

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ/ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DOI: <https://doi.org/10.60797/IRJ.2025.160s.1>

ИЗВЛЕЧЕНИЕ ФОРМАНТ И ДРУГИХ ХАРАКТЕРИСТИК И КЛАССИФИКАЦИЯ СОСТОЯНИЯ ЗДОРОВЬЯ ГОЛОСА МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Научная статья

Саламатов К.А.^{1,*}, Смирнов А.А.²

²ORCID : 0000-0002-7312-7219;

^{1,2}Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, Екатеринбург, Российская Федерация

* Корреспондирующий автор (kirillsalamatov20[at]gmail.com)

Аннотация

Данная статья посвящена разработке и анализу методов машинного обучения для автоматизированной диагностики нарушений голоса, что представляет собой актуальную задачу в современной медицине и логопедии. Голос играет ключевую роль в коммуникации, и его патологии могут существенно снижать качество жизни пациентов. Традиционные методы диагностики, включающие визуальный осмотр и эндоскопию, требуют участия специалистов и не всегда обеспечивают объективность. В связи с этим применение алгоритмов машинного обучения открывает новые возможности для повышения точности выявления речевых расстройств и их диагностики. В работе рассматриваются основные виды нарушений голоса, включая дисфонию, афонию, фонастению, брадилалию, тахилалию, заикание, дислалию и ринолалию. Для каждого из них анализируются этиология, симптоматика и существующие методы коррекции. Особое внимание уделяется акустическим параметрам голоса, таким как частота основного тона, jitter, shimmer и отношение сигнал/шум, которые могут служить маркерами патологий.

Ключевые слова: нарушения голоса, машинное обучение, акустический анализ, дисфония, афония, метод опорных векторов, градиентный бустинг, диагностика голосовых нарушений.

EXTRACTION OF FORMANTS AND OTHER CHARACTERISTICS AND CLASSIFICATION OF VOICE HEALTH STATUS BY MACHINE LEARNING METHODS

Research article

Salamatov K.A.^{1,*}, Smirnov A.A.²

²ORCID : 0000-0002-7312-7219;

^{1,2}Ural Federal University named after the first President of Russia B. N. Yeltsin, Ekaterinburg, Russian Federation

* Corresponding author (kirillsalamatov20[at]gmail.com)

Abstract

This article is devoted to the development and analysis of machine learning methods for the automated diagnosis of voice disorders, which is an urgent task in modern medicine and speech therapy. Voice plays a key role in communication, and its pathologies can significantly reduce the quality of life of patients. Traditional diagnostic methods, including visual examination and endoscopy, require the participation of specialists and do not always ensure objectivity. In this regard, the use of machine learning algorithms opens up new opportunities for improving the accuracy of speech disorders detection and diagnosis. The paper examines the main types of voice disorders, including dysphonia, aphonia, phonasthenia, bradylalia, tachylalia, stuttering, dyslalia and rhinolalia. The etiology, symptoms, and existing correction methods are analyzed for each of them. Special attention is paid to acoustic parameters of the voice, such as pitch frequency, jitter, shimmer, and signal-to-noise ratio, which can serve as markers of pathologies.

Keywords: voice disorders, machine learning, acoustic analysis, dysphonia, aphonia, support vector machine, gradient boosting, diagnostics of voice disorders.

Введение

Голос является одним из основных инструментов коммуникации человека, и его нарушения могут значительно повлиять на качество жизни, социальную адаптацию и профессиональную деятельность. Традиционные методы диагностики [1] нарушений голоса, такие как визуальная оценка и эндоскопия гортани, требуют участия опытных специалистов и могут быть субъективными. Однако с развитием технологий машинного обучения [2], [3], [4] появились новые возможности для автоматизации процесса диагностики, что позволяет повысить точность и объективность оценки.

Цель данной работы — исследование методов машинного обучения, применимых для выявления нарушений голоса. В рамках исследования были поставлены следующие задачи: рассмотреть основные виды нарушений голоса, выявить ключевые параметры голоса, которые могут быть использованы для анализа, разработать методику эксперимента и исследовать эффективность различных алгоритмов машинного обучения для классификации нарушений голоса.

