



**ПРОФЕССИОНАЛЬНОЕ  
ЗАВТРА**



VI Всероссийский сетевой конкурс студенческих проектов с участием студентов с инвалидностью

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО  
ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФГАОУ ВО «Уральский федеральный университет имени  
первого Президента России Б.Н. Ельцина»**

**Институт фундаментального образования  
Кафедра информационных систем и технологий**

**Направление «Профессиональное завтра в науке»**

**Номинация «Полезное изобретение»**

**«Адаптируемая система автоматического распознавания и коррекции речи  
для лиц с нарушениями речи»**

**Выполнил:  
Бредихин Борис Андреевич**

Екатеринбург, 2023

## **1. Обоснование актуальности проблемы**

Актуальность работы обусловлена значительным количеством людей с нарушениями речи, а также отсутствием продуктов для распознавания и коррекции русскоязычной аномальной речи. По данным, предоставленным Центром инклюзивного образования УрФУ, в университете обучаются 282 студента с инвалидностью (0,8% от всех студентов), из них 4% имеют нарушения слуха (и, вероятно, речи) и 28% — нарушения опорно-двигательного аппарата, иногда сопровождаемые нарушениями речи, таким образом, порядка 32 % людей с инвалидностью являются целевыми пользователями изобретения.

Стандартные аналоги изобретения, анализ которых приведён в соответствующем разделе, допускают значительное количество ошибок, меняющих смысл высказываний, при распознавании речи людей с нарушениями речи.

## **2. Потенциальные потребители**

На момент подготовки заявки непосредственными потенциальными пользователями изобретения являются люди с гиперкинетической дизартрией и их собеседники. Тестирование проводилось только на этом виде нарушений речи.

После качественного решения задачи для гиперкинетической дизартрии планируется добавить возможность распознавания других видов нарушений речи, в первую очередь речи людей с ДЦП и с нарушениями слуха.

Также изобретение возможно применять в медицинских целях для автоматической диагностики степени нарушений речи и для коммуникации с пациентами.

### **3. Название изобретения и его класс**

«Адаптируемая система автоматического распознавания и коррекции речи для лиц с нарушениями речи»

Классы МПК: А61F 11/04, G10L 13/033, G10L 15/04, G10L 15/14, G10L 15/16, G10L 15/187, G10L 15/20, G10L 21/057

Область изобретения: Альтернативная и дополненная коммуникация

### **4. Анализ аналогов**

Прямой аналог изобретения — Google Project Relate [1] позволяет по 500 фразам обучить индивидуальную модель распознавания речи, но судя по информации на официальном сайте, поддерживается только английский язык и ограниченный набор фраз.

Косвенные аналоги — это стандартные системы распознавания речи: Яндекс, ВК, Google, Apple Siri и другие. Эти системы не приспособлены для распознавания аномальной речи и не озвучивают распознанный текст, то есть их использование как средства дополненной коммуникации затруднено.

Прототипом изобретения является метод распознавания речи, предложенный L. Rabiner [2] в 1984 году, представляющий собой метод анализа фаз звуковой волны с применением скрытых марковских моделей, а также разработанная в 1952 году система распознавания 10 цифр по изменению звукового давления.

### **5. Сущность изобретения**

Изобретение представляет собой индивидуально обучаемый метод распознавания и коррекции аномальной речи с использованием списка всех допустимых слов и словоформ естественного языка (не следует путать с

ограниченным набором фраз).

## 5.1 Формула изобретения

1. Система распознавания речи с применением нейронных сетей на основе архитектуры wav2vec2 для распознавания фонем,
2. использующая индивидуальные особенности пользователей для создания скрытых марковских моделей,
3. использующая индивидуально обученные скрытые марковские модели для преобразования фонем в буквы,
4. использующая специальный алгоритм для разделения последовательности букв на слова и поиска ближайшего слова в словаре по нормализованному расстоянию Левенштейна.

## 5.2 Описание метода распознавания речи

Метод состоит из предобученной модели распознавания фонем XLS-R, далее обозначаемой как  $XLS_R$ , индивидуально обучаемой скрытой марковской модели  $HMM$  и словаря допустимых слов  $W$

Возможна работа в двух режимах: обучение и использование.

