

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

## INFORMATION TECHNOLOGIES



УДК 004.934.2 + 004.85  
<https://doi.org/10.37661/1816-0301-2023-20-1-104-114>

Оригинальная статья  
Original Paper

## Малоразмерные спектральные признаки для машинного обучения в задачах анализа и классификации голосового сигнала

Д. С. Лихачёв<sup>✉</sup>, М. И. Вашкевич, Н. А. Петровский, И. С. Азаров

Белорусский государственный университет  
информатики и радиоэлектроники,  
ул. П. Бровки, 6, Минск, 220013, Беларусь  
<sup>✉</sup>E-mail: [likhachov@bsuir.by](mailto:likhachov@bsuir.by)

### Аннотация

**Цели.** Решается задача разработки метода вычисления малоразмерных спектральных признаков, повышающего эффективность существующих систем машинного обучения для анализа и классификации голосовых сигналов.

**Методы.** Спектральные признаки извлекаются с помощью генеративного подхода, который предполагает вычисление дискретного спектра Фурье последовательности отчетов, сгенерированной с использованием авторегрессионной модели входного голосового сигнала. Сгенерированная последовательность, обрабатываемая дискретным преобразованием Фурье, учитывает периодичность преобразования, позволяя тем самым повысить точность спектральной оценки анализируемого сигнала.

**Результаты.** Предложен и описан генеративный метод вычисления спектральных признаков, предназначенных для применения в системах машинного обучения при анализе и классификации голосовых сигналов. Проведен экспериментальный анализ точности и стабильности представления спектра тестового сигнала с известным спектральным составом с использованием огибающих. Огибающие вычислялись с помощью предложенного генеративного метода и дискретного преобразования Фурье с различными окнами анализа (прямоугольным окном и окном Ханна). Проведенный анализ показал, что генеративный метод получения спектральных огибающих позволил добиться более точного представления спектра тестового сигнала по критерию минимума квадратичной ошибки. Проведено сравнение эффективности классификации голосового сигнала при использовании предложенных признаков и признаков на основе мел-частотных кепстральных коэффициентов. В качестве базовой тестовой системы для оценки эффективности предлагаемого подхода на практике использовалась система диагностики бокового амиотрофического склероза по голосу.

**Заключение.** Результаты экспериментов показали ощутимое повышение точности классификации при использовании предлагаемых признаков для вычисления признаков по сравнению с признаками на основе мел-частотных кепстральных коэффициентов

**Ключевые слова:** анализ голоса, генеративный метод, авторегрессия, машинное обучение, спектральные признаки, классификация

**Для цитирования.** Малоразмерные спектральные признаки для машинного обучения в задачах анализа и классификации голосового сигнала / Д. С. Лихачёв [и др.] // Информатика. – 2023. – Т. 20, № 1. – С. 104–114.

<https://doi.org/10.37661/1816-0301-2023-20-1-104-114>

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

---

Поступила в редакцию | Received 07.02.2023

Подписана в печать | Accepted 10.02.2023

Опубликована | Published 29.03.2023

---

---

## Small-size spectral features for machine learning in voice signal analysis and classification tasks

Denis S. Likhachov<sup>✉</sup>, Maxim I. Vashkevich, Nick A. Petrovsky, Elias S. Azarov

*Belarusian State University of Informatics and Radioelectronics,  
st. P. Brovki, 6, Minsk, 220013, Belarus*

<sup>✉</sup>E-mail: [likhachov@bsuir.by](mailto:likhachov@bsuir.by)

### Abstract

**Objectives.** The problem of developing a method for calculating small-sized spectral features that increases the efficiency of existing machine learning systems for analyzing and classifying voice signals is being solved.

**Methods.** Spectral features are extracted using a generative approach, which involves calculating a discrete Fourier spectrum for a sequence of samples generated using an autoregressive model of the input voice signal. The generated sequence processed by the discrete Fourier transform considers the periodicity of the transform and thereby increase the accuracy of spectral estimation of the analyzed signal.

**Results.** A generative method for calculating spectral features intended for use in machine learning systems for the analysis and classification of voice signals is proposed and described. An experimental analysis of the accuracy and stability of the spectrum representation of a test signal with a known spectral composition has been carried out using the envelopes. The envelopes were calculated using the proposed generative method and using the discrete Fourier transform with different analysis windows (rectangular window and Hanna window). The analysis showed that the spectral envelopes obtained using the proposed method more accurately represent the spectrum of the test signal according to the criterion of minimum square error. A comparison of the effectiveness of the voice signal classification with proposed features and the features based on the mel-frequency cepstral coefficients is carried out. A diagnostic system for amyotrophic lateral sclerosis was used as a basic test system to evaluate the effectiveness of the proposed approach in practice.

**Conclusion.** The obtained experimental results showed a significant increase of the classification accuracy when using the proposed approach for calculating features compared with the features based on the mel-frequency cepstral coefficients.

