Abstract: Heart diseases are one of the most frequent causes of death in the modern world. Therefore, the ECG signal features have been under peer review for decades to improve medical diagnostics. In this paper, we provide smoothing of the atrial premature complex (APC) of the electrocardiogram (ECG) signal using unbiased finite impulse response (UFIR) smoothing filtering. We investigate the P-wave distribution using the Rice law and determine the probabilistic confidence interval based on a database associated with normal heartbeats. It is shown that the abnormality in the APC is related to the P-wave morphology. Different filtering techniques employing predictive and smoothing filtering are applied to APC data and compared experimentally. It is demonstrated that UFIR smoothing provides better performance among others. We finally show that the P-wave confidence interval defined for the Rice distribution can be used to provide an automatic diagnosis with a given probability.

Key–Words: Atrial premature complex, UFIR smoothing, electrocardiogram, Rice distribution, confidence interval.

1 Introduction

Electrocardiogram (ECG) signals have been under profound investigation in medical sector for decades and many heart diseases were identified based on ECG. In past decades, methods of statistical signal processing (SSP) were involved to provide accurate decisions. Estimates provided by SSP algorithms strongly depend on confidence intervals (CIs) specified for ECG features [1]. However, measured ECG features are highly variable, because the ECG signals vary depending of the human state and environmental conditions. That imposes difficulties in the recognition heart diseases, among which the arrhythmia is associated with abnormality of the cardiac rhythm [2]. Different types of arrhythmias can be studied analyzing ECG features using SSP. But, in view of measurement noise, optimal algorithms are often required.

The first segment of the ECG signal called ‘depolarization’ corresponds to the P wave (atrial depolarization). From this feature, the atrial premature complex (APC) can be analysed via changes in the P wave morphology. However, the ECG signal features extraction is a complex process in view of noise and some measurement uncertainties [3–8]. Methods developed in the time and frequency domains include linear predictors, Fourier and wavelet transform-based analysis, etc. [9, 10]. Diverse techniques were also developed to provide noise reduction [11]. In general, denoising algorithms are designed such that the fundamental properties of the ECG signal are not violated. In this regard, optimal smoothing is most powerful due to the best denoising effect and optimal tracking ability. The most widely used smoother was developed by Savitsky and Golay [12]. It provides smoothing at the middle of the averaging horizon. A more general solution is known is the $p$–shift unbiased finite impulse response (UFIR) filter developed by Shmaliy et al. [13–15]. This approach implies that lag $q = -p > 0$ must be selected for each degree individually and not obligatorily at the middle of the horizon to provide best denoising. Smoothing filtering is organized here with $p < 0$, filtering with $p = 0$, and predictive filtering with $p > 0$. An optimal Savitsky-Golay smoother minimizing the minimum mean square error (MSE) is given in [16, 17].

In this paper, we apply the $p$-shift UFIR smoothing filter to ECG data and investigate the confidence interval for the P-wave using the Rice distribution. This investigation is based on the MIT-BIH Arrhythmia Database available for free use from [18, 19].
1.1 Atrial Premature Complex (APC)

The APC is considered by medical community as a independent predictor of atrial fibrillation, stroke, and death. The frequency of occurrence of APC in the general population is little-known. This abnormality is also related to some other cardiovascular risk factors. The APC is characterized by premature beating originated in the atria and can occur in different situations such as infection, myocardial ischemia, inflammation, usage of tobacco, alcohol, and caffeine, or anxiety and hypokalemia [2]. A fragment of the ECG heartbeats is shown in Fig. (1), in which one recognizes features such as P, QRS complex, and T. As can be seen, the APC is characterized by abnormal P-wave morphology providing a compensatory pause.

2 q-Lag UFIR Smoothing Filtering of Discrete Polynomial ECG Model

Let us define the ECG signal in discrete time index $n$ as $x_n$ and its measurement $s_n$ as

$$s_n = x_n + v_n,$$

where $v_n$ is an additive zero mean noise. Because noise is poorly known in such signals, we will ignore its statistics. Signal $x_n$ can be represented with a degree polynomial on a horizon of $N$ points, from $n - N + 1 - p$ to $n - p$, where $p \leq 0$ is a shift such that $p = 0$ means filtering, $p < 0$ smoothing filtering with lag $q = -p$, and $p > 0$ predictive filtering with step $p$. The $p$-shift UFIR filter can be designed in the convolution-based batch forms as [20]

