

Identification of Abnormal Heart Sounds

Sasan Yazdani^{*,1}, Silas Schlatter¹, Seyyed Abbas Atyabi², Jean-Marc Vesin¹

¹Applied Signal Processing Group, Swiss Federal Institute of Technology, Lausanne, Switzerland

²Cardiovascular Research Group, K. N. Toosi University of Technology, Terhan, Iran

Abstract

As a part of 2016 Physionet/CinC challenge, this work aims at the detection of abnormal phonocardiogram (PCG) recordings. Heart sound signal analysis has been an active research topic over the past decades with various studies such as heart sound segmentation and classification. We used the Physionet/CinC2016 challenge PCG database, which contains a large public collection of PCG recordings from a variety of clinical and nonclinical environments. The PCG classification in this work is performed in two steps. PCG heartbeats are first segmented and various heart sound markers are delineated. Then, a series of beat-specific features are extracted from the segmented heartbeats. Finally, PCG recordings are classified into normal and abnormal groups by performing classification based on tape-long features and by analyzing beat-extracted features from the PCG. Our method achieved an overall score of 80 in the unofficial phase of the challenge. In the official phase, the overall score of the proposed method was 82, with a sensitivity of 89%.

1. Introduction

Mechanical activity of the heart, which is triggered by the propagation of electrical activity within the heart, is audible at different locations on the chest wall. Phonocardiogram (PCG) is an audio recording of these mechanical activities recorded at the chest surface. Heart sounds provide valuable information which can help in the diagnosis of heart valve disorders. PCG comprises fundamental heart sounds from which $S1$ and $S2$ are the most prominent in normal recordings, respectively representing the beginning of ventricular contraction and the beginning of diastole [1]. In abnormal cases, various markers maybe present beside the fundamental heart sounds such as murmurs. Murmurs are noise-like high frequency sounds and long systolic murmurs as well as diastolic and continuous murmurs are generally pathologic [2].

Auscultation is a common cost-effective technique that provides valuable information about heart valve problems. However, by listening to mechanical activity of the heart

with a stethoscope, physicians cannot examine all physical characteristics of heart sounds, as it has been shown that only limited activity of heart sounds and murmurs are within human audibility range [1][3].

Over the years, various heart sound classification methods have been proposed. Generally, the analysis of heart sounds is performed on the heart cycle. To this end many PCG heart sound segmentation methods have been developed, enabling the detection of fundamental heart sound markers such as the beginning/end of $S1$, systole, $S2$ and diastole. These segmentation techniques use different approaches such as signal envelopes [4][5], frequency and amplitude features [6] as well as phase [7] and complexity features [8], hidden markov models (HMM) [9][10], and machine learning approaches [11][12].

With the segmented cardiac cycles, the classification of heart sound pathologies is made possible and several methods have been proposed over the last decades. Among these studies, artificial neural networks [13], support vector machines [14] and HMM based [15] approaches are common. Classification based on clustering has also been shown to be effective in heart sound pathology classification [16]. Typical features used in these studies comprise time-domain, frequency-domain, time-frequency and complexity-based features [13][14][16].

2. Methods

This study aims at the identification of abnormal PCG recorded in various conditions and locations, as a part of the Physionet/CinC 2016 Challenge. Our approach relies on heartbeat driven features that are reinforced with tape-long features for a robust classification of PCG. Block diagram of the proposed method is illustrated in Fig. 1.

2.1. Challenge Data

For this challenge, train/test datasets were created by collecting nine smaller datasets from several contributors around the world. Data were collected from different locations on the body and in clinical/nonclinical conditions. All recordings were resampled to 2000 Hz and standard-

