Noise Resistance of Several Top-Scored Heart Beat Detectors

Marcus Vollmer

Institute of Bioinformatics, University Medicine Greifswald, Germany

Abstract

Background: *PhysioNet provides the MIT-BIH Noise Stress Test Database, which consists of only two different clean ECG recordings for which realistic noise signals were added to the ECG in different signal-to-noise ratios (SNR). For the evaluation of the noise resistance of multi-signal heart beat detectors this dataset is not sufficient.*

Methods: *Our stress test is based on 100 clean records from the training dataset (set-p) of the PhysioNet/CinC Challenge 2014. Each record is composed of a different number of several biosignals (ECG, respiration, blood pressure). Multi-signal heart beat detectors were able to identify the heart beats easily without a predictive loss. We added noise to the 10 minute records using the* `nst`*-function of the WFDB Toolbox with SNR ranging from 24 to -6 dB. Since noise was added only between the 5th and 7th and after the 9th minute, clear segments can serve as learning periods.*

Results: *The evaluation of the noise resistance of several top-scored multi-signal heart beat detectors includes the algorithms by Pangerc et al., Johnson et al., Hoog Antink et al., de Cooman et al. and Vollmer. The noise resistance is expressed as sensitivity and positive predictivity in noisy segments of the beat detection at different SNR levels. The most resistant algorithms were published by Hoog Antink and Vollmer. We are able to observe the effect of signal quality based constraints in the algorithms by Pangerc, Johnson and Vollmer, which mitigates a further decrease in the positive predictive value.*

1. Introduction

The electrocardiogram (ECG) and other frequently derived physiological signals are used as noninvasive medical procedures for diagnosis of diseases and for monitoring the status of health in and outside hospitals. However, the measurements are exposed to various external disturbances, specifically during the recording of photoplethysmogram, ECG and respiration. The reasons of noise and artifacts are diverse, e.g. poor contact between sensor and body, power line interference, physical activity or cable rupture to name but a few. These artifacts can cause many false alarms during continuous long-term monitoring and could annoy and mislead medical doctors in the intensive care unit for instance. To avoid such situations, well defined and robust methods can help to reduce noise or to improve the further processing of the raw data. Standard filter techniques are based on the frequency spectrum to suppress irrelevant low and high frequencies with the use of low-pass, high-pass or notch filters. But in general, there is still a large quantity of artifacts remaining. In this regard PhysioNet and Computing in Cardiology arranged a competition in 2014 to detect heart beats in a robust way by using multiple simultaneous measured signals [1]. In light of this, I am going to analyze the robustness against noise of selected challenge algorithms.

2. Challenge 2014 detectors

The challenge took place in three phases with five submitted entries at maximum per author and each entry was evaluated on a hidden test set. The organizers offered the chance to evaluate improved algorithms after the official challenge phase – so called post-challenge phase. The source code of the challenge participants are located on the public challenge 2014 website[1]. The top entries are original PhysioNet/CinC Challenge 2014 entries and evaluated on the revised hidden test set within the post-challenge phase. The 210 records used for external evaluation consists of $152,478$ heart beats and included roughly 5% abnormal beats (more detailed information are presented in the lead article by Silva et al. [1]). It turned out that Urška Pangerc et al. [2] ranked as best with an overall score of 93.64. The overall score is calculated by averaging gross sensitivity, gross positive predictivity, average sensitivity, and average positive predictivity [1]. With short distance, a cluster of well performing algorithms had an overall score of about 90. Table 1 is giving the ranked list of competitors with overall score and detailed results (sensitivity SE and positive predictive value +P) of the training sets set-p and set-p2.

