



Apnea bradycardia detection based on new coupled hidden semi Markov model

Nasim Montazeri Ghahjaverestan, Mohammad Bagher Shamsollahi, Di Ge,
Alain Beuchée, A I. Hernández

► To cite this version:

Nasim Montazeri Ghahjaverestan, Mohammad Bagher Shamsollahi, Di Ge, Alain Beuchée, A I. Hernández. Apnea bradycardia detection based on new coupled hidden semi Markov model. Medical and Biological Engineering and Computing, 2021, 59 (1), pp.1-11. 10.1007/s11517-020-02277-8 . hal-03033759

HAL Id: hal-03033759

<https://hal.science/hal-03033759>

Submitted on 18 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Apnea Bradycardia detection based on new Coupled Hidden semi Markov Model

Nasim Montazeri Ghahjaverestan
Mohammad Bagher Shamsollahi Di Ge Alain Beuchée
Alfredo I. Hernández

Received: July 4-2020 / Accepted: October 12-2020 in Medical
& Biological Engineering & Computing (MBEC)

Abstract

In this paper, a method for apnea bradycardia detection in preterm infants is presented based on coupled hidden semi Markov model (CHSMM). CHSMM is a generalization of Hidden Markov Models (HMM) used for modeling mutual interactions among different observations of a stochastic process through using finite number of hidden states with corresponding resting time. We introduce a new set of equations for CHSMM to be integrated in a detection algorithm. The detection algorithm was evaluated on a simulated data to detect a specific dynamic and on a clinical dataset of electrocardiogram signals collected from preterm infants for early detection of apnea bradycardia episodes. For simulated data, the proposed algorithm was able to detect the desired dynamic with sensitivity of 96.67%, and specificity of 98.98%. Furthermore, the method detected the apnea bradycardia episodes with 94.87% sensitivity and 96.52% specificity with mean time delay of 0.73 seconds. The results show that the algorithm based on CHSMM is a robust tool for monitoring of preterm infants in detecting apnea bradycardia episodes.

Key Words: Coupled Hidden Semi Markov Model (CHSMM), Coupled hidden Markov model (CHMM), Forward-Backward (FB) algorithm, Electrocardiography (ECG), Apnea Bradycardia (AB).

1 Introduction

Apnea of prematurity is a common complication in preterm infants born prior 37 weeks of gestation. Apnea of prematurity is associated with repetitive long (>15 -20 sec) interruptions in respiratory airflow, which cause drop in blood oxygen saturation and occurrence of an arrhythmia called bradycardia characterized by reduction in heart rate [1]. According to American academy of pediatrics, an episode of apnea of prematurity is defined as a pause in breathing for 20 seconds or more, or shorter accompanied by bradycardia (less than 100 beats per minutes) [2]. Monitoring and treatment of apnea bradycardia require hospitalization that increases healthcare costs. If not treated, apnea bradycardia leads to higher risk of infant mortality [3] and in case of survival, can cause adverse neurobehavioral deficits during childhood [4]. Monitoring of preterm infants in neonatal intensive care unit (NICU) and early detection of apnea episodes prevent the fatal cardiorespiratory and neurobehavioral outcomes of apnea bradycardia.

The detection of apnea bradycardia has been investigated in a few studies. The primary approach for apnea bradycardia detection was to use a fixed or an adaptive threshold to mark drop in heart rate [5,6]. The thresholding approach was further improved by developing algorithms for detecting abrupt changes in heart rate [7]. However, these methods detect the apnea episodes with delay and are sensitive to the amplitude of heart rate. The heart rate signal can be affected by motion artifacts and noises that generate false alarm. Thus, recent studies has improved the accuracy of apnea bradycardia detection through using algorithms that learn the dynamic of cardiac activity. The learning algorithm extensively investigated to detect apnea bradycardia is Hidden Markov Model (HMM) [8–10].

HMM represents the dynamics of a system using a set of hidden states [11]. Transition to a state is related to the previous state and the transition probabilities. At each time instant, system transits to a state and generates an observation, based on a probability density function (Fig. 1(a)). Several generalizations have been proposed for HMM. For example, the behavior of a system can be explained by a Hidden semi Markov Model (HSMM) [12,13], in which the system can rest in a state for several time instants (resting time) (Fig. 1(b)). Alternatively, systems with several interacting components can be modeled by coupled hidden Markov model (CHMM) [14,15]. In CHMM, it is assumed that each observation recorded from the system is generated by a component, which can be modeled by a set of hidden states. In each

component, the transition from one state to another depends on its state transition probabilities as well as the current states of other components (Fig. 1(c)).

In our previous work [16], we presented a new frame of equations for implementing the supervised learning of model parameters and classifying (inference) of unknown observations based on CHMM assumptions. We evaluated that model in detecting the apnea bradycardia episodes. For that purpose, from single channel of electrocardiogram (ECG), we extracted three observations (features) including the distance between two consecutive R peaks, and the width and amplitude of QRS complex. We hypothesized that there were distinct dynamics and interactions among these observations during normal breathing compared to apnea bradycardia episodes modeled by two different CHMMs. By comparing the inference of the two CHMMs, the proposed detection algorithm was able to accurately detect segments of data, in which apnea bradycardia episode was occurred. Nevertheless, the structure of the proposed CHMM can be improved by considering HSMM assumptions. Incorporating the resting time for the hidden states reduces the sensitivity of the algorithm to noises and artifacts. Moreover, each apnea bradycardia can last more than 15 seconds, which means the dynamic of the model does not change at every time instants. Thus, the performance of the detection algorithm based on CHMM can be improved by adding another hyperparameter corresponding to the resting time of hidden states. The new model is the combination of CHMM and HSMM called Coupled Hidden semi Markov Model (CHSMM).

