

Способ автоматического обновления биометрических персональных данных в режиме онлайн по речевому сигналу пользователя биометрической системы

Андрей Владимирович Савченко¹, Владимир Васильевич Савченко²✉

¹ Национальный исследовательский университет «Высшая школа экономики», Нижний Новгород, Россия, avsavchenko@hse.ru, <https://orcid.org/0000-0001-6196-0564>

² Нижегородский государственный лингвистический университет, Нижний Новгород, Россия, vvsavchenko@yandex.ru✉, <https://orcid.org/0000-0003-3045-3337>

Аннотация. Рассмотрена проблема устаревания биометрических персональных данных и предложен способ её решения путём автоматического обновления указанных данных в хранилище биометрической системы. Обновление происходит по речевым сигналам зарегистрированных пользователей с запросами на их идентификацию и онлайн-обслуживание, при этом используется масштабно-инвариантный показатель качества голосовых образцов. Установлено, что применение масштабно-инвариантного показателя обеспечивает необходимый уровень значимости решений, принимаемых условным наблюдателем. С применением авторского программного обеспечения поставлен и проведён натурный эксперимент, дано практическое обоснование эффективности предложенного способа. Полученные результаты будут полезны при разработке новых и модернизации существующих систем и технологий автоматизированного контроля качества и обновления биометрических персональных данных.

Ключевые слова: акустика речи, акустические измерения, речевой сигнал, биометрия, биометрическая система, биометрические персональные данные

Финансирование: исследование выполнено при финансовой поддержке Российского научного фонда (проект № 20-71-10010).

Для цитирования: Савченко А. В., Савченко В. В. Способ автоматического обновления биометрических персональных данных в режиме онлайн по речевому сигналу пользователя биометрической системы // Измерительная техника. 2021. № 11. С. 60–66. <https://doi.org/10.32446/0368-1025it.2021-11-60-66>

Method for automatic on-line updating of personal biometric data based on speech signal of the biometric system user

Andrey V. Savchenko¹, Vladimir V. Savchenko²✉

¹ National Research University Higher School of Economics, Nizhny Novgorod, Russia, avsavchenko@hse.ru, <https://orcid.org/0000-0001-6196-0564>

² Nizhny Novgorod State Linguistic University, Nizhny Novgorod, Russia, vvsavchenko@yandex.ru✉, <https://orcid.org/0000-0003-3045-3337>

Abstract. The problem of “aging” of personal biometric data over time is considered. A method is proposed to overcome it by automatically updating the personal biometric data by 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 consequence, it is characterized by guaranteed reliability of the decisions made in the conditions of a wide speech signal dynamic range. It is shown that a guaranteed level of significance of decisions made by an observer is provided. The proposed method is implemented in a special software, which was used to experimentally estimate its effectiveness with real speech data. The obtained results are intended for use in the development of new and modernization of existing software and technologies for the personal biometric data automated quality control and updating.

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

Введение. В настоящее время в самых разных отраслях экономики и управления широко применяются биометрические системы идентификации (БСИ). Примером может служить созданная в 2018 г. по инициативе Банка России и Минкомсвязи РФ общенациональная программная платформа «Единая биометрическая система» (Единая биометрическая система:

[сайт]. URL: <https://bio.rt.ru>). Её основное назначение – дистанционная (в режиме онлайн) идентификация граждан РФ на основе предварительно собранных биометрических персональных данных (БПД). На первом этапе внедрения Единой биометрической системы в экономику страны приоритет отдан банковской сфере, в которой потребности в дистанционном

способе оказания услуг населению, в частности, по голосовым запросам пользователей в режиме онлайн, существуют давно и в наиболее острой форме. Специально для банков страны оператор Единой биометрической системы – Рос-телеком – создал мобильное приложение «Биометрия». С июля 2018 г. граждане РФ через все крупные банки страны получили возможность регистрировать БПД, причём в бимодальном варианте: фотография лица и запись голоса. В сборе БПД по проекту Единой биометрической системы участвует более 200 российских банков. Однако далеко не все проблемы в этом актуальном направлении на данный момент успешно решены.

Одной из наиболее острых проблем в биометрии является естественное устаревание БПД или частичная утрата их потребительских свойств с течением времени [1–3]. Её решение предполагает периодические обновления БПД в хранилище БСИ. Рекомендуемый период обновлений голосовых образцов находится в интервале от одного года до нескольких лет в зависимости от голосовых особенностей конкретного пользователя [4, 5]. Вариативность голоса пользователя сопровождается вариативностью речевой артикуляции, которая неизбежно проявляется в изменении черт (признаков) лица пользователя [6, 7]. В качестве частичного решения рассматриваемой проблемы можно предложить автоматическое обновление БПД в хранилище БСИ с учётом того, что скорость устаревания БПД зависит от физиологических особенностей личности пользователя [4], а речевой тракт – высокочувствительный индикатор таких особенностей [8]. Отсюда следует, что можно найти зависимость оценки качества БПД от акустических свойств и характеристик голосовых образцов зарегистрированных пользователей БСИ. По измеренным акустическим характеристикам голоса пользователя (по его речевому сигналу в режиме онлайн) можно судить о его текущем психофизическом состоянии [5, 9]. По существу, это новый подход к проблеме обновления БПД, при котором очередной момент их перезаписи для хранения в БСИ обуславливается лишь фактом нарушения установленных требований к акустическому качеству хранящихся голосовых образцов. При этом необходимо решить проблему масштабирования речевого сигнала по амплитуде с учётом широкого (десятки децибел) динамического диапазона на входе БСИ [3, 10].

