Automatic Segmentation of Heart Sound Signals Using Hidden Markov Models

AD Ricke[1,2], RJ Povinelli[2], MT Johnson[2]

[1]GE Healthcare, Milwaukee, WI, USA
[2]Marquette University, Milwaukee, WI, USA

Abstract

The monitoring of respiration rates using impedance plethysmography is often confused by cardiac activity. This paper proposes using the phonocardiogram as an alternative, since the process of respiration affects heart sounds. As part of this research, a technique is developed to segment heart sounds into its component segments, using Hidden Markov Models. The heart sounds data is preprocessed into feature vectors, where the feature vectors are comprised of the average Shannon energy of the heart sound signal, the delta Shannon energy, and the delta-delta Shannon energy. The performance of the segmentation system is validated using eight-fold cross-validation.

1. Introduction

Respiration is an important physiological signal often monitored in clinical settings. An inexpensive, non-invasive method of measuring respiration is impedance plethysmography. In impedance plethysmography, respiration is measured by detecting the variation in impedance between the ECG electrodes on each side of the chest [1]. A small, high frequency current is applied across these electrodes, and the patient monitoring device measures the change in voltage as the impedance changes between the leads. However, impedance plethysmography can accidentally detect cardiac activity as the impedance changes. This is of particular concern when the patient stops breathing, because the patient monitoring equipment can detect these cardiac artifacts as respiration. This can cause periods of apnea to go undetected. Also, obstructive apnea may go undetected because the chest wall continues to move as the patient struggles to breath. This is undesirable.

Heart sounds are affected by respiration; therefore, it is feasible that a patient monitor could record these sounds and use them to determine respiration rates. Previous research at GE Healthcare [2] proved the feasibility of this concept using an autocorrelation technique. This technique determined respiration at lower rates (7.5, 10, 15 Breaths per Minute (BrPM)), but failed at higher rates (30 BrPM). The purpose of this study is to develop an algorithm to determine the rate of respiration from the heart sound signal. Such a system could not be confused by the electrical impulses from cardiac activity, and it would be better suited for detecting an apnea condition than the traditional impedance method.

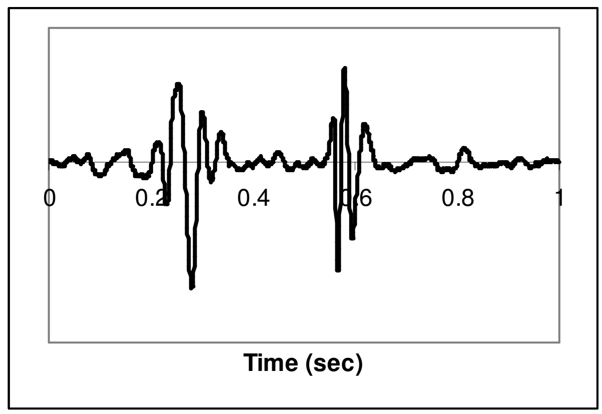

Figure 1. Example Phonocardiogram

The first phase in this project is to develop a robust technique for segmenting heart sounds (phonocardiogram) into its component segments: the S1 sound, the systole period, the S2 sound, and the diastole period. Fig. 1 provides an example of a phonocardiogram. The heart generates the first heart sound (S1) when tricuspid and mitral heart valves snap shut, and the heart generates the second heart sound (S2) when the pulmonary and aortic values snap shut.

As a person inhales, the lungs expand and apply pressure against the heart. The left lung puts greater pressure against the heart than the right lung because the heart sits on the left side of the chest. This pressure differential causes the valves on the left side of the heart to snap shut after the valves on the right side, which causes a split in the S2 sound.

Since the S2 sound splits during inhalation [3], segmenting the heart sounds into its components is important for detecting the heart sound split. The rate of the split of the S2 heart sound can be converted directly to a respiration rate. In addition, respiration causes a change in the beat-to-beat interval, called respiratory sinus arrhythmia. As the patient breaths, the beat-to-beat interval cycles up and down at the same frequency as respiration [4].

