

## Детектирование речевых нарушений у пациентов с БАС на основе анализа голосового сигнала<sup>1</sup>

### Резюме

В статье рассмотрен подход к выявлению речевых нарушений у пациентов с боковым амиотрофическим склерозом (БАС) на основе акустического анализа голосового сигнала. Целью работы являлась проверка пригодности речевого задания на протяжное произнесение гласного звука /а/ для автоматического выявления пациентов с БАС. Рассмотрены четыре группы информационных признаков голосового сигнала, получающиеся в результате анализа контура частоты основного тона (ЧОТ): джиттер, шиммер, статистические параметры ЧОТ и частотно-временные параметры ЧОТ. Проведенные эксперименты показали, что использование предложенных методов акустического анализа голоса и классификации по методу к ближайших соседей позволило получить систему детектирования речевых нарушений у пациентов с БАС, обладающую точностью на уровне 95,7% (при чувствительности 91,5% и специфичности 97,4%).

**Ключевые слова:** боковой амиотрофический склероз, акустический анализ голоса, классификация

### Введение

Боковой амиотрофический склероз (БАС) является нейродегенеративным неизлечимым заболеванием, приблизительно 50% пациентов с БАС умирают в течение 30 месяцев с момента появления симптомов [1]. В настоящий момент не выявлено биомаркеров БАС и диагноз ставится на основании большого числа клинических наблюдений. В среднем на постановку диагноза уходит более года [2]. Использование акустического анализа голоса и речи является перспективным способом совершенствования процесса выявления БАС и мониторинга прогрессирования данного заболевания [3]. Это становится возможным, так как бульбарные моторные изменения (т. е., трудности с речью или глотанием) являются первыми симптомами примерно у 30% лиц с БАС [4], а на более поздних стадиях бульбарные симптомы появляются почти у всех пациентов с БАС. Важно, что атипичные значения акустических параметров голоса были выявлены у больных БАС при помощи теста на протяжное произнесение гласного звука, в то время, как в их разговорной речи не чувствовалось ощутимых изменений [5].

Необходимость разработки средств диагностики голоса и речи пробудила интерес к разработке специальных методов акустического анализа. Одним из важных аспектов современных исследований, связанных с выявлением неврологических заболеваний (таких как болезнь Паркинсона и БАС), является то, что они нацелены на использование смартфонов и планшетов для записи голоса с помощью стандартного микрофона в различных домашних условиях [6, 7]. Продвижение в этом направлении может привести к созданию простых в использовании средств для детектирования, отслеживания и телемониторинга заболеваний.

В предыдущих исследованиях для детектирования БАС использовались различные подходы и речевые задания. Одним из общих подходов, стремящихся к решению проблемы дифференциальной диагностики, является классификация типа дизартрии по речевому/голосовому сигналу [8, 9]. Основная трудность этого подхода заключается в том, что он требует сбора репрезентативной выборки аннотированных данных для всех типов дизартрии. Более простой подход к обнаружению БАС основан на использовании речевой базы, содержащей патологические и нормальные образцы голоса/речи [4, 7, 10-12]. В ряде

---

<sup>1</sup> Вашкевич М.И., Рушкевич Ю.Н. Детектирование речевых нарушений у пациентов с БАС на основе анализа голосового сигнала // Цифровая обработка сигналов. – 2019. – №4. – С. 33-41.

исследований обнаружение БАС осуществляется с использованием кинематических датчиков, размещаемых на артикуляционных органах для последующего моделирования артикуляции и измерения просодических элементов, таких как длительность гласных или скорость речи [2, 4, 13-15].

В работе [10] для обнаружения БАС был применен репрезентативный подход машинного обучения с использованием сверточных нейронных сетей. В качестве низкоуровневых признаков использовались энергии субполосных сигналов, полученных в банке фильтров, а также их первые и вторые производные. Наилучшая точность была получена с помощью сверточной нейронной сети, работающей во временной области (чувствительность 71,6% и специфичность 80,9%). В работе [7] был проанализирован огромный набор речевых признаков, полученных с помощью программного пакета OpenSmile, наиболее информативными были признаны мелчастотные кепстральные коэффициенты (МЧКК) и признаки, основанные на фильтрации речевого сигнала в модуляционной области (англ. RASTA – Relative Spectral Analysis). Однако классификатор на основе метода опорных векторов с линейным ядром, представленный в [7], имел невысокую точность (79% для мужчин и 83% для женщин). В работе [11] для выявления БАС использовался совместный анализ гласных звуков /а/ и /и/, извлеченных из тестового сигнала беглой речи. Классификатор, основанный на линейном дискриминантном анализе, представленный в [11] имел точность 88% (чувствительность 90,5% и специфичность 84,6%). Диодохокинетический тест (быстрое повторение слогов) использовалась для автоматического обнаружения БАС в работе [4]. Использование параметра фрактального джиттера, предложенного в [4], наряду с МЧКК и артикуляционными признаками позволило получить классификатор на основе экстремального градиентного бустинга (XGBoost) с точностью 90,2% (чувствительность 94,5% и специфичность 85,1%).

#### **Цели исследования:**

- 1) проверить пригодность простого теста на протяжное произнесение гласного звука /а/ для обнаружения в голосе признаков БАС;
- 2) проверка различных подходов к решению задачи детектирования речевых нарушений при БАС – метода линейного дискриминантного анализа и метода k ближайших соседей;
- 3) выявления набора наиболее информативных акустических признаков для решения задачи детектирования речевых нарушений при БАС.

## **1 Методы**

### **1.1 Амплитудные и частотные пертурбационные параметры голоса**

Джиттер - это мера вариативности периода основного тона. Поскольку джиттер оценивает краткосрочную вариацию, она не может быть отнесена к контролируемым (произвольным) изменениям частоты основного тона (ЧОТ). Таким образом, джиттер является показателем стабильности работы фонаторной подсистемы речеобразования. Высокий уровень джиттера возникает в результате нарушения нейромоторной функции [16]. В простейшем случае, джиттер определяется как средняя разница между длительностями последовательных периодов, деленная на среднее значение периода основного тона.

$$J_{loc} = \frac{\frac{1}{N-1} \sum_{i=2}^N |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N |T_i|}, \quad (1)$$

где  $T_i$  – длительность  $i$ -го периода основного тона, а  $N$  – количество периодов основного тона в сигнале.

