2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Unsupervised Heart Sound Decomposition and State Estimation with Generative Oscillation Models

Ryohei Shibue[1], Masahiro Nakano[1], Tomoharu Iwata[1], Kunio Kashino[1] and Hitonobu Tomoike[1]

***Abstract*— This paper proposes a new generative probabilistic model for phonocardiograms (PCGs) that can simultaneously capture oscillatory factors and state transitions in cardiac cycles. Conventionally, PCGs have been modeled in two main aspects. One is a state space model that represents recurrent and frequently appearing state transitions. Another is a factor model that expresses the PCG as a non-stationary signal consisting of multiple oscillations. To model these perspectives in a unified framework, we combine an oscillation decomposition with a state space model. The proposed model can decompose the PCG into cardiac state dependent oscillations by reflecting the mechanism of cardiac sounds generation in an unsupervised manner. In the experiments, our model achieved better accuracy in the state estimation task compared to the empirical mode decomposition method. In addition, our model detected S2 onsets more accurately than the supervised segmentation method when distributions among PCG signals were different.**

***Index Terms*— Phonocardiogram, heart sound segmentation, state space model, oscillation decomposition, variational inference**

I. INTRODUCTION

Heart sounds are an important source of information to examine patients' cardiovascular status. A phonocardiogram (PCG), an electric recording of heart sounds, generally reflects various kinds of heart activities and has been used for heart disease detection. Many studies have tried to construct automatic heart sound analysis, but heart auscultation and its interpretation still largely depends on doctors' subjectivity.

The PCG has the following two features. First, it has a periodic structure comprising a small number of states and transitions among them. The heart repeatedly contracts and dilates to pump blood throughout the body. This cardiac cycle is mainly divided into four periods; the first heart sound (S1), systole period, second heart sound (S2), and diastole period. Second, the heart sound is composed of several factors that are attributable to their individual sources. S1 and S2 sounds originate from valve vibrations inside the heart: S1 comprises mitral and tricuspid valve closure, and S2 comprises aortic and pulmonary valve closure. Therefore, the heart sound can be viewed as a non-stationary signal consisting of multiple oscillations originated from valve vibrations.

These features have been used extensively in signal processing and machine learning methods for PCGs [1]. In the following, we briefly review previous research.

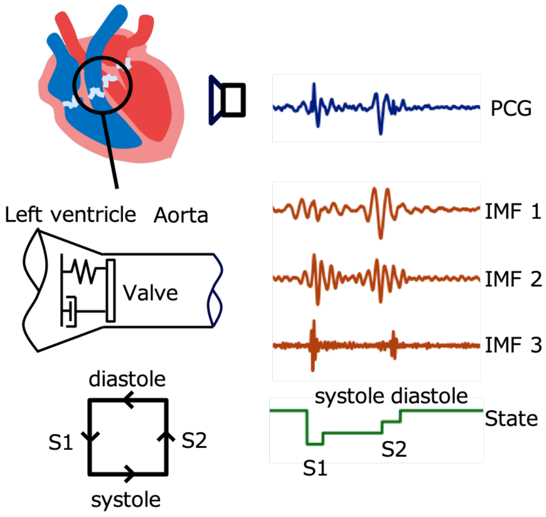


Fig. 1. Illustration of oscillation representation under our model. We distinguish four cardiac states: S1, systole, S2, and diastole (green). There are multiple hidden oscillations called intrinsic mode functions (IMFs) induced by valve closures, and their amplitude changes depending on the cardiac states (red). According to an idealized physical model for valve vibrations [2], we express the hidden oscillations as a second order autoregressive system. The PCG is observed as the summation of those oscillations (blue).

The state space models have been used for expressing cardiac state transitions, including hidden Markov models [3], [4], [5], [6], hidden semi Markov models [7], [8] and switching linear state space models [9]. On the other hand, the factorization representations have also been used for capturing latent factors from heart sounds, such as non-negative matrix factorization (NMF) [10], [11] and empirical mode decomposition (EMD) [12], [13], [14]. Other approaches have been studied, including frequency based features [15], [16], auto-correlation functions [17], [18], Shannon energy [19], and deep neural networks [20], [21], [22].

Recently, empirical mode decomposition (EMD) and its extensions have been considered as a promising direction of research in this field. EMD decomposes a one-dimensional signal into a sum of oscillations, called intrinsic mode functions (IMFs). Therefore, it has been expected to extract the non-stationary multiple oscillations from the valve vibrations in a data driven manner.

However, EMD has some limitations. First, it suffers from a mode mixing problem [23], which results in the loss of IMFs' physical meaning. Second, since EMD is calculated in a heuristic manner, incorporating generative mechanisms,

*This work was not supported by any organization

[1] Ryohei Shibue, Masahiro Nakano, Tomoharu Iwata, Kunio Kashino and Hitonobu Tomoike are with NTT Corporation, ryohei.shibue.vk@hco.ntt.co.jp

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

or knowledge about the heart, into decomposition is difficult. Third, the EMD-based methods are limited to one-dimensional PCG. During diagnosis, doctors change stethoscope locations on patients to obtain detailed information about heart activities and detect the sound propagation direction of abnormal heart sounds. Hence, extracting oscillation components shared by multidimensional PCG may be useful for diagnosis, though EMD has not been used for such signal analysis.

