ECG segmentation and fiducial point extraction using multi hidden Markov model

Mahsa Akhbari, Mohammad B. Shamsollahi, Omid Sayadi, Antonis A. Armoundas, Christian Jutten

ECG Segmentation and Fiducial Point Extraction Using Multi Hidden Markov Model

Mahsa Akhbari[a,b,*], Mohammad B. Shamsollahi[a], Omid Sayadi[c], Antonis A. Armoundas[c], Christian Jutten[b]

[a] *BiSIPL, Department of Electrical Engineering, Sharif university of Technology, Tehran, Iran.*
[b] *GIPSA-Lab, Grenoble, and Institut Universitaire de France, France.*
[c] *Cardiovascular Research Center, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129,USA.*

Abstract

In this paper, we propose a novel method for extracting fiducial points (FPs) of electrocardiogram (ECG) signals. We propose the use of multi hidden Markov model (MultiHMM) as opposed to the traditional use of Classic HMM. In the MultiHMM method, each segment of an ECG beat is represented by a separate ergodic continuous density HMM. Each HMM has different state number and is trained separately. In the test step, the log-likelihood of two consecutive HMMs is compared and a path is estimated, which shows the correspondence of each part of the ECG signal to the HMM with the maximum log-likelihood. Fiducial points are estimated from the obtained path. For performance evaluation, the Physionet QT database and a Swine ECG database are used and the proposed method is compared with the Classic HMM and a method based on partially collapsed Gibbs sampler (PCGS). In our evaluation using the QT database, we also compare the results with low-pass differentiation, hybrid feature extraction algorithm, a method based on the wavelet transform and three HMM-based approaches. For the Swine database, the root mean square error (RMSE) values, across all FPs for MultiHMM, Classic HMM and PCGS methods are 13, 21 and 40 msec, respectively and the MultiHMM exhibits smaller error variability than other methods.

*Corresponding author
Email addresses: akhbari.mahsa@gmail.com (Mahsa Akhbari), mbshams@sharif.edu (Mohammad B. Shamsollahi), sayadi.omid@mgh.harvard.edu (Omid Sayadi), aarmoundas@partners.org (Antonis A. Armoundas), christian.jutten@gipsa-lab.grenoble-inp.fr (Christian Jutten)

For the QT database, RMSE values for MultiHMM, Classic HMM, Wavelet and PCGS methods are 10, 17, 26 and 38 msec, respectively. Our results demonstrate that our proposed MultiHMM approach outperforms other benchmark methods that exist in the literature; therefore can be used in practical ECG fiducial point extraction.
Keywords: Electrocardiogram (ECG), Hidden Markov Model (HMM), MultiHMM, Segmentation, Fiducial Point (FP) Extraction.

1. Introduction

The electrocardiogram (ECG) is used for measuring the electrical activity of the heart. ECG signal is obtained non-invasively by a simple device and provides valuable information about the health and heart diseases in humans. Acquiring the ECG signal and using its information are inexpensive and helpful [1].

Measurements used by cardiologists for detecting pathological beats and heart diseases are actually based on features like heart rate variability, and various intervals or segments between waves of successive beats. In this purpose, it is mandatory to be able to accurately estimate onset, offset and peak locations of the P, Q, R, S and T waves of each ECG. ECG segmentation and finding the onset and offset of ECG waves are difficult task due to lack of precise definition for onset and offset of some ECG waves, for example, there is no exact definition for the offset of QRS complex and T-wave [1].

Several techniques have been proposed for QRS complex detection including filtering and derivation, adaptive filtering, dynamic programming, classification methods, mathematical morphology methods and transformations [2, 3]. Low pass differentiation (LPD) [4], hidden Markov models [5, 6, 7, 8, 9, 10, 11, 12, 13], partially collapsed Gibbs sampler (PCGS) [14, 15], wavelet transform [16, 17, 18], correlation analysis [19, 20], support vector machine (SVM) [21], empirical mode decomposition (EMD) [22] and extended Kalman filter (EKF) [23, 24, 25] are also used for ECG segmentation and fiducial point (FP) extraction.

Finding the onset, offset and peak of ECG waves is known as fiducial point extraction which can be used as a preprocessing step in many applications [26]. In [27], the authors first extract some features from ECG signals such as P-wave, QRS com-

plex, T-wave amplitude and duration. After that they used the extracted features for detection of fragmented QRS complex. In [28], the authors used the initial estimation of ECG waves and their onset and offset locations for mobile health care applications. They used both time and frequency analysis and called it as a hybrid feature extraction algorithm (HFEA). Onset and offset of the P-wave and QRS complex were used as the input to the model which was proposed by Bono et al. [29] for a "Selvester QRS scoring" system. Finally, Kumar et al. [30] used the onset and offset of ECG waves for ischemia detection.

Hidden Markov model (HMM) is a model for describing the process which is not directly observable but can be observed with sequence of symbols [31]. HMMs were used for several applications: speech recognition [32], apnea identification [33], apnea-bradycardia detection in preterm infants [34, 35, 36], segmentation of heart sound recordings [37], estimation of fetal cardiac timing events [38] and FP extraction [7].

HMM is one of the approaches which is used for ECG segmentation. In most of the previous HMM-based approaches [6, 9], each ECG beat is modeled with a single HMM and ECG waves and baselines are considered as states of a HMM model. In these approaches, ECG beats are considered as an observation of HMM model and parameters of HMM are found using training data set with supervised or unsupervised learning methods. In the test step, ECG segmentation is done using the inference algorithms.

Supervised learning methods require to accurately label the observations. In contrast, unsupervised learning methods work automatically and do not require the labels of observation symbols and the relevant hidden states, but these methods may suffer from falling into local maxima due to the ill-suited initial values. Hence, in some cases the obtained results are not accurate, especially for the ECG segmentation and fiducial point extraction [11].

It is worth noting that: (i) HMMs have been used in previous works, for ECG segmentation and detection of ECG waves [6, 7, 8, 9, 13], or for beat detection and classification [5, 7, 9, 11], while our work is focused on fiducial point extraction, which is a much more complex task. Only [7] proposed a HMM model for such purpose, but considering wavelet transform of the ECG signal, (ii) Most of these studies are based on supervised learning approach which need the accurate labels of expert and are time

consuming, (iii) In some works [6, 7, 8, 9] encoded ECG by the wavelet transform or the coefficients of wavelet in different scales are used as an observation of HMM models, (iv) Some works [6] use hidden semi-Markov model to improve the results and solve the "double beat segmentation" problem.

Conversely, we will show that the proposed approach has many advantageous over previous methods. It is used for ECG fiducial point extraction, it uses raw ECG signal as an observation of HMM and finally can solve the double beat segmentation problem and also can accurately estimate fiducial points for many pathological beats.

In this paper, the approach for extracting ECG fiducial points is based on HMM, too. It is called "MultiHMM" since one HMM model is considered for each ECG segment and in the training step, a rough segmentation is performed to define the training data for each HMM. Then, the Baum-Welch algorithm is used to find the parameters of each HMM, separately. Afterwards in the test step, the label of the current beat segment (i.e., the most appropriate HMM model) is estimated through comparison of log-likelihood of HMMs.