В более общем случае, для оценки джиттера используют среднюю разницу между текущим значением периода и усредненным значением периода на нескольких смежных циклах. Такая мера называется коэффициентом возмущений периодов (англ. PPQ – period perturbation quotient):

$$J_L = \frac{\frac{1}{N-L+1} \sum_{i=1+(L-1)/2}^{N-(L-1)/2} \left| T_i - \frac{1}{L} \sum_{n=i-(L-1)/2}^{i+(L-1)/2} T_n \right|}{\frac{1}{N} \sum_{i=1}^N |T_i|}, \quad (2)$$

где  $L$  – параметр, определяющий, сколько периодов используется для вычисления «скользящего» среднего. Чаще всего используют значения  $L = 3, 5$  и  $55$  [16,17].

Шиммер – это мера, характеризующая степень вариативности амплитуды акустической волны во время фонации. В простейшем случае шиммер ( $S_{loc}$ ) определяется как средняя абсолютная разница между амплитудами последовательных периодов основного тона, делённая на среднюю амплитуду, по аналогии с выражением (1) [16]. Однако, на  $S_{loc}$  влияет постепенный равномерный (естественный) спад интенсивности голоса (см. рисунок 1).

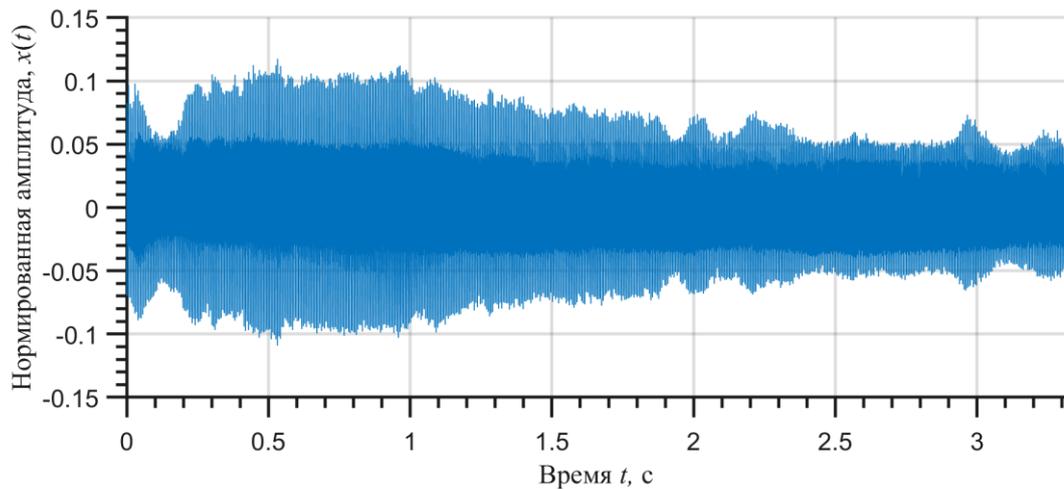


Рисунок 1 – Пример записи здорового человека с естественным спадом интенсивности голоса. Значение  $S_{loc} = 3,3\%$ , а коэффициент амплитудных возмущений  $S_3 = 1,8\%$

Для устранения эффекта "дрейфа" амплитуды сигнала и с целью получения более точной оценки шиммера обычно измеряют коэффициент амплитудных возмущений (англ. APQ – amplitude perturbation quotient) [16]:

$$S_L = \frac{\frac{1}{N-L+1} \sum_{i=1+(L-1)/2}^{N-(L-1)/2} \left| A_i - \frac{1}{L} \sum_{n=i-(L-1)/2}^{i+(L-1)/2} A_n \right|}{\frac{1}{N} \sum_{i=1}^N |A_i|}, \quad (3)$$

где  $A_i$  – амплитуда  $i$ -го периода основного тона,  $L$  – параметр, определяющий, сколько периодов используется для вычисления «скользящего» среднего. Чаще всего используют значения  $L = 3, 5, 11$  и  $55$  [16,17].

Как правило пертурбационные параметры (1)-(3) измеряются в процентах, поскольку абсолютные их значения невелики.

## 1.2 Энтропия периодов основного тона

В работе [18] предложен информационный признак – энтропия периодов основного тона (англ. pitch period entropy), который применялся в задаче классификации голосов пациентов с болезнью Паркинсона. В работе указывалось, что здоровые голоса обладают естественным плавным вибрато или микротремором, которые детектируются при помощи

параметра джиттера. Однако, общим симптомом дисфонии при болезни Паркинсона является нарушение способности контролировать стабильность частоты основного тона при протяжной фонации. Поэтому, используя традиционные параметры джиттера трудно отличить естественные, здоровые вариации периода основного тона от дисфонических вариаций, появляющихся в следствии болезни Паркинсона. Поскольку при БАС также нарушается способность контролировать стабильность частоты основного тона при протяжном произнесении звуков, то в данной работе предлагается также использовать энтропию периодов основного тона в качестве информационного признака.

Вычисление энтропии периодов основного тона (ЭПОТ) основано на следующих наблюдениях. Величина естественной вариации связана со средним значением частоты основного тона: человек, имеющий высокий голос обладает более выраженным вибрато, чем человек с низким голосом (в том случае, если вибрато оценивать по абсолютной частоте в Герцах). Поэтому более подходящим масштабом для оценивания отклонений вариаций в вибрато является перцептуальная логарифмическая (тональная) шкала. Также важно перед тем как оценивать отклонения в вариации вибрато удалить естественные (здоровые) вариации. В работе [18] описаны только общие шаги получения ЭПОТ. Ниже приводится алгоритм, применявшийся в данной работе для вычисления данного параметра.

**Шаг 1.** Расчет контура частоты основного тона  $f_o(m)$  для голосового сигнала  $s(n)$  выполняется при помощи алгоритма [19] с шагом по времени в 5 мс.

