Detection and Identification of Heart Sounds Using Homomorphic Envelogram and Self-Organizing Probabilistic Model

D Gill[1,3], N Gavrieli[2], N Intrator[1]

[1]Tel Aviv University, Tel Aviv, Israel
[2]Technion IIT, Haifa, Israel
[3]The Hebrew University, Jerusalem, Israel

Abstract

This work presents a novel method for automatic detection and identification of heart sounds. Homomorphic filtering is used to obtain a smooth envelogram of the phonocardiogram, which enables a robust detection of events of interest in heart sound signal. Sequences of features extracted from the detected events are used as observations of a hidden Markov model. It is demonstrated that the task of detection and identification of the major heart sounds can be learned from unlabelled phonocardiograms by an unsupervised training process and without the assistance of any additional synchronizing channels.

1. Introduction

Analyzing heart sounds by phonocardiogram (PCG) enables one to obtain valuable information about the cardiovascular system and especially about the heart valves functioning. Applications based on phonocardiography include detection of pulmonary hypertension, ventricular dysfunction, coronary artery disease, and cardiomyopathies. The cardiovascular system, which consists of the heart and the blood vessels, transports blood to all parts of the body in two circulations: pulmonary (lungs) and systemic (the rest of the body). The heart is a four-chambered pump with two *atria* (right and left), which collect the blood from the veins and two *ventricles* (right and left), which pump out the blood to the arteries. The right side of the heart pumps blood to the pulmonary circulation, and the left side pumps blood to the systemic circulation. The blood from the pulmonary circulation returns to the left atrium, and the blood from the systemic circulation returns to the right atrium. The blood flow is controlled by two sets of valves: the *Atrioventricular* (AV) valves (*tricuspid* and *mitral*) between the atria and the ventricles, and the *Semilunar valves* (aortic and pulmonary) between the ventricles and the arteries. The AV valves prevent backflow of blood from ventricles to atria during ventricular contraction (*systole*). The Semilunar valves prevent backflow of blood to ventricles during ventricular relaxation (*diastole*). The valves are opened and closed passively, and modulated by changes in the contractility of the heart, the compliance of the chamber walls and arteries and the developed pressure gradients.

Most investigators agree that heart sounds originate from the vibrations of the whole cardiovascular system and not only by the movement of the valve leaflets [1]. A normal cardiac cycle contains two major audible sounds: the *first heart sound* (S1) and the *second heart sound* (S2). As soon as the ventricular pressure exceeds the atrium pressure, the mitral and the tricuspid valves close and the vibrations of S1 begin. At the end of the ventricular systole and the beginning of ventricular relaxation, S2 occurs following the closure of the aortic and the pulmonary valves. The ventricle pressure drops steeply, and when it falls below the atrial pressure, the mitral valve opens, and the rapid filling phase begins, with a possibility of another audible sound, S3, at its end. A forth heart sound, S4, may be heard sometimes due to atrial contractions displacing blood into the distended ventricles. Murmurs are sounds caused by certain cardiovascular diseases and defects.

Phonocardiography is a noninvasive, low-cost but accurate monitoring method for valves functioning; it is easily repeatable with no risk to the patient. However, heart diagnosis by auscultation requires high skills and experience of the listener. Automatic detection and identification of heart sounds plays an essential role in automatic diagnosis of phonocardiograms, especially in cases when additionally synchronizing channels like Electrocardiograms (ECG) and carotid pulse (CP) are not present.

Segmentation refers to partitioning the PCG signal into cardiac cycles, and detection of the main events (e.g.: S1, S2, S3, S4, murmurs, and other sounds) and intervals (systole, diastole) in each cycle. Several methods of PCG segmentation have been developed. Some of them use the ECG recording and/or the carotid pulse as reference. Others are based on the PCG channel only. Some of these

0276–6547/05 $20.00 © 2005 IEEE

Computers in Cardiology 2005;32:957–960.

methods are based on the information conveyed by the envelope of the signal and attempt to detect certain events where they cross a predetermined threshold. These methods include envelope extraction using discrete wavelet decomposition and reconstruction [2] or using the magnitude of the analytic signal formed using the PCG and its Hilbert transform [3]. Both methods can lead to several problems like missing low energy events or detecting high-energy artifacts as cardiac activity or detecting extra peaks, which originate from S2 splits. Another method used statistical supervised learning approach, which was based on manually annotated training phonocardiograms [4].

