Computational Aspects of Models for Minimizing the Effects of Ectopic Beats on Heart Rate Variability

Martin Frank[1,2], Martin Bachler[1,2*], Siegfried Wassertheurer[2], Christopher Mayer[2]

[1] Institute of Analysis and Scientific Computing, Vienna University of Technology, Wiedner Haupstraße 8-10, 1040 Vienna, Austria; * martin.bachler@tuwien.ac.at

[2] AIT Austrian Institute of Technology GmbH, Health & Environment Department, Biomedical Systems, Donau-City-Straße 1, 1220 Vienna, Austria

Abstract. Several studies have highlighted the need of reliable markers to determine cardiac health. One of the most promising markers is the heart rate variability (HRV). However, artifacts, such as ectopic beats, have to be corrected before HRV parameters can be reliably used. This paper reviews several recently established models that are used for ectopic beat correction, with an emphasis on execution time and memory requirements. In general, physiologic models require far more execution time, while memory demand is compareable to simpler algorithms. Moreover, physiologic models result in better corrections, compared to simpler methods. Therefore, we conclude that physiologic models are prefferable since the execution time is low enough to use all models in an online approach.

Introduction

The latest reports of the "American Heart Association" and the "European Heart Network" highlight that the leading cause of death are heart diseases in the United States [1], and diseases of the heart and circulatory system in Europe [2]. Since early action is the key of surviving heart failure, reliable markers have to be established [1]. The so called heart rate variability (HRV) is one of the best suited markers for the relationship between the autonomic nervous system and cardiovascular mortality [3]. HRV parameters are immensely influenced by the presence of ectopic beats, which have

to be corrected before performing HRV analysis [3][7].

In the last few years several different processing and correction methods have emerged. However, no comprehensive comparative study is available yet. The approaches may be categorized in (1) interpolation and removal [8]-[10], (2) filtering [6],[11][14] and (3) model-based approaches [4][15]-19].

The goal of this work is to provide a comparative review of different model-based approaches for the correction of ectopic beats. Thus, different test cases are defined to perform an objective comparison using artifically corrupted error-free RR interval time series. In this work we focus on the computational aspects, i.e. the execution time and the required memory, of the models under investigation. A comprehensive evaluation of the medical parameters and inplications of ectopic beat correction on HRV measures can be found elsewhere.

1 Methods & Models

In total, seven model-based approaches were compared. All these algorithms were entirely self-implemented and tested in Matlab version 2007b (The Mathworks Inc., Natick, US) based on literature.

1.1 Ectopic Beat Correction models

The first correction model is called "buffer with combination rules" (BUFFER), a fast online correction algorithm that was introduced by Rand et al. [18]. It uses several combination and/or splitting rules of neighbouring RR intervals in the following pre-defined order, as given in Figure 1.

Correction	Description
Split	Missed heartbeat; divide IBI into two equal intervals
Split 3	Two missed heartbeats; Split IBI into three equal intervals
Combine	False trigger; combine two IBIs into one
Combine 2 / Split 2	Replace two IBIs with their average
Combine 3 / Split 2	Get two new IBIs as average of three
Combine 2 / Split 3	Get three new IBIs as average of two
Combine 3 / Split 3	Replace three IBI values with their average
Uncorrected	Could not apply any rule, but IBI appears faulty

Figure 1: BUFFER algorithm by Rand et al. [18].

Before applying the correction, its usefulness is determined by comparing the RR interval under consideration to statistical parameters of the output buffer containing *n* preceding correct RR intervals (five RR intervals were used, as suggested in the original work). In addtion, the new RR interval must fit to those of the intervals in the input buffer covering *m* unprocessed RR intervals accounting for at least 6s. The buffer is also able to detect ectopic beats on its own.

