Robust, real-time generic detector based on a multi-feature probabilistic method

Matthieu Doyen, Di Ge, Alain Beuchée, Guy Carrault, Alfredo Hernández

Robust, real-time generic detector based on a multi-feature probabilistic method

M. Doyen, D. Ge, *Member, IEEE*, A. Beuchée, G. Carrault and A.I. Hernández, *Member, IEEE.*

***Abstract*—*Objective:* Robust, real-time event detection from physiological signals acquired during long-term ambulatory monitoring still represents a major challenge. In this paper, we propose an original and generic multi-feature probabilistic detector (MFPD) and apply it to real-time QRS complex detection. *Methods:* The proposed method first derives relevant features from one or more physiological signals and estimates their probability distributions. Bayesian probabilities are then calculated for each feature before being merged through a data fusion node, with regard to the Kullback-Leibler divergence measure. To this end, we derived a KLD estimation method between two Generalized Normal Distributions (GND), in the general case. The efficiency of our method is validated on two noisy ECG databases : 1) a benchmark database by adding noise recordings[23] to the entire MIT-BIH arrhythmia database with noise levels ranging from -6dB to 24dB, 2) an exercise stress test database composed of 54 real ECG recordings (17.8 hours). *Results:* In both cases significant improvements in detection performance are obtained compared with reference methods. For the benchmark noise stress database, performance gains are consistent for all noise type and SNR levels, and are increasing in low SNR levels. For the exercise stress database, detection error criterion is lowered to 20.91%, as compared with Pan-Tompkins-based (33.08%) and wavelet-based QRS detector (29.02%). *Conclusion:* Thanks to its multi-feature aspect and its KLD-based decision method (able to adaptively adjust the relative contribution of each feature to the final decision in real-time), the proposed method yields some interesting results, especially on highly-artifacted signals. *Significance*: Robust, easy to implement and with a reasonable computational cost, MFPD offers some promising perspectives for various challenging monitoring situations.**

***Index Terms*—Robust probabilistic detection, Centralized fusion, QRS detection, Kullback-Leibler divergence (KLD).**

I. Introduction

EVENT detections from physiological signals are often faced with important noise perturbations, especially in clinical monitoring context. Main strategy is often focused on finding an efficient feature reliable in most cases. Generally, these methods get interesting results under low- to mid-level noise conditions [1], but performances decrease significantly with the signal-to-noise ratio (SNR) diminution or with a change in the noise type since all features have vulnerabilities to specific distortions. To circumvent this weakness, multi-feature detectors were proposed [2] but the decentralized fusion method does not permit to fully exploit feature informations. The objective of this paper is to propose a centralized event detection method using multiple features' distributions and to apply it to real-time QRS complex detection from electrocardiogram (ECG) signals.

QRS complex is the most prominent deflection in ECG signal and corresponds to the electrical depolarization of ventricles. The detection is often the first analysis performed on ECG signal processing, in order to estimate basic cardiac markers, such as heart rate or to perform further ECG segmentation and analysis. The QRS complex detection has been investigated for many decades [1] and yet remains a challenge [3] as an event detection problem from physiological signals. Many different methods have been proposed and a number of review publications have been dedicated to this subject [4] [5]. The main proposed methods are based on filtering and non linear transformations [1], fuzzy hybrid neural networks [6], S-Transform [7] or wavelet analysis [8], [9], [10], [11]. Although these QRS detection methods perform well in low- to mid-level noise conditions, their applications on long-term periods of ECG recordings in ambulatory care or in intensive care units still pose a significant challenge. Indeed, these ECG recordings are often prone to episodes of strong signal non-stationarity, sudden modifications of beat morphologies and most importantly the presence of several types of noise (baseline drift, saturation, power-line pickup, muscular contractions and motion artifacts [12]). Recent publications [13] [14] [15] and a recent PhysioNet challenge [3] have been focused on the specific problem of robust QRS detection. Furthermore, the emergence of wearable cardiac monitors, with a limited number of leads [16] [17] for long-term daily-life recordings [18] further revives the research interests on this subject.

In our previous works, we have proposed different methods to improve the robustness of QRS detection, through multi-sensor fusion [2], adaptive selection of QRS detectors as a function of the signal context [19] or through optimal detector parameter configuration, using evolutionary methods [20] [21]. More recently, we revisited this optimization process in order to identify optimal parameter configurations with respect to changes in signal noise [22].

In this paper, we propose and evaluate a novel, generic event detector, that provides improved robustness through the probabilistic combination of a set of signal features. Section II presents the general architecture of the proposed Multi-Feature Probabilistic Detector (MFPD) and a specific implementation adapted to robust QRS detection. Section III evaluates the

M. Doyen, D. Ge (correspondence e-mail: Di.Ge@univ-rennes1.fr), G. Carrault and A. I. Hernández are with the INSERM, U1099, Rennes, F-35000, France and also with Université de Rennes 1, LTSI, Rennes, F-35000, France.

A.Beuché is with with CHU Rennes, Rennes F-35000, France, with the INSERM, U1099, Rennes, F-35000, France and also with Université de Rennes 1, LTSI, Rennes, F-35000, France. Rennes F-35000, France

The authors would like to thank the french Ministry of Higher Education and Research for its support.

Binary executables of the QRS detection application and their user manuals are available at https://physionet.org/physiobank/database/nstdb/.

reliability of MFPD on two databases: the MIT noise stress database [23] and an exercise stress test database composed of 54 real ECG signals. Performance obtained by MFPD is compared with a wavelet-based [24] [22] and a Pan-Tompkins-based [20] QRS detectors, using both the default detector parameters and optimized versions of these parameters, obtained through evolutionary algorithm [20].

II. Methods

A. General architecture of the detector

The general architecture of the MFPD is depicted in Fig. 1. It is based on the following steps:

1) Pre-processing: Raw signals are processed in order to improve the SNR and to pre-select potential candidates (to be validated by the detector) of the events of interest at instants t.
2) Feature extraction: For every event candidate selected at instant t, a vector $C(t) = \{C_i(t)|i \in \mathcal{I}\}$ is created, where $\mathcal{I}$ is a set of complementary features extracted from the preprocessed signals.
3) Probability density estimation: The probability density functions (pdf), noted as $P_i(C_i(t); \Theta_{i0/1}, \mathscr{H}_{0/1})$ are used to model feature i on the observed candidate $C(t)$, with the 2 hypothesis :

$$\mathscr{H}_0 : D(t) = 0,$$
$$\mathscr{H}_1 : D(t) = 1,$$

where $D(t)$ is the final detection decision ($D(t) = 1$ for positive detection and $D(t) = 0$ otherwise) and $\Theta_{i0/1}$ the parameter set for each hypothesis. Note that for each feature, the two pdf belong to the same distribution family, whose parameters $\Theta_{i0/1}$ are initialized at the beginning of the recording and updated throughout the detection process.
4) Probabilistic characterization: The posterior probability $P_i(\mathscr{H}_1|C_i(t))$ is calculated by applying the Bayes law. Moreover, the Kullback-Leibler divergence (KLD) between each pdf pair characterizing feature i, D^i_{KL}, is calculated.
5) Decision fusion: The posterior probabilities, weighted by their respective KLD, are combined to build a binary decision $D(t)$ on whether candidate $C(t)$ is a valid event ($D(t) = 1$) or not ($D(t) = 0$). According to the decision for the current candidate, distribution parameters are updated to complete the real-time learning process.