In this study we present a sample based learning tool for automatic detection and identification of S1, S2 and other sounds in phonocardiograms without any additional assisting channels. Moreover, the model training is done by using phonocardiograms with no annotations in a self-organization process.

2. Methods

In the present approach, homomorphic filtering [5] is used to extract a smooth envelogram, which enables the detection of events that are suspected to be S1, S2 or others. The advantage of such *homomorphic envelogram* is its scalable smoothness, which handles the problems of splits and serrated peaks. We use hidden Markov model (HMM) whose states simulate the true events and the observations within each state include features extracted from the envelogram's peaks.

A monocomponent AM-FM signal can be expressed as a product of its amplitude modulation (AM) and frequency modulation (FM) components:

$$x(t) = a(t) \cdot f(t) \qquad a(t) > 0 \tag{1}$$

where $a(t)$ is the AM component or instantaneous amplitude (IA) and $f(t)$ is the FM component and has the form of :

$$f(t) = \sin(\phi(t)) \tag{2}$$

where $\phi(t)$ is the instantaneous phase and $\omega(t) = d\phi(t)/dt$ is the instantaneous frequency (IF). By assuming a simple approximated model by which the PCG is a narrow-band non-stationary signal, we can express it as a monocomponent AM-FM signal (see [6]).

We denote:

$$\hat{x}(t) = \ln|x(t)| . \tag{3}$$

In cases where $x(t) = 0$ we add a small positive value, and then we have

$$\hat{x}(t) = \ln a(t) + \ln|f(t)| . \tag{4}$$

By applying an appropriate linear low-pass filter on $\hat{x}(t)$ we can eliminate the FM component, which is characterized by rapidly variations in time. We denote the low pass filter system by L and the filtered signal by $\hat{x}_l(t)$. Because L is a linear system we obtain:

$$\hat{x}_l(t) = L[\hat{x}(t)] = L[\ln a(t)] + L[\ln|f(t)|] \tag{5}$$

By using a low-pass filter whose pass-band covers the typical frequencies of the AM component and attenuates the typical high frequencies of the FM component, we obtain:

$$\hat{x}_l(t) = L[\ln a(t)] + L[\ln|f(t)|] \approx \ln a(t) . \tag{6}$$

The reversal procedure is done by an exponential operation:

$$\exp[\hat{x}_l(t)] \approx \exp[\ln a(t)] = a(t) , \tag{7}$$

$f(t)$ can be recovered in a similar way. A signal and its recovered AM component (homomorphic envelogram) are shown in Fig.1.

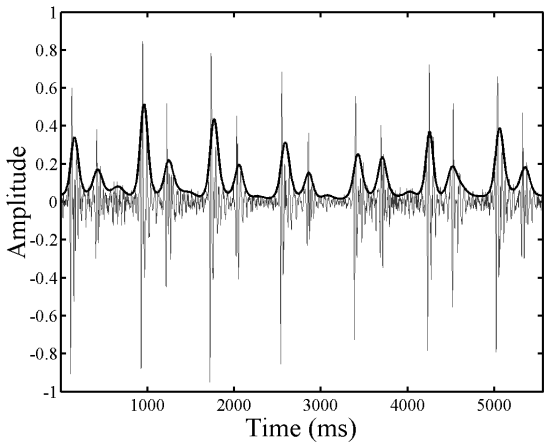

Figure 1. A phonocardiogram (gray) and its homomorphic envelogram (black).

The Shannon energy of a bounded signal $x(t)$, which has real values from –1 to 1, is defined as:

$$E(t) = -x^2(t) \cdot \log x^2(t) . \tag{8}$$

The Shannon energy emphasizes medium intensity

amplitudes and attenuates low amplitudes and amplitudes close to 1 [2]. Converting the signals (normalized by their maximal absolute value) to their Shannon energy prior to the homomorphic filtering was used due to its significant improvement of the final results. Due to the homomorphic filtering, there are small deviations between the envelogram peak and the maximum absolute value of the original peak. We define *event location* as the time of the maximal absolute value of the heart sound. In order to find the exact time of event occurrence, i.e., the time of the maximal absolute value of the signal within a symmetric window with time duration of 250ms surrounding the envelogram's peak, was marked. Choosing all event locations prevents misses of events as usually happens when using a static or dynamic threshold [2]. The peaks may not only be due to S1 or S2, but also due to other events. In order to identify the type of the detected event there is a need to run a parser on the selected events sequence.

