# Identifying Outliers in HRV–Seizure Signals using $p$-shift UFIR Baseline Estimates

JORGE MUNOZ–MINJARES, YURIY SHMALIY, MISAEL LOPEZ–RAMIREZ, JORGE M. CRUZ–DUARTE
Universidad de Guanajuato
Deparment of Electronic Engineering
36885, Salamanca, Gto
MEXICO
ju.munozminjares, shmaliy@ugto.mx

*Abstract:* Heart rate variability (HRV) is typically associated with neuroautonomic activity and viewed as a major non–invasive tool to detect seizures. The HRV has been assumed and analyzed as a stationary signal. However the presence of seizures can violate estimates of statistical parameters and conventional techniques intended to remove outliers can be inaccurate. A useful approach implies setting thresholds to compute the first and third quartiles from histogram data or residuals based on the estimated baseline. In this paper, we propose an accurate method to identify outliers in HRV measurements with partial epilepsy retaining relevant information. The baseline perturbed by the seizure in the HRV data is removed using the $p$-shift unbiased finite impulse response (UFIR) smoothing filter operating on optimal horizons. The residuals histogram is plotted and the upper bound (UB) and lower bound (LB) are computed as thresholds. A comparison is provided of a typical points detected in HRV/seizures based on several methods used to estimate the baseline. A time/frequency analysis is supplied to show the difference between the raw HRV and HRV without outliers. The method proposed is tested by partial seizures records taken from patients during continuous EEG/ECG and video monitoring.

*Key–Words:* Heart Rate Variability (HRV), outliers, seizures, $p$-shift (UFIR) smoothing filter.

## 1 Introduction

The term 'heart rate variability' (HRV) has been accepted to describe variations of both instantaneous heart rate and RR intervals, where R is a point corresponding to the peak of the QRS complex of the electrocardiogram (ECG) wave and RR is the interval between successive Rs [1]. Because the HRV is strongly coupled with the autonomic nervous system (ANS) [2], information from several physiological signals such as ECG is also used to detect seizures [3, 4]. Epileptic seizures are taken as indicators, which enable the diagnostic of ANS disorders. The HRV spectral content has frequency components, which are classified in three main peaks: very low frequency (VLF) $< 0.04$ Hz, low frequency (LF), $0.04 - 0.15$ Hz, and high frequency (HF) $0.15 - 0.4$ Hz [5]. The LF and HF powers are associated with sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) activity [6], respectively.

One of the main and oldest problems in biomedical signal processing is the noise reduction without disturbing the original signal, especially in ECG having a high content of vital information [7]. Special components of the disturbance, which immerse in noise signals, are outliers which affect significantly the performance of denoising algorithms. The outlier is defined as the measurement, which significantly deviates from the logical or normal pattern of sensed data [8]. Similarly, in statistics, an outlier is defined as a 'case that does not follow the same model as the rest of the data' [9].

There are many methods to remove outliers from data. The most simple and useful one is known as the Boxplot. This method displays batches of data [10] and only five values from a set are used: the extremes, the upper and lower hinges (quartiles), and the median [11]. However, this method is inefficient when data have non-stationary characteristics in some parts. Nevertheless, an increasingly common process to remove outliers is to estimate a model using several method, such as smoothing algorithms [12, 13, 14]. Complete the smoothing, the outliers are excluded at an arbitrary distance greater than $A\sigma$ from the model, where $A$ can be specified by the user. Because there are many algorithms to smooth bio-signals, the selection of an ideal method faces difficult. Furthermore,

errors in the baseline optimal estimates of HRV signals strongly depend on the smoother degree.

