

Министерство науки и высшего образования Российской Федерации

Томский государственный университет
систем управления и радиоэлектроники (ТУСУР)

На правах рукописи



Новохрестова Дарья Игоревна

**МЕТОДИКИ И АЛГОРИТМ АНАЛИЗА ДАННЫХ ПРИ ОЦЕНКЕ
КАЧЕСТВА ПРОИЗНОШЕНИЯ СЛОГОВ В ПРОЦЕССЕ РЕЧЕВОЙ
РЕАБИЛИТАЦИИ**

05.13.17 – Теоретические основы информатики

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель
доктор технических наук,
профессор, И.А. Ходашинский

Томск 2022

Содержание

Введение.....	4
1 Обзор существующих методов, подходов и алгоритмов оценки качества голоса и речи.....	13
1.1 Объективные методы оценки качества речи и голоса при передаче по каналам связи.....	15
1.2 Методы оценки в задачах синтеза и анализа речи.....	19
1.3 Субъективные методы оценки качества речи и голоса.....	23
1.4 Оценка речи в медицине. Применяемая методика оценки речи в реабилитации на основе ГОСТ Р 50840–95.....	26
1.5 Выводы по главе.....	31
2 Анализ базы данных записей речи пациентов.....	32
2.1 Анализ записей и оценок качества речи на основе существующей методики.....	35
2.2 Анализ спектрограмм аудиозаписей слогов, содержащих наиболее подверженные изменениям фонемы.....	41
2.3 Выводы по главе.....	47
3 Методика и алгоритм оценки речевых сигналов.....	48
3.1 Алгоритм нахождения количественного значения схожести двух речевых сигналов.....	48
3.1.1 Выбор алгоритмов временной нормализации и подсчета количественной оценки.....	49
3.1.2 Выбор метрики для оценки.....	55
3.1.3 Алгоритм нахождения количественной оценки схожести двух речевых сигналов.....	73
3.2 Методика оценки схожести звуковых сигналов.....	77
3.3 Выводы по главе.....	85
4 Методика оценки речевой реабилитации с использованием получаемых оценок и биологической обратной связи.....	87
4.1 Использование биологической обратной связи в процессе речевой реабилитации.....	87

4.2 Формирование положительных подкрепляющих стимулов.....	91
4.2.1 Кратковременные подкрепляющие стимулы	91
4.2.2 Долговременные подкрепляющие стимулы.....	94
4.3 Оценка качества речи в рамках речевой реабилитации с использованием биологической обратной связи	99
4.4 Выводы по главе.....	102
5 Разработка и внедрение программного комплекса	104
5.1 Структура базы данных записей и программного обеспечения для записи и оценка качества речи	104
5.1.1 База данных записей пациентов	104
5.1.2 Программа для оценки качества речи «Разборчивость речи».....	107
5.1.3 Подсчет количественной оценки аудиозаписей сеанса	111
5.1.4 Запись сеанса оценки произношения слогов с проведением оценки и БОС	114
5.2 Результаты тестирования программного комплекса на основе искусственно сформированного набора данных	121
5.3 Внедрение программного комплекса в процесс реабилитации и результаты внедрения.....	125
5.4 Выводы по главе.....	128
Заключение	129
Список использованной литературы.....	131
Приложение А – Патент на изобретение	150
Приложение Б – Свидетельства о государственной регистрации	151
Приложение В – Акты внедрения	156
Приложение Г – Сертификат гранта Американского акустического сообщества (ASA)	160
Приложение Д – Спектрограммы сигналов	161
Приложение Е – Матрицы расстояний между сигналами	165
Приложение Ж – Графическое представление аудиосигналов.....	168
Приложение И – Декомпозиции процессов оценки	170

Введение

Оценка качества речи и речевого сигнала применяется во многих областях науки и техники. Оценка речи используется при разработке и сравнении методов и алгоритмов передачи сигналов по каналам связи, в практических задачах идентификации и аутентификации по голосу в сфере защиты информации, при распознавании речевого сигнала в процессе речевой и голосовой реабилитации после заболеваний органов речеобразующего тракта. Одной из групп таких заболеваний является онкология, лечение которой в большинстве случаев требует хирургического вмешательства. За 2020 год зарегистрировано более 20 000 новых случаев онкологии органов полости рта и ротоглотки, среднегодовой темп прироста 2,5% [1, 2]. Проводится комбинированное лечение, включающее химиолучевую терапию и оперативное вмешательство в различной последовательности, где ведущим компонентом остается хирургический этап. Хирургическое лечение при раке полости рта и ротоглотки неизбежно приводит к нарушению таких жизненно важных функций как речь, жевание и глотание, в некоторых случаях нарушаются также голос и дыхание. Согласно концепции развития здравоохранения России рак полости рта и ротоглотки относится к группе наиболее социально значимых болезней, так как нарушение звучной речи в постоперационный период приводит к инвалидизации, снижая трудовой потенциал и качество жизни пациентов. Таким образом, после проведения хирургического лечения обязательным этапом является речевая реабилитация. Под речевой реабилитацией подразумевается комплекс медицинских и организационных мероприятий для восстановления разборчивой речи пациента. Под голосовой реабилитацией – набор мероприятий по восстановлению и формированию голоса пациента. Согласно национальному проекту «Здравоохранение» одним из приоритетных направлений является развитие персонализированной медицины. Поэтому оценка произношения в рамках

речевой реабилитации должна производиться с учетом особенностей речи пациента, который ее проходит.

Особенности решения задачи оценки речи зависят от сферы применения. Так в одних задачах целью является получение объективных количественных значений. Речь анализируется с точки зрения вычисления параметров сигнала и их изменении после проведения манипуляций (передача по каналам связи, сжатие сигнала, оценка защищенности). В таких случаях работа идет или с параметрами самого аудиосигнала (частота, меры зашумленности, энергия и др.) с учетом влияния измерительных приборов, или с параметрами голоса диктора (частота основного тона, длительность фонации и т.д.). Над решением подобных и смежных с ними задач работали такие ученые как Р.В. Шафер, Л. Рабинер, А. Асеро, В.П. Бондаренко, А.В. Аргановский, Р.К. Потапова, А.А. Карпов, А.Л. Ронжин, Л.А. Чистович, Т.К. Винцюк, Л.В. Бондарко, Л.В. Златоустова, В.Н. Трунин-Донской, Т.В. Шарий, А.С. Фадеев, Р.В. Мещеряков, Е.Ю. Костюченко, В.П. Коцубинский, С.Ю. Корнилов, И.А. Рахманенко, С.С. Харченко и многие другие. Другим актуальным направлением является решение задачи распознавания речи. Для ее решения применяются методы, основанные на искусственных нейронных сетях, скрытых марковских моделях, динамическом программировании, дискриминантном анализе и других подходах. Существенный вклад в развитие этих методов внесли Л.Е. Баум, Дж.К. Бейкер, Б.Т. Лоуэрр, Л.Р. Липорак, Б. Жуань, С.Е. Левинсон и др.

Однако применение методов, направленных на решение вышеописанных задач, для оценки речи в речевой реабилитации не является возможным из-за принципиальных отличий в сути проводимого сравнения и применяемых подходов. Например, в методах при передаче по каналам связи оцениваются два разных представления (до и после передачи) одной и той же реализации, а в методы распознавания речи эффективны на более крупных фонетических единицах (от слова и больше). Большинство применяемых методов анализа качества речи в медицине основаны на

экспертной оценке или сопоставлении с эталоном. Чем больше сходства с эталоном, тем лучше считается произношение и речь. В этом направлении публиковали свои работы такие ученые, как Д. Росситр, Д. Кьюли, Е.Е. Ляксо, Л.Н. Балацкая и др. Однако недостаток применяемых методов состоит в субъективности получаемых оценок, а применение эталона не учитывает важные особенности речи оцениваемого диктора. Поэтому построение методов и алгоритмов, позволяющих объективно оценивать речь диктора и в то же время учитывать его индивидуальность, является актуальной задачей.

Целью исследования является повышение эффективности процесса речевой реабилитации за счет автоматизации и сокращения времени оценки качества речи.

Для достижения поставленной цели необходимо решить следующие **задачи**:

1) выполнить анализ текущего состояния исследований в изучаемой предметной области: алгоритмы оценки качества и разборчивости речи, методики оценки речи в процессе речевой и голосовой реабилитации, алгоритмы анализа речевых сигналов;

2) разработать алгоритм количественной оценки схожести двух речевых сигналов, результаты работы которого согласуются с оценками, получаемыми по старой методике;

3) разработать методику оценки схожести речевых сигналов, учитывающую особенности речи дикторов (пациентов);

4) адаптировать методику голосовой реабилитации для речевой реабилитации с использованием биологической обратной связи и оценки качества речи;

5) реализовать и апробировать алгоритм и методики на практике в процессе речевой реабилитации.

Объектом исследования данной работы является процесс речевой реабилитации пациентов после комбинированного лечения онкологических заболеваний органов полости рта и ротоглотки.

Предметом исследования является алгоритмы и методики оценки качества речи.

Основные методы исследования, примененные в диссертационной работе – это методы моделирования, системного анализа, цифровой обработки сигналов и математической статистики.

Научная новизна результатов работы и проведенных исследований заключается в следующем:

1. Предложен алгоритм нахождения количественной оценки схожести двух речевых сигналов, отличающийся гибридной мерой совпадения, основанной на комбинации метрик DTW-расстояния, коэффициента корреляции и расстояния Минковского.

2. Предложена методика оценки схожести речевых сигналов до и после оперативного вмешательства, отличающаяся применением нескольких опорных речевых сигналов, отражающих одну и ту же фонетическую единицу.

3. Для проведения речевой реабилитации адаптирована методика голосовой реабилитации с использованием бионического принципа биологической обратной связи, отличающаяся новой гибридной мерой для оценки схожести речевых сигналов.

Теоретическая значимость исследования заключается в развитии технологии анализа и оценки качества речи. Предложенная методика позволила проводить количественную оценку схожести речевых сигналов на основе сравнения с несколькими опорными сигналами в заранее известном интервале значений.

Практическая значимость работы подтверждается применением полученных в ней результатов для решения практической задачи оценки качества произношения слогов в процессе речевой реабилитации после

хирургического лечения онкологических заболеваний органов полости рта и ротоглотки. Результаты работы внедрены в процесс лечения, проводимого на базе НИИ Онкологии томского НИМЦ. Методика речевой реабилитации может быть также применена в рамках лечения иных заболеваний со сходным характером вмешательства.

Положения, выносимые на защиту:

1. Предложенный алгоритм нахождения количественной оценки схожести двух речевых сигналов позволяет получить значение меры схожести двух различных по продолжительности аудиосигналов. Среднее количество совпадений оценок, получаемых по предложенному алгоритму, с оценками по разработанной ранее методике составляет 83%.

Соответствует пункту 5 паспорта специальности 05.13.17: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

2. Методика оценки схожести речевых сигналов до и после оперативного вмешательства, отличающаяся применением нескольких опорных речевых сигналов, позволяет учитывать вариативность произношения фонем и индивидуальные особенности речи пациента.

Соответствует пункту 5 паспорта специальности 05.13.17: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

3. Адаптированная методика голосовой реабилитации с использованием бионического принципа биологической обратной связи и методики оценки качества речи позволяет формировать подкрепляющий стимул на основе сравнения ретроспективных данных и сократить время оценки качества речи в речевой реабилитации более чем на 64 %.

Соответствует пункту 13 паспорта специальности 05.13.17: Применение бионических принципов, методов и моделей в информационных технологиях.

Обоснованность и достоверность результатов работы подтверждается их внутренней непротиворечивостью и положительным эффектом от внедрения научных исследований в работу действующего медицинского учреждения (ФГБНУ «Томский НИМЦ РАН» «Научно-исследовательский институт онкологии»), о чем свидетельствует соответствующий Акт о внедрении. Полученные результаты согласуются с оценками по ранее используемой методике на основе ГОСТ Р 50840–95 и результатами исследований других научных групп.

Внедрение результатов. Результаты диссертационной работы внедрены в процесс речевой реабилитации в рамках лечения онкологических заболеваний на базе НИИ Онкологии г. Томска, а также в учебный процесс Томского государственного университета систем управления и радиоэлектроники.

Личный вклад. Основные выносимые на защиту научные результаты получены лично автором. Часть опубликованных работ написана в соавторстве с сотрудниками научных групп и научным руководителем. Автором совместно с сотрудниками ФГБНУ «Томский НИМЦ РАН» «Научно-исследовательский институт онкологии» проведены внедрение и апробация результатов работы. Разработанные алгоритм и методики были реализованы в виде программного комплекса лично автором. Постановка изложенных в диссертации задач осуществлялась совместно научным руководителем д.т.н., профессором Ходашинским И.А. и консультантом к.т.н, доцентом Костюченко Е.Ю.

Апробация работы. Основные и промежуточные результаты исследования докладывались и обсуждались на следующих конференциях и семинарах:

- International Conference on Speech and Computer, SPECOM (2016, 2017, 2018, 2019, 2021);
- X Международной IEEE научно-технической конференции «Динамика систем, механизмов и машин» (Омск, Россия, 2016г.),
- II Конгрессе «Здравоохранение России. Технологии опережающего развития» КОНГРЕСС ЗДРАВ 2016 (Томск, 2016г);
- III Всероссийском конкурсе студенческих научных обществ и конструкторских бюро, секция «Биомедицина, медицинская физика», (Алтайский государственный университет, Барнаул, 2017);
- Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР» (ТУСУР, Томск, 2019, 2021, 2022);
- Международной научно-практической конференции «Электронные средства и системы управления» (ТУСУР, Томск, 2016, 2017, 2018, 2019, 2021);
- на заседаниях кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР и IEEE семинарах «Интеллектуальные системы моделирования, проектирования и управления» в г. Томске.

Были получены один патент на изобретение (приложение А), четыре свидетельства о государственной регистрации программ для ЭВМ, одно свидетельство государственной регистрации базы данных (приложение Б).

Работа выполнена при финансовой поддержке РФФИ (научный проект № 20–37-90082 «Модель и алгоритмы анализа данных при оценке качества произнесения слогов в процессе речевой реабилитации»), РНФ (научный проект № 16-15-00038) и гранта для научных проектов, выполняемых молодежными коллективами научно-исследовательских лабораторий образовательных организаций высшего образования, подведомственных Министерству науки и высшего образования Российской Федерации (номер проекта FEWM-2020-0042, АААА-А20-120111190016-9). Часть

исследований проводилась при поддержке стипендии для акустиков – студентов и аспирантов из России, полученной от Американского акустического общества (Приложение Г).

Публикации по теме диссертации. По результатам исследований опубликовано 33 работы, из них 4 статьи в журналах, входящих в перечень рекомендованных ВАК журналов. 14 публикаций проиндексированы в международной базе SCOPUS, 24 публикаций в материалах международных и всероссийских научных конференций.

Структура и объем диссертации. Диссертация содержит введение, 5 глав, заключение, 8 приложений и список источников из 211 наименований. Объем диссертационной работы: 171 страница, в том числе 23 таблицы и 50 рисунков.

Во введении обосновывается актуальность темы исследования, формулируется цель работы, излагаются полученные автором основные результаты проведенных исследований, показывается их научная новизна, теоретическая и практическая значимость, отражаются основные положения, выносимые на защиту.

В первой главе производится обзор проблемы исследования. Описываются алгоритмы и методы оценки качества речи, применяемые в различных сферах решаемых практических задач. Рассмотрен способ оценки качества речи, который использовался в качестве основы для проведения оценки речи в процессе речевой реабилитации до внедрения предложенных алгоритмов и методик. Описаны требования к разрабатываемым алгоритмам и методам, а также программному комплексу на их основе.

Во второй главе описывается анализ базы данных аудиозаписей пациентов, которые проходили лечение онкологических заболеваний органов рта и ротоглотки и речевую реабилитацию. Выделены основные проблемные фонемы, работе с которыми уделяется внимание в рамках речевой реабилитации, список проблемных фонем скорректирован с учетом

решаемой практической задачи. Описан визуальный анализ спектрограмм аудиозаписей слогов, содержащих проблемные фонемы, с выделением особенностей изменений, характерных для постоперационной речи пациентов.

В третьей главе описывается разработанный алгоритм нахождения количественного значения схожести двух речевых сигналов. Описаны подход к нормализации речевых сигналов к единой длине и формирование гибридной меры для нахождения количественного значения: подбор метрик и механизма их комбинации в гибридной мере. Описана методика оценки схожести звуковых сигналов с использованием нескольких опорных сигналов. Показано преимущество использования опорных сигналов того же диктора, для которого производится оценка. Описана практическая реализация методики для решения поставленной задачи речевой реабилитации.

В четвертой главе описывается методика речевой реабилитации и оценка качества речи с использованием количественных оценок и биологической обратной связи в рамках речевой реабилитации. Описан процесс оценки речи в рамках проведения сеансов реабилитации. Приводятся алгоритмы формирования кратковременных и долговременных подкрепляющих стимулов для реализации метода бионического принципа биологической обратной связи. Показаны результаты по повышению эффективности процесса речевой реабилитации за счет сокращения времени проведения оценки качества речи.

В пятой главе описана реализация предложенных алгоритма и методик в составе разработанного программного комплекса по оценке качества речи. Приводятся результаты тестирования работы комплекса на здоровых дикторах и результаты внедрения комплекса в процесс речевой реабилитации.

1 Обзор существующих методов, подходов и алгоритмов оценки качества голоса и речи

Понятия речь и голос в некоторых сферах являются разнозначными, в то время как в других сферах четко разделены. Решаемая научно-практическая задача относится к сфере медицины, поэтому воспользуемся определениями, которые дает Большая медицинская энциклопедия (1970). «Речь – высшая форма символически-выразительных функций», «речевой процесс может быть отождествлен с рефлекторным процессом, функции которого представлены процессами слышания и произношения (понимания) фонетических единиц (слог, слово, фраза и т.д.)». В то время как «голос – это всеобъемлющее понятие, которое объединяет в себя все звуки, исходящие из гортани человека, вне зависимости от их предназначения». Использование таких определений позволяет определить особенности, которые необходимо учитывать в рамках решаемой задачи. Так в речевой реабилитации происходит восстановление разборчивой речи пациентов, соответственно акцент сделан на изменение механизмов артикуляции при произношении конкретных фонем для формирования разборчивых, понятных для окружающих речевых конструкций. Необходимость изменения механизмов артикуляции вызвана невозможностью корректного использования ранее сформированных механизмов из-за изменения конструкций органов артикуляции.

Понятия качества речи и голоса также не имеют однозначных определений и зависят от контекста решаемой задачи. Так в задачах передачи сигналов по каналам связи качество речи может ограничиваться одной из характеристик – разборчивостью речи (для однозначности восприятия передаваемой информации) и в то же время включать в себя расчет характеристик самого аудиосигнала. В лингвистике под качеством речи может пониматься соответствие оцениваемой речи набору характеристик: точность, чистота, логичность, выразительность, богатство,

уместность и т.д. Также в рамках анализа используются различные подходы к оценке качества голоса [3]. В рамках решаемой научно-практической задачи под качеством речи понимается степень схожести оцениваемой речи с эталонным произношением, которое определяется для каждого пациента индивидуально в соответствии с его особенностями. Речевая реабилитация представляет собой комплекс медицинских, логопедических и психологических мероприятий для восстановления разборчивой речи пациента до предоперационного уровня после хирургического вмешательства в органы полости рта и ротоглотки. В процессе речевой реабилитации акцент идет на формирование или восстановления понятной для окружающих речи пациента, когда у него сохраняется возможность воспроизводить звуки, но частично или полностью отсутствует возможность формировать разборчивые фонетические единицы от слога и выше. Также в медицине есть понятие голосовой реабилитации, цель которой восстановление звучного голоса пациента. Под голосом в таком случае понимается возможность формировать звуки, например, при хирургическом удалении гортани пациенту предлагается освоить пищеводный голос. При комплексном нарушении может быть необходимо проведение и речевой, и голосовой реабилитации, такой тип реабилитации называется логопедическим, однако в некоторых источниках он соответствует термину речевой реабилитации.

Глобально методы оценки речи, голоса и их качеств можно разделить на три подгруппы: объективные, субъективные и оценка в рамках анализа и синтеза речи. Методы объективной оценки основаны на расчете характеристик сигнала и/или оценки изменений характеристик сигнала до и после проведения манипуляций над ним (к примеру, передача по каналу связи, фильтрация и т.д.). Субъективные методы основаны на оценке сигнала группой экспертов, сильно реже – одним экспертом. Основой оценки служит восприятие сигнала человеком путем прослушивания. Отдельной подгруппой можно выделить методы оценки речи в задачах

анализа и синтеза речи. Несмотря на то, что методы этой подгруппы являются объективными, однако их особенностью является не только контекст решаемых задач (распознавание речи, синтез речи, голосовое управление, идентификация диктора и т.д.), но и применение вероятностного подхода и методов классификации. Далее рассмотрены наиболее часто встречаемые методы оценки речи, а также подходы к оценке в рамках решения различных практических задач.

1.1 Объективные методы оценки качества речи и голоса при передаче по каналам связи

Разработка объективных методов оценки речи получила широкое распространение с развитием каналов передачи информации, в том числе аудиосигналов [4]. Многие предложенные меры были стандартизированы [5] и приняты в качестве эталона оценки речи. Однако все еще сохраняется ограничение объективных методов тем, что большинству требуется доступ к исходному речевому сигналу, а некоторые могут только моделировать низкоуровневую обработку (например, эффекты маскировки) слуховой системы. Группа методов, которым для оценки необходим доступ к исходному сигналу, называется интрузивными методами. Несмотря на недостатки, именно эта группа является самой изученной, а получаемые результаты оценки хорошо согласуются с оценками экспертов. Другой класс методов, а именно неинтрузивные методы, не требует доступа к исходному сигналу.

Большинство объективных методов основаны на сегментации речевого сигнала на кадры длительностью 10–30 мс, а затем вычислении меры искажения между исходным и обработанным сигналами. Итоговая оценка искажения речи вычисляется путем усреднения или сложения показателей искажения каждого речевого кадра. Более сложные объективные меры [6, 7] включают блок оценки временной задержки для выравнивания двух

сигналов перед вычислением меры искажения. Вычисление может быть выполнено либо во временной области (например, измерения отношения сигнал/шум), либо в частотной области (например, измерения спектрального расстояния). Для измерений в частотной области предполагается, что любые искажения или различия, обнаруженные в спектрах амплитуд, коррелируют с качеством речи.

Методы объективной оценки основаны на расчете одной или нескольких мер искажения. Обзоры рассчитываемых мер искажения можно найти в [8 – 12]. Сами используемые меры предложены довольно давно (с 70-90х гг. XX века), однако способы и подходы к их вычислению, модификациям и совместному использованию являются актуальными темами исследований и на текущий момент.

Если рассматривать непосредственно методы оценки качества речи, то можно выделить основные из них.

Самым часто встречающимся и используемым методом является Perceptual Evaluation of Speech Quality (PESQ). Он определен в рекомендации МСЭ-Т P.862 [13-15]. В [5] также предложен алгоритм оценки качества, называемый перцептивной мерой качества речи (PSQM). Область применения PSQM ограничена оценкой искажений, вносимых речевыми кодеками с более высокими битами, работающими по безошибочным каналам. Преимуществом этого метода является возможность предсказания результатов субъективной оценки речи экспертами. Данный метод является интрузивным методом. Модели оценки, созданные на основе этого метода, являются эталонными для разрабатываемых неинтрузивных методов [16, 17]. В [18] предлагается использование функции спектральных оценок речевых сигналов на основе статистических характеристик, но метод опять же предназначен для оценки речи при передаче по каналам связи, а качество речи характеризуется оценкой качества непрерывной речи PESQ. В [19] предложен способ устранения недостатка интрузивных методов относительно неинтрузивных.

Другим известным методом оценки речи является E-model [20]. Этот метод оценивает искажения, привносимые терминалам и сетью каждым в отдельности [21]. На выходе вычисляется значение R-фактора, в котором учитываются 20 параметров: однонаправленная задержка, коэффициент потери пакетов, потери данных, искажения из-за преобразования аналогового сигнала в цифровой и последующего сжатия, эхо, суммарная сквозная задержка, искажения вносимые оборудованием и др. Также в [21] описано сравнение результатов применения E-model и субъективных методов оценки.

Выделяется группа методов, основанных на расчете соотношения сигнал/шум (SNR) и сегментного соотношения сигнал/шум (segSNR) [22]. SNR - критерий общего отношения сигнал/шум по всей длине сигнала. Он имеет существенный недостаток – при недостаточной интенсивности полезного сигнала существует вероятность маскировки другим участком с большей интенсивностью. Это приводит к искажению общей оценки. segSNR представляет собой развитие SNR. В этом случае оценка отношения сигнал/шум производится на сегменте - интервале 15-20 мс, что позволяет исключить влияние неравномерной интенсивности сигнала [4].

Еще одним методом является Enhanced Modified Bark Spectral Distortion (EMBSD) [23]. Этот метод представляет собой объективную меру качества речи на основе слышимых искажений и модели познания. Этот метод показал результаты лучше, чем PESQ, в экспериментах с данными множественного доступа с временным разделением (TDMA), содержащими искажения, встречающиеся в реальных сетевых приложениях.

Преимуществом вышеуказанных объективных мер является то, что их легко реализовать и легко оценить. Однако их способность предсказывать субъективное качество ограничена, они не точно имитируют обработку сигнала на слуховой периферии. Нет явного моделирования или измерения нормальной частотной избирательности слуха, а также воспринимаемой громкости. В [24 – 30] описаны разработки объективных мер, основанных

на моделях слухового восприятия речи человека. Был достигнут значительный прогресс в моделировании нескольких стадий слуховой обработки на основе существующих знаний из психоакустики о том, как люди-слушатели обрабатывают тоны и полосы шума [31]. Одной из таких мер, учитывающих вышеуказанные знания, является the Bark distortion measure (BSD). Мера BSD была разработана в Калифорнийском университете [26]. Измерение BSD основано на предположении, что качество речи напрямую связано с громкостью речи, которая представляет собой психоакустический термин, определяемый как величина слухового ощущения. Для расчета громкости речевой сигнал обрабатывается по результатам психоакустических измерений. BSD оценивает общее искажение, используя среднее евклидово расстояние между векторами громкости эталонной и искаженной речи. BSD хорошо работает в тех случаях, когда искажение в вокализованных областях представляет собой общее искажение, поскольку обрабатывает только вокализованные области. В [27, 23, 32] описаны улучшения показателя BSD, что привело к разработке модифицированного показателя BSD (MBSD). Эксперименты в [26, 27] показали, что BSD и MBSD дают высокую корреляцию с результатами экспертной оценки по методу MOS (описан в разделе с субъективными методами).

Основными недостатками вышеописанных мер и методов относительно решаемой задачи можно обозначить следующие особенности:

1) суть вычисляемых мер и совместного использования их значений направлена на оценку искажений, вносимых в сигнал при проведении манипуляций (передача по каналам связи, сжатие и т.д.) над ним;

2) при парном сравнении происходит сравнение одной и той же реализации одной и той же фонетической единицы, сравнить разные реализации одной и той же единицы по этим методам невозможно;

3) основные меры направлены на вычисления характеристик самого аудиосигнала или качеств голоса, а не качеств речи.

1.2 Методы оценки в задачах синтеза и анализа речи

При решении задач синтеза и анализа речи основой применяемых методов является построение математических моделей речевых сигналов. В [33] описаны основные концепции, используемые в современных системах распознавания и понимания речи. Можно выделить следующие виды систем распознавания речи: контекстно-зависимые и контекстно-независимые системы, системы отдельной и непрерывной речи, диктор-зависимые и диктор-независимые системы и другие. Выбор того или иного типа системы зависит от сферы применения. Основным подходом к оценке качества в системах распознавания речи является вычисление точности распознавания как отношение количества корректно распознанных слов к общему числу слов [34]. В [35] описан подход к улучшению качества распознавания за счет применения методов поиска информации. Непосредственно оценка качества речи применяется в контексте соответствия аудиосигнала, содержащего речь, или условий записи аудиосигнала предъявляемым требованиям для достижения заявленного уровня точности распознавания. Отдельным направлением исследований выделяется улучшение систем распознавания для возможности работы с зашумленными сигналами и различными типами речи (к примеру, шепотной, с наличием дефектов речи и т.д.), как например в [36]. В [37] рассмотрены подходы к оценке речи на основе измерений разборчивости речи и узнаваемости голоса говорящего при передаче речи на фоне аддитивных помех. Оригинальная система распознавания русской речи SIRIUS и лежащий в ее основе подход к формированию морфемного уровня представления речи и языка описан в [38]. В [39] предложена модель автоматического распознавания речи на основе анализа звуковой и визуальной информации, а также модель компьютерного синтеза аудиовизуальной русской речи по произвольному тексту, представлен краткий обзор речевых технологий, применяемых в авионике.

Если говорить про системы синтеза речи, то основным методом оценки синтезированной речи является ее сравнение с естественной. Один из подходов описан в [40], где сравнивалась синтезированная в Text Aloud речь и естественная речь диктора. В [41] приведена модель описания гласных звуков с получением основных амплитудно-частотных характеристик и численные значения параметров на основе спектров и спектрограмм для дальнейшего применения в задачах синтеза речи.

Основной недостаток применяемых в вышеописанных системах и методах подходов относительно решаемой задачи это наличие вероятностного подбора, а также их применимость только к единицам речи от слова и больше. Можно дополнительно отметить слабую применимость систем распознавания к искаженной речи (в том числе к постоперационной) [42].

Если говорить про повышение качества распознавания речи, то математическим аппаратом выравнивания нераспознанных или пропущенных фонем в словах являются скрытые марковские модели (СММ). Этот аппарат был предложен Баумом Л.Е. [43] и был использован для распознавания речи в [44 – 46]. В [47] описано обобщение метода оценки параметров непрерывных СММ на случай использования смесей логарифмически вогнутых и/или эллиптически симметричных распределений. Пересчет параметров СММ для случая аддитивной суммы гауссовых распределений представлены в [48]. Была предложена параметрическая модель длительности состояний [49]. Подход к оценке речи в рамках решаемой задачи на основе методов распознавания описан в [50, 51].

Также были рассмотрены методы, направленные на анализ и оценку характеристик голоса диктора (анализ речевого сигнала). Для анализа требуется обработка сигнала: описание физических параметров восприятия речи на основе некоторой модели с последующим преобразованием полученного представления в нужную форму [52]. Задачи анализа речевого

сигнала вытекают из целей обработки речевого сигнала и их специфики [53]. Оценка качества речи может представлять собой оценку соответствия вычисленных параметров некоторым заранее заданным величинам. Соответственно, ставится задача не только вычисления самих параметров, но и подбор значений, соответствующих качественному «эталонному» сигналу. Зачастую вычисление параметров и вычисление эталонов привязывается к фонетическому строю языка [54] и механизмам восприятия речи человеческим слухом [55].

Для повышение качества вычисления параметров голоса диктора с уменьшением влияния уровня шума предлагается использовать методы фильтрации, в том числе применение спектра модуляции [56], автоматическое выделение дефектных участков сигнала [57] и предварительный анализ информативности [58]. Также перед вычислением параметров зачастую требуется проведение сегментации сигнала на вокализованные участки. Так в [59] предложен подход к параметрическому описанию потока речи: описана последовательная обработка входного сигнала, включающая сегментацию речевого сигнала, выделение вокализованных участков и параметрическое описание. Метод сегментации сигнала на вокализованные и невокализованные участки с одновременной маскировкой описан в [60]. Иной подход к сегментации сигнала и выделения участков с гласными описан в [61].

На практике для вычисления параметров используются различные представления сигнала [62, 63], применяемый метод представления зависит не только от вычисляемых параметров, но и от необходимой скорости и точности обработки сигнала. Одним из вариантов вычисляемых параметров является вычисление формант [64]. В [65, 66] рассматриваются недостатки действующего метода формантного выравнивания, а также приводится описание новой методики выделения формант и получения выравненных фрагментов фонограмм. Вычисление формант для форсированной, а не спокойной речи предложено в [67]. В [68] описаны закономерности

изменения формантной картины в зависимости от вида речи, а в [69] закономерности распределения в зависимости от диктора.

Другой часто используемой характеристикой является частота основного тона (ЧОТ). Различные методы и подходы к вычислению ЧОТ, а также применение вычисленных значений в контексте оценки качества речи описаны в [70 – 73]. Проблема повышения эффективности на этапе предварительной обработки в алгоритмах оценки ЧОТ рассмотрена в [74], сделан вывод, что наиболее адаптивным методом декомпозиции речевых сигналов является метод КМДЭМ (метод комплементарной множественной декомпозиции на эмпирические моды). Подходы комплексной оценки с вычислением нескольких параметров описаны в [75, 76]. В работах [77, 78] предложены алгоритмы мгновенного вычисления гармонических параметров речевого сигнала и ЧОТ. Расчет значений мгновенной энергии и выбор размера окна, оптимального для решаемой задачи описан в [79].

В [80] приводится формализованный алгоритм цифровой обработки сигнала с линейно-частотной модуляцией частотно-временным преобразованием Вигнера для оценки разборчивости речи, включающий в себя обобщенный алгоритм оценки защищенности и принятых мер защиты объекта информатизации, алгоритм снижения порогового эффекта сигнала с линейной частотной модуляцией и алгоритм оценки разборчивости речи. Оценка вычисляемых параметров речевого сигнала при их обработке в нейронных сетях представлена в [81].

Вычисленные параметры сигналов могут использоваться не только в задачах оценки речи, но и в задачах идентификации и верификации диктора. Так в задаче верификации диктора в [82] рассмотрено применение мелкепстральных коэффициентов (MFCC), пар линейного спектра (LSP), кепстральных коэффициентов перцептивного линейного предсказания (PLP), кратковременной энергии, формантных частот, частот основного тона, вероятностей вокализации (voicing probability) и др. в системе GMM-

УВМ, а в [83] рассмотрено применение свёрточных глубоких сетей доверия в качестве средства извлечения речевых признаков.

Применение описанных выше методов является возможным как часть алгоритма оценки качества речи в рамках речевой реабилитации, однако их особенностью является работа именно с характеристиками голоса диктора, а не речи, что соответствует основной поставленной задаче.

1.3 Субъективные методы оценки качества речи и голоса

В [84] рассматриваются вопросы структуры речи и особенности ее восприятия, проводится анализ известных методов слуховых и инструментальных измерений.

Самая распространенная и используемая методика субъективной оценки качества описана в рекомендации МСЭ Р.800 и известна как методика MOS (Mean Opinion Score) [85]. По методике MOS оценка формируется как арифметическое среднее оценок, выставленных слушателями, где основными оценочными параметрами являются: разборчивость, естественность звучания голоса и уровень усилий слушателя.

Тест MOS проходит в два этапа: обучение и оценка сигналов. Все образцы оцениваются аудиторами по шкале от 1 до 5 по заранее известным критериям качества. Число аудиторов строго не определено и может варьироваться в зависимости от опыта аудиторов. Минимально рекомендуемое число неопытных слушателей - 20 человек, если аудиторы являются экспертами, то 10 человек. Продолжительность аудирования составляет не более 20 минут, также предъявляются требования по наличию наушников или требования к помещению, где проходит аудирование. Именно эта методика является основной для оценки качества моделирования объективных методов оценки речи [86]. Применение MOS для оценки качества речи при передаче по каналам связи, а также при

имитационном моделировании и разработке систем фильтрации речевого сигнала описаны в [87 – 92]. В [93] определены недостатки модели MOS в сетях IP.

У методики MOS есть вариация Degradation mean opinion score (DMOS) [94], в рамках которой производится оценка восприятия ухудшенного сигнала в сравнении с эталонным. Аудиторы сначала слушают исходный аудиосигнал, а затем выбирают, какой рейтинг присвоить искаженному сигналу. Оценка, как и по методике MOS, производится по 5-ти бальной шкале. Пример расчета по методике DMOS и сравнение результатов с PESQ описан в [95].

В [96] описан метод Diagnostic Acceptability Measure (DAM). По этому методу качество речи оценивает аудиторами по трем шкалам: параметрической, метаметрической и изометрической. Всего оценивается 16 характеристик качества речи. Шкала оценки от 0 до 100 баллов. Метаметрическая и изометрическая шкалы позволяют оценить речь с точки зрения «разборчивости», «приятности» и «приемлемости». Параметрическая шкала оценивает искажение сигнала и фона. Баллы, собранные по всем трем шкалам, обрабатываются, рассчитывается средневзвешенное значение отдельных шкал. Преимуществом относительно MOS является более высокая точность, но эта точность достигается за счет предъявления большого количества требований к аудиторам (экспертам) и их обучении.

Стандарт ITU-T (P 835) [97] содержит в себе методику экспертной оценки как искажения сигнала, так и фонового шума при оценке общего качества сигнала. Оценка проводится следующим образом. Оцениваемые аудиосигналы имеют длительность приблизительно четыре с: одна секунда фонового шума, две секунды речи с шумом, одна секунда фона. Затем дается время на оценку. Оценка проводится по 5-ти бальной шкале отдельно для фонового шума, речи с шумом и фона. Предъявляются требования к аудиторам, в том числе требования к численности (от 32 человек).

Метод Absolute Category Rating (ACR) [98] используется для получения абсолютных, субъективных оценок и представляет собой общий подход к оценке. В рекомендации определены три шкалы: качество речи, понимание смысла предложений, оценка громкости. Оцениваемый сигнал должен содержать простые коротких предложения, выбранные случайным образом, которые являются легкими по смыслу для понимания и не связаны между собой. Длительность сигнала 2–3 секунды. Подобный метод описан в [99], он основывается на экспертной оценке качества речи в разговорных ситуациях.

Если говорить про методы парного сравнения (эталонного и обработанного сигнала), то выделяются два метода Degradation Category Rating (DCR) и Comparison Category Rating (CCR) [100]. В первом методе сначала демонстрируется эталон, затем оцениваемый сигнал, и аудитор оценивает ухудшение качества по 5-ти бальной шкале. Во втором методе порядок демонстрации сигнала заранее неизвестен, и шкала оценок может быть перечислена как: намного лучше, лучше, немного лучше, примерно то же самое, немного хуже, хуже, намного хуже.

Недостатком методов с явным указанием на эталонный сигнал является вероятность искажения оценок. В [101] описан метод MUSHRA со скрытым эталоном и якорем (отфильтрованный эталонный сигнал). Цель скрытого эталона является отбрасывание оценок экспертов, которые оценили эталон слишком низко. Аудитор оценивает все аудиозаписи по шкале от 0 до 100.

Если говорить про сравнение результатов субъективных и объективных методов оценки речи, то по результатам эксперимента, описанного в [102], при передаче речи в сетях IP-телефонии предпочтительным методом является MOS.

Предложенная в [103] методика субъективной оценки речи позволяет дать оценку фонетических особенностей речи говорящего и может быть использована в качестве самостоятельного раздела при разработке

комплексной методики оценки параметров речевых сигналов и их модификации под воздействием различных факторов.

1.4 Оценка речи в медицине. Применяемая методика оценки речи в реабилитации на основе ГОСТ Р 50840–95

Крупным направлением исследования голоса в медицине является диагностика психоэмоционального состояния, в том числе как часть медицинской диагностики заболеваний. В [104] описан перцептивно-слуховой эксперимент, в котором группа аудиторов оценивала характеристики речи (темп, паузы, ритмический рисунок) в различных стрессовых ситуациях. Были выделены сферы применения подобного измерения характеристик. Подобная же методика может быть использована и для выявления психоэмоционального состояния «агрессия», в [105] описывается перцептивно-слуховой «паспорт» состояния, близкого к состоянию агрессии, на основе материалов для языков. Оригинальный способ повышения эффективности оценки психоэмоционального состояния предлагается в [106]. В его основе лежит улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом и мел-частотного кепстрального анализа. На основании полученных результатов сделан вывод, что предложенный авторами способ может успешно применяться в системах дистанционного мониторинга оценки психоэмоционального состояния. Подход по определению уровня эмоций на основе применения классификации транскрипций, полученных с помощью системы распознавания, описан в [107]. В [108] описаны методы, используемые для определения депрессии в каждой из модальностей и в совокупности, наиболее популярными методами моделирования и распознавания депрессии в исследовании являются нейронные сети. Оценка речи и голоса используется в [109] для создания голосового портрета

ребенка с различными типами развития. Также при помощи оценки речи и голоса предлагается оценивать общее состояние ребенка [110].

В [111] предложена модель голосообразования при наличии патологий органов речепродуцирования, в том числе с учетом болевого синдрома. Данная модель также дает возможность моделирования голоса с имитацией наличия болевых ощущений. Использование реализаций подсистемы распознавания речи и подсистемы распознавания жестов и команд для создания медицинского приложения дополненной реальности "Автоматического мобильного сурдопереводчика" предложено в [112]. В [113] предложен способ оценки патологических изменений в речевом сигнале для диагностики неврологических заболеваний: выявление отклонений в речи предлагается выполнять путем совместного анализа различных гласных звуков тестируемого сигнала. Использование программно-аппаратного комплекса спектрального анализа голоса в дополнение к магнитно-резонансной томографии для диагностики патологий голоса предложено в [114]. Подход к диагностике и реабилитации пациентов с дефектами голосо-речевой функции описан в [115], определение диагноза пациента путем сравнения частотно-амплитудных характеристик голоса пациента с нарушениями голосо-речевой функции с частотно-амплитудными характеристиками голоса контрольных пациентов из базы данных и с учетом зависимости диагноза контрольных пациентов от частотно-амплитудных характеристик их голоса. Подход и программное обеспечение на его основе для анализа речевых сигналов при неврологических патологиях, в частности, бульбарных нарушениях предложен в [116]. В основе подхода - методы цифровой обработки сигналов, адаптированные к задаче анализа речи в клинической медицине: выполняется расчет и построение спектрограммы, гистограммы, кепстрограммы, частоты основного тона для записанных речевых тестов.

В процессе обучения речи применяются подходы, суть которых заключается в сравнении речи с эталоном. Чем больше сходство речи с

эталонном, тем лучше считается произношение. Для реализации таких подходов была разработана система [117], использующая коэффициенты линейного предсказания для отражения формы речевого тракта. Схожей системой является ISTRА [118, 119], ее особенностью является распознавание изолированных слов с возможностью индивидуальной настройки на диктора.

Что касается непосредственно медицинской реабилитации, в том числе голосовой и речевой, после онкологии органов головы и шеи, подходы и методики описаны в [120 – 125]. Среди методов оценки речи при речевой реабилитации основным является метод экспертной оценки, при этом требований к количеству аудиторов не предъявляется, однако есть требования к самому аудитору – аудитором может быть логопед или логопед-онколог. При голосовой реабилитации методом оценки является расчет параметров голоса и сравнение их с заранее заданным эталонным значением [126]. Также показана эффективность использования метода биологической обратной связи и персонализации в комплексной реабилитации пациентов [127 – 129]. Если говорить про другие заболевания, то в речевой реабилитации также применяется экспертный метод оценивания речи и расчет характеристик голоса [130 – 131].

Основная методика, которая ранее использовалась при реабилитации пациентов в НИИ Онкологии г. Томска, основана на ГОСТ Р 50840–95 «Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости» [133]. Стандарт описывает получение оценки качества передачи речи, основанной на разборчивости и узнаваемости голоса диктора. В речевой реабилитации применяется оценка слоговой разборчивости. В приложении Б ГОСТ Р 50840–95 приведены таблицы слогов (1000 таблиц с 50 слогами в каждой), которые диктор должен прочитать для записи, а также тестовые фразы и таблицы фраз (Приложения Г и Д ГОСТ Р 50840-95 – 100 таблиц по 50 фраз в каждой). В рамках сеанса диктор зачитывает $5*k$ таблиц, где $k=1,2,3\dots$. Пятерки таблиц должны иметь

номера 1–5, 6–10, 11–15 и т.д. Зачитывание неполной пятерки таблиц по ГОСТ не допускается. Далее аудитор прослушивает запись и оценивает разборчивость произнесенных слогов, за каждый слог ставится оценка 1 или 0, где 0 – слышимый слог не совпадает с табличным, 1 – совпадает. Слоговой разборчивостью сеанса признается отношение суммы баллов, выставленных аудитором, к общему количеству записанных в рамках сеанса слогов. Оценки усредняются по всем аудиторам. Предъявляется требование к обучению и к количеству дикторов и аудиторов: не менее 5 дикторов и 15 аудиторов.

Используемая в речевой реабилитации методика оценки разборчивости речи (далее - старая методика) основана на ГОСТ Р 50840–95. Пациент произносит слоги согласно таблицам слогов, эксперт-логопед оценивает произношение значением 1 или 0, где 1 – слог произнесен разборчиво и совпадает с табличным, 0 - иначе. Отличия применяемой методики от методики в ГОСТ 50840–95 следующие:

- 1) диктор (пациент) зачитывает только первую или первые две таблицы слогов;
- 2) предварительного обучения у дикторов (пациентов) нет;
- 3) оценки выставляет один аудитор: эксперт – логопед.

Непосредственное применение полной методики согласно ГОСТ в рамках речевой реабилитации невозможно: нет возможности набора полноценной (от 5 человек) группы аудиторов. Даже если такая возможность существовала, состав группы аудиторов сохранялся бы постоянным длительное время, а это приводит к быстрому заучиванию списка слогов аудиторами и снижению объективности оценок. Время, затрачиваемое на оценку речи пациента в рамках одного сеанса, составляет примерно утроенное суммарное время звучания всех записанных слогов (здесь учитывается время, затрачиваемое на включение записи, прослушивание, выставление оценки), также это время может увеличиться

за счет того, что некоторые слоги приходится прослушивать несколько раз ввиду особенностей записи и/или речи пациента.

