A.D. Cherkay, Yu.A. Vlasov. Linguistic Analysis of the Human Heart Rate. – The Issues of Temporal Organization of Living Systems. Department of Physiology, Academy of Sciences of the USSR. Moscow, Nauka Publishers, 1979, pp. 62 - 70.

The heart rate of cardiac contractions is one of the natural output signals issued by a biological system. Watching it closely can help to understand the general logic of the temporal design of biological signals. The actual biological signals contain elements of both determinacy and occasionality. One of the tasks of processing them is identifying and analyzing the determinate elements of such signals. Techniques of data processing being used vary, and one of them is a linguistic method. In recent years, its applications to the analysis and reproduction of the time signal curves and space curves and 3D images have been burgeoning [*Th. Pavlidis*, 1971; *I.B. Muchnik, R.B. Muchnik*, 1973; et al.]. The linguistic methods can be used to analyze the heart rate as well.

The point of linguistic analysis of the sequentially measured RR intervals on an ECG is to determine the pieces within their implementations that repeat to a certain accuracy – the *words* (or *patterns* or *roots sequences*) that are to form a dictionary. The way the specific implementations which constitute a whole language are built up from the words is governed by the formation rules of the language – its grammar. The steadiness of consequential pairs, triads and other groups RR intervals on an ECG [*Ye.N. Meshalkin, Yu.A. Vlasov, A.D. Cherkay*, et al., 1973] has allowed the linguistic approach to be used successfully [*A.D. Cherkay, Yu.A. Vlasov*, 1977] at the heart rate analysis.

We describe here one of the applied procedures of the linguistic analysis of ECG RR intervals. It is based on determining the repeating sequences of RR intervals belonging to an ε-tube for a given ε. The development of matrix sequences and theirs roots is discussed, and the specific examples of the ECG RR interval analysis are brought. We also give an estimation of the statistical significance of condensation of the repeating $X^{(i_k,m_k)}$ sequences and see into the heart rate simulation.

Now let's look upon one of the applied procedures of linguistic analysis of ECG RR intervals.

Let $X^{(1,N)}=\{X_1, \ldots, X_N\}$ be the sequence of N measurements of ECG RR intervals.

Then $X^{(i,m)} = \{X_i, X_{i+1}, \ldots, X_{i+m-1}\}$ is a chain of m elements of the sequence in question placed according to their sequence order in $X^{(1,N)}$ starting with the i-th element ($1 \le i \le N–m+1$, $2<m \le N$). Assign the numbers $k = 1, 2, \ldots$ to various chains $X^{(i_k,m_k)}$ according to the sequence order in $X^{(1,N)}$ of their first element X_{i_k}. Thus, for all $k = 1, 2, \ldots$ we have $i_k < i_{k+1}$.

The subsequences $X^{(i_k,m_k)}$ ($k = 1, 2, \ldots$), for which there exist such $j_k (i_k \le j_k \le i_k + m_k - 1, k = 1,2,...)$ that the following inequality holds for any $q = \underline{m}, \overline{m}$, $\underline{m} = \min_k (i_k - j_k)$, $\overline{m} = \max_k (m_k + i_k - j_k - 1)$

$$\max_k (X_{j_k+q}) - \min_k (X_{j_k+q}) \le \varepsilon, \qquad (1)$$

will be referred to as the chains belonging to the ε-tube. For the set of chains $X^{(i_k,m_k)}$ (k=1, 2, …) that belong to the ε-tube with the fixed j_k ($k = 1, 2, \ldots$), define the values $\max_k (X_{j_k+q})$ and $\min_k (X_{j_k+q})$ as the upper and lower bounds of q-th section of the ε-tube. Let Δm_l be the number

of different sections of a specific ε-tube, each of which contains within its upper and lower bounds not less than l ($l \geq 2$) elements from the chains that form the given ε-tube.

