

ISSN 1816-0301 (Print)  
ISSN 2617-6963 (Online)

**ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ**  
*SIGNAL, IMAGE AND SPEECH PROCESSING*

УДК 616-71 + 612.78  
<https://doi.org/10.37661/1816-0301-2020-17-1-78-86>

Поступила в редакцию 16.10.2019  
Received 16.10.2019

Принята к публикации 12.12.2019  
Accepted 12.12.2019

**Анализ акустических параметров голоса для выявления заболеваний гортани**

М. И. Вашкевич<sup>1✉</sup>, А. А. Бурак<sup>1</sup>, Н. С. Конойко<sup>2</sup>, В. С. Долдова<sup>2</sup>

<sup>1</sup>Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь  
✉E-mail: vashkevich@bsuir.by

<sup>2</sup>Республиканский научно-практический центр оториноларингологии, Минск, Беларусь

**Аннотация.** Приведены результаты анализа двух способов описания голосового сигнала для решения задачи выявления заболеваний гортани. Сравнивались параметры голоса, определяемые клинической системой lingWaves, и параметры, получаемые в результате мел-частотного кепстрального анализа голоса. Для определения пригодности данных параметров при решении задачи выявления заболеваний гортани на их основе строился классификатор с использованием вероятностной модели – логистической регрессии. Для обучения классификатора была записана база голосов 60 человек, 30 из которых составляли контрольную группу, а другие 30 имели различные заболевания гортани (узелки голосовых складок, паралич гортани или функциональную дисфонию). Показано, что точность классификатора на основе мел-частотных кепстральных параметров (83,8 %) выше, чем точность классификатора на основе параметров, полученных в системе lingWaves (60,4 %).

**Ключевые слова:** анализ голоса, акустические параметры голоса, кепстральный анализ, детектирование патологии в голосе, логистическая регрессия

**Для цитирования.** Анализ акустических параметров голоса для выявления заболеваний гортани / М. И. Вашкевич [и др.] // Информатика. – 2020. – Т. 17, № 1. – С. 78–86. <https://doi.org/10.37661/1816-0301-2020-17-1-78-86>

**Analysis of acoustic voice parameters for larynx pathology detection**

Maxim I. Vashkevich<sup>1✉</sup>, Anton A. Burak<sup>1</sup>, Natallia S. Kanoika<sup>2</sup>, Valeria S. Daldova<sup>2</sup>

<sup>1</sup>Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus  
✉E-mail: vashkevich@bsuir.by

<sup>2</sup>Republican Scientific and Practical Center of Otorhinolaryngology, Minsk, Belarus

**Abstract.** The comparative study of two types of voice signal representation for larynx pathology detection is presented. Parameters obtained in clinical system lingWaves compared to parameters obtained by mel-frequency cepstral analysis. The classifier based on the probabilistic model (logistic regression) was designed to determine the suitability of given parameters for the larynx pathology detection problem. To train the classifier, the base of voice samples of 60 persons was recorded, 30 of which constitute the control group, and the other 30 had various diseases of the larynx (nodules of the vocal folds, laryngeal paralysis, or functional dysphonia). The results show that the classifier based on mel-frequency cepstral parameters (83,8 %) higher than the classifier based on parameters obtained in lingWaves (60,4 %).

**Keywords:** voice analysis, acoustic voice parameters, cepstral analysis, voice pathology detection, logistic regression

**For citation.** Vashkevich M. I., Burak A. A., Kanoika N. S., Daldova V. S. Analysis of acoustic voice parameters for larynx pathology detection. *Informatics*, 2020, vol. 17, no. 1, pp. 78–86 (in Russian). <https://doi.org/10.37661/1816-0301-2020-17-1-78-86>

**Введение.** Человеческий голос, и как акустический феномен, и как анатомо-физиологическое действие, а также по своему социальному значению, – явление в своем роде уникальное. Голос служит не только средством передачи информации и общения между людьми, но и своеобразным орудием производства у представителей ряда профессий: преподавателей, экскурсоводов, диспетчеров и др. Пациенты с нарушениями голоса испытывают физические и психологические затруднения в общении, формирующие чувство неполноценности или более глубокие психические осложнения [1].

