

# A Data-Driven Feature Extraction Method for Enhanced Phonocardiogram Segmentation

Francesco Renna<sup>1</sup>, Jorge Oliveira<sup>2</sup>, Miguel T. Coimbra<sup>2</sup>

<sup>1</sup> University of Cambridge, UK

<sup>2</sup> Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal

## Abstract

*In this work, we present a method to extract features from heart sound signals in order to enhance segmentation performance. The approach is data-driven, since the way features are extracted from the recorded signals is adapted to the data itself. The proposed method is based on the extraction of delay vectors, which are modeled with Gaussian mixture model priors, and an information-theoretic dimensionality reduction step which aims to maximize discrimination between delay vectors in different segments of the heart sound signal.*

*We test our approach with heart sounds from the publicly available PhysioNet dataset showing an average  $F_1$  score of 92.6% in detecting  $S_1$  and  $S_2$  sounds.*

## 1. Introduction

Auscultation is one of the fundamental steps of the physical examination of a patient and the first line of screening of cardiovascular disease. Based on the clinical signs extracted from it, a clinician decides if additional exams are needed, typically an echocardiogram, which is more reliable, but also more expensive and requiring specific expertise to apply. The cost-effectiveness of auscultation contrasts with the worrying revelations of important studies quantifying that only around 20% of medical interns can actually perform cardiac auscultation effectively [1]. This observation fueled recent research efforts in automatizing part or the entire process of analysis of the phonocardiogram (PCG) signal.

A key step usually implemented to extract information from a heart sound signal consists in dividing each heart cycle into its fundamental components. Namely, a normal heart cycle is usually divided into first heart sound (denoted by  $S_1$ ), systole, second heart sound (denoted by  $S_2$ ) and diastole. Further sound components of interest are represented by murmurs, clicks, splitting of the first and/or second heart sounds and further third ( $S_3$ ) and fourth ( $S_4$ ) heart sounds.

Different approaches have been presented in the literature in order to perform PCG segmentation (see [2] for a general overview). In general, such methods are performed via a two-steps approach, that first extract features from the PCG signal and then assigns features to the different states corresponding to the different sound segments or components. In particular, features can be extracted in the time domain (e.g., Shannon energy [3]), in the frequency domain (e.g., Mel-frequency cepstral coefficients [4]) or other transform domains (wavelet transform, etc.).

Then, features corresponding to different time instants can be assigned to the different segments of the signal using various kinds of classifiers, as, for example, support vector machines (SVM) [5], artificial neural networks (ANN) [6], and, more recently, deep neural networks [7].

Another popular approach is based on the explicit modeling of the sequential nature of states in the heart sound signal. In particular, algorithms based on hidden Markov models (HMM) and hidden semi-Markov models (HSMM) are designed in order to segment heart sound signals. The use of HMM for heart sounds dates back to the work of Gill *et al.* [8]. On the other hand, HSMMs are a generalization of HMM which allow to model explicitly sojourn time distributions, i.e., the distribution of the time spent by the signal in a given state of the corresponding hidden Markov chain. HSMMs have been recently coupled with the use of logistic regression distribution to model emission probability distributions, thus obtaining state-of-the-art segmentation performance [9].

However, PCG segmentation still represents a challenging problem to solve when considering its application in real-world, noisy environments. In this work, we propose a technique to extract features from the heart sound signal, which is adapted to the data via training, and which aims to increase separability between segments corresponding to different signal states. The method provides a principled framework based on dynamical system analysis (delay vectors method [10]) and information-theoretic tools to enhance segmentation performance. In particular, the proposed algorithm is based on the following steps:

1. extract *delay vectors* from a heart sound signal;

2. model the delay vectors with state-dependent, multi-variate Gaussian mixture model (GMM) priors;
3. reduce the delay vector state space by using information-theoretic linear discriminant methods.

The proposed feature extraction technique can be coupled with different kinds of classifiers. In particular, it can be applied to HMMs or HSMMs, due to the amenability in computing the statistical quantities required for optimal estimation of the hidden state sequence.

