Approaches to Statistical Processing of Rhythmocardiosignal with Increased Resolution

Iaroslav Lytvynenko[1[0000-0001-7311-4103]], Serhii Lupenko[1[0000-0002-6559-0721]], Vyacheslav Kharchenko[2[0000-0001-5352-077X]1], Andrii Horkunenko[3[0000-0001-8644-0776]], Andrii Zozulya[1[0000-0003-1582-3088]]

[1]Ternopil Ivan Puluj National Technical University, Department of Computer Science, Ternopil, Ukraine
iaroslav.lytvynenko@gmail.com
[2]National Aerospace University, Kharkiv, Ukraine
v.kharchenko@csn.khai.edu
[3]I. Horbachevsky Ternopil National Medical University, Department of Medical Physics of Diagnostic and Therapeutic Equipment, Ternopil, Ukraine
horkunenkoab@tdmu.edu.ua

Abstract. The paper is devoted to statistical methods estimation of probabilistic characteristics of rhythmocardiosignal with increased resolution on the basis of its model in the form of a vector of stationary and stationary related random processes. The hypothesis about the normality of the law of components distribution of the rhythmocardiosignal with increased resolution is confirmed. It was made a decomposition of the statistical estimates of autocorrelation and inter-correlation functions allowed to obtain spectral and inter-spectral power density of vector components, that allowed to reduce the space dimension of diagnostic features in heart rate analysis systems based on rhythmocardiosignals with increased resolution. That allowed to substantiate the vector of diagnostic features in the systems of cardiac rhythm analysis based on the rhythmocardiosignals with increased resolution is substantiated.

Keywords: methods of statistical estimation, probabilistic characteristics, vector of stationary and stationary-related random sequences, electrocardiogram, rhythmocardiosignal, heart rate.

1 Introduction

Automated heart rhythm analysis systems make it possible to evaluate both the state of the cardiovascular system and the state of the adaptive capacity of the human body as a whole. Most modern heart rate analysis systems are based on the use of stochastic mathematical models of rhythmocardiosignal and methods of its statistical analysis by rhythmocardiogram, which is an ordered set of durations of R-R intervals in a registered electrocardiosignal [1-8].

However, this approach makes it impossible to detect subtle, more detailed features of the heart rhythm, since RR intervals reflect only the change in the duration of the cardiac cycles, and not the whole totality of time intervals between single-phase values of the electrocardiosignal, which makes it impossible to describe the rhythm of hearts in full.

In papers [9,10], in order to provide a more informative description of the heart rhythm, a new approach to its analysis based on rhythmocardiosignal with increased resolution has been substantiated. The classical rhythmocardiogram is embedded in the increased-resolution rhythmocardiogram, which is the basis for increasing the level of informativeness of the heart rate analysis in modern computer systems of functional diagnostics of the human heart condition based on the rhythmocardiogram with increased resolution.

In papers [9, 10], the use of a random variable vector as a mathematical model of rhythmocardiosignal with increased resolution is substantiated. However, this model is a relatively poor mathematical model of rhythmocardiosignal with increased resolution, since it does not allow to study its time dynamics. To take into account the time dynamics of rhythmocardiosignal with increased resolution, it is necessary to use a mathematical apparatus of the theory of random processes, that is, to consider it as a vector of discrete-time random processes.

In this paper, we develop methods for the statistical estimation of the probabilistic characteristics of a rhythmocardiogram with high resolution based on its model in the form of a vector of stationary and stationary related random sequences.

2 Methods

One of the simplest stochastic models that take into account the dynamics of rhythmocardiosignal with increased resolution is the vector of $\boldsymbol{\Xi}_L(\omega', m) = \left\{ T_l(\omega', m), \omega' \in \boldsymbol{\Omega}', l = \overline{1, L}, m \in \mathbf{Z} \right\}$ stationary and stationary related random processes. In this vector, the index m indicates the cycle number of the electrocardiosignal, and the index l - the reference number of the electrocardiogram within it m-th cycle. The number of samples L per electrocardiosignal cycle determines the resolution of the rhythmcardiosignal, and specifies the number of phases in the cycle of the electrocardiosignal that can be separated by methods of segmentation and detection in solving the problem of automatic formation of the rhythmocardiosignal from the electrocardiosignal.

The defining property of a vector $\boldsymbol{\Xi}_L(\omega', m)$ of stationary and stationary related random sequences is the invariance of its family of distribution functions to time shifts by an arbitrary integer $k \in \mathbf{Z}$. Namely, for any distribution function $F_{p_{T_{l_1} \ldots T_{l_p}}}\left(x_1, \ldots, x_p, m_1, \ldots, m_p\right)$ order p $\left(p \in \mathbf{N} \right)$ from the family of vector distribution functions $\boldsymbol{\Xi}_L(\omega', m)$ of stationary and stationary random sequences such equality occurs:

$$F_{p_{T_{l_1} \ldots T_{l_p}}}\left(x_1, \ldots, x_p, m_1, \ldots, m_p\right) = F_{p_{T_{l_1} \ldots T_{l_p}}}\left(x_1, \ldots, x_p, m_1 + k, \ldots, m_p + k\right),$$

$$x_1,\dots,x_p \in \mathbf{R},\ m_1,\dots,m_p \in \mathbf{Z},\ l_1,\dots,l_p \in \left\{\overline{1,L}\right\},\ k \in \mathbf{Z}\,. \qquad (1)$$

We present formulas that represent the convergence in the root-mean-square sense of the corresponding statistical estimates to the estimated probabilistic characteristics of the vector of $\mathbf{\Xi}_L(\omega', m) = \{T_l(\omega', m),\ \omega' \in \mathbf{\Omega}',\ l = \overline{1,L},\ m \in \mathbf{Z}\}$ stationary and stationary related random sequences.

An estimate that converge in the root-mean-square sense to the distribution function $F_{p_{T_{l_1}\dots T_{l_p}}}(x_1,\dots,x_p,\ m_1,\dots,m_p)$ of order p $(p \in \mathbf{N})$ of vector of $\mathbf{\Xi}_L(\omega', m)$ stationary and stationary related random sequences:

$$F_{p_{T_{l_1}\dots T_{l_p}}}(x_1,\dots,x_p,\ m_1,\dots,m_p) = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K}\prod_{j=1}^{p} H\left(x_j - T_{l_j}(\omega', m_j + k)\right),$$
$$x_1,\dots,x_p \in \mathbf{R},\ m_1,\dots,m_p \in \mathbf{Z},\ l_1,\dots,l_p \in \left\{\overline{1,L}\right\},\ k \in \mathbf{Z}. \qquad (2)$$

Function $H(x) = \begin{cases} 1, x \ge 0, \\ 0, x < 0. \end{cases}$ is a Heaviside step function, which is an indicator of a negative number.

