Proceedings

Automated P-wave quality assessment for wearable sensors[†]

Diogo Tecelão [1,*] and **Peter Charlton** [2]

[1] Departamento de Física, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Lisbon, Portugal

[2] Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, King's College London, King's Health Partners, London, UK; peter.charlton@kcl.ac.uk

* Correspondence: d.tecelao@campus.fct.unl.pt

† Presented at the 5th International Electronic Conference on Sensors and Applications, 15–30 November 2018; Available online: https://sciforum.net/conference/ecsa-5.

Abstract: Hospital patients recovering from major cardiac surgery are at risk of paroxysmal atrial fibrillation (AF), an arrhythmia which can be life-threatening. Wearable sensors are routinely used for electrocardiogram (ECG) monitoring in patients at risk of AF, providing real-time AF detection. However, wearable sensors could have greater impact if used to identify the subtle changes in P-wave morphology which precede AF. This would allow prophylactic treatment to be administered, potentially preventing AF. However, ECG signals acquired by wearable sensors are susceptible to artefact, making it difficult to distinguish between physiological changes in P-wave morphology, and changes due to noise. The aim of this study was to design and assess the performance of a novel automated P-wave quality assessment tool to identify high quality P-waves, for AF prediction. We designed a two-stage algorithm which uses P-wave template matching to assess quality. Its performance was assessed using the AFPDB, a database of wearable sensor ECG signals acquired from both healthy subjects and patients susceptible to AF. The algorithm's quality assessments of 97,989 P-waves were compared to manual annotations. The algorithm identified high quality P-waves with high sensitivity (93%) and good specificity (82%), indicating that it may have utility for identifying high quality P-waves in wearable sensor data. Measurements of P-wave morphology derived from high quality P-waves could be used to predict AF, improving patient outcomes and reducing healthcare costs. Further studies assessing the clinical utility of the presented tool are warranted for validation.

Keywords: electrocardiogram; P-wave; quality assessment; atrial fibrillation.

1. Introduction

Continuous electrocardiogram (ECG) monitoring using wearable sensors enables early identification of several types of arrhythmia. However, data acquired by wearable sensors are susceptible to artefacts (e.g., due to poor sensor contact or motion artefact). Hence, a challenge in the use of continuous wearable monitoring, where data are collected without clinical supervision, is ensuring that only high quality signals are used to derive clinical measurements. Physiological parameters extracted from artefact-corrupted signals may be inaccurate, which can lead to a high frequency of false alarms [1]. Therefore, assessment of signal quality is a crucial step for accurate and precise data analysis, such as extracting features, identifying deteriorations, and generating alerts.

Hospital patients recovering from major cardiac surgery are at high risk of paroxysmal atrial fibrillation (AF), an arrhythmia which can be life-threatening. Wearable sensors are routinely used to monitor patients at risk of AF. The sensors typically have 3 or 5 ECG leads which are attached

to the chest and an optional wired finger probe, and are carried in a bag tied to the patient. The sensors provide ECG, respiratory rate and arterial blood oxygen saturation monitoring. The data are transmitted in real-time to a central monitoring station, and alarms are sounded when arrhythmias, such as AF, are detected. Hence wearable sensors currently provide early recognition of AF, allow for it to be treated promptly, and help minimise the resulting complications. However, wearable sensors could have greater impact if used to predict AF before its onset, rather than simply detecting it when it occurs. This would allow prophylactic treatments to be administered, potentially preventing the arrhythmia, and consequently reducing the risk of complications, and healthcare costs.

Subtle alterations in P-wave morphology have been found to be predictive of AF [2,3]. However, the relatively low amplitude of the P-wave makes it highly susceptible to noise, severely affecting the extraction and quantification of its features. This makes it difficult to distinguish between physiological changes in morphology and changes due to noise. Thus, the clinical utility of techniques for predicting AF could be highly limited by parameters being estimated erroneously from low quality P-waves, causing false alerts. Therefore, it is essential to exclude artefact-corrupted P-waves from analyses. Previously, this has been performed either manually, where expert cardiologists excluded unreliable P-waves by hand [2], or automatically, with a conventional template-comparison method [3,4]. In the latter, individual P-waves were discarded if they had a cross-correlation coefficient lower than 0.7 with a template P-wave obtained with an averaging procedure [3,4].

In this study we built on the previous work by developing an optimised P-wave quality index (PQI) designed to identify high quality P-waves, from which clinical measurements can be reliably derived. This tool was designed using several P-wave quality assessment features, and its performance was assessed through comparison with manually annotated P-waves from a database of ECG recordings acquired using wearable sensors.