$$\hat{x}_{n|n-p} = \sum_{i=1}^{N-1+p} h_{li}(N, p)s_{n-i}, \quad (2)$$

where $h_{li}(p) \triangleq h_{li}(N, p)$ is the $(N, p)$-variant $l$-degree impulse response of the UFIR filter. In a compact matrix form, (2) can be represented as

$$\hat{x}_{n|n-p} = W_I^T(p)S_N,$$ 

where $S_N$ is the measurement vector and the UFIR filter gain matrix is given by

$$W_I^T(p) = [h_l(p)h_{l(1+p)}(p)\cdots h_{l(N-1+p)}(p)]. \quad (4)$$

In order for the estimate (2) to be unbiased, the following unbiasedness condition must be satisfied,

$$E\{\hat{x}_{n|n-p}\} = E\{x_n\}, \quad (5)$$

and the coefficients for gain (4) obeying (5) can be found as shown below.

2.1 $p$-Shift Gain for UFIR Smoothing Filter

The gain matrix $W_I^T(p)$ for the UFIR filter has the following fundamental properties:

$$\sum_{i=p}^{N-1+p} h_{li}(N, p) = 1, \quad (6)$$

$$\sum_{i=p}^{N-1+p} h_{li}(N, p)i^u = 0, \quad 1 \leq u \leq l. \quad (7)$$

representing the unit area (6) and zero moments (7). In matrix forms, these properties can be represented by the unbiasedness constraint

$$\tilde{W}_l^T(p)V(p) = J^T, \quad (8)$$

where

$$J = [1 \ 0 \ \cdots \ 0]^T \quad (9)$$

and $V(p)$ is the $(l+1) \times (l+1)$ Vandermonde matrix,

$$V(p) = \begin{bmatrix} 1 & p & p^2 \\ 1 & 1+p & (1+p)^2 \\ 1 & 2+p & (2+p)^2 \\ \vdots & \vdots & \vdots \\ 1 & N-1+p & (N-1+p)^2 \end{bmatrix}. \quad (10)$$

By operating the right-side of (8) with an identity $[V^T(p)V(p)]^{-1}V^T(p)V(p)$ and discarding the nonzero $V(p)$ from both sides, the $p$-shift UFIR filter gain becomes

$$W_I^T(p) = J^T[V^T(p)V(p)]^{-1}V^T(p). \quad (11)$$

The UFIR filter gain can be represented with the $l$-degree polynomial as [14,20]

$$h_{li}(p) = \sum_{j=0}^{l} a_{lj}(p)n^j, \quad (12)$$
where coefficient $a_{ij}$ can be defined by solving an equation
\begin{equation}
J = D(p)B(p),
\end{equation}
where
\begin{equation}
B = [a_0, a_1, \cdots, a_l]^T
\end{equation}
and the $(l + 1) \times (l + 1)$ matrix is
\begin{equation}
D(p) = V^T(p)V(p) = \begin{bmatrix}
d_0 & d_1 & \cdots & d_l \\
d_1 & d_2 & \cdots & d_{l+1} \\
\vdots & \vdots & \ddots & \vdots \\
d_l & d_{l+1} & \cdots & d_{2l}
\end{bmatrix},
\end{equation}

where $d_m(p) = \sum_{i=p}^{N-1} i^m$, $m \in [0, 2l]$, is a generic component. Most frequently, filters of low-degree $l = [1, 4]$ are used with the $p$-shift polynomial gain existing as [11]
\begin{equation}
h_{li}(N, p) = \begin{cases}
\text{nontrivial}, & p \leq i \leq N - 1 + p \\
0, & \text{otherwise}
\end{cases}
\end{equation}

\subsection{2.2 Low-Degree Polynomial gains}

For linear modeling of ECG signal on $[p, N - 1 + p]$, the ramp function $h_{1i}(p)$ is used [14],
\begin{equation}
h_{1i}(p) = a_{01}(p) + a_{11}(p)n,
\end{equation}
with the coefficients
\begin{align}
a_{01}(p) &= \frac{2(2N - 1)(N - 1) + 12p(N - 1 + p)}{N(N^2 - 1)}, \\
a_{11}(p) &= \frac{6(N - 1 + 2p)}{N(N^2 - 1)}.
\end{align}

This gain provides the best noise reduction inherent to simple averaging with lag $q = -p = N - \frac{1}{2}$.

For quadratic and cubic signal models, the polynomial gains are given by
\begin{align}
h_{2i}(p) &= a_{02}(p) + a_{12}(p)i + a_{22}(p)i^2, \\
h_{3i}(p) &= a_{03}(p) + a_{13}(p)i + a_{23}(p)i^2 + a_{33}(p)i^3,
\end{align}

with the coefficients given in [14, 20].

