Principal Component Regression Approach for QT Variability Estimation

P. A. Karjalainen* and M. P. Tarvainen
Department of Applied Physics
University of Kuopio
P.O.Box 1627, FIN-70211 Kuopio, Finland
*Email: Pasi.Karjalainen@uku.fi

T. Laitinen
Department of Clinical Physiology and Nuclear Medicine
Kuopio University Hospital
P.O.Box 1777, FIN-70211 Kuopio, Finland

***Abstract*— New algorithm for quantifying the variation in the QT interval of ECG recording is presented. The algorithm is based on the Principal Component Regression where the eigenvectors of the data correlation matrix are calculated. The eigenvectors are then used for calculation of the principal components and one of them is selected to represent the information about T wave variation. The algorithm is tested using high speed ECG recording.**

I. Introduction

QT interval in an electrocardiogram presents a measure of ventricular repolarization. It is known that it is controlled by neural regulatory system same way as heart rate (HR). Furthermore, it has been found that there is variation in the duration of the QT intervals same way than the heart rate varies during time [1], [2]. While this kind of variation is normal part of the regulation of cardiac system it has been argued [3] that abnormal variation in the repolarization duration could be a marker for a group of severe cardiac diseases such as ventricular arrhythmias. It has also been argued that the QT variability could yield such additional information which can not be observed from HR variability [4].

In this paper we propose a robust method for quantifying the variation in the QT interval of ECG recording. The method is based on the so called Principal Component Regression (PCR) and it does not necessitate the detection of T wave. The variation of the interval is seen to be proportional to the shape of an eigenvector of the autocorrelation matrix of an ensemble of dataset. It is further shown how these eigenvectors can be used in the estimation of each individual QT interval duration. The method is tested using a sample of single channel high speed ECG recording.

II. Methods

The estimation of QT interval is not always a simple task. T wave is a smooth waveform that is very hard to detect in conditions where the signal-to-noise ratio is not high enough. Several artefacts also affect the reliability of the detection remarkably. In our approach we do not detect each individual T wave separately. Our approach is to extract the information of the variation using an ensemble of cardiac cycles.

A. Principal component regression

In the principal component regression, the vector containing the measured signal is presented as a weighted sum of orthogonal basis vectors. The central idea in PCR is to reduce the dimensionality of the data set, while retaining as much as possible of the variance in the original data. The dimensionality of measurements can be estimated by the number of basis vectors needed to estimate measurements in a certain accuracy.

B. The algorithm

The ECG measurement is first divided into adequate epochs such that each epoch includes one QRS complex and the following T wave. The epochs are fixed according to the fiducial points of the QRS complexes. Let us denote the t'th such ECG epoch with a length N column vector

$$z_t = \begin{pmatrix} z_t(1) \\ \vdots \\ z_t(N) \end{pmatrix}. \tag{1}$$

As an observation model we use the so-called additive noise model

$$z_t = s_t + e_t \tag{2}$$

where s_t is the noiseless ECG signal corresponding to t'th epoch and e_t is additive measurement noise. The measurement noise is assumed to be a stationary zero mean process. Note that both s_t and e_t are random vectors. If we have M such epochs, the response signals s_t will span a vector space $\mathcal{S}$, which will be at most of $\min\{M, N\}$ dimensions. In the case that the ECG epochs are rather similar, the dimension of the vector space $\mathcal{S}$ will be $K \leq \min\{M, N\}$ and epochs can be well approximated with some lower dimensional subspace of $\mathcal{S}$. Thus, each epoch can be expressed as linear combination

$$z_t = H_{\mathcal{S}}\theta_t + e_t \tag{3}$$

where $H_{\mathcal{S}} = (\psi_1, \psi_2, \ldots, \psi_K)$ is a $N \times K$ matrix of basis vectors which span the K dimensional subspace of $\mathcal{S}$ and $\theta_t \in \mathbb{R}^K$ is a column vector of weights related to t'th epoch. By

defining a $N \times M$ measurement matrix $z = (z_1, z_2, \ldots, z_M)$ the observation model (3) can be written in the form

$$z = H_{\mathcal{S}}\theta + e \tag{4}$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_M)$ is a $K \times M$ and $e = (e_1, e_2, \ldots, e_M)$ is a $N \times M$ matrix. The critical point in the use of model (4) is the selection of basis vectors ψ_k. A variety of ways to select these basis vectors exist. Here a special case, i.e. principal component regression [5], [6], is considered. In PCR the basis vectors are selected to be the eigenvectors v_k of either the data covariance or correlation matrix. Here, the correlation matrix is utilized. The eigenvectors of the correlation matrix are orthonormal and, therefore, the ordinary least squares solution for the parameters θ becomes

$$\hat{\theta}_{\mathrm{PC}} = H_{\mathcal{S}}^T z. \tag{5}$$

Quantitatively the first basis vector is the best mean square fit of a single waveform to the entire set of epochs. Thus, the first eigenvector is similar to the mean of the epochs and the corresponding parameter estimates or principal components (PCs) $\hat{\theta}_t(1)$ reveal the contribution of the first eigenvector to each epoch ($t = 1, \ldots, M$). The second and third eigenvectors, on the other hand, cover the time variations in the QRS complex and T wave times. Thus, either the second or third eigenvector is expected to resemble the derivative of the T wave and the corresponding PCs to reflect the variability of the time difference between the QRS complex and the T wave. These PCs are taken as estimates of QT variability.