'Gross positioning of beats' (GP-IIA) was suggested by Mateo et al. as coarse positioning of beats [4]. This method also performs a classification of all beats. By means of Lagrange's interpolation formula the derivative of the instantaneous heart rate is estimated at the kth beat by

$$|\acute{r}_k| = 2\left|\frac{t_{k-1} - 2t_k + t_{k+1}}{(t_{k-1} - t_k)(t_{k-1} - t_{k+1})(t_k - t_{k+1})}\right| < U, \quad (1)$$

where U is a predefined threshold that is calculated by

$$U = 4.3 \cdot std(r_k), \quad (2)$$

and limited by 0.5. Based on six different test cases the beat type is determined. Single ectopic beats are simply shifted to an intermediate position between the previous and following normal sinus beat. False-positives (FPs) are deleted and false-negatives (FNs) are corrected by insertion of an intermediate evenly spaced beat. Consecutive ectopic beats are corrected by insertion of multiple evenly spaced beats.

In contrast to the former mentioned algorithms, the 'integral pulse frequency modulation' (IPFM) model relies on a physiological relationship. This model is able to predict the autonomous nervous system (ANS) activity on the sinoatrial (SA) node by simulating the series of cardiac events as firings of the SA node [20]. The IPFM model integrates the input signal until a beat is generated and is then reset to zero [4]. Hence, the kth beat can be interpreted as the integration of the instantaneous heart rate over the actual RR interval.

The index of the kth beat can be calculated by

$$k = \int_0^{t_k} \frac{1 + m(t)}{T} dt, \quad (3)$$

where t_k denotes the occurrence time of the kth beat, $m(t)$ is the modulating part of the heart rate and T is the mean RR interval length.

The IPFM model with s-parameter (IPFM-S) was introduced by Mateo et al. [4]. According to the IPFM model, the integrator is reset too early if an ectopic beat occurs, resulting in a lower integration value than expected (denoted as s). An indexing function of the beat occurrence times is introduced, which is split into a forward and a backward function. The forward function is based on the normal sinus beats prior to the ectopic beat and the backward function is based on the normal sinus beats afterwards. Both functions are extrapolated to the neighbouring beat until they overlay. The vertical difference of the two indexing functions is the s-parameter:

$$\hat{s} = \frac{1}{t(k_e) - t(k_e - 1 + s)} \cdot \\ \cdot \int_{t(k_e-1+s)}^{t(k_e)} (\hat{x}^f(t) - \hat{x}^b(t))dt, \quad (4)$$

where k_e denotes the index of the ectopic beat, $\hat{x}^f$ is the forward indexing function, and $\hat{x}^b$ is the backward indexing function.

The IPFM model with δ-parameter (IPFM-D) was suggested by Solem et al., since calculation of the s-parameter may be rather time consuming [15]. The δ-parameter corresponds to the time shift of the beat occurrence times followed by an ectopic beat and is related to the IPFM model by

$$\int_0^{t_k} 1 + m(\tau)\, d\tau = kT_0 + \delta. \quad (5)$$

Different δ - parameters may be used, dependent on how many beats prior to the ectopic beat are involved in the calculation (δ1, δ2 and δ3). Solem et al. mentioned that the consideration of more than one prior beat does not enhance the correction ability [15]. Hence, the δ1-parameter was used in this work. Its performance is nearly identical to that of the s-parameter, despite of a reduction in computation time.

The IPFM model with cost function (IPFM-C) was introduced by Brennan et al. [16]. It was designed to only account for single premature ventricular contractions (PVCs).

Basically, the cost function $C(t_e)$ calculates the quadratic deviation from the mean of the impulse height of the integration function in dependence of the beat occurrence time points:

$$C(t_e) = \sum_{S_k \in \psi} (S_k - E[\psi])^2, \qquad (6)$$

where S_k denotes the height of the integration value at each beat occurrence time at the reset point, and can be calculated as follows:

$$S_k = \frac{1}{2\pi f_c} \sum_{j=k-M}^{k+M+1} Si(2\pi f_c(t_{k+1} - t_j)) - Si(2\pi f_c(t_k - t_j)), (7)$$

where Si is the sinc function and f_c is calculated as the reciprocal of the doubled mean RR interval length $\bar{I}$:

$$f_c = \frac{1}{2 \cdot \bar{I}} \qquad (8)$$

Further, $\bar{I}$ is the threshold for the integrate-to-threshold process (the integration function). The mean integration height $E[\psi]$ may be approximated by the following equation:

$$E[\psi] \cong E[S_k] = \frac{1}{2M + 2} \sum_{k=e-M-1}^{e+M} S_k. \qquad (9)$$

Thereby, just the $\pm M$ adjacent impulse heights ψ, with respect to t_e, are considered. The new beat insertion time corresponds to the lowest costs.

