Lung sound classification using cepstral-based statistical features

Nandini Sengupta\textsuperscript{a,}\textsuperscript{*}, Md Sahidullah\textsuperscript{b}, Goutam Saha\textsuperscript{a}

\textsuperscript{a}Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India, Kharagpur-721 302.
\textsuperscript{b}Speech and Image Processing Unit, School of Computing, University of Eastern Finland, Joensuu, Finland, Finland-80101.

Abstract

Lung sounds convey useful information related to pulmonary pathology. In this paper, short-term spectral characteristics of lung sounds are studied to characterize the lung sounds for the identification of associated diseases. Motivated by the success of cepstral features in speech signal classification, we evaluate five different cepstral features to recognize three types of lung sounds: normal, wheeze and crackle. Subsequently for fast and efficient classification, we propose a new feature set computed from the statistical properties of cepstral coefficients. Experiments are conducted on a dataset of 30 subjects using the artificial neural network (ANN) as a classifier. Results show that the statistical features extracted from mel-frequency cepstral coefficients (MFCCs) of lung sounds outperform commonly used wavelet-based features as well as standard cepstral coefficients including MFCCs. Further, we experimentally optimize different control parameters of the proposed feature extraction algorithm. Finally, we evaluate the features for noisy lung sound recognition. We have found that our newly investigated features are more robust than existing features and show better recognition accuracy even in low signal-to-noise ratios (SNRs).

Keywords: Artificial Neural Network (ANN), Auscultation, Discrete Wavelet Transform (DWT), Mel-frequency Cepstral Coefficients (MFCCs), Spectral

\textsuperscript{*}Corresponding Author

Email addresses: nandinisengupta@iitkgp.ac.in (Nandini Sengupta), sahid@cs.uef.fi (Md Sahidullah), gsaha@ece.iitkgp.ernet.in (Goutam Saha)
1. Introduction

Lung sound characteristics and their diagnoses form an indispensable part of pulmonary pathology [1]. Medical practitioners use diverse techniques to identify the lung sound characteristics but most of the available methods are not always convenient. The simplest and most popular technique to determine the pathological conditions from chest auscultation is to use a stethoscope. However, it is unreliable owing to the factors like: (a) the inexperience of the physician leading to his/her inability to recognize the lung sound abnormalities and dysfunctions; and (b) low sensitivity of the human ear to the lower frequency band of a lung sound [2]. Advanced techniques like chest X-rays, spirogram, arterial blood gas analysis, which physicians heavily rely on, are also not without limitations. Arterial blood gas analysis is invasive and expensive; X-ray radiation is harmful for body and spirogram is a subjective process. Furthermore, the non-stationary nature of lung sounds, which gets severe for the case of abnormal subjects [2], makes the detection task an even more difficult for the physicians [3]. Ergo, more economical, patient-friendly, non-invasive, objective and automated or computerized techniques are desirable. Recent studies in computerized procedures have not only improved over simple auscultation processes, but they have also furnished new insights into the analysis of lung sounds for diagnostic purposes.

An automatic lung sound analysis system mainly progresses through the steps of pre-processing, feature extraction and classification. In pre-processing step, collected signal is prepared for the subsequent processing by reducing the heart sound effect [4], filtering or sample rate conversion and amplitude normalization. Feature extraction involves deriving a compact representation of a large set of data without losing distinguishable information. On the other hand, classification assigns different signals to their corresponding groups. A systematic and extensive study on various features and classifiers used for computer-
based lung sound recognition can be found in a recent paper [5]. Wavelet-based features proved to be the most popular among these, since they focus on non-stationarity of lung sounds [2]. Compared to this, however, cepstral features along with Gaussian mixture model (GMM) classifier have shown better classification results [6]. Though Gaussian mixture model (GMM) and mel-frequency cepstral coefficients (MFCCs) yielded better recognition accuracy in the case of crackle, it was when MFCCs were used along with artificial neural network (ANN), that comparatively better results for wheeze and normal lung sounds were registered [7]. Over the years, a number of studies have been published where wavelet and cepstral features are used for this task [8, 8–15]. Autoregressive (AR) coefficients, have also been used in certain cases [16–21]. But being highly sensitive to numerical precision, they were transformed into cepstral coefficients [22]. Another frequently used set of features for lung status recognition are those derived from the power spectrum of the signal [23–26, 26–29]. Yet, being overly sensitive to noise, those power-spectrum based features show poor results.

Motivated by the great success of cepstral features in speech pattern classification, in our work, we explore the cepstral features further for lung sound recognition. In speech processing, they are compact representation of spectral envelope that separates source and filter [30]. The source signal is due to the vibrations of the vocal cords whereas the vocal tract acts as a filter [22]. In the context of lung sounds, the chest wall or thoracic wall acts as a filter through which the lung sounds sift through [31, 32]. Besides, cepstral coefficients are suitable for pattern recognition task for they are uncorrelated with each other.