In particular, if in the formula (2) $p = 1$, that is $l_1 = l_2 = \dots = l_p = l$, then we will have one-dimensional $F_{1_{T_l}}(x) = F_{1_{T_l}}(x, m)$ distribution auto-function of stationary random sequence $T_l(\omega', m)$, for which from the formulas (2) it follows a convergence in the root-mean-square sense:

$$F_{1_{T_l}}(x) = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K} H\left(x - T_l(\omega', k)\right),\ x \in \mathbf{R},\ l \in \left\{\overline{1,L}\right\},\ k \in \mathbf{Z}. \qquad (3)$$

An estimate that converge in the root-mean-square sense to the mixed initial moment function of order $s = \sum_{j=1}^{p} s_j$:

$$c_{s_{T_{l_1}\dots T_{l_p}}}(m_1,\dots,m_p) = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K} T_{l_1}^{s_1}(\omega', m_1 + k)\cdot\ldots\cdot T_{l_p}^{s_p}(\omega', m_p + k),$$
$$m_1,\dots,m_p \in \mathbf{Z},\ l_1,\dots,l_p \in \{\overline{1,L}\},\ k \in \mathbf{Z}\,. \qquad (4)$$

If $s = 2$ and $p = 2$, then from the formula (4) follows such a convergence in the root-mean-square sense for the covariance function $c_{s_{T_{l_1}T_{l_2}}}(m_1, m_2)$ two stationary and stationary connected random sequences $T_{l_1}(\omega', m)$ and $T_{l_2}(\omega', m)$, describing the time distances between single-phase electrocardiosignal samples for l_1-st and l_2-d its phases, such as:

$$c_{2_{T_{l_1}T_{l_2}}}(m_1, m_2) = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K} T_{l_1}(\omega', m_1 + k)\cdot T_{l_2}(\omega', m_2 + k),$$

$$m_1, m_2 \in \mathbf{Z},\ l_1, l_2 \in \{\overline{1, L}\},\ k \in \mathbf{Z}\ . \tag{5}$$

If in the formula (5) $p = 1$, that is $l_1 = l_2 = ... = l_p = l$, then we have the convergence of the estimate in the root-mean-square sense to the one-dimensional initial moment function $c_{s_{T_l}}(m)$ s-th order, which is for a stationary random sequence $T_l(\omega', m)$ is a constant $c_{s_{T_l}} = c_{s_{T_l}}(m)$ (the initial moment s-th order), that is:

$$c_{s_{T_l}} = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K} T_l^s(\omega', k),\ l \in \left\{\overline{1, L}\right\}. \tag{6}$$

If in the formula (6) $s = 1$, then we will have the convergence of the estimate in the root-mean-square sense to the initial moment of the first order $c_{1_{T_l}} = c_{1_{T_l}}(m)$ (mathematical expectation) stationary random sequence $T_l(\omega', m)$, that is:

$$c_{1_{T_l}} = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K} T_l(\omega', k),\ l \in \left\{\overline{1, L}\right\}. \tag{7}$$

An estimate that converge in the root-mean-square sense to the mixed initial central function of order $s = \sum_{j=1}^{p} s_j$:

$$r_{s_{T_{l_1}\ldots T_{l_p}}}(m_1, ..., m_p) = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K}\left(T_{l_1}(\omega', m_1 + k) - c_{1_{T_{l_1}}}\right)^{s_1}\cdot ...\cdot\left(T_{l_p}(\omega', m_p + k) - c_{1_{T_{l_p}}}\right)^{s_p},$$

$$m_1, ..., m_p \in \mathbf{Z},\ l_1, ..., l_p \in \{\overline{1, L}\},\ k \in \mathbf{Z}\ . \tag{8}$$

If $s = 2$ and $p = 2$, then from the formula (8) such a convergence follows in the root-mean-square sense for the correlation function $r_{s_{T_{l_1}T_{l_2}}}(m_1, m_2)$ two stationary and stationary connected random sequences $T_{l_1}(\omega', m)$ and $T_{l_2}(\omega', m)$, describing the time intervals between single-phase electrocardiogram samples for l_1-th and l_2-d its phases, that is:

$$r_{2_{T_{l_1}T_{l_2}}}(m_1, m_2) = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K}\left(T_{l_1}(\omega', m_1 + k) - c_{1_{T_{l_1}}}\right)\cdot\left(T_{l_2}(\omega', m_2 + k) - c_{1_{T_{l_2}}}\right),$$

$$m_1, m_2 \in \mathbf{Z},\ l_1, l_2 \in \{\overline{1, L}\},\ k \in \mathbf{Z}\ . \tag{9}$$

If in the formula (9) $s = 2$ and $p = 1$, that is $l_1 = l_2 = ... = l_p = l$, then we will have a convergence of the estimate in the root-mean-square sense to the variance $r_{2_{T_l}}$ stationary random sequence $T_l(\omega', m)$, that is:

$$r_{2_{T_l}} = \underset{K\to\infty}{l.i.m.}\frac{1}{2K+1}\sum_{k=-K}^{K}\left(T_l\left(\omega',k\right)-c_{1_{T_l}}\right)^2,\ l\in\left\{\overline{1,L}\right\}. \qquad (10)$$

The above formulas reflect the convergence in the root-mean-square sense of the statistical estimates to the corresponding probabilistic characteristics of rythmo-cardisignal with increased resolution, and, therefore, the statistical estimates are consistent.

Since in real computer systems of cardiac rhythm analysis the finite number of cycles of electrocardiosignal is always recorded, this fact should be taken into account also in the statistical estimation of probabilistic characteristics of the rhythmocardiosignal with increased resolution. Namely, the statistical evaluation of the probabilistic characteristics of the rhythmocardiosignal with increased resolution is to obtain the realizations of statistical estimates that can be taken as pproximation to the corresponding probabilistic characteristics of the analyzed rhythmocardiosignal.