Изменения в звучании голоса могут предшествовать структурным изменениям в ЛОР-органах. Поэтому естественно полагать, что существенная информация о состоянии голосового аппарата пациента отражается в характере звучания его голоса. Извлечь эту информацию можно лишь путем соответствующего акустического анализа голоса [2]. Преимуществом такого анализа является то, что он позволяет получить объективное представление о качестве голоса [3].

В настоящее время в Республиканском научно-практическом центре оториноларингологии для акустического анализа голоса используется аппаратно-программный комплекс lingWaves ver. 2.5 (Германия). Самым трудным и важным этапом анализа является интерпретация данных акустического исследования, поскольку часто его результаты неоднозначны.

Цель работы – анализ акустических параметров, выделяемых программой lingWaves с точки зрения их пользы для задачи выявления заболеваний гортани. Для достижения поставленной цели ставится задача построения бинарного классификатора для определения голосов с патологией. Для сравнения предлагается использовать два набора признаков, описывающих голосовой сигнал: рассчитанных в системе lingWaves и полученных в результате мел-кепстрального представления голосового сигнала [4]. Эти наборы признаков предлагается применять для обучения детектора патологии в голосе на основе логистической регрессионной модели. Сравнительный анализ полученных детекторов позволит определить прогностическую ценность применяемых в них параметров.

**Анализ голоса в системе lingWaves.** При использовании системы lingWaves голос записывается с помощью микрофона, оснащенного встроенным измерителем уровня шума. Оптимальным режимом для записи голоса считается уровень шума не более 40 дБ. Микрофон следует располагать на расстоянии 30 см от рта исследуемого. В качестве тестового сигнала записывают протяжный звук /a/ с частотой дискретизации  $F_s$  44,1 кГц. Пример отчета, получаемого в результате анализа голоса в системе lingWaves, показан на рис. 1.

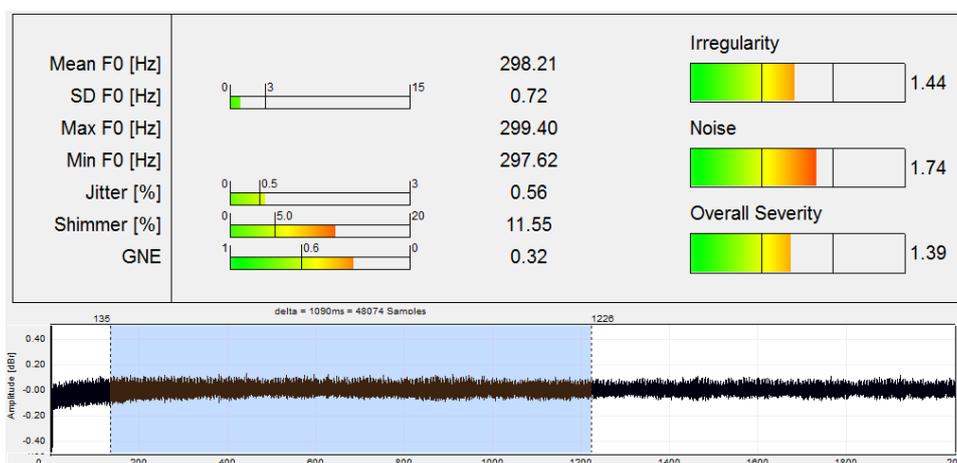


Рис. 1. Пример отчета lingWaves

Основными показателями, позволяющими оценить состояние голосовой функции, служат частота основного тона  $f_0$ ; степень частотной нестабильности вибрации голосовых складок, или джиттер (от англ. jitter – дрожание), и аналогичная мера амплитудной нестабильности – шиммер (от англ. shimmer – мерцание); соотношение шума и гармонических компонентов (harmonic/noise ratio, HNR); гортанно-шумовой коэффициент (glottal noise excitation ratio, GNE). Дополнительными обобщающими параметрами голоса являются нерегулярность (irregularity), шум (noise) и общая тяжесть дисфонии (overall severity). Приведенные параметры рассчитывались для исследуемой базы голосов и использовались для обучения детектора на основе логистической регрессии.

**Мел-кепстральное представление голосового сигнала.** В качестве альтернативного способа описания голосового сигнала в работе использовалось его мел-частотное кепстральное представление [4]. Расчет мел-частотных кепстральных коэффициентов (МЧКК) относится к методам кратковременного анализа голосового сигнала, которые предполагают разбиение сигнала на интервалы (кадры) анализа [5]. Как правило, в интервале от 10 до 30 мс голосовой сигнал можно считать стационарным. Рассматриваемый в настоящей статье анализ выполнялся на интервалах длительностью 20 мс, которые имели перекрытие 10 мс. Вычисление МЧКК производилось в частотной шкале мелов, которая учитывает специфику восприятия высоты звука человеческим ухом [6].