Виды нарушений голоса

Нарушения голоса могут быть вызваны различными причинами, включая механические повреждения голосового аппарата, воспалительные процессы, гормональные изменения, а также психосоматические заболевания. В зависимости от характера и степени выраженности, нарушения голоса делятся на несколько видов [5]:

1. Дисфония — это нарушение голоса, которое проявляется в изменении тембра, высоты тона и громкости. Дисфония может быть вызвана усталостью голосовых связок, стрессом, аллергиями или инфекциями. Симптомы включают хрипоту, изменение тембра и трудности при произношении слов.

2. Афония — полная или частичная потеря голоса. Человек с афонией может говорить только шепотом или издавать прерывистые звуки. Причины афонии включают воспаление гортани, травмы, инфекции или опухоли.

3. Фонастения — нарушение, связанное с дискоординацией работы дыхательного, артикуляционного и фонационного аппаратов. Проявления фонастении включают быстрое утомление голоса, дрожание, прерывание речи и снижение силы голоса.

4. Брадилалия и тахилалия — нарушения, связанные с изменением скорости речи. Брадилалия характеризуется замедленной речью, а тахилалия — ускоренной. Оба состояния могут затруднять коммуникацию и приводить к непониманию со стороны окружающих.

5. Заикание — нарушение, при котором человек испытывает трудности при произношении слов или фраз. Заикание может сопровождаться повторением звуков, слов или длинными паузами. Это нарушение может оказывать значительное влияние на социальную и эмоциональную жизнь человека.

6. Дислалия — нарушение звукопроизношения, при котором человек заменяет, искажает или смешивает звуки. Это может быть вызвано недостатком обучения или снижением слуховой восприимчивости.

7. Ринолалия — нарушение произношения, связанное с физиологическими дефектами речевого аппарата, такими как недостаточное закрытие небных дуг. Ринолалия проявляется в изменении тембра голоса, который становится гнусавым.

Диагностика нарушений голоса

Диагностика [1], [6] нарушений голоса включает несколько этапов. Для дисфонии и афонии используются голосовые тесты и эндоскопия гортани, которые позволяют оценить состояние голосовых связок. Для диагностики фонастении применяются неврологические тесты и анализ речи. Брадилалия и тахилалия диагностируются с помощью оценки скорости речи и произношения звуков. Заикание выявляется с помощью специальных тестов, таких как чтение текста или акустический анализ речи. Для диагностики дислалии и ринолалии используются методы анализа звукопроизношения и аудиологические исследования.

Лечение нарушений голоса

Лечение нарушений голоса зависит от их типа и причин. Для дисфонии и афонии часто применяются упражнения для укрепления голосовых связок, дыхательные техники и изменение высоты голоса. Фонастения лечится с помощью логопедической терапии, направленной на улучшение координации работы голосового аппарата. Брадилалия и тахилалия требуют терапии, направленной на нормализацию скорости речи. Заикание лечится с помощью психологической поддержки, техник управления стрессом и упражнений на улучшение дыхания и речи. Для дислалии и ринолалии применяются логопедические упражнения и, в некоторых случаях, хирургические методы.

Параметры голоса и акустический анализ

Голос представляет собой сложный акустический сигнал, который можно разложить на ряд параметров, отражающих его свойства. Основные параметры голоса включают:

– Частота общего тона (ЧОТ) [7], [8] — основная частота колебания голосовых связок. ЧОТ зависит от пола, возраста и эмоционального состояния человека. Обычно женский голос имеет частоту 180–250 Гц, а мужской — 100–130 Гц. Вычисляется с помощью дискретного преобразования Фурье (ДПФ).

– Форманты [8], [9] — акустические характеристики звука, которые определяют тембр и разборчивость речи. Они представляют собой резонансные частоты голосового тракта. Представлены на рисунке 1.

