Selection Parameters in the ECG Signals for Analysis of QRS Complexes

Iurii Krak[1,3 \[0000-0002-8043-0785\]], Anatolii Pashko[1\[0000-0001-6944-8477\]],
Oleg Stelia[1\[0000-0002-1453-501X\]], Olexander Barmak[2\[0000-0003-0739-9678\]],
Sergey Pavlov[4\[0000-0002-6473-9627\]]

[1]Taras Shevchenko National University of Kyiv, Ukraine
[2]Khmelnytskyi National University, Khmelnytskyi, Ukraine
[3]Glushkov Cybernetics Institute, Kyiv, Ukraine
[4]Vinnytsia National Technical University, Ukraine
`yuri.krak@gmail.com`
`aap2011@ukr.net`
`oleg.stelya@gmail.com`
`barmakov@khnu.km.ua`
`psv@vntu.vinnica.ua`

Abstract. An approach to processing and analyzing ECG data based on high-precision determination of R peaks in QRS complexes and a statistical analysis of the QRS complex duration and ECG signal dispersion online is studied. A method is proposed for effectively finding R peaks in the ECG. Conducted statistical sequential analysis to study the behavior of R peaks, the significant difference of which is that the number of observations necessary to make a decision on the hypothesis depends on the test results and is a random variable. The method of successive testing of a hypothesis involves at each stage of monitoring the state of the heart rhythm making a decision on the presence or absence of violations. It turned out that a consistent assessment of the variance of the cardiogram in a healthy person has a pronounced linear character. With a poor heart rate, the areas where the heart works ambiguously are clearly expressed.

Keywords: QRS complex, features extraction, R peaks, signal analysis, approximation, sequential analysis

1 Introduction

In the online monitoring problems of the state of the cardiovascular system, a huge amount of data is obtained, which needs automated processing for making medical decisions and recommendations. The quality of these decisions and the effectiveness of subsequent recommendations depends on the processing efficiency of the resulting data array.

By heart rhythm disturbances, or arrhythmias, is meant any heart rhythm that is not a regular sinus rhythm of a normal frequency. Arrhythmias can be caused by a change in one or more functions of the heart (see surveys [1],[2]).

Various pathologies can lead to heart rhythm disturbances and behavioral changes. Depending on the heart rhythm disturbances, several types of arrhythmia can be distinguished: tachycardia (an increase in the heart rate of more than 90 beats per minute at state of rest), bradycardia (a slowing of the pulse of less than 60 beats per minute), extrasystole (a violation of cardiac activity, which consists in premature reduction myocardium or its individual parts), atrial fibrillation (chaotic atrial contraction).

The occurrence of arrhythmia is primarily associated with a violation of the heart rhythm. It's mean that the formation of a sequence of data that reflects the rhythmic structure of the heart is so important. As a rule, the QRS complex acts as data elements [3]. Algorithms based on the differentiated ECG are computationally efficient for real-time analysis of datasets. A frequently used mathematical tool in various algorithms for QRS complex selection are derivatives.

When processing the data of the QRS complex, it is necessary to take into account possible omissions of the QRS complex and formation of a false sequence element, called false alarm. Also an important characteristic is the accuracy of the binding of data sequence elements to the time position of the QRS complex.

2   Related works

An important problem of ECG analysis is the construction of methods and computer algorithms for detecting and correctly detecting the QRS complex. For these purposes, an effective method for detecting the QRS complex is widely used — the Pan-Tompkins algorithm [4]. As characteristic features for highlighting QRS complexes, the slope, amplitude and width of the ECG signal are used. It is important to distinguish other characteristic features from the ECG signal. So in the paper [5], as characteristic features for the successful detection of anomalies, it is proposed to use amplitude, duration, pre-gradient, post-gradient and so on. In the paper [6] dynamic features would mean extracting RR interval features, heart rate, HRV and the R/P ratio. Note that an ECG signal is the combination of various peaks, waves, valleys, segments, intervals, complexes and points. It should be noted that the data obtained from the ECG signal are non-stationary functions [7], and its derivatives are calculated by differentiating the interpolation functions. In many studies, piecewise polynomial curves are chosen as such functions. An analysis of various interpolation methods is contained in papers [8] and [9]. Various interpolation methods — linear interpolation, Lagrange interpolation, Hermits interpolation, and cubic spline interpolation — have been studied to find the optimal method for choosing the RR interval when analyzing heart rate variability. As a result of experiments, it was shown that the third-order Lagrange interpolation polynomial is the most suitable algorithm for selecting the RR interval for estimating the autoregressive spectrum, since it requires a short processing time and shows the lowest error rates in calculating the heart rate. Cubic splines are widely used to interpolate heart rate, although despite a large number of works on this topic, new

ones are constantly being created and existing mathematical algorithms are developed for processing and analyzing ECG data. In paper [10], a combination of wavelet transform with cubic interpolation by splines is used to increase the accuracy of detection of QRS complexes. Additional requirements for mathematical algorithms are presented in the case of online ECG processing [11]. The identification of the R-peak is crucial in the analysis of the signal of the ECG. In the study [12] proposed an adaptive and time-efficient algorithm for detecting R-peaks for ECG processing.