This paper aims to obtain natural oscillation representation for PCGs with a generative model. To this end, we construct a generative probabilistic model for PCGs unifying the following approaches: (1) a state space model that expresses cardiac state transitions and oscillations with uncertainty, and (2) EMD that gives useful factored representations. The existing EMD-based methods (e.g. [13], [14]) decompose PCG into multiple oscillations without considering the state transition structure. Thus, the amplitude of the obtained oscillations does not align with cardiac state transitions. The proposed model expresses oscillations and cardiac state transitions simultaneously so that it provides more natural oscillation representations well-aligned with cardiac states.

The main contributions of this paper are as follows.

- We propose a state space model with EMD-based factor representation for PCGs newly introducing a physical generation mechanism to simultaneously capture oscillatory factors and state transitions in the cardiac cycle. Our model definition can take advantage of multi-channel PCGs recorded with multiple microphones.
- We derive a tractable evidence lowerbound based on the black box structured variational family [24] for parameter estimation. This lowerbound reduces the number of parameters to be estimated.

II. The proposed model

A. Model definition

Consider that PCG signals with length R are simultaneously observed at D microphones and denote as $x_r \in \mathbb{R}^D$, $r = 1, 2, \ldots, R$. For notational simplicity, we denote a concatenate of multiple variables along time indices as $x_{1:R} = \{x_r\}_{r=1}^R$.

We assume that the PCG is a summation of multiple random oscillations, each of which depends on the hidden cardiac states. Specifically, we adopt the switching linear state space representation to express the PCG. Figure 1 shows an overview of our model.

First, we introduce the notion of state transitions into the cardiac cycle. Let z_r be a hidden state of the PCG signals on the r-th timestep. We suppose $z_r \in \{1, 2, 3, 4\}$, which we expect to correspond to the four states of the cardiac cycle, i.e., S1, systole, S2, and diastole. Each state (e.g., S1) stays in the current state or only moves to a particular state (e.g., systole). Additionally, the duration in the current state is considered to be independent of each other in each state. To achieve them, we define the transition probability as

$$p(z_{r:r+\delta-1} = j, z_{r+\delta} \neq j \mid z_r = i) = A_{ij} p_j(\delta), \quad (1)$$

where $A = [[0, 1, 0, 0], [0, 0, 1, 0], [0, 0, 0, 1], [1, 0, 0, 0]]$ is a transition matrix and $p_j(\delta)$ is a duration distribution in state j. For the duration distribution, we choose a negative binomial distribution with a success probability θ_i, $i = 1, \ldots, 4$ and a fixed shape parameter m

$$p_i(\delta) = \begin{pmatrix} \delta + m - 2 \\ \delta - 1 \end{pmatrix} (1 - \theta_i)^m \theta_i^{\delta - 1}, \ \delta = 1, 2, \ldots,$$

to use the following HMM embedding technique [25]. Let $e_r \in \{1, \cdots, m\}$ be pseudo states and define augmented states $\bar{z}_r = (z_r, e_r)$. Then, there is a transition probability matrix $\bar{A}$ over $\bar{z}_r$ that holds

$$p(z_{1:R}) = \sum_{e_r} p(\{(z_r, e_r)\}_{r=1}^R) = \sum_{e_r} p(\bar{z}_{1:R}).$$

The summation is taken over all pseudo state sequences. This relation implies that the posterior of $\bar{z}_r$ can be calculated in an HMM manner, and marginalizing it over e_r gives the posterior of z_r. This embedding reduces the computational cost of calculating posterior probabilities from $O(R^2)$ to $O(R)$.

Next, we introduce a random oscillation model into the PCGs. As we mentioned in Introduction, S1 and S2 sounds originate from valve vibrations inside the heart. Previous research attempted to represent those oscillations with dynamical systems [26], [27], [28]. The center of the left column in Figure 1 shows the membrane displacement model [2] that approximates the aortic valve dynamics. At the beginning of the diastole period, the blood pressure on the aortic valve instantaneously increases. This pressure increase induces the membrane vibrations, which decay exponentially according to the damping parameter of this membrane. This situation can be expressed as the differential equation

$$M\ddot{u} + C\dot{u} + Ku = \Delta P,$$

where u is the membrane displacement, ΔP is a blood pressure on the membrane, M is a mass of vibration, C is a damping factor, and K is a stiffness factor. The solution of this differential equation becomes a damped oscillation

$$u(t) \propto \exp(-\alpha t) \sin(\omega t - \psi).$$

Using this solution and the proportionality relation between the amplitude of the PCG and the velocity $\dot{u}$, we can assume that the hidden damped oscillation induced by the valve closure is expressed as a second order autoregressive model. Let $y_{rl} = (y_{rl}^{(1)}, y_{rl}^{(2)}) \in \mathbb{R}^2$, $r = 1, \ldots, R$ be coordinates of analytic signal representation of the l-th oscillation and define its dynamics given $z_r = i$ as

$$\begin{pmatrix} y_{rl}^{(1)} \\ y_{rl}^{(2)} \end{pmatrix} = a_l \begin{pmatrix} \cos(\frac{2\pi f_l}{f_{\mathrm{S}}}) & \sin(\frac{2\pi f_l}{f_{\mathrm{S}}}) \\ -\sin(\frac{2\pi f_l}{f_{\mathrm{S}}}) & \cos(\frac{2\pi f_l}{f_{\mathrm{S}}}) \end{pmatrix} \begin{pmatrix} y_{r-1,l}^{(1)} \\ y_{r-1,l}^{(2)} \end{pmatrix}$$

