Automatic Wheezing Detection Based on Signal Processing of Spectrogram and Back-Propagation Neural Network

Bor-Shing Lin, PhD^{1*}; Huey-Dong Wu, MD² and Sao-Jie Chen, PhD³

¹Department of Computer Science and Information Engineering, National Taipei University, New Taipei City, Taiwan

²Department of Integrated Diagnostics and Therapeutics, National Taiwan University Hospital, Taipei, Taiwan

³Department and Graduate Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan

Submitted April 2015. Accepted for publication August 2015.

ABSTRACT

Wheezing is a common clinical symptom in patients with obstructive pulmonary diseases such as asthma. Automatic wheezing detection offers an objective and accurate means for identifying wheezing lung sounds, helping physicians in the diagnosis, long-term auscultation, and analysis of a patient with obstructive pulmonary disease. This paper describes the design of a fast and high-performance wheeze recognition system. A wheezing detection algorithm based on the order truncate average method and a back-propagation neural network (BPNN) is proposed. Some features are extracted from processed spectra to train a BPNN, and subsequently, test samples are analyzed by the trained BPNN to determine whether they are wheezing sounds. The respiratory sounds of 58 volunteers (32 asthmatic and 26 healthy adults) were recorded for training and testing. Experimental results of a qualitative analysis of wheeze recognition showed a high sensitivity of 0.946 and a high specificity of 1.0.

Keywords: Asthma, wheezing detection, bilateral filtering, order truncate average, back-propagation neural network

1. INTRODUCTION

In 2012, the number of noninstitutionalized adults and children in the United States having asthma are 18.7 million and 6.8 million, respectively [1]. In asthmatic or chronic obstructive pulmonary disease (COPD) patients, wheezes have been reported to be adventitious respiratory sounds generated during forced exhalation maneuvers. Wheezes are musical, adventitious, and continuous lung sounds. The waveform of a wheezing sound contains one or more sinusoidal components, explaining its musicality.

*Corresponding author: Bor-Shing Lin, Department of Computer Science and Information Engineering, National Taipei University, 151, University Rd., San Shia District, New Taipei City, 23741 Taiwan. Phone: +886-2-86741111 ext. 67123. Fax: +886-2-26744448. E-mail: bslin@mail.ntpu.edu.tw. Other authors: hdwu@ntuh.gov.tw; csj@ntu.edu.tw.

Thus, distinct peaks can be observed in the frequency domain [2]. According to the updated definitions in the most recent Computerized Respiratory Sound Analysis (CORSAs) standards, the dominant frequency of a wheeze is typically greater than 100 Hz, with the duration greater than 100 ms [3]. The transmission of a wheezing sound through the airway is better detected than the transmission through the lung to the surface of the chest wall. Thus, higher frequency sounds are detected more clearly over the trachea compared to the chest [4]. The high frequency components of breath sounds are absorbed mainly by the lung tissue [5]. The frequency of a wheeze lies in the range of 100–2500 Hz.

In previous research on the respiratory sounds of asthmatic patients, different algorithms have been developed to detect and analyze wheezes. The most straightforward methods for automatic detection involve searching for peaks in the frequency domain [6–10]. In spectral analysis, wheezes are seen as narrow peaks in the power spectrum, generally below 2000 Hz. Diagnostic failures often result from the shifting of the dominant frequency of a wheeze or the noise power being greater than that of wheezes. Algorithms proposed in the aforementioned studies are simple and fast; however, they are not very reliable and sensitive. To increase the sensitivity of wheeze detection based on successive spectra, combinations of algorithms and classification models have been proposed [11–14]. These approaches involve feature extraction and model comparison, both of which improve wheezing recognition. However, because the coefficients of classification models must be adjusted empirically, these approaches are inconvenient and increase the complexity of the algorithms. Thus, while these approaches enable precise wheezing detection, they are slow. Recent attempts to achieve higher sensitivity and efficient detection performance include the consideration of a set of criteria in the time-frequency domain [15–23]. These criteria pertain to the duration, pitch range, and magnitude of wheezes in the time-frequency representation of the wheezes obtained through spectrogram analysis. The objective of these studies was to automatically locate and identify wheezing episodes from sound recordings on the basis of well-defined criteria. These threshold criteria were used several times to empirically generate a normalized spectrum for detecting the maximum number of spectrum peaks unrelated to background noise. Thus, it is difficult to reproduce a wheezing detection system in different measuring environments and to obtain the maximal number of harmonics of wheezing episodes. In a recent study [24], a spirometer was employed to assist detection of wheezing. This mechanism provides the most accurate wheezing detection. However, it is inconvenient for long-term monitoring.

The objective of the current study is to investigate a method that involves image-processing techniques based on the normalized spectrogram recorded from lung sounds for identifying similar lung sounds. The proposed method enables visualizing wheezing characteristics, facilitating the search for horizontal or nearly horizontal edges of the spectrogram. The order truncate average (OTA) method is employed to overcome the drawbacks observed in the study of Homs-Corbera *et al.* [18]. The proposed method can be used for all sound levels, and enables identifying the most wheezing episodes in a short time. The back-propagation neural network (BPNN) is used to learn the features of wheezes in automatic wheezing recognition. After the BPNN is trained, it can precisely classify wheezing and non-wheezing without airflow data.