However, the challenge performance statistics itself is not adequate to evaluate the applicability in real-time situations since the blank overall scores does not show the

---

[1]http://physionet.org/challenge/2014/sources/

Table 1. CinC/Physionet-Challenge 2014 post-phase results (revised hidden test set)

Rank	Version	Score	Reference	challenge/set-p	challenge/set-p2
1	urska.pangerc-420.zip	93.64	Pangerc et al. [2]	99.97/99.95	96.32/95.39
2	alistairewj-425.zip	91.50	Johnson et al. [3]	99.82/99.80	93.68/88.37
3	hoog.antink-407.zip	90.70	Hoog Antink et al. [4]	99.96/99.96	90.51/90.38
4	thomas.decooman-420.zip	90.02	de Cooman et al. [5]	99.86/99.92	87.56/85.75
5	lj-405.tar.gz	89.73	Galeotti et al. [6]	NA	NA
6	marcus.vollmer-402.zip	89.55	Vollmer [7]	99.97/99.99	91.80/91.34

weaknesses and strengths of the detectors in detail. Indeed, it is hard to compare the algorithms without knowledge of signal composition, source of the data and patient or test person information. Only some of the authors have given some performance statistics on public data bases for comparison. Nevertheless an independent black box evaluation is important, too.

3. Noise stress test

In this regard, I selected the top-ranked algorithms of Pangerc et al. [2], Johnson et al. [3], Hoog Antink et al. [4], de Cooman et al. [5] and Vollmer [7] to analyze the robustness against noise in detail. A Linux system was appropriately configured (Lenovo IdeaCentre K330, Intel Core i5-2300 CPU @ 2.80GHz, 6GB RAM: openSUSE 13.2, Matlab 2015b, Compiler gcc-4.7.4, WFDB Toolbox von Physionet wfdb-10.5.24) to make detectors executable which were programmed in C or C++. The algorithms of Hoog Antink und Vollmer were prepared to be evaluated on a Windows system (Fujitsu Celsius W280, Intel Core i7-860 CPU @ 2.80GHz, 8GB RAM: Windows 7 Professional, Matlab 2015b, Matlab Support Package for MinGW-w64 from TDM-GCC (GCC version 4.9.2)). It was necessary to modify the source codes in order to assign annotation files to the authors and to save large annotation files using `wrann`, which caused problems without modification[2].

PhysioNet provides a public noise stress test, the MIT-BIH Noise Stress Test Database [8](NSTDB) which was generated using two clean recordings (118 and 119) from the MIT-BIH Arrhythmia Database. Noise was added in two-minute segments to the records, beginning from the fifth minute of each record such that clear and noisy parts were alternating. The WFDB-Toolbox function [9] `nst` was used to pair and combine three noisy records with clean records. The provided noisy records contains baseline wander, electrode motion artifacts and muscular noise and were derived from an active measurement using a Holter recorder and standard electrodes. Before pairing and combining these three typically observed noisy parts with clean records the signal-to-noise ratio (SNR) within

noisy segments can be set. According to the WFDB manual [9], S is measured by the peak-to-peak QRS amplitude as the 5% winsorized mean of the first 300 normal QRS complexes. The squared value will be divided by 8 to get a feasible estimate of S. N is determined for unscaled noise signals by squaring the 5% winsorized root mean squared noise amplitude of one-second chunks of the first 300 seconds of the noise record. According to the desired SNR the weighting of the noisy and clean record has been adjusted according to the estimates of S and N. Therefore, a correct detection of heart beats in noisy segments is harder while increasing SNR. The resulting annotation files were compared to the reference annotation with respect to a tolerance of $150\,\mathrm{ms}$ (gold standard). Table 2 is giving the average SE and +P of both records (118 and 119) for the selected algorithms and are listed for SNR=6 and $0\,\mathrm{dB}$. Vollmer shows balanced values of SE and +P with the best +P values in all analyzed algorithms, while de Cooman shows more emphasis on SE.

Table 2. MIT-BIH noise stress test
Sensitivity and positive predictive values

Algorithm	SNR 6dB	SNR0dB
Pangerc	99.91/95.91	83.97/68.92
Johnson	31.26/37.90	31.26/37.63
Hoog Antink	84.49/76.40	72.20/66.37
de Cooman	99.47/73.30	96.51/59.36
Vollmer	98.50/96.73	77.10/74.91

Unfortunately NSTDB consists of two records only, such that the interpretative power of the results, as shown in Table 2, is very limited. To take more variety of ECGs and the composition of biosignals into account, I decided to extend this approach. I generated noisy records by applying `nst`[3] to the 100 relatively clean records of the set-p training database for which all the selected algorithms shows a nearly perfect heart beat detection, cf. Table 1.

