ИНФОКОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ
INFOCOMMUNICATION TECHNOLOGIES

УДК 004.934.2
DOI:10.18413/2411-3808-2018-45-1-168-175

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ФОНЕТИЧЕСКИХ СВОЙСТВ РЕЧЕВОГО СИГНАЛА НА ОСНОВЕ ЕГО ТЕОРЕТИКО-ИНФОРМАЦИОННОЙ МОДЕЛИ

EXPERIMENTAL STUDY OF PHONETIC PROPERTIES OF THE SPEECH SIGNAL ON THE BASIS OF ITS INFORMATION-THEORETIC MODEL

В.В. Савченко, Т.А. Соловьева
V.V. Savchenko, T.A. Solovyova

Нижегородский государственный лингвистический университет,
Россия, 603155, г. Нижний Новгород, ул. Минина 31а

Nizhniy Novgorod State Linguistic University,
31a Minin St, 603155, Nizhny Novgorod, Russia

E-mail: svv@lunn.ru, appleta1@rambler.ru

Аннотация
На основе теоретико-информационной модели минимальных речевых единиц в метрике Кульбака-Лейблера исследованы фонетические (звуковые) свойства речевого сигнала в зависимости от индивидуальности диктора и условий речеобразования. Исследование проведено экспериментальным способом с использованием авторского программного комплекса «Voice Self-Analysis». Описаны программа и методика эксперимента, представлены полученные результаты. Показано, что фонетические свойства речи зависят от пола, возраста и образования диктора и поэтому могут служить его идентификации по голосу. Даны рекомендации по их применению в системах автоматической обработки речи.

Abstract
On the basis of the information-information model of minimal speech units in the Kulbak-Leibler metric, the phonetic (sound) properties of the speech signal are investigated depending on the speaker's individuality and the conditions for speech formation. The research was carried out in an experimental way using the author's software package "Voice Self-Analysis", intended for rapid testing of the speaker's emotional state by his speech signal. The statement of the problem of studying the phonetic properties of a speech signal has been made. The program and experimental procedure are described. Based on the results of the experiments, the results are presented. It is shown that the phonetic properties of speech depend on gender, age and the formation of the announcer, and therefore can serve as his voice identification. The efficiency of evaluating the properties of a speech signal in this way was proved. Conclusions are made and recommendations are given on their practical application in automatic speech processing systems.

Ключевые слова: речь, акустика речи, речевой сигнал, минимальная речевая единица, фонетический детектор лжи
Keywords: speech, acoustics of the speech, speech signal, minimum speech unit, phonetic lie detector

Введение

Статья посвящена исследованию фонетических (звуковых) свойств речевого сигнала, в том числе на выходе телефонного канала связи, с использованием достижений современной науки и информационных технологий. Указанные свойства играют важную роль при разработке и тестировании систем автоматической обработки речи, включая область голосового управления робототехникой [Savchenko, 2016]. Поэтому тема настоящей статьи представл яется актуальной как для теории, так и для практики автоматической обработки речи [Савченко, 2015].

Принцип действия большинства речевых систем и технологий основывается на последовательном членении речевого сигнала на короткие (10-20 мс) фреймы $\mathbf{x} = (x_1, x_2, ..., x_n)$ длиной в одну минимальную речевую единицу (МРЕ) с их последующим сопоставлением по тонкой, в частности, спектральной структуре [Savchenko, 2015] с соответствующим эталоном. Поэтому главной проблемой для таких систем является выбор и обоснование множества фонетических эталонов $\left\{x_r^*\right\}$

Основная часть

Известно [Savchenko, 2016], что любой диктор в силу ряда причин, например, из-за особенностей своей речи или слуха, в принципе не в состоянии в процессе речеобразования точно воспроизвести эталон той или иной (r-й) МРЕ. Выходом из такой ситуации может служить задание каждой МРЕ не одной, а одновременно несколькими допустимыми вариантами

