



Modelos Acústicos HMM Multimodales para Sonidos Cardiacos y Pulmonares

P. Mayorga Ortiz*
C. Drugalski **
J.E. Miranda Vega*
D.O. Calderas Ochoa*

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

** Elec. Eng. Dept.,
California State. University,
1250 Bellflower Blvd, Long
Beach, CA 90840, USA.

RESUMEN

Este artículo muestra el proceso de clasificación de señales bioacústicas normales y anormales registradas sobre el tórax humano lo cual incluye los sonidos de corazón y del pulmón. La idea específica es diseñar un sistema de clasificación de señales basado en técnicas de modelado acústico empleando particularmente modelos HMM para detectar secuencias de eventos, y GMM para modelar cúmulos que corresponden a los datos de los eventos. Las modalidades para extraer las características de los datos son vectores MFCC y Octiles. Esta aproximación tiene el potencial de mejorar la clasificación de la precisión en indicadores de diagnóstico auscultatorios, esto es interesante ya que los modelos HMM han demostrado ser menos sensibles al ruido en estudios previos. Resultados preliminares demuestran una precisión del 95% en clasificación de las señales de sonido evaluadas. Esto es particularmente crítico tomando en cuenta la interferencia ambiental en una variedad de consultorios médicos. Debido a que algunas frecuencias del sonido cardiaco son paralelas a los sonidos pulmonares, estas pueden ser modeladas a partir de un mismo registro. Resultados experimentales preliminares de esta aproximación demuestran que es factible el desarrollo de valoraciones de diagnóstico automatizado de pacientes mediante identificadores de diagnóstico auscultatorios en forma temprana usando tecnologías de bajo costo.

Palabras clave: Modelos Ocultos de Markov (HMM), Modelos Mezclados Gaussianos (GMM), Vectores MFCC, Vectores Octílicos, Sonidos Cardiacos y Pulmonares.

Correspondencia:
 Jesus Elias Miranda Vega
 Correo electrónico:
 Ing.jesus_elias@hotmail.com

Fecha de recepción:
1 de julio de 2014
Fecha de aceptación:
10 de octubre de 2014

ABSTRACT

This paper demonstrates classification processes of normal and abnormal bioacoustics signals recorded over a human thorax which encompasses heart and lung sounds. The specific aim is to design a signal classification system based on acoustical modeling techniques employing particularly HMM models to detect events' sequences, and GMM to model clusters corresponding to the data events. The modalities for extracting data characteristic are the MFCC and Octile vectors. These approaches have a potential of enhancing the classification accuracy of these auscultatory diagnostic indicators as the initial studies demonstrated that the HMM based models are less sensitive to the noise. Preliminary results demonstrate over 95% accuracy in classification of the evaluated sound signals. This is particularly critical taking into account environmental interference in a variety of medical care settings. As the heart sounds frequency components parallel those of the lungs sounds, but with a different periodicity, they can be modeled with the same recording. The preliminary experimental results are supportive of this approach and demonstrate feasibility of a development of an automated early diagnostic assessment of patients' auscultatory diagnostic indicators utilizing low cost technologies.

Keywords: Linear Markov Models (HMM), Gaussian Mixture Models (GMM), Octile vectors, MFCC vectors, lung sounds and heart sounds.

INTRODUCCIÓN

La bioacústica es una ciencia relacionada con diferentes disciplinas científicas, la cual investiga la producción y percepción de sonidos biológicos, incluidos los que produce y procesa el ser humano [2]. En este contexto, las características de los sonidos del pulmón (LS) y del corazón (HS) están relacionados con la salud de quien los produce.

Por otro lado, las condiciones ambientales y la contaminación en el aire se consideran factores críticos que contribuyen al incremento de enfermedades respiratorias y cardiovasculares [3, 4]. Particularmente, los expertos en asma consideran que los sonidos adventicios no son un elemento suficiente para el diagnóstico de asma, pero sí muy valioso para este propósito. Además, dependiendo del sonido podría inferirse la existencia de un problema en las vías respiratorias; en el caso del corazón, los sonidos

pueden evidenciar algunos padecimientos como Estenosis o Defecto Septal Ventricular (VSD).

Invariablemente, los sonidos y su auscultación mediante el estetoscopio u otros dispositivos han constituido una práctica estándar en medicina, de hecho este instrumento es indispensable en todo consultorio médico por modesto que este lo sea. El uso del estetoscopio tradicional acústico mecánico y consecuentemente la eficiencia del mismo, se circunscribe a las capacidades periciales y sensoriales del médico que ausculta, o en otros casos al de una enfermera ayudante del mismo.

Estas limitaciones están también relacionadas al uso de descripciones cualitativas de los sonidos característicos, los cuales frecuentemente se usan de manera muy subjetiva, especialmente cuando estos son revisados por diferente personal médico, y en periodos muy largos de tiempo. Inclusive, en

ocasiones es necesario interrogar a los padres de un infante sobre los síntomas o sonidos producidos por este, cuando se sospecha de alguna patología cardiovascular o respiratoria; desafortunadamente, los padres normalmente no están familiarizados o no tienen entrenamiento para emitir juicios sobre síntomas.

Una alternativa a los problemas mencionados, es hacer uso de la tecnología digital inteligente, es decir, implementar metodologías con reconocimiento de patrones, internet y aquellas que permitan ser aplicadas en zonas remotas, o con limitantes como las antes mencionadas.

Dentro de las metodologías existentes se pueden citar Maquinas de Soporte Vectorial (SVM), Análisis Discriminante Lineal, Modelos Mezclados Gaussianos (GMM) y en el caso de este artículo, Modelos Ocultos de Markov (HMM). La diferencia más importante es la interpretación de los datos, HMM los procesa como secuencias de eventos relacionados con los estados, mientras que las otras técnicas los procesan solo como cúmulos relacionados como patrones sin considerar la secuencia de los eventos. En otras áreas del conocimiento [5], se observa que la modelación con HMM es más susceptible a la pérdida de información que al ruido, mientras que la modelación en cúmulos es más susceptible al ruido que a la pérdida de información. En este sentido esa ventaja es la que se puede aprovechar con los modelos HMM, aunque pudiera ser un poco más costoso computacionalmente.

