ON DIRECT SEQUENTIAL ANALYSIS OF HRV SIGNALS

Dragana Bajic, Olijandra Šibarevic, Cedomir Stefanovic, *Faculty of Technical Sciences, Novi Sad*
Nina Žigon, *Medical faculty, Belgrade*

Abstract — *HRV analysis represents one of the most promising and the most commonly used quantitative measures of the cardiovascular autonomic regulatory system. The analysis includes traditional statistical analytical tools and a number of new methods based on nonlinear system theory, recently developed to give better insight into complex HR signals. This paper compares a new direct sequential analysis to the existing sequential analysis, in trials to penetrate the characteristics of children's HRV.*

1. INTRODUCTION

Heart rate variability (HRV) has become the conventionally accepted term to describe variations of interval between consecutive heartbeats, as well as the oscillations between consecutive instantaneous heart rates. It represents one of the most promising and the most commonly used quantitative measures of the cardiovascular autonomic regulatory system. The analysis include traditional statistical analytical tools both in time and spectral domain and a number of new methods based on nonlinear system theory, recently developed to give better insight into complex HR dynamics, such as fractal correlation properties, the slope of the power law relation and approximate entropy (ApEn). These methods might reveal malign abnormalities at early stage that may not be uncovered by traditional measures. However, the significance and meaning of these different measures of HRV are more complex than generally appreciated, and there is a potential for incorrect conclusions and for excessive or unfounded extrapolations. Besides, in spite of general opinion that HRV time series is easy to obtain, visual inspection and careful manual editing after automatic extraction is absolutely necessary. For 24 hours' holter signal it presents a cumbersome task. [1-4]

This paper discusses the possibility of application of various analyses based upon the sequence matching including the original, recently developed, one to children's HRV signals. The subsequent section gives a brief review of some of the existing methods based upon a template (sequence) matching. The third section gives a theoretical approach to the new direct sequence evaluations. The method is illustrated using analog ECG holter signals of clinically healthy children recorded at Children's hospital (Tirsova) are digitalized and HRV extracted at Faculty of Technical Sciences, Novi Sad.

2. SEQUENCES AND TEMPLATES

The HRV signal samples form a vector of length L, denoted by $\mathbf{y} = [y(j)], j=1,\ldots,L_S$. An analytical approach based upon the sequence (template) matching try to unveil if the similar set of samples is followed by other similar set of samples. Therefore, a "sequence" (or a "template" of length N is defined as a short vector $\mathbf{x}_N(i) = [y(i+k-1)]$, $k = 1,\ldots,N$, $i=1,\ldots,L_S-N+1$ that is a part of a long time series. For each pair of sequences a distance $d(\mathbf{x}_N(i), \mathbf{x}_N(j))$, $i,j=1,\ldots,L_S-N+1$ is defined. It can be maximal absolute distance, mean square distance or any other distance suitable for the current investigation.

2.1. Approximate entropy and its modifications

ApEn can be defined as a "regularity statistic" that quantifies the unpredictability of fluctuations in a time series. Intuitively, one may reason that the presence of repetitive patterns of fluctuation in a time series renders it more predictable than a time series in which such patterns are absent. A time series containing many repetitive p atterns has a relatively small ApEn; a less predictable (i.e., more complex) process has a higher ApEn. Therefore, ApEn estimates likelihood that patterns of certain length that are close one to another would remain close if the pattern length increases. The procedure for its evaluation from a time series $\mathbf{y}$ is simple: number of N-tuples (and $N+1$-tuples) for which $d(\mathbf{x}_N(i),\mathbf{x}_N(j))$, $i,j=1,\ldots,L_S-N+1$ is within a specified distance r are counted an processed:

$$C_i^N(r) = \frac{1}{L_S - N + 1} \cdot \sum_{j=1}^{L-N+1} z_j; \qquad z_j = \begin{cases} 0, & d(\mathbf{x}_N(i), \mathbf{x}_N(j)) \geq r \\ 1, & d(\mathbf{x}_N(i), \mathbf{x}_N(j)) < r \end{cases} \tag{1}$$

where $C_i^N(r)$ estimates the probability that any sequence $\mathbf{x}_N(j)$ is within a distance r from the template $\mathbf{x}_N(i)$. Then the approximate entropy can be estimated as:

$$\mathrm{ApEn}(N,r,L) = \frac{1}{L_S - N + 1} \cdot \sum_{i=1}^{L-N+1} \ln\left[C_i^N(r)\right] - \frac{1}{L_S - N} \cdot \sum_{i=1}^{L-N} \ln\left[C_i^{N+1}(r)\right] \tag{2}$$

Since this approach has its unbiased [4] and uncorrelated [5] modifications that yield different values (as could be seen in Fig. 1), the results cannot be regarded as reliable, so they are excluded from this research.