$$+ \begin{pmatrix} v_{rli}^{(1)} \\ v_{rli}^{(2)} \end{pmatrix}, \quad \begin{pmatrix} v_{rli}^{(1)} \\ v_{rli}^{(2)} \end{pmatrix} \sim \mathrm{N}\left(0, \sigma_{li}^2 I\right), \quad (2)$$

where a_l and f_l are a decay coefficient and a mean frequency of the l-th oscillation, respectively, and f_{S} is a sampling frequency. We here emphasize that variances σ_{li}^2 should vary depending on the hidden states, which reflects the changes of dominant oscillation in PCG according to the cardiac states. The larger system variances σ_{li}^2 means that the l-th oscillation is dominant in the state i. Note that the projection of $y_{rl}, r = 1, \dots, R$ onto the first coordinate corresponds to an IMF in EMD based methods. Hereafter, we denote $y_r = (y_{rl}^{(1)}, y_{rl}^{(2)}, \dots, y_{rL}^{(1)}, y_{rL}^{(2)})$.

PCG signals are summations of multiple oscillations filtered according to the relative distance and the internal body structure between the heart sound sources and mics. For simplicity, we assume that the observed PCG signals are expressed as the summations of the oscillations multiplied by the microphone-specific weights:

$$x_r = \begin{pmatrix} 1 & 0 & \cdots & 1 & 0 \\ g_{21}^{(1)} & g_{21}^{(2)} & \cdots & g_{2L}^{(1)} & g_{2L}^{(2)} \\ \vdots & & \ddots & & \vdots \\ g_{D1}^{(1)} & g_{D1}^{(2)} & \cdots & g_{DL}^{(1)} & g_{DL}^{(2)} \end{pmatrix} \begin{pmatrix} y_{r1}^{(1)} \\ y_{r1}^{(2)} \\ \vdots \\ y_{rL}^{(1)} \\ y_{rL}^{(2)} \end{pmatrix} + w_r,$$

$$w_r \sim \mathrm{N}(0, \tau^2 I). \tag{3}$$

Similar state space representations have been used for instantaneous phase estimation [29], [30]. Our model includes them as a special case: when $\sigma_{l1}^2 = \sigma_{l2}^2 = \sigma_{l3}^2 = \sigma_{l4}^2$, our model reduces to the linear Gaussian state space model without hidden state transitions in these studies.

B. Parameter estimation

We begin with a rough sketch of our Bayesian inference method. Our goal is to estimate model parameters maximizing the log marginal likelihood given observations. However, our model does not lead to a closed form expression for the marginal likelihood. Thus, we use a variational inference with the structured black-box variational family [24]. We devise a neural network that approximates the relationship between observations and hidden states directly. Adopting such an inference network reduces the number of parameters, which leads to a simple implementation and good practice mixing.

The key idea of variational inference is to approximate the posterior and the marginal likelihood through optimization. We first introduce variational posterior $q(y_{1:R}, \bar{z}_{1:R})$ to obtain the lowerbound for the log marginal likelihood [31]:

$$\log p(x_{1:R}) \\ \geq \mathrm{E}_{q(y_{1:R}, \bar{z}_{1:R})} \left[\log \frac{p(x_{1:R} \mid y_{1:R}) p(y_{1:R} \mid \bar{z}_{1:R}) p(\bar{z}_{1:R})}{q(y_{1:R}, \bar{z}_{1:R})} \right],$$

where $p(x_{1:R} \mid y_{1:R})$ is the observation model (3), $p(y_{1:R} \mid \bar{z}_{1:R})$ is the system model (2) and $p(\bar{z}_{1:R})$ is the prior distribution (1). This lowerbound is called evidence lowerbound and the equality holds when $q(y_{1:R}, \bar{z}_{1:R})$ equals to the true posterior $p(y_{1:R}, \bar{z}_{1:R} \mid x_{1:R})$. Instead of maximizing the log marginal likelihood, we maximize this evidence lowerbound with respect to model parameters and the variational posterior.

The variational posterior $q(y_{1:R}, \bar{z}_{1:R})$ should be a good approximation for the true posterior and should be easy to calculate the expectation. To satisfy these requirements, we first impose independence assumption as $q(y_{1:R}, \bar{z}_{1:R}) = q(y_{1:R}) q(\bar{z}_{1:R})$. Under this assumption, the variational posterior $q(y_{1:R})$ maximizes the lowerbound is uniquely determined and the evidence lowerbound reduces to

$$\log p(x_{1:R}) \geq \log \rho(x_{1:R}) + \mathrm{E}_{q(\bar{z}_{1:R})} \left[\log \frac{p(\bar{z}_{1:R})}{q(\bar{z}_{1:R})} \right], \tag{4}$$

where

$$\rho(x_{1:R}) \\ = \int p(x_{1:R} \mid y_{1:R}) \exp \left(\mathrm{E}_{q(\bar{z}_{1:R})} \left[\log p(y_{1:R} \mid \bar{z}_{1:R}) \right] \right) \mathrm{d}y_{1:R}.$$

Note that $\rho(x_{1:R})$ is not a normalized density function. Derivation of this lowerbound is shown in Appendix.