Цель исследования – разработка способа автоматического обновления БПД в режиме онлайн по речевым сигналам пользователей БСИ с использованием оригинальной авторской модификации [11] информационной меры Кульбака-Лейблера [12], инвариантной к масштабу (коэффициенту усиления) речевого сигнала на выходе канала связи.

Постановка задачи. В теоретической части настоящего исследования задействован оригинальный авторский математический аппарат в рамках информационной теории восприятия речи [13, 14]. Пусть $x(t)$ – речевой сигнал конечной длительности T с запросом пользователя БСИ на обслуживание. Разобьём этот сигнал на конечную последовательность квазистационарных сегментов (фреймов) небольшой длительности $\tau = 20...30$ мс. Рассмотрим случай дискретного

времени $t=1, 2, \dots$ с периодом взятия отсчётов $T=F^{-1}$, где F – частота дискретизации речевого сигнала. Зададим его текущий (наблюдаемый) фрейм вектором отсчётов $\mathbf{x}$ размерностью $N=\tau F$. На основе распространённой [14, 15] в системах автоматической обработки речи многомерной (n -мерной) гауссовой модели речевого сигнала для установления его закона распределения $\text{Norm}(K)$ с автокорреляционной матрицей (АКМ) K рассмотрим задачу проверки двух статистических гипотез

$$\begin{aligned} H: K &\subset \{K_r\}; \\ \bar{H}: K &\not\subset \{K_r\}, \end{aligned} \quad (1)$$

где K_r – АКМ r -й ($r \leq R$) минимальной речевой единицы типа фонемы в составе хранящегося (актуального) голосового образца определённого пользователя; R – количество контролируемых фонем; чертой над символом H обозначена операция логического отрицания.

Системе (1) эквивалентно равенство

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

где H_r – r -я (парциальная) составляющая гипотезы H о равенстве двух АКМ:

$$H_r: K = K_r, \quad r = \overline{1, R}; \quad (2)$$

Согласно работам [3, 4] контролируемые в БСИ фонемами может быть множество гласных фонем как наиболее содержательных звуков речи пользователя в теоретико-информационном смысле [11–14] и выделяемых из речевого сигнала $x(t)$ по амплитудному признаку [4] или по признаку стабильной частоты основного тона [16]. В таком случае для русского языка имеем $R=6$. При этом каждая гласная фонема может быть задана в хранилище БСИ множеством $\{\mathbf{x}\}$ своих аллофонов из речевого сигнала зарегистрированного пользователя в качестве его допустимых голосовых образцов.

Предположим, что рассматриваемый пользователь зарегистрирован в БСИ под своим персональным идентификатором Id (в противном случае ему будет автоматически отказано в запросе на обслуживание), а его наблюдаемый фрейм $\mathbf{x}$ отнесён к числу R контролируемых (гласных) фонем. В таком случае задача (1) имеет два варианта решений: $W(\mathbf{x})$ и $\bar{W}(\mathbf{x})$. В первом случае текущий наблюдаемый фрейм $\mathbf{x}$ признаётся БСИ (т. е. условным наблюдателем) безупречным гласным звуком речи данного пользователя, во втором случае – пограничным (маргинальным) звуком речи. В таком случае вектор $\mathbf{x}$ направляется в БСИ в качестве разрешённого звука речи пользователя в дополнение к его собранным на данный момент голосовым образцам. Одновременно в хранилище БПД может быть направлен и новый вектор $\mathbf{y}$ характерных признаков его лица, полученный в ходе текущего обращения пользователя с использованием специальных средств [6, 17]. С течением времени

часть сохраненных биометрических образцов $\mathbf{x}$, $\mathbf{y}$ из числа отслуживших установленный срок (например, три года для Единой биометрической системы [4]) автоматически удаляется из хранилища. По существу, это означает обновление БПД в режиме on-line. Проверке (1) после этого подлежит очередной фрейм речевого сигнала $x(t)$, вплоть до последнего фрейма.

Какое именно решение на каждом очередном шаге примет наблюдатель зависит от используемого им критерия или правила принятия решений. В работе [4] предложено использовать в этих целях критерий допустимой (пороговой) величины

$$W(\mathbf{x}): (\exists r \leq R) \rho_r(\mathbf{x}) \leq \rho_0 \quad (3)$$

информационного рассогласования в метрике Кульбака-Лейблера [12]:

$$\rho_r(\mathbf{x}) \triangleq 0,5 \left[n^{-1} \text{tr}(SK_r^{-1}) - n^{-1} \ln |SK_r^{-1}| - 1 \right], \quad (4)$$

