



<http://dx.doi.org/10.35596/1729-7648-2025-23-6-96-102>

УДК 004.934.2+534.784

ДЕТЕКТИРОВАНИЕ БОКОВОГО АМИОТРОФИЧЕСКОГО СКЛЕРОЗА НА ОСНОВЕ АКУСТИЧЕСКОГО АНАЛИЗА ГОЛОСА С ИСПОЛЬЗОВАНИЕМ БИБЛИОТЕКИ openSMILE

А. В. МИХНЕВИЧ, М. И. ВАШКЕВИЧ

*Белорусский государственный университет информатики и радиоэлектроники
(Минск, Республика Беларусь)*

Аннотация. Рассмотрена задача автоматического выявления признаков бокового амиотрофического склероза на основе анализа акустических характеристик голосового сигнала. Для извлечения функциональных акустических признаков голоса использовалась библиотека openSMILE с конфигурацией ComParE_2016. В качестве исходных данных использовались аудиозаписи из голосовой базы Minsk2020_ALS, включающей записи как здоровых пациентов, так и пациентов с боковым амиотрофическим склерозом. Проведены сравнение голосовых признаков между группами с использованием непараметрического критерия Манна – Уитни и FDR-коррекции множественных сравнений и раздельный анализ по полу. Выполнен эксперимент по классификации голосовых сигналов с использованием вложенной процедуры перекрестной проверки. Получены классификаторы, имеющие вероятность правильного обнаружения 75,0 % (для мужских голосов) и 74,2 % (для женских голосов). Выявлены статистически значимые акустические параметры, которые могут быть полезны в задачах автоматизированной диагностики и мониторинга бокового амиотрофического склероза.

Ключевые слова: боковой амиотрофический склероз, openSMILE, речевые признаки, акустический анализ, статистика, FDR-коррекция.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Для цитирования. Михневич, А. В. Детектирование бокового амиотрофического склероза на основе акустического анализа голоса с использованием библиотеки openSMILE / А. В. Михневич, М. И. Вашкевич // Доклады БГУИР. 2025. Т. 23, № 6. С. 96–102. <http://dx.doi.org/10.35596/1729-7648-2025-23-6-96-102>.

DETECTION OF AMYOTROPHIC LATERAL SCLEROSIS BASED ON ACOUSTIC VOICE ANALYSIS USING THE openSMILE LIBRARY

ALEXANDRINA V. MIKHNEVICH, MAXIM I. VASHKEVICH

Belarusian State University of informatics and Radioelectronics (Minsk, Republic of Belarus)

Abstract. This paper examines the problem of automatically detecting signs of amyotrophic lateral sclerosis based on the analysis of the acoustic characteristics of a voice signal. The openSMILE library with the ComParE_2016 configuration was used to extract functional acoustic features of the voice. Audio recordings from the Minsk2020_ALS voice database, which includes recordings of both healthy patients and patients with amyotrophic lateral sclerosis, were used as input. Voice features were compared between groups using the non-parametric Mann–Whitney test and FDR correction for multiple comparisons, with separate analysis by gender. An experiment on classifying voice signals was conducted using a nested cross-validation procedure. The resulting classifiers had a correct detection probability of 75.0 % (for male voices) and 74.2 % (for female voices). Statistically significant acoustic parameters that may be useful in automated diagnostics and monitoring of amyotrophic lateral sclerosis were identified.

Keywords: amyotrophic lateral sclerosis, openSMILE, speech features, acoustic analysis, statistics, FDR correction.

Conflict of interests. The authors declare that there is no conflict of interests.

For citation. Mikhnevich A. V., Vashkevich M. I. (2025) Detection of Amyotrophic Lateral Sclerosis Based on Acoustic Voice Analysis Using the openSMILE Library. *Doklady BGUIR*. 23 (6), 96–102. <http://dx.doi.org/10.35596/1729-7648-2025-23-6-96-102> (in Russian).