Then, we define $q(\bar{z}_{1:R})$ as the following structured expression:

$$q(\bar{z}_{1:R}) \propto p(\bar{z}_1) \prod_{r=2}^{R} p(\bar{z}_r \mid \bar{z}_{r-1}) \prod_{r=1}^{R} \psi(\bar{z}_r, x_{1:R}).$$

This variational posterior depends on the parameters in the state transition model (1) and the oscillation model (2), as well as on the newly introduced part $\psi(\bar{z}_r, x_{1:R})$. This $\psi(\bar{z}_r, x_{1:R})$ is the r-th node potential that provides probabilistic guess at each hidden state inferred from observations $x_{1:R}$. To capture the complex dependence between $\bar{z}_r$ and $x_{1:R}$, we adopt a neural network for $\psi(\bar{z}_r, x_{1:R})$. In the following experiment, we define this ψ as

$$\psi(\bar{z}_r = (i, \delta), x_{1:R}) = [\phi(x_{r-s+1:r})]_{(i,\delta)}, \tag{5}$$

where $\phi(\cdot)$ is a neural network that maps $x_{r-s+1:r} \in \mathbb{R}^{sd}$ to $\phi(x_{r-s+1:r}) \in \mathbb{R}_{+}^{4 \times m}$, s is a fixed window length, and $[\cdot]_{(i,\delta)}$ is an operator that returns the (i, δ)-th element of the input vector. The possible choice for this ϕ is a convolutional neural network (CNN). With CNN, the node potential first calculates the dominant sound features in the window $[r-s+1, r]$ and then converts them into a probabilistic guess about the location of this window relative to cardiac cycles. Other candidates for ψ are specific structured neural networks used for heart sound segmentation in a supervised manner (e.g., CNN [32] and RNN [33], [34]). The prior distribution $p(\bar{z}_1)$ and $p(\bar{z}_r \mid \bar{z}_{r-1})$, $r = 2, \dots, R$ connects neighboring node potentials in Markov manner.

With this variational posterior, the lowerbound (4) is easy to be calculated by utilizing message-passing algorithms. The first term can be expressed as a closed form by using the Kalman filter. The expectation with respect to $q(\bar{z}_{1:R})$ needs marginal posterior $q(\bar{z}_r)$, $r = 1, \dots, R$ and $q(\bar{z}_{r-1}, \bar{z}_r)$, $r = 2, \dots, R$, and these terms can also be calculated by using the forward-backward algorithm for hidden Markov models. In addition, introducing the noise ratio parameters $\tilde{\sigma}_{li}^2 = \sigma_{li}^2 / \tau^2$, and considering maximization with respect to $\tilde{\sigma}_{li}^2$ instead of σ_{li}^2, we obtain a closed form solution for τ^2 [35]. This replacement reduces the number of parameters to be etimated and makes estimation more stable.

In the estimation step, we maximize the lowerbound (4) with respect to the model parameters $(\{\theta_i\}, \{a_l\}, \{f_l\}, \{\tilde{\sigma}_{li}\}, \{g_{dl}\})$ and variational parameters in the node potential function ψ by using gradient ascent. The computational cost for calculating the objective function in each step is $O(R)$.

Structured black box variational families have been used for approximate inference in hierarchical state space models. For example, [36] chose a structured black box Gaussian family for variational posterior $q(y_{1:R})$ and calculated the evidence lowerbound with Monte-Carlo sampling. The main difference relative to such work is that the inference network in our method directly connects bottom hidden states and observations. This formulation provides the closed form evidence lowerbound, and we need not carry Monte-Carlo sampling for it.

III. Applications and Experiments

Since our proposed model can be fitted to the observed PCGs in an unsupervised manner, it could potentially be applied to a variety of applications, including anomalous sound detection and disease prediction, combined with other machine learning methods. As a telling example, this paper demonstrates the usefulness of the proposed model by taking one of the most fundamental applications, the heart sound segmentation task.

Approaches for the heart segmentation task are mainly divided into two ways: supervised and unsupervised approaches. Supervised approaches use human-annotated labels or reference ECG signals for model training and outputs the cardiac state labels for other PCG signals. On the other hand, unsupervised approaches use only PCG signal for model training and outputs cardiac state labels simultaneously. Supervised approaches generally achieve better segmentation performance compared to unsupervised approaches. However, their use is limited to the case where the training labels are given and the distributions of training and test signals are consistent.

Heart sound segmentation under our model is classified as an unsupervised approach. Our model can handle the difference in the distributions of PCG signals and does not require cardiac state labels for training. In the experiments, we compared our model to the existing unsupervised heart sound segmentation based on EMD decomposition. We also compared our model to LR-HSMM, the state art of supervised heart sound segmentation method.

A. Datasets

Five datasets were used for this experiment. The first and the second ones were normal and abnormal PCG signals under various kinds of symptoms attached to auscultation textbooks [37], [38]. In total, these datasets contain 119 signals. The onsets of S1 were annotated manually. Figure 2 shows examples of PCGs in these datasets. We denote these datasets as (a) and (b). The third one was a heart sound signal obtained from one of the authors with a microphone. The onsets of S1 and S2 were annotated by hand. We denote it as (c). The fourth one was a 2016 PhysioNet/CinC dataset [1], [39]. This dataset contains simultaneously recorded PCG and ECG signals. PCG signals in this dataset were divided into normal and abnormal heart sounds: normal sounds were from healthy subjects and abnormal sounds were from subjects with certain cardiac diseases. We denote it as (d). The last one was a multidimensional PCG signal measured at two microphones simultaneously. We denote it as (e).

