Computers in Biology and Medicine 39 (2009) 824–843

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes

Mohammed Bahoura*

Department of Engineering, University of Quebec at Rimouski, 300, allée des Ursulines, Rimouski, Que., Canada G5L 3A1

ARTICLE INFO

Article history:
Received 14 July 2007
Accepted 26 June 2009

Keywords:
Respiratory sounds
Wavelet transform
Mel-frequency cepstral coefficients
Linear predictive coding
Vector quantization
Gaussian mixture models
Multi-layer perceptron
Receiver operating characteristic
Statistical significance
McNemar's test

ABSTRACT

In this paper, we present the pattern recognition methods proposed to classify respiratory sounds into normal and wheeze classes. We evaluate and compare the feature extraction techniques based on Fourier transform, linear predictive coding, wavelet transform and Mel-frequency cepstral coefficients (MFCC) in combination with the classification methods based on vector quantization, Gaussian mixture models (GMM) and artificial neural networks, using receiver operating characteristic curves. We propose the use of an optimized threshold to discriminate the wheezing class from the normal one. Also, post-processing filter is employed to considerably improve the classification accuracy. Experimental results show that our approach based on MFCC coefficients combined to GMM is well adapted to classify respiratory sounds in normal and wheeze classes. McNemar's test demonstrated significant difference between results obtained by the presented classifiers ($p < 0.05$).

1. Introduction

The conventional method of auscultation with stethoscope provides the physician with useful information for diagnose respiratory disorders. However, it is a subjective method that depends on experience, ability and auditory perception among physicians to differentiate sound patterns. This process is also influenced by the variability in their verbal description. To overcome these limitations, quantitative techniques for an objective assessment of lung sound have been developed over the past two decades.

Respiratory sounds can be divided into normal and abnormal (or adventitious) categories according to their acoustic properties. Normal sounds include sounds generated in healthy lungs and airways by normal spontaneous breathing, that can subdivided on tracheobronchial and vesicular sounds, depending on the location where they are recorded (at the trachea or over the thorax) [1,2]. Adventitious sounds are divided into continuous and discontinuous sounds depending on their duration. Continuous sounds are further subdivided into wheezes and rhonchi. Discontinuous adventitious breath sounds are similarly divided into fine and coarse crackles. According to the earlier definition of the ATS (American Thoracic Society), sounds are considered "continuous" if their duration is longer than 250 ms; otherwise they are considered "discontinuous" [2]. The ATS defines wheezes as high-pitched continuous sounds (dominant frequency above 400 Hz) and rhonchi as low-pitched continuous sounds (dominant frequency of 200 Hz or less). According to the new definition of CORSA (Computerized Respiratory Sound Analysis) guidelines, the dominant frequency of wheeze is usually above 100 Hz and the duration greater than 100 ms [2].

Wheezes are continuous adventitious sounds, which are superimposed on normal breath sounds and often associated with bronchial airway obstruction. There are many circumstances leading to wheezing. They include all mechanisms narrowing airway caliber such as bronchospasm, mucosal edema, external compression by a tumor mass, or dynamic airway obstruction [3]. Thus, wheezes can be heard in several diseases, not only in asthma [2]. Regarding clinical applications of the wheezes analysis, previous studies have shown that loudness or intensity of wheezing sounds, monophonic versus polyphonic, inspiratory or expiratory timing did not correlate with airway obstruction [4]. The best correlations were the proportion of the respiratory cycle occupied by wheeze (T_w/T_{tot}), and the frequency content of the highest pitched wheeze.

Fig. 1 represents a typical pattern recognition system including two blocks: feature extraction and modeling/classification. They operate in two steps: training and testing. During the training step, each class data is modeled then a discriminant is determined to delimitate these classes. During the testing step, the new data are classified using the discriminant. A survey of literature from the last two decades shows that Fourier transform (FT) [5–8],

* Tel.: +1 418 723 1986.
E-mail address: Mohammed_Bahoura@uqar.qc.ca.

0010-4825/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2009.06.011

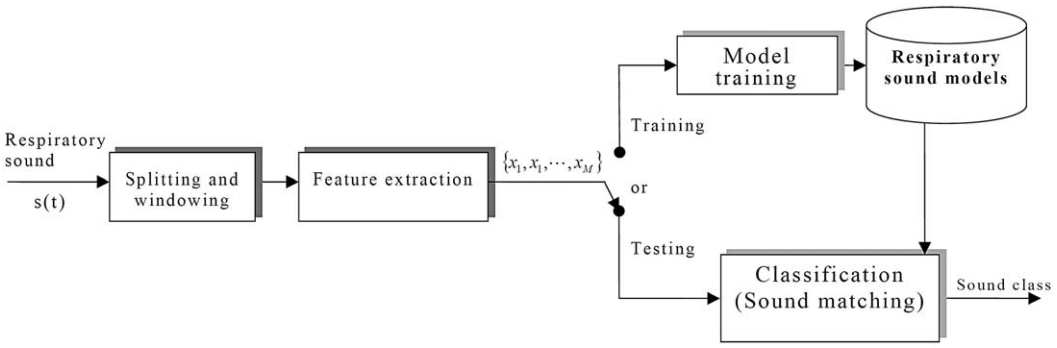

Fig. 1. Block diagram of a typical respiratory sounds classifier.

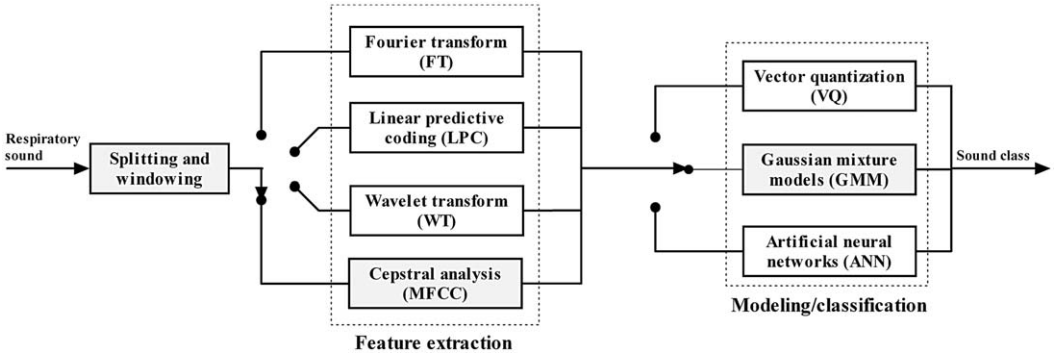

Fig. 2. Respiratory sounds classifiers obtained through various combinations of feature extraction and classification methods.

autoregressive modeling (AR) [9], wavelet transform (WT) [10–12] and Mel-frequency cepstral coefficients (MFCC) [13–15] have been used for feature extraction; while artificial neural networks (ANN) [5–8,11,12], *k*-nearest neighbor (*k*-NN) [9], vector quantization (VQ) [13] and Gaussian mixture models (GMM) [14,15] have been used for respiratory sounds classification.

In this paper, we present the feature extraction methods and the modeling techniques which have been proposed to classify respiratory sounds. We compare various combinations of these methods using receiver operating characteristic (ROC) curves. As illustrated in Fig. 2, we evaluate the feature extraction methods based on FT, LPC, WT and MFCC in combination with the classification methods based on VQ, GMM and ANN. Also, we propose the use of an optimized threshold to discriminate the wheezing class from the normal one. Finally, a post-processing filter is suggested to considerably improve the classification accuracy.

2. Feature extraction

Feature extraction is a process that transforms originally high-dimensional vectors into lower dimensional vectors. It is a mapping $f : \mathbb{R}^N \rightarrow \mathbb{R}^D$, where $D \ll N$. Feature extraction can be considered as a data reduction process that attempts to capture the essential characteristics of the analyzed signal with a small data rate. This operation is described in Fig. 3, where the analyzed signal is segmented in frames using short and overlapped window functions.

In this paper, we will briefly describe some feature extraction methods based on the Fourier transform, the linear predictive coding, the cepstral analysis, and the wavelet transform.

2.1. Fourier transform

The short-time Fourier transform (STFT) of a discrete-time signal $s[n]$ is defined as

$$S[m,k] = \sum_n s[n]w[n-m]e^{-j2\pi nk/N} \quad (1)$$

where $w[n]$ is a short-time windowing function of size L, centered at time location m and N is the number of discrete frequencies. Since the Fourier transform is a complex function, we use the power spectrum density (PSD) given by

$$P_s[m,k] = \frac{1}{N}|S[m,k]|^2 \quad (2)$$

At the sampling frequency f_S, each windowed signal (frame) is represented by N-points power spectrum covering the frequency range $[-f_S/2, f_S/2[$. The power spectrum cannot be used directly as feature vector because it contains a large size data (N/2 components).

To classify asthmatic breath sounds, Rietveld et al. [6] compute the power spectra from shorter intervals of approximatively 3 s, representing a full breathing cycle. The frequency range 100–1300 Hz is divided into 26 bands of approximately 46 Hz each. The feature vector of 26 components is obtained by averaging its power spectrum in each band. In this approach, noted R-FT method, the feature vector is defined by

$$\mathbf{x} = [\overline{P}_1, \overline{P}_2, \ldots, \overline{P}_{26}]^T \quad (3)$$

where $\overline{P}_k$ is the average power spectrum in the k th band and T is the transpose operation.

Fig. 4 gives an example of power spectrum of breath sounds obtained from both healthy and asthmatic subjects. There are significant differences between spectra in the upper frequency.

2.2. Linear predictive coding

Linear predictive coding (LPC), also called auto-regressive (AR) modeling, is widely used in signal processing and specially in speech processing applications. The linear prediction model assumes that each sample can be approximated by a linear combination of a few

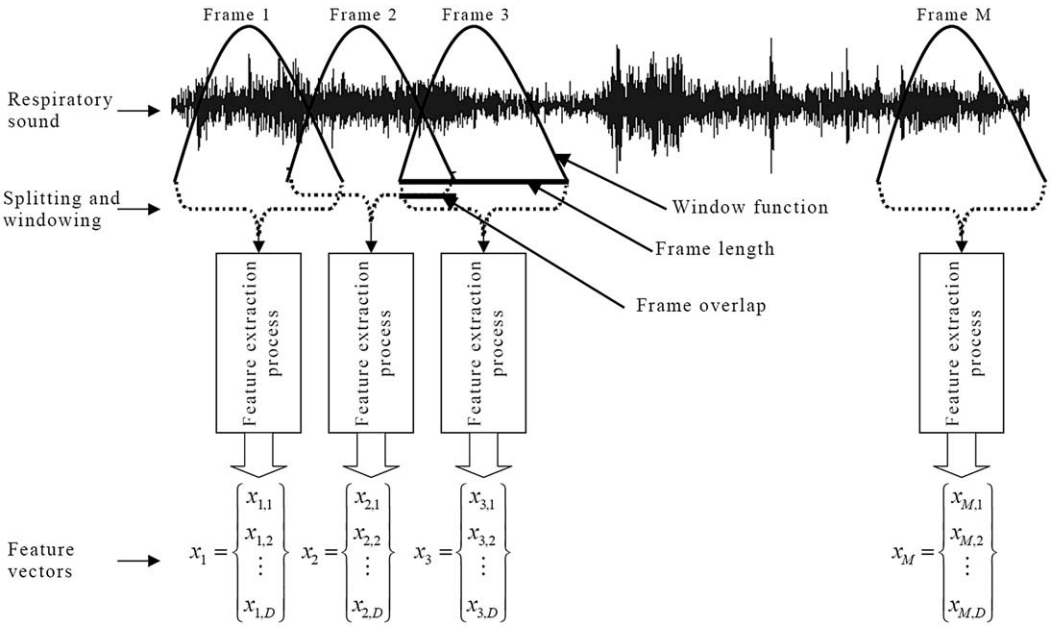

Fig. 3. Block diagram of a typical feature extraction process. The analyzed signal is segmented in frames using short and overlapped window functions. Each frame is then characterized by a reduced size feature vector.

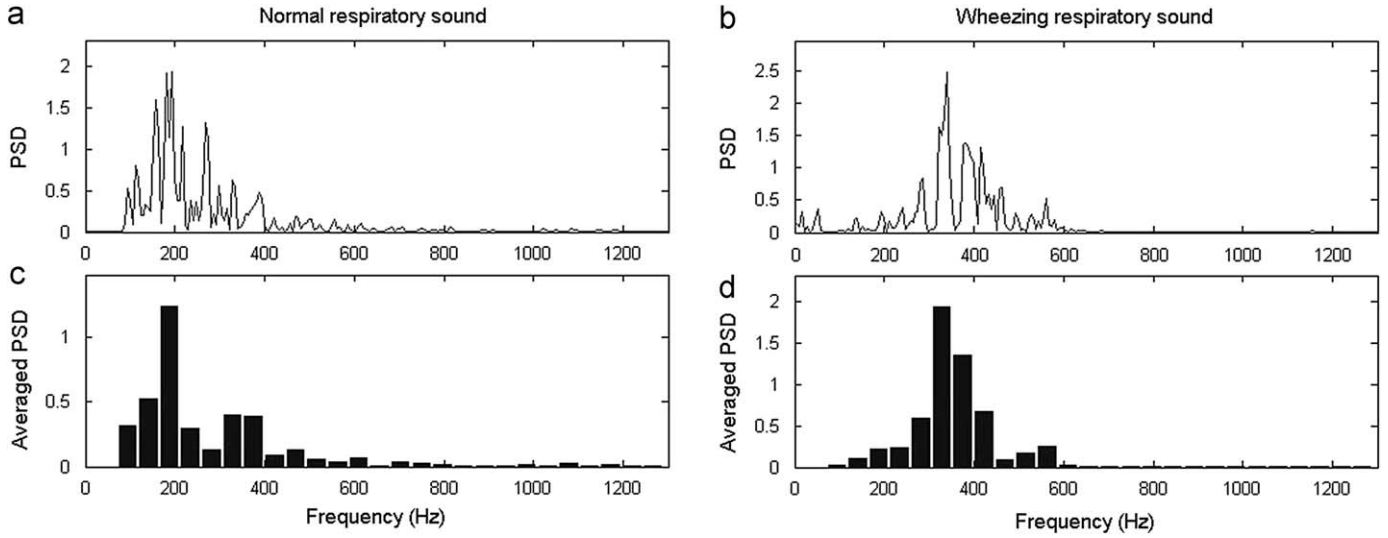

Fig. 4. Power spectrum density (PSD) of breath sound obtained from: (a) healthy subject and (b) asthmatic subject. The respective averaged PSD (c) and (d) are obtained by dividing the frequency range 100–1300 Hz into 26 bands of approximatively 46 Hz each [6].

past samples:

$$\hat{s}[n] = \sum_{k=1}^{p} a_k s[n-k], \quad n = 1, 2, \ldots, N \tag{4}$$

where $\hat{s}[n]$ is the prediction of the signal $s[n]$, and the vector $a = [a_1, \ldots, a_p]$ is the coefficients vector of a predictor of order p. The prediction error $e[n]$ for n th sample $x[n]$ is given by the difference between the actual sample and its predicted value:

$$e[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k] \tag{5}$$

The predictor coefficients a_k are solved by minimizing the mean-square value of the estimation error.