Otra ventaja, es que los sistemas inteligentes pueden ser aplicados considerando una simple Laptop, y que puede ser una tecnología al alcance en la gran mayoría de escenarios socioeconómicos. En este sentido, se puede hacer notar que el nivel socioeconómico predominante del mundo es bajo, por ejemplo en América Latina el 70 % de la población está debajo de la media del índice de pobreza [6].

Esto es relevante, porque muchas de las aportaciones no necesariamente están orientadas a consultorios tradicionales del tercer mundo [7-9], ya que su costo podría superar la decena de miles de dólares; por lo tanto, es necesario insistir en un enfoque más económico. Adicionalmente,

los consultorios están ubicados en zonas o espacios con ruido ambiental, el cual interfiere con el registro de señales LS u HS y repercute en el diagnóstico.

Un sistema inteligente (ya sea para voz, pulmón u otra señal) simula el proceso de reconocimiento que el ser humano efectúa sobre la señal acústica o evento, acústico para nuestro caso [10].

La mayoría de las aproximaciones en sistemas inteligentes para clasificación de señales LS u HS se basan en cúmulos (clusterización), filtrado clásico y otras aproximaciones susceptibles al ruido o a la atenuación de la señal importante durante el proceso de filtrado. Los Modelos Ocultos de Markov han demostrado ser menos susceptibles al ruido, que aquellos basados en cúmulos en áreas como el procesamiento digital de voz [11, 12].

La misma señal LS normalmente comparte un rango de frecuencia considerable con la señal HS, esto hace pensar que los HMM podrían ser útiles en este campo, ya que están orientados a detectar secuencia de eventos. El presente trabajo está relacionado con la aplicación de modelos HMM a señales LS y HS, además se utilizan dos modalidades en el extractor de características, MFCC ya exitoso en voz y otro inspirado en clasificación de patologías a partir de medidas de función pulmonar [13].

En este trabajo se aplican los Modelos Ocultos de Markov, los cuales han demostrado ser muy poderosos para clasificación de secuencias de eventos [5, 14]. Los modelos HMM tienen a su vez requerimientos en vectores de características para entrenar los modelos que son una secuencia de estados, y los estados a su vez son modelados mediante Modelos Mezclados Gaussianos.

Vectores de Características MFCC y Octiles

Los vectores de características tienen la función de resaltar particularidades de una clase mediante valores, de tal manera que mejoran la clasificación, en este apartado se describen las dos aproximaciones utilizadas para los experimentos que se muestran más adelante, es decir, MFCC y Octiles.

Vectores MFCC

Los vectores MFCC son una metodología exitosa en caracterización acústica, fundamentada en la denominada deconvolución Cepstral. El Cepstrum es un método fundado en la transformada de Fourier que permite aislar las frecuencias fundamentales propias de una señal acústica.

Los vectores MFCC son una extensión de los principios cepstrales, y su transformación a un espacio frecuencial no lineal está relacionado con la audición humana [15-18].

En MFCC, los sonidos son parametrizados, haciendo un preénfasis con filtros FIR, seguido por una ventana Hamming aplicada a cada trama de análisis. En este trabajo, se experimentó con ventanas Hamming de 400 ms y corrimientos de 300 ms para LS y 30 ms con 15 ms de corrimiento para 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 o la de Bark. A esto le sigue una etapa de logaritmo de los valores obtenidos de cada filtro y posteriormente la transformada inversa de Fourier (para este caso siendo valores reales es igual que calcular la transformada discreta cosenoidal). El resultado final es un vector de características llamado MFCC [15, 19]. En este trabajo se aplicaron los vectores MFCC a señales acústicas del pulmón (LS) y a señales acústicas del corazón (HS).

Vectores Octílicos

Otro tipo de vectores acústicos son los Cuantiles, que se basan en la función de distribución acumulativa (CDF). El Cuantil 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 algún valor p , donde p se encuentra entre $0 < p < 1$.

Esto puede ser calculado para el caso de una función de densidad de probabilidad continua $f(x)$ a través de la Ec. 1.

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

Para propósitos de caracterización acústica, se desea encontrar un número dado de coeficientes Cuantílicos q_p , por lo que se parte de la transformada inversa de la CDF. En el caso de señales acústicas como LS y HS es importante realizar los cálculos en el límite de estacionariedad; la respiración humana se encuentra en el rango de los 400 ms.

El límite de estacionariedad está determinado por el tiempo de duración en que se generan los eventos; en el caso de la respiración está relacionado con la duración de la inspiración, la espiración y los silencios existentes entre ellas, los cuales fueron determinados experimentalmente [20].

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). Una idea similar para HS, con 3 segundos considerando inspiración, silencio y espiración [21]. En este proceso, la primera etapa consiste en la lectura de la señal, partiendo de archivos *.wav; posteriormente, se aplica la FFT. Cumpliendo con un principio básico para una función de distribución de probabilidad, la distribución espectral se normaliza (Ec. 2).

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

La Ec. 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 Octiles, calculados aquí mediante la Ec. 3, cuyos valores frecuenciales $f_{0.125}, \dots, f_{0.875}$ correspondientes a cada uno de los respectivos coeficientes Octílicos son los que nos interesan en este trabajo. El cálculo del último Octil no es importante ya que siempre es igual a 1, lo que nos deja un vector de 7 dimensiones.

$$A_{0.125} = \int_{-\infty}^{f_{.125}} F_N(f) df, \dots, A_{0.875} = \int_{-\infty}^{f_{.875}} F_N(f) df \quad (3)$$

Algorítmicamente, A_p se calcula mediante una suma iterativa para obtener el área y detectar los valores frecuenciales para los cuales el área obtenida es $A=0.125, \dots, A=0.875$. Si bien éste fue el caso de Octiles, el mismo principio puede aplicarse a Cuartiles u otro tipo de Cuantil. Para una descripción más extensa, se puede revisar el trabajo [21].

