

# Метод измерений показателя динамики эмоционального состояния пользователя по речевому сигналу в режиме реального времени

Людмила Васильевна Савченко<sup>1</sup>, Андрей Владимирович Савченко<sup>2</sup>✉

<sup>1</sup>Национальный исследовательский университет «Высшая школа экономики», Нижний Новгород, Россия

<sup>1</sup>lsavchenko@hse.ru, <sup>2</sup>avsavchenko@hse.ru✉

**Аннотация.** Рассмотрены проблемы внедрения систем с голосовым интерфейсом для дистанционного обслуживания населения. Эффективность таких систем может быть повышена за счёт автоматического анализа изменений эмоционального состояния пользователя в ходе диалога. Для измерений показателя динамики эмоционального состояния в режиме реального времени предложено использовать эффект звуковой (фонетической) вариативности речи пользователя на интервалах наблюдений небольшой (доли минуты) длительности. На основе теоретико-информационного подхода разработан метод акустических измерений динамики эмоционального состояния в условиях малых выборок с использованием масштабно-инвариантной меры вариаций речевого сигнала в частотной области. Рассмотрен пример практической реализации данного метода в режиме мягкого реального времени. Показано, что в данном случае задержка получения результатов измерений не превышает 10–20 с. Результаты экспериментальных исследований подтвердили высокое быстродействие предложенного метода и его чувствительность к изменениям динамики эмоционального состояния под действием внешних возмущений. Разработанный метод можно использовать при внедрении автоматизированного контроля качества голосовых образцов пользователей единых биометрических систем. Также метод будет полезен для повышения безопасности путём бесконтактного выявления потенциально опасных лиц с краткосрочным нарушением психоэмоционального состояния.

**Ключевые слова:** измерение динамики эмоционального состояния, акустические измерения, акустика речи, автоматическая обработка речи, проблема малых выборок, теоретико-информационный подход

**Для цитирования:** Савченко Л. В., Савченко А. В. Метод измерений показателя динамики эмоционального состояния пользователя по речевому сигналу в режиме реального времени // Измерительная техника. 2021. № 4. С. 49–57. <https://doi.org/10.32446/0368-1025it.2021-4-49-57>

## The method of real-time acoustic measurement of dynamical changes in the speaker's emotional state

Lyudmila V. Savchenko<sup>1</sup>, Andrey V. Savchenko<sup>2</sup>✉

<sup>1, 2</sup> HSE University, Nizhny Novgorod, Russia

<sup>1</sup>lsavchenko@hse.ru, <sup>2</sup>avsavchenko@hse.ru✉

**Abstract.** In this paper we consider the issues in implementations of interactive voice response systems with remote access. Their efficiency can be improved by automatically analyzing changes in the user's emotional state during the dialogue. In order to measure the indicator of the dynamics of the emotional state in real time, it is proposed to use the effect of sound (phonetic) variability of the user's speech at short intervals (fractions of a minute). The novel method of acoustic measurements in conditions of small samples has been developed based on information-theoretic approach by using a scale-invariant gain-optimized dissimilarity measure of the speech signals in the frequency domain. An example of its practical implementation in soft real time is considered. It is shown that the delay in obtaining the measurement results does not exceed in this case 10–20 sec. The experimental results confirmed the high speed of the proposed method and its sensitivity to changes in the emotional state under the influence of external noise. The proposed method can be used for automated quality control of voice samples of users in unified biometric systems, as well as to improve safety by non-contact identification of potentially dangerous persons with short-term psycho-emotional disorders.

**Keywords:** measurement of dynamics in emotional state, acoustic measurements, speech acoustics, automatic speech processing, small-sample-size problem, information-theoretic approach

**Введение.** В связи с бурным развитием информационно-коммуникационных технологий в мире широко распространяются диалоговые системы с голосовым интерфейсом для дистанционного обслуживания населения [1, 2]. Их эффективность в значительной мере зависит от динамики эмоционального состояния пользователей (ДЭС) информационных систем и оперативного выявления изменений

этого состояния в ходе диалога в режиме реального времени [3, 4]. Поэтому актуальными являются измерения показателя ДЭС по речевому сигналу  $x(t)$  конечной длительности  $T_x < \infty$ . Исследования в этом научном направлении осуществляются на протяжении ряда последних лет, однако достигнутой точности измерений показателя ДЭС всё ещё недостаточно для широкого практического

применения систем дистанционного обслуживания населения в режиме речевого диалога [5–8]. Причина не столько в сложности рассматриваемой задачи, сколько в отсутствии до настоящего времени эффективного математического аппарата для её решения. Однако это не относится к автоматической обработке речи [9–12], основанной на математическом аппарате многослойных искусственных нейронных сетей с настройкой под объект наблюдений по технологии глубокого обучения [13]. В данном направлении достигнуты впечатляющие коммерческие результаты – программные разработки Apple (Siri), Google (Voice Search), Microsoft (Cortana) и других мировых производителей, включающие в себя элементы искусственного интеллекта [14, 15]. Однако существует ряд острых проблем, создающих серьёзные препятствия для действительно широкого применения современных речевых технологий [16]. Главная из них – многоресурсность (в смысле повышенных требований к вычислительной мощности используемого оборудования) глубоких нейронных сетей, порождающая проблему быстродействия и, как следствие, практической реализуемости сетевых алгоритмов автоматической обработки речи в задаче измерений показателя ДЭСП в режиме реального времени [17]. С этой главной проблемой тесно связаны проблемы информационной безопасности [18] и малых выборок наблюдений [19, 20], особенно актуальные для систем дистанционного обслуживания пользователей в режиме диалога. Именно в этом направлении исследований сосредоточены в настоящее время усилия большинства авторитетных компаний – разработчиков речевых технологий [1, 12].

Между тем, острота проблемы быстродействия автоматической обработки речи может быть сильно снижена при переходе к автономным (без выхода в сеть Интернет) алгоритмам с настройкой на голос каждого отдельного пользователя [20–22]. Принцип действия таких алгоритмов базируется на фонетическом анализе речи с последовательным членением речевого сигнала на фреймы небольшой ( $\tau=10\ldots15$  мс) длительности, не выходящей за пределы длительности минимальных речевых единиц (МРЕ) типа фонем и их аллофонов в устной речи диктора [2, 12]. В них роль многообъёмных (десятки и сотни гигабайт) речевых баз данных выполняют персональные фонетические базы данных, которые формируются на этапе подготовки данных по известной методике [2, 3]. Исследованию в этом актуальном направлении для акустических измерений показателя ДЭСП и посвящена настоящая статья.

Цель работы – определить объективный информационный показатель ДЭСП и разработать метод измерений этого показателя в режиме мягкого реального времени [23]. При учёте известной инерционности эмоционального состояния пользователя во времени [10, 11] задержка результата измерений информационного показателя ДЭСП не должна выходить за пределы долей минуты. Для достижения поставленной цели авторами настоящей статьи применён оригинальный математический аппарат, созданный в рамках информационной теории восприятия речи [24–27].

