Performance evaluation and enhancement of lung sound recognition system in two real noisy environments

Gwo-Ching Chang\textsuperscript{a,*}, Yung-Fa Lai\textsuperscript{b}

\textsuperscript{a} Department of Information Engineering, I-Shou University, No. 1, Sec. 1, Syuecheng Rd., Dashu Township, Kaohsiung County 840, Taiwan, ROC
\textsuperscript{b} Chest Division, Department of Internal Medicine, E-Da Hospital & I-Shou University, Kaohsiung, Taiwan, ROC

\textbf{ARTICLE INFO}

Article history:
Received 22 September 2008
Received in revised form
19 April 2009
Accepted 18 June 2009

Keywords:
Lung sound recognition
Autoregressive coefficients
Mel-frequency cepstral coefficients
Dual sensor spectral subtraction algorithm

\textbf{ABSTRACT}

This study investigates the problems associated with lung sound recognition under noisy conditions. Firstly, the effects of noise on the lung sound feature representation and the classification performance are analyzed. Two kinds of feature representations, autoregressive and mel-frequency cepstral coefficients, are used to characterize the lung sound signals. Dynamic time warping is used to categorize the lung sounds to one of the three: normal, wheezes, or crackles. Our experimental results indicate that additive noise produces a mismatch between training and recognition environments and deteriorates the classification performance with a decrease in the SNR levels. In order to compensate the degrading effect of noise on the lung sound recognition, a dual sensor spectral subtraction algorithm is applied to the lung sound signals before the extraction of lung sound features. It is observed that the proposed algorithm is capable of providing adequate performance in terms of noise suppression and lung sound signal enhancement.

\section{Introduction}

Lung sounds that are heard with the aid of a stethoscope can be classified into two categories: normal breathing sounds (when no respiratory disorders exist) and adventitious (abnormal) sounds (when a respiratory disease exists). Adventitious breathing sounds refer to extra or additional sounds during a regular breathing cycle. Two adventitious sounds that are commonly associated with respiratory disorders are wheezes and crackles \cite{1}. Wheezes are continuous musical sounds that are longer than 250 ms; they are high pitched, with a dominant frequency of 400 Hz or more \cite{1}. Crackles are discontinuous sounds that are usually shorter than 20 ms and show a wide frequency spectrum between 200 and 2000 Hz \cite{1}.

Lung sounds are highly non-stationary stochastic signals owing to the changing air flow rate and pulmonary volumes during a respiration cycle. Because of the stochastic nature of the lung sound, many parametric modeling methods and power spectral methods have widely been applied to characterize information on respiratory sound signals for computer-aided lung sound recognition, such as autoregressive coefficients (AR) \cite{2–4}, linear prediction cepstral coefficients (LPCC) \cite{5}, mel-frequency cepstral coefficients (MFCC) \cite{6,7}, wavelet coefficients \cite{8,9}, spectrograms \cite{10–12}, and higher-order spectra \cite{13}. Based on the specific feature representation, various pattern recognition algorithms, such as k-nearest neighbor algorithm \cite{4}, neural networks \cite{5,9} and Gaussian mixture model \cite{7}, and 2D bilateral filter \cite{14}, have been developed in lung sound classification. Another
classifier, dynamic time warping (DTW), is a technique for finding an optimal match between two sequences with certain non-linear variations in time or space. It has been successfully applied to speech recognition for a long time [15]. It is expected that DTW may be suitable for lung sound recognition.

Generally, the actual environments of lung sound recognition are noisy. Lung sound recognition may be difficult in noisy environments such as those in an ambulance, a busy emergency room, or a medical assistance airplane. The reduced clarity of lung sounds could render the use of lung sound recognition technology in applications in real environments impractical. Most of the previous studies pertaining to the classification of lung sounds did not analyze the influence of noisy environments on the experimental results. Therefore, for the lung sound recognition system to be used as a reliable bedside monitoring equipment, more researches are required to evaluate and enhance the classification performance of the system in real noisy environments.

