Lung Sound Classification based on Hilbert-Huang Transform Features and Multilayer Perceptron Network

Yun-Xia Liu¹,²,³, Yang Yang ⁴ and Yue-Hui Chen¹,²
¹School of Information Science and Engineering, University of Jinan, Jinan 250022, China
²Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan 250022, China
³ School of Control Science and Engineering, Shandong University, Jinan 250061, China
E-mail: ise_liuyx@ujn.edu.cn, yhchen@ujn.edu.cn Tel: +86-531-82767500
⁴ School of Information Science and Engineering, Shandong University, Jinan, China
E-mail: yyang@sdu.edu.cn Tel: +86-531-88364103

Abstract—Accurate classification of lung sounds plays an important role in noninvasive diagnosis of pulmonary diseases. A novel lung sound classification algorithm based on Hilbert-Huang transform (HHT) features and multilayer perceptron network is proposed in this paper. Three types of HHT domain features, namely the instantaneous envelope amplitude of intrinsic mode functions (IMF), envelop of instantaneous amplitude of the first four layers IMFs, and max value of the marginal spectrum are proposed for jointly characterization of the time-frequency properties of lung sounds. These proposed features are feed into a multi-layer perceptron neural network for training and testing of lung sound signal classification. Abundant experimental work is carried out to verify the effectiveness of the proposed algorithm.

I. INTRODUCTION

Classification of lung sound is the prerequisite and basis of correct diagnosis for a wide range of lung related diseases [1] such as pneumonia, pulmonary fibrosis, and etc. Clinically, the classification is usually performed by experienced doctors according to their empirical knowledge. Apparently, the traditional method suffers from great limitations of strong subjectivity, low reliability and is hard to generate reproducible results for the purpose of disease tracing.

With the fast improvement of diagnostic techniques and development of signal processing algorithms, it is an inevitable trend to make digital and scientific pulmonary auscultation [2]. The development of computerized lung sound analysis has attracted much attention from researchers in recent years, which has led to several implementations of machine learning algorithms [3], where feature extraction and classification are two key steps in machine learning based diagnosis of lung sound. Traditional feature extraction methods in lung sound classification can be approximately divided into three categories, namely the time domain methods, the frequency domain methods and the joint time-frequency domain methods, where autoregressive model based features, mel-frequency cepstral coefficient (MFCC), energy, entropy, spectral features, and wavelet coefficients [3] are typical representatives. However, as revealed by the Heisenberg-Gabor uncertainty principle, these spectrum analysis methods cannot simultaneously achieve high analysis resolution in both the time and frequency domain, thus have limited classification performance.

Hilbert-Huang transform (HHT) is an innovative signal processing method [4] which demonstrates excellent performance in analyzing non-stationary signals. Firstly, the empirical modal decomposition (EMD) is performed to transform the time domain signal into series of intrinsic mode functions (IMF), which can be regarded as different high-frequency components. Then, the Hilbert transform is applied to IMFs to enable reliable assessment of instantaneous frequency and amplitude at different timescale. There have been several HHT based features in the field of lung sound analysis, e.g. denoising of explosive lung sounds in [5] and enhancement of lung sounds in [6], verifying its effectiveness. For lung sound classification, the mean value of each IMF is used to classify the coarse and fine crackles [7], while the energy weight in various frequency bands is utilized as features for velcro rales identification [8] with a support vector machine classifier. However, these methods use either time or frequency domain features separately, which cannot fully exploit the advantages of HHT.

In this paper, we propose to jointly use time and frequency features in HHT domain for better lung sound classification. Three different types of features are discussed in this paper, namely the instantaneous envelope amplitude of IMF1-4, their HHT spectrum, and also the marginal spectrum. A multilayer perceptron network is then utilized as the classifier to identify different types of abnormal lung sound. Experimental results performed on the R.A.L.E. database [9-10] demonstrate the effectiveness of the proposed algorithm.

II. RELATED WORKS

In this section, we first introduce the various types of abnormal lung sounds and their different time-frequency domain properties. Based on a brief overview of the Hilbert Huang transform, we discuss the properties of several features of typical lung sound signals, which provide the basis for the
proposed HHT features that will be discussed in detail in section III.

A. Classification of lung sound

Accurate classification of the type of lung sound is helpful for the diagnosis of respiratory disease. Lung sound is a typical non-stationary signal with wide bandwidth varying from 20 to 2000Hz. Table 1 lists five typical types of abnormal lung sounds, their related possible lung diseases [11], and characteristics of the corresponding signals for comparison.

