

[dx.doi.org/10.17488/RMIB.39.1.6](https://doi.org/10.17488/RMIB.39.1.6)

Detección Automática y Clasificación de Eventos en Sonidos Cardiopulmonares de Sujetos Saludables

Automatic Detection and Classification of Cardiopulmonary Events

P. Mayorga-Ortiz¹, J. A. Valdez-Gonzalez¹, C. Druzgalski², V. Zeljkovic³

¹Instituto Tecnológico de Mexicali

²California State University

³The Lincoln University

RESUMEN

En este estudio, se presenta una metodología para evaluar lentes fáquicos intraoculares, cuando el flujo del humor La auscultación de señales basada en un estetoscopio estándar y/o electrónico no solo incluye sonidos internos del cuerpo, también incluye frecuentemente ruido externo de interferencia con componentes en el mismo rango. Esta forma de examinar es incluso afectada por los umbrales auditivos variantes de los profesionales de la salud y el grado de experiencia en reconocimiento de indicadores peculiares. Además, los resultados son a menudo caracterizados en términos cualitativos descriptivos sujetos a interpretaciones individuales. Para direccionar esta preocupación, los estudios presentados en este artículo contienen un procesamiento concurrente de las componentes dominantes de sonidos del corazón (HS) y del pulmón (HS), y una etapa de acondicionamiento que incluye la reducción de HS presente en señales LS. Específicamente, la transformada de Hilbert fue una técnica de caracterización para HS. En el caso de señales enfocadas a LS, las técnicas de detección de actividad de voz y el cálculo de umbrales de algunos componentes de los vectores acústicos de Coeficientes Cepstrales en Frecuencia Mel (MFCC), fueron útiles en la caracterización de eventos acústicos asociados. Las fases de inspiración y expiración fueron diferenciadas por medio de la sexta componente de MFCC. Con el fin de evaluar la eficiencia de esta aproximación, proponemos los Modelos Ocultos de Markov con Modelos Mezclados Gaussianos (HMM-GMM). Los resultados utilizando esta forma de detección son superiores cuando se desarrolla la clasificación con modelos HMM-GMM, la cual refleja las ventajas de la forma de detección cuantificable y clasificación sobre la aproximación clínica tradicional.

PALABRAS CLAVE: Transformada de Hilbert (HT); Detección de Actividad de Voz (VAD); Modelos Mezclados Gaussianos(GMM); Modelos Ocultos de Markov (HMM).

ABSTRACT

A standard and/or electronic stethoscope based auscultatory signals include not only the internal sounds of the body but also interfering external noise often with similar frequency components. This form of examination is also affected by varying thresholds of clinical practitioner's hearing and degree of experience in recognition of peculiar auscultatory indicators. Further, the results are often characterized in qualitative descriptive terms subject to individual's interpretation. To address these concerns, presented studies include concurrent processing of dominant heart (HS) and lung (LS) sounds components and a conditioning stage involving HS presence reduction within LS focused signals. Specifically as determined, the Hilbert transform was a technique of choice in HS characterization. In the case of LS focused signals, the speech activity detection techniques (VAD) and the thresholds calculation of some components of acoustic vectors of Cepstral Coefficients in Mel Frequency (MFCC), were useful in characterization of associated acoustic events. The phases of inspiration and expiration were differentiated by means of the sixth component of MFCC. In order to evaluate the efficiency of this approach, we propose Hidden Markov Models with Mixed Gaussian Models (HMM-GMM). The results utilizing this form of detection are superior when performing classification with HMM-GMM models, which reflect the advantages of presented form of quantifiable detection and classification over traditional clinical approach.

KEYWORDS: Hilbert Transform (HT); Voice Activity Detection (VAD); Mixed Gaussian Models (GMM); Hidden Markov Models (HMM).

Correspondencia

DESTINATARIO: Julio Alejandro Valdez Gonzalez
INSTITUCIÓN: Instituto Tecnológico de Mexicali
DIRECCIÓN: Av. Tecnológico S/N, Col. Elías Calles, C. P.
21376 , Mexicali, Baja California, México
CORREO ELECTRÓNICO: julito_valdez@hotmail.com

Fecha de recepción:

3 de agosto de 2017

Fecha de aceptación:

14 de noviembre de 2017

INTRODUCCIÓN

La salud es una tarea que concierne a varios campos del conocimiento, el avance de la misma fue incipiente sin el progreso de otras áreas como la ingeniería. Además, la tecnología facilitó y mejoró los procesos en la medicina. Tal es el caso de la auscultación cardiaca y respiratoria, que con el avance de la tecnología se ha logrado mejorar los estetoscopios y de esta manera tener una mejor percepción de los sonidos cardiacos y pulmonares. Pero aun así, cuando se colectan datos del pulmón, estos contienen elementos ajenos a la respiración normal y en condiciones normales existen agentes externos (ruido ambiente) e internos (sonidos cardiacos), que pueden dificultar la auscultación, por ello se buscan alternativas para reducir la información que no es de interés. Una alternativa de reducción es el filtrado (pasa banda, pasa baja, pasa altas) el cual permite eliminar frecuencias no necesarias y que afectan la detección de los segmentos de inhalación y exhalación [1]. Una situación similar se presenta cuando el objetivo es el análisis, procesamiento o reconocimiento de sonidos del corazón. Desafortunadamente, esta alternativa no resulta tan eficaz si hay traslape en el dominio de la frecuencia.

A través de la historia se han implementado técnicas para la obtención de sonidos pulmonares, desde la auscultación simple con estetoscopio tradicional o acústico, hasta métodos más especializados como lo son chalecos de auscultación, pero estos últimos suelen ser costosos y menos accesibles para consultorios comunes o de bajos recursos. Por lo anterior, se ha optado por métodos que estén al alcance de cualquier consultorio médico de uso general, es decir, que no cuenten con instalaciones altamente equipadas con aislamiento contra ruidos externos, para ello se han propuesto métodos como grabaciones de sonidos pulmonares utilizando micrófonos con condensador, es decir, convirtiendo los estetoscopios tradicionales en estetoscopios digitales [1]. Para lograr avances en el contexto de la segmentación y clasificación de las

señales del pulmón y del corazón, ha sido fundamental el uso del análisis computacional.

En el contexto de los sonidos del pulmón hay propuestas como el método supervisado del algoritmo de 'clusterización fuzzy C-means' (FCM), este fue desarrollado para identificar episodios de respiración, como sonidos con ronquido y sonidos sin ronquido, así como sonidos respiratorios, sonidos de *tragado*, y otros tipos de ruidos [1]. Pero estos requieren instalaciones dedicadas y supervisar al paciente durante el sueño.

En la actualidad existen distintos niveles de atención médica [2], algunos básicos y otros más especializados, los niveles de atención médica básicos cuentan con cierta funcionalidad, implicando al menos equipo como un estetoscopio; además por normatividad, se debe contar con una computadora para el expediente electrónico del paciente [3, 4], lo cual es asequible. Desafortunadamente, el estetoscopio presenta varios retos, como el ruido ambiental y el traslape de los sonidos del corazón (HS) con los sonidos del pulmón (LS).

Hoy en día, existen diferentes propuestas para identificación de los ciclos pulmonares los cuales dan una apreciación de sus eventos principales, ayudando a determinar si un paciente se encuentra en condiciones saludables o en si en estos eventos existe una anomalía que podría ser interpretada como enfermedad respiratoria. La forma clásica de identificación de sonidos pulmonares es la auscultación con estetoscopio simple, pero este método cuenta con la desventaja de que el oído humano tiene un rango limitado y hay frecuencias fuera de la percepción humana, lo que le dificulta al médico diagnosticar con certeza la existencia de alguna enfermedad [5]. Además, con el paso del tiempo el oído humano va perdiendo eficiencia en la percepción de los sonidos, por lo cual es necesario un sistema que no dependa del oído humano que pueda segmentar y clasificar sonidos cardiacos y sonidos pulmonares por métodos automatizados y computarizados.