In our study, we consider three types of lung sounds: normal, wheeze and crackle. Towards this end, five popular cepstral features are studied falling under two broad categories: (i) features based on all-pole modeling, and (ii) features based on filterbank method. Under the first head, \textit{linear prediction cepstral coefficient} (LPCC) and its perceptual form, \textit{perceptual linear prediction coefficient} (PLPCC) [33] are considered. In the case of filterbank-based approach, \textit{linear frequency cepstral coefficient} (LFCC), its perceptual form, \textit{mel-frequency...
and a complementary feature of MFCC, i.e., \textit{inverted mel-frequency cepstral coefficient} (IMFCC) \cite{35} are studied. Next, we have proposed a new feature extraction method that uses statistical properties of the cepstral coefficients. The proposed feature outperforms the wavelet-based features in terms of classification accuracy. In addition to that, the proposed features require significantly less time for classification task as compared to the baseline cepstral features. Finally, we optimize different parameters for our proposed features and evaluate their performance in presence of different noises.

The rest of the paper is organized as follows. Section 2 briefly describes the database used in our study. In Section 3, short-term spectral characteristics of lung sound are studied. In Section 4, cepstral features are described for the completeness of this paper. In Section 5, we study the statistical properties of cepstral coefficients and propose new features from this analysis. Experimental setups are described in Section 6. Results are discussed in Section 7. Finally, we conclude the paper in Section 8 by summarizing the work and by providing some future directions.

2. Database description

Systematic collection of lung sound samples with reliable ground truth is an important part of this research. Our database consists of recorded lung sounds obtained from three different resources: RALE database\(^1\), Audio and Bio-signal Processing Lab (IIT Kharagpur)\(^2\), Institute of Pulmocare and Research (Salt Lake, Kolkata)\(^3\). The sampling frequency of the original recording is 8000 Hz in all the three cases. The ground-truths were verified by experienced pulmonologists. The database includes three types of lung sounds: normal, crackle (both fine and coarse crackle) and wheeze. As the heart sound is mixed with the recorded lung sound, the lung sound separation is an essential step before the

\(\text{\footnotesize 1http://www.rale.ca/repository.htm}\)
\(\text{\footnotesize 2http://www.ecdept.iitkgp.ernet.in/index.php/home/labs/bio-sig-proc}\)
\(\text{\footnotesize 3http://www.pulmocareindia.org/}\)
samples are modeled and characterized. In this work, the effect of heart sound has been reduced from the collected data by adopting the method proposed in [36]. In this approach, the signal is first decomposed into intrinsic mode functions (IMFs) using empirical mode decomposition (EMD) technique. After that, an energy-based heart sound peak detection method is used in each IMFs within the LS signals followed by a boundary estimation algorithm to localize the spread of the heart sounds. Finally, localized heart sounds are reduced from the signal using filtering technique. The lung sound cycles were extracted using a Hilbert envelope based algorithm [37] and we use 72 cycles (24 from each class) for our experiments collected from 30 different subjects.

3. Spectral characteristics of lung sounds

Lung sounds can be broadly categorized into normal, abnormal, and adventitious [38, 39]. Normal lung sound refers to the respiratory sound of a healthy subject, hardly audible without a stethoscope. On the contrary, abnormal sound indicates the absence or decrease of normal sound [38] while adventitious sound appertains to the superposition of additional sound with normal sound. Adventitious sounds can again be continuous or discontinuous. Among the several adventitious lung sound instances we come across in the medical field viz. wheeze, crackle, cough, rhonchus, squawk, stridor and so on, two most common adventitious sounds wheeze (continuous) and crackle (discontinuous) are considered for detailed study in this paper. Often, either wheeze or crackle is found to be a characteristic feature of most of the other adventitious pulmonary sounds mentioned earlier. Acoustically, wheezes characterize cough sounds, even the asthmatic ones. Rhonchus is a low-pitched wheeze; squawk is a short wheeze preceded by a crackle; and stridor discernibly is low-frequency wheeze [39].

Wheeze: Wheeze is a continuous and adventitious waveform of duration 250 ms or more, perceptibly displaying a musical character [40]. Wheezing, usually the outcome of localized or diffused airway narrowing or obstruction in the passageway from the larynx to the small bronchi, can be induced by multiple
Figure 1: Time-domain characteristics and spectrogram of (a) normal, (b) wheeze, and (c) crackle lung sound cycle.

Figure 2: F-ratio of power-spectrum coefficients illustrating the separability of: (a) normal and wheeze, (b) normal and crackle, (c) wheeze and crackle, and (d) normal, wheeze and crackle.
causes like mucosal edema, external compression, partial obstruction by a tumor or foreign body, and so on [41]. The dominant frequency range typifying a wheeze normally exceeds 100 Hz [2, 41]. Wheezing is a symptom of diseases predominantly associated with the obstruction of airways. Whereas reversible obstruction characterizes asthmatic tendencies [42], irreversible obstruction denotes chronic obstructive pulmonary diseases (COPDs) [43], for example, emphysema (caused by smoking), chronic bronchitis and cystic fibrosis.

Crackle: Discontinuous in nature, crackles normally last for 1 to 10 ms [31, 40]. In general, crackles are caused by explosive opening of small airways [44]. They are short, explosive, non-musical sounds heard on inspiration and sometimes during expiration. It has a distinctly wide frequency range up to 2000 Hz [38, 44]. Two types of crackles are commonly found: fine and coarse. Fine crackles are high pitched and last for less than 5 ms as distinguished from coarse crackles, which are low pitched and have a duration of about 10 ms [45]. Crackles detected in patients can be symptomatic of alveolitis, interstitial lung diseases [46], congestive heart failure [47], carotid arterial stiffness [48], etc.

