Analysis of Surface Atrial Signals Using Spectral Methods for Time Series with Missing Data

R Sassi[1], VDA Corino[2], LT Mainardi[2]

[1]Dipartimento di Tecnologie dell'Informazione, Università di Milano, Crema, Italy
[2]Dipartimento di Bioingegneria, Politecnico di Milano, Milano, Italy

Abstract

In this work, the analysis of atrial signals recorded during atrial fibrillation was pursued using two spectral estimators designed for series with missing data: the Lomb periodogram (LP) and the Iterative Singular Spectrum Analysis (ISSA). The main aim is to verify if subtraction of the ventricular activity might be avoided by performing spectral analysis on those ECG intervals where such activity is absent, (i.e. the T-Q intervals), at least to estimate the dominant atrial Fibrillatory Frequency (FF).

Recordings coming from the 2004 *Computers in Cardiology Termination Challenge Database were analyzed. Fibrillatory frequencies were then compared with those obtained from the analysis of the correspondent atrial signals extracted using a modified Average Beat Substraction (ABS) technique. We observed that the mean absolute difference was* 0.42 ± 0.66 *Hz for LP, (mean±SD), and* 0.39 ± 0.64 *Hz for ISSA. We concluded that estimation of FF is feasible without applying QRS-T subtraction.*

1. Introduction

Analysis of fibrillatory waves extracted from the surface ECG on subjects undergoing atrial fibrillation (AF) have been documented to provide significant information on the properties of AF events [1]. In particular, the dominant fibrillatory frequency (FF), or the related "dominant atrial cycle length" (DACL), are features of clinical relevance to assess drugs treatment and to predict the outcome of cardioversion [2] or ablation therapy [3].

Quantification of the dominant FF is obtained through spectral analysis of Atrial Signal (AS), which is usually extracted from surface ECG by removing waves induced by ventricular activities. The derivation of AS requires advanced signal processing techniques, since atrial and ventricular activities, during AF, overlap in time and frequency. These techniques include averaged beat subtraction (ABS) [4], spatio-temporal QRS-T cancellation [5] or methods based on independent (ICA) or principal (PCA) component analysis [6, 7].

Once obtained the atrial signal, the detection of the dominant FF is often the main (and only) goal. Thus, one might wonder if subtraction of the ventricular activity might be avoided by performing spectral analysis on those ECG intervals where ventricular activity is absent, (i.e. the T-Q intervals). The idea was originally explored by Rosenbaum & Cohen [8] who averaged periodograms computed on all the several T-Q intervals at disposal. The technique, while limited in the scopes, does not need several concurrent leads, as it might be the case in Holter recordings. Limitations are the low spectral resolution, and the problem of vanishing T-Q intervals at high heart rates.

In order to overcome these limitations, we observed that the successions of T-Q intervals, obtained by removing the ECG portions affected by QRS-T waves, can be treated as a time-discontinuous (or unevenly sampled) signal. Looked from this perspective, the question might be recast into a problem of missing data in a long time series and proper methods might be applied. In particular, two techniques were explored in this work: i) the Lomb periodogram (LP) [9] and ii) the Iterative Singular Spectrum Analysis (ISSA) [10]. The objective of the work is to explore the capability of these methods in detecting the dominant FF from single-lead analysis of surface ECG recorded during AF.

2. Methods

2.1. Iterative SSA method

Singular spectrum analysis (SSA) is a technique used for state space reconstruction of nonlinear dynamical systems. It was also used for detecting oscillations of physical significance in a time series [11, 12]. The original time series $x(t_n)$ ($n = 1, 2, \ldots, N$, being $\Delta t_n = t_{n+1} - t_n =$ constant) is initially embedded in a space of large dimension M, where the embedding dimension M determines the longest periodicity captured. The M-lag correlation matrix $\mathbf{C}_x$ is computed in the vector space of delay coor-

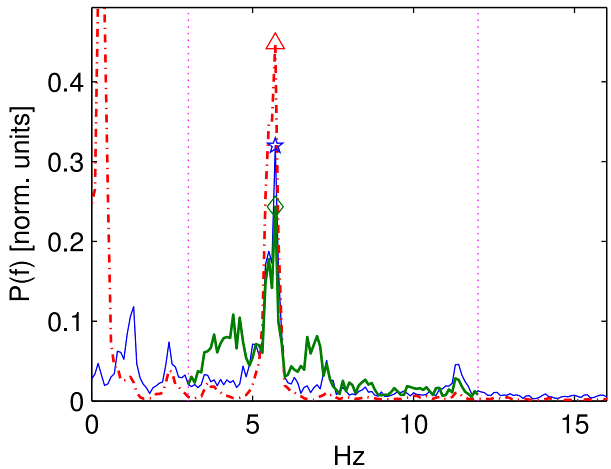

Figure 1. Spectra of the atrial fibrillation signal as obtained with the three methods employed in the work: iterative SSA (dash-dotted line), Lomb periodogram (thick line) and the modified ABS method as a reference (thin continuous line). Data refers to record `n02`, first channel.

