

The Physionet QT Database: Study on the Reliability of P-wave Manual Annotations Under Noisy Recordings

Francisco González¹, Raúl Alcaraz², José J Rieta¹

¹ BioMIT.org, Electronic Engineering Department, Universitat Politècnica de Valencia, Spain

² Research Group in Electronic, Biomedical and Telecomm. Eng., Univ. of Castilla-La Mancha, Spain

Abstract

Thanks to its manual annotations, the PhysioNet QT database (QTDB) has been widely used as the reference of ECG delineators. However, a significant percentage of its annotations have been reported as inaccurate. Thus, any precise ECG delineator will never be able to meet, without error, all its annotations. The present work analyzes these inaccuracies and also how noise altered the final timing of annotations. As this effect is higher for low amplitude waveforms, P-waves were studied through a robust P-wave delineator. Its delineation results were compared with manual annotations under two scenarios. Firstly, a direct comparison without ECG denoising was performed. Secondly, the P-waves were delineated after efficient Wavelet-based denoising. Results showed that automatic annotations were closer to manual annotations for noisy ECGs and farther in the case of denoised ECGs, thus proving that noise altered the timing of manual annotations. An unreal improvement in delineation performance for noisy ECGs was obtained for P-wave onset, peak and offset in 45.83%, 57.29% and 56.25% of the recordings, respectively. Thus, to improve delineators reliability, either the need to review the QTDB annotations or its replacement by a better annotated database are suggested.

1. Introduction

Any new ECG signal processing algorithm designed to be used in a clinical setting requires the evaluation of its performance [1]. To do so, as well as for its proper development, the availability of databases, whose size is large enough to cover the wide diversity of waveform patterns that ECG recordings may present, is required [1]. Moreover, for the validation of some specific methods such as detectors or delineators of particular waves, annotations defining the time instants when these events occur are needed. In order to consider these annotations trustworthy, they might be determined manually and carefully by

expert physicians. However, there exist many factors that can influence the result of these manual annotations and, therefore, compromise their reliability.

There exist a considerable number of standard databases available to researchers with their proper distinctive features that depend on the initial aim for which they were created. Two of these databases are the MIT-BIH arrhythmia database [2] and the AHA Database, which were developed to evaluate arrhythmia detectors. On the other hand, the European ST-T Database was born in response to the growing interest in the analysis of ST-T segment as indicative of myocardial ischemia [3]. However, none of the aforementioned databases contains manual annotations of the locations of boundaries and peaks of all the waveforms that can be found in a normal ECG recording, something necessary for the validation of delineation methods of these waves. Therefore, standard databases providing manual annotations of ECG waves boundaries and peaks are needed. Some examples of this are the CSE multilead measurement database (CSEDB) [4] and the QT-database (QTDB) [5]. However, due to its free access and mirrored availability, this latter database has been the most widely used as a reference for the validation of P-wave delineation algorithms during last years.

Furthermore, the additional characteristics that have caused the QTDB to reach that position are, mainly, the wide diversity of waveform morphologies that contains, as recordings were specifically selected to reflect the real world variability and the considerable amount of waveforms annotations made manually by expert physicians, much more than in the CSEDB. In addition, another significant difference between these two databases is that the QTDB is public, thus allowing any researcher to use it without the need of financial support. However, manual annotations may not always be perfectly accurate. In this study the quality of manual waveform timings in the QTDB has been evaluated, specifically for the case of the P-waves, and the origin of the possible errors in the precise location of these points has been analyzed.

2. The QT database

The QTDB consists of 105 fifteen-minutes excerpts of two channels ECG recordings. These signals were collected from other existing databases as the MIT-BIH Arrhythmia Database and the European ST-T Database, as well as other databases collected at Boston's Beth Israel Hospital. It contains manual annotations of the beginning, peak and end of P-waves; beginning and end of QRS-complexes; peak and end of T-waves and, if present, the peak and end of U-waves. These annotations are present in, at least, 30 beats per record. The annotated beats were selected among the signal to represent its more characteristic or dominant morphological pattern [5].

The QTDB contains two sets of manual annotations made by each expert physician. However, one of them include annotations for just 11 records and, consequently, has rarely been used for validation purposes. Hence, in this study only the larger set of annotation will be discussed.