$$x_{r,j}, \quad j = \overline{1; J_r},$$

где $r = \overline{1; R}$, а R – объем фонетической базы данных. В таком случае диктору будет достаточно приблизить свое произношение к любому из них, чтобы быть правильно понятым гипотетическим наблюдателем или слушателем. Этим существенно ослабляется рассматриваемая проблема вариативности устной речи: каждый конкретный диктор в процессе речеобразования выбирает наиболее удобный, достижимый для себя вариант эталонного произношения МРЕ из некоторого множества альтернатив $\left\{x_{r,j}\right\}$ [Savchenko, 2015]. Одновременно становится понятным и собственно критерий качества формируемого (на выходе речевого тракта диктора) речевого сигнала к эталону: он должен войти в границы J_r-множества X_r вариантов рассматриваемой МРЕ как полноправный, $\left(J_r + 1\right)$-й его элемент. Задача переходит, в таком случае, в сугубо предметную плоскость: сначала по каждой из R рассматриваемых МРЕ требуется сформировать множество (кластер) $X_r \equiv \left\{x_{r,j}\right\}$ ее допустимых образцов – на этапе обучения диктора [Савченко, 2016]. И после этого в процессе речеобразования тестировать текущий сигнал $\mathbf{x}$ согласно естественному требованию достаточной близости к ним - в среднем в пределах кластера X_r - в некоторой метрике $\rho(\mathbf{x}/\mathbf{x}_{r,j})$. При высокой степени близости качество речи диктора в отношении определенной фонемы можно оценить как высокое. И, наоборот, при нарушении указанного требования соответствующая (текущая) МРЕ должна быть забракована наблюдателем как ошибка речеобразования [Савченко, 2016].

В вычислительном отношении проще, однако, задаться аналогичным условием

$$\rho(\mathbf{x} / \mathbf{x}_r^*) \le \rho_0 \qquad (1)$$

тестирования вектора входного сигнала $\mathbf{x}$ относительно «центра массы» r-го кластера

$$\mathbf{x}_r^* = \mathbf{x}_{r,\nu} : \; J_r^{-1} \sum_{j=1}^{J_r} \rho(\mathbf{x}_{r,j} / \mathbf{x}_{r,\nu}) = \min_{i \le J_r} J_r^{-1} \sum_{j=1}^{J_r} \rho(\mathbf{x}_{r,j} / \mathbf{x}_{r,i}) \equiv \rho_r^*. \qquad (2)$$

где ρ_0 - допустимый пороговый уровень. В режиме реального времени вместо $J_r >> 1$ расстояний $\rho(\mathbf{x}/\mathbf{x}_{r,j})$ здесь вычисляется только одно расстояние от $\mathbf{x}$ в пределах кластера X_r: до его центра $\mathbf{x}_r^*$. Указанный центр – это своего рода эталон данного кластера, или эталон соответствующей фонемы. А множество таких эталонов $\{\mathbf{x}_r^*\}$ – экономный способ задания фонетической базы данных конкретного диктора, или звукового ряда $\{X_r\}$ (строя) его разговорного языка. В информационной теории восприятия речи [Savchenko, Savchenko, 2016.] в роли расстояний между аллофонами в (1) используются величина информационного рассогласования по Кульбаку-Лейблеру [Кульбак, 1967]. Для наглядности ниже на рис. 1 показана геометрическая интерпретация кластерной модели МРЕ (1), (2). Здесь жирной точкой обозначен информационный центр–эталон $\mathbf{x}_r^*$ фонетического кластера X_r.

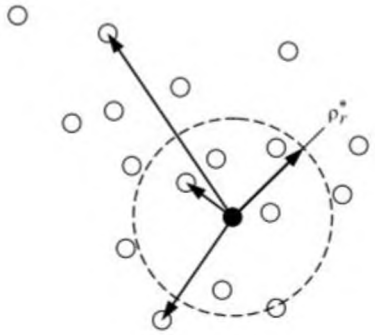

Рис. 1. Кластер реализаций фонемы и его информационный центр-эталон
Fig. 1. The cluster of phoneme realizations and its reference data center

Не трудно заметить, что в рамках рассматриваемой теоретико-информационной модели [Савченко, Акатьев, 2013] r-й МРЕ фонетическое качество речи (ФКР) характеризуется в количественном виде величиной среднего радиуса ρ_r^* соответствующего фонетического кластера относительно его информационного эталона $\mathbf{x}_r^*$. При увеличении этого радиуса в процессе речеобразования можно говорить об ухудшении ФКР диктора [Савченко, 2016], за счет увеличения вариативности его МРЕ. И, наоборот, при уменьшении ρ_r^* фонетическое качество речи объективно улучшается за счет повышения степени однородности фонетических единиц в пределах одного речевого потока [Savchenko, 2015]. Таким образом, величина ρ_r^* может рассматриваться как информационный показатель ФКР по каждой отдельной, r-й фонеме. Его возможности на множестве гласных фонем русского языка исследуются далее экспериментальным

способом. При этом используется авторский программный комплекс «Voice Self-Analysis [Савченко, 2017], в котором величина ρ_r^* определяется по конечному отрезку речевого сигнала в режиме реального мягкого времени.