dinates as

$$\mathbf{C}_x(i,j) = \frac{1}{N-|i-j|} \sum_{n=1}^{N-|i-j|} x(t_n)x(t_{n+|i-j|}) \quad (1)$$

with $0 \leq |i-j| < M$. $\mathbf{C}_x$ is a $M \times M$ Toeplitz matrix with constant diagonals. Then, principal component analysis (PCA) is carried on $\mathbf{C}_x$ computing its eigenvectors, $\mathbf{E}_l$, also known as empirical-orthogonal functions (EOF), and eigenvalues λ_l. Projecting the time-series onto each EOF produces $l \leq M$ principal components

$$A_l(t_n) = \sum_{j=1}^{M} x(t_{n+j})E_l(j) \quad (2)$$

where $0 \leq n \leq N-M$. Principal components are shorter $(N-M+1)$ and filtered version of the time series $x(t_n)$. The original time series can be expanded in an optimal way [12] as the sum of its M *reconstructed components*:

$$x(t_n) = \sum_{l=1}^{\Lambda} \mathcal{R}_l(t_n) \quad (3)$$

where

$$\mathcal{R}_l(t_n) = \sum_{j=1}^{M} \frac{A_l(t_{n-j})E_l(j)}{M} \quad (4)$$

is valid for $M \leq n \leq N-M+1$. (Reconstruction formula for the remaining points can be found in [12]). The eigenvectors corresponding to relatively small eigenvalues λ_l have little contribution on $x(t_n)$ and are likely to be generated by noise. Therefore, once sorted the eigenvalues in decreasing order, signal and noise are typically separated by inspecting the diagram of λ_l vs l and the reconstruction of $x(t_n)$ is performed with only m reconstructed components (and thus principal components).

Schoellhamer [13] suggested a modified version of the SSA algorithm to compute spectral estimates of time series with missing data. The main idea is that one can still compute (1) by ignoring any pairs of data points where a value is missing:

$$\widetilde{C}_x(i,j) = \frac{1}{K_l} \sum_{K_l} \widetilde{x}(t_k)\widetilde{x}(t_{k+|i-j|}) \quad (5)$$

where the summation is now extended to the K_l available pairs only, being $K_l \leq N-|i-j|$ and M determines the width of the gaps to be filled. In (5) $t_k \subset t_n$ and $\Delta t_k = t_{k+1} - t_k \neq$ constant. A few leading reconstructed components are then obtained with a singular value decomposition of the matrix $\widetilde{\mathbf{C}}_x$ and used to build a filtered version of the time series $\widetilde{x}(t)$. Finally the gaps of $x(t_n)$ are filled with the corresponding samples of $\widetilde{x}(t)$.

The method can be applied in an iterative way [10]. Firstly a subset of the available data (called *test set*) is removed from the series. The leading empirical orthogonal function E_1 is selected and the corresponding reconstructed component $\mathcal{R}_1$ is used to fill the gaps in the original time series. The empirical orthogonal function E_1 is recomputed and the procedure is iterated (inner loop) until a convergence criterium is met. At this point a second EOF is added for reconstruction, starting from the series with data filled in by $\mathcal{R}_1$, and the inner loop is again repeated until convergence. The process is stopped when adding new EOF does not improve significantly the reconstruction of the *test set*. At the end, the *test set* data are plugged back in and the inner loop repeated again until convergence. Summarizing, the inner loop is used to refine the reconstruction with a fixed number of reconstructed components. The outer loop instead is meant to increase the number of such components, if necessary.

2.2. Lomb periodogram

A classical method of spectral analysis for unevenly sampled data is the Lomb-Scargle periodogram (LP) [9]. A *slightly modified* version of the periodogram was given in [14] by:

$$P(\omega) = \frac{1}{2\sigma^2} \left\{ \frac{\left[\sum_{k=1}^{K} \left[\widetilde{x}(t_k) - \bar{x}\right] \cos\omega(t_k - \tau)\right]^2}{\sum_{k=1}^{K} \cos^2\omega(t_k - \tau)} + \frac{\left[\sum_{k=1}^{K} \left[\widetilde{x}(t_k) - \bar{x}\right] \sin\omega(t_k - \tau)\right]^2}{\sum_{k=1}^{K} \sin^2\omega(t_k - \tau)} \right\} \quad (6)$$

where $\omega = 2\pi f$, $\bar{x}$ and σ^2 are respectively the mean and variance of the series $\widetilde{x}(t_k)$. τ, defined as

$$\tan(2\omega\tau) = \frac{\sum_{k=1}^{K} \sin(2\omega t_k)}{\sum_{k=1}^{K} \cos(2\omega t_k)}, \tag{7}$$

is an offset which makes $P(f)$ independent of time-shifts. The Lomb periodogram analyzes periodic signals weighting data on available points, rather then on fixed time intervals (such as the DFT method does).

