

М.И. ВАШКЕВИЧ, И.С. АЗАРОВ
**ОПРЕДЕЛЕНИЕ ПАТОЛОГИИ ГОЛОСОВОГО АППАРАТА НА
ОСНОВЕ АНАЛИЗА МОДУЛЯЦИОННОГО СПЕКТРА РЕЧИ В
КРИТИЧЕСКИХ ПОЛОСАХ**

Вашкевич М.И., Азаров И.С. Определение патологии голосового аппарата на основе анализа модуляционного спектра речи в критических полосах.

Аннотация. Предложен метод определения патологических изменений голоса на основе анализа модуляционного спектра речевого сигнала. Модуляционный спектр вычисляется при помощи двух последовательных частотно-временных преобразований: первое переводит сигнал в частотную область, второе – в модуляционную. Особенностью предложенного метода является использование неравномерного частотно-временного плана при переводе сигнала в частотную область. Предложен способ оценки модуляционного спектра, основанный на принципах антропоморфической обработки сигналов, имитирующий неравномерное частотно-временное разрешение слуховой системы человека. Использован неравнополосный банк фильтров модулированный дискретным преобразованием Фурье, аппроксимирующий шкалу критических частот и повышающий временное разрешение в широких полосах. Банк фильтров использует фазовые (всепропускающие) звенья в качестве элементов задержки, что позволяет получить неравномерный частотно-временной план. Выполнено сравнение предложенного способа с более традиционным способом вычисления модуляционного спектра в критических полосах с фиксированным временным разрешением. Приведены данные экспериментов по использованию параметров модуляционного спектра для определения патологических изменений голоса, обусловленных: 1) боковым амиотрофическим склерозом (БАС); 2) заболеваниями гортани. Отбор и ранжирование параметров модуляционного спектра, использованных в качестве информационных признаков для классификации, выполнен при помощи метода LASSO. Классификация параметров на основе линейного дискриминантного анализа обеспечила высокую точность определения патологии (более 97%). Полученные результаты свидетельствуют о том, что использование неравномерного частотно-временного плана является предпочтительным в случае, когда анализируемый сигнал является протяжным гласным звуком, поскольку обеспечивает более высокую точность обнаружения патологии при меньшем числе модуляционных параметров.

Ключевые слова: анализ речевого сигнала, критические полосы, модуляционный спектр, модуляционные признаки, извлечение признаков, детектирование патологии по голосу.

1. Введение. Во многих практических задачах обработки речевых сигналов используется кратковременное спектральное представление, полученное в результате анализа коротких (5-50 мс), перекрывающихся временных интервалов аудиосигнала. Однако спектральное представление не описывает амплитудную модуляцию спектральных компонент сигнала, которая является важной для многих практических задач. Альтернативой классическому спектральному представлению является модуляционный

спектр, представляющий собой частотное разложение амплитудных огибающих спектральных компонент.

Низкочастотные амплитудные модуляции звука являются основными носителями информации в речи и музыке [1–3]. Так в работе [4] было показано, что в диапазоне от 1 до 12 Гц модуляционного спектра находится основная речевая информация и именно этот диапазон наиболее важен для распознавания речи с точки зрения слуховой системы человека. Кроме того, во многих работах показано, что представление речи в области модуляционных частот устойчиво к воздействию аддитивных шумов [5, 6]. Этой особенностью обусловлено широкое применение модуляционных параметров в системах распознавания речи [7–9], где устойчивость к воздействию шума является критически важным свойством. В частности, техника RASTA (англ. *RelAtive SpecTrAl*) является классическим методом анализа модуляционного спектра [10], который применяется в системах распознавания речи [7], а также в задаче идентификации диктора [11]. Параметры, полученные из модуляционного спектра, используются в задачах шумоподавления [12, 13] и классификации акустических сигналов [14, 15]. Модуляционный спектр также используют для предсказания разборчивости речи для людей с нормальным слухом и с тугоухостью [16].

Представление речи, фокусирующееся на низкочастотных модуляциях, также имеет прямую связь с динамикой работы артикуляционных органов, скорость движения которых находится в диапазоне от 2 до 12 Гц [17]. Нейрофизиологические исследования показали, что различные отделы головного мозга отвечают за восприятие различных модуляционных частот временной огибающей звукового сигнала [18]. Причем, обнаружена чувствительность к модуляционным частотам в низкочастотном диапазоне до 32 Гц (данный диапазон отвечает за восприятие манеры артикуляции и слоговую скорость – англ. *syllabic rate*) и в высокочастотном диапазоне от 64 до 256 Гц (данный диапазон отвечает за восприятие просодики и других акустических свойств голоса). Интересно, что в работе [18] не было обнаружено чувствительности к модуляционным частотам в диапазоне от 32 до 64 Гц.

В последнее десятилетие анализ модуляционного спектра начал применяться для выявления и классификации патологических изменений голоса [19–22]. Однако, в отличие от систем распознавания речи для выявления патологий голосового аппарата используется значительно более широкий диапазон (от 0 до 500 Гц), поскольку для диагностики важно описание процесса фонации, состоящего из повторяющихся сокращений голосовых связок. Исследования [19, 20] подтвердили, что информация, содержащаяся в расширенном диапазоне модуляционных частот от 0 до

500 Гц является важной для правильной диагностики патологических изменений голоса. Так, в [19] описана система выявления патологии голосового аппарата, использующая параметры модуляционного спектра и машину опорных векторов и имеющая точность 94,1%. Характерной особенностью методов [19, 20] является большая размерность параметров модуляционного спектра. Для сокращения размерности в [19] использовался метод, основанный на сингулярном разложении тензоров. В [20] проблема большой размерности данных решалась путем конструирования новых интегральных признаков на базе исходных модуляционных параметров. В настоящей работе предлагается другой подход, основанный на анализе модуляционного спектра в критических частотных полосах, который позволяет на начальной стадии существенно сократить размерность конечного представления голосового сигнала. Данный подход также соответствует парадигме антропоморфической обработки сигналов, согласно которой обработка информации должна строиться на тех же принципах, что и в слуховой системе человека. Следование данной парадигме привело к существенному улучшению систем автоматического распознавания речи [5, 7], что позволяет надеяться, что и для области детектирования патологий голосового аппарата она также окажется эффективной.

Рассмотрены два подхода к анализу модуляционного спектра сигнала в критических полосах: на основе дискретного преобразования Фурье (ДПФ) с объединением каналов [23] и на основе неравнополосного ДПФ-модулированного банк фильтров [24]. В разделе 6 сравниваются параметры модуляционного спектра, полученные при помощи указанных подходов применительно к задаче выявления патологии голосового аппарата.

2. Модуляционный спектр. Модуляционный спектр – трехмерное представление сигнала, отражающее информацию о распределении его энергии по модуляционным частотам, обнаруживаемым в спектральных огибающих. Абсцисса обычно отвечает за модуляционную частоту, ордината – за акустическую частоту, а аппликата – за акустическую энергию. В общем случае для получения модуляционного спектра требуются два преобразования: первое – для перехода от временного представления к акустическим частотам, а второе – для перехода от акустических частот к модуляционным [1] (см. рис. 1).

Часто в качестве обоих преобразований используются ДПФ [14, 19]. На первом шаге для дискретного сигнала $x(n)$ рассчитывается кратковременное преобразование Фурье (КВПФ) [19]:

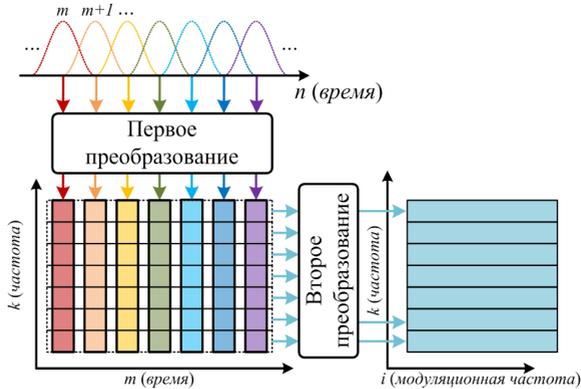


Рис. 1. Обобщенная схема оценки модуляционного спектра звукового сигнала

$$X_k(m) = \sum_{n=-\infty}^{\infty} h(mM - n)x(n)W_K^{kn}, \quad k = 0, \dots, K - 1, \quad (1)$$

где $W_K = e^{-j2\pi/K}$, $h(n)$ – оконная функция, используемая в процессе анализа акустических частот; M – перекрытие между соседними кадрами анализа. Перед тем как перейти к оценке модуляционного спектра, в каждой субполосе устраняется постоянная составляющая путем вычитания среднего значения огибающей сигнала $|X_k(m)|$. Далее применяют второе КВПФ к временным последовательностям огибающих субполосных сигналов:

$$X_l(k, i) = \sum_{m=-\infty}^{\infty} g(lL - m)|X_k(m)|W_I^{im}, \quad i = 0, \dots, I - 1, \quad (2)$$

где $g(m)$ – оконная функция, используемая в процессе анализа модуляционных частот; L – перекрытие между соседними кадрами анализа; k и i – индексы акустической и модуляционной частоты, соответственно.

