# Suppression of False Arrhythmia Alarms Using ECG and Pulsatile Waveforms

Paula Couto[1,2], Ruben Ramalho[3], Rui Rodrigues[1,2]

[1] Centro de Matemática e Aplicações (CMA), FCT, UNL,Portugal
[2] Departamento de Matemática, FCT, UNL
[3] Faculdade de Ciencias e Tecnologia da Universidade Nova de Lisboa, Portugal

## Abstract

*In the context of the 2015 PhysioNet/CinC Challenge we present an algorithm to detect false arrhythmia alarms in the Intensive Care Unit(ICU). Our focus is on life threatening arrythmia alarms: asystole, extreme bradycardia, extreme tachycardia, ventricular tachycardia and ventricular fibrillation or flutter.*

*Our method uses simultaneous ECG and pulsatile waveforms, photoplethysmogram or arterial blood pressure, to detect false alarms. QRS detectors produce for each signal a set of QRS detections that can be used to detect false alarms. Often the conclusions drawn from the results of QRS detectors on different signals are contradictory: some of the signals may be contaminated by noise or simply get lost for a while. Evaluating the signal quality of each waveform is necessary to can decide if we can trust the QRS detections obtained on that waveform. We describe a method to choose in each case which set of QRS detections should be used to conclude if the alarm is true or not. A set of rules is used for each alarm type.*

*The average score of our best entry in the Challenge was 78.65.*

## 1. Introduction

In the context of the 2015 PhysioNet/CinC Challenge we developed an algorithm to detect false arrhythmia alarms in the Intensive Care Unit(ICU). For a description of the challenge and the datasets used we refer to [1]. Given a record containing two EGC leads and one or two pulsatile signals, plethysmogram(PLETH) and arterial blood pressure(ABP), we must decide if an alarm triggered 5 minutes after the beginning of the record should be supressed.

Much effort as been dedicated to the problem of lowering false alarms and detection of signal corruption, see the introduction of [1] for references and further details.

## 2. Method Description

### 2.1. Signal Preprocessing

Each signal in the records was downsampled to 125 Hz.

In some records ECG channels were contaminated by pacemaker spikes. That noise can produce false QRS detections. We identified pacemaker spikes in the ECG comparing each sample with the average of the preceding and following sample. When a sequence of one or more ECG samples is associated with a pacemaker spike those values are replaced by a signal interpolation.

Baseline wander can be a liability, it is thus removed from all channels. An estimate of the baseline is obtained by applying a median filter with width of 125 samples, the same as the sampling frequency. We then correct the wandering by subtracting the baseline estimate from the signal.

NaNs in ECG were replaced by an arbitrary value, -5.

### 2.2. QRS detection on the different channels

In the two ECG channels of each record we always apply the open source QRS detectors 'gqrs' [2] and 'eplimited' [3], authored respectively by George Moody and Patrick Hamilton. Additionally, in the case of Ventricular fibrillation we also apply our own detector 'fibrillation peaks'. The 'eplimited' detector classifies beats as either Normal or Ventricular. If the alarm is Ventricular Tachycardia we also use a subset of all eplimited detections: those identified as ventricular beats. QRS detection in the pulsatile channels is done exactly in the same way for PLETH and ABP, we apply the open source detector 'wabp' [4] after a standard signal scaling. For each detector and each channel the QRS locations (samples) are grouped in a set of annotations. So for each ECG channel we get two or three annotation sets(depending on the alarm type), and for each pulsatile channel we get one set of annotations.

