

Evaluating Mathematical Models for Morphological Classification of the QRS Complex

João Paulo do Vale Madeiro¹, Délcio Barreto², João Alexandre Lôbo Marques, João Loures Salinet Júnior⁴

¹Department of Computing Science, Federal University of Ceará, Fortaleza, Brazil

²Department of Electrical Engineering, Federal University of Ceará, Fortaleza, Brazil

³Shenzhen Institutes of Advanced Technologies, Chinese Academy of Sciences, Shenzhen, China

⁴Center of Engineering, Modeling, and Applied Social Sciences, Federal University of ABC, São Paulo, Brazil

Abstract

The present work aims to present a comparative study of the performance of different kernels for mathematical modeling and morphological classification of the QRS complex of the ECG signal. Initially, we use a simulator to generate synthetic signals from dynamic models containing variations within a set of physiological parameters. From the generation of twenty different types of QRS morphology, computing tests for mathematical modeling of the beat waveform (Q, R and S waves) were performed. For that, the following mathematical functions are employed: Gaussian function, Mexican Hat function and Rayleigh probability density function. Subsequently, 10 real signal records from the MIT-BIH (Massachusetts Institute Technology - Beth Israel Hospital) Arrhythmia database have their QRS complex morphologies also modeled by the proposed mathematical functions. The preliminary results demonstrate the proposed mathematical functions with adjustable parameters can be applied together for modeling and automatic classification of some QRS morphologies commonly present in real signals, with efficiency and precision. The computing of normalized RMS error allows the identification of the model which is more appropriate to a given morphology, which can change over the same patient record.

1. Introduction

Automatic ECG feature extraction comprises a set of steps that begin with the detection and segmentation of their characteristic waves. Correct detection of QRS complexes and their delineation are fundamental conditions for the efficient detection and segmentation of the other characteristic ECG waves. In addition, they provide the basis for cardiac arrhythmia pattern recognition algorithms [1].

It is relevant to mention that the ECG signal tracing presents a plethora of possible formats, depending on the situations and physical-emotional conditions, the noise from the environment, each individual itself, the modulation of the amplitude of the signal, according to breathing and noise of electro-surgical devices.

In this sense, there have been many advances related to the automatic extraction of ECG signal parameters and currently there is a large number of pathophysiological diseases that can be diagnosed with the aid of engineering tools and techniques, allowing the increased use of the ECG signal in daily life.

Determining the polarity of QRS deflections has extremely relevant implications for several diagnostic criteria and is fundamentally important for the electrocardiogram electrical axis calculation. As far as polarity is concerned, the QRS complex may be positive, negative or isodiphasic. The QRS complex is positive when the height of the largest positive wave (R or R') is greater than the depth of the largest negative wave (Q or S). In turn, the QRS complex is negative when the height of the largest positive wave (R or R') is lower than the depth of the largest negative wave (Q or S). Finally, the QRS complex is isodiphasic when the height of the largest positive wave and the depth of the largest negative wave are similar [2].

The master thesis developed by [3], also available from the Physionet website, suggests twenty possible classes for QRS morphologies: qR, qRs, Rs, R, RS, rSR', rR', qrSr', RSr', rR's, rS, rSr', Qr, QS, QR, qrS, qS, rsR's', QRs, and Qrs. Thus, automatic recognition of different QRS morphologies through mathematical modelling provides to obtain a rich set of features related to QRS waveform and also discriminate certain types of arrhythmias and conduction disorders.

Based on this assumption, this work aims the modeling and morphological classification of the QRS complex of the ECG signal contributing to the dissemination of the

study of the electrocardiogram in the academic and scientific space.

2. Methodology

In the development of the research, the following steps were followed:

1. Computing implementation of artificial ECG signal generation algorithms with variable parameters, including mean and standard deviation of interval between beats, characteristic wave morphology (PQRST cycle) and signal to noise ratio, taking as initial reference the technique developed by MCSHARRY et al. in [4]. Simulated noise insertion algorithms were also elaborated on the synthesized ECG signals, corresponding to the baseline oscillations, mains interference (50-60 Hz carriers), muscle contraction and Gaussian white noise;

2. Literature review concerning computational methods for classical linear and nonlinear filters and application of the Wavelet transform to eliminate ECG noise;

3. Implementation of QRS complex detection and segmentation algorithms over artificial ECG signals. It was proposed to use the detection methodology based on the application of the first derivative function and the Wavelet and Hilbert transforms, in combination with the adaptive threshold technique [5].