2.2. Correlation dimension approach – D_C

Correlation dimension estimates fractal dimension of an attractor from a time series. An attractor dimension itself shows a statistical measure of the self-similarity of the geometry of the sets of points in the phase space, i.e. the number of degrees of freedom necessary to describe a process. In our case, each sequence (N-tuple) is a point in an N-dimensional phase space. D_C is estimated from a time series using the correlation integral $C_N(r)$ that measures the number of points correlated with each other in a sphere of radius r around the point $\mathbf{x}_N(i)$ [6]:

$$C_N(r) = \frac{1}{(L_S - N + 1)\cdot(L_S - N)} \cdot \sum_{i=1}^{L_S-N} \sum_{j=i+1}^{L_S-N+1} z_{ij}, \qquad z_{ij} = \begin{cases} 0, & d(\mathbf{x}_N(i), \mathbf{x}_N(j)) \geq r \\ 1, & d(\mathbf{x}_N(i), \mathbf{x}_N(j)) < r \end{cases} \tag{3}$$

This quantity is similar to probability estimate (Eq (1)). The differences are a) triangular rather than rectangular

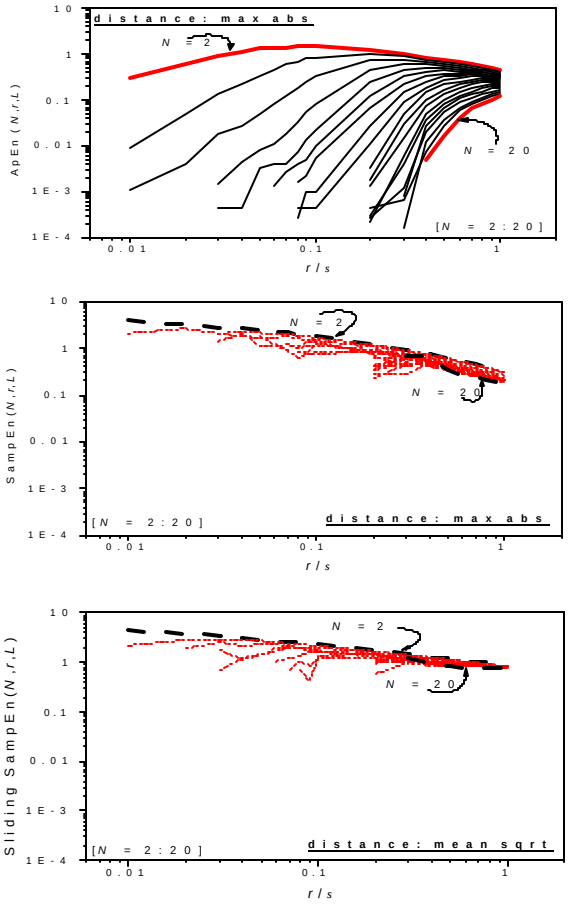

Fig. 1. *Entropy for child "L"*: Approximate (upper), sample (middle), sliding (lower part)

summation; b) squared instead of max. absolute distance; and c) for an ApEn approach, standard deviation of each separate signal is used as a scaling factor for *r*.

The correlation dimension is evaluated as:

$$D_C = \lim_{r \to 0} \frac{\mathrm{ld}(C_N(r))}{\mathrm{ld}(r)}. \tag{4}$$

Its value can be obtained by plotting ld($C_N(r)$) vs. ld(r). The slope of the resulting straight lines, for different N, tends to constant value D_C, as explained in [6].

The D_C analysis is closely related to multifractal property of HRV signal. Contrary to the monofractal signals that are homogeneous, multifractal signals can be decomposed into many subsets. The statistical properties of the different subsets are characterized by local Hurst exponents h that shows the local singular behavior and can be quantified by the function $D(h)$ - fractal dimension of the subset of the time series [7].

2.3. Illustrative examples and a note on stationarity

As previously mentioned, HRV time-series are obtained from the children's ECG, known to be extremely non-stationary and mutually different (although healthy). From each 24h signals two 15 min HRV sub-series were chosen, the ones that seemed (by mere visual inspection of medical doctor's eye) to be stationary. To verify the assumption, a stationarity test is performed [8].

