Cybernetics and Systems Analysis, Vol. 39, No. 3, 2003

RESTORATION OF A STANDARD SAMPLE OF CYCLIC WAVEFORMS WITH THE USE OF THE HAUSDORFF METRIC IN A PHASE SPACE

L. S. Fainzilberg

UDC 621.317.755

A stochastic model of generation of cyclic waveforms is considered. It is shown that this model is a generalization of models of periodic and quasi-periodic functions. A constructive method is proposed for estimation of a standard sample by a cyclic waveform observed in a phase space.

Keywords: *cyclic waveform, Hausdorff metric, phase space.*

Repetitive processes frequently run in engineering and biological systems. Such processes generate specific signal waveforms, which are usually called cyclic [1] or quasiperiodic [2] in the scientific literature. Typical examples of cyclic waveforms are an electrocardiogram (ECG), a rheogram, a magnetocardiogram, and many other physiological waveforms reflecting the cyclic character of functioning of the circulatory system of living organisms.

As is well known, existing computer systems of analysis and interpretation of cyclic waveforms, in particular, ECGs, do not provide the required reliability of results up to now [3]. According to [4], this is caused, first of all, by errors made during measurement of parameters (diagnostic attributes) in processing actual waveforms in the time domain. In an alternative method of analysis of such waveforms, which was proposed in [5] and then developed in some other works, in particular, in [6–8], a waveform is assumed to be represented and processed in a phase space.

In this article, a model of generation of cyclic waveforms is proposed and, based on this model, a new method of restoration of a standard sample of a cyclic waveform from its distorted realizations observed in the phase space is investigated.

Statement of the Problem. Let an observable waveform $y(t) = \Phi[\hat{y}_0(t), \zeta(t)]$ be the result of some distortion of a periodic process $\hat{y}_0(t)$ by a random disturbance $\zeta(t)$, where $\Phi(\cdot)$ is some function. We call the standard cycle $y_0(t)$ a part of the nonobservable function $\hat{y}_0(t)$ within any its period T_0. We pose the problem of estimation of the standard sample $y_0(t)$ based on its realization $y(t)$ observed on an interval $[0, T]$, $T >> T_0$.

A Stochastic Model of Generation of Cyclic Waveforms. Before passing to the solution of the problem formulated, we consider a possible model of generation of $y(t)$ on the basis of the standard sample $\hat{y}_0(t)$. We assume that the standard sample $y_0(t)$ can be represented in the form of a function piecewise specified on the interval $[0, T_0)$ by its individual fragments as follows:

$$y_0(t) = \begin{cases} y_0^{(1)}(t), \quad 0 \le t < t_0^{(1)}, \\ \dots \\ y_0^{(i)}(t), \ t_0^{(i-1)} \le t < t_0^{(i)}, \\ \dots \\ y_0^{(K_F)}(t), \ t_0^{(K_F-1)} \le t < t_0^{(K_F)} = T_0; \end{cases} \tag{1}$$

International Scientific-Educational Center of Information Technologies and Systems, Kiev, Ukraine, *fainzilberg@svitonline.com*. Translated from Kibernetika i Sistemnyi Analiz, No. 3, pp. 20-28, May-June, 2003. Original article submitted April 9, 2003.

1060-0396/03/3903-0338$25.00

the number of such fragments is assumed to be equal to $K_F \geq 2$. As applied to an ECG, such fragments correspond to the following stages of the process of excitation of definite areas of heart: atrial depolarization (the wave *P*), excitation (the complex *QRS*), and ventricular repolarization (the wave *T*) [1].

We represent the waveform $y(t)$ being observed in the form of a sequence of distorted standard samples (1), assuming that, during each *m*th cycle of such a sequence ($m = 1, 2, \ldots$), individual fragments $y_0^{(i)}(t)$ of the standard sample are linearly dilated (contracted) in time independently of one another and that the amplitude of the function $y_0(t)$ itself is linearly dilated (contracted). In other words, it is assumed that the process of generation of the *i*th fragment ($i = 1, \ldots, K_F$) of each *m*th cycle ($m = 1, 2, \ldots$) is based on the following functional transformation:

$$y_m^{(i)}(t) = a_m y_0^{(i)}\left(\frac{t - \tau_m^{(i)}}{b_m^{(i)}}\right), \tag{2}$$

where a_m and $b_m^{(i)}$ are, respectively, the parameters of linear amplitude and time dilatations (contractions) and $\tau_m^{(i)}$ is a shift in time. To provide the continuity of the waveform being generated, we assume that $y_0(0) = 0$. The fulfillment of the latter condition can always be provided by normalization of the standard sample $\tilde{y}_0(t) = y_0(t) - y_0(0)$ in advance.