2. Materials and Methods

2.1. Data description

The openly available paroxysmal AF prediction database (AFPDB) [5] was used to develop and assess the performance of the PQI. This database was originally assembled to develop techniques for predicting paroxysmal AF. The dataset contains wearable ECG recordings during sinus rhythm, acquired from both healthy controls and patients who subsequently developed AF. All ECG signals have a duration of 30 minutes, and are excerpts from recordings acquired with a two-channel long-term Holter wearable sensor at a sampling frequency of 128Hz and with a 12-bit resolution [5].

This study used 44 records (23 from controls and 21 from AF patients) from the AFPDB. After pre-processing with a bandpass filter with cut-off frequencies of 0.5 and 40Hz, the ECG lead in which P-waves were most visible for each record was chosen, and P-wave quality was manually annotated and double-checked. P-waves were classified into three distinct classes, as illustrated in Figure 1:

Class A: High quality clean P-waves;

Class B: Complete noise and absent P-waves. This included P-waves with *no resemblance* to normal P-wave morphology, either due to motion artefacts, severe baseline wander, muscular activation interference, or simply the absence of a P-wave;

Class C: Unreliable, noise-distorted P-waves. This included P-waves that had *some resemblance* to normal P-wave morphology, but were still unreliable (i.e., their morphology was still excessively affected by noise). The degree of distortion varied from mildly to heavily distorted (see Figure 1).

This resulted in 22 hours of recording, corresponding to 97,989 P-waves: 88,606 annotated as class A (90.4%), 5,102 as class B (5.2%), and 4,281 as class C (4.4%).
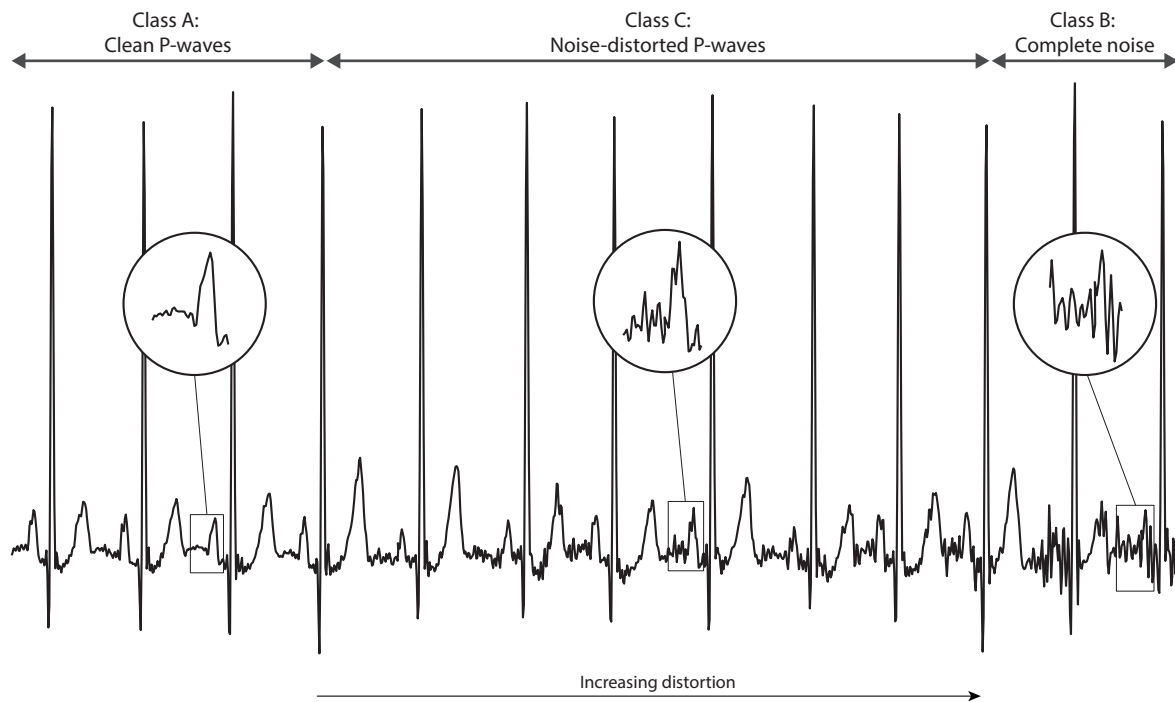
Figure 1. P-waves were classified into three different classes: high quality clean P-waves (class A), unreliable, noise-distorted P-waves (class C), and complete noise or absent P-waves (class B). Even though class C P-waves had some resemblance to normal P-wave morphology, they were still considered unreliable. For that class, the degree of distortion varied from mildly to heavily distorted.

