DOI 10.1007/s11018-022-02025-4
Measurement Techniques, Vol. 64, No. 11, February, 2022

ACOUSTIC MEASUREMENTS

METHOD FOR AUTOMATIC ONLINE UPDATING OF PERSONAL BIOMETRIC DATA BASED ON SPEECH SIGNAL OF THE BIOMETRIC SYSTEM USER

A. V. Savchenko[1] and V. V. Savchenko[2]

UDC 534.6:53.082.4

The article considers the problem of personal biometric data "aging" over time. A method has been proposed to overcome this problem by automatically updating the specified data in the biometric system storage using the speech signals of registered users obtained during latest requests for their identification and online service. The proposed method uses a scale-invariant indicator of the voice template quality. As a result, it is characterized by guaranteed reliability of the decisions made in the conditions of a wide speech signal dynamic range. It was established that the use of a scale-invariant indicator provides the guaranteed significance level of decisions made by a conventional observer. A full-scale experiment implementing the proposed method has been set up and carried out using an authoring software; practical justification for the effectiveness of the proposed method with real speech data has been given. The results obtained are intended for using in the development of new and modernization of existing systems and technologies for automated quality control and updating of personal biometric data.

Keywords: *speech acoustics, acoustic measurements, speech signal, biometrics, biometric system, personal biometric data, voice template.*

Introduction. Currently, biometric identification systems (BIS) are widely used in various sectors of economy and management. An example is the nationwide software platform "Unified Biometric System" (https://bio.rt.ru) created in 2018 at the initiative of the Bank of Russia and the Ministry of Communications of the Russian Federation. Its main purpose is remote (online) identification of Russian citizens based on preliminary collected personal biometric data (PBD). At the first stage of implementing the Unified Biometric System in the country's economy, priority was given to the banking sector, in which the needs for a remote method of providing services to the population, in particular, using online voice requests of users, have existed for a long time and in the most acute form. Especially for the banks of the country, Rostelecom, being the operator of the Unified Biometric System, has created a mobile application "Biometrics." Since July 2018, citizens of the Russian Federation, through all major banks of the country, have received the opportunity to register PBD in a bimodal version: a face photograph and a voice recording. More than 200 Russian banks are involved in the PBD collection for the Unified Biometric System project. However, not all problems in this topical direction have been successfully resolved at the moment.

One of the most acute problems in biometrics is the natural PBD aging or the partial loss of the PBD consumer properties over time [1–3]. Its solution involves periodic PBD updating in the BIS storage. The recommended period for updating voice templates is in the range from one year to several years, depending on the voice features of a particular user [4, 5]. The variability of the user's voice is accompanied by the variability of speech articulation, which inevitably manifests itself

[1] National Research University Higher School of Economics, Nizhny Novgorod, Russia; e-mail: avsavchenko@hse.ru.
[2] Nizhny Novgorod State Linguistic University, Nizhny Novgorod, Russia; e-mail: vvsavchenko@yandex.ru.

Translated from Izmeritel'naya Tekhnika, No. 11, pp. 60–66, November, 2021. Original article submitted June 09, 2021. Accepted July 20, 2021.

in a change in the features of the user's face [6, 7]. As a partial solution to the problem under consideration, it is possible to propose automatic PBD updating in the BIS storage, taking into account the fact that the PBD aging rate depends on the physiological features of the user's personality [4], and the vocal tract is a highly sensitive indicator of such features [8]. Hence, it follows that it is possible to find the dependence of the PBD quality estimate on the acoustic properties and characteristics of the voice templates of the registered BIS users. Using the measured acoustic characteristics of the user's voice (using his online speech signal), one can judge his current psychophysical state [5, 9]. In fact, this is a new approach to the problem of PBD updating, in which the next moment of its rewriting for the BIS storage is conditioned only by the fact of violation of the established requirements for the acoustic quality of the stored voice templates. At that, it is necessary to solve the problem of scaling the speech signal in amplitude, taking into account the wide (tens of decibels) dynamic range at the BIS input [3, 10].

The objective of this study is to develop a method for automatic online PBD updating based on the speech signals of the BIS users using the original authoring modification [11] of the Kullback–Leibler information standard [12], which is invariant to the scale (amplification factor) of the speech signal at the output of the communication channel.

