

Использование формантных характеристик гласных и согласных звуков русского языка для аутентификации пользователей

Е. П. Белова

Уфимский университет науки и технологий (УУНиТ)

Аннотация: В статье предлагается метод аутентификации пользователей на основе формантных характеристик. Разработан биометрический образ, состоящий из формантных характеристик 46 звуков русского языка (6 гласных, 4 дифтонгов и 36 согласных). Предложена архитектура системы аутентификации по формантным характеристикам и частоте лидирующей форманты гласных и согласных звуков. Разработан модуль выделения формантных характеристик гласных и согласных звуков. Представлены структурная схема модуля выделения формантных характеристик гласных и согласных звуков и блок-схемы его подмодулей. Алгоритм «случайный лес», метод k-ближайших соседей, логистическая регрессия, многослойный перцептрон и метод опорных векторов объединены в группу алгоритмов. Приведена структурная схема обучения и тестирования моделей на основе группы алгоритмов. На основе группы алгоритмов обучены модели. Получены результаты эксперимента и проведён их анализ.

Ключевые слова: биометрическая аутентификация по голосу, биометрическая система аутентификации, биометрический образ пользователя, формантные характеристики, гласные звуки, дифтонги, согласные звуки.

Для цитирования: Белова Е. П. Использование формантных характеристик гласных и согласных звуков русского языка для аутентификации пользователей // Вестник СибГУТИ. 2024. Т. 18, № 1. С. 59–69. <https://doi.org/10.55648/1998-6920-2024-18-1-59-69>.



Контент доступен под лицензией
Creative Commons Attribution 4.0
License

© Белова Е. П., 2024

Статья поступила в редакцию 07.08.2023;
переработанный вариант – 29.09.2023;
принята к публикации 01.10.2023.

1. Введение

В последнее время популярность набирают биометрические методы аутентификации. В отличие от парольных методов аутентификации и токена, они не требуют запоминания или хранения. В основе каждого биометрического метода аутентификации находится биометрический образ личности, основанный на одном или нескольких биометрических характеристиках. Каждый биометрический образ уникален, что обеспечивает высокую степень безопасности информационной системы. В большинстве случаев исключается использование биометрических характеристик пользователя другими лицами.

Биометрическая аутентификация по голосу не требует дополнительного оборудования и достаточно проста для применения пользователями. Одним из перспективных методов аутентификации по голосу является аутентификация по формантным характеристикам звуков речи человека.

Эффективный алгоритм выделения формантных характеристик представлен в работе [1]. Использование формантного анализа для задач биометрической аутентификации исследуется в статьях [2, 3]. Применению формантных характеристик звуков речи посвящены [4–6].

Исследованию формантных характеристик гласных звуков посвящены работы [7–8].

Целью данной работы является исследование формантных характеристик гласных и согласных звуков для повышения эффективности работы биометрической системы аутентификации.

2. Биометрический образ личности, основанный на формантных характеристиках гласных и согласных звуков

Целью данной работы является исследование формантных характеристик гласных и согласных звуков для повышения эффективности работы биометрической системы аутентификации.

Исходя из общей теории фонетики [9–11] и акустической теории речевого производства [12], под формантой понимается частотный диапазон звуков, содержащий усиленный звуковой компонент.

Форманта обычно образуется из-за резонанса в дыхательных путях при произнесении звука.

Под формантными характеристиками подразумеваются параметры формант, определяющие их частоту, амплитуду и ширину.

Биометрический образ личности в данном исследовании составляют следующие формантные характеристики:

1. Частоты первых 7 формант – $F_1, F_2, F_3, F_4, F_5, F_6, F_7$.
2. Начало и конец частотного диапазона каждой из 7 формант – $F_{1s}, F_{1f}, F_{2s}, F_{2f}, F_{3s}, F_{3f}, F_{4s}, F_{4f}, F_{5s}, F_{5f}, F_{6s}, F_{6f}, F_{7s}$ и F_{7f} .
3. Частота лидирующей форманты – F_l .