III. Results

The proposed method was tested with a high-speed ECG recording (sampling frequency 20 kHz) measured from a healthy young male in relaxed conditions. As a recording device a Compumedics Neuroscan measurement system was used. In order to remove measurement noise, the ECG recording was band-pass filtered (pass-band 1–30 Hz). The aim of the filtering was also to enable as unambiguous detection of the R and T wave maximums as possible. The RT intervals were then extracted for each consecutive beat as the time difference between the T and R wave maximums. The obtained values are here considered as the "true" RT intervals and the whole interval series is presented on bottom of Fig. 1.This series reflects the variability in the ventricular depolarization duration.

The ECG recording was then downsampled to 500 Hz with antialias filtering. RT segments were then aligned using the R peak as a marker (see the topmost axes of Fig. 1). The time window started 0.1 s before the R peak and ended 0.4 s after that. The length of each epoch was then $N = 250$ points. Each epoch was then normalized to unit norm. The normalization diminishes the influence of ECG amplitude level changes on the PCR. Thus, PCR becomes more sensitive to the waveshapes of the epochs which is desirable here.

The first three eigenvectors of the correlation matrix of the ensemble are shown in Fig. 2. In Fig. 3, the first and the third

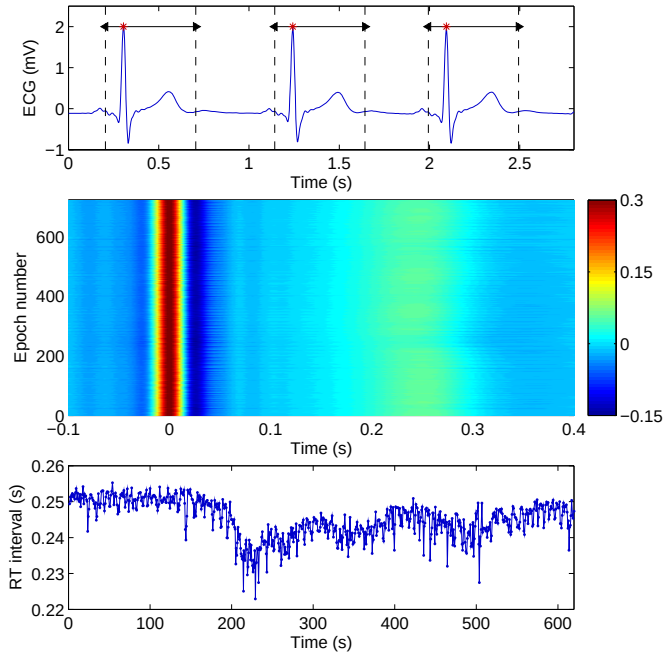

Fig. 1. ECG epochs obtained using windowing (top). The extracted normalized segments as an image (middle). RT intervals extracted from the original high speed recording (bottom).

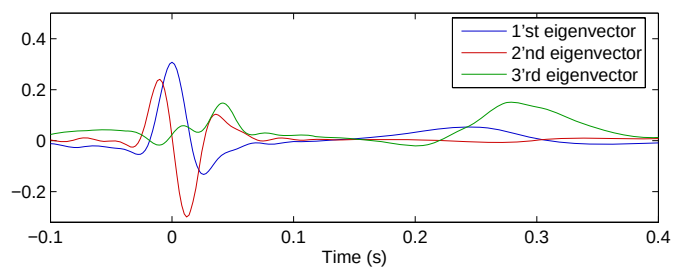

Fig. 2. Eigenvectors of the data correlation matrix corresponding the three largest eigenvalues $\lambda_1 = 0.9938$, $\lambda_2 = 0.0028$, and $\lambda_3 = 0.0019$.

eigenvectors are shown in different scale. The first eigenvector represents the mean of the ensemble. The peak value of the first eigenvector is thus close to the mean occurrence time of the T waves. It is easy to see that in the superposition of first and third eigenvector the peak value is moved to the right if the weight of the third eigenvector is positive. Correspondingly, using a negative weight moves the peak to the left. These two eigenvectors are thus capable to model the variation in T wave occurrence time. The shape and width of the waveform in the 3'rd eigenvector itself would serve as measures of variation of the QT interval.