Formulation of the problem. In the theoretical part of this study, the original authoring mathematical apparatus is used in the framework of the information theory of speech perception [13, 14]. Let $x(t)$ be a speech signal of finite duration T with a service request from the BIS user. Let us divide this signal into a finite sequence of quasi-stationary segments (frames) of short duration $\tau = 20$–30 ms. Let us consider the case of discrete time $t = 1, 2, \ldots$ with the sampling period $T = F^{-1}$, where F is the discretization frequency of the speech signal. Let us set its current (observed) frame by the sampling vector $\mathbf{x}$ of dimension $N = \tau F$. Based on the multidimensional (n-dimensional) Gaussian speech signal model widespread [14, 15] in the automatic speech processing systems in order to establish its distribution law Norm(K) with an autocorrelation matrix (ACM) K, let us consider the problem of testing two statistical hypotheses

$$\left.\begin{aligned} H &: K \subset \{K_r\}; \\ \bar{H} &: K \not\subset \{K_r\}, \end{aligned}\right\} \tag{1}$$

where K_r is the ACM of the rth ($r \le R$) minimal speech unit of the phoneme type as part of the stored (actual) voice template of a certain user; R is the number of controlled phonemes; the line above the symbol H denotes the operation of logical negation.

The system of equations (1) is equivalent to the equality

$$H = \bigcup_{r=1}^{R} H_r,$$

where H_r is the rth (partial) component of the hypothesis H about the equality of two ACMs:

$$H_r\colon K = K_r,\ r = 1, \ldots, R. \tag{2}$$

According to [3, 4], the phonemes controlled in the BIS can be a set of vowel phonemes as the most meaningful sounds of the user's speech in the theoretical information sense [11–14] and the most distinguished from the speech signal $x(t)$ by the amplitude feature [4] or by the feature of the stable frequency of the main tone [16]. In this case, $R = 6$ for the Russian language. At that, each vowel phoneme can be specified in the BIS storage by the set $\{\mathbf{x}\}$ of its allophones from the speech signal of the registered user as his admissible voice templates.

Let us suppose that the user in question is registered in the BIS under his personal identifier Id (otherwise he will be automatically denied a service request), and his observed frame $\mathbf{x}$ refers to the number R of the controlled (vowel) phonemes. In this case, Eq. (1) has two solutions: $W(x)$ and $\overline{W}(x)$. In the first case, the current observed frame $\mathbf{x}$ is recognized by the BIS (i.e., by a conventional observer) as a perfect vowel sound of the certain user's speech, in the second case, it is recognized as a boundary (marginal) sound of speech. In this case, the vector $\mathbf{x}$ is sent to the BIS as the allowed sound of the user's speech in addition to his currently collected voice templates. At the same time, a new vector $\mathbf{y}$ of the typical features of his face, obtained during the user's current appeal using special means [6, 17], can be sent to the PBD storage. Over time, part of the stored biometric templates $\mathbf{x}$, $\mathbf{y}$ from among those that have served a specified period (for example, three years for the Unified Biometric System [4]) are automatically deleted from the storage. Essentially, it means the online PBD updating. After that, the next frame of the speech signal $x(t)$, up to the last frame, is subject to testing using Eq. (1).

What kind of decision at each next step the observer will make depends on the criterion he uses or the decision making rule. In [4], it was proposed to use for this purpose the criterion of the admissible (threshold) value

$$W(\mathbf{x})\colon (\exists r \le R)\ \rho_r(\mathbf{x}) \le \rho_0 \tag{3}$$

of the information mismatch in the Kullback–Leibler metric [12]

$$\rho_r(\mathbf{x}) \triangleq 0.5[n^{-1}\mathrm{tr}(SK_r^{-1}) - n^{-1}\ln\left|SK_r^{-1}\right| - 1], \tag{4}$$

where $S \triangleq M^{-1}\sum_{m=1}^{M}\mathbf{x}_m\mathbf{x}_m^{\mathrm{T}}$ is the ACM statistical estimate of a centred speech signal over an observation range into one speech frame $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)$ over a sample of vector observations $\mathbf{x}_m$, $m \le M$, of a finite volume $M = [Nn^{-1}]$; $\mathbf{x}_m$ is the n-vector (column) of the speech signal $x(t)$ samples within its partial (mth) observation segment with a duration $\tau_0 = \tau M^{-1}$ (here and below, the symbols $\mathrm{tr}(\cdot)$ and $|\cdot|$ denote the trace and determinant of a square $(n \times n)$-matrix; $[\cdot]$ is the integer part of a rational number; $\triangleq$ is the equality by definition; $\exists$ is the existential quantifier). The threshold level ρ_0 in Eq. (3) is set by a conventional observer depending on the permissible probability of a type 1 error [15]:

$$\alpha \triangleq P\{\overline{W}(\mathbf{x})\left|H_0\right.\} = P\{\min_{r\le R}\rho_r(x) > \rho_0\left|H_0\right.\} = P\{\rho_\nu(x) > \rho_0\left|H_\nu\right.\} \triangleq \alpha(\rho_0), \tag{5}$$

where $\nu \le R$. In mathematical statistics [18], probability in Eq. (5) is defined as the significance level (reliability) of decisions $\overline{W}(\mathbf{x})$ of the observer about the PBD updating. The higher the threshold ρ_0, the lower the significance level.