In the following, we detail the realization of the above-mentioned MFPD, adapted for real-time detection of QRS complexes. From the generic approach of Fig. 1, the specific adaptions to QRS complex detection concern mainly the preprocessing (step 1) and the feature extraction (step 2), they are depicted in Fig. 2.

B. Pre-processing the ECG signal

As in many other QRS detection methods, the first step consists in applying to the raw ECG signal different transformations. Typically, a band-pass filter, a derivative filter, a non-linear transformation and a final smoothing filter are applied. Fig. 2 represents a diagram of these signal processing steps and Fig. 3 shows a representative example. In the following, the band-pass filtered (low-pass then high-pass filter) ECG signal will be denoted SAecg; and the output of a squared transformation followed by a derivative then a smoothing filter on signal SAecg will be denoted SFecg. Each local maximum (or positive peak) detected at an instant t on SFecg is considered as a potential QRS candidate. It is a common pre-processing method previously used in [1] [25]. A set of features ($\mathcal{I}$) is extracted from signals SAecg and SFecg around instant t in order to estimate, through the proposed MFPD, if the candidate at time t is a valid QRS ($D(t) = 1$) or not ($D(t) = 0$).

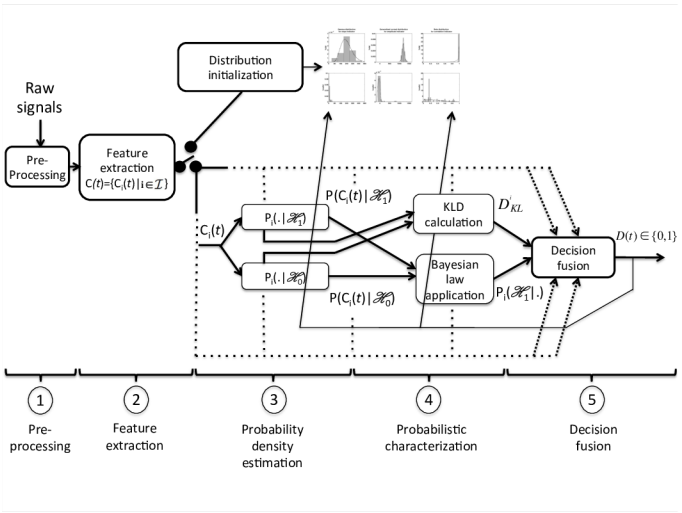

Figure 1: *Diagram representing the global architecture of the MFPD. In process step, data coming from ECG (raw and filtered) are converted into features. $P_i(.|\mathscr{H}_1)$ and $P_i(.|\mathscr{H}_0)$ are probabilistic parametric models of feature i, representing valid and invalid detections.*

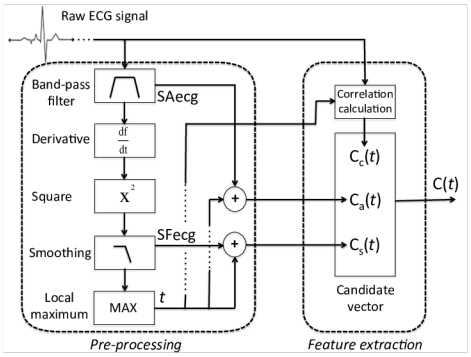

Figure 2: *Specification of signal pre-processing and feature extraction (steps 1 and 2 in Figure 1), for the application of robust QRS detection.*

C. Feature extraction and probability distributions

In the proposed QRS detection application, each candidate is characterized by a set of 3 features $\mathcal{I} = \{s, a, c\}$. The

following features and probability distribution functions have been selected for this application:

- The squared slope of the peak (s) is the value of SFecg signal at instant t. This feature is represented with the Gamma distribution with two degrees of freedom, defined as:

$$P_s(x;k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\Gamma(k)\theta^k}\mathbb{1}_{x>0}, \tag{1}$$

  where $k \in \mathbb{R}^+$ is the shape parameter, and $\theta \in \mathbb{R}^+$ the scale parameter. The indicator function $\mathbb{1}_{x>0}$ typically limits the function support to $\mathbb{R}^+$.

- The peak amplitude (a) is the value of SAecg signal at instant t. We characterized it using the Generalized Normal Distribution (GND) defined as:

$$P_a(x;\alpha;\beta;\mu) = \frac{\beta}{2\alpha\Gamma(1/\beta)}\, e^{-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}} \tag{2}$$

  where Γ is the gamma function

$$\Gamma(t) = \int_0^{\infty} x^{t-1}e^{-x}\,dx \tag{3}$$

  $\mu \in \mathbb{R}$ is the position parameter, $\alpha \in \mathbb{R}^+$ the scale parameter and $\beta \in \mathbb{R}^+$ the shape parameter. Note that both positive and negative peak values are considered with the GND model.

- the Bravais-Pearson correlation (c) is calculated between the candidate peak (represented by 50 ms of raw ECG signal centered 20 ms before the peak) at instant t and an adaptive template. In order to model this feature, we have chosen the Beta distribution, defined as:

$$P_c(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}\ \mathbb{1}_{[0,1]}(x). \tag{4}$$

  This can also be considered as a special case of the Dirichlet distribution, with two positive shape parameters α and β. Parameters are estimated using maximum a posteriori (MAP) method, further details are reported in section III-B.

Note again that, for each feature, the same probability model structure is proposed for both $\mathscr{H}_0$ and $\mathscr{H}_1$ while model parameters $\Theta_{i0/1}$ are initialized during the heating-up period then updated throughout the detection process according to the final decision.

D. Probabilistic characterization of the the candidates

According to the probability distributions defined in II-C, two probabilistic markers are calculated for each feature: the posterior probability and the KLD. The posterior probability of validating $\mathscr{H}_1$ for a given feature $C_i(t)$ is given by

$$P_i(\mathscr{H}_1|C_i(t)) = \frac{P_i(C_i(t)|\mathscr{H}_1)\,P_i(\mathscr{H}_1)}{P_i(C_i(t)|\mathscr{H}_1)P_i(\mathscr{H}_1) + P_i(C_i(t)|\mathscr{H}_0)P_i(\mathscr{H}_0)}$$

using the Bayes rule.

The KLD is a non-negative measure defined by:

$$D_{\mathrm{KL}}(p\|q) = \int_{-\infty}^{\infty} p(x)\log\frac{p(x)}{q(x)}\,dx \tag{5}$$

for continuous distributions. It is particularly well-suited to assess the distance between each distribution pair $P_i(C_i(t);\Theta_{i0/1};\mathscr{H}_{0/1})$. Analytic expressions for (5) can be found in the literature in the case of Beta [26] and Gamma distributions [27]. However for the GND case and to our best knowledge, no analytic expression can be found in the general case, especially when $\mu_p \neq \mu_q$. One theoretical contribution of this article is to efficiently calculate the KLD between 2 GND distributions in the case of general settings. While detailed derivation is reported in the appendix B, we give a summary of the main results here:

1) Analytic expressions for $\beta_q \in \mathbb{N}^+ \cup \{0\}$ have been derived;
2) We demonstrated that equation (5) is monotonous with respect to β_q;
3) The computational complexity requires $2 \times (\beta_q + 1)$ gamma function evaluations.