Each HRV series is divided into K intervals. For each interval a mean value m_i and variance σ_i^2:are estimated, i = 1,…,K. Then the following sums are evaluated, both for m_i and σ_i^2:

$$A = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} a_{ij}, \quad a_{ij} = \begin{cases} 1, & m_i > m_j \quad (\text{or } s_i^2 > s_i^2) \\ 0, & m_i \le m_j \quad (\text{or } s_i^2 \le s_i^2) \end{cases} \tag{5}$$

The acceptance region for the stationarity hypothesis at a α level of significance was considered done by:

$$[A_{K;1-\alpha/2} < A < A_{K;\alpha/2}].$$

The corresponding border values $A_{K;1-\alpha/2}$ and $A_{K;\alpha/2}$ are obtained from discrete distribution function:

$$\Pr\{A = i\}, \quad i = 0, \cdots, K \cdot (K-1)/2. \tag{6}$$

This probability was not available in [8], but it can be obtained knowing that probability that exactly b elements exceeds the value of $(K\text{-}n)^{\text{th}}$ element, given that m elements in total exceeds the value of $(K\text{-}n)^{\text{th}}$ one:

$$\Pr\{b / K - n, m\} = \frac{\binom{b}{n} \cdot \binom{K-n-1}{m-b}}{\binom{K-1}{m}}, \tag{7}$$

$$b = 0, \cdots, n; n = 1, \cdots, K-1, m = b, \cdots, K-n-1+b$$

$$\Pr\{A\} = \sum_{m,b1+b2+\ldots+bK-1=A} \Pr\{b_1 / K-1, m\} + \Pr\{b_2 / K-2, m\} + \ldots + \Pr\{b_{K-1} / 1, m\} \tag{8}$$

***K*=10**		
	m	s^2
A1	5	29
A2	18	20
C1	7	27
C2	9	39
D1	25	38
D2	24	21
L1	23	12
L2	30	15
S1	39	10
S2	25	20

Table 1: *Values of A for m and s^2; for α=0.05 and K=10 values should be 11<A<33*

Only 5 of time series passed both tests with α=0.05 level of significance (underlined values in Table 1). .

Fig. 2 shows the correlation dimension – dashed lines for non-stationary series. It is interesting to note that it was not

possible to extract D_C – neither automatically, nor manually, for one subject. Fractal dimension, again with dashed lines representing the non-stationary data, is shown in Fig. 3. Thick lines show the examples that fit the best and the worst the stationarity hypothesis.

It is known that children's HR is extremely variable. From the above figures no conclusions can be done. The fact that the children are healthy is according to the children's cardiologist opinion.

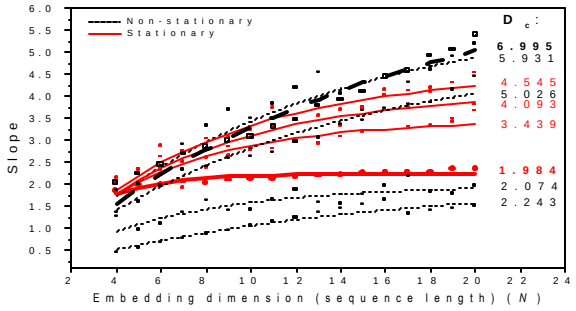

Fig. 2. *Correlation dimension D_C*

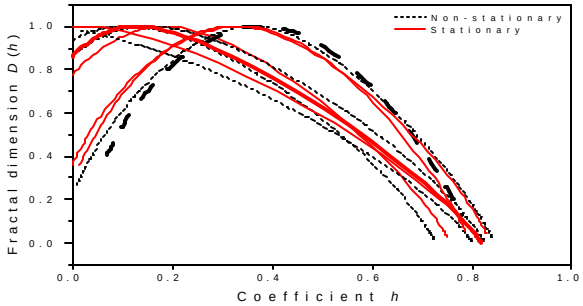

Fig. 3. *Fractal dimension D(h)*

3. DIRECT SEQUENTIAL ANALYSIS

Direct sequential analysis deals with the analysis of time parameters of the observed signals – the expected value of time units (number of samples) between the predefined set of M sequences (templates), and expected number of time units from the random starting position until one of the predefined M sequences is found. The first time distance is of the first type, the other one of the second type. The set of M sequences (templates) under consideration are the ones that match a certain criterion. For introductory explanation of this new analytical approach, the criterion for forming the sequence set would be the criterion of "smoothness" - the expected number of "rough" samples between the "smooth" intervals.

