Early Detection of Apnea-Bradycardia Episodes in Preterm Infants based on Coupled Hidden Markov Model

S. Masoudi¹, N. Montazeri¹², M.B. Shamsollahi¹, D. Ge²³, A. Beuchée²³⁴, P. Pladys²³⁴, A.I. Hernandez²³

¹ BiSIPL, School of Electrical Engineering, Sharif University of Technology, Tehran, Iran
² LTSI, Université de Rennes 1, Rennes, France
³ INSERM: U1099, Rennes, France
⁴ CHU Rennes, Pôle Médiwo-Chirurgical de Pédiatrie et de Génétique Clinique, Néonatologie, Rennes, France

Abstract—The incidence of apnea-bradycardia episodes in preterm infants may lead to neurological disorders. Prediction and detection of these episodes are an important task in healthcare systems. In this paper, a coupled hidden Markov model (CHMM) based method is applied to detect apnea-bradycardia episodes. This model is evaluated and compared with two other methods based on hidden Markov model (HMM) and hidden semi-Markov model (HSMM). Evaluation and comparison are performed on a dataset of 233 apnea-bradycardia episodes which have been manually annotated. Observations are composed of RR-interval time series and QRS duration time series. The performance of each method was evaluated in terms of sensitivity, specificity and time detection delay. Results show that CHMM has the sensitivity of 84.92%, specificity of 94.17% and time detection delay of 2.32±4.82 seconds, which are better than the reference methods.

I. INTRODUCTION

Apnea-bradycardia episode is defined as a respiratory pause accompanied with a heart rate reduction which is a common event in preterm infants [1]. The repetition of these episodes compromises oxygenation and tissue perfusion and may lead to neurological impairment or even short-term morbidity-mortality. Since apnea-bradycardia event consists of an apnea episode followed by a significant fall in heart rate, earlier detection of bradycardia means better detection of apnea. However, most of the works in this area try to provide an acceptable solution using only cardiac signals, limiting the use of the more noisy respiratory signals. Recent research in detection of apnea-bradycardia episodes, mostly concentrate on processing of a single variable extracted from the electrocardiogram (ECG): the cardiac cycle length (RR interval) [4].

Premature infants’ cardiac signals are continuously monitored in neonatal intensive care units (NICU), in order to detect apnea-bradycardia events and to initiate quick nursing actions such as manual or vibrotactile stimulation, oxygenation and ventilation through a mask or intubation [2,3]. Typically, when an infant presents a bradycardia event, an alarm is generated by a monitoring device, an available nurse or physician goes to the appropriate NICU room, and applies a manual stimulation to the infant in distress.

Conventional detectors used in NICU detect bradycardia through fixed or relative-threshold methods. In these methods high specificities (few false alarms) and low sensitivities (missed apnea-bradycardia episodes) are achieved. They also produce long detection delays that prolong the intervention delay. As an evidence for this conclusion the mean intervention delay measured from the activation of the alarm to the application of the therapy based on these approaches has been estimated to be around 33 seconds with mean manual stimulation duration of 13 seconds [4].

Literature on prognostic methods is extremely limited but the concept has been gaining importance in recent years. Unlike numerous methods available for diagnostics, prognostics is still in its infancy and researchers are yet to present a working model for effective prognostics. In our recent works, hidden Markov model (HMM) and hidden semi-Markov model (HSMM) have been proposed as useful tools for early detection of apnea-bradycardia [5, 6]. HMM, as a probabilistic approach, has been quite effective in some applications such as speech and biomedical signal processing. Although the statistical methods of hidden Markov modelling were initially introduced and studied in the late 1960s and early 1970s, they have become increasingly popular recently. There are two major reasons for this. First, the models have very rich mathematical structure and can form a solid theoretical foundation. Second, the models have many successful applications in practice. The fact that HMM-based models could be applied in the area of prognostics was first pointed out by Bunks et al [7] and Baruah and Chinnam [8] in machining processes field. HMM is a statistical model with unobserved states that produce a sequence of observations. Stochastic processes which can be modelled in HMMs are Markovian. In this network, change for a future state depends
solely on the current state. An inherent limitation of the HMM approach is that it does not provide adequate representation of systems in which the stage-phase in a given state cannot be represented by a geometrical law. One way to overcome this limitation is to apply Hidden Semi-Markov Models. HSMM’s structure is similar to a classic HMM but the main difference is that the unobserved process is semi-Markov in the sense that a change for a future hidden state depends on both the current state and the time spent on this state. Unlike the standard HMM in which one state produces one single observation, a state in a semi-Markov model is defined as consecutive repetitions of that state which generates a segment of observations.