У применяемой методики оценки можно выделить ряд недостатков:

1) существенное влияние мнения эксперта: в этот недостаток можно включить как особенности слуха специалиста-логопеда, так и его знание записываемых слогов – при знании того, что должно быть произнесено, возможно искажение восприятия звуковой информации;

2) большое время обработки (относительно суммарной длительности всех записей);

3) трудоемкость учета изначальных особенностей речи пациента: изначальные особенности произношения фонем и сочетаний фонем могут искажать произнесенный слог, что может влиять на слоговую разборчивость;

4) список слогов для записи общий, нет конкретизации под различные локализации заболевания;

5) нет возможности построения динамики реабилитации для отдельных слогов, фонем, сочетаний фонем;

6) отсутствие возможности построение сеансов с использованием биологической обратной связи для формирования положительного подкрепляющего сигнала;

7) отсутствие возможности изменять набор слогов при коррекции процесса речевой реабилитации. Если, к примеру, изначально была проблема с несколькими фонемами, и в рамках реабилитации произношение одной их скорректировалось до предоперационного уровня, то логично было бы в дальнейшем уделять больше времени оставшимся проблемным фонемам и убрать из сеанса записи фонемы, работа с которыми уже закончена. Однако, это не отменяет возможности наличия контрольных сеансов с присутствием всех изначально проблемных фонем.

1.5 Выводы по главе

В результате анализа алгоритмов и методов анализа и оценки речи и голоса сделан вывод, что несмотря на обилие существующих решений, применение объективных способов из области связи, а также анализа и синтеза речи для решения рассматриваемой задачи является невозможным, так как в них заключен принципиально другой подход к оценке речевого сигнала, а также наличие вероятностных методов к распознаванию пропущенных фонем. Недостатками же субъективных методов именно в рамках решения поставленной задачи является нерешенный вопрос выбора эталона при сравнении с ним (сюда же можно отнести вопрос выполнения требования по персонализации оценки речи) и требования к количеству аудиторов.

Исходя из этого были сформулированы требования к разрабатываемым алгоритмам и методикам, а также программному комплексу на их основе:

- 1) необходимо проводить оценку речи в автоматизированном режиме;
- 2) необходимо учитывать индивидуальные особенности речи пациента;
- 3) необходимо рассчитывать такие оценки, чтобы была возможность отследить динамику восстановления речи;
- 4) оценку необходимо производить в режиме, близком к реальному времени, для реализации метода бионического принципа биологической обратной связи.

2 Анализ базы данных записей речи пациентов

Речевой (произносительный) аппарат человек состоит из большого количества органов, его схема представлена на рисунке 2.1 [134]. При этом органы можно разделить на две подгруппы: одни отвечают за формирование голоса (голосовые связки, трахея и др.), другие (язык, небо, зубы, губы и др.) за счет набора различных положений, в совокупности называемого артикуляционной базой, формируют отдельные разборчивые (понятные слушателю) минимальные фонетические единицы – фонемы. Во время хирургического этапа лечения рака органов полости рта в большинстве случаев вмешательство осуществляется только в органы второй подгруппы. Поэтому после операции голосовая функция практически не меняется, изменяется только речевая функция в рамках изменения произношения фонем, и это приводит к необходимости проведения именно речевой реабилитации и, соответственно, необходимости алгоритмов анализа качества произношения отдельных фонем.

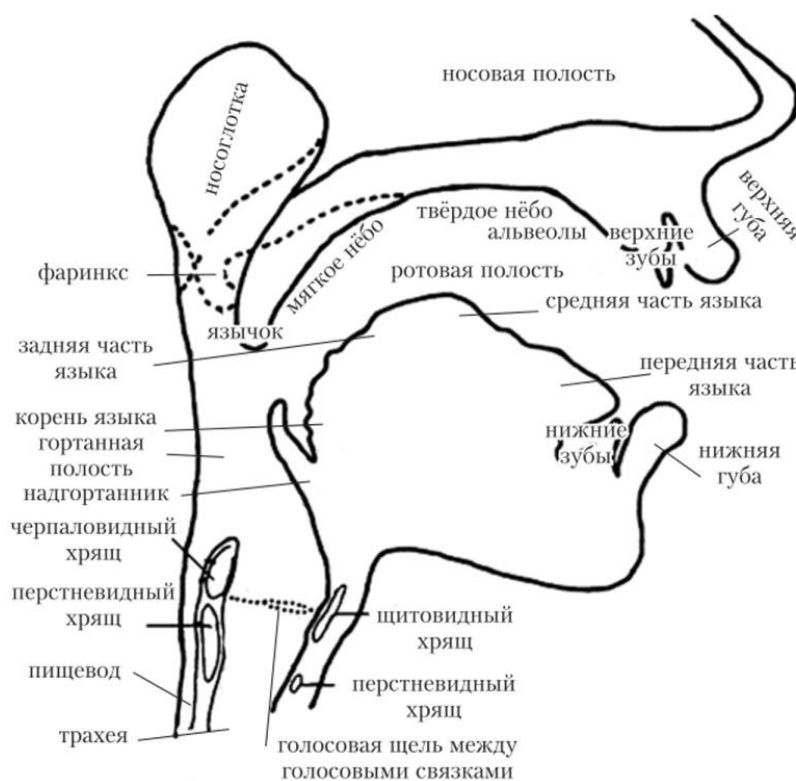


Рисунок 2.1 – Органы речевого аппарата.

Для оценки разборчивости речи в процессе речевой реабилитации производится запись произношения пациентом отдельных слогов, а непосредственно количественная оценка выставляется по методике оценки разборчивости речи (описана в разделе 1.4) на основе прослушивания сделанного аудиофайла. Схематичное представление процесса получения аудиофайлов отражено на рисунке 2.2.

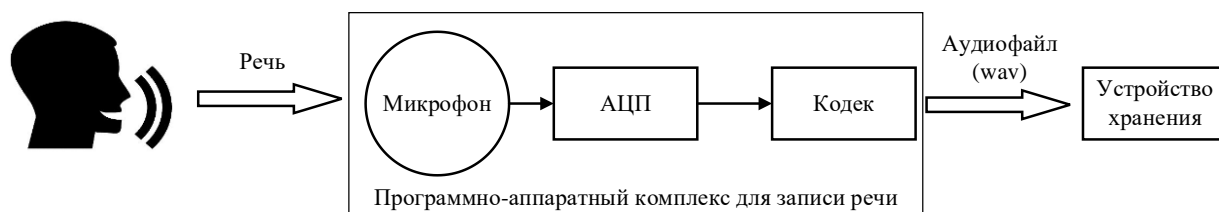


Рисунок 2.2 – Схема процесса записи аудиофайлов

Пациент произносит некоторый звук (в рамках решаемой задачи – слог), микрофон представляет считанные звуковые волны в виде электрического сигнала, который после прохождения АЦП (аналого-цифровой преобразователя) и кодека преобразовывается в последовательность значений, записанных в соответствии с необходимым форматом аудиофайла (в данном случае, wav). Какие конкретно записываются значения, зависит от используемого аудиоформата, а также характеристик аудиосигнала. В рамках решаемой научно-практической задачи работа осуществляется с уже готовыми аудиофайлами, поэтому технические моменты обработки сигнала в процессе всех преобразований не рассматриваются.

В процессе проведения сеанса оценки речи формируется набор аудиозаписей, каждая из которых содержит реализацию одного слога. Пример осциллограмм речевых сигналов с нормальным произношением и с искаженным произношением приведен на рисунках 2.3 и 2.4 соответственно.

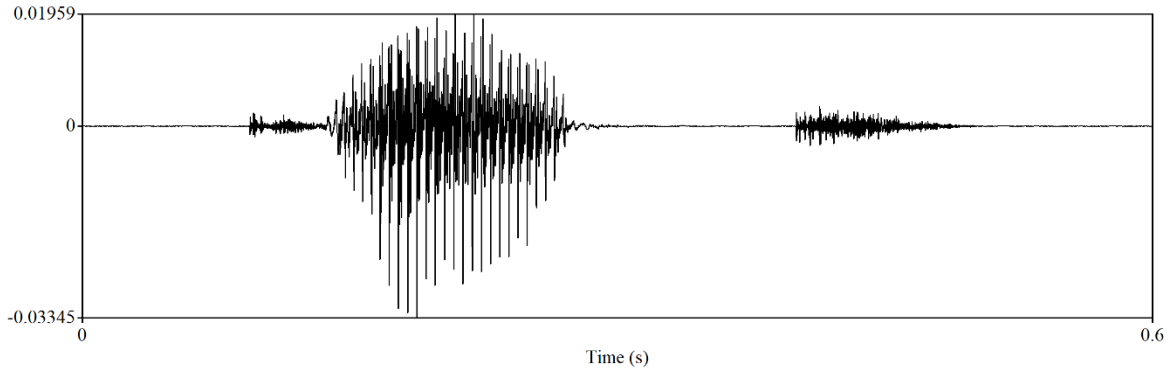


Рисунок 2.3 – Осциллограмма сигнала, содержащего произношение слога «кят» без искажений.

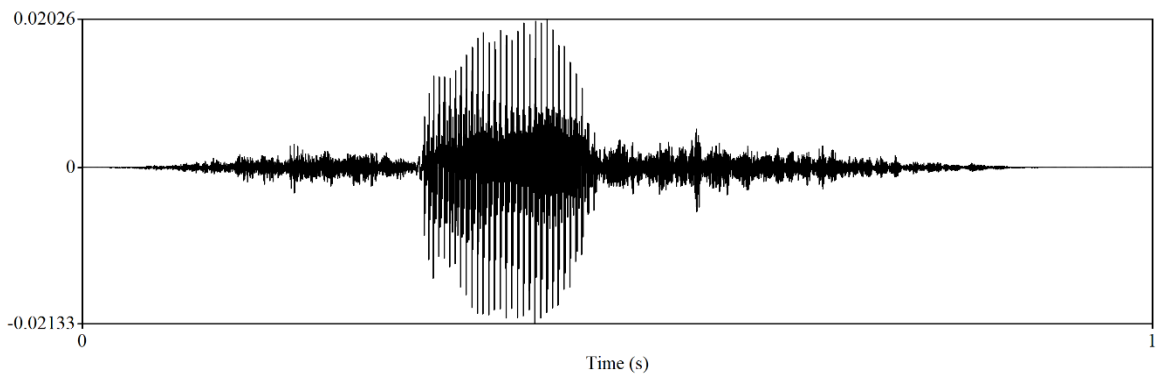


Рисунок 2.4 – Осциллограмма сигнала, содержащего произношение слога «кят» с искажением в произношении согласных.

На начальном этапе исследования уже существовала база данных с информацией о записях пациентов и набор самих аудиозаписей, а также программное обеспечение для работы с базой данных (просмотр и внесение информации о записях в базу данных), а также возможность внесения оценок записей, полученных с использованием адаптированной методики согласно ГОСТ Р 50840-95. Список слогов для записи был взят непосредственно из таблиц Приложения Б ГОСТа и не учитывал особенности решаемой практической задачи. Поэтому был проведен анализ записей речи пациентов для выявления закономерностей изменений в произношении фонем и слогов после проведения хирургического лечения, а также анализ особенностей отображения наиболее подверженных изменениям фонем на спектрограммах аудиосигналов.

2.1 Анализ записей и оценок качества речи на основе существующей методики

В имеющейся базе данных на момент начала исследований были собраны записи 15 пациентов с пятью различными видами рака с локализацией в ротовой полости. Чаще всего в базе данных представлен рак языка (6 больных, 13 тренировок). Нахождения статистически значимых оценок на небольшом объёме данных не представляется возможным, однако в процессе анализа были получены общие закономерности. Было проведено сравнения, как меняется речь пациента после проведения операции по сравнению с речью до операции.

На первом этапе анализа были выявлены буквы, которые пациенты произносили неправильно [135]. Подсчет ошибок для букв, а не для фонем, на данном этапе осуществлен из-за невозможности однозначности соотнесения «буква=фонема» при текущих условиях записи произношения слогов. Пациенту в процессе записи предлагался список слогов в текстовом виде, поэтому возникали разночтения в произношении предложенного слога. Если при работе с пациентом предлагать ему фонемную запись слова, это позволило бы снять проблемы, связанные с вариативностью прочтения слогов, однако не все пациенты могут корректно прочитать такую форму записи. Для решения этой проблемы предлагается добавить возможность дублировать записываемый слог его звукозаписью (однозначно определяющей его фонемный состав), это позволит избежать ошибок при чтении (за счет получения информации сразу по двум каналам – визуальному и акустическому), кроме того, позволит при обработке использовать именно фонемную запись, что позволит сразу выявлять проблемные фонемы. Для анализа были выбраны аудиозаписи 6 пациентов, соответствующих следующим критериям: есть запись сеанса до операции,

есть как минимум одна запись после операции, основной диагноз – рак языка.

Под ошибкой в произношении слогов с буквами Ъ и Ь подразумевается изменение в мягкости стоящих перед ними букв: в случае буквы Ъ – потеря мягкости в произношении, в случае буквы Ь – добавление мягкости.

На втором этапе было подсчитано количество ошибок для каждой из букв, соответствующих проблемным фонемам. Такие буквы были проранжированы в рамках одного сеанса одного пациента, где наибольший ранг у букв, в произношении которых чаще всего допускались ошибки.

Далее ранги были усреднены и буквы переранжированы по средним значениям. В результаты буквам, которые чаще всего произносились с ошибками, присвоен максимальный ранг, буквам, которые реже всего изменялись, присвоен минимальный ранг. Результаты оценки представлены в Таблице 2.1, содержащей рейтинг по убыванию рангов.

Таблица 2.1 – Результаты ранжирования букв по количеству ошибок в произношении.

id пациента/ Буква	16	15	3	19	11	13	Средний ранг	Итоговый ранг
р	32,5	30	33	30,5	32	33	31,8	33
т	25	32	29	32	33	30	30,2	32
с	32,5	33	28	27,5	23,5	25,5	28,3	31
ь	30,5	30	31	33	15,5	29	28,2	30
ф	28	30	23	24,5	23,5	31,5	26,8	29
ш	28	26	14,5	24,5	23,5	31,5	24,7	28
е	26	20	20,5	30,5	29	18,5	24,1	27
д	23,5	28	25	27,5	29	7,5	23,4	26
и	23,5	20	23	19,5	29	7,5	20,4	25
н	10,5	15	26,5	22	29	18,5	20,3	24
ы	17,5	26	6,5	21	29	18,5	19,8	23
л	30,5	20	6,5	29	5,5	25,5	19,5	22
к	28	20	32	7,5	5,5	18,5	18,6	21
ц	17,5	23,5	30	24,5	5,5	7,5	18,1	20
в	10,5	9,5	26,5	24,5	15,5	18,5	17,5	19
ж	17,5	15	11,5	15	15,5	25,5	16,7	18
м	17,5	15	23	2	15,5	25,5	16,4	17
г	21,5	9,5	17,5	7,5	15,5	25,5	16,2	16
ч	10,5	9,5	17,5	15	23,5	18,5	15,8	14,5

Продолжение таблицы 2.1.

щ	17,5	3,5	17,5	15	15,5	25,5	15,8	14,5
о	21,5	15	11,5	7,5	23,5	7,5	14,4	13
х	10,5	23,5	17,5	19,5	5,5	7,5	14,0	12
з	17,5	26	9,5	15	5,5	7,5	13,5	11
ё	3,5	15	14,5	15	15,5	7,5	11,8	10
я	3,5	20	6,5	15	15,5	7,5	11,3	9
ю	10,5	3,5	13	7,5	23,5	7,5	10,9	8
й	3,5	9,5	20,5	15	5,5	7,5	10,3	7
у	10,5	3,5	2,5	7,5	15,5	18,5	9,7	6
э	10,5	9,5	6,5	7,5	15,5	7,5	9,5	5
а	10,5	3,5	2,5	7,5	5,5	18,5	8,0	4
п	3,5	9,5	9,5	7,5	5,5	7,5	7,2	3
б	3,5	3,5	2,5	2	5,5	7,5	4,1	1,5
ь	3,5	3,5	2,5	2	5,5	7,5	4,1	1,5

Для оценки согласованности ранжирования букв по пациентам был подсчитан коэффициент конкордации [136] $W = 0,577$, значение коэффициента является значимым на уровне значимости 0,05: наблюдаемое значение 110,8 против критического 43,8. Принимается гипотеза о согласованности оценок.

Из данных таблицы видно, что наиболее часто подвержены изменениям фонемы, связанные со следующими 10 буквам (в порядке убывания ранга): Р, Т, С, Ь, Ф, Ш, Е, Д, И, Н.

Чтобы понять разницу между рангами была получена аналогичная таблица, но по частоте ошибок. В таблице указана частота ошибок для каждого пациента, среднее количество ошибок для каждой из букв. Далее буквы были проранжированы по убыванию среднего количества ошибок, чем больше ошибок в произношении фонемы, характерной для буквы, отмечено, тем больше у нее ранг. Результат показан в Таблице 2.2. Для такого метода ранжирования список букв выглядит как: Р, Т, С, Ь, Ф, К, Ш, Д, Е, Ц.

Таблица 2.2 – Результаты ранжирования букв по частоте ошибок в произношении.

id пациента/ Буква	16	15	3	19	11	13	Среднее количество ошибок	Итоговый ранг
р	10	10	25	11	7	11	12,3	33
т	5	16	15	12	8	4	10	32
с	10	19	13	8	2	2	9	31
ь	8	10	17	14	1	3	8,8	30
ф	7	10	9	6	2	7	6,8	29
к	7	3	23	1	0	1	5,8	28
ш	7	5	5	6	2	7	5,3	27
д	4	6	10	8	3	0	5,2	25,5
е	6	3	7	11	3	1	5,2	25,5
ц	2	4	16	6	0	0	4,7	24
л	8	3	1	9	0	2	3,8	22,5
н	1	2	11	5	3	1	3,8	22,5
и	4	3	9	3	3	0	3,7	21
в	1	1	11	6	1	1	3,5	20
м	2	2	9	0	1	2	2,7	18,5
ы	2	5	1	4	3	1	2,7	18,5
г	3	1	6	1	1	2	2,3	16,5
х	1	4	6	3	0	0	2,3	16,5
ч	1	1	6	2	2	1	2,2	14,5
щ	2	0	6	2	1	2	2,2	14,5
ж	2	2	3	2	1	2	2	13
з	2	5	2	2	0	0	1,8	11,5
о	3	2	3	1	2	0	1,8	11,5
ё	0	2	5	2	1	0	1,7	9,5
й	0	1	7	2	0	0	1,7	9,5
ю	1	0	4	1	2	0	1,3	8
я	0	3	1	2	1	0	1,2	7
э	1	1	1	1	1	0	0,8	6
п	0	1	2	1	0	0	0,7	4,5
у	1	0	0	1	1	1	0,7	4,5
а	1	0	0	1	0	1	0,5	3
б	0	0	0	0	0	0	0	1,5
ъ	0	0	0	0	0	0	0	1,5

Различия в списках могут быть объяснены разным количеством неправильно произнесенных фонем у разных пациентов, в результате в способе с использованием частот, усредняющих пациента с высоким уровнем ошибки вносят самый весомый вклад в итоговый результат.

На основании анализа полученного списка буква «Ь» исключается из итогового списка, потому что эта буква описывает не саму фонему, а ее особенность – мягкость. Наличие этой буквы в списке указывает на необходимость включения в итоговый список проблемных фонем как твердого, так и мягкого варианта каждой из них.

Полученные данные согласуются с данными, представленными в [137], со ссылкой на более ранние работы коллектива авторов (И. Болов, М. Соловьев, Л. Душак, Д. Подгорных, А. Шендеров, 1974). В рамках работы представлены результаты исследования группы больных с различными онкологическими заболеваниями полости рта после оперативного вмешательства. Установлено, что наибольшее количество искажений наблюдается при артикуляции зубных согласных (с, з, ц, т, д, н, л) и альвеолярных согласных звуков (р). В меньшей степени искажают передние (ш, ж, ч, щ) и небные согласные (к, г, х). Однако частота искажений небных согласных резко увеличивается при увеличении объема оперативного вмешательства (резекции части языка). Часто в артикуляции вышеуказанных согласных изменялся признак твердости. Аналогичные данные, но для другого диагноза, можно найти в [138]. Результаты, полученные в рамках этих исследований, согласуются с представленными результатами по списку наиболее склонных к изменению согласных. Кроме того, подтвердились явные изменения свойств твердости, выразившиеся в значительном количестве ошибок, связанных с буквами «Ь» (прямая замена твердости) и «Е» (по сравнению с «Э» — по существу замена твердости на предшествующий согласный).

Имеющиеся некоторые различия в частоте ранжирования можно объяснить тем, что другие исследования проводились на нескольких локализациях заболевания полости рта, в данном исследовании представлены данные только для рака языка. Кроме того, могут быть различия в способах подсчета, что может привести к замене некоторых записей в таблице наиболее склонными к изменению фонемами. Однако

отсутствие прямого противоречия с представленной работой позволяет говорить о достоверности полученных данных.

Согласно классификации согласных звуков [139] чаще всего подвергаются изменениям в произношении язычные согласные, в особенности переднеязычные – [т], [с], [ш]. Исходя из практических задач и реальных данных совместно со специалистами из НИИ Онкологии для дальнейшей работы в рамках построения алгоритмов и методик для оценок качества речи список проблемных фонем был скорректирован [140, 141]. Были выделены фонемы [к], [с], [т] и их мягких вариантов - [к´], [с´], [т´]. Включение фонемы [к] и ее мягкого варианта обусловлено необходимостью добавление в список одной из заднеязычных фонем, частота изменения которых возрастает с увеличением объёма хирургического вмешательства [142]. Существенное расширения списка для обработки всех подверженных изменениям фонем невозможно из-за наличия особенностей работы с пациентом в пред- и постоперационный периоды (медицинские и организационные моменты). Исключение фонемы [р] из списка обоснованно почти полным отсутствием возможности восстановления корректного произношения данной фонемы из-за особенностей методов лечения онкологии органов речеобразующего аппарата [143, 144].

На основе списка фонем был составлен список слогов для записи в рамках проведения сеансов оценка качества речи. Список слогов для записи представлен в Таблице 2.3. Для каждой из проблемных фонем в списке представлено по 15 слогов: 5 слогов, в которых проблемная фонема расположена в начале слога, 5 слогов, в которых проблемная фонема распложена в середине слога, и 5 слогов, в которых проблемная фонема распложена в конце слога.

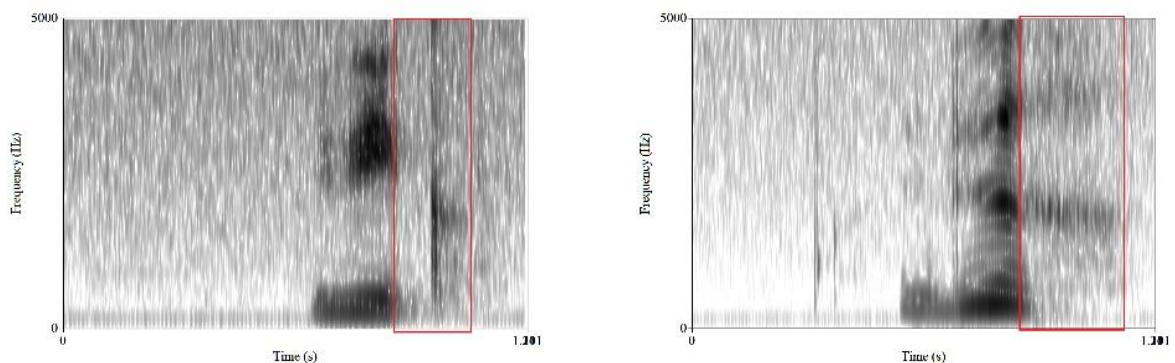
Таблица 2.3 – Таблица слогов для записи «ОнкоЯзык», сформированная на основе списка наиболее проблемных фонем.

[К]	[К']	[С]	[С']	[Т]	[Т']
Кась	Кясь	Сось	Сич	Тышь	Тют
Крус	Кюй	Свум	Сёсь	Трит	Тёр
Куп	Кясть	Стял	Сюд	Торь	Тить
Касть	Кёт	Сыч	Седь	Тэф	Тюм
Кроп	Кят	Сул	Сих	Трух	Тянь
Ркат	Скял	Фсэп	Фсен	Чтап	Стял
Докт	Шкюн	Псун	Ксет	Штай	Штяй
Скат	Фкён	Дсал	Псюн	Стыч	Стич
Фкон	Скюн	Чсэб	Гсех	Стар	Стяр
Шкун	Шкет	Рсат	Деял	Чтуп	Чтех
Няк	Викь	Дыс	Плисьь	Бят	Пость
Ёк	Някь	Гыс	Нысь	Пут	Чать
Фек	Фекь	Здес	Сось	Тют	Шеть
Вик	Плакь	Ныс	Шись	Трат	Мять
Плык	Фокь	Гос	Кузь	Фёт	Луть

2.2 Анализ спектрограмм аудиозаписей слогов, содержащих наиболее подверженные изменениям фонемы

После формирования набора наиболее проблемных фонем и списка слогов для записи на основе этого набора, была произведена запись аудиосигналов согласно составленному списку. Запись проводилась диктором женского пола, было записано два набора слогов: без изменений (с нормальным произношением проблемных фонем) и с изменениями. Изменения были искусственно сымитированы в соответствии с наиболее характерными для постоперационного периода особенностями произношения. Задачи анализа всех возможных вариантов искажения каждой из проблемных фонем не стояло. Для каждой аудиозаписи были построены спектрограммы: спектрограмма Фурье, спектрограмма по коэффициентам линейного предсказания (далее – КЛП) и автокорреляция. Также был произведен аналогичный анализ спектрограмм аудиозаписей слогов из базы данных записей пациентов НИИ Онкологии г. Томска.

В результате анализа были выявлены следующие особенности, одинаковые как для фонемы [к], так и для [к']: вне зависимости от расположения фонемы в слове после операции время непосредственного произнесения фонемы увеличивает, пропадает характерная для взрывных согласных пауза перед взрывом (фонемы ([к] и [к'] относятся к взрывным согласным звукам [139]), основные частоты смещаются выше (примерно с 1,3 кГц до 1,5 кГц), в верхней части спектра (от 2,5 кГц до 5кГц) звук становится более заметным и более выраженным. Спектрограммы Фурье аудиосигналов, содержащих слог «Вик» с наличием и отсутствием искажений в произношении, представлены на рисунке 2.5 (красным выделен фрагмент, отражающий проблемную фонему). Спектрограммы для аудиозаписей других слогов с фонемами [к] и [к'] приведены в приложении Д на рисунках Д.1 – Д.3.



Произношение фонемы без искажений

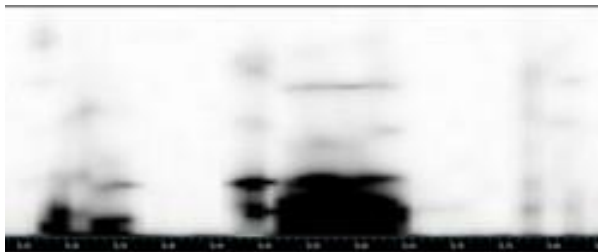
Произношение фонемы с искажением

Рисунок 2.5 – Сравнение спектрограмм Фурье аудиозаписей слога «Вик».

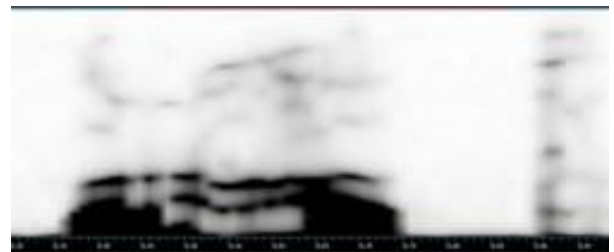
В [145] проведен анализ спектрограмм Фурье аудиозаписей, содержащих слоги с проблемными фонемами [т] и [т']: основные изменения схожи с фонемами [к] и [к']: полностью или частично пропадает пауза перед взрывом, уменьшается интенсивность произношения, при нахождении проблемной фонемы перед согласной спектр фонемы становится схож со спектром впереди стоящей согласной.

Дальнейший анализ был проведён на синтетически сформированном наборе данных. Для фонемы [к] изменения в спектрограммах КЛП

отличаются для различных положений фонемы в слоге [146]. На рисунке 2.6. а также в приложении Д на рисунке Д.4 представлены спектрограммы КЛП аудиозаписей слогов, содержащих фонему [к]. Здесь и далее по разделу на спектрограммах Фурье, КЛП и коррелограммах по оси ординат – частота в Гц, по оси абсцисс – время в секундах. Основные изменения, которые можно увидеть на спектрограмме КЛП, заключаются в увеличении длительности произношения фонемы и отсутствии ярко выраженного ударного момента (четкого разделения паузы перед взрывной фонемой и самого взрыва). Основная частота в большинстве случаев не изменяется. При всех трех положениях фонемы [к] присутствует сужение диапазона частот, явно это заметно при нахождении фонемы в середине и начале слога. У фонем, находящихся в середине или конце слога, начало участка спектрограммы, который отражает фонему, расположено сразу после окончания предыдущей фонемы, отсутствует характерная пауза перед фонемой. В некоторых случаях наблюдается уменьшение длительности гласного после изменённой фонемы.



Произношение без искажений



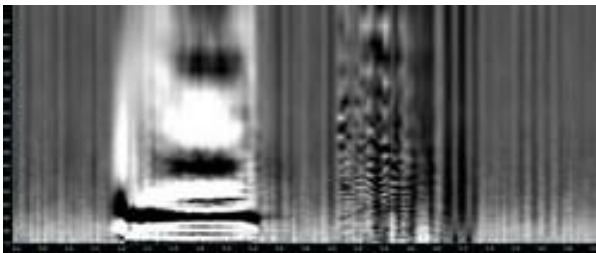
Произношение с искажениями

Рисунок 2.6 - Спектрограммы КЛП аудиозаписей слога, содержащего фонему [к] в середине слога.

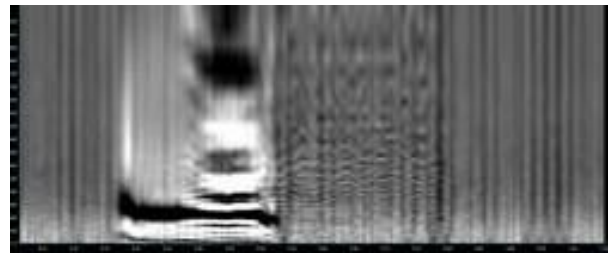
На рисунке Д.5 приложения Д изображены автокорреляции записей слогов. На них заметно увеличение длительности фонемы при искаженном произношении (что наблюдалось и на спектрограммах КЛП), также видно увеличение количества «полос» изображения автокорреляции фонемы при ее расположении в середине и начале слога.

Для фонемы [к'] на спектрограммах КЛП (рисунок Д.6 приложения Д) схожие с фонемой [к] изменения. Это позволяет дальнейшем при составлении модели объединить аудиозаписи с указанными фонемами в

одному группе для подбора параметров оценки. На рисунке 2.7 и в приложении Д на рисунке Д.7 представлены коррелограммы записей слогов с проблемной фонемой [к']. При изменениях в произношении характерно увеличение длительности измененной фонемы. Основная часть фонемы находится в более низких частотах относительно произношения без искажений. Также, как у фонемы [к], нет паузы между звуками, начало отрезка, соответствующего фонеме [к'], совпадает с концом отрезка, соответствующего предшествующей ей фонеме.



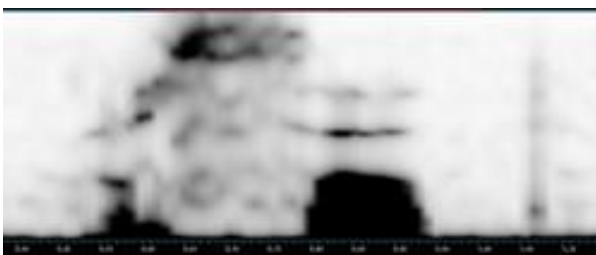
Произношение без искажений



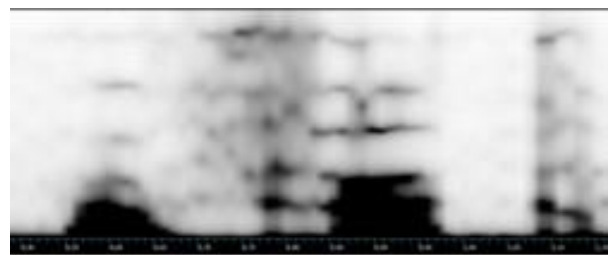
Произношение с искажениями

Рисунок 2.7 - Коррелограммы аудиозаписей слога, содержащего фонему [к'] в конце слога.

У фонем [с] и [с'] наблюдаются одинаковые изменения, их можно объединить в одну группу для анализа и дальнейшего изучения. На спектрограммах КЛП (рисунок 2.8, рисунки Д.8 и Д.9 приложения Д) заметно уменьшение интенсивности отрезка сигнала, содержащего проблемную фонему, в некоторых случаях отображение фонемы на спектрограмме становится неразличимым. Длительность фонемы не изменяется. На большинстве спектрограмм сигналов с искаженным произношением начали проявляться более низкие частоты, основная частота также снижена.



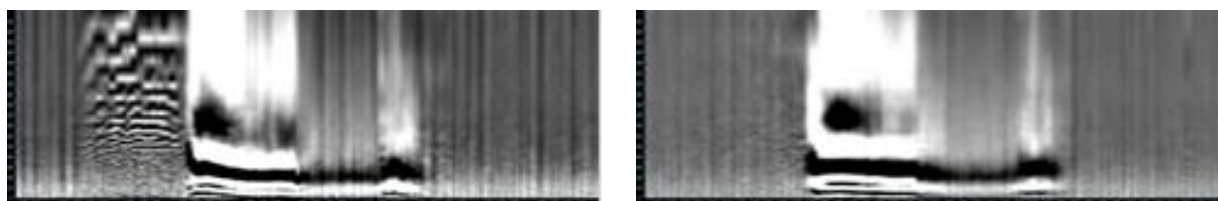
Произношение без искажений



Произношение с искажениями

Рисунок 2.8 - Спектрограммы КЛП аудиозаписей слога, содержащего фонему [с] в середине слога.

На коррелограммах аудиозаписей с проблемными фонемами [с] и [с'] (рисунок 2.9, рисунки Д.10 и Д.11 приложения Д) в некоторых случаях наблюдается увеличение интенсивности гласной после изменяющейся фонемы. Отмечено увеличение частоты «ряби» (отражением смены знака с положительного на отрицательный), характеризующей фонему на изображении автокорреляционной функции, а также уменьшение основной частоты, перемещение спектра в более низкий диапазон частот.

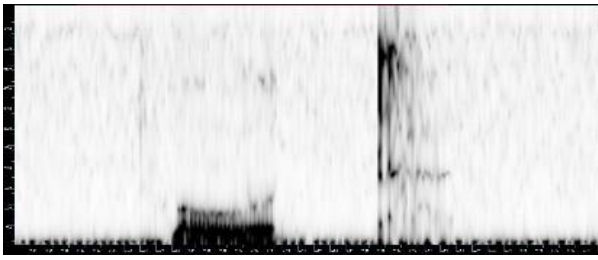


Произношение без искажений

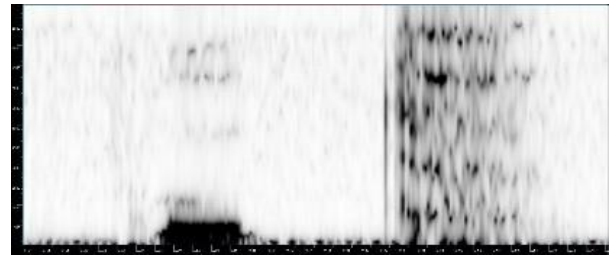
Произношение с искажениями

Рисунок 2.9 - Коррелограммы аудиозаписей слога, содержащего фонему [с'] в начале слога.

Для аудиозаписей с проблемными фонемами [т] и [т'] также были проанализированы спектрограммы КЛП и коррелограммы [147]. У фонемы [т] более выражены различия в изменениях в спектрограммах КЛП в зависимости от расположения фонемы в той или иной части слога. Если фонема расположена в начале и середине слога (рисунок Д.12, 1а-1б, 2а-2б), можно говорить об уменьшении ударного момента, характерного для фонемы [т]. Наблюдается уменьшение интенсивности и незначительное увеличение длительности. При расположении фонемы в конце слога (рисунок 2.10) появляются более низкие частоты. Увеличение длительности также характерно, но изменяется лишь момент окончания, момент начала произношения фонемы остается неизменным. Изменения в коррелограммах (рисунок Д.13) менее заметны по сравнению с другими проблемными фонемами. Наблюдается незначительное увеличение длительности и незначительное увеличение интенсивности средних частот при расположении фонемы в конце слога. Основная «рябь», характеризующая эту согласную, перемещается в нижнюю или среднюю полосу частот.



Произношение без искажений



Произношение с искажениями

Рисунок 2.10 - Спектрограммы КЛП аудиозаписей слога, содержащего фонему [т] в конце слога.

Анализ спектрограмм аудиозаписей слогов с проблемной фонемой [т'] (Рисунок Д.14) показал, что если фонема расположена в середине или начале слога, на изображениях с изменениями практически незаметны участки, характеризующие фонему: присутствуют лишь незначительные или вовсе отсутствуют признаки наличия в данном отрезке какой-либо фонемы. При расположении фонемы в конце слога, наблюдается уменьшение интенсивности, появление в звуке более низких частот, незначительное увеличение по длительности. На коррелограммах, построенных для аудиозаписей с фонемой [т'], (рисунок 2.11, рисунок Д.15 приложения Д) можно отметить те же изменения, что и для фонемы [т]. Это позволяет при дальнейшем анализе объединить фонемы [т] и [т'] в одну группу для изучения.



Произношение без искажений



Произношение с искажениями

Рисунок 2.11 - Коррелограммы аудиозаписей слога, содержащего фонему [т'] в середине слога.

2.3 Выводы по главе

Были проанализированы изменения в речи пациентов после проведения хирургического этапа комбинированного лечения онкологии органов полости рта. На основе статистических данных был сформирован список проблемных фонем, которые чаще всего подвергаются изменениям в постоперационной речи. На основе списка проблемных фонем предложен список слогов для записи в рамках проведения сеансов оценки произношения слогов. Был проведен визуальный анализ спектрограмм Фурье и КЛП, а также коррелограмм аудиозаписей слогов, содержащих проблемные фонемы. Анализ показал, что в дальнейшем исследовании допустимо объединение проблемных фонем в группы ввиду схожести постоперационных изменений в произношении.

3 Методика и алгоритм оценки речевых сигналов

3.1 Алгоритм нахождения количественного значения схожести двух речевых сигналов

Особенностью решаемой задачи является то, что необходимо найти схожесть двух временных последовательностей разной длины, то есть на вход алгоритму подаются два звуковых сигнала с различными продолжительностями. Шаблон алгоритма нахождения оценки может быть представлен следующей последовательностью шагов. Для каждого шага перечислены параметры, которые необходимо конкретизировать для составления алгоритма нахождения количественного значения схожести двух временных рядов.

Шаг 1. Преобразование аудиосигнала (аналоговое представление, оциллограмма преобразование Фурье, Вейвлет преобразование и т.д.) выделение фрагмента последовательности для дальнейшего анализа. Этот шаг выполняется для обеих аудиозаписей, для выполнения шага необходимо определить следующие параметры:

- вид преобразования сигнала: аналоговое представление, оциллограмма преобразование Фурье, Вейвлет преобразование и т.д.;
- фрагмент для анализа: анализировать весь слог или только проблемную фонему, выделенную с использованием алгоритмов сегментации.

Шаг 2. Приведение последовательностей к одной длине. Для выполнения шага необходимо определить следующие параметры:

- алгоритм приведения последовательности к одной длине – временная нормализация;
- обоснование необходимости или ее отсутствия в предварительной нормализации последовательностей: нормализация по интенсивности сигнала, сглаживание сигнала.

Шаг 3. Нахождение количественного значения меры схожести двух трансформированных временных рядов. Для выполнения шага необходимо определить следующие параметры:

– метрика или их комбинация (гибридная мера) для нахождения количественного значения.

Так как одно из требований к разрабатываемым алгоритму и методикам, а соответственно и программному комплексу на их основе, это необходимость в формировании подкреплений в рамках биологической обратной связи в режиме реального времени, то преимущество отдается более простым для реализации вариантам параметров при сопоставимости получаемых результатов. И одним из критериев выбора является время работа алгоритма.

3.1.1 Выбор алгоритмов временной нормализации и подсчета количественной оценки

При сравнении двух объектов обязательным условием является сопоставимость этих объектов, а в случае со звуковыми файлами, которые конвертируются в числовые массивы путем считывания амплитуд, необходимым условием является одинаковые длины. Соответственно, для сопоставления необходимо трансформировать последовательности к одной длине или использовать алгоритмы выбора сопоставимых элементов из каждого ряда. В контексте обработки дискретных данных наиболее эффективным считается алгоритм динамической трансформации временной шкалы – *dynamic time warping*, далее DTW-алгоритм, который специализируется на поиске оптимального соответствия временных последовательностей, к которым и относятся звуковые файлы [148, 149, 150]. Преимуществами этого алгоритма является простота его реализации, а также линейная сложность, что является важным параметром из-за

необходимости оценки в режиме реального времени для организации биологической обратной связи.

DTW-алгоритм основан на построении матриц расстояний между всеми элементами первой последовательности и всеми элементами второй последовательности (матрица d размера $m*n$, где $d_{i,j}$ – расстояние между точками t_i и q_i последовательностей T и Q соответственно) и построении матрицы деформаций на основе матрицы расстояний (матрица D размера $m*n$, где $D_{i,j}$ определяется по формуле 3.1). Расстояние $d_{i,j}$ (стоимость шага) в DTW-алгоритме принято считать на основе формул 3.2 или 3.3 [148]. Так как значения амплитуд звуковых файлов малы (среднее значение равно $1*10^{-5}$), то использование квадрата уменьшает абсолютную величину итогового расстояния, поэтому при реализации алгоритма будет использоваться расстояние городских кварталов (3.3).

$$D_{i,j} = d_{i,j} + \min(D_{i-1,j}, D_{i,j-1}, D_{i-1,j-1}) \quad (3.1)$$

$$d(t_k, q_h) = (t_k - q_h)^2, \quad (3.2)$$

$$d(t_k, q_h) = |t_k - q_h|, \quad (3.3)$$

где t_k и q_h точки последовательностей T и Q соответственно.

Последнее значение матрицы деформации - стоимость пути, а трансформированные последовательности получаются путем выбора элементов каждой из последовательности согласно индексам элементов, принадлежащих оптимальному пути.

Также преимуществом данного алгоритма является то, что в результате выполнения уже получается количественная оценка – DTW-расстояние, которое представляет собой стоимость оптимального пути между двумя последовательностями. Либо же количественная оценка произношения слога получается путем сравнения трансформированных последовательностей на основе различных мер различия. Для определения способа подсчета оценки было произведено сравнение результатов, полученных путем подсчета DTW-расстояния, коэффициента корреляции (3.4) и евклидова расстояния (3.5) [151].

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3.4)$$

где r_{XY} – коэффициент корреляции,

\bar{x} и \bar{y} – выборочные средние числовых последовательностей

$X = (x_1, \dots, x_n)$ и $Y = (y_1, \dots, y_n)$ соответственно,

n – длина последовательностей.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (3.5)$$

Для этого на основе списка слогов с наиболее подверженными изменениям фонемами было произведено 3 сеанса записи: первый и второй сеансы с нормальным произношением слогов, и третий сеанс с произношением слогов с имитацией характерных для постоперационного периода изменений (искаженное произношение). Аудиозаписи были отредактированы таким образом, чтобы периоды тишины до и после произношения были минимальные. На вход алгоритма подсчета подавались спектрограммы Фурье полученных записей (весовая функция – функция Хэмминга, размер окна 512, размер окна перекрытия 511, размерность выборки быстрого преобразования Фурье – 511). Были подсчитаны меры различия для пар слогов норма1 - норма2, норма1 - искажение, норма2 - искажение, где норма1 – запись слога первого сеанса, норма2 – запись этого же слога из второго сеанса, искажение - запись этого же слога из третьего сеанса. Если говорить о DTW-расстоянии и евклидовом расстоянии, количественная оценка пары норма1-норма2 должно быть меньше, чем оценка норма1-искажение и норма2-искажение, что можно обозначить как критерий соответствия для данных метрик. Для коэффициента корреляции рассматривалась близость коэффициента к 1, соответственно коэффициент для пары норма1-норма2 должен быть ближе к 1, чем коэффициент норма1-искажение и коэффициент норма2-искажение, что является критерием соответствия для метрики коэффициент корреляции. В таблице 3.1 приведены результаты работы алгоритма [152], указаны средние значения оценок для пар из исследуемых сеансов, значение отношений указанных

средних значений. Также рассчитано количество ошибок для каждой меры различия (суммарное количество ошибок, количество ошибок для пар норма1-искажение, количество ошибок для пар норма2-искажение), слоги, в которых были подсчитаны неверные оценки, время работы алгоритма (время подсчета оценки для 1 пары слогов, среднее время на одну пару слогов при подсчете оценок для трех сеансов).

Таблица 3.1 – Сравнение результаты работы алгоритмов оценки на основе различных мер различия.

		Мера различия		
		DTW- расстояние	Евклидово расстояние	Коэффициент корреляции
Среднее значение оценок, полученных для пар (запись из сеанса норма1) – (запись из сеанса норма2)		1,405	0,137	0,401
Среднее значение оценок, полученных для пар (запись из сеанса норма1 или норма 2) – (запись из сеанса искажение)		1,864	0,178	0,389
Отношение средних оценок		0,754	0,772	0,97
Соблюдение критерия соответствия		+	+	+
Количество ошибок	норма1-искажение	32	40	52
	норма2-искажение	2	16	46
	Одновременно в обеих парах	1	12	33
Время выполнения (сек на 1 оценку)		3,54	3,3	5,32

Каждая из метрик соответствуют обозначенному для нее критерию. Наименьшим количество ошибок обладает алгоритм, итоговой оценкой которого является DTW-расстояние. Чем меньше DTW-расстояние, тем качественнее (ближе к эталону) признается произнесенный слог.

