Revista Ingeniería Biomédica
ISSN 1909-9762 / E-ISSN 1909-9991
Volumen 13 / Número 25 / enero-junio de 2019 / pp. 25-34
Universidad EIA / Envigado, Colombia

Expanded VAD Guided Subdivision of Cardiopulmonary Sounds

Pedro Mayorga[1], Julio A. Valdez[2, ψ], Christopher Druzgalski[3], Vesna Zeljkovic[4], Gilberto Chavez[1], Monceni Perez[5]

[1]Instituto Tecnológico de Mexicali, Mexicali, México

[2]Universidad del Valle de México, Mexicali, México

[3]California State University, Long Beach, CA, USA

[4]The Lincoln University, PA, USA

[5]Univerisdad Autónoma de Baja California, Mexicali, México

Recibido 4 de abril de 2019. Aceptado 29 de julio de 2019

Abstract**—**Cardiopulmonary auscultation is a diagnostic procedure that has a challenging task since the components of heart rate and lung sounds overlap. There were many approaches to quantify the characteristics of these signals, and one of the newest is the voice activity detection (VAD) and the Gaussian Mixture Models (GMM). Considering the lung and heart sounds as acoustic events, this paper proposes a novel assessment methodology of these diagnostic indicators. Here, a new VAD based on GMM (VAD-GMM) was applied to detect and extract the main events in lung sound and heart sounds. VAD-GMM results were compared with other VAD methodology based on statistical approach, and it was found that VAD-GMM give more definite results. Since Mel Frequency Cepstral coefficients (MFCC) and Quartiles feature vectors, were already successful in pattern recognition, VAD-GMM was carried out using this kind of acoustic vectors. Therefore, this method could add in a transition from qualitative traditional auscultation to quantitative assessment and assisted computerized diagnosis by identifying abnormal acoustic indicators. Diagnosis by computerized detection promises to be a more efficient method than traditional methods, which are limited by the auditory capability and experience of a medical professional.

Keywords—Cardiopulmonary diagnosis, Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), Hilbert transform, Voice Activity Detection (VAD).

Subdivisión Guiada de VAD Ampliada de Sonidos Cardiopulmonares

Resumen**—**La auscultación cardiopulmonar es un procedimiento de diagnóstico que tiene una tarea difícil ya que los componentes de la frecuencia cardíaca y los sonidos pulmonares se superponen. Hubo muchos enfoques para cuantificar las características de

ψ Dirección para correspondencia: julito_valdez@hotmail.com
DOI: https://doi.org/10.24050/19099762.n25.2019.1317

estas señales, y uno de los más nuevos es la detección de actividad de voz (VAD) y los modelos mezclados gaussianos (GMM). Considerando los ruidos pulmonares y cardíacos como eventos acústicos, este artículo propone una nueva metodología de evaluación de estos indicadores de diagnóstico. Aquí, se aplicó un nuevo VAD basado en GMM (VAD-GMM) para detectar y extraer los eventos principales en el sonido pulmonar y cardíaco. Los resultados de VAD-GMM se compararon con otra metodología de VAD basada en el enfoque estadístico, y se descubrió que VAD-GMM da resultados más definitivos. Dado que los coeficientes cepstrales de frecuencia de mel (MFCC) y los vectores de características de cuartiles, ya tuvieron éxito en el reconocimiento de patrones, VAD-GMM se llevó a cabo utilizando este tipo de vectores acústicos. Por lo tanto, este método podría agregar una transición de la auscultación tradicional cualitativa a la evaluación cuantitativa y el diagnóstico computarizado asistido mediante la identificación de indicadores acústicos anormales. El diagnóstico por detección computarizada promete ser un método más eficiente que los métodos tradicionales, que están limitados por la capacidad auditiva y la experiencia de un profesional médico.

***Palabras clave*—**Diagnóstico Cardiopulmonar, Modelos Mezclados Gaussianos (GMM), Modelos Ocultos de Markov (HMM), Transformación de Hilbert, Detección de Actividad de Voz (VAD).

Subdivisão Guiada da DAV Prolongada dos sons Cardiopulmonares

***Resumo*—**A ausculta cardiopulmonar é um procedimento de diagnóstico que tem uma tarefa difícil e os componentes da frequência cardíaca e dos sons pulmonares se sobrepõem. Havia muitas abordagens para quantificar as características desses sinais, e uma das mais recentes é a detecção de atividade de voz (VAD) e modelos de mixagem gaussiana (GMM). Considerando os ruídos pulmonares e cardíacos como eventos acústicos, este artigo propõe uma nova metodologia para avaliar esses indicadores diagnósticos. Aqui, um novo VAD baseado em GMM (VAD-GMM) foi aplicado para detectar e extrair os principais eventos no som pulmonar e cardíaco. Os resultados do VAD-GMM foram comparados com outra metodologia do VAD com base na abordagem estatística, e verificou-se que o VAD-GMM fornece resultados mais definitivos. Como os coeficientes de frequência de mel cepstral (MFCC) e os vetores característicos do quartil já eram bem-sucedidos no reconhecimento de padrões, o VAD-GMM realizou o uso desse tipo de vetores acústicos. Portanto, esse método pode adicionar uma transição da ausculta qualitativa tradicional para avaliação quantitativa e diagnóstico computadorizado assistido, identificando indicadores acústicos anormais. O diagnóstico computadorizado promete ser um método mais eficiente do que os métodos tradicionais, limitados pela audição e experiência de um profissional médico.

***Palavras-chave*—**Diagnóstico Cardiopulmonar, Modelos Gaussianos de Mixagem (GMM), Modelos Ocultos de Markov (HMM), Transformação de Hilbert, Detecção de Atividade Vocal (DVA).

