Salamanca, Gto., a 2 de Noviembre del 2020.

M. en I. HERIBERTO GUTIÉRREZ MARTIN
JEFE DE LA UNIDAD DE ADMINISTRACIÓN ESCOLAR
P R E S E N T E.-

Por medio de la presente, se otorga autorización para proceder a los trámites de impresión, empastado de tesis y titulación al alumno(a) M. en I. Carlos Mauricio Lastre Domínguez del ***Programa de Doctorado en*** Ingeniería Eléctrica y cuyo número de ***NUA*** es: 145795 del cual soy director. El título de la tesis es: Denoising and Features Extration of ECG Signals using Unbiased FIR Estimation Techniques

Hago constar que he revisado dicho trabajo y he tenido comunicación con los sinodales asignados para la revisión de la tesis, por lo que no hay impedimento alguno para fijar la fecha de examen de titulación.

A T E N T A M E N T E

Prof. Yuriy S Shmaliy
NOMBRE Y FIRMA
DIRECTOR DE TESIS
SECRETARIO

Dr. Oscar G Ibarra-Manzano
NOMBRE Y FIRMA
DIRECTOR DE TESIS

Dr. Arturo García-Pérez
NOMBRE Y FIRMA
PRESIDENTE

Dr. Gilberto Múñoz Moreno
NOMBRE Y FIRMA
VOCAL

Dr. Miguel Octavio Arias-Estrada
NOMBRE Y FIRMA
VOCAL

Dr. J. Amparo Andrade-Lucio
NOMBRE Y FIRMA
VOCAL

UNIVERSIDAD DE GUANAJUATO

CAMPUS IRAPUATO SALAMANCA

DIVISIÓN DE INGENIERÍAS

Denoising and Features Extraction of ECG Signals using Unbiased FIR Estimation Techniques

TESIS DOCTORAL

QUE PARA OBTENER EL GRADO:

DOCTOR EN INGENIERÍA ELÉCTRICA

PRESENTA:

M.I. CARLOS MAURICIO LASTRE DOMÍNGUEZ

ASESORES:

Ph.D. YURIY S. SHMALIY

Ph.D. OSCAR G. IBARRA MANZANO

SALAMANCA, GTO. MAYO 2020

Acknowledgments.

Primordially, thank God. I would like to express my special thanks of gratitude to my advisers Yuriy S. Shmaliy and Oscar Ibarra Manzano who gave me the golden opportunity to do this wonderful project on the topic ECG signals processing, which also helped me in doing a lot of research and I came to know about so many new things I am really thankful to them. I thank Universidad de Guanajuato and CONACYT for the support. I would also like to thank my parents, brothers, my partner Guillermina, my son Mauricio, that they helped me a lot emotionally in finalizing this work.

To my son, Mauricio Itzae...

Satisfaction lies in the effort, not in the attainment. Full effort is full victory.

Mahatma Gandhi

Abstract

The electrocardiogram (ECG) signals bear fundamental information for deciding about heart diseases. So the scientific community has been performing many efforts during decades to extract features of heartbeats via ECG records with high accuracy and efficiency using different strategies and methods. However, the noise and artifacts provided by external factors avoid significant patterns associated with the ECG signals. These patterns play an important role to find specific abnormalities in ECG signals. Hence, techniques based on unbiased FIR (UFIR) filtering promises better results. In this dissertation, we have applied a model based on UFIR to ECG signals. Hence, we compare the proposed technique with traditional method such as predictors, standard filters (e.g. low-pass filter) wavelet filters, Savitsky-Golay filter. The UFIR method outperforms other studied techniques for ECG signals.

Contents

Title Page i

Acknowledgments ii

Abstract v

Contents ix

List of Figures xiv

List of Tables xv

Acronyms xv

1 Introduction **1**

1.1 Background 1

1.2 Motivation 2

1.3 Hypothesis 4

1.4 Objectives 4

1.4.1 General objective 4

1.4.2 Specific objectives 4

1.5 Scope 5

1.6 Highlights 5

2 Foundations **7**

2.1 ECG signals foundations 7

2.1.1 Database 9

2.1.2 Analysed pathologies 10

2.2 Estimation theory 14

2.2.1 Estimator evaluation 15

2.2.2 Unbiased estimation 17

2.2.3 Criterion of minimum variance 18

2.3 Previous work for ECG signals . . . 19
2.3.1 Time based techniques . . . 19
2.3.2 Frequency based techniques . . . 20
2.3.3 Time-frequency based techniques . . . 21
2.3.4 Empirical mode decomposition . . . 21

3 Denoising and features extraction for ECG signals using unbiased finite impulse response (UFIR) smoothing 22

3.1 p-shift UFIR smoothing filtering . . . 22
3.1.1 ECG Signal denoising on adaptive horizons . . . 24
3.1.2 General UFIR smoothing algorithm . . . 25
3.1.3 Computing N_{opt} for ECG data . . . 26
3.1.4 Denoising algorithm for ECG signals . . . 28
3.1.5 UFIR-based algorithm for features extraction . . . 29
3.1.6 Algorithm design for features extraction of ECG signals . . . 31
3.2 Testing of algorithms for estimating N_{opt} and denoising algorithm . . . 33
3.2.1 Critical evaluation of denoising algorithms . . . 34
3.2.2 Effect of SNR on the estimator MSE . . . 35
3.3 Applications to ECG signals . . . 36
3.3.1 APC heartbeats UFIR smoothing and P-wave features Analysis using Rice Distribution . . . 37
3.3.2 P-Wave detection algorithm . . . 40
3.3.3 Validation of P-Wave feature estimates . . . 42
3.3.4 Confidence interval for P-Wave . . . 43
3.4 Discussion . . . 44

4 Denoising and Features Extraction of ECG signals in State Space Using Unbiased FIR Smoothing 47

4.1 ECG signal model in discrete-time state-space . . . 49
4.1.1 UFIR filtering and smoothing of ECG signals . . . 49
4.1.2 Batch UFIR filter and smoother . . . 50
4.1.3 Adapted optimal horizon N_{apt} . . . 51
4.1.4 Iterative UFIR smoothing . . . 51

4.2 ECG signal features extraction in state space . . . 53
4.2.1 Fiducial points detection and features extraction . . . 54
4.3 Tuning and Testing . . . 55
4.3.1 Optimal lag for UFIR smoother . . . 56
4.3.2 Testing iterative UFIR algorithm . . . 57
4.4 Assessment of adaptive UFIR smoothing algorithm for ECG signals . . . 60
4.4.1 Filtering and artifact removal . . . 61
4.4.2 Computational complexity . . . 63
4.4.3 Features extraction and errors comparison . . . 64
4.4.4 Applications to normal and abnormal ECGs . . . 65
4.4.5 Classification . . . 65
4.5 Discussion . . . 68

5 ECG Signal Denoising and Features Extraction Using Harmonic Unbiased FIR Smoothing 80

5.1 State-Space Representation of ECG Signals using Harmonic Model . . . 80
5.1.1 Unbiased FIR Smoothing . . . 83
5.2 Optimal Horizon for UFIR smoother employing harmonic model . . . 84
5.3 Adaptive UFIR smoothing algorithm for ECG records . . . 86
5.4 Testing UFIR Smoother by Harmonic Model . . . 88
5.4.1 Signal-to-Noise Ratio (SNR) Analysis . . . 92
5.4.2 Percentage-root-mean-square difference (PRD) Analysis . . . 93
5.5 Discussion . . . 93

6 Conclusions and future work 96

6.1 Conclusions . . . 96
6.2 Future work . . . 97
6.2.1 High order harmonic UFIR smoother . . . 97
6.2.2 Segmentation of ECG heartbeats . . . 97
6.2.3 Pattern recognition . . . 97
6.2.4 Biomedical signals . . . 98
6.2.5 Hardware implementation . . . 98

Bibliography 99

Appendix A Low-Degree UFIR Functions $h_{li}(N,p)$ **111**

A.0.1 Ramp, $l = 1$: 111

A.0.2 Quadratic, $l = 2$: 111

Appendix B UFIR Filter for state space **112**

B.1 Extended model in state space 112

List of Figures

2.1 Representation of a Heartbeat. 8

2.2 The heartbeat pulse model represented with features (amplitudes and durations) of the P-wave, QRS complex, T-wave, U-wave, and ST angle. 9

2.3 Typical 'normal' and 'APC' heartbeats from ECG. 12

2.4 Representation of the normal and abnormal T-wave 13

2.5 Dependence de PDF on unknown parameters 14

2.6 Random data sample for A 16

2.7 Histogram for sample mean $\hat{A}$ (a) and first value $\tilde{A}$ (b) 17

3.1 Caption for LOF 25

3.2 Step-by-step events representing the strategy of the ECG signal denoising and features extraction: (a) original ECG signal, (b) smoothed ECG signal, (c) peak-value R, Q, and S, (d) P and T points, (e) P wave, (f) duration of QRS complex, (g) T wave, and (h) ST-angle. 30

3.3 Effect of N on the MSV with a) $l = 1$, b) $l = 2$, c) $l = 3$ and d) $l = 4$: the MSV is circled, $\sqrt{V_N}$ is a cubic approximation of the MSV, and $\frac{\partial}{\partial N}\sqrt{V_N}$ and $\sqrt{\frac{\partial}{\partial N}V_N}$ are the derivatives of $\sqrt{V_N}$. The optimal horizon $N_{\text{opt}} = 19$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N}V_N}$. 33

3.4 Effect of the UFIR filter degree l on the estimation accuracy. 34

3.5 Denoising of ECG signals: (a) heartbeat estimation with different methods such as ASmooth-UFIR (UFIR Adaptive-Smoothing filter), Linear Predict, Predict-UFIR (UFIR predictive filter), and Filter-UFIR (UFIR filter), and (b) segmental visualization of five estimates. 35

3.6 Errors produced by different estimators: (a) in the T-wave, (b) error boxplot in the T-wave, (c) in the QRS-complex, (d) error boxplot in the QRS-complex, (e) in the P-wave, and (f) error boxplot in the P-wave. Estimator 1 is Smooth UFIR, estimator 2 is ASmooth-UFIR, estimator 3 is the predict linear, estimator 4 is the Predict-UFIR and finally, estimator 5 is the Filter-UFIR 36

3.7 RMSEs of UFIR denoising estimators and linear predictor (Linear-Predict) as functions of SNR 37

3.8 Features of the ECG signal extracted using the UFIR smoothing filter (Smooth-UFIR), basic linear predictor (Linear Predict) [1], and UFIR predictive filter (Predict UFIR): (a) P_{amp}, (b) QRS_e, and (c) T_{amp}. 38

3.9 Features of the ECG signal extracted using the UFIR smoothing filter (ASmooth-UFIR), basic linear predictor (Linear Predict) [1], and UFIR predictive filter (Predict Ufir): (a) P_{dur}, (b) QRS_{dur}, (c) T_{dur}, and (d) ST_{angle} $\hat{\theta}$. The durations are sampled with 0.0028 seconds. Also ST-angle is referenced with 45 grades 39

3.10 Boxplot of features of the ECG signal extracted using the UFIR adaptive smoothing filter (Estimator 1: ASmooth-UFIR), UFIR predictive filter (Estimator 2: Predict Ufir), basic linear predictor (Estimator 3: Linear Predict), UFIR filter (Estimator 4: Filter-UFIR): (a) P_{amp}, (b) P_{dur}, (c) QRS_e, and (d) QRS_{dur},(e)T_{amp}, (f) T_{dur} and (g) ST_{angle} $\hat{\theta}$. 40

3.11 Applications of different filters to extract the ECG amplitude features (a) Amplitud 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 42

3.12 Normalized histograms and Rice-based approximations for P-wave features of healthy heartbeats. 43

3.13 Confidence intervals for APC (sold) and healthy (dashed) ECG heartbeats. 44

4.1 Fiducial points of heartbeat 48

4.2 Block-diagram of features extraction of the ECG signal in state space using an UFIR smoother. 53

4.3 Fiducial features of a single heartbeat extracted along with the spacial points in state space using the UFIR approach: (a) first state, (b) second state, and (c) third state. 56

4.4 Denoising of a test sinusoid signal (solid) corrupted by zero mean AWGN using the UFIR smoother (dotted with asterisk) with lag q1 (4.21) and (dash-dotted and marked square) with lag q2 (4.22). The Daubechies wavelet-based smoothers are: db6 (dashed with marked circle), db14 (dash with dot marked), sym4 (solid with marked cross), bior2.2 (dashed with marked plus sing), coif2 (solid with marked diamond). The standard filters are: band pass filter (bandpass, dashed with marked pentagon), low pass filter low-pass (solid with marked hexagon), median (medfilt, dotted with marked point), and notch (notchfil, dashed with marked point). 58

4.5 RMSEs corresponding to Fig. 4.4 and computed over 1000 iterations for the UFIR smoother, wavelet-based, and standard filters such as the low-pass, band-pass, median, and notch. 59

4.6 RMSEs of the UFIR smoother compared to the wavelets-based and standard filters. 60

4.7 Smoothing errors in terms of PRD produced by diverse filters. 61

4.8 ECG signal denoising: (a) heartbeat estimation with the UFIR smother, wavelet-based filters, and standard filters; (b) segmental visualization of ten estimates. . . 62

4.9 Measurement residuals produced by the UFIR smoother (q-lag1 and q-lag2), wavelet-based filters (db6, db14, sym4, bior2.2, and coif2), low-pass filter, median filter, and notch filter: (a) actual residuals and (b) error boxplot of heartbeat. 63

4.10 Estimates of the second state (first time-derivative) provided by the Savitsky-Golay smoother with lag q1 (4.21), UFIR smoother with lag q2, wavelet-based filters (db6, db14, sym4, bior2.2, and coif2), low-pass filter, median filter, and notch filter. . . . 72

4.11 Baseline removal using UFIR smoothing with N=1001 and q-lag2 73

4.12 Extracted features of the P-wave duration $\mathrm{Dur_p}$. Expert annotations (gold standard) [2] are represented with the upper and lower boundaries (solid lines) along with an expected average. 74

4.13 Features extracted using the UFIR smoother with q-lag1 (filter 1) and q-lag2 (filter 2) and other algorithms depicted as db6 (filter 3), db14 (filter 4), sym4 (filter 5), bior2.2 (filter 6), coif2 (filter 7), low-pass (filter 8), median (filter 9), and notch (filter 10): (a) duration $\mathrm{Dur_P}$ of P-wave, (b), duration $\mathrm{Dur_{QRS}}$ of QRS-complex, (c), duration $\mathrm{Dur_T}$ of T-wave, (d), amplitude $\mathrm{Amp_P}$ of P-wave, (e) amplitude $\mathrm{Amp_{QRS}}$ of QRS-complex, and (f) amplitude $\mathrm{Amp_T}$ of T-wave. Features are extracted from record 100 lead II of arrhythmia MIT-BIH database. 75

4.14 Features of the P-wave duration extracted using the wavelet-based filter with db6 from the AF and Normal ECG of MIT-BIH Arrhythmia Database: (a) $\mathrm{Dur_p}$, (b) $\mathrm{Dur_p}$ normalized histogram, (c) $\mathrm{Amp_p}$, and (d) $\mathrm{Amp_p}$ normalized histogram. . . . 76

4.15 Features of the P-wave duration extracted using the UFIR smoother from the AF and Normal ECG of MIT-BIH Arrhythmia Database: (a) $\mathrm{Dur_p}$, (b) $\mathrm{Dur_p}$ normalized histogram, (c) $\mathrm{Amp_p}$, and (d) $\mathrm{Amp_p}$ normalized histogram. 77

4.16 T-wave estimates provided by the unbiased FIR algorithm: a) First state: smoothed signal, b) Second state: first derivative and c) third state: second derivative. 78

4.17 boxplot of different amplitudes T-waves for different record. Negative amplitude means inverted T-waves and Positive negative means normal T-wave. The discriminate line is located in the zero. The record one (circled) , two, three and four are the 100, 101, 103 and 112 records of MIT-Arrhythmia data base. 79

5.1 An example of the centralized ECG record taken from MIT-BIH Arrhythmia Database. The record 100/MLII features are depicted as P, QRS complex and T waves. Applied a correction of baseline, the ECG measurement average is near to zero 81

5.2 Single sided Magnitude spectrum: fundamental frequency f_d and first harmonics . 82

5.3 Effect of N on the MSV (circled) for $m = 1$. A cubic approximation of the MSV is $\sqrt{V_N}$ and the optimal horizon $N_{\mathrm{opt}} = 14$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N} V_N}$. 85

5.4 Effect of N on the MSV (circled) for $m = 2$. A cubic approximation of the MSV is $\sqrt{V_N}$ and the optimal horizon $N_{\mathrm{opt}} = 14$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N} V_N}$. 86

5.5 Effect of N on the MSV (circled) for $m = 3$. A cubic approximation of the MSV is $\sqrt{V_N}$ and the optimal horizon $N_{\mathrm{opt}} = 14$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N} V_N}$. 87

5.6 Effect of N on the MSV (circled) for $m = 5$. A cubic approximation of the MSV is $\sqrt{V_N}$ and the optimal horizon $N_{\mathrm{opt}} = 14$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N} V_N}$. 88

5.7 UFIR smoothing of an ECG signal using the 1st harmonic (dotted), 3rd harmonic (dashed), and 5th harmonic (solid-dotted). The ECG data are depicted with a solid line. 89

5.8 Estimation of ECG signals with single ventricular premature complex (VPC) . . . 90

5.9 Denoising of a test harmonic signal (solid) corrupted by the AWGN using a harmonic filter (double dash-triple dots), UFIR smoother with q-lag 1 (dotted) [3] and with q-lag 2 (one dash-dotted) [4]. 91

5.10 RMSEs corresponding to Fig. 5.9 as computed over 1000 iterations for the harmonic filter, UFIR q-lag 1 smoother, and UFIR q-lag 2 smoother. 92
5.11 RMSEs of the UFIR smoother compared to the polynomial and harmonic model and standard filters. 93
5.12 PRD of the UFIR smoother compared to the polynomial and harmonic model. . . 94

List of Tables

2.1 11

2.2 Previous work for ECG signals 20

4.1 Computation Time Required by Diverse Algorithms 64

4.2 Performance of the AF for Normal ECG Heartbeats Based on Different Classifiers 66

4.3 Performance of the AF and Normal ECG Signals Applying PCA and Based on Different Classifiers: FG SVM is the Fine Gaussian SVM, CG SVM is the Coarse Gaussian SVM 67

4.4 Comparative Study of AF detection using Different Approaches and the UFIR Smoother (UFIRS) 67

4.5 Comparison of the ECG Signal Features Extracted Using Different Methods 69

Chapter 1

Introduction

1.1 Background

The electrocardiography (ECG) signals play a key role in diagnosing diverse kinds of heart diseases. World health organization (WHO) shows that cardiovascular diseases (CVDs) are the amount one cause of death globally: more people die annually from CVDs than from any other cause [5]. Because pulses produced by heart may have subtle differences from each other and noise affects the decision accuracy, the ECG is performed using precise electronic equipment [6]. Accurate measurements are required when data are used to extract features of ECG signals and make decisions about different heart diseases using special software. However, even precise measurements are contaminated by artefacts and noise. Artefacts may result from a variety of internal and external causes, such as the Parkinsonian muscle tremors drying electrode gel. Different noises may contaminate the ECG signal during its acquisition and transmission, such as the high frequency noise (electromyogram noise, additive white Gaussian noise, and power line interference) and low frequency noise (baseline wandering). Because noise may lead to wrong interpretation, ECG signal denoising is required. Therefore, meaningful attention has been paid during the last decades to develop mathematical methods and computation algorithms to extract the ECG features from regular (noisy) data with an accuracy sufficient for medical needs [7–17]. The Fourier transform-based approach has been developed in [18] to extract ECG signal features in the frequency domain. But, this method omits the time resolution, which affects the estimation accuracy. This issue has been circumvented in some other works by providing the time-frequency analysis without affecting the resolution. In [19–22], the wavelet transform-based algorithms were developed to find applications in some medical areas. In the wavelet domain, a compromise be-

tween the frequency and time resolutions is more easily achievable and one can select a proper wavelet to provide a reasonable accuracy. However, a choice of an optimal wavelet is still challenging [23] and the approach has low efficiency in smoothing ECG signals. Other algorithms tested for such needs include the principal component analysis (PCA) [24], linear discriminant analysis (LDA) [25], independent component analysis (ICA) [26], support vector machine [27], and neural networks [28].

One of the recognized approaches proposed in [29] provides noise reduction and features extraction from ECG data by employing linear prediction based on the theory developed in [1]. The approach suggests that principal features of ECG signals can be saved and gained using a one-step linear predictor. Accordingly, features extraction in the QRS complex (region of fast ECG excursions) is provided from an analysis of residual errors between the data and estimates. The approach has manifested itself as useful in the detection of arrhythmias. In other works employing one-step prediction [30, 31], automatic classification of the ECG cardiac abnormalities is provided using Gaussian mixtures. Later, the prediction-based approach has been recognized as one of the standard techniques suitable for ECG signals [32]. It has to be remarked now that, from the standpoint of optimal filtering, *prediction* is less accurate in noise reduction than *filtering* and much less accurate than *smoothing*. On the other hand, the ECG signal processing problems do not imply predicting future values and smoothing with some time-lag may be a better choice for cardiac analysis. A classical example is the Savitzky-Golay filter (smoother) [33], which has found wide applications in diverse areas [34–39].

1.2 Motivation

It is known that the electrocardiogram (ECG) signals bear essential information about different kinds of heart diseases. Therefore, different strategies have been developed during decades to investigate ECG signals and extract critical features with highest accuracy and efficiency [9, 14, 16, 40, 41]. Several algorithms have analysed the patterns extraction in ECG signals. Among these patterns, we can find fiducial features, rhythm variabilities, noise detection based on agglomerative clustering of morphological features and information about the atrium behaviour. Morphological characteristics associated to ECG signals are typically learned through the P, QRS, and T waves, using appropriate methods of ECG signal denoising and features extraction [17, 41–51]. Even so, it is still challenging to reach accurate results due to errors caused by data noise and artifacts induced by data acquisition equipment. In addition, methods developed for denoising

and features extraction based on the Fourier transform assume that ECG signals are stationary and ignore time resolution. A better trade off between the frequency and time is guaranteed by the wavelet transform-based algorithms, provided that a proper wavelet is chosen [42, 52–65]. Other methods can also be applied to ECG signals, such as the empirical mode decomposition (EMD) and Hadamard transform [66, 67]. To increase the accuracy, several authors combined the above methods with approaches such as the principal component analysis (PCA) [24, 26], support vector machine (SVM) [27], and neural networks or deep learning techniques [28, 65].

An optimal approach to provide smoothing and state estimation in linear models has been proposed in [68] to minimize the mean square error (MSE). A solution was found on a horizon $[m-p, n-p]$ of N data points, where n corresponds to a fixed discrete point of the ECG signal, $m = n - N + 1$, and p is a discrete shift. The derived optimal FIR (OFIR) filter becomes smoothing with lag $q = -p$ by $p < 0$, provides filtering with $p = 0$, and becomes p-step predictive when $p > 0$. However, the p-shift OFIR filter requires information about noise, which is not completely available for ECG signals.

A special case of the p-shift OFIR filter is the p-shift unbiased FIR (UFIR) filter [68–71], which completely ignores zero mean noise and is thus more suitable for ECG signals. As being more general, the p-shift UFIR filter generalizes the Savitzky-Golay filter by $p = -(N-1)/2$ and linear predictor with $p > 0$. Although such a filter does not require noise statistics except for the zero mean assumptions, it provides considerable optimal estimates when N is optima (N_{opt}). To calculate N_{opt}, it is essential to consider the minimizing of the MSE (mean square error) [68]. It is clear that a disadvantage of the batch p-shift UFIR filter [69] works in slow operation, which causes a computational burden and complexity in denoising and features extraction [72]. A disadvantage of the batch p-shift UFIR filter [69] resides in slow operation, which causes a computational burden and complexity in denoising and features extraction [72].

Moreover, the UFIR [69] and Savitsky-Golay [33–38] smoothers were designed to de-noise signal with no extra information about the ECG signal state required to facilitate features extraction. However, Let us also notice that the Savitsky-Golay filter was recently modified to be optimal in the minimum mean square error (MSE) sense [34, 35]. The modification is akin to the optimal UFIR filter [73], which produces a maximum likelihood estimate [74]. Because both these solutions require information about noise, which is not well studied in ECG signals, the use of the UFIR smoother becomes more preferable.

A more efficient state space iterative p-shift UFIR algorithm using recursions was designed by

Shmaliy in [72] and then developed and applied with different purposes in many papers [68, 69]. Although the iterative UFIR smoother [72] provides much more information than the batch UFIR [69] and Savitsky-Golay [33] structures, its development for features extraction in ECG signals still has not been addressed in the literature this motivates our work. In this case, two models for ECG signals have been considered: polynomial and harmonic.

1.3 Hypothesis

The UFIR filtering techniques works adequately in low frequency signals by unknowing its statistical parameters. Hence, if we apply UFIR filtering with horizon N_{opt} in ECG signals, we could find stronger estimation and extraction of temporal features. For validating this hypothesis, we establish the objectives following section.

1.4 Objectives

1.4.1 General objective

- Design a denoising and features extraction system of ECG signals using unbiased FIR (UFIR) estimation techniques.

1.4.2 Specific objectives

- Develop an adaptive-horizon UFIR smoothing filtering for denoising ECG signals and extracting features
- Employ an iterative UFIR smoother in state space to increase accuracy of extracting features and fiducial points detection in ECG signals
- Design an optimal q-lag state-space UFIR smoothing algorithm for ECG signals denoising and artifacts removal
- Implement a UFIR filter based on harmonic model to remove noise in ECG signals.
- Compare different unbiased FIR estimation techniques with proposed methods for denoising and features extraction in ECG signals

1.5 Scope

This investigation is divided into three parts:

In the first part, we develop an adaptive-horizon UFIR smoothing filtering algorithm for denoising ECG signals and extracting features. Here, we also investigate the trade-off between the UFIR smoothing filter, UFIR filter, and UFIR predictive filter and compare them to the standard linear predictor suggested in [29]. Focused on the design of efficient algorithms, this first work is limited to data associated with normal heartbeats and postpone an analysis of different kinds of heart diseases to future works.

In the second part, we develop and use an iterative UFIR smoother in state space. The principal aim is to increase accuracy of the features extraction and fiducial points detection. We develop an optimal q-lag state-space UFIR smoothing algorithm for ECG signals denoising and artifacts removal, an algorithm for ECG signal stable features extraction using different classifiers under unknown noise and finally, a high-accuracy patterns classification for ECG signals with atrial fibrillation (AF) and normal conditions. We test the results by different classifiers and compared to those available from several machine learning techniques.

In the third part, we develop a UFIR filter based on harmonic model. Here, we represent an ECG signal with a Fourier series, apply the UFIR smoother to real measurements of ECG signals, and provide a comparative analysis with the polynomial smoothers developed by Lastre *et. al* [4] [3] in terms of the denoising effect. In general, the polynomial and harmonic model have been implemented for designing the UFIR filters in ECG signals. We base this investigations on the MIT-BIH Arrhythmia and PTB Database available for free [75, 76].

1.6 Highlights

- Suboptimal denoising of ECG signals with no requirements to noise, except for the zero mean assumption.
- Unbiased filtering in the QRS region, in which the ECG signal demonstrates rapid excursions.

- An optimal q-lag state-space UFIR smoothing algorithm for ECG signals denoising and artifacts removal.
- An algorithm for ECG signal stable features extraction using different classifiers under unknown noise.
- High-accuracy patterns classification for ECG signals with atrial fibrillation (AF) and normal conditions.
- An optimal horizon N harmonic UFIR smoothing algorithm for ECG signals denoising.

Chapter 2

Foundations

This chapter describes the essential concepts used in this dissertation. Initially, we provide a brief definition of the electrocardiogram signals in the sense of its morphological characteristics. In addition, we detail the used database and some pathologies. We also recognise some concepts of estimation theory as a fundamental part of this work. Finally, although we had explained a brief idea about literature in the first chapter, we profound some techniques definition which compares with the proposed UFIR techniques.

2.1 ECG signals foundations

A heartbeat or the ECG complex contains different waves divided among themselves by distinct intervals [2] (Fig. 2.1). The P-wave represents a depolarization in the right and left atrial, which is provided by sinus node. Normally, the P-wave is positive in most of the leads. In LII (Lead II), the P-wave amplitude is registered to be larger [77]; it does not surpass $2,5\,\mathrm{mV}$ and its duration does not exceed $0.1\,\mathrm{s}$. The QRS complex follows the P-wave and represents the ventricular depolarization. This complex is composed by Q, R and S points (sometimes called waves) and the duration of QRS complex normally ranges from $0.06\,\mathrm{s}$ to $0.10\,\mathrm{s}$, although it varies with heartbeat rate (cardiac frequency) and is smaller in children. The T-wave starts from the isophasic line and can adopt several forms such as tall, pointed, flattened, inverted and biphasic. The T wave length varies considerably. However, habitually it mostly measures $2\,\mathrm{mm}$ and is positive in all of the leads, excepts for aVR that is negative. The nature of the U-wave is still not well understood and it is hard to recognize this wave in most of the leads. What follows from many measurements is that this wave is positive. Hence, The morphology of heartbeat

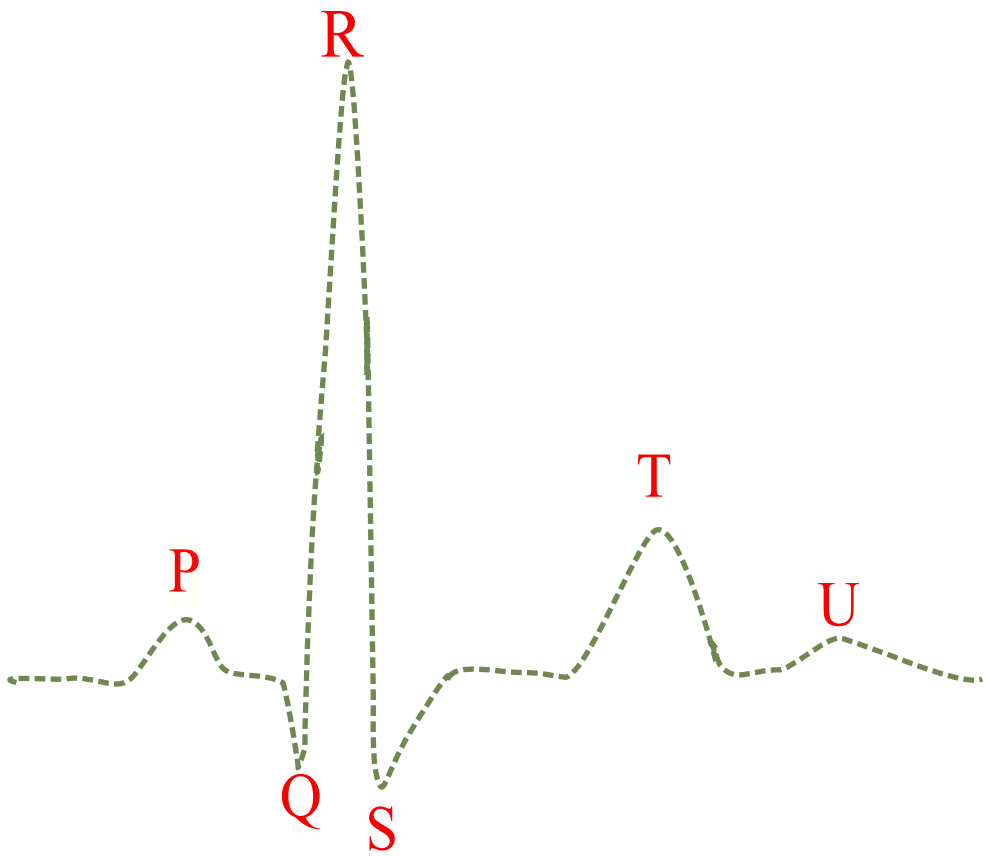

Figure 2.1: Representation of a Heartbeat.

is fundamental for extracting features of ECG signals, which are quasi-periodic as sketched in Fig. 2.2. The heartbeat pulse can be represented with four fundamental features: P-wave (left slow excursion), QRS-complex (central fast excursion), T-wave (first right slow excursion), and U-wave (second right slow excursion).

Several problems arise while processing ECG signals shown in Fig. 3.1

- Measurement data are commonly contaminated by noise, which may not be Gaussian and white.
- Standard features depicted in Fig. 3.1 must be estimated with highest accuracy to avoid medical mistakes.
- The ground truth (reference model) is not available to tune an estimator optimally.

Under such conditions, two approaches relying on accurate identification of heartbeat pulses are commonly considered to extract ECG signal features: *fiducial* and *non-fiducial*. The fiducial

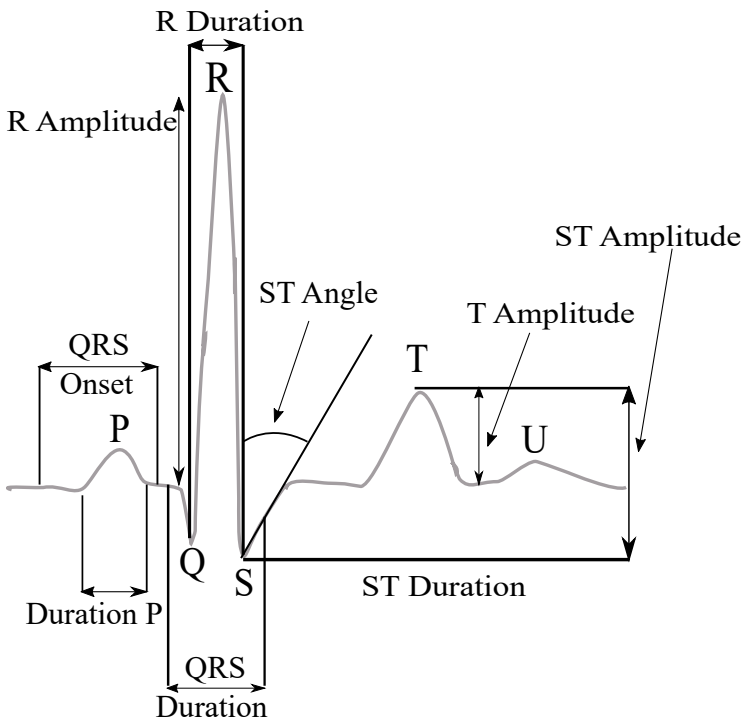

Figure 2.2: The heartbeat pulse model represented with features (amplitudes and durations) of the P-wave, QRS complex, T-wave, U-wave, and ST angle.

approach refers to the characteristics such as amplitude and heart rate, which are related to the duration, amplitude, and wave shape [78–83]. The non-fiducial approach refers to quasi-periodicity of ECG signals [32] and all features are separated into three main categories based on autocorrelation, phase-space, and frequency-domain analysis.

2.1.1 Database

Over years, the medical sector together with engineers has collected information about biomedical signal. These data are fundamental for detecting and prediction of some diseases. Among these data, the electrocardiogram (ECG) signals play a significant role in our society. We base our work on the PhysioNet resources platform also known as *Research Resource for Complex Physiologic Signals* and managed by members of the MIT Laboratory for Computational Physiology. The other core laboratory of the PhysioNet Resource is the Margret and H.A. Rey Institute for Nonlinear Dynamics at Beth Israel Deaconess Medical Center [76]. This platform contains several physiological and clinical records taken from different databases such as the MIT-BIH Arrhythmia (MITDB) and PTB Diagnostic ECG database. PTB means *Physikalisch-Technische Bundesanstalt* also it is known as National Metrology Institute of Germany.

MIT-BIH Arrhythmia

The MITDB describes 48 records with normal and abnormal rhythms taken from 47 subjects. The records are sampled to 360 Hz per lead with 11-bit resolution over a 10 mV range. This database provides the records in two leads, where the most common is the MLII (modified lead II). Other leads are also used, such as V1, V5, etc. A key issue is to choose the lead that most clearly reflects the ECG signal morphology.

PTB diagnostic ECG database

The PTB belong be the PhysioNet platform database [76]. This database contains 549 records from 290 subjects aged between 17 to 87 years old. Each record enclose 15 simultaneously measured signals by 15 leads. The signal is digitized at 1000 samples that corresponds to 10 KHz. The database contains the following diagnosis classes: Myocardial infarction, heart failure, bundle branch block, disrhythmia, myocardial hypertrophy, valvular heart disease, myocarditis, and healthy control (see 2.1).

In this collection, the ECGs were obtained using a non-commercial, PTB prototype recorder with the following features:

- Input channels, (14 for ECGs, 1 for respiration, 1 for line voltage)
- Input voltage: ± 16 mV, compensated offset voltage up to ± 300 mV Input resistance: 100 Ω (DC)
- Resolution: 16 bit with 0.5 μV/LSB (2000 A/D units per mV) Bandwidth: 0 - 1 kHz (synchronous sampling of all channels)
- Noise voltage: max. 10 μV(pp), respectively 3 μV (RMS) with input short circuit
- Online recording of skin resistance Noise level recording during signal collection

2.1.2 Analysed pathologies

Atrial fibrillation

In the ECG, the waves of atrial fibrillation (AF) could be difficult to localize to any repetitive and stable circuit in the atria. Most cases of AF are thought to originate in the area of the pulmonary vein left atrial junctions. The atrial tissue becomes involved in the active maintenance

Table 2.1:

Diasnostic Class	Number of patients
Myocardial infartion	148
Cardiomyopathy/Heart failure	18
Bundle branch block	15
Dysrhythmia	14
Myocardial hypertrophy	7
Valvular heart disease	6
Myocarditis	4
Miscellaneous	4
Healthy controls	52

of the arrhythmia, related to the simultaneous formation of multiple unstable entrant circuits throughout the atria. AF is also one of the most frequently observed arrhythmias in patients with structural heart disease. The predominance of this arrhythmia increase with older people. Common diagnostic include coronary artery disease, hypertensive heart disease, and valvular heart disease. Patients with coronary artery disease may experience AF for the first time during an acute myocardial infarction (MI) or commonly, as a consequence of chronic ischemic myocardial disease, possibly because of associated atrial dilation or fibrosis. In addition, the left atrial enlargement is frequently related to hypertensive pathology.

Atrial premature complex

Atrial premature complexes (APCs) occur when the ectopic stimuli and are heartbeats arising from loci in either the left or right atrium, or inter-atrial septum, but not the sinoatrial (SA) node itself. The atria, therefore, are depolarized from an ectopic site. After an atrial or junctional depolarization, the stimulus may spread normally through the His-Purkinje system into

the ventricles. APCs may occur often or sporadically. Two APCs occurring consecutively are referred to as an atrial couplet.

APCs may occur in people with normal hearts or with practically any type of heart disease. Hence, the presence of APBs does not imply that an individual suffers from cardiac disease. In normal people, these premature complexes can be associated with emotional stress, excessive intake of caffeinated drinks, or the administration of sympathomimetic agents (epinephrine, isoproterenol). APCs may also occur with hyperthyroidism. Even, APC produces palpitations; in this situation, patients may complain of feeling an irregular pulse. This pathology may also be seen with various types of structural heart disease such as atrial fibrillation. The APC is also characterized by premature beating originated in the atria and can occur in different situations such as infection, myocardial ishemia, inflamation, usage of tobaco, alcohol, and caffeine, or anxiety and hypokalemia [84]. A fragment of the ECG heartbeats is shown in Fig. (2.3), in which one recognizes features such as P, QRS complex, and T. As can be seen, the APC represents abnormal P-wave morphology providing a compensatory pause.

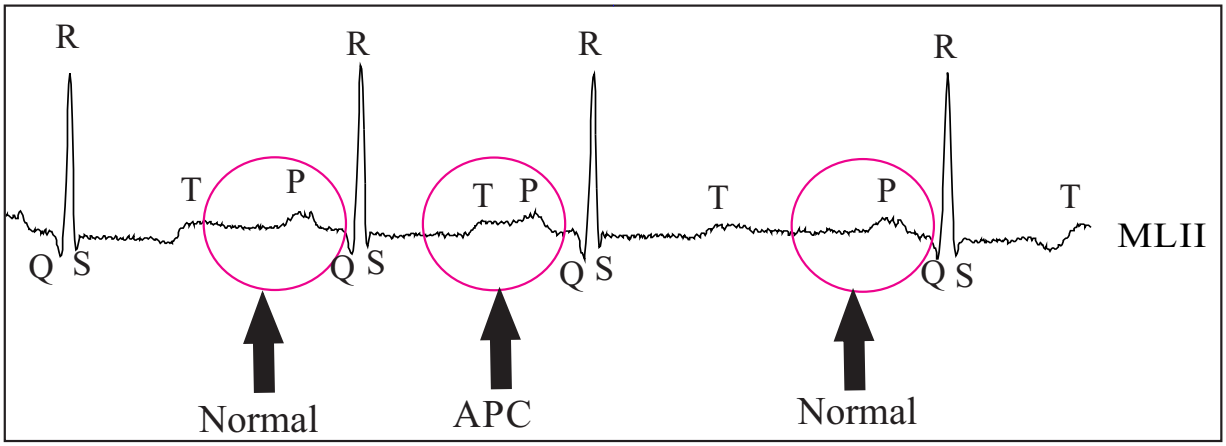

Figure 2.3: Typical 'normal' and 'APC' heartbeats from ECG.

Inverse T-wave

The T wave represents ventricular depolarization. This wave is located after QRS-complex. Ocasionally, the T-wave provides irregular behaviours (see fig 2.4). Because depicts the relative refractory period of ventricular depolarization, a period where the cells of heart are vulnerable to the aditional stimuli. Habitually, the T-wave is positive in normal conditions. May be negative en some leads as LIII and precordial leads V1 and V2. But, never must be negative in LI and

LII [85]. A peaked T-wave can indicate hyperkalemia or myocardial injury. Inverted T waves in some leads may represent myocardial ischemia.