Приведенное время, затрачиваемое на проведение одной оценки, не позволяет производить оценку в режиме реального времени, однако это вопрос оптимизации алгоритма и программного кода, а также вычислительных мощностей оборудования. Так как нет преимущества по времени у какой-либо из метрик, то все они могут быть в дальнейшем рассмотрены в качестве составных частей гибридной меры различия. Можно обозначить меру DTW-расстояния как основную меру для расчета итоговой количественной оценки из-за наименьшего числа ошибок для этой метрики.

Также была рассмотрена возможность нормализации или сглаживания временной последовательности перед применением DTW-алгоритма. Были сравнены результаты, полученные на основе последовательностей без нормализации и сглаживания, последовательностей с нормализацией по интенсивности (приведение интенсивности звука к константе), сглаженных на основе простого скользящего среднего (согласно 3.6) последовательностях, а также нормализованных по интенсивности сглаженных на основе простого скользящего среднего последовательностям.

$$\bar{y}(k) = \frac{1}{n} \sum_{t=k-\frac{n-1}{2}}^{t=k+\frac{n-1}{2}} y(t), \quad (3.6)$$

где n – размер окна сглаживания (обязательно нечетное число),

$y(t)$ - t -ый элемент временного ряда,

$\bar{y}(k)$ - k -ый элемент сглаженного временного ряда.

Результаты приведены в таблице 3.2: подсчитаны средние значения оценок, значение отношения средних значений, соответствие критерию, количество ошибок, среднее время выполнения одного расчета. Алгоритм подсчета средних значений и отношений и критерии соответствия для каждой из метрик аналогичны таблице 3.1.

Таблица 3.2 – Сравнение результатов работы алгоритм DTW с нормализацией и сглаживанием числовых последовательностей.

		Алгоритм			
		DTW	DTW с нормализацией по интенсивности	DTW со сглаживанием	DTW с нормализацией по интенсивности и сглаживанием
Среднее значение оценок, полученных для пар (запись из сеанса норма1) – (запись из сеанса норма2)		1,405	4,41	0,225	0,402
Среднее значение оценок, полученных для пар (запись из сеанса норма1 или норма 2) - (запись из сеанса искажение)		1,864	4,8	0,246	0,438
Отношение средних оценок		0,738	0,754	0,919	0,918
Соблюдение критерия соответствия		+	+	+	+
Количество ошибок	норма1-искажение	32	29	33	29
	норма2-искажение	2	37	42	36
	Одновременно в обоих парах	1	19	24	19
Время выполнения (сек на 1 оценку)		3,54	3,8	3,26	3,78

Из представленных данных сделан вывод, что применение «чистого» DTW-алгоритма дает лучший результат, соответственно, этот вариант и будет использован в качестве алгоритма временной нормализации при оценке качества речи. Также необходимо отметить время выполнения вычислений: здесь во время выполнения включены как непосредственно вычисления, так и преобразование аудиозаписей в спектрограммы Фурье. Среднее время вычисления составляет более 3 секунд, что приводит к

невозможности реализации биологической обратной связи в режиме реального времени при использовании текущей версии алгоритма.

Также в качестве возможных алгоритмов для временной нормализации были исследованы встроенная функция программного комплекса Matlab `resample`, предназначенная для передискретизации последовательности [153], а также подход с нахождением наиболее похожих участков последовательностей на основе подсчета коэффициента корреляции [154]. Оба этих подхода показали неэффективность работы с полными несегментированными аудиозаписями слогов в разрезе времени выполнения алгоритма и корректности нахождения схожих интервалов. Применение этих подходов показывает результат лучше при использовании сегментированных аудиозаписей (то есть работа не со всем слогом, а только с проблемной фонемой). Применение существующих алгоритмов сегментации к набору аудиозаписей и вывод об их применимости описан в [155]. Подобные алгоритмы работают корректно с аудиозаписями с правильным произношением, но при искаженном произношении результат не является адекватным. Поэтому иные подходы к временной нормализации могут быть исследованы повторно на применимость в случае разработки или модификации алгоритмов сегментации аудиозаписи слога для выделения участков реализации каждой из фонем.

3.1.2 Выбор метрики для оценки

Для выбора метрики был использован следующий алгоритм анализа.

1. Проводится анализ оценок, полученных для выбранных метрик, для одной из фонем или групп фонем (комбинации фонема и ее мягкий вариант).

2. Если для одной фонемы метрика показывает подходящие результаты, то проводится анализ оценок, получаемых с использованием этих метрик для аудиозаписей с другими проблемными фонемами.

Так как зачастую за оценку разности двух объектов понимается расстояние между этими объектами, были исследованы на применимость метрики, которые используются для подсчета расстояния. Были выбраны следующие метрики [151]:

- Евклидово расстояние (3.7);

$$d(T, Q) = \sqrt{\sum_{i=1}^n (t_i - q_i)^2}, \quad (3.7)$$

где T и Q – числовые последовательности.

- Расстояние Чебышева (3.8);

$$d(\mathbf{t}, \mathbf{q}) = \max_{i=1, \dots, n} |t_i - q_i|, \quad (3.8)$$

где \mathbf{t} и \mathbf{q} – вектора с координатами (t_1, t_2, \dots, t_n) и (q_1, q_2, \dots, q_n) соответственно.

- Расстояние Минковского (3.9);

$$d(T, Q) = \sqrt[p]{\sum_{i=1}^n |t_i - q_i|^p}, \quad (3.9)$$

где p – параметр метрики Минковского.

- Манхэттенское расстояние, оно же расстояние городских кварталов (3.10);

$$d(T, Q) = \sum_{i=1}^n |t_i - q_i| \quad (3.10)$$

- Расстояние Махаланобиса (3.11).

$$d(\mathbf{t}, \mathbf{q}) = \sqrt{(\mathbf{t} - \mathbf{q})^T S^{-1} (\mathbf{t} - \mathbf{q})} \quad (3.11)$$

где S – матрица ковариации.

На спектрах Фурье различия на записях до и после хирургического вмешательства отчетливо видны изменения, что позволяет использовать спектрограммы для подсчета оценок в рамках выбора критерия. То есть в качестве объектов, между которыми будет вычисляться расстояния по формулам 3.7–3.11, использованы спектрограммы аудиозаписей. Подходящей метрикой считается метрика, получаемые по которой оценки отличаются в случае нахождения расстояния между эталонными слогами и в случае нахождения расстояния между эталонными и оцениваемыми слогами.

Далее в среде программирования MATLAB был реализован модуль, на вход которого подавались записи пациентов, в котором осуществлялся расчет всех описанных ранее метрик, на выходе составлялись матрицы расстояний между слогами или фонемами в слогах.

Для применения выбранных метрик расстояния были записаны слоги: 5 версий одного слога без изменений, 5 версий слогов, в которых анализируемая фонема была произнесена с характерными для нее изменениями. Для фонемы [к] были выбраны слоги «Фек», «Кась», «Вик» (слоги выбраны из списка слогов из раздела 2.1). Оценка производилась путем вычисления расстояния между парами слогов, подсчитывалось среднее расстояние между слогами без изменений в произношении фонемы, среднее расстояние между слогами с изменениями в произношении фонемы, а также среднее между парой слогов, один из которых с фонемой без изменений, второй с фонемой с изменениями в произношении.

В приложении Е в таблицах приведены примеры итоговых матриц результатов для оценки слога «Фек» для всех анализируемых метрик. Были подсчитаны значения средних расстояний для каждой из выбранных метрик и подсчитано, насколько изменилось среднее расстояние в случае пар между эталонными и оцениваемыми слогами относительно среднего между эталонными слогами. Здесь и далее под эталонными слогами подразумеваются слоги с правильным произношением исследуемой фонемы, под оцениваемыми слогами – слоги с искаженным произношением исследуемой фонемы.

В таблице 3.3 представлены средние расстояния и процент изменения. Подходящие результаты у метрики Минковского ($p=3$) и метрик, являющихся частным случаем метрики Минковского. Для расстояния Чебышева и Махаланобиса разница не наблюдается, соответственно эти метрики не могут быть применены в качестве итоговой метрики или в составе гибридной меры.

Таблица 3.3 – Средние расстояния для слога для полного спектра Фурье и изменения средних расстояний для выбранных метрик

	Евклидово	Чебышева	Минковского	Манхэттенское	Махаланобиса
Среднее между эталонными слогами	312,42	3,80	45,49	137 790,00	559 650,00
Среднее между эталонными и оцениваемыми слогами	383,03	3,97	54,68	172 876,00	531 696,00
Среднее для оцениваемыми слогами	321,61	3,83	47,04	140 380,00	538 460,00

Так как расстояние Евклида и Манхэттенское расстояние являются частными случаями метрики Минковского, был проведен анализ по выбору оптимального параметра метрики Минковского, который будет применяться в дальнейших расчетах.

Для выбора параметра метрики Минковского был сформирован набор аудиосигналов: для каждой из групп фонем по 5 записей с корректным произношением проблемных фонем, по 3 записи фонем с искаженным произношением проблемных фонем. Были выбраны следующие слоги, включающие в себя все перечисленные выше проблемные фонемы: “какь” [как’], “сось” [сос’], “туть” [тут’]. Для временной нормализации использовался DTW – алгоритм, расстояние Минковского рассчитывалось между трансформированными последовательностями согласно формуле 3.9. Были получены значения оценок для параметра p от -10 до 10. Для каждого значения параметра p были получены матрицы оценок, пример такой матрицы представлен в таблице 3.4 для $p=2$ для группы фонем [к] и [к’]. В таблице приняты следующие обозначения: норм1 – норм5 – обозначение аудиозаписей с нормальным произношением, иск1 – иск3 – аудиозаписи с искаженным произношением. Так как метрика Минковского и DTW – алгоритм являются симметричными относительно порядка подаваемых

аудиосигналов на вход, то нижняя половина матрицы оценок для пар норма-норма не рассчитывалась.

Таблица 3.4 – Матрица расстояний между сигналами для $p=2$.

	норм1	норм2	норм3	норм4	норм5	иск1	иск2	иск3
норм1		0,04918	0,05346	0,05909	0,06058	0,07689	0,06821	0,08086
норм2			0,04231	0,04710	0,05146	0,08077	0,07255	0,08408
норм3				0,04354	0,04737	0,08078	0,07176	0,08531
норм4					0,04586	0,08509	0,07527	0,09018
норм5						0,08393	0,08104	0,09308

Обозначим $S1$ набор оценок, получаемых в парах нормальное – искаженное произношения (в таблице 3.4 окрашены в **темно серый** цвет), $S2$ – набор оценок, получаемых в парах нормальное – нормальное произношение (в таблице 3.4 окрашены в **светло серый** цвет). $n1$ – количество сигналов с нормальным произношением, $n2$ – количество сигналов с искаженным произношением. Для выбора параметра метрики Минковского для каждого возможного значения рассчитаны следующие критерии.

1. Отношение средних значений наборов $S1$ и $S2$ (3.12). Чем ближе это значение к 1, тем ближе искаженное произношение к нормальному;

$$Cr1 = \frac{\bar{l}_1}{\bar{l}_2}, \quad (3.12)$$

где \bar{l}_1 – среднее значение оценок в наборе $S1$,

\bar{l}_2 – среднее значение оценок в наборе $S2$.

$$\bar{l}_1 = \frac{\sum_{i \in S1} l_i}{n1 * n2}, \quad (3.13)$$

$$\bar{l}_2 = \frac{\sum_{i \in S2} l_i}{n1*(n1-1)}. \quad (3.14)$$

2. Отношение минимального значения оценок из набора $S1$ к максимальному значению оценок из набора $S2$ (3.15). Отношение

минимального различия исходного и модифицированного сигналов к максимальному различию исходных сигналов. Если это значение больше 1, то значения наборов не пересекаются и чем дальше они друг от друга, тем лучше сформированы различия. Если значение меньше 1, то чем ближе оно к 1, тем меньше интервал пересечения наборов значений;

$$Cr2 = \frac{l_{1min}}{l_{2max}}, \quad (3.15)$$

где l_{1min} – минимальное значение оценок в наборе $S1$,

l_{2max} – максимальное значение оценок в наборе $S2$.

$$l_{1min} = \min_{i \in S1} l_i, \quad (3.16)$$

$$l_{2max} = \max_{i \in S2} l_i. \quad (3.17)$$

3. Оценка количества вхождений значений оценок одного набора данных в интервал значений второго (рассчитывается для каждого из набора значений) (3.18, 3.19). Доля пар сигналов с нормальными произношением, расстояние между которыми превышает минимальное расстояние между парами нормальное - искаженное произношение, а также доля пар нормальное – искаженное произношение, расстояние между которыми меньше максимального между парами нормальное – нормальное произношение.

$$Cr31 = \frac{2 * \text{count}_{i \in S2}(l_i > l_{1min})}{n1 * (n1 - 1)}, \quad (3.18)$$

$$Cr32 = \frac{\text{count}_{i \in S1}(l_i < l_{2max})}{n1 * n2}. \quad (3.19)$$

Предъявляются следующие правила анализа полученных критериев для итогового выбора параметра:

- для критерия $Cr1$ – значение должно быть больше 1, при прочих равных выбирается параметр с большим значением этого критерия;
- для критерия $Cr2$ – чем больше значение критерия, тем лучше;
- для критериев $Cr31$ и $Cr32$ – в идеале значения обоих критериев должно быть равно 0 (тогда значение $Cr2$ будет больше 1), чем меньше значение критерия, тем лучше.

Полученные значения критериев для рассматриваемого диапазона значений параметра p представлены на диаграммах на рисунках 3.1–3.3. На Рисунке 3.1 отражены значения критериев для групп фонем [к] и [к']. На Рисунке 3.2 отражены значения критериев для групп фонем [с] и [с']. На Рисунке 3.3 отражены значения критериев для групп фонем [т] и [т']. На всех диаграммах значения критериев $Cr31$ и $Cr32$ построены по основной оси ординат (слева), значения критериев $Cr1$ и $Cr2$ по вспомогательной оси ординат (справа). По оси абсцисс расположены значения параметров p .

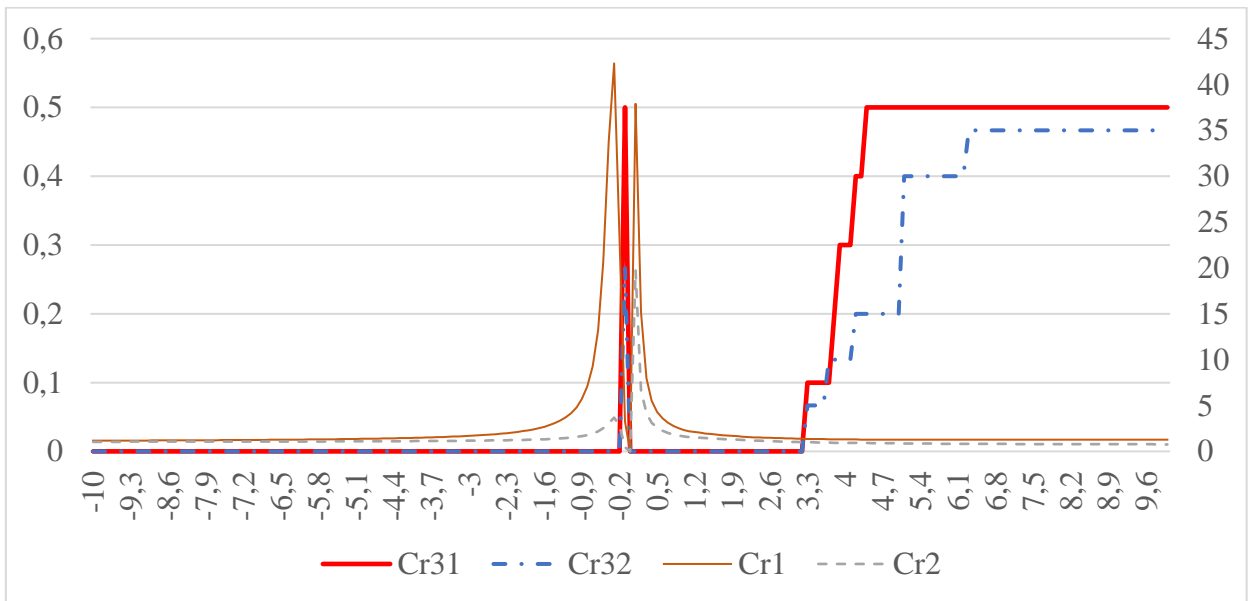


Рисунок 3.1 - Диаграмма значений критериев для рассматриваемого диапазона значений параметра p метрики Минковского для группы фонем [к] и [к'].

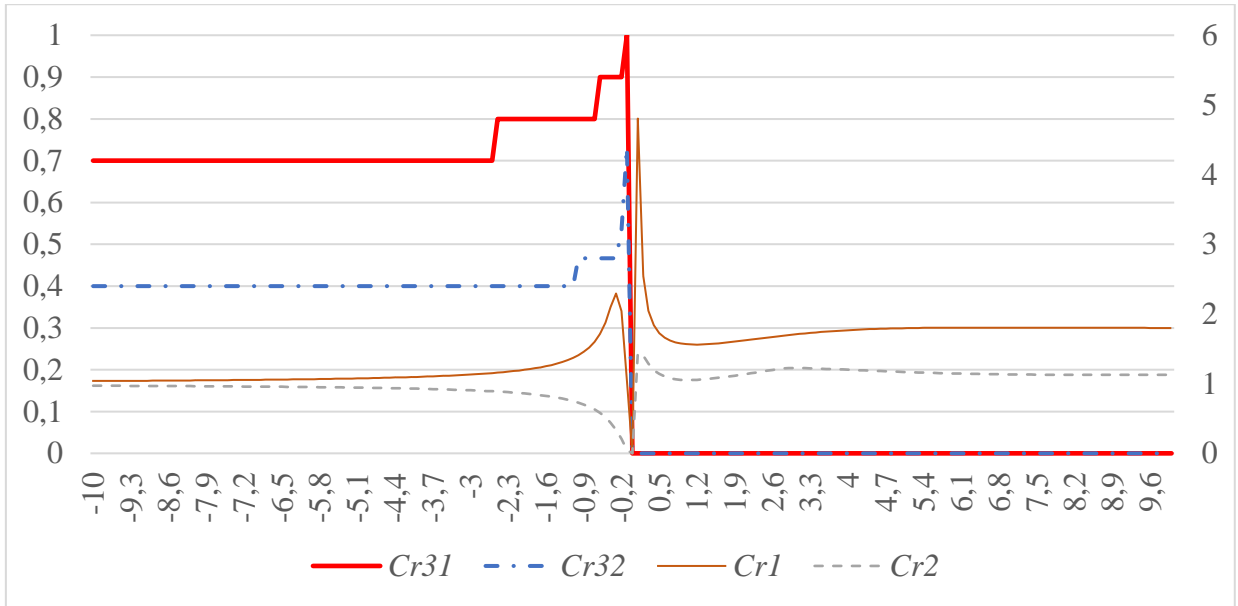


Рисунок 3.2 - Диаграмма значений критериев для рассматриваемого диапазона значений параметра p метрики Минковского для группы фонем [с] и [с'].

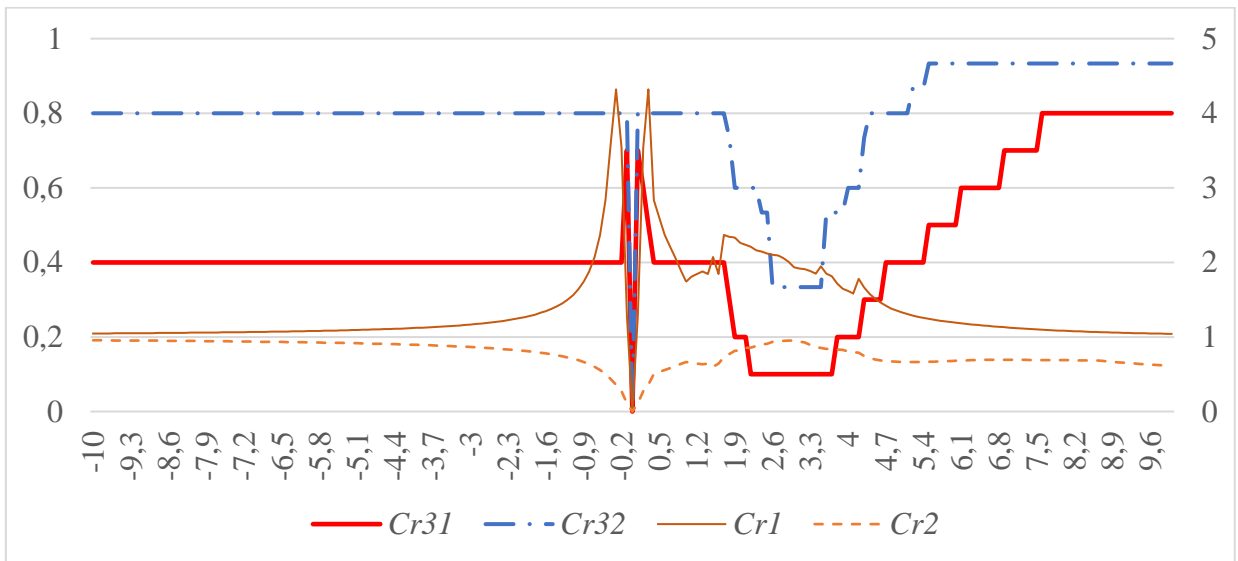


Рисунок 3.3 - Диаграмма значений критериев для рассматриваемого диапазона значений параметра p метрики Минковского для группы фонем [т] и [т'].

В соответствии с полученными значениями критериев и выделенными правилами анализа можно обозначит следующие возможные интервалы значений параметра p метрики Минковского для каждой из групп фонем: для группы фонем [к] и [к'] интервал $[-10; 3,2]$, для группы

фонем [с] и [с'] интервал [0,1; 10], для группы фонем [т] и [т'] интервал [2,4; 3,9]. Пересечением интервалов для всех групп фонем является интервал [2,4; 3,2]. Аналогичные исследования были проведены на аудиозаписях другого диктора [156], результаты показали оптимальный интервал значения p равный [1,6; 3,1], что пересекается с полученными выше значениями. В рамках дальнейших исследований для использования в метрике Минковского значение параметра p выбрано равно 3.

В качестве одной из возможных метрик для использования был предложен коэффициент корреляции как мера схожести двух последовательностей [157]. Для получения значения критерия корреляции используются значения нормированных по времени интенсивностей спектров Фурье до и после операции. При этом значение может рассчитываться как по отдельным фонемам, так и по целым слогам. Математически значение коэффициента корреляции можно записать в виде:

$$r_{XY} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2} * \sqrt{n \sum Y^2 - (\sum Y)^2}} \quad (3.20)$$

где X, Y – наборы значений, n – их длина.

Применительно к двумерному спектру сигнала с учетом реализации его вычисления на основе этой величины может быть сформирован критерий сравнительного качества речи K , записываемый в виде:

$$K = \frac{n_t n_f \sum_t \sum_f S_0 S_1 - (\sum_t \sum_f S_0) * (\sum_t \sum_f S_1)}{\sqrt{n_t n_f \sum_t \sum_f S_0^2 - (\sum_t \sum_f S_0)^2} * \sqrt{n_t n_f \sum_t \sum_f S_1^2 - (\sum_t \sum_f S_1)^2}} \quad (3.21)$$

где n_t — количество моментов времени после преобразования Фурье, n_f — количество значений частоты после преобразования Фурье, S_0 — сигнал до операции, S_1 — сигнал после операции, t – время, f – частота.

Чем ближе сигнал после операции к эталонному предоперационному сигналу, тем ближе значение этого критерия к 1. Как и ранее, эталонный сигнал – аудиозапись с правильным произношением исследуемой фонемы. Коэффициент линейной корреляции предназначен для выявления линейной

зависимости, что приводит к отсутствию необходимости нормировки сигналов по интенсивности.

Тестирование применимости коэффициента корреляции как меры схожести проводилось в два этапа: на модельном сигнале, имитирующем дисфункцию при использовании языка здоровым носителем, - «модельные пациенты» и на реальных картах пациентов до и после операции. Необходимость использования модельных сигналов обусловлена тем, что не все больные после операции остаются на процедуре речевой реабилитации, они предпочитают проходить ее по месту жительства. Это приводит к тому, что получается сделать только записи перед операцией и при выписке (но до того, как логопедом будут заданы даже базовые упражнения).

Полученные значения слоговой разборчивости по ГОСТ 50840–95, с использованием ранее исследованной метрики (метрика Минковского) и с использованием текущей метрики (коэффициент корреляции) для «модельных пациентов» 1, 2 и 3 представлены в Таблице 3.5 (левый столбец — речь до операции, правый столбец — речь после операции). Под «модельным пациентом» подразумевается здоровый диктор, имитирующий характерные для фонем изменения в произношении.

Таблица 3.5 - Средние значения критериев для модельных пациентов до и после операции (левый столбец для речи до операции, правый столбец для речи после операции)

«Модельный пациент»	Слоговая разборчивость (ГОСТ) Больше=лучше		Метрика Минковского Больше=хуже		Коэффициент корреляции Больше=лучше	
	1	0,6	97,45	126,91	0,62	0,52
2	1	0,8	131,12	165,27	0,51	0,44
3	1	0,6	118,73	144,14	0,55	0,49

Те же расчеты для реальных пациентов представлены в таблице 3.6. Как до, так и после операции было сделано пять записей.

Таблица 3.6 - Средние значения критериев для реальных пациентов до и после операции (левый столбец для речи до операции, правый столбец для речи после операции)

Реальный пациент	Слоговая разборчивость (ГОСТ) Больше=лучше		Метрика Минковского Больше=хуже		Коэффициент корреляции Больше=лучше	
1	1	0,6	127,41	141,54	0,56	0,48
2	1	0,6	115,21	139,28	0,56	0,51
3	1	0,6	141,81	165,53	0,43	0,35

Оценки, полученные по всем трем вариантам оценки качества речи, не содержат противоречий, что подтверждает их состоятельность. Но так как вычисления были произведены на основе всего 5 опорных сигналов, можно говорить лишь про возможность применения рассматриваемых метрик. Дальнейшие вычисления на большем объеме данных дают более точную оценку применимости и значимости оценок, получаемых с использованием выбранных метрик.

В рамках итогового выбора метрики или составления гибридной меры, составленной путем комбинации нескольких метрик, были отобраны следующие подходящие для использования варианты:

- DTW-расстояние, как уже готовая количественная оценка стоимости пути между двумя трансформированными последовательностями;
- коэффициент корреляции как оценка схожести двух временных рядов;
- метрика Минковского как количественная оценка расстояния между двумя объектами с использованием значения параметра p равного 3.

Первоначальные вычисления были проведены на синтетически сформированном наборе данных. Имеется два набора записей слога «ФЕК»: пять записей с нормальным произношением фонемы [к], пять записей с искаженным произношением фонемы [к]. Записи были сделаны здоровым диктором женского пола. Задачи достоверной имитации всех возможных

изменений в произношении проблемных фонем не стояло, диктор опирался на наиболее часто встречающееся изменение в произношении.

Обозначим оценку, получаемую в результате сравнения пары записей, в каждой из которых проблемная фонема произнесена правильно, оценкой "норма/норма", а оценку, получаемую для пары записей, в одной из которых фонема произнесена правильно, а во второй с искажением, оценкой "норма/искажение". Предполагается, что получаемые оценки "норма/норма" должны отличаться от оценок "норма/искажение". Соответственно, множества оценок не должны пересекаться.

Так как все записи имеют различные длины, то перед непосредственным нахождением количественной оценки схожести производилась временная нормализация. Ранее для нормализации предложен DTW-алгоритм. Все три проверяемые метрики являются симметричными, порядок в паре при подаче на вход алгоритма неважен. В качестве временных последовательностей используются значения амплитуд сигналов.

Полученные оценки представлены в таблицах 3.7–3.8 для DTW-расстояния, таблицах 3.9–3.10 для коэффициента корреляции и Таблицах 3.11–3.12 для расстояния Минковского ($p=3$). В таблицах записи с нормальным произношением обозначены н1-н5, с искаженным и1-и5.

Таблица 3.7 – DTW-расстояния для пар «норма/норма»

Норма/Норма	н1	н2	н3	н4	н5
н1	0				
н2	16,738	0			
н3	17,474	15,62	0		
н4	18,276	18,912	18,447	0	
н5	16,717	18,142	18,632	12,937	0

Таблица 3.8 – DTW-расстояния для пар «норма/искажение»

Норма/Искажение	и1	и2	и3	и4	и5
н1	25,487	28,537	34,649	28,837	33,159
н2	23,394	26,658	30,983	26,814	30,825
н3	22,82	25,25	31,246	26,361	31,626
н4	24,831	28,507	39,49	32,227	37,891
н5	24,953	28,006	40,15	32,013	37,976

Таблица 3.9 – Коэффициенты корреляции для пар «норма/норма»

Норма/Норма	н1	н2	н3	н4	н5
н1	1				
н2	0,053	1			
н3	0,072	0,209	1		
н4	0,187	0,102	0,085	1	
н5	0,109	0,106	0,194	0,462	1

Таблица 3.10 – Коэффициенты корреляции для пар «норма/искажение»

Норма/Искажение	и1	и2	и3	и4	и5
н1	0,124	0,163	0,045	0,072	0,063
н2	0,112	0,151	0,109	0,119	0,165
н3	0,125	0,1	0,394	0,169	0,774
н4	0,24	0,188	0,039	0,111	0,082
н5	0,176	0,17	0,094	0,151	0,149

Таблица 3.11 – Расстояния Минковского для пар «норма/норма»

Норма/Норма	н1	н2	н3	н4	н5
н1	0				
н2	0,466	0			
н3	0,441	0,41	0		
н4	0,394	0,381	0,368	0	
н5	0,399	0,348	0,243	0,266	0

Таблица 3.12 – Расстояния Минковского для пар «норма/искажение»

Норма/Искажение	и1	и2	и3	и4	и5
н1	0,378	0,373	0,556	0,408	0,544
н2	0,411	0,356	0,541	0,474	0,556
н3	0,403	0,412	0,498	0,453	0,331
н4	0,231	0,234	0,514	0,264	0,493
н5	0,253	0,245	0,259	0,252	0,264

Для DTW-расстояния оценки в парах «норма/норма» заключены в отрезок [12,937; 18,912] (здесь и далее не используются значения, получаемые в случае подачи на вход алгоритма двух идентичных записей), в парах «норма/искажение» [22,82; 40,15]. Данная метрика подходит под обозначенный критерий выбора, множества оценок не пересекаются.

Для коэффициента корреляции оценки пар «норма/норма» принадлежат отрезку [0,53; 0,462], оценки пар «норма/искажение» [0,039; 0,774]. Для расстояния Минковского оценки пар «норма/норма» [0,243; 0,466], оценки

пар «норма/искажение» [0,231; 0,556]. Для искусственно сформированного набора данных с учетом вышеописанного критерия выбора эти две метрики не являются применимыми для решения поставленной задачи. Однако эти метрики возможны к применению как составные части гибридной меры, сформированной на основе комбинации нескольких метрик с применением некоторого механизма комбинирования.

В рамках создания гибридной меры (или окончательного выбора одной из метрик как итоговой для применения) были произведены вычисления на основе записей сеансов речевой реабилитации реальных пациентов. Были взяты звуковые файлы, которые записывались в рамках процесса речевой реабилитации после комбинированного лечения онкологических заболеваний органов полости рта и ротоглотки, проводимого на базе НИИ Онкологии г. Томска. Запись сеансов осуществлялась согласно списку слогов для записи, описанного в разделе 2.1 настоящей работы. Были отобраны записи 15 пациентов. Были выбраны те пациенты, у которых более двух сеансов, и записи сеансов были произведены по составленному списку слогов с использованием первой версии разработанного программного комплекса. У четырех пациентов по четыре сеанса, у 11 пациентов по три сеанса. Первый сеанс у каждого пациента является эталонным, то есть именно с записями этого сеанса будут сравниваться соответствующие записи из остальных сеансов. В эталонном сеансе у пациента правильное (или максимально приближенное к правильному) произношение фонемы, в оцениваемом сеансе – произношение на момент записи аудио, его и требуется оценить. Таким образом, для подсчета оценок использовано 34 сеанса.

Все выбранные записи были оценены двумя способами.

Первый способ – старая методика (адаптированная методика согласно ГОСТ Р 50840–95 (описана в разделе 1.4). Оценка выставляется следующим образом: 1 если слог произнесен полностью верно, 0 – иначе. Итоговые оценки выбранных сеансов (кроме эталонных) были выставлены

следующим образом: 1 если и у записи в эталонном сеансе оценка 1, и записи в оцениваемом сеансе оценка 1, 0 – иначе.

Второй способ: с использованием разрабатываемого алгоритма для автоматизированной оценки качества речи. На вход алгоритму подается два звуковых файла: эталонная запись и оцениваемая запись. На выходе получается модельная (количественная) оценка. Для временной нормализации в алгоритме оценки использовался DTW-алгоритм (описание представлено в разделе 3.1.1).

По аналогии с предыдущими исследованиями считались три оценки с использованием трех разных метрик:

- DTW – расстояние,
- коэффициент корреляции,
- расстояние Минковского (параметр метрики p равен 3).

Критерий выбора метрики или комбинации метрик – наибольшая согласованность с оценками по старой методике.

Оценка, полученная по старой методике, является бинарной (либо 0, либо 1), а оценки по алгоритму являются вещественными: для DTW-расстояния и расстояния Минковского оценка является положительным вещественным числом, для коэффициента корреляции оценка заключена в интервал от -1 до 1. Для сравнения оценок между собой было принято решение о преобразовании всех оценок к бинарному виду.

Подбор комбинации метрик осуществлялся для каждой из групп фонем по отдельности. Использование группы фонем, а не одной из них, позволило получить больший набор числовых данных для анализа, а применимость данной группировки обоснована схожестью возникающих в постоперационных период искажений в произношениях фонем.

Ниже приведено описание эксперимента для группы фонем [к] и [к']. Суммарно с фонемами этой группы было исследовано 1020 записей. Для двух других групп ([т] и [т'], [с] и [с']) исследование проводилось аналогичным образом.

Обозначим следующие наборы данных:

— $X = \{X_1, X_2, \dots, X_{1020}\}$ – оценки, полученные по старой методике;

— $Y = \{Y_1, Y_2, \dots, Y_{1020}\}$ - оценки, полученные с использованием метрики DTW-расстояния;

— $Z = \{Z_1, Z_2, \dots, Z_{1020}\}$ - оценки, полученные с использованием коэффициента корреляции;

— $W = \{W_1, W_2, \dots, W_{1020}\}$ - оценки, полученные с использованием метрики расстояние Минковского).

Комбинированную оценку обозначим $C = \{C_1, C_2, \dots, C_{1020}\}$. Для ее расчета предложено использование нечеткого классификатора, алгоритмом оптимизации внутри которого является генетический алгоритм.

Можно сказать, что старая методика является классификацией: для записи выставляется оценка 1 или 0, тем самым набор аудиозаписей разбивается на два класса – если оценка 1, то аудиозапись относится к классу «произношение слога в аудиозаписи полностью разборчиво», если оценка 0, то аудиозапись относится к классу «произношение слога в аудиозаписи неразборчиво». Поэтому в качестве гибридной меры для оценки предложено использование нечеткого классификатора.

Нечеткая система может быть представлена [158] как

$$y = f(\mathbf{g}, \boldsymbol{\theta}, \mathbf{r}),$$

где \mathbf{g} – входной вектор значений $\|g_1, g_2, g_3\|$ (в рамках работы $g_1 = Y_i, g_2 = Z_i, g_3 = W_i$),

$\boldsymbol{\theta} = \|\theta_1, \dots, \theta_N\|$ — вектор параметров антецедентов,

$N = n * (\text{число параметров, описывающих одну функцию принадлежности}) * (\text{число термов, описывающих одну входную переменную}),$

y — скалярный выход системы,

$\mathbf{r} = \|r_1, \dots, r_R\|$ — вектор параметров консеквентов.

В нечеткой системе i -ое правило имеет следующий вид:

$$\text{IF } g_1=A_{1i} \text{ AND } g_2=A_{2i} \text{ AND } g_3=A_{2i} \text{ THEN } y = r_i,$$

где A_{ij} — лингвистический терм, которым оценивается переменная g_i ,
 r_i — значение консеквента i -го правила.

В качестве алгоритма оптимизации, используемого в нечетком классификаторе, используется генетический алгоритм. Генетический алгоритм представляет собой эвристический алгоритм поиска для решения задач оптимизации и моделирования путём последовательного подбора, комбинирования и вариации искомых параметров с использованием механизмов, напоминающих биологическую эволюцию [159].

На вход классификатору подается три значения: значение расстояния по метрике DTW-расстояния Y , значение коэффициента корреляции Z , значение расстояния по метрике Минковского W . На выходе получается значение оценки C , принадлежащее интервалу $[0;1]$. Интерпретация выходного значения сформулирована как «чем более схоже произношение с эталонными сигналами, тем больше оценка». Значение 1 соответствует оценке по старой методике, равной 1. Далее классификатор на основе подобранного в процессе обучения порогового значения определяет класс, к которому относится комбинация поданных на вход значений – 0 или 1.

Обучение классификатора произведено на основе набора оценок по старой методике X и получаемых по метрикам оценок. Для каждой группы фонем был обучен отдельный классификатор [160]. Точности классификации для тестовых и обучающих выборок для рассматриваемых групп фонем приведены в таблице 3.13. Точности были получены на 10-кратной кросс-валидации. Средняя точность для тестовых выборок составила 0,83.

Таблица 3.13 – Точности классификации для групп проблемных фонем.

	Точность на тестовой выборке	Точность на обучающей выборке
Группа [к] и [к']	0,87	0,88
Группа [с] и [с']	0,84	0,85
Группа [т] и [т']	0,79	0,81

Полученные значения точностей на тестовых выборках являются точечными оценками относительной частоты случаев правильно выставленной метки класса относительно общего объема тестовой выборки. Построим доверительный интервал для полученных значений согласно интервальной оценке неизвестной вероятности по относительной частоте [136]. Для уровня надежности, равного 0,95, интервальные оценки представлены в таблице 3.14.

Таблица 3.14 – Интервальные оценки точностей.

	Интервальная оценка точности
Группа [к] и [к']	$0,87 \pm 0,021$
Группа [с] и [с']	$0,84 \pm 0,022$
Группа [т] и [т']	$0,79 \pm 0,025$

В случае работы с иными фонемами (не входящими в список проблемных), предложено использование рассчитанного значения DTW-расстояния в качестве итоговой оценки, потому что именно эта метрика соответствует всем предъявляемым к разрабатываемой мере требованиям.

В работах [161, 162] рассмотрен подход к формированию гибридной меры на основе линейной свертки рассчитанных значений метрик с учетом коэффициентов значимости, такой подход может быть использован, однако его недостатком является необходимость ручного подбора пороговых значений преобразования количественного значения комбинированной меры для бинаризации и соотнесения с оценками по старой методике. При наличии возможности ручного подбора, значения, получаемые с использованием линейной свертки, согласуются с оценками по старой методике по критерию знаков на уровне значимости 0,01.

Также было оценено время выполнения вычислений. Так как при использовании любой из метрик сначала производится временная нормализация, время подсчета первой метрики (DTW – расстояния) равно времени выполнения временной нормализации. Поэтому основная задача в

оценке времени – оценить время, затрачиваемое на расчеты двух других метрик, в контексте общего времени обработки одной пары записей.

Среднее время вычисления количественного значения схожести звуковых сигналов, вычисленное на 5000 парах сигналов, составило 1,4952 секунды. Из этого времени 1,4578 секунды (97,5% от общего времени) – временная нормализация и вычисление DTW-расстояния и 0,023 секунд (1,5% от общего времени) – вычисление метрик Минковского и коэффициента корреляции. Это позволяет сделать вывод, что вклад времени вычисления двух метрик (Минковского и коэффициента корреляции) незначителен в общем времени вычислений, поэтому их использование для повышения точности оценки в рамках решения поставленной задачи является возможным.

3.1.3 Алгоритм нахождения количественной оценки схожести двух речевых сигналов

На вход алгоритму подаются два аудиосигнала с одинаковыми характеристикам самих сигналов (частота дискретизации, формат аудиофайла). Аудиосигналы содержат различные реализации одного и того же слога. Длины аудиосигналов могут быть различными. Также на вход подается указание на проблемную фонему в записанном слоге для применения построенного классификатора или выбора метрики DTW в качестве итоговой.

С учетом подобранных параметров итоговый алгоритм представляет собой следующую последовательность шагов.

Шаг 1. Преобразование сигналов в последовательность значений. Оба сигнала преобразуются в массив значений амплитуд сигналов – числовые последовательности.

Шаг 2. Временная нормализация двух числовых последовательностей – приведение последовательностей к единой длине. Этот шаг выполняется с

использованием алгоритма динамической трансформации временной шкалы – DTW – алгоритм.

Шаг 3. Фиксируется одна из количественных оценок – DTW-расстояние. Это значение стоимости пути между двумя последовательностями – последний элемент матрицы преобразований.

Шаг 4. На основе матрицы преобразований и построенного на шаге 2 оптимального пути составляются трансформированные числовые последовательности.

Шаг 5. Нахождение количественных оценок между трансформированными числовыми последовательностями с использованием двух метрик: коэффициента корреляции и метрики Минковского (параметр $p = 3$).

Шаг 6. Для выбранной проблемной фонемы выбирается нечеткий классификатор, обученный для данной группы проблемных фонем. Если фонема не входит в список выделенных проблемных фонем, то делается указание на это.

Шаг 7. Итоговая количественная оценка находится как выход нечеткого классификатора (на вход подаются значения DTW-расстояния, коэффициента корреляции и метрики Минковского) в случае оценки слога с одной из выделенных проблемных фонем или как DTW-расстояние иначе.

Шаг 8. Вывод итоговой оценки.

На рисунке 3.4 описанный выше алгоритм представлен в нотации Насси – Шнейдермана [163].

Получить числовую последовательность 1 – значения амплитуд сигнала 1			
Получить числовую последовательность 2 – значения амплитуд сигнала 2			
Применить DTW-алгоритм к числовым последовательностям 1 и 2			
Запомнить значение Y - DTW-расстояние (значение последнего элемента матрицы преобразований)			
Получить трансформированные числовые последовательности 1 и 2 согласно матрице преобразований и кратчайшему пути			
Вычислить значение Z – коэффициент корреляции Пирсона между трансформированными числовыми последовательностями 1 и 2			
Вычислить значение W – расстояние по метрике Минковского между трансформированными числовыми последовательностями 1 и 2			
Выбрать проблемную фонему			
[к] или [к']	[с] или [с']	[т] или [т']	Иначе
Получить C как выход нечеткого классификатора FC1 (вход – Y, Z, W)	Получить C как выход нечеткого классификатора FC2 (вход – Y, Z, W)	Получить C как выход нечеткого классификатора FC3 (вход – Y, Z, W)	$C = Y$
Вывести количественную оценку C			

Рисунок 3.4 – Алгоритм нахождения количественной оценки

Для практической реализации первого шага алгоритма требуется обработка аудиофайла. Эта подготовка представляет собой считывание информации из аудиофайлов с расширением .wav и ее преобразование в числовые последовательности, пригодные для дальнейшей обработки. Считывание файла производится встроенными средствами библиотек System.IO [164], преобразование считанного файла основано на знании структуры файлов используемым расширением, а именно то, что служебная информация содержится в первых 44 байтах, остальные данные — это значения амплитуд записанного звука. Именно эти амплитуды и

представляют собой временные последовательности, с которыми будут производиться дальнейшие преобразования.

Данный алгоритм может быть модернизирован для решения иных задач в рамках изменения следующих параметров.

1. В шаге 1 возможно преобразование в числовую последовательность сигнала путем построения спектра сигнала, например, спектрограмма Фурье, построения огибающей сигнала на основе расчета различных параметров звука.

2. При наличии адекватно работающих алгоритмов сегментации (подходящих под решаемую задачу) возможна работа только с частью сигнала. Тогда для временной нормализации могут быть применены иные алгоритмы, например функция передискретизации сигнала или подход на основе поиска наиболее схожих фрагментов с использованием корреляционного коэффициента.

3. Возможно применение фильтров шума или определенных частот, а также предварительной нормализации сигналов по интенсивности. Необходимость и возможность такой модификации зависят от решаемой задачи и параметров исследуемых сигналов.

4. Изменение механизма формирования гибридной меры (к примеру, использование линейной или иной свертки) и изменение состава метрик – возможно при проведении дополнительных исследований как в рамках решения текущей практической задачи (например, при расширении списка фонем), так и в рамках решения иных задач по оценке речи.

Возможные модификации и реализации предложенного алгоритма нахождения количественной оценки описаны в [165, 166, 167].

3.2 Методика оценки схожести звуковых сигналов

Предложенный алгоритм позволяет количественно оценить схожесть двух звуковых сигналов при их представлении в виде числовых последовательностей. Однако такую оценку, представляющую собой положительное действительное число, невозможно однозначно интерпретировать в контексте оценки схожести речи к эталонной или, в рамках данной задачи, к предоперационной речи конкретного пациента. Предлагаемая методика оценки схожести звуковых сигналов за счет применения нескольких опорных сигналов позволяет получить интерпретируемую оценку схожести произношения, заключенную в заранее известный интервал. Использование нескольких опорных сигналов позволяет учитывать вариативность речи до операции, так как в речи нет фиксированной продолжительности или точных значений параметров произношения фонетических единиц.

Подразумевается, что по предлагаемой методике можно оценить подобие звуковых сигналов, представляющих одинаковую по содержанию фонетическую единицу речи – слог. То есть, можно сравнивать между собой записи только одного и того же слога.