HMM is a stochastic finite state automaton, which generates a sequence of vector observations [7]. The model is parameterized by: a vector Π whose entries $\{\pi_i\}$ are the initial state probabilities, a matrix A, whose entries $\{a_{ij}\}$ are the transition probability from state i to state j and the probability densities $p(x|i)$ describing the distribution of the observation x emitted from state j.

The Baum-Welch algorithm estimates the HMM parameters, using unlabeled sequences of observations. The Viterbi algorithm computes the most probable sequence of states given the model parameters and a sequence of observations.

HMM seems natural for modeling cardiac activity [4,8], where each phase of the cardiac cycle corresponds to a state, but when considering the probabilistic character of it, the problem of the duration at each state arises. According to the transition probability definition, the standard discrete-time finite-state HMM allows duration probability in state i according to the geometric distribution:

$$p_i(d) = (a_{ii})^{d-1}(1-a_{ii}) \tag{9}$$

where a_{ii} is the self-transition probability of state i and d is the duration time. This distribution is not appropriate for PCG (nor ECG; see [8]) modeling. Another problem is the features selection for producing the observations sequence. The algorithm suggested here, handles the duration time probability problem by using time durations as features.

A features vector is extracted from each detected event. A comparative study was done in order to evaluate the contribution of different features, which led to a choice of six. The features extracted from each event, selected by the detection process, were: time durations form preceding peak to current and to the following one, amplitudes of preceding, current and following peaks and the second derivative of current peak.

When the training process is complete, the model states have to be identified according to their sound type. Since empirically the transition from detected S1 to detected S2 shows much higher probability than other transitions, we find the maximal transition probability, denoted by $a_{i^*j^*}$ and identify the state i^* as S1 and state j^* as S2.

3. Results

44 phonocardiograms were taken from 17 subjects, aged from 13 to 72 years, and were recorded for 30-60 seconds from several auscultation locations. The environmental conditions included: quiet office, hospital computed tomography (CT) room (noise from the CT scanner, patients moving, and talking), and hospital echo room (noise from patients moving, and people talking). The recording was done using a passive piezoelectric sensor at sampling rate of 4kHz. The recordings were filtered with a pass band from 20Hz to 250Hz.

Using the Baum-Welch algorithm, HMM with three states and a multivariate normal distribution of observations was trained. The training set consisted of sequences of features vectors that were extracted from all the event locations.

The algorithm performance evaluation cannot be done in the regular way of calculating the error percentage on a given test samples set. For each of S1 and S2 we define a true positive (TP) when the corresponding heart sound is detected within a symmetric appropriate window and is the nearest detected event to a human expert annotation. A false negative (FN) is considered when there is a failure to detect the event correctly within a window. A false positive (FP) is considered when there is an event detection outside the appropriate window or if the detection is within the appropriate window but is not the nearest detected event to the expert annotation. The sensitivity (Se) is defined as the ratio of the correctly detected events and all true events:

$$Se = \frac{TP}{TP+FN}. \tag{10}$$

Positive predictivity is defined as the ratio of correctly detected and all detected events:

$$PP = \frac{TP}{TP+FP}. \tag{11}$$

Due to differences in length of the test sequences, each of

the statistics defined in (10) and (11) was calculated for each sequence and averaged on all the test sequences giving each sequence an equal weight. The events S1 and S2 were marked manually on the test phonocardiograms using the aid of an ECG channel. A four-fold cross-validation was performed for model evaluation. The evaluation results are given in Table 1. The results on all recordings are given in histograms shown in Fig. 2.

Table 1.The algorithm results on the authors detected events.