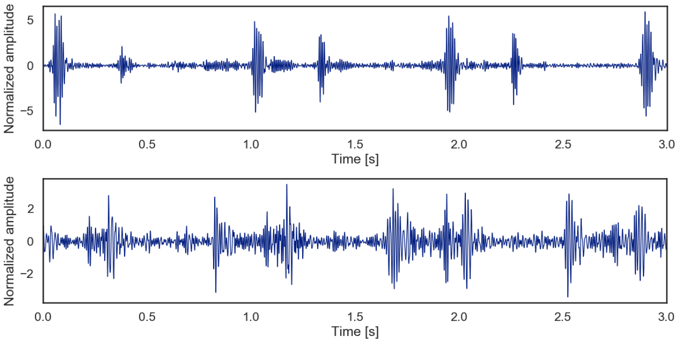


Fig. 2. Examples of PCG signals attached to auscultation textbooks [37], [38] in datasets (a) and (b). (Top) Normal. (Bottom) Mitral regurgitation.

B. Comparing methods

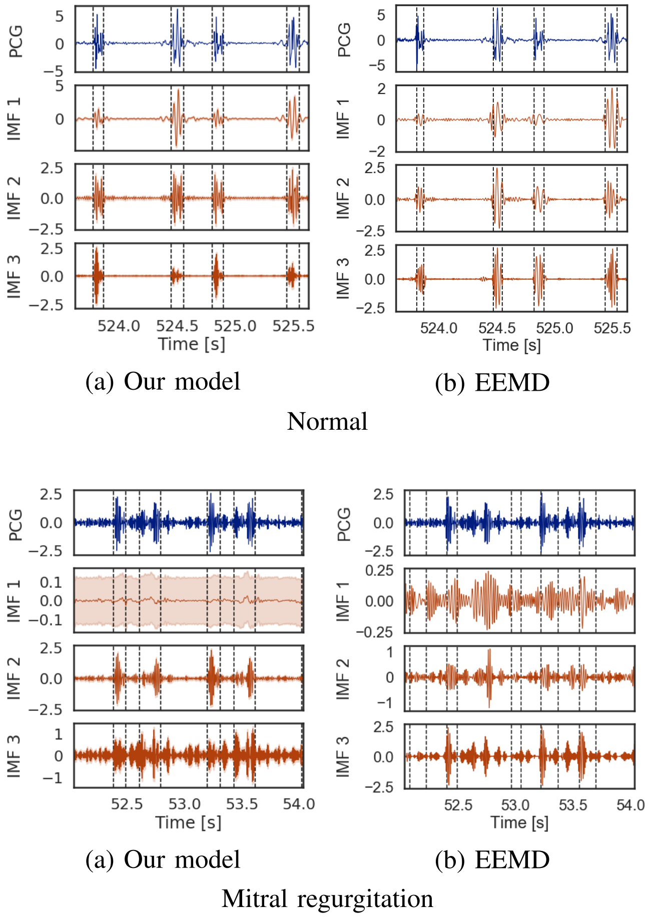
We chose the ensemble empirical mode decomposition method with a kurtosis feature [14] for comparison. Hereafter, we denote this method as EEMD. The assumption that multiple oscillations exist in PCGs also holds in this approach.

In EEMD, we first apply an ensemble version of EMD to the observed signals to extract IMFs [23]. During S1 and S2, the amplitude of each IMF instantaneously increases. To detect this increase, we calculated kurtosis values from IMFs using a sliding window. If the window contains the onsets of S1 and S2 sound, the marginal distribution of values in this window becomes heavy-tailed and its kurtosis grows. Therefore, finding peaks of the product of the kurtosis values along IMFs and different scale windows gives the estimates of S1 and S2 onsets.

LR-HSMM expresses PCG signals as hidden semi Markov model with logistic observation model. Parameters of this model should be estimated from training PCG signals and true cardiac labels beforehand in a supervised manner.

C. Procedure and evaluation

For datasets (a), (b), and (c), we first down-sampled all the PCGs to 2,000 Hz and applied a band-pass filter with cut off frequencies of 10 Hz and 150 Hz. Then, we applied our model and EEMD to those signals. Although the noise levels were different among the signals and the datasets, we used the same hyperparameter values to test the robustness of the two methods. In our model, we define the node potential



(a) Our model (b) EEMD

Normal

(a) Our model (b) EEMD

Mitral regurgitation

Fig. 3. Examples of oscillation decomposition and heart sound segmentation results with our model and EEMD. The blue lines indicate input PCG signals and the red lines indicate the obtained oscillations. The red bands show the 95% credible intervals with our model. The dashed lines indicate the estimated onsets and offsets of S1 and S2. (Top) Normal. (Bottom) Mitral regurgitation.

$\psi(\bar{z}_r, x_{1:R})$ as (5) and adopted a two-layer convolutional neural network for $\phi(x_{r-s+1:r})$.

