Probabilistic Models for Automated ECG Interval Analysis in Phase 1 Studies

Technical Report BSP 08-01

Nicholas P. Hughes and Lionel Tarassenko

Institute of Biomedical Engineering
University of Oxford
ORCRB, Off Roosevelt Drive
Oxford OX3 7DQ, UK

Email: lionel@robots.ox.ac.uk
Tel: +44 (0)1865 617674
Fax: +44(0)1865 617701
Abstract

The accurate measurement and assessment of the various ECG intervals, and in particular the QT interval, is currently the gold standard for evaluating the cardiac safety of new drugs. Automated methods for ECG interval analysis offer the potential for a more extensive evaluation of ECG data from clinical trials, and hence a more robust assessment of drug safety. In this paper we consider the use of hidden Markov models (HMMs) for this task, with a focus on the analysis of ECGs from phase 1 studies. We introduce a novel sample-wise encoding of the ECG based on the undecimated wavelet transform and derivative features, and show how the robustness of the HMM segmentations can be greatly improved through the use of state duration constraints. Most significantly, we show how the probabilistic nature of the HMM can be leveraged to provide a confidence measure in the model segmentations. This enables our algorithm to automatically highlight any ECG interval measurements which are potentially unreliable. The performance of our approach is evaluated on a substantial ECG data set from a clinical study, where we demonstrate that our algorithm is able to produce QT interval measurements to a similar level of accuracy as human experts.
INTRODUCTION

Drug safety is an area of increasing concern for both pharmaceutical companies and regulatory agencies. Of particular importance is the propensity for some non-cardiac drugs (such as antihistamines) to induce potentially fatal heart rhythms. The primary means by which this aspect of cardiac safety is assessed in clinical drug trials is through the analysis of the 12-lead electrocardiogram (ECG), and in particular, the various timing intervals within each individual heartbeat.

Fig. 1 shows a typical ECG waveform and the standard ECG intervals occurring within a given beat: namely, the PR interval, QRS duration (QRSd) and QT interval. Of particular significance is the QT interval, which is defined as the time from the start of the QRS complex to the end of the T wave (i.e. Toff − Q). This value corresponds to the total duration of electrical activity (both depolarisation and repolarisation) within the ventricles in a given heartbeat. The importance of the QT interval stems from its link to the cardiac arrhythmia torsade de pointes [1]. This arrhythmia is known to occur when the QT interval is prolonged significantly (so called long QT syndrome), which in turn can lead to ventricular fibrillation and sudden cardiac death.

Recently, the propensity for some non-cardiac drugs to prolong the QT interval, and hence cause torsade de pointes, has begun to emerge [2]. This issue was first highlighted in the case of the antihistamine terfenadine, which had the side-effect of significantly prolonging the QT interval in a number of patients. Unfortunately this effect was not adequately characterised in the clinical trials and only came to light after a large number of people had unexpectedly died whilst taking the drug.

In order to address the issue of drug-induced QT prolongation, the U. S. Food and Drug Administration (FDA) and Health Canada recently introduced a new type of phase 1 clinical study known as a “thorough phase 1 QT/QTc study” [3]. These studies are now a mandatory requirement for all new drug submissions and must be performed before approval can be granted by a regulatory body. A key requirement of the thorough QT study is the collection of a large number of ECGs (typically 20,000 or more), and the need for highly accurate and reliable QT interval measurements.

At the present time, ECG interval measurements for clinical trials are carried out manually by expert analysts. This procedure is both costly and labour intensive. Furthermore, manual ECG interval measurements are susceptible to occasional mistakes by the analysts. These drawbacks typically limit the total number of electrocardiograms which can be successfully analysed in a given trial. Automated methods for ECG interval analysis therefore offer
the potential for a more extensive evaluation of ECG data from clinical trials, and hence a more robust assessment of drug safety.

Previous Approaches to Automated ECG Analysis

Despite the advantages offered by automated approaches to ECG interval analysis, current automated systems are not sufficiently reliable to be used for the assessment of ECG data from clinical trials. The algorithms which underlie these systems typically fall into one of two categories: rule-based approaches and model-based approaches.

Most rule-based algorithms divide the ECG segmentation problem into a number of distinct stages. In the first stage, the locations of the R peaks in the 10-second ECG signal are estimated using a QRS detector, such as the Pan and Tompkins algorithm [4]. Once the R peaks have been identified, the next step is to search forwards and backwards from each peak to locate the QRS complex boundaries (i.e. the Q and J points). These points are typically determined as the first points (searching forwards or backwards from the peak) where the signal (or its derivative) crosses a pre-determined threshold [5].

Following the identification of the QRS boundaries, the next stage is to search for the peaks of the P and T waves. These can be estimated as the points with the maximum absolute amplitude (relative to the Q or J points) occurring within a given window to the left and the right of the R peak [5]. An alternative approach is to use template matching techniques to locate the peaks [6]. The advantage of this approach is that it offers an increased degree of robustness to noise in the ECG signal, compared with thresholding methods. The drawback however is that it requires the definition of one or more templates for the given waveform feature.

Following the determination of the P and T wave peaks, the final stage is to locate the P and T wave boundaries. Threshold methods represent the most common approach for this problem. The standard thresholding algorithm for ECG segmentation, known as “ECGPUWAVE”, applies a series of adaptive thresholds to the ECG derivative in order to determine the waveform feature boundaries [7]. In the case of the T wave offset, the associated threshold is first estimated from the peak of the ECG signal derivative within a given window (following the T wave peak). The derivative value at this point is then scaled according to a pre-determined scaling factor \( k \). The resulting value then defines the threshold for detecting the T wave offset in the ECG derivative.

An alternative rule-based approach to thresholding is given the tangent method. This approach was first introduced more than fifty years ago as a manual technique for T wave offset determination [8]. More recently it has been used as the basis for a number of automated techniques [6]. The tangent method determines the end of the T
wave as the point of intersection between the (estimated) isoelectric baseline and the tangent to the downslope of the T wave (or the upslope for inverted T waves). The tangent itself is computed at the point of maximum (absolute) gradient following the peak of the T wave. The isoelectric baseline is typically estimated from the region of baseline between the P wave offset and the QRS complex onset.

A significant disadvantage of the tangent method is that it is sensitive to the amplitude of the T wave. In particular, large amplitude T waves can cause the tangent to intersect the isoelectric baseline before the true end of the T wave. As a result, the automated QT intervals from the tangent method can significantly underestimate the “true” QT interval values (as determined by an expert analyst). To overcome this problem, Xu and Reddy apply a “nonlinear correction” factor (based on the T wave amplitude) to the T wave offset determined by the tangent method [6].

The application of wavelet transforms to the analysis of the ECG has been considered by a number of authors. In the context of ECG segmentation, wavelets are generally used in combination with standard thresholding methods [9, 10] (i.e. the thresholding is carried out in the wavelet domain). The advantage of wavelet thresholding approaches over similar methods which operate on the ECG signal itself (or its derivative) is that the resulting segmentations tend to be less sensitive to the effects of noise and artefact in the signal.

An alternative to the rule-based approaches to ECG signal analysis is to construct a statistical model of the ECG. This approach was first investigated by Coast et al., who demonstrated that hidden Markov models (HMMs) can be applied successfully to the problem of cardiac arrhythmia analysis [11]. To address this problem, the authors trained hidden Markov models (with Gaussian observation densities) to recognise ECG beats from three different types of heart rhythms. For each model, a bandpass filtered version of a single channel ECG signal was used as the input observation sequence to the HMM. Separate HMMs were trained on ECG beats produced by a normal heart rhythm, a ventricular heart rhythm, and a supraventricular heart rhythm. The three HMMs were then connected together to form a single “global” model for continuous ECG analysis.

For HMM training, Coast et al. used a patient-specific training strategy to estimate the model parameters for each of the three HMMs. In particular, for a given patient, the individual HMMs were trained using ECG data recorded from that patient only. In the first step of the training procedure, three ECG beats associated with each of the three different heart rhythms were identified manually in the ECG recording for the given patient. These beats were then segmented manually and the annotated ECG beats for each type of heart rhythm were used to initialise the parameters of the corresponding model. The model parameters for each HMM were then refined by running the EM algorithm on the three ECG beats associated with that particular model.
More recently, Andreão et al. have applied HMMs to the problem of ECG beat segmentation and classification for online ambulatory ECG analysis [12]. The authors used multiple hidden Markov models, with different HMMs employed to model different ECG waveform morphologies. Each individual HMM consisted of a left-right Markov chain with 3 states used for each of the isoelectric baseline, P wave and QRS complex regions of the ECG, 2 states for the PQ and ST segments, and 6 states for the T wave. The number of states was chosen in proportion to the average duration of each ECG feature. The observation models for each of the hidden states were chosen to be Gaussian densities.

For each model, Andreão et al. used a continuous wavelet transform representation of the ECG signal as the input observation sequence to the HMM. Specifically, the authors employed a Mexican Hat wavelet function, and used only the wavelet coefficients evaluated at the dyadic scales $2^2$, $2^3$, and $2^4$. The HMMs were then trained in an unsupervised manner with the EM algorithm, using ECG data taken from the MIT-BIH QT Database.

A limitation of the approach of Andreão et al. is that the wavelet representation used does not encode the full spectral content of each ECG. By selecting the coefficients from only a small number of scales of the continuous wavelet transform, the resulting encoding only captures the ECG signal content at the particular frequency bands corresponding to the given scales. As a result, waveform feature boundaries which consist of sharp transitions or “discontinuities” in the ECG signal (such as the Q and J points), and which therefore give rise to spectral content over a wide range of different frequency bands, may not be detected as accurately as possible (compared with an HMM trained on a complete representation of the data).

In the context of ECG interval analysis, a more serious limitation of the approaches of both Coast et al. and Andreão et al. is the use of unsupervised learning to estimate the hidden Markov model parameters. Although this is a common approach in many applications of HMMs (particularly in the field of automatic speech recognition - where it is not feasible to collect large volumes of labelled speech data for supervised learning), it is not guaranteed that the resulting model will produce ECG segmentations consistent with those of expert analysts. In particular, unsupervised learning with the EM algorithm merely seeks to find the model parameters which maximise the likelihood of the data. These parameter estimates may be very different from those which provide optimal segmentation performance (with respect to expert measurements). Thus, a more appropriate strategy is to estimate the HMM parameters in a supervised manner using ECG data annotated by expert cardiologists. This is the approach taken in this paper.
Shortcomings of Current Automated Systems

As discussed previously, current automated systems are not sufficiently robust to be used for the analysis of ECG data from clinical drug studies. This lack of robustness can be traced back to the approach to automated ECG interval analysis which underlies all current automated systems. Specifically, traditional ECG interval analysis algorithms are designed to compute ECG interval measurements regardless of the suitability for measurement of the particular signal under consideration.