Previously attempts at an algorithm for segmentation resulted in a 93% success rate [5]; however, implementing this algorithm is prone to error, and it is sensitive to changes in pre-processing and setup parameters. The purpose of this study is to develop a robust segmentation algorithm for segmenting heart sounds into its components using Hidden Markov Models.

2. Methods

2.1. Data acquisition

The physiological data was collected at GE Healthcare from nine different subjects, using CardioLab® system and a Dash® family patient monitor. The CardioLab® system sampled data from ECG leads I, II and III, and an electronic stethoscope signal at 977 samples per second. Each subject breathed according to a fixed protocol, with the assistance of a metronome.

2.2. Pre-processing

The original heart sound data was unlabeled. Using the PhysioNet toolkit [6], the heart sound data and ECG data were converted to the MIT file format, interpolated to a 1k sampling rate, and annotated with QRS and heart sound annotations. The heart sound annotations were manually adjusted for accuracy.

We divided the data files into a “clean” set and a “dirty” set. The “clean” data files have a high SNR and little or no visible noise, while the “dirty” data files have a low SNR and/or highly visible noise. This study concentrated on the “clean” data files (46 files, ~2286 seconds total).

The system filters the original heart sound signal using a band-pass filter with cutoff frequencies at 30 Hz and 200 Hz. Next, the signal is normalized according to the equation (1).

$$x_{norm}(k) = \frac{x(k)}{\max_i \left(|x(i)|\right)} \tag{1}$$

Then, it calculates the average Shannon energy in continuous 0.04-second segments, with 0.02 seconds of overlap per segment. As explained in [2], the Shannon energy emphasizes the medium intensity signals and attenuates the high intensity signals. This tends to make medium and high intensity signals similar in amplitude.

The system calculates the average Shannon energy of each frame, using equation (2), where x_{norm} is the normalized heart sound signal.

$$E_s = -1/N \sum_{i=1}^{N} x^2{}_{norm}(i) * \log x^2{}_{norm}(i) \tag{2}$$

Then, the system normalizes the average Shannon energy over all of the frames, using equation (3), where $E_s(t)$ is the average Shannon energy for frame *t*, $\xi(E_s(t))$ is the mean value of $E_s(t)$ and $\sigma(E_s(t))$ is the standard deviation of $E_s(t)$.

$$P_a(t) = \frac{E_s(t) - \xi(E_s(t))}{\sigma(E_s(t))} \tag{3}$$

The normalized average Shannon energy ($P_a(t)$) is the primary feature used in this study.

2.3. Mel-spaced filterbanks

Next, the system extracts the spectral characteristics from the heart sound signal. Since the average duration of the S1 sound is 0.16 seconds (empirical), the system divides the signal into 0.15-second frames, with 0.02 seconds of overlap for each frame.

According to [7], the frequency spectrum of S1 contains peaks in the 10 to 50 Hz range and the 50 to 140 Hz range, while the frequency spectrum of S2 contains peaks in the 10 to 80 Hz range, the 80 to 200 Hz range, and the 220 to 400 Hz range. As a result, this study limits the spectral feature extraction between the frequencies of 10 Hz and 430 Hz.

Mel-Spaced filter banks provide a simple method for extracting spectral characteristics from an acoustic signal. This method involves creating a set of triangular filter banks across the spectrum. The filterbanks are equally spaced along the mel-scale, as defined in equation (4).

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

Equal spacing on the mel-scale provides non-linear spacing on the normal frequency axis. This non-linear spacing means that there are numerous, small banks at the lower frequencies and sparse, large banks at the higher frequencies [7]. Since most of the energy of the heart sounds is in the lower frequency ranges, using a mel-scale matches the frequency spectrum of the heart sounds.

Each triangular filter is multiplied by the discrete Fourier transfer of the heart sound frame and summed. This creates a set of frequency bins, where each bin represents a portion of the frequency spectrum.

2.4. Regression coefficients

The final feature extraction step is to calculate a set of regression coefficients. Regression coefficients are used to represent the changes in each feature over time. The system computes the first order regression (delta coefficients) and the second order coefficients (delta-delta coefficients) using the following regression formula [7].