It is now worth mentioning that an ideal baseline for HRV signals does not exist and it is far less available in the presence of partial epilepsy. Therefore, an accurate method is required to estimate the optimal baseline that helps identifying the outliers in HRV measurements. One approach used in medical applications is based on the $p$-shift unbiased finite impulse response (UFIR) filter, which minimizes the mean square error (MSE) on optimal averaging horizons $N_{\rm opt}$. The $p$–shift UFIR filter does not require unavailable prior knowledge about the ECG signal statistics, but produces suboptimal estimates when $N_{\rm opt}$ is set properly. In [15], two ways have been proposed to find $N_{\rm opt}$: a) by minimizing the trace of the error covariance matrix and b) by minimizing the derivative of the trace of the mean square value (MSV) associated with the raw signal and estimate as function of $N$. Because an exact mathematical model of the HRV signal is not available, the only option is to find $N_{\rm opt}$ using the second approach.

In this paper, we develop an approach to estimate the baseline model of HRV with partial seizures and identify the outliers by employing the $p$-shift UFIR filter. We also compare the number of atypical points based on the residuals calculated using several algorithms: LOESS, RLOESS, LOWESS, Moving Average, and Spline. To test the algorithms, we use measurement data from Post-Ictal Heart Rate Oscillations in Partial Epilepsy available at MIT-BIH [17]. The paper is structured as follows. In Section 2, a brief description is given for the HRV based on the electrocardiogram model and database features used in this work. In Section 3, we design the proposed algorithm to estimate the baseline of HRV signals and develop an algorithm to calculate $N_{\rm opt}$. In Section 4, we find $N_{\rm opt}$ and identify the outliers using the $p$-shift UFIR filter and some other algorithms. Finally, conclusions and further work are discussed in Section 5.

## 2 HRV from Partial Seizures

In a typical ECG signal, a mixture of deflections Q, R, and S generated by the heart's muscles form a complex QRS. The RR interval, called tachogram, is obtained by computing the distance between the peaks R immerses in the complex QRS. The HRV is then represented with the instantaneous heart rate (IHR), which is the inverse of the RR interval [18].

This work is based on data of HRV measurements with partial seizures, which represent the shape alteration causing perturbations in the baseline. The HRV data are analyzed from partial seizures recorded in 5 female patients during continuous EEG/ECG and video monitoring considered as a *gold standard* [17]. The heart beat annotations were obtained using fully automated methods described in [20] and available in 'physionet'.

## 3 Baseline Estimate

The baseline estimate can be obtained using diverse FIR filtering algorithms [19]. However, smoothing allows for higher accuracy when the solution is not required in real time, as in our case. Therefore, we will base our investigation on the $p$-shift UFIER filter, which is briefly described below.

### 3.1 $p$-Shift UFIR Filtering.

Let us represent the ECG measurements as $s_n = x_n + \nu_n$, where $n$ is the discrete time index, $x_n$ is the ECG signal and $\nu_n$ is the random noise, which can be supposed to be white Gaussian. Assuming that $x_n$ is the baseline and $\nu_n$ is the measurement residual, we can find $\nu_n = s_n - x_n$ based on an accurate approximation of $x_n$. The residual $\nu_n$ contains outliers in white Gaussian noise environment. Now, $x_n$ can be expressed with a degree polynomial on a horizon $[n-N+1-p, n-p]$ of $N$ points, where $p \leqslant 0$ is a time–shift. The model can be estimated using the $p$-shift UFIR filtering algorithm proposed in [21]. In the convolution-based form, the estimate of $x_n$ can be found via data taken from $[n-N+1-p, n-p]$ as

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

where $h_{ln}(p) \triangleq h_{ln}(N,p)$ is the $\{N,p\}$-variant impulse response of the $l$-degree discrete FIR filter.

