Model-Based Determination of QT Intervals

GD Clifford, MC Villarroel

[1]Harvard-MIT Division of Health Sciences & Technology, 45 Carleton St., Cambridge, MA, USA

Abstract

A method is presented to determine the QT interval by fitting a nonlinear artificial ECG model to segmented regions of a human ECG. The model consists of a set of temporally Gaussian functions with different widths and heights. These parameters are fitted to a given ECG (segmented around the QRS complex to include the P and T wave) using a nonlinear least squares optimization routine. The Q onset and T offset can be determined precisely (in a statistical sense) from the parameters of the Gaussian. Since the fitted waveform contains no noise, the differential is smooth. Waveform boundaries can also be determined by searching for the minimum of a differential. Furthermore, the residual error provides an estimate of the confidence in the fit, and hence, the derived QT interval. Using the human expert-annotated PhysioNet QT database, various QT interval estimation schemes were compared using the model-fitted ECG to find an optimal marker of the QT interval. It was found that humans are inconsistent and almost always under-estimate the T-offset (if defined to be the end of any repolarization). This is probably due to the truncation of any human estimation when the T wave tail is consumed by noise. We therefore propose an alternative QT end point. Finally, an entry based on the most favourable technique was submitted in the PhysioNet / Computers in Cardiology Challenge 2006; QT Interval Measurement, which is intended to produce a comparison of several automatic and human annotators on the Physikalisch-Technische Bundesanstalt diagnostic ECG database. A follow-up paper to address differences between those generated by our method and the consensus of the other entries will be submitted shortly.

1. Introduction

The Expert Working Group (Efficacy) of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) has developed and has given its final ('step 4') endorsement in May, 2005 to a set of guidelines (ICH E14) for clinical evaluation of QT/QTc interval prolongation and pro-arrhythmic potential for non-antiarrhythmic drugs [1]. A major part of the motivation for the seventh annual *PhysioNet/Computers in Cardiology Challenge* [2] is to provide well-characterized data that might support modifications of the ICH E14 recommendations with respect to fully-automated methods. The Challenge is attempts to answer the important clinical question: *Can the QT interval be measured by fully automated methods with an accuracy acceptable for clinical evaluations?*

For this Challenge each contestant is required to submit an entry for a Q onset and accompanying T offset for one 'representative' beat in lead II of each of the 549 recordings in the Physikalisch-Technische Bundesanstalt Diagnostic ECG Database (PTBDB). There are essentially two categories; manual and automated. The score for each entry is calculated from the mean square difference between these intervals and the median of the manual (human) intervals, weighted by the number of records attempted. Further details can be found on PhysioNet [2] and in [3].

2. A model-based approach

Our approach is based upon the concept of fitting a realistic model to the ECG and extracting parameters from the model to determine waveform onsets and offsets.

2.1. Preprocessing

Before fitting a model to the ECG, it is necessary to perform a series of noise reduction and first-order segmentation steps, from which an adaptive procedure is used to *fine-tune* the segmentation, and allow precise feature localization. The preprocessing steps are:

- **Mains noise filtering**: The PTBDB exhibits significant mains noise interference at 50 and 100 Hz. Therefore, two FIR notch filters centered at $f_{notch} = 50$ Hz and $f_{notch} = 100$ Hz were implemented using a forward-backward zero-phase filtering procedure.
- **QRS detection**: A standard QRS detector was used [4] that locates a maximum in a smoothed, differentiated, squared and integrated ECG. A quality metric was also defined to reject QRS complexes that are too long or short.
- **Baseline wander removal**: Non-cardiac related low frequency baseline changes are generally removed by con-

structing a cubic spline interpolated signal using each beat's isoelectric point as a node for the spline procedure. Here we propose a new and robust method for isoelectric point location. Each beat is segmented in a window between $R(t_i) - \alpha(R(t_i) - R(t_{i-1})$ and $R(t_{i+1}) - \alpha(R(t_{i+1}) - R(t_i)$ with $\alpha = 0.4$. The data is then 5 point median filtered, quantized into $f_s/20$ levels and then the mode of the sequence is calculated. The latest value the segment with the value of this mode is taken to be the isoelectric point. A cubic spline is then constructed using these points at a rate of $f_{cubic} = 5$ Hz. The data is then antialias-resampled from f_{cubic} Hz to f_s Hz. For the PTBDB $f_s = 1000$ Hz. The resultant signal is then removed from the recorded ECG.