Thus, the problem of approximating ECG signals is important and relevant. In this study, a parametric piecewise polynomial continuous curve that has a continuous first derivative will be used to approximate ECG signals [13]. The advantages of this spline are the possibilities, using controlled parameters, of fitting its shape to the unsteady behavior of real ECG signals. Also, the proposed curve allows you to approximate the heart signal and calculate the first derivatives at arbitrary points. Of particular importance will be attached to the correct finding of R-peaks [14]. The article proposes methods of statistical analysis that allow recursively evaluating the characteristics of the ECG signal, which is especially important for real-time processing and verification and confirmation of some hypotheses about the distribution of R- peaks. The data for the study were taken from open access [15], as well as from [16], [17].

3 ECG signal spline approximation for QRS complex analysis

Let's introduce mathematical problem statement: will be to supose that the sequence of points be given on time interval $[a,b]$: $\Delta_\tau : a = \tau_1 < \tau_2 < ... < \tau_N = b$, and at points τ_i, the values of the cardiac signal F_i are given.

Along with the grid Δ_τ, a grid $\Delta_x : \tau_1 = x_1 < x_2 < ... < x_{N+1} = \tau_N$ is introduced, where

$$x_i = \eta_i \tau_{i-1} + (1 - \eta_i)\tau_i , \ 0 < \eta_i < 1 , \ i = \overline{3, N-1} , \ 0 < \eta_2 \le 1,$$

$$x_{i+1} = \gamma_i \tau_{i-1} + (1 - \gamma_i)\tau_i , \ 0 < \gamma_i < 1 , \ \eta_{i+1} = 1 - \gamma_i , \ i = \overline{2, N-1} , \ 0 < \gamma_{N-1} \le 1 .$$

For segments $[x_i, x_{i+1}], i = \overline{2, N-1}$ the interpolation polynomial is written as follows:

$$s_i(x) = f_{i+1} \frac{x - x_i}{x_{i+1} - x_i} + f_i \frac{x - x_{i+1}}{x_i - x_{i+1}} + (x - x_i)(x - x_{i+1})(Ax + B) , \tag{1}$$

where

$$A = \frac{f_i' + f_{i+1}'}{(h_i + h_{i+1})^2} - \frac{2(f_{i+1} - f_i)}{(h_i + h_{i+1})^3} ,$$

$$B = \frac{f_{i+1}' - f_i'}{2(h_i + h_{i+1})} - \frac{(f_{i+1}' + f_i')(x_i + x_{i+1})}{2(h_i + h_{i+1})^2} + \frac{(f_{i+1} - f_i)(x_i + x_{i+1})}{(h_i + h_{i+1})^3} ,$$

$$f_i = \eta_i F_{i-1} + (1 - \eta_i)F_i , \ f_{i+1} = \gamma_i F_{i+1} + (1 - \gamma_i)F_i , \ f_i' = \frac{F_i - F_{i-1}}{\tau_i - \tau_{i-1}} , \ f_{i+1}' = \frac{F_{i+1} - F_i}{\tau_{i+1} - \tau_i} ,$$

$$h_i = \eta_i(\tau_i - \tau_{i-1}) , \ h_{i+1} = \gamma_i(\tau_{i+1} - \tau_i) .$$