Various types of lung sounds have their different spectral characteristics [49]. From the spectrogram plot in Fig. 1, it can be observed that, crackles are distinguishable from normal and wheeze for their wide frequency range that usually extends up to 2000 Hz. This is also evident from the Fig. 1 that unlike crackle or wheeze, a normal respiratory sound has a narrower frequency range. Then we have done statistical analysis with power-spectrum coefficients to study the class-separability in different frequency components of short-term power spectrum. We use F-ratio based method which gives separability measure between the data of two or more classes [50, 51]. For a particular frequency component, it is computed as the ratio of between-class variance to within-class variance of power spectrum coefficients. For a given frequency component \( k \), it is defined as,

\[
F(k) = \frac{\sum_{i=1}^{C} \sum_{j=1}^{C_i} [\bar{S}_i(k) - \bar{S}(k)]^2}{\sum_{i=1}^{C} \frac{1}{C_i} \sum_{j=1}^{C_i} [S_{ij}(k) - \bar{S}_i(k)]^2}
\]
where $S^j_i(k)$ is the power spectrum coefficient of $j$-th sample ($j = 1, 2, ..., C_i$) of $i$-th class ($i = 1, 2, ..., C$) at frequency $k$. $\bar{S}_i(k)$ and $\bar{S}(k)$ are the mean of $k$-th frequency component of $i$-th class and all the classes respectively. We compute F-ratios for different frequency components for all pair of classes to analyze class separability in detail. All the lung sound cycles belonging to the three classes are used for this analysis. The variation of F-ratio with respect to frequency is shown in Fig. 2. The short-term power spectrum of each lung sound cycle is obtained using 20 ms frame length with 50% overlap and Hamming window.

The main observation from Fig. 2 is that short-term power spectrum in low-frequency region (i.e., from 0 Hz to 1000 Hz) contains more discriminative information than higher frequency region (i.e., from 1000 Hz to 2000 Hz). In pairwise F-ratio analysis, frequency content up to 1000 Hz is adequate to distinguish a signal between normal and wheeze class (Fig. 2(a)). However, entire frequency band (i.e., from 0 Hz to 2000 Hz) contains discriminative information for the separability of normal and crackle as shown in Fig. 2(b). We also notice that the discriminative information between wheeze and crackle are distributed in almost entire frequency band. This is also true when all the three classes are considered together as in Fig. 2(d).

From the above analysis, we can conclude that short-term features containing power spectrum information can be used to represent different lung sounds. Cepstral coefficients represent this spectral characteristics in a more robust manner. In the following section, we briefly review the computation method of several cepstral features.

4. Computation of cepstral features

Two all-pole modeling based features and three filterbank-based features are used in our study. Unlike wavelet-based features [9], cepstral features are based on short-term spectral analysis. Hence lung sounds are first divided into overlapping frames and features are computed in frame-wise manner.
4.1. All-pole modeling based features

All pole modeling is a method of representation of speech signal in terms of fewer parameters. It is based on autoregressive (AR) model of random process.

4.1.1. Linear prediction cepstral coefficients (LPCCs)

Short-term window of lung sound, \( s(n) \) (where \( n = 0, 1, 2, ..., M - 1 \) and \( M \) is the window size) is modeled using a \( q \)-th order linear prediction model:

\[
s(n) = \sum_{k=1}^{q} a(k) s(n-k) + e(n)
\]

where \( a(k) \) are the linear prediction coefficients (LPCs) and \( e(n) \) is the prediction error [52]. The value of \( q \) is chosen such that it can effectively capture the real and complex poles in the frequency range up to the Nyquist frequency. The predictor coefficients are determined by minimizing the prediction error in the mean-square sense. These LPCs typically are converted into a robust form popularly known as linear prediction cepstral coefficients (LPCCs) using the following recursive equation,

\[
c_0 = \ln G \\
c_m = a_m + \sum_{k=1}^{m-1} \left( \frac{k}{m} \right) c_k a_{m-k}, \quad 1 \leq m \leq q
\]

where \( G \) is the gain parameter of the LPC model. LPCCs are traditionally features in speech and speaker recognition [52, 53].

4.1.2. Perceptual linear prediction cepstral coefficients (PLPCCs)

In this case, lung sound is first converted into perceptually meaningful signal through several processing steps mimicking human auditory perception [33, 54, 55]. In the first step, the power-spectrum is integrated using critical band filters placed at an interval of one unit in psychoacoustical Bark scale [56]. Next, the integrated spectrum is weighted using equal loudness curve approximating
LFCC: Linear frequency cepstral coefficient
MFCC: Mel-frequency cepstral coefficient
IMFCC: Inverted mel-frequency cepstral coefficient

Figure 3: Illustration of filter bank frequency responses for (a) LFCC, (b) MFCC, and (c) IMFCC.

the non-linear sensitivity of human auditory system. Then, cube-root compression is performed on the weighted-energy coefficients. Finally, LPCCs of this coefficients are computed as PLPCCs.

4.2. Filterbank-based cepstral features

Filterbank-based parametrizations are widely accepted than the LP-based methods due to their computational advantage using fast Fourier transform (FFT). Three types of such features are studied in this work.

