

Determinación del Tamaño Óptimo de Modelos HMM-GMM para Clasificación de las Señales Bioacústicas

P. Mayorga-Ortiz¹, C. Druzgalski², J.E. Miranda Vega¹, V. Zeljkovic³

¹Depto. de Posgrado, Instituto Tecnológico de Mexicali. Av. Tecnológico S/N, Mexicali, B.C. 21396, México.

²Elec. Eng. Dept., California State University, Long Beach, CA 90840, USA.

³CPES Dept., The Lincoln University PA, USA

RESUMEN

Este artículo está relacionado con el análisis y la propuesta de una arquitectura HMM-GMM para clasificación de señales HS y LS, haciendo un énfasis en el tamaño del modelo. Actualmente, las enfermedades respiratorias y cardiovasculares son un problema a nivel mundial y con una alta mortandad, esto podría ser disminuido mediante un diagnóstico temprano y objetivo; las herramientas digitales y el empleo de reconocimiento de patrones ampliarían las perspectivas de aplicación. Particularmente, aquí se demuestra que los modelos HMM-GMM son eficientes para consultorios de atención primaria, así mismo los extractores de características tales como MFCC y Cuantiles mejoran la tarea de clasificación. Si bien la visualización con siluetas, dendrogramas y algoritmos tales como BIC no son concluyentes cuando se aplican GMM's, no obstante sí fue el punto de partida para dimensionar el tamaño del modelo, disminuyendo la cantidad de experimentos con distintos tamaños del mismo. Adicionalmente, se constata que la estructura de señales normales HS y LS cambian cuando hay patologías y permite la clasificación aplicando MFCC o Cuantiles. Además, se observa que con una gran cantidad de datos se podrían obtener modelos más robustos y adaptados, pero esto no es una limitante para el cálculo de los modelos.

Palabras clave: cuantíl, Coeficientes Cepstrales en Frecuencia Mel (MFCC), Modelos Ocultos de Markov (HMM), Modelos Mezclados Gaussianos (GMM).

Correspondencia:

Jesús Elías Miranda Vega

Depto. de Posgrado, Instituto Tecnológico de Mexicali. Av.
Tecnológico S/N, Mexicali, B.C. 21396, México.

Correo electrónico: Ing.jesus_elias@hotmail.com

Fecha de recepción:

21 de julio de 2015

Fecha de aceptación:

30 de diciembre de
2015

ABSTRACT

This paper demonstrates the analysis and proposed HMM-GMM models architecture to classify heart and lung sounds (HS and LS) signals emphasizing the model size optimization. Respiratory and cardiovascular diseases continue to represent one of the major worldwide healthcare problems associated with a high mortality rate, which can be reduced by an early and effective diagnosis; in this context, the use of digital tools utilizing signal pattern recognition allows efficient screening for abnormalities and their quantitative assessment. In particular, the HMM-GMM models demonstrated their efficiency in normal and traditionally noisy environments in light of very low intensities of these auscultation signals used as diagnostic indicators. Furthermore, applied MFCC and Quantiles feature extractors improve overall classification. While characterization with silhouettes, dendrograms and algorithms such as BIC was inconclusive when GMM was applied, however they were useful as a starting point in the determination of a size of the model as it allowed a reduction in the number of iterations considering different model size. In addition one can note that application of MFCC or Quantiles allowed differentiating the characteristics of normal HS and LS from those associated with pathological conditions. Furthermore, it was observed that a large amount of data leads to more robust and adapted models, but does not limit the calculation demand. Overall, this approach may enhance efficiency and precision of the diagnostic screening for abnormal auscultation indicators.

Keywords: quantil, Mel Frequency Cepstrum Coefficients (MFCC), Hidden Markov Models (HMM), Gaussian Mixture Models (GMM).

INTRODUCCIÓN

Los sonidos del pulmón (LS, por sus siglas en inglés) y los sonidos del corazón (HS, por sus siglas en inglés) son parte de las señales generadas por el cuerpo humano, las cuales son utilizadas para propósito de diagnóstico médico. En la actualidad existen distintos niveles de atención médica [1], pero incluso los niveles básicos deben contar con cierta funcionalidad, lo que involucra al menos equipo básico como un estetoscopio, e incluso por normatividad una computadora para el expediente electrónico [2]. Por lo tanto, las nuevas normatividades favorecen el uso de sistemas que permiten auscultar de manera digital los sonidos del corazón y del pulmón. Desafortunadamente, los distintos sonidos del corazón y del pulmón se traslapan en un rango de frecuencia importante [3-9], además de que el ruido ambiental durante las sesiones de auscultación dificulta aún más la tarea del diagnóstico.

Del párrafo anterior se remarcan algunos elementos importantes, el uso de computadora y estetoscopio, que son convenientes cuando se aplica reconocimiento de patrones como apoyo en la auscultación médica de primer contacto, para la detección de enfermedades respiratorias [1-6]. La auscultación del pecho con estetoscopio, constituye una herramienta portable de bajo costo ampliamente utilizada para detección de enfermedades respiratorias. Desafortunadamente, como ya se mencionó, el estetoscopio presenta varios retos, como el ruido ambiental y el traslape de los sonidos del corazón (HS) con los sonidos del pulmón (LS). Esto ha atraído la atención de la comunidad científica para enfocar esfuerzos en esta dirección. Existen técnicas con esquema jerárquico de Modelos Mesclados Gaussianos (GMM) y Maquinas de Soporte Vectorial (SVM), logrando resultados del 90% en sensibilidad y especificidad [7].

Algunas aproximaciones son dirigidas a patologías que son problemáticas regionales, como el enfoque de extraer características acústicas de la tos y crepitaciones para reforzar vectores de coeficientes cepstrales en frecuencia Mel (MFCC) y la aplicación de ondículas [8]. Otros trabajos se dirigen al análisis y monitoreo de las ondas sonoras del corazón [9]. Estos evidencian que hay características que cambian durante stress cardiaco, y que el cambio es más significativo para personas con problemas cardiacos [9]. Otros autores mencionan que si bien el uso de estetoscopio es una herramienta de bajo costo, la agitación sobre todo en niños contamina los registros de los sonidos [10]. Por lo tanto, ellos proponen un esquema automatizado multi-banda para supresión de ruido, y mejorar la calidad de las señales de auscultación contra fuerte contaminación de fondo [10].

El estetoscopio puede también ser de utilidad en la detección de enfermedades del corazón, i.e., los sonidos de las válvulas aortica (A2) y pulmonar (P2) pueden diagnosticarse con la duración de la energía de frecuencias instantáneas (EIF), en particular el segundo sonido del corazón (S2) [11].

Además, proponen medir los parámetros relevantes identificando el inicio y fin de A2 y P2, pero incluyendo el diagnóstico de los EIFs de A2 y P2 examinados. Por lo tanto, este método conduce explícitamente a distinguir los S2s Normales/Anormales y los tipos de separación [11]. Otra alternativa para separar la señal S2 de HS, es una descomposición no estacionaria que permite lidiar con traslapes y energía modelando subcomponentes de S2 [12]. Los autores proponen el método de descomposición de vibración de Hilbert (HVD) [12]. Además, proponen localizar A2 y P2 usando un suavizado con la distribución de Wigner-Ville seguida por el método de reasignación. Finalmente, las separaciones son calculadas tomando las diferencias entre el promedio de los índices de tiempo de A2 y

P2 [12].