We write down the expressions to calculate the realizations of the corresponding statistical estimates of the probabilistic characteristics of the vector of $\mathbf{\Xi}_L(\omega', m) = \left\{T_l(\omega', m), \omega' \in \mathbf{\Omega}', l = \overline{1,L}, m \in \mathbf{Z}\right\}$ stationary and stationary related random sequences when some long realization is given $\mathbf{\Xi}_{L_{\omega'}}(m) = \left\{T_{l_{\omega'}}(m),\ l = \overline{1,L}, m = \overline{1,M}\right\}$, where M - the number of registered complete cycles from which the rhythmocardiosignal with increased resolution is formed.

An expression for calculating the realization of a statistical estimate of a distribution function $F_{p_{T_{l_1}\dots T_{l_p}}}\left(x_1,\dots,x_p, m_1,\dots,m_p\right)$ of order p ($p\in\mathbf{N}$) of vector $\mathbf{\Xi}_L(\omega', m)$ stationary and stationary related random sequences looks like:

$$\hat{F}_{p_{T_{l_1}\dots T_{l_p}}}\left(x_1,\dots,x_p, m_1,\dots,m_p\right) = \frac{1}{M-M_1+1}\sum_{k=0}^{M-M_1}\prod_{j=1}^{p} H\left(x_j - T_{l_{j\omega'}}\left(m_j+k\right)\right),$$
$$x_1,\dots,x_p \in \mathbf{R},\ m_1,\dots,m_p \in \left\{\overline{1,M_1}\right\},\ l_1,\dots,l_p \in \left\{\overline{1,L}\right\}. \qquad (11)$$

Where $M_1 (M_1 \ll M)$ – maximum value of arguments $m_1,\dots,m_p$, which is selected depending on the number of averages in the realization of statistics to provide the required level of accuracy and assurance of statistical estimation.

In particular, if in the formula (11) $p=1$, that is $l_1 = l_2 = \dots = l_p = l$, then we will have an expression to calculate the realization of the statistical estimate $\hat{F}_{1_{T_l}}\left(x\right)$ one-dimensional auto-distribution function $F_{1_{T_l}}\left(x\right)$ stationary random sequence $T_l(\omega', m)$, that is:

$$\hat{F}_{1_{T_l}}\left(x\right) = \frac{1}{M-M_1+1}\sum_{k=0}^{M-M_1} H\left(x - T_{l_{\omega'}}\left(k\right)\right),\ x\in\mathbf{R},\ l\in\left\{\overline{1,L}\right\}. \qquad (12)$$

An expression to calculate the realization of a statistical estimate of a mixed initial moment function of order $s=\sum_{j=1}^{p} s_j$ is given:

$$\hat{c}_{s_{T_{l_1}\ldots T_{l_p}}}\left(m_1,\ldots,m_p\right)=\frac{1}{M-M_1+1}\sum_{k=0}^{M-M_1} T_{l_{1_{\omega'}}}^{s_1}\left(m_1+k\right)\cdot\ldots\cdot T_{l_{p_{\omega'}}}^{s_p}\left(m_p+k\right),$$

$$m_1,\ldots,m_p\in\{1,M_1\},\, l_1,\ldots,l_p\in\{\overline{1,L}\}\,. \quad (13)$$

If $s=2$ and $p=2$, then from formula (13) follows the expression to calculate the realization of the statistical estimate $\hat{c}_{s_{T_{l_1}T_{l_2}}}\left(m_1,m_2\right)$ of the covariance function $c_{s_{T_{l_1}T_{l_2}}}\left(m_1,m_2\right)$ two stationary and stationary-related random sequences $T_{l_1}(\omega',m)$ and $T_{l_2}(\omega',m)$, that describe the time distances between single-phase samples of electrocardiosignal for l_1-st and l_2-d its phases, in particular:

$$\hat{c}_{2_{T_{l_1}T_{l_2}}}\left(m_1,m_2\right)=\frac{1}{M-M_1+1}\sum_{k=0}^{M-M_1} T_{l_{1_{\omega'}}}\left(m_1+k\right)\cdot T_{l_{2_{\omega'}}}\left(m_2+k\right),$$

$$m_1,m_2\in\left\{\overline{1,M_1}\right\},\, l_1,l_2\in\left\{\overline{1,L}\right\}. \quad (14)$$

If in the formula (13) $p=1$, that is $l_1=l_2=\ldots=l_p=l$, then we get an expression to calculate the realization of the statistical estimate $\hat{c}_{s_{T_l}}$ the initial moment s-th order $c_{s_{T_l}}$ of stationary random sequence $T_l(\omega',m)$, in particular:

$$\hat{c}_{s_{T_l}}=\frac{1}{M-M_1+1}\sum_{k=0}^{M-M_1} T_{l_{\omega'}}^{s}\left(k\right),\, l\in\left\{\overline{1,L}\right\}. \quad (15)$$

If in the formula (15) $s=1$, then we get an expression to calculate the realization of the statistical estimate $\hat{c}_{1_{T_l}}$ of the initial moment of the first order $c_{1_{T_l}}$ (mathematical expectation) stationary random sequence $T_l(\omega',m)$, that is:

$$\hat{c}_{1_{T_l}}=\frac{1}{M-M_1+1}\sum_{k=0}^{M-M_1} T_{l_{\omega'}}\left(k\right),\, l\in\left\{\overline{1,L}\right\}. \quad (16)$$

An expression to calculate the realization of a statistical estimate of a mixed initial central function of order $s=\sum_{j=1}^{p} s_j$ is given:

$$\hat{r}_{s_{T_{l_1}\ldots T_{l_p}}}\left(m_1,\ldots,m_p\right)=\frac{1}{M-M_1+1}\sum_{k=0}^{M-M_1+1}\left(T_{l_{1_{\omega'}}}\left(m_1+k\right)-\hat{c}_{1_{T_{l_1}}}\right)^{s_1}\cdot\ldots\cdot\left(T_{l_{p_{\omega'}}}\left(m_p+k\right)-\hat{c}_{1_{T_{l_p}}}\right)^{s_p},$$

$$m_1,\ldots,m_p\in\{\overline{1,M_1}\},\, l_1,\ldots,l_p\in\{\overline{1,L}\}\,. \quad (17)$$