This represents a considerable drawback since not all the ECG signals recorded during the course of a clinical study are suitable for accurate and reliable interval measurements (either by machine or by human expert). As noted by Fenichel et al., “Some ECGs are technically defective (with muscle artifact, misplaced leads, or other problems), and one should not attempt to derive useful data from them. [...] Even in carefully conducted studies, about 1 in every 30+ ECGs does not lead to meaningfully measurable data.” [1].

The issue of the “suitability for measurement” of ECG signals is complicated since there exists a continuous spectrum between extreme cases (such as ECG “signals” recorded when the electrodes have been disconnected from the machine) and those where the signal is affected by only a small amount of noise. As such, “it is very difficult (if not impossible) to suggest any criteria for quick visual inspection of the ECGs in order to distinguish cases in which the automatic assessment should be excluded” [13].

Since a standard automated system will provide ECG interval measurements for any input ECG signal (regardless of the ECG signal quality or waveform normality), it is not possible therefore to differentiate between those automated measurements which are reliable and those which are not. This is the fundamental problem which hinders current approaches to automated ECG interval analysis and prevents automated systems from being utilised in the analysis of ECG data from clinical drug studies.

Overview and Contributions of this Paper

In this paper we consider the use of probabilistic models for automated ECG interval analysis, with a focus on the analysis of 10-second 12-lead ECGs from thorough phase 1 QT/QTc studies. In particular, we show how hidden Markov models can be successfully applied to this problem through a careful consideration of the choice of ECG representation and model architecture. Most significantly, we show how the probabilistic generative nature of the HMM can be leveraged to provide a confidence measure in the model segmentations.

Motivated by the observation independence assumption inherent in the HMM framework, we introduce a novel
sample-wise ECG encoding that is well suited to analysis by an HMM. This encoding, which is based on a combination of the undecimated wavelet transform and its associated derivative features, provides a representation of the ECG which facilitates accurate segmentations by a hidden Markov model. In addition, we analyse the problem of “double-beat segmentations” by an HMM, and show how the use of minimum state duration constraints serves to improve significantly the robustness of the model segmentations.

Although the use of hidden Markov models for ECG analysis has been explored previously in the literature [11, 12], the full probabilistic description these models provide has not previously been fully exploited. To this end, we show an HMM can be used to provide a statistical confidence measure in its analysis of a given ECG waveform. The level of confidence in the model segmentations can be assessed by evaluating the HMM joint log likelihood for each segmented ECG beat, with respect to the beat duration. The resulting confidence measures can then be used to differentiate between normal ECG waveforms (for which automated measurements may be reliably inferred), and abnormal or unusual ECG waveforms (for which automated measurements are frequently unreliable). Thus, by utilising a confidence-based approach to automated ECG interval analysis, we can automatically highlight those waveforms which are least suitable to analysis by machine (and thus most in need of analysis by a human expert).

We evaluate our approach on a substantial ECG data set from a thorough phase 1 QT/QTc study, and compare the performance with that of a standard threshold-based algorithm for ECG analysis. The results demonstrate the ability of our algorithm to overcome many of the problems associated with existing automated systems. Furthermore, we show that our algorithm is able to produce QT interval measurements to a similar level of accuracy as those of expert ECG analysts (taking into account the inter-analyst variability inherent in manual QT measurements).

**THOROUGH PHASE 1 QT/QTc STUDIES**

*Overview*

The phase 1 clinical study represents the first time that a drug being tested is administered to humans. The primary aim of a phase 1 study is to assess the safety of the drug. In particular, the study is designed to assess what happens to the drug in the human body, i.e. how it is absorbed, metabolised, and excreted. In addition, the phase 1 study aims to uncover any possible side-effects which may occur as the dosage levels of the drug are increased.

An important aspect of phase 1 studies is that they are usually carried out in a relatively small number of healthy normal volunteers, rather than the target population for the drug (who are the focus of subsequent phase 2 and phase
Healthy normal volunteers are typically males between the ages of 18 and 45 years old, or females who are not of child-bearing potential.

The thorough or definitive phase 1 QT/QTc study is a specific type of phase 1 study designed to provide a robust assessment of the effect of a drug on the QT and QTc intervals [3]. These studies are characterised by the following requirements:

- A large number of ECGs recorded at multiple time points, both at baseline and on drug
- A wide range of different doses of the drug under study
- The use of a positive control to assess the sensitivity of the study

In common with standard phase 1 studies, thorough phase 1 QT/QTc studies are usually carried out in healthy normal volunteers. As a result, the electrocardiogram signals recorded over the course of a thorough QT study are generally normal and exhibit standard waveform morphologies. In some instances however, the drug under consideration may affect the morphology of the ECG in an unusual manner (particularly at high dose levels).

An important aspect of the thorough QT study is that it is designed to detect small changes in the QT/QTc interval, rather than instances of torsade de pointes (or other cardiac arrhythmias) which may very occasionally result from a prolongation of the QT interval. The reason for this focus is that torsade de pointes is extremely rare, and hence little reassurance is provided by the lack of observation of this arrhythmia during the clinical trials [1]. However, because torsade de pointes is preceded by QT interval prolongation, measurement of the QT interval (and any drug-induced changes thereof) can be used as a “surrogate marker” to characterise the pro-arrhythmic potential of the drug under consideration.

Since the thorough QT study requires that multiple ECGs are recorded at multiple time points, both at baseline and on drug (or placebo), the total number of ECGs is typically much greater than that of a standard phase 1 study. In particular, a thorough QT study typically produces between 10,000 and 20,000 standard 10-second 12-lead ECGs, although some studies can give rise to more than 40,000 ECGs [14].

At the present time, the ECG signals resulting from a thorough phase 1 QT/QTc study must be analysed manually by human experts. The focus of this analysis is on the accurate measurement and assessment of the various ECG intervals. The precision with which the ECG intervals are measured is of particular importance because of the need to detect small changes in the QT interval. This requirement, together with the large quantity of ECG data generated by such studies, is driving an increasing demand for robust and reliable automated ECG interval analysis technologies.
Data Set

The data set used in this paper is taken from the placebo arm of a thorough phase 1 QT/QTc study. The data set consists of approximately 20,000 10-second 12-lead ECGs which were recorded (at a sampling rate of 500 Hz) from 380 healthy normal volunteers.

All of the ECGs collected in the study were analysed manually by a group of expert analysts. Specifically, each expert analyst was assigned to a unique subset of the subjects in the study and analysed only the ECG waveforms from these subjects.

For each 10-second 12-lead ECG, the longest lead was chosen for ECG interval measurements. With this approach, the expert analyst first determined (by visual inspection) the lead which contained the longest QT interval values, and this lead was then selected for ECG beat annotation. Three consecutive beats were annotated by the expert (in a small number of cases only two beats could be annotated), who identified the following points within each beat:

- P wave onset ($P_{on}$)
- QRS onset (Q)
- R peak (R)
- QRS offset (J)
- T wave offset ($T_{off}$)

The leads most frequently annotated in the study were limb lead II and chest lead V2. These two leads are also commonly employed in ECG interval analysis when measurements must be performed on a specific lead (which is selected in advance) across an entire ECG data set. Given the importance of leads II and V2 for ECG interval analysis, we focus therefore in the remainder of this paper on the development of probabilistic models for ECG signals recorded from these two leads. However, all of the techniques presented in this paper can be applied to any given lead of the ECG. Table 1 shows the number of unique ECG signals and annotated beats for leads II and V2 of the thorough QT study data set.

[TABLE 1 about here.]

Note that, whilst a number of previous papers on automated ECG analysis have made use of publicly available ECG databases such as the MIT-BIH QT Database [15] and the Common Standards for Electrocardiography (CSE)
Database [16], the ECG signals contained in these databases are not representative of those which are typically encountered in phase 1 clinical studies. This is because these ECGs were recorded from patient populations (as opposed to healthy normals), and therefore include a variety of beat morphologies, as well as other cardiac abnormalities. Hence, these databases are most appropriate for assessing the robustness of ECG analysis algorithms designed to work with a wide range of both normal and abnormal ECG waveforms. By contrast, the focus of this paper is on the development of techniques for the accurate measurement and assessment of the QT interval in large volumes of ECG data from healthy normals. Hence, we choose to evaluate our approach on the thorough QT data set described above.

PROBABILISTIC MODELLING OF THE ECG

Before we can apply the probabilistic modelling approach to the problem of automated ECG interval analysis, it is necessary to decide upon a suitable choice for the model itself. Specifically, we require a model that can be used to segment the ECG signal and which can also take advantage of the sequence ordering which exists between the waveform features of normal ECG heartbeats. These requirements motivate the use of a hidden Markov model for this problem.

In this section we provide an overview of hidden Markov models and the associated algorithms for learning and inference. We then compare and contrast a number of different model architectures for ECG interval analysis, and discuss the key assumptions inherent in the HMM framework and the validity of these assumptions in the context of ECG signal modelling.

Overview of Hidden Markov Models

A hidden Markov model, or HMM, is a probabilistic model which describes the statistical relationship between an observable sequence $O$ and an unobservable or “hidden” state sequence $S$. The hidden state itself is discrete and governed by an underlying Markov chain. The observation values however may be either continuous or discrete in nature [17].

An HMM (with $K$ hidden states) is defined by the following three parameters: an initial state distribution $\pi$, a state transition matrix $A$ (with elements $a_{ij}$), and a set of observation probability distributions $b_k$ (for each state $k$). The first two parameters govern the Markov model which describes the statistical properties of the hidden states. The observation or “emission” probability distributions provide the link between the statistical properties of the
observable values and the associated hidden states.

It is often useful to consider a hidden Markov model from a generative perspective. That is, we can consider the HMM as providing a “bottom up” description of how the observed sequence $O$ is produced or generated. Viewed as a generative model, the operation of an HMM is as follows:

1. Select the initial state $k$ by sampling from the initial state distribution $\pi$
2. Generate an observation value from this state by sampling from the associated observation distribution $b_k$
3. Select the state at the next time step according to the transition matrix $A$
4. Return to step 2

Hidden Markov models can be seen as generalisations of various statistical models [18]. One such view is to consider an HMM as a form of temporal mixture model. With a standard (static) mixture model, each data point is considered to have been “generated” by one of the $K$ mixture components independently of the other data points [19]. If we now relax the strong assumption of statistical independence between data points and allow the individual mixture components to possess Markovian dynamics, the result is a hidden Markov model (where the hidden state at each time step corresponds to the particular mixture component which is active at that time step).