**Keywords:** voice analysis, generative method, autoregression, machine learning, spectral features, classification

**For citation.** Likhachov D. S., Vashkevich M. I., Petrovsky N. A., Azarov E. S. *Small-size spectral features for machine learning in voice signal analysis and classification tasks*. *Informatika [Informatics]*, 2023, vol. 20, no. 1, pp. 104–114 (In Russ.). <https://doi.org/10.37661/1816-0301-2023-20-1-104-114>

**Conflict of interest.** The authors declare of no conflict of interest.

**Введение.** В настоящее время большое количество систем для обработки речевых и голосовых сигналов с использованием машинного обучения оперируют признаками в частотной области [1–4]. Во многих случаях извлечение признаков в таких системах сводится к получению в том или ином виде спектральной оценки анализируемого сигнала. Основным требованием

к полученной оценке является как можно более высокое спектральное разрешение при удовлетворительной точности оценки для целевой задачи. В большинстве применяемых в настоящее время методов вычисления признаков для машинного обучения в частотной области используется дискретное преобразование Фурье (ДПФ). Это преобразование хорошо подходит для частотно-временного анализа периодических или частично периодических сигналов, в том числе речевых и голосовых [4]. Спектральная оценка при помощи ДПФ является точной при соблюдении двух условий: анализируемый сигнал должен быть строго периодическим и стационарным на протяжении всего интервала анализа; период анализируемого сигнала должен точно совпадать с размером ДПФ. Для речевых и аудиосигналов два вышеприведенных требования в строгом виде не выполняются. Это обуславливает необходимость применения временного окна для обработки входных данных [5]. Обработка окном хотя и позволяет ослабить влияние боковых спектральных лепестков, но в то же время приводит к значительным искажениям полученной спектральной оценки по отношению к истинной [5].

В настоящей работе для получения спектральных признаков предлагается подход, использующий генеративный метод спектрального оценивания на базе авторегрессионной модели анализируемого сигнала. Как будет показано далее, применение данного метода позволяет отказаться от использования окна при выполнении спектрального оценивания и в конечном итоге существенно улучшить эффективность работы системы обработки речевого сигнала на основе машинного обучения для диагностики бокового амиотрофического склероза (БАС) [6].

**Генеративный метод вычисления дискретной оценки огибающей спектра авторегрессионного процесса.** Основная идея предлагаемого метода получения дискретной оценки огибающей спектра авторегрессионного процесса (ОСАП) состоит в том, чтобы использовать ДПФ без применения оконной функции для обработки входных данных, обеспечив при этом согласованность периода анализируемого сигнала и размера ДПФ. С этой целью в качестве входных данных для ДПФ предлагается использовать специальным образом сгенерированную последовательность, которая имела бы равный с размером ДПФ период и соответствующий анализируемому сигналу спектральный состав.

Схема вычисления дискретной оценки ОСАП в общем виде изображена на рис. 1.

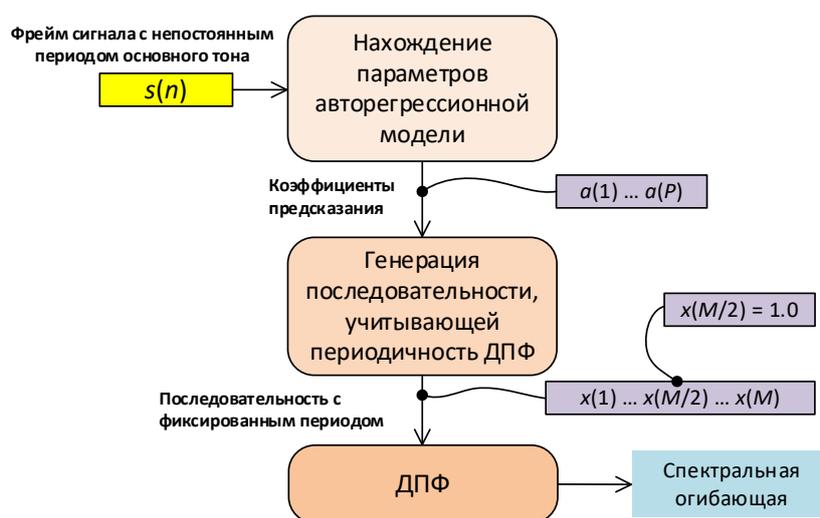


Рис. 1. Схема вычисления спектральной огибающей  
Fig. 1. Scheme for calculating the spectral envelope

Видно, что процесс вычисления дискретной оценки ОСАП включает несколько основных этапов:

1. По входному кадру анализируемого сигнала с использованием автокорреляционного метода определяются параметры авторегрессионной модели  $P$ -го порядка:  $a_k, k = \overline{1, P}$  [7].

2. С помощью авторегрессионной модели анализируемого процесса генерируется последовательность с периодом, равным размеру ДПФ. Размер ДПФ зависит от целевой размерности вектора признаков и покрываемого им частотного диапазона.