I. Introduction

Practically most initial cardiopulmonary evaluations encompass auscultation but up to now involving mostly descriptive terms rather than quantitative characterization [1]. In particular, cardiopulmonary auscultation is a challenging task as the frequency components of the heart and lung sounds overlap. The S1 and S2 are the two dominant components of the heart sounds. The S3 and S4 are not as easily detectable due to their magnitude, with S3 which could be normal in children, pregnancy, or well fit persons, while S4 very often is indication of abnormality. The inspiratory and expiratory phases of respiration cycles could also contain very useful diagnostic indicators, and for example indicate presence of wheezes or other abnormal sounds [1]. Different authors [2-6] propose diverse feature extraction methods of Lung Sounds (LS) and Heart Sounds (HS) [2]. The HS extraction was improved, applying Hilbert transform. Also, Heron's formula was used to obtain S1 and S2 components from the signal. Some authors [3, 4], propose methods of Voice Activity Detection (VAD) and Mel Frequencies Cepstral Coefficients (MFCC) to extract important events from LS. In general, the main drawback is the noise (environmental noise, cardiac noise or pulmonary noise depending on the diagnostic objective) [3].

The idea behind VAD is to find segments in a signal, which contain diagnostically useful information and at the same time avoid segments associated with silence, or background noise without useful information. Originally, this idea was supported by energy and zero-crossing principles, since the voice segments have more energy and less zero-crossings than the noise segments [7]. Later works applied statistical principles [8], taking into account that HS and LS are sounds containing information useful to discover some abnormalities. Therefore, utilization of VAD could contribute to improve diagnosis. In particular this work is focused on a novel method encompassing a computerized detection of inhalation and exhalation related acoustic events based on Gaussian Mixture Models (GMM).

Further filtering techniques allow separation and extraction of another extraneous sounds such as for example snoring which can be present during respiration [9]. The need of documenting diagnostic indicators and general basic infrastructure norms should be also taking into account a clinical setting [10, 11].

Unfortunately, auscultation with a commonly used traditional stethoscope presents several challenges, such as the presence of environmental noise and the overlap of the HS and LS frequency components. As a result, the perception of cardiac sounds is limited due to the hearing capability and experience of the medical practitioner. Also, these sounds may include frequency components and intensity levels outside of the human auditory range. For these reasons, it is difficult to diagnose the existence of certain abnormalities [12]. Therefore, a system that does not depend on human hearing, and which can detect and classify cardiopulmonary sounds utilizing automated computerized methods, would significantly contribute to improved diagnosis.

Some related diagnostic approaches are focused on endemic diseases, where the acoustic characteristics of cough and crepitation are used to compute vectors of MFCC, applying Wavelets [13].

In other approach, the cardiac sounds envelope is extracted by the Hilbert-Huang Transform (HHT), and the cardiac sounds are segmented by the double-threshold method [16]. In another study [18], a localization method for S1 and S2 is suggested, which is based on an algorithm involving frequency filtering, energy detection, and interval duration. The accuracy of the localization was evaluated by comparing the algorithm with a localization method based on traditional Hilbert transform (HT) [18].

In [19], a computerized method for segmentation and analysis of peak detections in HS patterns is proposed, with emphasis on the characteristics of HS envelopes and taking into account the properties of the Hilbert Transform. Through the use of MFCC as well as applying the VAD, the most important characteristics of the events are obtained [20]. Besides that, some authors propose the extraction of main characteristics by the Fast Fourier Transform (FFT), to carry out classification [21].

Here is proposed to detect events of S1, S2 in HS signals with the presence of S3 and S4. Applying Hilbert transform allows the detection of extreme points (maximum and minimum). In addition, with the support of VAD techniques, which are based on GMM models, computerized extraction is performed on LS and HS signals.

II. S1, S2, S3 and S4 Sounds

The heart sounds are composed of two main sounds S1 and S2 and on occasions there are two more signals identified as S3 and S4 which can be present in normal subjects or reflect pathological conditions. The first sound S1, and the second sound S2, are produced by opening the atrioventricular valves and the closure of the semilunar valve, respectively and vice versa. The sounds S3 and S4 occur at the end of S2 due to the vibration of the blood flow inside the ventricles, the fourth sound S4 is just before S1 due to the contraction of the atrium [22]. Table I summarizes the most relevant characteristics of HS regarding duration, frequency and other characteristics, Table I was obtained based on papers [23-25], and measurements carried out on signals.

III. Methodology

The fundamental concepts of the pre-processing, modeling and characteristics of signals encompassing lung sounds (LS) and heart sounds (HS) in presented experiments are explained in this section.

A. Hilbert Transform

When a signal is evaluated either in time or frequency, the real and the imaginary parts in the other domain are linked by the Hilbert transform [6] [44]. Formally, the Hilbert transform is defined as the convolution of f(t) with the function -1/πt:

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

Hilbert transform does not change the amplitude spectrum, only shifts π/2 for positive frequencies and -π/2 for negative frequencies. If one writes a complex function in the following way:

$$g(t) = f(t) + iHT\{f(t)\} \qquad (2)$$

The envelope E(t) of a function f(t) is defined, as the module of its analytic function:

$$E(t) = |g(t)| = \sqrt{f(t)^2 + HT\{f(t)\}^2} \qquad (3)$$

B. Acoustic Vectors (MFCC and Quartiles), GMM Modelling and VAD

In MFCC acoustic vectors, the sounds are parameterized by implementing a pre-emphasis with FIR filters, followed by a Hamming window applied to each analyzed frame [26-29]. In this project, the experiments were carried out using 50 ms (LS signals) and 130 ms (HS signals) Hamming windows with a 50% shift for both signals, to which the Fast Fourier Transform (FFT) was applied; subsequently, the module was obtained and then multiplied by a filter bank whose frequency range and central frequencies were distributed per the Mel or Bark scale.

