View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Essex Research Repository

IEEE Trans Biomed Eng, vol. 59, pp. 2828-2837, Oct. 2012. doi:10.1109/TBME.2012.2211356

The algorithm described in the paper can be tested online at: http://neurostat-mit.appspot.com/

A Real-Time Automated Point Process Method for Detection and Correction of Erroneous and Ectopic Heartbeats

Luca Citi, *Member, IEEE,* Emery N Brown, *Fellow, IEEE,* and Riccardo Barbieri, *Senior Member, IEEE*

***Abstract*—The presence of recurring arrhythmic events (also known as cardiac dysrhythmia or irregular heartbeats), as well as erroneous beat detection due to low signal quality, significantly affect estimation of both time and frequency domain indices of heart rate variability (HRV). A reliable, real-time classification and correction of ECG-derived heartbeats is a necessary prerequisite for an accurate on-line monitoring of HRV and cardiovascular control. We have developed a novel point process based method for real-time R-R interval error detection and correction. Given an R-wave event, we assume that the length of the next R-R interval follows a physiologically motivated, time-varying inverse Gaussian probability distribution. We then devise an instantaneous automated detection and correction procedure for erroneous and arrhythmic beats by using the information on the probability of occurrence of the observed beat provided by the model. We test our algorithm over two datasets from the Physionet archive. The Fantasia normal rhythm database is artificially corrupted with known erroneous beats to test both the detection and correction procedure. The benchmark MIT-BIH Arrhythmia database is further considered to test the detection procedure of real arrhythmic events and compare it with results from previously published algorithms. Our automated algorithm represents an improvement over previous procedures, with best specificity for detection of correct beats, as well as highest sensitivity to missed and extra beats, artificially misplaced beats, and for real arrhythmic events. A near-optimal heartbeat classification and correction, together with the ability to adapt to time-varying changes of heartbeat dynamics in an on-line fashion, may provide a solid base for building a more reliable real-time HRV monitoring device.**

***Index Terms*—Point-processes, heart rate variability, arrhythmias, ectopic beats, erroneous R-R intervals.**

I. Introduction

Heart rate variability (HRV) techniques [1]–[3] provide a window into the many physiological factors that modulate the normal heart rhythm. In particular, they have been found useful for non-invasive autonomic tone assessment in a wide range of clinical and non-clinical scenarios. In order to provide reliable results, these techniques require uninterrupted series of normal R-R intervals. Peak detection errors – when the algorithm misses a beat and/or detects one when there is none – and ectopic beats often determine abrupt changes in the R-R interval series that can lead to substantial deviations of the HRV indices.

The authors are with the Dept. Anesthesia, Massachusetts General Hospital – Harvard Medical School, Boston, MA, USA, and with the Dept. Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. e-mail: {lciti,enb,barbieri}@neurostat.mit.edu .

The increasing quality of electrocardiogram (ECG) recording systems together with better R-peak detection algorithms has reduced the problem of noisy R-R series in controlled clinical and research environment. Nevertheless, this issue is still of primary importance in new applications of HRV studies, such as ambulatory monitoring in very unstructured environments and extreme conditions, and when the ECG is acquired using single-channel miniature devices, smartphones, or wearable technology.

The simple approach of discarding all sections contaminated by irregular R-R intervals might pose some problems. For example, if the rate of occurrence of mis-detections or ectopic events is not uniform but it is dependent on specific parts of the protocol (e.g., those with frequent motion artifacts) or on the physiological state of the subject (e.g., states that may result in abnormal morphology of the QRS or in higher incidence of ectopic events), then the exclusion of these periods might skew the resulting HRV indices [4]–[6]. When the goal of the research is to study HRV in extreme conditions (e.g., during rescue operations, military missions, or hiking expeditions) removing all sections contaminated by R-peak misdetections caused by motion artifacts is simply not an option.

A better approach is to attempt to correct the corrupted R-R series in order to obtain a new artifact-free R-R series reflecting the underlying HRV dynamics. This new R-R series can then be used for further processing using the HRV analysis techniques of choice. To date, the detection and correction of irregular beats has been mainly achieved by direct human expert evaluation and, more recently, by automatic techniques [4]–[16].

We have developed a novel point process based method for real-time R-R interval error detection and correction. Given an R-wave event, we assume that the length of the next R-R interval follows a time-varying history-dependent inverse Gaussian (IG) probability distribution [17]. We then use this model for the detection and correction of erroneous and arrhythmic beats. A preliminary version of the detection step of this novel approach has been briefly sketched in our previous publication [18]. In Section II-C we present the detection algorithm in detail and later in the paper we show its performance in comparison with three established methods for the detection of erroneous (Section III-A) and ectopic (Section III-C) beats. Whenever a beat is classified as erroneous, the correction step attempts to fix it according to the point-process model. In Section II-D we present the correction algorithm and in Section III-B we assess its performance in

comparison with four reference methods.

A software tool that implements the algorithm described in this manuscript can be tested from our website [19]. It can be used as a preprocessing stage to detect and correct erroneous and ectopic beats in R-R series before analysing it with the chosen HRV technique [1]–[3], such as time-frequency analysis (LF/HF, ...), standard time-domain methods (SDNN, SDANN, RMSSD, ...), and non-linear indices (DFA, SampEn, MSE, ...). Some methods, such as the Lomb spectrogram [20] do not require the correction step but might still benefit from the accurate detection capabilities of our algorithm.

II. Algorithmic Framework

A. Physiologically-based probability model of heartbeat intervals

Each cardiac contraction is initiated by the synchronous depolarization of the heart's pacemaker cells beginning in the sino-atrial (SA) node of the right atrium and then propagating through its specialized conduction system to the left atrium and to the two ventricles. Following every depolarization, the transmembrane potentials of these cells return to their resting potentials and begin anew their spontaneous rise toward threshold [21]. Deterministic models of this integrate (rise of transmembrane potential)-and-fire (depolarization) mechanism, such as the integral pulse frequency modulation (IPFM) model, are used regularly to simulate heart beats [22]–[24] and for the analysis of HRV [25].

An elementary, stochastic integrate-and-fire model is the Gaussian random walk model with drift. We assume that the afferences to the SA node can be represented by random walk excitatory inputs responsible for the basal cardiac rhythm (B), and random walk excitatory (E) and inhibitory (I) inputs, reflecting the influence of the autonomic nervous system through the sympathetic and para-sympathetic branches. In this case, the membrane voltage equation is governed by:

$$dV(t) = \alpha_{\mathrm{B}}\, dN_{\mathrm{B}}(t) + \alpha_{\mathrm{E}}\, dN_{\mathrm{E}}(t) - \alpha_{\mathrm{I}}\, dN_{\mathrm{I}}(t)$$

where $N_{\mathrm{B}}(t) \sim P(\lambda_{\mathrm{B}})$, $N_{\mathrm{E}}(t) \sim P(\lambda_{\mathrm{E}})$ and $N_{\mathrm{I}}(t) \sim P(\lambda_{\mathrm{I}})$ are independent Poisson processes that govern the respective times of the basal, excitatory and inhibitory inputs, while α_{B}, α_{E} and α_{I} are the magnitude of steps up or down [26], [27]. Assuming $V(0)=0$, $m_0=\alpha_{\mathrm{B}}\lambda_{\mathrm{B}}$, $m_1=\alpha_{\mathrm{E}}\lambda_{\mathrm{E}}-\alpha_{\mathrm{I}}\lambda_{\mathrm{I}}$, and $\sigma^2=\alpha_{\mathrm{B}}^2\lambda_{\mathrm{B}}+\alpha_{\mathrm{E}}^2\lambda_{\mathrm{E}}+\alpha_{\mathrm{I}}^2\lambda_{\mathrm{I}}$, and using the properties of Poisson processes, the mean of $V(t)$ is $\mathrm{E}\left[V(t)\right]=(m_0+m_1)\,t$ [22], [23] and its variance $\mathrm{Var}[V(t)]=\sigma^2\,t$. A diffusion approximation of $dV(t)$ before hitting the threshold θ is given by the Wiener process with drift:

$$dV(t) = (m_0+m_1)\,dt + \sigma\, dW(t)$$

where $W(t)$ is a standard Wiener process (Brownian Motion). The probability density of the first passage time for this process, i.e., the times between threshold crossings, is given by the inverse Gaussian probability density [28], [29] defined as:

$$f(t) = \frac{\theta}{\sqrt{2\pi\sigma^2 t^3}}\, \mathrm{e}^{-\frac{(\theta-(m_0+m_1)\,t)^2}{2\sigma^2 t}}$$

which can be written in the usual form after performing the change of variables $\mu=\theta/(m_0+m_1)$ and $\lambda=\theta^2/\sigma^2$. The inverse Gaussian probability density has been used to study both neural processes [30] and heart beats [17], [31].