Thus, we are able to obtain a close approximation of the KLD value for all $\beta_q \in \mathbb{R}^+$. Note that the other parameters (α_p, α_q, β_p, μ_p, μ_q) have no effect on the calculation method.

E. Decision fusion

Based on the probabilistic markers, the following decision rule is applied

$$\begin{aligned} &D(t) = 1 \text{ iff} \\ &\sum_{i\in\mathcal{I}} \frac{\overline{D}^i_{\mathrm{KL}}.P_i(\mathscr{H}_1|C_i(t))}{\sum_{j\in\mathcal{I}}\overline{D}^j_{\mathrm{KL}}} > \lambda \end{aligned} \tag{6}$$

where the normalized $\overline{D}^i_{\mathrm{KL}}$ is calculated by letting $i^* = argmax\{D^i_{KL}\}$ and :

$$\overline{D}^j_{KL} = \begin{cases} min\left\{D^j_{KL}, 2\sum\limits_{h\neq i^*} D^h_{KL}\right\}, & j = i^* \\ D^j_{KL}, & j \neq i^* \end{cases}$$

such that the most significant contributor should not exceed 2/3 after normalization.

Intuitively, the decision rule represents the sum of all posterior probabilities, weighted by their normalized KLD, such that features that are better separated in distributions (between $\mathscr{H}_0$ and $\mathscr{H}_1$) have more weight in the final decision making.

F. Evaluation methodology

1) Database: Two noisy ECG databases were used:

a Simulation database: three noise types (baseline wander, muscle artifact and electrode motion artifact) from recordings acquired on physically active volunteers in the MIT noise stress database [23] were added to the first lead of the 48 ECG of the MIT-BIH Arrhythmia Database, with SNR levels from -6dB to 24dB and a constant step of

6dB. We thus created a benchmark simulated database composed of 864 noisy signals [1], for which the reference annotations are simply copies of those for the original clean ECGs in the MIT-BIH Arrhythmia Database. The purpose of this database is to provide a ground truth for detection performance comparison with different levels and types of noises.

b Test database: A real exercise stress test database composed of real ECG signals (17.8 hours) of 54 patients recorded using an ergocycle (sampling frequency 1000 Hz using a Cardionics system) at the University Hospital of Rennes in France was also used for validation. Manual QRS complex annotations were realized by a trained operator. We aimed at further validating the method on real ECG signals acquired under ambulatory situations with a dominance of muscle and electrode motion artifacts.

2) Comparison methods: As in recent publications [28] [13], performance of the proposed MFPD was compared with the following QRS detection methods:

- WBD: a wavelet-based QRS detector [24] for the test database
- PTM: an improved Pan-Tompkins detector with optimized Remez-based filter coefficients[20] for both the simulation and test database.

The performance of all three detectors (MFPD, WBD and PTM) are impacted by the tuning of several parameters (threshold adjustments, cutoff frequencies, order of filters, window sizes, etc). All these parameters were optimized by an evolutionary algorithm [20] on a subpart of the simulation database (25% for MFPD and PTM, see [22] for WBD). Detectors with the obtained optimal parameters are noted WBD*, PTM* and MFPD*. Note that the 3 detectors with default and optimized parameters were evaluated on the test database.

3) Performance criterion: To compare the reliability of our detector to others, we used the detection error criterion defined as:

$$C_{err} = \sqrt{(1-Se)^2 + (1-PPV)^2} \tag{7}$$

where Se is the sensitivity and PPV is the positive predictive value [22]. This criterion is preferred to overall accuracy, the latter being inappropriate in this case due to the very low prevalence of the QRS complex event [29]. Knowing that a QRS complex normally lasts $90-100$ ms in adults [30], a detection is considered as a true positive (TP) when it is located within a 50 ms window centered in the reference QRS complex annotation. All the other detections are considered as false positives (FP). Moreover, the ROC curve (with 1-PPV in abscissa) is depicted for experimental results to provide a complete comparison of the performances.

III. Results

In addition to the global performance evaluation and comparison summarized in III-D, we also provide here some intermediate results to illustrate the multi-feature complementarity in III-A, their distribution estimation results in III-B, and the importance of KLD weighting in the centralized decision making in III-C.

A. Multi-feature complementarity

Figure 3 shows an example of the processed ECG signals and the features extracted in this paper. Panel (a) shows an ECG segment from record MIT-101, with added electrode noise at 6 dB. Panels (b) and (c) represent, respectively, the SAecg and SFecg signals obtained from the ECG in panel (a). Each peak (QRS candidate) detected from the SFecg signal is marked with a symbol (either $\times$ or $\circ$) and these symbols have been projected to the other signals in the figure. Features s and a are directly computed for each candidate as the values of signals SFecg and SAecg at the instant of the corresponding peak. Panel (d) represents the Bravais-Pearson correlation (feature c), calculated for each detected peak. Using these features for each candidate, a global detection is performed. Peaks with symbol $\times$ are the ones that have been invalidated by our method as a QRS, while symbol $\circ$ represents the candidates that have been selected as QRS detections. In this example, all validated candidates were TP and all invalidated candidates were TN.

The vertical box in Figure 3 shows a set of features that, if analyzed individually with simple thresholds, would have produced a false positive. This example shows the utility of the proposed MFPD in this complex signal context.

B. Distribution estimation

Estimated and empirical distributions for each feature are depicted in Fig. 4. The dashed line represents $P_i(C_i(t);\Theta_{i0/1},\mathscr{H}_{0/1})$, depicted jointly with their normalized histogram in grey. In some cases, because of the high stability of the QRS complex morphology on mildly-artifacted signals, $C_c(t)$ can tend to 1 for most of valid candidates. Thus large α values are obtained for $P_c(C_i(t);\Theta_{c1},\mathscr{H}_1)$ yielding a rapid convergence towards a dirac-like distribution around 1, and a net separation in distribution (large D^c_{KL}) from $P_c(C_i(t);\Theta_{c0},\mathscr{H}_0)$ (cf third column in Fig. 4). As a consequence, future candidates must have a very high correlation (above 0.9999) to be validated, giving a decision with a high PPV, but with a low sensitivity in the case of sudden noise artifacts or even mild morphology change. In order to obtain a better trade-off between PPV and sensitivity, we propose to limit the estimated α parameter by imposing a conjugate prior law on the parameters:

$$P(\alpha,\beta) \propto B(\alpha,\beta)^K e^{-a\alpha} e^{-b\beta}$$

for $K,a,b \in \mathbb{N}^+$. The $\exp^{-a\alpha}$ term indeed forbids large values in estimating α to control the distribution shape of $\mathscr{H}_1$. Numerical implementation is detailed in appendix A .

C. KLD weighting

In this section, we present results showing the relevance of the proposed MFPD method, integrating a KLD, with respect to mono-featured method.