2. METHODS

2.1. Overview of the System

Various types of equipment and techniques exist for obtaining respiratory sounds [25]. The automatic wheezing detection system proposed in the current study was developed in accordance with CORSA standards and on the basis of previous studies [26–32]. Figure 1 shows a diagram of the system, consisting of hardware and software. The hardware consists of a sensor, a pre-amplifier, a band-pass filter (BPF), and a final amplifier prior to analog-to-digital converter (ADC). The purpose of the BPF is to reduce the heart, muscle, and contact noises. The designed bandwidth of BPF is from 60 Hz to 4 kHz for both lung and trachea sound analysis. A Butterworth low-pass filter (LPF) of fourth-order with 4 kHz cut-off and a fourth-order Bessel high-pass filter (HPF) are used to form the BPF. One single Quad Op-amp (MC34074, Motorola, Inc., USA) is used in both the HPF and LPF implementations, connected as 4 active filter stages (each of order 2) in cascade. The amplifiers increase the amplitude of the captured signal such that the full ADC range can be optimally used, and sometimes to adjust the impedance of the sensor. The sensor is realized using an electret condenser microphone (ECM, KEC-2738, Kingstate Electronics Corp., Taipei, Taiwan) with the bell of a stethoscope (3M Littmann Classic S. E.) fixed by hand between the skin surface and the microphone.

In clinical experiments, the hardware device collected and amplified the respiratory sounds. The digitization of sounds was performed by a soundcard (CS4297A) bounded in an IBM laptop (A22M, P-III 1 GHz). A 2 kHz bandwidth appears to be sufficient for studies of wheeze, but extending the bandwidth to 4 kHz is a perfect choice for the analysis of both adventitious sounds and upper-airway sounds. A standardized sampling rate applied in many industry standard sound facilities is 44.1 kHz, which is rather high for respiratory sound studies.

In this study, we focused on algorithms. All algorithms were implemented on a laptop by MATLAB 7.0 (The MathWorks, Inc., Natick, MA, USA). The software performed signal processing and involved using a neural network, as shown in Figure 1. The recorded respiratory sounds were first processed by an OTA signal-processing algorithm, and were then sent to the BPNN. Finally, the BPNN program performed training and wheezing recognition.

Figure 1. Block diagram of the automatic wheezing detection system.

Figure 2. Part of a spectrogram of a typical tracheal wheezing sound.

2.2. Signal Processing Using Order Truncate Average Method

Musical wheezing characteristics were determined using a fundamental frequency and its harmonics. Because wheezing is continuous, the resulting spectrogram contains quasi-horizontal lines indicating the strong presence of a determined frequency during a period of time. Figure 2 shows part of a spectrogram of a typical tracheal wheezing sound. It is clear that wheeze episodes (bounding boxes of dotted line in Figure 2) can be easily distinguished as edges distinct from the background sound components.

However, the edges that represent wheeze episodes are difficult for a computer to recognize because of blurring or spot formation resulting from noise. To enable automatic recognition by a computer, a wheeze detecting algorithm based on OTA filtering of a spectrogram was developed to retain edges defining wheeze episodes and to eliminate unwanted noise.

The objective of the proposed OTA algorithm is to preserve the maximum number of wheezing episodes. A schematic representation of the OTA algorithm is depicted in Figure 3. Each step and its respective results can be described as follows:

- 1) *Spectrogram*: Initially, each recorded respiratory sound file is loaded, and the length of the discrete Fourier transform (DFT), the type and length of the time window, and the overlapping percentage are defined. A DFT with 2048 points achieves an adequate frequency resolution of 2.15 Hz/pixel. The Hanning window acquires a rather smooth and acceptable spectral leakage, and its length is approximately 58 ms. The overlap of the window is approximately 50%. The time scale interval in the spectrogram is 29 ms, which provided an appropriate time resolution for wheezes.
- 2) *OTA Method*: The spectrogram is three-dimensional, indicating time, frequency, and power. The frequency-power plane of the spectrogram is acquired in sequence and then processed using the OTA method. Most frequency peaks are preserved after OTA processing.

Figure 3. Wheezing episode detection procedures based on the OTA method.

The OTA method is a type of spectrum analysis employed in signal detection systems to determine the presence and frequency of signals. By using the Fourier transform (FFT), the waveform sampled in the time domain was converted to the frequency domain. The magnitude spectrum is then obtained by computing the envelope of the frequency components in the FFT output. Large peaks in the magnitude spectrum indicate the presence of high noise values and signals.

To search for signals, noise background must be whitened, typically at unit height. The process of whitening the noise spectrum is called normalization, and is mathematically defined as

$$N_k = X_k / \mu_k, \quad (1)$$

where X_k is the magnitude in bin k and μ_k is the noise mean estimate in bin k . Bin k is the same as frequency k in the spectra. N_k is the normalized magnitude in bin k .

Typically, μ_k is a function of the spectral bin outputs in the neighborhood of bin k . If Ω_k denotes the set of bin numbers that can be used to estimate μ_k , then one possible definition of Ω_k is

$$\Omega_k = (k - M, k - M + 1, \dots, k + M - 1, k + M) \quad \text{for} \quad k \geq 0. \quad (2)$$

This definition assumes that the bin of interest is centered at Ω_k and that the number of bins in Ω_k is $K = 2M + 1$, where M is a positive integer.

Once an appropriate definition of Ω_k has been chosen, the manner of using the bin values in Ω_k to obtain the noise mean μ_k should be determined. The simplest estimator is the sample mean

$$\mu_k = \sum_{i \in \Omega_k} \frac{X_i}{K}. \quad (3)$$

If Ω_k were to contain only noise, this estimator would be the optimal linear estimator because the sample mean is the minimum variance unbiased linear estimator. However, when signals are present in Ω_k , μ_k can be biased severely upwards, thereby rendering the normalized outputs N_k in eqn. (1) considerably low.