2.2. P-wave Quality Index (PQI) algorithm

The PQI tool algorithm is depicted in Figure 2. Briefly, the process started with P-wave detection and signal extraction, and was followed by two different decision-making stages based on template-comparisons. The first decision stage aimed to remove P-waves with no resemblance to the normal P-wave morphology (class B), while the second was more refined, removing P-waves still excessively distorted by noise, and hence unreliable (class C).
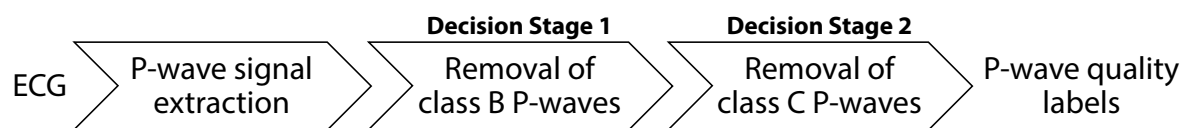

Figure 2. The P-wave quality index (PQI) tool consisted of three steps. In the first, the P-waves were identified in the electrocardiogram (ECG), and extracted. P-wave template morphologies and P-wave features were extracted from these P-waves, and were used in two different decision-making stages. This resulted in each P-wave being labelled as either high (class A) or low quality (classes B and C), allowing artefact-corrupted and unreliable P-waves to be excluded from the analysis.

2.2.1. P-wave signal extraction

The first step towards assessing the P-waves' quality was to detect them in the ECG and extract them. This signal extraction step was of high importance, as proper P-wave quality assessment was only possible with accurate P-wave signal extraction. Firstly, P-wave peaks were identified. To do so, each R-peak was identified using the widely used Pan, Hamilton and Tompkins algorithm [6], and then delineated using the phasor transform delineation algorithm [7]. Finally, P-wave signals were

then extracted by taking a window of width 300ms, centred on the corresponding P-wave peak. This window length ensured that the whole P-wave plus its surrounding baseline signal were captured.

2.2.2. Decision stage 1: Removal of completely noisy or absent P-waves

The aim of the first decision-making stage was to exclude P-waves which were heavily distorted by noise or absent P-waves (class B). This first stage had the additional purpose of removing P-waves that could influence the more refined templates created during the second decision-making stage.

A P-wave template was created for each 30-minute recordings as the average shape of the P-waves in that time period. Then, those P-waves were aligned with that template, and several features were extracted in order to retrieve information about them. These features were then used as candidate features for the PQI to assess P-wave quality. The following features were tested:

1. P-wave template comparison features:

    (a) Pearson's correlation coefficient between P-wave and template (T_{corr});
    (b) Root mean square difference between P-wave and template (T_{rmsd});
    (c) Pearson's correlation coefficient between P-wave and template derivative (dT_{corr});
    (d) Root mean square difference between P-wave and template derivative (dT_{rmsd}).

2. P-wave signal features:

    (a) Standard deviation of P-wave (P_{std});
    (b) Number of zero-crossings of P-wave (P_{zc});
    (c) Area of P-wave (P_{area});
    (d) Kurtosis of P-wave ($P_{kurtosis}$);
    (e) Skewness of P-wave ($P_{skewness}$);
    (f) Standard deviation of P-wave derivative (dP_{std}).

A decision tree was built using a selection of those features (Figure 3). This model was estimated assuming both classes' probability as equal, allowing one to balance sensitivity and specificity. Furthermore, the decision tree had a maximum of five splits (decisions), in order to avoid overfitting.

2.2.3. Decision stage 2: Removal of distorted P-waves

More refined P-wave templates could be created after removal of completely distorted P-waves. These were then used in a final decision stage with the aim of removing distorted and unreliable P-waves (class C), whilst being able to accommodate possible P-wave morphological variations.

This second decision stage was similar to the first, but with the single difference that a template was created, to which P-waves were aligned, every 20 P-waves (instead of every 30 minutes). This aimed to accommodate both physiological variations in the ECG over time and the greater P-wave variability that precedes AF [2,3]. Similarly to the first decision stage, a decision tree was built, but this time maximising the ability to correctly identify class A P-waves (Figure 3).