где $S \triangleq M^{-1} \sum_{m=1}^M \mathbf{x}_m \mathbf{x}_m^T$ – статистическая оценка АКМ централизованного речевого сигнала на интервале наблюдений в один речевой фрейм $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M)$ по выборке векторных наблюдений $\mathbf{x}_m$, $m \leq M$ конечного объёма $M = \lfloor Nn^{-1} \rfloor$; $\mathbf{x}_m$ – n -вектор (столбец) отсчётов речевого сигнала $x(t)$ в пределах его парциального (m -го) отрезка наблюдений длительностью $\tau_0 = \tau M^{-1}$ (символами $\text{tr}(\cdot)$ и $|\cdot|$ здесь и далее обозначены соответственно след и определитель квадратной $(n \times n)$ -матрицы; $[\cdot]$ – целая часть рационального числа; $\triangleq$ – равенство по определению; $\exists$ – квантор существования). Пороговый уровень ρ_0 в (3) устанавливается условным наблюдателем в зависимости от допустимой вероятности ошибки 1-го рода [15]:

$$\begin{aligned} \alpha &\triangleq P\{\bar{W}(\mathbf{x}) | H_0\} = P\{\min_{r \leq R} \rho_r(\mathbf{x}) > \rho_0 | H_0\} = \\ &= P\{\rho_v(\mathbf{x}) > \rho_0 | H_v\} \triangleq \alpha(\rho_0), \end{aligned} \quad (5)$$

где $v \leq R$. В математической статистике [18] вероятность (5) определяют как уровень значимости (надёжности) решений $\bar{W}(\mathbf{x})$ наблюдателя об обновления БПД. Чем выше порог ρ_0 , тем уровень значимости ниже.

Задача сводится к определению в явном виде зависимости $\alpha(\rho_0)$ в правой части выражения (5). Однако применительно к критерию (3) это практически неразрешимая задача. Проблема состоит в известной нестабильности, причём в широких пределах, масштаба или амплитуды речевого сигнала на входе БСИ [10, 19]. Учитывая, что используемые в (1) эталоны хранятся в виде R -множества АКМ голосовых образцов определённой (фиксированной) амплитуды, на практике необходимо соблюдать условие: при любой константе $c_r > 0$ в роли масштабного множителя должна выполняться система равенств

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

В противном случае критерий (3) утрачивает работоспособность, поскольку для любого уровня значимости α_0

согласно (5) необходимо соблюсти парадоксальное требование $\rho_0 \rightarrow \infty$ к пороговому уровню критерия (3) во всём диапазоне значений масштабного множителя $c_r \gg 1$. Для устранения данного препятствия модифицируем определение (4) информационного показателя качества голосовых образцов, наделив его свойством инвариантности (6) к масштабу речевого сигнала на входе.

Информационный показатель качества. В рамках общесистемного принципа минимума информационного рассогласования [12, 15] рассмотрим оптимизационную задачу поиска минимума величины $\rho_r(c_r \mathbf{x})$ по области определения масштабного множителя $c_r > 0$, где

$$\begin{aligned} \rho_r(c_r \mathbf{x}) &= 0,5 \left[n^{-1} c_r^2 \text{tr}(SK_r^{-1}) - n^{-1} \ln |c_r^2 SK_r^{-1}| - 1 \right] = \\ &= 0,5 \left[c_r^2 n^{-1} \text{tr}(SK_r^{-1}) - \ln c_r^2 + n^{-1} \ln |S^{-1}| - n^{-1} \ln |K_r^{-1}| - 1 \right]. \end{aligned} \quad (7)$$

Для этого в соответствии с методикой [11] сначала найдём первую производную от целевой функции $\rho_r(c_r) \triangleq \rho_r(c_r \mathbf{x})|_{\mathbf{x}}$:

$$\frac{d\rho_r(c_r)}{dc_r} = c_r n^{-1} \text{tr}(SK_r^{-1}) - c_r^{-1}.$$

Приравняем её нулю и получим оптимизационное уравнение

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

Найдём корень этого уравнения

$$c_r^* = \left[n^{-1} \text{tr}(SK_r^{-1}) \right]^{-0,5}$$

и подставим полученный результат в (7):

$$\begin{aligned} \rho_r(c_r^* \mathbf{x}) &= 0,5 \left[\ln(n^{-1} \text{tr}(SK_r^{-1})) + \right. \\ &\left. + n^{-1} \ln |S^{-1}| - n^{-1} \ln |K_r^{-1}| \right] \triangleq \rho_r^*(\mathbf{x}). \end{aligned} \quad (8)$$

Выражение (8) определяет модифицированный показатель качества r -го голосового образца как альтернативу (4) в роли оптимальной решающей статистики для критерия (3). Этот показатель обладает свойством масштабной инвариантности в формулировке (6):

$$\begin{aligned} \forall c_r > 0: \rho_r^*(c_r \mathbf{x}) &= 0,5 \left[\ln(n^{-1} \text{tr}(c_r^2 SK_r^{-1})) + n^{-1} \ln |c_r^2 S^{-1}| - \right. \\ &\left. - n^{-1} \ln |K_r^{-1}| \right] = 0,5 \ln \left[c_r^2 n^{-1} \text{tr}(SK_r^{-1}) |c_r^2 S^{-1}|^{1/n} |K_r^{-1}|^{-1/n} \right] = \\ &= 0,5 \ln \left[n^{-1} \text{tr}(SK_r^{-1}) \times |S^{-1}|^{1/n} |K_r^{-1}|^{-1/n} \right] = \\ &= 0,5 \left[\ln(n^{-1} \text{tr}(SK_r^{-1})) + n^{-1} \ln |S^{-1}| - n^{-1} \ln |K_r^{-1}| \right] = \rho_r^*(\mathbf{x}). \end{aligned}$$

Совокупность выражений (3), (4), (8) определяет способ обновления БПД по результатам обработки речевого сигнала в режиме онлайн. Предложенный способ можно дополнительно обосновать с практической точки зрения.