\subsection{2.3 Optimized Adaptive Denoising of ECG Signals}

An optimal horizon $N_{opt}$ is required by UFIR filtering to minimize the MSE. Because the ECG model is not available, we determine $N_{opt}$ via the measurement residual $s_n - \hat{x}_{n|n-p}(N)$ and the mean square value (MSV)
\begin{equation}
V(N) = E\{[s_n - \hat{x}_{n|n-p}(N)]^2\}
\end{equation}
by the relation [23]
\begin{equation}
N_{opt} = \arg\min_N \frac{\partial V(N)}{\partial N} + 1.
\end{equation}

The optimal horizon $N_{opt}$ serves for slow parts of the ECG beat shown in Fig. (2), by excluding fast excursions associated with the QSR complex. The MSV $V_n$ behaves similarly for different degrees $l$ and $N_{opt}$ has closely related values for $l = 1, 2, 3$. Because low degrees provide better denoising [23], we select $l = 2$ associated with $N_{opt} = 21$ (Table 1).

\begin{table}[h]
\centering
\begin{tabular}{c|cccc}
\hline
\text{l} & 1 & 2 & 3 & 4 \\
\hline
$N_{opt}$ & 19 & 21 & 20 & 27 \\
MSE\times10^{-4} & 1.7 & 1.29 & 1.14 & 0.78 \\
\hline
\end{tabular}
\caption{Effect of the UFIR degree $l$ on $N_{opt}$ and MSE}
\end{table}

\section{3 Comparison of Smoothing and Others Techniques}

The linear prediction method proposed in [24] and developed for ECG signals by Martis [9] was regarded
as one of the standard approaches to analyse features of digitized ECG signals. A heartbeat \( x_n \) is modeled here as a linear combination of its past input samples \( x_{n-k}, k = 1, 2, \ldots, p \), where \( p \) is considered as the order of prediction and \( a_k \) denotes the \( k \)th linear prediction coefficient. In the prediction technique, the estimation error is defined as \( e_n = x_n - \bar{x}_n \), where the predictive estimate given by \( \bar{x}_n \) is computed as \( \bar{x}(n) = \sum_{k=1}^{p} a_k x_{n-k} \). According to Martin, \( e_n \) is a portion of the ECG signal, which cannot be predicted by linear models. To provide UFIR predictive filtering, we configure the filter degree with \( p = 2 \) as stated in [25]. The UFIR filter predicts estimates with \( p > 0 \) and both the prediction method and the UFIR predictive filtering employ the discrete linear prediction of undergoing process via its noisy measurement. Due to the presence of noise, a part of the ECG signal taken from \([120:200]\) is difficult to predict UFIR filtering, filtering, and smoothing filter - different methods is required. Therefore, a comparative analysis of different methods is required.

In Fig. 3(a), we sketch typical errors produced by the predictive filtering, filtering, and smoothing filtering. A part of the ECG signal taken from [120:200] is zoomed in Fig. 3(b). As can be seen, all filters are successful in denoising the P-wave. However, the UFIR smoothing filter does it more precisely, while the prediction method and the UFIR predictive filtering employ the discrete linear prediction of undergoing process via its noisy measurement. Even so, there are some zones in the ECG picture where the linear prediction is not successful in extracting the ECG features. Therefore, a comparative analysis of different methods is required.

In this section, we develop another one based on the UFIR smoothing filtering algorithm.

The QRS complex (Fig.2) is the first to detect. To this end, the peak value R (ECG signal maximum) is estimated as \( \hat{R} \) and a window is introduced with two points, \( Q' \) and \( S' \). The estimate \( \hat{Q} \) of \( Q \) is obtained by the least value between \( Q' \) and \( \hat{R} \). In turn, the estimate \( \hat{S} \) of \( S \) is provided by the least value between \( \hat{R} \) and \( S' \). Provided \( \hat{Q}, \hat{R} \) and \( \hat{S} \), the QRS complex is suppressed to save only the P and T waves. Then the estimates \( \hat{P} \) of P and \( \hat{T} \) of T are found similarly by suppressing one of the waves.