Otros trabajos en segmentación de sonidos del corazón (HSS) para localizar el primer (S1) y segundo sonido del corazón (S2), aplican descomposición en modo ensamble empírico (EEMD) combinando curtosis y a la técnica le llaman HSS-EEMD/K [13]. Cuando se efectúan registros sobre el pecho, la interface tórax-micrófono genera una distorsión en la medición, con el propósito de restaurar el sonido del pulmón se propone una ecualización de sonidos crepitantes, para lo cual se aplican tonos entre 100 y 1200 Hz en la boca, midiéndolos en la boca y en el pecho, y generando curvas de atenuación promedio y un ecualizador de tiempo discreto en sonidos crepitantes [14]. Algunas técnicas no invasivas, se orientan en ayudar a regular la respiración mediante realidad virtual (VR), se apoyan en vectores MFCC con técnicas de segmentación. Además, usan detección de actividad de voz (VAD) y establecen umbrales lineales a la señal acústica de la respiración, capturándola y usando un micrófono para representar las diferencias entre inhalar y exhalar en el dominio de la frecuencia [15]. Otro enfoque para disminuir la interferencia de los sonidos del corazón en los sonidos del pulmón, es la localización de los componentes primarios de sonidos del corazón. Aquí, el análisis de espectro singular (SSA), es una técnica de análisis de series de tiempo que puede ser utilizada [16].

Un estudio sobre pérdida de paquetes en internet y su efecto en reconocimiento de voz y de locutor, desarrollada con una base de señales con 7080 registros de tres frases distintas leídas por 295 locutores, muestra el comportamiento basado en modelos acústicos con cúmulos como Modelos Mesclados Gaussianos (GMM) y modelos acústicos con Modelos Ocultos de Markov (HMM); estos últimos han mostrado fortaleza en la detección secuencias de eventos, i.e., palabras [17, 18]. Los resultados indican que los modelos acústicos con cúmulos son más sensibles a distorsión

constante como codificación, pero menos a la pérdidas de segmentos de información; mientras que los modelos acústicos con HMM son más sensibles a pérdida de información, pero menos a la distorsión constante [17-19]. Esto es de utilidad, si se piensa en zonas remotas o donde sería conveniente un escenario de internet, y que la información requiera codificarse o comprimirse para respaldarla o transmitirla. Además, los modelos HMM ya han demostrado su fortaleza en reconocimiento del habla, donde el significado depende de la secuencia de los fonemas [18], por lo cual se propone validar estas metodologías en sonidos del pulmón y del corazón, donde la secuencia de eventos es significativa.

METODOLOGÍA Y BASE DE DATOS

Los extractores de características destacan las diferencias entre clases mediante valores, de tal manera que mejoran el proceso de clasificación. Una buena clasificación requiere de modelos robustos por lo cual es importante la mejor configuración y tamaño. En esta sección se revisan algunas metodologías visuales como dendrogramas y siluetas como un primer criterio del tamaño. Además, el Criterio de Información Bayesiana (BIC), es también aplicado como un intento para dimensionar los modelos. Posteriormente, se describen los modelos GMM y HMM con base para la clasificación.

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 [20-23]. En este trabajo, se experimentó con ventanas Hamming de 30 ms y 15 ms de corrimiento en las señales HS, 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 posteriormente la transformada inversa de Fourier. Dado que la energía será real y par, la transformada de Fourier inversa es un producto interno donde subsisten únicamente las partes pares, resultando igual a calcular la Transformada Discreta Cosenoidal (DCT). El resultado final es un vector de características llamado MFCC [15, 18, 24]

VECTORES CUANTÍLICOS

Otro tipo de vectores acústicos son los Cuantiles, estos se basan en la función de distribución acumulativa (CDF). 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 la función de densidad de probabilidad continua $f(x)$ a través de la ecuación (1):

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

En caracterización acústica, la idea es encontrar un número dado de coeficientes Cuantílicos q_p , por lo que se parte de la transformada inversa de la CDF. Específicamente, en señales acústicas como LS y HS es necesario realizar los cálculos en el límite de la estacionariedad; este límite está determinado por el tiempo en que los eventos conservan sus características estadísticas. Considerando una tasa de 15 respiraciones por minuto (normalmente el rango está entre 12-20 respiraciones por minuto para sujetos adultos saludables, y mucho mayor para niños pequeños). 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. 400 ms por vector Cuartílico con corrimientos de 300 ms, nos arroja alrededor de 5 o más vectores por fase [25, 26]. 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 [27]. Esto posibilita vectores MFCC de 30 ms con corrimientos de 20 ms.

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

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

La Ecuación (2) garantiza que la suma de la distribución de valores frecuenciales obtenidos a partir de la FFT será igual a 1, por lo cual N implica la normalización. 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 a cada uno de los respectivos coeficientes Cuartílicos. El cálculo del último Cuantíl no es importante ya que siempre es igual a 1, lo que resulta en un vector de 3 dimensiones.

$$A_{0.25} = \int_{-\infty}^{f_{0.25}} F_N(f) df, \dots, \quad (3)$$

$$A_{0.75} = \int_{-\infty}^{f_{0.75}} F_N(f) df \quad (4)$$

Algorítmicamente, A_p se calcula mediante una suma iterativa para obtener el área y detectar los valores frecuenciales correspondientes a $A = 0.25, \dots, A = 0.75$. Si bien éste fue el caso de Cuartiles, el mismo principio puede aplicarse a Octiles u otro tipo de Cuantíl. Para una descripción más extensa, se puede revisar el trabajo [26].

ANÁLISIS DE LA CANTIDAD DE CÚMULOS PARA EL DESARROLLO DEL MODELO

Existen diferentes técnicas para determinar la cantidad de cúmulos existentes en una clase. Específicamente, visualizar la cantidad de conglomerados o cúmulos en una clase es importante para dimensionar la configuración de un modelo. En esta sección se revisan algunas técnicas para visualizar la cantidad de cúmulos por clase. Las técnicas aquí discutidas son el análisis con dendrogramas, siluetas y BIC.

Un dendrograma es básicamente un diagrama de árbol invertido que se basa en la distancia de cada uno de los datos con respecto a todos los demás, y busca asociar aquellos que se encuentran más cerca entre sí (considerando una métrica de distancia, e.g. Euclidiana). Partiendo de esto, los datos se van asociando uno a uno hasta haber asociado la totalidad de los datos en cúmulos como se muestra en la Figura 1 [28].

Las siluetas permiten representar el número de cúmulos existentes en una serie de tiempo. El índice de silueta es el indicador del número ideal de cúmulos en una clase. En nuestro estudio (Figura 2), se calcularon todos los vectores acústicos para todas las señales de una clase y se aplicaron las siluetas. Conforme se aproxima al valor a 1 en el eje x, indica que el número de cúmulos en el eje y es más representativo de la clase. Por el contrario, los datos del lado izquierdo presentan incertidumbre de pertenecer a la clase [29, 30].

Otra alternativa para determinar la cantidad de cúmulos en una clase, es aplicando el Criterio de Información Bayesiano (BIC), como se muestra en la Figura 3. El modelo que tenga el valor más alto de BIC se considera el “mejor” modelo [31]. BIC es útil para estimar qué tan bien se ajustan el tipo de covarianza y número de cúmulos a los datos, para nuestro caso en un GMM, pero además los vectores medias y ponderaciones [31].