• **Template construction and artefact rejection**: To construct a general template of a lead-specific beat, the first 40 beats in the given lead (II) are segmented as described above. The initial template, $\overline{s}$, is simply the mean of this set of beats. Abnormal and noisy beats are removed from the template if the correlation coefficient, ρ, between a given beat and $\overline{z}$ is less than 0.9. The new template is then the average of the remaining beats with $\rho \geq 0.9$. This procedure was first described in [5]. If more than 10 beats are removed, another 20 beats are added and this procedure is repeated until at least 30 beats are in the template.

2.2. Dynamic Gaussian model for waveform parameterization

The general model fitting procedure is described in [6] and [7]. Briefly, each of the symmetric turning points (Q, R, and S) in the ECG are characterised by one Gaussian and the asymmetric turning points (P and Q) are characterised by two Gaussians (to account for bi-phasic P waves and the asymmetric nature of the T wave at low to medium heart rates). Therefore, 21 parameters (7 Gaussians) can be used to accurately describe the ECG. From [6];

$$z = \sum_{i \in \{P_1,P_2,Q,R,S,T_1,T_2\}} (a_i/2b_i)\, e^{\frac{\Delta t_i^2}{2b_i^2}} + z_i t \qquad (1)$$

where P_1 and P_2 are the two Gaussians describing asymmetric/bi-phasic P wave, and T_1 and T_2 are the two Gaussians used to describe the asymmetric T wave. Functions other than Gaussians can be used, such as the Gumble or log-Normal functions, to reduce the number of parameters from 21, but flexibility and interpretation can be lost.

2.3. Fitting the model to features

To fit the model to an observation, a nonlinear gradient descent is performed to optimize the parameters, a_i, b_i and t_i to produce a sum of Gaussians that best fit the signal. The signal model (Eq. 1) is fitted to an observation $s(t)$, by minimizing the squared error ε_r, between s and the model output, z such that $\varepsilon_r = \min_{a_i,b_i,\theta_i} \|s(t) - z(t)\|_2^2$. This is achieved using a nonlinear gradient descent [6].

2.3.1. Model intialisation

The gradient descent can be accelerated by estimating the model parameters. The only parameters of the initial guess that require accurate estimation are the t_i. These can be simply estimated from the initial template, firstly through a simple peak and trough detection algorithm, limited by heart-rate adjusted refactory periods, and then through an initial fit to the average template. For a lead II configuration, t_Q and t_S are taken to be the minima in a 100 ms window either side of the R peak. t_P is then taken to be time at which s is maximal in the section before t_Q. t_T is taken to be time at which s is maximal in the section following t_S. For the asymmetric waves (P and T), their t_i's are taken to be ± 40 ms either side of the t_P and t_T. The a_i and b_i are initialized with an arbitrary small value of $10+5\vartheta$, where ϑ is a uniform distribution on the interval [0,...,1].

2.4. QT interval determination from model parameters

We considered two methods for determining wave boundaries. The first method is an intuitive interpretation of the end of repolarization, where we look for the gradient of the ECG to drop to zero. The second is a probabilistic interpretation of the end point, given as a fixed number of standard deviations from the central peak of a sum of Gaussians.

2.4.1. Absolute zero-gradient criteria

In practice, the digital ECG has a finite sampling frequency and resolution, so we must look for a point where the absolute gradient drops below some small constant, ϵ, that depends on the sampling frequency and resolution. The search region is limited to the peak of the final Gaussian (second portion of the T wave) and the first Gaussian (first portion of the P wave) of the next beat. That is, where

$$|\frac{dz}{dt}| < \epsilon, \;\; t_i^{T_2} < t < t_{i+1}^{P_1}. \qquad (2)$$

For a resolution of 16 bits and $f_s = 1$ kHz, $\epsilon = 10^{-3}$.

2.4.2. Probabilistic criteria

The second method we considered depends on locating the onset and offset using a statistical technique. Specifically, since the constituent waves within the ECG are represented by Gaussians, we can calculate the onsets and off-

sets as a set number of standard deviations, σ, away from the mean μ, the central location of the Gaussian. One problem however, is that each wave (and in particular the T wave) must be represented by more than one Gaussian.