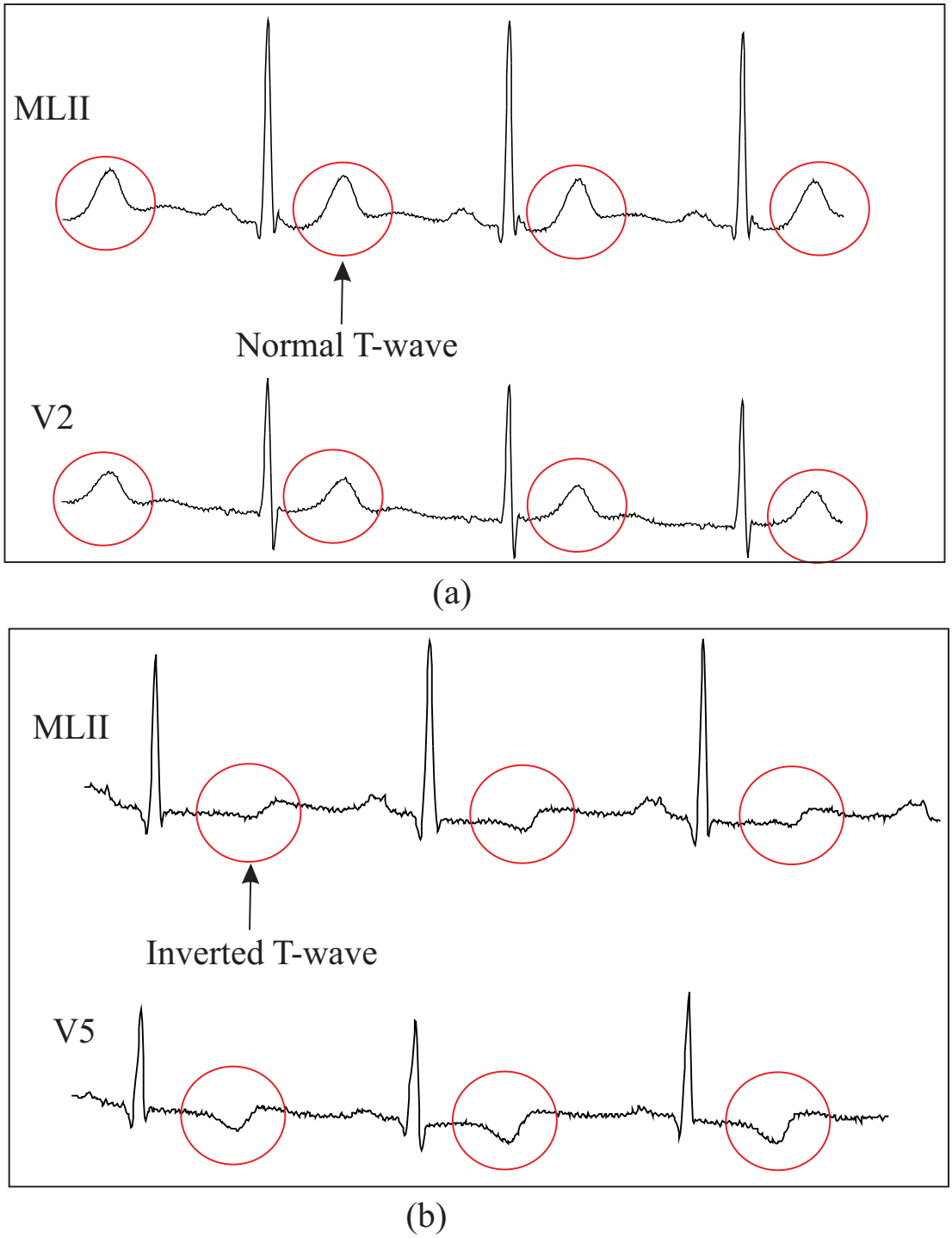

Figure 2.4: Representation of the normal and abnormal T-wave

2.2 Estimation theory

Electronic signal processing systems can be designed under foundations of estimation theory. These systems extract information and require estimating parameters by satisfying a determined model. Systems such as radar, communication, sonar, speech, seismology, image analysis, control and biomedicine share the trivial issue of needing to estimate the values of a group of parameters [86]. In this investigation the foundations about estimation theory play a important role for denoising and features extraction of electrocardiogram ECG signals.

Problem formulation: Modelling of estimation

The first step to determine a good estimator is to fit the data to a mathematical model. Considering that data are random, a probability density function (PDF) can be parametrized upon unknown parameters θ,

$$p(x[0],\, x[1],\, \ldots, x[N-1]; \theta), \tag{2.1}$$

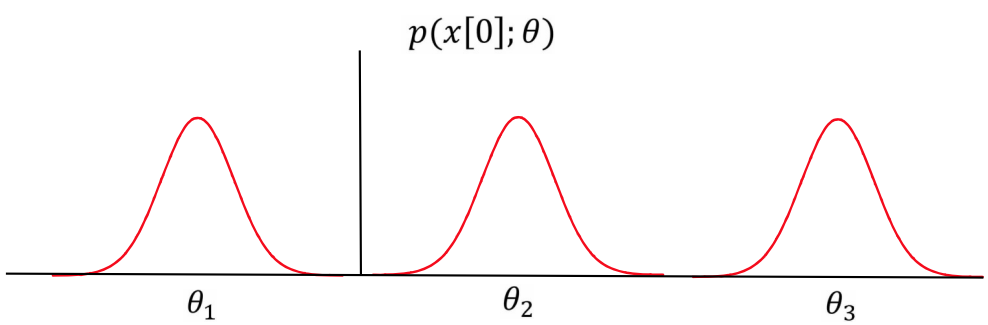

Figure 2.5: Dependence de PDF on unknown parameters

i.e., we stablish a class of PDFs determined by different values of θ. The semicolon denote this independence. To illustrate, if $N = 1$ and θ indicate the mean, then the PDF of data would be expressed as

$$p(x[0]; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[\frac{0.5}{\sigma}(x[0] - \theta)^2] \tag{2.2}$$

which is depicted in Figure 2.5 for distinct values of θ. It should be naturally obvious that because the value of θ influences the probability of $x[0]$, we should be able to assume the value of θ from the observed value of $x[0]$. For example, if $x[0] < 0$ then it is uncertain that $\theta = \theta_2$.

The value $\theta = \theta_1$ might be more reasonable. This specification of the PDF is fundamental for determining an appropriate estimator.

2.2.1 Estimator evaluation

Given the measurement signal $x[n]$ shown in the figure 2.1, where A is the reference signal corrupted by noise $w[n]$ with zero mean, we can express a model as follow

$$x[n] = A + w[n] \ . \tag{2.3}$$

Considering the data set $x[0], x[1], \ldots, x[N-1]$, it is possible to estimate A. Intuitively, the average de $x[0]$ can determine the value A. Then, we can estimate A as

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \tag{2.4}$$

Here, we consider a estimator of A, e.g $\tilde{A} = x[0]$. However, this information appropriately does not represent the data. An estimator should be a function of the data determined as random variables. Then, to evaluate of estimator performance, we employed a statistical analysis. Then, a significant number of experiments is necessary to reach the value closer of A. We consider the concentration of the histogram of the sample mean value $\hat{A}$ and sample $\hat{A}$ for 100 iterations (The figure 2.7 represents the histogram which describes the number of times the estimator produces a given range of values and is an approximation to the PDF). To demonstrate that $\hat{A}$ is better than $\tilde{A}$, we could establish that the variance is less. We assume a model with zero mean noise $w[n]$ and variance σ^2. Then, it is appropriate to demonstrate that the mean of each estimator is the true value,

$$\begin{aligned} E(\hat{A}) &= E(\frac{1}{N} \sum_{n=0}^{N-1} x[n]) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} E(x[n]) \\ &= A \end{aligned} \tag{2.5}$$

and, $E(\tilde{A}) = E(x[0])$, then $E(\tilde{A})$=A. The variances can be expressed as,

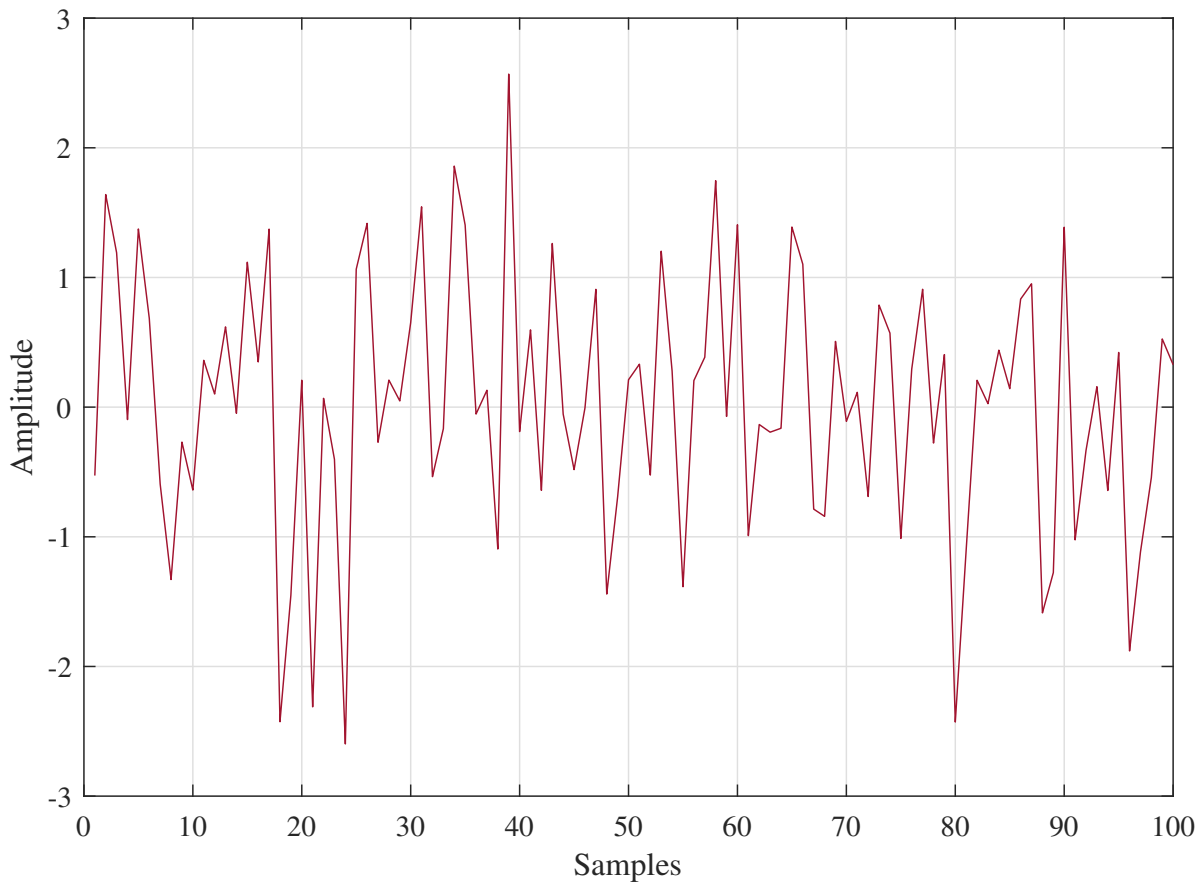

Figure 2.6: Random data sample for A

$$
\begin{aligned}
var(\hat{A}) = \mathrm{var}(\frac{1}{\mathrm{N}}\sum_{\mathrm{n=0}}^{\mathrm{N-1}}(\mathrm{x[n]})) &= \frac{1}{N^2}\sum_{n=0}^{N-1}\mathrm{var}(x[n]) \\
&= \frac{1}{N^2}N\sigma^2 \\
&= \frac{\sigma^2}{N}
\end{aligned}
\tag{2.6}
$$

give that $w[n]$ component is uncorrelated and thus

$$
\begin{aligned}
\mathrm{var}(\tilde{A}) = \mathrm{var}(x[n]) &= \sigma^2 \\
&> \mathrm{var}(\tilde{A}).
\end{aligned}
\tag{2.7}
$$

Finally, it is reasonable to say that an estimator is a random variable where its performance can be described by its probability density function (PDF).

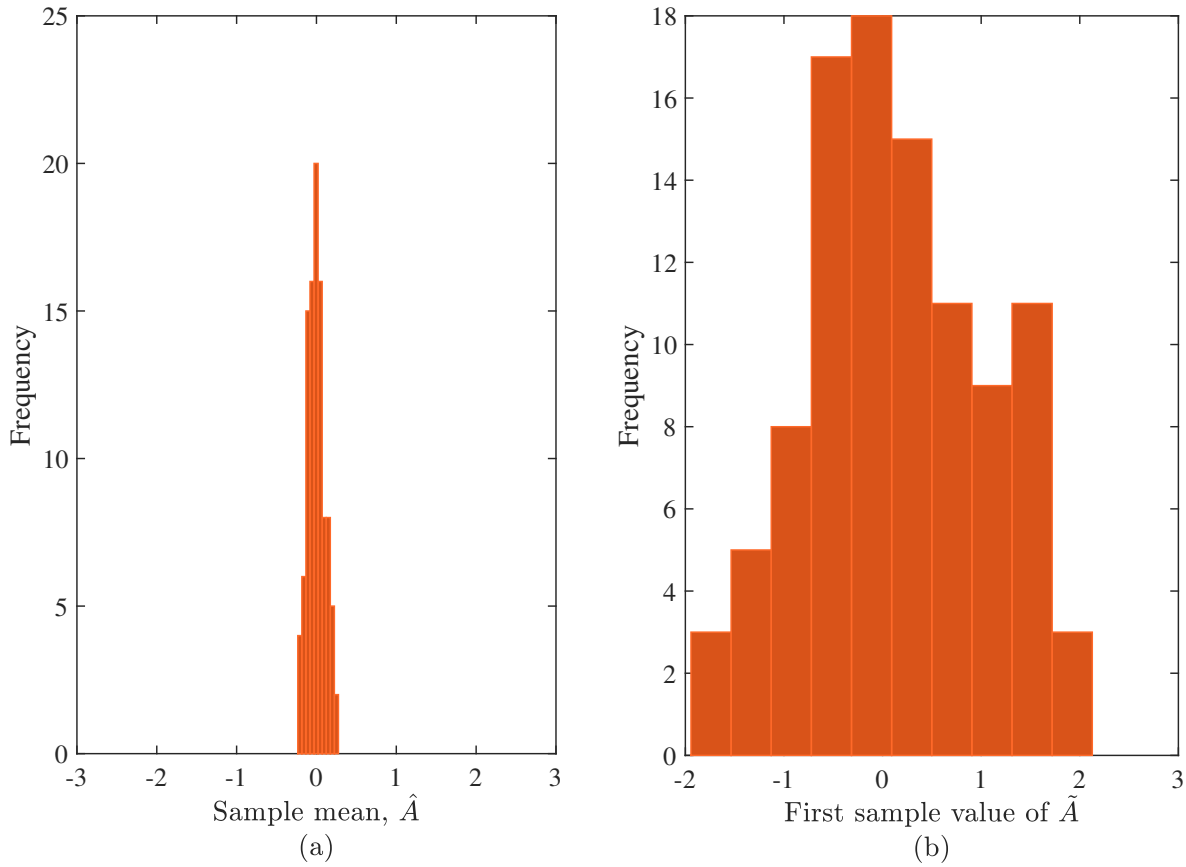

Figure 2.7: Histogram for sample mean $\hat{A}$ (a) and first value $\tilde{A}$ (b)

2.2.2 Unbiased estimation

For an estimator to be unbiased we rough that on the average the estimator will provide the true value of the unknown parameter. Since the parameter value may be anywhere in the interval $x_1 < \theta < x_2$, unbiasedness property argues that no matter what the true value of θ, our estimator will yield it on the average

The unbiasedness describe that no matter what the true value of θ, our estimator will yield it on the average . Analytically, an estimator is unbiased if

$$E(\tilde{\theta}) = \theta, \; x_1 < \theta < x_2 \tag{2.8}$$

where $x1$ and $x2$ denotes the range of possible values of θ. To illustrate, we consider the following observation,

$$x[n] = A + w[n] \; n = 0, 1, \ldots N-1 \tag{2.9}$$

Here, A is the parameter to be estimated and $w[n]$ is white Gaussian noise (WGN). The parameter A takes values in the interval $-\infty < A < \infty$. Hence, a estimator for the average value of $x[n]$ can be represented by

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \tag{2.10}$$

then, analysing the linearity properties of the expectation operator for all A. We employed the same analysis expressed in (2.5)

$$\begin{aligned} E(\hat{A}) &= E\left| \tfrac{1}{N} \textstyle\sum_{n=0}^{N-1} x[n] \right| \\ &= A \end{aligned} \tag{2.11}$$

Hence, the sample mean estimator is unbiased. Here, we consider an unbiased estimator that has symmetric PDFs centered on the real value of θ over the constraint given by (2.8)

2.2.3 Criterion of minimum variance

For finding the optimal estimators, a optimal criterion it is necessary. A natural one is the mean square error (MSE), defined as

$$mse(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{2.12}$$

Therefore, this defines the average mean squared deviation of the estimator from the true value. However, the assumption of this natural criterion induces to impossibly to achieve a realizable estimator. This cannot be defined justly as a function of the data. To understand the problem which arises the equation (2.12) can be rewrite as

$$\begin{aligned} mse(\hat{\theta}) &= E\left\{ \left| (\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \hat{\theta}) \right|^2 \right\} \\ &= \mathrm{var}(\hat{\theta}) + |E(\hat{\theta}) - \theta|^2 \\ &= var(\hat{\theta}) + \mathrm{bias}^2(\theta) \end{aligned} \tag{2.13}$$

which demonstrates that the mean square error (MSE) depends on the variance and bias of the estimator.

For practical effects, the minimum mean square estimator (MSE) is no required. A trustworthy alternative is to constrain the bias to zero and find the estimator which minimizes the variance. This estimator is called the minimum variance unbiased estimator (MVUE).

This estimator belongs to classical estimation methods such as: Best linear unbiased estimator, maximum likehood estimator (MLE) and Gauss's least square (Gauss's LS) among others. Unlike the above mentioned methods, Bayesian estimation methods such as the famous Kalman filter have been considered to be into of modern methods. however, this method it is considered when a model of problem is known. Then, a promising alternative is to apply a robust filter such as the unbiased FIR filter (UFIR) which could be considered as versatile structure depending of problem. It mean this structure can work like filter, smoothing and prediction. Also, the UFIR structure can be represented into two models: Polynomial and harmonic model. The chapters following describes the UFIR estimator developed by Carlos Lastre-Dominguez for electrocardiogram signals.

2.3 Previous work for ECG signals

2.3.1 Time based techniques

Techniques such as linear prediction, Savitsky-Golay smoothing, median filter, notch filter and standard filter such as low pass filter and passband filter have been widely employed for ECG signals. Each technique have been employed during decades for ECG signal features extraction. Among these methods, the linear predict approach proposed in [1] and developed by Martis [32] has been recognized as one of most efficient. The method employs the following model

$$\hat{\lambda}(n) = \sum_{i=1}^{q} \delta(i)\lambda(n-i)\,, \tag{2.14}$$

in which $\lambda(n)$ is the original ECG pulse, q is the estimator order, and $\delta(i)$ is the linear prediction coefficient. The estimate $\hat{\lambda}(n)$ is provided as a linear weighted combination of $\lambda(n-i)$, $i = [1, q]$. The residual error

$$\epsilon(n) = \lambda(n) - \hat{\lambda}(n)\,, \tag{2.15}$$

is considered as the ECG signal fraction, which cannot be predicted. To compare with the UFIR filter, we will assign $q = 2$ as suggested by Lin *et. al.* [29].

The UFIR filter predicts estimates with $p > 0$ and both the prediction estimator (2.14) and the UFIR predictive filter represented by equation (3.2a) and described in chapter 3, employ

Table 2.2: Previous work for ECG signals

Literature	Approach	Limitation	Database
Peyman *et. al.*[24]	EMD	Empirical	FECG
Manpreet *et. al.* [25]	DTW	Limited to mother wavelet	MIT-BIH Arrhythmia
Martis *et. al.* [26]	DTW	Limited to mother wavelet	MIT-BIH Arrhythmia
Martis *et. al.* [31]	Linear predict	Increase the signal noise	MIT-BIH Arrhythmia
Berk *et. al.* [39]	S-G smoothing	p-lag limited to middle of horizon N	MIT-BIH Arrhythmia
Golden *et. al.* [18]	TF	Resolution loss	Own database
Rajesh *et. al.* [23]	EMD	Empirical	MIT-BIH Arrhythmia
Goel *et. al.* [56]	DTW-PSD	Limited to mother wavelet	Own database

discrete linear prediction of the undergoing process via its noisy data . Even so, there are some zones in the ECG picture where linear predictors are unsuccessful in extracting ECG features. Therefore, a comparative analysis of different methods developed in [68–71] is required.

The Savitzky – Golay filter (SG) is a filter, first described in 1964 by Abraham Savitzky and Marcel JE Golay. The SG approach performs on the calculation of a k-degree polynomial localregression, with at least $k+1$ equally spaced points, to determine the new value of each point. The result will be an input function smoothed. The primary advantage of this approach is that it preserves characteristics of the initial distribution such as the relative maxima and minima, and the width of the peaks, which commonly disappear with other averaging approaches such as the moving average.

2.3.2 Frequency based techniques

Methods of ECG analysis in the time domain methods provide better visualization of the data, but the alterations on time in the amplitude and frequency of ECG are not particularly well defined. Many techniques such as describe significant classification efficiency, they have been assessed on noise free conditions. In real scenario, we should consider noise, baseline wander and artifacts. There is a requiring to analyze the information from these noisy ECG data . To analyze such hidden information in the ECG data, different frequency domain methods such as the Fourier transform and the power spectral density (PSD) methods have been investigated.

2.3.3 Time-frequency based techniques

In the case of wavelet transform (WT or DWT) , the frequency resolution may be prejudiced to some extent to raise the temporal resolution. Hence, the inverse of frequency called the scale is employed. The wavelet transform provides multiple resolutions, recognized as multi-resolution analysis. Wavelet transform can need discrimination between two distinct signals with the same spectrum magnitude (absolute amplitude of frequency spectrum. The translation and the dilation of basis function known as mother wavelet which allows the extensive spectrum of coordinates at every scale.

2.3.4 Empirical mode decomposition

Empirical mode decomposition (EMD) is a nonlinear and adaptive approach representing the detailed fluctuations in the signal. This technique extends a determined ECG signal into a few number of intrinsic mode functions (IMFs) to provided the instantaneous frequency. Instantaneous magnitude and its frequency are obtained from the IMFs. The IMFs are amplitude and frequency modulated waves. The table 2.2, we describe some previous works and its limitations. In the following chapter, we develop a Unbiased finite impulse response (UFIR) smoothing appropriated for ECG signals.

Chapter 3

Denoising and features extraction for ECG signals using unbiased finite impulse response (UFIR) smoothing

In this chapter, we develop an adaptive-horizon UFIR smoothing filtering algorithm for denoising ECG signals and features extraction. We also investigate the trade-off between the UFIR smoothing filter, UFIR filter, and UFIR predictive filter and compare them to the standard linear predictor suggested in [29]. We base our investigations on the MIT-BIH Arrhythmia Database available for free use from [75, 76]. Focused on the design of efficient algorithms, in this work we limit our investigations by data associated with normal heartbeats and postpone an analysis of different kinds of heart diseases to future investigations. The rest of the chapter is organized as follows. In the section 3.1, we develop the theory of the algorithms proposed. In the section **??**, we perform the validation by defining the methods based on linear prediction. The experimental results are showed in the Sections 3.2, 3.3, 3.3.3 where we provide a comparison between the UFIR, UFIR smoothing, and UFIR predictive filtering algorithms. A discussion of the results is provided in Section 3.4.

3.1 p-shift UFIR smoothing filtering

Let us suppose that the ECG signal x_n (Fig. 3.1) is contaminated by zero mean additive noise v_n with unknown statistics. Then measurement s_n of x_n can be represented in discrete time index

n as an additive sum of

$$s_n = x_n + v_n \,. \tag{3.1}$$

In view of the fact that noise v_n in (3.1) may not be white Gaussian and its statistics are commonly not well-known, the best way to avoid large estimation errors is using filters that do not require information about the statistics of noise. The p-shift UFIR filter, which completely ignores noise and the initial conditions, can thus be considered as a good candidate.

On a finite horizon $[m-p, n-p]$ of N points, the ECG signal can be represented with a degree polynomial and the p-shift UFIR filter [69] applied to remove noise. In accordance with [69], the UFIR estimate $\hat{x}_{n|n-p}$ of x_n via data s_n taken from $[m-p, n-p]$ can be found in the convolution-based form of

$$\hat{x}_{n|n-p} = \sum_{i=p}^{N-1+p} h_{li}(p) s_{n-i} \tag{3.2a}$$

$$= \mathbf{W}_l^T(p) \mathbf{S}_N(p) \,, \tag{3.2b}$$

where $h_{ln}(p) \triangleq h_{ln}(N,p)$ is the $\{N,p\}$-variant impulse response of the l-degree UFIR filter, the extended measurement vector $\mathbf{S}_N$ is

$$\mathbf{S}_N(p) = [\, s_{n-p} \; s_{n-1-p} \; \cdots \; s_{m-p} \,]^T \,, \tag{3.3}$$

and the filter gain matrix is given by

$$\mathbf{W}_l^T(p) = [\, h_{lp}(p) \, h_{l(1+p)}(p) \cdots h_{l(N-1+p)}(p) \,] \,. \tag{3.4}$$

If to satisfy the unbiasedness condition

$$E\{\hat{x}_{n|n-p}\} = E\{x_n\} \,, \tag{3.5}$$

where $E\{z\}$ means an average of z, then $h_{ln}(p)$ can be represented as [69, 87]

$$h_{li}(p) = \sum_{j=0}^{l} a_{jl}(p) i^j \,, \tag{3.6}$$

where $i \in [p, N-1+p]$ and the coefficients $a_{jl}(p)$ are defined by [69]

$$a_{jl}(p) = (-1)^j \frac{M_{(j+1)1}(p)}{|\mathbf{D}(p)|} \,. \tag{3.7}$$

Here, $|\mathbf{D}(p)|$ is the determinant of matrix $\mathbf{D}(p) = \mathbf{V}^T(p)\mathbf{V}(p)$, where $\mathbf{V}(p)$ is the $N \times (l+1)$ Vandermonde matrix,

$$\mathbf{V}(p) = \begin{bmatrix} 1 & p & p^2 & \dots & p^l \\ 1 & 1+p & (1+p)^2 & \dots & (1+p)^l \\ 1 & 2+p & (2+p)^2 & \dots & (2+p)^l \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & N-1+p & (N-1+p)^2 & \dots & (N-1+p)^l \end{bmatrix} \tag{3.8}$$

and $M_{(j+1)1}(p)$ is the minor of $\mathbf{D}(p)$. Function $h_{li}(N,p)$ has the following fundamental properties [69, 87]:

$$h_{li}(N,p) = \begin{cases} \text{nontrivial}, & p \leqslant i \leqslant N-1+p \\ 0, & \text{otherwise} \end{cases}, \tag{3.9}$$

$$\sum_{i=p}^{N-1+p} h_{li}(N,p) = 1\,, \tag{3.10}$$

$$\sum_{i=p}^{N-1+p} h_{li}(N,p) i^u = 0, \quad 1 \leq u \leq l\,. \tag{3.11}$$

For low-degrees, $l = 1$ and $l = 2$, one can find $h_{li}(N,p)$ in Appendix A. For higher degrees, $h_{li}(N,0)$ can be computed using a recurrence relation [88, 89] and then $h_{li}(N,p)$ is obtained by a projection. Of importance is that the UFIR estimate (3.2b) does not require the noise statistics and initial values. The zero mean noise v_n is allowed to have any distribution and covariance [90, 91] that is a fundamental difference with optimal estimates.

3.1.1 ECG Signal denoising on adaptive horizons

The determination of optimal horizon N_{opt} is critical in UFIR filtering and smoothing [92]. Because a reference signal is unavailable for ECG data, N_{opt} can be found following [93] via the mean square value (MSV) $V(N,p) = E\{\varepsilon_n(N,p)^2\}$ of the measurement residual $\varepsilon_n(N,p) = s_n - \hat{x}_{n|n-p}(N)$. It has been shown in [93] that $N_{\text{opt}}(p)$ can be estimated by minimizing the derivative $\partial V(N,p)/\partial N$ as

$$\hat{N}_{\text{opt}}(p) = \arg\min_N \frac{\partial V(N,p)}{\partial N} + 1\,. \tag{3.12}$$

To optimize the horizon, let us consider a single ECG pulse shown in Fig. 3.2. As can be seen, the ECG pulse is slowly changing, except for a fast excursion in the QRS region. The slow background requires an optimal horizon $N_{\text{opt}} \triangleq N_{\text{opt}}(p)$ in order to provide best denoising with

no essential bias. On the contrary, the QRS region requires a minimum horizon $N_{\min} \triangleq N_{\min}(p)$ of $l+1$ points to track the behavior exactly. The horizon N must thus be adaptive.

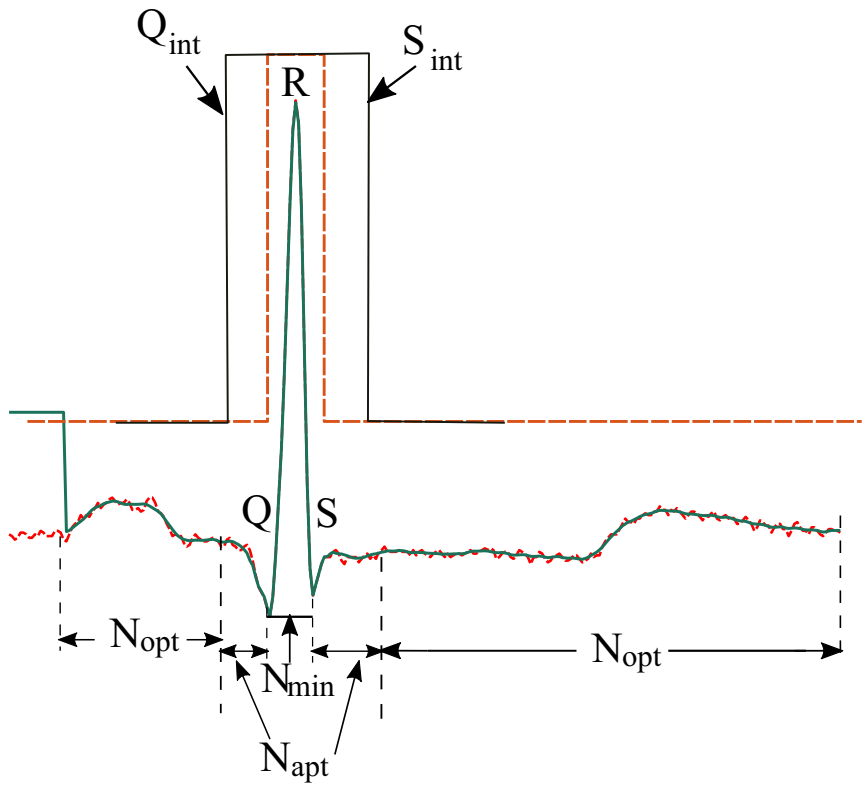

Figure 3.1: An original single ECG pulse corrupted by measurement noise (dashed) and the denoised pulse (solid line). Slow parts of the ECG pulse require denoising with N_{opt} and a fast excursion requires a minimum horizon length of $N_{\min} = l+1$ points. Here, Q and S are the morphology features of ECG signal, Q_{int} and S_{int} represent the window width allowed for the adaptation. The adaptive horizon N_{apt} ranges from N_{opt} to $N_{\min}$.

3.1.2 General UFIR smoothing algorithm

The general UFIR smoothing algorithm is represented with a pseudo code as Algorithm 1. It requires values of $\mathbf{S}_N$, N, and l described above. Function **CalculateG** provides a vector $\mathbf{G} = [1\,0\cdots 0]^T$ and **CalculateV** calculates matrix $\mathbf{V}$ given by (3.8). Vector $\mathbf{B}$ contains the UFIR filter coefficients (5.6). Provided $\mathbf{V}$ and $\mathbf{B}$, the l-degree matrix $\mathbf{W}_l$ is computed and estimate $\hat{x}_{n|n-p}(N)$ is provided by (3.2b). We will use this algorithm at different horizons as **smoothingUFIR** function.

Algorithm 1 General UFIR Smoothing Algorithm for Estimating $\hat{x}_{n|n-p}(N)$

Data: $\mathbf{S}_N$, l , N,

Result: $\hat{x}_{n|n-p}(N)$

1: **Begin** :

2: $\mathbf{G} = \mathbf{CalculateG}\ (l)$

3: $p{=}-(N-2)/2$

4: $\mathbf{V}{=}\mathbf{CalculateV}\ (p, l, N)$

5: $\mathbf{B}{=}(\mathbf{V}\mathbf{V}^T)\mathbf{G}$

6: $\mathbf{W}_l(p){=}\mathbf{V}^T\ \mathbf{B}$

7: $\hat{x}_{n|n-p}(N){=}\mathbf{W}_l(p)\ \mathbf{S}_N(p)$

3.1.3 Computing N_{opt} for ECG data

Optimal horizon N_{opt} is provided by the algorithm designed, which pseudo code is listed as Algorithm 2. This algorithm requires the following variables: heartbeats data s_i, filter degree l, a set of heartbeats $beats$, the number of heartbeats $Nbeats$, and the window width $Interval$.

By defining (3.8) and (3.9) and then analysing (3.12), the filter coefficients specified by (3.13) are obtained for given l, N, and p. Next, coefficients are computed for (5.11) and estimate (3.1) is provided as $\tilde{x}_{n|n-p}(N)$. Function **IntervalQRS** is introduced to detect Q and S via data s_i and a value called $Interval$, which is related to the window width.

The window covers a region including Q and S and is used as N_{min}. Because N_{opt} will produce highly biased estimates around Q and S, the window is split into three parts:

$$\tilde{x}_{n|n-p}(N)_1 = \tilde{x}_{n|n-p}(N)(1 : \text{Q}_{\text{int}} - 1)\,, \tag{3.13}$$

$$\tilde{x}_{n|n-p}(N)_2 = \tilde{x}_{n|n-p}(N_{\text{min}})(\text{Q}_{\text{int}} : \text{S}_{\text{int}})\,, \tag{3.14}$$

$$\tilde{x}_{n|n-p}(N)_3 = \tilde{x}_{n|n-p}(N)(\text{Q}_{\text{int}} : \text{T})\,, \tag{3.15}$$

where points Q_{int} and S_{int} determinate the window width for $Inteval$. The horizon N_{opt} is applied to the first part (3.13) and third part (3.15). In the second part (3.14), estimation is provided with N_{min}.

Function **Cat** is used to concatenate estimates (3.13)–(3.15) and compute the final estimate $\hat{x}_{n|n-p}(N)$. Provided $\hat{x}_{n|n-p}(N)$, function $V_n(N)$ is calculated for s_i in the N scale. This variable is saved as $V_i(N)$ to represent a whole set of data $V_n(N)$ for different heartbeats. Provided various values of MSV for each s_i, an average of $V_i(N)$ is computed as V_{avg}. Because V_{avg} is

Algorithm 2 Algorithm for estimating $\hat{N}_{\text{opt}}$ using measurements

Data: s_i , l , $beats$, $Nbeats$, $Interval$

Result: $\hat{N}_{\text{opt}}$

Begin :

for $i = 1$ to $Nbeats$ **do**

  s_i= $beats(i)$

  T=**length**(s_i)

  for $N = l + 1$ to N_{max} **do**

    $\tilde{x}_{n|n-p}(N) =$ **SmoothingUfir**$(s_i$,N,$l)$

    $[\mathrm{Q}_{int} \quad \mathrm{S}_{int}]$ =**IntervalQRS**$(s_i, Interval)$

    $\tilde{x}_{n|n-p}(N)_1 = \tilde{x}_{n|n-p}(N)(1 : \mathrm{Q}_{int} - 1)$

    $\tilde{x}_{n|n-p}(N)_2 = \tilde{x}_{n|n-p}(N_{min})(\mathrm{Q}_{int} : \mathrm{S}_{int})$

    $\tilde{x}_{n|n-p}(N)_3 = \tilde{x}_{n|n-p}(N)(\mathrm{S}_{int} + 1 : \mathrm{T})$

    $\hat{x}_{n|n-p}(N)$=**Cat**$(\tilde{x}_{n|n-p}(N)_1, \tilde{x}_{n|n-p}(N)_2,$
    $\tilde{x}_{n|n-p}(N)_3)$

    $V_n(N) = E\{[s_i - \hat{x}_{n|n-p}(N)]^2\}$

  end for

  $V_i(N) = V_n(N)$

end for

$V_{avg}(N)$ =**Average**$(V_i(N))$

$V(N) =$ **CubicFit**$(V_{avg}(N))$

$\hat{N}_{\text{opt}} = \arg\min_{N} \frac{\partial V(N)}{\partial N} + 1$

accompanied with ripples causing ambiguities, it is further approximated with a cubic polynomial using function **CubicFit**. The derivative applied to smoothed V_{avg} while solving the optimization problem (3.12) yields N_{opt}.

3.1.4 Denoising algorithm for ECG signals

Provided N_{min} and N_{opt}, the UFIR smoothing algorithm can be designed for ECG signals with a pseudo code listed as Algorithm 3 . In this algorithm, function **smoothingUFIR** is applied to different ECG signals parts with different horizons.

Algorithm 3 Algorithm for estimating $\hat{x}_{n|n-p}$ in ECG signals

Data: $\mathbf{S}_N$, l , N_{min}, N_{opt}

Result: $\hat{x}_{n|n-p}$

1: $p = \frac{-(N-1)}{2}$
2: **for** $N = N_{min}$ to N_{opt} **do**
3: $\quad \tilde{x}_{n|n-p}(N) =$ **smoothingUFIR** ($\mathbf{S}_N$, N, l)
4: **end for**
5: $[\mathrm{Q}_{int} \quad \mathrm{S}_{int} \quad \mathrm{Q} \quad \mathrm{S}] =$**IntervalQRS**$(\mathbf{S}_N, Interval)$
6: $\tilde{x}_{n|n-p}(N)_1 = \tilde{x}_{n|n-p}(N_{opt})(1 : \mathrm{Q}_{int} - 1)$
7: $\tilde{x}_{n|n-p}(N)_2 = \tilde{x}_{n|n-p}(N_{apt})(\mathrm{Q}_{int} : \mathrm{Q} - 1)$
8: $\tilde{x}_{n|n-p}(N)_3 = \tilde{x}_{n|n-p}(N_{min})(\mathrm{Q} : \mathrm{S})$
9: $\tilde{x}_{n|n-p}(N)_4 = \tilde{x}_{n|n-p}(N_{apt})(\mathrm{S} + 1 : \mathrm{S}_{int})$
10: $\tilde{x}_{n|n-p}(N)_5 = \tilde{x}_{n|n-p}(N_{opt})(\mathrm{S}_{int} + 1 : T)$
11: $\hat{x}_{n|n-p} =$ **Cat** $((\tilde{x}_{n|n-p}(N)_1, \tilde{x}_{n|n-p}(N)_2,$
$\tilde{x}_{n|n-p}(N)_3), \tilde{x}_{n|n-p}(N)_4), \tilde{x}_{n|n-p}(N)_5))$

Five parts of the ECG signal are recognized by function **smoothingUFIR** over points Q_{int}, S_{int}, S and Q:

$$\tilde{x}_{n|n-p}(N)_1 = \tilde{x}_{n|n-p}(N_{opt})(1 : \mathrm{Q}_{int} - 1)\,, \tag{3.16}$$

$$\tilde{x}_{n|n-p}(N)_2 = \tilde{x}_{n|n-p}(N_{apt})(\mathrm{Q}_{int} : \mathrm{Q} - 1)\,, \tag{3.17}$$

$$\tilde{x}_{n|n-p}(N)_3 = \tilde{x}_{n|n-p}(N_{min})(\mathrm{Q} : \mathrm{S})\,, \tag{3.18}$$

$$\tilde{x}_{n|n-p}(N)_4 = \tilde{x}_{n|n-p}(N_{apt})(\mathrm{S} + 1 : \mathrm{S}_{int})\,, \tag{3.19}$$

$$\tilde{x}_{n|n-p}(N)_5 = \tilde{x}_{n|n-p}(N_{opt})(\mathrm{S}_{int} + 1 : T)\,. \tag{3.20}$$

In Fig. 3.2, the first and fifth parts are defined by (3.16) and (3.20), respectively to apply N_{opt}. The third part represents an estimate, which is equal to the original ECG signal without noise reduction on [Q, S]. The adaptive horizon N_{apt} is applied to (3.17) and (3.19). Here, N is decreased from N_{opt} to with a one-time step in the QRS complex region. Beyond the QRS complex, N is gradually increased from N_{min} to N_{opt} with a one-time step. Finally, function **Cat** provides the ECG signal estimate at the last fifth part.

3.1.5 UFIR-based algorithm for features extraction

Provided denoising by Algorithm 3, in this section we develop an efficient computation algorithm for ECG signal features extraction. To this end, we first localize special points on the ECG heartbeat pulse and then compute relevant amplitudes, durations, and an angle. Unlike the approaches developed in [20, 94, 95], this algorithm is based on the p-shift and l-order UFIR smoothing filter exploited with $l = 2$ and $p < 0$. It was found out for data used that $N_{\text{opt}} = 21$ suites smooth parts of the discrete ECG signal and $N_{\text{min}} = 3$ fits the QRS complex. Note that N_{opt} and N_{min} must be specified for each of the measured ECG signals.

Step-by-step events representing the strategy of ECG signal denoising and features extraction are shown in Fig. 3.2. The original discrete-time ECG signal (a) is smoothed as (b) using Algorithm 3. Then the ECG signal features are extracted as in the following:

- Fig. 3.2(c): The peak value R (ECG signal maximum) is estimated as $\hat{\mathrm{R}}$ and a window is introduced with two points, Q' and S'. The estimate $\hat{\mathrm{Q}}$ of Q is found as the least in the interval between Q' and $\hat{\mathrm{R}}$. The estimate $\hat{\mathrm{S}}$ of S is found as the least between $\hat{\mathrm{R}}$ and S'.

- Fig. 3.2(d): Provided $\hat{\mathrm{Q}}$, $\hat{\mathrm{R}}$ and $\hat{\mathrm{S}}$, the QRS complex is suppressed to save only P and T waves. Then the estimates $\hat{\mathrm{P}}$ of P and $\hat{\mathrm{T}}$ of T are obtained similarly by suppressing one of the waves.