The problem is reduced to an explicit determination of the dependence $\alpha(\rho_0)$ on the right-hand side of Eq. (5). However, as applied to criterion in Eq. (3), it is an almost insoluble problem. The trouble lies in the known instability of the scale or amplitude of the speech signal within a wide range at the BIS input [10, 19]. Considering that the standards used in Eq. (1) are stored in the form of an R-set of ACMs of voice templates of a certain (fixed) amplitude, in practice it is necessary to observe the condition: for any constant $c_r > 0$ in the role of a scale factor, the following system of equalities should be fulfilled:

$$\overline{W}_r(\mathbf{x}) = \overline{W}_r(c_r\mathbf{x}),\ \ r = 1, ..., R. \tag{6}$$

Otherwise, the criterion in Eq. (3) loses its operating capacity, since for any significance level α_0, according to Eq. (5), it is necessary to observe the paradoxical requirement $\rho_0 \to \infty$ to the threshold level of criterion in Eq. (3) in the entire range of values of the scale factor $c_r \gg 1$. To eliminate this obstacle, let us modify the definition of the information quality indicator of voice templates in Eq. (4), endowing it with the invariance property of Eq. (6) to the speech signal scale at the input.

Information quality indicator. Within the framework of the general-system principle of minimum information mismatch [12, 15], let us consider the optimization problem of finding the minimum of the quantity $\rho_r(c_r\mathbf{x})$ over the definition domain of the scale factor $c_r > 0$, where

$$\begin{aligned}\rho_r(c_r\mathbf{x}) &= 0.5\left[n^{-1}c_r^2\mathrm{tr}(SK_r^{-1}) - n^{-1}\ln\left|c_r^2SK_r^{-1}\right| - 1\right] = \\ &= 0.5\left[c_r^2n^{-1}\mathrm{tr}(SK_r^{-1}) - \ln c_r^2 + n^{-1}\ln\left|S^{-1}\right| - n^{-1}\ln\left|K_r^{-1}\right| - 1\right].\end{aligned} \tag{7}$$

In order to do that, in accordance with the technique in [11], let us first find the first derivative of the objective function $\rho_r(c_r) \triangleq \rho_r(c_r\mathbf{x})|_{\mathbf{x}}$:

$$d\rho_r(c_r)/dc_r = c_rn^{-1}\mathrm{tr}(SK_r^{-1}) - c_r^{-1}.$$

Let us equate it to zero and obtain the optimization equation

$$c_r^2n^{-1}\mathrm{tr}(SK_r^{-1}) - 1 = 0.$$

Let us find the root of this equation

$$c_r^* = [n^{-1}\mathrm{tr}(SK_r^{-1})]^{-0.5}$$

and substitute the obtained result into Eq. (7):

$$\rho_r(c_r^*\mathbf{x}) = 0.5[\ln(n^{-1}\mathrm{tr}(SK_r^{-1})) + n^{-1}\ln|S^{-1}| - n^{-1}\ln|K_r^{-1}|] \triangleq \rho_r^*(\mathbf{x}). \tag{8}$$

Equation (8) defines the modified quality indicator of the rth voice template as an alternative to Eq. (4) in the role of the optimal decision statistics for criterion in Eq. (3). This indicator has the property of scale invariance in the formulation of Eq. (6):

$$\forall c_r > 0 : \rho_r^*(c_r\mathbf{x}) = 0.5\left[\ln\left(n^{-1}\mathrm{tr}(c_r^2 S K_r^{-1})\right) + n^{-1}\ln\left|c_r^{-2}S^{-1}\right| - n^{-1}\ln\left|K_r^{-1}\right|\right] =$$

$$= 0.5\ln\left[c_r^2 n^{-1}\mathrm{tr}(SK_r^{-1})\left|c_r^{-2}S^{-1}\right|^{1/n}\left|K_r^{-1}\right|^{-1/n}\right] = 0.5\ln\left[n^{-1}\mathrm{tr}(SK_r^{-1})\times\left|S^{-1}\right|^{1/n}\left|K_r^{-1}\right|^{-1/n}\right] =$$

$$= 0.5\left[\ln\left(n^{-1}\mathrm{tr}(SK_r^{-1})\right) + n^{-1}\ln\left|S^{-1}\right| - n^{-1}\ln\left|K_r^{-1}\right|\right] = \rho_r^*(\mathbf{x}).$$

The set of Eqs. (3), (4), (8) determines the method of PBD updating based on the results of a speech signal online processing. The proposed method can be additionally substantiated from a practical point of view.

