Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Hidden Markov model with heart sound signals for identification of heart diseases

Hang Wu, Sahong Kim, and Keunsung Bae

School of Electrical Engineering and Computer Science, Kyungpook National University, Daegu 702-701, Korea
phone: + (82) 53 940 8627, fax: + (82) 53 940 8827, email: whuwuhang@hotmail.com

PACS: 43.60.LQ, 43.80.VJ, 43.80.QF

ABSTRACT

Heart auscultation still remains a dominant method for the diagnosis of heart diseases caused by heart valve abnormalities. But it is very subjective and significantly relies on the interpretation or perception of well-trained physicians. Thus it would be very desirable to develop a computer-aided automated or semi-automated heart sound identification system that can provide more objective diagnositic results. Recently a hidden Markov model (HMM) has been used quite successfully for the classification of heart sounds. In this paper, we have investigated the classification performance of the MFCC-based HMM with heart sound signals by varying the model's number of states, number of mixtures, and analysis frame size in MFCC feature extraction. We carried out the classification experiments using the 325 heart sound data made up of 10 different types of heart sound signals. From this, maximum correct classification rate of 95.08% was achieved when the HMM has 4 states, 8 mixtures with analsys frame size of 20ms for feature extraction.

1. INTRODUCTION

Heart abnormality is often resulted from the turbulent flow of blood in heart vessels. Auscultation is the most widely used and cost effective non-invasive technique for the diagnosis of heart diseases among a variety of diagnostic techniques such as electrocardiogram (ECG), echo cardiogram, etc[1-2]. However, the auscultation based heart disease diagnosis method is very subjective since it largely relies on the interpretation or perception of physicians. To have precise diagnosis of heart diseases with auscultation, physicans need rich diagnostic experiences which will take years to acquire. Thus developing a computer-aided heart disorder diagnosis system is very desirable. Phonocardiogram (PCG), a visual display of the heart sound waveform has proved to provide valuable information of heart conditions such as major components of heart sound and cardiac murmurs for the diagnosis of heart diseases. During recent decades, enormous research efforts have been contributed to developing such automated diagnostic system for the PCG to assist clinicians in making a better diagnosis of heart disorders.

Feature extraction is important in the classification process. Time-frequency based methods have been reported to be suitable for classifying heart disorders and used to characterize the heart sound signals [3-5]. Recently, mel-frequency cepstral coefficient (MFCC) has been used with a hidden Markov model (HMM) for automatic heart sound auscultation and has shown good performance [6]. Furthermore, comparative classification experiments according to different dynamic feature sets based on time-frequency representation have shown that the MFCC features achieved the best classification performance than any other feature sets [7]. As a classifier for the heart sound signals, most researchers use an artificial neural network (ANN) [8-10]. The ANN with hidden layers shows good classification performance, but it is not effective for non-stationary time varying signals such as speech and heart sound. On the contrary, the HMM has proved to be effective for modelling the heart sound signals [11-13] due to its capability of modelling well the time varying and non-stationary signals. In [14], the authors found that HMM outperforms ANN over heart sound classification.

In the MFCC-based HMM automatic system for identification of heart diseases, factors such as the frame size in MFCC feature extraction, the number of HMM states and the number of Gaussian mixtures in each HMM state affect the classification performance, however, such a problem has not been addressed in detail at most referred papers. Thus, in this paper, we have investigated the classification performance of the HMM with MFCC features depending on the analysis frame size, number of states, and number of mixtures in each state. The 13-dimension MFCC including the log energy of the frame was used as feature parameters. Then experimental results are presented with our discussions.

The remaining of this paper is organized as follows. Section 2 will give a brief explanation about the MFCC feature extraction and HMM modelling algorithms for the heart sounds. Then section 3 presents experimental results with discussions. Finally conclusion is given in Section 4.