MODELADO APLICANDO GMM Y HMM

Los modelos GMM son exitosos en varias áreas del conocimiento, sobre todo en señales acústicas, en este trabajo se utilizan como una representación de los estados en modelos HMM.

Modelos GMM

Un modelo GMM es una tripleta compuesta por las medias, covarianzas y ponderaciones; generalmente, cada clase es representada por un modelo GMM Λ . Un modelo Λ se entrena a partir de vectores acústicos obtenidos a partir de los datos muestreados de las señales. El modelado GMM se sirve del algoritmo EM para calcular los modelos $\Lambda_i = \{m_i, \vec{\mu}_i, \Sigma_i\}$.

Este cálculo se efectúa sobre vectores acústicos extraídos de las distintas grabaciones (LS u HS).

La media $\vec{\mu}_i$ 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 [16, 22].

$$p(\vec{x}|\Lambda) = \sum_{i=1}^M m_i b_i(\vec{x}) \quad (4)$$

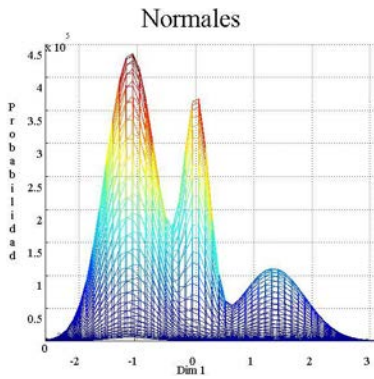
Donde $\vec{x}$ es un vector aleatorio D -dimensional (vectores acústicos MFCC u Octílicos), $b_i \forall i = 1, \dots, M$ son las densidades componentes y $m_i \forall i = 1, \dots, M$ son las ponderaciones de las mezclas. Cada densidad componente es una función Gaussiana D -dimensional [5, 23-25] como sigue:

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i) \right] \quad (5)$$

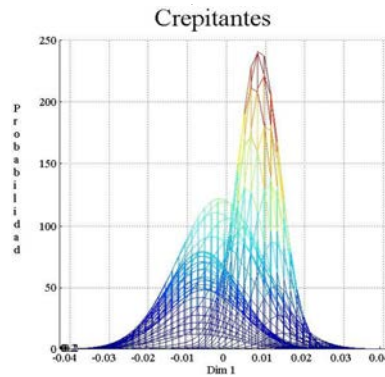
Aquí, $\vec{\mu}$ es un vector de medias y Σ_i es la matriz de covarianza. Por otro lado, las ponderaciones de las mezclas deben satisfacer la restricción $\sum_i^M m_i = 1$. El conjunto de modelos GMM son parametrizados por los vectores de medias, matrices covarianza y las ponderaciones de las mezclas para las densidades Gaussianas de cada estado dentro del modelo HMM.

El modelo que contiene los parámetros es representado por la siguiente expresión [5, 23-25]:

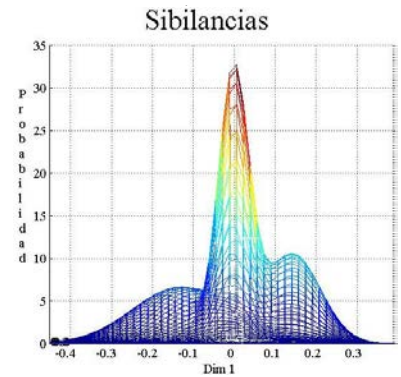
$$\Lambda = \{m_i, \vec{\mu}_i, \Sigma_i\} \forall i = 1, \dots, M \quad (6)$$



a)



b)



c)

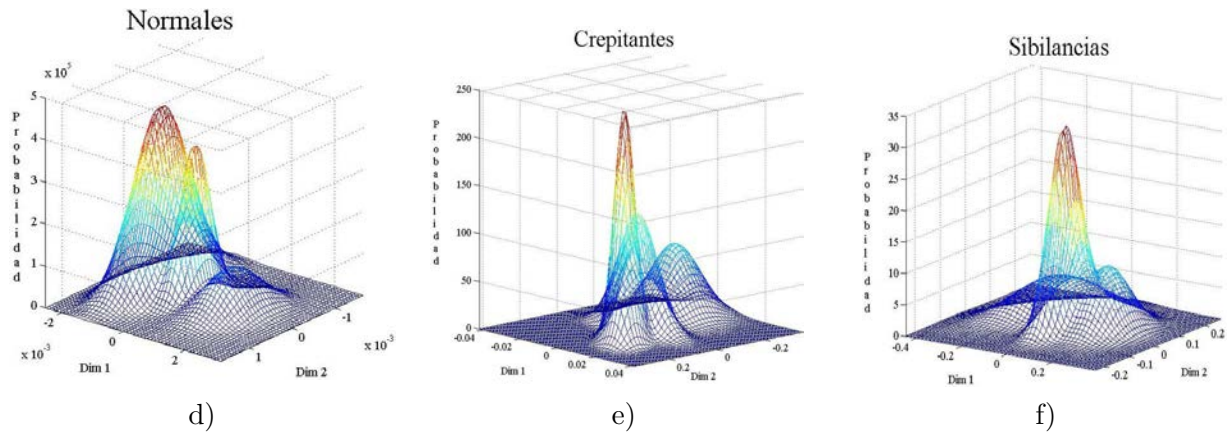


Fig. 1 Modelos GMM para Señales LS Normal (a y d), Crepitante (b y e) y Sibilante (c y f). Donde a), b) y c) son representaciones en 2 dimensiones, mientras que d), e) y f) corresponden a la representación de GMM de 3 dimensiones.