3.1. Time parameters and empirical series

Suppose that the HRV signal is shown in Fig. 4 and the smooth intervals are the ones for which the absolute value of sample do not exceed certain level. "Smooth" sequence of length N is the one that does not contain

more than L excursions beyond this specified level. If L is the number of allowed excursions, there would be exactly M different types of predefined sequences:

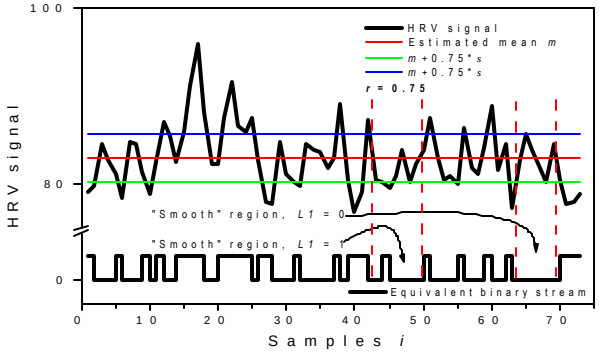

Fig. 4. *HRV signal, "smooth regions" and binary signal*

$$M = \sum_{n=0}^{L} \binom{N}{n}. \tag{9}$$

Similarly to (1) and (3), it is defined

$$z_{ij} = \begin{cases} 0, \mathrm{d}(x_N(i+j-1), m) \le r \\ 1, \mathrm{d}(x_N(i+j-1), m) > r \end{cases}, \quad j = 1, \cdots, N \tag{10}$$

and

$$z_i = \begin{cases} 1, & \sum_{j=1}^{N} z_{ij} \le L \\ 0, & \sum_{j=1}^{N} z_{ij} > L \end{cases}, \quad i = 1, \cdots, L_S - N + 1 \tag{11}$$

Then a sample time span between "smooth" intervals can be obtained as:

$$B_{1i}^{N}(r) = \sum_{\substack{z_i, \\ z_k \neq 1, i<k<j}}^{z_j} 1 \ , \tag{12}$$

for the time distance of the first type, and for the time distance of the second type, the same sample time would be:

$$B_{2k}^{N}(r) = \sum_{\substack{z_i = 1, \\ z_k \neq 1, i<k<j}}^{z_j} 1 \ , \tag{13}$$

Averaging the sample times does the estimate of the corresponding mean times. If $i_{\max}$ is the position of the last "smooth" interval within the time series, and

Averaging the sample times does the estimate of the corresponding mean times. If $i_{\max}$ is the position of the last "smooth" interval within the time series, and

$$z_{\max} = \sum_{i=1}^{L_S - N + 1} z_i \tag{14}$$

numbers of "smooth" intervals, then the corresponding estimates of mean times of type 1 and 2 are:

$$\overline{t_1(r)} = \frac{1}{i_{\max} - N + 1} \cdot \sum_{i=1}^{i_{\max} - N + 1} B_{1i}^{N}(r) \tag{15}$$

$$\overline{t_2(r)} = \frac{1}{z_{\max} - 1} \cdot \sum_{k=1}^{z_{\max}-1} B_{2k}^{N}(r) \qquad (16)$$

The main advantage of the direct sequential analysis is that the tests against surrogate data, disputed by many, and is avoided: there exists a theory against which it could be compared.

4. EXAMPLES AND DISCUSSION

For an illustrative example, some of the numerous results are shown in Fig. 4. Distance r is, through this investigation, normalized by standard deviation (the same as suggested for ApEn approach). Distance values for measured data are constrained by series length L_S: if its value is, e.g., 3000, then mean value cannot be greater. Upper part of figure 4 shows the absolute value of type 1 time; middle part – a relationship between the measured and theoretical data of type 1 evaluated as

$$\frac{E\{t_1\} - \overline{t_1}}{E\{t_1\}} \cdot 100\,. \qquad (26)$$

Most of the subjects had a positive relative error (26). an i.e. empirical value of mean time necessary to reach the smooth region is shorter for empirical sequences. The lower part compares theoretical and measured values of type 2 time distance, plotting the values

$$\frac{E\{t_2(N+1)\}}{E\{t_2(N)\}}, \quad \frac{\overline{t_2(N+1)}}{\overline{t_2(N)}} \qquad (27)$$

The measured data are in perfect accordance with the theoretical ones (this was not the case with time of type 1). The only exception was subject D1, who possess some other interesting properties: it was not possible to extract the correlation dimension, neither automatically nor manually. Besides, its value for distance type 1 has an interesting shape.