Provided \( \hat{P} \), the P wave is split into two segments, \( P_1 \) and \( P_2 \), where \( P_1 \) is extended from the initial point to \( P_2 \). In segment \( P_1 \), we apply the derivative. Next, we consider a small portion of the resulting signal and find a global maximum. This point is next considered as the beginning of P wave and called \( P_{onset} \). In segment \( P_2 \), we also apply the derivative, consider a small portion of the resulting signal, and find a global minimum. This minimum, which corresponds to the end of P wave, is called \( P_{offset} \). Values \( P_{onset} \) and \( P_{offset} \) are located at points \( P_{on} \) and \( P_{off} \), respectively. Then the duration of P wave is computed as \( P_{dur} = P_{off} - P_{on} \). A distance between \( \hat{P} \) and the baseline is calculated and called the wave amplitude.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Average Error</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smooth-UFIR</td>
<td>( 1.7997 \times 10^{-5} )</td>
<td>( 4.2073 \times 10^{-5} )</td>
</tr>
<tr>
<td>Linear predictor</td>
<td>-0.0074</td>
<td>2.1462 \times 10^{-4}</td>
</tr>
<tr>
<td>Predict UFIR</td>
<td>-2.4691 \times 10^{-4}</td>
<td>2.2799 \times 10^{-4}</td>
</tr>
</tbody>
</table>

**4 P wave Detection**

The known technique for ECG signal features extraction is based on the Pan Tompkins algorithm [26]. In this section, we develop another one based on the UFIR smoothing filtering algorithm.

The QRS complex (Fig.2) is the first to detect. To this end, the peak value R (ECG signal maximum) is estimated as \( \hat{R} \) and a window is introduced with two points, \( Q' \) and \( S' \). The estimate \( \hat{Q} \) of \( Q \) is obtained by the least value between \( Q' \) and \( \hat{R} \). In turn, the estimate \( \hat{S} \) of \( S \) is provided by the least value between \( \hat{R} \) and \( S' \). Provided \( \hat{Q}, \hat{R} \) and \( \hat{S} \), the QRS complex is suppressed to save only the P and T waves. Then the estimates \( \hat{P} \) of P and \( \hat{T} \) of T are found similarly by suppressing one of the waves.

Provided \( \hat{P} \), the P wave is split into two segments, \( P_1 \) and \( P_2 \), where \( P_1 \) is extended from the initial point to \( P_2 \). In segment \( P_1 \), we apply the derivative. Next, we consider a small portion of the resulting signal and find a global maximum. This point is next considered as the beginning of P wave and called \( P_{onset} \). In segment \( P_2 \), we also apply the derivative, consider a small portion of the resulting signal, and find a global minimum. This minimum, which corresponds to the end of P wave, is called \( P_{offset} \). Values \( P_{onset} \) and \( P_{offset} \) are located at points \( P_{on} \) and \( P_{off} \), respectively. Then the duration of P wave is computed as \( P_{dur} = P_{off} - P_{on} \). A distance between \( \hat{P} \) and the baseline is calculated and called the wave amplitude.
represents the reference line; $f_s$ is the data sample frequency; and $Interval$ is a value, which determines the window width to cover Q and S points. The algorithm output consists of estimates of the ECG signal features such as $\hat{P}$ of P, $P_{\text{amp}}$ of the P amplitude, $P_{\text{dur}}$ of the P duration. All these features are extracted from the smoothed signal.

Algorithm 1: A pseudo code of the algorithm to extract morphological features of P-wave

**Data:** $ss_i$, $N_b$, $Baseline$, $f_s$, $Interval$

**Result:** $\hat{P}$, $P_{\text{amp}}$, $P_{\text{dur}}$.

1. **Begin**:
2. for $i = 1$ to $N_b$ do
3. $ss_i = \text{beatss}(i)$
4. $[\hat{R}, \hat{R}_p] = \text{max}(ss_i)$
5. $[Q', S'] = \text{IntervalQRS}(ss_i, Interval)$
6. $[\hat{Q}, \hat{Q}_p] = \text{min}(ss_i(Q': R_p))$
7. $[\hat{S}, \hat{S}_p] = \text{min}(ss_i(R_p: S'))$
8. $ss_{\text{new}} = \text{suppress}(ss_i(Q': S'))$
9. $[\hat{P}, \hat{P}_p] = \text{max}(ss_{\text{new}}(1: Q'))$
10. $\hat{P}_1 = ss_{\text{ss}}(1 : P_p)$
11. $\hat{P}_2 = ss_{\text{ss}}(P_p : Q')$
12. $[P_{\text{onset}}, P_{\text{off}}] = \text{max}(\text{diff}(P_1))$
13. $[P_{\text{onset}}, P_{\text{off}}] = \text{min}(\text{diff}(P_2))$
14. $Baseline(i) = \text{length}(ss_i) = P_{\text{offset}}(i)$
15. $P_{\text{amp}}(i) = \hat{P} - Baseline(i)$
16. $P_{\text{dur}}(i) = P_{\text{off}} - P_{\text{on}} / fs$
17. **end for**