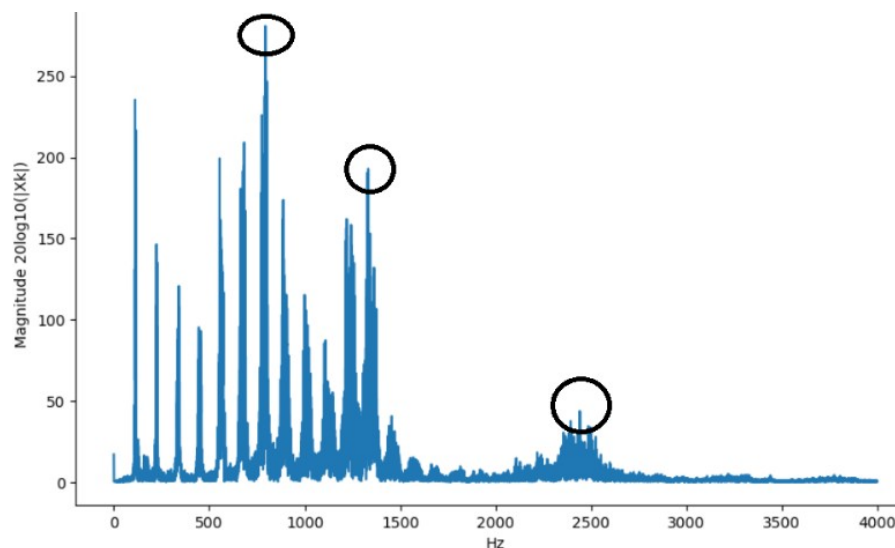


Рисунок 1 - Частотные характеристики голоса
DOI: <https://doi.org/10.60797/IRJ.2025.160s.1.1>

Примечание: слева направо выделены пики — форманты

– Jitter [7], [8] — изменение частоты голоса во времени, либо же изменение периода колебаний голосовых связок. Вычисляется по формуле 1. Jitter измеряется в процентах и отражает стабильность голоса. Точки для вычисления Jitter`а видно на рисунке 2.

$$\text{Jitter}(\%) = \frac{\frac{1}{1-N} \sum_{i=1}^{N-1} (T_i - T_{i+1})}{\frac{1}{N} \sum_{i=1}^N T_i}, \quad (1)$$

где N — количество периодов основного тона;

T_i — длина периода основного тона.

– Shimmer [7], [8] — изменение амплитуды голоса во времени. Shimmer также измеряется в процентах и указывает на вариабельность громкости голоса. Вычисляется по формуле 2. Точки для вычисления Shimmer`а видно на рисунке 2.

$$\text{Shimmer}(\%) = \frac{\frac{1}{1-N} \sum_{i=1}^{N-1} (A_i - A_{i+1})}{\frac{1}{N} \sum_{i=1}^N A_i}, \quad (2)$$

где N — количество периодов основного тона;

A_i — Амплитуда пика основного тона.

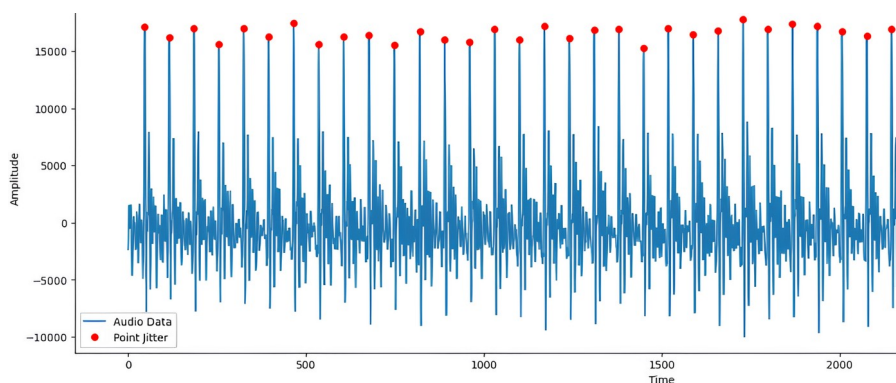


Рисунок 2 - Аудиозапись голоса
DOI: <https://doi.org/10.60797/IRJ.2025.160s.1.2>

Примечание: точками отмечены периоды и амплитуды основного тона

– Отношение сигнал/шум [7], [8] — отношение мощности гармоник к мощности шума. Этот параметр отражает чистоту голоса.