The process of manually annotating a signal is a complex, time consuming task. In addition, variables as the experience of the annotator, the degree of concentration and tiredness during the accomplishment of the task or the annotation tool accuracy and reliability, among others, can influence in the outcome of the procedure. Moreover, the presence of noise in the signal may further complicate this task, especially for low amplitude waves such as the P-wave. Due to this evidence, and the increasing attention that has gained the delineation of P-waves within the ECG because of the demonstrated relationship between different morphological characteristics of this waveform and clinical conditions, such as the recurrence of atrial fibrillation [6], the P-wave in particular has been selected in this study to evaluate the accuracy or the QTDB manual annotations.

3. Reliability analysis of the QTDB

3.1. The full scope concern

During the manual annotation process of the QTDB both leads were displayed simultaneously and the location of the annotations was established common to both channels [5]. This adopted procedure presents a serious drawback when it comes to knowing which wave was taken into account for each particular annotation, which is basic for the validation of single-lead delineation algorithms. In fact, due to the projection of the cardiac vector under each lead, it is well known that the ECG waves timing will be different as a function of the considered lead [7].

Other recently available database, the Chinese Cardiovascular Diseases Database (CCDD) [8], has followed the same full-lead scope annotation procedure, in which the annotation of boundaries and peaks positions are common

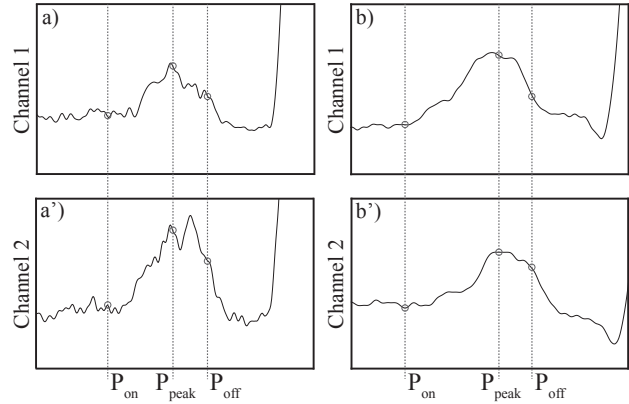


Figure 1. Examples of arguable annotated P-waves from the QTDB. a) and a') show the P-wave in both channels corresponding to the 17th annotated beat from recording *sele303*. b) and b') show both channels of the 17th annotated beat in recording *sel33*. The full scope fiducial points locations are indicated with a dotted line. Remark that, specially the offset annotations, are more than debatable.

for all leads. This specific case is even more challenging to deal with, as all 12 leads are provided. To overcome this difficulty, researchers has adopted two different strategies [9]. On the one hand, take as reference for each recording the lead in which their automatic delineation result is closer to manual annotations, thus assuming that only one lead was effectively taken into account to perform manual annotations. On the other hand, take as effective reference the closest automatically delineated point to each manual annotation, regardless of the lead where this takes place. Both strategies are oriented to get an unfair advantage of the database under test because of its full scope annotations.

3.2. Variable annotation criteria

Because of the criteria followed by physicians to locate the fiducial points is unclear, as there exists a lack of consensus in the scientific community about the precise location of the boundaries of a P-wave, generally it is not possible to categorically label an annotation as erroneous or imprecise. However, in some specific cases it is hardly debatable that the annotations are severely defective.

This occur when the annotated point is significantly far from the intuitive area in which an annotation could be considered as well located, as shown in Fig. 1, where common annotations to both channels for two P-waves can be observed. The waves were taken, among many other similar examples found, to exhibit that this is not a particular recording issue. Both leads are displayed because of the aforementioned lack of information about the channel taken into account during the annotation procedure.

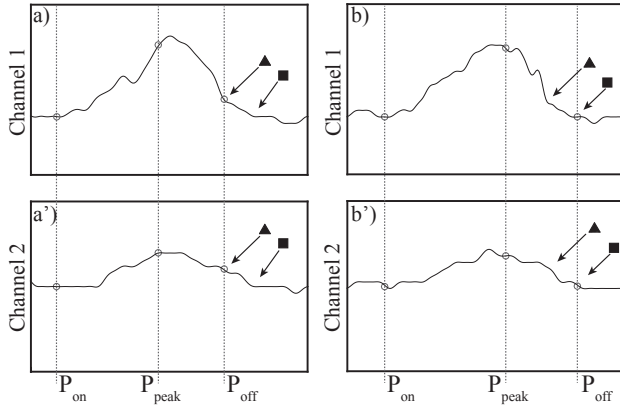


Figure 2. Examples of variable annotation criteria (triangle and square) in two P-waves from recording *sel803* of the QTDB. a) and a') show both channels of the third annotated P-wave, whereas b) and b') show the seventh P-wave. The full scope fiducial points locations are indicated with a dotted line. Remark the dissimilar criteria in a) and b).