Subsequently, in a simple HMM there might be several observation sequences. This is the case when a number of observation sequences are extracted from one single signal, i.e. the RR series and the QRS durations from the same ECG signal in the current application. For this purpose, it is essential to study the coupling effects between the channel dynamics behind the observation sequences. In such a coupled hidden Markov model (CHMM), states in multiple channels evolve with mutual dependences while each observation sequence follows an emission distribution conditionally to the channel state only. This implies a Cartesian product form of states in multiple channels for the Markovian transition law, to potentially increase the amount of information extracted from data.

The objective of this paper is to apply the CHMM framework for apnea-bradycardia detection. Also, in order to provide a better insight about the added value of a CHMM approach, a quantitative performance comparison with respect to our previous works on HMM and HSMM-based apnea-bradycardia detectors is presented. These three methods are evaluated on the same database. The implementation of CHMM in this work is based on the framework proposed by Rezek et al. [9].

This paper is organised as follows: Section II will introduce basic concepts of CHMM. Application of this method and the comparison with the others in detection of apnea-bradycardia will be discussed in section III. Concluding remarks will be presented in section IV.

II. CHMM

CHMM is a statistical signal model with a coupled multi-channel structure. These channels are parallel to each other, each of them acts like a Markov chain and represents a random sequence of observations. These observations are based on probabilistic functions given the unobserved states [10]. A sequence will be generated as a result of a random walk in transition networks of the states. The transitions are reflected in the corresponded chain and this causes the production of an observation at each visit of a state [11, 12]. A general CHMM model with C channels is depicted in Fig.1 having the following elements, for c=1,…,C:

1. A Set of states: $S^c = \{s_1^c, s_2^c, s_3^c, ..., s_{l_c}^c\}$. (Number of states: $l_c$)
2. A Set of individual observations: $V^c = \{v_1^c, v_2^c, v_3^c, ..., v_{K_c}^c\}$. (Number of observations: $K_c$)
3. The state transition probability distribution $A^c = \{a_i^c\}_{i=1}^{l_c}$ which is a C+1 ordered tensor needed to encode the state transition law in each channel c and each tensor has $l_c \times l_c$ elements.

$$a_{i_1, i_2, ..., i_{l_c}} = P(q_{t+1} = s_{i_2}^c | q_t = s_{i_1}^c, ..., q_{t_l} = s_{i_{l_c}}^c) = \sum_{j=1}^{l_c} a_{i_1, i_2, ..., i_{l_c}} = 1$$

4. The observation probability distribution in state $s_i^c$, $B^c = \{b_i^c[k]\}$, where

$$b_i^c[k] = P(O_t = v_k^c | q_t = s_i^c) = \sum_{k=1}^{K_c} b_i^c[k] = 1$$

5. The initial state distribution $\pi^c = \{\pi_i^c\}$ where

$$\pi_i^c = P(q_1 = s_i^c) = \sum_{t=1}^{l_c} \pi_i^c \sum_{t=1}^{l_c} = 1$$

So for each channel the model can be described by the compact notation set of parameters, often used in the literature:

$$\lambda^c = (A^c, B^c, \pi^c)$$

The final model for general CHMM can be expressed as in Eq.5.

$$\lambda = (\lambda^1, \lambda^2, ..., \lambda^C)$$

Fig. 1 A general CHMM with 2 channels

According to these characteristics of the CHMM’s structure, the following equations can be deduced. One essential probabilistic description of the above system is expressed as
\[ P(q_{i+1}^c, \ldots, q_C^c | q_i^c, \ldots, q_C^c) = \prod_{c=1}^C P(q_{i+1}^c | q_i^c, \ldots, q_C^c) \]  

(6)

which declares that the states at time \( t+1 \) \( (q_{i+1}^c, c=1, \ldots, C) \) given all the states at time \( t \), are independent. We also have the independence of observations at time \( t \) \( (o_t^c, c=1, \ldots, C) \) when their corresponding states are specified.

\[ P(o_1^c, \ldots, o_T^c | q_1^c, \ldots, q_T^c) = \prod_{c=1}^C P(o_1^c | q_1^c) \]  

(7)

For continuous observation problems, some restrictions have to be forced on the form of the emission probability. One of the most general representation is a finite mixture of Gaussian distributions, and is applied in this work:

\[ P(o_1^c, \ldots, o_T^c | q_1^c, \ldots, q_T^c) = \prod_{c=1}^C \sum_{\mu_j^c \in S^c} \sum_{\sigma_j^c \in S^c} \pi_j^c \phi(o_1^c, \ldots, o_T^c | \mu_j^c, \sigma_j^c) \]  

(8)

where \( \mathcal{G} \) represents the Gaussian distribution and \( C_j^c \) stands for its weight coefficient. The number of mixed Gaussian distributions with the mean and variance are noted \( M^c, \mu_j^c \), and \( \sigma_j^c \) respectively.