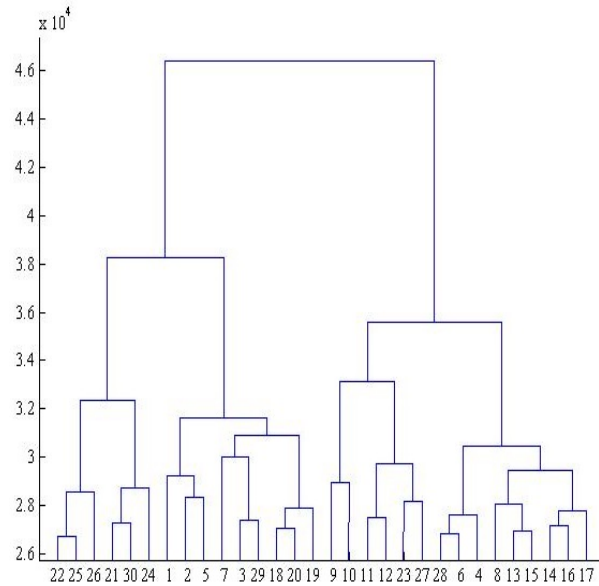


Figura 1. Dendrograma para la clase LS Normal

Este criterio se aproxima con:

$$p(X|M_g) \approx BIC_g = 2 \log p(X|\hat{\theta}_g M_g) - m_g \log(n) \quad (5)$$

Donde X son las observaciones por clase, M_g es el modelo, m_g es el número de parámetros independientes, $\hat{\theta}_g$ los parámetros estimados y n el número de vectores MFCC o cuartiles. En la Figura 3, cada curva corresponde a un GMM con distinto tipo de covarianza para la clase LS Normal.

Si bien las Figura 1, 2 y 3 corresponden todos los vectores acústicos obtenidos para señales LS Normales, el procedimiento se puede efectuar para cada clase de LS, o HS.

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 EM para calcular las tripletas $\Lambda_i = \{m_i, \bar{\mu}, \Sigma_i\}$. Este cálculo se efectúa sobre vectores acústicos extraídos de las distintas grabaciones (LS o HS). La media $\bar{\mu}$ representa el promedio de todos los vectores, mientras que la matriz de covarianza Σ_i modela la variabilidad de las características en una clase

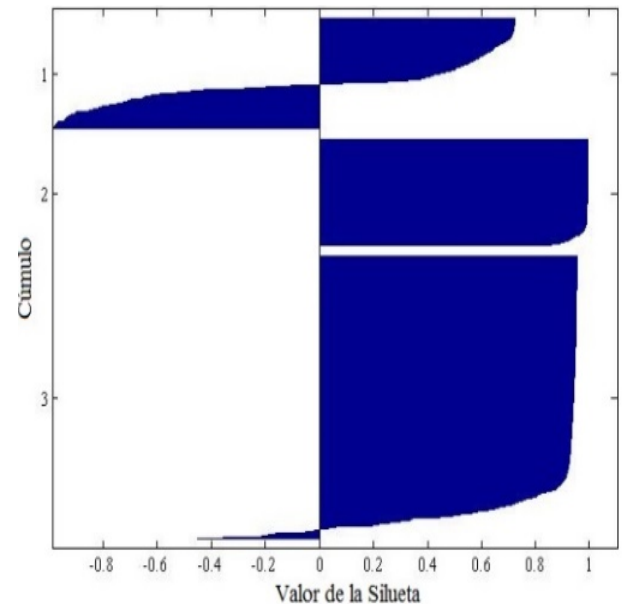


Figura 2. Silueta obtenida con las observaciones de la clase LS Normal.

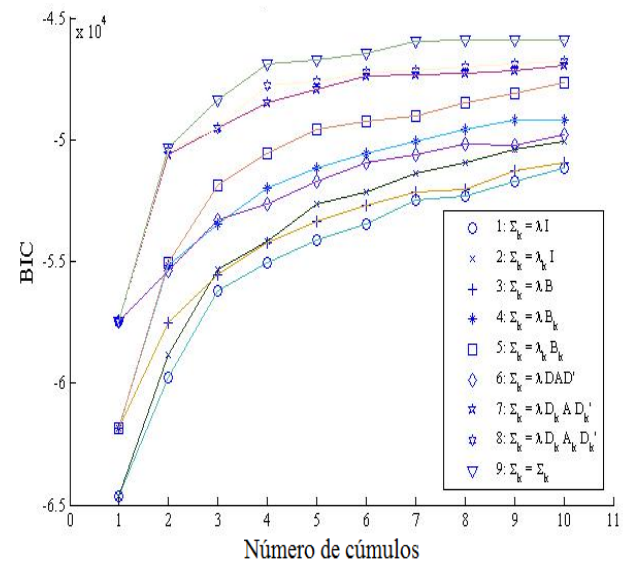


Figura 3. Curvas de BIC que muestran distinto número de cúmulos por tipo de covarianza para clase LS Normal.

acústica [21].

$$p(\vec{x}|\Lambda) = \sum_{i=1}^M m_i b_i(\vec{x}) \quad (6)$$

En la ecuación 5, $\vec{x}$ es un vector aleatorio D-dimensional (vectores acústicos MFCC, Cuartílicos u Octílicos), $b_i, \forall i = 1, \dots, M$ son las densidades componentes 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, 32-34]. 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 6 [18, 32-34].

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i) \right] \quad (7)$$

MODELOS OCULTOS DE MARKOV (HMM)

Un HMM es un autómata finito basado en estados que no son directamente observados. En nuestra metodología, cada estado en un HMM está constituido por un GMM, el cual modela las observaciones correspondientes a ese estado. Formalmente, un HMM es definido con los siguientes componentes [35]:

- $S = \{S_1, S_2, \dots, S_N\}$ el conjunto finito de los posibles estados (ocultos);
- La matriz de transiciones $A = \{a_{i,j}, 1 \leq i, j \leq N\}$, correspondiendo a la probabilidad de transitar de un estado S_i a un estado S_j $a_{i,j} = P[q_{t+1} = S_j | q_t = S_i], 1 \leq i, j \leq N$, con la restricción $a_{i,j} \geq 0, \sum_{j=1}^N a_{i,j} = 1$, donde q_t denota la probabilidad de un estado ocupado en un tiempo t .
- La función de emisión de cada estado j , $B = \{b(O|S_j)\}$ (en nuestro caso un GMM), denota la probabilidad de emitir una observación O cuando el sistema está en el estado S_j .
- $\pi = \{\pi_i\}$, 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 $\sum_{i=1}^N \pi_i = 1$.

En el caso de los experimentos de este estudio, las observaciones pueden ser vectores acústicos MFCC, Cuartiles u Octiles.

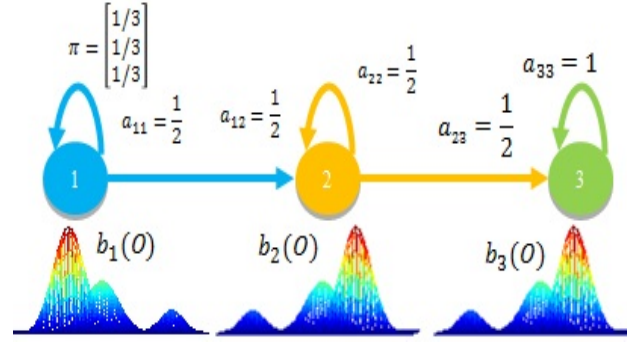


Figura 4. Modelo HMM con la tripleta $\lambda = (A, B, \pi)$.