2. MFCC AND HMM

2.1 Mel-Frequency Cepstral Coefficient

The MFCC, a perceptual representation of the power spectrum of a sound signal, is obtained by taking a discrete cosine transform of logarithmic power spectrum on a nonlinear mel

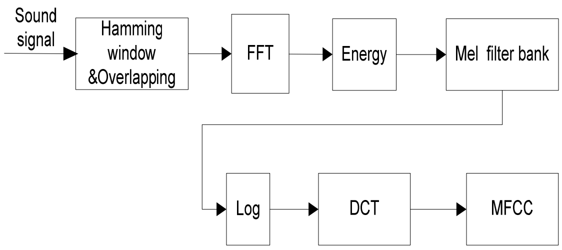

Figure 1. Block diagram of MFCC extraction process

scale of frequency. Figure 1 shows the block diagram of the MFCC feature extraction process.

The signal is first segmented into short frames. Then Hamming window is applied to each frame to reduce the edge effect. The Hamming window with length N is given by eq.(1):

$$w(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}) \quad (1)$$

Denoting the m th frame of the input signal by $x_m(n)$, then its DFT is given like eq.(2):

$$X_m(k) = \sum_{n=0}^{N-1} x_m(n) w(n) e^{-j\frac{2\pi nk}{N}} \quad (2)$$

Then the output energy of the i th filter of the filter bank is calculated using the eq.(3). The filter bank is made up of a certain number of triangular shaped bandpass filters which are centered on equally spaced in the mel frequency domain.

$$E_m(i) = \sum_{l=0}^{N_1-1} | X_m(l) |^2 \ H_i(l) \quad (3)$$

where $H_i(l)$ is the l th weight of the i th triangular filter in the filter bank, and N_1 is the associated number of weights. The mapping from linear frequency to mel frequency is expressed as

$$Mel(f) = 2595\log_{10}(1 + \frac{f}{700}) \quad (4)$$

By taking the discrete cosine transform of the log energy of each filter calculated by eq.(3), we can get the MFCCs as given in eq.(5):

$$c_{m,q} = \sum_{i=1}^{M} \log(E_m(i)) \bullet \cos[\frac{\pi q(i-0.5)}{M}] \quad (5)$$

where M is the number of filters in the filter bank and q is the order of the MFCC. In our work we set q to 12 as used in speech recognition. Including the log energy of the frame as given in eq.(6), we then have 13-dimension MFCCs as feature parameters.

$$E_0 = \log \sum_{n=1}^{N} x_m^2(n) \quad (6)$$

2.2 HMM classification system

A hidden Markov model is a stochastic finite state machine which can control the selection of the states of a sequence of observations as well as the transition probabilities between states. The probability distribution of observations in each state, saying j , is modelled by Gaussian mixtures, expressed as eq.(7)

$$b_j(o) = \sum_{m=1}^{M} w_{jm} N(o \mid \mu_{jm}, U_{jm})$$
$$where \sum_{m=1}^{M} w_{jm} = 1, w_{jm} \geq 0, 1 \leq j \leq N \quad (7)$$

Where M is the number of Gaussian mixtures, N is the number of HMM states, w_{jm}, μ_{jm}, U_{jm} are the weight of m th mixture of j th state, mean vector, covariance matrix, respectively, $N(o \mid \mu_{jm}, U_{jm})$ is a multivariate Gaussian probability distribution function. The transition from state i to state j is controlled by the transition probability a_{ij} .Given the HMM model $\lambda = (A, B, \pi)$ in which A represents the state transition matrix, B represents the probability distribution of observations, and π represents the initial state distribution,and a sequency of observations $O = \{o_1, o_2, ...o_T\}$, classification is carried out by calculating the likelihood score of $P(O \mid \lambda)$.

Figure 2 shows the procedure of training to generate HMMs of the heart sound signals. Using the Baum-Welch algorithm with MFCC features obtained from the same type of heart sound signals, each HMM corresponding to the specific type of heart disease is constructed. In the classification procedure, the MFCCs extracted from the test signal are applied to each HMM and calculate the correspondng $P(O \mid \lambda)$. Then the model which gives the highest value is selected as the classification result. Figure 3 shows the block diagram of the classification procedure.