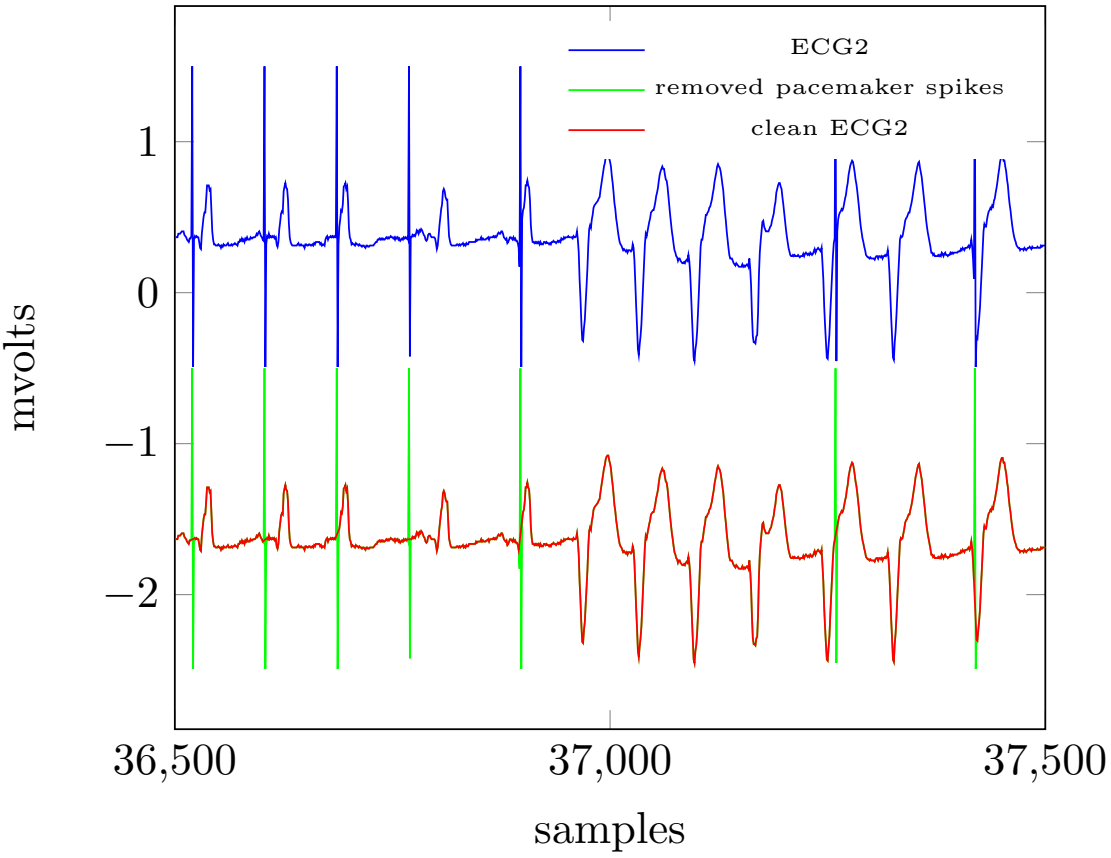


Figure 1. Pacemaker spikes removal in record v842s (2nd channel ECG)

## 3. Choosing the best set of QRS detections

### 3.1. Critical time

For each alarm type there is a time segment that we call *critical time*. It finishes 5 minutes after the beginning of the record(alarm time) and lasts 14 seconds for asystole, 10 seconds for bradycardia, 17.3 seconds for tachycardia, 10 seconds for ventricular tachycardia and 14 seconds for ventricular flutter/fibrilation. The algorithm's decision is based on what happens during this time interval.

### 3.2. Initial Quality Index assigned to each channel

In this work we have little interest on the evaluation of the general clinical utility of the signals. Neither we are interested in evaluating the possibility to estimate the QT interval in the ECG nor we give much importance to extract reliable blood pressure values from ABP signal. Instead, we aim only to measure the possibility to extract reliable QRS locations from the different channels. A measure of signal quality enables us to rank the different signals in one record: if we assign quality indexes x and y, with x>y, to channels A and B respectively, we expect that channel A gives better QRS detections than channel B. Given that we are dealing with different types of signals like ECG, ABP and PLETH, deriving such a quality index is a very challenging task.

Next we describe the signal quality index that we assigned to the different types of signals.