- Fig. 3.2(e): Provided $\hat{\mathrm{P}}$, 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 section of the resulting signal and find a global maximum. We consider it as a start point of P wave and call it $\mathrm{P}_{\text{onset}}$. 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 $\mathrm{P}_{\text{offset}}$. Values of $\mathrm{P}_{\text{onset}}$ and $\mathrm{P}_{\text{offset}}$ are

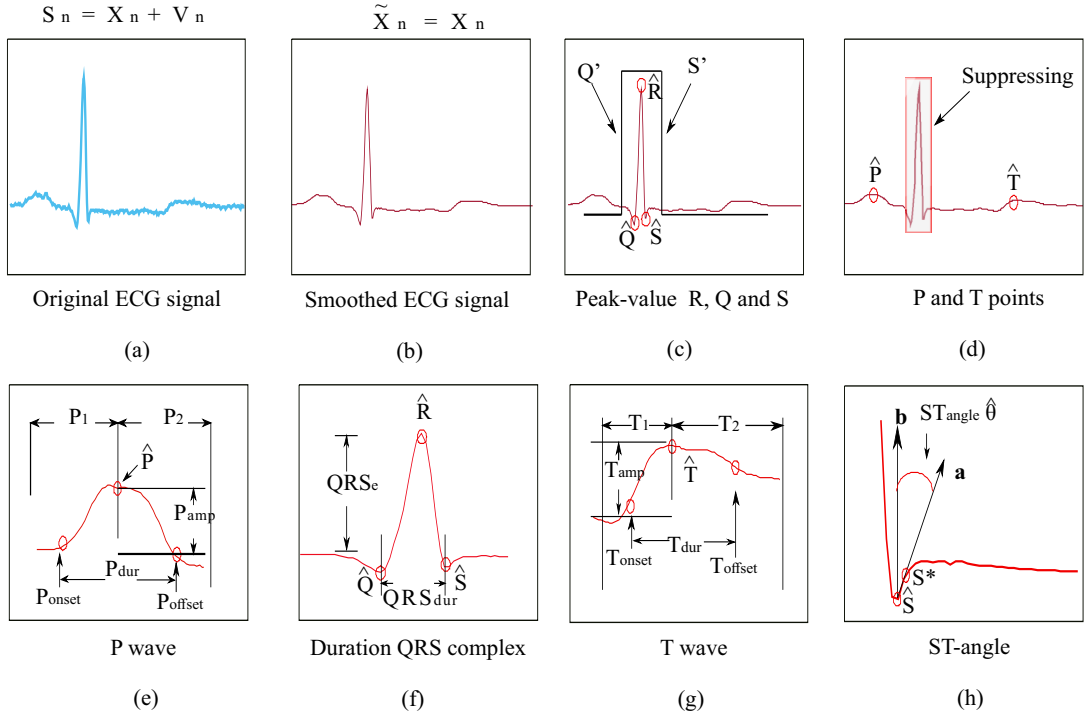

Figure 3.2: Step-by-step events representing the strategy of the ECG signal denoising and features extraction: (a) original ECG signal, (b) smoothed ECG signal, (c) peak-value R, Q, and S, (d) P and T points, (e) P wave, (f) duration of QRS complex, (g) T wave, and (h) ST-angle.

located at points[1] P_p^{on} and P_p^{off}, respectively. Then, the duration of P wave is computed as $P_{dur} = P_p^{off} - P_p^{on}$. A distance between $\hat{P}$ and the baseline is calculated and called the wave amplitude.

- Fig. 3.2(f): The QRS complex duration is obtained by the distance between points $\hat{Q}$ and $\hat{S}$. The QRS complex amplitude is provided by a distance between the baseline and $\hat{R}$.
- Fig. 3.2(g): Similarly, points T_{onset} and T_{offset} are obtained for the T wave by splitting this wave into two segments, T_1 and T_2.
- Fig. 3.2(h): The ST-angle θ is computed by

$$\frac{\mathbf{a}.\mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \cos\theta\,, \tag{3.21}$$

[1] Although P_p^{on} and P_p^{off} are omitted in Fig. 3.1, their values represent the temporal line in the ECG signal. These points can be used to compute features of the duration and applied to R_p, Q_p S_p, P_p, T_p, T_p^{on}, T_p^{off}, S_p and S_p^*, which are described above in Algorithm 4.

where **a** and **b** are vectors created from $\hat{\mathrm{S}}$ and S^*. These values are localized in $\mathrm{S_p}$ and $\mathrm{S_p^*}$. Vectors **a** and **b** have two components dependent on $\hat{\mathrm{S}}$ and S^*. We consider a flat part, where $\mathrm{S_p}$ and $\hat{\mathrm{S}}$ represent the origin zero point. We sum a temporal unity from the origin, obtain S^* and $\mathrm{S_p^*}$, and rename S^* as $\mathrm{S_y^*}$ and $\mathrm{S_p^*}$ as $\mathrm{S_x^*}$ from xy plane. We then compute $\mathbf{a} = \mathrm{S_x^*} + \mathrm{S_y^*}$ and $\mathbf{b} = 0_x + \mathrm{S_y^*}$ and estimate θ via (3.18).

3.1.6 Algorithm design for features extraction of ECG signals

A pseudo code of the algorithm designed to extract features of ECG signal is shown as Algorithm 4. Here, ss_i is the smoothed ECG signal represented as $\tilde{\mathrm{x}}$ in Fig 3.2; N_b is the number of heartbeats; *Baseline* is a variable, which 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 (Fig 3.2). The algorithm output consists of estimates of the ECG signal features such as $\hat{\mathrm{P}}$ of P, $\mathrm{P_{amp}}$ of the P amplitude, $\mathrm{P_{dur}}$ of the P duration, QRS_e of the QRS amplitude, $\mathrm{QRS_{dur}}$ of the QRS duration, $\hat{\mathrm{T}}$ of T, $\mathrm{T_{amp}}$ of the T amplitude, $\mathrm{T_{dur}}$ of the T duration, and $\hat{\theta}$ of the ST angle θ. All these features are extracted from the smoothed signal ss_i.

The algorithm starts by computing $\hat{\mathrm{R}}$ as the ECG signal maximum, using function **max**. Function **IntervalQRS** is applied to compute points Q' and S'. The *Interval* variable determines the window width to cover the QRS complex and obtain $\hat{\mathrm{Q}}$ and $\hat{\mathrm{S}}$ as two minima between points Q' and S'. Function **min** is used to find the above-mentioned points. The **supress** function is used to suppress the QRS complex. Function **max** is used to estimate P and T. Function **diff** is introduced to compute the derivatives in the P_1, P_2, T_1 and T_2 intervals. Functions **max** and **min** with function **diff** are used to find $\mathrm{P_{onset}}$, $\mathrm{P_p^{on}}$, $\mathrm{P_{offset}}$, $\mathrm{P_p^{off}}$, $\mathrm{T_{onset}}$ $\mathrm{T_p^{on}}$ and $\mathrm{T_{offset}}$ $\mathrm{T_p^{off}}$. Provided the above-mentioned values, the duration is estimated of P and T features. Function **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 $\mathrm{P_{offset}}$. Function **vector** is used to provide vectors **a** and **b** based on $\mathrm{S_p}$, $\hat{\mathrm{S}}$, $\mathrm{S^*}_\mathrm{p}$ and S^*. Finally, function **arcos** is used to compute an angle between vectors **a** and **b**. Note that all the above introduced functions are available from the authors by request.

Algorithm 4 A pseudo code of the algorithm to extract morphological features of ECG signals

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

Result: $\hat{\mathrm{P}}$, $\mathrm{P_{amp}}$, $\mathrm{P_{dur}}$, $\hat{\mathrm{R}}$, QRS_e, $\mathrm{QRS_{dur}}$, $\hat{\mathrm{T}}$, $\mathrm{T_{amp}}$, $\mathrm{T_{dur}}$, $\hat{\theta}$

Begin :

for $i = 1$ to N_b **do**

$ss_i = beatss(i)$

$\left[\hat{\mathrm{R}}, \mathrm{R_p}\right] = \mathbf{max}(ss_i)$

$[\mathrm{Q}' \quad \mathrm{S}'] = \mathbf{IntervalQRS}(s_i, Interval)$

$\left[\hat{\mathrm{Q}}, \mathrm{Q_p}\right] = \mathbf{min}(ss_i(\mathrm{Q}' : \mathrm{R_p}))$

$\left[\hat{\mathrm{S}}, \mathrm{S_p}\right] = \mathbf{min}(ss_i(\mathrm{R_p} : \mathrm{S}'))$

$ss_{new} = \mathbf{suppress}(ss_i(\mathrm{Q}' : \mathrm{S}'))$

$\left[\hat{\mathrm{P}}, \mathrm{P_p}\right] = \mathbf{max}(ss_{new}(1 : \mathrm{Q}'))$

$\left[\hat{\mathrm{T}}, \mathrm{T_p}\right] = \mathbf{max}(ss_{new}(\mathrm{S}' : \mathbf{length}(ss_{new})))$

$\mathrm{P}_1 = ss_i(1 : \mathrm{P_p})$

$\mathrm{P}_2 = ss_i(\mathrm{P_p} : \mathrm{Q}')$

$\left[\mathrm{P_{onset}}, \mathrm{P_p^{on}}\right] = \mathbf{max}(\mathbf{diff}(\mathrm{P}_1))$

$\left[\mathrm{P_{offset}}, \mathrm{P_p^{off}}\right] = \mathbf{min}(\mathbf{diff}(\mathrm{P}_2))$

$\mathrm{T}_1 = ss_i(S' : \mathrm{T_p})$

$\mathrm{T}_2 = ss_i(\mathrm{T_p} : \mathbf{length}(ss_i))$

$\left[\mathrm{T_{onset}}, \mathrm{T_p^{on}}\right] = \mathbf{max}(\mathbf{diff}(\mathrm{T}_1))$

$\left[\mathrm{T_{offset}}, \mathrm{T_p^{off}}\right] = \mathbf{min}(\mathbf{diff}(\mathrm{T}_2))$

$Baseline(1 : \mathbf{length}(ss_i)) = \mathrm{P_{offset}}(i)$

$\mathrm{P_{amp}}(i) = \hat{\mathrm{P}} - Baseline(i)$

$\mathrm{QRS_{amp}}(i) = \hat{\mathrm{R}} - Baseline(i)$

$\mathrm{T_{amp}}(i) = \hat{\mathrm{T}} - Baseline(i)$

$\mathrm{P_{dur}}(i) = (\mathrm{P_p^{off}} - \mathrm{P_p^{on}})/fs$

$\mathrm{QRS_{dur}}(i) = (\mathrm{S_p} - \mathrm{Q_p})/fs$

$\mathrm{T_{dur}}(i) = (\mathrm{T_p^{off}} - \mathrm{T_p^{on}})/fs$

$\mathrm{S^*}_\mathrm{p} = \mathrm{S_p} + 1$

$\mathrm{S}^* = ss_i(\mathrm{S^*}_\mathrm{p})$

$[\mathbf{a}, \mathbf{b}] = \mathbf{vector}(\mathrm{S_p}, \hat{\mathrm{S}}, \mathrm{S^*}_\mathrm{p}, \mathrm{S}^*)$

$\hat{\theta}(i) = \mathbf{arcos}(\mathbf{a}, \mathbf{b})$

end for

3.2 Testing of algorithms for estimating N_{opt} and denoising algorithm

To test Algorithm 2 experimentally, we selected healthy heartbeats with 301 samples and estimated errors by allowing $N_{\min} \leq N \leq 10^3$ for $l = 1$ (Fig. 3.3a), $l = 2$ (Fig. 3.3b), $l = 3$ (Fig. 3.3c), and $l = 4$ (Fig. 3.3d).

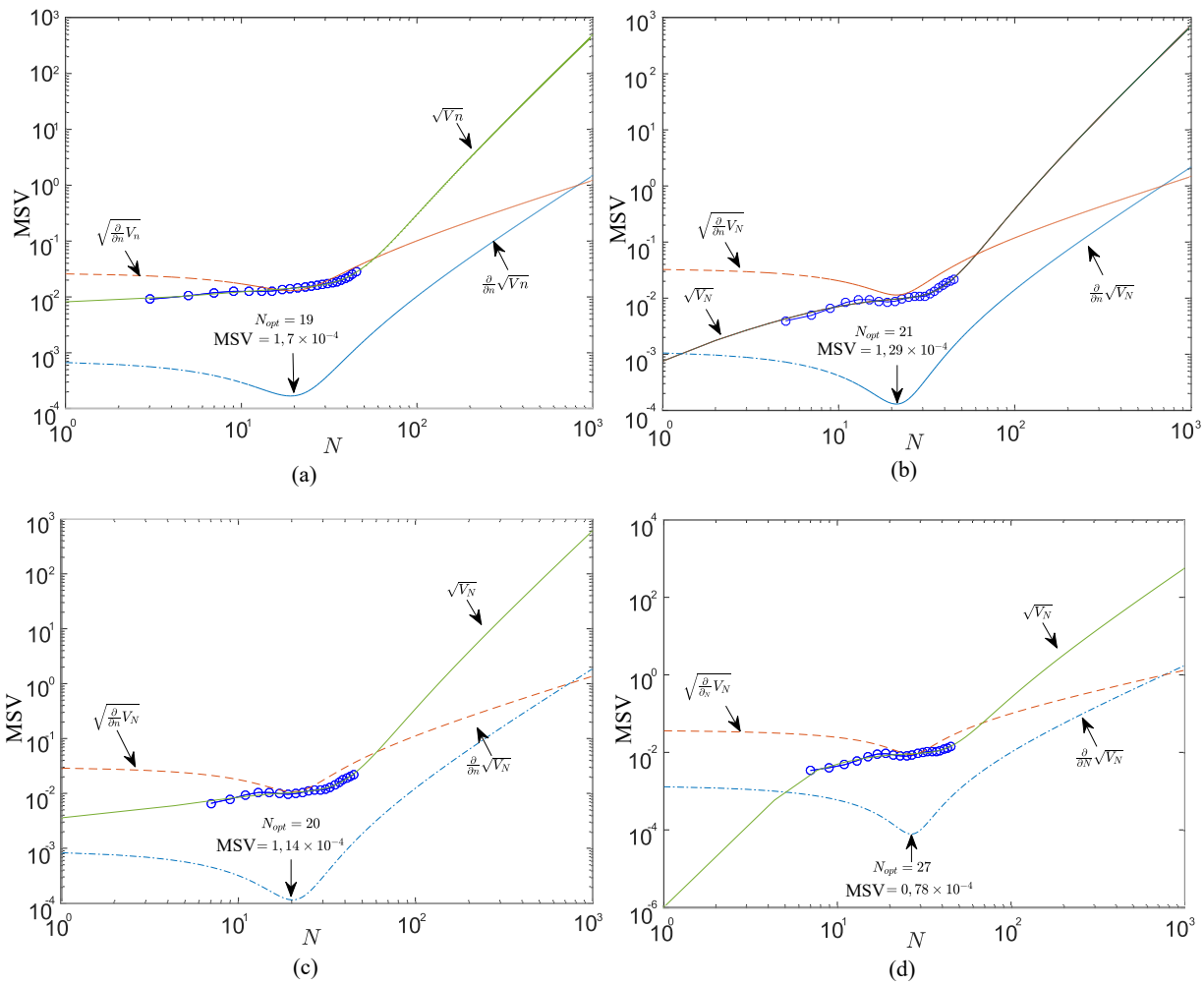

Figure 3.3: Effect of N on the MSV with a) $l = 1$, b) $l = 2$, c) $l = 3$ and d) $l = 4$: the MSV is circled, $\sqrt{V_N}$ is a cubic approximation of the MSV, and $\frac{\partial}{\partial N}\sqrt{V_N}$ and $\sqrt{\frac{\partial}{\partial N}V_N}$ are the derivatives of $\sqrt{V_N}$. The optimal horizon $N_{\text{opt}} = 19$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N}V_N}$.

As can be seen, V_n behaves similarly for different degrees l. It can also be observed that N_{opt} generally grows with l and elevates to $N_{\text{opt}} = 27$ when $l = 4$. Particularly in the algorithm

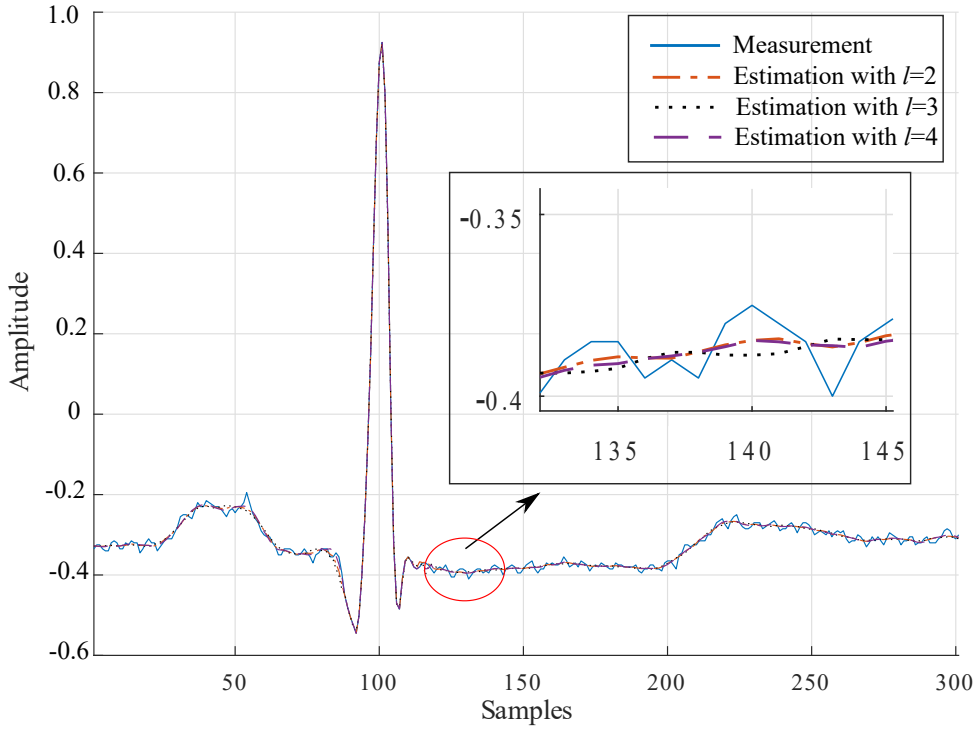

Figure 3.4: Effect of the UFIR filter degree l on the estimation accuracy.

3, an analysis of estimation errors produced by the 2-degree and 3-degree UFIR filters reveals no significant differences, except for the horizon length, which inherently grows with l. This is explained by the fact that $p = \frac{-(N-1)}{2}$ makes the noise power gain (NPG) of both filters equal [69]. The role of p on the smoothing filter NPG has been studied by Shmaliy *et. al.* in [69]. However, choosing $l = 2$ reduces the computational complexity, while saving the estimation accuracy, and we accept $l = 2$ as near optimal. Effect of l on the estimation accuracy is illustrated in Fig. 3.3.

3.2.1 Critical evaluation of denoising algorithms

In Fig 3.4a, we illustrate typical denoising errors produced by the predictive filter, filter, and smoothing filter, all having batch structures. A part of the ECG signal taken from [120:200] is zoomed in Fig. 3.4b. The denoising errors are sketched in Fig. 3.5.

As can be seen, all UFIR filters are successful in denoising with consistent errors. Even so, the UFIR smoothing filter does it more precisely while the predictive filter produces more errors. The median of errors produced by the algorithms and represented with the dispersion are listed in Fig. 3.6. This figure suggests that the UFIR smoothing filter outperforms both the UFIR filter and the standard linear predictor developed in [32] for ECG signals. An analysis of the

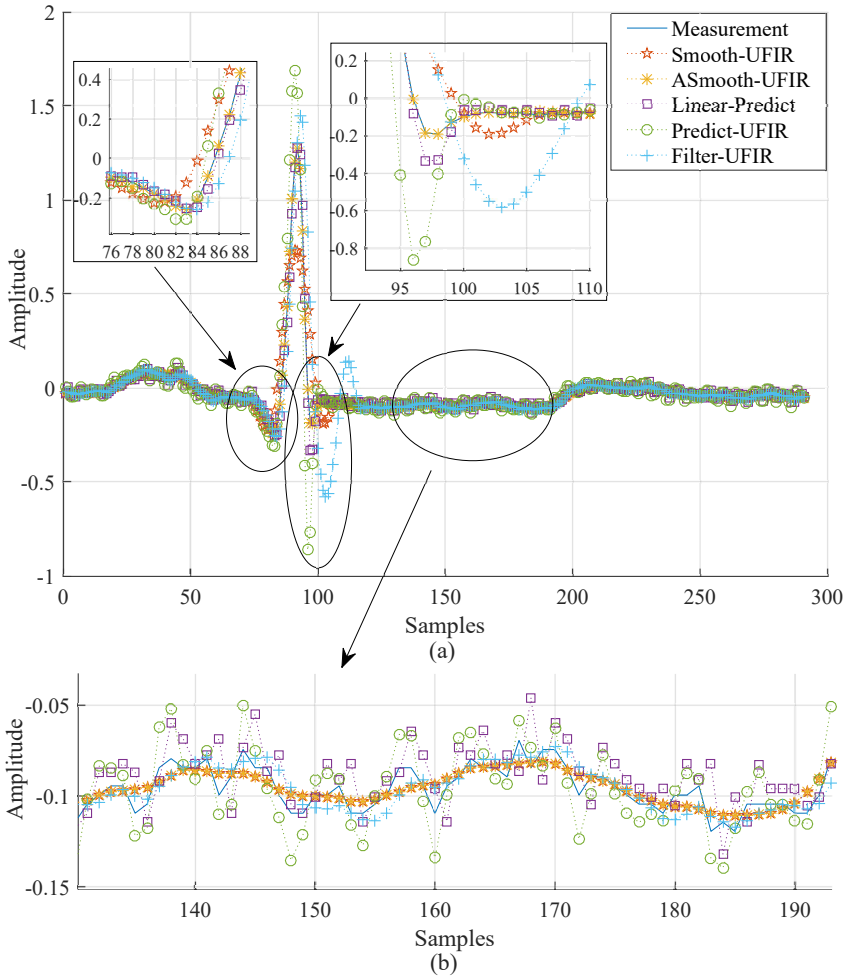

Figure 3.5: Denoising of ECG signals: (a) heartbeat estimation with different methods such as ASmooth-UFIR (UFIR Adaptive-Smoothing filter), Linear Predict, Predict-UFIR (UFIR predictive filter), and Filter-UFIR (UFIR filter), and (b) segmental visualization of five estimates.

signal-to-noise rations (SNRs) at the filters outputs will be provided next.

3.2.2 Effect of SNR on the estimator MSE

The root MSEs (RMSEs) are shown in Fig. 3.7 as functions of the SNR depicted in decibels (dB) at 18 discrete points with a step of 5dB. It follows that the UFIR smoothing filter outperforms other solutions in a wide range of SNR values. For $0 \leqslant \text{SNR} < 15\,\text{dB}$, higher accuracy is achieved with a constant N and, for $\text{SNR} > 15\,\text{dB}$, with an adaptive N.

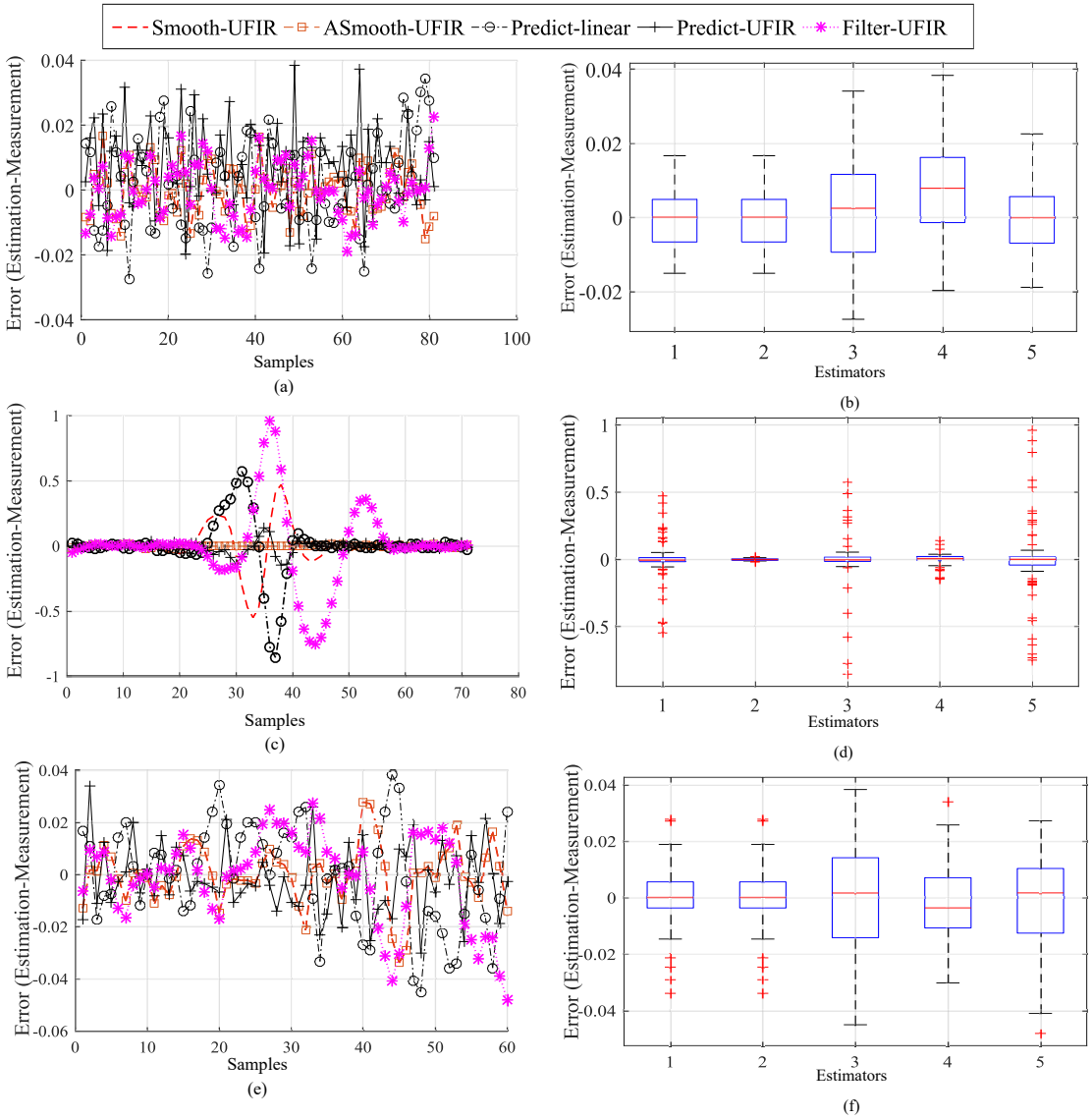

Figure 3.6: Errors produced by different estimators: (a) in the T-wave, (b) error boxplot in the T-wave, (c) in the QRS-complex, (d) error boxplot in the QRS-complex, (e) in the P-wave, and (f) error boxplot in the P-wave. Estimator 1 is Smooth UFIR, estimator 2 is ASmooth-UFIR, estimator 3 is the predict linear, estimator 4 is the Predict-UFIR and finally, estimator 5 is the Filter-UFIR

3.3 Applications to ECG signals

Based upon the above developed UFIR-based approach, we now apply Algorithm 3 to the ECG signal database and extract special features depicted in Fig. 3.1. The results obtained using the designed UFIR smoothing Algorithm 2 (UfirSmooth), UFIR predictive algorithm (Predictor

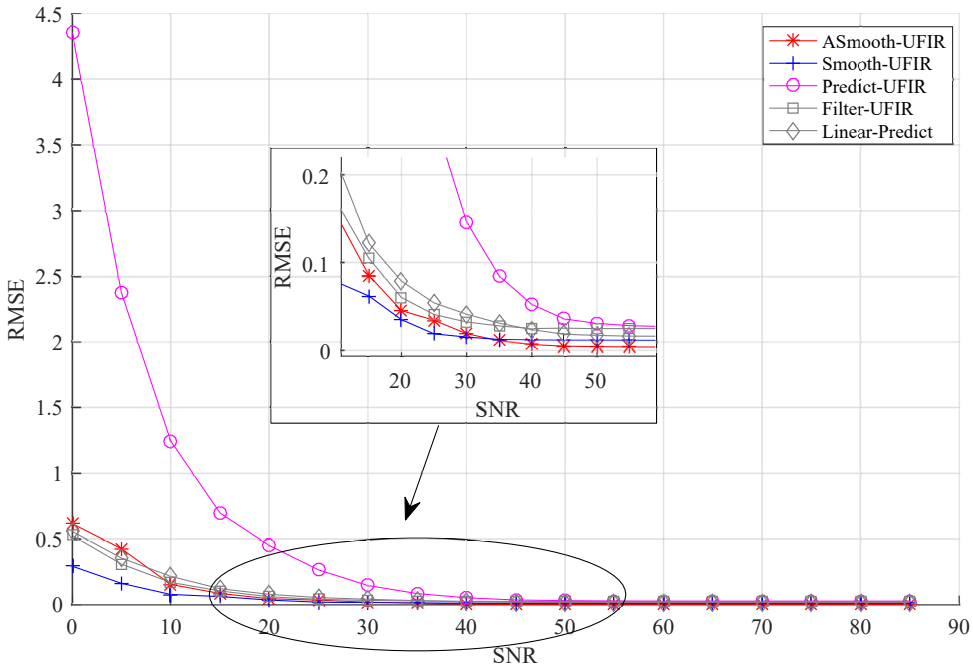

Figure 3.7: RMSEs of UFIR denoising estimators and linear predictor (Linear-Predict) as functions of SNR

Ufir), and basic linear predictor (Linear Predict) [1] are sketched in Fig. 3.8 and Fig. 3.9. In these figures, 100 synthetic heartbeats are processed at each time index. This synthetic ECG signal is contaminated by AWGN at 35 dB with properties similar to the original data.

In Fig. 3.10, we show dispersions and concentrations of the estimated features about their means. Shadowed areas represent features extracted by smoothing and it follows that the outputs of the filter and linear predictor are more vulnerable. Furthermore, noise dominates in the predictive filters outputs. This experiment was based on healthy records of MIT-Arrhythmia database (lead MLII) analysing 1000 heartbeats. Overall, the UFIR smoothing approach developed in this work always produced better estimates than by other linear methods considered.

3.3.1 APC heartbeats UFIR smoothing and P-wave features Analysis using Rice Distribution

Heart diseases are one of 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 work, we provide smoothing of the atrial premature complex (APC) of the electrocardiogram (ECG) signal using unbiased finite impulse response (UFIR) smoothing filtering. We investigate

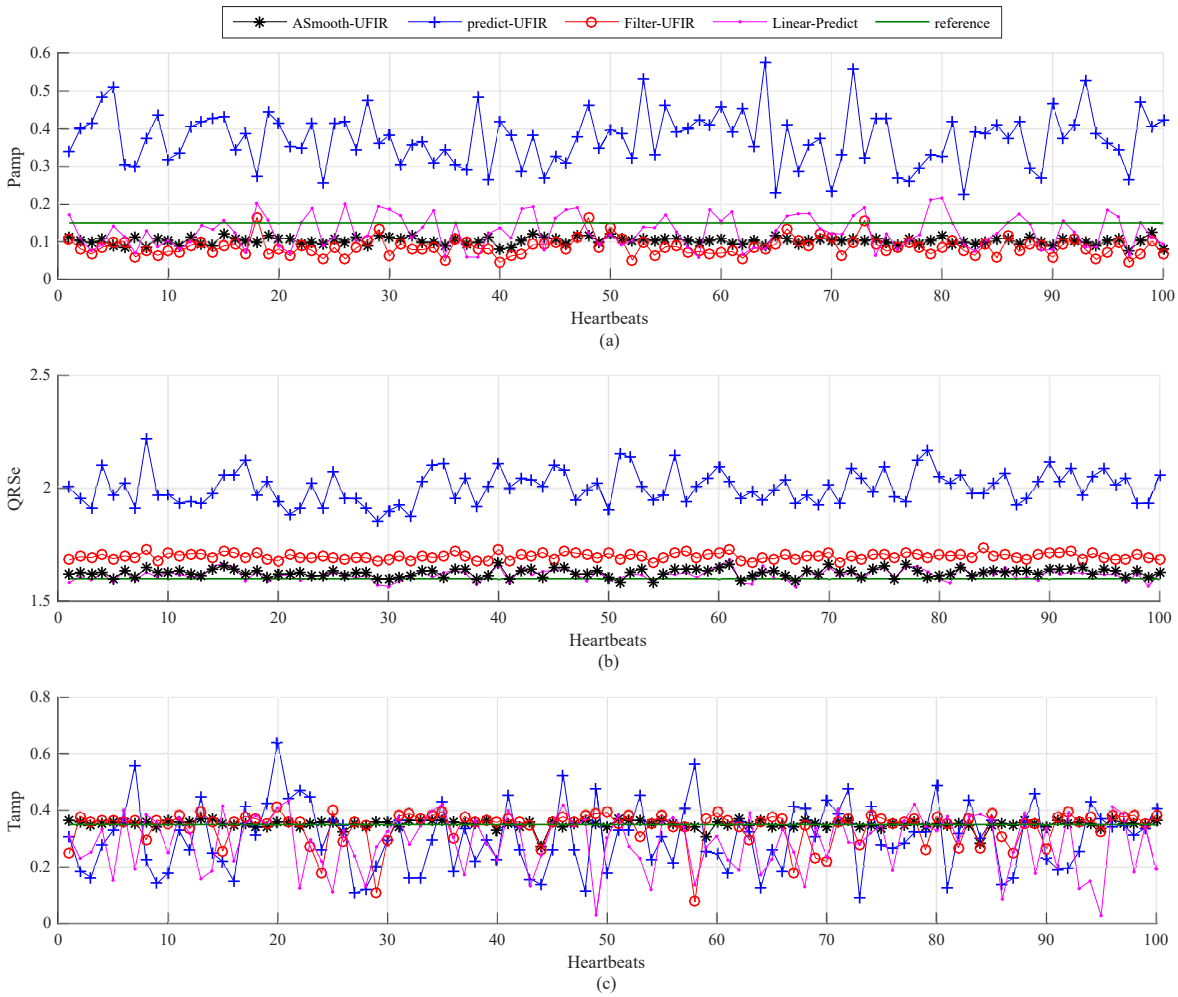

Figure 3.8: Features of the ECG signal extracted using the UFIR smoothing filter (Smooth-UFIR), basic linear predictor (Linear Predict) [1], and UFIR predictive filter (Predict UFIR): (a) P_{amp}, (b) QRS_e, and (c) T_{amp}.

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 the UFIR smoothing filter has better performance than the others ones. 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.

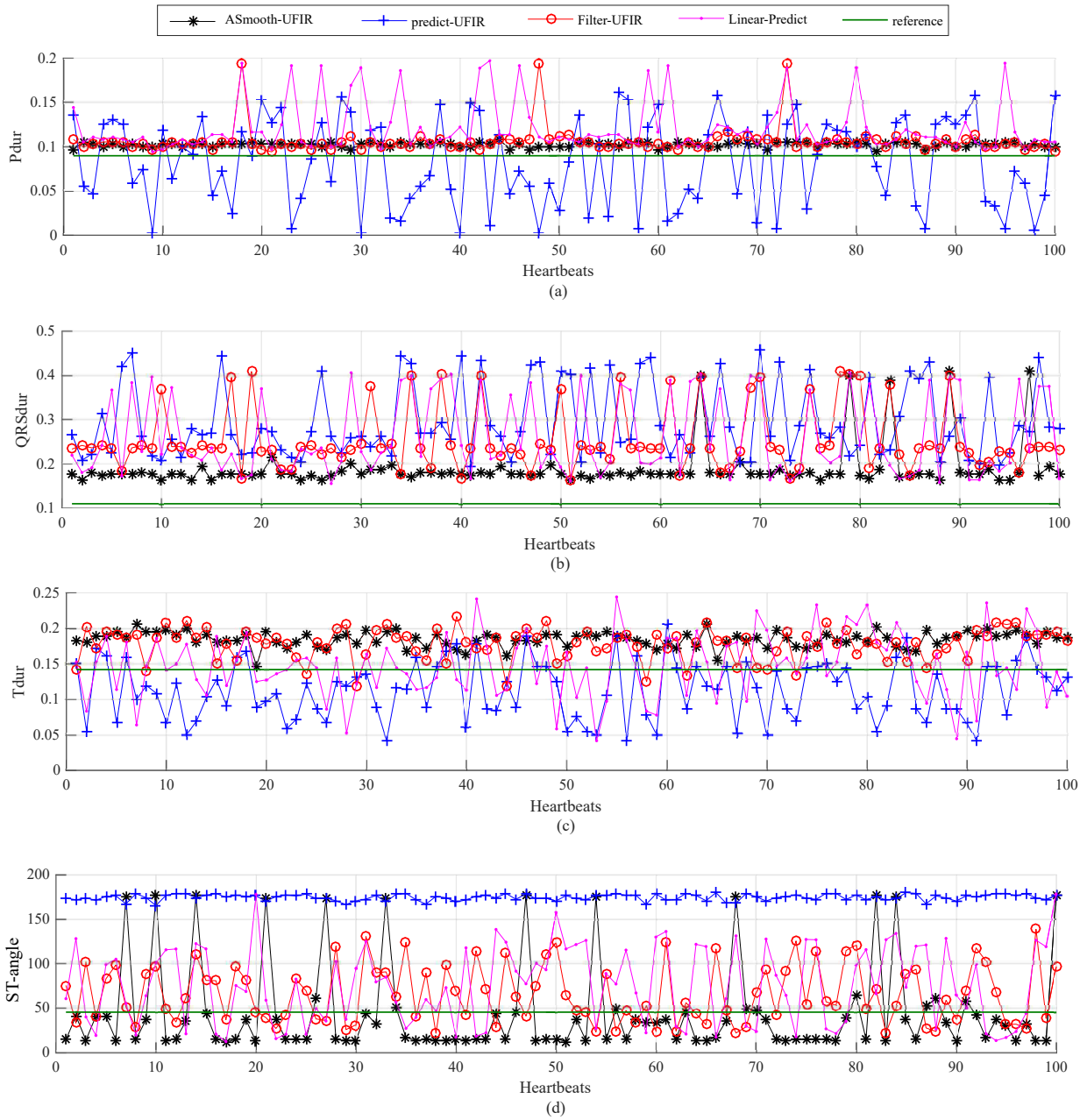

Figure 3.9: Features of the ECG signal extracted using the UFIR smoothing filter (ASmooth-UFIR), basic linear predictor (Linear Predict) [1], and UFIR predictive filter (Predict Ufir): (a) P_{dur}, (b) QRS_{dur}, (c) T_{dur}, and (d) ST_{angle} $\hat{\theta}$. The durations are sampled with 0.0028 seconds. Also ST-angle is referenced with 45 grades

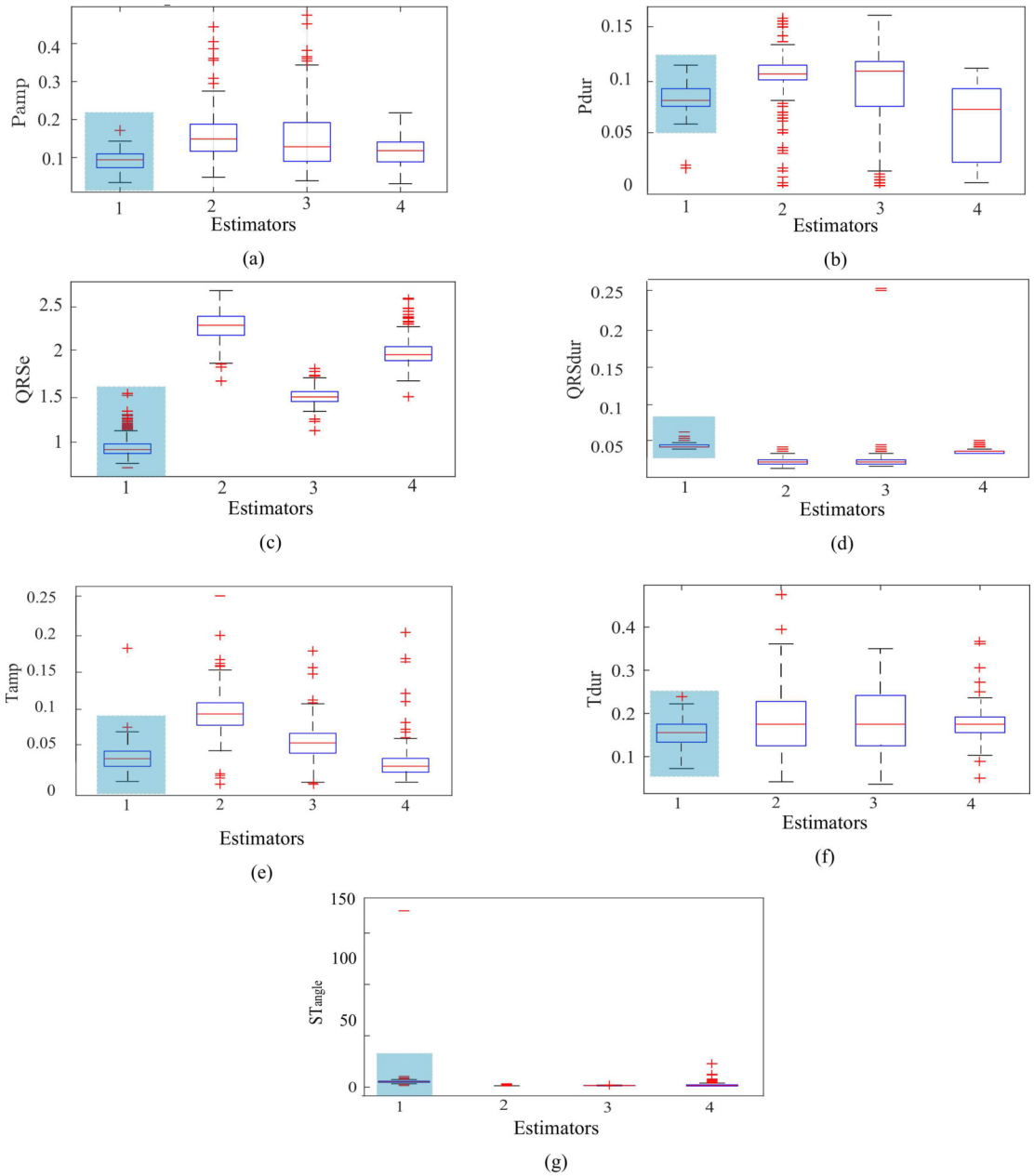

Figure 3.10: Boxplot of features of the ECG signal extracted using the UFIR adaptive smoothing filter (Estimator 1: ASmooth-UFIR), UFIR predictive filter (Estimator 2: Predict Ufir), basic linear predictor (Estimator 3: Linear Predict), UFIR filter (Estimator 4: Filter-UFIR): (a) P_{amp}, (b) P_{dur}, (c) QRS_e, and (d) QRS_{dur},(e)T_{amp}, (f) T_{dur} and (g) ST_{angle} $\hat{\theta}$.