Satisfied the unbiasedness condition of $E\{\hat{x}_{n|n-p}\} = E\{x_n\}$, the $p$–variant $h_{li}(p)$ can be represented as [22, 21]

$$h_{li}(p) = \sum_{j=0}^{l} a_{jl}(p) i^j\,, \tag{2}$$

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

$$a_{jl}(p) = (-1)^j \frac{M_{(j+1)1}(p)}{|\mathbf{D}(p)|}\,, \tag{3}$$

where $|\mathbf{D}(p)|$ is the determinant of matrix $\mathbf{D}(p) = \mathbf{V}^{\mathbf{T}}(p)\mathbf{V}(p)$, in which $\mathbf{V}(p)$ is the $(l+1)\times(l+1)$ Vandermonde matrix. The region of existence of $h_{li}(p)$ is

the following [21, 22],

$$h_{li}(N,p)=\begin{cases}\text{nontrivial}, & p\le i\le N-1+p\\ 0, & \text{otherwise}\end{cases}. \tag{4}$$

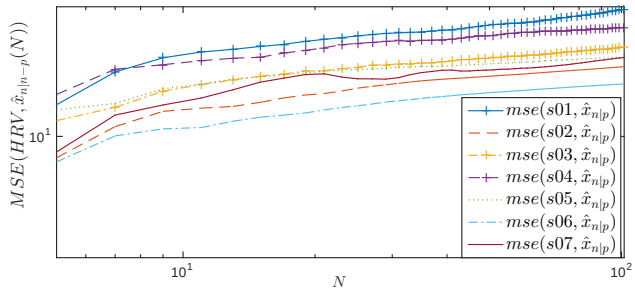
Although the $p$-shift filtering estimate can be provided for any UFIR filter degree, the low-degree filters have found most applications owing to better noise reduction.

### 3.1.1 Optimimal Horizon

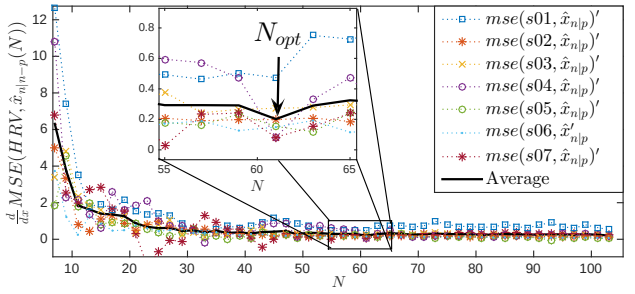
The most important parameter of the UFIR filter is the optimal horizon $N_{\rm opt}$. Because the mathematical model of HRV and baseline signal is not available, we find $N_{\rm opt}$ follow the method proposed in [15, 16]. The horizon $N_{\rm opt}$ can be estimated from the measurement residual $s_n-\hat{x}_{n|n-p}(N)$ by minimizing the derivative $\partial\eta(N)/\partial N$, where $\eta$ is the MSV $\eta(N)=E\{[s_n-\hat{x}_{n|n-p}(N)]^2\}$, and solving the optimization problem [15]

$$\hat{N}_{\rm opt}=\arg\min_N\frac{\partial\eta(N)}{\partial N}+1\,. \tag{5}$$

Provided $N_{\rm opt}$, the baseline of HRV van be estimates using the $p$-shift UFIR filtering approach.




(a)




(b)

Figure 1: Estimation of optimal horizon as $N_{\rm opt}=61$: a) MSE of HRV/seizures data (s01, s02,..., s07) for the estimate $x_{n|n-p}(N)$ and b) minimization of $\partial\eta(N)/\partial N$ to solve the optimization problem (4).

## 4 Baseline Estimates and Outliers Detection

In order to estimate the baseline of HRV with partial seizures, $N_{\rm opt}$ is required. Provided $N_{\rm opt}$, the shift $p$ can be specified for each degree as shown in [21]. Following the suggestion given in [16], we select $l=2$ and specify $p$ by $p=-(N_{\rm opt}-1)/2$. Using the physionet database, we next compute the MSE of HRVseizures (s01, s02, ..., s07) for the estimate $x_{n|n-p}(N)$ as function of $N[1:2:105]$. The results are plotted in Fig. 1a.