Стоит отметить, что при доработке алгоритмов сегментации и модернизации алгоритма оценки возможно будет сравнивать разные слоги, содержащие одну и ту же проблемную фонему, однако оценка подобия будет производиться только для фрагментов, содержащих эту фонему, а не слогов целиком.

Здесь и далее приняты следующие обозначения:

— эталонный сеанс - сеанс записи речи пациента, проводимый до проведения хирургического лечения, речь близка к нормальной, возможны изменения, связанные с медицинскими аспектами онкологических заболеваний данной локализации;

— эталонный набор записей/аудиосигналов – записанные в процессе проведения эталонного сеанса оценки речи аудиосигналы, содержащие близкое к нормальному произношению слогов и фонем;

— эталонная запись – одна запись из эталонного набора;

— оцениваемый сеанс – сеанс записи речи пациента на одном из этапов речевой реабилитации (до, в процессе или после), речь пациента может быть как искажена, так и близка к нормальной в зависимости от медицинских аспектов (диагноз, локализация новообразования, объем хирургического вмешательства и др.);

— оцениваемый набор записей – записанные в процессе проведения оцениваемого сеанса оценки речи аудиосигналы, оценка которого в сравнении с эталонными должна быть получена по описываемой методике;

— оцениваемая запись - одна запись из оцениваемого набора.

Опорными сигналами служат эталонные записи из нескольких эталонных сеансов. С точки зрения теоретического решения поставленной задачи количество опорных сигналов ограничено только вычислительными возможностями, большее количество опорных сигналов позволит учитывать большее количество вариантов нормального произношения одной и той же единицы речи. С точки зрения решения практической задачи в рамках речевой реабилитации рекомендуемым количеством опорных сигналов, а следовательно, и эталонных сеансов, является 2 сигнала и 2 сеанса соответственно. Это количество позволяет учитывать вариативность речи, и в то же время запись такого количества сеансов возможна в процессе проведения предоперационной подготовки. Запись такого количества сеансов также является необременительной для пациентов, подверженных болевому синдрому из-за онкологического заболевания, однако проведение сеансов записи рекомендовано проводить либо в разные дни, либо с перерывом в несколько часов в течение одного дня. Такие рекомендации способствуют уменьшению нагрузки на органы речепродуцирования.

Методика заключается в последовательном выполнении следующей последовательности действий:

1. Выбор нескольких эталонных записей из нескольких эталонных сеансов и оцениваемой записи из оцениваемого сеанса. Все записи представляют собой реализацию одного и того же слога.

2. Нахождение количественных оценок между парами эталонных записей по алгоритму, описанному в разделе 3.1.3.

3. Нахождение среднего значения количественных оценок, полученных в пункте 2.

4. Нахождение количественных оценок между каждой из эталонных записей и оцениваемой записью по алгоритму, описанному в разделе 3.1.3.

5. Нахождение среднего значения количественных оценок, полученных в пункте 4.

6. Нахождение отношения среднего значения оценок между эталонными и оцениваемым сигналами к среднему значению оценок между эталонными сигналами – относительная оценка оцениваемого сигнала.

7. Вывод о схожести оцениваемого сигнала и эталонных по относительной оценке.

Относительная оценка заключена в интервал от 0 до 1 (включая обе границы) и чем ближе результат к 1, тем качественнее (= более похоже на эталон) произнесен слог.

Относительная оценка может быть равна 0 при условии полной идентичности сигналов, то есть на вход алгоритму, реализующему методику, в качестве эталонных и оцениваемых подан один и тот же звуковой файл.

Само отношение средних оценок теоретически может быть больше 1, это означает что оцениваемая речь полностью схожа с эталонной и речь пациента не нуждается в речевой реабилитации с точки зрения анализа количественной оценки схожести речевых сигналов. В таком случае окончательное решение о необходимости проведения реабилитации

принимается лечащим врачом – логопедом. Если отношение в процессе вычисления получается больше 1, то согласно алгоритмам расчета относительная оценка принимается равной 1.

Относительная оценка может принимать значение 1 при условии полного восстановления речи до предоперационного уровня. Если рассматривать с точки зрения наличия предпосылок, все из них говорят об ухудшении речи пациента после проведения оперативного вмешательства и невозможности достижения предоперационного уровня. Проведенное оперативное вмешательство (зачастую резекция одного из органов речеобразующего аппарата) приводит к невозможности полного восстановления правильного произношения всех проблемных фонем. Практический анализ аудиозаписей пациентов до и после хирургического лечения, в процессе речевой реабилитации, а также через 3, 6 и 12 месяцев после операции подтверждает описанное выше.

На примере наличия двух эталонных сигналов проведение оценки сеанса записи с использованием описанной выше методики может быть представлено в виде следующего алгоритма.

1. Выбрать два эталонных сеанса ε_1 и ε_2 (опорные сеансы);
2. Выбрать сеанс для оценки (оцениваемый сеанс) o ;

Для первого слога:

3. Найти оценку слогу по алгоритму для пар: $o-\varepsilon_1$, $o-\varepsilon_2$, $\varepsilon_1-\varepsilon_2$:

$$C_{o,\varepsilon_1}, C_{o,\varepsilon_2}, C_{\varepsilon_1,\varepsilon_2};$$

4. Найти относительную оценку сигнала (слога) d :

$$4.1 \quad C_{o,\varepsilon} = \frac{C_{o,\varepsilon_1} + C_{o,\varepsilon_2}}{2} \quad \text{— среднее значение оценок в парах}$$

(оцениваемая-эталонная);

$$4.2 \quad \text{относительная оценка } C_o = \frac{C_{o,\varepsilon}}{C_{\varepsilon_1,\varepsilon_2}};$$

Оценка C_o в интервале $[0;1]$. Чем ближе к 1, тем лучше.

5. Повторить шаги 3–4 для остальных слогов в сеансе;

6. Найти оценку сеанса, как среднее арифметическое оценок слогов, входящих в оцениваемый сеанс

Описанная последовательность действий в виде NS-диаграммы представлен на рисунке 3.5.

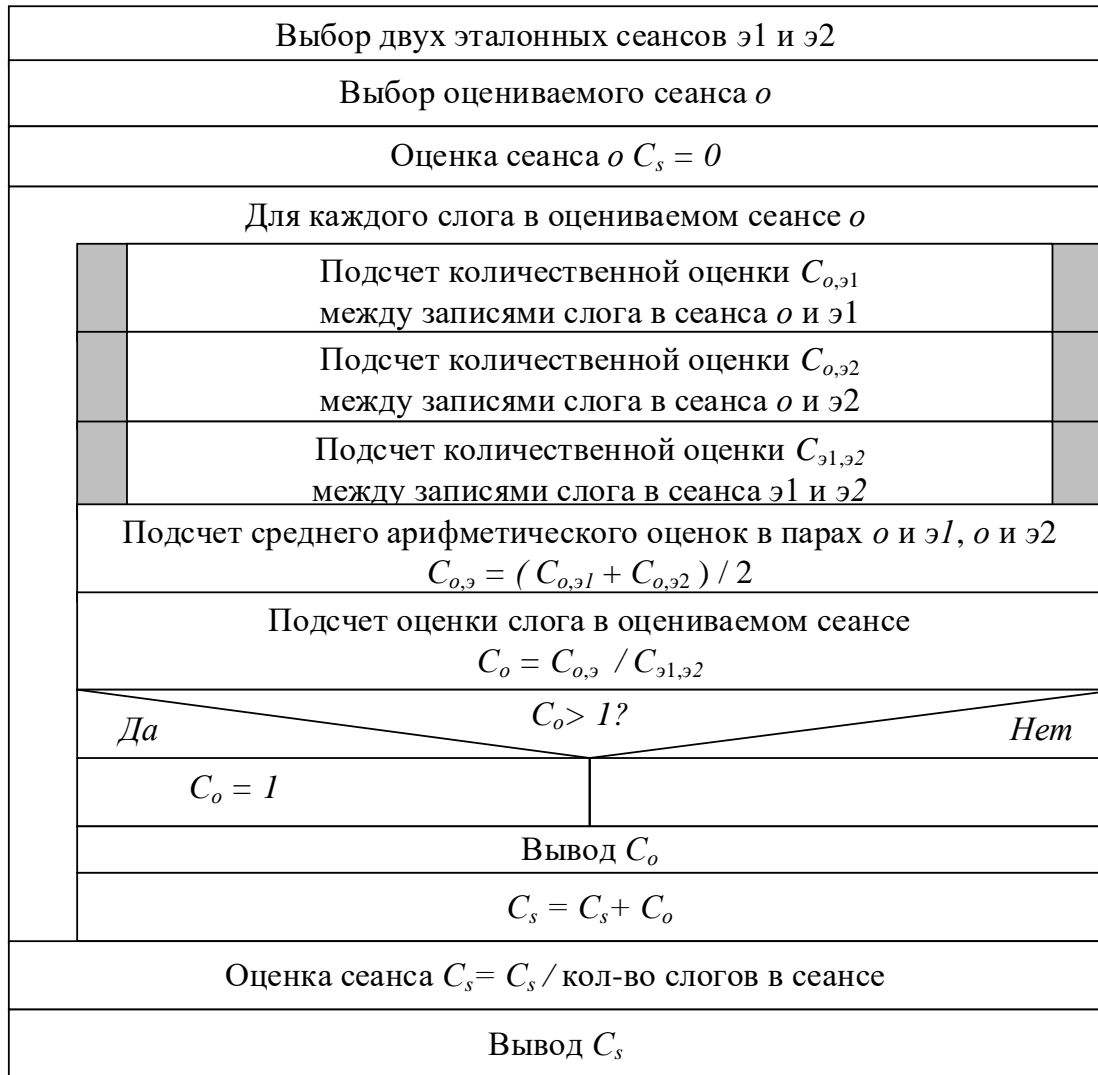


Рисунок 3.5 – Последовательность действий для двух эталонных сеансов согласно методике оценки схожести звуковых сигналов

Использование в качестве эталонных сигналов записей речи того же пациента, чью речь необходимо оценить, позволяет учитывать особенности речи пациента, в то время как использование в качестве эталона записей некоторого человека с идеальным произношением приведет к существенному занижению оценки качества речи.

Было рассчитано, как изменяется оценка сеанса в зависимости от набора эталонных сеансов. Были выбраны сеансы, в которых речь

пациентов и дикторов близка к идеальной, то есть оценка по старой методике большинства записей в сеансе равна 1 (допускалось не более пяти 0 в оценках). В таблице 3.15 приведен пример сравнения оценок. В сеансах 1 и 2 диктора 1 и сеансах 4 и 5 диктора 2 все слоги были произнесены правильно (оценки произношения равны 1), в сеансе 3 диктора 1 было допущено 4 ошибки (в слоге была неправильно произнесена проблемная фонема, остальные фонемы – правильно).

Таблица 3.15 – Сравнение оценок сеанса с использованием эталонных сеансов разных дикторов.

Оцениваемый сеанс	Эталонные сеансы			
	Диктор 1			Диктор 2
	сеанс 1+ сеанс 2	сеанс 1 + сеанс 3	сеанс 2 + сеанс 3	сеанс 4 + сеанс 5
сеанс 1 (диктор 1)	-	-	0,716	0,252
сеанс 2 (диктор 1)	-	0,982	-	0,296
сеанс 3 (диктор 1)	0,992	-	-	0,383

Из данных таблицы видно, что использование в качестве эталона сеансов другого диктора приводит к существенному занижению оценок произношения слогов. Для установления и проверки направленности изменений получаемых оценок были получены оценки сеансов 4 пациентов, для каждого из которых была проведена оценка на основе сравнения с сеансами этого же пациента и других пациентов. Проверка производилась с использованием критерия Вилкоксона [136]. Оценки представлены в таблице 3.16. Сформулирована нулевая гипотеза: оценка сеанса пациента, сформированная на основе сравнения с эталонными сеансами этого же пациента, превышает оценку этого же сеанса, сформированную на основе сравнения с эталонными сеансами других пациентов. Альтернативная гипотеза: оценки сеансов пациента, получаемых на основе сравнения с эталонными сеансами этого же пациента, не превышают оценки этого же сеанса, полученных на основе сравнения с эталонными сеансами других пациентов.

Таблица 3.16 – Оценки сеансов в зависимости от принадлежности эталонных сеансов разным пациентам.

Оценка сеанса (эталонные сеансы этого же пациента)	Оценка сеанса (эталонные сеансы другого пациента)	Разность	Номер ранга разности
0,716	0,252	0,464	34
0,982	0,296	0,686	3
0,992	0,383	0,609	6
0,716	0,297	0,419	35
0,982	0,311	0,671	4
0,992	0,402	0,59	9
0,716	0,307	0,409	36
0,982	0,354	0,628	5
0,992	0,432	0,56	18
0,716	0,231	0,485	33
0,982	0,258	0,724	1
0,992	0,304	0,688	2
0,857	0,309	0,548	21
0,952	0,375	0,577	13,5
0,988	0,406	0,582	12
0,857	0,284	0,573	16
0,952	0,388	0,564	17
0,988	0,413	0,575	15
0,857	0,357	0,5	30
0,952	0,442	0,51	26
0,988	0,451	0,537	23
0,857	0,305	0,552	19
0,952	0,375	0,577	13,5
0,988	0,402	0,586	10
0,904	0,388	0,516	25
0,989	0,407	0,582	11
0,996	0,458	0,538	22
0,904	0,355	0,549	20
0,989	0,396	0,593	8
0,996	0,402	0,594	7
0,904	0,399	0,505	27
0,989	0,465	0,524	24
0,996	0,498	0,498	31
0,904	0,403	0,501	29
0,989	0,487	0,502	28
0,996	0,506	0,49	32

Для подсчета значения критерия были получены разности значений оценок (столбец Разность таблицы 3.16), а также проведена ранжировка полученных значений разности (столбец Номер ранга таблицы 3.16). Наблюдаемое значение критерия Вилкоксона представляет собой сумму рангов нетипичных сдвигов. По данным таблицы, все сдвиги (разности) имеют одинаковую направленность, поэтому наблюдаемое значение Т-критерия Вилкоксона равно 0. Критическое значение для объема выборки $n = 36$ и уровня значимости 0,01 равно 185. Принимается нулевая гипотеза о том, что оценки сеансов пациента, полученные с использованием эталонных сеансов этого же пациента, превосходят оценки, получаемые с использованием эталонных сеансов других пациентов.

Поэтому использование в качестве эталона предоперационной речи самого пациента, согласно описанной выше методике, позволяет учитывать индивидуальные особенности речи пациента.

В рамках практического применения возможна модификация методики оценки схожести звуковых сигналов в случае невозможности записи двух эталонных сеансов до операции по организационным или медицинским причинам. В этой ситуации возможно проведение оценки на основе одного эталонного сеанса. Методика оценки модифицируется в следующую последовательность шагов.

Шаг 1. Выбор эталонной записи слога из эталонного сеанса и оцениваемой записи этого же слога из оцениваемого сеанса.

Шаг 2. Вычисление количественной оценки между оцениваемой записью и эталонной записью по алгоритму из раздела 3.1.3.

При такой модификации оценка сеанса будет представлять собой среднее значение оценок записей, входящих в этот сеанс.

При использовании только 1 эталонного сеанса получаемая оценка представляет собой действительное число в интервале от 0 до 1. Относительно предложенной гибридной меры, возможно применение интерпретации «чем больше оценка, тем лучше». При сравнении сеансов

между собой более качественно произнесенным признается сеанс с большей оценкой. При использовании такой модификации в рамках оценки изменения речи в процессе речевой реабилитации можно говорить только о самом факте наличия изменений и относительной близости к эталонным записям.

3.3 Выводы по главе

Предложенный алгоритм нахождения количественного значения схожести позволяет оценить подобие двух речевых сигналов различных длин. Сформированная гибридная мера, основанная на расчете количественных значений по трем выбранным метрикам (DTW-расстояние, коэффициент корреляции, расстояние Минковского) и применении нечеткого классификатора на основе генетического алгоритма, позволяет получить оценку в известном интервале значений. Были построены три классификатора для каждой из групп проблемных фонем, на вход которым подаются три рассчитанных значения по метрикам, на выходе количественная оценка схожести в интервале от 0 до 1 и метка класса (для проведения соответствия с оценками по старой методике). Для классификатора, построенного для группы фонем [к] и [к'], точность составила $0,87 \pm 0,021$ для тестовой и 0,88 для обучающей выборки. Для классификатора группы фонем [с] и [с'] соответствующие точности составили $0,84 \pm 0,022$ и 0,85. Для классификатора группы фонем [т] и [т'], точность составила $0,79 \pm 0,025$ для тестовой и 0,81 для обучающей выборки. Точность классификации была получена на 10-кратной кросс-валидации. Средняя точность для всех фонем составила 0,83.

Предложена методика оценки схожести звуковых сигналов с применением нескольких опорных сигналов. За счет применения нескольких опорных сигналов производится учет вариативности произношения. Также методика позволяет учитывать индивидуальные

особенности речи пациента, имеющиеся до проведенного оперативного вмешательства, например, такие особенности как картавость, гнусавость и другие [168]. При проведении сравнения с максимально правильным произношением другого диктора приводит к существенному занижению получаемых оценок. Однако этот вариант может быть рассмотрен при отсутствии возможности проведения сеанса записи до оперативного вмешательства при учете описанных выше особенности. Описана модификация методики с условием наличия только одного опорного сигнала.

4 Методика оценки речевой реабилитации с использованием получаемых оценок и биологической обратной связи

4.1 Использование биологической обратной связи в процессе речевой реабилитации

Метод биологической обратной связи (БОС) заключается в графическом или аудио представлении пациенту на устройствах вывода значений его физиологических показателей, определяемых клиническим протоколом [169 – 171]. В рамках БОС могут быть представлены как текущие значения показателей пациента, так и считываемые в течение некоторого момента времени до отображения. Под протоколом понимается набор условий, обуславливающих процедуру БОС. Все протоколы биологической обратной связи разделяются на два направления [172 – 176]:

— “neurofeedback” – модификация различных параметров электроэнцефалограммы (ЭЭГ) головного мозга (амплитуды, мощности, когерентности основных ритмов ЭЭГ – обозначается также термином “neurotherapy”);

— “biofeedback” – модификация показателей вегетативной (симпатико-парасимпатической) активации (проводимость кожи, кардиограмма, чсс, дыхание, параметры речи и голоса, электромиограмма, температура, фотоплетизмограмма и др.)

В данной работе под биологической обратной связью (БОС) рассматриваются методы, включенные во второе направление, а именно оценка параметров речи.

В рамках реабилитации использование БОС направлено на активизацию внутренних резервов организма для восстановления или совершенствования физиологических навыков. Его конечная цель — эффективная саморегуляция важных физиологических функций организма

[172, 177 – 180]. Исследования показали, что эффективность БОС в реабилитации зависит от степени мотивации, социальной адаптации, коммуникабельности и мало связана с возрастом, течением болезни, социальными факторами [181]. Применимость и эффективности использования принципа методов БОС в процессе речевой реабилитации и работе с речевыми нарушениями описана в [172, 182 – 184].

БОС всегда включает компонент подкрепления, при этом подкрепление может быть направлено на удовлетворение не только первичных потребностей, но и потребности в информации, в оценке, в положительной самооценке и т.д. Таким образом, само получение информации об успехе или неудаче может служить подкреплением [185].

Различаются следующие виды подкреплений:

- положительное подкрепление – в качестве подкрепления используются фото, видео и/или аудио образы, вызывающие позитивную реакцию организма;

- отрицательное подкрепление – в качестве подкрепления используются фото, видео и/или аудио образы, вызывающие негативную реакцию организма;

- смешанное подкрепление – использование положительного и отрицательного подкрепления в зависимости от рассчитанных параметров (например, при достижении заданной громкости голоса – позитивное подкрепление, в случае недостижения – отрицательное).

Подкрепление может состоять из одно или нескольких стимулов (сигналов), каждый из которых представляет собой отдельный образ (картинка, цвет, звук, видео). Повышенная эффективность какого-либо из видов подкреплений в отношении остальных, а также преимущества какого-либо из каналов восприятия информации над другими не выявлена в исследованиях [186 – 189].

Для процесса речевой реабилитации предложена методика оценки качества речи с возможностью формирования положительных

подкрепляющих стимулов на основе подсчета количественных оценок качества произношения слогов по методике оценки схожести звуковых сигналов. Существующая методика оценки голоса в процессе голосовой реабилитации описана в [182, 190]. Для применения в речевой реабилитации была проведена адаптация в рамках изменения следующих моментов:

1) объект оценки сменился с голоса на речь, соответственно вместо расчетов параметров голоса (ЧОТ, длительность фонации) рассчитываются оценки качества речи;

2) параметры принятия решения при формировании определенного типа подкреплений (решение о формировании отрицательного или положительного подкрепления) теперь рассчитываются индивидуально для каждого пациента с учетом его особенностей речи (ранее рассчитывались общие значения для всех пациентов);

3) в качестве подкрепления предложено формирование двух видов визуальных подкреплений: кратковременное и долговременное. Данная терминология отражает время отображения сформированного подкрепления пациенту. В методике голосовой реабилитации использовался один тип подкрепления, состоящий из аудио и визуального стимулов.

Сама методика реабилитации с точки зрения медицинских мероприятий и применение методики оценки качества речи в качестве ее составной части описаны в [191].

Применение оценки качества речи и реализация принципа биологической обратной связи предлагается с использованием следующих шагов. Здесь и далее рассматривается вопрос оценки качества речи и использования методики оценки в рамках реабилитации, медицинские аспекты реабилитации не включены.

1. Запись двух сеансов оценки произношения слогов согласно составленному списку слогов до оперативного вмешательства.

2. После проведение операции, на первом приеме специалиста-логопеда запись сеанса оценки произношения слогов, подсчет оценки сеанса

на основе сравнения его с эталонными сеансами и подсчет среднего. Если оценка сеанса равна 1, то с точки зрения количественной оценки качества речи можно говорить о том, что необходимости в проведении речевой реабилитации нет, но окончательное решение данного вопроса остается на усмотрение лечащего врача-логопеда-онколога. Если уровень ниже, то пациенту рекомендуется пройти полную речевую реабилитацию.

3. При стандартной продолжительности речевой реабилитации на каждом следующем приеме проведение записи сеанса оценки произношения слогов, сравнение с оценками предыдущего сеанса, вывод динамики восстановления речи.

На основе динамики у специалиста-логопеда есть возможность корректировать план по тренировке речи. Если по некоторым фонемам оценки стали ухудшаться относительно предыдущих сеансов, то согласно рекомендациям по использованию программного комплекса предлагается скорректировать тренировку в сторону увеличения количества упражнений на проработку данных фонем. Если оценка фонемы достигла эталонного уровня (средняя оценка по фонеме равна 1), уменьшить количество упражнений. Совсем убирать упражнения на фонему не рекомендуется из-за необходимости комплексности тренировок, окончательное решение принимается специалистом-логопедом.

Биологическую обратную связь предлагается использовать в рамках реализации этапа 3. Схема информационных потоков реализации БОС в составе речевой реабилитации представлена на рисунке 4.1.

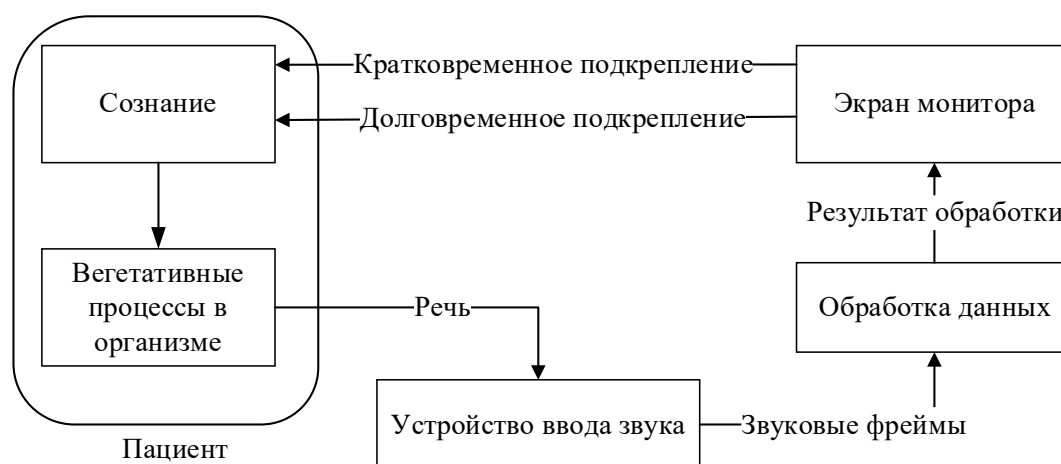


Рисунок 4.1 - Схема информационных потоков в методах БОС с использованием двух видов визуального подкрепления.

4.2 Формирование положительных подкрепляющих стимулов

Методика предполагает формирование двух видов подкреплений: кратковременное подкрепление и долговременное [192]. Кратковременное подкрепление предлагается использовать в процессе записи сеанса оценки произношения слогов для отражения сравнения оценки записи слога в текущем сеансе и одним из предыдущих. Долговременное подкрепление предлагается для отображения усредненных результатов по всем проведенным сеансам оценки произношения слогов в течение речевой реабилитации.

4.2.1 Кратковременные подкрепляющие стимулы

Кратковременные подкрепляющие стимулы (подкрепления) основаны на сравнении оценки записи произношения пациентом конкретного слога в текущем сеансе и оценки записи произношения этого же слога в одном из предыдущих сеансов.

Подкреплением является цветовая окраска оценки предыдущей записи, отражаемая в режиме реального времени в процессе проведения сеанса

записи слогов. При переходе на следующий слог стимул формируется заново. Такой вид стимулов позволяет удовлетворять потребность в мгновенной оценке в течение сеанса.

При отсутствии необходимости в реализации БОС оценка отображается на экране стандартным черным цветом (как и остальной текст на экране). В случае реализации БОС оценка может быть окрашена в два цвета:

— если оценка текущей записи слога меньше оценки этого же слога в выбранном для сравнения одном из предыдущих сеансов, то оценка отображается в стандартном черном цвете (нет явной реализации отрицательного подкрепляющего стимула).

— если оценка текущей записи больше или равна оценке этого же слога в выбранном для сравнения одном из предыдущих сеансов, то оценка отображается в зеленом цвете (сформированный положительный подкрепляющий стимул).

Пример отображения оценки представлен на рисунке 4.2.

Оценка: 0,7170

а – вид оценки при наличии положительного подкрепления

Оценка: 0,3420

б – вид оценки при отсутствии положительного подкрепления

Рисунок 4.2 – пример отображения оценки в случае реализации БОС

Обозначение положительного подкрепления зеленым цветом обосновано тем, что зеленый цвет нормализует давление, приводит к нормализации дыхания и пульса, создает несильный, но прочный подъем умственной работоспособности, благоприятствует концентрации внимания, число правильно решенных задач увеличивается на 10% при сокращении числа ошибок на 20% [193]. Также зеленый цвет считается цветом «правильности, одобрения», ассоциируется со здоровьем [194].

Структура процесса записи слога, его оценки и БОС представлена в виде NS-диаграммы на рисунке 4.3.



Рисунок 4.3 – Последовательность действий процесса записи слога с формированием кратковременного подкрепления.

Для реализации подсчета оценок и формирования подкрепляющих положительных стимулов необходимо соблюдение следующих правил:

- Проводимый сеанс является минимум вторым после операции, то есть в базе данных уже есть данные о предыдущих минимум 3 сеанса (два сеанса до операции – эталонные и один сеанс после операции);
- Отмечена необходимость в проведении оценки сеанса в режиме реального времени (более подробно описано в Главе 5 Раздел 5.1.4);
- Отмечена необходимость в реализации биологической обратной связи (более подробно описано в Главе 5 Раздел 5.1.4);
- Выбран сеанс для сравнения в рамках реализации биологической обратной связи, этот сеанс должен быть уже оценен и его оценки записаны в базе данных;

— Для проведения оценки выбраны те же эталонные сеансы, что и для сеанса, с которым производится сравнение в рамках реализации БОС.

4.2.2 Долговременные подкрепляющие стимулы

Долговременные подкрепляющие стимулы (подкрепления) основаны на сравнении усредненных по фонемам или группам фонем оценок из разных сеансов. Эти стимулы формируются после проведения сеанса записи слогов и отражают динамику изменения усредненных оценок по всем выбранным сеансам. В зависимости от особенностей речи пациента можно проводить усреднение как по отдельным проблемным фонемам, так и по группам проблемных фонем. Такой вид подкрепления направлен на оценку общего состояния речи пациента и позволяет повысить мотивацию пациента к дальнейшим тренировкам по восстановлению речи. На основе динамики логопед может при необходимости корректировать программу тренировок в случае отсутствия положительной динамики по какой-либо фонеме. Само подкрепление представляет собой графическое представление оценок и динамики их изменений – пример на рисунке 4.4.

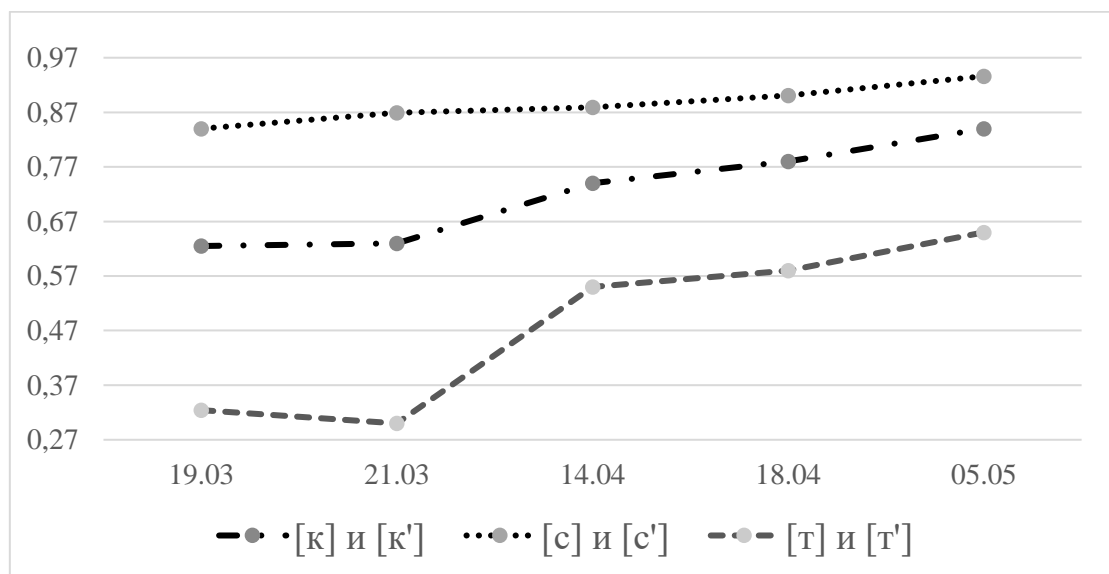


Рисунок 4.4 - Долговременное подкрепление – отражение динамики изменения оценок по сеансам.

В случае долговременного подкрепления это подкрепление является смешанным, так как его нельзя заранее характеризовать как положительное или отрицательное. Вид стимула зависит от динамики оценок: если оценки уменьшаются, то можно назвать стимул отрицательным, если оценки увеличиваются, то положительным. Также нет графического различия в отображении отрицательных и положительных стимулов.

Процесс построения долговременного подкрепления выполняется с учетом нижеперечисленных правил.

1. Для отображения выбираются сеансы, записанные с использованием одного и того же списка слогов.

2. Для отображения выбираются уже оцененные сеансы. В случае, если необходимо отобразить не оцененный сеанс, перед построением проводится оценка сеанса.

3. Для отображения выбираются сеансы, оцененные с использованием одного и того же набора эталонных сеансов (одно и того же эталонного сеанса – в случае использования только одного эталонного сеанса).

4. При использовании нескольких эталонных сеансов в отображения не включаются сами эталонные сеансы.

Алгоритм построения долговременного подкрепления с точки зрения специалиста-логопеда представляет собой выполнение следующей последовательности шагов.

1. Выбрать сеансы для отображения на графике, представляющем собой долговременное подкрепление. На усмотрение специалиста-логопеда возможно включение не всех имеющихся сеансов пациента, а выбранных по некоторым условиям.

2. Проверить выбранные сеансы на соблюдение описанных выше правил.

3. Выбрать порядок построения динамики оценок. На усмотрение специалиста-логопеда возможно построение динамики от времени начала реабилитации до даты окончания и наоборот.

4. Определить необходимость объединения проблемных фонем в группы.

5. Запустить процесс отображения динамики оценок.

6. Сохранить изображение динамики (при необходимости).

Алгоритм построения графика с динамикой оценок (долговременное подкрепление) с точки зрения вычислений представляет собой следующую последовательность шагов. Здесь рассмотрен процесс с использованием сеансов, записанных на основе списка слогов наиболее подверженных изменениям (описан в главе 2 разделе 2.1).

1. Выбрать сеанс из списка выделенных для построения динамики сеансов.

2. Из базы данных сохранить оценки записей сеанса, выбранного в пункте 1.

3. В зависимости от необходимости объединения фонем в группы:

3.1. При необходимости объединения в группы – усреднить значения по группе фонем, т.е. найти средние значения для записей с 1 по 30, с 31 по 60, с 61 по 90 (группы фонем [к] и [к'], [с] и [с'], [т] и [т'] соответственно).

3.2. При работе с отдельными фонемами – усреднить значения группы записей, содержащих проблемную фонему, то есть найти средние значения для записей с 1 по 15 для фонемы [к], с 16 по 30 для фонемы [к'], с 31 по 45 для фонемы [с], с 46 по 60 для фонемы [с'], с 61 по 75 для фонемы [т], с 76 по 90 для фонемы [т'].

4. Добавить усреднённые значения с указанием даты проведения сеанса в массив данных для отображения.

5. Повторить пункты 1–4 для всех выделенных сеансов.

6. Отобразить массив данных на графике.

Более подробно алгоритм представлен в виде NS-диаграммы на рисунке 4.5.

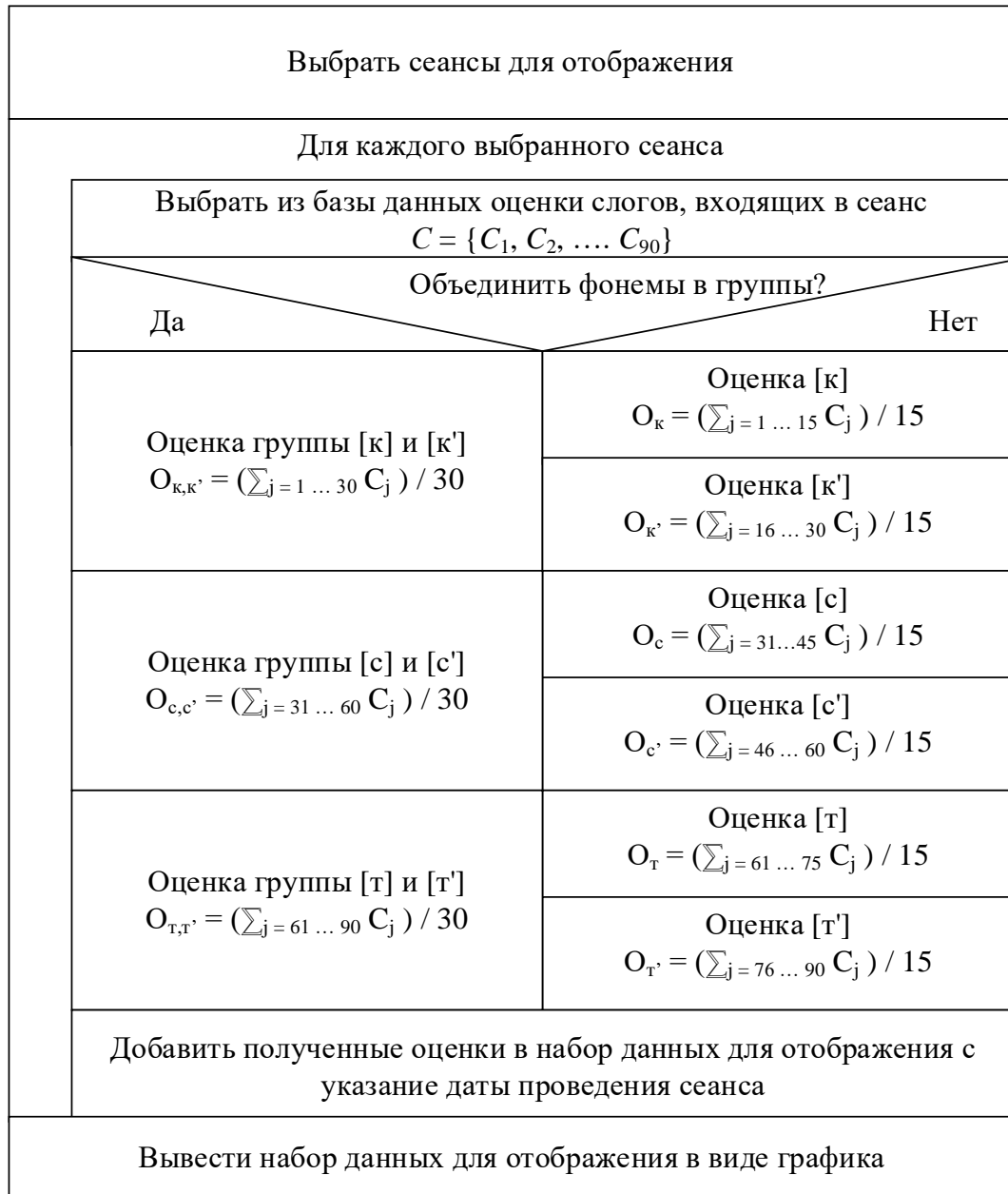


Рисунок 4.5 – Расчет значений, формирующих долговременное подкрепление

Вид долговременного подкрепления, построенного на синтетически сформированном наборе данных (набор описан в Главе 5 Раздел 5.1.4), представлен на рисунках 4.6–4.7. Графическое представление (осциллограммы и спектрограммы) аудиозаписей произношения слога «куп» из каждого из сеансов сформированного набора представлено в Приложении Ж.

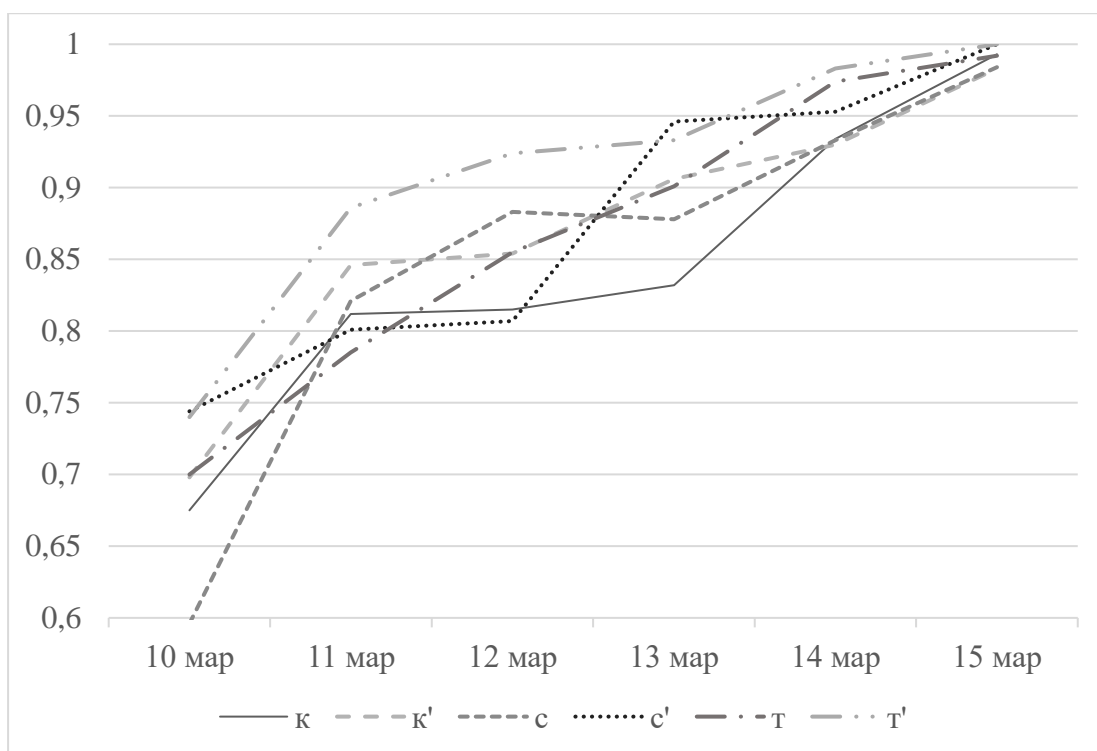


Рисунок 4.6 – Долгосрочное подкрепление – изображения динамики средних оценок по отдельным проблемным фонам

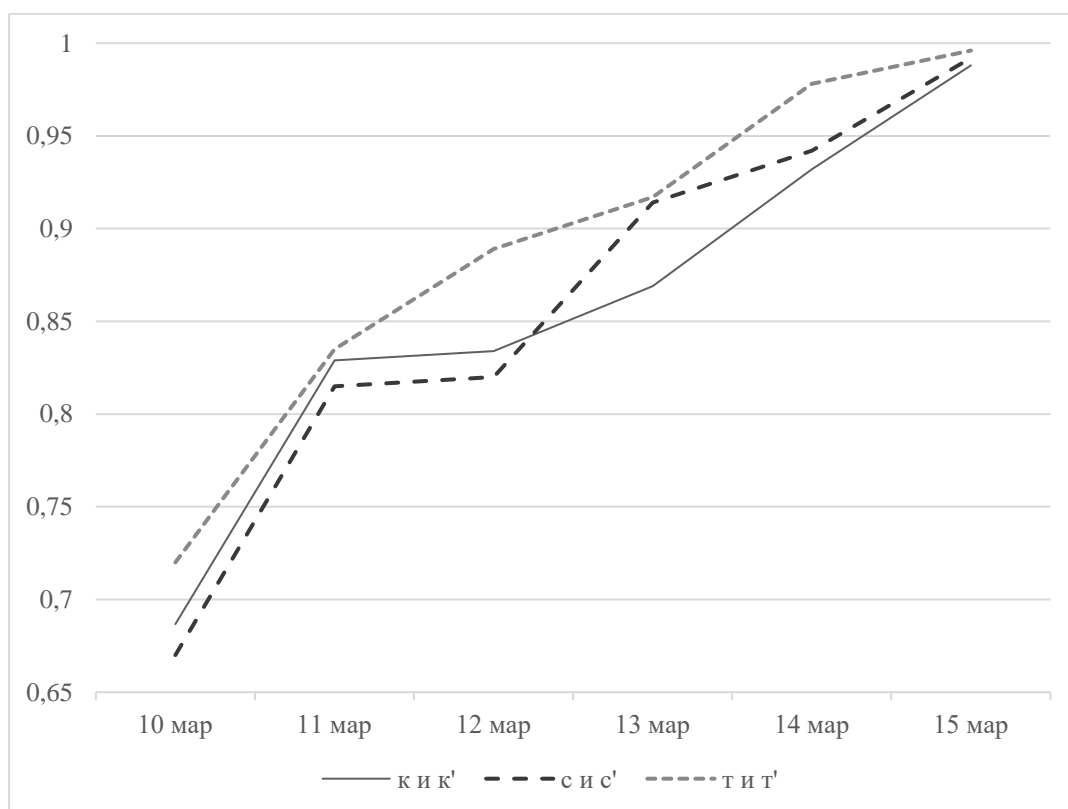


Рисунок 4.7 – Долгосрочное подкрепление – изображения динамики средних оценок по группам проблемных фонем

4.3 Оценка качества речи в рамках речевой реабилитации с использованием биологической обратной связи

Процесс речевой реабилитации представляет собой сочетание приемов специалиста-логопеда и самостоятельные тренировки пациента по составленному логопедом плану. Прием специалиста-логопеда включает себя следующие моменты:

- проведение медицинских процедур;
- проведение тренировок речи;
- оценка качества речи.

В оценку качества речи включается запись речи пациента (под речью понимается произношение слогов, фраз и других элементов речи) и непосредственно оценка записанных аудиозаписей.

Использование описанных выше алгоритма и методик предполагается в рамках оценки качества речи. Программный комплекс, содержащий в себе реализацию алгоритма и методик, (подробно описан в главе 5) позволяет минимизировать участие специалиста-логопеда в пункте оценки качества речи. Методика оценки качества речи в процессе реабилитации заключается в следующей последовательности шагов.

1. После проведения медицинских процедур и тренировок речи провести сеанс записи речи пациента для оценки качества речи.

2. В процессе записи сеанса провести оценку качества произношения слогов в режиме реального времени с формированием кратковременного подкрепления биологической обратной связи для отражения изменений в речи пациента и повышения его мотивированности в дальнейшей работе.

3. После окончания записи продемонстрировать пациенту динамику оценок его речи, тем самым реализовав долговременное подкрепление биологической обратной связи, дать пояснения и при необходимости скорректировать план самостоятельных тренировок.

В рамках такой реализации используются оба предложенные типа подкреплений – кратковременные подкрепляющие стимулы и долговременные подкрепляющие стимулы.

Возможна модификация предложенной реализации, она представляет собой последовательное выполнение нижеописанных шагов.

1. После проведения медицинских процедур и тренировок речи провести сеанс записи речи пациента без проведения оценки в режиме реального времени.

2. После записи сеанса провести оценку сеанса.

3. После окончания подсчета оценок продемонстрировать пациенту динамику оценок его речи, дать пояснения и при необходимости скорректировать план самостоятельных тренировок пациентов.

В рамках этой модификации используется только долговременное подкрепление – отражение динамики оценок речи.

За счет использования программного комплекса, в котором реализуются описываемые алгоритмы и методики, сокращается время, затрачиваемого на оценку качества речи.