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta (c_{t+\theta} - c_{t-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{4}$$

The system combines the Shannon energy, the spectral features, and the regression coefficients into a single feature vector per frame. It stores these feature vectors for later use in the training and testing of the heart sound Hidden Markov Models.

2.5. Heart sound Hidden Markov Model

An Hidden Markov Model (HMM) is a probabilistic state machine where the states of the machine are unobservable, but the outputs of the state machine are observable. A HMM can model signals where the outputs are discrete or continuous. An example of a discrete HMM is a HMM that models the series of heart sound labels over time. An example of a continuous HMM is a HMM that models the Shannon energy feature over time.

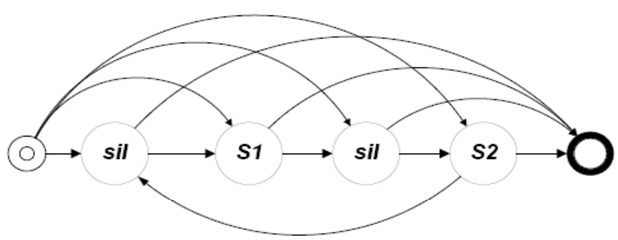

Figure 2. Heart Sound Markov Model

One can model the phonocardiogram signal as a four state HMM. The first state corresponds to the S1 sound, the second state corresponds to the silence during the systolic period, the third state corresponds to the S2 sound, and the fourth state corresponds to the silence during the diastolic period (see Figure 1). This model ignores the possibility of the S3 and S4 heart sounds, because these heart sounds are not germane to the task of recognizing respiration rates from heart sound data. Additionally, these sounds are difficult to hear and record; therefore, they are most likely not noticeable in our heart sound data.

This four state HMM is useful for modeling the sequence of symbols (or labels) of the phonocardiogram; however, it is too simple to accurately model the transitions between sound and silence. One solution is to embed another HMM inside of each of the heart sound symbol states. The embedded HMM models the signal as it traverses a specific labeled region of the signal. Using this combined approach, we can model both the high-level state sequence of our signal (S1-sil-S2-sil) and the continuous transitions of the signal. This type of model is similar to how a speech processing system has a high-level probabilistic grammar to model the transition of words or phonemes, and an embedded HMM for each phoneme [8].

All of the experiments utilized an eight state HMM for the S1 sounds, a six state HMM for the S2 sound, and a three state HMM for each silence period. The number of states where calculated by taking the average duration of each heart sound, and dividing by the frame duration. For example, the S1 sound has an average duration of 160 milliseconds and the frame step size is 20 milliseconds; therefore, it can be represented by eight states (160 ms / 20 ms = 8).

In addition, the experiments utilized a four state grammar that represented the state model given in Figure 1. The probabilities for this model were learned using a discrete HMM where the label files were used to train the model. The resultant HMM represents the symbol transitions of the phonocardiogram. We manually translated the discrete HMM into a grammar for use with the HTK toolset [8].

2.6. Validation

Two different methods for measuring the performance of the system are employed: *frame error rate* and *model error rate*. To determine the frame error rate, we compare each frame of the labeled signal to the output signal. We calculate the error rate of the system by dividing the number of mismatched frames by the total number of frames in the system.

To determine the model error rate, we calculate the center of the heart sound label and the center of the learned heart sound, and calculating the difference between these centers. The system marks a labeling as a success if the delta between these centers is less than 50 milliseconds. Then, the error rate is the number of mismatched S1 or S2 labels divided by the total number of sound labels in the system.

We measure both the frame error rate and the model error rate, for both the training of the system and the validation of the system. Since there were only clean files for eight of the patients, eight-fold cross-validation was used. Finally, the noisy files were validated against the model where the model was trained with only clean files.

3. Results

To determine the best features to use for this study, experiments were run using various combinations of energy, Shannon energy, Mel-spaced filterbanks and Mel-frequency Cepstral coefficients. The lowest training error rate came from systems where we only used the Shannon energy (SE), the delta Shannon energy (D), and the acceleration of the Shannon energy (A). Table 1 clearly shows that this feature vector has the lowest frame error rate over the training set.

