An Algorithm for Filtering Electrocardiograms to Improve Nonlinear Feature Extraction

Mohammad Reza BAHMANYAR
School of Engineering and Design
Brunel University
Uxbridge, Middlesex, UB8 3PH, UK

and

Wamadeva BALACHANDRAN
School of Engineering and Design
Brunel University
Uxbridge, Middlesex, UB8 3PH, UK

Email: mohammad.bahmanyar@brunel.ac.uk

ABSTRACT

This paper introduces an algorithm for removing high frequency noise components from electrocardiograms (ECGs) based on Savitzky-Golay finite duration impulse response (FIR) smoothing filter. The peaks of R waves and the points at which Q waves end and S waves start are detected for all beats. These points are used to separate the low amplitude parts of the ECG in each beat, which are most affected by high frequency noise. The Savitzky-Golay smoothing algorithm is then applied to these parts of the ECG and the resultant filtered signals are added back to their corresponding QRS parts. The effect of high frequency noise removal on nonlinear features such as largest Lyapunov exponent and minimum embedding dimension is also investigated. Performance of the filter has been compared with an equiripple low pass filter and wavelet de-noising.

Keywords: Electrocardiograms, Nonlinear Feature Extraction, Savitzky-Golay algorithm, Lyapunov exponent, Taken's delay coordinates method, embedding dimension.

1. INTRODUCTION

Electrocardiography is an important tool in diagnosis of heart disease because it is a non-invasive, cheap, quick and relatively accurate medical test. Electrocardiograms (ECG's) are voltage signals taken from the surface of body and reflect electrical activity of heart. These signals are often interpreted by cardiologists but there are some occasions (as in intensive care units and smart pacemakers) in which real time monitoring and automatic interpretation of the patient's ECG is vital. In these occasions, time is extremely critical and a decision often shall be made within few seconds (e.g. in using Automatic External Defibrillators, AEDs). On the other hand, beat by beat interpretation of long ambulatory ECG's is tedious and time consuming. That is why automatic interpretation of electrocardiograms (ECG's) has been the subject of intensive research in recent years.

Most of the existing automatic ECG classifiers use either time or frequency domain representation of the ECG to obtain features required in the intelligent classification algorithms. Although the performance of these machines varies depending on the specific feature sets and algorithms used, they all face one major problem, which is rooted in the wild variations in the morphologies of ECG waveform. Due to these variations, accurate feature extraction from time domain signal is difficult. Also because of inherent nonlinearities of ECG signals, frequency spectrums are wide and continuous and not very informative.

Techniques from nonlinear dynamical systems theory have been recently used to extract features from the reconstructed phase space of the ECG time series [1-3]. Almost all reconstruction methods are based on Taken's delay coordinates embedding theorem [4]. Features extracted by this method are either based on calculating invariant measures such as correlation dimension [5], Lyapunov exponent [6], etc. or in the form of determining the distribution of points (states) in the phase space [7, 8].

High frequency noise components can affect the extracted nonlinear features and must be removed. On the other hand, these features should be preserved by the adopted filtering method.

In this paper, we will introduce an algorithm for filtering ECG signals based on Savitzky-Golay FIR smoothing filter with minimal distortion of ECG waves. We will argue that this type of filtering is suitable in terms of preserving features extracted from the reconstructed state space. The effect of noise on nonlinear features is also investigated. The results are compared with filtering by an equiripple low pass FIR filter as well as wavelet de-noising.

2. METHODOLOGY

Visual inspection of sampled ECG signal reveals that the slowly varying parts of ECGs are most affected by high frequency noise. These parts lie between the points where the Q wave of one beat ends and the S wave of the subsequent beat starts. The application of any type of filtering to the whole ECG signal removes the high frequency noise but unfortunately distorts the ECG waves to some extent, particularly the QRS complex. The amount of this distortion depends on the type and order of the filter used and often a compromise must be made between the distortion and the smoothness of the ECG. The resulting distortion may not be crucial if a cardiologist inspects the ECG, but is important if automated beat classification algorithms are employed. Such classification algorithms are based on features extracted from the ECG, and any distortion or loss of information may result in inaccurate features and lead to misclassification. To minimize distortion to QRS complexes, the proposed algorithm detects the end points of the Q waves and the starting points of the S waves throughout the ECG signal and applies the SG smoothing filter to the slowly varying parts. These smoothed parts are then added back to the corresponding QRS complexes.

