Support Vector Machine Hidden Semi-Markov Model-based Heart Sound Segmentation

David B Springer[1], Lionel Tarassenko[1], Gari D Clifford[1,2]

[1] University of Oxford, Oxford, UK
[2] Emory University, Atlanta, USA

Abstract

The segmentation of the primary heart sounds within a phonocardiogram (PCG) is an essential step in the classification of pathological cardiac events.

Recently, probabilistic models, such as hidden Markov models, have been shown to surpass the segmentation capabilities of previous methods. These models are further improved when a priori information about state duration is incorporated into the model, such as in a hidden semi-Markov model (HSMM).

This paper addresses the problem of the accurate segmentation of heart sounds within noisy, real-world PCGs using a HSMM, extended with the use of support vector machines (SVMs) for emission probability estimation.

A database of 123 patients with over 20,000 labelled heart sounds were used to train and test the algorithm. Best reported alternatives in the literature were also implemented and tested on the same data. On out-of-sample test data, our method outperforms previously reported methods with sensitivities of 94.9% and 91.0% and positive predictivities of 95.2% and 90.9% for first and second heart sounds respectively.

1. Introduction

The identification of the primary heart sounds in the phonocardiogram (PCG), or heart sound segmentation, is an essential step in the automatic analysis of heart sound signals. The primary heart sounds refer to the first and second heart sounds, S_1 and S_2. S_1 occurs immediately after the QRS complex of the ECG, while S_2 occurs at approximately the end of the T-wave [1], shown in Fig. 1. Accurate identification of these sounds allows subsequent classification of pathologies in the PCG.

Segmentation becomes a difficult task when PCGs are corrupted by in-band noise. Common noise sources include talking, motion artefacts, background noise from machinery and physiological sounds (such as clicks, intestinal activity, breathing, and additional cardiac activity such as heart murmurs, S_3 and S_4 sounds).

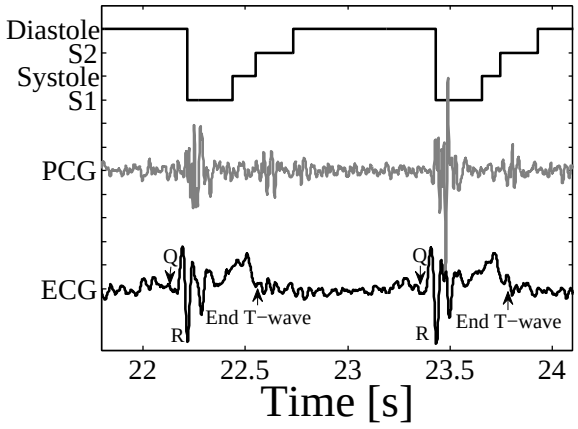

Figure 1. Example of an ECG-labelled PCG recording. The ECG, PCG and the four states of the heart cycle (S1, Systole, S2 and Diastole) are shown. The Q wave, R-peak and T-wave end point are also labelled in the ECG as references for defining the S1 and S2 onsets.

Approaches to heart sound segmentation have included envelope and energy threshold-based approaches [2, 3], neural networks [4] and most promisingly, hidden Markov models (HMMs) [5–7]. The work presented in this article builds on that of Schmidt *et al.* [7] by exploring a wider range of features and a machine learning approach to deriving the emission probabilities. We compare our results to those of Schmidt *et al.* [7] on a substantially larger database, collected in a real-world setting.

2. Methods

2.1. Dataset

The dataset consisted of 405, 20-30 s, synchronous PCG and ECG recordings from 123 adult patients attending the Massachusetts General Hospital for cardiac screening, 83 of whom were found to have murmurs [8]. A Welch-Allyn Meditron Elite electronic stethoscope was used.

In order to ensure accurate reference points for the S_1 and S_2 sounds in the PCG, poor quality synchronous ECG signals were excluded. The quality of the ECG signals was computed using the classifier derived by Behar *et al.* [9]. By computing the average probability of being low-quality, 108 ECGs were found to be below a probability of 0.5 and excluded. From the remaining recordings, 150 were randomly selected for a training set (10,807 heart sounds), leaving 147 recordings (10,975 sounds) for the test set. It was ensured no recordings from a single patient appeared in both training and test sets.

The reference positions of the S_1 and S_2 sounds were found in the synchronous ECG recordings.