## 2. Methods

### 2.1. Delay vectors

We denote with  $x(t), t = 0, 1, \dots, T$  the samples of the observed heart sound signal, and with  $s(t), t = 0, 1, \dots, T$  the corresponding labels, i.e.,  $s(t) \in \{1, \dots, I\}$ , where the values  $1, \dots, I$  correspond to different states of the PCG signal. In particular, the PCG states considered in this work are:  $S_1$ , systole,  $S_2$  and diastole.

In dynamical system analysis, delay vectors are used to study the evolution of nonlinear systems, and they are shown to contain the full information about the system dynamics under mild assumptions [10]. Delay vectors are extracted from the observed signal as follows: given a time lag  $\tau \in \mathbb{N}$  and a vector dimension  $n \in \mathbb{N}$ , the delay vector associated to time  $t$  is defined as  $\mathbf{x}(t) = [x(t), x(t-\tau), \dots, x(t-(n-1)\tau)]^T$ , where  $(\cdot)^T$  denotes the transpose operator.

Then, delay vectors corresponding to a given state  $i$  are modeled via a multivariate GMM, i.e., we assume that the probability density function (pdf) of the vector  $\mathbf{x}(t)$  conditioned on  $s(t) = i$  is given by

$$b_i(\mathbf{x}(t)) = p(\mathbf{x}(t)|s(t) = i) = \sum_{k=1}^{K^{(i)}} p_k^{(i)} \mathcal{N}(\boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)}), \quad (1)$$

where  $K^{(i)}$  is the number of Gaussian components in the GMM emission distribution associated to state  $i$  and  $p_k^{(i)}$  is the probability of the  $k$ -th component in state  $i$ , so that  $0 \leq p_k^{(i)} \leq 1$  and  $\sum_k p_k^{(i)} = 1$ . Moreover,  $\mathcal{N}(\boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)})$  denotes a Gaussian multivariate distribution with mean  $\boldsymbol{\mu}_k^{(i)}$  and covariance matrix  $\boldsymbol{\Sigma}_k^{(i)}$ .

Note that the statistical description of a delay vector  $\mathbf{x}(t)$  conditioned on state  $i$  does not depend on the time instant  $t$ . For this reason, we can drop the time index and simply write  $b_i(\mathbf{x}) = p(\mathbf{x}|s = i)$ . Moreover, the marginal distribution of a delay vector  $\mathbf{x}$  is obtained by averaging the conditional distributions as

$$p(\mathbf{x}) = \sum_{i=1}^I \pi_i \sum_{k=1}^{K^{(i)}} p_k^{(i)} \mathcal{N}(\boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)}), \quad (2)$$

where  $\pi_i$  denotes the prior probability of being in state  $i$ .

Finally, we assume that the parameters  $\pi_i, p_k^{(i)}, \boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)}$ ,  $i = 1, \dots, I$ , that define the statistical description of the delay vectors, are estimated from annotated data in the training phase. Then, the distributions  $p(\mathbf{x})$  and  $p(\mathbf{x}|s = i)$  are used in the next step of the proposed feature extraction process, that is, dimensionality reduction.

### 2.2. Information-theoretic dimensionality reduction

Dimensionality reduction consists in finding a linear transformation which aims to: i) reduce the dimension of the delay vectors in order to decrease computational complexity and ii) enhance discrimination between delay vectors belonging to different states.

In particular, when  $m \in \mathbb{N}$  features are extracted from each delay vector, on denoting by  $\Phi \in \mathbb{R}^{m \times n}$  the linear transformation applied to the delay vectors, we quantify the level of discrimination between transformed delay vectors in different states by using Shannon's mutual information (MI) [11], i.e, we write

$$\mathbb{I}(\Phi \mathbf{x}; s) = h(\Phi \mathbf{x}) - \sum_{i=1}^I \pi_i h(\Phi \mathbf{x}|s = i), \quad (3)$$

