Comparing Hidden Markov Model and Hidden Semi-Markov Model Based Detectors of Apnea-Bradycardia Episodes in Preterm Infants

Miguel Altuve[1,2,3], Guy Carrault[2,3,4], Alain Beuchée[2,3,4,5],
Cyril Flamand[6], Patrick Pladys[2,3,4,5], Alfredo I. Hernández[2,3,4]

[1] Grupo de Bioingeniería y Biofísica Aplicada, Universidad Simón Bolívar, Caracas, Venezuela
[2] Inserm U1099, Rennes, F-35000, France
[3] Université de Rennes 1, LTSI, Rennes, F-35000, France
[4] INSERM CIC-IT 804, Rennes, F-35000, France
[5] CHU Rennes, Pôle Médico-Chirurgical de Pédiatrie et de Génétique Clinique, Néonatologie, Rennes, F-35000, France
[6] CHU Nantes, Service de réanimation pédiatrique, néonatologie et CIC, Nantes, France

Abstract

In this paper, we propose, evaluate and compare two detectors of apnea-bradycardia episodes, based on hidden Markov models (HMM) and hidden semi-Markov models (HSMM). Evaluation is performed on a database of 233 apnea-bradycardia episodes manually annotated. The acquired ECG signals are processed to obtain RR series. The proposed detectors, applied to these RR series, are composed of two HMM or HSMM models, each one representing two distinct physiopathological states: absence and presence of apnea-bradycardia. A learning phase is firstly applied to each model in order to estimate their parameters from a learning dataset. Then, using a sliding window, the models are applied to a set of new observations, to compute the log-likelihood of each model for each time instant. Detection of the events of interest is based on the comparison of log-likelihoods with respect to a threshold. The optimal detection configuration was obtained in terms of sensitivity, specificity and detection delay. Results show that the analysis of the dynamics of RR series, through the HSMM, allows for a significant improvement of sensitivity (90.38% vs 88.42%) and specificity (92.23% vs 89.67%), with a reduction of the detection delay (0.92±3.56s vs 1.60±3.72s).

1. Introduction

In preterm infants, a respiratory pause accompanied with a heart rate reduction is an apnea-bradycardia episode. The repetition of apnea-bradycardia episodes are associated with short-term morbi-mortality and neurological impairment during childhood [1,2]. For these reasons, in neonatal intensive care units (NICU), preterm infants undergo continuous cardiorespiratory monitoring to detect early signs of bradycardia and to initiate quick nursing actions (manual or vibrotactile stimulation, oxygenation, ...). Apnea-bradycardia episodes are commonly detected by processing a single variable extracted from the electrocardiogram (ECG): the cardiac cycle length (RR interval) [3–5]. However, conventional detectors based on thresholds (fixed and relative) used in NICU, produce high specificities (few false alarm) but low sensitivities (missed apnea-bradycardia episodes) and long detection delays that prolongs the intervention delay (33 s approximately, measured from the activation of the alarm to the application of the therapy [6]).

Recently, an apnea-bradycardia detector based on hidden semi-Markov models (HSMM) with different preprocessing methods has been proposed [7]. The detector exploits the instantaneous values of the RR interval and the dynamics of the RR interval time series. The detector, evaluated on a database of 233 apnea-bradycardia episodes, showed a detection performance higher than conventional detectors used in NICU, while reducing the time detection delay. In this work, an equivalent apnea-bradycardia detector based on hidden Markov models (HMM) is proposed and evaluated on the same database. The objective here is to compare the detection performance of HMM and HSMM based detectors, when different preprocessing methods, like the quantization of the time series and the inclusion of delayed versions of the observations, are employed.

The rest of this work is organized as follows: the next section introduces the models to perform the apnea-bradycardia detection, the preprocessing stage and the evaluation methodology. The evaluation of the proposed detector is presented in section 3. Finally, the conclusions and future works are outlined in the last section.

