QRS Detectors Performance Comparison in Public Databases

Mariano Llamedo, Juan Pablo Martínez

BSICoS, Aragon Inst of Eng Research, IIS Aragón, Univ of Zaragoza, Aragon, Spain
CIBER of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Spain

Abstract

Automatic QRS detection remains a challenging task in certain types of recordings, limiting the capacity of automating subsequent tasks that heavily depends on proper heartbeat location. Performance estimation of these algorithms is calculated almost exclusively in a few databases, ignoring the generalization to other more complex situations. In this work, we evaluated six QRS detection algorithms in 13 ECG databases. Four out of the six algorithms, and 11 out of 13 databases are publicly available. The databases were categorized into 5 groups: normal sinus rhythm, arrhythmia, ST and T morphology changes, stress-test and long-term. The best evaluated algorithm was gqrs*, achieving S of 95 (85-98) (median and percentile range 5-95) and* P^+ *of 93 (90-96) across all databases. When analyzing the performance by groups of databases, this algorithm obtained the first rank in 4 out of 5 groups. The algorithm developed in our group achieved a performance close to* gqrs*, and obtained the best performance in the stress group. This evaluation setup includes a broad variety of recordings, being useful to estimate the actual performance of QRS detection algorithms, not only in a global sense but also specific to specific type of recordings.*

1. Introduction

The analysis of electrocardiographic (ECG) signals provides a noninvasive and inexpensive technique to analyze the heart function for different cardiac conditions. One of the most frequent analysis performed in first place is the detection of heartbeats or QRS complexes, and the subsequent construction of the RR interval sequence. In the last decades big efforts were made to perform this analysis automatically, and as a result, many algorithms for QRS complex detection were published [?, ?, ?, 1, 2] and some of them are open-source or freely available [?, 2].

The performance achieved by these algorithms reported average sensitivities (S) and positive predictive values (P^+) well above the 90%, when evaluated in several public databases [?]. As was also discussed in [?], most of the algorithms presented were trained and evaluated in the same database, typically the MIT-BIH Arrhythmia database (mitdb) [2], this fact is well-known to optimistically bias the performance estimation. Despite the good performance reported in the works referenced in [?], low SNR recordings, e.g. stress-tests, long-term or arrhythmia, remain challenging scenarios for automatic algorithms. As a result, QRS locations calculated should be manually or automatically reviewed before subsequent processing, resulting in a trade-off between automaticity and performance.

The objective of this work is to develop an evaluation setup for automatic QRS detector algorithms, comprising several types of databases, in order to perform a broad and more realistic performance estimation.

2. Material and methods

In this work we used 13 ECG databases grouped in 5 categories: normal sinus rhythm (NSR), arrhythmia (AR), ST and T morphology changes (STT) stress (STR) and long-term (LT). Of all the databases used, 11 are publicly available online at [2] or [3] websites. With respect to the non-free databases, ahadb is distributed by ECRI institute [4], and biosigna is distributed by Biosigna GmbH [5]. All the databases has expert-reviewed QRS complexes locations, serving as gold-standard for the performance evaluation. Some details of the databases are summarized in Table 1.

In this work we evaluated a set of publicly-available algorithms representative of the state of the art, and the algorithms that were developed, or used in the past by our group. The algorithms evaluated in this work are summarized in Table 2.

The evaluation of each algorithm is performed lead-by-lead, and in multilead mode if the algorithm allows it. The configuration of each algorithm was with its default values, in order to recreate the actual performance that any user could achieve out-of-the-box. The performance is evaluated for each lead l, by means of the sensitivity

$$S_l = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Table 1. Databases characteristics

group	name	fs (Hz)	length	# rec	leads	ref
NSR	nsrdb	128	1 day	18	2	[2]
AR	ahadb	250	30 m	155	2	[4]
	biosigna	500	1 h	50	12	[5]
	mitdb	360	30 m	48	2	[2]
	svdb	128	30 m	78	2	[2]
	incartdb	257	30 m	75	12	[2]
STT	edb	250	2 h	90	2	[2]
	ltstdb	250	21-24 h	86	2-3	[2]
STR	thew15	1000	15 m	909	12	[3]
	stdb	360	10-40 m	28	2	[2]
LT	ltdb	128	14-22 h	7	2	[2]
	nsrdb	128	1 day	18	2	[2]
	ltstdb	250	21-24 h	86	2-3	[2]
	ltafdb	128	1 day	84	2	[2]

Table 2. Algorithms description

name	evaluated in	multilead	ref
wavedet	mitdb, edb, CSE, qtdb	yes	[1]
gqrs	N/A	no	[2]
sqrs	N/A	no	[2]
wqrs	N/A	no	[2]
pantom	mitdb	no	[**?**, **?**]
aristotle	mitdb	yes	[**?**]

and positive predictive value

$$P_l^+ = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}.$$

Then for each database in Table 1, the median performances were calculated pooling together all leads performances. Finally the median performances for the 5 groups of databases were calculated. This same procedure is repeated, but considering only the best performing lead for each recording. The best performing lead was selected based on the metric $q_l = 2S_l + P_l^+$, given that we preferred more sensitive detectors. The criterion to rank the algorithms consisted in estimating the lower performance achieved. This was done by calculating the q_l criterion, but using the 5th percentile across all the database, or groups of databases.

