Feature-Based Cardiac Cycle Segmentation in Phonocardiogram Recordings

Jussi Taipalmaa[1], Morteza Zabihi[1], Serkan Kiranyaz[2], Moncef Gabbouj[1]

[1] Tampere University of Technology, Tampere, Finland
[2] Qatar University, Doha, Qatar

Abstract

Phonocardiogram (PCG) conveys crucial information for cardiac health evaluation in ambulatory care and is an essential diagnostic test for heart assessment. Thus, identification and positioning of the first and second heart sound within PCG is a vital step in automatic heart sound analysis.

This study proposes a solution for individual cardiac cycle segmentation of PCG recordings. It extracts a rich set of features that are used for the segmentation of each cardiac cycle in a PCG recording by localizing the PCG peaks, S1 and S2. To accomplish this objective, a rich set of 66 discriminative features are selected and extracted from each frame in a PCG recording and several classifiers are evaluated to find out the one that achieves the highest segmentation accuracy. Finally, a post-processing method is proposed to reduce the classification noise and hence improve the segmentation performance

Contrary to the earlier methods proposed in the literature, this method is evaluated on one of the largest datasets available consisting of 48 877s PCG recordings. The proposed method has achieved F_1-score of 93.45%, and Sensitivity and Specificity values of 94.23% and 98.16% respectively. Moreover, it has been tested on the Pascal benchmark dataset, and has achieved Sensitivity and Specificity values of 96.42% and 98.12%, respectively.

1. Introduction

According to World Health Organization (WHO) report in 2015, an estimated 17.7 million people died from cardiovascular diseases only in 2015, representing 37% of all premature deaths worldwide. The most common, and possibly the cheapest and the earliest clinical examination is heart auscultation, which can reveal several cardiac anomalies. However, there are several limitations and practical problems in the human auditory system when it comes to phonocardiogram (PCG) signal analysis, despite the cognitive skills and expertise of the medical examiner. This brought the need for an automated, cost-effective and robust anomaly detection method for PCG signals.

A typical cardiac cycle consists of two main components: S1 heart sound, occurring at the beginning of the systole, and the second heart sound (S2), which marks the beginning of diastole. The PCG segmentation for locating each cardiac cycle is crucial for an accurate anomaly detection. Although it has been shown in the recent studies that anomaly detection can be performed without segmentation (e.g. [1]), providing a robust segmentation method can lead to more medically interpretable results.

Some earlier PCG segmentation methods include threshold-based methods [2, 3] and envelope extraction based methods [4, 5]. More recently, methods such as a Hidden Markov Model (HMM) and a Hidden Semi-Markov Model (HSMM) have been used [6, 7]. Currently, the most advanced method is the Logistic Regression-HSMM-based algorithm [8]. These methods have achieved reasonable results in heart sound segmentation, but many of them have been evaluated only on small or hand-picked datasets.

In order to address the aforementioned limitations and drawbacks, in this study a novel PCG segmentation method is proposed by framing the PCG recordings into non-overlapping frames of 250 ms and classifying each frame into one of the three classes (S1, S2 , other). The main contributions are: 1) proposing a large set of discriminative features for PCG segmentation (Section 2.2). 2) introducing a systematic post-processing approach to decrease the amount of noise in the predicted labels (Section 2.4). The proposed method is evaluated on the largest publicly available dataset and the results are discussed and compared with the other state-of-the-art methods in this domain in Section 3.

2. Methods and material

2.1. Benchmark dataset

Within this study, we used the Physionet/Computing in Cardiology 2016 challenge dataset [9, 10]. In this study, the focus is to classify frames in normal heart sounds. All the PCG recordings containing S1 and S2 peaks closer to each other than our selected frame length (125 ms) were

left out of the dataset (1% of the recordings). This gives us a dataset of a total of 2200 different PCG recordings with 13.3 hours of recorded PCG data with a sampling frequency of 2000 Hz. In this dataset, the number of S1 and S2 heart sounds is 63 094 and 62 985 respectively.

2.2. Feature Engineering

In this study, first, the PCG recordings are windowed into non-overlapping frames of 250 samples (125 ms). Then, 66 features in the time, frequency, and time-frequency domains are extracted from each frame. The extracted features can be categorized into seven types:

• **Shannon energy:** The Shannon energy of each frame is calculated.

• **Linear Predictor Coefficients (LPC):** The coefficients of 8th order linear predictor (8 coefficients).

• **Polynomial fitting:** A 12th degree polynomial is fitted to each frame and coefficients are used as features.