The segmentation process was based on the calculation of an indicator related to the area covered by the QRS envelope. The envelope corresponds to the analytical signal obtained from the serial application of the Wavelet transform and the first derivative function on the ECG signal (real part) and the application of the Hilbert transform on the resulting signal (imaginary part) [5].

4. Modeling of the QRS complex on synthesized artificial ECG signals based on the applied mathematical models. The mathematical models used for evaluation and experimental modeling test of the QRS complex are: Gaussian function, Mexican Hat function and combination of Rayleigh Probability Density functions. To determine the performance of the tested kernels, the normalized RMS (Root Mean Square) error is computed.

5. Implementation of QRS complex detection, segmentation and modeling algorithms over ten 30-minute reference real-time signal records from the MIT-BIH (Beth Israel Hospital) Arrhythmia database.

2.1. Proposed Mathematical Models

Having in mind that real ECG waveforms are intrinsically asymmetric and even distorted by interference and physiological noise, it is intuitive that mathematical functions should acquire a certain degree of distortion for becoming able to fit physiological signal.

Thus, the first mathematical model corresponds to the

combination of two Gaussian functions. A single discrete-time Gaussian function may be defined as [6]

$$G_{\sigma_1}[k] = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{k^2}{2\sigma_1^2}\right). \quad (1)$$

The mathematical model resultant from combining two discrete-time Gaussian functions with individual width parameters σ_1 and σ_2 is, therefore, given as

$$G_{\sigma_1, \sigma_2}[k] = \begin{cases} G_{\sigma_1}[k], & -x_1 \leq k \leq 0 \\ G_{\sigma_2}[k], & 0 \leq k \leq x_2 \end{cases} \quad (2)$$

The second mathematical model corresponds the second derivative of the first mathematical model, that is, a modified Mexican Hat function, once the original one is symmetric. Therefore, given a function $G_{\sigma_1, \sigma_2}[k]$ derived from the combination of two Gaussian functions, the second mathematical model $G_{\sigma_1, \sigma_2}^d[k]$ may be obtained as a second order progressive divided difference

$$G_{\sigma_1, \sigma_2}^d[k] = \frac{G_{\sigma_1, \sigma_2}[k+1] - 2G_{\sigma_1, \sigma_2}[k] + G_{\sigma_1, \sigma_2}[k-1]}{T_s^2} \quad (3)$$

where T_s represents the sampling period.

Finally, we define a discrete-time Rayleigh probability density function $R_{\sigma_1}[k]$ within the interval $0 \leq k \leq h$

$$R_{\sigma_1}[k] = \frac{k}{\sigma_1^2} \exp\left(-\frac{k^2}{2\sigma_1^2}\right) \quad (4)$$

where σ_1 is the corresponding variable parameter.

Then, the next three mathematical models corresponds to three different combinations of two Rayleigh functions: Rayleigh cycle positive/negative, Rayleigh cycle negative/positive and Rayleigh cycle positive/positive.

Concerning the third mathematical model, we obtain a composition of two Rayleigh functions, denoted $R_{\sigma_1, \sigma_2}[k]$, $0 \leq k \leq 2h$. We depart from the definition of two discrete-time Rayleigh probability density functions $R_{\sigma_1}[k]$ and $R_{\sigma_2}[k]$ so that

$$R_{\sigma_1, \sigma_2}[k] = \begin{cases} R_{\sigma_1}[h-k], & 0 \leq k \leq h \\ -R_{\sigma_2}[k-h], & h \leq k \leq 2h \end{cases} \quad (5)$$

The other two mathematical models correspond to variations of the third mathematical model, being the fourth one (Rayleigh cycle negative/positive) given as

$$R_{\sigma_1, \sigma_2}[k] = \begin{cases} -R_{\sigma_1}[h-k], & 0 \leq k \leq h \\ R_{\sigma_2}[k-h], & h \leq k \leq 2h \end{cases} \quad (6)$$

and the fifth one (Rayleigh cycle positive/positive) given

as

$$R_{\sigma_1, \sigma_2}[k] = \begin{cases} R_{\sigma_1}[h - k], & 0 \leq k \leq h \\ R_{\sigma_2}[k - h], & h \leq k \leq 2h \end{cases} \quad (7)$$

