros y los micrófonos son los dispositivos que más se reportan en investigación. Sin embargo, no existe consenso frente a la confiabilidad y validez de estos dispositivos [14]. Por otro lado, la AC digital utilizando estetoscopio (AC-S) tiene como ventaja que trabaja bajo los mismos criterios utilizados por los evaluadores clínicos, de tal forma que les permite escuchar los sonidos tal y como se perciben con un estetoscopio analógico [13], lo que puede generar una mayor aceptación en el personal asistencial. Por otro lado, el dispositivo es relativamente barato, fácil de movilizar, tiene alta disponibilidad, su posicionamiento es sencillo y no requiere cooperación [4].

Usualmente, los reportes sobre el uso de técnicas automáticas para interpretar objetivamente las señales de AC -y correlacionarlas con los sonidos deglutorios- están orientados a la clasificación biclase entre sonidos normales y anormales asociados a desórdenes de la deglución [15]. De acuerdo con Dudik et al. [6], una de las necesidades actuales para mejorar estos métodos automáticos es utilizar más de dos clases en la clasificación con el fin de poder distinguir eventos no asociados a la deglución que pueden enmascarar los sonidos deglutorios, tales como sonidos de voz o sonidos considerados como otras fuentes de ruido, por ejemplo, el aclaramiento de la garganta. La identificación de estos eventos facilitaría la discriminación entre sonidos deglutorios normales y anormales.

En la literatura se han reportado trabajos orientados a la clasificación entre sonidos deglutorios y otras fuentes de ruido, principalmente dirigidos al monitoreo de la ingesta de alimentos [16]. Estos trabajos tienen la limitación de utilizar ventanas de análisis muy grandes, entre 500 ms y 1.5 s [7] [17]. Estas ventanas, aunque son adecuadas para la detección del evento deglutorio completo, tienen muy baja resolución temporal, dificultando la identificación de las diferentes componentes del sonido deglutorio: ascenso laríngeo, apertura del esfínter esofágico superior y relajamiento glótico post-deglutorio [18]. Aunque el evento deglutorio completo tiene una duración de 732 ± 201 ms [19], el sonido de doble clic producido por el cierre glótico tiene una duración aproximada de 33 ms [15]. Además, existe consenso que las señales fisiológicas de corta duración tienen estacionariedad local, lo que en deglución implica tiempos en el orden de los milisegundos [20].

En este trabajo se propone un esquema para la detección de sonidos deglutorios a partir de la caracterización tiempo-frecuencia de las señales adquiridas con estetoscopio y la utilización de algoritmos de aprendizaje de máquina. El esquema propuesto utiliza ventanas de análisis de 60 ms, de tal forma que sea posible la identificación temporal de los segmentos de clic que identifican las diferentes componentes de la señal de auscultación cervical. La metodología fue evaluada en sujetos sanos y se incluyeron diferentes fuentes de ruido que pueden ser generadas por el paciente durante la adquisición de los datos: sonidos de voz de corta duración (un solo fonema) y sonidos correspondientes a aclaramiento de garganta.

METODOLOGÍA

Protocolo de Toma de Datos

En el presente estudio participaron de forma voluntaria 10 sujetos sanos (6 hombres y 4 mujeres), teniendo como criterios de exclusión que no presentaran ningún desorden en la deglución, ni procesos inflamatorios activos en la boca o garganta. La edad promedio fue de 27.3 ± 5.4 años y todos firmaron consentimiento informado, aprobado por el Comité de Ética del Instituto Tecnológico Metropolitano. A cada sujeto se le solicitó la ejecución de cuatro tareas: deglución de 5 mL de agua, deglución de 10 mL de agua, pronunciación del fonema /a/ durante 1 s (dos repeticiones), y aclaramiento de la garganta (dos repeticiones). Cada sujeto realizó tres repeticiones completas de las cuatro tareas solicitadas. Se hizo variación de volumen pero no de consistencia, ya que la duración de la señal AC-S solo se ve afectada por el primer factor [21]. Durante la ejecución de las acciones, se registraron las señales de AC-S por medio de un estetoscopio digital conectado a un equipo de adquisición de señales (ver Adquisición de señales). El estetoscopio se posicionó en la garganta, de forma lateral al cartílago cricoides [3]. Simultáneamente, se registró la señal de un pulsador presionado por el evaluador al observar el ascenso y descenso de la laringe en el caso de la deglución, o al momento de emitir los sonidos requeridos en cada prueba. El pulsador se utilizó como señal de referencia para validación de los segmentos correspondientes a cada acción en la señal de AC-S.