Table 1. Main Characteristics In HS Sounds

Sound	Auscultation point	Freq.	Characteristics	Duration	Form of auscultation
S1	Mitral with greater intensity than the tricuspid	30- 120 Hz	Caused by systole	*0.08 – 0.16 sec (0.14 sec)*	Stethoscope diaphragm
S2	Mitral	70-150 Hz	By the aortic valve closure	*0.06 – 0.12 sec. (0.11 sec)*	Stethoscope diaphragm
S3	Mitral	27-70 Hz	Diastole due to ventricular dysfunction	*0.04 - 0.08 sec*	Stethoscope bell
S4	Mitral	10-50 Hz	Auricular noise due to voltage in valves	*0.03 – 0.06 sec*	Stethoscope bell

This was followed by calculating a log stage of the values previously obtained from each filter and subsequently the Inverse Fourier Transform. As the outcome a feature vector called MFCC was obtained [4, 30, 31]. Feature vectors called MFCC were applied [4, 30, 31] with 13 coefficients per vector.

Other experiments were based in quartiles, in this case the duration of the phase of inhalation (~1.5 s.) and the phase of exhalation (~2.5 s.) for the most LS signals used [32]. In quartiles vectors, each frequency value f0.25,..., f0.75 corresponds to its respective quartile coefficient as shown below [32].

$$A_{0.25} = \int_{-\infty}^{f_{0.25}} F_N(f)df \,, \ldots, A_{0.75} = \int_{-\infty}^{f_{.75}} F_N(f)df \quad (4)$$

A Gaussian mixture model (GMM) is a probabilistic model which states that all generated data points are derived from a mixture of a finite Gaussian distributions that has no known parameters. The parameters for Gaussian mixture models are computed either with the maximum a posteriori estimation or the iterative expectation-maximization algorithm. Mathematically, GMM are a weighted sum of component Gaussian densities. GMM are used in biometric systems where the parametric model helps in understanding the behavior of experiment or event.

In our experiments GMM modeling uses the expectation–maximization (EM) algorithm to train the models $\Lambda_i = \{m_i, \vec{\mu_i}, \Sigma_i\}$. The average $\vec{\mu}$ represents the average of all vectors, while the $\sum_i$ covariance matrix models the variability of characteristics of an acoustic class [33]. In equation (5), $\vec{x}$ it is an MFCC or Quartile vector, while b_i ,$\forall_i = 1, \ldots, M$ are the weights of each density in the model [27, 30].

$$p(\vec{x}|\Lambda) = \sum_{i=1}^{M} m_i b_i(\vec{x}) \quad (5)$$

A Voice Activity Detector (VAD) is used to identify speech presence or speech absence in audio, or in our case sound presence or absence related with LS and HS signals. Basic VAD algorithms are based in energy and zero-crossing rate measures of data frames, but now there are alternative algorithms. In perfectly clean conditions even a simple energy detector will do a perfect task at detecting LS-HS; unfortunately, perfectly clean signals are not possible to get in hospitals or doctor offices. That is why proposed VAD is based on GMM. Normally, a VAD is used to classify voiced and unvoiced parts of speech as well as silence. The features introduced on this work are suited to classify activity (LS or HS), noise and silence.

C. VAD based on GMM Modelling

The common voice activity detection (VAD) algorithm is the VAD Rabiner-Schur algorithm [7,8], but others authors, have contributed in this area [34-37]. GMM was successful classifying voice [38-40]. For the lungs, the events to consider are the inhalation and exhalation; for the heart the S1, S2, S3 and S4 sounds are events of interest. Here, a version of VAD based on Gaussian Mixture Models (GMM) is proposed, this allows us to detect the active segments of interest in the signals.

The sets of HS signals were filtered, centered and bleached. A Butterworth low pass filter of order 7 and a cutoff frequency of 150 Hz was applied. Initially, active signals segments were cut manually from the original corpus, theirs end-points were detected visually from their graphics; a new set of signals that contain only events (S1, S2, S3, S4, or inhalation, exhalation depending on the case) were obtained, and these were called manual segmentation. Manual segmentation means to separate segments visually from the graphics of the cardiopulmonary signals sound. With this new set, acoustic vectors were computed, and then the GMM models corresponding to each class (HS or LS). These GMM models were the base of the VAD method proposed, and to determine activity zones (corresponding to the event) and zones of non-activity (noise or silence).

In the detection with GMM of the signal, a value of 1 was assigned to activity zones and 0 to silence or noise; this was multiplied by the original signal, by obtaining a new signal composed by activity only.

When the VAD system was applied to the complete signals, MFCC vectors were computed over complete signals, and the proposed VAD determined if each MFCC corresponded to an activity region or to a non-activity (silence or background noise). The signals and their cuts were converted to vectors MFCC for purposes of calculating VAD-GMM models. Since the sixth MFCC component has more energy in inhalation than in exhalation segments [4], the mean of the sixth components of the MFCC vectors was applied as a threshold to distinguish between inhalation and exhalation frames.

