



GMM y LDA Aplicado a la Detección de Enfermedades Pulmonares

P. Mayorga Ortiz*,**
C. Druzgalski**
M.A. Criollo Arellano*
O.H. González Arriaga*

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

**Electrical Engineering Department, California State University, 1250 Bellflower Blvd, Long Beach, CA 90840, USA.

RESUMEN

El propósito de este artículo es presentar metodologías que pueden ser usadas para la valoración cuantitativa de los sonidos del pulmón, así como los indicadores de desórdenes respiratorios. En este contexto, se realizaron experimentos utilizando señales normales y anormales de la respiración (LS), las cuales fueron modeladas y evaluadas utilizando principalmente la base de datos RALE y señales de sujetos saludables y no saludables, logrando hasta un 98 % de eficiencia. En la práctica médica la evaluación de enfermedades respiratorias involucra a la auscultación, pero la aplicación de métodos de análisis cuantitativos de señales podría mejorar estas valoraciones. En particular, se sugiere una metodología de evaluación acústica basada en representaciones de vectores acústicos MFCC (Coeficientes Cepstrales en Frecuencia Mel), GMM (Modelos Mezclados Gaussianos) y LDA (Análisis Discriminante Lineal). Estas técnicas podrían asistir en un análisis más amplio, identificación y diagnóstico de desórdenes pulmonares manifestados por sonidos respiratorios peculiares tales como sibilancias, crepitancias y asma, y distinguiéndolos de los sonidos respiratorios normales.

Palabras clave: análisis discriminante lineal (LDA), modelos mezclados Gaussianos (GMM), crepitancias, sibilancia.

Correspondencia:
 Marco Antonio Criollo Arellano
 Correo electrónico:
 Criollo_300@hotmail.com

Fecha de recepción:
 23 de Noviembre de 2012
Fecha de aceptación:
 11 de Junio de 2013

ABSTRACT

This study presents experimentally tested methods, which can be used for a quantitative assessment of respiratory sounds as the indicators of pulmonary disorders. In particular, conducted experiments considered both normal and abnormal lung sounds (LS). As a part of the RALE Database, signals were recorded from healthy subjects and those with respiratory disorders. Current medical practices including evaluation of respiratory diseases often involve qualitative and frequently subjective auscultation. However, the application of quantitative signal analysis methods could improve the assessments of these diseases. In particular, we utilized acoustic evaluation methodologies based on the MFCC (Mel frequency Cepstral Coefficients) acoustic vectors representation, GMM (Gaussian Mixed Models), and LDA (Linear Discriminant Analysis). To assure the validity of determined class models representing diagnostic classification, the LS signals were cross validated within sequential sets of respiratory cycles for a given subject as well as cross correlated within the specific groups of subjects representing particular conditions of normal or given class of abnormal pulmonary functions. Higher order MFCC vectors, including 9, 10 and 11 Gaussian mixtures, resulted in improved classification of the LS attributes, reached up to 98 % of efficiency recognition. This documented automated classification of LS makes it suitable for a more efficient mass screening of respiratory disorders. In particular, the presence of peculiar sounds such as crackles and wheezes lead to more robust models thus reflecting the useful applicability of the presented diagnostic tool. These techniques can assist in broader analysis, identification, and diagnosis of pulmonary disorders manifested by peculiar auscultatory findings.

Keywords: Linear Discriminant Analysis (LDA), Gaussian Mixture Models (GMM), crackles, wheezes.

INTRODUCCIÓN

Las condiciones ambientales y la contaminación en el aire se consideran factores críticos que contribuyen al incremento en enfermedades respiratorias, incluyendo asma, lo cual continúa siendo un problema debido al crecimiento poblacional. Los expertos en asma consideran que los sonidos adventicios no son un elemento suficiente para el diagnóstico de asma; de cualquier manera, los sonidos pulmonares continúan siendo fundamentales para el diagnóstico de enfermedades respiratorias [1]. La literatura científica en señales del pulmón, frecuentemente relaciona las sibilancias con el asma, así como con enfermedades crónicas

obstructivas [1-5]. Diversos estudios han demostrado que los padres de los niños no han sido capaces de identificar los episodios de las sibilancias durante las pruebas médicas o cuando un médico los interroga [6-13]. Además, es esencial tener un historial médico bien documentado, preferentemente con una evaluación de sonidos adventicios [6-14]. La auscultación que usa el estetoscopio tradicional tiene muchas limitaciones debido a que depende de las características del estetoscopio, el cual prácticamente elimina las bajas frecuencias de los sonidos pulmonares (LS, por sus siglas en inglés) que el oído humano no puede percibir. Otras limitantes están relacionadas con las capacidades

físicas del oído y la habilidad para distinguir, por parte del auscultador, diferentes patrones de sonido [2-5].

En particular, las herramientas digitales proveen una manera más precisa de identificar niños con desarrollo temprano de sonidos pulmonares anormales y fortalecer diagnósticos médicos objetivos [4-14]. Algunos experimentos relacionados con la evaluación del oído humano, muestran la existencia de una relación entre la edad y la deficiencia en la detección de cualquier sonido peculiar durante la auscultación [4, 5]. Por el contrario, los métodos digitales, tales como el reconocimiento de patrones y el procesamiento digital de señales han mostrado resultados más precisos [2, 4]. Actualmente, algunas comunidades científicas como *Computerized Respiratory Sound Analysis (CORSA)* y la *International Lung Sounds Association (ILSA)*, promueven el uso de métodos digitales para el tratamiento y detección de enfermedades respiratorias.

