On modifying the temporal modeling of HSMMs for pediatric heart sound segmentation

Jorge Oliveira
Instituto de Telecomunicações,
Faculdade de Ciências da Universidade do Porto
Email: oliveira_jorge@dcc.fc.up.pt

Theofrastos Mantadelis
CRACS & INESC TEC,
Faculdade de Ciências da Universidade do Porto
Email: theo.mantadelis@dcc.fc.up.pt

Francesco Renna
Department of Applied Mathematics
and Theoretical Physics, University of Cambridge
Email: fr330@cam.ac.uk

Pedro Gomes and Miguel Coimbra
Instituto de Telecomunicações,
Faculdade de Ciências da Universidade do Porto
Email: {ptmgomes,mcoimbra}@dcc.fc.up.pt

***Abstract*—Heart sounds are difficult to interpret because a) they are composed by several different sounds, all contained in very tight time windows; b) they vary from physiognomy even if the show similar characteristics; c) human ears are not naturally trained to recognize heart sounds. Computer assisted decision systems may help but they require robust signal processing algorithms. In this paper, we use a real life dataset in order to compare the performance of a hidden Markov model and several hidden semi Markov models that used the Poisson, Gaussian, Gamma distributions, as well as a non-parametric probability mass function to model the sojourn time.**

Using a subject dependent approach, a model that uses the Poisson distribution as an approximation for the sojourn time is shown to outperform all other models. This model was able to recreate the "true" state sequence with a positive predictability per state of 96%. Finally, we used a conditional distribution in order to compute the confidence of our classifications. By using the proposed confidence metric, we were able to identify wrong classifications and boost our system (in average) from an $\approx$ 83% up to $\approx$ 90% of positive predictability per sample.

I. Introduction

The phonocardiogram (PCG) signal is recorded during an auscultation using an electronic stethoscope. The PCG contains important information concerning the mechanical activity of the heart valves [1]. Signal processing of a PCG has two main goals: the first one is to split the PCG into heart cycles. Each heart cycle is mainly composed by the first heart sound ($S1$), the systolic period ($siSys$), the second heart sound ($S2$), and the diastolic period ($siDia$). The second goal is the detection of other sounds such as the third and fourth heart sounds ($S3$ and $S4$ respectively), heart murmurs, snaps, etc.

The methods used for heart sound segmentation can be divided depending on which domain they are applied: the time domain (Shannon energy [2]), the frequency domain (homomorphic filter [3]), etc. Moreover, different approaches have been proposed to assign features extracted from the PCG to the different segments/states, e.g., Artificial Neural Networks (ANN) [4], Support Vector Machines (SVM) [5] and hidden Markov models (HMM).

Among these, HMMs and their variations have the advantage of naturally modeling the sequential nature of heart sound signals. Recently, HMMs have shown to be very effective in modeling the heart sound signals: in Gill et al. [6], the signal is pre-processed and a subset of candidates (peaks) are extracted from the homomorphic envelogram, and these candidates are classified using a discrete-time HMM, where the state-distribution is modeled using the time-duration from the preceding candidate to the current one. Chung [7], detected and classified heart sounds using first a left-right HMM model (the first state is assumed to be known) and later a fully-connected HMM. The variability in each state is modeled by using multiple mixtures of a Gaussian multivariate distribution.

Schmidt et al. [8] implemented a hidden semi Markov model (HSMM) [9] using the homomorphic filtering envelogram as an observation to the system. This has the advantage (compared to the traditional HMM) that every state duration is explicitly modeled in the state transition matrix. The state duration distribution function is modeled by a Gaussian distribution, where the systolic ($siSys$) and diastolic ($siDia$) duration parameters are estimated through autocorrelation analysis of the homomorphic filtering envelogram. Springer et al. [10] expanded Schmidt's algorithm mainly on the study of the emission probability distribution. He explored a wider range of features and machine learning approaches to model the emission probabilities.