B. Point-process model of the heartbeat series

An ordered set $\{u_j\}_{j=1}^{J}$ of consecutive heartbeat times (for example, but not limited to, R-waves detected from an ECG) recorded in an observation interval $(0,T]$ can be interpreted as a realization of a point process [32], a type of stochastic process for which any one realization consists of a set of isolated points. For $t \in (0,T]$, the cádlág function $N(t)=\max\{k : u_k \le t\}$ is the associated counting process. Its differential, $dN(t)$, denotes a continuous-time indicator function, where $dN(t)=1$, when there is an event (such as the ventricular contraction) or $dN(t)=0$, otherwise.

In this paper we assume a history dependent, time-varying, model of the the probability distribution of the waiting time until the next R-wave according to a physiologically motivated (see Section II-A) inverse Gaussian distribution. Specifically, given any beat event u_k, the probability density function (PDF) of the length of the next R-R interval, $\tau-u_k$, is

$$f\left(\tau-u_k \,|\, \mu\left(\boldsymbol{H}_k,\boldsymbol{\theta}(t)\right), \lambda\left(\boldsymbol{\theta}(t)\right)\right) = \sqrt{\frac{\lambda\left(\boldsymbol{\theta}(t)\right)}{2\pi(\tau-u_k)^3}} \exp\left\{-\frac{1}{2}\frac{\lambda\left(\boldsymbol{\theta}(t)\right)\,\left(\tau-u_k-\mu\left(\boldsymbol{H}_k,\boldsymbol{\theta}(t)\right)\right)^2}{(\tau-u_k)\,\mu^2}\right\}$$

where the shape parameter, λ, and the mean of the distribution, μ, depend on a vector of time-varying parameters $\boldsymbol{\theta}(t)=\{\theta_1(t),\ldots,\theta_{P+1}(t)\}$ and a history vector $\boldsymbol{H}_k=\{w_k, w_{k-1},\ldots,w_{k-(P-1)}\}$ containing information about P previous R-R intervals $w_i=u_i-u_{i-1}$. While λ is simply $\lambda\left(\boldsymbol{\theta}(t)\right)=\theta_{P+1}(t)$, the history-dependent mean is a regression of the past P R-R intervals with time-varying weights:

$$\mu\left(\boldsymbol{H}_k,\boldsymbol{\theta}(t)\right) = \sum_{i=1}^{P} \theta_i(t)\, w_{k-i}\,. \tag{1}$$

A local maximum likelihood method [17] is used to estimate the unknown time-varying parameter set $\boldsymbol{\theta}(t)$. Given a local observation interval $(t-W,t]$ of duration W, we consider the subset of R-wave events within the interval, $U_{m:n}=\{u_k : m\le k\le n\}$ with $m=N(t-W)+1$ and $n=N(t)$. At each time t, we find the parameter vector $\tilde{\boldsymbol{\theta}}(t)$ that maximizes the local log-likelihood, given the R-wave events recorded in the local observation interval:

$$L(\boldsymbol{\theta}(t)\,|\,U_{m:n}) = \sum_{k=m+P}^{n-1} \omega(t-u_{k+1}) \log\left[f\left(u_{k+1}-u_k \,|\, \mu(\boldsymbol{H}_k,\boldsymbol{\theta}(t)), \lambda(\boldsymbol{\theta}(t))\right)\right] \tag{2}$$

where $\omega(\tau)=\mathrm{e}^{-\alpha\,\tau}$ is an exponential weighting function for the local likelihood [17], [33].

C. Detection of erroneous heartbeats

After observing the $(k+1)$-th R-wave, at time u_{k+1}, it is possible to assess whether this observation is in agreement with the model resulting from the most recent parameter vector $\tilde{\boldsymbol{\theta}}(u_k)$. For conciseness, in the following, let $\mu_1 = \mu(\boldsymbol{H}_k, \tilde{\boldsymbol{\theta}}(u_k))$ denote the mean for the distribution of the first R-R interval following u_k, given the history and the model at time u_k, and let $\lambda_1 = \lambda(\tilde{\boldsymbol{\theta}}(u_k))$ denote the shape parameter, given the model at time u_k.

Normal beat: A straightforward way to test the hypothesis that the beat at time u_{k+1} is in agreement with the model is to evaluate the log probability density of observing the event u_{k+1} given the previous event u_k, the history $\boldsymbol{H}_k$ and the model $\tilde{\boldsymbol{\theta}}(u_k)$: $p = \log f(u_{k+1} - u_k \,|\, \mu_1, \lambda_1)$. In [14], this probability p was compared with a fixed threshold and, if lower, a beat was either inserted or deleted. While this approach works reasonably well, it has some glitches. In fact, a very low value of p might indicate an erroneous R-wave detection, an ectopic heartbeat, but also a sudden physiologic change in autonomic control. Ideally, we would like to be able to detect and possibly correct the former two cases while preserving the latter. For this reason, in this work, we take a different approach: instead of setting a threshold for the score p, we compare its value with the values obtained from alternative hypotheses, assuming that the event at time u_{k+1} is erroneous or arrhythmic. Below, we analyze these cases in detail while in Fig. 1 we give a schematic representation.

1) Extra beat: One hypothesis that we consider is that the beat at time u_{k+1} is an extra event that was erroneously placed between two consecutive heartbeats. This can happen, for example, when the detection algorithm mistakenly identifies a T wave as a ventricular contraction. The score of this event is $p_\mathrm{e} = \log f(u_{k+2} - u_k \,|\, \mu_1, \lambda_1)$ where u_{k+2} is the time of the event following u_{k+1}. If $p_\mathrm{e} > p + \eta_\mathrm{e}$, where η_e is a pre-defined threshold, the event at time u_{k+1} is labelled as an extra ("e") beat.

2) Missed beat: Another hypothesis is that between the events u_k and u_{k+1} there was an additional event that has been missed. This can happen, for example, when the detection algorithm misses an R-wave because a superimposed artifact has drastically changed its morphology. In this case, the time elapsed between u_k and u_{k+1} should be compatible with the prediction of the model for the sum of two R-R intervals given the history at time u_k. We assume that there are two stochastic events following u_k, the first at time τ' and the second at time τ, of which only the latter was observed and mistakenly labeled as u_{k+1}. According to our model, the distribution of the first R-R interval, $\tau' - u_k$, is simply $f(\tau' - u_k \,|\, \mu_1, \lambda_1)$. The distribution of the second R-R interval, $\tau - \tau'$, also depends on the first R-R interval. In fact, the unknown term $\tau' - u_k$ enters the history vector $\boldsymbol{H}'(\tau' - u_k) = \{\tau' - u_k, w_k, w_{k-1}, \dots\}$ and, through the regression in (1), affects the first moment of the next R-R interval, $\mu_2 = \mu(\boldsymbol{H}'(\tau' - u_k), \tilde{\boldsymbol{\theta}}(u_k))$, that we write as $\mu_2(\tau' - u_k)$ to stress this dependency. The PDF of the second R-R interval can be written as $f(\tau - \tau' \,|\, \mu_2(\tau' - u_k), \lambda_2)$ where $\lambda_2 = \lambda_1$. The total probability density function can be obtained marginalizing the unknown τ' in the joint probability density:

$$\int_{u_k}^{\tau} f(\tau - \tau' \,|\, \mu_2(\tau' - u_k), \lambda_2)\, f(\tau' - u_k \,|\, \mu_1, \lambda_1)\, \mathrm{d}\tau' \,. \quad (3)$$

Rather than attempting to solve (3), here we make the simplifying assumption that the resulting probability density function can be approximated by an IG distribution with mean $\mu_{1+2} = \mu_1 + \mu_2(\mu_1)$ and shape parameter

$$\lambda_{1+2} = \lambda_1 \, \frac{(\mu_1 + \mu_2(\mu_1))^3}{(1 + \tilde{\theta}_1(u_k))^2 \, \mu_1^3 + \mu_2^3(\mu_1)} \quad (4)$$

whose mathematical derivation is shown in the Appendix. Using these results, we can now evaluate the log probability density of observing a beat at u_{k+1} after a missed beat at an unknown time $\tau' \in (u_k, u_{k+1})$: $p_\mathrm{s} = \log f(u_{k+1} - u_k \,|\, \mu_{1+2}, \lambda_{1+2})$. If $p_\mathrm{s} > p + \eta_\mathrm{s}$, where η_s is a pre-defined threshold, we assume that there might be a missed (skipped, "s") beat.

