Validation of Baseline Wander Removal and Isoelectric Correction in Electrocardiograms Using Clustering

Kjell Le,
Trygve Eftestøl,
and Kjersti Engan
Department of Electrical Engineering
and Computer Science
University of Stavanger
Email: kjell.le@uis.no

Øyunn Kleiven
Department of Cardiology
Stavanger University Hospital

Stein Ørn*†
*Department of Cardiology
Stavanger University Hospital
†Department of Electrical Engineering
and Computer Science
University of Stavanger

***Abstract*—The removal of baseline wander from the electrocardiogram does not always correct the isoelectric level of the signal. That is, to enforce reference points for amplitude measurements to 0 volt. In this work new parameters and a modified clustering method is proposed to find the isoelectric bias, which is the amplitude difference between the reference point and 0 volt prior to the isoelectric correction. Validation with old parameters and previous iteration of the clustering method are also performed on the QT database from PhysioNet. Both methods are viable to find the isoelectric bias in the PQ-segment. 3 assessment criteria are used. The first criterion (A) is based on the location of the isoelectric bias with respect to the PQ-segment. The second criterion (B) is based on the absolute difference between the isoelectric bias and median amplitude of the PQ-segment. And the third criterion (C) is the mean and standard deviation of the absolute difference between the isoelectric bias and median amplitude of the PQ-segment.**

Best results are obtained with the new proposed method with a probability of true detection (PD) of 99.62 pct (A), PD of 98.40 pct (B), mean of 6.28 microvolt (C) and standard deviation of 12.03 microvolt (C) respectively for the criteria mentioned above.

***Index Terms*—Baseline wander, Isoelectric correction**

I. Introduction

The removal of baseline wander, a low frequency noise, from the electrocardiogram (ECG) does not always correct the isoelectric level of the signal [1], [2]. The isoelectric correction is useful in wave amplitude measurements, and improves mean electrical axis calculation [1], [3]. The correction is achieved by enforcing specific reference points [4], [5] to $0\,\mathrm{V}$. Typical amplitude measurements for example P wave and QRS complexes are measured with respect to P-onset and QRS-onset respectively [5]. The baseline wander removal and the isoelectric correction can be regarded as two separate problems [1]. The baseline wander removal is the removal of low frequency noise while the isoelectric correction is to enforce the reference points to $0\,\mathrm{V}$. The term *isoelectric bias* is used to describe the amplitude offset of the reference point prior to the isoelectric correction. See for example in figure 2 and 3 for unadjusted and adjusted respectively.

Because of annotation errors, either automatic or manual [6], [7], and for robustness of point selection the isoelectric bias is often estimated from regions of low electrical activity, like the PQ-segment (PR-segment) [4], [7], the ST-segment and the TP-segment [8]. Even though the ST-segment exhibits low electrical activity in a normal ECG there is a potential ST-elevation indicating ischemic heart disease. Therefore the ST-segment is in general unreliable to be used to find the reference point.

Other methods which remove the baseline wander and correct the isoelectric level is the cubic spline interpolation (CSI) method [9] and the quadratic variation reduction (QVR) method [2]. Both methods use a point with a fixed location relative to the QRS complex as a reference point. Such points are also known as knots. The CSI method would fail in a signal averaged template because there is only one knot in the template. Increasing the number of knots derived from the TP-segment would still not be viable because of edge effect [1], where the end points are not necessarily knots. There is also the difficulty of finding the T-offset [4], [10]. The QVR has been shown to perform well with one knot [2]. However, there is no guarantee that all reference points are enforced to $0\,\mathrm{V}$ when only one knot is used [1].

A preliminary study of the proposed, clustering, method has been published recently in [1]. In the previous work the proposed method was applied to the PQ-segment within a signal averaged template. The isoelectric bias is to be considered constant. This is to ensure that the isoelectric level of the template is adjusted to $0\,\mathrm{V}$ [1]. Calculation of the mean electrical axis (MEA) [3] was used to validate the isoelectric correction in [1]. The MEA can be calculated from any pair of leads in the frontal plane. Therefore a low variance across the measurements from all the lead pairs indicates consistent calculation of the MEA [1], [3].

The proposed method was not validated with a public

978-9-0827-9703-9/19/$31.00

database. A sensible validation is to check if the isoelectric bias is within the PQ-segment prior to the isoelectric correction. This means that the reference point is in the PQ-segment. Because the clustering method finds the isoelectric bias it is not directly comparable with other segmentation methods which find the P-offset and QRS-onset first.