Учитывая быструю сходимость (со скоростью неуклучшаемого порядка $1/M$) статистических оценок АКМ по формуле

выборочного среднего [18], следует ожидать, что эмпирическое распределение $\text{Norm}(S)$ должно не сильно отличаться от своего эталона $\text{Norm}(K_r)$ при справедливости парциальной гипотезы (2) в условиях конечных (объёмом $M \gg 1$) выборок наблюдений. Например, при длительности речевого фрейма $\tau = 30$ мс, частоте дискретизации речевого сигнала $F = 8$ кГц и размерности его распределения $n = 20$ (типичные значения параметров для алгоритмов автоматической обработки речи [3, 4]) размерность фрейма будет $N = 30 \cdot 8 = 240$ и, следовательно, имеем $M = 240/20 = 12$ непересекающихся отрезков сигнала $x(t)$. При сокращении длительности фрейма до $\tau = 20$ мс получим $M = 20 \cdot 8/20 = 8$ отрезков речевого сигнала, что тоже немало.

Метод измерений. Раскроем выражение (8) в явном виде через N -вектор $\mathbf{x}$ отсчётов речевого сигнала на интервале наблюдений длительностью в один фрейм с использованием авторегрессионной (АР) модели r -й фонемы пользователя в качестве его одноимённого голосового образца [13]:

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

где через $a_r(i)$ обозначен i -й коэффициент ($i \leq p$) авторегрессии p -го порядка; $\{\eta_r(t)\}$ – порождающий процесс типа белого гауссова шума в дискретном времени t ; $\sigma_r^2 = \text{const}$ – дисперсия $\{\eta_r(t)\}$.

Дисперсия σ_r^2 здесь определяет минимально достижимое значение среднего квадрата погрешности линейного предсказания случайного временного ряда (9) на один шаг вперёд [14, 20]. При условии $p < n$ эта дисперсия равна обратной величине первого элемента одноимённой (r -й) обратной АКМ [15]:

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

Здесь $\mathbf{e} = (1, 0, \dots, 0)^T$ – индикаторный n -вектор-столбец, составленный из одних нулей, за исключением единицы на первой позиции. Аналогичным образом может быть определён и соответствующий вектор АР-коэффициентов $\mathbf{a}_r$:

$$(1; -\mathbf{a}_r^T)^T = \sigma_r^2 K_r^{-1} \mathbf{e} = \frac{K_r^{-1} \mathbf{e}}{\mathbf{e}^T K_r^{-1} \mathbf{e}} \triangleq \mathbf{b}_r,$$

состоящий из взятых с коэффициентом σ_r^2 элементов первого столбца одноимённой обратной АКМ, исключая её первый элемент. Здесь $\mathbf{b}_r$ – вектор коэффициентов линейного обеляющего фильтра, настроенного (на этапе подготовки данных) на сигнал r -й фонемы $x_r(t)$. Обеляющий фильтр – ключевое звено предлагаемого метода измерений. Его порядок равен $p = n - 1$. Динамика фильтра описывается инверсным по отношению к (9) выражением [15]:

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

Дисперсия $\sigma_r^2(\mathbf{x}) \triangleq \langle z_r^2(t) \rangle$ выходного сигнала $z_r(t)$ ($\langle \cdot \rangle$ – математическое ожидание случайной величины) отвечает соотношению общего вида $\sigma_r^2(\mathbf{x}) \geq \sigma_r^2$. Строгое равенство здесь

достигается лишь в асимптотике (при $N \rightarrow \infty$), когда на вход обеляющего фильтра (10) поступает сигнал (9) одноимённой фонемы. Эмпирическая (по выборке) оценка данной дисперсии определяется по формуле выборочного среднего квадрата отклика $z_r(\mathbf{x}_m) = \mathbf{b}_r^T \mathbf{x}_m$ r -го обеляющего фильтра (10) на m -й вектор отсчётов $\mathbf{x}_m$ централизованного речевого сигнала [18, 21]:

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

Дополним выражение (11) асимптотическим равенством [20]:

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

а также его двумя статистическими аналогами [15]:

$$n^{-1} \text{tr}(S K_r^{-1}) \Big|_{n \rightarrow \infty} = \frac{\hat{\sigma}_r^2(\mathbf{x})}{\sigma_r^2}, \quad n^{-1} \ln |S| \Big|_{n \rightarrow \infty} = \ln \hat{\sigma}_r^2(\mathbf{x}),$$

где $\hat{\sigma}_r^2(\mathbf{x}) = (\mathbf{e}^T S^{-1} \mathbf{e})^{-1}$ – эмпирическая дисперсия речевого сигнала на выходе адаптивного обеляющего фильтра, настроенного на входной сигнал $x(t)$ по выборке наблюдений $\mathbf{x}$ в режиме «скользящего окна» длиной τ в один речевой фрейм. При учёте данных равенств запишем согласно (8) выражение

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

Полученный результат совместно с (10), (11) определяет метод измерений парциального (r -го) показателя (8) качества БПД со свойством масштабной инвариантности (6) по отношению к речевому сигналу $x(t)$. Вычислительная сложность предложенного метода состоит главным образом из затрат порядка n^3 на операцию обращения $(n \times n)$ -матрицы S [21], что вполне приемлемо для его применения в режиме мягкого (с задержкой на длительность T сигнала [14]) реального времени.