В результате кратковременного анализа формируется большой набор МЧКК, которые описывают локальную структуру сигнала. Для перехода к более глобальному представлению и уменьшению объема данных получающиеся МЧКК усредняются для формирования надсегментного вектора признаков.

В работе использовался следующий алгоритм мел-частотного кепстрального анализа:

*Шаг 1.* Коррекция спектра сигнала, заключающаяся в выравнивании энергий высокочастотной и низкочастотной составляющих при помощи фильтра, который имеет подъем амплитудно-частотной характеристики приблизительно 6 дБ на октаву:

$$s(n) = x(n) - 0,82 \cdot x(n-1).$$

*Шаг 2.* Разбиение сигнала на кадры длительностью  $N$  отсчетов (и с перекрытием в  $N/2$  отсчетов) и взвешивание их с помощью окна Хэмминга:

$$w(n) = 0,54 + 0,46 \cdot \cos(2\pi n / N).$$

Далее для всех кадров  $s_t(n)$ , где  $t$  – номер кадра, выполняются шаги 3–6.

*Шаг 3.* Расчет кратковременного спектра при помощи быстрого преобразования Фурье (БПФ) для каждого анализируемого кадра:

$$S_t(k) = \sum_{n=0}^{N-1} w(n) \cdot s_t(n) \cdot e^{-j2\pi kn / N}, \quad k = 0, 1, \dots, N-1.$$

*Шаг 4.* Расчет набора из  $M$  фильтров ( $m = 1, 2, \dots, M$ ) треугольной формы:

$$H_m(k) = \begin{cases} 0, & k < f(m-1); \\ \frac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \leq k \leq f(m); \\ \frac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) \leq k \leq f(m+1); \\ 0, & k > f(m+1), \end{cases}$$

которые используются для усреднения спектра вблизи центральных частот  $f(m)$ , расположенных равномерно в шкале мелов (рис. 2). В данном расчете параметр  $M = 19$  обозначает число критических полос в анализируемом частотном диапазоне.

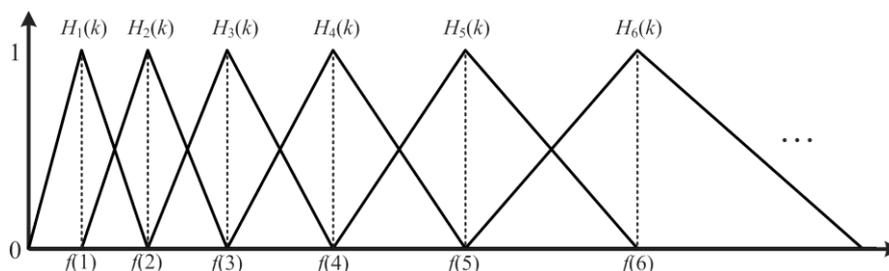


Рис. 2. Треугольные фильтры для вычисления мел-спектра

Если  $f_l$  и  $f_h$  – нижняя и верхняя границы частотного диапазона (Гц), который покрывает набор фильтров, то граничные частоты  $f(m)$ , равномерно расположенные в шкале мелов, рассчитываются по формуле [6]

$$f(m) = \frac{N}{F_s} B^{-1} \left( B(f_l) + m \frac{B(f_h) - B(f_l)}{M + 1} \right),$$

где функция  $B$  осуществляет переход от шкалы герцев к шкале мелов:

$$B(f) = 1127 \cdot \ln(1 + f / 700).$$

Значения  $f_l$  и  $f_h$  выбирались равными 50 и 5300 Гц соответственно. Указанный диапазон частот соответствует первым 19 критическим полосам слуха. Кроме того, в него попадает основная энергия анализируемого в данной работе гласного звука /a/.