Algunas aproximaciones son dirigidas a enfermedades endémicas, en donde se utilizan las características acústicas de la tos y crepitaciones para reforzar vectores de Coeficientes Cepstrales en Frecuencia Mel (por sus siglas en inglés, MFCC) y se aplican las ondículas (Wavelets) ^[6]. Otros trabajos se destinan al análisis y monitoreo de las ondas sonoras del corazón ^[7]. Estos evidencian algunas características que se modifican durante el stress cardiaco, y que el cambio es más significativo para personas con problemas cardiacos ^[7]. Otros autores mencionan que si bien el uso de estetoscopio es una herramienta de bajo costo, los movimientos de los pacientes sobre todo en niños contamina los registros de los sonidos ^[8]. Por lo cual, se propone un esquema automatizado multi-banda para supresión de ruido, y mejorar la calidad de las señales de auscultación contra fuerte contaminación de fondo ^[8].

Un procesamiento previo y básico en el análisis de los sonidos pulmonares es algún método de cancelación de ruido cardíaco, localizando primeramente los componentes de sonido del corazón. El Análisis de Espectro Singular (por sus siglas en inglés, SSA) es una técnica de análisis de series de tiempo. A pesar de la superposición frecuencial de los componentes del corazón con los sonidos del pulmón, se pueden diferenciar dos tendencias en los espectros con valores propios, estos conducen a un sub-espacio que contiene más información sobre el sonido subyacente del corazón. En un experimento se mezclaron artificialmente señales respiratorias reales para evaluar la eficiencia del método ^[9]. Seleccionando la longitud adecuada para una ventana SSA que resulta en una descomposición de buena calidad y bajo costo computacional del algoritmo. En lo general los algoritmos de detección de sonido del corazón existentes tienen menor desempeño en las señales anormales ^[9].

En ^[10], se propone un algoritmo basado en un método de doble umbral para una detección robusta de los sonidos cardiacos S1 y S2. La señal original del sonido

del corazón (HSS) es filtrada aplicando una ventana de Hamming. La envolvente de sonido cardiaco se extrae mediante la Transformada de Hilbert-Huang (HHT) y es segmentado el sonido cardiaco por el método de doble umbral ^[10].

En ^[11], se presenta un método de baja complejidad para la detección del primer y segundo sonidos del corazón (S1 y S2) y los períodos de sístole y de diástole sin necesidad de utilizar una referencia electrocardiográfica. El algoritmo utiliza el modo de descomposición empírica (EMD, de sus siglas en inglés) que produce envolventes de intensidad de los principales sonidos del corazón en el dominio del tiempo ^[11].

En otro estudio ^[12], se sugiere un método de localización para S1 y S2, basado en un algoritmo que implica el filtrado en frecuencia, detección de energía, y la duración de intervalo. La exactitud de la localización se evaluó comparando el algoritmo con el método de localización basado en transformar de Hilbert tradicional ^[12].

En ^[13], se propone un método automático para segmentación y análisis de detección de pico en patrones de sonidos cardíaco (HS), con especial atención a las características de las envolventes de HS y teniendo en cuenta las propiedades de la Transformada de Hilbert (HT). La segmentación y localización del pico se lleva a cabo mediante la envolvente (ET) de la señal HS que se obtiene con un énfasis en S1 y S2. Luego, basándose en las características de ET y las propiedades de la HT de las funciones convexa y cóncava, se aplica la transformada de Hilbert modificada en tiempo corto (STMHT) para segmentar y localizar automáticamente los puntos pico para HS mediante los cruce por cero de la STMHT ^[13].

Algunos autores proponen métodos basados en análisis multi-escala de frecuencias instantáneas (IF) y envolventes instantáneas (IE) que son calculadas después de aplicar un conjunto de modos de descomposi-

ción empírica (EEMD) [5]. Aquí, las señales respiratorias son descompuestas en una serie de componentes, llamadas funciones de modo intrínseco (IMFs), para los cuales la IF puede ser definida en cada punto. Los clasificadores están basados en el hecho de que la dispersión de las IFs de las señales respiratorias disminuye marcadamente cuando los ruidos continuos adventicios (CAS) aparecen en los ciclos de respiración. Dado que los segmentos de CAS tienen menor dispersión de frecuencias instantáneas que los segmentos que contienen respiración normal, los segmentos que son considerados como CAS pueden ser detectados con una ventana deslizante, la cual se desliza sobre la totalidad de las secuencias IFs [5]. Sin embargo, sigue siendo no evidente obtener la componente correcta.

Existen técnicas para obtener los componentes principales del ciclo respiratorio como lo son inhalación, exhalación, de una forma automática. Algunos trabajos proponen el uso de realidad virtual (VR) para monitorear y distinguir entre fases de inspiración y expiración de acuerdo a las características de las señales. A través del uso de coeficientes cepstrales en frecuencia Mel (MFCC) se obtienen las características más importantes de los eventos, así como la detección de actividad de voz (VAD), la cual es una herramienta que facilitaría detectar segmentos de actividad y de silencio [14].

Otra alternativa en detección de eventos de ciclos respiratorios es por medio del uso de análisis de componentes principales (PCA), los cuales permiten obtener las características principales y necesarias para la detección de segmentos de actividad que se encuentran dentro del ciclo de respiración. Algunos autores proponen la extracción de características principales por medio de la transformada rápida de Fourier (FFT), para llevar a cabo la clasificación [15].

Como hemos mencionado en los párrafos anteriores, el análisis de los sonidos de pulmón (LS) se complica con la interferencia de sonidos cardíacos (HS) y vice-

versa, estos tienden a enmascarar características importantes de LS o HS según la señal a analizar. Por otra parte, un HS patológico interfiere de manera más significativa en LS que un HS normal. Por esto aquí se proponen algunos pasos de procesamiento previo a la reducción de HS, remarcando la importancia de la localización de los componentes de HS. Además, se plantea abordar dos problemas al mismo tiempo, el de HS y el de LS. Por un lado, en HS se utiliza la transformada de Hilbert para la detección de puntos extremos (máximos y mínimos), por otro lado, en LS la aplicación de técnicas de detección de actividad de voz (VAD) y el cálculo de umbrales de algunas componentes de vectores acústicos MFCC, son útiles en la detección y localización de eventos. Incluso, los ciclos de inspiración y expiración podrían también ser diferenciados por medio de la sexta componente de MFCC, la cual contiene información importante para llevar a cabo su detección. Con el propósito de evaluar la eficiencia de la detección automática, se implementaron modelos HMM-GMM.

La sección 2 explica los principios que subyacen en las técnicas y métodos aplicados en nuestra propuesta. La sección 3 muestra los resultados de los experimentos realizados, así como la eficiencia obtenida con la metodología propuesta. La sección 4 expone las conclusiones obtenidas en base a los experimentos de la metodología propuesta.

METODOLOGÍA

La detección automática de los eventos S1 y S2 del corazón se apoya fuertemente en la transformada de Hilbert; la detección de la respiración normal inhalación y exhalación, aquí se apoya en herramientas tradicionales del procesamiento digital de voz, como lo es la detección de actividad de voz (VAD). Además, es importante determinar si la detección fue eficaz, por lo cual se proponen modelos HMM-GMM para evaluar la eficiencia de la detección manual y automática. Las características acústicas de estos modelos fueron obte-

nidas aplicando vectores MFCC y cuartiles. La detección automática se llevó a cabo con una serie de pasos que pueden resumirse en las Figura 1 y Figura 2, cuyas etapas serán explicadas posteriormente.