[1]For reproducibility, we provide a quick access to this benchmark test database (https://physionet.org/physiobank/database/nstdb/) though it can also be constructed using the method described in [23]

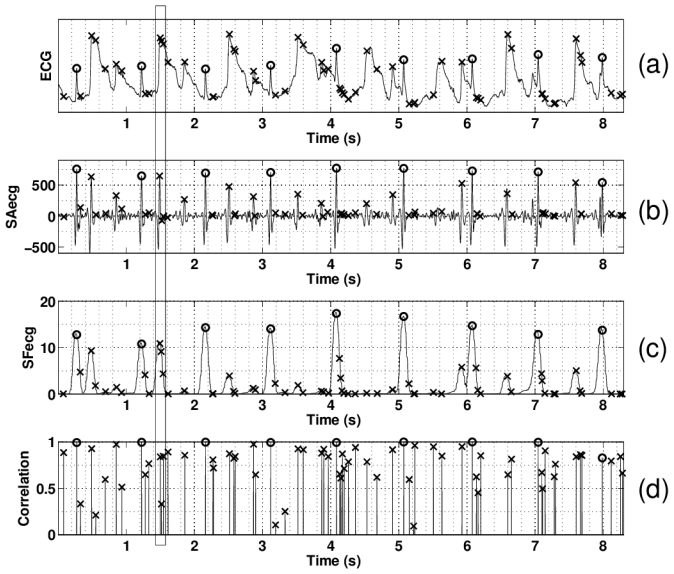

Figure 3: *Example of the processed ECG signals and the features extracted from each QRS candidate. a) Raw ECG segment from record MIT-101 with added electrode noise (SNR=6 dB). The raw ECG is processed to obtain signals SAecg and SFecg, represented respectively in (b) and (c). Symbols $\times$ and $\circ$ represent QRS candidates, detected as peaks in signal SAecg. Events annotated as true positives are marked with symbol ($\circ$). Features a and s are computed for each candidate from signals SAecg and SFecg, while the values of feature c (the Bravais-Pearson correlation) are presented in panel (d). In this example, all validated candidates are TP and all invalidated candidates are TN. The vertical box shows a temporal support during which the application of a simple threshold to each feature independently would have caused FP or FN.*

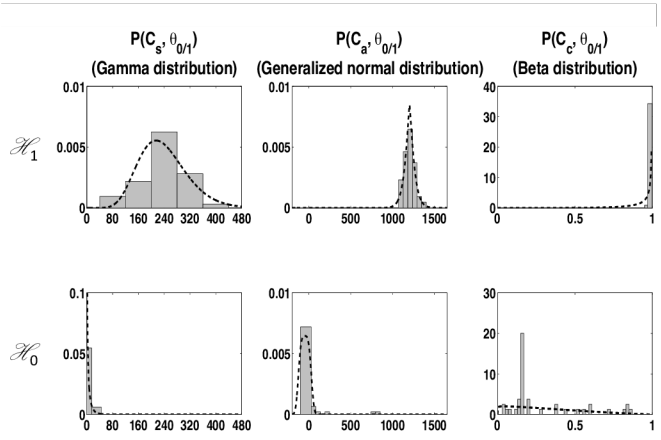

Figure 4: *Distribution estimation extracted from record MIT-101 with added baseline noise (SNR=12 dB). Dash lines representing the estimated distributions are compared with the normalized histograms of the data samples (gray boxes). Each column corresponds to a particular feature; at the top, distributions for validated candidates ($\mathscr{H}_1$) and at the bottom, distributions for invalidated candidates ($\mathscr{H}_0$).*

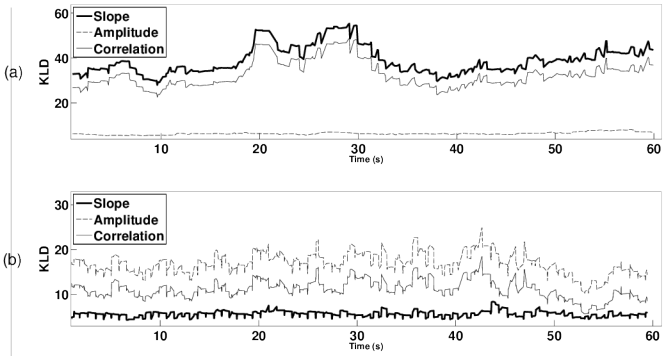

Figure 5: *KLD normalized variations for successive candidates (QRS complex or not) computed on signal segments of 60 s. a) segment from record MIT-109 with baseline noise added with SNR=18 dB. b) segment from record MIT-102 with added muscle artifact with SNR=6 dB. Bold, dashed and solid lines represent respectively the variations of $\overline{D}^s_{KL}$, $\overline{D}^a_{KL}$ and $\overline{D}^c_{KL}$.*

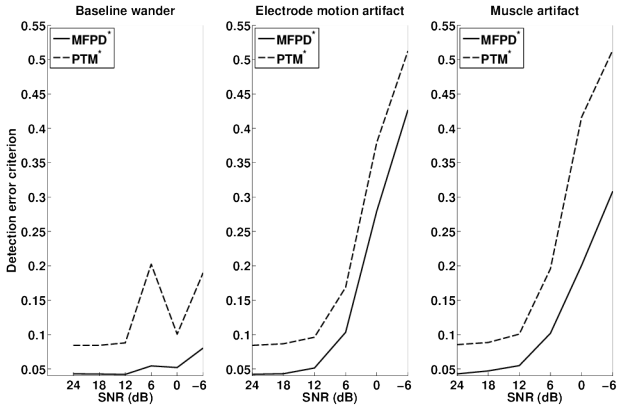

Figure 6: *Performance of PTM* and MFPD* on the whole MIT noise stress database, with different types of noise and with different SNR levels.*

Fig. 5a represents the time series of $\overline{D}^{\mathcal{I}}_{KL}$ during 60 s obtained from record MIT-109 (baseline noise added with SNR=18 dB). $\overline{D}^s_{KL}$ values (bold line) are higher than those of $\overline{D}^a_{KL}$ (dash line) and close to $\overline{D}^c_{KL}$ values (solid line). In this example, the performance of the PTM* detector, which only exploits the square slope feature, (PPV = 100%, sensitivity= 97.67%) is close to that obtained by the proposed MFPD* (PPV = 99.56%, sensitivity= 99.09%). This result was expected, since the values of $\overline{D}^s_{KL}$ are high in this case. On the contrary, in Fig. 5b (MIT-102, muscle artifact with SNR=6 dB), $\overline{D}^s_{KL}$ values are lower than those of $\overline{D}^a_{KL}$ and $\overline{D}^c_{KL}$, leading to a degraded performance of the PTM* detector (PPV = 88.66%, sensitivity = 70.10%) in comparison with the MFPD* method (PPV = 91.36%, sensitivity = 91.65%). These examples further prove the concept of multi-feature complementarity to confront different noise types and the importance of KLD weighting in the centralized decision making to improve the overall detection performance.

D. Performance evaluation

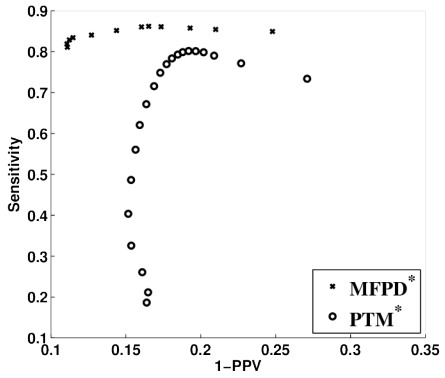

Figure 7: *Comparison of MFPD* and PTM* performances on [23], with 1-PPV in abscissa and sensitivity in ordinate.*

1) Evaluation 1 on the simulation database: A first performance analysis of MFPD* and PTM* was applied using the simulation database. Fig. 6 compares the detection error criterion (C_{err}) for different noise types and noise levels. Firstly, results show that MFPD* performances are always higher than that of PTM* for all noise types and SNR levels. Secondly, for all noise types, its performance gain increases with decreasing SNR, especially in the case of muscle artifact ($C_{err} = 51.37\%$ for PTM* vs 30.84% for MFPD*). Finally, we can notice that performances on baseline wander noise are higher than with other kinds of noise for both detectors since QRS complex morphology is less affected. These preliminary results demonstrate the MFPD viability and highlight its efficiency in noisy context.