In this study, we investigated the problems that arise when lung sound recognition is performed under noisy conditions. Firstly, we studied the effect of noise on the lung sound representation and the classifier performance. Secondly, we propose a dual sensor spectral subtraction method for compensating the effect of noise on the lung sound representation. To evaluate the effect of noise on the lung sound representation, we added two types of noise (babble noise for simulating the interference in a busy emergency room environment and ambulance car noise for simulating the disturbance in an ambulance) to each lung sound signal in the test database. We compared the performance of the methods that use the AR coefficients with those that use the MFCC coefficients to characterize the lung sound signal. Dynamic time warping was performed to categorize the lung sounds to one of three: normal, wheezes, or crackles.

### 2. Materials and methods

#### 2.1. Data collection

Normal respiratory sounds were collected from twelve healthy subjects. Wheeze lung sounds were recorded from thirteen patients with asthma or chronic obstructive pulmonary disease. Crackles lung sounds were obtained from eleven patients with pulmonary fibrosis, pneumonia, or congestive heart failure. All wave files were divided into 298 respiratory sound samples in the test database. We down sampled the noise sample from 19.98 kHz to 8 kHz and then added them artificially to the lung sound signals in the testing set for SNR values of 40, 30, 20, 10, 5, 0 dBs.

#### 2.2. Lung sound processing and classification

The end point detection based on volume was applied to extract the region of interest from the raw lung sound signal. When the predefined threshold was exceeded, the signal was then subdivided into frames of 512 samples and overlapped at 50%. Each frame was analyzed to extract the 8-order AR coefficients and the 13-order MFCC coefficients. The two types of feature sets were tested using DTW.

The leave-one-out method [17] was used to estimate the recognition rate. Each time a sample was tested, the remaining samples were used as the training samples. Every sample in the entire sample set was tested once. Since every test sample was excluded from the training sample set, the test set and the training set were independent of each other. In the following subsections, the lung sound processing and classification methods are briefly described.

##### 2.2.1. Autoregressive model

In the AR model, each sample $x(k)$ of the lung sound signal is described as a linear combination of the previous samples plus an error term $e(k)$ that is independent of the past samples. A pure AR time-series model is given as follows:

$$x(k) = \sum_{i=1}^{p} a_i x(k-i) + e(k), \quad k = 0, 1, 2, \ldots$$

where $x(k)$ denotes the samples of the modeled signal, $a_i$ denotes the AR coefficients, $p$ is the order of the AR model and residual $e(k)$ represents a white noise sequence. The AR coefficients are obtained by minimizing the variance of $e(k)$ [18]. Lung sound recognition is then reduced to identifying the parameters $a_i (i=0, 1, 2, \ldots, p)$.

##### 2.2.2. Mel-frequency cepstral coefficients

The mel-frequency cepstral coefficients are the inverse discrete cosine transform of the logarithm of the short-term power spectrum of a signal [6]. They are defined as

$$c_n = \sum_{k=1}^{M} \log(E_k) \cos \left( n(k - 0.5) \frac{\pi}{M} \right), \quad n = 1, 2, \ldots, L$$

where $M$ is the number of subbands and $L$ is the desired length of the cepstrum. $E_k$ is the filter bank energy after the signal passes through a triangular bandpass filter. The frequency bands are computed using the mel-frequency scale.

##### 2.2.3. Dynamic time warping

Suppose we have two vectors – a vector $T$ of length $n$ and a vector $R$ of length $m$ – where

$$T = t_1, t_2, \ldots, t_i, \ldots, t_n$$

$$R = r_1, r_2, \ldots, r_j, \ldots, r_m$$

To classify these two vectors using DTW, we first construct an $n \times m$ matrix in which the $(i, j)$th element of the matrix corresponds to the squared distance $d(t_i, r_j) = (t_i - r_j)^2$ between the two points $t_i$ and $r_j$. The best match between these two
vectors can be found through the following recursive equation [19]:

\[
D(i, j) = d(t_i, r_j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\}
\]  (5)

2.2.4 Dual sensor spectral subtraction method
The spectral subtraction method was first proposed by Boll [20], and it has widely been utilized in various enhancement applications such as speech and audio enhancement in addition to being used for the removal of scratches and artifacts produced on CD or DVD recordings. This technique operates in the frequency domain and is based on the direct estimation of the short time duration of a clean signal. The research in [21] showed that a dual sensor spectral subtraction technique could perform better than a single sensor method due to the ability to recognize not only the spectral energy distribution of noise but also its time varying property. For this reason, the former is applied before the AR or MFCC feature extraction to reduce the deteriorating effect of the noise added to the lung sound signals.