If $s=2$ and $p=2$, then from the formula (17) follows the expression to calculate the realization of the statistical estimation of the correlation function $r_{s_{T_{l_1}T_{l_2}}}\left(m_1,m_2\right)$ of two stationary and stationary related random sequences $T_{l_1}(\omega', m)$ and $T_{l_2}(\omega', m)$, that describe the time distances between single-phase samples of electrocardiosignal for l_1 -st and l_2 -d its phases, that is:

$$\hat{r}_{2_{T_{l_1}T_{l_2}}}(m_1, m_2) = \frac{1}{M - M_1 + 1} \sum_{k=0}^{M-M_1} \left(T_{l_{1_{\omega'}}}(m_1 + k) - \hat{c}_{1_{T_{l_1}}} \right) \cdot \left(T_{l_{2_{\omega'}}}(m_2 + k) - \hat{c}_{1_{T_{l_2}}} \right),$$

$$m_1, m_2 \in \{\overline{1, M_1}\}, l_1, l_2 \in \{\overline{1, L}\}. \tag{18}$$

Since for stationary and stationary random sequences, correlation functions are functions of only one integer argument u, which is equal to $u = m_1 - m_2$, then their statistical estimates also depend on only one argument u. In this case, assuming the ergodicity of the stationary components of the vector $\Xi_L(\omega', m)$, then the formula (18) will look like this:

$$\hat{r}_{2_{T_{l_1}T_{l_2}}}(u) = \hat{r}_{2_{T_{l_1}T_{l_2}}}(m_1 - m_2) = \frac{1}{M - M_1 + 1} \sum_{k=0}^{M-M_1} \left(T_{l_{1_{\omega'}}}(k) - \hat{c}_{1_{T_{l_1}}} \right) \cdot \left(T_{l_{2_{\omega'}}}(k + u) - \hat{c}_{1_{T_{l_2}}} \right),$$

$$u = \overline{0, M_1 - 1}, m_1, m_2 \in \{\overline{1, M_1}\}, l_1, l_2 \in \{\overline{1, L}\}. \tag{19}$$

If in the formula (19) $u = 0$, a $l_1 = l_2 = l$, then we will have an expression to calculate the realization of the variance estimate $r_{2_{T_l}}$ stationary random sequence $T_l(\omega', m)$, that is:

$$\hat{r}_{2_{T_l}} = \frac{1}{M - 1} \sum_{k=1}^{M} \left(T_l(\omega', k) - c_{1_{T_l}} \right)^2, l \in \left\{ \overline{1, L} \right\}. \tag{20}$$

3 Normality hypothesis test of the vector components

The most comprehensive information on the probabilistic characteristics of a increased-resolution rhythmocardiogram is contained in the distribution function family $\left\{ F_{p_{T_{l_1} \dots T_{l_p}}}(x_1, \dots, x_p, m_1, \dots, m_p), p \in \mathbf{N}, l_1, \dots, l_p \in \left\{ \overline{1, L} \right\} \right\}$ of vector $\Xi_L(\omega', m)$ stationary and stationary-related random sequences, and all the other probabilistic characteristics (mixed, central, initial moment functions of different orders) are derived from this family. However, due to the high computational complexity of the methods of statistical estimation of multidimensional vector distribution functions $\Xi_L(\omega', m)$, it is necessary to study increased-resolution rhymocardialsignals to substantiate their types of distribution, in particular, to test the statistical hypothesis for the normality (Gaussian) of the stationary components of a vector, which, if confirmed, will allow us to apply the model of the studied rhythmocardiogram within the framework of spectral-correlation theory, in particular instead of a tedious, computationally complex estimation of distribution functions, to apply simpler computational procedures for estimating spectral-correlation characteristics of rhythmocardiosignals with increased resolution.

Let's test the hypothesis for normality of the distribution law of components of a vector $\Xi_L(\omega', m)$. To do this, we apply the Pearson's agreement criterion (χ^2 -test), that allow to establish consistency (or inconsistency) of empirical and theoretical distributions of vector components $\Xi_L(\omega', m)$. Empirical distribution of vector components $\Xi_L(\omega', m) = \left\{ T_l(\omega', m), \omega' \in \mathbf{\Omega}', l = \overline{1, L}, m = \overline{1, M} \right\}$, estimated by making a histogram. That is, the interval at which all the values of the realization fall into $T_{l_{\omega'}}(m)$ l -th component $T_l(\omega', m)$ divided into I subintervals $\{(S_i^l, S_{i+1}^l), i = \overline{1, I}\}$ with durations $\{\Delta_i^l = S_{i+1}^l - S_i^l, i = \overline{1, I}\}$ and for each interval S_i^l the number is calculated h_i^l (empirical frequency), which is equal to the ratio of the number of realization values $T_{l_{\omega'}}(m)$ l -th component $T_l(\omega', m)$, falling into the interval Δ_i^l, to their total number M , that is:

$$h_i^l = \frac{n_i^l}{\Delta_i^l \cdot M}, i = \overline{1, I}, l = \overline{1, L} \cdot \quad (21)$$

The set of pairs $\{(\Delta_i^l, h_i^l), i = \overline{1, I}\}$ for realization $T_{l_{\omega'}}(m)$ l -th component $T_l(\omega', m)$ can be presented either as a table or graphically as a histogram, for example in Figure 2 shows the results of such calculations for the components of the vector.
In the test χ^2 - square as a measure of the empirical frequency deviation h_i^l from the corresponding theoretical probability p_i^l the value is used

$$\chi^2 = \sum_{i=1}^{I} \frac{\left(\frac{h_i^l}{M} - p_i^l \right)^2}{p_i^l}. \quad (22)$$

Value χ^2 in expression (22) is a random variable, the distribution of which at $M \to \infty$, tend to χ^2 - distribution $P_q(x)$, which depends on the parameter q, which is called the number of freedom degrees equal to:

$$q = I - s - 1, \quad (23)$$

where s - the number of theoretical distribution parameters, against which the hypothesis about the consistency of empirical and theoretical distributions is tested. In the case of the normal distribution of the stationary components of the vector $\Xi_L(\omega', m)$ $s = 2$.

Application of χ^2 - test implies that some level of significance is given previously α (for example, $\alpha = 0.01; \alpha = 0.05$), which makes it possible to calculate the quantile $\chi_{q\alpha}^2$ of distribution χ^2 for a given α and q. If the value χ^2, calculated by the formula (22), more than $\chi_{q\alpha}^2$, then it is considered, that a theoretical distribution (for example, normal) is in poor agreement with the results of observations at a given level of significance α . Conversely, if the value is calculated χ^2 less than $\chi^2_{q\alpha}$, then it is considered, that the theoretical and empirical distributions is in good agreement.