Акустический анализ голоса позволяет извлекать эти параметры и использовать их для диагностики нарушений. Например, повышенный уровень Jitter и Shimmer может указывать на наличие дисфонии или афонии.

Методология эксперимента

Для анализа нарушений голоса использовалась база данных VOICED [10], содержащая записи голоса 208 человек, из которых 58 были здоровыми, а 150 — с патологиями. Для извлечения параметров голоса использовались библиотеки языка программирования Python [11], такие как Numpy [12], Librosa [13]. Эти библиотеки позволяют анализировать аудиозаписи и извлекать такие параметры, как ЧОТ, Jitter и Shimmer, форманты. Пример извлеченных параметров, для обучения моделей машинного обучения, представлен на рисунке 3.

| | f0 | db_f0 | f1 | db_f1 | f2 | db_f2 | f3 | db_f3 | jitter | shimmer | lables |
|---|---------|---------|---------|-------|----------|-------|----------|-------|---------|----------|--------|
| 0 | 266.67 | 2002.41 | 8950.00 | 8.76 | 13383.33 | 4.00 | 18600.00 | 2.87 | 24.8929 | 6.311449 | 1 |
| 1 | 1100.00 | 2119.64 | 6550.00 | 40.28 | 11666.67 | 9.75 | 21233.33 | 0.73 | 7.1284 | 3.647323 | 1 |

Рисунок 3 - Пример таблицы параметров
DOI: <https://doi.org/10.60797/IRJ.2025.160s.1.3>

Существующие подходы машинного обучения

Для классификации нарушений голоса были рассмотрены различные алгоритмы машинного обучения [2], [3], [4]:

1. Метод k-ближайших соседей (KNN) — алгоритм, который классифицирует объекты на основе majority vote среди k-ближайших соседей в пространстве признаков. KNN не строит явную модель, но требует хранения всех обучающих данных. Эффективен для задач с четкой кластерной структурой.

2. Линейная регрессия (Linear Regression) — алгоритм для задач регрессии, который строит линейную зависимость между целевой переменной и входными признаками, минимизируя сумму квадратов ошибок. Позволяет оценить важность признаков через коэффициенты.

3. Стохастический градиентный спуск (SGDClassifier) — алгоритм оптимизации, используемый для обучения линейных моделей (логистической регрессии, SVM и др.). Обновляет веса на небольших пакетах данных, что делает его эффективным для больших наборов данных.

4. Логистическая регрессия — алгоритм, используемый для бинарной классификации. Он моделирует связь между входными признаками и вероятностью принадлежности к классу.

5. Дерево решений — алгоритм, который строит структуру в виде дерева для классификации или регрессии. Каждый узел дерева представляет признак, а ветви — возможные значения.

6. Случайный лес — ансамбль деревьев решений, который строит множество моделей и усредняет их прогнозы. Случайный лес снижает риск переобучения и повышает устойчивость к шуму в данных.

7. Градиентный бустинг — метод, который последовательно строит слабые модели для исправления ошибок предыдущих. К этой категории относятся такие алгоритмы, как LightGBM, CatBoost и XGBoost.

Результаты анализа

Наиболее эффективными оказался метод линейной классификации: логистическая регрессия, который показали точность 70%. Наименее эффективными были дерево решений (56%). Анализ важности признаков показал, что наиболее значимыми параметрами для классификации являются Shimmer, Jitter (рисунок 4). Использовалась метрика ассигасу, также все модели показывали одинаковую метрику AUC-ROC, равную 60%.