In this work a further developed method from [1] is presented. The proposed method is computationally more efficient by simplification in the input and decision of the main cluster. The input is cut in half while the decision is simplified into one rule only. A larger set of parameters are tested. The method is thoroughly tested and validated on the QT database [11] with different assessment criteria.

II. Materials and Methods

A. The QT Database

The QT database [11] from PhysioNet consists of 105 ECG records with a sampling frequency of 250 Hz, and physical unit in millivolt. Each record contains two leads. The database has two annotators. Because of lack of annotations from the second annotator, only the first annotator's second pass (annotator q1c) is used. Only annotation sequence "p)(N", which corresponds to P peak, P-offset, QRS-onset and normal beat respectively, is considered. This is because there are no other noteworthy and representative combinations in the annotation set.

B. Preprocessing

The baseline wander is removed first with a simple highpass filter (IIR 4th order Butterworth) with a cutoff frequency $f_c = 0.67$ Hz. Forward-backward filtering is used to ensure linear/zero phase. Normally high frequencies above 150 Hz should be removed according to the recommendation in [12]. This is not necessary because the Nyquist frequency is 125 Hz.

C. Finding the Isoelectric Bias

The algorithm to find the isoelectric bias for a beat is outlined in algorithm 1. The input to the algorithm is a sorted set of amplitudes, S, of length L (milliseconds) prior to the R peak. The amplitude, y_i, is ordered from lowest to highest amplitude to ensure that the algorithm gives the same result every time it is executed [1]. The R peaks are found from manual annotation. The output of the algorithm is the (local) isoelectric bias $\hat{b}$ for a beat. Local isoelectric correction is performed by subtracting the isoelectric bias with the signal.

Finding d in the for loop requires only one calculation since the elements in the input set S are ordered [1]. The main cluster C_{iso} is simply the cluster with the highest number of elements. The isoelectric bias, $\hat{b}$, is the average of amplitudes in C_{iso},

$$\hat{b} = \frac{1}{|C_{iso}|} \sum_{y \in C_{iso}} y, \tag{1}$$

where $|C_{iso}|$ is the total number of elements in the cluster.

The algorithm groups the amplitudes such that each cluster contains samples which ideally does not classifies as a wave.

Algorithm 1 Find the isoelectric bias

Input: $S = \{y_0, y_1, \ldots, y_{N-1}\}, y_0 \leq y_1 \leq \cdots \leq y_{N-1}$
Output: $\hat{b} = |C_{iso}|^{-1} \sum_{y \in C_{iso}} y$

```
k ⇐ 0
C_k ⇐ {y_0}
for i=1 to i=N-1 do
    d ⇐ max{|y − y_i|}, ∀y ∈ C_k
    if d < ϵ then
        C_k ⇐ C_k ∪ {y_i}
    else
        k ⇐ k + 1
        C_k ⇐ {y_i}
    end if
end for
C_iso ⇐ arg max_{C_k} {|C_k|}, ∀k
```

The cluster that has the highest number of elements must have samples from the isoelectric segment. A line estimation without a slope is simply the averaged of the amplitudes.

Previously in [1] the input segment was a (sorted) segment within a neighbourhood L of the R peak, where the elements on the right side of the R peak were included in the clustering as well. However, to find C_{iso} i.e. a) having the highest number of elements on the left side of the R peak and b) having the highest total number of elements in the cluster, will most of the time simplify to the new simplified decision. For brevity the two methods are referred as R-left (this work) and R-dual (in [1]) depending on their inputs. The inputs are visualized in figure 1. The R-left method uses samples from the solid red line as input. The R-dual method uses samples from the solid red- and dashed green line.