Под частотой лидирующей форманты подразумевается частота, которая соответствует наибольшему значению амплитуды выброса в любой из фиксируемых на спектрограмме формант [8].

В данном исследовании используются формантные характеристики 46 звуков (6 гласных, 4 дифтонгов и 36 согласных звуков), то есть все звуки русского языка.

Стоит отметить, что согласные звуки имеют менее выраженные форманты. Поэтому они рассматриваются главным образом посредством их артикуляции.

На рис. 1 приведена спектрограмма звука «Э».

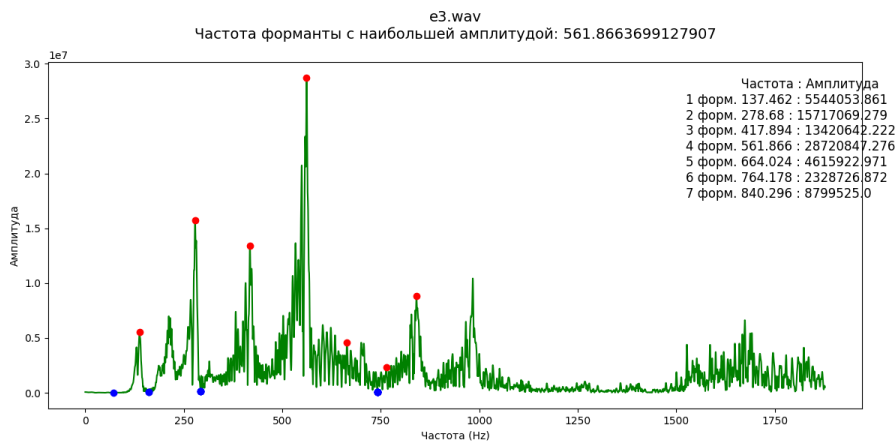


Рис. 1. Спектрограмма звука «Э»

На рис. 2 представлена спектрограмма звука «Т».



Рис. 2. Спектрограмма звука «Т»

Как видно из рисунков, спектрограмма звука «Т» имеет ярко выраженные форманты.

3. Разработка модуля выделения формантных характеристик гласных и согласных звуков

На рис. 3 представлена архитектура системы аутентификации по формантным характеристикам и частоте лидирующей форманты гласных и согласных звуков.