CHSMM was proposed for the first time by Natarajan et al. [17]. They analyzed the sequence of data in a set of segments defined as *bricks*. Each *brick* was shorter sequence of data generated by one hidden state. The resting time of each state was modeled with the length of the *brick*. However, implementing the CHSMM using the *bricks* adds more complexity and computational costs to the algorithm.

The main goal of this paper was to present a simple set of equations for CHSMM that can be implemented with less computational costs to detect apnea bradycardia. The proposed structure and equations of CHSMM is presented in section II. Section III details the detection algorithm based on CHSMM. The performance of the detection algorithm was evaluated using a simulated dataset as well as the clinical data to detect apnea bradycardia in preterm infants. The performance of the CHSMM is compared to other HMM-based methods in section IV. Finally, section V concludes the paper.

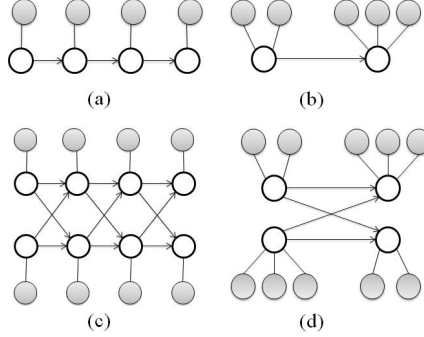


Figure 1: Structures of: a) HMM, b) HSMM, c) CHMM, d) CHSMM. Empty circles represent hidden Markov states, while filled circles show the observed time series of features.

2 Methods

2.1 CHSMM structure

In the proposed CHSMM, each dimension of observations (channel) is modeled by a Markovian process with a set of hidden states. The state transition depends on both the previous state in that channel and the states of other channels. The inter-channel coupling relationship can be simplified using a product format [18]:

$$P(q_t^\varsigma | q_{t-1}^1, q_{t-1}^2, \dots, q_{t-1}^C) = \prod_{c=1}^C P(q_t^\varsigma | q_{t-1}^c) \quad (1)$$

where q_t^ς is the current state in channel ς at time t and C is the total number of channels (equal to the number of observations). q_t^ς can be one of the hidden states, $S_1^\varsigma, S_2^\varsigma, \dots, S_{M(\varsigma)}^\varsigma$ in channel ς with $M(\varsigma)$ as the maximum number of states. Accordingly, $a_{nm}^{\varsigma c} = P(q_t^\varsigma = S_m^\varsigma | q_{t-1}^c = S_n^c)$ is defined as the effect of channel c being in state n at time $t-1$ on the transition to state m in channel ς at time t . Furthermore, π_m^c denotes the initial probability of state m at $t=1$. For the rest of the paper, the superscript of the state variable is omitted due to simplicity.

For a given channel ς , after each transition, it remains for d samples of time before the next transition with probability of $p_m^\varsigma(d) = P(q_{t:t+d-1}^\varsigma = S_m^\varsigma)$, for $d \in \{1, 2, \dots, D(\varsigma)\}$ with $D(\varsigma)$ as the maximum resting time. For

simplicity, we use the notation $v_t^c(m, d)$ for the event $q_{t:t+d-1}^c = S_m$ before the next transition.

For each time instant $t = \{1, 2, \dots, T\}$, $o_t = \{o_t^1, \dots, o_t^C\}$ is the set including observations of all channels, where o_t^c is the sample of observation at time t in channel c . The conditional emission probability as in Fig.1 writes $b_m(o_t^c) = P(o_t^c | q_t^c = S_m)$. Based on this structure, a CHSMM is specified by its model parameters, $\lambda = \{\{\pi_m^c\}, \{p_m^c(d)\}, \{a_{nm}^{cs}\}, \{b_m(o_t^c)\}\}$. All the probabilities in the following parts are conditioned on the model parameters that will be omitted for the rest of the paper for simplicity.

We summarise here the three assumptions in the current CHSMM. First, it is assumed that the observed data of each channel, o_t^c , depends only on state, q_t^c . Second, inter-channel influences are independent as in (1). Third, inter-channel state influences only occur upon transition. In the following, the equations related to model parameter re-estimation and inference will be derived by generalizing the coupling version of these methods introduced in our previous work on CHMM [16], we presented a coupling version of these equations, which are generalized by considering the parameter related to resting time in this paper.

2.1.1 Forward and Backward Recursions

To establish learning of the model parameters and inferring the unknown observations in Markovian models, *Forward* and *Backward* variables are required to be defined and calculated, recursively.

Forward Parameter: It is defined as the probability of being in state m of channel c between time t to $t + d$ subjected to the observations of all channels from the beginning to time x .

$$\alpha_{t|x}^c(m, d) \triangleq P(v_t^c(m, d) | o_{1:x}) \quad (2)$$

The *predicted*, *filtered* and *smoothed* definition of the Forward variable can be formed using $x = t - 1, t$ and T , respectively. According to [12], since the Markovian chain either remains in the current state or transits to another one, we can write:

$$\alpha_{t|t-1}^c(m, d) = \mathcal{A}_t^c(m, d) + \mathcal{B}_t^c(m, d). \quad (3)$$

where $\mathcal{A}_t^c(m, d) = P(v_t^c(m, d), \tau_{t-1}^c > 1 | o_{1:t-1})$ for no transition at t and $\mathcal{B}_t^c(m, d) = P(v_t^c(m, d), \tau_{t-1}^c = 1 | o_{1:t-1})$ when the transition occurs. Note

that τ_t^ς defines the remaining time of the current state (q_t^ς) before transition to another state or choose to return to the current state at time $t + \tau_t^\varsigma$. Thus, when $\tau_{t-1}^\varsigma > 1$, no transition occurs at t while $\tau_{t-1}^\varsigma = 1$ a transition is forced.