Let, within each *m*th cycle, the parameter a_m assumes a fixed value

$$a_m = 1 + \xi_m, \quad m = 1, 2, \ldots, \tag{3}$$

where ξ_m is a sequence of realizations of independent random quantities that, with the zero expectation $M[\xi\} = 0$, are distributed on an interval $[-\varepsilon_0, \varepsilon_0]$, where a fixed number $\varepsilon_0 \in [0, 1)$. We also assume that the parameter $b_m^{(i)}$ assumes the following fixed value during generation of each *i*th fragment of the *m*th cycle:

$$b_m^{(i)} = 1 + \delta_m^{(i)}, \quad m = 1, 2, \ldots, \quad i = 1, \ldots, K_F, \tag{4}$$

where $\delta_m^{(i)}$ is a sequence of realizations of independent random quantities that, with zero expectation $M\{\delta\} = 0$, are distributed over intervals $[-\Delta_0^{(i)}, \Delta_0^{(i)}]$, where fixed numbers $\Delta_0^{(i)} \in [0, 1)$.

Under such assumptions, the duration of the *i*th fragment of the *m*th cycle of the waveform $y(t)$ is connected with the duration $\Delta T_0^{(i)} = t_0^{(i)} - t_0^{(i-1)}$ of the corresponding fragment of the standard sample by the relation

$$\Delta T_m^{(i)} = (t_0^{(i)} - t_0^{(i-1)})(1 + \delta_m^{(i)}).$$

Hence, the total duration of the *m*th cycle of the generated waveform $y(t)$ is determined by the expression

$$T_m = T_0 + \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)})\delta_m^{(i)},$$

to the beginning of the *m*th cycle corresponds the moment of time

$$t_m = (m-1)T_0 + \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)})\delta_j^{(i)},$$

and to the beginning of the *i*th fragment of the *m*th cycle corresponds the moment of time

$$t_m^{(i-1)} = (m-1)T_0 + \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)})\delta_j^{(i)} + \sum_{l=1}^{i-1} (t_0^{(l)} - t_0^{(i-1)})(1 + \delta_m^{(l)}). \tag{5}$$

We apply functional transformation (2) to the *i*th fragment of the standard sample $y_0(t)$, setting the shift parameter $\tau_m^{(i)} = t_m^{(i-1)} - (1 + \delta_m^{(i)})t_0^{(i-1)}$. Then it follows from (2) and relations (3)–(5) that the process of generation of the *i*th fragment of the *m*th cycle can be represented in the form

$$y_m^{(i)}(t) = (1+\xi_m) y_0^{(i)}(\theta), \quad m = 1, 2, \ldots, \quad i = 1, \ldots, K_F \ , \tag{6}$$

where

$$\theta = \frac{t-(m-1)T_0 + (1+\delta_m^{(i)}) t_0^{(i-1)} - \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)}) \delta_j^{(i)} - \sum_{l=1}^{i-1} (t_0^{(l)} - t_0^{(l-1)})(1+\delta_m^{(l)})}{1+\delta_m^{(i)}}. \tag{7}$$

The proposed model that describes the distortions of the standard sample $y_0(t)$ that are irregular in time is more suitable for description of actual cyclic waveforms, in particular, ECGs, than its simplified variant

$$y_m^{(i)} = (1+\xi_m) y_0^{(i)} \left[\left(t - (m-1)T_0 - T_0 \sum_{j=1}^{m-1} \delta_j \right) (1+\delta_m)^{-1} \right], \quad m = 1, 2, \ldots,$$

obtained under the assumption that the random parameter δ in (7) depends only on the number m of a cycle but does not depend on the number i of a fragment.

It is easily verified that the stochastic model (6) is a direct generalization of well-known models of strictly periodic and almost-periodic processes. In fact, setting $\delta_j^{(i)} = 0 \ \ \forall j = 1, \ldots, m, \ \ i = 1, \ldots, K_F$, in formula (7), we can represent model (6) in the form of the relation

$$| y_m[t] - y_0[t-(m-1)T_0] | \le \varepsilon_0, \quad m = 1, 2, \ldots,$$

which describes an almost-periodic process [9] and, under the additional condition $\varepsilon_0 = 0$, is reduced to the model of a strictly periodic function $y[t-(m-1)T_0] = y_0[t]$.