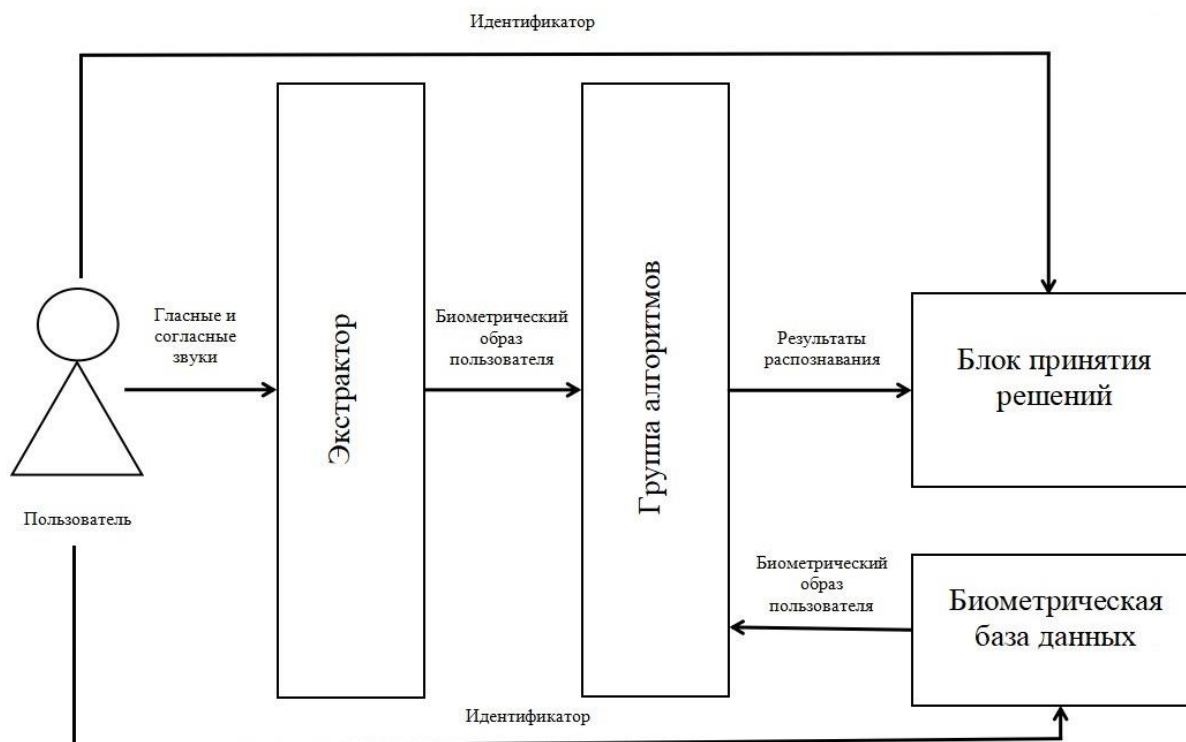


Рис. 3. Архитектура системы аутентификации по формантным характеристикам и частоте лидирующей форманты гласных и согласных звуков

Группа алгоритмов получает биометрический образ пользователя из экстрактора, также называемого модулем выделения формантных характеристик гласных и согласных звуков.

Структурная схема модуля выделения формантных характеристик гласных и согласных звуков приведена на рис. 4.



Рис. 4. Структурная схема модуля выделения формантных характеристик гласных и согласных звуков

Модуль выделения формантных характеристик гласных и согласных звуков состоит из следующих подмодулей:

- main – задаёт основное меню;
- recognition – строит спектрограммы, ищет форманты и определяет их характеристики;
- detect_peaks – непосредственно занимается поиском пиков.

Также модуль содержит папку Results, в которую записываются результаты.

Блок-схема подмодуля main представлена на рис. 5.

Подмодуль main создаёт окно программы, папку для записи результатов и кнопки меню. После того как происходит запись звука, подмодуль main запускает в работу подмодуль recognition.

Блок-схема подмодуля recognition представлена на рис. 6.

Подмодуль recognition запускает алгоритм поиска пиков через подмодуль detect_peaks и согласно полученным результатам строит спектрограммы.

На рис. 7 представлена блок-схема подмодуля detect_peaks.

Подмодуль detect_peaks осуществляет поиск пиков при помощи быстрого преобразования Фурье.

Выгрузка формантных характеристик гласных и согласных звуков в таблицу формата csv идёт последовательно. Порядок следования гласных и согласных звуков задаётся в подмодуле main.

Для обработки в приложении названия гласных и согласных звуков должны быть написаны латиницей. Латинские названия звуков представлены в столбце «Название папки».

В результате работы модуля выделения формантных характеристик гласных и согласных звуков получена выборка.

4. Проектирование группы алгоритмов

Группа алгоритмов состоит из:

- 1) алгоритма «случайный лес»;
- 2) метода k-ближайших соседей;
- 3) логистической регрессии;
- 4) многослойного перцептрона;
- 5) метода опорных векторов.