Consider the examples of using polynomials (1) to approximate ECG data and determine the R peak. In computational experiments, we used three data sets that did not undergo preliminary processing. In determining the position of the R peaks, the values of the first derivatives are used. The search for the peak was carried out among a group of points, the values of the first derivatives at which exceeded a certain threshold. In this case, the values of the derivatives were calculated not only at the starting points, but also at the intermediate ones. In the experiments presented, nine intermediate points were used. As shown by computational experiments, this approach is justified, since the values of the derivative exceeding a given threshold can be located between the nodes of the original mesh. In Fig.1., Fig.2., Fig.3., solid lines show graphs of cardiac signals [15] on which the detected peaks of R are indicated by dots. It $\eta_i = \gamma_i = 0.5$, $i = \overline{1, N-1}$ was assumed in the calculations.
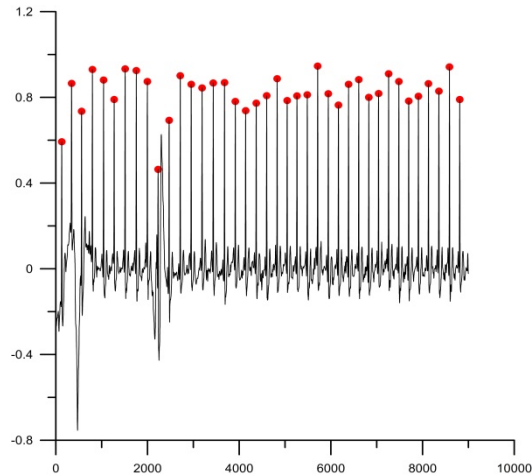
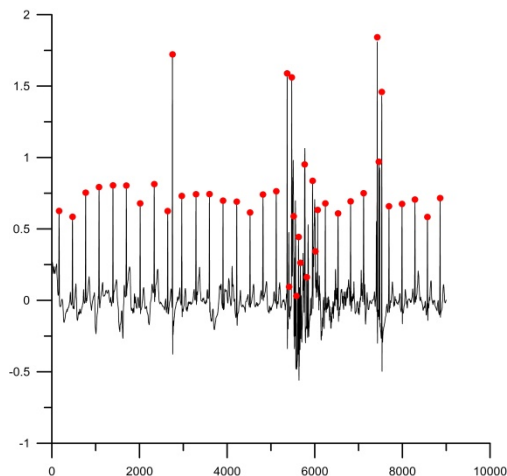**Fig.1.** Spline approximation of the ECG signal
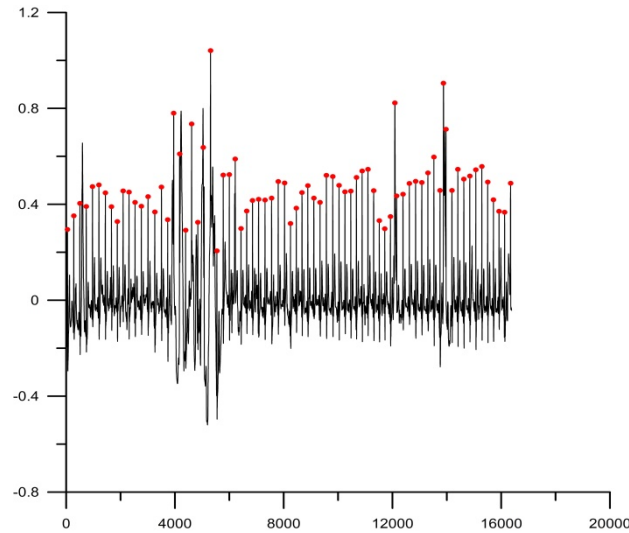
Fig.2. Spline approximation of the ECG signal

Fig.3. Spline approximation of the ECG signal

Note that it is the squares of the first derivatives that play a key role in the algorithm for detecting R peaks. After separating the intervals with the values of the square of the derivative exceeding the specified threshold, the maximum value on the interval and the corresponding abscissa value are extracted from the signal. To automate the peak extraction process, a certain threshold value is set.

4   Sequential analysis of R-R intervals

It is believed that the distribution of the lengths of the R-R intervals has a normal distribution for the cardiogram of a healthy person.

For various cardiac arrhythmias, it becomes necessary to test the hypothesis of equality of means for the normal distribution, as well as the hypothesis of equality of variances with an equal average.

In this paper, use the method of sequential analysis to solve these problems is proposed. A significant difference of this method is that the number n of observations $z_1, ..., z_n$ required to make a decision on the hypothesis depends on the outcome of the tests and is a random variable. The method of consecutively testing a hypothesis assumes at each stage of observation the adoption of one of the possible solutions: accept the hypothesis, reject it, or continue observing.

Before planning a sequential analysis procedure, acceptable values of the probabilities of permissible errors are assigned: α - the probability of accepting the hypothesis H_1, when the hypothesis is true H_0 (error of the first kind) and β - the probability of accepting the hypothesis H_0, when the hypothesis is true H_1 (error of the second kind).

Sequential analysis suggests that $\alpha << \beta$, or $\beta << \alpha$, when α, β are quantities of different order of smallness.

To test the null hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu = \mu_1$, $(\mu_0 < \mu_1)$ when the variances are equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$, they are calculated:

$$C = \frac{\sigma^2}{\mu_1 - \mu_0} \ln\left(\frac{\beta}{1-\alpha}\right) + n\frac{\mu_1 + \mu_0}{2}; \ D = \frac{\sigma^2}{\mu_1 - \mu_0} \ln\left(\frac{1-\beta}{\alpha}\right) + n\frac{\mu_1 + \mu_0}{2}.$$

If $\sum_{i=1}^{n} z_i \leq C$, then a hypothesis $H_0 : \mu = \mu_0$ is accepted.

If $\sum_{i=1}^{n} z_i \geq D$, then a hypothesis $H_1 : \mu = \mu_1$ is accepted.

If $C < \sum_{i=1}^{n} z_i < D$, then the observations continue.