The performance of the proposed method is compared with previously published methods, including Wavelet [17], LPD [4], PCGS [14], HFEA [28], three HMM-based approaches [7] and "Classic HMM". Validation and comparison are done on the Physionet QT database [39, 40] and an annotated Swine ECG database [41].

The rest of this paper is organized as follows: Related work, essentially methods used in performance comparison, are described in Section 2. The proposed method is explained in Section 3. Section 4 presents the experimental results, and finally section 5 concludes the paper.

2. Related Work

2.1. A method based on wavelet transform

In [17], a method based on the wavelet transform is used for finding the fiducial points of ECG waves. In this method, wavelet decomposition into 5 scales ($2^1 - 2^5$) is used. Because most of the energy of QRS complexes lies in scales $2^1 - 2^4$ and for P and T waves, most of the energy lies within scales $2^4 - 2^5$. Local maxima, minima and

zero crossings at different scales are used to detect the QRS complexes, P- and T-waves and their peak, onset and offsets.

2.2. Partially Collapsed Gibbs Sampler Method (PCGS)

Lin et al. [14] proposed a method based on partially collapsed Gibbs sampler (PCGS) to delineate P- and T-waves and find their peak, onset and offset. In this model, the proposed algorithm first detects the QRS complexes, then constructs two search blocks for P- and T-waves, finally uses Bayesian inference in each block to delineate the P- and T- waves. This model uses prior distribution of wave locations, amplitude and waveform coefficients. Detection of P and T waves are based on using theses prior distributions and the likelihood of observed data.

2.3. HMM-based Methods

2.3.1. Review on mathematical equations of HMM

A discrete density HMM is characterized by the following parameter set: $\lambda = (A, B, \pi)$ where A is the matrix of state-transition probabilities, B is the observation probability, and π is the initial state probability [32].

In some applications, the observations are continuous signals (or vectors) and it would be advantageous to be able to use HMMs with continuous observation densities [32]. The most general representation of the model probability density function (pdf) is a finite mixture of the form:

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm}\, \mathscr{N}[\mathbf{O}, \mu_{jm}, \mathbf{U}_{jm}], 1 \leq j \leq N \tag{1}$$

where $\mathbf{O}$ is the vector being modeled, c_{jm} is the mixture coefficient for the m^{th} mixture in state j and $\mathscr{N}$ is Gaussian model, with mean vector μ_{jm} and covariance matrix $\mathbf{U}_{jm}$ for the m^{th} mixture component in state j. The usual observation model is a weighted mixture of Gaussian distributions. The mixture gains c_{jm} satisfy the stochastic constraint

$$\begin{aligned} &\sum_{m=1}^{M} c_{jm} = 1, 1 \leq j \leq N \\ &c_{jm} \geq 0, 1 \leq j \leq N, 1 \leq m \leq M \end{aligned} \tag{2}$$

so that the pdf is properly normalized, i.e.,

$$\int_{-\infty}^{\infty} b_j(x)dx = 1, 1 \leq j \leq N \tag{3}$$

We use the compact notation $\lambda = (A, \mu_{jm}, \mathbf{U}_{jm}, \pi)$ to indicate the complete parameter set of the model.

2.3.2. Previous HMM-based methods

In 1990, Coast et al. [5] proposed a Markov model for cardiac arrhythmia analysis. Hughes et al. [6] used HMM for ECG segmentation. In their first model, they considered raw ECG as an observation of HMM. After that, they improved the results by applying HMM on the wavelet encoded ECG and also applying hidden semi-Markov model (HSMM) on the wavelet encoded ECG. Andreao et al. [7] proposed three HMM-based approaches for finding the onset and offset of ECG waves: (i) generic HMM training, (ii) individual's HMM training, and (iii) generic HMM adapted to each individual. Krimi et al. [8] used the combination of the wavelet transform and HMM for ECG segmentation. They first used the wavelet transform to find the edge and peaks of ECG signal, then the features extracted from the edges serve as inputs for the HMM [8]. Andreao et al. [9] also combined the wavelet transform and HMM for ECG beat segmentation and classification. Thomas et al. [10] proposed two HMM-based approaches for ECG interval analysis. In the first one, called generic, a global model which is a concatenation of six HMMs, is built. The resulting global HMM can be regarded as a hierarchical HMM and the decision for a new ECG beat is made using this HMM model. In the second one, called clustering, ten classes of ECG beats are generated and the decision for a new ECG beat is made after clustering it. Liang et al. [11] proposed a two-layered HMM algorithm for ECG feature extraction and classification. In the first HMM layer, the ECG signals are segmented into baseline intervals, P-wave, QRS complex and T-wave, respectively. Then the corresponding interval features are used to classify the ECG into normal or abnormal types in the second HMM layer [11]. Li et al. [13] proposed an HMM-based approach for ECG segmentation. They first estimated the QRS complexes. After that, based on the detected R peaks, the ECG data are segmented. By using a heuristic rule segmented ECG is classified to N groups. The

classification is based on the length of the RR-intervals and each group includes ECG data with similar RR-intervals and temporal features. A separate HMM is defined for each group and is only used for extracting the ECG characteristic waves of signals of that group. The authors presented the sensitivity and positive predictive for detecting ECG waves but they did not estimate the exact location of ECG fiducial points. Altuve et al. [36] proposed a model with several hidden semi-Markov models for online apnea bradycardia detection in preterm infants.

Here, we discuss a widely-used ECG FP extraction method based on HMM. In this model, which referred to as "Classic HMM", a left-right continuous density HMM with seven states, corresponding to B_1, P, PQ, QRS, ST, T and B_2 segments of an ECG beat, is considered (Fig. 1). This structure is almost similar to the structure which has been used in [6, 8, 9, 11] although the aim of these works are not FP extraction. Fig. 2 shows these seven ECG segments. The four baselines are defined as below: B_1: segment from beginning of beat to P_{on}, PQ: segment from P_{off} to QRS_{on}, ST: segment from QRS_{off} to T_{on} and B_2: segment from T_{off} to end of beat.

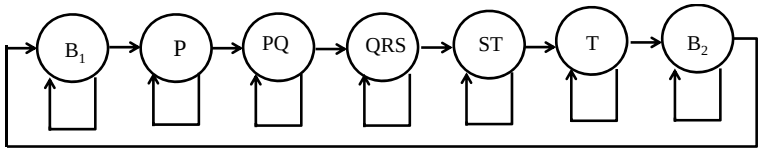


Figure 1: A left-right continuous density HMM with 7 states for Classic HMM.