Предположим, необходимо сгенерировать требуемую входную последовательность  $x(m)$  для ДПФ размером  $M$ . Искомая последовательность не должна быть нулевой, т. е. должно выполняться условие

$$\sum_{m=1}^M |x(m)| \neq 0. \quad (1)$$

Зная параметры авторегрессионной модели анализируемого процесса, можно записать следующее равенство [7]:

$$x(m) = -\sum_{k=1}^P a_k \cdot x(m-k), \quad m = \overline{1, M+P}. \quad (2)$$

Положив  $x(m) = 0$  при  $m = \overline{M+1, M+P}$  и  $m = \overline{-P+1, 0}$ , выражение (2) для  $m = \overline{1, M}$  можно записать в виде равенства

$$\begin{pmatrix} -1 & 0 & 0 & \dots & 0 & 0 \\ -a_1 & -1 & 0 & \dots & 0 & 0 \\ -a_2 & a_1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & -a_p \end{pmatrix} \cdot \begin{pmatrix} x(1) \\ x(2) \\ x(3) \\ \vdots \\ x(M) \end{pmatrix} = 0. \quad (3)$$

Введя дополнительное ограничение  $x(M/2) = 1$  для выполнения условия (1) и решив полученную из (3) систему линейных уравнений относительно  $x$ , можно найти искомую последовательность  $x(m), m = \overline{1, M}$ , которая будет затухать к краям кадра аналогично оконной функции.

3. С помощью ДПФ вычисляется дискретный спектр одного периода сгенерированной на этапе 2 последовательности:

$$S_{est}(k) = \left| \sum_{m=0}^{M-1} x(m) e^{-i \frac{2\pi}{M} km} \right|, \quad k = \overline{1, \frac{M}{2}}. \quad (4)$$

Полученный дискретный спектр  $S_{est}(k)$  является искомой дискретной оценкой ОСАП.

**Анализ эффективности предлагаемого метода вычисления дискретной оценки ОСАП.** С целью анализа эффективности предлагаемого метода проведено сравнение точности и стабильности представления спектра тестового сигнала с помощью спектральных огибающих, полученных тремя способами: предложенным методом, ДПФ без применения окна, ДПФ с применением окна Ханна.

В качестве тестового сигнала использовался полигармонический речеподобный сигнал с заданными параметрами (рис. 2). В ходе эксперимента генерировались сегменты тестового сигнала с постоянной частотой основного тона и длительностью до 200 мс. Каждый сегмент тестового

сигнала анализировался кадрами длительностью от 1,5 до 20 мс с шагом 0,068 мс. Частота основного тона варьировалась от 85 до 240 Гц с шагом 1 Гц. Для корректного сопоставления частотной сетки спектра тестового сигнала и полученных огибающих использовалась линейная интерполяция. Все огибающие при сравнении нормировались по энергии. Пример спектральных огибающих, вычисленных для одного кадра тестового сигнала, приведен на рис. 2.

