УДК 004.934.1
doi: 10.21685/2307-5538-2023-4-5

МЕТОДЫ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ

В. В. Козлов[1], А. А. Трофимов[2], Е. А. Фокина[3], В. Н. Пономарев[4], Т. О. Жуков[5]

[1, 2, 3, 4, 5] Пензенский государственный университет, Пенза, Россия
[1] val.iit@mail.ru, [2, 5] iit@pnzgu.ru, [3] ekaterina.isay1997@gmail.com, [4] revik2296@gmail.com

Аннотация. *Актуальность и цели.* Технология автоматического распознавания речи сталкивается с большим количеством трудностей. Сложная структура человеческой речи в совокупности с различными шумовыми эффектами делает эту задачу очень трудоемкой. Для упрощения данной задачи применяют предварительную обработку речевых сигналов, которая может осуществляться с помощью различных методов. *Материалы и методы.* Предварительная обработка речевых сигналов включает в себя ряд техник и методов, направленных на улучшение качества речевых данных и подготовку их для дальнейшего анализа. Рассмотрены методы предварительной обработки речевых сигналов и приведены направления применения данных методов. Представлен обзор некоторых техник предварительной обработки речевых сигналов. *Результаты и выводы.* Представлен анализ наиболее подходящих методов для предварительной обработки речевых сигналов.

Ключевые слова: речевые сигналы, предварительная обработка, декомпозиция на эмпирические моды, вейвлет-преобразование, преобразование Фурье

Для цитирования: Козлов В. В., Трофимов А. А., Фокина Е. А., Пономарев В. Н., Жуков Т. О. Методы предварительной обработки речевых сигналов // Измерение. Мониторинг. Управление. Контроль. 2023. № 4. С. 43–49. doi: 10.21685/2307-5538-2023-4-5

METHODS OF PREPROCESSING SPEECH SIGNALS

V.V. Kozlov[1], A.A. Trofimov[2], E.A. Fokina[3], V.N. Ponomarev[4], T.O. Zhukov[5]

[1, 2, 3, 4, 5] Penza State University, Penza, Russia
[1] val.iit@mail.ru, [2, 5] iit@pnzgu.ru, [3] ekaterina.isay1997@gmail.com, [4] revik2296@gmail.com

Abstract. *Background.* Automatic speech recognition technology faces a lot of difficulties. The complex structure of human speech combined with various noise effects makes this task very time-consuming. To simplify this task, preprocessing of speech signals is used, which can be carried out using various methods. *Materials and methods.* Preprocessing of speech signals includes a number of techniques and methods aimed at improving the quality of speech data and preparing them for further analysis. The article discusses the methods of preprocessing speech signals and provides directions for the application of these methods. An overview of some techniques of preprocessing speech signals is presented. *Results and conclusions.* The analysis of the most suitable methods for preprocessing speech signals is presented.

Keywords: speech signals, preprocessing, empirical modes decomposition, wavelet transform, Fourier transform

For citation: Kozlov V.V., Trofimov A.A., Fokina E.A., Ponomarev V.N., Zhukov T.O. Methods of preprocessing speech signals. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measuring. Monitoring. Management. Control.* 2023;(4):43–49. (In Russ.). doi: 10.21685/2307-5538-2023-4-5

Введение

Технология автоматического распознавания речи сталкивается с большим количеством трудностей. Из-за сложности речевого сигнала время его обработки и анализа оказывается значительным. При занижении качества сигнала, для уменьшения времени обработки, происходит снижение точности распознавания. В настоящее время технология распознавания речи способна достичь высокой точности только в идеальных условиях. Сложная структура человеческой речи в совокупности с различными шумовыми эффектами делает эту задачу одной из самых трудоемких областей компьютерных наук, которая включает в себя лингвистику, математику

и статистику. Распознаватели речи состоят из нескольких компонентов, таких как ввод речи, предварительная обработка, извлечение признаков, векторы признаков, декодер и вывод слов.

Предварительная обработка речевых сигналов является важным этапом в процессе их анализа и распознавания. Она включает в себя ряд техник и методов, направленных на улучшение качества речевых данных и подготовку их для дальнейшего анализа. В данной статье рассмотрим основные аспекты предварительной обработки речевых сигналов.