To test hypothesis $H_0 : \sigma^2 = \sigma_0^2$ against alternative $H_1 : \sigma^2 = \sigma_1^2 (\sigma_1 > \sigma_0)$ with a known average μ, are calculated

$$C = \frac{2\sigma_0^2 \sigma_1^2}{\sigma_1^2 - \sigma_0^2} \ln\left[\frac{\beta}{1-\alpha}\left(\frac{\sigma_1}{\sigma_0}\right)^n\right]; \ D = \frac{2\sigma_0^2 \sigma_1^2}{\sigma_1^2 - \sigma_0^2} \ln\left[\frac{1-\beta}{\alpha}\left(\frac{\sigma_1}{\sigma_0}\right)^n\right].$$

If $\sum_{i=1}^{n} z_i^2 \leq C$, then a hypothesis $H_0 : \sigma^2 = \sigma_0^2$ is accepted.

If $\sum_{i=1}^{n} z_i^2 \geq D$, then a hypothesis $H_1 : \sigma^2 = \sigma_1^2$ is accepted.

If $C < \sum_{i=1}^{n} z_i^2 < D$, then the observations continue.

Note, that if the mean and variance values are unknown, their estimates are used.

During the online processing of measurement results, the sums $S1_n = \sum_{i=1}^{n} z_i$ and $S2_n = \sum_{i=1}^{n} z_i^2$ are calculated.

4.1 R-R intervals distribution for healthy person

To further study the distribution of R-R intervals, we consider the ECG of a healthy person [15], i.e. cardiogram where there is no heart rhythm disturbance. We studied a ECG lasting one minute, the data were taken in step $\Delta t = 0,003\,\text{sec}$ and contain $N = 21600$ points. The graph of the initial ECG is shown in Fig. 4. Using the method proposed in this study (see Section 3), we find the R peaks in the original cardiogram and calculate the lengths of the R-R intervals for this cardiogram (see Fig. 5). Distribution of lengths of R-R intervals is demonstrated on the Fig. 6.

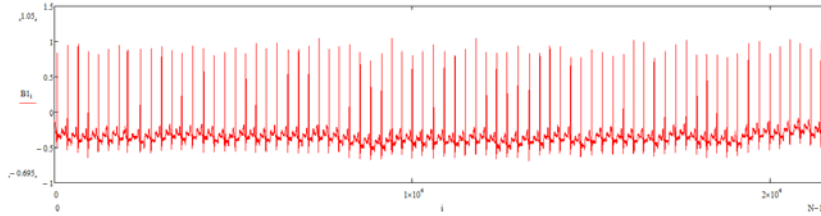Note, that investigated ECG contains $M = 74$ R-R intervals.
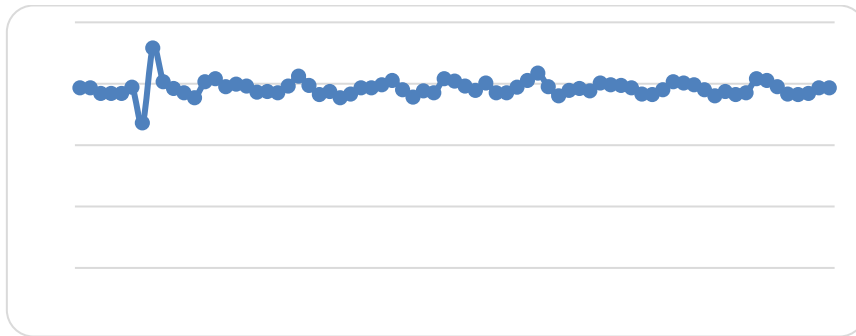
Fig.4. Initial ECG of a healthy person

Fig.5. Lengths of the R-R intervals of the initial ECG
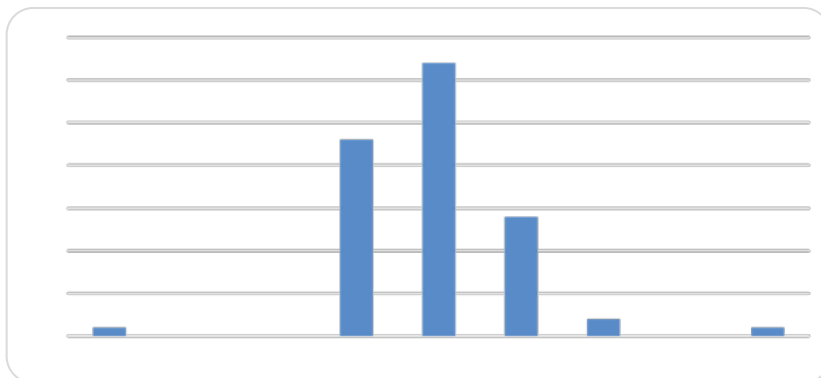

Fig.6. Histogram of distribution of lengths of R-R intervals of the initial ECG

The graph of the R-R intervals lengths distribution is given in Fig. 6 demonstrates that their distribution is close to the normal distribution. The average value is equal $\mu_0 = 292.42$, variance $\sigma^2 = 182.30$, coefficient of variation $kv = \dfrac{\sigma}{\mu_0} = 0.046$. At a confidential level $\varepsilon = 0.95$ we can conclude that the hypothesis that the distribution of the lengths of the R-R intervals has a normal distribution is confirmed.