4.2.1. Linear frequency cepstral coefficients (LFCCs)

In order to compute the filterbank-based features, short-term power spectrum is computed from the lung sound frame, \( s(n) \), using \( N_{FFT} \)-point FFT
\[ |Y(k)|^2 = \left| \sum_{n=1}^{M} s(n) \cdot e^{\frac{j2\pi nk}{\text{FFT}}} \right|^2, \quad (5) \]

where, \(1 \leq k \leq \frac{\text{FFT}}{2} + 1\). Next, filterbank with linearly spaced triangular filters in the Hz scale is imposed on the spectrum as shown in Fig. 3(a). The outputs \(\{e(i)\}_{i=1}^{Q}\) of the linearly spaced band-pass filters can be calculated by a weighted summation between respective filter response \(\psi_i(k)\) and the energy spectrum \(|Y(k)|^2\) as:

\[ e(i) = \sum_{k=1}^{\frac{\text{FFT}}{2}+1} |Y(k)|^2 \cdot \psi_i(k) \quad (6) \]

Finally, discrete cosine transform (DCT) is performed on the log filterbank energies \(\{\log[e(i)]\}_{i=1}^{Q}\) and the \(m\)-th LFCC coefficient is computed by written as:

\[ C_m = \sqrt{\frac{2}{Q}} \sum_{l=0}^{Q-1} \log[e(l+1)] \cdot \cos \left[ m \cdot \left( \frac{2l - 1}{2} \right) \cdot \frac{\pi}{Q} \right]. \quad (7) \]

### 4.2.2. Mel-frequency cepstral coefficients (MFCCs)

The computation of MFCC is almost the same as LFCC except the fact that here the filters are spaced in non-linear in mel scale [57, 58]. According to the psychophysical studies, human perception of the frequency content of sounds follows this nonlinear scale defined as,

\[ f_{\text{mel}}(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right) \quad (8) \]

where \(f\) is the original frequency in Hz. Here filters are spaced more closely in low-frequency region than high-frequency region as shown in Fig. 3(b). We expect that MFCCs will perform better than LFCCs as more emphasize is given on highly discriminative section of power spectrum.
4.2.3. Inverted mel-frequency cepstral coefficients (IMFCCs)

IMFCCs are computed with inverted mel scale. Here, the filters are more densely spaced in higher-frequency region [35] as illustrated in Fig. 3(c). IMFCCs capture more detailed information from the high frequency region. From the F-ratio analysis in Fig. 2, we have seen that high-frequency cues could be useful in discriminating normal and crackle as well as wheeze and crackle lung sounds. For this reason, we study IMFCC features in lung sound classification problem.

5. Proposed statistical features and class-separability analysis

5.1. Feature formulation

Standard cepstral features are computed in a frame-wise manner. In speech processing applications, such as speech and speaker recognition, they are expressed as mixture of multivariate Gaussian distributions with GMMs since each mixture approximately represents one acoustic class [59]. GMM is suitable there as different speech sounds belong to different acoustic class. Here, we hypothesize that each lung sound cycle of a particular disease class is produced due to a specific physical structure of lung, and hence each cycle can be modeled with a unimodal multivariate normal distribution function, i.e., a single Gaussian distribution per cycle.

First we study the histogram of cepstral features for different categories of lung sounds for full database (i.e., considering all the cycles). A representative figure is illustrated in Fig. 4. It can be observed that they approximately follow the normal distribution with different mean and variance. We then perform Jarque-Bera test to check the normality of the cepstral features within a cycle. This uses the values of skewness and kurtosis of distribution curve to check the normality [60]. It gives a test decision whether a data comes from a normal distribution with an unknown mean and variance or not. We have found that in most cases the cepstral coefficients follow normal distribution. The plot of p-values for different features are shown in Fig. 5. We have also used normal
We choose MFCCs for this representative plot. Probability plot, a graphical technique to verify if the cepstral data follows normal distribution or not [61]. If all the data points approximate the straight line, then corresponding features follow normal distribution. Figure 6 is the plot of a single dimension of cepstral coefficients of all the frames of an arbitrary chosen cycle. We notice from the illustration that the cepstral coefficients approximately follow normal distribution.

From the above analysis, we can conclude that cepstral features from each cycle can be approximated with normal distribution. This leads us to define the probabilistic model of the cepstral coefficients of \( k \)-th cycle as,

\[
p_k(x) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right\}
\]  \hspace{1cm} (9)

where \( \mu \) and \( \Sigma \) are respectively the mean and covariance of the the \( d \)-dimensional feature vector \( x \), where \( d \) is the number of coefficients. Hence, we can parameterize the cepstral features of each cycle with their sample mean.
Figure 5: The illustration of p-values obtained from Jarque-Bera test of different feature set at the 5% significance level. Plot of p-values are shown for all 24 cycles of each class for each dimension of feature vector.

Figure 6: A representative figure showing normal probability plot.
and variance across frames. We call these as statistical features and they are computed as follows.

Let $T$ be the number of frames in a given lung sound cycle and $N_F$ be the number of coefficients, we compute mean, $\mu$ and standard deviation, $\sigma$ (i.e., square-root of the diagonal elements of $\Sigma_k$) of the cycle as,

\[ \mu_i = \frac{1}{T} \sum_{j=1}^{T} x_{ij} \] (10)

and

\[ \sigma_i = \left[ \frac{1}{T} \sum_{j=1}^{T} (x_{ij} - \mu_i)^2 \right]^{1/2} \text{ where } i = 1, 2, \ldots N_F \] (11)

where $x_{ij}$ is the $i$-th cepstral coefficient of $j$-th frame of the cycle. Note that as the cepstral coefficients are uncorrelated and hence, we can approximate the covariance matrix as diagonal [34, 59]. We propose to use this mean and standard deviation as the features for lung sound classification. The block diagram of this feature extraction method is shown in Fig. 7. The main advantage of this formulation is that frame-wise cepstral features can be represented as a single vector and it will lead to a faster training and classification for cepstral features. In the next subsection, we analyze the suitability of our proposed statistical features for class separability.