Taking into account the rapid convergence (with a rate of unimprovable order $1/M$) of the ACM statistical estimates using the sample mean formula [18], one should expect that the empirical distribution Norm(S) should not differ much from its standard Norm(K_r) at the validity of the partial hypothesis of Eq. (2) in the conditions of finite (with a volume $M \gg 1$) samples of observations. For example, with the speech frame duration $\tau = 30$ ms, the speech signal sampling rate $F = 8$ kHz, and the dimension of its distribution $n = 20$ (typical parameter values for automatic speech processing algorithms [3, 4]), the frame dimension will be $N = 30{\cdot}8 = 240$ and, therefore, $M = 240/20 = 12$ of the disjoint signal segments $x(t)$. When reducing the frame duration to $\tau = 20$ ms, one obtains $M = 20{\cdot}8/20 = 8$ segments of the speech signal, which is also a lot.

Measurement technique. Let us expand Eq. (8) in explicit form through the N-vector $\mathbf{x}$ of the speech signal samples over an observation interval of one frame duration using the autoregressive (AR) model of the rth phoneme of the user as his voice template with the same name [13]:

$$x_r(t) = \sum_{i=1}^{p} a_r(i)x_r(t-i) + \eta_r(t), \quad t = 1, 2, ..., \tag{9}$$

where $a_r(i)$ denotes the ith coefficient ($i \le p$) of the autoregression of the pth order; $\{\eta_r(t)\}$ is the generating process of the white Gaussian noise type in discrete time t; $\sigma_r^2 = \mathrm{const}$ is the dispersion of $\{\eta_r(t)\}$.

Dispersion σ_r^2 here determines the minimum achievable value of the mean square of the linear prediction error for a random time series in Eq. (9) one step ahead [14, 20]. Under the condition $p < n$, this dispersion is equal to the reciprocal of the first element of the same-name (rth) inverse ACM [15]:

$$\sigma_r^2 = (\mathbf{e}^{\mathrm{T}}K_r^{-1}\mathbf{e})^{-1}.$$

Here, $\mathbf{e} = (1, 0, \ldots, 0)^{\mathrm{T}}$ is the indicator n-vector column composed of only zeros, with the exception of one in the first position. Similarly, the corresponding vector of AR coefficients $\mathbf{a}_r$ can be determined as

$$(1; -\mathbf{a}_r^{\mathrm{T}})^{\mathrm{T}} = \sigma_r^{\mathrm{T}}K_r^{-1}\mathbf{e} = (K_r^{-1}\mathbf{e})/(\mathbf{e}^{\mathrm{T}}K_r^{-1}\mathbf{e}) \triangleq \mathbf{b}_r,$$

consisting of the elements of the first column of the same-name inverse ACM taken with a coefficient σ_r^2, excluding its first element. Here, $\mathbf{b}_r$ is the vector of coefficients of the linear whitening filter tuned (at the stage of data preparation) to the signal of the rth phoneme $x_r(t)$. The whitening filter is a key link in the proposed measurement technique. Its order is $p = n - 1$. The dynamics of the filter is described by the expression inverse with respect to Eq. (9) [15]:

$$z_r(t) = x(t) - \sum_{i=1}^{n-1} a_r(i)x(t-i), \quad t = 1, 2, \ldots. \tag{10}$$

The dispersion $\sigma_r^2(\mathbf{x}) \triangleq \langle z_r^2(t)\rangle$ of the output signal $z_r(t)$ ($\langle\cdot\rangle$ is the mathematical expectation of a random variable) corresponds to the general relation $\sigma_r^2(\mathbf{x}) \ge \sigma_r^2$. The strict equality is achieved here only in the asymptotics (at $N \to \infty$), when the same-name phoneme signal of Eq. (9) arrives at the input of the whitening filter of Eq. (10). The empirical (by sample) estimate of this dispersion is determined by the formula of the sample mean square of the response $z_r(\mathbf{x}_m) = \mathbf{b}_r^{\mathrm{T}}\mathbf{x}_m$ of the rth whitening filter of Eq. (10) on the mth sample vector $\mathbf{x}_m$ of the centred speech signal [18, 21]

$$\hat{\sigma}_r^2(\mathbf{x}) = M^{-1}\sum_{m=1}^{M} z_r^2(\mathbf{x}_m). \tag{11}$$