Similar to the HMM [14], three basic problems are to be addressed for the CHMM:

1. Evaluation (also called Classification): Given C observation sequences \( o^c = o_1^c, o_2^c, o_3^c, \ldots, o_T^c \) for each of C channels with C set of parameters \( \lambda^c \), where \( c=1, \ldots, C \), what is the probability of the all C observation sequences with the given model, i.e., \( P(o_1^c, o_2^c, \ldots, o_T^c | \lambda^1, \lambda^2, \ldots, \lambda^C) \)?

2. Decoding (also called Recognition or Inference): Given C observation sequences \( o^c = o_1^c, o_2^c, o_3^c, \ldots, o_T^c \) for each of C channels with C set of parameters \( \lambda^c \), where \( c=1, \ldots, C \), what sequences of hidden states: \( q^c = q_1^c, q_2^c, q_3^c, \ldots, q_T^c \), \( c=1, \ldots, C \), will generate these given sequences of observations with maximum probability?

3. Learning (also called Training): Given C observation sequences \( o^c = o_1^c, o_2^c, o_3^c, \ldots, o_T^c \) for each of C channels, how do we adjust the model parameters \( \lambda^c = (A^c, B^c, \pi^c), c=1, \ldots, C \), to maximize \( P(o_1^c, o_2^c, \ldots, o_T^c | \lambda^c) \)?

Different algorithms have been developed for these problems and are based on the simplification of the CHMM structure. As a comprehensive and precise CHMM method, Rezek et al. [9] framework is employed in this work. They substituted the C channel CHMM with a large single-channel HMM-like structure and each state can be viewed as a Cartesian product of the states in C channels: \( s = [s_{1t}, s_{2t}, s_{3t}, \ldots, s_{Ct}] \) where \( s_{kt} \) is a member of the set \( S^c \) for any \( c=1, \ldots, C \).

Therefore, there would be total \( N = \prod_{c=1}^C I_c \) possible states in this large HMM-like at every time instant. Based on the idea of Rezek et al. [9] an N by N matrix called \( A \) is formed, in which each element represents probability of state transition from one C-fold state to another state of the respective structure in a large HMM-like. Strictly speaking, a large HMM is something more general.

In Rezek’s CHMM, observation in one time instant can be considered as a Cx1 vector \( v = [v_1^c, v_2^c, v_3^c, \ldots, v_C^c] \) where \( v_k^c \) is a member of the set \( V^c \) for any \( c=1, \ldots, C \). So there are \( M = \prod_{c=1}^C K_c \) possible observations at every time instant of the large HMM. An N by M matrix \( B \) can be defined to denote the emission probabilities of the CHMM. Thus this large HMM can be described as \( \lambda = (A, B, \pi) \).

Practically there is no necessity to estimate the full NxDN state transition matrix \( A \), however each element of the transition matrix \( A \) can still be obtained as a factor of the elements of the original transition matrices \( (A_i) \) of the C-channel CHMM.

Consequently solving the three aforesaid problems for a CHMM is equivalent to solving the same problems in a single-channel HMM context.

The forward–backward procedure [10,11] is used to solve the evaluation problem, whereas for the decoding problems the Viterbi algorithm is applied to find the most likely state sequence given the observation sequence in the maximum a posteriori probability sense by maximizing \( P(O, \lambda) \). As for the learning problem, no analytical solution exists in the literature. An iterative procedure such as the Baum–Welch method (or equivalently the Expectation-Maximization algorithm) that locally maximizes \( P(O|\lambda) \) is applied to allow adjustment of model parameters \( \lambda = (A, B, \pi) \) [15].