BASE DE DATOS Y CARACTERÍSTICAS DE LOS SONIDOS

En lo general, hay dos tipos de LS. Primeramente los sonidos traqueales, los cuales pueden ser escuchados sobre la tráquea, con un volumen relativamente alto y con rango frecuencial de 0 a 2 kHz; sus componentes en el espectro de frecuencias declinan rápidamente para valores mayores que 850-900 Hz. En segundo lugar los sonidos vesiculares, los cuales pueden ser escuchados sobre el pecho, en las proximidades a las vías aéreas centrales y con un rango de frecuencia desde 0 a 600 Hz [2-5]. Los sonidos del corazón pueden ser filtrados parcialmente cuando se usa un filtro pasa-alta, ya que estos se traslapan con los LS a bajas frecuencias. Algunos de los sonidos pulmonares normales están típicamente debajo de los 400 Hz [6, 9, 11, 12, 14]. Con el fin de obtener una amplia cobertura en el rango de frecuencias, las señales deben ser obtenidas a través de un estetoscopio electrónico con un rango de frecuencias que va de 20 Hz hasta 5 kHz. Sin embargo, las frecuencias

de muestreo normalmente aplicadas en señales de audio, van desde 4 kHz hasta 22 kHz. Un sistema de reconocimiento (sea específico para voz o para pulmón) es un sistema inteligente y automático que simula el proceso de reconocimiento que un ser humano efectúa sobre la señal acústica. Se propone el uso de los vectores acústicos MFCC, como método de extracción de características de la señal acústica, debido a que estos son una forma de emular la respuesta del oído humano a las diferentes frecuencias del ambiente (por lo que la extracción de características con MFCC sobre voz, podría ser igualmente efectiva sobre señales del pulmón). El análisis en tiempo típicamente para audio está entre 20 ms y 50 ms, significando que para una frecuencia de muestreo de 10 kHz, un bloque de señal debería tener 256, 512 o 1024 muestras, ya que el algoritmo FFT solamente opera en 2^n datos [4], por lo que será necesario agregar “camillas de ceros” para ajustar el tamaño de trama necesario.

Sonidos adventicios

Para propósitos de este estudio, sólo los siguientes sonidos adventicios serán descritos:

- a) Crepitancias: Son LS discontinuos, con una duración de al menos 20 ms y un rango de frecuencia típicamente desde 10 hasta 2000 Hz [6, 8, 9].
- b) Sibilancias: Sus frecuencias dominantes van alrededor de los 100 a 2000 Hz y sus rangos de duración comprenden desde los 80 hasta los 250 ms [6, 9, 11, 12, 14].

Repositorio de sonidos respiratorios RALE

Desafortunadamente, hay muy pocas alternativas cuando se trata de bases de señales LS, y esta es la razón por la que la base de datos RALE fue usada para los experimentos. Esta base de datos fue elaborada por la Universidad de Manitoba, Winnipeg, Manitoba, Canadá. Es un repositorio de grabaciones obtenidas de pacientes los cuales exhibían sonidos respiratorios normales, crepitancias, sibilancias y otros sonidos pulmonares. RALE contiene los

siguientes registros de sonidos pulmonares: 5 de asma, 4 de crepitancia, 8 normales, 4 de estridor y 7 de sibilancias, los cuales pueden ser útiles en el modelado; por último, contiene 24 grabaciones no etiquetadas que pueden ser útiles para evaluar la eficiencia del sistema. Estas señales fueron tratadas (previamente a ser catalogadas dentro del repositorio) con un filtro pasa-alta en 7.5 Hz para suprimir cualquier desplazamiento o compensación en DC, utilizando un filtro Butterworth de primer orden. Adicionalmente, un segundo filtro Butterworth pasa-baja de 8vo orden fue aplicado en 2.5 kHz para evitar el efecto de "Aliasing" en la señal. Todas las señales de RALE fueron muestreadas a 10 kHz.

Por último, aunado al uso de las señales dentro del repositorio RALE, se utilizaron 28 señales grabadas de alumnos en el Instituto Tecnológico de Mexicali (ITM). Estas señales incorporadas fueron catalogadas como normales y fueron capturadas bajo las mismas consideraciones que las señales contenidas en el repositorio RALE, con la intención de seguir con los mismos estándares para todos los datos utilizados en el presente trabajo.

MODELADO GMM Y LDA DE SONIDOS RESPIRATORIOS

En sistemas de reconocimiento, es ideal contar con una gran cantidad de señales por cada caso, tanto para modelado como para experimentación, la cual es dividida en dos particiones. La primera es usada para modelos acústicos computacionales, mientras que la segunda es usada para propósitos de evaluación. Debido a que se utilizó una base de datos limitada en los experimentos, el método de validación cruzada fue el más apropiado [15]. Iniciaremos esta sección hablando de la metodología con GMM y posteriormente pasaremos a discutir la metodología con LDA.

Vectores de Características MFCC

Un sistema médico útil y fácil de implementar debe incluir un diccionario de modelos

acústicos de patologías que permitan hacer una discriminación de las señales LS, a fin de proveer una hipótesis con respecto a la salud del paciente o posibles patologías [16-18]. En este enfoque, los LS obtenidos de la base de datos RALE y los de los alumnos del ITM, fueron parametrizados (tanto en la fase inspiratoria como espiratoria), realizando un pre-énfasis con un filtro FIR (con coeficientes 1 y -0.97) sobre una ventana de Hamming de 30 ms cada 10 ms (*i.e.*, la señal es descompuesta en n tramas, determinadas por el tamaño de la ventana y su respectivo traslape). Posteriormente, el algoritmo de la FFT (con 512 puntos ya que esto permite observar las bandas de frecuencia más importantes) es aplicado trama por trama de la señal; en seguida el módulo es calculado sobre este resultado; el modulo correspondiente es multiplicando por un banco de filtros distribuidos en la escala *Mel*; finalmente, el resultado es nombrado vector de características de Coeficientes Cepstrales en Frecuencia Mel (MFCC por sus siglas en inglés) [16, 19-22], como se muestra en la figura 1.

La ventaja de aplicar MFCC, es que el proceso está relacionado con la respuesta del oído a las componentes frecuenciales del sonido, caracterizándolo como coeficientes dentro de un vector acústico.