3. Results

The results obtained for all the evaluations are summarized in the following tables. Table 3 shows the three best performing algorithms for each database, while Table 4 presents the performance achieved for each group of databases. The *gqrs* algorithm was the best performing in NSR, AR and STT groups, while *wavedet* was in STR and LT groups.

4. Discussion and conclusions

In this work we presented an evaluation setup for QRS detection algorithms. This setup includes 13 databases, 11 of them publicly available online [2, 3], including ECG recordings from several types, such as normal sinus rhythm, arrhythmia, ST-T wave changes, stress test and long-term. This setup was used for the evaluation of 6 QRS detection algorithms, 4 of them publicly available [**?**, 2].

The results suggest that the best overall algorithm was achieved by *gqrs* [2], which obtained a S of 95 (85-98) (median and percentile range 5-95) and $P^+$of 93 (90-96) across all databases. When analyzing the performance by groups of databases, this algorithm obtained the first rank in NSR, AR and STT groups, as can be seen in Table 4. The *wavedet* algorithm achieved a slightly lower performance in those groups, but outperformed *gqrs* in STR and LT groups. These results suggest that although *gqrs* is the best overall detector, *wavedet* is more convenient for long term or stress-test recordings. The fact that *gqrs* and *wavedet* were among the best three algorithms for all the evaluated databases, suggest that these algorithms are the best performing in this evaluation set. The *gqrs* algorithm has the additional advantage of being faster than *wavedet*, however *wavedet* also provides the wave delineation of the ECG. The third performing detector was *aristotle*, which performed worst in biosigna, incartdb and edb, but was top 3 in the rest of databases, as is shown in Table 3. Other interesting result is the improvement achieved by selecting the best performing lead, as is shown in the right columns of Tables 3 and 4. If this could be achieved automatically, the performance could be improved to a S of 95 (85-98) and $P^+$of 93 (90-96) across all databases. The fact that there is not a clear winner for all groups of databases, and the improvement obtained by selecting the proper lead, motivated the development of quality detection metrics for selecting the proper algorithm and lead for any type of recording. An algorithm dealing with this problem was presented in this same conference by the same authors.

The evaluation setup presented in this work includes a broad variety of signals that QRS detection algorithms must deal with in real-world applications. This fact make it useful to better estimate the actual performance, and therefore track performance improvements, not only in a global sense, e.g. across all databases, but also specific to certain classes of recordings.

Acknowledgments

This work was supported by projects TEC2010-21703-C03-02, TEC2013-42140-R from MINECO (Spain) and by Grupo Consolidado BSICoS (T96) from DGA (Aragon) and European Social Fund (EU). The CIBER of Bioengineering, Biomaterials and Nanomedicine is an initiative of

Table 3. Best performing algorithms per database. Results expresed as median and percentile range 5-95

group	name	detector	All leads grouped together				Best lead only			
			S		P^+		S		P^+	
NSR	nsrdb	gqrs	86	72- 97	88	82-100	96	80- 97	92	82-100
		aristotle	88	77- 98	86	77- 96	96	82- 98	87	78- 99
		wavedet	86	73- 98	85	78- 93	97	83- 98	87	78- 98
AR	ahadb	pantom	100	92-100	64	36- 92	100	98-100	68	40- 92
		wavedet	100	63-100	64	12- 84	100	99-100	68	44- 93
		gqrs	100	71-100	64	20- 80	100	98-100	64	40- 86
	biosigna	sqrs	96	0- 97	95	36-100	97	92- 99	99	75-100
		wavedet	94	21- 98	81	33-100	97	74- 99	98	60-100
		wqrs	71	2- 97	82	32-100	97	85- 97	99	44-100
	mitdb	sqrs	97	94-100	100	78-100	97	96-100	100	84-100
		wavedet	95	26- 99	100	32-100	97	93-100	100	88-100
		pantom	96	30- 99	100	35-100	97	87-100	100	87-100
	svdb	gqrs	100	99-100	100	98-100	100	99-100	100	98-100
		wqrs	100	98-100	100	89-100	100	98-100	100	89-100
		wavedet	100	99-100	99	70-100	100	100-100	100	88-100
	incartdb	gqrs	92	57- 94	100	85-100	94	92- 94	100	99-100
		wqrs	92	57- 94	100	86-100	94	93- 96	100	98-100
		aristotle	93	59- 94	99	80-100	94	92- 95	100	98-100
STT	edb	sqrs	99	93-100	100	91-100	99	98-100	100	97-100
		pantom	99	79-100	100	90-100	99	98-100	100	96-100
		aristotle	99	77-100	100	83-100	99	97-100	100	85-100
	ltstdb	gqrs	96	81- 97	100	92-100	96	89- 97	100	93-100
		aristotle	96	80- 98	99	87-100	96	85- 98	99	88-100
		pantom	96	78- 97	100	90-100	96	85- 97	100	89-100
STR	thew15	wavedet	100	90-100	97	80-100	100	99-100	99	91-100
		gqrs	99	97-100	48	40- 59	100	99-100	49	41- 61
		wqrs	94	54-100	48	38- 61	99	95-100	46	38- 54
	stdb	gqrs	97	87-100	99	80-100	97	87-100	99	80-100
		wavedet	97	29-100	100	37-100	97	88-100	100	89-100
		sqrs	96	42-100	99	45-100	96	42-100	99	45-100
LT	ltdb	pantom	87	39- 95	94	40- 99	87	39- 95	94	40- 99
		gqrs	86	38- 97	95	40- 99	86	38- 97	95	40- 99
		wavedet	85	33- 95	95	40- 99	87	34- 95	95	40- 99
	nsrdb	gqrs	86	72- 97	88	82-100	96	80- 97	92	82-100
		aristotle	88	77- 98	86	77- 96	96	82- 98	87	78- 99
		wavedet	86	73- 98	85	78- 93	97	83- 98	87	78- 98
	ltstdb	gqrs	96	81- 97	100	92-100	96	89- 97	100	93-100
		aristotle	96	80- 98	99	87-100	96	85- 98	99	88-100
		pantom	96	78- 97	100	90-100	96	85- 97	100	89-100
	ltafdb	gqrs	75	33- 96	89	72-100	75	33- 96	89	72-100
		pantom	74	44- 94	89	62- 99	74	44- 94	89	62- 99
		wavedet	68	43- 96	85	59- 96	73	48- 96	88	71- 96