Using Rezek’s form, the forward variable \( \alpha_t(s) \) is defined as

\[ \alpha_t(s) = P(o_1^c, o_2^c, \ldots, o_t^c | s_t = s_1^c, q_1^c = s_1^c) \]  

(9)

which is employed to calculate the probability of entire \( O = o_1^c, o_2^c, \ldots, o_T^c \) given the whole model \( \lambda \). This definition leads to produce the joint likelihood for a CHMM model, i.e., \( P(O|\lambda) \).

Based on work of Rezek et al. [9] there is another forward variable \( \alpha_t^c(s) \) defined as

\[ \alpha_t^c(s) = P(o_1^c, o_2^c, \ldots, o_{t-1}^c | s_t = s_1^c, q_1^c = s_1^c) \]  

(10)

By using this statement which was elicited Eq.7, in evaluation phase there would be one likelihood for each channel. Then in a C channel CHMM the joint probability of \( o_1^c, o_2^c, \ldots, o_T^c \) given the model \( \lambda^c \) can be the multiplication of all C likelihoods. It can be shown that \( \alpha_t^c(s) \) can be obtained as follows:

\[ \alpha_t^c = \sum_{c=1}^C \sum_{c=1}^C \frac{\alpha_t^c(s)}{\prod_{c=1}^C b_k^c(o_t^c)} \]  

(11)

In this paper former definition is applied for evaluation process.
III. APPROACHES

Early detection in this work is actually a classification problem as proposed in [5], i.e., for each newly observed data we choose a model among several competing models and the likelihood evaluation from problem (1) allows us to choose the best match based on the observations [14]. The implementation details for these three methods are discussed in following paragraphs.

A. C-channel CHMM

Proposed CHMM based method consists of $K$ models, each representing a certain class. If $K$ equals two, then these classes can be the situations of absence and presence of an event. Decision about attribution of one model to a certain instant $t$ will be made using a sliding window of length $T$. Every model has $C$ channels, each channel produces a one dimensional observation sequence. Given the observations of a segment of size $T$, the marginal log likelihood of the $k$th model can be written as:

$$L_k^t = \log P(O_{t-T+1:t}^1 | \delta_k^t)$$

where $O_{t-T+1:t}^1$ represents the observations of all channels in a segment from $t-T+1$ to $t$.

Detection of event $a$ among $K$ incidents is performed by comparing the difference between these log likelihoods to a certain threshold:

$$L_a^t - L_k^t > \delta_{a,k}$$

B. HMM/HSMM with C dimensional observation

HMM and HSMM based methods also have $K$ models to present K prenominated possible situations. These models have a sequence of single hidden states. In this paper, multiple observations will be exploited in the HMM/HSMM algorithm. Similar to what was stated in the previous section about model selection for every segment of observation, log likelihood for each model of HMM or HSMM is defined as in Eq.12. The only difference here is that each of HMM and HSMM models has one channel. This fact reduces the number of states and the complexity of transitions, which therefore reduces the computational load [5, 6].

IV. EVALUATION

Inspiring from [6], the analysis have been done over simulated and real observation signals. The evaluation on simulated signals aims at optimizing and validating the proposed methods while experiments on the real signals acquired in NICU, for which apnea-bradycardia episodes are manually annotated, are used for performance evaluation. For event detection we use two models (presence and absence of the event) and for simplicity each model has two states per channel.

A. Simulated data

There is a need to generate two observation signals, to describe an event with potentially different dynamics. Synthetic time series, which are produced by Fitz Hugh-Nagumo model [6] have been used in our work.

$$\frac{dv}{dt} = 3 \left( v - \frac{1}{3}v^3 + r + t \right)$$

$$\frac{dr}{dt} = -\frac{1}{3}(v - a + 0.8r)$$

Two groups of these signals, one consisting of rest intervals and the other having perturbation are used to train two models. Each group contains 40 signals of 10 seconds.

A group of 40 signals of 400s consisting both rest and perturbations intervals, are generated for simulation test. Log likelihoods of both models are evaluated from a sliding window of 10s.

B. Real data

RR-interval time series and QRS duration time series, extracted from ECG of 32 preterm infants, are the two-channel observations in the present study and we believe the explicit modelling and estimation of the coupling effect between the two channels might better take advantage of the system dynamics. These observations have been manually divided into two groups. The first group includes 148 segments of 7s in normal situation, and is used to train the normal model (in absence of AB events); the other group includes 30 segments of the same size containing bradycardia events and is used to train the other model (in presence of the AB events).

A set of 40 signals with different length are used as test data in order to evaluate the performance of the algorithms. A 7-second window slides sample by sample on the observation sequence for making a decision about occurrence of the bradycardia event at the sampled time $t_0$.