From a purely statistical perspective, an HMM (with parameter set $\lambda$) defines a joint probability distribution over observation sequences $O$ and hidden state sequences $S$, i.e. $p(O, S \mid \lambda)$. Given this joint distribution, an important quantity of interest is the conditional distribution $P(S \mid O, \lambda)$. In particular, it is often of interest to find the particular state sequence $S^*$ which maximises this conditional distribution. This corresponds to the state sequence which is most likely to have “generated” the given observation sequence, and therefore enables us to associate particular observation “segments” with particular model states (i.e. signal segmentation). However, before the model can be used for this purpose, we require a method to learn the “optimal” model parameters from a given data set (such that the HMM provides a useful statistical model of the data). We now briefly consider the solution to each of these problems in turn.

**Learning in HMMs**

The learning problem in hidden Markov models is concerned with the estimation of the “optimal” model parameters given a particular training data set. In this paper we focus on the case of supervised learning, in which the training
data set is labelled (or “annotated”), i.e. it consists of both the observation sequences and the corresponding hidden state sequences (which generated the observed data). In the case of unsupervised learning (where the data set consists of observation sequences only), the EM algorithm can be used to estimate the HMM parameters [19, 17].

In supervised learning, we can estimate the initial state distribution by evaluating the proportion of the hidden state sequences that commence in each of the given model states at the first time step of each sequence. Denoting the total number of hidden state sequences which commence in state \( i \) at the first time step by \( n_{\text{init}}(i) \), we then have the following estimator for the \( i \)th element of the initial state distribution:

\[
\hat{\pi}_i = \frac{n_{\text{init}}(i)}{\sum_{k=1}^{K} n_{\text{init}}(k)}
\]  

(1)

For some applications, this estimator may not be appropriate. For example, the training data set used in this paper consists of annotated ECG waveforms which commence with the onset of the P wave. Hence, using the estimator given in Eq. (1) will result in the P wave state being assigned an initial state probability of one, and the remaining states a initial state probability of zero. Since a 10-second ECG recording can commence during any phase of the cardiac cycle, this estimator for \( \pi \) is clearly inappropriate. In order to overcome this problem, the following modified estimator can be used:

\[
\hat{\pi}_i = \frac{n(i)}{\sum_{k=1}^{K} n(k)}
\]  

(2)

where \( n(i) \) is the total number of times that state \( i \) occurs in the set of hidden state sequences. Thus, Eq. (2) estimates the initial state probability for each model state \( j \) as the fraction of times the model occupies state \( i \) compared to the total number of state occupancies.

To estimate the elements of the transition matrix, we evaluate the proportion of each possible state transition over all the hidden state sequences. Denoting the total number of transitions from state \( i \) to state \( j \) over all the hidden state sequences by \( n_{\text{trans}}(i, j) \), we have the following estimator for the \((i, j)\)th element of the transition matrix:

\[
\hat{a}_{ij} = \frac{n_{\text{trans}}(i, j)}{\sum_{k=1}^{K} n_{\text{trans}}(i, k)}
\]  

(3)

The exact estimator for the parameters of the observation models depends on the specific functional form chosen for these models. Typical choices for the observation models are Gaussian or Gaussian mixture densities. However, the general estimation procedure for the observation models is straightforward. For each hidden state \( i \), we first “extract” all the observation sequences associated with this particular state. The parameters of the observation
model for state $i$ are then estimated from this particular subset of the observation data.

### Inference in HMMs

The inference problem in hidden Markov models can be viewed as one of determining the “optimal” state sequence for a given observation sequence. In particular, we are interested in computing the single most probable state sequence given the observed data, i.e.:

$$S^* = \arg\max_S \{ P(S \mid O, \lambda) \}$$

$$= \arg\max_S \{ p(S, O \mid \lambda) \}$$  \hspace{1cm} (4)

which follows from Bayes’ theorem. Hence it suffices to find the state sequence $S^*$ which maximises the joint distribution $p(S, O \mid \lambda)$. The solution to this problem is provided by the Viterbi algorithm [20, 17].

In the “forward pass” of the Viterbi algorithm, we compute the quantity $\delta_t(i)$, given by:

$$\delta_t(i) = \max_{s_1 s_2 \cdots s_{t-1}} \left\{ p(s_1 s_2 \cdots s_t = i, O_1 O_2 \cdots O_t \mid \lambda) \right\}$$  \hspace{1cm} (5)

This is the likelihood of the most probable state sequence that accounts for the first $t$ observations and ends in state $i$ at time $t$. This quantity can be computed in an efficient manner using the following recursion:

$$\delta_{t+1}(i) = \max_j \{ \delta_t(j) a_{ji} \} b_i(O_{t+1})$$  \hspace{1cm} (6)

It is also useful to record the particular “maximising” state which maximises the value of $\delta_{t+1}(i)$, i.e.:

$$\psi_{t+1}(i) = \arg\max_j \{ \delta_t(j) a_{ji} \}$$  \hspace{1cm} (7)

Having computed Eqs. (6) and (7) for all times $t$ and states $i$, we can then compute the optimal value for the hidden state at the final time step as $s_T^* = \arg\max_i \{ \delta_T(i) \}$. Using this value, we can work back to uncover the optimal state value at the previous time step $t = T - 1$. This is easily found as the previous “maximising state” for $\delta_T(s_T^*)$, i.e. $\psi_T(s_T^*)$. Based on this value, we can then follow a similar procedure to uncover the optimal state value at time step $T - 2$. This general backtracking approach can be performed successively as a “look up” procedure.
using the stored $\psi$ values to uncover the full optimal hidden state sequence $S^*$:

\[ s^*_t = \psi_{t+1}(s^*_{t+1}) \quad t = T - 1, T - 2, \cdots, 1 \]  

(8)

For a fully-connected (i.e. ergodic) HMM with $K$ states, the time complexity of the Viterbi algorithm is $O(TK^2)$. In many practical applications however, it is common to use an HMM architecture in which only certain pairs of states are connected together. In this case, when computing Eq. (6) we need only maximise over those states which can be reached from state $i$. When the HMM state space is large but the average state connectivity is small, computing $\delta_t(i)$ in this manner can lead to a considerable computational saving. This is particularly the case when using an HMM with built-in “duration constraints”, as considered later in this paper.

**HMM Architectures for ECG Interval Analysis**

We can assign the various regions of the ECG to specific HMM states in a number of different ways. Fig. 2(a) shows a three-state model with separate states* for the PQ, QT, and baseline region of the ECG. Alternatively, to estimate the QT interval only, we could use a simple two-state model with a single state for the QT interval and an additional “catch-all” state (X) for the remainder of the ECG, as shown in Fig. 2(b). The advantage of this type of model architecture is that the compact state space results in a smaller number of model parameters (compared with larger state space models) and more efficient inference. The disadvantage of this approach however is that by modelling multiple waveform features within a single state (e.g. the QRS complex and the T wave), the accuracy of the model segmentations is typically sub-optimal since the observation models are spread over a wider region of the data space [22].

*For the purposes of clarity, we use the HMM state label “PQ” to refer to the region of the ECG from the onset of the P wave to the onset of the QRS complex. However, the timing interval defined by this region is commonly referred to as the “PR interval” (see Fig. 1) [21].
interval, the HMM provides a better generative model of the ECG, which in turn helps to improve the accuracy of the segmentations produced by the model [22].

**Validity of HMM Assumptions for ECG Modelling**

The HMM framework for signal modelling is based on a number of key assumptions. We now consider the two most important assumptions in turn and discuss their validity in the context of ECG signal modelling.

**Observation Independence Assumption**

In the standard formulation of a hidden Markov model, the observation values “within” a given state are considered to be independent and identically distributed (i.i.d.). Hence, when an observation value is generated by a particular state at a given time step, it is generated independently of any previous samples which may have been generated from that same state in previous time steps. As a result of this assumption, the conditional probability

\[ p(O_{1:T} \mid S_{1:T}, \lambda) = \prod_{t=1}^{T} p(O_t \mid S_t, \lambda) \]  

(9)

The assumption of statistical independence between successive observations (within a state) is perhaps the most serious limitation of the standard hidden Markov model, and presents a significant challenge to the use of HMMs for ECG signal modelling. In particular, ECG samples from a given waveform feature clearly exhibit strong correlations over the range of the feature. Indeed, it is these statistical dependencies between successive samples which give rise to the characteristic waveform patterns which make up a typical ECG signal. By ignoring such dependencies, the standard hidden Markov model is unlikely to provide a faithful statistical description of the ECG. Furthermore, since these statistical dependencies give rise to the shape of the waveform features, they are therefore essential in determining the feature boundaries for the purpose of ECG interval analysis.

A number of different approaches have been suggested to overcome the assumption of statistical independence between observations within a given HMM state. One possibility is to make use of an autoregressive (AR) model for the observation density within each HMM state [17, 23]. With this approach, the AR model is used to model the observations within a given state, and the corresponding observation probabilities are then derived from the Gaussian noise process for the AR residuals. However, the performance of this form of HMM is governed by the degree to which the observations from each model state can successfully be represented as an AR process.
An alternative approach to incorporating information from temporal dependencies within each model state is to encode this information in the observation vector itself. This is the dominant approach in modern speech recognition research with hidden Markov models. In particular, “contextual” information from neighbouring samples of speech is captured in two different ways. Firstly, the speech signal is encoded using a technique based on the short-time Fourier transform. The resulting feature vectors therefore incorporate spectral information over the length of the window centred on each signal sample. Secondly, the dependencies between adjacent feature vectors are captured through the use of derivative features or so-called “delta coefficients” [24]. The use of time-frequency transforms and derivative features to provide a more suitable representation of the ECG for hidden Markov modelling is discussed in the following section.

Transition Stationarity Assumption

The transition stationarity assumption is concerned with the time-invariant nature of the state transition probabilities. Thus, for any two times \( t_1 \) and \( t_2 \), we assume that:

\[
P(s_{t_1+1} = j \mid s_{t_1} = i) = P(s_{t_2+1} = j \mid s_{t_2} = i)
\]

(10)

An important consequence of time-invariant transition probabilities is that the individual HMM state durations follow a geometric distribution. This distribution is a poor match to the true state duration distributions of many real-world signals, including the ECG, and can lead to physiologically implausible segmentations. This issue, and its associated remedy, are discussed in more detail in the Duration Constraints section.