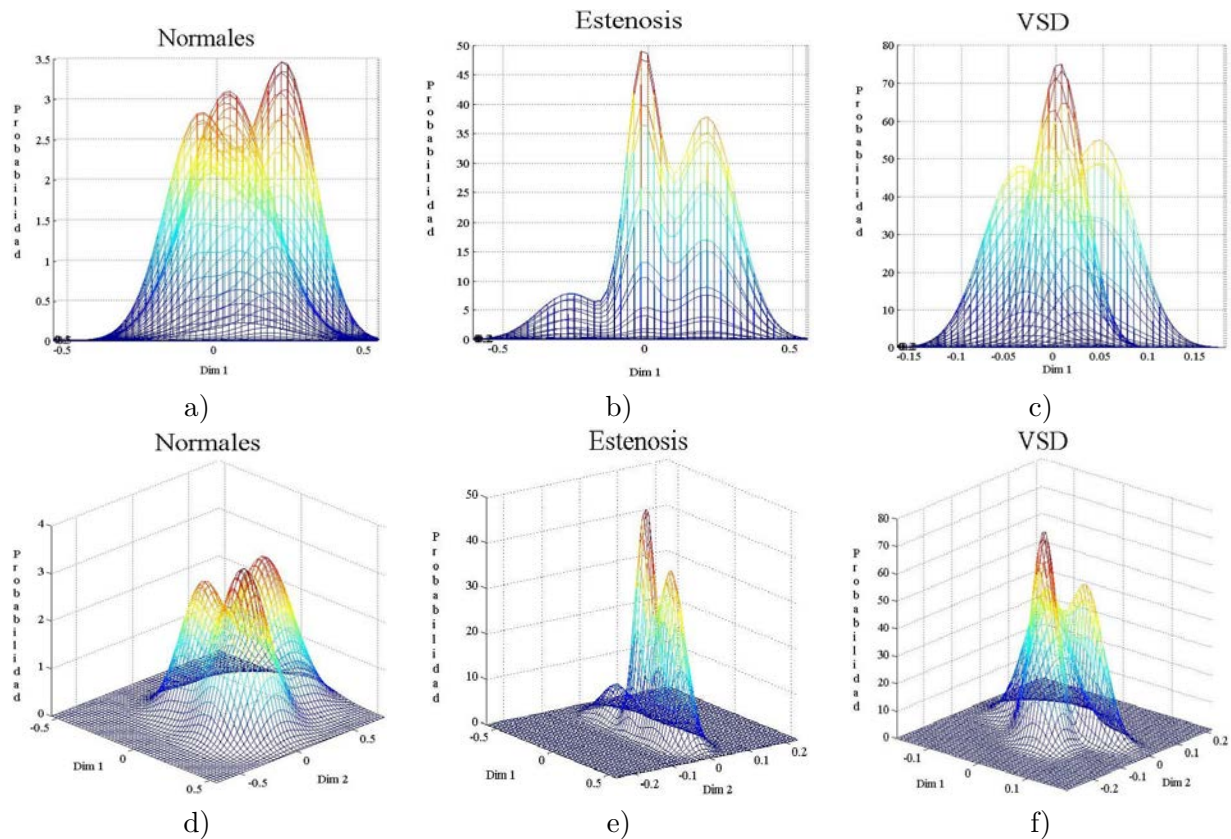


Fig. 2 Modelos GMM para Señales HS Normal (a y d), estenosis (b y e) y VSD (c y f). Donde a), b) y c) son representaciones en 2 dimensiones, mientras que d), e) y f) corresponden a la representación de GMM de 3 dimensiones.

Modelos HMM

El proceso del ciclo cardíaco es secuencial y cíclico, el cual consiste en dos eventos dominantes, conocidos como el primer ruido cardíaco “S1” y el segundo ruido cardíaco “S2”,

S1 se refiere al cierre de las válvulas mitral y tricúspide, mientras se genera S2 por la interrupción de las válvulas aórtica y pulmonar, como se observa en la Fig. 3. En el caso de los sonidos del pulmón (LS), también es un

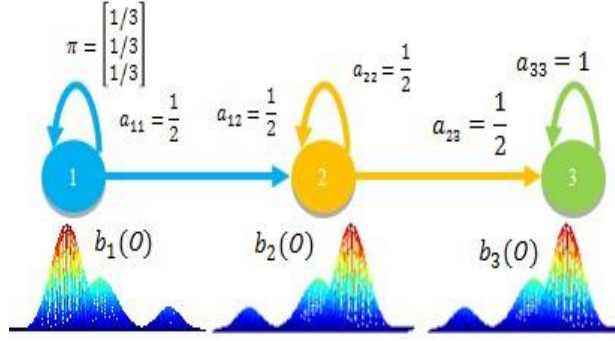


Fig. 3. Matriz de probabilidad de transición $A\{a_{ij}\}$ para inicializar los modelos HMM.