Программный комплекс «Voice Self-Analysis» (от английского «самоанализ») – это новая версия программы «Фонетический детектор лжи» [Савченко, Васильев, 2014], предназначенная для экспресс-тестирования (самоанализа) эмоционального состояния диктора по его речевому сигналу. Принцип действия комплекса основывается на измерении текущего фонетического (звукового) качества речи диктора. При этом используются известные [Savchenko, Savchenko, 2016] информационные корреляты МРЕ в метрике Кульбака-Лейблера [Кульбак, 1967]. А показатель акустического качества речи (2) отображается на экране компьютера в своем относительном (процентном) выражении

$$\delta_r^* = 100 \Big/ \left(1+\rho_r^*\right), \ \% . \qquad (3)$$

Его измерения проводятся в программе последовательно по коротким (20-30 сек.) отрезкам речевого сигнала. При этом были используются исключительно гласные звуки речи: "А", "О", "У", "И", "Ы" и "Э" как наиболее информативные среди всех других звуковых единиц в акустико-артикуляционном смысле [Белов, Белов, 2008; Савченко, 2017]. Преимуществом данной программы является высокая помехоустойчивость и минимальные требования к длительности анализируемого фрагмента речевого сигнала [Savchenko, 2017; Савченко, 2017]. В настоящий момент программа находится в открытом доступе по ссылке в Интернет https://sites.google.com/site/frompldcreators/. Ее главное окно показано на рис. 2.

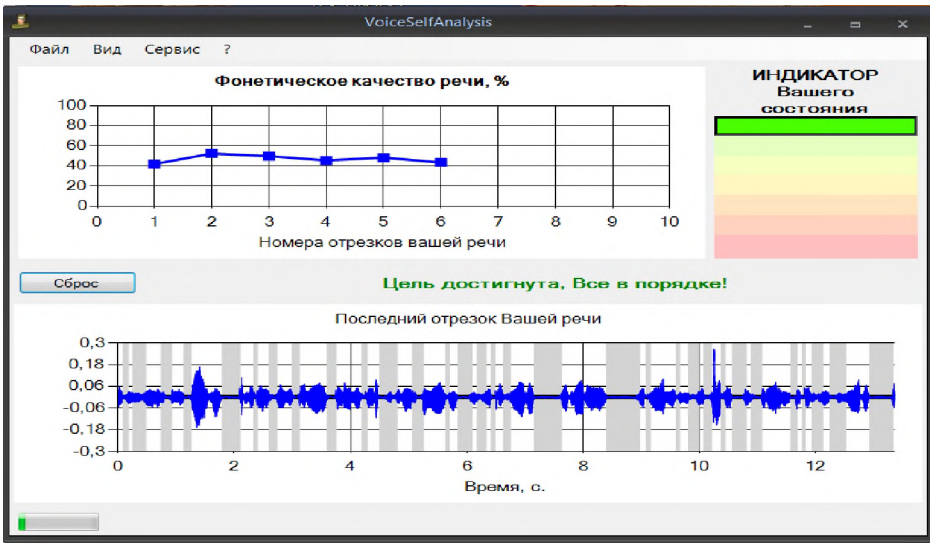

Рис. 2. Главное окно программы
Fig. 2. Main window of the computer program

График в верхней части окна отображает динамику показателя (3) в отношении одного (устанавливается пользователем программы) из шести гласных звуков речи. В данном случае это был звук "А". Последний из отрезков речевого сигнала показан во временной диаграмме в нижней части окна. А индикатор в его правой части предназначен для фиксации момента достижения данным диктором достаточно высокого качества в процессе непрерывного чтения. Указанный момент определяется в программе автоматически по принципу допустимой амплитуды колебаний показателя (3). Так, в случае, отображенном на рис. 2, вся процедура заняла менее двух минут.