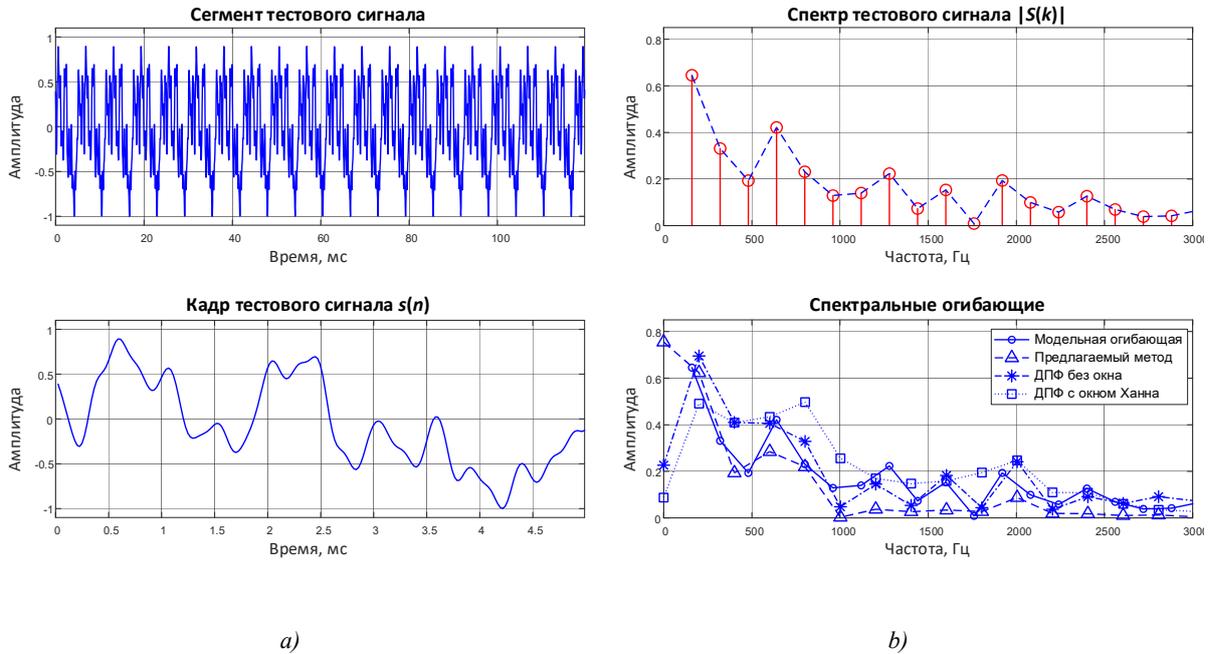


Рис. 2. Тестовый сигнал (a) и соответствующие спектральные огибающие для одного кадра (b)  
Fig. 2. Test signal (a) and corresponding spectral envelopes for one frame (b)

В качестве критериев точности и стабильности представления спектра тестового сигнала с помощью полученных огибающих использовались относительные квадратичные ошибки (%), вычисляемые с помощью следующих выражений:

$$SE_{acc}^{i,j} = \frac{\sum_{k=1}^K (S_{mod}^i(k) - S_{est}^{i,j}(k))^2}{\sum_{k=1}^K (S_{mod}^i(k))^2} \cdot 100, \quad SE_{stab}^i = \frac{\sum_{k=1}^K (S_{est}^{i,1}(k) - S_{est}^{i-1,F}(k))^2}{\sum_{k=1}^K (S_{est}^{i-1,F}(k))^2} \cdot 100, \quad (5)$$

где  $SE_{acc}^{i,j}$  – относительная квадратичная ошибка для  $j$ -го кадра сигнала из  $i$ -го сегмента;  $SE_{stab}^i$  – относительная квадратичная ошибка для  $i$ -го сегмента;  $S_{mod}^i(k)$  –  $k$ -й частотный отсчет спектра тестового сигнала для  $i$ -го сегмента;  $S_{est}^{i,j}(k)$  –  $k$ -й отсчет полученной спектральной огибающей для  $j$ -го кадра сигнала из  $i$ -го сегмента;  $S_{est}^{i,1}(k)$  и  $S_{est}^{i-1,F}(k)$  –  $k$ -й частотный отсчет спектральной огибающей для первого кадра из  $i$ -го сегмента и последнего кадра из  $i-1$ -го сегмента;  $K$  – количество частотных отсчетов, участвующих в оценке;  $F$  – количество кадров сигнала в сегменте.

Для получения итогового показателя точности  $\overline{SE_{acc}}$  вычисленные по выражению (5) ошибки  $SE_{acc}^{i,j}$  усреднялись по всем сегментам и кадрам. Итоговое значение показателя стабильности

находилось путем усреднения ошибок  $SE_{stab}^i$  по всем сегментам, за исключением первого. Полученные результаты представлены в табл. 1 и 2, жирным шрифтом выделены лучшие показатели.

Таблица 1

Средняя квадратичная ошибка  $\overline{SE}_{acc}$  для кадра анализа с разной длительностью, %

Table 1

Mean square error  $\overline{SE}_{acc}$  for an analysis frame with different duration, %

| Метод получения спектральной огибающей<br><i>Method for obtaining the spectral envelope</i> | Длительность кадра анализа, мс<br><i>Analysis frame duration, ms</i> |             |             |             |             |
|---|--|-------------|-------------|-------------|-------------|
|   | 1,5  | 3,2         | 5,0         | 10          | 20          |
| Предложенный метод<br><i>Suggested method</i>   | <b>24,8</b>  | <b>22,0</b> | <b>14,3</b> | <b>12,2</b> | <b>16,2</b> |
| ДПФ без применения окна<br><i>DFT without windowing</i>                                     | 46,2   | 33,1        | 28,1        | 29,3        | 42,6        |
| ДПФ с применением окна Ханна<br><i>DFT using Hann window</i>                                | 39,4   | 29,1        | 19,4        | 19,8        | 36,6        |

Таблица 2

Средняя квадратичная ошибка  $\overline{SE}_{stab}$  для кадра анализа с разной длительностью, %

Table 2

Mean square error  $\overline{SE}_{stab}$  for an analysis frame with different duration, %

| Метод получения спектральной огибающей<br><i>Method for obtaining the spectral envelope</i> | Длительность кадра анализа, мс<br><i>Analysis frame duration, ms</i> |             |             |            |            |
|---|--|-------------|-------------|------------|------------|
|   | 1,5  | 3,2         | 5,0         | 10         | 20         |
| Предложенный метод<br><i>Suggested method</i>   | <b>22,6</b>  | <b>23,4</b> | <b>19,5</b> | <b>2,4</b> | <b>0,7</b> |
| ДПФ без применения окна<br><i>DFT without windowing</i>                                     | 28,9   | 27,3        | 19,71       | 8,8        | 6,6        |
| ДПФ с применением окна Ханна<br><i>DFT using Hann window</i>                                | 28,8   | 30,8        | 27,41       | 9,0        | 2,2        |

Как показали результаты экспериментов, спектральные огибающие, полученные с помощью предложенного метода, более точно по сравнению с остальными методами представляют спектр тестового сигнала с точки зрения минимальной квадратичной ошибки.