Цели предварительной обработки

Речевой сигнал является аналоговым, поэтому для дальнейшей обработки его сначала преобразуют в дискретные сигналы, а затем представляют в виде зависимости амплитуды от дискретных отсчетов времени. Из-за сложности речевого сигнала, а именно из-за его нестационарности, предварительная обработка – один из важных шагов при распознавании речи, которая преследует следующие цели.

Одной из важных задач является устранение шумов, так как шум в речевых сигналах может сильно снижать качество распознавания и анализа. Предварительная обработка включает в себя методы фильтрации и устранения шумовых компонент. Сначала происходит удаление сильных фоновых шумов, например, если речь записана на фоне музыки или шума движения, то можно произвести фильтрацию или усреднение для уменьшения влияния этих шумовых компонент. Затем при необходимости производится фильтрация не интересующих нас частотных компонент. При использовании включают фильтры нижних и верхних частот, фильтры скользящего среднего и фильтры Калмана.

Для неравномерных по уровню сигналов проводится нормализация громкости, так как исследуемый сигнал может иметь разную громкость в разные моменты времени, а нормализация громкости приводит сигнал к стандартизированному уровню громкости (например, приведения его к определенному уровню или максимальной амплитуде) для более надежного анализа.

Зачастую для более качественной обработки речевого сигнала производится извлечение признаков (Feature Extraction), т.е. таких характеристик, как спектральные признаки или временные характеристики, чтобы создать векторы признаков для упрощения дальнейшей классификации или распознавания. При преобразовании временного сигнала в набор характеристик выбираются те, которые легче анализировать. Данный процесс может включать в себя выделение мел-кепстральных коэффициентов (MFCC), энергии сигнала, скорости изменения и др.

Также в процессе распознавания могут применяться некоторые техники предварительной обработки речевых сигналов:

– амплификация (усиление сигнала) – представляет собой увеличение амплитуды сигнала, что может помочь улучшить отношение сигнал-шум и делать его более слышимым;

– удаление силенсов и пауз – данная техника позволяет удалить периоды тишины и пауз в речи;

– устранение эха – данное действие помогает в случае использования микрофонов и акустических систем, где может потребоваться устранение эха для улучшения качества аудио;

– разделение на фразы и слова – представляет собой разбиение речевого сигнала на отдельные фразы и слова, что может потребоваться для более точного анализа и распознавания;

– выравнивание продолжительности – при анализе нескольких речевых сигналов, например в системах распознавания речи, может потребоваться выравнивание продолжительности сигналов для сравнения;

– компенсация шума – использование алгоритмов для выделения речи и уменьшения влияния окружающего шума.

Выбор методов предварительной обработки зависит от конкретной задачи и характеристик сигнала. Эти методы могут быть комбинированы для достижения наилучших результатов в конкретном контексте.

Методы предварительной обработки речевых сигналов

Одним из классических является метод преобразования Фурье, который представляет сигнал, заданный во временной области, в виде разложения по ортогональным базисным

функциям (синусам и косинусам), выделяя, таким образом, частотные составляющие [1]. Результат преобразования Фурье – амплитудно-частотный спектр, по которому можно определить присутствие некоторой частоты в анализируемом сигнале. Преобразование Фурье дает достаточно простые для расчетов формулы и прозрачную интерпретацию результатов:

$$F(\omega)=\int_{-\infty}^{+\infty} f(t)e^{-j\omega t}dt;\ f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)e^{j\omega t}d\omega,$$

где $F(\omega)$ – сигнал в частотной области; $f(t)$ – сигнал во временной области; j – мнимая единица.

Недостаток преобразования заключается в том, что частотные компоненты не могут быть локализованы во времени, что накладывает ограничения на применимость данного метода к ряду задач (например, в случае анализа динамики изменения частотных параметров сигнала на временном интервале).

Преобразование Фурье может использоваться в предварительной обработке речевых сигналов для анализа и улучшения их частотных характеристик. В данном контексте преобразование Фурье может применяться для следующих целей:

1. Построение спектрограммы. Преобразование Фурье можно применить к речевому сигналу, чтобы получить его частотное представление. Это позволяет создать спектрограмму, которая отображает, какие частоты преобладают в сигнале в разные моменты времени. Спектрограмма полезна для визуальной оценки частотных характеристик речи и может помочь в выделении формант (резонансных частотных областей), что важно для распознавания речи.