In Classic HMM, the labeled data set of ECG waveforms is used and a HMM model is trained. The observations of a HMM are a continuous signal, modeled by a Gaussian mixture model (GMM). In order to find the suitable number of Gaussians for GMM, the Akaike information criterion (AIC) [42] or the Bayesian information criterion (BIC) [43] is used. Once the model has been trained, the Viterbi algorithm [32] is used to infer the optimal state sequence for each beat of the signals in the test set. The obtained optimal state sequence (estimated path) has seven levels, each one associated to one segment. Levels 1 to 7 represent the B_1, P, PQ, QRS, ST, T and B_2 segments, respectively. The proposal to find the onset and offset of waves from the

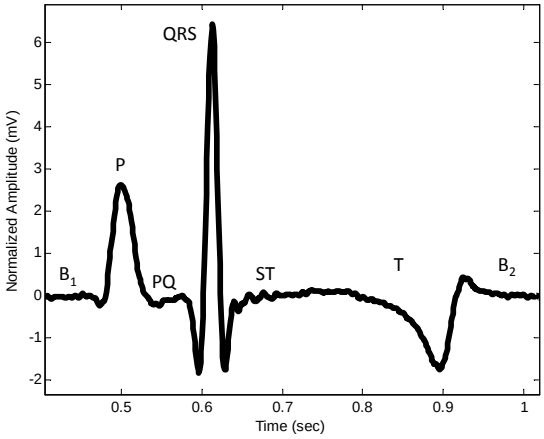


Figure 2: Segments of a single ECG beat.

estimated path is as follows:

- P_{on}: The point in which the path transits from level 1 to 2.
- P_{off}: The point in which the path transits from level 2 to 3.
- QRS_{on}: The point in which the path transits from level 3 to 4.
- QRS_{off}: The point in which the path transits from level 4 to 5.
- T_{on}: The point in which the path transits from level 5 to 6.
- T_{off}: The point in which the path transits from level 6 to 7.

Since the peaks can be positive or negative, peak position of waves ($P_{peak}, R_{peak}, T_{peak}$) are defined as the maximum of absolute value of signal between onset and offset.

3. Proposed Method (MultiHMM)

3.1. Methodology of MultiHMM

In the MultiHMM method, each segment of an ECG beat (Fig.2) is represented by a separate ergodic continuous density HMM. Similar state numbers are not assumed for different HMMs. The AIC or BIC criterion is used to obtain a rough estimation of the

number of states, and the exact number of states in each HMM is found experimentally in the training step. First we detect the R-peaks of ECG beats and associate a linear phase between $-\pi$ to π to it, similar to Sameni et al. [44] (R-peaks have phase equal to 0, beginning and end of the beats have phase equal to $-\pi$ and π, respectively.) According to the phase transitions from π to $-\pi$, we can find the beginning and end of beats. The onset and offset of ECG waves are annotated by physicians and from the ECG segments, we can construct the train data for each HMM as follows: training data of the first HMM is constructed from the B_1 segments of all beats and training data of the second HMM is constructed from the P segments of all beats, etc. We use the Baum-Welch algorithm [32] to find the HMM parameters: $\lambda_{B_1}, \lambda_P, \lambda_{PQ}, \lambda_{QRS}, \lambda_{ST}, \lambda_T$ and λ_{B_2} $(\lambda_1, \cdots, \lambda_7)$. λ_k is defined as $\lambda_k = (A_k, \mu_{jmk}, \mathbf{U}_{jmk}, \pi_k)$, $k = 1, 2, \ldots, 7$. We use the HMM toolbox written by Kevin Murphy [45] for training the HMMs.

Fig. 3 shows the blockdiagram of our proposed MultiHMM approach for finding the peak, onset and offset of ECG characteristic waveforms.

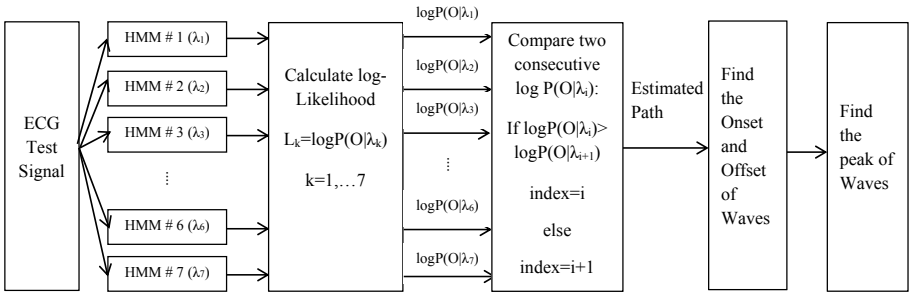


Figure 3: Blockdiagram of the proposed MultiHMM approach for finding the peak, onset and offset of ECG characteristic waveforms.

After training all HMMs, we use test data and define a sliding window with length "n_w" and consider the data inside the window as the observation of HMMs (O with length n_w). The length of sliding window is fixed. Each window has $n_w - 1$ overlapping samples with previous window and only one sample differs between two consecutive windows. We then compute the log-likelihood of each HMM as:

$$L_k = \log P(O_{1:n_w}|\lambda_k),\ k \in \{1, 2, \cdots, 7\} \tag{4}$$

where $P(O_{1:n_w}|\lambda_k)$ is the probability that the observation sequence $O_{1:n_w} = O_1, O_2, \cdots, O_{n_w}$ is generated by the model with parameters λ_k. Afterwards, we compare the log-likelihood of two consecutive HMMs and choose the HMM with the maximum log-likelihood:

$$index = \underset{k}{\operatorname{argmax}} \ \log P(O_{1:n_w}|\lambda_k),\ k \in \{i, i+1\} \tag{5}$$

where i is the number of the current HMM.

The procedure of finding the path is done for each ECG beat separately. Since each ECG beat starts with B_1 segment, hence we assume that the first observation sequence $O_{1:n_w}$ is in B_1 and at the beginning we set "index=1". Then, we compare the log-likelihood of two consecutive HMMs: HMM_1 and HMM_2, i.e. $k \in \{1,2\}$ in (5) and the result will be $index = 1$ or $index = 2$. We start to compare the next two HMMs ($k \in \{2,3\}$ in (5)), when we achieve $index = 2$ for at least mm times. mm is a parameter which is defined experimentally smaller than n_w and prevents oscillations between two successive indexes. Finally, a path is estimated which shows the correspondence of each part of the ECG signal to the HMM with the maximum log-likelihood. The estimated path has seven levels, each one associated to one HMM (one ECG segment). Levels 1 to 7 represent the B_1, P, PQ, QRS, ST, T and B_2 segments, respectively. The onset and offset of the P-wave, QRS complex and T-wave are found from the transitions of one level to upper level in this path (same as for the Classic HMM which is explained in Section 2.3). Peak position of waves are defined as the maximum absolute value of signal between onset and offset of waves.

3.2. Data and Evaluation Metrics

To evaluate the performance of the proposed method in extracting ECG fiducial points, the following two databases are used which include ECG signal annotations by physicians: the Physionet QT database (human ECG) [39, 40] and a Swine ECG database (Swine ECG) [41]. The Swine database includes ECG signals acquired during acute myocardial infarction, which exhibit significant morphologic changes (such as ST elevation and QT prolongation). Records of this database are sampled at 1000 Hz and each record of each subject has 200 annotated beats. Records of the QT database are sampled at 250 Hz (one sample=4 ms) and each record has 30-50 annotated beats.

As a pre-processing step, the ECG mean is removed and its variance is set to one. The baseline wander of signal is also removed by median filter which is available in the "open-source electrophysiological toolbox (OSET)" [46], and its length is $0.3f_s$ (f_s is sampling frequency).