The S_1 reference positions were identified as synchronous with the ECG R-peak, found using a standard open source peak detector applied to the ECG [10]. In addition, these positions were used to find the reference heart rate. The S_2 reference positions were identified as the point synchronous with the end of the ECG T-wave using the method of Vazquez-Seisdedos *et al.* [11].

The S_1 and S_2 durations were set to their mean, expected duration, derived in Section 2.2. The periods between the S_1 and S_2 sounds were labelled as systole, while those periods between S_2 and S_1 were labelled as diastole. An example of a labelled PCG recording is shown in Fig. 1.

2.2. Hidden Semi-Markov Models

HMMs are a statistical framework used to describe sequential data. They operate by making inferences about the likelihood of being in, transitioning between, and seeing observations in certain discrete hidden states. In this case, the HMM is first order, with the hidden sequence consisting of the four states of the heart, while the observations are features derived from the PCG.

An HMM can defined as a function of A, B and π, where A is the transmission matrix, governing the probability of transitioning between states, B is the emission or observation distribution, defining the probability of seeing an observation in each state, and π is the initial state probability distribution [12].

The utility of the HMM for heart sound segmentation is finding the most likely state sequence, given a HMM, $\lambda = (A, B, \pi)$, and an observation sequence, $\mathbf{O}$. This is derived using a dynamic programming method called the Viterbi algorithm [12].

A HMM does not incorporate information about the expected duration of each state, meaning that the state durations are governed only by the self-transition probabilities. This is poorly suited for PCG analysis as the heart sound duration is governed strongly by the cardiac dynamics. In order to improve the duration modelling, an extra parameter is introduced:

Let us define the new model as $\lambda = (A, B, \pi, p)$, where

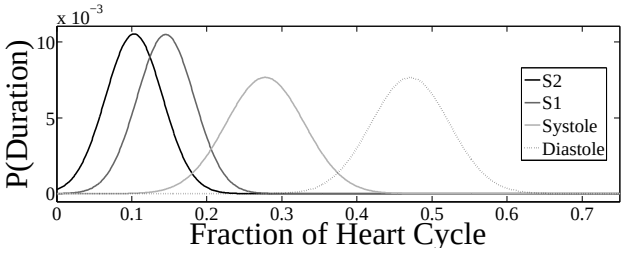

Figure 2. Timing distributions of the four major heart states shown as the fraction of the entire heart cycle for a HR of 60 beats per minute.

$p = \{p_i(d)\}$ is the explicitly defined probability of remaining in state i for duration d. This is then called a hidden semi-Markov model (HSMM) [13].

A key component of the HSMM for heart sounds is an estimate of the amount of time expected to remain in each state. In this case the four states are: 1) S_1 2) systole (between S_1 and S_2), 3) S_2 and 4) diastole (between S_2 and S_1). These durations were modelled as Gaussian distributions, following Schmidt *et al.* [7].

From the largest study of the duration of S_1 and S_2 [14], the duration distributions of these sounds, D_{S_1} and D_{S_2} in seconds, were defined as:

$$D_{S_1} \sim \mathcal{N}\left(0.146, (0.038)^2\right) \tag{1}$$

$$D_{S_2} \sim \mathcal{N}\left(0.104, (0.038)^2\right) \tag{2}$$

The QS_2 time, or electromechanical systole, is the duration from the Q-wave in the ECG to the start of the S_2 sound. The linear relationship between HR and the QS_2 time [15] can be used with knowledge of D_{S_1} and D_{S_2} to derive the systolic and diastolic durations of a heart cycle, given the HR. Therefore, the systolic and diastolic durations, D_{sys} and D_{dias}, can be found using:

$$D_{sys} = QS_2 - D_{S_1} \tag{3}$$

$$D_{dias} = D_{cycle} - QS_2 - D_{S_2} \tag{4}$$

where D_{cycle} is the duration of a HR-derived cardiac cycle. These duration distributions can be seen in Fig. 2.

2.3. Support Vector Machine HSMM

Earlier work on a HSMM-based approach to PCG segmentation used Gaussian distributions [7] to derive emission probabilities which do not allow accurate discrimination between states. Here we introduce SVM-derived observation probability estimates instead.

The SVM is a supervised binary classifier that attempts to find a separating hyperplane between two sets

of data [16]. A SVM has two parameters: γ, which defines the width of the Gaussian kernel and influences the flexibility of the classifier; and C, which controls the width of the margin each side of the separating hyperplane. This influences the classifier's misclassification tolerance. Cross-validation was used to set γ and C (see Section 2.5).