Как видим из этого рисунка, минимум акустического качества речи [Савченко, Акатьев, 2017] имеет место в самом начале периода чтения текста, когда диктор еще

недостаточно сосредоточен. Напротив, максимум качества речи чаще всего фиксируется ближе к концу периода чтения контрольного текста диктором – в моменты максимальной концентрации его внимания. И, что характерно, в самом конце этого периода показатель (3), как правило, не сильно, но все же понижается – по мере естественного утомления диктора в процессе непрерывного чтения. Отметим, что зафиксированные на рис. 2 абсолютные значения показателя отражают индивидуальные особенности речи конкретного диктора

Программа экспериментальных исследований включала в себя оценку ФКР в пределах контрольной группы из 10 дикторов: [Савченко и др., 2017] все – примерно одинакового возраста (25-30 лет) и одного (высшего технического) уровня образования, без явно выраженных дефектов речи. Каждым из них был прочитан в среднем темпе один и тот же художественный текст – из первой главы романа А.С. Пушкина "Евгений Онегин" – объемом в одну стандартную машинописную страницу. Исследования выполнялись в три этапа:

- оценка ФКР контрольной группы дикторов в комфортных условиях;
- оценка степени влияния физической нагрузки на диктора на качество его речи;
- оценка степени влияния эмоционального напряжения диктора на качество его речи.

Для вычислений использовался современный ноутбук Asus N61D, 4 Гбайт ОЗУ, Windows 10, а также комплекс специальных аппаратных и программных средств, в том числе микрофон Sony. Частота дискретизации встроенного АЦП была установлена равной 8 кГц – стандартное значение при обработке разговорной речи. Это дало 160 тысяч отсчетов на каждый отрезок речевого сигнала длиной 20 сек. (нижняя временная диаграмма на рис. 2), или не менее L=80 тысяч отсчетов (50%) в расчете на гласные звуки. В пересчете к каждой из гласных на интервале квазистационарности речевого сигнала (10-15) мс в среднем получаем более 100 ее аллофонов ($J \geq 100$) на каждый отрезок. В таком случае точность оценки информационного показателя (3) может быть охарактеризована согласно работе [Савченко, 2015] погрешностью измерений (5-10)% на уровне значимости 0,1 и ниже. И это весьма высокий результат при учете суммарной длительности речевого сигнала 1-2 мин.

На первом этапе были созданы наиболее комфортные для каждого диктора условия: в домашних условиях и в отсутствии внешних шумов. Полученные результаты представлены на рис. 3 в виде трех гистограмм показателя ФКР (3) по результатам трех подряд испытаний одного и того же диктора.

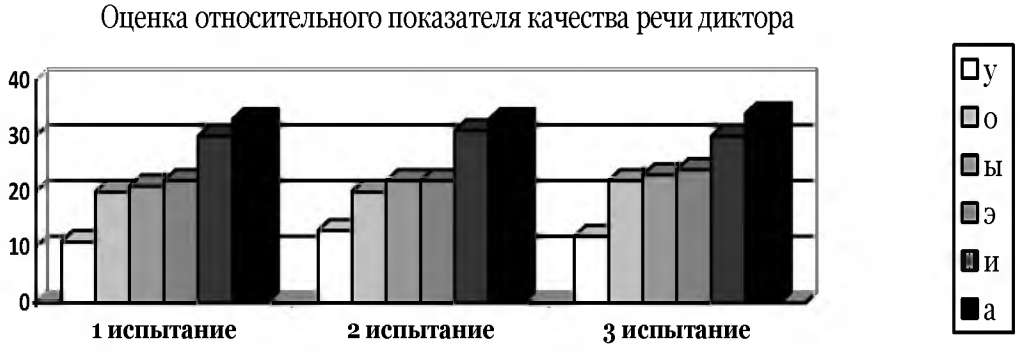

Рис. 3. Оценка фонетического качества речи диктора
Fig. 3. Evaluation of phonetic speech quality of speaker

Как видим, распределение ФКР по разным фонемам имеет ярко выраженный личностный характер и вместе с тем по каждой отдельной фонеме информационный показатель (2) практически инвариантен к виду текста, времени и длительности чтения. Все остальные дикторы из контрольной группы были охарактеризованы столь же устойчивыми во времени распределениями величины ФКР по фонемам.

