A Multi-HMM Approach to ECG Segmentation

Julien Thomas, Cédric Rose, François Charpillet

A Multi-HMM Approach to ECG Segmentation

Julien Thomas, Cédric Rose and François Charpillet
INRIA-LORIA, B.P.239
54506 Vandoeuvre-lès-Nancy, France
{jthomas,rose,charpillet}@loria.fr

Abstract

Pharmaceutic studies require to analyze thousands of ECGs in order to evaluate the side effects of a new drug. In this paper we present a new approach to automatic ECG segmentation based on hierarchic continuous density hidden Markov models. We applied a wavelet transform to the signals in order to highlight the discontinuities in the modeled ECGs. A training base of standard 12-lead ECGs segmented by cardiologists was used to evaluate the performance of our method. We used a Bayesian HMM clustering algorithm to partition the training base, and we improved the method by using a multi-model approach. We present a statistical analysis of the results where we compare different automatic methods to the segmentation of the cardiologist.

1. Introduction

The measurement of the cardiac electric activity is widely used in order to obtain information about the heart behavior. The analysis of an electrocardiogram (ECG) requires the identification of the different waves composing a cardiac cycle. The precise segmentation of ECG waves is not trivial and can be subject to discussions even among cardiologists. Pharmaceutic studies require to analyze thousands of ECGs in order to evaluate the side effects of a new drug. Ad hoc algorithms have been developed in order to help cardiologists to segment large amounts of ECGs. But these algorithms do not provide a precise segmentation, and repetitive corrections have to be made. Wavelet parametrization is known to highlight discontinuities in the signal, and has proven to give good results for ECG segmentation [9]. Hidden Markov models (HMM) are often used in signal modeling for speech recognition, but they have also been applied to ECG analysis. We propose a new method based on HMMs to learn segmenting an ECG from validated examples. This method relies on a hierarchic modeling, and Bayesian clustering of ECGs.

2. Application context

The development of a new drug is subject to various test phases (fundamental research in laboratory, animal experimentation, therapeutic test) before its validation and its marketing. The first phase of tests on human subjects, called phase I [5], is undoubtedly the most crucial. It enables to highlight the toxicity and tolerance levels, as well as undesirable side effects related to the administration of various amounts of this drug on a healthy organism.

One of the controled side effects is the cardiovascular impact of the molecule, to prevent possible heart attacks. Pharmaceutical laboratories have to carry out a series of measurements on the cardiac system using an ECG in order to prevent cardiac injuries.

The standard 12-lead ECG measures the electric potential which reflects the cardiac activity [2]: contraction (depolarization) and phase of rest (repolarization) of the heart cavities. The 12 leads, also called derivations, are named dI, dII, dIII, aVR, aVL, aVF, V1, V2, V3, V4, V5 and V6. They allow to visualize the activity of the heart under 12 different angles.

To know if drugs have an impact on the heart, not all the recorded periods are analysed at the same time. An ECG is splitted in ECG complexes, which are defined as a waveform succession representing one specific cardiac cycle period. Three intervals are more particularly studied (see figure 1).

The PR interval, measured from the beginning of the P wave up to the beginning of the QRS complex, represents the time interval between the beginning of atrial depolarization and the beginning of ventricular depolarization. The QRS complex represents the ventricular depolarization. The QT interval, measured from the beginning of the QRS complex up to the end of the T wave, represents the time interval between the beginning of ventricular depolarization and the end of ventricular repolarization.

The ECG monitoring allows to trace the evolution of the impact of a new drug on the heart. Cardiologists compare the duration of the various intervals before, during, and after

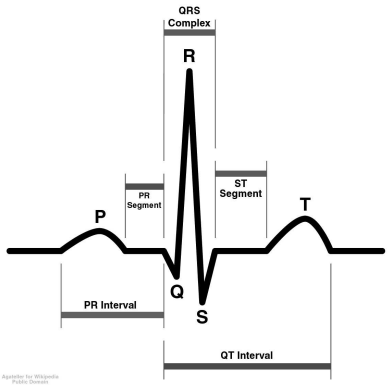

Figure 1. Schematic representation of normal ECG.

treatment.