Было оценено время, затрачиваемое на оценку качества речи в процессе речевой реабилитации. Был проведен замер времени, которое требуется специалисту на оценку сеанса по старой методике (адаптация ГОСТ Р 50840–95), а также времени проведения оценки с использованием описываемых алгоритма и методик (новая методика). Время измерялось в секундах, было проведено 49 подсчетов для обоих способов оценки. При использовании методики оценки схожести расчеты производились при наличии двух эталонных сеансов (при модификации методики под один эталонный сеанс время оценки существенно ниже).

На полученных двух выборках были подсчитаны средние значения и проверена гипотеза о равенстве средних значений по критерию Стьюдента [136]. Средние значения приведены в таблице 4.1. Значение критерия Стьюдента для полученных выборок равно 26,77, критическое значение 1,98

(уровень значимость принят 0,05), следовательно, гипотеза о равенстве средних отклоняется.

Таблица 4.1 – Средние значения времени на оценку одного сеанса.

	Старая методика	Новая методика
Среднее значение времени оценки 1 сеанса (в секундах)	371,55	130,57

В результате применения методики оценки качества речи время, затрачиваемое на оценку уже записанного (и внесенного в базу данных) сеанса, сократилось на 64,86%.

Если говорить про весь комплекс проводимых мероприятий (запись сеанса, его обработка и оценка), то при проведении «ручной» оценки качества речи по старой методике (описана в главе 1 разделе 1.4) в присутствии пациента проводилась только запись слогов, оценка производилась в свободное от приема пациентов время. Время оценки существенно зависело от качества речи пациента, также сюда включалось время внесения оценок в базу данных. При использовании программного комплекса время проведения оценки не в процессе записи практически не изменяется в зависимости от речи пациента, в процессе записи зависит от количества ошибок в течение записи, однако разбег во времени не такой значительный, как при ручном методе. Сводные значения времени, затрачиваемом на оценку, представлены в таблице 4.2.

Таблица 4.2 – Сравнение времени проведения сеансов оценки произношения слогов по старой и новой методикам (время на один сеанс из 90 слогов)

		Старая методика	Новая методика
Время записи сеанса без озвучки* и без оценки	среднее	3 мин 18 с	2 мин 45 с
	интервал	2 мин 46 с - 5 мин 15 с	2 мин 33 с - 3 мин 1 с
Время записи сеанса с озвучкой* и без оценки	среднее	5 мин 2 с	4 мин 48 с
	интервал	4 мин 43 с – 6 мин 54 с	4 мин 30 с – 5 мин 9 с

Продолжение таблицы 4.2.

		Старая методика	Новая методика
Время записи сеанса с озвучкой*, оценкой и БОС**	среднее	15 мин ***	6 мин 39 с
	интервал	8 мин – 20 мин	6 мин 00 с – 7 мин 04 с

* под озвучкой понимается: в случае адаптированной методики – озвучка специалистом-логопедом правильного произношения слога, в случае методики оценки качества речи и использования ПО – демонстрация по аудиоканалу правильного произношения слога.

** реализация БОС производилась только при использовании методики оценки качества речи.

*** суммарное время проведения следующих процедур: записи сеанса, обработка аудиозаписей, оценка аудиозаписей путем прослушивания, внесение оценок в систему. Оценка времени производилась специалистом-логопедом.

Время на запись сеанса без оценки с использованием новой методики и программного комплекса сократилось на 16,7% без озвучки и на 5% с озвучкой. Общее время на проведение записи сеанса сократилось на 55,7% (с 15 минут до 6 минут 39 сек).

4.4 Выводы по главе

Предложена методика оценки речевой реабилитации с использованием количественных оценок и применением метода бионического принципа биологической обратной связи. Методика является адаптацией успешно применяемой методики голосовой реабилитации после ларинготомии, также используемой на базе НИИ Онкологии г. Томска. Произведена смена объекта оценки с голоса на речь. Предложенный алгоритм расчета пороговых значений для принятия решения о формировании или отсутствии формирования положительного подкрепления в рамках биологической обратной связи является индивидуальным для каждого пациента. Предложено использование двух видов визуального подкрепления: кратковременного и долговременного. Кратковременное подкрепление

формируется как положительное подкрепление при увеличении оценки произношения слога в текущем сеансе в сравнении с оценкой этого же слога в предыдущем сеансе. Подкрепление представляет собой цветовую окраску демонстрируемой оценки в зеленый цвет. При отсутствии увеличения оценки подкрепление не формируется, и оценка демонстрируется в стандартном черном цвете. Долговременное подкрепление представляет собой отражение динамики усреднённых по фонемам или группам фонем оценок произношения слогов. Такое подкрепление позволяет отслеживать изменения в речи пациента и способствует повышению мотивации к дальнейшим самостоятельным тренировкам пациента. Применение предложенной методики оценки позволило сократить время проведения оценки произношения более чем на 64%, что повысило эффективность процесса речевой реабилитации.

5 Разработка и внедрение программного комплекса

5.1 Структура базы данных записей и программного обеспечения для записи и оценка качества речи

Для использования описанных ранее алгоритма и методик в процессе речевой реабилитации они были реализованы в составе программного комплекса по оценке качества речи. Разработанный программный комплекс представляет собой базу данных и программу для работы с базой данных.

5.1.1 База данных записей пациентов

На момент начала исследований на базе НИИ Онкологии использовалась база данных записей пациентов [195], однако ее структура требовала доработки для хранения всей необходимой информации. На разработанную базу данных было получено свидетельство о регистрации базы данных (Приложение Б) [196]. В базе данных содержится краткая информация по каждому пациенту, информация о сеансах оценки качества речи, информация о записях слогов и фраз, согласно адаптированной методике ГОСТ Р 50840–95, информация о записях слогов по составленному списку с наиболее проблемными фонемами, подсчитанные оценки сеансов. Частичная структура базы данных в виде ERR – диаграммы представлена на рисунке 5.1. Приведена часть используемой базы данных, которая имеет отношение к решаемой практической задаче.

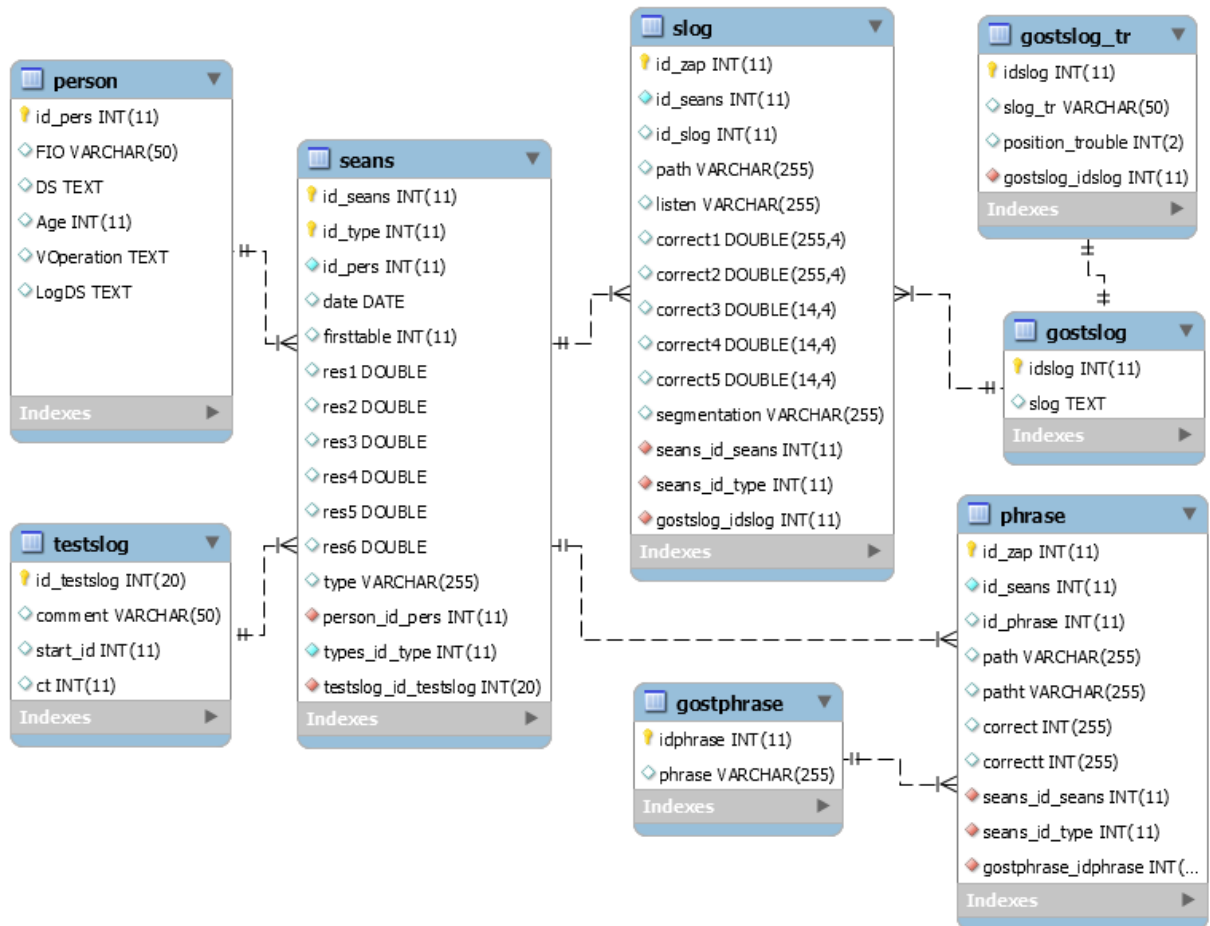


Рисунок 5.1 – Частичная структура доработанной базы данных в виде ERR – диаграммы.

Для решения поставленной задачи были сформированы и доработаны основные сущности и содержащиеся в них поля. Сущность «Пациент» (person) предназначена для хранения информации для идентификации пациента и его медицинских данных. Сущность представлена следующими полями:

- уникальный номер пациента в базе данных;
- фамилия имя отчество пациента;
- диагноз;
- год рождения пациента;
- объем хирургического вмешательства;
- логопедический диагноз.

Сущность «Сеанс» (seans) предназначена для хранения информации о проведенном сеансе оценки речи в рамках речевой реабилитации. Сущность представлена следующими полями:

- уникальный номер сеанса в базе данных;
- номер типа сеанса: в соответствии со списком типов сеанса в базе данных;
- идентификационный номер пациента;
- дата проведения сеанса;
- номер первой таблицы (для проведения сеансов согласно ГОСТ Р 50840–95);
- поля хранения оценок сеанса;
- текстовое поле для комментариев типа проводимого сеанса: уточнение времени и обстоятельств проведения записи.

Сущность «Списки слогов для записи» (testslog) предназначена для хранения различных списков слогов для записи сеансов слоговой разборчивости и оценки качества произношения слогов. Поля сущности:

- уникальный номер списка;
- название списка слогов;
- номер первого слога из общего набора слогов;
- количество слогов в списке.

Сущность «Сеанс записи слогов» (Slog) предназначена для хранения информации о записи конкретного слога в рамках проведения сеансов. Поля сущности:

- уникальный номер записи;
- номер сеанса, которому принадлежит запись;
- номер слога из списка слогов для записи;
- путь до аудиозаписи;
- поля для оценок;
- поле для сегментации.

Сущность «Слоги» (gostslog) предназначена для хранения общего набора слогов, доступных для записи. Поля сущности:

- номер слога в общем списке;
- сам слог.

Сущность «Транскрипции слогов» (gostslog_tr) предназначена для хранения транскрипций слогов, доступных для записи, а также обозначения расположения проблемной фонемы в слоге. Поля сущности:

- номер слога в общем списке;
- транскрипция слога;
- позиция проблемной фонемы в слоге.

Текущая структура базы данных является дополненной структурой базы данных, существовавшей на момент начала исследований [197]. Модернизация старой части структуры не производилась ввиду наличия другого программного средства, которое взаимодействует с базой данных.

Некоторый набор слогов составляет сеанс, по его итогам выставляется средняя оценка качества произнесения, сравнение таких оценок позволяет оценивать динамику изменения качества речи пациентов в процессе всего процесса реабилитации. Программа обеспечивает взаимодействием (облегчением взаимодействия) специалиста-логопеда с базой данных, а также выполняет количественную оценку путем попарного сравнения записей одного и того же слога в разных сеансах (оценка фразовой разборчивости хоть и предусмотрена ГОСТ Р 50840–95 и имеются записанные сеансы со фразовой разборчивостью, но в рамках данного исследования задачи оценки фразовой разборчивости не стоит).

5.1.2 Программа для оценки качества речи «Разборчивость речи»

Общая задача, для решения которой разрабатывался программный комплекс, состоит в сокращении времени, затрачиваемого на обработку

сеансов оценки качества речи пациентов в процессе реабилитации, что, соответственно, приводит к достижению цели использования программного комплекса - сокращению общего времени реабилитации. Были сформулированы подзадачи программы (как части программного комплекса) в рамках решаемой в работе задачи:

- осуществление количественной оценки произношения слогов;
- осуществление записи слогов с их последующей обработкой для занесения в базу данных;
- осуществление взаимодействия пользователя (специалиста-логопеда) и базы данных.

Были сформулированы следующие функции программы:

- обработка информации, содержащей в базе данных (в том числе создание новых записей, редактирование и удаление информации);
- воспроизведение записанных слогов, информация о которых содержатся в базе данных;
- экспертное оценивание качества произношения слогов (наличие возможности экспертного оценивания);
- запись слогов в рамках сеанса с последующим занесением информации о сеансе и записях в базу данных;
- количественное оценивание качества произношения слогов в режиме реального времени при записи слогов с применением методов бионического принципа биологической обратной связи (БОС);
- количественное оценивание сеанса записи путем сравнения слогов в сеансе со слогами в эталонном сеансе (возможность оценить ранее записанный сеанс полностью);
- подсчет средней оценки за сеанс на основе оценок слогов в этом сеансе

– графическое отображение динамики количественных оценок записей слогов в нескольких сеансах в рамках речевой реабилитации как подкрепления в рамках реализации методов бионического принципа БОС.

Некоторые моменты разработки и реализации описаны в [198, 199, 200, 201]. На главной форме приложения «Разборчивость речи» пользователю предлагается выбрать интересующего его пациента из базы пациентов, отображенной в таблице, и перейти к сеансам этого пациента. Под пользователем понимается специалист-логопед. По каждому пациенту представлена краткая информация: ФИО, год рождения, диагноз и логопедический диагноз, а также информация о проведенном лечении. Также на главной форме есть возможность настройки параметров записи голоса для адаптации к используемому устройству ввода (микрофону). Главная форма представлена на рисунке 5.2.

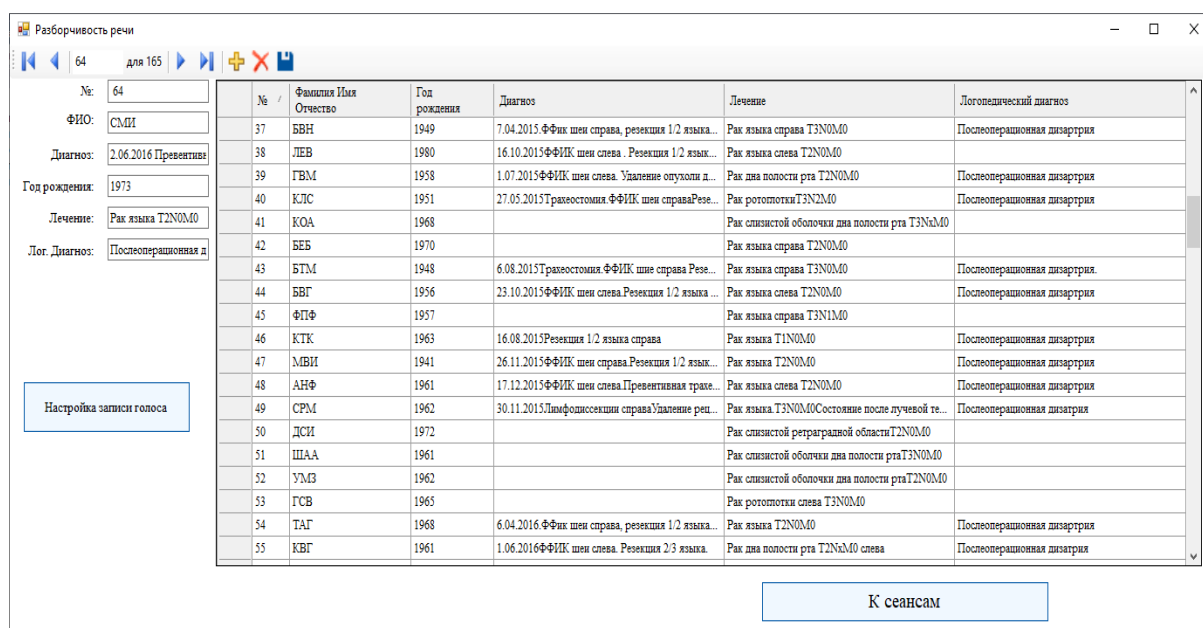


Рисунок 5.2 – Главная форма приложения «Разборчивость речи»

У каждого пациента в процессе реабилитации появляется несколько сеансов записи. На форме «Сеансы» пользователю предоставляется возможность просмотра, выбора, редактирования и удаления существующих сеансов пациента, выбранного на главной форме, а также добавление для него новых сеансов. Внешний вид формы «Сеансы» представлен на рисунке 5.3. В полях res1 и res2 содержатся оценки сеанса –

относительная оценка сеанса согласно методике, описанной в Главе 3 разделе 3.2 и оценка сеанса по старой методике (основанной на ГОСТ Р 50840–95 – экспертные оценки) соответственно. Для перехода к просмотру конкретного сеанса требуется выделить этот сеанс и нажать кнопку «Перейти к сеансу». Для создания нового сеанса этого пациента необходимо указать дату проведения сеанса (по умолчанию – текущая дата) и список слогов для записи (список слогов с наиболее проблемными фонемам обозначен как ОнкоЯзык) и нажать кнопку «Создать новый сеанс». Программа выведет сообщение об успешном создании сеанса и предложит приступить к записи слогов.

Также на этой форме доступно построение долгосрочного подкрепления в рамках реализации метода бионического принципа БОС. Для этого требуется выделить необходимые для отображения сеансы и нажать кнопку «Построить графики».

id_seans	id_type	date	res1	res2
1319	10	09.03.2022	0,857	
1324	10	10.03.2022	0,693	
1325	10	11.03.2022	0,855	
1326	10	12.03.2022	0,762	
1327	10	12.03.2022	0,833	
1328	10	13.03.2022	0,894	
1329	10	14.03.2022	0,951	
1330	10	15.03.2022	0,992	

Рисунок 5.3 - Форма «Сеансы» для работы со списком сеансов пациента.

При нажатии на кнопку «Перейти к сеансу» открывается форму «Слоговая разборчивость» (Рисунок 5.4). В таблице можно получить информацию обо всех записях слогов, входящих в этот сеанс, а также прослушать конкретную аудиозапись путем двойного клика на строку с требуемой записью. Также в таблице отражаются оценки записей, сохраненные в базе данных. Сам процесс оценки описан в разделе 5.1.3. Поле Segmentation предназначено для сохранения интервалов сегментации слога. Поля correct1 и correct2 предназначены для хранения оценок записи.

id_zap	id_seans	id_slog	path	listen	correct1	correct2	segmentation
29884	1324	52	c:\Sound\Slog H...		0,641		
29885	1324	53	c:\Sound\Slog H...		0,838		
29886	1324	54	c:\Sound\Slog H...		0,702		
29887	1324	55	c:\Sound\Slog H...		0,81		
29888	1324	56	c:\Sound\Slog H...		0,816		
29889	1324	57	c:\Sound\Slog H...		0,571		
29890	1324	58	c:\Sound\Slog H...		0,993		
29891	1324	59	c:\Sound\Slog H...		1		
29892	1324	60	c:\Sound\Slog H...		0,627		
29893	1324	61	c:\Sound\Slog H...		0,925		
29894	1324	62	c:\Sound\Slog H...		0,68		
29895	1324	63	c:\Sound\Slog H...		0,89		
29896	1324	64	c:\Sound\Slog H...		0,591		
29897	1324	65	c:\Sound\Slog H...		0,838		

Рисунок 5.4 - Форма «Слоговая разборчивость» для работы со списком сеансов пациента.

5.1.3 Подсчет количественной оценки аудиозаписей сеанса

На форме «Слоговая разборчивость» предусмотрена возможность оценки всех аудиозаписей сеанса. Оценка, подсчитываемая в автоматизированном режиме, производится согласно алгоритму и методике, описанными в Главе 3.

Рассмотрим процесс оценки сеанса с использованием программного комплекса. Контекстная диаграмма процесса представлена на рисунке 5.5.

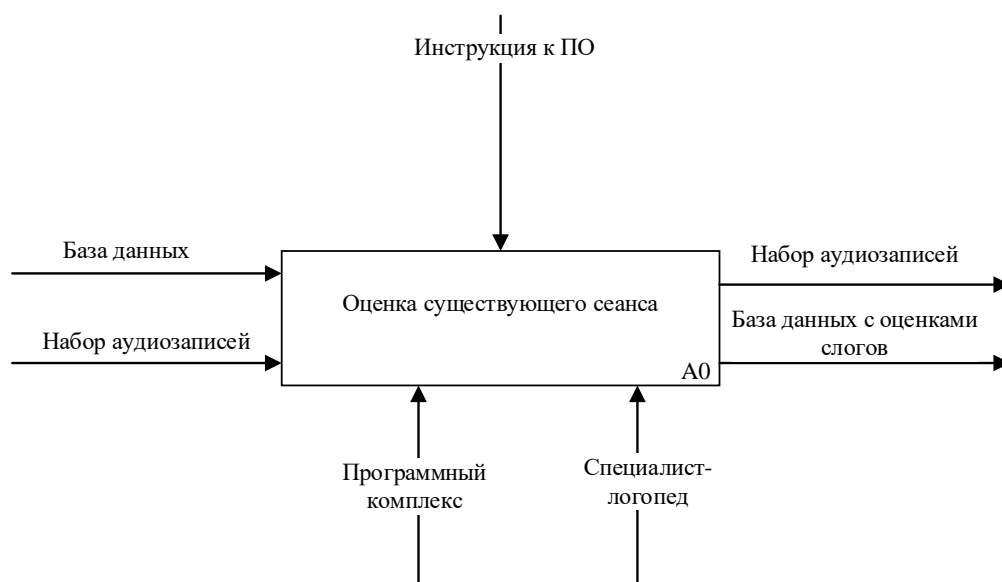


Рисунок 5.5 - Контекстная диаграмма процесса оценки уже занесенного в базу данных сеанса

Входными данными являются записи слогов для всех пациентов (интересующие нас будут выбираться в процессе оценки), а также

информация в базе данных. Здесь подразумевается, что пациент при оценке уже существующего сеанса никак не взаимодействует с системой, а всем управлением и настройками занимается специалист-логопед. В результате процесса в базе данных появляются количественные оценки качества произношения слогов.

В приложении И на рисунке И.1 представлена декомпозиция первого уровня, на которой более детально описан процесс оценки уже существующего сеанса. Пользователь системы (или, как ранее отмечалось, специалист-логопед) принимает участие лишь в настройке оценки, которая заключается в выборе сеанса с эталонным записям слогов. Посчитанные оценки отражаются в поле correct1 и автоматически сохраняются в базу данных. Для сохранения в базе данных средней оценки сеанса или же оценок, внесенных в поле correct2, требуется нажать кнопку «Сохранить».

Процесс оценки с использованием разработанного программного средства представлен на рисунках 5.6–5.10. Перед непосредственным проведением оценки требуется произвести настройку – выбрать сеанса для сравнения. Без выполнения настройки подсчет оценок недоступен.

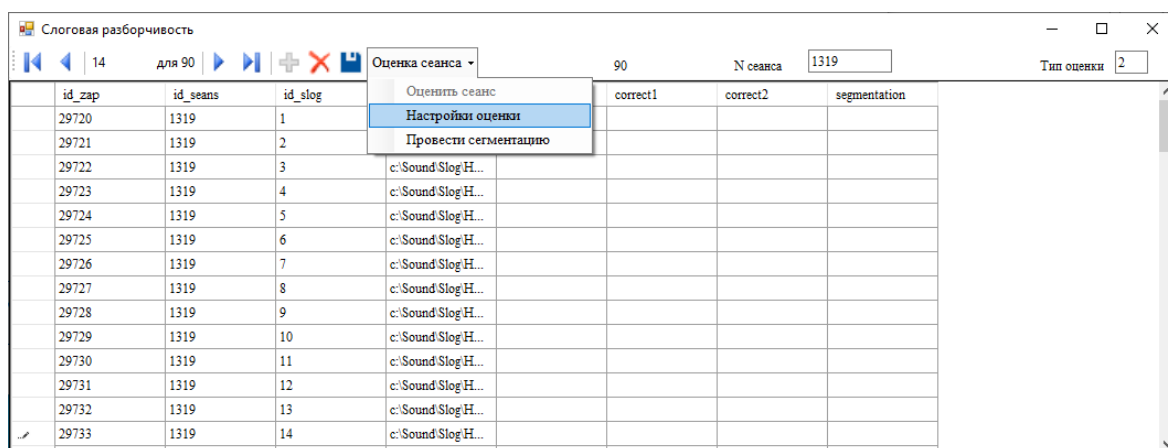


Рисунок 5.6 – Переход в меню настроек оценки

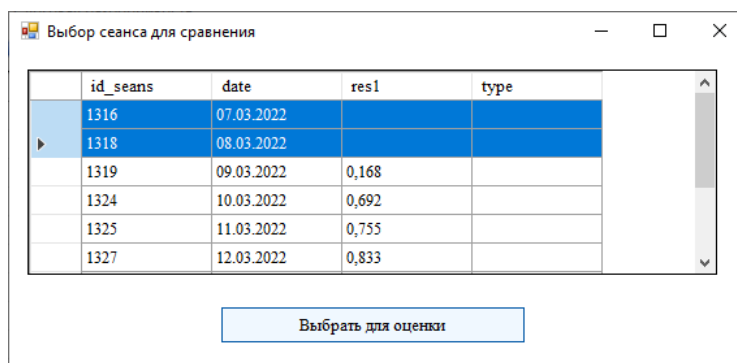


Рисунок 5.7 – Настройка оценки: выбор эталонных сеансов

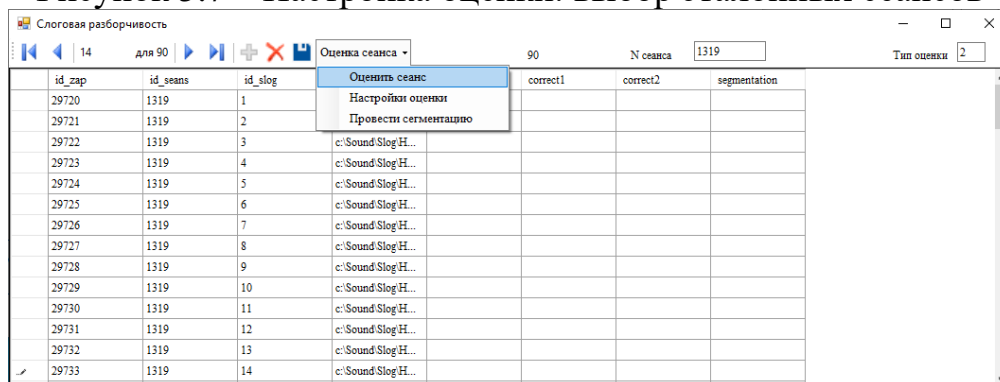


Рисунок 5.8 – Оценка сеанса: после выполнения настроек активна кнопка «Оценить сеанс»

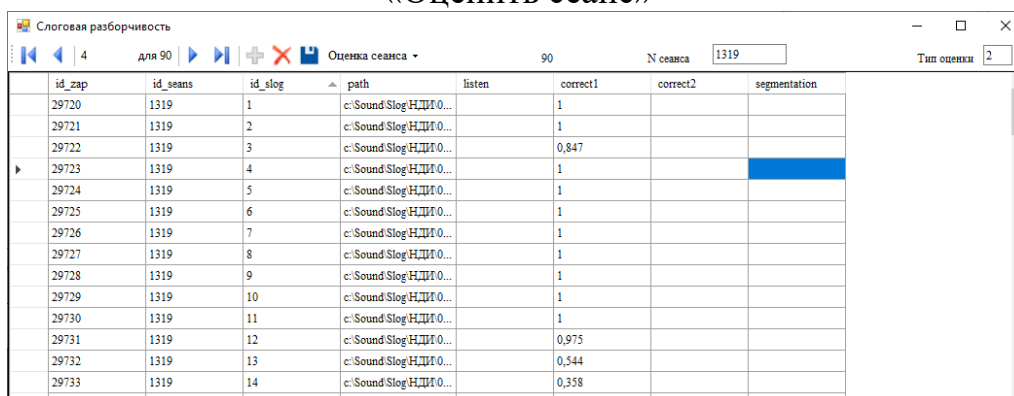


Рисунок 5.9 – Результат оценки: оценка для каждой записи

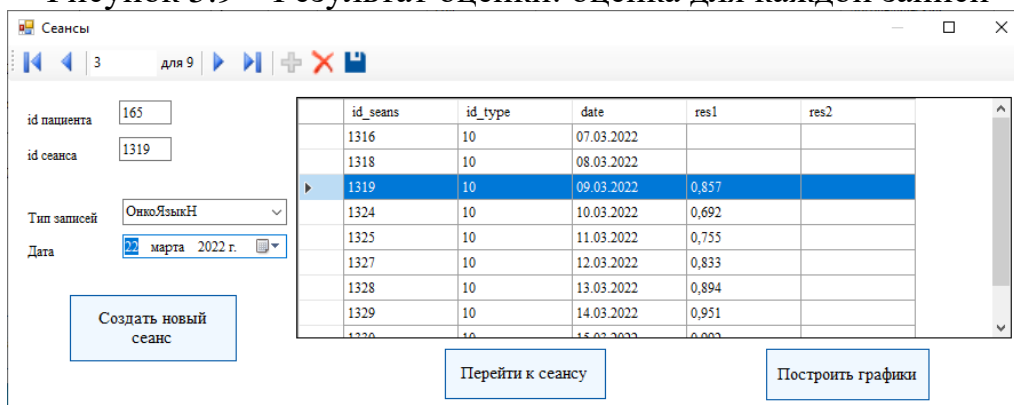


Рисунок 5.10 – Оценка сеанса: подсчет средней оценки сеанса

5.1.4 Запись сеанса оценки произношения слогов с проведением оценки и БОС

Запись сеанса оценки произношения слогов подразумевает под собой комбинацию следующих процессов:

- последовательная запись произношения слогов согласно списку слогов для записи с переключением между слогами в автоматическом режиме;
- (при необходимости) демонстрация правильного произношения по аудиоканалу;
- (при необходимости) оценка только что произведенной записи согласно алгоритму и методике;
- (при необходимости) реализация биологической обратной связи.

Для упрощения процедуры записи слогов во время сеанса в программе используется алгоритм обнаружения голосовой активности (voice activity detection - VAD) [202]. Его применение обосновывается в необходимости ускорения процесса записи слогов за счет его автоматизации: нажатие кнопки «Начать запись» является сигналом к началу записи, а взамен нажатия кнопки «Следующий слог» для перехода на следующий слог VAD сам определяет момент произношения слога, а также принимает решение о переходе на следующий слог. Для окончания записи сеанса необходимо нажать кнопку «Завершить сеанс записи», причем если был записан последний слог согласно типу записи и списку слогов, то программа выведет сообщение о завершении записи сеанса и необходимости нажатия кнопки завершения. Применяемый алгоритм VAD основан на сравнении параметров текущего фрейма звукового потока в буфере записи с заранее рассчитанными параметрами. Превышении порогового значения хотя бы по одному из используемых параметров является признаком наличия голоса в записанном фрейме аудиосигнала. В реализованном алгоритме VAD для подсчета используются следующие параметры:

– значение краткосрочной энергии сигнала:

$$E = \sum_{n=T}^{T+N-1} (s(n) * w(n))^2 \quad (5.1)$$

где $s(n)$ — амплитуда сигнала в отсчете n ,

$w(n)$ — оконная функция (в данной реализации использовано прямоугольное окно $w(n) = 1$),

N – размер окна сигнала, для которого рассчитывается E ,

размер перекрытия между окнами равен $N-1$.

– мера спектральной плоскостности (SFM — Spectral Flatness Measure):

$$SMF = 10 * \log_{10} \frac{Gm}{Am} \quad (5.2)$$

где Gm - среднее геометрическое спектра сигнала,

Am - среднее арифметическое спектра сигнала.

Любой громкий шум также будет воспринят как голос, поэтому предъявляются требования к записывающей аппаратуре, а также к помещению, в котором проводится речевая реабилитация и поведению пациента и логопеда на самом сеансе записи. После нажатия кнопки начала записи запускается VAD и «слушает» голос. Если голос не обнаруживается, запись не производится, если же VAD улавливает акустическую активность, начинается запись аудиофайла. Когда уровень звука падает до уровня тишины, запись прекращается, записанный аудиофайл сохраняется, и происходит переключение на следующий слог, который необходимо произнести.

Запись звука осуществляется при помощи сторонней библиотеки NAudio [203], которая была выбрана из-за прозрачности настройки обработчика записываемого потока данных, в который и был встроен алгоритм VAD. Реализованный в программном комплексе VAD описан в [204, 205].

Для облегчения понимания того, что требуется произнести, было принято решение о необходимости дублирования слога по аудио каналу

помимо отображения слога и его транскрипции на форме. Для этого были записаны правильные произношения слогов согласно ГОСТ Р 50840–95, а также по составленному ранее списку слогов с наиболее подверженными изменениями фонемами. Включение функции демонстрации правильного произношения осуществляется на форме «Запись слога» путем отметки в пункте «Воспроизводить слог».

Помимо оценки уже существующих слогов также есть необходимость в оценке слогов в рамках записи нового сеанса в режиме реального времени и реализации биологической обратной связи. Процесс записи нового сеанса с подсчетом оценки представлен на контекстной диаграмме на рисунке 5.11. Декомпозиция первого уровня представлена в Приложении И, рисунок И.2.

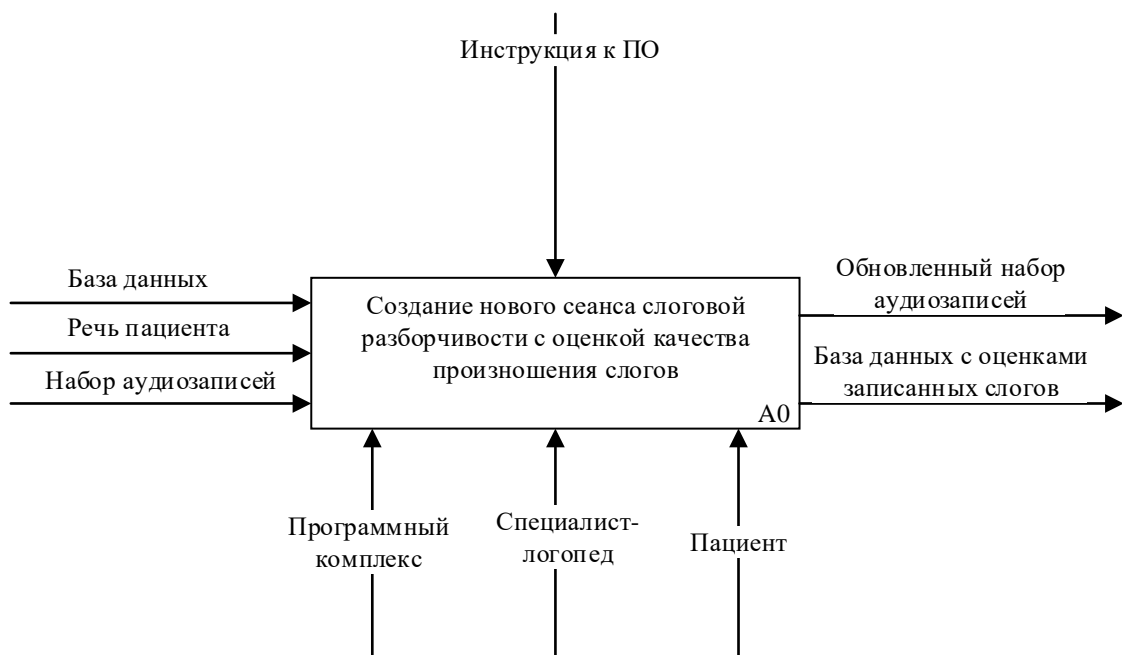


Рисунок 5.11 - Контекстная диаграмма процесса записи нового сеанса с подсчетом оценки

На выходе получают внесенные записи в базу данных, а также набор звуковых файлов, записанных в рамках сеанса. Так как обращение к модулю оценки происходит в рамках обработчика звукового потока (если голос уже был, а сейчас в течение некоторого времени тишина, то следует закрыть файл записи, оценить его и перейти к следующему слогу).

Форма программы «Запись слогов» представлена на рисунке 5.12.

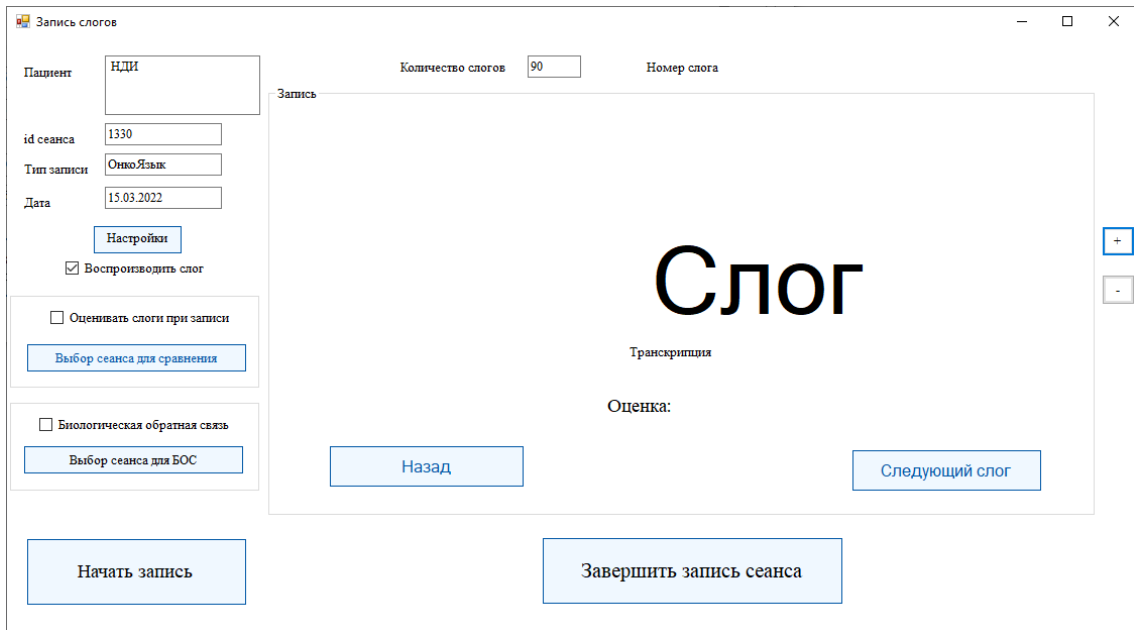


Рисунок 5.12 – Окно программы, в котором осуществляется запись сеансов.

Настройка процесса записи представляет собой последовательное выполнение следующих шагов:

1. Настройка записи голоса – индивидуальная настройка параметров звука для VAD для текущего оборудования и пациента.

1.1. По нажатию на кнопку «Настройки» переход в форму «Настройка записи голоса». Вид формы приведен на рисунке 5.13.

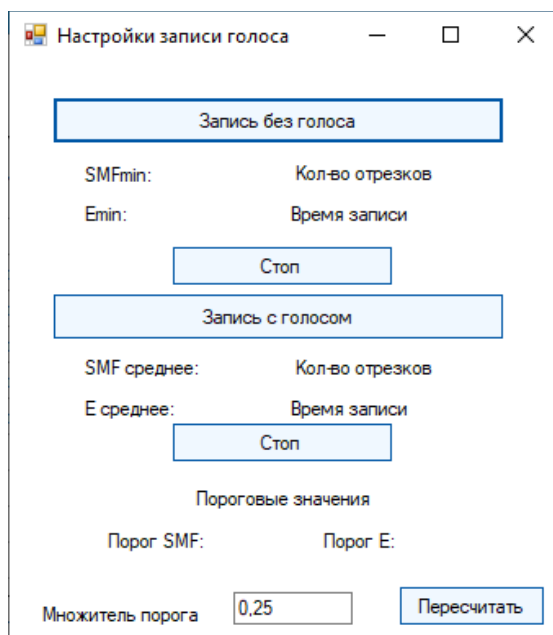


Рисунок 5.13 – Вид формы «Настройка записи голоса»

1.2. Запись тишины, то есть без голоса пациента – нажатие на кнопку «Запись без голоса», 3–5 секунд ожидания, нажатие на кнопку «Завершить» (рисунок 5.14).

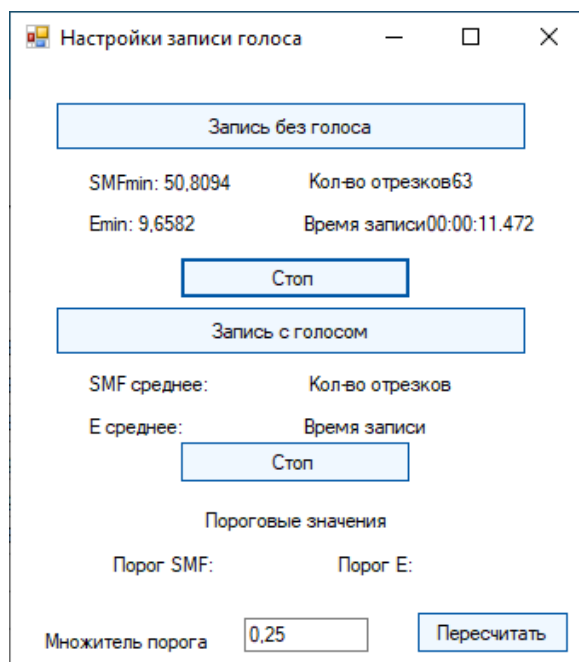


Рисунок 5.14 – Параметры аудиосигнала, содержащего фоновый шум

1.3. Запись голоса пациента – нажатие на кнопку «Запись с голосом», 3–5 секунд непрерывной речи пациента, нажатие на кнопку «Завершить».

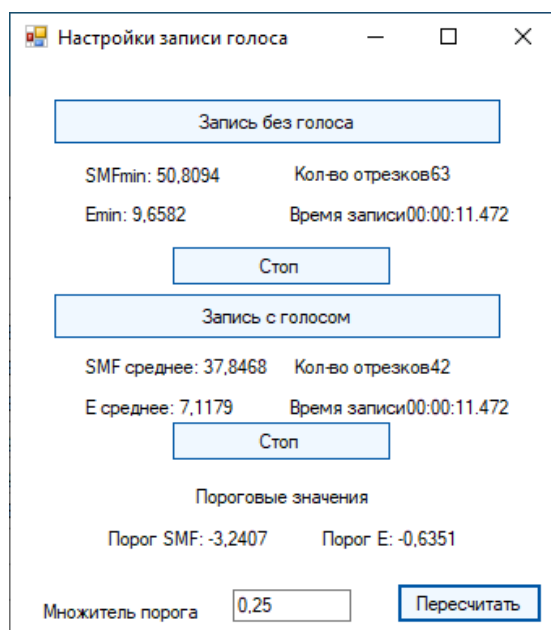


Рисунок 5.15 – Параметры аудиосигнала, содержащего непрерывную речь

1.4. Пересчет параметров VAD при необходимости с учетом коэффициента. На практике подобраны следующие коэффициенты: коэффициент 0,5 применим при использовании стандартного записывающего оборудования, коэффициент 0,25 применим при использовании профессионального и полупрофессионального записывающего оборудования.

2. Отметка о включение/выключении функции воспроизведения правильного произношения слога.

3. При необходимости провести оценку – настройка параметров оценки.

3.1. Отметка о необходимости проведения оценки – отметить пункт «Оценивать слоги при записи»

3.2. При нажатии на кнопку «Выбор сеансов для сравнения» переход в форму «Выбор сеансов для сравнения». Вид формы представлен на рисунке 5.16.

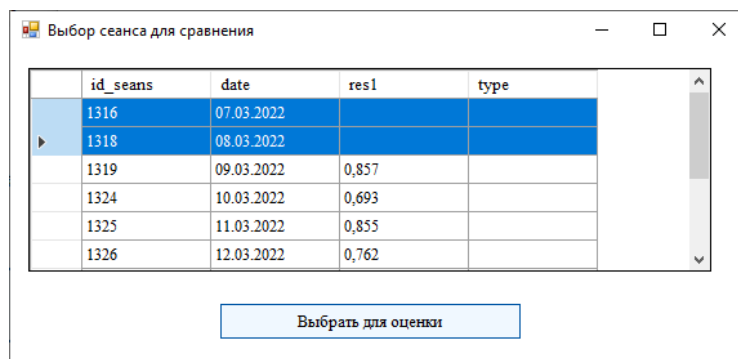


Рисунок 5.16 – Форма выбора сеансов в качестве эталонных.

3.3. Выбор 2 эталонных сеансов с нажатием клавиши Ctrl, нажатие на кнопку «Выбрать для оценки».

4. При необходимости реализации БОС – настройка БОС.

4.1. Отметка о необходимости реализации БОС – отметить пункт «Биологическая обратная связь»

4.2. При нажатии на кнопку «Выбор сеанса для БОС» переход в форму «Выбор сеанса для сравнения». Вид формы представлен на рисунке 5.17.