Получаемая таким образом модуляционная спектрограмма отображает модуляционную спектральную энергию $|X_l(k, i)|$ на объединенной акустическо-модуляционной плоскости. Пример модуляционной спектрограммы, вычисленной для фрагмента протяжного произнесения звука [а] здоровым мужчиной (22 года), приведен на рисунке 2. В процессе

вычислений использовались следующие значения параметров в выражениях (1) и (2): $M = 44$ (что соответствует 1 мс на частоте дискретизации 44,1 кГц), $L = 78$, $h(n)$ – 132-точечное (или 3 мс) окно Хемминга, $g(m)$ – 78-точечное окно Хемминга.

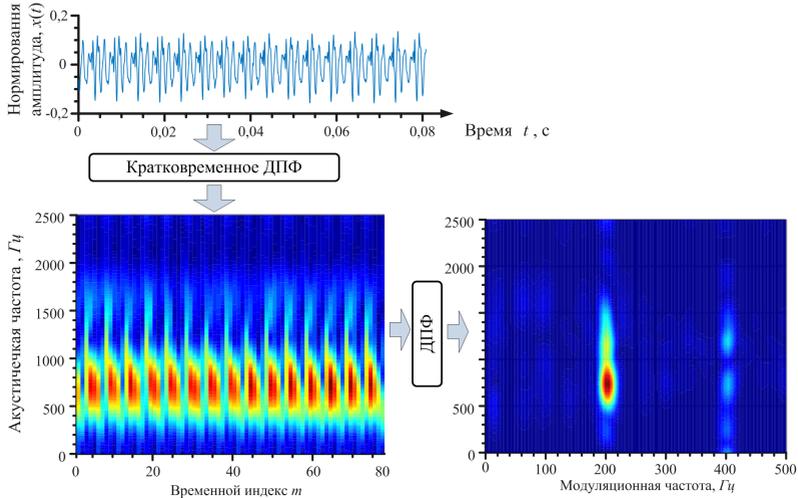


Рис. 2. Пример вычисления модуляционной спектрограммы для фрейма длительностью 80 мс

Протяжное произнесение звука не обладает синлабической (слоговой) и фонетической временной структурой и, соответственно, в нем отсутствуют типичные для речи модуляционные частоты в диапазоне 1-20 Гц. Поскольку при анализе использовалось временное окно $h(n)$ с длиной короче, чем период основного тона, то в спектральном представлении не различима гармоническая структура звука. Данная особенность приводит к тому, что основная энергия (компонент модуляционного спектра) концентрируется в области частоты основного тона (~ 200 Гц). Энергия локализуется на акустических частотах, соответствующих формантам звука [а] (на рисунке 2 первая форманта на ~ 750 Гц, а вторая на ~ 1100 Гц).

Недостатком рассмотренного подхода к представлению голосового сигнала в модуляционной области в контексте задачи классификации является большая размерность результирующего представления [1, 14]. В рассмотренном примере (как и в [19]) использовалось $K = 257$ акустических частот и $I = 257$ модуляционных частот, соответственно,

результатирующее представление является матрицей 257×257 , то есть содержит 66 049 параметров.

3. Оценка модуляционного спектра в критических частотных полосах. Одним из вариантов решения указанной выше проблемы большой размерности является использование вместо шкалы акустических частот (измеряемой в герцах), какой-либо психоакустической шкалы (например, мелов или барков) [14, 25], которая учитывает конфигурацию критических частотных полос. Аналогичный подход к преобразованию частоты выполняется при расчете мел-частотных кепстральных коэффициентов [7]. Так, если использовать число критических полос, равное $B = 20$, то в указанном выше примере число параметров модуляционного спектра будет равно $B \times I = 20 \times 257 = 5140$, то есть в 26 раз меньше первоначального.

Рассматриваются два подхода к оценке спектра сигнала в психоакустической шкале барков (1 барк = 1 критической полосе):

1) на основе вычисления ДПФ с последующим объединением спектральных коэффициентов [7, 23];

2) на основе пропускания сигнала через неравнополосный ДПФ-модулированный банк фильтров [23, 24].

Особенностью первого способа является его простота и вычислительная эффективность в силу существования быстрых алгоритмов вычисления ДПФ. Отличием второго способа является возможность анализа сигнала с различным временным разрешением в зависимости от анализируемой полосы частот. Как будет показано ниже, импульсные характеристики банка фильтров имеют большую длину на низких частотах и меньшую на высоких частотах, что позволяет более точно описывать частотно-временную структуру сигнала. В этом отношении анализ сигнала с использованием ДПФ менее гибкий, поскольку имеет фиксированное окно анализа для всех частотных полос.

3.1. Оценка спектра сигнала в шкале барков с использованием ДПФ. Для вычисления энергии спектра в шкале барков ДПФ-спектр сигнала, получаемый в (1), суммируется в критических полосах, которые измеряются в барках [23]:

$$E(b) = \sum_{k=kl(b)}^{kh(b)} |X(k)|^2, \quad b = 0, \dots, B - 1, \quad (3)$$

где $kl(b)$, $kh(b)$ – номер начального и конечного спектрального отсчета, попадающих в b -ю критическую полосу.

На рисунке 3 представлен пример оценки спектра сигнала в шкале барков с использованием выражения (3).

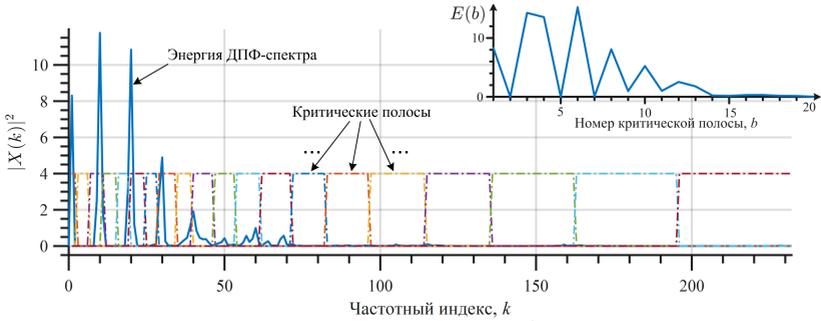


Рис. 3. Оценка энергии спектра сигнала в шкале барков: сплошной линией показана энергия ДПФ-спектра речевого сигнала, а штрихпунктирной – границы критических полос в шкале барков; в верхнем правом углу показана полученная оценка спектра в шкале барков

3.2. Неравнополосный ДПФ-модулированный банк фильтров. Неравнополосный ДПФ-модулированный банк фильтров образуется из своей равнополосной версии путем деформации частотной оси при помощи фазового преобразования [24]. Равнополосный банк фильтров (в виде набора импульсных характеристик) можно получить, используя комплексную модуляцию НЧ фильтра-прототипа:

$$h_b(n) = h_0(n)W_B^{bn}, \quad b = 0, \dots, B - 1, \quad (4)$$

где $n = 0, \dots, N - 1$; B – число полос банка фильтров; $h_0(n)$ – фильтр-прототип порядка N с конечной импульсной характеристикой (КИХ) и частотой среза $\omega_c = 2\pi/B$.

Если обозначить z -преобразование фильтра-прототипа $h_0(n)$ как $H_0(z)$, то выражение (4) представляется в z -области в виде:

$$H_b(z) = H_0(zW_B^b). \quad (5)$$

Если в выражении (5) выполнить замену $z = e^{j\omega}$, то получится выражение для комплексной частотной характеристики банка фильтров:

$$H_b(e^{j\omega}) = H_0(e^{j\omega}W_B^b) = H_0(e^{j(\omega - 2\pi b/B)}), \quad (6)$$

которое показывает, что банк фильтров состоит из сдвинутых по частоте копий фильтра-прототипа.

Название «ДПФ-модулированный» банк фильтр, описываемый выражением (4), получил вследствие того, что его эффективная реализация, получаемая в результате полифазного представления фильтра-прототипа, сводится к вычислению ДПФ размера B . Подробное описание реализации ДПФ-модулированного банка фильтров можно найти в [23].

Неравнополосная версия банка фильтров образуется в результате применения фазового преобразования [26], которое заключается в замене всех элементов задержки в структуре банка фильтров фазовыми звеньями:

$$z^{-1} \rightarrow A(z), \quad (7)$$

где $A(z)$ – передаточная характеристика фазового (всепропускающего) фильтра. В данной работе использовались фазовые фильтры первого порядка [27]:

$$A(z) = \frac{z^{-1} + \alpha}{1 + \alpha z^{-1}}, \quad \alpha \in \mathbb{R}, \quad |\alpha| < 1, \quad A(e^{j\omega}) = e^{j\varphi(\omega)},$$

где

$$\varphi(\omega) = \omega - 2 \arctan \frac{\alpha \sin \omega}{1 + \alpha \cos \omega}. \quad (8)$$

В результате замены (7) в банке фильтров происходит деформация частотной оси, описываемая выражением (8). Данный процесс проиллюстрирован на рисунке 4.