4.2 Dispersion distribution of the ECG for healthy person

Next, we turn to the study of the dispersion of the cardiogram for a healthy person (see Fig. 3.). Consider the variance $\sigma^2(m,n)$ of the cardiogram as an observed parameter, the observation of variance of length m. The variance was estimated by the next formula

$$\sigma^2(n,m) = \frac{1}{\min(n-1,m-1)}\left(\sum_{k=\max(0,n-m)}^{n} z_k^2 - \frac{\left(\sum_{k=\max(0,n-m)}^{n} z_k\right)^2}{\min(m,n)}\right), \quad n = 0,1...N-1. \quad (2)$$

The length of the observation segment may change. we investigate in the given paper two segments: segment of the length $m = 400$ (one cycle of the heart rate) and $m = 1200$ (three cycles of the heart rate) are considered. On the Fig. 7, and Fig.8. the changes in the variance of the cardiogram are shown.
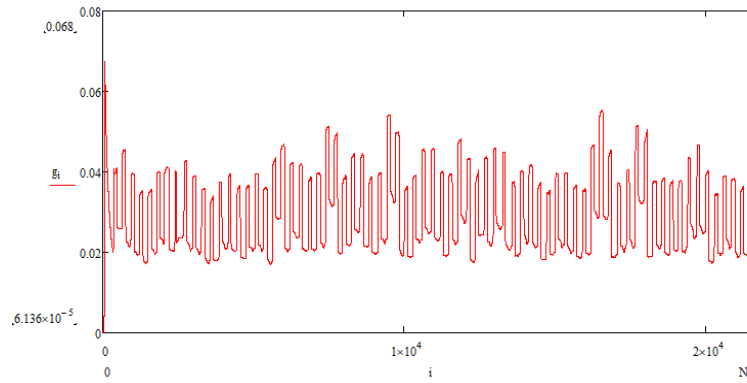

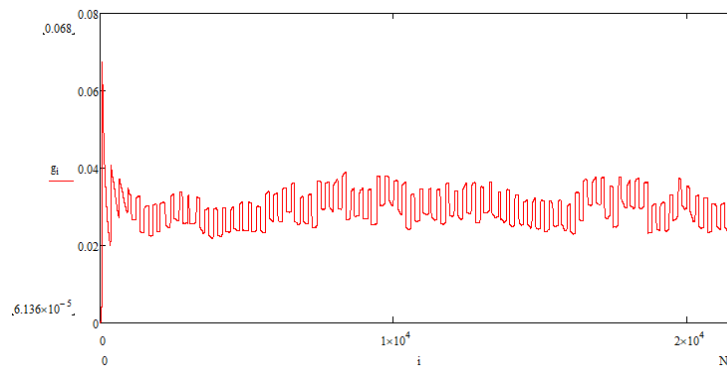**Fig.7.** Dispersion changes for one cycle of the heart rate for health person.

Fig.8. Dispersion changes for three cycles of the heart rate for health person.

5 ECG analysis for person with changes in heart rate

Let's consider a ECG for person with changes in heart rate [15] with the aim of investigate distribution of lengths of R-R intervals and dispersion changes base on approach given Section 4. ECG lasting 4.5 minutes, the data was taken in steps $\Delta t = 0,0025\,\mathrm{sec}$ and contains $N = 108000$ points. A graph of a given cardiogram is presented on the Fig. 9.

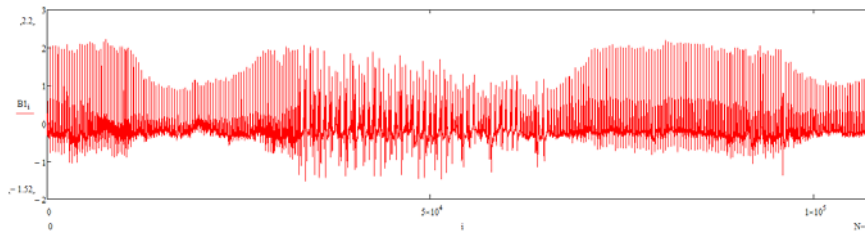In fig. 10 and fig. 11 is the variance estimate for $m = 400$ and $m = 1200$, respectively.