В режиме обучения пользователь отправляет на вход аудиозапись и содержащийся в ней текст. По аудиозаписи распознаются фонемы  $P_i, i = \overline{1, n}$ , а текстовая строка  $S$  приводится к нижнему регистру, удаляются пробелы, немые буквы («ъ», «ь» и другие случаи) и знаки препинания. после этого составляются тройки  $\langle p_i, p_{i+1}, s_i \rangle$ , в соответствии с которыми обновляется матрица переходов скрытой марковской модели по алгоритму Баума-Велша.

В режиме использования пользователь отправляет на вход аудиозапись и содержащийся в ней текст. По аудиозаписи распознаются фонемы, которые преобразуются скрытой марковской моделью в текст по алгоритму Витерби.

Затем с помощью анализа подстрок распознанного текста и их сравнению по расстоянию Левенштейна с элементами множества  $W$  формируется итоговый текст.

### 5.3 Прототип изобретения

В соответствии с формулой изобретения и описанным методом распознавания речи было разработано web-приложение для индивидуального обучения и использования описанного метода.

### 5.4 Тестирование изобретения

Предложенный метод был протестирован человеком с гиперкинетической дизартрией на отрывке из романа А. С. Пушкина "Евгений Онегин". После обучения на трёх строках система смогла корректно распознавать речь.

## 6. Скриншоты и архитектурные диаграммы

### Запишите аудио для распознавания

▶ 0:00 — 🔊

Начать    Распознать

Распознанные фонемы: *keelnafazeno tsitlekakstovo*

Распознанный текст: /

Введите правильный текст:

Сохранить

Рисунок 1 — Пример распознавания фонем и коррекции результата

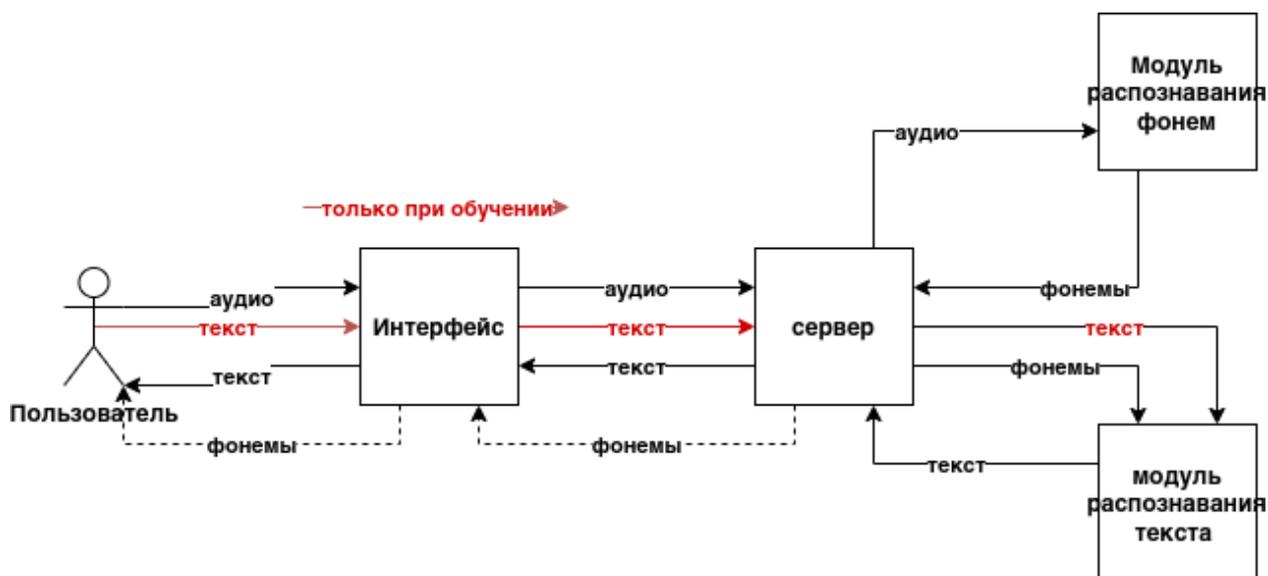


Рисунок 2: Архитектура системы

## СПИСОК ЛИТЕРАТУРЫ

1. Project Relate – Google Research [Electronic resource] // Project Relate – Google Research. URL: <https://sites.research.google/relate/> (accessed: 11.06.2023).
2. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition // Proc. IEEE. 1989. Vol. 77, № 2. P. 257–286.