When using HMMs, the sojourn time (waiting time) is geometrically distributed over all states. This is an unrealistic assumption in heart sound signals, due to physiological time constraints that exist in the cardiac cycle. For example, the cardiac muscle, like any excitable tissue, exhibits a refractory period to re-stimulation. During this time interval, normal

This article is a result and funded by the project NanoSTIMA, 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 internal project SmartHeart in scope of project UID/EEA/50008/2013. T. Mantadelis is funded by SMILeS (PL02024) project. F. Renna was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 655282.

cardiac impulse cannot re-excite an already excited area of the cardiac muscle [1]. On the other hand, the geometric distribution decreases monotonically, as a result the most probable sojourn time duration a priori is equal to one sample. In the current state-of-the-art for heart sound classification using HSMMs [8], [10], only the Gaussian distribution is examined as an approximation for the sojourn time distribution. But the standard Gaussian distribution is not strictly positively defined and therefore it is not the most advisable distribution, since the sojourn times are by nature strictly positive. Imposing the positivity constraint, would cause the variance to approach zero for very fast events. Motivated by the above observations, this work proposes to enhance the performance of PCG segmentation for pediatric subjects via the following contributions:

1) the study of different distributions for approximating the sojourn time in HSMMs;
2) the experimental validation of the performance of each presented model over a real-life pediatric dataset;
3) finally, we propose a novel confidence metric based on conditional probability.

The paper is organized as follows: in section 2 we present HMMs and HSMMs. In section 3, we explain our methodology. In section 4, we present our experimental results. Finally, we draw conclusions.

II. Modeling Heart Sounds

A. Hidden Markov Models

HMMs are probabilistic models, where the observation sequence $X = x_1, x_2, \cdots, x_n$ depends on the underlying hidden state sequence $S = s_1, s_2, \cdots, s_n$ and the unobserved Markov process [11]. A homogeneus hidden Markov model assumes that the state transition probability matrix Γ is constant over time. In this case, the i^{th} row and j^{th} column entry of Γ, i.e.,

$$\gamma_{ij} = Pr(s_t = j | s_{t-1} = i), \tag{1}$$

is the probability of being in state j knowing that the previous state was i, and such probability is independent of current evaluation time t [7]. It is assumed that each state of a HMM corresponds to an element of the heart sound signal because the signal characteristics in each element are thought to be homogeneous. For simplicity, our model ignores $S3$, $S4$ and murmur sounds. The likelihood [11] of a state sequence S of a HMM with an observation sequence X is:

$$P(X, S, \Theta) = \pi_1 \left\{ \prod_{t=2}^{n} \gamma_{s_{t-1} s_t} \right\} \prod_{l=1}^{n} P_{s_l}(x_l), \tag{2}$$

where $\Theta = \{\pi_1, \Gamma, B\}$ denotes the model parameters such as the initial state distribution π_1, state transition probability matrix Γ and state depended distribution matrix[1] B. In this work, B is assumed to be a continuous Gaussian function:

$$p(x_t | \mu_{s_k}, \nu_{s_k}) = \frac{1}{\nu_{s_k}\sqrt{2\pi}} e^{-(x_t - \mu_{s_k})^2 / (2\nu_{s_k}^2)}, \tag{3}$$

[1]Note that in case the emissions are modeled by continuous random variables B contains the parameters of the pdf for each state.

with ν_{s_k} being the standard deviation and μ_{s_k} the mean value in state s_k.

B. Hidden Semi Markov Models

In HMMs, the sojourn time (expressed in number of time samples) is geometrically distributed over all states [10]. This is an unrealistic assumption in heart sound signals, since the state transition probabilities are constantly changing over time. The solution we consider is to model explicitly the sojourn time by using a HSMM [9].

We first define, D as the sojourn time distribution matrix. The entries of D are $d_{s_k}(u_k)$ which is the probability of spending u_k units of time in the state $s_k \in \mathcal{S} = \{S1, siSys, S2, siDia\}$. We use five different approaches to model the sojourn time. Four of them are represented by parametric distributions, whereas the last one is a non-parametric probability mass function:

- Parametric sojourn time distributions:
  - Poisson:

$$d_{s_k}(u_k | \lambda_{s_k}) = \frac{e^{-\lambda_{s_k}} \lambda_{s_k}^{u_k}}{u_k!}, \tag{4}$$

  where λ_{s_k} is the expected sojourn time in the state s_k.
  - Gaussian:

$$d_{s_k}(u_k | \lambda_{s_k}, \sigma_{s_k}) = \frac{1}{\sigma_{s_k}\sqrt{2\pi}} \cdot e^{-\frac{(u_k - \lambda_{s_k})^2}{2\sigma_{s_k}^2}}, \tag{5}$$