The OTA normalizer was developed by Wolcin in 1978 [33]. The following steps describe the OTA method:

- (a) The K bin values in Ω_k are ordered to form a new sequence $(Y_1, Y_2, \dots, Y_k)$, where Y_1 is the smallest bin value and Y_k is the highest bin value.
- (b) The sample median Y_M is identified, and all bins having values greater than rY_M are excluded (the value for r is given later). Assume that L bins remain after the exclusion process.
- (c) The noise mean estimate μ_k is then obtained using the L remaining bins:

$$\mu_k = \sum_{i=1}^L \frac{Y_i}{L}. \quad (4)$$

- (d) In a similar manner, the shearing threshold for the OTA normalizer is defined as

$$\begin{aligned} T = rY_M &= [MM + (2 \times SGMM)]Y_M \\ &= \left\{ \frac{1}{2}(\pi/\ln 2)^{1/2} + 2[(4 - \pi)/4M \ln 2]^{1/2} \right\} Y_M, \end{aligned} \quad (5)$$

where MM is the theoretical mean-to-median ratio and $SGMM$ is the theoretical sigma-to-median ratio.

- 3) *Multiplication with Power Strength*: After OTA processing, we obtained many frequency peaks in each frequency-power plane. Some peaks were of interest, but other peaks were not required. To retain only the peaks of interest, each point on the frequency-power plane was multiplied by its original strength. Wheezes always have high strength, and therefore, wheezing peaks are likely to increase upon multiplication with their original strength. In this manner, more wheezing episodes can be preserved, similar to the preservation of wheeze harmonics.
- 4) *Shearing Using a Threshold*: To preserve the high-amplitude components, a limiter algorithm was developed. Because different sounds can be recorded using different techniques, different recorded signals may correspond to different recording levels. To ensure that recording levels have the same limitation, an adapted threshold is required. To achieve the optimal performance, the threshold should be appropriate for the properties of breath

sounds. Figures 4 and 5 show typical spectral variations during one breath. Figure 4 represents the power spectra of a wheezing subject, while Figure 5 represents those of a healthy subject. Normal tracheal sounds correspond to a broad peak and appear almost randomly. By contrast, wheezing produces a small number of well-defined peaks in the power spectrum. This difference

Figure 4. Power spectra over the trachea for a wheezing sound.

Figure 5. Power spectra over the trachea for a normal sound.

emphasizes the importance of amplitude criteria, such as the criterion used in the current study for distinguishing between normal and abnormal spectra. From the aforementioned properties of breath sounds, we can infer that the wheezing sounds have a larger standard deviation of power spectra than that of normal sounds. Thus, we defined an optimal threshold (Th_1) as follows:

$$Th_1 = \frac{C_1 \times m_{local}}{\sigma_{local}}, \quad (6)$$

where C_1 is a constant obtained from experiments, m_{local} is the mean of all points in a frequency-power plane, and σ_{local} is the standard deviation of all points in the frequency-power plane.

- 5) *Adjustments Involving N Iterations*: After the preceding step, the spectrogram still contains considerable noise and unwanted episodes. Therefore, a new threshold (Th_2) is used to further filter unwanted episodes. Considering the difference between the properties of wheezing and those of normal sounds, eqn. (6) is adjusted to obtain

$$Th_2 = \frac{C_2 \times m_{local}}{\sigma_{local}}, \quad (7)$$

where C_2 is a constant obtained from experiments. C_1 and C_2 are estimated experimentally as constant values with the goal of preserving the most complete shapes and maximum number of wheezing episodes.

After the power plane is sheared using this threshold, the noise power spectra of normal breath sounds are eliminated. In particular, peaks such as the wheezing peaks are preserved. This shearing method can be reused by employing the same threshold.

- 6) *Removal of Small Spots and Grouping*: After three shearing iterations, many quasi-horizontal lines and small spots are likely to exist. First, the small spots are removed and then the broken wheezing episodes are grouped. An algorithm is applied to connect separate episodes when the start and end points of the two episodes are close on the spectrogram. The grouping algorithm developed in this study considers the time, frequency, and amplitude proximity of the previously detected wheezing peaks. This algorithm scans the time-frequency plane and searches for ungrouped wheezing peaks. When a wheeze is found, the algorithm attempts to group it with other peaks as follows:
- (a) It searches for ungrouped peaks at a time distance of 29 ms. If there are peaks at this distance, the algorithm performs a frequency proximity check, only retaining wheezes within 50 Hz of the original wheeze.
 - (b) If there are no peaks fulfilling these conditions, the algorithm searches for ungrouped peaks at a time distance of 58 ms. A frequency proximity check is then performed, and only the peaks within 65 Hz are retained.
 - (c) The wheeze with the amplitude closest in value to the amplitude of the original wheezing peak is grouped with a new longer wheeze.

- (d) The entire process is repeated by considering the final grouped peak as the new starting peak, until no peak close to this one is found. Once the process is terminated, the entire wheeze is defined.
- 7) *Removal of Unacceptable Objects*: According to the definition of wheezing, all wheezes with durations shorter than 100 ms were eliminated.
 - 8) *Preservation of Wheezing Episodes*: By combining figures obtained from above steps, we obtain a final figure. This figure presents all detected wheezes with their corresponding strengths.

2.3. Back-Propagation Neural Network

An artificial neural network (ANN) is a powerful data-modeling tool that can capture and represent complex input/output relationships. Motivation for the development of neural network technology stemmed from a desire to develop an artificial system that could perform intelligent tasks comparable to the human brain. Currently, the back-propagation architecture is the most frequently applied, effective, and user-friendly model for complex, multilayered networks. Its most notable advantage is that it can be used for obtaining nonlinear solutions to ill-defined problems [34–36].

