Assessing Effect of Beat Detector on Detection Dependent Signal Quality Indices

Chathuri Daluwatte[1], Lars Johannesen[2], Loriano Galeotti[1,3], Jose Vicente[1,2,4], David G Strauss[2], Christopher G Scully[1]

[1]Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, US FDA, Silver Spring, MD, USA
[2]Office of Clinical Pharmacology, Office of Translational Sciences, Center for Drug Evaluation and Research, US FDA, Silver Spring, MD, USA
[3]Office of Device Evaluation, Center for Devices and Radiological Health, US FDA, Silver Spring, MD, USA
[4]BSICoS Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, Zaragoza, Spain

Abstract

Patient monitoring algorithms which use multimodal physiological waveforms are needed to reduce alarm fatigue by alarming only for physiologic events and not signal artifacts. When combining information from multiple ECG signals, computational approaches that automatically identify artifacts in ECG signals play an important role. Signal quality indices (SQIs) have been derived which can differentiate between ECG signal artifacts and normal QRS morphology. Some of these SQIs are derived using beat detections and might be affected by the beat detector used. Using ECG signals from the PhysioNet/Computing in Cardiology Challenge 2015 training set, we studied the effect of beat detector on previously reported ECG SQIs derived using beat detections. We found that, while being affected by the beat detector, some of these SQIs can predict beat detector failure. Using beat detector specific SQIs can improve the designs of robust monitoring algorithms.

1. Introduction

Alarm fatigue refers to the high number of clinically irrelevant alarms in Intensive Care Units (ICU). Alarm fatigue is a top medical device hazard [1] and it has been shown, in one study, that up to 88.8% of arrhythmia alarms in the ICU are false [2]. Alarm fatigue disturbs patient rest, is burdensome to the caregiver staff, and with desensitization puts patients at risk with delayed reaction times from caregivers [3]. Inappropriate user settings, patient conditions, noise and motion artifacts, and algorithm performance have been identified as factors which contribute to alarm fatigue [4]. Clinicians use complementary information available on physiological signals from different monitors to recognize false alarms. Similar approaches can be used in automatic algorithms to reduce false alarms. For example, the Computing in Cardiology (CinC) 2015 challenge focused specifically on reducing false arrhythmia alarms in the ICU using patient monitoring algorithms which use multimodal physiological waveforms [5]. However combining information from multiple physiological signals introduces a new potential risk of adding noise artifacts from a low quality signal onto information from a high quality signal (e.g. During an asystole episode, one ECG signal will have no beats detected due to asystole while another noisy ECG can have erroneous beat detections and combining these beats will cause the monitoring algorithm to miss the true asystole episode). Therefore computational approaches that automatically identify artifacts in ECG signals play an important role in multimodal physiological monitoring algorithm development. Signal quality indices (SQIs) have been derived which can differentiate between artifacts which occur in ECG signals and normal QRS morphology. Some of these SQIs are derived using beat detections and might have an effect from the beat detector used. Using ECG signals from the PhysioNet/Computing in Cardiology Challenge 2015 training set, we studied the effect of beat detector on previously reported ECG SQIs which are derived using beat detections. We characterized the distributions of SQIs to assess if previously reported SQIs derived using beat detections 1) are affected by the beat detector used and 2) can predict beat detector performance.

2. Methods

2.1. Dataset

For our analysis on ECG SQIs we used the ECG signals in the CinC 2015 training set [5]. We focused our analysis on asystole, bradycardia, and tachycardia records since ventricular tachycardia and ventricular fibrillation/flutter result in extreme modification of the ECG waveform. For these alarms, three reviewers annotated beat locations during the 10 s periods: during alarm and immediately before the alarm. The reference annotations were generated if two out of three reviewers marked annotations within ±150 ms [6]. We pooled data from these three alarm types to get an enriched dataset with a wide *mean heart rate* range where median, 2.5th and 97.5th percentile *mean heart rate* was 98, 29 and 162 beats per minute (bpm) respectively.

During the annotation process some of the signals were identified to contain pacemaker pulses and since pre-processing stages specific for detection and removal of pacemaker pulses are available we excluded these records from our analysis. This resulted in 648 total ECG signals used from the dataset providing 1296 ECG epochs corresponding to during alarm and immediately before the alarm periods.

2.2. Selection of signal quality indices and beat detectors

SQIs that classify ECG signals as either high or low quality and derived using beat detections were identified from the literature for this study. These were selected based on the information available to implement the computational method on 10 s single-lead ECG signals. The list of SQIs selected for this study is summarized in table 1. In the original literature, 10 s epochs were used to calculate each of these SQIs from the ECG.

We selected a set of beat detector algorithms with open source implementation to detect beats in the ECG epochs we annotated. This set consisted of the Zong [7], Afonso [8], Pan [9], Hamilton [10], and Johannesen [11] beat detectors. Selected SQIs were calculated using beats detected by each of these detectors and reference beat annotations. The F1 score to detect beats for two 10 s periods with respect to reference beat annotations was calculated as $\frac{2TP}{2TP+FN+FP}$ [15] (TP [true positive] is the number of beats which are within ±150 ms of reference beat annotations, FP [false positive] is the number of beats the detector incorrectly identified which does not have a reference beat annotation within ±150 ms, and FN [false negative] is the number of reference beat annotations which does not have beat detection within ±150 ms). For periods where the number of reference annotations and number beats detected by a detector is both zero, F1 score was set to 100%. Signal analysis was performed in Matlab R2014b (The Mathworks, Natick, MA).