The variation study of the QT interval duration is very important in the validation process of a new drug. This duration varies according to the heart rate. The QT interval prolongation is a surgate of torsade de pointes (TdP) [4], ventricular arrhythmia which can degenerate into ventricular fibrilation, and lead to a sudden death.

The phase I studies relate to a range of 20 to 80 subjects (depending on the drug) which receive several doses until the tolerance level is evaluated. Throughout the tests, thousands of ECGs are recorded and require to be analyzed. Currently, an ad-hoc algorithm provides a first segmentation, and each ECG has to be corrected manually by a cardiologist. A very repetitive work!

The purpose of our tool is to propose an automatic segmentation, based on the expertise of cardiologists. More precisely, we want to obtain a marker for the beginning and the end of each wave: Ponset, Poffset, QRSonset, QRSoffset, and Toffset. This tool cannot replace the cardiologist who must validate the segmentation, but it will help him in the choice of these five markers.

3. Previous work on automated ECG segmentation

An extensive review and comparison of QRS detection algorithms can be found in [10]. Many approaches consist in comparing signal features to a threshold. In spite of the good results on QRS detection, interval analysis remains a difficult task for these threshold based methods. The use of wavelets for ECG delineation has been studied in [12] and [9]. Hidden Markov tree models have been applied in [6] to the segmentation of ECG signals encoded with the discrete wavelet transform. In [8] semi-Markov models with explicit state duration modeling were proposed for ECG interval analysis.

4. Tools

4.1. Wavelet representation

The goal of our tool is to segment ECG complexes using HMMs. These models require the most relevant information in order to be efficient. Several mathematical transforms make it possible to capture the transitions from a signal more or less efficiently.

The Fourier Transform (FT) allows to analyse singularites. It however provides only a global representation of the signal. Integrating over the whole temporal domain $(-\infty,+\infty)$ erases information to locate the transitions.

This lake of temporal precision could be corrected using the Windowed Fourier Transform (WFT). It uses a temporal window to allow temporal localization. The size of the window must be fixed, which is very difficult to do without a priori information about the signal.

To highlight nonstationary properties of a signal without any information about it, we used the Continuous Wavelet Transform (CWT) [3]:

$$\gamma(\tau, s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{|s|}} \overline{\psi(\frac{t-\tau}{s})} dt$$

The CWT is performed using the convolution between the signal f and the time-localized mother wavelet ψ, which will be translated by a factor τ and scaled (or dilated) by a factor s.

The scale factor plays an important role in this wavelet transform and can be seen as the resolution of a map. Small scales, which means a high frequency version of the mother wavelet, give rich temporal information but poor frequency information. Large scales, which means a low frequency version of the mother wavelet, give rich frequency information but poor temporal information.

The choice of the number of scales is a strategic one. The first scales can be rejected to denoise the signal but they are important to temporally localize discontinuities. The large scales are not interesting for temporal localization but can detect artefacts which are not visible with the temporal representation of the signal. Several tests were carried out to define the scale band containing the most useful information.

The mother wavelet selected here is the Haar wavelet, for its facility of implementation:

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq t < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$

The representation of the Haar wavelet transform of the lead dII (see figure 2) highlights the relation between the signal discontinuities and the coefficients packages.

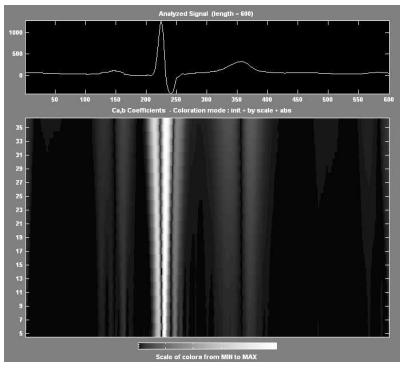

Figure 2. Haar wavelet coefficients for lead dII.

