Ruiz, A.; Arias, MA.; Pachón, MI.; Langley, P.; Rieta, JJ.; Alcaraz, R. (2019). Thorough Assessment of a P-wave Delineation Algorithm Through the Use of Diverse Electrocardiographic Databases. IEEE. 1-4. https://doi.org/10.1109/EHB47216.2019.8970078

Thorough Assessment of a P-wave Delineation Algorithm Through the Use of Diverse Electrocardiographic Databases

Antonio Ruiz[1], Miguel A. Arias[2], María Inmaculada. Pachón[2], Philip Langley[3], José J. Rieta[4], Raúl Alcaraz[1]

[1]Research Group in Electronic, Biomedical and Telecommunication Engineering, University of Castilla-La Mancha,Cuenca, Spain; {antonio.ruiz, raul.alcaraz}@uclm.es

[2]Cardiac Arrhythmia Department, Hospital Virgen de la Salud, Toledo, Spain

[3]School of Engineering and Computer Science, University of Hull, Hull, UK; p.langley@hull.ac.uk

[4]BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Valencia, Spain; jjrieta@upv.es

***Abstract*—Detailed analysis of the electrocardiogram (ECG) recording allows to obtain relevant information about the heart state and its behavior. Indeed, the P-wave study has proven to be helpful in diagnosing many atrial electrical conduction disorders. However, manual delineation of this wave is a very time-consuming task, and a wide variety of automatic algorithms have been proposed for that purpose. The performance of these methods has been mostly validated on the QT database, since it contains manual annotations from clinical experts and is freely available at PhysioNet. Unfortunately, in the last years several authors have shown some major limitations of this database and this work hence presents a more robust validation of a previously published P-wave delineator. Thus, two novel databases which have been very recently made available to the scientific community have been analyzed. The results show that the methods ability to detect P-waves is close to 100% in stable sinus rhythm episodes, but it decreases to 94% when the heart rhythm is irregular or consecutive P-waves exhibit notable morphological variations. Regarding the error between the annotations obtained automatically and those available as a reference, no relevant differences have been observed in terms of standard deviation, thus suggesting that the algorithm achieves a consistent delineation of the P-waves. Contrarily, a notable error has been noticed in average values, thus highlighting the need of a more extensive training for the P-wave delineator than that offered by the QT database.**

***Keywords*—Electrocardiogram; P-wave Detection; P-Wave Delineation; Robust Validation**

I. Introduction

In the surface electrocardiogram (ECG) recording the P-wave reflects atrial activity, being therefore the best source of non-invasive information about its electrical conduction and the different associated pathologies [1]. In fact, nowadays it is clinically accepted that some parameters obtained directly from this wave, including its duration, its dispersion or its morphology, provide information on some atrial conduction disorders, such as atrial fibrillation [2]. However, it is still necessary to make a strong effort to standardize the way to extract such information, since the P-wave presents very low amplitude and, consequently, is greatly affected by the presence of noise. Additionally, still there is no a single definition of its fiducial points, that is, of the points marking its beginning, its peak and its end [3]. Another aspect to be highlighted is that manual delineation of the P-wave is a very expensive task in time and effort for clinical experts, depending largely on their experience [2].

To mitigate these limitations, in recent years a broad variety of algorithms have been proposed to automatically detect and delineate P-waves in long-term ECG recordings. Most of them are based on transforming the ECG signal to enhance the P-wave and thus facilitate its detection and delineation. Some examples are the methods based on the phasor transform [4] or on the Wavelet transform [5]. However, they can cause morphological alterations in the P-wave, thus obtaining inaccurate delineations in some scenarios. For this reason, other algorithms have been recently proposed, which are based on modeling the P-wave using different patterns, so that its initial morphology is not altered, and the noise effect is minimized [6].

The performance of these methods has been mostly validated making use of the QT database, which is freely available at PhysioNet [7] and also contains manual annotations from experts on the beginning, the peak and the end of the P-waves [8]. However, in the last years several authors have evidenced some major limitations of this database, such as the lack of information on the specific channel used for delineating each P-wave, the diversity of criteria used in the annotation process, or the noise effect on the result of the manual annotation [9], [10]. As a consequence, they have also suggested the need for a more robust validation of the P-wave delineation algorithms on new carefully annotated databases [11]. In this context, the present work validates the

performance of a previously published P-wave delineator [6] with two novel databases, which have been recently made available to the scientific community [11].