Detection of key points

The peaks of the R waves of all beats in the ECG are detected first. As a property of peak points, the difference between each peak point to its immediate preceding and following points are both positive. On the other hand, due to the higher slope and amplitude of the R wave, these differences are considerably greater compared with other peaks happening in the beats. These properties of R wave peaks are used as conditions in the peak detection procedure. This is done by searching for points for which the averages of their differences with *m* preceding and following points are greater than a critical value *k*. The reason for choosing *m* preceding and following points lies in the fact that occasionally sampled beats have two points of equal ordinate at their R peaks. This is more likely to happen in ECGs sampled at higher frequency, and is a source of error. The constants *k* and *m* should be chosen with respect to the sampling frequency. To present this procedure in mathematical terms, suppose that the ECG time series is denoted by the sequence $\{x_n\}$. The following conditions should be satisfied for R wave peaks:

$$2x_i - x_{i-m} - x_{i+m} \geq k \qquad (1)$$

$$x_i - x_{i+m} > 0 \qquad (2)$$

$$x_i - x_{i-m} > 0 \qquad (3)$$

We chose k as a fraction of the difference between the maximum and the mean of the normalized ECG time series:

$$k = \alpha(\max(x_n) - mean(x_n)), \quad 0 < \alpha < 1 \qquad (4)$$

At the second stage, the points at which Q waves end and S waves start are detected by calculating the changes in the slopes of the waveform in each R wave. These points are shown as P1 and P2 in figure 1. These points are used to separate the low amplitude parts of the ECG in each beat, which are most affected by high frequency noise. These parts of the ECG are extracted for filtering. After filtering, these parts are inserted back into their corresponding QRS complexes.

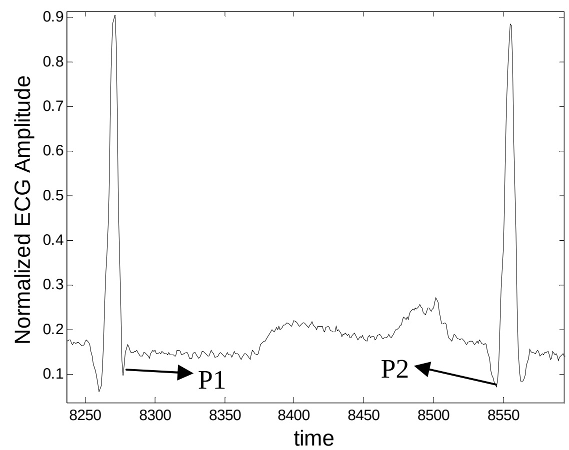

Figure 1: Key points in ECG beats

Smoothing Filter

Various types of filters may be used to smooth these parts. However, careful visual inspection of the results from applying different types of filters shows that Savitzky-Golay smoothing filters are good candidates, producing a good degree of smoothing and preserving useful high frequency components of these parts of ECGs. Another feature of these filters as a FIR filter is their ability to preserve nonlinear features from a reconstructed phase space. Later on, this point will be addressed and explained in more detail.

These filters which are also known as polynomial smoothing or least-square smoothing filters are generalizations of the FIR averaging filters that optimally fit a set of data points to polynomials of different degrees.

By proper selection of polynomial degree and span, one can achieve very good smoothness while preserving the high frequency content of the signal, as illustrated in Figure 4.1.2. Optimal values of degree and span may be obtained by trial and error.

In order to smooth parts of the ECG which lie between QRS complexes by a SG filtering method, these intervals are divided into subintervals containing an odd number of data points (span) *N*. The SG filter then fits the set of $N=2M+1$ data points $\mathbf{x}$ to a polynomial $\hat{x}_m$ of degree *d*:

$$\mathbf{x} = \left[x_{-M}, ..., x_0, ..., x_M \right]^T \qquad (5)$$

$$\hat{x}_m = \sum_{i=0}^{d} c_i m^i, \quad -M \leq m \leq M \qquad (6)$$

The points $\hat{x}_{-M}, \hat{x}_{-M+1}, \ldots, \hat{x}_M$ which are the projections of the points $x_{-M}, x_{-M+1}, \ldots, x_M$ on the polynomial may be found as follows:

$$\begin{pmatrix} \hat{x}_{-M} \\ \hat{x}_{-M+1} \\ \vdots \\ x_0 \\ \vdots \\ \hat{x}_{M-1} \\ \hat{x}_M \end{pmatrix} = \begin{pmatrix} 1 & -M & \cdots & (-M)^d \\ 1 & -M+1 & \cdots & (-M+1)^d \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & M-1 & \cdots & (M-1)^d \\ 1 & M & \cdots & M^d \end{pmatrix} \times \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_d \end{pmatrix} \quad (7)$$

Or, in short:

$$\hat{\mathbf{x}} = \mathbf{S}\mathbf{c} \quad (8)$$

here:

$$\mathbf{S} = \left(m^i\right)_{N\times(d+1)}; \quad -M \le m \le M; \quad 0 \le i \le d \quad (9)$$