To obtain a separate pattern in the shape of a matrix chain considering $\varepsilon = K_\varepsilon \overline{\mathrm{X}}\left(\overline{\mathrm{X}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{X}_i, K_\varepsilon = 0{,}02 \div 0{,}05\right)$, one should select such triads $\mathrm{X}^{(j_k,3)}(k = 1,2,...)$ from $\mathrm{X}^{(1,N)}$ that they belong to the same ε-tube while the difference between elements in each triad is more than $\varepsilon_1(\varepsilon_1 \geq \varepsilon)$, i.e. one should pick out a set of triads $\mathrm{X}^{(j_k,3)}(k = 1,2,...)$ from $\mathrm{X}^{(1,N)}$ that comply with both inequality (1) for all $q = \overline{0,2}$ and the following inequality

$$\left|\mathrm{X}_{j_k+q} - \mathrm{X}_{j_k+\eta}\right| > \varepsilon_1, \text{ for all } q,\eta = \overline{0,2}, q \neq \eta, k = 1,2,... \qquad (2)$$

Let n such triads be selected from $\mathrm{X}^{(1,N)}$. By expanding these triads rightwards and leftwards by way of adding elements from the sequence in question according to their sequence order $\mathrm{X}^{(1,N)}$ we can form the chains $\mathrm{X}^{(i_k,m_k)} \supseteq \mathrm{X}^{(j_k,3)}(k = 1,2,...)$ belonging to the ε-tubes (complying with inequality (1)) in which each separate element from $\mathrm{X}^{(1,N)}$ with its fixed number can be found in $\mathrm{X}^{(1,N)}$ not more than once, i.e. $i_k + m_k < i_{k+1}(k = \overline{1,n-1})$ chains.

Then from the sets of these chains belonging to separate ε-tubes, such sets are selected that their maximum value is Δm_n; afterwards, from the obtained chain sets with $\Delta m_n = \max$ such sets are taken that their maximum value is Δm_{n-1}; and so on until the $\Delta m_2 = \max$ sets are reached.

Now let the thus gained chain sets belonging to the ε-tubes be called the ε-chain sets with the maximum length and maximum concentration of their elements.

Tied in to the selected ε-chain sets with the maximum length and maximum concentration of their elements are then all the additional chains with three or more elements that can be included in the ε-tubes being built. The augmented chain sets are expanded rightwards and leftwards like triads, while expanded longest chain sets with maximum concentration of their elements are formed, and such sets are picked out from the sets of n' chains $\mathrm{X}^{(i_k',m_k')}$ ($k = \overline{1,n'}$) where $n' > n$, that contain maximum quantity of nonrepeatable elements from $\mathrm{X}^{(1,N)}$. As we enumerate the $\mathrm{X}^{(i_k',m_k')}$ chains according to the order of their appearance in $\mathrm{X}^{(1,N)}$, we have the chains sets where $i_k' + m_k' < i_{k+1}'(k = \overline{1,n-1})$ and $\sum_{k=1}^{n'} m_k' = \max$.

After that those new sets of newly formed ε-chains are further complemented by means of new additional chains from $\mathrm{X}^{(1,N)}$ that are contained in the ε-tubes with the chains already selected on the fragments three or more elements long, and all the chains being used here are expanded again rightwards and leftwards, etc., until among triads from $\mathrm{X}^{(1,N)}$ not yet used in building up the pattern, there is no more such triads that are contained in the tube of any of the chain sets already selected.

For different sets of such chains for each $q = 0,\pm1,\pm2,...(\underline{m} \leq q \leq \overline{m})$, the number of their X_{j_k+q} elements is re-calculated with different $k(n_q)$ and the average $\overline{\mathrm{X}_q}$ are evaluated.

The individual assessed average $\overline{\mathrm{X}_{\underline{m}}}, \overline{\mathrm{X}_{\underline{m}+1}}, ..., \overline{\mathrm{X}_{\overline{m}}}$ values chain composed in the above way is what we use as *word* (*pattern, root sequences*).

In the general case, other sets of patterns can be selected in a similar way, either in order of frequency of occurrence of the $X^{(1,N)}$ triads they are based on, with exception of $X^{(1,N)}$ elements that were already used in building the patterns from the further analysis, or in order of frequency of occurrence in $X^{(1,N)}$ of the chains longer than those triads that belong to the ε-tubes and contain the triads satisfying the condition (2) with or without the additional chains being added to expand their set.