The model proposed can be easily generalized to the description of the process of generation of more complicated cyclic waveforms, in particular, ECGs with varying morphology of individual cycles (extrasystoles) [10]. To this end, it suffices to introduce into consideration not one but $G > 1$ standard samples $y_{01}(t), \ldots, y_{0G}(t)$ and to assume that each mth cycle is generated by similar distortions of one of these standard samples chosen at random with probabilities $P_g \ \left(\sum_{g=1}^{G} P_g = 1 \right)$.

A Generator of Cyclic Sequences. Let us consider a sufficiently simple algorithm of construction (generation) of discrete cyclic sequences from standard samples. Let each of $G > 1$ standard samples $y_{0g}(t) \ (g = 1, \ldots, G)$ be represented by a finite number N_g of discrete values $y_{0g}[1], \ldots, y_{0g}[N_g]$ fixed with a constant sampling interval. We will specify the total number $K_F^{(g)}$ of fragments of each standard sample and the numbers of points $N_g^{(i)} \in [1, N_g], \ i = 1, \ldots, K_F^{(g)} - 1$, that determine the bounds of the ith and $(i+1)$th fragments of the gth standard sample. For such initial data, the procedure of generation of a cyclic sequence is reduced to the steps considered below.

Step 1. Specify the total number M of cycles of the sequence being generated.

Step 2. Compute the number M_g of cycles generated by the gth standard sample by the formula $M_g = R[P_g M]$; here and in the sequel, $R[\eta]$ is the operation of rounding to the nearest integer η.

Step 3. Choose the number g_m of the standard sample that generates the mth cycle $(m = 1, \ldots, M)$ based on the value of a realization λ_m of an integer random quantity λ distributed over the interval $[1, G]$, i.e., we have $g_m = \lambda_m$.

Step 4. If $M_g = 0$, then repeat step 3.

Step 5. Compute the number of points of the ith fragment of the mth cycle by the formula

$$N_m^{(i)} = R[(N_g^{(i)} - N_g^{(i-1)})(1+\delta_m^{(i)}],$$

where $\delta_m^{(i)}$ is a realization of the random quantity δ distributed with zero expectation over the interval $[-\delta, \delta_0], \ \ \delta_0 \in [0, 1)$.

Step 6. Using any interpolation method, compute the values of the sequence being generated in $N_m^{(i)}$ points from the discrete values of the ith fragment of the gth standard sample in $N_g^{(i)} - N_g^{(i-1)}$ nodes.

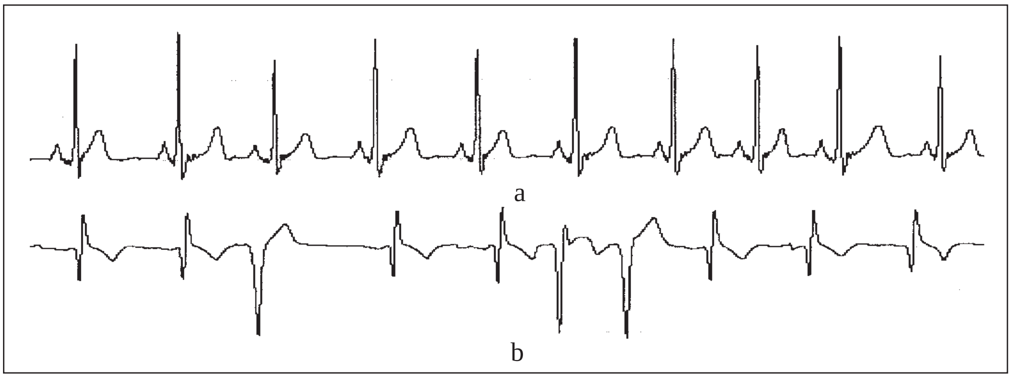

Fig. 1. The ECG waveform generated by model (6) on the basis of one standard sample (a) and two standard samples (b).

Step 7. Modify each computed value of y, using the multiplicative procedure $\tilde{y} = y(1+\xi_m)$, where ξ_m is a realization of a random quantity ξ that is distributed with zero expectation over the interval $[-\varepsilon_0, \varepsilon_0]$, $\varepsilon_0 \in [0,1)$.