**Постановка задачи.** В момент появления сбоев (фальши) в голосе пользователя артикуляция резко меняет динамику. Это естественный эффект от известного насилия человеческой воли над здравым смыслом в критических ситуациях. В поступающем речевом сигнале требуется обнаружить моменты сбоев, вызванных внешними помехами или дефектами речи пользователя на основе анализа фонетической вариативности речи. Решение должно приниматься в режиме мягкого реального времени на интервалах вокализации речевого сигнала по данным выборочных наблюдений за динамическими изменениями артикуляции пользователя.

Под фонетической вариативностью речи специалисты понимают неконтролируемые изменения акустических свойств и характеристик речевого сигнала  $x(t)$  в текущем времени  $t$  [24, 28]. Из этого феномена следует, в частности, понятие множества реализаций  $r$ -й МРЕ  $X_r = \{\mathbf{x}_{r,j}, j=1, J_r\}$  в исходящем от пользователя (условного диктора) речевом потоке [27]. Здесь  $\mathbf{x}_{r,j}$  – вектор отсчётов  $j$ -й реализации  $r$ -й МРЕ ( $r \leq R$ );  $R$  – число фонем/МРЕ в анализируемом списке звуков речи;  $J_r$  – число реализаций  $r$ -й МРЕ. Например, при анализе русских гласных фонем, как наиболее содержательных в информационно-коммуникационном смысле [26, 29], имеем  $R=6$ . Фонетическую вариативность МРЕ характеризует значение

$$\rho_r^* \triangleq J_r^{-1} \sum_{j=1}^{J_r} \rho_r(\mathbf{x}_{r,j}) \quad (1)$$

среднего рассогласования реализаций  $\rho(\mathbf{x}_{r,j}|\mathbf{x}_r^*) \triangleq \rho_r(\mathbf{x}_{r,j})$  относительно информационного центра-эталона  $\mathbf{x}_r^*$ ;  $\triangleq$  – символ равенства по определению [2, 25, 26].

Эталонная реализация для  $r$ -й МРЕ определяется на основе принципа минимума суммарного расстояния до остальных реализаций:

$$\mathbf{x}_r^* \triangleq \mathbf{x}_{r,v}: \sum_{j=1}^{J_r} \rho(\mathbf{x}_{r,j}|\mathbf{x}_{r,v}) = \min_{q \leq J_r} \sum_{j=1}^{J_r} \rho(\mathbf{x}_{r,j}|\mathbf{x}_{r,q}), \quad (2)$$

где  $v$  – номер эталонной реализации.

Гипотеза  $H_r$ :  $\mathbf{x} \in X_r$  о появлении на выходе канала связи сигнала  $r$ -й МРЕ принимается по принципу допустимого рассогласования наблюдаемого речевого фрейма  $\mathbf{x}$  относительно эталона  $r$ -й МРЕ [27]:

$$\rho_r(\mathbf{x}) \leq \rho_0. \quad (3)$$

Пороговый уровень  $\rho_0$  устанавливается в зависимости от заданной вероятности ошибки 1-го рода [25, 30]:

$$\alpha_r \triangleq P\{\rho_r(\mathbf{x}) > \rho_0 | H_r\} \quad (4)$$

или уровня значимости принимаемых согласно правилу (3) решений. Чем меньше порог  $\rho_0$ , тем выше в общем случае требования условного наблюдателя к степени однородности кластера  $X_r$ . При этом не исключается возможность

срабатывания решающего правила (3) одновременно для нескольких МРЕ из фонетических баз данных  $\{X_r\}$  одного и того же диктора.

Проблема состоит в том, что акустические свойства и характеристики речевого сигнала значительно варьируются в зависимости от условий речеобразования и эмоционального состояния диктора. Поэтому при использовании параметра  $\rho_r^*$  из (1) в качестве акустического коррелята ДЭСП следует периодически обновлять составы фонетических кластеров  $X_r$  во времени  $t=t_1, t_2, \dots, t_k, \dots$ . Здесь  $k$  – порядковый номер периода наблюдений. Причём период  $T_x=t_k-t_{k-1}=\text{const}$  такого обновления должен быть рассчитан на определённое количество  $J_{r,k} \gg 1$  фреймов  $\mathbf{x}_{r,j}=\mathbf{x}_{r,j,k}$  по каждой из  $R$  контролируемых МРЕ в пределах отрезка  $x_k(t)$  речевого сигнала конечной длительностью  $T_x$ . Одновременно каждый раз согласно (2) пересчитывается номер  $v=v_k$  эталонной реализации и средний радиус  $\rho_r^*=\rho_{r,k}^*$  обновленного фонетического кластера  $X_r=X_{r,k}$ . В таком случае показателем фонетической вариативности речи диктора может служить усреднённое значение

$$\rho^*(k) = \frac{\sum_{r=1}^R J_{r,k} \rho_{r,k}^*}{\sum_{r=1}^R J_{r,k}} \quad (5)$$

фонетического радиуса  $\rho_{r,k}^*$  в пределах фонетической базы данных данного диктора со свойством  $\rho^*(k) \leq \rho_0$ , так как в соответствии с формулами (2), (3) выполняется соотношение  $\rho_{r,k}^* \leq \rho_0 \quad \forall r \leq R$ .

**Информационный показатель.** Объективным показателем ДЭСП может служить относительная разность показателей качества речи диктора (5) в двух периодах наблюдений за его речевым сигналом:

$$P(k) = \frac{\rho^*(k-1) - \rho^*(k)}{\rho_0} \cdot 100, \quad k=1, 2, \dots \quad (6)$$

В комфортных условиях речеобразования уровень значимости (4) всех МРЕ будет стабильным. Как следствие, в соответствии с (5) выполняется система приближенных равенств  $\rho^*(k) \approx \rho^*(k-1)$ , поэтому показатель (6) будет близок к нулю. При появлении внешних возмущений в процессе речепроизводства средний радиус (5) неизбежно увеличивается как показатель акустического качества речи диктора в некомфортных условиях [24]. При этом в соответствии с (6) неравенство  $P(k) < 0$  будет признаком отрицательной ДЭСП. В случае  $P(k) > 0$  можно говорить о положительной динамике в эмоциональном состоянии пользователя и повышении акустического качества речевого сигнала.