2. Фильтрация. Применение фильтрации в частотной области с использованием преобразования Фурье позволяет удалять шумы и нежелательные частотные компоненты из речевого сигнала. Например, можно использовать фильтры нижних и верхних частот для улучшения качества сигнала и уменьшения влияния шума [2].

3. Извлечение частотных характеристик. Преобразование Фурье также может использоваться для извлечения конкретных частотных характеристик из речевого сигнала. Например, можно извлекать пики в спектре для определения наиболее выраженных частотных компонентов, которые могут содержать информацию о звуках речи.

4. Обработка в частотной области. В некоторых случаях можно провести обработку и модификацию сигнала в частотной области с использованием преобразования Фурье, например, уменьшение амплитуды или удаление частотных компонентов, которые несут ненужную информацию, может улучшить качество речевого сигнала [3].

Применение преобразования Фурье в предварительной обработке речевых сигналов зависит от конкретной задачи и требований. Оно может помочь в анализе и улучшении частотных характеристик речи, что может быть полезно в распознавании и классификации аудиоданных.

Другим методом временной обработки сигналов является вейвлет-преобразование, обладающее самонастраивающимся подвижным частотно-временным окном, который одинаково хорошо выявляет как низкочастотные, так и высокочастотные характеристики сигнала на разных временных масштабах. В этом случае сигнал анализируется путем разложения по базисным функциям, полученным из некоторого прототипа путем сжатий, растяжений и сдвигов. Функция «прототип» называется материнским, или анализирующим, вейвлетом. В общем случае вейвлет-преобразование функции $f(t)$ выглядит так:

$$W(x,s)=\frac{1}{s}\int_{-\infty}^{\infty}\Psi^{*}\left(\frac{t-x}{s}\right)f(t)dt,$$

где t – ось времени; x – момент времени; s – параметр, обратный частоте; Ψ – функция анализирующего вейвлета; $(^{*})$ – комплексно-сопряженное значение.

Благодаря хорошей приспособленности к анализу нестационарных сигналов вейвлет-преобразование стало мощной альтернативой преобразованию Фурье. Недостатком вейвлет-преобразования является необходимость априорной информации об исследуемом сигнале для правильного подбора материнского вейвлета.

Вейвлет-преобразование также может быть полезным в предварительной обработке речевых сигналов, так как обладает способностью анализировать сигналы на разных временных и частотных масштабах, что делает его мощным инструментом для извлечения информации из речевых данных. Вейвлет-преобразование можно использовать при предварительной обработке речевых сигналов следующим образом:

1. Выделение временных и частотных характеристик. Вейвлет-преобразование позволяет разложить речевой сигнал на различные компоненты разных масштабов. Это может быть полезно для выделения как быстрых изменений в речи (короткие всплески), так и более медленных изменений (модуляции интонации). Выделение различных временных и частотных характеристик может помочь в анализе и распознавании особенностей речи.

2. Сжатие данных. Вейвлет-преобразование может использоваться для сжатия речевых данных, удаляя несущественные детали и сохраняя важные компоненты. Это может быть полезно для уменьшения объема данных и улучшения эффективности хранения и передачи аудиосигналов.

3. Уменьшение шума. Вейвлет-преобразование может использоваться для удаления шумовых компонентов из речевого сигнала на разных масштабах, что помогает улучшить качество сигнала.

4. Выделение формант. Форманты являются резонансными частотами, которые содержат важную информацию о произношении звуков и гласных. Вейвлет-преобразование может помочь в выделении этих формант из речевых сигналов.

5. Анализ изменений интонации. Вейвлет-преобразование может помочь в анализе изменений интонации в речи, что может быть полезно в задачах, связанных с выделением эмоциональной окраски или акцентов.

Применение вейвлет-преобразования в предварительной обработке речевых сигналов зависит от конкретных целей и задачи анализа речи. Этот метод предоставляет множество возможностей для извлечения информации из аудиоданных и может быть эффективным инструментом в области обработки речи [3].

Подробный анализ известных способов обработки в частотно-временной области выявил перспективность использования способов на основе преобразования Гильберта – Хуанга, в частности с использованием декомпозиции на эмпирические моды [3, 4].