Введение

Боковой амиотрофический склероз (БАС) – нейродегенеративное заболевание, которое поражает двигательные нейроны головного и спинного мозга, приводя к прогрессирующей мышечной слабости и нарушению двигательных функций. Одним из ранних симптомов БАС может быть изменение речевой функции, что делает голос важным биомаркером на ранних стадиях заболевания [1]. Автоматизированный анализ речи становится все более востребованным направлением в контексте скрининга, мониторинга и поддержки диагноза БАС [2, 3].

Несмотря на активное развитие методов машинного обучения, в частности, глубоких нейронных сетей, их применение в задаче детектирования БАС на основе речевого сигнала сталкивается с рядом ограничений. В [4, 5] подчеркивается, что нейросетевые подходы демонстрируют ограниченную эффективность при работе с малыми выборками и высокой вариативностью в данных, что является характерным для медицинских наборов речевых данных. Отмечается, что даже при использовании современных архитектур, таких как сверточные нейронные сети, результаты моделей зачастую не обладают достаточной обобщающей способностью.

Глубокие модели требуют значительного объема данных для обучения, что становится критичным при работе с редкими заболеваниями, такими как БАС. В дополнение к проблеме дефицита данных существенную роль играет интерпретируемость модели. Для реального применения важен не только корректный результат детектирования патологии, но и возможность установить, изменение каких именно характеристик голоса на него повлияло. В случае нейросетей такой интерпретируемости зачастую достичь сложно, поскольку они функционируют как черный ящик.

Альтернативой акустическим методам анализа речи являются подходы, основанные на применении кинематических сенсоров, таких как электромагнитная артикулография [6]. Эти методы позволяют регистрировать движения языка и губ с высокой точностью и применяются в клинических исследованиях нарушений речи. Однако использование таких технологий требует дорогостоящего оборудования, проведения обследования в лабораторных условиях. Это существенно ограничивает их применимость в массовой диагностике и регулярном мониторинге.

В статье рассмотрена возможность построения системы детектирования БАС на основе акустического анализа голосового сигнала. Методологическая основа исследования включает три этапа. Во-первых, для извлечения акустических признаков применяется библиотека openSMILE [7], обеспечивающая комплексное описание речевого сигнала при помощи стандартизированного набора параметров и доказавшая свою эффективность в задачах медицинской диагностики [8]. Во-вторых, анализ голосового сигнала проводится раздельно для мужской и женской групп, что позволяет учесть физиологические различия в строении голосового аппарата и минимизировать вариативность в данных. В-третьих, для детектирования используются интерпретируемые статистические классификаторы, которые дают возможность идентифицировать наиболее значимые диагностические признаки.

Описание базы голосов

Использовалась база данных Minsk_ALS_2020 [9], содержащая записи устойчивого произнесения гласных «а» и «и», собранная в Республиканском научно-практическом центре неврологии и нейрохирургии (Минск). Выборка включала 64 диктора (31 БАС, 33 группы контроля) и содержала 128 аудиозаписей. Группа БАС включала 17 мужчин ($(61,1 \pm 7,7)$ лет) и 14 женщин ($(57,3 \pm 7,8)$ лет). Контрольная группа включала 13 мужчин ($(50,2 \pm 13,8)$ лет) и 20 женщин ($(56,1 \pm 9,7)$ лет). Протокол записи предусматривал произнесение гласных на комфортной высоте и громкости с максимальной продолжительностью. Все записи были стандартизированы по длительности и приведены к единому формату (wav, 44,1 кГц). Аудиоматериалы и метаданные доступны в репозитории проекта на GitHub [10].

В качестве иллюстрации на рис. 1 приведен вид аудиосигналов из рассматриваемой базы во временной и мел-частотной областях.