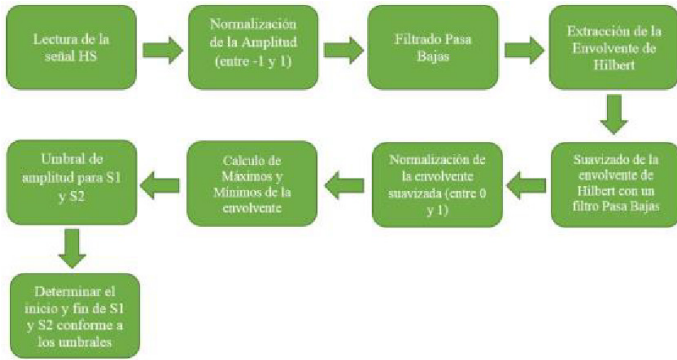


FIGURA 1. Diagrama de bloques propuesto para la detección de eventos cardiacos.

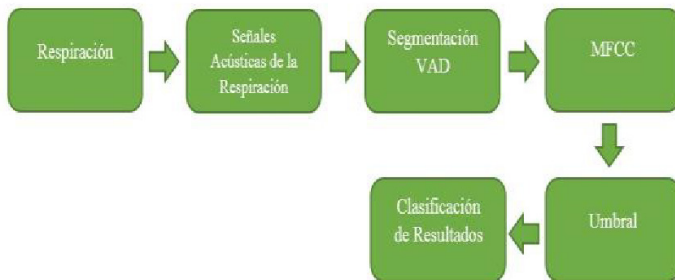


FIGURA 2. Diagrama de bloques propuesto para la detección de eventos respiratorios.

Vectores MFCC

En MFCC, los sonidos son parametrizados, haciendo un preénfasis con filtros FIR, seguido por una ventana Hamming aplicada a cada trama de análisis [16-18]. En este trabajo, se experimentó con ventanas Hamming de 30 ms y 15 ms de corrimiento en las señales HS-LS, a las cuales se aplica la Transformada Rápida de Fourier (FFT); posteriormente, se obtiene el módulo y se multiplica por un banco de filtros donde sus rangos de frecuencia y frecuencias centrales están distribuidos en la escala de Mel. A esto le sigue una etapa de logaritmo de la energía obtenida de cada filtro y consecutivamente la transformada inversa de Fourier. Dado que la energía será real y par, la transformada inversa

de Fourier es un producto interno donde subsisten únicamente las partes par, por lo tanto, resulta equivalente a calcular la Transformada Discreta Cosenoidal (DCT). El resultado final es un vector de características llamado MFCC [14, 19, 20].

Vectores Cuantílicos

El Cuantíl q_p de una variable aleatoria está definido como el número q más pequeño, tal que la función de distribución acumulativa es mayor o igual a una probabilidad p , donde p se encuentra entre $0 < p < 1$. Esto se puede definir con una función de densidad de probabilidad continua $f(x)$ a través de la Ecuación (1):

$$p = \int_{-\infty}^{q_p} f(x) dx \quad (1)$$

La estacionariedad de LS está relacionada con la duración de la fase de inspiración (≈ 1.5 s.) y la fase de espiración (≈ 2.5 s.) para la mayoría de las señales LS utilizadas [21]. En HS la duración promedio de S1 es de 0.1 s a 0.12 s, y S2 de 0.8 a 0.14 s [22]. Esto posibilita vectores de 30 ms con corrimientos de 10 ms. Es decir, el tamaño de los vectores tiene que estar dentro del rango de estacionariedad.

Para el cálculo de los cuartiles, el primer paso es la lectura de la señal, partiendo de archivos de audio *.wav; posteriormente, se aplica la FFT. Cumpliendo con un principio básico para una función de densidad de probabilidad, la distribución espectral se normaliza como se muestra en la Ecuación (2).

$$F_N(f) = \frac{\int_{-\infty}^{\infty} f(t)e^{-j2\pi ft} dt}{\text{area}(F(f))} \quad (2)$$

Un ejemplo particular de Cuantiles son los Cuartiles, calculados aquí mediante la Ecuación (3), cuyos valores frecuenciales $f_{0.25}, \dots, f_{0.75}$ corresponden al valor frecuencial donde se acumula 0.25, ..., 0.75 del área espectral por trama, que es lo que denotamos como coeficientes Cuantílicos. El cálculo del último Cuantil

no es importante ya que siempre es igual a 1. Lo que resulta en un vector de 3 dimensiones o 3 componentes por vector resultante de una trama [23].

$$\mathbf{A}_{0.25} = \int_{-\infty}^{f_{0.25}} \mathbf{F}_N(f) df, \dots, \mathbf{A}_{0.75} = \int_{-\infty}^{f_{0.75}} \mathbf{F}_N(f) df \quad (3)$$

Transformada de Hilbert

La transformada de Hilbert de una señal produce un adelanto de su fase de $\pi/2$ radianes. Cuando una señal es causal en un dominio, ya sea tiempo o frecuencia, la parte real y la imaginaria en el otro dominio estarán vinculadas por la transformada de Hilbert [24]. Se define la transformada de Hilbert como la convolución con la función $-1/\pi t$:

$$\text{HT}\{f(t)\} = f(t) * \frac{-1}{\pi t} = \frac{-1}{\pi} \int_{-\infty}^{\infty} \frac{f(\tau)}{t-\tau} d\tau \quad (4)$$

Convolución en tiempo por $-1/\pi t$, es equivalente a multiplicar en frecuencia por $i \cdot \text{sign}(w)$, es decir no se modifica el espectro en amplitud, solo se efectúa un corrimiento $\pi/2$ para frecuencias positivas y de $-\pi/2$ para frecuencias negativas. Podemos escribir a la función exponencial compleja de la siguiente forma:

$$e^{i\omega t} = \cos \omega t + i \sin \omega t = \cos \omega t - i \text{HT}\{\cos \omega t\} \quad (5)$$

Generalizando esta idea y creando una función compleja a partir de una función real, cuya parte imaginaria tenga un retardo en fase de 90° respecto de su parte real, es decir:

$$g(t) = f(t) - i \text{HT}\{f(t)\} \quad (6)$$

La función $g(t)$ es conocida como función analítica asociada a $f(t)$. Dada una función en tiempo cuya parte imaginaria sea igual a menos la transformada de Hilbert de su parte real, su transformada de Fourier será causal. Análogamente, si una función temporal es causal, la parte real e imaginaria de su transformada de Fourier estarán vinculadas por la

transformada de Hilbert. Se define a la envolvente $E(t)$ de una función $f(t)$ como el módulo de su función analítica:

$$E(t) = |g(t)| = \sqrt{f(t)^2 + (\text{HT}\{f(t)\})^2} \quad (7)$$

Detección de Actividad en la Señal

El objetivo con algoritmos VAD es detectar tramas de actividad como en nuestro caso la inspiración y espiración. La técnica VAD es usada aquí con propósitos de segmentación, distinguiendo entre zonas de silencio y zonas activas de respiración, pero sin especificar cuáles son de inspiración y cuáles de espiración, para esto último se aplicara la sexta componente de los vectores MFCC. Una suposición en el algoritmo de VAD es que el espectro de la señal activa cambia más rápido que el ruido de fondo el cual tiene cambios que se producen lentamente. Además, que la magnitud de los segmentos activos es generalmente más alta que la magnitud del ruido de fondo [25, 26]. El algoritmo VAD fue adaptado para obtener datos que incluyen las duraciones y puntos de inicio y fin. El algoritmo VAD se implementa primero para filtrar componentes de baja frecuencia no deseados, y luego se calcula la potencia con diferentes tamaños de ventana de la transformada rápida de Fourier (FFT) de la señal de entrada. Suponiendo que $X(K)$ es la FFT de la señal de entrada $x(n)$ [14]. El cálculo de energía E_n , de la señal es como sigue:

$$E_n = \sum_{K=K_1}^{K_2} |X(K)|^2 \quad (8)$$

La segmentación con VAD se basó en el algoritmo de [25]. Si suponemos que la respiración esta degradada por ruido aditivo no correlacionado, se pueden plantear dos hipótesis:

- › H_0 : ausencia de respiración: $Y=N$.
- › H_1 : respiración presente: $Y=S+N$.