5.2. Class separability analysis using proposed features

It is of interest to see which type of the statistical feature – mean or standard deviation – has greater influence on the lung sound classification. Figure 4 shows

![Figure 7: Block diagram of proposed feature extraction method.](image-url)
the distribution plot of an arbitrarily chosen cepstral feature. We observe that means relatively more different than the standard deviation parameter. As is shown in the figure, the numerical values of standard deviation, i.e, square root of variance of baseline cepstral features, show smaller differences than the numerical values of the mean among classes. When comparing normal and wheeze sounds, we find that their standard deviations are 0.0677 and 0.0617, respectively, whereas their means are -0.0212 and 0.0391. Figure 8 shows average of statistical coefficients for different classes. We notice in Fig. 8 that crackle is significantly separable in case of standard deviation of MFCCs for all the coefficients. On the other hand, wheeze and normal are not separable. However, when using mean-based features of MFCCs, lower-order coefficients (particularly from 3 to 11) are distinguishable for different classes.

To quantify the discriminative ability of the two distribution parameters, we have calculated the *Rayleigh Quotient* [62, p. 120]. This is a measure of compactness within each class and separability among classes and it is used in linear discriminant analysis (LDA). In LDA, higher dimensional data are projected to lower dimensional space to maximize the separability in the projected space. The direction of projection is determined such that the Rayleigh Quotient is maximized. We have computed this measure for both features separately. It is computed by,

\[ J(w) = \frac{|w^T S_B w|}{|w^T S_W w|} \]  \hspace{1cm} (12)

where \( w \) is the direction of projection that maximizes the discrimination among different classes, \( S_B \) and \( S_W \) are the between-class and within-class scatter matrices before projection respectively. \( w \) is evaluated by solving the following generalized eigenvalue problem,

\[ S_B w = \Lambda S_B w \]  \hspace{1cm} (13)

where \( \Lambda \) is a diagonal matrix containing eigenvalues as the diagonal elements.
Better separability among classes and compactness within each class is possible by maximizing $J(w)$ [62].

We have calculated this Rayleigh Quotient for both type of statistical features. For mean-based, we get 0.79 and for standard deviation based, it is 0.47 on full database. Therefore, we can say that mean-based features are more distinctive than standard deviation based features. These observations are in agreement with the results in Table 1 and Table 2 as discussed in Section 7.

6. Experimental setup

6.1. Pre-processing & feature extraction

In the pre-processing stage, signals are first down-sampled at 4000 Hz as most of the lung sound information exists within 2 kHz. Then we perform amplitude normalization to reduce the instrumental noises. We have used 20 ms long frame length with 50% overlap for extracting short-term features. We use 1024-pt FFT for computing FFT-based power spectrum. As we use 20 filters for spectral analysis and all-pole modeling, the feature dimension is 20 for LFCCs, MFCCs,
IMFCCs and LPCCs. But for PLPCCs as 13 filters are used in bark scale for the range 0-2000 Hz, the feature dimension is 13. Sample mean and standard deviation of each with each set of coefficients are calculated across frames as in Eq. (10) and (11). This leads to 20 dimensional feature vector for LPCC, LFCC, MFCC, and IMFCC and 13 dimensional feature vector for PLPCC. For better efficiency, prior to classification, feature vectors are normalized between [-1,1] [2].

6.2. Classifier and performance evaluation

ANN with multilayer perceptron (MLP) has been used as classifier [63, 64]. This consists of 40 hidden layer nodes and 3 output layer nodes. Resilient back propagation (RP) [2, 65] is used here for training ANN. Activation functions of hidden layer and output layer chosen are tan-sigmoid and log-sigmoid for better classification accuracy and efficiency [2]. We have iterated the whole classification method 25 times and average accuracy is calculated.

We use leave-one-out cross validation to evaluate our methodology. In leave-p-out cross-validation, p observations are used as the test set and the remaining observations in the training class models [18, 66]. In our experiments, one of the available 72 cycles is used as test sample while the remaining are used for training. This method is repeated 72 times to test cycles. The classification accuracy is computed as,

\[
\text{Classification accuracy} = \frac{\text{No. of correctly classified cycles}}{\text{Total no. of cycles under test}} \times 100.
\]  

(14)

We further compute the sensitivity and specificity metric [67]. Sensitivity measures the proportion of abnormal cycles that are correctly identified as abnormal and can be written as,

\[
\text{Sensitivity} = \frac{\text{No. of correctly classified abnormal cycle}}{\text{Total no. of abnormal cycles under test}} \times 100.
\]  