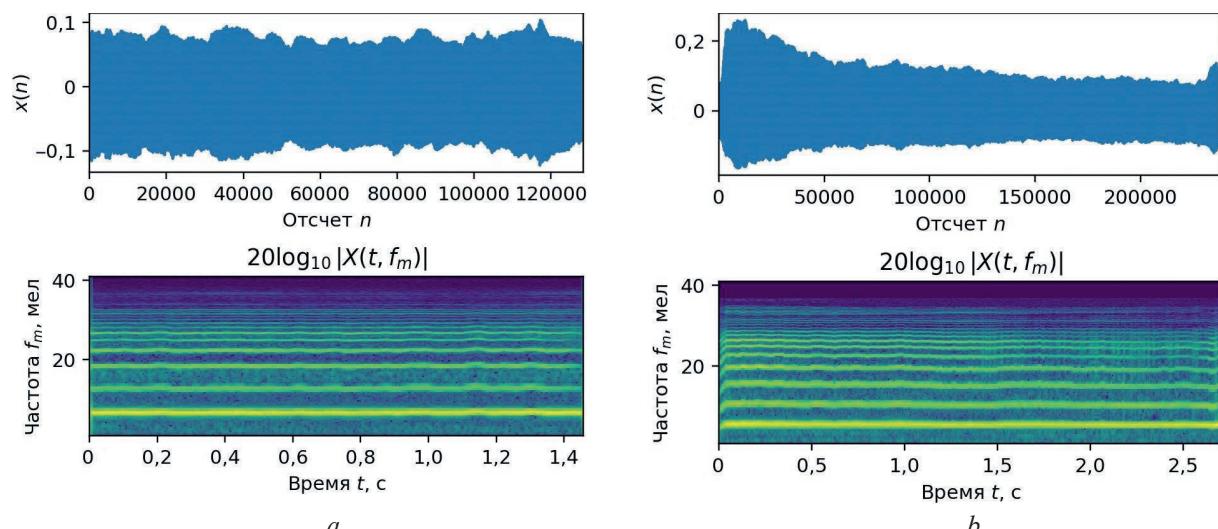


Рис. 1. Пример записей женских голосов из базы Minsk_ALS_2020 по возрасту:
 а – 64 года (группа контроля); б – 55 лет (пациентка с боковым амиотрофическим склерозом)
Fig. 1. Example of female voice recordings from the Minsk_ALS_2020 database by age:
 a – 64 years old (control group); b – 55 years old (patient with amyotrophic lateral sclerosis)

Приведенные на рис. 1 сигналы, несмотря на внешнее сходство, имеют различную динамику частотных компонент, которые позволяют отличить голос человека в норме и при патологии.

Извлечение признаков и статистический анализ

Анализ акустических признаков голосовых сигналов проводили с помощью программной библиотеки openSMILE (open-source Speech and Music Interpretation by Large-Space Extraction) [7], представляющей собой один из наиболее широко используемых инструментов в задачах автоматического анализа речи. Данная библиотека – это гибкий и модульный фреймворк для извлечения широкого спектра параметров речи, включая мел-частотные кепстральные коэффициенты (МЧКК), амплитудный и модуляционный спектры, а также статистические характеристики, вычисленные на основе данных первичных параметров.

В исследовании применялась конфигурация ComParE_2016, входящая в стандартный набор openSMILE и ориентированная на извлечение надсегментных признаков. Общее число извлекаемых признаков составило 6373. На этапе статистического анализа для каждого признака проводилось сравнение между группой контроля и БАС с использованием непараметрического U-критерия Манна – Уитни, поскольку распределения признаков не удовлетворяли требованиям нормальности. Для учета эффекта множественных сравнений применялась коррекция полученных p -значений по методу Бенджамина – Хохберга ($q \leq 0,05$). В табл. 1 приведены наиболее значимые признаки, найденные для групп мужчин и женщин.