The coefficients $c_0, c_1, \ldots, c_d$ of the polynomials are unknown and must be determined. To determine these, the equation $\mathbf{x} - \hat{\mathbf{x}} = 0$ needs to be solved to find the vector **c**. This leads to solving the following matrix equation for **c**:

$$\mathbf{x} = \mathbf{S}\mathbf{c} \quad (10)$$

The set of equations (4.10) consisting N equations with a (smaller) number of unknowns *(d+1)* is inconsistent but can be solved in least square sense. It can be shown [115] that the least-square solution to an inconsistent system $\mathbf{x} = \mathbf{S}\mathbf{c}$ satisfies the following equation:

$$\mathbf{S}^T\mathbf{x} = \mathbf{S}^T\mathbf{S}\mathbf{c} \quad (11)$$

The coefficient vector **c** and the fitted data vector $\hat{\mathbf{x}}$ may then be calculated as follows:

$$\mathbf{c} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{x} \quad (12)$$

$$\hat{\mathbf{x}} = \mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{x} \quad (13)$$

3. RESULTS

Based on the described methodology, an algorithm was developed in Matlab. The algorithm was tested on a number of ECG signals from the MIT-BIH arrhythmia database [10]. Visual inspection of filtered ECG showed successful removal of high frequency noise without distortion of ECG waves, particularly QRS complex. A portion of a sample ECG and the filtered signal are shown in Figure 2.

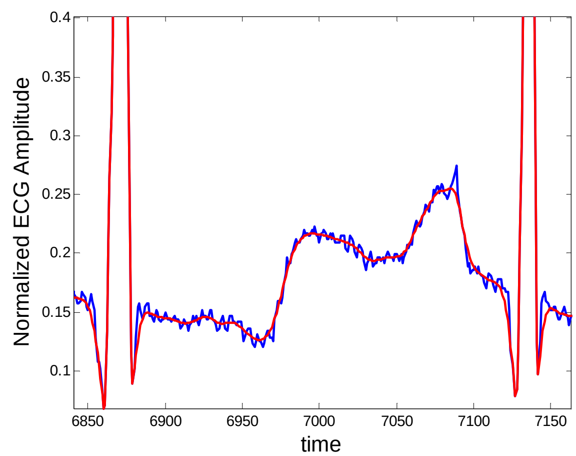

Figure 2: Original and filtered ECG.

A part of record No. 100 (from MIT-BIH database) consisting 21300 data points and its filtered version were used to calculate the largest Lyapunov exponent (LLE) and minimum embedding dimension using TSTOOL package [11]. The results are shown in figures 3-6 below.

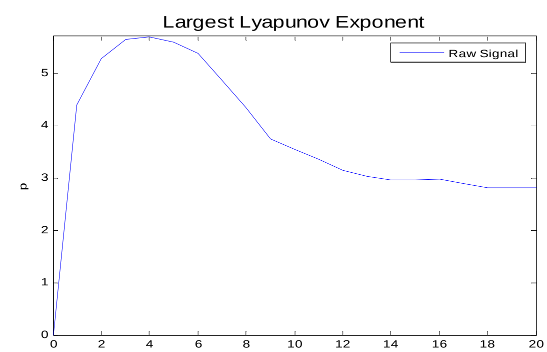

Figure 3: Largest Lyapunov Exponent of the raw signal.

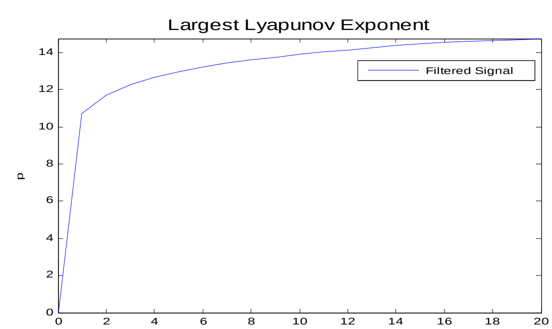

Figure 4: Largest Lyapunov Exponent after filtering.

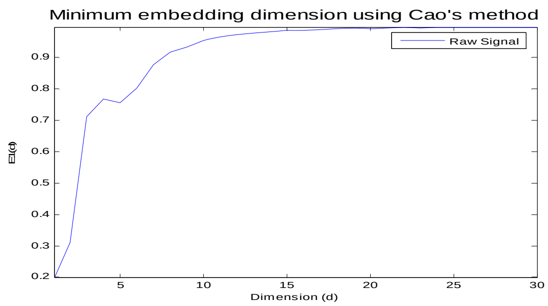

Figure 5: Minimum embedding dimension (raw signal).