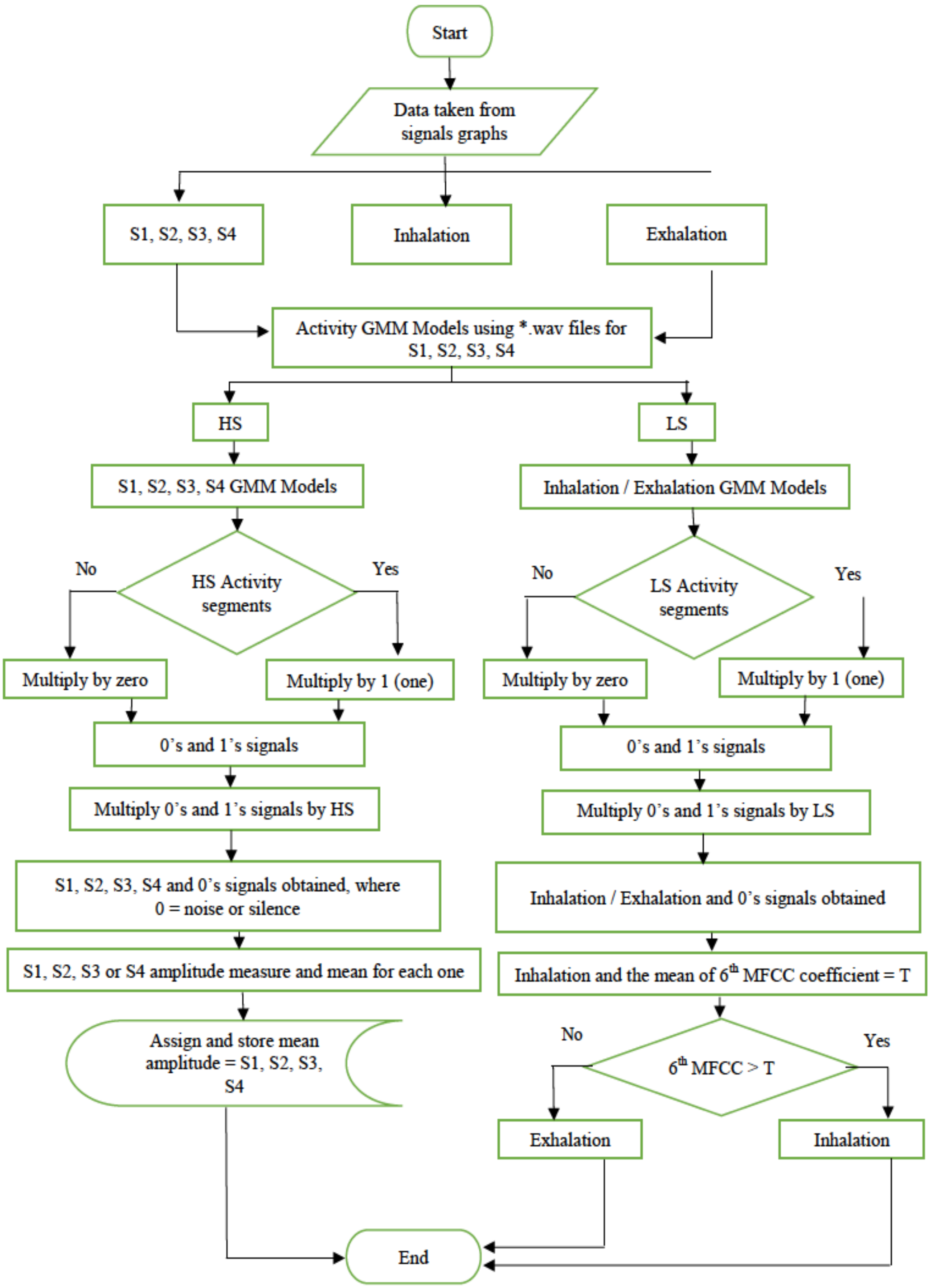

Fig. 1. VAD-GMM System.

D. S1 and S2 detection

Two models based on the VAD-GMM algorithm were applied to the set of acoustic signals to detect and segment automatically S1, S2 and the silence zones. After computerized detection of S1 and S2, the Hilbert envelope of the signal was computed, where the envelope was compared with the original signal (i.e., the normalized signal). Subsequently, the Hilbert envelope was smoothed by a Butterworth filter; after several experiments with each frequency between 7-25 Hz, 8 Hz was selected as the best choice [6]. Since the amplitude for S1 is higher than for S2, it can be used to distinguish between them. The thresholds to differentiate between S1 and S2 is basically the mean of their amplitude. If the amplitude is higher than the threshold is the case of S1, in other case S2. Thresholds based on amplitude are computed to establish the peaks corresponding to S1 and S2. This is accomplished by applying minimums and maximums, considering a minimum as the start of one peak and another minimum as the end of the same peak, it can be extracted automatically. In the algorithm was important to distinguish between systolic and diastolic segments that is why S1 and S2 could be identified by their particular sounds, and allow to determinate which one was the first in the signal recording. The purpose is to separate systole and diastole in pair or unpaired signals when the signal starts with S1 the unpaired segment corresponds to systole, while the pairs correspond to diastole; if the signal starts with S2, the pairs correspond to systole and unpaired correspond to diastole. From these signals, it is possible to obtain the time durations (width) of S1 and S2.

E. Database

Signals from RALE database [45], were filtered with 7.5 Hz band-pass Butterworth filter to suppress any DC offset. Besides, an eight-order Butterworth low-band filter was applied at 2.5 kHz to avoid overlapping; these signals were sampled at 11025 Hz. The LS normal signals from the original RALE database were segmented, theirs end-points were detected visually from their graphics (by the authors); this step was done to obtain only segments of inhalation-exhalation from signals, making a total of 20 inhalations and 20 exhalation recordings in wav format. The HS set signals used for experiments come from [28, 46]. Hence, 20 normal signals were segmented, obtaining 20 sounds for S1 and 20 for S2; the sampling frequency was 11025 Hz. The signals were partitioned in training set and evaluation set. Experimentally, it was found that the time length intervals of the signal phases for LS were 1.5 seconds for inhalation, and 2.5 seconds for exhalation.