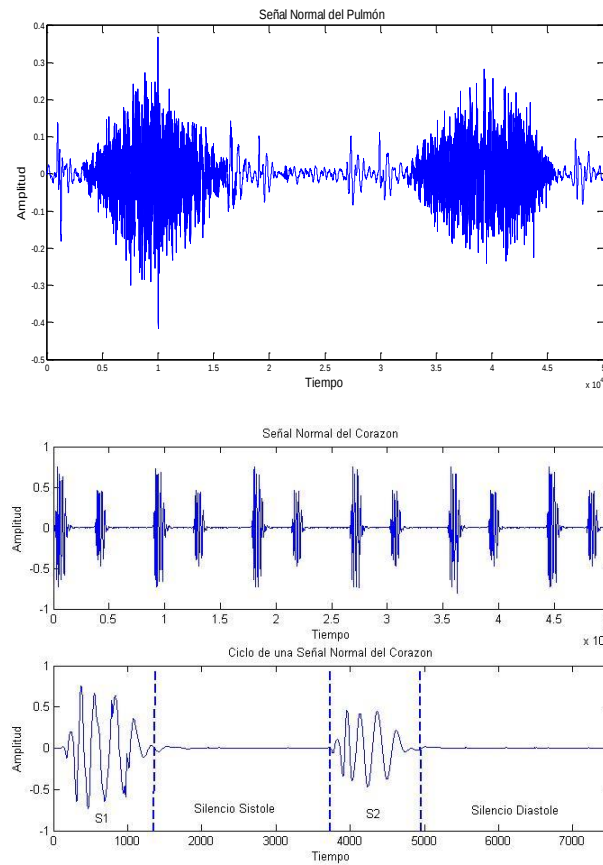


Fig. 4 Señal de LS (a), HS Normal y un Ciclo HS normal (b).

proceso cíclico que consta fundamentalmente de dos eventos, la inspiración y la espiración.

Tanto en el caso de señales HS como LS hay silencios entre los eventos principales y estos son distintos dependiendo del evento que le preceda y anteceda, por lo tanto en ambos casos son secuencias de eventos que pueden variar dependiendo de las circunstancias, de la salud y porque no del estado de ánimo.

En HS, S1 y S2 así como los silencios, podrían ser etapas de la señal a modelar, mientras que en LS, la inspiración, la espiración y los silencios serían las etapas de la señal a modelar (Fig. 4). Luego, estos son procesos que podrían ser modelados mediante Modelos Ocultos de Markov o HMM que es un autómata de estados finito. En este trabajo experimentamos con HS y LS por separado para poder proponer en un futuro una arquitectura HMM-GMM para HS-LS conjunta. A título ilustrativo, las Fig. 1 y Fig. 2 muestran los GMM's para cada clase de LS y HS que permiten visualizar las clases.

Un HMM puede ser visto como un modelo basado en estados que no son directamente observados, de hecho cada estado está caracterizado por un GMM, el cual modela las observaciones correspondientes a ese estado.

Formalmente, un HMM es definido con los siguientes componentes [26]:

- $S = \{S_1, S_2, \dots, S_N\}$ el conjunto finito de los posibles estados (ocultos);
- La matriz de transiciones $A = \{a_{i,j}, 1 \leq j \leq N\}$, 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 un estado ocupado en un tiempo t .

- La función de emisión en un estado j (en nuestro caso un GMM) $B = \{b(O|S_j)\}$, denota la probabilidad de emisión de una observación $O \in V$ cuando el sistema está en el estado S_j ; V corresponde a observaciones continuas, en el caso de los experimentos de este trabajo, se aplicaron vectores acústicos MFCC u Octiles, y por lo tanto $b(O|S_j)$ es un modelo GMM.

- $\pi = \{\pi_i\}$, 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 $\sum_{i=1}^N \pi_i = 1$.

Al igual que en el caso de los modelos GMM, es convencional expresar un modelo HMM como

una tripleta $\lambda = (A, B, \pi)$. Un ejemplo de modelo HMM para señales acústicas es como lo muestra la Fig. 3.

El entrenamiento o aprendizaje de los parámetros HMM, dado un conjunto de secuencias $\{O_i\}$, es normalmente realizado aplicando el algoritmo Baum-Welch [26], 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* [26].

BASE DE DATOS

El conjunto de señales HS utilizadas para los experimentos provienen de bases de datos las cuales están disponibles para propósitos académicos o científicos [27, 28]. Con lo anterior se logró un repositorio de grabaciones de HS, el cual consta de 204 HS con las especificaciones descritas en la Tabla I, 68 HS son Normales (NL), de esta base de datos solo se utilizaron las que tienen una duración de 9 a 12 segundos.

Las señales originales fueron capturadas a una tasa de 44 kHz y 22 kHz, pero se sub-muestrearon a 11 kHz. De las mismas fuentes se obtuvieron 68 HS etiquetadas como Estenosis (ES); igualmente, fueron sub-muestreadas a 11 kHz. Las señales fueron particionadas para entrenar modelos acústicos y para la etapa de evaluación. Una partición de 17 señales de cada clase se utilizó de prueba, y 51 por clase para entrenamiento. Las etapas de evaluación y modelado fueron llevadas a cabo con distintas configuraciones intercambiando señales de prueba y de entrenamiento, *i.e.*, aplicando validación cruzada.

Con respecto a los experimentos de señales LS se utilizaron dos bases de datos, RALE y BDITM. La primera base de datos es RALE y consiste en un conjunto de grabaciones .wav de sonidos LS adventicios y normales, la cual fue desarrollada por la universidad de Winnipeg, Canadá. Las señales de RALE 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. En este trabajo solo utilizamos las señales adventicias Crepitantes y Sibilantes (en inglés: Crackles y Wheeze). La otra base de datos utilizada fue BDITM, que se compone de grabaciones de respiración normal de estudiantes cuyo rango de edad se encuentra entre 18 y 25 años.

En este caso, la evaluación fue efectuada con validación cruzada, dejando una señal para evaluar, el resto para el cálculo del modelo y cambiando las configuraciones hasta agotar todas las posibilidades. Para los experimentos se tomaron 7 LS normales de BDITM, 7 crepitantes y 7 sibilantes de RALE.

RESULTADOS

Los vectores Octílicos fueron calculados sobre tramas de 400 ms y corrimientos de 300 ms (*i.e.* 100 ms de traslape) en señales LS. En lo que se respecta a vectores MFCC, estos se calcularon sobre tramas de 400 ms con corrimientos de 300 ms y 13 coeficientes Cepstrales (sin coeficientes Δ o $\Delta\Delta$'s) para el caso de señales LS; para señales HS las tramas fueron de 30 ms, corrimientos de 15 ms y 12 coeficientes Cepstrales (sin coeficientes Δ o $\Delta\Delta$'s). En los experimentos no se efectúa reducción de dimensionalidad, ya que en ensayos previos no nos proporcionó una mejoría sensible.