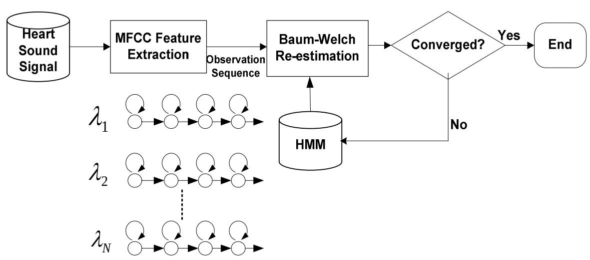

Figure 2. Block diagram of HMM training procedure

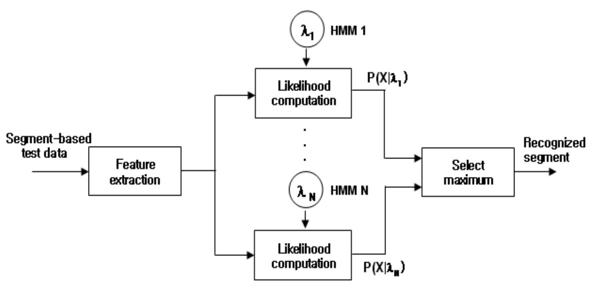

Figure 3. Block diagram of HMM classification procedure

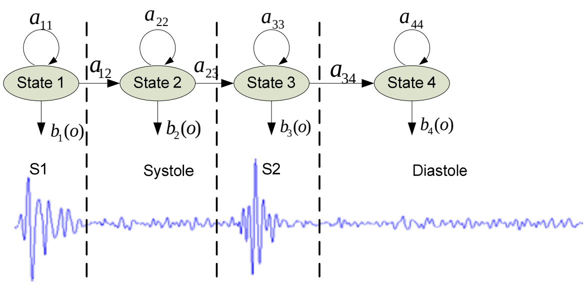

Figure 4. A four-state left-to-right HMM for a cycle of normal heart sound signal

One period of a heart sound cycle consists of four components, namely S1, systole, S2 and diastole, as shown in Figure 4. So, in modelling the heart sound signal with HMMs, a four-state left-to-right HMM is first adopted. Each state in Figure 4 can be considered as each sound component since the signal characteristics in the state can be viewed to be homogeneous [14].

3. EXPERIMENTS AND DISCUSSIONS

In this work, 325 manually segmented heart sound cycles corresponding to 10 types of heart diseases, namely, Normal sound (NM), Innocent Murmur (IM), Splitting (SPI), Mitral Value Prolapse (MVP), Gallop (GAL), Ventrical Septal Defect (VSD), Aortic Stenosis (AS), Mitral Stenosis (MS), Pulmonic Stenosis (PS), Coarctation of Aorta (CA), were used for classification experiments. The original signals were obtained from the clinical audio CDs [15], and were resampled to 8KHz with 16bit resolution, mono format. The numbers of sound cycles associated with their types of heart diseases are listed in Table 1.

Table 1. Number of data samples for each heart disease type

TP	*NM*	*IM*	*SPI*	*MVP*	*GAL*	*VSD*	*AS*	*MS*	*PS*	*CA*
Num	20	20	30	28	38	30	39	40	40	40

To overcome the data insufficiency problem, leave-one-out method was used in the classification. That is, only one data sample is used for test while the left data samples are used for training, and that process is repeated until all the data are tested once. The classification performance is evaluated in the form of correct classification accuracy rate (CCAR). We carried out comparative experiments with respect to different parameter values of analysisi window size in MFCC feature extraction, different number of HMM states and different number of Gaussian mixtures to investigate the influence of these factors on the classification performance.