The results are promising, although no firm conclusion can be made: the children's data are variable and there was neither diagnosis, nor a medicine applied, influence of which could be observed within our data. Therefore, future investigation would be based upon laboratory rats with applied treatment (already in progress).

AKNOWLEDGMENT

The authors are grateful to Prof. Goran Vukomanovic who has kindly provided the children's signals, and to Tanja Loncar and Vladimir Minic that have extracted HRV signals and then performed a cumbersome task to correct the errors inherent to commercial software. The research was performed with the aid of the research grant 1774.

REFERENCES

[1] S.M. Pincus: "Approximate entropy as a measure of system complexity", *Proc Natl Acad Sci USA*;Vol. **88**:pp 2297-2301, 1991

[2] A.L. Goldberger, S.M. Pincus: "Physiological time-series analysis: What does regularity quantify?" *Am J Physiol*, Vol. **266** (Heart Circ Physiol), pp H1643-H1656, 1994.

[3] J.S. Richman, J.R. Moorman: "Physiological time-series analysis using approximate entropy and sample entropy" *Am J Physiol Heart Circ Physiol* Vol. **278**(6), ppH2039-H2049, 2000.

[4] T.Loncar, D. Bajic, N. Zigon: "On Information-Theoretic Approach To Hrv Time-Series Analysis", 48. Konferencija ETRAN-a, Cacak, juni 2004.

[5] R. Karvajal et al.: "Dimensional Analysis of HRV in Hypertrophic Cardiomyopathy Patientes", *IEEE Engineering in Medicine and Biology*, **21**(4), pp 71-78, July 2002.

[6] Plamen Ch. Ivanov et al.: Multifractality in human heartbeat dynamics", Nature, Vol. 399, pp 461-465, June 1999.

[7] Bendat and Piersol: *Random data analysis and measurement procedures*, New York – Wiley Interscience, 1986.

[8] D. Bajic, C. Stefanovic, D. Vukobratovic: "On probabilistic approach to bifix indicators", *Proceedings of 2003 IEEE International Symposium on Information Theory* **ISIT-2005**, Australia, July 2005,

[9] D. Bajic: "New simple method for solving first passage time problem", Electronics Letters, 1991, Vol. 27. No. 16, pp 1419-1421.

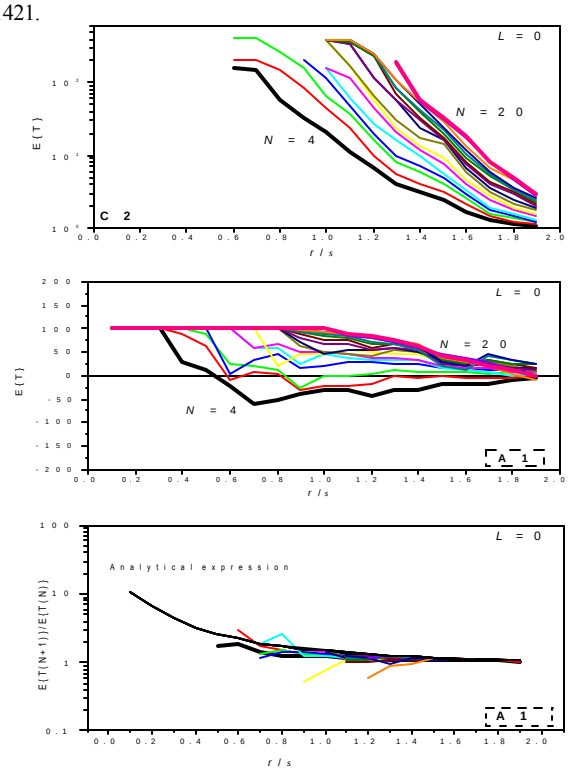

Fig. 4. *Results of the sequencial analysis, type 1 and 2*

Sažetak — *Analiza promenljivosti srcanog ritma predstavlja jednu od najcešcih kvantitativnih metoda za merenje autonomne regulacije kardiovaskularnog sistema. Analize ukljucuju tradicionalne statisticke postupke ali i brojne nove metode zasnovane na nelinearnim teorijama. U ovom radu se porede rezultati dobijeni poznatim metodama sekvencijalnog uopredivanja i nove metode koja ne zahteva poredenje sa veštacki generisanim signalima.*

O DIREKTNOJ SEKVENCIJALNOJ ANALIZI SIGNALA PROMENLJIVOSTI SRCANOG RITMA

Dragana Bajic, Nina Žigon, Olijandra Šibarevic, Cedomir Stefanovic