After detecting each QRS complex within a given ECG record, it is possible to establish search windows centered in each QRS fiducial point (R-wave peak) aiming to investigate which model is more appropriate for a given predominant morphology. By an exhaustive search, we achieve to fit each model within a given morphology and compute a normalized RMS error between a given QRS segment and a given fitted model. The winner mathematical model is the one which allows to obtain the lowest normalized RMS error, which may be computed as

$$\varepsilon = \sqrt{\frac{\sum_{k=1}^L |W_X[k] - \tilde{W}_X[k]|^2}{\sum_{k=1}^L W_X[k]^2}} \quad (8)$$

where $W_X[k]$ refers to the signal window for which we desire to fit a model, and $\tilde{W}_X[k]$ refers to the evaluated kernel.

3. Results and Discussion

Based on parameter adjustments of the dynamic models proposed by [4], it was possible to generate 60 samples of each of the QRS morphologies predominantly found in the literature, varying amplitude and phase of the waves Q, R and S: qR, qRs, Rs, R, RS, rSR', rR', qrSR', RSr', rR's, rS, rSr', Qr, QS, QR, qrS, qS, rSR's', QRs and Qrs.

For each of the tested mathematical models, a morphology, among the twenty considered, was identified in which a given model has the lowest RMS error in the Ranking, as shown in Table 1. Figure 1 illustrate three different QRS morphologies generated by the implemented computing simulator: rS, rSr' and qrS..

Table 2 presents the detailed modeling results for each of the 20 QRS morphologies considered on the synthetic signals, with the corresponding winning kernel.

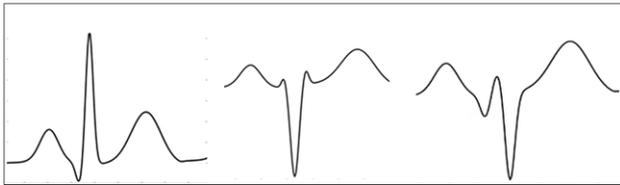


Figure 1. QRS morphologies rS, rSr' and qrS from synthetic ECG signals.

Table 1. Ranking of lowest RMS error.

Mathematical Model	QRS Morphology	RMS Error
Gaussian	R	4,67% ± 1,22 %
Mexican-hat	rsR's'	18,36% ± 0,94 %
Rayleigh	RS	10,23% ± 3,1 %

Table 2. Modeling results for each of the twenty common QRS morphologies on synthetic ECG signals

QRS Morphology	Winning kernel	RMS Error
qR	Gaussian	17,94% ± 0,72%
qRs	Mexican-hat	20,06% ± 1,43%
Rs	Rayleigh +/-	20,23% ± 4,16%
R	Gaussian	4,67% ± 1,22%
RS	Rayleigh +/-	10,23% ± 3,10%
rSR'	Rayleigh -/+	17,02% ± 3,91%
rR'	Rayleigh +/-	27,08% ± 2,76%
qrSR'	Rayleigh +/-	19,56% ± 1,61%
RSr'	Rayleigh +/-	15,01% ± 3,46%
rR'S	Rayleigh +/-	26,97% ± 2,27%
rS	Rayleigh -/+	18,37% ± 0,55%
rSr'	Mexican-hat	19,93% ± 2,88%
Qr	Rayleigh +/-	11,95% ± 1,79%
QS	Gaussian	6,09% ± 1,69%
QR	Rayleigh -/+	11,96% ± 3,78%
qrS	Rayleigh -/+	14,71% ± 2,47%
qS	Rayleigh +/-	23,29% ± 1,77%
rsR's'	Mexican-hat	18,36% ± 0,94%
QRs	Rayleigh -/+	18,80% ± 2,05%
Qrs	Rayleigh +/-	14,95% ± 1,08%

Table 3 presents the results of applying the QRS complex modeling algorithm over 10 real signal records from the MIT-BIH Arrhythmia Database, considering the mathematical models Gaussian function (Model 1), Mexican Hat function (Model 2), Rayleigh Positive + Negative Cycle (Model 3), Rayleigh Negative + Positive Cycle (Model 4) and Rayleigh Positive + Positive Cycle (Model 5). It is relevant to note that in all modeling application results detailed in Table 5, only the first channel of the ECG signals was used, and from the five mathematical models tested, the winning model presented a percentage of occurrence above 70% for all cases.