To compare unsupervised and supervised segmentation approaches, we applied our model and LR-HSMM to dataset (d). We first divided the dataset into 50 pairs of normal heart sounds and abnormal heart sounds. In LR-HSMM, we estimated parameters from the 40 second normal heart sounds and the cardiac labels, and decomposed 10 second abnormal heart sounds using the estimated parameters. In our model, we shared the convolutional neural network in $\psi(\bar{z}_r, x_{1:R})$ among PCGs and decomposed these normal and abnormal PCGs simultaneously. Accuracy was calculated from the segmentation results of abnormal heart sounds.

We judged that S1 and S2 onsets are correctly estimated if the estimated onset and the ground-truth onset are located within the 100 millisecond interval. To evaluate the segmentation performance objectively, we calculated the F_1 score:

$$F_1 = \frac{2 \times P_+ \times S_e}{P_+ + S_e},$$

where P_+ and S_e are precision and recall respectively.

TABLE I

ACCURACY OF S1 AND S2 DETECTION. THE NUMBERS OF S1 AND S2 (N), TRUE POSITIVES (TP), FALSE POSITIVES (FP), FALSE NEGATIVES (FN) AND F_1-SCORES ARE LISTED.

			N	TP	FP	FN	$F_1(\%)$
(a)	S_1	Our model	1089	703	459	386	**62.46**
		EEMD		488	365	601	50.26
(b)	S_1	Our model	2002	1660	464	342	**80.47**
		EEMD		1171	582	831	62.37
(c)	S_1	Our model	1709	1660	87	49	**96.06**
		EEMD		1399	127	310	86.49
	S_2	Our model	1719	1681	86	38	**96.44**
		EEMD		1678	517	41	87.24
(d)	S_1	Our model	600	546	49	54	91.38
		LR-HSMM		586	16	14	**97.50**
	S_2	Our model	587	504	90	83	**85.35**
		LR-HSMM		453	135	134	77.11

D. *Results*

Figure 3 shows a typical example of segmentation results for normal and abnormal PCG obtained with our model and EEMD. Although both methods detected roughly the same S1 and S2 sections in these signals, the obtained IMFs are different. That is, the amplitudes of IMFs rapidly increase in S1 and S2 intervals in the proposed method, compared with the IMFs obtained with EEMD. This is because our model utilizes the cardiac state transition structure in its oscillation model.

Table I lists the accuracy of S1 and S2 onset detection. This table shows that our model estimated S1 and S2 more accurately than EEMD for all three datasets. The accuracy of both methods for the datasets (a) and (b) were relatively lower than the dataset (c). This is because the datasets (a) and (b) contain abnormal PCG signals that are difficult to divide into four states. We also experienced that it was not straightforward to optimize the multiple hyperparameters in the EEMD case. For example, the number of the oscillations, the length of the sliding windows, and the detection threshold must be determined in advance but highly affect the segmentation results. The results obtained from the dataset (d) show the difference in segmentation between unsupervised and supervised approaches. LR-HSMM detected S1 onsets better than our model, whereas our model detected S2 onsets more accurately than LR-HSMM. This is because the shapes of S2 differ between normal and abnormal heart sounds compared to S1, and the assumption that training and test signals distribute to the same distribution does not hold in S2 intervals. In our model, parameters other than the inference network are estimated separately for each PCG, which allows us to capture the differences in S2 sounds. On the other hand, LR-HSMM reduced false positives and false negatives in S1 detection since S1 sounds have similar shapes and amplitudes between training and test signals.

Figure 4 shows an oscillation decomposition example obtained from (e). Our model can extract hidden oscillations shared by multiple signals measured at different micro-

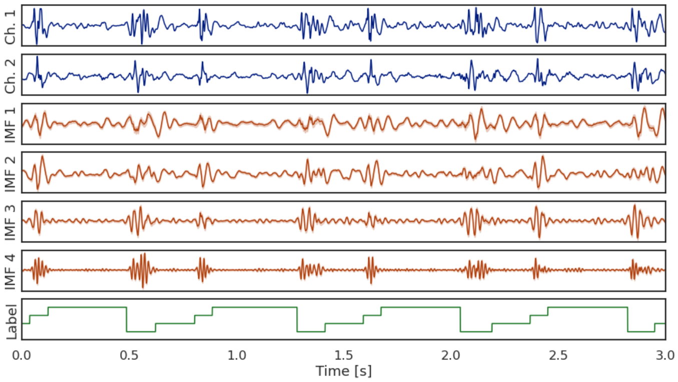


Fig. 4. An example of oscillation decomposition obtained from a two dimensional PCG. The blue lines indicate the row signals, the red lines indicate the obtained oscillations with 95% credible intervals, and the green line indicates the estimated label.

phones. We believe that using multidimensional PCG signal improves the oscillation estimation accuracy and gives clues about the location of the heart sound source, and such local analysis is very useful for understanding and analyzing the three-dimensional cardiac activity.

IV. Conclusions

We proposed a new generative probabilistic model for multidimensional PCGs. The proposed model can simultaneously capture oscillatory factors and the cardiac state transitions. This model is inspired by the oscillation decomposition representation [29], [30]. We extended this representation to the case of non-stationary signals, combining it with the hidden semi-Markov model, and introduced the physical generation mechanism of PCG for increased resolutions of the analysis. To derive the tractable evidence lowerbound for Bayesian inference, we adopted the structured black box variational family that directly connects hidden states and observations.