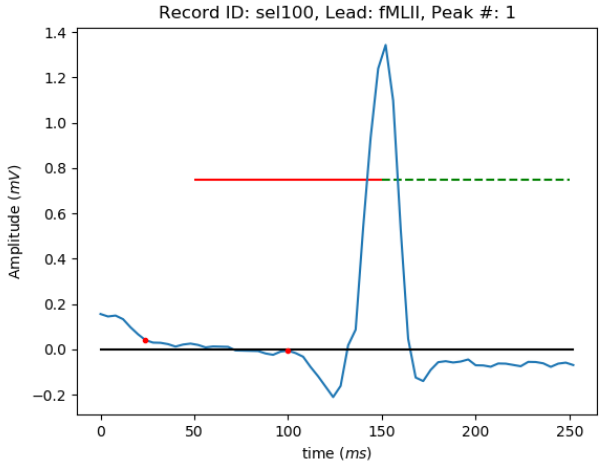

Fig. 1. An example of QRS complex of record ID sel100. Red dots are P-offset and QRS-onset from manual annotations (q1c). The red solid line shows L (100 ms) length prior to the R peak, used as input (sorted) in the R-left method. The red solid line with the green dashed line show the L neighbourhood of the R peak, used in the R-dual method. Most of the samples in the PQ-segment are within the red solid line.

D. Cubic spline interpolation knots selection

The CSI knots selection method in [9] is used for comparison. The method selects a fixed point prior the QRS complex. For robustness the knot is an averaged of a neighbourhood of 16 ms (4 samples, 9 samples total) of the fixed point. The QVR method shares the same knots as in the CSI method.

E. Assessment

A sensible verification of the method is to test whether the isoelectric bias is *within* the PQ-segment prior to isoelectric correction. Since ϵ should be below 30 µV [1] and amplitude measurement error in the order of 20-40 µV (depending on the peak's amplitude) is acceptable [5], a location based criterion such as

$$\min(PQ) \leq \hat{b} \leq \max(PQ), \tag{2}$$

where PQ is the set of amplitudes in the PQ-segment derived from manual annotation can be used. Alternatively there exist a zero-crossing, or zero, in the PQ-segment after the isoelectric correction. This criterion is denoted as criterion A.

However, the annotations are not perfect [6], an example of wrong annotation is shown by the red marks representing the P-offset and QRS-onset in figure 2. To make the assessment fair to that type of problem, criterion B is defined: The absolute difference between the isoelectric bias, $\hat{b}$, and the median of the amplitudes in the PQ-segment is less than σ_{th},

$$|\operatorname{median}(PQ) - \hat{b}| < \sigma_{th}, \tag{3}$$

then $\hat{b}$ is considered to be within the PQ-segment. Criterion B is an amplitude based criterion. The threshold parameter σ_{th} should be in the range of the acceptable amplitude measurement error or less. To simulate errors in the location of P-offset and QRS-onset the PQ-segment derived from manual annotation is extended by a few samples on both direction.

The probability of true detection (P_D) is used as an assessment criterion. The P_D is calculated as

$$P_D = \frac{TP}{N}, \tag{4}$$

where TP is the true positive, i.e. the number of isoelectric bias that is within its respective PQ-segment prior to isoelectric correction, and N is the total number of (input) segments which are evaluated. Since it is not certain which lead the annotation is based on [6], the best result from each lead is reported [13]. (N is therefore based on one of the leads and not both in each record). The P_D which is based on (2), criterion A, shows the number of isoelectric bias which are within (location based) the PQ-segment. The P_D which is based on (3), criterion B, shows the number of isoelectric bias which are within (amplitude based) a threshold between the bias and median amplitude of the PQ-segment. These P_D will be denoted as P_D A and P_D B based on criterion A and B, respectively.

Another assessment is done based on the mean and standard deviation of the absolute difference between the isoelectric bias and the median of the amplitudes in the PQ-segment. The lower value the better.

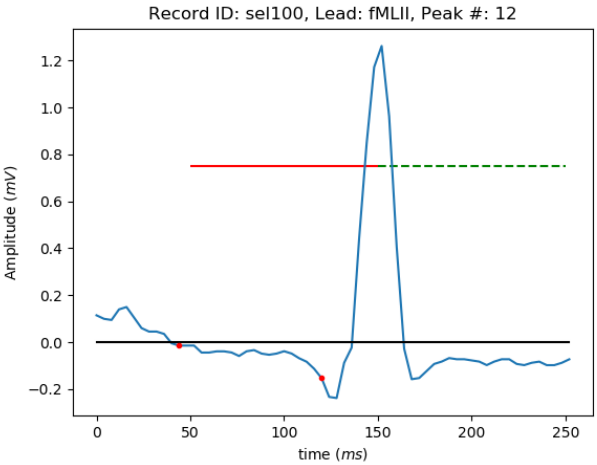

Fig. 2. Similar figure as in figure 1. However, the manual annotations (red dots) are incorrectly placed. Comparison with figure 1 the QRS-onset should be at 100 ms.