*Шаг 5.* Применение набора фильтров, полученного на шаге 4, для расчета логарифма от энергии на выходе каждого фильтра:

$$Y_t(m) = \ln \left( \sum_{k=0}^{N-1} |S_t(k)|^2 \cdot H_m(k) \right), \quad 0 \leq m < M.$$

*Шаг 6.* Вычисление МЧКК при помощи дискретного косинусного преобразования:

$$c_t(k) = \sum_{m=0}^{M-1} Y_t(m) \cdot \cos(\pi k(m + 1/2) / M),$$

где  $k = 0, 1, \dots, M - 1$ .

В формировании вектора признаков участвовали  $P$  начальных коэффициентов ( $k = 1, 2, \dots, P$ ), использовалось значение  $P = 12$ .

*Шаг 7.* Вычисление конечных разностей МЧКК, которые применяются наряду с МЧКК в качестве информационных признаков:

$$\Delta c_t(k) = c_t(k) - c_{t-1}(k),$$

где  $c_t(k)$  – это  $k$ -й коэффициент, вычисленный для кадра с номером  $t$ .

*Шаг 8.* Выполнение расчета надсегментных признаков на основе МЧКК для получения характеристического вектора. Для этого последовательности  $c_t(k)$  и  $\Delta c_t(k)$  разбиваются на сегменты длительностью 0,8 с, по которым вычисляются их средние значения  $c^\mu(k)$ ,  $\Delta c^\mu(k)$  и среднеквадратические отклонения (СКО)  $c^\sigma(k)$  и  $\Delta c^\sigma(k)$ .

В результате для каждого сегмента был получен следующий вектор признаков:

$$\mathbf{x} = [c^\mu(1), \dots, c^\mu(P); \Delta c^\mu(1), \dots, \Delta c^\mu(P); c^\sigma(1), \dots, c^\sigma(P); \Delta c^\sigma(1), \dots, \Delta c^\sigma(P)]. \quad (1)$$

Описанный процесс анализа голосового сигнала проиллюстрирован на рис. 3.

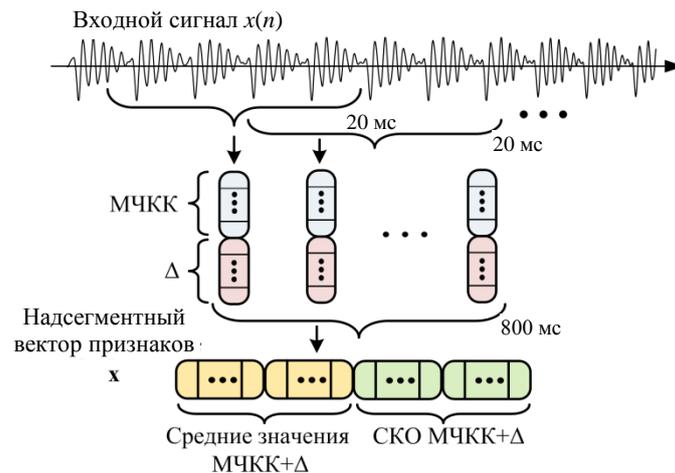


Рис. 3. Получение вектора надсегментных признаков на основе МФКК

**Детектор патологии в голосе на основе логистической регрессии.** Определение наличия патологии в голосе является задачей бинарной классификации. В данной работе классификатор строился на основе распространенной вероятностной дискриминантной модели – логистической регрессии [7].

Логистическая регрессия моделирует апостериорную вероятность принадлежности характеристического вектора  $\mathbf{x}$  к классу  $y=1$ :

$$P(y=1 | \mathbf{x}) = f(z), \quad z = \boldsymbol{\theta}^T \mathbf{x}, \quad f(z) = \frac{1}{1 + e^{-z}}, \quad (2)$$

где  $\mathbf{x}$  – вектор-столбец признаков (1), к которому добавлена компонента с единичным значением для перехода к однородным координатам;  $\boldsymbol{\theta}$  – вектор-столбец параметров регрессии. Для определения  $\boldsymbol{\theta}$  использовался метод, основанный на максимизации функции условного правдоподобия:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} CL(\boldsymbol{\theta}), \quad CL(\boldsymbol{\theta}) = \prod_{i=1}^m P(y_i | \mathbf{x}^{(i)}) = \prod_{i=1}^m \hat{p}(\mathbf{x}^{(i)})^{y_i} (1 - \hat{p}(\mathbf{x}^{(i)}))^{(1-y_i)},$$

где  $m$  – количество примеров в обучающей выборке.