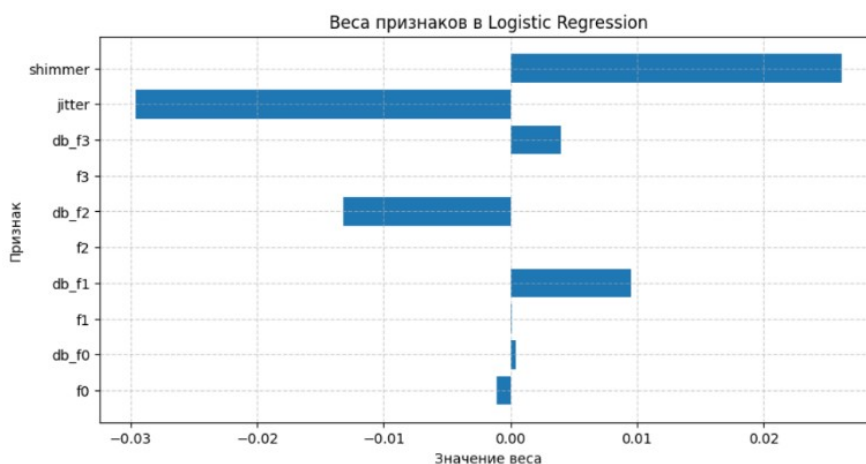


Рисунок 4 - Веса признаков в логистической регрессии
DOI: <https://doi.org/10.60797/IRJ.2025.160s.1.4>

Заключение

Исследование подтвердило перспективность применения машинного обучения для выявления нарушений голоса. Методы линейной классификации показали высокую эффективность, а анализ важности параметров позволил выявить ключевые признаки для дальнейшего улучшения моделей. Дальнейшие исследования могут привести к созданию автоматизированных систем диагностики и мониторинга нарушений голоса, что улучшит качество реабилитации пациентов.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Аникеева З.И. Современные методы диагностики и комплексного лечения респираторного тракта у профессионалов голоса в амбулаторных условиях / З.И. Аникеева. — Москва: Граница, 2011. — 416 с.
2. 10 самых популярных алгоритмов машинного обучения // Алгоритмы машинного обучения. — 2020. — URL: <https://cloud.vk.com/blog/samye-populyarnye-algoritmy-mashinnogo-obucheniya/> (дата обращения: 10.03.24)
3. Пять классических алгоритмов машинного обучения // Библиотека программиста. — 2022. — URL: <https://proglib.io/p/5-klassicheskikh-algoritmov-mashinnogo-obucheniya-o-kotoryh-vam-obyazatelno-sleduet-znat-2022-08-16> (дата обращения: 20.03.24)
4. 90+ алгоритмов и моделей машинного обучения, и их реализация с помощью языка программирования Python // Алгоритмы машинного обучения. — URL: <https://biconsult.ru/products/90-algoritmov-i-modeley-mashinnogo-obucheniya-i-ih-realizaciya-s-pomoshchyu-yazyka> (дата обращения: 30.03.2024)
5. Волкова Л.С. Логопедия : учебник для студентов дефектологических факультетов / Л.С. Волкова, С.Н. Шаховская. — Москва: Владос, 2004. — 704 с.
6. Лаврова Е.В. Логопедия. Основы фонопедии / Е.В. Лаврова. — Москва: Академия, 2007. — 144 с.
7. Лысак А.П. Акустический анализ голоса в норме и патологии. / А.П. Лысак // Речевые технологии. — 2012. — № 4. — С. 25–32.
8. Bäckström T. Introduction to Speech Processing / T. Bäckström, O. Räsänen, A. Zewoudie et al. — 2nd. ed. — 2022. — URL: <https://speechprocessingbook.aalto.fi> (accessed: 30.03.24). DOI: 10.5281/zenodo.6821775
9. Кодзасов С.В. Общая фонетика / С.В. Кодзасов, О.Ф. Кривнова. — Москва: изд-во РГГУ, 2001. — 592 с.
10. Cesari U. A new database of healthy and pathological voices / U. Cesari, G. De Pietro, E. Marciano et al. — Naples: Computers & Electrical Engineering, 2018. — 11 p. doi: 10.1016/j.compeleceng.2018.04.008
11. Python programming language // Python. — URL: <https://www.python.org> (accessed: 10.03.2024)
12. Python Nampai programming language framework // Nampai. — URL: <https://numpy.org> (accessed: 11.03.2024)
13. Libros Python programming language framework // Libros. — URL: <https://librosa.org/doc/latest/index.html> (accessed: 12.03.2024)