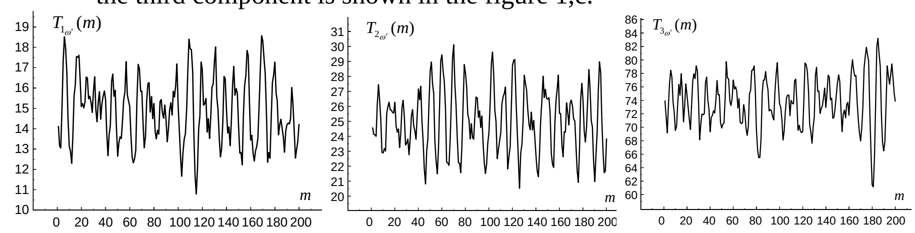
4 The results of statistical analysis

In order to obtain the probable result of verification for the normality of the law of distribution of the rhythmocardialsignal with the increased resolution, the realization of the electrocardiosignal in the second lead, which contained 245 cardiac cycles and was generated by the work of the heart of the patient with a conditional norm, was processed. From the registered electrocardiogram according to the method of automatic formation of rhythmocardiogram with high accuracy, the realization received $\mathbf{\Xi}_{3_{\omega'}}(m) = \{T_{l_{\omega'}}(m), l = \overline{1,3}, m = \overline{1,245}\}$ of tricomponent vector $\mathbf{\Xi}_3(\omega', m) = \{T_l(\omega', m), \omega' \in \mathbf{\Omega}', l = \overline{1,3}, m = \overline{1,245}\}$ of stationary and stationary-related random sequences.

The first component $T_1(\omega', m)$ of this vector is a random stationary sequence, which describes the duration P - intervals in the electrocardiosignal for all of its 245 recorded cycles.

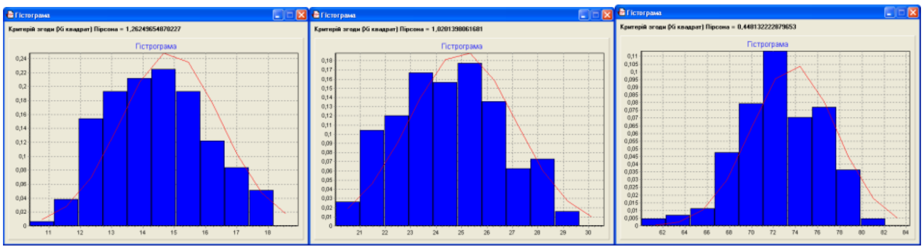
The plot of realization $T_{1_{\omega'}}(m)$ of this component is shown in Figure 1,a. Second component $T_2(\omega', m)$ of this vector is a random stationary sequence, which describes the duration R - intervals in the electrocardiosignal.

The plot of realization $T_{2_{\omega'}}(m)$ of the second component is shown in the figure 1,b. The third component $T_3(\omega', m)$ of this vector is a random stationary sequence, which describes the duration T - intervals in the electrocardiosignal. Plot of realization $T_{3_{\omega'}}(m)$ the third component is shown in the figure 1,c.



a) b) c)

Fig. 1. Plot of realization: a) $T_{1_{\omega'}}(m)$ the first component $T_1(\omega', m)$, which describes the duration P - intervals in the electrocardiosignal; b) $T_{2_{\omega'}}(m)$ the second component $T_2(\omega', m)$, which describes the duration R - intervals in the electrocardiosignal; c) $T_{3_{\omega'}}(m)$ third component $T_3(\omega', m)$, which describes the duration T - intervals in the electrocardiosignal



a) b) c)

Fig. 2. Histogram for: a) the first component $T_1(\omega', m)$, which describes the duration P - intervals in the electrocardiosignal; b) the second component $T_2(\omega', m)$, which describes the duration R - intervals in the electrocardiosignal; c) third components $T_3(\omega', m)$, which describes the duration T - intervals in the electrocardiosignal

The number of freedom degrees was chosen equal $q = 7$, level of signification $\alpha = 0.05$, i, respectively quantile χ^2 - distribution with q freedom degrees $\chi^2_{0.95,7} = 14.07$. Figures 2, a-c show histograms for realizations $T_{1_{\omega'}}(m)$, $T_{2_{\omega'}}(m)$ and $T_{3_{\omega'}}(m)$ the corresponding three stationary components of the vector $\Xi_3(\omega', m)$.

Table 1 shows the results of application χ^2 - test for checking the normality of the law of distribution of three stationary components of a vector $\Xi_3(\omega', m)$, that set the rhythmcardiosignal with increased resolution.

Table 1. The results of application

Stationary component number	**Quantile value $\chi^2_{q\alpha}$ at $\alpha = 0.05$ and $q = 7$**	**The value of realization a random variable χ^2**	**Hypothesis testing result**
1	14,067	1,26	$\chi^2 < \chi^2_{q\alpha}$ (confirmed)
2	14,067	1,02	$\chi^2 < \chi^2_{q\alpha}$ (confirmed)
3	14,067	0,49	$\chi^2 < \chi^2_{q\alpha}$ (confirmed)

Thus, based on the results of normality hypothesis testing of the distribution of the stationary components of a random vector $\Xi_L(\omega', m)$ by Pearson's criterion, it is found that these results do not contradict the hypothesis for normality of its distribution. Normality of the vector $\Xi_L(\omega', m)$ is the basis for the substantiation of diagnostic features in systems of cardiac rhythm analysis according to rhythmocardiogram with increased resolution within the spectral-correlation theory, which significantly reduces the computational complexity of such analysis. In this case, to evaluate the probabilistic structure of the vector $\Xi_L(\omega', m)$ of stationary and stationary-related random sequences, it is sufficient to statistically evaluate only the vector

$\mathbf{C}_L^1 = \{c_{1_{T_l}}, l = \overline{1,L}\}$ his mathematical expectations according to formula (16) and the matrix of correlation functions $\mathbf{R}_T = [r_{2_{T_{l_1}T_{l_2}}}(u), l_1, l_2 = \overline{1,L}]$ according to formula (19).