Donde, N denota el ruido y S denota respiración sin ruido en el dominio espectral respectivamente.

Las funciones de densidad de probabilidad par H_0 y H_1 son las siguientes:

$$H_0: P_N(Y) = f(Y|H_0) = \frac{\exp\left(-\frac{|Y|^2}{2\sigma_N^2}\right)}{2\pi\sigma_N^2} \quad (9)$$

$$H_1: P(Y) = f(Y|H_1) = \frac{\exp\left(-\frac{|Y|^2}{2(\sigma_N^2 + \sigma_s^2)}\right)}{2\pi(\sigma_N^2 + \sigma_s^2)} \quad (10)$$

Donde σ_N^2 σ_s^2 son las varianzas del ruido y la señal de respiración limpia, las cuales se calculan a priori con las señales. Asignando $P_N(Y)/P(Y) < \varepsilon$, partiendo de que $P(Y) > P_N(Y)$, calculando heurísticamente ε y suponiendo que el cociente de las varianzas $k = \sigma_s^2/\sigma_N^2$, algebraicamente se puede llegar a:

$$|Y|^2 > -\frac{2(\sigma_s^2 + \sigma_N^2)}{k+2} \ln\left(\varepsilon \sqrt{\frac{1}{1+k}}\right) \quad (11)$$

Si se define

$$|Y_\varepsilon|^2 \triangleq -\frac{2(\sigma_s^2 + \sigma_N^2)}{k+2} \ln\left(\varepsilon \sqrt{\frac{1}{1+k}}\right) \quad (12)$$

Luego Y_ε puede aplicarse como umbral. La bandera de nivel en VAD puede ser definida como:

$$\text{flag} = \begin{cases} 1, & |Y|^2 > |Y_\varepsilon|^2 \\ 0, & |Y|^2 \leq |Y_\varepsilon|^2 \end{cases} \quad (13)$$

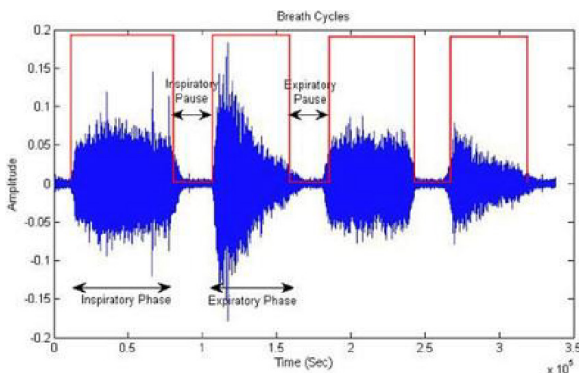


FIGURA 3. Segmentación de la señal acústica de la respiración usando VAD [26].

Análisis de Componentes Principales (PCA)

PCA es una técnica estadística útil aplicada en muchos campos para identificar patrones en datos con una gran dimensión y expresar los datos de tal manera que resalte sus similitudes y diferencias [27]. Otra ventaja es que PCA reduce el número de dimensiones con menor pérdida de información. Los tres pasos importantes para el cálculo de PCA son:

Paso 1: Cálculo de la media de la nube de puntos U . En este paso, se calcula la media de cada dimensión (por ejemplo, los vectores acústicos):

$$\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i \quad (14)$$

Paso 2: Cálculo de matriz de covarianza denotada como $CovU$ mediante el producto externo de la matriz U consigo misma:

$$CovU = \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})(u_i - \bar{u})^T \quad (15)$$

Paso 3: En este paso, se calculan los eigenvectores y eigenvalores de la matriz de covarianza $CovU$. Primero se calcula A , la matriz diagonal de los eigenvalores de $CovU$. Enseguida, se busca la matriz E de eigenvectores que diagonalizan la matriz $CovU$.

$$E^{-1} * CovU * E = A \quad (16)$$

Los eigenvectores E de la matriz de covarianza $CovU$ son los componentes principales de la nube de puntos U .

Análisis de Componentes Independientes (ICA)

ICA es un algoritmo para separar fuentes mezcladas linealmente; para este algoritmo es importante algunos pasos de pre-procesamiento antes de aplicarse como filtrado, centrado, y blanqueado [28, 29]. Centrar implica restar la media de todos los datos a cada uno de los datos. Blanquear es un paso que remueve cualquier

correlación de los datos, es decir cuando se aplican extractores de características, de-correlaciona sus componentes. Lo anterior significa que la matriz de covarianza tendrá unos en su diagonal principal y ceros fuera de ella. ICA produce una rotación en los datos obtenidos minimizando la gaussianidad de los mismos. Cuando medimos $X=AS$, X es la matriz con los datos obtenidos por sensores o mediciones, A es la matriz mezcladora y S es la matriz de las fuentes. Por lo tanto, ICA obtiene $S=WX$, donde S es la matriz con las fuentes, W la matriz des-mezcladora y X los datos originalmente medidos. En nuestro caso tenemos dos fuentes, LS y HS, pero en la lectura con dos canales, inicialmente cada lectura de canal tiene mezclados los datos LS y HS, por lo cual aplicar ICA implicaría separar LS en un canal y HS en otro canal.

Modelos Mezclados Gaussianos (GMM)

Un modelo GMM es una tripleta Λ compuesta por las medias, covarianzas y ponderaciones. El modelado GMM se sirve del algoritmo de Expectation-Maximization (algoritmo EM) para calcular las tripletas $\Lambda_i = \{m_i, \mu, \Sigma_i\}$. Este cálculo se efectúa sobre vectores acústicos extraídos de las distintas grabaciones (LS u HS). La media μ representa el promedio de todos los vectores, mientras que la matriz de covarianza Σ_i modela la variabilidad de las características en una clase acústica [21], o la interdependencia de una componente de otra.

$$p(\vec{x}|\Lambda) = \sum_{i=1}^M m_i b_i(\vec{x}) \tag{17}$$

En la ecuación 17, x es un vector aleatorio D-dimensional (vectores acústicos MFCC, o Cuartiles). $b_i, \forall_i=1, \dots, M$ son las densidades de probabilidad de la mezcla y $m_i, \forall_i=1, \dots, M$ son las ponderaciones de cada densidad en el modelo. Cada densidad componente es una función Gaussiana D-dimensional [18, 19]. Además, las ponderaciones de las mezclas deben satisfacer la restricción $\sum_i^M m_i = 1$. Cada densidad Gaussiana contiene los parámetros representados en la Ecuación 18.