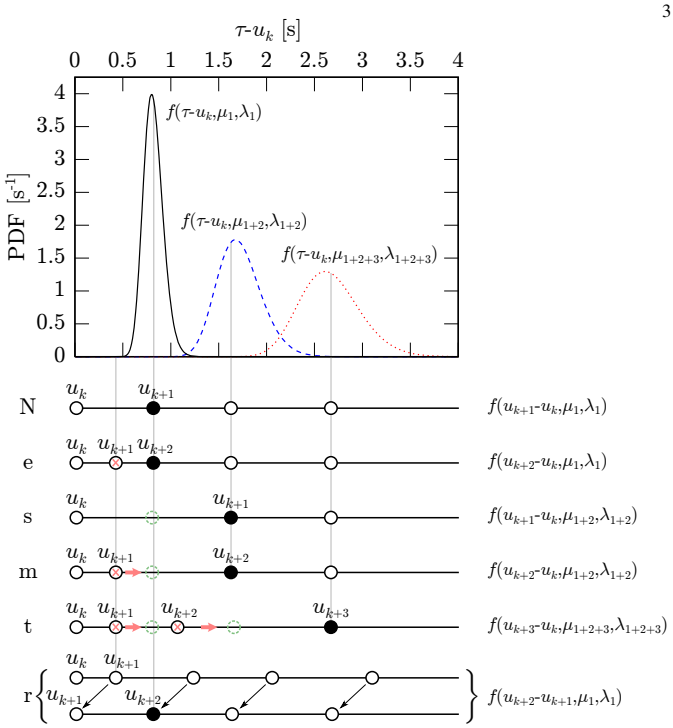

Fig. 1. Schematic representation of the different hypotheses that we make for the beats following u_k. The plot at the top shows the IG distribution of the time of occurrence of the first beat after u_k (black solid line), the second one (blue dashed line) and the third one (red dotted line). All three probabilities are estimated at time u_k and this explains why predictions that are further in the future present a wider variance. In the lower part of the figure, sample configurations of the beat times representing the different hypotheses ("N", ..., "r") are reported. The black full circles indicate the beats for which the PDF on the right presents a high value; the crosses mark the beats that are going to be removed while the dashed circles are beats that need to be inserted. For example in the "e" case the beat u_{k+2} has high probability of being the first beat after u_k (black bell) and for this reason the beat u_{k+1} will be removed (red cross). In the "m" case, instead, u_{k+1} has low probability of being the first beat after u_k (black bell) while u_{k+2} has high probability of being the second beat after u_k (blue bell) and, therefore, the beat u_{k+1} is removed from its current position (red cross) and shifted to its most likely position (green dashed circle).

3) Misplaced beat: A further hypothesis is that the beat at time u_{k+1} is misplaced, i.e., a beat at a different time in the interval (u_k, u_{k+2}) has been mistakenly assigned to time u_{k+1}. This can happen, for example, when an artifact deforms the QRS complex and makes the detector misinterpret some random deflection for an R wave and miss the true R wave. For simplifying purposes, we will consider in this category also some types of ectopic beats, such as premature ventricular contractions that are followed by a complete compensatory pause. Even if, strictly speaking, ectopic beats are not misdetections, it is often desirable to treat them this way. In fact, ectopic beats may significantly impair the results of many heart rate analysis techniques (including most time-domain and spectral estimation methods) while, on the other hand, acceptance of only ectopic-free, short-term recordings may introduce significant selection bias [4]–[6]. For this reason, it is desirable to detect irregular beats and, possibly, replace them with fictitious regular beats in order to carry on the desired analysis. A beat u_{k+1} is likely to be a misplaced one whenever the corresponding R-R interval does not fit the model while the sum of the two R-R intervals, $u_{k+2} - u_k$, does. Therefore, the score for the case of a misplaced beat is: $p_{\mathrm{m}} = \log f(u_{k+2} - u_k \,|\, \mu_{1+2}, \lambda_{1+2})$, and if $p_{\mathrm{m}} > p + \eta_{\mathrm{m}}$, where η_{m} is a pre-defined threshold, we assume that there might be a misplaced ("m") beat.

4) Two misplaced beats: This hypothesis accounts for the case that two beats in a row, u_{k+1} and u_{k+2}, are misplaced. This can happen, for example, in some types of arrhythmias where two shorter beats are followed by a long compensatory pause. This case is handled similarly to the case of one misplaced beat. The beats u_{k+1} and u_{k+2} are likely to be misplaced whenever the corresponding R-R intervals do not fit the model while the sum of the three R-R intervals, $u_{k+3} - u_k$, does. Making the approximation that $u_{k+3} - u_k$ follows an IG distribution of parameters μ_{1+2+3} and λ_{1+2+3} (defined in the Appendix), the score for the case of two misplaced beats is $p_{\mathrm{t}} = \log f(u_{k+3} - u_k \,|\, \mu_{1+2+3}, \lambda_{1+2+3})$. If the condition for "m" holds and, additionally, $p_{\mathrm{t}} > p_{\mathrm{m}} + \eta_{\mathrm{t}}$, where η_{t} is a pre-defined threshold, we assume that there might be two misplaced ("t") beats. In other words, we require the first beat to be detected as misplaced (because this case is a special case of "m"), and, additionally, that the second beat is misplaced with respect to the prediction of the model for the first beat.

5) Resetting beat: The last hypothesis accounts for resetting ectopic beats, i.e., ectopic beats not followed by a compensatory pause. As a result of the absence of the compensatory pause, not only the R-R interval $u_{k+1} - u_k$ does not fit the model but neither do the sum of the two R-R intervals, $u_{k+2} - u_k$, or the sum of three, $u_{k+3} - u_k$. As a matter of a fact, the R-R interval after the ectopic beat, $u_{k+2} - u_{k+1}$, is more likely to fit the model than any of previous hypotheses. For this reason, defining $p_{\mathrm{r}} = \log f(u_{k+2} - u_{k+1} \,|\, \mu_1, \lambda_1)$, the beat u_{k+1} is likely to be a resetting ("r") ectopic beat if $p_{\mathrm{r}} > \max\{p, p_{\mathrm{e}}, p_{\mathrm{s}}, p_{\mathrm{m}}, p_{\mathrm{t}}\} + \eta_{\mathrm{r}}$, where η_{r} is a pre-defined threshold.

Bootstrap detection: As the algorithm uses a sliding window of duration W to fit the model (see end of Section II-B), no model is available before time W from the beginning of the recording. A simpler bootstrap detection algorithm is used during this period to provide some results even for beats $u_k \in (0, W)$ and also to avoid training the algorithm on non-normal R-R intervals. This bootstrap algorithm is based on outlier rejection where beats whose distance from the median R-R is more than 7 times the median absolute deviation of the R-R intervals are marked as erroneous.

D. Correction of erroneous heartbeats

After an outlying beat is detected, our algorithm attempts to correct it. The correction strategy depends on the type of erroneous beat detected, as reported in detail below. Finally, the attempted solution is accepted only if it is an improvement according to the improvement check that we describe in the following section.

1) Extra beat: The case of an extra beat is simply addressed by removing the beat.