Next, we solve the optimization problem (5) and find $N_{\rm opt}=61$ as shown in Fig.1b. Typical values are labeled when the point is larger than the upper bound $UB=Q_3+w(Q_3-Q_1)$ or smaller than the lower bound $LB=Q_1-w(Q_3-Q_1)$, where $Q_1$ and $Q_3$ are specified in the $25\%$ and $75\%$ sense, respectively. The value $w=1.5$ corresponds to $\approx\pm2.7\sigma$ and the probability of $0.993$ for normal distributed data.

The baseline for each measurement can now be obtained using $N_{\rm opt}$ to calculate the residuals and detect outliers for the given upper and lower bounds. In Fig. 2a we show a signal s07 (doted) and the baseline estimated (solid) along with the the data residual (dashed). The normalized histogram (H1) of sample s07 and the outliers identified using the boxplot method are plotted in Fig. 2b and Fig. 2c, respectively. Build similarly, the histogram (H2) of the residuals and its typical values are sketched in Fig. 2d and Fig. 2e.

An comparison is next provided of the baseline estimates obtained by the UFIR smoothing filter and by several other techniques such as the moving average filter (MAF), Splines, LOESS, LOWESS, and RLOESS. The LOWESS and LOESS methods employ the locally weighted linear regression to smooth data using linear polynomial and quadratic polynomial respectively. A robust version of the LOESS is called RLOESS. Taking as a benchmark the baseline estimates sketched in Fig. 3a and Fig. 3b, we show the outliers identified during the seizure of signal s07. As can be observed, the loss of information using LOESS, RLOESS, and LOWESS is greater than by others methods.

Finally, Table 1 generalizes the results. The MAF ansd SPLINE algorithms identify the minimal number of outliers, while the LOESS, RLOESS, and LOWES ones the maximum number. Unlike in other methods, the number of points detected by the $p$-shift UFIR smoothing filter is located in the middle of data. Herewith, the identification based on the boxplot of raw