4.2. Hidden Markov Models

Bayesian approaches have been successfully applied for signal modeling. In particular, stochastic signal models work very well for building recognition systems and are commonly used in speech recognition. A HMM is a stochastic finite automaton (see figure 3) that represents a partially observable Markov process. In a first order Markov process, the state S of the system at a time t is conditionally independent of the state history knowing the immediate previous state at time $t-1$, i.e. $P(S_t|S_{t-1}, S_{t-2}, ..., S_1) = P(S_t|S_{t-1})$. A continuous density HMM of size N can be fully described by the triplet $\lambda = (A, B, \Pi)$ where:

- A is a (N, N) matrix of transition probabilities such as $A(i, j) = P(S_t = j|S_{t-1} = i)$,
- B is a set of N observation probability density functions $B = (B_1, .., B_N)$ where $B_i = P(O_t|S_t = i) = (\mu_i, \sigma_i)$,
- Π is an initial probability distribution over the HMM states: $\Pi = P(S_0)$.

The three basic problems for HMMs and their solution are presented in [13].

1. Compute the probability $P(O|\lambda)$ of an observation sequence $O = O_1, ..., O_t$ given the model λ. This can be viewed as the problem of scoring how well a given observation sequence matches a given model. The *Forward-Backward* procedures can be used to answer this question.

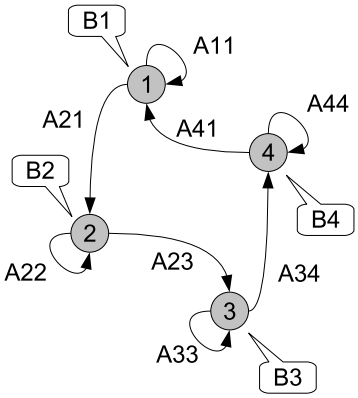

Figure 3. A HMM as a probabilistic automaton. Each state is associated with an emitted signal.

2. Given an observation sequence O and a model λ, find the most probable state sequence. The *Viterbi* algorithm is used to attempt to uncover the hidden part of the model.

3. Given a state sequence O, find the parameters of the model λ that maximize $P(O|\lambda)$. This is the training problem. The *Baum-Welch* method is an iterative procedure for optimizing the model so as to best explain a given observation sequence.

5. Using HMM for ECG segmentation

ECG segmentation is a recognition problem. We need to recognize and precisely localize the different waves of the cardiac cycle inside the signal. Our approach to ECG segmentation consists in building a model λ of the signal, and in using the most likely state sequence for a given observation sequence in order to find the wave transitions.

The parameters of the model are learned from reference ECG complexes, segmented by a cardiologist. For our experimentation we used a data collection of 1800 properly segmented ECG complexes. As represented in figure 4, the ECG complex is split into 6 sub-waves (Base1, P, Base2, QRS, T, Base3) and a specific wave model is trained for each segment.

5.1. Wave models

Each wave of the ECG complex is represented by a 5 states left-right continuous density HMM as represented on figure 5. Six different HMMs are trained : λ_{Base1},λ_P,λ_{Base2},λ_{QRS},λ_T,λ_{Base3}.

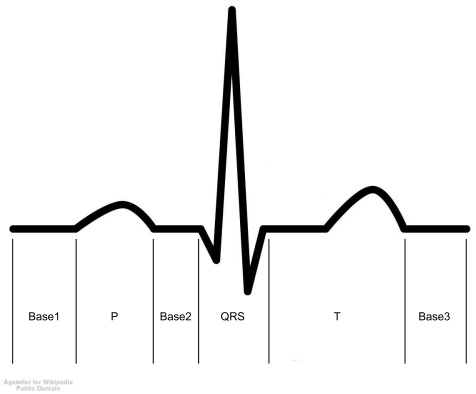

Figure 4. ECG sub waves.

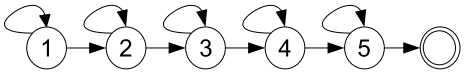

Figure 5. Single wave model.