II. Methods

A. Delineation algorithm of the P-wave

Although the P-wave delineator has been previously described in [6], to render the work easier to read, its performance is briefly summarized. Thus, the method begins by preprocessing the ECG signal to eliminate common noise from its acquisition and by identifying all R-peaks [6].

Next, the delineation process is mainly based on the creation of a gaussian model of the P-wave to serve as a reference. Hence, each wave is not treated independently, but certain parameters computed from the preceding waves guide the location of its fiducial points. These parameters include the approximate distance between the corresponding R-peak and the P-wave (dPR); its approximate duration (durP), which defines a search window expanding its width on each side by 25% (vbP); the type of morphology; the distance in time and amplitude between the maximum peak and each of the wave limits (amplIni, amplFin, tIni, tFin); and finally some initial coefficients of the gaussian function to facilitate the search for the best new adjustment. Some of these parameters are shown in Figure 1. Thus, these parameters are updated for each new wave processed with an impact ratio of 20% on the existing values. In addition, before that, it is checked if the difference between the two values is greater than 25%. In that case, the wave is classified as abnormal and the values will not be updated, thus obtaining a real-time recording of the P-waves presenting an abnormal morphology.

The initial values for these parameters are obtained through an initialization stage, where a reference P-wave is generated. Thus, the ECG segments preceding the first five R-peaks are taken, and a reference interval is created by averaging. Figure 1(a) shows an example of this segment, on which the reference P-wave will be first detected, its morphology will be later identified, and its fiducial points will be finally detected. These three processes are next described. Nonetheless, it should be noted that the first and third ones will be recursively used on each new P-wave.

For the P-wave detection, the largest amplitude peak is sought in a window that begins at the QRS complex onset and extends backwards with a duration equal to one third of the average RR-interval. Around the maximum peak detected, a segment is isolated to delineate the P-wave. The length of this interval is set at 180ms, although it is extended or reduced by 20ms when the average RR-interval is too long or short, respectively. Subsequently, the detected P wave is classified as positive monophasic, negative monophasic or biphasic, with either positive-negative or negative-positive polarity. For this purpose, a decision tree based on gaussian fitting is used [6].

Finally, for the P-wave delineation, the generated gaussian model is divided in half, and the two halves are separately processed. Then, for each case a slope threshold is computed to locate the limits of the wave as the first points exceeding

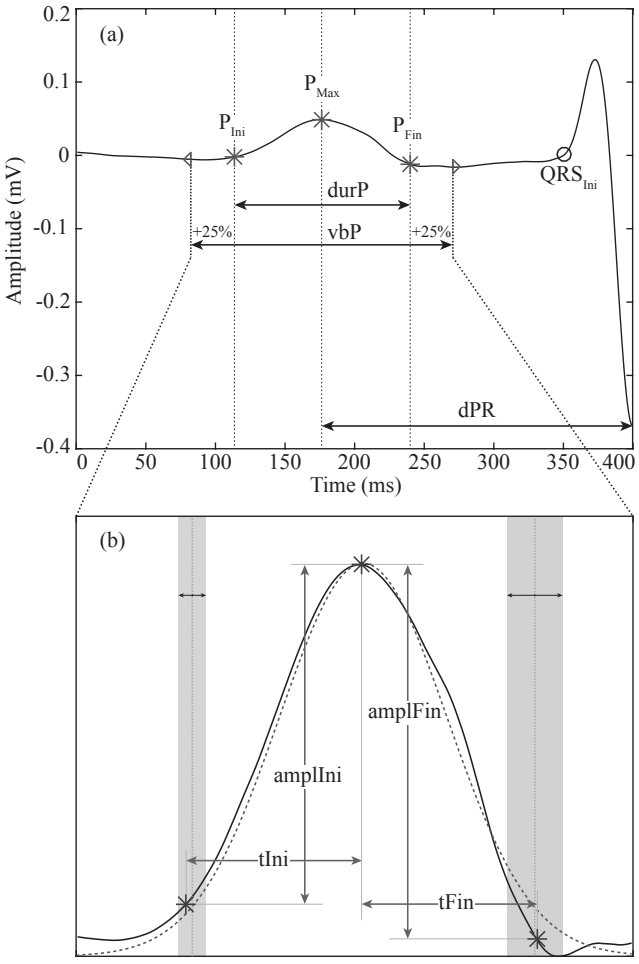

Fig. 1. Example of the P-wave delineation. In the panel (a) the fundamental parameters used in the delineation process are represented. In the panel (b) the real P-wave (solid line) and its gaussian model (dot line) are displayed, along with the search windows and fiducial points established.