Устройство обновления БПД. Структурная схема устройства, реализующего предложенный метод, представлена на рис. 1. Устройство содержит последовательно соединённые элементы: селектор 1 гласных звуков речи; формирователь 2 речевого фрейма $\mathbf{x}$; вычислитель 3 парциальных показателей качества (12); блок 4 выбора минимума; пороговый элемент 5; многоканальный ключ 6; хранилище БПД 7; генератор 8 тактовых импульсов с периодом τ . Двойными стрелками обозначены многомерные (векторные) функциональные связи. Регулируемые параметры устройства – длительность фрейма τ и пороговый уровень ρ_0 .

Процедура обновления БПД запускается в действие по сигналу БСИ об успешной идентификации Id зарегистрированного пользователя. Обновление БПД осуществляется автоматически путём добавления в хранилище по адресу данного пользователя – наблюдаемого фрейма $\mathbf{x}$ при условии, что для него не выполнен критерий (3). Причём ранее собранные БПД в хранилище также сохраняются, по крайней мере до истечения установленного в БСИ контрольного срока, как не вполне утратившие свою актуальность для

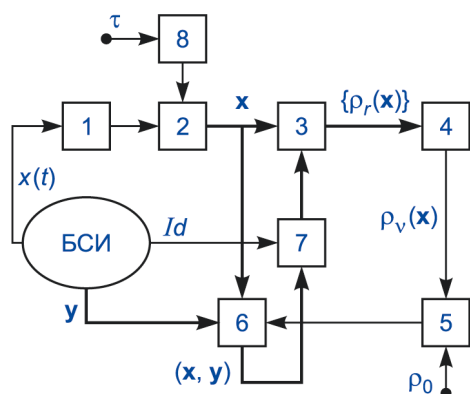


Рис. 1. Структурная схема устройства обновления БПД:

1 – селектор гласных звуков речи; 2 – формирователь речевого фрейма; 3 – вычислитель парциальных показателей качества; 4 – блок выбора минимума; 5 – пороговый элемент; 6 – многоканальный ключ; 7 – хранилище БПД; 8 – генератор тактовых импульсов

Fig. 1. Block diagram of the device for updating of PBD (personal biometric data):

1 – vowel sounds selector; 2 – frame forming unit; 3 – partial quality indicator calculation unit; 4 – minimum selection unit; 5 – threshold unit; 6 – multi-channel key; 7 – PBD storage; 8 – clock pulse generator

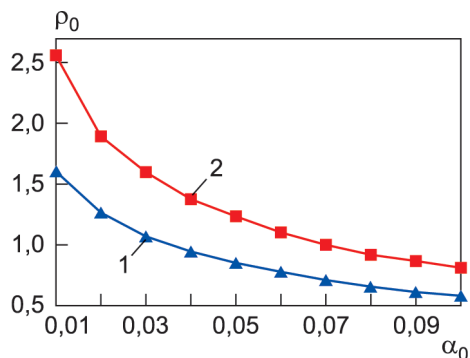

 Рис. 2. Графики зависимости порогового уровня ρ_0 (15) от уровня значимости α_0 при $M=12$; 8 (кривые 1, 2 соответственно)

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

идентификации пользователя. Уровень значимости принимаемых согласно критерию (3) решений регулируется наблюдателем с помощью порогового уровня ρ_0 .

Регулируемый уровень значимости. В соответствии с (12) запишем выражение для верхней границы r -го показателя качества

$$\rho_r^*(\mathbf{x}) = 0,5 \ln \left[\frac{\hat{\sigma}_r^2(\mathbf{x})}{\hat{\sigma}_x^2(\mathbf{x})} \right] \leq 0,5 \left[\frac{\hat{\sigma}_r^2(\mathbf{x})}{\hat{\sigma}_x^2(\mathbf{x})} - 1 \right] \triangleq \sup_r^*(\mathbf{x}).$$

Данная граница в соответствии с (5) определяет гарантированный уровень значимости принимаемых решений:

$$\alpha(\rho_0) = P\{\rho_v^*(\mathbf{x}) > \rho_0 | H_v\} \leq P\left\{ \frac{\hat{\sigma}_v^2(\mathbf{x})}{\hat{\sigma}_x^2(\mathbf{x})} > 1 + 2\rho_0 \middle| H_v \right\}. \quad (13)$$

При справедливости гипотезы H_v обе дисперсии в (13) рассчитываются по формуле среднего квадрата случайной гауссовой величины $\text{Norm}(K_v)$. Поэтому по аналогии с работой [14] опишем указанные дисперсии двумя χ^2 -распределениями (Пирсона) с M степенями свободы каждое [18]. В предположении о статистической независимости двух χ^2 -величин [10, 15] запишем выражение:

$$\alpha(\rho_0) \leq P\left\{ \frac{\chi_1^2(M)}{\chi_2^2(M)} > 1 + 2\rho_0 \right\} = 1 - \Phi_{M,M}(1 + 2\rho_0), \quad (14)$$

где $\Phi_{M,M}(\cdot)$ – интегральная функция F -распределения Фишера с двумерным числом (M, M) степеней свободы.