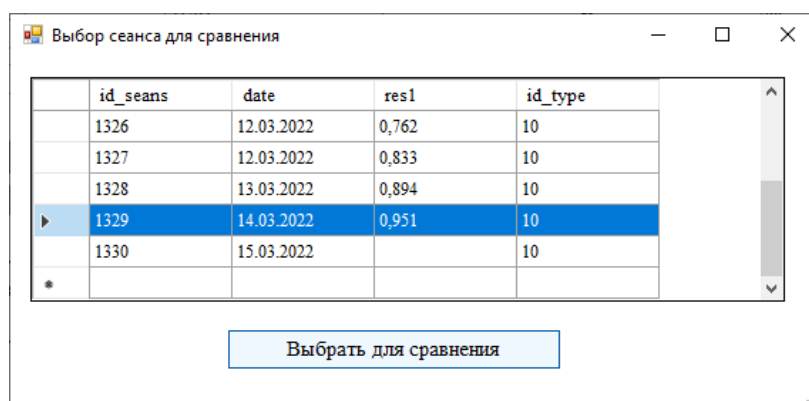


Рисунок 5.17 – Форма выбора сеанса для сравнения в рамках БОС.

4.3. Выбор 1 сеанса, с оценками которого будут сравниваться оценки записываемого сеанса, нажатие на кнопку «Выбрать для сравнения».

5. При необходимости – индивидуальная настройка размера шрифта отображения слога.

После настройки при нажатии на кнопку «Начать запись» начинается непосредственно процесс записи слогов. Для каждого слога повторяется набор следующих действий: демонстрация произношения по аудиоканалу – произнесение пациентом слога – оценка слога – демонстрация сформированного подкрепления БОС – переход к следующему слогу.

В процессе записи форма «Запись слога принимает следующий вид (рисунки 5.18 и 5.19).

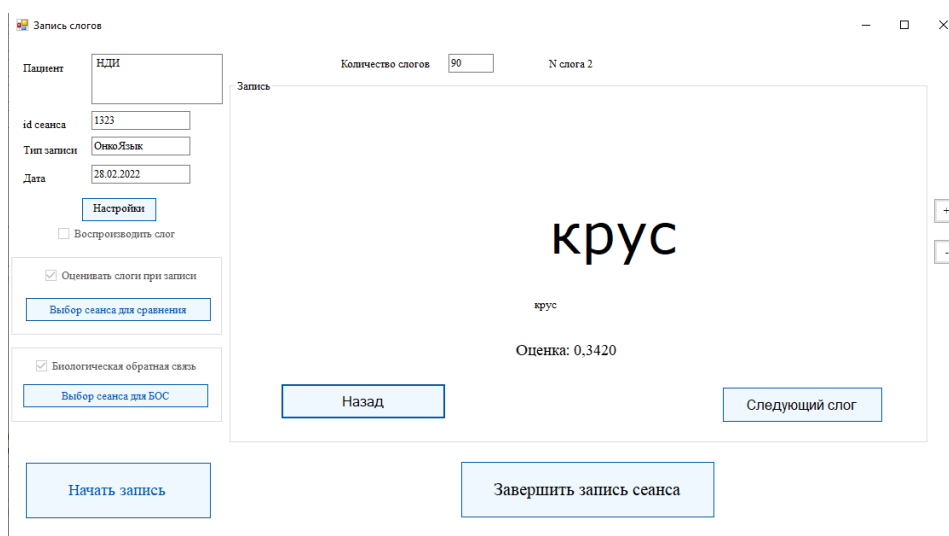


Рисунок 5.18 – Вид формы «Запись слогов» в процессе записи слогов с проведением оценки и реализацией БОС при отсутствии положительного кратковременного подкрепления.

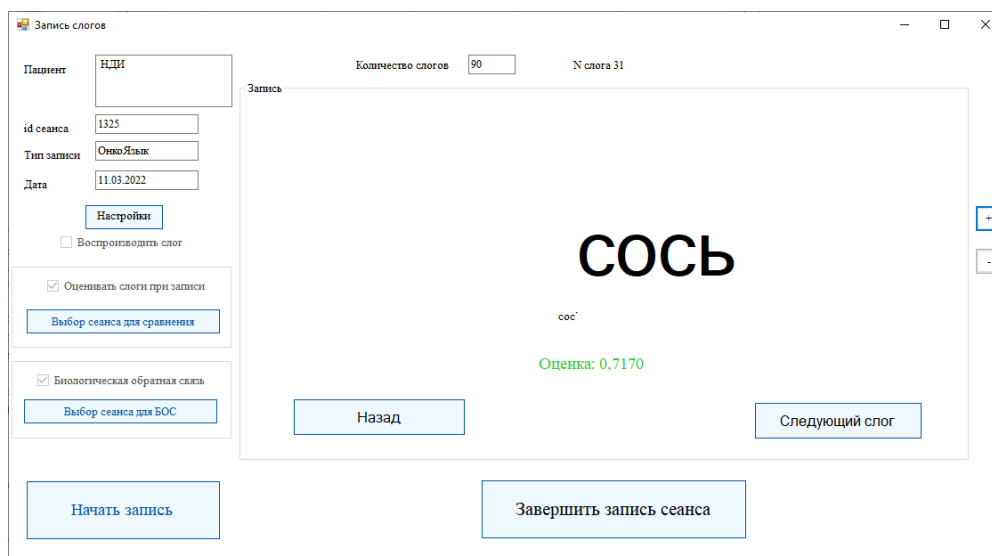


Рисунок 5.19 – Вид формы «Запись слогов» в процессе записи слогов с проведением оценки и реализацией БОС с отображением сформированного положительного кратковременного подкрепления.

При завершении записи всех сеансов выводится сообщение о результате и необходимости нажатия кнопки «Завершить запись сеанса» (Рисунок 5.20).

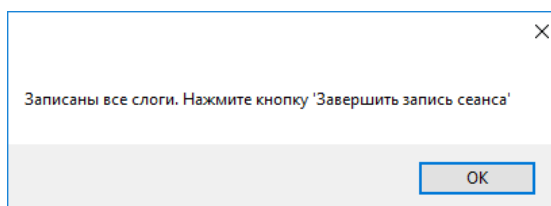


Рисунок 5.20 – Сообщение об окончании сеанса записи

5.2 Результаты тестирования программного комплекса на основе искусственно сформированного набора данных

Для тестирования реализованных алгоритмов и методик было проведено несколько сеансов записи, причем особенное внимание уделялось корректной записи слогов, в составе которых есть глухие шумные согласные, а также взрывные согласные, которые характеризуют наличием паузы перед «взрывом». Внимание таким согласным было уделено ввиду особенностей алгоритма VAD (а именно переключение на следующий слог при детектировании тишины при уже звучащем голосе). После записи все аудиофайлы были прослушаны и сделан вывод о

корректности работы алгоритма. Также корректность работы модуля записи была протестирована в процессе записи речи пациентов. Был протестирован модуль оценки, в том числе и подсчете оценки между двумя одинаковыми сеансами (эталонным для сеанса был задан этот же сеанс). Согласно методике оценки схожести, результаты попарных оценок всех слогов должны были совпадать и равняться единице (полное совпадение произношений). Результат совпал с ожидаемым, форма «Слог» после такой оценки представлена на рисунке 5.21.

id_zap	id_seans	id_slog	path	listen	correct1	correct2	segmentation
29720	1319	1	c:\Sound\Slog H...		1		
29721	1319	2	c:\Sound\Slog H...		1		
29722	1319	3	c:\Sound\Slog H...		1		
29723	1319	4	c:\Sound\Slog H...		1		
29724	1319	5	c:\Sound\Slog H...		1		
29725	1319	6	c:\Sound\Slog H...		1		
29726	1319	7	c:\Sound\Slog H...		1		
29727	1319	8	c:\Sound\Slog H...		1		
29728	1319	9	c:\Sound\Slog H...		1		
29729	1319	10	c:\Sound\Slog H...		1		
29730	1319	11	c:\Sound\Slog H...		1		
29731	1319	12	c:\Sound\Slog H...		1		
29732	1319	13	c:\Sound\Slog H...		1		
29733	1319	14	c:\Sound\Slog H...		1		
29734	1319	15	c:\Sound\Slog H...		1		
29735	1319	16	c:\Sound\Slog H...		1		
29736	1319	17	c:\Sound\Slog H...		1		
29737	1319	18	c:\Sound\Slog H...		1		
29738	1319	19	c:\Sound\Slog H...		1		
29739	1319	20	c:\Sound\Slog H...		1		

Рисунок 5.21 – Результат оценки сеанса с самим собой для проверки правильности работы алгоритма

Далее был протестирован процесс записи и оценки с имитацией процессе восстановления речи в процессе речевой реабилитации. Был создан набор сеансов, описание приведено в таблице 5.1.

Таблица 5.1 – Описание сеансов искусственно сформированного набора данных.

Номер сеанса в наборе	Описание сеанса	Дата сеанса	Соответствие сеансу речевой реабилитации
0-1	Сеанс с правильными произношением всех слогов	07.08.22	Сеанс до операции – эталонный сеанс
0-2	Сеанс с правильным произношением всех слогов	08.03.22	Сеанс до операции – эталонный сеанс

Продолжение таблицы 5.1.

Номер сеанса в наборе	Описание сеанса	Дата сеанса	Соответствие сеансу речевой реабилитации
1	Сеанс с неправильным произношением всех проблемных фонем в слоге	10.03.22	Сеанс после операции – оценка уровня речи после операции и перед речевой реабилитацией
2	Сеанс с неправильным произношением проблемных фонем в 4 из 5 слогах каждого расположения проблемной фонемы	11.03.22	Сеанс в процессе речевой реабилитации
3	Сеанс с неправильным произношением проблемных фонем в 3 из 5 слогах каждого расположения проблемной фонемы	12.03.22	Сеанс в процессе речевой реабилитации
4	Сеанс с неправильным произношением проблемных фонем в 2 из 5 слогах каждого расположения проблемной фонемы	13.03.22	Сеанс в процессе речевой реабилитации
5	Сеанс с неправильным произношением проблемных фонем в 1 из 5 слогах каждого расположения проблемной фонемы	14.03.22	Сеанс в процессе речевой реабилитации
6	Сеанс с правильным произношением всех слогов	15.03.22	Сеанс по окончанию речевой реабилитации

Согласно описанному алгоритму и методикам оценка сеансов 1–4 должна возрастать от сеанса к сеансу, так как уменьшается количество ошибок в произношении. Были получены следующие оценки сеансов (таблица 5.2). Графическое представление (осциллограммы и спектрограммы) аудиозаписей произношения слога «куп» из каждого из сеансов сформированного набора, а также оценки аудиозаписей представлены в Приложении Ж. Полученные оценки согласуются с

предполагаемыми результатами. Оценки сеансов близки к единице, так как при произношении допущены ошибки были только в проблемных фонемах, тогда как на практике ошибки возможны во всех входящих в слог фонемах.

Таблица 5.2 – Оценки сеансов сформированного набора данных.

Сеанс	Оценка сеанса
1	0,693
2	0,755
3	0,833
4	0,894
5	0,951
6	0,992

Также на этом наборе данных были сформированы долгосрочные подкрепления. Они представлены на рисунках 5.22 и 5.23 в том виде, в котором подкрепления отражаются в программном комплексе. Полученные значения возрастающих средних оценок по фонемам и группам фонем также согласуются с предполагаемыми результатами.

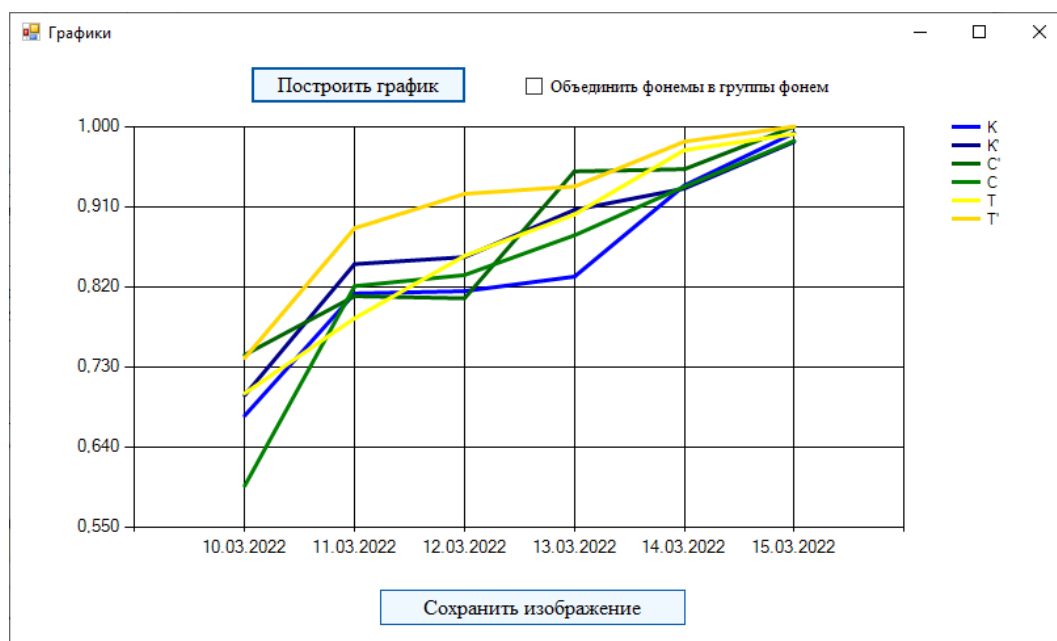


Рисунок 5.22 – Долгосрочное подкрепление – динамика оценок по проблемным фонемам по сеансам

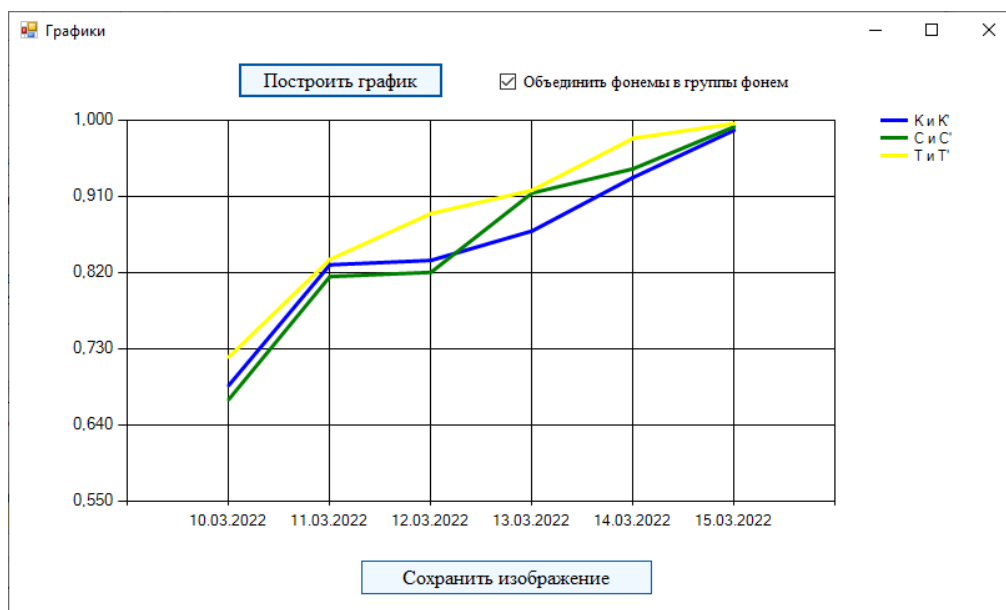


Рисунок 5.23 - Долгосрочное подкрепление – динамика оценок по группам проблемных фонем по сеансам

5.3 Внедрение программного комплекса в процесс реабилитации и результаты внедрения

Версии программного обеспечения для ЭВМ были зарегистрированы (приложение Б) [206 – 209] и внедрены в процесс речевой реабилитации (приложение В). Применение программного комплекса и промежуточные результаты его внедрение описано в [210, 211].

Внедрение осуществлялось постепенно. На первой стадии внедрения программный комплекс использовался для работы с базой данных и осуществления записей слогов. В дальнейшем были реализованы функции расчета оценок и отражения подкреплений в рамках биологической обратной связи. За время полноценного (с возможностью подсчета оценок) внедрения были записаны и обработана аудиозаписи 62 пациентов, 43 из которых проходили речевую реабилитацию с использованием программного комплекса, то есть было проведено два и более сеансов оценка качества произношения слогов. При этом записано 148 сеансов, 136 из которых согласно списку слогов с наиболее подверженным изменения после

хирургического вмешательства фонемами [к], [с], [т] и их мягкими вариантами- [к'], [с'], [т']. Всего записано 12810 аудиозаписей произношения слогов.

На рисунке 5.26 представлен отражение списка сеансов пациента с подсчитанными оценками на основе использования одного эталонного сеанса. Можно увидеть, что оценка в начале речевой реабилитации меньше, чем в процессе реабилитации, что говорит об улучшении качества произношения слогов (приближении речи к предоперационному уровню).

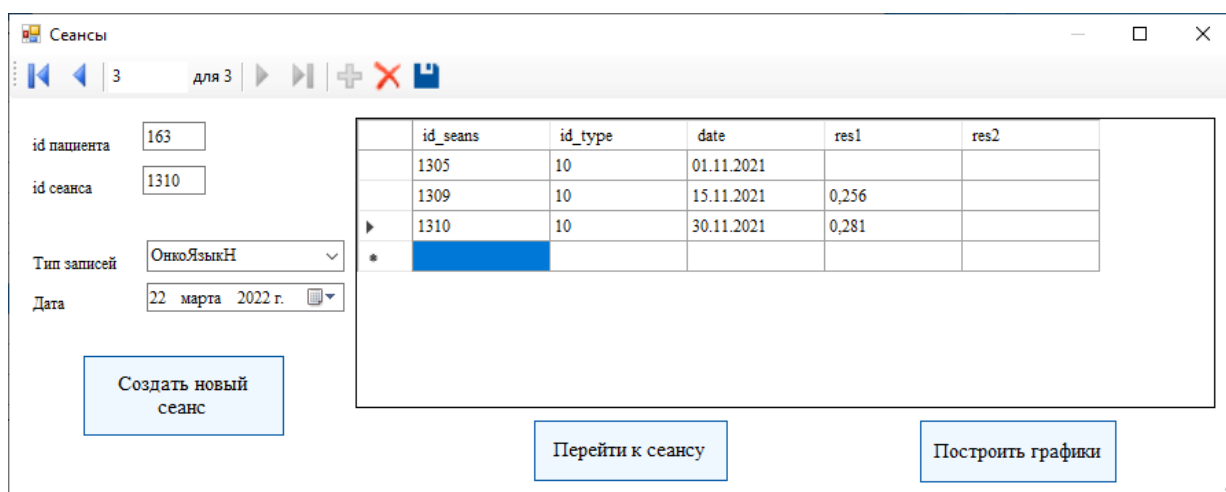


Рисунок 5.26 – Отражение списка пациента с подсчитанными оценками сеансов.

На рисунке 5.27 представлен вид сформированного долговременного подкрепления для пациента, который проходил полноценную речевую реабилитацию после рака языка. Пациент мужского пола, 34 года на момент первичной постановки диагноза рак языка, проведена трахеостомия, частичная глоссэктомия, логопедический диагноз механическая полимфная дислалия. Сеанс до операции был проведен 13.05.2020, первый сеанс после операции 04.06.2020. Расчет оценок производился на основе модификации методики оценки схожести звуковых сигналов с использованием одного эталонного (опорного) сеанса. Можно увидеть постепенный рост оценок по каждой из проблемных фонем. Однако для некоторых фонем наблюдается уменьшение оценок от сеанса к сеансу. Это может быть объяснено тем, что

речь не статичная единица и зависит не только от непосредственно состояния органов речепродуцирования, но и от психоэмоционального состояния и многих других факторов. Также наблюдается фактор ухудшения оценок из-за изменения скорости произношения, что является частью процесса восстановления речи у некоторых пациентов. Уменьшение оценки не является однозначным признаком ухудшения речи, а служит лишь сигналом, который специалист-логопед принимает во внимание во время построения схемы дальнейшей речевой реабилитации.

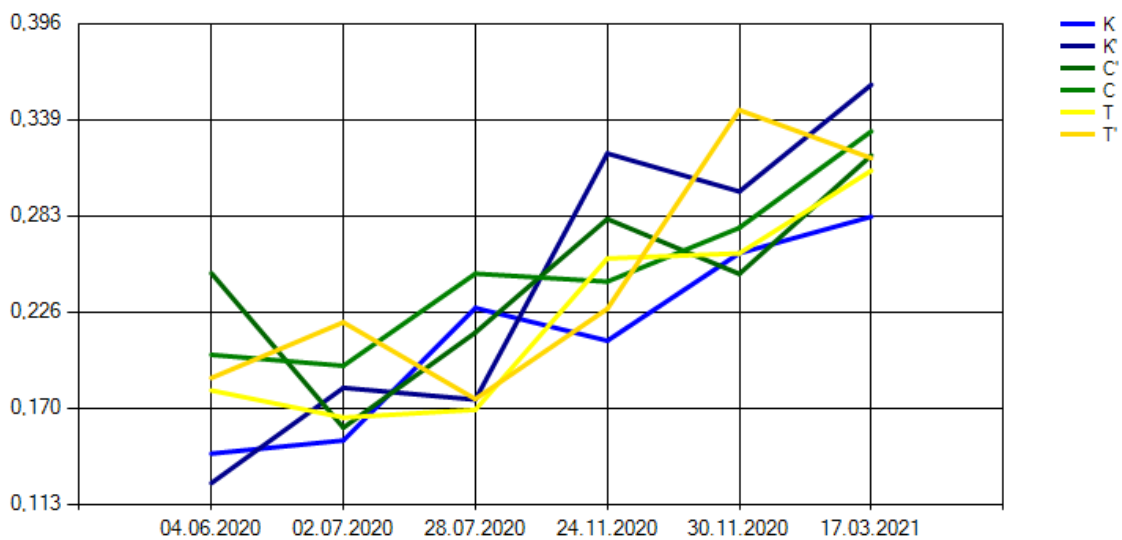


Рисунок 5.27 – Динамика усредненных по проблемным фонемам оценок для реального пациента НИИ Онкологии.

Использование биологической обратной связи позволило демонстрировать в понятном виде оценки качества произношения слогов и их изменение в процессе речевой реабилитации, что повышает мотивацию пациентов к продолжению самостоятельных тренировок речи.

5.4 Выводы по главе

Предложенные алгоритм и методики были реализованы в составе программного комплекса для оценки качества речи. Программный комплекс представляет собой программное обеспечение и базу данных. Структура базы данных была доработана для реализации всех требуемых функций. Программное обеспечение позволяет просматривать и редактировать информацию о пациентах, проводит сеансы оценки произношения слогов – производить запись слогов с индивидуально настроенными параметрами записи звука, рассчитывать оценки произношения в режиме реального времени на основе одного или двух эталонных сеансов, реализовывать метод биологической обратной связи для отражения текущего уровня качества речи пациента.

Разработанный программный комплекс был внедрен в процесс речевой реабилитации после комбинированного лечения онкологических заболеваний органов полости рта и ротоглотки, проводимого в НИИ Онкологии Томского НИМЦ. Внедрение позволило сократить время обработки аудиосигналов и расчета количественных оценок. Оценки произношения слогов позволяют пациенту отслеживать динамику восстановления речи до предоперационного уровня, а специалисту-логопеду при необходимости корректировать процесс речевой реабилитации. За время внедрения речевую реабилитацию с использованием программного комплекса прошло 43 пациента, было записано 168 сеансов и более 12 000 аудиозаписей.

Заключение

В результате работы было достигнуто повышение эффективности речевой реабилитации по восстановлению речи у пациентов после хирургического лечения онкологических заболеваний за счет внедрения алгоритма и методик анализа аудиосигналов и оценки качества произношения слогов, а также использования визуальных подкреплений на основе метода бионического принципа биологической обратной связи. Время оценки речи в процессе речевой реабилитации сократилось на 64,86%. Основные результаты работы представлены ниже.

1. Проведен анализ методов, методик и алгоритмов анализа и оценки речи, который показал, что существующие методы в рамках решения поставленной задачи не могут быть применены или имеют ряд существенных недостатков.

2. Разработан алгоритм количественной оценки схожести двух речевых сигналов с применением временной нормализации и гибридной меры оценки схожести. Сформулирована гибридная мера оценки схожести, основанная на расчете DTW-расстояния, коэффициента корреляции и расстояния Минковского и применении нечеткого классификатора к рассчитанным значениям. На основе оценок аудиозаписей пациентов НИИ онкологии обучены классификаторы для каждой из групп проблемных фонем. Точности классификации для тестовых и обучающих выборок составили: 0,87 и 0,88 для группы фонем [к] и [к'], 0,84 и 0,85 для группы фонем [с] и [с'], 0,79 и 0,81 для группы фонем [т] и [т'].

3. Разработана методика оценки схожести речевых сигналов с применением нескольких опорных сигналов, учитывающая вариативность произношения слогов и особенности речи дикторов (пациентов). Получаемые по методике оценки записей слогов и средние оценки сеансов представляют собой интерпретируемые оценки в заранее известном интервале. Описано прикладное применение методики оценки схожести в процессе речевой реабилитации.

4. Адаптирована методика голосовой реабилитации для проведения речевой реабилитации с применением методов бионического принципа

биологической обратной связи. Описан процесс оценки качества произношения слогов в рамках проведения сеансов речевой реабилитации. Описан процесс формирования кратковременных и долговременных подкрепляющих стимулов в рамках реализации биологической обратной связи. Время оценки качества произношения слогов сократилось на 64,86% в сравнении с ранее применяемой методикой, время записи сеанса с проведением оценки и реализацией биологической обратной связи сократилось на 55,7%.

5. Предложенные методики и алгоритмы реализованы в программном комплексе по оценке качества речи. Программный комплекс предназначен для хранения информации о пациентах и их сеансах оценки речи, проведения самих сеансов – запись произношения слогов и фраз, оценке качества записанных слогов, построения динамики восстановления речи пациентов. Оценка в программном комплексе производится в автоматизированном режиме и рассчитывается в режиме реального времени, также доступна возможность реализации метода бионического принципа биологической обратной связи.

6. Разработанный программный комплекс был протестирован и внедрен в процесс речевой реабилитации пациентов после комбинированного лечения онкологических заболеваний органов рта и ротоглотки. За время внедрения 43 пациента прошли реабилитацию с использованием программного комплекса, было записано 148 сеансов и 12810 аудиозаписей слогов.

Факт успешного внедрения результатов диссертационного исследования в процесс речевой реабилитации пациентов, проходящих лечение онкологических заболеваний на базе НИИ Онкологии Томского НИМЦ подтверждается актом внедрения. Кроме того, результаты диссертационного исследования используются в учебном процессе ФГБОУ ВО «Томский государственный университет систем управления и радиоэлектроники», что также подтверждается актом внедрения.

Список использованной литературы

1. Злокачественные новообразования в России в 2020 году (заболеваемость и смертность) // Под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой – М.: МНИОИ им. П.А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, – 2021. – илл. – 252 с.
2. Состояние онкологической помощи населению России в 2020 году // Под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой – М.: МНИОИ им. П.А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2021. – илл. – 239 с.
3. Методы анализа звучащей речи: новые измерения и результаты / Т. И. Шевченко, Е. А. Бурая, И. Е. Галочкина [и др.]. – Дубна: Феникс+, 2017. – 248 с.
4. Quackenbush, S.R. Objective Measures of Speech Quality / S.R. Quackenbush, T.P. Barnwell, M.A. Clements. – Prentice Hall, 1988. – 408 p.
5. ITU-T Recommendation Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. – 2000. – 862 p.
6. Perceptual evaluation of speech quality (PESQ) - A new method for speech quality assessment of telephone networks and codecs / A. Rix, J. Beerends, M. Hollier, A. Hekstra // Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2001. – vol. 2. – 749–752 pp.
7. Voran, S. Objective estimation of perceived speech quality - Part I: Development of the measuring normalizing block technique // IEEE Transactions on Speech and Audio Processing, 1999. – № 7(4). – 371–382 pp.
8. Dimolitsas, S. Objective speech distortion measures and their relevance to speech quality assessments // IEE Proceedings - Vision, Image and Signal Processing, 1989. – № 136(5). – 317–324 pp.
9. Kubichek, R. Advances in objective voice quality assessment / R. Kubichek, D. Atkinson, A. Webster // Proceedings Global Telecommunications Conference, 1991. – № 3. – 1765–1770 pp.
10. Kitawaki, N. Quality assessment of coded speech // Advances in Speech Signal Processing, 1991. – 357–385 pp.
11. Barnwell, T. Objective measures for speech quality testing // The Journal of the Acoustical Society of America, 1979. – № 66(6). – 1658–1663 pp.
12. Application of Objective Speech Quality Metrics / S. K. Isabelle, S.J. Guzman, R. Saliba, R. Novorita // First ISCA ITRW on Auditory Quality of Systems, 2003. – 39-44 pp.
13. Rix, A.W. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs // 2001 Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. – № 2. – 749–752 pp.
14. Perceptual evaluation of speech quality (PESQ): The new ITU standard for end-to-end speech quality assessment. part I - time-delay

compensation / A.W. Rix, M.P. Hollier, A.P. Hekstra, J.G. Beerends // Journal of the Audio Engineering Society, 2002. – № 50(10). – 755-764 pp.

15. Perceptual evaluation of speech quality (PESQ): The new ITU standard for end-to-end speech quality assessment. part II - psychoacoustic model / J.G. Beerends, A.P. Hekstra, A.W. Rix, M.P. Hollier // Journal of the Audio Engineering Society, 2002. – № 50(10). – № 765-778 pp.

16. Sun, L. Voice quality prediction models and their application in VoIP networks / L. Sun, E.C. Ifeachor // IEEE Trans Multimedia, 2006. – № 8(4). – № 809-820 pp.

17. VoIP call quality assesment based on RPROP neural networks / M. Voznak, J. Rozhon, F. Rezac, E. Gresak // Proceedings 11th International Symposium on Telecommunications, ВИНТЕЛ 2016. –1-5 pp.

18. Сандовал-Ибарра, Ю. Улучшение качества речи с использованием адаптивных спектральных оценок [Электронный ресурс]/ Ю. Сандовал-Ибарра, В. Диаз-Рамирез, В.И. Кобер, В.Н. Карнаухов // Информационные процессы. – 2015 – № 3. – Режим доступа: <http://www.jip.ru/2015/314-323-2015.pdf> (дата обращения 1.04.2022).

19. Терехов, А. Н. Недостаток интрузивного метода оценки качества передачи речи и способ его устранения / А. Н. Терехов // Фундаментальные проблемы радиоэлектронного приборостроения. – 2016. – Т. 16. – № 5. – С. 98-102.

20. G.107: The E-model: a computational model for use in transmission planning, 2015.

21. Полторак, В. П. Оценка качества передачи речи в IP-телефонии [Электронный ресурс]/ В. П. Полторак, О. М. Моргаль, Ю. А. Заика // Молодой ученый. – 2014. – №4. – С. 121-123. Режим доступа: <https://moluch.ru/archive/63/9849/> (дата обращения 1.04.2022).

22. Способы оценки субъективного качества речи [Электронный ресурс]. – Режим доступа: <https://habrahabr.ru/post/177099/> (дата обращения 1.04.2022)

23. Yang, W. Enhanced modified bark spectral distortion (embsd): an objective speech quality measure based on audible distortion and cognition model / phd. — USA: Temple University, 1999. – 163 p.

24. Klatt, D. Prediction of perceived phonetic distance from critical band spectra // Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1982. — № 7. — 1278–1281 pp.

25. Karjalainen, M. A new auditory model for the evaluation of sound quality of audio system // Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1985. — № 10. — 608–611 pp.

26. Wang, S. An objective measure for predicting subjective quality of speech coders / S. Wang, A. Skey, A. Gersho. // IEEE Journal on Selected Areas in Communications, 1992. — 10(5). — 819–829 pp.

27. Yang, W. Performance of the modified Bark spectral distortion as an objective speech quality measure / W. Yang, M. Benbouchta, R. Yantorno //

Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1998. — 541–544 pp.

28. Karjalainen, M. Sound quality measurements of audio systems based on models of auditory perception // Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1984. — № 9. — 132–135 pp.

29. Chen, G. Loudness pattern-based speech quality evaluation using Bayesian modelling and Markov chain Monte Carlo methods / G. Chen, V. Parsa // Journal of the Acoustical Society of America, 2007. — № 121(2). — 77–83 pp.

30. On the use of Bayesian modeling for predicting noise reduction performance / N. Pourmand, D. Suelzle, V. Parsa, Y. Hu et al // Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2009. — 3873–3876 pp.

31. Moore, B. An introduction to the psychology of hearing // London: Academic Press, 2003. — 5th edn. — 413 p.

32. Novorita, B. Incorporation of temporal masking effects into bark spectral distortion measure // Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. — № 2. — 665–668 pp.

33. Acero, A. Speech recognition and understanding // NAACL-Tutorials '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Tutorials, 2003. — № 5. — 3 p.

34. Экспериментальная оценка качества распознавания разговорной русской речи / А. А. Кибалко, А. Г. Данилов, А. С. Мышкин [и др.] // Речевые технологии. — 2012. — № 4. — С. 23–33.

35. Xiao, X. Information retrieval methods for automatic speech recognition / X. Xiao, J. Droppo, A. Acero // 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010. — 5550–5553 pp.

36. Ронжин, А. Л. Особенности дистанционной записи и обработки речи в автоматах самообслуживания / А. Л. Ронжин, А. А. Карпов, И. А. Кагиров // Информационно-управляющие системы. — 2009. — № 5(42). — С. 32–38.

37. Аналитическая оценка качества речи на выходе систем низкоскоростного кодирования при воздействии акустических помех / С. Н. Кириллов, Ю. Н. Ромашкин, Я. О. Картавенко, Т. В. Дмитриев // Речевые технологии. — 2012. — № 4. — С. 16–22.

38. Карпов, А. А. SIRIUS - система дикторнезависимого распознавания слитной русской речи / А. А. Карпов, А. А. Ронжин, И. В. Ли // Известия ТРТУ. — 2005. — № 10(54). — С. 44–54.

39. Карпов, А. А. Аудиовизуальный речевой интерфейс для систем управления и оповещения / А. А. Карпов // Известия ЮФУ. Технические науки. — 2010. — № 3(104). — С. 218–222.

40. Наседкина, З. А. Новый метод оценки качества синтезированной речи / З. А. Наседкина // Национальная Ассоциация Ученых. – 2016. – № 3-2(19). – С. 37-38.
41. Лань, Г. Синтез фрагментов голоса человека на основе модели реконструкции частотных спектров / Г. Лань, А. С. Фадеев, А. Н. Моргунов // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2021. – Т. 24. – № 2. – С. 14-20.
42. The evaluation process automation of phrase and word intelligibility using speech recognition systems / E. Kostuchenko, D. Novokhrestova, M. Tirskaaya [et al.] // Lecture Notes in Computer Science. – 2019. – Vol. 11658 LNAI. – P. 237-246.
43. Baum, L.E. Statistical inference for probabilistic functions of finite state Markov chains / L.E. Baum, T. Petrie // Annals of Mathematical Statistics, 1966. – Vol. 37:1966. – 1554-1563 pp.
44. Baker, J.K. Stochastic modeling for automatic speech understanding" in Speech Recognition // New York: Academic Press, 1975. – 521 -542 pp.
45. Baker, J.K. The DRAGON system An overview. // IEEE Transactions on Acoustics, Speech, and Signal Processing, 1975. – vol. ASSP-23. – 24 - 29 pp.
46. Lowerre, B.T. The HARPY speech understanding system / Unpublished Ph.D. dissertation // Dept. of Computer Science, Carnegie-Mellon Univ., Pittsburg, PA. –1976.
47. Liporace, L.R. Maximum likelihood estimation for multivariate observations of Markov sources. // IEEE Transactions on Information Theory, 1982. – Vol. IT-28. – 729-734 pp.
48. Juang, B.H. Maximum likelihood estimation for multivariate mixture observations of Markov chains / B.H. Juang, S.E. Levinson, S.M. Sondhi // IEEE Transactions on Information Theory, 1986. – Vol. IT-32. – 307-309 pp.
49. Левинсон, С.Е. Структурные методы автоматического распознавания речи // ТИИЭР, 1985. – т. 73, № 11. –100-128 с.
50. Assessment of syllable intelligibility based on convolutional neural networks for speech rehabilitation after speech organs surgical interventions / E. Kostuchenko, D. Novokhrestova, S. Pekarskikh [et al.] // Lecture Notes in Computer Science. – 2019. – Vol. 11658 LNAI. –359-369 pp.
51. Kostyuchenko, E. Evaluation of syllable intelligibility through recognition in speech rehabilitation of cancer patients / E. Kostyuchenko, D. Novokhrestova, L. Balatskaya // Integrating Research Agendas and Devising Joint Challenges: International Multidisciplinary Symposium ICT Research in Russian Federation and Europe, Stavropol - Dombay, 15–20 Oct 2018. – Stavropol - Dombay: North-Caucasian Federal University, 2018. – 75-80 pp.
52. Айфичер, Э. Цифровая обработка сигналов. Практический подход / Айфичер, Б. Джервис. - М.: Вильяме, 2004. - 992 с.

53. Аграновский, А.В. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов / А.В. Аграновский, Д.А. Леднов. — М.: Радио и связь, 2004. - 150 с.
54. Общая и прикладная фонетика / Л. В. Златоустова, Р. К. Потапова, В. В. Потапов, В. Н. Трунин-Донской. — Москва : Издательство Московского государственного университета, 1997. — 416 с.
55. Чистович, Л.А. Физиология речи. Восприятие речи человеком / Л.А. Чистович, А.В. Венцов, М.П. Грамстрем и др. - Л: Наука, 1976. - 388 с.
56. Черкашин, Е. И. Применение спектра модуляции речевого сигнала для распознавания голосовых команд / Е. И. Черкашин, Т. В. Шарий // Донецкие чтения 2018: образование, наука, инновации, культура и вызовы современности : Материалы III Международной научной конференции, Донецк, 25 октября 2018 года / Под общей редакцией С.В. Беспаловой. — Донецк: Донецкий национальный университет, 2018. — С. 250-251.
57. Шарий, Т. В. Метод автоматического выявления звуковых дефектов в аудиосигналах / Т. В. Шарий // Донецкие чтения 2019: образование, наука, инновации, культура и вызовы современности : материалы IV Международной научной конференции, Донецк, 29–31 октября 2019 года. — Донецк: Донецкий национальный университет, 2019. — С. 269-271.
58. Шарий, Т. В. Анализ информативности участков звуковых сигналов для распознавания речи в условиях шума / Т. В. Шарий // Вестник Донецкого национального университета. Серия Г: Технические науки. — 2020. — № 1. — С. 37-45.
59. Бондаренко, В. П. Сегментация и параметрическое описание речевого сигнала / В. П. Бондаренко, А. А. Конев, Р. В. Мещеряков // Известия высших учебных заведений. Приборостроение. — 2007. — Т. 50. — № 10. — С. 3-7.
60. Конев, А. А. Сегментация речевых сигналов на вокализованные и невокализованные участки на основе одновременной маскировки / А. А. Конев, Р. В. Мещеряков, Е. Ю. Костюченко // Автметрия. — 2018. — Т. 54. — № 4. — С. 51-57.
61. Сорокин В.Н. Сегментация и распознавание гласных / В.Н. Сорокин. // Информационные процессы. - 2004. - Т.4. - №2. - С. 202-220.
62. Rabiner, L. R. Digital processing of speech signals / L. R. Rabiner, R. W. Schafer // Prentice-Hall, Englewood Cliffs, NJ. —1978. — 512 p.
63. Винцюк, Т. К. Анализ, распознавание и интерпретация речевых сигналов // Киев: Наук. думка, 1987. — 262 с.
64. Формантное представление речевого сигнала для решения задач обработки речи / А. В. Аграновский, М. Ю. Зулкарнеев, Д. А. Леднов, П. М. Сулима // Известия ТРТУ. — 2002. — № 2(25). — С. 92-95.
65. Гураков, И. А. Автоматизация выделения формант и поиска выравненных фрагментов фонограмм при подготовке к проведению фоноскопической экспертизы / И. А. Гураков, Е. Ю. Костюченко, Д. И.

Новохрестова // Автоматика и программная инженерия. – 2018. – № 2(24). – С. 45–50.

66. Алгоритм выделения формант и поиска выровненных фрагментов при подготовке к проведению фоноскопической экспертизы / И. А. Гураков, Е. Ю. Костюченко, Д. И. Новохрестова, М. П. Силич // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2018. – Т. 21. – № 2. – С. 48-53.

67. Исследование спектра формант форсированной речи / А. В. Иванов, В. А. Трушин, Г. В. Маркелова, И. Л. Рева // Научный вестник Новосибирского государственного технического университета. – 2015. – № 4(61). – С. 63-73.

68. Евдокимова, В. В. Вариативность формантной картины гласных в разных видах речи / В. В. Евдокимова // Первый междисциплинарный семинар "Анализ разговорной русской речи" (АРЗ - 2007), Санкт-Петербург, 29 августа 2007 года / Составители А.Л. Ронжин, И.А. Кагиров. – Санкт-Петербург: Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 2007. – С. 49-54.

69. Статистические распределения формант различных дикторов при проведении фоноскопических экспертиз методом формантного выравнивания / И. А. Гураков, Е. Ю. Костюченко, Д. И. Новохрестова, А. А. Шелупанов // Информационные технологии в управлении (ИТУ-2018) : материалы конференции, Санкт-Петербург, 02–04 октября 2018 года. – Санкт-Петербург: "Концерн "Центральный научно-исследовательский институт "Электроприбор", 2018. – С. 501-509.

70. Колоколов, А. С. Способ определения частоты основного тона речевого сигнала на основе автокорреляции / А. С. Колоколов, И. А. Любинский // DSPA: Вопросы применения цифровой обработки сигналов. – 2017. – Т. 7. – № 2. – С. 31-34. Савченко, В. В. Метод Измерения частоты основного тона с межпериодным накоплением речевого сигнала / В. В. Савченко // Цифровая обработка сигналов. – 2017. – № 2. – С. 44-48.

71. Вольф, Д. А. Модель процесса сингулярного оценивания частоты основного тона речевого сигнала / Д. А. Вольф, Р. В. Мещеряков // Акустический журнал. – 2016. – Т. 62. – № 2. – С. 216.

72. Азаров, И. С. Алгоритм оценки мгновенной частоты основного тона речевого сигнала / И. С. Азаров, М. И. Вашкевич, А. А. Петровский // Цифровая обработка сигналов. – 2012. – № 4. – С. 49-57.

73. Вишнякова, О. А. Гибридный алгоритм выделения частоты основного тона / О. А. Вишнякова, Д. Н. Лавров // Математические структуры и моделирование. – 2016. – № 1(37). – С. 59-65.

74. Алимуратов, А. К. Исследование частотно-избирательных свойств методов декомпозиции на эмпирические моды для оценки частоты основного тона речевых сигналов / А. К. Алимуратов // ТРУДЫ МФТИ.

Труды Московского физико-технического института (национального исследовательского университета). – 2015. – Т. 7. – № 3(27). – С. 56-68.

75. Сафонов, А. П. Сравнительный анализ алгоритмов спектрального вычитания шума из речевых сигналов на основе объективных критериев оценки / А. П. Сафонов // Актуальные проблемы развития современной науки и образования : Сборник научных трудов по материалам Международной научно-практической конференции в 5 частях, Москва, 30 апреля 2015 года / ООО "АР-Консалт". – Москва: Общество с ограниченной ответственностью "АР-Консалт", 2015. – С. 117-118.

76. Яшин, А. В. Методический аппарат оценки статистических характеристик речевых сигналов / А. В. Яшин, Д. П. Бачурин, И. В. Красавин // Моделирование авиационных систем: материалы Всероссийской научно-практической конференции, Москва, 21–22 ноября 2013 года. – Москва: Государственный научно-исследовательский институт авиационных систем, 2013. – С. 122-123.

77. Азаров, И.С. Вычисление мгновенных гармонических параметров речевого сигнала / И.С. Азаров, А.А. Петровский // Речевые технологии. - 2008. -№ 1.-С. 67-77.

78. Азаров, И.С. Алгоритм оценки мгновенной частоты основного тона речевого сигнала / И.С. Азаров, М.И. Вашкевич, А.А. Петровский // Цифровая обработка сигналов. - 2012. -№ 4. - С. 49-57.

79. Novokhrestova, D. Choice of Signal Short-Term Energy Parameter for Assessing Speech Intelligibility in the Process of Speech Rehabilitation / D. Novokhrestova, E. Kostyuchenko, R. Meshcheryakov // Lecture Notes in Computer Science. – 2018. – Vol. 11096 LNAI. – P. 461-469

80. Железняк, В. К. Цифровая обработка сигнала с линейно-частотной модуляцией частотно-временным преобразованием Вигнера для оценки разборчивости речи / В. К. Железняк, К. Я. Раханов // Вестник Полоцкого государственного университета. Серия С. Фундаментальные науки. – 2019. – № 4. – С. 16-25.

81. Оценка информативности параметров речевого сигнала при обработке с использованием нейронных сетей / Е. Ю. Костюченко, В. П. Коцубинский, А. А. Авдеев [и др.] // Речевые технологии. – 2010. – № 3. – С. 39-44.

82. Рахманенко, И. А. Анализ идентификационных признаков в речевых данных с помощью GMM-UBM системы верификации диктора / И. А. Рахманенко, Р. В. Мещеряков // Труды СПИИРАН. – 2017. – № 3(52). – С. 32-50.

83. Рахманенко, И. А. Автоматическая верификация диктора по произвольной фразе с применением свёрточных глубоких сетей доверия / И. А. Рахманенко, А. А. Шелупанов, Е. Ю. Костюченко // Компьютерная оптика. – 2020. – Т. 44. – № 4. – С. 596-605.