2) Missed beat: In case a missed beat is detected, a new beat is inserted between the event u_k and the event u_{k+1}. The time of the inserted event τ' is:

$$\tau' = \underset{\tau' \in (u_k, u_{k+1})}{\arg\max}\; f(\tau' - u_k \,|\, \mu_1, \lambda_1)\; f(u_{k+1} - \tau' \,|\, \mu_2(\tau' - u_k), \lambda_2) \tag{5}$$

where μ_1, λ_1, $\mu_2(\tau' - u_k)$ and λ_2 are defined as before. Inserting an event τ' in (u_k, u_{k+1}) splits this interval into two R-R intervals; therefore, equation (5) attempts to find the optimal τ' accounting for the probability distribution of the first interval given the P R-R intervals preceding it (first factor of the product) but also for the probability distribution of the second R-R interval given its P previous R-R intervals (second factor), including $\tau' - u_k$. This is an improvement over the approach in [14] where the beat was inserted at the time maximizing only the first of the two factors in (5).

3) Misplaced beat: The case of a misplaced beat is similar to the previous case. The event at time u_{k+1} is moved at time τ' given by:

$$\tau' = \underset{\tau' \in (u_k, u_{k+2})}{\arg\max}\; f(\tau' - u_k \,|\, \mu_1, \lambda_1)\; f(u_{k+2} - \tau' \,|\, \mu_2(\tau' - u_k), \lambda_2)\;. \tag{6}$$

Equations (5) and (6) are solved numerically using a Newton-Raphson algorythm.

4) Two misplaced beats: To correct the case of two misplaced beats and find the optimal times of the two unknown beats we use a *greedy algorithm*. Each one of the beat times is optimized in turn, using (6) for the first beat and a similar equation for the second beat, while keeping the other fixed. This process is iterated until convergence.

5) Resetting ectopic beat: The case of a resetting beat cannot be corrected like the other cases because it would require shifting all the following beats. Therefore our algorithm defers the decision about the action to take to the user, who can choose the best strategy for the type of HRV analysis that follows (for example omit the beat or suspend adaptive causal estimation).

E. Improvement check

In the cases above, the correction step changes the series of events $U = \{u_j\}$ into an alternative series $\hat{U} = \{\hat{u}_j\}$ by

removing, inserting, or moving one or two events. For the case of a resetting beat, "r", we consider the virtual correction (that we do not take) of shifting back by $u_{k+1} - u_k$ all the beats starting from u_{k+1}.

Before the new series is accepted, we evaluate the probability of observing Q events following u_k for both series:

$$\mathbf{P}(\hat{U}_{k-P:k+Q}) = \sum_{j=k}^{k+Q-1} \log\left[f\left(\hat{u}_{j+1} - \hat{u}_j \,|\, \hat{\boldsymbol{H}}_j, \tilde{\boldsymbol{\theta}}(u_k)\right)\right] ,$$

$$\mathbf{P}(U_{k-P:k+Q}) = \sum_{j=k}^{k+Q-1} \log\left[f\left(u_{j+1} - u_j \,|\, \boldsymbol{H}_j, \tilde{\boldsymbol{\theta}}(u_k)\right)\right] .$$

The corrected series is accepted only if $\mathbf{P}(\hat{U}_{k-P:k+Q}) > \mathbf{P}(U_{k-P:k+Q}) + \hat{\eta}_a$, where a is one of $\{\mathrm{e}, \mathrm{s}, \mathrm{m}, \mathrm{t}, \mathrm{r}\}$ depending on the action was taken to correct the beats. This ensures not only that the correction fits the past values (through $\boldsymbol{H}_k$) but also that it is a good regressor for the future R-R intervals (through $\hat{\boldsymbol{H}}_j$, $j > k$). If the corrected series passes the test, then $\hat{U} = \{\hat{u}_j\}$ replaces $U = \{u_j\}$, a new updated parameter vector $\tilde{\boldsymbol{\theta}}(u_{k+1})$ is estimated using (2), and the detection procedure starts again for the next event, u_{k+2} (u_{k+3} in the case "t").

F. Values of the parameters

Empirical sensitivity analysis conducted on in-house recordings showed that the model provides robust results for parameters within meaningful ranges of values.

Without conducting an exhaustive research on the whole parameter space, we found that the following settings provide excellent results in most conditions: $P = 5$, $W = 60\,\mathrm{s}$, $\alpha = 0.02\,\mathrm{s}^{-1}$, $\eta_\mathrm{e} = 3$, $\hat{\eta}_\mathrm{e} = 8$, $\eta_\mathrm{s} = 0$, $\hat{\eta}_\mathrm{s} = 4$, $\eta_\mathrm{m} = 2$, $\hat{\eta}_\mathrm{m} = 7$, $\eta_\mathrm{t} = 8$, $\hat{\eta}_\mathrm{t} = 28$, $\eta_\mathrm{r} = 6$, $\hat{\eta}_\mathrm{r} = 14$, and $Q = 3$. These values of the parameters were also used during the analysis described in the following of this manuscript.

III. Validation and Performance Assessment

A. Detection on artificially-corrupted R-R series

We tested the performance of our detection algorithm on a set of real R-R timeseries with artificially deleted, shifted, and inserted beats and compared our results with three well-established approaches.

Methods: The first method that we implemented for comparison is the well-established algorithm by Berntson et al. [9] which uses two criteria to assess whether a heartbeat is an artifact: Maximum Expected Difference (MED) and Minimal Artifact Difference (MAD). In order to allow the algorithm to adapt to non-stationary data, for each R-event u_k, MED and MAD were estimated using events in a 60 s sliding window preceding u_k. The second algorithm that we used is based on the important contribution by Mateo and Laguna [13]. For each R-event u_k we evaluated equation (1) from the original work (where t_k corresponds to our u_k) using a time-varying threshold $U_k = 4.3\,\sigma_k$ where σ_k is the standard deviation of the w_j values obtained for events in a 60 s sliding window preceding u_k. Then we classified the beat using the set of rules described in Section II.A of the referenced paper. Finally, we implemented the more recent method proposed by Rand et al. [15]. We followed the algorithm described in the original paper and used the values of parameters that they found optimal.

For this comparison, we used a subset of the Fantasia database [34] from the Physionet archive [35]. We selected all the records that, according to the annotation database, contained no more than two non-"N" heartbeats.[1] We tested all four algorithms on each pristine record and recorded the number of "N" heartbeats correctly classified as such and the number of beats which triggered a false alarm. Then we tested the four algorithms' performance in the detection of the three main types of erroneous beats: extra, missed, and misplaced beats. To test the detection of "e" (extra) beats, after inserting a beat before every beat of index $k = 100\,n$ ($n \in \{1, 2, 3, \dots\}$) of each record, we recorded how many of these beats were correctly identified as erroneous beats. Similarly, we measured the performance in the detection of "s" (missed) beats by removing from the original record every beat of index $k = 100\,n$.

Finally, we assessed each algorithm's efficacy in the detec-

[1]The records used were: f1o01, f1o05, f1o10, f1y01, f1y03, f1y08, f2o04, f2y03, f2y06.

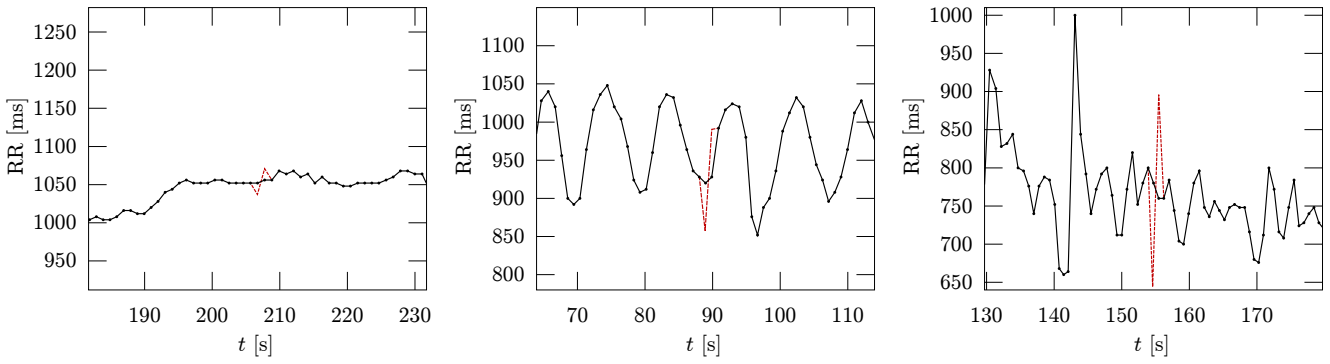

Fig. 2. The figure shows examples from three subjects of the original R-R series (continuous line with markers) with one artificially misplaced beat (dashed line) obtained as described in Section III-A. The R-R series on the left has a very low beat-to-beat variability, the one in the centre presents a marked respiratory sinus arrhythmia (RSA), while the series on the right has an even higher variability but the RSA is less predictable. In all three cases, the same value $q = 2$ was used to determine the record-dependent displacement $\Delta t_r = q \cdot \mathrm{RMSSD}_r$. This allowed us to average the algorithms performance across subjects on a task of comparable difficulty. If we had moved the beats by a fixed displacement, e.g., 50 ms, the detection would have been trivial for the R-R series on the left and very hard for the series on the right.