Декомпозиция на эмпирические моды (ДЭМ) представляет собой адаптивную итерационную вычислительную процедуру, в результате которой исходный сигнал раскладывается на внутренние функции (частотные составляющие), называемые эмпирическими модами (ЭМ). Разложение на ЭМ позволяет анализировать локальные особенности сигнала, поэтому данный метод может быть использован при обработке нестационарных данных.

В основе метода ДЭМ заключается построение гладких огибающих по максимумам и минимумам функции сигнала и дальнейшее вычитание среднего значения этих огибающих из исходного сигнала. Для этого производится поиск экстремумов и методом сплайн аппроксимации этих точек определяются верхняя и нижняя огибающие. ДЭМ не имеет строгого математического описания, а аналитическое выражение имеет следующий вид:

$$x(n)=\sum_{i=1}^{I} IMF_i(n)+r_I(n),$$

где $x(n)$ – исходный сигнал; $IMF_i(n)$ – ЭМ; $r_I(n)$ – конечный остаток, $i = 1, 2, \ldots, I$ – номер ЭМ, n – дискретный отсчет времени [1].

Так как последний метод не требует точного математического описания сигнала, а все данные берет из самого сигнала, этот метод наилучший при обработке речевых сигналов из-за их нестационарности.

Результаты разложения речевых команд на моды с помощью улучшенной полной множественной декомпозиции на эмпирические моды с адаптивным шумом представлены на рис. 1.

При предварительной обработке речевых сигналов этот метод используется для анализа и разложения временных рядов и представляет собой адаптивный и ориентированный на данные метод разложения сигнала на компоненты, называемые интринсическими модами.

Он был разработан для анализа нестационарных сигналов, таких как речь, и может быть полезен в различных задачах обработки речевой информации. Его используют в области анализа и обработки сигналов, включая обработку речи, для следующих задач:

1. Извлечение признаков. ДЭМ может использоваться для извлечения интринсических признаков из речевых сигналов, которые могут быть полезными для задач распознавания речи, классификации или анализа эмоциональной окраски [1].

2. Фильтрация и удаление шума. ДЭМ позволяет разделять сигналы от шума и удалить нежелательные компоненты из речевых записей [5].

3. Исследование нестационарности. ДЭМ помогает анализировать временные изменения в речи, такие как изменения в частоте или амплитуде, что может быть важно для понимания динамики речевых сигналов.