2) Evaluation 2 on the test database: The objective here is to further validate the MFPD* detector on an independent database with ECG signals recorded using different materials on different subjects and under different clinical settings. The ROC curves were plotted for MFPD* and PTM* in Fig. 7 (information not provided in [22] for WBD*). This figure clearly shows the better performance of the proposed MFPD method, especially when high PPV values are considered. Indeed, the PTM* method cannot be used for PPV higher than 85% without sacrificing heavily on the sensitivity. Finally in table I, all detection performances in C_{err} were compared using the test database before and after parameter optimization obtained with the experimental database. MFPD provides always a better performance than PTM and WBD detectors, with or without parameter optimization.

Detector	C_{err}
PTM	38.46%
PTM*	33.08%
WBD	30.84%
WBD*	29.02%
MFPD	22.52%
MFPD*	20.91%

Table I: C_{err} computed on the whole test database.

IV. Discussion

In this paper, a novel, generic and robust detector combining different features extracted from the signal of interest has been proposed. The original aspects of this method concern particularly i) the probabilistic approach with online learning ii) the multi-feature design iii) the centralized fusion method based on KLD.

In the proposed method, the pdf of each feature is patient, device and even experience specific, a particular attention has been paid to find an appropriate online estimation method. Parametric probability models are particularly suited with regard to the real-time constraints of our application. It avoids both the tuning of the number of bins and widths as in histogram based approaches, and the increasing evaluation costs, inherent to variable-bandwith kernel density estimation [31]. Our proposed online learning method requires a small data sample (40 validated candidates) to initiate the parametric estimation, given the appropriate prior laws of the statistical models. Among the wide variety of pdf families, the GND seems well-suited for uncentered features, but to our best knowledge, no analytic expression of the KLD between 2 GND can be found in the general case. Existing solutions [32] are limited to cases where the mean values (μ) are equal (see II-D). We proposed in this paper an innovative estimation method of the KLD in the general case. With a reasonable computational cost, it can be used in real-time context.

The proposed MFPD QRS detector has been evaluated using two different databases and compared with two other detectors from the literature. Results show that the selected features provide complementary information to improve detection performance compared with other single featured QRS detectors in the literature (see Fig. 5), particularly under noisy conditions. This is essentially due to the fact that different kinds of noise or uncommon pathological phenomena affect extracted features in different ways. Previously proposed methods based on decentralized fusion [2] and algorithm-switching [19] also prove the relevance of multi-feature approaches.

The proposed KLD weighting method is able to adaptively adjust the relative contribution of each feature to the final decision. To our knowledge, this is the first real-time QRS detection method integrating such an adaptive, multi-feature, centralized decision fusion. Furthermore, MFPD method is more compact and easy to implement than [2] or [19]. Quantitative comparisons with the well-known Pan & Tompkins detector, as well as a wavelet-based detector, have shown that the proposed approach provides the best performance in all noise conditions. These results are particularly encouraging for challenging monitoring situations, in which the heterogeneity and levels of noise may be particularly high.

Even though this method was implemented in the case of a single-lead application, it can be easily extended to the multi-lead and multi-source case. Indeed, further improvements in detection robustness are expected from the combined use of multiple ECG leads, but also by integrating other physiological signals (pulse oximetry, phonocardiography, etc..) or other sensors sensitive to noise (accelerometers for movement noise, etc..). Future works will be directed towards the extension and evaluation of the method in these multi-channel, multi-source contexts. Finally, in addition to the qualitative results of the selected features' relevance shown in Fig. 3 and III-D2, an independence measure/test of the features shall be performed

quantitatively to further improve the detector robustness under noise conditions.

V. Conclusion

We proposed an original multi-feature probabilistic detector working in real-time and applicable to different physiological signal applications. The method, illustrated on QRS complex detection, has been compared to two common detectors of the the literature, using the MIT noise stress database and an exercise stress test database. Parameters of all detectors were optimized using an evolutionary algorithms [21] [22]. The proposed MFPD has achieved significant performance improvements compared with reference detectors, especially on highly-artifacted signals. These performance improvements are mainly due to the multi-feature aspect and the probabilistic approach using a KLD-based decision method that adaptively adjust the relative contribution of each feature in real-time.

Besides the theoretical contribution and experimental validation, this new approach boasts a reasonable computational cost and thus can be embedded into low-power devices offering interesting possibilities in the current context of connected health applications.

Appendix A
Application of the Karush-Kuhn-Tucker on the MAP of Beta distribution

1) About the KKT: Consider the non-linear optimization problem: maximize $f(x), \mathbb{R}^n \to \mathbb{R}$ (the cost function) subject to m inequality constraints and l equality constraints :

$$g_i(x) \leq 0, \text{ for i=1,..., m}$$
$$h_j(x) = 0, \text{ for j=1, ..., l.}$$

Suppose further that both the objective function $f(x)$ and the constraint functions $g_i(x)$, $h_j(x)$ are continuously differentiable at a point $\tilde{x}$. If $\tilde{x}$ is a local maximum of $f(x)$ that satisfies some regularity conditions, then there exist the KKT multipliers:

$$\mu_i \in \mathbb{R}, i = 1, \ldots, m \text{ and } \lambda_j \in \mathbb{R}, j = 1, \ldots, l, \text{ such that:}$$

$$\nabla f(\tilde{x}) = \sum \mu_i \nabla g_i(\tilde{x}) + \sum \lambda_j \nabla h_j(\tilde{x}); \quad (8)$$
$$g_i(\tilde{x}) \leq 0, \ \mu_i \geq 0, \ \mu_i g_i(\tilde{x}) = 0 \text{ for all } i = 1, \ldots, m \quad (9)$$
$$h_j(\tilde{x}) = 0 \text{ for all } j = 1, \ldots, l \quad (10)$$

We note that in the particular case of $m = 0$, the KKT conditions are reduced to the Lagrange conditions. Typically, if both g_i and h_j are affine functions (the case of MLE for the beta distribution), then no other condition is needed.