A block diagram of the dual sensor spectral subtraction method is shown in Fig. 1. The primary sensor is responsible for detecting the corrupted lung sound signal and the reference sensor is responsible for recording the environmental noise. Let \(x(k)\) and \(n(k)\) be samples of a clean lung sound signal and environmental noise, respectively, and let \(y(k) = x(k) + n(k)\) be the samples of noisy lung sounds. The time domain signals were segmented and windowed with half-overlapping Hamming windows, and the resultant data were subjected to short time fast Fourier transform (STFFT), as shown in Fig. 1. The magnitude and phase of the signals were computed from the latter and the phase of the environmental noise was discarded from them. The reconstruction of the clean lung sound signal was performed by using only the corrupted lung sound phase. In order to enhance the lung sound, an estimate of the environmental noise magnitude was obtained; this value was then subtracted from the corrupted lung sound magnitude:

\[
\hat{X}(e^{j\omega}) = (|Y(e^{j\omega})| - |\hat{N}(e^{j\omega})|)e^{j\Phi(e^{j\omega})}
\]  (6)

where \(\hat{X}(e^{j\omega})\), \(|Y(e^{j\omega})|\), \(|\hat{N}(e^{j\omega})|\), and \(\Phi(e^{j\omega})\) are the estimated clean lung sound spectrum, corrupted lung sound magnitude, estimated noise magnitude, and the corrupted lung sound phase respectively. After spectral subtraction, the inverse short time fast Fourier transform (ISTFFT) was performed for each window to obtain the enhanced lung sound signal.

3. Results and discussions

3.1 Effects of noise on lung sound feature representation
The effects of noise were analyzed depending on the lung sound feature representation and the type and level of noise. Two types of lung sound feature representation, AR and MFCC coefficients, were calculated for the three categories of clean lung sounds (normal, wheeze, and crackle) as a noise-free reference feature set. In order to investigate the effects of noise on the lung sound feature representation under noisy conditions, we added babble noise and ambulance car noise at various SNR levels to each lung sound signal in the test database. Subsequently, the AR and MFCC coefficients of the corresponding noisy lung sound were computed. After the feature parameters were calculated for each clean and noisy lung sound, the scatter plots of the first two dimensions of these feature parameters were obtained and superimposed for the different kinds of lung sounds, feature representations, and for the different types and levels of noise. Fig. 2 shows the effect of the two real noises on two feature representations for a noisy wheeze signal, as a typical illustration, to observe the effects of additive noise. The two top subplots of Fig. 2(a) and (b) illustrates the effect of babble noise on the AR feature and the MFCC feature, respectively. Fig. 2(c) and (d) shows the effect of ambulance car noise on two feature representations. As shown in Fig. 2, additive noise causes a non-linear distortion in the feature representations in feature space. The value of distortion progressively increases as the SNR level of noise decreases. By comparing the top two subplots with the bottom two subplots in Fig. 2, we can observe

![Fig. 1 – Block diagram of dual sensor spectral subtraction algorithm.](image-url)
that babble noise leads to a greater distortion as compared to ambulance car noise, in the case of both the AR and MFCC feature representations. The distortion caused by ambulance car noise is relatively small. This result indicates that different types of environmental noise can generate distinct effects on feature representations. The technique used for reducing distortion in lung sound feature representations requires to be improved so as to be applicable in all possible noise conditions, not just for specific noise. This is our objective for studying robust lung sound recognition techniques under different environments. In the same way, similar investigations were performed for the other two types of lung sound-crackle and normal. Figs. 3 and 4 show the influences of additive noises for the other two types of lung sound-crackle and normal, respectively. The effects of the non-linear distortion caused by two types of real noise in the feature representations in feature space can also be observed from these two figures.