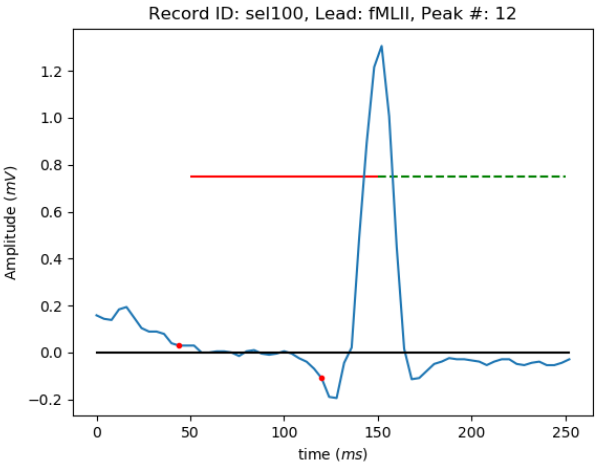

Fig. 3. Same figure as in figure 2, but isoelectric adjusted i.e. reference point is enforced to 0 V.

III. Experiment and Results

The experiment is performed on the QT database with both the R-left and R-dual methods. The following parameters are used for testing: The ϵ variable used for clustering is varied among the values **10**, **20**, and 30 µV. **Bold** values are the same parameters used in [1]. The σ_{th} used in (3) is 30 µV. The segment neighbourhood (R-dual) or length of the segment (R-left) L is varied among the values **100**, 110 120, 130, 140. The number of offset samples to extend the PQ-segment derived from manual annotation is varied among 0 to 3 samples, in each direction. This corresponds to 0 to 12 ms.

To not be overwhelmed by the results with all the possible parameter values, only the top 5 by various assessment criteria are shown. Table I, II, and III show results for the R-left and R-dual methods, with P_D A, P_D B, and mean and standard deviation as assessment criteria respectively. Table IV, V, and VI respectively as above but with only parameters used in [1].

Table VII shows results for no isoelectric correction, when the isoelectric bias is per definition set to $0\,\mathrm{V}$, a baseline result. Table VIII shows the results for knots selection method in [9]. The number of total beats is $N = 3193$. Figure 3 shows an example of isoelectric correction. Figure 2 shows the same figure without an isoelectric correction.

TABLE I
TOP 5 RESULTS FOR P_D A BASED ON (2).

L (ms)	ϵ (µV)	Offset samples	Method	P_D A (%)
130	30	3	R-left	99.62
130	20	3	R-left	99.53
120	30	3	R-left	99.40
130	10	3	R-dual	99.31
130	20	2	R-left	99.31

TABLE II
TOP 5 RESULTS FOR P_D B BASED ON (3).

L (ms)	ϵ (µV)	Offset samples	Method	P_D B (%)
130	30	0	R-left	98.40
140	30	2	R-left	98.40
140	30	1	R-left	98.37
130	30	2	R-left	98.34
140	30	0	R-left	98.34

TABLE III
TOP 5 RESULTS BASED ON MEAN AND STANDARD DEVIATION.

L (ms)	ϵ (µV)	Offset samples	Method	μ (µV)	σ (µV)
130	30	0	R-left	6.28	12.03
130	30	1	R-left	6.36	12.04
130	30	2	R-left	6.53	12.11
130	20	0	R-dual	6.54	13.76
140	30	0	R-left	6.57	13.66

TABLE IV
TOP 5 RESULTS FOR P_D A BASED ON (2), WITH PARAMETERS IN [1].

L (ms)	ϵ (µV)	Offset samples	Method	P_D A (%)
100	20	3	R-left	97.75
100	20	3	R-dual	97.71
100	10	3	R-dual	97.68
100	10	3	R-left	97.56
100	20	2	R-left	97.34

TABLE V
TOP 5 RESULTS FOR P_D B BASED ON (3), WITH PARAMETERS IN [1].

L (ms)	ϵ (µV)	Offset samples	Method	P_D B (%)
100	20	2	R-dual	96.05
100	20	1	R-dual	95.96
100	20	2	R-left	95.93
100	20	0	R-left	95.83
100	20	1	R-left	95.83

TABLE VI
TOP 5 RESULTS BASED ON MEAN AND STANDARD DEVIATION, WITH PARAMETERS IN [1].