Рис. 5. Блок-схема подмодуля main

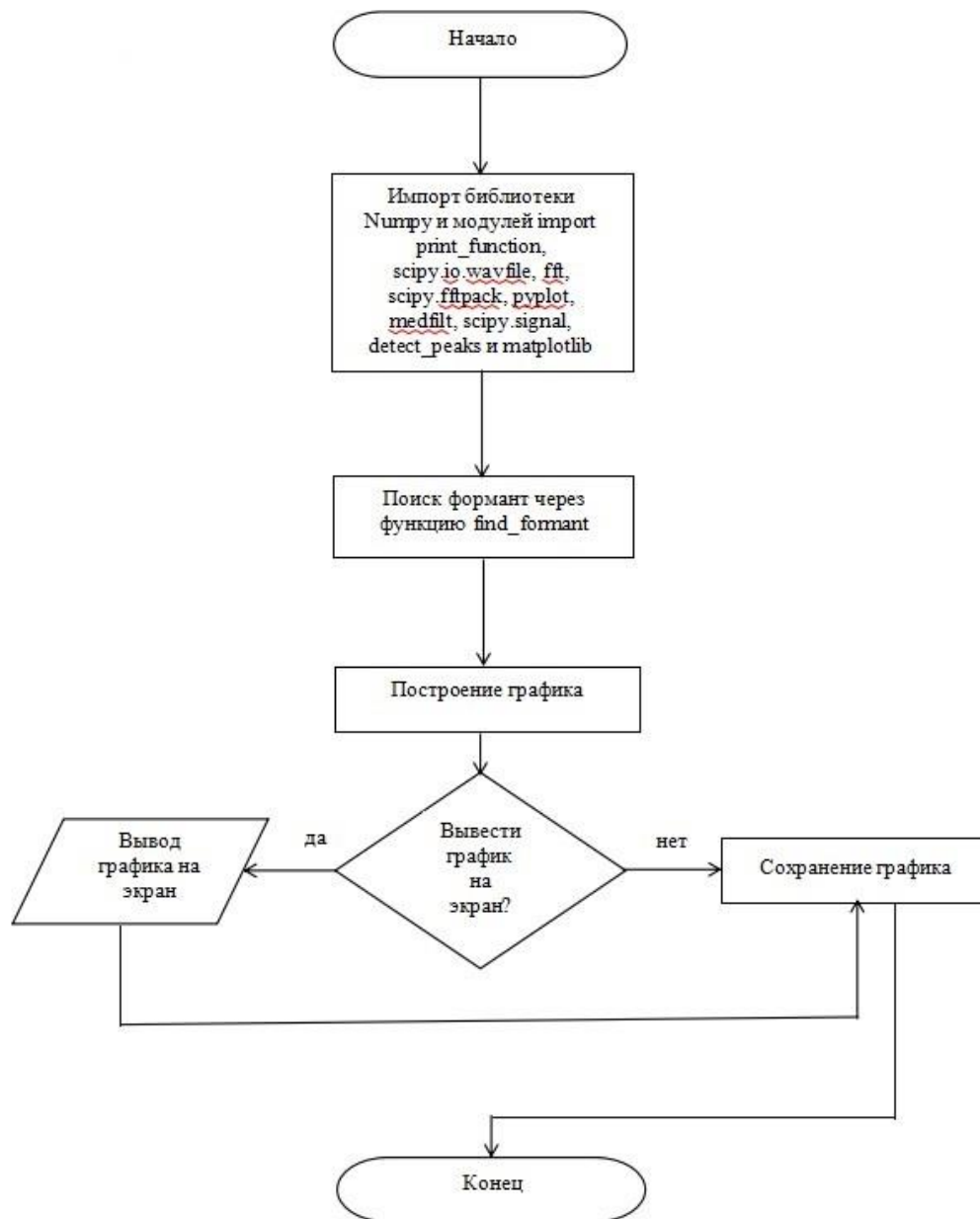


Рис. 6. Блок-схема подмодуля recognition



Рис. 7. Блок-схема подмодуля detect_peaks

На рис. 8 представлена структурная схема обучения и тестирования моделей на основе группы, состоящей из метода опорных векторов (SVM), логистической регрессии (Log Reg), метода k-ближайших соседей (KNN), алгоритма «случайный лес» (RF) и многослойного перцептрона (MLP).