Step 8. If $i < K_F^{(g)}$, then go to step 5.

Step 9. Assign $M_g := M_g - 1$.

Step 10. If $m < M$, then go to step 3.

The results of simulation testify to the effeciency of the considered algorithm in imitating actual cyclic waveforms (Fig. 1).

A Method of Estimation of a Standard Sample by its Distorted Realization. Let cyclic waveform (6) be represented by a sequence $Y^{(K)} = \{y[k],\ k=1,\dots,K\}$ of discrete values observed during $M > 2$ cycles. We assume that, for each kth value, there is an estimate of the derivative $\dot{y}[k]$. We normalize

$$y_k^* = \frac{y[k] - \min\limits_{1\le k\le K} y[k]}{\max\limits_{1\le k\le K} y[k] - \min\limits_{1\le k\le K} y[k]} \quad \forall k=1,\dots,K,$$

$$\dot{y}_k^* = \frac{\dot{y}[k] - \min\limits_{1\le k\le K} \dot{y}[k]}{\max\limits_{1\le k\le K} \dot{y}[k] - \min\limits_{1\le k\le K} \dot{y}[k]} \quad \forall\ k=1,\dots,K,$$

and form the set $Q = \{z_k = (y_k^*, \dot{y}_k^*),\ k=1,\dots,K\}$ of points belonging to the trajectory of the waveform being observed in the two-dimensional normalized phase space $y - \dot{y}$.

Let the numbers of points $k = k_m$, $m = 1,\dots,M$, that correspond to the beginning of the mth cycle be also known. (The algorithm of determination of the numbers k_m is not considered in this article.) Then the set Q can be subdivided into M subsets $Q_1,\dots,Q_M$ of normalized vectors z_k whose ends are on phase trajectories of individual cycles.

We will estimate the distance between any two subsets $Q_i \subset Q$ and $Q_j \subset Q$, $i \ne j$, by the Hausdorff metric [11]

$$R_H(Q_i, Q_j) = \max\left\{ \max_{z_j\in Q_j} \min_{z_i\in Q_i} \rho(z_i, z_j),\ \max_{z_i\in Q_i} \min_{z_j\in Q_j} \rho(z_i, z_j) \right\}, \tag{8}$$

where $\rho(z_i, z_j) = ||z_i - z_j|| = \sqrt{(y_i^* - y_j^*)^2 + (\dot{y}_j^* - \dot{y}_j^*)^2}$ is the Euclidean distance between points z_i and z_j.

We call the reference cycle the subset $Q_0 \subset Q$ of vectors z_k that has the minimal total distance (8) from the other $M-1$ subsets, i.e., we have

$$Q_0 = \arg\min_{\substack{1\le j\le M \\ j\ne i}} \sum_{i=1}^{M} R_H(Q_i, Q_j), \tag{9}$$

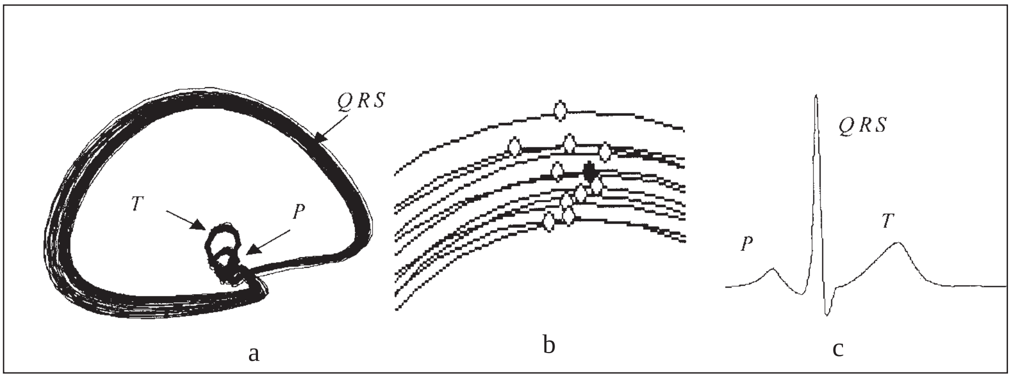

Fig. 2. An illustration of the algorithm that estimates of a standard sample by an example of an ECG; here, (a) are phase trajectories, (b) are fragments of trajectories, and (c) is the standard cycle.

and will estimate the standard sample (the average cycle) by averaging the points of different trajectories located in a vicinity of points $z_1, \ldots, z_{K_0}$ of the reference cycle.