Concerning HS, the S1 lasts 0.1 to 0.12 seconds; S2 is between 0.8 to 0.14 seconds [47]. The evaluation was performed by means of leave-one-out method, where a signal is left for evaluation, and using the remaining signals to calculate the model and changing the settings until all possibilities were exhausted.

F. Classification

The cardiac cycle consists of two main acoustic events, the first heart sound, "S1" and the second heart sound "S2". The lung sounds (LS) occurrence is also a cyclical process formed by two main events, inhalation and exhalation. The HS and LS signals have silences between their main events, and both cases are sequences that may vary depending on health conditions, and even the person's mood. Then, these are modeled as Hidden Markov Models (HMM). An HMM is a state-based model, in which each state is characterized by a GMM. HMM is explained exhaustively in [7]; the HMM are expressed as triplets λ= (A, B, π) Fig. 2..

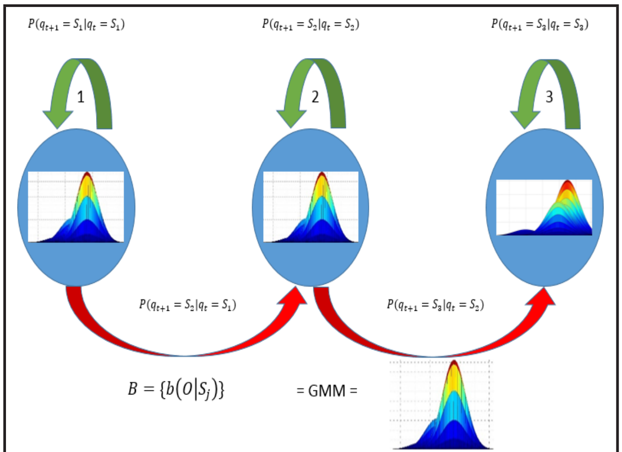

Fig. 2. HMM model with λ= (A, B, π)

Where A is a matrix which gives us the probability of transition from one state to other, B (in our case GMM) gives us the probability of acoustic vector (MFCC or Quartile vector) which was generated from one state, and π is the probability to start in one state. The training for the HMM parameters were computed with the forward-backward algorithm [7].

Since leave-one-out method was applied and the efficiency measured in error rate, the inhalation and exhalation, S1, S2, S3 and S4 signals' sets were used to compute models per class and its evaluation. In leave-one-out method, during each test one signal from the set n (n = 20) is used for evaluation, while the remaining n-1 signals are used to compute the model. Considering n signals, n evaluations are done, but in each evaluation test, the signal's test and the n-1 remaining signals are used to compute the model

IV. Results and Discussion

Once the methodologies described in the previous section are applied, it is important to check if they are really effective. For this purpose, experiments of classification applying Hidden Markov Models (HMM) were carried out. Applying computerized detection with VAD-GMM a graph shown in Fig. 2 is obtained, at the same time this gives us the beginning and the end index of the events extracted from this data. For the first classification experiment a database of 20 inhalations and 20 signals of exhalation was used, which were obtained by computerized detection applying VAD.

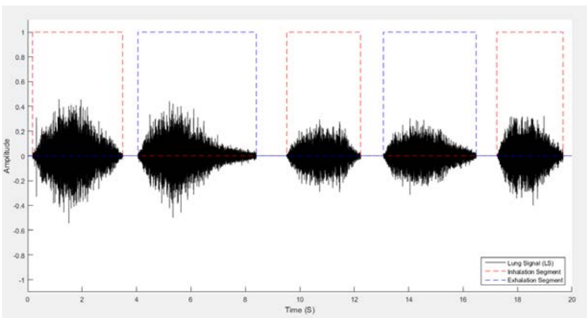

Fig. 3. Computerized segmentation LS with VAD-GMM algorithm

To evaluate the efficiency of this process, it was experimented with different architecture configurations of HMM models, as well as two kinds of acoustic vectors (quartiles and MFCC). The results of classification efficiency are shown in Table 2.

Table 2. Efficiency of LS classification with VAD applying automatic detection LS [4]

# of states	# of Gaussians	Acoustic vectors	Classification efficiency
3	3	Quartiles	77.5%
3	3	MFCC	76.25%
2	3	Quartiles	75%
2	3	MFCC	70%

For the second experiment, the same sets of LS signals that previous experiments were utilized, but in this case VAD-GMM was applied, as shown in Table 3.

Table 3. Efficiency of LS Classification of Automatic Detection in LS with VAD-GMM

# of states	# of Gaussians	Acoustic vectors	Classification efficiency
3	3	Quartiles	85.63%
3	3	MFCC	95%
2	3	Quartiles	88.13%
2	3	MFCC	91.25%

In the third experiment, a set of S1 and S2 signals obtained by computerized detection of HS (Fig. 4) were used. As well VAD-GMM was applied; the best classification result for S1 and S2 was 92.7%. As can be seen in Table 4 both acoustic vectors were used to compute models, with two HMM-GMM architectures configurations and the best results were attained with 3 states and 3 Gaussian by state.