Выбор параметра фазового преобразования α в выражении (8) зависит от частоты дискретизации f_s , на которой работает банк фильтров, а также от психоакустической шкалы, которую банк аппроксимирует. Согласно [28] для аппроксимации шкалы барков параметр фазового преобразования должен выбираться по следующей формуле:

$$\alpha(f_s) = 1.048 \left[\frac{2}{\pi} \arctan(0,07212 f_s) \right]^{1/2} - 0.1957. \quad (9)$$

4. Практические аспекты реализации схем анализа модуляционного спектра сигнала. Для практических целей, рассматриваемых в данной работе, разработаны две схемы анализа модуляционного спектра сигнала – первая основана на вычислении ДПФ с последующим объединением каналов (везде далее – «схема на основе ДПФ»), а вторая на основе

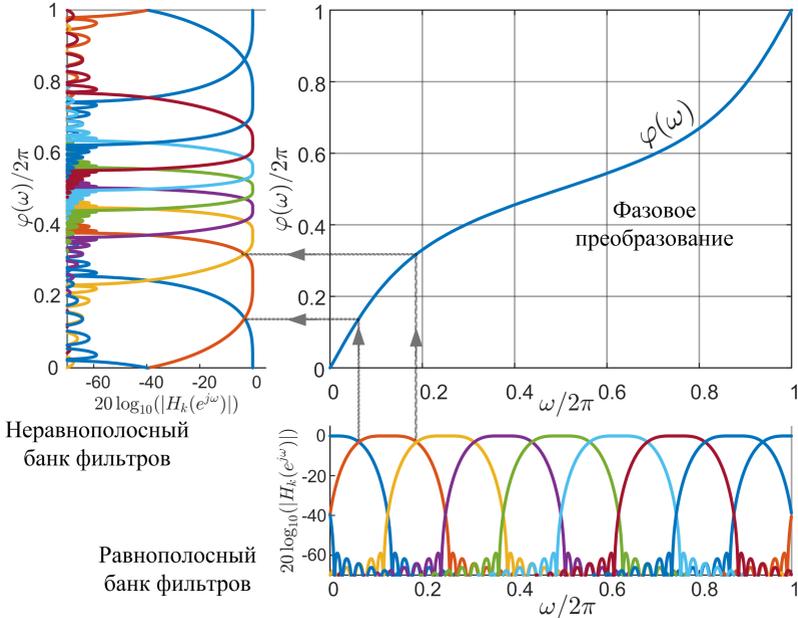


Рис. 4. Процесс получения неравнополосного банка фильтров с использованием фазового преобразования, коэффициент определяющий степень деформации частотной оси $\alpha = -0,4$

неравнополосного банка фильтров (везде далее – «схема на основе банка фильтров»)¹. Первая схема анализа представлена на рисунке 5.

На вход схемы поступает сигнал $x(n)$ длительностью T мс, если исходный сигнал имеет большую длительность, то он разбивается на сегменты по T мс, каждый из которых обрабатывается отдельно. Входной сигнал разбивается на фреймы размером N отсчетов с перекрытием h_{size} . Для каждого фрейма рассчитывается энергия ДПФ-спектра и производится объединение каналов для перехода к шкале барков. Полученные таким образом оценки спектра помещаются во временный буфер. В блоке «Выч. ср» выполняется вычитание среднего значения временной последовательности спектральных оценок в каждой полосе (среднее вычисляется после того, как временный буфер полностью заполнен). Далее происходит децимация временных спектральных последовательностей с коэффициентом D . На заключительном этапе рассчитывается модуляционный

¹Matlab-реализация данных схем размещена в общедоступном репозитории https://github.com/Mak-Sim/modulation_spectra_in_CB

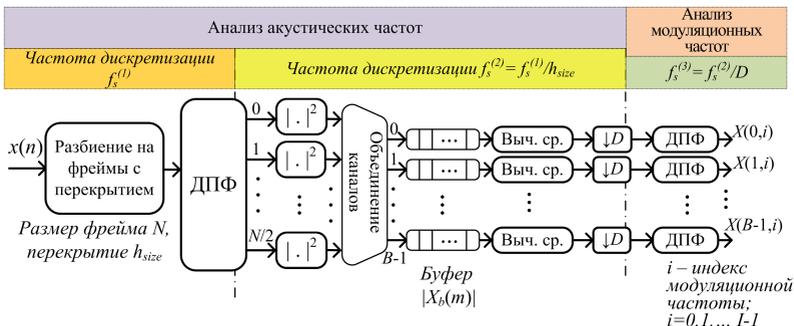


Рис. 5. Схема анализа модуляционного спектра сигнала: анализ акустических частот в шкале барков выполняется при помощи ДПФ с объединением каналов

спектр при помощи ДПФ. В работе использовались следующие значения параметров: частота дискретизации входного сигнала $f_s^{(1)} = 44100$ Гц, длительность входного сигнала $T = 800$ мс, длина фрейма $N = 1764$ (что соответствует окну анализа 40 мс), перекрытие $h_{size} = 21$ (или 0,5 мс), число каналов $B = 20$ (т.е. анализировались первые 20 критических полос), $f_s^{(2)} = 44100/h_{size} = 2100$ Гц, коэффициент децимации D имеет рациональное значение $\frac{525}{128}$, $f_s^{(3)} = 2100/D = 512$ Гц, число модуляционных частот $I = 257$ (для чего размер ДПФ выбирался равным 512). В результате анализа для каждой критической полосы выполнялась оценка модуляционного спектра в диапазоне от 0 до 256 Гц с шагом 1 Гц.

Вторая схема анализа модуляционного спектра, использующая неравнополосный банк фильтров, показана на рисунке 6. Входной сигнал подается аналогичным образом, как и в предыдущей схеме. Далее выполняется его децимация с коэффициентом R_1 таким образом, чтобы в результирующий частотный диапазон попадало ровно B критических полос. Далее сигнал подается в банк фильтров, где пропускается через цепочку фазовых звеньев (обеспечивающих деформацию частотной оси), умножается на коэффициенты фильтра-прототипа с последующим суммированием для получения выходов полифазных компонент фильтра. После чего выполняется децимация канальных сигналов с коэффициентом R_2 , за которой следует вычисление ДПФ, завершающее формирование выходов банка фильтров. Последующие элементы имеют такое же значение, как и в схеме на основе ДПФ (рис. 5).

Коэффициенты децимации R_1 и R_2 выбраны таким образом, чтобы выполнялось равенство $R_1 R_2 = h_{size}$. Этим достигается согласованность частот дискретизации рассматриваемых схем. Параметры схемы имели

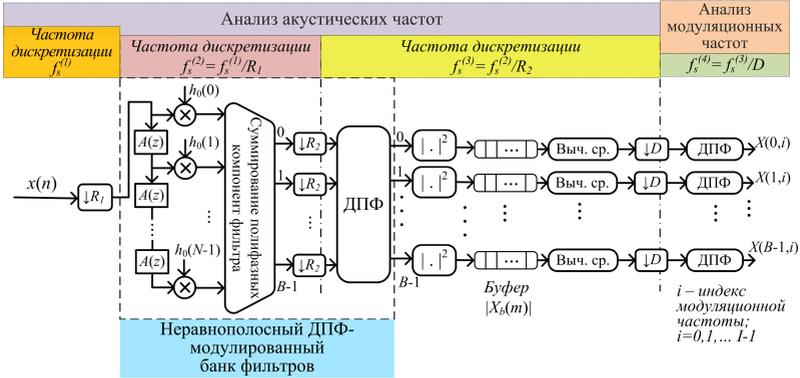


Рис. 6. Схема анализа модуляционного спектра сигнала: анализ акустических частот в шкале барков выполняется при помощи ДПФ с объединением каналов

следующие значения: $R_1 = 7/2$, $R_2 = 6$, $f_s^{(2)} = 44100/R_1 = 12600$ Гц, $f_s^{(3)} = 12600/R_2 = 2100$ Гц. Оставшиеся параметры имели такие же значения, как и в схеме на рисунке 5.

В качестве фильтра-прототипа для банка фильтров рассчитан КИХ-фильтр 200 порядка [29]. Амплитудно-частотная характеристика (АЧХ) результирующего неравнополосного банка фильтров показана на рисунке 7.

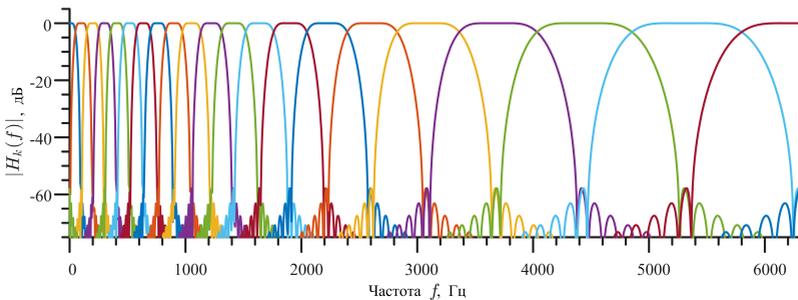


Рис. 7. АЧХ 20-канального неравнополосного ДПФ-модулированного банка фильтров, используемого в схеме на рисунке 6; коэффициент фазового преобразования $\alpha = 0,52$ (выбирался согласно выражению (9))

Как упоминалось ранее, отличительная черта неравнополосного ДПФ-модулированного банка фильтров состоит в том, что временной масштаб, анализирующих функций (фильтров) не является постоянным, но зависит от анализируемой полосы частоты. Низкие частоты анализиру-

ются на большем временном окне, а с возрастанием частоты окно анализа уменьшается. Таким образом происходит имитация представления сигнала периферийной частью слуховой системы человека. На рисунке 8 показаны выборочные примеры импульсных характеристик фильтров, входящих в состав банка.