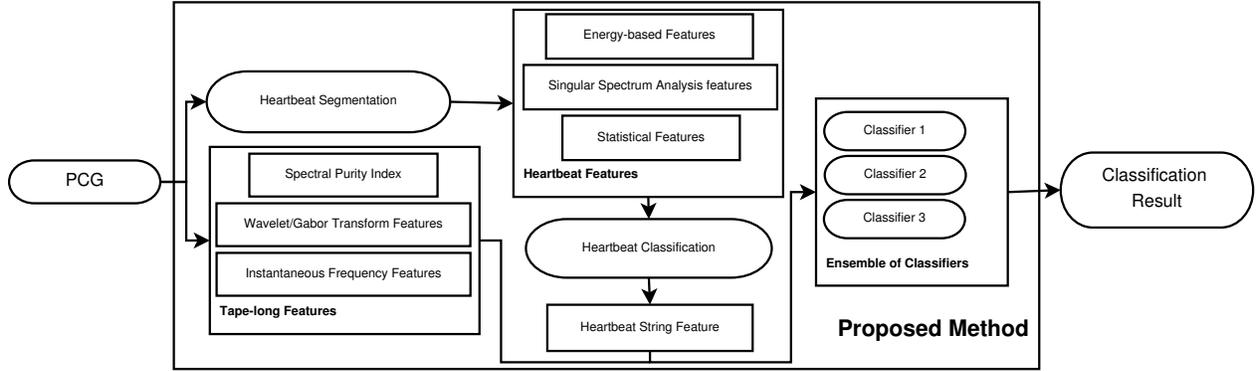


Figure 1. Block diagram of the proposed method.

ized for this challenge. Out of the nine different datasets in this challenge, a subset of six were used for training and six were used as hidden test set. Training and hidden test subsets contained two exclusive databases and four non-overlapping shared databases. For training set an overall 3239 recordings was considered from 764 subject with 79.5% and 20.5% of the dataset corresponding to respectively normal and abnormal instances. The hidden test set comprised of 1353 recording from 308 subjects with 11.3% abnormal, 72.8% normal and 15.9% unsure instances [2].

2.2. Beat to Beat Analysis

Heartbeat Segmentation. In order to analyze heartbeats and extract beat-driven features, the first step is to perform heartbeat segmentation on PCG recording in order to extract various heart sound markers. In order to alleviate this issue, we used the provided state-of-the-art PCG segmentation algorithm for this challenge [9]. This duration-dependent HMM segmentation algorithm is able to delineate the S1, S2, systole and diastole phases of PCG which are essential for the beat-to-beat analysis of the PCG.

Heartbeat Features. After PCG segmentation, each heartbeat is scrutinized for beat-driven features. First, time series representing each beat is non-parametrically decomposed into several sub-signals using singular spectrum analysis (SSA) [17]. Observations have shown that generally, 99% of the original signal can be recovered using the five largest components and therefore only these levels were further processed. A series of features were extracted directly from these components such as variances and central frequencies. Furthermore, the mean, energy, ratio between maxima in the systolic phase, six highest component maxima (heuristically chosen) inside the systolic or diastolic phase, the sparsity of compositions (calculated as number of samples higher than a threshold), the mean and variance of the instantaneous frequency of components.

2.3. Tape-long Features

The scheme described here uses benefits from features which describe the overall behavior of the PCG alongside beat-features. Although the state-of-the-art mostly focuses on beat-driven features, our work examined a wide range of tape-based features that could be suitable for abnormal PCG classification. Moreover, tape-long features are preferable to beat-driven features as generally their computation does not require the overhead of segmentation and beat-by-beat feature extraction.

Murmurs can be mostly considered as high frequency noise-like activity in PCG. Therefore, we studied how well the normal/abnormal PCGs are represented using a single frequency. In order to quantify this, we used the spectral purity index (SPI) [18]. SPI is defined based on running second- and fourth-order spectral moments and ranges between $[0, 1]$, with one representing a pure sinusoid. Observations of SPI over the PCG signals in the training set led to the conclusion that there is a noticeable difference between normal and abnormal PCGs in terms of mean and probability distribution function of SPI. Therefore, the mean SPI and the normalized number of samples within 10% of mean SPI were used as features. Fig. 2 shows the feature space created from SPI features.

Wavelet based features are commonly used for segmentation and PCG classification. In the proposed scheme, PCGs were analyzed using a 64-level daubechies WT and the total energy of each decomposition were used as features.

Instantaneous frequency (IF) is another suitable measure to describe the change of frequency in time. Since different murmurs contain specific frequencies, especially in the systolic and diastolic phases, the instantaneous frequencies of normal and abnormal tapes are separable in these phases. First, the instantaneous frequency is calculated the Hilbert transform of the PCG and then the mean and variance in the systolic phase and the diastolic phase are extracted as features.