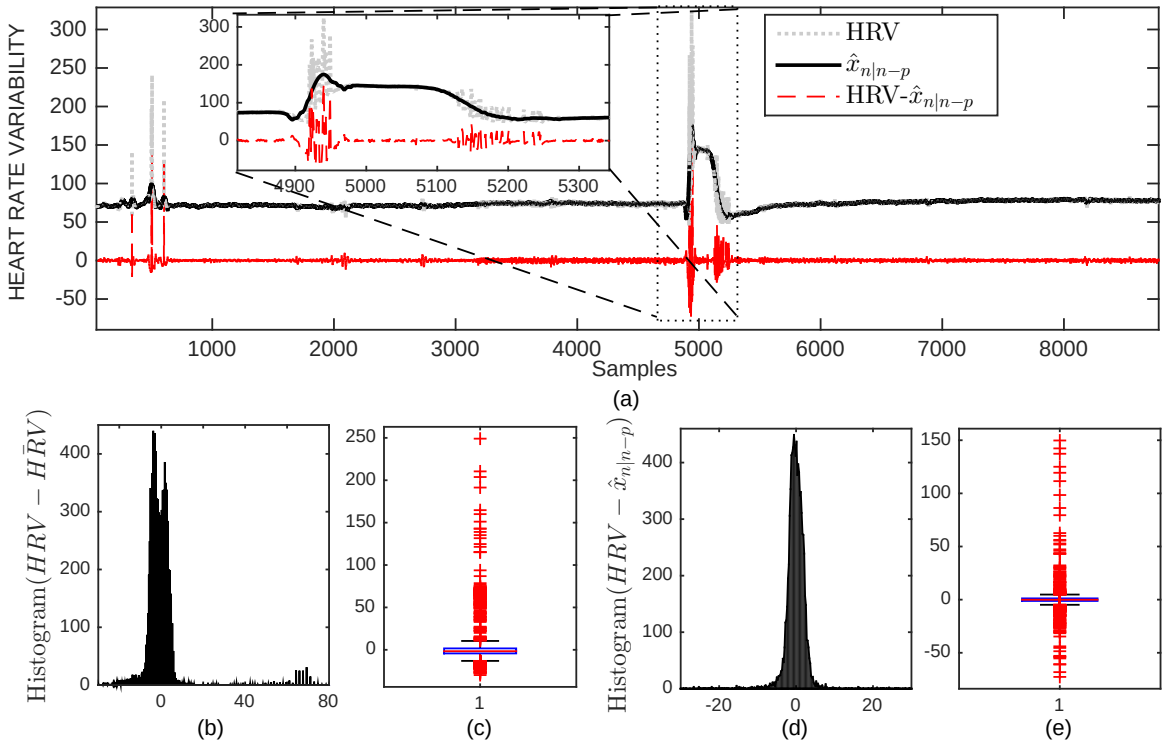



Figure 2: Baseline estimation and outliers detection using the $p$-shift UFIR filter with $N_{\text{opt}} = 61$ (solid) from HRV seizures measurements s07 (doted) and residuals (dashed): (a) baseline estimates; (b)–(e) histograms (H1) and typical points for raw HRV with an extracted average.

HRV seizures has demonstrated an erratic behavior. It is also worth mentioning that the MSEs produced by these models is proportional to the number of the identified outliers.

## 4.1 Time/Frequency Comparison

We finish our investigation by using the time/frequency (TF) technique to analyze a signal in the time and frequency domains simultaneously. We provide it for the normalized sample frequency $Fs = 1$ and a 'flattop' window using the command 'tftb_window' from Matlab. A TF comparison of HRV in *Pre* and *Post* –Seizure is showed in Fig. 4a and Fig. 4b. In this example, data correspond to a 48-year-old woman with partial epilepsy. The *Pre*-Seizure ($< 0$min) demonstrates a relative HF of $0.15 - 0.4$ Hz, while in Post-Seizure ($> 3$min), we have indicated a frequency transition between 3 and $4.5$ min. Removed the outliers, the mean difference between the raw HRV and HRV is in the lapse seizure and we notice the frequency weakness from Fig. 4a to Fig. 4b. However, the principal peaks power remain almost intact.

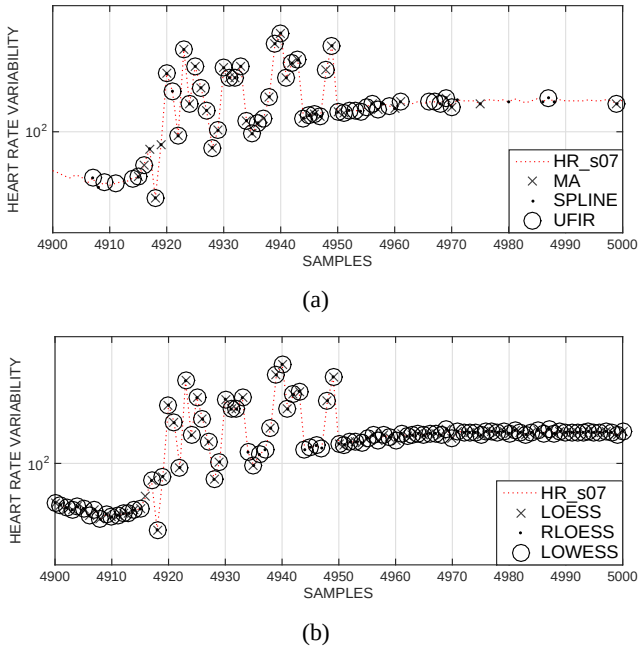



Figure 3: Atypical points identified at the beginning of epileptic attack in s07 data and outliers found with estimated baseline using: (a) MA (cross), Spline (point) and $p$-shift UFIR (circle) and (b) LOESS (cross), RLOESS (point) and LOWESS (circle) algorithms.