We picked out l ($l< l\geq \overline{m}-m$) consecutive elements from an individual pattern beginning with the q-th (q =0,±1,…) position – $\overline{X^{(q,l)}}$ – and built up two tubes with the radii equal to $\varepsilon/2$ and $3\varepsilon/2$. Then we calculated the number (n) of subsequences taken from $X^{(1,N)}$ with the length l that were contained in the smaller ($n=n_1$) and larger ($n=n_2$) tubes.

The evaluation of the significance of the condensation (concentration) around point $\overline{X^{(q,l)}}$ resolves itself to the use of binominal distribution. The probability for n_1 or more subsequences to accidentally get into the smaller tube when the number of subsequences contained in the larger tube equals to n_2 can be defined by the following relation:

$$P_i = \sum_{i=n_1}^{n_2} C_{n_2}{}^{i} P^i (1-P)^{n_2-i}, \text{ where } P = (1/3)^l.$$

There are statistical tables for P_i. We selected the non-random areas (where $P_i < 0.05$) of the patterns. Let their fragments that are repeated in other patterns with the ε-accuracy be called roots. The roots reveal themselves at picking out the most frequent ε-chains longer than three elements. Furthermore, several ε-tubes can have shared fragments being three or more consecutive elements long.

The procedure described above was used to select *patterns* (*roots sequences*) from the RR interval sequences that comprised 6,070 elements all together taken from seven different surveyed persons.

Table 1 shows an example of extraction of the repeating fragments and their corresponding root sequence from the 750-RR-interval sequence taken from surveyed person A's ECG.

Table 1. Surveyed person A's RR interval fragments
corresponding to an individual root sequence

Numbers of repeating sequence elements	Repeating RR element sequences					
	1	2	3	4	5	6
89-93	0,67	0,70	0,98	0,89	0,70	–
221-225	0,67	0,70	0,97	0,88	–	–
250-254	0,68	0,69	0,98	0,88	0,71	0,71
284-289	0,70	0,70	0,98	0,89	0,70	0,72
336-369	-	0,70	0,99	0,89	0,68	–
Root sequence	**0,68**	**0,70**	**0,98**	**0,89**	**0,70**	0,71

Table 2 displays the root sequence set of this person with the number shown of repeating fragments within $X^{(1,N)}$ used to build it up. Highlighted are the most frequent parts sequences (roots words) from $X^{(1,N)}$.

Table 2. Surveyed person A's set of roots sequences

Quantity of the identified repeating sequences	**Roots sequences (patterns) of RR intervals**									
	1	2	3	4	5	6	7	8	9	10
6	0,68	**0,71**	**0,99**	**0,89**	**0,71**	0,78	0,93	0,92		
5	0,71	0,78	0,96	0,85	0,70	0,80	0,91	0,75	0,76	
7	0,90	0,77	0,95	0,90	0,71	0,79	1,01	0,96	0,85	
7	0,94	0,86	**0,69**	**0,73**	**0,98**	**0,90**	**0,71**	0,75	0,99	
5	**0,68**	**0,70**	**0,98**	**0,89**	**0,70**	0,71				
5	0,68	0,64	**0,68**	**1,02**	**0,92**	**0,71**	0,74			
10	0,93	0,87	0,68	**0,72**	**1,02**	**0,98**	**0,71**	0,74	1,00	0,92

On the basis of three RR interval sequences taken from surveyed person B's ECG, we have set $\varepsilon = 0.05\overline{X}$ and managed to pick out 16 root sequences (patterns) 3 to 21 elements long that we have built using (see Table 3) almost all intervals that were contained in the initial rows 540, 514, and 514 RR-interval long, respectively.