2.3. P-wave Quality Index (PQI) performance assessment

The performance of the PQI was assessed on the 44 records through comparison with the manual annotations. Class A clean P-waves were considered as the positive class, while classes B and C were merged into one unique negative class. In addition, given the higher variability in P-wave morphology present in patients susceptible to AF, a sub-group analysis of the performance in controls and patients who experienced AF after the recording was conducted, and significance was assessed using a two-sample *t*-test at the 5% significance level. Sensitivity and specificity metrics were used to assess performance, as they are independent of class distributions, and therefore are able to provide comprehensive assessment of imbalanced learning problems, such as the present one.
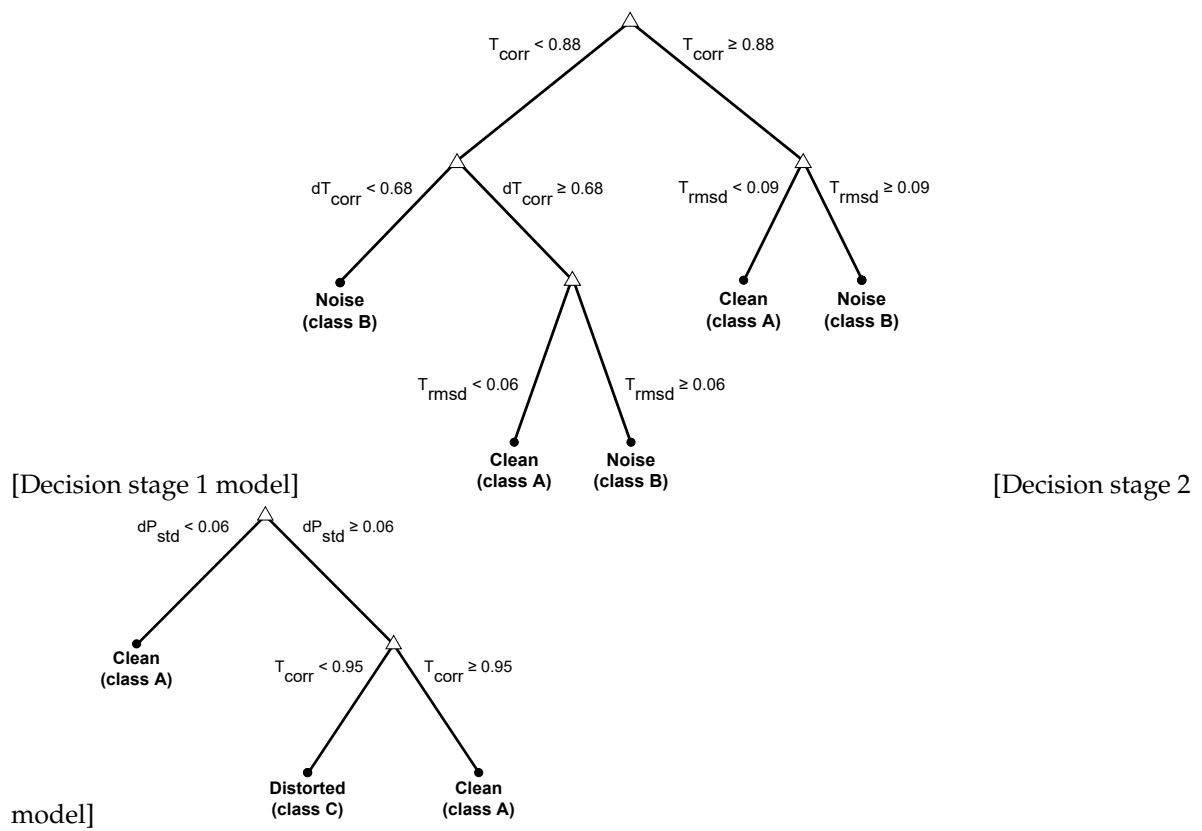
[Decision stage 1 model]

[Decision stage 2 model]

Figure 3. Decision trees used in the decision-making stages of the P-wave quality index (PQI) tool. These models used features obtained from the P-wave signal and P-wave template comparisons.

3. Results

3.1. P-wave Quality Index (PQI) performance assessment

The PQI was able to identify high quality P-waves with high sensitivity (93%) and good specificity (82%). Furthermore, no statistically significant difference was found in sensitivity or specificity between the control and AF groups (Table 1).

Table 1. Performance of the P-wave quality index (PQI) tool on all 44 records, and comparison of performance between healthy controls and patients susceptible to atrial fibrillation (AF).

	Total	Controls	AF Patients	*P*-value
Sensitivity	93%	92%	95%	0.33
Specificity	82%	76%	86%	0.08

4. Discussion

Hospital patients recovering from major cardiac surgery are at risk of AF, which can be life-threatening. Wearable sensors are routinely used for ECG monitoring in the postoperative period, and could have greater impact if used to identify the subtle changes in P-wave morphology which are predictive of AF. This would potentially allow AF to be prevented, reducing risks and costs. However, ECG signals acquired by wearable sensors are susceptible to artefact, making it difficult to distinguish between physiological changes in P-wave morphology, and changes due to noise. Hence, the future implementation of techniques that use the P-wave to predict AF will rely on proper P-wave quality assessment.