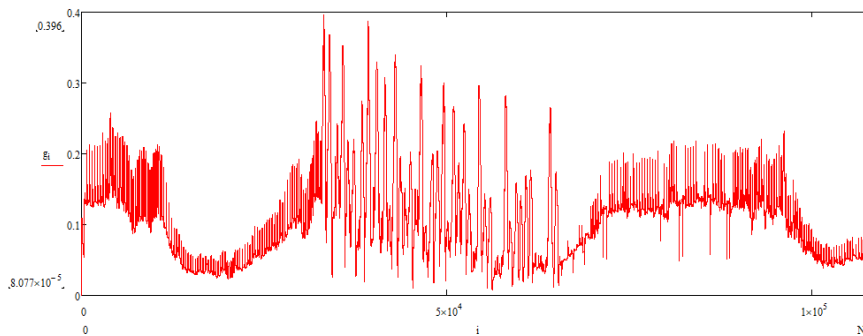**Fig.9.** Initial ECG of a person with changes in heart rate.

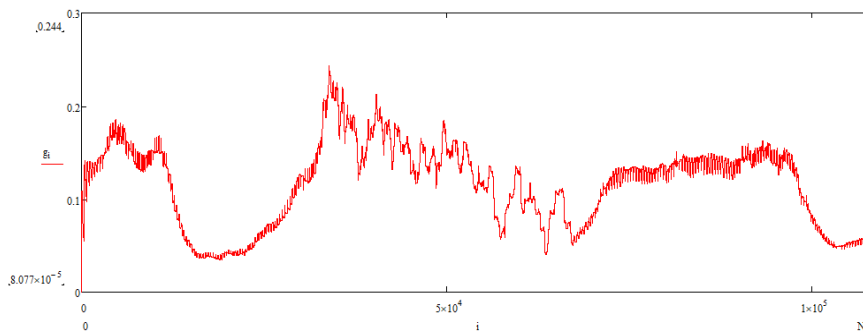**Fig.10.** Dispersion changes for one cycle for person with changes in heart rate.

Fig.11. Dispersion changes for three one cycle for person with changes in heart rate.

The variance changes shown in the graphs show at least three different sections of the heart rhythm. The ECG elements of these areas are shown in Fig. 12, Fig. 13., Fig.14 respectively.
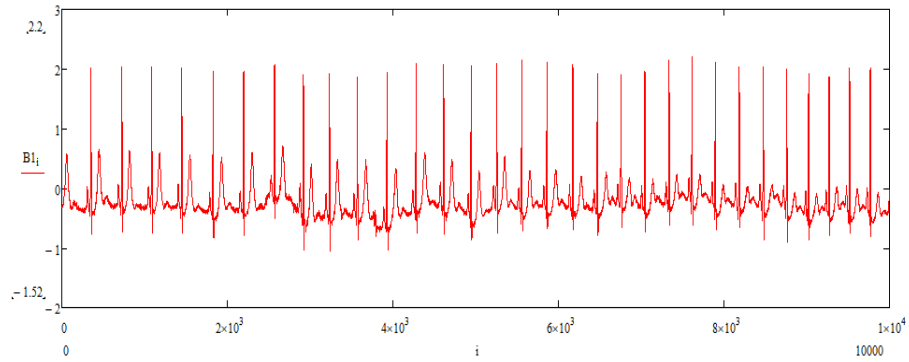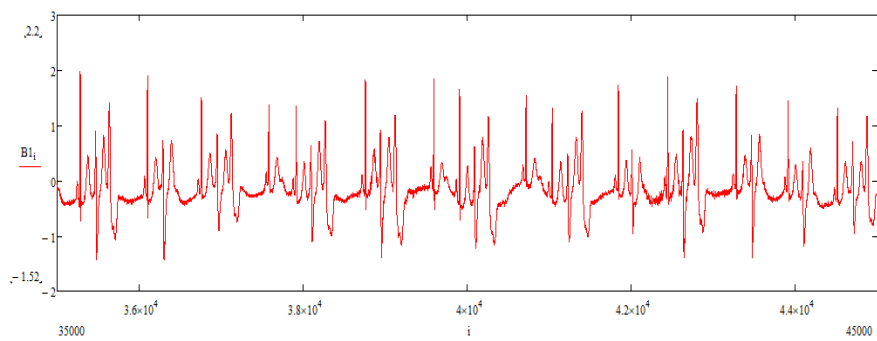

Fig.12. ECG segment for 1-25 sec.

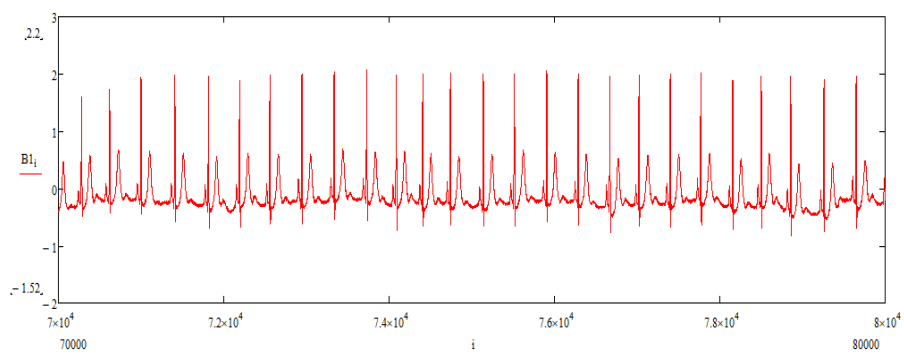**Fig.13.** ECG segment for 88-113 sec.