Table 3. Modeling algorithm application results for a set of real signals from the MIT-BIH Arrhythmia Database

Record	Winning Model	Hit Percentage	RMS Error
234	1	94,25%	7,56% ± 1,74 %
220	3	99,21%	22,01% ± 3,58%
219	3	79,55%	9,71% ± 3,34 %
217	1	79,78%	12,74% ± 8,15%

215	3	94,31%	20,87% ± 4,79%
212	1	81,90%	12,41% ± 3,86%
209	3	95,41%	18,07% ± 4,74%
202	1	91,88%	10,81% ± 17,16%
201	1	75,57%	19,55% ± 29,05%
123	3	80,59%	19,45% ± 3,85%

As an illustrative example, Figure 2 presents the predominant morphology in record 209 from MIT-BIH Arrhythmia Database, modeled by Rayleigh Positive + Negative Cycle (Model 3).

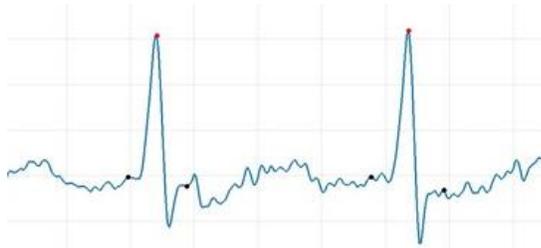


Figure 2. QRS predominant morphology from record 209 (MIT-BIH Arrhythmia Database) modeled by Rayleigh Positive + Negative Cycle (Model 3).

4. Conclusion

The wide diversity of QRS complex morphologies (around 20, according to the literature), characterizes the task of classification and recognition as extremely challenging. From the results achieved, it was observed that the proposed mathematical functions with adjustable parameters can be applied together for modeling and automatic classification of beat morphologies commonly present in real signals, with efficiency and precision. The normalized RMS error computing allows the differentiation and identification of the most appropriate model for a given morphology, which can alternate over the same patient record. The next step of the research is to use the parameters provided by the mathematical models as well as the normalized RMS error values computed for each model as input data for a classifier to be constructed from machine learning techniques for cardiac disease pattern recognition.

Acknowledgments

The work is supported by Brazilian Agencies CNPq, Grant No. 426002/2016-4, and FUNCAP, Grant No. BP3-0139-00284.01.00/18.

References

- [1] J. P. do Vale Madeiro, P. C. Cortez J. L. Salinet Jr., R. C. Pedrosa, J. M. da Silva Monteiro Filho, and A. R. A. Brayner, "Classical and modern features for interpretation of ECG signal," in *Developments and Applications for ECG Signal Processing*. Elsevier, 2019, pp. 1-28.
- [2] J. P. do Vale Madeiro, J. M. da Silva Monteiro Filho, and P. R. F. Rodrigues, "Delineation of QRS complex: challenges for the development of widely applicable algorithms," in *Developments and Applications for ECG Signal Processing*. Elsevier, 2019, pp. 119-139.
- [3] Lugovaya, T. S. "Biometric human identification based on electrocardiogram," [Master's thesis] Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russian Federation; June 2005.
- [4] Mcsharry, P. E., Clacor, G.D., Tarassenko, L., Smith, L. A. "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289 – 294, 2003.
- [5] Madeiro, J. P. do V.; Cortez, P. C.; Marques, J.A.L.; Seisdedos, C.R.V.; Sobrinho, C.R.M.R. "An innovative approach of QRS segmentation based on first-derivative, Hilbert and wavelet transforms," *Med. Eng. Phys.*, vol. 34 pp. 1236– 1246. 2012.
- [6] J. P. do Vale Madeiro, E. M. B. E. dos Santos, P. C. Cortez, J. H. da Silva Felix, and F. S. Schlindwein, "Evaluating gaussian and Rayleigh-based mathematical models for T and P-waves in ECG," *IEEE Latin America Transactions*, vol. 15, no. 5, pp. 843-853, 2017.

Address for correspondence:

Dr. João Paulo do Vale Madeiro.
 Department of Computing Science, Federal University of Ceará,
 Campus do Pici, 60440-900, Fortaleza, Ceará, Brazil.
 E-mail: jpaulo.vale@dc.ufc.br