L (ms)	ϵ (µV)	Offset samples	Method	μ (µV)	σ (µV)
100	20	1	R-dual	16.87	104.49
100	20	0	R-dual	16.90	104.41
100	20	2	R-dual	17.04	104.87
100	20	3	R-dual	17.40	105.19
100	20	0	R-left	18.69	117.28

TABLE VII
NO ISOELECTRIC CORRECTION. THE ISOELECTRIC BIAS IS PER DEFINITION $0\,\mathrm{V}$.

Offset samples	P_D A (%)	P_D B (%)	μ (µV)	σ (µV)
3	67.18	40.90	47.35	38.95
2	54.68	40.40	48.01	39.45
1	44.00	39.74	48.67	39.93
0	35.48	39.30	49.31	40.31

TABLE VIII
CUBIC SPLINE KNOTS SELECTION. AVERAGED OF 9 SAMPLES, $32\,\mathrm{ms}$.

Offset samples	P_D A (%)	P_D B (%)	μ (µV)	σ (µV)
3	98.56	94.68	10.81	18.49
2	97.81	94.49	10.71	18.91
1	96.87	94.64	10.65	19.20
0	95.24	94.74	10.67	19.67

IV. DISCUSSION

An example of no isoelectric correction and isoelectric correction can be seen in figure 2 and 3 respectively. Similar results were evident when the R-left and R-dual methods were compared based on the same parameters as in [1] (see table IV, V, and VI). However, for other parameters as in table I, II, and III, the R-left method dominates, but the R-dual method is amongst the top five. Except for P_D B based on (3) as shown in table II.

For parameters used in [1] with the R-dual method, table IV, the best result for P_D A is $97.71\,\%$, with parameters $L = 100\,\mathrm{ms}$ and $\epsilon = 20\,\mu\mathrm{V}$. This means that the isoelectric bias is $97.71\,\%$ of the time within the PQ-segment, with 3 offset samples. For P_D B, table V, the best result is $96.05\,\%$ with parameters $L = 100\,\mathrm{ms}$ and $\epsilon = 20\,\mu\mathrm{V}$, with 2 offset samples. It means that $96.05\,\%$ isoelectric biases are within an acceptable error, $30\,\mu\mathrm{V}$, for amplitude measurements. For the mean and standard deviation, the best result is $\mu = 16.87\,\mu\mathrm{V}$ and $\sigma = 104.49\,\mu\mathrm{V}$, with parameters $L = 100\,\mathrm{ms}$, $\epsilon = 20\,\mu\mathrm{V}$, and 1 offset sample as shown in table VI. The mean is acceptable, but the standard deviation is a bit high. This might be because of outliers. Using a robust dispersion measurement such as median absolute deviation (MAD) might be preferable in this case. Though, the results for parameters used in [1] with the R-dual method do show that the R-dual method is a valid isoelectric correction method, i.e. it finds an acceptable isoelectric bias.

The parameters $L = 130\mathrm{ms}$ and $\epsilon = 30\mu\mathrm{V}$ show the best results across the assessment criteria with the R-left method. In table I the best P_D A is $99.62\,\%$, with 3 offset samples. In

table II the best P_D B is $98.40\,\%$, with 0 offset samples. And for the mean and standard deviation, table III, the best results are $\mu = 6.28\,\mu\text{V}$ and $\sigma = 12.03\,\mu\text{V}$, with 0 offset sample. The mean is well within an acceptable range and the same can be said with the standard deviation. These new parameters and clustering method should be run through the MEA method as a further fidelity validation as in [1]. This is necessary because of potential over fitting [13]. A high P_D with a high ϵ might suggest that some records are subjected with white noise.

Nevertheless, wrong annotations are also a source of error. The P_D A would benefit from error such as shown in figure 2, where the QRS-onset is slightly moved to the right. The P_D B does not benefit from the same error. As seen in table I and II, the best result for the P_D A has 3 offset samples, but the best result for the P_D B has no offset samples. The same trend can be seen in table VII and VIII.

Table VII shows the baseline evaluation results when no isoelectric correction is performed. The results show a very poor P_D. Best results are with 3 simulated offset samples. The P_D A is $67.18\,\%$, P_D B is $40.90\,\%$, mean is $47.35\,\mu\text{V}$, and standard deviation is $38.95\,\mu\text{V}$. It can be concluded that $0\,\text{V}$ cannot be used as a reference point in an ECG signal without an isoelectric correction, even after a baseline wander removal.