Table 1. Error Rates for Various Features

Feature Types	Frame error rate	Model error rate
SE	0.093 ± 0.047	0.032 ± 0.045
SE_D	0.091 ± 0.049	0.026 ± 0.050
SE_D_A	0.089 ± 0.035	0.021 ± 0.035
MELSPEC_SE_D_A	0.098 ± 0.034	0.019 ± 0.026

We performed an eight-fold cross validation using an SE-Delta-Acceleration feature vector with and without the Mel-spaced filterbanks. Also, we tested the clean model using the noisy phonocardiogram files. Table 2 and Table 3 provide the cross validation results and the noisy file results, respectively.

Table 2. Cross-Validation Results, Clean Files

Feature Types	Frame error rate	Model error rate
SE_D_A	0.093 ± 0.035	0.024 ± 0.035
MELSPEC_SE_D_A	0.10 ± 0.042	0.026 ± 0.035

Table 3. Test Results, Noisy Files

Feature Types	Frame error rate	Model error rate
SE_D_A	0.23 ± 0.17	0.19 ± 0.25
MELSPEC_SE_D_A	0.22 ± 0.15	0.16 ± 0.19

4. Discussion

Shannon Energy features with and without the Mel-spaced filterbank features are nearly identical in performance. Shannon Energy features are better suited for lowering the frame error rate while Mel-spaced filterbanks are better suited for lowering the model error rate. Mel-spaced filterbanks are marginally better as features for noisy phonocardiograms than clean phonocardiograms. The selection of the feature set becomes a tradeoff between processing speed and noise immunity.

The noisy file set has an increased error rate for both frame error rates and model error rates. Since the heart sound spectrums and the typical ambient noise spectrums are in similar frequency ranges, this is expected. Further work will include using a subtraction technique to remove ambient noise from the heart sound signal. This could be in the form of a second electronic stethoscope that is placed on the body away from the thorax. Subtract the secondary stethoscope's signal from the primary stethoscope's signal to remove the additive noise in the system.

5. Conclusions

The goal of this study was to develop a robust segmentation algorithm for segmenting heart sounds into its components, using Hidden Markov Models. We have accomplished this goal by achieving a 9% frame error rate rate, and a 2% model error rate rate, during an eight-fold cross validation. The model accuracy of 98% from the Hidden Markov Model approach is an improvement over the 93% correctness in the automatic identification of S1 and S2 from [5]. Using this segmentation technique, we can extract additional features that are useful for extracting the respiration rate from the phonocardiogram, such as detecting the splits in the S2 sound or the beat-to-beat interval.

Acknowledgements

The authors are grateful to GE Healthcare and Don Brodnick for providing the data for this project.

References

[1] Bronzino JD. The Biomedical Engineering Handbook. New York: CRC Press;1995.

[2] Thomas A. Heart Sound Variation White Paper. 2004. GE Healthcare. Internal Use Only.

[3] Physical Examination of the Heart [online]. 2004 [cited 2004 Sep 13]. Available from: URL: http://www2.umdnj.edu/~shindler/exam.html

[4] Lingen Z, Reisman S, Findley T. Respiration Derived from the Electrocardiogram during Heart Rate Variability Studies. Computers in Cardiology 1994; 123-4.

[5] Lang H, Lukkarinen S, Hartimo I. Heart Sound Segmentation Algorithm Based on Heart Sound Evenlogram. Computers in Cardiology 1997; 105-8

[6] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13).

[7] Laing H, Lukkarinen S, Hartimo I. A Heart Sound Segmentation Algorithm Using Wavelet Decomposition and Reconstruction. Proceedings of the 19th Annual International Conference of the IEEE 1997. 4:1631.

[8] Young et. al. The HTK Book (for HTK Version 3.2.1). Cambridge University Engineering Department; 2002.

Address for correspondence

Anthony Ricke
GE Healthcare
8200 West Tower Avenue
Milwaukee, WI 53223, USA
Anthony.Ricke@med.ge.com