Sankur et al. [9] used a 6th order AR-model in combination with a k-nearest neighbor (k-NN) classifier in order to distinguish pathological respiratory sounds from healthy respiratory sounds. In this approach, named S-AR method, the feature vector is constructed from the LPC coefficients a_k and mean-square prediction error ε_p:

$$\mathbf{x} = [a_1, a_2, \ldots, a_p, \varepsilon_p]^T \tag{6}$$

2.3. Cepstral analysis

Initially [16], the cepstrum analysis was motivated by the need to separate the signal content of glottal speech $s[n]$ excitation from that of vocal tract response $h[n]$:

$$y[n] = s[n] * h[n] \tag{7}$$

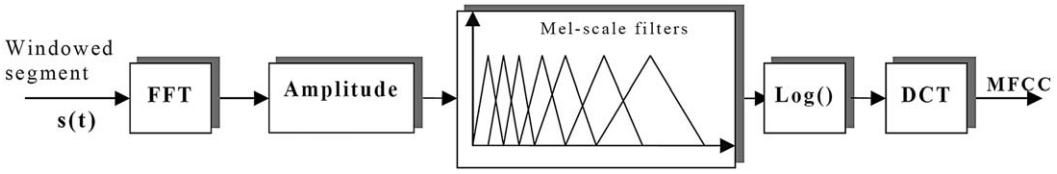

Fig. 5. Block diagram for Mel-Frequency Cepstral Coefficient (MFCC) feature extraction.

In the frequency domain, the convolution relationship becomes a multiplication relationship:

$$Y(e^{j\Omega}) = S(e^{j\Omega})H(e^{j\Omega}) \tag{8}$$

It is very difficult to separate these signals in the time or frequency domain without any knowledge of $h[n]$ or $s[n]$. However, the logarithmic transformation in the frequency domain will lead to $h[n]$ and $s[n]$ separation:

$$\log(Y(e^{j\Omega})) = \log(S(e^{j\Omega})) + \log(H(e^{j\Omega})) \tag{9}$$

For most signal processing applications, we require only the amplitude spectra. Eq. (9) is reformulated as

$$\log(|Y(e^{j\Omega})|) = \log(|S(e^{j\Omega})|) + \log(|H(e^{j\Omega})|) \tag{10}$$

The cepstral coefficients are given by the inverse transformation to time domain:

$$c[n] = DFT^{-1}\log(|S(e^{j\Omega})|) + DFT^{-1}\log(|H(e^{j\Omega})) \tag{11}$$

The cepstral coefficients $c[n]$ represent the deconvolution result. In the cepstral domain, lower coefficients describe the envelope structure (vocal tract response) and the higher coefficients the harmonic structure (glottal speech excitation). These coefficients have been previously used to study the generation and transmission process of the respiratory sound [17].

One of the most important features used among various kinds of speech applications is the Mel-frequency cepstral coefficients (MFCC). This approach is motivated by the need to separate the signal content of glottal speech excitation from that of vocal tract response [16]. The Mel-cepstrum exploits auditory system principles, as well as the decorrelating property of the cepstrum.

The MFCC computation is illustrated in Fig. 5. The spectrum $Y[k]$ of the windowed waveform is computed using the discrete Fourier transform (DFT). The magnitude of $Y[k]$ is weighted by the series of the L filters frequency responses. The real cepstrum associated with the resulting energies $E[k]$ is referred to as the MFCC and is computed for the given frame as

$$C[n] = \sum_{k=1}^{L} \log(E[k]) \cos\left(n(k-0.5)\frac{\pi}{M}\right) \tag{12}$$

where $n = 0, 1, \ldots, M$ and M is the number of desired MFCC coefficients.

Recently, we have proposed the cepstral analysis to characterize the respiratory sounds [13–15]. In this approach, named B-MFCC, the feature vector is constructed from MFCC coefficients:

$$\mathbf{x} = [C[1], C[2], \ldots, C[M]]^T \tag{13}$$

where M is the number of MFCC coefficients.

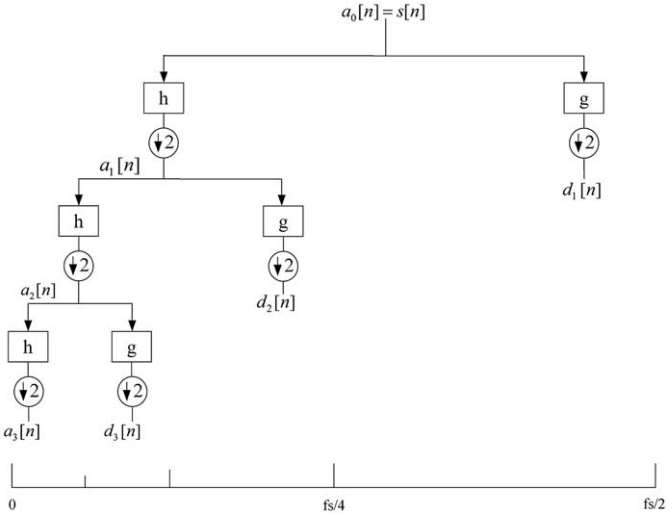

Fig. 6. The 3-level wavelet transform decomposition tree. Each unit consists of a low-pass (h) and high-pass (g) filter pair, followed by a decimation (↓). The frequency range is divided according to the dyadic scale.

2.4. Wavelet transform

The continuous wavelet transform (CWT) of a signal $s(t)$ is defined by

$$CWT_s^{\psi}(\lambda, \tau) = \int_{-\infty}^{+\infty} s(t)\psi_{\lambda,\tau}^{*}(t)\,dt \tag{14}$$

where $(*)$ denotes the complex conjugate, $\lambda \in \mathbb{R}^+$ represents the scale parameter, and $\tau \in \mathbb{R}^+$ represents the time translation. The function $\psi_{\lambda,\tau}^{*}$ is obtained by shifting and scaling the prototype wavelet $\psi(t)$, and is defined as

$$\psi_{\lambda,\tau}(t) = \frac{1}{\sqrt{\lambda}}\psi\left(\frac{t-\tau}{\lambda}\right) \tag{15}$$

This transform is often discretized in λ and τ on a dyadic grid, with the time remaining continuous. In general, the scale factor λ is chosen 2^j in order that the frequency axis be decomposed in octaves. The time translation τ is chosen as $n2^j$:

$$w_j[n] = 2^{-j/2}\int_{-\infty}^{+\infty} s(t)\psi(2^{-j}t - n)\,dt \tag{16}$$

where $j \in \mathbb{Z}$ represents the level and $n \in \mathbb{Z}$ the time shift. The discrete wavelet transform (DWT) is implemented using a complementary low-pass/high-pass filtering followed by down-sampling by 2. As represented in Fig. 6, this operation is recursively applied to the low-pass branch [18]. The outputs of the high-pass filters are named details $d_j[n]$ and those of the low-pass filters are named approximations $a_j[n]$. They are, respectively, given by

$$d_{j+1}[n] = \sum_m g[m-2n]a_j[m] \tag{17}$$

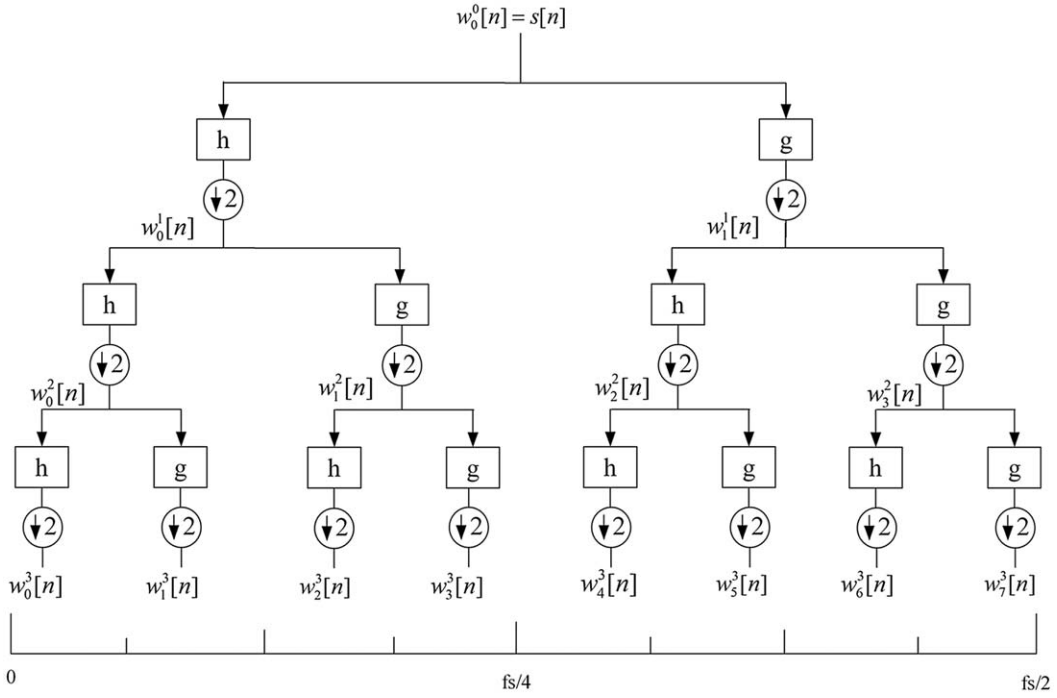

Fig. 7. The 3-level wavelet packet decomposition tree. Each unit consists of a low-pass (h) and high-pass (g) filter pair, followed by a decimation (↓). The frequency band is uniformly divided.

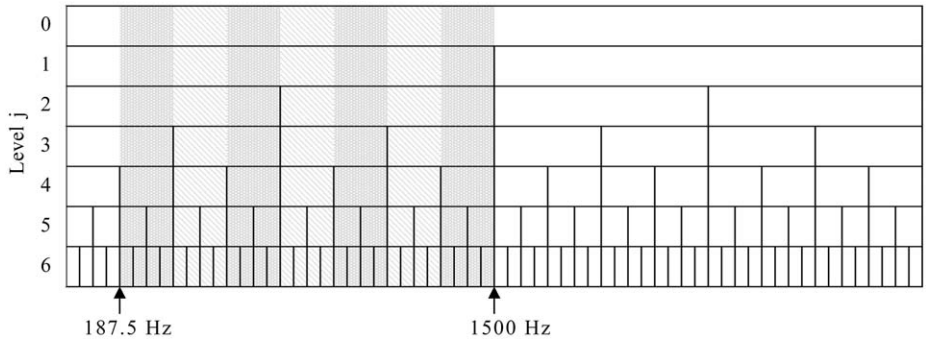

Fig. 8. The seven shaded regions of the wavelet packet decomposition tree are used by Pesu et al. [10] for feature extraction. The frequency bands are obtained with a sampling frequency of 6 kHz.

$$a_{j+1}[n] = \sum_m h[m-2n]a_j[m] \tag{18}$$

Kandaswamy et al. [11] proposed the use wavelet coefficients, combined to a neural network, to classify lung sounds. The respiratory signal is divided into segments before applying the wavelet transform (WT) for each segment. In this work, the respiratory signals are sampled at 11 025 Hz and the WT is computed for $j = 7$. The lung sounds are characterized by the wavelet coefficients of five subbands, d_3, d_4, d_5, d_6 and d_7, which correspond to the frequency band 43.07–1378.13 Hz. In this approach, named K-WT, the feature vector of 19 components is constructed as follows [11]:

- Mean of the absolute values of the coefficients in each subband (μ_{d_i}).
- Average power of the wavelet coefficients in each subband (p_{d_i}).
- Standard deviation of the coefficients in each subband σ_{d_i}.
- Ratio of the absolute mean values of adjacent subbands ($\mu_{d_i}/\mu_{d_{i+1}}$).

$$\mathbf{x} = [\mu_{d_i}, p_{d_i}, \sigma_{d_i}, \mu_{d_i}/\mu_{d_{i+1}}]^T, \quad i = 3, \ldots, 7 \tag{19}$$

In the present work, we adapt this algorithm to the sampling frequency of 6000 Hz. The wavelet transform is computed for $j = 6$. In this case, the respiratory sounds are characterized by the five subbands, d_2, d_3, d_4, d_5 and d_6, which correspond to the frequency band 46.87–1500 Hz. The feature vector of 19 components is defined by

$$\mathbf{x} = [\mu_{d_i}, p_{d_i}, \sigma_{d_i}, \mu_{d_i}/\mu_{d_{i+1}}], \quad i = 2, \ldots, 6 \tag{20}$$

The wavelet packet transform (WPT), proposed by Coifman and Wickerhauser [19], can be considered as an extension of the wavelet transform (see Fig. 7). Unlike the wavelet transform, which is obtained by iterating the low-pass branch $a_j[n]$, the wavelet packet transform is obtained by iterating either branch $a_j[n]$ and $d_j[n]$ at any level j. The wavelet packet decomposition is equivalent to a multichannel filtering where the number of filters and their bandwidths are related to the level j. The wavelet packet coefficients $w_k^{j+1}[n]$ corresponding to even and odd subband k are, respectively, given by

$$w_{2p}^{j+1}[n] = \sum_m h[m-2n]w_p^j[m] \quad \text{even subband } k \tag{21}$$

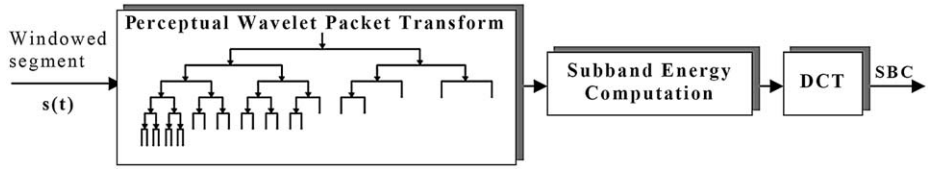

Fig. 9. Block diagram for subband based cepstral (SBC) feature extraction.

$$w_{2p+1}^{j+1}[n] = \sum_m g[m-2n]w_p^j[m] \quad \text{odd subband } k \tag{22}$$

where $p = 0, \ldots, 2^j - 1$ refers to the subband of level j.