**Использование малоразмерных спектральных признаков на основе дискретной оценки ОСАП.** Для проверки предлагаемого подхода в качестве тестовой взята система анализа и классификации голосового сигнала с использованием мел-частотных кепстральных коэффициентов (МЧКК), описанная в работе [6] (рис. 3). Основная задача данной системы заключается в выполнении бинарной классификации. В качестве входных данных используется голосовой сигнал, который представляет собой протяжный звук /a/. По результатам классификации входной образец голосового сигнала относят к норме или патологии. В соответствии со схемой на рис. 3 на этапе предобработки выполняется фильтрация сигнала. Далее происходит выделение признаков и рассчитываются надсегментные спектральные (кепстральные) признаки. Полученные из набора надсегментных спектральных признаков векторы используются в классификаторе для отнесения исходного голосового образца к одному из классов. Для решения задачи классификации применялся метод опорных векторов SVM (Support Vector Machine) [8].



Рис. 3. Тестовая система анализа и классификации голосового сигнала

Fig. 3. Voice signal analysis and classification test system

Выделение и формирование признаков показаны на рис. 4. Исходный голосовой сигнал обрабатывался сегментами длительностью до 1 с. Для каждого такого сегмента проводился кратковременный анализ с вычислением спектральных признаков, которые достаточно точно описывают локальную структуру спектра сигнала на временном промежутке от 10 до 40 мс. Однако в результате такого кратковременного анализа формируется чрезмерно большой набор признаков для эффективной классификации с помощью машинного обучения [8]. С целью уменьшения объема данных, используемых для работы классификатора, полученные в результате кратковременного анализа признаки усреднялись и уже на их основе формировался надсегментный вектор признаков.

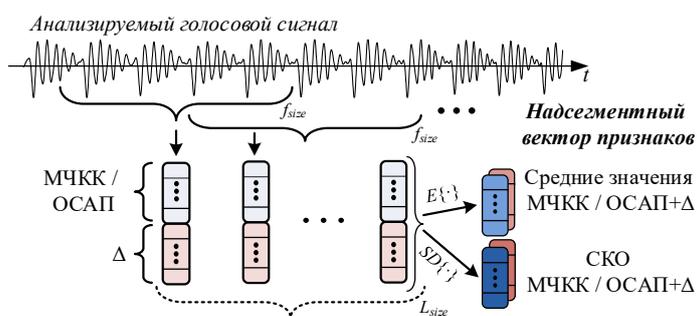


Рис. 4. Принципы формирования надсегментных признаков

Fig. 4. Principles of formation of suprasegmental features

Для формирования надсегментных признаков из наборов кратковременных спектральных признаков по каждому набору рассчитывались их средние значения и СКО.

Кратковременные спектральные признаки использовались в двух вариантах:

1. Малоразмерные спектральные признаки на основе дискретной оценки ОСАП, вычисляемые в соответствии с выражением (4).

2. Малоразмерные спектральные признаки на основе МЧКК, вычисляемые с использованием метода, описанного в работе [3]. Данный метод базируется на вычислении энергетического спектра сигнала с помощью ДПФ, который переводится в мел-шкалу путем взвешивания на треугольные окна с равной шириной в шкале мелов. Для вычисления МЧКК от энергетического спектра вычисляется логарифм и применяется дискретное косинусное преобразование второго типа.

На этапе проектирования классификатора с целью повышения эффективности его работы производился отбор признаков с использованием метода LASSO (Least Absolute Shrinkage and Selection Operator) [9]. В этом случае применяется регрессионный анализ, который хорошо подходит для отбора признаков, имеющих сильную корреляцию друг с другом. При использовании метода LASSO решается задача линейной регрессии со штрафной функцией, накладываемой на абсолютные значения коэффициентов модели. Задача решается итеративно для определенного числа возрастающих значений параметра регуляризации, причем веса при некоторых признаках модели приближаются к нулю или становятся равными нулю. По порядку, в котором обнуляются параметры модели (признаки), можно судить о степени влияния того или иного набора признаков на качество работы классификатора.

Для обучения и тестирования системы использовалась база, содержащая 54 образца голосовых сигналов (протяжный звук /a/), описанная в работе [10]. База содержит записи 39 здоровых человек (23 мужчин, 16 женщин) и 15 больных БАС (6 мужчин, 9 женщин).

**Экспериментальные исследования эффективности предложенного формирования спектральных признаков.** С помощью описанной выше системы анализа и классификации голосового сигнала (см. рис. 3) было проведено сравнение качества классификации для кратковременных спектральных признаков, полученных с применением предлагаемого метода и на основе МЧКК.