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (\vec{x} - \bar{\mu}_i)^T \Sigma_i^{-1} (\vec{x} - \bar{\mu}_i) \right] \tag{18}$$

Un HMM es un autómata finito basado en un conjunto de estados $S = \{S_1, S_2, \dots, S_N\}$ que no son directamente observables (ocultos). En nuestros experimentos, cada estado de un HMM está compuesto por un GMM, el cual modela las observaciones correspondientes a ese estado. Un HMM está definido por los siguientes componentes (Figura 4) [30]:

- › Los elementos de la matriz de transiciones $A = \{a_{i,j}, 1 \leq i, j \leq N\}$, corresponden a la probabilidad de transitar de un estado S_i a un estado S_j , es decir, $a_{i,j} = P(q_{t+1} = S_j | q_t = S_i), 1 \leq i, j \leq N$.
- › La función de emisión de cada estado $j, B = \{b(O|S_j)\}$ (un GMM), denota la probabilidad de que una observación sea generada en el estado S_j . En el caso de los experimentos de este estudio, las observaciones correspondieron a vectores acústicos MFCC o Cuartiles.
- › $\pi = \{\pi_i\}$, las probabilidades de estar inicialmente en un estado $i, \pi_i = P[q_1 = S_i], 1 \leq i \leq N$ con $\pi_i \geq 0$ y $\sum_i^N \pi_i = 1$.

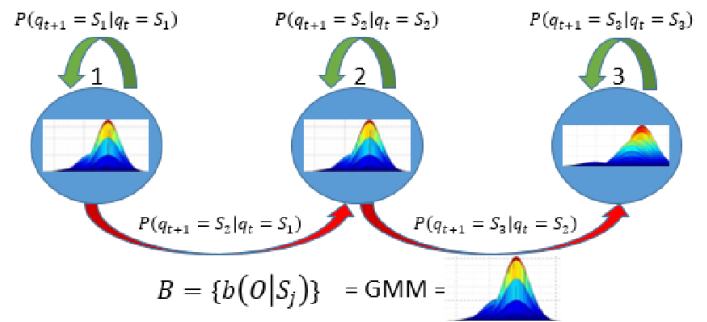


FIGURA 4. Modelo HMM de 3 estados representados por modelos GMM.

Para el entrenamiento de los parámetros HMM, se requiere una secuencia de observaciones $\{O\}$, aplicando el algoritmo Baum-Welch [31], el cual determina

los parámetros maximizando la *verosimilitud* o probabilidad $P(O_i|A)$. En la etapa de evaluación, se requiere calcular $P(O|A)$, dado el modelo A y una secuencia O de observaciones; aquí se aplicó algoritmo de forward-backward [31]. La arquitectura HMM fue de tipo izquierda-derecha (Bakis). Se puede destacar que las probabilidades de transición y de estado inicial, fueron inicializadas aleatoriamente.

Base de Datos

Con respecto a los experimentos de señales LS se usó la base de datos RALE. RALE consiste de un conjunto de grabaciones *wav de sonidos LS [32], la cual fue desarrollada por la universidad de Winnipeg, Canadá. Dichas señales fueron filtradas con un pasa-altas a 7.5 Hz para suprimir cualquier offset DC mediante un filtro Butterworth de primer orden. Además, se aplica un filtro Butterworth pasa-bajas de octavo orden a 2.5 kHz para evitar traslape. Las señales en la base de datos están muestreadas a 11025 Hz. Además se efectuaron grabaciones de señales LS en estudiantes universitarios con las mismas características que RALE para aumentar el corpus. Las señales normales LS fueron segmentadas manualmente y automáticamente para obtener registros correspondientes solo a la inspiración o a la espiración de cada una de las señales, obteniendo un total de 15 señales de inspiración y 15 señales de espiración.

El conjunto de señales HS utilizadas para los experimentos provienen de bases de datos que están disponibles para propósitos académicos o científicos [32, 33]. De aquí se utilizaron 15 señales normales, las cuales fueron segmentadas manual y automáticamente para obtener 15 sonidos S1 y 15 S2. La frecuencia de muestreo para los registros es de 11025 Hz, con un formato tipo *wav monoaural. Las señales originales fueron capturadas a una tasa de 44 kHz y 22 kHz, pero se sub-muestrearon a 11 kHz. Las señales utilizadas fueron particionadas para la etapa de entrenamiento y de evaluación aplicando Validación Cruzada (VC) [18, 34].

Por mediciones experimentales, se obtuvo que la duración en fases de las señales del repositorio de datos RALE está en el orden de 1.5 segundos para la inspiración y 2.5 para la espiración. Para HS, la fase S1 tiene una duración de alrededor de 0.1 a 0.12 segundos; la fase S2 se encuentra entre 0.8 a 0.14 segundos [22].

RESULTADOS Y DISCUSIÓN

Primeramente, se explicará el proceso de detección automática de S1 y S2, luego en otro apartado se explicará la detección automática de la inhalación y la exhalación con la ayuda de VAD, y posteriormente se mostrarán los resultados de clasificación utilizando los modelos HMM-GMM.

Detección de S1 y S2

En las señales HS, primero se normalizó la señal en el rango de ± 1 , luego se aplicó un filtro pasa bajas Butterworth con una frecuencia de corte de 150 Hz, y con un orden de 10. Después del filtrado, se calcula la envolvente de Hilbert de la señal, el resultado se muestra en la Figura 5, donde la envolvente de Hilbert se compara con la señal original (es decir, la señal normalizada). Posteriormente, se suaviza la envolvente de Hilbert mediante un filtro Butterworth de orden 5, con una frecuencia de corte de entre 7-25 Hz. Aquí, se efectuaron varios experimentos para determinar la mejor frecuencia de corte, resultando en 8 Hz, como se aprecia en la Figura 6. A continuación, se segmentaron las partes de la señal que corresponden al corazón como se observa en la Figura 7.

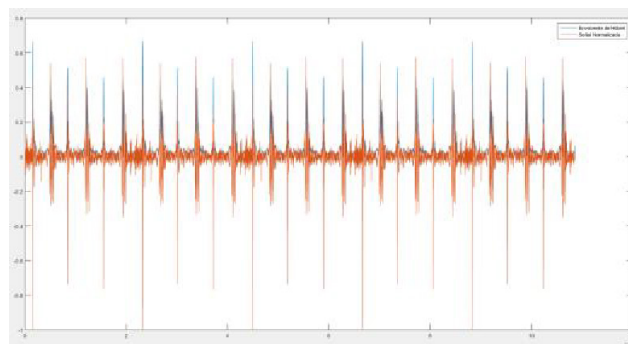


FIGURA 5. Envolvente de Hilbert.

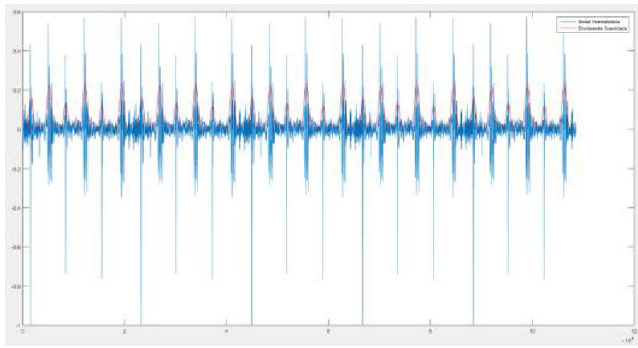


FIGURA 6. Envoltiva suavizada de Hilbert.

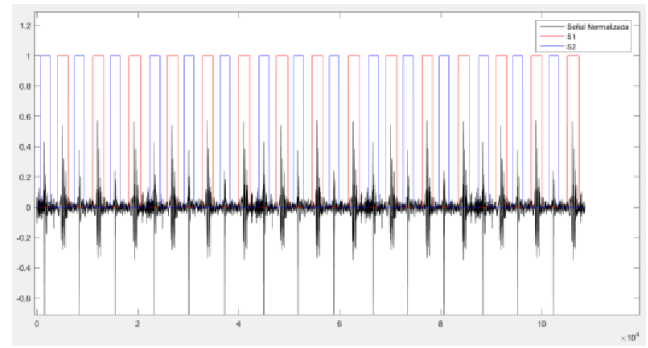


FIGURA 8. Identificación de S1 y S2.