These variables can be calculated as follows:

$$\mathcal{A}_t^\varsigma(m, d) = \alpha_{t-1|t-2}^\varsigma(m, d+1) \tilde{b}_m^\varsigma(o_{t-1}) \quad (4)$$

$$\begin{aligned} \mathcal{B}_t^\varsigma(m, d) &= p_m^\varsigma(d) \sum_{n_\varsigma=1}^{M(\varsigma)} \{a_{n_\varsigma m}^{\varsigma\varsigma} \alpha_{t-1|t-2}^\varsigma(n_\varsigma, 1) \tilde{b}_{n_\varsigma}^\varsigma(o_{t-1})\} \\ &\times \prod_{c \neq \varsigma}^C \sum_{n_c, d_c} \{a_{n_c m}^{c\varsigma} \alpha_{t-1|t-2}^c(n_c, d_c) \tilde{b}_{n_c}^c(o_{t-1})\}. \end{aligned} \quad (5)$$

where

$$\tilde{b}_m^\varsigma(o_t) \triangleq \frac{\alpha_{t|t}^\varsigma(m, d)}{\alpha_{t|t-1}^\varsigma(m, d)} \quad (6)$$

By considering the initialization of this recursion as: $\alpha_{1|0}^\varsigma(m, d) = \pi_m^\varsigma p_m^\varsigma(d)$, the *Forward* variable can be calculated, recursively.

Backward variable: It is defined as the ratio of the smoothed over predicted version of *Forward* variable:

$$\beta_t^\varsigma(m, d) \triangleq \frac{\alpha_{t|T}^\varsigma(m, d)}{\alpha_{t|t-1}^\varsigma(m, d)}. \quad (7)$$

In case of no transition, *Backward* variable can be calculated using:

$$\beta_t^\varsigma(m, d) = \tilde{b}_m^\varsigma(o_{t-1}) \beta_{t+1}^\varsigma(m, d-1), d \neq 1, \quad (8)$$

and if there is a transtion ($d = 1$), the current state $v_t^\varsigma(m, 1)$ transits to $v_{t+1}^\varsigma(n_\varsigma, d_\varsigma)$. Thus, the *backward* variable can be recursively updated as follows:

$$\begin{aligned} \beta_t^\varsigma(m, 1) &= \tilde{b}_m^\varsigma(o_t) \sum_{n_\varsigma, d_\varsigma} \{ \beta_{t+1}^\varsigma(n_\varsigma, d_\varsigma) a_{mn}^{\varsigma\varsigma} p_{n_\varsigma}^\varsigma(d_\varsigma) \\ &\prod_{c \neq \varsigma}^C \sum_{n_c, d_c} a_{n_c n}^{c\varsigma} p_{n_c}^c(d_c) \alpha_{t|t-1}^c(n_c, d_c) \tilde{b}_{n_c}^c(o_t) \} \end{aligned} \quad (9)$$

with the initial value $\beta_t^\varsigma(m, 1) = \tilde{b}_m^\varsigma(o_T)$ for all values of d . See Supplementary document section I for the proof of these equations.

2.1.2 Defining more variables

In addition to *Forward* and *Backward* variables, four more variables are required for the implementation of model parameters learning (re-estimation). The first variable is the smoothed *Forward* variable as $\alpha_{t|T}^\varsigma(m, d) = \alpha_{t|t-1}^\varsigma(m, d)\beta_t^\varsigma(m, d)$. By summing $\alpha_{t|T}^\varsigma(m, d)$ over d , we can define $\gamma_t^\varsigma(m)$. Furthermore, the probability of transition to state m at time t and resting till $t+d$ given the previous observations:

$$\begin{aligned}\delta_t^\varsigma(m, d) &\triangleq P(\tau_{t-1}^\varsigma = 1, v_t^\varsigma(m, d) | o_{1:t-1}) \\ &= \beta_t^\varsigma(m, d)p_m^\varsigma(d)\mathcal{B}_t^\varsigma(m, d).\end{aligned}\quad (10)$$

And finally, we introduced, the smoothed probability for channel ς to transit to state m while the channel c is in state n given the observation of all channels:

$$\begin{aligned}\Upsilon_t^{c\varsigma}(n, m) &\triangleq \sum_{d=1}^{D(\varsigma)} P(q_{t-1}^c = S_n, v_t^\varsigma(m, d), \tau_{t-1}^\varsigma = 1 | o_{1:T}) \\ &= \sum_{d=1}^{D(\varsigma)} \beta_t^\varsigma(m, d) a_{mn}^{c\varsigma} p_m^\varsigma(d) \sum_{d'=1}^{D(c)} \alpha_{t-1|t-2}^c(n, d') \tilde{b}_n^c(o_{t-1}).\end{aligned}\quad (11)$$

For more information on the extraction of the variables see Supplementary document section II.

2.2 Re-estimation

The closed form of re-estimation equations to learn model parameters are summarized as follows:

$$\begin{aligned}\bar{a}_{m'm}^{c'c} &= \frac{\sum_{t=1}^T \Upsilon_t^{c'c}(m', m)}{\sum_{t=1}^T \sum_{m''=1}^{M(c)} \Upsilon_t^{c'c}(m', m'')}, \\ \bar{\pi}_m^c &= \frac{\sum_{d=1}^{D(c)} \alpha_{1|T}^c(m, d)}{\sum_{m'=1}^{M(c)} \sum_{d=1}^{D(c)} \alpha_{1|T}^c(m', d)}, \\ \bar{p}_m^c(d) &= \frac{\sum_{t=1}^T \delta_t^c(m, d)}{\sum_{t=1}^T \sum_{d'=1}^{D(c)} \delta_t^c(m, d')}, \\ \bar{\mu}_m^c &= \frac{\sum_{t=1}^T \gamma_t^c(m) o_t^c}{\sum_{t=1}^T \gamma_t^c(m)}, \quad \bar{\sigma}_m^c = \frac{\sum_{t=1}^T \gamma_t^c(m) (o_t^c - \bar{\mu}_m^c)^2}{\sum_{t=1}^T \gamma_t^c(m)}.\end{aligned}\quad (12)$$

The proof of these equations has been included in Supplementary document section III.