Al igual que en el caso de los modelos GMM, es convencional expresar los modelos HMM como tripletas $\lambda = (A, B, \pi)$. Un ejemplo de modelo HMM para señales acústicas se muestra en la Figura 4 [36].

El entrenamiento o aprendizaje de los parámetros HMM, dado un conjunto o secuencia de observaciones $\{O_i\}$, es típicamente efectuado aplicando el algoritmo Baum-Welch [35], el cual determina los parámetros maximizando la *verosimilitud* o probabilidad $P(O_i|\lambda)$. En la etapa de evaluación, se requiere calcular $P(O|\lambda)$, dado el modelo λ y una secuencia O de observaciones; aquí se aplicó el algoritmo de forward-backward [35].

La arquitectura HMM fue de tipo izquierda-derecha (*Bakis*), como lo muestra la Figura 4. Aquí, el vector π denota las probabilidades iniciales (*a priori*) de estar en alguno de los estados q ; los valores $a_{i,j}$ son las probabilidades de transición entre estados, mientras que $b_i(O)$ es la probabilidad de que la observación O (vector acústico) haya sido emitido en el estado q_i (para este caso un GMM). Se puede destacar que las probabilidades de transición y de estado inicial, fueron inicializadas aleatoriamente.

BASE DE DATOS

El conjunto de señales HS utilizadas para los experimentos provienen de bases de datos que están disponibles para propósitos académicos o científicos [37, 38]. De ellas, se utilizó un

conjunto de grabaciones HS, el cual consta de 21 señales HS, con una frecuencia de muestreo de 11025 Hz, una duración de entre 9 y 12 segundos, formato tipo **wav* monoaural. Las señales originales fueron capturadas a una tasa de 44 kHz y 22 kHz, pero se submuestrearon a 11 kHz. De éstas señales se tomaron 7 HS Normales, 7 de Estenosis y 7 de Defecto Septal-Ventricular (VSD). Las señales utilizadas fueron particionadas para la etapa de entrenamiento y de evaluación aplicando Validación Cruzada (VC).

Con respecto a los experimentos de señales LS se utilizaron dos bases de datos: RALE y BDITM. RALE consiste en un conjunto de grabaciones **wav* de sonidos LS adventicios y Normales, 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. De RALE, sólo se utilizaron señales adventicias Crepitantes y Sibilantes (en inglés: Crackles y Wheeze).

La otra base de datos utilizada fue BDITM. Sus señales fueron obtenidas a partir de auscultaciones realizadas a 7 estudiantes del Instituto Tecnológico de Mexicali, todas con un estetoscopio digital. El rango de edad de los individuos fue de 21 a 26 años. Se realizaron de una a cuatro sesiones por sujeto, totalizando 13 sesiones. Cada sesión comprendió una auscultación traqueal, 8 puntos de auscultación en el pecho, y 8 puntos de auscultación en la espalda (es decir, 17 puntos en total), lo cual coincide con el protocolo de RALE [26]. El tamaño de muestra fue de 16 bits, en formato **wav* monoaural. Adicionalmente, se les removió la componente en DC.

En LS también se utilizó validación cruzada para la evaluación. Esto consistió en dejar una señal para evaluación y el resto

para el cálculo de los modelos. Pero la señal de evaluación cambió de manera iterativa, de modo que se efectuaron tantas evaluaciones como señales existentes en todas las clases. El corpus utilizado en LS consistió de 7 LS Normales de BDITM, así como 7 crepitantes y 7 sibilantes de RALE.

Por observaciones experimentales, se concluye que la duración en fases de las señales del repositorio de datos RALE y de BDITM está en el orden de 1.5 segundos para la inspiración y 2.5 para la espiración. Para HS, la fase S1 está alrededor de 0.1 a 0.12 Segundos; la fase S2 se encuentra entre 0.8 a 0.14 Segundos [27]. Considerando estos tiempos, los vectores Cuartílicos de 400 ms y corrimiento de 300 ms están en el rango de estacionariedad de las fases de la respiración. Por su parte, los vectores MFCC de 30 ms y corrimientos de 20 ms están en el rango de estacionariedad de S1 y S2.

RESULTADOS

La Tabla 1 muestra los experimentos con el mejor valor de BIC con señales LS Normal, Crepitancias y sibilancias, donde el número de densidades corresponde a 10, 3 y 4 respectivamente. Para señales HS, los resultados se muestran en la Tabla 2, para el mejor valor de BIC. Para ambas tablas se aplicaron vectores Cuartílicos con tramas de 400 ms, corrimientos de 300 ms y una covarianza tipo completa.

En la evaluación de la eficiencia de HMM con señales LS y HS (mostrados en la Tabla 3 y Tabla 4), los vectores Octílicos y Cuartílicos

Tabla 1: Número de cúmulos y respectivo tipo de covarianza que arrojan los mejores valores de BIC por clase LS, aplicando cuartiles

Señal	Tipo de Covarianza	Número de Cúmulos
Normal	Completa	10
Crepitancia	Completa	3
Sibilancia	Completa	4

Tabla 2: Número de cúmulos y respectivo tipo de covarianza que arrojan los mejores valores de BIC por clase HS, aplicando cuartiles

Señal	Tipo de Covarianza	Número de Cúmulos
Normal	Completa	6
Estenosis	Completa	8
VSD	Completa	3

fueron calculados sobre tramas de 400 ms y corrimientos de 300 ms (*i.e.*, 100 ms de traslape). En lo que se respecta a vectores MFCC, éstos se calcularon sobre tramas de 30 ms con corrimientos de 15 ms y 12 coeficientes Cepstrales. En los experimentos no se efectuó reducción de dimensión.

Los valores de eficiencia mostrados a partir de la Tabla 3 están representados en términos de Probabilidad de Correcta Clasificación (PCC). Debido al limitado número de señales adventicias, se aplicó VC en las evaluaciones (tipo *leave one-out*) [34]. Esto consistió en 21 evaluaciones; es decir, 7 evaluaciones por cada clase. En cada evaluación se utiliza una señal de una de las tres clases para evaluar los tres modelos. A su vez, el modelo de clase se entrena con las señales restantes de su respectiva clase. Se efectuará una evaluación por cada una de las señales existentes en el corpus, (*i.e.* 21). Este proceso es el mismo para la evaluación de HS y LS.

En cada evaluación se considera un acierto si el clasificador identifica la señal correctamente con su clase, de otra manera, se toma como un error de clasificación. En total, se efectúan 21 evaluaciones con modelos HMM para señales LS, así como 21 evaluaciones para señales HS, siguiendo el mismo proceso para ambas.

La Tabla 4 muestra los mejores resultados en términos de eficiencia de PCC de los experimentos realizados con señales HS. En este caso, también se evaluó al clasificador HMM como en la Tabla 3, *i.e.* con VC. Sin embargo, las clases utilizadas en HS son:

Tabla 3: Mejores resultados con modelos HMM con 3 Gaussianas por estado, 3 estados y 3 iteraciones para todas las señales LS.

Tipo de covarianza	Tipo de Vector	Eficiencia
Completa	Cuartil	95.23 %
Diagonal	Cuartil	100 %
Esférica	Octil	95.238

Tabla 4: Mejores resultados con modelos HMM con 3 Gaussianas por estado, 3 estados y 3 iteraciones para todas las señales HS.