Adquisición de Señales

Para la adquisición de señales de sonido por AC-S, se usó un estetoscopio electrónico (E-scope® Cardionics). Este dispositivo se conectó al polígrafo PowerLab 16/35 (AD Instruments Inc.). La frecuencia de muestreo fue de 4 kHz. La frecuencia de muestreo se seleccionó teniendo en cuenta que otros autores han reportado que la banda de interés para el análisis de los sonidos deglutorios se encuentra entre 50Hz y 2500Hz [15]. Sin embargo, el diafragma del estetoscopio funciona como filtro pasabajos con frecuencia de corte en 1000 Hz [13]. Con estos criterios, se seleccionó un filtro pasabanda entre 80Hz y 2000 Hz. La frecuencia de corte baja se seleccionó con el fin de filtrar al mismo tiempo el ruido de 60Hz y la frecuencia de corte alta se seleccionó como filtro *antialiasing*. Las señales, tanto de audio como del pulsador de referencia, fueron exportadas a un formato

compatible con MATLAB (The Mathworks, USA). La Figura 1 ilustra una señal AC-S donde se evidencian las componentes del sonido deglutorio.

Conjunto de Entrenamiento y Validación

Para seleccionar los segmentos de señal utilizados en los conjuntos de entrenamiento y validación, se hizo identificación visual de la señal y se miró la correspondencia con los eventos marcados con el pulsador de referencia. Se asignaron etiquetas para discriminación de línea base, eventos deglutorios, sonidos de voz y sonidos de aclaramiento de garganta, estos dos últimos considerados como fuentes de ruido para la detección del cierre glótico. Adicional al pulsador de referencia, todas las señales fueron reproducidas en audio con el fin de confirmar la etiqueta asignada a cada evento. Se seleccionaron 216 segmentos, correspondientes a 54 segmentos por tarea, de tal forma que los grupos estuvieran balanceados. El ancho de la ventana de evaluación de los segmentos se estableció en 60 ms, lo que corresponde a 240 muestras. Con esta ventana se asegura que el evento de doble clic quede contenido en la misma. Una vez seleccionados los segmentos, se construyó una matriz de características de 216 filas (segmentos) y 17 columnas (características). Se utilizaron como características 10 coeficientes cepstrales, la energía de 6 coeficientes de detalle generados por la transformada wavelet discreta y la entropía de Shannon.

Extracción de Características

Coeficientes Cepstrales

Los coeficientes cepstrales en la escala de frecuencia Mel (MFCC - Mel-*frequency cepstral coefficient*) constituyen un método muy utilizado para el procesamiento de audio, especialmente en esquemas de reconocimiento de voz. MFCC utiliza un banco de filtros triangulares escalados logarítmicamente (escala de Mel) [22].

Las frecuencias centrales en la escala Mel de cada filtro están determinadas por:

$$Mel(f) = 1127 \ln\left(1 + \frac{f}{700}\right) \quad (1)$$

donde f es la frecuencia a re-escalar.

Se aplicó un filtro preénfasis FIR de primer orden y, posteriormente, se calculó la transformada discreta de Fourier de cada segmento:

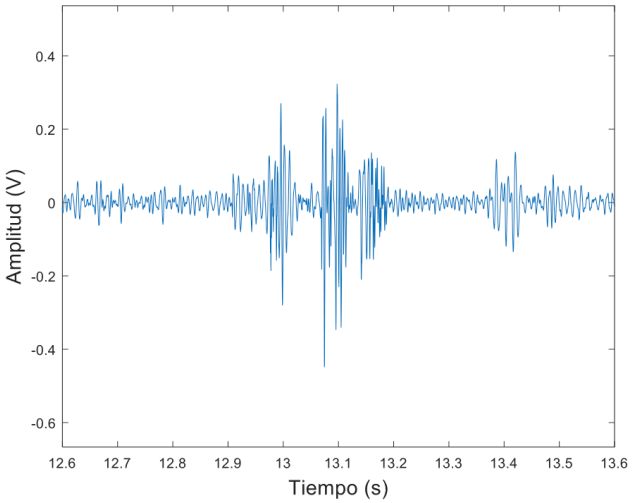

FIGURA 1. Evento deglutorio en una señal de auscultación cervical con estetoscopio.

$$X[k] = \sum_{n=0}^{N-1} w[n]s[n]e^{-j\frac{2\pi kn}{N}} \quad (2)$$

Donde k es el contador de frecuencias, n es el contador de muestras, $s[n]$ es cada segmento, N es la longitud de cada segmento y $w[n]$ es una ventana de Hamming descrita por:

$$w[n] = \begin{cases} 0.54 - 0.46\cos\frac{2\pi n}{(N-1)}, & 0 \le n \le N-1 \\ 0, & en\ otro\ caso \end{cases} \quad (3)$$

Para cada escala, la salida de los filtros se expresa de forma logarítmica y se calcula mediante la multiplicación de la magnitud del espectro de frecuencia de la señal por la respuesta en frecuencia de su filtro triangular correspondiente, tal como lo indica la siguiente ecuación [23]:

$$Xf[m] = \ln\left(\sum_{k=0}^{N-1} |X[k]|\, H_m[k]\right) \quad (4)$$

donde m es un contador de filtro entre 1 y el número de filtros utilizados M, en este caso 10, N es la longitud de cada segmento, k es un contador de frecuencia, y $H_m[k]$ representa la respuesta en frecuencia de la magnitud de los filtros pasa-banda triangulares.