Modelado con GMM

Un Modelo Mezclado Gaussiano (GMM) es caracterizado por sus medias, covarianzas y ponderaciones; y cada caso es representado por un modelo GMM (λ). Los modelos acústicos son obtenidos para cada caso o patología en la Etapa de Entrenamiento. Los vectores MFCC son extraídos del conjunto de grabaciones para cada posible patología o señal normal dependiendo del caso. Una vez que todos los vectores están reunidos, el algoritmo de Maximización-Expectación (EM) es aplicado para estimar los correspondientes modelos acústicos; una explicación extendida puede ser encontrada en [16, 17, 23, 24]. Una Mezcla de Densidades Gaussianas es una suma ponderada de M densidades componentes, como se muestra en la siguiente ecuación:

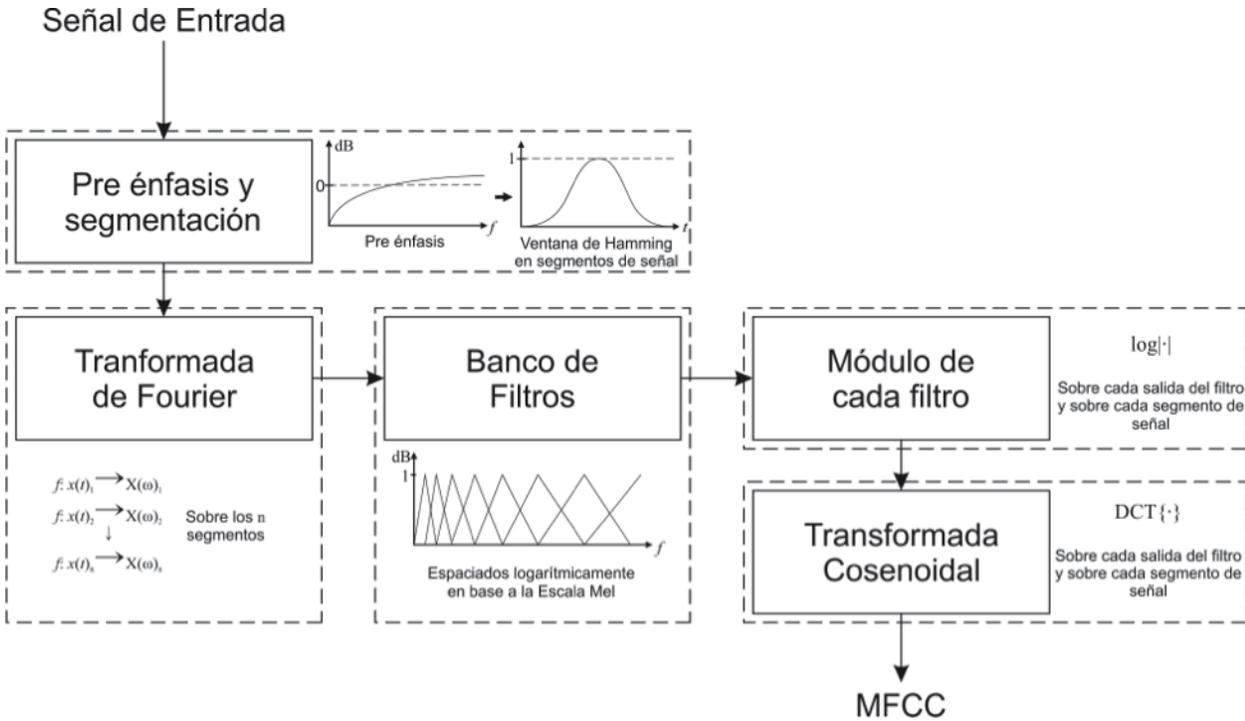


Fig. 1. Procedimiento para la obtención de vectores MFCC.

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

Donde \vec{x} es un vector aleatorio D -dimensional (vector acústico MFCC), $b_i \forall i = 1, \dots, M$ 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 [16-18, 23-25] como sigue:

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

Aquí, $\vec{\mu}$ es un vector de medias y Σ_i es la matriz de covarianza. Por otro lado, las ponderaciones de las mezclas satisfacen la restricción $\sum_i^M m_i = 1$. El conjunto de modelos GMM es parametrizado por los vectores de medias, matrices covarianza y las ponderaciones de las mezclas para las densidades Gaussianas. El modelo que contiene los parámetros es representado por la siguiente expresión [16, 17, 23, 24]:

$$\lambda = \{m_i, \vec{\mu}_i, \Sigma_i\} \forall i = 1, \dots, M \quad (3)$$

Usando la regla de decisión de Bayes [23], la fórmula fundamental, la cual presenta la

mejor hipótesis en el proceso de reconocimiento automático de señales, es como sigue:

$$C = \arg_{1 \leq i \leq I} \max p(x|\lambda_i) p(\lambda_i) \quad (4)$$

Debido a la gran cantidad de grabaciones, un gran conjunto de vectores de características son obtenidos, y suponiendo independencia estadística entre los vectores MFCC, la regla para seleccionar la mejor hipótesis está dada por [23]:

$$\prod_{t=1}^T p(\vec{x}_t|\lambda_i) > \prod_{t=1}^T p(\vec{x}_t|\lambda_r), \forall r = 1, \dots, I \quad (5)$$

En el contexto científico de procesamiento de voz y el reconocimiento estadístico de patrones [23], $\prod_{t=1}^T p(\vec{x}_t|\lambda_i)$ es conocido como una función de similitud. Dado que la función logaritmo es monótona, es válido aplicarla en la ecuación (5) [23]. Consecuentemente, la productoria es transformada en una sumatoria, y la nueva expresión es conocida como *regla de decisión de máxima similitud*:

$$L(\lambda_I) = \sum_{t=1}^T \log p(\vec{x}|\lambda_I) \quad (6)$$

La ecuación (6) es usada en el proceso comparativo, en el cual la señal de entrada se asocia al modelo acústico de nuestro diccionario que arroje la mayor verosimilitud.