In this study we proposed the PQI, a novel and optimised tool for automatically identifying low quality P-waves. Briefly, the algorithm starts by detecting and extracting the P-waves' signal, which is then used in two decision-making stages: the first with the aim of removing highly noisy or absent P-waves, and the second with the aim of removing less distorted, but still unreliable, P-waves. The PQI identified high quality P-waves with high sensitivity (93%) and good specificity (82%), and performed similarly on healthy subjects and patients susceptible to AF, indicating that it was able to accommodate the P-wave variability that precedes AF [2,3]. The high performance of the PQI suggests that it may have utility for identifying high quality P-waves in wearable sensor data, which could be used to perform unsupervised predictions of AF.

The proposed tool was built and tested using a large dataset containing almost 100,000 manually annotated P-waves across different morphologies, and trialled 10 different P-wave quality assessment features, some of them novel. In addition, the proposed tool was built using simple decision tree models, making it more likely to perform well on novel data. The proposed tool can work on a near real-time basis, an important feature of quality assessment tools for use in continuous monitoring. Despite requiring 30-minute signal portions, such time precision is enough to predict and act upon atrial arrhythmias such as AF.

Even though the presented tool exhibited high performance, future assessments of its clinical utility are warranted for validation. For instance, future studies should investigate whether the use of the PQI results in more precise measurements of P-wave features, and whether the PQI can be used to improve unsupervised predictions of AF. In addition, future studies should assess the PQI on independent datasets containing heart rhythms other than sinus rhythm and AF. Finally, the methodology of the PQI might be further improved in future studies. For instance, it is assumed that, during the creation of a P-wave template each 30 minutes, noise is cancelled out and, therefore, the obtained signal reflects a clean P-wave. This may not be the case during poor electrode contact for long periods of time. This can be safeguarded against in future works with the addition of a

template-verification stage where, for example, the obtained template signals could be compared with a Gaussian function.

5. Conclusions

This paper presented a novel P-wave quality assessment tool, which was able to identify high quality P-waves with high sensitivity (93%) and good specificity (82%). Measurements of P-wave morphology derived from high quality P-waves could be used to predict AF using wearable ECG monitoring, potentially improving patient outcomes and reducing healthcare costs. Further studies assessing the clinical utility of the PQI tool are warranted for validation.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Orphanidou, C. *Signal Quality Assessment in Physiological Monitoring: State of the Art and Practical Considerations*, 1 ed.; Springer International Publishing: Cham, 2018; p. 63. doi:10.1007/978-3-319-68415-4.
2. Martínez, A.; Alcaraz, R.; Rieta, J.J. Morphological variability of the P-wave for premature envision of paroxysmal atrial fibrillation events. *Physiological Measurement* **2014**, *35*, 1–14. doi:10.1088/0967-3334/35/1/1.
3. Censi, F.; Corazza, I.; Reggiani, E.; Calcagnini, G.; Mattei, E.; Triventi, M.; Boriani, G. P-wave Variability and Atrial Fibrillation. *Scientific Reports* **2016**, *6*, 1–7. doi:10.1038/srep26799.
4. Censi, F.; Ricci, C.; Calcagnini, G.; Triventi, M.; Ricci, R.P.; Santini, M.; Bartolini, P. Time-Domain and Morphological Analysis of the P-Wave. Part I : Technical Aspects for Automatic Quantification of P-Wave Features. *PACE: Pacing and Clinical Electrophysiology* **2008**, *31*, 874–883. doi:10.1111/j.1540-8159.2008.01102.x.
5. Moody, G.B.; Goldberger, A.L.; McClennen, S.; Swiryn, S.P. Predicting the onset of paroxysmal atrial fibrillation: The computers in cardiology challenge 2001. *Computers in Cardiology* **2001**, *28*, 113–116. doi:10.1109/CIC.2001.977604.
6. Pan, J.; Tompkins, W.J. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering* **1985**, *3*, 230–236. doi:10.1109/TBME.1985.325532.
7. Martínez, A.; Alcaraz, R.; Rieta, J.J. Application of the phasor transform for automatic delineation of single-lead ECG fiducial points. *Physiological Measurement* **2010**, *31*, 1467–1485. doi:10.1088/0967-3334/31/11/005.