For two Gaussians X_1 with probability p_1 and X_2 has probability p_2, with $p_1 + p_2 = 1$, can be considered to be linear combinations such that $X = p_1X_1 + p_2X_2$. The mean is then $\mu = p_1\mu_1 + p_2\mu_2$. For the T wave, $\mu_1 = t_{T_1}$, $\mu_2 = t_{T_2}$, $\sigma_1 = b_{T_1}$ and $\sigma_2 = b_{T_2}$. The variance is given by $p_1^2\sigma_1^2 + p_2^2\sigma_2^2$ and so the standard deviation is $\sigma = (A_1^2b_1^2 + B_2^2b_2^2)^{\frac{1}{2}}$, where A_1 and B_1 are normalisation constants given by $A_1 = \frac{p_1}{p_1+p_2}$ and $A_2 = \frac{p_2}{p_1+p_2}$). The relative probabilities of the Gaussians, p_1 and p_2, that form the T wave are simply the areas under each Gaussian and can be derived from an analytical integration of two exponentials to give $p_1 = a_{T_1}b_{T_1}\sqrt{2\pi}$ and $p_2 = a_{T_2}b_{T_2}\sqrt{2\pi}$

Therefore the end point of the T wave is defined as

$$\begin{aligned} T_{\text{offset}} &= \mu_T + \aleph\sigma_T \\ \mu_T &= (a_{T_1}b_{T_1}\sqrt{2\pi})t_{T_1} + (a_{T_2}b_{T_2}\sqrt{2\pi})t_{T_2} \\ \sigma_T &= (A_1^2\sigma_{T_1}^2 + A_2^2\sigma_{T_2}^2)^{\frac{1}{2}}. \end{aligned} \quad (3)$$

The start of the Q wave is defined for one Gaussian in a simple way; $Q_{\text{onset}} = \mu_Q - \aleph\sigma_Q$. We chose $\aleph = 2$ as a convenient, although non-optimal definition for the onset/offset of the Q and T waves.

2.5. Beat selection

The competition rules suggest that the first 'representative' beat in lead II should be selected for analysis. However, the problems with selecting such a beat are numerous:

- Rapid nonstationary changes; beats following abnormal beats will have non-representative QT intervals.
- Slow nonstationary changes; beats later in the recording may have significantly different QT intervals.
- Noise on beats; rejecting beats with the template cross correlation $\rho < 0.9$ removes unrepresentative beats.
- Inter-lead differences; QT dispersion across leads means that each lead will provide a different length QT interval.

These criteria preclude using the first beat in each file. To obtain the first non-noisy beat we select the second beat with $\rho \geq 0.9$ with the 30-beat template. This precludes any beat that follows and abnormal beat, and any noisy or abnormally shortened/elongated beat. Of our three entries, our final entry also used information from all the standard 12 leads, as described below. (The VCG leads were not used, since they usually give longer QT intervals than the standard 12 leads.)

2.5.1. Robust multi-lead comparisons & quality metrics

QT dispersion (QTd) is defined as the difference between the longest QT and the shortest QT interval measured on a standard 12 lead configuration [8]. Although great controversy exists surrounding the phenomenon of QTd [9], it is clear that it inter-lead differences do exist. The reason for QT interval differences between leads is thought to be due to the differing spatial orientation of the ECG lead vectors and their differing sensitivity, which modifies T wave amplitudes differently on each lead [9]. It is sometimes considered to be both an observational error in measurement, and a real effect due to inhomogeneities in the line of conduction between the heart and the electrodes.

A realistic QT interval should be assessed as either the longest (for LQTS) or the shortest (for SQTS) calculated from all 12 standard leads. However, this approach in automated ECG analysis is sensitive to noise. In particular, choosing just one lead (such as lead II) is unlikely to give accurate QT interval measurements for all morphologies. Therefore, by taking the median QT interval from all 12 leads, we provide a more accurate (and more human-like) analysis of the QT interval.

Furthermore, the longest QT interval is determined by taking the median of the QT intervals longer than the overall median. The shortest QT interval is similarly determined by taking the median of all the QT intervals shorter than the overall median QT interval. This *double-median* QT interval determination method leads to a more robust automated QTd measurement.

The quality of the fit is defined to be one minus the normalised residual error in the model-fit per sample. However, since the QT intervals for this data are currently unknown, calibration of this quality index on this data set was not possible. Therefore, all 549 records in the PTBDB were annotated regardless of quality.

3. Competition entries

Three different sets of QT onset/offset pairs were entered into the competition:

1. Where $|dz/dt| \approx 0$.
2. At the 2σ point on the T wave of a lead II beat.
3. Median of all 12 clinical leads determined by (2).

A deliberate choice was made to enter these three methods in this order, since we expected that each method would approximate the human decision process. $\aleph = 2$ standard deviations was chosen to define the end of the T wave since this has a statistically justifiable interpretation. No effort was made to incrementally optimize this parameter, $\aleph$, in order to achieve closer results to the median of the manual scores, since we felt this detracted from the philosophical point of our approach. That is, we wished to determine how far a series of well-founded, mathematically justifiable, locations are from the consensus human standard.

