Research on Recognition of CHD Heart Sound Using MFCC and LPCC

ZHU Lili*, PAN Jiahua², SHI Jihong¹, YANG Hongbo² and WANG Weilian¹
¹School of Information, Yunnan University, Kunming, Yunnan 650500, China
²Yunnan Fuwai Cardiovascular Disease Hospital, Kunming, Yunnan 650102, China

*E-mail:857407372@qq.com

Abstract. Congenital heart disease (CHD) is a disease that seriously harms the children and family. It needs to be diagnosed and treated in time. The initial diagnosis way of CHD is cardiac auscultation in which the rich experience and expertise are needed. A kind of recognition algorism was proposed to help the initial diagnosis and screening of CHD in this work. The heart sounds were analyzed and the feature extraction by using MFCC and LPCC methods in which the frame length was 2048. The BP neural network was selected as classifier. Results show that the specificity and sensitivity of recognition ratios for CHD are 93.02% and 88.89% by using MFCC, and 86.96% and 86.96% by using LPCC respectively. The features extracted by using MFCC method are better than one by using LPCC.

1. Introduction

Congenital heart disease (CHD) is a disease that seriously harms the health of children. There are 16.5 million new born children in China every year, in which 0.8 percent of them may suffer from CHD and the incidence of CHD is going on the rise [1]. CHD not only harms the health of children seriously, but also imposes a great burden on the patient’s family. [2] A patient will be saved, if CHD can be diagnosed and treated timely. Save a child may save a family.

By now the cardiac auscultation through a stethoscope is basic way for doctors to diagnose and screen CHD initially. It requires a both of rich experience and expertise for a doctor to make auscultation. To accumulate the clinic experience and expertise need take a long time. Analysis of heart sound using a digital signal processing method may be helpful to realization of machine assisted auscultation of CHD.

In recent years, MFCC (Mel-Frequency Cepstral Coefficients) has had a good application in the feature extraction of cardiac sounds [3], [4], [5]. BP (Back Propagation) neural network also has a good effect on the classification and recognition of heart sounds [6], [7]. In [8] seven types of cardiac signal were decomposed into five layers. The normalized sub-band energy of the best wavelet basis was extracted as the feature vector. 70 heart sound samples were used to test the trained seven BP neural networks. The correct recognition rate of both normal heart tone and soft first heart sound was 90%, the correct identification rate of second heart sound split was 80%, and for other conditions was 70%. However, the average recognition rate of this method was 77.14%, then the experimental test samples were too small to explain problem. In [9] the first heart sound (S1) and the second heart sound (S2) were segmented by a time threshold, and the Shannon entropy of S1 and S2 were...
calculated. The calculated amount of information was combined as a feature value into a matrix containing a plurality of elements. It acted as an input of the BP neural network to identify heart sound signals. Correct rates of mitral stenosis and mitral insufficiency were 86.67% and 93.33%. But, the correct identification rates of simple aortic insufficiency and aortic insufficiency were just 73.33% and 80.00% only. The features extracted from the Shannon entropy of S1 and S2 were not good enough to explain the information.

Mel-Frequency Cepstrum Coefficient (MFCC) was used to extract features in [10], and support vector machines (SVM) was used for classification identification. The length of the frame selected for the MFCC feature extraction was 20 to 30 ms, and the sampling rate was 44100 Hz. When the number of test samples was 10, 20, 30, their true positive rates were 95%, 98.7%, 96.6%, and the true negative rate were 99.4%, 99.3%, 99.8% respectively. However, the number of tests is too small to demonstrate the reliability of the SVM's recognizer. The 13-dimensional MFCC parameter was used as signal feature, and the hidden markov model (HMM) was used as recognition classifier in [11]. The length of the MFCC window selected in the experiment was 20 ms and the sampling rate was 8000 Hz. When the HMM had 4 states, the maximum correct classification rate was 95.08%. However, in the experiment, 10 types of identification were required, and the maximum number of test samples for each type of identification was less than 40 cases. As a classifier, HMM only depends on each state and its corresponding observation object. And the cardiac signal is a signal that has a strong correlation. In [12], 100 heart sound signals of 50 people were selected to compare the characteristic parameters of Linear Predictive Cepstral Coefficients (LPCC) and MFCC. Using gaussian mixture model (GMM) as classifier, the experiment shows that LPCC parameter was more effective than MFCC parameter. But, due to the data volume and the defects of GMM itself, it was not enough to explain the universality of the method.