TABLE I

RESULTS OF THE DETECTION OF IRREGULAR BEATS ON ARTIFICIALLY-CORRUPTED R-R SERIES. OUR METHOD (PP) IS COMPARED WITH THREE REFERENCE METHODS (RAND2007, MATEO2003, AND BERNTSON1990) AND THE PERCENTAGE OF CORRECT DETECTION (CORR) OF NORMAL ("N") BEATS AND ARTIFICIALLY CORRUPTED BEATS ("s", "e", "m", SEE TEXT) IS REPORTED.

Series	$\overline{\Delta t_r}$ [ms]	Tot #	PP		Rand2007		Mateo2003		Berntson1990	
			Corr [%]	CT [%]	Corr [%]	CT [%]	Corr [%]	CT [%]	Corr [%]	CT [%]
N		60017	99.985	-	99.880	-	99.923	-	99.938	-
s		602	100.000	99.502	100.000	94.518	95.681	94.186	100.000	-
e		602	100.000	99.834	100.000	96.379	99.668	99.668	100.000	-
m $q = 2$	68	602	40.864	36.877	11.628	1.661	40.199	33.223	5.316	-
m $q = 4$	137	602	96.013	93.189	54.153	17.110	96.844	83.887	18.106	-
m $q = 8$	322	602	100.000	98.173	94.352	68.771	100.000	90.033	74.917	-
m $q = 16$	466	602	100.000	99.336	100.000	93.355	99.834	84.385	95.515	-

tion of "m" (misplaced) beats, by shifting each beat of index $k = 100\,n$ by a temporal displacement Δt. In order to account for the fact that the detectability of misplaced beats depends on the heart rate variability of the R-R series, we used a record-dependent displacement $\Delta t_r = q \cdot \mathrm{RMSSD}_r$ where RMSSD_r is the RMSSD measure [1] for the record r (see Fig. 2). If we think of the RMSSD_r of a normal sinus series as the noise amplitude and Δt_r as the amplitude of the signal we are trying to detect, q^2 represents a form of signal-to-noise ratio (SNR). For obvious reasons, we capped Δt_r at $0.75\,\overline{\mathrm{NN}}_r$ where $\overline{\mathrm{NN}}_r$ is the average N-N interval of the record r. The test was repeated for the values of $q \in \{2, 4, 8, 16\}$.

Results: Table I reports, for the normal series and for each artificially-corrupted series, the number of beats tested (columns "Tot #") and the percentage of these that were correctly detected (columns labelled "Corr") by the four algorithms. For the "m" beats, the average displacement $\overline{\Delta t} = \mathrm{mean}_r\, \Delta t_r$ among all records is also reported.

As one further goal of our algorithm is to try to correct erroneous beats, it is important that the algorithm is also able to classify the type of error ("e", "s", and "m") and, as a result, the type of action needed to attempt to fix it. All algorithms considered, except the one by Berntson et al., provide this type of information. Table I reports, in the columns labelled "CT", the percentage of beats for which each algorithm correctly classified the type of error.

These results show that our method represents an improvement over the previous algorithms. In fact it presents the best false detection rate (row "N") with only 9 normal beats out of 60017 reported as erroneous (which correspond to 99.985% correct classification), while the second best method, by Berntson, triggered a false detection in 37 cases. Our method also shows perfect detection of missed (row "s") and extra beats (row "e"), and better detection of misplaced beats at all levels of SNR (rows "m" of Table I and Fig. 3). In terms of sensitivity, only the method by Mateo and Laguna provides similar results to our approach but at the cost of a much lower specificity.

We also performed a Monte Carlo analysis of performance by randomly sampling the test beats with probability 1%, obtainining results very similar to those reported here. We decided to report the results of the deterministic sampling of the test beats (every hundredth beat), rather than those from random sampling, to support reproducibility of results and allow other scientists to test and compare their algorithms on exactly the same recordings and beats.

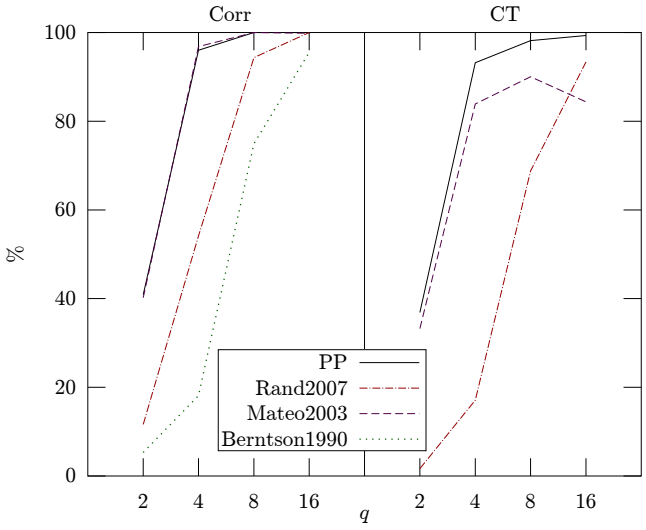

Fig. 3. The left side of the plot (Corr) reports the percentage of the beats shifted by $\Delta t_r = q \cdot \mathrm{RMSSD}_r$ (as detailed in Section III-A) that were correctly detected as erroneous. The right side of the plot (CT) shows the percentage of the same beats that were correctly classified as type "m" beats. For example, a shifted beat detected as extra beat would count as a hit for "Corr" but as a fail for "CT".

B. Correction on artificially-corrupted R-R series

In this section we report the results of evaluating the performance of the correction phase of our algorithm and compare it with existing methods.

Methods: We wanted to find out how accurately our model is able to estimate the unknown time of occurrence of a beat (either a missed or a misplaced beat). This was simply evaluated by using (6) to estimate the beat location $\hat{u}_k^{\mathrm{PP}}$ of every beat of index $k = 100\,n$ from each original record described in Section III-A and then measuring the difference, $\hat{u}_k^{\mathrm{PP}} - u_k$, between the predicted time of the beat and the actual time of the original beat.

We compared our algorithm with four existing methods of estimating the position of a missing beat. As first method we used the "nguess" function available from the WFDB library of the PhysioToolkit [35] and called its estimate $\hat{u}_k^{\mathrm{N}}$. The second method consisted in splitting the interval around the missing beat in two identical R-R intervals by inserting a beat at time $\hat{u}_k^{\mathrm{H}}$ such that $u_{k+1} - \hat{u}_k^{\mathrm{H}} = \hat{u}_k^{\mathrm{H}} - u_{k-1}$. Then we tested the family of $\hat{\delta}_N$ estimators given by equation (45) of [36] and

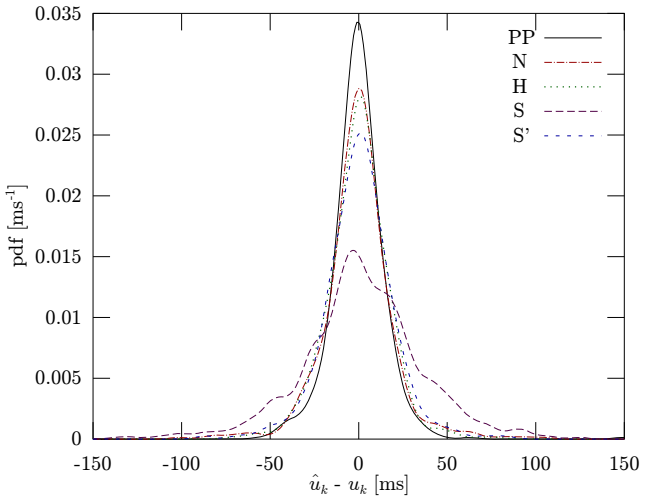

Fig. 4. Probability density function of the difference between the estimated beat time $\hat{u}_k$ and the actual time of the original beat u_k for our point-process algorithm (PP) and for the four reference methods (N, H, S, S'). The PDF was estimated using kernel density estimation with Gaussian kernels and a bandwidth of 5 ms.

found, in agreement with their findings, that the first order estimator ($\hat{\delta}_1$) is the one giving best results. Therefore, as third method we used the $\hat{\delta}_1$ estimator which returns a beat time $\hat{u}_k^{\mathrm{S}}$ such that the new R-R interval repeats the previous R-R interval, i.e., $\hat{u}_k^{\mathrm{S}} - u_{k-1} = u_{k-1} - u_{k-2}$. Finally, the fourth method that we used for comparison is a non-causal version of $\hat{\delta}_1$ which is simply the average of the beat time estimated using the previous R-R interval and that estimated using the next R-R interval, i.e., $\hat{u}_k^{\mathrm{S}'} = u_{k+1} + u_{k-1} - (u_{k-2} + u_{k+2})/2$.