Fig.14. ECG segment for 175-200 sec.

As can be seen from the graphs, the hypothesis of inequality of the average variance in these areas is confirmed.

Consider, as an obsession, the lengths of R-R intervals. For the entire ECG, using spline analysis, R-R intervals were isolated and their lengths were calculated. An estimate of the density of the distribution of the length of R-R intervals is constructed.

On the Fig.15. a graph of the lengths of R-R intervals is presented, and a graph of histogram of the distribution of lengths of R-R intervals is presented in the Fig.16.
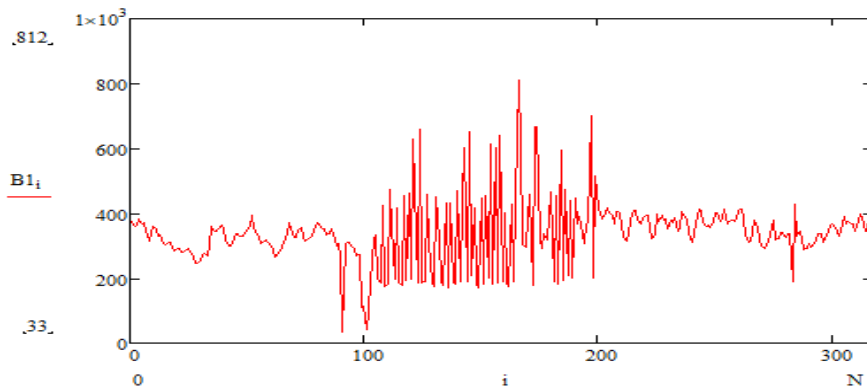

Fig.15. Lengths of R-R intervals for person with changes in heart rate.

It can be seen that the lengths of the R-R intervals have pronounced three sections of different behaviour.

The average value of the R-R intervals is 338.9, the variance is 10239.8, and the standard deviation is 101.2.

From the graphs of the R-R intervals it can be seen that in the first segment (points 1 ... 90):
- the average value of R-R intervals is 323.8,
- the variance is 1142.8,
- coefficient of variation 0.10;

on the second segment (points 91 ... 203):
- the average value of R-R intervals is 329.2,
- the variance is 25957.2,
- the coefficient of variation is 0.49;

on the third segment (points 204 ... 318):
- the average value of R-R intervals is 358.9,
- the variance is 1363.5,
- coefficient of variation 0.10.

Thus, the hypothesis was confirmed that the average is equal in these areas and that the variance is different.

The heart rate variability in the first segment corresponded to the norm, in the second and third heart rate variability has features.

What caused these features is unknown, a person could run, and then stop, or perform some kind of action.
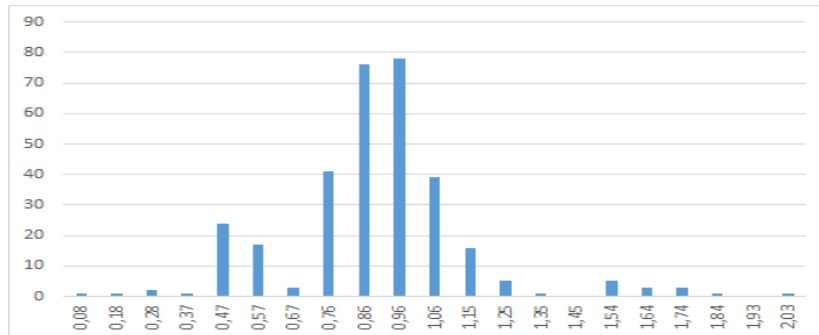
Fig.16. Histogram of distribution of lengths of R-R intervals of the ECG for person with changes in heart rate.

The use of these algorithms allows you to select areas of changes in heart rate online. Depending on the state of the person for whom these measurements are carried out, the results can be used to make decisions about the state of the person.

6 Conclusion

The article proposes an approach to the processing and analysis of data in cardiology, based on high-precision determination of R peaks in QRS complexes and a statistical analysis of the QRS complex duration and ECG signal dispersion online.

In this investigation proposes the use of statistical sequential analysis for these problems, the significant difference of which is that the number of observations necessary to make a decision on the hypothesis depends on the outcome of the tests and is a random variable. The method of successive testing of a hypothesis involves at each stage of monitoring the state of the heart rhythm making a decision on the presence or absence of violations.

It was found that a consistent assessment of the variance of the cardiogram of a healthy person has a pronounced linear character. With a bad heart rate, areas where the heart works ambiguously are clearly expressed.

The use of these algorithms allows you to select areas of changes in heart rate online. The results can be used to make decisions about the condition of a person whose heart rate is being monitored. It is unknown what caused these features, a person could run away, and then stop, or perform some kind of action, or he has heart problems.

