Detection of Biological Signals with Chaotic Properties through Assessment of Conventional Entropy

Liudmila A. Manilo
Saint Petersburg Electrotechnical University "LETI"
St. Petersburg, Russia
lmanilo@yandex.ru

***Abstract*— Theoretical problems and results of experiments on detection of biological signals with the aid of methods based on assessment of conventional entropy were considered. The corrected entropy parameters were shown to effectively serve description of regular as well as chaotic components of the processes under analysis as well as solution of medical diagnosis tasks.**

Keywords*— analysis of biomedical signals; chaotic components; conventional entropy; recognition of arrhythmias

I. Introduction

Study of specifics of the biological-medical signals generated by living systems is of interest because it develops new approaches to analysis of complex systems differing from each other in their nonlinear behavioural dynamics and prompts solution of a number of applied problems of diagnosis of living organism. Presence of mathematical models describing behaviour of a complex system makes it possible to effectively construct trajectories of its movement within the phase space for any initial conditions and even to assess in time the probability of approximation of such a system to a balanced condition [1]. In real tasks, however, particularly when analyzing living systems, it is often difficult to construe the mathematical model taking into consideration all specific features of the organism condition under study. Therefore one has to assess by experimental data both regular and fortuitous components of the behaviour trajectory. At the same time, the task of detecting regular and chaotic components describing in complex the process under study is particularly urgent in analysis of biological signals. Such an analysis can be directed to detection of chaos in the much diversified electrophysiological data that result from interaction among many variables every one of which has its own statistical law of distribution. These can be the tasks of studying the electrical activity of the brain in different stages of sleep and in wakefulness, revealing of EEG pathological changes, classification of heart rhythm disturbances by the ECG [2, 3].

Modern theory of chaos provides a number of parameters (dimension of attractor, Liapunov's parameter, Kolmogorov's entropy) that describe properties of the system and in theory enable one to reconstruct the equations generating the process under analysis. In practice, however, their application to tasks of biological signal analysis is limited by difficulty of approximation of these parameters by the samples of final length data [1].

Another way of analysis of chaotic process is based on the information theory proposed by E. Shannon [4]. The information theory gives us a quantitative measure of information contained in the sequence of symbols from a given alphabet and it can be applicable to any discrete sequences. According to Shannon, information serves as a measure of "unexpectedness" of revealing each new letter in the symbol sequence under analysis. The conventional entropy [5] may be used as a measure of "unexpectedness" of appearance of new sequences not coordinated with the preceding those, while the entropy change in time, i.e. in elongation of the symbol chain representing the process under analysis, may serve as effective evaluation of the system behaviour's chaotic state.

II. Calculation of the Conventional Entropy Evaluations

Calculation of the conventional entropy involves carrying out steps sequence as follows. First. for the given sample of counts $\{x(i)\}, i=1,2,\dots,N$ a number of L-dimensional sequences will be formed

$$\{X_L(i)=x(i),\, x(i+1), \dots x(i+L-1)\},$$

$$i=1,2,\dots,(N-L+1).$$

Then by means of quantification operation and subsequent transformation from the number one into symbolic form this aggregate can be represented in the form of an aggregate of sequences of the states

$$\{Z_L(i)\}, i=1,2,\dots,(N-L+1).$$

This research was supported by RFBR, research projects Nos 16-01-00159 and 18-07-00264.

978-1-5386-5824-6/18/$31.00 ©2018 IEEE

And each element of the sequence $Z_L(i) = z(i), z(i+1), \ldots, z(i+L-1)$ will be coded with the symbols from arbitrary alphabet $A = \{a_p\},\ p = 1, \ldots, \xi$.

Such a transformation results in regulated i sequence of events. They are connected with the interval evaluation of each count $x(i)$ within the range of $(x_{\min,}\ x_{\max})$ and further interpreted as respective chains of symbols of the length *L*. Changing the *L* parameter one can form the equal length sequences of events.

The entropy conditions will be determined in the form:

$$E(L|L-1) = -\sum_{L-1} p_{L-1} \sum_{L|L-1} p_{L|L-1} \log p_{L|L-1},$$

where $p_{L|L-1}$ specifies the probability of appearance of a concrete symbol in the last *L*-th cell of the *L* length chain under condition of discriminating all identical chains of a shortened length $L-1$.