Pesu et al. [10] are the pioneers in using the wavelet packet decomposition (WPD), combined with the learning vector quantization (LVQ), to classify the respiratory sounds. In this approach, named P-WPT, the feature vector is constructed using the following algorithm [10]:

- Divide the respiratory signal into segments.
- Calculate the wavelet packet decomposition ($j = 6$) for each segment.
- Search for the best basis in the WPD tree.
- Subdivide the WPD tree into seven regions (Fig. 8).
- Construct a feature vector $\mathbf{x}$ of 14 components that represents the coast ζ and the level ℓ of each region.

$$\mathbf{x} = [\zeta_1, \zeta_2, \ldots, \zeta_7, \ell_1, \ell_2, \ldots, \ell_7]^T \tag{23}$$

With a sampling frequency of 6 kHz, these regions approximately correspond to seven frequency bands between 187.5 and 1500 Hz.

Other feature extraction methods, based on the wavelet transform, can be used to characterize respiratory sounds. In our previous works [14,15], we used subband based cepstral (SBC) parameters that approximate the Mel-scale frequency division [20]. This kind of decomposition is also called the perceptual wavelet packet transform. The block diagram to compute the SBC parameters is given in Fig. 9. The perceptual wavelet packet transform is processed for the given decomposition tree. The energy of each subband is computed and then scaled by the number of coefficients in that subband:

$$E[k] = \frac{1}{N_k} \sum_{m=1}^{N_k} (w_k[m])^2 \tag{24}$$

$w_k[m]$ is the m th wavelet coefficient of the k th subband, k the subband frequency index ($k = 1, \ldots, L$), and N_k the number of coefficients in the k th subband.

Finally, the SBC parameters are derived from subband energies by applying the discrete cosine transformation (DCT)

$$SBC[n] = \sum_{k=1}^{L} \log(E[k]) \cos\left(n(k-0.5)\frac{\pi}{L}\right), \quad n = 0, \ldots, N \tag{25}$$

where N is the number of SBC parameters and L is the number of the frequency bands. In this approach, named B-SBC method, the feature vector is constructed from the SBC coefficients:

$$\mathbf{x} = [SBC[1], SBC[2], \ldots, SBC[N]]^T \tag{26}$$

where T represents the transpose operation.

3. Modeling/classification

In this paper, we describe three classification methods based on the vectorial quantification (VQ), Gaussian mixture models (GMM), and perceptron multi-layer (MLP), respectively. These supervised classification methods require the reconstruction of a classifier, that is, a function that affects a model for each respiratory sound class described by its characteristic vectors. There are two phases to constructing classifier. In the training phase, each respiratory sound class is modeled using a training data, then a discriminant is established to delimit these classes. In the test phase, the unknown sound is analyzed and the best matching model is reached.

3.1. Vector quantization (VQ)

3.1.1. Principle

The vector quantization (VQ) is a process of mapping vectors from a large space to a finite number of regions in that space. Each region is called a cluster and can be represented by its centroid $\mathbf{c_i}$ called *codeword* [13,21]. For a given respiratory sound class, the resulting *codewords* constitute the *codebook* of this class $\mathscr{C} = \{\mathbf{c_1}, \mathbf{c_2}, \ldots, \mathbf{c_N}\}$. There are two phases in this classification system: training and recognition (see Fig. 10). In the training phase, an acoustical model (*codebook*) is constructed for each class of respiratory sound and the models are stored in a database. Thus, there are K *codebooks* $\mathscr{C}_1, \mathscr{C}_2, \ldots, \mathscr{C}_K$ generated for K reference sounds. There are many different algorithms to create a *codebook*. The most extensively used is LBG (Linde–Buzo–Gray) algorithm [22]. In the recognition phase, the distortion between a set of testing feature vectors $X = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}\}$ and each codebook is computed, then an average quantization distortion Q_k to the k th codebook ($\mathscr{C}_k$) is performed [23], according to

$$Q_k = \frac{1}{M} \sum_{i=1}^{M} \min_{1 \le j \le N} d(\mathbf{x_i}, \mathbf{c}_j) \tag{27}$$

where $d(\mathbf{x_i}, \mathbf{c}_j)$ is the distortion measure (usually a Euclidean distance) between the input vector $\mathbf{x_i}$ and a centroid $\mathbf{c}_j$ of the k th codebook ($\mathscr{C}_k$). The unknown sound is then identified as the reference sound with the minimum average distortion measure

$$\widehat{k} = \arg \min_{1 \le k \le K} \{Q_k\} \tag{28}$$

where K is the number of the sound classes. In this study, unknown sound is classified segment-by-segment ($M = 1$).

3.1.2. Two-class classification

To illustrate this, we attempt to classify respiratory sounds in two classes, $k \in \{normal, wheezing\}$. To achieve that, we propose to define a score function $Sc(X)$ by the difference of distorsions:

$$Sc(X) = Q_{normal} - Q_{wheezing} \tag{29}$$

The classification decision is made by comparing the score function $Sc(X)$ to a threshold θ

$$Sc(X) \underset{normal}{\overset{wheezing}{\gtrless}} \theta \tag{30}$$

Eq. (28) is a particular case of Eq. (30), where $\theta = 0$.

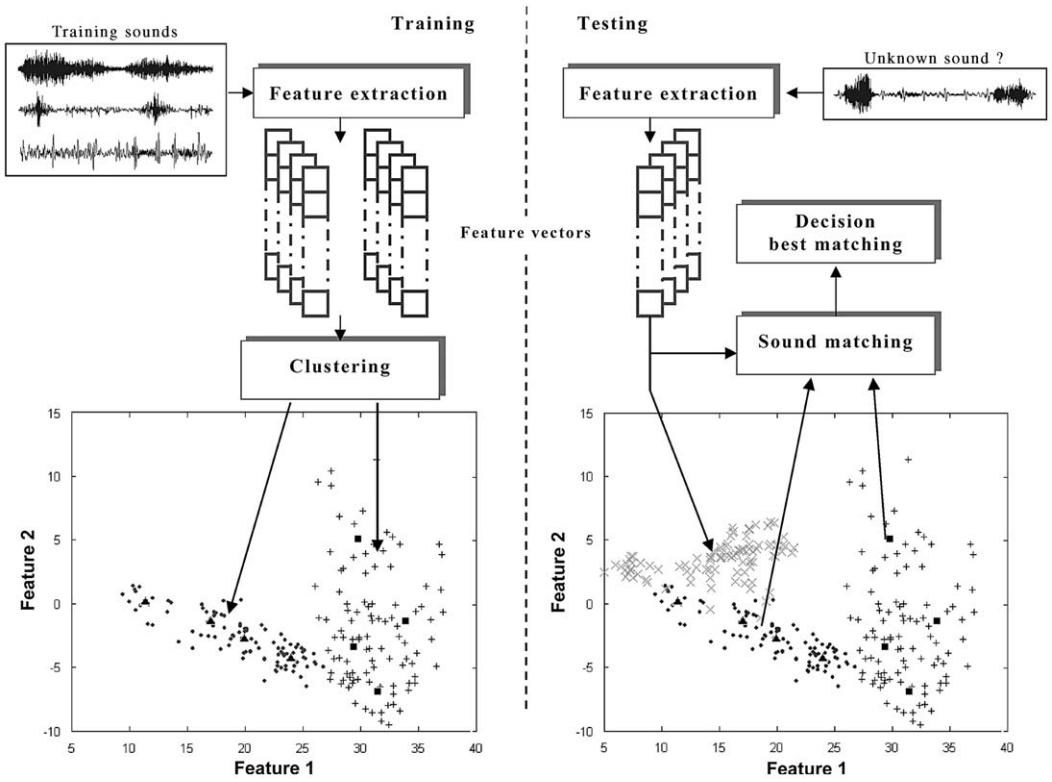

Fig. 10. Block diagram of VQ-based classifier in the feature space [13]. The feature vectors of the *normal* sound (·) are represented by four (▲) centroids, while those of the *wheezing* sound (+) are modeled by four centroids (■). The unknown respiratory sound (×) is compared to the codebook of these classes.

3.2. Gaussian mixture models (GMM)

3.2.1. Principle

A powerful statistical method massively used for speaker identification/verification is the Gaussian mixture models (GMM) [24]. In this approach, the distribution of the feature vectors extracted from respiratory sound class is modeled by a weighted sum of N Gaussian densities:

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{N} w_i p_i(\mathbf{x}) \tag{31}$$

where $\mathbf{x}$ is D-dimensional feature vector or observation vector, λ is the sound class model, $p_i(\mathbf{x})$, $i = 1, \ldots, N$, are the components densities characterized by the mean μ_i and the covariance matrix Σ_i. The mixture weights w_i satisfy the constraint that $\sum_{i=1}^{N} w_i = 1$. Each component density is a D-variate Gaussian function of the form

$$p_i(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_i)'\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right\} \tag{32}$$

The mean vectors, covariance matrix and mixture weights of all Gaussian functions together parameterize the complete Gaussian mixture density. These parameters are collectively represented by

$$\lambda = \{w_i, \mu_i, \Sigma_i\}, \quad i = 1, \ldots, N \tag{33}$$

In the respiratory sound context, each sound class is modeled by a specific GMM that is trained using a set of data with the Expectation-Maximization (EM) algorithm [15,21]. During the test phase, an unknown sound is compared to all the GMM models and the classification decision is based on the maximum likelihood (ML) criterion. Fig. 11 shows an example of respiratory sounds modeling. Assume we have K sound classes represented by the GMM models $\lambda_1, \lambda_2, \ldots, \lambda_K$. The objective is to find the sound model λ_k which maximizes the *a posteriori* probability for a given observation sequence $X = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}\}$:

$$\widehat{k} = \arg \max_{1 \le k \le K} p(\lambda_k|X) \tag{34}$$

Using the Bayes rules, Eq. (34) can be expressed as

$$\widehat{k} = \arg \max_{1 \le k \le K} \frac{p(X|\lambda_k)p(\lambda_k)}{p(X)} \tag{35}$$

Assuming equally likely classes and noting that $p(X)$ is the same for all classes, the classification task can be simplified to finding

$$\widehat{k} = \arg \max_{1 \le k \le K} p(X|\lambda_k) \tag{36}$$

Using the *log* function and assuming independence between observations, the classification rule becomes

$$\widehat{k} = \arg \max_{1 \le k \le K} \sum_{i=1}^{M} \log p(\mathbf{x_i}|\lambda_k) \tag{37}$$

In this study, unknown sound is classified segment-by-segment ($M = 1$).

3.2.2. Two-class classification

In our previous work [15], we used a GMM-based system to classify respiratory sounds into normal and wheezing. Each segment feature is tested and a scoring function is given by the likelihood ratio in the *log* domain [25]:

$$Sc(x) = \log\{p(X|\lambda_{wheezing})\} - \log\{p(X|\lambda_{normal})\} \tag{38}$$

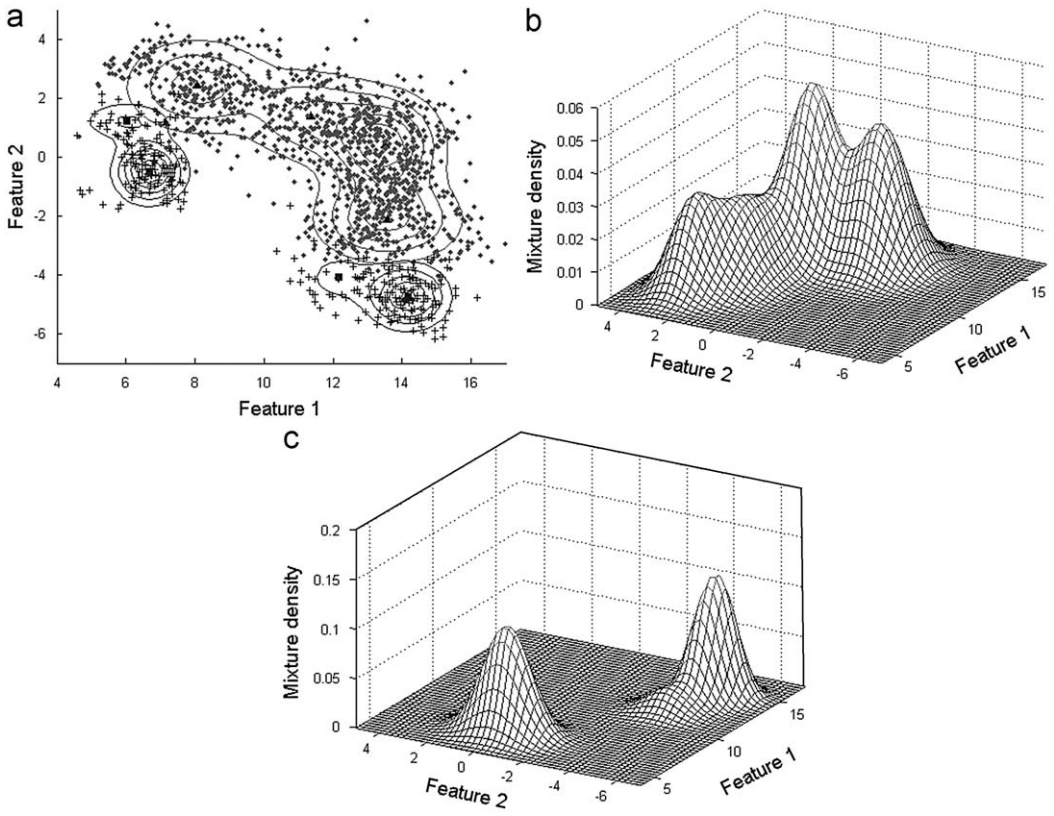

Fig. 11. Example of 4th order Gaussian mixture models (GMM) in the 2-D representation (a), the means of the Gaussian functions modeling the *normal* (·) and the *wheezing* (+) classes are, respectively, represented by the (▲) and (■) symbols. The contour lines reflect the variance of the Gaussian functions. The 3-D representations of (b) and (c) give the mixture density of the *normal* and the *wheezing* classes, respectively.