Для раскрытия показателя (6) в явном виде воспользуемся универсальным теоретико-информационным подходом, в рамках которого метрика  $\rho_r(\mathbf{x})$  определяется как информационное рассогласование Кульбака-Лейблера [31]. В работе [26] показано, что при обработке речевых сигналов в частотной области информационное рассогласование принимает вид симметричной формы дивергенции Итакуры-Саито или COSH-расстояния [32]:

$$\rho_{r,k}(\mathbf{x}) = F^{-1} \int_{-F/2}^{F/2} \left( \frac{G_{x,k}(f)}{G_r(f)} + \frac{G_r(f)}{G_{x,k}(f)} - 2 \right) df \geq 0, \quad r=1, \overline{R}. \quad (7)$$

Здесь  $G_r(f)$  – спектральная плотность мощности эталона  $r$ -й МРЕ;  $G_{x,k}(f)$  – её текущая оценка по речевому фрейму  $\mathbf{x}$  в пределах  $k$ -го интервала наблюдений  $[t_{k-1}; t_k]$ ;  $F$  – частота временной дискретизации речевого сигнала  $x_k(t)$ ;  $f$  – частота колебаний (в герцах). Однако в таком виде выражение (7) не может определять меру фонетических вариаций относительно эталона. Варьирование одноимённых МРЕ искажает результаты измерений таких основных акустических характеристик, как высота, тон и тембр звука речи, поэтому используем модификацию COSH-расстояния [20]:

$$\rho_{r,k}(\mathbf{x}) = \sqrt{\left( F^{-1} \int_{-F/2}^{F/2} \frac{G_{x,k}(f)}{G_r(f)} df \right) \left( F^{-1} \int_{-F/2}^{F/2} \frac{G_r(f)}{G_{x,k}(f)} df \right)} - 1 \geq 0. \quad (8)$$

Найденное таким образом рассогласование инвариантно к масштабу сигналов [33]: для любых спектральных плотностей мощности  $G^*(f) > 0$  и константы  $c > 0$  справедливо равенство

$$\rho_{r,k}(\mathbf{x})|_{G_{x,k}(f)=G^*(f)} = \rho_{r,k}(\mathbf{x})|_{G_{x,k}(f)=cG^*(f)}.$$

Заметим, что мера близости (7) этим свойством не обладает. Таким образом, выражение (8) с учётом формул (5), (6) определяет информационный показатель ДЭСП, рассчитанный на автоматическую нормировку речевого сигнала по интенсивности в варианте “gain optimization” [32]. В основу акустических измерений ДЭСП положен цифровой спектральный анализ.

**Описание метода акустических измерений динамики эмоционального состояния.** В задачах автоматической обработки речи распространено использование линейной авторегрессионной модели [34–36]:

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

где  $\mathbf{a}_r = \{a_r(i)\}$  – вектор коэффициентов порядка  $p > 1$ ;  $\eta_r(t)$  – порождающий процесс типа белого шума.

С учётом модели (9) оценим спектральную плотность мощности речевого сигнала для подстановки в (8) по принципу максимума энтропии [37]:

$$G_r(f) \triangleq \sigma_{\eta}^2 F \left| 1 - \sum_{i=1}^p a_r(i) \exp \frac{-j\pi i f}{F} \right|^{-2}, \quad |f| \leq 0,5F, \quad (10)$$

где  $\sigma_{\eta}^2$  – дисперсия порождающего процесса.

Для авторегрессионного анализа речевого сигнала в пределах текущего фрейма  $\mathbf{x}$  размерности  $n$  можно использовать, например, процедуру Берга-Левинсона [38] – оценку вектора авторегрессионных коэффициентов  $\mathbf{a}_r$  по выборке конечного объёма  $n \geq (2 \dots 4)p$ . Данное условие выполняется, например, в случае равенств  $n=80$ ;  $p=20$  [3, 27].

Отдельного внимания заслуживает вопрос об определении оптимального порогового уровня  $\rho_0$  (3). Для этого перепишем соотношение (8) в эквивалентном виде [20]:

$$\rho_r(\mathbf{x}) = \sqrt{\frac{\sigma_r^2(\mathbf{x})}{\sigma_r^2(\mathbf{x}_r^*)} \frac{\sigma_x^2(\mathbf{x}_r^*)}{\sigma_x^2(\mathbf{x})}} - 1 = \sqrt{\frac{\sigma_r^2(\mathbf{x})}{\sigma_x^2(\mathbf{x})} \frac{\sigma_x^2(\mathbf{x}_r^*)}{\sigma_r^2(\mathbf{x}_r^*)}} - 1, \quad (11)$$

где  $\sigma_r^2(\mathbf{x})$  – эмпирическая дисперсия отклика на речевой сигнал  $x(t)$  линейного обеляющего фильтра [25], настроенного на авторегрессионные коэффициенты  $r$ -й МРЕ;  $\sigma_r^2(\mathbf{x}_r^*)$  – эмпирическая дисперсия отклика линейного обеляющего фильтра на сигнал эталона  $x_r(t)$ ;  $\sigma_x^2(\mathbf{x})$ ,  $\sigma_x^2(\mathbf{x}_r^*)$  – эмпирические дисперсии откликов обеляющего фильтра, настроенного по наблюдаемому фрейму  $\mathbf{x}$ , на речевой сигнал и  $r$ -й эталон соответственно.

Обе дисперсии связаны с дисперсией  $\sigma_\eta^2$  порождающего белого шума (дисперсией погрешности линейного предсказания [38]):

$$\sigma_r^2(\mathbf{x}_r^*) \approx \sigma_\eta^2 \leq \sigma_r^2(\mathbf{x}).$$

При справедливости гипотезы  $H_r$  о появлении сигнала  $r$ -й МРЕ из выражения (11) в первом приближении можно записать [20]:

$$\rho_r(\mathbf{x}) \leq \sigma_r^2(\mathbf{x}) / \sigma_x^2(\mathbf{x}) - 1,$$

откуда в соответствии с (4) запишем [24, 28]:

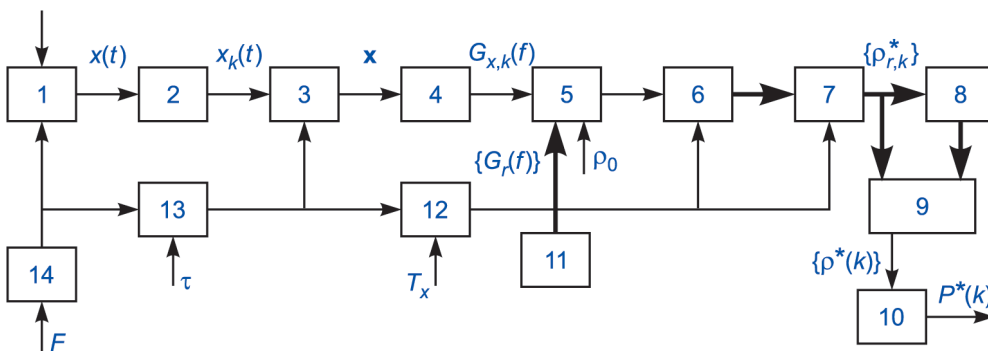


Рис. 1. Структурная схема устройства акустических измерений:

1 – канал речевой связи; 2 – аналого-цифровой преобразователь; 3 – формирователь речевого фрейма длительностью  $\tau$ ; 4 – цифровой спектральный анализатор; 5 – блок распознавания фонем с порогом  $\rho_0$ ; 6 – блок хранения фонетической базы данных пользователя; 7 – блок вычислений информационного рассогласования; 8 – блок памяти на период  $T_x$ ; 9 – блок вычислений парциальных показателей по формуле (5); 10 – блок вычислений интегрального показателя (6); 11 – блок записи и хранения фонетических эталонов; 12, 13 – соответственно второй и первый делители частоты тактовых импульсов; 14 – генератор тактовых импульсов частотой  $F$

Fig. 1. Block diagram of the acoustic measurement device:

1 – voice communication channel; 2 – analog-to-digital converter; 3 – shaper of a speech frame of duration  $\tau$ ; 4 – digital spectrum analyzer; 5 – phoneme recognition unit with a threshold  $\rho_0$ ; 6 – user phonetic database storage unit; 7 – unit for information discrimination calculation; 8 – memory unit for the  $T_x$  period; 9 – unit for calculating partial indicators (5); 10 – block for calculating the integral indicator (6); 11 – ideal phonemes recording and storing unit; 12 – second frequency divider; 13 – the first divider of the clock pulse frequency; 14 – clock pulse generator with frequency  $F$

$$\alpha_r \leq P\{\rho_r(\mathbf{x}) \geq \rho_0 | H_r\} = P\{F_{n-p, n-p} \geq 1 + \rho_0\} = 1 - \Phi_{n-p, n-p}(1 + \rho_0) \triangleq \alpha_0.$$

Здесь  $F_{n-p, n-p}$  – случайная величина с  $F$ -распределением Фишера со степенями свободы  $(n-p, n-p)$ ;  $\Phi_{n-p, n-p}(\cdot)$  – интегральная функция распределения этой случайной величины. Например, при уровне значимости  $\alpha_0 = 0,01$  и при  $n = 80$ ,  $p = 20$  получим пороговый уровень  $\rho_0 = 0,84$ .

Таким образом, выражение (6) совместно с выражениями (3), (5), (8)–(11) определяет метод акустических измерений информационного показателя ДЭСП. Большое значение при этом имеют его высокие динамические свойства. Они проявляются, во-первых, в предельно малой длительностью  $\tau = 10 \dots 15$  мс речевого фрейма  $\mathbf{x}$ , а также в известном [см. 24] эффекте «насыщения» фонетического кластера  $X_{r, k}$  реализациями  $\mathbf{x}_{r, j, k}$  при конечном объеме  $J_r$  наблюдений. Задержка в выдаче результатов измерений не превышает длительности одного интервала наблюдений  $T_x = 10 \dots 20$  с. Это является ключевым требованием к системе автоматической обработке речи при работе в режиме реального времени. Техническая реализация предлагаемого метода показана на рис. 1 в виде структурной схемы устройства акустических измерений.

Речевой сигнал пользователя по каналу связи 1 последовательно поступает на обработку в аналого-цифровой преобразователь 2, формирователь 3 речевого фрейма  $\mathbf{x}$  и цифровой спектральный анализатор 4, в котором измеряется текущая (кратковременная)

спектральная оценка  $G_{x, k}(f)$ . Частота  $F$  взятия отсчётов сигнала устанавливается в генераторе 14 тактовых импульсов. По значению  $G_{x, k}(f)$  в блоке 5 автоматического распознавания МРЕ с порогом  $\rho_0$  определяется порядковый номер  $r \leq R$  сигнала наиболее близкой фонемы из блока 6 хранения фонетической базы данных пользователя. Для этого используется критерий (3) и множество альтернативных спектральных плотностей мощности  $\{G_r(f)\}$  с выхода блока 11 записи и хранения фонетических эталонов. Содержание данного блока формируется наблюдателем по известной методике [2, 25] на этапе подготовки данных в задаче измерений ДЭСП. По результатам распознавания соответствующий кластер  $X_r$  мгновенно пополняется новым элементом. В момент окончания текущего ( $k$ -го) периода наблюдений – по сигналу с выхода делителя частоты 12 – в блоке 7 согласно формуле (1) пересчитываются

информационные радиусы всех  $R$  фонетических кластеров. Результаты пересчёта  $\{\rho_{r,k}^*\}$  поступают на вход блока 9 вычисления показателя (5) фонетической вариативности речи непосредственно и с задержкой на один период  $T_x$  в блоке памяти 8. В соответствии с формулой (6) в блоке 10 вычисляется текущее ( $k$ -е) значение информационного показателя ДЭСП.

Перед сигналом пуска в устройстве устанавливают значения его параметров: частоты дискретизации  $F$ , длительности фрейма  $\tau$ , длительности периода  $T_x$ . По ним определяют коэффициенты деления частоты тактовых импульсов  $n = F\tau$  и  $N = T_x/\tau = FT_x/n$  в блоках 11, 13 (см. рис. 1) соответственно. Если учесть паузы между словами (примерно 40–50 % длительности сигнала), а также долю чётко проговорённых гласных в пределах одного речевого потока (примерно 20–25 %), то за один период наблюдений будем иметь по  $J_r = (0,12...0,15)T_x/(\tau R)$  образцов каждой МРЕ. Например, при  $\tau = 10$  мс,  $T_x = 10$  с и  $R = 6$  (для гласных звуков русского языка)  $J_r = (0,12...0,15) \cdot 10 \cdot (10 \cdot 10^{-3} \cdot 6) = 20...25$  образцов  $\mathbf{x}_{r,j,k}$ . Теоретически этого вполне достаточно для формирования представительного фонетического кластера  $X_r$  в динамике [24].

**Результаты экспериментального исследования.** Для подтверждения и развития выводов теоретического исследования авторами настоящей статьи проведён натурный эксперимент для реальных речевых сигналов. При этом использовано авторское программное обеспечение (Voice Self-Analysis [сайт]. URL: <https://sites.google.com/site/frompldcreators/VoiceSelfAnalysis.zip>), пользовательский интерфейс подробно описан в работе [24] и представлен на сайте авторов настоящей статьи. Принцип действия программы основан на алгоритме автоматической обработке речи (3), (8), (10) и методе акустических измерений информационного показателя ДЭСП согласно выражениям (5), (6) с применением авторегрессионной модели (9) достаточно высокого порядка  $p = 20$ .