Modelado con LDA

Supóngase que se tiene un conjunto de vectores de prueba $\{x_1, x_2, \dots, x_n\}$ (estos son, los vectores MFCC), cada uno de los cuales es asignado a una de dos clases, ω_1 o ω_2 (en nuestro caso pueden ser LS normales y anormales). Usando el conjunto designado, se busca un vector de pesos w y un umbral w_0 tal que se pueda obtener la siguiente ecuación de clasificación:

$$w^T x + w_0 \begin{cases} > 0 \\ < 0 \end{cases} \Rightarrow x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases} \quad (7)$$

Con el fin de construir modelos, se aplica el criterio basado en la regla de clasificación de Fisher [18], como sigue:

$$\begin{aligned} w &= S_W^{-1} (m_1 - m_2) \\ w_0 &= -\frac{1}{2} (m_1 + m_2)^T S_W^{-1} (m_1 + m_2) \\ &\quad - \log \left(\frac{p(\omega_2)}{p(\omega_1)} \right) \end{aligned} \quad (8)$$

Donde m_i es la media de la clase i , S_W es la sumatoria de las covarianzas intra-clases, y $p(\omega_i)$ corresponde a la probabilidad *a priori* de cada clase [23].

Cuando la expectativa de clasificación abarca más de dos clases, es posible encontrarse con varias alternativas para poder elaborarla. La primera es la construcción de tantas funciones discriminantes como para poder comprender la interacción de todas las clases, demandando un alto más costo computacional (*i.e.*, en una aproximación, $\sum_{k=2}^C k - 1$ funciones discriminantes para C clases).

Una segunda alternativa consiste en crear una función discriminante para cada clase, asignando x_i a la clase para la cual la probabilidad *a posteriori* $p(x|\omega_i)$, o equivalentemente $\log(p(\omega_i|x))$, es mayor. Sin embargo, si se supusiera que los datos son normalmente distribuidos, con matrices de covarianza iguales (igual a la matriz de

covarianza intra-clase, S_W) en cada clase y medias m_i , entonces la regla de discriminación es: asignar x a la clase ω_i si $g_i \geq g_j$ para todos los $j \neq i, j = 1, \dots, C$, donde

$$g_i = \log(p(\omega_i)) - \frac{1}{2} (x - m_i)^T S_W^{-1} (x - m_i) \quad (9)$$

Si despreciamos los términos cuadráticos de x , resulta [23]

$$g_i = \log(p(\omega_i)) - \frac{1}{2} m_i^T S_W^{-1} m_i + x^T S_W^{-1} m_i \quad (10)$$

La cual es conocida también como *función discriminante de base normal lineal* [23]. Considerando el criterio de Fisher [23] en una aproximación multiclase, tenemos

$$J_F = \frac{a^T S_B a}{a^T S_W a} \quad (11)$$

donde las estimaciones basadas en el muestreo de S_B y S_W están dadas por

$$\begin{aligned} S_B &= \sum_{i=1}^C \frac{n_i}{n} (m_i - m)(m_i - m)^T \\ S_W &= \sum_{i=1}^C \frac{n_i}{n} \widehat{\Sigma}_i \end{aligned} \quad (12)$$

tal que m_i y $\widehat{\Sigma}_i, \forall i = 1, \dots, C$, son las medias muestra y las matrices de covarianzas de cada clase (con n_i muestras), y m es la media muestra de todas las clases. Aquí, se busca un conjunto de vectores de características a_i que maximice J_F , sujeto a la restricción de normalización $a_i^T S_W a_j = \delta_{ij}$ (los vectores de clase centralizados en el espacio transformado son no correlacionados). Esto conduce a la ecuación del eigenvector generalizado simétrico [23]

$$S_B A = S_W A \Lambda \quad (13)$$

donde A es la matriz cuyas columnas son a_i y Λ es la matriz diagonal de eigenvalores. Si S_W^{-1} existe, se puede escribir como

$$S_W^{-1} S_B A = A \Lambda \quad (14)$$

Los eigenvectores corresponden a los eigenvalores más grandes usados en la extracción de

características. Si A es la transformación de discriminante lineal, entonces S_W^{-1} puede ser escrito como

$$S_W^{-1} = AA^T + A_{\perp}A_{\perp}^T \quad (15)$$

donde $A_{\perp}^T m_j = 0$ para todos los valores de j . Usando la expresión anterior para S_W^{-1} en la ecuación (10), se obtiene la siguiente función discriminante

$$g_i = \log(p(\omega_i)) - \frac{1}{2}(y(x) - y_i)^T (y(x) - y_i) \quad (16)$$

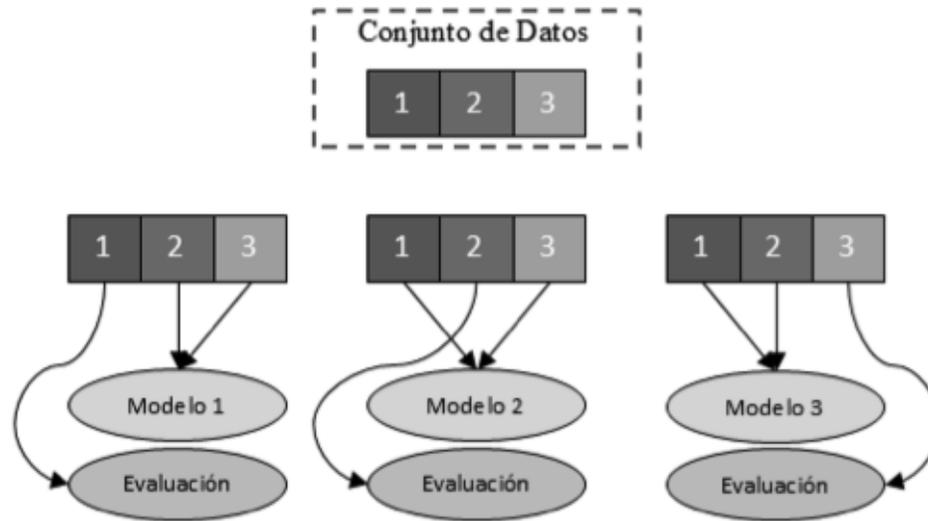
e ignorando los términos que son constantes a través de las clases, la discriminación está basada en

$$g_i = \log(p(\omega_i)) - \frac{1}{2}y_i^T y_i + y^T(x) y_i \quad (17)$$

Se obtiene un *Clasificador de Media-Clase más Cercana* en el espacio transformado basado en el criterio de Fisher [23], donde $y_i = A^T m_i$ y $y(x) = A^T x$.