• **Sub-band energy:** The frequency spectrum of S1 contains components within the interval of 10-50 Hz and 50-140 Hz, while for S2 the frequency bands 10-80 Hz, 80-200 Hz and 220-400 Hz are considered more informative [3]. Sub-band energies are extracted within an interval of 0-300 Hz using equally divided spectrum with a spacing of 50 Hz. This gives us a total of 6 sub-band energy features.

• **Mel-scale features:** Mel-spaced filter banks offer a simple way of extracting spectral characteristics from an acoustic signal. In this method, an equally spaced triangular filter banks are created across the spectrum. They are defined in Eq. (1),

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

where $Mel(f)$ is the frequency f in Mel-scale. Because the most of the energy of heart sounds is located at the lower frequencies, this gives us a good way to extract features [11]. A filter bank of 15 triangular filters was used in the range of 10-430 Hz. Each filter is multiplied with discrete Fourier transform of a heart sound frame and then summed. This gives us the total of 15 features from each heart sound frame.

• **Regression coefficients:** Regression coefficients represent changes in each feature over time [11]. Regression coefficients are calculated from the Mel-scale features for each frame with Eq. (2),

$$d_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2 \sum_{n=1}^{N} n^2} \qquad (2)$$

where c_t is the Mel-scale feature of the tth frame. We use $N = 2$ to refer for the previous and next 2 frames.

• **Wavelet energy:** Discrete wavelet transform (Daubechies 4) is applied to each PCG frame and the approximation coefficients of level 7 (a_5) and the detail coefficients of level

1 to 7 ($d_1, \ldots, d_7$) are as features as we calculate the energy of each level of approximation and detail coefficients. This gives us a total of 8 wavelet energy features.

2.3. Classification

The performances of different classifiers were evaluated in 10-fold cross-validation experimentation. The performance metrics used in the evaluation are average of sensitivity (Se), specificity (Sp) and accuracy (Acc) in all folds. The evaluation metrics are defined as:
$Se = TP/(TP + FN)$, $Sp = TN/(TN + FP)$ and $Acc = (TP + TN)/(TP + TN + FP + FN)$, where T/F denotes true/false and P/N denotes positive/negative.

In the first phase, the performance was evaluated by separating S2 peaks from the rest of the data. Tested methods included k-nearest neighbor classifier (k-NN), Linear Discriminant Analysis (LDA) and Neural networks (NN) with different architectures (best results with 2 hidden layers with 25 neurons on both hidden layers). The performance of different classifiers can be seen in Table 1.

Table 1. Comparison among different classifiers.

Classifier	avg Se	std Se	avg Sp	std Sp	avg Acc	std Acc
k-NN	0.8785	0.0062	0.8805	0.0088	0.8801	0.0088
LDA	0.8446	0.0038	0.7600	0.0025	0.7768	0.0017
NN	0.8068	0.0061	0.8109	0.0043	0.8100	0.0041

The best performance was found with k-NN classifier using Manhattan distance and value k = 16.

The training data was balanced in a way that there is an equal number of elements in each class using oversampling. The size of the class with a lower number of elements is increased by randomly duplicating elements until all classes are equal in size. [12]

2.4. Post-processing

The used classifier has a tendency to detect extra peaks, but the number of missed peaks is relatively meager. The first step in post-processing is to reject extra peaks. The second step is to try to recover the missed peaks. In practice, it was noticed that detecting missed peaks is much more difficult than rejecting extra peaks. This means that the performance of a classifier that has a low number of missed peaks can be improved significantly. The final part of post-processing is to determine if detected peaks are S1 peaks or S2 peaks. All of the post-processing steps are performed to one PCG recording at a time.

• **Rejecting extra peaks:** The set of rules to reject and merge extra peaks in [2] was used as a guideline:
1) Calculate the intervals between adjacent peaks. At this point, we ignore the peak labels S1 and S2 and just ob-

serve if a peak is detected or not. Then a low-level limit is calculated, defined in Eq. (3).

$$\text{Low-level limit} = \mu - c_1 \cdot \sigma \qquad (3)$$

Where μ and σ are the mean and the standard deviation of the calculated intervals. The constant c_1 is obtained in grid search. All the peaks with an interval shorter than the low-level limit are taken to further inspections and the peak with lower amplitude is rejected.

2) Merge peaks by finding a local maximum. This is performed by inspecting every detected peak and finding the greatest amplitude within the interval of 55 ms.

3) Find two S1 or S2 peaks in consecutive frames and remove the one with lower amplitude.