  where λ_{s_k} is the expected sojourn time in the state s_k and $\sigma_{s_k}^2$ is the variance of the sojourn time in the state s_k.[2]
  - Gamma:

$$d_{s_k}(u_k | \alpha_{s_k}, \beta_{s_k}) = \frac{1}{\Gamma(\alpha_{s_k}) \beta_{s_k}^{\alpha_{s_k}}} \cdot u_k^{(\alpha_{s_k} - 1)} \cdot e^{-\frac{u_k}{\beta_{s_k}}}, \tag{6}$$

  where $u_k, \alpha_{s_k}, \beta_{s_k} > 0$; $\alpha_{s_k}, \beta_{s_k}$ are the shape and scale for the state s_k, respectively.
- The non-parametric probability mass function is represented as a matrix $N \times M$, where N is the number of states and M the number of discretized values.

Then, we define, $d^*_{s_k}(u_k)$ as the survivor function of the sojourn time.

$$d^*_{s_k}(u_k) = \sum_{v \geq u_k} d_{s_k}(v). \tag{7}$$

We also define r as the total number of state transitions that occurred until time n; and finally, we define $N(t)$ as the current state at time t.

The likelihood of a state sequence S of a HSMM is then given by:

$$\begin{aligned} P(X, S, \Theta) =& \pi_1 \left\{ \prod_{k=2}^{r-1} \gamma_{s_{k-1} s_k} d_{s_k}(u_k) \right\} \gamma_{s_{r-1} s_r} d^*_{s_r}(u_r) \\ & \cdot \prod_{l=1}^{n} P_{s_{N(l)}}(x_l), \end{aligned} \tag{8}$$

[2]Note that λ_{s_k} and $\sigma_{s_k}^2$ are chosen so that the probability that the corresponding sojourn time is negative is negligible.

where s_k is the k^{th} visited state and u_k is the sojourn time of the k^{th} state. Therefore, a HSMM is specified by the quadruple $\Theta = \{\pi_1, \Gamma, B, D\}$ [12].

C. Initializing the parameters of HMM and HSMM

We use a subject dependent approach, meaning that we train and test with mutually exclusive heart beats of each subject. Parameters are initialized in the training phase, using annotated samples from a given subject. During the testing phase we further optimize our parameters by using different non-annotated samples from the subject. We used exhaustive cross-validation from 1 to 7 training heart beats but we constrained our training sets to those that produce continuous test set. For example, if we train with heart beats $1, 2, 8$ we test with heart beats $3, 4, 5, 6, 7$.

For both HMMs and HSMMs, the initial states distribution (π_1) are initialized with equal starting probabilities. The Γ parameters are fixed to:

$$\Gamma = \begin{array}{c} \\ S1 \\ siSys \\ S2 \\ siDia \end{array} \begin{array}{c} \begin{array}{cccc} S1 & siSys & S2 & siDia \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array}, \tag{9}$$

because in a normal cardiac system the state sequence $\{S1 \rightarrow siSys \rightarrow S2 \rightarrow siDia \rightarrow S1\}$ is fixed. For the purposes of this paper we will ignore any skipped beats, extra sounds or murmurs.

To initialize B we use the annotated samples and compute the parameters $\mu_s, \sigma_s \forall s \in S$ by using the corresponding maximum likelihood estimators.

To compute the initial parameters D for the:

- Parametric sojourn time distributions
  - Poisson: $\lambda^0_{s_k} \forall s_k \in \mathcal{S}$, we use the average annotated time lapse between the beginning and the end of the corresponding state s_k.
  - Gaussian: $\lambda^0_{s_k}, \sigma^0_{s_k} \forall s_k \in \mathcal{S}$, we use the average and standard deviation annotated time lapse between the beginning and the end of the corresponding state s_k, respectively.
  - Gamma: $\alpha^0_{s_k}, \beta^0_{s_k} \forall s_k \in \mathcal{S}$, we use the maximum likelihood estimator (MLE) by Choi over the annotated events [13].
- Non-parametric probability mass function is initialized as uniformly distributed $U(a^0_{s_k}, b^0_{s_k})$ where: $a^0_{s_k} \forall s_k \in \mathcal{S}$ is the minimum sojourn time annotated for state s_k; and $b^0_{s_k} \forall s_k \in \mathcal{S}$ is the maximum sojourn time annotated for state s_k.