The term $p(X|\lambda_{wheezing})$ is the likelihood that the tested feature vectors X belongs to the wheezing class and $p(X|\lambda_{normal})$ is the likelihood that the tested feature vectors X belongs to normal class. The likelihood ratio is compared to a threshold θ to discriminate the *wheezing* from the *normal* class:

$$Sc(X) \underset{normal}{\overset{wheezing}{\gtrless}} \theta \tag{39}$$

Eq. (37) is a particular case of Eq. (39), where $\theta = 0$.

3.3. Multi-layer perception (MLP)

3.3.1. Principle

The multi-layer perception (MLP) is a collection of neurones (nodes), arranged together in layers in feed-forward manner (Fig. 12). Signals pass into the input layer nodes, progress forward through the hidden layers and finally emerge from the output layer [26]. Fig. 12 represents an example of an MLP network characterized by D inputs, one hidden of N nodes, and K outputs. Each node j, in the hidden layer, receives the output of each node i from the input layer through a connection of weight $w^h_{j,i}$ and then produce a corresponding response z_j which is forwarded to the output layer. In fact, each node j performs a weighted sum which is transferred by a nonlinear function f_h according to

$$z_j = f_h\left(\sum_{i=0}^{D} w^h_{j,i} x_i\right), \quad j = 1, \ldots, N \tag{40}$$

where $x_0 = 1$ is the bias input of the hidden layer.

In the same manner, the output of each node j, in hidden layer, is given by

$$y_j = f_o\left(\sum_{i=0}^{N} w^o_{j,i} z_i\right), \quad j = 1, \ldots, K \tag{41}$$

where f_o is the transfer function, $w^o_{j,i}$ is the connection weight, and $z_0 = 1$ is the bias input of the output layer. In this study, we use an *hyperbolic tangent sigmoid* function $f_h(x) = (1 - e^{-2x})/(1 + e^{-2x})$ for the hidden layer and a *logistic sigmoid* function $f_o(t) = (1 + e^{-x})^{-1}$ for the output layer.

The connection weights $\mathbf{w} = \{w^h_{j,i}, w^o_{j,i}\}$ are determined in the training phase using a set of inputs for which the desired outputs are known $\{(\mathbf{x_1}, \mathbf{d_1}), (\mathbf{x_2}, \mathbf{d_2}), \ldots, (\mathbf{x_p}, \mathbf{d_p})\}$. To accomplish this, the back-propagation algorithm is commonly used. The details of the algorithm can be found in [27]. When the i th pattern (x_i, d_i) is presented, the error of training is defined as [28]

$$E(\mathbf{w}) = \frac{1}{2}\sum_{j=1}^{K}(d_{i,j} - y_{i,j})^2 \tag{42}$$

where $d_{i,j}$ an $y_{i,j}$ are the desired response and the actual response value of the output neurone j. Error is reduced by updating the weights using the gradient descent method:

$$\Delta w_{j,i} = -\eta \frac{\partial E}{\partial w_{j,i}} \tag{43}$$

where $(0 < \eta < 1)$ is the *learning rate*. A small value of η can guarantee convergence but involves a slow learning. On the other hand, a

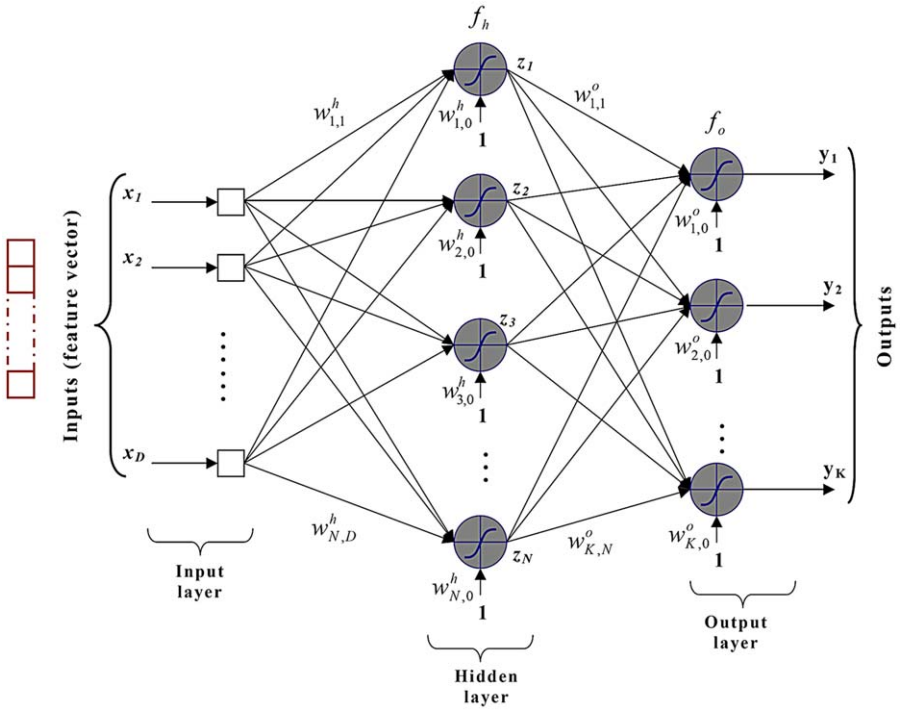

Fig. 12. A typical example of multi-layer perception (MLP). This network is characterized by D inputs, one hidden of N nodes, and K outputs. $w^h_{j,i}$ and $w^o_{j,i}$ are the connection weight of the hidden layer and the output layer, respectively.

large value of this parameter involves a rapid learning but can lead to oscillation or even divergence. To overcome this limitation, many variations of this algorithm have been introduced for training neural networks. In this study, we use RPROP (Resilient propagation) which is one of the faster algorithms [29]. The basic principle of this algorithm is to eliminate harmful influences of the derivatives' magnitude on the weight update. The direction of the weight update is specified by using the sign of the derivative, while the size of the weight change is determined by a separate update value [29].

The backpropagation algorithm can be implemented in two different ways: sequential mode and batch mode [27]. In sequential mode, the weights are updated after each training example (pattern) is applied to the network. In batch mode, all the training examples, that constitute an epoch, are applied to the network before the weights are updated. For a given epoch, a cost function is defined as the average squared error of Eq. (42)

$$E_{av}(\mathbf{w}) = \frac{1}{2P}\sum_{i=1}^{P}\sum_{j=1}^{K}(d_{i,j} - y_{i,j})^2 \tag{44}$$

where P is the number of the training examples (patterns) per epoch and K is the number of the outputs. The synoptic weight $w_{j,i}$, connecting neuron i to neurone j, is updated using

$$\Delta w_{j,i} = -\eta\frac{\partial E_{av}}{\partial w_{j,i}} \tag{45}$$

This training procedure should be repeated until an acceptable error rate is achieved or until a certain number of iterations (Ni) are completed using training examples. In this experiment, we take $\eta = 0.01$, $E_{av} = 0.001$ and $Ni = 5000$. As the connections $w_{j,i}$ are initialized with random values, the learning and testing process is repeated 50 times in order to take the average value.

Unlike VQ and GMM methods that require one model per class, the MLP network needs only one model with a number of outputs equal to the number of the sound classes (K).

In order to classify an unknown sound, the observation sequence $X = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}\}$ is passed through the network yielding the actual output sequence $Y = \{\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_M}\}$. For each output k, the network provides a set of M values $\{y_{i,k}\}$, $i = 1, \ldots, M$. The sound class is identified as the reference sound that corresponds to the largest mean value of the accumulated outputs:

$$\widehat{k} = \arg\max_{1 \le k \le K}\{\overline{y}_k\} \tag{46}$$

where the mean values $\overline{y}_k$ are computed over the M segments:

$$\overline{y}_k = \frac{1}{M}\sum_{i=1}^{M} y_{i,k}, \quad k = 1, \ldots, K \tag{47}$$

In this study, unknown sound is classified segment-by-segment ($M = 1$).

3.3.2. Two-class classification

Fig. 13 gives an example of MLP that we have used to classify the respiratory sounds in two classes $s \in \{normal, wheezing\}$ [15]. This network is characterized by 24 inputs, 30 neurons in the hidden layer and 2 outputs [15]. Obviously, the number of neurons in the input layer must be adjusted to the feature vector length. In this classifier, we propose to define a score function $Sc(X)$ by the difference of outputs:

$$Sc(X) = y_{wheezing} - y_{normal} \tag{48}$$

The classification decision is made by comparing the score function $Sc(X)$ to a threshold θ

$$Sc(X) \underset{normal}{\overset{wheezing}{\gtrless}} \theta \tag{49}$$

Eq. (46) is a particular case of Eq. (49), where $\theta = 0$.

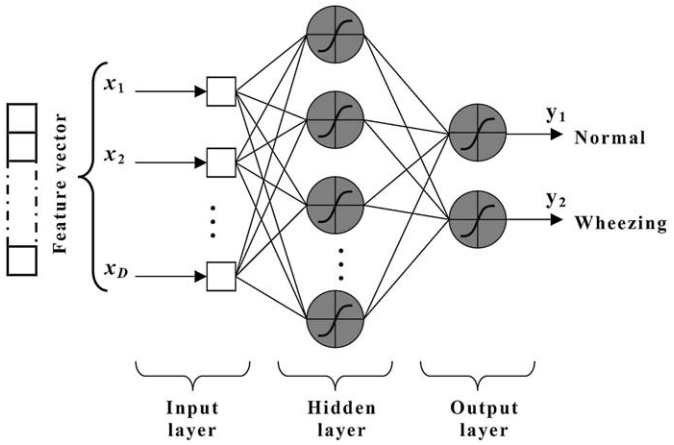

Fig. 13. Neural network structure that we have used to classify respiratory sounds in normal and wheezing classes [15]. This network is characterized by 30 neurons in the hidden layer and two outputs. The number of neurones in the input layer must be adjusted to the feature vector length.

This MLP network can be simplified using only one output [15]. For example, 0 and 1 of the output y correspond to normal and wheezing classes, respectively. In this case, $Sc(X) = y$.

Fourier transform combined to neural networks have been used by Forkheim et al. [5], Rietveld et al. [6], Waitman et al. [7], and Guler et al. [8]. Linear predictive coding combined to k-nearest neighbor method are used by Sankur et al. [9]. Wavelet and neural networks are used by Pesu et al. [10], Dokur et al. [12] and Kandaswamy et al. [11]. Finally, the combination of cepstral analysis and Gaussian mixture models has been proposed by Bahoura et al. [14,15].

4. Evaluation criteria

4.1. Definition

The performances of the classifiers are evaluated by comparing their output to the physicians' diagnosis. The performance of each test is characterized in terms of its ability to identify true positives while rejecting false positives [30]. The true positive ratio (TPR) and the false positive ratio (FPR) are given by the following equations:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{FN} + \mathrm{TP}} \tag{50}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}} \tag{51}$$

where TP, FN, FP, and TN are the number of the true positive, false negative, false positive, and true negative test results, respectively (Fig. 14a). Two other quantities are also used: *sensitivity* and *specificity*, which are defined by Eqs. (52) and (53), respectively:

$$\text{Sensitivity (SE)} = \mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{FN} + \mathrm{TP}} \tag{52}$$

$$\text{Specificity (SP)} = 1 - \mathrm{FPR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{53}$$

Sankur et al. [9], Waitman et al. [7] and Forkheim et al. [5] have used both sensitivity and specificity. Pesu et al. [10] and Lu et al. [21] have also used both sensitivity and positive predictivity, whereas Bahoura et al. [13] and Kandaswamy et al. [11] have used only sensitivity.

The receiver operating characteristic (ROC) curve is a useful technique for organizing classifiers and visualizing their performance [31]. It has been used to display the relationship between sensitivity (true positive rate) and 1-specificity (false positive rate) across all possible cutoff values that define the positivity of the test (Fig. 14). They are frequently used in biomedical research to evaluate classification and prediction models [32]. The area under curve (AUC) serves a single measure that summarizes the discriminative ability of the classifier across a full range of cutoffs [33]. It has been used by Guler et al. [8] to compare the diagnostic accuracy of their classifiers.

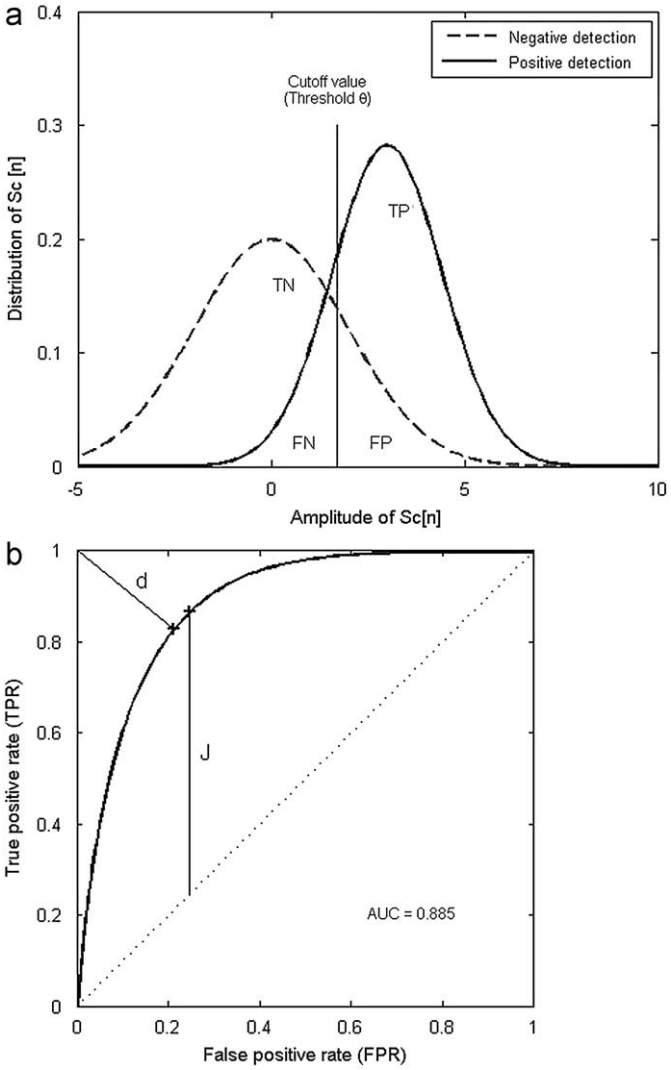

Fig. 14. (a) Simulated Gaussian distributions of actually positive (full line) and negative (dashed line) signal [30]. The cutoff value (vertical line) misclassifies the left-hand part of the actually positive distribution as false negatives (FN), and the right-hand part of the actually negative distribution as false positives (FP). (b) The corresponding ROC curve that is used to compute the area under curve parameter (AUC = 0.885).