3.3.2 P-Wave detection algorithm

A pseudo code of the algorithm designed for the ECG signal features extraction is shown as Algorithm 3.3.2. Here, ss_i is the smoothed ECG signal; N_b is the number of heartbeats; *Baseline*

is a variable, which 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 $\hat{\mathrm{P}}$ of P, $\mathrm{P_{amp}}$ of the P amplitude, $\mathrm{P_{dur}}$ of the P duration. All these features are extracted from the smoothed signal.

Algorithm 5 A pseudo code of the algorithm to extract morphological features of P-wave

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

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

Begin :
for $i = 1$ to N_b **do**
  $ss_i = beatss(i)$
  $\left[\hat{\mathrm{R}}, \mathrm{R_p}\right] = \mathbf{max}(ss_i)$
  $[\mathrm{Q}' \quad \mathrm{S}'] = \mathbf{IntervalQRS}(s_i, Interval)$
  $\left[\hat{\mathrm{Q}}, \mathrm{Q_p}\right] = \mathbf{min}(ss_i(\mathrm{Q}' : \mathrm{R_p}))$
  $\left[\hat{\mathrm{S}}, \mathrm{S_p}\right] = \mathbf{min}(ss_i(\mathrm{R_p} : \mathrm{S}'))$
  $ss_{new} = \mathbf{suppress}(ss_i(\mathrm{Q}' : \mathrm{S}'))$
  $\left[\hat{\mathrm{P}}, \mathrm{P_p}\right] = \mathbf{max}(ss_{new}(1 : \mathrm{Q}'))$
  $\mathrm{P_1} = ss_i(1 : \mathrm{P_p})$
  $\mathrm{P_2} = ss_i(\mathrm{P_p} : \mathrm{Q}')$
  $\left[\mathrm{P_{onset}}, \mathrm{P_p^{on}}\right] = \mathbf{max}(\mathbf{diff}(\mathrm{P_1}))$
  $\left[\mathrm{P_{offset}}, \mathrm{P_p^{off}}\right] = \mathbf{min}(\mathbf{diff}(\mathrm{P_2}))$
  $Baseline(1{:}\mathbf{length}(ss_i)) = \mathrm{P_{offset}}(i)$
  $\mathrm{P_{amp}}(i) = \hat{\mathrm{P}} - Baseline(i)$
  $\mathrm{P_{dur}}(i) = (\mathrm{P_p^{off}} - \mathrm{P_p^{on}})/fs$
end for

The algorithm begins with computing $\hat{\mathrm{R}}$ as the ECG signal maximum, using function **max**. Function **IntervalQRS** is applied to compute Q' and S'. The *Interval* variable determines the window width to cover the QRS complex and obtain $\hat{\mathrm{Q}}$ and $\hat{\mathrm{S}}$ as two minima between Q' and S'. Function **min** is used to find the above-mentioned points. The **supress** function is used to suppress the QRS complex. Function **max** is used to estimate P. Function **diff** is introduced to compute the derivatives in the $\mathrm{P_1}$, $\mathrm{P_2}$ intervals. Functions **max** and **min** with **diff** are used to find $\mathrm{P_{onset}}$, $\mathrm{P_p^{on}}$, $\mathrm{P_{offset}}$, and $\mathrm{P_p^{off}}$. Provided these values, the duration is estimated of P features. Function **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_{offset}. As can see in the Fig. 3.11, the estimates provided by the UFIR smoothing filter are more consistent to the average P-wave than by other techniques.

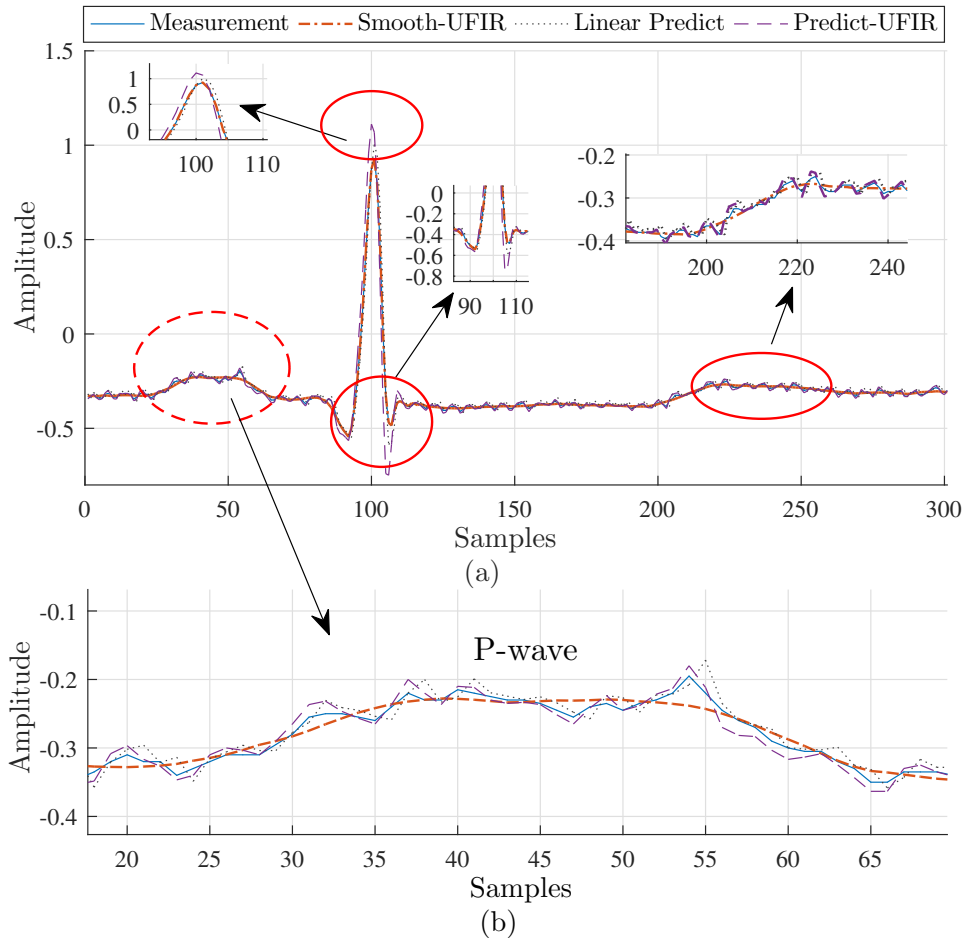

Figure 3.11: Applications of different filters to extract the ECG amplitude features (a) Amplitud of P-wave and (b) duration of P-wave. Here Smooth-UFIR is the UFIR smoothing filter, Predict UFIR is the linear predictor and Predict-UFIR is the UFIR predictive filter

3.3.3 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 by using the Rice probability density function (pdf), which corresponds to the envelop of a harmonic signal in Gaussian noise [96, 97],

$$p(r) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2+A}{2\sigma^2}\right) I_0\left(\frac{Ar}{\sigma^2}\right) , \tag{3.22}$$

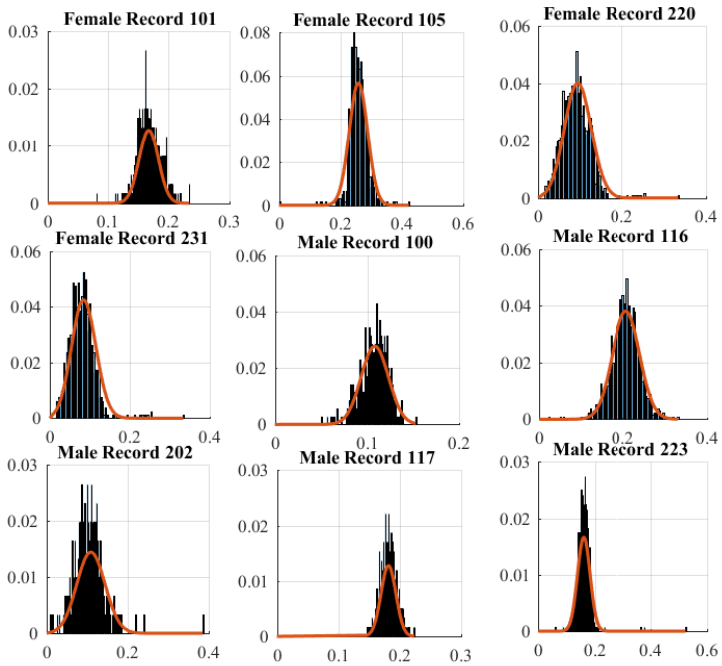

Figure 3.12: Normalized histograms and Rice-based approximations for P-wave features of healthy heartbeats.

where σ^2 is the variance of the 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

$$p(v) = v \ \mathtt{exp}\left(-\frac{v^2 + a^2}{2}\right) I_0(Av)\,, \tag{3.23}$$

where $v = \frac{r}{\sigma}$ and $a = \frac{A}{\sigma}$. Note that (3.23) reduces to the Rayleigh distribution [98] when $a = 0$ and, by large a, it becomes Gaussian.

3.3.4 Confidence interval for P-Wave

To specify the confidence interval for the P-wave estimates provided by Algorithm 3.3.2, we have investigated the P-wave histograms for normal and abnormal heartbeats taken from different persons as shown in Fig. 3.3.4.

We employed a register of different ages between female and male gender with 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

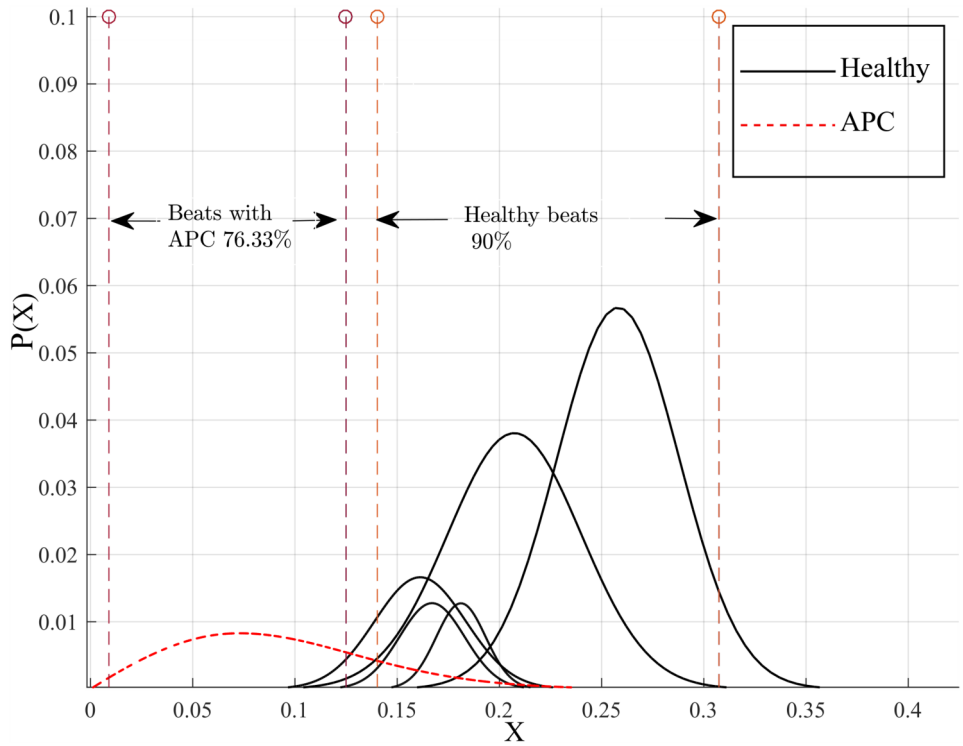

Figure 3.13: Confidence intervals for APC (sold) and healthy (dashed) ECG heartbeats.

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.

Confidence Interval for Normal Heartbeats

As can seen in the Fig. 3.3.4, 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.

3.4 Discussion

The purpose of this study is denoising the attached noise in ECG signals using a UFIR smoothing filter for features extraction. This work is focused in the morphological features extraction individual ECG signal processing with normal rhythm. A principal findings in the applying the proposed method is the considerably reduction of noise with an optimum and adaptive horizon for real ECG data. This reduction contributes a determine with better precision the features associated to the heartbeat.

From analysis of errors variability in real ECG signals and SNRs based on ECG synthetic data in different estimators has shown that the UFIR smoothing filter with adaptive horizon outperforms the linear predictor between 0 and 50 of SNR with RMSE not more than 0.5 [1, 30–32] and other UFIR solutions such as the UFIR filter and UFIR predictive filter on MIT-BIH arrhythmia dataset. Let us notice again that the approaches based on linear prediction were recognized as standard for the ECG signal features extraction [32]. In this regard, better performance of the smoothing algorithm developed in this work opens new horizons in achieving higher accuracy and reliability in detecting different kinds of heart diseases.

The UFIR smoothing filter performance was optimized by making the averaging horizon adaptive. Note that such an opportunity has not been used in the design of known linear predictors for ECG data. As a results, we have achieved the following improvements:

1. Suboptimal denoising of ECG signals with no requirements to noise, except for the zero mean assumption.

2. Unbiased filtering in the QRS region, in which the ECG signal demonstrates rapid excursions.

Such abilities of the UFIR smoothing filter have resulted in higher estimation accuracy, namely in smaller variability of the estimated features around their mean values. In this regard, let us notice that larger variability in the standard linear predictor is due to larger errors and instability caused by unknown future data and errors in the determination of the predictor coefficients determined by the correlation method. Accordingly, errors in the determination of the prediction function lead to larger prediction errors (random and regular). This has appeared to be particularly true for the P_{amp} and T_{amp} values, which are estimated by other methods with much larger errors. Estimates of QRS_e and QRS_{dur} by different methods have appeared to be consistent, because these values are not affected by noise as much as other features. Nevertheless, the UFIR smoothing has demonstrated smaller errors even for QRS_{amp}. In the cases of both T_{dur} and angle θ, one watches for highly unstable estimates provided by the prediction-based filters, while the proposed UFIR smoothing filter has produced acceptable estimates. Also, it is important clarify that the evaluation of features are analysed from the consistence of data near of average of the measurement. However, in this scenario, the quality features is no strictly analysed because the ECG signal used is just under normal conditions. We consider that estimate given by the UFIR approach is satisfactory but it is not enough in the sense to extract better features. Because we

have employed a external function that provides the derivative of the estimation with attached noise and reduced length. This is a limitation to determine with precision the features. Hence, in the following chapter, we propose a UFIR smoothing in the state-space.

Chapter 4

Denoising and Features Extraction of ECG signals in State Space Using Unbiased FIR Smoothing

Referring to the first results obtained in [3, 92], where the batch UFIR smoothing filter has demonstrated a better performance than several other well-recognized estimators, in this chapter we have employed and developed an iterative UFIR smoother in state space. The principal objective is to increase accuracy of the features extraction and fiducial points detection.

The main contributions of this work are the following:

- An optimal q-lag state-space UFIR smoothing algorithm for ECG signals denoising and artifacts removal.
- An algorithm for ECG signal stable temporal features extraction using different classifiers under unknown noise.
- High-accuracy patterns classification for ECG signals with atrial fibrillation (AF) and normal conditions.

To reach the goal, we first provide denoising of ECG signals and compare the results obtained by the wavelet-based and some standard filters. We then extract features of the ECG-waves and analyse confidence intervals for particular ECG records. The results are tested by different classifiers and compared to those available from several machine learning techniques. The rest of the chapter is organized as follows. Section 5.1.1 presents the discrete-time state-space UFIR

filtering and smoothing approaches. In Section **??**, we design an adaptive UFIR smoothing algorithm for ECG signal features extraction. Specifics of the UFIR smoother optimal tuning and testing are given in Section **??**. Assessment of adaptive UFIR smoothing algorithm for ECG signals to ECG signals are given in Section 4.4. Applications for T-wave Features Analysis are given in Section **??**

Discussion of the results is provided in Section 4.5.

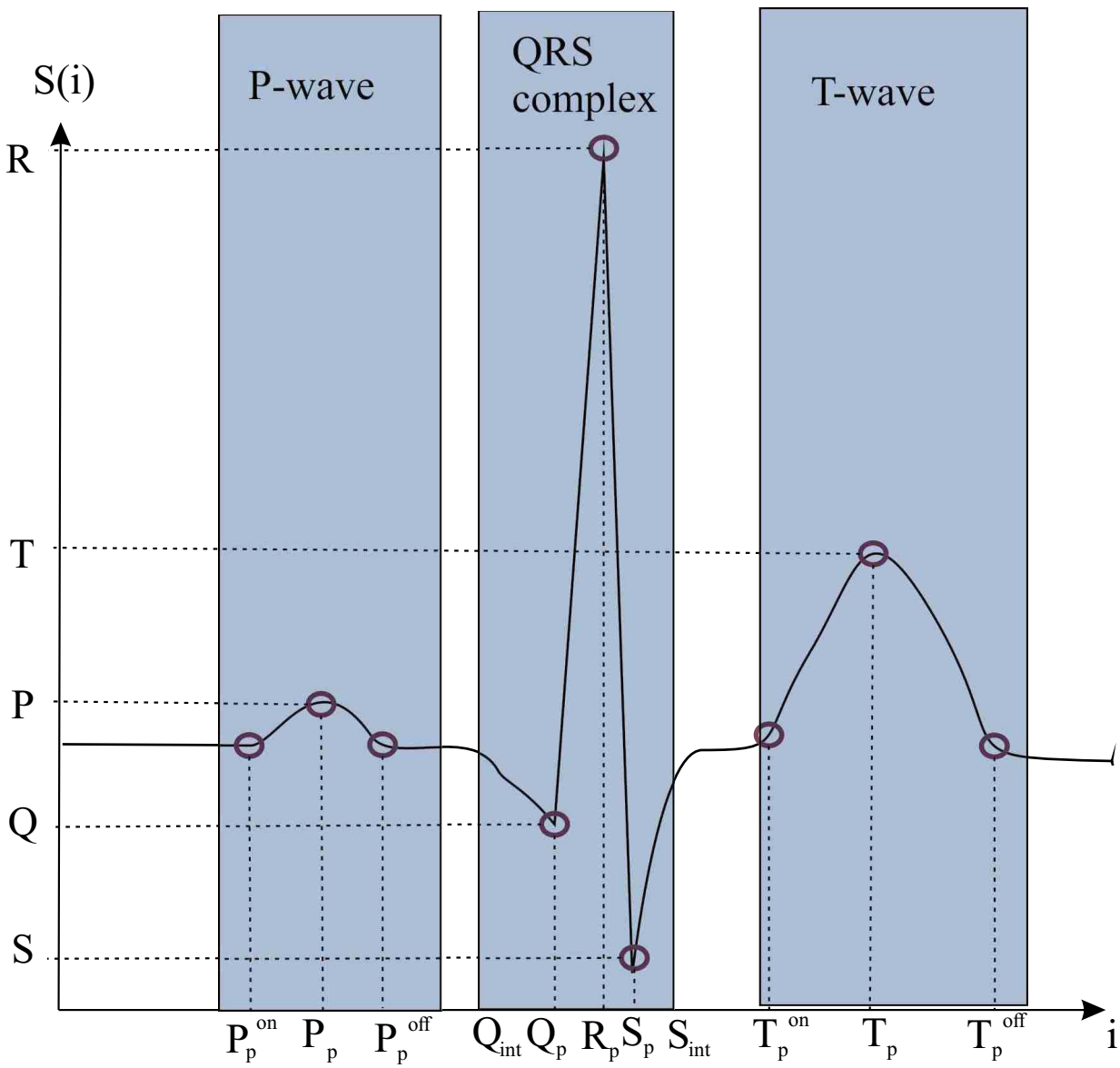

Figure 4.1: Fiducial points of heartbeat

4.1 ECG signal model in discrete-time state-space

To provide efficient denoising and features extractions, in this subsection we model an ECG signal upon fiducial points (see Fig. 4.1) in discrete-time state-space. We represent an ECG signal on a horizon $[m, n]$ of N points, from $m = n - N + 1$ to n, where n is the discrete time index, with a degree polynomial as shown in [3]. The inherent ECG noise is still not well understood and its incorrect description may cause estimation errors. Therefore, we suppose that the underlying process in each ECG pulse is time-invariant and deterministic. We also suppose that scalar measurements of the ECG signal are provided in the presence of zero mean noise having an unknown distribution (not obligatorily Gaussian) and covariance.

Under such assumptions, we represent an ECG signal in discrete-time state-space with the following state and observation equations, respectively,

$$\mathbf{x}_n = \mathbf{F}\mathbf{x}_{n-1}\,, \tag{4.1}$$

$$y_n = \mathbf{H}\mathbf{x}_n + v_n\,, \tag{4.2}$$

where $\mathbf{x}_n \in \mathbb{R}^K$ is the ECG process state vector, y_n is the scalar observation, v_n is the scalar measurement noise, $\mathbf{F} \in \mathbb{R}^{K\times K}$ is the system matrix projecting the initial state $\mathbf{x}_{n-1}$ to $\mathbf{x}_n$ and given by [99]

$$\mathbf{F} = \begin{bmatrix} 1 & \tau & \frac{\tau^2}{2} & \dots & \frac{\tau^{K-1}}{(K-1)!} \\ 0 & 1 & \tau & \dots & \frac{\tau^{K-2}}{(K-2)!} \\ 0 & 0 & 1 & \dots & \frac{\tau^{K-3}}{(K-3)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \tag{4.3}$$

For a scalar measurement, we assign the observation matrix as $\mathbf{H} = [1\ 0 \cdots 0] \in \mathbb{R}^{1\times K}$ and suppose that noise v_n is zero mean with unknown distribution and other statistics. The batch UFIR filter can now be applied to (4.1) and (4.2) to provide state estimates as in the following.

4.1.1 UFIR filtering and smoothing of ECG signals

Provided modeling of an ECG signal in discrete-time state space, in this section we discuss the UFIR filter and smoother first in the batch form and then in a fast iterative form using recursions. Because the optimal averaging horizon is shape-varying for ECG signals, we also discuss its adaptive structure.

4.1.2 Batch UFIR filter and smoother

On a horizon $[m, n]$ of N ECG data points, the batch UFIR filtering estimate $\mathbf{x}_n \triangleq \hat{\mathbf{x}}_{n|n}$ of $\mathbf{x}_n$ is given by [72]

$$\hat{\mathbf{x}}_n = (\mathbf{W}_{m,n}^T \mathbf{W}_{m,n})^{-1} \mathbf{W}_{m,n}^T \mathbf{Y}_{m,n} \,, \tag{4.4}$$

where the extended observation vector (This model is described in Appendix B) $\mathbf{Y}_{m,n}$ and augmented measurement matrix $\mathbf{W}_{m,n}$ are, respectively,

$$\mathbf{Y}_{m,n} = [\, y_m^T \; y_{m+1}^T \ldots \; y_n^T \,]^T \,, \tag{4.5}$$

$$\mathbf{W}_{m,n} = \begin{bmatrix} \mathbf{H}(\mathbf{F}^{n-m})^{-1} \\ \vdots \\ \mathbf{H}\mathbf{F}^{-1} \\ \mathbf{H} \end{bmatrix} . \tag{4.6}$$

In the discrete convolution-based form, estimate (5.4) can be represented as

$$\hat{\mathbf{x}}_n = \mathcal{H}_{m,n} \mathbf{Y}_{m,n} \,, \tag{4.7}$$

where the UFIR filter gain matrix $\mathcal{H}_{m,n}$ given by

$$\mathcal{H}_{m,n} = (\mathbf{W}_{m,n}^T \mathbf{W}_{m,n})^{-1} \mathbf{W}_{m,n}^T \tag{4.8}$$

can be rewritten as

$$\mathcal{H}_{m,n} = \mathbf{G}_n \mathbf{W}_{m,n}^T \,, \tag{4.9}$$

where $\mathbf{G}_n$ is the generalized noise power gain (GNPG),

$$\mathbf{G}_n = \mathcal{H}_{m,n} \mathcal{H}_{m,n}^T = (\mathbf{W}_{m,n}^T \mathbf{W}_{m,n})^{-1} \,. \tag{4.10}$$

Given the UFIR filtering estimate $\hat{\mathbf{x}}_n \triangleq \hat{\mathbf{x}}_{n|n}$ of $\mathbf{x}_n$ by (4.7), the q-lag UFIR smoothing estimate can be obtained by projecting $\hat{\mathbf{x}}_n$ into $\hat{\mathbf{x}}_{n-q}$ as shown in [69],

$$\hat{\mathbf{x}}_{n-q|n} = \mathbf{F}^{-q} \hat{\mathbf{x}}_{n|n} \,, \tag{4.11}$$

where $q_{\text{opt}} = \lfloor \frac{N}{2} \rfloor$ is a digital optimal lag for odd-degree UFIR smoothers and q_{opt} must be set individually following Fig. 8 in [69] for each even-degree. Let us notice again that the Savitsky-Golay solution ignores this specific and suggests taking lags from the middle points of $[m, n]$ for all degrees that introduces errors.

4.1.3 Adapted optimal horizon N_{apt}

Of importance is that the UFIR filter is able to minimize the MSE on $[m, n]$, if the horizon N is set optimally as N_{opt} [93]. To make it possible in the absence of the reference signal (ground truth), we follow [93] and find N_{opt} for ECG signals by minimizing the trace of the derivative of the mean square value (MSV) of the measurement residual matrix $\mathbf{V}(N)$ as

$$\hat{N}_{\text{opt}} = \arg\min_{N} \frac{\partial \operatorname{tr} \mathbf{V}(N)}{\partial N} + 1\,. \tag{4.12}$$

A solution to the optimization problem (5.12) has been provided in our early paper together with an algorithm [3], which we will employ further. It has been found out in [3] that an optimal horizon $N_{\text{opt}} = 21$ serves for the 2-degree polynomial corresponding to three states, $K = 3$, and database [76] exploited in this paper.

An important specific is that N_{opt} varies on different parts of the ECG signals [92]. Therefore, we will make N_{opt} adaptive (N_{apt}) to range from $N_{\text{min}} = K = 3$ to N_{opt} as

$$N_{\text{min}} \leqslant N_{\text{apt}} \leqslant N_{\text{opt}}\,,$$

where N_{min} is a minimum horizon applied to a fast excursion between Q_{p} and S_{p} (Fig. 4.1). To this end, we recognize five parts in the ECG signal separated with the following points in Fig. 4.2: Q_{int}, Q_{p}, S_{p}, and S_{int}. Up to Q_{int}, a smooth part of the ECG signal is processed with N_{opt}. Between Q_{int} and Q_{p}, the horizon N_{apt} linearly reduces from N_{opt} to N_{min}. The QRS complex, between Q_{p} and S_{p}, is processed with N_{min} to follow exactly a fast excursion around R_{p}. From S_{p} to S_{int}, the horizon N_{apt} linearly increases from N_{min} to N_{opt}, The horizon finally becomes N_{opt} above S_{int}. Accordingly, adaptive UFIR smoothing is provided as

$$\hat{\mathbf{x}}_{n-q|n} = \begin{cases} \hat{\mathbf{x}}_{n-q|n}(N_{\text{opt}})\,, & 1 \leqslant n \leqslant \text{Q}_{\text{int}} - 1\,, \\ \hat{\mathbf{x}}_{n-q|n}(N_{\text{apt}})\,, & \text{Q}_{\text{int}} \leqslant n \leqslant \text{Q}_{\text{p}} - 1\,, \\ \hat{\mathbf{x}}_{n-q|n}(N_{\text{min}})\,, & \text{Q}_{\text{p}} \leqslant n \leqslant \text{S}_{\text{p}}\,, \\ \hat{\mathbf{x}}_{n-q|n}(N_{\text{apt}})\,, & \text{S}_{\text{p}} + 1 \leqslant n \leqslant \text{S}_{\text{int}}\,, \\ \hat{\mathbf{x}}_{n-q|n}(N_{\text{opt}})\,, & \text{S}_{\text{int}} + 1 \leqslant n \leqslant T\,, \end{cases} \tag{4.13}$$

where T represents the heartbeat length. Provided N_{apt}, we can next design an iterative UFIR smoothing algorithm using recursions, which reduces the computational load.

4.1.4 Iterative UFIR smoothing

Like the Kalman filter (KF), iterative computation of the batch UFIR estimate (5.7) is provided recursively in two phases: *predict* and *update* [72]. In contrast to the KF, the UFIR algorithm

does it with no requirements for the noise statistics and initial values and is thus more suitable for ECG signals in view of generally unknown heartbeat noise.

At the predict phase, the UFIR algorithm computes the prior state estimate $\hat{\mathbf{x}}_n^- = \mathbf{F}\hat{\mathbf{x}}_{n-1}$ and ignores the prior error covariance, unlike the KF. At the update phase, the UFIR algorithm updates the GNPG $\mathbf{G}_n$ as $\mathbf{G}_n = [\mathbf{H}^T\mathbf{H} + (\mathbf{F}\mathbf{G}_{n-1}\mathbf{F}^T)^{-1}]^{-1}$, the measurement residual $z_n = y_n - \mathbf{H}\hat{\mathbf{x}}_n^-$, the bias correction gain $\mathbf{K}_n = \mathbf{G}_n\mathbf{H}^T$, and the state estimate $\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n^- + \mathbf{K}_n z_n$. A pseudo code of the UFIR smoothing algorithm [72] adapted to ECG signals is listed as Algorithm 1.

Algorithm 6 Adaptive Iterative UFIR Smoothing Algorithm for ECG Signals

Data: y_n, $N = N_{\text{apt}}$, $q = q_{\text{opt}}$

Result: $\hat{\mathbf{x}}$

Begin :
for $n = N-1, N, \ldots$ **do**
  $m = n - N + 1$, $s = n - N + K$
  $\mathbf{G}_s = (\mathbf{W}_{m,s}^T\mathbf{W}_{m,s})^{-1}$
  $\tilde{\mathbf{x}}_s = \mathbf{G}_s\mathbf{W}_{m,s}^T\mathbf{Y}_{m,s}$
  for $l = s+1$ to n **do**
    $\tilde{\mathbf{x}}_l^- = \mathbf{F}\tilde{\mathbf{x}}_{l-1}$
    $\mathbf{G}_l = [\mathbf{H}^T\mathbf{H} + (\mathbf{F}\mathbf{G}_{l-1}\mathbf{F}^T)^{-1}]^{-1}$
    $\mathbf{K}_l = \mathbf{G}_l\mathbf{H}^T$
    $\tilde{\mathbf{x}}_l = \tilde{\mathbf{x}}_l^- + \mathbf{K}_l(y_l - \mathbf{H}\tilde{\mathbf{x}}_l^-)$
  end for
  $\hat{\mathbf{x}}_n = \tilde{\mathbf{x}}_n$
  $\hat{\mathbf{x}}_{n-q} = \mathbf{F}^{-q}\hat{\mathbf{x}}_n$
end for

Provided ECG data y_n, adaptive horizon N_{apt}, and optimal lag q_{opt} for a chosen filter degree, Algorithm 1 starts self-computing the initial GNPG G_s and initial state $\tilde{\mathbf{x}}_s$ at s, which corresponds to a short initial horizon of K points. This is required to overcome singularities in the UFIR filter gain on shorter horizons. Estimate $\hat{\mathbf{x}}_n$ at time index n is computed iteratively, using an auxiliary time variable l, which starts with $l = n - N_{\text{apt}} + K + 1$ and finishes when $l = n$. The estimate $\hat{\mathbf{x}}_n$ obtained in such a way minimizes the MSE and is called the optimal UFIR estimate. Provided

$\hat{\mathbf{x}}_n$, the UFIR smoothing estimate with a lag q is obtained by a projection from n to $n-q$ as [72]

$$\hat{\mathbf{x}}_{n-q} = \mathbf{F}^{-q}\hat{\mathbf{x}}_n \tag{4.14}$$

and we notice again that lag q must be set optimally as q_{opt} to reach minimum possible smoothing errors.

4.2 ECG signal features extraction in state space

Features extraction from ECG signals in state space using Algorithm 4.1.4 is provided in five stages (Fig. 4.2): 1) detrending, 2) QRS-complex detection, 3) segmentation, 4) adaptive iterative UFIR smoothing, and 5) windowing of ECG waves.

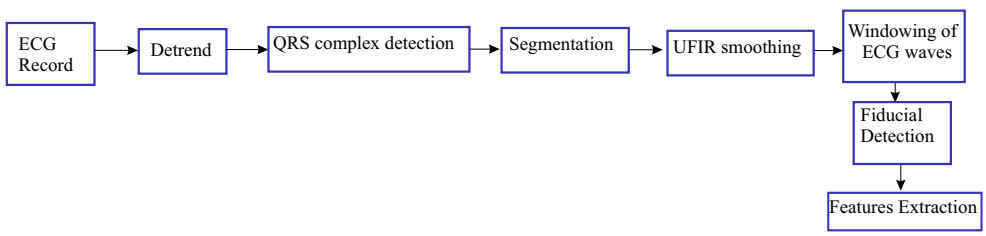

Figure 4.2: Block-diagram of features extraction of the ECG signal in state space using an UFIR smoother.

Detrending

At this state, Algorithm 4.1.4 is applied on a large horizon $N \gg N_{\text{opt}}$ to remove artifacts from the external systems.

QRS-complex detection

The QRS-complex is detected using annotations of the arrhythmia MIT-BIH database following the approach proposed by Tompkins and Pan *et. al.* [94]. Note that a majority of annotations detect the QRS complex with a probability of 99.3%.

Segmentation

Localized the QRS-complex, a closest point to the R-peak is detected in each heartbeat. Next, by taking 100 samples to the left and 200 samples to the right, a window is created to outline

a heartbeat as in Fig. 4.1. If this window does not cover all points of interest (P-wave, QRS-complex, and T wave), its width is increased. The segmentation process is organized heuristically with the aim of analysing the morphological waves. We refer to this technique described in [100–102].

Iterative UFIR smoothing

Specified N_{opt} and N_{apt} for the database used as shown in [3], the horizon $N_{\text{opt}} = 21$ is applied beyond the QRS complex. To avoid large bias errors, N_{apt} specified by (4.13) is applied over all EGC signal. Provided UFIR filtering, smoothing with a lag q is organized using (4.14).

Windowing of ECG waves

The UFIR smoother provides denoising and estimation of the ECG signal three states as shown in Fig. 4.3 for the first state (de-noised ECG signal), second state (time derivative of the de-noised signal), and third state (second time derivative of the de-noised signal). Using information about the ECG signal states, the R-peak, QRS_{max}, and QRS_{min} are determined and a window is applied to cover the QRS complex. The P-waves detection is provided beginning from Q until the heartbeat ends. Here, a window is applied to cover P_{on} , P-peak, P_{off} points, which are determined by P_{max} and P_{min} in the second state. Similarly, the T-wave is detected, in which case T_{on} and T_{off} are covered by a window created for T_{max} and T_{min} (Fig. 4.3b).

4.2.1 Fiducial points detection and features extraction

In this section, we use the above results to provide fiducial points detection and features extraction, such as the durations and amplitudes of different detected fiducial points. Provided windowing of the P-wave, QRS-complex, and T-wave, we use the fiducial point P_{on} as an initial point of P-wave, P as a P-peak, P_{off} as a final point of P-wave, Q as an initial point of QRS complex, R as a R-peak, S as a final point of QRS complex, T_{on} as a initial point of T-wave, T as T-peak, and T_{off} as a final of T-wave. The fiducial points are extracted as follows.

QRS-complex

The fiducial points for a QRS-complex are determined by finding a maximum QRSmax and a minimum QRSmin in the second state (Fig. 4.3b), which are corroborated by the third state at zero cross points (Fig. 4.3c). Two variables “dqrs1” and “dqrs2” are introduced to calculate the

initial and final points of a QRS-compex. The R-peak is detected as $\hat{\mathrm{R}}$ at a zero cross point of the second state and is corroborated by QRSmin of the third state.

P and T waves

The fiducial points for P and T waves are determined by finding Pmax, Tmax, Pmin, and Tmin in the second state (Fig. 4.3b), which are corroborated by the third state at zero cross points (Fig. 4.3c). Two variables "dp1" and "dp2" are introduced to calculate the initial and final points of the P-wave. Similarly, two variables "dt1" and "dt2" are introduced for the T-wave. The P-peak and T-peak assigned as $\hat{\mathrm{P}}$ and $\hat{\mathrm{T}}$, respectively, are detected at the zeros cross of the second state. These points are confirmed by Pmin and Tmin in Fig. 4.3.

Provided the fiducial points $\hat{\mathrm{P}}_{\mathrm{p}}^{\mathrm{on}}$, $\hat{\mathrm{P}}$, $\hat{\mathrm{P}}_{\mathrm{p}}^{\mathrm{off}}$, $\hat{\mathrm{Q}}$, $\hat{\mathrm{R}}$, $\hat{\mathrm{T}}_{\mathrm{p}}^{\mathrm{on}}$, $\hat{\mathrm{T}}$, $\hat{\mathrm{T}}_{\mathrm{p}}^{\mathrm{off}}$ to represent the relevant points[1] in Fig. 4.1, the ECG wave durations and amplitudes are calculated for the P-wave as

$$\mathrm{P}_{\mathrm{dur}} = \mathrm{P}_{\mathrm{p}}^{\mathrm{off}} - \mathrm{P}_{\mathrm{p}}^{\mathrm{on}} \cong \hat{\mathrm{P}}_{\mathrm{p}}^{\mathrm{off}} - \mathrm{S}\hat{\mathrm{P}}_{\mathrm{p}}^{\mathrm{on}}\,, \tag{4.15}$$

$$\mathrm{P}_{\mathrm{amp}} = \mathrm{S}(\mathrm{P}_{\mathrm{p}}) - \mathrm{S}(\mathrm{P}_{\mathrm{p}}^{\mathrm{on}}) \cong \hat{\mathrm{P}} - \mathrm{S}(\hat{\mathrm{P}}_{\mathrm{p}}^{\mathrm{on}})\,, \tag{4.16}$$

for the QRS-complex as

$$\mathrm{QRS}_{\mathrm{dur}} = \mathrm{S}_{\mathrm{p}} - \mathrm{Q}_{\mathrm{p}} \cong \hat{\mathrm{S}}_{\mathrm{p}} - \hat{\mathrm{Q}}_{\mathrm{p}}\,, \tag{4.17}$$

$$\mathrm{QRS}_{\mathrm{amp}} = \mathrm{R} - \mathrm{S}(\mathrm{Q}_{\mathrm{p}}) \cong \hat{\mathrm{R}} - \mathrm{S}(\hat{\mathrm{Q}}_{\mathrm{p}})\,, \tag{4.18}$$

and for the T-wave by

$$\mathrm{T}_{\mathrm{dur}} = \mathrm{T}_{\mathrm{p}}^{\mathrm{off}} - \mathrm{T}_{\mathrm{p}}^{\mathrm{on}} \cong \hat{\mathrm{T}}_{\mathrm{p}}^{\mathrm{off}} - \hat{\mathrm{T}}_{\mathrm{p}}^{\mathrm{on}}\,, \tag{4.19}$$

$$\mathrm{T}_{\mathrm{amp}} = \mathrm{S}(\mathrm{T}_{\mathrm{p}}) - \mathrm{S}(\mathrm{T}_{\mathrm{p}}^{\mathrm{on}}) \cong \hat{\mathrm{T}} - \mathrm{S}(\hat{\mathrm{T}}_{\mathrm{p}}^{\mathrm{on}})\,. \tag{4.20}$$

Note that estimates $\hat{\mathrm{P}}_{\mathrm{p}}^{\mathrm{on}}$, $\hat{\mathrm{P}}_{\mathrm{p}}^{\mathrm{off}}$, $\mathrm{S}(\hat{\mathrm{Q}}_{\mathrm{p}})$, $\hat{\mathrm{T}}_{\mathrm{p}}^{\mathrm{on}}$, and $\hat{\mathrm{T}}_{\mathrm{p}}^{\mathrm{off}}$ represent points, which belong to the base line of an ECG signal.

4.3 Tuning and Testing

In order to achieve the best smoothing effect, in this section we tune the UFIR smoother to the ECG signals in terms of optimal lags related to optimal horizons. As benchmarks, we will employ the wavelet-based, low-pass, high-pass, median, and notch filters employed in [21, 22, 32, 43, 63, 103].

[1] The points P_{p}, Q_{p}, R_{p}, S_{p} and T_{p} are time indexes that determine the peaks P-peak, Q-point, R-peak, S-point and T-peak in the heartbeat.

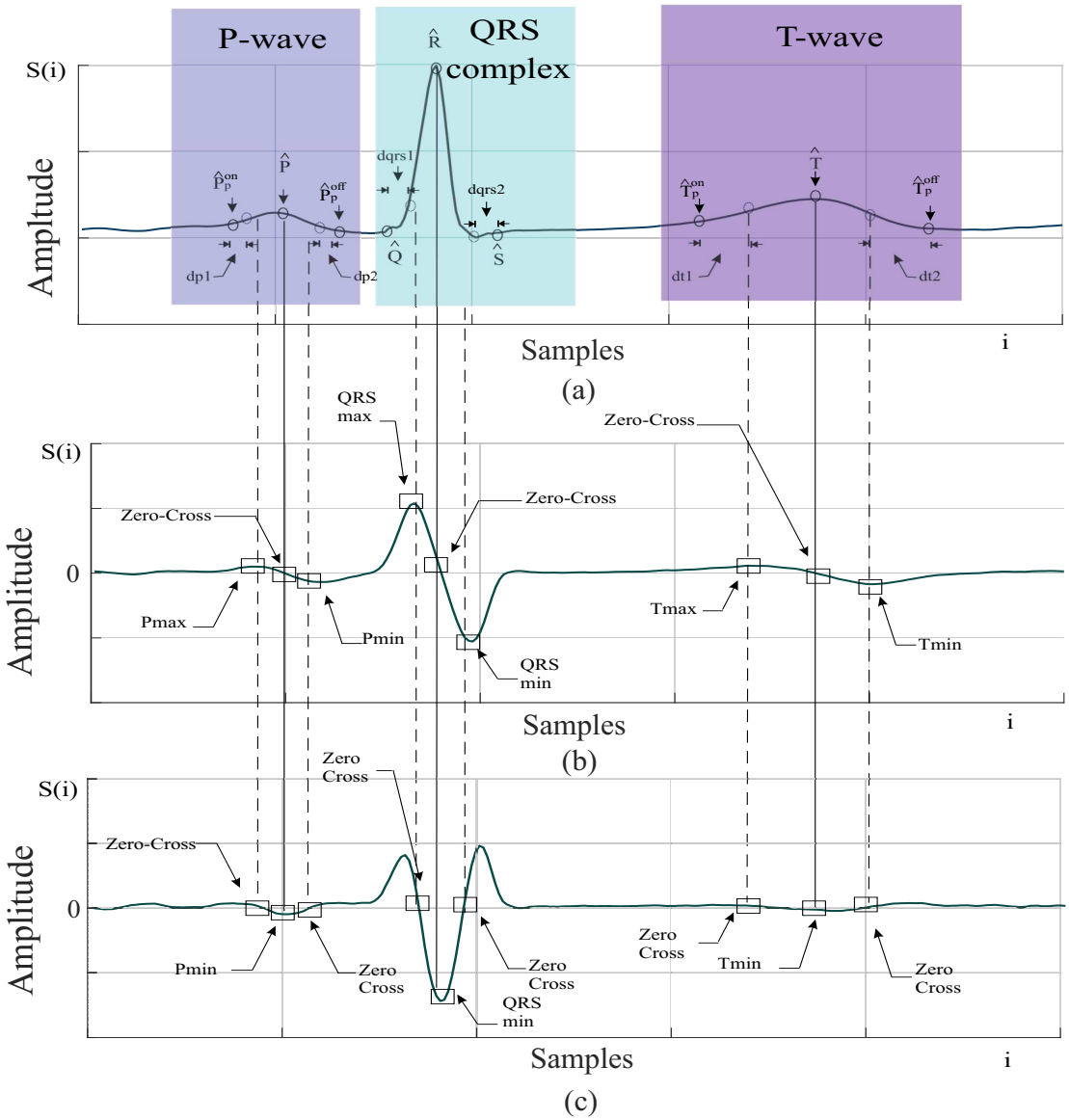

Figure 4.3: Fiducial features of a single heartbeat extracted along with the spacial points in state space using the UFIR approach: (a) first state, (b) second state, and (c) third state.