Another circumstance in which an annotation can be unequivocally recognized as erroneous, despite the annotation criterion, takes place when the criterion itself varies between two waves of similar morphology. Figure 2 displays an example of two waves collected from the same recording. Observe how the offset annotation in both waves has been carried out according to different criteria. The two approximate positions according to each annotation criterion are indicated with two different symbols.

3.3. The relevance of noise

Finally, it must be taken into consideration that the signals, and specifically the P-waves because of their generally low amplitude, are deeply affected by noise. Thus, annotating under noisy conditions may render to undesired mistaken fiducial points. Although expert physicians are trained to overcome these adverse conditions and, in many times, they success in ignoring the presence of noise, in other occasions, they may fail despite making similar decisions. This can be easily revealed by a simple denoise process, as the applied in Figure 3, in which a wavelet-based denoising approach has been applied [10]. Two different P-waves are shown, a) and b), with their respective manual annotations that seems to be approximately well located. However, after noise reduction, a different reality is shown for a'), where a late onset has been mistakenly annotated. By contrast, the fiducial points in b') seem to preserve a reasonable location.

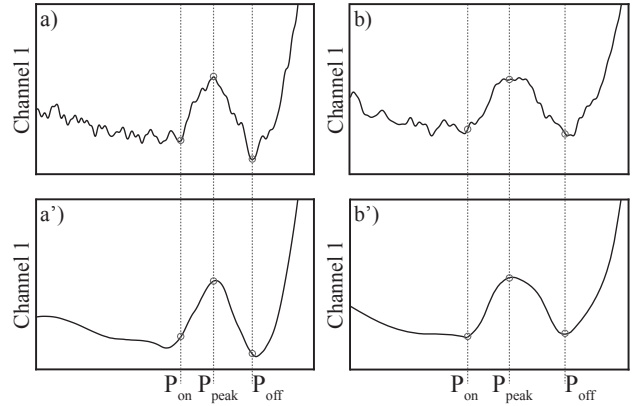


Figure 3. Example of noise effect in annotation reliability for two consecutive P-waves in the first channel of recording *sel230* from the QTDB. a) and b) P-waves of beats 30 and 31. a') and b') Resulting waves and annotations after ECG denoising.

4. Methodology and results

All recordings of the QTDB, to a greater or lesser extent, are affected by noise and consequently also their annotations. To approximately quantify this fact and other issues in the QTDB, an automatic delineator has been applied to P-waves before and after Wavelet-based denoising [10]. Next, the average absolute difference between automatic and manual annotations was computed. Both channels were delineated and the error in each case was computed as the time difference from automatic to the nearest manual annotation.

An adaptive P-wave delineation method, based on fitting P-waves by Gaussian functions, has been applied due to its robustness under noisy recordings [11]. A total of 96 recordings from the QTDB were analyzed, as those with less than three consecutive annotated P-waves were discarded. Results showed that automatic annotations were closer to manual annotations for noisy ECGs and farther in the case of denoised ECGs. An unreal improvement in delineation performance for noisy ECGs was obtained for P-wave onset, peak and offset in 45.83%, 57.29% and 56.25% of the recordings, respectively. Therefore, the influence of noise and (maybe) other aspects leading to mistaken annotations in the QTDB are highly relevant and should be considered seriously.

5. Discussion

As shown in Figures 1 to 3, the existence of inaccurate annotations of P-waves within the QTDB is undeniable, as previous studies have also reported [12]. In fact, other studies have manually reannotated the QTDB before the application of ECG delineation algorithms [13]. Further-

more, defective annotations are not the only source of error for developing automatic delineation methods. Thus, for cases where different criteria are adopted indistinctly, as shown in Figure 2, even if the correctness of both locations could be considered, no automatic delineator could ever meet the two criteria. This situation would involve an error for the algorithm in any case when comparing with the reference of manual annotation. It is, therefore, necessary to define a clear annotation criterion.