The layout of an ANN filter, shown in Figure 6, is similar to that of the human neural system. It comprises numerous interconnected processing elements (PEs). The ANN filter typically consists of an input layer of input nodes, one or more hidden layers of PEs, and an output layer that also consists of PEs. In this study, a feed-forward multilayer perceptron, which produces an output response to input signals in the network by propagating in the forward direction only, was used.

Figure 6. Back-propagation neural network.

Figure 7. Definition of parameters used for extracting wheezing features.

In this study, five input parameters are extracted from processed wheezing episodes. These parameters are presented in Figure 7 and defined as follows:

- 1) V_T : The time duration of a wheezing episode.
- 2) V_F : The frequency range of a wheezing episode.
- 3) V_{Ext} : The area/boundary of a wheezing episode $\left(\frac{A}{V_T \times V_F} \right)$, where A is the area of a wheezing episode.
- 4) V_{Std} : The normalized power spectra, equivalent to the local standard deviation of a wheezing episode/global standard deviation of the entire spectrogram.
- 5) V_{Slope} : The slope of a wheezing episode.

V_T , V_F , and V_{Ext} can provide the shape of an episode. According to the shapes, the wheezes present in the episode can be effectively detected. If V_{Std} is high, it indicates the possibility of a wheezing episode being present in the spectrogram. The final parameter, V_{Slope} , provides the slope of a wheezing episode. In most wheezing cases, the shape of wheezing episode appears as a quasi-horizontal line in a spectrogram. If the slope is close to 0 or 1, the episode may not be a wheezing episode.

After choosing appropriate parameters to extract processed wheezing episodes, we built a BPNN for training and testing the respiratory sound samples. To enhance the BPNN performance, we chose an appropriate training set size and an appropriate network structure. A common approach used for BPNN training in medical domains is to divide the collection of data samples into two groups based on a cutoff date; the training samples correspond to earlier dates, and the test samples correspond to later dates, simulating the prospective use of the BPNN.

All steps in the wheezing recognition process are presented in Figure 8. In the proposed BPNN, we required two groups of recorded respiratory sounds; one was for training and the other was for testing. After training the BPNN, we fixed the weights and biases of the BPNN. The test samples were then sent to the BPNN for classification.

Figure 8. Flow chart for the wheezing recognition process involving the BPNN.

In this study, the following conditions were considered for determining the training set size and BPNN structure:

- 1) *Input Nodes and Output Node*: By following all signal processing steps for spectrograms presented in Section 2.2, we can identify wheezing episodes. From these episodes, we choose the longest ten based on time duration. For each selected episode, we can extract three features such as (V_T, V_F, V_{Ext}) , four features such as $(V_T, V_F, V_{Ext}, V_{Std})$ and $(V_T, V_F, V_{Ext}, V_{Slope})$, or five features such as $(V_T, V_F, V_{Ext}, V_{Std}, V_{Slope})$. Thus, we can obtain 30, 40, or 50 input nodes according to the number of features chosen. Furthermore, we have only one output node that outputs “1” for a wheezing respiratory sound, and “0” for a normal respiratory sound.
- 2) *Hidden Layers and Neuron Numbers*: A typical BPNN has an input layer, an output layer, and at least one hidden layer. There is no theoretical limit on the number of hidden layers, but it is typically one or two. Another factor to consider is the neuron number in each hidden layer. When designing a neural network, one crucial parameter difficult to determine is the number of neurons in the hidden layers. The hidden layer is responsible for the internal representation of the data and information transformation in the input and output layers. Therefore, an optimal design for the number of neurons in the hidden layer is required.

In this study, we used (in, n_1, out) and (in, n_1, n_2, out) to construct a BPNN. Here, in is the number of input nodes, n_1 (between 15 and 200) is the number of neurons in Hidden Layer 1, n_2 (between 30 and 100) is the number of neurons in Hidden Layer 2, and out is the number of nodes in the output layer.

- 3) *Learning Factor*: To address the problem of training speed, attention was devoted to the learning factor during the development of the back-propagation software. As suggested by McClelland [37], the weights were updated once every complete cycle through the training samples rather than after each training sample. This process is known as batch training, or weight update by epoch, and it reduces the

number of computations required at each step. The back-propagation algorithm, similar to other numerical algorithms, can become unstable if the steps are too large. McClelland recommends $1/n$ as an appropriate size for the learning rate, where n is the total number of nodes in the network.

In the current study, we ignored the training time. Thus, we used a small and fixed learning factor to achieve stable convergence. In all experiments, the learning factor was 0.02, and the accumulated absolute error was less than 5×10^{-8} .

3. RESULTS

3.1. Participants

The respiratory sounds of 58 volunteers in the National Taiwan University Hospital (NTUH) were recorded to prepare for training and testing. The physician grouped 58 volunteers into training and test subjects, and marked all recorded sounds as wheezing or non-wheezing. The training subjects were of two types, as shown in Table 1. The training subjects consisted of 13 stable, asthmatic adults who had been without any acute exacerbation for 2 months, and 10 normal adults without any reported respiratory pathology. The test subjects were 35 volunteers, including 19 asthmatic adults and 16 normal adults, as shown in Table 2. After training the BPNN, we fixed the weights and biases of the BPNN. The test samples were then sent to the BPNN for classification.