Таблица 1. Наиболее значимые речевые признаки
Table 1. The most significant speech features

Пол	Название признака	p -значение
Женский	mfcc_sma[9]_minSegLen	8,6245e-05
	audSpec_Rfilt_sma[16]_minSegLen	2,9027e-04
	pcm_fftMag_spectralSlope_sma_risetime	3,2555e-04
	audSpec_Rfilt_sma[22]_minSegLen	3,3778e-04
	mfcc_sma_de[4]_minSegLen	3,5855e-04
Мужской	audSpec_Rfilt_sma[15]_range	4,9817e-04
	audSpec_Rfilt_sma_de[17]_range	4,9817e-04
	audSpec_Rfilt_sma[17]_lpc0	4,9817e-04
	audSpec_Rfilt_sma[7]_lpc1	4,9817e-04
	audSpec_Rfilt_sma[7]_stddevRisingSlope	5,7128e-04

Следует заметить, что ни один признак не прошел проверку на статическую значимость после коррекции Бенджамини – Хохберга. С другой стороны, признаки имеют p -значения менее 0,001, что указывает на существенные различия между анализируемыми группами. Из табл. 1 видно, что признаки в двух группах отличаются. Это говорит об акустических различиях речевого тракта у мужчин и женщин. Признак `mfcc_sma[9]_minSegLen`, например, измеряет минимальную продолжительность сегмента, в котором (сглаженный) девятый коэффициент МЧКК остается выше определенного порога. Низкое значение $p = 8,6\text{e-}05$ указывает на статистически значимую разницу между сравниваемыми группами по стабильности среднего спектрального контура в области среднего мел-диапазона.

На рис. 2 представлена ящиковая диаграмма признака `pcm_fftMag_spectralSlope_sma_risetime`, отражающего время нарастания (англ. *rise time*) спектрального наклона (англ. *spectral slope*) до 50 % от максимальной величины. Значение $p = 3,26\text{e-}04$ для данного признака указывает на существенные различия в темпах формирования спектрального наклона в группе женских голосов.

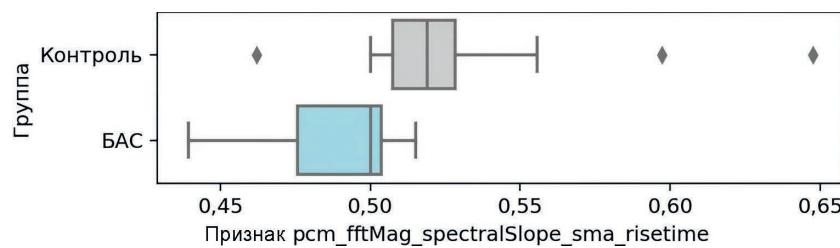


Рис. 2. Ящиковая диаграмма признака `pcm_fftMag_spectralSlope_sma_risetime`
Fig. 2. Boxplot of `pcm_fftMag_spectralSlope_sma_risetime` feature

Следует отметить, что обе группы включают признаки из группы `audSpec_Rfilt`, которые вычисляются на основе модуляционного спектра и RASTA-фильтрации [11]. Таким образом, можно сделать обобщающий вывод, что различия голосов в анализируемых группах проявляются в спектротемпоральной области.

Полученные на первом этапе признаки были стандартизированы ($\mu = 0, \sigma = 1$) для выполнения экспериментов с классификацией набора данных. Для тестирования были выбраны четыре базовых классификатора, которые хорошо себя зарекомендовали в задачах детектирования патологий по речи [2]: машина опорных векторов (МОВ) с линейным ядром, МОВ с радиальным ядром [12], линейный дискриминантный анализ (ЛДА) [9] и деревья решений [12]. Оценка производительности классификаторов выполнялась при помощи вложенной перекрестной проверки, внешний цикл которой использовался для расчета правильности классификатора, а внутренний – для подбора параметров классификатора (в частности, для определения оптимального числа признаков).