Tabla I Características del formato en las señales HS

Tipo de canal	Frecuencia de Muestreo	Duración por HS	Formato de Archivo/ Extensión
Monoaural	11025 Hz 16 Bits	9 - 12 Segundos	Waveform Audio Format/ .wav

Tabla II Modelos HMM para LS y vectores MFCC; modelos GMM con mejores resultados (GMM3, GMM4).

	Estados por HMM # NS	# de Coef. sin Δ 's	Iteraciones:	GMM3: %Recon.	GMM4: %Recon.
Diagonal	3	12	2	95.238	100
Esférica	3	12	2	95.238	85.714
Completa	3	12	2	85.714	57.142

Tabla III Modelos HMM para LS y vectores Octílicos; modelos GMM con mejores resultados (GMM3, GMM4).

	Estados por HMM # NS	# de Coef.	Iteraciones:	GMM3: %Recon.	GMM4: %Recon.
Diagonal	3	7	2	95.238	100
Esférica	3	7	2	95.238	95.238
Completa	3	7	2	38.095	33.333

Los valores obtenidos en las Tabla II y Tabla III, son valores considerando tres clases de señal: 7 normales, 7 crepitantes y 7 sibilancias. Las señales fueron particionadas en 4 conjuntos, 3 para el cálculo de los HMM de cada clase y 1 señal para evaluación, de las 3 clases alguno contaba solo con 6 señales para entrenar el modelo, mientras que los otros 2 contaban con 7. En este proceso la señal de prueba se va alternando hasta agotar todas las posibilidades.

Por lo tanto, se efectuaron 21 evaluaciones. En LS, después de muchos experimentos, los modelos HMM fueron calculados con dos tipos de configuraciones: modelos HMM con 3 estados ($Ns = 3$), 3 y 4 Gaussianas ($k = 3$) por estado. En ambas configuraciones la convergencia en el cálculo de los modelos se logró con dos iteraciones ($it = 2$).

Los modelos HMM fueron tipo izquierda-derecha (Bakis), y los estados fueron representados con modelos GMM como lo muestra la Fig. 3.

Aquí, el vector π denota las probabilidades iniciales (a priori) de estar en alguno de los estados q , y los valores a_{ij} son las probabilidades de transición entre estados, mientras que b_i (O) son las funciones de emisión de una observación (para este caso un GMM).

Se puede destacar que las probabilidades de transición y de estado inicial, fueron inicializadas aleatoriamente.

En la Tabla II se muestran los resultados de los experimentos realizados con vectores MFCC correspondientes a señales LS, obteniendo el mejor resultado con 4 Gaussianas por estado y covarianza diagonal, correspondiendo al 100% de eficiencia. En la Tabla III el mejor resultado es con una configuración similar.

La evaluación de Tabla III también fue efectuada con validación cruzada (VC), pero a diferencia de la Tabla II, se aplican Octiles [21, 29].

Aunque la Tabla II y la no muestran los porcentajes explícitos por clase de LS, los mejores resultados fueron del 100% para cada una de las tres clases, con vectores Octílicos y MFCC, aplicando covarianza diagonal y GMM con 4 Gaussianas por estado y 3 estados por HMM. Se puede observar que la matriz de covarianza de los modelos GMM influye en los resultados, resaltando que la covarianza diagonal tuvo más éxito.

La covarianza es importante para modelar la forma, volumen y orientación de los cúmulos de datos en señales; el tipo de covarianza completa, implica más cálculos, más iteraciones, más datos y no siempre se logra la convergencia en los modelos, por lo cual es necesario un análisis de los modelos antes de proponer una arquitectura final HMM - GMM.

Tabla IV Experimentos en HS con HMM, vectores MFCC y modelos GMM para representar los estados de HMM.

	Estados por HMM # NS	# de Coef. sin Δ 's	Iteraciones:	GMM3: %Recon.	GMM4: %Recon.
Normal	4	13	98.5	100	100
Estenosis	4	13	100	100	100
VSD	4	13	99.5	100	98

Tabla V Experimentos en HS con HMM, vectores Octílicos y modelos GMM para representar los estados de HMM.

	Estados por HMM # NS	# de Coef.	Iteraciones:	GMM3: %Recon.	GMM4: %Recon.
Normal	4	7	95.0	98.0	90
Estenosis	3	7	55.5	98.5	55.0
VSD	4	7	75.5	80.5	90.5

Sin embargo, para validar la observación anterior convendría contar con un número más abundante de señales por clase, ya que la covarianza completa es más versátil.

El cálculo de modelos HMM-GMM más complejos demanda más señales para una buena convergencia. Lo que se puede comentar de las Tabla II y Tabla III, es que la covarianza diagonal con Octiles fue ligeramente mejor.

Los resultados con HMM mostraron robustez al ruido e interferencia, dado que BDITM (utilizada aquí) es una base de datos con ruido, pero en situaciones reales de consultorios médicos sencillos, identificar una señal LS, con ruido será lo habitual.

La Tabla IV muestra resultados de experimentos de clasificación en HS con vectores MFCC, mientras que la Tabla V son experimentos con HS pero aplicando vectores Octílicos. El cálculo del modelo HMM en ambos casos de señales HS, requirió 3 iteraciones ($It = 3$) para lograr la convergencia. Los experimentos fueron realizados con diferentes configuraciones en número de estados ($NS=2,3$ y 4) y número de Gaussianas en el GMM ($K=2,3$ y 4) por estado, sin embargo aquí solo se muestran los mejores resultados. La matriz de covarianza en HS utilizada fue la diagonal.

