Sensitivity of QRS Detection Accuracy to Detector Temporal Resolution

Katarzyna Heryan[1], Wojciech Reklewski[1], Andrzej Szaflarski[2], Maciej Ordowski[1], Piotr Augustyniak[3], Marek Miśkowicz[1]

AGH University of Science and Technology, Cracow, Poland
[1] Department of Measurement and Electronics
[2] Faculty of Computer Science, Electronics and Telecommunications
[3] Department of Biocybernetics and Biomedical Engineering

Abstract

The reliable evaluation of the QRS detection algorithm requires comparability and reproducibility. Although it is commonly accepted to test QRS detection accuracy by standard binary classification parameters, much less attention is paid to the temporal accuracy of the detector. A variety of temporal tolerance values are used in the literature for performance evaluation of QRS detection, ranging from 60 ms to 160 ms, which sometimes result in comparison of algorithms with different temporal resolutions. This paper addresses a problem of the dependence of the accuracy of QRS detection algorithms represented by detection error rate, sensitivity, and positive predictivity, on the temporal resolution of the detection defined by Detector Temporal Tolerance (DTT). In this work, the classification statistics achieved for three state-of-the-art low complexity algorithms in a broad range of DTT (from 160 ms to 8 ms) for the entire standard MIT-BIH Arrhythmia Database are compared with the performance of the Pan-Tompkins algorithm. The analysis shows that along with decreasing value of DTT, the classification statistics for R-peak detection algorithms deteriorate, while the deterioration rate is characteristic of a given algorithm. In addition, the algorithms change their positions in the detection accuracy ranking with changing DTT value. The performed analyses proved that DTT is an integral parameter of QRS complex detection that determines the reproducibility of test results and fair comparative study.

1. Introduction

The QRS complex detection is a very active research area extensively studied over the last decades with more than 500 papers reported per annum [1]. The accuracy of QRS detection is evaluated conventionally by the binary classification parameters such as Sensitivity $S = TP/(TP + FN)$, Positive Predictivity $+P = TP / (TP + FP)$ and Detection Error Rate $DER = (FP + FN)/(TB)$, where TP is the number of correctly detected R-peaks (True Positive), FN is the number of omitted R-peaks (False Negative), FP is the number of incidences that were wrongly classified as R-peaks (False Positive), and TB is a number of annotated R-peaks in a database (Total Beats). In QRS detection, the number of incidences that are not annotated as R-peaks is not considered, so TN (True Negative) is undefined.

Most research works on QRS detection compete to gain top statistics for these metrics. At the same time, much lower attention is paid to the temporal accuracy of the detector [2]. Meanwhile, the numerical values of binary classification parameters depend on the adopted temporal tolerance of QRS detection. Let us define the Detector Temporal Tolerance (DTT) as the maximum allowed time difference between the peaks fiducial points detected by the algorithm (t_D) and the corresponding annotations from the reference database (t_A). The R-peak detected anywhere within the tolerance window of $2DTT$ length centered at t_A is classified as correct (TP):

$$TP: \;\; t_D \in t_A \pm DTT \tag{1}$$

The higher the DTT, the better the numerical results of the aforementioned accuracy metrics. However, the R-peak detected within a long tolerance window that is distant from its reference annotation, is still classified as correct (TP), although it might even be located in the intersection of tolerance windows of two adjacent QRS complexes (Fig. 1). Therefore, for a reliable classification, the tolerance window should not be longer than the duration of QRS complex (compare Fig. 2). Otherwise, it is possible that the detected R-peaks that fall within the window $2DTT$ and are treated as TP, should contribute to both FP (the wrong sample was detected as R-peak) and FN (the right R-peak was not detected) at the same time.