При проведении экспериментов для вычисления признаков использовались следующие значения параметров системы анализа:

- длительность сегмента около  $L_{size}$  800 мс;
- межсегментное перекрытие  $f_{size}$  75 %;
- количество параметров в частотной области  $M$ , которые используются для формирования вектора признаков, от 10 до 60. В случае вычисления МЧКК значение  $M$  равно числу кепстральных коэффициентов, а в случае вычисления спектральной огибающей – количеству отсчетов в частотной области в целевом диапазоне частот;
- размерность вектора надсегментных признаков равна  $1 \cdot 4 \cdot (M - 1)$  и варьируется от 36 до 236. Так, например, для  $M = 20$  она равна 1·76. (Кепстральные коэффициенты и отсчеты спектральной огибающей с нулевым индексом не используются для формирования вектора признаков.);
- нижняя и верхняя границы частотного диапазона, используемого для формирования признаков, от 50 до 6400 Гц;
- размер кадра анализа от 10 до 40 мс.

Для оценки производительности классификатора использовался метод перекрестной проверки (англ. cross-validation) LOSO (Leave-One-Subject-Out) [11]. Относительно рассматриваемой системы применение данного метода предполагало итеративное тестирование классификатора. На каждой итерации в тестовый набор включались данные, соответствующие одному диктору, а в обучающий – все оставшиеся. Количество итераций равнялось количеству дикторов. На каждой итерации определялась сбалансированная точность классификации:

$$BA = 0,5 \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right), \quad (6)$$

где  $TP$ ,  $TN$ ,  $FP$ ,  $FN$  – количество истинно положительных, истинно отрицательных, ложноположительных и ложноотрицательных результатов классификации.

Оценки работы классификатора для каждой итерации использовались далее для вычисления среднего значения по всем итерациям  $\overline{BA}$ . Полученные результаты представлены в табл. 3 и 4, жирным шрифтом выделены лучшие показатели. Видно, что применение признаков на основе дискретной оценки ОСАП позволило добиться более высокого значения полноты классификации для всех размеров кадра анализа, применявшихся в экспериментах. Наибольшая сбалансированная точность классификации при использовании признаков на основе дискретной оценки ОСАП достигалась при количестве частотных отсчетов около 20.

Таблица 3  
Максимальная средняя сбалансированная точность классификации при использовании МЧКК признаков в зависимости от количества частотных отсчетов и длительности кадра

Table 3  
The maximum average balanced accuracy of the classification when using MFCC features, depending on the number of frequency samples and the duration of the frame

| Длительность кадра анализа, мс | Сбалансированная точность классификации $\overline{BA}$ , % |      |      |      |      |      |      |      |      |      |      |
|--------------------------------|---|------|------|------|------|------|------|------|------|------|------|
|                                | Balanced classification accuracy $\overline{BA}$ , %        |      |      |      |      |      |      |      |      |      |      |
|                                | M≤10  | M≤15 | M≤20 | M≤25 | M≤30 | M≤35 | M≤40 | M≤45 | M≤50 | M≤55 | M≤60 |

| Analysis frame duration, ms |      |      |      |             |             |      |      |             |             |             |      |
|-----------------------------|------|------|------|-------------|-------------|------|------|-------------|-------------|-------------|------|
| 10                          | 69,3 | 70,5 | 71,7 | 71,7        | <b>73,2</b> | 67,2 | 70,3 | 72,4        | 71,9        | 70,7        | 69,7 |
| 15                          | 70,1 | 70,2 | 70,1 | 70,1        | 69,1        | 71,6 | 71,8 | 73,0        | <b>73,7</b> | 69,9        | 70,8 |
| 20                          | 71,3 | 68,3 | 71,3 | 72,5        | 73,9        | 72,6 | 70,8 | <b>74,4</b> | 71,2        | 72,1        | 72,1 |
| 25                          | 72,5 | 68,8 | 72,2 | 72,7        | <b>75,7</b> | 69,3 | 70,4 | 72,3        | 69,4        | 72,0        | 70,8 |
| 30                          | 73,0 | 67,5 | 74,3 | <b>74,7</b> | 70,9        | 69,7 | 70,2 | 70,1        | 70,3        | 72,2        | 69,7 |
| 35                          | 74,3 | 67,3 | 69,3 | 75,4        | 71,2        | 69,7 | 75,0 | 69,6        | 72,3        | <b>77,3</b> | 69,5 |
| 40                          | 73,9 | 67,1 | 70,9 | <b>75,0</b> | 67,9        | 69,3 | 71,4 | 68,8        | 72,5        | 71,3        | 72,8 |

Таблица 4

Максимальная средняя сбалансированная точность классификации при использовании предлагаемых признаков в зависимости от количества частотных отсчетов и длительности кадра