Результаты исследований и их обсуждение

Для мужской группы голосов наилучшую производительность показал классификатор на основе деревьев решений, его вероятность правильного обнаружения составила 75,0 %, для женской – наилучший результат у классификатора на основе ЛДА со схожей вероятностью правильного обнаружения – 74,2 % (рис. 3). Оба классификатора значительно превосходят классификатор на основе случайного выбора, который для бинарной классификации имеет вероятность правильного обнаружения 50 %. Ширина доверительных интервалов (ДИ) примерно одинакова, что указывает на сопоставимую надежность полученных оценок правильности. В целом полученные результаты демонстрируют эффективность речевых признаков для детектирования БАС.

Дополнительно были проанализированы признаки, которые имели наибольшую информативность, т. е. чаще всего отбирались для формирования классификатора. Для классификации мужских голосов наибольшую значимость имел признак `audSpec_Rfilt_sma[17]_lpc2`. Согласно принципам именования библиотеки `openSMILE`, префикс `audSpec_Rfilt` указывает, что признак получен на основе энергии узкополосного спектра (вычисленного в критических полосах слуха) после RASTA-фильтрации; 17 означает 17-ю полосу спектра; `sma` (англ. *simple moving average*) – сглаживание простым скользящим средним по кадрам; `lpc2` – второй коэффициент линейного предсказания, рассчитанный для 17-й полосы. На рис. 4, *a* приведена ящиковая диаграмма этого признака.

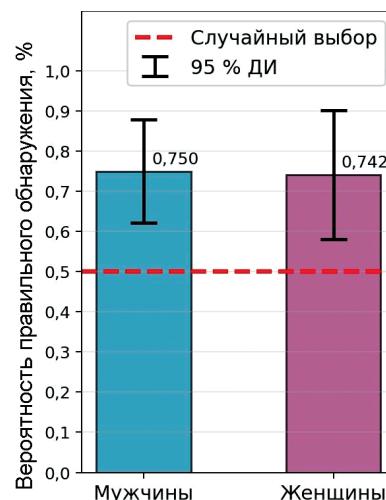


Рис. 3. Производительность классификаторов для мужских и женских голосов
Fig. 3. Classification performance for male and female voices

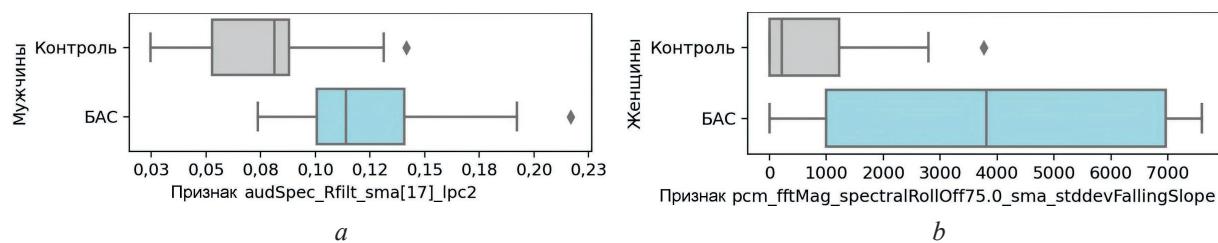


Рис. 4. Ящиковыe диаграммы наиболее информативных речевых признаков для групп:
a – мужчин; **b** – женщин
Fig. 4. Box plots of the most informative speech features for groups: *a* – men; *b* – women

Высокое абсолютное значение lpc2 (рис. 4, *a*) указывает на выраженные «изломы» модуляционного спектра в 17-й полосе, что отражает нестабильную динамику тембра голоса. Нужно учесть, что анализируются именно записи устойчивой фонации звука «а», и в идеальном случае колебания в модуляционном спектре должны быть минимальными. Низкое значение lpc2 говорит о более однородном распределении энергии в полосе (без ярко выраженных пиков) – это признак ровного звучания, который характерен для здоровых голосов. Ящиковая диаграмма на рис. 4, *a* подтверждает приведенные выше рассуждения: у пациентов с БАС признак audSpec_Rfilt_sma[17]_lpc2 имеет более высокое значение, чем в группе контроля.