5 Choice substantiation of diagnostic features in cardiac rhythm analysis systems by rhythmocardiosignals with increased resolution

An important step in the development of information systems of cardiac rhythm analysis is a substantiated choice of diagnostic features set, which will be used for automated procedure for diagnostic decision making. There are mainly two requirements for this set of diagnostic features. The first requirement is the informativeness requirement of many diagnostic features, and the second - the requirement of minimality of their number.

The first requirement regarding the informative nature of the diagnostic features is the ability to distinguish between different state of the system under study on these features. Such informativeness of diagnostic features is determined by two of their characteristics, that is, sensitivity of diagnostic features to change of a state of regulatory mechanisms of cardiovascular system and organism as a whole, and also insensitivity to various non-informative noise factors (interferences) which are always present in a rhythmocardiosignal. One of the possible quantitative informativeness indicators of diagnostic features is the ratio of the average distance between the diagnostic classes (training sets) and the average diameter of the corresponding classes that corresponds to different states of the cardiovascular system in the metric space of diagnostic features. If this ratio is significant, then the components of the diagnostic feature vector are considered informative.

The minimum number requirement of diagnostic features provides the minimum dimension of diagnostic features space, which, as a consequence, provides the minimum computational complexity of algorithms for diagnostic decision making.

Let’s substantiate the diagnostic features set to evaluate the state of regulatory mechanisms of the cardiovascular system and the organism as a whole, that is, such sets of diagnostic features, which, on the one hand, are informative, and on the other - have a minimum number. First, let's focus on the procedure for providing the minimum number of diagnostic features by rhythmocardiosignals with increased-resolution.

Since the hypothesis for normality of distribution of the rhythmocardiosignal with increased resolution was previously confirmed, as described above, the initial set of diagnostic features is a numeric vector $\hat{\mathbf{C}}_L^1 = \{\hat{c}_{1_{T_l}}, l = \overline{1,L}\}$ point estimates of mathematical expectations calculated according to expression (16) and an estimates matrix of correlation functions $\hat{\mathbf{R}}_T = [\hat{r}_{2_{T_{l_1}T_{l_2}}}(u), l_1, l_2 = \overline{1,L}]$, which were calculated according to the formula (19). One of the obvious ways to reduce the number of diagnostic features of the rhythmocardiosignal is to take into account the fact of symmetry ($\hat{r}_{2_{T_{l_1}T_{l_2}}}(u) = \hat{r}_{2_{T_{l_2}T_{l_1}}}(u),\ l_1, l_2 = \overline{1, L}$) estimates matrix of correlation functions $\hat{\mathbf{R}}_T = [\hat{r}_{2_{T_{l_1}T_{l_2}}}(u),\ l_1, l_2 = \overline{1, L}]$, indicating that it is sufficient to evaluate only those elements of the matrix $\hat{\mathbf{R}}_T$, what lie on its diagonal and above the diagonal, that is, such an ordered set $\hat{\mathbf{R}}_T = [\hat{r}_{2_{T_{l_1}T_{l_2}}}(u),\ l_1 = \overline{1, L},\ l_2 = \overline{l_1, L}]$. On the diagonal of this matrix, when $l_1 = l_2$, autocorrelation functions estimates are placed, and the elements of the matrix $\hat{\mathbf{R}}_T$, which are placed above its diagonal, that is, when $l_1 < l_2$, are estimates of cross-correlation functions.

Therefore, the matrix $\hat{\mathbf{R}}_T = [\hat{r}_{2_{T_{l_1}T_{l_2}}}(u),\ l_1, l_2 = \overline{1, L}]$ without losing the informativeness, we can replace with the triangular matrix $\hat{\mathbf{R}}_T = [\hat{r}_{2_{T_{l_1}T_{l_2}}}(u),\ l_1 = \overline{1, L},\ l_2 = \overline{l_1, L}]$.

Another way to reduce the number of diagnostic features in cardiac rhythm analysis information systems on the basis of rhythmocardiosignal with increased resolution is to use spectral decompositions of the triangular matrix elements themselves $\hat{\mathbf{R}}_T = [\hat{r}_{2_{T_{l_1}T_{l_2}}}(u),\ l_1 = \overline{1, L},\ l_2 = \overline{l_1, L}]$, in particular, by using a discrete Fourier transform of autocorrelation estimates and cross-correlation functions from this matrix. That is, instead of a triangular matrix $\hat{\mathbf{R}}_T = [\hat{r}_{2_{T_{l_1}T_{l_2}}}(u),\ l_1 = \overline{1, L},\ l_2 = \overline{l_1, L}]$ correlation functions can be used by a triangular matrix $\hat{\mathbf{S}}_T = [\hat{S}_{2_{T_{l_1}T_{l_2}}}(\nu),\ l_1 = \overline{1, L},\ l_2 = \overline{l_1, L}]$, elements of which are Fourier-images of the corresponding estimates of the correlation functions from the matrix $\hat{\mathbf{R}}_T$. That is, Fourier-images from the matrix $\hat{\mathbf{S}}_T$ are calculated as follows:

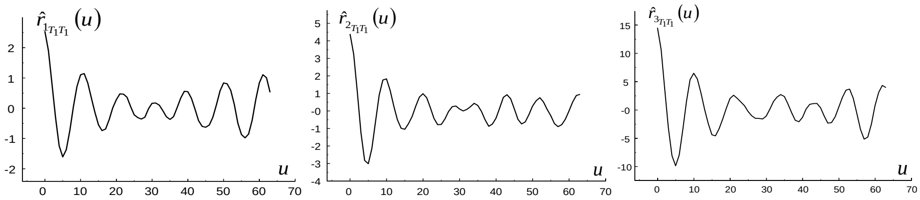
$$\hat{S}_{2_{T_{l_1}T_{l_2}}}(\nu) = \sum_{u=0}^{M_1-1} \hat{r}_{2_{T_{l_1}T_{l_2}}}(u) \cdot e^{\frac{-j2\pi u\nu}{M_1}},\ \nu = \overline{0, M_1 - 1},\ l_1 = \overline{1, L},\ l_2 = \overline{l_1, L},\ j = \sqrt{-1}\,. \qquad (24)$$

Based on the Bessel inequality, as diagnostic features we will not choose the whole set $\left\{\hat{S}_{2_{T_{l_1}T_{l_2}}}(\nu),\ \nu = \overline{0, M_1 - 1}\right\}$ samples of functions $\hat{S}_{2_{T_{l_1}T_{l_2}}}(\nu)$, but only a subset of their first M_2 ($M_2 << M_1$) samples $\left\{\hat{S}_{2_{T_{l_1}T_{l_2}}}(\nu),\ \nu = \overline{0, M_2 - 1}\right\}$, which contribute to the full energy of evaluation $\hat{r}_{2_{T_{l_1}T_{l_2}}}(u)$ correlation function is not less than 95%.