3.2. Experiments

In this subsection, the effectiveness of the proposed algorithm is shown and discussed. First, wheeze recognition based on OTA filtering of the spectrogram was performed to classify segmented respiratory sounds. The high accuracy and robustness of wheeze recognition were demonstrated using wheezing and normal data sets. For example, we used Figures 9(a) to 12(d) to represent all steps in Section 2.2 to enhance the features

Table 1. Demographics of the training subjects

	Asthmatics (N=13)	Normal (N=10)
Age (years)	42.92±17.71	42.13±16.31
Height (m)	1.63±0.07	1.63±0.06
Weight (kg)	67.27±11.11	60.25±10.28
BMI (kg/m ²)	25.22±3.24	22.64±2.79

Table 2. Demographics of the test subjects

	Asthmatics (N=19)	Normal (N=16)
Age (years)	37.32±16.18	38.38±14.09
Height (m)	1.67±0.11	1.64±0.12
Weight (kg)	65.86±12.31	65.83±14.86
BMI (kg/m ²)	23.54±4.02	24.29±4.61

Figure 9. Results obtained using the OTA method for a wheezing sound: (a) original spectrogram, (b) strength before and after the OTA method was applied, (c) spectrogram obtained after shearing by using a threshold, and (d) all point values of a frequency-power plane before and after shearing using a threshold.

of wheezes in a wheezing respiratory sound. Many wheezes are evident at the dominant frequencies where harmonics are clear. By contrast, the spectrogram of a normal respiratory sound is almost bare or contains only a few objects. The spectrogram obtained is shown in Figure 9(a). The result of OTA processing is shown in Figure 9(b). Figure 9(c) shows a spectrogram in which each frequency-power plane is sheared using a threshold. Figure 9(d) shows all point values of a frequency-power plane before and after shearing using the threshold. Figures 10(a) to (d) and 11(a) to (d) show the spectrogram after one to four iterations as well as all point values of a frequency-power plane before and after shearing using the Th_2 threshold. In the proposed system, three iterations are sufficient to eliminate unwanted episodes and preserve peak components. The results of removing small spots and grouping are shown in Figures 12(a) and 12(b), respectively. Figure 12(c) shows the result of removal of unacceptable objects. Eventually, by combining figures obtained from Figures 9(a) to 12(c), we obtain Figure 12(d). This figure presents all detected wheezes with their corresponding strengths.

After signal processing of the spectrogram, the extracted features of preserved peak components are sent to BPNN for training and classification. Two-layer and three-layer BPNNs with one and two nonlinear hidden layers, respectively, were applied. In the two-layer BPNN, Hidden Layer 1 was formed using a log-sigmoid transfer function, and the output layer was formed using a linear transfer function. In the three-layer BPNN, Hidden Layer 1 was formed using a tan-sigmoid transfer function, Hidden Layer 2 was formed using a tan-sigmoid transfer function, and the output layer was formed using a linear transfer function. The value of the output at an output neuron represents the probability of wheeze occurrence; “1” represents “most likely,” whereas “0” represents “most unlikely.”

Eventually, a series of experimental tests were conducted by considering different input nodes, different numbers of neurons in the hidden layers, and different numbers of layers. Results based on the 30 inputs extracted from the three features (V_T , V_F , V_{Ext}) are shown in Table 3. Results based on the 40 inputs extracted from the four features (V_T , V_F , V_{Ext} , V_{Std}) are presented in Table 4. Results based on the 40 inputs extracted from the four features (V_T , V_F , V_{Ext} , V_{Slope}) are shown in Table 5. Results based on the 50 inputs extracted from the five features (V_T , V_F , V_{Ext} , V_{Std} , V_{Slope}) are presented in Table 6. In Tables 3 through 6, the estimated system performance (PER) is dependent on sensitivity (SE) and specificity (SP) as defined below:

$$Sensitivity(SE) = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)} \quad (8)$$

$$Specificity(SP) = \frac{True\ Negative(TN)}{True\ Negative(TN) + False\ Positive(FP)} \quad (9)$$

$$Performance\ (PER) = \sqrt{SE \times SP} \quad (10)$$

Figure 10. Results obtained using the OTA method for a wheezing sound: (a) spectrogram obtained after one iteration, (b) all point values of a frequency-power plane before and after shearing using a threshold, (c) spectrogram obtained after two iterations, and (d) all point values of a frequency-power plane before and after shearing using a threshold.

Figure 11. Results obtained using the OTA method for a wheezing sound: (a) spectrogram obtained after three iterations, (b) all point values of a frequency-power plane before and after shearing using a threshold, (c) spectrogram obtained after four iterations, and (d) all point values of a frequency-power plane before and after shearing using a threshold.

Figure 12. Results obtained using the OTA method for a wheezing sound: (a) spectrogram obtained after removing small spots, (b) spectrogram acquired after grouping, (c) spectrogram obtained after removing unacceptable objects, (d) spectrogram containing all detected wheezes with their corresponding strengths.

Table 3. Wheeze recognition results for the three features (V_T , V_F , V_{Ext})

BPNN Structure	Sensitivity (SE)	Specificity (SP)	Performance (PER)
$(in, n_1, out)=(30, 15, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, out)=(30, 30, 1)$	15/19 (0.79)	16/16 (1.00)	0.89
$(in, n_1, out)=(30, 60, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, out)=(30, 90, 1)$	16/19 (0.84)	16/19 (0.84)	0.84
$(in, n_1, out)=(30, 120, 1)$	15/19 (0.79)	15/16 (0.94)	0.86
$(in, n_1, n_2, out)=(30, 30, 30, 1)$	15/19 (0.79)	16/16 (1.00)	0.89
$(in, n_1, n_2, out)=(30, 60, 30, 1)$	16/19 (0.84)	14/16 (0.88)	0.86
$(in, n_1, n_2, out)=(30, 60, 60, 1)$	16/19 (0.84)	13/16 (0.81)	0.82
$(in, n_1, n_2, out)=(30, 90, 30, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, n_2, out)=(30, 90, 60, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, n_2, out)=(30, 120, 30, 1)$	17/19 (0.90)	16/16 (1.00)	0.95
$(in, n_1, n_2, out)=(30, 120, 60, 1)$	17/19 (0.90)	14/16 (0.88)	0.89