In the information aspect, this expression evaluates on the average the loss of information associated with appearance of the *L*-th symbol under condition of obtaining the whole subchain of preceding symbols. In other words, the value $E(L|L-1)$ identifies additional information necessary for prediction of the *L*-th event in any of the sequences $\{Z_L(i)\},\ i = 1, 2, \ldots, (N-L+1)$, if the whole sequence of preceding events is known. The conventional entropy tends to zero with *L* amplification for the processes with obvious regular component and takes up a constant positive value for the processes generated by systems with chaotic properties. The latter means a decrease of the prognostication level for behaviour of dynamic system in presence of noise in the process under study.

The value $E(L|L-1)$ can be calculated as an increment obtained by unconditioned entropy $E(L)$ in transition from the event sequence of the length $(L-1)$ to the length *L*:

$$E(L|L-1) = E(L) - E(L-1) = -\sum_{L} p_L \log p_L + \sum_{L-1} p_{L-1} \log p_{L-1}.$$

In analysis of a limited count sample, however, the replacement of the conventional probability $p_{L|L-1}$ distribution with the values of the hit $x(i)$ frequency in respective intervals leads to a considerable shift of the $E(L|L-1)$ evaluation. In the result of this, the value $E(L|L-1)$ with the growth *L* will tend to zero irrespective of the type of the process under analysis.

It is possible to overcome this difficulty by means of correction of the conventional entropy evaluation:

$$\tilde{E}_1(L|L-1) = E(L|L-1)\left(1 + \frac{N_{L-1}^{(1)}}{\left(N_L - N_{L-1}^{(1)}\right)}\right), \qquad (1)$$

or

$$\tilde{E}_2(L|L-1) = E(L|L-1) + E(1) \cdot \frac{N_{L-1}^{(1)}}{N_L}, \qquad (2)$$

where N_L is a number of the length symbol chains L under analysis, $N_{L-1}^{(1)}$ is the number of the length symbol chains $L-1$ met only once.

In expression (1), an entropy average increment is given to the "unknown" chains, the increment being evaluated by $\left(N_L - N_{L-1}^{(1)}\right)$ number of informative sequences of events.

In the formula (2), the rare events are given weights equal to $E(1)$, i.e. corresponding to entropy of single events owing to which the "unknown" chains are regarded as fortuitous and not as the regular sequences. This makes it possible to avoid a false idea of unconfirmed regular changes in the processes under study which is connected to limitation of the data sample size by the value *N*.

Introduction of weights given by the value $E(1)$ was justified in the work [5]; they, however, were given to all single events revealed on the *L*-th step which led to a considerable shift of the corrected entropy. In our works, into the range of absolutely fortuitous events, only those chains are transferred that were represented by a single vector only in the cells $(L-1)$ – the dimensional space of states.

The analysis of dependences of the conventional entropy evaluations under consideration in model signals made it possible to suggest the ways of choosing the informative parameters as follows. First, these could be pointed evaluations of the corrected curve $\tilde{E}_1(L|L-1)$ or $\tilde{E}_2(L|L-1)$ with the values $L = 2, 3, 4,$ i.e. in the area where contribution of singular chains is insignificant. This will enable one to further use them as discriminant signs when solving the classification tasks. Secondly, the lower border of the curve $\tilde{E}_2(L|L-1)$ can be used as measure of process irregularity, the curve taking on small significance for regulatory processes and a great significance in case of presence of obvious chaotic changes in the signals. In the publication [5], to this purpose use of the value of the conventional entropy of $E(1)$ corrected function maximum shift value was substantiated; therefore in this work also evaluation $ME = E(1) - \min_{L=2\ldots10}\left\{\tilde{E}_2(L|L-1)\right\}$ has been introduced.

III. EXPERIMENTAL STUDIES

These considered evaluations were used for analysis of heart rhythm disturbances, particularly in the task of detecting cardiac fibrillation by the cardiac cycle sequence. Fig. 1 shows examples of rhythmograms for three classes of heart rhythm: normal rhythm (*a*), atrial fibrillation (*b*) and frequent ventricular extrasystole (*c*).