Results: Table II reports a comparison between the different methods in terms of root-mean-square (RMS) error between the estimated time and the original beat time. The first row shows the result of pooling all 602 beats from all subjects and then evaluating the pooled RMS. The second row reports the result of finding the RMS error for each subject separately and then averaging the results. Similarly, the last row shows the median value of the RMS error for the individual subjects.

Our algorithm (first column) clearly outperforms all four reference methods in all measures and provides accurate estimates with an RMS error of under 15 ms. This is also confirmed by the probability density function of the difference between the estimated beat time $\hat{u}_k$ and the actual time of the original beat u_k (Fig. 4).

TABLE II
RESULTS OF THE ESTIMATION OF UNKNOWN BEAT TIMES. THE ROOT MEAN SQUARE (RMS) OF THE ERROR BETWEEN THE BEAT TIME ESTIMATED BY OUR METHOD $\hat{u}_k^{\mathrm{PP}}$ AND THE TRUE BEAT TIME u_k IS COMPARED WITH THAT OF THE OTHER METHODS REPORTED IN THE TEXT.

RMS	$\hat{u}_k^{\mathrm{PP}} - u_k$	$\hat{u}_k^{\mathrm{N}} - u_k$	$\hat{u}_k^{\mathrm{H}} - u_k$	$\hat{u}_k^{\mathrm{S}} - u_k$	$\hat{u}_k^{\mathrm{S}'} - u_k$
pooled	14.98 ms	21.78 ms	18.79 ms	34.59 ms	19.69 ms
average	14.21 ms	19.98 ms	17.66 ms	33.08 ms	18.59 ms
median	12.10 ms	17.81 ms	17.68 ms	32.92 ms	15.71 ms

C. Detection on arrhythmia database

As mentioned before, the algorithm presented here can also be used to detect ectopic events based on the R-R intervals alone, not for the purpose of classifying the different types of arrhythmias but simply to try to replace them with fictitious normal beats and carry on with the desired type of HRV analysis.

Methods: The performance of our algorithm in the detection of ectopic beats was evaluated on the MIT-BIH Arrhythmia database [35], [37]. For this analysis we selected the recordings having less than 10 non-normal beats in any 20-second window.[2] The main reason is that we are interested in R-R series that are mildly affected by erroneous or ectopic events, and that can be recovered and further processed with HRV techniques. In a recording with an excessive number of irregular beats, the information about the underlying sinus rhythm cannot be reliably recovered and therefore the HRV indices cannot be assessed.

For each record of the MIT-BIH Arrhythmia Database considered, we ran the four algorithms described in Section III-A. As some of the algorithms use a 60 s sliding window to adaptively estimate their thresholds, we did not test the detection in the first minute of the record. We also ignored types of ectopic beats that merely affect the morphology of the contraction potential but not its timing because all the algorithm that we tested do not have access to the raw ECG data but only to the series of R-R intervals.

Results: Table III reports the confusion matrices obtained for the four algorithms, whereas Table IV shows their performance measures. Our method outperforms the previous algorithms, especially in terms of a lower number of false detections, resulting in higher specificity and positive predicted value. Figure 5 shows a few examples of real arrhythmic events from this database and how our algorithm detects and corrects them.

[2]The records used were: 100, 101, 103, 105, 108, 112, 113, 114, 115, 116, 117, 121, 122, 123, 215, 230.

TABLE III
CONFUSION MATRICES FOR THE POINT-PROCESS METHOD PRESENTED HERE (PP) AND FOR THE THREE STATE-OF-THE-ART METHODS CONSIDERED AGAINST THE EXPERTS' ANNOTATIONS IN THE DATABASE (ANNOT). THE LABEL "N" REPRESENT NORMAL BEATS WHILE "$\neq N$" CONTAINS ALL OTHER TYPES OF BEATS.

		PP		Rand2007		Mateo2003		Berntson1990	
		$\neq N$	N	$\neq N$	N	$\neq N$	N	$\neq N$	N
Annot	$\neq N$	389	24	355	58	351	62	387	26
	N	5	32596	99	32502	16	32585	20	32581

TABLE IV
PERFORMANCE MEASURES OF THE FOUR METHODS.

	PP	Rand2007	Mateo2003	Berntson1990
Accuracy	99.91%	99.52%	99.76%	99.86%
Sensitivity	94.19%	85.96%	84.99%	93.70%
Specificity	99.98%	99.70%	99.95%	99.94%
Pos.Pred.Val	98.73%	78.19%	95.64%	95.09%

IV. Discussion and Conclusions

We present a novel paradigm for detection and correction of ectopic and erroneous heartbeats. At the core of the method is the ability of the underlying probabilistic model to indicate with accurate precision the degree of probability of having a heartbeat at each moment in time. This feature allows to create a tree-like decisional structure based on quantitative comparisons between the observed events and the probability of having the event at that time. More specifically, we test the alternative hypotheses that: a) a beat has been correctly identified, b) a beat has been missed, c) the observed beat is spurious, d) one or two beats need to be moved to a most probable time according to the probability distribution, or e) a beat is a resetting ectopic beat. Since the model is defined in continuous time and is updated recursively in an on-line fashion, the devised algorithm is able to process the decisional tree in real-time using only short-time past information, and then finalizing the correction within at maximum three beats in the future.

Our method, similarly to the IPFM-based method, is based on a physiologically motivated model, with the further advantage that the point process framework allows for dynamically updating the model by tracking the time-varying heartbeat variations, including an objective goodness-of-fit framework which can validate the decisional outcome based on optimal model selection. Importantly, the parameter vector is here computed only to estimate the likelihood, but this is the same vector that defines the instantaneous time and frequency domain point-process HRV as in [17], [38]–[40], thus providing a simultaneous assessment of cardiovascular control that is unbiased by artifactual or arrhythmic events.

Our results provide evidence of the efficacy of the detection method in recognizing missing, spurious, mis-detected, and irregular heartbeats with remarkable accuracy even in scenarios that require discerning very small time resolution errors at millisecond scales. They further point at our method as more accurate than three previously published detection methods and four correction methods. In particular, our automated method achieves 99.985% specificity for detection of correct beats, 100% sensitivity to missed and extra beats, 96.01% sensitivity for beats artificially misplaced by 137 ms on average, and 98.73% positive prediction value for real arrhythmic events. Our algorithm represents an improvement over previous procedures: only one other method provides similar results but at the cost of a much lower specificity. Our correction procedure provides accurate estimates of the replacement time of an unobserved beat, with and error below 15 ms, outperforming other existing methods of estimating the time of missing beats.

One important feature of the technique is that it works with predetermined decisional thresholds that have been fixed once for all during our analyses, and have been demonstrated to work on a wide range of experimental recordings, with high robustness to noise and ability to cope with inter-subject variability. In particular, the algorithm can be used for exclusive pre-processing purposes of any R-R interval series and consequent application of any kind of HRV analysis. The routine is very fast, easy to run, and available online at no cost for research purposes [19]. We hope dissemination will further corroborate the robustness and efficacy of the proposed method in an even wider range of physiological and clinical settings.