Table 1: Implemented Signal Quality Indices

SQI	Feature
meanhr [12]	Mean heart rate for 10 s
maxrri [12]	Maximum RR interval for 10 s
maxrr2minrr [12]	Ratio of maximum RR interval to minimum RR interval for 10 s
avecorr [12]	Average template matching correlation coefficient: Average of the correlation coefficients of each QRS complex with mean QRS complex
qrsa [14]	Median value of the peak-to-nadir amplitude difference of the QRS complexes in 10 s
sdrr2meanrr [13]	Ratio of the standard deviation of RR interval to mean RR interval
rangeqrs [13]	Range of signal amplitude around QRS detection: Maximum minus the minimum signal amplitude within a QRS complex
bw [11]	Baseline wander estimation using cubic spline
pln [11]	Power line noise estimation using regression-subtraction
residual [11]	Residual noise by subtracting the estimated signal (median over 10 s) after subtracting baseline wander and power line noise

(The Mathworks, Natick, MA).

2.4. Statistical Analysis

The hypothesis that an SQI calculated using different beat detectors would have the same distribution was tested using Kruskal-Wallis test followed by post-hoc Dunn's test with Bonferroni correction for multiple comparisons.

An SQI might have a different distribution based on the beat detector used but still be able to predict the detector performance. We studied the ability of these SQIs to predict beat detector performance, using area under the receiver operating characteristic curve (AUC) to discriminate epochs which have "F1 score ≤ 90%" (considered poor detector performance) vs. "F1 score > 90%" (considered acceptable detector performance). All statistical calculations were conducted in R version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria).

3. Results

All SQIs had $p < 0.001$ in Kruskal-Wallis test suggesting at least one SQI had a different distribution

due to the beat detector used. Table 2 summarized the Bonferroni corrected Dunn's test *p*-value when compared to the SQI calculated using reference annotations. Signal quality indices *minimum RR interval, maximum RR interval to minimum RR interval ratio, standard deviation of RR interval to mean RR interval ratio* and *average template matching correlation coefficient* had Bonferroni corrected Dunn's test $p < 0.05$ across all beat detectors indicating that they have different distributions compared to the distribution derived using reference annotations depending on the beat detector used. For all other SQIs whether they had a different distribution than that from reference depended on the beat detector.

The AUC values for the SQIs to classify into groups "F1 score ≤ 90%" vs. "F1 score > 90%" are reported in table 3. *Average template matching correlation coefficient* had one of the three highest AUC values for each detector with an AUC closer to 0.9 except for Afonso [8] for which the AUC = 0.7. Figure 1 shows distributions of F1 score for each detector and distribution of *average template matching correlation coefficient* for each group: "F1 score ≤ 90%" and "F1 score > 90%" for each beat detector.

4. Discussion

Signal quality indices derived using beat detections have the potential to have different distributions depending on the beat detector used which might affect their ability to discriminate high quality ECG segments from those of low quality. We studied which SQI have beat detector dependent distributions and whether they can still differentiate ECG segments where the beat detector had superior beat detection performance in an enriched dataset with arrhythmia and wide range of *mean heart rate*.

Although distribution of *average template matching correlation coefficient* depends on the beat detector used, it can consistently discriminate between performance groups for four out of five detectors. For detectors Johannesen [11], Hamilton [10] and Pan [9], SQIs, *maximum RR interval to minimum RR interval ratio* and *standard deviation of RR interval to mean RR interval ratio* also discriminated between the two performance groups (AUC > 0.7). Therefore when these beat detectors are used in monitoring systems, the SQIs, *average template matching correlation coefficient, maximum RR interval to minimum RR interval ratio* and *standard deviation of RR interval to mean RR interval ratio* may have utility to identify high quality ECG epochs. Similarly for detector Zong [7], the combination of SQI which can best detect detector failure is *average template matching correlation coefficient, mean heart rate* and *minimum RR interval.* The highest AUC value for the detector Afonso [8] is 0.72 which suggests that none of the studied SQI can successfully predict the detector failure. The detector Afonso [8] also has an F1 score distribution showing inferior performance than the other

Table 2: Bonferroni corrected Dunn's test *p*-value when SQI calculated using each beat detector is compared to the SQI calculated using reference annotations.