(15)
Table 1: Lung sound classification performance for wavelet and baseline cepstral features. Classification accuracy (in %) is shown for crackle, wheeze and normal lung sounds. Sensitivity, specificity and overall accuracy are also computed. The last column shows the computational time (in minutes).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Crackle</th>
<th>Wheeze</th>
<th>Normal</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Overall</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelet (31.25-1000 Hz)</td>
<td>100.00</td>
<td>91.00</td>
<td>74.83</td>
<td>95.91</td>
<td>74.83</td>
<td>88.61</td>
<td>4.22</td>
</tr>
<tr>
<td>Wavelet (0-2000 Hz)</td>
<td>100.00</td>
<td>90.66</td>
<td>82.83</td>
<td>95.50</td>
<td>82.83</td>
<td>91.16</td>
<td>3.76</td>
</tr>
<tr>
<td>Baseline LPCC</td>
<td>100.00</td>
<td>95.83</td>
<td>91.67</td>
<td>97.91</td>
<td>91.67</td>
<td>95.83</td>
<td>723.56</td>
</tr>
<tr>
<td>Baseline PLPCC</td>
<td>100.00</td>
<td>95.83</td>
<td>91.67</td>
<td>97.91</td>
<td>91.67</td>
<td>95.83</td>
<td>691.36</td>
</tr>
<tr>
<td>Baseline LFCC</td>
<td>100.00</td>
<td>95.83</td>
<td>91.67</td>
<td>97.91</td>
<td>91.67</td>
<td>95.83</td>
<td>717.70</td>
</tr>
<tr>
<td>Baseline MFCC</td>
<td>100.00</td>
<td>95.83</td>
<td>91.67</td>
<td>97.91</td>
<td>95.83</td>
<td>97.20</td>
<td>725.50</td>
</tr>
<tr>
<td>Baseline IMFCC</td>
<td>100.00</td>
<td>95.83</td>
<td>87.5</td>
<td>97.91</td>
<td>87.5</td>
<td>94.44</td>
<td>730.98</td>
</tr>
</tbody>
</table>

LPCC: Linear prediction cepstral coefficient
PLPCC: Perceptual linear prediction cepstral coefficient
LFCC: Linear frequency cepstral coefficient
MFCC: Mel-frequency cepstral coefficient
IMFCC: Inverted mel-frequency cepstral coefficient.

On the other hand, specificity measures the proportion of normal cycles that are correctly identified as normal and it can be expressed as,

\[
\text{Specificity} = \frac{\text{No. of correctly classified normal cycle}}{\text{Total no. of normal cycles under test}} \times 100.
\]

(16)

6.3. System setup

All algorithms are implemented using MATLAB (Version 2013b) programming language. The experiments are conducted with a Lenovo MT-M 3492-H20 desktop computer using Intel(R) Core(TM) i5-347OS CPU, 8GB RAM and Windows 7 operating system.

7. Results & discussions

7.1. Comparison with baseline methods

We have conducted experiments with our proposed features as well as wavelet-based approach [2]. First, wavelet coefficients are calculated using Daubechies
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Feature</th>
<th>Crackle</th>
<th>Wheeze</th>
<th>Normal</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Overall</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>LPCC</td>
<td>99.83</td>
<td>95.66</td>
<td>95.00</td>
<td>97.91</td>
<td>95.00</td>
<td>96.80</td>
<td>3.59</td>
</tr>
<tr>
<td></td>
<td>PLPCC</td>
<td>100.00</td>
<td>93.50</td>
<td>91.00</td>
<td>97.25</td>
<td>91.00</td>
<td>94.83</td>
<td>3.70</td>
</tr>
<tr>
<td></td>
<td>LFCC</td>
<td>99.50</td>
<td>95.00</td>
<td>98.30</td>
<td>97.41</td>
<td>98.43</td>
<td>97.61</td>
<td>3.69</td>
</tr>
<tr>
<td></td>
<td>MFCC</td>
<td>99.67</td>
<td>94.50</td>
<td>97.50</td>
<td>97.41</td>
<td>97.50</td>
<td>97.20</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td>IMFCC</td>
<td>100.00</td>
<td>89.16</td>
<td>87.67</td>
<td>95.41</td>
<td>87.66</td>
<td>92.27</td>
<td>3.99</td>
</tr>
<tr>
<td>Standard</td>
<td>LPCC</td>
<td>94.00</td>
<td>73.00</td>
<td>73.83</td>
<td>88.58</td>
<td>73.83</td>
<td>80.27</td>
<td>7.02</td>
</tr>
<tr>
<td>Deviation</td>
<td>PLPCC</td>
<td>98.50</td>
<td>70.60</td>
<td>72.16</td>
<td>84.66</td>
<td>72.16</td>
<td>80.44</td>
<td>3.12</td>
</tr>
<tr>
<td></td>
<td>LFCC</td>
<td>96.60</td>
<td>77.16</td>
<td>83.16</td>
<td>90.25</td>
<td>83.16</td>
<td>85.6</td>
<td>3.31</td>
</tr>
<tr>
<td></td>
<td>MFCC</td>
<td>95.83</td>
<td>76.50</td>
<td>68.30</td>
<td>87.67</td>
<td>68.33</td>
<td>80.22</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td>IMFCC</td>
<td>97.66</td>
<td>77.30</td>
<td>71.16</td>
<td>90.00</td>
<td>71.17</td>
<td>82.05</td>
<td>3.99</td>
</tr>
</tbody>
</table>

LPCC: Linear prediction cepstral coefficient
PLPCC: Perceptual linear prediction cepstral coefficient
LFCC: Linear frequency cepstral coefficient
MFCC: Mel-frequency cepstral coefficient
IMFCC: Inverted mel-frequency cepstral coefficient.

