

New Geometric Method of Heart Rate Variability Estimation based on the Multiscale Correlation Analysis Representation

V.E. Antsiperov

Kotelnikov Institute of Radioengineering and Electronics of Russian Academy of Sciences,
Mokhovaya 11–7, Moscow, Russia
Contact: antsiperov@cplire.ru

Introduction

The nature and characteristics of ECG signals have been the subject of research for more than 100 years. This is because ECG signals contain a valuable diagnostic information about the functioning of the heart and, indirectly, about some other organs. Now, despite the impressive track record of the diagnostics and recommendations already developed [1], ECG signals continue to be the subject of intensive study both in the field of cardiology and biomedical engineering. Cardiologists search for new diagnostic properties of ECG signals, while engineers are interested in new approaches and methods for such signals processing and analysing. New methods imply, in particular, new techniques of noise suppression, efficient signal representation and feature extraction.

Feature extraction is an essential step in ECG processing. It consists in formation of some feature patterns – a set of descriptors most adequately describing the signal. Among these features / descriptors the most important are those with diagnostic properties. As a rule, they represent time domain parameters, though sometimes they can be defined in the frequency domain. Basing on ECG patterns, one can make some initial diagnosis [1]. For example, an “irregular” QRS complex without P wave is the hallmark of atrial fibrillation, the shape of the QRS with a left bundle branch block indicates a risk of cardiomyopathy, etc.

ECG patterns, especially P–QRS–T complexes, are also the basis for the subsequent heart rate analysis. Usually it is a more complex phase of ECG signal processing. However, now there are many methods and approaches related, for example, to the heart rate variability (HRV) analysis [2]. Quantitative evaluation of the HRV is usually based on the calculation of some indices connected with the variability of sequential NN intervals (normal-to-normal intervals between adjacent QRS complexes). Some of these indices are determined in the time domain – SDNN, CV, RMSSD, PNN50, etc. Others – in the frequency domain – VLF, LF, HF spectral power components, IC index, etc. It should be noted that, due to the complexity of the ECG signal, its fractal-like nature, the values of almost all indices can essentially diverge for different lengths of time intervals analysed, even for the same record. Therefore, the European Society of Cardiology suggests differentiating the diagnostic methodologies based on HRV indices obtained from the short-term recordings (of 2 to 5 min) and from the long-term recordings (of entire 24-h period) [3].

The above discussion is illustrated in Figure 1 by cardioin-

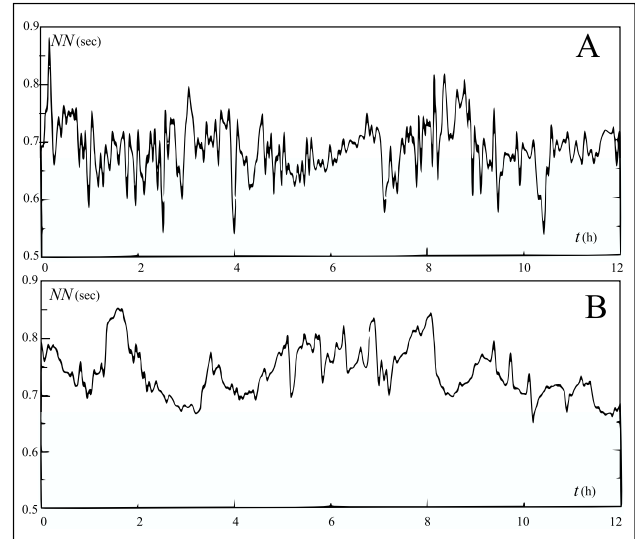


Figure 1: The interbeat interval NN after low-pass filtering for (A) a healthy subject and (B) a patient with severe cardiac disease (dilated cardiomyopathy). The healthy CIG shows more complex fluctuations compared to the more “smoother” diseased heart rate fluctuation pattern. Drawing is adapted from [4].

tervalograms (CIG, the NN interval dependences on time) for a healthy person (A) and a heart disease person (B). It is clear from the figure that both CIGs are qualitatively different. For example, at a small temporal scale, the variability (spread) of NN intervals in a healthy person is greater, which is a well known fact [5] – but on large scales the variability becomes equal in both cases. This observation suggests that the diagnosis based on HRV is adequate, but diagnostic procedures (automatic in particular) may be non-trivial. This circumstance gives high relevance to the development of new ECG processing methods based on long records – more than 10 minutes, for example, Holter.