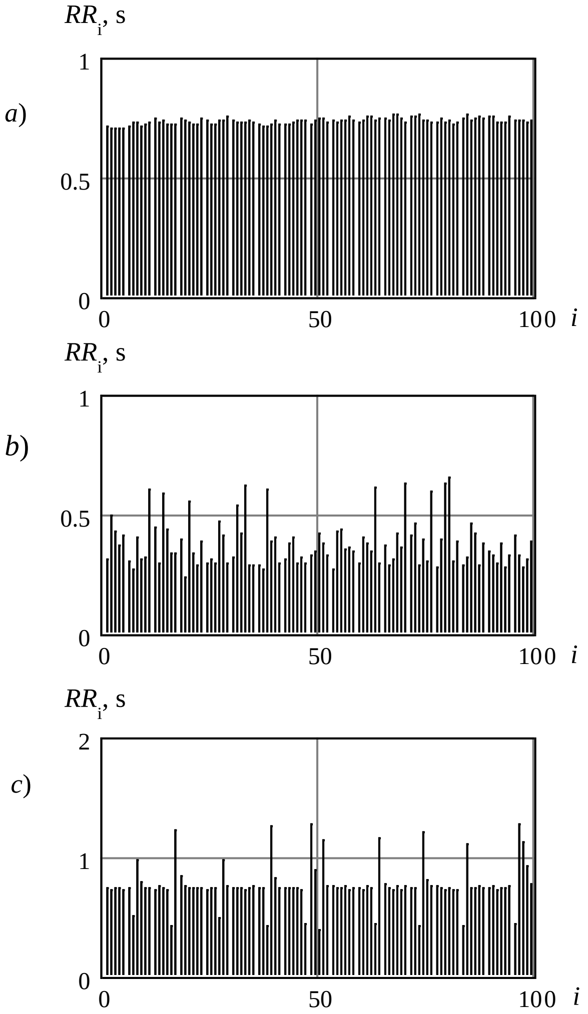

Fig. 1. Examples of rhythmograms for different types of heart rhythm: normal rhythm (a), atrial fibrillation (b) and frequent ventricular extrasystole (c)

The dependences obtained for these realizations $E\left(L|L-1\right)$, $\tilde{E}_1\left(L|L-1\right)$, $\tilde{E}_2\left(L|L-1\right)$; $L=1,\ldots,10$; in the graphical form are shown in Fig. 2. The length of the fragments analyzed was N = 300 samples of the rhythmogram.

As the figures show, the normal rhythm having a wave structure of the rhythmogram is characterised by the parameter *ME* considerably transcending in cardiac fibrillation (*ME* = 0.93, Fig. 2*a*). In case of frequent extrasystoles specific by their short-time regular changes in the RR-interval sequence, a considerable change of the function $\tilde{E}_2\left(L|L-1\right)$ occurs, and the *ME* parameter value is also rather great (*ME* = 0.74, Fig. 2*c*). And only in cardiac fibrillation alone specific by its irregular distribution of the cardiac cycle durations, the *ME* parameter is considerably lesser (*ME* = 0.28, Fig. 2*b*).

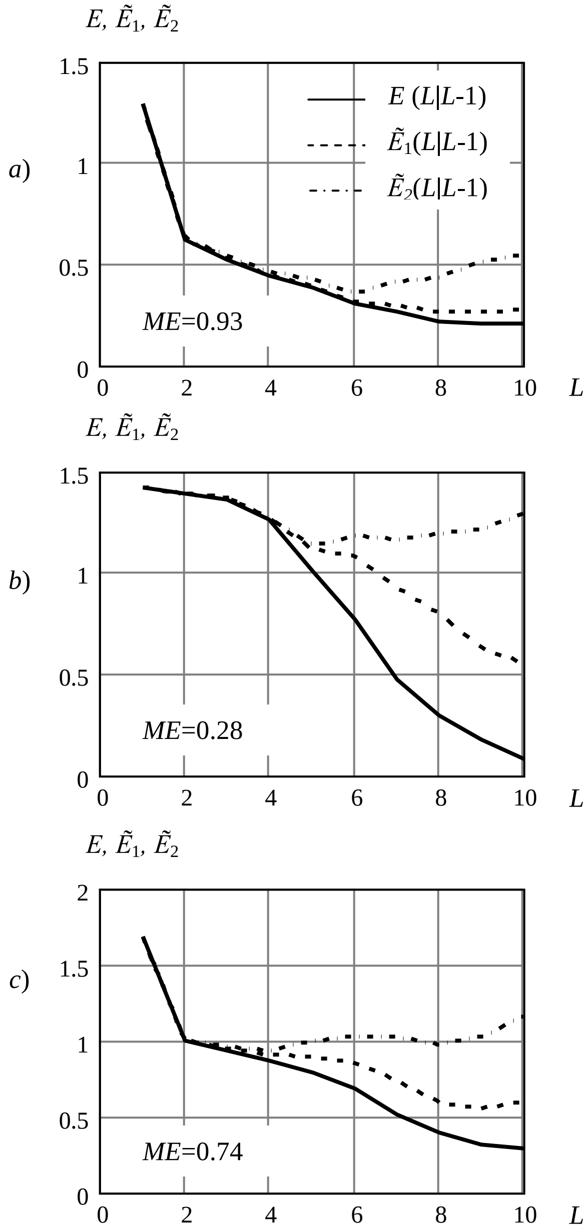

Fig. 2. Dependences of the estimates of the conditional entropy $E\left(L|L-1\right)$, $\tilde{E}_1\left(L|L-1\right)$, $\tilde{E}_2\left(L|L-1\right)$ on the length of chains L for three types of rhythmograms: normal rhythm (*a*), atrial fibrillation (*b*) and frequent ventricular extrasystole (*c*)