For quantitative evaluation of a FP extraction method, we calculate estimation error defined as time differences between cardiologist annotations (considered as ground truth) and results of the method. Quantitative results are reported using common metrics: mean (m), standard deviation (s) and root mean square error (RMSE), defined as:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(e_j)^2} = \sqrt{(m^2+s^2)} \tag{6}$$

where $e_j = \hat{y}_j - y_j$ is denoted as the j^{th} element of the estimation error vector and N is the length of the error vector (number of annotations). y_j and $\hat{y}_j$ are the j^{th} cardiologist annotation and estimated point, respectively. m, s and $RMSE$ are given in millisecond (ms). Since the RMSE considers both mean and standard deviation of error, it is a more relevant parameter for comparing the methods.

Some authors considered the values given by the "CSE working party [1]" in [47, 48] as a reference for delineation error tolerances. In [47], it is stated that "the standard deviation of the differences [of an algorithm results] from the reference (s) should not exceed certain limits ($2s_{CSE}$)". The limits given in [47], are obtained as two standard deviations of the differences (in ms) between the median of the individual readers and the final referee estimates [17]. These results take into account the large variability in expert annotations.

As a consequence, we can consider that, for being competitive with a good expert, an algorithm must achieve $s < 2s_{CSE}$ ("loose criteria") or strictly $s < s_{CSE}$ ("strict criteria"): in Sections 4.1 and 4.2, we will discuss about these criteria for the records of the Swine and QT database, respectively.

To assess the degree of agreement between each of the automated methods and

[1] Common standards for quantitative electrocardiography (CSE) is an international project initiated by European community.

the manual annotations, we use the Bland-Altman approach [49] to estimate the mean difference and the standard deviation of the difference among all annotations of physicians, across all subjects. The mean of the estimation error and the limits of agreement (defined as twice the standard deviation of the estimation error) are estimated for different methods and discussed in Sections 4.1 and 4.2 for both databases.

We will also use the Wilcoxon rank-sum test with Bonferroni correction [50] to statistically compare all method pairs.

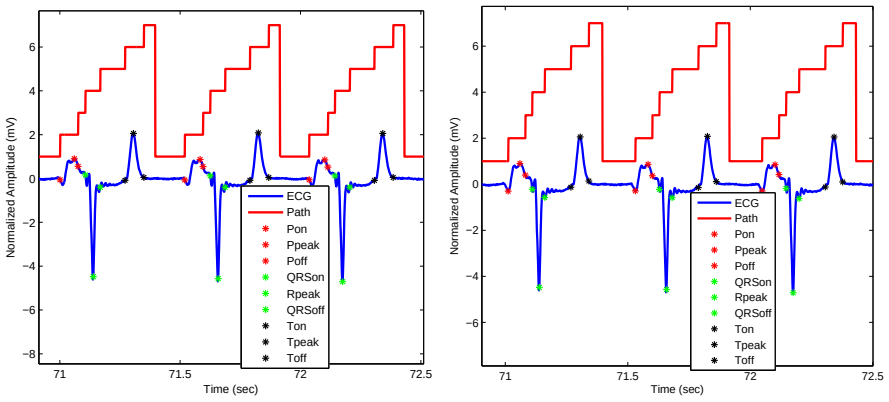
4. Results

4.1. Results for the Swine database

Fig.4.(a) shows the estimated path by the Classic HMM for a small segment of the record Ischemia09 of the Swine database. It also shows the estimated fiducial points by the Classic HMM which are found from the estimated path. Fig.4.(b) shows the estimated path and FPs by the MultiHMM approach for this record. It is worth to mention that these subfigures are illustrative examples of what the estimated path looks like and clarify how the onset and offset of waves can be found from the transition of one level to upper level in a multi-level estimated path. In this example, the two methods achieve good (and thus similar) results in FP estimation.

Here, we use 5-fold cross validation [51] for training the MultiHMM for each subject, i.e., for each record. The performance of different methods for ECG FP extraction in the Swine database are compared in Table 1, where the best results of RMSE values are denoted as bold. We see that for all FPs except T_{off}, the MultiHMM achieves the least RMSE value and exhibits smaller error variability than others.

The mean and standard deviation of aggregate results across all FPs are calculated and Bland-Altman analysis, which is briefly presented in Section 3.2, is performed. The mean of estimation error and the limits of agreement (twice of standard deviation) estimated for MultiHMM, Classic HMM and PCGS methods are equal to 2.2 ± 27, 1.6 ± 42 and 6.9 ± 78 ms, respectively. The RMSE values across all FPs for above-mentioned methods are equal to 13, 21 and 40 ms, respectively. We observe that the



(a) (b)

Figure 4: Estimated path and fiducial points by the (a) Classic HMM and (b) MultiHMM method for the record Ischemia09 of the Swine database.

limits of agreement and RMSE values for MultiHMM are smaller than those for others, indicating the superior performance of the proposed method in extracting FPs.

For all FPs, standard deviation of the MultiHMM method is below the CSE loose criteria (last row of Table 1), which is not the case for the other methods. It means that results provided by the MultiHMM method is competitive with result obtained by a good physician expert. The MultiHMM also satisfies the "strict criteria" for T_{off}.

Finally, pairwise comparisons using the Wilcoxon rank-sum test show a statistically significant difference between any two methods (p-value < 0.0001).

Table 1: Mean $\pm$ standard deviation (first line) and RMSE (second line) of error in ms between estimated FPs and manual annotations for signals of the Swine database (fs=1000Hz), (N.A.: Non Available)

Method	P_{on}	P_{peak}	P_{off}	QRS_{on}	R_{peak}	QRS_{off}	T_{on}	T_{peak}	T_{off}
MHMM	7 ± 6	2 ± 2.7	-2.3 ± 7	-0.7 ± 6	0.9 ± 0.3	0.3 ± 11	25 ± 21	0.03 ± 4	-13 ± 8
	10	**3**	**7**	**6**	**1**	**11**	**32**	**4**	15
CHMM	6 ± 15	2 ± 2.7	-6 ± 8	-0.5 ± 11	0.9 ± 0.3	0.05 ± 23	22 ± 47	0.07 ± 4	-10 ± 10
	16	**3**	10	11	**1**	23	52	**4**	**14**
PCGS	4 ± 19	3 ± 6.7	16 ± 16	*N.A*	*N.A*	*N.A*	37 ± 52	-8 ± 50	-11 ± 42
	19.5	7	22	*N.A*	*N.A*	*N.A*	64	51	44
$\mathbf{2s_{CSE}}$	10.2	–	12.7	6.5	–	11.6	–	–	30.6

4.2. Results for the QT database

Here, we use 2-fold cross validation for training the MultiHMM for each subject, i.e., for each record of the QT database. Each record has 30-50 annotated beats. We separate the data into two parts with equal size, we then train on first part and test on second part, followed by training on second part and testing on first part, and finally find the estimation error vector for each record. After that we aggregate the error vector for all records and find the mean, standard deviation and RMSE of total error across all records. The performance of different methods for ECG FP extraction in the QT database are compared in Table 2. Since in the QT database, the physician annotations for T_{on} are not available, therefore we can not estimate the estimation error for T_{on}. The table 2 is split in different parts which differ by the number of records of the QT database used in each experiment. In this table, rows 1 to 4 represent the results obtained using MultiHMM, Classic HMM, Wavelet and PCGS methods, respectively, on records of Arrhythmia and Normal Sinus Rhythm databases which are annotated in the QT database (19 records). "*" in rows 3 and 4 of this table, indicates that these results are obtained by using MATLAB codes provided by the authors of [17] and [14] for 19 records. The least RMSE values among rows 1 to 4 are denoted in bold.