Для классификации женских голосов наибольшую значимость имел признак pcm_fftMag_spectralRollOff75.0_sma_stddevFallingSlope (рис. 4, *b*). В данном случае pcm_fftMag указывает, что признак вычислялся на основе амплитудного спектра сигнала; spectralRollOff75 – это частота, ниже которой сосредоточено 75 % общей энергии спектра; stddevFallingSlope – стандартное отклонение углов наклона на тех сегментах, где признак убывает. Из ящиковской диаграммы на рис. 4, *b* можно видеть, что у пациентов с БАС значение данного признака значительно выше, чем в группе контроля. В целом высокое значение этого признака говорит о нерегулярности и резких изменениях в движении границы 75 % энергии: в областях спада энергия спектра быстро «скатывается» вниз. Низкое stddevFallingSlope-значение, наоборот, указывает на плавный спад частотной границы 75 % энергии и является характеристикой ровного, спокойного голоса.

Полученные результаты также находятся в согласии с ранее опубликованными данными. Так, в [2], где библиотека openSMILE тоже использовалась для решения задачи детектирования голосов пациентов с БАС, было выявлено, что признак audSpec_Rfilt_sma[19]_lpc4 входит в список наиболее значимых признаков для анализа мужских голосов. Помимо этого, среди наиболее информативных признаков в [2] выделен audSpec_Rfilt_sma[10]_maxPos. В целом можно сделать обобщающий вывод, что признаки, полученные на основе анализа спектра, вычисленного в критических полосах, с использованием RASTA-фильтрации имеют существенное значение для решения задачи детектирования БАС с использованием акустического анализа.

Заключение

1. Проведено исследование применимости признаков, вычисляемых библиотекой openSMILE, для задачи детектирования голосов пациентов с боковым амиотрофическим склерозом с использованием набора данных, собранных при помощи смартфона. Анализ и классификация голосовых сигналов выполнялись с учетом пола. Для мужской группы голосов лучшую производительность показал классификатор на основе деревьев решений (75,0 % правильных результатов). Для женской группы голосов наилучший результат у классификатора на основе линейного дискриминантного анализа (74,2 % правильных результатов). Полученные показатели вероятности правильного обнаружения оставляют большое пространство для улучшения. Это можно объяснить тем, что библиотека openSMILE является универсальным средством извлечения речевых признаков и не учитывает специфику конкретной голосовой патологии.

2. Важным практическим результатом исследования является установление доминирующей роли признаков, вычисляемых на основе модуляционного спектра сигнала (RASTA-фильтрация) в задаче детектирования бокового амиотрофического склероза. Полученные данные позволяют предположить, что целенаправленное проектирование признакового пространства на основе анализа модуляционных спектра может существенно повысить эффективность автоматизированных систем детектирования бокового амиотрофического склероза по речевым сигналам.