Geometric Method of HRV Estimation

As it is mentioned above, variations in heart rate may be evaluated by a number of indices, which can be calculated by a number of methods. Perhaps the most popular and at the same time the simplest to perform are the so-called geometric methods [2], [3].

Geometric Methods

Geometrical methods include, as a rule, some geometric pattern formed by the sequence of NN intervals, such as the sample density distribution, sample density distribution of differences between adjacent intervals, Lorenz plot of NN or RR intervals, etc., and a simple formula, which judges the variability based on the geometric and/or graphic properties of the formed pattern. A number of popular geometric methods can be found in [3], [6].

The main advantage of geometric methods is their relative insensitivity to the accuracy of NN intervals measuring. The main drawback is the need for a large amount of NN intervals for constructing a geometric pattern. In practice, it takes at least 20 minutes (but preferably 24 hours) to ensure correct operation of geometric methods, that is, the measurements of not less than ~ 1500 NN intervals are necessary. In addition to the fact that this is hard routine work (which, however, with certain precautions can be assigned to a computer), it makes known geometric methods inappropriate to assess short-term changes in HRV.

MCA based Method

In order to avoid the limitations of classical geometric methods (but to inherit their attractive properties), we developed a new estimating HRV method based on the multiscale correlation analysis (MCA) representations [7]. The idea of this method springs in connection with the study of statistical characteristics of point processes and was first announced in [8].

Let's note from the beginning, that the main instrument of multiscale correlation analysis – the MCA representation (pattern) – is formed not on the basis of NN intervals, as in the classical methods, but on the basis of some special estimation of autocorrelation function (ACF) calculated directly by the ECG record. The basic definitions of MCA representations for analyzing the rhythm of the ECG in the general form can be found in [7].

Further, let's accept for ECG the special model $z(t')$ of a pulse process:

$$z(t') = \sum_{k=-\infty}^{\infty} A_k g(t' - t'_k) \quad (1)$$

where sequence $-\infty < \dots < t_0 < t < t_1 < \dots < \infty$ is the random time moments of pulse occurrences, specially numbered with respect to some fixed time t (the point of the current analysis), $\dots, A_0, A_0, \dots$ are the random amplitudes of the pulses, which we assume to be a stationary sequence of independent random variables with mean $\bar{A}$, and $g(t')$ is a non-random waveform, like the P-QRS-T complex, the same for all pulses. The statistics of the sequence $\{t_k\}$ is given by some point process (PP), whose model is specified below.

Within the model (1) framework, the general definition of MCA representation [7] takes the following special form:

$$\hat{R}(t, \tau) = \frac{1}{\tau} \sum_{k=-\infty}^0 \sum_{k'=1}^{\infty} A_k A_{k'} \int_{-\infty}^{\infty} |G(f)|^2 \times \exp\{2\pi i f \tau\} \exp\{2\pi i f(t_k - t_{k'})\} df \quad (2)$$

where τ is the scale parameter of MCA representation and $G(f)$ is the Fourier spectrum of $g(t')$. The equation (2) expresses the MCA representation in terms of random variables $\{t_k\}$, $\{A_k\}$ and thereby fully determines the $\hat{R}(t, \tau)$ as a random value. However, even with the simplifications made (independence of the amplitudes and pulse intervals, stationarity, etc.), it is completely unrealistic to obtain a complete statistical description of $\hat{R}(t, \tau)$ in the form, for example, of its probability distribution. Therefore, it is necessary to confine the description of MCA representation to some simple characteristics, namely, to mathematical expectation of MCA $\langle \hat{R}(t, \tau) \rangle$, without taking into account possible random fluctuations around it. The corresponding formula for the mean MCA representation can be obtained by averaging (2) (over all random variables and under all assumptions made) and has the form announced in [8]:

$$\langle \hat{R}(t, \tau) \rangle = \frac{\bar{A}^2}{\tau} \int_{-\infty}^{\infty} |G(f)|^2 \exp\{2\pi i f \tau\} \times \left(\sum_{k=-\infty}^0 \sum_{k'=1}^{\infty} \langle \exp\{2\pi i f(t_k - t_{k'})\} \rangle \right) df \quad (3)$$