Rows 5 to 7 of Table 2, represent the obtained results of Wavelet, PCGS and LPD methods, respectively, for all records of the QT database which are reported in [17], [14] and [4], respectively. Row 8 of the Table represents the results of hybrid feature extraction algorithm (HFEA) method for 27 records of the QT database which is reported in [28]. Finally, rows 9 to 11 of this Table, represent the results of HMM-based approaches which are reported in [7]. They considered three cases: (i) generic HMM training, (ii) individual's HMM training and (iii) generic HMM adapted to each individual. In rows 5 to 11, red values show the RMSE values which are less than RMSE values of MultiHMM method [2].

According to the results of Table 2, for the MultiHMM method, the mean errors for all FPs are smaller than or around one sample (4 ms). The standard deviations are

[2]It is worth to mention that the number of beats used by the different authors are quite different (and we do not know how the beats are selected or rejected) and consequently the comparison is not very easy.

Table 2: Mean ± Standard deviation (first line) and RMSE (second line) of error in ms between estimated FPs and manual annotations for signals of the QT database (fs=250Hz), (N.A.: Not Available). '*' in rows 3 and 4 indicates that these results are obtained for 19 records of the QT database.

Method	P_{on}	P_{peak}	P_{off}	QRS_{on}	R_{peak}	QRS_{off}	T_{peak}	T_{off}
MHMM	4 ± 12	0.2 ± 3.5	-3 ± 11	-5 ± 10	0 ± 0.2	1.5 ± 11.5	-0.4 ± 5.6	-5 ± 14
	12.5	**3.4**	**12.2**	**11**	**0.2**	**11.6**	**5.6**	**14.7**
CHMM	10 ± 16	0.2 ± 4.2	-5 ± 11	13.5 ± 12.7	0 ± 0.2	-3 ± 33	-0.5 ± 5.6	-14 ± 15
	19	4.3	12.6	18.5	0.2	33.2	5.63	21
WT*	-9 ± 37	-2 ± 29	3 ± 15	13 ± 14	1.5 ± 2	-1.8 ± 13	2 ± 30	8 ± 38
	38.4	29.1	15.8	19.3	2.6	13.3	30.3	38.4
PCGS*	-35 ± 31	5 ± 8	24 ± 15	*N.A*	*N.A*	*N.A*	5.6 ± 33	27 ± 49
	47	9.8	28.5	*N.A*	*N.A*	*N.A*	33.2	56.2
LPD	14 ± 13.3	4.8 ± 10.6	-0.1 ± 12.3	-3.6 ± 8.6	*N.A*	-1.1 ± 8.3	-7.2 ± 14.3	13.5 ± 27
	19.3	11.6	12.3	**9.3**	*N.A*	**8.4**	16	30.2
WT	2 ± 14.8	3.6 ± 13.2	1.9 ± 12.8	4.6 ± 7.7	*N.A*	0.8 ± 8.7	0.2 ± 13.9	-1.6 ± 18.1
	14.93	13.7	13	**9**	*N.A*	**8.7**	13.9	18.2
PCGS	3.7 ± 17.3	4.1 ± 8.6	-3.1 ± 15.1	*N.A*	*N.A*	*N.A*	1.3 ± 10.5	4.3 ± 20.8
	17.7	9.5	15.4	*N.A*	*N.A*	*N.A*	10.6	21.2
HFEA	-6 ± 12	5 ± 9	3 ± 16	4 ± 8	4 ± 10	12 ± 16	-15 ± 29	-16 ± 21
	14	10.7	16.3	**8.6**	10.5	20.6	33	26.6
HMM(i)	16 ± 18	*N.A*	-2 ± 15	11 ± 8	*N.A*	3 ± 10	*N.A*	3 ± 30
	24.2	*N.A*	15.4	14.4	*N.A*	**10.9**	*N.A*	30
HMM(ii)	1 ± 14	*N.A*	-5 ± 11	4.7 ± 7.8	*N.A*	-4 ± 9	*N.A*	15 ± 24
	14.5	*N.A*	12.3	**9.1**	*N.A*	**9.8**	*N.A*	28.3
HMM(iii)	12 ± 14	*N.A*	-6 ± 12	9 ± 8	*N.A*	2 ± 10	*N.A*	12 ± 21
	18.6	*N.A*	13.2	11.8	*N.A*	**10.5**	*N.A*	24.7
$\mathbf{2s_{CSE}}$	**10.2**	**–**	**12.7**	**6.5**	**–**	**11.6**	**–**	**30.6**

around three samples for the onset and offset of waves and around one sample for the peak of waves. Median of estimation error for all FPs except QRS_{on} are equal to zero. Any variation at the level of one sample is not significant.

Comparison of rows 1 to 4 of Table 2 shows that the RMSE values of MultiHMM for all FPs are less than others, especially for P_{on} and T_{off}. We observe that for all FPs MultiHMM has also smaller standard deviation than others: it means that the proposed MultiHMM can find FPs more accurately than previously described methods.

The comparison of the RMSE values of the MultiHMM with results of rows 5 and 6 of Table 2 shows that for all FPs except QRS_{on} and QRS_{off}, the MultiHMM method

achieves lower RMSE values than LPD and Wavelet methods and can estimate FPs more precisely. Comparison of the RMSE values of MultiHMM with results of PCGS, in row 7, shows that for all FPs, the MultiHMM method has better results than PCGS.

Comparing the results of the MultiHMM with results of the HFEA method, in row 8 of Table 2, shows the superiority of the MultiHMM for all FPs except QRS_{on}. Finally, comparison of the RMSE values of the MultiHMM with results of rows 9 to 11 of this Table shows that our proposed MultiHMM has better results than "generic HMM training" and "generic HMM adapted to each individual" approaches in rows 9 and 11 (except for QRS_{off}). We observe that for all FPs except QRS_{on} and QRS_{off}, MultiHMM has less RMSE than "individual's HMM training" approach in row 10.

The last row of Table 2 shows the CSE tolerance, which is described in Section 3.2. We see that for all FPs, RMSE of the MultiHMM is always smaller than those for other methods, and its standard deviation usually less or very close (except QRS_{on}) to CSE tolerance. The MultiHMM also satisfies the "strict criteria" for T_{off}.

Mean and standard deviation of aggregate results across all FPs are estimated for MultiHMM, Classic HMM, Wavelet and PCGS methods as -1 ± 10, 0.2 ± 17.6, 1.9 ± 26.2 and 5.5 ± 38 ms, respectively. RMSE values across all FPs for above-mentioned methods are estimated as 10.1, 17.6, 26.3 and 38.5 ms, respectively. We observe that standard deviation and RMSE values for the MultiHMM are smaller than others.

Pairwise comparisons using the Wilcoxon rank-sum test show a statistically significant difference between any two methods (p-value < 0.0001).