Los coeficientes cepstrales se calculan con la transformada coseno discreta aplicada sobre $Xf[m]$ de acuerdo con la siguiente ecuación:

$$Cc_l = \sum_{m=0}^{M-1} Xf[m] cos\left(\frac{l\pi}{M}\left(m - \frac{1}{2}\right)\right) \quad l = 1, \ldots, M. \quad (5)$$

El cálculo de los coeficientes cepstrales se realizó utilizando ventanas deslizantes con solapamiento del 50%. En la aplicaciones prácticas usualmente se utilizan entre 5 y 15 coeficientes [22]. Para este trabajo se calcularon 10 coeficientes cepstrales, debido a que, después de evaluar 15 coeficiente, se encontró que solo los primeros 10 proporcionaban información relevante de la señal. Finalmente, a partir de la técnica MFCC se construyó el subconjunto de características [Cc_1,..., Cc_{10}].

Energía de los Coeficientes Wavelet

Se implementó la Transforma Wavelet Discreta (DWT-*Discrete Wavelet Transform*) [23]. Se calcularon los coeficientes de detalle y aproximación a partir de la descomposición wavelet con 6 niveles utilizando la wavelet madre *Daubechies* de orden 6.

A partir de la DWT se estimó como característica la energía relativa wavelet de los coeficientes de detalle (EcD), que representa la energía que cada nivel de detalle aporta al total de la energía de la señal [24]. El subconjunto de características construido es el siguiente: {EcD_1,..., EcD_6}.

Entropía de Shannon

La entropía de Shannon (H) mide la incertidumbre de una fuente de información. La entropía de Shannon es máxima cuando todos los valores de la señal tienen la misma probabilidad. Se parte entonces de la hipótesis de que los datos menos probables contienen más información. El cálculo de la entropía de Shannon está definido por [25]:

$$H(x) = -\sum_i p(x_i) \log_2 p(x_i) \quad (6)$$

Donde $p(x_i)$ es la probabilidad de ocurrencia de los valores de una variable x en el rango i. Los rangos i se definen para la construcción del histograma.

Análisis de Relevancia

Para evaluar la relevancia de cada característica respecto a la selección de las tres clases definidas, se registró la distribución en diagramas de cajas. Asimismo, se calculó el radio discriminante de Fisher (FDR) [26]. El FDR permite cuantificar la capacidad de una característica para separar las clases en un problema específico. Está definido por la siguiente ecuación:

$$FDR = \sum_j^C \sum_{k \neq i}^C \frac{(j-\mu_k)^2}{\sigma_j^2 + \sigma_k^2} \quad (7)$$

Donde $\{\mu_j, \sigma^2_j\}$ y $\{\mu_k, \sigma^2_k\}$ son las medias y varianzas de las clases j y k, respectivamente. C es el número de clases.

Clasificación

Se implementó una máquina de vectores de soporte (SVM) con *kernel* lineal [27]. La selección de las características utilizadas para entrenamiento y validación del modelo se realizó a partir de la evaluación del índice FDR. Se utilizó validación cruzada con cinco particiones para el reporte de los resultados de rendimiento del clasificador.

Se implementaron tres esquemas de clasificación: a) clasificación multiclase para la discriminación entre sonidos deglutorios y fuentes de ruido; b) clasificación biclase para la detección de eventos sonoros; y c) clasificación biclase para la detección de sonidos degluto-

rios. Para el esquema multiclase se definieron las siguientes etiquetas: clase 0, correspondiente a segmentos de línea base; clase 1, correspondiente a sonidos deglutorios; y clase 2, que contiene segmentos con sonidos de voz y sonidos de aclaramiento de garganta, ambos considerados como otras fuentes de ruido. El entrenamiento de la SVM multiclase se realizó bajo el método uno contra uno. Para el esquema b, las etiquetas se definieron así: clase 0, para la línea base; y clase 1, que contiene cualquier evento sonoro. Para el clasificador c las etiquetas son: clase 0, que contiene tanto línea base como otras fuentes de ruido; y la clase 1, que contiene solo sonidos deglutorios.