4.2. Conventional discriminatory threshold

As shown in Fig. 14, the previous parameters depend on the threshold value θ. We define $\mathrm{SE_c}$ and $\mathrm{SP_c}$, respectively, the values of sensitivity and specificity that correspond to the conventional threshold $\theta = 0$ in Eqs. (30), (39) and (49). These values correspond to the conventional definitions given by Eqs. (28), (37) and (46).

4.3. Optimal discriminatory threshold

A perfect classifier would be able to identify all wheezing sounds (100% sensitivity) and to identify all normal sounds (100%

Table 1
Normal respiratory sounds (N01–N12) recorded on 12 healthy subjects and wheezing sounds (W01–W12) recorded on 12 asthmatic subjects.

Normal sounds			Wheezing sounds		
File	Duration (s)	Number of segments	File	Duration (s)	Number of segments
N01	15.68	183	W01	6.73	77
N02	17.14	199	W02	4.62	53
N03	32.39	378	W03	9.63	111
N04	10.10	117	W04	2.76	31
N05	16.84	196	W05	2.71	30
N06	17.77	207	W06	17.53	204
N07	7.68	88	W07	4.21	48
N08	8.22	94	W08	12.36	143
N09	6.84	179	W09	6.21	71
N10	7.24	183	W10	8.07	93
N11	9.16	106	W11	4.14	47
N12	7.91	91	W12	6.72	77
Total	157.01	1822	Total	85.71	985

Recorded from the trachea, the respiratory sounds are down-sampled to 6000 Hz and normalized in amplitude. Sounds are uniformly divided into overlapped (50%) segments, 1024 samples each.

specificity). However, in practice, few classifiers are perfect, and a balance must be found between sensitivity and specificity. Two methods to optimize the discriminatory threshold are commonly used [33,34]:

(1) Choose the threshold that will make the resulting classification as close to a perfect classifier as possible. As shown in Fig. 14b, the closest point of the ROC curve to the upper left corner is given by the minimum of the Euclidean distance.

$$d = \min\left\{\sqrt{[\mathrm{FPR}(\theta)]^2 + [1-\mathrm{TPR}(\theta)]^2}\right\} = \min\left\{\sqrt{[1-\mathrm{SE}(\theta)]^2 + [1-\mathrm{SP}(\theta)]^2}\right\} \quad (54)$$

The optimal threshold is defined by $\theta_1^* = \arg\{d\}$.

(2) Choose the threshold that will make the resulting classification as far from a noninformative as possible. As shown in Fig. 14b, the maximum vertical distance between the ROC curve and the diagonal or chance line is given by

$$J = \max\{\mathrm{FPR}(\theta) - \mathrm{FPR}(\theta)\} = \max\{\mathrm{SE}(\theta) + \mathrm{SP}(\theta) - 1\} \quad (55)$$

In fact, this distance (J) is equal to Youden's index [35]. The optimal threshold is defined by $\theta_2^* = \arg\{J\}$.

The choice of method is mostly dependent on the field of application. In the case of a medical diagnosis, optimal generally means *close to ideal*, whereas in the case of engineering, it usually means *far from random* [34]. Thus, we define $\mathrm{SE_o}$, and $\mathrm{SP_o}$, respectively, the values of sensitivity and specificity that correspond to the optimal threshold (θ_1^*).

5. Statistical significance

In this paper, McNemar's test was used to test whether any difference in performance between two classification algorithms, A_1 and A_2, tested on the same dataset is statistically significant. The joint performance of the two algorithms can be summarized in a 2×2 contingency table containing the number of examples correctly classified by both algorithms (n_{00}), neither algorithm (n_{11}), only the first one (n_{01}), and only the second one (n_{10}):

		A_2	
		Correct	Incorrect
A_1	Correct	n_{00}	n_{01}
	Incorrect	n_{10}	n_{11}

where $n = n_{00} + n_{01} + n_{10} + n_{11}$ is the total number of examples in the test set.

Under the null hypothesis $\mathrm{H_0}$, the two algorithms should have the same error rate, which means that $n_{01} = n_{10}$. McNemar's test statistic is given as [36]

$$Z = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \quad (56)$$

where this quantity (Z) is distributed approximately as chi-squared χ_1^2 statistic with 1 degree of freedom; it incorporate a *continuity-correction* term (of -1 in the numerator) to take into account the fact that the statistic is discrete while χ_1^2 distribution is continuous [37].

If this quantity is greater than the critical value $\chi^2_{1,1-\alpha}$, defined for the desired significance level α, then $\mathrm{H_0}$ is rejected in favor of the hypothesis that the two classifiers have different performance. For the commonly used values of α, 0.05, 0.01 and 0.001, their corresponding critical values are 3.84, 6.63 and 10.82, respectively. The hypothesis can also be tested by calculating a probability called p-value. The p-value is defined as $p = \mathrm{Prob}(Z > z)$ where Z is a random variable with distribution χ_1^2 and z is the realization value of McNemar's test Z. The null hypothesis $\mathrm{H_0}$ is rejected if $p < \alpha$.

6. Experimental results

6.1. Database

Our database is constructed from real respiratory sounds obtained from various sources: (1) sounds recorded on healthy and asthmatic patients [38], (2) the R.A.L.E. database-CD [39], (3) the ASTRA database-CD [40], and (4) and various Internet sites of laboratories working in this field. The classifiers are tested and evaluated using two respiratory sound classes: normal and wheezing. The normal class is obtained from sounds recorded on 12 healthy subjects, while the wheezing class is obtained from sounds recorded on 12 asthmatic subjects. The wheezing sounds are manually pre-processed

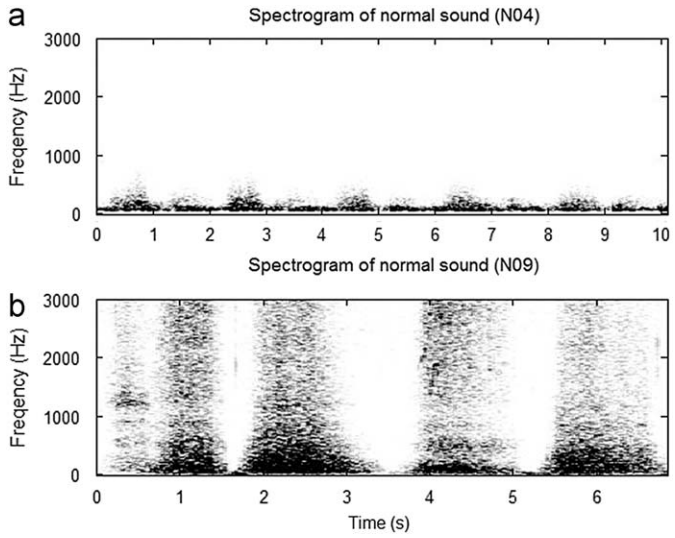

Fig. 15. Spectrograms of two normal respiratory sounds having distinct frequency content and respiratory cycle duration. (a) Spectrogram of normal sound (N04). (b) Spectrogram of normal sound (N09).

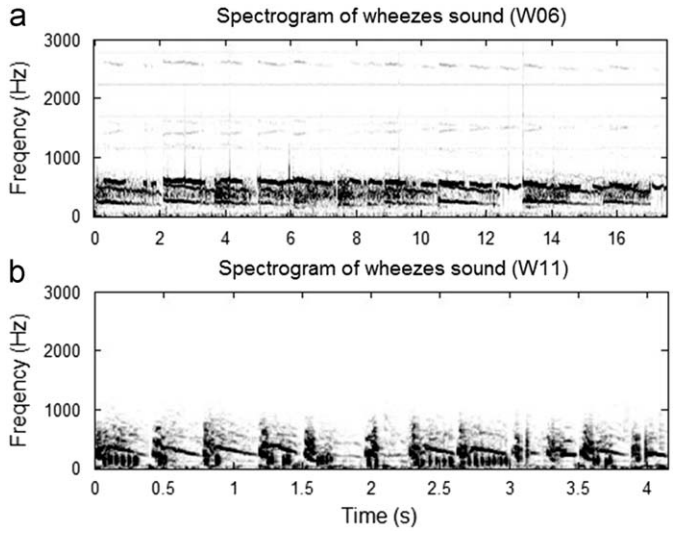

Fig. 16. Spectrograms of two respiratory sounds including monophonic and polyphonic wheezes. (a) Spectrogram of wheezes sound (W06). (b) Spectrogram of wheezes sound (W11)

in order to eliminate eventual non-wheezing segments with *Adobe Audition* [41] software by listening and visualizing their spectrograms [13]. Recorded from the trachea, the respiratory sounds are down-sampled to 6000 Hz and normalized in amplitude. As shown in Table 1 and the examples of Figs. 15 and 16, we have purposefully taken files which are short but representative of the spectral variabilities of respiratory sounds in order to avoid bias in the classification process by presenting similar respiratory phases (redundant data). The frequency content of these sounds is very heterogenous, where the upper frequency varies from 600 to 3000 Hz. Also, the wheezing sound class contains monophonic and polyphonic wheezes. The characteristics due to sex, age, physiology, recording position or any special recording conditions were not taken into consideration because the ultimate goal is to perform a classification system insensitive to these parameters. In fact, training and testing phases are performed on respiratory sound segments. In this study, we are only interested in the presence/abscence of wheezes in the tested segment not by their number nor their frequencies.

6.2. Protocol

We recall that the classifiers operate in two steps: the training phase and the testing phase (Fig. 1). For a rigorous evaluation, the sound used in the training must not be used in the test. In the training phase, sounds recorded on 11 subjects of each class are used in the training while the sound recorded on the remaining one is used in the recognition phase. This is repeated for each sound (*leave-one-out* method). For example, when sounds N02-N12 and W02-W12 are used for training, the remaining ones N01 and W01 are used for testing. As indicated previously, training and testing phases are performed on respiratory sound segments.

6.3. Parameters affecting classification

The performance of the constructed classifier depends both on: (1) the combination of feature extraction and modeling methods, and (2) the parameters associated to each of these methods.

The feature extraction methods can be affected by the following parameters:

- window type;
- frame length;
- frame overlap;
- feature vector length;

while the modeling methods can be affected by:

- number of centroids and number of iterations for VQ;
- number of mixtures, type of initialization, and covariance matrix for GMM;
- number of neurones in hidden layer and number of iterations for MLP.

We have tested all possible combinations of characterization and modeling methods described in Sections 2 and 3, respectively, but it is impossible to experiment all possible values taken by the parameters described previously. Table 2 gives the values of the parameters, giving the best performances.

6.4. Postprocessing

Considering the fact that wheezes are continuous adventitious sounds, it is very interesting to take into account the variation of the score over adjacent segments [15]. This would be done by smoothing the score function. The filtered function $Sc_f[n]$ is obtained by convoluting the original function $Sc[n]$ with a low-pass filter $h[n]$:

$$Sc_f[n] = Sc[n] * h[n] \tag{57}$$

The best results are obtained with a Hamming window $h[n]$ of a length corresponding to 0.5 s.

6.5. Results and discussion

In this work, we evaluate 17 classifiers obtained by combining six feature extraction methods (R-FT, S-AR, B-MFCC, B-SBC, K-WT and P-WPT) and three modelization/classification techniques (VQ, GMM and MLP), as illustrated in Figs. 1 and 2. We are not be able to

Table 2
Optimized values of parameters affecting classification.