**Шаг 2.** Преобразование  $f_o(m)$  в логарифмическую шкалу полутонов:

$$p(m) = \frac{\ln \frac{f_o(m)}{f_{low}}}{\ln \sqrt[12]{2}},$$

где  $f_{low}$  – нижняя граница октавы, рассчитанной исходя из того, что среднее значение частоты основного тона ( $\mu_{f_o}$ ) находится в середине октавы:

$$f_{low} = \frac{\mu_{f_o}}{\sqrt{2}}.$$

**Шаг 3.** Применение «отбеливающего» фильтра к сигналу  $p(m)$  для удаления естественных вариаций:

$$r(m) = \sum_{i=0}^M a_i p(m-i), \quad a_0 = 1,$$

где  $a_i$  – коэффициенты фильтра линейного предсказания [20].

**Шаг 4.** Расчёт дискретного распределение вероятности появления относительных тоновых (англ. semitone) вариаций  $P(r)$ , путем расчета нормализованной гистограммы на интервале от -1.5 до 1.5, который разбит на  $N = 30$  равных интервалов.

**Шаг 5.** Расчёт энтропии для дискретного распределения вероятности  $P(r)$ :

$$PPE = - \sum_{i=1}^N P(r_i) \log_2 P(r_i)$$

Чем больше мера энтропии, тем больше наблюдаемые вариации превышают естественный уровень вариации основного тона в здоровом голосе.

На рисунке 2 приведен пример, иллюстрирующий процесс вычисления энтропии периодов основного тона для голоса здорового человека и пациента с БАС.

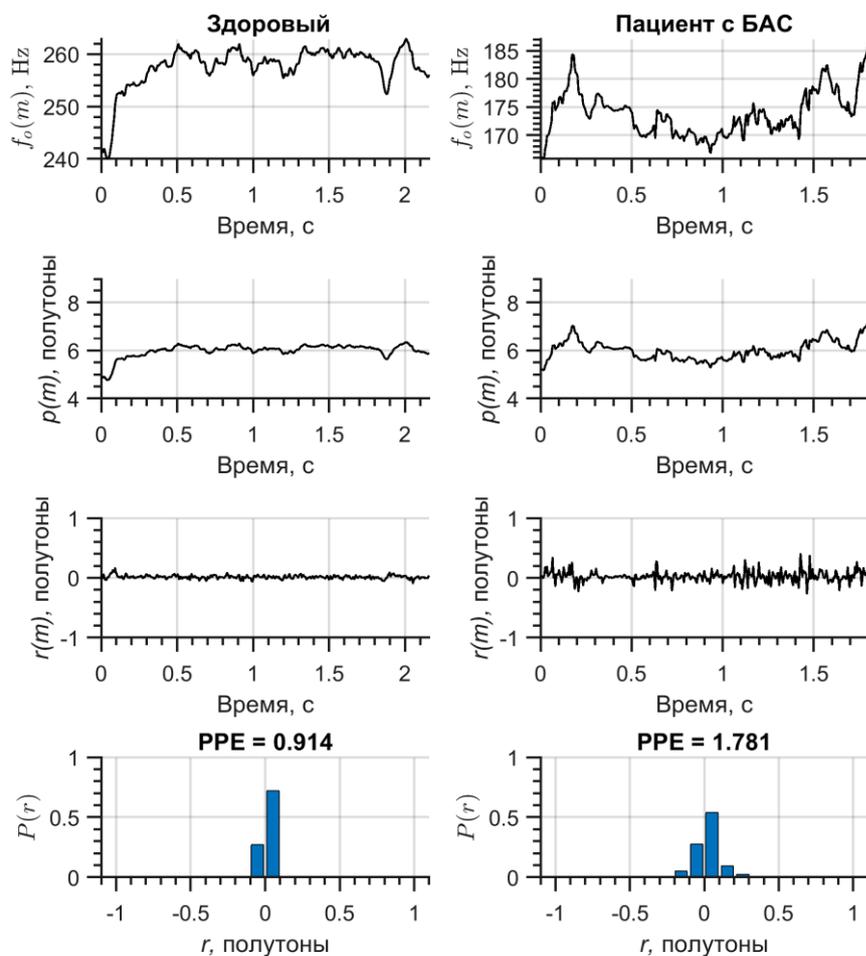


Рисунок 2 – Процесс вычисления ЭПОТ. Левая панель – здоровый человек; правая панель – пациент с БАС. Обозначения:  $f_o(m)$  – контур частоты основного тона в Гц,  $p(m)$  – частота основного тона в логарифмической шкале полутонов,  $r(m)$  – результат спектрального «отбеливания»  $p(m)$ ,  $P(r)$  – дискретная плотность распределения вероятности остаточных периодов основного тона  $r$ ,  $PPE$  – значение параметра ЭПОТ

### 1.3 Анализ вибрато

Вибрато – это быстрое и регулярное колебание ЧОТ, возникающее во время продолжительной стабильной фонации. Непосредственный анализ спектрограмм голосов здоровых людей и больных БАС показал, что существует значительная разница в частоте у данных групп. Оценка степени патологических изменений вибрато основана на наблюдении, что для здоровых голосов вибрато лежит в диапазоне 5-8 Гц [21], в то время как для больных БАС характерно наличие высокочастотных составляющих в диапазоне 9-14 Гц [22]. В данном исследовании мы используем следующий метод оценки индекса патологичности вибрато (англ. PVI – pathology vibrato index) [23]:

**Шаг 1.** Расчет контура частоты основного тона  $f_o(m)$  для голосового сигнала  $s(n)$  выполняется при помощи алгоритма [19] с шагом по времени в 5 мс.

**Шаг 2.** Нормализация контура частоты основного тона:

$$f'_o(m) = \frac{f_o(m)}{\mu_{f_o}},$$

где  $\mu_{f_o}$  – среднее значение частоты основного тона.

**Шаг 3.** Фильтрация  $f'_o(m)$  с помощью полосового фильтра Баттерворта с полой пропускания [9 14] Гц для сохранения частот в интересующем диапазоне.