4.3. Classic HMM Limitation (double-beat segmentation)

For some (usually pathological) signals, the Classic HMM cannot estimate a suitable path and suffers from a problem which is named "double-beat segmentation". Such segmentations occur when the model incorrectly infers two (or more) beats where there is only a single beat present in that part of the signal [6].

Fig. 5.(a) shows the estimated path by the Classic HMM method for the record Ischemia05 of the Swine database. We see that, during a unique beat, the estimated path goes from 1 to 2, then 3,.... and reaches 7 and again goes to 1, 2,.... and reaches 7. In the second part of the estimated path the transitions between levels are so fast that

levels 3, 4, 5 and 6 appear only for one sample. In Fig.5.(a) the preliminary estimated onset and offset points which are found from the estimated path are shown. We see that for each onset or offset, two points are estimated, one of which with a wrong location should be canceled. Fig.5.(c) shows the final estimated onset and offset points (after omitting incorrect points) using colorful points and the physician labels using vertical lines. According to this figure, the Classic HMM achieves high error in estimating the QRS_{off} and T_{on}.

Fig. 5.(b) shows the estimated path, onset and offset points by the MultiHMM method for the same record Ischemia05. We see that the path and points are estimated correctly. Fig.5.(d) shows the estimated onset and offset points using colorful points and the physician labels using vertical lines. Consequently, conversely to the Classic HMM method, the MultiHMM can solve the double-beat segmentation problem and achieves a good FP extraction.

4.4. PCGS limitation in FP estimation of biphasic waves

Fig. 6 shows the estimated FPs by PCGS and MultiHMM methods for the records Ischemia06 and Ischemia07 of the Swine database. Here, the original labels are shown using color vertical lines and estimated onset, offset and peak of T waves are shown using stars.

In Fig.6 left, we see that the record Ischemia06 has a biphasic T-wave and physicians considered the positive peak as a label for T_{peak}, whereas the PCGS method estimates only the first part of the T-wave (negative peak). Therefore, the estimation error of the PCGS method for T_{peak} and T_{off} will be very high. Here, the MultiHMM method estimates T_{on}, T_{peak} and T_{off} more exactly than the PCGS method.

Fig.6 right, shows that the record Ischemia07 has also a biphasic T-wave and physicians considered the negative peak as a label for T_{peak}, whereas the PCGS method estimates only the last part of the T-wave (positive peak). Hence, the estimation error for T_{on} and T_{peak} will be very high. Also in this case the MultiHMM method estimates T_{on}, T_{peak} and T_{off} more precisely than the PCGS method. These figures show the superiority of the MultiHMM approach in estimating the fiducial points of signals with biphasic waves.

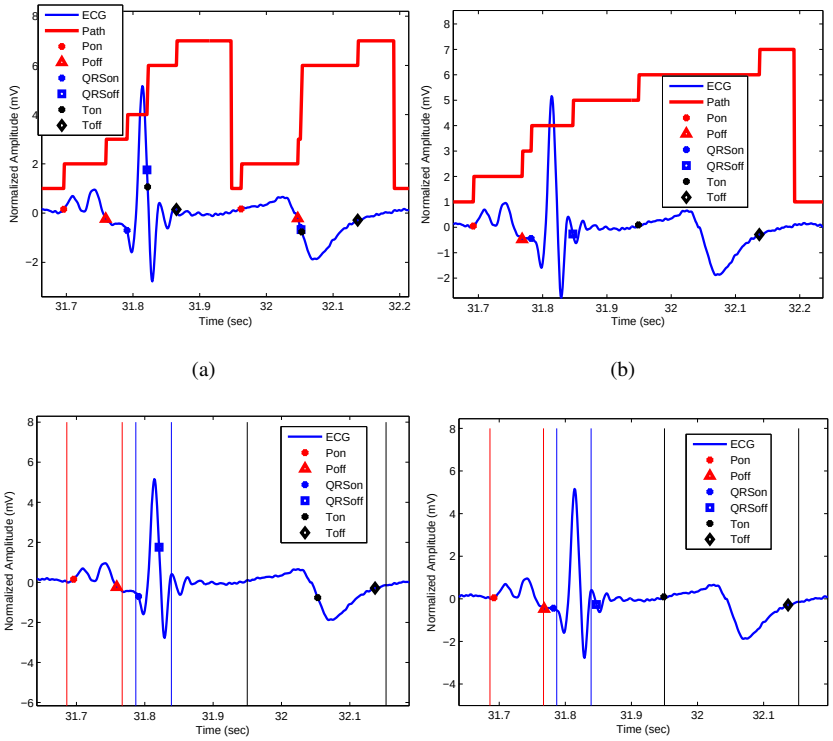
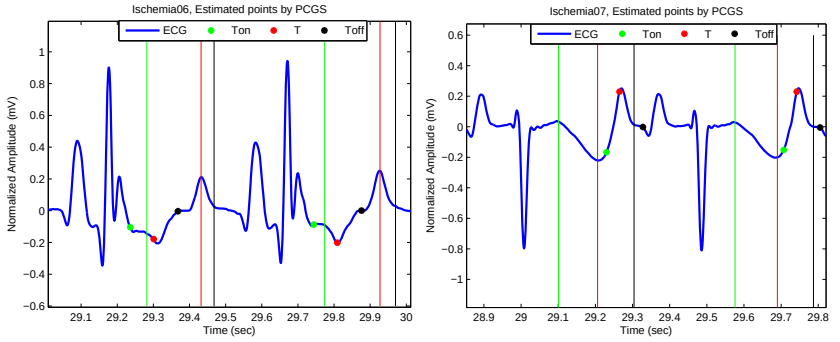


Figure 5: (a) Estimated path and preliminary FPs by Classic HMM (b) Estimated path and FPs by MultiHMM (c) Original and final FPs by the Classic HMM (d) Original and estimated FPs by MultiHMM for the record Ischemia05 of the Swine database. In this figure, the physician labels are shown using vertical lines.

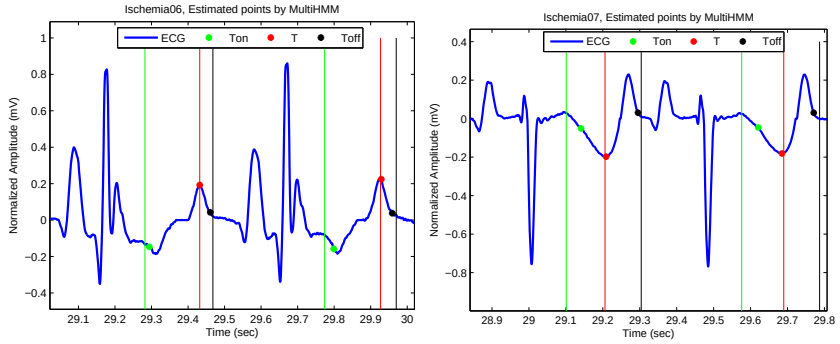
5. Discussion and Conclusions

In this paper, a novel method (MultiHMM) for ECG fiducial point extraction is proposed. Experiments carried out on ECG signals from QT and Swine databases show that the MultiHMM performance is better than the state of the art ECG delineators such as Classic HMM, PCGS, LPD, HFEA, Wavelet and three HMM-based approaches.