Список литературы / References

1. Bowden M., Beswick E., Tam J., Perry D., Smith A., Newton J., et al. (2023) A Systematic Review and Narrative Analysis of Digital Speech Biomarkers in Motor Neuron Disease. *NPJ Digital Medicine*. 6 (1), 1–19.
2. Norel R., Pietrowicz M., Agurto C., Rishoni S., Cecchi G. (2018) Detection of Amyotrophic Lateral Sclerosis (ALS) Via Acoustic Analysis. *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*. 377–381.
3. Folgado D., Rocha P. S., Matias P., Liebeatrau D., Nunes F., Rocha T., et al. (2025) HomeSenseALS: A Mobile Sensing and Self-Monitoring System to Monitor Patients with ALS in Everyday Life. *ACM Transactions on Computing for Healthcare*. 25 (3), 1–32.
4. An K., Kim M., Teplansky K., Green J., Campbell T., Yunusova Y., et al. (2018) Automatic Early Detection of Amyotrophic Lateral Sclerosis from Intelligible Speech Using Convolutional Neural Networks. *Proceedings of Interspeech*. 1913–1917.
5. Illa A., Patel D., Yamini B. K., Veeramani P.-K., Polavarapu K., Nashi S., et al. (2018) Comparison of Speech Tasks for Automatic Classification of Patients with Amyotrophic Lateral Sclerosis and Healthy Subjects. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 6014–6018.
6. Yunusova Y., Green J. R., Greenwood L., Wang J., Pattee G. L., Zinman L. (2012) Tongue Movements and Their Acoustic Consequences in Amyotrophic Lateral Sclerosis. *Journal of Speech, Language, and Hearing Research*. 55 (3), 673–687.
7. Eyben F., Wöllmer M., Schuller B. (2010) openSMILE: The Munich Versatile and Fast Open-Source Audio Feature Extractor. *Proceedings of the 18th ACM International Conference on Multimedia*. 1459–1462.
8. Bocklet T., Steidl S., Nöth E., Skodda S. (2013) Automatic Evaluation of Parkinson's Speech-Acoustic, Prosodic and Voice-Related Cues. *Proceedings of Interspeech*. 1149–1153.
9. Vashkevich M., Rushkevich Yu. (2021) Classification of ALS Patients Based on Acoustic Analysis of Sustained Vowel Phonations. *Biomedical Signal Processing and Control*. 65, 1–14.
10. Minsk2020_ALS_database: Classification of ALS Patients Based on Acoustic Analysis of Sustained Vowel Phonations. *GitHub*. Available: https://github.com/Mak-Sim/Minsk2020_ALS_database (Accessed 6 October 2025).
11. Hermansky H., Morgan N. (1994) RASTA Processing of Speech. *IEEE Transactions on Speech and Audio Processing*. 2 (4), 578–589.
12. Hastie T., Tibshirani R., Friedman J. H. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer*.

Поступила 10.10.2025

Принята в печать 18.11.2025

Received: 10 October 2025

Accepted: 18 November 2025

Вклад авторов

Михневич А. В. осуществила извлечение признаков с использованием библиотеки openSMILE, провела статистический анализ полученных результатов, выполнила эксперименты по классификации голосовых сигналов. Участвовала в обсуждении результатов и подготовке текста статьи.

Вашкевич М. И. разработал концепцию исследования, определил методику анализа акустических признаков, участвовал в интерпретации результатов и подготовке текста статьи.

Authors' contribution

Mikhnevich A. V. performed feature extraction using the openSMILE library, conducted statistical analysis of the results, and performed experiments on voice signal classification. Participated in the discussion of the results and the preparation of the article.

Vashkevich M. I. developed the study concept, defined the acoustic feature analysis methodology, and participated in the interpretation of the results and the preparation of the article.

Сведения об авторах

Михневич А. В., асп. каф. электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники

Вашкевич М. И., д-р техн. наук, проф. каф. электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники

Адрес для корреспонденции

220013, Республика Беларусь,
Минск, ул. П. Бровки, 6
Белорусский государственный университет
информатики и радиоэлектроники
Тел.: +375 17 293-84-78
E-mail: vashkevich@bsuir.by
Вашкевич Максим Иосифович

Information about the authors

Mikhnevich A. V., Postgraduate at the Department of Electronic Computing Facilities, Belarusian State University of Informatics and Radioelectronics

Vashkevich M. I., Dr. Sci. (Tech.), Professor at the Department of Electronic Computing Facilities, Belarusian State University of Informatics and Radioelectronics

Address for correspondence

220013, Republic of Belarus,
Minsk, P. Brovki St., 6
Belarusian State University
of Informatics and Radioelectronics
Tel.: +375 17 293-84-78
E-mail: vashkevich@bsuir.by
Vashkevich Maxim Iosifovich