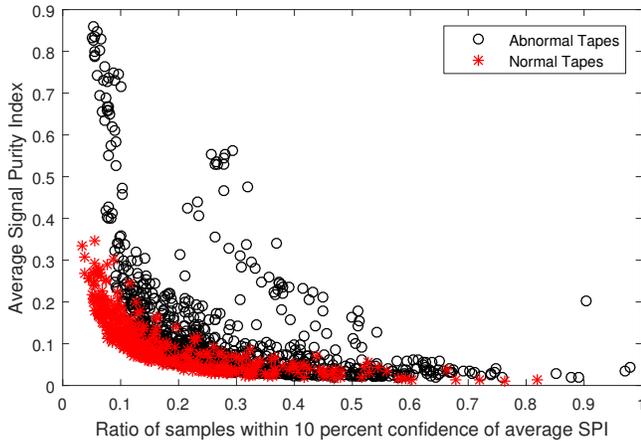


Figure 2. Signal purity index (SPI) feature space.

2.4. Beat Classification and Beat-to-Tape Feature Linking

After extraction of the beat-driven features, a multivariate analysis was performed in order to create a normal/abnormal heartbeat classifier. Since the number of normal instances were vastly higher than the normal instances, the multivariate analysis was performed using an abnormal-misclassification cost three times greater than that of normal beat misclassification. Using the bagging meta-algorithm with decision trees as base classifier and a 10-fold cross-validation, an ensemble of classifiers was created.

As PCG recordings in the training database vary in duration, i.e. from a couple of second up to more than 100 seconds, it was necessary to further process the output of the heartbeat classifier in order to use this information alongside the tape-long features. Therefore, a series of features were extracted from each PCG heartbeat string. a Heartbeat string is defined as a binary array containing the output of the heartbeat classifier for all heartbeats in a recording. From the heartbeat string, the ratio of abnormal beats, sum, and the standard deviation of the heartbeat string were extracted as features and added to the tape-long features from Sec. 2.3.

2.5. Tape Classifier

The final feature set is composed of the tape-long features and heartbeat string features. Using this feature set, a multivariate analysis was performed on the training PCGs. As the number of normal and abnormal tapes in the training set is unbalanced, three sub-training sets were created with all abnormal instances and the same number of normal instances randomly chosen without replacement. For each sub-training set a bagging ensemble was created based on a 10-fold cross-validation. The final decision is

obtained through a pessimistic voting the output of these classifiers. in which the tape is considered abnormal if any classifier declare the instance as an abnormal case, as illustrated in Fig. 1.

3. Results

The Physionet/CinC2016 challenge was divided into an unofficial and an official phase. For the unofficial phase the training set comprised recordings from one exclusive dataset and the shared training/test datasets. The hidden test set contained only the recordings from shared datasets. At the beginning of the unofficial phase the second exclusive dataset was added to the training set and two exclusive datasets were added to the hidden test set.

The score in this challenge was computed based on the sensitivity (Se) and the specificity (Sp) of the proposed schemes. In the unofficial phase, Se and Sp were calculated through Eq. 1 using values reported in Table 1. The overall score considered as $(Sp + Se)/2$.

Table 1. Sensitivity and Specificity Parameters.

Reference Label	Scheme Output			
	Abnormal	Abnormal	Unsure	Normal
		Normal	Na	Nq

$$Se = \frac{Aa + w \times Aq}{Aa + Aq + An} \quad \text{and} \quad SP = \frac{Nn + w \times Nq}{Nn + Nq + Na} \quad (1)$$

where ($w = 0.5$) represents a weight set to penalize incorrect 'unsure' labels less than PCG misclassifications. About two weeks before the final deadline, the scoring function was changed for the official phase as described in [2].

Table 2 shows the results obtained for the unofficial and official phases of this challenge. Although the proposed scheme does not reject low-quality signals, i.e. PCGs are only classified into abnormal and normal groups, the obtained score for the unofficial phase was 80. It is noteworthy that in the unofficial phase of the challenge, the proposed method was solely based on the tape-long features. With the start of the official phase, the best entry in the unofficial phase was resubmitted and an overall score of 76 was achieved. Then, beat-driven features were added to scheme which resulted in a score of 82, that is an increase of 6 percent to the overall score. However, after introducing the new scoring function a final score of 79.9 was obtained. Final performance on the training set brought a sensitivity of 80.34% and a specificity of 86.33% which led to a score of 83.34. It is noteworthy that the organizer of the challenge also proposed a simple benchmark classifier [2]. This logistic regression classifier obtained a sensitivity and a specificity of respectively 62% and 70%, with an overall score of 66 on the training data.

Table 2. Performance details of the proposed scheme.