Method	R-FT	S-AR	B-MFCC	B-SBC	K-WT	S-WPT
Feature extraction						
Window type	Hamming	Hamming	Hamming	Hamming	Hamming	Hamming
Frame length	1024	1024	1024	1024	1024	1024
Frame overlap	50%	50%	50%	50%	50%	50%
Feature vector length	26	7	24	24	14	19
Method	VQ		GMM		MLP	
Modeling/classification						
Number of centroid	32		–		–	
Number of mixtures	–		8		–	
Nodes in hidden layer	–		–		30	
Number of iterations	50		10		5000	
Type of covariance matrix	–		Diag.		–	
Type of initialization	–		LBG		–	

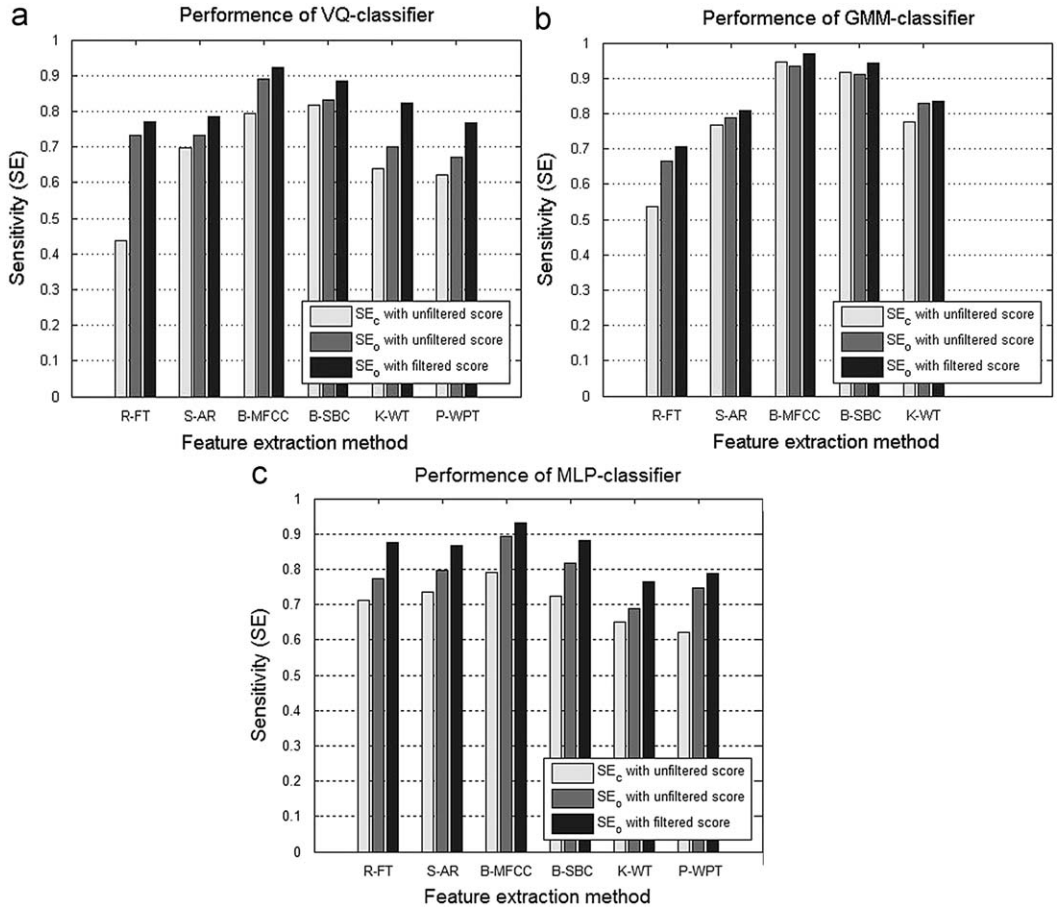

Fig. 17. Sensitivity (SE) obtained with (a) vector quantization (VQ), (b) Gaussian mixture models (GMM) and (c) multi-layer perceptron based classifiers for various feature extraction methods.

obtain the results of the combination P-WPT/GMM because the EM algorithm cannot converge properly. This numerical difficulty can be associated to outliers or to repeated data samples [42]. We believe that the second half of the feature vector (Eq. (23)) composed of integer values ($1 \leq \ell \leq 6$) can be considered as repeated data. These classifiers are implemented in Matlab using neural network [43], wavelab [44] and netlab [45] toolboxes.

Using the values of Table 2, the computed sensitivity, specificity and AUC of these classifiers are reported on Figs. 17, 18 and 19, respectively. It can be seen that the conventional threshold provides an unbalanced sensitivity and specificity for most of the feature extraction techniques. The large difference is obtained with the R-FT/VQ classifier ($SE_c = 43.7\%$ and $SP_c = 92.8\%$), while the small difference is obtained with the B-MFCC/GMM classifier ($SE_c = 94.6\%$ and $SP_c = 91.9\%$). However, the optimal threshold allows a better balance between sensitivity and specificity. Also, it can be seen that the smoothing process improves the performance of classifiers. In fact, the smoothing filter prevents against false detections by eliminating

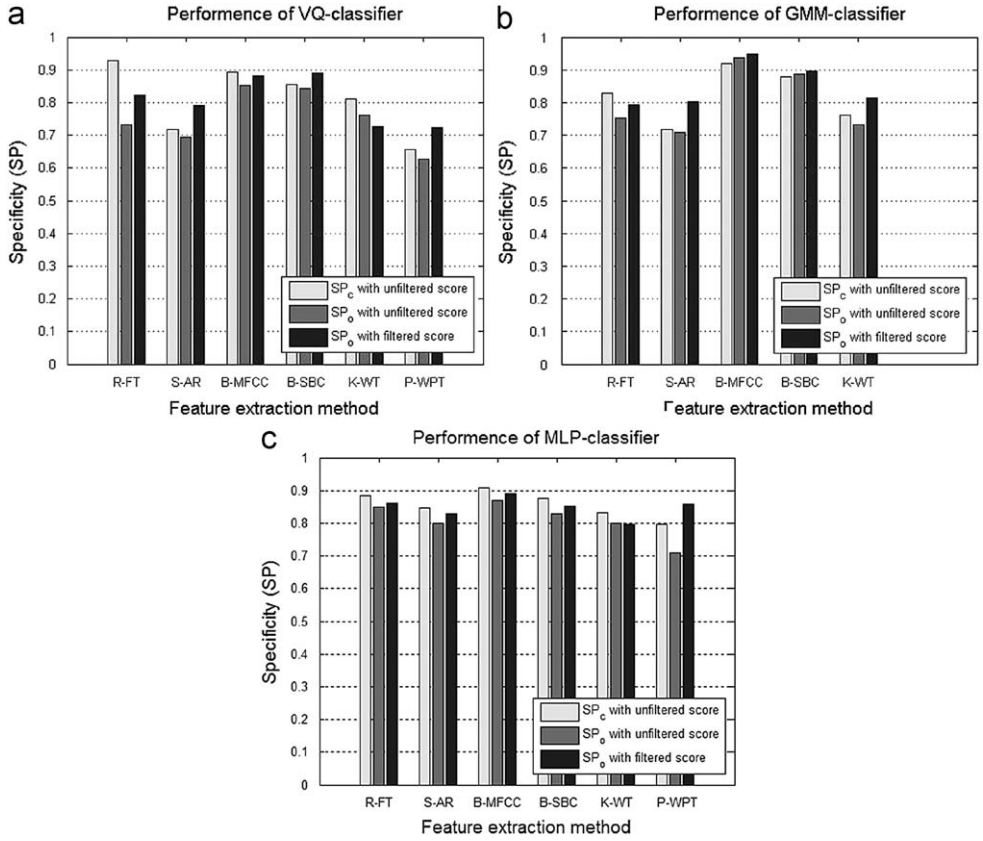

Fig. 18. Specificity (SP) obtained with (a) vector quantization (VQ), (b) Gaussian mixture models (GMM) and (c) multi-layer perceptron (MLP) based classifiers for various feature extraction methods.

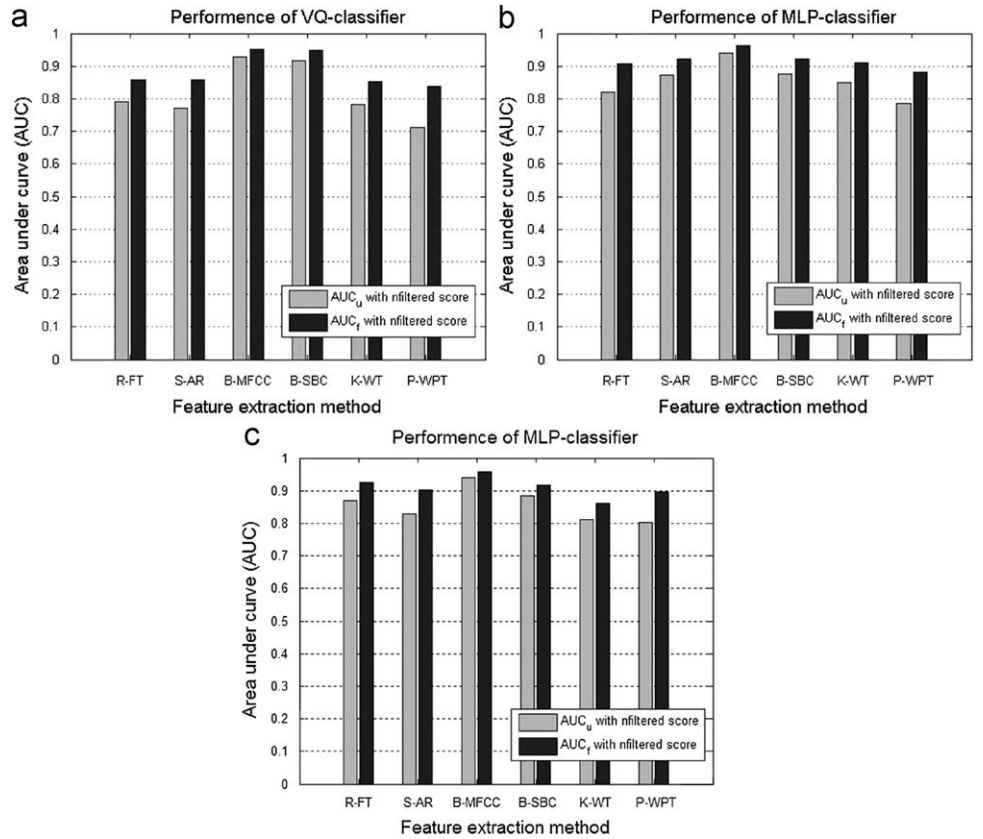

Fig. 19. Area under curve (AUC) obtained with (a) vector quantization (VQ), (b) Gaussian mixture models (GMM) and (c) multi-layer perceptron (MLP) based classifiers for various feature extraction methods.

Table 3
McNemar's test for pair-wise method efficiency comparison using unfiltered score function and conventional threshold.

Methods	S-AR/VQ	B-MFCC/VQ	B-SBC/VQ	K-WT/VQ	P-WPT/VQ	R-FT/GMM	S-AR/GMM	B-MFCC/GMM	B-SBC/GMM	K-WT/GMM	R-FT/MLP	S-AR/MLP	B-MFCC/MLP	B-SBC/MLP	K-WT/MLP	P-WPT/MLP
R-FT/VQ	15.21	110.96	75.35	0.37	94.60	8.793	2.79	327.61	185.15	0.88	46.12	29.01	176.71	83.09	10.11	18.30
S-AR/VQ	–	223.25	174.01	12.37	29.26	2.66	7.80	451.64	347.29	25.02	85.19	119.18	225.08	160.89	44.50	0.02
B-MFCC/VQ		–	5.03	123.39	362.31	166.94	161.95	89.59	23.50	95.17	25.59	33.82	4.54	5.72	62.10	206.54
B-SBC/VQ			–	96.84	306.85	120.35	120.53	131.05	54.61	72.98	8.98	13.24	14.42	0.002	41.17	164.96
K-WT/VQ				–	81.31	3.94	1.39	378.75	234.11	3.68	52.00	37.41	185.10	94.51	17.51	11.68
P-WPT/VQ					–	55.04	58.98	621.27	473.58	112.12	213.95	205.27	407.80	303.41	149.44	42.92
R-FT/GMM						–	0.29	374.67	242.98	12.88	82.84	64.10	210.82	128.61	31.72	2.50
S-AR/GMM							–	380.10	291.11	7.67	51.68	71.57	165.67	111.49	20.72	4.24
B-MFCC/GMM								–	37.09	335.77	218.30	182.22	69.25	138.65	298.71	446.06
B-SBC/GMM									–	196.54	83.19	101.25	4.86	54.38	155.52	305.56
K-WT/GMM										–	28.21	20.41	142.11	63.48	5.72	25.00
R-FT/MLP											–	0.06	67.36	10.28	13.08	101.02
S-AR/MLP												–	44.64	11.63	6.86	88.80
B-MFCC/MLP													–	22.29	118.50	244.87
B-SBC/MLP														–	39.33	162.72
K-WT/MLP															–	52.34

A value greater than 3.84 indicates that difference between the classification performances of the two methods is statistically significant.

Table 4
McNemar's test for pair-wise method efficiency comparisons using unfiltered score function and optimal threshold.

Methods	S-AR/VQ	B-MFCC/VQ	B-SBC/VQ	K-WT/VQ	P-WPT/VQ	R-FT/GMM	S-AR/GMM	B-MFCC/GMM	B-SBC/GMM	K-WT/GMM	R-FT/MLP	S-AR/MLP	B-MFCC/MLP	B-SBC/MLP	K-WT/MLP	P-WPT/MLP
R-FT/VQ	1.26	120.06	123.28	10.59	46.31	0.007	0.71	300.85	218.39	19.92	92.64	26.90	205.74	105.44	46.58	7.64
S-AR/VQ	–	126.88	173.22	17.24	27.91	1.50	6.68	373.84	324.00	31.17	76.17	49.76	153.87	111.24	47.65	0.93
B-MFCC/VQ		–	3.78	55.27	259.78	113.16	80.15	101.51	56.22	31.02	4.14	30.37	7.93	0.70	17.50	168.23
B-SBC/VQ			–	81.51	299.32	125.94	123.70	79.93	41.55	60.00	12.75	55.48	0.34	7.29	38.50	189.20
K-WT/VQ				–	88.60	10.41	3.60	268.92	186.78	3.71	37.36	3.71	109.63	48.44	14.17	30.28
P-WPT/VQ					–	46.82	54.27	518.70	432.67	117.11	196.49	120.67	315.91	236.05	148.18	24.69
R-FT/GMM						–	0.62	295.69	223.50	21.31	87.14	24.90	177.92	105.46	41.80	7.06
S-AR/GMM							–	313.71	272.61	12.19	45.38	19.42	102.77	68.70	23.90	9.17
B-MFCC/GMM								–	15.47	237.27	149.76	195.18	66.70	120.50	212.81	380.25
B-SBC/GMM									–	156.25	78.82	150.76	22.28	75.40	132.56	303.73
K-WT/GMM										–	16.66	0.09	69.14	25.44	3.80	45.89
R-FT/MLP											–	10.26	28.51	1.73	7.11	126.49
S-AR/MLP												–	50.74	23.44	1.31	55.41
B-MFCC/MLP													–	14.40	53.22	227.98
B-SBC/MLP														–	13.32	156.79
K-WT/MLP															–	153.07

A value greater than 3.84 indicates that difference between the classification performances of the two methods is statistically significant.

Table 5
McNemar's test for pair-wise method efficiency comparisons using filtered score function and conventional threshold.