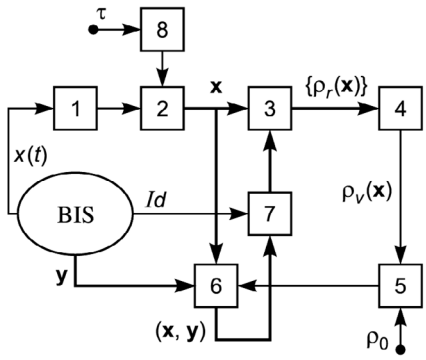

Fig. 1. Block diagram of the PBD updating device: *1*) vowel speech sounds selector; *2*) speech frame forming unit; *3*) partial quality indicator calculation unit; *4*) minimum selection unit; *5*) threshold unit; *6*) multichannel key; *7*) PBD storage; *8*) clock pulse generator.

Let us supplement Eq. (11) with the asymptotic equality [20]

$$n^{-1}\ln|K_r|\Big|_{n\to\infty} = \ln\sigma_r^2,$$

and its two statistical images [15]

$$n^{-1}\mathrm{tr}(SK_r^{-1})\Big|_{n\to\infty} = \frac{\hat{\sigma}_r^2(\mathbf{x})}{\sigma_r^2}; \qquad n^{-1}\ln|S|\Big|_{n\to\infty} = \ln\hat{\sigma}_x^2(\mathbf{x}),$$

where $\hat{\sigma}_x^2(\mathbf{x}) = (\mathbf{e}^{\mathrm{T}}S^{-1}\mathbf{e})^{-1}$ is the empirical dispersion of the speech signal at the output of the adaptive whitening filter, tuned to the input signal $x(t)$ based on a sample of observations $\mathbf{x}$ in the "sliding window" mode of one speech frame length τ. Taking into account these equalities, let us write, according to Eq. (8), the expression

$$\rho_r^*(\mathbf{x}) = 0.5\ln\left[\frac{\sigma_r^2}{\hat{\sigma}_x^2(\mathbf{x})}\times\frac{\hat{\sigma}_r^2(\mathbf{x})}{\sigma_r^2}\right] = 0.5\ln\left[\frac{\hat{\sigma}_r^2(\mathbf{x})}{\hat{\sigma}_x^2(\mathbf{x})}\right]. \tag{12}$$

The obtained result together with Eqs. (10), (11) determines the method for measuring the partial (rth) indicator of Eq. (8) of the PBD quality with the property of scale invariance of Eq. (6) with respect to the speech signal $x(t)$. The computational complexity of the proposed method consists mainly of costs of the n^3 order for the inverting operation of the $(n\times n)$ matrix S [21], which is quite acceptable for its application in the real time soft mode (with a delay equal to the signal duration T [14]).

PBD updating device. The block diagram of a device that implements the proposed method is shown in Fig. 1. The device contains series-connected elements: vowel speech sounds selector *1*; speech frame x forming unit *2*; partial quality indicator calculation unit *3* of Eq. (12); minimum selection unit *4*; threshold unit *5*; multichannel key *6*; PBD storage *7*; clock pulse generator *8* with a period τ. Double arrows indicate multidimensional (vector) functional connections. Adjustable parameters of the device are the frame duration τ and the threshold level ρ_0.

The PBD updating procedure is triggered by the BIS signal about the successful identification *Id* of a registered user. The PBD is updated automatically by adding the observed frame $\mathbf{x}$ to the storage at the address of the given user, provided that the frame does not meet the criterion of Eq. (3). Moreover, the previously collected PBDs are also stored in the storage, at least until the expiration of the revision period established in the BIS, as they have not completely lost their relevance for user identification. The significance level of decisions made according to criterion of Eq. (3) is adjusted by the observer using the threshold level ρ_0.