Tipo de covarianza	Tipo de Vector	Eficiencia
Completa	Octil	47.619 %
Diagonal	MFCC	76.190 %
Esférica	MFCC	100.00 %

Normales, estenosis y VSD. De la misma manera, se utilizaron 7 señales de cada clase, obteniendo una base de datos de 21 señales. Al final, se efectuaron 21 evaluaciones distintas, en donde los modelos HMM (tipo izquierda-derecha) fueron calculados con 3 estados ($N_s = 3$) y 3 Gaussianas ($k = 3$) por estado, para cada tipo de covarianza.

Para efectos de observar una tendencia en términos de eficiencia de clasificación variando el número de Gaussianas por estado y el número de estados, se efectuaron numerosos experimentos, pero aquí sólo se muestran configuraciones de 2 a 4 Gaussianas por estado y de dos a cuatro estados por modelo HMM. La Tabla 5 muestra estos experimentos para señales LS aplicando covarianza diagonal con las 21 señales de las 3 clases, resaltando el mejor resultado con 3 Gaussianas por estado y 3 estados. La Tabla 6 muestra un experimento similar, pero para señales HS, obteniendo también uno de los mejores resultados con la configuración de 3 Gaussianas por estado y 3 estados para el modelo HMM.

Tabla 5: Eficiencia en términos de PCC con distintos números de Gaussianas por estados y cantidad de estados, aplicando vectores Cuartílicos (400, 300) para señales LS.

ks	Ns		
	2	3	4
2	85.7143	80.9524	90.4762
3	95.2381	100%	100%
4	95.2381	90.4762	95.2381

Tabla 6: Eficiencia en términos de PCC con distintos números de Gaussianas por estados y cantidad de estados, aplicando vectores MFCC (30, 15) para señales HS.

ks	Ns		
	2	3	4
2	95.2381	100	100
3	95.2381	100	100
4	100	95.2381	100

A partir de los resultados de la Tabla 5 y la Tabla 6, se decidió adoptar la configuración 3-Gaussianas/estado y 3 estados por HMM para sucesivos experimentos.

Se puede observar en la Tabla 5, que se logró una eficiencia de hasta 100%, utilizando Cuartiles con una covarianza diagonal para señales LS. Por otro lado, en la Tabla 6 se logra hasta un 100% utilizando vectores MFCC con varias configuraciones.

La covarianza es importante para modelar la forma, volumen y orientación de los cúmulos de las clases. La covarianza completa es teóricamente más versátil, pero al ser completa, implica más parámetros que calcular y se requieren más datos en la etapa de entrenamiento. Sin embargo, cuando no se cuenta con estas condiciones, es más difícil lograr la convergencia de los modelos.

Las imágenes de la Figura 5 a 7 intentan visualizar la cantidad de cúmulos existentes por clase, aunque sólo se muestran los experimentos para la clase LS Normal.

Posteriormente, se efectúan experimentos con distintas configuraciones de número de Gaussianas por estado y cantidad de estados en un modelo HMM, observando resultados alentadores con HMM de 3 Gaussianas por estado y 3 estados.

De manera complementaria, se pretendió visualizar la cantidad de cúmulos presentes en cada clase, por lo que se realizaron gráficas 3-dimensionales utilizando vectores Cuartílicos. De la Figura 5 a la Figura 10 se muestran los cúmulos de las clases LS Normal, sibilante y crepitante (respectivamente), donde cada clase está construida por todas las señales de ésta. Cada clase cuenta con dos representaciones: la primera es el 1er y 2do cuartil con probabilidad en el eje z; mientras que la segunda es el 2do y 3er cuartil. De la Figura 11 a la Figura 16 se muestran los cúmulos de las clases HS Normal, regurgitación y estenosis. Al igual que con LS, las gráficas se presentan con dos vistas de la misma señal.

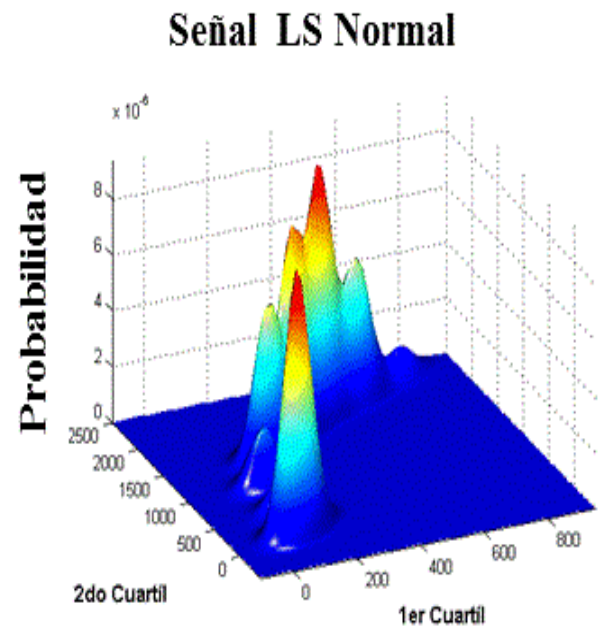


Figura 5. Visualización de cúmulos en la clase LS Normal con 1er y 2do cuartil.

Señal LS Normal

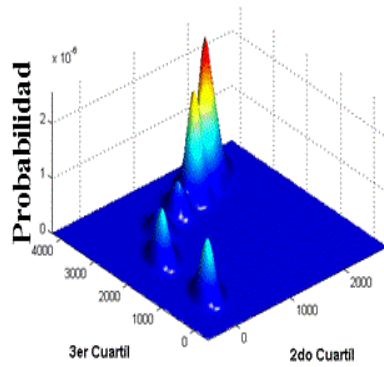


Figura 6. Visualización de cúmulos en la clase LS Normal con 2do y 3er cuartil.

Señal LS Sibilante

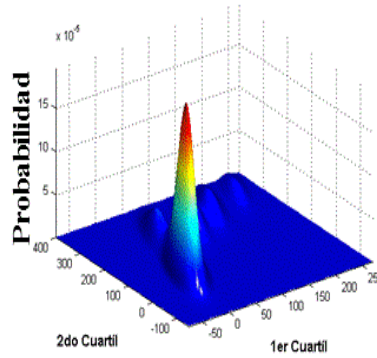


Figura 7. Visualización de cúmulos en la clase LS Sibilante con 1er y 2do cuartil.

Señal LS Sibilante

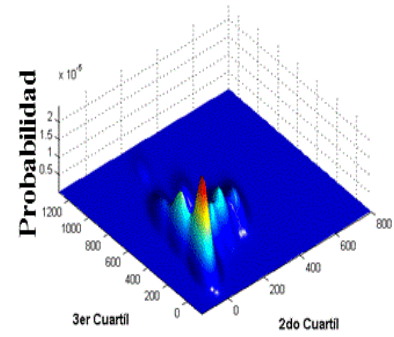


Figura 8. Visualización de cúmulos en la clase LS Sibilante con 2do y 3er cuartil.

Señal LS Crepitante

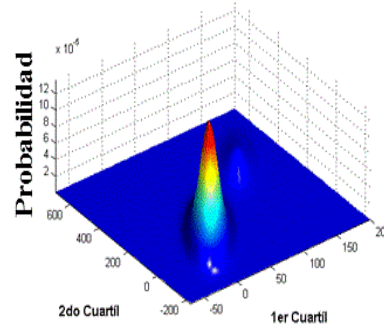


Figura 9. Visualización de cúmulos en la clase LS Crepitante con 1er y 2do cuartil.