Table 2: Lung sound classification performance for proposed statistical features. Classification accuracy (in %) is shown for crackle, wheeze and normal lung sounds. Sensitivity, specificity and overall accuracy are also computed. The last column shows the computational time (in minutes).

mirror filters of order 8 (db8) with five levels, encompassing a frequency range of 31.25-1000 Hz. Then the mean of absolute subbands, standard deviation, average power and ratio of computed means of adjacent subbands are calculated to formulate 19-dimensional feature vectors. These are used with neural network classifier. For an even more appropriate comparison, the frequency range is changed to 0-2000 Hz. Here, the feature vector dimension is 27. These are again applied to neural network classifier using the same configuration. To evaluate the performance of baseline cepstral features, frame-wise cepstral coefficients are separately considered as input to the neural network. The output of the neural network is combined across all the frames to determine the target class for each lung sound cycle. The classification results of all the baseline methods are shown in Table 1. Table 1 demonstrates that cepstral features perform better than the wavelet-based features. This is possibly due to the fact that short-term spectral characteristics can represent the lung sound in-
Table 3: Effect of dynamic coefficients on lung sound classification accuracy for proposed MFCC features.

<table>
<thead>
<tr>
<th></th>
<th>Static</th>
<th>Static + Δ</th>
<th>Static + Δ + ΔΔ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crackle (%)</td>
<td>99.67</td>
<td>100.00</td>
<td>99.50</td>
</tr>
<tr>
<td>Wheeze (%)</td>
<td>94.50</td>
<td>91.50</td>
<td>91.30</td>
</tr>
<tr>
<td>Normal (%)</td>
<td>97.50</td>
<td>99.00</td>
<td>94.80</td>
</tr>
<tr>
<td>Overall (%)</td>
<td>97.20</td>
<td>96.80</td>
<td>95.20</td>
</tr>
</tbody>
</table>

formation in a more effective manner unlike statistical measures of lung sound signals in wavelet-domain. Beside, cepstral features are more robust in presence of nuisance variation in the signal. We also study this in Section 7.3. In case of sensitivity and specificity, cepstral features outperform the baseline wavelet based features. Further from Table 1, we also observe that among all the studied cepstral features, baseline MFCCs yield the best results. This could be due to the fact that MFCCs capture information relevant to the auditory perception. In medical diagnosis field, the expert pulmonologists are able to detect the differences between normal and adventitious sounds by listening it through stethoscope. Therefore, auditory perception specific information captured by MFCCs are suitable for the recognition purpose. On the other hand, baseline IMFCCs give the worst performance among the cepstral features. The reason behind this may be that resolution is more in high-frequency than that of low frequency in inverted mel scale. Due to the higher-spacing of filters at lower-frequencies, this feature fails to capture the discriminative information of the three classes as discussed in Section 3. We also observe that crackle sounds are accurately classified using all the studied features whereas identification of normal and wheeze are relatively difficult.

Results in Table 2 shows lung sound recognition performance with proposed statistical features. We found that the proposed mean feature gives better results than standard deviation feature. The performance of proposed mean feature is also comparable to the performance obtained using baseline cepstral coefficients. We also note that the computational time required for our proposed
Figure 9: Effect of frame length on lung sound recognition accuracy for proposed MFCC features.

approach for training and testing is significantly reduced as compared to the baseline cepstral features. For instance, it requires 725.50 minutes for baseline MFCC features where as for its statistical counterpart, the computational time is 3.43 minutes. This is expected as the features are evaluated in a frame-wise manner for the baseline method. This means if a lung cycle consists of $T$ frames, then $T$ data-samples per cycle are processed for training and test. On the other hand, the proposed features are represented by a single vector similar to the wavelet features, and their dimensions are equal to the number of filters in the filterbank. The standard deviation based features gives poor performance because of their lower discriminative capability as discussed in Section 5.2. Note that the proposed features are studied so far for an ad hoc configuration used in speech processing. In the next section, we optimize the configuration parameters of this features for lung sound recognition.
7.2. Experimental optimization of cepstral features

7.2.1. Effect of delta and double-delta features

Dynamic features of MFCC - delta and double-delta features - include temporal information, i.e., the trajectories of MFCCs over time. Proposed MFCCs along with appended delta and double-delta features are considered for lung sound classification. The experimental results are shown in Table 3. It is observed that MFCCs as stand-alone features show higher accuracy than when combined with any of the delta or double-delta coefficients. Noticeably, double-delta features show even lower performance improvement than delta features. We conclude that temporal information of our proposed features are not useful for lung sound recognition. Therefore, in our subsequent experiments, the delta and double-delta features are not considered.