Table 2 shows the classification results with the number of HMM states equal to 4. The first column represents different number of mixtures while the first row represents different analysis frame size, i.e., window length with fixed window shifting rate. By observing the results of each column, we can find that the best performance is achieved when the number of mixtures is set to 8. In the light of this observation, we can say that neither too small nor too large number of mixtures could produce the best results because too small number of mixtures couldn't model the spectral variability of heart sounds well in each HMM state while too big number would result in mixture-heavies problem. Similarly, we can observe the results in each row to investigate the influence of the window length on the classification performance. Firstly by observing the results of the first row and the fourth row, we can find that the performance decreases as the window length gets smaller. This might be due to that a large window length could capture enough distinctive characteristic information of the signal within each frame when the number of mixtures is 4. Similarly, for the case in which the number of mixtures is 24, the severity of mixture-heaviest problem within a larger window length is smaller than that of within a smaller window length even though they are both affected by mixture-heaviest problem. However, excluding these two extreme cases, for the two cases with the number of mixtures equals to 8 or 16, the previous changing trend doesn't hold any more, indicating that we should make a compromise between the number of mixtures and the window length. The maximum corrct classification rate of 95.08% was achieved when the pair parameters of window length and window shift rate are set to 20ms/10ms and the number of mixtures is set to 8.

Table 2. CCAR (%), MFCC=13, NumState=4

	50ms/10ms	**25ms/10ms**	**20ms/10ms**	**15ms/10ms**
4	93.28	91.48	91.17	90.23
8	***94.22***	***94.51***	***95.08***	***92.10***
16	93.01	93.62	91.78	90.29
24	92.10	89.62	88.73	87.24

Table 3. CCAR (%), MFCC=13, NumState=5

	50ms/10ms	**25ms/10ms**	**20ms/10ms**	**15ms/10ms**
4	92.70	90.27	91.49	89.67
8	***94.85***	***94.55***	***94.55***	***93.03***
16	93.94	92.37	90.87	91.82
24	92.43	89.03	88.08	86.63

Table 4. CCAR (%), MFCC=13, NumState=3

	50ms/10ms	**25ms/10ms**	**20ms/10ms**	**15ms/10ms**
4	88.35	85.31	86.24	82.86
8	92.00	88.73	89.66	***89.96***
16	93.28	***92.10***	***91.21***	89.09
24	***94.53***	89.70	89.70	87.58

Table 3 shows the classification results when the number of HMM states equals to 5. By analyzing the results in Table 3 in a similar way like Table 2, we can find that the maximum classification accuracy in each column is achieved when the number of mixtures equals to 8, but there merely exists slight difference between the performances with respect to the window length parameter. The best accuracy rate of 94.85% was achieved in this table. The performance changing trend versus the window length parameters with the number of mixtures fixed is not the same as before.

Table 4 shows the classification results when the number of HMM states equals to 3. From this table, the CCAR changing trend is significantly different from that of the previous two tables due to the effect of the number of HMM states, set to 3. It's obvious that the overall performance decreases signifi-

Table 5. Classifictication confusion matrix under the condition that the Number of HMM states equal to 4, the number of Guassian mixtures equal to 8, the window length and window shift values equal to 20ms/10ms.

	NM	*IM*	*SPI*	*MVP*	*GAL*	*VSD*	*AS*	*MS*	*PS*	*CA*
NM	**20**									
IM		**20**								
SPI			**25**		***5***					
MVP				**22**	***6***					
GAL					**38**					
VSD						**28**				***2***
AS	***2***						**36**	***1***		
MS								**40**		
PS									**40**	
CA										**40**

cantly compared to the previous two cases of the number of HMM states.We may infer that the 3-state left-to-right HMM in this case couldn't model precisely the underlying structure of one heart sound cycle which includes four components, namely S1, systole, S2 and diastole in order.

In short, a 4-state HMM with the number of mixtures 8, and the window length with window shifting rate 20ms/10ms are appropriate among all the cases we have investigated. Classification confusion matrix in this case is given in Table 5. The column represents the real type of heart disease, while the row represents the category into which the heart sound signal was classified. From this matrix, the two types of heart diseases MVP, SPI are relatively prone to be wrongly assigned into the GAL type due to their resemblance in shape, i.e. there exsits splitting or click sound in their waveforms.