The main contribution of this paper is proposing a MultiHMM model for ECG FP extraction, which for each ECG wave or segment, a separate HMM is considered and the parameters of each HMM are trained separately. The number of states for HMM of baselines are 2-3, of P-wave and T-wave are 2-6 and of QRS complex are 4-8. It means



(a) (b)



(c) (d)

Figure 6: Original and estimated FPs by the PCGS for (a) Ischemia06 and (b) Ischemia07. Original and estimated FPs by the MultiHMM for (c) Ischemia06 and (d) Ischemia07. In this figure, estimated FPs are shown using stars and the physician labels are shown using vertical lines.

that for segments which have more complex shape like QRS complex, more states are required for modeling that segment by HMM.

Two parameters are defined in this paper: n_w which is the length of the window and mm which is a parameter smaller than n_w, used for preventing oscillations between two successive indexes. The value of these parameters are defined experimentally and for each record individually. For the records of Swine database, we have these values: f_s (sampling frequency)= 1 KHz, n_w=31 and $mm = 12$. For the records of QT database, f_s=250Hz, n_w= 21 or 16 (for some records n_w is 21 and for others is 16) and $mm = 6$.

After training all HMMs, we use test data and define a sliding window with length

"n_w" and consider the data inside the window as the observation of HMMs (O with length n_w). We then compute the log-likelihood of each HMM. Afterwards, we compare the log-likelihood of two consecutive HMMs and choose the HMM with the maximum log-likelihood. Finally, a path is estimated which shows the correspondence of each part of the ECG signal to the HMM with the maximum log-likelihood. The onset and offset of the P-wave, QRS complex and T-wave are found from the transitions of one level to upper level in this path. Peak position of waves are defined as the maximum absolute value of signal between onset and offset of waves.

The advantages of the proposed model are: (i) the ability to estimate the ECG FPs from raw ECG signals, while in several related work ([6, 7, 8, 9]), encoded ECG by the wavelet transform or the coefficients of wavelet in different scales are used as an observation of HMM models; (ii) the ability to successfully segment pathological beats (biphasic waves), while the PCGS method fails under these conditions; (iii) the ability to solve the double-beat segmentation problem, while for some signals, the Classic HMM suffers from this problem and obtains a high estimation error; (iv) the MultiHMM is not very sensitive to the number of each HMM's states while the Classic HMM is weakly sensitive to the number of Gaussian functions of GMM.

For the Swine database, the RMSE values across all FPs for MultiHMM, Classic HMM and PCGS methods are 13, 21 and 40 ms, respectively: the MultiHMM method is then much more accurate than other methods.

For the QT database, RMSE values across all FPs for MultiHMM, Classic HMM, Wavelet and PCGS methods are 10, 17, 26 and 38 ms, respectively: again the MultiHMM method is much more accurate than other methods. For the MultiHMM, the mean errors for all FPs are smaller than or around one sample (4 ms). The standard deviations are around three samples for the onset and offset of waves and around one sample for the peak of waves. Median of estimation error for all FPs except QRS_{on}, are equal to zero, which shows the superiority of the MultiHMM method over others.

For both databases, standard deviation of the MultiHMM is less than the CSE tolerance ($s < 2s_{CSE}$), which means that it can be competitive with a good physician expert.

The run-time of the proposed method for a 15 seconds record takes about 3.5 seconds for training and 21.5 seconds for test step (using a Core i3, 2.53 GHz CPU),

suggesting that this method is almost fast. It is worth to mention that our simulations are done in MATLAB, which is not a very fast language, and it could be improved by C implementation.

Acknowledgment

This work has been partly supported by the PhD scholarship of the French Embassy and the European project ERC-2012-AdG-320684-CHESS.

References

[1] A. Gacek, W. Pedrycz, ECG Signal Processing, Classification and Interpretation: A Comprehensive Framework of Computational Intelligence, Springer London, 2012, ch.11 (volutionary Optimization of ECG Feature Extraction Methods: Applications to the Monitoring of Adult Myocardial Ischemia and Neonatal Apnea Bradycardia Events, by A. I. Hernandez and J. Dumont and M. Altuve and A. Beuchee and G. Carrault).

[2] B. Kohler, C. Hennig, R. Orglmeister, The principles of software QRS detection: Reviewing and comparing algorithms for detecting this important ECG waveform, IEEE Engineering in Medicine and Biology (2002) 42–57.

[3] R. J. Martis, U. R. Acharya, H. Adeli, Current methods in electrocardiogram characterization, Computers in Biology and Medicine 48 (2014) 133–149.

[4] P. Laguna, R. Jane, R. Caminal, Automatic detection of wave boundaries in multilead ECG signals: validation with the CSE data-base, Computers and Biomedical Research 27 (1994) 45–60.

[5] D. A. Coast, R. M. Stern, G. G. Cano, S. A. Briller, An approach to cardiac arrhythmia analysis using hidden Markov models, IEEE Transaction on Biomedical Engineering. 37 (9) (1990) 826–836.

[6] N. P. Hughes, Probabilistic models for automated ECG interval analysis, Ph.D. thesis, Department of Engineering Science, University of Oxford (2006).

[7] R. Andreao, B. Dorizzi, J. Boudy, ECG signal analysis through hidden Markov models, IEEE Transaction on Biomedical Engineering 53 (8) (2006) 1541–1549.

[8] S. Krimi, K. Ouni, N. Ellouze, An approach combining wavelet transform and hidden Markov models for ECG segmentation, in: International Conference on Information and Communication Technologies: From Theory to Applications, 2008, pp. 1018–1023.

[9] R. Andreao, J. Boudy, Combining wavelet transform and hidden Markov models for ECG segmentation, EURASIP Journal on Advances in Signal Processing (2006) 1–6.

[10] J. Thomas, C. Rose, F. Charpillet, A multi-HMM approach to ECG segmentation, in: International Conference on Tools with Artificial Intelligence, 2006.

[11] W. Liang, Y. Zhang, J. Tan, Y. Li, A novel approach to ECG classification based upon two-layered HMMs in body sensor networks, Sensors 14 (2014) 5994–6011.

[12] R. V. Andreao, S. M. T. Muller, J. Boudy, B. Dorizzi, T. F. Bastos-Filho, M. Sarcinelli-Filho, Incremental HMM training applied to ECG signal analysis, Computers in Biology and Medicine 38 (2008) 659–667.

[13] H. Li, J. Tan, ECG segmentation in a body sensor network using hidden Markov models, in: International Workshop on Wearable and Implantable Body Sensor Networks, 2008, pp. 285–288.

[14] C. Lin, C. Mailhes, J. Y. Tourneret, P- and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler, IEEE Transaction on Biomedical Engineering 57 (12) (2010) 2840–2849.

[15] C. Lin, G. Kail, J. Y. Tourneret, C. Mailhes, F. Hlawatsch, P and T wave delineation and waveform estimation in ECG signals using a block Gibbs sampler, in: International Conference on Acoustics, Speech and Signal Processig, 2011, pp. 537–540.

[16] C. Li, C. Zheng, C. Tai, Detection of ECG characteristic points using wavelet transforms, IEEE Transaction on Biomedical Engineering 42 (1) (1995) 21–28.