D. Optimizing the HSMM parameters

The parameters Θ are tuned using the expectation maximization (EM) method [11]. This maximizes the likelihood (8) by iterating over the following steps:

- **Expectation (E-step)**: The conditional expectations of the missing data given the observations and given the current estimate of Θ are computed. Namely, the forward vector α_t [14] defined as:

$$\alpha_t^{s_k} = P(S_{t+1} \neq s_k, S_t = s_k | X_0^t = x_0^t), \tag{10}$$

and the backward vector β_t [14] defined as:

$$\beta_t^{s_k} = \frac{P(S_{t+1} \neq s_k, S_t = s_k | X_0^T = x_0^T)}{P(S_{t+1} \neq s_k, S_t = s_k | X_0^t = x_0^t)}, \tag{11}$$

where we have used the compact notation X_0^t to denote the observations sub-sequence $X_0, \ldots, X_t$.
Using these fundamental quantities, it is possible to compute expected number of times η_{iu_k} that the model remains in state i for u_k time steps. The η_{iu_k} is defined as:

$$\begin{aligned} \eta_{iu_k} =& P(s_{u_k} \neq i, s_{u_k - v} = i, v = 1, ..., u_k | X, \Theta) + \\ & \sum_{t=1}^{n} P(s_{t+u_k+1} \neq i, s_{t+u_k-v} = i, \\ & v = 0, ..., u_k - 1, s_t \neq i | X, \Theta). \end{aligned} \tag{12}$$

- **Maximization (M-step)**: Maximize, with respect to Θ, the likelihood (8). The D parameters are re-estimated, according to the distribution used:
  - Poisson:

$$\hat{\lambda}_{i,\beta} = \sum_{v=1}^{n} \frac{\eta_{iv}}{\eta_i} (v - \beta), \tag{13}$$

where β is all possible shifts parameters, $\beta = 1, \ldots, \min(u_k : \eta_{iu_k} > 0)$.
  - Gaussian:

$$\hat{\sigma}^2_{i,\beta} = \sum_{v=1}^{n} \frac{\eta_{iv}}{\eta_i} (\lambda_{i,\beta} - v)^2, \tag{14}$$

The maximization step chooses the β, equation (14) and (13), that maximizes likelihood in equation (8).
  - Gamma: The $\hat{\alpha}_i, \hat{\beta}_i$ are obtained as in [12].
  - Non-parametric probability mass function: In the maximization step for the non-parametric probability mass function we re-estimate the matrix D as:

$$\hat{d}_i(u_k) = \frac{\eta_{iu_k}}{\sum_{v=1}^{n} \eta_{iv}}. \tag{15}$$

The emission parameters B ($\hat{\lambda}_{s_k}, \hat{\sigma}_{s_k} \forall s_k \in S_k$) are re-estimated as:

$$\hat{\lambda}_{s_k} = \frac{\sum_{t=1}^{n} \alpha_t^{s_k} x_t}{\sum_{t=1}^{n} \alpha_t^{s_k}}, \tag{16}$$

and

$$\hat{\sigma}^2_{s_k} = \frac{\sum_{t=1}^{n} \alpha_t^{s_k} (x_t - \hat{\lambda}_{s_k})^2}{\sum_{t=1}^{n} \alpha_t^{s_k}}. \tag{17}$$

Finally, the initial transition probabilities are re-estimated as: $\hat{\pi}_{s_k} = \alpha_0^{s_k}$.

E. Decoding algorithm

In this paper, we use the Viterbi algorithm [15] to determine the hidden state sequences corresponding to heart beat components. We recall that the Viterbi algorithm does not attempt to classify every observation sample separately, but instead, it returns the hidden state sequence that maximizes the likelihood function reported in (2) for HMM and in (8) for HSMM.

F. Computing a confidence metric point-by-point

Not all sample classifications have the same degree of confidence. For example, samples between states are harder to classify, since it is a hard task to identify the exact location where one state ends and another begins. Similarly, high level noise could be easily misinterpreted as heart sounds ($S1$ or $S2$) because of its high amplitude in the homomorphic envelogram. These samples should have inherently low confidence in their classifications. On the other side, samples in the middle of states have very high amplitude and are easier to classify, providing to their classifications a higher confidence. We propose a measure of sample confidence based on the conditional probability distribution P_r[3].