SAMPLE-WISE ECG ENCODING

In this section we describe the combination of the undecimated wavelet transform and associated derivative features, which we use to generate a sample-wise encoding of the ECG for subsequent hidden Markov modelling.

Wavelet Representations

Wavelets are a class of functions that are well localised in both the time and frequency domains, and which form a basis for all finite energy signals. The atoms used in wavelet analysis are generated by scaling and translating a
single mother wavelet $\psi$. The general wavelet transform (WT) of a signal $x \in L^2(\mathbb{R})$ is given by:

$$ W(s, \tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t) \psi^* \left( \frac{t - \tau}{s} \right) dt $$

(11)

where $W(s, \tau)$ are the wavelet coefficients at scale $s$ and time-shift $\tau$, and $\psi^*$ is the complex conjugate of the mother wavelet. The specific way in which the mother wavelet is scaled and translated leads to a number of different wavelet transform algorithms.

Table 2 shows a comparison of the key properties of the three main types of wavelet transform. The continuous wavelet transform, or CWT, makes use of the set of continuous scales and time-shifts of the mother wavelet. Specifically, the CWT is defined for all positive scales $s \in \mathbb{R}^+$ and all time-shifts $\tau \in \mathbb{R}$. This leads to a highly redundant representation of the original signal, since successive wavelet atoms (across neighbouring scales or time-shifts) are strongly correlated. Although this redundancy can be advantageous for visualisation (where the “richness” of the coefficients aids the visual interpretation), it can be problematic for many signal analysis and pattern recognition techniques which depend upon compact low-dimensional representations of the data. For example, if we wish to model the pdf of the wavelet coefficients using a Gaussian mixture model, the number of model parameters which must be estimated for each Gaussian is proportional to the square of the dimension of the data.

TABLE 2 about here.

An alternative to the CWT is given by the discrete wavelet transform, or DWT, which decomposes the signal of interest over wavelet atoms that are mutually orthogonal. This is achieved by using only dyadic scales and time-shifts in the wavelet transform (the term dyadic refers to values based on “powers of two”). The wavelet representation produced by the DWT is considerably more efficient than that for the CWT since the signal is projected onto orthogonal basis functions. Furthermore, the DWT can be computed in a computationally efficient manner using a multirate filter bank structure [25].

Although the DWT is well suited to applications such as data compression, it suffers from a number of important drawbacks which limit its use for signal modelling and segmentation problems. Firstly, the DWT does not provide a constant-length vector of wavelet coefficients for each individual time sample. Instead, the coefficients corresponding to different scales (or levels) of the DWT algorithm are governed by a sampling rate which is unique to that particular scale. This stems from the restriction that the wavelet atoms are translated in time by integer multiples of the (dyadic) scale, which leads to different numbers of wavelet coefficients at different scales.
An important consequence of the variation in sampling of the coefficients at different levels of the DWT is that the wavelet coefficients are not translation-invariant. Translation-invariance refers to the property that “when a pattern is translated, its numerical descriptors should be translated but not modified” [25]. For many pattern recognition applications, it is important to construct representations of the input data that are robust with respect to translations or shifts of the input data. In the case of the DWT however, translated (i.e. circularly shifted) versions of a given signal can correspond to wavelet representations which are numerically very different, even after the translations have been corrected for [26].

Given the preceding discussion, we can identify four properties which must be satisfied if a given time-frequency representation is to be effective for the task of ECG signal modelling and segmentation:

- Constant-length vector of coefficients at each time sample
- Compact low-dimensional representation at each time sample
- Translation-invariance of coefficients
- Computationally efficient implementation

A representation which meets all of the above conditions is provided by the undecimated wavelet transform [27, 28, 26]. The undecimated wavelet transform, or UWT, restricts the scale parameter $s$ to the dyadic range $s = 2^k$ (for $k \in \mathbb{Z}$) but allows the time-shift parameter $\tau$ to take on any value, i.e. $\tau \in \mathbb{R}$. This results in a representation which is redundant in the time domain, but orthogonal in the frequency domain. More formally, the set of wavelet atoms employed by the UWT algorithm form a frame or an overcomplete basis.

Importantly, the UWT provides a constant-length vector of wavelet coefficients for each time sample of the original signal. These coefficients are guaranteed to be translation-invariant, such that a (circular) time-shift of the signal will result in an identical (circular) time-shift of the wavelet coefficients [25]. In addition, for a signal of length $T$, the UWT can be computed in $O(T \log T)$ using a fast filter bank algorithm [28]. These properties make the UWT an effective representation for ECG signal modelling and segmentation.

†More precisely, for any given scale the UWT atoms at different time-shifts will not (in general) be orthogonal, however for any given time-shift the UWT atoms at different scales will be orthogonal.
Choice of Wavelet

The first step in applying the undecimated wavelet transform to the ECG is to choose an appropriate wavelet for the analysis. For most practical problems of interest, the choice of wavelet depends primarily on the particular application at hand. In the specific case of time-series segmentation however, we can identify a number of requirements that an effective wavelet should satisfy:

- Good time localisation
- Appropriate degree of smoothness for the signal under consideration
- Realisation as a conjugate mirror filter (CMF) pair
- CMFs possess a near-linear phase response

The first requirement can be achieved by simply minimising the width of the wavelet (and hence the length of the corresponding filters). The second requirement is designed to ensure that the wavelet provides a good match to the characteristic features of the signal under consideration. In general, wavelets with a longer width tend to offer a greater degree of smoothness, whereas wavelets with a shorter width are better suited to the analysis of sharp transients [28]. In the case of the ECG, we wish to detect both sharp transients (i.e. the onset and offset of the QRS complex) as well as more smooth features (i.e. the end of the T wave). Hence the smoothness of the wavelet is a compromise between these two competing requirements.

The third requirement relates to the type of wavelet transform algorithm which can be used with the given wavelet. In particular, if we wish to perform an undecimated (or discrete) wavelet transform, then the wavelet must admit a realisation as a conjugate mirror filter pair (this is a lowpass and a highpass filter that satisfy particular conditions [25]). The final requirement of near-linear phase filters is designed to ensure that the wavelet coefficients at different scales of the UWT algorithm can be aligned in time by means of a simple (circular) time-shift.

Taken together, the above requirements imply that a suitable wavelet for ECG analysis should possess a relatively short width, offer a reasonable amount of smoothness, and provide a CMF pair which offer a close approximation to linear phase filters. In practice, the first wavelet of the Coiflet family, commonly termed C(6) since the corresponding filters are of length 6, satisfies all of these requirements [28].

Previous work on the application of wavelets to ECG signal analysis has primarily focused on wavelets which are either the first or second derivative of a Gaussian. In particular, Li et al. [9] and Sahambi et al. [10] used the first
derivative of a Gaussian as an analysing wavelet to decompose the ECG signal. More recently, Andreao et al. [12] used the second derivative of a Gaussian (known as the Mexican Hat function) to decompose the ECG signal.

Unfortunately, wavelets defined as a derivative of a Gaussian cannot be realised as conjugate mirror filters. This restricts their use to the continuous wavelet transform only. We can however approximate the first and second derivatives of a Gaussian using wavelets which can be realised as conjugate mirror filters. In particular, the biorthogonal spline wavelet BIOR(6,2) with lowpass and highpass filters of length 6 and 2 respectively, provides a reasonable approximation to the first derivative of a Gaussian. Similarly, the “least asymmetric” Daubechies wavelet with a filter length of 12, commonly termed LA(12), provides a useful approximation to the second derivative of a Gaussian.

Fig. 3 shows the three different wavelets which we investigate in this paper for ECG analysis (see Experiments section). To encode an ECG signal with the UWT (for a given wavelet basis), we compute the wavelet coefficients at seven different levels/scales of the UWT filter bank. The result of this encoding is a 7-D vector of wavelet transform coefficients for each time sample in the ECG signal.

[FIGURE 3 about here.]

Derivative Features

A significant limitation of the standard wavelet representation of a signal is the manner in which phase information about the signal is encoded. To illustrate this point, consider the first level UWT coefficients produced in the region of the T wave of an ECG. If the filters associated with the given wavelet are approximately linear phase (as is the case with many commonly used wavelet families), then the corresponding impulse responses will be close to symmetrical. If we now assume that the T wave itself is approximately symmetrical, then the resulting wavelet coefficients at the first level of the UWT analysis will also be approximately symmetrical about the peak of the T wave. As a result, the phase of the ECG signal at each time sample is not well characterised by the wavelet coefficients.

The issue of phase encoding with wavelet representations is particularly significant in the context of automated ECG interval analysis. Since we wish to detect the onset and offset points of the various waveform features as accurately as possible, it is essential that the chosen representation of the ECG enables the automated system to differentiate between the upslope and the downslope of each of the ECG features.

A particularly simple and effective technique for encoding a measure of phase from a signal is to make use of
the signal derivative. Since this quantifies the local gradient of the signal at each time point, it can be used to aid in the differentiation of the various regions of a given waveform. The use of derivative features (or so-called “delta coefficients”) has been found to produce a substantial improvement in the performance of HMM-based automatic speech recognition systems [24, 29].

In order to implement derivative features in practice, it is necessary to first smooth the signal of interest such that the estimate of the derivative is not unduly affected by high frequency noise. However, if the signal is to be encoded with a wavelet transform, then one or more frequency bands from the encoding can be used in place of the original signal for the derivative computations. This approach is advantageous since the frequency bands can be chosen to minimise the effects of noise and emphasise the underlying signal content.

In order to incorporate derivative information into the UWT encoding of the ECG, we augmented the 7-D wavelet coefficient vector at each time sample with the derivative of the coefficients from levels 4, 5 and 6 (i.e. scales $2^4$, $2^5$ and $2^6$) of the UWT filter bank. These levels correspond to the frequency range of 3 Hz to 34 Hz, which in turn corresponds to the dominant spectral content of the ECG. The derivative of the wavelet coefficients at each level was computed using a simple first-order central difference:

$$\Delta(s, \tau) = \frac{W(s, \tau + 1) - W(s, \tau - 1)}{2}$$  \hfill (12)

where $\Delta(s, \tau)$ is the derivative feature associated with the wavelet coefficient at scale $s$ and time-shift $\tau$.

Incorporating derivative features into the UWT encoding leads to a 10-D feature vector of UWT coefficients and the associated derivative features at each time sample.