Model Specification

As follows from (3), in order to obtain the final, suitable for analysis mean MCA representation, it is necessary to know the averages $\chi_{k'k}(f) = \langle \exp\{-2\pi i f s_{k'k}\} \rangle$, where $s_{k'k} = (t_{k'} - t_k)$ are the intervals between ECG pulses (P-QRS-T complexes) occurring in $t_{k'}$ and t_k . Let us note in this connection that $\chi_{k'k}(f)$ also represents the characteristic function of $s_{k'k}$, which, in turn, is the sum $s_{k'k} = s_k + \dots + s_{k'-1}$, where $s_j = (t_{j+1} - t_j)$ are the intervals between consecutive time moments. So, recalling that the characteristic function of the sum of independent random variables is the product of the characteristic functions of the terms, we arrive at the idea of choosing the cyclic renewal process – point process with independent intervals [9] – as associated with (1) PP.

As the concept of cyclic renewal process does not have widespread usage let's give here its exact definition with reference to [9]. Namely, consider a stochastic process $y(t')$ that cycles in some time moments $0 = t_0 < \dots < t_j < \dots < \infty$ through states $0 \rightarrow 1 \rightarrow \dots \rightarrow K-1 \rightarrow 0 \rightarrow \dots$ in that order, again and again. The time intervals $s_j = (t_{j+1} - t_j)$ with which the process remains in the states j are called sojourn times of these states. If all these sojourn times are independent, distributed with their own probability densities $\rho_j(s)$, then $y(t')$ is called the cyclic renewal process. It is clear that the process obtained by decimating $y(t')$ in such a way that the result contains only the moments of the

state 0 arrivals, then the resulting process will be the usual renewal process.

If we re-index the process time moments $\{t_j\}$ with respect to the current analysis time moment t in the same way as in the definition (1), then, it can be shown, that when t unlim- itedly increases, any set of intervals $s_k, \dots, s_{k'-1}$ becomes stationary, it is a well-known fact in the theory of (usual) re- newal processes. Moreover, in this asymptotically station- ary case the probability distribution density of the intervals is greatly simplified and its explicit expression becomes the following:

$$\rho(s_k, \dots, s_{k'-1}) = \sum_{l=0}^{K-1} P_l \prod_{j=k}^{k'-1} \rho_{(l+j) \bmod(K)}(s_j) \\ k \leq 0 < k', \quad P_l = \bar{s}_l / \bar{S}, \quad \bar{S} = \bar{s}_0 + \dots + \bar{s}_{K-1} \quad (4)$$

where $\bar{s}_l$ are the mean values of sojourn time intervals in the states l , i.e. mathematical expectations of $\rho_l(s)$. In this context $\bar{S}$ is the duration of the mean overall cycle or the mean period of the cyclic renewal process considered.

It follows from (4) that stationary joint distribution of in- tervals $s_k, \dots, s_{k'-1}$ does not decompose into a product of individual interval distributions. So, formally, stationary cyclic process ceases to be a renewal process (except for the case $K = 1$). At the same time, since (4) can be considered as a formula of total probability for alterna- tives $l = 0, 1, \dots, K-1$, conditional (provided that s_0 is dis- tributed in accordance with the distribution $\rho_l(s)$ of the l 'th alternative) distributions of intervals will continue to be represented by products. Thus, a process considered would be more correctly treated as a conditionally renewal cycli- cal process. Incidentally, we note that the probabilities of the alternatives P_l are determined by the relative sojourn times in the consequent states l .