We used multivariate Gaussian distributions as observation probability density functions,

$$B_S(\mathbf{w}) = P(\mathbf{w}|S) = N(\mathbf{w}, \mu_S, \Sigma_S)$$

where μ_S and Σ_S are the mean vector and the covariance matrix of the Gaussian distribution associated with state S,

$$N(\mathbf{w}, \mu_S, \Sigma_S) = p \times exp(-\frac{1}{2}(\mathbf{w} - \mu_S)'\Sigma_S^{-1}(\mathbf{w} - \mu_S))$$

where $p = (2\pi)^{-N/2}|\Sigma_S|^{-1/2}$ is the normalizing constant that ensures $\int_{\mathbf{w}} P(\mathbf{w}|S) = 1$. ($N$ is the dimension of $\mathbf{w}$).

The training of the HMMs is realized using Viterbi training (VT) followed by the Baum-Welch algorithm (BW). Both algorithms are based on the Expectation-Maximization principle and maximize the probability $p(O|\lambda)$. VT gives a first estimation of the HMM parameters with a reduced computational cost. The solution is then refined using BW. The difference between VT and BW is that during expectation step VT uses the Viterbi estimation of the most likely state sequence instead of the standard Forward-Backward computation as in BW.

5.2. Models Aggregation

Once we have a HMM for each wave, we need to build a global model for the whole ECG complex. The global model Λ is a concatenation of the wave HMMs. The probability transition between two successive models is taken from the outgoing probability of the first HMM.

The resulting global HMM can be regarded as a hierarchical HMM (see figure 6), where each state of the super model corresponds to a particular wave of the ECG.

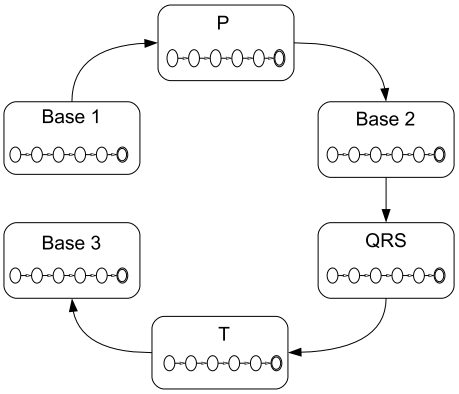

Figure 6. Hierarchical HMM integrating the wave models.

For a given wavelet coefficient sequence, we can compute the most likely corresponding state sequence applying the Viterbi algorithm on the model Λ. Marker positions are deduced from state transitions inside model Λ.

This method allows us to learn the way of segmenting an ECG complex from an expert. The resulting model can be more or less specialized according to the variety of training cases. A specialized model will result in an accurate segmentation on cases close to the training base, but will cover a limited number of cases. On a particular subject the experimentation has shown that the segmentation obtained using a single generic model was not as good as the segmentation resulting from a model trained specifically on this subject.

6. Multi-model approach (clustering)

In order to take advantage of the accuracy of specialized models, and still be able to treat a large variety of cases, we developed a multi-model approach based on the clustering of ECG complexes. A cluster is defined as a class of signals with some similarities. We used a Bayesian HMM clustering algorithm to constitute the classes of ECGs maximizing the overall likelihood of the database. This algorithm consist in finding a set of K models $(\Lambda_1, .., \Lambda_K)$ that maximizes:

$$P(O|\Lambda_1, .., \Lambda_K) = \prod_{i=1}^{N} P(O_i|\Lambda_1, .., \Lambda_K)$$

where $O = (O_1, ..O_N)$ is a set of observation sequences.

The likelihood of a particular observation sequence is given by:

$$P(O_i|\Lambda_1, .., \Lambda_K) = \max_{j \in [1..K]} P(O_i|\Lambda_j)$$

This HMM clustering algorithm (see table 1) is simply a variation of the k-mean algorithm, where the clusters are de-

fined by HMMs rather than by centers in the data space. For a given class, the notion of distance to the center is replaced by the likelihood $P(O_i|\Lambda_j)$ of the observation sequence. The number of classes K is fixed arbitrarily.