**Шаг 4.** Оценка амплитудного спектра  $A_{f'_o}(f)$  методом Уэлча с временным окном в 1 секунду и перекрытием 95 %.

**Шаг 5.** Вычисление индекса патологичности вibrато:

$$PVI = \sum_{f \in [9,14] \text{ Гц}} A_{f'_o}(f).$$

На рисунке 3 приведен пример, иллюстрирующий процесс вычисления индекса патологичности вibrато для голоса здорового человека и пациента с БАС.

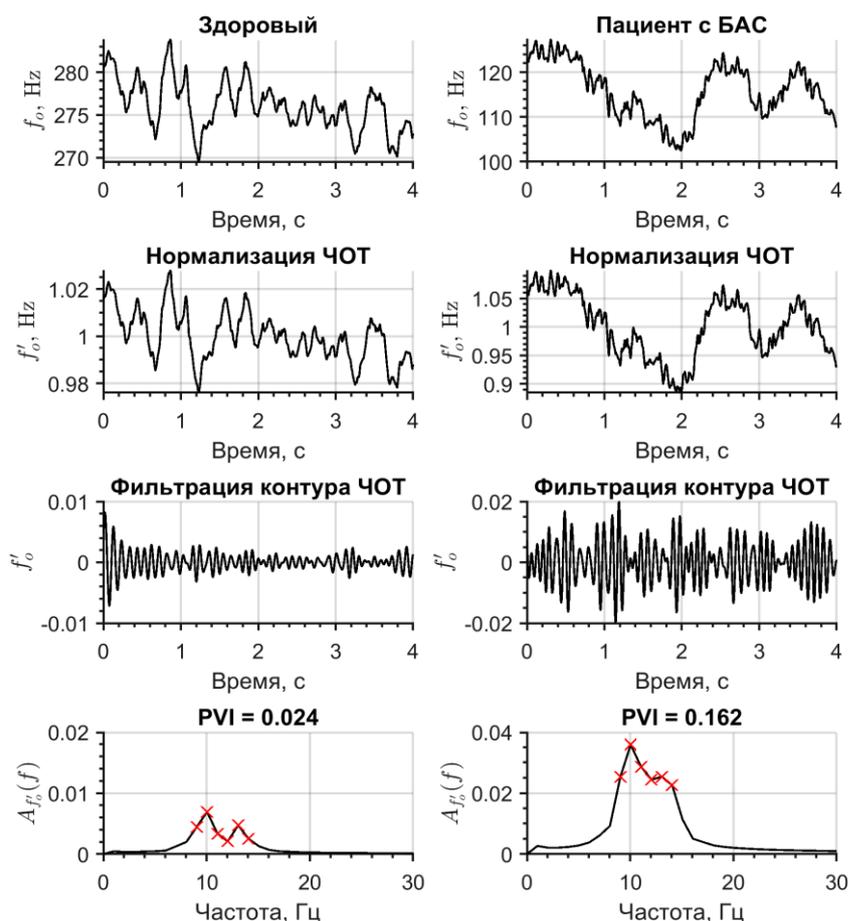


Рисунок 3 – Процесс вычисления индекса патологичности вibrато. Левая панель – здоровый человек; правая панель – пациент с БАС.

#### 1.4 Дополнительные признаки, получаемые путем анализа ЧОТ

В качестве информационного признака может использоваться параметр среднеквадратичного отклонения частоты основного тона –  $SD_{f_o}$  [16].

Также использовался частотный диапазон фонации (англ. PFR – phonatory frequency range), определяемый как

$$PFR = \frac{\log_{10} \frac{\max f_o}{\min f_o}}{\log_{10} 2},$$

который характеризует отношение максимального значения ЧОТ к минимальному во время фонации.

В работе анализировался направленный коэффициент возмущений (англ. DPF – directional perturbation factor) [17]:

$$DPF = \frac{N_{\Delta\pm}}{N} \times 100,$$

где  $N_{\Delta\pm}$  – число периодов на которых разница между последовательными циклами основного тона изменяла свой знак.

## 1.5 Классификация

Для различения двух классов нормальных и патологических голосов использовались два распространенных подхода машинного обучения: метод линейного дискриминантного анализа (ЛДА) и метод  $k$ -ближайшие соседи (англ.  $k$ -NN –  $k$  nearest neighbors).

### 1.5.1 Линейный дискриминантный анализ

Классификационная функция в методе ЛДА имеет следующий вид:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T \cdot \mathbf{x} + b),$$

где  $\mathbf{w}$  – вектор, определяющий нормаль к разделяющей гиперплоскости,  $\mathbf{x}$  – вектор информационных признаков (вычисленный для аудио-файла с записью голоса),  $b$  – смещение,  $\text{sign}(a)$  – функция возвращающая знак числа  $a$ . Если  $f(\mathbf{x}) = 1$ , то вектор  $\mathbf{x}$  относят к классу больных БАС, а при  $f(\mathbf{x}) = -1$  к классу здоровых.

Параметр  $\mathbf{w}$  ищется путем максимизации критерия Фишера:

$$J(\mathbf{x}) = \frac{(\mu_{\mathbf{w}}^+ - \mu_{\mathbf{w}}^-)^2}{(\sigma_{\mathbf{w}}^+)^2 + (\sigma_{\mathbf{w}}^-)^2},$$

где  $\mu_{\mathbf{w}}^+$  и  $\mu_{\mathbf{w}}^-$  средние значения проекций положительных и отрицательных обучающих векторов на вектор  $\mathbf{w}$ , а  $\sigma_{\mathbf{w}}^+$  и  $\sigma_{\mathbf{w}}^-$  – соответствующие среднеквадратичные отклонения. Таким образом, критерий Фишера минимизирует внутриклассовую вариативность и максимизирует межклассовую вариативность. Практическое руководство по вычислению  $\mathbf{w}$  и  $b$  можно найти в [24].