Having a complete statistical description (4) of any set of intervals $s_k, \dots, s_{k'-1}$, it is possible, in the frames of the model accepted, to find all necessary for calculating (3) characteristic functions $\chi_{k'k}(f) = \langle \exp\{-2\pi i f s_{k'k}\} \rangle$:

$$\chi_{k'k}(f) = \langle \exp\{-2\pi i f \sum_{j=k}^{k'-1} s_j\} \rangle = \\ = \sum_{l=0}^{K-1} P_l \prod_{j=k}^{k'-1} \chi_{(l+j) \bmod(K)}(f) \quad (5)$$

where $\chi_j(f) = \langle \exp\{-2\pi i f s_j\} \rangle$ are the characteristic func- tions of sojourn time interval distributions in the states j – distribution densities $\rho_j(s)$.

For one important case, which is implied in the cardiac NN intervals modeling, expression (5) can be essentially simplified. This is the case of a narrow enough densities $\rho_j(s)$, where the narrowness is understood as the small $Q_j = \sigma_j / \bar{s}_j$ values – standard deviation to the mean value ratios. Since in this case $\chi_j(f) = \langle \exp\{-2\pi i f s_j\} \rangle$ are well approximated by Gaussian characteristic functions

$$\chi_j(f) \simeq \exp\{-2\pi i f \bar{s}_j\} \exp\{-\frac{1}{2} (2\pi f \sigma_j)^2\},$$

then formula (5) can be approximated by the expression:

$$\chi_{k'k}(f) = \\ = \sum_{l=0}^{K-1} P_l \exp\{-2\pi i f \bar{s}_{l+k}^{l+k'-1}\} \exp\{-\frac{1}{2} (2\pi f \Sigma_{l+k}^{l+k'-1})^2\} \quad (6)$$

where:

$$\bar{s}_{l+k}^{l+k'-1} = \sum_{j=k}^{k'-1} \bar{s}_{(l+j) \bmod(K)} \\ \Sigma_{l+k}^{l+k'-1} = \sqrt{\sum_{j=k}^{k'-1} \sigma_{(l+j) \bmod(K)}^2} \quad (7)$$

Geometric Pattern of mean MCA representation

To obtain the main instrument of the geometric method based on MCA, it only remains for us to substitute ex- pression (6) in mean MCA representation (3), calculate the sums of internal series and take the corresponding integral. The series in (3) can be calculated exactly:

$$\sum_{k=-\infty}^0 \sum_{k'=1}^{\infty} \langle \exp\{2\pi i f (t_k - t_{k'})\} \rangle = \sum_{n=1}^{\infty} H_n(f) \quad (8)$$

where:

$$H_n(f) = \frac{1}{\bar{S}} \sum_{l=0}^{K-1} \bar{s}_l^{l+n-1} \times \\ \times \exp\{-2\pi i f \bar{s}_l^{l+n-1}\} \exp\{-\frac{1}{2} (2\pi f \Sigma_l^{l+n-1})^2\} \quad (9)$$

The integral in (3) can not be calculated exactly, since we do not know the exact form of the spectrum $|G(f)|^2$. Nev- ertheless, it can be found approximately if we take into account the characteristic properties of ECG signals, that are prototypes of model (1) process. As HRV manifests itself in a noticeable change of NN intervals duration, the width of the distributions $\rho_l(s)$ is greater than the duration of the characteristic fragments of P–QRS–T complex. But this implies that in the frequency domain, on the contrary, $H_n(f)$ is much narrower than the spectrum $|G(f)|^2$. There- fore, taking $|G(f)|^2$ from the integral (3) as the constant $|G(0)|^2$ and calculating the remaining Gaussian integrals, we finally obtain:

$$\langle \hat{R}(t, \tau) \rangle = \frac{\bar{A}^2 |G(0)|^2}{\bar{S}} \sum_{n=1}^{\infty} \sum_{l=0}^{K-1} \frac{\exp\left\{-\frac{(\tau - \bar{s}_l^{l+n-1})^2}{2(\Sigma_l^{l+n-1})^2}\right\}}{\sqrt{2\pi} (\Sigma_l^{l+n-1})^2} \quad (10)$$

With regard to formula (10), let us make an almost obvious observation – although in the left-hand site of the MCA representation denotation there remains a formal depen- dence on t , it is absent on the right-hand side. This is a natural consequence of the assumption of model stationar- ity.