Table 1: Outliers identified in HRV data in the presence of seizures from s01 to s07 using different methods estimating the baseline. The 'Normal' column means atypical values localized with the boxplot technique applied to raw HRV with an extracted average.

| | OUTLIERS BY METHOD | | | | | | |
|---|---|---|---|---|---|---|---|
| HR Seizure | Normal | UFIR | MA | SPLINE | LOESS | RLOESS | LOWESS |
| s01 | 58 | 689 | 414 | 428 | 740 | 1049 | 796 |
| s02 | 1261 | 426 | 185 | 405 | 828 | 970 | 890 |
| s03 | 1578 | 866 | 508 | 580 | 1580 | 1965 | 1721 |
| s04 | 403 | 300 | 244 | 250 | 389 | 359 | 403 |
| s05 | 637 | 450 | 493 | 572 | 686 | 726 | 909 |
| s06 | 518 | 424 | 111 | 231 | 493 | 568 | 497 |
| s07 | 484 | 356 | 159 | 251 | 711 | 529 | 870 |

## 5 Conclusions

The $p$-shift UFIR smoothing filter operating on optimal horizon has demonstrated an ability to estimate the baseline in spite of statistical changes caused by the seizures. The residual distribution at the UFIR smoothing filter output is more normalized than by other techniques. Therefore, the discrimination of data has appeared to be most accurate. Because the $l = 2$ degree UFIR smoothing filter is invariant to sharp changes, the proposed solution allows erasing atypical values with a minimal loss of information. Finally, the time/frequency analysis has shown that the proposed method does not alter the ECG principal frequencies.

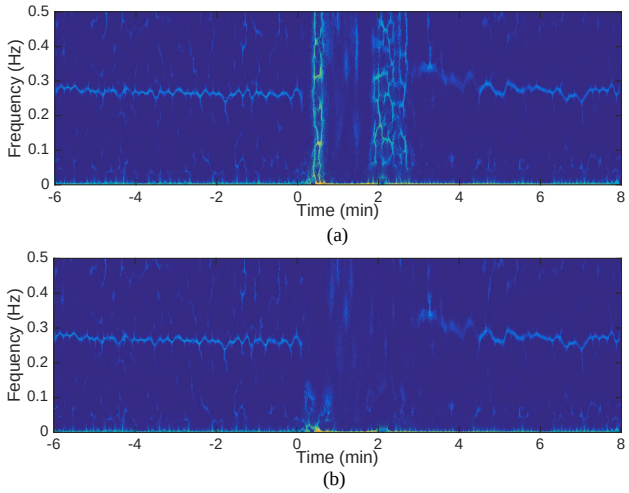



Figure 4: Time/frequency test of the raw HRV with an extracted average: (a) a pre and post epylepsy seizure and (b) TF analysis of HRV without atypical points, which are removed based on the smooth UFIR method in the same lapse.

*References:*

[1] American Heart Association Inc, Heart rate variability, Standards of measurement, physiological interpretation, and clinical use, *European Society of Cardiology*, 1996.

[2] F. Jurysta, P. van de Borne, P.-F. Migeotte, M. Dumont, J.-P. Lanquart, J.-P. Degaute, P. Linkowski, A study of the dynamic interactions between sleep EEG and heart rate variability in healthy young men, In Clinical Neurophysiology, *Clinical Neurophysiology*, Vol. 114, 2003, pp. 2146–2155.

[3] S. R. Quint, J. A. Messenheimer, M. B. Tennison, and H. T. Nagle, Assessing autonomic activity from the EKG related to seizure onset detection and localization, *Proc. 2nd Annual IEEE Symposium on Computer-Based Medical Systems*, 1989.

[4] S. J. Tavernor, S. W. Brown, R. M. Tavernor, and C. Gifford, Electrocardiograph QT lengthening associated with epileptiform EEG discharges a role in sudden unexplained death in epilepsy?, *Seizure*, Vol. 5, 1996, pp. 79–83.