ISCIII.

References

[1] Martínez JP, Almeida R, Olmos S, Rocha A, Laguna P. A wavelet-based ECG delineator: Evaluation on standard databases. IEEE Transactions on Biomedical Engineering 2004;51:570–581.

[2] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, 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.

Table 4. Ranking of algorithms per group of databases. Results expresed as median and percentile range 5-95

group	detector	All leads grouped together				Best lead only			
		S		P^+		S		P^+	
NSR	gqrs	97	58- 99	99	82-100	99	51- 99	99	83-100
	wavedet	99	78-100	94	63-100	99	85-100	99	82-100
	pantom	98	76- 99	95	80-100	99	83- 99	99	79-100
	aristotle	98	80-100	89	64-100	99	86-100	98	57-100
	sqrs	93	60-100	90	60-100	99	73-100	99	64-100
	wqrs	88	48- 99	94	77-100	99	27- 99	99	78-100
AR	gqrs	93	32-100	99	60-100	99	83-100	98	56-100
	wavedet	94	32-100	96	60-100	100	93-100	96	60-100
	pantom	93	29-100	99	60-100	98	64-100	98	56-100
	aristotle	93	32-100	98	60-100	97	90-100	96	56-100
	sqrs	93	22-100	98	60-100	97	17-100	98	56-100
	wqrs	93	32-100	98	60-100	99	91-100	97	60-100
STT	gqrs	97	78- 99	100	94-100	97	92- 99	100	96-100
	wavedet	97	78-100	100	77-100	98	89-100	100	88-100
	pantom	97	77- 99	100	90-100	98	87- 99	100	93-100
	aristotle	97	79-100	100	84-100	98	90-100	100	87-100
	sqrs	97	83-100	100	88-100	97	89-100	100	94-100
	wqrs	97	71- 99	100	87-100	97	82- 99	100	90-100
STR	gqrs	99	97-100	48	40- 59	100	99-100	49	41- 65
	wavedet	100	89-100	97	80-100	100	99-100	99	91-100
	pantom	2	0-100	86	0-100	100	0-100	100	0-100
	aristotle	69	59- 80	51	41- 63	78	71- 89	57	47- 75
	sqrs	0	0- 1	33	0-100	0	0- 1	100	33-100
	wqrs	94	54-100	48	38- 61	99	94-100	46	38- 57
LT	gqrs	93	58- 97	97	79-100	94	54- 97	96	77-100
	wavedet	90	51- 97	93	71-100	93	56- 98	93	74-100
	pantom	89	56- 97	96	75-100	88	52- 97	95	73-100
	aristotle	91	52- 97	95	75-100	90	48- 98	94	73-100
	sqrs	87	51- 97	94	71-100	84	48- 97	91	69-100
	wqrs	85	54- 97	95	75-100	85	46- 97	93	71-100

[3] Couderc JP. The telemetric and holter ECG warehouse initiative (thew). URL `thew-project.org`.

[4] American Heart Association. American heart association ECG database, 2010. URL `https://www.ecri.org`.

[5] Fischer R, Sinner M, Petrovic R, Tarita E, Kääb S, Zywietz TK. Testing the quality of 12 lead holter analysis algorithms. In Computers in Cardiology, volume 35. 2008; 453–456.

Address for correspondence:

Mariano Llamedo Soria, llamedom@unizar.es
I3A - I+D Building, C/ Mariano Esquillor S/N
Despacho 4.0.05 – CP: 50018, Zaragoza, España.