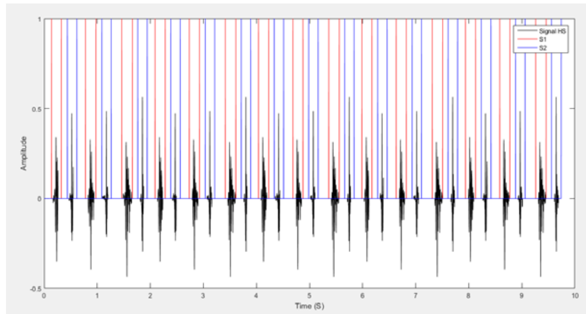

Fig. 4. Computerized segmentation of HS signals with VAD-GMM algorithm

As shown in Table 2, the best classification result what was obtained is 77.5% with VAD (not based on GMM) while with VAD-GMM 95% achieved as shown TABLE III, and this demonstrates the superiority of VAD-GMM. Even in HS events

(S1 and S2) detection and extraction VAD-GMM performed well as shown in the TABLE IV; it means that VAD-GMM can determine to which event belongs a vector and no matter what kind of vector (Quartile or MFCC). GMM utilization improves the capacity of VAD to associate an acoustic vector with its correct class. This is reasonable since in speaker recognition schemes the identification to what class belongs an acoustic vector is common done with GMM. In addition to computerized detection and extraction, VAD-GMM could provide a documented record for long term monitoring and comparative analysis.

Table 4. Efficiency of LS Classification of Automatic Detection of HS With VAD-GMM

# of states	# of Gaussians	Acoustic vector	Classification efficiency
3	3	Quartiles	92.7%
3	3	MFCC	90.38%
2	3	Quartiles	87.31%
2	3	MFCC	91.25%

The results of classification for S1, S2, S3 and S4 were obtained with partitions completed with leave-one-out method and measured in error rate. To measure the efficiency of the VAD-GMM proposed method, classification was made with the events of new signals set obtained by computerized detection. Classification experiments were carried out applying Hidden Markov Models (HMM) on the new set of signals. Applying VAD-GMM computerized detection Fig. 5 and 6 were obtained, this also shows the start and end indexes of each extraction event. HS classification experiments were carried out, a database of 20 signals of S1, S2 and S3; the same experiments were carried out with 20 signals of S1, S2 and S4. In both cases, the signals were extracted automatically with VAD-GMM.

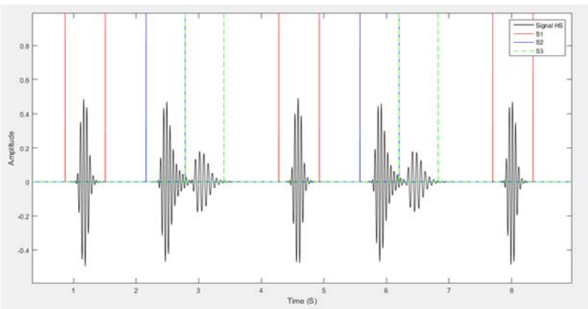

Fig. 5. Detection of events S1, S2 and S3 in signal HS

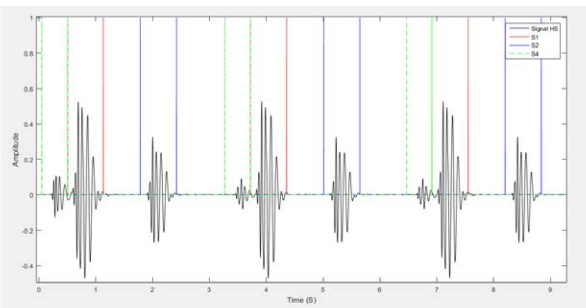

Fig. 6. Detection of events S1, S2 and S4 in signal HS

The best results for S1, S2 and S3 of classification efficiency are shown in Table 5, being an architecture composed of 2 states and 3 Gaussians by state, obtaining up to 96.98% of classification efficiency.

Table 5. Classification Efficiency with VAD Applying Computerized Detection in HS Signals with S1, S2 and S3

# of States	# of Gaussians	Acoustic vector	Classification efficiency
3	3	Quartiles	90.48%
3	3	MFCC	95.40%
2	3	Quartiles	85.08%
2	3	MFCC	96.98%

For the second experiment with HS, the same set of signals was used as in the previous experiment (this time for S1, S2 and S4), obtaining better classification results with a composition of 2 states and 3 Gaussians by state using MFCC vectors, as shown in Table 6.

Table 6. Efficiency of classification with VAD applying computerized detection in HS signals with S1, S2 and S4

# of states	# of Gaussians	Acoustic vector	Classification efficiency
3	3	Quartiles	90.21%
3	3	MFCC	96.11%
2	3	Quartiles	91.34%
2	3	MFCC	97.22%

In order to compare results, the experiments with manual segmentation for LS and HS signals are shown in Table 7. Next, these segments are stored as independent wav files.

Table 7. Classification Efficiency of the Manual Detection of LS and HS with VAD-GMM

Signal	Acoustic vector	# of Gaussians	# of states	Classification efficiency
Insp-Exp	MFCC	3	3	88.84
		2	3	83.07
		2	2	86.15
	Quartiles	3	3	79.23
		2	3	82.30
		2	2	71.93
S1-S2	MFCC	3	3	84.61
		2	3	84.61
		2	2	84.61
	Quartiles	3	3	52.5
		2	3	68.46
		2	2	57.69

First, a set with manual segmentation of the events was obtained, this new set was used to calculate GMM models as the basis of the proposed VAD method. Next, the VAD-GMM was applied to segment LS-HS signals, while evaluating efficiency, HMM models were calculated using the set obtained with the proposed VAD that has been modified using GMM. The results obtained with computerized segmentation indicate that manual segmentation performed by a person, can be substituted with automated outcome as the result of VAD-GMM. In addition, this process is more objective and less dependent on the auditory and visual capacities of a health professional performing auscultation, which is subjective by nature if the observations are only recorded verbally and if the process does not involve technological support as proposed above.

V. Conclusion

The classification efficiency was augmented applying VAD-GMM computerized detection. With the first VAD algorithm, 77.5% of efficiency was achieved, while with VAD-GMM reached 95%.