RESULTADOS Y DISCUSIÓN

En la Figura 2A se observa la señal de AC-S en azul y la señal de referencia en rojo. Para el proceso de asignación de etiquetas, el desfase entre el pulsador y los eventos representados en la señal se ajustaron a partir de la señal de audio. Todos los registros tienen la siguiente secuencia de eventos: deglución de 5 mL de agua, deglución de 10 mL de agua, dos eventos de voz y dos eventos de aclaramiento de garganta.

La Figura 2B muestra que Cc_2 incrementa su amplitud y presenta picos positivos por encima de la línea base en los intervalos con sonido. Por otro lado, la Figura 2C muestra un ejemplo del comportamiento de Cc_4, el cual toma valores positivos donde se presentó una deglución de 5mL (aproximadamente a los 5 s), mientras que en la deglución de 10mL (aproximadamente a los 13 s) toma valores positivos y negativos, y para voz y aclaramiento de garganta (últimos cuatro eventos después de los 30 s) solo toma valores negativos. El comportamiento anterior fue común en la mayoría de los registros.

La Tabla 1 presenta el resultado de la evaluación de la relevancia de las características utilizando el índice FDR. A partir de estos resultados, las características seleccionadas fueron Cc_2, Cc_3, Cc_4, EcD_5, EcD_6 y $H(x)$. Es de notar que, aunque la característica EcD_1 presentó un índice FDR superior al índice de Cc_4, esta característica fue descartada porque el coeficiente EcD_1 en la transformada wavelet está asociado usualmente a ruido de alta frecuencia.

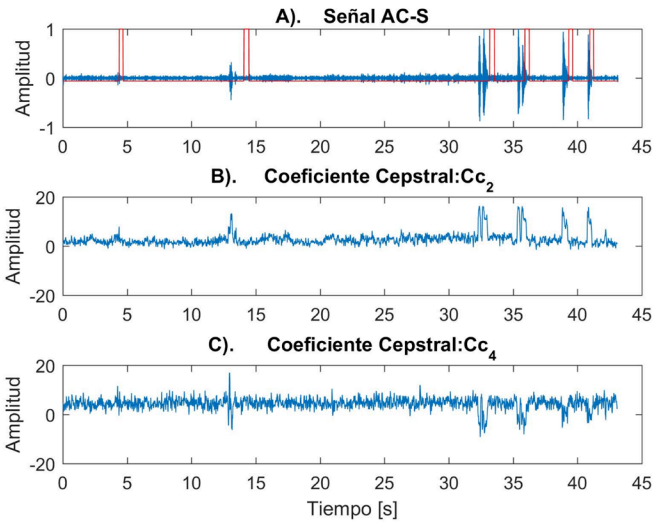

FIGURA 2. Ejemplo de coeficientes cepstrales 2 y 4 de una señal de AC-S.

La Figura 3 ilustra la distribución en diagramas de cajas de las características seleccionadas. Se puede observar que las características que presentan una mayor distancia entre las medias y una mejor separación de clases son Cc_2 y EcD_6, lo que corresponde a las características con un mayor FDR. Las demás características seleccionadas presentan una distribución que contribuye a discriminar alguna de las clases: Cc_3 presenta valores altos para la clase 0, en Cc_4 los valores más altos se presentan en la clase 1, EcD_5 presenta valores bajos para la clase 2, mientras en $H(x)$ la clase 2 presenta los valores más altos y dicha clase se separa claramente de las demás. Cabe destacar que para $H(x)$, aunque el grafico de cajas muestra una buena separación de las clases, las medias de la clase 0 y la clase 1 están muy cercanas, lo que afecta el índice FDR.

La Figura 4 ilustra la distribución del espacio de características utilizando las dos características con mayor relevancia -Cc_2 y EcD_6- de acuerdo con el índice

FDR. En esta representación, se observa que la separación entre línea base (clase 0) y demás sonidos está más definida que la separación entre sonidos deglutorios (clase 1) y ruido (clase 2). Al implementar el clasificador solo con estas dos características se obtiene una tasa de aciertos de 82.9%. Al probar el clasificador con las seis características seleccionadas, presentó una tasa de acierto del 91.7%. Para confirmar que el descarte de la característica EcD_1 fue correcto, se implementó el clasificador incluyendo esta característica y el rendimiento disminuyó a 91.2%.

La Tabla 2 presenta los resultados de los tres esquemas de clasificación implementados. La clasificación entre eventos sonoros y línea base (esquema b) fue 97.7%. La clasificación entre deglución y las otras clases (esquema c) fue 90.3%.

La Tabla 3 muestra la matriz de confusión para el caso multiclase. Se puede observar que, con el esquema propuesto, la discriminación de la línea base presenta una tasa de aciertos muy elevada. El mayor reto está en la clasificación entre sonidos deglutorios y otras fuentes de ruido.