Here is an example of the statistical evaluation of vector elements $\mathbf{C}_L^1 = \{c_{1_{T_l}},\ l = \overline{1, L}\}$ mathematical expectations, elements of the matrix of correlation functions $\mathbf{R}_T = \left[r_{2_{T_1 T_2}}(u), l_1, l_2 = \overline{1, L} \right]$ and matrix elements of the Fourier-images $\hat{\mathbf{S}}_T = \left[\hat{S}_{2_{T_1 T_2}}(\nu), l_1 = \overline{1, L}, l_2 = \overline{l_1, L} \right]$ by one realization $\Xi_{3_{\omega'}}(m) = \{T_{l_{\omega'}}(m), l = \overline{1,3}, m = \overline{1,245}\}$ tricomponent vector $\Xi_3(\omega', m) = \{T_l(\omega', m), \omega' \in \mathbf{\Omega}', l = \overline{1,3}, m = \overline{1,245}\}$ stationary and stationary-related random sequences.

Figure 3 shows the plot of realization $\hat{r}_{2_{T_1 T_1}}(u)$ statistical estimation of autocorrelation function $r_{2_{T_1 T_1}}(u)$ ($l_1 = l_2 = 1$) of three vector components $\Xi_3(\omega', m)$.

Table 2 presents the statistical evaluation results of the mathematical expectations of the stationary components of the vector $\Xi_3(\omega', m)$.



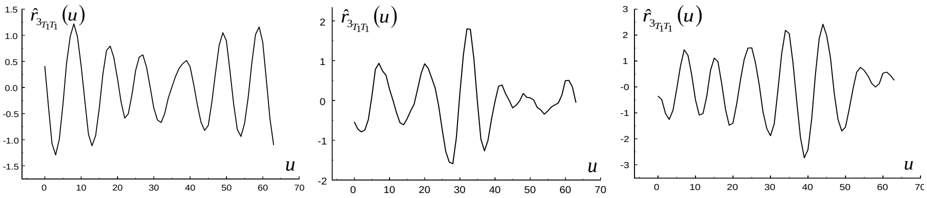
a) b) c)

Fig. 3. Plot of realization: a) $\hat{r}_{2_{T_1 T_1}}(u)$ statistical estimation of autocorrelation function $r_{2_{T_1 T_1}}(u)$ ($l_1 = l_2 = 1$), the first component $T_1(\omega', m)$, which describes the duration P - intervals in the electrocardiosignal; b) $r_{2_{T_2 T_2}}(u)$ statistical estimation of autocorrelation function $r_{2_{T_3 T_3}}(u)$ ($l_1 = l_2 = 2$) the second component $T_2(\omega', m)$, which describes the duration R - intervals in the electrocardiosignal; c) $\hat{r}_{2_{T_3 T_3}}(u)$ statistical estimation of autocorrelation function $r_{2_{T_3 T_3}}(u)$ ($l_1 = l_2 = 3$) third component $T_3(\omega', m)$, which describes the duration T - intervals in the electrocardiosignal

Table 2. The statistical evaluation results

Stationary component number	Statistical estimation realization value of mathematical expectation
1	$c_{1_{T_1}} = 14{,}88$
2	$c_{1_{T_2}} = 25{,}02$
3	$c_{1_{T_3}} = 73{,}82$

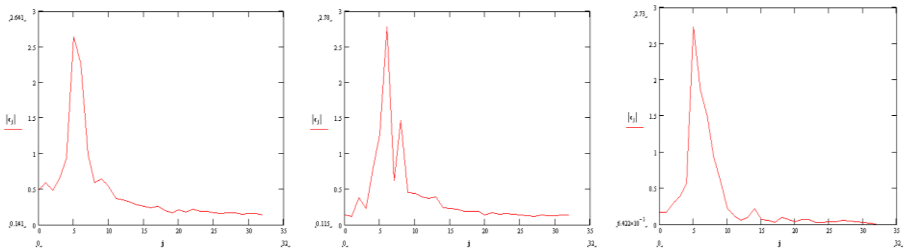
Figure 4 shows the graphs of realization of statistical estimates of the cross-correlation functions of the vector components $\Xi_3(\omega', m)$.



a) b) c)

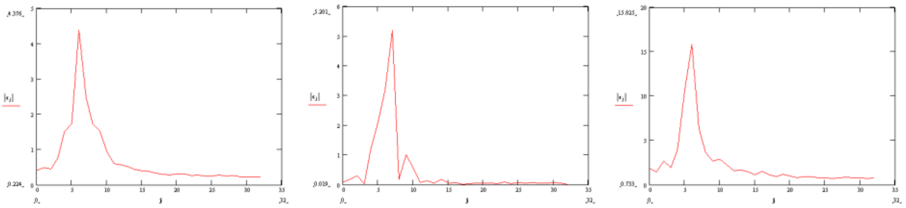
Fig. 4. Plot of realization: a) $\hat{r}_{2_{T_1T_2}}(u)$ statistical estimation of the cross-correlation function $r_{2_{T_1T_2}}(u)$ ($l_1 = 1, l_2 = 2$) first $T_1(\omega', m)$ and second $T_2(\omega', m)$ vector components $\Xi_3(\omega', m)$; b) $\hat{r}_{2_{T_1T_3}}(u)$ statistical estimation of the cross-correlation function $r_{2_{T_1T_3}}(u)$ ($l_1 = 1, l_2 = 3$) first $T_1(\omega', m)$ and third $T_3(\omega', m)$ vector components $\Xi_3(\omega', m)$; c) $\hat{r}_{2_{T_1T_3}}(u)$ statistical estimation of the cross-correlation function $r_{2_{T_1T_3}}(u)$ ($l_1 = 1, l_2 = 3$) first $T_1(\omega', m)$ and third $T_3(\omega', m)$ vector components $\Xi_3(\omega', m)$