Señal LS Crepitante

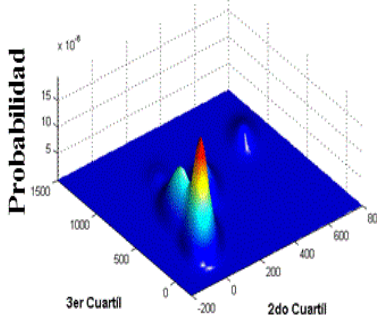


Figura 10. Visualización de cúmulos en la clase LS Crepitante con 2do y 3er cuartil.

Señal HS Normal

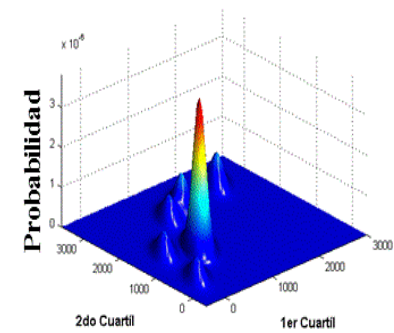


Figura 11. Visualización de cúmulos en la clase HS Normal con 1er y 2do cuartil.

Señal HS Normal

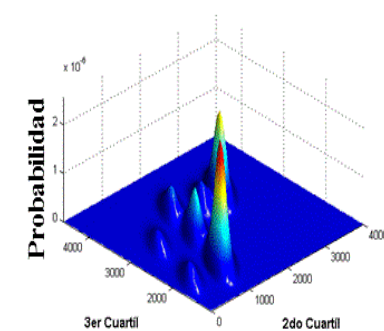


Figura 12. Visualización de cúmulos en la clase HS Normal con 2do y 3er cuartil.

Señal HS Regurgitaciones

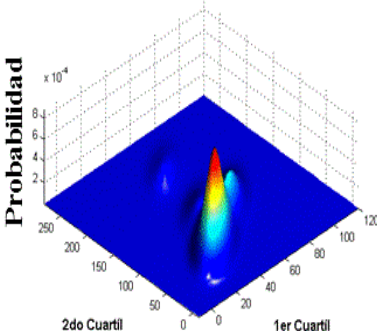


Figura 13. Visualización de cúmulos en la clase HS Regurgitaciones con 1er y 2do cuartil.

Señal HS Regurgitaciones

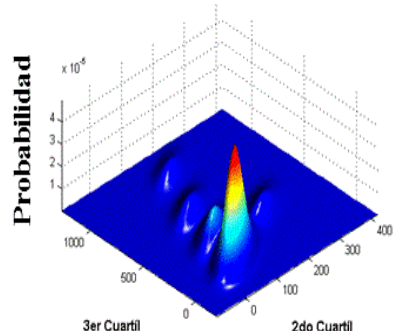


Figura 14. Visualización de cúmulos en la clase HS Regurgitaciones con 2do y 3er cuartil.

Señal HS: Estenosis Mitral

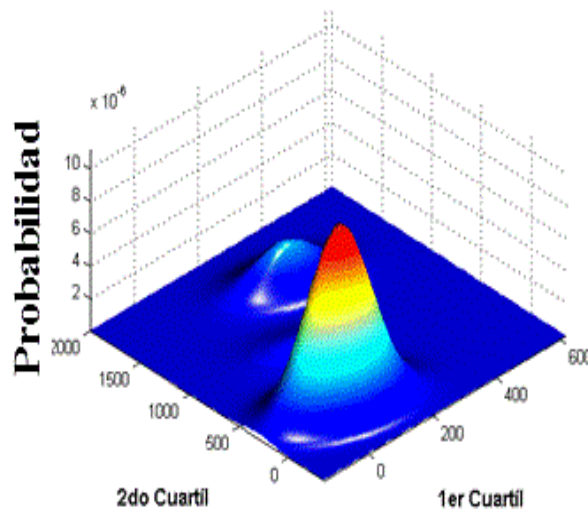


Figura 15. Visualización de cúmulos en la clase HS Estenosis Mitral con 1er y 2do cuartil.

Señal HS: Estenosis Mitral

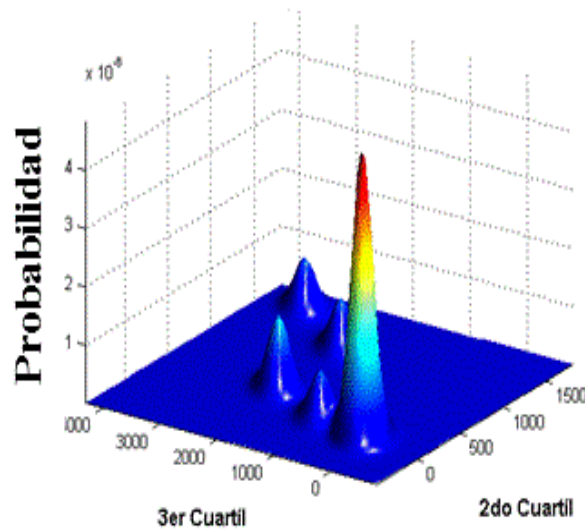


Figura 16. Visualización de cúmulos en la clase HS Estenosis Mitral con 2do y 3er cuartil.

Concerniente a señales **LS Normales** se observa una mayor cantidad de cúmulos con el 1er y 2do cuartil. Esto hace pensar que el mayor contenido de frecuencias y energía se encuentra en una región frecuencial correspondiente a los primeros dos cuartiles.

Tabla 7: Resumen de cúmulos observados por clase en LS y par de coeficientes Cuartílicos.

Señal	Coeficientes	Cúmulos Resultantes
Normal	1,2	7
Normal	2,3	6
Sibilante	1,2	5
Sibilante	2,3	8
Crepitante	1,2	2
Crepitante	2,3	5

Tabla 8: Resumen de cúmulos observados por clase en HS y par de coeficientes Cuartiles.

Señal	Coeficientes	Cúmulos Resultantes
Normal	1,2	6
Normal	2,3	7
Regurgitaciones	1,2	4
Regurgitaciones	2,3	5
Estenosis	1,2	3
Estenosis	2,3	5

En señales LS sibilantes, se visualizan más cúmulos definidos aplicando el 2do y 3er cuartil, lo que implica que hay más contribución en frecuencias más altas que en la señal Normal. Un comentario similar se puede extender al caso de las crepitaciones, pero aquí hay plegamiento de Gaussianas con la aparición de contribuciones intermedias. La Tabla 7 resume la cantidad de cúmulos observados por cada clase y en términos de los pares de 1ro-2do y 2do-3er cuartil.

En señales HS aplicando cuartiles, la clase LS Normal exhibe más cúmulos con la combinación del 2do y 3er cuartil, lo que implica más variaciones en la parte superior de su espectro. En la clase regurgitación presenta más aglomeraciones con el 2do y 3er cuartil. Además, las aglomeraciones están bien definidas aunque es una la que destaca. La Estenosis presenta más cúmulos bien definidos en el 2er y 3er cuartil. El resumen de estas observaciones se muestra en la Tabla 8.

Las figuras de cúmulos para señales LS y HS con vectores MFCC no se presentan en este trabajo, pero los resultados se muestran en la Tabla 9 y Tabla 10.