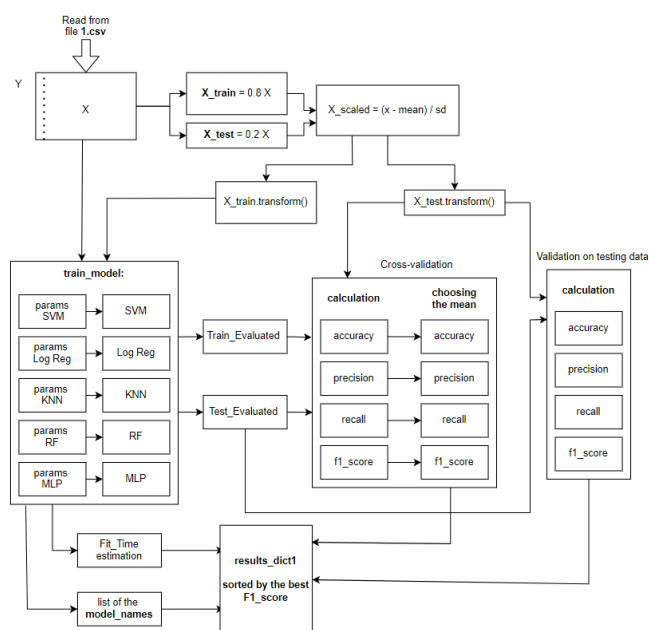


Рис. 8. Структурная схема обучения и тестирования моделей на основе группы алгоритмов

Проведение процедуры аутентификации при помощи группы алгоритмов состоит из следующих этапов:

1. На вход группы алгоритмов подаётся выборка, состоящая из формантных характеристик гласных (в том числе и дифтонгов) и согласных звуков.
2. Осуществляется разделение выборки на обучающую и тестовую в соотношении 80 % к 20 %.
3. Представленные данные проходят через процедуру усреднения.
4. Обучающая выборка служит для обучения моделей на основе метода опорных векторов (SVM), логистической регрессии (Log Reg), метода k-ближайших соседей (KNN), алгоритма «случайный лес» (RF) и многослойного перцептрона (MLP), для этого она поступает на входы соответствующих формируемых моделей.
5. Для каждого выбранного алгоритма во время обучения происходит расчёт метрик и выбор лучшей из них.
6. Формируется отчёт по обучению моделей на основе группы алгоритмов.
7. На вход обученных моделей поступает тестовая выборка.
8. Данные тестовой выборки при помощи метрик сравниваются с результатами обучения.
9. На основе полученных результатов формируется отчёт по тестированию моделей.

5. Анализ полученных результатов

Для анализа эффективности работы группы алгоритмов выбраны следующие параметры:

1. Среднее время подсчёта – это время, за которое свёрточная искусственная нейронная сеть решает 1 пример. Измеряется в секундах.
2. Среднее значение точности – это значение точности работы искусственной нейронной сети на обучающих и тестовых примерах.
3. Среднее значение точности на тестовой выборке.

В табл. 1 представлены результаты использования биометрического образа пользователя, основанного на формантных характеристиках гласных и согласных звуков.

Таблица 1. Результаты использования биометрического образа пользователя, основанного на формантных характеристиках гласных и согласных звуков

Model Name	Среднее время подсчёта	Среднее значение точности	Среднее значение точности на тестовой выборке
Алгоритм «случайный лес»	0.856373	0.871467	0.891304
Метод k-ближайших соседей	0.003199	0.704348	0.729348
Логическая регрессия	0.138321	0.533424	0.513043
Многослойный перцептрон	34.335736	0.797554	0.816304
Метод опорных векторов	0.149714	0.508152	0.548913

Большее количество времени занял подсчёт результатов у многослойного перцептрона – 34.335736 с. на решение одного примера.

Максимальное среднее значение точности (0.871467) и максимальное среднее значение точности на тестовой выборке (0.891304) получены с помощью алгоритма «случайный лес».

6. Заключение

Данное исследование посвящено изучению использования формантных характеристик гласных и согласных звуков русского языка для процедуры аутентификации по голосу. Всего использовано 46 звуков: 10 гласных (в том числе 4 дифтонга) и 36 согласных.

Для выделения формантных характеристик использовался модуль выделения формантных характеристик гласных и согласных звуков.