RESULTADOS

El enfoque utilizado para el particionamiento de la base de datos, y la posterior evaluación del sistema, fue la Validación Cruzada, el cual fue implementado con el objetivo de probar el sistema con señales diferentes (tanto para experimentos de GMM, como de LDA). Estas señales fueron aplicadas en la etapa de aprendizaje (*i.e.*, conjunto de entrenamiento), para poder obtener un punto de vista objetivo de la eficiencia del sistema. El procedimiento, en particular utilizado en este trabajo, fue Validación Cruzada por K -fold (aplicado para cada clase); el concepto básico es dividir los datos en K particiones de igual tamaño (una señal por partición). Una partición es reservada para las pruebas, y el resto de los datos son usados para ajustar el modelo. Este procedimiento es repetido hasta que las K particiones han sido usadas como conjunto de prueba; esta es la razón por la que en la figura 2, la señal i fue reservada para prueba.



En general, si el Conjunto de Datos
 $D = x_1, x_2, \dots, x_i, \dots, x_N$
 El Modelo i usa $M(i) = V, \forall V \subset D, x_i \notin V$,
 ya que x_i es usada para evaluación $E = x_i$

Fig. 2. Algoritmo de evaluación VC.

Para el reconocimiento de LS, cada caso es representado por un GMM y es designado como un modelo λ para ese caso. Es razonable suponer que las señales respiratorias parametrizadas con MFCC pueden ser caracterizadas por un modelo. La forma del espectro y los histogramas obtenidos de cada grupo de datos pueden ser representados por varias densidades Gaussianas ponderadas. Cada Gaussiana tiene media $\vec{\mu}_i$, y las variaciones del promedio de la forma del espectro pueden ser representadas por la matriz de covarianza $\vec{\Sigma}_i$. Esto es importante, porque la misma señal capturada varias veces, aun del mismo paciente, no tiene exactamente la misma forma. Una señal, de la misma patología, debe ser grabada con múltiples pacientes para obtener modelos más robustos. En los experimentos, las clases corresponden a los datos obtenidos de los eventos acústicos con LS normales y adventicias; la complejidad de la señal puede influenciar el número necesario de densidades en los modelos GMM.

Uno de los primeros experimentos con reconocimiento de señales LS fue realizado

calculando diferentes tamaños de modelos GMM, extrayendo observaciones de la base de datos RALE como vectores acústicos MFCC, y entonces calculando un diccionario de modelos usando señales normales, crepitantes, sibilantes y asmáticas, como se muestra en la Tabla 1. Otros experimentos fueron realizados para obtener el tamaño mínimo razonable para un modelo GMM, como se muestra en las Tablas 1 y Tabla 2. Básicamente, en estos experimentos se varía el tamaño del vector y del modelo, teniendo también en cuenta la inclusión o exclusión de las derivadas del vector MFCC. Los coeficientes del vector MFCC están relacionados a la energía resultante de un conjunto de filtros, los cuales están linealmente distribuidos hasta el orden de los 1000 Hz, pero logarítmicamente para frecuencias mayores. La primera y la segunda derivada son sensibles al ruido, y pueden ser confundidas con crepitancias, mientras que éste no es el caso con las otras señales. No obstante, las derivadas son útiles para inspeccionar la dinámica o la evolución de la señal.

Tabla 1. Tamaño mínimo aceptable para GMM (Número de densidades), y vectores MFCC (Número de coeficientes); modelos GMM con mejores resultados (GMM9, GMM10, GMM11).

	Mínimo # Den ^a	# de Coef. ^b sin Δ 's	GMM9: %Recon. ^c	GMM10: %Recon. ^c	GMM11: %Recon. ^c
Normal	6	7	48	44.2	46.1
Crepitante	6	7	90.3	98	98
Asma	6	7	50	51.9	50
Sibilante	6	7	25	25	26.9

^aNúmero de densidades; ^bNúmero de coeficientes y ^cReconocimiento.

Tabla 2. Eficiencia de reconocimiento con vectores MFCC, con y sin derivadas (deltas), para promedio de valores obtenidos para 1-20 densidades de modelos GMM.

Experimento:	13 Coef.	13 Coef.+13 Δ .	13 Coef.+13 $\Delta\Delta$.
Normal	46.2	48.7	51.2
Crepitantes	90	87.5	86.2
Asma	61.2	53.7	50
Sibilancia	30	31.2	32.5

El mejor comportamiento de los resultados fue obtenido aplicando 11 Gaussianas (GMM11), pero los cambios obtenidos no fueron substanciales comparados con GMM9 y GMM10; sin embargo, es evidente que los resultados mejoraron cuando los vectores MFCC fueron mayores a cuatro coeficientes para modelar, excluyendo algunos “comportamientos” irregulares (para GMM9, y al estabilización en GMM10). Esto es razonable porque vectores más grandes incluyen características relacionadas con bandas de energía más significativas, como se muestra en la Tabla 2 (MFCC=13).