Methods	S-AR/VQ	B-MFCC/VQ	B-SBC/VQ	K-WT/VQ	P-WPT/VQ	R-FT/GMM	S-AR/GMM	B-MFCC/GMM	B-SBC/GMM	K-WT/GMM	R-FT/MLP	S-AR/MLP	B-MFCC/MLP	B-SBC/MLP	K-WT/MLP	P-WPT/MLP
R-FT/VQ	7.89	73.70	45.62	4.84	94.48	11.30	6.62	293.22	126.53	0.66	24.45	10.53	124.68	52.97	0.11	18.75
S-AR/VQ	–	155.60	111.20	0.89	38.89	0.0874	0.070	376.10	252.16	4.549	43.60	60.61	149.08	93.09	6.79	0.34
B-MFCC/VQ		–	4.55	114.61	307.36	128.51	146.27	117.88	22.40	93.65	23.65	40.60	4.67	6.47	83.53	154.06
B-SBC/VQ			–	94.89	250.46	89.14	107.61	160.58	47.90	74.21	7.98	17.32	13.10	0.04	67.37	124.84
K-WT/VQ				–	54.89	0.51	0.47	432.01	203.92	2.38	53.09	29.28	188.88	89.48	4.59	2.99
P-WPT/VQ					–	51.48	43.33	577.00	391.57	77.17	168.33	148.69	344.06	257.28	84.03	41.14
R-FT/GMM						–	0.004	341.00	177.49	3.87	58.20	36.40	173.95	97.78	6.66	1.09
S-AR/GMM							–	370.67	268.81	3.68	40.11	56.44	145.46	95.16	5.79	0.68
B-MFCC/GMM								–	72.73	393.75	258.02	211.74	103.91	170.23	389.97	401.65
B-SBC/GMM									–	178.42	71.85	100.72	4.00	51.04	178.95	232.00
K-WT/GMM										–	31.16	17.11	156.42	63.80	0.41	8.06
R-FT/MLP											–	0.46	66.37	9.08	32.90	70.56
S-AR/MLP												–	45.47	13.22	13.73	54.47
B-MFCC/MLP													–	24.89	156.47	187.64
B-SBC/MLP														–	63.40	124.56
K-WT/MLP															–	15.08

A value greater than 3.84 indicates that difference between the classification performances of the two methods is statistically significant.

Table 6
McNemar's test for pair-wise method efficiency comparisons using filtered score function and optimal threshold.

Methods	S-AR/VQ	B-MFCC/VQ	B-SBC/VQ	K-WT/VQ	P-WPT/VQ	R-FT/GMM	S-AR/GMM	B-MFCC/GMM	B-SBC/GMM	K-WT/GMM	R-FT/MLP	S-AR/MLP	B-MFCC/MLP	B-SBC/MLP	K-WT/MLP	P-WPT/MLP
R-FT/VQ	2.13	74.11	135.92	13.08	24.09	3.16	0.10	397.75	203.25	10.32	51.44	31.92	113.00	94.88	12.05	1.57
S-AR/VQ	–	39.30	110.14	2.36	38.59	0.08	2.04	366.96	235.08	3.50	12.89	23.24	37.11	50.98	1.24	6.05
B-MFCC/VQ		–	25.39	24.10	157.61	54.02	49.56	220.95	93.23	16.31	7.06	4.59	0.67	2.38	31.28	100.96
B-SBC/VQ			–	85.92	253.99	113.61	130.02	146.27	32.43	78.51	44.82	47.10	13.75	13.80	104.88	178.78
K-WT/VQ				–	61.62	5.46	6.26	387.68	162.75	0.11	9.44	6.63	40.35	41.55	0.28	23.49
P-WPT/VQ					–	38.00	26.08	562.24	374.83	70.81	101.78	100.50	168.88	187.62	59.26	16.99
R-FT/GMM						–	0.55	348.16	174.09	4.89	30.01	22.71	74.43	76.26	3.51	8.55
S-AR/GMM							–	399.57	290.41	9.40	19.55	33.43	47.26	66.06	4.61	1.43
B-MFCC/GMM								–	78.98	387.75	315.20	250.41	233.25	212.22	406.88	424.80
B-SBC/GMM									–	176.05	114.30	122.85	65.12	78.00	194.28	242.30
K-WT/GMM										–	4.10	4.35	23.96	28.87	0.66	19.12
R-FT/MLP											–	0.05	19.86	18.13	17.30	66.91
S-AR/MLP												–	6.00	10.89	9.23	52.84
B-MFCC/MLP													–	0.35	56.24	119.97
B-SBC/MLP														–	57.12	125.50
K-WT/MLP															–	24.35

A value greater than 3.84 indicates that difference between the classification performances of the two methods is statistically significant.

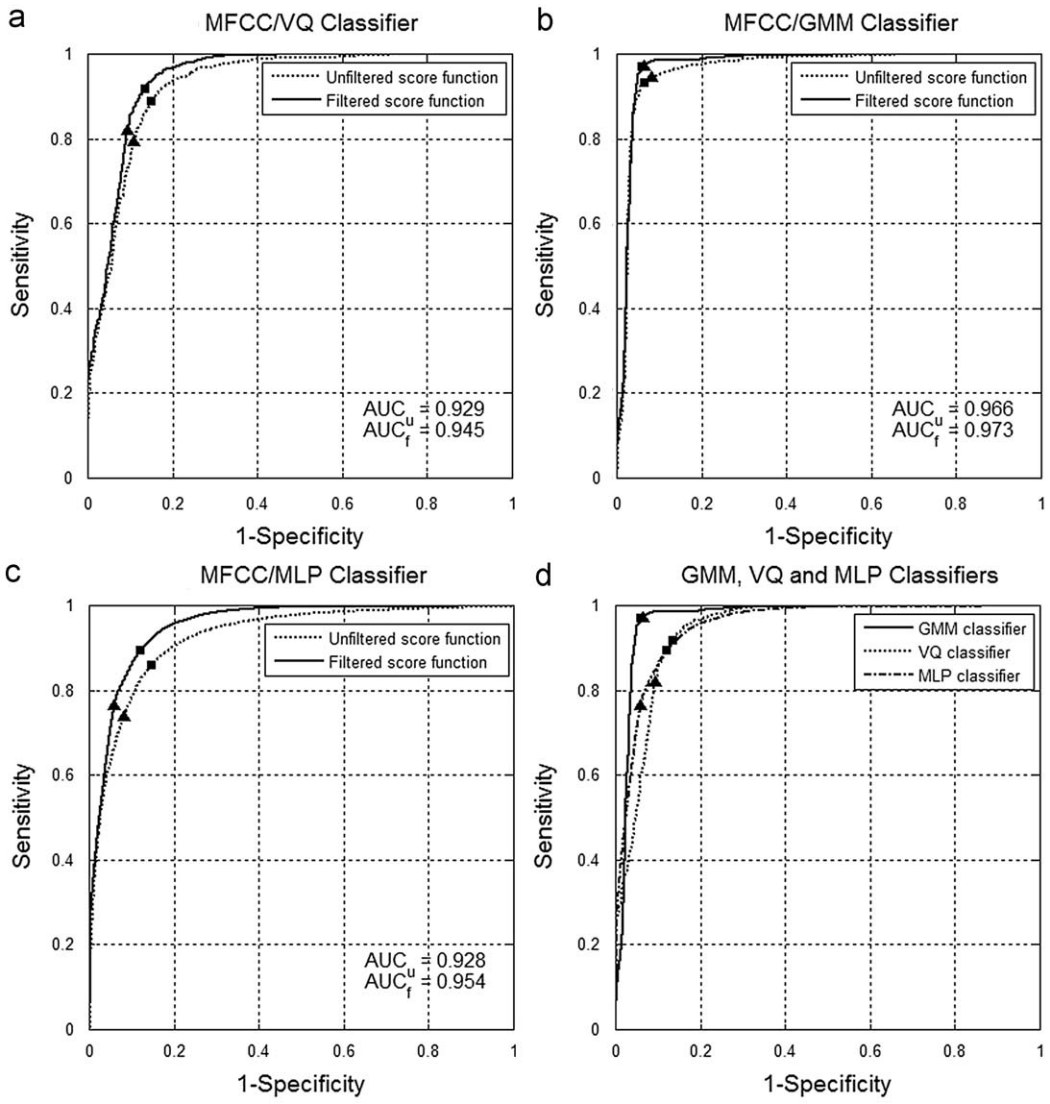

Fig. 20. ROC curves obtained for (a) VQ, (b) GMM and (c) MLP based classifiers for unfiltered (dashed line) and filtered (full line) score function. The conventional and optimum values of *sensitivity* and *specificity* are indicated by the symbols (▲) and (■), respectively. The results of the filtered score functions are summarized in (d). These curves are obtained with the parameters of Table 2. Area under curve for unfiltered and filtered score functions are given by AUC_u and AUC_f, respectively.

isolated peaks in the score function. Figs. 17, 18 and 19 show that the proposed characterization methods, based on the cepstral and perceptual analysis (B-MFCC and B-SBC), give the best results. The best performances are obtained with B-MFCC/GMM classifier using the filtered score function (SE_o=97.2%, SP_o=94.2% and AUC_f=0.974). The performance of this approach can be attributed to its ability to exploit the auditory system principle, as well as the decorrelating property of the cepstrum [13].

Tables 3–6 show the pair-wise method comparison results for the presented classifiers using McNemar's test. As in Figs. 17, 18 and 19, the validation data set is composed of 1822 normal and 985 wheeze sound segments. Each value in a cell is a McNemar's test value calculated using Eq. (56). At the significance level α=0.05, most of these values are greater than 3.84, which lead to the rejection of the null hypothesis. The obtained McNemar's test values prove that the overall results presented in Figs. 17, 18 and 19 are statistically significant ($Z > 3.84$ or $p < 0.05$). It can also be seen that the performances obtained by the proposed approach (B-MFCC/GMM) are significantly different from those obtained by the other methods ($Z > 3.84$ or $p < 0.05$).

Figs. 20a–c represent the ROC curves obtained with B-MFCC feature extraction method in combination with VQ, GMM and MLP based classifiers, respectively. Unlike comparing combinations of sensitivity and specificity at various threshold values, the area under curve (AUC) constitutes an objective parameter of the diagnostic accuracy that is not influenced by the threshold value. Sensibility and specificity are considerably affected by the threshold optimization in the case of the VQ and MLP classifiers, but with a lesser extent in the case of the GMM.

The results obtained with the tested database show that:

(1) the cepstral analysis (B-MFCC) offers the best performances for the three classifiers,
(2) the B-MFCC/GMM combination is well-adapted to classify respiratory sounds,
(3) the post-processing filter improves the classification performances.

The best performance obtained by GMM is due to its ability to capture the intra-class variability by considering each class as a

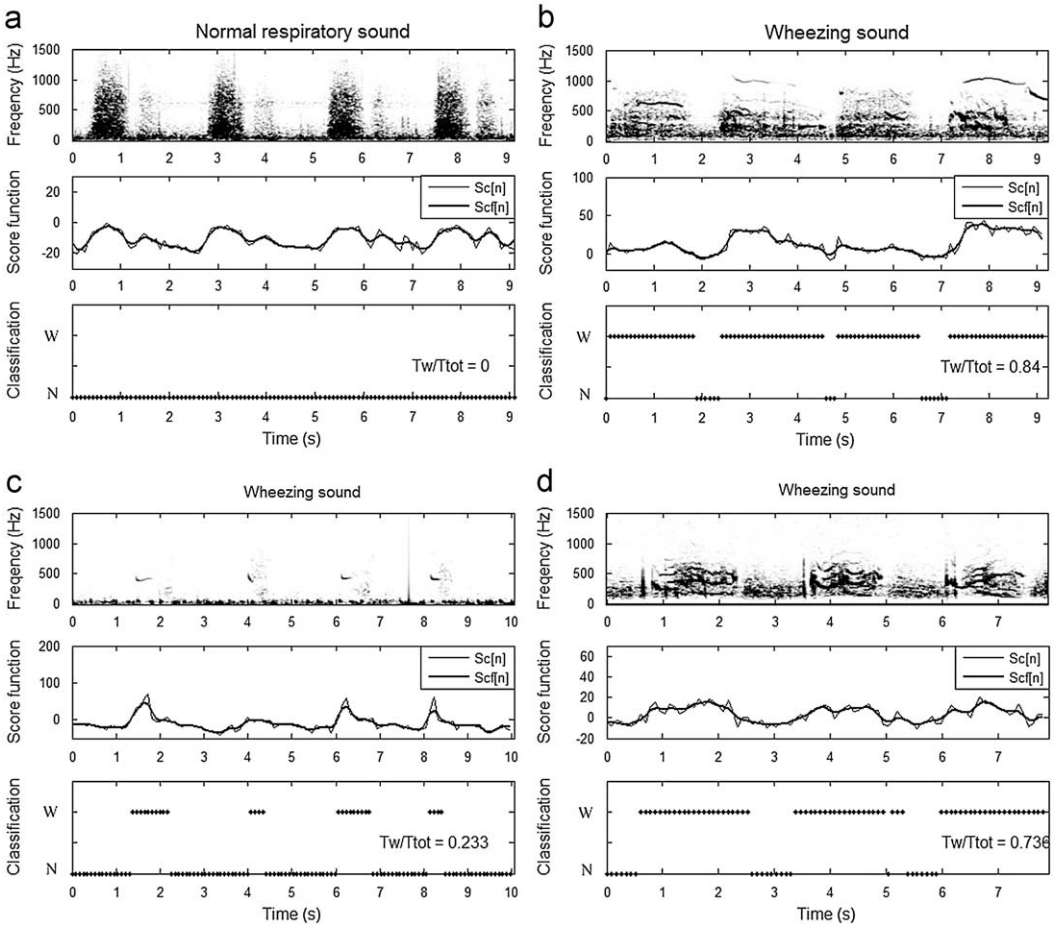

Fig. 21. Classification of (a) normal and (b–d) wheezing respiratory sounds. Each subfigure includes the spectrogram of the tested sound (top), the score functions (middle) and the classification of the sound segments (bottom) in normal (N) or wheezes (W) classes. The proportion of the respiratory cycle occupied by wheezes (T_w/T_{tot}) is computed for each sound.

mixture of subclasses of Gaussian distributions. MLP classifier provides performances slightly higher than those obtained with VQ because the nonlinear nature of its discriminant functions used for separate classes [46].