Для упрощения формул лучше использовать логарифм функции правдоподобия:

$$LCL(\boldsymbol{\theta}) = \ln CL(\boldsymbol{\theta}) = \sum_{i=1}^m y^{(i)} \ln f(\boldsymbol{\theta}^T \mathbf{x}^{(i)}) + (1 - y^{(i)}) \ln(1 - f(\boldsymbol{\theta}^T \mathbf{x}^{(i)})). \quad (3)$$

Поиск оптимального значения параметра  $\boldsymbol{\theta}$ , минимизирующего (3), выполнялся посредством метода градиентного спуска [7]:

$$\boldsymbol{\theta}' = \boldsymbol{\theta} + \alpha \nabla \ln CL(\boldsymbol{\theta}) = \boldsymbol{\theta} + \alpha \sum_{i=1}^m (y^{(i)} - f(\boldsymbol{\theta}^T \mathbf{x}^{(i)})) \mathbf{x}^{(i)}, \quad \alpha > 0.$$

После обучения, в результате которого находится оптимальное значение параметра  $\boldsymbol{\theta}$ , процесс бинарной классификации выполняется следующим образом. Для поступающего вектора признаков  $\mathbf{x}$  вычисляется вероятность по формулам (2). Если полученное значение вероятности больше 0,5, то  $\mathbf{x}$  относился к классу «патология» ( $y=1$ ), в противном случае – к классу «норма» ( $y=0$ ).

**Отбор признаков и оценка производительности классификатора.** Отбор признаков для классификатора является важной задачей, от решения которой во многом зависит качество его работы. В исследовании для отбора признаков использовался метод LASSO (от англ. least absolute shrinkage and selection operator) [8]. Этот метод основан на решении задачи линейной регрессии со штрафной функцией, накладываемой на абсолютные значения коэффициентов

линейной модели. Задача решается для различных (возрастающих) значений параметра регуляризации  $\lambda$ , в результате чего веса при некоторых признаках (предикторах) линейной модели приближаются к нулю или становятся равными нулю. Фиксируя порядок, в котором модель отбрасывает признаки, можно ранжировать их по значимости, так как первыми исключаются наименее значимые признаки.

Для оценки производительности классификатора использовался метод перекрестной проверки по  $K$  блокам (англ. *K-fold cross-validation*) [7], 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 таким образом, чтобы каждый блок один раз выступил в роли тестового набора. Метки, присвоенные классификаторами, для тестовых наборов сохраняются, и по ним выполняется оценка производительности классификатора. В качестве основных характеристик классификатора вычислялись точность (accuracy), чувствительность (sensitivity) и специфичность (specificity):

$$\text{acc} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \text{sens} = \frac{TP}{TP + FN}, \quad \text{spec} = \frac{TN}{TN + FP},$$

где  $TP$ ,  $TN$ ,  $FP$ ,  $FN$  – количество истинно положительных и истинно отрицательных, ложно положительных и ложно отрицательных результатов классификации соответственно. Процедура перекрестной проверки повторялась 40 раз, после чего вычислялись выборочное среднее и выборочное СКО для оценок точности, чувствительности и специфичности.

**Результаты экспериментов.** В исследовании использовалась база голосов, записанная в фониатрическом отделении Республиканского научно-практического центра оториноларингологии. Всего были выполнены записи образцов голосов 60 человек, 30 из которых были здоровыми, а 30 имели различные заболевания гортани (18 – узелки голосовых складок, 6 – парез или паралич гортани, 8 – функциональную дисфонию). Все записанные звуковые файлы были обезличены, им присвоены буквенно-цифровые коды. Также записи звука не содержали никаких персональных данных (имя, возраст и пр.).

Для получения предварительного представления о статистических свойствах признаков, описывающих голосовой сигнал, выполнялся их корреляционный анализ. Для всех пар  $(x_{i,j}, y_i)$  вычислялся коэффициент корреляции  $R$ , где  $x_{i,j}$  –  $j$ -й признак  $i$ -го примера из соответствующего набора, а  $y_i$  – метка класса для  $i$ -го примера (голоса с патологией имели метку «1», а здоровые – «-1»). Результаты корреляционного анализа для первого набора признаков приведены в табл. 1.