The distribution of X_t conditioned on all observations $X_{\setminus t} = (X_1, \ldots, X_{t-1}, X_{t+1}, \cdots, X_T)$ is given by:

$$P_r = P(X_t = x_t | X_{\setminus t} = x_{\setminus t}) = \sum_{i \in S_k} ln\left(\epsilon_t^i\right) \cdot e^{\zeta_t^i}, \qquad (18)$$

where $\epsilon_t^i = \frac{e^{-(\alpha_t^i \times \Gamma + \beta_t^i)}}{\sum_{j \in S_k} e^{-(\alpha_t^j \times \Gamma + \beta_t^j)}}$, $\zeta_t^i = \frac{-ln(p^i(x_t))}{\sum_{k=1}^{T} ln(p^i(x_k))}$ and $p^i(x_t)$ is the associated probability of observing x_t in the state i. The exponentiation of the α, β and the logarithms of $p^i(x_t)$ are used in order to reduce the chance of underflow and overflow respectively.

III. Methology

A. Materials

The DigiScope dataset is composed of samples from 29 different healthy individuals, ranging in age from six months to 17 years old. The recordings have a minimum, maximum and average duration of $\approx 2, 20$ and 8 seconds, respectively. This is a very challenging dataset given the highly varying heart rates of individuals in this age range. A dataset with healthy adults is potentially easier to process, given its heart rate stability and the full maturity of the heart, which motivated us to focus on this pediatric dataset. Heart sounds have been collected in Real Hospital Português (Recife, Brasil) using a Littmann 3200 stethoscope embedded with the DigiScope Collector [16] technology, recorded at 4000 Hz. The heart sounds have all been collected from the mitral spot. These sounds were then manually annotated by cardiopulmonologists using the audacity software[4]. In the annotations we have the beginning and ending stages of $S1$ and $S2$ during a variable number of heart cycles.

B. Pre-processing

Following previous literature [3], [6], [17], the system first normalizes and scales the signal to the $[0, 1]$ range. The scaled signal is filtered using a Butterworth lowpass filter of order 10 with a cutoff frequency of 100 Hz, since the majority of the frequency content of the $S1$ and $S2$ (for the DigiScope dataset) is contained in the range $30 - 80$ Hz, as it shown in Figure 1. Similar pre-processing methods are also used in [18].

[3]For the rest of the paper we shorthand $P(X_t = x_t | X_{\setminus t} = x_{\setminus t})$ as P_r.

[4]www.audacityteam.org

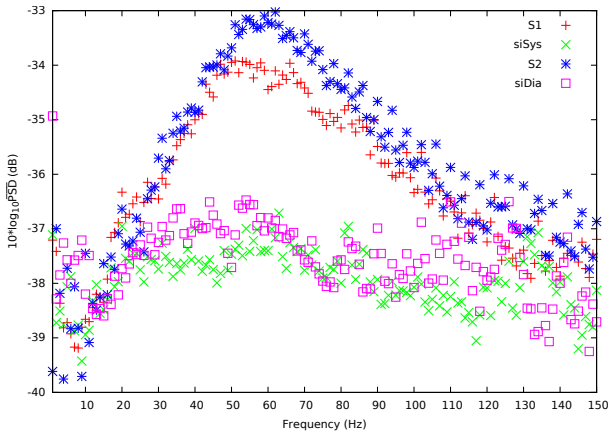

Fig. 1. Average power spectral density (PSD) for each state over the frequency range $[0, 150]$ Hz. The $S1$ peak is ≈ 50 Hz and the $S2$ peak is ≈ 60 Hz.

From the filtered signal, the homomorphic envelogram is computed as in [6]. In our previous work [17], we experimented several different envelograms and confirmed that the homomorphic envelogram is suitable for pediatric heart sound signals.

C. Performance metrics

The performance of the HSMM and HMM was measured as the model's capacity to recreate the state sequence annotated by the cardiacpulmonologists. We first compute the positive predictability per sample (P^+_{sample}) as:

$$P^+_{sample} = \frac{TP_{samples}}{TP_{samples} + FP_{samples}}, \qquad (19)$$

where $TP_{samples}$, $FP_{samples}$ is the count of true and false positive samples respectively. A sample at time t is a true positive when the predicted state sample and the annotated state sample are the same ($s_t^{model} = s_t^{expert}$), otherwise it is a false positive[5].