2.3 Inference

Knowing the parameters of a model, an unknown observation can be inferred through finding the optimal sequence of states for each of the channels. For inference, $\delta_t^c(m, d)$ is calculated for $t = 1, 2, \dots, T$ over all states and durations in all channels. Then, starting from $t_0 = 1$, the best m^* and d^* which maximize $\delta_{t_0}^c(m, d)$ are found and the optimal state sequence is assigned as $Q_{max}^c(t_0 : t_0 + d) = m^*$. The same analysis is repeated for the samples after $t = d^*$ to $t = T$.

2.4 Detection Algorithm

By establishing the equations for implementing re-estimation and inference, the CHSMM can be used in the detection algorithm, as explicitly described in [9, 16]. Briefly, for on-line detection of events, the data of the subjects were randomly categorized into train set and test set. The train set was segmented into intervals with duration of T_W samples. For each event type, a CHSMM was trained using segments starting from the onset of the event. In addition to the CHSMMs trained by the dynamics of different types of events, another CHSMM was trained using the segments void of all types of events. For simplicity, the detection algorithm is explained for two classes.

For each recording in the test set, a window of length T_W moves every samples (t) to choose a segment of data. Each channel of the chosen segment was analyzed and inferred separately by the two models, $\lambda_{1,2}$. By applying inference algorithm, a sequence of states (Q^c) was assigned to each channel of segment that was used for calculating *likelihood signals* defined as $ll_k^c(t) = \log(P(o_{t:t+T_W-1}^c | \lambda_k, Q^c))$ for each model ($\lambda_k, k = 1, 2$), where the inside argument can be computed using:

$$P(o_{t:t+T_W-1}^c | Q^c) = \prod_{t'=t}^{t+T_W-1} b_{Q^c(t')}(o_{t'}^c). \quad (13)$$

By comparing the *likelihood signals* of the two competing models, the segment was classified to class of the model with higher likelihood. To implement such

comparison, total log-likelihood of class 1 compared to class 2 (ll_{total}) can be defined as follows:

$$ll_{total}(t) = \sum_{c=1}^C (ll_1^c(t) - ll_2^c(t)). \quad (14)$$

An event corresponding to class 1 occurred at time t , if the following condition is satisfied:

$$ll_{total}(t) \geq \kappa \quad (15)$$

where κ is a constant threshold that was optimized using Receiver Operation Curves (ROC) [19] (for more information see 2.6).

2.5 Initialization

For each competing model ($\lambda_{1,2}$), parameters a_{nm}^{cs} and π_m^s were initialized based on uniform probabilities. To initialize the observation probability ($b_m(o_t^c)$), k-means clustering was applied to the observation of each channel. The number of clusters was assumed equal to the number of hidden states dedicated for the corresponding channel. Each cluster was characterized by a Gaussian distribution that initialized μ_m^c and σ_m^c .

The number of states for each channel was chosen based on Bayesian Information Criterion (BIC) algorithm [20], which was applied on the train set selected in the first round of cross-validation. BIC can be defined as:

$$BIC = -2 \log(L) + \nu \times \log(T), i = \{1, 2\}, \quad (16)$$

where L and ν are likelihood and the number of states, respectively. In this paper, for L the difference of average value of ll_{total} during event from that of normal status was calculated. Then, the number of states, which generated lower BIC was selected. BIC prevents overfitting through maintaining a trade-off between the complexity of the model quantified as the number of hidden states and the increase in the likelihood. the average value of the likelihood signal in each class and the total number of states used in each model were used to calculate BIC. Finally, a grid search is applied on specific intervals (see results section) and for each grid point, the BIC value was calculated. The configuration presenting the optimal BIC (minimum) value was selected.

2.6 Evaluation metrics

To evaluate the performance of the detection algorithm, we performed cross-validation with 5 folds. Sensitivity ($SEN = TP/(TP + FN)$) and specificity ($SPC = TN/(TN + FP)$) were calculated. TP , FP , FN and TN denote the number of true positives, false positives, true negatives and false negatives, respectively. TP occurs when a detection falls within a window with a length equal to the moving window and centered at a given annotation. The classification of the segments were compared to the reference annotations.

To optimize κ in (15), ROC and the area under the ROC curve (AUC) were tracted. In this curve, the optimum κ is the related treshold of the perfect detection (PD), defined as $PD = argmax\{SEN \times SPC\}$. Also, the distance to PD noted as $\sqrt{(1 - SEN)^2 + (1 - SPC)^2}$ [9] were calculated from the average ROC. Moreover, we compute the detection time delay (TD) [9, 16], defined as the time elapsed from the detected to annotated onsets. Also, for clinical data Positive Windows (PW) [16] was calculated as the ratio of the number of detections occurring before the annotated onset, over the total number of detected apnea bradycardia. The lower values of PW (below 50%) signifies the prediction of episodes.