Illustrations

In order to illustrate the important characteristics of a synthesized instrument that can be used in ECG diagnostics, we calculated the mean MCA representation $\langle \hat{R}(t, \tau) \rangle$ (10) for the model case in which the sojourn time intervals are given by a simple sinusoidal formula:

$$\bar{s}_l = \frac{\bar{S}}{K} (1 + \delta \sin(2\pi l/K)), \quad 0 \leq l < K \quad (11)$$

where the value $\bar{S}/K$ determines the average level of all sojourn time intervals $\bar{s}_l$, and the dimensionless "modulation index" δ determines the degree of their variability. Figure 2 shows two graphs of $\langle \hat{R}(t, \tau) \rangle$ (10) – one (A) for the considerable variability index $\delta = 0.3$ and the other (B) for its small value $\delta = 0.03$. It is not difficult to see that in the first case (A) the noticeable "beats" of the $\langle \hat{R}(t, \tau) \rangle$ oscillations are clearly manifested, but in the second case (B) the dependence on τ of the mean representation envelope is practically monotonic. Thus, extracting the envelope of the mean representation and estimating its oscillations, one can detect the presence of rhythm variability, and even find the related to it parameters, for example, the process cycle.

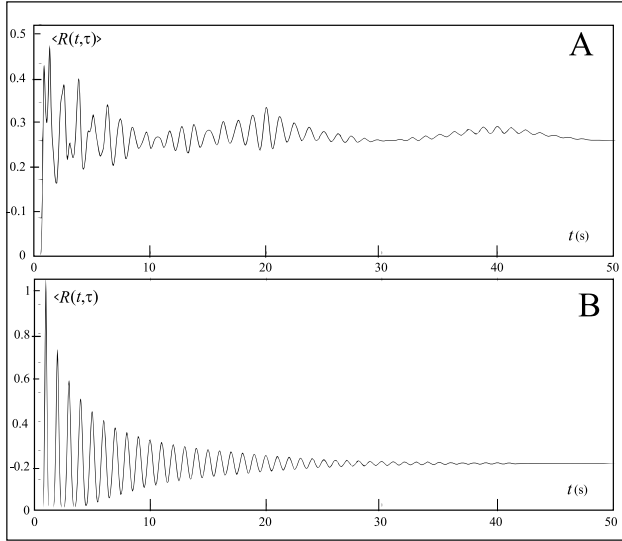


Figure 2: The graphics of mean MCA representation $\langle \hat{R}(t, \tau) \rangle$ (10) with the same parameters $\bar{S}/K = 1$, $\bar{A}|G(0)|/\bar{S} = 1$, but different heart variability parameter δ (A) $\delta = 0.3$ and (B) $\delta = 0.03$.

To illustrate how the discussed features of the synthesized instrument manifest themselves in the processing of real data, Figure 3 shows two graphs of the $\langle \hat{R}(t, \tau) \rangle$ (10) calculated and averaged over the ECG records from the Physionet portal databases NSRDB and SHAREE [10]. In both cases, with a time step $\Delta = 2$ sec a set of $N = 1000$ ACFs $\hat{R}(t, \tau)$ (2) were calculated along each record and resulting representation $\langle \hat{R}(t, \tau) \rangle$ was obtained as their arithmetic mean.