4.3.1 Optimal lag for UFIR smoother

It has been shown in [69] that an optimal lag q_{opt} for odd-order polynomial UFIR smoothers must be taken from the middle of an optimal averaging horizon of N_{opt} points. Accordingly, we specify q_{opt} as

$$q_{\text{opt}} = \left\lfloor \frac{N_{\text{opt}} - 1}{2} \right\rfloor , \tag{4.21}$$

where $\lfloor x \rfloor$ means the floor of x, i.e. the largest integer less than or equal to x.

For even-order polynomials, [69] suggests that q_{opt} must be set individually. Specifically, the

minimum optimal lag for the second-order UFIR smoother is [69]

$$q_{\text{opt}} = \frac{N_{\text{opt}} - 1}{2} - \frac{1}{2}\sqrt{\frac{N_{\text{opt}}^2 + 1}{5}}\,. \tag{4.22}$$

In this regard, let us notice that even though the Savitsky-Golay smoother [33] was derived from different prospectives, it has a similar structure with the UFIR smoother and similar properties such as adaptability to signal variations and robustness to noise. An advantage of the UFIR approach is that it suggests optimal lag for each smoother degree [69] that was not provided by Savitsky and Golay.

4.3.2 Testing iterative UFIR algorithm

The three-state polynomial model was shown in [3] to be near optimal for ECG signals. Referring to [3], we represent the system and measurement matrices as, respectively,

$$\mathbf{F} = \begin{bmatrix} 1 & \tau & \frac{\tau^2}{2} \\ 0 & 1 & \tau \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{H} = [\,1\ 0\ 0\,], \tag{4.23}$$

where a discrete time-step $\tau = 1/f$ is due to the sampling frequency of $f = 360\,\text{Hz}$ used in DataBase MIT-BIH Arrhythmia. For (4.23), the augmented measurement matrix becomes

$$\mathbf{W}_{m,n} = \begin{bmatrix} \mathbf{H}\mathbf{F}^{-2} \\ \mathbf{H}\mathbf{F}^{-1} \\ \mathbf{H} \end{bmatrix}. \tag{4.24}$$

At these stage, we compare performances of the UFIR smoother relying on q_{opt} (4.21) and (4.21) and several other available filters. To test estimators, we generate a signal $s(n) = \sin(n)$ corrupted by an additive zero mean white Gaussian noise (WGN) having the variance $\sigma^2 = 0.0625$ and sketch the results in Fig. 4.4. As can be seen, the UFIR smoother with q_{opt} (4.22) is most successful in accuracy, since its estimate ranges most close to the generated signal.

The root mean square errors (RMSEs) corresponding to Fig. 4.4 and computed over 1000 iterations and are shown in Fig. 4.5. One observes that all wavelet-based and standard filters produce much larger errors than the UFIR smoothers irrespective of the wavelet chosen. Among the two UFIR smoothers used, the second one performs better due to the optimal lag (4.22). This simulation confirms the fact that the lag must be chosen optimally for all even-order smoothers, unlike for the Savitsky-Golay filter.

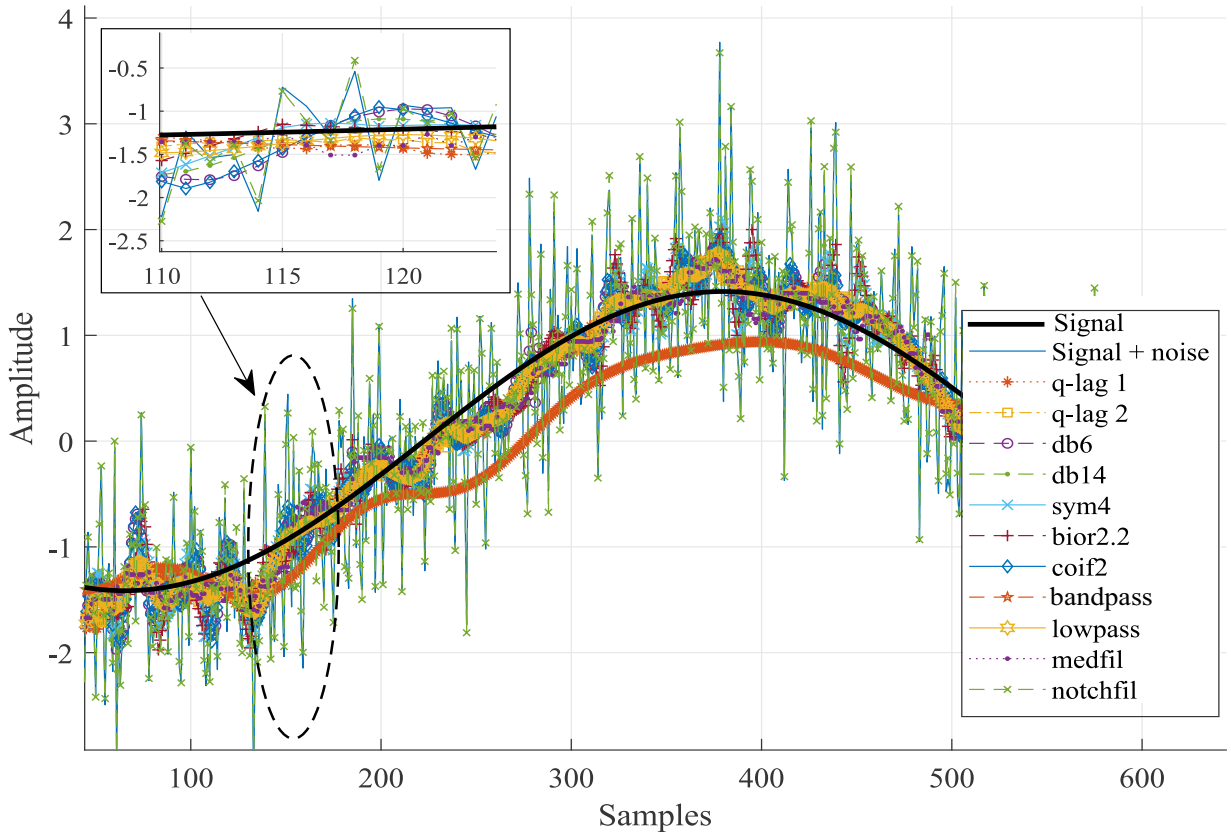

Figure 4.4: Denoising of a test sinusoid signal (solid) corrupted by zero mean AWGN using the UFIR smoother (dotted with asterisk) with lag q1 (4.21) and (dash-dotted and marked square) with lag q2 (4.22). The Daubechies wavelet-based smoothers are: db6 (dashed with marked circle), db14 (dash with dot marked), sym4 (solid with marked cross), bior2.2 (dashed with marked plus sing), coif2 (solid with marked diamond). The standard filters are: band pass filter (bandpass, dashed with marked pentagon), low pass filter low-pass (solid with marked hexagon), median (medfilt, dotted with marked point), and notch (notchfil, dashed with marked point).

We next provide an analysis of the signal-to-noise ratios (SNRs) at the filter outputs in terms of the percentage root mean square (PRD). In doing so, a synthetic ECG signal is considered with known ECG signal and noise. The pass-band filter is set aside due to the instability (Fig 4.5). As can be seen in Fig. 4.6, both UFIR smoothers (q-lag1 and q-lg2) are most successful in accuracy for small and large SNR values. The notch filter produce considerable errors when the SNR drops below 10 dB. The wavelet-based smoothers perform well when the SNR exceed 20 dB and the low-pass filter performs similarly. It is also seen that the median filter is less accurate

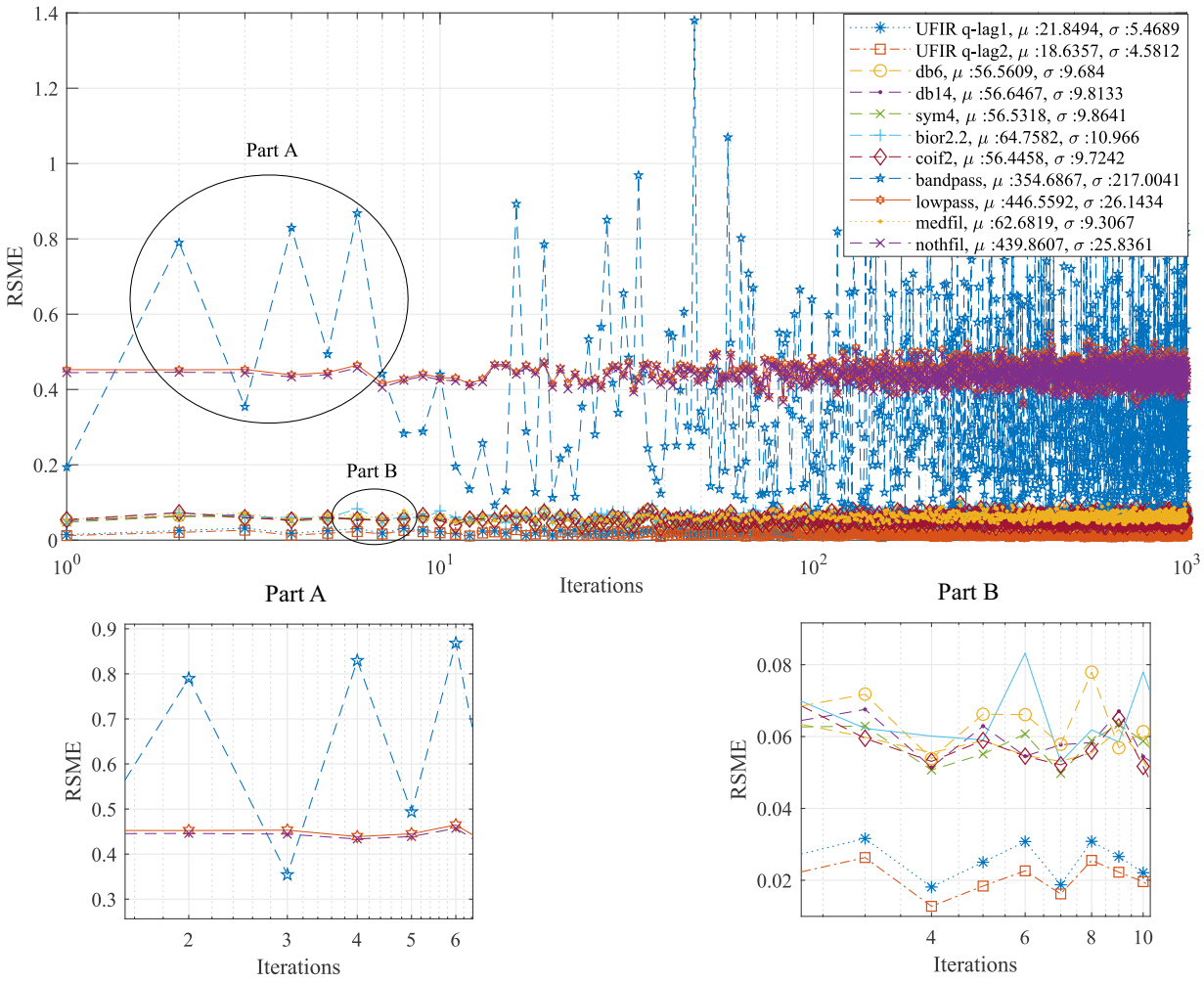

Figure 4.5: RMSEs corresponding to Fig. 4.4 and computed over 1000 iterations for the UFIR smoother, wavelet-based, and standard filters such as the low-pass, band-pass, median, and notch.

among other solutions when the SNR exceeds 20 dB.

Another experiment has been conducted to analyse the error variability with respect to the signal energy. The results are sketched in Fig. 5.9 in terms of the PRD. Again we notice that both UFIR smoothers produce smallest errors among other solutions.

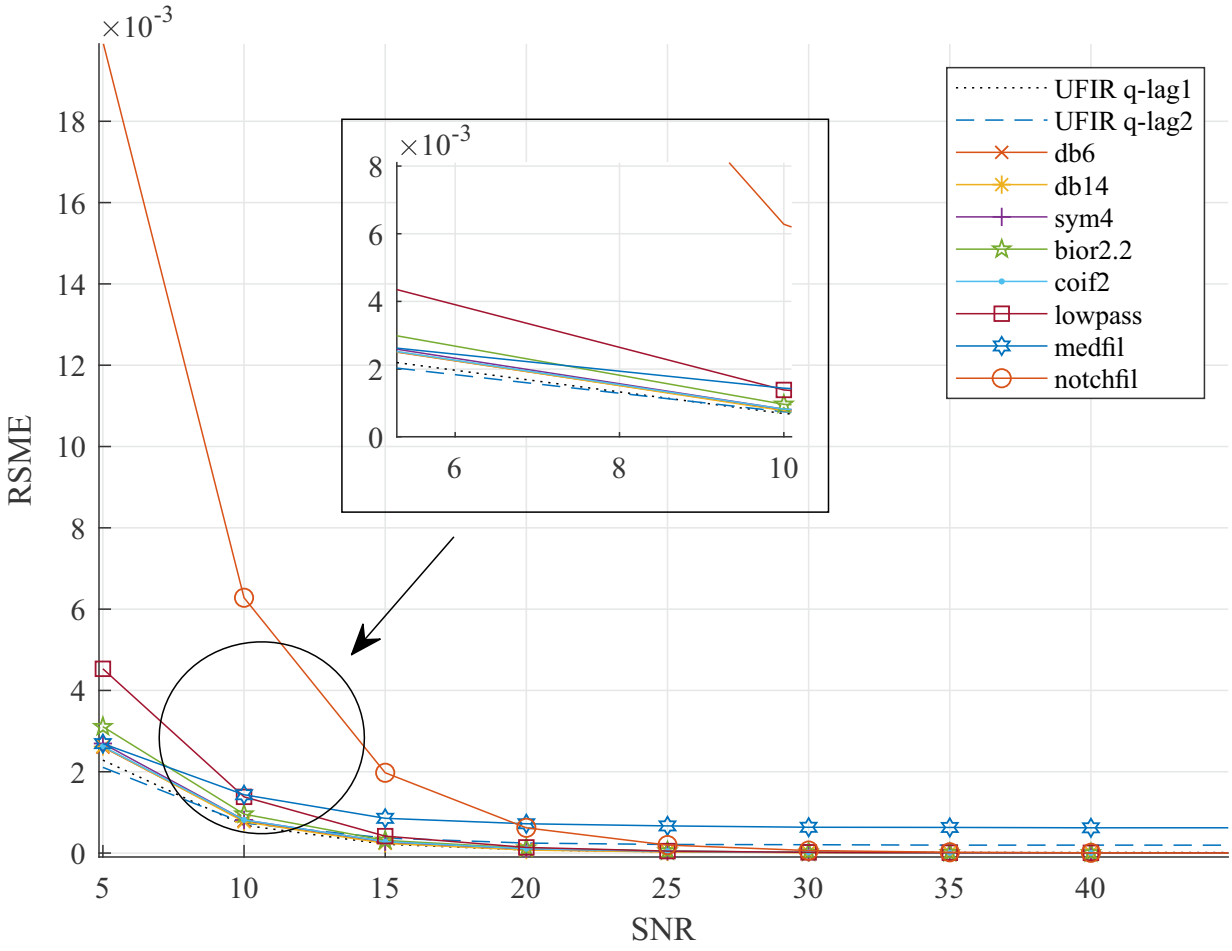

Figure 4.6: RMSEs of the UFIR smoother compared to the wavelets-based and standard filters.

4.4 Assessment of adaptive UFIR smoothing algorithm for ECG signals

In this section, we make efforts to extract features of ECG signals with a highest available accuracy provided by the adaptive UFIR smoothing algorithm designed based on the MIT-BIH Arrhythmia benchmark [76], which contains several records taken from different databases such as the MIT-BIH Arrhythmia (MITDB). The wavelet-based filters with several mother wavelets will be used as benchmarks.

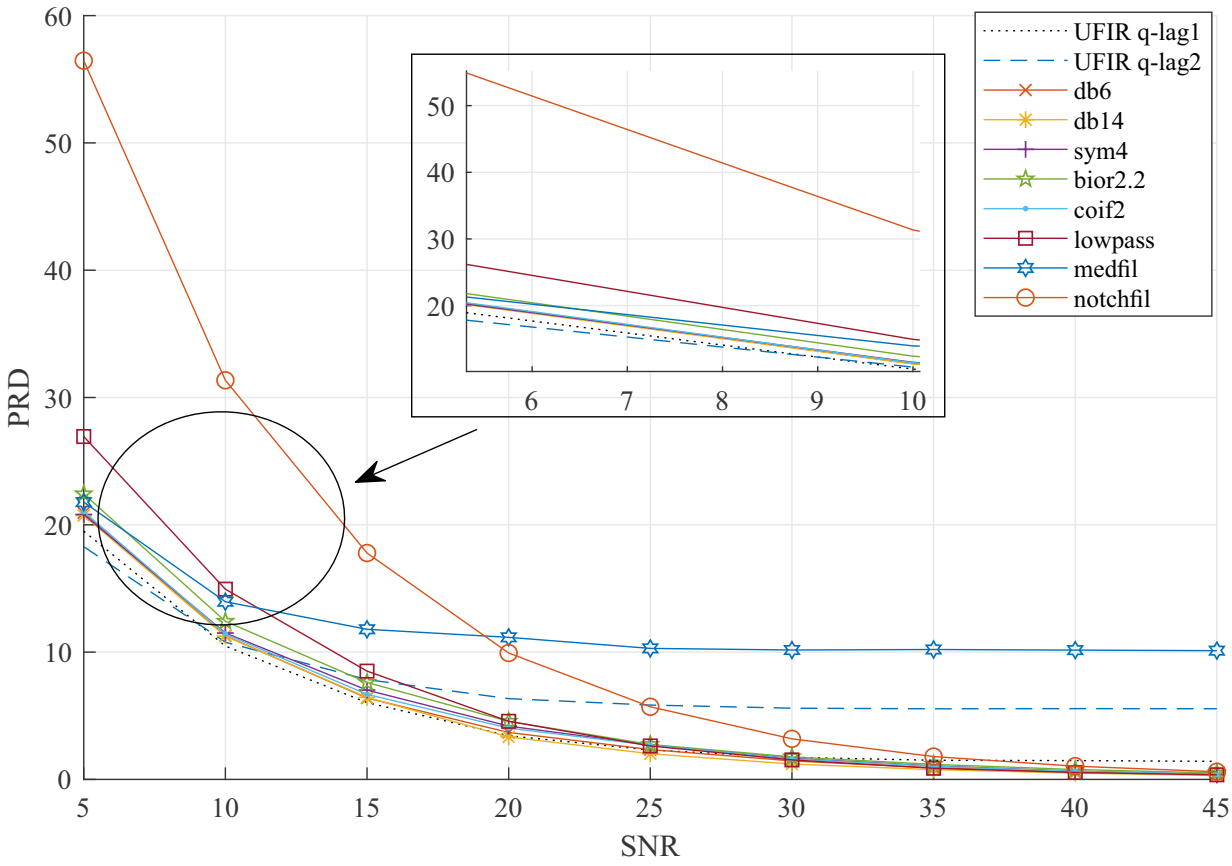

Figure 4.7: Smoothing errors in terms of PRD produced by diverse filters.

4.4.1 Filtering and artifact removal

What we expect from the estimates of the first state is that the outputs of the UFIR smoother with lags (4.21) and (4.22) and the outputs of the wavelet-based filters will not get away significantly from one another. Herewith, we suppose that errors in the estimates of the second and third states provided by the Savitsky-Golay smoother and wavelet-based filters will range higher than in the UFIR smoother, because the former estimates the high-order states via the derivatives, while the later makes it in state space concurrently. Our expectations are confirmed in Fig. **??**, where we also highlight a part with clearly seen bias errors when an ECG signal changes rapidly within the QRS-complex.

To sketch a more clear error picture, in Fig. 4.9 we give the measurement residuals produced by different estimators. What follows from this figure is that the UFIR smoother outperforms the wavelet-based and standard filters over all data, especially within the QRS complex. In Fig. 4.10, we give estimates of the second state provided by the estimators within and beyond the

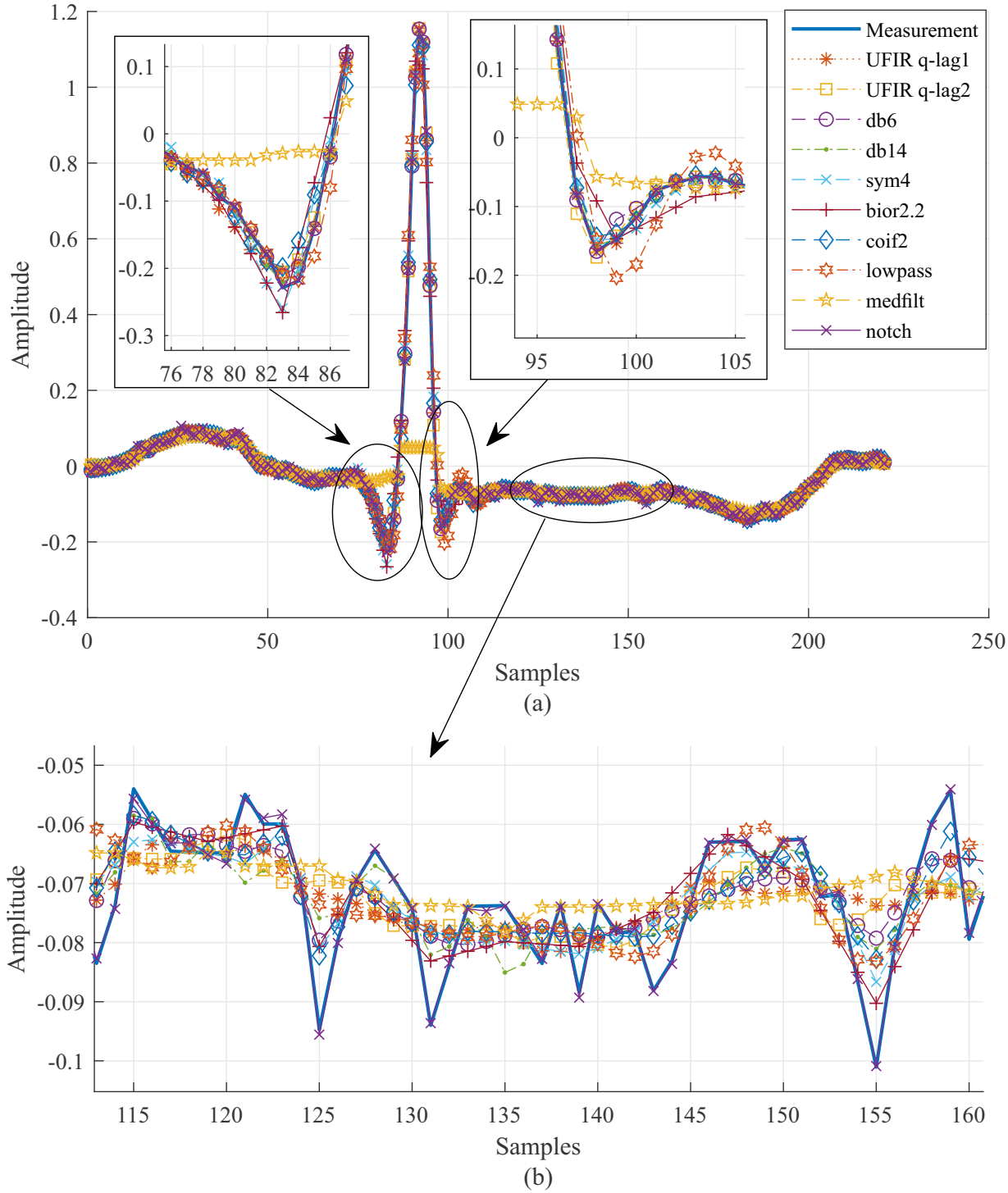

Figure 4.8: ECG signal denoising: (a) heartbeat estimation with the UFIR smother, wavelet-based filters, and standard filters; (b) segmental visualization of ten estimates.

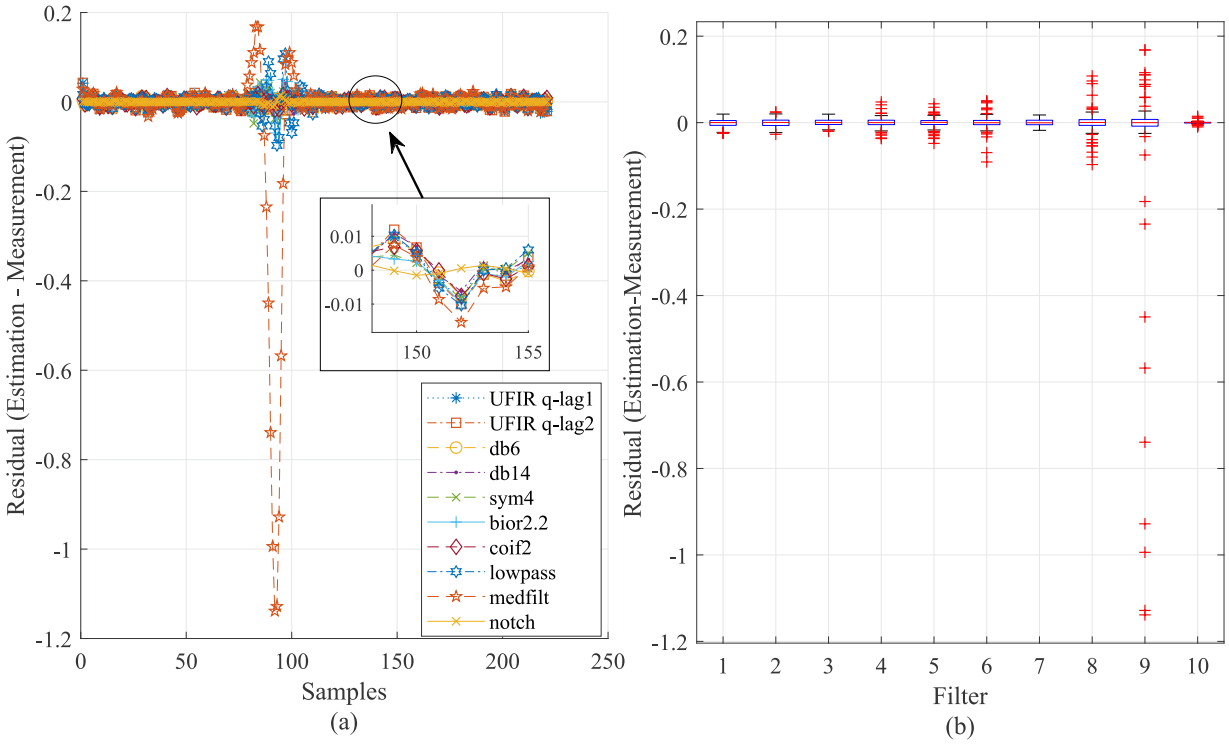

Figure 4.9: Measurement residuals produced by the UFIR smoother (q-lag1 and q-lag2), wavelet-based filters (db6, db14, sym4, bior2.2, and coif2), low-pass filter, median filter, and notch filter: (a) actual residuals and (b) error boxplot of heartbeat.

QRS-complex. This figure also confirms that the UFIR smoother is most accurate among other solutions. An important specific is that the UFIR smoother is able to remove efficiently artifacts as shown in Fig. 4.11. This property is useful to detrend the process, such as that shown in Fig. 4.1.

4.4.2 Computational complexity

Although the computation time is not strictly limited in ECG signals processing, an issue may arise when the consumed time is unacceptably large for medical needs. To find out how fast each algorithm operates under the same conditions, we next process an ECG record of 30 seconds with 1000 iterations. We base the computation time measurement on the MATLAB R2019 operating on a computer with intel core i7-4510U CPU (2.60) GHz and 16.0 GB RAM. The consumed times are listed in Table 4.1 and it is seen that an increase in the accuracy in the UFIR smoother is achieved at expense of the computation time, which is largest among other solutions, because the

UFIR algorithm 4.1.4 has the $\mathcal{O}(N)$ complexity [72]. Even so, the time consumed by the UFIR smoother can be acceptable for medical needs, provided that the results demonstrate highest accuracy. Note that the UFIR algorithm still was not optimized in terms of fast operation and the computation time can be significantly reduced in special implementations.

Table 4.1: Computation Time Required by Diverse Algorithms

Algorithm	Average time (sec)	Parameter
UFIR q-lag 1	6.81	N=21
UFIR q-lag 2	5.48	N=21
db6	0.1610	level=3
db14	0.1521	level=3
sym4	0.1424	level=3
bior 2.2	0.1379	level=3
Coif2	0.1331	level=3
lowpass	0.3556	30Hz
medfilt	0.0029	–
notchfil	0.0015	60Hz

4.4.3 Features extraction and errors comparison

Provided estimates of the ECG signal states, we next conduct accurate features extraction following the above discusses scheme, in which relations (4.19) and (4.20) are used to extract features of the P-wave, (4.21) and (4.22) to compute the QRS-complex duration and amplitude, and (4.23) and (4.24) to extract features of the T-wave.

An extraction of the P-wave duration using the UFIR smoother is illustrated in Fig.4.12, where we also sketch estimates provided by some wavelet-based filters. The estimates are given along with the expert annotations (gold standard) taken from [2, 100, 104] and shown as the upper and lower boundaries corresponding to the confidence interval of the probability of 95%.

Several features extracted using the UFIR smoother and other algorithms are generalized in Fig. 4.13. Again we see that the UFIR smoother provides estimates consistent with the gold standard, while the wavelet-based and standard filters are not always successful and their estimates undergo hight variabilities leading to inconsistent outputs. A more deep investigation respect to the P-wave will be provided next for normal and abnormal ECG signals.

4.4.4 Applications to normal and abnormal ECGs

As an example of applications, we now extract features of the P-wave related to records with normal rhythm and atrial fibrillation and illustrate the results obtained using the wavelet filter with Db6 (Fig. 4.14) and UFIR smoother (Fig. 4.15). The following observations follow from an analysis of these figures:

- The UFIR smoother puts the extracted features within the confidence interval that allows getting a strong discrimination between the normal and abnormal records as will be shown latter.
- All other algorithms produce unstable estimates (Fig. 13), which range out of the gold standard boundaries. Thus, making a good determination between the normal and abnormal records is more problematic by these filters.

4.4.5 Classification

We now evaluate features provided by (4.15), (4.16), (4.17), (4.18), (4.19) and (4.20) using nine classifiers. Considering 29266 heartbeats including healthy and abnormal heartbeats, we first train the classifiers by the cross-validation process 10 considering records 100, 103, 105, 201, 203, 210 from arrhythmia MITDB. Next, the classifiers are tested by new data 106, 112,113, 219, 221 taken from arrhythmia MITDB. All data are divided into several balanced sets to avoid biases produced by imbalanced data (a specific class set is larger than other). The metrics used for performance assessment are accuracy (Acc.), specificity (Spec.), and sensitivity (Sens.),

$$\text{Acc} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \tag{4.25}$$

$$\text{Spec} = \frac{\text{TN}}{\text{TP}+\text{FN}}, \tag{4.26}$$

$$\text{Sens} = \frac{\text{TP}}{\text{TN}+\text{FP}}, \tag{4.27}$$

where, TP (true positives) means that healthy heartbeats are correctly classified, TN (true negatives) means that abnormal heartbeats are correctly classified, FN (false negatives) means that healthy heartbeats are classified as abnormal heartbeats, and FP (false positives) means that abnormal heartbeats are classified as healthy heartbeats.

By these metrics, the performance of each classifier turns out to be averaged that is seen in Table 4.2 representing the general classifier performance provided by the tree model (complex tree,

medium tree, and simple tree), logistic regression, ensemble model (bagged tree, support vector machine (SVM) (linear, quadratic, and cubic), and subspace k-nearest neighbour (KNN). Note

Table 4.2: Performance of the AF for Normal ECG Heartbeats Based on Different Classifiers

Classifier	Acc.	Spec.	Sens.
Complex Tree	0.9664	0.9710	0.9964
Medium Tree	0.9664	0.9710	0.9964
Simple Tree	0.9664	0.9614	0.9952
Logistic Regression	0.9160	0.9814	0.9611
Bagged Tree	0.9934	0.9996	0.9855
Linear SVM	0.9275	0.9889	0.8827
Quadratic SVM	0.8331	0.9889	0.7604
Cubic SVM	0.8072	0.8949	0.7887
Subspace KNN	0.8072	0.9823	0.5988

that the best classifiers were selected during the initial training. A similar process was organized by applying the principal component analysis (PCA) (See Table 4.3) and comparing the effects. It follows from both cases that the UFIR smoothing approach provides a considerably better performances that follows from Table 4.4, where a comparison is provided using the empirical mode decomposition (EMD), autoregressive model (AR), Hadamard transform (HT), wavelet transform (WT), and convolutional neural networks (CNN).

Features extraction in T-wave

The T-wave features extraction begins with the suppressing of the P-wave and QRS complex. This part of signal is detect by analysing some zeros cross before to the T-wave. Detected the P-wave and QRS complex, a new signal from ECG signal estimate depicts just the T-wave (See fig.4.3).

Applying the iterative UFIR filtering with K states ($K = 3$ in our case),here, three signals represents the estimation(First state fig. 5.5a), first derivative(Second state fig. 5.5b) and second derivative(Third state 5.5c) of T-wave.

Different features of the T-wave can be extracted from these estimates. Points T_{on} and T_{off} are obtained via the peak values, T_{max} and T_{min}. Similarly, features as amplitude and duration of the T-wave are calculated in this stage. The duration is the diference between T_{off} and T_{on} and

Table 4.3: Performance of the AF and Normal ECG Signals Applying PCA and Based on Different Classifiers: FG SVM is the Fine Gaussian SVM, CG SVM is the Coarse Gaussian SVM

Classifier	Acc.	Spec.	Sens.
Complex Tree	0.7144	0.7973	0.6221
Medium Tree	0.6942	0.7201	0.6655
Simple Tree	0.6757	0.5575	0.8073
Logistic Regression	0.8410	0.7514	0.9043
Linear SVM	0.8244	0.7527	0.7441
Quadratic SVM	0.8272	0.9019	0.4940
FG SVM	0.7804	0.8526	0.7848
CG SVM	0.7887	0.7560	0.6957
Bagged Tree	0.7801	0.8126	0.8474

Table 4.4: Comparative Study of AF detection using Different Approaches and the UFIR Smoother (UFIRS)

Studies	Acc	Sens.	Spec.	Approach
Maji *et al.* [67]	–	96.0%	–	EMD
Padmavathi *et al.* [44]	100%	–	–	AR
Lee *et al.* [66]	99.5%	99.9%	98.7%	HT
Annavarapu *et al.* [45]	–	97.2%	95.9%	RR-interval
Runnan He *et al.* [65]	99.2%	99.41%	98.9%	WT + CNN
Tateno *et al.* [46]	–	91.20%	96.08%	RR-interval
Alcaraz *et al.* [43]	88.84%	–	–	notch filter
Proposed method	99.3%	99.6%	99.9%	UFIRS

amplitude is the difference between value of baseline (In this case may be T_{off} or T_{on}) and the value of T-peak estimated $\hat{T}$. This technique is applied for normal and inverted T-waves(See fig. 5.6).

Finally, we demonstrate the performance of the approach by comparing with other available methods.

Evaluation of the proposed method for fiducial points

The proposed method is validated using the PTB and MIT-arrhythmia database. Both Benchmark are recognized and containing records with different diagnosis and pathologies such as: Arrhythmias, Myocardial infarction, heart failure, bundle branch block, myocardial hypertrophy, valvular heart disease, myocarditis, and healthy control . Also, we consider accepted tolerances according to [106] for T-wave. Then a comparison with other approached is provided.

Comparison with other methods

Determined T_{on}, T_{off} errors can be computed between the reference pulse and the estimates. The reference value is calculated by approximating the real one. Errors are averaged to determine the mean value μ and standard deviation σ.

In the table 4.5 summarizes the results for μ and σ given in milliseconds. In [106], the reference is determined via measurements related to boundaries of the fiducial points. We apply the UFIR-based algorithm and other methods described in the literature; namely, the threshold detector (TD) [107], wavelet detector (WD) [19], and morphological transform (MMD) [108]. As can be seen, the fiducial features estimated using the UFIR-based algorithm and represented with T_{off} have the smallest standard deviation indicating that the features are clustered closely around the media.

Detection of abnormalities

Given the MIT-BIH arrhythmia database, we select four records for analysing the concentration of the amplitude in the the T-wave. The first record has a inverted T-wave (as can see in fig 4.17, it seek as circled) , according to the literature may be a ischemic disease heart. The second and third records have normal behaviour. Finally, the fourth record has irregular waves with possible inversions in the T-wave. Known the conditions of data, as can see in fig. 4.17, a considerable discrimination is presented between abnormal and normal T-wave.

4.5 Discussion

The purpose of this investigation was to remove the measurement noise and extract concurrently features of ECG signals in state space using the q-lag UFIR smoother. This smoother does not require the noise statistics and initial values and is thus more suitable for ECG signals, whose noise

Table 4.5: Comparison of the ECG Signal Features Extracted Using Different Methods

Method	Parameter	T_{on}	T_{off}
TD [107]	μ(ms)	23.3	18.7
	σ(ms	28.3	29.8
MMD [108]	μ(ms)	7.9	8.3
	σ(ms	15.8	12.4
WD [19]	μ(ms)	-4.8	-8.9
	σ(ms	13.5	18.8
UFIR	μ(ms)	40.8	10
	σ(ms)	9.63	16.9
CSE(ref)	σ(ms)	$-$	30.6

is still not well understood. We were focused on the morphological features of individual ECG signals with normal rhythm and atrial fibrillation. To reach the highest accuracy allowed by the UFIR smoothing approach, we have developed an efficient algorithm and tested it by diverse ECG data in a comparison with other available techniques. The test has confirmed our expectations. Namely, the UFIR smoother considerably outperformed several standard algorithms in noise reduction and accuracy. That has become possible by setting optimal lags and adaptive horizons to the UFIR smoothing algorithm.

As benchmarks, we employed several wavelet-based filters and standard filters such as the low-pass, pass-band, median, and notch. A comparative analysis has shown that the UFIR smoother extracts the ECG signal specific features with higher accuracy. That was also expected, since the wavelet-based algorithms do not allow for time-varying dynamic optimization similar to the adaptive UFIR structures, at least we did not find relevant solutions suggested for ECG signals in the wavelet area. A critical advantage of the state-space UFIR approach is that, unlike in the Savitsky-Golay and wavelet-based filters, noise reduction and state estimation are provided

simultaneously. This presumes higher efficiency in noise reduction and better accuracy in features extraction. Note that the Savitsky-Golay and wavelet-based filters are not state-space estimators. Estimation of higher-order states can be provided using these filters *a posteriori* via the time-derivatives applied to the first state estimate that is typically accompanied with larger noise. As a result, even for the second state, the UFIR smoother produced much more accuracy in the estimation of extreme points of MIT-BIH arrhythmia database.

It worth noticing again that the Savitsky-Golay and wavelet-based filters were already recognized as standard approaches for ECG signals [42, 55–64]. In this regard, better performance of the UFIR smoothing algorithm developed in this paper opens new horizons in accurate and precise features extraction from measurements of ECG signals having normal and abnormal heartbeat characteristics.

The UFIR smoother optimized for ECG signals by setting optimal lags and adaptive horizons for each individual degree-polynomial has essentially outperformed the Savitsky-Golay filter, which does not suggest such an optimization [35, 36, 38]. Accordingly, the following main results were achieved:

1. Optimal denoising and artifacts removal with q_{opt}-lag assigned for each optimal horizon N_{opt}.
2. High accuracy in ECG signal denoising achieved using an adaptive optimal horizon N_{apt}.
3. High accuracy in features extraction achieved taking advantages of the state-space approach.

What left behind is to notice some particular differences between the UFIR and wavelet-based approaches. It has been revealed that errors produced by the wavelet-based filters are more dispersed in the extracted features. We explain it by the fact that the available wavelet shapes are not optimal for ECG signals. Furthermore, the wavelet-based filters are not state-space estimators. Therefore, even confusing results can be expected from wavelets. Another specific is that features extracted using the UFIR approach have appeared to be more stable than by the machine learning techniques. That has been demonstrated in a comparison with the EMD, PCA, HT, RR-interval analysis, WT + CNN, and notch filter.

Summarising, we state that the proposed UFIR smoothing approach is more suitable for ECG signals then other techniques and methods considered in this paper. This have been depicted in the table 4.4, where our method obtained 99.3% of accuracy, 99.6% of sensibility and 99.9% of specificity. A flaw is in the computational time, which is largest among other approaches. Thus,

it is still challenging to design fast UFIR smoother-based algorithms, although the computation time of several seconds is not an issue for medical needs.