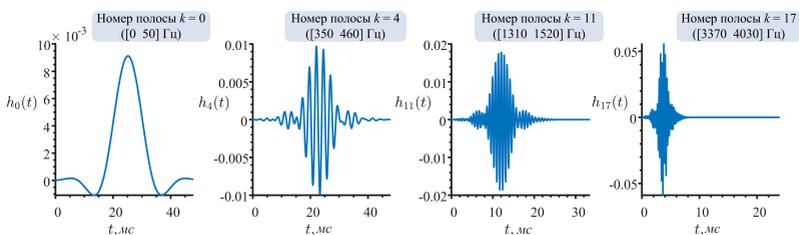


Рис. 8. Вид импульсных характеристик неравнополосного банка фильтров (на графиках показаны только действительная часть импульсных характеристик)

Анализ сигнала в неравнополосном ДПФ-модулированном банке фильтров имеет сходство с вейвлет-анализом (временной масштаб окна анализа зависит от анализируемой ширины полосы частот). Разница заключается в том, что в неравнополосном банке фильтров за управление шириной полос и временным масштаб импульсных характеристик фильтров отвечает один коэффициент фазового преобразования α , в то время как в вейвлет-анализе необходимо непосредственно задавать коэффициент масштабирования для каждой отдельной полосы. Кроме того, как показано в [30], точность аппроксимации психоакустической шкалы при использовании пакета дискретного вейвлет преобразования ниже, чем в случае использования банка фильтров с фазовым преобразованием.

5. Пример анализа речевого сигнала. На рисунке 9 показан пример анализа одного и того же речевого сигнала при помощи ДПФ с объединением каналов и при помощи неравнополосного банка фильтров. Анализ выполнялся согласно схемам на рисунках 5 и 6. Данное представление сигнала (рис. 9) соответствует тому, которое подается на анализ модуляционных частот (для улучшения визуализации операция вычитания среднего значения не выполнялась, а сами значения показаны в логарифмическом масштабе).

Приведенные графики выявляют особенности анализа, присущие рассматриваемым схемам. Схема анализа на основе ДПФ формирует сглаженное частотно-временное распределение энергии сигнала за счет относительно широкого (40 мс) временного окна анализа. В свою очередь, схема анализа на основе неравнополосного банка фильтров позволяет

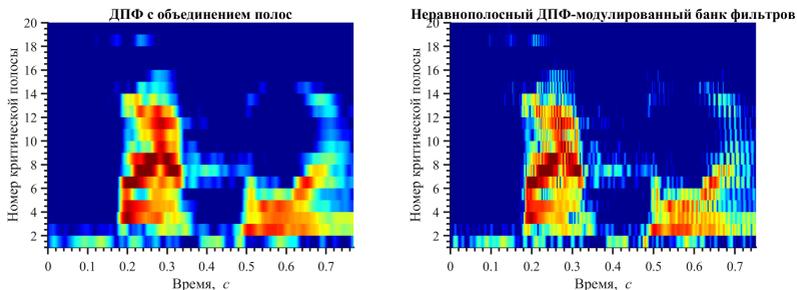


Рис. 9. Представление речевого сигнала (слово «Цапля», произнесенное мужским голосом) в системах анализа на основе ДПФ с объединением полос (слева) и на основе неравнополосного банка фильтров (справа)

получить более детализированное представление. На низких частотах (1-6 критические полосы) рассматриваемые представления схожи, поскольку эффективная длина окон анализа в обеих схемах для данных частот приблизительно совпадает. Однако, на более высоких частотах, в схеме на основе ДПФ наблюдается размытость спектральных составляющих, в то время как в схеме на основе неравнополосного банка фильтров высокочастотные компоненты имеют хорошую временную локализацию.

6. Применение параметров модуляционного спектра для выявления патологии голосового аппарата. Акустический анализ голоса используется при построении систем детектирования неврологических заболеваний [31], таких как болезнь Паркинсона [32,33], боковой амиотрофический склероз [34–37] и другие. [38]. Его также применяют для выявления заболеваний гортани, таких как узлы, полипы, парезы [39,40], рак [41] и прочие [42]. В большинстве подобных работ исследователи используют методы анализа во временной области (джиттер, шиммер) [37,41,43], а также методы анализа в частотной и кепстральной областях [31,32,34,40], которые первоначально использовались в системах распознавания речи. Имеется относительно немного работ, описывающих применение параметров модуляционного спектра к задаче выявления патологии голосового аппарата. Например, в [19] описывается применение признаков, полученных в из модуляционного спектра, для построения систем детектирования и классификации патологий голосового аппарата. Использование модуляционных признаков в задаче детектирования патологии позволило получить такую же точность (~94%), как и использование кепстральных признаков. Однако в задаче классификации патологии, когда классификатору нужно было отличить одну патологию от другой, модуляционные признаки поз-

волили получить точность значительно большую, чем признаки на основе кепстра.

В настоящей работе анализировалось применение параметров модуляционного спектра, полученных при помощи схем, показанных на рисунках 5 и 6 для детектирования патологии голосового аппарата. Использовались три различные речевые базы данных: 1) база записей голосов (протяжный звук [а]) пациентов с БАС [37]; 2) база коротких речевых фраз пациентов с БАС [44]; 3) база записей голосов пациентов с заболеванием гортани [40].

6.1. Описание эксперимента. Эксперимент проводился в три этапа: 1) подготовка обучающей выборки; 2) ранжирование признаков; 3) обучение и тестирование классификатора с использованием различного числа признаков.

На первом этапе исходная база звуковых файлов переводилась в параметры модуляционного спектра (согласно схемам на рисунках 5 и 6). Поскольку сигнала анализировался сегментами длительностью 0,8 с, а средняя длительность анализируемых звуковых файлов превышала 3 с, то для каждого файла получалось несколько выходных значений модуляционного спектра. Каждое такое выходное значение $X(b, i)$ преобразовывалось в вектор $\mathbf{x} = [x_j]$, следующим образом:

$$X(b, i) \mapsto x_j, \quad j = (i - 1)B + b,$$

где $b = 1, \dots, B$ – номер критической полосы; $i = 1, \dots, I$ – номер модуляционной частоты; $j = 1, \dots, J$ – номер компоненты вектора \mathbf{x} ; $J = B \times I$ – общее число параметров модуляционного спектра. Каждому характеристическому вектору $\mathbf{x}^{(m)}$ ставилась в соответствие метка класса $y^{(m)}$ (равная «1», если в голосе есть патология и «-1» в противном случае). В результате анализа всей базы голосов формировалась обучающая выборка $\{\mathbf{x}^{(m)}, y^{(m)}\}$, $m = 1, \dots, M$, где M – общее число векторов в обучающей выборке.

На втором этапе для полученной обучающей выборки выполнялось ранжирование признаков (т.е. параметров модуляционного спектра) методом LASSO (описание данного метода приводится ниже). Результатом ранжирования является перестановка

$$\sigma(z) = j, \quad z, j \in \{1, \dots, J\},$$

которая переупорядочивает индексы параметров модуляционного спектра в порядке их значимости. Например, $\sigma(1)$ указывает индекс наиболее

значимого параметра, $\sigma(2)$ – следующий по значимости параметр и так далее.

На третьем этапе выполнялось обучение и тестирование классификатора методом перекрестной проверки по K блокам ($K = 4$). Причем разбиение на блоки выполнялось на уровне дикторов и, таким образом, тестовый и обучающий наборы содержали вектора, относящиеся к голосам разных дикторов. Процедура перекрестной проверки выполнялась 75 раз, для того, чтобы проверить работу классификатора с различным числом признаков (от 1 до 75). При этом на k -м запуске для классификатора выбирались признаки, имеющие ранг с 1 по k . По результатам перекрестных проверок строился график, который показывал, как изменяется точность классификатора при увеличении числа признаков.

Описанный эксперимент выполнялся для данных, полученных в результате анализа звуковых сигналов схемой на основе ДПФ и схемой на основе банка фильтров.

6.2. Отбор признаков и классификация. Как указывалось ранее, в результате анализа входного сигнала длительностью $T = 800$ мс при помощи схем, показанных на рисунках 5 и 6, получается число параметров модуляционного спектра, равное $B \times I = 20 \times 257 = 5140$. Очевидно, что не все они являются информативными для задачи выявления патологии голосового аппарата. Для отбора признаков разработан ряд методов таких, как, например, Relief [45], mRMR [46], LASSO [47], QoV [48]. В процессе проведения экспериментов были испробованы все указанные методы, однако результаты, полученные с использованием метода LASSO, значительно превосходили все прочие, поэтому в статье описаны только они. Отбор подходящего подмножества признаков не только позволяет ускорить процесс обучения классификатора, поскольку уменьшается число параметров обучения, но и помогает избежать переобучения [45]. Кроме того, анализируя отобранные признаки, можно понять какие именно параметры наиболее важны для правильной классификации.

Метод LASSO (англ. *least absolute shrinkage and selection operator*), используемый в данной работе, основан на решении задачи линейной регрессии [47]:

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left(\frac{1}{2} \sum_{m=1}^M (y^{(m)} - \beta_0 - \sum_{j=1}^p x_j^{(m)} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right), \quad (10)$$

где M – число примеров в обучающей выборке; y_i – метка класса i -го примера («1» – голос с патологией, «-1» – голос без патологии); $x_j^{(m)}$ – j -й признак m -го примера; β_j – j -й коэффициент линейной модели; λ – параметр регуляризации. Решение (10) при достаточно большом λ приводит к тому, что часть коэффициентов (предикторов) β становятся в точности нулевыми. Поэтому, решая (10) для ряда возрастающих значений параметра регуляризации λ и фиксируя порядок, в котором модель «отбрасывает» признаки, можно ранжировать их по значимости (первыми отбрасываются наименее значимые признаки).