84. Михайлов, В. Г. Измерение параметров речи / В. Г. Михайлов, Л. В. Златоустова. – Москва : Научно-техническое издательство "Радио и связь", 1987. – 168 с.
85. Streijl, R.C. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives / R.C. Streijl, S. Winkler, D.S. Hands // *Multimedia Systems*, 2016. – Vol. 22, № 2. – P. 213–227
86. Rix, A.W. Perceptual speech quality assessment - A review // *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings*, 20042004. – p. III1056-9.
87. Парфенов, В. И. Повышение эффективности шумоподавления при винеровской фильтрации речевых сигналов / В. И. Парфенов, М. М. Жуков, Е. А. Кривцов // *Вестник Воронежского института МВД России*. – 2019. – № 4. – С. 137-145.
88. Трифонов, М. А. Имитационное моделирование передачи речи по технологии VOIP / М. А. Трифонов // *Вектор науки Тольяттинского государственного университета*. – 2010. – № 3(13). – С. 31-33.
89. Лебединский, А. К. Оценка качества передачи речи в сети ОБТС / А. К. Лебединский // *Автоматика, связь, информатика*. – 2012. – № 2. – С. 5-10.
90. Вовк, И. Ю. Система синтеза естественной речи для русского языка на базе глубоких нейронных сетей / И. Ю. Вовк, В. С. Гогорян // *Новые информационные технологии в автоматизированных системах*. – 2019. – № 22. – С. 142-150.
91. Пентелейчук, А. В. Влияние искажений речевого сигнала в тракте приемо-передачи на оценку качества услуг подвижной связи / А. В. Пентелейчук, Н. Ю. Фудина // *Фундаментальные проблемы радиоэлектронного приборостроения*. – 2009. – Т. 9. – № 4. – С. 234-238.
92. Кириллов, С. Н. Обработка и кодирование речевых сигналов с помощью искусственных нейронных сетей / С. Н. Кириллов, Е. С. Сазонова // *Цифровая обработка сигналов*. – 2017. – № 1. – С. 27-32.
93. Бахтиярова, Е. А. О методах оценки качества передачи речи / Е. А. Бахтиярова, К. Е. Досболаева // *Вестник Казахской академии транспорта и коммуникаций им. М. Тынышпаева*. – 2013. – № 2(81). – С. 30-35.
94. Cote, N. *Integral and diagnostic intrusive prediction of speech* // Springer-Verlag Berlin Heidelberg, 2011. – 267 p.
95. Объективное и субъективное оценивание качества речевых сигналов с ограниченной полосой частот / К. С. Замша, Б. В. Лозинский, Ю. А. Митяй [и др.] // *Электроника и связь*. – 2016. – Т. 21. – № 1(90). – С. 18-26.
96. Loizou, P.C. *Speech Quality Assessment* // *Multimedia Analysis, Processing and Communications* / Berlin, Heidelberg: Springer, 2011. – P. 623–654.
97. P.835: Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm, 2003.

98. Naderi, B. Speech Quality Assessment in Crowdsourcing: Comparison Category Rating Method [Электронный ресурс]/ B. Naderi, S. Möller, R. Cutler// Режим доступа: <https://arxiv.org/abs/2104.04371> (дата обращения 01.04.2022).
99. Friedemann Koster Introducing a new Test-Method for Diagnostic Speech Quality Assessment in a Conversational Situation [Электронный ресурс] / Режим доступа: https://www.researchgate.net/publication/303339436_Introducing_a_new_Test-Method_for_Diagnostic_Speech_Quality_Assessment_in_a_Conversational_Situation (дата обращения 01.04.2022).
100. P.800: Methods for subjective determination of transmission quality, 1996.
101. RECOMMENDATION ITU-R BS.1534-1 - Method for the subjective assessment of intermediate quality level of coding systems, 2015. – 36 p.
102. Бахтиярова, Е. А. Критерии и методы оценки качества передачи речи / Е. А. Бахтиярова, М. Н. Балгабаева // Инновационные технологии на транспорте: образование, наука, практика: Материалы ХLI Международной научно-практической конференции, Алматы, Казахстан, 03–04 апреля 2017 года / Под редакцией Б.М. Ибраева. – Алматы, Казахстан: Казахская академия транспорта и коммуникаций им. М. Тынышпаева, 2017. – С. 38–43.
103. Потапова, Р. К. Методика оценки индивидуальных фонетических параметров сигнала / Р. К. Потапова, В. В. Потапов, М. В. Хитина // Вестник Московского государственного лингвистического университета. – 2014. – № 1(687). – С. 197–205.
104. Потапова, Р. К. Некоторые прикладные аспекты исследования звучащей речи / Р. К. Потапова, В. В. Потапов // Вестник Московского государственного лингвистического университета. – 2011. – № 607. – С. 164–186.
105. Потапова, Р. К. О возможности перцептивно-слухового распознавания состояния "агрессия" по устной речи / Р. К. Потапова, Л. Р. Комалова // Вестник Московского государственного лингвистического университета. – 2014. – № 13(699). – С. 202–214.
106. Алимуратов, А. К. Оценка психоэмоционального состояния человека на основе декомпозиции на эмпирические моды и кепстрального анализа речевых сигналов / А. К. Алимуратов, А. Ю. Тычков, П. П. Чураков // Вестник Пензенского государственного университета. – 2018. – № 2(22). – С. 89–95.
107. Двойникова, А. А. Автоматическое определение эмоционального состояния участников предметных разговоров по транскрипциям речи / А. А. Двойникова, Д. Ю. Мамонтов, А. А. Карпов // Альманах научных работ молодых учёных университета ИТМО, Санкт-Петербург, 01–04 февраля 2021 года. – Санкт-Петербург: федеральное

государственное автономное образовательное учреждение высшего образования "Национальный исследовательский университет ИТМО", 2021. – С. 63–68.

108. Величко, А. Н. Аналитический обзор систем автоматического определения депрессии по речи / А. Н. Величко, А. А. Карпов // Информатика и автоматизация. – 2021. – Т. 20. – № 3. – С. 497–529.

109. Голосовой портрет ребенка с типичным и атипичным развитием / Е. Е. Ляксо, О. В. Фролова, С. В. Гречаный [и др.]. – Санкт-Петербург: Издательско-полиграфическая ассоциация высших учебных заведений, 2020. – 204 с.

110. К вопросу об оценке состояния ребенка по характеристикам его голоса и речи / Е. Е. Ляксо, О. В. Фролова, А. С. Григорьев [и др.] // Центральные механизмы речи : Сборник материалов IX Всероссийской (с международным участием) научной конференции, посвященной памяти проф. Н.Н. Трауготт, Санкт-Петербург, 11–13 ноября 2019 года / Под общ. ред. А. Н. Шеповальникова. – Санкт-Петербург: ООО "Издательство ВВМ", 2019. – С. 60.

111. Тиунов, С. Д. Моделирование голосообразования при патологии / С. Д. Тиунов // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. – 2014. – № 2. – С. 19–23.

112. Худоев, И. В. Медицинские приложения дополненной реальности / И. В. Худоев, Р. Я. Пирмагомедов, М. А. Маколкина // Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь (DCCN-2018) : материалы XXI Международной научной конференции, Москва, 17–21 сентября 2018 года / Российский университет дружбы народов; Под общей редакцией В. М. Вишневого и К.Е. Самуйлова. – Москва: Российский университет дружбы народов (РУДН), 2018. – С. 149–161.

113. Выделение признаков неврологических заболеваний из речевого сигнала / В. И. Вашкевич, Ю. Н. Рушкевич, И. С. Азаров, А. А. Петровский // DSPA: Вопросы применения цифровой обработки сигналов. – 2018. – Т. 8. – № 1. – С. 193–197.

114. Фаттахова, М. Я. Разработка методики реабилитации пациентов с нарушениями голосо-речевой функции / М. Я. Фаттахова, Р. Ш. Хабипов // Казанский физико-технический институт имени Е.К Завойского. Ежегодник. – 2021. – Т. 2020. – С. 92–95.

115. Патент № 2738660 С1 Российская Федерация, МПК А61В 5/00. Способ диагностики и реабилитации пациентов с нарушениями голосо-речевой функции: № 2020113670: заявл. 16.04.2020: опубл. 15.12.2020 / М. Я. Фаттахова, Я. В. Фаттахов, Р. Ш. Хабипов, В. Н. Красножен.

116. Куль, Т. П. Адаптация методов цифровой обработки сигналов к задаче анализа речи при неврологических патологиях / Т. П. Куль, Ю. Н. Рушкевич, С. А. Лихачев // Доклады Белорусского государственного

университета информатики и радиоэлектроники. – 2018. – № 7(117). – С. 128–132.

117. Rossiter, D. A realtime LPC based vocal tract area display for voice development / D. Rossiter, D.M. Howard, M. Downes // *Journal of Voice*, 1995 - №8(4). – 314-319 pp.

118. Kewley, P. D. The Indiana Speech Training Aid ISTR: A microcomputer-based aid using speaker-dependent speech recognition / P. D. Kewley, C.S. Watson, P.A. Cromer // *Synergy '87, The 1987 ASHF Computer Conference, Proceedings*, 1987. – 94 - 99 pp.

119. Watson, C.S. Advances in Computer-based speech training (CBST): Aids for the profoundly hearing impaired / P. D. Kewley, C.S. Watson, P.A. // *Research on the Use of Sensory Aids for Hearing-Impaired Persons*, N. McGarr, (Ed.), *Volta Review*, 1989. – № 91(4). – 29-45 pp.

120. Проблемы и перспективы комплексной психолого-педагогической реабилитации и ресоциализации пациентов после хирургического лечения опухолей головы и шеи / М. Ш. Магомед-Эминов, О. С. Орлова, Д. В. Уклонская, Ю. М. Хорошкова // *Специальное образование*. – 2018. – № 2(50). – С. 50-62.

121. Патент № 2286812 С1 Российская Федерация, МПК А61N 2/00, G09B 5/04. Способ восстановления звучной речи у больных после полного удаления гортани: № 2005108699/14 : заявл. 28.03.2005: опубл. 10.11.2006 / Л. Н. Балацкая, Е. Л. Чойнзонов, Е. А. Красавина [и др.].

122. Кожанов, А. Л. Современные аспекты лечения и реабилитации больных при раке гортани / А. Л. Кожанов // *Опухоли головы и шеи*. – 2016. – Т. 6. – № 2. – С. 17-25.

123. Патент № 2722818 С1 Российская Федерация, МПК А61N 2/06, А61N 5/067. Способ реабилитации больных после оперативных вмешательств при опухолях головы и шеи с нарушением голосовой и речевой функций : № 2019104447 : заявл. 18.02.2019 : опубл. 04.06.2020 / Т. Я. Кучерова, Л. Н. Балацкая, Е. Л. Чойнзонов [и др.].

124. Методы формирования устной речи уларингэктомированных пациентов (обзор литературы) / Н. А. Дайхес, Е. В. Осипенко, О. С. Орлова [и др.] // *Оториноларингология. Восточная Европа*. – 2021. – Т. 11. – № 3. – С. 334-347.

125. Уклонская, Д. В. Психологические особенности лиц с удаленной гортанью как фактор успешности реабилитации речевой функции / Д. В. Уклонская, Ю. М. Хорошкова // *Современные проблемы науки и образования*. – 2016. – № 3. – С. 353.

126. Устранение нарушений голосовой функции при открытых диагональных резекциях гортани / А. Л. Кожанов, Л. Г. Кожанов, В. Н. Сорокин, Е. С. Романова // *Опухоли головы и шеи*. – 2018. – Т. 8. – № 4. – С. 32-38.

127. Применение биологической обратной связи в комплексной реабилитации больных после полного удаления гортани / Л. Н. Балацкая, Е.

Л. Чойнзонов, Е. А. Красавина [и др.] // Сибирский онкологический журнал. – 2004. – № 4. – С. 17-21.

128. Бердникович, Е. С. Персонафицированный подход в речевой реабилитации: фокус на пациенте / Е. С. Бердникович, О. С. Орлова, Д. В. Уклонская // Специальное образование. – 2022. – № 1(65). – С. 20-34.

129. Управление процессом речевой реабилитации на основе биологической обратной связи / А. Б. Миронов, С. А. Пахандрин, Д. С. Иванов, В. П. Бондаренко // Научная сессия ТУСУР - 2006 : Материалы докладов Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых "Научная сессия ТУСУР - 2006", посвященной 75-летию Ф. И. Перегудова: в 5 частях, Томск, 04–07 мая 2006 года. – Томск: В-Спектр, 2006. – С. 37-39.

130. Колмогорова, А. В. Языковая / речевая биография пациента с афазией как фактор, определяющий специфику процесса его речевой реабилитации / А. В. Колмогорова, С. А. Лямзина, И. Л. Киселев // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2020. – Т. 18. – № 2. – С. 103-116.

131. Поливара, З. В. Анализ механизмов речи в реабилитации: роль речевых кинестезий для развития фонематических процессов / З. В. Поливара, И. И. Чайковская // Университетская медицина Урала. – 2018. – Т. 4. – № 1(12). – С. 76-78.

132. Харитонов, Д. Ю. Особенности речевой реабилитации пациентов после хирургических вмешательств в зубочелюстной системе / Д. Ю. Харитонов, Н. Е. Митин, А. Е. Устюгова // Современные проблемы науки и образования. – 2017. – № 5. – С. 26.

133. ГОСТ Р 50840–95 Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости [Электронный ресурс] / Режим доступа: <http://docs.cntd.ru/document/1200027288> (дата обращения 01.04.2022).

134. Куликова, И. С. Введение в языкознание в 2 ч. Часть 1: учебник для вузов / И. С. Куликова, Д. В. Салмина. — М: Юрайт, 2018. — 366 с.

135. Speech quality measurement automation for patients with cancer of the oral cavity and oropharynx / R. V. Meschryakov, E. Y. Kostyuchenko, D. I. Ignatieva [et al.] // 2016 International Siberian Conference on Control and Communications, SIBCON 2016 - Proceedings, Moscow, 12–14 мая 2016 года. – Moscow, 2016. – P. 5.

136. Гмурман В. Е. Теория вероятностей и математическая статистика: Учебное пособие для вузов. – 10-е издание, стереотипное. – М.: Высшая школа, 2004. – 479 с.

137. MedFind. Oncology. Plastic surgery in the surgical treatment of tumors of the face and jaws [Электронный ресурс] / Режим доступа: <http://medfind.ru/modules/sections/index.php?op=viewarticle&artid=324> (дата обращения 5.09.2021).

138. Korzhov I.S. Features of diagnostics of speech disorders in cancer patients with acquired maxillary defects // *Siberian Journal of Oncology*, 2015. – №1. –84-88 pp.

139. Бондарко, Л.В. Звуковой строй современного русского языка // М.: Просвещение, 1977. – 175 с.

140. Возможности речевой реабилитации больных раком полости рта и ротоглотки после реконструктивно-пластических операций / Е. А. Красавина, Е. Л. Чойнзонов, Д. Е. Кульбакин [и др.] // *Сибирский онкологический журнал*. – 2020. – Т. 19. – № 5. – С. 35-43.

141. Тризна, Н. М. Речевые нарушения у пациентов после хирургического лечения опухолей полости рта и ротоглотки / Н. М. Тризна, Ж. В. Колядич, А. А. Жалейко // *Оториноларингология. Восточная Европа*. – 2021. – Т. 11. – № 1. – С. 29-40.

142. Патент № 2760173 С1 Российская Федерация, МПК А61Н 1/00. Способ исправления дефектов произношения заднеязычных звуков, возникших после хирургического лечения у больных раком органов полости рта и ротоглотки: № 2020137816 : заявл. 17.11.2020 : опубл. 22.11.2021 / Е. А. Красавина, Е. Л. Чойнзонов, Д. Е. Кульбакин, Л. Н. Балацкая.

143. Тризна, Н. М. Восстановление речевой функции у пациентов после хирургического лечения опухолей полости рта и ротоглотки / Н. М. Тризна, Ж. В. Колядич, Е. А. Жалейко // *Онкология и радиология Казахстана*. – 2020. – № 4(58). – С. 47-52.

144. Тризна, Н. М. Медицинская реабилитация пациентов с опухолями полости рта и ротоглотки / Н. М. Тризна, А. Г. Жуковец, Ж. В. Колядич // *Онкологический журнал*. – 2021. – Т. 15. – № 1(57). – С. 57-62.

145. Пятков, А.В. Изменения на спектрограммах Фурье при произнесении фонем [т] и [т'] при речевой реабилитации / А.В. Пятков, Д.И. Игнатьева, Е.Ю. Костюченко // *Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции*. – 2016. – № 1–1. – С. 168-170.

146. Игнатьева, Д. И. Изменения при произнесении фонем [к] и [к'] на спектрограммах КЛП и изображениях автокорреляции / Д. И. Игнатьева, А. В. Пятков, Е. Ю. Костюченко // *Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции*. – 2016. – № 1-1. – С. 173-175.

147. Игнатьева, Д. И. Изменения в спектрограммах КЛП и на изображениях автокорреляции у фонем [т] и [т'] / Д. И. Игнатьева, А. В. Пятков, Е. Ю. Костюченко // *Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции*. – 2016. – № 1-1. – С. 166-168.

148. Романенко, А.А. Выравнивание временных рядов: прогнозирование с использованием DTW [Электронный ресурс] / Режим доступа: <http://jmla.org/papers/doc/2011/no1/Romanenko2011Dynamic.pdf>. (дата обращения: 22 03 2022).

149. Keogh, E. J. Scaling up Dynamic Time Warping to Massive Datasets. Principles of Data Mining and Knowledge Discovery / E. J. Keogh, M. J. Pazzani // Prague: Springer Berlin Heidelberg, 1999. – 1-11 pp.
150. Salvador, S. Fastdtw: Toward accurate dynamic time warping in linear time and space / S. Salvador, P. Chan // Workshop on Mining Temporal and Sequential Data, Seattle, 2004. – 70-80 pp.
151. Теслер, Г.С. Метрики и нормы в иерархии категориальных семантик и функций / Г.С. Теслер // Математические машины и системы. – 2005 - № 2. С. 63 -75
152. Новохрестова, Д. И. Временная нормализация слогов алгоритмом динамической трансформации временной шкалы при оценке качества произнесения слогов в процессе речевой реабилитации / Д. И. Новохрестова // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2017. – Т. 20. – № 4. – С. 142-145.
153. Resample uniform or nonuniform data to new fixed rate [Электронный ресурс] / Режим доступа: <https://www.mathworks.com/help/signal/ref/resample.html> (дата обращения 1.04.2022).
154. Correlation normalization of syllables and comparative evaluation of pronunciation quality in speech rehabilitation / E. Kostyuchenko, R. Meshcheryakov, D. Ignatieva [et al.] // Lecture Notes in Computer Science. – 2017. – Vol. 10458 LNAI. – P. 262-271.
155. Харченко, С. С. Проблема сегментации на фонемы при оценке качества произнесения слогов в рамках речевой реабилитации / С. С. Харченко, Д. И. Новохрестова, Е. Ю. Костюченко // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. – 2018. – № 1-1. – С. 223-226.
156. Evaluation of the speech quality during rehabilitation after surgical treatment of the cancer of oral cavity and oropharynx based on a comparison of the fourier spectra / E. Kostyuchenko, R. V. Mescheryakov, D. Ignatieva [et al.] // Lecture Notes in Computer Science. – 2016. – Vol. 9811. – P. 287-295.
157. Correlation criterion in assessment of speech quality in process of oncological patients rehabilitation after surgical treatment of the speech-producing tract / E. Kostyuchenko, R. Meshcheryakov, D. Ignatieva [et al.] // Advances in Intelligent Systems and Computing. – 2019. – Vol. 759. – P. 209-216
158. Ходашинский, И. А. Идентификация нечетких систем на основе прямого алгоритма муравьиной колонии / И. А. Ходашинский, П. А. Дудин // Искусственный интеллект и принятие решений. – 2011. – № 3. – С. 26-33.
159. Гладков, Л.А. Генетические алгоритмы / Л.А. Гладков, В.В. Курейчик, В. М. Курейчик // Под ред. В. М. Курейчика. 2-е изд., испр. и доп. М.: ФИЗМАТЛИТ, 2006. – 320 с.
160. Новохрестова, Д. И. Гибридная мера для сравнения записей слогов / Д. И. Новохрестова // Материалы XXVII Международной научно-

технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2022». – Томск, 18-20 мая 2022 г. (в печати).

161. Experimental Analysis of Expert and Quantitative Estimates of Syllable Recordings in the Process of Speech Rehabilitation / D. Novokhrestova, E. Kostuchenko, I. Hodashinsky, L. Balatskaya // Lecture Notes in Computer Science. – 2021. – Vol. 12997 LNAI. – P. 483-491

162. Новохрестова, Д. И. Комбинированная оценка качества произношения слогов / Д. И. Новохрестова // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. – 2021. – № 1–2. – С. 299-301.

163. Nassi, I. Flowchart Techniques for Structured Programming / I. Nassi, B. Shneiderman // SIGPLAN Notices 8, 1973.

164. System.IO Пространство имен [Электронный ресурс]. – Режим доступа: [https://msdn.microsoft.com/ru-ru/library/mt481548\(v=vs.110\).aspx](https://msdn.microsoft.com/ru-ru/library/mt481548(v=vs.110).aspx) (дата обращения 01.04.2022)

165. Новохрестова, Д. И. Оценка разборчивости произношения слогов: метод и алгоритмы / Д. И. Новохрестова, Е. Ю. Костюченко, А. В. Пятков // Автоматика и программная инженерия. – 2018. – № 3(25). – С. 30-34.

166. Model of system quality assessment pronouncing phonemes / E. Kostyuchenko, D. Ignatieva, R. Mescheryakov [et al.] // 2016 Dynamics of Systems, Mechanisms and Machines, Dynamics 2016, Omsk, 15–17 ноября 2016 года. – Omsk, 2016. – P. 7819016

167. Методы и модель оценки качества произношения слогов при речевой реабилитации / Д. И. Новохрестова, Е. Ю. Костюченко, Л. Н. Балацкая, Е. Л. Чойнзонов. – Москва: ООО "Научный инновационный центр Международный институт стратегических исследований", 2019. – 69 с.

168. Волкова, Л.С. Логопедия: Учебник для студентов дефектологических факультетов педагогических высш. учеб. Заведений // Владос, 2009. – 703 с.

169. Биологическая обратная связь при обучении устной речи / Л.Н. Балацкая и др. // Сборник трудов XVI сессии Российского акустического общества. – 2005. – С. 7-10.

170. Волкова, С.В. Анализ результатов коррекционного воздействия с применением бос технологий в комплексной реабилитации детей с нарушениями речи / С.В. Волкова, С.В. Ястребова, Н.О. Карасева // Наука и образование: новое время. – 2016. – № 5. – С. 115-123.

171. Макарецва, Е.В. Использование метода биологической обратной связи (БОС) в работе с заикающимися детьми // Актуальные проблемы коррекционной педагогики и специальной психологии. Материалы VII Всероссийской (с международным участием) научно-практической конференции. – Череповец. – 2012. – С. 266-267.

172. Ивановский, Ю.В. Принципы использования метода биологической обратной связи в системе медицинской реабилитации / Ю.В.

Ивановский, А.А. Сметанкин // Биологическая обратная связь, 2000. – Т. 3. – 2-9 с.

173. Vibrotactile Biofeedback System and Bilateral Vestibular Loss: Pilot Study [Электронный ресурс] // Режим доступа: <http://medcraveonline.com/JOENTR/JOENTR-04-00124.php> (дата обращения 15.03.2022).

174. Egner, T. EEG signature and phenomenology of alpha/theta neurofeedback training versus mock feedback / T. Egner, E. Strawson, J.H. Gruzelier // Applied psychophysiology and biofeedback, 2002. – №27(4). – 261-270 pp.

175. Blanchard, B. Biofeedback treatments of essential hypertension / B. Blanchard // Biofeedback and Selfregulation, 1990. – №15(3). – 209-228 pp.

176. Biofeedback & Self-Control. An Aldine Annual on the Regulation of Bodily Processing and Consciousness // Chicago: Aldine Publishing Company, 1976. – 581 p.

177. Боксер, О.Я. Функциональные биотехнические системы / О.Я. Боксер // Вестник РАМН, 1999. – №6. – С. 29–34.

178. Захарова, В. В. Биоуправление. Итоги и перспективы развития / В. В. Захарова, К. Роберт, Э. М. Сохадзе // Биоуправление-2: теория и практика, 1993. – С. 13–19.

179. Биоуправление в психоневрологической практике / А.Б. Скок, О.В. Филатова, М.Б. Штарк, О.А. Шубина // Бюллетень Сибирского отделения РАМН, 1999. – №91(1). – С. 30-35.

180. Agnihotri, H. Biofeedback approach in the treatment of generalized anxiety disorder / H. Agnihotri, M. Paul, J.S. Sandhu // 2007. 2007 г., Iran: Psychiatry, 2007. – №2. – 90-95 pp.

181. Tsutsui, S. Biofeedback therapy in chronic headaches — prognostic investigation / S. Tsutsui, K. Tsuboi, Y. Nakagawa // Current Biofeedback Research in Japan, 1993. – 97 - 102 pp.

182. Корнилов А.Ю. Управление процессом речевой реабилитации с использованием биологической обратной связью: дис. ... канд. техн. наук: 05.13.01 / Корнилов Александр Юрьевич. – Томск, 2005. – 139 с.

183. Орлова, О. С. Комплексный персонифицированный подход в реабилитации больных с дисфониями / О.С. Орлова, П.А. Эстрова, А.С. Калмыкова // Современные технологии в диагностике и лечении патологии гортани и смежных дисциплинах. – Омск: ОмГМУ, 2016. – С. 51.

184. Клачкова, О. В. Влияние метода биологической обратной связи (БОС) на общее развитие, формирование речи у детей с ограниченными возможностями / О.В. Клачкова // Отечественный и международный опыт в решении современных педагогических проблем: Материалы международной заочной электронной научной конференции. – Магадан: ООО "Центр научного сотрудничества "Интерактив плюс", 2015. – СС. 97-103.

185. Biofeedback, Fourth Edition: A practitioner's guide / edited by M. Schwartz and F. Andrasik // NY: Guilford Press, 2017.

186. Богданов, О.В. Эффективность различных форм сигналов обратной связи в ходе лечебных сеансов функционального / О.В. Богданов, Д.Ю. Пинчук, Е.Л. Михайленок // 1990. №1, 1990 г., Физиология человека, 1990. – №1, Т. 16. – С. 13-17.

187. Активность мотивационных систем положительного и отрицательного подкрепления и фоновые показатели артериального давления у человека/ Афтанас Л. И., Сидорова П. В., Павлов С. В. [и др.] // Российский физиологический журнал им. И.М. Сеченова, 2007. – №12, Т. 93. – С. 1362 - 1373.

188. Зубкова, Е. И. Влияние фрагмента стресс-реализующего гормона АКТГ на системы положительного и отрицательного подкрепления / Е. И. Зубкова, О. И. Кириллова, А. П. Салей // Эмоции и поведение: системный подход. – М.: Министерство здравоохранения СССР, 1984.

189. Нейробиологические механизмы систем награды и наказания в головном мозге при активации прилежащего ядра / М.В. Шевелева, А.А. Лебедев, Р.О. Роик, П. Д. Шабанов // Обзоры по клинической фармакологии и лекарственной терапии, 2013. – №3, Т. 11. – С. 3-19.

190. Харченко С.С. Математическая электроакустическая модель псевдоголоса и программный комплекс голосовой реабилитации пациентов после ларингэктомии на основе бионических принципов: дис. ... канд. техн. наук: 05.13.01 / Харченко Сергей Сергеевич. – Томск, 2017. – 114 с.

191. Способ восстановления речевой функции у больных раком полости рта и ротоглотки после органосохраняющих операций: пат. 2694516 С1 Рос. Федерация: МПК7 А61Н 1/00 // Л. Н. Балацкая, Е. Л. Чойнзонов, Е. А. Красавина, С. Ю. Чижевская, Д. И. Новохрестова, Е. Ю. Костюченко.

192. Новохрестова, Д. И. Реализация биологической обратной связи в рамках оценки качества речи / Д. И. Новохрестова // Материалы XXVII Международной научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2022». – Томск, 18-20 мая 2022 г. (в печати).

193. Серов, Н. Светоцветовая терапия. Смысл и значение цвета / Н. Серов. – 2-е издание, переработанное. – Санкт-Петербург: Речь, 2002. – 160 с.

194. Саламатова, А. А. Психологические основы восприятия зеленого цвета / А. А. Саламатова // Приоритетные направления развития науки в современном мире: Сборник статей по материалам I международной научно-практической конференции, Уфа, 13 мая 2019 года. – Уфа: Общество с ограниченной ответственностью "Научно-издательский центр "Вестник науки", 2019. – С. 145-150.

195. Структура и база данных программного обеспечения оценки качества и разборчивости речи в процессе реабилитации после операции при лечении рака полости рта и ротоглотки, челюстнолицевой области / Е.

Ю. Костюченко, Р. В. Мещеряков, Л. Н. Балацкая, Е. Л. Чойнзонов // Труды СПИИРАН. – 2014. – № 1(32). – С. 116-124.

196. Свидетельство о государственной регистрации базы данных № 2018620852 Российская Федерация. Звукозаписи пациентов с заболеваниями органов речеобразующего тракта OnkoSpeechDB : № 2018620017 : заявл. 09.01.2018 : опубл. 13.06.2018 / Е. Ю. Костюченко, Р. В. Мещеряков, Д. И. Новохрестова [и др.].

197. Костюченко, Е. Ю. Формирование базы данных пациентов при речевой реабилитации после комбинированного лечения онкологических заболеваний органов речеобразующего тракта / Е. Ю. Костюченко, Д. И. Новохрестова, А. В. Пятков // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. – 2017. – № 1-2. – С. 245-247.

198. Программный комплекс по оценке разборчивости речи / Д. И. Новохрестова, Е. Ю. Костюченко, Е. С. Катаева [и др.] // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2019. – Т. 22. – № 3. – С. 43-48.

199. Novokhrestova, D. Automation of speech quality assessment in speech rehabilitation / D. Novokhrestova, E. Kostyuchenko, E. Kosenko // International Journal of Advanced Trends in Computer Science and Engineering. – 2019. – Vol. 8. – No 6. – P. 3026-3030.

200. Software for an objective evaluation of the quality of syllables's pronunciation in speech rehabilitation / E. Y. Kostyuchenko, R. V. Mescheryakov, D. I. Novokhrestova [et al.] // Proceedings of 2017 IEEE 2nd International Conference on Control in Technical Systems, CTS 2017 : 2, St. Petersburg, 25–27 октября 2017 года. – St. Petersburg, 2017. – P. 267-270.

201. Программное обеспечение для объективной оценки качества произнесения слогов при речевой реабилитации / Е. Ю. Костюченко, Р. В. Мещеряков, Д. И. Новохрестова [и др.] // Международная научная конференция по проблемам управления в технических системах. – 2017. – Т. 1. – С. 277-280.

202. Простой, но эффективный Voice Activity Detection алгоритм реального времени [Электронный ресурс]. – Режим доступа: <https://habrahabr.ru/post/192954/> (дата обращения 24.03.2022)

203. NAudio. Audio and MIDI library for .NET [Электронный ресурс]. – Режим доступа: <https://github.com/naudio/NAudio> (дата обращения 22.03.2022)

204. Алгоритм детектирования голосовой активности / Е. Ю. Костюченко, Р. В. Мещеряков, Д. И. Новохрестова [и др.] // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. – 2017. – № 1-2. – С. 250-252.

205. Новохрестова, Д. И. Алгоритм детектирования голосовой активности в программном комплексе по оценке разборчивости речи / Д. И. Новохрестова, Е. Ю. Костюченко // Электронные средства и системы

управления. Материалы докладов Международной научно-практической конференции. – 2019. – № 1-2. – С. 138-140.

206. Е. Ю. Костюченко, Р. В. Мещеряков, Д. И. Новохрестова [и др.] *Speech quality assessment* // Свидетельство о государственной регистрации программы для ЭВМ № 2018613005. Дата регистрации в реестре: 09.01.2018.

207. Е. Ю. Костюченко, Л. Н. Балацкая, Д. И. Новохрестова [и др.] *OnkoSpeech v1.0* // Свидетельство о государственной регистрации программы для ЭВМ № 2019610957. Дата регистрации в реестре: 29.12.2018.

208. Д. И. Новохрестова, Е. Ю. Костюченко, С. С. Харченко [и др.] Программное обеспечение для объективной оценки качества произнесения слогов при речевой реабилитации "OnkoSpeech v2.0" // Свидетельство о государственной регистрации программы для ЭВМ № 2020610774. Дата регистрации в реестре: 25.12.2019.

209. Е. Ю. Костюченко, Д. И. Новохрестова, С. С. Харченко [и др.] *OnkoSpeech v3.0* // Свидетельство о государственной регистрации программы для ЭВМ № 2021611404. Дата регистрации в реестре: 19.01.2021.

210. Kostyuchenko, E. *Speech rehabilitation after combined treatment of cancer and the formation of a set of syllables for assessing speech quality* / E. Kostyuchenko, D. Novokhrestova // *CEUR Workshop Proceedings: Proceedings of the 3rd International Conference on R. Piotrowski's Readings in Language Engineering and Applied Linguistics, PRLEAL 2019, Saint Petersburg, 27 ноября 2019 года.* – Saint Petersburg, 2020. – P. 73-84.

211. Восстановление речевой функции у больных раком полости рта и ротоглотки с использованием инновационных технологий / Л. Н. Балацкая, Е. Л. Чойнзонов, Е. А. Красавина, Е.Ю. Костюченко, Д.И. Новохрестова // *Вопросы онкологии.* – 2020. – Т. 66. – № 3. – С. 247-251.

Приложение А – Патент на изобретение

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2694516

**Способ восстановления речевой функции у больных раком
полости рта и ротоглотки после органосохраняющих
операций**

Патентообладатель: *Федеральное государственное бюджетное
научное учреждение "Томский национальный
исследовательский медицинский центр" Российской
академии наук" (Томский НИМЦ) (RU)*

Авторы: *Балацкая Лидия Николаевна (RU), Чойнзонов Евгений
Лхамациренович (RU), Красавина Елена Александровна (RU),
Чижевская Светлана Юрьевна (RU), Новохрестова Дарья
Игоревна (RU), Костюченко Евгений Юрьевич (RU)*

Заявка № 2019104446

Приоритет изобретения 18 февраля 2019 г.

Дата государственной регистрации в

Государственном реестре изобретений

Российской Федерации 15 июля 2019 г.

Срок действия исключительного права

на изобретение истекает 18 февраля 2039 г.



Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Израиль

Приложение Б – Свидетельства о государственной регистрации

РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU2021611404**

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):

2021611404

Дата регистрации: 27.01.2021

Номер и дата поступления заявки:

2021610348 19.01.2021

Дата публикации и номер бюллетеня:

27.01.2021 Бюл. № 2

Автор(ы):

Костюченко Евгений Юрьевич (RU),

Новохрестова Дарья Игоревна (RU),

Харченко Сергей Сергеевич (RU),

Чойнзонов Евгений Лхаматирович (RU),

Балацкая Лидия Николаевна (RU)

Правообладатель(и):

Федеральное государственное бюджетное

образовательное учреждение высшего

образования «Томский государственный

университет систем управления и

радиоэлектроники» (RU)

Название программы для ЭВМ:

OnkoSpeech v3.0

Реферат:

Программа проводит непосредственную запись файлов в формате WAV, которые могут содержать произнесение слогов или фраз из ГОСТ 50840-95. Возможна запись слогов из специальных таблиц, ориентированных на работу с фонемами, наиболее подверженными изменению после проведения операций на органах речеобразующего тракта. Оценивается качество произнесения слогов и качество произношения фраз с использованием системы распознавания речи с оценкой фразовой и словесной разборчивости, а также разборчивости на уровне символов (звуков). Программа разработана в рамках гранта РФФИ № 16-15-00038 «Восстановление речевой функции с использованием технических методов и математического моделирования у больных раком полости рта и ротоглотки после хирургического лечения». Регистрационный номер НИОКТР АААА-А19-119112990048-3. Тип ЭВМ: IBM PC-совмест. ПК. ОС: Windows 7/8/10.

Язык программирования:

C#

Объем программы для ЭВМ:

351 КБ

РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU2020610774**

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства): 2020610774 Дата регистрации: 20.01.2020 Номер и дата поступления заявки: 2019667335 25.12.2019 Дата публикации и номер бюллетеня: 20.01.2020 Бюл. № 1	Автор(ы): Новохрестова Дарья Игоревна (RU), Костюченко Евгений Юрьевич (RU), Харченко Сергей Сергеевич (RU), Чойнзонов Евгений Лхамцэренович (RU), Балацкая Лидия Николаевна (RU) Правообладатель(и): федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (RU)
--	---

Название программы для ЭВМ:
Программное обеспечение для объективной оценки качества произнесения слогов при речевой реабилитации «OnkoSpeech v2.0»

Реферат:

Программа проводит непосредственную запись таких файлов, которые могут содержать произнесение слогов или фраз из ГОСТ 50840-95 Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. Кроме того, возможна запись слогов из специальных таблиц, ориентированных именно на работу с фонемами, наиболее подверженными изменению после проведения конкретного вида операции на органах речеобразующего аппарата. Программа способна проводить оценку качества произнесения слогов, полученных в процессе речевой реабилитации, по сравнению с эталонными записями, осуществляемыми перед проведением операции. Также имеется возможность проведения оценки качества произношения фраз на основе распознавания фраз с использованием системы распознавания речи с оценкой как фразовой, так и словесной разборчивости. Тип ЭВМ: IBM PC-совмест. ПК; ОС: Windows 7/8/10.

Язык программирования: C#
Объем программы для ЭВМ: 13307 Кб

РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU 2019610957**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ**

Номер регистрации (свидетельства): 2019610957 Дата регистрации: 18.01.2019 Номер и дата поступления заявки: 2018665509 29.12.2018 Дата публикации и номер бюллетеня: 18.01.2019 Бюл. № 1 Контактные реквизиты: Нет	Автор(ы): Костюченко Евгений Юрьевич (RU), Балацкая Лидия Николаевна (RU), Новохрестова Дарья Игоревна (RU), Пятков Александр Владиславович (RU), Чойнзонов Евгений Лхаматцренович (RU) Правообладатель(и): Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР) (RU)
--	---

Название программы для ЭВМ:
OnkoSpeech v1.0

Реферат:

Объектом программы являются файлы, содержащие звукозаписи сеансов, полученные при работе с пациентами. Программа предназначена для автоматизации процесса проведения речевой реабилитации пациентов после оперативного вмешательства при лечении органов речеобразующего тракта. Проведен принципиальный ряд доработок по сравнению с программным комплексом, используемым для оценки качества произнесения слогов: добавлен модуль автоматической сегментации слога на фонемы. В данной версии в дальнейшую работу идут только фонемы, выделенные в начале слога, остальная же часть слога обрабатывается целиком (из-за большого количества ошибок алгоритма при других положениях фонем). Модифицирован модуль получения количественной оценки для возможности ее расчета в реальном времени с учетом добавленного модуля сегментации. Добавлен модуль визуализации результатов для наглядного их представления пациенту при организации биологической обратной связи в рамках процедуры речевой реабилитации. Вследствие принципиальной значимости изменений, позволивших от простого получения оценки качества перейти именно к процедуре речевой реабилитации изменено название программного обеспечения. Работа выполнена в рамках гранта Российского Научного Фонда «Восстановление речевой функции с использованием технических методов и математического моделирования у больных раком полости рта и ротоглотки после хирургического лечения», № 16-15-00038. Регистрационный номер проекта НИОКТР АААА-А17-117070620027-7.

Язык программирования: C#
Объем программы для ЭВМ: 40864 Кб

РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU 2018620852**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ БАЗЫ ДАННЫХ, ОХРАНЯЕМОЙ
АВТОРСКИМИ ПРАВАМИ**

Номер регистрации (свидетельства): 2018620852 Дата регистрации: 13.06.2018 Номер и дата поступления заявки: 2018620017 09.01.2018 Дата публикации и номер бюллетеня: 13.06.2018 Бюл. № 6 Контактные реквизиты: нет	Автор(ы): Костюченко Евгений Юрьевич (RU), Мещеряков Роман Валерьевич (RU), Новохрестова Дарья Игоревна (RU), Пятков Александр Владиславович (RU), Чойнзонов Евгений Лхаматренович (RU), Балацкая Лидия Николаевна (RU) Правообладатель(и): Федеральное государственное бюджетное образовательное учреждение высшего образования «Гомский государственный университет систем управления и радиоэлектроники» (ТУСУР) (RU)
---	--

Название базы данных:

Звукозаписи пациентов с заболеваниями органов речеобразующего тракта OnkoSpeechDB**Реферат:**

Объектом базы данных являются файлы, содержащие записи пациентов, сделанные перед началом комбинированного лечения и после его прохождения – перед началом речевой реабилитации и по мере ее прохождения. Содержит данные о пациентах, заболевании и объеме операции, описание проводимой оценки, ссылки на записи, проведенные в процессе работы с пациентами и оценки качества произнесения, полученные различными способами. Предназначена для проведения исследований в области речевой реабилитации пациентов после комбинированного лечения органов речеобразующего тракта. Выполнена в рамках гранта Российского Научного Фонда «Восстановление речевой функции с использованием технических методов и математического моделирования у больных раком полости рта и ротоглотки после хирургического лечения», № 16-15-00038. Регистрационный номер проекта НИОКТР АААА-А17-117070620027-7.

Вид и версия системы управления базой данных: MySQL

Объем базы данных: 344,7 Мб

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2018613005

SPEECH QUALITY ASSESSMENT

Правообладатель: *Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР) (RU)*

Авторы: *Костюченко Евгений Юрьевич (RU), Мещеряков Роман Валерьевич (RU), Новохрестова Дарья Игоревна (RU), Пятков Александр Владиславович (RU), Чойнзонов Евгений Лхамцыренович (RU), Балацкая Лидия Николаевна (RU)*


Заявка № 2018610111

Дата поступления 09 января 2018 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 01 марта 2018 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ивлиев



Приложение В – Акты внедрения

Федеральное государственное
бюджетное научное учреждение
«Томский национальный
исследовательский медицинский центр
Российской академии наук»



**НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ
ИНСТИТУТ ОНКОЛОГИИ
(НИИ онкологии)**

Кооперативный пер., д. 5, Томск, 634009
Тел. (3822) 51 10 39 / 51 33 06
Факс (3822) 51 33 06
E-mail: onco@tntmc.ru

ОКПО 15601567, ОГРН 1027000861568
ИНН / КПП 7019011979 / 701745014

№ _____
На № _____ от _____



АКТ

о внедрении результатов диссертационной работы
Новохрестовой Дарьи Игоревны

Комиссия в составе:

Председатель комиссии:

Чойнзонов Е.Л. - д.м.н., профессор, академик РАН, директор НИИ онкологии Томского НИМЦ, заведующий отделением опухолей головы и шеи НИИ онкологии Томского НИМЦ.

Члены комиссии:

Балацкая Л.Н. - д.б.н., ведущий научный сотрудник НИИ онкологии Томского НИМЦ;

Красавина Е.А. - к.б.н., логопед НИИ онкологии Томского НИМЦ.

составили настоящий акт о нижеследующем.

Результаты диссертационной работы Новохрестовой Д.И. «Методики и алгоритм анализа данных при оценке качества произношения слогов в процессе речевой реабилитации», представленной на соискание ученой степени кандидата технических наук, внедрены в деятельность НИИ онкологии ФГБНУ «Томский НИМЦ РАН» в процессе речевой реабилитации после комбинированного лечения опухолей органов полости рта и ротоглотки.

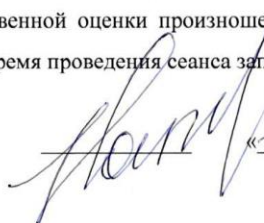
Программный комплекс по оценке речи, включающий в себя разработанные методики и алгоритмы, используется в процессе проведения сеансов речевой реабилитации для записи речи пациентов и для количественной оценки качества произношения слогов.

Применение программного комплекса позволило автоматизировать процесс записи сеансов слоговой разборчивости, учитывать индивидуальные особенности речи пациента за счет

применения его же речи в качестве эталонной, а также применять в процессе записи сеанса биологическую обратную связь для отражения динамики восстановления речи пациента.

За время внедрение программного комплекса была записана речь 62 пациентов, 43 из которых проходили речевую реабилитацию с его использованием, записано 136 сеансов оценки произношения слогов согласно разработанному в рамках диссертационного исследования списку слогов с наиболее проблемными фонемами и более 12 500 аудиозаписей слогов. Применение разработанных методик и алгоритмов количественной оценки произношения слогов позволило сократить время проведения оценки на 64,86%, время проведения сеанса записи слогов на 55,7%.

Чойнзонов Е.Л.
д.м.н., профессор, академик РАН,
директор НИИ онкологии Томского НИМЦ,
заведующий отделением опухолей головы
и шеи НИИ онкологии Томского НИМЦ

 «15» 04 2022 г.

Балацкая Л.Н.
д.б.н., ведущий научный сотрудник
НИИ онкологии Томского НИМЦ

 «15» 04 2022 г.


Красавина Е.А.
к.б.н., логопед НИИ онкологии Томского НИМЦ

 «15» 04 2022 г.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ
И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)

УТВЕРЖДАЮ

Проректор по учебной работе ТУСУР

 П.В. Сенченко

« 8 » 04 2022 г.

АКТ

о внедрении результатов диссертационной работы
Новохрестовой Дарьи Игоревны в учебный процесс

Комиссия в составе:

Давыдова Е.М., к.т.н., декан факультета безопасности ТУСУР –
председатель комиссии;

Конев А.А., к.т.н., доцент кафедры КИБЭВС ТУСУР;

Рахманенко И.А., к.т.н., доцент кафедры БИС ТУСУР

составила настоящий акт о нижеследующем.

Результаты диссертационной работы Новохрестовой Д.И. используются в учебном процессе на факультете безопасности ТУСУР при чтении курса лекций и проведении практических занятий по дисциплинам «Методы анализа данных», «Методы оптимизации», «Наборы данных в задачах искусственного интеллекта» и «Математические методы теории сигналов и систем» для подготовки специалистов по защите информации, обучающихся по специальностям «10.05.04 – Информационно-аналитические системы безопасности» и «10.05.02 – Информационная безопасность телекоммуникационных систем».