Another metric adopted is the positive predictability per state (P^+_{state}) which is computed as in (19) where TP_{states}, FP_{states} is the count of true and false positive states respectively. A classification is a true positive when the model's state is equal to the closest expert's annotated state. Otherwise it is considered a false positive. In particular, we define $B(i) = \underset{k \leq i}{\text{argmin}}(S_k = S_i)$ as the beginning time of state s_i. We also define the ending time of a state s_i as $E(i) = \underset{k \geq i}{\text{argmax}}(S_k = S_i)$. The middle time of a state s_i is $M(i) = \frac{B(i)+E(i)}{2}$. Finally, we count true positives as:

$$TP_{states} = \sum_{j \in M^{expert}} \underset{i \in M^{model}}{\text{argmin}} \left(|i - j|\right) \wedge s_i^{model} = s_j^{expert}. \qquad (20)$$

For the sake of completeness and in order to allow a fair comparison with other results from the literature about PCG

[5]Note that, when computing positive predictability true/false negatives are not defined.

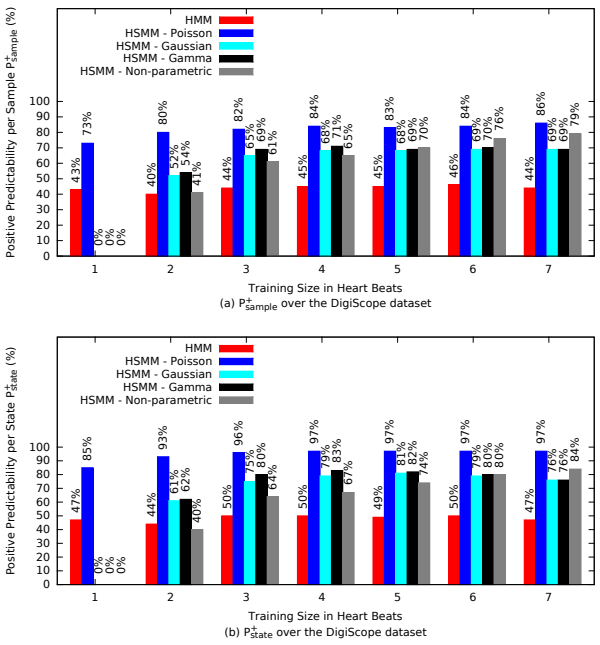

Fig. 2. Subject dependent results. Average positive predictability (a) per sample P^+_{sample} and (b) per state P^+_{state} for the tested HMM, HSMM models over the DigiScope dataset.

segmentation, we also report performance metrics related to the capacity of the systems of detecting the precise position of the principal heart sounds $S1$ and $S2$. In this case true, false positives (TP_S, FP_S) and true, false negatives (TN_S, FN_S) will be determined according to a definition akin to that used to define positive predictability per state. The performance metrics considered for principal heart sounds detections are:

$$\text{Precision} = \frac{TP_S}{TP_S + FP_S} \quad (21)$$

$$\text{Recall} = \frac{TP_S}{TP_S + FN_S} \quad (22)$$

$$\text{F-measure} = 2\,\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (23)$$

$$\text{Accuracy} = \frac{TP_S + TN_S}{TP_S + TN_S + FP_S + FN_S}. \quad (24)$$

IV. Results

We conducted experiments both with HMM and HSMM. The HMM is not as capable as the HSMM in detecting the right sequence and duration of states, as it can be seen in Figure 2, where the HMM average positive predictability per sample P^+_{sample} and per state P^+_{state} is considerably lower than the P^+_{sample} and P^+_{state} of any HSMM that we tested.

As it is shown in Figure 2, the HSMM using the Poisson distribution outperformed significantly the ones based on Gaussian, Gamma distributions and the non-parametric probability mass function for P^+_{sample} and P^+_{state}. Furthermore, we can see that the non-parametric probability mass function, while it starts with weak performance, it improves significantly

TABLE I
Performance of HSMM with Poisson sojourn time distribution in detecting $S1$ and $S2$.