To this end, we choose the trajectories to be averaged by determining the subset $\Omega_0 \subset Q$ of all the trajectories whose Hausdorff distance from the reference one is smaller than a specified quantity δ_H, i.e., we have $\Omega_0 = \{Q_j : R_H(Q_j, Q_0) < \delta_H\}$. To improve the estimate, we represent the reference cycle Q_0 and the other cycles $Q_j \in \Omega_0$ by a sequence of extended vectors $u_k = (z_k, \tau_k)$ that, in addition to normalized phase coordinates $z_k = (y_k^*, \dot{y}_k^*)$, contain an additional component τ_k. The quantity τ_k is computed at each lth point of the jth trajectory by the formula

$$\tau_l^{(j)} = \frac{l - k_m^{(j)}}{K_j - 1}, \quad l = k_m^{(j)}, \ldots, k_m^{(j)} + K_j - 1; \quad j = 1, \ldots, M,$$

where $k_m^{(j)}$ is the number of the first point of the jth trajectory consisting of K_j points.

The introduction of the additional component τ_k makes it possible, in averaging points, to estimate their proximity not only from the viewpoint of values of phase coordinates y_k^* and $\dot{y}_k^*$ but also taking into account their synchronism in time. To this end, we propose to determine the Euclidean distance $\rho_\tau(u_0, u_j) = ||u_0 - u_j|| = \sqrt{(y_0^* - y_j^*)^2 + (\dot{y}_0^* - \dot{y}_j^*)^2 + (\tau_0 - \tau_j)^2}$ between extended vectors $u_0 = (z_0, \tau_0)$ of the reference trajectory and extended vectors $u_j = (z_j, \tau_j)$ of the other trajectories $Q_j \in \Omega_0$ and use the following relation to estimate the sequence of points $\hat{z}_i$, $i = 1, \ldots, K_0$, of the average cycle:

$$\hat{z}_i = \frac{z_0^{(i)} + \sum_{j=1}^{M_0} z_{ij}^{\min}}{M_0 + 1}, \quad i = 1, \ldots, K_0, \tag{10}$$

where $z_{ij}^{\min}$ is the point that belongs to the jth trajectory (which is not the reference one) and whose Euclidean distance ρ_τ from a point $z_0^{(i)} \in Q_0$ of the reference trajectory is minimum, i.e., we have

$$z_{ij}^{\min} = \underset{k_m^{(j)} \le l \le k_m^{(j)} + K_j - 1}{\arg\min} \rho_\tau(u_0^{(i)}, u_j^{(l)}), \quad i = 1, \ldots, K_0.$$

The sequence of vectors $\hat{z}_i = (\hat{y}_j, \hat{\dot{y}}_j)$, $i = 1, \ldots, K_0$, which is computed by formula (10), gives an estimate of the sought-for nonobservable standard sample in the phase space, and the corresponding sequence $\hat{y}_1, \ldots, \hat{y}_{K_0}$ gives an estimate of the standard cycle in the time domain (Fig. 2).

A Model Example. Let a standard sample $y_0(t)$ be of the form of an isosceles triangle (Fig. 3a) specified by two fragments represented in the form of linear functions as follows:

$$y_0(t) = \begin{cases} t, & 0 \le t < 3, \\ 6 - t, & 3 \le t < 6. \end{cases} \tag{11}$$