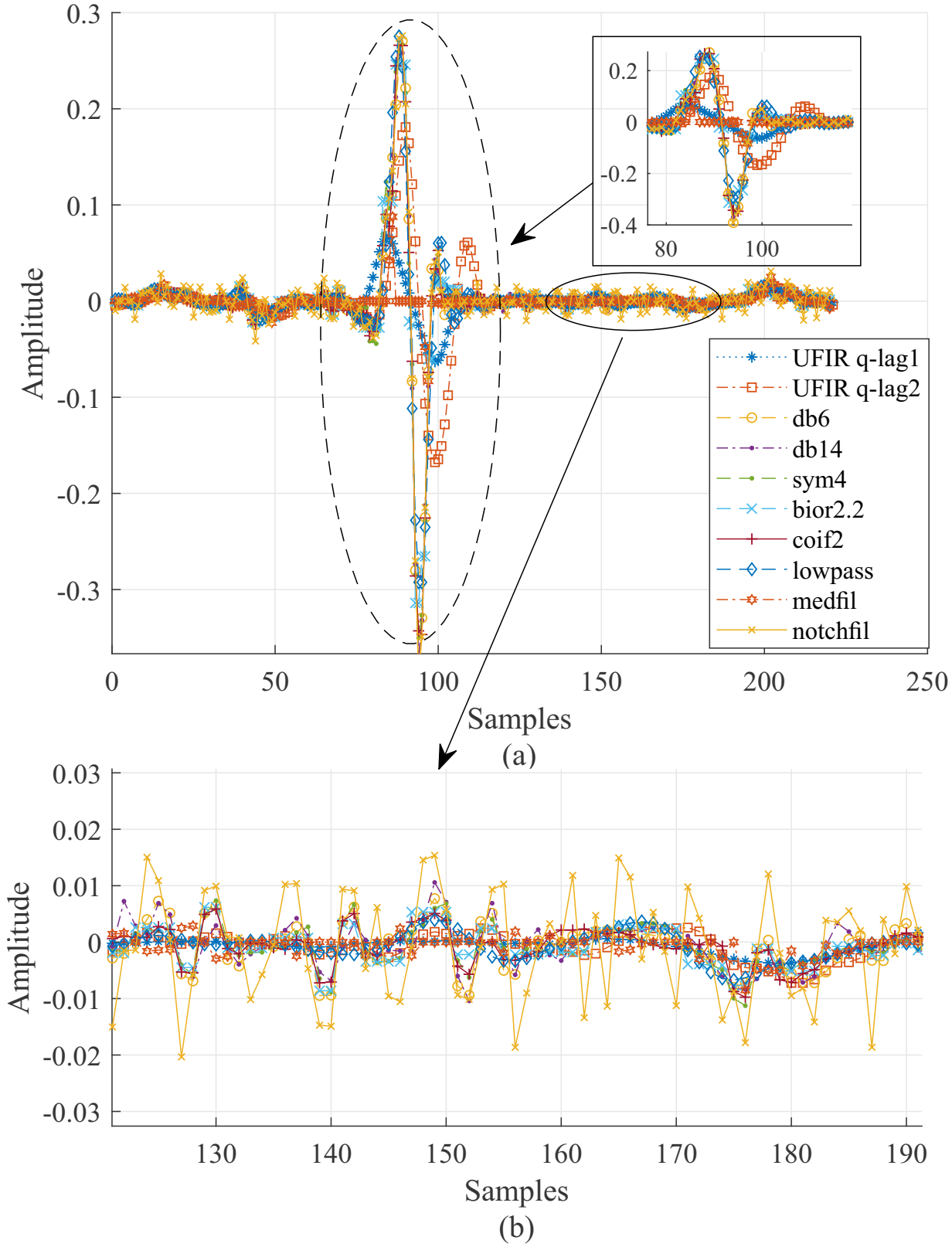

Figure 4.10: Estimates of the second state (first time-derivative) provided by the Savitsky-Golay smoother with lag q1 (4.21), UFIR smoother with lag q2, wavelet-based filters (db6, db14, sym4, bior2.2, and coif2), low-pass filter, median filter, and notch filter.

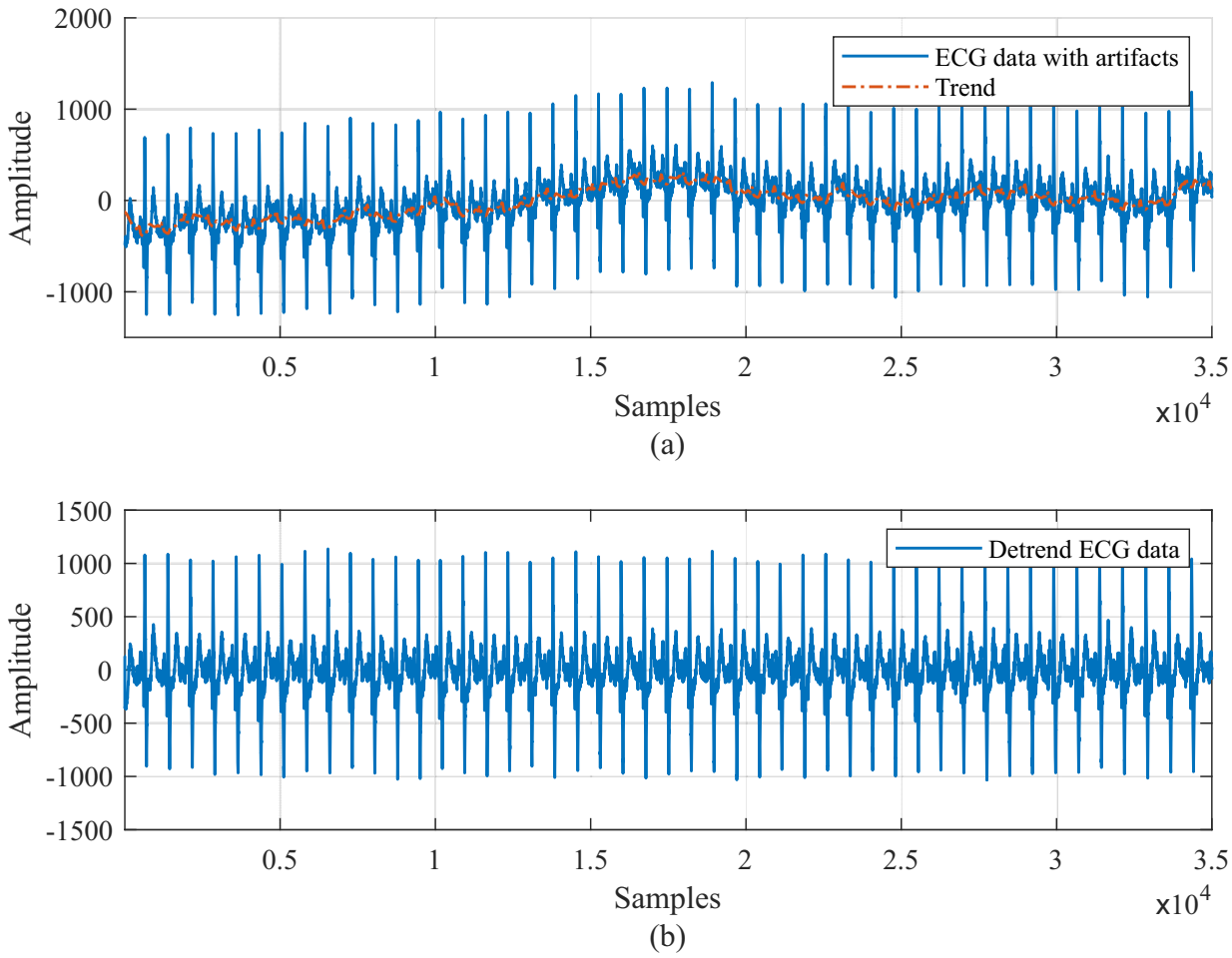

Figure 4.11: Baseline removal using UFIR smoothing with N=1001 and q-lag2

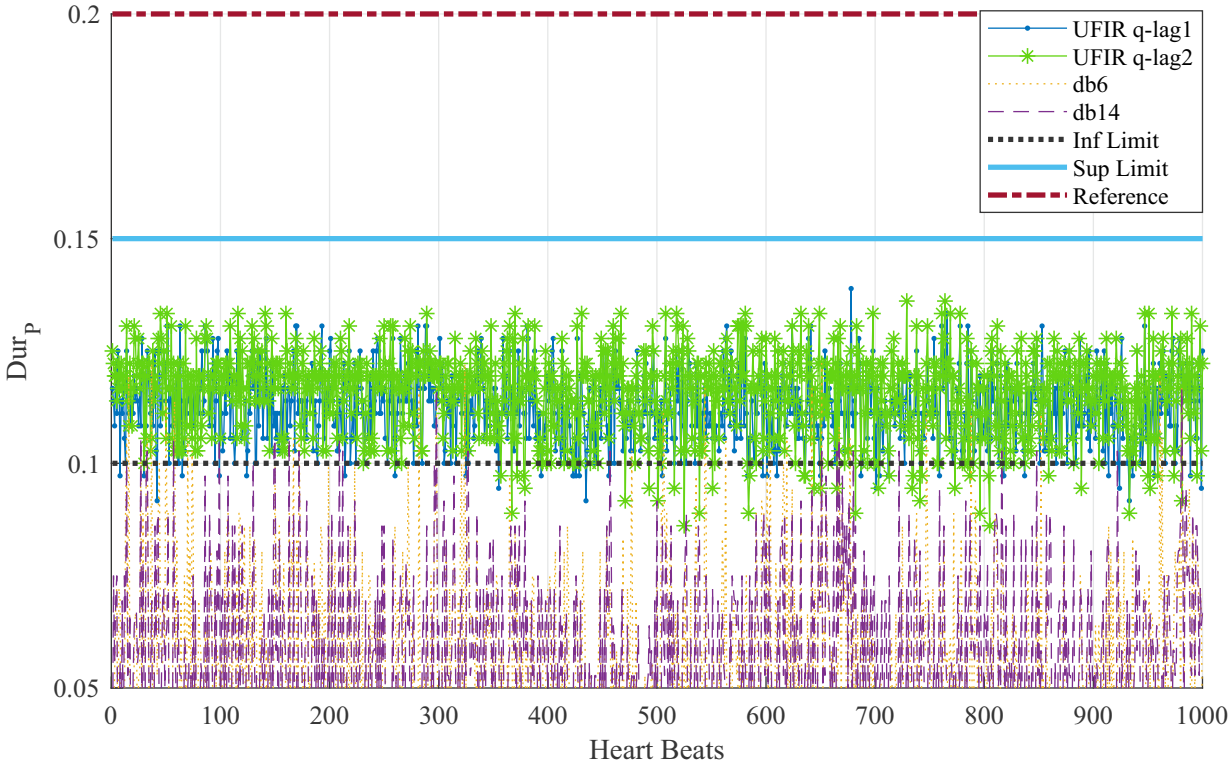

Figure 4.12: Extracted features of the P-wave duration $\mathrm{Dur_p}$. Expert annotations (gold standard) [2] are represented with the upper and lower boundaries (solid lines) along with an expected average.

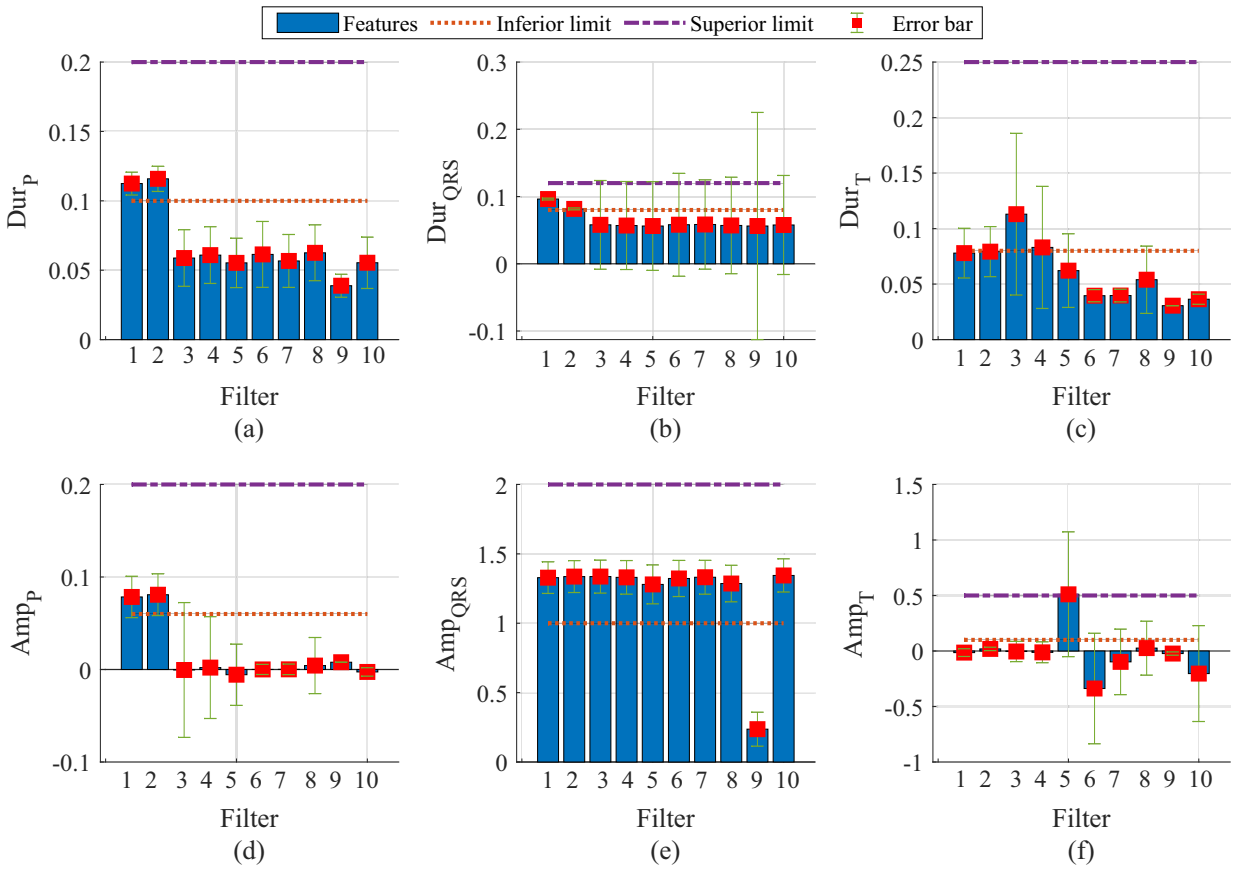

Figure 4.13: Features extracted using the UFIR smoother with q-lag1 (filter 1) and q-lag2 (filter 2) and other algorithms depicted as db6 (filter 3), db14 (filter 4), sym4 (filter 5), bior2.2 (filter 6), coif2 (filter 7), low-pass (filter 8), median (filter 9), and notch (filter 10): (a) duration $\mathrm{Dur_P}$ of P-wave, (b), duration $\mathrm{Dur_{QRS}}$ of QRS-complex, (c), duration $\mathrm{Dur_T}$ of T-wave, (d), amplitude $\mathrm{Amp_P}$ of P-wave, (e) amplitude $\mathrm{Amp_{QRS}}$ of QRS-complex, and (f) amplitude $\mathrm{Amp_T}$ of T-wave. Features are extracted from record 100 lead II of arrhythmia MIT-BIH database.

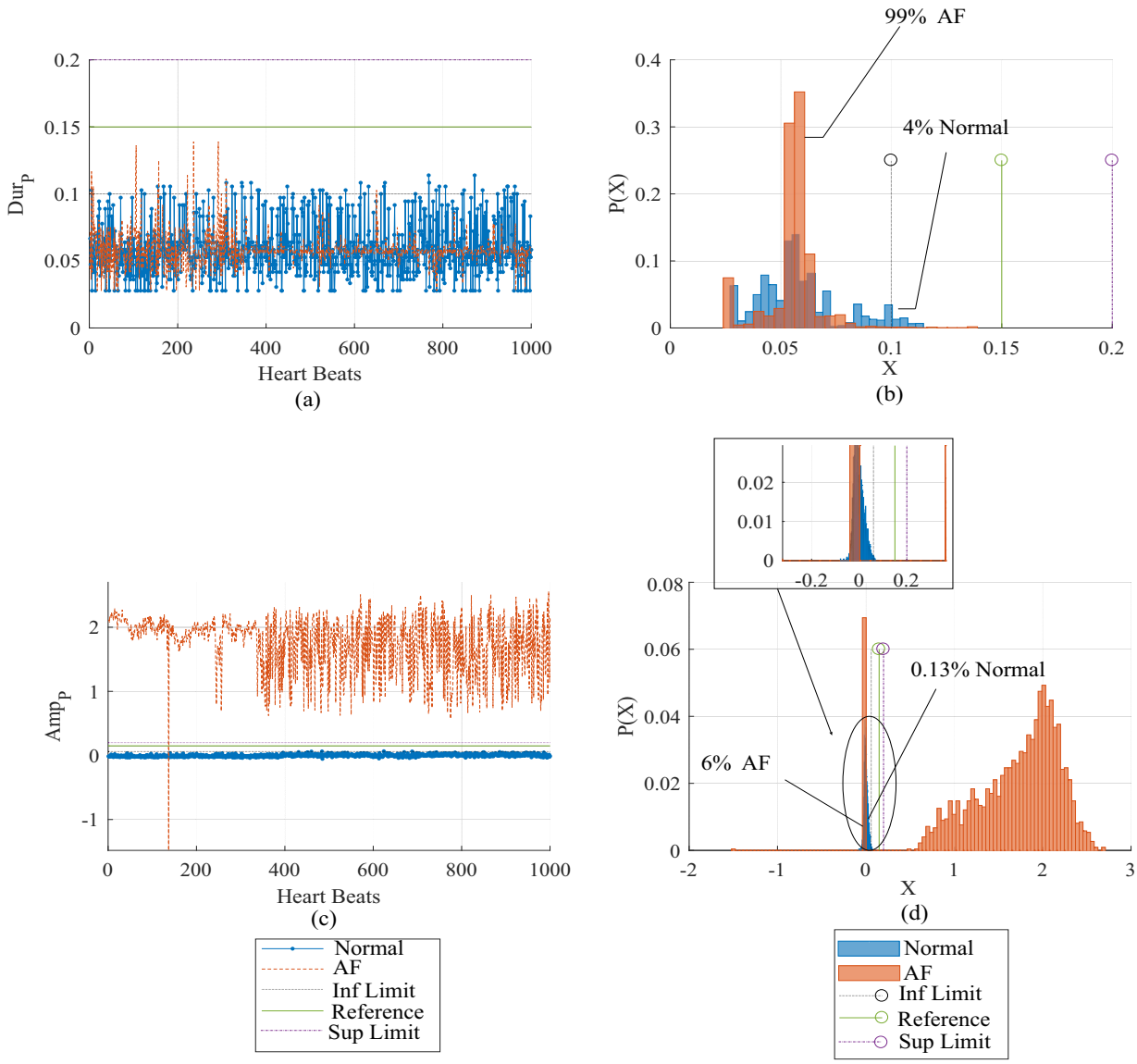

Figure 4.14: Features of the P-wave duration extracted using the wavelet-based filter with db6 from the AF and Normal ECG of MIT-BIH Arrhythmia Database: (a) $\mathrm{Dur_p}$, (b) $\mathrm{Dur_p}$ normalized histogram, (c) $\mathrm{Amp_p}$, and (d) $\mathrm{Amp_p}$ normalized histogram.

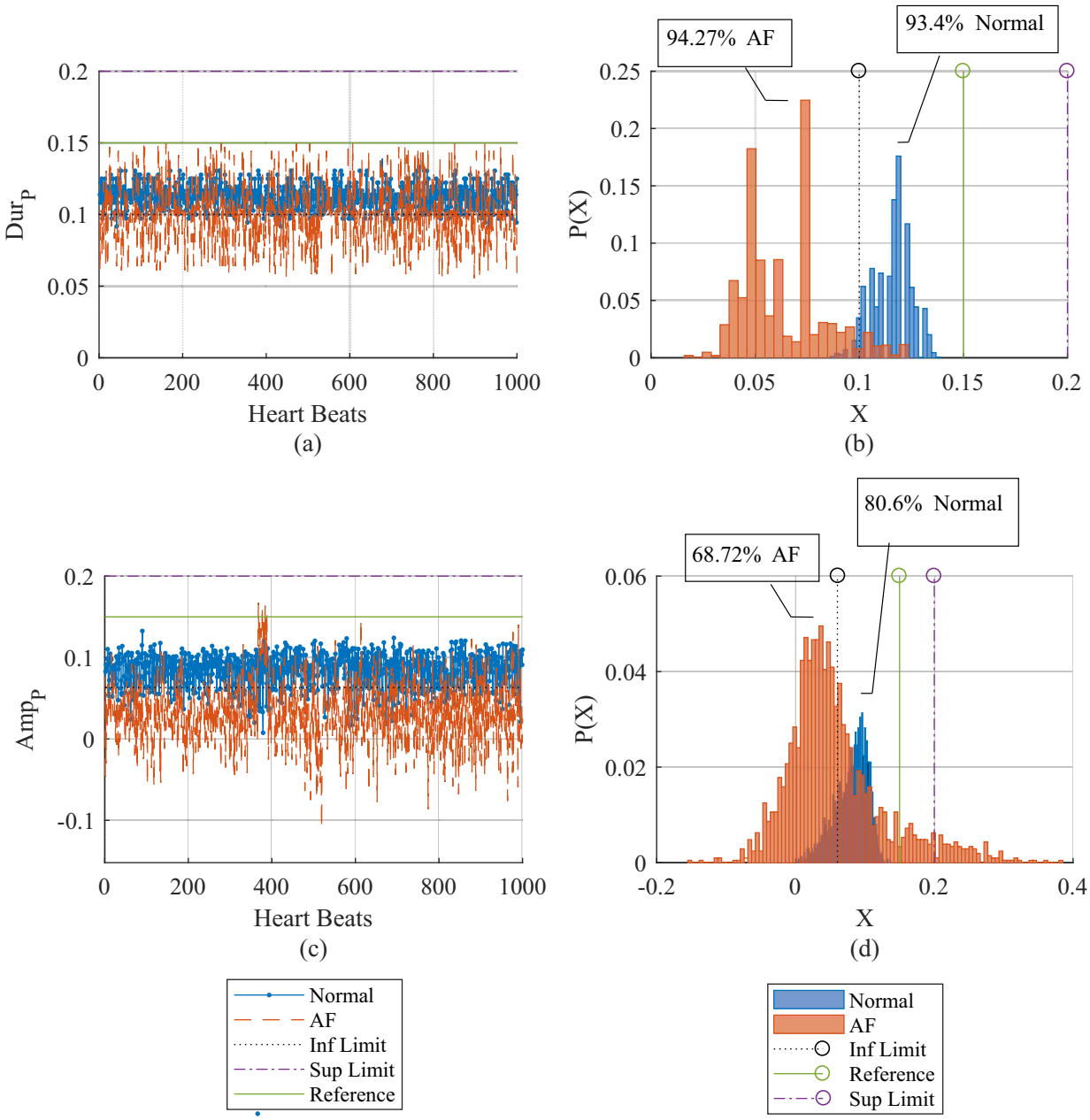

Figure 4.15: Features of the P-wave duration extracted using the UFIR smoother from the AF and Normal ECG of MIT-BIH Arrhythmia Database: (a) $\mathrm{Dur_p}$, (b) $\mathrm{Dur_p}$ normalized histogram, (c) $\mathrm{Amp_p}$, and (d) $\mathrm{Amp_p}$ normalized histogram.

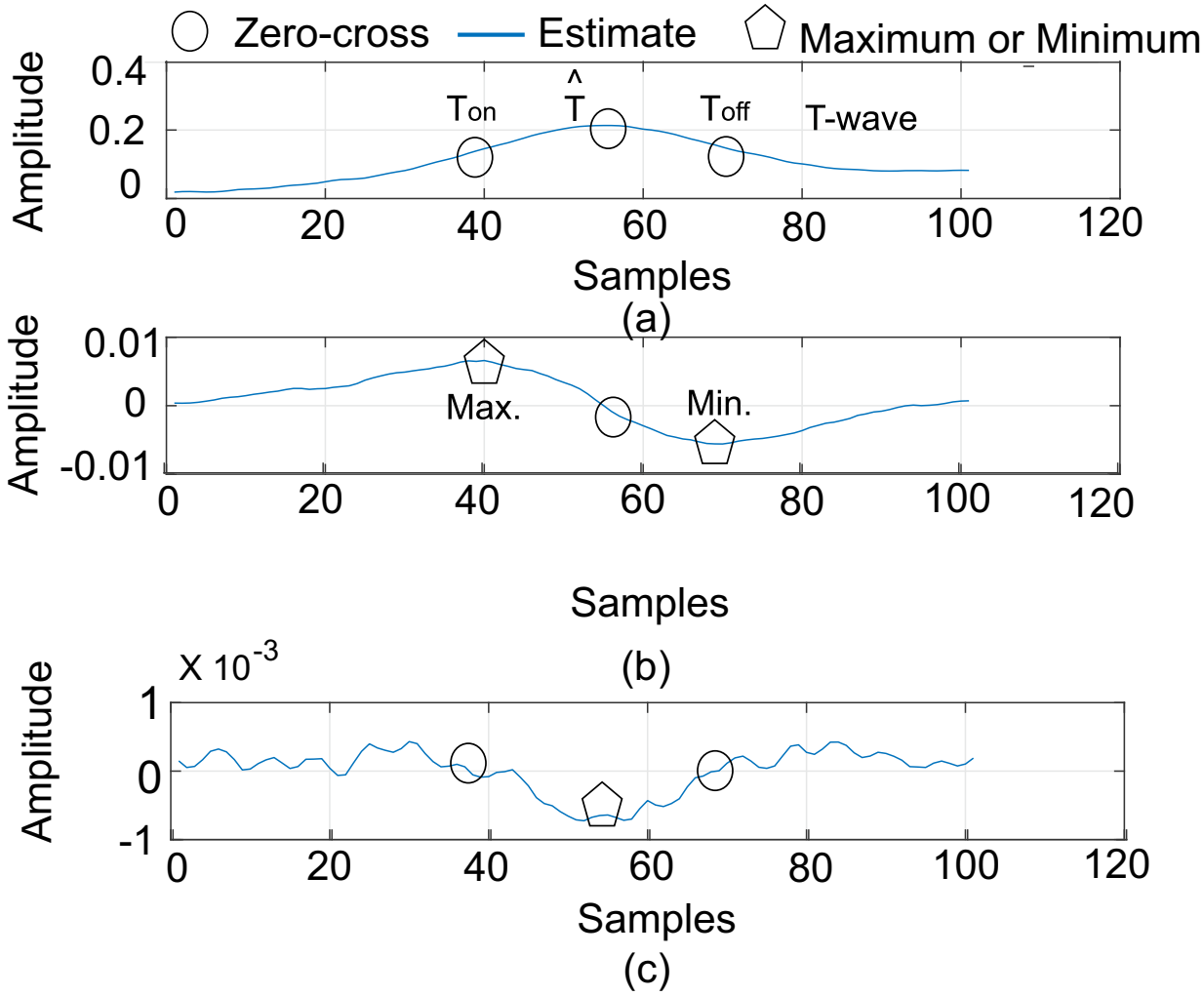

Figure 4.16: T-wave estimates provided by the unbiased FIR algorithm: a) First state: smoothed signal, b) Second state: first derivative and c) third state: second derivative.

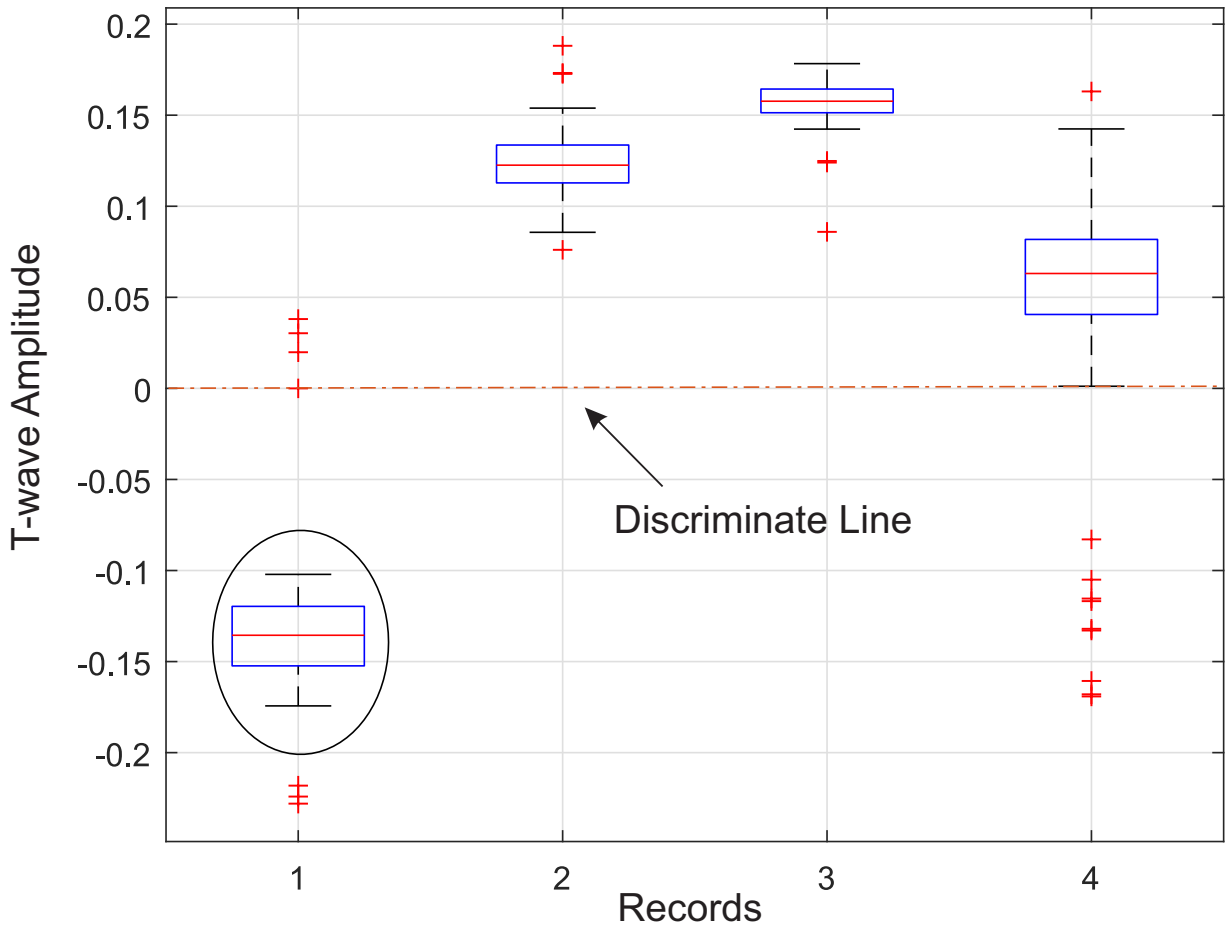

Figure 4.17: boxplot of different amplitudes T-waves for different record. Negative amplitude means inverted T-waves and Positive negative means normal T-wave. The discriminate line is located in the zero. The record one (circled) , two, three and four are the 100, 101, 103 and 112 records of MIT-Arrhythmia data base.

Chapter 5

ECG Signal Denoising and Features Extraction Using Harmonic Unbiased FIR Smoothing

In this chapter, we represent an ECG signal with a Fourier series, apply the UFIR smoother to real measurements of ECG signals, and provide a comparative analysis with the polynomial smoothers in terms of the denoising effect. We use the MIT-BIH Arrhythmia Database [75, 76], from which in this paper we take only the normal and premature ventricular complex heartbeats. In section 5.1, a brief description of harmonic model for ECG signals. In section 5.2, a optimal horizon calculation is defined by employing harmonic model . In section 5.3, an adaptive UFIR smoothing algorithm is designed for ECG records. In section 5.4, a evaluation for harmonic model is developed using real and simulated data. Finally, a discussion is performed about obtained results and literature.

5.1 State-Space Representation of ECG Signals using Harmonic Model

In view of a quasi-periodic nature of heartbeats, we represent an ECG signal (see Fig. 5.1) with a Fourier series corrupted by noise,

$$y(t) = A_{0,t} + \sum_{m=1}^{M} A_{m,t} \cos(m\omega t) + v(t)\,, \tag{5.1}$$

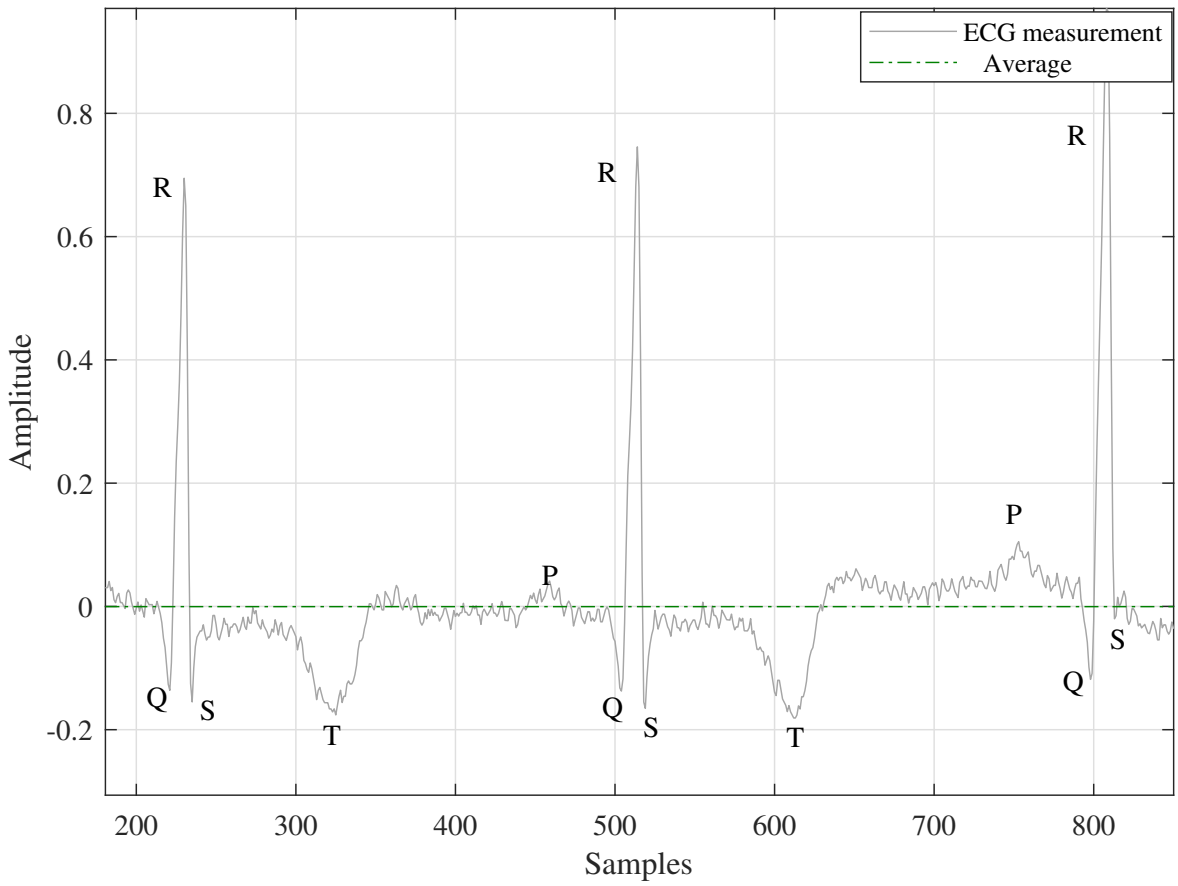

Figure 5.1: An example of the centralized ECG record taken from MIT-BIH Arrhythmia Database. The record 100/MLII features are depicted as P, QRS complex and T waves. Applied a correction of baseline, the ECG measurement average is near to zero

where ω is the fundamental angular frequency widely given as $\omega = 2\pi f_d$, fd is the fundamental frequency, M is the number of harmonics (see Fig. 5.2), $A_{m,t}$ are time-varying amplitudes associated with the mth harmonic, and $v(t)$ is an additive zero mean white Gaussian noise. Before applying an estimator, we make efforts to centralize the ECG data about zero. Therefore, the DC offset component becomes zero, $E\{A_{0,t}\} = 0$, and we will further omit $A_{0,t}$. If we introduce the discrete time t_k, $k = 0, 1 \ldots$, a time step $\Delta t = t_k - t_{k-1}$, and set $A_{0,t} = 0$, then the model (5.1) can be represented as

$$y_k = \sum_{m=1}^{M} A_{m,k} \cos(m\omega\Delta tk) + v_k \,. \tag{5.2}$$

Assuming that the mth harmonic components $\Delta A_{m,k}$ is random, the magnitude of the mth

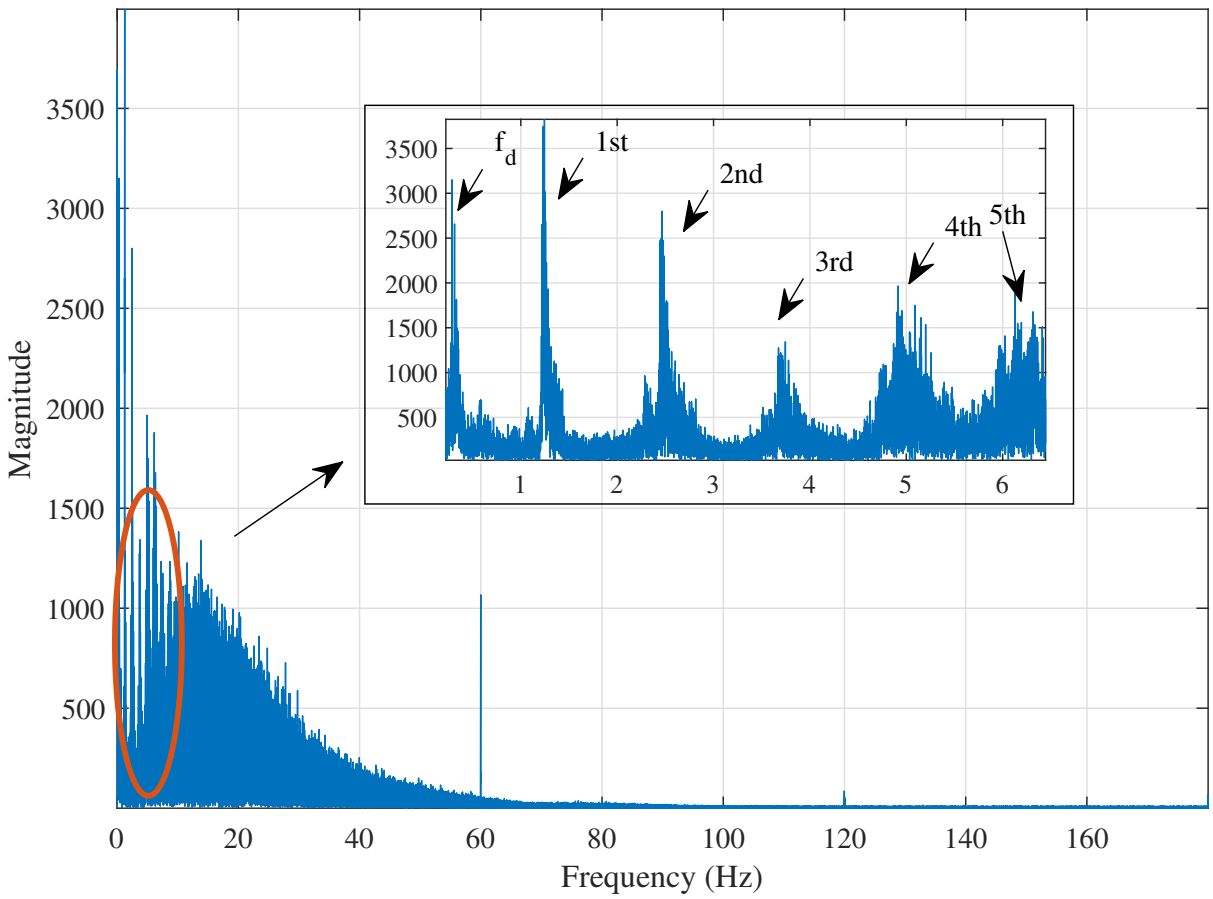

Figure 5.2: Single sided Magnitude spectrum: fundamental frequency f_d and first harmonics

harmonic component at time index $k+1$ can be represented as

$$A_{m,k+1} = A_{m,k} + \Delta A_{m,k}\,. \tag{5.3}$$

Referring to (5.2), we now introduce a $(2M)$-state vector $\mathbf{x}_k$,

$$\mathbf{x}_k = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{2M,k} \\ x_{2M+1,k} \end{bmatrix} = \begin{bmatrix} A_{1,k}\cos(w\Delta tk) \\ A_{1,k}\sin(w\Delta tk) \\ \vdots \\ A_{M,k}\cos(Mw\Delta tk) \\ A_{M,k}\sin(Mw\Delta tk) \end{bmatrix}, \tag{5.4}$$

where the components $x_{i,k}$, $i \in [1, M]$, are specified by the model (5.2). Accordingly, the time-

invariant state space model of an ECG signal can now be written as

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_k \,, \tag{5.5}$$

$$y_k = \mathbf{C}\mathbf{x}_k + v_k \,, \tag{5.6}$$

where matrix $\mathbf{A}$ is specified as

$$\mathbf{A} = [\text{blockdiag}(\Theta(m), m = 1, \cdots, M)] \,, \tag{5.7}$$

where "blkdiag" in (5.7) means a block of matrices $\Theta(m)$ represented by

$$\Theta(m) = \begin{bmatrix} \cos(m\omega) & -\sin(m\omega) \\ \sin(m\omega) & \cos(m\omega) \end{bmatrix} . \tag{5.8}$$

For this model, the observation vector becomes

$$\mathbf{C} = [1\ 0 \ldots 1\ 0]$$

and we think that the measurement noise v_k has zero mean, $E\{v_k\} = 0$, and unknown distribution and statistics, as it usually is in practice. Given the state-space model (5.5) and (5.6), the UFIR smoothing algorithm can be used as in the following.

5.1.1 Unbiased FIR Smoothing

The p-shift UFIR filtering approach [109] described by chapter 3 as convolution and chapter 4 as matrix form, suggests that 1) UFIR filtering must be provided as $\hat{\mathbf{x}}_k$ and 2) the q-lag UFIR smoothing organized by projecting the filtering estimate $\hat{\mathbf{x}}_k$ to $k-q$ as $\hat{\mathbf{x}}_{k-q} = \mathbf{A}^{-q}\hat{\mathbf{x}}_k$.

On a horizon $[m, n]$ of N points, from $m = n - N + 1$ to n, the UFIR filter processes N past ECG data y_n. To provide a near optimal output, the horizon length must be set optimally as N_{opt}. For the sake of best denoising with the minimum mean square error (MSE), it is also required to make the horizon adaptive around the QRS complex (Fig. 5.1).