Полученные результаты показывают, что лишь три признака (нерегулярность, шум и GNE) имеют статистически значимый ( $p < 0,05$ ) коэффициент корреляции (в табл. 1 они выделены жирным шрифтом). Для признаков, связанных с частотой основного тона, низкая корреляция ожидаема, поскольку они сильно зависят от пола и возраста. Однако низкая корреляция параметров «джиттер» и «шиммер» – неожиданное явление, так как считается, что при патологии голоса эти показатели имеют повышенные значения. Признак общей тяжести дисфонии не использовался, поскольку было установлено, что он является линейной комбинацией параметров «нерегулярность» и «шум»:

$$\text{общая тяжесть дисфонии} = 0,8 \cdot \text{нерегулярность} + 0,135 \cdot \text{шум}.$$

В табл. 2 представлены результаты корреляционного анализа для второго набора признаков: перечислены девять признаков, имеющих наибольшие коэффициенты корреляции. Приведенные результаты показывают, что признаки, полученные на основе анализа МЧКК, имеют гораздо большую корреляционную связь с наличием патологии в голосе, чем параметры, получаемые в lingWaves. Следует заметить, что выполненный корреляционный анализ направлен на выявление лишь линейной зависимости между параметрами голосового сигнала и меткой класса голоса. В данной работе не предпринимались попытки поиска более сложных нелинейных зависимостей.

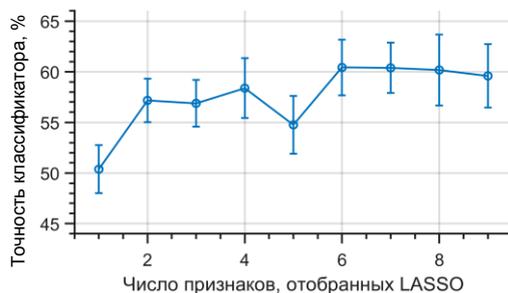
Таблица 1

| Признак        | Коэффициент корреляции $R$ (уровень значимости $p$ ) | Границы 95 % доверительного интервала |
|----------------|--|---------------------------------------|
| Нерегулярность | <b>0,27 (<math>p = 0,03</math>)</b>                  | [0,022; 0,49]                         |
| Шум            | <b>0,27 (<math>p = 0,04</math>)</b>                  | [0,015; 0,49]                         |
| GNE            | <b>-0,27 (<math>p = 0,04</math>)</b>                 | [-0,49; -0,014]                       |
| Джиттер        | 0,19 ( $p = 0,14$ )                                  | [-0,07; 0,42]                         |
| $\min f_o$     | 0,18 ( $p = 0,18$ )                                  | [-0,08; 0,41]                         |
| СКО $f_o$      | 0,17 ( $p = 0,20$ )                                  | [-0,09; 0,41]                         |
| Среднее $f_o$  | 0,14 ( $p = 0,28$ )                                  | [-0,12; 0,38]                         |
| $\max f_o$     | 0,12 ( $p = 0,37$ )                                  | [-0,14; 0,36]                         |
| Шиммер         | 0,06 ( $p = 0,65$ )                                  | [-0,20; 0,31]                         |

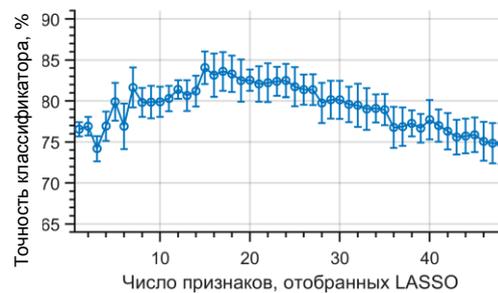
Таблица 2

| Признак               | Коэффициент корреляции $R$ (уровень значимости $p$ ) | Границы 95 % доверительного интервала |
|-----------------------|--|---------------------------------------|
| $\Delta c^\sigma(10)$ | <b>0,46 (<math>p = 1,8 \cdot 10^{-16}</math>)</b>    | [0,37; 0,55]                          |
| $c^\mu(7)$            | <b>-0,44 (<math>p = 4,6 \cdot 10^{-15}</math>)</b>   | [-0,53; -0,34]                        |
| $c^\mu(8)$            | <b>-0,42 (<math>p = 3,0 \cdot 10^{-13}</math>)</b>   | [-0,51; -0,31]                        |
| $\Delta c^\sigma(8)$  | <b>0,41 (<math>p = 3,4 \cdot 10^{-13}</math>)</b>    | [0,31; 0,51]                          |
| $\Delta c^\sigma(9)$  | <b>0,40 (<math>p = 1,6 \cdot 10^{-12}</math>)</b>    | [0,30; 0,50]                          |
| $\Delta c^\sigma(6)$  | <b>0,40 (<math>p = 2,4 \cdot 10^{-12}</math>)</b>    | [0,30; 0,49]                          |
| $\Delta c^\sigma(12)$ | <b>0,40 (<math>p = 3,6 \cdot 10^{-12}</math>)</b>    | [0,29; 0,49]                          |
| $\Delta c^\sigma(7)$  | <b>0,39 (<math>p = 9,3 \cdot 10^{-12}</math>)</b>    | [0,29; 0,49]                          |
| $\Delta c^\sigma(11)$ | <b>0,38 (<math>p = 3,4 \cdot 10^{-11}</math>)</b>    | [0,28; 0,48]                          |