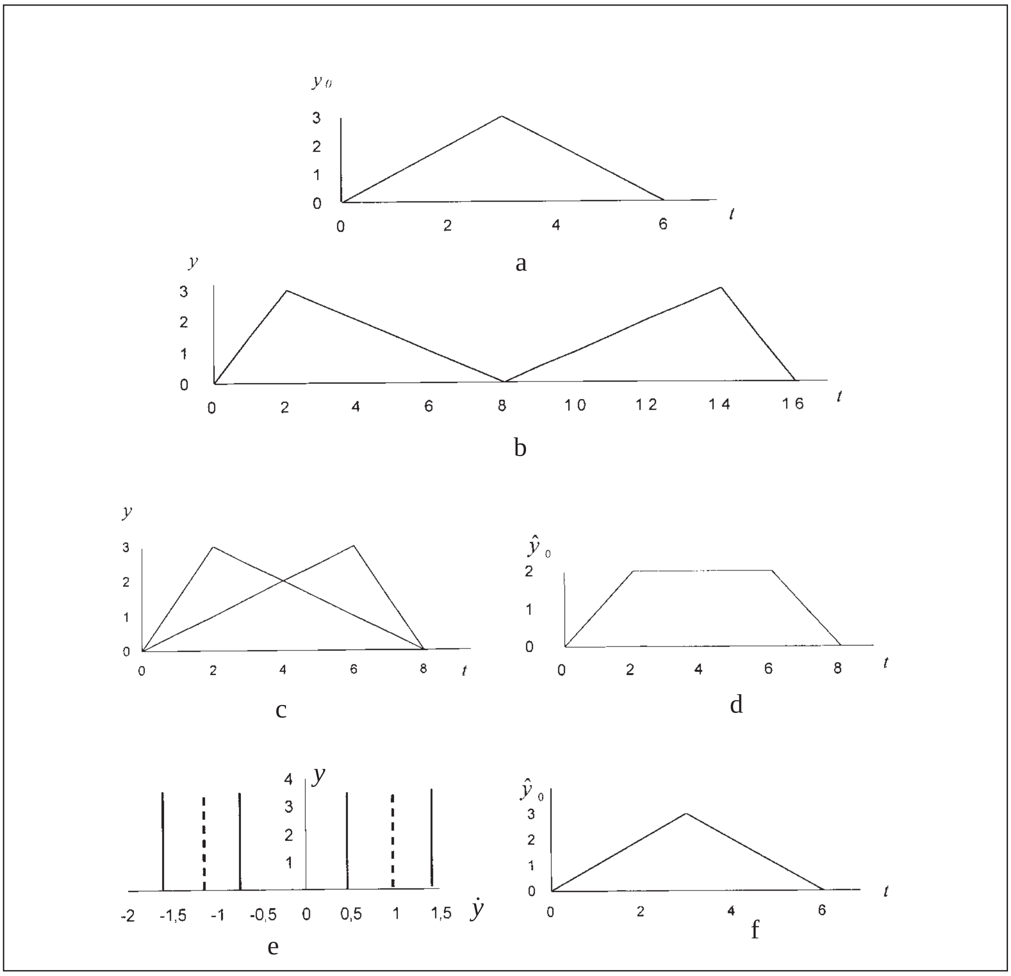

Fig. 3. An illustration to the model example; here, (a) is a standard sample, (b) is an observable waveform, (c) are cycles overlapped in time, (d) is an estimate of the standard sample in the time domain, (e) is a phase trajectory, and (f) is an estimate of the standard sample in the phase space.

We assume that two cycles of a waveform are observed that are constructed, in accordance with model (6), from standard sample (11) and, at the same time, after the first cycle, the parameters of time dilatation assume the values $b_1^{(1)} = 2/3$ and $b_1^{(2)} = 2$ and, after the second cycle, they assume the values $b_2^{(1)} = 2$ and $b_2^{(2)} = 2/3$. As a result, the waveform being observed is described by the function

$$y(t) = \begin{cases} 1.5t, & 0 \le t < 2, \\ 4 - 0.5t, & 2 \le t < 8, \\ 0.5t - 4, & 8 \le t < 14, \\ 24 - 1.5t, & 14 \le t < 16, \end{cases} \tag{12}$$

whose plot is depicted in Fig. 3b.

To estimate the standard sample in the time domain, we overlap the cycles being observed (Fig. 3c) on the interval [0, 8] and average them in the time domain. It is easy to see that the following estimate (Fig. 3d) is obtained in this case:

$$\hat{y}_0(t) = \begin{cases} t, & 0 \le t < 2, \\ 2, & 2 \le t < 6, \\ 8 - t, & 8 \le t < 8, \end{cases}$$

whose form does not correspond to that of the standard sample. We note that the averaging of the same cycles in the phase space (Fig. 3e) with subsequent passage to the time domain (Fig. 3f) makes it possible to precisely restore the standard sample (11).

Practical Results. The proposed method of estimation of standard cycles was practically used in developing new computer systems of processing ECGs in the phase space [12–14]. These systems were medically tested at the N. D. Strazhesko Cardiology Institute in Kiev.

