Neurocomputing 123 (2014) 362–371

Contents lists available at ScienceDirect

Neurocomputing

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

New approaches for spectro-temporal feature extraction with applications to respiratory sound classification

F. Jin [a,*], F. Sattar [b], D.Y.T. Goh [c]

[a] Department of Electrical & Computer Engineering, Ryerson University, Toronto, Canada M5B 2K3
[b] Department of Electrical & Computer Engineering, University of Waterloo, Waterloo, Canada ON N2L 3G1
[c] Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 15 February 2013
Received in revised form
11 May 2013
Accepted 15 July 2013
Communicated by R. Capobianco Guido
Available online 20 August 2013

Keywords:
Respiratory sounds
Feature extraction
Instantaneous kurtosis
Sample entropy
Separability index
Support vector machine

ABSTRACT

Auscultation based diagnosis of pulmonary disorders relies on the presence of adventitious sounds. In this paper, we propose a new set of features based on temporal characteristics of filtered narrowband signal to classify respiratory sounds (RSs) into normal and continuous adventitious types. RS signals are first decomposed in the time–frequency domain and features are extracted over selected frequency bins containing distinct signal characteristics based on auto-regressive averaging, the recursively measured instantaneous kurtosis, and the sample entropy histograms distortion. The presented features are compared with existing features using a modified clustering index with different distance metrics. Mean classification accuracies of 97.7% and 98.8% for inspiratory and expiratory segments respectively have been achieved using Support Vector Machine on real recordings.

1. Introduction

Pulmonary auscultation has been the key technique to detect and evaluate respiratory dysfunctions for many years [1]. In order to remove the subjectivity inherent to auscultation, attempts have been made recently in the development of automatic respiratory sound (RS) recording and analysis methods. According to the Computerized Respiratory Sound Analysis (CORSA) guidelines [2], RS can be divided into normal breath sounds and adventitious sounds [3]. Normal breath sounds are characterized by broadband noise, while adventitious sounds are additional RS superimposed on the normal breath sounds. An important category of the latter refers to continuous adventitious sounds (CASs) such as wheeze and stridor (i.e. low-frequency wheeze) which have pseudo-periodic character. They have dominant frequency usually greater than 100 Hz and duration longer than 100 ms.

Since CASs can be viewed as narrow spectral peaks in the power spectrum, nonparametric frequency domain approaches have been widely used for feature extraction. These include the extraction of percentile frequencies by estimating power spectral densities [4], cepstral analysis for the extraction of MFCC (Mel-Frequency Cepstral Coefficients) [5,6], as well as signal coherence measuring the spectral stability of the signal in terms of amplitude and phase [7]. Among all, spectral analysis is most widely adopted for CAS detection. Relatively high detection rate has been reported in [8] but it is less effective when the magnitude difference between the spectrum of breath sound with and without CASs is small (which is common for mild wheeze). Furthermore, time-domain nonparametric feature extraction methods such as those based on Katz fractal dimension and variance fractal dimension to measure signal complexity are presented in [9]. On the other hand, parametric approaches such as those described in [10,11] adopt AR (autoregressive) modeling where AR parameters are estimated using the Levinson–Durbin recursion algorithm. As a crucial step in classification, feature extraction reduces the dimensionality of the pattern vector and provides a set of measurements with more discriminatory information and less redundancy to the classifier [12]. Some of the feature extraction techniques based on Fourier transform, linear predictive coding and MFCC have been evaluated and compared in [13]. A method for RS classification based on signal coherence is presented in [7]. Moreover, a stochastic approach using hidden Markov models and bigram models is presented in [21] for the classification of normal and abnormal respiratory sounds giving 84.2% classification accuracy. In [22], five feature sets are extracted using time–frequency and time-scale analysis and fed into support vector machine (SVM) classifier for detection of crackles. As a first feature set, the outputs of TF analysis are integrated over frequency, and so the behavior of signal upon time

* Corresponding author. Tel.: +1 4088982242.
E-mail addresses: jinfeng.ac@gmail.com, jinfeng2@hotmail.com (F. Jin), fsattar@uwaterloo.ca (F. Sattar), paegohyt@nus.edu.sg (D.Y.T. Goh).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.07.033

is obtained. As a second feature set, the outputs of TF analysis are integrated over time, and the behavior of signal upon frequency is obtained. As the third and fourth feature sets, the outputs of TS analysis are integrated over scale and time, and the behaviors of signal upon time and scale, respectively, are obtained. The original signals are used as the fifth feature set. Although the highest 97.20% overall accuracy is obtained with the proposed method, the processing is complex due to the large size feature sets. As mentioned in [13], artificial neural networks (ANN) [14,15], k-nearest neighbor (k-NN) [16], vector quantization (VQ) [17,18] as well as Gaussian mixture models (GMM) [19,20] have also been used for RS classification.

In particular, the accuracies of the classification using the energy based features of AR parameters, MFCC, signal coherence, and percentile frequencies, depend largely on the amplitude of the CAS. These tend to produce less discriminative patterns that make the choice of the classifier more crucial for obtaining reliable classification. On the contrary, features such as those measuring signal complexity are less dependent on the signal energy and therefore provide more accurate results. With the extracted features we could separate and classify adventitious sounds from normal lung sounds/tracheal breath sounds. Thereby, we are able to identify/classify different kinds of respiratory disease, for example pulmonary (lung) disease. This also allows us the scientific and clinical validation for several pathologies such as asthma, bronchitis [1].