where  $h(\mathbf{x}) = -\int p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}$  is Shannon's differential entropy. Then, the desired  $\Phi$  is obtained as the solution of the optimization problem

$$\underset{\Phi}{\text{maximize}} \mathbb{I}(\Phi \mathbf{x}; s), \quad \text{subject to } \text{tr}(\Phi \Phi^T) \leq m, \quad (4)$$

where  $\text{tr}(\cdot)$  is the trace operator and the constraint in (4) is imposed in order to guarantee stability with respect to measurement noise. The solution of (4) can be approximated by using a gradient descent algorithm that leverages the expression of the gradient of the mutual information  $\mathbb{I}(\Phi \mathbf{x}; s)$  with respect to  $\Phi$ . In particular, a method to efficiently compute such gradient is reported in [11] and it is shown to require the computation of a minimum mean-squared error (MMSE) matrix via Monte Carlo integration. However, we note that, since mutual information is not a concave function with respect to  $\Phi$ , gradient descent is not guaranteed to converge to the global optimum. Moreover, Monte Carlo integration can lead to numerical instability for high-dimensional delay vectors  $\mathbf{x}(t)$ .

For this reason, we propose also the use of an alternative method which is based on the following approximation of Shannon's mutual information:

$$\mathbb{I}_G(\Phi \mathbf{x}; s) = h_G(\Phi \mathbf{x}) - \sum_{i=1}^I \pi_i h_G(\Phi \mathbf{x}|s = i), \quad (5)$$

where  $h_G(\mathbf{x})$  is the differential entropy of a Gaussian vector having the same mean and covariance matrix as the

vector  $\mathbf{x}$ . Such approximation goes under the name of information discriminant analysis (IDA) [12]. It is also possible to show that  $\mathbb{I}_G(\Phi; s)$  is a concave function of  $\Phi$  and its gradient with respect to  $\Phi$  can be expressed in closed form. In this way, convergence to a global optimum can be guaranteed when using gradient descent [12].

### 2.3. Proposed features with HSMMs

The proposed method for feature extraction can be easily incorporated in algorithms for heart sound segmentation which are based on HSMMs. In particular, we recall that a HSMM is completely defined by a set of initial state probabilities, sojourn time distributions and emission probability distributions. Then, on denoting by  $\mathbf{y}(t) = \Phi \mathbf{x}(t)$ ,  $t = 0, 1, \dots, T$  the sequence of observed, transformed delay vectors, these represent the observed emissions in the HSMM. Moreover, the corresponding emission pdfs can be easily computed from (1) as

$$b_i(\mathbf{y}) = p(\mathbf{y}|s = i) = \sum_{k=1}^{K^{(i)}} p_k^{(i)} \mathcal{N}(\Phi \boldsymbol{\mu}_k^{(i)}, \Phi \boldsymbol{\Sigma}_k^{(i)} \Phi^T). \quad (6)$$

Finally, the state sequence that maximizes the log-likelihood associated with the sequence of observed feature vectors  $\mathbf{y}(t)$  can be efficiently computed via the use of the Viterbi algorithm [13].

## 3. Experimental methodology

In this section, we present an application example of the proposed feature extraction framework for PCG segmentation. In particular, we consider 792 heart sounds recordings of 135 patients from the publicly available PhysioNet dataset.<sup>1</sup> Among those, 406 sounds are collected from patients with cardiac pathologies (most commonly mitral valve prolapse), as assessed by echocardiography. The remaining 386 sounds are collected from healthy patients. Sound recordings have variable durations in the range from 1s to 35.5s. They are collected from several spots over the chest and they are corrupted by different sources of noise. The annotations provided with the dataset are computed via the analysis of synchronous electrocardiogram (ECG) recordings, based on the agreement between five different automatic R-peak and end-T-wave detectors.

Experiments are conducted as follows: the considered 792 heart sound signals are splitted randomly so that half of the sounds are used for training and half of the sounds are used for testing. Random splits are forced to ensure that sounds belonging to the same patient cannot appear in both the training and testing sets.