En señales LS con MFCC, la cantidad de cúmulos observados no cambió entre la visualización de los pares 1er-2do coeficientes y 2do-3ero. Esto está relacionado con bandas de frecuencia del orden de 50 hasta 350 Hz, donde no existe mucha variabilidad entre LS Normales y adventicias. Con respecto a la Tabla 10, se trabajó con señales HS Normales y adventicias aplicando MFCC: se hizo un análisis utilizando del 1er hasta el 5to coeficiente MFCC (barriendo desde 50 hasta 500 Hz), que es la mayor parte del espectro donde se encuentran las señales HS (incluyendo los casos patológicos). Debido a la naturaleza de los vectores MFCC, se pueden observar menos cúmulos por par

Tabla 9: Resumen de cúmulos observados por clase en LS y par de coeficientes MFCC.

Señal	Coeficientes	Cúmulos Resultantes
Normal	1,2	2
Normal	2,3	2
Crepitante	1,2	4
Crepitante	2,3	3
Sibilancia	1,2	5
Sibilancia	2,3	5

Tabla 10: Resumen de cúmulos observados por clase en HS y par de coeficientes MFCC.

Señal	Coeficientes	Cúmulos Resultantes
Normal	1,2	5
Normal	2,3	4
Normal	4,5	3
Regurgitación	1,2	2
Regurgitación	2,3	2
Regurgitación	4,5	2
Estenosis	1,2	9
Estenosis	2,3	7
Estenosis	4,5	6

Tabla 11: Tabla de contingencia aplicando HMM de 3 estados y 3 Gaussianas por estado, para señales Normales HS-LS (p) vs HS-LS con patología (n), con cuartiles.

Hipótesis	Entrada	Sensibilidad	Especificidad
	p	n	
P	28	0	1
N	0	28	1

de coeficientes, ya que MFCC está compuesto por más coeficientes que los cuartiles [15, 26].

En el caso de señales HS, requerimos menos coeficientes MFCC que para señales LS debido al rango de frecuencias que forma su espectro.

Otra manera de representar la eficiencia de clasificación, en la práctica clínica [39], es en términos de tablas de contingencia. La Tabla 11 es una tabla de contingencia donde la señal de entrada es denotada por p y n , mientras la hipótesis del sistema está dada por P y N .

En estas evaluaciones, se trabajó con 14 señales LS Normales, así como 14 HS Normales. En cuanto a patologías, se contó con 7 sibilantes, 7 crepitantes, 7 regurgitaciones y 7 estenosis. En la tabla puede observarse un excelente compromiso del sistema para detectar correctamente a los verdaderos positivos (Normales) de los verdaderos negativos (Patologías).

Aparentemente, la configuración de 3 densidades por evento (o estado) y 3 estados por modelo HMM-GMM, resultó suficiente para modelar la mayoría de las clases.

DISCUSIÓN

El análisis con BIC, siluetas y dendrogramas fue el punto de partida para construir un criterio antes de proponer una arquitectura final de reconocimiento HMM-GMM. En este trabajo se utilizó más la covarianza diagonal. En teoría, la covarianza completa es más versátil; sin embargo se requeriría contar con un número mayor de señales por clase, lo cual

permitirá determinar el comportamiento de las distintas covarianzas.

A fin de formar un criterio preliminar para dimensionar el modelo, se hicieron pruebas visuales (siluetas, dendrograma y BIC). Desafortunadamente, no se contó con una gran cantidad de señales, lo que implica que con más datos la configuración de la arquitectura probablemente tenga que reconfigurarse. Sin embargo, los resultados sí dejan ver el potencial de los modelos HMM-GMM.

El grafico de la Figura 3 muestra en el eje x la cantidad de cúmulos existentes en la clase y el valor de BIC en el eje y. BIC evalúa la probabilidad *a-posteriori* de la pertenencia de los datos con un modelo, pero a su vez penaliza la complejidad del modelo, como se observa en la Ecuación (4). Esto se evidencia con un punto de inflexión, el cual no se obtuvo en las curvas de BIC (Figura 3). Probablemente, para nuestros fines se requeriría una cantidad mayor de datos para obtener un punto de inflexión en las curvas BIC.

De la Figura 5 a Figura 16, se muestra la cantidad de cúmulos por clase en términos de pares de coeficientes. Si bien las cantidades no corresponden a las de los modelos experimentados, si está en el rango, lo que evita considerar arquitecturas HMM-GMM muy alejadas de las calculadas.

Los sonidos LS son parecidos a los fonemas /s/ o /f/, mientras que los HS están más relacionados con fonemas con frecuencias más definidas. En la literatura los sonidos HS se referencian como 'lub-dub'. MFCC fue concebido a partir del paradigma del funcionamiento del oído humano, el cual está relacionado con bandas a las cuales es sensible la audición humana [15, 40, 41]. Por tal motivo, los vectores MFCC podrían ser más aptos para el reconocimiento de HS.

Los modelos HMM-GMM son muy eficientes en señales acústicas, habiendo tenido un amplio éxito en señales del habla o reconocimiento de voz [17, 19, 41]. Además,

las señales utilizadas por clase son de buena calidad, salvo las de BDITM que fueron colectadas en condiciones normales de un consultorio de atención médica primaria. De hecho, considerando la poca variedad de clases, facilita la tarea de los modelos HMM-GMM en clasificación, y dados los resultados de la Tabla 11, se observa que tienen gran potencial para este tipo de patrones.

CONCLUSIÓN

Se efectuó un análisis de cúmulos y tamaño de modelo con siluetas, dendrogramas y BIC. Si bien no fue concluyente, resultó de utilidad para proponer una arquitectura HMM-GMM para clasificar señales LS y HS. En dicha arquitectura se aplicaron vectores Octílicos, Cuartílicos y MFCC, logrando hasta el 100 % de eficiencia de clasificación en LS y HS. El tipo de covarianza utilizada impactó el cálculo de un modelo GMM óptimo, que a su vez representó los estados en el modelo HMM. En términos de eficiencia, los mejores casos arrojaron 3 densidades por modelo GMM, en modelos de 3 estados, como una arquitectura genérica para cada una de las clases. Invariablemente, HS y LS presentan diferentes constituciones en términos de cúmulos entre sus clases Normales y Anormales.

Los modelos HMM-GMM mostraron su potencial para clasificación tanto para LS como HS; sin embargo, para validar una arquitectura genérica sería recomendable contar con un corpus extenso de registros normales y anormales para reconfigurar una arquitectura definitiva.

En un futuro sería interesante extender la clasificación a enfermedades definidas y a sectores específicos de la población local, ya que algunos sectores están más expuestos a contaminación y al clima que otros. Adicionalmente, sería conveniente incluir otro tipo de factores, tales como medidas de flujo de la respiración, con el fin de valorar su impacto en el diagnóstico de enfermedades específicas.