На втором этапе исследований на каждого диктора была воздействована определенная физическая нагрузка: по 10 -20 приседаний с гантелями 3 кг в руках перед каждым сеансом работы с программой (по каждой гласной фонеме – отдельный сеанс). И только после этого диктор читал свой текст – с паузами между сеансами не менее 10 минут. Полученные оценки для одного из дикторов представлены на рис. 4.

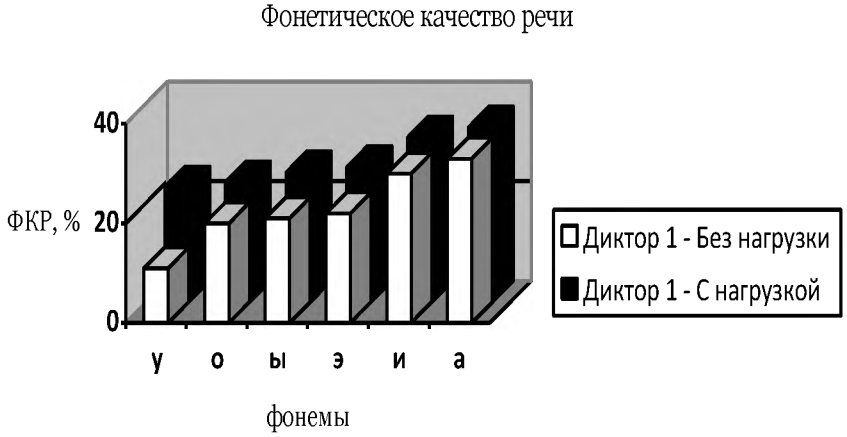

Рис. 4. Зависимость ФКР от физической нагрузки на диктора
Fig. 4. Dependence of phonetic speech quality on physical activity on speaker

Видно, что при воздействии физической нагрузки на диктора его ФКР пропорционально повысилось, причем, дружно по всем фонемам. Аналогичные результаты были получены по всем остальным дикторам из контрольной группы при очевидной оговорке на допустимые пределы физической нагрузки. Причем, для каждого диктора этот предел разный.

На третьем, заключительном этапе исследований был изменен характер нагрузка на дикторов: каждый из них прослушал достаточно громкую, энергичную музыку (в стиле "хард рок") – через наушники в течение 5 минут перед каждым сеансом чтения. Результаты проведенных на данном этапе измерений отражены на рис. 5.

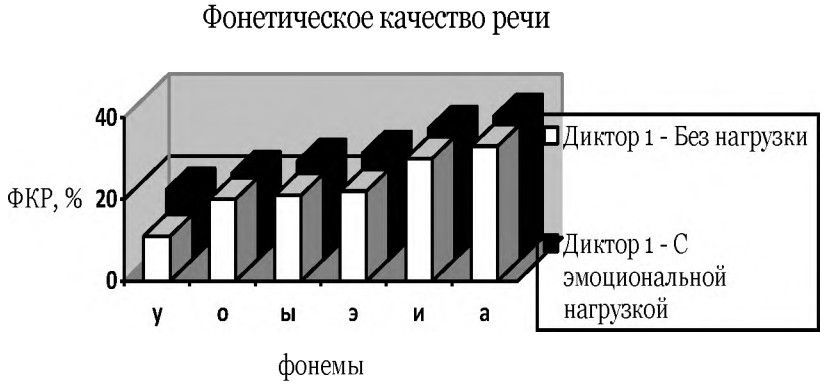

Рис. 5. Зависимость ФКР от эмоциональной нагрузки на диктора
Fig. 5. Dependence of phonetic speech quality on emotional activity on speaker

Хорошо видно, что эмоциональная нагрузка на диктора оказывает выборочное влияние на его ФКР в отношении разных звуков речи. Чаще всего ФКР повышается, предположим, если эмоциональная нагрузка не подавляет, а бодрит диктора.

Выводы

Таким образом, по результатам проведенных исследований в целом можно сделать следующие выводы:

- экспериментально подтверждена устойчивость информационного показателя ФКР диктора;

- экспериментально установлена прямо пропорциональная зависимость ФКР от интенсивности физической нагрузки на диктора, если она не выходит за допустимые пределы;

- установлена высокая чувствительность ФКР по отношению к эмоциональным нагрузкам на диктора в процессе чтения.