4) Calculate the average amplitude of S1 and S2 peaks. Set a Threshold and reject all the peaks below the corresponding threshold.

• **Recovering missed peaks:** Recovering the missed peaks is the opposite operation of rejecting extra peaks. First, the intervals of the peaks are calculated again. At this point, the most of the extra peaks are already rejected so we can focus on intervals between two peaks with the same label (S1-S1 or S2-S2). This time a high-level limit is calculated, defined in Eq. (4).

$$\text{High-level limit} = Q3 - c_2 \cdot IQR \qquad (4)$$

Where $Q3$ is the boundary of upper 75th percentile and IQR is the difference between the 75th and 25th percentiles of the intervals. Again, the constant c_2 is obtained in grid search. If the interval between two peaks is longer than the high-level limit, we can assume that there is a missed peak between the two detected peaks.

• **Determine peak labels by intervals between peaks:** The last step of post-processing is to determine the labels of the detected peaks. In this case, we are trying to find three consecutive peaks labeled the same (S1-S1-S1 or S2-S2-S2) and determine the labels correctly by using the fact that the distance S2-S1 is greater than the distance S1-S2.

3. Discussion and Results

3.1. Classification Performance Evaluation

F1-scores were used to evaluate the performance of segmentation. The evaluation method is obtained from Physionet/Computing in Cardiology Challenge 2017 [9].

Different frame lengths were tested and the results are presented in Table 2.

The classifier performance was evaluated with the leave-one-out cross-validation method. The post-processing methods were applied to each classified recording individually. The post-processing efficiently reduces the number of de-

tected extra peaks. Figure 1 illustrates the effects of post-processing to a single PCG recording.

Table 2. F-scores for different frame lengths before and after post-processing.

PP	FL	No.	$F_1(O)$	$F_1(S1)$	$F_1(S2)$	F_1
-	250	2200	0.8859	0.7850	0.7428	0.8046
x	**250**	**2200**	**0.9769**	**0.9340**	**0.8924**	**0.9345**
-	300	2160	0.8786	0.8122	0.7644	0.8184
x	300	2160	0.9663	0.9289	0.8790	0.9248
-	350	2082	0.8758	0.8219	0.7704	0.8227
x	350	2082	0.9455	0.9032	0.8322	0.8936

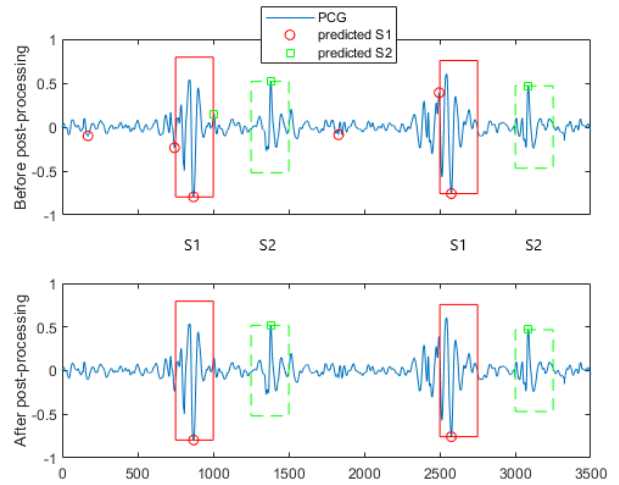PP = post-processing, FL = frame length, No. = number of signals

Figure 1. Detected peaks before and after post-processing.

The classified frames are presented in Table 3, where numbers in the confusion matrix denote to the number of frames predicted that belongs in certain reference class. In terms of detecting S1 or S2 peaks, the proposed method has achieved $Se = 94.23\%$ and $Sp = 98.16\%$.

Table 3. The classification results before and after post-processing.

		Predicted Classification					
		Before post-processing			After post-processing		
		Other	S1	S2	Other	S1	S2
	Other	209458	23025	24455	252215	1603	3120
Reference	S1	1749	59033	2312	1784	58451	2859
Classification	S2	4703	5258	53024	5406	2010	55569

The system performance was also evaluated with PASCAL dataset B [13]. The dataset consists of 90 normal PCG recordings. The sampling frequency of PASCAL dataset B is 4 000 Hz so the frame length used here is 500 samples. The confusion matrix obtained over the PASCAL dataset B is presented in Table 4. In this dataset, the proposed method achieved $Se = 0.9642$ and $Sp = 0.9812$.

Table 4. Peak detection on PASCAL dataset.