	Detection misses	Se	PP
S1	0%	98.6%	96.9%
S2	0%	98.3%	96.5%

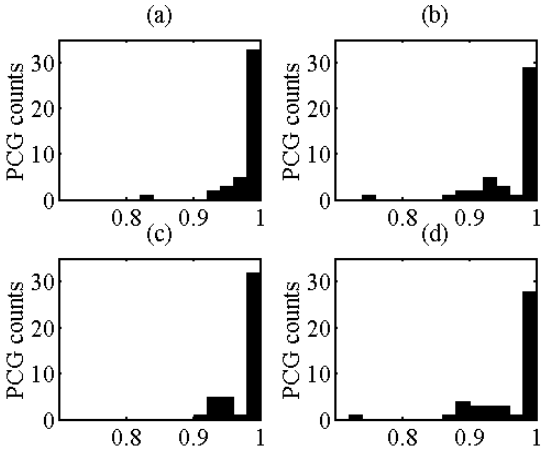

Figure 2. Histograms of the results on the entire data set: (a) Se of S1. (b) PP of S1. (c) Se of S2. (d) PP of S2.

4. Discussion and conclusions

The aim of this study was to demonstrate the usefulness of the homomorphic envelogram for robust automated detection and identification of the major components of phonocardiograms. It was shown that simple features extracted from the homomorphic envelogram and the raw signal are sufficient for consistent self-organization of the HMM, which in addition, provides probability measure for the states deciphering. The method is based upon very little prior knowledge and does not use pre-determined threshold heuristics or prior estimation of systolic and diastolic period durations. Moreover, the model parameter estimation was done in an unsupervised manner and based on bare phonocardiograms with no accompanying annotations. The training consisted of four stages each: envelogram extraction, detecting its peaks, features extraction from the peaks neighborhood and unsupervised training for probabilistic modeling. For segmentation, the features were extracted in the same way and then the most probable sequence of states was estimated. As shown in the results, the homomorphic filtering is sensitive enough to reveal low SNR events and still overcome the problem of detecting too many events caused by heart sounds splits, serrated peaks and noisy environment. The results show that this method competes successfully with other state of the art methods [2,4] and yet it is simple and unsupervised. Adding supervised learning and time-frequency representation features can strengthen the method and enlarge the variety of problems it can handle such as automatic detection of other heart sounds as S3, S4, murmurs, and valves defects. In addition, the method can be generalized to other biomedical signals.

Acknowledgements

The authors would like to thank Guy Amit from the School of Computer Science in Tel-Aviv University for his help with the data collection. Thanks to the Deutsch Institute for the financial support.

References

[1] Durand LG, Pibarot P. Digital signal processing of the phonocardiogram: review of the most recent advancement. Critical Reviews in Biomedical Engineering 1995;23(3/4):163-219.

[2] Huiying L, Sakari L, Liro H. A heart sound segmentation using wavelet decomposition and reconstruction. Proceedings of the 19th Annual International Conference of the IEEE-EMBS, Chicago 1997;1630-1633.

[3] Sarkady AA, Clarck RR, Williams R. Computer analysis techniques for phonocardiogram diagnosis. Comput Biomed Res. 1976 Aug;9(4):349-63.

[4] Gamero LG, Watrous R. Detection of the first and second heart sound using probabilistic Models. Proceedings of the 25th Annual International Conference of the IEEE-EMBS, Cancun, Mexico, 2003; 2877-2880.

[5] Oppenheim AV. Generalized linear filtering. In: Gold B, Rader CM, editors. Digital Processing of Signals, New-York, McGraw-Hill Book Company, 1969; ch. 8.

[6] Xu J, Durand LG, Pibarot P. Nonlinear transient chirp signal modeling of the aortic and pulmonary components of the second heart sound, IEEE Trans. Biomed. Eng. 2000; 47:1328–1335.

[7] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 1989; 77(2):257-286.

[8] Hughes NP, Tarrassenko L, Roberts SJ. Markov models for automated ECG interval analysis, Advances in Neural Information Processing Systems 16, 2003; Cambridge MA, MIT Press, 2004.

Address for correspondence

Daniel Gill
Department of Statistics
The Hebrew University
Mount Scopus, 91905 Jerusalem, Israel
gill@mta.ac.il