Inaccurate R-peak detection stems from the finite precision of detection algorithms, noise, and abnormal heartbeats. For example, some algorithms classify as the R-

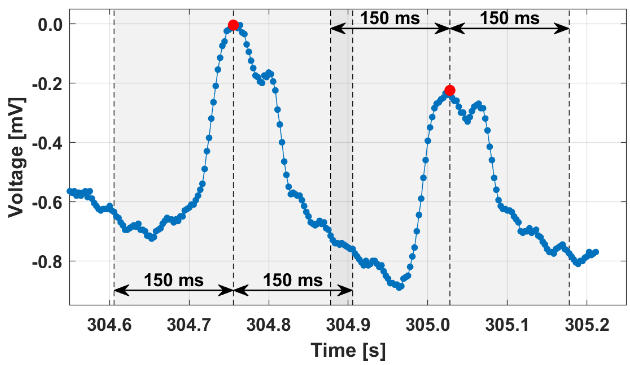

Figure 1. Excerpt from 205 ECG record of the MIT-BIH Arrhythmia Database. The tolerance windows defined around adjacent R-peak annotations overlap.

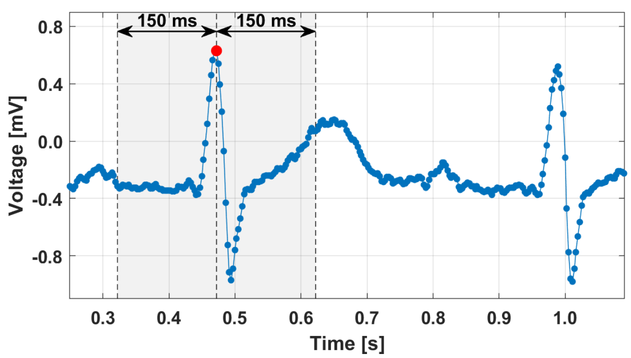

Figure 2. Excerpt from 215 ECG record of the MIT-BIH Arrhythmia Database containing a QRS complex that lasts for 67 ms (24 samples), while the tolerance window 150 ms (ANSI recommendation). A potential TP detection can be distant from the actual QRS complex.

peak the first local maximum in the QRS searching window, while the others consider the same for the absolute extremum. Both methods may imply unnoticed false detections. Additionally, the filtering operations applied at preprocessing stage introduce extra lag defined by the filter group delay [2].

Nevertheless, an accurate R-peak temporal localization is crucial for subsequent ECG signal analysis, since any advanced test is based on accurate R-peak detection and a possible error in this respect is further propagated. For example, the inaccurate R-peak localization affects not only the assessment of heart rate (HR) and, consequently, the heart rate variability (HRV) but also ML-based analysis. In the case of the latter, the surrounding of the wrongly identified QRS complex, which is an entirely different excerpt of the ECG signal, is the basis of the calculations of morphological parameters. Moreover, prominent accuracy statistics do not necessarily translate into the diagnostic relevance of such R-peak detectors. From the clinical point of view, the ability of the algorithm to recognize normal R-peaks is inadequate, as it is more important to correctly diagnose pathological phenomena. These are certainly present less frequently in ECG records and therefore may not significantly affect the classification statistics. However, the correctness of their interpretation depends on the temporal accuracy of R-peak detectors, and as proved in [2], it is for abnormal beats that the largest difference $|t_A - t_D|$ indeed occurs.

A variety of DTT values have been chosen in the literature for performance evaluation of QRS detection algorithms: 60 ms [3], 75 ms [4], 100 ms [5], 150 ms [6, 7] and 160 ms [8]. But, even more frequently, this parameter is not explicitly provided, e.g. [9–11], which may result in a comparison of algorithms with a different temporal resolutions of the detection. On the one hand, the ANSI recommendation defines a value of 150 ms as temporal tolerance for QRS detection [12]. Its objective is to provide a criterion for algorithms comparison. However, the value of 150 ms is rather a moderate requirement. As seen in (Fig. 1), which is an exemplary excerpt of ECG signal from the MIT-BIH Arrhythmia database, the tolerance windows for two adjacent QRS complexes overlap. On the other hand, for HRV analysis, the expected temporal precision is in the range of a few milliseconds [13–15]. This value is more than an order of magnitude lower than ANSI recommendation and DTT reported in the majority of research papers on QRS detection [3–8].