2) MAP for the Beta distribution: We derive here the numerical method to achieve the maximum-likelihood estimator (MLE) and MAP estimator given N independent samples of the Beta distribution, whose probability density function writes:

$$P(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \mathbb{1}_{0 \leq x \leq 1}.$$

where $B(\alpha, \beta)$ is a normalization constant. For the MLE, the function to maximize is the joint log-likelihood function:

$$f(\alpha, \beta) = \sum_{i=1}^{N} \log P(x_i; \alpha, \beta) = (\alpha - 1)X + (\beta - 1)Y - N \log B(\alpha, \beta),$$

where the sufficient statistics (X, Y) are :

$$X = \sum_{i=1}^{N} \log x_i, \ Y = \sum_{i=1}^{N} \log(1 - x_i).$$

In a Bayesian setting to avoid the dirac-like shape of the beta distribution, a prior law can be added :

$$P(\alpha, \beta) \propto B(\alpha, \beta)^K e^{-a\alpha} e^{-b\beta} \text{ for } K < N, a, b \in \mathbb{R}^+. \quad (11)$$

The objective function to achieve the MAP estimator becomes:

$$f^*(\alpha, \beta) = \alpha(X - a) + \beta(Y - b) - (N - K) \log B(\alpha, \beta) + C^t \quad (12)$$

for which inequality constraints are :

$$g_1 = \epsilon - \alpha \leq 0, \, g_2 = \epsilon - \beta \leq 0$$

where $\epsilon > 0$ is close to zero to form a closed space. Note that these constraints are affine functions to satisfy the regularity conditions in the KKT. We thus search the solution of :

$$F_1(\boldsymbol{z}) = (X - a) - (N - K)\left(\psi(\tilde{\alpha}) - \psi(\tilde{\alpha} + \tilde{\beta})\right) + \tilde{\mu}_1 = 0,$$
$$F_2(\boldsymbol{z}) = (Y - b) - (N - K)\left(\psi(\tilde{\beta}) - \psi(\tilde{\alpha} + \tilde{\beta})\right) + \tilde{\mu}_2 = 0,$$
$$\epsilon - \tilde{\alpha} \leq 0, \ \tilde{\mu}_1 \geq 0, \ F_3(\boldsymbol{z}) = \tilde{\mu}_1(\epsilon - \tilde{\alpha}) = 0, \quad (13)$$
$$\epsilon - \tilde{\beta} \leq 0, \ \tilde{\mu}_2 \geq 0, \ F_4(\boldsymbol{z}) = \tilde{\mu}_2(\epsilon - \tilde{\beta}) = 0, \quad (14)$$

where $\psi(\cdot)$ represents the di-gamma function.

We apply the Newton Raphson method on $F(\boldsymbol{z}) = [F_1, F_2, F_3, F_4]^t$ by updating iteratively :

$$\boldsymbol{z}^{(n+1)} - \boldsymbol{z}^{(n)} = -\mathbf{J}(\boldsymbol{z}^{(n)})^{-1} F(\boldsymbol{z}^{(n)}), \quad (15)$$

and checking the inequality constraints in (13) and (14).

We next show that $\mathbf{J}(\boldsymbol{z})$ is always invertible given the inequality constraints. The Jacobian $\mathbf{J}(\boldsymbol{z}) = [\frac{\partial F_i}{\partial \boldsymbol{z}_j}]$ writes:

$$\mathbf{J}(\boldsymbol{z}) = \left[\begin{array}{cc|c} (N-K)\begin{bmatrix} \psi^{(1)}(\alpha+\beta) - \psi^{(1)}(\alpha) & \psi^{(1)}(\alpha+\beta) \\ \psi^{(1)}(\alpha+\beta) & \psi^{(1)}(\alpha+\beta) - \psi^{(1)}(\beta) \end{bmatrix} & & \mathbf{I}_2 \\ \hline & \begin{matrix} -\mu_1 & 0 \\ 0 & -\mu_2 \end{matrix} & \begin{matrix} -\alpha & 0 \\ 0 & -\beta \end{matrix} \end{array}\right]$$

where $\psi^{(1)}(\cdot)$ is the tri-gamma function (second derivative of the log-gamma). Its determinant is :

$$\det \mathbf{J}(\boldsymbol{z}) = \det \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \det AD - BC = (N-K)^2 \alpha\beta$$
$$\left(\psi^{(1)}(\alpha)\psi^{(1)}(\beta) - (\psi^{(1)}(\alpha) + \psi^{(1)}(\beta))\psi^{(1)}(\alpha+\beta)\right)$$
$$+ (N-K)\mu_1\beta(\psi^{(1)}(\beta) - \psi^{(1)}(\alpha+\beta))$$
$$+ (N-K)\mu_2\alpha(\psi^{(1)}(\alpha) - \psi^{(1)}(\alpha+\beta))$$

The second equality is due to the fact that C and D commute (i.e. $CD = DC$). From the relation

$$\psi^{(1)}(\alpha)\psi^{(1)}(\beta) > (\psi^{(1)}(\alpha) + \psi^{(1)}(\beta))\psi^{(1)}(\alpha+\beta)$$

it can be verified that $\det \mathbf{J}(\boldsymbol{z}) > 0$. Thus the Jacobian matrix is always invertible.

APPENDIX B
KULBACK-LEIBER DIVERGENCE FOR GENERALIZED NORMAL DISTRIBUTIONS

Recall that the probability density of the generalized normal distribution writes:

$$P(x;\alpha,\beta,\mu) = \frac{\beta}{2\alpha\Gamma(1/\beta)}\, e^{-(|x-\mu|/\alpha)^{\beta}} \text{ for } \alpha,\beta>0$$

Thus, the Kullback-Leibler divergence is:

$$\begin{aligned} D_{\mathrm{KL}}(P_1\|P_2) &= \int_{\mathbb{R}} \frac{\beta_p}{2\alpha_p\Gamma(1/\beta_p)}\, e^{-(|x-\mu_p|/\alpha_p)^{\beta_p}} \\ &\times \log\left(\frac{\frac{\beta_p}{2\alpha_p\Gamma(1/\beta_p)}\, e^{-(|x-\mu_p|/\alpha_p)^{\beta_p}}}{\frac{\beta_q}{2\alpha_q\Gamma(1/\beta_q)}\, e^{-(|x-\mu_q|/\alpha_q)^{\beta_q}}}\right) dx \\ &= \log\left(\frac{\beta_p\alpha_q\Gamma(1/\beta_q)}{\beta_q\alpha_p\Gamma(1/\beta_p)}\right) + \int_{\mathbb{R}} \frac{\beta_p}{2\Gamma(1/\beta_p)}\, e^{-(|x-\mu_p|/\alpha_p)^{\beta_p}} \\ &\times \left(-(|x-\mu_p|/\alpha_p)^{\beta_p} + (|x-\mu_q|/\alpha_q)^{\beta_q}\right)\frac{dx}{\alpha_p}. \end{aligned}$$

Let $t=\frac{x-\mu_p}{\alpha_p} \Leftrightarrow dx=\alpha_p dt$. Since $\alpha_p>0$, we have:

$$\begin{aligned} D_{\mathrm{KL}}(P_1\|P_2) &= \log\left(\frac{\beta_p\alpha_q\Gamma(1/\beta_q)}{\beta_q\alpha_p\Gamma(1/\beta_p)}\right) \\ &- \int_{\mathbb{R}} \frac{\beta_p}{2\Gamma(1/\beta_p)}\, e^{-|t|^{\beta_p}}\, |t|^{\beta_p} dt \\ &+ \underbrace{\int_{\mathbb{R}} \frac{\beta_p}{2\Gamma(1/\beta_p)}\, e^{-|t|^{\beta_p}}\, (|t\alpha_p+\mu_p-\mu_q|/\alpha_q)^{\beta_q} dt}_{(*)} \end{aligned}$$

With the definition of the gamma function in (3) and $\Gamma(z+1)=z\,\Gamma(z)$, the second term of the above equation can be further simplified :

$$\int_0^{+\infty} \frac{\beta_p}{\Gamma(1/\beta_p)}\, e^{-t^{\beta_p}}\, t^{\beta_p} dt = \frac{\Gamma(1/\beta_p+1)}{\Gamma(1/\beta_p)} = \frac{1}{\beta_p}$$