Similarly to the NSTDB approach, I set the parameters of `nst` in such a way, that noise was added in standard settings to the 10-minute set-p records from the fifth

---

[2]for more details see https://github.com/ikarosilva/wfdb-app-toolbox/issues/52

[3]nst: noise stress test in version of the 6 October 2009, see http://www.physionet.org/physiotools/wfdb/app/nst.c for more details.
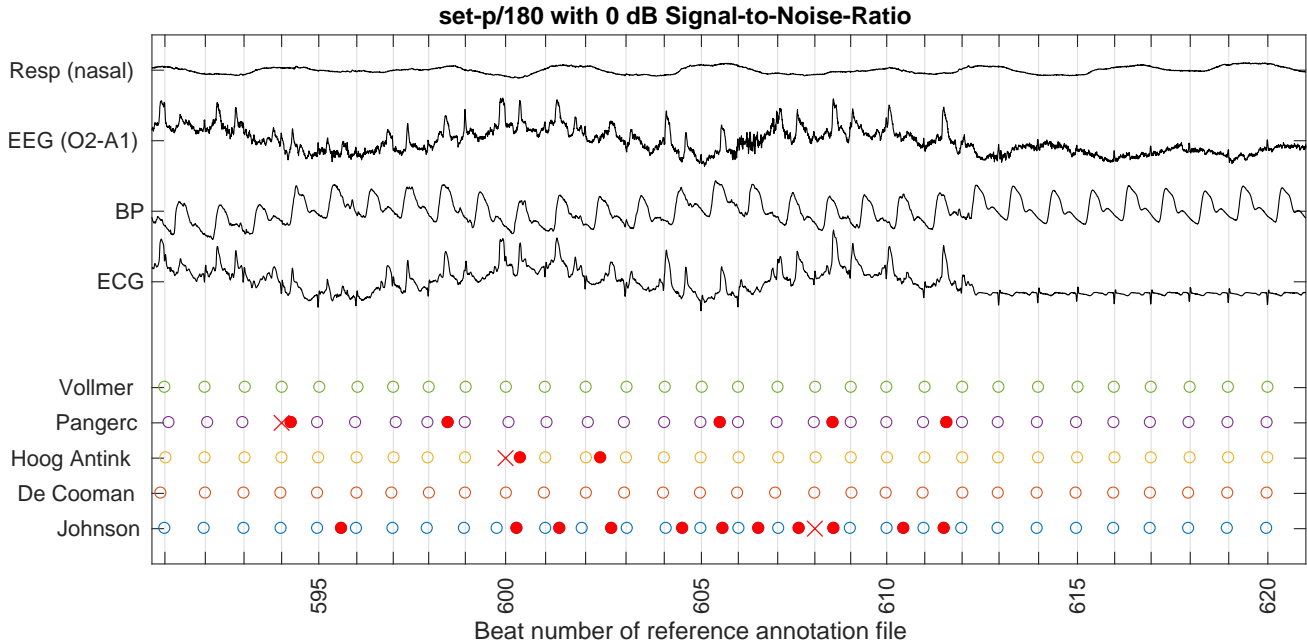
Figure 1. Transition area of noisy and clean parts of a record. The noisy segment ends in the seventh minute with beat number 612. The ticks are set according to the reference annotation file. The added noise causes a large number of false positive (red dots) and false negative annotations (red crosses).

to seventh minute and ninth to tenth minute. The clean segment till the fifth minute can be considered as a learning period to train the algorithms, although the algorithms were not prepared and planned to deal with such shifts. I set signal-to-noise ratios running from $24\,\mathrm{dB}$ to $-6\,\mathrm{dB}$ in order to compare SE and +P statistics in the course of the SNR range. To get a feeling of SNR values, I would like to mention that $10\,\mathrm{dB}$ corresponds to 10-fold stronger signal than noise (10:1), while $3\,\mathrm{dB}$ corresponds to a 2:1 relation and $-6\,\mathrm{dB}$ to a 1:4 relation. It it worth to figure out, how the algorithms perform in the transition areas and in noisy segments. Figure 1 shows such a transition area when applying `nsr` with SNR=$0\,\mathrm{dB}$ to record set-p/180. The heavy weight of noise until beat number 612 is visible when comparing the clean ECG signal to the noisy segment before. The baseline wander and many artifacts are visible. Although the blood pressure signal (BP) made a perfect annotation possible, only de Cooman and Vollmer are giving a correct annotation file.