Для решения задачи бинарной классификации использовался метод на основе линейного дискриминантного анализа (ЛДА). В методе ЛДА строится функция принадлежности к классу:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{j=1}^p w_j x_j + w_0 \right), \quad (11)$$

где w_j – весовые коэффициенты, $\mathbf{x} = [x_1 x_2 \dots x_p]$ – вектор признаков; w_0 – смещение; $\text{sign}(a)$ – функция возвращающая, знак числа a . Если $f(\mathbf{x}) = 1$, то вектор \mathbf{x} относят к классу голосов с патологией, а при $f(\mathbf{x}) = -1$ к классу здоровых голосов. Подробности реализации метода ЛДА можно найти в [47].

6.3. Оценка точности классификатора. Для оценки производительности классификатора использовался метод перекрестной проверки по K блокам (англ. *K-fold cross-validation*) [45]. Исходная обучающая выборка перемешивалась случайным образом и разбивалась на K блоков. Далее выполнялось обучение классификатора, причем один из блоков выступал как тестовый набор, а оставшиеся $K - 1$ блоков составляли обучающий набор. Эта процедура повторялась K раз так, чтобы каждый блок один раз выступил в роли тестового набора. Метки, присвоенные классификаторами, для тестовых наборов сохранялись, и выполнялась оценка точности работы классификатора:

$$\text{Acc} = \frac{TP + TN}{TP + FP + FN + TN}, \quad (12)$$

где TP , TN , FP , FN – количество истинно положительных, истинно отрицательных, ложно положительных и ложно отрицательных результатов классификации соответственно. Процедура перекрестной проверки повторялась 40 раз, после чего вычислялись выборочное среднее и среднеквадратичное отклонение (СКО) параметра точности классификатора.

6.4. Результаты экспериментов. В первом эксперименте по детектированию патологии голосового аппарата использовалась база голосов из исследования [37], содержащая записи протяжного звука [a] 15 пациентов с БАС и 39 здоровых людей. В [37] использовались голоса пациентов с БАС, чья речь обладала хорошей разборчивостью, но имела признаки патологии, слышимые для врача-специалиста, но трудно различимые для неподготовленного слушателя. В результате перевода звуковой базы в модуляционные параметры получена выборка из 877 обучающих векторов (650 – группа здоровых, 227 – группа с БАС). На рисунке 10 приводится график, показывающий, как изменялась точность классификации при увеличении числа признаков в характеристическом векторе. На графике указано среднее значение точности, а также ее СКО.

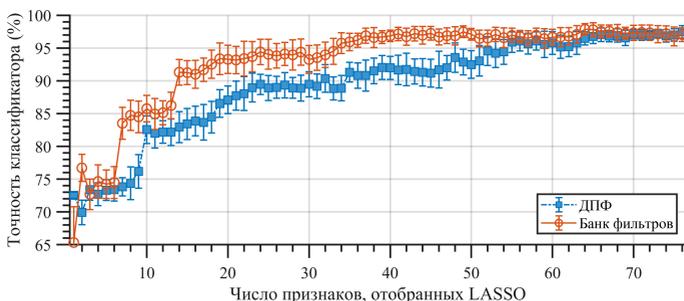


Рис. 10. Точность классификатора, при использовании параметров модуляционного спектра, полученных в схемах на основе ДПФ и банка фильтров

Анализируя график на рисунке 10, можно заметить, что признаки, полученные в схеме на основе банка фильтров, обеспечивают лучшую точность. При числе признаков $p = 55$ и более классификаторы показывают приблизительно равную точность.

В таблице 1 приведена точность классификации в зависимости от схемы получения признаков, при ограничении на размерность характеристического вектора. Данное ограничение обусловлено тем, что при построении классификатора всегда желательно иметь меньшее число признаков, но обеспечивающих лучшую точность.

Следует отметить, что эффективность модуляционных параметров, полученных в схеме с банком фильтров, выражается в том, что с увеличением числа признаков точность классификации возрастает быстрее, чем в случае использования модуляционных параметров, оцененных в схеме на основе ДПФ. Так, например, для достижения точности 93,4% классификатору требуется 19 модуляционных параметров, полученных в

Таблица 1. Точность классификации в зависимости от используемого числа признаков (p) для базы голосов [37]. Для каждого интервала в скобках указано число признаков, при котором достигается наибольшая точность

	Точность классификации, %						
	$p \leq 10$	$p \leq 20$	$p \leq 30$	$p \leq 40$	$p \leq 50$	$p \leq 60$	$p \leq 70$
Схема на основе ДПФ, %	82,6 (10)	87,1 (20)	89,6 (24)	92,0 (39)	93,5 (48)	96,1 (58)	97,0 (66)
Схема на основе банка фильтров, %	85,7 (10)	93,4 (19)	94,4 (24)	96,9 (37)	97,5 (49)	97,5 (49)	97,9 (64)

схеме на основе банка фильтров. Аналогичная точность классификации достигается лишь при использовании 48 модуляционных параметров из схемы на основе ДПФ.

Второй эксперимент заключался в построении классификаторов для решения задачи детектирования патологии голосового аппарата на основе анализа речевых сигналов. Использовалась база речевых сигналов из работы [44], которая содержала 15 записей речи пациентов с БАС и 15 записей речи здоровых людей. Каждая запись представляла собой последовательность из трех слов «раз, два, три». Речь пациентов с БАС обладала хорошей разборчивостью, но имела признаки заболевания, слышимые для врача-специалиста, но слабо ощутимыми для неподготовленного слушателя. В результате перевода речевой базы в модуляционные параметры получена выборка из 221 обучающего вектора (106 – группа здоровых, 115 – группа с БАС). На рисунке 11 показан график изменения точности классификации при увеличении числа модуляционных параметров. В таблице 2 приведена точность классификации при ограничении числа признаков, полученных в схемах на основе ДПФ и банка фильтров.

Анализируя график на рисунке 11, можно заметить, что признаки, полученные в схеме на основе ДПФ, обеспечивают лучшую точность. Тем не менее, при числе признаков $p \geq 67$ классификаторы имеют приблизительно равную точность.

В третьем эксперименте строился классификатор для выявления патологии гортани на основе анализа голосового сигнала (протяжного произнесения звука [а]). Использовалась база голосов из работы [40], содержащая 30 записей голосов с различными видами патологии гортани и 30 записей голосов здоровых людей. В результате перевода звуковой базы в модуляционные параметры получена выборка из 280 обучающих векторов (185 – группа здоровых, 95 – группа с патологиями гортани).

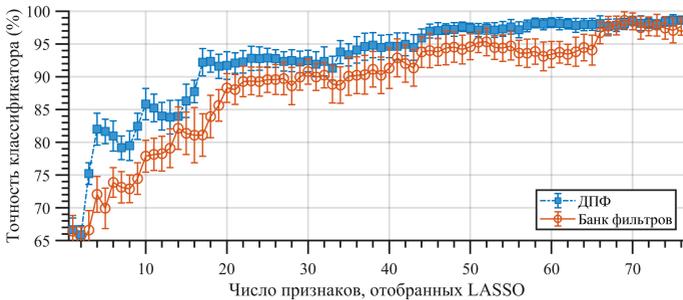


Рис. 11. Точность классификаторов, при использовании параметров модуляционного спектра, полученных в схемах основе ДПФ и банка фильтров

Таблица 2. Точность классификации в зависимости от используемого числа признаков (p) для речевой базы [44]. Для каждого интервала в скобках указано число признаков, при котором достигается наибольшая точность

	Точность классификации, %						
	$p \leq 10$	$p \leq 20$	$p \leq 30$	$p \leq 40$	$p \leq 50$	$p \leq 60$	$p \leq 70$
Схема на основе ДПФ, %	85,8 (10)	92,3 (18)	92,9 (23)	94,8 (39)	97,6 (49)	98,2 (60)	98,5 (70)
Схема на основе банка фильтров, %	77,9 (10)	88,3 (20)	90,8 (30)	91,3 (40)	94,6 (50)	95,3 (52)	98,6 (69)

На рисунке 12 показан график точности полученных классификаторов. Точность классификации в зависимости от числа признаков и схемы их извлечения приведена в таблице 3.

Анализируя график на рисунке 12, можно заметить, что признаки, полученные в схеме на основе банка фильтров, во всех случаях обеспечивают большую точность. Однако при числе признаков $p \geq 58$ классификаторы имеют приблизительно равную точность.