### 1.5.2 Классификация методов $k$ ближайших соседей

Идея метода  $k$ -NN состоит в назначении вектору  $\mathbf{x}$  метки, учитывая  $k$  ближайших к нему векторов из обучающей базы. В данной работе использовался следующий подход: для классификации нового вектора находилось  $K$  ближайших векторов из положительного класса  $\mathbf{x}_{1...K}^+$  и  $K$  ближайших векторов из отрицательного класса  $\mathbf{x}_{1...K}^-$ . Затем использовалась процедура голосования на основе взвешивания расстояний, т.е. чем ближе эталон (т.е. вектор из обучающей базы) к классифицируемому вектору, тем весомее его голос. В данной работе использовалась следующая функция:

$$f(\mathbf{x}) = \text{sign} \left( \sum_{k=1}^K \frac{1}{d(\mathbf{x}, \mathbf{x}_k^+)} + \frac{-1}{d(\mathbf{x}, \mathbf{x}_k^-)} \right),$$

где  $d(\mathbf{x}, \mathbf{y})$  – представляет собой расстояние Махаланобиса:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{y})},$$

где  $\mathbf{\Sigma}$  – ковариационная матрица, рассчитанная на обучающем наборе. Необходимость применения расстояния Махаланобиса объясняется тем, что используемые информационные признаки имеют ненулевую корреляцию.

### 1.5.3 Кросс-проверка

Для оценки качества классификаторов использовался метод перекрестной проверки по  $K$ -блокам (англ. *K-fold cross-validation*), которая заключается в следующем. Исходный набор данных перемешивается случайным образом и разбивается на  $K$  блоков. Далее выполняется обучение классификатора, причем один из блоков выступает, как тестовый набор, а оставшиеся  $K - 1$  в совокупности составляют обучающий набор. Эта процедура повторяется  $K$  раз, так, чтобы каждый блок один раз выступил в роли тестового набора. Метки, присвоенные классификаторами, для тестовых наборов сохраняются и по ним выполняется оценка производительности классификатора [24]. В качестве основных характеристик классификатора вычислялись точность (Acc), чувствительность (Sens), специфичность (Spec) и средняя полнота ( $R_{ave}$ ). Ниже приводятся выражения для их вычисления:

$$\text{Acc} = \frac{TP + TN}{TP + FP + TN + FN}, \quad \text{Sens} = \frac{TP}{TP + FN}, \quad \text{Spec} = \frac{TN}{TN + FP},$$
$$R_{ave} = \frac{1}{2} (\text{Sens} + \text{Spec}),$$

где  $TP, TN, FP, FN$  – истинно положительные, истинно отрицательные, ложно положительные, ложно отрицательные результаты классификации (под положительным результатом понимают верную постановку диагноза БАС). Точность характеризует частоту правильных решений, выносимых классификатором. Однако данный параметр не дает полного представления о качестве работы классификатора, особенно если в исходном наборе данных классы здоровых голосов и голосов с патологией не сбалансированы. Параметр чувствительности показывает способность классификатора детектировать патологию, если она есть. Специфичность характеризует способность классификатора определять отсутствие патологии, когда она действительно отсутствует. Средняя полнота – показывает общую способность классификатора относить голос к правильному классу.

## 2 Материалы

Записи голосов, использованные в данном исследовании, были собраны в Республиканском научно-клиническом центре неврологии и нейрохирургии (Минск, Беларусь). Всего было записано 54 человека, из них 39 здоровых (23 мужчины, 16 женщин) и 15 больных БАС с признаками бульбарных нарушений (6 мужчин, 9 женщин). Средний возраст в здоровой группе составил 41,9 лет (СКО 16,3), а средний возраст в группе БАС-57,7 лет (СКО 9,0). Всем участникам было предложено произносить протяжный гласный звук /а/ как можно дольше с комфортной высотой и громкостью. Фонация производилась на одном дыхании. Голосовые сигналы записывались с помощью смартфона с гарнитурой (частота дискретизации 44,1 кГц) и сохранялись в виде несжатых 16-битных wav-файлов. Средняя продолжительность записей составила 4,1 с. База голосов и Matlab-функции, используемые для анализа голоса, размещены в общедоступном репозитории<sup>2</sup>.

---

<sup>2</sup> <https://github.com/Mak-Sim/Troparion>

### 3 Результаты и обсуждение

#### 3.1 Статистический анализ акустических признаков

Для визуализации параметров голоса в контрольной группе и группе пациентов с БАС были рассчитаны статистические графики нескольких признаков. Для описания распределения применялись диаграммы размаха, которые позволяют увидеть медиану распределения и интервал, в который попадает 25% и 75% всех данных, а также максимальное и минимальное значение в выборке; также строился график плотности вероятности, получаемый путем сглаживания с ядром (англ. kernel smoothing).

На рисунке 4 представлены статистические графики коэффициента возмущений периодов  $J_3$  и коэффициента амплитудных возмущений  $S_3$ . Хотя по-видимому значения данных параметров в группах здоровых и пациентов с БАС сильно перекрываются, тем не менее данные параметры часто отбирались при автоматическом поиске наилучшего вектора информационных признаков. Это может говорить о том, что, хотя сами по себе эти параметры не позволяют выполнить разделение на больных и здоровых, они могут содержать дополнительную информацию, способную улучшить качество классификации.