4. CONCLUSION

In this paper, we have investigated the classification performance of the HMM with MFCC features depending on the analysis frame size, number of states, and number of mixtures in each state. Using the 325 manually segmented heart sound cycles corresponding to 10 types of heart diseases, we achieved maximum correct classification rate of 95.08% when the HMM has 4 states, 8 mixtures with frame size of 20ms. Even though a 4-state HMM model with appropriate parameter values of window length and the number of Gaussian mixtures has shown to produce the best result, there is only slight difference from that of the 5-state HMM. This may be due to that many heart sound signals have splitting or click components as well as mumur components in silent durations of systole and diastole, which makes the one cycle mostly viewed to have 5 components. From this perspective, a 5-state HMM to some extent can be more appropriate and robust to model the various types of heart sound signals.

In our work, we think the size of the data was not enough to validate the results. Thus, more classification experiments with large size of the various types of heart disorder PCG data are necessary for the future work. In addition to that, classification experiments were done using the manually segmented heart sound cycles. So future work will be also dedicated to developing a robust automatic segmentation algorithm and combining it with MFCC feature based HMM classifier to implement a complete automatic diagnositic system of heart disorders.

ACKNOWLEDGEMENT

This work was supported by the Grant of the Korean Ministry of Education, Science and Technology (The Regional Core Research Program/Anti-aging and Well-being Research Center).

REFERENCES

[1] A. Leatham, Auscultation of the Heart and Phonocardiography, second edi. Churchill Livingstone, London, Edinburgh, 1975.
[2] N.C. Nanda, Doppler Echocardiography, second edi., Lea &Febiger, London, 1993.
[3] Debjasis, F., Durand, l.g. "Time Frequency Analysis of Heart Murmurs. Part II: Optimization of Time-Frequency Representations and Performance Evaluation," Medical & Biological Engineering&Computing 35, 480-485,1997.
[4] Bently, P.M.,. "Time Frequency Analysis of Native and Prosthetic Heart Valve Sounds," PhD Thesis, Electrical and Electronics Department, University of Edinburgh, 1996.
[5] Leung, T.S., White, P.R., Collis, W.B., Brown, E., Salmon, A.P., "Classification of Heart Sounds Using Time Frequency Method and Artificial Neural Networks," Proceedings of the 22nd Annual International Conference of the IEEE-EMBS, Chicago, 988-991, 2000.
[6] Sunita Chauhan, Ping Wang, etc. "A Computer-aided MFCC-based HMM System for Automatic Ausculatation," Computers in Biology and Medicine, Vol, 38 , pp. 221-233. 2008.
[7] A.F. Quiceno-Manrique, J.I. Godino-Llorente,etc., "Selection of Dynamic Features Based on Time-Frequency Representations for Heart Murmur Detection from Phonocardiographic Signals," Annals of Biomedical Engineering, Vol.38, pp.118-137, 2010.
[8] Cathers, I., "Neural Network Assisted Cardiac Auscultation," Artificial Intelligence. Med. 7, pp.53-66, 1995.
[9] Degroff C, Bhatikar S, Hertzberg J, Shandas R,etc, "Artificial Neural Network-based Method of Screening Heart Murmur in Children," Circulation 103, pp. 2711-2716. 2001.
[10] Bhatikar, S.R., DeGroff, C., "A Classifier Based on Artifical Neural Network Approach for Cardiac Auscultation in Pediatrics," Artificial Intelligence.Med. 33, pp. 251-260, 2005.
[11] M. El-Hanjouri, W.Alkhaldi, N.Hamdy, O.Abdel Alim, "Heart Diseases Diagnosis Using HMM," IEEE MELECON, pp. 489-491. 2002.
[12] Ricke, A.D., Povinelli, R.J., Johnson, M.T , "Automatic Segmentation of Heart Sound Signals using Hidden Markov Models," Computers in Cardiology. pp. 953-956. 2005.
[13] SE Schmidt, E Toft, C Holst-Hansent,ect. "Segmentation of Heart Sound Recordings From An Electronic Stethoscope by a Duration Dependent Hidden-Markov Model," Computers in Cardiology, 35, pp. 345-349. 2008.
[14] Chung, Y., "A Classificaiton Approach for Heart Sound Signals Using Hidden Markov Models," SSPR/SPR 375-383, 2006.
[15] Daniel Mason, "Listening to the heart: A Comprehensive Collection of Heart Sounds and Murmurs," F. A. Davis Company, Philadelphia, 2000.