In the following, we treat the term in $(*)$. First define $\tilde{u}=\frac{|\mu_p-\mu_q|}{\alpha_p}$ and it can be written as :

$$\begin{aligned} (*) &= \left(\frac{\alpha_p}{\alpha_q}\right)^{\beta_q} \frac{\beta_p}{2\Gamma(1/\beta_p)} \int_{\mathbb{R}} e^{-|t|^{\beta_p}} |t+\tilde{u}|^{\beta_q} dt \\ &= k_1 \int_{\mathbb{R}} e^{-|t|^{\beta_p}} |t+\tilde{u}|^{\beta_q} dt \\ &\overset{t'=t/\tilde{u}}{=} k_1 \int_{\mathbb{R}} e^{-(\tilde{u}t')^{\beta_p}}\, |\tilde{u}(t'+1)|^{\beta_q} \tilde{u}\, dt' \\ &= k_1\, \tilde{u}^{\beta_q+1} \int_{\mathbb{R}} e^{-\tilde{u}^{\beta_p}|t|^{\beta_p}}\, |t+1|^{\beta_q}\, dt \\ &= k_2 \int_{\mathbb{R}} e^{-\xi|t|^{\beta_p}} |t+1|^{\beta_q}\, dt \end{aligned}$$

with $k_1=\left(\frac{\alpha_p}{\alpha_q}\right)^{\beta_q}\frac{\beta_p}{2\Gamma(1/\beta_p)}$, $\xi=\tilde{u}^{\beta_p}$ and $k_2=k_1\,\tilde{u}^{\beta_q+1}>0$.

Since both $|t|$ and $|t+1|$ exist in the expression, we further decompose the integral into :

$$\begin{aligned} (*) =& k_2 \int_0^1 e^{-\xi t^{\beta_p}} \left[(t+1)^{\beta_q} + (1-t)^{\beta_q}\right] dt \\ &+ k_2 \int_1^{+\infty} e^{-\xi t^{\beta_p}} \left[(t+1)^{\beta_q} + (t-1)^{\beta_q}\right] dt \end{aligned}$$

Notice that when β_q is even, the two functions inside the integrals are identical since $(t-1)^{\beta_q}=(1-t)^{\beta_q}$. For instance, if $\beta_q=2$:

$$\begin{aligned} & k_2 \int_0^1 e^{-\xi t^{\beta_p}} 2\left[t^2+1\right] dt + k_2 \int_1^{+\infty} e^{-\xi t^{\beta_p}} 2\left[t^2+1\right] dt \\ &= 2k_2 \int_0^{\infty} e^{-\xi t^{\beta_p}} \left[t^2+1\right] dt = \frac{2k_2}{\beta_p}\left(\frac{\Gamma(3/\beta_p)}{\xi^{3/\beta_p}} + \frac{\Gamma(1/\beta_p)}{\xi^{1/\beta_p}}\right), \end{aligned}$$

by exploiting the relation :

$$\int_0^{+\infty} e^{-\xi t^{\beta_p}} t^{\beta_q}\, dt = \frac{1}{\beta_p\, \xi^{\frac{\beta_q+1}{\beta_p}}} \Gamma\left(\frac{\beta_q+1}{\beta_p}\right). \tag{16}$$

This result applies also for $\beta_q=0$ though it is supposed to be strictly positive by definition. Let's then investigate the case of $\beta_q=2n+1$ with $n\in\mathbb{N}$. First, for $\beta_q=1$,

$$\begin{aligned} (*) =& k_2 \int_0^1 e^{-\xi t^{\beta_p}} (2)\, dt + k_2 \int_1^{+\infty} e^{-\xi t^{\beta_p}} (2t)\, dt \\ =& 2k_2 \left[\int_0^1 e^{-\xi t^{\beta_p}}\, dt + \int_1^{+\infty} e^{-\xi t^{\beta_p}} t\, dt\right] \\ \overset{y=\xi t^{\beta_p}}{=}& \frac{2k_2}{\beta_p}\left[\left(\frac{1}{\xi}\right)^{\frac{1}{\beta_p}} \int_0^{\xi} e^{-y} y^{\frac{1}{\beta_p}-1} dy + \left(\frac{1}{\xi}\right)^{\frac{2}{\beta_p}} \int_{\xi}^{\infty} e^{-y} y^{\frac{2}{\beta_p}-1} dy\right] \\ =& \frac{2k_2}{\beta_p}\left[\left(\frac{1}{\xi}\right)^{\frac{1}{\beta_p}} \gamma(\frac{1}{\beta_p},\xi) + \left(\frac{1}{\xi}\right)^{\frac{2}{\beta_p}} \Gamma(\frac{2}{\beta_p},\xi)\right] \end{aligned}$$

in which

$$\gamma(s,x) = \int_0^x t^{s-1}\, e^{-t}\, \mathrm{d}t, \quad \Gamma(s,x) = \int_x^{\infty} t^{s-1}\, e^{-t}\, \mathrm{d}t$$

are the lower and upper incomplete gamma functions respectively. Similarly, for $\beta_q=3$, we can develop $(1+t)^3+(1-t)^3$ and $(1+t)^3+(t-1)^3$ respectively to get :

$$(*) = k_2\left(\int_0^1 e^{-\xi t^{\beta_p}}(2+6t^2)dt + \int_1^{+\infty} e^{-\xi t^{\beta_p}}(6t+2t^3)dt\right),$$

so that the same technique using $y=\xi t^{\beta_p}$ can be applied to coerce $(*)$ into a sum of lower and upper incomplete gamma functions. In a similar manner as (16), we obtain the following relations for the incomplete gamma integrals:

$$\int_0^1 e^{-\xi t^{\beta_p}} t^{\beta_q} dt = \frac{1}{\beta_p \xi^{\frac{\beta_q+1}{\beta_p}}} \gamma\left(\frac{\beta_q+1}{\beta_p},\xi\right), \tag{17}$$

$$\int_1^{\infty} e^{-\xi t^{\beta_p}} t^{\beta_q}\, dt = \frac{1}{\beta_p \xi^{\frac{\beta_q+1}{\beta_p}}} \Gamma\left(\frac{\beta_q+1}{\beta_p},\xi\right). \tag{18}$$

To generalize, $(*)$ is a sum of weighted gamma functions using (16) for even β_q whereas for odd β_q it is a sum of weighted upper and lower incomplete gamma functions using (17) and (18). In both cases, binomial coefficients of β_q-degree polynomials are needed to weight the sum.

Next we show that $(*)$ is monotonously increasing with respect to β_q or :

$$\frac{\partial(*)}{\partial\beta_q} = k_2 \int_{\mathbb{R}} e^{-\xi|t|^{\beta_p}} |1+t|^{\beta_q} \log|1+t| dt > 0$$

in which k_2 and $e^{-\xi|t|^{\beta_p}}|1+t|^{\beta_q}$ are positive for $t \in \mathbb{R}$. $\log|1+t|$ is negative for t $\in [-2,0]$, and positive otherwise. A sufficient condition is :

$$\int_{-2}^{2} e^{-\xi t^{\beta_p}}|1+t|^{\beta_q}\log|1+t|\ dt > 0 \tag{19}$$

By splitting the integral into 2 parts and letting $y=-t$ in the first part, the integral of the above inequality becomes

$$\int_0^2 e^{-\xi y^{\beta_p}}|1-y|^{\beta_q}\log|1-y|dy+\int_0^2 e^{-\xi y^{\beta_p}}(1+y)^{\beta_q}\log(1+y)dy$$
$$=\int_0^2 e^{-\xi y^{\beta_p}}\underbrace{\left(|1-y|^{\beta_q}\log|1-y|+(1+y)^{\beta_q}\log(1+y)\right)}_{G_{\beta_q}(y)}dy$$

Note that the function $G_{\beta_q}(y)$ is a continuous function of y and differentiable by piece. It can be shown that $G_{\beta_q}(y) \geq 0$ for all $y \in [0,2], \beta_q > 0$, which validates the condition in (19). It is then evident that $(*)$ is a monotonously increasing function of β_q.