To illustrate the principle of segmental classification in more detail, four additional respiratory sounds are tested using B-MFCC/GMM classifier. The model is trained using the previous 24 respiratory sounds described in Section 6.1. Fig. 21 presents the obtained results, where each subfigure includes the spectrogram of the tested sound (top), the score function (middle) and the classification results of the sound segments (bottom) in normal (N) or wheezes (W) classes. Despite the variability of the frequency content of the respiratory sounds, normal and wheezing segments are correctly classified. In addition, the classification by segment presents two advantages: (1) possibility of a real-time implementation, (2) assessment of the severity of airway obstruction by calculating the proportion of the respiratory cycle occupied by wheezes (T_w/T_{tot}).

6.6. Comparison to reference methods

The proposed classification method B-MFCC/GMM is compared to two reference methods:

- S-AR/*k*-NN [9] is a method based on the feature vector described by Eq. (6) in combination with a *k*-NN classifier. In this experiment, the feature vectors are computed independently of the respiratory phase. The *k*-NN classifier is based on the seven nearest neighbors using Euclidean distance.
- K-WT/K-MLP [11] is a technique based on the feature vector described by Eq. (20) in combination with a MLP classifier. This method was initially proposed to classify respiratory sounds into one of six classes: normal, wheeze, crackle, squawk, stridor, or rhonchus. In this experiment, the network is characterized by 19 inputs, 40 neurones in the hidden layer and 2 output.

The lack of a score function in the case of the *k*-NN classifier does not permit the building of its ROC curves. Therefore, the comparison between the three methods is limited to the parameters of sensitivity and specificity. For the GMM and K-MLP based classifiers, these parameters are computed on unfiltered score function using a conventional threshold.

The obtained results (Table 7) show that the performance of the proposed method GMM/MFCC is considerably higher than those of the reference methods ($SE_c = 94.6\%$ and $SP_c = 91.9\%$). As shown in Table 8, McNemar's test demonstrated significant difference between results obtained by the three methods ($Z > 10.82$ or $p < 0.001$).

7. Conclusion

In this paper, different pattern recognition methods used to classify respiratory sounds into normal and wheeze classes are

Table 7
Comparison of the proposed method B-MFCC/GMM to S-AR/*k*-NN [9] and K-WT/K-MLP [11] reference methods.

Tested files	S-AR/*k*-NN [9]					K-WT/K-MLP [11]					B-MFCC/GMM				
	True positives	False negatives	True negatives	False positives	Sensitivity SE_c (%)	Specificity SP_c (%)	True positives	False negatives	True negatives	False positives	Sensitivity SE_c (%)	Specificity SP_c (%)	True positives	False negatives	True negatives	False positives	Sensitivity SE_c (%)	Specificity SP_c(%)
W1, N1	34	43	159	24	44.1	86.9	32.9	44.1	177.8	5.2	42.8	97.1	73	4	167	16	94.8	91.2
W2, N2	48	5	192	7	90.6	96.5	50.5	2.5	170.6	28.4	95.3	85.7	53	0	180	19	100	90.4
W3, N3	60	51	317	61	54.0	83.8	86.7	24.3	319.9	58.1	78.1	84.6	110	1	356	22	99.1	94.2
W4, N4	19	12	117	0	61.3	100	6.2	24.8	105.2	11.8	20.0	89.9	15	16	117	0	48.4	100
W5, N5	15	15	196	0	50.0	100	20.9	9.1	174.4	21.6	69.8	89.0	30	0	196	0	100	100
W6, N6	194	10	163	44	95.1	78.7	60.4	143.6	206.5	0.5	29.6	99.8	197	7	207	0	96.6	100
W7, N7	38	10	43	45	79.2	48.8	43.7	4.3	48.0	40.0	91.2	54.5	48	0	71	17	100	80.7
W8, N1	43	100	14	81	30.1	14.7	98.5	44.5	86.4	8.6	68.9	90.9	133	10	89	6	93.0	93.7
W9, N9	33	38	56	23	46.5	70.9	67.2	3.8	54.3	24.7	94.7	68.7	71	0	79	0	100	100
W10, N10	89	4	51	32	95.7	61.4	90.1	2.9	16.4	66.6	96.9	19.7	93	0	18	65	100	21.7
W11, N11	41	6	64	42	87.2	60.4	27.5	19.5	76.1	29.9	58.6	71.7	32	15	105	1	68.1	99.1
W12, N12	64	13	91	0	83.1	100	71.7	5.3	88.7	2.3	93.1	97.5	77	0	91	0	100	100
Total	678	307	1463	359	**68.8**	**80.3**	656.6	328.4	1524.4	297.6	**66.7**	**83.6**	932	53	1676	146	**94.6**	**91.9**

For the K-WT/K-MLP method, the learning and testing process is repeated 50 times in order to take the average value of each evaluation parameter.

Table 8
McNemar's test for pair-wise comparison results using the same data test as Table 7.

Method		K-WT/K-MLP [11]	B-MFCC/GMM
S-AR/*k*-NN [9]	–	1208.44	1867.99
K-WT/K-MLP [11]		–	298.71

presented. Various combinations of feature extraction and classification methods are evaluated using receiver operating characteristic (ROC) curves. The obtained results show that the proposed approach B-MFCC/GMM outperforms all other commonly used methods ($p < 0.05$). A significant performances improvement is obtained by smoothing the score function by reducing false detections. Also, the proposed system can be used to quantify the severity of airway obstruction by computing the proportion of the respiratory cycle occupied by wheezing.

Finally, the proposed system presents superior performances, which permit the practitioner to achieve a better diagnosis of the pulmonary diseases with the possibility of real-time implementation.

Conflict of interest statement

None declared.

Acknowledgment

This research is financially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] F. Dalmay, M.T. Antonini, P. Marquet, R. Menier, Acoustic properties of the normal chest, European Respiratory Journal 8 (10) (1995) 1761–1769.
[2] A.R.A. Sovijärvi, L.P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, J.E. Earis, Characteristics of breath sounds and adventitious respiratory sounds, European Respiratory Review 10 (2000) 591–596.
[3] N. Meslier, G. Charbonneau, J.L. Racineux, Wheezes, European Respiratory Journal 8 (11) (1995) 1942–1948.
[4] R.G. Loudon, Clinical application of wheeze analysis, Monaldi Archives of Chest Disease 48 (5) (1993) 583–585.
[5] K.E. Forkheim, D. Scuse, H. Pasterkamp, A comparison of neural network models for wheeze detection, in: IEEE WESCANEX 95 Proceedings, vol. 1, New York, USA, 1995, pp. 214–219.
[6] S. Rietveld, M. Oud, E.H. Dooijes, Classification of asthmatic breath sounds: preliminary results of the classifying capacity of human examiners versus artificial neural networks, Computers and Biomedical Research 32 (1999) 440–448.
[7] L.R. Waitman, K.P. Clarkson, J.A. Barwise, P.H. King, Representation and classification of breath sounds recorded in an intensive care setting using neural networks, Journal of Clinical Monitoring and Computing 16 (2000) 95–105.
[8] I. Guler, H. Polat, U. Ergun, Combining neural network and genetic algorithm for prediction of lung sounds, Journal of Medical Systems 29 (3) (2005) 217.
[9] B. Sankur, Y.P. Kahya, E.C. Guler, T. Engin, Comparison of AR-based algorithms for respiratory sound classification, Computers in Biology and Medicine 24 (1994) 67–76.
[10] L. Pesu, P. Helisto, E. Adomovic, J.C. Pesquet, A. Saarinen, A.R.A. Sovijarvi, Classification of respiratory sounds based on wavelet packet decomposition and learning vector quantization, Technology and Health Care 6 (1) (1998) 65–74.
[11] A. Kandaswamy, C. Sathish Kumar, Rm.Pl. Ramanathan, S. Jayaraman, N. Malmurugan, Neural classification of lung sounds using wavelet coefficients, Computers in Biology and Medicine 34 (2004) 523–537.
[12] Z. Dokur, T. Olmez, Classification of respiratory sounds by using an artificial neural network, International Journal of Pattern Recognition and Artificial Intelligence 17 (4) (2003) 567–580.
[13] M. Bahoura, C. Pelletier, New parameters for respiratory sound classification, in: Electrical and Computer Engineering, 2003, Canadian Conference on IEEE CCECE 2003, vol. 3, Montreal, Canada, May 4–7, 2003, pp. 1457–1460.
[14] M. Bahoura, C. Pelletier, Respiratory sounds classification using Gaussian mixture models, in: Electrical and Computer Engineering, 2004, Canadian Conference on IEEE CCECE 2004, vol. 3, Niagara Falls, Canada, May 2–5, 2004, pp. 1309–1312.
[15] M. Bahoura, C. Pelletier, Respiratory sounds classification using cepstral analysis and Gaussian mixture models, in: 26th Annual Conference of the IEEE EMBS, San Francisco, CA, September 1–5, 2004, pp. 9–12.

[16] A.M. Noll, Cepstrum pitch determination, Journal of the Acoustical Society of America 41 (1967) 293–309.
[17] J. Xu, J. Cheng, Y. Wu, A cepstral method for analysis of acoustic transmission characteristics of respiratory system, IEEE Transactions on Biomedical Engineering 45 (5) (1998) 660–664.
[18] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 674–693.
[19] R.R. Coifman, M.V. Wickerhauser, Best-adapted wave packet bases, Preprint, Yale University, 1990.
[20] R. Sarikaya, B. Pellom, J.H.L. Hansen, Wavelet packet transform features with application to speaker identification, in: NORSIG-98 IEEE Norsic Signal Processing Symposium, Vigso, Denmark, June 1998, pp. 81–84.
[21] X. Lu, M. Bahoura, An integrated automated system for crackles extraction and classification, Biomedical Signal Processing and Control 3 (3) (2008) 244–254.
[22] Y. Linde, A. Buzo, R.M. Gray, An algorithm for vector quantizer design, IEEE Transactions on Communications 28 (1) (1980) 84–95.
[23] T. Pham, M. Wagner, Ambiguity reduction in speaker identification by the relaxation labeling process, Pattern Recognition 32 (7) (1999) 1249–1254.
[24] D. Reynolds, R. Rose, Robust test-independent speaker identification using Gaussian mixture speaker models, IEEE Transactions on Speech and Audio Processing 3 (1995) 72–83.
[25] D. Reynolds, Speaker identification and verification using Gaussian mixture speaker models, Speech Communication 17 (1995) 91–108.
[26] Q. Chen, Y.W. Chan, K. Worden, Structural fault diagnosis and isolation using neural networks based on response-only data, Computers & Structures 81 (22–23) (2003) 2165.
[27] S. Haykin, Neural Networks: A Comprehensive Foundation, second ed, Prentice-Hall, Upper Saddle River, NJ, 1999.
[28] S. Jung, J. Ghaboussi, Characterizing rate-dependent material behaviors in self-learning simulation, Computer Methods in Applied Mechanics and Engineering 196 (1–3) (2006) 608–619.
[29] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: the RPROP algorithm, in: Proceedings of the IEEE International Conference on Neural Networks, ICNN'93, San Francisco, CA, 1993.
[30] N.N. Khodarev, J. Park, Y. Kataoka, E. Nodzenski, S. Hellman, B. Roizman, R.R. Weichselbaum, C.A. Pelizzari, Receiver operating characteristic analysis: a general tool for DNA array data filtration and performance estimation, Genomics 81 (2) (2003) 202–209.
[31] T. Fawcett, ROC graphs: notes and practical considerations for data mining researchers, Technical Report HPL-2003-4, HP Labs, 2003.
[32] T.A. Lasko, J.G. Bhagwat, K.H. Zou, L. Ohno-Machado, The use of receiver operating characteristic curves in biomedical informatics, Journal of Biomedical Informatics 38 (5) (2005) 404–415.
[33] A.K. Akobeng, Understanding diagnostic tests 3: receiver operating characteristic curves, Acta Pædiatrica 96 (5) (2007) 644–647.
[34] M. Gönen, Analyzing Receiver Operating Characteristic Curves with SAS, SAS Publishing, 2007.
[35] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (1) (1950) 32–35.
[36] Q. McNemar, Note on the sampling error of the difference between correlated proportions or percentages, Psychometrika 12 (2) (1947) 153–157.
[37] T.G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, Neural Computation 10 (7) (1998) 1895–1923.
[38] M. Bahoura, Analyse des Signaux Acoustiques Respiratoires: Contribution à la Détection Automatique des Sibilants par Paquets d'Ondelettes, Ph.D. Thesis, Université de Rouen, January 1999.
[39] The R.A.L.E. lung sound 3.0. ⟨www.rale.ca⟩, 2003.
[40] L'auscultation pulmonaire, CD audio des laboratoires Astra France. ⟨www.ifrance.com/y-detauriac/cdrom/pneumo.htm⟩.
[41] Adobe audition. ⟨www.adobe.com/products/audition/⟩.
[42] C. Archambeau, J.A. Lee, M. Verleysen. On convergence problems of the EM algorithm for finite Gaussian mixtures, in: ESANN'2003 Proceedings—European Symposium on Artificial Neural Networks, Bruges, Belgium, 2003, pp. 99–106.
[43] M. Beale, H. Demuth, M. Hagan, Neural Network Toolbox User's Guide, The MathWorks Inc., Natick, MA, 2007.
[44] Wavelab 850, Stanford Statistics Department, Stanford University ⟨www-stat.stanford.edu/∼wavelab/⟩, 2005.
[45] I.T. Nabney, Netlab: Algorithms for Pattern Recognition, Springer, Berlin, 2004.
[46] L. Zhang, C. Liu, C.J. Davis, A mixture model-based approach to the classification of ecological habitats using forest inventory and analysis data, Canadian Journal of Forest Research 34 (2004) 1150–1156.

Mohammed Bahoura received a B.Sc. degree in Electronics from the University of Science and Technology of Algiers, Algeria, in 1990, a M.Sc. degree (D.E.A.) in Instrumentation and Control from the University of Rouen, France, in 1994 and a Ph.D. in Electrical Engineering, from the same University, in 1999. He was a Postdoctoral Fellow at the University of Quebec at Chicoutimi, Canada, from 1999 to 2001. Since 2001, he has been with the University of Quebec at Rimouski, where is a Professor of Electrical Engineering. His research interests focus on digital signal processing and pattern recognition.