Given the aforementioned incoherence regarding DTT, the aim of this work is to examine how sensitive the QRS detection algorithms are to DTT. For this reason, we have performed an experiment in which the temporal tolerance is an input parameter of QRS detection and the outputs are the classification metrics TP, TN, FN.

2. Experiment specification

The algorithms selected to examine the sensitivity of QRS detection accuracy to DTT have been designed for real-time operation. The Pan-Tompkins algorithm [16] was chosen as it is a key reference in QRS detection. For the sake of this research, its simplified version with publicly available source code was used [17]. The other selected algorithms [6–8] belong to the category of low computational complexity and high energy efficiency methods. Here, these were implemented in Python based on their descriptions provided in the original papers.

The algorithms [7, 8, 16] work along with similar subsequent signal processing steps: filtering of the ECG signal, feature signal determination, and designation of heartbeat fiducial points in QRS complex with adaptive thresholding of the feature signal. However, the Pan-Tompkins algorithm has more elaborate input signal filtering, employs more numerical operations, and uses additional resources for recovering possible missed R-peaks in the search back procedure.

The operation of [6] is different since it uses the level-crossing sampling [18] that provides information on ECG local extrema. These extrema are subsequently classified as either R-peak or noise peak based on peak width with thresholds that are adjusted to input signal properties.

To provide comprehensive analysis, the algorithms have been tested for the entire standard MIT-BIH Arrhythmia Database that comprises 48 ECG recordings lasting 30 minutes each, acquired with sampling frequency equal to 360 Hz and with 11-bit resolution covering 10 mV range, having TB=109 494 [19, 20]. This database is especially dedicated to the QRS detection test as it constitutes a good mixture of normal and pathological R-peaks with various morphology, as well as noise artifacts challenging the detection process.

3. Test results

We conducted an experiment to learn how the aforementioned accuracy metrics ($+P$, S, DER) change with decreasing DTT for selected algorithms. These results are presented in detail in Table 1, for DTT varying from 163.89 ms to 8.33 ms with a step equal to 11.12 ms (4 samples). In order to visualize the observed dependencies, the DER is also presented as the function of DTT in Fig. 3.

As expected, along with decreasing DTT (improving the detector temporal resolution), the statistics for each algorithm gradually deteriorate. For DTT = 152.78 ms, which is close to ANSI recommendation, DER ranges between 1% [7, 8, 16] and around 2% [6], and for DTT = 108.33 ms, DER is still below 3% for all tested algorithms. With further reduction of DTT until 60 ms, DER increases up to 6% for all except the Pan-Tompkins algorithm [16]. For high temporal resolution (low DTT), DER may exceed 100%, which represents an extreme case when the number of errors (FN and FP) is higher than the number of R-peaks annotated in the database.

The deterioration rate along with decreasing DTT is an individual characteristic of a given algorithm. Both [6] and [7] methods experience slight steadily performance decline for the whole evaluated DTT range. While the other two [8, 16] demonstrate dramatic drops at certain values of DTT. For the Pan-Tompkins algorithm [16], a sharp increase of DER occurs for a relatively high DTT equal to 108 ms. But in this particular case, it may be a consequence of original method simplification [17]. Consequently, which is worth noting, particular algorithms change their position in the accuracy ranking with DTT varying from 60 ms to 160 ms (compare the bold values of DER scores in Table 1 and Fig. 3).

The natural conclusion from the conducted analysis is that the mere provision of the algorithm statistics ($+P$, S, DER), without specifying the DTT for which they were determined, precludes a reliable assessment of the QRS detection performance, as well as reproducibility of the test results. Furthermore, a lack of DTT specification makes a fair comparative performance analysis between algorithms impossible.