Finally, the presence of noise affected considerably the annotations. Thus, any delineator applied to noisy or de-noised ECGs will render different fiducial point locations. In fact, delineation under noisy conditions provided better results with respect to manual annotations, but they will be, actually, far from being accurate, thus affecting adversely to outstanding P-wave delineators.

It is difficult to know to what extent these inaccurate annotations have influenced previous results in the literature [9], or even if any of them have could taken advantage of these misplacements. However, what seems to be fairly evident is that this validation method is deficient and its improvement may be presented as the first obstacle to overcome for the development of reliable and clinically useful tools for the delineation of ECG waveforms.

6. Conclusions

In this study, several defects related to the annotations of the QTDB, the generally recognized reference for the validation of ECG delineation algorithms, have been highlighted. There have been identified, among others, the lack of information about the specific lead annotated in each case, the diversity of criteria during the annotation process or the effect of noise in the result of manual annotation. This latter source of error has been demonstrated to have a decisive effect on the outcome of automatic P-wave delineation methods. Thus, to improve the development of reliable and precise delineation methods, as well as to be able to trust on their validation process, one out of two conditions have to be satisfied: either the review of manual annotations on the QTDB, by minimizing all deficiencies highlighted in this study, or the adoption of another alternative more accurately annotated database as a reference.

Acknowledgements

Research supported by grants TEC2014-52250-R and DPI2017-83952-C3 MINECO/AEI/FEDER, UE.

References

[1] Sörnmo L, Laguna P. Ch. 6 - the electrocardiogram-a brief background. In *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Biomedical Engineering.

Burlington: Academic Press. ISBN 978-0-12-437552-9, 2005; 411 – 452.

[2] Moody GB, Mark RG. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* May 2001;20(3):45–50. ISSN 0739-5175.

[3] Taddei A, Distanti G, Emdin M, Pisani P, Moody GB, Zeelenberg C, Marchesi C. The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *European Heart Journal* 1992;13(9):1164–1172.

[4] Willems JL, Arnaud P, Bommel JHV, Bourdillon PJ, Degani R, Denis B, Graham I, Harms FMA, Macfarlane PW, Mazzocca G, Meyer J, Zywiets C. A reference data base for multilead electrocardiographic computer measurement programs. *Journal of the American College of Cardiology* 1987;10(6):1313 – 1321. ISSN 0735-1097.

[5] Laguna P, Mark RG, Goldberg A, Moody GB. Database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In *Computers in Cardiology*. 1997; 673–676.

[6] Magnani JW, Williamson MA, Ellinor PT, Monahan KM, Benjamin EJ. P-wave indices: current status and future directions in epidemiology, clinical, and research applications. *Circ Arrh and electroph* Feb 2009;2:72–9.

[7] Rieta JJ, Alcaraz R. The Genesis of the Electrocardiogram (ECG). In *Wiley Encyclopedia of Electrical and Electronics Engineering*. Hoboken, NJ, USA: John Wiley & Sons, Inc., February 2017; 1–15.

[8] Zhang JW, Liu X, Dong J. CCDD: An enhanced standard ECG database with its management and annotation tools. *Int Jml on Artificial Intelligence Tools* 2012;21(05).

[9] Beraza I, Romero I. Comparative study of algorithms for ECG segmentation. *Biomedical Signal Processing and Control* April 2017;34:166–173.

[10] Bora PK, Sinha R, Yadav SK. Electrocardiogram signal denoising using non-local wavelet transform domain filtering. *IET Signal Processing* February 2015;9(1):88–96.

[11] González F, Alcaraz R, Rieta JJ. Electrocardiographic P-wave delineation based on adaptive slope gaussian detection. In *Computing in Cardiology Conference (CinC)*, volume 44. IEEE, 2017; In press.

[12] Lenis G, Pilia N, Oesterlein T, Luik A, Schmitt C, Dössel O. P-wave detection and delineation in the ECG based on the phase free stationary wavelet transform and using intracardiac atrial electrograms as reference. *Biomedizinische Technik* 2016;61(1):37–56.

[13] Martínez A, Alcaraz R, Rieta JJ. Application of the phasor transform for automatic delineation of single-lead ECG fiducial points. *Physiol Meas* 2010;31(11):1467–1485.

Address for correspondence:

Francisco González Molina
Electronic Engineering Department, Building 7F
Universidad Politécnica de Valencia
Camino de Vera, s/n, 46022, Valencia, Spain
E-mail: fgmolina@upv.es