The eigenvectors can further be used as basis vectors in linear fitting of the measurements using eq. (5). The variation of the duration of each individual QT interval should then be seen in the variation of the parameter corresponding to the 3'rd eigenvector. We can observe this clearly in bottom of Fig. 4. It can be clearly seen that the variation in the third principal components is similar to the true QT-intervals. The

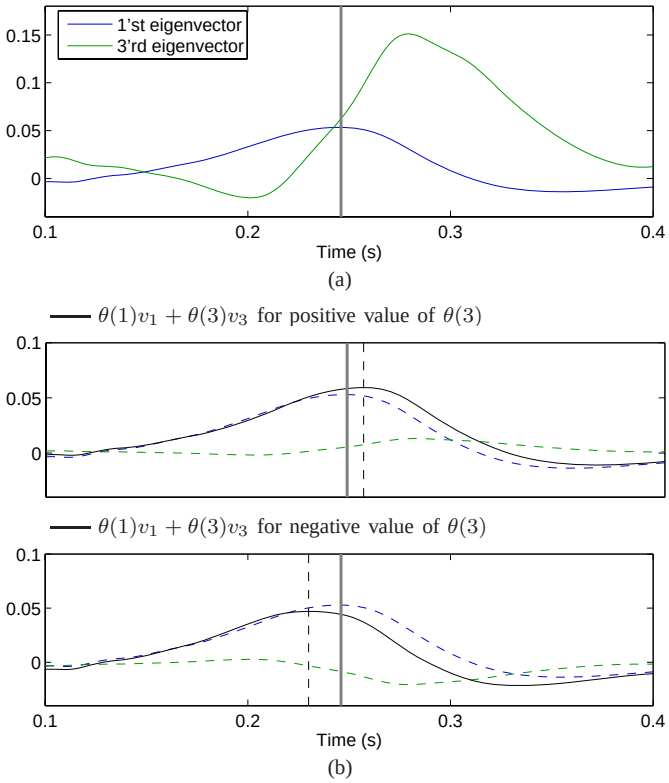

Fig. 3. The superposition of the first and third eigenvectors with different weights. (a) The first and third eigenvectors and (b) the superposition of these eigenvectors for weights $\theta(1) = 0.993$ and $\theta(3) = 0.087$ (top) and for weights $\theta(1) = 0.990$ and $\theta(3) = -0.136$ (bottom). The peak positions of the resulting superpositions are indicated with dashed vertical lines and the peak position of the first eigenvector itself with vertical gray line.

actual QT interval times (or RT interval in this case) can be further estimated if necessary.

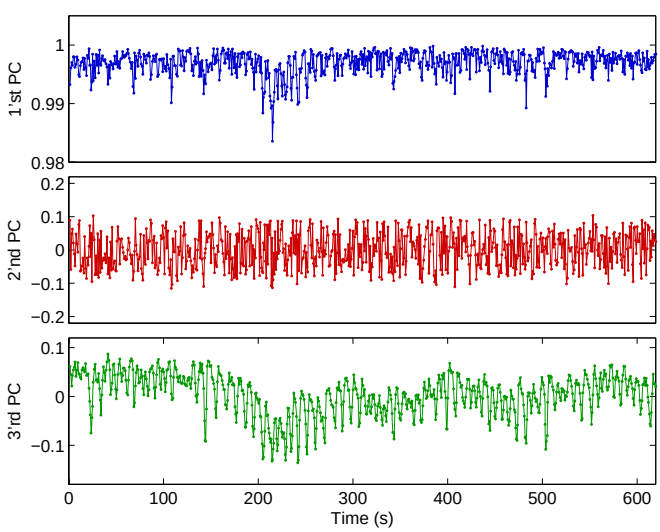

Fig. 4. Principal components corresponding the first, second, and third eigenvector (top, middle, and bottom, respectively).

IV. Discussion

We have proposed a new algorithm for quantifying QT interval variability. We have shown that using a principal component regression based approach there is no need to detect the T wave. This makes the algorithm more reliable e.g. in the situations where the signal to noise ratio is not very high. The algorithm can further be developed to take into account time variation in the repolarization duration. The eigenvectors may be used as basis vectors in some appropriate adaptive algorithm.

References

[1] M. Merri, J. Benhorin, M. Alberti, E. Locati, and A. Moss, "Electrocardiographic quantitation of ventricular repolarization," vol. 80, no. 5, pp. 1301–1308, 1989.

[2] M. Merri, M. Alberti, and A. Moss, "Dynamic analysis of ventricular repolarization duration from 24-hour Holter recordings," *IEEE Trans Biomed Eng*, vol. 40, pp. 1219–1225, December 1993.

[3] R. Berger, "Qt variability," *J Electrocardiol*, vol. 36 Suppl, pp. 83–87, 2003.

[4] R. Negoescu and S. Dinca-Panaitescu, "Mental stress enhances the sympathetic fraction of QT variability in an rr-independent way," *Integr Physiol Beh Sci*, vol. 32, no. 3, pp. 220–227, 1997.

[5] P. Brown, *Measurement, Regression, and Calibration*. Oxford Science Publications, 1993.

[6] I. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.