Randomly form a partition of K cluster
DO
Train a HMM for each cluster
Distribute observation sequence to clusters with the highest likelihood
WHILE Cluster configuration is not stable

Table 1. Bayesian clustering algorithm.

The HMM structure used in the clustering process is the same as the one used for ECG segmentation. As a results the cluster HMMs ($\Lambda_1, .., \Lambda_K$) can directly be used for ECG segmentation. In order to analyze a new signal O, we simply need to select the model with the highest likelihood (the model Λ_i that maximizes $P(O|\Lambda_i)$).

The computational cost of this clustering algorithm is high, but it can easily be parallelized, since the training of the K HMMs can be done simultaneously. In our experiments, the classification of 1800 ECG complexes into 10 classes took between 20 to 30 iterations.

Note that the training of the K models is done only once. The computation overhead for the segmentation of a new signal is multiplied by a factor K.

7. Segmentation of multi-channel ECGs

As seen before, an ECG is made up of 12 leads which show the same electric activity under 12 different angles. Rather than selecting only one lead (dII is most frequently selected) each 12 lead is segmented. The most probable segmentation for an ECG complex is selected by computing the median value among the 12 proposals. Aberrant values, due to bad recordings for example, can be filtered out by using this method.

The training base is decomposed into 12 sub bases, in order to train a model for each category of leads. As a result we have 12 models for the "generic" approach, and $K \times 12$ models for the "cluster" approach.

For the "generic" approach the use of these 12 models provides 12 different segmentations for an ECG complex. The choice of the marker nearest to the expected segmentation is carried out by determining the median value of the 12 proposals.

For the "cluster" approach first we use the maximum likelihood estimate to find the class to which a lead belongs. When the 12 leads are classified and segmented, we use the median value choice.

8. Experimental method and results

8.1. Methodology

Several cardiologists manually segmented a batch of approximately 1800 ECG complexes to constitute a training base. Two approaches are considered: a method for which the training relates to the whole of these files, called the "generic" approach, and a clustering method with $K = 10$ classes (this number was fixed arbitrarily) refered to the "cluster" approach.

To evaluate their efficiency, these 2 semgentations are compared to the automatic segmentation of the ad hoc algorithm, and to the manual segmentation of the cardiologist considered as the gold standard. The statistical analysis is carried out by a paired-sample Student's t-test ($\alpha = 0.05$) [1]. These 4 segmentation methods are applied on 173 ECG complexes.

Statistical analysis are performed with SAS software (version 9.1; SAS Institute Inc., Cary, NC).

8.2. Comparative results

For each interval (PR interval, QRS complex and QT interval), 2 tables show on the one hand a descriptive statistical analysis of the 4 segmentation methods (by the ad hoc algorithm, manually by a cardiologist, with "generic" and "cluster" approaches), and on the other hand a comparison of the different methods with the paired-samples averages. All the durations are expressed in milliseconds.

Tables 2, 4 and 6 can also be represented by a box plot graph to illustrate the dispersion phenomenon. The bottom of the rectangle is the first quartile, the top is the third quartile and the line in the middle is the median. The lower line is the minimum value and the higher the maximum.

For the PR interval (see tables 2 and 3), the mean and median values for the cardiologist and the ad hoc algorithm are very close. But even if the mean of the differences between the two segmentation methods is less than 2 milliseconds, the standard deviation is approximately two times higher for the segmentation of the ad hoc algorithm compared to the cardiologist. The mean value and the mean of the differences with the cardiologist for the "generic" approach is higher compared to the two other automatic methods. The mean and median values, and the dispersion of the "cluster" approach are very close to ones of the cardiologist.

For the QRS complex (see tables 4 and 5), the three automatic methods have mean and median values lower than the values of the cardiologist.

PR	Mean	Min.	Max.	Median	SD[1]
Ad hoc alg.	160.6	110	408	158	33.3
Cardiologist	159.2	114	204	158	19.0
Generic	170.1	119	370	166	31.7
Cluster	161.9	120	210	161	18.6

Table 2. PR interval statistical analysis.