'Trend predict correction' (TPC-HT) was developed by Wen et al. as a method that corrects ectopic beats based on trend correlation of the heart timing signal [17]. Each predicted RR interval is composed of two parts, the trend and the turbulence:

$$RRI_{pred} = RRI_{trend} + RRI_{turb}. \qquad (10)$$

The trend is simply calculated as the weighted mean of the n previous NN intervals:

$$RRI_{trend} = \sum_{t=t_{k-n}}^{t_{k-1}} w(t) \cdot RR(t), \qquad (11)$$

where $w(t)$ are time dependent exponential weights as described by Citi et al. [19]. The turbulence can be seen as the slope of the previous NN intervals and is approximated by the following calculation:

$$RRI_{turb} = I[t_k] \cdot E[t_k], \qquad (12)$$

where $I[t_k]$ is just the sign of the turbulence and $E[t_k]$ is the quantity. $I[t_k]$ can be judged by the signs of the slopes at t_{k-1} and t_{k-2}:

$$I[t_k] = \frac{k_1 \cdot k_2 \cdot (k_1 + k_2)}{|k_1 \cdot k_2 \cdot (k_1 + k_2)|}. \qquad (13)$$

The slopes are determined by the following calculation (a tiny value is added to avoid zero, not shown):

$$k_1 = \frac{y(t_{k-1}) - y(t_{k-2})}{y(t_{k-1}) + y(t_{k-2})} \text{ and} \qquad (14)$$

$$k_2 = \frac{y(t_{k-2}) - y(t_{k-3})}{y(t_{k-2}) + y(t_{k-3})}. \qquad (15)$$

The quantity of the turbulence $E[t_k]$ is calculated by consideration of the two slopes k_1 and k_2, RRI_{trend} and the standard deviation of the previous RR intervals SD_{RRI}:

$$E[t_k] = RR_{trend} \cdot \frac{\sqrt{|k_1 \cdot k_2|}}{\frac{a+b}{SD_{RRI}}}. \qquad (16)$$

The two coefficients a and b are not specified by Wen et al. [17] and thus were approximated by comparison of several values in different magnitudes.

'Point process with history dependent inverse Gaussian distribution' (PPHDIG) is an approach based on a physiologically motivated model, as the IPFM model. Citi et al. mentioned that the Gaussian random walk model with drift is an elementary, stochastic integrate-and-fire model that is able to reflect afferences to the SA node [19]. These excitatory inputs are responsible for the basal cardiac rhythm and the influence of the autonomic nervous system through the sympathetic and para-sympathetic inputs. They used a history-dependent, time-varying model based on the inverse Gaussian probability distribution of the waiting time until the next beat occurs. The probability of the length of the next RR interval, $\tau - u_k$, is described at any beat event u_k by the probability density function (PDF):

$$f\left(\tau - u_k \middle| \mu(H_k, \theta(t)), \lambda(\theta(t))\right) =$$
$$= \sqrt{\frac{\lambda(\theta(t))}{2\pi(\tau - u_k)^3}} \cdot e^{-\frac{1}{2}\frac{\lambda(\theta(t))(\tau - u_k - \mu(H_k, \theta(t)))^3}{(\tau - u_k)\mu^2}} \qquad (17)$$

H_k is the history vector and contains the P previous RR intervals (standard parameter $P = 5$). Further, λ denotes the shape parameter and μ the mean of the inverse Gaussian distribution.

Both depend on the time varying parameters $\theta(t) = \theta_1(t), \dots, \theta_{P+1}(t)$, whereby $\lambda(\theta(t))$ is simply $\theta_{P+1}(t)$. The history dependent mean is a regression of the past P RR intervals with time-varying weights:

$$\mu(H_k, \theta(t)) = \sum_{i=1}^{P} \theta_i(t) w_{k-i}. \qquad (18)$$

The unknown time-varying parameter set $\theta(t)$ is estimated by a local maximum likelihood method. At each time t, the parameter vector that maximizes the local log likelihood in a given observation interval $U_{m:n}$ is obtained. m denotes the index of the first beat in this interval and n the index of the last beat.

$$L(\theta(t)|U_{m:n}) = \sum_{k=m+P}^{n-1} \omega(t - u_{k+1}) \cdot \log[f(u_{k+1} - u_k| \; \mu(H_k, \theta(t)), \lambda(\theta(t)))], \quad (19)$$

where $\omega(\tau) = e^{-\alpha\tau}$ is an exponential weighting function for the local likelihood.