Fig. 4 illustrates the impact of derivative features on the accuracy of the T wave offset segmentation for a particular ECG signal. The upper plots show the T wave region of the ECG, together with the T wave offset point determined by a.) an HMM with the UWT encoding only (upper left plot), and b.) an HMM with the UWT + derivative features encoding (upper right plot). In each case the T wave offset point determined by an expert analyst is also shown. For the HMM with the UWT encoding only, it is evident that the model infers the T wave offset close to the peak of the T wave, and much earlier than the true T wave offset location. By incorporating derivative features into the ECG encoding however, the accuracy of the T wave offset segmentation is considerably improved.

[FIGURE 4 about here.]

In order to understand how derivative features serve to improve the accuracy of the segmentations produced by
an HMM, it is useful to consider the log observation probabilities for the various HMM states. The lower plots in Fig. 4 show, for each HMM, the difference between the log observation probabilities for the JT interval state and those for the Baseline state, evaluated over the course of the T wave (shown in the upper plots). For the HMM without derivative features (lower left plot), the log observation probabilities for the Baseline state exceed those for the JT interval state significantly before the actual end of the T wave. This results in a substantial error in the T wave offset location inferred by the Viterbi algorithm. Conversely, for the HMM with derivative features (lower right plot), the log observation probabilities for the JT interval state exceed those for the Baseline state throughout the downslope of the T wave, resulting in a significantly more accurate segmentation.

In the Experiments section we present detailed results comparing the performance of hidden Markov models trained with different wavelet encodings, and demonstrate the improvement in ECG segmentation accuracy which can be gained with derivative features.

**DURATION CONSTRAINTS FOR ROBUST SEGMENTATIONS**

As indicated previously, a significant limitation of the standard HMM framework is the manner in which state durations are modelled. In particular, for a given state $i$ with self-transition coefficient $a_{ii}$, the probability mass function of the state duration $d$ is a geometric distribution, given by:

$$P_i(d) = (a_{ii})^{d-1}(1 - a_{ii})$$  \hspace{1cm} (13)

In the case of the ECG, the geometric distribution provides only a poor match to the statistical characteristics of the true state durations. Fig. 5 shows the duration distribution curves for the JT interval, based on i.) a gamma density fitted to the JT interval durations derived from the expert analyst annotations (dashed line), and ii.) the geometric duration distribution for the corresponding HMM state (solid line). It is evident that much of the probability mass of the geometric distribution lies to the left of the minimum duration for the JT interval.

[FIGURE 5 about here.]

This mismatch between the statistical characteristics of the model and those of the ECG gives rise to the problem of *double-beat* segmentations, an example of which is shown in Fig. 6. Such segmentations occur when the model incorrectly infers two (or more) beats when only a single beat is present in a particular region of an ECG [22]. The “rogue” beat segmentations are characterised by hidden state sequences of a very short duration.
Hughes & Tarassenko Technical Report BSP 08-01

(typically much shorter than the minimum state duration observed with genuine ECG signals). This in turn leads to automated ECG interval measurements which are highly unreliable.

[FIGURE 6 about here.]

There are two possible approaches to improving the duration characteristics of a hidden Markov model and hence the robustness of the associated ECG segmentations. The first approach is to restrict each of the hidden states within the model such that they can only generate “runs” of states which satisfy a given minimum duration constraint. The second approach is to relax the geometric duration distribution for each HMM state and to enable the model to represent arbitrary state duration distributions. This latter approach leads to a class of statistical model known as a hidden semi-Markov model, or HSMM [30]. Although these models offer considerable flexibility for state duration modelling, in practice the accuracy of the resulting ECG segmentations is very similar to that produced by an HMM with minimum state duration constraints [22]. Furthermore, the improved duration modelling of the HSMM comes at the expense of a considerable increase in the time complexity of the associated Viterbi inference procedure. We therefore focus in this paper on the use of duration constraints for improving the robustness of the HMM segmentations.

Fig. 7 illustrates how a minimum state duration constraint can be incorporated into the architecture of an HMM. For a given state $k$ with a minimum duration of $d_{\text{min}}(k)$ (where $d_{\text{min}}(k) > 1$), we augment the model with $d_{\text{min}}(k) - 1$ additional states directly preceding the original state $k$. Each additional state has a self-transition probability of zero, and a probability of one of transitioning to the state immediately to its right. Thus taken together these states form a simple linear Markov chain, where each state in the chain is only occupied for at most one time sample (during any run through the chain).

[FIGURE 7 about here.]

An important feature of this chain is that the observation probability distribution for each additional state is tied to the distribution for the original state $k$. Thus the observations associated with the $d_{\text{min}}$ states associated with the original state $k$ are governed by a single set of parameters (which is shared by all $d_{\text{min}}$ states). The use of parameter “tying” serves to reduce the number of free parameters in the resulting HMM.

From Fig. 7 it is easy to understand the effect of the duration constraints upon the generative nature of the model. Once the model has transitioned into state $k_1$ it is then forced to emit $d_{\text{min}}(k)$ samples from the observation probability distribution associated with state label $k$ before it has the opportunity to enter a state with a different...
observation distribution. The result is that the model is prevented from generating “runs” of state $k$ whose duration is less than $d_{\text{min}}(k)$.

In practice, the use of an appropriate set of minimum state duration constraints with an HMM prevents the model from inferring double-beat segmentations, and leads to a considerable improvement in the robustness of the associated ECG interval measurements. Table 3 shows the number of single-beat and double-beat segmentations on leads II and V2 of the test set (see Experiments section for more details), for a standard HMM and an HMM with duration constraints. For both leads II and V2, the standard HMM produces a significant number of double-beat segmentations. However by incorporating duration constraints into the HMM architecture such rogue beat segmentations are eliminated.

[TABLE 3 about here.]

**CONFIDENCE MEASURES FOR HMM SEGMENTATIONS**

The main focus of the work presented so far has been on improving the accuracy and robustness of the ECG segmentations produced by a hidden Markov model. This is clearly an important and worthwhile goal, since any automated system for ECG interval analysis must reach a certain performance level if it is to be used successfully in a real-world setting.

In practice however, it is not realistic to expect the model (or any other automated system) to maintain a consistently high level of segmentation accuracy for every possible ECG waveform with which it might be presented. Thus, we must accept that the model may produce unreliable segmentations when analysing ECG waveforms which are sufficiently novel with respect to the waveforms on which the model was trained.

More generally, it is often the case that only a subset of the heartbeats contained within a given ECG data set are suitable for accurate and reliable ECG interval measurements (either by an expert or by an automated system). ECG signals often contain beats which are corrupted by noise or artefact, such that the various intervals cannot be measured to a high degree of accuracy [1]. In this case, it is essential that the automated system has the ability to highlight those beats for which it is least confident in its analysis.

We now focus therefore on the problem of automatically detecting unreliable ECG segmentations, and hence improving the level of sophistication of the ECG analysis produced by the trained model. More generally, we wish to enable the model to detect automatically those ECG beats which are least suitable to analysis by machine, and thus most in need of analysis by an expert cardiologist. In this way, we aim to combine the advantages of automated
analysis with those of manual analysis.

**Assessing Confidence Using the Joint Log Likelihood**

In the context of automated ECG interval analysis, an effective confidence measure should provide an assessment of the quality of the waveform segmentations produced by the automated system and hence the reliability of the associated interval measurements. More generally, we can view the confidence measure as quantifying the “normality” of the ECG waveform under consideration (with respect to the waveforms in the training set). Intuitively we should have more confidence in the segmentations of ECG waveforms which are similar to those on which the model was trained, compared with the segmentations of waveforms which are markedly different from those on which the model was trained.

By formulating the confidence measure in terms of the normality of the ECG waveform, we can utilise the full probabilistic nature of the models considered in this paper in order to derive a suitable quantitative definition. Specifically, since the models considered in this paper are generative models, they inherently define a joint probability distribution over observation signals and hidden state sequences, i.e. \( p(O, S \mid \lambda) \). This probability distribution can therefore be exploited in order to assess the normality of a given ECG waveform and the associated segmentation produced by the model.

More formally, for an observation signal \( O \) with corresponding Viterbi state sequence \( S^* \), we can compute the joint log likelihood as:

\[
\log p(O_{1:T}, S^*_{1:T} \mid \lambda) = \log \pi_{s^*_1} + \sum_{t=1}^{T-1} \log a_{s^*_t s^*_{t+1}} + \sum_{t=1}^{T} \log b_{s^*_t}(O_t) \tag{14}
\]

In practice, a 10-second ECG signal will typically contain a number of heartbeats and hence a number of interval measurements. We would therefore like to compute a separate log likelihood value for each individual heartbeat in an ECG (with onset time \( t_1 \) and offset time \( t_2 \)). This can be achieved by simply redefining the limits of the summation in Eq. (14):

\[
\log p(O_{t_1:t_2}, S^*_{t_1:t_2} \mid \lambda) = \log \pi_{s^*_1} + \sum_{t=t_1}^{t_2-1} \log a_{s^*_t s^*_{t+1}} + \sum_{t=t_1}^{t_2} \log b_{s^*_t}(O_t) \tag{15}
\]
Modelling the Joint Log Likelihood

When assessing the joint log likelihood value for a given ECG heartbeat, we must also consider the duration of the beat under consideration (i.e. $t_2 - t_1$). To understand why this is necessary, it is useful to consider a simple example. Consider an HMM $\lambda$, which generates a signal $O$ of length $T$ according to the hidden state sequence $S$. The joint log likelihood of the signal and the state sequence is then given by Eq. (14). If at the following time step the model transitions to hidden state $s_{T+1}$ and generates a signal sample $O_{T+1}$ (by sampling from the observation density for that state), then the joint log likelihood will change by a factor of $\log a_{s_T s_{T+1}} + \log b_{s_{T+1}}(O_{T+1})$.

Crucially, the most significant contribution to the change in the log likelihood will be due to the observation probability term $b_{s_{T+1}}(O_{T+1})$. This is because the transition probabilities span the range $[0, 1]$, whereas the observation probabilities (which are densities) span the range $[0, \infty)$. As a result, the dynamic range of the log observation probabilities is much greater than that of the log transition probabilities.

Given the preceding analysis, we might expect an approximately linear relationship between the joint log likelihood $p(O_{t_1:t_2}, S^*_{t_1:t_2} | \lambda)$ and the duration of the ECG beat under consideration ($t_2 - t_1$). Fig. 8 shows a scatter plot of the joint log likelihood values against the beat duration for the annotated ECG beats from the training set for lead V2. As expected, there is a linear trend to the data points shown in the scatter plot.

[FIGURE 8 about here.]