In order to overcome the disadvantages listed in the above, a novel method was proposed for recognition of CHD heart sound, which was based on the classical speech signal processing recognition process. Firstly, the heart sound signals were pre-processed, and then the characteristic parameters of cardiac signal were extracted using MFCC and LPCC. It was liked the feature extraction algorithm of speech in [13] and [14] respectively. A BP neural network was selected to classify the normal and abnormal heart sound signals in this algorithm. Training and testing were carried out on a certain amount of data. The results showed that the algorithm had a good performance on recognition of CHD heart sound.

2. Methodology

2.1. Data Collection
The abnormal data used in this paper came from the clinical collection of Yunnan Fuwai Cardiovascular Hospital. Normal data collected from healthy children in primary schools in Mangshi, Dehong Prefecture, Yunnan Province. A case of heart sound signal was of 30 seconds, which was acquired from five auscultation position of each of the collected persons. The database was diagnosed by the medical expert during clinical collection to ensure the accuracy of the data. The One stethoscope from Thinkslabs Medical was used during the auscultation. The labview platform was used to store the signals and the files were saved in lvm format. The sampling rate of the heart sound signal was 5000Hz

2.2. Pre-Processing heart sounds
Before analysis of heart sound signals, the heart sound signal should be pre-processed. There are some environmental noises during heart sound collection. The wavelet transform was used to denoise the original heart sound. According to the selected wavelet base, threshold, and threshold function, the signal is decomposed and denoised. Threshold processing based on selected fixed thresholds and soft threshold functions. In this work, db6 wavelet was selected for 8 layers decomposition, denoising, and reconstructing signal.
2.3. **Heart sound segmentation**

A complete cardiac cycle consists of S1, S1-S2 interval, S2 and S2-S1 intervals. The S1, S1-S2 interval constitutes a systolic phase, and the S2 and S2-S1 intervals constitute a diastolic phase, as shown in Figure 1.

![Figure 1. A complete cardiac cycle](image)

It can be observed from the figure 1 that the systolic phase is shorter than the diastolic phase, the S1 origin can be effectively found according to this phenomenon. In order to better segment the position, the Shannon enrichment energy was used to extract the envelope of the heart sound signal. Segmentation was performed according to the extracted rich envelope and mapped to the original signal to find S1. The pseudo peak was removed by setting the local threshold to find the local maximum point and the local minimum point. S1 was found according to the length of the S2-S1 interval. Taking the S1 as the starting point, the signal of 6 seconds was intercepted for further analysis.

2.4. **Feature extraction and selection**

For higher recognition rate in the next step of identification, the feature extraction and selection is very important. Since the original heart sound signals usually contain some noise, it is need to discard the noise and keep the valid messages. Feature extraction is a process of signal attribute extraction. In speech processing, the parameters are estimated by signal waveforms, which are called feature parameters. The heart sound signal is the same as the voice signal, and it is a waveform signal. In [15] the feature extraction method of the speech waveform worked well, so it can also be tried in the heart sound processing. Moreover in [16] and [17], the feature extraction methods were also applied to the estimation of heart sound characteristic parameters. In this work, the feature extraction was done by using MFCC and LPCC methods which were relatively mature methods in speech processing.

2.4.1. **MFCC**

MEL frequencies are based on the characteristics of the human ear and are relatively close to the sounds that the human ear can recognize [18]. The process of MFCC consists of pre-weighting of the signal, adding a window in frames, doing FFT (fast Fourier transform), filtering through MEL filter, making logarithmic energy transform, and doing DCT (discrete cosine transform). The flow chart of MFCC feature extraction was shown in figure 2.

![Figure 2. The process of MFCC feature extraction](image)

The signal is pre-weighted to improve the signal to noise ratio of high frequency band to get flat signal, which is conducive to the next analysis. The frame length with complete information and stable signal should be selected when framing. The length of the frame selected in this paper was 2048 and the frame shift was 1024, which was decided by experiment. In order to avoid frequency leakage,
A hamming window was selected. A frame of original signal and the signal windowed were shown in figure 3.

![Figure 3](image_url)

**Figure 3.** (a) One frame original heart sound signal and (b) one frame plus window heart sound signal

The heart sound signals were transformed from time-domain to frequency-domain by using FFT. Then the signal was filtered and scaled by the MEL triangular filter group, and transformed to the MEL inverse spectrum domain. The relationship between MEL scale and frequency was shown in figure 4.