<table>
<thead>
<tr>
<th>Possible Lung Diseases</th>
<th>characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crackles</td>
<td>duration is less than 20 ms, typically narrow frequency content</td>
</tr>
<tr>
<td>Wheezes</td>
<td>periodic waveforms with a dominant frequency usually over 100 Hz and with a duration of 100 ms</td>
</tr>
<tr>
<td>Rhonchi</td>
<td>very loud wheezes, a prominent peak at about 1000 Hz</td>
</tr>
<tr>
<td>Squawks</td>
<td>always occur along with crackles, and often begin with a crackle</td>
</tr>
<tr>
<td>Stridors</td>
<td>with dominant frequency less than 200 Hz</td>
</tr>
</tbody>
</table>

It can be seen from Table 1 that, joint time and frequency properties should be specified to determine the presence of a certain type of abnormal lung sound. For example, the duration of crackles is about 20ms, while the wheezes is about 100 ms. Another example is that the prominent peak in stridors is about 1000 Hz while the Rhonchi is less than 200Hz. The crackles begin with an oscillation, and gradually expanded, then stabilized, the sudden change characteristic of visible moment. As the time and frequency domain features related to each other and are not easy to capture simultaneously, single feature is impossible for effective presentation of these complex signals. Multiple adaptive feature extraction schemes should be designed for better discrimination of lung sounds.

B. Hilbert Huang Transform

The Hilbert Huang Transform (HHT) [4] is an effective time-frequency analysis tool that is consisted of two steps.

Firstly, apply the empirical mode decomposition (EMD) to decompose the signal \( s(t) \) into a series of intrinsic mode functions (IMF) according to their timescale,

\[
s(t) = \sum_{i=1}^{D} c_i(t) + r_D(t),
\]

where \( c_i(t) \) is the \( i \)-th IMF signal (i.e. IMF\(_i\)), \( D \) is number of decomposition layers, \( r_D(t) \) is a trend term, representing a slow trend in the signal, or a constant. Fig.1 depicts the IMFs of a typical lung sound signal. We can see that the time domain resolution changes from fine to coarse with the increase of \( i \), where the first layer IMF\(_1\) corresponds to the IMF with the fastest oscillations.

For the second step, one applies the Hilbert transform to these IMF components to obtain the corresponding instantaneous frequency and instantaneous envelope. The Hilbert Huang spectrum is constructed as

\[
z(t) = c_i(t) + jH[ c_i(t) ] = A_i(t)e^{j\phi_i(t)},
\]

where \( A_i(t) \) and \( \phi_i(t) \) represents the instantaneous envelope and instantaneous phase, respectively. The input signal \( s(t) \) can then be expressed as

\[
H(\omega, t) = \sum_{i=1}^{D} A_i(t)e^{j\phi_i(t)},
\]

then the original signal \( s(t) \) can be obtained as:

\[
s(t) = \text{Re} \sum_{i=1}^{D} A_i(t)e^{j\phi_i(t)}.\]

Finally, we can further obtain the marginal spectrum by integration along the time axis on the time-frequency plane as

\[
H(\omega) = \int_{0}^{T} H(\omega, t) dt,
\]

where \( T \) denotes the total data length. Fig.2 shows the marginal spectrum of the five types of abnormal lung sound signal as well as the normal one, with decomposition levels setting to four. One can see the consistence of their characteristics as discussed in Table 1, which is quite stable in our abundant experiments.

![Fig. 1 IMFs of a typical lung sound signal with crackles.](image)

![Fig. 2 Marginal spectrum of six types of lung sounds.](image)

From the discussion above, we see that the HHT framework provides a wide range of time and frequency domain features for characterization of signals. How to extract efficient features that have strong discrimination power is important to success lung sound classification.
III. PROPOSED LUNG SOUND CLASSIFICATION ALGORITHM

In this section, based on in-depth analysis of the characteristics of lung sounds, we discuss the three types of stable features that could be used for lung sound classification. The overall architecture is illustrated in Fig.3.

![Flowchart of the proposed lung sound classification algorithm.](Image)

Firstly, we choose the instantaneous envelope amplitude of intrinsic mode functions for the time domain features. As we have observed in Fig.2 that EMD decomposition is equivalent to an adaptive high pass filter. The lower layer IMF components demonstrate characteristics of rapid and fast decay and the remaining IMF components as the background sound including normal breathing and some low-frequency pleural friction sound, which has lower discrimination power. Thus the first four IMFs are utilized for feature extraction. Instead of using the summation of the first three layers as in [8] utilized for further feature extraction, we propose to use the layer-wise features. The proposed IMF feature vector $I$ is consisting of max and mean values of IMF1-4 as

$$I = [\text{max}(IMF_1(t)), ..., \text{max}(IMF_4(t)), \text{mean}(IMF_1(t)), ..., \text{mean}(IMF_4(t))]. \quad (5)$$