References

1. Bhoi, A. K., Sherpa, K.: QRS Complex Detection and Analysis of Cardiovascular Abnormalities: A Review. Int. J. BIOautomation 18(3): 181-194 (2014)
2. Luz, E., Schwartz, W., Cámara-Chávez, G., Menotti, D.: ECG-based heartbeat classification for arrhythmia detection: A survey, Computer Methods and Programs in Biomedicine, Vol. 127, 144-164 (2016) https://doi.org/10.1016/j.cmpb.2015.12.008

3. Kaplan Berkaya, S., Uysal, A., Sora Gunal, E., Ergin, S., Gunal, S., Gulmezoglu, M.: A survey on ECG analysis, Biomedical Signal Processing and Control, Vol. 43, 216-235, (2018) https://doi.org/10.1016/j.bspc.2018.03.003

4. Pan J, Tompkins, W.: A Real-Time QRS Detection Algorithm, IEEE Transactions on Biomedical Engineering, Vol. BME-32(3) 230-236 (1985) DOI: 10.1109/TBME.1985.325532

5. Soorma, N, Jaikaran Singh, J., Tiwari, M.: Feature Extraction of ECG Signal Using HHT Algorithm, International Journal of Engineering Trends and Technology V8(8): 454-460 (2014). doi: 10.14445/22315381/IJETT-V8P278

6. Mayapur, P.: Classification of Arrhythmia from ECG Signals using MATLAB, International Journal of Engineering and Management Research 8(6): 115-129 (2018) DOI: doi.org/10.31033/ijemr.8.6.11

7. Kaplan Berkaya, S., Uysal, A., Sora Gunal, E., Ergin, S., Gunal, S., Gulmezoglu, M.: A survey on ECG analysis, Biomedical Signal Processing and Control, Vol. 43, 216-235, (2018) https://doi.org/10.1016/j.bspc.2018.03.003

8. Jang, D., Hahn, M., Jang, J., Farooq U., Park, S.:A comparison of interpolation techniques for RR interval fitting in AR spectrum estimation, IEEE Biomedical Circuits and Systems Conference (BioCAS), Hsinchu, pp. 352-355 (2012) DOI: 10.1109/BioCAS.2012.6418424

9. Ellis, R., Zhu, B., Koenig, J., Thayer, J., Wang, Y.: A careful look at ECG sampling frequency and R-peak interpolation on short-term measures of heart rate variability, Physiol. Meas. 36, 1827-1852 (2015) doi:10.1088/0967-3334/36/9/1827

10. Rodrigues, L., Marengoni, M.: Detecting QRS complex in ECG using wavelets and cubic spline interpolation, Biomedical Engineering (BioMed 2012), February 15 – 17, 2012, Innsbruck, Austria, 6 p. (2012) DOI: 10.2316/P.2012.764-135

11. Sahoo,S., Kanungo, B., Behera, S., Sabut, S.: Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities, Measurement, Vol. 108, 55-66 (2017). https://doi.org/10.1016/j.measurement.2017.05.022

12. Qin, Q., Li, J., Yue, Y., Liu, C.: An Adaptive and Time-Efficient ECG R-Peak Detection Algorithm, Journal of Healthcare Engineering Article ID5980541, 14pages (2017) https://doi.org/10.1155/2017/5980541

13. Stelia, O., Krak, I., Potapenko, L.: Controlled Spline of Third Degree: Approximation Properties and Practical Application. In: Lytvynenko V., Babichev S., Wójcik W., Vyno-kurova O., Vyshemyrskaya S., Radetskaya S. (eds) Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2019. Advances in Intelligent Systems and Computing, vol 1020. Springer, Cham, pp.215-224 (2020)

14. Krak, I., Pashko, A., Khorozov, O., Stelia, O.: Physiological Signals Analysis, Recognition and Classification Using Machine Learning Algorithms. In: Sergey Subbotin ed. Computer Modeling and Intelligent Systems. CEUR-WS.org. Vol-2608. pp.966-965 (2020)

15. Physionet ECG data. https://github.com/mathworks/physionet_ECG_data/ [Accessed: 2020-05-05]

16. Khorozov, O., Krak, I., Kulias, A., Kasianiuk, V., Wojcik, W., Tergeusizova, A.: Monitoring vital signs using fuzzy logic rules. In: Information technology in Medical Diagnostics II - - Proceedings of the International Scientific Internet Conference on Computer Graphics and Image Processing and 48th International Scientific and Practical Conference on Application of Lasers in Medicine and Biology, 2018 CRC Press/Balkema. pp. 237-244 (2019)

17. Tan, R., Perkowski, M.: Toward Improving Electrocardiogram (ECG) Biometric Verification using Mobile Sensors: A Two-Stage Classifier Approach. Published online 2017 Feb 20. doi: 10.3390/s17020410 (2017)