The real data experiments enable one to establish the fact that the function $\tilde{E}_2\left(L|L-1\right)$ and the parameter *ME* can be effectively used for evaluation of regularity grade of the changes occurring in temporal rows of a final length and can also be used in the task of detecting the rhythm disturbances with a complex dynamics of cardiac interval distribution. In the study, over 150 records of the cardiac intervals, obtained from the electrocardiographic base MIT-BIH [6], and for them rhythmograms were constructed and the relative minimum of the more precise entropy $\tilde{E}_2\left(L|L-1\right)$ was calculated.

The experimental result analysis showed that the *ME* parameter has stable interval evaluations at the significance level $\alpha = 0.05$ for different types of the rhythmograms under consideration, the value of this parameter being considerably lesser in cardiac fibrillation. Thus for the normal rhythm this value equals (0.85 ± 0.09), in frequent ventricular extrasystoles it is somewhat lesser (0.74 ± 0.10), and it sharply decreases in fits of cardiac fibrillation (0.17 ± 0.11).

At the same time, detection of the rhythmogram chaotic fragments against the background of the rhythm alternative groups specific by their obvious regular changes in cardiac cycle durations will be expedient if performed within the space of the following parameters: $E\left(L\middle|L-1\right),\, L = 2, 3, 4;\, ME$.

The construction of decisive functions was carried out using linear discriminant analysis based on the Fischer criterion [7]. To minimise the classification errors, correction of the dividing function [8, 9] was introduced. This involved finding of the second vector optimising the position of the dividing area borders. The experimental results showed that transition from the parameter *ME* analysis to description of the signals with the proposed set of parameters enabled one to 1.5-fold enhance the efficacy of classification. When recognising the cardiac fibrillation fragments, the mean error of classification did not exceed the value $\varepsilon = 1\%$.

IV. Conclusion

The theoretical and experimental results obtained suggest this approach for identifying the processes of final length with different intensities of discriminated and chaotic components. The parameters $E\left(L\middle|L-1\right),\, L = 2, 3, 4;\, ME$ in particular can be effectively used when recognising cardiac fibrillation fragments in the course of tracing the patient's ECG which is an extremely important task of the cardiomonitoring.

References

[1] Schuster H. G. Deterministic Chaos – An Introduction, 2nd ed., VCH Publishers: Weinheim, Germany, 1988, 270 p.

[2] Rangayyan Rangaraj M. Biomedical Signal Analysis: A Case-Study Approach. Wiley-IEEE Press. 2002. 552 p.

[3] Nonlinear biomedical signal processing /Edited by Metin Akay. Vol. 2, Dynamic analysis and modelling. New York: IEEE. 2001. 341 p.

[4] Shannon C. E. A mathematical theory of communication In: Bell System Technical Journal. 1948. Vol. 27, № 3, pp. 379–423; Vol. 27, № 4, pp. 623–656.

[5] Porta A., Baselli G., Liberati D., Montano N., Cogliati C., Gnecchi-Ruscone T., Malliani A., Cerutti S. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow // Biological Cybernetics. 1998, 78, pp. 71–78.

[6] MIT-BIH Arrhythmia Database. Available from MIT-BIH database distribution // Massachusetts institute of technology. URL: http://www.physio-net.org/physiobank/database/mitdb; http://ecg.mit.edu.

[7] Duda R.O., Hart P.E., Stork D.H. Pattern Classification (2nd ed.). Wiley Interscience. 654 p. (2001).

[8] Manilo L.A. Ordering of spectral attributes by empirical estimations of intergroup distance in problems of biosignals classification // Proceedings of the Russian Universities: Radioelectronics. 2006. № 3. pp. 20–29 (in Russian).

[9] Nemirko A.P. Transformation of feature space based on Fisher’s linear discriminant. Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. 2016. Vol. 26, No 2. pp. 257–261.