SQI	Afonso	Hamilton	Johannesen	Pan	Zong
meanhr	0.062	*	1.0	*	*
sdrr	*	*	0.001	*	1.0
sdrr2meanrr	*	*	*	*	*
maxrri	*	1.0	0.715	1.0	*
minrri	*	*	*	*	*
maxrr2minrr	*	*	*	*	*
avecorr	*	*	*	*	*
rangeqrs	0.002	0.002	*	0.017	1.0
qrsa	1.0	0.952	1.0	*	*
bw	*	0.049	0.210	1.0	*
pln	*	0.036	0.032	0.529	0.103
resn	*	0.001	1.0	0.003	*

*: $p < 0.001$, shaded: $p > 0.05$

Table 3: AUC values for SQIs for discrimination between groups "F1 score ≤ 90%" vs. "F1 score > 90%". Three highest AUC values for each detector (column) are shaded.

SQI	Afonso	Hamilton	Johannesen	Pan	Zong
meanhr	0.66	0.80	0.50	0.66	0.83
sdrr	0.66	0.77	0.74	0.76	0.60
sdrr2meanrr	0.71	0.87	0.77	0.82	0.72
maxrri	0.52	0.63	0.68	0.67	0.57
minrri	0.71	0.83	0.67	0.78	0.90
maxrr2minrr	0.72	0.86	0.75	0.81	0.77
avecorr	0.69	0.89	0.87	0.87	0.95
rangeqrs	0.51	0.53	0.54	0.53	0.56
qrsa	0.50	0.54	0.49	0.52	0.57
bw	0.60	0.60	0.59	0.58	0.64
pln	0.61	0.70	0.65	0.67	0.68
resn	0.65	0.70	0.64	0.65	0.63

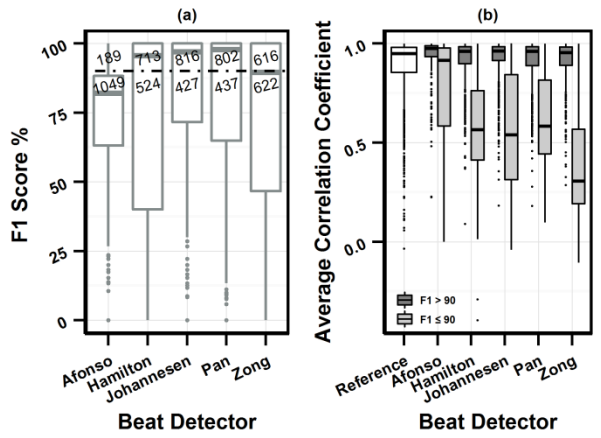

Fig. 1: (a) F1 score distribution for each beat detector. The number of ECG epochs in each group: "F1 score ≤ 90%" and "F1 score > 90%" for each detector is shown on the figure. (b) Distribution of *average template matching correlation coefficient* (*avecorr*) calculated using reference annotations and for each group "F1 score ≤ 90%" and "F1 score > 90%" for each beat detector.

four detectors in this dataset (figure 1a).

The ability of SQIs to predict detector failure is detector dependent, since the method used to detect heart beats by each detector is different. We found this to be true in our study as the best set of SQIs changed with the detector used. Therefore choosing SQIs based on the beat detector, or assessing the best SQI for the selected beat detector, could improve the design of robust monitoring systems.

Acknowledgements

This project was supported in part by U.S. Food and Drug Administration's Medical Countermeasures Initiative, Critical Path Initiative, Office of Women's Health and appointments to the Research Participation Programs at the Oak Ridge Institute for Science and Education through an interagency agreement between the Department of Energy and FDA.

Disclaimer

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

References

[1] ECRI Institute,Top 10 Health Technology Hazards for 2016. Health Devices 2015.

[2] Drew BJ, et al. Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients. PLoS ONE 2014; 9(10): e110274.

[3] Schmid F, et al. Patient monitoring alarms in the ICU and in the operating room. Crit Care. 2013;. 17: 216

[4] Borowski M, et al. Medical device alarms Biomedizinische Technik. 2011; 56:73-83.

[5] Clifford G D, et al. The PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU Comput Cardiol.. 2015; 273-6

[6] ANSI/AAMI EC57: 2012 Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms

[7] Zong W, et al. A robust open-source algorithm to detect onset and duration of QRS complexes. Comput Cardiol 2003;30:737-40.

[8] Afonso, V, et al. ECG beat detection using filter banks. IEEE Trans. Biomed. Eng. 1999; 46(2):192-202.

[9] Pan J, et al. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985;32:230-6.

[10] Hamilton PS, et al. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. IEEE Trans Biomed Eng 1986;33:1157-65.

[11] Johannesen L, et al. Ecglib: library for processing electrocardiograms. Comput Cardiol. 2013; 951-4.

[12] Orphanidou C, et al. Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. IEEE J Biomed Health Inform. 2015; 19:832-8.

[13] Hayn D, et al. QRS detection based ECG quality assessment. Physiol Meas. 2012; 33:1449-61.

[14] Di Marco LY, et al. Evaluation of an algorithm based on single-condition decision rules for binary classification of 12-lead ambulatory ECG recording quality. Physiol Meas. 2012; 33:1435-48.

[15] Sasaki Y. The truth of the F-measure. Technical Report University of Manchester, School of Computer Science 2007.

Address for correspondence:
Chathuri Daluwatte
10903 New Hampshire Avenue
Silver Spring, Maryland, 20993
USA
chathuri.daluwatte@fda.hhs.gov