#### 3.2.1. ECG

We use the bSQI quality index[5] with the two aforesaid QRS detectors. We consider the two sets of QRS locations given by 'gqrs' and 'eplimited' contained in the critical time segment when applied to one ECG channel. The quality index of that channel is the fraction of matched locations divided by the total number of QRS locations. This quality index is thus a number in the zero to one interval. Zero if there are no matches between both sets and one if the locations of the two sets perfectly match. When we match locations, a tolerance of 150 milliseconds is set.

### 3.2.2. PLETH and ABP

Signal quality defined in the same way for both types of signals. We follow the method used in [6] exclusively for PLETH but we made some simplifications and adaptations. As a first step the signal is scaled in a standard way. After applying the detector wabp, we call 'wave' the signal segment limited by two detections. A wave template was created using the all record. Next, we computed for each wave two coefficients. The first is the simple correlation coefficient between the template and the present wave, the second is dynamic time warping distance between the template and the present wave (see [7]). This two coefficients are then averaged and a wave quality index is assigned to each wave. Finally, the quality index of a pulsatile channel is the average wave quality indexes during the critical time segment.

## 3.3. Final trust assignment to each set of annotations

Fusion of annotations from different detectors and channels was an option, tough we don't rely on any annotation fusion, except for the case when annotations perfectly match during critical time. The initial trust we assign to each set of annotations is preset as the quality index of the corresponding channel. When two sets of QRS annotations from different channels coincide in the *critical time* the signal quality from the two channels add to increase the trust on both annotation sets. For example, if one set of annotations from some ECG channel completely matches the annotations from the PLETH, the trust we assign to those annotations will be the sum of the quality indexes of that ECG channel and the PLETH.

## 4. A set of rules for each alarm type

Given a record associated with a particular alarm type and each set of annotations a conclusion, or result, is implicit about whether the alarm should be supressed or not. For example, if the alarm is asystole and there is an interval of more than 4 seconds between consecutive QRS detections in critical time we conclude that it is a true alarm. Given the different sets of annotations associated to the different channels and different detectors, contradictory conclusions may occur and a decision as to be taken. For each alarm type we applied a set of rules that uses the result and the trust assign to each set of detections to arrive to a final conclusion about the supression of the alarm.

For all alarm types we use a constant called *trust threshold*. For asystole and ventricular tachycardia the value of *trust threshold* is 0.82 and for the other alarm types the value is 0.9.

Next we describe the different sets of rules we used.

*Asystole, Bradycardia and Tachycardia:* If the highest trust value is smaller than the trust threshold, we do not take into account the results of the different annotations, and we also don't suppress the alarm. Otherwise, we take the result of the annotations set with the best trust index if it is comes from a pulsatile channel. In the case the highest trust value is above the trust threshold but it is associated with an ECG channel we only suppress the alarm if the annotation of both detectors on that channel indicate it is a false alarm.

*Ventricular Tachycardia:* If the trust of the best annotation set is above the trust threshold and the result is negative(false alarm) then we conclude that the alarm should be suppressed. If the result for that set of annotations is positive, the alarm should be suppressed only if the results of the two sets of annotations that use only ventricular beats are negative. If the trust of the best annotation set is smaller than the trust threshold, the alarm should not be suppressed.

*Ventricular fibrillation or flutter:* If the highest trust value is smaller than the trust threshold, we suppress the alarm if the detector 'fibrilation peaks' gives negative result on both ECG channels. Otherwise, we do not suppress the alarm. In the case the highest trust value is above the trust threshold, and the annotation set is from a pulsatile channel, the final result is the one of that annotation set. If the highest trust value is above the trust threshold, and the annotation set is from an ECG channel, we suppress the alarm only if both gqrs and eplimited give a negative result on that channel.

## 5. Results

The global results of the Top-scoring entries(Plešinger et al.) in the Unofficial Phase are given in table 1. The score function is given by the formula

$$score = \frac{TP + TN}{TP + TN + FP + 5 * FN}$$

where TP and FP are algorithm true and false positives respectively and TN and FN are algorithm true and false negatives. We refer to [1] for further details on the challenge competition namely for 'real time' and 'retrospective' categories. Our algorithm did not make any distinction between the records of those two categories.