7.2.2. Optimization of frame length

The short-term features from speech signals are usually extracted with a window of 20 ms. In this section, we analyze how the variation in frame lengths during the process of computing MFCCs can affect the lung sound recognition accuracy. We conduct experiments by varying frame lengths between 10 ms to 200 ms. Figure 9 illustrates the effect of varying the frame length on the overall classification accuracy and also the accuracies for each of the three classes. We observe that classification accuracy of crackle is consistently better than normal or wheeze accuracy for different length of analysis window. Lung sound detection accuracy in the case of wheeze is the lowest compared to the other two. Further, we have found that with the variation in frame length, the individual accuracies for normal, wheeze and crackle sounds are not much affected. The accuracy of normal sound is found to be the best (i.e., 99.16%) at a window length of 50 ms. On the other hand, we obtain best recognition accuracy of wheeze obtained as 95.17% at slightly higher frame length (i.e., 70 ms). In Fig. 9, the solid line representing overall accuracy attains maximum of 97.83% at a frame length of 50 ms.
Figure 10: A representative figure showing the noise effect (Additive white Gaussian of 10 dB SNR) on wavelet-based baseline (in left) and sample mean based proposed (in right) features for normal, wheeze, and crackle sounds.
In real-world, lung sounds can be distorted with noise due to the transmission channel effects, errors in data acquisition, etc. In this section, we evaluate the performance of proposed and baseline features in presence of three additive noises: white Gaussian, babble and volvo noise. Figure 10 illustrates the effect of white Gaussian noise with 10 dB SNR on the proposed mean-based statistical features and wavelet-based features for the three classes. We observe that proposed mean-based features are relatively more robust than wavelet-based features. We evaluate the performance in presence of the those types of noises by artificially degrading the test sample for different SNR levels (0 dB–40 dB). The comparative results are shown in Table 4. For white Gaussian noise, the proposed feature outperforms wavelet-based features for low SNR (less than 20 dB). On the other hand, it yields better results in presence of babble noise for higher SNR (greater than 10 dB). Furthermore for volvo noise, we have obtained improved performance at all SNR levels. Interestingly for white noise, the performance of the proposed features drops suddenly at 20 dB SNR. This could be possibly due to the regularization effect as a result of noise contamination [68]. Considering the overall performance in all the three noisy conditions, we conclude that the proposed mean-based features outperform the wavelet-based features.

Table 4: Performance (overall accuracy in %) of lung sound classification with wavelet and mean-based proposed features (computed from MFCCs) in presence of additive white Gaussian noise, babble noise and volvo noise.

<table>
<thead>
<tr>
<th>Noise</th>
<th>Feature</th>
<th>0 dB</th>
<th>10 dB</th>
<th>20 dB</th>
<th>30 dB</th>
<th>40 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>Wavelet</td>
<td>35.49 ± 0.22</td>
<td>49.03 ± 0.32</td>
<td>83.73 ± 0.28</td>
<td>89.16 ± 0.22</td>
<td>90.48 ± 0.09</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>63.31 ± 1.05</td>
<td>84.90 ± 1.13</td>
<td>75.35 ± 0.98</td>
<td>85.90 ± 1.06</td>
<td>94.98 ± 0.30</td>
</tr>
<tr>
<td>Babble</td>
<td>Wavelet</td>
<td>55.07 ± 1.34</td>
<td>83.82 ± 0.49</td>
<td>87.87 ± 1.38</td>
<td>89.89 ± 0.41</td>
<td>90.49 ± 0.10</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>37.92 ± 1.01</td>
<td>67.04 ± 2.40</td>
<td>88.09 ± 1.47</td>
<td>97.16 ± 0.42</td>
<td>97.83 ± 0.17</td>
</tr>
<tr>
<td>Volvo</td>
<td>Wavelet</td>
<td>59.58 ± 1.26</td>
<td>87.77 ± 0.82</td>
<td>89.60 ± 0.36</td>
<td>90.17 ± 0.15</td>
<td>90.52 ± 0.25</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>75.72 ± 0.89</td>
<td>93.33 ± 3.69</td>
<td>97.57 ± 0.10</td>
<td>97.74 ± 0.11</td>
<td>97.73 ± 0.12</td>
</tr>
</tbody>
</table>
8. Conclusions

Acoustic signals produced by the lungs during inspiration and expiration render useful information regarding the lung status. In this study, the competency of the statistical parameters of cepstral features in recognizing three types of lung sounds is investigated. Irrespective of subjective factors like auditory sensitivity or the physician’s inexperience, procuring and utilizing spectral information related to lung sounds and thereby segregating different lung sounds can be done efficiently. Conventional methods of classification that use the wavelet-based method or baseline MFCC based method are, no doubt, satisfactory. But cepstral features not only yield good results compared to wavelet but also, the statistical features from cepstral coefficient consume much less computational overhead in comparison to the baseline cepstral features. The proposed features are further optimized and their performance in the presence of three different additive noises are also studied separately. We have found that cepstral features are better than wavelet-based features in terms of classification accuracy. The statistical features extracted from cepstral coefficients are good to approximate the class-specific information and they reduce the computational time appreciably. The sample mean is found to be more suitable for classification than the standard deviation. During the experimental optimization of the proposed features, we observe that dynamic features are not useful for lung sound recognition. We have also found that the optimum size of analysis window for short-term feature extraction of lung sound is moderately longer than what frequently used in speech analysis. In the experiments with noisy lung sounds, we have found that the proposed optimized features outperform existing features at most of the noise-levels studied in this paper.

In the present study, we have emphasized on the formulation of feature vectors and their experimental optimization whereas the classifier is fixed. As a continuation of this work, research will be conducted to find the best classifier suitable with proposed features. The proposed feature can also be studied for the analysis of other biomedical signals such as heart sound.
Conflict of interest statement

The authors declare that they have no conflict of interest.

Acknowledgement

The work was supported by Ministry of Human Resource & Development, Govt. of India. We are thankful to Dr. Ashok Mondal for his help in preparing the database for the experiments. We would also like to thank the reviewers for their careful reading of the manuscript and helpful comments. The work of the second author was partially supported Academy of Finland grants (proj. no. 253120 and 283256).

References