For the CHMM approach, RR interval and the QRS duration, both extracted from the raw ECG signal are used as the two-channelled observation data. Since every channel utilizes a particular type of observation, its corresponding channel states can be defined as certain amplitude ranges. Typically, the 2-D HMM/HSMM algorithm including RR interval and QRS duration can’t interpret certain dynamic characteristics since the observations consist of two incongruous emissions [5, 6].

C. Performance evaluation

The detection results with the annotated data for each of the three algorithms are compared. For this purpose, we define four parameters, i.e. true positive (TP), true negative (TN), false positive (FP) and false negative (FN). True positives is attributed to detections of the bradycardia occurrence within a 20s window, centered at the annotations.

Finally estimating sensitivity (SEnable)(TP/(TP+FN)) and specificity (SNCable) for different threshold values, leads to variant quantities of these measures which are represented by means of ROC curve. Time detection delay is also calculated. It is defined as the time elapsed between annotation instant and the detection. This definition produces a
negative value when detection situates before the annotated begin of bradycardia. Consequently there would be a number of time detection delays for each threshold. By estimating the average and standard deviations of these time detection delays, dd can be defined (dd= Mean ±Stdev). Optimum performance detection and its respective threshold are evaluated using the criterion introduced as perfect detection (PD) according to Eq.15.

\[
PD = \text{max}(SEN \times SPC)
\]  

(15)

V. RESULTS

The average results over ten times of executions on simulated data are shown in table 1. Performance measures including specificity, sensitivity and detection delay are calculated for all three methods. Sensitivity and specificity values are acceptable and satisfying for all of them, however CHMM based method has the ability to detect earlier than the labelled onset of the event of interest.

<table>
<thead>
<tr>
<th>Variable</th>
<th>SEN(%)</th>
<th>SPC(%)</th>
<th>dd(seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM [16]</td>
<td>98.37%</td>
<td>92.80%</td>
<td>0.81± 0.03</td>
</tr>
<tr>
<td>HSMM [17]</td>
<td>97.35%</td>
<td>97.93%</td>
<td>0.30± 0.02</td>
</tr>
<tr>
<td>CHMM [9]</td>
<td>97.43%</td>
<td>92.24%</td>
<td>0.69± 1.31</td>
</tr>
</tbody>
</table>

All three methods are also analyzed with a real database of 233 apnea-bradycardia episodes. Achieved results for sensitivity vs specificity are represented in ROC curve in Fig.2. The ROC curve of the CHMM proposed method is placed above two other curves which indicates the better performance of this method. The optimum points obtained using Eq.15 are shown with character “X” in Fig.2. Superiority of CHMM is also confirmed by PD points.

![Fig. 2 ROC curves for sensitivity and specificity](image)

Considering the importance of time detection delay in this type of prediction problem, obtained time detection delays versus specificity changes are depicted in Fig.3. Results clearly state that the earlier detection of apnea-bradycardia with a trustable level of accuracy is achieved by CHMM. Optimum points are also represented in this figure.

![Fig. 3 Specificity and mean detection delay of apnea-bradycardia episodes.](image)

Related values to the PD points, which are obtained for the thresholds of \( \delta^{a,r} = 300, 3900, 27900 \) respectively, are listed in table.2. All performance measures for CHMM based method are better than two other algorithms. In particular, calculated detection delay and specificity of CHMM has a significant improvement in comparison with other models. It is worthy of noting that same amounts of specificity results in better time detection delay for CHMM as it can be seen in Fig.3.

![Table 2: Performance measures by evaluating PD for real data.](image)

VI. CONCLUSIONS

Evaluation results show a reduced time detection delay and higher detection performance in Coupled Hidden Markov Model based method rather than conventional detectors based on Hidden Markov Model and Hidden Semi Markov Model. Employing two channels instantaneously, to take into account their reciprocal effects on each other and construct a strong structure. These characteristics of CHMMs provide the capability of reliable performance and make them as strong tools for prediction in medical applications such as apnea-bradycardia detection. On the other hand, the order of calculations in current implementation of CHMM is \( O(M^{2C}T) \) where \( M = \max(I_1, \ldots, I_C) \) which dramatically increases the calculation load by large number of channels. So our future work could focus on the optimisation of these algorithms.

![Diagram](image)
work will be dedicated to present a new more efficient method for CHMM framework.

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

The authors would like to thank Dr. Iead Rezek for his guidance and comments.

REFERENCES