7. Обсуждение результатов. Важным практическим результатом проведенных экспериментов является то, что достигаемая в них точность классификации, выполненная на основе параметров модуляционного спектра, выше точности классификаторов, описанных в исходных работах, откуда взяты базы голосов. Это говорит о том, что параметры модуляционного спектра являются важными признаками для задач выявления патологий голосового аппарата. Полученные результаты согласуются с предшествующими работами в данной области. Так, в [19] отмечалось,

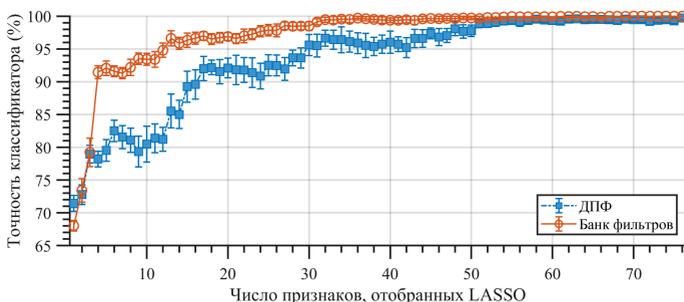


Рис. 12. Точность классификаторов, при использовании параметров модуляционного спектра, полученных при помощи ДПФ и на основе банка фильтров. Использована база голосов для детектирования заболеваний гортани [40]

Таблица 3. Точность классификации в зависимости от используемого числа признаков (p) для базы голосов [40]. Для каждого интервала в скобках указано число признаков, при котором достигается наибольшая точность

	Точность классификации, %						
	$p \leq 10$	$p \leq 20$	$p \leq 30$	$p \leq 40$	$p \leq 50$	$p \leq 60$	$p \leq 70$
Схема на основе ДПФ, %	82,3 (6)	92,2 (18)	95,6 (30)	96,6 (32)	98,0 (48)	99,6 (58)	99,6 (63)
Схема на основе банка фильтров, %	93,5 (9)	97,0 (17)	98,5 (27)	99,7 (36)	99,7 (49)	99,9 (55)	100,0 (70)

что параметры модуляционного спектра хорошо подходят для оценки и классификации патологий гортани. Также в работе [35] отмечалось, что наиболее информативными признаками для выявления заболевания БАС по голосу (как мужчин, так и женщин) являлись признаки, полученные на основе RASTA анализа.

Сравнивая две рассмотренных в работе схемы извлечения параметров модуляционного спектра, нельзя сделать вывод, что какая-то из них однозначно лучше или хуже. Так, в случае детектирования болезни БАС по речи (эксперимент 2) параметры, полученные в схеме на основе ДПФ, при меньшем числе признаков позволяли более точно выполнять классификацию. Отсюда следует вывод, что использование схемы на основе ДПФ предпочтительнее в случае, когда анализируемый сигнал является речью. Возможно, что в случае речевого сигнала параметры,

получаемые в схеме на основе банка фильтров, имеют большую вариацию и поэтому хуже подходят для детектирования патологии. Однако в случае, когда анализируемым сигналом являлось протяжное произнесение звука [a] (эксперименты 1 и 3) признаки, полученные в схеме на основе банка фильтров, позволили выполнять классификацию более эффективно (большая точность, при меньшем числе признаков). Таким образом, можно сделать вывод, что схема извлечения модуляционных параметров на основе банка фильтров лучше подходит для описания процесса фонации в задаче детектирования патологии голосового аппарата, чем схема на основе ДПФ.

В таблице 4 приводятся результаты сравнения точности классификации, достигнутые в исходных работах, откуда были взяты речевые базы с точностью, полученной в данной работе. В качестве точности, достигнутой в данной работе, бралась максимальная точность, полученная при использовании параметров из схемы на основе банка фильтров (правый столбец последней строки в таблицах 1–3). Низкая точность классификации в оригинальных статьях объясняется тем, что в них использовались относительно простые акустические признаки, которые не позволяли точно разделить голоса на норму и патологию.

Таблица 4. Сравнение точности классификации

Речевая база данных	Оригинальная работа	Данная работа
База голосов пациентов с БАС [37]	91,6± 2,3 %	97,9± 1,5 %
Речевая база пациентов с БАС [44]	84,8± 3,5 %	98,6± 1,2 %
База голосов пациентов с патологиями гортани [40]	83,3± 1,2 %	100,0± 0,1 %

8. Заключение. Исследован вопрос анализа модуляционного спектра звукового сигнала на основе неравнополосной обработки. Рассмотрены две схемы получения параметров модуляционного спектра, отличающиеся в части анализа акустических частот. В первом случае использовалось ДПФ с объединением каналов, а во втором – неравнополосный ДПФ-модулированной банк фильтров. Показаны особенности анализа сигнала обеими предложенными схемами. В экспериментальном исследовании выполнено применение разработанных схем анализа для извлечения параметров модуляционного спектра в задаче детектирования патологии голосового аппарата. Для эксперимента использовались три различные речевые базы. Оказалось, что параметры модуляционного спектра, полученные в схеме на основе ДПФ, более предпочтительны для выявления патологии по речевому сигналу. Напротив, параметры, полученные в схеме

на основе банка фильтров, более эффективны в задаче детектирования патологии на основе анализа голосового сигнала (протяжного произнесения звука [а]). Во всех случаях полученные классификаторы имели высокую точность >97%, что свидетельствует о несомненной практической значимости параметров модуляционного спектра для построения медицинских систем анализа и оценки патологических изменений в голосе.

Авторы благодарят главного научного сотрудника неврологического отдела РНПЦ неврологии и нейрохирургии (г. Минск), д.м.н. Ю.Н. Рушкевич, а также заведующую фониатрическим отделением РНПЦ оториноларингологии (г. Минск) Н.С. Конойко за доступ к базам речевых сигналов.

Литература

1. *Atlas L., Shamma S.A.* Joint acoustic and modulation frequency // *EURASIP Journal on Applied Signal Processing*. 2003. vol. 7. pp. 668–675.
2. *Ding N. et al.* Temporal modulations in speech and music // *Neuroscience & Biobehavioral Reviews*. 2017. vol. 81. pp. 181–187.
3. *Elhilali M.* Modulation Representations for Speech and Music // *Timbre: Acoustics, Perception, and Cognition*. 2019. pp. 335–359.
4. *Arai T. et al.* Syllable intelligibility for temporally filtered LPC cepstral trajectories // *The Journal of the acoustical society of America*. 1999. vol. 105. no. 5. pp. 2783–2791.
5. *Greenberg S., Kingsbury B.E.D.* The modulation spectrogram: In pursuit of an invariant representation of speech // *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 1997. vol. 3. pp. 1647–1650.
6. *Tyagi V. et al.* Mel-cepstrum modulation spectrum (MCMS) features for robust ASR // *2003 IEEE Workshop on automatic speech recognition and understanding*. 2003. pp. 399–404.
7. *Рылов А.С.* Анализ речи в распознающих системах // Бестпринт. 2003. 264 с.
8. *Yan B. C. et al.* Exploring Low-Dimensional Structures of Modulation Spectra for Robust Speech Recognition // *INTERSPEECH*. 2017. pp. 3637–3641.
9. *Moritz N. et al.* Multi-channel speech enhancement and amplitude modulation analysis for noise robust automatic speech recognition // *Computer Speech & Language*. 2017. vol. 46. pp. 558–573.
10. *Hermansky H., Morgan N.* RASTA processing of speech // *IEEE Transactions on speech and audio processing*. 1994. vol. 2. no. 4. pp. 578–589.
11. *Nidhyananthan S. S., Kumari R. S. S., Selvi T. S.* Noise robust speaker identification using RASTA–MFCC feature with quadrilateral filter bank structure // *Wireless Personal Communications*. 2016. vol. 91. no. 3. pp. 1321–1333.
12. *Петровский А.А., Башун Я.М.* Пре-процессор повышения качества зашумленной и ревербирующей речи для систем улитковой имплантации // *Цифровая обработка сигналов*. 2002. № 2. С. 48–61.
13. *Mowlae P., Blass M., Kleijn W. B.* New results in modulation-domain single-channel speech enhancement // *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2017. vol. 25. no. 11. pp. 2125–2137.
14. *Sukittanon S., Atlas L.E., Pitton J.W.* Modulation-scale analysis for content identification // *IEEE Transactions on signal processing*. 2004. vol. 52. no. 10. pp. 3023–3035.
15. *Avila A. R. et al.* Feature pooling of modulation spectrum features for improved speech emotion recognition in the wild // *IEEE Transactions on Affective Computing*. 2018.