Given this linear trend, we can model the relationship between the log likelihood and the beat length using standard linear regression techniques [31]. In particular, we assume a simple linear model of the form:

$$\log p(O_{t_1:t_2}, S^*_{t_1:t_2} | \lambda) = v_0 + v_1(t_2 - t_1) + \epsilon_i$$  \hspace{1cm} (16)

where $v_0$ and $v_1$ are the parameters of the model (the intercept and gradient, respectively), and $\epsilon$ is a Gaussian noise process with zero mean and variance $\sigma^2$. These parameters can be estimated in a standard manner using least-squares.

From the linear regression model we can evaluate the lower confidence bound at the 99% level of significance (indicated by the dashed line in Fig. 8). This lower bound then provides a suitable threshold which can be used to determine if ECG beats segmented by the model are sufficiently novel as to cast doubt on the reliability of the associated interval measurements.
Normalised Confidence Measures

In a practical setting, it is desirable to produce a normalised confidence measure for each ECG beat. If each log likelihood value is normalised to the range [0, 1] (where a value of 0 represents minimum confidence, and a value of 1 represents maximum confidence), then the normalised values can be used directly (without the need for graphical display of the log likelihoods) and can be readily interpreted by cardiologists.

In order to normalise each log likelihood value, we need to map the log likelihood range \([0, \infty)\) to the normalised range \([0, 1]\). Specifically, we require a normalisation mapping which is appropriate for the distribution of log likelihood values for each beat duration. Such a mapping can be achieved through the use of the sigmoid function [19]:

\[
y = \frac{1}{1 + e^{-\nu(x-\mu)}}
\]

(17)

where \(x\) is the input to the function (i.e. the joint log likelihood value), \(y\) is the output of the function (i.e. the normalised confidence value), and \(\mu\) and \(\nu\) are parameters which govern the characteristic form of the function.

In order to set appropriate values for the sigmoid parameters (at a given beat duration), we must first select two distinct normalised confidence values corresponding to two distinct log likelihood values. We set the normalised confidence value at the 99% lower confidence bound to 0.7, and the normalised confidence value at the mean regression line to 0.85. Given these values, we can then rearrange Eq. (17) and solve for the sigmoid parameters \(\mu\) and \(\nu\) at a given beat duration.

With the above procedure, we can convert the log likelihood value (at a given beat duration) to a normalised confidence value by simply applying the sigmoid mapping function. If we then set the “confidence threshold value” (i.e. the confidence value below which we will label segmented beats as low confidence) to 0.7, we can automatically exclude from any subsequent analysis those ECG beats whose segmentations are deemed to be unreliable.

In the following Section, we evaluate the performance on our confidence-based approach to ECG interval analysis on the thorough QT study data set.

EXPERIMENTS

Construction of Training and Test Sets

In order to evaluate the performance of HMMs for automated ECG interval analysis, it is desirable to use independent training and test sets. In particular, if we wish to assess how an HMM-based automated system is likely to
perform in practice when used to analyse ECG data from a clinical trial, we should evaluate the performance of the model on ECGs from subjects who were not present in the training set. This reflects the way in which we would expect the automated system to be used in practice (it is possible however for the model to be “re-trained” for each new study, however this would require manual annotations for a subset of the ECG data from the study). In this way, we can evaluate the ability of the trained model to generalise to ECG data from new subjects.

Given the above, we partitioned the thorough QT study data set into approximately equally-sized training and test sets in the following manner. For each of leads II and V2, the ECGs in the data set with manual annotations on that particular lead were first selected together with the corresponding subject identifiers. From this subset of the ECGs in the study, the 10-second ECG signals from a random selection of half of the subjects were used to form the training set. The 10-second ECGs recorded from the remaining subjects were then used to form the test set. Following the partitioning, the training and test sets for each lead consisted of approximately 6000 annotated ECG beats, with the ECGs in each set recorded from different subjects.

ECG Encoding Schemes

We evaluated four different sample-wise ECG representations, namely:

1. Bandpass filtered ECG time-series

2. Short-time Fourier transform (STFT)

3. Undecimated wavelet transform (UWT)

4. UWT + derivative features

The bandpass filtered ECG time-series representation was formed by simply bandpass filtering each ECG signal using a linear-phase FIR filter with lower and upper half-power cutoff frequencies at 1.5 Hz and 35 Hz respectively.

For the short-time Fourier transform, we used a sliding window procedure to form a sample-wise encoding of the ECG power spectrum. Specifically, for each time sample we first applied a Hann window (of length 256 samples, centred on the given sample) to the signal and then computed the power spectrum of the resulting windowed signal using the FFT. This procedure was repeated for successive ECG samples (by sliding the window along by one sample) until the entire 10-second ECG signal had been fully encoded. Finally, we formed an 8-dimensional representation of the power spectrum (at each time sample) by averaging “blocks” of consecutive coefficients.
For the undecimated wavelet transform encoding, we computed the wavelet coefficients at seven different levels/scales of the UWT filter bank (generating a 7-D wavelet coefficient vector at each time sample). As discussed previously, we experimented with three different wavelet bases: the C(6) Coiflet wavelet, the BIOR(6,2) biorthogonal wavelet, and the LA(12) “least asymmetric” Daubechies wavelet. The latter two wavelets were chosen due to their similarity to the first and second derivatives of a Gaussian, respectively (see Choice of Wavelet section).

Finally, we selected the wavelet basis with the best overall test set performance (based on the segmentation accuracy for the Q, R and Toff points), and augmented the 7-D wavelet coefficient vector at each time sample with the derivative of the coefficients from levels 4, 5 and 6 of the UWT filter bank (see Derivative Features section). This resulted in a 10-D encoding of UWT coefficients and the associated derivative features.

Model Training and Testing

For each ECG encoding scheme, we trained a hidden Markov model in a supervised manner on the annotated ECG waveforms from the training set (we trained separate models for leads II and V2). We used the five-state HMM shown in Fig. 2(c) and incorporated minimum duration constraints into the model architecture as described previously. In particular, the minimum duration for each model state was set to 0.8 of the minimum duration present in the training set for that state (this was value found to be effective in preventing double-beat segmentations whilst allowing accurate segmentations of short ECG waveform features). The initial state distribution and transition matrix for the HMM were then estimated according to Eqs. (2) and (3). We used Gaussian mixture models (GMMs) with full covariance matrices as the observation densities for each of the hidden states, due to their flexibility in modelling multi-modal data. The parameters for each GMM were estimated using a combined model order selection and EM algorithm due to Figueiredo & Jain [32].

The final stage of the model training procedure consisted of learning the HMM confidence model from the annotated ECG waveforms. This was achieved by first computing the joint log likelihood value, according to Eq. (15), for each annotated ECG beat in the training set. A linear regression model was then fit to the set of \{beat duration, log likelihood\} pairs, and the mean regression line and 99% lower confidence bound determined. For a segmented beat from the test set, a normalised confidence measure could then be computed by mapping the log likelihood score through a sigmoid function fit to the mean regression and lower confidence values at the given beat duration (see Confidence Measures section for further details).

In the model testing phase, the trained HMM was used to segment each 10-second ECG signal in the test set.
The segmentation was performed by using the Viterbi algorithm to infer the most probable underlying sequence of hidden states for the given signal. Note that the full 10-second ECG signal was processed, as opposed to just the manually annotated ECG beats, in order to mimic the way an automated system would be used for ECG interval analysis in practice. Next, for each segmented ECG beat, we computed the normalised HMM confidence measure using the learnt confidence model, as described previously. The model segmentations corresponding to the particular beats which had been manually annotated were then extracted, and the differences between the model segmentations and the corresponding expert analyst annotations were computed for those beats with a confidence measure $\geq 0.7$ (i.e. high confidence beats only).

The overall training and testing procedure (for each of leads II and V2) is summarised in Fig. 9.

[FIGURE 9 about here.]

In terms of computational complexity, our algorithm consists of three key stages: 1.) computation of the undecimated wavelet transform, 2.) computation of the Viterbi algorithm, and 3.) computation of confidence measures. Each of these stages can be performed in an efficient manner. In particular, for an ECG signal of length $T$, the undecimated wavelet transform of the signal can be evaluated in $O(T \log T)$, and the model segmentations using the Viterbi algorithm in $O(T K C_{avg})$, where $K$ and $C_{avg}$ are the total number of hidden states in the HMM and the average state connectivity, respectively. Furthermore, the joint log likelihood values required for the HMM confidence measures can be computed in $O(T)$. Thus, our method is computationally efficient and can be readily used in practice with very large data sets of ECG signals (such as those collected in thorough QT studies).

Results

Table 4 shows the mean errors (and the corresponding standard deviations in parentheses) between the hidden Markov model segmentations and the expert analyst annotations on leads II and V2 of the test set. For each ECG annotation point, the smallest mean error (across the six different encodings) is highlighted in bold.

[TABLE 4 about here.]

In our experiments we found that the use of the bandpass filtered ECG time-series as the observation sequence for an HMM led to inaccurate and unreliable segmentations on the test set. In particular, the trained model tended to produce very early P wave onset and T wave offset segmentations, most noticeably on the ECGs from lead II.
This poor performance is most likely a result of the considerable overlap in the distributions of the signal samples for the various ECG wave shapes, which in turn leads to a lack of discrimination between the Gaussian mixture observation models for the corresponding HMM states.

The three ECG transform encodings (STFT, UWT, and UWT + derivatives) each resulted in a substantial improvement in the model segmentation performance compared with the filtered time-series representation. For both leads II and V2, there is a marked improvement in the accuracy of the segmentations with the UWT encoding, compared with the STFT. This highlights the advantage of using an adaptive time-frequency representation over a fixed window approach.

For the three different wavelet bases investigated, the BIOR(6,2) wavelet resulted in the best overall performance. Most significantly, the BIOR(6,2) encoding led to a considerable improvement in the R peak accuracy, compared with that for the C(6) and LA(12) encodings. These results can be understood in terms of the similarity of the BIOR(6,2) wavelet to the first derivative of a Gaussian (see Fig. 3 previously). The latter is known to be optimal for the detection of discontinuities in signals and images (which have been corrupted by Gaussian noise), and is commonly used for edge detection in image processing [33].

The use of derivative features with the UWT BIOR(6,2) encoding adds a further level of accuracy to the HMM segmentations. This improvement is most pronounced for lead V2, for which the accuracy of the segmentations for both the QRS onset and T wave offset points is improved significantly. For lead II, there is a small improvement in the T wave offset error with the addition of derivative features. This indicates that the BIOR(6,2) wavelet alone is able to capture most of the salient information in the ECG signal for this particular lead.