Table 4

The maximum average balanced accuracy of the classification when using proposed features, depending on the number of frequency samples and the duration of the frame

| Длительность кадра анализа, мс<br>Analysis frame duration, ms | Сбалансированная точность классификации $\overline{BA}$ , %<br>Balanced classification accuracy $\overline{BA}$ , % |      |             |      |      |      |      |      |      |      |      |
|---|---|------|-------------|------|------|------|------|------|------|------|------|
|   | M≤10  | M≤15 | M≤20        | M≤25 | M≤30 | M≤35 | M≤40 | M≤45 | M≤50 | M≤55 | M≤60 |
| 10  | 72,5  | 81,0 | <b>81,6</b> | 80,4 | 74,2 | 74,6 | 73,1 | 78,5 | 78,2 | 78,5 | 75,7 |
| 15  | 69,2  | 79,0 | <b>80,3</b> | 79,5 | 76,6 | 74,4 | 74,7 | 76,6 | 78,5 | 77,6 | 78,9 |
| 20  | 67,4  | 78,6 | <b>80,5</b> | 79,4 | 77,9 | 75,2 | 74,8 | 77,9 | 79,7 | 79,4 | 77,7 |
| 25  | 67,1  | 77,0 | <b>81,5</b> | 77,5 | 77,7 | 75,4 | 73,3 | 75,6 | 78,1 | 78,6 | 77,2 |
| 30  | 67,0  | 74,4 | <b>82,7</b> | 76,2 | 77,4 | 73,9 | 72,7 | 74,1 | 78,6 | 77,5 | 77,1 |
| 35  | 66,8  | 74,6 | <b>81,0</b> | 76,2 | 78,7 | 74,1 | 72,2 | 73,7 | 77,5 | 77,2 | 76,8 |
| 40  | 66,2  | 74,5 | <b>81,5</b> | 75,7 | 77,5 | 73,6 | 71,7 | 72,9 | 76,5 | 78,3 | 75,6 |

**Закключение.** В работе предложены спектральные признаки на основе дискретной оценки ОСАП, которые являются альтернативой классическим мел-спектральным признакам. Результаты проведенных экспериментов для определения патологических изменений голоса свидетельствуют, что представленный метод позволяет более эффективно извлекать полезную для классификации информацию из имеющихся обучающих данных и ощутимо повышать точность классификации.

Характерной особенностью задачи определения патологических изменений голоса является крайне малый объем доступных обучающих данных, что обусловлено высокой сложностью их получения [10]. Это делает использование сложных моделей классификации, например на основе глубоких нейросетей, нецелесообразным вследствие явления переобучения [8]. Повышение точности классификации в таких условиях возможно при использовании простых, но более адекватных речевому процессу моделей, что является важной практической составляющей полученного результата. Таким образом, предложенный подход имеет хорошую перспективу применения для задач обработки речевых и голосовых сигналов, в частности, в системах машинного обучения с малым объемом обучающих данных.

**Вклад авторов.** Д. С. Лихачёв проработал идею использования генеративного подхода на основе авторегрессионной модели для вычисления малоразмерных признаков и провел исследование ее эффективности. М. И. Вашкевич и Н. А. Петровский приняли участие в программной реализации системы, предназначенной для постановки экспериментов. И. С. Азаров определил цели и задачи исследования. Все авторы принимали участие в подготовке текста статьи, анализе и интерпретации результатов экспериментов.

#### Список использованных источников

1. Towards robust voice pathology detection / P. Harar [et al.] // Neural Computing and Applications. – 2020. – Vol. 32, no. 20. – P. 15747–15757.

2. Benba, A. Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis / A. Benba, A. Jilbab, A. Hammouch // *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. – 2016. – Vol. 24, no. 10. – P. 1100–1108.
3. Vashkevich, M. Classification of ALS patients based on acoustic analysis of sustained vowel phonations / M. Vashkevich, Y. Rushkevich // *Biomedical Signal Processing and Control*. – 2021. – Vol. 65. – P. 1–14.
4. Rabiner, L. R. *Fundamentals of Speech Recognition* / L. R. Rabiner, B. H. Juang. – Pearson Education, 1993. – 570 p.
5. Harris, F. J. On the use of windows for harmonic analysis with the discrete Fourier transform / F. J. Harris // *Proc. of the IEEE*. – Jan. 1978. – Vol. 66, no. 1. – P. 51–83.
6. Вашкевич, М. И. Система анализа и классификации голосового сигнала на основе пертурбационных параметров и кепстрального представления в психоакустических шкалах / М. И. Вашкевич, Д. С. Лихачёв, И. С. Азаров // *Доклады БГУИР*. – 2022. – Т. 20, № 4. – С. 73–82. <https://doi.org/10.35596/1729-7648-2022-20-1-73-82>
7. Markel, J. D. *Linear Prediction of Speech* / J. D. Markel, A. H. Gray. – Berlin, N. Y. : Springer-Verlag, 1976. – 290 p.
8. Flach, P. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data* / P. Flach. – Cambridge University Press, 2012. – 416 p.
9. *An Introduction to Statistical Learning with Applications in R* / G. James [et al.]. – Springer, 2013. – 440 p.
10. Vashkevich, M. Bulbar ALS detection based on analysis of voice perturbation and vibrato / M. Vashkevich, A. Petrovsky, Y. Rushkevich // *IEEE Intern. Conf. on Signal Processing: Algorithms, Architectures, Arrangements, and Applications, Poznan, Poland, 18–20 Sept. 2019*. – Poznan, 2019. – P. 267–272.
11. The necessity of leave one subject out (LOSO) cross validation for EEG disease diagnosis / S. Kunjan [et al.] // *Brain Informatics*. – Springer, 2021. – P. 558–567. [https://doi.org/10.1007/978-3-030-86993-9\\_50](https://doi.org/10.1007/978-3-030-86993-9_50)