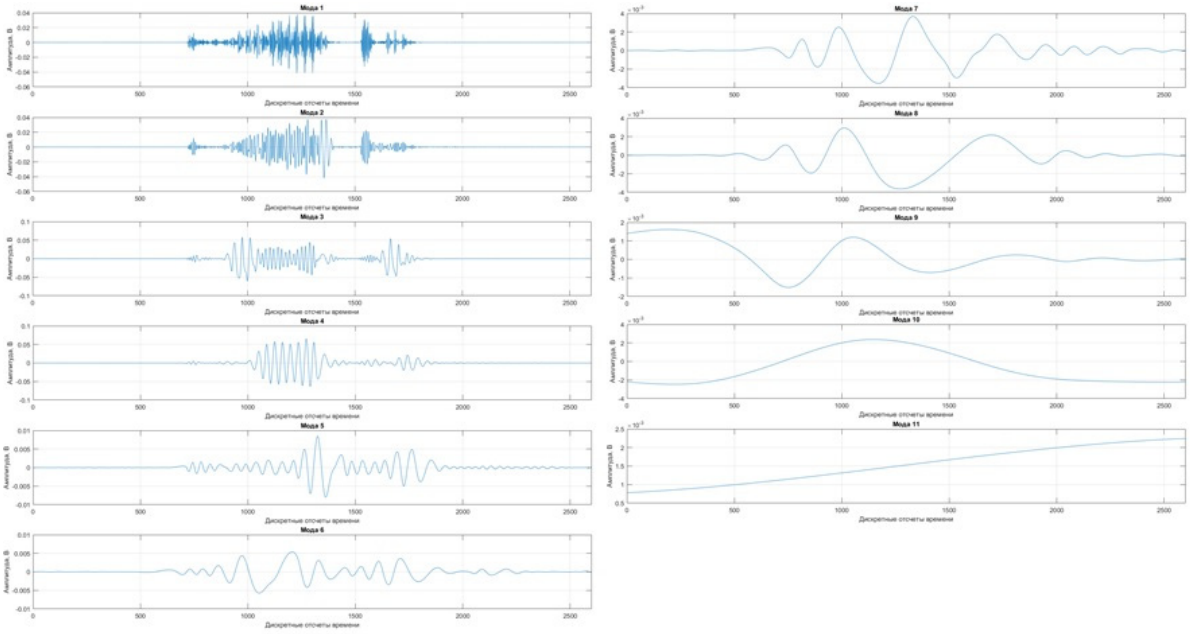

Рис. 1. Разложение на эмпирические моды речевого сигнала

Процесс декомпозиции на эмпирические моды включает в себя разложение сигнала на интринсические моды, каждая из которых представляет собой частотную компоненту с различной временной длительностью. Эти моды можно затем анализировать и использовать в дальнейших задачах обработки речи.

Однако следует учесть, что эффективность данного метода предварительной обработки речевых сигналов может зависеть от конкретной задачи и характеристик сигнала. Важно также учитывать, что существуют различные модификации ДЭМ, такие как множественная декомпозиция на эмпирические моды, которые могут быть более стабильными и устойчивыми к шуму.

Заключение

Предварительная обработка речевых сигналов играет ключевую роль в обеспечении высокой точности распознавания и анализа речи. Эффективная предварительная обработка позволяет извлечь информацию из акустических сигналов и улучшить качество последующих этапов анализа, таких как распознавание речи или классификация речевых команд.

Анализ методов показал, что наиболее подходящим методом предварительной обработки речевых сигналов может служить декомпозиция на эмпирические моды, а также различные модификации данного метода, например, множественной декомпозиции с адаптивным шумом.

Благодаря применению алгоритма предварительной обработки, основанного на декомпозиции на эмпирические моды, может быть увеличена скорость выделения информативной части речевого сигнала, что в дальнейшем можно будет применять при решении различных задач.

Список литературы

1. Козлов В. В., Фокина Е. А., Трофимов А. А. Предварительная обработка сигнала при распознавании голосовых команд методом улучшенной полной множественной декомпозиции на эмпирические моды // Измерение. Мониторинг. Управление. Контроль. 2022. № 3. С. 56–61. doi: 10.21685/2307-5538-2022-3-6
2. Фокина Е. А., Трофимов А. А., Козлов В. В. [и др.] Устройство распознования речевых сигналов на основе искусственной нейронной сети // Методы, средства и технологии получения и обработки

измерительной информации («Шляндинские чтения – 2022») : сб. ст. по материалам XIV Междунар. науч.-техн. конф. с элементами научной школы и конкурсом научно-исследовательских работ для обучающихся и молодых ученых (г. Пенза, 24–26 октября 2022 г.). Пенза : Изд-во ПГУ, 2022. С. 190–195. EDN: EQHJOA
3. Козлов В. В. Предварительная обработка сигнала методом разложения на собственные числа для распознавания голосовых команд // Методы, средства и технологии получения и обработки измерительной информации («Шляндинские чтения – 2021») : сб. ст. по материалам XIII Междунар. науч.-техн. конф. с элементами научной школы и конкурсом научно-исследовательских работ для студентов, аспирантов и молодых ученых (г. Пенза, 28–30 октября 2021 г.) / под ред. Е. А. Печерской. Пенза : Изд-во ПГУ, 2021. С. 147–150. EDN: JHSRXE
4. Bodin O. N., Kozlov V. V., Nefed'ev D. I. [et al.] Pre-processing voice signals for voice recognition systems // 18th International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices EDM 2017 : Conference Proceedings (Erlagol, Altai, 29 June – 03 July 2017). Erlagol, Altai: IEEE Computer Society, 2017. P. 242–245. doi: 10.1109/EDM.2017.7981748. EDN: PRLTAH
5. Бердибаева Г. К., Бодин О. Н., Громков Н. В. [и др.] Применение искусственных нейронных сетей для распознавания речевых команд // Измерение. Мониторинг. Управление. Контроль. 2017. № 2. С. 77–84. EDN: YUECPT