2. Methods

2.1. HMM and HSMM based detectors

An HMM is a statistical model with unobserved states that produce a sequence of observation, in which the stochastic process to be modeled is Markovian, implying that the change to a future state depends solely on the current state. An HSMM is similar to a classic HMM, but the main difference is that the unobserved process is semi-Markov in the sense that a change to a future hidden state depends on both the current state and the time spent on this state.

HMM and HSMM are characterized by a number of M states and the set of parameters $\lambda \triangleq \{a_{ij}, b_i, \pi_i\}$ for HMM and $\lambda \triangleq \{a_{ij}, b_i, \pi_i, p_i\}$ for HSMM, where a_{ij} is the transition probability between states i and j ($a_{ii} = 0$ for HSMM), b_i is the probability of emission of observations, π_i is the probability of the initial state and p_i is the probability of duration for state i.

In this work, M is calculated using Bayesian information criterion (BIC) [8]. b_i and p_i are represented by a Gaussian distribution: $b_i(\vec{\mu}, \Sigma)$ and $p_i(\mu_d, \sigma_d)$, where $\vec{\mu}$ and Σ correspond to the centers and covariance matrix of observations and μ_d and σ_d are the mean and the standard deviation of the duration of the states (in seconds and truncated in zero). In order to avoid negative durations, Gaussian distribution for $p_i(\mu_d, \sigma_d)$ are always positives and truncated in zero.

In this work, K HMM and K HSMM are used to model K different observation dynamics, associated with distinct physiopathological states or events, to be discriminated. From the sequences of observations provided in a learning dataset, the parameters of the models λ^k, $\forall\, k \in \{1, 2, \ldots, K\}$ for HMM and HSMM, are estimated in a learning phase. During this phase, a_{ij} and π_i are initialized with uniform probabilities and $b_i(\vec{\mu}, \Sigma)$ is initialized by a Gaussian mixture model where the center of each Gaussian, which corresponds to the barycenter of each state in the observation space, have been initialized by the K-means algorithm. Additionally, λ^k for HSMM is initialized from the parameters of the equivalent HMM, in which the Viterbi algorithm is used to estimate $p_j(\mu_d, \sigma_d)$. The Viterbi algorithm and its extension to HSMM are subsequently applied to obtain the final value of λ^k for HMM and HSMM, respectively, through an expectation-maximization stage [9].

Learning is achieved when the log-likelihood

$$L^k = \log P(O_{1:T}|\lambda^k) \tag{1}$$

converges to a maximum value, where $P(O_{1:T}|\lambda^k)$ is the probability that the observation sequence $O_{1:T} = O_1, O_2, \ldots, O_T$ is generated by the model with parameters λ^k. Once the learning phase is completed, the K HMM and the K HSMM are applied in a test phase to a set of observations on a specific dataset, in order to perform event detection. In this case, the log-likelihood for instant t and model k,

$$L_t^k = \log P(O_{t-T+1:t}|\lambda^k) \tag{2}$$

is determined using a sliding window of length T (see Figure 1).

On-line detection of event $\alpha \in \{1, 2, \ldots, K\}$ is finally performed when the following equation is verified:

$$L_t^\alpha - L_t^k > \delta^{\alpha,k} \quad \forall\, k \in \{1, 2, \ldots, K\}, k \neq \alpha \tag{3}$$

where $\delta^{\alpha,k}$ are fixed thresholds that have to be optimized.