Значения интегральной функции подробно табулированы [21], в том числе в электронном виде. Особо отметим, что правая часть выражения (14) не зависит от номера фонемы v и поэтому распространяется на все R контролируемых фонем зарегистрированного пользователя. Приравняв правую часть (14) заданному уровню значимости α_0 , получим в соответствии с (5) уравнение $\alpha(\rho_0) = \alpha_0$, из решения которого найдём пороговый уровень:

$$\rho_0(\alpha_0) = 0,5 \left[\Phi_{M,M}^{-1}(1 - \alpha_0) - 1 \right] \quad (15)$$

для подстановки в правую часть критерия (3). Полученный результат определяет в явном виде величину ρ_0 через уровень значимости α_0 .

С использованием электронных таблиц Excel на рис. 2 построены графики зависимости (15) для двух разных значений параметра M . В обоих случаях пороговый уровень ρ_0 монотонно возрастает с уменьшением уровня значимости α_0 . Это признак снижения требований наблюдателя к акустическому качеству голосовых образцов в хранилище БСИ. Например, при равенствах $\alpha_0 = 0,05; 0,01$ для случая $M=12$ будем иметь $\rho_0 = 0,84; 1,61$ соответственно. Отметим, что межфонемное информационное рассогласование (4) в пределах множества гласных фонем превышает данный порог более чем на порядок [4, 14].

Таким образом, предложенный метод измерений (10)–(12) показателя качества (8) голосовых образцов зарегистрированного пользователя предоставляет наблюдателю возможность регулировки в широком диапазоне значений уровня значимости α_0 принимаемых согласно критерию (3) решений $\hat{W}(\mathbf{x})$ об обновлении БПД. Указанная возможность проиллюстрирована далее результатами проведённого авторами эксперимента.

Программа и результаты эксперимента. Объектом экспериментального исследования служил речевой сигнал $x(t)$ одного из авторов настоящей статьи в роли контрольного диктора. Этот сигнал содержал условный голосовой пароль в формате Единой биометрической системы [3] – в виде последовательности из десяти слов с названиями арабских цифр 0, 1, ..., 9 в порядке их возрастания. Такая последовательность в ударных слогах содержит пять из шести гласных звуков русской речи. Для их выделения из речевого сигнала на входе устройства (см. рис. 1) по аналогии с работой [4]

использовался стандартный амплитудный селектор. Каждый выделенный звук сопоставлялся с голосовыми образцами сигналов $x_i(t)$ шести гласных фонем, записанными тем же диктором в 2019 г.

Предмет исследования – предложенный выше способ автоматического обновления БПД в режиме on-line с использованием информационного показателя качества (8). Техническая реализация способа при заданных значениях параметров $F=8$ кГц, $\tau=30$ мс, $N=240$ и $n=20$ осуществлена в программном виде на базе авторской компьютерной программы “Phoneme Training”. Указанная программа размещена на сайте авторов в режиме свободного доступа (Информационная система фонетического анализа и обучения речи: [сайт]. URL: <https://sites.google.com/site/frompldcreators/produkty-1/phonemetraining>). Интерфейс программы подробно описан в работах [4, 10].

На этапе подготовки эксперимента по записям сигналов гласных фонем $x_i(t)$ в программе была создана необходимая фонетическая база данных условного пользователя в виде набора $\{K_r\}$ АКМ его голосовых образцов. При этом применялась классическая формула эмпирического коэффициента корреляции [18, 21]. Одновременно путём обращения соответствующих АКМ с помощью высокоскоростной вычислительной процедуры Левинсона-Дурбина [20] рассчитывались векторы коэффициентов R обеслаивающих фильтров (10). Затем программа переводилась в рабочий режим. Основное содержание рабочего режима – вычисления по формуле (11) эмпирических дисперсий $\hat{\sigma}_r^2(\mathbf{x})$, $\hat{\sigma}_x^2(\mathbf{x})$ перед их подстановкой в выражение (12). Результат вычислений фиксировался в рабочем окне «Сегментирование». Значение порогового уровня ρ_0 в критерии (3) при этом регулировалось наблюдателем с использованием вкладки «Параметры» в меню компьютерной программы.

Полученные результаты представлены на рис. 3 в виде трёх временных диаграмм для трёх разных значений порога ρ_0 ; цветным фоном отмечены отрезки гласных звуков речи, выделенные для обновления БПД согласно критерию (3). Чем темнее цвет фона, тем выше степень информационного расхождения (8) текущего фрейма $\mathbf{x}$ по отношению к его одноимённому (однофонемному) голосовому образцу из фонетической базы данных условного пользователя.

Из сопоставления временных диаграмм друг с другом можно сделать вывод о подтверждении обратно пропорциональной зависимости (15) между уровнем значимости (6) принимаемых решений об обновлении БПД согласно критерию (3), с одной стороны, и установленного в нем порогового уровня ρ_0 – с другой. Сделанный вывод имеет очевидное практическое значение с точки зрения возможной регулировки условным наблюдателем надёжности используемого критерия (3) при учёте в каждом конкретном случае особенностей решаемых БСИ задач.