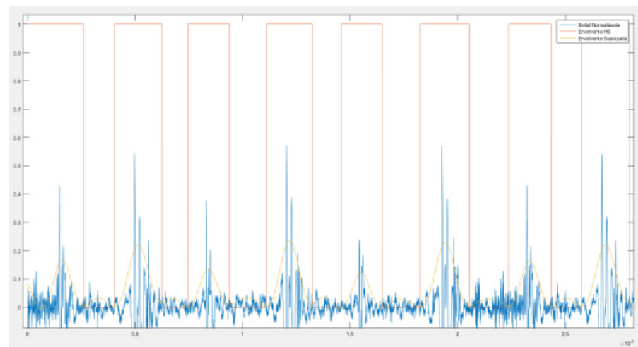


FIGURA 7. Identificación de los segmentos correspondientes al corazón.

Detección de Inspiración y Espiración

En lo que concierne a la detección de inspiración y espiración, primero se lee una señal capturada con el estetoscopio de dos canales, ambas señales de cada canal son centradas y blanqueadas, como se aprecia en la Figura 9. Inicialmente, las señales no están bien definidas si son LS o HS, debido a que los estetoscopios con los que fueron grabadas no estaban bien identificados pero aplicando ICA y correlación cruzada, las señales son etiquetadas a la clase que más se acerquen (LS o HS).

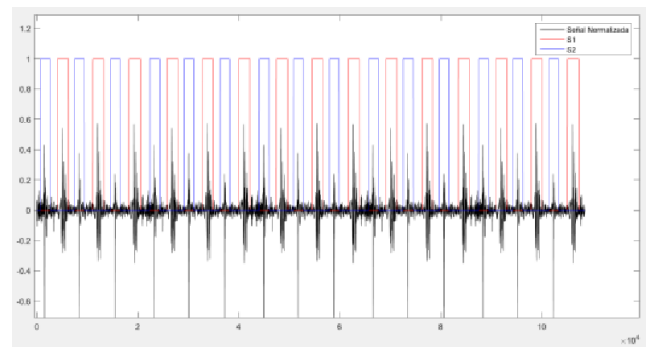


FIGURA 9. Señales HS y LS centradas y blanqueadas.

El siguiente paso es determinar la duración de S1, S2, Sístole y Diástole. Se inicia con el cálculo de umbrales para determinar la amplitud y establecer cuáles picos corresponden a S1 y cuáles a S2. Esto se logra aplicando máximos y mínimos, considerando un mínimo como inicio y otro mínimo como el final de un pico de la señal. Para distinguir cual segmento corresponde a sístole, y cual a diástole, primero se identifica con que sonido inicia la grabación, *i.e.*, S1 o S2. Esto es debido a que separamos la señal que contiene sístole y diástole en señales pares e impares, ya que si la señal comienza con S1 las señales impares corresponden a sístole, mientras que la señal par a diástole; así mismo, si la señal comienza con S2, la señal par representa sístole y la señal impar a diástole. El propósito de separar las señales en sístole y diástole es para tener una aproximación de cuantas pulsaciones por minuto tiene la persona de la cual se grabó la señal del corazón. Y a partir de estas señales, se calculan las anchuras (tiempo de duración) de S1 y S2, obteniendo la siguiente figura.

Se aplica un filtro Chebyshev Tipo 1 pasa altas con una frecuencia de corte de 150 Hz, lo que atenúa las componentes del corazón S1 y S2 que aun puedan persistir en la señal LS. A continuación se aplica el análisis de componentes independientes (ICA) para separar lo más posible las dos señales que se encuentran mezcladas en ambos canales, y al resultado de este proceso se le aplica correlación cruzada con cada una de las señales de entrada para determinar su correspondiente

clase, debido a que después de utilizar ICA se entre mezclan las señales y los otros de las variables con las que están asociadas. En este paso se selecciona la señal HS para ser correlacionada con las señales resultantes de la aplicación de ICA, ya que HS es más evidente. Aquellas señales con una correlación más alta implican que son de la misma clase, es decir pertenecen a la misma clase HS. En la Figura 10 se puede apreciar el resultado de aplicar ICA a nuestras señales, como se puede apreciar en la imagen, para el caso de LS resulta realmente conveniente, debido a que la forma que adopta después de este proceso, es como la que se encuentra en la literatura para LS, algo muy diferente de la señal original de entrada mostrada en la Figura 9. Para el caso HS, podemos observar que ICA cambio la forma normalmente mostrada en la literatura por otra muy diferente más similar al ruido, pero de momento esto no es preocupante ya que estamos solo interesados en obtener la señal LS.

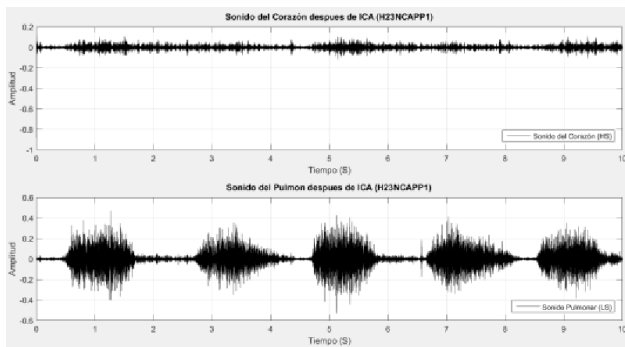


FIGURA 10. Señales LS y HS después de aplicar ICA.

En seguida, se le aplica VAD a la señal LS con 5 distintos umbrales para detectar todo lo que corresponda a la actividad de respiración, es decir, inspiración y espiración. Después, se aplica una función para limpiar los falsos positivos, o sea, todo aquello de la señal que no corresponda a inspiración y espiración pero que por algún motivo fueron detectados por VAD como actividad erróneamente. Ya limpias, las 5 señales son analizadas para determinar nuevamente a que clase (inspiración, espiración) pertenecen de una forma mayoritaria, es decir, si un segmento de actividad aparece en 4 de las

5 señales este segmento se toma como bueno y pasar a formar parte de una de las dos clases. A partir de esta señal, se construyen los segmentos que corresponden a las dos fases de la respiración (inspiración y espiración) que hasta este punto no se había determinado. En la Figura 11, podemos apreciar esta primera segmentación donde está identificada la actividad de respiración.

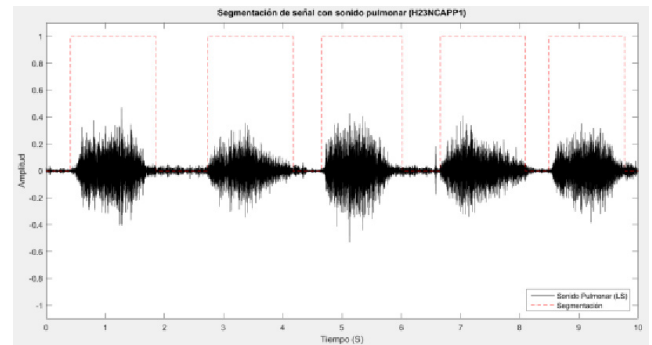


FIGURA 11. Señal LS en donde esta segmentada la actividad respiratoria.

Al final, a cada uno de los segmentos que se observan en la Figura 11, se le calculan los vectores MFCC, haciendo énfasis en la sexta componente MFCC de cada vector, cada componente MFCC es sumada para obtener un promedio, el cual es utilizado como umbral. Este umbral permitirá distinguir la inspiración de la espiración. Los segmentos con valores en su sexta componente menores al umbral se consideran como espiración, de lo contrario inspiración. El resultado de este proceso se puede apreciar en la Figura 12, donde la señal LS está totalmente segmentada en inspiración y espiración.