Table 4. Wheeze recognition results for the four features (V_T , V_F , V_{Ext} , V_{Std})

BPNN Structure	Sensitivity (SE)	Specificity (SP)	Performance (PER)
$(in, n_1, out)=(40, 20, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, out)=(40, 40, 1)$	14/19 (0.74)	16/16 (1.00)	0.86
$(in, n_1, out)=(40, 80, 1)$	15/19 (0.79)	16/16 (1.00)	0.89
$(in, n_1, out)=(40, 120, 1)$	15/19 (0.79)	16/16 (1.00)	0.89
$(in, n_1, out)=(40, 160, 1)$	16/19 (0.84)	13/16 (0.81)	0.82
$(in, n_1, n_2, out)=(40, 40, 40, 1)$	17/19 (0.90)	11/16 (0.69)	0.79
$(in, n_1, n_2, out)=(40, 80, 40, 1)$	14/19 (0.74)	16/16 (1.00)	0.86
$(in, n_1, n_2, out)=(40, 80, 80, 1)$	16/19 (0.84)	13/16 (0.81)	0.82
$(in, n_1, n_2, out)=(40, 120, 40, 1)$	17/19 (0.90)	15/16 (0.94)	0.92
$(in, n_1, n_2, out)=(40, 120, 80, 1)$	16/19 (0.84)	14/16 (0.88)	0.86
$(in, n_1, n_2, out)=(40, 160, 40, 1)$	17/19 (0.90)	16/16 (1.00)	0.95
$(in, n_1, n_2, out)=(40, 160, 80, 1)$	15/19 (0.79)	16/16 (1.00)	0.89

4. DISCUSSION

The proposed wheeze detection system has high sensitivity and high specificity, but also shows erroneous detection. The factors related to the erroneous detection will be discussed below. The performance of the BPNN is affected by many factors, including the number of input nodes, the number of hidden layers, and the number of neurons. We analyzed the experimental results to obtain the optimal parameters of the BPNN.

Table 5. Wheeze recognition results for the four features (V_T , V_F , V_{Ext} , V_{Slope})

BPNN Structure	Sensitivity (SE)	Specificity (SP)	Performance (PER)
$(in, n_1, out)=(40, 20, 1)$	15/19 (0.79)	16/16 (1.00)	0.89
$(in, n_1, out)=(40, 40, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, out)=(40, 80, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, out)=(40, 120, 1)$	17/19 (0.90)	16/16 (1.00)	0.95
$(in, n_1, out)=(40, 160, 1)$	16/19 (0.84)	15/16 (0.94)	0.89
$(in, n_1, n_2, out)=(40, 40, 40, 1)$	15/19 (0.79)	16/16 (1.00)	0.89
$(in, n_1, n_2, out)=(40, 80, 40, 1)$	17/19 (0.90)	16/16 (1.00)	0.95
$(in, n_1, n_2, out)=(40, 80, 80, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, n_2, out)=(40, 120, 40, 1)$	17/19 (0.90)	13/16 (0.81)	0.85
$(in, n_1, n_2, out)=(40, 120, 80, 1)$	16/19 (0.84)	13/16 (0.81)	0.82
$(in, n_1, n_2, out)=(40, 160, 40, 1)$	15/19 (0.79)	16/16 (1.00)	0.89
$(in, n_1, n_2, out)=(40, 160, 80, 1)$	17/19 (0.90)	13/16 (0.81)	0.85

Table 6. Wheeze recognition results for the five features (V_T , V_F , V_{Ext} , V_{Std} , V_{Slope})

BPNN Structure	Sensitivity (SE)	Specificity (SP)	Performance (PER)
$(in, n_1, out)=(50, 25, 1)$	15/19 (0.79)	16/16 (1.00)	0.89
$(in, n_1, out)=(50, 50, 1)$	14/19 (0.74)	16/16 (1.00)	0.86
$(in, n_1, out)=(50, 100, 1)$	17/19 (0.90)	16/16 (1.00)	0.95
$(in, n_1, out)=(50, 150, 1)$	17/19 (0.90)	15/16 (0.94)	0.92
$(in, n_1, out)=(50, 200, 1)$	15/19 (0.79)	13/16 (0.81)	0.80
$(in, n_1, n_2, out)=(50, 50, 50, 1)$	17/19 (0.90)	15/16 (0.94)	0.92
$(in, n_1, n_2, out)=(50, 100, 50, 1)$	17/19 (0.90)	16/16 (1.00)	0.95
$(in, n_1, n_2, out)=(50, 100, 100, 1)$	17/19 (0.90)	16/16 (1.00)	0.95
$(in, n_1, n_2, out)=(50, 150, 50, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, n_2, out)=(50, 150, 100, 1)$	16/19 (0.84)	14/16 (0.88)	0.86
$(in, n_1, n_2, out)=(50, 200, 50, 1)$	16/19 (0.84)	16/16 (1.00)	0.92
$(in, n_1, n_2, out)=(50, 200, 100, 1)$	16/19 (0.84)	16/16 (1.00)	0.92

4.1. Input Nodes

In Section 2.3, we propose four types of input nodes. After the experimental tests were conducted, we found that the 50 input nodes extracted from the five features (V_T , V_F , V_{Ext} , V_{Std} , V_{Slope}) showed the highest average performance, and the 40 input nodes extracted from the four features (V_T , V_F , V_{Ext} , V_{Std}) showed the poorest average performance. The average performance comparison is presented in Figure 13. The aforementioned experimental results show that the shape (V_T , V_F , V_{Ext}) and slope (V_{Slope})

Figure 13. Average performance comparison of input nodes.