Fig. 2 – Scatter plots of the first two dimensions of AR and MFCC coefficients for noisy wheeze signal. The two top subplots (a) and (b) illustrate the effect of babble noise on the AR and MFCC coefficients at different SNR levels, respectively. The two bottom subplots (c) and (d) show the effect of ambulance car noise on the AR and MFCC coefficients at different SNR levels, respectively.
3.2. Effect of noises over recognition performance

According to the abovementioned results, additive noise produces distortion in feature representation space. Due to this distortion, the feature parameters representing the clean lung sound do not appropriately represent the noisy lung sound. This will give rise to a mismatch between training (clean) and recognition (noisy) environments. We can predict that the mismatch will worsen the classification rate during lung sound recognition because the corrupted lung sound is recognized by using the clean feature representation. In order to understand the impact of noise on classification rate, the practical recognition experiments were performed using the DTW technique under different noise conditions. Firstly, we evaluated our lung sound recognition system by using noiseless training sets and noise-free test sets and achieved 91.95% correct identifications with two types of feature parameters. Next, recognition experiments were performed under two real noise disturbances at various SNR levels using clean feature models. The recognition performances for the two types of features are presented in Fig. 5. The accuracies are plotted against the SNR levels ranging from 0 to 40 dB. It can be observed that additive babble noise significantly deteriorates the identification rate in the lung sound recognition process for both

![Fig. 3 - Scatter plots of the first two dimensions of AR and MFCC coefficients for noisy crackling signal. The two top subplots (a) and (b) illustrate the effect of babble noise on the AR and MFCC coefficients at different SNR levels, respectively. The two bottom subplots (c) and (d) show the effect of ambulance car noise on the AR and MFCC coefficients at different SNR levels, respectively.](image)
the AR coefficients and MFCC coefficients at low SNR levels. For example, the recognition rate reduces by about 60% at SNR of 0 dB for both types of feature representations. The average recognition rate is 62.37% for the AR feature and 50.81% for the MFCC feature for babble noise disturbance. This coincides with previous observations: babble noise leads to a more severe inconsistency between training and recognition environments than ambulance car noise. The average recognition rate is 79.19% for the AR feature and 65.77% for the MFCC feature for interference from ambulance car noise.

Additionally, in Fig. 5, the performance of the AR coefficients is observed to be better than that of the MFCC coefficients in our lung sound identification system at SNR levels above 5 dB. The mean recognition rate when the AR coefficients are used is superior to that when the MFCC coefficients

Fig. 4 – Scatter plots of the first two dimensions of AR and MFCC coefficients for noisy normal respiratory sound. The two top subplots (a) and (b) illustrate the effect of babble noise on the AR and MFCC coefficients at different SNR levels, respectively. The two bottom subplots (c) and (d) show the effect of ambulance car noise on the AR and MFCC coefficients at different SNR levels, respectively.
are employed by 11.55% for lung sound signal contaminated by babble noise. When the lung sound signals in the test set were corrupted by ambulance car noise, the mean recognition rate obtained by using the AR coefficients is better than that obtained by employing the MFCC coefficients by 13.42%. From the experimental results, it can be shown that the AR feature set is more robust for lung sound recognition in the presence of noisy interference. The MFCC feature set is more vulnerable to noise disturbance. However, the mean recognition rates of AR coefficients merely lie in between 62.37% and 79.19% under two types of real noise. These accuracy rates do not appear to be sufficient and need to be improved by using some signal enhancement algorithms if the lung sound recognition technology is to be utilized in real environments.