Программа эксперимента включала в себя оценку информационного показателя (6) при чтении контрольного текста двумя тестируемыми пользователями – мужчиной (диктором 1) и женщиной (диктором 2). Контрольный текст – первая глава романа А. С. Пушкина «Евгений Онегин» прочитан в среднем темпе, объём текста примерно 15–20 тыс. знаков и, соответственно, длительность  $T_0$  речевого сигнала каждого диктора не превышала 20–25 мин. В программе «Voice Self-Analysis» этот сигнал подвергался аналого-цифровому преобразованию с частотой дискретизации  $F = 8$  кГц. Полученный цифровой сигнал разделялся на последовательность фреймов  $\mathbf{x}$  длительностью  $\tau = 10$  мс. По фреймам в режиме «скользящего окна» согласно алгоритму (8)–(11) вычислялись текущие значения информационного рассогласования  $\{\rho_{r,k}(x)\}$ . По этим значениям в пределах каждого отдельного фрейма  $\mathbf{x}$  с использованием решающего правила (3) детектировались гласные фонемы  $\mathbf{x}_r$  диктора (случай  $R = 6$ ,  $\rho_0 = 0,8$ ), которые направлялись на хранение и обработку в соответствующий кластер  $X_r$  его фонетической базы данных. Процесс наполнения указанной базы данных с периодом  $T_x = 10$  с повторялся для нахождения согласно (1)

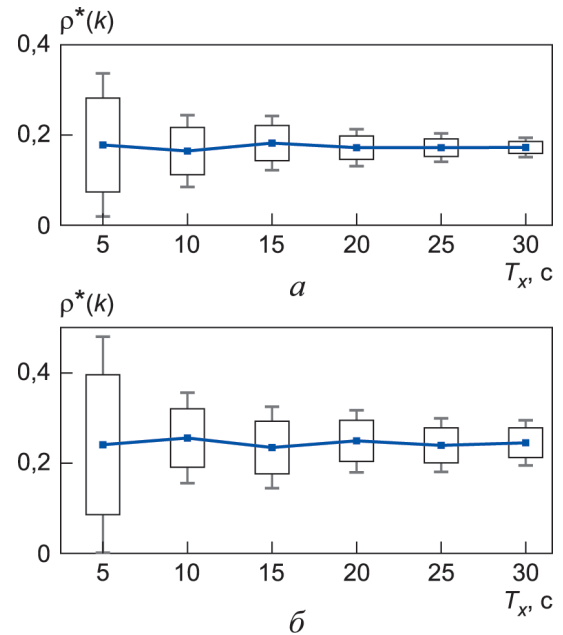


Рис. 2. Диаграмма размаха показателя (см. формулу (5)) вариативности речевого сигнала  $x_k(t)$  первого (а) и второго (б) контрольных дикторов в зависимости от длительности интервала наблюдений  $T_x$

Fig. 2. Indicator (5) of the speech signal variability for the first (a) and the second (b) control speakers, depending on the duration of the observation interval  $T_x$

значений информационных эталонов  $\{\mathbf{x}_{r,k}^*\}$  и радиусов  $\{\rho_{r,k}^*\}$  в пределах каждого отдельного  $k$ -го интервала наблюдений. При этом каждый раз по формулам (5), (6) вычислялись  $R$  показателей фонетической вариативности речи дикторов, а также текущие значения их интегральных показателей ДЭСП. Результаты измерений выводились в режиме on-line в главное окно программы в виде графика зависимости  $P(k)$  в дискретном времени  $k = 1, 2, \dots$ . Содержимое фонетической базы данных каждые  $T_x = 10$  с автоматически обнулялось в целях её подготовки к очередному наполнению по той же схеме вычислений.

Исследование по каждому диктору проводилось в два этапа. На первом этапе изучали динамические свойства метода измерений информационного показателя (6), на втором – чувствительность метода к влиянию помех в процессе речепроизводства. В качестве помех выступал акустический шум, созданный программным способом и действующий на диктора через наушники во время чтения контрольного текста. Интенсивность помехи  $h^2$  варьировалась в диапазоне 0–30 дБ выше средней мощности фонового шума. Относительная погрешность измерений при доверительной вероятности 0,9 составила  $\delta = 1,65/\sqrt{L}$  [26, 27], где  $L = T_0/T_x$  – число интервалов наблюдений в пределах контрольного речевого сигнала. При  $T_x = 10$  с это число  $L = (20...25) \cdot 60 : 10 = 120...150$ , следовательно, погрешность  $\delta = 13...15$  %.

На рис. 2 представлена диаграмма размаха показателя вариативности (5) на множестве МРЕ для контрольных

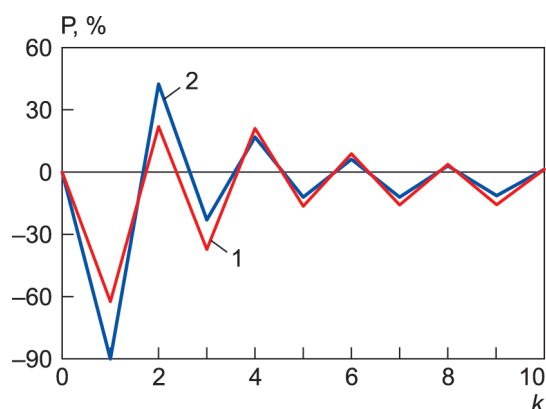


Рис. 3. Графики показателя (см. формулу (6)) динамики ЭСП для первого (1) и второго (2) контрольных дикторов при интенсивности акустической помехи  $h^2=20$  дБ

Fig. 3. Indicator (6) for the first (1) and the second (2) control speakers at the intensity of acoustic interference  $h^2=20$  dB

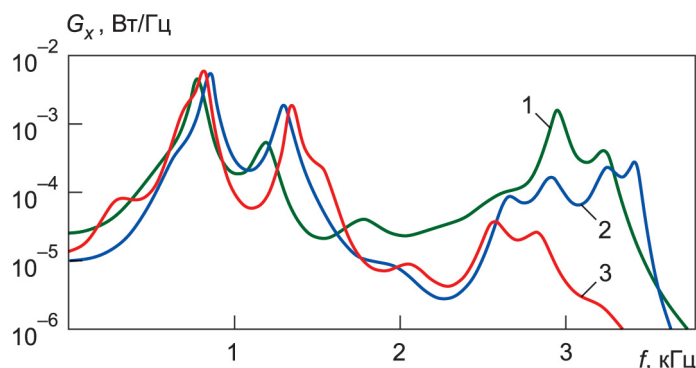


Рис. 4. Частотные зависимости спектральной плотности шума фонемы «а» по речевому фрейму  $x$  первого контрольного диктора при интенсивности акустической помехи  $h^2=0; 10; 20$  дБ – кривые 1–3 соответственно

Fig. 4. Estimates of the power spectral density of the phoneme «a» for the speech frame  $x$  from the first control speaker depending on the intensity of the acoustic interference  $h^2=0; 10; 20$  dB – 1–3 curves

дикторов в комфортных условиях речеобразования (случай  $h^2=0$ ) в зависимости от длительности интервала наблюдений  $T_x$ . При увеличении  $T_x$  амплитуда размаха  $\delta$  в обоих случаях монотонно сокращается до минимума и стабилизируется на стационарном уровне при  $T_x=10...20$  с. Различия между дикторами состоят в данном случае лишь в значении стационарного уровня показателя  $p^*(k)$  и дисперсии погрешности его измерений в динамике: у диктора-мужчины они меньше.