When assessing the performance of the trained HMMs, it is advantageous to consider the inter-analyst variability inherent in the expert analyst ECG annotations. Given that a total of 21 expert analysts were used to annotate the thorough QT data set, we would naturally expect a degree of variability in the manual QT measurements produced by different analysts. Previous investigations into this issue have shown that the inter-analyst variability (defined as the standard deviation of the differences in the QT intervals, for two analysts measuring the same set of ECG waveforms) is typically of the order of 10 ms [34, 35].

For the HMM trained with the UWT BIOR(6,2) + derivative features encoding, the standard deviation of the errors in the QT interval (i.e. automated QT - manual QT) for leads II and V2 is 12.4 ms and 10.4 ms, respectively. Taking the inter-analyst variability in the manual QT measurements to be approximately 10 ms, we can conclude that the model for both leads is performing close to the optimum level possible (given the variability inherent in the data set annotations).
Analysis of Confidence Measures

Fig. 10 shows a plot of the cumulative distribution of the HMM confidence measures for the manually annotated beats from lead II of the test set. A notable feature of the cumulative confidence distribution is that only a small proportion of the manually annotated ECG beats have a low confidence value. This is to be expected however since the beats chosen for manual analysis by an expert will typically correspond to clean ECG waveforms with little noise or artefact. Applying a confidence threshold of 0.7 results in 7.4% of the model segmentations for ECG beats in the lead II test set being labelled as low confidence, and 8.3% of the beats in the lead V2 test set.

[FIGURE 10 about here.]

Fig. 11 shows a selection of the low confidence beats (i.e. confidence < 0.7) from the HMM segmentations of the full 10-second ECG signals for the lead II test set. In each plot, the confidence value specified is an average of the confidence values for the beat segmentations shown. For the ECGs in plots A, B and C, the Viterbi algorithm was unable to locate any valid beats in the 10-second ECG signal. This was because the state observation probabilities were extremely small, and thus $\delta_t(i)$ for each state (as computed in the Viterbi algorithm) was equal to zero for much of the duration of the signal. As a result, it was not possible to segment these ECG signals and the confidence values were therefore set to zero.

[FIGURE 11 about here.]

The ECGs shown in Fig. 11 exhibit a variety of different ECG noise sources and waveform abnormalities. In particular, we can identify the following in each of the respective plots:

A. Poor electrode contact

B. Electrode contact noise

C. High-frequency mains interference

D. Low amplitude noisy T waves

E. Large amount of movement artefact (between the second and third beats)

F. Abnormal QRS complex morphology
Fig. 12 shows a selection of the high confidence beats (i.e. confidence $\geq 0.7$) from the 10-second ECG signals for the lead II test set. Again, the confidence value specified in each plot is an average of the confidence values for the beat segmentations shown. We can see from Fig. 12 that the high confidence beats are associated with noise-free ECG signals which contain well-defined QRS complexes and T waves. Visual inspection of the model annotations (i.e. the Q and T_{off} points) for these beats indicates that the model is performing to a high degree of accuracy.

Comparison with Threshold-based Algorithm

We compared the performance of our HMM algorithm for automated ECG interval analysis (using the UWT BIOR(6,2) + derivative features encoding scheme) with that of ECGPUWAVE – the standard thresholding algorithm for ECG segmentation [7] (see Introduction). In order to compare and contrast the performance of the two approaches, we produced Bland-Altman plots of agreement between the automated QT intervals (for each of the two methods) and the corresponding manual QT intervals (as determined by the expert analysts). The Bland-Altman plot displays the difference between two sets of measurements against the mean of the measurements [36]. This enables any systematic differences between the measurement techniques to be visualised, together with any trend in the differences.

Fig. 13 shows the Bland-Altman plots for the HMM segmentations (with confidence $\geq 0.7$) and the ECGPUWAVE segmentations, evaluated on lead V2 of the test set. The plot for the ECGPUWAVE algorithm shows a considerable number of outliers, indicated by the large cloud of data points which lie to the upper right of the main data density. These points correspond to significant over-estimates of the QT interval by the algorithm. In contrast, the data points on the HMM plot are well clustered around the mean difference, and there is only a relatively small number of outliers lying outside the two standard deviation range.

DISCUSSION

The introduction of the thorough phase 1 QT/QTc study, and the large volumes of ECG data which are now being generated as a result, has led to a growing demand for accurate and reliable methods for ECG interval analysis. In this context, automated systems offer the potential for a more extensive evaluation of ECG data from clinical trials,
and hence a more robust assessment of drug safety.

In this paper we have presented a new approach to automated ECG interval analysis based on the use of probabilistic models. One of the most powerful features of this approach is the ability to learn a model for ECG segmentation from expert annotated ECG data. Such data-driven learning allows us to sidestep neatly the problem of having to specify an explicit rule for determining the end of the T wave in the electrocardiogram (as with, for example, threshold-based methods).

Perhaps the most important aspect of the probabilistic modelling approach is the ability of the model to generate a statistical confidence measure in its analysis of a given ECG waveform. Current automated ECG interval analysis systems are limited by their inability to differentiate between normal ECG waveforms (for which automated measurements may be reliably inferred), and abnormal or unusual ECG waveforms (for which automated measurements are frequently unreliable). By utilising a confidence-based approach to automated ECG interval analysis however, we can automatically highlight those waveforms which are least suitable to analysis by machine (and thus most in need of analysis by a human expert). This strategy therefore provides an effective way to combine the twin advantages of manual and automated ECG interval analysis.

Although the use of hidden Markov models for ECG analysis has been explored previously in the literature [11, 12], the joint probabilistic description these models provide has not previously been fully exploited. We have shown how the confidence in the model segmentations can be assessed by evaluating the HMM joint log likelihood \( p(S^*, O) \) with respect to the beat duration. Furthermore, by normalising this log likelihood value in an appropriate manner, it is possible to derive normalised confidence values which can be readily interpreted by cardiologists and medical practitioners.

In order for the automated ECG interval analysis system to produce both accurate segmentations together with robust and reliable confidence measures, it is necessary that the HMM provides a good model of the statistical characteristics of normal ECG waveforms (i.e. the trained HMM is a suitable description of ECG waveform normality). Much of the research described in this paper is driven by the requirement to improve the “fit” between the statistical properties of the standard hidden Markov model and those of the electrocardiogram signal.

Motivated by the observation independence assumption inherent in the HMM framework, we have demonstrated that a combination of the undecimated wavelet transform and associated derivative features (evaluated over a number of UWT scales) provides a representation of the ECG which is well suited to analysis by an HMM. Furthermore, the robustness of the segmentations produced by an HMM can be improved considerably by incorporating minimum state duration constraints into the model architecture. In this way the “duration constrained”
HMM is prevented from inferring waveform segmentations that are physiologically implausible.

The results on the thorough QT data set demonstrate that our algorithm is able to produce QT interval measurements to a similar level of accuracy as expert analysts (taking into account the inherent variability in the measurements of different experts). However, to quantify more precisely the sensitivity of our automated system, it would be necessary to evaluate the performance of the model on an ECG data set with a small drug-induced change in the QT/QTc interval (i.e. 5 to 10ms). This level of change can be brought about by the drug moxifloxacin, which is commonly used as a “positive control” in thorough QT studies. Demonstrating the sensitivity of our approach on such data is a key focus of our current research.

ACKNOWLEDGEMENTS

This work was supported in part through the Engineering and Physical Sciences Research Council (EPSRC) under Grant GR/N14248/01 and the UK Medical Research Council (MRC) under Grant No. D2025/31, and in part by Oxford BioSignals Ltd. The authors would like to thank Jay Mason, Iain Strachan, Mahesan Niranjan, Tony Robinson and Steve Young for many helpful comments on this work, and Covance Inc. for providing the clinical data.

References


TABLE 1. Number of 10-second ECG signals and annotated beats for leads II and V2 of the thorough QT study data set.

<table>
<thead>
<tr>
<th>Lead</th>
<th>Num. of 10-second ECGs</th>
<th>Num. of annotated ECG beats</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>4135</td>
<td>12291</td>
</tr>
<tr>
<td>V2</td>
<td>4109</td>
<td>12307</td>
</tr>
</tbody>
</table>
TABLE 2. A comparison of the three main types of wavelet transform.

<table>
<thead>
<tr>
<th>WAVELET TRANSFORM TYPE</th>
<th>Scale $s$, Time-shift $\tau$</th>
<th>Translation-invariant</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous (CWT)</td>
<td>$s \in \mathbb{R}^+, \tau \in \mathbb{R}$</td>
<td>Yes</td>
<td>Highly redundant</td>
</tr>
<tr>
<td>Discrete (DWT)</td>
<td>$s = 2^k, \tau = 2^k l, (k, l) \in \mathbb{Z}^2$</td>
<td>No</td>
<td>Orthogonal basis</td>
</tr>
<tr>
<td>Undecimated (UWT)</td>
<td>$s = 2^k, \tau \in \mathbb{R}, k \in \mathbb{Z}$</td>
<td>Yes</td>
<td>Frame</td>
</tr>
</tbody>
</table>
TABLE 3. The effect of duration constraints on the HMM segmentations.

(a) Lead II test set (5930 beats)

<table>
<thead>
<tr>
<th>MODEL:</th>
<th>Number of segmentations of type:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single-beat</td>
</tr>
<tr>
<td>Standard HMM</td>
<td>5196</td>
</tr>
<tr>
<td>HMM with duration constraints</td>
<td>5930</td>
</tr>
</tbody>
</table>

(b) Lead V2 test set (6742 beats)

<table>
<thead>
<tr>
<th>MODEL:</th>
<th>Number of segmentations of type:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single-beat</td>
</tr>
<tr>
<td>Standard HMM</td>
<td>6343</td>
</tr>
<tr>
<td>HMM with duration constraints</td>
<td>6742</td>
</tr>
</tbody>
</table>
TABLE 4. Mean errors in milliseconds (standard deviations in parentheses) between the “duration-constrained” HMM segmentations (with different ECG encodings) and the manual expert annotations, for leads II and V2 of the test set.