Otra aproximación con la que se realizaron experimentos sobre reconocimiento está relacionada con funciones discriminantes, aplicando LDA. Específicamente, se llevaron a cabo dos tipos de experimentos: primeramente clasificar señales normales y anormales; consecutivamente, clasificar en forma más exhaustiva las señales adventicias. En este caso, se emplearon 3 clases diferentes de señales, señales normales, señales sibilantes y señales crepitantes (contenidas en el repositorio RALE, y de la base de datos de señales normales. Adicionalmente a lo anterior, las evaluaciones fueron elaboradas con el enfoque de VC (para tener una referencia objetiva del reconocimiento, Figura 2). También se construyeron 3 clasificadores discriminantes, los necesarios para poder hacer discriminación entre por lo menos dos clases de la base de datos (*i.e.*, N-C, N-S, C-S). Por lo tanto, en la primera etapa de clasificación, la señal era introducida a los 2 clasificadores que relacionaban la clase normal con el resto de las clases (N-C y N-S), de resultar la hipótesis del sistema, que la clase es normal, la señal era considerada normal y el experimento terminaba. En el caso de que la hipótesis del sistema fuese clase no normal (C o S), se hacía una segunda etapa de clasificación, donde la señal es sometida al clasificador C-S y entonces detectada la patología correspondiente. Al igual que en GMM, el vector de características utilizado en la clasificación fue MFCC. Los resultados obtenidos en la primera etapa (mostrados en la Tabla 3) muestran que la clasificación entre normal y anormal arroja porcentajes del 100 % de eficiencia. Una vez

efectuada la clasificación entre señal normal y anormal, se aplicó una siguiente etapa para identificar el tipo de señal adventicia (Tabla 4), remarcando el hecho de que esta etapa se aplica solo si la señal fue reconocida como adventicia. En las evaluaciones con VC, la eficiencia de reconocimiento entre las dos clases de señales adventicias disminuyó en algunos casos hasta el 50 % (crepitancias, Tabla 4), incluso con buenos resultados de hasta 90 % (C1, S1 y S2). Esto se explica por la limitada cantidad de señales con las que se contaba para los experimentos.

Los resultados en la tabla 4 podrían mejorarse al incrementar el número de señales en el entrenamiento, obteniendo funciones discriminantes más robustas (modelos mejor entrenados). Otro aspecto a resaltar que se observó, es el bajo costo computacional en comparación con GMM (los clasificadores LDA son modelos menos complejos matemáticamente que los GMM), haciendo de LDA una excelente alternativa cuando el objetivo es diferenciar entre señales normales y anormales.

Tabla 3. Primeros experimentos utilizando LDA y MFCC en señales respiratorias, para separar señales normales de adventicias

Clase	Hipótesis	
	Normal	Adventicia
N1 - N14	100%	0%
C1 - C4	0%	100%
S1 - S3	0%	100%

Tabla 4. Segundos experimentos utilizando LDA y MFCC, realizados para identificar el tipo de señal adventicia

Clase	Hipótesis	
	Crepitante	Sibilante
C1	91.03%	8.97%
C2	73.69%	26.3%
C3	42.74%	57.26%
C4	42.3%	57.69%
S1	13.22%	86.77%
S2	9.83%	90.16%
S3	99.19%	0.8%

DISCUSIÓN

En el contexto del reconocimiento de voz, es necesaria una gran cantidad de datos con el fin de computar modelos más robustos [15-18]. Desafortunadamente, la base de datos RALE carece de un número deseado de grabaciones para representar enfermedades respiratorias del paciente, o detalles acústicos de las vías aéreas respiratorias. No obstante, la función del sistema respiratorio sirve como una fuente útil de información acústica para la aproximación mostrada en este trabajo. Adicionalmente, las señales adventicias de RALE no están relacionadas con un grupo de edad particular o características individuales, o una localización definida del cuerpo. Pero a pesar de estas limitantes, los experimentos mostraron una precisión superior al 50% en evaluaciones VC, lo cual confirma el potencial de esta metodología [20]. La Tabla 5 muestran la incertidumbre (hipótesis erróneas) durante la etapa de evaluación, conteniendo la cantidad de coeficientes incluidos en los vectores MFCC, número de densidades con las que se elaboraron los modelos GMM, así como también las señales de evaluación (N, C, A, W o S); considerando una señal de entrada a reconocer (IN) y la hipótesis de salida del sistema (OUT). Las incertidumbres fueron: 1 con un modelo de cuatro densidades (*i.e.*, una señal Sibilante *W* de entrada, fue reconocida por el sistema como una señal crepitante *C*, utilizando un vector de 4 coeficientes MFCC y un modelo de 4 Gaussianas), 3 con modelos de 5 densidades (N-W, N-A, W-A), y dos con modelos de siete densidades (N-W, N-W).

Una vez más, los modelos con menos coeficientes MFCC y menos densidades GMM producen mayor errores de reconocimiento (4 y 5 coeficientes MFCC, 4 y 5 densidades GMM). Por otra parte, las señales con el peor grado de reconocimiento fueron los sonidos normales y sibilantes, pero el rendimiento general mejoró con el incremento del número de los coeficientes y densidades en sus modelos. Una explicación es que los modelos con más coeficientes MFCC y densidades GMM representan más eficientemente las características acústicas y la estructura fina del tracto respiratorio.

Tabla 5. Resumen de incertidumbre

Configuración:	Número de reconocimiento (Señales IN) - (Hipótesis Out)			
	4	5	6	7
4 MFCC	W^c	W^c	0	N^a
	C^b	N^a		W^c
5 MFCC		A^d	0	
	0	W^c		N^a
		A^d		W^c
		0		

^a(N) Normal; ^b(C) Crepitancia; ^c(W) Sibilancia; ^d(A) Asma

En la mayoría de las evaluaciones, los modelos con el peor grado de reconocimiento fueron aquellos computados para respiración normal y sibilancias; la razón para esto podría ser que la respiración normal tiene un patrón menos definido que las señales adventicias. Debido a que las características particulares de la anatomía entre niños y adultos mayores influyen en las características de la frecuencia de la señal, es necesaria una gran cantidad de diversas grabaciones de diferentes pacientes, en diversas situaciones, en la misma localización del cuerpo, y preferiblemente sobre grupos de edades para validar de manera más contundente los resultados.