TABLA 1. Evaluación de la relevancia de las características utilizando el índice FDR. En negrilla se presentan las características seleccionadas para a implementación del clasificador.

Coeficientes Cepstrales		Energía de detalle Wavelet	
Cc_1	0.7704	EcD_1	1.8096
Cc_2	**23.2831**	EcD_2	0.1614
Cc_3	**4.8351**	EcD_3	1.1282
Cc_4	**1.5228**	EcD_4	1.1847
Cc_5	0.2693	**EcD_5**	**2.3545**
Cc_6	0.4742	**EcD_6**	**9.8992**
Cc_7	0.9071	**Entropía de Shannon**	
Cc_8	0.1762	$H(x)$	**4.3845**
Cc_9	0.6038		
Cc_{10}	0.3179		

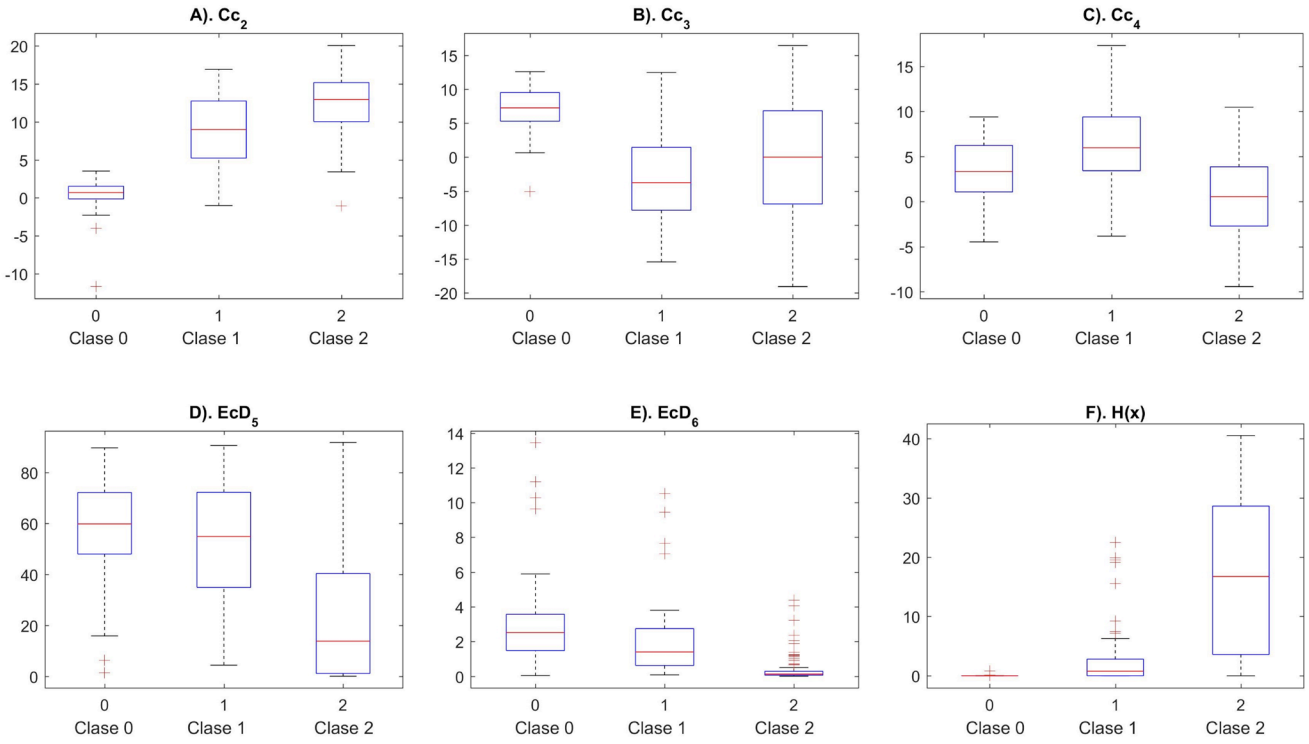

FIGURA 3. Diagramas de cajas de las características con mayor FDR.

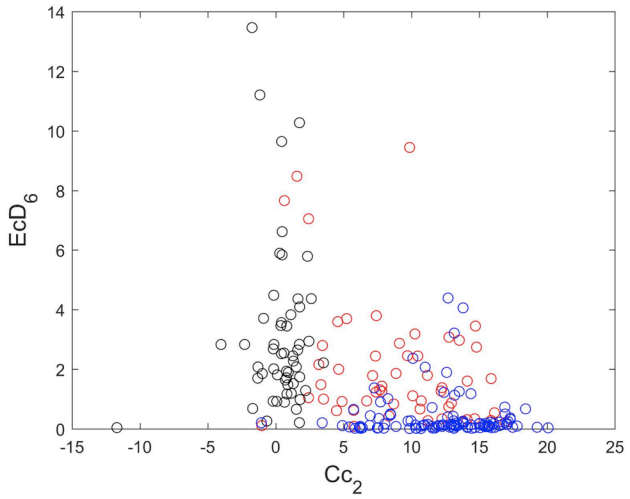

FIGURA 4. Espacio de las dos características con mayor FDR. Los círculos negros representan la línea base, los rojos deglución y el azul otros eventos acústicos (voz y aclarar garganta).