Список литературы на английском языке / References in English

1. Anikeeva Z.I. Sovremennyye metody' diagnostiki i kompleksnogo lecheniya respiratornogo trakta u professionalov golosa v ambulatorny'x usloviyax [Modern methods of diagnosis and complex treatment of the respiratory tract among voice professionals in outpatient settings] / Z.I. Anikeeva. — Moscow: Granicza, 2011. — 416 p. [in Russian]
2. 10 samikh populyarnikh algoritmov mashinnogo obucheniya [10 most popular machine learning algorithms] // Machine learning algorithms. — 2020. — URL: <https://cloud.vk.com/blog/samye-populyarnye-algoritmy-mashinnogo-obucheniya/> (accessed: 10.03.24) [in Russian]
3. Pyat' klassicheskikh algoritmov mashinnogo obucheniya [Five classical machine learning algorithms] // Programmer's library. — 2022. — URL: <https://proglib.io/p/5-klassicheskikh-algoritmov-mashinnogo-obucheniya-o-kotoryh-vam-obyazatelno-sleduet-znat-2022-08-16>. (accessed: 20.03.24) [in Russian]
4. 90+ algoritmov i modeley mashinnogo obucheniya, i ih realizaciya s pomoshh'ju yazyka programmirovaniya Python [90+ algorithms and models of machine learning, and their implementation using the Python programming language] // Machine learning algorithms. — URL: <https://biconsult.ru/products/90-algoritmov-i-modeley-mashinnogo-obucheniya-i-ih-realizaciya-s-pomoshchyu-yazyka> (accessed: 30.03.2024) [in Russian]
5. Volkova L.S. Logopediya : uchebnik dlya studentov defektologicheskix fakul'tetov [Speech therapy : a textbook for students of defectological faculties] / L.S. Volkova, S.N. Shaxovskaya. — Moscow: Vlados, 2004. — 704 p. [in Russian]
6. Lavrova E.V. Logopediya. Osnovy' fonopedii [Speech therapy. Fundamentals of phonopedia] / E.V. Lavrova. — Moscow: Akademiya, 2007. — 144 p. [in Russian]
7. Ly'sak A.P. Akusticheskij analiz golosa v norme i patologii [Acoustic analysis of voice in norm and pathology]. / A.P. Ly'sak // Speech technology. — 2012. — № 4. — P. 25–32. [in Russian]
8. Bäckström T. Introduction to Speech Processing / T. Bäckström, O. Räsänen, A. Zewoudie et al. — 2nd. ed. — 2022. — URL: <https://speechprocessingbook.aalto.fi> (accessed: 30.03.24). DOI: 10.5281/zenodo.6821775

9. Kodzasov S.V. Obshchaya fonetika [General Phonetics] / S.V. Kodzasov, O.F. Krivnova. — Moscow: publishing house RGGU, 2001. — 592 p. [in Russian]
10. Cesari U. A new database of healthy and pathological voices / U. Cesari, G. De Pietro, E. Marciano et al. — Naples: Computers & Electrical Engineering, 2018. — 11 p. doi: 10.1016/j.compeleceng.2018.04.008
11. Python programming language // Python. — URL: <https://www.python.org> (accessed: 10.03.2024)
12. Python Nampai programming language framework // Nampai. — URL: <https://numpy.org> (accessed: 11.03.2024)
13. Libros Python programming language framework // Libros. — URL: <https://librosa.org/doc/latest/index.html> (accessed: 12.03.2024)