The algorithm begins with computing $\hat{R}$ as the ECG signal maximum, using function $\text{max}$. Function $\text{IntervalQRS}$ is applied to compute $Q'$ and $S'$. The $Interval$ variable determines the window width to cover the QRS complex and obtain $\hat{Q}$ and $\hat{S}$ as two minima between $Q'$ and $S'$. Function $\text{min}$ is used to find the above-mentioned points. The $\text{suppress}$ function is used to suppress the QRS complex. Function $\text{max}$ is used to estimate $P$. Function $\text{diff}$ is introduced to compute the derivatives in the $P_1, P_2$ intervals. Functions $\text{max}$ and $\text{min}$ with $\text{diff}$ are used to find $P_{\text{onset}}, P_{\text{off}}$, $P_{\text{on}}$, and $P_{\text{off}}$. Provided these values, the duration is estimated of P features. Function $\text{length}$ is introduced to compute the signal length. The $Baseline$ variable determines the reference line for computing the amplitude features. This variable is equal to $P_{\text{offset}}$. As can see in the Fig. 4, the estimates provided by the UFIR smoothing filter are more consistent to the average P-wave than by other techniques.

Fig. 4: Applications of different filters to extract the ECG amplitude features: (a) amplitude of P-wave and (b) duration of P-wave. Here "Smooth-UFIR" is the UFIR smoothing filter, "Predict UFIR" is the linear predictor and "Predict -UFIR" is the UFIR predictive filter.

## 5 Validation of P-Wave Feature Estimates

Because the positive-valued P-values vary for normal and abnormal heartbeats in a wide range, it is required to specify the confidence interval for the P-wave estimate to be valid with a given probability. We do it using the Rice probability density function (pdf), which corresponds to the envelop of a harmonic signal in Gaussian noise [27, 28],

$$p(r) = \frac{r}{\sigma^2} \exp \left( -\frac{r^2 + A}{2\sigma^2} \right) I_0 \left( \frac{A r}{\sigma^2} \right), \quad (24)$$

where $\sigma^2$ is the variance of acting noise, $A$ is the harmonic signal amplitude, and $I_0(x)$ is the modified Bessel function of the first kind and zeroth order. The normalized Rice pdf is given by

$$p(v) = v \exp \left( -\frac{v^2 + a^2}{2} \right) I_0(Av), \quad (25)$$

where $v = \frac{r}{\sigma}$ and $a = \frac{A}{\sigma}$. Note that (25) reduces to the Rayleigh distribution [29] when $a = 0$ and, by large $a$, it becomes Gaussian.

### 5.1 Confidence Interval for P-Wave

To specify the confidence interval for the P-wave estimates provided by Algorithm 4.1, we have investigated the P-wave histograms for normal and abnormal heartbeats taken from different persons as shown in Fig. 5. We employed a register of different ages between female and male gender with
Fig. 5: Normalized histograms and Rice-based approximations for P-wave features of healthy heartbeats.

9159 healthy and 2540 APC beats. The name of the signal is referenced based on the Arrhythmia database. In this case the records used are [101, 105, 220, 231, 100, 116, 202, 117, 223]. Other registers were also processed, but we select only those records, which are most close to the reference signal. The number of bins varies for each histogram between 500 and 1000. Also, we consider more representative features, because the ECG record has a considerable variability. Values $v$ and $a$ for the Rice pdf were chosen in the minimum MSE sense.

5.1.1 Confidence Interval for Normal Heartbeats

As can seen in the Fig. 6, the confidence interval (CI) for the APC heartbeats has two boundaries, 0.009 and 0.125, with the probability of 76.33%. For healthy beats, the boundaries were found to be 0.1403 and 0.3074.

6 Conclusions

In this paper, we have presented the UFIR smoothing filter as an estimator of the ECG signal features and shown that it has better performance than the known predictive and filtering solutions. We also specified the confidence interval for the ECG heartbeats using the Rice-based approximation of the P-wave distribution. Because the confidence intervals specified for expert’s probabilities play a key role in automatic diagnosis of heart diseases, we consider this topic as a next important part of our work.

References:


Fig. 6: Confidence intervals for APC (sold) and healthy (dashed) ECG heartbeats.