Adjustable significance level. In accordance with Eq. (12), let us write down the expression for the upper boundary of the rth quality indicator

$$\rho_r^*(\mathbf{x}) = 0.5\ln\left[\frac{\hat{\sigma}_r^2(\mathbf{x})}{\hat{\sigma}_x^2(\mathbf{x})}\right] \le 0.5\left[\frac{\hat{\sigma}_r^2(\mathbf{x})}{\hat{\sigma}_x^2(\mathbf{x})} - 1\right] \triangleq \sup\rho_r^*(\mathbf{x}).$$

This boundary, in accordance with Eq. (5), determines the guaranteed significance level of the decisions made:

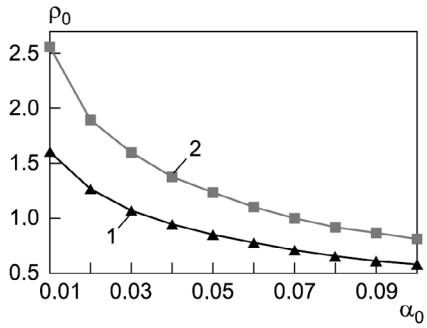

Fig. 2. Dependence of the threshold level ρ_0 of Eq. (15) on the significance level α_0 at $M = 12, 8$ (curves *1*, *2*, respectively).

$$\alpha(\rho_0) = P\{\rho^*_\nu(x) > \rho_0 \,|H_\nu\} \le P\left\{\frac{\hat\sigma^2_\nu(\mathbf{x})}{\hat\sigma^2_x(\mathbf{x})} > 1 + 2\rho_0 \,\middle|\, H_\nu\right\}. \tag{13}$$

If the hypothesis H_ν is valid, both dispersions in Eq. (13) are calculated using the formula for the mean square of a random Gaussian quantity Norm(K_ν). Therefore, similar as in [14], let us describe the indicated dispersions by two χ^2-distributions (Pearson) with M degrees of freedom each [18]. Assuming the statistical independence of two χ^2-quantities [10, 15], let us write the expression

$$\alpha(\rho_0) \le P\left\{\frac{\chi^2_1(M)}{\chi^2_2(M)} > 1 + 2\rho_0\right\} = 1 - \Phi_{M,M}(1 + 2\rho_0), \tag{14}$$

where $\Phi_{M,M}(\cdot)$ is the integral function of the Fischer F-distribution with a two-dimensional number (M, M) of degrees of freedom.

The values of the integral function are tabulated in detail in [21], including in electronic form. It should be especially noted that the right side of Eq. (14) does not depend on the phoneme number ν; therefore, it can be applied to all R controlled phonemes of the registered user. Equating the right side of Eq. (14) to the given significance level α_0, one obtains, in accordance with Eq. (5), the equation $\alpha(\rho_0) = \alpha_0$, from the solution of which the threshold level can be found as

$$\rho_0(\alpha_0) = 0.5[\Phi^{-1}_{M,M}(1 - \alpha_0) - 1] \tag{15}$$

for substituting to the right side of criterion of Eq. (3). The result obtained explicitly determines the value of ρ_0 through the significance level α_0.

Using Excel spreadsheets, dependences of Eq. (15) are plotted in Fig. 2 for two different values of the parameter M. In both cases, the threshold level ρ_0 monotonically increases with decreasing significance level α_0. This is a sign of a decrease in the observer's requirements for the acoustic quality of voice templates in the BIS storage. For example, when $\alpha_0 = 0.05$, 0.01 and $M = 12$, $\rho_0 = 0.84$, 1.61, respectively. Let us note that the inter-phonemic information mismatch of Eq. (4) within the set of vowel phonemes exceeds this threshold by more than an order of magnitude [4, 14].

Thus, the proposed technique of Eqs. (10)–(12) for measuring the quality indicator of Eq. (8) of the registered user's voice templates provides the observer with the ability to adjust the decisions $\overline{W}(\mathbf{x})$ about PBD updating made according to criterion of Eq. (3) in a wide range of values of the significance level α_0. This possibility is illustrated below by the results of the experiment carried out by the authors.

Experimental procedure and its results. The object of the experimental study was the speech signal $x(t)$ of one of the authors of this article in the role of a control speaker. This signal contained a conditional voice password in the format of the Unified Biometric System [3] in the form of a sequence of ten words with the names of Arabic numerals 0, 1, ..., 9 in ascending order. Such a sequence contains five of the six vowel sounds of Russian speech in stressed syllables. To select them from the speech signal, similar as in [4], a standard amplitude selector was used at the input of the device (see Fig. 1). Each selected sound was compared with voice templates of signals $x_r(t)$ of six vowel phonemes recorded by the same speaker in 2019.