16. *Dau T.* Predicting speech intelligibility based on the modulation spectrum and modulation frequency selectivity // *The Journal of the Acoustical Society of America.* 2018. vol. 143. no. 3. pp. 1916–1916.
17. *Smith C. L. et al.* Extracting dynamic parameters from speech movement data // *The Journal of the acoustical society of America.* 1993. vol. 93. no. 3. pp. 1580–1588.
18. *Giraud A.L. et al.* Representation of the temporal envelope of sounds in the human brain // *Journal of neurophysiology.* 2000. vol. 84. no.3. pp. 1588–1598.
19. *Markaki M., Stylianou Y.* Voice pathology detection and discrimination based on modulation spectral features // *IEEE Transactions on audio, speech, and language processing.* 2011. vol. 19. no. 7. pp. 1938–1948.
20. *Moro-Velázquez L. et al.* Modulation spectra morphological parameters: a new method to assess voice pathologies according to the GRBAS Scale // *BioMed Research International.* 2015. vol. 2015. 13 p.
21. *Benba A. et al.* Using RASTA-PLP for discriminating between different neurological diseases // 2016 International Conference on Electrical and Information Technologies (ICEIT). pp. 406–409.
22. *Moro-Velázquez L., Gómez-García J.A., Godino-Llorente J.I.* Voice pathology detection using modulation spectrum-optimized metrics // *Frontiers in bioengineering and biotechnology.* 2016. vol. 4. pp. 1–13.
23. *Петровский А.А.* Анализаторы речевых и звуковых сигналов: Методы, алгоритмы и практика // Бестпринт. 2009. 456 с.
24. *Kappelan M., Strauss B., Vary P.* Flexible nonuniform filter banks using allpass transformation of multiple order // *Proceedings of 8th european signal processing conference (EUSIPCO).* 1996. pp. 1–4.
25. *Moritz, N., Anemuller J., Kollmeier B.* An auditory inspired amplitude modulation filter bank for robust feature extraction in automatic speech recognition // *IEEE Transactions on audio, speech, and language processing.* 2015. vol. 23. no. 11. pp. 1926–1937.
26. *Оппенгейм А., Джонсон Д., Штейнлиц К.* Вычисление спектров с неравномерным разрешением посредством быстрого преобразования Фурье // *ТИИЭР.* 1971. Т. 59. № 2. С. 138–139.
27. *Galijasevic E., Klierer J.* Design of allpass-based non-uniform oversampled DFT filter banks // *Proceedings of IEEE International conference on acoustics, speech, and signal Processing (ICASSP).* 2002. vol. 2. pp. 1181–1184.
28. *Smith J.O., Abel J.S.* Bark and ERB bilinear transforms // *IEEE Transactions on speech and audio processing.* 1999. vol. 7. no. 6. pp. 697–708.
29. *Lin Y.P., Vaidynathan P.P.* A kaiser window approach for the design of prototype filters of cosine modulated filterbanks // *IEEE signal processing letters.* 1998. vol. 5. no 6. pp. 132–134.
30. *Gülzow T., Engelsberg A., Heute U.* Comparison of a discrete wavelet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement // *Signal processing.* 1998. vol. 64. no. 1. pp. 5–19.
31. *Картов А. А., Кайа Х., Салах А. А.* Актуальные задачи и достижения систем паралингвистического анализа речи // *Научно-технический вестник информационных технологий, механики и оптики.* 2016. Т. 16. №. 4. С. 581–591.
32. *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.
33. *Soumaya Z. et al.* A hybrid method for the diagnosis and classifying parkinson’s patients based on time–frequency domain properties and K-nearest neighbor // *Journal of Medical Signals and Sensors.* 2020. Т. 10. №. 1. pp. 60–66.

34. *Гвоздович А.Д., Рушкевич Ю.Н., Вашкевич М.И.* Детектирование бульбарных нарушений при боковом амиотрофическом склерозе на основе анализа речевого сигнала // Доклады Белорусского государственного университета информатики и радиоэлектроники. 2018. Т. 116. № 6. С. 52–58.
35. *Norel R. et al.* Detection of Amyotrophic Lateral Sclerosis (ALS) via Acoustic Analysis // Proceedings of Interspeech'2018. 2018. pp. 377–381.
36. *Куль Т. П. и др.* Методическое и аппаратно-программное обеспечение для регистрации и обработки речевых сигналов с целью диагностики неврологических заболеваний // Информатика. 2019. Т. 16. № 2. С. 27–39.
37. *Vashkevich M., Petrovsky A., Rushkevich Y.* Bulbar ALS detection based on analysis of voice perturbation and vibrato // Proceedings of international conference Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA'2019). 2019. pp. 267–272.
38. *Hlavníčka J. et al.* Characterizing vocal tremor in progressive neurological diseases via automated acoustic analyses // Clinical Neurophysiology. 2020.
39. *Кожанов А. Л. и др.* Функциональные результаты после открытых вертикальных резекций гортани // Вестник оториноларингологии. 2019. Т. 84. № 1. С. 46–50.
40. *Вашкевич М.И. и др.* Анализ акустических параметров голоса для выявления заболеваний гортани // Информатика. 2020. Т. 17. № 1. С. 78–86.
41. *Лысак А.П.* Акустический анализ голоса в норме и патологии // Речевые технологии. 2012. № 4. С. 68–82.
42. *Тычков А. Ю., Алимуратов А. К., Чураков П. П.* Метод адаптивной обработки сигналов для диагностики органов речевого аппарата // Измерительная техника. 2016. № 5. С. 26–29.
43. *Свистушкин В.М. и др.* Изменения акустических параметров голоса после хирургического лечения хронического тонзиллита // Folia Otorhinolaryngologiae et Pathologiae Respiratoriae. 2018. Т. 24. № 2. С. 46–55.
44. *Vashkevich M., Gvozдович А., Rushkevich Y.* Detection of bulbar dysfunction in ALS patients based on running speech test // Communications in Computer and Information Science. 2019. vol. 1055. pp. 192–204.
45. *Флах П.* Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных // М.: ДМК Пресс. 2015. 400 с.
46. *Tsanas A., Little M.A., McSharry P.E.* A methodology for the analysis of medical data // Handbook of Systems and Complexity in Health. 2013. pp. 113–125.
47. *Джеймс Г. и др.* Введение в статистическое обучение с примерами на языке R // М.: ДМК Пресс. 2016. 450 с.
48. *Liu R., Gillie D.F.* Feature selection using order statistics // Proceedings of international conference on Pattern Recognition and Information Processing (PRIP). 2011. pp. 195–199.

Вашкевич Максим Иосифович — канд. техн. наук., доцент; доцент кафедры электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники (БГУИР). Область научных интересов: методы и алгоритмы цифровой обработки звуковых и речевых сигналов, применение методов математики и информатики в сфере медицинской диагностики. Число научных публикаций — 75. vashkevich@bsuir.by; БГУИР, ул. П. Бровки 6, г. Минск, 220013, РБ; р.т. +37517-2-93-8478.

Азаров Илья Сергеевич — д-р техн. наук., доцент; заведующий кафедры электронных вычислительных средств БГУИР. Область научных интересов: цифровая обработка певческого голоса, системы конверсии голоса, распознавание речи, идентификация диктора, синтез речи по тексту. Число научных публикаций — 90. azarov@bsuir.by; БГУИР, ул. П. Бровки 6, г. Минск, 220013, РБ; р.т. +37517-2-93-8805.

M. VASHKEVICH, I. AZAROV
**VOICE PATHOLOGY DETECTION BASED ON ANALYSIS OF
MODULATION SPECTRUM IN CRITICAL BANDS**

Vashkevich M., Azarov I. Voice Pathology Detection based on Analysis of Modulation Spectrum in Critical Bands.

Abstract. The paper presents an approach to the analysis of the modulation spectrum of a voice signal, in which the primary acoustic analysis is performed in bands of unequal width. Nonuniform analysis corresponds to the psychoacoustic laws of human perception of sound information. In the context of the analysis of the modulation spectrum, the considered approach can significantly reduce the resulting number of parameters, which greatly simplifies the task of detecting pathological changes in the voice signal based on the analysis of the parameters of the modulation spectrum. For frequency decomposition of a signal into bands of unequal width, two methods are considered: 1) DFT with channel combination and 2) the use of a nonuniform filter bank. The first method is characterized by a fixed time window for the analysis of all frequency components, while in the second method the time-frequency analysis plan is consistent with the critical frequency scale of the barks. For each method, a practical signal analysis circuit has been developed and described. The paper presents the experimental data on the application of the developed schemes for the analysis of the modulation spectrum to the problem of detecting pathology in a speech signal. The parameters of the modulation spectrum acted as information signs for a classifier built on the basis of linear discriminant analysis. Three different voice bases were used in the experiment (in two cases, the pathology was neurological ALS disease (amyotrophic lateral sclerosis), and in the third case, diseases of the larynx). The parameters of the modulation spectrum obtained in the DFT-based scheme with channel combining turned out to be more preferable for classification with a small number of features, however, greater accuracy (with an increase in the number of features) made it possible to obtain the parameters obtained in the scheme based on an unequal filter bank. In all cases, the obtained classifiers were highly accurate (more than 97%). The obtained results show that the use of nonuniform time-frequency representation is preferable in the case when the analyzed signal is a sustained vowel phonation, since it provides higher accuracy of pathology detection using fewer modulation parameters.

Keywords: speech signal analysis, critical bands, modulation spectra, feature extraction, voice pathology detection

Vashkevich Maksim — Ph.D., Associate Professor, Associate Professor, Computer Engineering Department, Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: methods and algorithms of digital signal processing, application of mathematics and computer science methods in the medical diagnostics. The number of publications — 75. vashkevich@bsuir.by; BSUIR, P. Brovkiy str. 6, Minsk, 220013, Belarus; office phone +37517-2-93-8478.