В курсах «Методы анализа данных» и «Методы оптимизации» используются результаты работы Новохрестовой Д.И. по исследованию метрик и составлению гибридной меры, позволяющие студентам ознакомиться с применимостью метрик к решению практической задачи анализа речи и подходом к формированию


гибридной меры на основе комбинации метрик с подбором коэффициентов значимости с использованием методов оптимизации.

В курсе «Наборы данных в задачах искусственного интеллекта» используется составленный в процессе выполнения работы набор аудиозаписей, предназначенный для исследования речи в том числе методами искусственного интеллекта.


В курсе «Математические методы теории сигналов и систем» используются результаты работы Новохрестовой Д.И. по разработке алгоритма количественной оценки схожести двух речевых сигналов, позволяющие студентам ознакомиться с возможностью применения различных спектрограмм аудиосигналов в рамках решения практической задачи оценки речи.

Кроме того, студенты факультета безопасности имеют возможность ознакомиться с результатами диссертационного исследования в ходе выполнения групповых проектов, научно-исследовательских и дипломных работ и использовать их в практических работах по анализу речевого сигнала.


Давыдова Е.М.
к.т.н., декан факультета
безопасности ТУСУР

 « 7 » 04 2022 г.

Конев А.А.
к.т.н., доцент кафедры
КИБЭВС ТУСУР

 « 7 » 04 2022 г.

Рахманенко И.А.
к.т.н., доцент кафедры
БИС ТУСУР

 « 7 » 04 2022 г.

**Приложение Г – Сертификат гранта Американского акустического
сообщества (ASA)**

Acoustical Society of America



The Acoustical Society of America recognizes

Darya Igorevna Novokhrestova
Tomsk State University of Control Systems and Radioelectronics

as recipient of the

ASA International Student Grant

to assist the research of promising graduate students in acoustics

20 December 2020

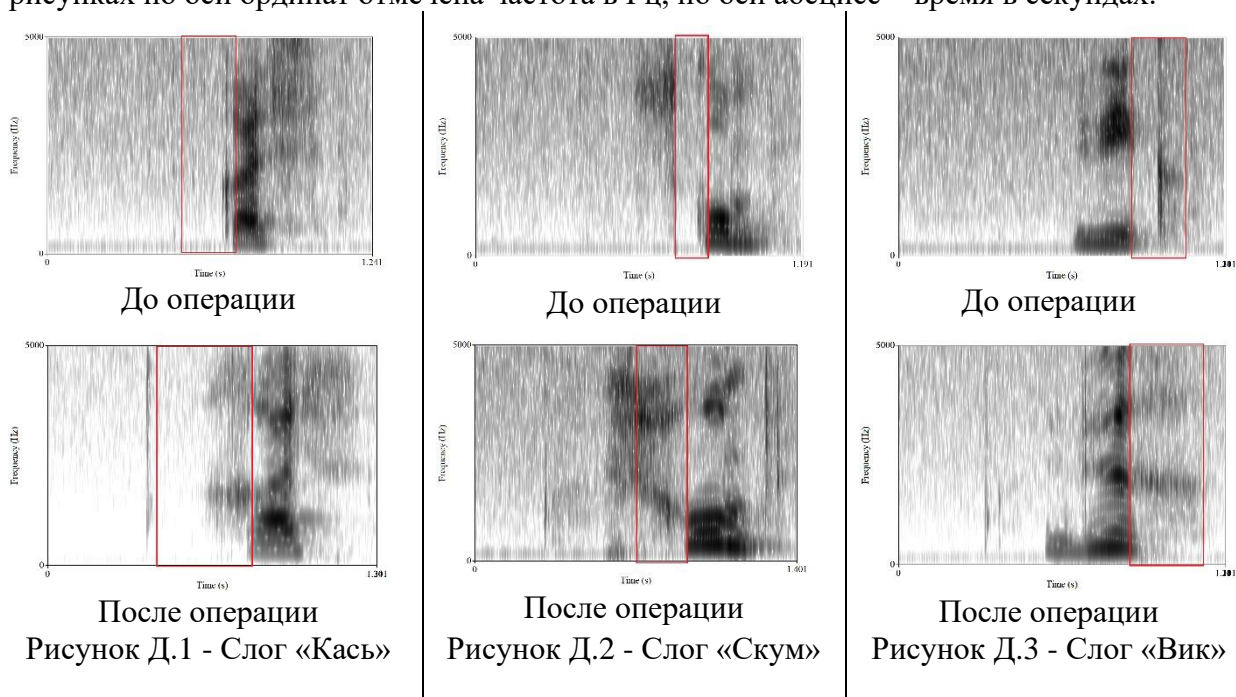
Diane M Port
ASA President



Vera Khokhlova
Chair, Committee on International Research and Education

Приложение Д – Спектрограммы сигналов

На рисунках Д.1 - Д.3 представлены спектрограммы Фурье звуков с указанием записанного слога и выделением фонем [к] и [к'] в составе слога до и после операции. На рисунках по оси ординат отмечена частота в Гц, по оси абсцисс – время в секундах.



Далее: на рисунках под цифрами 1 – слог, где изменяемая фонема в начале, 2 – фонема в середине, 3 – фонема в конце. Рисунки 1а-3а – слоги без изменения в произношении фонемы, рисунки 1б-3б – с изменением в произношении фонемы. На каждом рисунке по оси ординат – частота в Гц, по оси абсцисс – время в секундах.

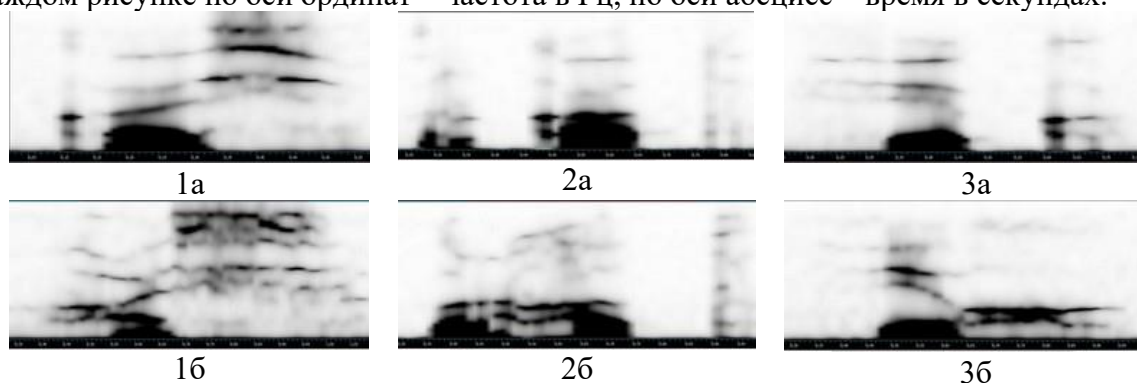


Рисунок Д.4 – спектрограммы КЛП записей слогов, содержащих фонему [к].

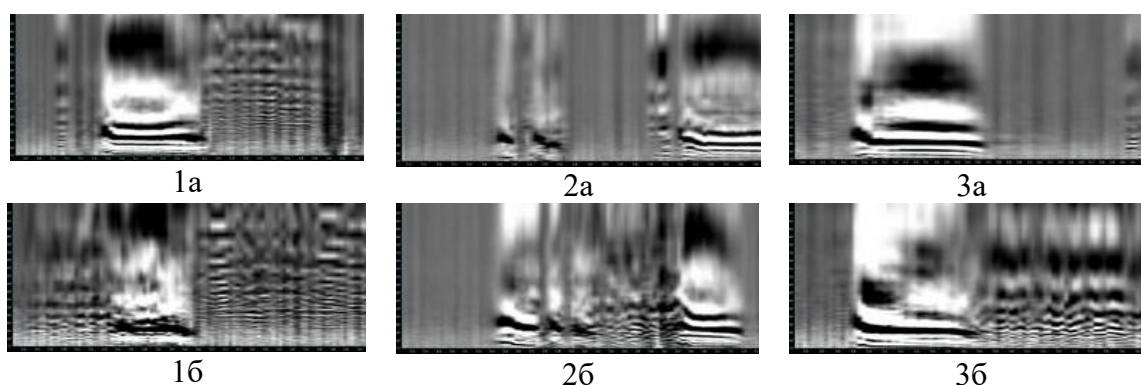


Рисунок Д.5 – Автокорреляции для записей слогов, содержащих фонему [к].

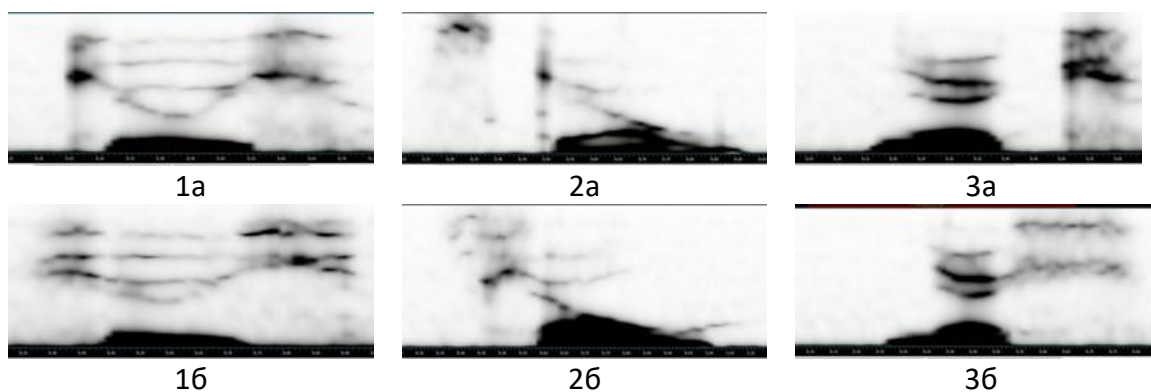


Рисунок Д.6 – Спектрограммы КЛП для записей слогов, содержащих фонему [к’].

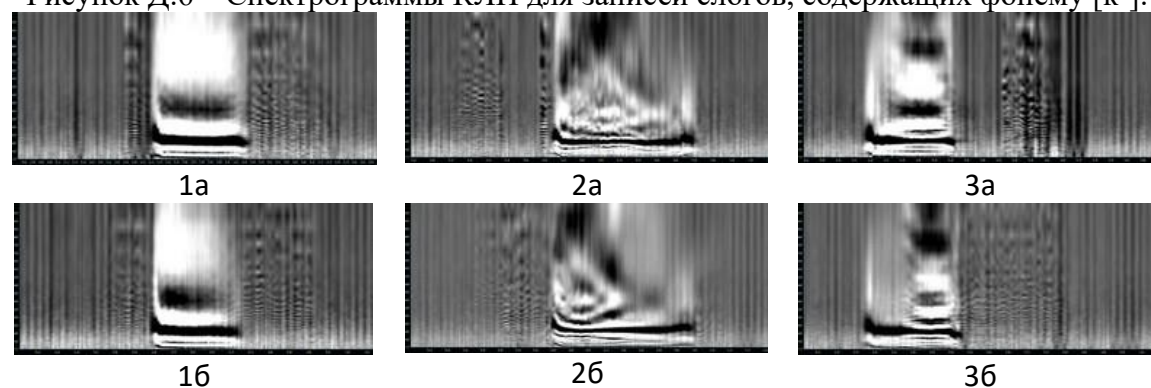


Рисунок Д.7 – Автокорреляция для записей слогов, содержащих фонему [к’].

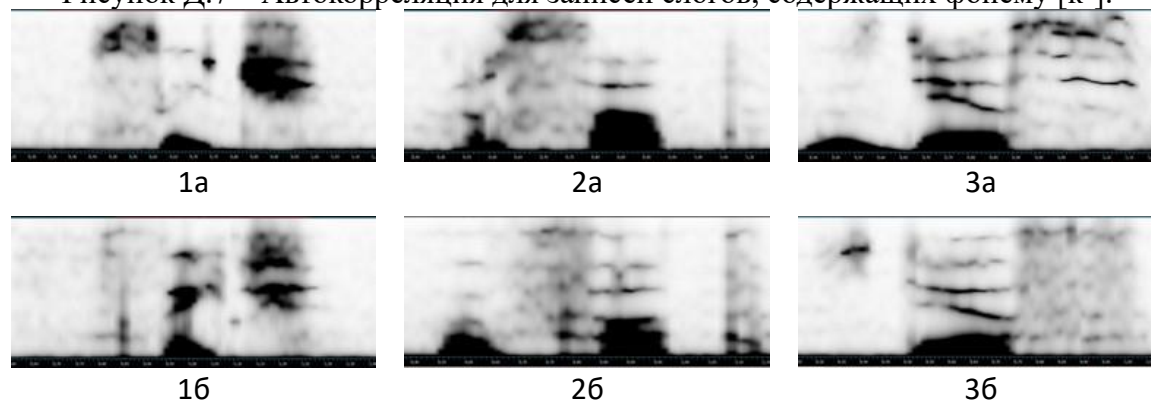


Рисунок Д.8 – Спектрограммы КЛП для записей слогов, содержащих фонему [с].

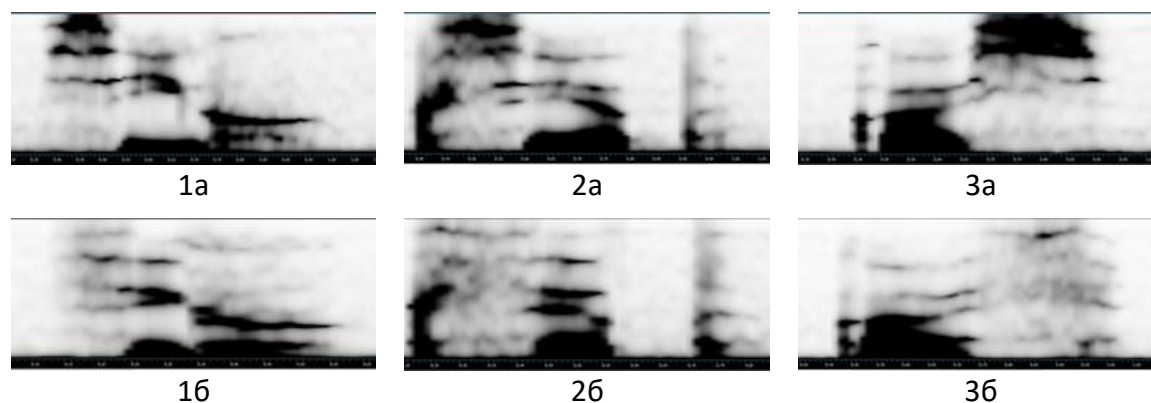


Рисунок Д.9 – Спектрограммы КЛП для записей слогов, содержащих фонему [с’].

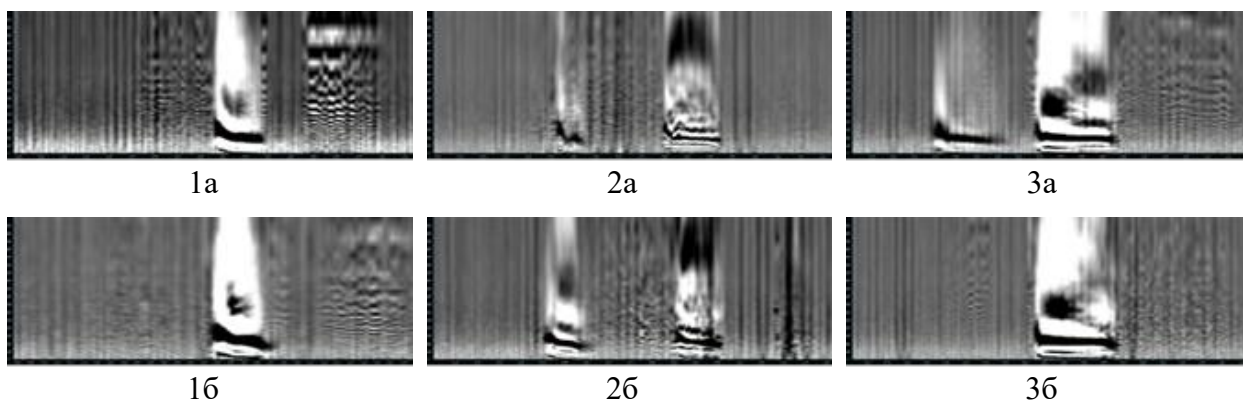


Рисунок 2.10 – Автокорреляции для записей слогов, содержащих фонему [с].

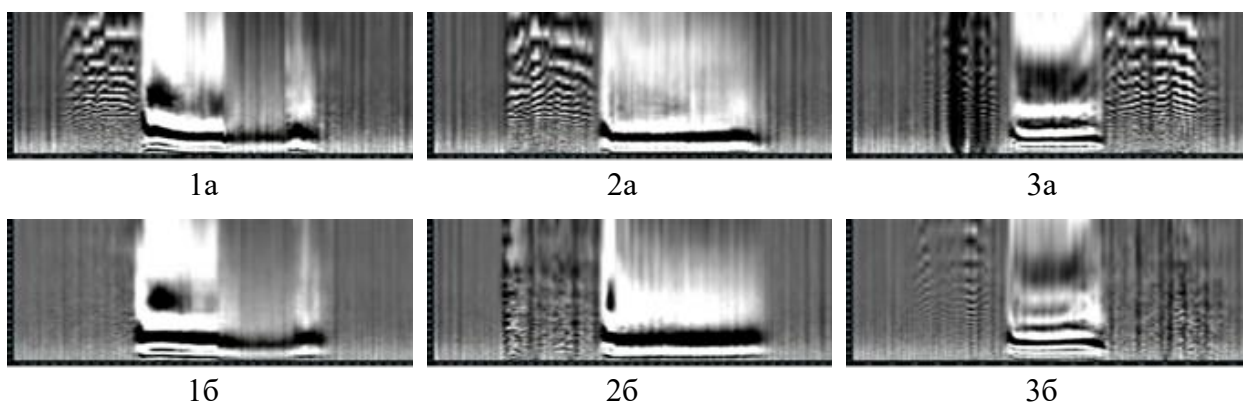


Рисунок 2.11 – Автокорреляции КЛП для записей слогов, содержащих фонему [с'].

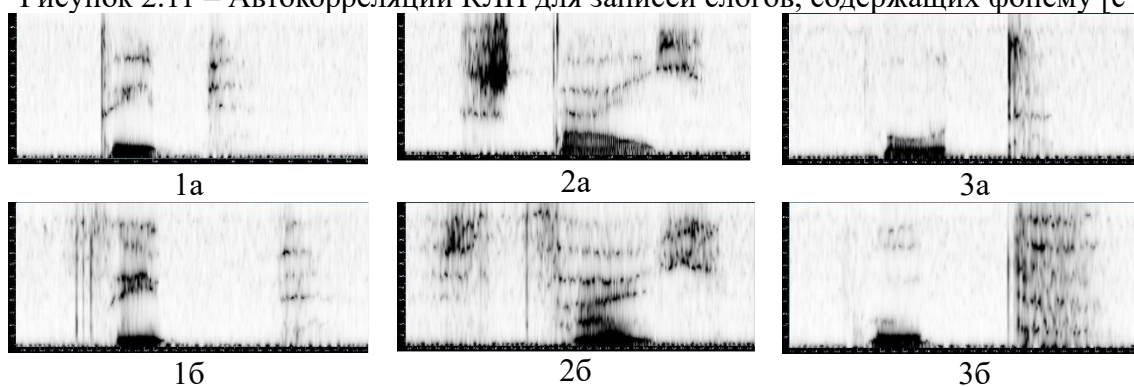


Рисунок 2.12 – Спектрограммы КЛП для записей слогов, содержащих фонему [т].

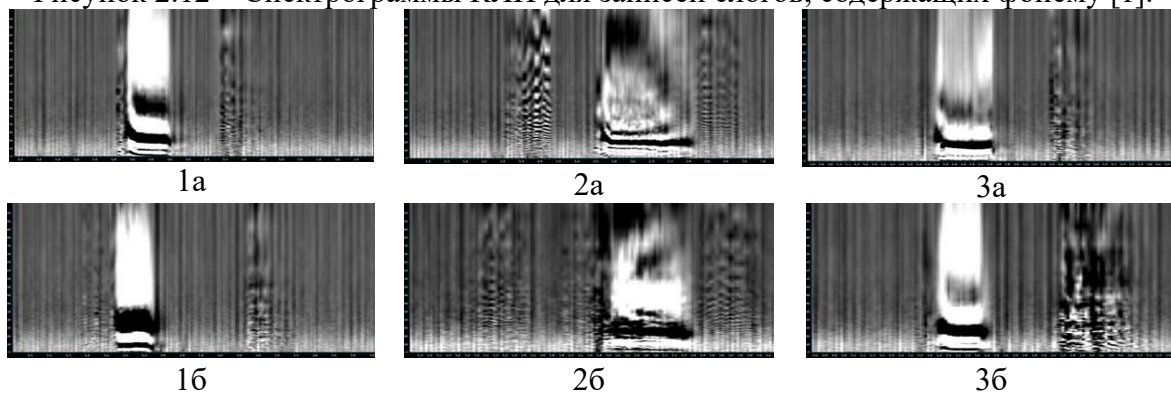


Рисунок 2.13 – Автокорреляции для записей слогов, содержащих фонему [т].

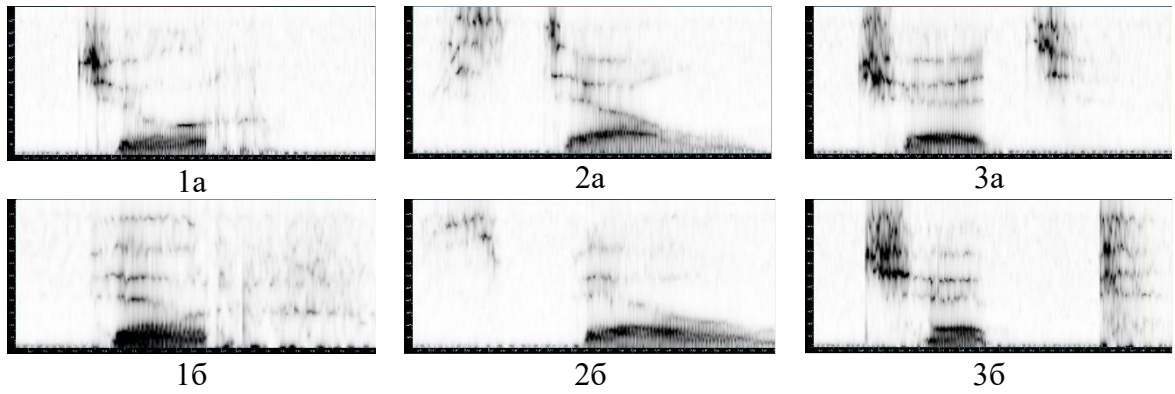


Рисунок Д.14 – Спектрограммы КЛП для записей слогов, содержащих фонему [тʰ].

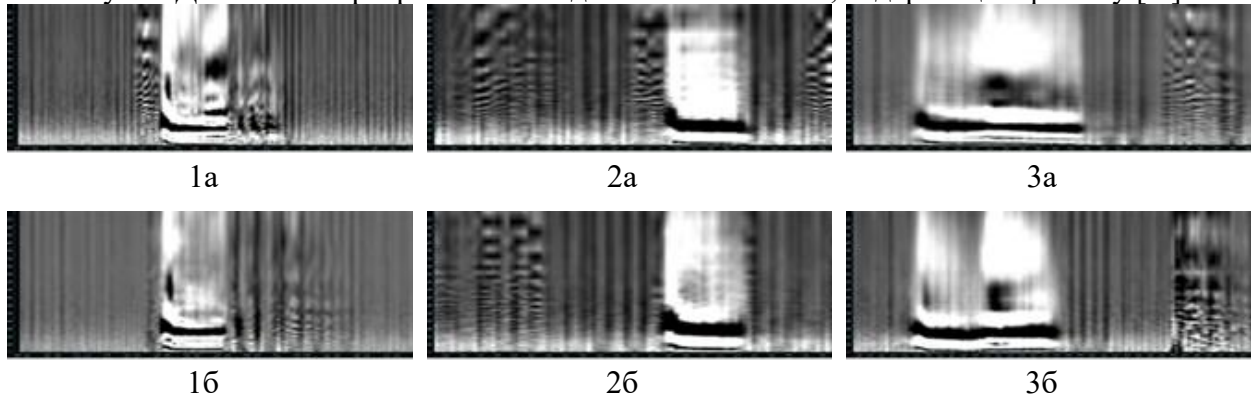


Рисунок Д.15 – Автокорреляции для записей слогов, содержащих фонему [тʰ].

Приложение Е – Матрицы расстояний между сигналами

Таблица Е.1 – Матрица расстояний для евклидова расстояния для всего слога «Фек» для полного спектра Фурье

Евклидово расстояние						
	1_1	1_2	1_3	1_4	1_5	
1_1						
1_2	342,2					
1_3	309,2	324,2				
1_4	312,4	302,2	272,6			Среднее значение
1_5	284,4	342,1	320,8	314,1		312,42
	2_1	2_2	2_3	2_4	2_5	
2_1						
2_2	335,3					
2_3	356,9	298,2				
2_4	365	330	312,9			Среднее значение
2_5	352	315,7	260	290,1		321,61
	1_1	1_2	1_3	1_4	1_5	
2_1	357	359,2	447	311,7	338,4	
2_2	381,6	359,4	429,8	369	358,4	
2_3	425,1	361,5	382,9	376,6	403,3	
2_4	366,9	408,4	460,4	358,2	360	Среднее значение
2_5	358,6	427,8	457,6	367	349,9	383,028

Здесь и далее в таблицах записи слога имеют следующее обозначение i_j , где:

- i – обозначение группы типа произношения слога, 1 если слог произнесен правильно, 2 если слог произнесен с искажениями.
- j – обозначение номера записи в группе от 1 до 5 (по 5 аудиозаписей с правильным и искаженным произношением).

Таблица Е.2 – Матрица расстояний для слога для полного спектра Фурье, вычисленных на основе метрики Чебышева

Чебышева						
	1_1	1_2	1_3	1_4	1_5	
1_1						
1_2	4,154					
1_3	4,078	3,732				
1_4	3,817	3,455	3,191			среднее значение
1_5	3,938	4,317	3,669	3,649		3,8

	2_1	2_2	2_3	2_4	2_5	
2_1						
2_2	4,412					
2_3	3,961	3,928				
2_4	3,376	3,62	3,98			среднее значение
2_5	3,526	3,815	4,085	3,547		3,825
	1_1	1_2	1_3	1_4	1_5	
2_1	4,354	4,287	3,92	3,148	4,365	
2_2	4,409	3,603	3,907	3,458	3,608	
2_3	4,159	3,476	4,263	3,835	4,25	
2_4	4,302	4,043	4,125	3,71	3,662	среднее значение
2_5	4,273	3,758	4,535	4,307	3,418	3,967

Таблица Е.3 – Матрица расстояний для слога для полного спектра Фурье, вычисленных на основе метрики Минковского ($p=3$)

Минковского						
	1_1	1_2	1_3	1_4	1_5	
1_1						
1_2	50,36					
1_3	45,68	47,25				
1_4	45,69	43,67	39,3			среднее значение
1_5	41,12	49,54	46,72	45,59		45,492
	2_1	2_2	2_3	2_4	2_5	
2_1						
2_2	48,65					
2_3	51,74	43,6				
2_4	53,2	48,95	46,41			среднее значение
2_5	50,72	46,22	38,38	42,49		47,036
	1_1	1_2	1_3	1_4	1_5	
2_1	52,09	50,67	63,02	44,7	49,32	
2_2	54,47	51,43	60,79	52,49	51,66	
2_3	59,99	51,44	55,55	53,59	57,27	
2_4	53,38	57,45	64,52	51,79	51,81	среднее значение
2_5	52,01	60,16	64,34	52,85	50,28	54,6828

Таблица Е.4 – Матрица расстояний для слога для полного спектра Фурье, вычисленных на основе метрики городских кварталов (Манхэттенское расстояние)

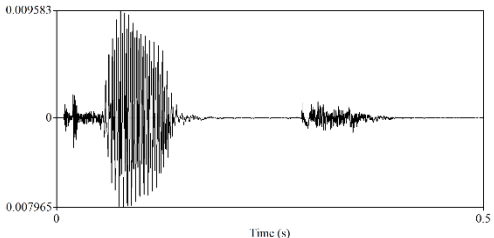
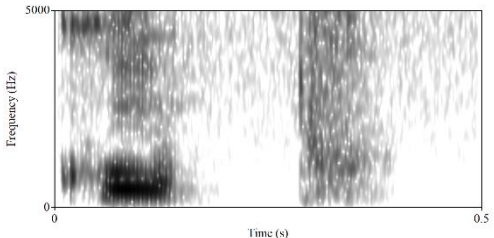
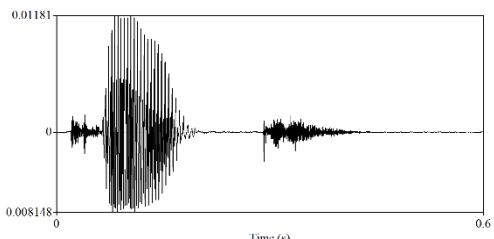
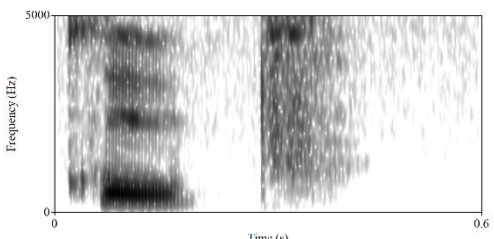
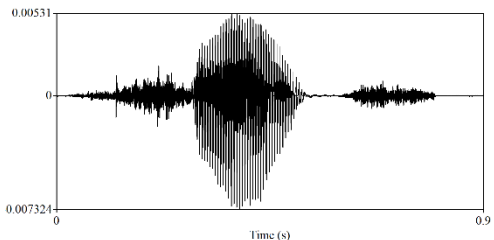
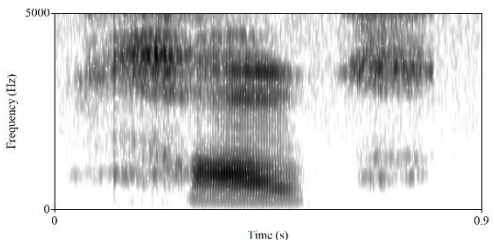
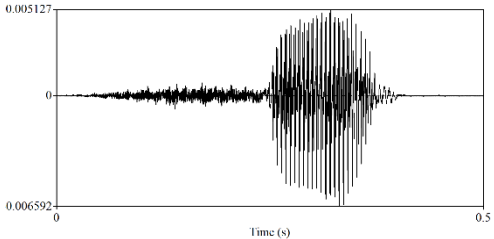
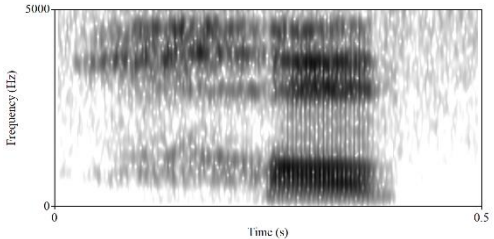
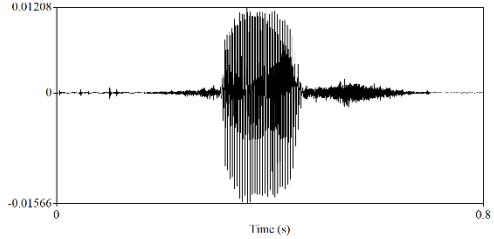
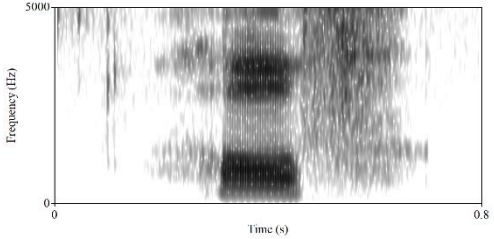
Манхэттенское						
	1_1	1_2	1_3	1_4	1_5	
1_1						
1_2	1,478E+05					
1_3	1,340E+05	1,420E+05				
1_4	1,378E+05	1,338E+05	1,217E+05			среднее
1_5	1,269E+05	1,533E+05	1,412E+05	1,394E+05		1,378E+05

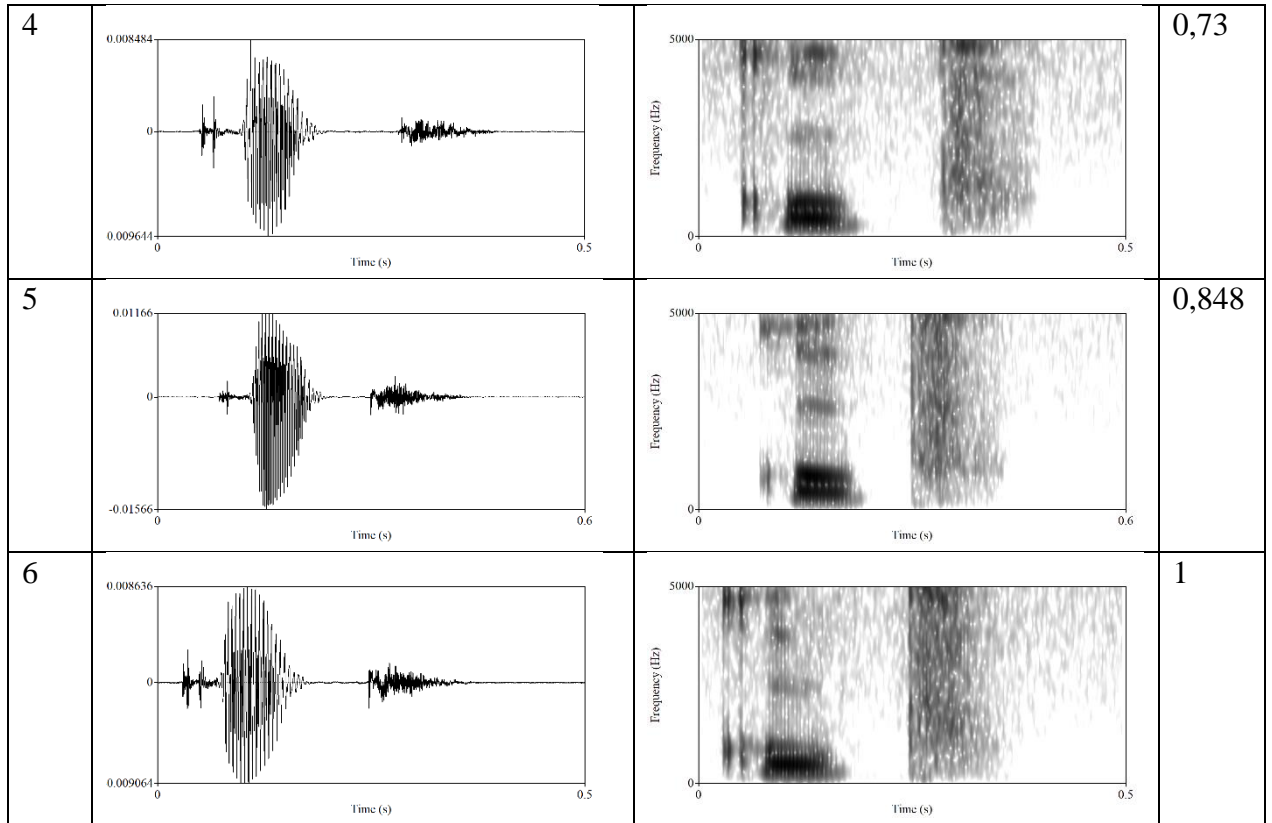
	2_1	2_2	2_3	2_4	2_5	
2_1						
2_2	1,471E+05					
2_3	1,574E+05	1,299E+05				
2_4	1,595E+05	1,418E+05	1,343E+05			среднее
2_5	1,566E+05	1,381E+05	1,125E+05	1,266E+05		1,404E+05
	1_1	1_2	1_3	1_4	1_5	
2_1	1,566E+05	1,641E+05	2,045E+05	1,399E+05	1,497E+05	
2_2	1,720E+05	1,617E+05	1,950E+05	1,672E+05	1,603E+05	
2_3	1,952E+05	1,633E+05	1,694E+05	1,707E+05	1,837E+05	
2_4	1,620E+05	1,875E+05	2,127E+05	1,597E+05	1,613E+05	среднее
2_5	1,587E+05	1,953E+05	2,093E+05	1,642E+05	1,579E+05	1,729E+05

Таблица Е.5 – Матрица расстояний для слога для полного спектра Фурье, вычисленных на основе метрики Махаланобиса

Махаланобиса						
	1_1	1_2	1_3	1_4	1_5	
1_1						
1_2	5,092E+05					
1_3	5,395E+05	5,383E+05				
1_4	5,591E+05	5,714E+05	6,131E+05			среднее значение
1_5	5,997E+05	5,544E+05	5,471E+05	5,647E+05		5,597E+05
	2_1	2_2	2_3	2_4	2_5	
2_1						
2_2	5,312E+05					
2_3	5,214E+05	5,556E+05				
2_4	5,048E+05	5,135E+05	5,261E+05			среднее значение
2_5	5,361E+05	5,439E+05	5,865E+05	5,655E+05		5,385E+05
	1_1	1_2	1_3	1_4	1_5	
2_1	5,153E+05	5,641E+05	5,078E+05	5,805E+05	5,405E+05	
2_2	5,300E+05	5,445E+05	5,068E+05	5,463E+05	5,375E+05	
2_3	5,259E+05	5,471E+05	5,071E+05	5,407E+05	5,292E+05	
2_4	5,162E+05	5,363E+05	5,136E+05	5,339E+05	5,393E+05	среднее значение
2_5	5,243E+05	5,154E+05	5,011E+05	5,324E+05	5,566E+05	5,317E+05

**Приложение Ж – Графическое представление аудиосигналов с
произношением слога «куп»**

Номер сеанса	Осциллограмма	Спектрограмма	Оценка
0-1			-
0-2			-
1			0,396
2			0,565
3			0,616



Приложение И – Декомпозиции процессов оценки

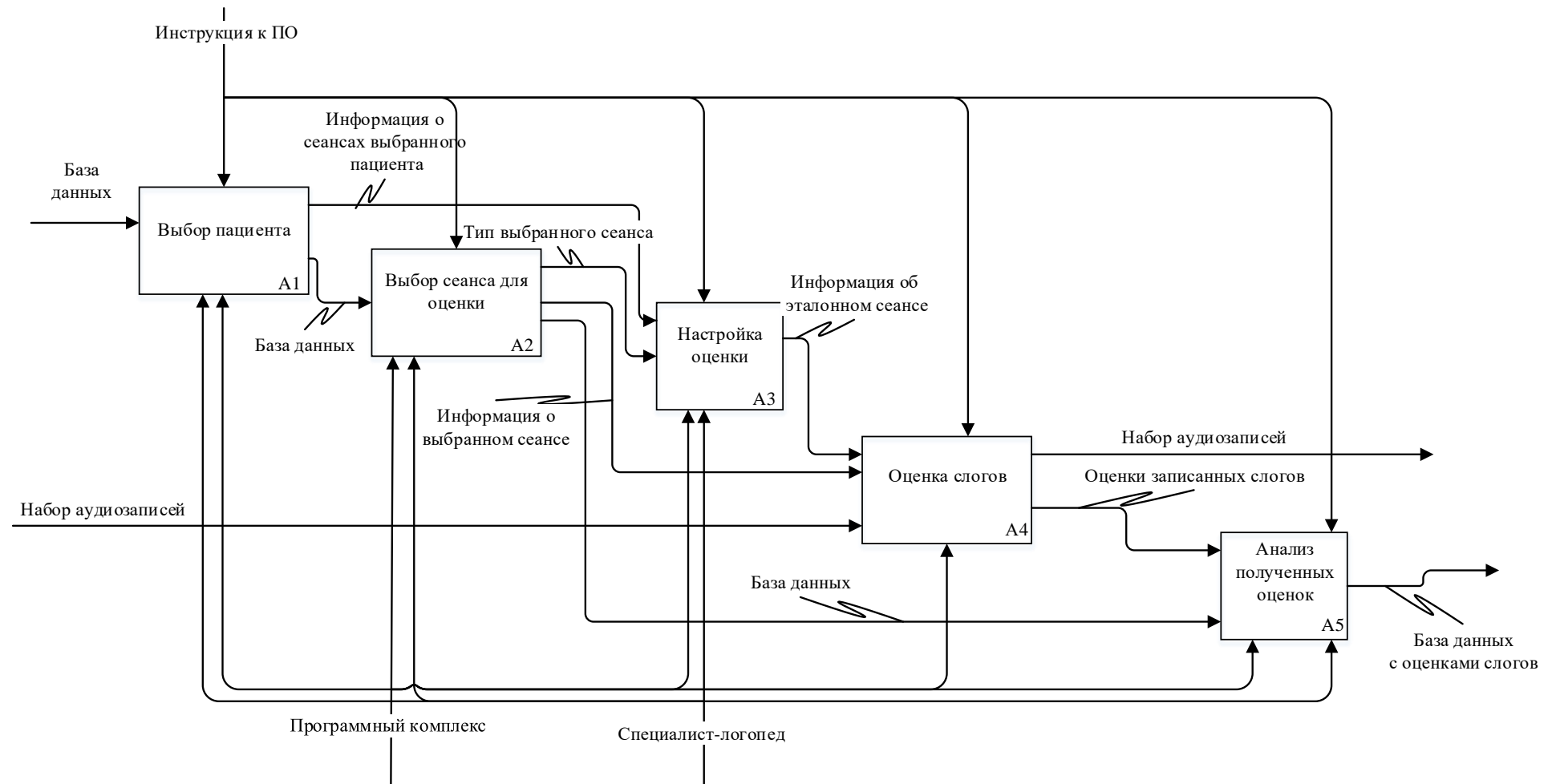


Рисунок И.1 – Декомпозиция процесса оценки существующего в базе данных сеанса произношения слогов

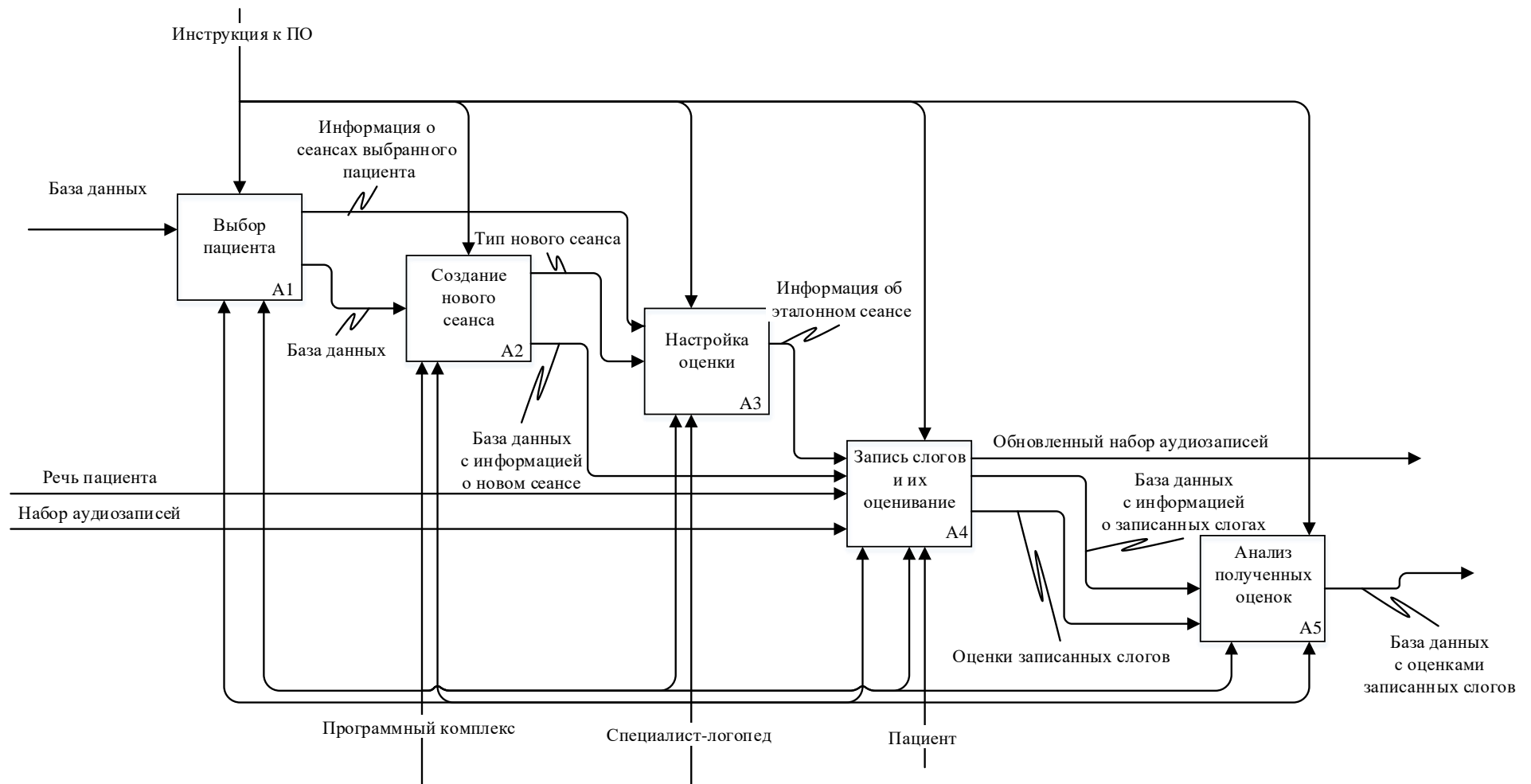


Рисунок И.2 – Декомпозиция процесса записи и оценки нового сеанса произношения слогов