На рис. 4 показаны результаты классификации в зависимости от набора и числа информационных признаков. Из приведенных графиков видно, что точность классификатора с признаками на основе МЧКК в целом выше, чем точность классификатора на основе признаков lingWaves. В табл. 3 приведены параметры классификаторов с максимальной точностью, полученных на основе набора признаков lingWaves (оптимальное число признаков – 6) и набора надсегментных МЧКК (оптимальное число признаков – 15). Интересно, что добавление признака «шиммер» (шестого по счету на рис. 4, а) значительно улучшает точность правильной классификации, хотя корреляционный анализ (см. табл. 1) не выявил его значимости. Это говорит о том, что данный параметр содержит важную дополнительную информацию в контексте других признаков.



а)



б)

Рис. 4. Производительность классификатора в зависимости от числа используемых признаков: а) набор признаков lingWaves; б) набор признаков на основе МЧКК

Таблица 3

Результаты классификации, %

| Набор признаков    | Подмножество признаков (в порядке убывания значимости)   | Точность       | Чувствительность | Специфичность  |
|--------------------|--|----------------|------------------|----------------|
| lingWaves          | Нерегулярность, шум, GNE, $\max f_o$ , СКО $f_o$ , шиммер  | $60,4 \pm 2,7$ | $55,3 \pm 4,7$   | $65,6 \pm 4,0$ |
| Надсегментные МЧКК | $\Delta c^\sigma(10)$ , $c^\mu(7)$ , $c^\mu(8)$ , $c^\mu(11)$ , $c^\mu(5)$ , $c^\mu(1)$ , $c^\sigma(8)$ , $c^\mu(2)$ , $c^\sigma(4)$ , $c^\sigma(11)$ , $c^\mu(3)$ , $c^\sigma(1)$ , $c^\sigma(2)$ , $c^\sigma(3)$ , $c^\sigma(6)$ | $83,8 \pm 1,2$ | $74,5 \pm 1,8$   | $88,7 \pm 1,7$ |

**Заключение.** Проведенное исследование показало, что параметры голоса, определяемые в системе lingWaves, в целом дают недостаточно информации, чтобы выявить наличие патологических изменений в голосе. Напротив, использование надсегментных признаков на основе МЧКК позволяет значительно повысить вероятность правильного детектирования патологии

голоса. Преимущество параметров, выделяемых системой lingWaves, заключается в том, что они имеют ясную интерпретацию и понятны для врача-фоноатра. Следовательно, существует насущная задача конструирования таких признаков патологии голоса, которые бы, с одной стороны, повышали вероятность выявления патологии, а с другой – имели вполне ясную интерпретацию для врача-специалиста.

Выполнен анализ акустических параметров голоса применительно к задаче выявления заболеваний гортани. Проанализированы наборы признаков, получаемых в системе lingWaves, которая используется во многих медицинских центрах СНГ и Европы, а также признаков, получаемых в результате кепстрального анализа голосового сигнала. Для сравнения данных наборов признаков на их основе выполнено обучение классификаторов на базе логистической регрессии. Точность классификатора на основе признаков lingWaves составила 60,4 %, а на основе признаков, полученных в результате кепстрального анализа, – 83,8 %. Несмотря на то что кепстральные признаки обладают большей информативностью, следует отметить и их недостаток – отсутствие ясной интерпретации, понятной для врача-фоноатра.