3.3. Compensation of noise by dual sensor spectral subtraction algorithm

In this study, the dual sensor spectral subtraction algorithm was applied to enhance the contaminated lung sound and reduce the mismatch between the training and the recognition environments. A number of spectrograms are presented in order to graphically display the effectiveness of the proposed algorithm in enhancing noisy lung sound signals. An
example of typical results obtained by using the algorithm is shown in Fig. 6. The spectrogram of the enhanced wheeze lung sound is presented in Fig. 6(d). From this figure, the algorithm compensated the noisy wheeze lung sound (Fig. 6(c)) affected by babble noise at an SNR of 20 dB (Fig. 6(b)) and it reconstructed the majority of the clean wheeze lung sounds shown in Fig. 6(a). Before the AR or MFCC feature extraction, all the noisy lung sound signals were preprocessed by using the dual sensor spectral subtraction method to suppress the effect of the noise added to the lung sound signals. Fig. 7 shows the scatter plots of the feature parameters of enhanced lung sound signals. It can be observed that after compensation, the feature parameters are shifted nearer to the clean feature parameters at SNR levels above 10 dB. As compared to the results in Fig. 2, the improvement in the distortion is between 29.6% and 51.4% for the AR coefficients and between 27.7% and 41.6% for the MFCC coefficients. The results show that the proposed algorithm is able to compensate the mismatch between the training and the recognition environments. However, the distortion is still significant at low SNR levels (5 dB and 0 dB). This is due to the fundamental limitations of the spectral subtraction algorithm [22]. In a future research, we intend combining the other methods of noise reduction and lung sound enhancement to increase the com-

Fig. 7 – Scatter plots of the first two dimensions of the AR and MFCC coefficients for an enhanced wheeze signal. The two top subplots (a) and (b) illustrate the compensating effect of babble noise on the AR and MFCC coefficients at different SNR levels, respectively. The two bottom subplots (c) and (d) show the compensating effect of ambulance car noise on the AR and MFCC coefficients at different SNR levels, respectively.
Fig. 8 – Recognition results after the noisy lung sound signal enhancement by dual sensor spectral subtraction. Noisy lung sound signal contaminated with (a) babble noise and (b) ambulance car noises at different SNR levels. The results correspond to AR and MFCC coefficients.

Table 1 – Comparison of average recognition rates after lung sound signal enhancement for two feature types with babble noise and ambulance car noise.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Average recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Noiseless</td>
</tr>
<tr>
<td>AR coefficients</td>
<td>91.95%</td>
</tr>
<tr>
<td>MFCC coefficients</td>
<td>91.95%</td>
</tr>
</tbody>
</table>

The recognition experiments were also evaluated for enhanced lung sound signals. The classification results are shown in Fig. 8. The recognition performance of the AR feature representation is still observed to be superior to that of the MFCC feature representation. Meanwhile, the average recognition rates for the enhanced are listed in Table 1. The values given in parentheses are the improvements obtained relative to the average recognition results without the lung sound signal enhancement. From Table 1, it is observed that the dual sensor spectral subtraction method improves the recognition performance of both feature representations when two types of real noise are present. For example, it improves 24.04% accuracy from 62.37% to 86.41% for the AR feature representation when the lung sound signals are corrupted by babble noise. Similar results can be observed in Table 1. The results presented in the table show that the application of the dual sensor spectral subtraction algorithm is a feasible method for improving the recognition performance in noisy environments.

4. Conclusions

In this study, the effect of noise on lung sound recognition was investigated under two real noise environments. It can be noted that additive noise produces a distortion in the feature representation and progressively degrades the recognition accuracies for low SNR levels. We proposed the dual sensor spectral subtraction method for compensating the effect of noise on the lung sound feature representation and for promoting the degraded classification results. The results are encouraging as they show that the proposed dual sensor spectral subtraction algorithm is able to provide adequate performance with regard to noise compensation and lung sound signal enhancement. However, it is notable that our study has not taken account of all the types of respiratory sounds. More experiments should be included to discriminate all the types of respiratory sounds.

Conflict of interest

The authors declare that no financial and personal relationships with any other people or organizations have inappropriately influenced the work here submitted.

Acknowledgement

The authors gratefully acknowledge the funding provided to this project by the National Science Council (NSC 96-2221-E-214-067), Taiwan, ROC.
REFERENCES