[5] V. Kviesulaitis, A. Puodziukynas, D. Haroldas Pauza, V. Zabiela, T. Kazakevicius, Heart rate variability after radiofrequency ablation of epicardial ganglionated plexuses on the ovine left atrium, *Cardiovasc Disord*, Vol. 17, 2017 pp. 292.

[6] F. Shaffer, J.P. Ginsberg,An Overview of Heart Rate Variability Metrics and Norms, *Frontiers in*

*Public Health*, Vol. 5, 2017 pp. 258.

[7] J. Mateo, E.M. Snchez–Morla, J.L. Santos, A new method for removal of powerline interference in ECG and EEG recordings *Computers & Electrical Engineering*, Vol. 45, 2015, pp. 235–248.

[8] P. Feng, L. Li, H. Zhang, G. Qian, Outliers detection in ICA, *Int. Conf. on Communications, Circuits and Systems (ICCCAS)*, 2013.

[9] S. Weinsberg, Applied Linear Regression. *John Wiley and Sons*, 1985.

[10] Tukey, J. W, Exploratory Data Analysis. *Addison–Wesley Publishing Co*, First Edition, 1977.

[11] R., McGill, J. W. Tukey, and W. A. Larsen, Variations of Boxplots, *American Statistician*, Vol. 32, 1978, pp. 12–16

[12] Y. S. Shmaliy, Y. Neuvo, S. Khan, Review of unbiased FIR filters, smoothers, and predictors for polynomial signals, *Front. Signal Process.* Vol. 2, 2018, pp. 1–29.

[13] R. Ganguli, Noise and outliers removal from jet engine health signals using weighted FIR median hybrid filters, *Systems and Signal Process.* Vol. 16, 2002, pp. 967–978.

[14] R. C. Hardie, C. Boncelet, LUM filters: a class of rank-order-based filters for smoothing and sharpening, *IEEE Trans. Signal Process.* Vol. 41, 1993, pp. 1061–1076.

[15] F. Ramirez-Echeverria, A. Sarr, and Y. S. Shmaliy, Optimal memory for discrete-time FIR filters in state-space, *IEEE Trans. Signal Process.* Vol. 62, 2014, pp. 557–561.

[16] C. Lastre-Domnguez, Y. S. Shmaliy, O. Ibarra–Manzano, M. Vazquez–Olguin, L. J. Morales–Mendoza, Unbiased FIR denoising of ECG data for features extraction, *IEEE Int. Autumn Meeting on Power, Electronics and Computing (ROPEC)*, 2017.

[17] I. C. Al–Aweel, K. B. Krishnamurthy, J. M. Hausdorff, J.E. Mietus, J. R. Ives, A. S. Blum, D. L. Schomer, A.L. Goldberger, Post-ictal heart rate oscillations in partial epilepsy, *Neurology*, Vol. 53, 1999.

[18] M. B. Malarvili, M. Mesbah and B. Boashash. "Time-Frequency Analysis of Heart Rate Variability for Neonatal Seizure Detection". *EURASIP Journal on Advances in Signal Processing*, 2007.

[19] Y. S. Shmaliy, O. Ibarra-Manzano, L. Arceo-Miquel, J. Munoz-Diaz, A thinning algorithm for GPS–based unbiased FIR estimation of a clock TIE model, *Measurement* Vol. 41, 2008, pp. 538–550.

[20] K. K. L., Ho, G.B., Moody, C. K., Peng, et al, Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics, *Circulation*, 1997.

[21] Y. S. Shmaliy, L. Morales-Mendoza, FIR smoothing of discrete-time polynomial models in state space, *IEEE Trans. Signal Process.*, Vol. 58, 2010, pp, 2544–2555.

[22] Y. S. Shmaliy, An unbiased p-step predictive FIR filter for a class of noise free discrete-time models with independently observed states, *Signal Image Video Process.*, Vol. 3, 2009, pp. 127–135.