2.7 Datasets

2.7.1 Simulated data

For evaluation of the proposed CHSMM a simulated data was generated from the Fitzhugh-Nagemo model [21]. It is a generic model consisting of two state variables, used to describe an excitable system (such as a neuron) from the following equations:

$$\begin{aligned}\frac{dv}{dt} &= 3(v - \frac{1}{3}v^3 + r + I) \\ \frac{dr}{dt} &= -\frac{1}{3}(v - a + 0.8r)\end{aligned}\tag{17}$$

where v is the membrane potential (fast variable), r is the recovery variable (slow variable), I is the value of the excitation (goes from 0 to 1 during excitation). The generated simulated data included rest condition and two excitations with 10 second duration, with different dynamics defined as a_1 , and a_2 . The difference between the two dynamics was in their morphologies, while the range of their amplitude were similar. For more information about

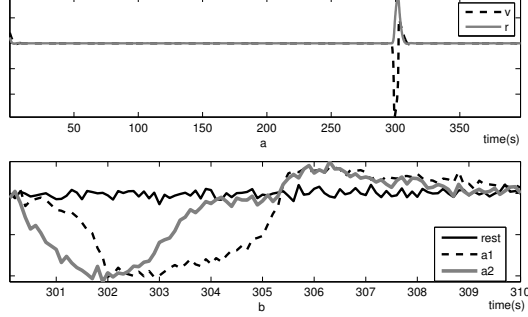


Figure 2: a: A generated simulated data by the FitzHugh-Nagumo model (state variables v and r), while a disturbance is applied during 300-305 s without noise. b: v dimension with 5 db additive noise of different dynamics. The dynamics (a_1 and a_2) are chosen from 300-310 s. The rest condition is taken from other intervals (200-210 s) and plotted for comparison. (y axis is the normalized amplitude.)

the simulation data, see [9, 16]. Same dataset has been used for evaluation of previously proposed HMM-based models. Thus, better comparison between CHSMM and other methods can be provided by the simulated dataset. Moreover, the generated data includes two patterns of excitations from rest status, which can assess the performance of the proposed CHSMM in detecting a dynamic from resting status as well as differentiating between two different dynamics.

Using the simulated data, the training data included 40 ten-second segments for each dynamic and rest condition. The segments corresponding to the dynamics were chosen from the beginning of the excitation. The test data were analyzed sample by sample using a moving ten-second window. $D(c)$, the maximum value of resting time in channel c , was considered as 5 samples of time for simulated data analysis. First, using a threshold on $ll_{total}^{\{a_1\}\{rest\}}$ (where $j \equiv \{rest\}$ and $i \equiv \{a_1\}$ according to (14)) a disturbance was detected. Then, second threshold on $ll_{total}^{\{a_1\}\{a_2\}}$ differentiated the dynamics a_1 and a_2 . Due to narrow range of changes in rest status, the number of hidden states was kept lower in the related model compared to that of the two dynamics. Accordingly, BIC was calculated for each member of the set, $\{rest-a_1-a_2|2-3-3, 2-4-4, 2-5-5, 2-6-6, 2-7-7, 3-4-4, 3-5-5, 3-6-6, 3-7-7, 4-5-5, 4-6-6, 4-7-7\}$, and the one with the lowest BIC was selected for each

model.

2.7.2 Real clinical data

Thirty-two preterm infants with birth weight of 1.23(1.07-1.36) kg, 31(29-32) weeks of gestational age and postnatal age of 12(7-20) days, hospitalized in NICU at the University Hospital of Rennes, France were participated in this study [9]. The study was approved by the ethical committee.

The parents of each included infant signed a consent form. From the participating infants, 148 single-channel ECG signals (400 Hz) with 26 ± 11 minutes length were recorded. These records were processed using a notch filter for removing 50 Hz noise. Then, Pan and Tompkins algorithm [22] was applied for R peak detection of ECG signals. Three features including the RR interval (RR), R-waveform amplitude (RAMP) and QRS duration (QRSd) were extracted from each detected heart beat, leading to a feature matrix of dimension $3 \times T$, where T is the length of the recording [23]. The extracted features were then upsampled using spline interpolation technique to 10 Hz. Fig. 3 depicts an example of these features before, during and after an apnea bradycardia episode.

One clinical expert manually annotated all bradycardia episodes from the whole database, leading to two annotated classes (bradycardia or not bradycardia). Overall, the database includes 233 annotated episodes of bradycardia with a mean duration of 21.5 ± 6.1 s. The rest of the data was considered into the "no bradycardia" class. Each infant presented between 0 to 39 (median of 4) bradycardia episodes.

The dataset was divided into train and test dataset based on 5-fold cross-validation. As described in [16], from train dataset, 30 seven-second segments starting from the apnea bradycardia onset and 30 seven-second segments void of events were randomly selected. The length of these segments (7 s) are chosen based on the average time measured from the beginning of the bradycardia to the peak RR value within the bradycardia episodes [9].

The effect of apnea can be observed in the RAMP feature before the start of the bradycardia event due to the modulation between the RAMP feature signal and respiratory activities [24]. Hence, there is a time delay between the RAMP and the other features, which can be compensated by a synchronization time delay (T_{sync}). According to the underlying physiological association between cardiac and respiratory activities [25], the delay was optimally chosen between 3 – 4.5 s for each HMM-based method. To choose