For comparison, table VIII shows the results for the CSI knots selection method in [9]. However, it cannot be explained why the P_D B is less than in table V, but table VI shows a higher mean and standard deviation. This reflects the need for different assessment criteria of what is within a PQ-segment. The best results are from various offset samples. The best P_D A is $98.56\,\%$ (3 offset samples), P_D B is $94.74\,\%$ (0 offset sample), $\mu = 10.67\,\mu\text{V}$ (0 offset sample) and $\sigma = 18.49\,\mu\text{V}$ (3 offset samples). The results show a poorer result than the best results from the proposed method.

V. Conclusion

In this work an extension of parameters were tested on two methods, R-left and R-dual. The R-left method shows a satisfactory results in all assessment criteria. The R-dual method with parameters as in [1] shows acceptable results. The P_D of the two within criteria are relatively high. The mean is low, however, the standard deviation is a bit high. The R-left method should be preferred over the R-dual method because of its simplicity and computationally more efficient. The method also shows a better result than the CSI knots selection method in [9]. The best results are obtained with the R-left method with a P_D A of $99.62\,\%$, P_D B of $98.40\,\%$, mean of $6.28\,\mu\text{V}$ and standard deviation of $12.03\,\mu\text{V}$. No isoelectric correction does not show satisfactory results. Reference points must be found if no isoelectric correction is performed.

References

[1] K. Le, T. Eftestøl, K. Engan, Øyunn Kleiven, and S. Ørn, "Baseline wander removal and isoelectric correction in electrocardiograms using clustering," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[2] A. Fasano and V. Villani, "ECG baseline wander removal with recovery of the isoelectric level," in *Computing in Cardiology Conference (CinC), 2015*. IEEE, 2015, pp. 577–580.

[3] K. Le, T. Eftestøl, K. Engan, Øyunn Kleiven, and S. Ørn, "Invariant mean electrical axis in electrocardiogram," in *Computing in Cardiology Conference (CinC)*, 2018.

[4] G. D. Clifford, F. Azuaje, P. McSharry *et al.*, *Advanced methods and tools for ECG data analysis*. Artech house Norwood, MA, 2006.

[5] I. E. Commission *et al.*, "IEC 60601-2-25: 2011: Medical electrical equipment–part 2-25: Particular requirements for the basic safety and essential performance of electrocardiographs [international standard]," 2011.

[6] R. Jane, A. Blasi, J. Garcia, and P. Laguna, "Evaluation of an automatic threshold based detector of waveform limits in holter ECG with the QT database," in *Computers in Cardiology 1997*, Sep. 1997, pp. 295–298.

[7] C. Zywietz, J. Willems, P. Arnaud, J. Van Bemmel, R. Degani, and P. Macfarlane, "Stability of computer ECG amplitude measurements in the presence of noise," *Computers and Biomedical Research*, vol. 23, no. 1, pp. 10–31, 1990.

[8] J. Stephenson, "Detection of isoelectric baseline and high frequency noise within electrocardiographique signal," *ESPCI, Laboratoire d'Electronique*, 2000.

[9] C. Meyer and H. Keiser, "Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques," *Computers and Biomedical Research*, vol. 10, no. 5, pp. 459–470, 1977.

[10] D. Nairn, P. Bonizzi, J. Karel, and A. R. A. Hernandez, "Performance analysis of T-wave-offset detection algorithms on patients with cardiac diseases," in *Computing in Cardiology*, 2018.

[11] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG," in *Computers in cardiology 1997*. IEEE, 1997, pp. 673–676.

[12] P. Kligfield, L. S. Gettes, J. J. Bailey, R. Childers, B. J. Deal, E. W. Hancock, G. van Herpen, J. A. Kors, P. Macfarlane, D. M. Mirvis, O. Pahlm, P. Rautaharju, and G. S. Wagner, "Recommendations for the standardization and interpretation of the electrocardiogram, part I: The electrocardiogram and its technology," *Circulation*, vol. 115, no. 10, pp. 1306–1324, 2007.

[13] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ECG delineator: evaluation on standard databases," *IEEE Transactions on biomedical engineering*, vol. 51, no. 4, pp. 570–581, 2004.