---

## References

1. Harar P., Galaz Z., Alonso-Hernandez J. B., Mekyska J., Burget R., Smekal Z. Towards robust voice pathology detection. *Neural Computing and Applications*, 2020, vol. 32, no. 20, pp. 15747–15757.
2. Benba A., Jilbab A., Hammouch A. Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2016, vol. 24, no. 10, pp. 1100–1108.
3. Vashkevich M., Rushkevich Y. Classification of ALS patients based on acoustic analysis of sustained vowel phonations. *Biomedical Signal Processing and Control*, 2021, vol. 65, pp. 1–14.
4. Rabiner L. R., Juang B. H. *Fundamentals of Speech Recognition*. Pearson Education, 1993, 570 p.
5. Harris F. J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, January 1978, vol. 66, no. 1, pp. 51–83. <https://doi.org/10.1109/PROC.1978.10837>
6. Vashkevich M. I., Likhachov D. S., Azarov E. S. *Voice analysis and classification system based on perturbation parameters and cepstral presentation in psychoacoustic scales*. *Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioelektroniki [Reports of the Belarussian State University of Informatics and Radioelectronics]*, 2022, vol. 20, no. 1, pp. 73–82 (In Russ.). <https://doi.org/10.35596/1729-7648-2022-20-1-73-82>
7. Markel J. D., Gray A. H. *Linear Prediction of Speech*. Berlin, New York, Springer-Verlag, 1976, 290 p.
8. Flach P. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge University Press, 2012, 416 p.
9. James G., Witten D., Hastie T., Tibshirani R. *An Introduction to Statistical Learning with Applications in R*. Springer, 2013, 440 p.
10. Vashkevich, M., Petrovsky A., Rushkevich Y. Bulbar ALS detection based on analysis of voice perturbation and vibrato. *IEEE International Conference on Signal Processing: Algorithms, Architectures, Arrangements, and Applications, Poznan, Poland, 18–20 September 2019*. Poznan, 2019, pp. 267–272.
11. Kunjan S., Grummett T. S., Pope K. J., Powers D. M. W., Fitzgibbon S. P., ..., Lewis T. W. The necessity of leave one subject out (LOSO) cross validation for EEG disease diagnosis. *Brain Informatics*, Springer, 2021, pp. 558–567. [https://doi.org/10.1007/978-3-030-86993-9\\_50](https://doi.org/10.1007/978-3-030-86993-9_50)

*Лихачёв Денис Сергеевич*, кандидат технических наук, доцент кафедры электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники.  
E-mail: [likhachov@bsuir.by](mailto:likhachov@bsuir.by)

*Вашкевич Максим Иосифович*, кандидат технических наук, доцент кафедры электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники.  
E-mail: [vashkevich@bsuir.by](mailto:vashkevich@bsuir.by)

*Петровский Николай Александрович*, кандидат технических наук, доцент кафедры электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники.  
E-mail: [nick.petrovsky@bsuir.by](mailto:nick.petrovsky@bsuir.by)

*Азаров Илья Сергеевич*, доктор технических наук, заведующий кафедрой электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники.  
E-mail: [azarov@bsuir.by](mailto:azarov@bsuir.by)

*Denis S. Likhachov*, Ph. D. (Eng.), Assoc. Prof., Belarusian State University of Informatics and Radioelectronics, Computer Engineering Department.  
E-mail: [likhachov@bsuir.by](mailto:likhachov@bsuir.by)

*Maxim I. Vashkevich*, Ph. D. (Eng.), Assoc. Prof., Belarusian State University of Informatics and Radioelectronics, Computer Engineering Department.  
E-mail: [vashkevich@bsuir.by](mailto:vashkevich@bsuir.by)

*Nick A. Petrovsky*, Ph. D. (Eng.), Assoc. Prof., Belarusian State University of Informatics and Radioelectronics, Computer Engineering Department.  
E-mail: [nick.petrovsky@bsuir.by](mailto:nick.petrovsky@bsuir.by)

*Elias S. Azarov*, D. Sc., Assoc. Prof., Head of Computer Engineering Department, Belarusian State University of Informatics and Radioelectronics.  
E-mail: [azarov@bsuir.by](mailto:azarov@bsuir.by)