During these tests, some new diagnostic attributes of ECGs in the phase space were established, which allow one to diagnose the patients with rheumatoid arthritis even if their ECGs are considered to be unchanged in the traditional time-domain analysis [12]. The method proposed makes it possible to reveal subtle varations in the morphology of ECG cycles and thereby to increase the sensitivity and specificity of diagnoses during mass prenosological diagnostics. It can be used for estimation of functional states of operators working under increased risk (drivers of means of transportation, air-traffic controllers, pilots, etc.) [13] and also for studying the influence of parameters of the external environment on the ECG of a healthy person [14].

Thus, the stochastic model (6) of the process of generation of cyclic waveforms is proposed, which is a direct generalization of well-known models of periodic and almost-periodic functions. This model can be easily generalized to the case of generation of cyclic waveforms with varying morphology of individual cycles.

Despite the fact that the model proposed is based on linear operations applied to fragments of standard sample (1), it describes distortions irregular in time for individual cycles of the waveform being generated, which is peculiar to many actual cyclic waveforms, in particular, ECGs.

The possibility of obtaining a reasonable estimate of a nonobservable standard sample from a realization of a cyclic waveform with the help of a constructive algorithm of averaging trajectories of individual cycles in the phase space with the use of the Hausdorff metric is shown. In computer systems for processing ECGs, the method considered made it possible to increase the sensitivity and specificity of ECG diagnostics.

REFERENCES

1. P. P. Kanjilal, J. Bhattacharya, and G. Saha, "Robust method for periodicity detection and characterization of irregular cyclical series in terms of embedded periodic components," Phys. Rev., **59**, 4013–4025 (1999).
2. E. T. Whittaker and G. N. Watson, "Quasi-Periodic Functions," in: A Course in Modern Analysis, Cambridge University Press, Cambridge (England) (1990), pp. 445–447.
3. A. M. Berkutov, S. G. Gurzhin, A. A. Dunayev, and Ye. M. Proshin, "Increasing the efficiency of registration of the form of electrocardiographic waveforms by correlation processing in digital oscillography," Biomedical Technologies and Radio Electronics, No. 7, 4–13 (2002).
4. A. K. Valuzhis and A. P. Rashimas, "Statistical algorithm for the structural analysis of an electrocardiographic waveform," Kibernetika, No. 3, 91–95 (1979).
5. N. M. Amosov, B. T. Agapov, and Yu. V. Panichkin, "Investigation of the contractile myocardium function by the method of phase coordinates," Dokl. Akad. Nauk SSSR, **202**, No. 1, 245–247 (1972).
6. L. L. Frumin and M. B. Shtark, "On a phase portrait of an electrocardiogram," Avtometriya, No. 2, 51–54 (1993).
7. L. S. Fainzilberg and T. P. Potapova, "Computer analysis and recognition of an electrocardiographic image in a cognitive phase space," in: Proc. 6th Intern. Conf. on Computer Analysis of Images and Patterns (CAIP'95), Prague (Czech Republic), 668–673 (1995).
8. L. S. Fainzilberg, "Heart functional state diagnostics using pattern recognition of ECG-images in a phase space," in: Proc. 6th European Congress on Intelligent Techniques and Soft Computing (EUFIT'98, Aachen, Germany, 1998), No. B-27, **3** (1998), pp. 1878–1882.
9. V. G. Lapa, Mathematical Foundations of Cybernetics [in Russian], Vyshcha Shkola, Kiev (1974).
10. V. V. Murashko and A. V. Strutinskii [in Russian], Electrocardiography, Meditsina, Moscow (1991).
11. V. A. Skvortsov, Examples of Metric Spaces [in Russian], MTsNMO, Moscow (2002).
12. L. S. Fainzilberg, A. F. Klubova, L. A. Stadnyuk, I. A. Chaikovskii, and D. Lerkhe, "A new method of analysis of ECGs of patients with rheumatoid a rthritis," Ukr. Revmatologich. Zh., No. 2, 48–51 (2001).
13. L. S. Fainzilberg, "An information technology for diagnostics of functional states of operators," USiM, No. 4, 40–45 (1998).
14. V. V. Vishnyevskii, M. V. Ragul'skaya, and L. S. Fainzilberg, "The action of solar activity on morphological parameters of the ECG of the heart of a healthy person," Biomedical Technologies and Radio Electronics, No. 3, 3–12 (2003).