На рис. 3 приведены информационные показатели ДЭСП для первого и второго контрольных дикторов при интенсивности акустической помехи  $h^2=20$  дБ. Здесь  $k=0$  – точка отсчёта по оси абсцисс – момент появления помехи в наушниках дикторов. Реакция обоих дикторов на помеху носит ожидаемо затухающий характер: сначала резкий скачок

вниз зависимости  $P(k)$  как внутренний протест дикторов в ответ на внешнее возмущение и далее постепенное «привыкание» к помехе. Различия в степени протеста дикторов, а также в скорости стабилизации показателя (6) в окрестностях нулевого уровня можно объяснить различиями их пола и темперамента [7, 29]. Можно предположить, что в общем случае амплитуда скачка и длительность «привыкания» дикторов к внешнему возмущению пропорциональны интенсивности действующей помехи.

На рис. 4 представлено семейство оценок спектральной плотности мощности фонемы «а» первого контрольного диктора, полученных для разных отношений сигнал/шум  $h^2$ . Из сопоставления этих оценок можно сделать вывод в отношении эмоциональной реакции дикторов на внешние возмущения: наиболее чувствительны форманты в спектре речевого сигнала из области обертонов, отвечающие за тембр речи диктора [27, 28].

**Заключение.** Предложенный информационный показатель ДЭСП на основе масштабно-инвариантной модификации метрики Кульбака-Лейблера и метод его измерения можно применять для разных дикторов, различных речевых потоков и условий речеобразования. Важной отличительной особенностью разработанного метода измерения является высокая скорость сходимости информационного показателя (6), (8) в условиях малых выборок наблюдений. Теоретическое и экспериментальное обоснования метода измерений информационного показателя ДЭСП выполнены через оценку в режиме реального времени его акустического коррелята – фонетической вариативности речевого сигнала. В силу взаимосвязи речи с мозговой деятельностью диктора [9, 10] любые изменения эмоционального состояния пользователя мгновенно ухудшают акустическое качество его речевого сигнала. Отслеживая и фиксируя такие изменения в динамике, наблюдатель получает возможность оперативно и адекватно реагировать на те или иные колебания в эмоциональном состоянии собеседника. Полученные результаты подтверждены экспериментальными данными и будут полезны исследователям и разработчикам систем дистанционного обслуживания, рассчитанных на работу в режиме речевого диалога с пользователем. Предложенное решение задачи измерения и отслеживания показателя ДЭСП может повысить качество обслуживания клиентов, в частности, при внедрении автоматизированного контроля качества голосовых образцов пользователей единых биометрических систем [2, 3], а также в перспективе противодействовать терроризму и угрозам национальной безопасности за счёт бесконтактного выявления потенциально опасных лиц с краткосрочным нарушением психоэмоционального состояния.

Исследование выполнено при поддержке Российского научного фонда (проект № 20-71-10010).

#### Список литературы

1. Davis S. K. et al., *Personality and Individual Differences*, 2020, vol. 160, no. 109938. <https://doi.org/10.1016/j.paid.2020.109938>

2. 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>
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. Галышина Е. И. Актуальные проблемы идентификации лиц по фонограммам телефонных переговоров // Сборник материалов XXIII международной научно-практической конференции «Деятельность правоохранительных органов в современных условиях». В 2-х томах. Восточно-Сибирский институт Министерства внутренних дел Российской Федерации. Иркутск, 2018. С. 141–146. URL: <https://istina.msu.ru/publications/article/167326015> (дата обращения: 14.08.2020).
5. Falagiarda F., Collignon O., *Cortex*, 2019, vol. 119, pp. 184–194. <https://doi.org/10.1016/j.cortex.2019.04.017>
6. Akbulut F. P., Perros H. G., Shahzad M., *Computer Methods and Programs in Biomedicine*, 2020, vol. 195, no. 105571. <https://doi.org/10.1016/j.cmpb.2020.105571>
7. Shaqra F. A., Duwairi R., Al-Ayyoub M., *Procedia Computer Science*, 2019, vol. 151, pp. 37–44. <https://doi.org/10.1016/j.procs.2019.04.009>
8. Arana J. M. et al., *Computers in Human Behavior*, 2020, vol. 104, no. 106156. <https://doi.org/10.1016/j.chb.2019.106156>
9. Bourguignon M. et al., *NeuroImage*, 2020, vol. 216, no. 116788. <https://doi.org/10.1016/j.neuroimage.2020.116788>
10. Liu Z. et al., *Brain and Language*, 2020, vol. 203, no. 104755. <https://doi.org/10.1016/j.bandl.2020.104755>
11. Schuller B., *Voice and Speech Analysis in Search of States and Traits*, in: Salah A. A., Gevers T. (eds.) *Computer Analysis of Human Behavior*, Springer, Heidelberg, 2011, 227 p. [https://doi.org/10.1007/978-0-85729-994-9\\_9](https://doi.org/10.1007/978-0-85729-994-9_9)
12. Cardona D. et al., *Neurocomputing*, 2017, vol. 265, pp. 78–90. <https://doi.org/10.1016/j.neucom.2016.09.140>
13. Yu D., Deng L., *Automatic Speech Recognition: A Deep Learning Approach*, Springer, 2014, 321 p. <https://doi.org/10.1007/978-1-4471-5779-3>
14. Schuster M., *Lecture Notes in Computer Science*, 2010, vol. 6230, pp. 8–10. [https://doi.org/10.1007/978-3-642-15246-7\\_3](https://doi.org/10.1007/978-3-642-15246-7_3)
15. Rammohan R. et al., *Journal of Allergy and Clinical Immunology*, 2017, vol. 139, iss. 2, no. ab250. <https://doi.org/10.1016/j.jaci.2016.12.804>
16. Володин Н. А., Ермоленко Т. В., Семенюк В. В. Исследование эффективности применения нейронных сетей для распознавания эмоций человека по голосу // В сборнике: Доноски чтения 2019: образование, наука, инновации, культура и вызовы современности. Материалы IV Международной научной конференции. 2019. С. 221–223. URL: [https://elibrary.ru/download/elibrary\\_41422521\\_75290048.pdf](https://elibrary.ru/download/elibrary_41422521_75290048.pdf) (дата обращения: 14.08.2020).
17. Grachev A. M., Ignatov D. I., Savchenko A. V., *Applied Soft Computing*, 2019, vol. 79, pp. 354–362. <https://doi.org/10.1016/j.asoc.2019.03.057>
18. Устинов Р. А. Особенности современных систем защиты речевой информации. *Безопасность информационных технологий*. [Электрон. журнал]. 2017. Т. 24. № 4. <https://doi.org/10.26583/bit.2017.4.08>
19. Cui S., Li E., Kang X., Autoregressive model based smoothing forensics of very short speech clips, *2020 IEEE International Conference on Multimedia and Expo (ICME)*, London, United Kingdom, 2020, pp. 1–6. <https://doi.org/10.1109/ICME46284.2020.9102765>
20. Savchenko V. V., *Radioelectronics and Communications Systems*, 2020, vol. 63, no 1, pp. 42–54. <https://doi.org/10.3103/S0735272720010045>
21. Савченко В. В., Савченко А. В. Критерий гарантированного уровня значимости в задаче автоматической сегментации речевого сигнала // Радиотехника и электроника. 2020. Т. 65. № 11. С. 1060–1066. <https://doi.org/10.31857/S0033849420110157>
22. Hautamäki R. G. et al., *Speech Communication*, 2017, vol. 95, pp. 1–15. <https://doi.org/10.1016/j.specom.2017.10.002>
23. 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. URL: <http://naukarus.com/akusticheskie-harakteristiki-rechevogo-signal-a-kak-pokazatel-funktsionalnogo-sostoyaniya-cheloveka> (дата обращения: 14.08.2020).
24. Savchenko V. V., *Journal of Communications Technology and Electronics*, 2018, vol. 63, no. 1, pp. 53–57. <https://doi.org/10.1134/S1064226918010126>
25. Savchenko A. V., Savchenko V. V., *Journal of Communications Technology and Electronics*, 2016, vol. 61, no. 4, pp. 430–435. <https://doi.org/10.1134/S1064226916040112>
26. Савченко В. В. Метод измерения показателя акустического качества речи на основе теоретико-информационного подхода // Измерительная техника. 2018. № 1. С. 60–64. <https://doi.org/10.32446/0368-1025it.2018-1-60-64>
27. 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>
28. Savchenko L. V., Savchenko A. V., *Journal of Communications Technology and Electronics*, 2019, vol. 64, no. 3, pp. 238–244. <https://doi.org/10.1134/S1064226919030173>
29. Савченко А. В., Савченко В. В. Метод измерения частоты основного тона речевого сигнала для систем акустического анализа речи // Измерительная техника. 2019. № 3. С. 59–63. <https://doi.org/10.32446/0368-1025it.2019-3-59-63>
30. Savchenko A. V., Sequential Three-Way Decisions in Efficient Classification of Piecewise Stationary Speech Signals, in: Polkowski L. et al. (eds) *Rough Sets. IJCRS 2017. Lecture Notes in Computer Science*, 2017, vol. 10314, Springer, Cham. [https://doi.org/10.1007/978-3-319-60840-2\\_19](https://doi.org/10.1007/978-3-319-60840-2_19)
31. Kullback S., *Information Theory and Statistics*, N. Y., Dover Publications, 1997, 432 p. URL: <https://www.amazon.com/dp/0486696847> (дата обращения: 14.08.2020).