Azarov Ilya — Ph.D., Dr.Sci., Associate Professor, Head of Department, Computer Engineering Department, Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: digital processing of singing voice, voice conversion systems, speech recognition, speaker identification, text to speech synthesis. The number of publications — 95. azarov@bsuir.by; 6, P.Brovkiy str. Minsk, 220013, Belarus; office phone: +37517293-23-40

References

1. Atlas L., Shamma S.A. Joint acoustic and modulation frequency. *EURASIP Journal on Applied Signal Processing*. 2003. vol. 7. pp. 668–675.
2. Ding N. et al. Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*. 2017. vol. 81. pp. 181–187.
3. Elhilali M. Modulation Representations for Speech and Music. *Timbre: Acoustics, Perception, and Cognition*. 2019. pp. 335–359.
4. Arai T. et al. Syllable intelligibility for temporally filtered LPC cepstral trajectories. *The Journal of the acoustical society of America*. 1999. vol. 105. no. 5. pp. 2783–2791.
5. Greenberg S., Kingsbury B.E.D. The modulation spectrogram: In pursuit of an invariant representation of speech. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 1997. vol. 3. pp. 1647–1650.
6. Tyagi V. et al. Mel-cepstrum modulation spectrum (MCMS) features for robust ASR. *2003 IEEE Workshop on automatic speech recognition and understanding*. 2003. pp. 399–404.
7. Rilov A.S. *Analiz rechi v raspoznajushhiih sistemah*[Speech analysis in recognition systems]. Minsk: Bestprint, 2003. 264 p. (in Russian)
8. Yan B.C. et al. Exploring Low-Dimensional Structures of Modulation Spectra for Robust Speech Recognition. *INTERSPEECH*. 2017. pp. 3637–3641.
9. Moritz N. et al. Multi-channel speech enhancement and amplitude modulation analysis for noise robust automatic speech recognition. *Computer Speech & Language*. 2017. vol. 46. pp. 558–573.
10. Hermansky H., Morgan N. RASTA processing of speech // *IEEE Transactions on speech and audio processing*. 1994. vol. 2. no. 4. pp. 578–589.
11. Nidhyanathan S. S., Kumari R. S. S., Selvi T. S. Noise robust speaker identification using RASTA–MFCC feature with quadrilateral filter bank structure. *Wireless Personal Communications*. 2016. vol. 91. no. 3. pp. 1321–1333.
12. Petrovsky A.A., Bashun Ya.M. [Pre-processor for improving the quality of noisy and reverberant speech for cochlear implant systems]. *Cifrovaja obrabotka signalov - Digital Signal Processing*. 2002. no. 2. pp. 48–61.
13. Mowlaee P., Blass M., Kleijn W. B. New results in modulation-domain single-channel speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2017. vol. 25. no. 11. pp. 2125–2137.
14. Sukittanon S., Atlas L.E., Pitton J.W. Modulation-scale analysis for content identification. *IEEE Transactions on signal processing*. 2004. vol. 52. no.10. pp. 3023–3035.
15. Avila A. R. et al. Feature pooling of modulation spectrum features for improved speech emotion recognition in the wild. *IEEE Transactions on Affective Computing*. – 2018.
16. Dau T. Predicting speech intelligibility based on the modulation spectrum and modulation frequency selectivity. *The Journal of the Acoustical Society of America*. 2018. vol. 143. no. 3. pp. 1916–1916.
17. Smith C. L. et al. Extracting dynamic parameters from speech movement data. *The Journal of the acoustical society of America*. 1993. vol. 93. no. 3. pp. 1580–1588.
18. Giraud A.L. et al. Representation of the temporal envelope of sounds in the human brain. *Journal of neurophysiology*. 2000. vol. 84. no.3. pp. 1588–1598.
19. Markaki M., Stylianou Y. Voice pathology detection and discrimination based on modulation spectral features. *IEEE Transactions on audio, speech, and language processing*. 2011. vol. 19. no. 7. pp. 1938–1948.
20. Moro-Velázquez L. et al. Modulation spectra morphological parameters: a new method to assess voice pathologies according to the GRBAS Scale. *BioMed Research International*. 2015. vol. 2015. 13 p.

21. Benba A. et al. Using RASTA-PLP for discriminating between different neurological diseases *Proceedings of International Conference on Electrical and Information Technologies (ICEIT)*. 2016. pp. 406-409.
22. Moro-Velázquez L., Gómez-García J.A., Godino-Llorente J.I. Voice pathology detection using modulation spectrum-optimized metrics. *Frontiers in bioengineering and biotechnology*. 2016. vol. 4. pp. 1–13.
23. Petrovsky A.A. *Analizatory rechevyh i zvukovyh signalov: metody, algoritmy i praktika* [Analyzers of speech and audio signals: methods, algorithms and practices]. Bestprint. 2009. 455 p. (In Russ.).
24. Kappelan M., Strauss B., Vary P. Flexible nonuniform filter banks using allpass transformation of multiple order. *Proceedings of 8th european signal processing conference (EUSIPCO)*. 1996. pp. 1-4.
25. Moritz N., Anemuller J., Kollmeier B. An auditory inspired amplitude modulation filter bank for robust feature extraction in automatic speech recognition. *IEEE Transactions on audio, speech, and language processing*. 2015. vol. 23. no. 11. pp. 1926-1937.
26. Oppenheim A., Johnson D., Steiglitz K. Computation of spectra with unequal resolution using the fast Fourier transform. *Proceedings of the IEEE*. 1971. vol. 59, no. 2. pp. 299–301. (Russ. ed.: Oppengejm A., Dzhonson D., SHtejglic K. *Vychislenie spektrov s neravnomernym razresheniem posredstvom bystrogo preobrazovaniya Fur'e*. TIIER. 1971. vol. 59. no. 2. pp. 138–139).
27. Galijasevic E., Kliewer J. Design of allpass-based non-uniform oversampled DFT filter banks. *Proceedings of IEEE International conference on acoustics, speech, and signal processing (ICASSP)*. 2002. pp. 1181–1184.
28. Smith J.O., Abel J.S. Bark and ERB bilinear transforms. *IEEE Transactions on speech and audio processing*. 1999. vol. 7. no. 6. pp. 697-708.
29. Lin Y.P., Vaidynathan P.P. A kaiser window approach for the design of prototype filters of cosine modulated filterbanks. *IEEE signal processing letters*. 1998. vol. 5. no 6. pp. 132-134.
30. Gülzow T., Engelsberg A., Heute U. Comparison of a discrete wavelet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement. *Signal processing*. 1998. vol. 64. no. 1. pp. 5-19.
31. Karpov A. A. et al. [State-of-the-Art Tasks and Achievements of Paralinguistic Speech Analysis Systems]. *Nauchno-texnicheskij vestnik informacionny'x texnologij, mexaniki i optiki – Journal Scientific and Technical Of Information Technologies, Mechanics and Optics*. 2016. vol. 104. no. 4. pp. 581–592. (In Russ.).
32. 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.
33. Soumaya Z. et al. A hybrid method for the diagnosis and classifying parkinson's patients based on time–frequency domain properties and K-nearest neighbor. *Journal of Medical Signals and Sensors*. 2020. T. 10. №. 1. pp. 60–66.
34. Gvozdoch A.D., Rushkevich Yu.N., Vashkevich M.I. [Detection of bulbar amyotrophic lateral sclerosis based on speech analysis]. *Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioelektroniki - Reports of the Belarusian State University of Informatics and Radioelectronics*. 2018. vol. 116. no 6. pp. 52–58. (In Russ.).
35. Norel R. et al. Detection of Amyotrophic Lateral Sclerosis (ALS) via Acoustic Analysis. *Proceedings of Interspeech'2018*. 2018. pp. 377–381.
36. Kul T. P. et al. [Methodical and hardware-software for recording and processing speech signals for diagnosis of neurological diseases]. *Informatika - Informatics*. 2019. vol. 16, no. 2, pp. 27–39. (In Russ.).

37. Vashkevich M., Petrovsky A., Rushkevich Y. Bulbar ALS detection based on analysis of voice perturbation and vibrato. *Proceedings of international conference Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. 2019. pp. 267–272.
38. Hlavnička J. et al. Characterizing vocal tremor in progressive neurological diseases via automated acoustic analyses // *Clinical Neurophysiology*. – 2020.
39. Kozhanov A. L. et al [The functional outcomes of open vertical resections of the larynx]. *Vestnik otorinolaringologii - Bulletin of Otorhinolaryngology*. 2019. vol. 84. no. 1. pp. 46–50. (In Russ.).
40. Vashkevich M. et al. [Analysis of acoustic voice parameters for larynx pathology detection]. *Informatika - Informatics*. 2020. vol.17. no 1. pp. 78–86. (In Russ.).
41. Lysak A. P. [Acoustic voice analysis in normal condition and pathologies]. *Rechevytekhologii - Speech Technologies*. 2012. no. 4. pp. 68–82.
42. Tychkov A. Yu., Alimuradov A. K., Churakov P. P. [Adaptive signal processing method for the diagnosis of organs of the speech apparatus] *Izmeritel'naya tekhnika. - Measurement Techniques* 2016. no. 5. pp. 26–29.
43. Svistushkin V.M. et al [Changes in the acoustic parameters of the voice after the surgical treatment of the chronic tonsillitis] // *Folia otorhinolaryngologiae et pathologiae respiratoriae*. 2018. vol. 24. no 2. pp. 46-55. (In Russ.).
44. Vashkevich M., Gvozdovich A., Rushkevich Y. Detection of bulbar dysfunction in ALS patients based on running speech test. *Communications in Computer and Information Science*. 2019. vol. 1055. pp. 192–204.
45. Flach P. *Machine learning: The art and science of algorithms that make sense of data*. Cambridge University Press. 2012. 409 p. (Russ. ed.: Flah P. *Mashinnoe obuchenie. Nauka i iskusstvo postroeniya algoritmov, kotorye izvlekayut znaniya iz dannyh*. M.: DMK Press. 2015. 400 p.).
46. Tsanas A., Little M.A., McSharry P.E. A methodology for the analysis of medical data // *Handbook of Systems and Complexity in Health*. 2013. pp. 113–125.
47. James G. et al. *An introduction to statistical learning with applications in R*. – New York : Springer, 2013. p. 426. (Russ. ed.: Dzhajms G. et al. *Vvedenie v statisticheskoe obuchenie s primerami na yazyke R*. M.: DMK Press. 2016. 450 p.).
48. Liu R., Gillie D.F. Feature selection using order statistics *Proceedings of international conference on Pattern Recognition and Information Processing (PRIP)*. 2011. pp. 195-199.