it. Next, a search interval is defined around of each identified point, as can be seen in Figure 1(b) in the form of shaded areas. Thus, the same process is repeated on the preprocessed ECG, but the search is restricted to the established interval, whose width depends on how well the model fits the ECG signal. Note that the more similar the Gaussian fitting to the P-wave, the narrower this search interval. As an example, Figure 1(b) shows how this interval is narrower for the P-wave onset than for its end, because the first half of the wave better fits to the gaussian model.

B. Fine-tuning of the delineation algorithm

As also described in [6], the P-wave delineation method was trained using some ECG recordings from the QT database (QTDB). In fact, to establish the relationship between the maximum slope and the slope in the boundaries in each P-wave defining the aforementioned threshold, 60 P-waves extracted from different recordings were analyzed. In addition, the overall performance of the method was evaluated with the complete database, which contains 105 ECG signals with two

leads and 15 minutes in length, where at least 30 beats were manually annotated.

C. Validation of the delineation algorithm

For a more robust and independent validation of the P-wave delineator, in the present work two novel databases, which have been recently made available to the scientific community, have been analyzed. Thus, the dataset MIT-BIH P-wave Annotations (MBPDB), available at PhysioNet [7], has been firstly considered. It consists of 12 signals extracted from the database MIT-BIH Arrhythmia database, where two experts have marked the presence of all P-waves. It should be noted that these recordings have a duration of 30 minutes, in which sinus rhythm episodes are alternated with arrhythmias, such as atrial fibrillation, tachycardia, bradycardia, etc., thus involving a very challenging scenario for every P-wave detector.

On the other hand, the LU electrocardiography database (LUEBD), described in [11] and freely available through a Lobachevsky University repository, has also been analyzed. It contains 200 12-lead ECG recordings with 10 seconds in length, which have been acquired at a sample rate of 500 Hz from healthy volunteers and patients with different cardiovascular diseases. Unlike the global annotations jointly obtained for the 2-lead ECG recordings in the QTDB [8], the boundaries and peaks of QRS, P and T waves were manually determined by visual inspection of each individual lead. In total, the database contains 16,797 annotated P waves, that is about five times greater than the widely used QTDB.

III. Results

As in the previous work where the P-wave delineator was defined [6], its performance has been summarized through two indices. On the one hand, the sensitivity (Se) has been computed as the number of P-waves correctly detected regarding the total of those manually identified. On the other hand, the distance between the annotations automatically obtained by the method and those manually marked by the experts were calculated and expressed in terms of its mean absolute value (μ) and standard deviation (σ) for the three fiducial points of the P-wave, i.e., for its onset ($\mathrm{P_{on}}$), its peak ($\mathrm{P_{peak}}$) and its end ($\mathrm{P_{end}}$).

The results obtained for the two analyzed databases, as well as those originally obtained from the QTDB [6], are displayed in Table I. The latter information is included to serve as a reference and make comparison easier. As can be seen, for the MBPDB only results about P-wave detection are presented, since in this case the experts only indicated the presence or absence of the wave for each beat and did not delineate its fiducial points. Contrarily, values of sensitivity and location error for P-wave delineation are presented from the other two datasets, i.e. LUEDB and QTDB.

More precisely, all P-wave fiducial points were detected from the LUEDB with a sensitivity quite similar to the QTDB, only losing a reduced number of marks. Similarly, for both datasets no significant differences were noticed in terms of the location error standard deviation (σ) for the three analyzed points, i.e., for $\mathrm{P_{on}}$, $\mathrm{P_{peak}}$, and $\mathrm{P_{en}}$. Contrarily, a similar finding is not applicable for the mean location error, because notable deviations between both datasets were seen.