Logarithmic probabilities of the following beat types are calculated [19]: Extra beat, missed beat, misplaced beat, two misplaced beats and resetting beat. The beat is only then classified as normal, if none of a set of hypotheses holds. Erroneous beats are corrected by deletion, insertion or shifting of beats. Before acceptance of a correction, an improvement check is performed. This ensures that the new RR interval time series is always more reliable than the original one.

2 Population and Tests

2.1 Study population

The used dataset contains 151 recordings, obtained from 17 women, aged 24 to 84, 109 men, aged 30 to 84, and from 25 unknown subjects. All known subjects, except of 22 patients, where just the ECG is described, suffer from at least one of the following heart diseases: Myocardial infarction, coronary artery disease, resting angina, effort angina, mixed angina or 1-, 2-, or 3-vessel disease. Only 5-min excerpts of ectopic free regions were used. All tests were performed 10 times with independently corrupted signals (N = 1510). The data is available via physionet [21], an online free-access-database of physiological datasets, and is composed of the European ST-T Database [22], the MIT-BIH Arrhythmia Database [23], and the QT Database [24].

2.2 Test cases

Two test cases were designed to determine the computational performance of each correction approach. In more detail, we monitored the computation time and peak memory using the built-in profiler function of Matlab. Further, we determined the correction ability of each method, whereby detailed results can be found elsewhere.

Test 1 contains only RR time series with single ectopic beats at a moderate density (about one to five ectopic beats with or without compensatory pause per 5-min signal).

Test 2 includes various ectopic beat types (PVC, premature atrial contraction (PAC), PVC couplets and triplets, bi- and trigeminy, sustained and non-sustained ventricular tachycardia).

3 Results

3.1 Test 1: Weakly corrupted RR interval time series

We could observe that physiologic models required far more computation time than simpler ones. The mean execution time is highest for PPHDIG (30 ms), followed by IPFM-C (26 ms, see Figure 2). In contrast, TPC-HT shows the lowest computation time (0.5 ms). However, we could not see this trend in peak memory. It is highest for IPFM-C ($9.7 \cdot 10^5$ kB), whereas all other algorithms result in a rather similar memory usage, ranging from $0.8 \cdot 10^5$ kB to $2.6 \cdot 10^5$ kB (see Figure 3).
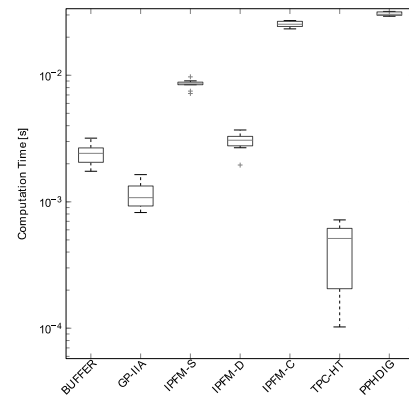

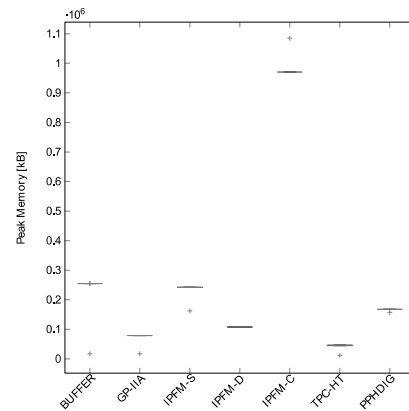**Figure 2**: Box plot of computation time of test 1.

Figure 3: Box plot of peak memory of test 1.

Further, we could demonstrate that all presented models result in a remarkable reduction of the errors in the HRV (data not shown). Correction of single ectopic beats is best achieved by IPFM-S, IPFM-D, TPC-HT and PPHDIG, illustrating the better correction ability of physiologically based models.