References

1. Kozlov V.V., Fokina E.A., Trofimov A.A. Signal preprocessing in voice command recognition by the method of improved complete multiple decomposition into empirical modes. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measurement. Monitoring. Management. Control*. 2022;(3):56–61. (In Russ.). doi: 10.21685/2307-5538-2022-3-6
2. Fokina E.A., Trofimov A.A., Kozlov V.V. et al. Speech signal recognition device based on an artificial neural network. *Metody, sredstva i tekhnologii polucheniya i obrabotki izmeritel'noy informatsii («Shlyandinskie chteniya – 2022»): sb. st. po materialam XIV Mezhdunar. nauch.-tekhn. konf. s elementami nauchnoy shkoly i konkursom nauchno-issledovatel'skikh rabot dlya obuchayushchikhsya i molodykh uchenykh (g. Penza, 24–26 oktyabrya 2022 g.) = Methods, tools and technologies for obtaining and processing measuring information ("Shlyandinsky readings – 2022") : collection of articles based on the materials of the XIV International Scientific and Technical a conference with elements of a scientific school and a competition of research papers for students and young scientists (Penza, October 24–26, 2022).* Penza: Izd-vo PGU, 2022:190–195. (In Russ.). EDN: EQHJOA
3. Kozlov V.V. Signal preprocessing by the method of decomposition into eigenvalues for recognition of voice commands. *Metody, sredstva i tekhnologii polucheniya i obrabotki izmeritel'noy informatsii («Shlyandinskie chteniya – 2021»): sb. st. po materialam XIII Mezhdunar. nauch.-tekhn. konf. s elementami nauchnoy shkoly i konkursom nauchno-issledovatel'skikh rabot dlya studentov, aspirantov i molodykh uchenykh (g. Penza, 28–30 oktyabrya 2021 g.) = Methods, means and technologies for obtaining and processing measuring information ("Shlyandinsky readings – 2021") : collection of articles based on materials of the XIII International Scientific and Technical A conference with elements of a scientific school and a competition of research papers for students, postgraduates and young scientists (Penza, October 28-30, 2021).* Penza: Izd-vo PGU, 2021:147–150. (In Russ.). EDN: JHSRXE
4. Bodin O.N., Kozlov V.V., Nefed'ev D.I. et al. Pre-processing voice signals for voice recognition systems. *18th International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices EDM 2017: Conference Proceedings (Erlagol, Altai, 29 June – 03 July 2017)*. Erlagol, Altai: IEEE Computer Society, 2017:242–245. doi: 10.1109/EDM.2017.7981748. EDN: PRLTAH
5. Berdibaeva G.K., Bodin O.N., Gromkov N.V. et al. Application of artificial neural networks for speech command recognition. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measurement. Monitoring. Management. Control*. 2017;(2):77–84. (In Russ.). EDN: YUECPT

Информация об авторах / Information about the authors

Валерий Валерьевич Козлов
кандидат технических наук,
доцент кафедры информационно-измерительной техники и метрологии,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: val.iit@mail.ru

Valeriy V. Kozlov
Candidate of technical sciences, associate professor of the sub-department of information and measuring equipment and metrology,
Penza State University
(40 Krasnaya street, Penza, Russia)

Алексей Анатольевич Трофимов
доктор технических наук, доцент,
профессор кафедры информационно-
измерительной техники и метрологии,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: iit@pnzgu.ru

Aleksey A. Trofimov
Doctor of technical sciences, associate professor,
professor of the sub-department of information
and measuring equipment and metrology,
Penza State University
(40 Krasnaya street, Penza, Russia)

Екатерина Александровна Фокина
аспирант,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: ekaterina.isay1997@gmail.com

Ekaterina A. Fokina
Postgraduate student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Владислав Николаевич Пономарев
аспирант,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: revik2296@gmail.com

Vladislav N. Ponomarev
Postgraduate student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Тимофей Олегович Жуков
студент,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: iit@pnzgu.ru

Timofey O. Zhukov
Student,
Penza State University
(40 Krasnaya street, Penza, Russia)

**Авторы заявляют об отсутствии конфликта интересов /
The authors declare no conflicts of interests.**

Поступила в редакцию/Received 26.09.2023
Поступила после рецензирования/Revised 23.10.2023
Принята к публикации/Accepted 21.11.2023