The purpose of this paper is to develop new reliable feature extraction methods to provide more accurate classification of even low intensity RS into normal and CAS types to improve detecting of respiratory diseases. New discriminating feature sets based on the ratio of short-term and long-term AR averaging, and the averaged block-wise instantaneous kurtosis are extracted from the computed narrowband signals of the selected frequency bin to identify the presence of the CAS signals. Additionally, new feature set based on sample entropy (*SampEn*) and its histogram distortion between different frequency bins is proposed. The incorporation of histogram as a statistical tool removes the dependency of classification accuracy on the intensity of the adventitious sound. The proposed feature sets are then evaluated and compared with other conventional RS classification features in terms of a modified clustering index defined as separability index (SI). By investigating SI values and the classification results using SVM classifier, the proposed feature sets are shown to give promising performances both in terms of their large cluster separability and high classification accuracy. Thus the contributions of this paper lay in the following highlighting issues: (1) three novel feature extraction methods are proposed to classify RS sounds based on instantaneous kurtosis, discriminating function and sample entropy; (2) the new non-energy based features are sensitive to low amplitude abnormal RS being masked by loud normal RS providing high classification accuracies for both lung sounds and tracheal breath sounds of real recordings.

The paper is organized as follows. Section 2 provides a detail description of experimental RS dataset as well as the experimental setup for data acquisition. Section 3 introduces the proposed extraction methods for the three new feature sets with illustrative examples, together with the corresponding feature selection and feature evaluation schemes. Section 4 elaborates and explains the classification performance for presented feature sets using various SVM classifiers. Finally, Section 5 draws conclusion and discusses future work.

2. Data

Real tracheal breath sounds (TBSs) recordings were carried out in an audio laboratory with the subjects in sitting position. Single electret condenser microphone (ECM-77B, Sony Inc., Japan) was inserted into a hemispherical rubber chamber of 2 cm in diameter, and placed over suprasternal notch. The recording environment and equipments were chosen according to the standard given by [2] to suppress the environmental noises. The respiratory sound recordings were saved in mono-channel at sampling frequency $F_s = 11\,025$ Hz. Test subjects were asked to hold the breath for 15s then breathe normally with no targeted flow, 600 s recording was saved for each subject. In this study, the recorded dataset consists of TBS from 7 healthy and 14 pathological subjects with different degrees of airway obstruction (8 males/13 females, 15 ± 9 years old). However, the characteristics due to sex, age, weight were not taken into consideration. The recorded lung sounds (LSs) were extracted from databases [24–26], which consist of recordings captured over upper/lower lungs/chest from 5 healthy and 19 pathological subjects (14 males/10 females, 11 ± 17 years old). Types of CAS signals recorded include wheeze, stridor, rhonchi, and mixture of these.

The recorded RS signals were first segmented into individual inspiratory/expiratory segments according to the simultaneous flow meter reading, and manual classification results for these inspiratory/expiratory segments were obtained from an experienced doctor in Singapore National University Hospital by listening to each individual RS segment. RS segments belonging to normal breath sounds were labeled as group 1 and those belonging to CAS were labeled as group 2. The segmented and pre-classified inspiratory/expiratory segments were then used as the inputs to extract the proposed feature sets followed by RS classification into normal breath sounds and CAS.

A total of 72 expiratory normal TBS segments, 72 inspiratory normal TBS segments, 96 expiratory CAS segments, and 99 inspiratory CAS segments have been extracted from the TBS experiment dataset. At the same time, 30 expiratory normal LS segments, 59 inspiratory normal LS segments, 83 expiratory CAS (LS) segments, and 67 inspiratory CAS (LS) segments have been extracted from the LS experiment dataset. Due to the various size of each inspiratory/expiratory phase, the approximate durations of the segments vary between $[0.5, 3]$ s.

3. Feature extraction

The overall block diagram of the proposed feature extraction scheme is presented in Fig. 1. The RS signals are segmented into respective inspiration/expiration segments before decomposed into narrowband signals using time–frequency (TF) analysis as described in Section 3.1. Three sets of new features denoted by DP, KP, and SEP as obtained based on discriminating function, instantaneous kurtosis, and *SampEn* respectively, are extracted from the decomposed narrowband signals. Detailed description of each feature set is elaborated in Sections 3.2.2–3.4. A separability index *SI* based on various distance metrics is further proposed to evaluate the extracted feature sets followed by RS classification.

3.1. Time–frequency analysis

Fast Gabor TF distribution is applied here to transform the input signal into narrowband signals which can be processed separately (see Appendix A for details). Since consecutive spectral peaks with frequency proximity less than 50 Hz are considered to be part of the same CAS signal [8], a minimum frequency resolution of 50 Hz is required for the TF analysis. The total number of frequency bins is therefore chosen to be 256 to ensure this minimum frequency resolution where $F_s/256 = 11\,025/256 \approx 43$ Hz. The parameters used for Gabor spectrogram, i.e. the variance of the Gaussian analysis window (i.e. α parameter for