Table 1. Top-scoring entry in the unofficial phase, global result

| | Real-time | Retrospective |
|---|---|---|
| Score | 76.6 | 78.5 |

The results of our top-scoring entry in the Unofficial Phase are given in table 2. TPR means true positive rate,

algorithm correct positives divided by all algorithm positive answers. TNR means true negative rate, algorithm correct negative answers divided by all algorithm negative answers.

Table 2. Results of our Top-scoring entry in the unofficial phase

| Alarm | TPR | TNR | Score |
|---|---|---|---|
| Asystole | 72% | 94% | 81.14 |
| Bradycardia | 90% | 78% | 70.80 |
| Tachycardia | 90% | 60% | 63.23 |
| Ventricular Flutter Fib | 44% | 76% | 52.56 |
| Ventricular Tachycardia | 50% | 70% | 46.67 |
| | | | |
| Real Time | 75% | 76% | 57.93 |
| Retrospective | 76% | 81% | 60.0 |

In the official phase we made a clear improvement as it is shown by the results of our best-scoring entry(table 3).

Table 3. Results of our Top-scoring entry in the official phase

| Alarm | TPR | TNR | Score |
|---|---|---|---|
| Asystole | 78% | 94% | 83.63 |
| Bradycardia | 95% | 66% | 71.43 |
| Tachycardia | 100% | 80% | 99.10 |
| Ventricular Flutter Fib | 89% | 96% | 88.71 |
| Ventricular Tachycardia | 69% | 95% | 71.67 |
| | | | |
| Real Time | 89% | 91% | 79.02 |
| Retrospective | 88% | 92% | 78.28 |

## 5.1. Discussion

We described an algorithm to suppress false arrhythmia alarms in the ICU. Improvements should be made namely to avoid the suppression of real alarms. The public availability of medical records containing ECG, PLETH, ABP and other signals is fundamental for the development of algorithms to reduce the number of false alarms in the ICU. Physionet as been doing an important work in that direction [8, 9] and this Challenge provided a set of data annotated by specialist that is invaluable. In the case of ventricular tachycardia there was an important lack of publicly ECG signals containing this type episode and annotated by specialists. That problem is mitigated with the training data of this challenge. The beat classifier we used, eplimited, could be tuned to improve the detection of ventricular beats in the context of ventricular tachycardia, that would reduce the rate of false negatives which was not very significant in test records of this alarm type but was higher in the training set.

## Acknowledgements

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00297/2013 (Centro de Matemática e Aplicações).

## References

[1] Clifford GD, Silva I, Moody B, Li Q, Kella D, Shahin A, Kooistra T, Perry D, Mark RG. The PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU. In Computing in Cardiology, 2015. September 2015; .

[2] Moody GB. gqrs man page. URL http://www.physionet.org/physiotools/wag/gqrs-1.htm.

[3] Hamilton P. Open source ecg analysis. In Computers in Cardiology 2002, volume 29. 2002; 101104.

[4] Zong W, Heldt T, Moody G, Mark R. An open-source algorithm to detect onset of arterial blood pressure pulses. In Computers in Cardiology 2003. 2003; 259–262.

[5] Q Li and RG Mark and GD Clifford. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter. Physiological measurement 2008;29:15–32.

[6] Li Q, Clifford GD. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. Physiological Measurement 2012;33(9):1491. URL http://stacks.iop.org/0967-3334/33/i=9/a=1491.

[7] Sakoe H. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics Speech and Signal Processing 1978;26:43–49.

[8] 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 (June 13);101(23):e215–e220. Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[9] Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman LW, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG. Multiparameter intelligent monitoring in intensive care ii (mimic-ii): A public-access intensive care unit database. Critical Care Medicine May 2011;39:952–960.

Address for correspondence:

Rui Rodrigues
Departamento de Matemática, Faculdade de Ciéncias e Tecnologia/ UNL
Campus da Caparica, 2829-516 Caparica, Portugal
rapr@fct.unl.pt