A one-vs-all approach was implemented using the LIBSVM library [17], training one SVM for the observations from each state in the model. The probability of a state j, ξ_j, given each observation at time t, $b_j(\xi_j|O_t)$ was derived using the method derived by Wu *et al.* [18], used in the LIBSVM library. Thereafter, the probability of an observation given a state, $b_j(O_t|\xi_j)$, as needed in the HSMM, was found using Bayes' rule:

$$b_j(O_t|\xi_j) = \frac{b_j(\xi_j|O_t) \times P(O_t)}{P(\xi_j)} \quad (5)$$

where $P(O_t)$ is found from a multivariate normal distribution derived from the training data, and $P(\xi_j)$ is found from π, the initial state probability distribution.

The input to the SVM were a combination of features, derived in Section 2.4.

2.4. Feature Extraction

Before feature extraction, all recordings were down-sampled to 1 kHz using a poly-phase anti-aliasing filter. The frequency content of the fundamental heart sounds is below 500 Hz [19] and hence the Nyquist-Shannon sampling criterion was satisfied.

Former HSMM-based heart sound segmentation methods used only the homomorphic envelogram [7]. This paper investigated three further features:

1. Amplitude of PCG signal envelope, extracted using the Hilbert transform [20], motivated by heart sounds generally having the highest energy in the PCG.
2. The absolute value of the detail coefficients from the discrete wavelet transform decomposition of the PCG, extracted using the Daubechies 10 wavelet at decomposition level 3. This wavelet was chosen due to its similarity in shape to the primary heart sounds and its centre frequency of 85.53 Hz, being within the expected range of frequencies within S_1 and S_2 (10 to 200 Hz) [19].
3. The maximum peak below 200 Hz in the power spectral density (PSD), found in 0.05 s windows of the PCG with 50% overlap after Hamming windowing. This is motivated by the majority of the frequency content of the primary heart sounds being below 200 Hz [19]18], while murmurs and other pathological sounds have higher frequency content that extends up to 600 Hz [21].

The feature vectors for each recording were individually normalised by subtracting their mean and dividing by their standard deviation.

Table 1. Segmentation performance (Se and P_+) for S_1 and S_2 for training (grey) and test (dark text). [7]* indicates performance of [7] using the same features as described in this article, while SVM-HSMM† indicates the performance of the SVM-HSMM using the feature from [7].

Algorithm	Se^{S_1}	$P_+^{S_1}$	Se^{S_2}	$P_+^{S_2}$
[7]	93.0	92.7	87.8	87.0
	92.6	92.3	89.0	88.3
[7]*	87.8	85.8	83.2	80.7
	88.6	86.0	85.8	83.1
SVM-HSMM†	92.2	92.9	85.7	85.7
	93.5	94.1	88.0	87.9
SVM-HSMM	94.8	95.4	88.6	88.5
	94.9	95.2	91.0	90.9

Following Schmidt *et al.* [7], the feature vectors were down-sampled further to 50 Hz poly-phase anti-aliasing filter, in order to increase the speed of computation.

2.5. Model Training & Evaluation

The parameters of the HSMM were computed using the ECG-labelled PCG sequences from the training set of 150 recordings.

In order to optimise the γ and C parameters when using the SVM-HSMM, the training set was further randomly split into three equal-sized, cross-validation sets. γ and C were set as 2^n, $n \in [-0.5, 0, 1, 2]$, with the parameters that led to the highest average binary classification accuracy of the four states over the three cross-validation folds being chosen as final values to be used in testing.

An S_1 sound was labelled correctly identified if the start of the segmented S_1 sound was found to be within 100 ms of the R-peak of the ECG. This tolerance is based on the recognised ECG R-peak detection tolerance [22]. An S_2 sound was labelled as correctly segmented if the centre of this S_2 sound was found to be within 100 ms of the end of the corresponding T-wave.

3. Results

The results of the SVM-HSMM algorithm on both the training and test sets, can be seen in Table 1. This table illustrates the sensitivity, (Se) and positive predictivity $(P+)$ for both the primary heart sounds, averaged across all the recordings in the training and test sets. The results of the state-of-the-art algorithm, developed by [7], can be seen in the first line of the table. The results of the SVM-HSMM algorithm, developed in this paper, can be seen in the last line of the table.