Revisando ambas tablas, se observa que los MFCC arrojaron mejor clasificación para las 3

clases de señales, incluso los vectores MFCC con modelos HMM-GMM, lograron hasta el 100 % de eficiencia clasificando señales del corazón (HS). Comparando los resultados HMM-GMM para el caso de señales LS, si descartamos la covarianza completa no hay una superioridad evidente de alguna modalidad de vector con respecto a la otra; mientras que en señales HS, si hay una superioridad indudable en HMM-GMM aplicando vectores MFCC.

Lo anterior hace pensar que dos modalidades (MFCC y Octílicos) no siempre arrojan resultados consistentes en todos los casos. Los vectores MFCC se apoyan en un banco de filtros distribuido en un rango de frecuencias relacionadas con la audición humana, lo cual está hecho para distinguir detalles relacionados con la fonética. Los vectores Octílicos se inspiraron en medidas de función pulmonar relacionados con estadísticas de dispersión.

Esto significa que los sonidos HS están más cerca de señales acústicas fonéticas y los MFCC son capaces de revelar detalles finos de estas señales. Por otro lado, las señales LS están más cercanas al ruido y tienen un espectro más disperso, con formantes menos definidas que en el caso de señales acústicas fonéticas. Por lo tanto, los vectores Octílicos son eficientes para las señales LS, pero menos eficientes en el caso de HS con respecto a vectores MFCC.

Tabla VI Tabla de contingencia aplicando HMM de 3 estados y 3 gaussianas por estado, para señales LS Normales (p) vs Crepitantes (n), con cuartiles

Hipótesis	p	n	Sensibilidad	Especificidad
P	7	0	1	1
N	0	7		

Tabla VII Tabla de contingencia aplicando HMM de 3 estados y 3 gaussianas por estado, para señales LS Normales (p) vs Sibilancias (n), con cuartiles

Hipótesis	p	n	Sensibilidad	Especificidad
P	7	0	1	1
N	0	7		

Tabla VIII Tabla de contingencia aplicando HMM de 4 estados y 3 gaussianas por estado, para señales HS Normales (p) vs Estenosis (n), con MFCC

Hipótesis	p	n	Sensibilidad	Especificidad
P	17	0	1	1
N	0	17		

Tabla IX Tabla de contingencia aplicando HMM de 4 estados y 3 gaussianas por estado, para señales HS Normales (p) vs SVD (n), con MFCC

Hipótesis	p	n	Sensibilidad	Especificidad
P	17	0	1	1
N	0	17		

A partir de los resultados se observa cierta robustez de los modelos HMM al ruido, lo cual hace pensar que para la secuenciación de eventos en presencia de ruido es fundamental un HMM y es menos afectado por el ruido que en el caso de modelado únicamente con GMM. Los resultados aquí obtenidos son similares pero superiores a otros realizados con anterioridad con discriminantes lineales [10].

Después de varios experimentos con distintas modalidades de vectores acústicos (Octiles, Cuartiles y MFCC), y buscando expresar y mejorar los resultados en términos de tablas de

contingencia, las opciones fueron Cuartiles para el caso de señales LS y MFCC para señales HS, como se muestra en las tablas VI a IX. En estas tablas la señal de entrada es denotada por p y n, y la señal hipotética por el sistema es P y N. A partir de las tablas se puede determinar un excelente compromiso del sistema para detectar correctamente a los verdaderos positivos (Normales) y a los verdaderos negativos (Adventicias). Lo anterior refuerza la idea de que los Cuantiles son una excelente modalidad para señales LS y los MFCC para señales HS. Otro aspecto interesante a denotar es que en este experimento se aumentó la cantidad de señales HS para entrenamiento, lo que condujo a modelos HMM más robustos.

CONCLUSIONES

El presente trabajo muestra la capacidad de los modelos HMM-GMM para clasificar señales LS y HS con modalidades de vector Octílicos, Cuartiles y MFCC. Aquí se utilizaron varias configuraciones, en ambos casos de señales LS y HS se alcanzó el 100 % de eficiencia de clasificación con MFCC. HMM-GMM logró distinguir señales normales y patologías con 3 estados para LS y para HS. En lo que respecta a GMM, fue suficiente un modelo con 3 Gaussianas para capturar las variaciones de cada uno de los estados en HS. Consecuentemente, si se desarrollara un sistema para clasificar señales LS y HS combinadas, aplicando una arquitectura apoyada en HMM-GMM, tal vez convendría considerar vectores MFCC.

Otra consideración importante que se desprende de las Tablas II y III, es que el tipo de covarianza influye en los resultados independientemente de si se utilizan MFCC o vectores Octílicos; en nuestro caso la covarianza diagonal fue mejor, pero esto amerita más experimentos y datos para una validación más rigurosa.

Como ya se ha discutido en otros trabajos [30, 31], la modelación de secuencias de eventos, en este caso LS y HS con HMM, parece superior al modelado únicamente con GMM, ya que de acuerdo a los resultados HMM es menos sensible al ruido, y considera HS como parte de la

interferencia dentro de la secuencia de eventos en una señal LS.

Lo anterior es ventajoso en un escenario realista de auscultación cuantitativa para consultorios médicos con escasos recursos, así como para una proyección por médicos generales, o quienes tengan un limitado nivel de entrenamiento.

En un futuro sería interesante extender la clasificación a enfermedades definidas y a sectores específicos de la población local, ya que la contaminación afecta más a algunos sectores dependiendo del clima. Igualmente, implementar una arquitectura incluyente para clasificar señales LS y HS.