Direct implementation of equation (6) leads to computationally intensive algorithms, which might become unfeasible when a large number of frequencies need to be explored. While fast approximate algorithms are available [15], in the following we rather preferred to simply evaluate equation (6) on a small number of evenly spaced frequencies ($\Delta f = 0.1$ Hz) in the range $[3, 12]$ Hz, where the dominant FF is supposed to be located.

2.3. AF database

To test the methods on real ECG recordings, we considered the data coming from the 2004 Computers in Cardiology AF Termination Challenge Database [16]. The Database includes 80 records extracted from long-term ECG Holter of patients undergoing AF episodes. Each recording is 1-minute long, sampled at 128 Hz and two simultaneous ECG leads are available.

3. Results

3.1. Reference FF computation

To assess the actual performances of the methods under analysis on real ECG data, a reference value for the fibrillation frequencies was needed. To this extent, a modified ABS method was employed. Details on the algorithm can be found in a previous paper [17]: briefly, it uses two separate templates for QRS and T waves to compensate for QT variability. The spectral power density was computed with the Welch's periodogram (windows length of 1280 points with 1120 samples overlap; spectral resolution 0.1 Hz) and the dominant oscillation was located by direct search in the range $[3\ 12]$ Hz. For each record, FF was estimated independently for each channel. To ensure that the dominant oscillation was correctly estimated (*e.g.* avoid mild dominant peaks in a relatively broad-band spectrum), those records for which the spectral concentration of the peak was lower than 0.12 were excluded (15 traces). Also, cases in which the heart rate was higher than 140 bpm were set apart as these would have been worst case scenario situations for the LP and ISSA methods (7 subjects).

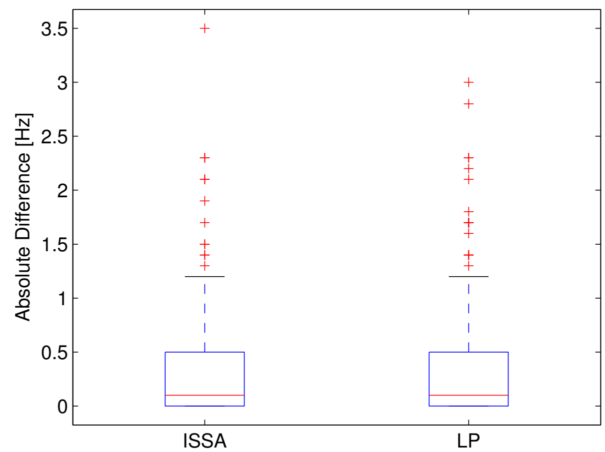

Figure 2. Box plot of the absolute differences between FF estimated obtained with ISSA or LP and the reference ABS method (See section 3.1 for details). Lines at the lower quartile, median, and upper quartile values were plotted. Possible outliers are marked with crosses.

3.2. Lomb and ISSA spectral estimate

ECG segments containing QRS-T waves were eliminated with a simple rule. We fixed $QT_c = 550$ ms and estimated each QT interval with the inverse Bazett formula: $QT = QT_c RR^2$. Then, starting from the Q onset, all the points contained in the QT segment were removed from the series. Ventricular beats and artifact were excluded completely (*i.e.* all the points up to the next beat were eliminated).

Spectral estimates were obtained for the 131 recordings with the Lomb and ISSA method. The variance of the Lomb estimates was decreased with an approach similar to the one used for the Welch's periodogram. $\widetilde{x}(t_k)$ was split into portions of 1280 points each with a 1120 samples overlap. Figure 1 compares the fibrillatory spectra obtained with the three methods.

3.3. Performances

Figure 2 summarizes the results obtained. The absolute differences between the estimated FFs and the reference values computed with the ABS method are shown. In general, the dominant oscillation was detected by all the methods and there was an overall good agreement: the discrepancy with the reference ABS method was ≤ 0.5 Hz for most cases ($101/131$ recordings for ISSA and $99/131$ for LP). Only in a few situations the observed absolute difference was higher than 1 Hz ($17/131$ cases for ISSA and $22/131$ cases for LP). The agreement between ISSA and ABS method is slightly superior than the one between LP and ABS. Overall, the mean absolute difference was

0.39 ± 0.64 Hz, (mean$\pm$SD), for ISSA and 0.42 ± 0.66 Hz for LP.