Заключение. Предложен способ автоматического обновления БПД в хранилище БСИ по речевым сигналам зарегистрированных пользователей с голосовыми запросами на идентификацию и онлайн-обслуживание. В отличие от способа, описанного авторами в предыдущей работе [4], в предложенном способе используется масштабно-инвариантный

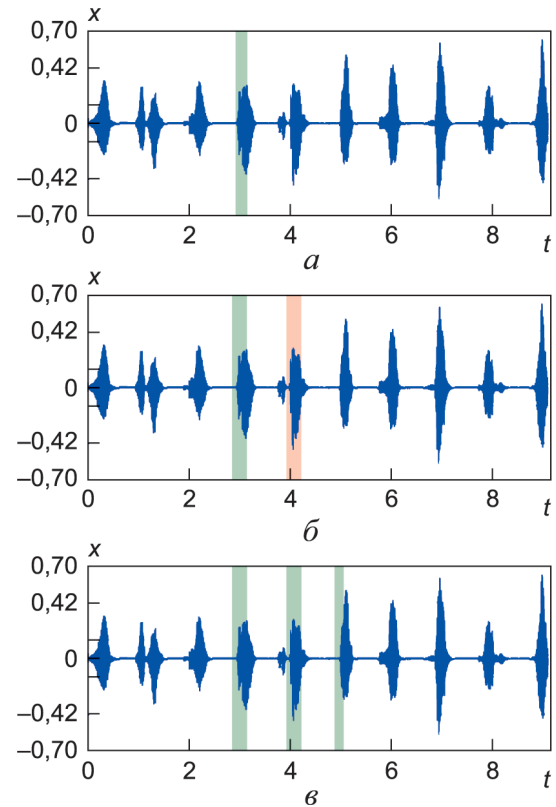


Рис. 3. Временные диаграммы речевого сигнала при значениях порога $\rho_0=1,0; 0,8; 0,6$ (а–в соответственно)

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

показатель акустического качества голосовых образцов, что обеспечивает наблюдателю гарантируемую надёжность принимаемых решений в условиях широкого динамического диапазона речевого сигнала на входе БСИ.

Полученные результаты будут полезными при разработке новых и модернизации существующих систем и технологий автоматизированного контроля качества и обновления БПД.

Список источников

1. Chawla S. K., Lamba V., Jangra S., *International Journal of Innovative Technology and Exploring Engineering*, 2019, vol. 9, no. 11, pp. 2278–3075. <https://doi.org/10.35940/ijitee.A4756.119119>
2. Manjani I. et al., *IEEE 8th International Conference on Biometrics Theory, Applications and Systems*, 2016. <https://doi.org/10.1109/BTAS.2016.7791202>
3. Савченко В. В., Савченко А. В. Метод измерений показателя акустического качества аудиозаписей, подготовленных для регистрации и обработки в Единой биометрической системе // *Измерительная техника*. 2019. № 12. С. 40–47. <https://doi.org/10.32446/0368-1025it.2019-12-40-46>
4. Савченко В. В., Савченко А. В. Способ обновления голосовых образцов в Единой биометрической системе в режиме реального времени // *Измерительная техника*. 2020. № 5. С. 58–65. <https://doi.org/10.32446/0368-1025it.2020-5-58-65>