Computerized detection of events in HS signals was improved by using VAD-GMM technique combined with Hilbert transform. A set of 20 signals (for each of the events) composed of S1, S2, and S3 was obtained by computerized detection; similarly, another set composed of S1, S2 and S4 was evaluated. The classification of signals was carried out applying two sets, one obtained by manual selections, and another by computerized detection. The classification was done with HMM, attaining up to 96.9% efficiency for the sounds of S1, S2 and S3; while for S1, S2 and S4 was 97%.

The 8 Hz edge frequency to smooth the signal envelope could change, due to the sampling frequency. However, model parameters could be better with a broader database; even the methodology would still be valid.

The VAD-GMM application adds to potential transition from qualitative auscultation to quantitative assessment and assisted computerized diagnosis by identifying abnormal acoustic indicators. Diagnosis of these indicators aided by computerized detection could be a more efficient and beneficial than traditional auscultation, which is also hindered by the auditory capability and experience of a medical professional. This method could also be used in general practice or utilized in a nursing home for screening of selected patients.

References

[1]. M. K. Fatma Ayaria, Ali T. Alouanib, "Lung sound extraction from mixed lung and heart sounds FASTICA algorithm " *Electrical Engineering Department, National School of Engineers Tunis (ENIT), Electrical Engineering Department, Tennessee Technological University,* 2012

[2]. F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing Heart Sounds in Respiratory Signals Using Singular Spectrum Analysis", *IEEE Transactions on Biomedical Engineering,* vol. 58, pp. 3360-3367, 2011.

[3]. N. S. K. Jongseo Sohn, Wonyong Sung, "A Statistical Model-Based Voice Activity Detection," *IEEE Signal Processing Letters,* vol. 6, January 1999.

[4]. A. A. a. M. Faezipour, "Acoustic Signal Classification of Breathing Movements to Virtually Aid Breath Regulation," *IEEE Journal of Biomedical And Health Informatics,* vol. 17, March 2013.

[5]. Y. X. H. Wang, and M. Li, "Study on the MFCC similarity-based voice activity detection algorithm," *in Proc. 2nd IEEE Int. Conf. Artif. Intell. Manage. Sci. Electron. Commerce,* pp. 4391–4394, Aug 2011.

[6]. P. B. a. G. S. Ashok M onda, "An automated tool for localization of heart sound components S1, S2, S3 and S4 in pulmonary sounds using Hilbert transform and Heron's formula," *Department of Electronics and Electrical Communication Engineering, IndianInstitute of Technology, Kharagpur, Kharagpur-721 302, India,* 2013.

[7]. L. R. Rabiner and B. H. Juang, *Fundamentals of speech recognition*. Englewood Cliffs, N.J.: PTR Prentice Hall, 1993.

[8]. D. K. Kim and J. H. Chang, "Statistical voice activity detection in kernel space," *J Acoust Soc Am,* vol. 132, pp. EL303-9, Oct 2012.

[9]. A. Azarbarzin and Z. M. K. Moussavi, "Automatic and Unsupervised Snore Sound Extraction From Respiratory Sound Signals," *IEEE Transactions on Biomedical Engineering,* vol. 58, pp. 1156-1162, 2011.

[10]. N. O. M. NOM-016-SSA3-2012, " Que establece las características mínimas de infraestructura y equipamiento de hospitales y consultorios de atención médica especializada.," 8 de enero de 2013.

[11]. P. Forgacs, "Lung sounds," *Br J Dis Chest,* vol. 63, pp. 1-12, 1969.

[12]. M. Lozano, J. A. Fiz, and R. Jané, "Automatic Differentiation of Normal and Continuous Adventitious Respiratory Sounds Using Ensemble Empirical Mode Decomposition and Instantaneous Frequency," *IEEE Journal of Biomedical and Health Informatics,* vol. 20, pp. 486-497, 2016.

[13]. K. Kosasih, U. R. Abeyratne, V. Swarnkar, and R. Triasih, "Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis," *IEEE Transactions on Biomedical Engineering,* vol. 62, pp. 1185-1194, 2015.

[14]. J. Herzig, A. Bickel, A. Eitan, and N. Intrator, "Monitoring Cardiac Stress Using Features Extracted From S1 Heart Sounds," *IEEE Transactions on Biomedical Engineering,* vol. 62, pp. 1169-1178, 2015.

[15]. D. Emmanouilidou, E. D. McCollum, D. E. Park, and M. Elhilali, "Adaptive Noise Suppression of Pediatric Lung Auscultations With Real Applications to Noisy Clinical Settings in Developing Countries," *IEEE Transactions on Biomedical Engineering,* vol. 62, pp. 2279-2288, 2015.

[16]. H. H.-L. CHEN Ji, LUO Liang-Cai, C. Yun, "Automatic Identification Method for the First and Second Heart Sound Based on Double-threshold," *Journal Computer Engineering,* vol. 38, pp. 174-177, 2012.

[17]. K. Bajelani, M. Navidbakhsh, H. Behnam, J. D. Doyle, and K. Hassani, "Detection and identification of first and second heart sounds using empirical mode decomposition," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine,* vol. 227, pp. 976-987, 2013.

[18]. S. D. Min and H. Shin, *A Localization Method for First and Second Heart Sounds Based on Energy Detection and Interval Regulation* vol. 10, 2015.

[19]. S. Sun, Z. Jiang, H. Wang, and Y. Fang, "Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform," *Computer Methods and Programs in Biomedicine,* vol. 114, pp. 219-230, 2014/05/01/ 2014.