![Figure 4](image_url)

**Figure 4.** The relationship between MEL scale and frequency

Mel scale represents the nonlinear characteristic of human ear's perception of frequency. The conversion relation between MEL frequency and linear frequency is shown in formula (1).

\[
Mel(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)
\]  

From each MEL triangular filter, the logarithmic energy of signal was calculated and the heart sound spectrum was compressed. Then the processed signal was transformed into discrete cosine transform (DCT) domain. The characteristic parameters of MFCC were calculated by the formula of discrete cosine transform, which was shown in formula (2).

\[
Mel(n) = \sum_{m=1}^{M} s(m) \cos \left( \frac{\pi n (m - 0.5)}{M} \right); \quad (0 \leq m \leq M)
\]

Where \(Mel(n)\) is the characteristic coefficient, \(s(m)\) is the logarithmic energy obtained, and \(M\) is the number of filters.

2.4.2. LPCC. The extraction of LPCC characteristic parameters is a common method in signal processing. Linear predictive inversion coefficient is the representation of LPC coefficient in inversion spectrum. The signal was pre-weighted and framed firstly, then the LPC coefficient was calculated, and was converted it into the LPCC coefficient. The LPCC parameter extraction process is shown in figure 5.
Figure 5. The process of LPCC feature extraction

A relatively stable signal is obtained by pre-weighting, which is beneficial to the next step, frame segmentation. The frame length was 2048 and the frame shift was also 1024. The LPC coefficient was calculated for each frame signal after framing. LPC coefficient prediction was based on the weighted sum of the past p sample values of the signal to predict the current sample values. The prediction of LPC coefficient was shown in formula (3).

$$\hat{x}(n) = \sum_{i=1}^{p} a_i x(n - i); \ (n = 1, 2, \ldots, n)$$

(3)

Where $\hat{x}(n)$ is the predicted value of $x(n)$, $a_i$ is the predicted linear coefficient, and at this time, it is the p-order linear prediction.

The linear predictive inverse spectral coefficient was calculated by taking the Fourier transform of the signal, and then taking the inverse Fourier transform of the digital analogy function. The formula of the LPCC coefficient obtained from the LPC coefficient was shown in formula (4).

$$\text{lpcc} = \begin{cases} 
\alpha_n ; & n = 1 \\
\alpha_n + \sum_{k=1}^{n-1} \frac{k\hat{h}(k)\alpha_{n-k}}{n} ; & 1 < n \leq p + 1 \\
\sum_{k=1}^{n-1} \frac{k\hat{h}(k)\alpha_{n-k}}{n} ; & n > p + 1 
\end{cases}$$

(4)

Where $\alpha_n$ is the LPC coefficient and $\hat{h}(k)$ is the inverse spectrum of the impulse function.

2.5. Ensemble classification model

Classification recognition is an important step in experiment. The choice of classifiers is even more crucial. In recent years, neural network as an identifier has become a hot research topic. In this paper, BP neural network was still selected, which can transmit information forward, feedback error back, and correct weight to achieve the convergence of the network. Three layers of neural networks were used in the experiment. The characteristic parameters of MFCC and LPCC obtained by feature extraction were 32 dimensions, so 32 input nodes of neural network were selected. Because the purpose of the network was to identify the heart sound signal if normal or abnormal, the output node of the network was 2. The nodes of the hidden layer are based on the experimental results obtained, and then the number of hidden layer nodes was selected to be 10. For the rigor of experimentation, whether the MFCC feature or the LPCC feature was used as a network input, both of their network structure was the same.

3. Results

200 cases of the heart sound signals were randomly selected from the heart sound database, in which 100 were normal heart sound signals and 100 were abnormal heart sound signals. As the training set of the network, 120 cases of cardiac signal (60 cases of normal signal and 60 cases of abnormal signal) were selected. And 80 cases of cardiac signal (40 cases of normal signal and 40 cases of abnormal signal) were selected to detect the network. The network test results with MFCC characteristic parameters as input were shown in Table 1. The network test results of the LPCC feature parameters as inputs were shown in Table 2.

| signal | Correct num | Correct rate(%) | Sensitivity(%) | Specificity(%) |
|--------|-------------|-----------------|----------------|---------------|
| normal | 37          | 92.50           | 88.89          |               |
| abnormal | 35         | 87.50           | 93.02          |               |

In table 1, the correct recognition number of normal heart sounds is 37 cases, and the correct recognition number of abnormal heart sounds is 35 cases. When the MFCC characteristic parameter was used as the input of the BP neural network, the correct recognition rate of the normal heart sound
of the test result was 92.50%, and the correct recognition rate of the abnormal heart sound was 87.50%. At this time, the sensitivity of the network was 88.89%, and the specificity was 93.02%.