REFERENCIAS

1. J. Earis, "Lung sounds," *Thorax*, vol. 47, pp. 671-2, Sep 1992.
2. P. Forgacs, "Lung sounds," *Br J Dis Chest*, vol. 63, pp. 1-12, Jan 1969.
3. H. S. Hira, "Lung sounds," *J Assoc Physicians India*, vol. 41, pp. 33-7, Jan 1993.
4. R. Loudon and R. L. Murphy, Jr., "Lung sounds," *Am Rev Respir Dis*, vol. 130, pp. 663-73, Oct 1984.
5. M. Mori, "[Origin of normal breath sounds and abnormal lung sounds (crackles and wheezes)]," *Kokyu To Junkan*, vol. 31, pp. 493-501, May 1983.
6. P. Roudebush, "Lung sounds," *J Am Vet Med Assoc*, vol. 181, pp. 122-6, Jul 15 1982.
7. I. Sen, M. Saraclar, and Y. P. Kahya, "A Comparison of SVM and GMM-Based Classifier Configurations for Diagnostic Classification of Pulmonary Sounds," *Biomedical Engineering, IEEE Transactions on*, vol. 62, pp. 1768-1776, 2015.
8. K. Kosasih, U. R. Abeyratne, V. Swarnkar, and R. Triasih, "Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis," *Biomedical Engineering, IEEE Transactions on*, vol. 62, pp. 1185-1194, 2015.
9. J. Herzig, A. Bickel, A. Eitan, and N. Intrator, "Monitoring Cardiac Stress Using Features Extracted From S1 Heart Sounds," *Biomedical Engineering, IEEE Transactions on*, vol. 62, pp. 1169-1178, 2015.
10. 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," *Biomedical Engineering, IEEE Transactions on*, vol. 62, pp. 2279-2288, 2015.
11. S. Barma, C. Bo-Wei, J. Wen, J. Feng, and W. Jhing-Fa, "Measurement of Duration, Energy of Instantaneous Frequencies, and Splits of Subcomponents of the Second Heart Sound," *Instrumentation and Measurement, IEEE Transactions on*, vol. 64, pp. 1958-1967, 2015.
12. S. Barma, C. Bo-Wei, M. Ka Lok, and W. Jhing-Fa, "Quantitative Measurement of Split of the Second Heart Sound (S2)," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 12, pp. 851-860, 2015.
13. C. D. Papadaniil and L. J. Hadjileontiadis, "Efficient Heart Sound Segmentation and Extraction Using Ensemble Empirical Mode Decomposition and Kurtosis Features," *Biomedical and Health Informatics, IEEE Journal of*, vol. 18, pp. 1138-1152, 2014.
14. D. Ferreira da Ponte, C. A. Faria da Rocha, D. C. Hizume, and R. Moraes, "Equalization of Crackle Sounds to Compensate Thorax Attenuation," *Instrumentation and Measurement, IEEE Transactions on*, vol. 63, pp. 1983-1990, 2014.
15. A. Abushakra and M. Faezipour, "Acoustic Signal Classification of Breathing Movements to Virtually Aid Breath Regulation," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, pp. 493-500, 2013.
16. F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing Heart Sounds in Respiratory Signals Using Singular

- Spectrum Analysis," *Biomedical Engineering, IEEE Transactions on*, vol. 58, pp. 3360-3367, 2011.
17. L. Besacier, A. Ariyaeinia, J. S. Mason, J. F. Bonastre, P. Mayorga, C. Fredouille, et al., "Voice Biometrics over the Internet in the Framework of COST Action 275," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, pp. 466-479, Apr 2004 2004.
 18. P. Mayorga, L. Besacier, R. Lamy, and J.-F. Serignat, "Audio packet loss over IP and speech recognition," *Automatic Speech Recognition and Understanding, 2003. ASRU '03. 2003 IEEE Workshop on*, 2003, pp. 607-612.
 19. L. Besacier, P. Mayorga, J.-F. Bonastre, C. Fredouille, and S. Meignier, "Overview of compression and packet loss effects in speech biometrics," *Vision, Image and Signal Processing, IEE Proceedings*, vol. 150, pp. 372-376, 2003.
 20. D. M. Istrate, "Detection et Reconnaissance des Sons pour la Surveillance Médicale, These Doctorale, France," PhD, ENSERG, INPG, Grenoble, France 2003.
 21. P. Mayorga, C. Druzgalski, and J. Vidales, "Quantitative models for assessment of respiratory diseases," in *Health Care Exchange (PAHCE), 2010 Pan American*, pp. 25-30, 2010.
 22. B. Milner and S. Semnani, "Robust Speech Recognition over IP Networks," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP2000*, Istanbul, Turkey, Jun 2000.
 23. 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*, pp. 131-134, 2001.
 24. J. S. Yoon, G. H. Lee, and H. K. Kim, "A MFCC-based CELP speech coder for server-based speech recognition in network environments," *IEEE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, vol. E90a, pp. 626-632, Mar 2007.
 25. P. Mayorga, M. Olguín, O. H. González, N. Flores, and V. Luis, "Quantile Acoustic Vectors vs. MFCC Applied to Speaker Verification," in *International Journal of Advanced Robotic Systems*, 2013.
 26. P. Mayorga, C. Druzgalski, O. H. Gonzalez, and H. Lopez, "Modified classification of normal Lung Sounds applying Quantile Vectors," presented at the *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, San Diego, California USA,, 2012.
 27. A. F. L. Julian David Echeverry, Juan Fernando López. (2007, 28 / octubre / 2015) "Reconocimiento de valvulopatías cardíacas en señales de fonocardiografía empleando la transformada Gabor," *Scientia Et Technica*. 139-144. Available: <http://www.redalyc.org/articulo.oa?id=84934024>
 28. K. Benabdeslem and Y. Bennani, "Dendogram based SVM for multi-class classification.," in *Journal of Computing and Information Technology-CIT 14*, 2006.
 29. Xue Mei Lu; Sung Jong Eun; Taeg Keun Whangbo, "Vector Silhouette Extraction for Generating Blueprint.," in *Automation and Logistics, 2007 IEEE International Conference on*, 18-21 Aug. 2007, pp. 2946-2951.

30. Xuejun Li; Jianguang Sun; Changgui Yang, "Extracting silhouette curves of NURBS surfaces by tracing silhouette points," in *Tsinghua Science and Technology*, June 1998, pp. 1005-1008.
31. J. Zhao, "Efficient Model Selection for Mixtures of Probabilistic PCA Via Hierarchical BIC," *IEEE Transactions on*, vol. vol.44, pp. 1871-1883, Oct. 2014.
32. Pearce D., "An Overview of ETSI Standards Activities for Distributed Speech Recognition Front-Ends," in *AVIOS 2000: The Speech Applications Conference*, San Jose, CA, USA, May 22-24, 2000.
33. Reynolds D. A., "Gaussian Mixture Modeling Approach to Text-Independent speaker Identification," Thesis from Georgia Institute of Technology, Georgia Institute of Technology, Georgia, 1992.
34. Webb Andrew R., *Statistical Pattern Recognition*, John Wiley & Sons Ltd., 2002.
35. L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, N.J.: PTR Prentice Hall, 1993.
36. P. Mayorga Ortiz, C. Druzgalski, J. E. Miranda Vega, and D. O. Calderas Ochoa. (2014) "Modelos Acústicos HMM Multimodales para Sonidos Cardiacos y Pulmonares," *Revista Mexicana de Ingeniería Biomédica*. 197-210.
37. B. C. s. Hospital. Boston Children's Hospital <http://www.childrenshospital.org/>. Available: <http://www.childrenshospital.org/>
38. T. H. Institute. Texas Heart Institute <http://www.texasheartinstitute.org/AboutUs/index.cfm>.
39. J. Cerda L and L. Cifuentes A, "Uso de tests diagnósticos en la práctica clínica (Parte 1): Análisis de las propiedades de un test diagnóstico," *Revista Chilena de Infectología*, vol. 27, pp. 205-208, 2010.
40. L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, N.J.: Prentice-Hall, 1975.
41. L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, N.J.: PTR Prentice Hall, 1993.