3.2   Test 2: Strongly corrupted RR interval time series

Figure 4 illustrates that increasing the amount of artifacts results in a longer median computation time, with respect to test 1. It increased for PPHDIG from 30 to 285 ms, for IPFM-C from 26 to 222 ms and is lowest for GP-IIA (2 ms). On the contrary, peak memory nearly stayed constant. IPFM-C required $9.7 \cdot 10^5$ kB, whereas all other algorithms still resulted in a rather similar memory demand, ranging from $0.8 \cdot 10^5$ kB to $2.6 \cdot 10^5$ kB (see Figure 5).

Similarly to test 1, the correction ability decreased, especially for physiologically motivated approaches.

4   Discussion

Our findings clearly demonstrate the effectiveness of all algorithms to correct both, weakly and strongly corrupted RR interval time series in less than 300 ms for one 5 min ECG signal. Therefore, all of the presented algorithms may also be used in an online fashion. However, attention has to be paid that some algorithms, like the PPHDIG model, require an input of 50 s ECG-signal before the actual correction starts. Thus, some of the models induce a time lag, while still be able to perform an online correction.

The complexity of the different algorithms is well reflected by the median computation time. The PPHDIG model and all IPFM models require more computation time, since they rely on a physiologic relationship. Increasing the error density in test 2 results in a nearly 10-fold rise in computation time for PPHDIG and IPFM-C. This result is not only caused by the higher error density, but also by the fact that both methods were not designed to deal with more complicated artifacts. Therefore, physiologic models just perform best when correcting those specific artifacts they were actually designed for.

Interesstingly, we could not observe any relationship between the complexity of an algorithm and its required peak memory. Most models require a rather similar amount of peak memory, except of IPFM model with cost function.
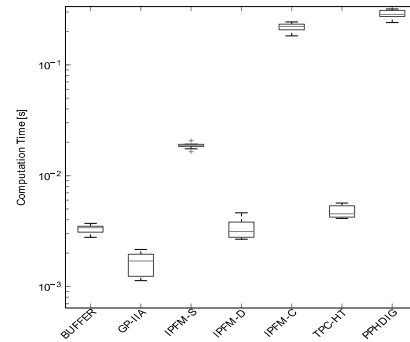

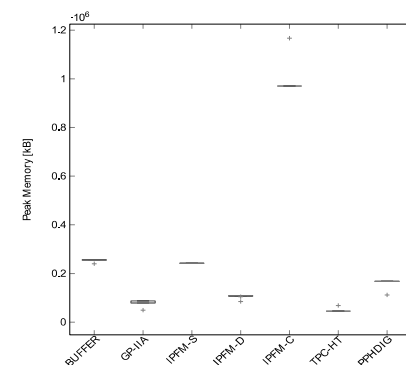**Figure 4**: Box plot of computation time of test 2.

Figure 5: Box plot of peak memory of test 2.

The reason is that the calculation of sinc functions is very memory consuming. Therefore we used a look-up table to improve the computation. Although we could decrease the computation time and the peak memory, the values were still much higher than those of the other algorithms. The original design of this algorithm to correct just single PVCs seems to be responsible for the high computational effort when correcting other types of artifacts.

As stated by Solem et al. IPFM model with δ-parameter requires much less computation time and peak memory [15]. However, the correction performance is compareable only for single ectopic beats, since the range for IPFM-D is much higher when correcting multiple ectopic beats (data not shown).

4.1   Limitations

Though all methods under investigation were implemented to the best of our knowledge and belief, one cannot completely rule out the possibility of programming errors. However, as our results do not show any unexpected outliers, we consider the chances for this scenario negligible. Further, the memory measurement of the MATLAB profiler is an undocumented feature and thus not officialy supported by the Mathworks Company.

5 Conclusion

Physiologically motivated models are best suited to correct single ectopic beats at a compareable memory demand as simpler algorithms. Considering the still very low computation time (below 300 ms per 5 min ECG-signal), these models are also capable to be used in an online fashion.

References

[1] Roger VL, et al. *Heart disease and stroke statistics 2012 update a report from the American heart association.* Circulation, 125(1):e2–e220, 2012.

[2] Nichols M, Townsend N, Luengo-Fernandez R, Leal J, Gray A, Scarborough P, Rayner M. (2012). *European cardiovascular disease statistics 2012.* European Heart Network, Brussels, European Society of Cardiology, Sophia Antipolis, P104.