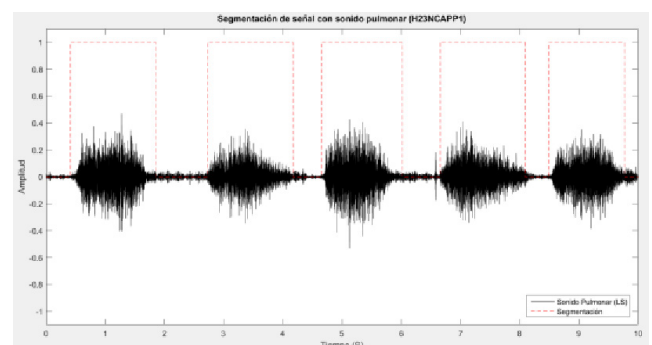


FIGURA 12. Segmentación Automática de la inspiración y espiración.

En esta etapa, al final se multiplica por uno los segmentos que pertenecen a inspiración y espiración y por cero el resto. Además, cada segmento de inspiración o espiración es respaldado en un archivo *.wav por separado, obteniendo archivos de inhalación y archivos de exhalación.

Evaluación en eficiencia de clasificación

Con el propósito de evaluar la eficacia de los métodos implementados, resulta útil aplicar modelos HMM para clasificar las clases. Los resultados gráficos previos están relacionados con inspiración-espiración o S1-S2, por lo cual, se decidió realizar algunas evaluaciones de clasificación confrontando las alternativas de segmentación manual y automática de estos eventos. Los resultados se muestran más adelante aplicando clasificación sin PCA, y luego con PCA a fin de analizar si esto incrementa la eficiencia.

TABLA 1. Resultados de la Clasificación de Insp-Exp, S1-S2 con segmentación manual.

Señal	Vector	Gaussianas	Est.	Efic. s/PCA (%)	Efic. c/PCA (%)
Insp-Exp	MFCC	3	3	65.76	88.84
		2	3	61.15	83.07
		2	2	61.92	86.15
	Cuartil	3	3	78.46	79.23
		2	3	75.76	82.30
		2	2	68.46	71.93
S1-S2	MFCC	3	3	96.92	84.61
		2	3	93.46	84.61
		2	2	91.92	84.61
	Cuartil	3	3	50	52.5
		2	3	50	68.46
		2	2	82.30	57.69

La Tabla 1 muestra resultados sobre señales segmentadas manualmente considerando distintas configuraciones en lo que respecta al número de gaussianas por estado y la cantidad de estados por modelo HMM.

Además, se aplicaron dos distintos extractores de características, es decir, MFCC y Cuartiles.

En la Tabla 1, la parte superior especifica el tipo de señal, vector, configuración de modelo HMM (# Gaussianas, # Estados), antes de aplicar PCA y aplicando PCA. Las columnas de la izquierda especifican las clases, la segunda columna el tipo de vector de características empleado, en color amarillo se encuentra el mejor resultado con las distintas combinaciones posibles de los métodos utilizados. En lo que respecta a la clasificación de segmentos de inspiración y espiración obtenidos manualmente, MFCC arroja los mejores resultados con una configuración de 3 Gaussianas por Estado y 3 Estados por modelo HMM aplicando PCA.

Incluso, tanto con MFCC como Cuartiles hubo un aumento de eficiencia de clasificación después de aplicar PCA, lo que puede significar que las señales tenían ruido, o no era el mejor espacio de coordenadas y PCA proyectó los vectores en otro espacio de una manera más eficiente y des-correlacionado.

Con relación a S1 y S2, MFCC arroja los mejores resultados en clasificación y con 3 gaussianas por estado y 3 estados en HMM como el caso de inspiración-espiración. Tal parece que las señales HS utilizadas en los experimentos estaban más limpias o no había correlación importante entre componentes vectoriales, ya que de hecho el efecto de PCA fue adverso. En ambos experimentos de clasificación, los mejores resultados fueron superiores al 80% con segmentación manual.

En la Tabla 2 se presentan los resultados de la eficiencia de clasificación obtenidos aplicando segmentación automática, apoyándose en la transformada de Hilbert así como Máximos y Mínimos para el caso del corazón, y VAD en el caso del pulmón. Es decir, las señales para los modelos y la evaluación fueron obtenidas automáticamente.

TABLA 2. Resultados de la Clasificación de Insp, Exp, S1, S2 con segmentación automática.

Señal	Vector	Gaussianas	Est.	Efic. s/PCA (%)	Efic. c/PCA (%)
Insp-Exp	MFCC	3	3	73.66	73
		2	3	80.33	74.33
		2	2	77.66	63.66
	Cuartil	3	3	79	79
		2	3	74	85
		2	2	74	76.33
S1-S2	MFCC	3	3	99.66	98.66
		2	3	99	96.33
		2	2	97.33	96.66
	Cuartil	3	3	77.33	99
		2	3	99.33	97
		2	2	99.33	98

En la Tabla 2, las columnas tienen el mismo significado que en el caso de la Tabla 1. La primera observación que se obtiene de la Tabla 1 y Tabla 2, es que la segmentación automática sin PCA mejora la eficiencia de clasificación, lo cual puede significar que las técnicas aplicadas son más eficientes que el oído y el ojo humano al identificar un evento en las señales.

Con segmentación automática y para el caso de MFCC, la eficiencia es mejor sin aplicar PCA a señales LS-HS, tal vez porque son señales más limpias, luego PCA no es necesario. De hecho, el efecto de PCA disminuye en segmentación automática la eficiencia salvo en el caso de cuartiles para Insp-Exp.

Lo anterior podría indicar, que aunque la segmentación automática tiene señales más limpias correspondientes a los eventos que la segmentación manual, aun así las señales LS conservan aun un poco de ruido, el cual PCA logra paliar.

De cualquier manera, el efecto de PCA no es tan adverso, ya que en promedio no disminuye la eficiencia sistemáticamente en todos los casos, lo cual es algo que habría que explorar en un futuro.

CONCLUSIONES

En este trabajo se presentaron experimentos y una metodología para la detección, localización y extracción de eventos en sonidos del pulmón así como sonidos del corazón, específicamente inspiración-espирación en LS y S1-S2 en HS.

La detección manual permitió extraer archivos .wav de los eventos en LS-HS los cuales fueron utilizados para calcular modelos HMM-GMM y efectuar la clasificación, arrojando valores promedio superior al 80%, mientras que con PCA mejoraron ligeramente en forma promedio. En cuanto a los experimentos con detección automática, en promedio logran casi el 90%. Es importante remarcar que la detección automática fue más eficiente que la detección manual, lo cual es alentador ya que las capacidades sensoriales de un médico no siempre son suficientes durante la auscultación y esta alternativa podría mejorar un diagnóstico. En promedio, la arquitectura HMM-GMM más exitosa fue con 3 Gaussianas (3G) por Estado y 3 Estados (3E) por Modelo.

Para el caso de las señales HS, se comprobó que la eficiencia con segmentación automática mejoró sensiblemente casi logrando el 100% de eficiencia. En el caso de LS, la eficiencia con segmentación automática se mantuvo más o menos igual con segmentación manual, a excepción de la opción con 2G y 3E con PCA, en la cual los Cuartiles alcanzan un 85% de eficiencia; esta fue la eficiencia de clasificación de eventos más alta lograda para Insp-Exp. Incluso, PCA en este caso mejoró la eficiencia con señales LS, lo cual hace pensar que las señales LS contienen más contaminación que las señales HS utilizadas en nuestros experimentos.