Los mejores resultados, indistintamente del experimento, fueron obtenidos con crepitancias; estas señales fueron las mejor reconocidas. Esto sugiere que tienen una banda de frecuencias características con una buena contribución de energía, lo cual hace de ellas fácil de diferenciar entre asma, sibilancias y señales normales. En otras palabras, los picos característicos de las crepitancias contribuyen suficientemente de manera estadística para distinguirlos de las otras. Consecuentemente, esto contribuye en gran medida a la varianza y la media, resultando modelos GMM más robustos. Las señales relacionadas con el asma son muy complejas; el concepto en sí mismo evoluciona y diverge de acuerdo a numerosas opiniones médicas. El diagnóstico de asma a partir únicamente de sonidos adventicios no es lo suficientemente preciso, ya que las grabaciones de asma utilizadas

fueron pocas y heterogéneas. Estas señales no tienen estadísticas distintivas conclusivas, o bandas con una contribución de energía particular. No obstante, obtuvimos precisión en los resultados por encima del 50% en cada caso (Tabla 1, Tabla 2), lo cual alienta al uso de estas técnicas para la evaluación del asma.

La naturaleza melódica de las sibilancias no fue capturada propiamente en las estadísticas y las bandas de energía, como se muestra en las Tablas 1, 2 y 5. Las derivadas podrían mejorar los resultados destacando una banda en particular o una evolución suavizada de la señal, pero este tipo de señal merece un análisis más a profundidad en futuros trabajos.

Las incertidumbres en la Tabla 5 son debidas principalmente a la similitud entre los parámetros estadísticos de las señales normales y sibilantes. Tratando de diferenciar entre estos sonidos, el modelo normal fue calculado usando grabaciones adicionales de LS obtenidos de estudiantes de ingeniería, con el fin de confirmar que la falta de señales normales ocasionó las incertidumbres, i.e. modelos menos robustos. Esta última observación fue confirmada computando un modelo, el cual incluía señales normales adicionales de LS normal, arrojando un menor número de incertidumbres (Tabla 6).

En otros trabajos encontrados en la literatura [20-22], se comenta que el rendimiento en reconocimiento entre señales LS normales y sibilantes puede ser mejorado substancialmente (estos trabajos reportan una eficiencia de reconocimiento entre normal y sibilante de un 90%), donde el uso de GMM y MFCC fueron también las aproximaciones utilizadas. Es importante considerar la diferencia entre las características y condiciones de los modelos, así como las señales utilizadas (i.e. diferente número de mezclas Gaussianas, características substancialmente diferentes entre las señales de las bases de datos, entre otros).

En [22], se reportan modelos GMM de alrededor de 16 mezclas, mientras que en el

presente trabajo se lograron modelos eficientes con tamaños del orden de 9 a 11 Gaussianas, donde se evidencia el potencial de los GMM. Además, en este trabajo también se presenta una situación no observada en los otros trabajos, que corresponde a la clasificación entre LS normal y crepitante. Un punto importante en el que varios trabajos concuerdan [20-22], es el análisis por tramas cuando la extracción de características es con MFCC. Esto es necesario debido a que normalmente se trabaja con señales cuasi-estacionarias o estacionarias solamente en ciertos límites de tiempo.

En otros experimentos, nos apoyamos en el método de Análisis Discriminante Lineal (LDA) con la misma base de datos, y los resultados fueron mejorados en el caso de las señales normales en relación a los obtenidos con GMM, lo cual hace pensar que LDA es más poderoso en problemas de discriminación de clases.

Una explicación puede ser que las señales normales utilizadas fueron de un conjunto diferente (exclusivamente de los estudiantes de ingeniería), y las mediciones fueron obtenidas de un solo punto de auscultación, i.e., el más cercano al pulmón, donde más señales coincidieran (Tabla 3 y 4). Por el contrario, las señales crepitantes y sibilantes mostraron peores resultados; esto puede ser debido a que se usaron diferentes puntos de auscultación en comparación con los otros experimentos (de la base de datos RALE). Asimismo, es importante ver las características propias de cada señal; en el caso de C4, ésta contiene crepitancias gruesas (distinguidas de las finas C1, C2 y C3 por sus componentes frecuenciales bajos), así como S3 que contiene *rhonchus* (sibilancias con contenidos frecuenciales bajos). Lo anterior puede ser un indicio del porqué el sistema de reconocimiento (LDA) confundió estos dos tipos de señales. De cualquier manera, es obvio que este método mejora la clasificación entre dos clases (normal y anormal), pero es necesario llevar a cabo más experimentos para ser más concluyentes en el caso de más de dos clases.

Tabla 6. Reconocimiento usando un modelo LS normales mejorado calculado con RALE más señales LS. Las incertidumbres marcan el número de hipótesis error dados por nuestro sistema GMM, por señal.

Parámetros	Señales Adventicias	# de grabaciones disponibles	# de grabaciones reconocidas	Incertidumbres:			
				N	C	A	S
Mezclas	Normal	36	34/36	-	1	0	1
Gauss ^a :10	Crepitancia	4	4/4	0	-	0	0
Vectores MFCC:	Asma	5	5/5	0	0	-	0
13	Sibilancia	7	7/7	0	0	0	-

^aGaussianas

CONCLUSIONES

Los modelos GMM podrían aplicarse a la auscultación no solo en adultos, sino también podrían ser particularmente útiles en casos de niños por debajo de los 5 años de edad, evitando someterlos a respiraciones forzadas, como en el caso de la espirometría; o en situaciones donde la auscultación es realizada por una persona con experiencia limitada. Estos resultados prometedores demuestran que esta alternativa permite evitar la ambigüedad en el caso del reconocimiento de señales adventicias. Durante los experimentos con dos aproximaciones como lo son GMM y LDA, el tamaño del modelo fue variado con el fin de obtener el mejor equilibrio entre precisión y demanda computacional (complejidad del modelo). Los mejores resultados en GMM fueron obtenidos con 9, 10 y 11 Mezclas Gaussianas. En consecuencia, se encontró que estas técnicas tienen un alto potencial en el diagnóstico de enfermedades en las vías respiratorias y reconocimiento de problemas de salud relacionados. En los experimentos elaborados con anterioridad, fue lograda una precisión en los resultados del 52.5 % aplicando evaluación por Validación Cruzada; mientras que en el reconocimiento de referencia fue lograda una precisión de 98.75 %.