The second type of feature utilized for lung sound classification is envelopes of instantaneous amplitude of the first four layers IMFs. We have experimentally observed that the peak values of corresponding IMFs are quite stable. For example, the peak values of crackles of the first four layers are 0.13, 0.16, 0.19 and 0.13, while those of bronchial signal are 0.0434, 0.0395, 0.1312 and 0.0850. (As a pre-processing procedure, all signals are normalized into the range of [0,1], which is assumed to be the default settings hereafter.) This is in consistency with the fact that high frequency component in bronchovesicular breathing is small and most of the energy is concentrated in the lower frequency bands. The mean and max values of these four layers contribute 8 dimensions in the proposed feature vector:

$$E = [\text{max}(A_1(t)), ..., \text{max}(A_4(t)), \text{mean}(A_1(t)), ..., \text{mean}(A_4(t))]. \quad (6)$$

The last important feature is the marginal spectrum, which is quite distinguishing of different types of lung signals as shown in Fig.2. From the statistical point of view, it represents the frequency of the amplitude (energy) accumulated in time. As discussed above that the summation of the first few IMFs reflects the high-frequency components, thus is informative in discriminating different types of lung sound signal. We propose to use the max value of the marginal spectrum of the summation of the first four IMFs

$$M = \max \int_0^T H(\omega,t)dt \quad (7)$$

as last dimension of the proposed feature vector.

In summary, three types of HHT domain features are extracted to form a 17 dimensional feature vector

$$F = [I, E, M] \quad (8)$$

for the purpose of lung signal classification.

To demonstrate the effectiveness of the proposed HHT features, a relative simple multilayer perceptron (MLP) neural network is adopted as the classifier. A three layers MLP with back prorogation algorithm is utilized for stochastically solutions of the supervised learning problem in our case. The influence of parameter settings on classification accuracy are discussed in detail in section IV.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We test the effectiveness of the proposed algorithm on the largest commercial lung sound database, namely R.A.L.E [9-10] on a PC with Intel Core I3-4150 CPU @3.50GHz, 8G RAM, Matlab R2017a. There are totally 51 pieces of lung signal sampled at 11025 Hz, including 11pieces with crackle, 3 pieces with wheeze, 4 pieces with stridor, 2 pieces with squawk, 2 pieces with rhonchi as well as 29 pieces normal ones, where each piece is about 5-10 seconds.

To extend the database we segment each signal into 100 overlapped sub-pieces with 50000 points in each sub-piece. Thus we get 51×100 = 5100 pieces of signal in total. We randomly choose one half as the training set and the other half as the testing set. Each figure in the following tables are mean value of 100 times evaluation to reduce the influence of noise.

A. Performance comparison of various features

To compare the classification performance of various features, we compare their classification rate (%) in Table II. For the IMF and envelope features $I$ and $E$, max and mean features are evaluated separately. The number of nodes in MLP in hidden layer $N$ is set to be 100, while total number of loop time is set to be 1000.

It can be seen from Table II that it is difficult to achieve high accuracy for any single feature, while the accuracy of the proposed combined features is much higher. In the rest of the paper, we choose the combination as the optimal features. An average accuracy of 94.82% is achieved, which is the best result on the R.A.L.E. to the best of our knowledge.

B. Performance comparison on different MLP setting
The second series of experiments are carried out to examine the influence of different parameter settings of the multilayer perceptron classifier for classification performance. With the proposed HHT feature, Table III compares the classification accuracy with varying number of nodes \( N \) in the hidden layer, while the total number of loop time is set to be 1000. It can be seen that as the increase of the hidden layer nodes, the classification accuracies is increased at first. However, when the number is more than 100, the accuracies tend to decrease, suggesting that \( N \) is a key parameter influencing the final result. In real applications, the value of \( N \) should be carefully determined depending on the scale of lung sound data collected.

Table IV compares the classification accuracy with varying number of loop time when \( N \) is fixed to 100. It can be seen that the increase of the loop time always lead to performance improvement. However, the improvement gain is trivial when loop time further increases above 10\(^5\). One have to carefully tradeoff between computation complexity and performance, in both training and testing stages.

### V. CONCLUSIONS

In this study, we propose a novel HHT domain feature for classification of crackle, wheeze, stridor, squawk and rhonchi from normal lung sound signals to assist diagnosis of related diseases. Aiming at joint time-frequency characterization of the lung sound signal, three types of features consisting of statistics of IMF, instantaneous envelope and Hilbert spectrum of high frequency components are utilized. An averaged classification accuracy of 95.84\% is achieved on the R.A.L.E. database with a multi-layer perceptron classifier, suggesting promising applications in pulmonary auscultation.

Current results are obtained by simulation on the R.A.L.E. database. In the future, we plan to collect more clinical data to improve the robustness of the algorithm.

### ACKNOWLEDGMENT

This work was supported by the National Nature Science Foundation of China (No. 61305015, No. 61203269), the Shandong Province Key Research and Development Program, China (Grant No. 2016GGX101022), and the Postdoctoral Science Foundation of China (No. 2015M580591).

### REFERENCES