The research subject is the above-proposed technique for automatic online PBD updating using the information quality indicator of Eq. (8). The technical implementation of the technique for the given values of the parameters $F = 8$ kHz, $\tau = 30$ ms, $N = 240$, and $n = 20$ was carried out in software form based on the authoring software "Phoneme Training." The

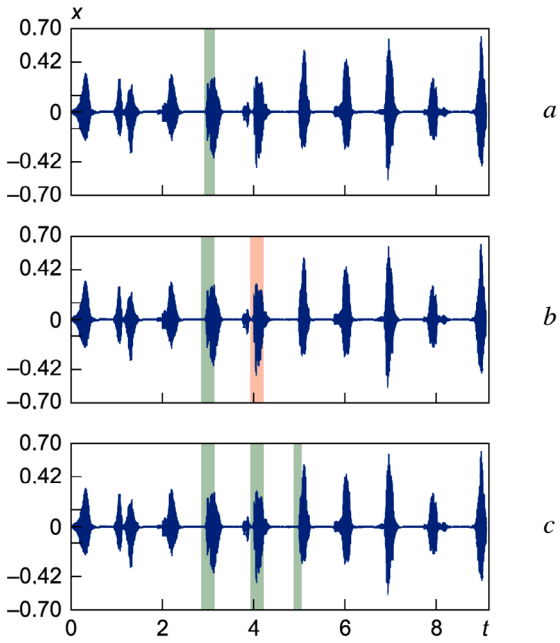

Fig. 3. Timing diagrams of a speech signal at threshold values $\rho_0 = 1.0, 0.8, 0.6$ (*a*–*c*, respectively).

specified software is posted on the authors' website in free access (Information System for Phonetic Analysis and Speech Training, https://sites.google.com/site/frompldcreators/produkty-1/phonemetraining). The software interface is described in detail in [4, 10].

At the stage of preparing the experiment on recording the signals of vowel phonemes $x_r(t)$, the necessary phonetic database of a conventional user was created in the software in the form of a set $\{K_r\}$ of ACMs of his voice templates. At that, the classical formula of the empirical correlation factor was used [18, 21]. At the same time, vectors of the coefficients R of whitening filters of Eq. (10) were calculated by inverting the corresponding ACMs using the high-speed Levinson–Durbin computational procedure [20]. Then the software was switched to operating mode. The main content of the operating mode is the calculation according to Eq. (11) of the empirical dispersions $\hat{\sigma}_r^2(\mathbf{x})$, $\hat{\sigma}_x^2(\mathbf{x})$ before their substitution into Eq. (12). The result of calculations was recorded in the operating window "Segmentation." The value of the threshold level ρ_0 in criterion of Eq. (3) was then adjusted by the observer using the "Parameters" tab in the software menu.

The results obtained are shown in Fig. 3 in the form of three timing diagrams for three different threshold values ρ_0; the coloured background marks the segments of vowel speech sounds, selected for PBD updating according to criterion of Eq. (3). The darker the background colour, the higher the degree of informational mismatch of Eq. (8) of the current frame $\mathbf{x}$ in relation to its same-name (same-phoneme) voice template from the phonetic database of a conventional user.

Comparing the timing diagrams with each other, one can conclude that the inversely proportional relationship of Eq. (15) between the significance level of Eq. (6) of the decisions made on PBD updating according to criterion of Eq. (3), on the one hand, and the threshold level ρ_0 set in it, on the other hand, is confirmed. This conclusion is of obvious practical importance from the point of view of possible adjustment by a conventional observer of the reliability of the used criterion of Eq. (3), taking into account the features of the problems solved by the BIS in each specific case.

Conclusion. A technique for automatic PBD updating in the BIS storage by the speech signals of registered users with voice requests for identification and online servicing has been proposed. In contrast to the technique described by the authors in the previous work [4], the proposed technique uses a scale-invariant indicator of the acoustic quality of voice templates, which provides the observer with guaranteed reliability of the decisions made in a wide dynamic range of the speech signal at the BIS input.

The results obtained will be useful in the development of new and upgrading of existing systems and technologies for automated PBD quality control and updating.