The experiments showed that the proposed model achieved better performance than the EMD-based method in the heart sound segmentation task. Compared to the supervised segmentation method, the proposed model was able to detect S2 onsets more accurately when the distributions of PCG signals were different. We also demonstrated that the improved resolution in the proposed method is useful for the analysis of multi-channel PCGs observed with multiple microphones. We believe that the proposed method can be a useful tool for spatiotemporal monitoring of cardiac activity using only inexpensive and convenient devices. We plan to discuss this point quantitatively in future work.

Appendix

A. Derivation of the evidence lowerbound

We here derive the evidence lowerbound (4). The evidence lowerbound under the variational posterior $q(y_{1:R})q(\bar{z}_{1:R})$ can be expressed as

$$\mathrm{E}_{q(y_{1:R})q(\bar{z}_{1:R})}\left[\log \frac{p(x_{1:R} \mid y_{1:R})p(y_{1:R} \mid \bar{z}_{1:R})p(\bar{z}_{1:R})}{q(y_{1:R})q(\bar{z}_{1:R})}\right]$$

$$= \mathrm{E}_{q(y_{1:R})}\Big[\log p(x_{1:R} \mid y_{1:R}) + \mathrm{E}_{q(\bar{z}_{1:R})}\left[\log p(y_{1:R} \mid \bar{z}_{1:R})\right] - \log q(y_{1:R})\Big] + \mathrm{E}_{q(\bar{z}_{1:R})}\left[\log \frac{p(\bar{z}_{1:R})}{q(\bar{z}_{1:R})}\right]$$

$$= \log \rho(x_{1:R}) - \mathrm{KL}(q(y_{1:R})||\rho(y_{1:R} \mid x_{1:R})) + \mathrm{E}_{q(\bar{z}_{1:R})}\left[\log \frac{p(\bar{z}_{1:R})}{q(\bar{z}_{1:R})}\right]$$

$$\leq \log \rho(x_{1:R}) + \mathrm{E}_{q(\bar{z}_{1:R})}\left[\log \frac{p(\bar{z}_{1:R})}{q(\bar{z}_{1:R})}\right].$$

Here,

$$\rho(x_{1:R}) = \int p(x_{1:R} \mid y_{1:R}) \exp\left(\mathrm{E}_{q(\bar{z}_{1:R})}\left[\log p(y_{1:R} \mid \bar{z}_{1:R})\right]\right) \mathrm{d}y_{1:R},$$

$$\rho(y_{1:R} \mid x_{1:R}) = \frac{p(x_{1:R} \mid y_{1:R}) \exp\left(\mathrm{E}_{q(\bar{z}_{1:R})}\left[\log p(y_{1:R} \mid \bar{z}_{1:R})\right]\right)}{\int p(x_{1:R} \mid y_{1:R}) \exp\left(\mathrm{E}_{q(\bar{z}_{1:R})}\left[\log p(y_{1:R} \mid \bar{z}_{1:R})\right]\right) \mathrm{d}y_{1:R}},$$

and $\mathrm{KL}(q(y_{1:R})||\rho(y_{1:R} \mid x_{1:R}))$ is the Kullback-Leibler divergence between $q(y_{1:R})$ and $\rho(y_{1:R} \mid x_{1:R})$. The last equality holds when $q(y_{1:R}) = \rho(y_{1:R} \mid x_{1:R})$. Hence, the maximization of the left hand side of this inequality with respect to $q(y_{1:R})$ and $q(\bar{z}_{1:R})$ reduces to the maximization of the right hand side with respect to $q(\bar{z}_{1:R})$. This concludes the derivation of (4).

References

[1] C Liu, D Springer, Q Li, B Moody, R A Juan, F J Chorro, F Castells, J M Roig, I Silva, A E W Johnson, Z Syed, S E Schmidt, C D Papadaniil, L Hadjileontiadis, H Naseri, A Moukadem, A Dieterlen, C Brandt, H Tang, M Samieinasab, M R Samieinasab, R Sameni, R G Mark, and G D Clifford, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, 2016.

[2] E F Blick, H N Sabbah, and P D Stein, "One-dimensional model of diastolic semilunar valve vibrations productive of heart sounds," *J. Biomech.*, vol. 12, no. 3, pp. 223–227, 1979.

[3] L G Gamero and R Watrous, "Detection of the first and second heart sound using probabilistic models," *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 3, pp. 2877–2880, 2003.

[4] D Gill, N Gavrieli, and N Intrator, "Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model," *Proc. Comput. Cardiol.*, pp. 957–960, 2005.

[5] A D Ricke, R J Povinelli, and M T Johnson, "Automatic segmentation of heart sound signals using hidden markov models," *Proc. Comput. Cardiol.*, pp. 953–956, 2005.

[6] P Sedighian, A W Subudhi, F Scalzo, and S Asgari, "Pediatric heart sound segmentation using hidden markov model," *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 5490–5493, 2014.

[7] S E Schmidt, C Holst-Hansen, C Graff, E Toft, and J J Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden markov model," *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, 2010.

[8] D B Springer, L Tarassenko, and G D Clifford, "Logistic regression-HSMM-based heart sound segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 822–832, 2016.

[9] F Noman, S-H Salleh, C-M Ting, S B Samdin, H Ombao, and Hi Hussain, "A markov-switching model approach to heart sound segmentation and classification," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 3, pp. 705–716, 2020.

[10] F Sattar, F Jin, A Moukadem, C Brandt, and A Dieterlen, "Time-scale-based segmentation for degraded PCG signals using NMF," in *Non-negative Matrix Factorization Techniques: Advances in Theory and Applications*, pp. 179–194. Springer, 2016.