TABLE I
THE ALGORITHM'S ABILITY TO DETECT AND DELINEATE P-WAVES IN THE ANALYZED DATABASES

Database	Performance Index	$\mathrm{P_{on}}$	$\mathrm{P_{peak}}$	$\mathrm{P_{end}}$
MBPDB	Se (%)		94.71	
LUEDB	Se (%)	99.71	100	99.90
	$\mu \pm \sigma$ (ms)	-12.4 ± 10.1	1.6 ± 5.0	5.1 ± 9.7
QTDB	Se (%)	100	100	100
	$\mu \pm \sigma$ (ms)	4.7 ± 9.6	2.8 ± 6.7	0.6 ± 9.8

IV. Discussion

The ECG recordings collected by the MBPDB have been obtained from patients suffering different types of arrhythmias, so that they present a fairly realistic scenario where ever P-wave delineation algorithm will commonly have to work. In fact, the P-wave is usually analyzed in patients prone to supraventricular arrhythmias and, therefore, during their long-term monitoring it is highly probable to find arrhythmic episodes interspersed with others of sinus rhythm. In this context, the analyzed algorithm has still shown a promising capacity to detect P-waves, since a value of sensitivity greater than 94% has been reported. However, this outcome is notable lower than that reported in sinus rhythm conditions.

The main cause of this loss of sensitivity could be due to the fact that the algorithm is based on the distance between consecutive R-peaks, as well as on the adaptive gaussian modeling of each P-wave from its preceding waveforms. Although this way of working offers good results in stable sinus rhythm episodes, irregular heart rates or remarked morphological alterations in consecutive P-waves might lead the algorithm to make relevant errors. As an example, Figure 2 shows how P-waves are incorrectly detected and delineated in presence of ventricular ectopic beats, which make the heart rhythm totally irregular. Hence, for a better performance in a long-term monitoring context, the analyzed algorithm should be slightly modified to identify abrupt heart rate alterations, as well as sudden changes in P-wave morphology, and then take specific measures to successfully detect and delineate P-waves.

On the other hand, the values of sensitivity reported for the LUEDB have been very similar to those obtained with the QTDB, thus highlighting that the method properly works in short and stable ECG intervals from a variety of cardiac disorders. Similarly, the location error standard deviation (σ) has proven to be similar for both datasets, thus suggesting a fairly consistent algorithm's behavior in both cases. Moreover, it is interesting note that this performance index was always lower than the well-accepted two standard deviation tolerances, recommended by the CSE working party [12].

Contrarily, the mean location error have provided notable differences for both databases. This could be due to the fact that P-waves were manually annotated through two different

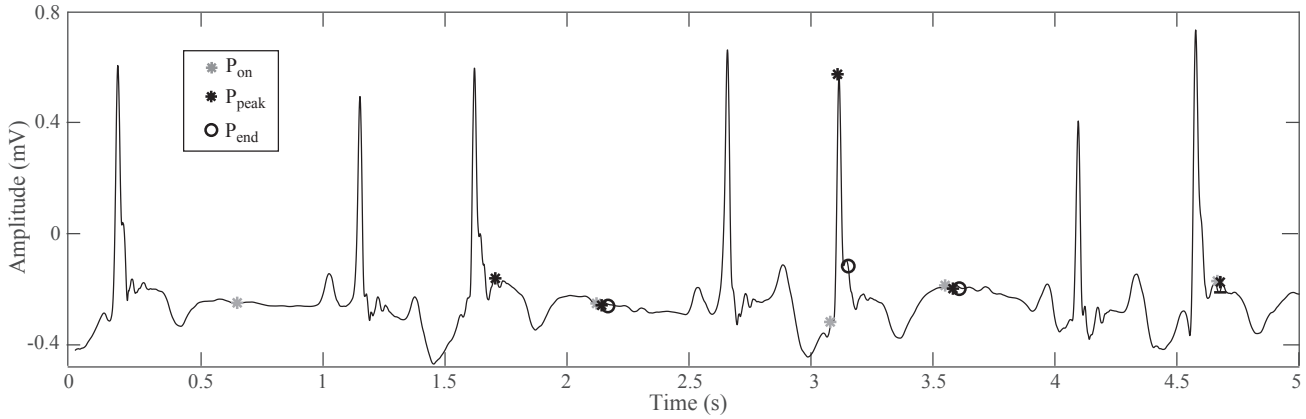

Fig. 2. Example where the analyzed algorithm unsuccessfully detects and delineates P-waves. The heart rate is irregular due to the presence of ventricular ectopic beats.