Полученные результаты и сделанные по ним выводы позволяют рекомендовать теоретико-информационную модель МРЕ и данный программный комплекс для практического применения при разработке и тестировании современных систем автоматической обработки речи. В частности, фонетические свойства речи могут служить надежной автоматической идентификации дикторов по голосу в системах защиты конфиденциальной информации.

Список литературы
References

1. Белов С.П., Белов А.С. 2008. О различиях частотных свойств информационных и неинформационных звуковых сигналов речевого диапазона. Научные ведомости БелГУ: Серия «Экономика. Информатика». 10(8): 86-93.

Belov S.P., Belov A.S., 2008. About distinction of frequency properties of informative and uninformative voice signal of vocal range. Nauchnye vedomosti Belgorodskogo gosydarstvennogo universiteta: Seriya “Ekonomika. Informatika” [Scientific bulletins of Belgorod state university: Series “Economy. Computer science”] 10(8): 86-93. (in Russian)

2. Савченко В.В. 2017. Метод измерения частоты основного тона с межпериодным накоплением речевого сигнала. Цифровая обработка сигналов. 2: 44-48.

Savchenko V.V. 2017. Method of measuring the pitch frequency with interperiodic accumulation of the speech signal. Tchifrovaya obrabotka signala. [Digital signal processing] 2: 44-48. (in Russian)

3. Савченко В.В. 2017. Метод фонетического декодирования слов с подавлением фонового шума Радиотехника и электроника. 62(7): 681-686.

Savchenko V.V. 2017. Method of phonetic decoding of words with suppression of background noise. Radiotechnika i elektronika. [Journal of communications technology and electronics] 62(7): 681-686. (in Russian)

4. Савченко В.В. 2015. Новая концепция программного обеспечения статистической обработки информации на основе прогностической функции теории вероятностей Научные ведомости БелГУ. Серия «Экономика. Информатика». 7(204): 84-88.

Savchenko V.V. 2015. A new concept of statistical information processing software based on the predictive function of probability theory. Nauchnye vedomosti Belgorodskogo gosydarstvennogo universiteta: Seriya “Ekonomika. Informatika” [Scientific bulletins of Belgorod state university: Series “Economy. Computer science”] 7(204): 84-88.

5. Савченко В.В. 2015. Определение объема контрольной выборки в условиях априорной неопределенности по принципу гарантированного результата. Научные ведомости БелГУ: Серия «Экономика. Информатика». 1(198): 74-78.

Savchenko V.V. 2015. The determination of sample size conditions of a priori uncertainty of the principle of guaranteed result. Nauchnye vedomosti Belgorodskogo gosydarstvennogo universiteta: Seriya “Ekonomika. Informatika” [Scientific bulletins of Belgorod state university: Series “Economy. Computer science”] 1(198): 74-78. (in Russian)

6. Савченко В.В. 2016. Повышение помехоустойчивости системы голосового управления робототехникой на основе метода фонетического декодирования слов Радиотехника и электроника. 12: 1196-1201.

Savchenko V.V. 2016. Enhancement of the noise immunity of a voice-activated robotics control system based on phonetic word decoding method. Radiotechnika iI elektronika [Journal of communications technology and electronics] 12: 1196-1201. (in Russian)

7. Савченко В.В. 2016. Распознавание речевых команд методом фонетического декодирования слов с подавлением фонового шума. Информационные технологии. 1: 76—80.

Savchenko V.V. 2016. The speech recognition method of phonetic decoding of words with background noise cancellation. Informatcionnye technologii [Information technology] 1: 76—80. (in Russian)

8. Савченко В.В. 2016. Распознавание речи на фоне шума методом фонетического декодирования слов. Телекоммуникации. 9: 16-25.

Savchenko V.V. 2016. Speech recognition in noise background by method of phonetic decoding of words Telekommunikatchii [Telecommunications and Radio Engineering] 9: 16-25. (in Russian)

9. Савченко В.В. 2017. Тестирование вокодера по критерию минимума требуемой избыточности речевого сигнала. Телекоммуникации. 1: 17-25.