Algunos de los casos que arrojaron mejores resultados con crepitancias fueron con GMM, en contraste la evaluación de señales normales basada en clasificación por LDA fue más exitosa, lo que hace pensar su superioridad en experimentos con dos clases (normales y anormales). En general, los resultados

fueron mejorados aplicando una etapa de pre-procesamiento, usando la base de datos RALE y otras grabaciones de LS adicionales. Es altamente recomendable una base de datos más amplia, como observamos para el modelo de LS normal (Tabla 6); esta deberá contener un número extendido de grabaciones para lograr un análisis más refinado de las bandas más importantes, y determinar un conjunto óptimo de coeficientes MFCC. Estas consideraciones conducen a la obtención de modelos más robustos y modelos acústicos representativos para LS normales o adventicias. Grabaciones LS adicionales tales como la tos podría ampliar las capacidades de la metodología y también el alcance de las aplicaciones.

REFERENCIAS

1. Serra Valdés M A. "Evaluación de la Terapia Inhalada en Pacientes con Asma Bronquial Persistente".
2. Hadjileontidais L J. "Biosignal and compression Standards". M-Health Emerging Mobile Health systems, Topics in Biomedical Engineering', Springer 2006 International Book Series, 2006; 277-292.
3. Gross V, Dittmar A, Penzel T, Schutter F, Von Wichert P. "The Relationship between Normal Lung Sounds, Age, and Gender". Am J Respir Crit Care Med, 2000; 162(905-909).
4. Sovijärvi A R A, Vanderschoot J, Earis J E. "Standardization of computerized

- respiratory sound analysis". *Eur Respir Rev* 2000, 2000; 10(77): 585.
5. Earis J E, Cheetham B M G. "Current methods used for computerize respiratory sound analysis". *Eur Respir Rev* 2000, 2000;
 6. Sovijärvi A R A, Malmberg L P, Charbonneau G, Vanderschoot J, Dalmaso F, Sacco C, et al. "Characteristics of breath sounds and adventitious respiratory sounds". *ERS Journals Ltd* 2000, 2000; 10(77): 591-596
 7. Bush A. "Diagnosis of asthma in children under five". *Prim Care Respir J*, 2007; 16(1): 7-15.
 8. Rossi M, Sovijärvi A R A, Piirilä P, Vannuccini L, Dalmaso F, Vanderschoot J. "Environmental and subject conditions and breathing manoeuvres for respiratory sound recordings". *Eur Respir Rev* 2000, 2000; 10(77): 611-615.
 9. Charbonneau G, Ademovic E, Cheetham B, Malmberg L P, Vanderschoot J, Sovijärvi A R A. "Basic techniques for respiratory sound analysis". *Eur Respir Rev* 2000, 2000; 10(77): 625-635.
 10. Druzgalski C. "Potentials and Barriers of Extensive Auscultatory Databases". 27th International Conference on Lung Sounds, 2000; Sweden
 11. Druzgalski C. "Distributed Analysis of Signal Integrity Impediments in Respiratory Acoustic Signatures WC2003". The World Congress on Medical Physics and Biomedical Engineering, 2003; Sydney, Australia
 12. Druzgalski C, Shenoy N, Kumar S. "Enhanced Pulmonary Function Testing and Segmental Respiratory Performance Evaluation". The WC on Medical Physics and Biomedical Engineering 2006, 2006; Seoul, Korea
 13. Schreur H J, Vanderschoot J, Zwinderman A H, Dijkman J H, Sterk P J. "Abnormal lung sounds in patients with asthma during episodes with normal lung function". *Chest*, 1994; 106(1): 91-9.
 14. Pasterkamp H. "State of Art: Respiratory Sounds Advances beyond the Stethoscope". *Am J Respir Crit Care Med*, 1997; 156(3): 974-987.
 15. Martinez W L, Martínez A R. *Computational Statistics Handbook With Matlab*. Chapman & Hall/CRC 2002.
 16. 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
 17. Mayorga P, Besacier L, Lamy R, Serignat J F. "Audio packet loss over IP and speech recognition". *Automatic Speech Recognition and Understanding*, 2003. ASRU '03. 2003 IEEE Workshop on, 2003; St. Thomas, Virgin Islands, USA
 18. Mayorga P, Druzgalski C, Vidales J. "Quantitative Models for Assessment of Respiratory Diseases". PAHCE 2010, 2010; Lima, Perú
 19. Sahidullah M, Saha G. "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition". *Speech Communication*, 2012; 54(4): 543-565.
 20. Bahoura M, Pelletier C. "Respiratory sounds classification using cepstral analysis and Gaussian mixture models". *Engineering in Medicine and Biology Society*, 2004. IEMBS '04. 26th Annual International Conference of the IEEE, 2004;
 21. Bahoura M, Pelletier C. "Respiratory sounds classification using Gaussian mixture models". *Electrical and Computer Engineering*, 2004. Canadian Conference on, 2004;

22. Jen-Chien C, Huey-Dong W, Fok-Ching C, Chung I L. "Wheeze Detection Using Cepstral Analysis in Gaussian Mixture Models". Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, 2007;
23. Webb A R. Statistical Pattern Recognition. John Wiley & Sons Ltd, 2002.
24. Reynolds D A. "A Gaussian Mixture Modeling Approach to Text-Independent speaker Identification". Georgia Institute of Technology, 1992;
25. Mayorga P, Druzgalski C, Morelos R L, González O H, Vidales J. "Acoustics Based Assessment of Respiratory Diseases using GMM". EMBC 2010 IEEE, 2010; Buenos Aires, Argentina.