Figure 14. Comparison of the average performance with different numbers of hidden layers.

of a wheezing episode have strong effects on wheezing recognition. However, the normalized power spectra (V_{Std}) has a weaker effect. We infer that both the noise and wheezing episodes have high power, which explains the weaker effect of the normalized power spectra.

4.2. Hidden Layers and Number of Neurons

There are no rules for selecting the number of hidden layers and the number of neurons. However, the experimental results demonstrate that two hidden layers show higher average performance than the others. The average performance comparison is shown in Figure 14. We selected 50 inputs and a three-layer BPNN structure. The performance comparison for neuron selections is shown in Figure 15.

Figure 15. Performance comparison for different numbers of neurons.

From the series of experimental tests, we finally chose 50 input nodes (in, n_1, n_2, out) = (50, 100, 50, 1) as the BPNN structure. With this simple structure, the proposed system has high sensitivity and high specificity for wheezing detection. The method effectively adapts to different sound volumes from different recording machines and resists the interference of environment noise. Depending on the wheezing properties, the physician can add more features to improve the rate of wheezing recognition. Regarding erroneous recognition, when reviewing the incorrect recognitions, these wheezing sounds appear very weak even for the physician to recognize. Therefore, a weak or noisy wheezing sound is the limitation in our proposed system. Reasons for erroneous recognition are discussed below:

- 1) Erroneous wheezing episodes may be preserved using the OTA method. In the signal- processing algorithm, the OTA method was used to preserve the maximum number of wheezing episodes. However, high-power noise may be preserved in some thresholds. To avoid the preservation of erroneous wheezing episodes and high-power noise, we should improve the OTA method and use noise reduction techniques.
- 2) Appropriate wheezing features should be chosen for extracting wheezing episodes. The experimental results revealed that the shape (V_T, V_F, V_{Ext}) and slope (V_{Slope}) of a wheezing episode have strong effects on wheezing recognition. To enhance wheezing recognition performance, we should identify new wheezing features to be used as inputs to the BPNN, such as T_w/T_{Cycle} , where T_w is the duration of a wheezing episode and T_{Cycle} is the duration of a respiratory cycle.
- 3) A larger number of subjects are required to improve the validation of the proposed wheeze recognition system. In the future, we intend to include a larger number of subjects for training and testing. We can even exchange the training and test subjects to achieve cross-verification, making the proposed system more accurate in wheezing recognition.

In this study, the proposed method not only provides a visual and auditory tool for clinicians, but also helps them to develop advanced diagnosis tools for pulmonary diseases. In clinic, some weak wheezing sounds are hard to recognize, especially for young physicians. The senior physicians can utilize our system to teach junior physicians in visual and auditory forms. After clinic, junior physicians can review the patients' records using our proposed system. The source code of our system was developed in Matlab, and can be easily modified to develop advanced algorithms for the diagnosis of pulmonary diseases.

5. CONCLUSIONS

A novel algorithm based on the OTA method was developed to detect wheezes with high performance, and to overcome the drawbacks in previous studies. The algorithm provides not only an automatic diagnosis, but also processed data to physicians. The treated spectrogram is shown on a computer screen before automatic recognition. The results of the experiments indicate that this algorithm can be useful in clinical diagnostics, mainly when the analysis is to be repeated for a number of respiratory cycles of a patient. The proposed wheeze detecting algorithm showed high sensitivity (0.946) and specificity (1.0) in the qualitative analysis of wheezes without the use of airflow data. Improvements are required for increased accuracy in detecting the duration of wheeze episodes. New wheezing features should be identified for use in the algorithm based on the OTA method. Also, a larger number of subjects should be included for training and testing.

ACKNOWLEDGEMENTS

This research was partly supported by Ministry of Science and Technology in Taiwan (R. O. C.), under grants MOST 103-2218-E-305-001, MOST 103-2218-E-305-003, and MOST 104-2221-E-305-006.

CONFLICT OF INTEREST

The authors indicated no potential conflicts of interest.

REFERENCES

- [1] Centers for Disease Control and Prevention (CDC). Asthma. 2012. <http://www.cdc.gov/nchs/fastats/asthma.htm>. Accessed March 8, 2015.
- [2] Sovijärvi ARA, Malmberg LP, Charbonneau G, Vanderschoot J. Characteristics of breath sounds and adventitious respiratory sounds. *European Respiratory Review*. 2000, 77(10):591–596.
- [3] Sovijärvi ARA, Dalmaso F, Vanderschoot J, Malmberg LP, Righini G, Stoneman SAT. Definition of terms for applications of respiratory sounds. *European Respiratory Review*. 2000, 77(10):597–610.
- [4] Fenton TR, Pasterkamp H, Tal A, Chernick V. Automated spectral characterization of wheezing in asthmatic children. *IEEE Transactions on Biomedical Engineering*. 1985, 32(1):50–55.
- [5] Wodicka GR, Stevens KN, Golub HL, Cravalho EG, Shannon DC. A model of acoustic transmission in the respiratory system. *IEEE Transactions on Biomedical Engineering*. 1989, 36(9):925–934.
- [6] Shabtai-Musih Y, Grotberg JB, Gavriely N. Spectral content of forced expiratory wheezes during air, He, and SF₆ breathing in normal humans. *Journal of Applied Physiology*. 1992, 72(2):629–635.