[17] J. P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, P. Laguna, A wavelet-based ECG delineator: Evaluation on standard databases, IEEE Transaction on Biomedical Engineering 51 (4) (2004) 570–581.

[18] J. Dumont, A. I. Hernandez, G. Carrault, Improving ECG beats delineation with an evolutionary optimization process, IEEE Transaction on Biomedical Engineering 57 (2010) 607–615.

[19] M. R. Homaeinezhad, M. ErfanianMoshiri-Nejad, H. Naseri, A correlation analysis-based detection and delineation of ECG characteristic events using template waveforms extracted by ensemble averaging of clustered heart cycles, Computers in Biology and Medicine 44 (2014) 66–75.

[20] A. Karimipour, M. R. Homaeinezhad, Real-time electrocardiogram P-QRS-T detection-delineation algorithm based on quality-supported analysis of characteristic templates, Computers in Biology and Medicine 52 (2014) 153–165.

[21] S. S. Mehta, N. S. Lingayat, Combined entropy based method for detection of QRS complexes in 12-lead electrocardiogram using SVM, Computers in Biology and Medicine 38 (2008) 138–145.

[22] S. Pal, M. Mitra, Empirical mode decomposition based ECG enhancement and QRS detection, Computers in Biology and Medicine 42 (2012) 83–92.

[23] M. Akhbari, M. B. Shamsollahi, C. Jutten, Fiducial points extraction and characteristic waves detection in ECG signal using a model-based bayesian framework, in: International Conference on Acoustics, Speech, and Signal Processing, 2013, pp. 1257–1261.

[24] O. Sayadi, M. B. Shamsollahi, A model-based bayesian framework for ECG beat segmentation, Physiological Measurement 30 (2009) 335–352.

[25] M. Akhbari, M. B. Shamsollahi, C. Jutten, ECG fiducial points extraction by extended kalman filtering, in: International Conference on Telecommunications and Signal Processing, 2013, pp. 628–632.

[26] R. O. Bonow, D. L. Mann, D. P. Zipes, P. Libby, Braunwald's Heart Disease :A Textbook of Cardiovascular Medicine, 9th Edition, Elsevier, 2011, ch.13 (Electrocardiography, by D. M. Mirvis and A. L. Goldberger).

[27] S. Maheshwari, A. Acharyya, P. E. Puddu, E. B. Mazomenos, G. Leekha, K. Maharatna, M. Schiariti, An automated algorithm for online detection of fragmented QRS and identification of its various morphologies, Journal of the Royal Society 10 (2013) 1–18.

[28] E. B. Mazomenos, D. Biswas, A. Acharyya, T. Chen, K. Maharatna, J. Rosengarten, J. Morgan, N. Curzen, A low-complexity ECG feature extraction algorithm for mobile healthcare applications, IEEE Journal of Biomedical and Health Informatics 17 (2013) 459–469.

[29] V. Bono, E. B. Mazomenos, T. Chen, J. A. Rosengarten, A. Acharyya, K. Maharatna, J. M. Morgan, N. Curzen, Development of an automated updated selvester QRS scoring system using SWT-based QRS fractionation detection and classification, IEEE Journal of Biomedical and Health Informatics 18 (2014) 193–204.

[30] A. Kumar, M. Singh, Ischemia detection using isoelectric energy function, Computers in Biology and Medicine 68 (2016) 76–83.

[31] L. R. Rabiner, B. Juang, An introduction to hidden Markov models, IEEE ASSP Magazine 3 (1986) 4–16.

[32] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (2) (1989) 257–286.

[33] C. M. Travieso, J. B. Alonso, M. Pozo, J. R. Ticay, G. Castellanos-Dominguez, Building a cepstrum-HMM kernel for apnea identification, Neurocomputing 132 (2014) 159–165.

[34] M. Altuve, G. Carrault, A. Buchee, P. Pladys, A. I. Hernandez, Online apnea-bradycardia detection using hidden semi Markov models, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 4374–4377.

[35] M. Altuve, G. Carrault, A. Buchee, C. Flamand, P. Pladys, A. I. Hernandez, Comparing hidden Markov model and hidden semi-Markov model based detectors of apnea-bradycardia episodes in preterm infants, in: Computing in Cardiology, 2012, pp. 389–392.

[36] M. Altuve, G. Carrault, A. Buchee, P. Pladys, A. I. Hernandez, Online apneabradycardia detection based on hidden semi-Markov models, Medical and Biological Engineering and Computing 53 (2015) 1–13.

[37] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, J. J. Struijk, Segmentation of heart sound recordings by a duration-dependent hidden Markov model, Physiological Measurement 31 (2010) 513–529.

[38] F. Marzbanrad, Y. Kimura, K. Funamoto, R. Sugibayashi, M. Endo, T. Ito, M. Palaniswami, A. H. Khandoker, Automated estimation of fetal cardiac timing events from doppler ultrasound signal using hybrid models, IEEE Journal of Biomedical and Health Informatics 18 (4) (2014) 1169–1177.

[39] http://www.physionet.org/physiobank/database/qtdb.

[40] P. Laguna, R. G. Mark, A. Goldberg, G. B. Moody, A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG, IEEE, Computers in Cardiology 24 (1997) 673–676.

[41] O. Sayadi, D. Puppala, N. Ishaque, R. Doddamani, F. M. Merchant, C. Barrett, J. P. Singh, E. K. H. T. Mela, J. P. Martinez, P. Laguna, A. A. Armoundas, A novel method to capture the onset of dynamic electrocardiographic ischemic changes and its implications to arrhythmia susceptibility, Journal of the American Heart Association (JAHA) 3 (2014) 1–14.

[42] J. E. Cavanaugh, A large-sample model selection criterion based on kullback's symmetric divergence, Statistics and Probability Letters 44 (1999) 333–344.

[43] G. Schwartz, Estimating the dimension of a model, The Annals of Statistics 6 (2) (1978) 461–464.

[44] R. Sameni, M. B. Shamsollahi, C. Jutten, G. D. Clifford, Nonlinear Bayesian filtering framework for ECG denoising, IEEE Transaction on Biomedical Engineering 54 (12) (2007) 2172–2185.

[45] HMM Toolbox,1998, https://www.cs.ubc.ca/murphyk/Software/HMM/hmm.html.

[46] R. Sameni, The Open-Source Electrophysiological Toolbox (OSET), version 3.1, 2014. [online], Available: http://www.oset.ir.

[47] CSE, Recomendations for measurement standards in quantitative electrocardiography, European Heart Journal: The CSE Working Party 6 (1985) 815 –825.

[48] J. L. Willems, P. Arnaud, J. H. van Bemmel, P. J. Bourdillon, C. Brohet, S. D. Volta, J. D. Andersen, R. Degani, B. Denis, M. Demeester, Assessment of the performance of electrocardiographic computer programs with the use of a reference data base, Circulation, Journal of the American Heart Association 71 (3) (1985) 523–534.

[49] J. M. Bland, D. G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, International Journal of Nursing Studies 47 (2010) 931–936.

[50] H. B. Mann, D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Annals of Mathematical Statistics 18 (1) (1947) 50–60.

[51] R. Kohavi, A study of cross validation and bootstrap for accuracy estimation and model selection, in: International Joint Conference on Artificial Intelligence, 1995.