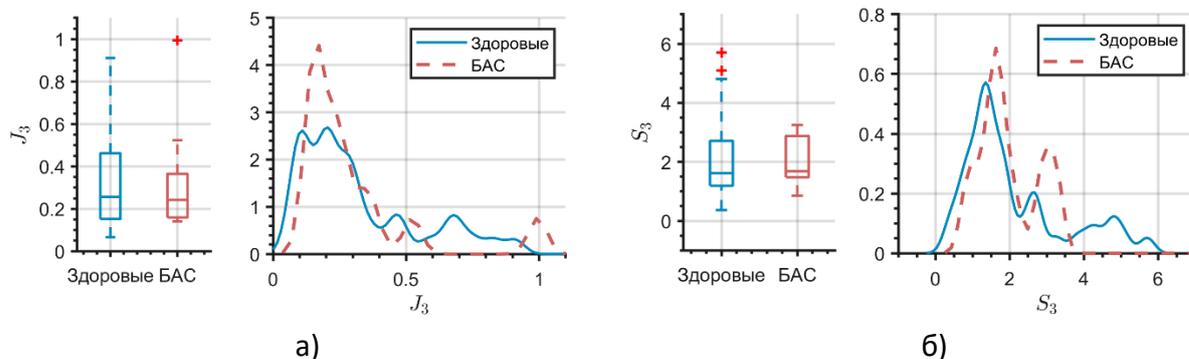


Рисунок 4 – Диаграмма размаха и плотность вероятности: а) коэффициент возмущения периодов ЧОТ  $J_3$ ; б) коэффициент амплитудных возмущений  $S_3$

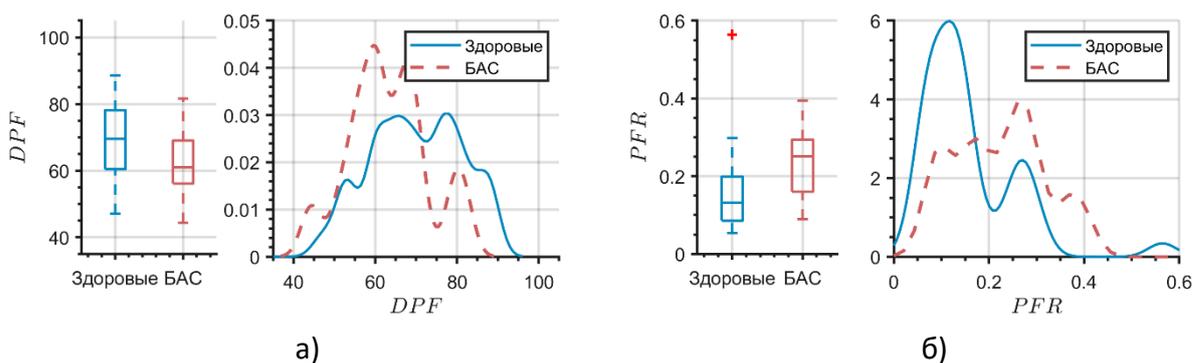


Рисунок 5 – Диаграмма размаха и плотность вероятности: а) направленный коэффициент возмущений  $DPF$ ; б) частотный диапазон фонации  $PFR$

На рисунке 5 приведены статистические графики направленного коэффициента возмущений и частотного диапазона фонации. В группе больных БАС медианное значение частотного диапазона фонации значительно выше, чем в группе здоровых. Это может объясняться тем, что пациенту с БАС сложнее сохранять стабильность ЧОТ во время фонации, что приводит к кратковременным отклонениям в меньшую или большую сторону от среднего значения ЧОТ.

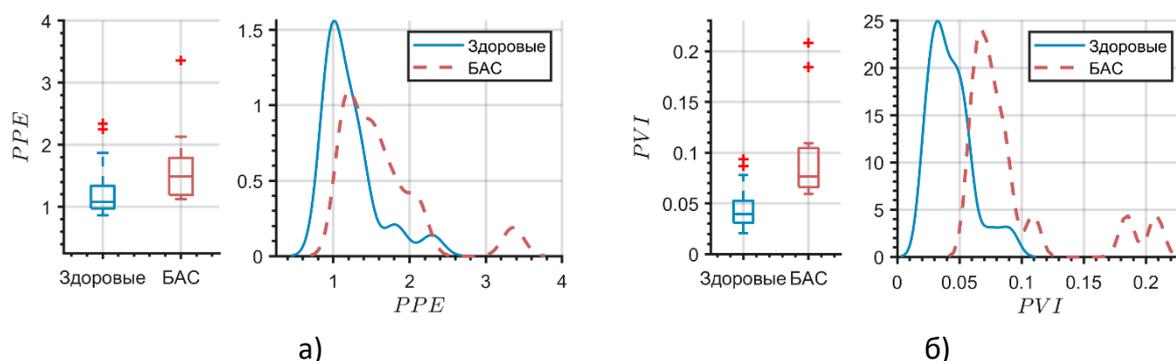


Рисунок 6 – Диаграмма размаха и плотность вероятности: а) энтропии периодов основного тона  $PPE$ ; б) индекса патологичности вибрато  $PVI$

На рисунке 6 приведены статистические графики энтропии периодов основного тона и индекса патологичности вибрато. В группе больных БАС медианное значение энтропии периодов основного тона значительно выше, чем в группе здоровых. Индекс патологичности вибрато показал наилучшую способность к разделению исходной выборки на здоровых и больных.

### 3.2 Результаты классификации

Были проведены эксперименты по классификации собранной базы голосов по методу кросс-проверки по  $K$  блокам ( $K = 6$ ) с использованием классификатора на основе линейного дискриминантного анализа и метода  $k$  ближайших соседей. Вектор информационных признаков состоял из следующих четырнадцати параметров, которые можно разбить на четыре категории:

- джиттеры ( $J_{loc}, J_3, J_5, J_{55}$ );
- шиммеры ( $S_{loc}, S_3, S_5, S_{11}, S_{55}$ );
- статистические параметры ЧОТ ( $SD_{f_0}, PFR$ );
- параметры, полученные путем анализа контура ЧОТ, т.е. частотно-временные параметры ( $DFP, PPE, PVI$ ).

Размерность вектора информационных признаков  $D = 14$  в решаемой задаче является относительно низкой, что позволило провести полный перебор всех возможных вариантов подмножеств признаков ( $2^D - 1$  вариантов), чтобы найти лучший из них.