4. Conclusions

The SVM-based HSMM can be seen to outperform the former state-of-the-art algorithm in Table 1. The poor results of the [7]* algorithm in this table indicate that it is not the addition of the new features alone that led to improved performance.

The results from both the SVM-HSMM† and SVM-HSMM algorithms in Table 1 indicate that it is the addition of the SVM that led to improved segmentation results, while the combination of the SVM and the new features led to best performance. Therefore, the greater discrimination between states afforded by the SVM classifier's non-linear kernel function led to the improved segmentation results.

It can be concluded that the incorporation of the SVM into an HSMM, based on literature-derived heart sound durations and incorporating Hilbert envelope, wavelet envelope and PSD features, is able to accurately segment the fundamental heart sounds within a large dataset of noisy, real-world PCGs.

Acknowledgements

David Springer is funded by the Rhodes trust. Many thanks to Dr. Francesa Nesta, Prof. John Guttag, Dr. Zeeshan Syed and Dorothy Curtis for the use of the MIT auscultation data.

References

[1] Tilkian AG, Conover MB. Understanding heart sounds and murmurs: with an introduction to lung sounds. Saunders, 2001.

[2] Liang H, Lukkarinen S, Hartimo I. Heart sound segmentation algorithm based on heart sound envelogram. In Computers in Cardiology, volume 24. Lund, Sweden, 1997; 105–108.

[3] Chen T, Kuan K, Celi L, Clifford G. Intelligent heartsound diagnostics on a cell phone using a hands-free kit. In AAAI Spring Symposium on Artificial Intelligence for Development. Stanford University, 2010; 26–31.

[4] Oskiper T, Watrous R. Detection of the first heart sound using a time-delay neural network. In Computers in Cardiology. Memphis, Tennesee, August 2002; 537–540.

[5] Ricke A, Povinelli R, Johnson M. Automatic segmentation of heart sound signals using hidden Markov models. In Computers in Cardiology, volume 24. Lyon, France, 2005; 953–956.

[6] Gill D, Gavrieli N, Intrator N. Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model. In Computers in Cardiology. Lyon, France, 2005; 957–960.

[7] Schmidt SE, Holst-Hansen C, Graff C, Toft E, Struijk JJ. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. Physiological Measurement April 2010;31(4):513–29.

[8] Syed Z, Leeds D, Curtis D, Nesta F, Levine RA, Guttag J. A framework for the analysis of acoustical cardiac signals. IEEE Transactions on Biomedical Engineering 2007; 54(4):651–662.

[9] Behar J, Oster J, Li Q, Clifford GD. ECG signal quality during arrhythmia and its application to false alarm reduction. IEEE Transactions on Biomedical Engineering June 2013;60(6):1660–6.

[10] Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Transactions on Biomedical Engineering March 1985;32(3):230–6.

[11] Vázquez-Seisdedos CR, Neto JaE, Marañón Reyes EJ, Klautau A, Limão de Oliveira RC. New approach for T-wave end detection on electrocardiogram: performance in noisy conditions. Biomedical Engineering Online January 2011;10(1):77.

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

[13] Yu SZ. Hidden semi-Markov Models. Artificial Intelligence February 2010;174(2):215–243.

[14] Luisada AA, Mendoza F, Alimurung MM. The duration of normal heart sounds. British Heart Journal January 1949; 11(1):41–7.

[15] Weissler AM, Harris WS, Schoenfeld CD. Systolic Time Intervals in Heart Failure in Man. Circulation February 1968;37(2):149–159. ISSN 0009-7322.

[16] Cortes C, Vapnik V. Support-vector networks. Machine Learning September 1995;20(3):273–297.

[17] Chang C, Lin C. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2011;2(3):27:1—-27:27.

[18] Wu T, Lin C, Weng R. Probability estimates for multi-class classification by pairwise coupling. The Journal of Machine Learning Research 2004;5:975–1005.

[19] Abbas AK, Bassam R. Phonocardiography Signal Processing. Morgan & Claypool Publishers, 2009.

[20] Rezek I, Roberts S. Envelope Extraction via Complex Homomorphic Filtering. Technical Report TR-98-9. Technical report, Imperial College, London, 1998.

[21] Williams H, Dodge H. Analysis of heart sounds. Archives of Internal Medicine 1926;38(6):685–693.

[22] American National Standards Institute. Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, 2012.

Address for correspondence:

David Springer
University of Oxford, Wellington Square, Oxford OX1 2JD
david.springer@eng.ox.ac.uk