The batch UFIR filtering estimate can be represented as follow [109],

$$\hat{\mathbf{x}}_k = \mathbf{H}_{m,k}\mathbf{Y}_{m,k} \tag{5.9a}$$

$$= (\mathbf{W}_{m,k}^T\mathbf{W}_{m,k}^T)^{-1}\mathbf{W}_{m,k}^T\mathbf{Y}_{m,k} \,, \tag{5.9b}$$

where $\mathbf{Y}_{m,k}$ represents the observation vector, $\mathbf{H}_{m,k}$ is the UFIR filter gain, and $\mathbf{W}_{m,k}$ is an

auxiliary matrix,

$$\mathbf{Y}_{m,k} = [y_m^T \; y_{m+1}^T \dots y_k^T]^T , \tag{5.10}$$

$$\mathbf{W}_{m,k} = \begin{bmatrix} \mathbf{C}(\mathcal{A}^{m+1})^{-1} \\ \mathbf{C}(\mathcal{A}^{m+2})^{-1} \\ \vdots \\ \mathbf{C}A^{-1} \\ \mathbf{C} \end{bmatrix} , \tag{5.11}$$

where the product $\mathcal{A}$ is specified as

$$\mathcal{A}_r^g = \begin{cases} \mathbf{A}^{r-g+1} , & g < r+1 , \\ \mathbf{I} , & g = r+1 , \\ \mathbf{0} , & g > r+1 . \end{cases} \tag{5.12}$$

The discrete convolution-based batch UFIR filter estimate thus appears as $\hat{\mathbf{x}}_k = \mathbf{H}_{m,k}\mathbf{Y}_{m,k}$, where the UFIR filter gain is computed by $\mathbf{H}_{m,k} = (\mathbf{W}_{m,k}^T\mathbf{W}_{m,k}^T)^{-1}\mathbf{W}_{m,k}^T = \mathbf{G}_{m,k}\mathbf{W}_{m,k}^T$, where $\mathbf{G}_{m,k}$ is the generalized noise power gain (GNPG)

$$\mathbf{G}_k = \mathbf{H}_{m,k}\mathbf{H}_{m,k}^T = (\mathbf{W}_{m,k}\mathbf{W}_{m,k})^{-1} , \tag{5.13}$$

which is responsible for an optimal balance between the regular (bias) and random errors. Provided $\hat{\mathbf{x}}_k$, the q-lag UFIR smoothed estimate appears by projecting the filtering estimate to $k-q$ as $\hat{\mathbf{x}}_{k-q} = \mathbf{A}^{-q}\hat{\mathbf{x}}_k$.

The batch estimate (5.9a) can also be computed iteratively using recursions [3, 72, 109], like in the Kalman filter. Since our concern in this work is to reduce smoothing errors, we will not consider the computational complexity of (5.9a) and postpone it to the next stage.

5.2 Optimal Horizon for UFIR smoother employing harmonic model

To achieve the best denoising effect in ECG signals, the UFIR smoother must operate on optimal averaging horizons of N_{opt} points, which can be found for the UFIR filter following a methodology worked out in [3, 4].

To specify N_{opt}, we select 10000 samples of healthy heartbeats and consider several harmonics of the Fourier series (5.2). Following [93], we compute the measurement residual as a difference

between the ECG data and the filter output, compute the mean square value (MSV) of the residual as a function of $N_{\min} \leqslant N \leqslant 10^3$, and approximate this function with a cubic polynomial as V_N. We next apply the derivative $\partial V_N/\partial N$, find its minimum, and find N_{opt} at this point for several harmonics ($m = 1, 2, 3, 5$) as shown in Fig. 5.2–Fig. 5.6. As can be seen, V_n behaves similarly for all of the harmonics selected, although the value of N_{opt} decreases from 14 for $m = 1$ to 7 for $m = 5$.

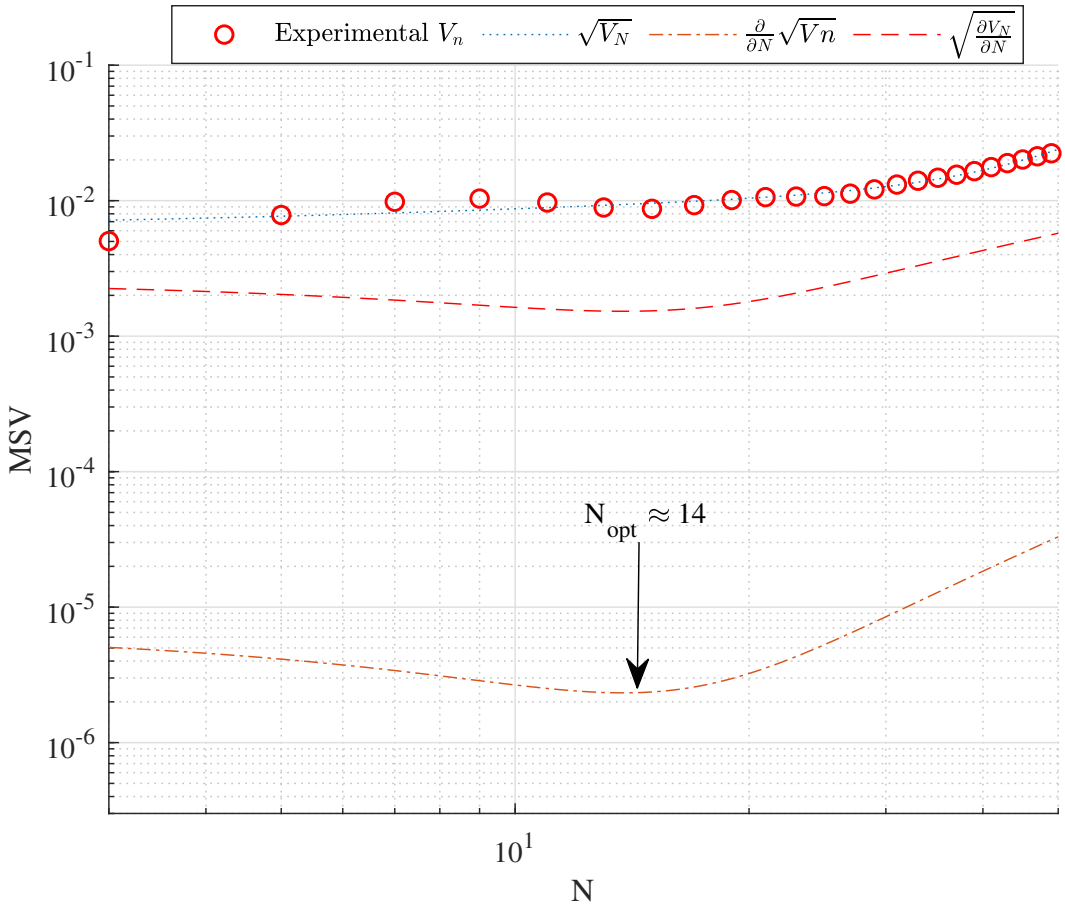

Figure 5.3: Effect of N on the MSV (circled) for $m = 1$. A cubic approximation of the MSV is $\sqrt{V_N}$ and the optimal horizon $N_{\text{opt}} = 14$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N} V_N}$.

Visually, an analysis of the smoothing errors [110] produced by the 1st and 3rd harmonics reveals no significant differences, except for the horizon length, which inherently grows with m. This fact can be explained by the observation that a lag $q = \frac{N-1}{2}$ makes the noise power gain (NPG) in both smoothers equal [69]. The role of q-lag on the NPG of the p-shift UFIR filter, $q = -p$, has been studied in [69]. The optimal shifts were implemented in [3, 4] as the p-shift 1

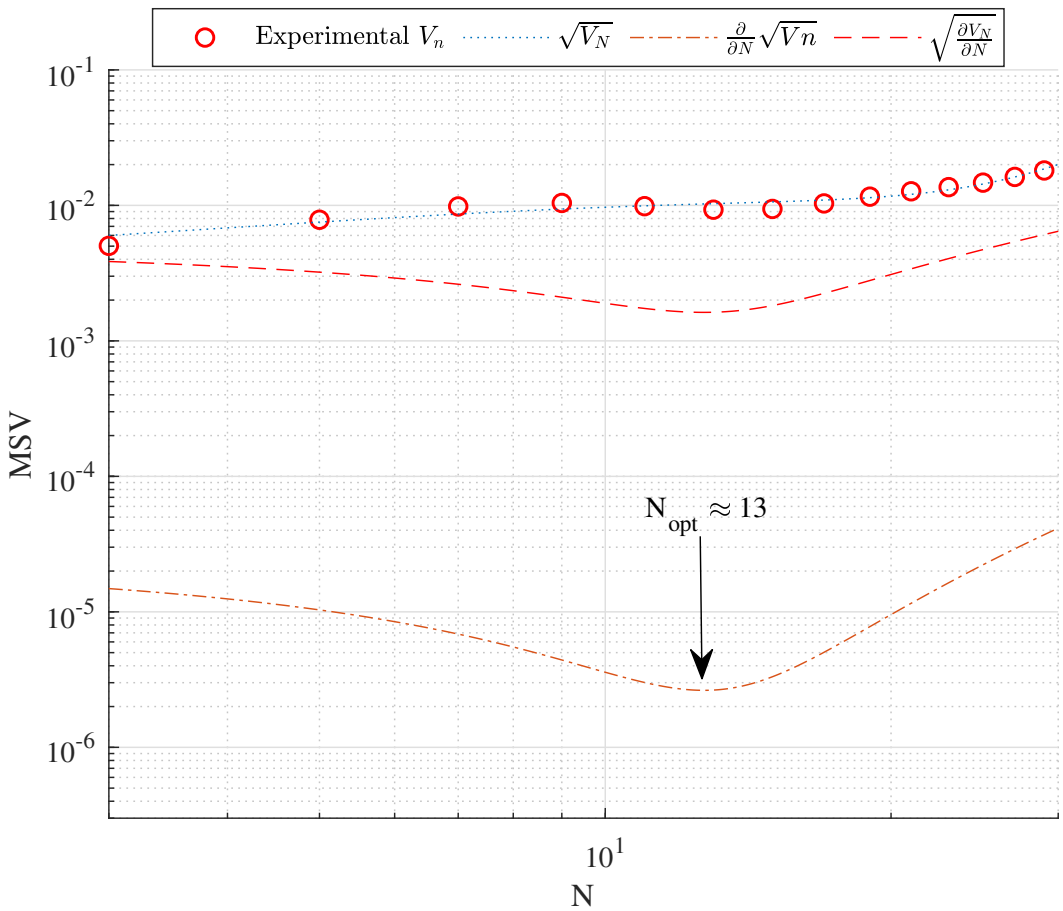

Figure 5.4: Effect of N on the MSV (circled) for $m = 2$. A cubic approximation of the MSV is $\sqrt{V_N}$ and the optimal horizon $N_{\text{opt}} = 14$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N} V_N}$.

and p-shift 2.

Given $\mathbf{A}$ by (5.7) and $\mathbf{C}$ (see Section 5.1), Fig. 5.7 sketches smoothed estimates provided using the 1st, 3rd, and 5th harmonics of an ECG signal.

It follows that the estimates are accurate in the slow part. To smooth a fast excursion around the QRS complex, an adaptive similar algorithm developed in [3, 4] is applied to avoid bias errors in ECG records.

5.3 Adaptive UFIR smoothing algorithm for ECG records

Unlike adaptive algorithm developed in [3, 4], the algorithm iterative UFIR smoothing with QRS complex correcting (see algorithm 5.3) bears the problem associated to fast excursion near to QRS

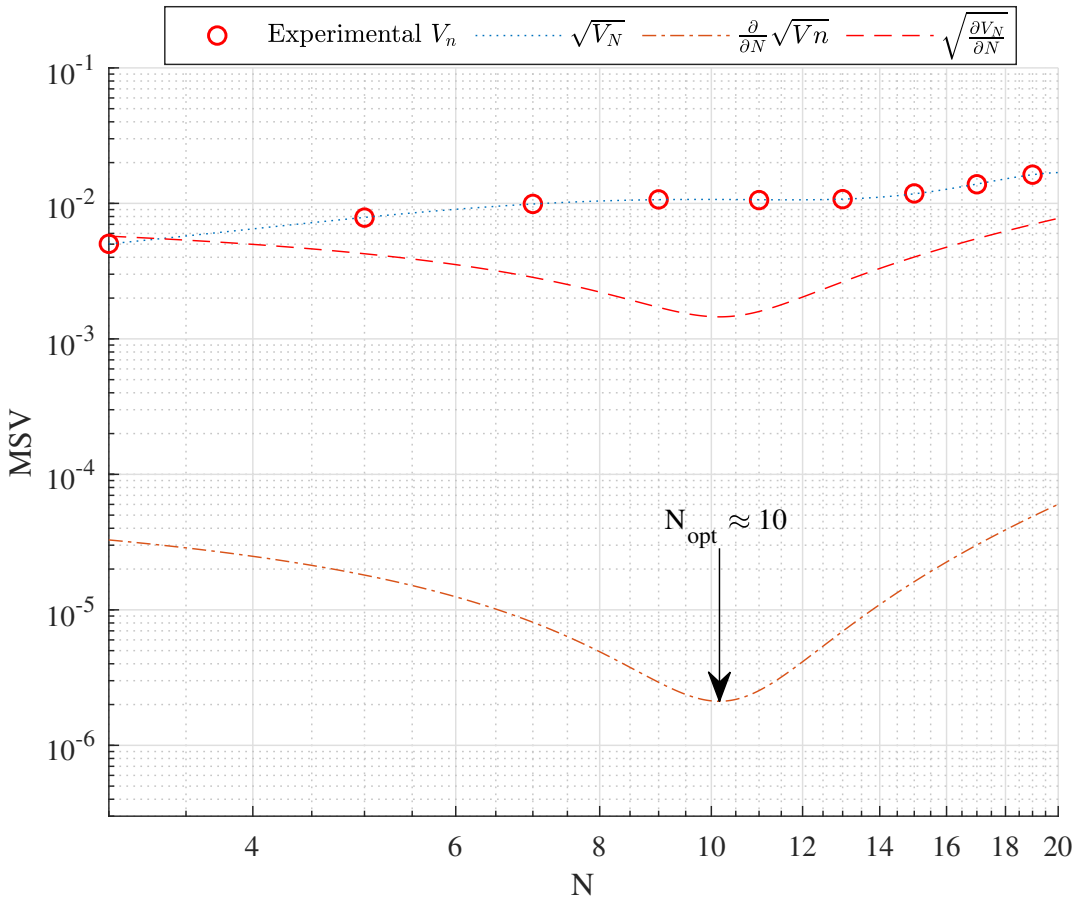

Figure 5.5: Effect of N on the MSV (circled) for $m = 3$. A cubic approximation of the MSV is $\sqrt{V_N}$ and the optimal horizon $N_{\text{opt}} = 14$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N} V_N}$.

complex by using the estimate standard deviation. We primordially apply the algorithm 4.1.4 proposed in section 4 with optimal horizon N_{opt} to ECG records. Hence considering the estimate $\mathbf{x}_{\mathbf{N_{opt}}}$, the EC measurement S_i, window parameters WindLeft and WindRight, we develop a adaptive automatic algorithm that estimates the ECG records. The algorithm receives the above parameters to calculate the estimate $\tilde{\mathbf{x}}$. Initially, we calculate the difference between the estimate $\mathbf{x}_{\mathbf{N_{opt}}}$ and measurement S_i which the residual difference is denoted by D_i. After, we create a function called **QRSCorrect** (see algorithm 5.3) which receives $\tilde{\mathbf{x}}_\mathbf{i}$, S_i, WindLeft and WindRight. We consider the residual D_i to calculate the standard deviation by **std** function. Here, we stablish frontiers conditions to correct the bias of estimate. The windows parameter WindLeft and WindRight comprehend the width of window where the bias prevails in the ECG records. Experimentally, we have consider $3rd$ harmonic, WindLeft= 2 and WindRight =5 to estimate

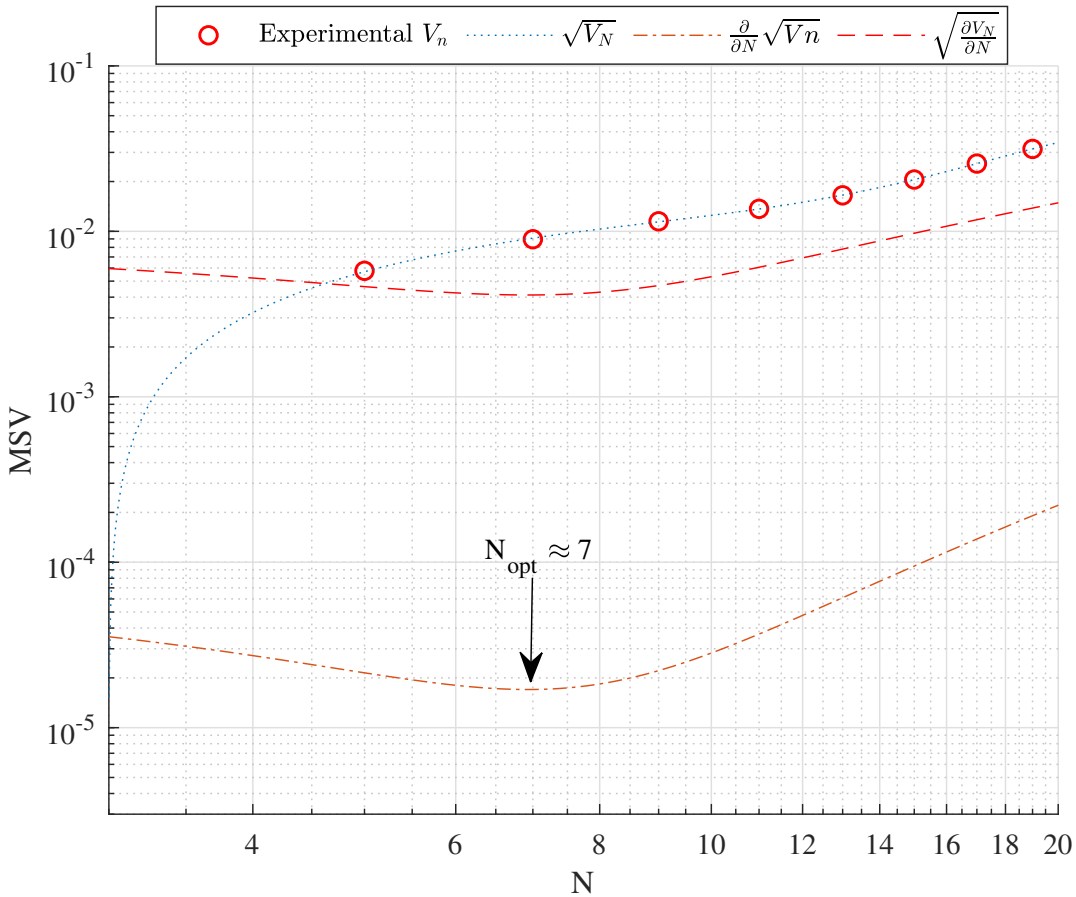

Figure 5.6: Effect of N on the MSV (circled) for $m = 5$. A cubic approximation of the MSV is $\sqrt{V_N}$ and the optimal horizon $N_{\text{opt}} = 14$ corresponds to the minimum of $\sqrt{\frac{\partial}{\partial N} V_N}$.

two ECG records each other with different pathology (see 5.7 and 5.8.)

5.4 Testing UFIR Smoother by Harmonic Model

All tests of synthetic data are provided using a special software designed on the MATLAB platform. The simulated signals employ the Fourier series assuming that data are corrupted at different levels by white Gaussian noise. Hence, we now provide an experimental test of the UFIR smoother performance by a harmonic model with the following system matrix

$$\mathbf{A} = \begin{bmatrix} \cos(\omega_1 \Delta t k) & -\sin(\omega_1 \Delta t k) \\ \sin(\omega_1 \Delta t k) & \cos(\omega_1 \Delta t k) \end{bmatrix} \tag{5.14}$$

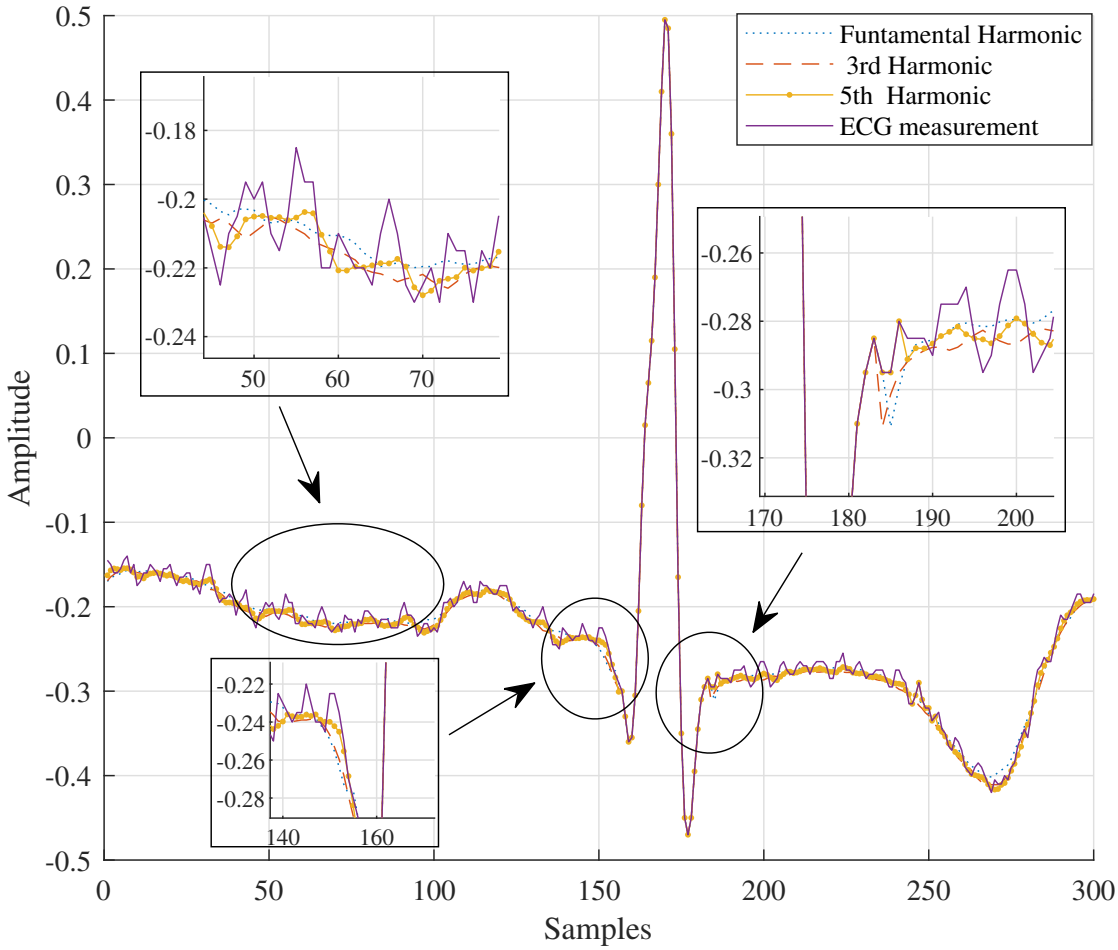

Figure 5.7: UFIR smoothing of an ECG signal using the 1st harmonic (dotted), 3rd harmonic (dashed), and 5th harmonic (solid-dotted). The ECG data are depicted with a solid line.

and observation matrix $\mathbf{C} = [1\,,0]$, where ω_1 is a chosen angular fundamental frequency. For a periodic signal $y = \cos\theta + \sin\theta$ corrupted by an additive white Gaussian noise (AWGN) with mean zero and the variance $\sigma^2 = 0.0625$, the results are sketched in Fig. 5.9. A comparison is provided with respect to the polynomial model discussed in [3, 4]. As can be seen, both UFIR smoothers are most successful in accuracy, since their estimates range most close to the generated signal. To support this conclusion, the smoother RMSEs computed over 1000 iterations are sketched in Fig. 5.10.

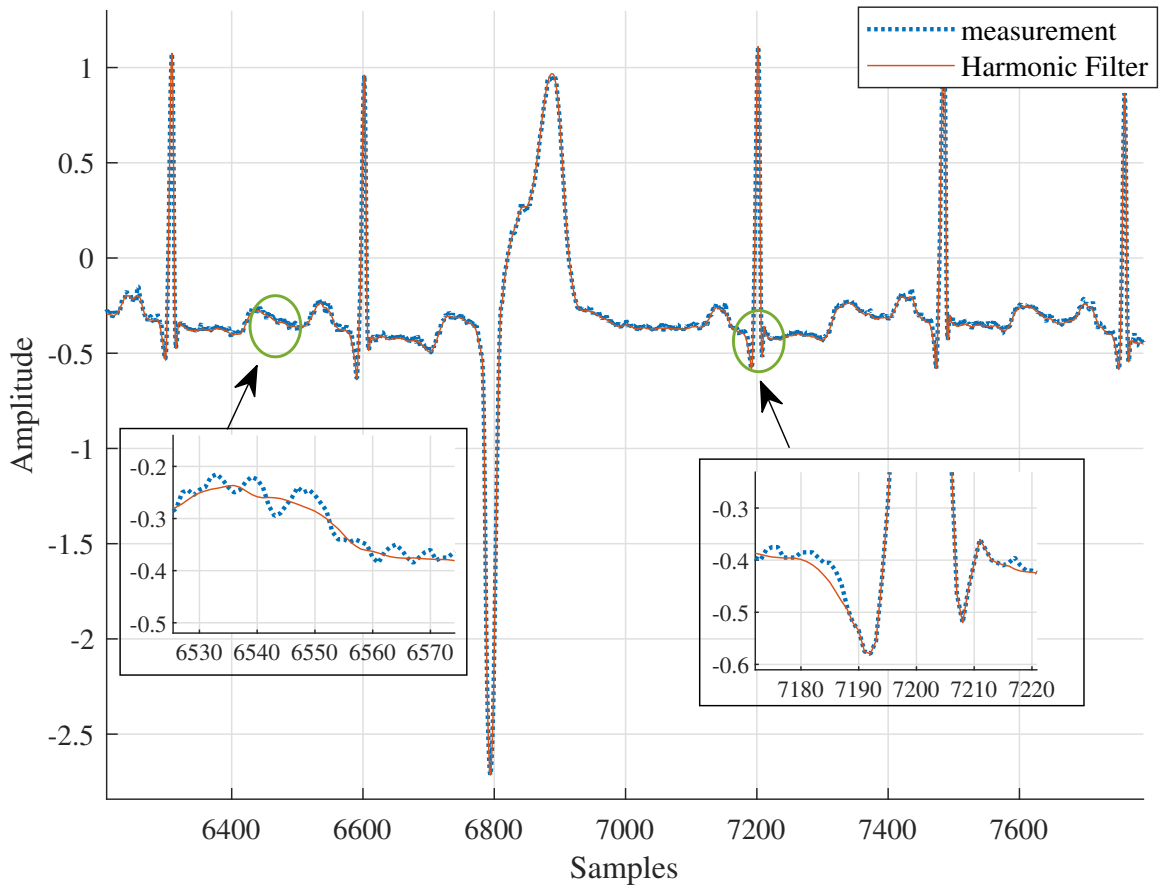

Figure 5.8: Estimation of ECG signals with single ventricular premature complex (VPC)

Algorithm 7 Automatic iterative UFIR Smoothing Algorithm for ECG records with QRS complex correcting

Data: S_i, $\hat{\mathbf{x}_{\mathbf{opt}}}$, WindLeft, WindRight

Result: $\tilde{\mathbf{x}}$

1: $\tilde{\mathbf{x}} = \hat{\mathbf{x}}_{\mathbf{N_{opt}}}$

2: **Begin** :

3: **for** $i = K + 1 :$ **length**(S_i) **do**

4: D_i=$\hat{\mathbf{x}}_{N_{opt}}(:, 1) - S_i$

5: x_est= **QRSCorrect**(S_i, $\tilde{\mathbf{x}}$, WindLeft, WindRight)

6: **end for**

Algorithm 8 QRScorrect

1: **if** $D_i > \mathbf{std}(D_i)$ || $D_i < -\mathbf{std}(D_i)$ **then**
2: **for** $j = 0 : \text{WindLeft}$ **do**
3: $\tilde{\mathbf{x}}(i-j, 1) = S_{(i-j)}$
4: **end for**
5: **for** $k = 1 : \text{WindRight}$ **do**
6: $\tilde{\mathbf{x}}(i+k, 1) = S_{(i+k)}$
7: **end for**
8: **end if**

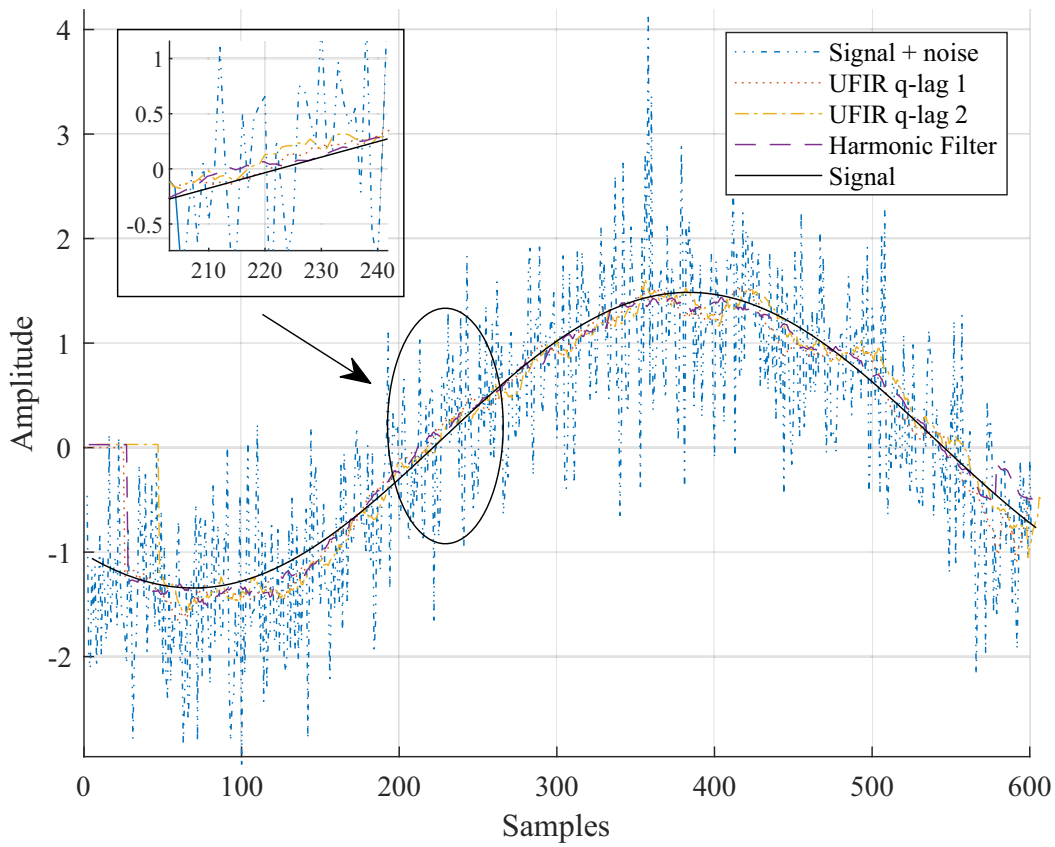

Figure 5.9: Denoising of a test harmonic signal (solid) corrupted by the AWGN using a harmonic filter (double dash-triple dots), UFIR smoother with q-lag 1 (dotted) [3] and with q-lag 2 (one dash-dotted) [4].

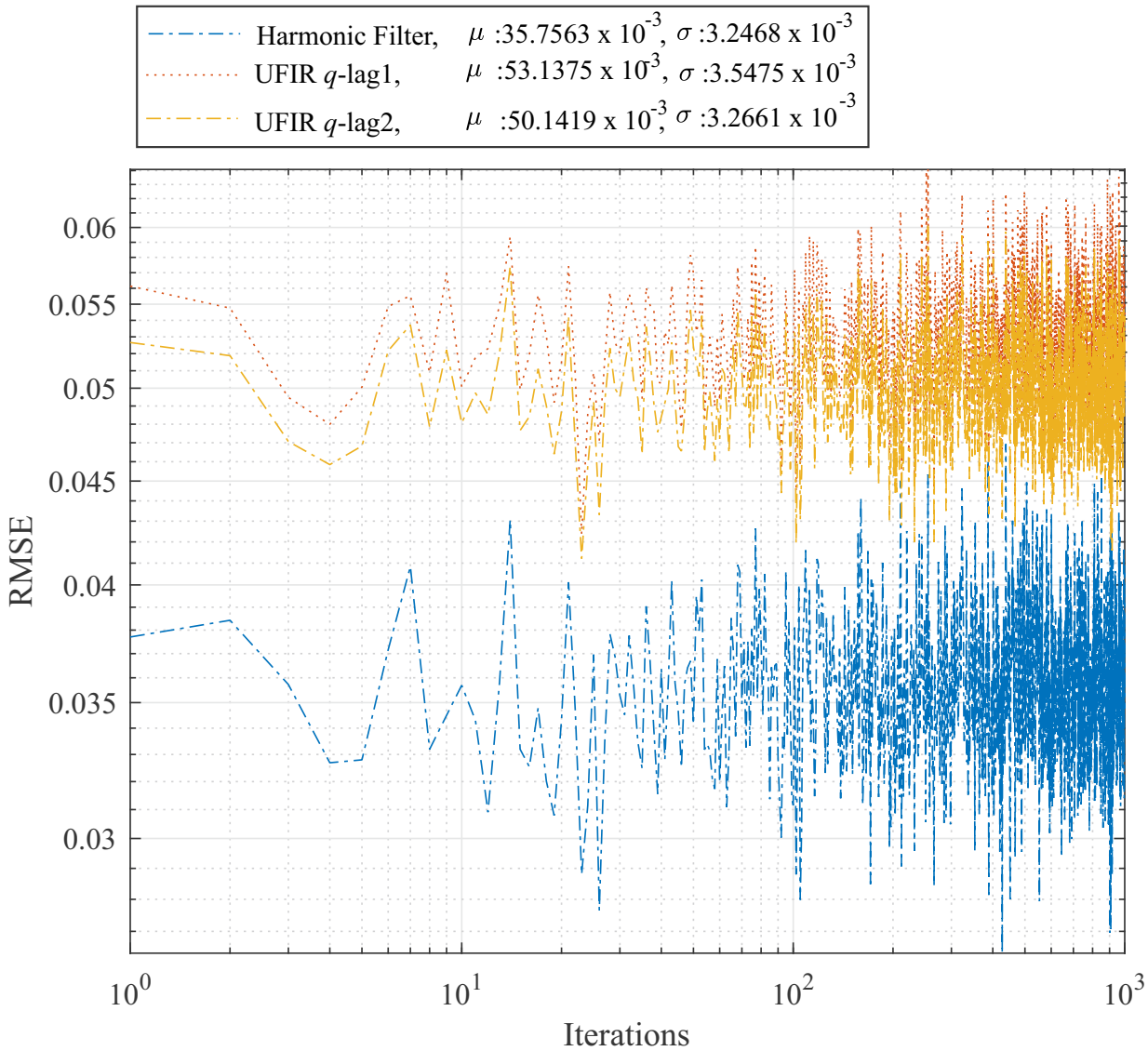

Figure 5.10: RMSEs corresponding to Fig. 5.9 as computed over 1000 iterations for the harmonic filter, UFIR q-lag 1 smoother, and UFIR q-lag 2 smoother.

5.4.1 Signal-to-Noise Ratio (SNR) Analysis

We next provide an analysis of the signal-to-noise ratios (SNRs) at the filter outputs in terms of root mean square error (RMSE). A synthetic ECG signal having characteristics similar to a real ECG is considered with known noise. For a comparison, we consider as well a polynomial model. As can be seen in Fig. 5.11, the harmonic model-based filter outperforms both the polynomial model-based UFIR smoothers.

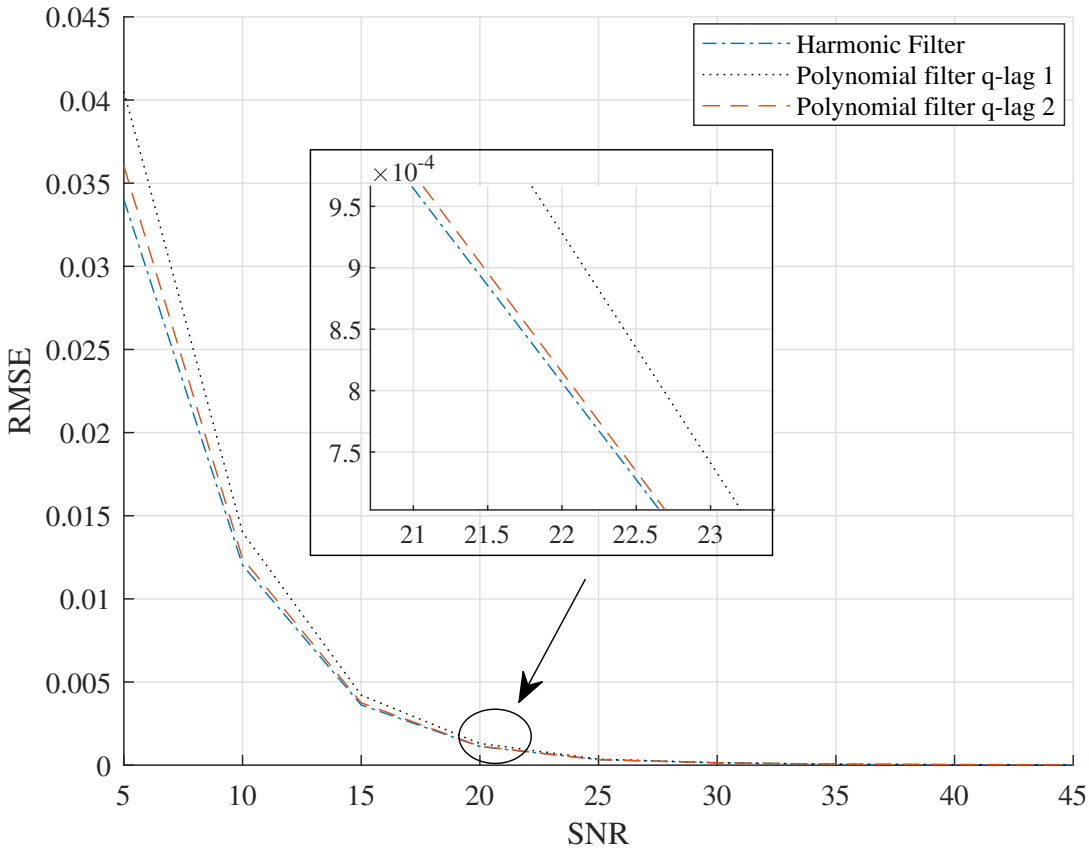

Figure 5.11: RMSEs of the UFIR smoother compared to the polynomial and harmonic model and standard filters.

5.4.2 Percentage-root-mean-square difference (PRD) Analysis

Initially, the polynomial filter q-lag 1RMSE shows to be major to polynomial filter q-lag 2 and harmonic filter RMSE. The two last mentioned filters outperform to polynomial filter q-lag. Among 20 and 25 of signal noise to ratio (SNR) the percentage root-mean-square difference represents 5% to 3% for three UFIR smoothers. However, the harmonic UFIR smoother highlights over the polynomial UFIR smoother. We depict this analysis in the figure 5.12.

5.5 Discussion

Profound learning of the electrocardiogram (ECG) signal features play a crucial role in medical sector, allowing finding cardiac disease patterns. The problem one meets here is that noise

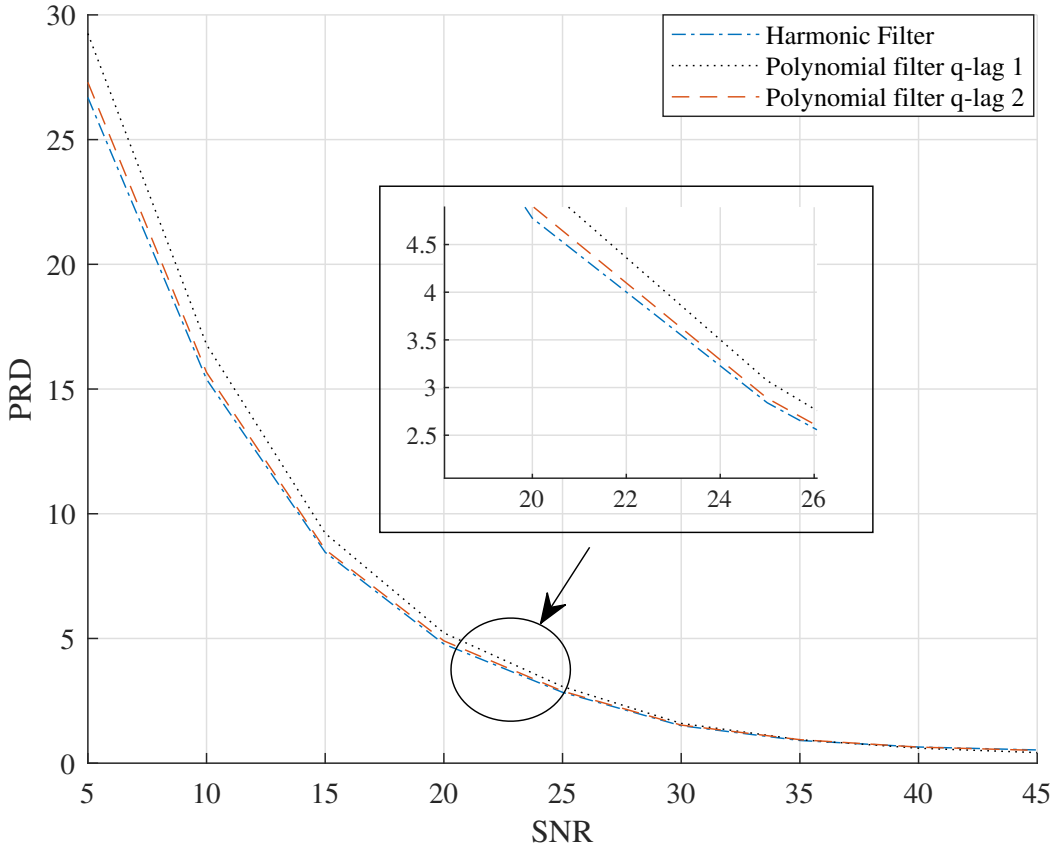

Figure 5.12: PRD of the UFIR smoother compared to the polynomial and harmonic model.

and artifacts, which are caused by the acquisition and breathe and other body affects, make an automatic detection of ECG signal patterns unreliable and modern signal processing algorithms are required. Therefore, many denoising algorithms have been designed for ECG signals [6, 7, 9] to felicitate features extraction. The standard low-pass and band-pass filters developed for ECG signals in [94] implying that an ECG signal is stationary. The approach provides a satisfactory denoising in the frequency domain, but overlooks the time resolution. This flaw was overcome in [26, 56] using the wavelet transform and a properly chosen wavelet. Yet another method based on empirical mode decomposition was developed in [59].