(a) Lead II test set (5930 beats)

<table>
<thead>
<tr>
<th>ECG Encoding</th>
<th>Mean segmentation error (ms)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P_{on}</td>
<td>Q</td>
<td>R</td>
<td>J</td>
<td>T_{off}</td>
</tr>
<tr>
<td>BP filtered time-series</td>
<td>-461.1</td>
<td>-31.5</td>
<td>-62.1</td>
<td>-98.3</td>
<td>-208.2</td>
</tr>
<tr>
<td></td>
<td>(196.7)</td>
<td>(59.6)</td>
<td>(60.7)</td>
<td>(60.6)</td>
<td>(65.7)</td>
</tr>
<tr>
<td>STFT</td>
<td>-8.3</td>
<td>-40.5</td>
<td>-29.2</td>
<td>41.3</td>
<td>-46.0</td>
</tr>
<tr>
<td></td>
<td>(25.4)</td>
<td>(33.6)</td>
<td>(69.4)</td>
<td>(30.6)</td>
<td>(44.5)</td>
</tr>
<tr>
<td>UWT [C(6) wavelet]</td>
<td>-0.8</td>
<td>-2.6</td>
<td>4.4</td>
<td>5.6</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>(16.6)</td>
<td>(4.7)</td>
<td>(10.4)</td>
<td>(7.3)</td>
<td>(19.5)</td>
</tr>
<tr>
<td>UWT [LA(12) wavelet]</td>
<td>-1.3</td>
<td>-3.7</td>
<td>5.8</td>
<td>4.4</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>(15.5)</td>
<td>(5.1)</td>
<td>(11.2)</td>
<td>(7.7)</td>
<td>(17.2)</td>
</tr>
<tr>
<td>UWT [BIOR(6,2) wavelet]</td>
<td>-1.0</td>
<td>-3.0</td>
<td>0.2</td>
<td>5.6</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>(16.4)</td>
<td>(4.4)</td>
<td>(1.5)</td>
<td>(7.1)</td>
<td>(12.6)</td>
</tr>
<tr>
<td>UWT [BIOR(6,2)] + derivatives</td>
<td>1.6</td>
<td>-3.6</td>
<td>0.4</td>
<td>4.5</td>
<td>-2.2</td>
</tr>
<tr>
<td></td>
<td>(8.7)</td>
<td>(5.0)</td>
<td>(1.6)</td>
<td>(7.5)</td>
<td>(11.2)</td>
</tr>
</tbody>
</table>

(b) Lead V2 test set (6742 beats)

<table>
<thead>
<tr>
<th>ECG Encoding</th>
<th>Mean segmentation error (ms)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P_{on}</td>
<td>Q</td>
<td>R</td>
<td>J</td>
<td>T_{off}</td>
</tr>
<tr>
<td>BP filtered time-series</td>
<td>-94.8</td>
<td>-3.3</td>
<td>8.2</td>
<td>-17.3</td>
<td>-76.0</td>
</tr>
<tr>
<td></td>
<td>(254.6)</td>
<td>(45.6)</td>
<td>(3.8)</td>
<td>(17.5)</td>
<td>(42.6)</td>
</tr>
<tr>
<td>STFT</td>
<td>-10.7</td>
<td>-27.2</td>
<td>21.2</td>
<td>18.7</td>
<td>-27.4</td>
</tr>
<tr>
<td></td>
<td>(35.7)</td>
<td>(31.8)</td>
<td>(57.2)</td>
<td>(28.7)</td>
<td>(36.1)</td>
</tr>
<tr>
<td>UWT [C(6) wavelet]</td>
<td>-7.8</td>
<td>-3.5</td>
<td>1.9</td>
<td>3.9</td>
<td>-5.2</td>
</tr>
<tr>
<td></td>
<td>(23.3)</td>
<td>(4.0)</td>
<td>(3.2)</td>
<td>(6.9)</td>
<td>(16.2)</td>
</tr>
<tr>
<td>UWT [LA(12) wavelet]</td>
<td>-10.4</td>
<td>-4.2</td>
<td>2.2</td>
<td>5.3</td>
<td>-7.0</td>
</tr>
<tr>
<td></td>
<td>(19.3)</td>
<td>(4.2)</td>
<td>(3.1)</td>
<td>(7.1)</td>
<td>(16.6)</td>
</tr>
<tr>
<td>UWT [BIOR(6,2) wavelet]</td>
<td>-4.8</td>
<td>-2.0</td>
<td>0.5</td>
<td>3.6</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>(28.3)</td>
<td>(4.0)</td>
<td>(4.1)</td>
<td>(6.7)</td>
<td>(12.0)</td>
</tr>
<tr>
<td>UWT [BIOR(6,2)] + derivatives</td>
<td>-5.6</td>
<td>-1.4</td>
<td>1.0</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>(22.7)</td>
<td>(3.9)</td>
<td>(2.7)</td>
<td>(6.7)</td>
<td>(9.7)</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A typical human ECG waveform together with the standard ECG intervals.</td>
</tr>
<tr>
<td>2</td>
<td>Hidden Markov model architectures for ECG interval analysis.</td>
</tr>
<tr>
<td>3</td>
<td>The three different wavelets considered in this paper (C = Coiflet, BIOR = Biorthogonal, LA = Least Asymmetric).</td>
</tr>
<tr>
<td>4</td>
<td>The impact of <em>derivative features</em> on the accuracy of the T wave offset segmentation. The upper plots show the T wave region of an ECG, together with the T wave offset point (blue dash-dot line) determined by a.) an HMM with the UWT encoding only (upper left plot), and b.) an HMM with the UWT + derivative features encoding (upper right plot). In each case the T wave offset point determined by an expert analyst is also shown (red dashed line). The corresponding lower plots show, for each HMM, the difference between the log observation probabilities for the JT interval and Baseline states evaluated over the course of the T wave.</td>
</tr>
<tr>
<td>5</td>
<td>A comparison of the duration distribution curves for the JT interval, based on i.) a gamma density fitted to the JT interval durations derived from the expert analyst annotations (dashed line), and ii.) the geometric duration distribution for the corresponding HMM state (solid line).</td>
</tr>
<tr>
<td>6</td>
<td>An example of a <em>double-beat</em> segmentation by a hidden Markov model. The first beat segmentation inferred by the model (in blue) correctly identifies the onset of the P wave, but incorrectly “locates” a QRS complex (of duration one sample) and the T wave offset within the PR segment. The second beat segmentation inferred by the model (in green) then correctly identifies the boundaries of the QRS complex and the offset of the T wave in the ECG signal.</td>
</tr>
<tr>
<td>7</td>
<td>Graphical illustration of incorporating a duration constraint into an HMM (the dashed box indicates that the observation densities for all the states in the chain are <em>tied</em>).</td>
</tr>
<tr>
<td>8</td>
<td>A scatter plot of the joint log likelihood values against the beat duration for the annotated ECG beats from lead V2 of the training set.</td>
</tr>
<tr>
<td>9</td>
<td>Schematic illustration of the model training and testing procedure.</td>
</tr>
<tr>
<td>10</td>
<td>Cumulative distribution of the HMM confidence measures for the manually annotated ECG beats from lead II of the test set.</td>
</tr>
<tr>
<td>11</td>
<td>A selection of the <em>low</em> confidence ECGs (i.e. confidence &lt; 0.7) from the lead II test set.</td>
</tr>
<tr>
<td>12</td>
<td>A selection of the <em>high</em> confidence ECGs (i.e. confidence ≥ 0.7) from the lead II test set.</td>
</tr>
<tr>
<td>13</td>
<td>Bland-Altman plots of agreement between automated (HMM - left plot, ECGPUWAVE - right plot) and manual QT interval measurements for lead V2 of the test set. The solid horizontal line indicates the mean QT difference, and the dashed lines indicate 2 x standard deviation error bars.</td>
</tr>
</tbody>
</table>
FIGURE 1. A typical human ECG waveform together with the standard ECG intervals.
(a) HMM for PR and QT interval segmentation

(b) HMM for QT interval segmentation

(c) HMM for full ECG interval segmentation

FIGURE 2. Hidden Markov model architectures for ECG interval analysis.
FIGURE 3. The three different wavelets considered in this paper (C = Coiflet, BIOR = Biorthogonal, LA = Least Asymmetric).
FIGURE 4. The impact of derivative features on the accuracy of the T wave offset segmentation. The upper plots show the T wave region of an ECG, together with the T wave offset point (blue dash-dot line) determined by a.) an HMM with the UWT encoding only (upper left plot), and b.) an HMM with the UWT + derivative features encoding (upper right plot). In each case the T wave offset point determined by an expert analyst is also shown (red dashed line). The corresponding lower plots show, for each HMM, the difference between the log observation probabilities for the JT interval and Baseline states evaluated over the course of the T wave.
FIGURE 5. A comparison of the duration distribution curves for the JT interval, based on i.) a gamma density fitted to the JT interval durations derived from the expert analyst annotations (dashed line), and ii.) the geometric duration distribution for the corresponding HMM state (solid line).
FIGURE 6. An example of a double-beat segmentation by a hidden Markov model. The first beat segmentation inferred by the model (in blue) correctly identifies the onset of the P wave, but incorrectly “locates” a QRS complex (of duration one sample) and the T wave offset within the PR segment. The second beat segmentation inferred by the model (in green) then correctly identifies the boundaries of the QRS complex and the offset of the T wave in the ECG signal.
FIGURE 7. Graphical illustration of incorporating a duration constraint into an HMM (the dashed box indicates that the observation densities for all the states in the chain are tied).
FIGURE 8. A scatter plot of the joint log likelihood values against the beat duration for the annotated ECG beats from lead V2 of the training set.
Encode training set ECGs using UWT + derivative features

Incorporate duration constraints into HMM architecture

Learn HMM parameters \( \{A, \pi, b\} \) using supervised learning estimators

Compute HMM log likelihood scores for annotated ECG beats

Learn HMM confidence model by fitting linear regression to log likelihoods

Encode test set ECGs using UWT + derivative features

Segment 10-second ECG signals using HMM + Viterbi algorithm

Compute HMM confidence measure for each segmented ECG beat

Extract HMM segmentations for ECG beats which were manually annotated

Compute segmentation errors for high confidence ECG beats only

FIGURE 9. Schematic illustration of the model training and testing procedure.
FIGURE 10. Cumulative distribution of the HMM confidence measures for the manually annotated ECG beats from lead II of the test set.
FIGURE 11. A selection of the low confidence ECGs (i.e. confidence < 0.7) from the lead II test set.
FIGURE 12. A selection of the *high* confidence ECGs (i.e. confidence ≥ 0.7) from the lead II test set.
FIGURE 13. Bland-Altman plots of agreement between automated (HMM - left plot, ECGPUWAVE - right plot) and manual QT interval measurements for lead V2 of the test set. The solid horizontal line indicates the mean QT difference, and the dashed lines indicate 2 x standard deviation error bars.