El resultado de este trabajo es comparable con los resultados reportados por Aboofazeli et al. [28], quienes obtuvieron una tasa de acierto de 91% para la detección del sonido de la deglución tanto en sujetos sanos como en pacientes. Sin embargo, Aboofazeli et al. no incluyeron fuentes de ruido en su trabajo ni implementaron estetoscopio. Asimismo, nuestros resultados pueden ser comparados con el trabajo de Sejdic et al. [10], aunque ellos usaron otro tipo de sensor. Dicho trabajo está orientado a identificar degluciones con aspiración y degluciones sanas, a partir de señales de acelerometría medidas en el cartílago cricoides; emplearon análisis de discriminante lineal como clasificador y *wavelet packet* para caracterización, logrando 90% de exactitud [10].

Lazareck et al. [29], hicieron un análisis más robusto de señales adquiridas con acelerómetros a partir de un espacio de características mayor. Evaluaron la capacidad de clasificación para varios tipos de bolos, encontrando una especificidad de 100%, pero una sensibilidad reducida (70%) en alimentos semisólidos [15]. Lazareck et al., al igual que en nuestro trabajo, incluyen ventanas de evaluación cortas de 50 ms con el fin de detectar los eventos de clic característicos del cierre glótico; sin embargo, no incluyen otras fuentes de ruido y su trabajo emplea sensores inerciales, con los cuales no están familiarizados los especialistas en fonoaudiología.

Con respecto a otros trabajos que han incluido otras fuentes de ruido, se destaca el de Sazonov et al. [7], quienes reportan el desarrollo de algoritmos de detección automática de la deglución a partir de sonidos, y en el cual incluyeron ruido de la voz y del ambiente. Ellos reportan 96.8% de acierto en detección de eventos sonoros y 84.7% para la detección de sonidos deglutorios. El trabajo de Sazanov et al. emplea ventanas de tiempo muy grandes comparado con nuestro trabajo (1.5 s vs 60 ms), lo cual dificulta el desarrollo de algoritmos que detecten los diferentes eventos tipo clic.

Yagi et al. [30], aplicaron un sistema para la detección de la deglución utilizando no solo información de los sonidos de auscultación cervical obtenidos con micrófono, sino también información de sensores de flujo respiratorio y sensores de movimiento de la laringe. Ellos obtuvieron una exactitud del 98.2% en la detección de eventos deglutorios, pero dicha exactitud bajó al 88.3% cuando incluyeron otras fuentes de ruido como el habla y los artefactos de movimiento.

TABLA 2. Resultados de los clasificadores.

Clasificador	Tasa de aciertos
SVM multiclase (a) Uno contra uno	91.7%
SVM biclase (b) Detección de eventos acústicos	97.7%
SVM biclase (c) Detección de sonidos deglutorios	90.3%

Ellos utilizaron un esquema de caracterización con MFCC y análisis de componentes principales y un clasificador SVM. Reportan una exactitud en la detección de eventos deglutorios intra-sujeto de 80.4%, pero dicha exactitud cae a 66.7% en el caso inter-sujeto.

TABLA 3. Matriz de confusión del clasificador multiclase.

		Clase estimada		
		C0	C1	C2
	C0	**54**	**0**	**0**
	C1	**3**	**41**	**10**
	C2	**0**	**5**	**103**

Olubanjo and Ghonvanloo [17], reportaron un esquema de detección de la deglución utilizando un micrófono para la auscultación cervical. En el experimento incluyeron eventos como hablar, masticar, toser y aclarar garganta. La ventana de observación fue reducida a 500ms, pero en una prueba con cuatro sujetos la precisión fue solo del 67.6%.

En nuestro trabajo se reporta un esquema de clasificación de sonidos deglutorios, respecto a la línea base y a eventos de ruido, con una mejora en la resolución temporal respecto a los trabajos anteriormente mencionados, ya que se utilizan ventanas de tiempo de corta duración. Igualmente, la tasa de acierto alcanzado con el esquema propuesto en este trabajo es mayor a la reportada en los trabajos previos que incluyen fuentes de ruido y es comparable a la alcanzada a partir de otros dispositivos de adquisición. Sin embargo, se deben tener en cuenta algunas limitaciones. Una de ellas está relacionada con la forma en que se seleccionaron los segmentos.