5. Smallman M., *Biometric Technology Today*, 2017, vol. 2017, no. 1, pp. 5–7. [https://doi.org/10.1016/S0969-4765\(17\)30013-9](https://doi.org/10.1016/S0969-4765(17)30013-9)
6. Crosswhite N. et al., *Image and Vision Computing*, 2018, vol. 79, pp. 35–48. <https://doi.org/10.1016/j.imavis.2018.09.002>
7. Orrù G., Marcialis G. L., Roli F., *Pattern Recognition*, 2020, vol. 100, 107121. <https://doi.org/10.1016/j.patcog.2019.107121>
8. Singh M., Singh R., Ross A., *Information Fusion*, 2019, vol. 52, no. 12, pp. 187–205. <https://doi.org/10.1016/j.inffus.2018.12.003>
9. Лебедева Н. Н., Каримова Е. Д. Акустические характеристики речевого сигнала как показатель функционального состояния человека // Успехи физиологических наук. 2014. Т. 45. № 1. С. 57–95.
10. Савченко В. В., Савченко А. В. Метод измерения искажений речевого сигнала, переданного по каналу связи в биометрическую систему идентификации // Измерительная техника. 2020. № 11. С. 65–72. <https://doi.org/10.32446/0368-1025it.2020-11-65-72>
11. Savchenko A. V., Savchenko V. V., Savchenko L. V., *Mathematical Optimization Theory and Operations Research MOTOR 2020, Lecture Notes in Computer Science*, Springer, Cham, 2020, 12095. https://doi.org/10.1007/978-3-030-49988-4_30
12. Kullback S., *Information Theory and Statistics*, New York, Dover Publications, 1997, 432 p.
13. Savchenko V. V., *Journal of Communications Technology and Electronics*, 2019, vol. 64, no. 6, pp. 590–596. <https://doi.org/10.1134/S1064226919060093>
14. Савченко В. В., Савченко Л. В. Метод измерений показателя разборчивости речевого сигнала в информационной метрике Кульбака–Лейблера // Измерительная техника. 2019. № 9. С. 59–64. <https://doi.org/10.32446/0368-1025it.2019-9-59-64>
15. Savchenko V. V., Savchenko A. V., *Journal of Communications Technology and Electronics*, 2020, vol. 65, no. 11, pp. 1311–1317. <https://doi.org/10.1134/S1064226920110157>
16. Kashani H. B., Sayadiyan A., Sheikhzadeh H., *Speech Communication*, 2017, vol. 91, pp. 28–48. <https://doi.org/10.1016/j.specom.2017.04.008>
17. Savchenko A. V., *PeerJ Computer Science*, 2019, 5e:197. <https://doi.org/10.7717/peerj-cs.197>
18. Боровков А. А. *Математическая статистика*. Санкт-Петербург: Лань. 2010, 704 с.
19. Meng Zh., Umair Bin Altaf M., Juang B.-H., *Digital Signal Processing*, 2020, vol. 101, 102672. <https://doi.org/10.1016/j.dsp.2020.102672>
20. Marple S. L., *Digital Spectral Analysis with Applications*, 2nd ed., Mineola, New York, Dover Publications, 2019, 432 p. <https://www.goodreads.com/book/show/19484239>
21. Muller P. H., Neumann P., Storm R., *Tafeln der mathematischen statistik*, Leipzig, VEB Fachbuchverlag, 1973, 279 p. <https://doi.org/10.1002/bimj.19740160816>
2. Manjani I. et al., *IEEE 8th International Conference on Biometrics Theory, Applications and Systems*, 2016. <https://doi.org/10.1109/BTAS.2016.7791202>
3. Savchenko V. V., Savchenko A. V., *Measurement Techniques*, 2019, vol. 62, no. 12, pp. 1071–1078. <https://doi.org/10.1007/s11018-020-01736-w>
4. Savchenko V. V., Savchenko A. V., *Measurement Techniques*, 2020, vol. 63, no. 5, pp. 391–400. <https://doi.org/10.1007/s11018-020-01800-5>
5. Smallman M., *Biometric Technology Today*, 2017, vol. 2017, no. 1, pp. 5–7. [https://doi.org/10.1016/S0969-4765\(17\)30013-9](https://doi.org/10.1016/S0969-4765(17)30013-9)
6. Crosswhite N. et al., *Image and Vision Computing*, 2018, vol. 79, pp. 35–48. <https://doi.org/10.1016/j.imavis.2018.09.002>
7. Orrù G., Marcialis G. L., Roli F., *Pattern Recognition*, 2020, vol. 100, 107121. <https://doi.org/10.1016/j.patcog.2019.107121>
8. Singh M., Singh R., Ross A., *Information Fusion*, 2019, vol. 52, no. 12, pp. 187–205. <https://doi.org/10.1016/j.inffus.2018.12.003>
9. Lebedeva N. N., Karimova E. D., Akusticheskie kharakteristiki rechevogo signala kak pokazatel' funktsional'nogo sostoyaniya cheloveka, *Uspekhi fiziologicheskikh nauk*, 2014, vol. 45, no. 1, pp. 57–95. (In Russ.)
10. Savchenko V. V., Savchenko A. V., *Measurement Techniques*, 2021, vol. 63, no. 11, pp. 917–925. <https://doi.org/10.1007/s11018-021-01864-x>
11. Savchenko A. V., Savchenko V. V., Savchenko L. V., *Mathematical Optimization Theory and Operations Research MOTOR 2020, Lecture Notes in Computer Science*, Springer, Cham, 2020, 12095. https://doi.org/10.1007/978-3-030-49988-4_30
12. Kullback S., *Information Theory and Statistics*, New York, Dover Publications, 1997, 432 p.
13. Savchenko V. V., *Journal of Communications Technology and Electronics*, 2019, vol. 64, no. 6, pp. 590–596. <https://doi.org/10.1134/S1064226919060093>
14. Savchenko V. V., Savchenko L. V., *Measurement Techniques*, 2019, vol. 62, no. 9, pp. 832–839. <https://doi.org/10.1007/s11018-019-01702-1>
15. Savchenko V. V., Savchenko A. V., *Journal of Communications Technology and Electronics*, 2020, vol. 65, no. 11, pp. 1311–1317. <https://doi.org/10.1134/S1064226920110157>
16. Kashani H. B., Sayadiyan A., Sheikhzadeh H., *Speech Communication*, 2017, vol. 91, pp. 28–48. <https://doi.org/10.1016/j.specom.2017.04.008>
17. Savchenko A. V., *PeerJ Computer Science*, 2019, 5e:197. <https://doi.org/10.7717/peerj-cs.197>
18. Borovkov A. A. *Matematicheskaya statistika*. St. Petersburg, Publ., 2010, 704 p. (In Russ.)
19. Meng Zh., Umair Bin Altaf M., Juang B.-H., *Digital Signal Processing*, 2020, vol. 101, 102672. <https://doi.org/10.1016/j.dsp.2020.102672>
20. Marple S. L., *Digital Spectral Analysis with Applications*, 2nd ed., Mineola, New York, Dover Publications, 2019, 432 p. <https://www.goodreads.com/book/show/19484239>
21. Muller P. H., Neumann P., Storm R., *Tafeln der mathematischen statistik*, Leipzig, VEB Fachbuchverlag, 1973, 279 p. <https://doi.org/10.1002/bimj.19740160816>

References

Статья поступила в редакцию 09.06.2021; одобрена после рецензирования 13.07.2021; принята к публикации 20.07.2021.
The article was submitted 09.06.2021; approved after reviewing 13.07.2021; accepted for publication 20.07.2021.