The evaluation of obtained data has let us conclude that the share of $X^{(1,N)}$ elements used to build a separate pattern up to 60%, and on average more 23% of the volume of the piece of initial sequence taken to pick out the pattern from. In other surveyed persons' root sequences the high frequency of elements can also be observed within their initial RR interval sequences. Especially frequent are the roots of the words (patterns) – they are repeated in the samples in question with ε-accuracy in dozens. It suggests a great share of determinacy in the ECG RR interval sequences. In this respect, a slight influence (within ε) of 'random' factors can take place at every step, while their significant influence is also possible but not indispensable at forming every ECG RR interval. Thus, the heart rate is sort of given in advance, i.e. it is set beforehand with a specified degree of accuracy for several step forward. In the simplest case, it allows to represent the ECG RR interval sequence as the repeating fragments of the sequence with the external influence (determinate or accidental) that occur as one moves from one fragment to another and define the order of change for those repeating fragments. This is the simplest kind of hierarchical signals.

Table 3. Characteristics of the repeating RR interval sequence fragments from surveyed person B's ECG

Number of initial sequence	Number of selected pattern	Initial sequence fragment boundaries	Quantity of elements used to pick out patterns from, N	Quantity of the pattern-shaping elements, n	n/N·100%
I	(1)	1-521	521	90	17,3
	(2)	4-546	543	146	26,9
	(3)	31-543	513	96	18,7
	(4)	39-535	497	54	10,9
	(5)	99-383	65	35	53,8
	(6)	58-122	285	12	4,2
	(7)	134-541	408	81	19,9
	(8)	226-463	238	9	3,8
	(9)	340-499	160	17	10,6
Total				540	
II	(10)	1-496	496	49	9,9
	(11)	12-514	503	141	28
	(12)	20-510	491	265	53,9
	(13)	96-409	314	59	18,8
Total				514	
III	(14)	1-518	518	319	61,6
	(15)	8-507	500	139	27,8
	(16)	120-420	301	56	18,6
Total				514	
Average					24,68

It is of interest to draw confirmation of the results achieved through a simpler method. For this purpose, 342 sequences taken from 309 persons at rest at different times of day and year were processed, each of sequences comprising 2,000 ECG RR interval measurements. For each such sequence $X^{(1,N)}$ (N = 2000), the points (X_1,X_2), (X_2,X_3),…,(X_{N-1},X_N) were consecutively marked on the plane with coordinates (x, y), thus creating the complementary interval distributions. Individual areas that have their positions, configurations, and tightnesses are identified within these distributions.

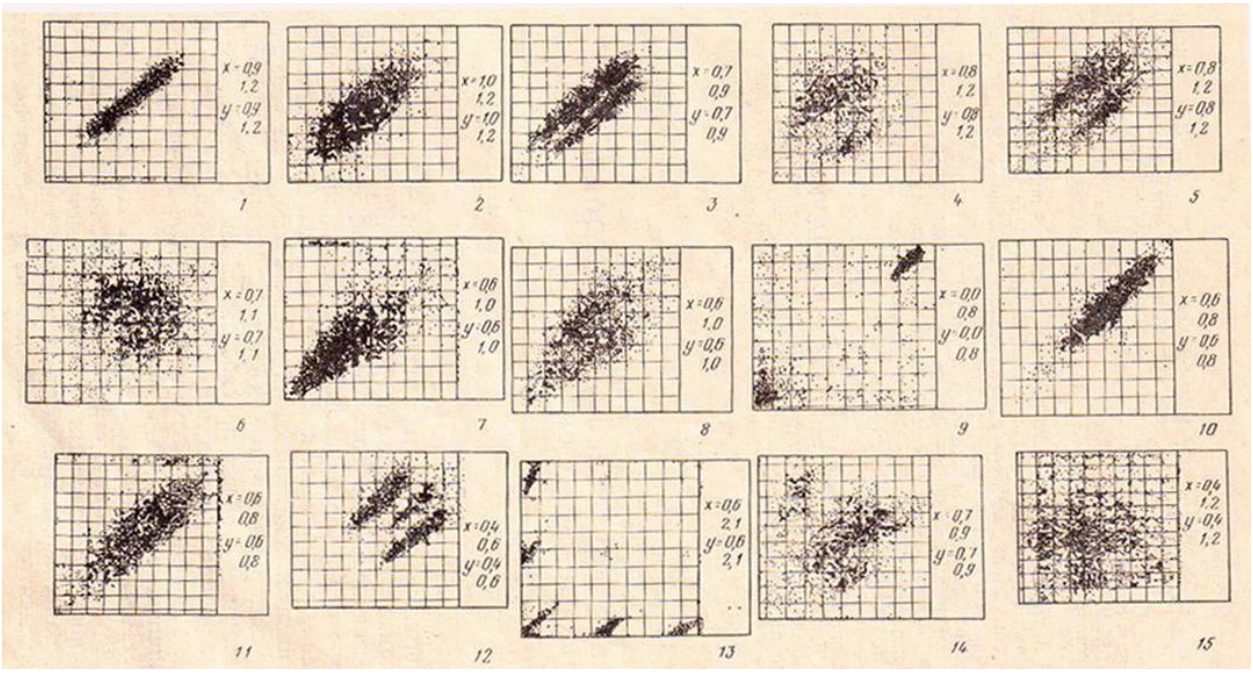
The distributions were divided into 15 types depending on their quantity, position, or appearance. They can be repeated within the data gathered from one person or among various persons.