Таблица 1 – Результаты классификации методом ЛДА

Вектор информационных признаков	Средняя полнота, %	Точность, %	Чувствительность, %	Специфичность, %
$S_{loc} S_3 S_{11} S_{55} PVI$	<b>89,9</b>	<b>91,3±1,9</b>	86,7±1,5	<b>93,1±2,4</b>
$S_{loc} S_3 S_{11} PVI$	89,8	91,0±2,5	87,0±1,4	92,6±3,6
$S_3 S_5 PVI DFP$	89,6	89,7±2,6	89,7±5,0	89,7±2,9
$S_3 S_5 PVI$	89,3	90,2±2,9	87,1±4,8	91,5±3,6
$S_3 PVI$	87,9	86,9±2,4	<b>90,0±4,5</b>	85,8±2,6

В таблице 1 приведены результаты классификации с использованием метода ЛДА. Результаты отсортированы по убыванию средней полноты, потому что наша цель получить классификатор, обладающий одновременно высокой специфичностью и чувствительностью. Первые три строчки в таблице соответствуют трем лучшим найденным комбинациям признаков, четвертая строка – лучшая комбинация, состоящая из трех признаков, а пятая

строка – лучшая комбинация из двух признаков. Нужно отметить тот факт, что все наборы признаков, представленные в таблице 1 включают параметр шиммера  $S_3$  и индекс патологичности вибрато (PVI).

В таблице 2 приведены результаты классификации с использованием метода k ближайших соседей, результаты также отсортированы по убыванию средней полноты. Первые три строчки в таблице соответствуют трем лучшим найденным комбинациям признаков, четвертая строка – лучшая комбинация, состоящая из двух признаков. Следует заметить, что, как и в случае с классификатором на основе ЛДА, во всех приведенных комбинация признаков присутствует параметр PVI, что говорит о его ключевом значении для правильной классификации.

Таблица 2 – Результаты классификации методом k ближайших соседей

Вектор информационных признаков	Средняя полнота, %	Точность, %	Чувствительность, %	Специфичность, %
$J_3$ PVI PPE	<b>94,4</b>	<b>95,7±2,5</b>	<b>91,5±3,0</b>	97,4±2,6
$J_5$ PFR PVI	92,1	94,8±1,2	86,0±2,9	<b>98,2±1,2</b>
$J_3$ PFR PVI	91,9	94,5±0,8	86,0±2,0	97,7±0,9
$J_3$ PVI	88,4	91,9±2,1	80,5±5,3	96,2±2,3

Общей особенностью классификаторов с использованием метода k ближайших соседей является то, что их специфичность значительно превышает чувствительность. В целом, если судить по значению средней полноты, то классификация методом k ближайших соседей более предпочтительна, так как позволяет достигнуть значения  $R_{ave} = 94,4\%$  по сравнению с 89,9% в методе ЛДА.

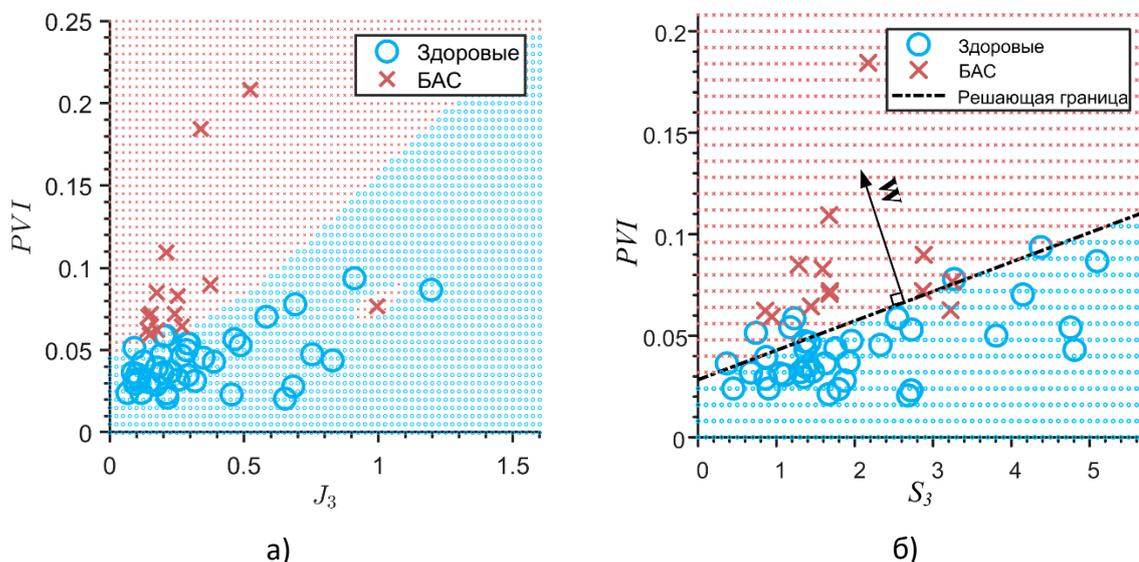


Рисунок 7 – Пример классификации: а) методом k ближайших соседей; б) методом ЛДА

Чтобы объяснить разницу между рассмотренными подходами обратимся к графической интерпретации процесса классификации. На рисунке 7 приведены примеры классификации по методу k ближайших соседей и методу ЛДА для случая, когда характеристический вектор признаков состоит из двух компонент. Из рисунка 7,б видно, что в методе ЛДА все векторное пространство разбивается на две части, после чего решение о том какому классу принадлежит вектор новых данных принимается в зависимости от того в какую об-

ласть он попадает. В случае метода к ближайших соседей решающая граница имеет нелинейный вид, кроме того, количество областей на которые разбивается все векторное пространство может быть больше двух (см. рисунок 7,а). Именно данной особенностью (т.е. большей гибкостью) объясняется преимущество метода к ближайших соседей над ЛДА.

### **Заключение**

В исследовании представлен подход к анализу голосового сигнала для обнаружения речевых нарушений при БАС. Рассмотрены два подхода к классификации голосовых сигналов – метод линейного дискриминантного анализа и метод к ближайших соседей. Наилучший достигнутый результат - точность 95,7% (чувствительность 91,5% и специфичность 97,4%) получен с применением метода к ближайших соседей. Наиболее релевантным признаком для построения классификаторов явился индекс патологичности вибрато (PVI). Представленные методы анализа голоса также могут быть полезны при распознавании различных степеней тяжести БАС. В будущем планируется изучить и другие дизартрические заболевания, чтобы определить насколько показатель PVI специфичен именно для БАС.