Savchenko V.V. 2017. Vocoder testing by criterion of minimum required redundancy of speech signal. Telekommunikatchii [Telecommunications and Radio Engineering] 1: 17-25. (in Russian)

10. Савченко В.В. 2017. Исследование стационарности случайных временных рядов с использованием принципа минимума информационного рассогласования. Известия вузов. Радиофизика. 60(1): 89-96.

Savchenko V.V. 2017. A study of stationarity of the random time series using the principle of the information divergence minimum.Izvestiya vyshich uchebnych zavedeniy. Radiofizika [Radiophysics and Quantum Electronics]. 60(1): 89-96. (in Russian)

11. Савченко В.В., Акатьев Д.Ю. 2013. Адаптивная кластерная модель минимальных речевых единиц в задачах анализа и распознавания речи. Наука и образование. 2: 323-334.

Savchenko V.V., Akatyev D.Yu. 2013. Adaptive cluster model of minimal speech units in analysis and speech recognition problems. Nauka I obrazovanie [Science and education]. 2:323-334. (in Russian)

12. Савченко В.В., Акатьев Д.Ю. 2017. Информационная технология речевого профайлинга Научные ведомости БелГУ. Серия: «Экономика. Информатика». № 9 (258): 157-165.

Savchenko V.V., Akatyev D.Yu. 2017. Information technology of speech profiling Nauchnye vedomosti Belgorodskogo gosydarstvennogo universiteta: Seriya “Ekonomika. Informatika” [Scientific bulletins of Belgorod state university: Series “Economy. Computer science”] 9 (258): 157-165. (in Russian)

13. Савченко В.В., Акатьев Д.Ю. 2017. Информационная технология психокоррекции эмоционального состояния пользователя на основе его устного чтения. Информационные технологии. 23(11): 771-775.

Savchenko V.V., Akatyev D.Yu. 2017. Information Technology of Psychocorrection of an User Emotional State Based on his Oral Reading Informatcionnye technologii [Information technology]. 23(11): 771-775. (in Russian)

14. Савченко В.В., Васильев Р.А. 2014. Анализ эмоционального состояния диктора по голосу на основе фонетического детектора лжи. Научные ведомости БелГУ: Серия «Экономика. Информатика». 32 (21): 186-194

Savchenko V.V., Vasyliev R.A. 2014. The analysis of the emotional condition of the announcer on the voice on the basis of the phonetic lie detector. Nauchnye vedomosti Belgorodskogo gosydarstvennogo universiteta: Seriya “Ekonomika. Informatika” [Scientific bulletins of Belgorod state university: Series “Economy. Computer science”] 32 (21): 186-194. (in Russian)

15. Савченко В.В., Акатьев Д.Ю., Кривошеев Д.В. 2017. Экспериментальные исследования акустического качества речевого сигнала на основе его информационных коррелятов Телекоммуникации. 7: 12-16.

Savchenko V.V., Akatyev D.Yu., Krivosheyev D.V., 2017. Experimental investigations of acoustic quality of speech signal, based on its information correlates Telekommunikatchii [Telecommunications and Radio Engineering] 7: 12-16. (in Russian)

16. Kullback S. 1967. Information Theory and Statistics. Professional Lecturer in Statistics The George Washington University, 404

17. Savchenko V.V. 2017. Study of the Stationarity of Random Time Series Using the Principle of the Information-Divergence Minimum. Radiophysics and Quantum Electronics. Vol. 60. No. 1. pp. 89–96.

18. Savchenko V.V. 2015. The Principle of the Information-Divergence Minimum in the Problem of Spectral Analysis of the Random Time Series Under the Condition of Small Observation Samples. Radiophysics and Quantum Electronics. Vol. 58. No. 5. P. 373-379.

19. Savchenko V.V. 2016. Enhancement of the Noise Immunity of a Voice-Activated Robotics Control System Based on Phonetic Word Decoding Method. Journal of Communications Technology and Electronics. Vol. 61, No. 12. p. 1374 -1379.

20. Savchenko V.V. 2017. Words Phonetic Decoding Method with the Suppression of Background Noise. Journal of Communications Technology and Electronics. Vol. 62, No. 7. pp. 788–793.

21. Savchenko V.V., Savchenko A.V. 2016. Information Theoretic Analysis of Efficiency of the Phonetic Encoding–Decoding Method in Automatic Speech Recognition. Journal of Communications Technology and Electronics. Vol. 61. No. 4. P. 430-435.