Distribution type	Its frequency	Distribution type	Its frequency	Distribution type	Its frequency
1	24	6	4	11	12
2	19	7	43	12	36
3	15	8	13	13	29
4	8	9	41	14	19
5	13	10	23	15	43

By connecting the consecutive points on the plane with the line segments, we obtained the motion paths of these points. It was noticed that some parts of the paths coincided or went very close, thus defining the type of the picture being observed.

By so doing we have managed to prove the existence of the repeating and similar sequences (patterns or words) in $X^{(1,N)}$ that define the kind of the regularity typical of one or another kind of sequences that can be found in various people.

Since the discussed types of complementary interval distributions can be the same in different people, we can expect them to be discovered by other experimenters as well. Indeed, the distributions similar to types 1, 12, 15 were known before [*P.A. Goldstein, G.O. Barnett*, 1967; *W.K. Haisty et al*., 1972], while distributions close to types 12 ad 13 were encountered even when they were evaluating the neuronal activity [*M. Biederman-Thorson*, 1966; *E. David et al*., 1968]. The latter fact tells us that various impulse signals of our body can be possibly generated in a similar way.



Identifying 15 types of joint distributions of ECG RR intervals

As they are being built, the i-th RR interval is put on the horizontal axis, and the next, (i+1)-th RR interval (i=1, 2, …, 1999) is put on the vertical one.

REFERENCES

Ye.N. Meshalkin, Yu.A. Vlasov, A.D. Cherkay, N.N. Glushkov, N.N. Melnikova. One approach to identification of random and regular elements in the heart rate management strategy. – Krovoobrashcheniye, 1973, **6**, №5, c. 12-16.

I.B. Muchnik, R.B. Muchnik. Formation algorithms of the curve description language. -- Avtomatika I Telemekhanika, № 5, 1973.

A.D. Cherkay, Yu.A. Vlasov. Linguistic analysis of the sequences of time intervals between the heart beats – in Theory and Practice of Electrocardiographic and Clinical Research Automatization. – Kaunas, 1977.; pp. 128 - 131.

M. Biederman-Thorson Source mechanisms for unit activity in isolated crayfish central nervous system. – J. Gen. Physiol., 1966, **49**, N 4, p. 597-612.

E. David, P. Finkenzeller, S. Kallet, W.D. Keidel. Die Bedeutung der temporalen Hemmung im Bereich der akustischen informationsverarbeitung. – Pflűgers Arch., 1968, **298**, N 4, p. 322-335.

P.A. Goldstein, G.O. Barnett. A statistical study of atrial fibrillation. – Comput. and Biomed. Res., 1967, **5**, N 3, p. 247-255.

W.K. Haisty, C. Batchlot, J. Cornfield, H.V. Pipberger. Discriminant function analysis of RR interval: an algoritm for on-line arrhythmia diagnosis. – Comput. and Biomed. Res., 1972, **5**, N 3, p. 247-255.

Th. Pavlidis. Linguistic analysis of waveforms. – Software Eng., 1971, 2, N 4, p. 203-225.