	Unofficial Phase	Official Phase	
Sensitivity	0.70	0.76	0.746 *
Specificity	0.90	0.89	0.850 *
Overall Score	80	82	75.1 *

* based on the new scoring function

4. Discussion and Conclusion

In this scheme, several approaches were used to improve the overall PCG classification robustness. From the basic statistical features such as mean and variance of samples in a heartbeat to more elaborate features such as spectral purity index, time-frequency analysis and wavelet transform analysis, we aimed at extracting a set of features that could be descriptive enough to discriminate abnormal heart sound recordings from the normal ones. It is noteworthy that several approaches, such as denoising methods, interbeat interval analysis were performed during this challenge. Although these features had some discriminative capabilities, they were not as powerful as other features described throughout this paper and therefore were discarded in the feature selection analysis.

While the state-of-the-art is mostly focused on beat-driven features, results show that tape-long features reinforce the overall classification performance. The multivariate analysis on both beat-driven and tape-long feature levels can be challenging as the datasets provided for this challenge are not balanced. In this challenge we tried different approaches such as defining specific cost functions, creating balanced training subsets and a hybrid of both. Several classification methods such as decision trees, support vector machines and k-nearest neighbors were considered and at each and in the end the classifier with most balanced outputs was selected.

Acknowledgements

This study was performed in the framework of the Nano-Tera initiative supported by the Swiss National Science Foundation (SNSF).

References

[1] Leatham A. Auscultation of the heart and phonocardiography: Churchill livingstone 1975;.

[2] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AE, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement* 2016;37(9).

[3] Mangione S, Nieman LZ. Cardiac auscultatory skills of in-

ternal medicine and family practice trainees: a comparison of diagnostic proficiency. *JAMA* 1997;278(9):717–722.

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

[5] Ari S, Kumar P, Saha G. A robust heart sound segmentation algorithm for commonly occurring heart valve diseases. *Journal of medical engineering technology* 2008; 32(6):456–465.

[6] Naseri H, Homaeinezhad M. Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric. *Ann Biomed Eng* 2013; 41(2):279–292.

[7] Varghees VN, Ramachandran K. A novel heart sound activity detection framework for automated heart sound analysis. *Biomed Signal Process Control* 2014;13:174–188.

[8] Nigam V, Priemer R. Accessing heart dynamics to estimate durations of heart sounds. *Physiol Meas* 2005;26(6):1005.

[9] Schmidt S, Holst-Hansen C, Graff C, Toft E, Struijk JJ. Segmentation of heart sound recordings by a duration-dependent hidden markov model. *Physiol Meas* 2010; 31(4):513.

[10] Springer DB, Tarassenko L, Clifford GD. Logistic regression-hmm-based heart sound segmentation. *IEEE Trans Biomed Eng* 2016;63(4):822–832.

[11] Gupta CN, Palaniappan R, Swaminathan S, Krishnan SM. Neural network classification of homomorphic segmented heart sounds. *Appl Soft Comput* 2007;7(1):286–297.

[12] Tang H, Li T, Qiu T, Park Y. Segmentation of heart sounds based on dynamic clustering. *Biomed Signal Process Control* 2012;7(5):509–516.

[13] Akay YM, Akay M, Welkowitz W, Kostis J. Noninvasive detection of coronary artery disease. *IEEE Eng Med Biol Mag* 1994;13(5):761–764.

[14] Ari S, Hembram K, Saha G. Detection of cardiac abnormality from pcg signal using lms based least square svm classifier. *Expert Syst Appl* 2010;37(12):8019–8026.

[15] Wang P, Lim CS, Chauhan S, Foo JYA, Anantharaman V. Phonocardiographic signal analysis method using a modified hidden markov model. *Ann Biomed Eng* 2007; 35(3):367–374.

[16] Avendano-Valencia L, Godino-Llorente J, Blanco-Velasco M, Castellanos-Dominguez G. Feature extraction from parametric time–frequency representations for heart murmur detection. *Ann Biomed Eng* 2010;38(8):2716–2732.

[17] Vautard R, Yiou P, Ghil M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D Nonlinear Phenomena* 1992;58(1):95–126.

[18] Goncharova II, Barlow JS. Changes in eeg mean frequency and spectral purity during spontaneous alpha blocking. *Electroencephalogr Clin Neurophysiol* 1990;76(3):197–204.

Address for correspondence:

Sasan Yazdani
 EPFL SCI STI JMV - ELD 224 - Station 11
 CH-1015 Lausanne - Switzerland.
 E-mail address: sasan.yazdani@epfl.ch