Figure 5 shows graphs of realization of statistical estimates of the cross-correlation functions of the vector components $\Xi_3(\omega', m)$ Figure 6 shows realization plot $\hat{S}_{2_{T_1T_2}}(\nu)$ statistical estimation of cross-spectral power density $S_{2_{T_1T_2}}(\nu)$ ($l_1 = 1, l_2 = 2$) first $T_1(\omega', m)$ and second $T_2(\omega', m)$ vector components $\Xi_3(\omega', m)$.



a) b) c)

Fig. 5. Plot of realization: a) $\hat{S}_{2_{T_1T_2}}(\nu)$ statistical estimation of power spectral density $S_{2_{T_1T_2}}(\nu)$ ($l_1 = l_2 = 1$) the first component $T_1(\omega', m)$, which describes the duration P - intervals in the electrocardiosignal; b) $\hat{S}_{2_{T_2T_2}}(\nu)$ statistical estimation of power spectral density $S_{2_{T_2T_2}}(\nu)$ ($l_1 = l_2 = 2$) the second component $T_2(\omega', m)$, which describes the duration R - intervals in the electrocardiosignal; c) $\hat{S}_{2_{T_3T_3}}(\nu)$ statistical estimation of power spectral density $S_{2_{T_3T_3}}(\nu)$ ($l_1 = l_2 = 3$) third component $T_3(\omega', m)$, which describes the duration T - intervals in the electrocardiosignal



a) b) c)

Fig. 6. Plot of realization: a) $\hat{S}_{2_{T_1T_2}}(\nu)$ statistical estimation of cross-spectral power density $S_{2_{T_1T_2}}(\nu)$ ($l_1 = 1, l_2 = 2$) first $T_1(\omega', m)$ and second $T_2(\omega', m)$ vector components $\Xi_3(\omega', m)$; b) $\hat{S}_{2_{T_1T_3}}(\nu)$ statistical estimation of cross-spectral power density $S_{2_{T_1T_3}}(\nu)$ ($l_1 = 1, l_2 = 3$) first $T_1(\omega', m)$ and third $T_3(\omega', m)$ vector components $\Xi_3(\omega', m)$; c) $\hat{S}_{2_{T_2T_3}}(\nu)$ statistical estimation of cross-spectral power density $S_{2_{T_2T_3}}(\nu)$ ($l_1 = 2, l_2 = 3$) second $T_2(\omega', m)$ and third $T_3(\omega', m)$ vector components $\Xi_3(\omega', m)$.

6 Conclusions

The methods of statistical estimation of probabilistic characteristics of rhythmocardiosignal with increased resolution on the basis of model in the form of a vector of stationary and stationary related random sequences are developed in the paper. Conducted statistical experiments confirmed the hypothesis for the normality of the law of distribution of components of the vector rhythmocardiosignal. The conducted decomposition of statistical estimates of autocorrelation and cross-correlation functions made it possible to obtain spectral and cross-spectral power densities of the vector components, which allowed to reduce the dimension space of diagnostic features in systems of analysis of cardiac rhythm according to the rhythmocardiosignals with increased resolution.
The developed statistical methods can be used in developing of specialized software in automated cardio-diagnostic complexes, in particular, rhythm analysis subsystems.

References

1. Indu Saini, Dilbag Singh, Arun Khosla. (2013) QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases. Journal of Advanced Research, Volume 4, Issue 4. July 2013. -pp. 331-344. doi.org/10.1016/j.jare.2012.05.007
2. M.K. Bhaskar, S.S. Mehta, N.S. Lingayat (2013) Probabilistic Neural Network for the Automatic Detection of QRS-complexes in ECG using Slope. International Journal of Emerging Technology and Advance Engineering Volume 3, Issue 6, June 2013. –pp.255-261. ISSN 2250-2459. ISO 9001:2008.
3. M. Rahimpour, M. E. Asl and M. R. Merati (2016) ECG fiducial points extraction using QRS morphology and adaptive windowing for real-time ECG signal analysis, 2016 24th Iranian Conference on Electrical Engineering (ICEE), Shiraz. -pp. 1925-1930. doi: 10.1109/IranianCEE.2016.7585836.

4. Israa Shaker Tawfic, Sema Koc Kayhan. (2017) Improving recovery of ECG signal with deterministic guarantees using split signal for multiple supports of matching pursuit (SSMSMP) algorithm, Computer Methods and Programs in Biomedicine, vol. 139, 2017. – pp. 39-50. doi.org/10.1016/j.cmpb.2016.10.014
5. Ciucurel C., Georgescu L., Iconaru E.I. (2018), ECG response to submaximal exercise from the perspective of Golden Ratio harmonic rhythm, Biomedical Signal Processing and Control, 40. –pp. 156-162. doi.org/10.1016/j.bspc.2017.09.018.
6. Fumagalli F., Silver A.E., Tan Q., Zaidi N., Ristagno G. (2018), Cardiac rhythm analysis during ongoing cardiopulmonary resuscitation using the Analysis During Compressions with Fast Reconfirmation technology, Heart Rhythm, 15(2).-pp.248-255. doi: 10.1016/j.hrthm.2017.09.003.
7. Napoli N.J., Demas M.W., Mendu S., Stephens C.L., Kennedy K.D, Harrivel A.R, Bailey R.E., Barnes L.E. (2018), Uncertainty in heart rate complexity metrics caused by R-peak perturbations, Computers in Biology and Medicine, 103.-pp. 198-207. doi: 10.1016/j.compbiomed.2018.10.009.
8. F.Shaffer, J.P.Ginsberg (2017), An Overview of Heart Rate Variability Metrics and Norms, Frontiers in Public Health, Volume 5, Article 258, September 2017. – pp. 1-17, doi: 10.3389/fpubh.2017.00258.
9. S. Lupenko, N. Lutsyk, O. Yasniy and Ł. Sobaszek, "Statistical analysis of human heart with increased informativeness," Acta mechanica et automatica, vol. 12, 2018, pp. 311-315.
10. Serhii Lupenko, Nadiia Lutsyk, Oleh Yasniy, Andriy Zozulia The Modeling and Diagnostic Features in the Computer Systems of the Heart Rhythm Analysis with the Increased Informativeness. 2019 9th International Conference on Advanced Computer Information Technologies (ACIT). IEEE, 2019. pp. 121-124.