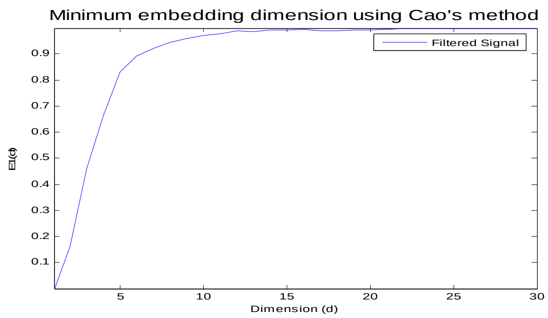

Figure 6: Minimum embedding dimension (filtered signal).

Figure 7 shows the performance of the described filter compared with an equiripple low pass FIR filter of order 49 with f_p=10 Hz, f_s=50 Hz as well as wavelet de-noising.

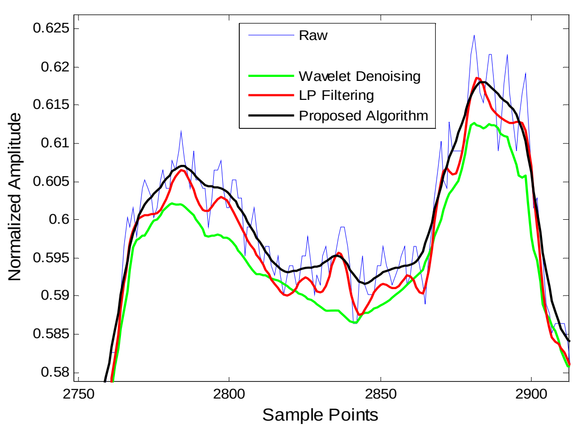

Figure 7: Comparison of different filtering methods: equi-ripple FIR low pass filter (red), wavelet de-noising (green) and proposed algorithm (black).

4. DISCUSSION

Results shown in figures 3-6 show that high frequency noise plays a part in the calculated nonlinear features and to ensure that the features are informative they must be removed. The filtering algorithm should also ensure preserving these features. It is shown that although FIR filters are linear in nature, they preserve all information that can be extracted from embedding techniques [12].

Visual inspection of the filtered signals obtained by different filters show that the performance of the proposed method is better in terms of distortion of ECG waves and smoothness of the signal.

5. REFERENCES

[1] Fojt, O., and Holcik, J., "Applying nonlinear dynamics to ECG signal processing", IEEE Enginerrring in Medicine and Biology, March/April 1998.

[2] M. I. Owis, A.H. Abou-Zeid, A. M. Yousef and Y. M. Kadah "Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification," *IEEE Trans.Biomed. Eng., vol 49, No. 7, pp. 733-736 July 2002.*

[3] Richter, M., and Schreiber, T., "Phase space embedding of electrocardiograms" Phys. Rev. E **58**, 6392, 1998.

[4] E Takens, "Dynamical Systems and Turbulence", eds. D. Rand and L.-S. Young (Springer-Verlag, Berlin, 1981).

[5] Casaleggio, A., Corana, A. and Ridella, R., "Correlation dimension estimation from electrocardiograms" *Chaos, Solitons & Fractals, Volume 5, Issues 3-4, March-April 1995, Pages 713-726.*

[6] Casaleggio, A. and Braiotta, S., "Estimation of Lyapunov exponents of ECG time series—The influence of parameters", *Chaos, Solitons & Fractals, Volume 8, Issue 10, October 1997, Pages 1591-1599.*

[7] Srinivasan, N., Wong, M. T., and Krishnan, S. M., "A new phase space analysis algorithm for cardiac arrhythmia detection" Proc. 25th IEEE EMBS, pp. 82-85, September 2003.

[8] Figueiredo, JCB, and Furuie, SS, "Statistical distance of quadtree partitioned attractors: Detecting changes in the behavior of ECG signals", Computers in Cardiology, 28, pp. 521-524, 2001.

[9] Orfandis, S. J., "Introduction to signal processing", Prentice Hall, 1996.

[10] R. Mark and G. Moody. (1997, May) MIT-BIH Arrhythmia Database. [Online]. Available: http://ecg.mit.edu/dbinfo.html.

[11] Merkwirth, C., Parlitz, U., and Lauterborn, W., "Tools for nonlinear data analysis", Int. Symp. Nonlinear Theory and its Applications, Switzerland, 1998.

[12] Broomhead, D. S., Huke, J. P., and Muldoon, M. R., "Linear filters and nonlinear systems", J. R. Stastist. Soc. B, 54, No. 2, pp. 373-382, 1992.