Appendix

In this Appendix we find the probability distribution of $\tau - u_k$ which is the sum of two unknown intervals $\tau' - u_k$ and $\tau - \tau'$ distributed according to $f(\tau' - u_k \,|\, \mu_1, \lambda_1)$ and $f(\tau - \tau' \,|\, \mu_2(\tau' - u_k), \lambda_2)$. Rather than finding the exact solution given by (3), we approximate it as an inverse Gaussian distribution of mean μ_{1+2} and shape parameter λ_{1+2}, that we are about to introduce. Let us define $\hat{w}_{k+1} = \tau' - u_k = \mu_1 + \epsilon_1$ and $\hat{w}_{k+2} = \tau - \tau' = \mu_2(\hat{w}_{k+1}) + \epsilon_2$ where ϵ_1 and ϵ_2 are

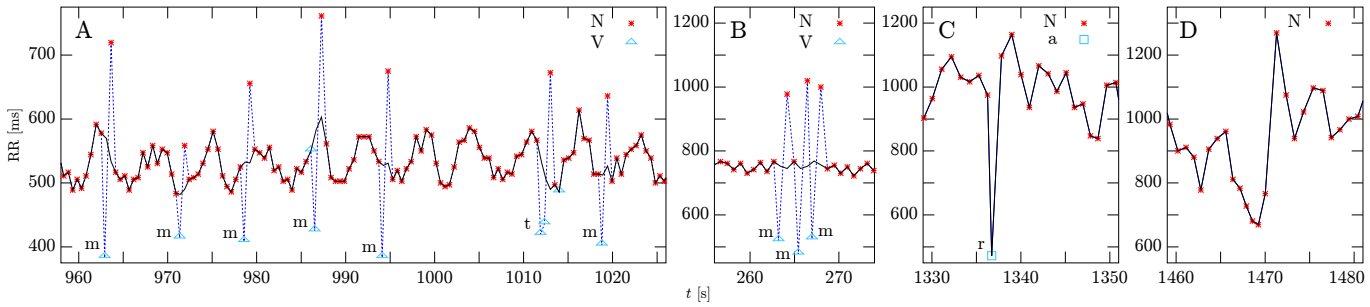

Fig. 5. The figure shows examples of beats from the MIT-BIH Arrhythmia database and how they were detected and corrected by our algorithm. The markers (little star, triangle, and square) and the associated labels ("N", "V", "a") correspond to the database annotations and they mean, respectively: normal beat, premature ventricular contraction, and aberrated atrial premature beat. The "m", "t", and "r" letters next to some beats are the output of our detection algorithm (all other beats were detected as normal). The dashed blue line corresponds to the input series of R-R intervals while the black line is the output series produced by our correction algorithm (when the two coincide, only the black line is visible). On the left, subplot A shows examples of "m" and "t" beats and the remarkable ability of our algorithm to correct the R-R series and recover the underlying rhythmic structure. Two ectopic beats were not detected by any of the algorithms because they do not cause a sufficient deviation from the normal rhythm. Very likely they were labelled as ectopic based on their morphology (unavailable to our algorithm). The letter "t" marks the case of two ectopic beats in a row followed by a compensatory pause. This is a particularly challenging case, as it requires predicting the time of occurrence of two beats at once. Subplot B shows a rapid sequence of three ectopic beats and how they were detected and corrected by the algorithm. Subplot C shows a resetting ectopic beat that our algorithm correctly detects as such and flags for further correcting action. Finally, in subplot D, there is an example of a sudden physiologic change in autonomic control that might trigger a false detection. In fact, our previous algorithm presented in [14] as well as all three reference algorithm presented here detect one or more of the beats around $t = 1470$ s as arrhythmic, while the new algorithm presented here does not.

both zero-mean noise terms. The variance of ϵ_1 is the same as that of $\hat{w}_{k+1}$, i.e., the variance of an inverse Gaussian distribution of parameters μ_1 and λ_1: μ_1^3/λ_1. The variance of ϵ_2 is $\mu_2^3(\hat{w}_{k+1})/\lambda_2$. We make the assumption that it can be approximated as $\mu_2^3(\mu_1)/\lambda_2$ and that ϵ_1 and ϵ_2 are uncorrelated. Using (1) we have $\hat{w}_{k+2} = \mu_2(\mu_1) + \tilde{\theta}_1(u_k)\,(\hat{w}_{k+1} - \mu_1) + \epsilon_2$. Therefore,

$$\tau - u_k = \hat{w}_{k+1} + \hat{w}_{k+2} = \mu_1 + \epsilon_1 + \mu_2(\mu_1) + \tilde{\theta}_1(u_k)\,\epsilon_1 + \epsilon_2 \quad (7)$$

which leads to $\mu_{1+2} = \mathrm{E}[\tau - u_k] = \mu_1 + \mu_2(\mu_1)$ and $\mathrm{Var}[\tau - u_k] = (1 + \tilde{\theta}_1(u_k))^2\,\mathrm{Var}[\epsilon_1] + \mathrm{Var}[\epsilon_2]$. This means $\mu_{1+2}^3/\lambda_{1+2} = (1 + \tilde{\theta}_1(u_k))^2\,\mu_1^3/\lambda_1 + \mu_2^3(\mu_1)/\lambda_2$. Using $\lambda_1 = \lambda_2$ and solving for λ_{1+2}, we obtain (4).

Following a similar reasoning, we obtain that the probability distribution of the sum of three unknown intervals can be modelled as an IG distribution with parameters $\mu_{1+2+3} = \mu_1 + \mu_2(\mu_1) + \mu_3(\mu_1, \mu_2)$ and

$$\lambda_{1+2+3} = \frac{\lambda_1\,\mu_{1+2+3}^3}{(1 + \tilde{\theta}_1 + \tilde{\theta}_2)^2\mu_1^3 + (1 + \tilde{\theta}_1)^2\mu_2^3(\mu_1) + \mu_3^3(\mu_1, \mu_2)}\,.$$

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Roger Mark for his valuable comments and suggestions that were of great help during the first stages of this work.

This work was supported by National Institutes of Health (NIH) through grants R01-HL084502 (R.B.) and DP1-OD003646 (E.N.B.).

REFERENCES

[1] A. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. Cohen, and P. Coumel, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology." *Circulation*, vol. 93, no. 5, p. 1043, 1996.

[2] U. Acharya, P. Joseph, N. Kannathal, C. Lim, and J. Suri, "Heart rate variability: a review," *Med. Biol. Eng. Comput.*, vol. 44, pp. 1031–1051, 2006.

[3] U. Acharya, J. Suri, J. Spaan, and S. Krishnan, *Advances in cardiac signal processing*. Springer Verlag, 2007.

[4] N. Lippman, K. M. Stein, and B. B. Lerman, "Comparison of methods for removal of ectopy in measurement of heart rate variability," *Am. J. Physiol. — Heart and Circulatory Physiology*, vol. 267, no. 1, pp. H411–H418, Jul. 1994.

[5] M. V. Kamath and E. L. Fallen, "Correction of the heart rate variability signal for ectopics and missing beats," in *Heart rate variability*, M. Malik and A. J. Camm, Eds., 1995, pp. 75–85.

[6] G. Clifford, P. McSharry, and L. Tarassenko, "Characterizing artefact in the normal human 24-hour RR time series to aid identification and artificial replication of circadian variations in human beat to beat heart rate using a simple threshold," in *Proceedings of Computers in Cardiology*, Memphis, Sep. 2002, pp. 129–132.

[7] M. N. Cheung, "Detection of and recovery from errors in cardiac interbeat intervals," *Psychophysiology*, vol. 18, no. 3, pp. 341–346, May 1981.

[8] M. Malik, T. Cripps, T. Farrell, and A. Camm, "Prognostic value of heart rate variability after myocardial infarction. a comparison of different data-processing methods," *Med. Biol. Eng. Comput.*, vol. 27, no. 6, pp. 603–611, 1989.

[9] G. G. Berntson, K. S. Quigley, J. F. Jang, and S. T. Boysen, "An approach to artifact identification: Application to heart period data," *Psychophysiology*, vol. 27, no. 5, pp. 586–598, Sep. 1990.

[10] E. J. M. Weber, P. C. M. Molenaar, and M. W. Molen, "A nonstationarity test for the spectral analysis of physiological time series with an application to respiratory sinus arrhythmia," *Psychophysiology*, vol. 29, no. 1, pp. 55–62, Jan. 1992.