Δ PR	Mean	SD	Min.	Max.	p value
Cardio-Ad hoc	-1.3	28.0	-254	36	0.5296
Cardio-Generic	-10.9	27.3	-233	18	0.0001
Cardio-Cluster	-2.7	5.8	-26	21	0.0001
Cluster-Ad hoc	1.3	28.3	-258	32	0.5322
Generic-Ad hoc	9.5	39.3	-255	233	0.0016
Cluster-Generic	8.2	25.8	-7	219	0.0001

Table 3. Comparison of different methods for PR interval.

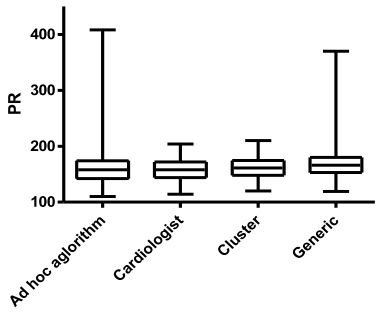

Figure 7. PR interval dispersion.

QRS	Mean	Min.	Max.	Median	SD[1]
Ad hoc alg.	84.2	64	122	84	9.5
Cardiologist	95.5	76	116	96	8.3
Generic	89.3	76	107	89	6.1
Cluster	88.3	69	109	87	6.9

Table 4. QRS complex statistical analysis.

Δ QRS	Mean	SD	Min.	Max.	p value
Cardio-Ad hoc	11.3	7.2	-16	32	0.0001
Cardio-Generic	6.1	6.4	-15	26	0.0001
Cardio-Cluster	7.1	6.9	-12	30	0.0001
Cluster-Ad hoc	4.1	8.0	-25	24	0.0001
Generic-Ad hoc	5.2	7.2	-21	25	0.0001
Cluster-Generic	1.0	4.2	-20	10	0.0018

Table 5. Comparison of different methods for QRS Complex.

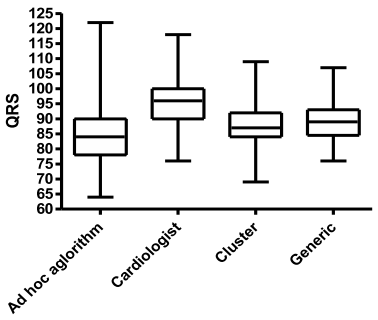

Figure 8. QRS complex dispersion.

For the QT interval (see tables 6 and 7), the results obtained by the 4 segmentation methods are similar but the "cluster" approach has the least difference with the cardiologist according to the mean, median and dispersion values.

QT	Mean	Min.	Max.	Median	SD[1]
Ad hoc alg.	396.5	306	454	398	25.8
Cardiologist	400.5	330	474	402	25.6
Generic	403.1	339	483	404	26.3
Cluster	399.0	330	481	400	25.3

Table 6. QT interval statistical analysis.

Δ QT	Mean	SD	Min.	Max.	p value
Cardio-Ad hoc	4.0	7.0	-18	24	0.0001
Cardio-Generic	-2.5	16.3	-142	36	0.0442
Cardio-Cluster	1.5	8.8	-30	27	0.0217
Cluster-Ad hoc	2.5	9.6	-29	34	0.0009
Generic-Ad hoc	6.5	17.9	-54	166	0.0001
Cluster-Generic	4.1	14.2	-25	142	0.0002

Table 7. Comparison of different methods for QT Interval.

9. Discussion

The goal of our tool is not to replace the cardiologist intervention for the ECGs analysis. We try to give him more useful information than the ad hoc algorithm currently used.

Compared with the ad hoc algorithm, the "cluster" approach segmentation is closer to the cardiologist results. Especially for PR and QT intervals, the mean and median are very close to the cardiologist values. Despite higher computation time than for the "generic" approach, the multi-model one gives a more accurate segmentation. With the "generic" approach the dispersion for this two intervals is higher.