Although the considered PCG signals are collected at 1 kHz sampling rate, in order to speed up computation, we

further downsample the signals at 50 Hz, as done in [9]. More precisely, in our experiment, the samples  $x(t)$  are collected from the homomorphic envelope of the PCG signals sampled at 50 Hz. Then, delay vectors are extracted from both training and testing data by choosing  $\tau = 2$  and  $n = 32$ , so that each delay vector spans approximately 1.24s of the recorded PCG signal. In this sense, we have observed experimentally that the best segmentation performance is achieved when allowing the delay vectors to span a large portion of the PCG signal, approximately an entire heart beat. In fact, delay vectors containing larger portions of the PCG signal seem to capture more information about the dynamic evolution of the signal, thus providing more robust segmentation results.

Annotated training data are used to determine the delay vector distributions  $b_i(\mathbf{x})$ . In particular, we choose the number of Gaussian components in each state-dependent GMM as  $K^{(i)} = 10$ ,  $i = 1, \dots, I$  and the parameters  $p_k^{(i)}$ ,  $\boldsymbol{\mu}_k^{(i)}$  and  $\boldsymbol{\Sigma}_k^{(i)}$  are estimated for each state  $i$  by using the expectation maximization (EM) algorithm [13].

Then, we fix  $m = 7$  and we compute the matrix  $\Phi \in \mathbb{R}^{7 \times 32}$  by following the steps described in Section 2.2, that is by maximization of Shannon’s mutual information or by using the IDA approach. Such  $\Phi$  is applied to the delay vectors extracted from the sounds in the testing set, and the emission probability distributions are obtained as in (6).

Sojourn time distributions are assumed to be Gaussian, with means and covariances that depend on the heart rate and that are estimated via the method used in [9].

Performance of the proposed segmentation method is measured in terms of the following metrics. First, we consider the positive predictive value per sample ( $P_s$ ), which accounts for the general capacity of the algorithm in discriminating sounds belonging to  $S_1$ , systole,  $S_2$  and diastole intervals. We also report further four standard performance metrics concerning the detection of the principal heart sounds  $S_1$  and  $S_2$ : precision ( $P$ ), recall ( $R$ ), accuracy ( $A$ ) and  $F_1$  score. In this case, true/false positives and true/false negatives are determined based on the comparison of the centers of the  $S_1$  and  $S_2$  sounds in the state sequence returned by the segmentation algorithm and the state sequence associated to the ground truth annotation. For example, a true positive happens when, given the center of an  $S_1$  ( $S_2$ ) sound in the output sequence, the closest sound center in the annotation state sequence is also associated to an  $S_1$  ( $S_2$ ) sound.

### 3.1. Results

The aforementioned performance metrics are computed on a patient basis, and then averaged over the testing set. Then, the means of these values over 100 different random training/testing splits are reported in Table 1. In particular, we report the results obtained when using delay vec-

<sup>1</sup><https://physionet.org/physiotools/hss/>.

Table 1. Performance obtained with the proposed feature extraction framework without dimensionality reduction (NDR), with MI maximization and with IDA, compared with the corresponding results obtained by the algorithm in [9]. Mean and standard deviation (%).

	NDR	MI	IDA	[9]
$P_s$	$80.6 \pm 1.7$	$83.7 \pm 1.7$	$85.5 \pm 1.3$	$86.6 \pm 1.2$
$P$	$89.6 \pm 2.0$	$91.5 \pm 1.9$	$92.7 \pm 1.5$	$93.9 \pm 1.3$
$R$	$89.6 \pm 2.0$	$91.7 \pm 1.8$	$92.8 \pm 1.4$	$94.4 \pm 1.2$
$A$	$90.9 \pm 1.6$	$92.6 \pm 1.5$	$93.7 \pm 1.1$	$94.6 \pm 1.1$
$F_1$	$89.4 \pm 2.0$	$91.4 \pm 1.9$	$92.6 \pm 1.5$	$94.1 \pm 1.2$

tors without linear dimensionality reduction, when maximizing Shannon’s MI with the method presented in [11], and the results obtained with the IDA approach [12]. For comparison, we report also the corresponding values of the performance metrics obtained with the state-of-the-art algorithm presented in [9]. Such algorithm couples HSMMs with the use of logistic regression to model emissions associated to four different features extracted from the PCG signal (homomorphic envelope, Hilbert envelope, power spectral density and discrete wavelet transform).