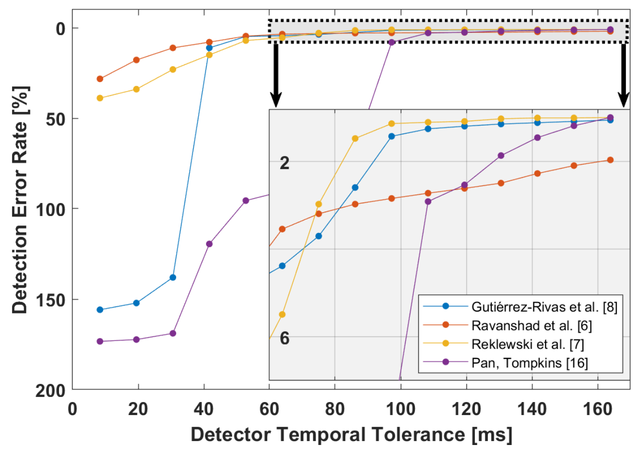

Figure 3. Comparison of evaluated QRS detection algorithms in terms of DER presented as a function of DTT. For a better visibility, the selected fragment of the plot is zoomed.

4. Conclusion

This paper addresses a problem of the sensitivity of the accuracy of QRS detection algorithms to the Detector Temporal Tolerance. The results achieved for three state-of-the-art low complexity algorithms in a broad range of DTT (from 160 ms to 8 ms) have been compared with each other and with the Pan-Tompkins algorithm which is a commonly accepted reference.

As the DTT value decreases, the classification statistics for R-peak detection algorithms deteriorate, while this deterioration rate is an individual characteristic of a given algorithm. Some algorithms experience graceful degradation, while others suffer an avalanche-like drop for particular DTT values. Moreover, the algorithms change positions in the detection accuracy ranking in various intervals of DTT.

The performed analyses proved that DTT is an integral parameter of QRS complex detection algorithms that determines the reproducibility of test results and fair comparative study.

Acknowledgments

This work was supported by the Polish National Center of Science under grant DEC- 2018/31/B/ST7/03874

Table 1. Comparison of algorithms performance ($+P$ [%], S [%], DER [%]) under increasing DTT expressed in [ms] and number of samples. The best method in terms of DER for particular DTT is bold.

DTT [ms] (samples)		[8]	[6]	[16]	[7]
8.33	+P	22.15	85.80	13.35	80.68
(3)	S	22.21	85.97	13.36	80.38
	DER	155.86	**28.25**	173.35	38.88
19.44	+P	23.99	91.01	13.86	83.13
(7)	S	24.06	91.19	13.87	82.82
	DER	152.16	**17.81**	172.34	33.98
30.56	+P	31.04	94.34	15.57	88.61
(11)	S	31.13	94.53	15.58	88.28
	DER	138.03	**11.14**	168.91	23.07
41.67	+P	94.31	95.92	40.26	92.65
(15)	S	94.58	96.11	40.29	92.30
	DER	11.13	**7.97**	119.50	15.02
52.78	+P	97.44	97.58	52.22	96.62
(19)	S	97.71	97.78	52.26	96.26
	DER	4.85	**4.65**	95.55	7.10
63.89	+P	97.68	98.14	54.78	97.43
(23)	S	97.95	98.33	54.82	97.07
	DER	4.38	**3.54**	90.44	5.49
75.00	+P	98.02	98.31	57.56	98.70
(27)	S	98.29	98.51	57.60	98.33
	DER	3.70	3.19	84.88	**2.97**
86.11	+P	98.57	98.42	66.38	99.45
(31)	S	98.84	98.61	66.43	99.08
	DER	2.59	2.97	67.22	**1.47**
97.22	+P	99.15	98.48	95.93	99.62
(35)	S	99.43	98.68	96.00	99.25
	DER	1.42	2.84	8.08	**1.13**
108.33	+P	99.24	98.55	98.51	99.63
(39)	S	99.52	98.74	98.58	99.26
	DER	1.25	2.72	2.91	**1.10**
119.44	+P	99.27	98.60	98.70	99.64
(43)	S	99.54	98.79	98.77	99.27
	DER	1.19	2.61	2.53	**1.08**
130.56	+P	99.29	98.66	99.03	99.68
(47)	S	99.57	98.86	99.11	99.30
	DER	1.14	2.49	1.86	**1.02**
141.67	+P	99.31	98.77	99.24	99.68
(51)	S	99.58	98.97	99.31	99.31
	DER	1.11	2.27	1.45	**1.00**
152.78	+P	99.32	98.86	99.37	99.69
(55)	S	99.60	99.06	99.45	99.32
	DER	1.08	2.09	1.18	**1.00**
163.89	+P	99.34	98.92	99.47	99.69
(59)	S	99.61	99.12	99.54	99.32
	DER	1.05	1.96	**0.99**	**0.99**

References

[1] F. M. Melgarejo-Meseguer *et al.*, "On the beat detection performance in long-term ECG monitoring scenarios," *Sensors*, vol. 18, no. 5, p. 1387, 2018.