REFERENCES

[1] J. Pan, WJ. Tompkins, "Biomedical Engineering, A real-time QRS detection algorithm," *Biomedical Engineering, IEEE Transactions on*, vol. 32, pp. 230-236, Mar. 1985.

[2] A. Hernández *et al.*, "Multisensor fusion for atrial and ventricular activity detection in coronary care monitoring," *IEEE Trans Biomed Eng.*, vol. 46, pp. 1186-90, Oct. 1999.

[3] Robust Detection of Heart Beats in Multimodal Data: the PhysioNet/ Computing in Cardiology Challenge 2014 - https://www.physionet.org/challenge/2014/

[4] B.U. Kohler *et al.*, "The principles of software QRS detection," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21(1), pp.42-57, Aug. 2002.

[5] S. Jain *et al.*, "QRS detection using adaptive filters: A comparative study," *ISA Transactions*, vol. 66, pp. 362-375, Jan. 2017

[6] S. Osowski, T.H. Linh, "ECG beat recognition using fuzzy hybrid neural network," *Biomedical Engineering, IEEE Transactions on*, vol. 48, pp. 1265-1271, Nov. 2011.

[7] Z. Zidelmala *et al.*, "QRS detection using S-Transform and Shannon energy," *Computer Methods and Programs in Biomedicine*, vol. 116, pp. 1-9, Aug. 2014.

[8] C. Li *et al.*, "Detection of ECG characteristic points using wavelet transforms," *Biomedical Engineering, IEEE Transactions on*, vol. 42, pp. 21-28, Jan.1995.

[9] D. Benitez *et al.*, "The use of Hilbert transform in ECG signal analysis," *Computers in Biology and medecine*, vol. 31, pp. 399-406 Sep. 2001.

[10] Z. Zidelmal *et al.*, "QRS detection based on wavelet coefficients," *Computer Methods and Programs in Biomedicine*, vol. 107, pp. 490-496, Sep. 2012.

[11] R. Rani *et al.*, "Automated Detection of QRS Complex in ECG Signal using Wavelet Transform," *International Journal of Computer Science and Network Security*, vol. 15, no.1, Jan. 2015.

[12] G.M. Friesen, "A comparison of the noise sensitivity of nine qrs detection algorithms," *Biomedical Engineering, IEEE Transactions on*, vol. 37, pp. 85-98, Jan. 1990.

[13] H. Khamis *et al.*, "QRS Detection Algorithm for Telehealth Electrocardiogram Recordings," *IEEE Trans Biomed Eng* vol.63 no.7), pp. 1377-1388, Jul. 2016.

[14] M. Elgendi *et al.* (2017 Mar), "Efficient ECG Compression and QRS Detection for E-Health Applications," *Scientific Reports*. [Online]. 7(1), Article no. 459. Available https://www.ncbi.nlm.nih.gov/pubmed/28352071.

[15] J. Kim, H. Shin (Mar. 2016), "Simple and Robust Realtime QRS Detection Algorithm Based on Spatiotemporal Characteristic of the QRS Complex", " *PLoS One*. [Online]. 11(3), Article no. e0150144. Available https://doi.org/10.1371/journal.pone.0150144

[16] M.M. Baig *et al.*, "A comprehensive survey of wearable and wireless ECG monitoring systems for older adults," *Med Biol Eng Comput*, vol. 51, pp. 485-495, May. 2013.

[17] C.J. Deepu *et al.*, (2010, Jan.). An ECG-on-Chip for Wearable Cardiac Monitoring Devices. Presented at Fifth IEEE International Symposium on Electronic Design, Test and Applications. [Online]. http://ieeexplore.ieee.org/document/5438685/

[18] D.M. Goldenholz *et al.*, "Long-term monitoring of cardiorespiratory patterns in drug-resistant epilepsy," *Epilepsia*, vol. 58, pp. 77-84, Jan. 2017.

[19] P. Portet *et al.*, "Evaluation of real-time QRS detection algorithms in variable contexts," *Medical and Biological Engineering and Computing*, vol. 43, 379-385, May. 2005.

[20] J. Dumont *et al.*, "Improving ECG Beats Delineation With an Evolutionary Optimization Process," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 607-615, Mar. 2010.

[21] M. Altuve *et al.*, "Analysis of the QRS complex for apnea-bradycardia characterization in preterm infants," in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medecine and Biology Society (EMBS)*, Minneapolis, E.U. 2009, pp. 946-949.

[22] D. Hernando *et al.*, "QRS Detection Optimization in Stress Test Recordings Using Evolutionary Algorithms," *Computing in Cardiology Conference (CinC)*, pp. 737-740, Sep. 2014.

[23] G.B. Moody *et al.*, "A noise stress test for arrhythmia detectors," *Computers in Cardiology*, vol. 11, pp. 381-384, Jan. 1984.

[24] J.P. Martinez *et al.*, "A wavelet-based ECG delineator: evaluation on standard databases," *IEEE Transactions on Biomedical Engineering*, vol.51, pp. 570-581, Apr. 2004.

[25] P. S. Hamilton and W. J. Tompkins, "Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database," *IEEE Transactions on Biomedical Engineering*, vol. BME-33, pp. 1157-1165, Dec. 1986. doi: 10.1109/TBME.1986.325695

[26] M. Gil *et al.*, "Rényi Divergence Measures for Commonly Used Univariate Continuous Distributions," *Information Sciences*, vol. 249, pp. 124-131, Nov. 2013

[27] J.R. Mathiassen *et al.*. (2002, May) Texture Similarity Measure Using Kullback-Leibler Divergence between Gamma Distributions. Presented at Computer Vision - ECCV. [Online]. https://link.springer.com/chapter/10.1007%2F3-540-47977-5_9

[28] M. Elgendi *et al.* (Sep. 2013), "Fast QRS Detection with an Optimized Knowledge-Based Method: Evaluation on 11 Standard ECG Databases,", " *PLoS One*. [Online]. 8(9), Article no. e73557. Available https://doi.org/10.1371/journal.pone.0073557

[29] A.J. Alberg *et al.*, "The Use of "Overall Accuracy" to Evaluate the Validity of Screening or Diagnostic Tests," *J Gen Intern Med*, vol. 19, pp. 460-465, May 2004.

[30] P.R. Rijnbeek *et al.*, "Normal values of the electrocardiogram for ages 16–90 years," *Journal of Electrocardiology*, vol. 47, pp. 914-921, Nov 2014.

[31] B.E. Hanson. (2004, May). University of Wisconsin. Madison, US. "Bandwidth selection for nonparametric distribution estimation," [Online]. Available: http://www.ssc.wisc.edu/ bhansen/papers/smooth.pdf

[32] N.D. Minh, M. Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance," *IEEE Transactions on signal processing*, vol. 11, pp. 146-158, Feb. 2002.