### **Литература**

1. Amyotrophic lateral sclerosis / M.C. Kiernan [et al.] // *Lancet*. – 2011. – vol.377 (issue 9769). – P. 942-955.
2. Detection of bulbar ALS using a comprehensive speech assessment battery/ Y. Yunusova [et al.] // *Proceedings of the International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications* – 2013. – P. 217-220.
3. Relation of automatically extracted formant trajectories with intelligibility loss and speaking rate decline in amyotrophic lateral sclerosis / R.L. Horwitz-Martin [et al.] // *Proceedings of Interspeech*. – 2016. – P. 1215–1219.
4. Fractal features for automatic detection of dysarthria / T. Spangler [et al.] // *IEEE EMBS International Conference on Biomedical Health Informatics*. – 2017 – P. 437–440.
5. Tomik, B. Dysarthria in amyotrophic lateral sclerosis: A review / B. Tomik, R.J. Guiloff // *Amyotrophic Lateral Sclerosis* – 2010. – vol. 11, № 1-2. – P. 4-15.
6. Benba, A. Discriminating between patients with Parkinsons 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.
7. Detection of amyotrophic lateral sclerosis (ALS) via acoustic analysis / R. Norel [et al.] // *Proceedings of Interspeech* – 2018.– P. 377–381.
8. Guerra, C. A modern approach to dysarthria classification / C. Guerra, D. Lovey // *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS)*. – 2003. – vol.~3. – P. 2257–2260.
9. Liss, J. M. Discriminating dysarthria type from envelope modulation spectra / J.M. Liss, S. LeGendre, A. J. Lotto // *Journal of Speech, Language, and Hearing Research*. – 2011. – № 5 (53). – P. 1246-1255.
10. Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks/ K. An [et al.] // *Proceedings of Interspeech* – 2018 – P. 1913-1917.
11. Features extraction for the automatic detection of ALS disease from acoustic speech signals/ M. Vashkevich [et al.] // *Proceedings of inter. conf. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. – 2018. – P. 321-326.
12. Детектирование бульбарных нарушений при боковом амиотрофическом склерозе на основе анализа речевого сигнала / Гвоздович А.Д. [и др.] // *Доклады БГУИР*.– 2018. – № 6 (116) –С. 52-58.

13. Comparison of speech tasks for automatic classification of patients with amyotrophic lateral sclerosis and healthy subjects/ Illa, A. [et al.]// Proceedings of IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP). – 2018 – P. 6014-6018.
14. Towards Automatic Detection of Amyotrophic Lateral Sclerosis from Speech Acoustic and Articulatory Samples/ J. Wang, [et al.] // Proceedings of Interspeech. – 2016 – P. 1195-1199.
15. Classification of bulbar ALS from kinematic features of the jaw and lips: Towards computer-mediated assessment / A. Bandini, [et al.] // Proceedings of Interspeech. – 2017 – P.1819-1823.
16. Baken, R.J., R.F. Orlikoff Clinical measurement of speech and voice, 2<sup>nd</sup> edition / R.J. Baken, Orlikoff R.F. // Thomson Learning, 2000. – 864 p.
17. Telephony-based voice pathology assessment using automated speech analysis / R. J. Moran [et al.] // IEEE Transactions on Biomedical Engineering – 2006. – vol. 53, no. 3. – P. 468-477.
18. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease / M. A. Little, [et al.] // IEEE Transactions on Biomedical Engineering. – 2009. – vol. 56, no. 4. – P. 1015-1022.
19. Азаров, И.С. Алгоритм оценки мгновенной частоты основного тона речевого сигнала / И.С. Азаров, М.И. Вашкевич, А.А. Петровский // Цифровая обработка сигналов. – 2012. – Вып. 4. – С. 49-57.
20. Рылов А.С. Анализ речи в распознающих системах – Мн.: Бестринт – 2003. – 264 с.
21. Nakano, T. An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features / T. Nakano, M. Goto, Y. Hiraga // Proceedings of Interspeech – 2006. – P. 1706–1709.
22. Rapid voice tremor, or flutter, in amyotrophic lateral sclerosis / A. E. Aronson [et al.] // Annals of Otology, Rhinology & Laryngology. – 1992. – vol. 101, no. 6. – P. 511–518.
23. Vashkevich, M. Bulbar ALS detection based on analysis of voice perturbation and vibrato / M. Vashkevich, A. Petrovsky, Yu. Rushkevich // Proceedings of inter. conf. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA). – 2019. – P. 267-272.
24. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах : пер. с англ. – М.: ДМК Пресс, 2015. – 400 с.

M.I. Vashkevich, Yu.N. Rushkevich

## Detection of impairment of speech function in patients with ALS based on analysis of voice signal

### Abstract

The paper presents an approach to the detection impairment of speech function in patients with amyotrophic lateral sclerosis (ALS) based on acoustic analysis of a voice signal. The aim of the work was to check the suitability of the sustain vowel /a/ phonation test for the automatic detection of patients with ALS. Four groups of acoustic features of a voice signal are considered that are obtained as a result of the analysis of the fundamental frequency (pitch): jitter, shimmer, statistical parameters of pitch and time-frequency parameters of the pitch contour. The experiments showed that the use of the proposed methods of acoustic voice analysis and classification according to the method of k nearest neighbors made it possible to obtain the detection system with accuracy of 95.7% (sensitivity 91.5% and specificity of 97,4%).

**Keywords:** amyotrophic lateral sclerosis, acoustic analysis, jitter, shimmer, classification

**Сведения об авторах**

Вашкевич М.И., к.т.н., доцент кафедры электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники  
vashkevich@bsuir.by

Рущкевич Ю. Н., к.м.н., доцент, ведущий научный сотрудник неврологического отдела РНПЦ Неврологии и нейрохирургии  
rushkevich@tut.by

**Information about the authors**

Vashkevich M.I., PhD, associate professor of computer engineering department of Belarusian state university of informatics and radioelectronics

Rushkevich Y.N. PhD, leading researcher of the neurology department of the Scientific and Technical Center for Neurology and Neurosurgery