4. Results

Next, I applied the selected algorithms on these modified set-p records with the range of values for SNR. All annotation files were evaluated by comparing the reference beats within the noisy segments to the annotations given by the detectors. The number of false positive and false negative beats were stored in spreadsheets and gross sensitivity

and gross positive predictivity was calculated. The course of noise resistance are shown in Figure 2 for each detector and leaves more room for interpretation. Remarkable reliable are the annotations of Hoog Antink und Vollmer. The SE and +P statistics sticks at almost $100\,\%$ even until $0\,\mathrm{dB}$, whereas in the case of Pangerc the sensitivity is massively decreasing, starting already at $12\,\mathrm{dB}$ until SE=$85\,\%$ has been reached at $-6\,\mathrm{dB}$. At the same time the +P value is decreasing as well, stops at $2\,\mathrm{dB}$ and increases interestingly until $-3\,\mathrm{dB}$. Johnson's +P takes a similar course, but has its minimum of 76% at $5\,\mathrm{dB}$. That means that the algorithms of Pangerc and Johnson are annotating many artifacts (false-positives) until a threshold has been reached. A higher SNR will result in the behavior that both algorithms will generally detect less heart beats, such that the +P is going to increase again at the costs of SE. As a reason it is to note that the "noise detection function" g_n, as implemented in the detector by Pangerc, rejects possible heart beats in noisy segments (see [10]). Responsible in the detector by Johnson et al. is the signal quality based fusion procedure (see 2.3.1 in [3]) which lowers the acceptance of heart beats when the estimated signal quality drops down. Vollmer [7] has used some signal quality index based on the regularity of RR interval series and an annotation threshold based on the difference between a smoothed windowed maximum and smoothed windowed minimum. The effect as seen in Figure 2 results in both, SE and +P tends to remain static as of $-2\,\mathrm{dB}$.
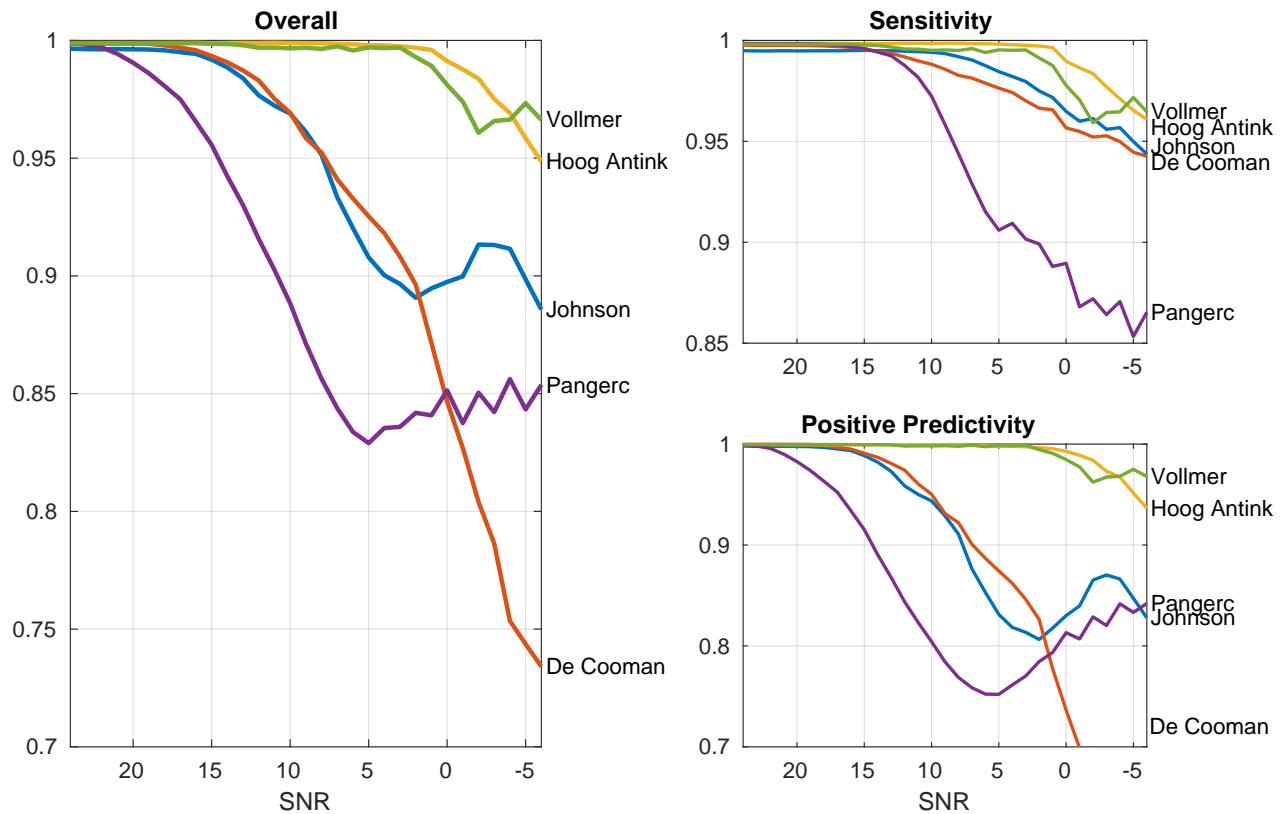
Figure 2. Noise resistance of several heart beat detectors participated in the PhysioNet/CinC Challenge 2014 evaluated on noisy set-p records at different signal-to-noise ratios. Left: overall score. Right: average sensitivity and positive predictivity within the noisy segments.