La matriz de entrenamiento y validación contiene solo 56 segmentos de línea base (con el fin de ajustar el balance de clases), lo que pudo haber excluido segmentos de línea base que podrían contener otras fuentes de ruido diferentes a las evaluadas, por ejemplo, los artefactos de movimiento. En futuros trabajos, se propone utilizar segmentos extraídos del registro completo para entrenar el clasificador. Igualmente, es necesaria una evaluación posterior con mayor número de eventos acústicos, tales como diferentes fonemas, otras fuentes de ruido, el acto de toser, o ruido externo al sujeto proveniente del entorno. Igualmente, se requiere una validación comparando la detección de los eventos deglutorios contra la prueba gold standard, en este caso, la videofluoroscopia, para validar los hallazgos y su posible utilización en la práctica clínica.

Este trabajo constituye una primera etapa para el desarrollo de un algoritmo robusto para clasificación de sonidos deglutorios entre sujetos sanos y pacientes con desórdenes de la deglución, para fines de diagnóstico automático. Con este fin, y teniendo como base la alta resolución temporal, en futuros proyectos se deben analizar los sonidos respiratorios que aparecen inmediatamente después de la deglución, ya que cuando ocurre aspiración o penetración laríngea estos sonidos sufren alteraciones. Se debe estudiar la tasa de clasificación de otros eventos deglutorios propios de la evaluación clínica que realizan los terapeutas, tales como el pre-clic, el "lub-dub" (parecido al latido del corazón durante la deglución) [4], y el de la respiración.

CONCLUSIONES

En el presente trabajo se presenta la evaluación de un esquema de clasificación de señales de AC-S mediante SVM, a partir de la caracterización en dominios de frecuencia y tiempo-frecuencia. Se analizaron características extraídas a partir de los MFCC y los coeficientes DWT, además de la entropía de Shannon, con lo cual se alcanzó una tasa de aciertos del 91.7% para detección de sonidos deglutorios debido al cierre glótico, en presencia de otras fuentes de ruido (en particular pronunciación de un fonema y el sonido de aclaramiento de garganta). El aporte de este trabajo está orientado a la utilización de un esquema que mejora la resolución temporal respecto a otros trabajos basados en sonidos deglutorios adquiridos mediante estetoscopios, el cual es un equipo de uso común en la práctica clínica. La exactitud reportada en este trabajo es comparable con otros trabajos similares que utilizan ventanas de tiempo con menor resolución temporal o que adquieren la señal con otros dispositivos.

Futuros trabajos deben ir orientados a la evaluación de otras posibles fuentes de ruido. Una vez resuelto este problema, se debe aplicar la metodología implementada a un grupo de pacientes para determinar diferencias, entre las características de los sonidos en degluciones sanas y patológicas. Esto ayudará a mejorar el entendimiento de las relaciones existentes entre los eventos acústicos que se detectan con AC-S y la fisiología de la deglución.

AGRADECIMIENTOS

Los autores de este trabajo agradecen al:

Departamento Administrativo de Ciencias, Tecnología e Innovación de la República de Colombia - COLCIENCIAS, por el apoyo brindado a este trabajo a través de la financiación del proyecto:

No. 115071149746

REFERENCIAS

[1] Stegemann S, Gosch M, Breitkreutz J. Swallowing dysfunction and dysphagia is an unrecognized challenge for oral drug therapy. Int J Pharm 2012;430(1-2):197-206.

[2] Alvo A, Olavarría C. Decannulation and assessment of deglutition in the tracheostomized patient in non-neurocritical intensive care. Acta Otorrinolaringol Esp 2014;65(2):114-9.

[3] Fonseca MAlB. Guia de prática basada en la evidencia para la auscultacion cervical en disfagia orofaringea. 2008;

[4] Leslie P, Drinnan MJ, Zammit-Maempel I, Coyle JL, Ford GA, Wilson JA. Cervical auscultation synchronized with images from endoscopy swallow evaluations. Dysphagia 2007;22(4):290-8.

[5] Bolzan GDP, Christmann MK, Berwig LC, Rocha RM. Contribution of the cervical auscultation in clinical assessment of the oropharyngeal dysphagia. Rev CEFAC 2013;15(2):455-65.

[6] Dudik JM, Coyle JL, Sejdi E. Dysphagia Screening : Contributions of Cervical Auscultation Signals and Modern Signal-Processing Techniques. IEEE Trans Human-Machine Syst 2015;45(4):1-13.

[7] Sazonov E, Makeyev O, Schuckers S, Lopez-Meyer P, Melanson E, Neuman M. Automatic Detection of Swallowing Events by Acoustical Means for Applications of Monitoring of Ingestive Behavior. IEEE Trans Biomed Eng 2010;57(3):626-33.