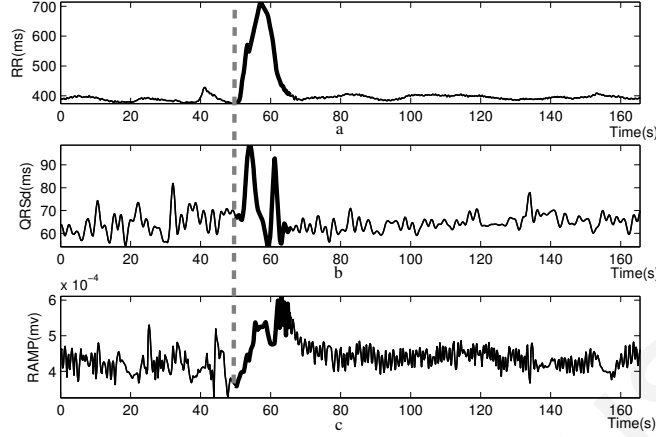


Figure 3: A segment of feature time series (shown separately) in real data, which includes an AB episode. a: RR (feature of RR intervals), b: QRSd (feature associated with the duration of QRS complex) c: RAMP (R-waveform amplitude). Grey dash line shows AB episode onset.

the optimal T_{sync} , two models with 2 states at each channel were trained using RR and RAMP with different time shifts. Each model was separately optimized using the training set of the first round of cross-validation. The time shift with the best performance was selected for T_{sync} . $D(c)$ was considered as 10 samples of time. We explored the set, $\{2, 3, 4, 5, 6, 7\}$, as the number of hidden states with same number of states in different channels to optimize the number of hidden states.

3 Results

3.1 Detection of dynamics of simulated data

The results are presented for detection of dynamic a_1 , which assesses the performance of the model to first detect an excitation from rest status and then identify dynamic a_1 from a_2 . Accordingly, the results related to the detection of a_2 reflects similar capacity and, thus were excluded. Table 1 compares the performance of the CHSMM with HMM [11], HSMM [12] and CHMM [16]. The proposed CHSMM demonstrated the best performance in terms of SEN (96.34%), SPC (99.71%) compared to other models. Also, in

Table 1: The cross-validation results of dynamic a_1 detection in simulated data using optimal state number comparing to models

Model	# states (rest- a_1 - a_2)	SEN(%)	SPC(%)	mean delay(s)	std delay(s)
HMM	2-4-4	85.23 ± 6.74	99.68 ± 0.48	10.83 ± 5.11	0.38 ± 0.52
HSMM	3-5-5	87.25 ± 1.56	97.47 ± 2.36	6.08 ± 4.71	1.42 ± 1.75
CHMM	2-5-5	93.43 ± 3.48	99.66 ± 0.09	1.59 ± 0.09	0.16 ± 0.06
CHSMM	2-5-5	96.34 ± 0.22	99.71 ± 0.49	1.72 ± 0.09	0.30 ± 0.50

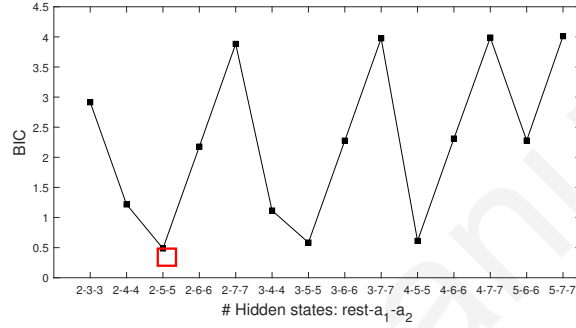


Figure 4: An example of optimization analysis for determining the number of hidden states based on BIC for the CHSMM trained by simulated data. The best number of states is associated with the lowest BIC value, marked by red square.

coupling methods lower average TD was obtained (1.72 s in CHSMM and 1.59 s in CHMM vs. 4.71 s in HSMM and 5.11 s in HMM). An example of optimizing number of hidden states for the CHSMM using BIC is shown in figure 4. Similar approach was used for the rest of the models applied to simulated data as well as those to real data.

3.2 Detection of apnea-bradycardia in real data

In Fig. 5, the ROC of different models are traced for different subset of features as $\{RR-RAMP\}$, $\{RR-QRSd-RAMP\}$ and $\{RR-QRSd\}$. The time delay is also shown in terms of $(1 - SPC)$ for each model. As illustrated, CHSMM method demonstrated the best performance in terms of AUC in all combination of features. The optimal thresholds (PD) of the ROCs are reported in Table 2. These best results were obtained based on $T_{sync} = 4.5$ s for CHSMMs and HSMMs, and $T_{sync} = 4$ s for HMMs and CHMMs.

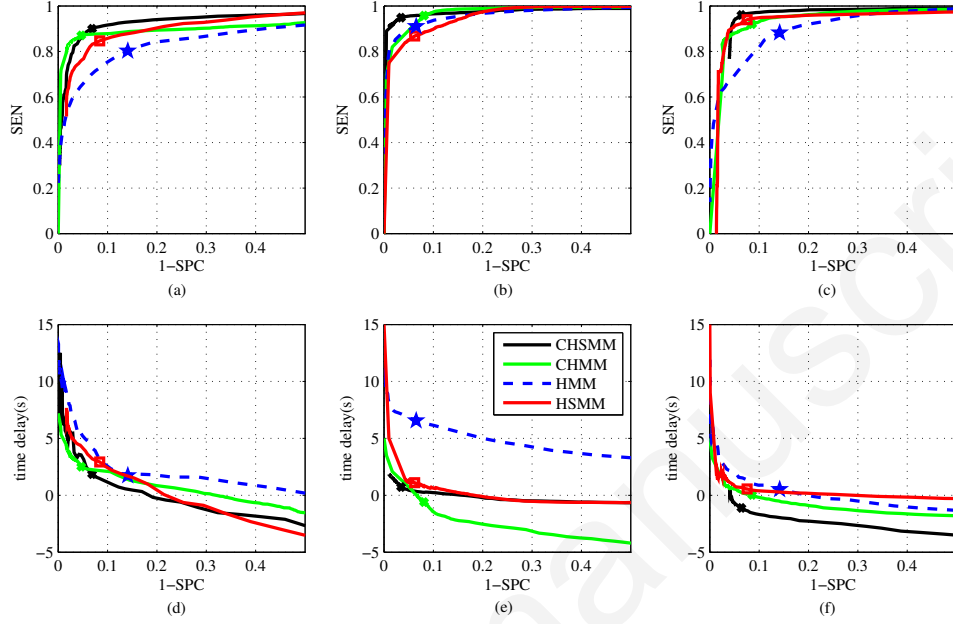


Figure 5: The performance of the proposed CHSMM in AB detection compared to other frameworks. a-c: ROC curves. d-f: time delay curves versus error. a, d: RR and QRSd. b, e: RR, QRSd and RAMP. c, f: RR and RAMP.