		Predicted Classification	
		S1 or S2	Other
Reference	S1 or S2	1266	38
Classification	Other	47	1988

3.2. Comparative Results

The current state-of-the-art method is Logistic Regression HSMM-based method [8]. In their paper they present F-scores: $F_1(S1) = 0.9695 \pm 0.0090$, $F_1(S2) = 0.9429 \pm 0.0108$ and $F_1 = 0.9563 \pm 0.0085$. This method was tested with a dataset that consists of 2.83 hours of PCG recordings. In the Table 5 is a summary of relevant methods previously tested. The table presents the dataset sizes and numerical results [8].

Table 5. Summary of different methods.

Auth.	Dataset description	Metrics and results
[2]	37 recordings (515 cycles) from children with murmurs (14 being pathological)	93.0% Acc
[3]	77 (1165 cycles) recordings from children with both pathological and physiological murmurs	94.6% Acc
[4]	Nine recordings (less than 5 s). 55% pathological	Cycles: 99.0% Acc
[5]	9426.8 s of recordings, normal (22.2%) and various pathologies	S1: 98.53% Acc S2: 98.31% Acc Cycles: 97.37% Acc
[6]	80 patients, 8 pathological. Recordings of 20 s from four auscultation sites (10045 S1, 9818 S2 sounds)	S1: 94.6% Se, 97.7% P_+ S2: 95.2% Se, 96.1% P_+
[7]	113 recordings of 8 s, 8% with coronary artery disease	98.8% Se, 98.6% P_+
[8]	112 patients, 10 172s of recordings.	95.34 $\pm$ 0.88% Se 95.92 $\pm$ 0.83% P_+ 92.52 $\pm$ 1.33% Acc

Acc denotes accuracy, Se sensitivity, Sp specificity and P_+ positive predictivity

4. Conclusions

This study proposes a solution for cardiac cycle segmentation of PCG recordings. It extracts a rich set of features that are used for the segmentation of each cardiac cycle in a PCG recording by localizing the peaks, S1 and S2. In the proposed approach a systematic post-processing is performed in order to reduce the classification noise in the predicted labels. According to the comparative evaluations performed over the other state-of-the-art methods, the proposed approach achieves competitive results using a simple and compact classifier ($F_1(S1) = 0.9340$, $F_1(S2) = 0.8924$, $F_1 = 0.9345$, $Se = 94.23\%$, $Sp = 98.16\%$). The system was evaluated with one of the largest datasets publicly available (PhysioNet/CinC Challenge 2016 dataset). This demonstrates the potential of this method and the results could be improved in the future with more sophisticated feature selection and classification methods.

References

[1] Zabihi M, Rad AB, Kiranyaz S, Gabbouj M, Katsaggelos AK. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. In Computing in Cardiology Conference. 2016; 613–616.

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

[3] Liang 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 Engineering in Medicine and Biology Society 1997;4:1630–1633.

[4] Yan Z, Jiang Z, Miyamoto A, Wei Y. The moment segmentation analysis of heart sound pattern. Computer Methods and Programs in Biomedicine 2010;98:140–150.

[5] Sun S, Jiang Z, Wang H, Fang Y. Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified hilbert transform. Computer Methods and Programs in Biomedicine 2014;114:219 – 230.

[6] 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 Engineering in Medicine and Biology Society 2003;3:2877–2880.

[7] Schmidt S, Holst-Hansen C, Graff C, Toft E, Struijk J. Segmentation of heart sound recordings by a duration-dependent hidden markov model. Physiological Measurement 2010;31:513–529.

[8] Springer DB, Tarassenko L, Clifford GD. Logistic regression-hsmm-based heart sound segmentation. IEEE Transactions on Biomedical Engineering 2016;63:822–832.

[9] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, et. al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101:e215–e220.

[10] Liu C, Springer D, Li Q, Moody B, et. al. An open access database for the evaluation of heart sound algorithms. Physiological Measurement 2016;37:2181–2213.

[11] Ricke AD, Povinelli RJ, Johnson MT. Automatic segmentation of heart sound signals using hidden markov models. Computers in Cardiology 2005;953–956.

[12] Rahman M, Davis DN. Addressing the class imbalance problem in medical datasets. International Journal of Machine Learning and Computing 2013;3:224–228.

[13] Bentley P, Nordehn G, Coimbra M, Mannor S. The PASCAL Classifying Heart Sounds Challenge 2011.

Address for correspondence:

Jussi Taipalmaa
P.O.Box 527, FI-33720 Tampere, FINLAND
jussi.taipalmaa@tut.fi