32. Gray R. M. et al., *IEEE Transactions on Signal Processing*, 1980, vol. 28, no. 4, pp. 367–377.

<https://doi.org/10.1109/TASSP.1980.1163421>

33. Savchenko A. V., Savchenko V. V., Savchenko L. V., Optimization of Gain in Symmetrized Itakura-Saito Discrimination for Pronunciation Learning, in: Kononov A., Khachay M., Kalyagin V., Pardalos P. (eds), *Mathematical Optimization Theory and Operations Research. MOTOR 2020, Lecture Notes in Computer Science*, 2020, vol. 12095, Springer, Cham.

[https://doi.org/10.1007/978-3-030-49988-4\\_30](https://doi.org/10.1007/978-3-030-49988-4_30)

34. Vestman V. et al., *Speech Communication*, 2018, vol. 99, pp. 62–79. <https://doi.org/10.1016/j.specom.2018.02.009>

35. Candan Ç., *Signal Processing*, 2020, vol. 166, no. 107256. <https://doi.org/10.1016/j.sigpro.2019.107256>

36. Tuncel K. S., Baydogan M. G., *Pattern Recognition*, 2018, vol. 73, pp. 202–215.

<https://doi.org/10.1016/j.patcog.2017.08.016>

37. Savchenko V. V., Savchenko A. V., *Radioelectronics and Communications Systems*, 2019, vol. 62, pp. 276–286.

<https://doi.org/10.3103/S0735272719050042>

38. Marple S. L., *Digital Spectral Analysis with Applications*, 2nd ed. Mineola, New York, Dover Publications, 2019, 432 p. URL: <https://www.goodreads.com/book/show/19484239> (дата обращения: 14.08.2020).

## References

1. Davis S. K. et al., *Personality and Individual Differences*, 2020, vol. 160, no. 109938.

<https://doi.org/10.1016/j.paid.2020.109938>.

2. Savchenko V. V., Savchenko A. V., *Measurement Techniques*, 2020, vol. 62, no. 5, pp. 458–465.

<https://doi.org/10.1007/s11018-020-01702-1>

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. Galyashina E. I., Aktual'nye problemy identifikatsii lits po fonogrammam telefonnyh peregovorov, *Proceedings of the XXIII International Scientific and Practical Conference "Deyatel'nost' pravoohranitel'nyh organov v sovremennyh usloviyah"*, in 2 volumes, Irkutsk, Vostochno-Sibirskiy institut Ministerstva vnutrennih del Rossijskoj Federatsii Publ., 2018, pp. 141–146, available at: <https://istina.msu.ru/publications/article/167326015/> (accessed: 14.08.2020). (In Russ.)