We note that MI maximization and the IDA approach play an important role in enhancing discrimination between delay vectors of different signal states, thus leading to better segmentation performance. Moreover, the IDA approximation is shown to guarantee more reliable segmentation, possibly due to its numerical stability when compared with MI maximization. Finally, we observe that the proposed framework can guarantee competitive results compared with current state-of-the-art approaches, as the one proposed in [9].

## 4. Conclusion

In this work, we have proposed a novel method to extract features from the PCG signal to enhance segmentation performance. The method, which is based on the extraction of delay vectors and on information-theoretic linear dimensionality reduction, is shown to produce promising results.

The proposed framework, which combines GMM distributions to model emission vectors and mutual information maximization, opens the way also for further advances in PCG segmentation. In particular, future work along this line of research includes the addition of side information signals (e.g., ECG) in the model in order to improve segmentation performance, as well as the study of more general nonlinear dimensionality reduction methods.

## Acknowledgements

This work is a result and funded by the project NanoS-TIMA (NORTE-01-0145-FEDER-000016), supported by

Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). It is also a result of the internal project Smart-Heart in scope of project UID/EEA/50008/2013. The work of F. Renna was funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 655282.

## References

- [1] Mangione S. Cardiac auscultatory skills of physicians-in-training: a comparison of three english-speaking countries. *The American journal of medicine* 2001;110(3):210–216.
- [2] Liu C, et al. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement* 2016; 37(12):2181–2213.
- [3] Liang H, Lukkarinen S, Hartimo I. Heart sound segmentation algorithm based on heart sound envelopogram. In *Computers in Cardiology*. 1997; 105–108.
- [4] Delgado-Trejos E, Quiceno-Manrique A, Godino-Llorente J, Blanco-Velasco M, Castellanos-Dominguez G. Digital auscultation analysis for heart murmur detection. *Annals Biomed Eng* 2009;37(2):337–353.
- [5] Vepa J. Classification of heart murmurs using cepstral features and support vector machines. In *IEEE EMBC*. 2009; 2539–2542.
- [6] Leung T, White P, Collis W, Brown E, Salmon A. Classification of heart sounds using time-frequency method and artificial neural networks. In *IEEE EMBC*. 2000; 988–991.
- [7] Chen TE, Yang SI, Ho LT, Tsai KH, Chen YH, Chang YF, Lai YH, Wang SS, Tsao Y, Wu CC. S1 and S2 heart sound recognition using deep neural networks. *IEEE Trans Biomed Eng* 2017;64(2):372–380.
- [8] Gill D, Gavrieli N, Intrator N. Detection and identification of heart sounds using homomorphic envelopogram and self-organizing probabilistic model. In *Computers in Cardiology*. 2005; 957–960.
- [9] Springer DB, Tarassenko L, Clifford GD. Logistic regression-HSMM-based heart sound segmentation. *IEEE Trans Biomed Eng* 2016;63(4):822–832.
- [10] Kantz H, Schreiber T. *Nonlinear time series analysis*, volume 7. Cambridge university press, 2004.
- [11] Chen M, Carson W, Rodrigues M, Calderbank R, Carin L. Communications inspired linear discriminant analysis. In *Int. Conf. Machine Learn*. Jun.-Jul. 2012; 1507–1514.
- [12] Nenadic Z. Information discriminant analysis: Feature extraction with an information-theoretic objective. *IEEE Trans Patt Anal Mach Intell* Aug. 2007;29(8):1394–1407.
- [13] Bishop C. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

Address for correspondence:

Name: Francesco Renna

Full postal address: Wilberforce Rd, Cambridge CB3 0WA, UK

E-mail address: fr330@cam.ac.uk