### Список использованных источников

1. Шиленкова, В. В. Дисфонии и голос / В. В. Шиленкова. – Ярославль : Аверс Плюс, 2018. – 256 с.
2. Ермолаев, В. Г. Руководство по фоноатрии / В. Г. Ермолаев, Н. Ф. Лебедева, В. П. Морозов. – Л. : Медицина, Ленинград. отд., 1970. – 271 с.
3. Коротченко, В. В. Акустический анализ голоса у детей в норме и при заболеваниях гортани : автореф. дис. ... канд. мед. наук / В. В. Коротченко ; Мин-во здравоохран. и соц. развития РФ, Ярославская гос. мед. академия. – М., 2012. – 24 с.
4. Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex / R. Fraile [et al.] // *Folia Phoniatrica Logopaedica*. – 2009. – Vol. 61. – P. 146–152.
5. Рылов, А. С. Анализ речи в распознающих системах / А. С. Рылов. – Минск : Бестринт, 2003. – 264 с.
6. Huang, X. *Spoken Language Processing: a Guide to Theory, Algorithm and System Development* / X. Huang, A. Acero, H.-W. Hon. – New Jersey : Prentice Hall, 2001. – 1009 p.
7. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : пер. с англ. / П. Флах. – М. : ДМК Пресс, 2015. – 400 с.
8. Tibshirani, R. Regression shrinkage and selection via the Lasso / R. Tibshirani // *J. of the Royal Statistical Society*. – 1996. – Vol. 58, no. 1. – P. 267–288.

### References

1. Shilenkova V. V. *Disfonii i golos. Dysphonies and Voice*. Yaroslavl, Avers Plus, 2018, 256 p. (in Russian).
2. Ermolayev V. G., Lebedeva N. F., Morozov V. P. *Rukovodstvo po foniatrii. Phoniatics Guide*. Leningrad, Medicina, Leningradskoe otdelenie, 1970, 271 p. (in Russian).
3. Korotchenko V. V. *Akusticheskij analiz golosa u detej v norme i pri zabolevaniyax gortani. Avtoref. dis. ... kand. med. nauk. Acoustic Analysis of Voice in Children Is Normal and with Diseases of the Larynx. Cand. med. sci. diss. abstr.* Ministerstvo zdravooхранenija i social'nogo razvitija Rossijskoj Federacii, Jaroslavskaja gosudarstvennaja medicinskaja akademija. Moscow, 2012, 24 p. (in Russian).
4. Fraile R., Sáenz-Lechón N., Godino-Llorente J., Osmá-Ruiz V., Fredouille C. Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex. *Folia Phoniatrica Logopaedica*, 2009, vol. 61, pp. 146–152.
5. Rilov A. S. *Analiz rechi v raspoznajushhix sistemah. Speech Analysis in Recognition Systems*. Minsk, Bestprint, 2003, 264 p. (in Russian).
6. Huang X., Acero A., Hon H.-W. *Spoken Language Processing: a Guide to Theory, Algorithm and System Development*. New Jersey, Prentice Hall, 2001, 1009 p.
7. Flach P. *Machine Learning. The Art and Science of Algorithms that Make Sense of Data*. Cambridge University Press, 2012, 409 p.
8. Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society*, 1996, vol. 58, no. 1, pp. 267–288.

**Информация об авторах**

*Вашкевич Максим Иосифович*, кандидат технических наук, доцент кафедры электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь.

E-mail: vashkevich@bsuir.by

*Бурак Антон Андреевич*, студент, Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь.

*Конюко Наталья Сергеевна*, заведующий фониатрическим отделением, Республиканский научно-практический центр оториноларингологии, Минск, Беларусь.

*Долдова Валерия Сергеевна*, врач-фониатр фониатрического отделения, Республиканский научно-практический центр оториноларингологии, Минск, Беларусь.

**Information about the authors**

*Maxim I. Vashkevich*, Cand. Sci. (Eng.), Associate Professor of Computer Engineering Department, Belarusian State University of Informatics and Radioelectronics Minsk, Belarus.

E-mail: vashkevich@bsuir.by

*Anton A. Burak*, Student, Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus.

*Natallia S. Kanoika*, Head of the Phoniatic Department, Republican Scientific and Practical Center of Otorhinolaryngology, Minsk, Belarus.

*Valeria S. Daldova*, Phoniatriest of the Phoniatic Department, Republican Scientific and Practical Center of Otorhinolaryngology, Minsk, Belarus.