The results of apnea bradycardia detection are detailed in Table 2 for different feature combinations. The results compared CHSMM with the other Markovian models. Extracted from ROCs, the reported values for SEN and SPC of the PD points were averaged over five iterations of cross-validation. Also, among the CHSMMs, the one that was trained with $\{RR-QRSd-RAMP\}$ resulted in higher accuracy (95.69%) compared to $\{RR-RAMP\}$ and $\{RR-QRSd\}$ with accuracy of 94.91% and 91.48%, respectively. The accuracy of CHSMM trained with $\{RR-QRSd-RAMP\}$ was also higher than 93.81% for CHMM, 90.30% for HSMM, and 92.24% for HMM. Compared to the other models, CHSMM showed the least distance to PD and the highest AUC in all feature combinations. PW was lower for CHSMM than other models showing that most events were predicated.

Table 2: The cross-validation results of CHSMM in different configurations of features compared to other frameworks

Features	Method	optimal # states	SEN(%)	SPC(%)	mean delay(s)	std delay(s)	PW(%)	AUC	Distance to <i>PD</i>
RR RAMP	HMM	6	85.82 ± 6.81	88.34 ± 2.21	0.77 ± 0.97	2.14 ± 0.98	64.30	0.94	0.18
	HSMM	3	93.74 ± 0.40	91.86 ± 0.23	0.56 ± 0.07	1.49 ± 0.03	60.00	0.95	0.10
	CHMM	5	87.64 ± 0.94	92.13 ± 6.44	0.59 ± 0.01	3.11 ± 1.94	47.29	0.95	0.15
	CHSMM	3	95.99 ± 0.31	93.84 ± 0.24	-1.11 ± 0.04	2.56 ± 0.03	35.00	0.97	0.07
RR QRSd	HMM	2	82.28 ± 16.87	85.32 ± 3.78	1.81 ± 0.84	4.10 ± 4.09	73.72	0.89	0.23
	HSMM	5	84.74 ± 0.36	91.33 ± 0.18	2.93 ± 0.03	3.08 ± 0.03	87.50	0.93	0.18
	CHMM	5	87.04 ± 3.78	95.31 ± 1.36	2.51 ± 0.93	3.02 ± 0.18	83.22	0.93	0.14
	CHSMM	5	89.97 ± 0.21	92.99 ± 0.31	1.85 ± 0.03	3.14 ± 0.04	78.95	0.95	0.12
RR RAMP QRSd	HMM	2	91.02 ± 0.64	93.47 ± 0.08	6.55 ± 0.69	2.29 ± 0.79	63.13	0.97	0.11
	HSMM	7	86.83 ± 25.62	93.77 ± 4.44	1.10 ± 2.74	1.64 ± 1.18	44.10	0.96	0.16
	CHMM	2	95.74 ± 0.82	91.88 ± 0.31	-0.59 ± 0.21	2.79 ± 0.06	44.23	0.97	0.09
	CHSMM	4	94.87 ± 3.18	96.52 ± 0.64	0.73 ± 0.60	2.45 ± 0.69	35.00	0.98	0.06

4 Discussion and Conclusion

In this study, a new HMM-based algorithm was presented to detect apnea bradycardia using single channel ECG in preterm infants. The main findings of the study are: 1) presenting a new set of equations for implementation of CHSMM in a supervised learning algorithm; 2) the CHSMM learning algorithm was integrated in a detection algorithm; 3) the detection algorithm was successfully validated on a simulated dataset; and 4) the detection algorithm was able to improve the accuracy of apnea bradycardia detection using the extracted features from ECG.

The proposed CHSMM was successfully applied in detecting apnea bradycardia episodes in preterm infants with the highest accuracy compared to other HMM-based models that were presented for this purpose. Previously, we presented a set of equations for CHMM [16], which expanded the HMM structure to model coupling among different observations of a system. In this study, the CHMM was further generalized by incorporating the resting time to the characteristics of hidden states. According to the CHSMM structures, the *Forward-Backward* variables along with other variables were re-defined and used to implement the re-estimation and inference. The re-definition of equations for CHSMM were inspired by [12], where they were calculated based on the conditional probability of states subjected to observation. In

contrast, in the CHSMM structure proposed by Natarajan [17], the equations were derived through using joint probability in the definition of *bricks*, which adds more complexity to the computations.

In our CHSMM, the *Forward* variable required $TMDC$ calculations for $\mathcal{B}_t^s(m, d)$ and $\mathcal{A}_t^s(m, d)$, which include $4 + MD + (MD + 2)(C)$ and $MD + 1$ multiplications, respectively. Hence, the order of calculations in this step is $O(TM^2D^2C^2)$. This computational cost was lower than the one that was reported in Natarajan's CHSMM [17], as $(O(TM^2C^2D^3))$. The recursive structure of the equations and sample by sample windowing process make the CHSMM computationally tractable, and suitable for online applications. The proof of the equations have been reported in Supplementary file for better flow.