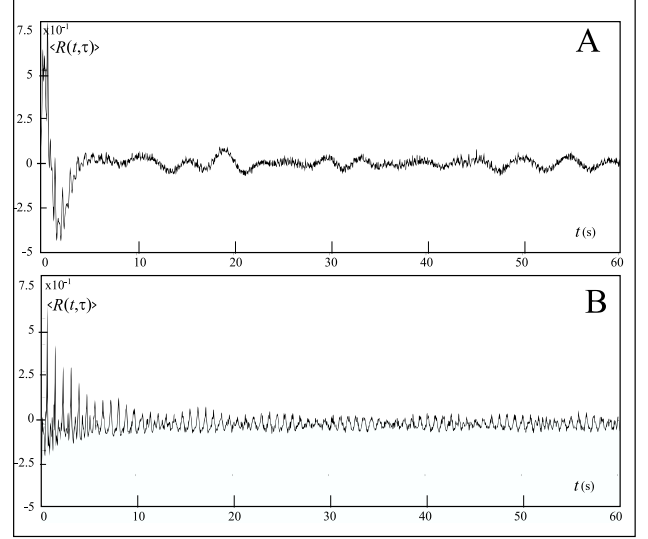


Figure 3: The graphics of mean MCA representation calculated and averaged over the real ECG records: (A) record of a healthy subject from NSRDB and (B) record of a patient with myocardial infarction from SHAREE [10].

The first graph (A) in Figure 3 is obtained from NSRDB ECG record and corresponds to the healthy person case. The second (B) is formed with the help of SHAREE data and shows the features of the disease person (myocardial infarction) representation. With the naked eye, we can see that the features of representations noted above for the model case also occur in the processing of real data. In particular, the shape of the envelope of representations allows us to accurately discriminate the case of a healthy patient from a patient with severe cardiac disease.

Conclusions

The synthesized instrument (10) showed very promising results on real records processed. Two types of ECG records from NSRDB and SHAREE databases [10] of a PhysioBank – the largest and growing archive of well-characterized digital recordings of physiologic signals were used. The proposed method made it possible to reliably discriminate the records from NSRDB, where no significant arrhythmias or any other HR disturbances were found from SHAREE records of hypertensive subjects at higher risk to develop vascular events.

Although there is still a lot of work to bring the method to the status of a useful and reliable tool for the analysis and diagnosis of heart disease, the first encouraging results suggest the optimistic prospects associated with it.

Acknowledgement

The author is grateful for the financial support of the work by Russian Science Foundation (grant 18-29-02108 mk).

References

- [1] J.S. Coviello. *ECG interpretation made incredibly easy!* Philadelphia: Wolters Kluwer, 3 edition, 2017.
- [2] R.M. Baevsky and A.G. Chernikova. Heart rate variability analysis: physiological foundations and main methods. *Cardiometry*, (10):66–76, 2017.
- [3] Task Force of the European Society of Cardiology, the North American Society of Pacing, and Electrophysiology. Heart rate variability. standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, 7(3):354–381, 1996.
- [4] A.L. Goldberger and H.E. Stanley et al. Long-range anticorrelations and non-gaussian behavior of the heartbeat. *Phys Rev Letters*, 70(9):1343–1346, 1996.
- [5] S.S. Yamamoto J.F. Thayer and J.F. Brosschot. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *Phys Rev Letters*, 141(2):122–131, 2010.
- [6] M. Malik. Geometrical methods for heart rate variability assessment. In M. Malik and A. Camm, editors, *Heart Rate Variability*, pages 47–61. Armonk, Futura Publishing Company -New York, 1995.
- [7] V.E. Antsiperov and I.V. Zabrosaev. New results for the pvc / spb detection using based on the mca heart rhythm estimation method. In *Proceedings of the 12th Russian-German Conference on Biomedical Engineering*, pages 182–186. Vladimir State University, jul 2016.
- [8] V.E. Antsiperov. Syclic renewal point processes for heart rate variability modeling. In *Extended abstracts of the second Russian Conference with international participation "Physics for Life Sciences"*, page 61. Ioffe Institute science-tech information department, sen 2017.
- [9] R. Serfozo. *Basics of Applied Stochastic Processes*. Springer, -Berlin, -Heidelberg, 2009.
- [10] L.A.N. Amaral A.L. Goldberger and et al. P.h. Ivanov. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.