[11] D. Sapoznikov, M. Luria, Y. Mahler, and M. Gotsman, "Computer processing of artifact and arrhythmias in heart rate variability analysis," *Computer methods and programs in biomedicine*, vol. 39, no. 1-2, pp. 75–84, 1992.

[12] M. Brennan, M. Palaniswami, and P. Kamen, "A new model-based ectopic beat correction algorithm for heart rate variability," in *Proceedings of the 23rd Annual International Conference of the IEEE EMBS*, vol. 1, 2001, pp. 567– 570.

[13] J. Mateo and P. Laguna, "Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 334–343, Mar. 2003.

[14] R. Barbieri and E. N. Brown, "Correction of erroneous and ectopic beats using a point process adaptive algorithm," in *Proceedings of the 28th Annual International Conference of the IEEE EMBS*, Sep. 2006, pp. 3373–3376.

[15] J. Rand, A. Hoover, S. Fishel, J. Moss, J. Pappas, and E. Muth, "Real-Time correction of heart interbeat intervals," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 946–950, May 2007.

[16] D. Widjaja, S. Vandeput, J. Taelman, M. Braeken, R. Otte, B. Van den Bergh, and S. Van Huffel, "Accurate R peak detection and advanced preprocessing of normal ECG for heart rate variability analysis," in *Proceedings of Computing in Cardiology*, Belfast, Sep. 2010, pp. 533–536.

[17] R. Barbieri, E. C. Matten, A. R. Alabi, and E. N. Brown, "A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability," *Am. J. Physiol. — Heart and Circulatory Physiology*, vol. 288, pp. H424–435, 2005.

[18] L. Citi, E. N. Brown, and R. Barbieri, "A point process local likelihood algorithm for robust and automated heart beat detection and correction," in *Proceedings of Computing in Cardiology*, Hangzhou, Sep. 2011.

[19] L. Citi and R. Barbieri, "Point process detection and correction of erroneous and ectopic heart beats," 2012. [Online]. Available: http://users.neurostat.mit.edu/barbieri/ncspu

[20] G. Moody, "Spectral analysis of heart rate without resampling," in *Proceedings of Computers in Cardiology*. IEEE, 1993, pp. 715–718.

[21] A. Guyton and J. Hall, *Textbook of medical physiology*. Saunders Philadelphia, PA, 1991, vol. 9.

[22] B. Hyndman and R. Mohn, "A model of the cardiac pacemaker and its use in decoding the information content of cardiac intervals," *Automedica*, vol. 1, pp. 239–252, 1975.

[23] R. DeBoer, J. Karemaker, and J. Strackee, "Description of heart-rate variability data in accordance with a physiological model for the genesis of heartbeats," *Psychophysiology*, vol. 22, no. 2, pp. 147–155, 1985.

[24] R. Berger, S. Akselrod, D. Gordon, and R. Cohen, "An efficient algorithm for spectral analysis of heart rate variability," *IEEE Trans. Biomed. Eng.*, no. 9, pp. 900–904, 1986.

[25] J. Mateo and P. Laguna, "Improved heart rate variability signal analysis from the beat occurrence times according to the ipfm model," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 8, pp. 985–996, 2000.

[26] E. Brown, *Theory of point processes for neural systems*, ser. Methods and Models in Neurophysics. Elsevier, 2005, ch. 14, pp. 691–726.

[27] J. Feng, *Computational neuroscience: a comprehensive approach*. CRC press, 2004.

[28] R. Chhikara and L. Folks, *The inverse Gaussian distribution: theory, methodology, and applications*. CRC, 1989, vol. 95.

[29] G. Gerstein and B. Mandelbrot, "Random walk models for the spike activity of a single neuron," *Biophysical Journal*, vol. 4, no. 1, pp. 41–68, 1964.

[30] H. Tuckwell, *Introduction to theoretical neurobiology: Nonlinear and stochastic theories*. Cambridge Univ Pr, 2005, vol. 2.

[31] G. Stanley, K. Poolla, and R. Siegel, "Threshold modeling of autonomic control of heart rate variability," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1147–1153, 2000.

[32] P. Andersen, *Statistical models based on counting processes*. Springer Verlag, 1993.

[33] C. Loader, *Local regression and likelihood*. Springer Verlag, 1999.

[34] N. Iyengar, C. Peng, R. Morin, A. Goldberger, and L. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *Am. J. Physiol. — Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 4, pp. R1078–R1084, 1996.

[35] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[36] K. Solem, P. Laguna, and L. Sörnmo, "An efficient method for handling ectopic beats using the heart timing signal," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 13–20, 2006.
[37] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, Jun. 2001.
[38] R. Barbieri and E. Brown, "Analysis of heartbeat dynamics by point process adaptive filtering," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 4–12, 2006.
[39] Z. Chen, E. Brown, and R. Barbieri, "Assessment of autonomic control and respiratory sinus arrhythmia using point process models of human heart beat dynamics," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1791–1802, 2009.
[40] ——, "Characterizing nonlinear heartbeat dynamics within a point process framework," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1335–1347, 2010.

Luca Citi (S'06–M'09) was born in Pistoia, Italy, in 1979. He received a laurea (MS) degree in Electronic Engineering with major in Biomedical Engineering, in 2004 from Università degli Studi di Firenze (Florence, Italy). In 2009, he received his PhD in Biorobotics Science and Engineering jointly offered by IMT Lucca (Lucca, Italy) and Scuola Superiore Sant'Anna (Pisa, Italy).

He is currently a post-doctoral research fellow at Massachusetts General Hospital / Harvard Medical School and affiliated with the Massachusetts Institute of Technology (Cambridge, MA, U.S.A.) where he is part of the "Neuroscience Statistics Research Laboratory" led by Prof. Emery Brown, and the "Neuro–Cardiovascular Signal Processing Unit" led by Dr. Riccardo Barbieri. He works on the statistical analysis of human cardiovascular rhythms and of neural spike trains. His work on HRV focuses on devising and refining stochastic models of cardiovascular control where heart beat intervals are governed by a history-dependent point process. Other interests include signal processing in computational neuroscience and neural interfaces.

Emery N Brown (M'01, SM'06, F'08) received the B.A. degree from Harvard College, Cambridge, MA, the M.D. degree from Harvard Medical School, Boston, MA, and the A.M. and Ph.D. degrees in statistics from Harvard University, Cambridge, MA. He is currently a Professor of computational neuroscience and health sciences and technology in the Department of Brain and Cognitive Sciences and the Division of Health Sciences and Technology, Harvard-Massachusetts Institute of Technology, the Warren M. Zapol Professor of Anesthesia at Harvard Medical School, and an Anesthesiologist in the Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital. His research interests include signal processing algorithms for the study of neural systems, functional neuroimaging, and electrophysiological approaches to study general anesthesia in humans. Dr. Brown is a member of the Association of University of Anesthesiologists. He is also a Fellow of the American Statistical Association, a Fellow of the American Association for the Advancement of Science, a member of the Institute of Medicine, and a Fellow of the American Academy of Arts and Sciences. He was a recipient of a 2007 National Institutes of Health Director's Pioneer Award and the 2011 recipient of the National Institute of Statistical Sciences Sacks Award for Outstanding Cross-Disciplinary Research.

Riccardo Barbieri (M'00, SM'08) was born in Rome, Italy, in 1967. He received the M.S. degree in Electrical Engineering from the University of Rome "La Sapienza", Rome, Italy, in 1992, and the Ph.D. in Biomedical Engineering from Boston University, Boston, MA, in 1998.

Dr. Barbieri is currently Assistant Professor of Anaesthesia at Harvard Medical School, Massachusetts General Hospital, Research Affiliate at Massachusetts Institute of Technology and Principal Investigator of the Neuro-Cardiovascular Signal Processing Unit within the Neuroscience Statistics Research Laboratory. His main research interests include: modeling of biological systems (with emphasis on the cardiovascular system, the central nervous system and the autonomic nervous system); biomedical signal processing (with emphasis on cardiovascular control and cardiovascular variability signals); and signal processing in computational neuroscience (EEG, Evoked Potentials, single cell neuronal recordings).