[1] Standard deviation

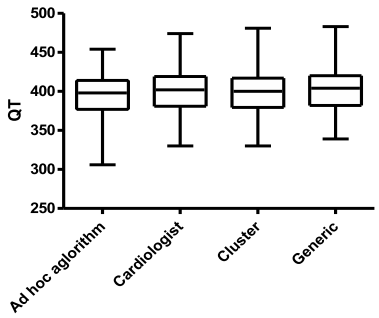

Figure 9. QT interval dispersion.

For the QRS complex however, both the "generic" and "cluster" approaches are not efficient enough: the QRS complex durations obtained by these rwo methods are lower than those obtained by the cardiologist. In fact the QRSonset marker is at the same time the end of the PR segment and the beginning of the QRS complex. The PR segment of a ECG lead (see figure 1) is sometimes very short. The states number for the PR segment model (λ_{Base2}) could be decreased so the QRSonset marker could be detect earlier. We expect that the QRS complex durations will be closer to the cardiologist ones.

To optimize the segmentation of the 3 intervals and more particularly for the QRS complex, several improvments are envisaged. We will enrich the training by increasing the number of ECG complexes provided to the training base. We plan to determine the optimal states number for each waves by using a criterion such as Bayesian information criterion and implement a method to perform the clustering method initialisation. The adatpation for a specific subject could also be considered in the case of a large pool of ECGs per patient.

Even though our algorithm has some problems segmenting the QRS complex, we have much better results with the "cluster" approach than the ad hoc algorithm for the three intervals.

10. Conclusion

We have presented a machine learning approach to ECG segmentation. We trained Continuous Density HMMs on ECGs segmented by cardiologists. The Haar continuous wavelet transform was used to encode the ECGs. A "cluster" approach based on the k-mean clustering algorithm was implemented in order to generate ECGs classes among the training base. Because of the specialization of each class, we have obtained segmentations nearer to the cardiologist than the "generic" approach.

A set of 1800 ECG complexes segmented by a cardiologist was used for the training of the models. The accuracy of the generic and the cluster methods was measured on 173 ECG complexes using the cardiologist segmentation as a reference. The results were also compared to a commercial ad hoc algorithm. The "cluster" approach gave significantly better results than the "generic" method but presents the drawback of having a higher computation cost.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1964.

[2] B. Aehlert. *ECG's Made Easy 2nd Edition*. C.V. Mosby, St. Louis, 2001.

[3] I. Daubechies. *Ten lectures on Wavelet*. SIAM, Philadelphia, 1992.

[4] F. Dessertenne. Ventricular tachycardia with 2 variable opposing foci. *Arch Mal Coeur Vaiss*, 59(2):263–72, February 1966.

[5] Food And Drug Administration. *General Considerations For The Clinical Evaluation Of Drugs*. U.S. Department Of Health, Education And Welfare, 1997.

[6] S. Graja and J.-M. Boucher. Markov models for automated ECG interval analysis. In *WISP 2003: IEEE International Symposium on Inteligent Signal Processing*, pages 105–109, 2003.

[7] D. Heckerman, D. Geiger, and D. M. Chikering. A tutorial on learning with bayesian networks. *Machine Learning*, 20:197–243, 1995.

[8] N. P. Hughes, L. Tarassenko, and S. J. Roberts. Markov models for automated ECG interval analysis. In S. Thrun, L. Saul, and B. Schoelkopf, editors, *Advances in Neural Information Processing Systems 16*, Cambridge, MA, 2004. MIT Press.

[9] A. Kawaja, S. Sanyal, and O. Dössel. A wavelet-based multi-channel ECG delineator. In *The 3rd European Medical and Biological Engineering Conference*, 2005.

[10] B.-U. Köhler, C. Hennig, and R. Orglmeister. The principles of software QRS detection. *Engineering in Medicine and Biology Magazine, IEEE*, 2002.

[11] C. Li and G. Biswas. Bayesian clustering for temporal data using hidden markov model representation. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pages 543–550, 2000.

[12] J. P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna. A wavelet-based ECG delineator: Evaluation on standard databases. In *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING BME*, 2004.

[13] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989.