The better performance of CHSMM and CHMM compared to HSMM and HMM implies that the coupling approach outperforms in learning the dynamics of the signals. In a study by Altuve et al. [9], it was shown that HSMM improved the accuracy of apnea bradycardia compared to HMM. This study showed that CHSMM demonstrated higher performance compared to CHMM in terms of detection accuracy.

In summary, we presented a new set of equations for implementing CHSMM. The detection algorithm based on CHSMM was evaluated on a simulated data with two different dynamics and a clinical data for automatic on-line detection of apnea bradycardia in premature infants. Such accurate detection algorithm can be implemented as a monitoring and alarm system in NICU for early detection of respiratory events to prevent the adverse outcomes of undiagnosed and untreated apnea of prematurity.

Acknowledgements

This work has been supported by the Center for International Scientific Studies and Collaboration (CISSC) and by Egide-Gundishapour program.

References

- [1] Jing Zhao, Fernando Gonzalez, and Dezhi Mu. Apnea of prematurity: from cause to treatment. *European journal of pediatrics*, 170(9):1097–1105, 2011.

- [2] Eric C Eichenwald et al. Apnea of prematurity. *Pediatrics*, 137(1), 2016.
- [3] TJ Mathews and Marian F MacDorman. Infant mortality statistics from the 2009 period linked birth/infant death data set. *National vital statistics reports*, 61(8), 2013.
- [4] Annie Janvier, May Khairy, Athanasios Kokkotis, Carole Cormier, Denise Messmer, and Keith J Barrington. Apnea is associated with neurodevelopmental impairment in very low birth weight infants. *Journal of perinatology*, 24(12):763–768, 2004.
- [5] Christian F Poets, Valerie A Stebbens, Martin P Samuels, and David P Southall. The relationship between bradycardia, apnea, and hypoxemia in preterm infants. *Pediatric research*, 34(2):144–147, 1993.
- [6] François Portet, Feng Gao, Jim Hunter, and Somayajulu Sripada. Evaluation of on-line bradycardia boundary detectors from neonatal clinical data. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 3288–3291. IEEE, 2007.
- [7] J Cruz, AI Hernández, S Wong, G Carrault, and A Beuchee. Algorithm fusion for the early detection of apnea-bradycardia in preterm infants. In *Computers in Cardiology, 2006*, pages 473–476. IEEE, 2006.
- [8] Jérôme Dumont, Alfredo I Hernández, Julien Fleureau, and Guy Carrault. Modelling temporal evolution of cardiac electrophysiological features using hidden semi-markov models. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 165–168. IEEE, 2008.
- [9] M Altuve, G Carrault, A Beuchée, P Pladys, and A I Hernández. On-line apnea-bradycardia detection based on hidden semi-markov models. *Medical & biomedical engineering & computing*, pages 1–13, 2014.
- [10] S Masoudi, N Montazeri, MB Shamsollahi, D Ge, Alain Beuchee, P Pladys, and AI Hernandez. Early detection of apnea-bradycardia episodes in preterm infants based on coupled hidden markov model. In *Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on*, pages 000243–000248. IEEE, 2013.

- [11] Lawrence Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [12] Shun-Zheng Yu and Hisashi Kobayashi. Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model. *Signal Processing, IEEE Transactions on*, 54(5):1947–1951, 2006.
- [13] Jack D Ferguson. Variable duration models for speech. In *Proc. Symposium on the application of hidden Markov models to text and speech*, pages 143–179, 1980.
- [14] Matthew Brand, Nuria Oliver, and Alex Pentland. Coupled hidden markov models for complex action recognition. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 994–999. IEEE, 1997.
- [15] Nathan Brewer, Nianjun Liu, Olivier De Vel, and Terry Caelli. Using coupled hidden markov models to model suspect interactions in digital forensic analysis. In *Integrating AI and Data Mining, 2006. AIDM’06. International Workshop on*, pages 58–64. IEEE, 2006.
- [16] Nasim Montazeri Ghahjaverestan, Samira Masoudi, Mohammad Shamsollahi, Alain Beuchee, Patrick Pladys, Di Ge, and Alfredo Hernandez. Coupled hidden markov model based method for apnea bradycardia detection. *Accepted in IEEE Journal of Biomedical and Health Informatics*, DOI: 10.1109/JBHI.2015.2405075, 2015.
- [17] Pradeep Natarajan and Ramakant Nevatia. Coupled hidden semi markov models for activity recognition. In *Motion and Video Computing, 2007. WMVC’07. IEEE Workshop on*, pages 10–10. IEEE, 2007.
- [18] Matthew Brand. Coupled hidden markov models for modeling interacting processes, 1997. Technical Report 405.
- [19] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

- [20] Andrew A Neath and Joseph E Cavanaugh. The bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2):199–203, 2012.
- [21] Jeffrey Rauch and Joel Smoller. Qualitative theory of the fitzhugh-nagumo equations. *Advances in mathematics*, 27(1):12–44, 1978.
- [22] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *Biomedical Engineering, IEEE Transactions on*, (3):230–236, 1985.
- [23] Miguel Altuve, Guy Carrault, Julio Cruz, Alain Beuchée, Patrick Pladys, and Alfredo Hernandez. Multivariate ecg analysis for apnoea? bradycardia detection and characterisation in preterm infants. *International Journal of Biomedical Engineering and Technology*, 5(2):247–265, 2011.
- [24] K Haskova, K Javorka, M Javorka, K Matasova, and M Zibolen. Apnea in preterm newborns: determinants, pathophysiology, effects on cardiovascular parameters and treatment. *Acta Medica Martiniana*, 13(3):5–17, 2013.
- [25] Christian F Poets. Apnea of prematurity: What can observational studies tell us about pathophysiology? *Sleep medicine*, 11(7):701–707, 2010.