Summary A new noise stress test database was generated by applying the WFDB function `nsr` to records of the training data set set-p. The most resistant algorithms were published by Hoog Antink and Vollmer. Nevertheless the noise added by `nsr` should be applied to ECG signals only and does not add realistic noise to other signals such as BP. The results as shown in this paper is therefore limited to the robustness against ECG noise in multimodal data.

References

[1] Silva I, Moody B, Behar J, Johnson A, Oster J, Clifford GD, Moody GB. Robust detection of heart beats in multimodal data. Physiological Measurement 2015;36(8):1629.

[2] Pangerc U, Jager F. Robust detection of heart beats in multimodal records using slope- and peak-sensitive band-pass filters. Physiological Measurement 2015;36(8):1645.

[3] Johnson AEW, Behar J, Andreotti F, Clifford GD, Oster J. Multimodal heart beat detection using signal quality indices. Physiological Measurement 2015;36(8):1665.

[4] Hoog Antink C, Brüser C, Leonhardt S. Detection of heart beats in multimodal data: a robust beat-to-beat interval estimation approach. Physiological Measurement 2015; 36(8):1679.

[5] de Cooman T, Goovaerts G, Varon C, Widjaja D, Willemen T, van Huffel S. Heart beat detection in multimodal data using automatic relevant signal detection. Physiological Measurement 2015;36(8):1691.

[6] Galeotti L, Scully CG, Vicente J, Johannesen L, Strauss DG. Robust algorithm to locate heart beats from multiple physiological waveforms by individual signal detector voting. Physiological Measurement 2015;36(8):1705.

[7] Vollmer M. Robust Detection of Heart Beats using Dynamic Thresholds and Moving Windows. In Computing in Cardiology. 2014; 569–572.

[8] Moody GB, Muldrow W, Mark RG. A noise stress test for arrhythmia detectors. Computers in Cardiology 1984; 11(3):381–384.

[9] Moody GB. WFDB applications guide. Harvard-MIT Division of Health Sciences and Technology, 10 edition, 2003.

[10] Pangerc U, Jager F. Robust detection of heart beats in multimodal data using integer multiplier digital filters and morphological algorithms. In Computing in Cardiology. 2014; 285–288.

Address for correspondence:

Marcus Vollmer / marcus.vollmer@uni-greifswald.de

Institute of Bioinformatics / University Medicine Greifswald

Walther-Rathenau-Str. 48 / 17475 Greifswald / Germany