5. Falagiarda F., Collignon O., *Cortex*, 2019, vol. 119, pp. 184–194. <https://doi.org/10.1016/j.cortex.2019.04.017>

6. Akbulut F. P., Perros H. G., *Computer Methods and Programs in Biomedicine*, 2020, vol. 195, no. 105571.

<https://doi.org/10.1016/j.cmpb.2020.105571>

7. Shaqra F. A., Duwairi R., Al-Ayyoub M., *Procedia Computer Science*, 2019, vol. 151, pp. 37–44.

<https://doi.org/10.1016/j.procs.2019.04.009>

8. Arana J. M. et al., *Computers in Human Behavior*, 2020, vol. 104, no. 106156. <https://doi.org/10.1016/j.chb.2019.106156>

9. Bourguignon M. et al., *NeuroImage*, 2020, vol. 216, no. 116788. <https://doi.org/10.1016/j.neuroimage.2020.116788>

10. Liu Z. et al., *Brain and Language*, 2020, vol. 203, no. 104755. <https://doi.org/10.1016/j.bandl.2020.104755>

11. Schuller B., Voice and Speech Analysis in Search of States and Traits, in: Salah A. A., Gevers T. (eds) *Computer Analysis of Human Behavior*, Springer, Heidelberg, 2011, 227 p. [https://doi.org/10.1007/978-0-85729-994-9\\_9](https://doi.org/10.1007/978-0-85729-994-9_9)

12. Cardona D. et al., *Neurocomputing*, 2017, vol. 265, pp. 78–90. <https://doi.org/10.1016/j.neucom.2016.09.140>

13. Yu D., Deng L., *Automatic Speech Recognition: A Deep Learning Approach*, Springer, 2014, 321 p.

<https://doi.org/10.1007/978-1-4471-5779-3>

14. Schuster M., *Lecture Notes in Computer Science*, 2010, vol. 6230, pp. 8–10. [https://doi.org/10.1007/978-3-642-15246-7\\_3](https://doi.org/10.1007/978-3-642-15246-7_3)

15. Rammohan R. et al., *Journal of Allergy and Clinical Immunology*, 2017, vol. 139, iss. 2, no. ab250.

<https://doi.org/10.1016/j.jaci.2016.12.804>

16. Volodin N. A., Ermolenko T. V., Semenyuk V. V., Issledovanie effektivnosti primeneniya nejronnyh setej dlya raspoznaniya emocij cheloveka po golosu, *Proceedings of the Conference International Scientific Conference "Doneckie chteniya 2019: obrazovanie, nauka, innovacii, kul'tura i vyzovy sovremenosti"*, 2019, pp. 221–223, available at: [https://elibrary.ru/download/elibrary\\_41422521\\_75290048.pdf](https://elibrary.ru/download/elibrary_41422521_75290048.pdf) (accessed: 14.08.2020). (In Russ.)

17. Grachev A. M., Ignatov D. I., Savchenko A. V., *Applied Soft Computing*, 2019, vol. 79, pp. 354–362.

<https://doi.org/10.1016/j.asoc.2019.03.057>

18. Ustinov R. A., *Bezopasnost' informacionnyh tekhnologij*, 2017, vol. 24, no. 4. (In Russ.) <https://doi.org/10.26583/bit.2017.4.08>

19. Cui S., Li E., Kang X., 2020 *IEEE International Conference on Multimedia and Expo (ICME)*, London, United Kingdom, 2020, pp. 1–6. <https://doi.org/10.1109/ICME46284.2020.9102765>

20. Savchenko V. V., *Radioelectronics and Communications Systems*, 2020, vol. 63, no. 1, pp. 42–54.

<https://doi.org/10.3103/S0735272720010045>

21. Savchenko V. V., Savchenko A. V., *Journal of Communications Technology and Electronics*, 2020, vol. 65, no. 11, pp. 1060–1066. <https://doi.org/10.31857/S0033849420110157>

22. Hautamäki R.G. et al., *Speech Communication*, 2017, vol. 95, pp. 1–15. <https://doi.org/10.1016/j.specom.2017.10.002>

23. Lebedeva N. N., Karimova E. D., *Uspekhi fiziologicheskikh nauk*, 2014, vol. 45, no. 1, pp. 57–95, available at: <http://naukarus.com/akusticheskie-harakteristiki-rechevogo-signala-kak-pokazatel-funktsionalnogo-sostoyaniya-cheloveka> (accessed: 14.08.2020). (In Russ.)

24. Savchenko V. V., *Journal of Communications Technology and Electronics*, 2018, vol. 63, no. 1, pp. 53–57.

<https://doi.org/10.1134/S1064226918010126>

25. Savchenko A. V., Savchenko V. V., *Journal of Communications Technology and Electronics*, 2016, vol. 61, no. 4, pp. 430–435. <https://doi.org/10.1134/S1064226916040112>

26. Savchenko V. V., *Measurement Techniques*, 2018, vol. 61, no. 1, pp. 79–84. <https://doi.org/10.1007/s11018-018-1391-8>

27. 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>

28. Savchenko L. V., Savchenko A. V., *Journal of Communications Technology and Electronics*, 2019, vol. 64, no. 3, pp. 238–244. <https://doi.org/10.1134/S1064226919030173>
29. Savchenko A. V., Savchenko V. V., *Measurement Techniques*, 2019, vol. 62, no. 3, pp. 282–288. <https://doi.org/10.1007/s11018-019-01617-x>
30. Savchenko A. V., Sequential Three-Way Decisions in Efficient Classification of Piecewise Stationary Speech Signals, in: Polkowski L. et al. (eds) *Rough Sets. IJCRS 2017, Lecture Notes in Computer Science*, 2017, vol. 10314, Springer, Cham. [https://doi.org/10.1007/978-3-319-60840-2\\_19](https://doi.org/10.1007/978-3-319-60840-2_19)
31. Kullback S., Information Theory and Statistics, N. Y., *Dover Publications*, 1997, 432 p., available at: <https://www.amazon.com/dp/0486696847> (accessed: 14.08.2020).
32. Gray R. M. et al., *IEEE Transactions on Signal Processing*, 1980, vol. 28, no. 4, pp. 367–377. <https://doi.org/10.1109/TASSP.1980.1163421>
33. Savchenko A. V., Savchenko V. V., Savchenko L. V., Optimization of Gain in Symmetrized Itakura-Saito Discrimination for Pronunciation Learning, in: Kononov A., Khachay M., Kalyagin V., Pardalos P. (eds), *Mathematical Optimization Theory and Operations Research, MOTOR 2020, Lecture Notes in Computer Science*, 2020, vol. 12095, Springer, Cham. [https://doi.org/10.1007/978-3-030-49988-4\\_30](https://doi.org/10.1007/978-3-030-49988-4_30)
34. Vestman V. et al., *Speech Communication*, 2018, vol. 99, pp. 62–79. <https://doi.org/10.1016/j.specom.2018.02.009>
35. Candan Ç., *Signal Processing*, 2020, vol. 166, no. 107256. <https://doi.org/10.1016/j.sigpro.2019.107256>
36. Tuncel K. S., Baydogan M. G., *Pattern Recognition*, 2018, vol. 73, pp. 202–215. <https://doi.org/10.1016/j.patcog.2017.08.016>
37. Savchenko V. V., Savchenko A. V., *Radioelectronics and Communications Systems*, 2019, vol. 62, pp. 276–286. <https://doi.org/10.3103/S0735272719050042>
38. Marple S. L., *Digital Spectral Analysis with Applications*, 2nd ed. Mineola, New York, Dover Publications, 2019, 432 p., available at: <https://www.goodreads.com/book/show/19484239> (accessed: 14.08.2020).

Статья поступила в редакцию 14.09.2020; принята к публикации 14.10.2020.