Процедура биометрической аутентификации проходила при помощи группы алгоритмов: алгоритма «случайный лес», метода k-ближайших соседей, логистической регрессии, многослойного перцептрона и метода опорных векторов. Каждый из этих алгоритмов имеет свои достоинства для решения задач по распознаванию пользователя.

Результаты, полученные при тестирования разработанных по данным алгоритмам моделей, находятся на оптимальном уровне.

Для повышения эффективности распознавания в следующих работах группа алгоритмов будет объединена в ансамбль. То есть все модели, обученные по алгоритмам, будут взаимосвязаны между собой и будут обучаться с учётом результатов друг друга. Планируется проведение работ по уменьшению ошибки 1-го рода, также будет проведён анализ и вычисление ошибки 2-го рода.

Литература

1. *Оганесян А. Г.* Эффективный алгоритм выделения формант из спектра речевого сигнала // Вычислительная техника и информатика. 2006. Т. LIX, № 1. С. 177–183.
2. *Сысоев О. В., Глаздовский В. В., Курпатов В. Л.* Исследование возможности применения формантных характеристик в задачах биометрической аутентификации // Сибирский информационно-коммуникационный журнал. 2017. № 1 (63). С. 93–97.
3. *Шохин А. А., Седельников С. Ю.* Методика формантного анализа для биометрической аутентификации пользователей // Информатика и вычислительная техника. 2018. № 4. С. 36–41.
4. *Peterson G. E. and Barney H. L.* Control methods used in a study of the vowels // The Journal of the Acoustical Society of America. 1952. V. 24 (2). P. 175–184.
5. *Бабосов Е. Г., Авдеев А. В., Зуева Е. В.* Исследование возможности использования формантных характеристик звуков речи для аутентификации пользователей // Медико-биологические и технические системы. 2015. № 3 (79). С. 172–178.
6. *Никитин А. И., Поливодский А. А.* Исследование возможности использования формантных характеристик гласных в процессе аутентификации // Новые информационные технологии. 2016. № 1. С. 37–43.
7. *Машкина И. В., Белова Е. П.* Разработка нейросетевой базы данных биометрических образов для системы аутентификации по голосу // Проблемы информационной безопасности. Компьютерные системы. 2019. № 2. С. 86–93.
8. *Belova Ye. P., Mashkina I. V.* Architecture and Work Algorithm of Biometric Authentication System by Voice // Automatic Control and Computer Sciences. 2021. V. 55, №. 8. P. 924–931. DOI: 10.3103/S014641162108006X.
9. *Ladefoged P. and Johnson K.* A course in phonetics. 6th ed. Wadsworth Cengage Learning, 2011.
10. *Stevens K. N.* Acoustic phonetics (Current studies in linguistics series). MIT Press, 1998.
11. *Ohala J. J.* Phonetic explanations for the development of tones // Language. 1972. V. 48 (2). P. 245–261.
12. *Fant G.* Acoustic theory of speech production. Mouton, 1960.

Белова Елена Петровна

старший преподаватель кафедры управления информационной безопасностью, Уфимский университет науки и технологий (УУНиТ), e-mail: yelenabelova92@gmail.com, ORCID ID: 0000-0001-5488-943X.

Автор прочитала и одобрила окончательный вариант рукописи.

Автор заявляет об отсутствии конфликта интересов.

Using Formant Characteristics of Russian Vowels and Consonants for User Authentication

Yelena P. Belova

Ufa University of Science and Technology (UUST)

Abstract: The article proposes a method for user authentication based on formant characteristics. A biometric image has been developed consisting of the formant characteristics of 46 sounds of the Russian language (6 vowels, 4 diphthongs and 36 consonants). An authentication system architecture based on formant characteristics and frequency of the leading formant of vowels and consonants is proposed. A Module for extracting the formant characteristics of vowels and consonants has been developed. A block diagram of the Module for extracting the formant characteristics of vowels and consonants and block diagrams of its submodules are presented. The Random Forest algorithm, k-nearest neighbors, logistic regression, multilayer perceptron and Support Vector Machine are combined into a cluster of algorithms. A block diagram of training and testing models based on a cluster of algorithms is given. Models were trained on the basis of algorithms clusters. The results of the experiment were obtained and their analysis was carried out.