[11] N Dia, J Fontecave-Jallon, P Gumery, and B Rivet, "Quasi-periodic non-negative matrix factorization for phonocardiographic signals denoising," *Proc. Sensor Array and Multi. chan. Sig. Proces.*, pp. 390–394, 2018.

[12] S Ari and Go Saha, "Classification of heart sounds using empirical mode decomposition based features," *Int. J. Med. Eng. Inform.*, vol. 1, no. 1, pp. 91–108, 2008.

[13] H Sun, W Chen, and J Gong, "An improved empirical mode decomposition-wavelet algorithm for phonocardiogram signal denoising and its application in the first and second heart sound extraction," *Proc. Int. Conf. Biomed. Eng. Inform.*, pp. 187–191, 2013.

[14] C D Papadaniil and L J Hadjileontiadis, "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1138–1152, 2014.

[15] H Naseri and M R Homaeinezhad, "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric," *Ann. Biomed. Eng.*, vol. 41, no. 2, pp. 279–292, 2013.

[16] A Castro, T T V Vinhoza, S S Mattos, and M T Coimbra, "Heart sound segmentation of pediatric auscultations using wavelet analysis," *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2013, pp. 3909–3912, 2013.

[17] M S Manikandan and K P Soman, "Robust heart sound activity detection in noisy environments," *Electron. Lett.*, vol. 46, no. 16, pp. 1100–1102, 2010.

[18] J Pedrosa, A Castro, and T T V Vinhoza, "Automatic heart sound segmentation and murmur detection in pediatric phonocardiograms," *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2014, pp. 2294–2297, 2014.

[19] V N Varghees and K I Ramachandran, "A novel heart sound activity detection framework for automated heart sound analysis," *Biomed. Signal Process. Control*, vol. 13, pp. 174–188, 2014.

[20] T Oskiper and R L Watrous, "Detection of the first heart sound using time-delayed neural network," *Proc. Comput. Cardiol.*, vol. 29, pp. 537–540, 2002.

[21] T-E Chen, S-I Yang, L-T Ho, K-H Tsai, Y-H Chen, Y-F Chang, Y-H Lai, S-S Wang, Y Tsao, and C-C Wu, "S1 and S2 heart sound recognition using deep neural networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 372–380, 2017.

[22] M Tschannen, T Kramer, G Marti, M Heinzmann, and T Wiatowski, "Heart sound classification using deep structured features," *Proc. Comput. Cardiol.*, pp. 565–568, 2016.

[23] Z Wu and N E Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009.

[24] M J Johnson, D K Duvenaud, A Wiltschko, R P Adams, and S R Datta, "Composing graphical models with neural networks for structured representations and fast inference," *Proc. Adv. Neural Inf. Process. Syst.*, pp. 2946–2954, 2016.

[25] M J Johnson and A S Willsky, "Stochastic variational inference for bayesian time series models," *Proc. Int. Conf. Mach. Learn.*, pp. II–1854–II–1862, 2014.

[26] D L Sikarskie, P D Stein, and M Vable, "A mathematical model of aortic valve vibration," *J. Biomech.*, vol. 17, no. 11, pp. 831–837, 1984.

[27] H Ozcan Gulcur and Y Bahadirlar, "Estimation of systolic blood pressure from the second heart sounds," *Proc. Int. Biomed. Eng. Days*, pp. 39–41, 1998.

[28] X-Y Zhang and Y-T Zhang, "Model-based analysis of effects of systolic blood pressure on frequency characteristics of the second heart sound," *Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2006, pp. 2888–2891, 2006.

[29] T Matsuda and F Komaki, "Multivariate time series decomposition into oscillation components," *Neural Comput.*, vol. 29, no. 8, pp. 2055–2075, 2017.

[30] H Soulat, E P Stephen, A M Beck, and P L Purdon, "State space methods for phase amplitude coupling analysis," *bioRxiv*, 2019.

[31] M I Jordan, Z Ghahramani, T S Jaakkola, and L K Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, pp. 105–161, 1998.

[32] F Renna, J Oliveira, and M T Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2435–2445, 2019.

[33] E Messner, M Zohrer, and F Pernkopf, "Heart sound segmentation-an event detection approach using deep recurrent neural networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1964–1974, 2018.

[34] T Fernando, H Ghaemmaghami, S Denman, S Sridharan, N Hussain, and C Fookes, "Heart sound segmentation using bidirectional LSTMs with attention," *IEEE J. Biomed. Health. Inform.*, vol. 24, no. 6, pp. 1601–1609, 2020.

[35] G Kitagawa, *Introduction to Time Series Modeling*, CRC Press, 2010.

[36] S Linderman, A Nichols, D Blei, Ml Zimmer, and L Paninski, "Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in c. elegans," *bioRxiv*, 2019.

[37] T Sawayama, *Auscultation Training with CD; Heart Sounds (in Japanese)*, Nankodo, 1994.

[38] T Sawayama, *Master of Auscultation; Heartbeat Shower with CD (in Japanese)*, Japan Medical Publisher, 1998.

[39] A L Goldberger, L A N Amaral, L Glass, J M Hausdorff, P C Ivanov, R G Mark, J E Mietus, G B Moody, C K Peng, and H E Stanley, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, pp. E215–E220, 2000.