[20]. A. Abushakra and M. Faezipour, "Acoustic Signal Classification of Breathing Movements to Virtually Aid Breath Regulation," *IEEE Journal of Biomedical and Health Informatics,* vol. 17, pp. 493-500, 2013.

[21]. N. J. Hupendra Devangan, "A Review on Classification of Adventitious Lung Sounds," *International Journal of Engineering Research & Technology,* vol. 4, 2015.

[22]. A. Mondal, P. Bhattacharya, and G. Saha, "An automated tool for localization of heart sound components S1, S2, S3 and S4 in pulmonary sounds using Hilbert transform and Heron's formula," *SpringerPlus,* vol. 2, p. 512, October 05 2013.

[23]. L. Matias. *La Biblioteca Médica*. Available: www.biblioteca-medica.com.ar

[24]. A. G. Rudnitskiĭ, "Two-Channel Processing of Signals for the Separation of Breath and Cardiac Sounds," *Acoustical Physics,* vol. 47, pp. 353-360, May 2001 2001.

[25]. E. Ç. Güler, B. Sankur, Y. P. Kahya, and S. Raudys, "Two-stage classification of respiratory sound patterns," *Computers in Biology and Medicine,* vol. 35, pp. 67-83.

[26]. D. M. Istrate, «Detection et Reconnaissance des Sons pour la Surveillance Médicale,» These pour obtenir le grade de docteur de l'INPG: spécialité Signal, Image, Parole, Télécoms docteur Institut National Polytechnique de Grenoble, Grenoble, France, 2003.

[27]. P. Mayorga, C. Druzgalski, and J. Vidales, "Quantitative Models for Assessment of Respiratory Diseases," *Pan American Health Care Exchange (PAHCE),* pp. 25 - 30, 2010.

[28]. Boston Children's Hospital [Online]. Available: http://www.childrenshospital.org/

[29]. D. Pearce, "Developing the ETSI Aurora advanced distributed speech recognition front-end and what next?," in *Automatic Speech Recognition and Understanding, 2001. ASRU '01. IEEE Workshop on*, 2001, pp. 131-134.

[30]. P. Mayorga, L. Besacier, R. Lamy, and J. F. Serignat, "Audio packet loss over IP and speech recognition," in *Automatic Speech Recognition and Understanding, 2003. ASRU '03. 2003 IEEE Workshop on*, 2003, pp. 607-612.

[31]. J. S. a. W. Sung, "A Voice Activity Detector Employing Soft Desicion Based Noise Spectrum Adaptation," 1998, pp. 365-368.

[32]. P. Mayorga, C. Druzgalski, O. H. González, and H. S. Lopez, "Modified classification of normal Lung Sounds applying Quantile Vectors," *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 4262-4265.

[33]. F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier*, et al.*, "A Tutorial on Text-Independent Speaker Verification," *EURASIP Journal on Advances in Signal Processing,* vol. 2004, pp. 1-22, 2004.

[34]. G. Lee, S. D. Na, J. H. Cho, and M. N. Kim, "Voice activity detection algorithm using perceptual wavelet entropy neighbor slope," *Biomed Mater Eng,* vol. 24, pp. 3295-301, 2014.

[35]. J. Park, W. Kim, D. K. Han, and H. Ko, "Voice activity detection in noisy environments based on double-combined fourier transform and line fitting," *ScientificWorldJournal,* vol. 2014, p. 146040, 2014.

[36]. F. Lindstrom, K. Ren, H. Li, and K. P. Waye, "Comparison of two methods of voice activity detection in field studies," *J Speech Lang Hear Res,* vol. 52, pp. 1658-63, Dec 2009.

[37]. J. Ramirez, J. M. Gorriz, and J. C. Segura, "Statistical voice activity detection based on integrated bispectrum likelihood ratio tests for robust speech recognition," *J Acoust Soc Am,* vol. 121, pp. 2946-58, May 2007.

[38]. L. Besacier, P. Mayorga, J. F. Bonastre, C. Fredouille, and S. Meignier, "Overview of compression and packet loss effects in speech biometrics," *IEE Proceedings - Vision, Image and Signal Processing,* vol. 150, pp. 372-376, 2003.

[39]. L. Besacier, A. M. Ariyaeeinia, 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 J. Appl. Signal Process.,* vol. 2004, pp. 466-479, 2004.

[40]. S. Jongseo, K. Nam Soo, and S. Wonyong, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters,* vol. 6, pp. 1-3, 1999.

[41]. Martinez W. L and M. A. R, *Computational Statistics Handbook with Matlab*, Second ed., 2008.

[42]. Wendy L. Martinez, Angel R. Martinez, and J. L. Solka, *Exploratory Data Analysis with MATLAB* vol. Second Edition, 2011.

[43]. P. Mayorga Ortiz, C. Druzgalski, M. A. Criollo Arellano, and O. H. González Arriaga, "GMM y LDA aplicado a la detección de enfermedades pulmonares," *Revista Mexicana de Ingeniería Biomédica,* vol. 34, pp. 131-144, 2013.

[44]. A. Mondal, A. K. Kumar, P. S. Bhattacharya, and G. Saha, "Boundary estimation of cardiac events S1 and S2 based on Hilbert transform and adaptive thresholding approach," in *2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT)*, 2013, pp. 43-47.

[45]. RALE, Lung Sounds Repository [Online]. Available: http://www.rale.ca/Repository.htm

[46]. Texas Heart Institute [Online]. Available: http://www.texasheartinstitute.org/AboutUs/index.cfm

[47]. J. D. Echeverry, A. F. López, and J. F. López, "Reconocimiento de valvulopatías cardíacas en señales de fonocardiografía empleando la transformada Gabor," *Scientia Et Technica,* vol. XIII, pp. 139-144, 2007.