4. Discussion and conclusions

In this work we tested two methods devised for computing spectral estimates from series with missing data. The analysis was performed with ECG signals recorded during atrial fibrillation. The main focus was on the quantification of the dominant fibrillation frequency, a key characteristic of the AS, widely used for clinical evaluation of AF events.

The results showed that both algorithms were able to provide reliable estimates of the FF with a level of agreement consistent with similar comparative works, where techniques for separating the atrial signal from ventricular waves were considered (*e.g.* see figure 3 in [18]).

The ability to manage missing data proved a practical advantage with AS: the development of accurate algorithms for the cancelation of the ventricular waveforms might be avoided as long as ECG segments including those activities are discharged.

Surely more work is needed though (for example to include the cases with heart rate > 140 bpm). But several aspects render these results promising: i) the algorithms are simpler and their implementations straightforward; ii) estimation of the dominant FF can be performed on a single-lead basis (which is a plus in long-term monitoring when only a few leads are usually recorded); iii) the ECG signals we considered were sampled at a low frequency and were not free of ectopic ventricular beats as in related comparative works [18].

References

[1] Bollmann A, Husser D, Mainardi L, Lombardi F, Langley P, Murray A, Rieta J, Millet J, Olsson B, Stridh M, Sörnmo L. Analysis of surface electrocardiograms in atrial fibrillation: techniques, research, and clinical applications. Europace 2006;8:911–926.

[2] Tai C, Chen S, Liu A, Yu W, Ding Y, Chang M, Kao T. Spectral analysis of chronic atrial fibrillation and its relation to minimal defibrillation energy. Pacing Clin Electrophysiol 2002;25:1747–1751.

[3] Haissaguerre M, Sanders P, Hocini M, Hsu L, Shah D, Scavee C, Takahashi Y, Rotter M, Pasquie J, Garrigue S, Clementy J, Jais P. Changes in atrial fibrillation cycle length and inducibility during catheter ablation and their relation to outcome. Circulation 2004;109:3007–3013.

[4] Slocum J, Sahakian A, Swiryn S. Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity. J Electrocardiol 1992;25:1–8.

[5] Stridh M, Sornmo L. Spatiotemporal QRST cancellation techniques for analysis of atrial fibrillation. IEEE Transactions on Biomedical Engineering 2001;48(1):105–111.

[6] Raine D, Langley P, Murray A, Dunuwille A, Bourke JP. Surface atrial frequency analysis in patients with atrial fibrillation: a tool for evaluating the effects of intervention. J Cardiovasc Electrophysiol 2004;15:1021–1026.

[7] Rieta JJ, Castells F, Sanchez C, Zarzoso V, Millet J. Atrial activity extraction for atrial fibrillation analysis using blind source separation. IEEE Transactions on Biomedical Engineering 2004;51(7):1176–1186.

[8] Rosenbaum DS, Cohen RJ. Frequency based measures of atrial fibrillation in man. In Engineering in Medicine and Biology Society, 1990. Proceedings of the Twelfth Annual International Conference of the IEEE. 1990; 582–583.

[9] Lomb NR. Least-squares frequency analysis of unequally spaced data. Astrophysics and Space Science 1976;39:447–462.

[10] Kondrashov D, Ghil M. Spatio-temporal filling of missing points in geophysical data sets. Nonlin Processes Geophys 2006;13:151–159.

[11] Vautard R, Ghil M. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Physica D 1989;35:395–424.

[12] Vautard R, Yiou P, Ghil M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. Physica D 1992; 58:95–126.

[13] Schoellhamer DH. Singular spectrum analysis for time series with missing data. Geophys Res Lett 2001; 28(16):3187–3190.

[14] Scargle JD. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. Astrophysical Journal 1982;263:835–853.

[15] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in C: the Art of Scientific Computing. Second edition. Cambridge Univ Press, 1992.

[16] Moody GE. Spontaneous termination of atrial fibrillation: a challenge from physionet and computers in cardiology 2004. In Computers in Cardiology, 2004, volume 31. 2004; 101–104.

[17] Mainardi LT, Matteucci M, Sassi R. On predicting the spontaneous termination of atrial fibrillation episodes using linear and nonlinear parameters of ECG signal and RR series. In Computers in Cardiology, 2004, volume 31. 2004; 665–668.

[18] Langley P, Rieta JJ, Stridh M, Millet J, Sornmo L, Murray A. Comparison of atrial signal extraction algorithms in 12-lead ECGs with atrial fibrillation. IEEE Transactions on Biomedical Engineering 2006;53(2):343–346.

Address for correspondence:

Roberto Sassi
Dipartimento di Tecnologie dell'Informazione
Università degli Studi di Milano
via Bramante 65, 26013 Crema (CR), Italy
E-mail address: sassi@dti.unimi.it