2.2. Preprocessing stage

Two preprocessing methods are employed to improve the detection performance of the proposed detectors. The first preprocessing method is the quantization of the observations. This quantization phase is justified since *i*) the models are constituted of a discrete number of states M representing the dynamic range of the signal amplitude and *ii*) the observations used as inputs to the proposed detector are the result of an automatic ECG segmentation phase that may produce segmentation errors. Two quantization methods are used: a uniform, characterized by a constant quantization step Δ_{QU} and a non uniform, characterized by a quantization step $\vec{\Delta}_{\mathrm{QNU}}$ which depends on the distribution of the signal. Vector $\vec{\Delta}_{\mathrm{QNU}}$ is found by comparing the cumulative sum of values of the normalized histogram with respect to a threshold δ_{QNU}. Different values of Δ_{QU} and δ_{QNU} have been tested to obtain the optimum value that improves the detection performance. Variables noted with UQ and with NUQ indicate its uniform and non uniform quantization, as for $\mathrm{RR}_{\mathrm{UQ}}$ and $\mathrm{RR}_{\mathrm{NUQ}}$.

The second preprocessing method concern the integration of delayed versions of the measures into the observation vector. In the case of the RR series, the application of such time-delay presents a physiological argument, since it is known that the RR interval is modulated by the autonomic nervous function, with different dynamics for the sympathetic and parasympathetic systems. In this case, the observation matrix becomes:

$$\mathbf{O} = \begin{bmatrix} O_{t-T+1:t} \\ O_{t-\tau-T+1:t-\tau} \end{bmatrix}, \tag{4}$$

where τ is the predefined time delay. Different values of τ have been tested to obtain the optimum value that improves the detection performance. These observation matrices will be represented in bold in this paper, such as **RR**.

2.3. Evaluation methodology

The proposed detectors were evaluated with real signals acquired in the Rennes NICU, in which apnea-bradycardia

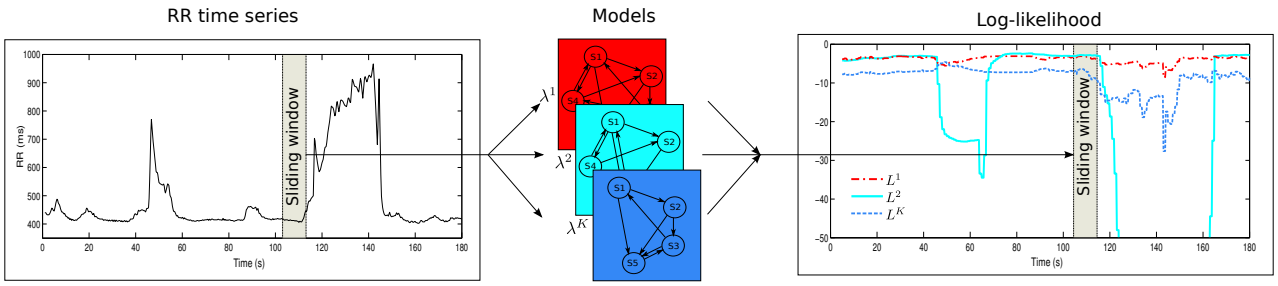

Figure 1. Methodology for event detection using HMM or HSMM. A sliding window is used to select a portion of the RR time series to be passed through the K models in order to obtain the log-likelihood.

episodes have been manually annotated by experts. RR time series were extracted from the ECG of 32 preterm infants with frequent apnea-bradycardia episodes, as described in [10, 11]. RR time series were uniformly resampled at 10 Hz. 233 bradycardia episodes were manually annotated from 148 RR series.

For the learning phase, two datasets were constructed:

- LS1: composed of 30 segments taken randomly and including an apnea-bradycardia event, taken at the beginning of the bradycardia with duration of $T = 7$ s.
- LS2: consisting of 300 segments at rest, taken randomly from the series, with duration of $T = 7$ s.

The length of these segments (7 s) corresponds to the average time measured from the beginning of the bradycardia to the peak RR value within the bradycardia episode. In order to reduce the variability of the first sample of the series, the mean, determined within 5 s before the start of each segment, was removed for all segments. This is particularly important for the estimation of the first state (π_i) of the models.

In the test phase, HMM and HSMM based detectors are applied to the totality of the 148 RR time series (series duration = 26.25 ± 11.37 minutes) with 233 bradycardia episodes. Detectors are applied to these series as described in the previous section, with a sliding window of size $T = 7$ s.

True positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) were determined for each sample by comparing the obtained detections with the available annotations. TP occur when a detection falls within a 20 s window, centered at a given annotation. Even if this window seems particularly large, we justify this choice by the fact that we are also evaluating the detection delay as an important marker for the selection of the optimal detection method.

Detection performance was evaluated by estimating the sensitivity (Se) and specificity (Sp) of each detector, for different detection thresholds, and represented by means of ROC curves. The detection delay (dd) is also determined, and defined as the time elapsed between the annotation instant and the detection. Optimal performances were evaluated using the shortest distance to perfect detection (D) according to Eq. 5. Detection delay results are measured in seconds.

$$D(\delta^{\alpha,k}) = \min_{\delta^{\alpha,k}} \sqrt{(1 - Se(\delta^{\alpha,k}))^2 + (1 - Sp(\delta^{\alpha,k}))^2} \tag{5}$$

3. Results

In this section, K is set to two for HMM and HSMM. Each parameter set λ^k was estimated to reproduce the dynamics of LSk, $k \in \{1, 2\}$. Only one threshold is thus used, $\delta^{1,2}$. M^k was estimated using the BIC. $T = 7$ s sliding window was used to construct $O_{t-T+1:t}$ and to determine L_t^k.

ROC curves are shown in Fig. 2 and its respective mean detection delays are shown in Fig. 3. Optimal parameters were tuned: $\tau = 0,67$ s, $\Delta_{\text{UQ}} = 1300$ and $\delta_{\text{QNU}} = 0.05$. Conventional detection methods (fixed and relative thresholds) are also shown as a reference. In these figures, the point where D occurs is shown with character "x".

Detection results were similar when using the approaches based on HMM or HSMM. The performance attained with these methods is higher than that obtained from conventional detection methods used in NICU, with a detection time occurring 2 seconds earlier. In term of sensitivity and specificity, the best detection performance is achieved using the HMM based detector when the RR time series are quantized uniformly, and it is slightly higher than that obtained by the equivalent HSMM approach when time-delay versions of the RR time series are taken into account. In term of detection delay, the HSMM based detector with time-delay versions of the RR time series is the most predictive.

4. Conclusions and future works

In this work, hidden Markov models and hidden semi-Markov models are used to characterize the dynamics of RR time series, extracted from the ECG of 32 preterm

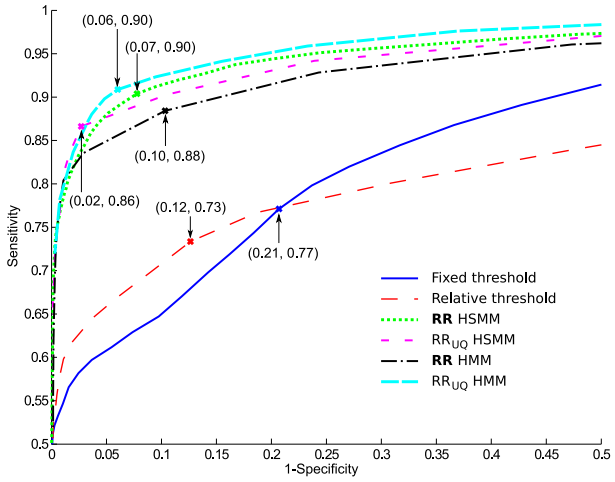

Figure 2. ROC curves for apnea-bradycardia detection. Character "x" represents the point where D occurs.