	Training Size in Heart Beats						
	1	2	3	4	5	6	7
Precision	80%	91%	95%	96%	97%	97%	97%
Recall	96%	98%	99%	99%	99%	99%	99%
F-measure	87%	94%	97%	98%	98%	98%	98%
Accuracy	88%	95%	97%	98%	98%	98%	98%

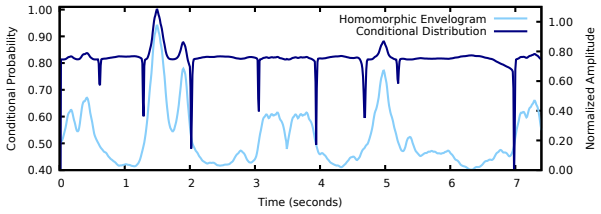

Fig. 3. The conditional distribution P_r generated by the HSMM using the Poisson distribution.

as the size of the training set increases.[6] For the reminder of the paper, we chose to use the Poisson distribution as an approximation of the sojourn time distribution, since it was shown to outperform the others.

Table I reports the performance obtained by the proposed HSMM approach in detecting the principal heart sounds when using the Poisson distribution to approximate sojourn times. Such results are expressed via the metrics described in Section III-C, for each cross-fold iteration. These results suggest that 4 heart beats are enough to capture the relevant information associated to a given recording.

Finally, we report results about the confidence metric function described in Section II-F. Figure 3 shows an example of the proposed confidence metric function as an overlay over the homomorphic envelogram. One can notice that the conditional distribution exhibits sudden low peaks around the transitions between different states. Furthermore, we notice that noise has lower probabilities compared to waveforms corresponding to heart sound segments.

Motivated by the above observations, we used the samples from all Poisson cross-fold iterations in order to compute the P^+_{sample} as a discrete function of P_r, the results are presented in Figure 4. Furthermore, one can notice from the plot that the majority of the samples follow a linear trend. In Figure 4, the circles are centered around P^+_{sample} and the color intensity (from low to high) indicates the number of samples (from few to many, respectively). Furthermore, using the nonlinear least-squares (NLLS) Marquardt-Levenberg algorithm with a first degree polynomial, we get the following regression function:

$$RP^+(x_t) = 0.98 \cdot P_r - 0.10, \quad (25)$$

with a weighted Pearson product-moment correlation coefficient (WPCC) of 0.93. The regression line is presented in

[6]We also tested the logarithmic distribution which performed worse than all the other distributions tested for HSMMs; for the sake of brevity, we excluded the corresponding results from this paper.

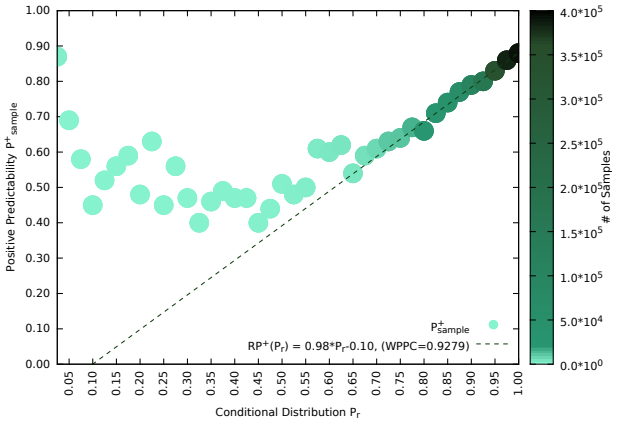

Fig. 4. Relationship of the conditional distribution P_r with positive predictability P^+_{sample}, RP^+, EP^+ and confidence P_c in our dataset.

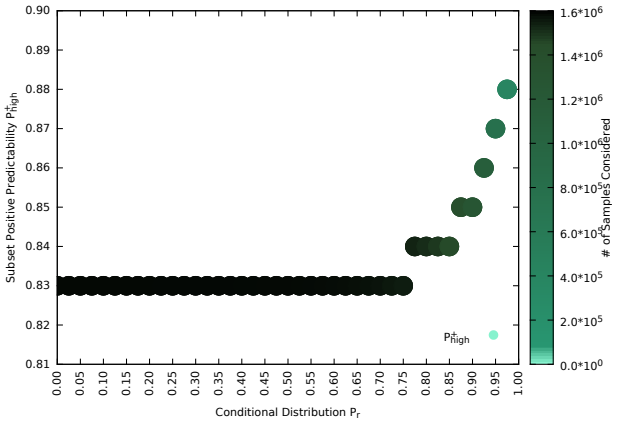

Fig. 5. Relationship of the conditional distribution P_r and the subset positive predictability which is obtained using a high threshold in our dataset.