Table 2. The test results of LPCC characteristic parameters were taken as the input

| signal   | Correct num | Correct rate(%) | Sensitivity(%) | Specificity(%) |
|----------|-------------|-----------------|----------------|----------------|
| normal   | 34          | 85.00           | 86.96          |                |
| abnormal | 34          | 85.00           | 86.96          |                |

The results in table 2 are the test results which use the LPCC characteristic parameters as input to the BP network. At this time, the number of correct recognition of the normal heart sound signal and the abnormal heart sound signal by the network were both 34 cases. The correct recognition rate of normal heart sound was 85.00%, and the correct recognition rate of abnormal heart sound was 85.00% also. The sensitivity and specificity of the network were both 86.96%.

![Figure 6. The sensitivity and specificity of network](image)

4. Conclusion
The algorithm in this work can help the machine assist the implementation of auscultation, and provide help for the initial diagnosis and screening of congenital heart disease. By comparing the results of MFCC feature and LPCC feature, the features extracted by using MFCC method are better than one by using LPCC. Also the BP neural network is suitable for classification and identification of cardiac signal.

5. Acknowledgment
Thanks to the help of Yunnan Fuwai Cardiovascular Hospital and its team. Thanks to the help of my tutor and other students in the lab.

References
[1] Qu Y, Liu X, Zhuang J, et al. Incidence of Congenital Heart Disease: The 9-Year Experience of the Guangdong Registry of Congenital Heart Disease, China J. Plos One, 2016, 11(7):e0159257.
[2] Miranovi V. The incidence of congenital heart disease: previous findings and perspectives J. Srpski Arhiv Za Celokupno Lekarstvo, 2014, 142(3-4):243-248.
[3] Fu W, Yang X, Wang Y. Heart sound diagnosis based on DTW and MFCC[C]// International Congress on Image and Signal Processing. IEEE, 2010:2920-2923.
[4] Nogueira D M, Ferreira C A, Jorge A M. Classifying Heart Sounds Using Images of MFCC and Temporal Features[C]// Portuguese Conference on Artificial Intelligence. Springer, Cham, 2017:186-203.

[5] Arslan A, Yildiz O. Automated Auscultative Diagnosis System for Evaluation of Phonocardiogram Signals Associated with Heart Murmur Diseases J. Gazi University Journal of Science, 2018, 31(1).

[6] Li L, Wang X, Du X, et al. Classification of heart sound signals with BP neural network and logistic regression[C]// Chinese Automation Congress. 2017:7380-7383.

[7] Li H T, Shi A S, Li K Z. A Modal of the Heart Disease Severity Diagnosis and Evaluation Based on Rough Set Theory and BP Neural Network[C]// the 2nd international conference on software engineering and data mining. 2010:532-537.

[8] Zhang G, Yuan Z. Method of Heart Sound Recognition Based on Wavelet Packet and BP Network J. Research Journal of Applied Sciences Engineering & Technology, 2013, 5(5):1568-1572.

[9] Liu S, Chen G, Chen G. The Application of Wavelet Analysis and BP Neural Network for the Early Diagnosis of Coronary Heart Disease J. Lecture Notes in Electrical Engineering, 2013, 236(1):165-172.

[10] Swati Verma, Tanuja Kashyap. ANALYSIS OF HEART SOUND AS BIOMETRIC USING MFCC & LINEAR SVM CLASSIFIER J. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2014, 3(1):6626-6633.

[11] Wu H, Kim S, Bae K. Hidden Markov model with heart sound signals for identification of heart diseases J. 2010.

[12] Zhong L, Wan J, Huang Z, et al. [Research on biometric method of heart sound signal based on GMM] J. Chinese Journal of Medical Instrumentation, 2013, 37(2):92-95.

[13] Jing X, Ma J, Zhao J, et al. Speaker recognition based on principal component analysis of LPCC and MFCC[C]// IEEE International Conference on Signal Processing, Communications and Computing. IEEE, 2014:403-408.

[14] Kumar P. Automatic Speaker Recognition using LPCC and MFCC J. 2015, 3(4):2106-2109.

[15] Gulzar T, Singh A, Sharma S. Comparative Analysis of LPCC, MFCC and BFCC for the Recognition of Hindi Words using Artificial Neural Networks J. Acta Materialia, 2014, 56(13):3132-3145.

[16] Çelebi G, Sözeri G, Yılmaz A, et al. Mel-frequency cepstral based heart sound signal segmentation for decision support system[C]// Signal Processing and Communications Applications Conference. IEEE, 2017.

[17] Sh-Hussain, Salleh, Kamarulafizam I, et al. Classification of heart sound based on multipoint auscultation system[C]// International Workshop on Systems, Signal Processing and Their Applications. IEEE, 2013:174-179.

[18] Liu D, Wang X, Zhang J, et al. Feature extraction using Mel frequency cepstral coefficients for hyperspectral image classification J. Applied Optics, 2010, 49(14):2670-2675.