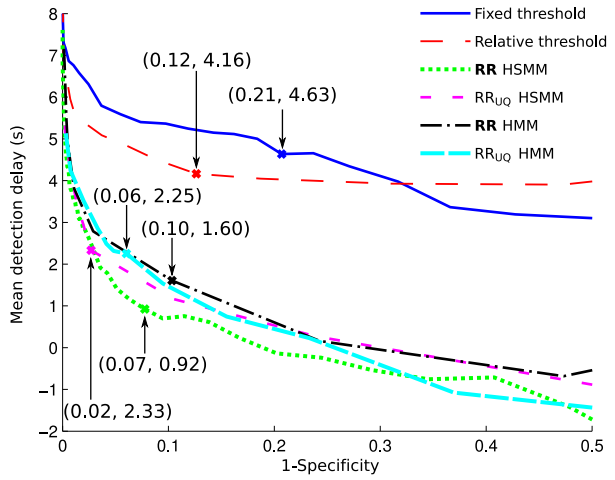

Figure 3. Mean detection delay of apnea-bradycardia episodes. Character "x" represents the point where D occurs.

infants, monitored in a neonatal intensive care unit. These models were subsequently employed to detect 233 apnea-bradycardia episodes presented on these RR time series, through the on-line comparison of log-likelihoods of two competitive models (reference and bradycardia models) with respect to a threshold. Additionally, a preprocessing phase is employed to improve the detection performance.

Results obtained show that the performance detection are comparable when hidden Markov models or hidden semi-Markov models are used as detectors. However, hidden semi-Markov model approach produce the best compromise in term of sensitivity, specificity and detection delay, when the RR time series and the delay version of it are integrated into the observation matrix. The performance obtained from both of these methods is higher than that obtained from traditional approaches, based on fixed or relative thresholds.

Current work is directed towards the clinical evaluation of these methods in a prospective trial.

Acknowledgements

The authors would like to thank the ECOS NORD project V09S04 for partly support this work.

References

[1] Janvier A, Khairy M, Kokkotis A, Cormier C, Messmer D, Barrington KJ. Apnea is associated with neurodevelopmental impairment in very low birth weight infants. J Perinatol 2004;24(12):763–8.

[2] Pillekamp F, Hermann C, Keller T, von Gontard A, Kribs A, Roth B. Factors influencing apnea and bradycardia of prematurity-implications for neurodevelopment. Neonatology 2007;91(3):155.

[3] Poets CF, Stebbens VA, Samuels MP, Southall DP. The relationship between bradycardia, apnea, and hypoxemia in preterm infants. Pediatr Res 1993;34(2):144–7.

[4] Cruz J, Hernandez A, Wong S, Carrault G, Beuchée A. Algorithm fusion for the early detection of apnea-bradycardia in preterm infants. In Computers in Cardiology, volume 33. 2006; 473–476.

[5] Portet F, Gao F, Hunter J, Sripada S. Evaluation of on-line bradycardia boundary detectors from neonatal clinical data. Conf Proc IEEE Eng Med Biol Soc 2007;2007:3288–91.

[6] Pichardo R, Adam JS, Rosow E, Bronzino J, Eisenfeld L. Vibrotactile stimulation system to treat apnea of prematurity. Biomed Instrum Technol 2003;37(1):34–40.

[7] Altuve M, Carrault G, Beuchée A, Pladys P, Hernández A. On-line apnea-bradycardia detection using hidden semi-markov models. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. 2011; 4374–4377.

[8] Schwarz G. Estimating the dimension of a model. The annals of statistics 1978;6(2):461–464.

[9] Forney Jr G. The viterbi algorithm. Proceedings of the IEEE 1973;61(3):268–278.

[10] Altuve M, Carrault G, Cruz J, Beuchée A, Pladys P, Hernández A. Multivariate ecg analysis for apnoea-bradycardia detection and characterisation in preterm infants. International Journal of Biomedical Engineering and Technology 2011;5(2/3):247–265.

[11] Dumont J, Hernández AI, Carrault G. Improving ecg beats delineation with an evolutionary optimization process. IEEE Trans Biomed Eng 2010;57(3):607–15.

Address for correspondence:

Miguel Altuve
Grupo de Bioingeniería y Biofísica Aplicada
Universidad Simón Bolívar, Caracas 1080, Venezuela.
maltuve@usb.ve