- [7] Xu J, Chen Q, Zhang Y, Liu S. Spectrum analysis of lung sounds. *Proceedings of the 11th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 1989, 5:1676–1677.
- [8] Hadjileontiadis LJ, Panas SM. Nonlinear analysis of musical lung sounds using the bicoherence index. *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 1997, 3:1126–1129.
- [9] Jané R, Salvatella D, Fiz JA, Morera J. Spectral analysis of respiratory sounds to access bronchodilator effect in asthmatic patients. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 1998, 6(6):3203–3206.
- [10] Jané R, Cortés S, Fiz JA, Morera J. Analysis of wheezes in asthmatic patients during spontaneous respiration. *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2004, 2:3836–3839.
- [11] Forkheim KE, Scuse D, Pasterkamp H. A comparison of neural network models for wheeze detection. *Proceedings of IEEE Communication, Power, and Computing Conference*. 1995, 1:214–219.
- [12] Bahoura M, Pelletier C. New parameters for respiratory sound classification. *Proceedings of IEEE Electrical and Computer Engineering Conference*. 2003, 3:1457–1460.
- [13] Bahoura M, Pelletier C. Respiratory sounds classification using Gaussian mixture models. *Proceedings of IEEE Electrical and Computer Engineering Conference*. 2004, 3:1309–1312.
- [14] Oletic D, Arsenali B, Bilas V. Low-power wearable respiratory sound sensing. *Sensors*. 2014, 14(4):6535–6566.
- [15] Waris M, Helistö P, Haltsonen S, Saarinen A, Sovijärvi ARA. A new method for automatic wheeze detection. *Technology and Health Care*. 1998, 6(1):33–40.
- [16] Taplidou SA, Hadjileontiadis LJ, Penzel T, Gross V, Panas SM. WED: An efficient wheezing-episode detector based on breath sounds. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2003, 3:2531–2534.
- [17] Taplidou SA, Hadjileontiadis LJ, Kittsas IK, Panoulas KI. On applying continuous wavelet transform in wheeze analysis. *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2004, 2:3832–3835.
- [18] Homs-Corbera A, Fiz JA, Morera J, Jané R. Time-frequency detection and analysis of wheezes during forced exhalation. *IEEE Transactions on Biomedical Engineering*. 2004, 51(1):182–186.
- [19] Lin BS, Lin BS, Wu HD, Chong FC, Chen SJ. Wheeze recognition based on 2D bilateral filtering of spectrogram. *Biomedical Engineering Applications, Basis & Communications*. 2006, 18:128–137.
- [20] Lin BS. Using back-propagation neural network for automatic wheezing detection. PhD dissertation, National Taiwan University, Taiwan, 2006.
- [21] Taplidou SA, Hadjileontiadis LJ. Analysis of wheezes using wavelet higher order spectral features. *IEEE Transactions on Biomedical Engineering*. 2010, 57(7):1596–1610.
- [22] Jin F, Krishnan S, Sattar F. Adventitious sounds identification and extraction using temporal–spectral dominance-based features. *IEEE Transactions on Biomedical Engineering*. 2011, 58(11):1596–1610.
- [23] Uwaoma C, Mansingh G. Detection and Classification of Abnormal Respiratory Sounds on a Resource-constraint Mobile Device. *International Journal of Applied Information Systems*. 2014, 7(11):35–40.
- [24] Kwan AM, Fung AG, Jansen PA, Schivo M, Kenyon NJ, Delplanque JP, Davis CE. Personal lung function monitoring devices. *IEEE Sensors Journal*. 2015, 15(4):2238–2247.
- [25] Earis JE, Cheetham BMG. Current methods used for computerized respiratory sound analysis. *European Respiratory Review*. 2000, 77(10):586–590.
- [26] Cheetham BMG, Charbonneau G, Giordano A, Helistö P, Vanderschoot J. Digitization of data for respiratory sound recordings. *European Respiratory Review*. 2000, 77(10):621–624.
- [27] Piirlä P, Sovijärvi ARA, Earis JE, Rossi M, Dalmaso F, Stoneman SAT, Vanderschoot J. Reporting results of respiratory sound analysis. *European Respiratory Review*. 2000, 77(10):636–640.

- [28] Jones A, Jones D, Kwong K, SC S. Acoustic performance of three stethoscope chest pieces. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 1998, 6(6):3219–3222.
- [29] Scanlon MV. Acoustically monitor physiology during sleep and activity. *Proceedings of the 1st Joint BMES/EMBS Conference*. 1999, 2:787.
- [30] Moussavi Z. Vocal noise cancellation from respiratory sounds. *Proceedings of the 23th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2001, 3:2001–2003.
- [31] Jamieson G, Cheetham BMG, Moruzzi JL, Earis JE. Digital signal processing of lung sound. *IEE Colloquium on Digital Signal Processing in Instrumentation*. 1992, (9):7/1–7/4.
- [32] Sun XQ, Cheetham BMG, Evans KG, Earis JE. Estimation of analogue pre-filtering characteristics for CORSA standardization. *Technology and Health Care*. 1998, 6(4):275–283.
- [33] Struzinski WA, Lowe ED. A performance comparison of four noise background normalization schemes proposed for signal detection systems. *The Journal of the Acoustical Society of America*. 1984, 76(6):1738–1742.
- [34] Long X, Cleveland WL, Yao YL. A new preprocessing approach for cell recognition. *IEEE Transactions on Information Technology in Biomedicine*. 2005, 9(3):407–412.
- [35] Shen S, Sandham W, Granat M, Sterr A. MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *IEEE Transactions on Information Technology in Biomedicine*. 2005, 9(3):459–467.
- [36] Walczak S. Artificial neural network medical decision support tool: predicting transfusion requirements of ER patients. *IEEE Transactions on Information Technology in Biomedicine*. 2005, 9(3):468–474.
- [37] Heermann PD, Khazenie N. Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Transactions on Geoscience and Remote Sensing*. 1992, 30(1):81–88.