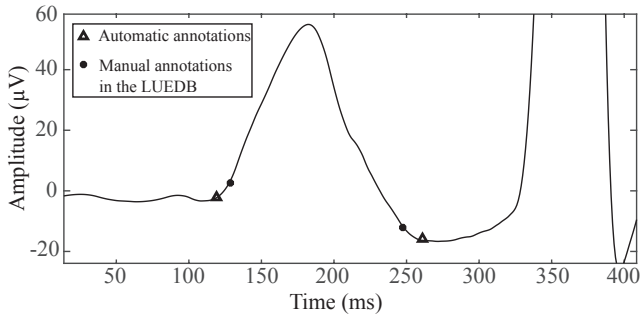

Fig. 3. Differences between the automatic annotations provided by the analyzed algorithm and those manually marked in the LUEDB.

approaches. Thus, whereas the experts made global annotations for the 2-lead ECG recordings available in the QTDB, each lead was individually analyzed for the LUEDB [11]. To this respect, Figure 3 shows how the algorithm tends to locate the fiducial points at the base of the P-wave, but manual annotations in the LUEDB are placed at points presenting a higher slope. Hence, this result suggests that the algorithm requires a much broader validation than that allowed by the QTDB to avoid its overfitting and allow its successfully performance in single-lead ECG recordings.

V. Conclusions

A robust validation of a previously published P-wave delineator has been presented by analyzing two novel databases of annotated ECG recordings, beyond the well-known QT dataset. The results have shown that the algorithm successfully works for a variety of cardiac diseases, but it should still be modified to identify irregular heart rates and abrupt changes in the P-wave morphology. In this way, its ability to delineate P-waves in challenging scenarios that often appear in long-term monitoring of patients prone to supra-ventricular arrhythmias could be improved. Moreover, the obtained outcomes also suggest that every P-wave detector should be validated on a wider set of ECG recordings than that offered by the QT database to reduce its overfitting as much as possible.

Acknowledgments

Research supported by grants DPI2017–83952–C3 from MINECO/AEI/FEDER UE, SBPLY/17/180501/000411 from JCCLM and AICO/2019/036 from GVA.

References

[1] P. G. Platonov, "Atrial conduction and atrial fibrillation: what can we learn from surface ECG?" *Cardiol J*, vol. 15, no. 5, pp. 402–7, 2008.

[2] J. W. Magnani, M. A. Williamson, P. T. Ellinor, K. M. Monahan, and E. J. Benjamin, "P wave indices: current status and future directions in epidemiology, clinical, and research applications," *Circ Arrhythm Electrophysiol*, vol. 2, no. 1, pp. 72–9, Feb 2009.

[3] K. J. Hari, T. P. Nguyen, and E. Z. Soliman, "Relationship between P-wave duration and the risk of atrial fibrillation," *Expert Rev Cardiovasc Ther*, vol. 16, no. 11, pp. 837–843, Nov 2018.

[4] A. Martínez, R. Alcaraz, and J. J. Rieta, "Application of the phasor transform for automatic delineation of single-lead ECG fiducial points," *Physiol Meas*, vol. 31, no. 11, pp. 1467–85, Nov 2010.

[5] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ECG delineator: evaluation on standard databases," *IEEE Trans Biomed Eng*, vol. 51, no. 4, pp. 570–81, Apr 2004.

[6] F. González, R. Alcaraz, and J. J. Rieta, "Electrocardiographic P-wave delineation based on adaptive slope Gaussian detection," in *Computing in Cardiology*, 2017, pp. 1–4.

[7] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–20, Jun 2000.

[8] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG," in *Computers in Cardiology*, 1997, pp. 673–676.

[9] F. González, R. Alcaraz, and J. J. Rieta, "The physionet QT database: Study on the reliability of P-wave manual annotations under noisy recordings," in *Computing in Cardiology*, 2017, pp. 1–4.

[10] L. Y. Di Marco and L. Chiari, "A wavelet-based ECG delineation algorithm for 32-bit integer online processing," *Biomed Eng Online*, vol. 10, p. 23, Apr 2011.

[11] A. I. Kalyakulina, I. I. Yusipov, V. A. Moskalenko, A. V. Nikolskiy, A. A. Kozlov, K. A. Kosonogov, N. Y. Zolotykh, and M. V. Ivanchenko, "LU electrocardiography database: a new open-access validation tool for delineation algorithms," *arXiv preprint arXiv:1809.03393v2*, 2018.

[12] "Recommendations for measurement standards in quantitative electrocardiography. The CSE Working Party," *Eur Heart J*, vol. 6, no. 10, pp. 815–25, 1985.