In many cases, denoising of ECG signals requires more accurate (optimal and robust) methods due to the not well-known origin of the heart noise and data artifacts. In this regard, smoothing techniques are recognized as most powerful to remove noise while retaining fundamental properties of ECG signals. An example is the smoothing technique developed by Savitsky and Golay (SG)

[33], which is widely applied to ECG signals [34, 35, 38]. But, as a fairly old polynomial technique, the SG smoother has two flaws: its batch form is computationally inefficient and the output is related to the middle of the averaging horizon that does not hold for all degrees.

A more general unbiased finite impulse response (UFIR) smoothing technique has been developed in [69] to generalize the SG smoother in a special case. The UFIR smoother operates in discrete-time state-space, has a fast iterative algorithm using recursions, and suggests that a suboptimal smoothing can be provided if to set an averaging horizons optimally as N_{opt} and choose an optimal lag q_{opt}. It is shown that an optimal lag q_{opt} corresponds to the middle of the averaging horizon only for odd-degree polynomials. Otherwise, q_{opt} must chosen individually for each even degree [68–70]. Note it as an essential difference with the SG smoother, which suggests the middle horizon point for all degrees.

In the chapter 3 and 4, the polynomial approach has been developed in [3, 4] for models employing the Taylor series expansion. It has been shown that the polynomial UFIR smoother produces more accuracy than other techniques. However, the polynomial model does not fit well with a quasi periodic ECG signal and further improvements have been achieved if to represent an ECG signal with the Fourier series. This deduction follows from the fact that the harmonic model is widely used in power systems, where signals are also quasi periodic [111–119]. Even thought, we find only a few works applying a harmonic model to ECG signals [120, 121].

Therefore, we have shown from RMSE and SNR analysis that the harmonic model is much more appropriate than the polynomial model for ECG signals with a significant difference of approximately 2%. This analysis is validated in Fig. 5.10 where the RMSE for Harmonic is 35.7×10^{-3} and for the q-lag polynomial model is 53.1×10^{-3}. But we limit the work by considering that the ECG signals are quasi-periodic signals. Therefore, in scenarios where the signal is aperiodic, it is not reasonable to use this model.

Chapter 6

Conclusions and future work

6.1 Conclusions

The UFIR smoothing filtering approach developed in this paper for ECG signals denoising and features extraction has demonstrated an ability to outperform the linear predictor-based on [1], which is recognized as one of the standard techniques for ECG signals. That has become possible by optimizing the order and averaging horizon for the UFIR smoothing filter in a way such that the horizon has become adaptive to different parts of ECG signals. A comparison of the UFIR predictive, filtering, and smoothing estimates has revealed a considerable difference in denoising in favor of the smoothing one. The results have also indicated that features extracted using the smoothing filter are more reliable and less prone to large deviations from average values. This is definitely an important advantage for medical needs.

The state-space UFIR smoothing approach developed in this paper for ECG signal denoising and features extraction has demonstrated better results than approaches employing the Savitsky-Golay smoother, wavelet-based filtering, and standard filters such as the low-pass, high-pass, notch, and median. That has become possible by designing an adaptive UFIR smoothing algorithm operating with optimal lags on optimal averaging horizons and approximating ECG signals with optimal degree-polynomials. Based upon this algorithm, the extracted features were evaluated by different classifiers and compared to performances provided by other methods.

Applications given for the P-wave features extraction based on detected fiducial points, has also shown a potential of the approach in a comparison with other methods. A modified UFIR smoother based on the harmonic model has been developed for ECG signals and an optimal horizon determined for the first, third, and fifth harmonics. It has been demonstrated that the

UFIR smoothers relying of the first and third harmonics produce the best denoising effect in the ECG signal. The harmonic model-based UFIR smoother also outperforms the polynomial model-based UFIR smoothers in terms of the SNR and MSE.

6.2 Future work

Although, as we have just said, we consider that the objectives of the thesis have been accomplished. There are many improvements that could be made to achieve better results. In the following section, we present for each of the aspects of the investigation carried out in this thesis, some open topics that deserve further investigation. Note that it is basically a compilation of the future work sections of each of the previous chapters (chapters 3, 4, and 5).

6.2.1 High order harmonic UFIR smoother

For the harmonic UFIR smoother, we have developed estimates in low order harmonics. However, we lack analysing the second state which means the estimated derivative, here, we could consider the derivative to increase the accuracy for feature extraction in ECG records. In addition, we plan to change the UFIR smoother approach for high order harmonics. This change could provide a far more appropriate scenario to find the optimal Nth harmonic for the ECG records.

6.2.2 Segmentation of ECG heartbeats

The ECG signals segmentation is an open problem in this investigation because of the variability provided by some arrhythmias. Here, the segmentation has been automatic by considering a window fixed that covers an ECG heartbeat. This method can work for some abnormalities when the ECG signals period does not vary significantly. But all ECG heartbeats cannot possess this same nature. There are specific cases where e.g. the P-wave disappears, making that heartbeat be drastically short on time. Hence, to use a fixed window it is not enough to solve all the problems associated with ECG signals. A heart rate variability analysis (HVRA) could be useful.

6.2.3 Pattern recognition

In this research work, we tested ECG signals features by applying some machine learning techniques. Here, the employed techniques provide favourable results. Hence, we only analysed the ECG signal in space-time, but we believe the performance of the classifiers could be better if we

continue in the exploration of extra features in different spaces. Even a mixture of temporal features with features extracted by a different space might be an interesting experiment that could provide important results. In addition, we could consider the UFIR estimate as preprocessing step to apply after deep learning technique.

6.2.4 Biomedical signals

We uniquely developed UFIR filtering for ECG signals. However, we could explore other signals with similar characteristics. To illustrate, we could find biomedical signals such as EEG and EMG in low frequency. This filter has a great potentiality in this frequency band. In this sense, we could provide promising results.

6.2.5 Hardware implementation

According to the literature review, the UFIR techniques have been developed in applications such as GPS, sensors, and biomedical signals. Initially, the UFIR filter required high memory as a batch structure. As time passed, a new UFIR filter with iterative structure was developed which its required memory depends on the optimal horizon. This is a promising scenario that motivates us to implement this filter in cost medium or low electronic devices.

Bibliography

[1] John Makhoul. Stable and Efficient Lattice Methods for Linear Prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(5):423–428, 1977.

[2] P. W. Macfarlane *et al.* *Comprehensive Electrocardiology.* Springer, London, 2010.

[3] Carlos Lastre-Domínguez, Yuriy S Shmaliy, Oscar Ibarra-Manzano, Jorge Munoz-Minjares, and Luis J. Morales-Mendoza. ECG Signal Denoising and Features Extraction Using Unbiased FIR Smoothing. *BioMed Research Intern.*, 2019:1–16, 2019.

[4] Carlos Lastre-Dominguez, Yuriy S Shmaliy, Oscar Ibarra-Manzano, and Miguel Vazquez-Olguin. Denoising and Features Extraction of ECG Signals in State Space Using Unbiased FIR Smoothing. *IEEE Access*, 7:152166–152178, 2019.

[5] World Health Organisation. About cardiovascular diseases, *https://www.who.int/cardiovascular_diseases/about_cvd/en/*, 2011.

[6] Manuel Blanco-Velasco, Binwei Weng, and Kenneth E. Barner. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine*, 38(1):1–13, 2008.

[7] Reza Sameni, Mohammad B. Shamsollahi, Christian Jutten, and Gari D. Clifford. A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Trans. Biomed. Eng.*, 54(12):2172–2185, 2007.

[8] Lena Biel, Ola Pettersson, Lennart Philipson, and Peter Wide. ECG analysis: a new approach in human identification. *IEEE Trans. Instrum and Meas.*, 50(3):808–812, 2001.

[9] Lukas Smital, V Martin, Jiri Kozumplık, and Ivo Provaznık. Adaptive wavelet Wiener filtering of ECG signals. *IEEE Trans. Biomed. Eng.*, 60(2):437–445, 2013.

[10] Roberto Olivera Reyna, Jorge Ulises Muñoz Minjares, Osbaldo Vite Chávez, Reyne Olivera Reyna, and José Manuel Cervantes Viramontes. Numerical comparison between three finite impulse response (FIR) filters in the heart rate Estimation Problem in an ECG Signal Corrupted by Additive White Gaussian Noise. In *First International Congess On Instrumentation and Applied Sciences*, pages 1–7, Cancun, Mexico, 2010.

[11] W Lu, H Hou, and J Chu. Feature fusion for imbalanced ECG data analysis. *Biomedical Signal Processing and Control*, 41:152–160, 2018.

[12] Tohid Yousefi Rezaii, Soosan Beheshti, Mahdi Shamsi, and Siavash Eftekharifar. ECG signal compression and denoising via optimum sparsity order selection in compressed sensing framework. *Biomedical Signal Processing and Control*, 41:1–27, 2018.

[13] K.N.V.P.S. Rajesh and R. Dhuli. Classification of ECG heartbeats using nonlinear decomposition methods and support vector machine. *Computers in Biology and Medicine*, 87:271–284, 2017.

[14] Yan Sun, Kap Luk Chan, and Shankar Muthu Krishnan. ECG signal conditioning by morphological filtering. *Computers in Biology and Medicine*, 32(6):465–479, 2002.

[15] Mahsa Akhbari, Mohammad B. Shamsollahi, Omid Sayadi, Antonis A. Armoundas, and Christian Jutten. ECG segmentation and fiducial point extraction using multi hidden Markov model. *Computers in Biology and Medicine*, 79(June):21–29, 2016.

[16] João Rodrigues, David Belo, and Hugo Gamboa. Noise detection on ECG based on agglomerative clustering of morphological features. *Computers in Biology and Medicine*, 87:322–334, 2017.

[17] A.A. Suárez-León, C. Varon, R. Willems, S. Van Huffel, and C.R. Vázquez-Seisdedos. T-wave end detection using neural networks and Support Vector Machines. *Computers in Biology and Medicine*, 96:116–127, 2018.

[18] Donald P. Golden, Roger A. Wolthuis, and G. W. Hoffler. A spectral analysis of the normal resting electrocardiogram. *IEEE Trans. Biomed. Eng.*, BME-20(5):366–372, 1973.

[19] C Li, C Zheng, and C Tai. Detection of ECG characteristic points using wavelet transforms. *IEEE Trans. Biomed. Eng.*, 42(1):21–28, 1995.

[20] Maxime Yochum, Charlotte Renaud, and Sabir Jacquir. Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. *Biomedical Signal Processing and Control*, 25:46–52, 2016.

[21] Zahia Zidelmal, Ahmed Amirou, Mourad Adnane, and Adel Belouchrani. QRS detection based on wavelet coefficients. *Computer Methods and Programs in Biomedicine*, 107(3):490–496, 2012.

[22] Brij N. Singh and Arvind K. Tiwari. Optimal selection of wavelet basis function applied to ECG signal denoising. *Digital Signal Processing*, 16(3):275–287, 2006.

[23] Kandala N V P S Rajesh and Ravindra Dhuli. Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. *Biomedical Signal Processing and Control*, 41:242–254, 2018.

[24] Peyman Ghobadi Azbari, Saeed Mohaqeqi, Niloofar Ghanbarzadeh Gashti, and Mohammad Mikaili. Introducing a combined approach of empirical mode decomposition and PCA methods for maternal and fetal ECG signal processing. *The Journal of Maternal-Fetal & Neonatal Medicine*, 29(19):3104–3109, 2016.

[25] Manpreet Kaur and A. S. Arora. Classification of ECG signals using LDA with factor analysis method as feature reduction technique. *Journal of Medical Engineering & Technology*, 36(8):411–420, 2012.

[26] Roshan Joy, U Rajendra Acharya, and Lim Choo. Biomedical Signal Processing and Control Technical note ECG beat classification using PCA , LDA , ICA and Discrete Wavelet Transform. *Biomedical Signal Processing and Control*, 8(5):437–448, 2013.

[27] Dhouha Rezgui and Zied Lachiri. Ecg biometric recognition using svm-based approach. *IEEJ Trans. on Electrical and Electronic Eng.*, 11:S94–S100, 2016.

[28] Pepino A. Sansone M., Fusco R. and Sansone C. Electrocardiogram pattern recognition and analysis based on artificial neural networks and support vector machines: A review. *Journal of Healthcare Eng.*, 4(4):465–504, 2013.

[29] K P Lin and W H Chang. QRS feature extraction using linear prediction. *IEEE Trans. Biomed. Eng.*, 36(10):1050–1055, 1989.

[30] Z. E. Hadj Slimane and F. Bereksi Reguig. Detection of the QRS complex by linear prediction. *Journal of Medical Engineering and Technology*, 30(3):134–138, 2006.

[31] Roshan Joy Martis, Chandan Chakraborty, and Ajoy K Ray. A two-stage mechanism for registration and classification of ECG using Gaussian mixture model. *Pattern Recognit.*, 42(11):2979–2988, 2009.

[32] Roshan Joy Martis, U. Rajendra Acharya, and Hojjat Adeli. Current methods in electrocardiogram characterization. *Computers in Biology and Medicine*, 48(1):133–149, 2014.

[33] Abraham. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964.

[34] S. Hargittai. Savitzky-Golay least-squares polynomial filters in ECG signal processing. *Computers in Cardiology*, 32:763–766, 2005.

[35] Sunder Ram Krishnan and Chandra Sekhar Seelamantula. On the selection of optimum Savitzky-Golay filters. *IEEE Trans. Signal Process.*, 61(2):380–391, 2013.

[36] Yibin Hong and Yong Lian. A memristor-based continuous-time digital FIR filter for biomedical signal processing. *IEEE Trans. Circuits. and Syst. I, Reg. Papers*, 62(5):1392–1401, 2015.

[37] Ivan A Dotsinsky and Georgy S Mihov. Tremor suppression in ECG. *Biomedical engineering online*, 7:29, 2008.

[38] Reza Sameni. Online filtering using piecewise smoothness priors: Application to normal and abnormal electrocardiogram denoising. *Signal Processing*, 133(October 2016):52–63, 2017.

[39] Cemal Kavalcioglu Berk Dagman and Glu. Filtering maternal and fetal savitzky-golay filter and adaptive least mean. *Mathematics, Informatics, Physics,*, 9(2):109–124, 2016.

[40] Ho-Ming Su, Tsung-Hsien Lin, Po-Chao Hsu, Wen-Hsien Lee, Chun-Yuan Chu, Chee-Siong Lee, Wen-Ter Lai, Sheng-Hsiung Sheu, and Wen-Chol Voon. Incremental prognostic value of identifying mitral L wave in patients with atrial fibrillation. *Int. J. Cardiology*, 168(4):4501–4503, 2013.

[41] M. Sabarimalai Manikandan and K. P. Soman. A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomed. Signal Process. Control*, 7(2):118–128, 2012.

[42] Federica Censi, Ivan Corazza, Elisa Reggiani, Giovanni Calcagnini, Eugenio Mattei, Michele Triventi, and Giuseppe Boriani. P-wave Variability and Atrial Fibrillation. *Nature Publishing Group*, pages 1–7, 2016.

[43] Raúl Alcaraz, Fernando Hornero, and José Joaquín Rieta. Dynamic time warping applied to estimate atrial fibrillation temporal organization from the surface electrocardiogram. *Medical Engineering Physics*, 35(9):1341–1348, sep 2013.

[44] K. Padmavathi and K. Sri Ramakrishna. Classification of ECG Signal during Atrial Fibrillation Using Autoregressive Modeling. *Procedia Computer Science*, 46:53–59, 2015.

[45] Ambika Annavarapu and Padmavathi Kora. ECG-based atrial fibrillation detection using different orderings of Conjugate Symmetric–Complex Hadamard Transform. *Int. J. Cardiovascular Academy*, 2(3):151–154, sep 2016.

[46] K Tateno and L Glass. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and Δ RR intervals. 39:664–671, 2001.

[47] Hongqiang Li, Xiuli Feng, Lu Cao, Enbang Li, Huan Liang, and Xuelong Chen. A New ECG Signal Classification Based on WPD and ApEn Feature Extraction. *Circuits, Systems, Signal Process.*, 35(1):339–352, 2016.

[48] Hongqiang Li, Huan Liang, Chunjiao Miao, Lu Cao, Xiuli Feng, Chunxiao Tang, and Enbang Li. Novel ECG Signal Classification Based on KICA Nonlinear Feature Extraction. *Circuits, Systems, and Signal Process.*, 35(4):1187–1197, 2016.

[49] Hongqiang Li, Danyang Yuan, Xiangdong Ma, Dianyin Cui, and Lu Cao. Genetic algorithm for the optimization of features and neural networks in ECG signals classification. *Scientific Reports*, 7:1–12, January 2017.

[50] Hongqiang Li, Xiaofei Wang, Lei Chen, and Enbang Li. Denoising and R-peak detection of electrocardiogram signal based on EMD and improved approximate envelope. *Circuits, Systems, Signal Process.*, 33(4):1261–1276, 2014.

[51] Hongqiang Li and Xiaofei Wang. Detection of electrocardiogram characteristic points using lifting wavelet transform and Hilbert transform. *Trans. Inst. Measur. Control*, 35(5):574–582, 2013.

[52] Guoqiang Han and Zhijun Xu. Electrocardiogram signal denoising based on a new improved wavelet thresholding. *Review Sci. Instrum.*, 87(8), 2016.

[53] Choujun Zhan, Lam Fat Yeung, and Zhi Yang. A wavelet-based adaptive filter for removing ECG interference in EMGdi signals. *J. Electromyography Kinesiology*, 20(3):542–549, 2010.

[54] Md Ashfanoor Kabir and Celia Shahnaz. Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains. *Biomed. Signal Process. Control*, 7(5):481–489, 2012.

[55] Kevin Kærgaard, Søren Hjøllund Jensen, and Sadasivan Puthusserypady. A comprehensive performance analysis of EEMD-BLMS and DWT-NN hybrid algorithms for ECG denoising. *Biomed. Signal Process. Control*, 25:178–187, 2016.

[56] Supriya Goel, Pradeep Tomar, and Gurjit Kaur. An optimal wavelet approach for ECG noise cancellation. *Int. J. Bio-Science Bio-Technology*, 8(4):39–52, 2016.

[57] Jianbo Gao, Hussain Sultan, Jing Hu, and Wen Wen Tung. Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison. *IEEE Signal Process. Let.*, 17:237–240, March, 2010.

[58] Santosh Kumar Yadav, Rohit Sinha, and Prabin Kumar Bora. Electrocardiogram signal denoising using non-local wavelet transform domain filtering. *IET Signal Process.*, 9(June 2014):88–96, 2016.

[59] Salim Lahmiri. Comparative study of ECG signal denoising by wavelet thresholding in empirical and variational mode decomposition domains. *Healthcare Technology Letters*, 1(3):104–109, 2014.

[60] Regis Nunes Vargas and Antônio Cláudio Paschoarelli Veiga. Electrocardiogram signal denoising by clustering and soft thresholding. *IET Signal Process.*, 12(9):1165–1171, 2018.

[61] Manuel García, Miguel Martínez-Iniesta, Juan Ródenas, José J. Rieta, and Raúl Alcaraz. A novel wavelet-based filtering strategy to remove powerline interference from electrocardiograms with atrial fibrillation. *Physiological Measur.*, 39(11), 2018.

[62] Matteo D'Aloia, Annalisa Longo, and Maria Rizzi. Noisy ECG signal analysis for automatic peak detection. *Information (Switzerland)*, 10(2):1–12, 2019.

[63] Wei Li. Wavelets for electrocardiogram: Overview and taxonomy. *IEEE Access*, 7:25627–25649, Nov., 2019.

[64] Riqing Chen, Yingsong Huang, and Jian Wu. Multi-window detection for P-wave in electrocardiograms based on bilateral accumulative area. *Computers in Biology and Medicine*, 78(January):65–75, 2016.

[65] Runnan He, Kuanquan Wang, Na Zhao, Yang Liu, Yongfeng Yuan, Qince Li, and Henggui Zhang. Automatic Detection of Atrial Fibrillation Based on Continuous Wavelet Transform and 2D Convolutional Neural Networks. *Frontiers in Physiology*, 9(AUG):1–11, aug 2018.

[66] Jinseok Lee, Bersain A. Reyes, David D. McManus, Oscar Maitas, and Ki H. Chon. Atrial Fibrillation Detection Using an iPhone 4S. *IEEE Trans. Biomed Eng.*, 60:203–206, jan Jan., 2013.

[67] U. Maji, M. Mitra, and S. Pal. Automatic Detection of Atrial Fibrillation Using Empirical Mode Decomposition and Statistical Approach. *Procedia Technology*, 10:45–52, 2013.

[68] Yuriy S. Shmaliy. Suboptimal FIR filtering of nonlinear models in additive white gaussian noise. *IEEE Transactions on Signal Processing*, 60(10):5519–5527, oct 2012.

[69] Yuriy S Shmaliy and Luis J. Morales-Mendoza. FIR smoothing of discrete-time polynomial signals in state space. *IEEE Trans. Signal Process.*, 58(5):2544–2555, 2010.

[70] Yuriy S Shmaliy, Oscar Ibarra-Manzano, Luis Arceo-Miquel, and Jorge Munoz-Diaz. A thinning algorithm for GPS-based unbiased FIR estimation of a clock TIE model. *Measurement*, 41(5):538–550, 2008.

[71] Yuriy S Shmaliy, Yrjö Neuvo, and Sanowar Khan. Review of Unbiased FIR Filters , Smoothers , and Predictors for Polynomial Signals. *Frontiers in Signal Processing*, 2(1):1–29, 2018.

[72] Yuriy S Shmaliy, Shunyi Zhao, and Choon K I Ahn. Unbiased FIR filtering: an iterative alternative to Kalman filtering ignoring noise and initial conditions. *IEEE Control Syst. Mag.*, 37:70–89, Oct., 2017.

[73] S. Zhao, Y. S. Shmaliy, and F. Liu. Fast Kalman-like optimal unbiased FIR filtering with applications. *IEEE Trans. on Signal Process.*, 64:2284–2297, May, 2016.

[74] S. Zhao and Y. S. Shmaliy. Unified maximum likelihood form for bias constrained FIR filters. *IEEE Signal Process. Lett.*, 23(12):1848–1852, Dec. 2016.

[75] G. B. Moody and R. G. Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.*, 20(3):45–50, 2001.

[76] Ary L Goldberger, Luis A N Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-kang Peng, and H Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101:215–220, 2000.

[77] A. L. Goldberger *et al.* *Goldberger's clinical electrocardiography: a simplified approach.* Elsevier, 8th ed.,, Philadelphia, 2013.

[78] Juan Sebastian Arteaga Falconi, Hussein Al Osman, and Abdulmotaleb El Saddik. ECG Authentication for Mobile Devices. *IEEE Trans. Instrum. Meas.*, 65(3):591–600, 2016.

[79] Maryamsadat Hejazi, S.A.R. Al-Haddad, Yashwant Prasad Singh, Shaiful Jahari Hashim, and Ahmad Fazli Abdul Aziz. ECG biometric authentication based on non-fiducial approach using kernel methods. *Digital Signal Processing*, 52:72–86, 2016.

[80] Jin Wang, Mary She, Saeid Nahavandi, and Abbas Kouzani. Human identification from ECG signals via sparse representation of local segments. *IEEE Signal Process. Lett.*, 20(10):937–940, 2013.

[81] Francesco Gargiulo, Antonio Fratini, Mario Sansone, and Carlo Sansone. Subject identification via ECG fiducial-based systems: Influence of the type of QT interval correction. *Comput. Methods and Programs in Biomedicine*, 121(3):127–136, 2015.

[82] Steven A. Israel, John M. Irvine, Andrew Cheng, Mark D. Wiederhold, and Brenda K. Wiederhold. ECG to identify individuals. *Pattern Recognit.*, 38(1):133–142, 2005.

[83] Mario Sansone, Antonio Fratini, Mario Cesarelli, Paolo Bifulco, Alessandro Pepino, Maria Romano, Francesco Gargiulo, and Carlo Sansone. Influence of QT correction on temporal and amplitude features for human identification via ECG. In *IEEE Workshop on Biometric*

Measurements and Systems for Security and Medical Applications, BioMS - Proceedings, pages 22–27, 2013.

[84] Ramalingam B. Rao PB, Singh N. Dexmedetomidine - induced atrial premature complex. *Ann Card Anaesth*, 19:347–350, 2016.

[85] Michael L. Armstrong. *Electrocardiograms: A systematic method of reading them.* Stanford California, United States, second edition, 1972.

[86] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*, volume 37. Prentice Hall, nov 1995.

[87] Yuriy S Shmaliy. An unbiased p-step predictive FIR filter for a class of noise free discrete-time models with independently observed states. *Signal, Image and Video Processing*, 3(2):127–135, 2009.

[88] L Morales-Mendoza, H Gamboa-Rosales, and Yuriy S Shmaliy. A new class of discrete orthogonal polynomials for blind fitting of finite data. *Signal Processing*, 93(7):1785–1793, 2013.

[89] H Gamboa-Rosales, L Morales-Mendoza, and Yuriy S Shmaliy. Unbiased impulse responses – A class of discrete orthogonal polynomials. *ICIC Express Letters*, 7(7):2005–2010, 2013.

[90] Miguel Vazquez-Olguin, Yuriy S. Shmaliy, and Oscar G. Ibarra-Manzano. Distributed unbiased FIR filtering with average consensus on measurements for WSNs. *IEEE Trans. Ind. Informat.*, 13(3):1440–1447, Jun 2017.

[91] Miguel Vazquez-Olguin, Yuriy S Shmaliy, Choon Ki Ahn, and Oscar G. Ibarra-Manzano. Blind robust rstimation with missing data for smart sensors using UFIR filtering. *IEEE Sensor Journal*, 17(6):1819–1827, 2017.

[92] Carlos Lastre Dominguez, Yuriy S. Shmaliy, Oscar Ibarra Manzano, and Luis J. Morales Mendoza. Unbiased FIR denoising of ECG signals. In *14th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, pages 1–6. IEEE, oct 2017.

[93] Felipe Ramirez-Echeverria, Amadou Sarr, and Yuriy S. Shmaliy. Optimal memory for discrete-time FIR filters in state-space. *IEEE Transactions on Signal Processing*, 62(3):557–561, feb 2014.

[94] Jiapu Pan and Willis J. Tompkins. A Real-Yime QRS Detection Algorithm. *IEEE Trans Biomed. Eng.*, BME-32(3):230–236, 1985.

[95] Can Ye, B. V K Vijaya Kumar, and Miguel Tavares Coimbra. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Trans Biomed. Eng.*, 59(10):2930–2941, 2012.

[96] Simon Haykin. *Communication Systems*. Wiley, 2001.

[97] Yuriy S. Shmaliy. Limiting phase errors of passive wireless SAW sensing with differential measurement. *IEEE Sensors Journal*, 4(4):819–827, 2004.

[98] G. Stock. Rcian Envelope Estimation and Confidence Intervals in Low Signal/Noise Levels. *IEEE Electronics Letters*, 23(16):832–834, 1987.

[99] Y. S. Shmaliy. An unbiased FIR filter for TIE model of a local clock in applications to GPS-based timekeeping. *IEEE Trans. Ultrason., Ferroelect., and Freq. Control*, 53(5):862–869, 2006.

[100] Idoia Beraza and Iñaki Romero. Comparative study of algorithms for ECG segmentation. *Biomed. Signal Process. Control*, 34:166–173, 2017.

[101] Eduardo Jose Luz, David Menotti, and William Robson Schwartz. Evaluating the use of ECG signal in low frequencies as a biometry. *Expert Systems with Applications*, 41(5):2309–2315, apr 2014.

[102] Javier Aspuru, Alberto Ochoa-Brust, Ramón Félix, Walter Mata-López, Luis Mena, Rodolfo Ostos, and Rafael Martínez-Peláez. Segmentation of the ECG Signal by Means of a Linear Regression Algorithm. *Sensors*, 19(4):775, feb 2019.

[103] Foo, Jong. Yong A. Comparison of wavelet transformation and adaptive filtering in restoring artefact-induced time-related measurement. *Biomedical Signal Processing and Control*, 1(1):93–98, 2006.

[104] Eduardo José Luz, William Robson Schwartz, Guillermo Cámara-Chávez, and David Menotti. ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine*, 127:144–164, 2016.

[105] Zachary D Ary L. Goldberger, Goldberger and Alexei Shvilkin. *Goldberger's Clinical Electrocardiography*. Philadelphia, PA, United States, eigh edition, 2013.

[106] Jocelyne Fayn, Paul Rubel, and Peter W. Macfarlane. Can the lessons learned from the assessment of automated electrocardiogram analysis in the Common Standards for quantitative Electrocardiography study benefit measurement of delayed contrast-enhanced magnetic resonance images? *Journal of Electrocardiology*, 40(3):246–250, 2007.

[107] Ivan K. Daskalov and Ivaylo I. Christov. Automatic detection of the electrocardiogram T-wave end. *Medical and Biological Engineering and Computing*, 37(3):348–353, 1999.

[108] Yan Sun, Kap Luk Chan, and Shankar Muthu Krishnan. Characteristic wave detection in ECG signal using morphological transform. *BMC Cardiovascular Disorders*, 5:1–7, 2005.

[109] Yuriy S. Shmaliy, Shunyi Zhao, and Choon Ki Ahn. Unbiased finite impluse response filtering: an iterative alternative to Kalman filtering ignoring noise and initial conditions. *IEEE Control Systems Mag.*, 37(5):70–89, oct 2017.

[110] Carlos Lastre-Dominguez. Unbiased FIR denoising of ECG signals. In *2017 14th Int. Conf. Elect. Eng., Comp. Sci. Auto. Control (CCE)*, number 1, pages 1–6. IEEE, oct 2017.

[111] Michael Sheng-Pu Hwang and Alan R. Wood. Harmonic state-space modelling of a controlled HVdc converter. *Electric Power Systems Research*, 124:65–73, jul 2015.

[112] Robert W. Stark. Bistability, higher harmonics, and chaos in AFM. *Materials Today*, 13(9):24–32, sep 2010.

[113] Jordan Rel C. Orillaza and Alan R. Wood. Harmonic State-Space Model of a Controlled TCR. *IEEE Transactions on Power Delivery*, 28(1):197–205, jan 2013.

[114] Bo-kyu Kwon, Soohee Han, and Kwang Lee. Robust Estimation and Tracking of Power System Harmonics Using an Optimal Finite Impulse Response Filter. *Energies*, 11(7):1811, jul 2018.

[115] Jing Lyu, Marta Molinas, and Xu Cai. Harmonic State Space Modeling of a Three-Phase Modular Multilevel Converter. In *2018 IEEE Int. Power Electro. and App. Conf. Exp. (PEAC)*, number June, pages 650–655, 2018.

[116] Nicolas *et. al.* Blin. A comparison of harmonic modeling methods with application to control of switched systems with active filtering. In *2019 18th European Control Conf., ECC 2019*, pages 4198–4203, 2019.

[117] J.J. Rico, Manuel Madrigal, and Enrique Acha. Dynamic harmonic evolution using the extended harmonic domain. *IEEE Trans. Power Delivery*, 18(2):587–594, 2003.

[118] Butt, Maryam *et. al.* Power Line Interference tracking in ECG signal using State Space RLS. In *Proc. 2013 IEEE 8th Conf. Ind. Electro. App., ICIEA 2013*, pages 211–215. IEEE, 2013.

[119] Nauman Razzaq, Shafa-At Ali Sheikh, Muhammad Salman, and Tahir Zaidi. An Intelligent Adaptive Filter for Elimination of Power Line Interference From High Resolution Electrocardiogram. *IEEE Access*, 4:1676–1688, 2016.

[120] Amjad Zia Khan and Imran Shafi. Removing Artifacts from Raw Electrocardiogram Signals Using Adaptive Filter in State Space. *Circuits, Syst., and Signal Process.*, (May), may 2019.

[121] *et. al.* Sang Hwan Park. Short-time harmonic analysis via the state-space optimal FIR filter. In *SICE '95. Proceedings of the 34th SICE Annual Conference. International Session Papers*, pages 1205–1210. Soc. Instrum. & Control Eng, 1995.

Appendix A

Low-Degree UFIR Functions $h_{li}(N,p)$

A.0.1 Ramp, $l = 1$:

$$h_{1i}(p) = a_{01}(N,p) + a_{11}(N,p)i\,, \tag{A.1}$$

where

$$a_{01}(N,p) = \frac{2(2N-1)(N-1) + 12p(N-1+p)}{N(N^2-1)}\,, \tag{A.2}$$

$$a_{11}(N,p) = \frac{6(N-1+2p)}{N(N^2-1)}\,. \tag{A.3}$$

A.0.2 Quadratic, $l = 2$:

$$h_{2i}(p) = a_{02}(N,p) + a_{12}(N,p)i + a_{22}(N,p)i^2\,, \tag{A.4}$$

where

$$a_{02}(N,p) = 3\frac{\begin{array}{l} 3N^4 - 12N^3 + 17N^2 - 12N + 4 \\ + 12(N-1)(2N^2-5N+2)p + 12(7N^2-15N+7)p^2 \\ + 120(N-1)p^3 + 60p^4 \end{array}}{N(N^2-1)(N^2-4)}\,, \tag{A.5}$$

$$a_{12}(N,p) = -18\frac{\begin{array}{l} 2N^3 - 7N^2 + 7N - 2 \\ + 2(7N^2-15N+7)p + 30(N-1)p^2 + 20p^3 \end{array}}{N(N^2-1)(N^2-4)}\,, \tag{A.6}$$

$$a_{22}(N,p) = 30\frac{N^2 - 3N + 2 + 6(N-1)p + 6p^2}{N(N^2-1)(N^2-4)}\,. \tag{A.7}$$

Appendix B

UFIR Filter for state space

B.1 Extended model in state space

Let me consider the model following,

$$\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{E}_n\mathbf{U}_n + \mathbf{B}_n\mathbf{w}_n, \tag{B.1}$$

$$y_n = \mathbf{C}\mathbf{x}_n + v_n\,, \tag{B.2}$$

Then, a state space transformation is given by,

$$\begin{aligned}
\mathbf{x}_n &= \mathbf{A}_n\mathbf{x}_{n-1} + \mathbf{E}_n\mathbf{U}_n + \mathbf{B}_n\mathbf{w}_n \\
\mathbf{x}_{n-1} &= \mathbf{A}_{n-1}\mathbf{x}_{n-2} + \mathbf{E}_{n-1}\mathbf{U}_{n-1} + \mathbf{B}_{n-1}\mathbf{w}_{n-1} \\
\mathbf{x}_{n-2} &= \mathbf{A}_{n-2}\mathbf{x}_{n-3} + \mathbf{E}_{n-2}\mathbf{U}_{n-2} + \mathbf{B}_{n-2}\mathbf{w}_{n-2} \\
&\vdots \\
\mathbf{x}_{m+2} &= \mathbf{A}_{m+1}\mathbf{x}_{m+1} + \mathbf{E}_{m+2}\mathbf{U}_{m+2} + \mathbf{B}_{m+2}\mathbf{w}_{m+2} \\
\mathbf{x}_{m+1} &= \mathbf{A}_{m+1}\mathbf{x}_{m} + \mathbf{E}_{m+1}\mathbf{U}_{m+1} + \mathbf{B}_{m+1}\mathbf{w}_{m+1} \\
\mathbf{x}_{m} &= \mathbf{A}_{m}\mathbf{x}_{m} + \mathbf{E}_{m}\mathbf{U}_{m} + \mathbf{B}_{m}\mathbf{w}_{m}
\end{aligned} \tag{B.3}$$

Hence,

$$
\begin{aligned}
\mathbf{x}_n &= \mathbf{A}_n\mathbf{A}_{n-1}\mathbf{x}_{n-1} + \mathbf{A}_n\mathbf{E}_{n-1}\mathbf{U}_{n-1} + \mathbf{A}_n\mathbf{B}_{n-1}W_{n-1} + \mathbf{E}_n\mathbf{U}_n + \mathbf{B}_n\mathbf{w}_n \\
\mathbf{x}_n &= \mathbf{A}_n\mathbf{A}_{n-1}\mathbf{A}_{n-2}\mathbf{x}_{n-3} + \mathbf{A}_n\mathbf{A}_{n-1}\mathbf{E}_{n-2}\mathbf{U}_{n-2} + \mathbf{A}_n\mathbf{A}_{n-1}\mathbf{B}_{n-2}\mathbf{W}_{n-2} \\
&\quad + \mathbf{A}_n\mathbf{E}_{n-1}\mathbf{U}_{n-1} + \mathbf{A}_n\mathbf{B}_{n-1}\mathbf{w}_{n-1} + \mathbf{E}_n\mathbf{U}_n + \mathbf{B}_n\mathbf{w}_n \\
\mathbf{x}_n &= \mathbf{A}_n\mathbf{A}_{n-1}\ldots\mathbf{A}_{m+1}\mathbf{x}_m + \mathbf{A}_n\mathbf{A}_{n+1}\ldots\mathbf{A}_{m+2}\mathbf{E}_{m+1}\mathbf{U}_{m+1} + \ldots \\
&\quad + \mathbf{A}_n\mathbf{E}_{n-1}\mathbf{U}_{n-1} + \mathbf{E}_n\mathbf{U}_n + \mathbf{A}_n\mathbf{A}_{n-1}\ldots\mathbf{A}_{m+2}\mathbf{B}_{m+1}\mathbf{w}_{m+1} + \ldots\mathbf{A}_n\mathbf{B}_{n-1}\mathbf{w}_{n-1} + \mathbf{B}_n\mathbf{w}_n
\end{aligned}
\tag{B.4}
$$

Then, each matrix in general form can be represented as follow,

$$
\mathbf{X}_{m,n} = \begin{bmatrix} \mathbf{x}_m \\ \mathbf{x}_{m+1} \\ \vdots \\ \mathbf{x}_n \end{bmatrix}, \ \mathbf{U}_{m,n} = \begin{bmatrix} \mathbf{U}_m \\ \mathbf{U}_{m+1} \\ \vdots \\ \mathbf{U}_n \end{bmatrix}, \ \mathbf{W}_{m,n} = \begin{bmatrix} \mathbf{w}_m \\ \mathbf{w}_{m+1} \\ \vdots \\ \mathbf{w}_n \end{bmatrix}. \tag{B.5}
$$

Considering the augmented matrix,

$$
\mathcal{A}_r^g = \begin{cases} \mathbf{A}_n, \mathbf{A}_{r-1}\ldots\mathbf{A}_g, & g < r+1 \\ \mathbf{I}, & g = r+1 \\ 0 & g > r+1 \end{cases} \tag{B.6}
$$

$$
\mathbf{F}_{m,n} = \begin{bmatrix} \mathbf{I} \\ \mathbf{A}^{m+1} \\ \vdots \\ \mathcal{A}_n^{m+1} \end{bmatrix}, \tag{B.7}
$$

$$
\mathbf{S}_{m,n} = \begin{bmatrix} \mathbf{E}_m & 0 & \ldots & 0 & 0 \\ \mathbf{A}^{m+1}\mathbf{E}_m & \mathbf{E}_{m+1} & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{A}_{n-1}^{m+1}\mathbf{E}_m & \mathcal{A}_{n-1}^{m+2}\mathbf{E}_{m+1} & \ldots & \mathbf{E}_{n-1} & 0 \\ \mathcal{A}_n^{m+1}\mathbf{E}_n & \mathcal{A}_{n-1}^{m+2}\mathbf{E}_{m+1} & \ldots & \mathcal{A}_n^{m+1}\mathbf{E}_{n-1} & \mathbf{E}_n \end{bmatrix}, \tag{B.8}
$$

$$\mathbf{D}_{m,n} = \begin{bmatrix} \mathbf{B}_m & 0 & \dots & 0 & 0 \\ \mathbf{A}^{m+1}\mathbf{B}_m & \mathbf{B}_{m+1} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{A}_{n-1}^{m+1}\mathbf{B}_m & \mathcal{A}_{n-1}^{m+2}\mathbf{B}_{m+1} & \dots & \mathbf{B}_{n-1} & 0 \\ \mathcal{A}_{n}^{m+1}\mathbf{B}_n & \mathcal{A}_{n-1}^{m+2}\mathbf{B}_{m+1} & \dots & \mathcal{A}_{n}^{m+1}\mathbf{B}_{n-1} & \mathbf{B}_n \end{bmatrix}, \tag{B.9}$$

Hence, we have a general definition of the model in state space,

$$\mathbf{X}_{m,n} = \mathbf{F}_{m,n}\mathbf{x}_m + \mathbf{S}_{m,n}\mathbf{U}_{m,n} + \mathbf{D}_{m,n}\mathbf{W}_{m,n} \tag{B.10}$$

$$\mathbf{Y}_{m,n} = \mathbf{H}_{m,n}\mathbf{x}_m + L_{m,n}\mathbf{U}_{m,n} + \mathbf{G}_{m,n}\mathbf{W}_{m,n} + \mathbf{V}_{m,n} \tag{B.11}$$

The measurement $\mathbf{y}$ have the definition following,

$$\begin{aligned} \mathbf{y}_n &= \mathbf{C}_n\mathbf{x}_n + \mathbf{v}_n \\ \mathbf{y}_{n-1} &= \mathbf{C}_{n-1}\mathbf{x}_{n-1} + \mathbf{v}_{n-1} \\ \mathbf{y}_{n-2} &= \mathbf{C}_{n-2}\mathbf{x}_{n-2} + \mathbf{v}_{n-1} \\ &\vdots \\ \mathbf{y}_m &= \mathbf{C}_m\mathbf{x}_m + \mathbf{v}_m \end{aligned} \tag{B.12}$$

Where,

$$\mathbf{Y}_{m,n} = \begin{bmatrix} \mathbf{y}_m \\ \mathbf{y}_{m+1} \\ \vdots \\ \mathbf{y}_n \end{bmatrix}, \ \mathbf{V}_{m,n} = \begin{bmatrix} \mathbf{v}_m \\ \mathbf{v}_{m+1} \\ \vdots \\ \mathbf{v}_n \end{bmatrix}, \tag{B.13}$$