Keywords: biometric voice authentication, biometric authentication system, biometric user image, formant characteristics, vowels, diphthongs, consonants.

For citation: Belova Ye. P. Using formant characteristics of Russian vowels and consonants for user authentication (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 1, pp. 59-69. <https://doi.org/10.55648/1998-6920-2024-18-1-59-69>.



Content is available under the license
Creative Commons Attribution 4.0
License

© Belova Ye. P., 2024

The article was submitted: 07.08.2023;
revised version: 29.09.2023;
accepted for publication 01.10.2023.

References

1. Oganessian A. G. Effektivnyi algoritm vydeleniya formant iz spektra rechevogo signala [Efficient algorithm for extracting formants from the spectrum of a speech signal]. *Vychislitel'naya tekhnika i informatika*, Moscow, 2006, vol. LIX, no.1, pp. 177-183.
2. Sysoev O. V., Glauzdovskii V. V., Kurpatov V. L. Issledovanie vozmozhnosti primeneniya formantnykh kharakteristik v zadachakh biometricheskoi autentifikatsii [Study of the possibility of using formant characteristics in biometric authentication tasks] *Sibirskii informatsionno-kommunikatsionnyi zhurnal*, 2017, no. 1 (63), pp. 93-97.

3. Shokhin A. A., Sedel'nikov S. Yu. Metodika formantnogo analiza dlya biometricheskoi autentifikatsii pol'zovatelei [Methods of formant analysis for biometric user authentication] *Informatika i vychislitel'naya tekhnika*, 2018, no. 4, pp. 36-41.
4. Peterson G. E., Barney H. L. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 1952, no. 24(2), pp. 175-184.
5. Babosov E. G., Avdeev A. V., Zueva E. V. Issledovanie vozmozhnosti ispol'zovaniya formantnykh kharakteristik zvukov rechi dlya autentifikatsii pol'zovatelei [Study of the possibility of using the formant characteristics of speech sounds for user authentication] *Mediko-biologicheskie i tekhnicheskie sistemy*, 2015, no. 3 (79), pp. 172-178.
6. Nikitin A. I., Polivodskii A. A. Issledovanie vozmozhnosti ispol'zovaniya formantnykh kharakteristik glasnykh v protsesse autentifikatsii [Study of the possibility of using the formant characteristics of vowels in the process of authentication]. *Novye informatsionnye tekhnologii*, 2016, no. 1, pp. 37-43.
7. Mashkina I. V., Belova E. P. Razrabotka neirosetevoi bazy dannykh biometricheskikh obrazov dlya sistemy autentifikatsii po golosu [Development of a neural network database of biometric images for a voice authentication system]. *Problemy informatsionnoi bezopasnosti. Komp'yuternye sistemy*, St. Petersburg, 2019, no. 2, pp. 86-93.
8. Belova Ye. P., Mashkina I. V. Architecture and Work Algorithm of Biometric Authentication System by Voice. *Automatic Control and Computer Sciences*, 2021, vol. 55, no. 8, pp. 924-931, DOI: 10.3103/S014641162108006X.
9. Ladefoged P., Johnson K. *A course in phonetics* (6th ed.), Wadsworth Cengage Learning, 2011.
10. Stevens K. N. *Acoustic phonetics* (Current studies in linguistics series), MIT Press, 1998.
11. Ohala J. J. Phonetic explanations for the development of tones. *Language*, 1972, no. 48(2), pp. 245-261.
12. Fant, G. *Acoustic theory of speech production*, Mouton, 1960.

Yelena P. Belova

Senior lecturer of the Department of Information Security Management, Ufa University of Science and Technology (UUST), e-mail: yelenabelova92@gmail.com, ORCID ID: 0000-0001-5488-943X.