REFERENCES

1. S. Kumer, V. K. Lamba, and S. Jangra, "EAgeBioS: Enhanced Biometric System to handle the Effects of Template Aging," *Int. J. Innov. Technol. Explor. Eng.*, **9**, No. 11, 3669–3677 (2019), https://doi.org/10.35940/ijitee.A4756.119119.
2. I. Manjani, H. Sumerkan, P. J. Flynn, and K. W. Bowyer, "Template aging in 3D and 2D face recognition," *2016 IEEE 8th Int. Conf. on Biometrics Theory, Applications and Systems* (2016), https://doi.org/10.1109/BTAS.2016.7791202.
3. V. V. Savchenko and A. V. Savchenko, "Method of measuring the acoustic quality indicator of audio recordings prepared for registration and processing in the Unified Biometric System," *Izmer. Tekhn.*, No. 12, 40–46 (2019), https://doi.org/10.32446/0368-1025it.2019-12-40-46.
4. V. V. Savchenko and A. V. Savchenko, "Method of real-time updating of voice templates in the Unified Biometric System," *Izmer. Tekhn.*, No. 5, 58–65 (2020), https://doi.org/10.32446/0368-1025it.2020-5-58-65.
5. M. Smallman, "Why voice is getting stronger in financial services," *Biometric Technol. Today*, **2017**, No. 1, 5–7 (2017), https://doi.org/10.1016/S0969-4765(17)30013-9.
6. N. Crosswhite, J. Byrne, et al., "Template adaptation for face verification and identification," *Image Vision Comput.*, **79**, 35–48 (2018), https://doi.org/10.1016/j.imavis.2018.09.002.
7. G. Orrù, G. L. Marcialis, and F. Roli, "A novel classification-selection approach for the self updating of template-based face recognition systems," *Pattern Recogn.*, **100**, 107121 (2020), https://doi.org/10.1016/j.patcog.2019.107121.
8. M. Singh, R. Singh, and A. Ross, "A comprehensive overview of biometric fusion," *Inform. Fusion*, **52**, No. 12, 187–205 (2019), https://doi.org/10.1016/j.inffus.2018.12.003.
9. N. N. Lebedeva and E. D. Karimova, "Acoustic characteristics of the speech signal as an indicator of the functional state of a person," *Usp. Fiziol. Nauk*, **45**, No. 1, 57–95 (2014).
10. V. V. Savchenko and A. V. Savchenko, "Method for measuring distortion of a speech signal during its transmission over a communication channel to a biometric identification system," *Izmer. Tekhn.*, No. 11, 65–72 (2020), https://doi.org/10.32446/0368-1025it.2020-11-65-72.
11. A. V. Savchenko, V. V. Savchenko, and L. V. Savchenko, "Optimization of Gain in Symmetrized Itakura-Saito Discrimination for Pronunciation Learning," in: *Mathematical Optimization Theory and Operations Research,* Springer, Cham (2020), pp. 440–454, https://doi.org/10.1007/978-3-030-49988-4_30.
12. S. Kullback, *Information Theory and Statistics*, Dover Publ., New York (1997).
13. V. V. Savchenko, "Itakura-Saito divergence as an element of the information theory of speech perception," *J. Commun. Technol. El.*, **64**, No. 6, 590–596 (2019), https://doi.org/10.1134/S1064226919060093.
14. V. V. Savchenko and L. V. Savchenko, "Method for measuring the speech signal intelligibility indicator in the Kullback–Leibler information metric," *Izmer. Tekhn.*, No. 9, 59–64 (2019), https://doi.org/10.32446/0368-1025it.2019-9-59-64.
15. V. V. Savchenko and A. V. Savchenko, "Guaranteed significance level criterion in automatic speech signal segmentation," *J. Commun. Technol. El.*, **65**, No. 11, 1311–1317 (2020), https://doi.org/10.1134/S1064226920110157.
16. H. B. Kashani, A. Sayadiyan, and H. Sheikhzadeh, "Vowel detection using a perceptually-enhanced spectrum matching conditioned to phonetic context and speaker identity," *Speech Commun.*, **91**, 28–48 (2017), https://doi.org/10.1016/j.specom.2017.04.008.
17. A. V. Savchenko, "Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet," *PeerJ Comput. Sc.*, **5e**, 197 (2019), https://doi.org/10.7717/peerj-cs.197.
18. A. A. Borovkov, *Mathematical Statistics*, Lan', St. Petersburg (2010).
19. Z. Meng, M. U. B. Altaf, and B. H. F. Juang, "Active voice authentication," *Digit. Signal Process.*, **101**, 102672 (2020), https://doi.org/10.1016/j.dsp.2020.102672.
20. S. L. Marple, *Digital Spectral Analysis with Applications*, Dover Publ., Mineola, New York (2019), 2nd ed., https://www.goodreads.com/book/show/19484239.
21. P. H. Müller, P. Neumann, and R. Storm, *Tables for Mathematical Statistics*, VEB Fachbuchverlag, Leipzig (1973), https://doi.org/10.1002/bimj.19740160816.