REFERENCIAS

1. Kosmidou V E, Hadjileontiadis L J. "Sign language recognition using intrinsic-mode sample entropy on sEMG and accelerometer data". *IEEE Trans Biomed Eng*, 2009; 56(12): 2879-90.
2. Gaunt S L, Nelson D A, Dantzker M S, Budney G F, Bradbury J W, Zink R. "New directions for bioacoustics collections". *The Auk*, 2005; 122(3): 984-987.
3. Reyna-Carranza M A, Moreno-Flores A P, Lopez-Avitia R. "Acute respiratory infections and its Spearman correlation with meteorological and air pollutants data from Mexicali, BC, Mexico". *PAHCE-2010 (Pan American Health Care Exchanges Conf. 2010)*, 2010; Lima, Peru
4. Reyna-Carranza M A, Merida-Palacio J V, Soria-Rodriguez C. "Acute effects of the PM10 and O3 in the healthy school children pulmonary function from Mexicali, BC, Mexico". *PAHCE-2010 (Pan American Health Care Exchanges Conf. 2010)*, 2010; Lima, Peru
5. Mayorga P, Besacier L, Lamy R, Serignat J F. "Audio packet loss over IP and speech recognition". *IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU '03*. 2003
6. (CEPAL) C E P A L Y E C. "Panorama Social de América Latina <http://www.eclac.org/publicaciones/xml/5/48455/PanoramaSocial2012DocI-Rev.pdf>". 2012;
7. Moussavi Z. "Respiratory sound analysis". *IEEE Engineering in Medicine and Biology Magazine*, 2007; 26 (1): 15.
8. Gnitecki J, Moussavi Z, Pasterkamp H. "Diagnostic potential in state space parameters of lung sounds". *Medical & biological engineering & computing*, 2008; 46(1): 93-99.
9. Charleston-Villalobos S, Martinez-Hernandez G, Gonzalez-Camarena R, Chillem G, Carrillo J G, Aljama-Corrales T. "Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients". *Comput Biol Med*, 2011; 41(7): 473-82.
10. Mayorga Ortiz P, Druzgalski C, Criollo Arellano M, González Arriaga O. "GMM y LDA Aplicado a la Detección de Enfermedades Pulmonares". *Revista Mexicana de Ingeniería Biomédica*, 2013; 34(2): 131-144
11. Solé-Casals J, Zaiats V. "Advances in nonlinear speech processing". *International Conference on Nonlinear Speech Processing, NOLISP 2009*, 2009;
12. Milner B, James A. "Robust speech recognition over mobile and IP networks in burst-like packet loss". *Audio, Speech, and Language Processing, IEEE Transactions on*, 2006; 14(1): 223-231.
13. Kiyokawa H, Pasterkamp H. "Volume-dependent variations of regional lung sound, amplitude, and phase". *Journal of Applied Physiology*, 2002; 93(3): 1030-1038.
14. Besacier L, Mayorga P, Bonastre J F, Fredouille C, Meignier S. "Overview of compression and packet loss effects in

- speech biometrics". Vision, Image and Signal Processing, IEE Proceedings, 2003; 150(6): 372-376.
15. Istrate D M. "Detection et Reconnaissance des Sons pour la Surveillance Médicale, These Doctorale, France". INPG, 2003
 16. Mayorga P, Druzgalski C, Vidales J. "Quantitative Models for Assessment of Respiratory Diseases". PAHCE-2010 (Pan American Health Care Exchanges Conf. 2010), 2010; Lima, Peru
 17. Milner B, Semnani S. "Robust Speech Recognition over IP Networks". IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP2000, Jun 2000; Istambul, Turkey
 18. Pearce D. "An Overview of ETSI Standards Activities for Distributed Speech Recognition Front-Ends". AVIOS 2000: The Speech Applications Conference, May 22-24 2000; San Jose, CA, USA
 19. Fredouille C, Mariéthoz J, Jaboulet C, Hennebert J, Mokbel C, Bimbot F. "Behavior of a Bayesian Adaptation Method for Incremental Enrollment in Speaker Verification". IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP2000), 2000; Istambul, Turkey
 20. Mayorga P, Olguín M, González O H, Flores N, Luis V. "Quantile Acoustic Vectors vs. MFCC Applied to Speaker Verification". Int J Adv Robot Syst, 2014, 11:21 | doi: 10.5772/56256
 21. Mayorga P, Druzgalski C, González O H, Lopez H S. "Modified classification of normal Lung Sounds applying Quantile Vectors". Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, 2012;
 22. Mayorga P, Druzgalski C, González O H, Zazueta A, Criollo M A. "Expanded Quantitative Models for Assessment of Respiratory Diseases and Monitoring". PAHCE-2011 (Pan American Health Care Exchanges Conf. 2011), March 2011; Rio de Janeiro, Brazil
 23. Pearce D. "An Overview of ETSI Standards Activities for Distributed Speech Recognition Front-Ends". AVIOS 2000: The Speech Applications Conference, 2000.; San Jose, California, USA
 24. Reynolds D. A. "Gaussian Mixture Modeling Approach to Text-Independent speaker Identification ". Georgia Institute of Technology, 1992;
 25. Webb Andrew R. Statistical Pattern Recognition. John Wiley & Sons Ltd, 2002.
 26. Rabiner L R, Juang B H. Fundamentals of speech recognition. PTR Prentice Hall (Englewood Cliffs, N.J.), 1993.
 27. "Boston Children's Hospital <http://www.childrenshospital.org/>".
 28. "Texas Heart Institute <http://www.texasheartinstitute.org/AboutUs/index.cfm>".
 29. Mayorga P, Druzgalski C, Gonzalez O H. "Quantile vectors based verification of normal lung sounds". PAHCE-2012 (Pan American Health Care Exchanges Conf. 2012), 2012; Miami, Florida
 30. Mayorga P, Druzgalski C, Miranda J, Zeljkovic V, Gonzalez O H. "The HMM diagnostic models of respiratory sounds". PAHCE-2014 (Pan American Health Care Exchanges Conf. 2014) 2014; Brasilia, Brazil
 31. Mayorga P, Druzgalski C, Calderas D, Zeljkovic V. "Multimodal classification of heart sounds attributes". PAHCE-2014 (Pan American Health Care Exchanges Conf. 2014) 2014; Brasilia, Brazil