Figure 4. For a $P_r \lesssim 0.50$, we do not have enough data to withdraw any conclusions, although we can safely argue, that in our dataset, the conditional distribution $P_r \gtrsim 0.50$ gives a good estimate of our P^+_{sample}.

Furthermore, we define P^+_{high} as:

$$P^+_{high} = \frac{TP^{high}_{samples}}{TP^{high}_{samples} + FP^{high}_{samples}}, \tag{26}$$

where $TP^{high}_{samples}$, $FP^{high}_{samples}$ are the correctly and wrongly classified samples respectively, where their conditional probability P_r is above a threshold. By setting such high thresholds and computing the P^+_{high} we plot Figure 5. For a high threshold of $P_r \geq 0$ we consider all data to our computation and we computed a $P^+_{high} \simeq 0.83$. By setting thresholds we observe that we increase our positive predictability to almost 90%. From Figure 5 we conclude, that by setting a high threshold in P_r, we can still select the majority of the sampling points, and at the same time, be more selective and confident with respect to set classification.

V. Conclusion

In this paper, we used HSMMs to decode the "true" state sequence of events in a PCG signal. We observed, that the HSMM always outperforms the HMM. We use an ensemble of distributions to approximate the sojourn time distribution. Our experiments using a subject dependent approach, showed that the Poisson clearly outperformed the Gaussian and Gamma distribution and the non-parametric probability mass function. We concluded that using information concerning the sojourn time distribution in each state is a compulsory step when modeling heart sound signals. Furthermore, we presented a novel way (based on conditional probability) to compute a confidence metric for sample classifications.

References

[1] J. E. Hall and A. C. Guyton, *Textbook of medical physiology*, 12th ed. Philadelphia, Pa.: Saunders/Elsevier, 2011.

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

[3] C. N. Gupta, R. Palaniappan, S. Swaminathan, and S. M. Krishnan, "Neural network classification of homomorphic segmented heart sounds," *Appl. Soft Comput.*, vol. 7, no. 1, pp. 286–297, 2007.

[4] T. Leung, P. White, W. Collis, E. Brown, and A. Salmon, "Classification of heart sounds using time-frequency method and artificial neural networks," in *IEEE EMBC*, 2000, pp. 988–991.

[5] J. Vepa, "Classification of heart murmurs using cepstral features and support vector machines," in *IEEE EMBC*, 2009, pp. 2539–2542.

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

[7] Y.-J. Chung, *Pattern Recognition and Image Analysis, Iberian Conference*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, ch. Classification of Continuous Heart Sound Signals Using the Ergodic Hidden Markov Model, pp. 563–570.

[8] S. Schmidt, E. Toft, C. Holst-Hansen, C. Graff, and J. Struijk, "Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent hidden-markov model," in *Computers in Cardiology*, 2008, pp. 345–348.

[9] S. zheng Yu, "Hidden semi-Markov models," *Artificial Intelligence*, 2010.

[10] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Support vector machine hidden semi-markov model-based heart sound segmentation," in *Computing in Cardiology Conference (CinC), 2014*, Sept 2014, pp. 625–628.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.

[12] J. OĆonnell and S. Højsgaard, "Hidden semi markov models for multiple observation sequences: The mhsmm package for r," *Journal of Statistical Software*, vol. 39, no. 1, pp. 1–22, 2011, https://cran.rproject.org/web/packages/mhsmm/mhsmm.pdf.

[13] S. C. Choi and R. Wette, "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias," *Technometrics*, vol. 11, no. 4, pp. 683–690, 1969.

[14] Y. Guédon, "Estimating Hidden Semi-Markov Chains from Discrete Sequences," *Journal of Computational and Graphical Statistics*, vol. 12, no. 3, pp. 604–639, 2003.

[15] G. D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[16] D. Pereira, F. Hedayioglu, R. Correia, T. Silva, I. Dutra, F. Almeida, S. Mattos, and M. Coimbra, "DigiScope - Unobtrusive collection and annotating of auscultations in real hospital environments," in *IEEE EMBC*, 2011, pp. 1193–1196.

[17] J. Oliveira, T. Mantadelis, and M. Tavares Coimbra, "Why should you model time when you use markov models for analysing heart sounds," in *IEEE EMBC*, 2016.

[18] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-hsmm-based heart sound segmentation." *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2016.