[2] S. Gradl *et al.*, "Temporal correction of detected R-peaks in ECG signals: A crucial step to improve QRS detection algorithms," in *37th Annual Int. Conf. of the IEEE Eng. in Med. and Biol. Soc.*, 2015, pp. 522–525.

[3] I. I. Christov, "Real time electrocardiogram QRS detection using combined adaptive threshold," *Biomedical engineering online*, vol. 3, no. 1, pp. 1–9, 2004.

[4] B. U. Kohler, C. Hennig, and R. Orglmeister, "QRS detection using zero crossing counts," *Progress Biomed. Research*, vol. 8, no. 3, pp. 138–145, 2003.

[5] L. Smital *et al.*, "Robust QRS detection using combination of three independent methods," in *2020 Computing in Cardiology*, vol. 47, 2020.

[6] N. Ravanshad, H. Rezaee-Dehsorkh, R. Lotfi, and Y. Lian, "A level-crossing based QRS-detection algorithm for wearable ECG sensors," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 1, pp. 183–192, 2014.

[7] W. Reklewski, K. Heryan, M. Miśkowicz, and P. Augustyniak, "Real time ECG R-peak detection by extremum sampling," in *6th Int. Conf. on Event-Based Control, Commun. and Signal Process.*, Sept. 2020, pp. 1–7.

[8] R. Gutiérrez-Rivas, J. J. García, W. P. Marnane, and A. Hernández, "Novel real-time low-complexity QRS complex detector based on adaptive thresholding," *IEEE Sensors J.*, vol. 15, no. 10, pp. 6036–6043, 2015.

[9] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. 33, no. 12, pp. 1157–1165, 1986.

[10] G. Nallathambi and J. C. Principe, "Integrate and fire pulse train automaton for QRS detection," *IEEE Trans. Biomed. Eng*, vol. 61, no. 2, pp. 317–326, 2014.

[11] M. Elgendi, "Fast QRS detection with an optimized knowledge-based method: evaluation on 11 standard ECG databases," *PloS one*, vol. 8, no. 9, p. e73557, 2013.

[12] *ANSI/AAMI EC57: Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, AAMI Recommended Practice*, ANSI Std., 1998. [Online]. Available: http://www.aami.org

[13] P. Augustyniak, "Recovering the precise heart rate from sparsely sampled electrocardiograms," in *1999 Computers in Medicine*, vol. 59, 1999.

[14] Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[15] R. E. Gregg, R. Firoozabadi, and S. Babaeizadeh, "Can peak-picked ECG be used for heart rate variability?" in *2018 Computing in Cardiology*, vol. 45, 2018.

[16] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, pp. 230–236, 1985.

[17] L. Howell and B. Porr. (2019, June) A collection of 7 ECG heartbeat detection algorithms implemented in Python. Developed in conjunction with a new ECG database: http://researchdata.gla.ac.uk/716/. [Online]. Available: https://github.com/berndporr/py-ecg-detectors/blob/master/ecgdetectors.py

[18] M. Miśkowicz, "Reducing communication by event-triggered sampling," in *Event-based control and signal processing*. CRC Press, 2016, pp. 37–58.

[19] A. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[20] *The impact of the MIT-BIH Arrhythmia Database*, vol. 20, no. 3, May-June 2001. [Online]. Available: https://physionet.org/content/mitdb/1.0.0/

Address for correspondence:

Katarzyna Heryan

AGH-UST, al. Mickiewicza 30, 30-059 Krakow, POLAND

heryan@agh.edu.pl