4. Results and discussion

Our first entry, based upon the zero-gradient criterion, resulted in a score of 163.39 ms with a positive bias of 95.32. This poor score reflected the fact that humans truncate the estimate of T wave offset at an early point due to the difficulty in locating the end of the exponential decay in the noise floor. Therefore, the zero-gradient criterion, although philosophically the most justifiable, poorly mimics human annotation.

Our second entry, based upon the 2σ point of the combined T wave Gaussian representation of a single representative beat in lead II, resulted in a much improved score of 38.51 ms (31.45 after correcting for the +23.87 bias). The residual error in the model fit indicated that many of the beats used to identify the end of the T wave were of low quality and therefore we entered the median of the model fit across all 12 clinical leads.

This resulted in a third and final score of 28.23 ms, or 30.79 ms with a bias of $+12.28$ ms. It is interesting to compare this to what is considered a significant error in the QT interval. Although there is no clear consensus, a figure of about 10 to 30 ms could be considered a reasonable error, since Ivaylo *et al.* [10] found that in many cases, experts often disagreed by this amount. It should be noted that the Challenge score uses a root mean square error metric and is therefore highly sensitive to errors in just one or two of the 549 subjects. It is therefore difficult to infer significance in the errors of these results. The significant positive bias suggests that the $\aleph\sigma$ criterion could be adjusted to $\aleph < 2$ in order to improve our score. Since $\aleph$ is a nonlinear, patient-specific parameter that leads to a QT interval that depends on the morphology of the T wave, the adjustment to mimic human annotation intervals may best be done by altering the estimate of the T offset as a function of T wave height and asymmetry, as well as width. However, further blind adjustment of this threshold in an incremental fashion for the competition is simply an attempt to over-tune on the data set to win the competition and does not aid the real development of the algorithm.

Rather, we wish to systematically explore which facets of the algorithm require further improvement. Moreover, the actual bias in the QT interval is not important, as long as one can adjust for the bias, or the biased measure is as consistent (or more consistent) than other ad-hoc methods for delineating the Q onset and T offset.

5. Conclusions and further work

In this work we have shown that a model-based approach can lead to a robust and accurate method for determining the QT interval in a large range of patients. Our approach is useful on one or multiple leads, and can be tuned to incorporate individual biases in human preferences of T wave end points (by adjusting the number of standard deviations that defines the T wave end).

We have also demonstrated that humans appear to consistently under-estimate reasonable (statistical) definitions of the 'true' T wave end-point. Future work will focus on exploring the true T wave end point using realistic conduction models and improving the stability of estimates for the repolarization period and the stability of QT proxies.

Of course, the final word in the selection of a QT interval estimation process will be its clinical utility.

Acknowledgments

The authors would like to acknowledge support from the NIH/NBIB grant R01 EB001659.

References

[1] Guidance for industry E14. Clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. Technical report, US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, 2004.

[2] www.physionet.org/challenge/2006/.

[3] Moody G, Koch H, Steinhoff U. The PhysioNet / Computers in Cardiology Challenge 2006: QT Interval Measurement. Computers in Cardiology 2006;33. In Press.

[4] Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985;32(3):220–236.

[5] Clifford GD, Tarassenko L, Townsend N. Fusing conventional ECG QRS detection algorithms with an auto-associative neural network for the detection of ectopic beats. In 5^{th} International Conference on Signal Processing. IFIP, Beijing, China: World Computer Congress, August 2000; 1623–1628.

[6] Clifford GD, Shoeb A, McSharry PE, Janz BA. Model-based filtering, compression and classification of the ECG. International Journal of Bioelectromagnetism May 2005; 7(1):158–161.

[7] Clifford G. A novel framework for signal representation and source separation. Journal of Biological Systems June 2006;14(2):169–183.

[8] Mirvis D. Spatial variation of QT intervals in normal persons and patients with acute myocardial infarction. J of Am Coll of Cardiol 1985;5:625–631.

[9] Rautaharju PM. QT and Dispersion of Ventricular Repolarization: The Greatest Fallacy in Electrocardiography in the 1990s. Circulation 1999;99(18):2476c–2479.

[10] Christov I, Dotsinsky I, Simova I, Prokopova R, Trendafilova E, Naydenov S. Dataset of manually measured QT intervals in the electrocardiogram. BioMedical Engineering OnLine 2006;5(1):31.

Address for correspondence:

Gari D. Clifford (http://alum.mit.edu/www/gari/)
HST / 45 Carleton St., / Cambridge MA 02139, / USA