[3] Task Force of ESC and NASPE. *Heart rate variability: standards of measurement, physiological interpretation and clinical use.* Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation. Mar 1996; 93(5):1043–1065

[4] Mateo J and Laguna P. *Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal.* IEEE Trans Biomed Eng. Mar 2003; 50(3):334–343

[5] Thuraisingham RA. *Preprocessing RR interval time series for heart rate variability analysis and estimates of standard deviation of RR intervals.* Comput Methods Programs Biomed. Jul 2006; 83(1):78–82

[6] Kumaravel N, Santhi C. Nonlinear filters for preprocessing heart rate variability signals. *International Journal of Computer Science and Network Security.* 2010; 10:250–254

[7] Colak OH. Preprocessing effects in time--frequency distributions and spectral analysis of heart rate variability. *Digital Signal Processing*, Elsevier, 2009; 19: 731-739

[8] Peltola MA. Role of editing of R-R intervals in the analysis of heart rate variability. *Front Physiol.* 2012; 3:148

[9] Lippman N, Stein KM, Lerman BB. Nonlinear predictive interpolation. A new method for the correction of ectopic beats for heart rate variability analysis. *J. Electrocardiol.*1993; 26(Suppl.:14–19)

[10] Peltola MA. Role of editing of R-R intervals in the analysis of heart rate variability. *Front Physiol.* 2012; 3:148

[11] Mietus JE. *Time domain measures: from variance to pnnx.* http://physionet.org/events/hrv-2006/mietus-1.pdf, 2006. [Online; Accessed: 03/2014].

[12] McNames J, Thong T, Aboy M. Impulse rejection filter for artifact removal in spectral analysis of biomedical signals. *Conf Proc IEEE Eng Med Biol Soc.* 2004;1:145–148

[13] Begum S, Islam MS, Ahmed MU, Funk P. K-nn based interpolation to handle artifacts for heart rate variability analysis. In Signal Processing and Information Technology (ISSPIT), *2011 IEEE International Symposium on.* 2011; P. 387–392

[14] Keenan DB. Detection and correction of ectopic beats for hrv analysis applying discrete wavelet transforms. *Int. J. Inf. Technol.* 2005; 2:54–60

[15] Solem K, Laguna P, Sornmo L. An efficient method for handling ectopic beats using the heart timing signal. *IEEE Trans Biomed Eng.* Jan 2006; 53(1):13–20

[16] Brennan M, Palaniswami M, Kamen P. A new model-based ectopic beat correction algorithm for heart rate variability. In Engineering in Medicine and Biology Society, 2001. *Proceedings of the 23rd Annual International Conference of the IEEE*, volume 1, pages 567–570 vol.1, 2001.

[17] Wen F, He FT. *An efficient method of addressing ectopic beats: new insight into data preprocessing of heart rate variability analysis.* J Zhejiang Univ Sci B. Dec 2011; 12(12):976–982

[18] Rand J, Hoover A, Fishel S, Moss J, Pappas J, Muth E. Real-time correction of heart interbeat intervals. *IEEE Trans Biomed Eng.* May 2007; 54(5):946–950

[19] Citi L, Brown EN, Barbieri R. A real-time automated point-process method for the detection and correction of erroneous and ectopic heartbeats. *IEEE Trans Biomed Eng.* Oct 2012; 59(10):2828–2837

[20] Ward S, Heneghan C, Nolan P. An integrate-and-fire based model of PP and PR variability in the human electrocardiogram. In Neural Engineering, 2003. Conference Proceedings. *First International IEEE EMBS Conference* on, pages 297–300, 2003.

[21] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, . Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, Physiotoolkit, and Physionet: components of a new research resource for complex physiologic signals. *Circulation.* 2000; 101(23):e215–e220

[22] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi. The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *Eur. Heart J.,* 13(9):1164–1172, Sep 1992.

[23] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. *Engineering in Medicine and Biology Magazine*, IEEE. May 2001; 20(3):45–50

[24] Laguna P, Mark RG, Goldberg A, Moody GB. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. *Computers in Cardiology* 1997, P. 673–676. IEEE, 1997.