En un futuro, sería interesante ampliar los experimentos a patologías, para lo cual habrá que aumentar el corpus de señales LS, e incluir sibilancias y crepitaciones. Concerniente a las señales HS, los eventos considerados son S3 y S4, comunes en patologías. Además, sería conveniente investigar versión mejorada de VAD.

REFERENCIAS

- [1] A. Azarbarzin and Z. M. K. Moussavi, "Automatic and Unsupervised Snore Sound Extraction From Respiratory Sound Signals," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 1156-1162, 2011.
- [2] J. Earis, "Lung sounds," *Thorax*, vol. 47, pp. 671-672, 1992.
- [3] "Características mínimas de infraestructura y equipamiento de hospitales y consultorios de atención médica especializada", 8 de enero de 2013.
- [4] P. Forgacs, "Lung sounds," *Br J Dis Chest*, vol. 63, pp. 1-12, 1969.
- [5] M. Lozano, J. A. Fiz, R. Jan, and x00E, "Automatic Differentiation of Normal and Continuous Adventitious Respiratory Sounds Using Ensemble Empirical Mode Decomposition and Instantaneous Frequency," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, pp. 486-497, 2016.
- [6] K. Kosasih, U. R. Abeyratne, V. Swarnkar, and R. Triasih, "Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 1185-1194, 2015.
- [7] J. Herzig, A. Bickel, A. Eitan, and N. Intrator, "Monitoring Cardiac Stress Using Features Extracted From S1 Heart Sounds," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 1169-1178, 2015.
- [8] D. Emmanouilidou, E. D. McCollum, D. E. Park, and M. Elhilali, "Adaptive Noise Suppression of Pediatric Lung Auscultations With Real Applications to Noisy Clinical Settings in Developing Countries," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 2279-2288, 2015.
- [9] F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing Heart Sounds in Respiratory Signals Using Singular Spectrum Analysis," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 3360-3367, 2011.
- [10] CHEN Ji, HOU Hai-Liang, LUO Liang-Cai, and C. Yun, "Automatic Identification Method for the First and Second Heart Sound Based on Double-threshold," *Journal Computer Engineering*, vol. 38, pp. 174-177, 2012.
- [11] K. Bajelani, M. Navidbakhsh, H. Behnam, J. D. Doyle, and K. Hassani, "Detection and identification of first and second heart sounds using empirical mode decomposition," *Journal of Engineering in Medicine* vol. 227 pp. 976-987, Aug 9, 2013 2013.
- [12] M. S. Dong and S. Hangsik, "A Localization Method for First and Second Heart Sounds Based on Energy Detection and Interval Regulation," *Journal of Electrical Engineering and Technology*, vol. 10, pp. 2126-2134, 2015.
- [13] S. Sun, Z. Jiang, H. Wang, and Y. Fang, "Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform," *Comput. Methods Prog. Biomed.*, vol. 114, pp. 219-230, 2014.
- [14] A. Abushakra and M. Faezipour, "Acoustic Signal Classification of Breathing Movements to Virtually Aid Breath Regulation," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, pp. 493-500, 2013.
- [15] Hupendra Devangan and N. Jain, "A Review on Classification of Adventitious Lung Sounds," *International Journal of Engineering Research & Technology*, vol. 4 2015.
- [16] D. M. Istrate, "Detection et Reconnaissance des Sons pour la Surveillance Médicale," These pour obtenir le grade de docteur de l'INPG: spécialité Signal, Image, Parole, Télécoms docteur Institut National Polytechnique de Grenoble, Grenoble, France, 2003.
- [17] D. Pearce, "Developing the ETSI Aurora advanced distributed speech recognition front-end and what next?," in *Automatic Speech Recognition and Understanding*, 2001. ASRU '01. IEEE Workshop on, 2001, pp. 131-134.
- [18] P. Mayorga, C. Druzgalski, and J. Vidales, "Quantitative models for assessment of respiratory diseases," in *Health Care Exchange (PAHCE)*, 2010 Pan American, 2010, pp. 25-30.
- [19] P. Mayorga, L. Besacier, R. Lamy, and J. F. Serignat, "Audio packet loss over IP and speech recognition," in *Automatic Speech Recognition and Understanding*, 2003. ASRU '03. 2003 IEEE Workshop on, 2003, pp. 607-612.
- [20] J. S. Yoon, G. H. Lee, and H. K. Kim, "A MFCC-Based CELP Speech Coder for Server-Based Speech Recognition in Network Environments," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E90-A, pp. 626-632, 2007.
- [21] P. Mayorga, C. Druzgalski, O. H. González, and H. S. Lopez, "Modified classification of normal Lung Sounds applying Quantile Vectors," *Engineering in Medicine and Biology Society (EMBC)*, pp. 4262 - 4265, 2012.
- [22] J. D. Echeverry, A. F. López, and J. F. López, "Reconocimiento de valvulopatías cardíacas en señales de fonocardiografía empleando la transformada Gabor," *Scientia Et Technica*, vol. XIII, pp. 139-144, 2007.
- [23] M. P., O. M., G. O. H., F. N., and L. V., "Quantile Acoustic Vectors vs. MFCC Applied to Speaker Verification," *International Journal of Advanced Robotic Systems*, 2013.
- [24] A. Mondal, A. K. Kumar, P. S. Bhattacharya, and G. Saha, "Boundary estimation of cardiac events S1 and S2 based on Hilbert transform and adaptive thresholding approach," in *Medical Informatics and Telemedicine (ICMIT)*, 2013 Indian Conference on, 2013, pp. 43-47.
- [25] N. S. K. Jongseo Sohn, Wonyong Sung, "A Statistical Model-Based Voice Activity Detection," *IEEE SIGNAL PROCESSING LETTERS*, vol. 6, January 1999.
- [26] A. A. a. M. Faezipour, "Acoustic Signal Classification of Breathing Movements to Virtually Aid Breath Regulation," *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, vol. 17, MARCH 2013.
- [27] P. Mayorga Ortiz, C. Druzgalski, M. A. Criollo Arellano, and O. H. González Arriaga, "GMM y LDA aplicado a la detección de enfermedades pulmonares," *Revista mexicana de ingeniería biomédica*, vol. 34, pp. 131-144, 2013.
- [28] Martinez W. L and M. A. R, *Computational Statistics Handbook with Matlab*, Second ed., 2008.

- [29]** J. P. Rosca, Independent component analysis and blind signal separation : 6th international conference, ICA 2006, Charleston, SC, USA, March 5-8, 2006 : proceedings. Berlin ; New York: Springer, 2006.
- [30]** P. Kannadaguli and V. Bhat, "A comparison of Gaussian Mixture Modeling (GMM) and Hidden Markov Modeling (HMM) based approaches for Automatic Phoneme Recognition in Kannada," in Signal Processing and Communication (ICSC), 2015 International Conference on, 2015, pp. 257-260.
- [31]** L. R. Rabiner and B. H. Juang, Fundamentals of speech recognition. Englewood Cliffs, N.J.: PTR Prentice Hall, 1993.
- [32]** RALE, Lung Sounds Repository [Online]. Available: <http://www.rale.ca/Repository.htm>
- [33]** Texas Heart Institute [Online]. Available: <http://www.texasheartinstitute.org/AboutUs/index.cfm>
- [34]** P. Mayorga, C. Druzgalski, O. H. Gonzalez, A. Zazueta, and M. A. Criollo, "Expanded quantitative models for assessment of respiratory diseases and monitoring," Pan American Health Care Exchange (PAHCE), pp. 317 - 322, 2011.