[8] Zenner PM, Losinski DS, Mills RH. Using cervical auscultation in the clinical dysphagia examination in long-term care. Dysphagia 1995;10(1):27-31.

[9] Movahedi F, Kurosu A, Coyle JL, Perera S, Sejdić E. A comparison between swallowing sounds and vibrations in patients with dysphagia. Comput Methods Programs Biomed 2017;144:179-87.

[10] Sejdic E, Steele CM, Chau T. Classification of penetration-aspiration versus healthy swallows using dual-axis swallowing accelerometry signals in dysphagic subjects. IEEE Trans Biomed Eng 2013;60(7):1859-66.

[11] Dudik JM, Coyle JL, El-Jaroudi A, Mao Z-H, Sun M, Sejdić E. Deep learning for classification of normal swallows in adults. Neurocomputing [Internet] 2018;0:1-9. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0925231218300201

[12] Klahn MS, Perlman AL. Temporal and durational patterns associating respiration and swallowing. Dysphagia 1999;14(3):131-8.

[13] Hamlet S, Penney DG, Formolo J. Stethoscope acoustics and cervical auscultation of swallowing. Dysphagia 1994;9(1):63-8.

[14] Leslie P, Drinnan MJ, Finn P, Ford GA, Wilson JA. Reliability and validity of cervical auscultation: A controlled comparison using videofluoroscopy. Dysphagia 2004;19(4):231-40.

[15] Lazareck LJ, Moussavi ZMK. Classification of normal and dysphagic swallows by acoustical means. IEEE Trans Biomed Eng 2004;51(12):2103-12.

[16] Makeyev O, Lopez-Meyer P, Schuckers S, Besio W, Sazonov E. Automatic food intake detection based on swallowing sounds. Biomed Signal Process Control [Internet] 2012;7(6):649-56. Available from: http://dx.doi.org/10.1016/j.bspc.2012.03.005

[17] Olubanjo T, Ghovanloo M. Real-time swallowing detection based on tracheal acoustics. In: Conference, Ieee International Processing, Signal. 2014. page 4417-21.

[18] Hanna F, Molfenter SM, Cliffe RE, Chau T, Steele CM. Anthropometric and demographic correlates of dual-axis swallowing accelerometry signal characteristics: A canonical correlation analysis. Dysphagia 2010;25(2):94-103.

[19] Honda T, Baba T, Fujimoto K, Goto T, Nagao K, Harada M, et al. Characterization of swallowing sound: Preliminary investigation of normal subjects. PLoS One 2016;11(12).

[20] Chau T, Chau D, Casas M, Berall G, Kenny DJ. Investigating the Stationarity of Paediatric Aspiration Signals. IEEE Trans Neural Syst Rehabil Eng 2005;13(1):99-105.

[21] Hammoudi K, Boiron M, Hernandez N, Bobillier C, Moriniere S. Acoustic study of pharyngeal swallowing as a function of the volume and consistency of the bolus. Dysphagia 2014;29(4):468-74.

[22] Sigurdson S, Petersen KB, Larsen J. Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music. Proc 7th Int Conf Music Inf Retr 2006;(m):3-6.

[23] Poornachandra S. Wavelet-based denoising using subband dependent threshold for ECG signals. Digit Signal Process [Internet] 2008 [cited 2012 Jan 25];18(1):49-55. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1051200407001388

[24] González Castañeda EF, Torres-García AA, Reyes-García CA, Villaseñor-Pineda L. Aplicación de la Sonificación de Señales Cerebrales en Clasificación Automática. Rev Mex Ing Biomed 2015;36(3):235-50.

[25] Dudik JM, Jestrovi I, Luan B, Coyle JL, Sejdi E. A comparative analysis of swallowing accelerometry and sounds during saliva swallows. Biomed Eng Online 2015;14(3):1-15.

[26] Sergios Theodorisdis and Konstantinos Koutroumbas. Pattern Recognition. 4th ed. Academic Press; 2009.

[27] Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. Data Min Knowl Discov [Internet] 1998;2(2):121-67. Available from: http://www.springerlink.com/index/Q87856173126771Q.pdf

[28] Aboofazeli M, Moussavi Z. Analysis and classification of swallowing sounds using reconstructed phase space features. ICASSP, IEEE Int Conf Acoust Speech Signal Process - Proc 2005;V:421-4.

[29] Lazareck LJ, Moussavi Z. Swallowing sound characteristics in healthy and dysphagic individuals. Conf Proc IEEE Eng Med Biol Soc 2004;5:3820-3.

[30] Yagi N, Nagami S, Lin M kuan, Yabe T, Itoda M, Imai T, et al. A non-invasive swallowing measurement system using a combination of respiratory flow, swallowing sound, and laryngeal motion. Med Biol Eng Comput 2017;55(6):1001-17.